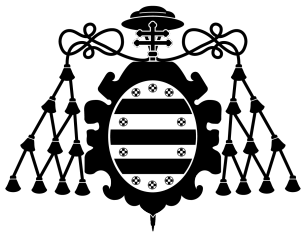


Universidad de Oviedo

Departamento de Ingeniería Eléctrica, Electrónica,
de Computadores y Sistemas

PhD Thesis

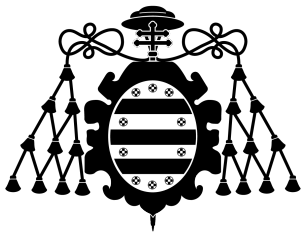
Monitoring and analysis of dynamic processes using visualization principles and techniques

Tesis Doctoral

Supervisión y análisis de procesos con dinámica mediante principios y técnicas de visualización

Daniel Pérez López

Febrero 2015



Universidad de Oviedo

Departamento de Ingeniería Eléctrica, Electrónica,
de Computadores y Sistemas

PHD THESIS

MONITORING AND ANALYSIS OF DYNAMIC PROCESSES USING VISUALIZATION PRINCIPLES AND TECHNIQUES

TESIS DOCTORAL

SUPERVISIÓN Y ANÁLISIS DE PROCESOS CON DINÁMICA MEDIANTE PRINCIPIOS Y TÉCNICAS DE VISUALIZACIÓN

Memoria presentada para la obtención del grado de Doctor por la Universidad de Oviedo

Autor: Daniel Pérez López
Director: Ignacio Díaz Blanco

Gijón, Febrero 2015

RESUMEN

La capacidad actual de adquisición y almacenamiento de datos hace que sea posible disponer de medidas de muchos de los sistemas que tenemos a nuestro alcance. Los nuevos sensores permiten medir una multitud de variables de cualquier sistema que nos interese. Los procesos que pueden resultar estimulantes son aquellos sistemas complejos, con dinámicas no lineales, como por ejemplo procesos industriales, que incluyan varios subsistemas (eléctrico, mecánico, etc) o consumos eléctricos en edificios, donde intervienen varios factores (térmicos, temporales, etc). Esos datos recogidos son una potencial fuente de información sobre el sistema y podrían utilizarse para estudiarlo en detalle.

Existen técnicas automáticas que analizan los datos y pueden resultar útiles para varias aplicaciones. Por ejemplo, crean modelos de los sistemas que describen, encuentran patrones y extraen relaciones entre las variables, o predicen comportamientos futuros del sistema. Esta información sería interesante presentarla de manera eficiente. La visualización transmite esta información a través de representaciones gráficas de los datos, en donde se amplifica la cognición del ser humano no sólo para apoyar el razonamiento sobre esos datos (en el desarrollo y evaluación de hipótesis) sino también para facilitar la comunicación de la información.

Durante muchos años se han desarrollado algoritmos de reducción de la dimensión, los cuales permiten obtener una visualización basada en una proyección de los datos, donde las muestras (representadas por puntos) son distribuidas por la similitud entre ellas, es decir, puntos cercanos representan muestras similares, y viceversa. Esta tesis se centra en este tipo de técnicas. A lo largo de su desarrollo se estudiarán los distintos tipos que existen y las formas para evaluar las proyecciones resultantes. Se han realizado proyecciones de sistemas complejos con diferentes dinámicas. Por ejemplo, proyecciones para los estados dinámicos de un proceso de laminación, que representan un mapa visual de las condiciones de funcionamiento. En estos mapas se diferencia un fallo denominado *chatter*, que consiste en una potente vibración durante la laminación y pone en riesgo el proceso productivo. También se han realizado estudios para el análisis de consumos eléctricos en edi-

ficios universitarios. Por ejemplo, la obtención de una proyección de datos de todo un año, representando con el color magnitudes como los precios de la tarifa eléctrica o la energía activa consumida.

La combinación de la capacidad de cálculo de los ordenadores con las ventajas de la visualización, a través de herramientas interactivas, permite introducir al humano en el proceso de análisis de datos. Este enfoque se ha aplicado en el desarrollo de un prototipo de interfaz web para la exploración de datos de consumos eléctricos, en el cual el usuario, a través de diferentes vistas, puede navegar entre diferentes representaciones de la información organizadas por temporalidades o similitudes. Con un propósito similar, también se ha estudiado la introducción de conocimiento previo del analista (como información de clases) en el proceso de reducción de la dimensión. Para ello se transforman los datos de partida mediante una extensión de características con la información de grupos. Esta información es introducida por el usuario de forma interactiva, lo cual modifica la proyección original de los datos revelando dicha información. Se han realizado varios experimentos con el método propuesto en diferentes escenarios de aplicación. Además, las proyecciones resultantes son evaluadas con medidas cuantitativas para ayudar al analista a decidir el grado de transformación en la proyección final, teniendo en cuenta su mejora visual y la preservación de la estructura original de los datos.

ABSTRACT

The current capacity of data acquisition and storage facilitates measurements of many systems around us. New sensors are capable to measure a multitude of variables of any system. The challenging processes are those complex systems with non-linear dynamics, such as industrial processes that include various subsystems (electric, mechanic, etc) or electrical consumptions in buildings, where several factors are involved (thermal, temporal, etc). The recorded data are a potential source of information about the system and can be used to perform a detailed study of this system.

There are automatic techniques for data analysis that can be useful for several applications. For instance, they create models of the described systems, find patterns and extract relationships between variables, or predict future behaviours of the system. This information would be interesting if it were presented efficiently. Data visualization conveys information through graphical representations, where human cognition is amplified not only to support reasoning about the data (in the development and evaluation of hypotheses) but also to facilitate the communication of the information.

Dimensionality reduction algorithms have been developed for many years. They allow obtaining a visualization based on data projection where samples (represented by points) are distributed according to their similarity, that is, close points represent similar samples, and vice versa. This thesis is focused on this type of techniques. The different kinds of the existing techniques are studied and also the ways to evaluate the resulting projections. Visual maps of complex systems with different dynamics have been obtained. For example, projections for dynamic states of a cold rolling mill process, where a visual map represents different operating conditions. In these maps a fault called *chatter*, that produces a powerful vibration in the rolling mill and can risk the process is detected. Besides, studies have been made for electric consumptions analysis in university buildings. For example, a projection of one-year data, representing with color the prices of the electric bill or consumed active energy.

The combination of the computation capabilities of the computers with the advantages of the visualization, through interactive visual interfaces, allows introducing the human in the loop of data analysis. This approach has been applied to the development of a prototype web interface for the exploration of electric consumption data, where the user can navigate through different views that represent the information for several temporalities or similarities. Furthermore, the introduction of analyst's prior knowledge (like classes information) into dimension reduction process has been studied. Input data are transformed by means of a feature extension using class information. This information is introduced interactively, which modifies the original data projection revealing that information. Various experiments have been performed using the proposed method in different scenarios. Besides, the resulting projections are evaluated with quantitative measures to support the analyst in deciding the degree of transformation of the final projection, taking into account the visual improvements and the structure preservation of the original data.

ÍNDICE GENERAL

1 INTRODUCCIÓN

En este primer capítulo se presentan el entorno, la motivación y los objetivos básicos marcados en la presente Tesis. Finalmente se detalla la estructura del documento.

1.1 INTRODUCCIÓN

Durante los últimos años la generación de datos ha experimentado un crecimiento extraordinario. Cada día se crean grandes cantidades de datos de muchas formas distintas. Algunos ejemplos se pueden apreciar con los propios datos geográficos que genera nuestro teléfono móvil que nos localizan en cualquier instante o los medios que poseen los actuales deportistas para registrar el rendimiento durante sus entrenamientos, impensables para sus homólogos de hace 20 años.

Un término utilizado para considerar estos aspectos es "*big data*", el cual se ha popularizado recientemente a raíz del informe McKinsey [131]. Para una definición objetiva del término, es habitual considerar las 3 *v*'s [135], descritas originalmente en 2001 por Douglas Laney [115]. El *volumen* se refiere al tamaño de los datos; la *variedad* se refiere a los diversos formatos de los datos, que con frecuencia proceden de diferentes fuentes; y la *velocidad* se refiere al ritmo al que son generados. Más tarde, IBM añadió una cuarta *v*, la *veracidad* para considerar la incertidumbre generada en el conjunto total de los datos.

Dichos aspectos están presentes en escenarios tan diversos como pueden ser el sector energético o en los procesos industriales, donde los requisitos de calidad del producto final son tan exigentes que obligan a continuas mejoras de los procesos. Los equipos actuales de adquisición de datos permiten un almacenamiento masivo, y en tiempo real, de casi todas las variables que se deseen, y para todos los procesos que tienen lugar en una instalación industrial. En el caso del sector energético, los distintos tipos de sensores instalados recogen datos de distintas fuentes, prácticamente a la misma velocidad a la que se generan. Esto se da en el caso de la medición de los consumos de energía que se producen dentro de edificios, en donde intervienen factores no sólo eléctricos, sino también térmicos (calefacción) o auxiliares (seguridad), en principio independientes

pero que afectan directamente a la demanda final. Esta información también se puede fusionar con predicciones meteorológicas que ayuden a valorar y predecir el consumo eléctrico.

Ambos casos los conforman sistemas complejos, con dinámicas no lineales, donde se pueden producir fenómenos que estén conectados entre sí o que procedan de un origen desconocido, como por ejemplo algún tipo de fallos que se puedan producir en un proceso industrial.

Los ordenadores poseen una gran potencia de cálculo, muy adecuada para el análisis automático por medio de algoritmos, que utilizando los datos, proporcionan información sobre el sistema que describen; por ejemplo qué tipo de relaciones existen entre algunos factores del sistema, crear modelos que definan de manera general el problema estudiado, o incluso predecir comportamientos futuros. Esta capacidad real para el análisis de todos los datos disponibles supone una oportunidad para utilizarlos como fuente de conocimiento nuevo, no sólo en el mundo empresarial (*business intelligence*), donde puede proporcionar una ventaja competitiva, sino también en el sector público, donde cada vez más los gobiernos proporcionan datos (*open data*) cuyo análisis puede mejorar los recursos y por tanto la calidad de vida de la sociedad. Por ejemplo, esto permitiría mejorar la productividad de una planta industrial, estudiar la eficiencia de algún servicio público como el transporte, o definir una política de actuación óptima para el consumo energético en edificios públicos.

Sin embargo, la sobrecarga de información puede no resultar tan beneficiosa, lo cual puede contradecir la idea de que a mayor cantidad de información mejores condiciones tendremos para encontrarnos más próximos a solucionar un problema. Aunque la cantidad de información esté aumentando, la información útil no aumenta tan rápido. Hay mayor número de datos que analizar pero solo una cantidad constante que determina la solución del problema analizado. Lo que aumenta más rápidamente son otros factores ajenos al problema, que pueden resultar incluso perjudiciales para su análisis, tales como variables no relacionadas con el problema, ruido en las medidas, etc. Esto afecta directamente a las interpretaciones que se realicen de los resultados.

Aunque existen muchas maneras de analizar datos de manera automática, para que el análisis se pueda llevar a cabo adecuadamente son necesarios algunos requisitos que determinen el problema correctamente, como por ejemplo en la organización o su procesamiento, que estén estructurados claramente, etc. Sin embargo, estos requisitos no están siempre bien definidos, lo que hace que un

análisis automático funcione correctamente en un limitado número de aplicaciones.

En cambio, nuestras mentes son rápidas, están diseñadas para detectar patrones de manera efectiva, y conectar ideas con la experiencia. Nuestro entendimiento se basa principalmente en el sentido de la vista, en el que un diagrama es más efectivo que una descripción textual para resolver problemas [116]. Una imagen en la retina se analiza mediante procesos relacionados con la atención, en los que se extraen patrones para ayudarnos a entender cualquier tarea que estemos haciendo [203]. Para un proceso de exploración de datos es interesante combinar la habilidad para el análisis y abstracción que posee el conocimiento humano con la capacidad de almacenamiento y procesamiento de los ordenadores actuales [105].

Representar la información multidimensional en un plano visual no es una tarea trivial. El proceso de diseño requiere considerar factores que dependen de aspectos como la percepción visual, no sólo para una correcta codificación de la información que queremos mostrar en elementos gráficos, sino también de su interpretación. Si el proceso de decodificar esa información falla, la visualización no tiene sentido.

Un enfoque visual proporciona una serie de ventajas, como el manejo de distintos tipos de datos heterogéneos mediante diferentes representaciones; un rápido e intuitivo análisis por medio del cual se pueden obtener mejores resultados que en aquellas ocasiones donde la utilización de algoritmos matemáticos complejos no resulta satisfactoria; la visualización de los datos en un contexto permite conectar con otras fuentes de conocimiento previo, estableciendo enlaces que facilitan la obtención de conocimiento de los datos, generar nuevas hipótesis para su verificación, e inferir conclusiones. Esto permite aumentar la capacidad para explotar dicho conocimiento previo de la persona dotándolo de un elemento clave para el análisis de datos y la interpretación de resultados.

Estas ventajas que ofrece la visualización pueden utilizarse en el contexto de los procesos industriales. El estado de un proceso se puede caracterizar a través de sus variables, y un conjunto de estados similares definen una determinada condición dinámica del proceso. Esta información presentada visualmente no sólo sirve para supervisar el proceso, sino también, por ejemplo, para explotar el conocimiento de expertos o la detección temprana de nuevos fallos. También se pueden aplicar a los consumos energéticos, donde una representación visual de la demanda eléctrica proporcionaría al usuario información de lo que está consumiendo, generándole

una conciencia energética que le permite evaluar la gestión actual de la energía y sugerir estrategias para mejorarla.

Sin embargo, aunque el cerebro humano es extraordinario en procesar información, tenemos que ser selectivos en cuanto a la cantidad de datos que elegimos para recordar, dados los volúmenes que se generan diariamente. Con estas grandes cantidades de datos, puede ocurrir que su visualización simultánea no sea efectiva; aunque tengamos mayores pantallas no podremos representar toda la información a la vez. Otro problema sobre nuestro instinto de percepción visual es que puede llevarnos a ver patrones donde en realidad no los hay, por lo que resultaría necesario comprobar los resultados para que las interpretaciones fueran lo más precisas posibles.

La combinación de una adecuada visualización basada en principios de percepción con algoritmos avanzados de tratamiento de datos [47, 105] y mecanismos de interacción que se pueden desarrollar [173, 87] facilita el razonamiento analítico sobre el sistema descrito por los datos. Su desarrollo mediante el uso de interfaces visuales define el campo denominado analítica visual (*visual analytics*, VA) [106, 183]. Aunque en algunas ocasiones resulta difícil decidir las tareas automáticas que debe realizar el ordenador o qué tareas interactivas se permiten al usuario, lo cierto es que el desarrollo de sistemas de analítica visual está creciendo no sólo en instituciones académicas (donde se originó) sino también en pequeñas empresas (como por ejemplo en *spin-offs* de grupos de investigación) realizando herramientas específicas para algún campo de aplicación determinado. Esto ha permitido el desarrollo de aplicaciones no sólo de código abierto [80], como por ejemplo *Gephi*, sino nuevos productos software como *Tableau* y que también multinacionales desarrollen nuevos componentes para el análisis eficiente de datos, como por ejemplo *General Electric* (http://visualization.geblogs.com/). En el trabajo publicado en [213] se realiza una comparación entre una selección de sistemas comerciales, que poseen varias funcionalidades de analítica visual, mediante la evaluación de su funcionalidad y rendimiento.

Este enfoque puede aplicarse a los datos procedentes de procesos industriales y/o tecnológicos, en los cuales se precisa incrementar su rendimiento y calidad. Mediante una representación gráfica interactiva se puede realizar una supervisión visual del proceso, con una mayor cantidad de datos. Estos datos poseen un número elevado de variables implicadas en el proceso y varios estados posibles de funcionamiento, que pueden permanecer ocultos, contenidos en un espacio de alta dimensión. Por medio de algoritmos automá-

ticos, se puede extraer información de esos datos y representarla de manera que revele visualmente conocimiento sobre un problema determinado o el rendimiento general del proceso. Ejemplo de esto son las técnicas de reducción de la dimensión, que permiten calcular una proyección de los datos en un mapa visual. Estas técnicas, combinadas con una adecuada visualización mediante el uso de interfaces interactivos, permiten obtener información extraída de los datos de una manera efectiva e identificarla rápidamente. Con esto, el usuario sería parte del proceso de análisis, y la exploración de los datos le ayudaría a confirmar hipótesis y a tomar decisiones sobre el sistema estudiado.

Con todo esto, el problema que se plantea es la posibilidad de estudiar procesos de naturaleza compleja por medio de un análisis intuitivo de los datos que los describen. Este problema abre una serie de preguntas como cuáles son los métodos más adecuados para extraer información de los datos útil para la tarea que se esté realizando, qué técnicas de visualización pueden representar esa información de la manera más efectiva para dicha tarea, o qué mecanismos se pueden combinar para proporcionar al usuario más control sobre el análisis de los datos, de manera que mejore su exploración.

La exploración y presentación de los datos de manera visual permite al ser humano la utilización de sus capacidades cognitivas para un análisis eficiente. Esta integración entre la extracción de información y su interpretación de manera visual con capacidades interactivas abren interesantes oportunidades de investigación que constituyen una de las principales motivaciones de esta tesis.

1.2 OBJETIVOS

A raíz del problema planteado, los principales objetivos tratados en esta tesis se pueden resumir en los siguientes puntos:

- Revisión de algoritmos de aprendizaje automático de datos, susceptibles de ser utilizados en herramientas de analítica visual, como por ejemplo, técnicas de reducción de la dimensión con las que se pueden obtener proyecciones de los datos. Elaboración del estado del arte de métodos de análisis para extraer información de los datos y crear modelos útiles del proceso. En concreto, dentro del ámbito de los procesos industriales y en aplicaciones de eficiencia energética.

- Revisión de los fundamentos de visualización, que determinen las maneras más intuitivas de representar la información,

así como las técnicas para visualizar datos multidimensionales. Estudio de mecanismos de interacción que aporten funcionalidades a las interfaces visuales y permitan al usuario una mejor exploración de los datos.

- El desarrollo y aplicación de métodos de analítica visual para el modelado y supervisión de procesos complejos. El objeto es incrementar la eficiencia de sistemas mediante un enfoque de análisis visual. Estos métodos incluyen técnicas de reducción de la dimensión que revelen condiciones de funcionamiento del proceso estudiado, así como métodos de clasificación de los datos de manera que permitan identificar visualmente esa información.

- Estudio y desarrollo de métodos que aborden algunas de las limitaciones que poseen las técnicas de proyección estudiadas, teniendo en cuenta su coste computacional. Ejemplos de estas limitaciones son el número de puntos a proyectar o la posibilidad de proyectar nuevos puntos sobre una proyección previamente calculada, como realizan las técnicas *out of sample*.

- Desarrollo de herramientas interactivas, que mejoren la exploración visual de los datos, aportando al usuario más control sobre las tareas de análisis. Estudio de métodos de transformación de los datos que permitan la incorporación en las proyecciones de conocimiento previo, como por ejemplo información de clases. Además, la aplicación de medidas cuantitativas para evaluar la calidad de esas proyecciones de manera que proporcionen al usuario más información para sus interpretaciones.

1.3 ESTRUCTURA DEL DOCUMENTO

El documento está dividido en varios capítulos que agrupan distintas facetas de la investigación. A continuación se indican estos capítulos con una breve descripción de los contenidos que incluyen.

En el capítulo 2, se revisan fundamentos de visualización, presentando variables visuales y algunas formas adecuadas para codificar la información de los datos en una representación visual. Además, se presentan algunos ejemplos representativos de visualizaciones de datos de varias dimensiones.

En el capítulo 3, se revisan varios métodos automáticos de análisis de datos, prestando más atención a las técnicas de reducción

de la dimensión, que proyectan datos multidimensionales en un espacio visual. Finalmente se estudia la evaluación de la calidad de las proyecciones resultantes por medio de medidas cuantitativas.

En el capítulo 4, se describen varios procedimientos desarrollados para el análisis visual de procesos y sistemas con comportamientos no lineales. Los métodos son aplicados a datos procedentes de casos reales, como el estudio de un fallo en un tren de laminación y la exploración de consumos eléctricos en edificios.

En el capítulo 5, se presenta una técnica interactiva, que transforma los datos modificando las proyecciones, de manera que permite introducir información de clases en la proyección para mejorar la interpretación y el conocimiento de los datos.

En el capítulo 6, se resumen las conclusiones finales de las investigaciones y se describe el trabajo futuro dentro de las posibles líneas de investigación abiertas.

2

PRINCIPIOS Y TÉCNICAS DE VISUALIZACIÓN

En este capítulo se revisan varios fundamentos desarrollados en el área de la visualización como pueden ser unas guías para el diseño visual o algunos conceptos de percepción visual. Aunque un repaso exhaustivo está fuera del alcance de esta tesis, también se presentan algunas técnicas que sirven como ejemplos representativos para una efectiva visualización de datos.

2.1 INTRODUCCIÓN

La tecnología actual permite la creación y almacenamiento de grandes cantidades de datos. Fenómenos como el denominado *big data* [135] demuestran mediante la velocidad de creación, el volumen y la variedad de esos datos una realidad presente en nuestros días, siendo además una gran oportunidad como fuente para el descubrimiento de conocimiento nuevo y apoyo a tomar mejores decisiones. Su manejo, transformación en información y visualización para hacerlo útil a la gente suponen un reto.

La visualización expresa información contenida en los datos a través de representaciones gráficas. Sirve para registrar visualmente la información y analizarla mediante la exploración de esos datos [104]. Esto permite desarrollar hipótesis, encontrar patrones o descubrir errores. También es muy útil en la comunicación a los demás; compartir ideas o contar historias se han realizado con frecuencia mediante el apoyo de una explicación visual [169, 186]. Para ello es recomendable conocer la estructura de los datos para representarlos de la mejor forma posible, además de los posibles medios para una exploración efectiva. En este sentido, el diseño es un proceso esencial en el que no sólo intervienen aspectos subjetivos sino también elementos cognitivos. Aunque no es posible mostrar todos los datos disponibles al mismo tiempo, el ser humano posee potentes cualidades visuales, como una amplia percepción visual con un alto ancho de banda, que a través de un adecuado diseño, pueden aprovecharse de una manera eficiente para el entendimiento de los datos.

Uno de los mejores ejemplos para ilustrar que los gráficos pueden revelar aspectos interesantes de los datos y que incluso pueden ser más precisos que cálculos estadísticos convencionales, es el lla-

Cuarteto de Anscombe

	I		II		III		IV	
	$\mathbf{x}_1$	$\mathbf{y}_1$	$\mathbf{x}_2$	$\mathbf{y}_2$	$\mathbf{x}_3$	$\mathbf{y}_3$	$\mathbf{x}_4$	$\mathbf{y}_4$
	10.0	8,04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6,95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
media	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
var.	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
corr.	0.82		0.82		0.82		0.82	

Tabla 1.: Conjuntos de datos que constituyen el cuarteto de Anscombe.

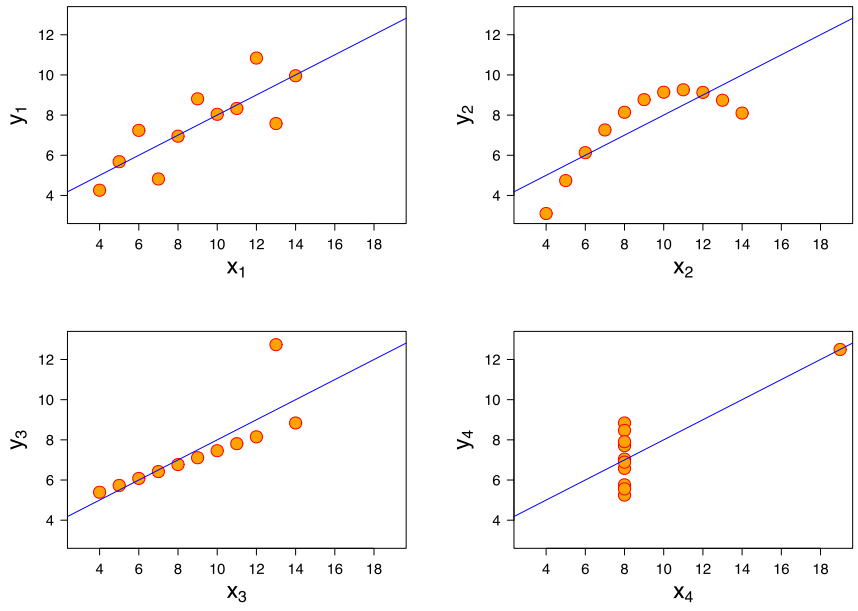


Figura 2.1.: Cuarteto de Anscombe son cuatro conjuntos de datos con idénticas propiedades estadísticas. Sin embargo, la inspección visual muestra diferentes estructuras.

mado *cuarteto de Anscombe*. Consta de cuatro conjuntos de datos, de dos variables, que tienen idénticos valores de medias, varianzas, correlación entre variables y modelo de regresión lineal, entre otros. En la tabla 1 se pueden ver los valores que forman estos conjuntos de datos. Sin embargo, una inspección visual de los cuatro conjuntos de datos por medio de sus representaciones gráficas (ver Fig. 2.1) muestran diferentes estructuras. Por ejemplo, se pueden observar valores atípicos (*outliers*) en los diagramas de la parte inferior de la Fig. 2.1; o también un patrón no lineal entre las dos variables, en la parte superior derecha de misma figura.

Son muchos los trabajos que han establecido las bases de este reciente campo [33, 34, 42, 128] y que, a lo largo de la historia [71, 206], han demostrado los beneficios de la visualización [62, 193], lo cual ha inspirado a numerosos investigadores a desarrollar nuevas técnicas [84] para representar la información. Estas técnicas sirven como ejemplos para entender los conceptos que hay detrás de la visualización de datos multivariables, mediante las soluciones, propuestas por otros, a problemas reales de diseño.

2.2 PRINCIPIOS DE DISEÑO

Uno de los aspectos más importantes en la visualización de la información es el diseño, es decir, la elección de la opción más adecuada entre todo el abanico de posibilidades para representar un conjunto de datos.

Edward R. Tufte expone en sus libros [187, 185, 186] varios fundamentos de diseño para la representación visual de la información. Mediante ejemplos se exponen varios principios para un adecuado diseño de gráficos estadísticos. Tufte es considerado un pionero en el diseño de visualizaciones de datos y sus principios son seguidos actualmente por muchos expertos. Introduce la *excelencia gráfica* como la comunicación de ideas complejas con claridad, precisión y eficiencia. Un gráfico que contiene esta excelencia aporta el mayor número de ideas en el más corto periodo de tiempo con el menor uso de tinta y espacio. Además, para llevar a cabo la excelencia gráfica es importante mostrar los datos tan directamente como sea posible, es decir, no mentir con la representación. Existen muchos casos en los que el diseño del gráfico contiene artefactos que no representan correctamente los valores de los datos. Un ejemplo es la representación de una variable unidimensional en más dimensiones mediante áreas o volúmenes, o el uso inadecuado de perspectivas que distorsionan la percepción del tamaño de los elementos gráfi-

cos. Estos errores pueden medirse cuantitativamente con el *factor mentira* definido de la siguiente manera:

$$\text{Factor mentira} = \frac{\text{Tamaño del efecto en el gráfico}}{\text{Tamaño del efecto en los datos}} \tag{1}$$

Por tanto, una representación gráfica debería poseer principios de *integridad gráfica* con el objeto de reducir este factor mentira. Las variaciones mostradas por el gráfico deberían ser directamente proporcionales a las cantidades numéricas de los propios datos que representan. Además también debería tener un etiquetado detallado para prevenir ambigüedades y aclarar las posibles distorsiones gráficas.

Otro concepto es la relación *datos–tinta* que define como la tinta de un gráfico que representa unos datos determinados.

$$\text{Datos–tinta} = \frac{\text{Tinta de los datos}}{\text{Tinta total utilizada en el gráfica}} \tag{2}$$

Tufte afirma que las buenas representaciones gráficas maximizan la relación datos–tinta tanto como sea posible. De esta manera todos los elementos visuales del gráfico que no son necesarios para entender la información representada (denominado *chartjunk*) deberían evitarse para no distraer el mensaje de la representación, se puede ver un ejemplo en Fig. 2.2. También se recomienda aumentar la *densidad de datos* de un gráfico, que consiste en la proporción del tamaño total del gráfico que es dedicada a la representación de los datos y se define de la siguiente manera:

$$\text{Densidad de datos} = \frac{\text{Números de objetos}}{\text{Área de datos en el gráfico}} \tag{3}$$

Otro ejemplo de buenas prácticas en la representación es la *separación* para enfatizar algún aspecto y su representación por capas cuando los elementos poseen diferencias sustanciales, teniendo en cuenta la relación apropiada de la información entre las capas. Un ejemplo es *Google Maps* donde se pueden añadir capas en un mapa que representen distintos tipos de información, como por ejemplo el transporte público.

También se proponen algunas técnicas. Por ejemplo se recomienda el uso de una serie repetida de pequeños gráficos similares entre sí denominados *small multiples*, los cuales son una gran herramienta para visualizar grandes cantidades de datos con muchas dimensiones. En la Fig. 2.3 se muestra un ejemplo de *small multiples* para representar los índices de desempleo en la zona euro publicado en el diario *The Washington Post*. En este caso se han utilizado

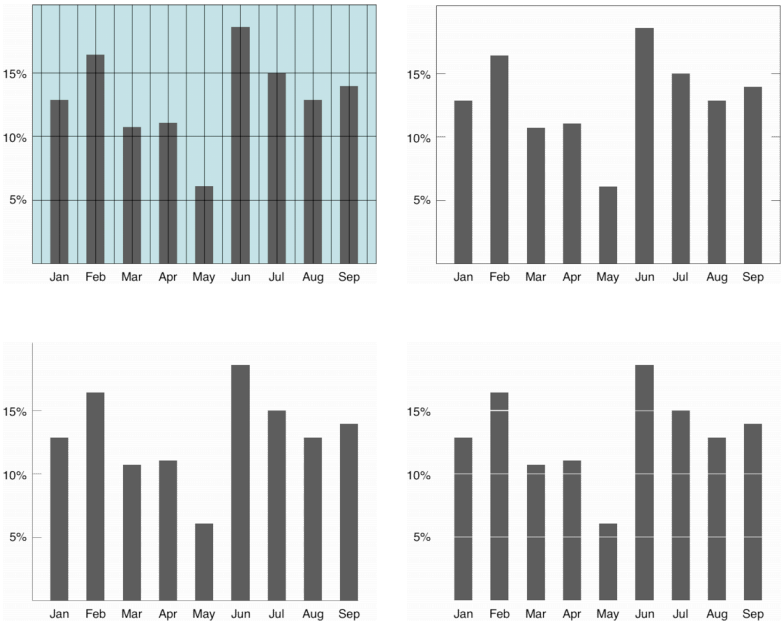


Figura 2.2.: Ejemplo de eliminación de *chartjunk* en un gráfico. Fuente: Tim Bray (http://www.tbray.org/ongoing/)

sparklines, que son pequeñas gráficas (sin ejes) mostrando la variación general de una medida a lo largo de un período de tiempo determinado, en este caso, el valor del índice de desempleo en la zona euro durante los años 2007 y 2013.

Por tanto, las guías generales de diseño aportadas por Tufte se podrían resumir con las siguientes pautas:

- Maximizar la relación datos–tinta
- Evitar *chartjunk*
- Aumentar la densidad de datos
- Separación por capas

Aunque Tufte es uno de los nombres más emblemáticos, existen muchos más autores en este campo [41, 34]. Por ejemplo en los trabajos de Stephen Few [63, 64, 65] se recogen fundamentos sólidos y excelentes consejos de visualización mediante guías de selección del adecuado gráfico dependiendo del mensaje que se quiere transmitir. Colin Ware [202, 203] estudió el comportamiento de nuestro cerebro cuando vemos elementos gráficos, lo cual conviene tener

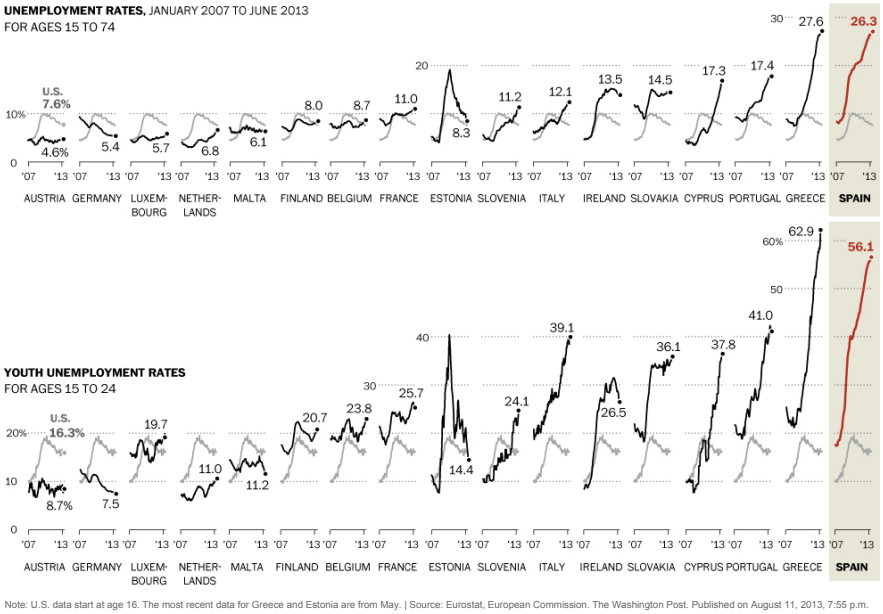


Figura 2.3.: Ejemplo de *small multiples* utilizando *sparklines*. Representación de índices de desempleo en la zona euro entre 2007 y 2013. Publicado por *The Washington Post* en Agosto 2013.

siempre presente a la hora de diseñar representaciones. Además recientes autores han revisado métodos, y presentado conceptos de forma que facilitan el aprendizaje del diseño visual de la información [32, 109, 137, 142].

2.3 PRINCIPIOS DE CODIFICACIÓN VISUAL

2.3.1 *Tipos de datos*

Muchas de las consideraciones que se realizan en el diseño de una visualización están condicionadas por el tipo de datos que se pretenden representar. Por ejemplo, es diferente la forma de presentar una tabla de números o una localización en un mapa. En el libro [171] Tamara Munzner realiza una distinción entre datos en forma de *tabla*, comúnmente utilizados en el análisis de datos y que trataremos en esta tesis; relacionales o *grafos*, donde nodos son enlazados por las relaciones que tienen entre sí formando redes, por ejemplo en forma de árbol; y *espaciales* como una situación geográfica o un campo de medidas tridimensionales como los utilizados en imágenes médicas. Cuando en el presente documento se indique

cualquier conjunto general de datos se referirá a un conjunto en forma de tabla. En una tabla de datos se suelen considerar las filas como objetos o muestras y las columnas como atributos o variables de cada muestra. Los tipos de atributos se pueden interpretar en términos de escalas de medidas, de la siguiente forma:

- Nominal o categórica: cuyos elementos describen objetos iguales o distintos entre sí, como por ejemplo la fruta (manzanas, naranjas, etc).
- Ordinal: En el que obedece a una relación de ordenación, como por ejemplo la talla de una camiseta (pequeña (S), mediana (M), grande (L), etc).
- Cuantitativa: Valores numéricos con los que se puede operar, como por ejemplo altura (180 cm), peso (80 Kg), etc

Estas escalas se basan en el trabajo original de Stevens [177], donde se clasifican las escalas de medida de una manera más detallada. Los tipos nominal y ordinal describen los datos mientras que las cuantitativas son números para ser analizados y dependientes entre sí. Este debate sobre las escalas de medida también se extiende en el trabajo de Wilkinson [204].

Además existen algunas estrategias para la reducción de datos como el *filtrado*, que elimina algunas muestras o algunos atributos o también la *agregación*, representar un conjunto de elementos por otro nuevo elemento calculado a partir de ese grupo.

2.3.2 *Canales visuales*

Las variables (o canales) visuales son un conjunto de elementos gráficos aplicados a datos para transmitir información mediante su codificación visual. Uno de los primeros trabajos relacionados fue realizado por Bertin [15], donde se definen unidades básicas denominadas *marcas*, que llevan información mediante canales visuales. Un punto es una marca adimensional, una línea es de una dimensión, una marca de dos dimensiones es un área, y de tres dimensiones es un volumen. Con estas unidades se desarrollan una serie de métodos dando lugar a 7 canales visuales que codifican la información: posición, color, tamaño, forma, orientación, textura y cambios de valor en una escala de grises. Más tarde, se estudió la decodificación de variables visuales, es decir, la precisión para distinguir la información codificada gráficamente. Por ejemplo, en [42] se establecen fundamentos de percepción gráfica en la identificación de tareas elementales usando datos cuantitativos. Tanto las

variables visuales como la precisión en su percepción visual fueron extendidas por Jock D. Mackinlay [128]. Las capacidades visuales para una correcta distinción dependen de si el tipo de datos es cuantitativo, ordinal, o categórico. En Fig. 2.4 se representa la clasificación de las variables visuales con respecto a la precisión en las tareas de percepción.

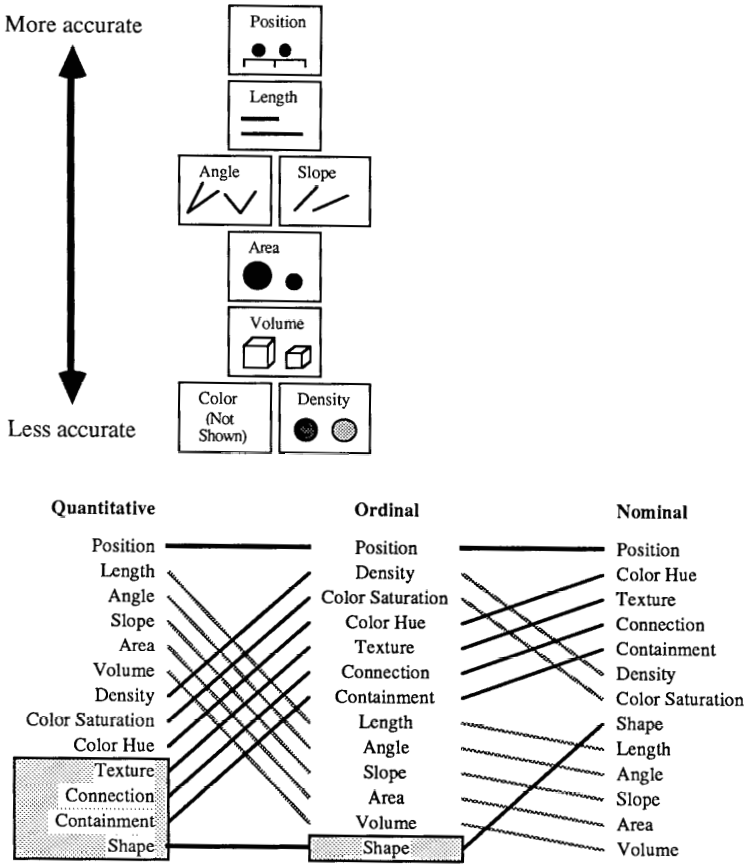


Figura 2.4.: Ranking de variables visuales respecto a la precisión en la percepción de datos cuantitativos probado en [42] (arriba) y para los distintos tipos de datos (abajo) en [128].

La posición es el canal visual más preciso para los tres tipos de datos, por tanto domina nuestra percepción de codificación visual. Por eso las dos dimensiones más importantes son normalmente representadas con las posiciones horizontal y vertical. A veces no es posible utilizar la posición o puede producir una difícil percepción, por ejemplo en la situación confusa de ver muchos puntos superpuestos (fenómeno denominado *cluttering*). La eficacia del uso de

otros canales depende de los distintos tipos de datos utilizados, por ejemplo los canales de longitud y ángulo son efectivos para datos cuantitativos pero no para datos categóricos. El tamaño y la longitud son buenas variables para comparar diferencias, aunque son malas para cambios mediante el área. Sin embargo, el color es muy preciso para datos categóricos pero mediocre para datos cuantitativos sin una adecuada escala. La forma es también adecuada para reconocer muchas clases sin ningún tipo de orden.

Muchos canales visuales se pueden utilizar simultáneamente para codificar diferentes dimensiones de los datos. Por ejemplo, en un diagrama de puntos (*scatterplot*) es habitual el uso de posición horizontal y vertical, color y tamaño para representar cuatro dimensiones. También más de un canal puede utilizarse para codificar la misma dimensión de forma redundante, se transmite menos información pero de forma más clara.

2.3.3 *Percepción visual*

El sistema visual humano utiliza regiones del cerebro donde se llevan a cabo tareas para procesar la información, la cual se extrae de diversas maneras, involucrando desde procesos cognitivos primarios hasta altos niveles de procesamiento donde se combina con conocimiento previo. Colin Ware propone en [202, 203] un modelo de percepción visual dividido en tres etapas:

- Etapa 1: millones de características básicas se procesan en paralelo simultáneamente.
- Etapa 2: procesado más lento para la extracción de patrones y estructuras.
- Etapa 3: procesado orientado a la tarea con información reducida que se retiene en la memoria visual para formar la base de pensamiento visual.

El cerebro combina la información mediante dos tipos de procesos. En los procesos *bottom-up* la información se selecciona y filtra de forma que características de bajo nivel en la primera etapa forman patrones en la segunda y objetos en la tercera mientras que los procesos *top-down* modulan la atención en función de la necesidad para cumplir un objetivo, por ejemplo simplemente entender una idea expresada en un diagrama. Esta atención causa una tendencia en favor de las señales que buscamos. Por ejemplo, si buscamos puntos rojos en una imagen, se da mayor prioridad a

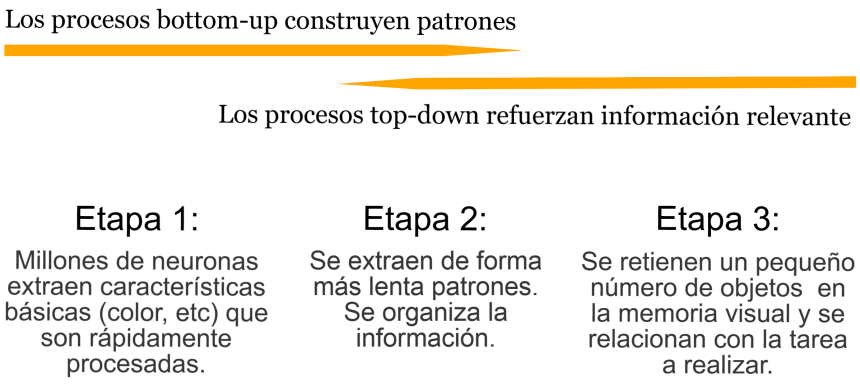


Figura 2.5.: Modelo de tres etapas de percepción visual. Imagen basada en [203].

los receptores de puntos rojos frente a otros. En el esquema representado en la Fig. 2.5 se simplifica el modelo propuesto por Ware [203].

El procesamiento de pre-atención sucede muy rápidamente (milisegundos) realizando una extracción de características básicas visuales (etapa 1) en paralelo. Este procesamiento es previo a la atención consciente y se refiere a la detección de cosas en lo que comúnmente llamamos *de un vistazo*. Los diseñadores pueden utilizar este tipo de características para hacer que información relevante sobresalga sobre el resto en las visualizaciones, efecto denominado *pop-out*. Por ejemplo, en la parte izquierda de la Fig. 2.6 el círculo verde se localiza rápidamente respecto a los demás círculos. Esto nos permite ejecutar tareas como la identificación de aspectos importantes en un gráfico de manera más eficiente. Muchos canales poseen esta propiedad, como por ejemplo el color, la forma, la curvatura, o la dirección de la luz. Sin embargo, solamente se puede aprovechar por un canal a la vez, por ejemplo en la parte derecha de la Fig. 2.6 los 3 cuadrados verdes no muestran un efecto de *pop-out*, incluso aunque se sepa lo que se está buscando. Este tipo de búsquedas, implicando más de un canal simultáneamente, se denomina *búsqueda conjuntiva visual*, y en la mayoría de los casos, son difíciles de realizar. Estos procesos de búsqueda son secuenciales (no paralelos) y, en estos casos, el tiempo que lleva encontrar los objetos crece proporcionalmente con el número total de objetos de la escena.

Si se desean buscar varias cosas al mismo tiempo fácilmente, la solución es utilizar canales diferentes, pero algunos canales visuales

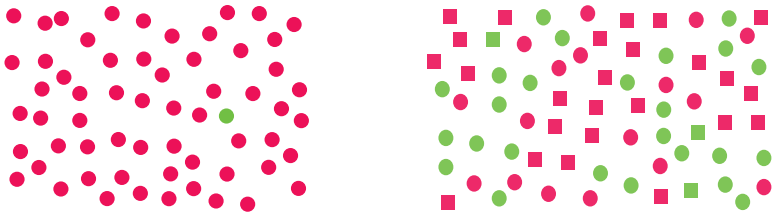


Figura 2.6.: Efectos de *pop-out*. Izq: el punto verde sobresale. Dcha: 3 cuadrados verdes no sobresalen sobre el resto. Tanto color y forma pueden tener efecto *pop-out* pero no los dos a la vez [203].

ejercen una interacción dentro del nivel consciente del ser humano, se integran entre sí dificultando su interpretación, por lo que no son buena elección para codificar diferentes dimensiones. Para esto, los canales más adecuados son aquellos que no tengan interacción en el procesado visual, es decir, deberían ser *separables*. Por ejemplo, el color y la posición son muy separables, también se puede ver que el tamaño horizontal y vertical no son tan fáciles de separar, porque nuestro sistema visual los integra en el área de percepción. El tamaño interacciona con muchos canales, por ejemplo cuando un objeto se hace más pequeño, más difícil se distinguen el color y su forma.

A través de la discriminación (igual-distinto) se detectan y diferencian elementos y patrones, los cuales son esenciales para organizar la información visual. También se han propuesto unos principios conocidos como *leyes de Gestalt*, explicadas también en detalle en la obra de Ware [202], que describen las maneras en que detectamos patrones y cómo integrar unidades individuales para una percepción coherente.

El color

El color puede ser un potente canal visual si es utilizado de manera adecuada, por lo que sus propiedades deberían entenderse correctamente [127]. Puede considerarse el color con respecto a tres canales visuales de percepción separados: tonalidad, saturación, y luminosidad.

La tonalidad es lo que comúnmente asociamos a los nombres de los colores (azul, rojo, etc) y es un efectivo canal para codificar datos categóricos, pero el rango de colores es limitado. La gente puede distinguir sobre una docena de colores distintos, por lo que

no permite codificar demasiadas clases. Además, los tonos de colores no poseen ordenación ninguna, por lo que no es un buen canal para representar datos ordenados. La saturación se refiere a la viveza del tono de color, en donde valores más bajos corresponden a tonos más oscuros del color. La luminosidad es una medida relativa que describe la cantidad de luz de un objeto comparado con lo que se representa en blanco. Se puede ordenar, por lo que se puede hablar de una escala desde valores claros a oscuros dentro de un mismo color. Por tanto, para datos ordinales la saturación y luminosidad son canales más efectivos por la propia ordenación que tienen implícita.

El rango de valores que codifican datos cuantitativos se denomina mapa de colores (*colormap*). Su diseño también debe tenerse en cuenta para mostrar correctamente las cantidades mediante una escala lineal, por ejemplo evitando escalas como la de arco iris (*rainbow*). El artículo [23] describe las características que hacen de una escala de arco iris una mala opción. Para tareas de comparación de valores, el mapa de colores debería seguir una ordenación perceptual, como por ejemplo una escala de grises. Sin embargo, la escala de arco iris está ordenada con los colores de la longitud de onda de la luz, pero no está ordenada de manera intuitiva, lo que puede resultar confuso. También oculta la variación real de los datos a través de su ineficacia para representar pequeños detalles dentro de su escala cuantitativa, y puede engañar introduciendo artefactos en la visualización, por ejemplo los cambios bruscos entre sus tonos de color podrían percibirse como transiciones bruscas en los datos que en realidad no existen. La aplicación *ColorBrewer* (www.colorbrewer.org) [27] es un recurso muy utilizado para la construcción de diversos mapas de colores de manera adecuada.

2.4 PRINCIPIOS DE INTERACCIÓN

Aparte de la representación de los datos en un gráfico, el componente de la interacción supone el diálogo entre el usuario y el sistema de exploración de los datos mostrados. La interacción es un aspecto esencial a la hora de considerar cambios en una visualización por parte del usuario. Los fundamentos de la interacción se apoyan dentro del área de la interacción persona-computador (*human-computer interaction*, HCI). Aunque estos conceptos se suelen estudiar separados de la visualización, no son excluyentes entre sí, por ejemplo una interacción del usuario puede activar un cambio de la representación.

Aunque una imagen estática diseñada adecuadamente puede resultar valiosa [143, 185], su utilidad puede estar limitada al tamaño de los datos y sus variables. A través de la interacción se pueden superar algunos de los límites que tiene una simple representación y amplificar el conocimiento del usuario [54]. Por tanto, explorar mayor cantidad de información que en una imagen estática es una de las potentes ventajas que permite la interacción. Sin embargo, ésta requiere un coste en cuanto al tiempo y la atención por parte del humano. Si un usuario debe comprobar todas las posibilidades del sistema, la interacción puede resultar ineficaz. Por otra parte, si la tarea se puede resolver automáticamente entonces no sería necesario ningún tipo de interacción. Siempre existe un compromiso para encontrar aspectos automáticos y la introducción del humano dentro del lazo de un proceso de análisis de datos. Además, la respuesta de los mecanismos de interacción afecta directamente en la calidad del proceso de exploración de los datos. En [57] se introduce el concepto *fluidez* en este campo y se proponen unas guías prácticas para el diseño de mecanismos de interacción, apoyadas con ejemplos explicativos.

Shneiderman propuso una taxonomía de técnicas interactivas [173], además de su influyente mantra: "*Overview first, zoom and filter, then details on demand*". Una visión en conjunto ayuda al usuario a identificar regiones donde un análisis puede ser interesante, en el que se puede posteriormente acceder a ellas mediante filtrado o navegando para solicitar detalles que pueden presentarse de muchas maneras, por ejemplo con la posición del cursor, la pulsación de un botón, etc. Existen más taxonomías en la literatura [54, 204, 105, 65] que describen tipos de interacción para realizar diversas tareas sobre los datos. Aunque muchos de estos estudios comparten elementos comunes, se diferencian en algunos aspectos. Mientras que unos se enfocan en técnicas de interacción de bajo nivel en el sistema [54], otros se centran en operaciones tales como navegación, selección y distorsión, así como en la identificación de espacios donde tales operaciones pueden aplicarse (pantalla, dato, etc) [201]. También existen taxonomías orientadas a las tareas que realiza el usuario como asociar, comparar, etc. [214].

En [210] se realiza una extensa revisión de trabajos en este campo y se sugieren fundamentos para entender el papel que juega la interacción en la visualización de la información. De su estudio emergen 7 categorías de interacción centrándose en la tarea que el usuario quiere alcanzar a través de una técnica de interacción específica. A continuación se describen estas categorías.

2.4.1 *Categorías de interacción*

Con la idea de que la interacción está siendo realizada por una persona para un propósito, se diferencian 7 categorías: seleccionar, explorar, reconfigurar, codificar, abstraer, filtrar y conectar.

Seleccionar: marcar algo como interesante

Proporciona al usuario la habilidad para marcar objetos de interés, los cuales se pueden hacer visualmente distinguibles y seguirlos, incluso en grandes conjuntos de datos, o si la representación cambia.

Las técnicas de selección parecen funcionar frecuentemente como acciones previas a otras operaciones. Ejemplos de aplicaciones donde se pueden ver selecciones son *Dust & Magnet* [211] cuyos objetos se seleccionan y se etiquetan en rojo, lo que facilita su seguimiento en posteriores operaciones, o *TableLens* [154] que visualiza datos numéricos mediante diagramas de barras en una vista de tabla, en donde la técnica de interacción es similar resaltando muestras en la tabla en lugar de etiquetarlas. En estos ejemplos se observa que su acoplamiento con otras técnicas enriquece la exploración del usuario.

Explorar: mostrar algo más

Esta interacción permite examinar un subconjunto diferente de los datos. Cuando se visualizan un conjunto de datos se suele ver solo una parte de ellos debido a las limitaciones del tamaño de la pantalla o de la propia percepción. Los usuarios normalmente ganan entendimiento moviendo los datos en la pantalla. La técnica más común se denomina *panning* que se refiere al movimiento de la cámara sobre la escena, o de la escena mientras la cámara permanece fija. Se realizan con simples movimientos del ratón o mediante barras de desplazamiento. Muchos sistemas poseen este tipo de interacción por ejemplo los citados anteriormente o también *Spotfire* [3] y *Vizster* [85].

Reconfigurar: mostrar una diferente disposición

Facilita al usuario diferentes perspectivas de los datos mediante cambios en la disposición espacial de las representaciones. Un objetivo principal es revelar características ocultas de los datos y las relaciones entre ellos. Una buena representación podría servir

para este propósito pero en ocasiones no proporciona la perspectiva suficiente. Por eso muchas herramientas permiten al usuario cambiar la disposición de los datos o la alineación para aportar diferentes perspectivas de ellos.

Las operaciones de ordenación en *TableLens* [154] son un ejemplo de este tipo de interacción. También la capacidad de cambiar atributos presentados en los ejes de la vista *scatterplot* de *Spotfire* [3], o en los mecanismos de exploración de datos multidimensionales presentados en [56], donde se puede navegar por varios *scatterplots* mediante transiciones animadas y también ordenar las dimensiones. En todo estos casos se cambian las perspectivas de los datos que mejoran la visualización de las relaciones existentes en ellos.

Codificar: mostrar una representación diferente

Esta categoría permite alterar la representación visual fundamental de los datos incluyendo la apariencia visual (color, tamaño, etc) de cada elemento. Los elementos visuales juegan un papel importante no solo por su rápida identificación sino también porque explican las relaciones y distribuciones en los datos.

Cambiar la forma en la que los datos están representados es un ejemplo de codificar, donde se espera que el usuario descubra nuevos aspectos de los mismos. Muchos sistemas presentan múltiples representaciones posibles de los datos, por ejemplo *Spotfire* [3] o *Xmdv tool* [199] tienen esta característica, en la que los mismos datos multidimensionales pueden visualizarse por medio de técnicas diferentes (diagramas de barras, *scatterplots*, etc). Otros sistemas alteran la codificación del color, como *Dust & Magnet* [211] o en *Attribute Explorer* [176] que es una técnica de codificación de color que ayuda a los usuarios a entender distribuciones de múltiples variables.

Abstraer/Elaborar: mostrar más o menos detalle

Proporciona al usuario el mecanismo para ajustar el nivel de abstracción en la propia representación de los datos. Este tipo de interacciones permiten alterar la representación desde una visión de conjunto a detalles de casos individuales con muchos niveles entre ambos.

Ejemplos de técnicas dentro de esta categoría son aquellas operaciones de detalles en demanda (como el *focus and context* [114]) incluidos en sistemas como *TableLens* [154] que permite centrarse en detalles, los cuales emergen en forma de texto con los valores

reales. Otro ejemplo es *zooming*, en el cual el usuario cambia la escala de una única representación de forma que se puede ver el conjunto total o el detalle de un subconjunto más pequeño de los datos.

Filtrar: mostrar algo condicionalmente

Un filtrado permite cambiar el conjunto de los datos a representar mediante unas condiciones específicas. Se determina una condición, de manera que solamente se representan datos que cumplen ese criterio. Los objetos que no lo cumplen se ocultan o se muestran de otra forma, de manera que cuando se elimina el criterio son recuperados otra vez. No se cambia la perspectiva en los datos, sino que se especifican condiciones en los que se muestran.

Los controles de peticiones dinámicas como los desarrollados en *Spotfire* [3] son un ejemplo representativo de este tipo de interacción. En *Attribute Explorer* [176] se extiende esta capacidad cambiando los colores de los datos filtrados en lugar de quitarlos de la pantalla.

Conectar: mostrar objetos relacionados

Esta interacción se refiere a técnicas que se usan para resaltar relaciones entre objetos de los datos ya representados. Cuando varias vistas se usan para mostrar diferentes representaciones de los mismos datos puede ser difícil identificar objetos correspondientes en otras vistas. El *brushing* se usa para resaltar la representación de objetos de los datos en otras vistas dibujadas, estableciendo una conexión de los objetos seleccionados en una de las vistas y mostrados en el resto. Aunque también se puede aplicar en situaciones con una vista, en *Vizster* [85] se visualizan redes sociales mediante grafos y el paso del cursor resalta nodos relacionados cambiando la opacidad del resto.

También sirve para mostrar elementos ocultos que son relevantes a un objeto específico y que no fueron mostrados inicialmente. En *Vizster* un doble clic en un nodo causa una expansión del nodo añadiendo todos sus nodos relacionados.

2.4.2 *Animación*

Una animación muestra cambios en el tiempo en una representación. Se distingue cuando fotogramas sucesivos pueden reproducirse y/o pararse mediante un control interactivo. La animación

tiene varios pros y contras. Un uso eficiente requiere un buen conocimiento del rol de la animación en el proceso de percepción y en la cognición [188, 86].

Aunque la animación puede ser efectiva durante una narración a veces se puede usar inconscientemente en una visualización [188]. Puede parecer obvio mostrar datos que cambian en el tiempo mediante una animación. Sin embargo, la gente tiene dificultad en hacer comparaciones específicas entre fotogramas que no son contiguos cuando ven una animación. La capacidad limitada de la memoria hace que seamos peores en comparar cosas memorizadas que hemos visto que comparar aquellas que están en nuestro campo de visión. Para tareas que requieren comparaciones entre varias docenas de imágenes, la comparación lado a lado es más efectiva que una animación. Además si el número de los objetos que cambian con el tiempo es grande, la gente apenas tendrá tiempo para seguir todo lo que ocurre [157].

No obstante, la presentación de datos animados puede ser beneficiosa cuando se asegura que los datos cuentan una historia limpia y se diseña evitando que muchas acciones ocurran simultáneamente. Se ha demostrado que las transiciones animadas pueden ser más eficientes que los saltos discontinuos, ayudando a seguir cambios en posiciones de objetos en gráficos estadísticos [86]. Para el caso especial de dos imágenes, una simple animación entre una y otra puede ser una manera útil para identificar diferencias entre ellas. Finalmente, en [51] se proporciona una animación controlada por el usuario entre varias representaciones de datos, cuyos estados intermedios pueden tener significado exploratorio y resultar útiles para el análisis de patrones eléctricos.

2.5 TÉCNICAS DE VISUALIZACIÓN DE DATOS

Anteriormente se han explicado algunos fundamentos de visualización para entender nuestra percepción visual y utilizar un diseño gráfico adecuado a la hora de representar información. A continuación , sin pretender realizar una lista exhaustiva, se muestran algunas técnicas de visualización de datos con varias dimensiones. Estas técnicas muestran diversos ejemplos sobre cómo otros resolvieron problemas reales de diseño de una manera efectiva.

2.5.1 *Matriz de scatterplots*

El diagrama de puntos (*scatterplot*) es la técnica más común en análisis de datos para la representación de dos variables, en la

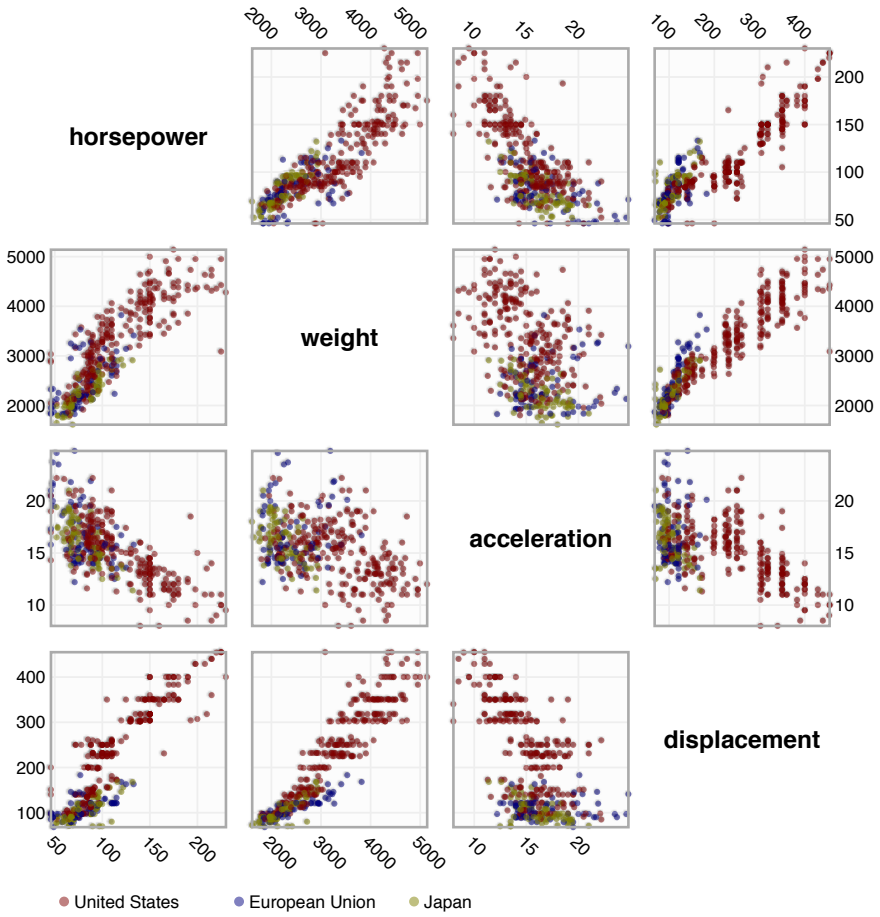


Figura 2.7.: Vista de matriz de scatterplots para el conjunto de datos de automóviles. Captura de pantalla del ejemplo interactivo en la versión web del artículo [84]

cual los objetos se representan por puntos donde el valor de una variable está representado por la posición en el eje horizontal, y el valor de otra variable determina la posición en el eje vertical. Para la representación de datos con varias variables se propuso utilizar pequeños diagramas de este tipo para cada pareja de variables de los datos en forma de matriz [35]. Esta técnica denominada matriz de scatterplots (*scatterplot matrix*, SPLOM) permite la inspección visual de las relaciones entre cualquier par de variables. Además se pueden utilizar técnicas de interacción (como *brushing* y *linking*) que conectan las vistas y facilitan la exploración de patrones en los datos. También incluso se puede usar una navegación más avanzada para su exploración mediante animaciones [56].

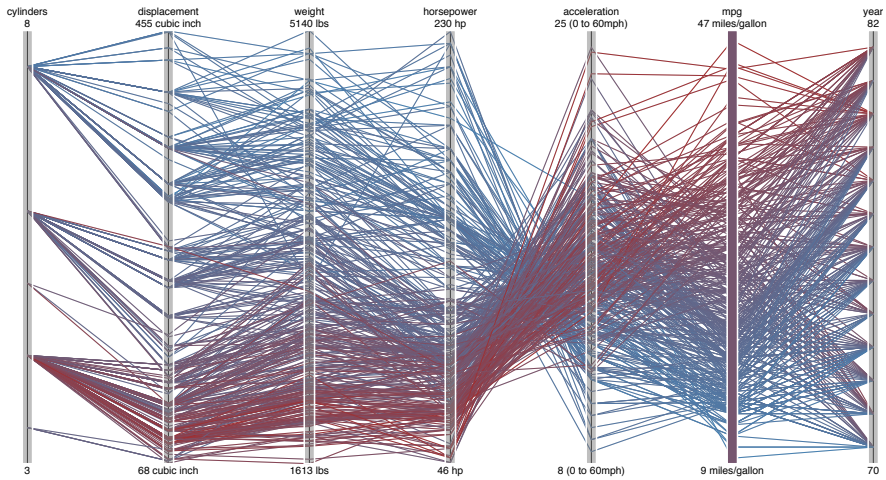


Figura 2.8.: Vista de coordenadas paralelas para el conjunto de datos de automóviles. Captura de pantalla del ejemplo interactivo en la versión web del artículo [84]

2.5.2 *Coordenadas paralelas*

Otro método de visualización son las coordenadas paralelas (*parallel coordinates*, PC). Este método fue presentado inicialmente por Inselberg [94] y utilizado en varias herramientas de visualización [199, 97]. Cada dimensión de los datos se representa por una línea vertical, cada línea conectada entre dichos ejes representa los valores correspondientes a una muestra de los datos. Por tanto, líneas que se cruzan indican relaciones inversas entre esas variables. Cambiar el orden de las dimensiones, así como seleccionar un subconjunto de muestras, lo cual se puede realizar mediante interacción, puede ayudar a encontrar patrones en datos con varias dimensiones.

2.5.3 *Técnicas orientadas de píxel*

Estas técnicas [103] representan tantos objetos como sea posible en la pantalla al mismo tiempo, utilizando un píxel coloreado para describir cada valor numérico (por ejemplo la dimensión j de la muestra i). Se obtiene una imagen, para cada una de las dimensiones de los datos, agrupando todos los píxeles en una dimensión determinada. Esto permite la visualización de grandes conjuntos de datos dividiendo la pantalla por cada una de las dimensiones de los datos. Pueden resultar útiles para encontrar propiedades interesantes y grupos distintos que se encuentren en grandes bases

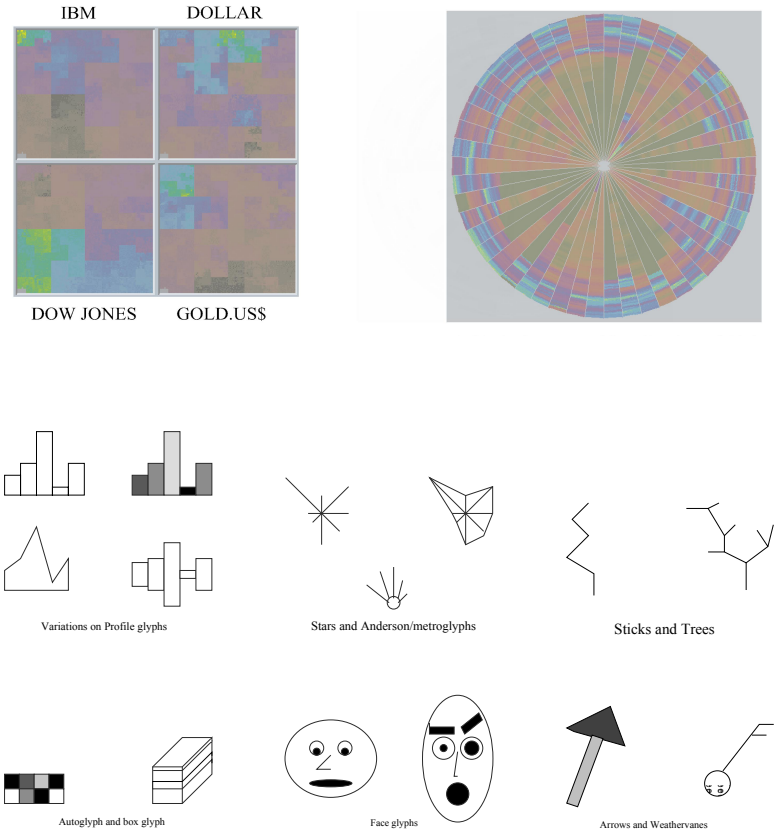


Figura 2.9.: Ejemplo de visualizaciones con técnicas orientadas de píxel usando curvas de Peano-Hilbert (superior, izquierda) y segmentos circulares (superior, derecha) de datos financieros, ejemplos de glifos (abajo). Fuentes [6, 200]

de datos. En la parte superior de la Fig. 2.9 se pueden ver representaciones de estas técnicas. En la parte izquierda se presentan datos financieros (como acciones de IBM, el índice Down Jones, el precio del oro o la divisa) desde el año 1987 hasta 1995, utilizando la curva de Peano-Hilbert. En la parte superior derecha de la misma figura se muestra la técnica denominada segmentos circulares (*circle segments*) [6], donde se visualizan los valores de 50 acciones alemanas en el periodo entre enero de 1974 y abril de 1995.

2.5.4 *Técnicas basadas en glifos*

Los glifos son símbolos gráficos que expresan varios valores de los atributos. Comparten la misma representación para cada objeto de los datos diferenciándose en la magnitud codificada de cada variable representada en el icono (por ejemplo, longitud, color). Un ejemplo que ilustra este concepto pueden ser las caras de Chernoff [39] que codifican diferentes variables mediante variaciones geométricas (escala, curvatura, rotación, etc) en las características de una cara humana. En [200] se revisa de manera detallada este tipo de representaciones, algunos ejemplos se representan en la parte inferior de la Fig. 2.9.

2.5.5 *Mapa auto-organizado*

El mapa auto-organizado (*self-organizing map*, SOM) [110] es un algoritmo que utiliza un tipo de red neuronal, cuya visualización es una herramienta eficiente para representar datos multidimensionales. Dicha red está compuesta por un conjunto de prototipos distribuidos de manera que representan a los datos de entrada y se visualiza mediante una proyección que preserva las propiedades topológicas de los datos. De esta forma, se obtiene una representación discreta en un espacio de menor dimensión, normalmente en un plano rectangular de dos dimensiones.

El SOM posee numerosas ventajas, útiles no solo para el análisis de datos con varias dimensiones, sino también para su visualización [197]. La estructura de la nube de puntos en el espacio de entrada proporciona información para entender el proceso. El SOM permite captar las características más relevantes de esa estructura mediante la preservación de la topología, donde objetos cercanos en la proyección, son también cercanos en el espacio de entrada.

El SOM proporciona numerosas formas de representar la información contenida en la estructura de los datos. Una de las técnicas más comunes es el uso de los *planos de componentes*, para realizar

un estudio detallado de los vectores prototipo. Cada plano de componente consiste en el mapa del SOM, en el que cada nodo de la retícula rectangular es codificado mediante una escala de color con los valores de cada variable del vector de datos. Proporciona una idea de la variabilidad de los valores en cada uno de los componentes de los datos. También pueden utilizarse para buscar fácilmente correlaciones entre parejas de variables, las cuales se revelan con similares patrones en posiciones idénticas de los planos de componentes. Pueden considerarse como los *small multiples* descritos anteriormente, en los que se pueden establecer vistas coordinadas, de manera que varias visualizaciones estén conectadas y mediante técnicas de interacción (como *linking* y *brushing*) puedan realizarse cambios en una vista y que se reflejen en otra. En el medio de la Fig. 2.10 se representan unos planos de componentes como ejemplo, en el que se remarcan tres planos cuyas variables están relacionadas entre sí.

Otra técnica para mostrar los grupos existentes en los datos es mediante matrices de distancias. La más utilizada se denomina *U-Matrix* [189], en la que se calculan las distancias de cada elemento del mapa con sus vecinos y se visualizan en una escala de grises, donde el tono más oscuro indica mayor distancia. Por tanto, los grupos contenidos en los datos pueden apreciarse como áreas más claras en el mapa con bordes oscuros. Por otra parte, las distancias también se pueden codificar por el tamaño o las formas de las unidades del mapa. Otra ventaja de este enfoque es mostrar la similitud de las unidades del mapa. Esto se puede realizar dando un color a cada unidad de forma que áreas con colores similares se encuentran cerca en el espacio original. En la parte superior de la Fig. 2.10 se pueden ver este tipo de representaciones de la U-Matrix. Los datos utilizados se encuentran en [197] y constan de información de fábricas de papel del mundo, donde la dimensión del espacio es de 75 incluyendo varios grupos de datos solapados.

Finalmente, también se han desarrollado numerosos mecanismos para la representación de nuevos datos con respecto a un mapa. Normalmente se realiza con las distancias entre las muestras y el vector prototipo más cercano, denominado *best matching unit* (BMU). Para varios vectores, se obtiene un histograma de los datos, que se puede visualizar por ejemplo representando un hexágono negro de tamaño proporcional al valor del histograma en la unidad correspondiente, representado en la parte inferior izquierda de la Fig. 2.10. Otro enfoque es mostrar en el mapa información sobre la posición del nuevo punto mediante la precisión, de forma que para un conjunto nuevo de datos se muestra la media del error de

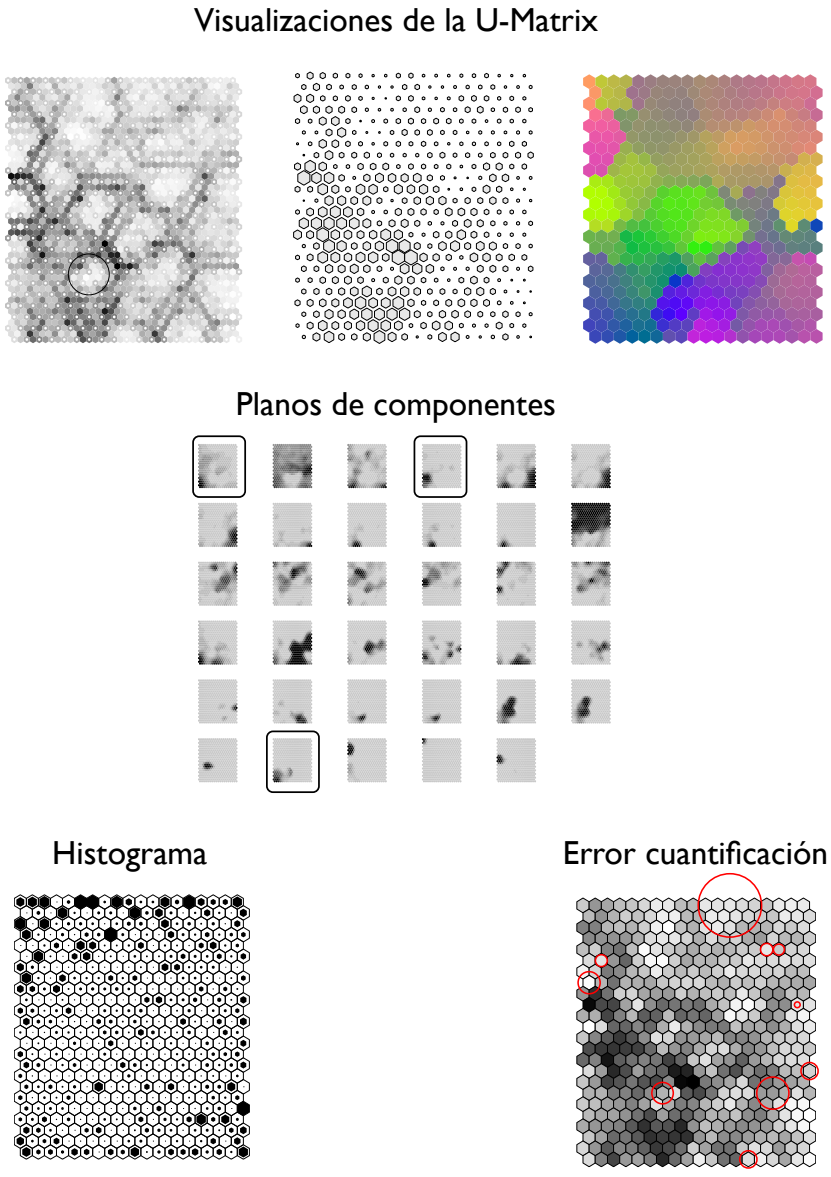


Figura 2.10.: Visualizaciones realizadas utilizando el SOM. Fuente [197].

cuantificación [1] representado de manera sencilla. Por ejemplo, en la parte inferior derecha de la Fig. 2.10, el diámetro del círculo se escala por la distancia media de cada unidad a sus vecinos. Si el círculo es más pequeño que el hexágono, la BMU está más cerca al dato (en valor medio) que a sus vecinos.

2.5.6 *Otras técnicas*

Además de las técnicas presentadas anteriormente, como las geométricas (SPLOM o PC), de pixel, el SOM o de las basadas en glifos, existen más formas para representar la información. En muchas aplicaciones es esencial conocer la estructura o alguna cualidad de los datos que se quieren representar (si indican valores temporales, geográficos, etc) para seleccionar una técnica adecuada.

En [4] se realiza una revisión de métodos que realizan tareas de análisis para atributos temporales en los datos. Para datos con una estructura jerárquica es habitual una representación en forma de árbol [167], como por ejemplo *Treemaps* [172], y también técnicas donde se representan redes [11] como aquellas basadas en grafos para la visualización de datos donde son importantes las relaciones entre sí, como pueden ser diagramas con nodos para representar elementos y líneas para las conexiones entre ellos. Además, estas técnicas no solo pueden ser combinadas con mecanismos de interacción, sino en ocasiones también con algunas técnicas de distorsión [123].

Finalmente reseñar la representación de datos que muestran una distribución espacial mediante posiciones relativas de sus componentes. Es el caso de los *mapas* que representan las relaciones de los datos mediante sus distribuciones geográficas. Suelen estar diseñadas utilizando diferentes proyecciones para transformar la información en un plano y dotadas de escalas como medida del grado de reducción entre las distancias reales y las representadas. Con su aplicación se pueden obtener técnicas de análisis muy útiles, incluso también combinando la información espacial con la temporal [5]. Uno de los primeros ejemplos es la demostración que realizó el Dr. John Snow, mostrando que el violento brote de cólera producido en Londres en 1854 fue causado por el consumo de aguas contaminadas. Para ello cartografió en un plano las bombas de agua con la mortalidad semanal del distrito, localizando el epicentro del brote. Aunque no fue el primero en utilizar mapas para

1 Distancia entre la muestra de los datos con su BMU correspondiente.

el estudio de enfermedades, ayudó a sentar las bases del método revelando patrones en epidemias en un contexto espacial.

3

ANÁLISIS DE DATOS MULTIDIMENSIONALES

En este capítulo se realiza un breve repaso relacionado con el análisis de datos multidimensionales. Se introducen varios conceptos relacionados con el tratamiento de datos, especialmente las técnicas automáticas que aprenden de los datos. Se describen algunos algoritmos para este propósito que se utilizarán más adelante, haciendo especial hincapié en las técnicas de reducción de la dimensión.

3.1 INTRODUCCIÓN

El análisis de datos es un término amplio en el que están implicados varios procesos aplicados a datos con el objeto de descubrir información, extraer conclusiones, sugerir nuevas hipótesis, y apoyar la toma de decisiones. Estos procesos incluyen tareas como la inspección, limpieza, modelado de los datos, así como su representación visual.

Para la adecuada realización de dichas tareas no sólo se deben aplicar principios correctos de diseño visual sino también adecuados algoritmos de cálculo y procesado de los datos. Con ellos se pueden realizar modelos de los datos en los que se puede extraer conocimiento. La *minería de datos* [205, 61] es el área que utiliza elementos de otros campos como son la inteligencia artificial, la exploración de bases de datos o el reconocimiento estadístico de patrones, todos ellos aplicados para el estudio de grandes conjuntos de datos y la extracción de información útil para mejorar y descubrir conocimiento referido a esos datos analizados.

Muchos de los algoritmos utilizados para este propósito son aquellos que realizan un **aprendizaje automático** de los datos [18, 55], el cual sirve para crear un modelo de ellos. En la mayoría de las ocasiones no somos totalmente conscientes de la frecuencia con la que estos algoritmos se aplican. Por ejemplo cada vez que se envía o recibe un *email*, se utiliza una tarjeta electrónica en una transacción bancaria, o se compra un artículo en alguna página web, existe un algoritmo de aprendizaje detrás que utiliza los datos para mejorar la eficiencia de la tarea, como por ejemplo optimizar el uso del correo, detectar si una transacción es fraudulenta o realizar

una recomendación de algún producto en base a los intereses del comprador.

En un escenario típico se desea predecir una variable de salida, cuantitativa o cualitativa, usando como datos de entrada muestras definidas por una serie de variables, atributos, o características. Con un conjunto de datos de entrenamiento, se construye un modelo que permite predecir tal salida para nuevos datos de entrada.

La gran variedad de algoritmos existentes se pueden categorizar de diferentes formas, una de ellas podría ser dependiendo del tipo de aprendizaje que se aplique [81]. En el aprendizaje *supervisado*, los algoritmos aprenden con unos datos de entrada y con sus respuestas conocidas, para luego inferir un modelo que proporcionará respuestas a nuevos datos. En el aprendizaje *no supervisado*, el propio algoritmo aprende únicamente de los datos de entrada y determinará directamente estructuras y patrones contenidos en ellos sin medidas conocidas de la variable de salida.

Existen más tipos de aprendizaje que los descritos anteriormente, por ejemplo una mezcla de ambos denominado semi-supervisado [36], pero se ha realizado de esta forma general para ilustrar de manera simplificada el concepto. Otros ejemplos más recientes de diferentes tipos de aprendizaje son:

- *Aprendizaje por refuerzo* [8, 101]. En el cual se define un determinado entorno como un proceso de decisión de Markov (MDP), que consta de un conjunto de estados, acciones, reglas de transición entre los estados y de recompensas. En dicho entorno un agente cambia su estado mediante una acción a través de su correspondiente transición entre estados, evaluados por funciones de recompensas. El objetivo principal es elegir acciones en el tiempo para maximizar el valor esperado de una función total de recompensa.

- Los *sistemas de recomendación* [155, 1] producen una lista de recomendaciones mediante dos enfoques generales: filtrado colaborativo, es decir, construir un modelo con comportamientos pasados de usuarios y sus preferencias sobre objetos, que pueden ser interesantes para nuevos usuarios con similares comportamientos; y filtrado en base del contenido, el cual usa una serie de características de un objeto para recomendar otro de similares características. También existen más variantes, como por ejemplo una combinación de ambos enfoques formando un sistema híbrido [31].

- *Aprendizaje ensemble* [52] utiliza varios modelos de forma conjunta para obtener una predicción global. Se apoya en el aprendizaje supervisado, combinando un conjunto de varias hipótesis para un problema particular, de forma que crea predictores sencillos para producir en conjunto una predicción mejor. Ejemplos que utilizan este tipo de aprendizaje son métodos como *bagging* [25] y *boosting* [69] que combinan múltiples modelos construidos de diferente forma los cuales se complementan unos a otros, o *random forest* [26] que construyen una multitud de árboles de decisión que se combinan para realizar una predicción final.

Existen muchos tipos de algoritmos disponibles para aplicar un aprendizaje automático a los datos de un determinado problema [207, 79]. En [55] se argumenta que en el gran espacio de algoritmos disponibles, se comparten fundamentalmente tres componentes:

- *Representación.* Consiste en la forma de representar formalmente el problema de manera que pueda manejarse por un ordenador. Comúnmente se denomina espacio de hipótesis. Una cuestión relacionada es cómo representar los datos de entrada y salida, que se discute más adelante.

- *Evaluación.* Una función de evaluación, también denominada función objetivo, se necesita para definir el rendimiento del algoritmo. La función de evaluación interna puede ser distinta de la función externa que queremos que se optimice (también llamada función de coste).

- *Optimización.* Finalmente, se necesita un método para encontrar la mejor solución final. La elección de una técnica de optimización es crucial para la eficiencia del método, y también para decidir si posee más de una solución posible.

3.2 APRENDIZAJE SUPERVISADO

A continuación se ilustrarán los componentes comentados anteriormente con simples ejemplos de aprendizaje supervisado.

Suponiendo un grupo de datos de entrada que consta de un número de muestras N, se define un vector $\mathbf{x}_i$ de características, en el que cada muestra i está compuesta por D atributos o variables, que determina la dimensión del conjunto de datos ($\mathbf{X} = [x_{ij}] \in \mathbb{R}^{N \times D}$). De la misma forma, suponemos una variable objetivo de salida y_i para cada muestra ($\mathbf{y} \in \mathbb{R}^N$). A partir

de un grupo de entrenamiento de n datos de entrada ($n < N$) y sus conocidas respuestas de salida, se tiene la pareja $\{\mathbf{x}_i, y_i\}$ con $i = 1, \ldots, n$. El algoritmo "aprenderá" la función de un modelo supuesto que mejor produzca las salidas en función de las entradas. Dependiendo de que esta variable de salida sea cualitativa (también categórica o discreta) o cuantitativa (continua), la tarea de predicción se denomina *clasificación* o *regresión* respectivamente.

Una etapa fundamental es obtener una **representación** formal, por ejemplo mediante la función $f(\mathbf{x}_i)$, de forma que la estimación de la salida sea $\hat{y}_i = f(\mathbf{x}_i)$. Esta función se compone de un conjunto de transformaciones h del vector de entrada, que expresan el conjunto de hipótesis, donde se asocian un conjunto de parámetros $(\theta_0, \ldots, \theta_q)$ que se modifican para ajustarse a los datos de entrenamiento. Para simplificar, suponemos para el parámetro θ_0 (también denominado *bias*) un término igual a 1. Una formulación muy común es el modelo lineal en los parámetros:

$$f(\mathbf{x}_i) = \theta_0 + \sum_{j=1}^{q} \theta_j \cdot h_j(\mathbf{x}_i) \tag{4}$$

Por ejemplo, en el caso sencillo de una regresión lineal, las hipótesis están descritas por las propias variables de los datos, de tal forma que $h_j(\mathbf{x}_i) = x_{ij}$, con $j = 0, \ldots, q$ (siendo en este caso $q = D$) y suponiendo $x_{i0} = 1$ como se indicó anteriormente. Por tanto, la función representada sería una combinación lineal de los atributos de los datos $f(\mathbf{x}_i) = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_D x_{iD} = \theta^T \mathbf{x}_i$. Otros ejemplos para las funciones h pueden ser polinómicas (x_{i1}^2, $x_{i1} x_{i2}^2$), trigonométricas ($\cos(\mathbf{x}_i)$), etc. También pueden ser no lineales con más parámetros asociados, como por ejemplo en el caso de una *sigmoide* (Ec. 5) utilizada para redes neuronales o regresión logística.

$$h(\mathbf{x}_i, \gamma) = \frac{1}{1 + e^{-\gamma^T \mathbf{x}_i}} \tag{5}$$

Existen otras maneras diferentes de realizar esta representación. Algunos ejemplos son los árboles de decisión [153], un conjunto de reglas de asociación [38], usando el teorema de Bayes [70], o los k-vecinos más próximos [44, 2] mediante una función de distancia.

Otro componente importante en la creación del modelo es la **evaluación** del algoritmo automático aplicada a los datos de entrenamiento. Para ello, es esencial tener en cuenta una *función objetivo* (o de coste) que evalúe los errores producidos. Por ejem-

plo, uno de los más populares para un modelo lineal es la suma de los errores cuadrados:

$$J(\theta) = \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2 \tag{6}$$

También es posible realizar una evaluación, una vez finalizado el modelo, con un nuevo conjunto de datos de *test* para determinar su rendimiento. Esto permite elegir qué método es más adecuado aplicar a un determinado problema. Otros ejemplos de evaluación incluyen cualquier medida numérica que evalúe la predicción, parámetros como la precisión y sensibilidad [159], curvas ROC [60], o mediante el estudio de probabilidades como en el caso de la divergencia de *Kullback-Leibler* [113].

Finalmente los parámetros son estimados mediante una etapa de **optimización** donde se busca obtener la parametrización que proporcione el mejor resultado posible en la evaluación. Por tanto, los coeficientes θ pueden calcularse minimizando el error dado por la función de coste mediante numerosos métodos [67, 24]. Para el caso del modelo lineal en los parámetros descrito en (4), se define una matriz $\mathbf{H} \in \mathbb{R}^{n \times q}$ de tal forma $\mathbf{H} = [h_j(\mathbf{x}_i)]$, y desarrollando de manera algebraica la ecuación (6), se obtienen las *ecuaciones normales* cuya solución de forma matricial es la siguiente:

$$\theta = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y} = \mathbf{H}^+\mathbf{y} \tag{7}$$

donde $\mathbf{H}^+$, es la matriz *pseudoinversa* de $\mathbf{H}$.

Cuando la función no es lineal en los parámetros, otros métodos de optimización pueden utilizar heurísticos o requerir una etapa de convergencia iterativa. Un método muy utilizado es el llamado *descenso de gradiente*, y sus variantes, que utiliza el cálculo de derivadas para alcanzar un mínimo local en su convergencia. Este método es aplicado en muchos algoritmos que minimizan funciones de coste, por ejemplo en el llamado *backpropagation*, muy utilizado en el entrenamiento de redes neuronales [17]. Otra clase de métodos son los denominados *quasi-Newton* que calculan máximos y mínimos locales de funciones de varias variables encontrando sus ceros por el método de Newton, en donde también se pueden incluir métodos numéricos como por ejemplo el algoritmo Broyden–Fletcher–Goldfarb–Shanno (*BFGS*) [67, 24] o el Levenberg–Marquardt [132].

A continuación se describen varias técnicas automáticas de análisis que fueron utilizadas en la realización de esta tesis y serán mencionadas en posteriores capítulos.

3.2.1 *Redes de base radial*

Las redes de funciones de base radial (*radial basis functions*, RBF) [17, 82] son un popular tipo de redes neuronales con una topología típica de una capa de entrada, una capa oculta y otra de salida. Considerando n vectores de entrada $\mathbf{x}$, el conjunto h se representa mediante unas funciones de activación llamadas funciones de *base radial*, $\phi_j(\mathbf{x}) = g(||\mathbf{x} - \mathbf{c}_j||)$, que calculan la distancia de un punto determinado a un centro $\mathbf{c}_j$, con $j = \{1, \ldots, q\}$, siendo un caso particular la elección de q igual al número n de puntos de entrada. La salida que proporciona la red corresponde a una combinación lineal de las funciones $\phi_j(\mathbf{x})$ de las entradas con los parámetros w_j de los pesos de salida.

$$F(\mathbf{x}) = \sum_{j=1}^{q} w_j \cdot \phi_j(\mathbf{x}) = \sum_{j=1}^{q} w_j \cdot g(||\mathbf{x} - \mathbf{c}_j||) \tag{8}$$

Entre los tipos más utilizados de funciones RBF se encuentra la Gaussiana $\phi_j(\mathbf{x}) = e^{-\frac{||\mathbf{x}-\mathbf{c}_j||^2}{2\sigma^2}}$. Mientras que las neuronas de la capa oculta poseen un carácter local con transformaciones no lineales, las neuronas de salida realizan una combinación lineal de las activaciones de las neuronas ocultas. De manera matricial, definiendo $\mathbf{y} \in \mathbb{R}^n$ como variable objetivo, y la matriz $\mathbf{H} = \phi_j(\mathbf{x}_i) = g(||\mathbf{x}_i - \mathbf{c}_j||) \in \mathbb{R}^{n \times q}$, entonces la matriz de pesos $\mathbf{W} \in \mathbb{R}^q$ se puede calcular mediante la pseudoinversa de $\mathbf{H}$, como en la ecuación (7), de la forma $\mathbf{W} = \mathbf{H}^+\mathbf{y}$.

Este tipo de redes se aplicaron inicialmente para problemas de interpolación, también se han utilizado en multitud de campos como por ejemplo procesamiento de imágenes, o análisis de series temporales. Una ventaja de este tipo de redes es que los centros y anchos de las funciones RBF se pueden determinar independientemente de los pesos de salida. Un inconveniente es que dan la misma importancia a todas las variables de los datos, a menos que sean incluidos en la etapa de optimización. Por tanto, sin una modificación adecuada no podrían tratar con variables irrelevantes de forma efectiva.

3.2.2 *Extreme learning machine*

Aunque las redes neuronales basadas en modelos no lineales, como los perceptrones de dos o más capas ocultas, han sido utilizadas en numerosas aplicaciones, poseen algunos inconvenientes como la convergencia a mínimos locales o el excesivo tiempo que

emplean en su entrenamiento. El algoritmo *extreme learning machine* (ELM) [92] consiste en una red neuronal "feedforward" de perceptrón multicapa compuesta por q neuronas en la capa oculta, que aproxima cualquier función continua con error nulo. Considerando n vectores de entrada $\mathbf{x}_i$ y las variables objetivo $\mathbf{t}_i \in \mathbb{R}^m$, la red estándar se modela mediante las funciones de activación g, utilizando unos pesos de entrada $\mathbf{w}_j$, y los parámetros umbral b_j, tal que $g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j)$, y también mediante unos pesos de salida β_j. Que esta red aproxime las n muestras con un error nulo significa que existen β_j, $\mathbf{w}_j$ y b_j tal que

$$\sum_{j=1}^{q} \beta_j \cdot g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = \mathbf{t}_i, \quad i = 1, \ldots, n. \tag{9}$$

Los valores para los pesos de entrada $\mathbf{w}_j$ y los parámetros b_j se pueden iniciar con valores aleatorios si las funciones de activación son diferenciables. Estas ecuaciones se pueden escribir como $\mathbf{H}\beta = \mathbf{T}$ donde $\mathbf{H} \in \mathbb{R}^{n \times q}$ es la matriz de salida de la capa oculta de la red neuronal, $\beta \in \mathbb{R}^{q \times m}$ es la matriz de pesos de salida y $\mathbf{T} \in \mathbb{R}^{n \times m}$ la matriz objetivo de los n casos de entrenamiento. Por tanto, la red puede considerarse como un sistema lineal cuyos pesos de salida (aquellos que enlazan la salida con la capa oculta) se pueden determinar analíticamente mediante una operación generalizada de inversa de matrices. Por tanto, los pesos de salida se calculan $\hat{\beta} = \mathbf{H}^{+}\mathbf{T}$, siendo $\mathbf{H}^{+}$ la pseudoinversa de la matriz $\mathbf{H}$.

El ELM mejora el rendimiento de redes neuronales de perceptrón multicapa ya que proporciona un entrenamiento rápido de manera eficiente. El parámetro necesario que se debe establecer es el número de neuronas de la capa oculta, aunque existen algoritmos automáticos para su selección óptima [138].

3.3 APRENDIZAJE NO SUPERVISADO

El otro gran tipo de aprendizaje automático en algoritmos que modelan datos de entrada, es el aprendizaje no supervisado, el cual se realiza directamente con los datos de entrada, sin tener en cuenta etiquetas o variables de salida que lo "supervisen". El principal objetivo es la detección de patrones incluidos en los datos.

Entre los principales problemas que se destacan se encuentra el análisis de grupos (*clusters*), en los que se pueden obtener descripciones que definen diferentes clases en los datos. Los posibles estudios que pueden hacer de las propiedades de los datos incluyen su caracterización estadística, por medio de la estimación de

la distribución de probabilidad que los definen, o la densidad de área que ocupan los datos en el espacio multidimensional.

Una técnica clásica para el modelado de la función de densidad de probabilidad de grandes volúmenes de datos es la *cuantización vectorial* (VQ). Esta técnica divide el conjunto de vectores de datos en un menor grupo de prototipos, los cuales representan de manera fiable el conjunto total. Originalmente fue desarrollado para la compresión de datos pero también se utiliza para el análisis de agrupamientos en los datos y la identificación de una clasificación natural.

3.3.1 *K-means*

Uno de los algoritmos más populares para encontrar agrupamientos en los datos es *K-means* [129], el cual ha inspirado posteriormente numerosos desarrollos dentro del análisis no supervisado de grupos en los datos [95].

Dado un grupo de N muestras $\mathbf{x}_i$, donde cada una es un vector D-dimensional, se considera un valor para el parámetro K del algoritmo ($K \leq N$) que indica el número de grupos (*clusters*) $Q = \{Q_1, Q_2, \ldots, Q_K\}$ en los que agrupará las N muestras. Para ello se minimiza la función de coste de la manera siguiente:

$$\arg\min_{\mathbf{Q}} \sum_{i=1}^{K} \sum_{\mathbf{x}_j \in Q_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2 \tag{10}$$

donde $\boldsymbol{\mu}_i$ es la media de los puntos en Q_i, denominados los *centroides* de cada grupo. En cada etapa se actualiza la posición de estos centroides hasta que llega a la convergencia.

Un resumen básico del algoritmo podría ser el siguiente:

1. Colocar de manera aleatoria tantos puntos como grupos (K), que serán los *centroides* de cada grupo $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_K$.

2. Se calculan las distancias a cada punto de los centroides, y se asignan los más cercanos a cada uno de ellos.

3. Se calculan las medias de todos los correspondientes puntos asignados a cada centroide.

4. Repetir pasos 2 y 3 hasta que no se modifiquen (convergencia).

En la Fig. 3.1 se representan unos datos a modo de ejemplo (300 puntos de dos dimensiones) y la progresión de 3 centroides (en

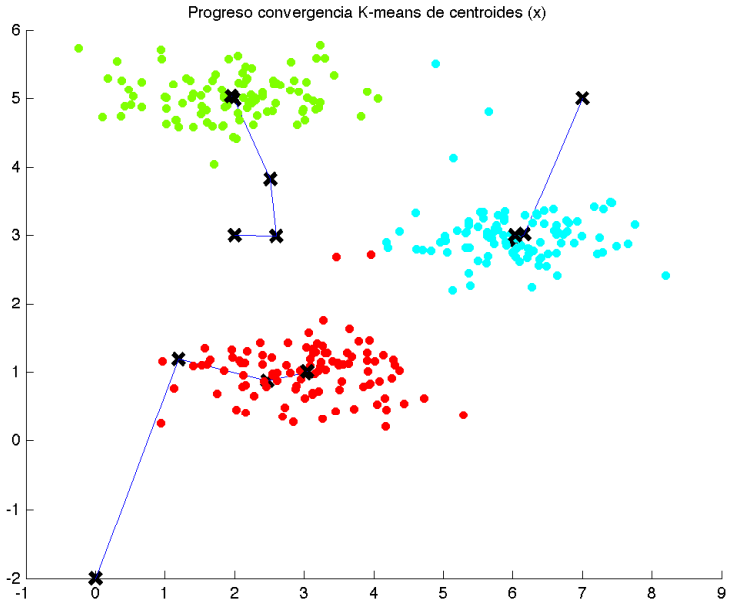


Figura 3.1.: Ejemplo de convergencia (10 iteraciones) del algoritmo *K-means* aplicado para 3 grupos (centroides en forma de x).

forma de x), en 10 iteraciones del algoritmo unidos con una línea representando su posición en cada iteración. Se puede ver que los centroides comienzan en una posición aleatoria y van cambiando su posición para representar 3 agrupaciones en los datos.

No obstante existen varias limitaciones. La inicialización aleatoria puede llevar a que el algoritmo converja en soluciones distintas dependiendo de esa etapa inicial, es decir, que la convergencia finalice en distintos óptimos locales de la función de coste válidos para esa etapa en concreto, pero que no son la mejor solución (óptimo global). Se pueden hacer múltiples ejecuciones con distintas inicializaciones y considerar las que dan muchos resultados iguales, puesto que es más probable que sea el óptimo global. Otra limitación es el propio parámetro K que debe ser fijado de antemano, lo cual hace necesario fijar cuántos grupos representan los datos, cosa que puede ser desconocida.

3.3.2 *Neural gas*

El algoritmo *neural gas* [133, 134] aproxima datos en un espacio de entrada en forma de vectores de características, mediante un

conjunto de prototipos los cuales se adaptan para que representen su distribución de probabilidad. Es muy utilizado en compresión de datos o cuantización vectorial, y también en el análisis de grupos aportando más robustez que el algoritmo *K-means* en la etapa de convergencia.

Dada una distribución de probabilidad $P(\mathbf{x})$ de vectores de datos $\mathbf{x}$ y un número finito de vectores prototipo $\mathbf{w}_i$ con $i = 1, \ldots, m$, se calculan los órdenes de los prototipos en relación a su proximidad a cada dato de entrada $(\mathbf{w}_{i0}, \mathbf{w}_{i1}, \ldots, \mathbf{w}_{im-1})$ siendo $\mathbf{w}_{i0}$ el más cercano al dato $\mathbf{x}$. Cada vector prototipo se adapta con la relación siguiente para cada paso t de cada iteración:

$$\mathbf{w}_{i_k}^{t+1} = \mathbf{w}_{i_k}^{t} + \alpha(t) \cdot e^{-k/\lambda} \cdot (\mathbf{x} - \mathbf{w}_{i_k}^{t}) \tag{11}$$

donde $\alpha(t)$ es el coeficiente de convergencia, λ el rango de vecindad y k el orden del vector prototipo $\mathbf{w}_i$ con respecto al dato $\mathbf{x}$.

Este algoritmo se ha planteado más adelante mediante un cálculo por lotes (modo *batch*) [43], de forma que la adaptación se hace para un conjunto de datos (o el total) a la vez:

$$\mathbf{w}_i^{t+1} = \frac{\sum_{j=1}^{N} h_i(\mathbf{w}_i, \mathbf{x}_j) \cdot \mathbf{x}_j}{\sum_{j=1}^{N} h_i(\mathbf{w}_i, \mathbf{x}_j)} \tag{12}$$

Siendo $h_i(\mathbf{w}_i, \mathbf{x}_j)$ la función de vecindad, para el caso anterior de la ecuación (11) la función de vecindad sería $h_i = e^{-k/\lambda}$. Nótese la analogía existente con el método *K-means* puesto que en este caso la función de vecindad sería 1 si pertenece al grupo asignado y 0 en otro caso, lo que daría el cálculo de la media de los puntos del grupo correspondiente.

3.4 TÉCNICAS DE REDUCCIÓN DE LA DIMENSIÓN

Considerar un conjunto de datos con un extenso número de variables nos permite ejecutar una gran variedad de tareas complejas a la hora de analizar los datos de un determinado proceso. Tener en cuenta la independencia entre variables y realizar una simplificación en un pequeño número con menor redundancia ayuda a un eficiente manejo y entendimiento de dichos datos. Esto hace que la **reducción de la dimensión** (DR) [119] sea una de las herramientas más importantes en el análisis de datos multivariables.

Idealmente, un método DR debería estimar el número de las variables latentes en los datos, determinada por la dimensión intrínseca de los datos, la cual revela su estructura topológica. Si

esa dimensión es menor que la propia de los datos, los puntos se encuentran en un subespacio, variedad dimensional o también denominado *manifold*. La estimación de esta estructura es lo que realizan los métodos DR con el objeto de un procesado más sencillo de los datos y una representación más compacta de los mismos. Por tanto, entre las aplicaciones típicas se encuentran la compresión de datos y su visualización.

El problema puede plantearse [192] suponiendo que tenemos N puntos/muestras de datos multivariables $\mathbf{x}_i$ en un espacio de entrada de dimensión D, formando una matriz $\mathbf{X} \in \mathbb{R}^{N \times D}$. Dichos datos poseen una dimensionalidad intrínseca d ($d < D$) lo cual quiere decir que los puntos se encuentran formando un *manifold* integrado en el espacio de entrada. Los métodos DR estiman esta estructura y proyectan los datos en el espacio d-dimensional, las coordenadas de los puntos proyectados se denotan como $\mathbf{y}_i \in \mathbb{R}^d$. La reducción de la dimensión organiza los datos de tal forma que la estructura latente se conserva. El problema radica en la caracterización de la estructura del *manifold* para conservarlo.

A continuación se revisarán, de manera breve, un grupo de las técnicas más importantes que son capaces de realizar una reducción de la dimensión.

3.4.1 *Análisis de componentes principales*

El análisis de componentes principales (*principal component analysis*, PCA) [100] es la primera y más popular técnica DR lineal, muy utilizada para análisis exploratorio de datos. Presentado inicialmente por Pearson [146], Hotelling la desarrolló [90] en el campo de la psicometría, más tarde fue redescubierto por Karhunen [102] y Loève [126], por lo que también se le conoce como la transformación "Karhunen-Loève".

Básicamente, PCA reduce las dimensiones de los datos de entrada encontrando combinaciones lineales ortogonales de los datos originales con mayor varianza. Se realiza mediante el cálculo de la matriz de covarianzas centrada en el origen, es decir, de los datos con media cero. Para centrar los datos, se puede hacer restando la esperanza de los datos en cada muestra, aproximada por la media de cada variable. Siendo la covarianza entre dos variables aleatorias x, y,

$$cov(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$$

entonces la matriz de covarianzas ($\mathbf{\Sigma}$) consta de las covarianzas entre las dimensiones (i, j) de los datos, de tal forma que:

$$\mathbf{\Sigma} = c_{ij} \quad \text{con} \quad c_{ij} = cov(\mathbf{x}^i, \mathbf{x}^j) \tag{13}$$

siendo $\mathbf{x}^i$, la columna i de la matriz $\mathbf{X}$. Considerando la matriz $\mathbf{X}$ de datos centrados, también se puede expresar la matriz $\mathbf{\Sigma}$ (simétrica y definida positiva) mediante una descomposición espectral $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, siendo $\mathbf{\Lambda} = diag(\lambda_1 \dots \lambda_D)$ la matriz diagonal de valores propios ordenados y $\mathbf{U}$ la matriz ortogonal $D \times D$ conteniendo los vectores propios, cuyos primeros p vectores se utilizan para la proyección de los datos en el nuevo espacio de dimensión p.

$$\mathbf{y}_i = \mathbf{U}^T\mathbf{x}_i \tag{14}$$

A esta descomposición espectral en vectores y valores propios también se denomina *descomposición de valores singulares* (SVD). Este método DR se puede enfocar por medio de diferentes criterios obteniéndose el mismo resultado, como el error de reconstrucción mínimo y máxima preservación de la varianza, utilizados por Pearson y Hotelling respectivamente.

Su simplicidad y la rapidez para obtener el resultado hace que PCA sea el primero de los métodos a utilizar para el análisis de datos multivariables, también es aplicable para la estimación de la dimensión intrínseca, dada por el salto producido en los valores propios, para la proyección de una reducción de la dimensión, y la separación de las variables latentes. La principal limitación del método es que supone una dependencia lineal de las variables observadas, así como una separación de variables más restrictiva para otras distribuciones distintas a la Gaussiana. Estos inconvenientes llevan a la búsqueda de nuevos métodos para la consideración de dichas distribuciones o para una reducción de la dimensionalidad no lineal.

3.4.2 *Escalamiento multidimensional*

La familia de métodos de escalamiento multidimensional (*multidimensional scaling*, MDS) [45] representa una colección de técnicas no lineales que crean un mapa entre los datos en alta y baja dimensión (espacio de entrada y salida respectivamente) conservando, tanto como sea posible, la distancia de los puntos entre ambos espacios. Existen varios tipos de distancias, la más utilizada es la distancia euclídea por las ventajosas propiedades que posee. Por

tanto, para los puntos de datos $\{\mathbf{x}_i\}_{i=1,\dots,N}$, con $\mathbf{x}_i \in \mathbb{R}^D$, la distancia euclídea entre la pareja de puntos i y j viene definida por

$$d(\mathbf{x}_i, \mathbf{x}_j) = ||\mathbf{x}_i - \mathbf{x}_j|| = \sqrt{\sum_{k=1}^{D} (x_{ik} - x_{jk})^2} \tag{15}$$

El primer enfoque realizado es el *clásico MDS métrico*, en el que los puntos proyectados $\{\mathbf{y}_i\}_{i=1,\dots,N} \in \mathbb{R}^d$ se obtienen calculando la descomposición espectral de la matriz de productos escalares o *matriz de Gram* centrada. En el caso de distancia euclídea proporciona resultados similares a PCA, minimizando el mismo criterio ya comentado, es decir, la diferencia de distancias en ambos espacios, expresado matemáticamente en la siguiente función de coste:

$$E_{MDS} = \sum_{ij} (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 \tag{16}$$

Tal como sucede en el caso de PCA, la solución se obtiene de manera algebraica, siendo también en este caso lineal. Por tanto, este método posee las mismas ventajas e inconvenientes que PCA, además de un mayor consumo de memoria por el mayor tamaño que tiene la matriz de Gram. Existen diferentes variantes del método, teniendo en cuenta enfoques distintos al producto escalar, minimizada con respecto a distancias entre puntos para obtener una representación espacial. A continuación se presenta el mapa no lineal de *Sammon* por su amplia aplicación en diferentes áreas.

Mapeo no lineal de Sammon

El algoritmo del mapeo no lineal de *Sammon* (*Sammon's nonlinear mapping*, NLM) [160] reduce la dimensión de un grupo finito de puntos de manera similar que un método MDS, minimizando la siguiente función de coste:

$$E_{NLM} = \frac{1}{\sum_{i<j} d(\mathbf{x}_i, \mathbf{x}_j)} \sum_{i<j} \frac{(d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2}{d(\mathbf{x}_i, \mathbf{x}_j)} \tag{17}$$

Es una de las numerosas variantes de los métodos MDS, considerando la anterior función de coste, denominada también función de *stress*, que también sirve para medir la calidad del mapa de puntos obtenido en la reducción. Su principal ventaja es que es un método no lineal, y su inconveniente es que su método de optimización puede ser lento para algunos conjuntos de datos, incluso quedándose en un mínimo local.

3.4.3 *Los métodos kernel*

Los métodos *kernel* [166, 164] son una clase de algoritmos para el análisis general de patrones en diversos tipos de datos. Usando funciones que miden la similaridad no lineal entre patrones, el espacio de entrada se transforma en un espacio euclídeo, definido por medio de los productos escalares entre sus puntos.

Se puede definir una función simétrica kernel $k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$, suponiendo que existe un mapa $\Phi : \mathbb{R}^D \to \mathcal{H}$ en un espacio de características $\mathcal{H}$, definido para toda pareja de puntos $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^D$ de forma que:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \tag{18}$$

Las funciones kernel deben ser continuas y simétricas, semidefinidas positivas, verificándose que $\langle \cdot, \cdot \rangle$ es un producto interno (escalar). Existen varios tipos de funciones kernel, los casos más típicos son polinómicas $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + c)^p$ siendo $c \geq 0$ y p un entero positivo, o gaussianas $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/\sigma^2)$, y su correcta elección depende del problema que se trate, en un primer lugar puede ser arbitraria o intuitiva dependiendo de la información que esperamos extraer de los datos. Este método, también denominado truco del kernel (*kernel trick*) permite trabajar implícitamente en espacios de dimensión infinita donde el problema es lineal, y utilizar técnicas de minimización con restricciones. Tendremos, en la dimensión D, una solución no lineal al problema.

Aunque su aplicación más conocida se refiere a los *support vector machines* (SVM) [30], en tareas de clasificación y regresión, este enfoque también puede ser utilizado en otros métodos de forma general aplicado como truco del kernel (*kernel trick*). Por ejemplo, su aplicación al tradicional PCA da lugar al llamado *Kernel PCA* (KPCA) [165] mediante el cálculo de la descomposición espectral de una matriz construida por una función de kernel en lugar de una matriz de covarianzas.

3.4.4 *Mapa topológico auto-organizado*

El mapa auto-organizado (*self-organizing map*, SOM) de Kohonen [110] es una de las redes neuronales más utilizadas para el análisis y visualización de datos multidimensionales. En primer lugar, el SOM realiza una cuantización vectorial, reemplazando el conjunto de puntos por un grupo menor de *prototipos*, es decir, neuronas representativas conectadas entre sí formando una malla. Estos prototipos se adaptan aproximando la distribución de los datos de entrada. Intuitivamente, esa malla se deforma de manera

ordenada. Considerando una retícula, homóloga a los prototipos, como un espacio de salida discreto, el SOM realiza un proyección no lineal de los datos preservando su topología.

Para ello, un conjunto de datos en el espacio de entrada $\mathbf{X} \in \mathbb{R}^{N\times D}$, se considera representado por el conjunto de prototipos $\{\mathbf{m}_i\}_{i=1,\dots,M}$ con $\mathbf{m}_i \in \mathbb{R}^D$ y por sus correspondientes vectores homólogos en el espacio de salida $\{\mathbf{g}_i\}_{i=1,\dots,M}$ con $\mathbf{g}_i \in \mathbb{R}^d$, habitualmente dispuestos formando una retícula bidimensional rectangular ($d = 2$). Esta retícula es la que define la proyección de los datos, de forma que para cada punto de entrada $\mathbf{x}$, su proyección viene dada por el punto de la retícula $\mathbf{y} = \mathbf{g}_c$, cuyo correspondiente prototipo $\mathbf{m}_c$ es el más cercano a $\mathbf{x}$ en el espacio de entrada.

$$c = \arg\min_i\{d(\mathbf{x}, \mathbf{m}_i)\} \tag{19}$$

Para determinar las coordenadas de $\mathbf{m}_i$ de manera que mejor definan la distribución de probabilidad de los datos de entrada, se minimiza iterativamente el error de la cuantificación en un determinado número de *épocas*. Para ello se propone en [110] un método similar a un descenso de gradiente usando una función de vecindad $h_{ci}(t)$ definida por una función de kernel en el espacio de salida, generalmente gaussiana. Para mejorar la eficiencia en el entrenamiento del SOM, existe una versión del algoritmo por lotes (*batch*) en la que se utilizan todos los datos simultáneamente, mediante el cálculo de una media ponderada. Por tanto, para un conjunto $\{\mathbf{x}_j\}_{j=1,\dots,N}$ con $\mathbf{x}_j \in \mathbb{R}^D$, se obtienen los coeficientes de los prototipos de la forma siguiente:

$$\mathbf{m}_i(t+1) = \frac{\sum_{j=1}^{N} h_{c(k)i}(t) \cdot \mathbf{x}_j}{\sum_{j=1}^{N} h_{c(k)i}(t)} \tag{20}$$

De esta manera se obtiene un mapa con una cuantización vectorial mediante un proceso simultáneamente competitivo y colaborativo, que ajusta a los datos de entrada una estructura preservando a la vez propiedades topológicas de la retícula definida por h_{ci}, en la que se reduce la dimensión de forma no lineal para la visualización de la información contenida en los datos. Es un método robusto y simple, que no requiere una gran complejidad de computación, aplicable a una numerosa cantidad de situaciones.

Mapa topográfico generativo

El mapa topográfico generativo (*generative topographic mapping*, GTM) [19] proporciona una alternativa al SOM utilizando un enfoque probabilístico. GTM calcula una función no lineal para realizar

un mapeo entre el espacio de visualización (una rejilla discreta de puntos similar al SOM) y el espacio de entrada, construyendo un modelo de los datos mediante una mezcla de funciones Gaussianas. La función de máxima verosimilitud del modelo es optimizada por medio del método *expectation-maximization* (EM) [49], adecuado para el caso de trabajar con mezclas de kernel gaussianos.

3.4.5 *Análisis de componentes curvilíneas*

Demartines y Hérault propusieron en 1997 el análisis de componentes curvilíneas (*curvilinear component analysis*, CCA) [48] como mejora del mapa auto-organizativo de Kohonen cuando se utiliza como método DR. Se encuentra dentro de los métodos que realizan una reducción de la dimensión no lineal mediante preservación de distancias, minimizando una función de coste como la siguiente:

$$E_{CCA} = \frac{1}{2}\sum_{\substack{i=1\\j=1}}^{N}(d(\mathbf{x}_i,\mathbf{x}_j) - d(\mathbf{y}_i,\mathbf{y}_j))^2 F_\lambda(d(\mathbf{x}_i,\mathbf{x}_j)) \qquad (21)$$

Siendo $F_\lambda(d(\mathbf{x}_i,\mathbf{x}_j)$ una función más general, elegida normalmente monótona decreciente en función de su argumento λ para favorecer la preservación topológica local. Ejemplos que se pueden utilizar para esta función pueden ser una función binaria, exponencial decreciente o sigmoide.

Calculadas las distancias entre los puntos de los datos, las coordenadas de la proyección se inicializan de manera aleatoria o por medio de sus componentes principales. Dado un factor de aprendizaje $\alpha(t)$ y un radio de vecindad λ para las distintas épocas de entrenamiento, se selecciona un punto y actualiza el resto durante la época actual, de acuerdo a las siguientes expresiones:

$$\begin{aligned} \mathbf{y}_j &\leftarrow \mathbf{y}_j - \alpha(t)\beta(i,j)\frac{\mathbf{y}(i)-\mathbf{y}(j)}{d(\mathbf{y}_i,\mathbf{y}_j)} \qquad (22)\\ \text{con } \beta(i,j) &= 2F_\lambda(d(\mathbf{y}_i,\mathbf{y}_j))(d(\mathbf{x}_i,\mathbf{x}_j) - d(\mathbf{y}_i,\mathbf{y}_j)) - \\ &\quad -(d(\mathbf{x}_i,\mathbf{x}_j) - d(\mathbf{y}_i,\mathbf{y}_j))^2 F'_\lambda(d(\mathbf{y}_i,\mathbf{y}_j)) \end{aligned}$$

Esto hace de CCA un método más flexible, por ejemplo en la elección de la función F_λ, o con una etapa opcional de cuantización vectorial.

3.4.6 *Isomap*

Isomap [182] sigue un procedimiento similar al clásico MDS, pero con el uso de una métrica distinta, una distancia de grafo en el espacio de entrada en lugar de la distancia euclídea, lo cual proporciona al método que sea no lineal.

Suponiendo una estructura curvilínea de los datos, el objeto de reducir la dimensión para la visualización de esa estructura de manera intuitiva es desenrollarla, pero usando la distancia euclídea, puntos que son lejanos en la estructura se calculan erróneamente como cercanos. Una forma más apropiada es calcular las distancias a lo largo del *manifold* con una medida distinta llamada distancia *geodésica*. Inicialmente se construye un grafo, en el que cada punto se conecta con sus vecinos más cercanos. Las distancias geodésicas se calculan de manera aproximada por medio del camino más corto entre los puntos del grafo, utilizando algoritmos como por ejemplo el de Dijkstra [53] o Floyd [68]. Las coordenadas de la proyección se obtienen aplicando el método MDS a la matriz de distancias resultante.

Es un potente método no lineal, pero no aplicable a todos los *manifold*, puesto que puede fallar con los que posean agujeros y también en aquellos que sean no conexos. También se pueden producir conexiones erróneas al construir el grafo, produciendo una inestabilidad topológica en forma de cortocircuitos.

La distancia geodésica también se ha combinado con otros métodos dando lugar a nuevas técnicas, como por ejemplo *geodesic NLM* [208], utilizando el algoritmo del mapeo no lineal de *Sammon* o también *curvilinear distance analysis* [117] que es la versión de CCA utilizando las distancias geodésicas en lugar de las euclídeas.

3.4.7 *Locally linear embedding*

A diferencia de otros métodos, *locally linear embedding* (LLE) [158] no utiliza un modelo, construye una representación de grafo (similar a *Isomap*) preservando las propiedades locales de un conjunto de N puntos. LLE calcula estas propiedades para cada punto $\mathbf{x}_i$ por medio de una combinación lineal de sus K vecinos más cercanos. Para ello se utiliza el error de reconstrucción $\varepsilon(w)$ del punto $\mathbf{x}_i$ a partir de los $\mathbf{x}_j$ con los pesos $w_{i,j}$:

$$\varepsilon(w) = \sum_{i=1}^{N} ||\mathbf{x}_i - \sum_{j} w_{i,j}\mathbf{x}_j||^2 \tag{23}$$

Donde los pesos $w_{i,j}$ indican la contribución del punto j de los datos para la reconstrucción del punto i y se calculan minimizando dicho error de la ecuación (23). En el espacio de salida, LLE calcula los puntos $\mathbf{y}_i$ que permiten reconstruir esta combinación lineal suponiendo linealidad local. Por tanto, el error de reconstrucción análogo viene dado por la siguiente función de coste:

$$E_{LLE} = \sum_{i=1}^{N} ||\mathbf{y}_i - \sum_j w_{i,j}\mathbf{y}_j||^2 \tag{24}$$

Siendo $w_{i,j}$ los pesos de reconstrucción para el punto i calculados previamente en el espacio de entrada. La ecuación (24) puede minimizarse mediante un problema de valores propios, lo que da lugar finalmente a las proyecciones $\mathbf{y}_i$ de los datos.

3.4.8 *Laplacian eigenmaps*

De forma similar, el método *laplacian eigenmaps* (LE) [12] construye una proyección preservando las propiedades locales de los datos de entrada. LE mantiene las mismas condiciones de vecindad calculando la distancia entre los vecinos más cercanos. Esto se realiza con la construcción de un grafo ponderado con tantos nodos como número de puntos tengan los datos. Estos nodos se conectan a los vecinos más cercanos formando un grafo de adyacencia. A partir de estas conexiones se eligen sus pesos y se recogen en una matriz de pesos $\mathbf{W}$. Por ejemplo, una opción es utilizar funciones de kernel entre los puntos que estén conectados. Con la matriz de pesos $\mathbf{W}$, se crea también la matriz diagonal $\mathbf{D}$ a partir de la suma de sus filas (o columnas, puesto que $\mathbf{W}$ es simétrica). Teniendo en cuenta el laplaciano $\mathbf{L} = \mathbf{D} - \mathbf{W}$, la función de coste para el cálculo de la proyección $\mathbf{y}_i$ cumple la siguiente condición:

$$E_{LE} = \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 w_{ij} = 2\mathbf{Y}^t\mathbf{L}\mathbf{Y} \tag{25}$$

Cuya minimización puede obtenerse calculando la descomposición espectral del laplaciano $\mathbf{L}$. Los d vectores correspondientes a los valores propios más pequeños no nulos proporcionan la proyección $\mathbf{Y}$ de los puntos en el espacio d-dimensional de salida.

3.4.9 *Autoencoders*

Los *autoencoders* [111, 89] son redes neuronales (similares al *multilayer perceptron*) pero con la diferencia de que están construidas con el mismo número de nodos D en la entrada que en

la salida, coincidentes con la dimensión de los datos de entrada $X \in \mathbb{R}^D$. El proceso de aprendizaje de la red reconstruye en la salida las propias entradas. Los valores de un número menor de nodos d en una capa oculta, correspondientes para un punto de entrada $\mathbf{x}_i$, determinan la proyección $\mathbf{y}_i$ en el espacio d-dimensional de salida. Si las funciones de activación en la red neuronal son lineales, el autoencoder es similar a una proyección PCA, mientras que el uso de sigmoides lo convierte en un método DR no lineal.

La red posee unas capas de codificación (de X a Y) entrenadas normalmente por *Restricted Boltzmann Machines* (RBM's), después tiene una etapa de reconstrucción de la red (decodificador) formada por la inversa de las capas anteriores.

Este enfoque ha sido utilizado recientemente en el conjunto de algoritmos *deep learning* [14] que crean niveles de representación de las características de los datos mediante el aprendizaje en múltiples capas con transformaciones no lineales.

3.4.10 *Métodos neighbor embedding*

Más adelante han aparecido técnicas *neighbor embedding* (NE) que tienen en cuenta una preservación más genuina de las similitudes, lo cual comenzó con la técnica *stochastic neighbor embedding* (SNE) [88]. A diferencia de los métodos espectrales que convierten similitudes entre puntos de alta dimensión en productos internos, SNE combina similitudes calculadas en alta y baja dimensión. Este enfoque ha sido extendido con el desarrollo de diversas variantes, como por ejemplo *t-distributed SNE* (t-SNE)[190], NeRV [196], JSE [118], etc.

Teniendo un conjunto N de puntos se define $\{\mathbf{x}_i\}_{i=1,\dots,N}$ con $\mathbf{x}_i \in \mathbb{R}^D$ como su representación en un espacio D-dimensional, y de manera análoga $\{\mathbf{y}_i\}_{i=1,\dots,N}$ con $\mathbf{y}_i \in \mathbb{R}^d$ su representación en un espacio d-dimensional con $d < D$, y $d = 2$ o 3 para su visualización. La similitud se refiere generalmente a una cantidad inversamente relacionada con la distancia, normalmente euclídea.

La vecindad de los puntos de entrada se codifica mediante una matriz P, donde P_{ij} es proporcional a la probabilidad de que $\mathbf{x}_j$ sea un vecino de $\mathbf{x}_i$. De manera análoga, la matriz Q expresa la vecindad en el espacio de proyección de baja dimensión. La etapa de optimización consiste en minimizar la diferencia entre estas dos matrices de probabilidad $\mathcal{D}(P||Q)$ mediante una determinada divergencia $\mathcal{D}$, por ejemplo la comúnmente utilizada divergencia de *Kullback-Leibler* (KL). Con el resultado de la minimización, se obtienen las coordenadas de los puntos proyectados $\mathbf{y}_i$. Las distintas

elecciones de P, Q y $\mathcal{D}$ dan lugar a los diferentes métodos *neighbor embedding*. Por ejemplo, la matriz P se calcula de la forma siguiente:

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2/2\sigma_i^2)} \tag{26}$$

Puesto que la densidad de puntos varía, no existe un valor óptimo de σ_i para todos los puntos, es decir, para regiones densas sería más apropiado valores pequeños de σ_i y viceversa. Para solucionarlo, se puede realizar una búsqueda binaria del valor de σ_i que produce una distribución de probabilidad P_i con una medida constante fijada por el usuario, denominada *perplejidad*, y definida como $Perp(P_i) = 2^{H(P_i)}$, siendo $H(P_i)$ la entropía de *Shannon* utilizada en teoría de la información $H(P_i) = -\sum_x p_{j|i}\log_2 p_{j|i}$.

Para el caso t-SNE [190] se utiliza una versión simétrica del SNE, para ello se convierte $p_{ij} = \frac{p_{j|i}+p_{i|j}}{2N}$. También se diferencia en el uso de una distribución de probabilidad t-Student en lugar de la gaussiana (que se usan en los métodos SNE y NeRV) para calcular las similitudes q_{ij} en el espacio de salida. Por tanto, utilizando dicha distribución para un grado de libertad, las probabilidades para el mapa de puntos $\{\mathbf{y}_i\}_{i=1,\dots,N} \in \mathbb{R}^d$ vienen definidas como

$$q_{ij} = \frac{(1+\|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k\neq l}(1+\|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}} \tag{27}$$

Una vez calculadas estas distribuciones de probabilidad, sus diferencias son evaluadas mediante divergencias. De esta manera la proyección viene determinada por la minimización de una función de coste construida con esta divergencia. Por ejemplo, en las técnicas SNE y t-SNE se utiliza la divergencia *Kullback-Leibler* (KL) entre P y Q, definida por

$$D_{KL}(P_i||Q_i) = \sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}} \tag{28}$$

siendo la función de coste de la forma $E = \sum_i D_{KL}(P_i||Q_i)$. En NeRV [196], se utiliza una mezcla de dos divergencias KL como sigue

$$D_{KL}^{\lambda}(P_i||Q_i) = \lambda D_{KL}(P_i||Q_i) + (1-\lambda)D_{KL}(Q_i||P_i) \tag{29}$$

siendo $0 \leq \lambda \leq 1$ el parámetro que controla la importancia de ambos términos. De forma que $E_{NeRV} = \sum_i D_{KL}^{\lambda}(P_i||Q_i)$ define su función de coste. Otra manera de combinar divergencias es de la forma

$$D_{JS}^{\lambda}(P_i||Q_i) = \lambda D_{KL}(P_i||Z_i) + (1-\lambda)D_{KL}(Q_i||Z_i) \tag{30}$$

siendo $Z_i = \lambda P_i + (1-\lambda)Q_i$ y $0 \leq \lambda \leq 1$. Esta mezcla se conoce como la divergencia generalizada de *Jensen-Shannon*. Este tipo de mezcla de divergencias define una nueva función de coste que ha sido utilizada en la reciente técnica *Jensen-Shannon embedding* (JSE)[118].

La etapa de optimización se suele realizar minimizando la función de coste, por ejemplo usando el método de descenso de gradiente, para el caso t-SNE se minimiza $D_{KL}(P_i||Q_i)$ entre las distribuciones P y Q (Eq. 26 y 27) quedando de la forma final siguiente

$$\frac{\delta C}{\delta \mathbf{y}_i} = 4\sum_j (p_{ij} - q_{ij})(\mathbf{y}_i - \mathbf{y}_j)(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1} \tag{31}$$

Puesto que t-SNE utiliza distintas definiciones de las similitudes en alta y baja dimensión hace que no pueda proporcionar proyecciones isométricas para *manifold* lineales. NeRV combinando dos divergencias distintas, y JSE con una mezcla más compleja de divergencias, llevan a una función de coste paramétrica que mejora las proyecciones resultantes [118]. Además, en [209] se ha propuesto una optimización para la aplicación de estas técnicas a grandes conjuntos de datos, calculando la contribución de puntos cercanos de forma individual y de puntos lejanos aproximando por su "centro de masas".

3.5 TÉCNICAS SUPERVISADAS DE PROYECCIÓN

Aunque el enfoque DR clásico comenzó como un problema no supervisado, se han desarrollado varios algoritmos utilizando un aprendizaje supervisado. A continuación se describen algunos ejemplos de estos desarrollos.

3.5.1 *Linear discriminant analysis*

Linear discriminant analysis (LDA) [66] utiliza la información de clases y maximiza la separación entre los puntos que pertenecen a diferentes clases de manera lineal. Por tanto, la representación en el espacio de salida representa una separación lineal de las clases de los datos. Se obtiene la proyección lineal $\mathbf{y}$ de los datos $\mathbf{x}$ de forma que se cumpla una relación de este tipo $\mathbf{y} = \mathbf{w}^T\mathbf{x}$.

Este criterio define la separación como matrices de dispersión *entre las clases* (S_b) y dentro de la clase *interna* (S_w) de la forma siguiente

$$S_b = \sum_c (\boldsymbol{\mu}_c - \bar{\mathbf{x}})(\boldsymbol{\mu}_c - \bar{\mathbf{x}})^T \tag{32}$$

$$S_w = \sum_c \sum_{i \in c} (\mathbf{x}_i - \boldsymbol{\mu}_c)(\mathbf{x}_i - \boldsymbol{\mu}_c)^T \tag{33}$$

Siendo $\bar{\mathbf{x}}$ la media de los datos y $\boldsymbol{\mu}_c$ la media de cada clase. Se considera el criterio de *Fisher* para la definición de la función objetivo

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \tag{34}$$

La maximización de esta función puede resolverse como un problema de valores singulares general $S_b\mathbf{w} = \lambda S_w\mathbf{w}$. Donde los vectores propios forman las columnas de la transformación lineal que se aplica para la proyección en el espacio de salida de dimensión d, tomando los d valores propios más grandes.

LDA ha sido utilizado en tareas de clasificación en numerosas aplicaciones como predicción de quiebra bancaria, reconocimiento facial, marketing, etc.

Generalized Discriminant Analysis

De la misma forma que se describió con los métodos kernel anteriormente, *generalized discriminant analysis* (GDA) [9] es la reformulación del método LDA en el espacio de entrada utilizando una función kernel. De manera similar, GDA intenta maximizar el criterio de *Fisher*, construyendo el espacio de alta dimensión mediante este tipo de funciones. Por tanto, con esta técnica se construyen mapas no lineales maximizando la separabilidad de las clases de los datos.

3.5.2 *Neighbourhood components analysis*

Neighbourhood components analysis (NCA) [75] es un método supervisado que aprende una métrica de distancia encontrando una transformación lineal de los datos de forma que el rendimiento medio de una clasificación se maximiza en el espacio transformado. Esta clasificación predice la clase de un punto de los datos mediante consenso de sus vecinos más próximos usando una función de distancia determinada. Suponiendo la transformación A, se aprende

una métrica $Q = A^T A$ tal que $d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T Q(\mathbf{x}_i - \mathbf{x}_j) = (A\mathbf{x}_i - A\mathbf{x}_j)^T(A\mathbf{x}_i - A\mathbf{x}_j)$.

Para ello, se define la probabilidad entre puntos distintos de entrada de la manera siguiente

$$p_{ij} = \frac{\exp(-\|A\mathbf{x}_i - A\mathbf{x}_j\|^2)}{\sum_{k \neq i} \exp(-\|A\mathbf{x}_i - A\mathbf{x}_k\|^2)} \tag{35}$$

Siendo nula para puntos iguales $p_{ii} = 0$. Con esto se puede calcular la probabilidad p_i de que cada punto i sea clasificado correctamente, denotando los puntos de la misma clase como $C_i = \{j : c_i = c_j\}$, sería $p_i = \sum_{j \in C_i} p_{ij}$. Por tanto, la función objetivo se define de la forma siguiente

$$f(A) = \sum_i \sum_{j \in C_i} p_{ij} = \sum_i p_i \tag{36}$$

La etapa de optimización se puede realizar con algún método basado en el cálculo del gradiente, pudiendo resultar de gran coste computacional para grandes conjuntos de datos. Los puntos proyectados resultantes vienen dados por $\mathbf{y}_i = A \cdot \mathbf{x}_i$. NCA unifica tareas de aprendizaje de una métrica de distancia y reducción de la dimensión con un método DR lineal no paramétrico.

3.5.3 *Maximally collapsing metric learning*

Maximally collapsing metric learning (MCML) [74] utiliza también el enfoque de aprender una métrica con una idea similar a la propuesta en el método NCA. Sin embargo, a diferencia de la técnica NCA el problema de optimización aquí es convexo por lo que la solución que proporciona es única. Sin embargo, diferentes inicializaciones y técnicas de optimización podrían afectar a la velocidad en el cálculo de dicha solución.

Puesto que el enfoque es muy parecido al método NCA, se calcula la distribución de probabilidad de la misma manera que en la ecuación (35). El método supone que cada clase se puede colapsar en un punto, de manera aproximada, por lo que todos los puntos de una misma clase corresponden a un punto del mapa e infinitamente lejos de puntos de diferentes clases, teniendo una distribución de probabilidad ideal p^*_{ij} de dos niveles.

$$p^*_{ij} \propto \begin{cases} 1 & \text{si } y_i = y_j \\ 0 & \text{si } y_i \neq y_j \end{cases} \tag{37}$$

Con esto la solución se obtiene minimizando la divergencia de *Kullback-Leibler* entre las dos distribuciones de probabilidad.

El algoritmo puede ser utilizado para el cálculo de proyecciones con un buen rendimiento mejorando problemas computacionales de métodos previos, además también puede extenderse utilizando funciones kernel.

3.5.4 *Local Fisher discriminant analysis*

Local Fisher discriminant analysis (LFDA) [178] es un reciente método que mejora la reducción de la dimensión no lineal supervisada en casos reales como la visualización de datos y en tareas de clasificación. Presenta una combinación del análisis de *Fisher* [66] (LDA o FDA) y una técnica no supervisada llamada *locality-preserving projection* (LPP) [83], la cual encuentra una transformación de forma que vecindades locales se conserven en la proyección.

LFDA combina las ideas de ambos métodos lo cual permite una proyección con datos etiquetados, maximizando la separación entre clases y preservando las vecindades locales dentro de cada clase. El método extiende el cálculo de proyecciones supervisadas en cualquier espacio, sin la limitación de LDA que la dimensión del espacio de salida sea menor que el número de clases.

El criterio utilizado en este método es invariante a transformaciones lineales, por lo que el rango de la matriz de transformación se puede determinar de manera única, pero la métrica de distancia no. Una interesante línea futura apuntada en [178] es desarrollar un eficiente método para determinar la métrica de distancia como por ejemplo en los casos de NCA y MCML.

3.6 EVALUACIÓN DE LA CALIDAD DE LA PROYECCIÓN

A diferencia de la gran cantidad de técnicas de reducción de la dimensión desarrolladas, la investigación sobre la evaluación de las proyecciones resultantes no ha tenido tanta actividad [16].

El objetivo principal de las técnicas DR [119] es preservar tantas propiedades de los datos en el mapa proyectado tanto como sea posible, como son vecindades, similitudes, o proximidades. Una de las maneras más sencillas de evaluar cuantitativamente la calidad de la proyección es medir la preservación de las distancias que se produce en la reducción. Como se ha descrito previamente, la convergencia de muchas técnicas DR consiste en minimizar una función de coste. Un enfoque directo es tomar el valor de la función de coste al final de la convergencia, para valorar el error que posee la proyección. Ejemplos de estos son las medidas del *stress* en técnicas MDS [112, 45], una de las más conocidas es la utiliza-

da en el mapa no lineal de Sammon [160], denominado *stress de Sammon*:

$$E = \frac{1}{\sum\limits_{i<j} d^*_{ij}} \sum_{i<j} \frac{(d^*_{ij} - d_{ij})^2}{d^*_{ij}} \tag{38}$$

donde d^*_{ij} y d_{ij} son la distancias entre dos puntos i y j en el espacio de entrada (alta dimensión) y de salida (de baja), respectivamente. Otros criterios son el error de reconstrucción, para técnicas DR que proporcionen un mapa $\mathcal{M}$ y sean capaces de calcular $\mathcal{M}^{-1}$, utilizado por ejemplo en los autoencoders [111]; y también los índices como el error de clasificación utilizando las etiquetas de datos.

Otro enfoque para evaluar la calidad de una proyección cuantitativamente se basa en medir la preservación de la estructura de los datos sin tener en cuenta la función de coste. Los primeros intentos en esta dirección se realizaron en los mapas auto-organizados [110], donde se definen medidas como el producto topográfico [10] y la función topográfica [198]. Más recientes son las medidas de confianza y continuidad (T&C) de Venna [195], el metacriterio de continuidad local (LCMC) [37], la media de errores relativos de rango (MRRE) [119] o las curvas y comportamiento de calidad [120, 121].

Todos estos criterios tienen en cuenta las ordenaciones de las distancias calculadas antes y después de la reducción de la dimensión, es decir, en los espacios de entrada y salida, respectivamente. Estas medidas analizan la preservación estructural mediante un parámetro K de vecindad permitiendo el uso de rangos. Un criterio que unifica este tipo de medidas de calidad de las proyecciones ha sido propuesto en los trabajos de Lee [120, 121, 122], permitiendo una simple comparación de la calidad de las proyecciones. Por ello a continuación se describirá este tipo de evaluación basándose en este criterio unificado.

3.6.1 *La matriz de co-ranking*

Denotando D_{ij} y d_{ij} la distancia entre los puntos i y j en el espacio de entrada y salida respectivamente, se define el rango de $\mathbf{x}_j$ con respecto de $\mathbf{x}_i$ en un espacio multidimensional como

$$\rho_{ij} = |\{k : D_{ik} < D_{ij} \text{ (o } D_{ik} = D_{ij} \text{ y } 1 \leq k < j \leq N)\}| \tag{39}$$

donde $|A|$ denota la cardinalidad del conjunto A, es decir, ρ_{ij} representa el número de puntos más cercanos a i de lo que se en-

cuentra j. De la misma forma se define el rango de $\mathbf{y}_j$ con respecto de $\mathbf{y}_i$ en el espacio de salida (de baja dimensión) como

$$r_{ij} = |\{k : d_{ik} < d_{ij} \text{ (o } d_{ik} = d_{ij} \text{ y } 1 \leq k < j \leq N)\}| \quad (40)$$

Estos rangos no son necesariamente simétricos, mientras que los rangos reflexivos son igualados a cero ($\rho_{ii} = r_{ii} = 0$). Por tanto los rangos no reflexivos ($i \neq j$) son únicos y pertenecen al conjunto $\{1, \ldots, N-1\}$. Las K-vecindades para $\mathbf{x}_i$ y $\mathbf{y}_i$ se denotan v_i^K y n_i^K, respectivamente, definidas por la lista de los K vecinos más próximos a i de la siguiente forma

$$\begin{aligned} v_i^K &= \{j : 1 \leq \rho_{ij} \leq K\} \\ n_i^K &= \{j : 1 \leq r_{ij} \leq K\} \end{aligned} \quad (41)$$

Con todo esto la **matriz de co-ranking** ($\mathbf{Q}$) se define de la manera siguiente

$$\mathbf{Q} = [q_{kl}]_{1 \leq k,l \leq N-1} \text{ con } q_{kl} = |\{(i,j) : \rho_{ij} = k \text{ y } r_{ij} = l\}| \quad (42)$$

la cual forma un histograma conjunto de los rangos. Con una escala de grises apropiada, la matriz de co-ranking puede ser representada e interpretada de forma similar a un diagrama de Shepard [170], utilizado normalmente para evaluar los resultados de un escalamiento multidimensional (MDS). La analogía con este enfoque se basa en el hecho de que el diagrama de Shepard sugiere centrarse en la parte superior e inferior de la matriz de co-ranking dividiéndola por la diagonal principal.

Siguiendo esa línea, se define el *error de rango* como $\rho_{ij} - r_{ij}$, llamaremos *intrusión* a un error de rango positivo para la pareja de puntos (i, j), y de manera análoga, *extrusión* a un error negativo siendo la amplitud de esta intrusión o extrusión el valor absoluto de dicho error. La intrusión y extrusión se corresponden con la submatrices triangulares inferior y superior de la matriz de co-ranking, respectivamente. De la misma forma se puede asociar a una intrusión (extrusión), un tamaño de vecindad K, siendo K-intrusión (K-extrusión) con las K filas y columnas que se consideren de la matriz.

3.6.2 *Criterio de calidad basado en rangos*

La matriz de co-ranking contiene toda la información referida a la preservación de los rankings en una representación en baja dimensión de los datos, pero su interpretación y lectura no es trivial.

El enfoque general consiste en calcular sumas ponderadas en algunos bloques de la matriz para un valor determinado K de vecindad. Por tanto, trabajos relacionados previos consideran distintas partes de la matriz.

En la Fig. 3.2 se representan las correspondencias de cada criterio con las partes sombreadas de la matriz $\mathbf{Q}$, dividida en bloques con respecto a K. Por ejemplo, el LCMC considera la matriz de tamaño $K \times K$ con el número de vecinos considerado. Las medidas de confianza y continuidad [195] (T&C) y la media de errores de rango relativos (MRRE) [119] se centran en diferentes partes de la matriz. Con el criterio MRRE la parte de la matriz de tamaño $K \times K$ se utiliza dos veces.

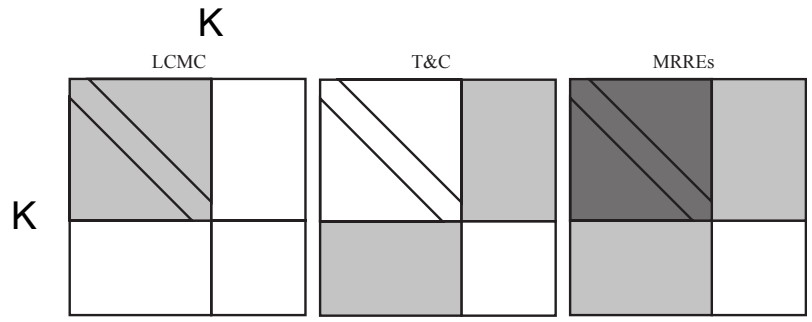


Figura 3.2.: Ilustración de las correspondencias de la matriz de co-ranking con criterios de calidad previos. Imagen modificada del trabajo [121].

Estos criterios conllevan una ponderación de las sumas por asuntos de normalización, sin embargo LCMC [37] cubre un bloque sencillo de la matriz de $\mathbf{Q}$, que no requiere dicha ponderación pero se pierde la diferencia entre extrusiones e intrusiones. Para ello en [121] se propone la definición de dos criterios. El primero definido como

$$Q_{\mathrm{NX}}(K) = \frac{1}{KN} \sum_{k,l \leq K} q_{kl} \tag{43}$$

en el que se evalúa la calidad general de una proyección, varía entre 0 y 1, donde el valor más alto indica la proyección ideal. Mide la preservación de las vecindades, por lo que también puede definirse por la siguiente ecuación

$$Q_{\mathrm{NX}}(K) = \sum_{i=1}^{N} \frac{|n_i^K \cap v_i^K|}{KN} \ , \tag{44}$$

El segundo criterio tiene en cuenta la diferencia entre las K-intrusiones y K-extrusiones,

$$B_{\mathrm{NX}}(K) = \frac{1}{KN}\Big(\sum_{\substack{k<l \\ k,l\leq K}} q_{kl} - \sum_{\substack{k>l \\ k,l\leq K}} q_{kl}\Big) \tag{45}$$

en donde el signo indica el "comportamiento" de la proyección si es intrusiva o extrusiva, siendo positivo o negativo respectivamente.

Estas medidas permiten una interpretación de la calidad de las proyecciones de manera escalar, permitiendo una rápida comparación entre ellas. Se pueden representar en una línea con respecto al parámetro K, de esta forma pueden valorarse las propiedades locales al inicio de las curvas y las globales al final. Un buen método DR debería tener valores altos para todos los valores de K, pero sobre todo para los bajos considerando que en la literatura tiene más importancia la preservación local de las propiedades [119]. Para esto, se pueden resumir los valores calculando la media para todos los valores de K de la forma siguiente

$$Q_{avg} = \frac{1}{N-1}\sum_{K=1}^{N-1} Q_{\mathrm{NX}}(K) \tag{46}$$

$$B_{avg} = \frac{1}{N-1}\sum_{K=1}^{N-1} B_{\mathrm{NX}}(K) \tag{47}$$

Estas cantidades varían de $[0, 1]$ y $[-1, 1]$ respectivamente, representando una perfecta proyección para $Q_{avg} = 1$ y $B_{avg} = 0$.

3.6.3 *Visualización de la medida de calidad estructural*

Las medidas descritas anteriormente evalúan el grado de fidelidad de una proyección con los datos originales. Un enfoque interesante es representar la preservación estructural de manera que sea accesible desde el espacio de la proyección. Esto facilita al usuario a valorar la pérdida de información que se produce en el proceso de reducción de la dimensión. Por ejemplo, en [7] se presentan varias maneras para visualizar errores de conservación de la estructura en una proyección, lo que permite evaluar las distorsiones producidas respecto a la topología original.

Un reciente trabajo [139] extiende la interpretación de la matriz de co-ranking (Ec. 42) para mejorar su interpretación visual y un mayor control de la medida por parte del usuario. La evaluación se realiza calculando la medida anterior ($Q_{NX}(K)$) con respecto

a cada punto de la proyección de forma que se pueda visualizar en la propia proyección.

Para esto la matriz de co-ranking $\mathbf{Q}$ se descompone por las N matrices de permutación $\mathbf{Q^{x_i}}$ que la forman, para cada punto $\mathbf{x}_i$ con $\mathbf{Q} = \sum_{i=1}^{N} \mathbf{Q^{x_i}}$, de la forma siguiente

$$\mathbf{Q^{x_i}} = [q_{kl}^{x_i}]_{1 \leq k,l \leq N-1} \text{ con } q_{kl}^{x_i} = |\{j : \rho_{ij} = k \text{ y } r_{ij} = l\}| \quad (48)$$

De esta forma las contribuciones a la medida de calidad $Q_{NX}(K)$ de cada punto vienen dadas por

$$Q_{NX}^{x_i}(K) = \sum_{k \leq K} \sum_{l \leq K} q_{kl}^{x_i} / K \quad (49)$$

Calculando su valor medio sobre todos los puntos, se obtiene la medida de calidad total $Q_{NX}(K) = \sum_{i=1}^{N} Q_{NX}^{x_i}(K)$. Con estos valores, cada punto del mapa puede ser coloreado con el valor que aporta a la medida total de la calidad para un determinado valor de K.

Por tanto este interesante trabajo [139] presenta la medida $Q_{NX}(K)$ por cada punto aportando al usuario una evaluación visual por medio del color de la proyección para una vecindad controlada también por el mismo.

3.6.4 *Medidas visuales de calidad de la proyección*

Las anteriores medidas de calidad de la proyección valoran la preservación en la estructura de los datos que se produce en el proceso de reducción. Aparte de este tipo de medidas, también se han desarrollado medidas que evalúan, aritméticamente, la calidad de una proyección desde un punto de vista visual. En este tipo de medidas se incluyen, por ejemplo las propuestas en [180], el *Histogram Density Measure* que hace un ranking de visualizaciones *scatterplot* de datos multidimensionales y *Class Density Measure* que evalúa la separación de las clases de una proyección dada, y también las medidas de consistencia de clase propuestas en [174].

Para que los patrones revelados en una proyección de datos puedan percibirse fácilmente, conviene evitar la superposición masiva de objetos, como por ejemplo puntos o grupos de puntos que pertenecen a diferentes clases. Las dos medidas de solapamiento definidas en [162] calculan el área de superposición entre grupos incluidos en los datos y la densidad de los puntos superpuestos en una proyección visual de datos multidimensionales. Dado un conjunto de datos etiquetados que incluyen un número g de grupos,

se describen los límites de la región que contenga los objetos que pertenezcan a un mismo grupo. Para ello existen algunos métodos que calculan regiones definidas por puntos, como aquellos basados en *convex hull* [77, 96]. Aquí se utiliza un método menos conocido denominado *concave hull* [141], en el cual por cada punto se busca la conexión entre sus k-vecinos más cercanos para asegurar un resultado tan compacto como sea posible, no necesariamente tiene que ser convexo. Una vez que la región de cada grupo está definida, se calcula la región superpuesta $intersect(i, j)$ para cada pareja de grupos i y j. El *área de solapamiento* suma el área de todas las regiones de intersección ($intersect(i, j)$) entre pares de grupos para el conjunto g de grupos, quedando definida de la siguiente forma

$$ov_{reg} = \sum_{i=1}^{|g|-1} \sum_{j=i+1}^{|g|} intersect(i, j) \tag{50}$$

La *densidad de solapamiento* tiene en cuenta la densidad de puntos superpuestos en la representación visual. El mapa se divide en unidades de rejilla donde la ocupación simultánea por dos o más clases se determina por medio de funciones gaussianas G en la función f descrita a continuación

$$f(G_{ip}, G_{jp}) = \begin{cases} 1 & si \quad G_{ip} > 0 \; y \; G_{jp} > 0 \\ 0 & resto \end{cases} \tag{51}$$

La función f se activa con 1 en el caso donde un cuadro de unidad de rejilla se ocupe por dos clases distintas usando el modelo gaussiano descrito. Con la suma de dichas rejillas por parejas de clases se obtiene la medida total, la cual, denotando por K el número de clases y en una imagen de P píxels, se define como

$$ov_{density} = \sum_{i=1}^{|K|-1} \sum_{j=i+1}^{K} \sum_{p=1}^{P} f(G_{ip}, G_{jp}) \tag{52}$$

En los resultados presentados en esta tesis los valores para esta medida se han fijado para una resolución de rejilla de 3 píxels y un valor de σ en el modelo gaussiano de 12, siendo utilizados para todos los experimentos de manera que sean todos comparables entre sí.

3.7 EJEMPLO ILUSTRATIVO DE PROYECCIÓN DE DATOS

A continuación se presenta un ejemplo para ilustrar los resultados que se obtienen de algunas de las técnicas DR explicadas

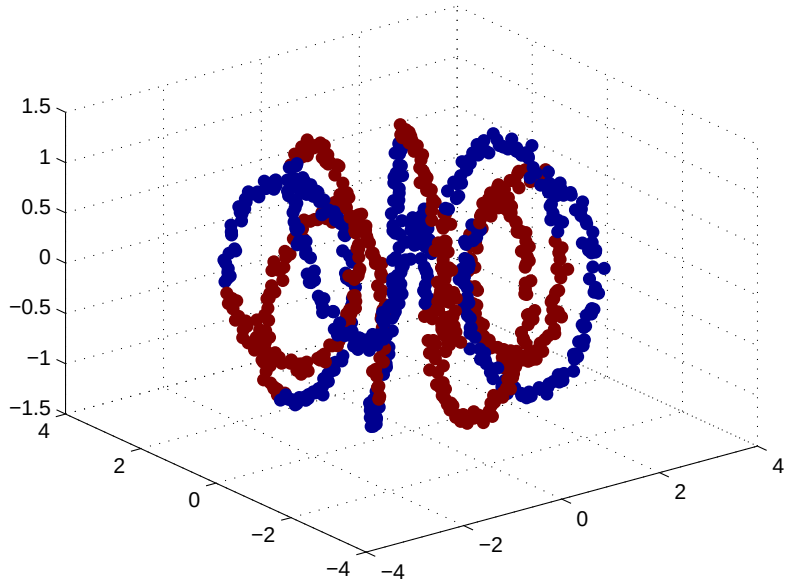


Figura 3.3.: Representación del conjunto *hélice* de datos

anteriormente, además también se calcula una evaluación de la calidad.

Se elige un conjunto de datos sintético y tridimensional denominado *hélice*, el cual consta de 1000 puntos. En la Fig. 3.3 se puede ver representada la geometría del conjunto de datos. Se han calculado proyecciones mediante algunas de las técnicas explicadas anteriormente. En la Fig. 3.4 aparecen representadas las proyecciones de los datos calculadas con PCA (superior, izquierda), el mapeo no lineal de Sammon (superior, derecha), LLE (inferior, izquierda), y t-SNE (inferior, derecha).

En estas proyecciones se visualizan los datos en un mapa de dos dimensiones en donde se revela la estructura latente de los datos originales. Aunque sea inevitable alguna pérdida de información que contienen los datos, se pueden extraer patrones de los mismos que revelen conocimiento de los datos en un espacio de alta dimensionalidad.

Utilizando la proyección calculada con PCA y los datos originales en tres dimensiones, se calcula la medida de calidad mediante el criterio de *rankings* explicado en el apartado 3.6. De esta manera, se obtiene la medida de calidad $Q_{\mathrm{NX}}(K)$ por medio de la ecuación (43) considerando todos los valores del conjunto K de posibles vecinos. En la Fig. 3.5 se representa las curvas con estos valores en la que se puede evaluar cómo se comporta cada técnica

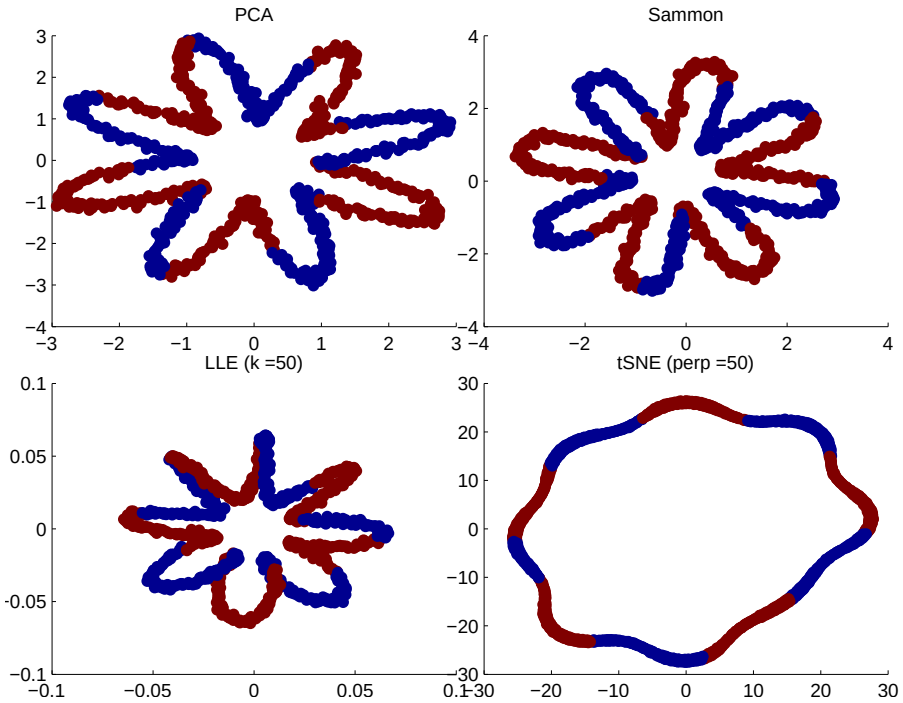


Figura 3.4.: Proyecciones del conjunto *hélice* de datos

para los diferentes vecinos K, es decir, una evaluación general del comportamiento de la técnica tanto a nivel local (valores bajos de K) como a nivel global (valores altos de K). El valor Q_{avg}, considerado en la ecuación (46), corresponde al valor medio de esta curva para cada técnica.

Además, también se puede visualizar la calidad de la proyección como se explica en el apartado 3.6.3. Con este enfoque se puede calcular la calidad media de cada punto y codificarla mediante el color en la propia proyección. En Fig. 3.6 se representa esta calidad por cada punto en cada una de las proyecciones calculadas. De esta manera se evalúa de manera visual la proyección, por ejemplo identificando rápidamente los puntos con más calidad o los que poseen peor medida de calidad. Finalmente, en la Fig. 3.7 se puede visualizar la calidad media de cada técnica de proyección en la vista 3D de los datos originales.

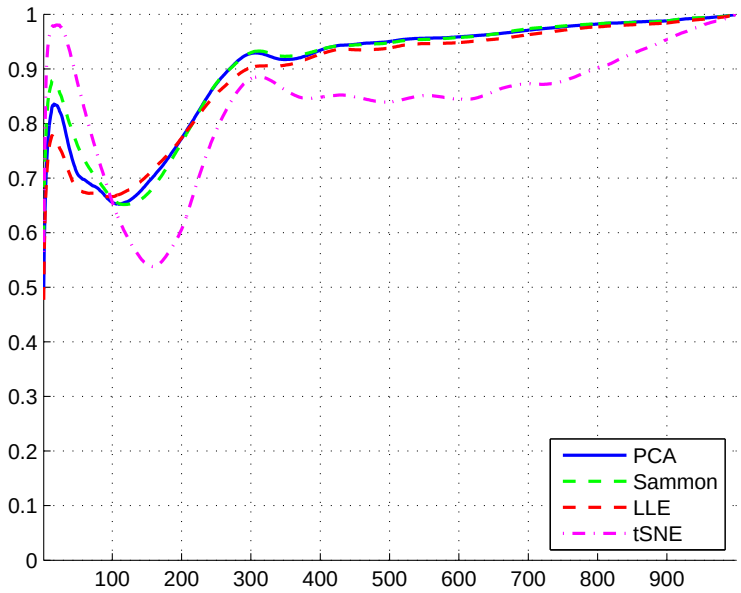


Figura 3.5.: Curva de la medida de calidad Q_{NX} en función del número de vecinos del conjunto *hélice* de datos

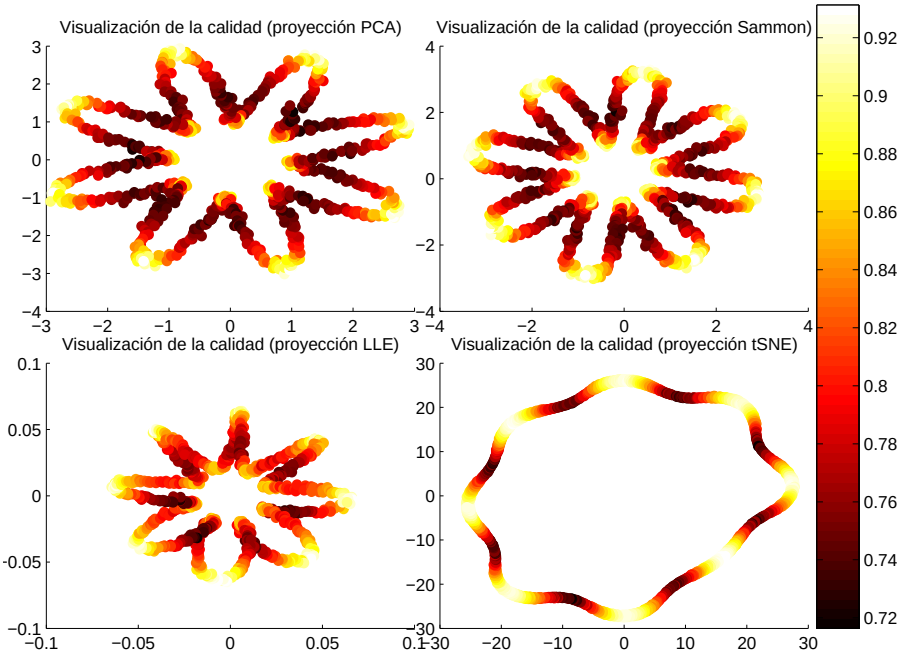


Figura 3.6.: Visualización de la medida de calidad Q_{NX} de cada punto en cada una de las proyecciones calculadas.

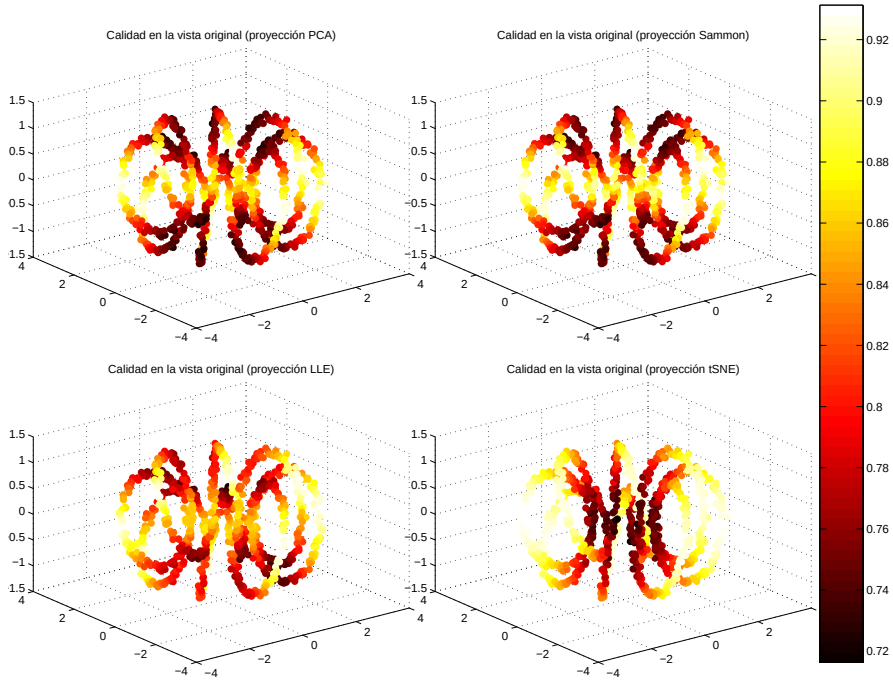


Figura 3.7.: Visualización de la medida de calidad Q_{NX} de cada punto en cada una de las proyecciones calculadas.

4

ANÁLISIS VISUAL DE PROCESOS CON DINÁMICA

En este capítulo se describen varios métodos que utilizan diferentes algoritmos de análisis de datos aplicados a problemas complejos. Los resultados se presentan de forma visual, lo cual facilita su interpretación, al crear nuevas hipótesis y apoyar decisiones de actuación que afectan directamente a la eficiencia de los procesos analizados. Se aplican a dos casos reales como el análisis de patrones de consumos eléctricos en edificios universitarios y en un proceso de laminación en frío.

4.1 INTRODUCCIÓN

La tecnología actual facilita la adquisición y el almacenamiento masivo de datos en la mayoría de los escenarios de nuestro entorno. Sin embargo, en numerosas ocasiones estos datos permanecen almacenados sin estudiarse, pudiendo ocultar información útil. Determinados procesos pueden poseer fenómenos de naturaleza difícil de determinar o características que solamente pueden tratarse mediante el análisis de sus datos. Con la aplicación de algoritmos específicos, se pueden obtener resultados que aborden problemas complejos y puedan apoyar decisiones de actuación en la eficiencia del proceso.

Una presentación visual de estos resultados es esencial, puesto que no sólo sirve como apoyo para comunicar de manera eficiente ideas complejas a los demás sino también facilita una rápida interpretación y, por tanto, actuación sobre un determinado problema. Además, el uso adecuado de mecanismos de interacción ayuda al usuario a la exploración de los datos y a un mejor entendimiento del proceso analizado.

A continuación se describen varios métodos desarrollados para el análisis inteligente y la proyección de datos multidimensionales aplicados a dos procesos distintos: el análisis de los consumos eléctricos que se producen en edificios universitarios, y el estudio de un fallo, denominado *chatter*, el cual es una potente vibración que ocurre en los trenes de laminación en frío. En ambos casos, se proponen diversas formas de procesado para la extracción de

información útil y la representación visual de proyecciones de sus datos para facilitar su interpretación.

4.2 ANÁLISIS VISUAL DE PATRONES ELÉCTRICOS

La energía es un aspecto importante en nuestra sociedad y su adecuada gestión implica un gran impacto en diferentes áreas como la economía o el medio ambiente. Además de los esfuerzos realizados para una producción limpia, renovable o sostenible, también se debería actuar en la demanda mediante un uso racional del consumo, lo cual mejoraría el balance energético.

Las recientes tecnologías permiten aumentar fácilmente la cantidad y calidad de los sensores instalados para medir la demanda energética. Sin embargo, esta información acaba con frecuencia almacenada o presentada de manera agregada sin revelar consumos ineficientes. En muchas ocasiones se toman decisiones importantes sin tener en cuenta esta información oculta. El conocimiento sobre patrones en el consumo eléctrico debería presentarse de manera que el usuario pueda evaluar rápidamente la demanda en un determinado periodo y que también pueda ayudar a la toma de posibles decisiones como cambios en los hábitos de consumo.

La visualización de patrones eléctricos temporales ha sido estudiada previamente. En [194] se agrupan patrones diarios similares y se muestran sus valores medios en gráficos conectados con los correspondientes días en una vista de calendario. En [76] se plantea un sistema para predecir el consumo de energía en edificios basado en redes neuronales. En [140] se analizan los perfiles de consumo eléctrico en edificios públicos, proyectados en un mapa 2D, así como la influencia de las variables ambientales en ellos.

A continuación se detallan dos aplicaciones para el análisis visual de patrones de consumos eléctricos recogidos en edificios universitarios. La primera de ellas, propone una proyección de una gran cantidad de muestras de la demanda energética. En la segunda, se presenta una aplicación web para la exploración interactiva de distintas vistas representando información sobre los consumos eléctricos.

4.2.1 *Proyección visual de consumos de potencia*

La enorme cantidad de datos disponibles, procedentes de las bases de datos y equipos de adquisición, supone una dificultad para la visualización. En el caso del cálculo de proyecciones de los datos, las técnicas de reducción de la dimensión poseen un

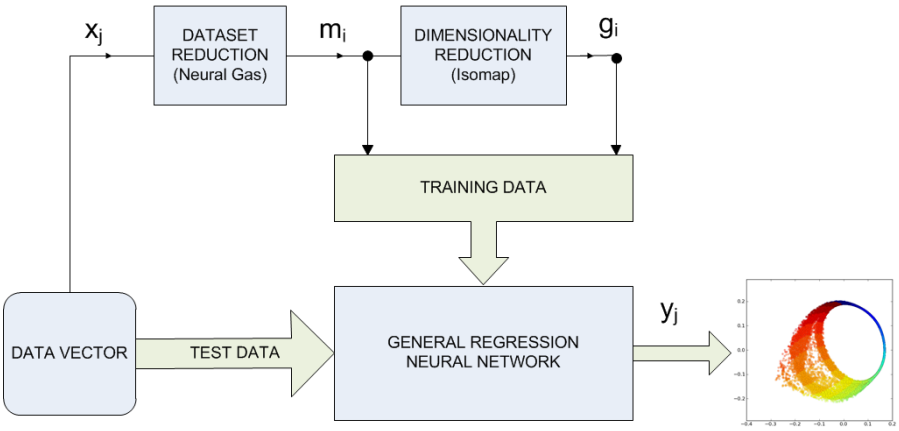


Figura 4.1.: Descripción gráfica del método propuesto en [148].

importante coste computacional. Por ejemplo, las técnicas *MDS* o *Isomap* tienen un coste de $\mathcal{O}(N^3)$ siendo N el número de muestras.

En el trabajo propuesto en [148] se tratan estas limitaciones, donde se obtiene una proyección de un grupo extenso de datos en un espacio de visualización para el análisis de patrones de consumo de potencia eléctrica. Para ello, en primer lugar el tamaño de muestra de los datos se reduce mediante un método de cuantización vectorial, el cual calcula una representación óptima de los datos describiendo homogéneamente la geometría más importante del espacio de entrada. De esta forma, con la aplicación de un método como *neural gas* [134], dado un vector de datos $\mathbf{x}_j \in \mathbb{R}^D$ con $j = 1, \ldots, N$ se obtienen un número n más reducido de vectores $\mathbf{m}_i \in \mathbb{R}^D$ llamados "*prototipos*". Este grupo de prototipos representan todas las muestras del conjunto de datos describiendo su función de densidad de probabilidad.

En una segunda etapa se calcula una reducción de la dimensión de los prototipos $\mathbf{m}_i$. En este caso, utilizando la técnica *Isomap* se obtienen los puntos $\mathbf{g}_i$ de la proyección, los cuales definen la estructura de dichos prototipos en un espacio de baja dimensión. Finalmente, la proyección total $\mathbf{y}_j$ se estima mediante una red neuronal llamada *general regression neural network* (GRNN) [175]. Este algoritmo estima una regresión continua $\hat{\mathbf{y}}(\mathbf{x})$ de la forma siguiente:

$$\hat{\mathbf{y}}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{g}_i \exp\left(-\frac{(\mathbf{x}-\mathbf{m}_i)^T(\mathbf{x}-\mathbf{m}_i)}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{(\mathbf{x}-\mathbf{m}_i)^T(\mathbf{x}-\mathbf{m}_i)}{2\sigma^2}\right)} \tag{53}$$

Siendo σ el parámetro que determina la uniformidad de la función de densidad estimada. Esta red calcula la salida interpolando

entre los valores de entrenamiento, por lo que puede perder cierta precisión, sin embargo, permite el cálculo de una proyección visual para todas las muestras de los datos. En Fig. 4.1 se representa un diagrama con los pasos del método aplicado en [148].

Experimento

Los datos utilizados en este experimento, para validar el método propuesto, fueron recogidos en el edificio de la Escuela de Ingeniería Agraria de la Universidad de León, entre las fechas de 11/09/2010 a 30/10/2010 con un periodo de muestreo de 2 minutos. Las variables analizadas en este estudio son los consumos de potencia eléctrica (potencia activa p_a y reactiva p_r), normalizadas con media cero y varianza unitaria. Además se añadieron variables temporales $t(i)$ en forma senoidal multiplicadas por un peso sobre el resto ($w = 5$), lo cual proporciona una periodicidad horaria a las características. De esta manera, el vector de características es de la forma siguiente:

$$\mathbf{x}_j = \left[\cos\left(2\pi\frac{t(i)}{24}\right)\cdot w, \sin\left(2\pi\frac{t(i)}{24}\right)\cdot w, p_a(i), p_r(i)\right] \qquad (54)$$

El algoritmo *neural gas* se aplica en su versión por lotes (*batch*) [43] con un número de 300 unidades y una vecindad de 0.1, utilizando 10 épocas en su etapa de convergencia. La reducción de la dimensión se calcula para los 300 puntos resultantes mediante la técnica *Isomap* [182] con un número de vecinos de $k = 5$. Entonces la red *GRNN* utiliza los 300 prototipos y la proyección obtenida para calcular la total de todo el conjunto de datos, con un valor del parámetro $\sigma = 3$.

La factura del consumo eléctrico se divide en tres periodos de facturación dependiendo de la hora y el día de la semana que sea.

- *Periodo 1 o punta* (P1) : Desde las 10 a 16 horas de lunes a viernes.
- *Periodo 2 o llano* (P2): Desde las 8 a 10 horas y 16 a 0 horas. Fines de semana de 18 a 0 horas.
- *Periodo 3 o valle* (P3): De 0 a 8 horas. Fines de semana de 0 a 18 horas.

En la Fig. 4.2 se representa gráficamente la distribución de estos periodos de facturación aplicados a los consumos eléctricos. La demanda tiene un diferente precio en cada uno de los periodos

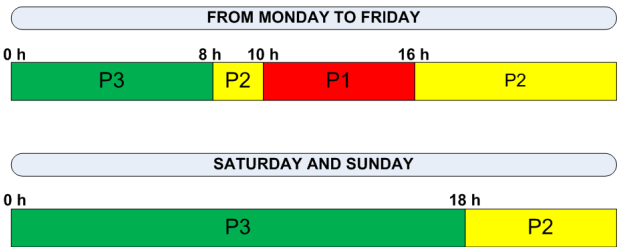


Figura 4.2.: Distribución de los periodos de facturación.

(siendo P1 el más caro y P3 el más barato) y se factura en forma de tres términos: el término de potencia P_f se calcula con el valor de la potencia contratada y el máximo de la demanda; el término de energía activa se factura directamente de las medidas recogidas; y un término de energía reactiva se añade cuando su valor es más del 33 % de la energía activa, es decir, cuando el factor de potencia es menor de 0,95.

Las proyecciones resultantes más importantes se representan en la Fig. 4.3 donde se puede ver que los puntos proyectados representan un círculo en forma de reloj (debido a la periodicidad introducida de 24 horas) rotado de forma que las 0 horas se encuentren en la posición más alta y las 12 horas en la baja. En la Fig. 4.3 (arriba, izquierda) el color representa la hora del día, lo cual ayuda a identificarlas fácilmente en la proyección. Los colores de la Fig. 4.3 (arriba, derecha) representan los periodos de distribución (descritos en Fig. 4.2). El área de color rojo corresponde al periodo 1, de mayor coste, el amarillo al periodo 2 con un precio intermedio y con el color verde el periodo 3 de menor coste.

En la parte inferior de la figura se representan las proyecciones con los colores correspondientes del término de la energía activa (izquierda) y reactiva (derecha). Se observan las horas del día con alta demanda que además tienen el precio más alto (P1) y los fines de semana con baja demanda. Además también se observa que no existe una relación directa del consumo de potencia reactiva con los periodos de facturación, puesto que el término de energía reactiva se factura en cualquier periodo.

En dicho trabajo [148] se muestra experimentalmente la viabilidad del cálculo de una proyección mediante una reducción de la dimensión sin la importante restricción en el tamaño de las muestras de los datos de entrada, que plantean las técnicas de reducción de la dimensionalidad. Se ha utilizado como ejemplo la visualización de patrones eléctricos donde se pueden identificar fácilmente los consumos y relacionarlos con los distintos periodos de su facturación. Esto permite el rápido análisis de variables implicadas en

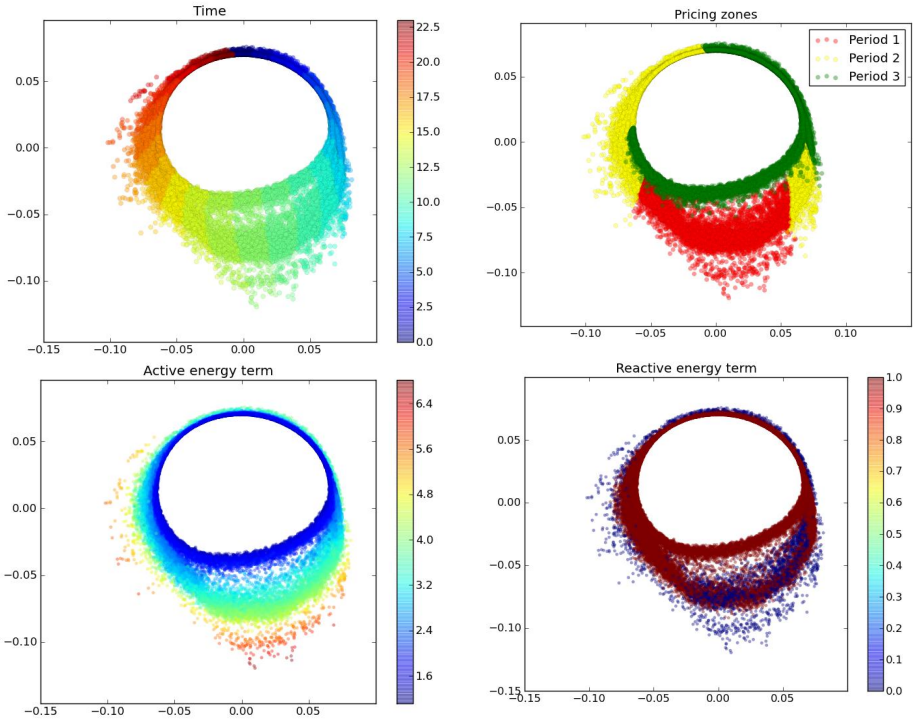


Figura 4.3.: Distribución de los periodos de facturación de consumos eléctricos.

el coste final del consumo, y como soporte para tomar decisiones con respecto a mejorar la eficiencia energética.

4.2.2 *Exploración interactiva de datos en aplicación web*

En el trabajo presentado en [147] se sugiere el uso de métodos web, considerando principios y técnicas de visualización, en el desarrollo de aplicaciones para supervisar procesos complejos. Se implementa un prototipo para facilitar el análisis visual y la interpretación de la demanda de potencia eléctrica en edificios universitarios.

Un problema inicial en este tipo de análisis es la concurrencia de distintos tipos de periodicidades, regulares o no regulares procedentes de la actividad de eventos específicos (deportivos, huelgas, etc). Además, la demanda de la potencia puede verse afectada por condiciones especiales, como las atmosféricas o periodos de exámenes en el caso de centros educativos. Por tanto, la aplicación debería ayudar al usuario a identificar patrones temporales regulares, especiales, o grupos de días con similares consumos. Para ello, se deberían proporcionar diferentes vistas con los distintos tipos de información de los datos y permitir establecer conexiones men-

tales entre ellas. Para su exploración, las vistas se pueden mostrar en una misma pantalla o dar al usuario la posibilidad de añadirlas, eliminarlas o intercambiarlas.

La interacción es una parte esencial en las interfaces visuales, permitiendo al usuario manipular las representaciones de diferentes maneras y concentrarse en los aspectos más importantes de los datos. Mecanismos habituales como *zooming* y *panning* permiten el cambio en la misma codificación visual; representar información contextual situada sobre el ratón del ordenador proporciona detalles bajo demanda; o el *brushing* relaciona una selección de un subconjunto para resaltarlos. La interacción proporciona movimiento a una imagen estática, lo cual involucra al usuario a seguir los objetos visualmente. En [188] se sugieren principios para una efectiva animación y se discuten sus ventajas e inconvenientes. En [86] se exploran las transiciones animadas para mejorar la percepción gráfica en datos estadísticos. Un reciente enfoque, denominado *Morphing Projections* [51], combina transiciones continuas entre codificaciones 2D con varias técnicas de interacción para la exploración de consumos eléctricos. Mediante este enfoque, la transición se controla mediante un parámetro $\lambda \in [0, 1]$ pudiendo revelar estados intermedios interpretables.

Caso de estudio

Como caso de estudio se ha desarrollado una aplicación web, implementada usando la librería D3 en Javascript, para la exploración interactiva de consumos de potencia. Los datos originales se han obtenido del sistema de adquisición de dos edificios de la Universidad de Oviedo durante un año, con un periodo de muestreo de 15 minutos. Se realiza una reducción 4:1 de los datos promediando el consumo en una base de horas. De esta manera, los datos constan de 8760 muestras (365 x 24). Las variables de los datos son las potencias activa (P), reactiva (Q) y aparente (S) en ambos edificios universitarios. Además también se tienen en cuenta el factor de potencia ($cos\varphi$) y un residuo (R), calculados de la siguiente forma:

$$\cos\varphi = \frac{P}{S} \tag{55}$$

$$R = S^2 - P^2 - Q^2 \tag{56}$$

El residuo R debería ser cero en condiciones ideales ($S^2 = P^2 + Q^2$) sin distorsión armónica. Este atributo ayuda a identi-

ficar desviaciones de las condiciones ideales. Las visualizaciones presentadas utilizan diferentes *scatterplots*, donde las muestras de los datos son representadas por puntos, cuya posición en el plano codifica un tipo de información.

Una codificación visual $\mathbf{p}_A$ viene definida por un conjunto de puntos $\mathbf{p}_A(i)$, que describen información de la muestra i mediante la posiciones espaciales que ocupan, normalmente con las magnitudes x e y del plano.

$$\mathbf{p}_A = \{\mathbf{p}_A(1), \mathbf{p}_A(2), \cdots, \mathbf{p}_A(N)\}, \quad \mathbf{p}_A(i) \in \mathbb{R}^2 \tag{57}$$

Los tipos de información codificados por las posiciones de los puntos incluyen vistas de los días agrupadas en forma de calendario, por similitud mediante la proyección obtenida de una reducción de la dimensión, o agrupadas en forma circular similar a un reloj para describir periodicidades temporales. Por ejemplo, para las horas del día $h(i)$, las posiciones de un punto son de la forma:

$$\mathbf{p}_D(i) = \left[\cos\left(2\pi\frac{h(i)}{24}\right), \sin\left(2\pi\frac{h(i)}{24}\right)\right] \tag{58}$$

La operación de *morphing* explicada en [51] consiste en mezclar dos (o más) codificaciones visuales en una nueva intermedia de ambas. Dadas dos codificaciones diferentes $\mathbf{p}_A(i)$ y $\mathbf{p}_B(i)$ para $i = 1, \ldots, N$, y $\lambda \in [0, 1]$ el coeficiente de mezcla, la operación entre $\mathbf{p}_A(i)$ y $\mathbf{p}_B(i)$ viene dada por la ecuación siguiente:

$$\mathbf{p}(i, t) = \lambda(t)\mathbf{p}_A(i) + (1 - \lambda(t))\mathbf{p}_B(i) \tag{59}$$

para $i \in \{1, \cdots N\}$.

Por tanto, el usuario puede controlar la transición animada entre las dos vistas, mejorando la conexión visual entre ellas. Además, un valor adecuado de mezcla puede revelar resultados intermedios interpretables para el análisis. Por ejemplo, con la mezcla entre "día de la semana" y "hora diaria" se obtiene una distribución de las "horas por día de la semana", como se muestra en la parte izquierda de la Fig. 4.4.

La combinación de los mecanismos de interacción permite analizar diferentes periodos mediante peticiones visuales. Por ejemplo, la información de una selección realizada se puede representar en forma de diagrama de barras para el estudio detallado de los datos. En la parte izquierda de la Fig. 4.4 se puede ver resaltada en rojo una selección (viernes a las 12 horas), mantener esta selección entre las vistas facilita el estudio de los puntos seleccionados

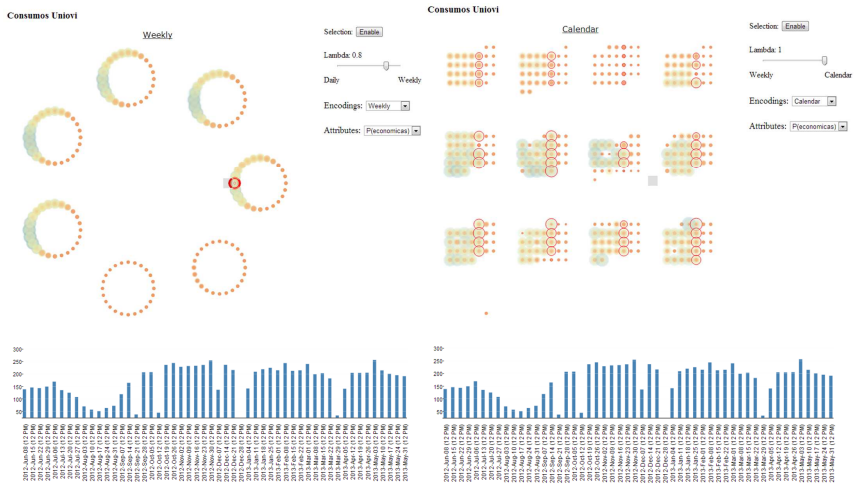


Figura 4.4.: Estados intermedios entre día de la semana y horas del día (izq.) y vista de calendario con selección de puntos (dcha.)

mediante su conexión visual (ver Fig. 4.4, derecha) en la vista de calendario, con los anteriores puntos resaltados en rojo.

Este caso de estudio muestra un ejemplo sencillo y eficiente de una aplicación web interactiva, para la exploración visual de consumos eléctricos, que podría ser extendido para supervisar otros procesos complejos mediante tecnología web.

4.3 SUPERVISIÓN DE UN PROCESO DE LAMINACIÓN EN FRÍO

En un proceso de laminación se reduce el espesor de una banda o desbaste de acero al hacerlo pasar a través de dos rodillos giratorios. Aunque este proceso es estándar, posee un alto grado de exigencia dentro del sector metalúrgico. Las condiciones típicas de funcionamiento incluyen grandes valores de fuerzas y tensiones a altas temperaturas que pueden llevar a situaciones extremas. En algunos casos, pueden producirse fallos que pueden causar importantes pérdidas económicas o incluso daños más graves, lo que hace que supervisar el proceso sea una tarea esencial.

Uno de los fallos más severos que puede producirse en este tipo de procesos es el denominado *chatter* [212], el cual consiste en una inesperada y potente vibración que afecta a la calidad del material producido causando una variación del espesor inadmisible o incluso, en casos extremos, puede producirse la rotura de alguna

de las partes implicadas en el proceso. Este complejo fenómeno es el resultado de un conjunto de interacciones dinámicas entre la caja de laminación y la banda de acero que se lamina. Varios trabajos previos han ayudado a entender el fallo [179, 184], mostrando que el chatter es una vibración auto-excitada cuya eliminación se suele alcanzar mediante una reducción de la velocidad de laminación. En la Fig. 4.5 se representan varias señales del proceso durante un episodio de chatter. Sin embargo, reducir la velocidad un tiempo y volver a subirla una vez que el efecto desaparece hace que la productividad de la instalación disminuya. Por esta razón, tanto la caracterización como la detección temprana del chatter tienen un gran interés para la eficiencia del proceso.

Se puede realizar una identificación del chatter mediante la densidad de potencia espectral en la banda de frecuencias en las que este fallo aparece (normalmente entre 100-300 Hz). En [72, 73] se propone la comparación y visualización de espectrogramas para un proceso de laminación en caliente mediante el cálculo de su proyección. En el mapa resultante se diferencian bobinas sin fallos de aquellas en las que fallos de chatter aparecieron durante su laminación.

Una forma de predecir el fallo es calculando un modelo del proceso [91]. Sin embargo, la complejidad matemática dificulta la obtención de un modelo preciso, además de la configuración que requieren sus parámetros. En [149] se propone un nuevo enfoque para estudiar el chatter con el análisis del comportamiento dinámico de un modelo basado en datos mediante sus proyecciones. Se hace un análisis de sensibilidad perturbando las entradas del modelo con una pequeña señal aleatoria, en distintos puntos de trabajo del proceso. De esta forma, se obtienen modelos de pequeña señal calculando las funciones de repuesta de frecuencia (FRF) mediante la técnica del periodograma de Welch aplicada a la salida y la entrada del modelo perturbado [149]. Estos vectores son proyectados en un plano mediante el algoritmo *t-Stochastic Neighbour Embedding* (*t*-SNE) [190] de manera que cambios en los comportamientos dinámicos del proceso se visualizan de manera espacial en el mapa, obtenido en la proyección.

Además, con el objeto de proyectar puntos nuevos, se desarrolló una versión *out-of-sample* de esta técnica para el cálculo de nuevas proyecciones, en el mapa 2D calculado previamente. Para ello, se aplica dicha técnica de proyección nuevamente para todos los puntos asegurando que las matrices de probabilidad sumen 1 para cada nuevo punto añadido. Las diferencias entre las distribuciones de probabilidad se calculan de la misma forma que el algoritmo

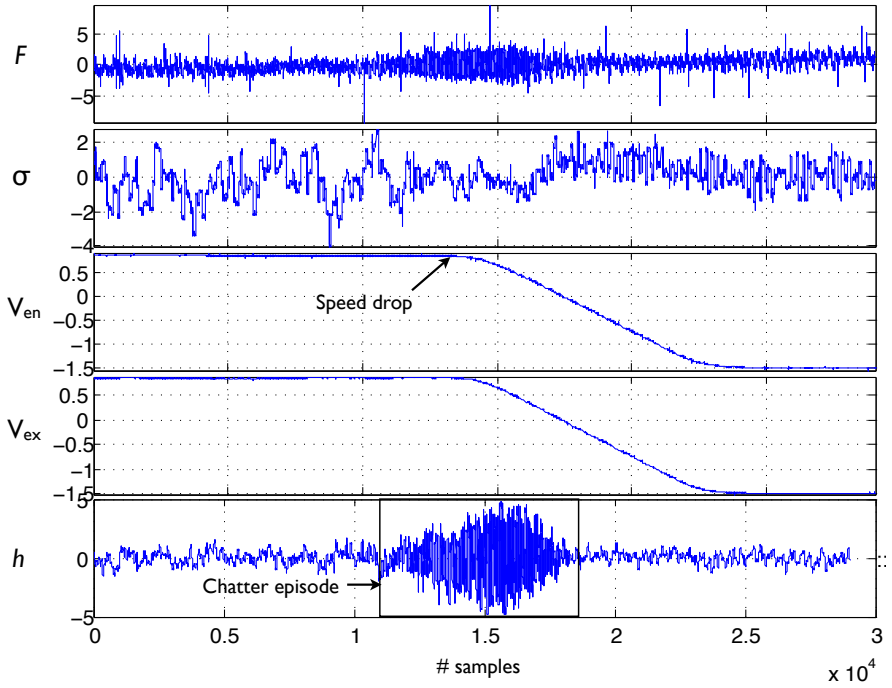


Figura 4.5.: Señales de varias variables: fuerza (F); tension (σ); velocidades de entrada (V_{en}) y salida (V_{ex}) y el espesor de salida (h) durante un episodio de chatter. Fuente [149].

original (mediante la divergencia de *Kullback-Leibler*) y también el método para minimizar esta función de coste, pero en este caso manteniendo las posiciones de los puntos de entrenamiento fijas.

Una revisión de este enfoque se extiende en [150], donde se realiza una caracterización de fallos de chatter, por medio de la visualización de estados dinámicos calculados del proceso en un espacio visual de baja dimensión. Las proyecciones de los datos en el dominio de la frecuencia permiten una detección visual de este tipo de fallos. A continuación se describe esta extensión de manera detallada.

4.3.1 *Descripción del modelo*

Los modelos clásicos de laminación en frío intentan calcular la fuerza y tensión necesaria para una determinada reducción del espesor. La complejidad del proceso es tan alta que para construir un modelo sencillo se necesitan hacer varios supuestos, tales como considerar las mismas direcciones de las vibraciones producidas en la parte superior e inferior, con el objeto de establecer una simetría en el modelo del tren o en otros casos suponer una velocidad horizontal constante del material [91, 156]. Un modelo hace posible analizar fallos como el chatter en las condiciones de funcionamiento, donde se producen variaciones en la fuerza que llevan al sistema a un estado inestable. Por tanto, es necesario generar un modelo donde se tengan en cuenta diferentes factores. Como se explica en [144], el fenómeno del chatter procede de una interacción entre diversas variables: la velocidad de entrada, las tensiones de entrada y salida, la fuerza de la banda de acero en los rodillos, y el espesor de salida. Si se añade un modelo dinámico para el tren, se puede construir un modelo para estudiar dicho fenómeno [136, 108].

Por tanto, un modelo simplificado de acuerdo a relaciones mecánicas podría ser $y = f(F, \sigma_{en}, V_{en}, V_{ex})$, donde y es el espesor de salida, F es la fuerza de laminación, σ_{en} es la tensión de entrada (la tensión de salida se considera constante) y finalmente V_{en} y V_{ex} son las velocidades de entrada y salida respectivamente. Para tener en cuenta el comportamiento dinámico del proceso, se obtiene una función de transferencia linealizada

$$Y(s) = \frac{1}{A(s)} \left[k_1 F(s) + k_2 \sigma_{en}(s) + k_3 V_{en}(s) + k_4 V_{ex}(s) \right] \quad (60)$$

En este modelo, $A(s)$ representa los polos del sistema físico y por tanto define los modos transitorios de vibración, mientras que los términos $k_1 F(s)$, $k_2 \sigma_{en}(s)$, $k_3 V_{en}(s)$ y $k_4 V_{ex}(s)$ constituyen

la dinámica forzada. En este caso, cabe hacer la hipótesis de que $k_1F(s)$ contiene la mayor parte de la información de vibraciones del proceso, puesto que las velocidades se utilizan para obtener el punto de funcionamiento y la señal $\sigma_{en}(s)$ tiene una frecuencia de muestreo que no puede proporcionar información frecuencial sobre el chatter. Por tanto, se elige la fuerza de laminación F para analizar la dinámica del proceso.

4.3.2 *Caracterización del comportamiento dinámico*

El estado de vibración se investiga analizando el comportamiento dinámico del proceso de la laminación mediante un análisis de bandas en frecuencia (*frequency-band analysis*, FBA), el cual ha sido utilizado previamente para el análisis de maquinaria eléctrica y mecánica [50, 181, 13, 163]. La razón principal de utilizar este enfoque reside en que proporciona representaciones más dispersas para las señales periódicas que las representaciones en el dominio temporal. Para tales señales, la energía aparece concentrada en determinadas frecuencias, siendo de menor valor para el resto.

Dada una variable $x(t)$, medida en intervalos regulares de tiempo con un periodo de muestreo T, produce una secuencia $x_k = x(kT)$. Considerando ventanas solapadas de longitud N_w, cada una desplazada $n_d < N_w$ muestras, se obtiene una ventana o *buffer* de datos

$$\mathbf{x}_n = [x_{n\cdot n_d},\, x_{n\cdot n_d+1},\, \cdots, x_{n\cdot n_d+N_w-1}]$$

donde $n_d = (1 - \frac{L}{100}) \cdot N_w$ siendo L el porcentaje de muestras solapadas, y produciendo N segmentos solapados.

La transformación discreta de Fourier (DFT) con ventanas se utiliza de acuerdo con la siguiente expresión:

$$X_i = \sum_{k=0}^{N_w-1} w(k)x_k e^{-\mathrm{j}2\pi ik/N_w}, \qquad i = 0, \cdots, N_w - 1 \qquad (61)$$

donde x_k es una secuencia de datos real o compleja, X_i es una secuencia compleja que describe las amplitudes y fases de los armónicos de la señal y $w(k)$ es la función de ventana (*Hanning* en este trabajo). Para la ventana n, las energías en bandas de m centros de frecuencias $f_1, f_2, \cdots, f_m$ con anchos de bandas predefinidos $B_1, B_2, \cdots, B_m$ se pueden calcular sumando los cuadrados

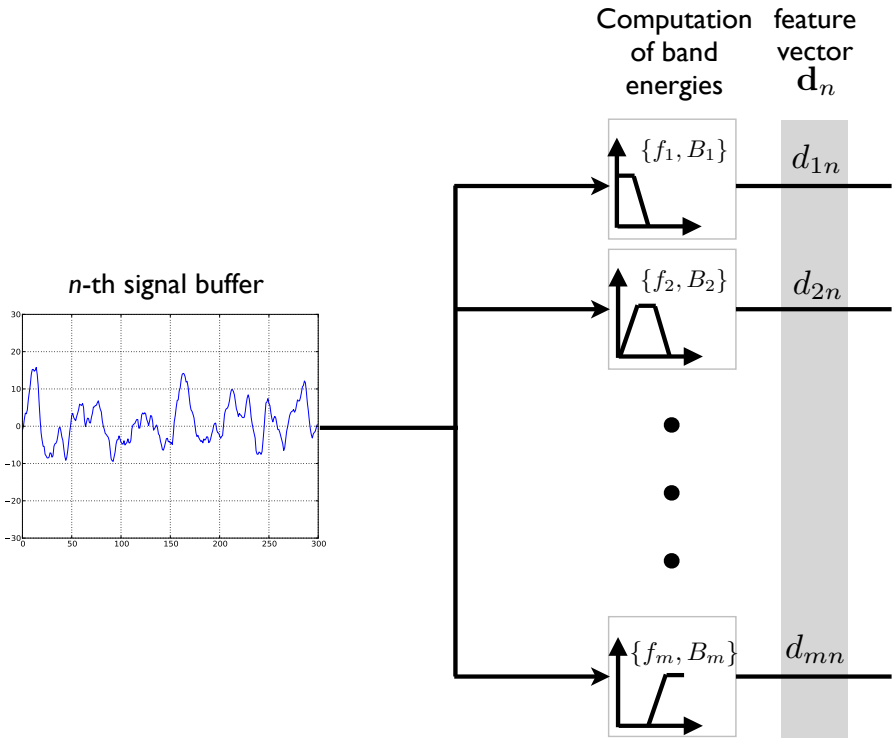


Figura 4.6.: Diagrama de bloques de la extracción de características para una señal. Fuente [150].

de los armónicos dentro de las bandas para obtener un *vector de características* de dimensión m.

$$\mathbf{d}_n = [d_{1n}, d_{2n}, \cdots d_{mn}]^T; \quad d_{jn} = \sqrt{\sum_{\frac{i}{NwT} \in [f_j - \frac{B_j}{2}, f_j + \frac{B_j}{2}]} \|X_i\|^2} \tag{62}$$

Los vectores de características se pueden agrupar en una matriz de datos $\mathbf{D}_x = (d_{jn}) = [\mathbf{d}_1, \mathbf{d}_2, \cdots \mathbf{d}_N]$, donde d_{jn} representa la energía en la banda $\{f_j, B_j\}$ – con centro de frecuencia f_j y ancho B_j– para la ventana n de la señal $x(t)$. El proceso de extracción de características para el *buffer* n se resume gráficamente en la Fig. 4.6.

La información de vibraciones del proceso se construye realizando el FBA descrito para las variables $y(t)$ y $F(t)$, con lo que se obtienen las matrices $\mathbf{D}_y$ y $\mathbf{D}_F$, respectivamente. Finalmente, con estas matrices se construye la matriz $\mathbf{H}$ como sigue

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots \mathbf{h}_N] = \begin{bmatrix} \mathbf{D}_y \\ \mathbf{D}_F \end{bmatrix} \qquad \mathbf{H} \in \mathbb{R}^{2m \times N}, \ \mathbf{h}_n \in \mathcal{D} = \mathbb{R}^{2m} \tag{63}$$

donde $\mathbf{h}_n$ es el vector que contiene el FBA de $y(t)$ y $F(t)$ para la ventana n, y el espacio de datos de entrada $\mathcal{D} = \mathbb{R}^{2m}$ contiene el conjunto total de los estados de vibración del proceso de laminación, caracterizados según el procedimiento descrito.

4.3.3 *Visualización del comportamiento dinámico mediante técnicas de reducción de la dimensión*

Como resultado de esto, se obtienen N vectores de $2m$ dimensiones para cada bobina que describen su estado de vibración. Se parte de la hipótesis de que estos vectores se encuentran en un *manifold* de menor dimensionalidad, por lo que se utilizan técnicas de reducción de la dimensión (DR) para extraer la estructura de los datos y representarla en un espacio de visualización.

Entre las técnicas DR existentes (ver capítulo 3) se ha elegido *t-Stochastic Neighbour Embedding* (t-SNE) [190], reciente técnica no lineal que calcula las similitudes, contenidas en los datos, mediante probabilidades y ha dado buenos resultados en la reducción a 2 o 3 dimensiones de conjuntos reales de datos para su visualización. Sin embargo, algunas limitaciones que posee dicha técnica, como por ejemplo sus tiempos de ejecución para un gran número de muestras (10000), dificulta su aplicación.

Para resolver esto, como se ha descrito previamente, se calcula una cuantización vectorial de los N vectores contenidos en $\mathbf{H}$ mediante el algoritmo *neural gas*. Se obtienen n_r vectores $\{\mathbf{w}_i\}_{1\leq i\leq n_r}$ cuya distribución describe la función de densidad de probabilidad conjunta de los vectores de entrada originales. Estos vectores se proyectan mediante el algoritmo t-SNE con una perplejidad P, comparable al número efectivo de vecinos más cercanos, dando lugar a un conjunto de puntos $\{\mathbf{q}_i\}_{1\leq i\leq n_r}$ de 2 dimensiones. La pareja $(\mathbf{w}_i, \mathbf{q}_i)$ relaciona el espacio de datos de vibraciones en alta dimensión $\mathcal{D}$ con el espacio de visualización $\mathcal{V}$ en baja.

Debido a la mayor rapidez de ejecución respecto a la extensión *out-of-sample* para t-SNE realizada en [149], se ha utilizado como alternativa una red de función de base radial (RBF) para generalizar el mapa a N puntos, y también para calcular aproximaciones para nuevas bobinas de test. Dado un conjunto de prototipos de entrada $\mathbf{w}_i$ y sus correspondientes puntos en la proyección $\mathbf{q}_i$, la red estima la proyección resultante para un nuevo vector de caracterís-

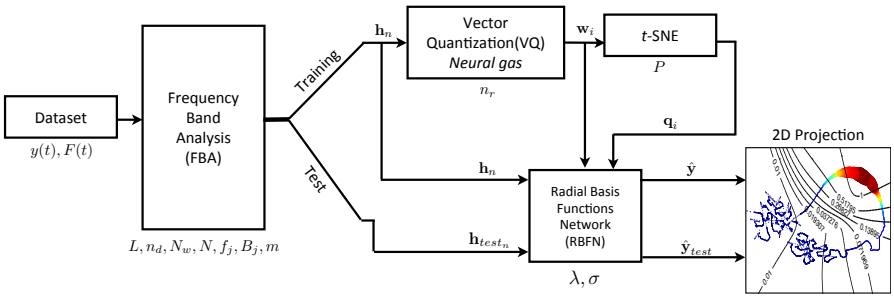


Figura 4.7.: Gráfico del método propuesto. Fuente [150].

ticas $\mathbf{h}$ como $\hat{\mathbf{y}} = \sum_i \mathbf{a}_i \psi_i(\mathbf{h})$ donde $\psi_i(\mathbf{h})$ son *kernels* Gaussianos normalizados.

$$\psi_i(\mathbf{h}) = \frac{\exp\left(-\frac{\|\mathbf{h}-\mathbf{w}_i\|^2}{2\sigma^2}\right)}{\sum_i \exp\left(-\frac{\|\mathbf{h}-\mathbf{w}_i\|^2}{2\sigma^2}\right)} \tag{64}$$

tal que $\sum \psi_i(\mathbf{h}) = 1$, donde σ es el ancho del kernel y los coeficientes $\mathbf{a}_i$ se pueden determinar mediante mínimos cuadrados regularizados $\mathbf{A} = (\mathbf{a}_i)^T = \|\mathbf{\Psi}^T\mathbf{\Psi} + \lambda\mathbf{I}\|^{-1}\mathbf{\Psi}^T\mathbf{Q}$, donde $\mathbf{\Psi}$ es una matriz de kernel tal que $\mathbf{\Psi}_{ji} = (\psi_i(\mathbf{w}_j))$, $\mathbf{Q} = (\mathbf{q}_j)^T$ y λ es el factor de regularización. Un resumen gráfico de los pasos realizados por el método propuesto se representa en la Fig. 4.7.

4.3.4 *Experimentos*

Para una etapa de validación, este método se ha utilizado con datos reales procedentes de un tren de laminación en frío. Como se ha explicado anteriormente, este análisis se enfoca en la relación dinámica entre el espesor de salida $y(t)$ y la fuerza de laminación $F(t)$. Estas variables fueron tomadas mediante un sistema de adquisición de datos con una frecuencia de muestreo de $F'_m = 2000$ Hz ($T' = 5 \cdot 10^{-4}s$). Dado que el chatter aparece en bandas de frecuencia entre 100 y 300 Hz, un diezmado con relación $r = 2$ fue aplicado a las señales para una eliminación regular de muestras, lo que permite obtener un conjunto de muestras de menor tamaño con una frecuencia de muestreo final de $F_m = 1000$ Hz ($T = 10^{-3}s$).

En la Fig. 4.8, se representan las señales consideradas (espesor de salida y fuerza de laminación) durante un episodio de chatter (arriba), así como los efectos que se producen en el espectro de frecuencias (abajo).

Los datos de entrenamiento y de test están compuestos de varias bobinas, incluyendo episodios de chatter y condiciones normales

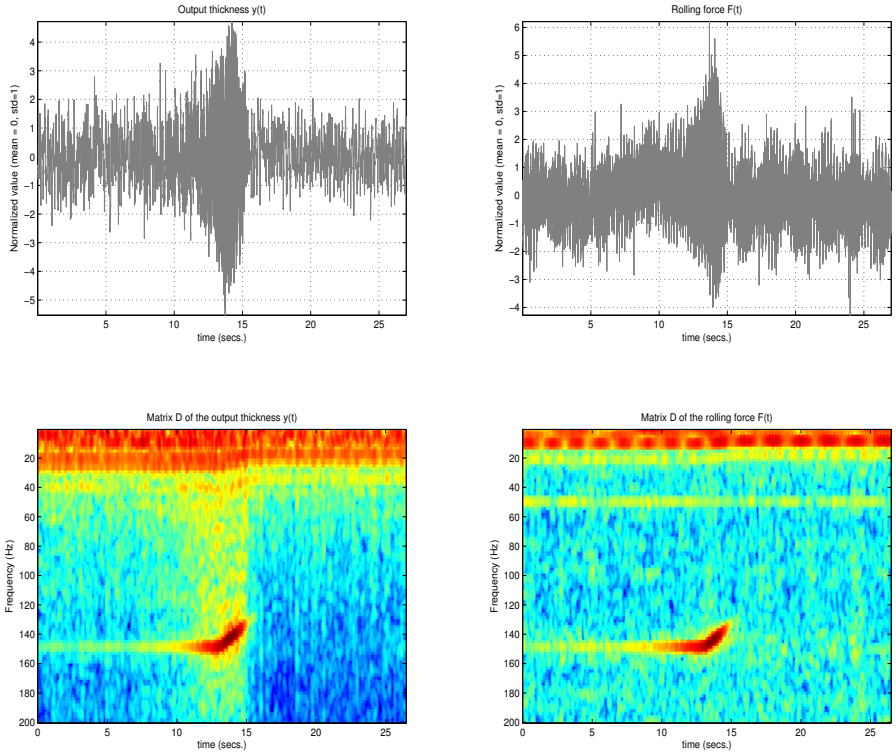


Figura 4.8.: Señales de espesor de salida (izquierda) y fuerza (derecha) de un episodio de chatter (superior) y una visualización tipo espectrograma de la matriz de datos (inferior). Fuente [150].

de funcionamiento, en cada una de ellas. Todas las señales se dividen en ventanas de longitud $N_w = 512$ con un solapamiento de $L = 99\,\%$ $(n_d = 5)$. El FBA se realiza sobre $m = 200$ frecuencias $(f_i = 1, 2, \ldots, 200$ Hz) y anchos $B_1 = B_2 = \cdots = B_{200} = 5$ Hz. Estos valores tienen el objeto de obtener una transición progresiva del contenido frecuencial de las señales. El tamaño total del conjunto de datos de entrenamiento es de $N = 4415$ vectores. La cuantización vectorial mediante el algoritmo *neural gas* reduce su tamaño a $n_r = 600$. Para la etapa DR, se aplica el algoritmo t-SNE con $P = 30$. Finalmente, la RBFN se entrena con un ancho del kernel $\sigma = 100$ y un parámetro de regularización $\lambda = 10^{-9}$.

Aquí, el número de prototipos fue elegido para alcanzar razonables tiempos de cálculo. Puesto que el objetivo del método es proporcionar un análisis visual, los parámetros para el t-SNE y RBFN fueron establecidos buscando un compromiso entre la suavidad de la proyección y la capacidad para separar condiciones de chatter y estados normales en las visualizaciones para las bandas

de entrenamiento y validación. En general, la experimentación sugirió el uso de pequeños valores del factor de regularización λ y valores del ancho σ dentro del mismo orden de magnitud que los datos en el espacio de entrada. Los parámetros utilizados en los experimentos se resumen en la Tabla 2.

Adquisición de datos		
Nombre	*Símbolo*	*Valor*
Frecuencia de muestreo inicial	F'_m	2000 Hz.
Periodo de muestreo inicial	T'	$5 \cdot 10^{-4}$ s
Relación de diezmado	r	2
Frecuencia de muestreo inicial	F_m	1000 Hz.
Periodo de muestreo final	T	10^{-3} s
Análisis de bandas en frecuencia (FBA)		
Nombre	*Símbolo*	*Valor*
Tamaño de ventana	N_w	512
Solapamiento	L	99 %
Desplazamiento	n_d	5
# de frecuencias	m	200
Centros de frecuencias	f_i	$1, 2, 3, \ldots, 200$ Hz.
Anchos de banda	B_i	5 Hz.
# de puntos	N	4415
Reducción de la dimensión (DR)		
Nombre	*Símbolo*	*Valor*
# de prototipos	n_r	600
Perplejidad	P	30
Ancho de kernel RBFN	σ	100
Parámetro regularización RBFN	λ	10^{-9}

Tabla 2.: Resumen de los parámetros utilizados en los experimentos.

4.3.5 *Resultados y discusión*

De acuerdo con la metodología descrita, se calculó la proyección para el conjunto de entrenamiento seleccionado y se representa en la parte superior de la Fig. 4.9. Las diferentes condiciones dinámicas están directamente relacionadas con la posición de los puntos en el mapa obtenido. Puntos que se encuentran cercanos en el mapa representan comportamientos dinámicos similares y viceversa, con el color y el tamaño indicando la energía media en la banda de frecuencias considerada para la aparición del fallo. Las proyecciones muestran una trayectoria fuera de las condiciones normales de funcionamiento, la cual representa la evolución dinámica de un episodio de chatter durante la laminación de una banda de acero.

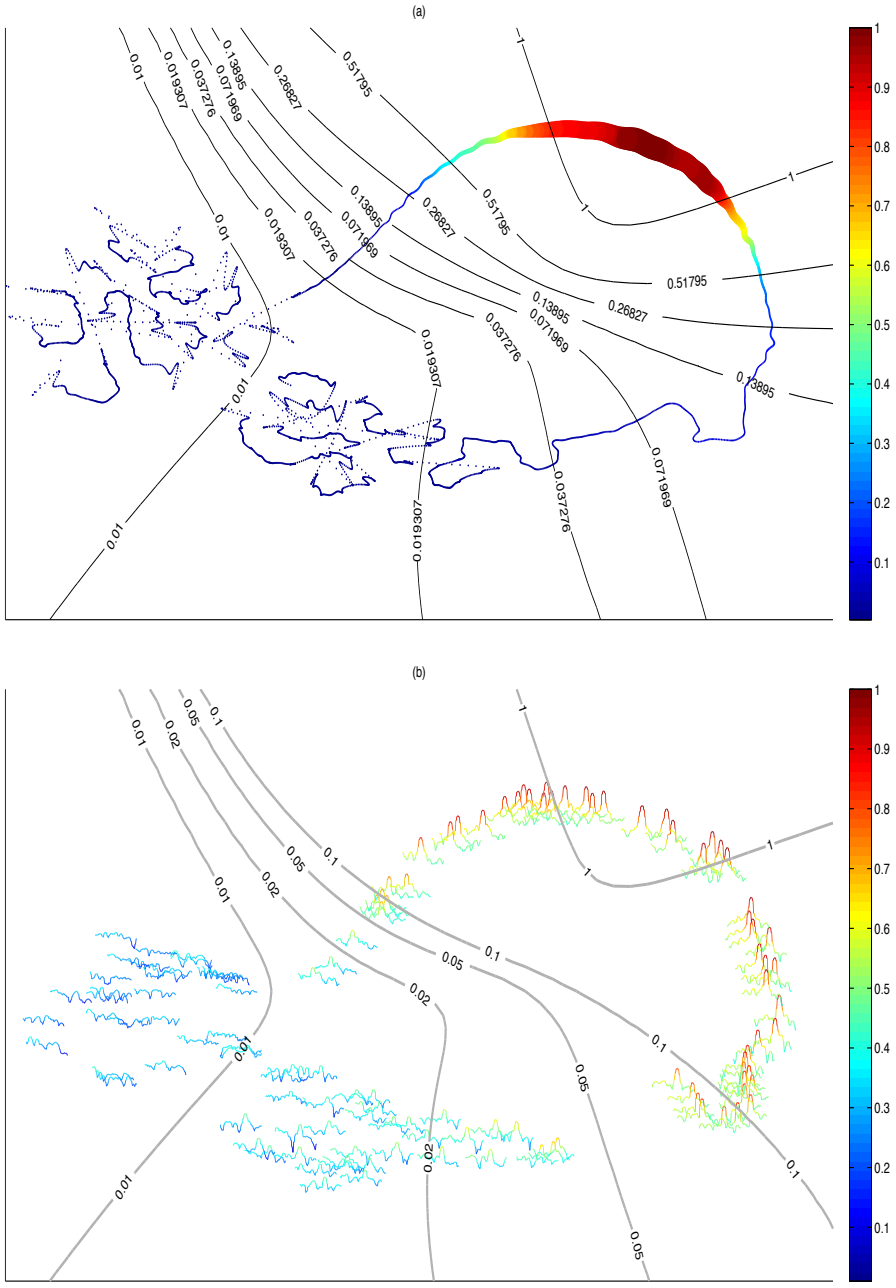


Figura 4.9.: Mapa de la proyección de los datos de entrenamiento. Las líneas de contorno representan la energía media en la banda de frecuencias de chatter codificada con color y tamaño (arriba)y con los perfiles de frecuencia de cada estado (abajo). Fuente [150].

Otra red RBF se utiliza para estimar el valor esperado de la energía de chatter mediante una malla regular en puntos del espacio de proyección 2D. Esto permite construir *isolíneas* de contorno de chatter (líneas de contorno de energía media en la banda de frecuencias para el modo de vibración del fallo) que sirven como fronteras de decisión para la predicción del chatter. Como se explica en los experimentos, todas las bobinas de validación abarcan una región de condiciones normales de funcionamiento y se mueven a una región de chatter, volviendo después a la condición normal. La región entre las dos condiciones, normales y de chatter, junto con las líneas de contorno, resultan ser pistas visuales para una detección temprana de chatter.

En la Fig. 4.9 (b) se representa en cada punto del mapa un glifo de la distribución de las energías de las bandas de frecuencia entre 100 y 180 Hz. Tal representación da una vista completa del contenido de frecuencias en el que el chatter aparece. En la Fig. 4.10, se representan varias proyecciones de diferentes bandas del proceso utilizadas como test y representadas con una escala de color, sobre la proyección de entrenamiento, la cual se muestra en una escala de grises. Las proyecciones de los puntos de un conjunto nuevo de datos se sitúan en zonas correspondientes a comportamientos dinámicos similares de las proyecciones de entrenamiento. Puede comprobarse que puntos de una nueva banda sin episodio de chatter, permanecerán en el área correspondiente a condiciones de operación normales mientras que si ocurre un fallo, se representará en el área correspondiente de chatter. Las fronteras de decisión establecidas en el mapa ayudan a detectar posibles episodios de chatter en la producción de nuevas bandas. Para todas las pruebas realizadas, se puede ver que las fronteras entre los valores 0,05 y 0,7 corresponden a una situación donde la aparición del chatter es inminente –una condición de pre-chatter– lo que puede resultar útil para una detección temprana del fallo.

4.4 CONCLUSIONES

Las aplicaciones descritas muestran casos reales en los que se realiza un análisis visual de diversos procesos con dinámica. Se procesan datos multidimensionales procedentes de un proceso para una extracción de información útil mediante algoritmos inteligentes y se proyecta en un mapa, que sirve para supervisar visualmente el proceso. Además, el uso de mecanismos de interacción mejora la experiencia del usuario en la exploración visual de los datos.

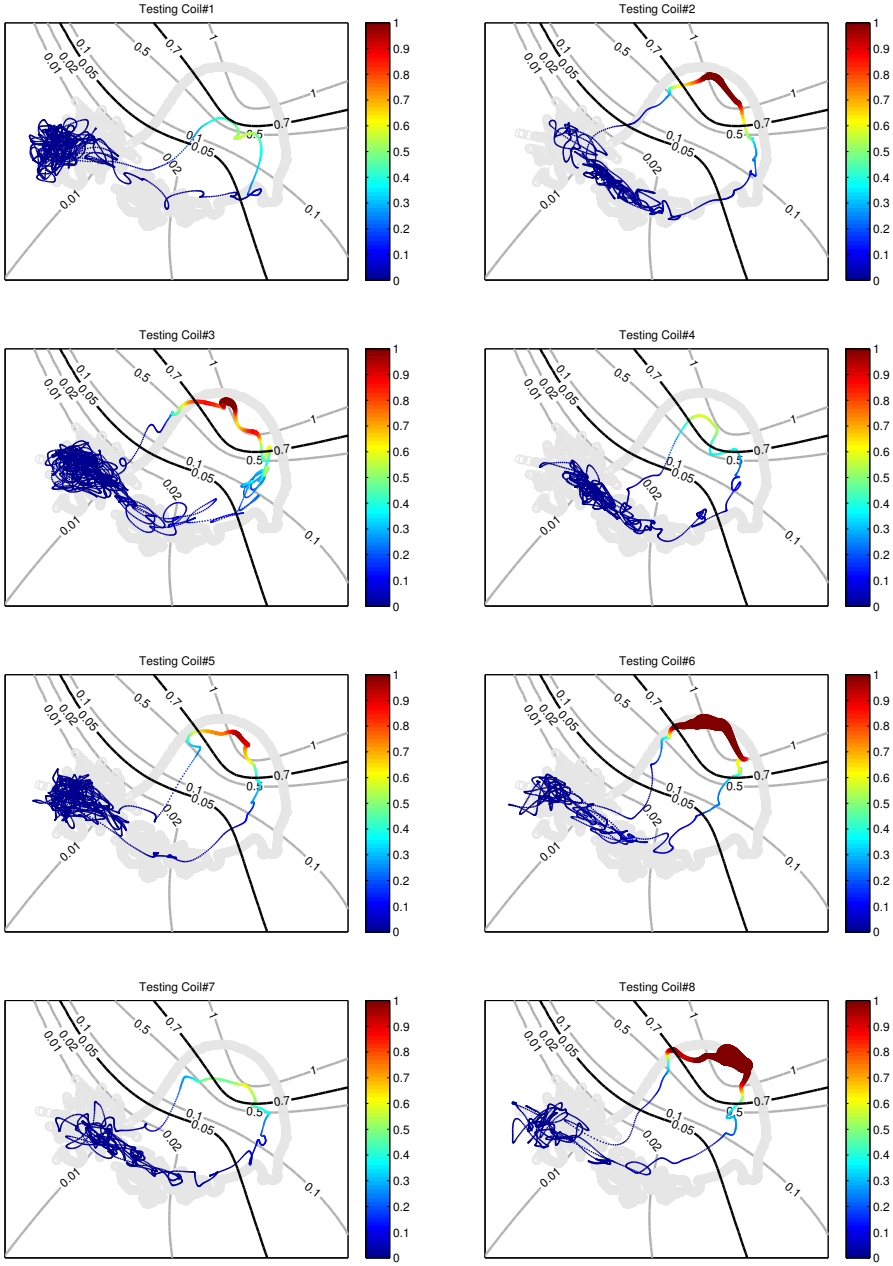


Figura 4.10.: Proyecciones de varias bobinas sobre un mapa de entrenamiento. Fuente [150].

Se han presentado ejemplos para supervisar patrones de consumos eléctricos que facilitan su rápida identificación visual y relación con diferentes periodos temporales. Esto puede ser utilizado para actuar en los hábitos de demanda tomando decisiones respecto a eficiencia energética. También se ha estudiado el fallo de chatter que puede ocurrir en un tren de laminación en frío, ayudando a supervisar este proceso de laminación con este tipo de fallos severos. Los contornos representados en las proyecciones facilitan la detección de una condición temprana del chatter y puede ayudar a una posible predicción del fallo.

La metodología propuesta puede aplicarse no sólo a los problemas planteados de chatter en la laminación en frío o el análisis de la demanda eléctrica sino también para otros, como por ejemplo procesos industriales con comportamientos variantes de operación o en casos más generales modificando la primera etapa de extracción de características.

5

EXPLORACIÓN INTERACTIVA DE PROYECCIONES DE DATOS MULTIDIMENSIONALES

En este capítulo se presenta una técnica que por medio de una transformación de los datos, permite al usuario introducir conocimiento nuevo, como información de clases, y realizar una exploración interactiva de las proyecciones.

5.1 INTRODUCCIÓN

El análisis de datos multidimensionales mediante el estudio de sus proyecciones es un enfoque muy utilizado para la identificación eficiente de patrones. Con la utilización del *principio de espacialización* [58] se crea una representación visual en la que cada muestra de los datos se presenta como un punto en un mapa en dos o tres dimensiones, donde la posición de dichos puntos se asocia a la similitud entre muestras, de manera que puntos cercanos representan objetos similares y lejanos objetos distintos. Estos mapas se calculan mediante técnicas de reducción de la dimensión (DR) [119, 192] que, basándose en una función de distancia para evaluar las similitudes entre los puntos, estiman la estructura latente de los datos en un espacio de menor dimensionalidad. En el capítulo 3 se expone un resumen de la amplia variedad de técnicas DR existentes.

A pesar de la gran cantidad de métodos propuestos para el cálculo de proyecciones de datos multidimensionales [119], los resultados que se obtienen no son siempre satisfactorios, sobre todo en escenarios reales donde los puntos proyectados aparecen superpuestos y las fronteras entre grupos de puntos solapadas. Las razones pueden ser varias, como el gran tamaño de los datos con estructuras latentes complejas, el ruido existente en ellos o variables irrelevantes que no representan correctamente los datos. Estos hechos son muy comunes en casos reales de análisis, y pueden llevar a que distancias entre las muestras no queden claramente visibles para estos algoritmos.

Esto hace que se obtengan proyecciones que no muestran partes de los datos de interés específico del analista o dónde existen patrones significativos. Además puede darse también el caso de que

los datos no contengan por sí solos toda la información útil para el análisis, como por ejemplo información categórica de clases. Existen técnicas DR supervisadas que tienen en cuenta este tipo de información en su algoritmo para la estimación de la estructura de los datos [66, 74, 75, 178]. Sin embargo, en muchos casos no puede garantizarse una proyección satisfactoria. Para mejorar la claridad visual de las proyecciones se ha propuesto recientemente un enfoque [151, 152], el cual realiza una extensión de características usando información previa de clases, y modifica los mapas visuales para su mejor interpretación. El desarrollo de este enfoque es el objeto del presente capítulo.

5.2 PROYECCIÓN Y TRANSFORMACIÓN INTERACTIVA DE DATOS MULTIDIMENSIONALES

Las técnicas DR clásicas [192, 119, 161] estiman la estructura latente en la que datos de alta dimensión se encuentran situados con una menor dimensión topológica. Esta dimensión intrínseca es el número mínimo de parámetros necesarios para definir la geometría de la estructura. Cuando estas técnicas DR se utilizan con propósitos de visualización, se obtiene una proyección en un mapa 2D que puede ser insatisfactoria. La estimación de esa estructura latente, realizada en la reducción, implica cierta pérdida de información con respecto a los datos de entrada. Además, si la dimensión intrínseca de los datos es mayor que dos, la estructura calculada puede visualizarse en el mapa ocluida por los distintos grupos existentes en los datos. Para algunos algoritmos automáticos resulta difícil decidir qué información es más relevante para las tareas específicas del análisis.

Como se ha descrito en el capítulo 2, la interacción integra al ser humano en el proceso de análisis permitiéndole adaptar la configuración de una visualización usando su conocimiento actual [193]. Esta manipulación de las representaciones, combinada con el valor que posee la propia visualización de la información [62], proporciona una potente ayuda en el proceso exploratorio de los datos y puede acelerar la obtención de conocimiento nuevo. Por medio de la interacción, el analista puede utilizar sus habilidades cognitivas para conducir el proceso de análisis, extraer conclusiones o incluso generar nuevas hipótesis de manera eficiente. Aunque una mera animación no es garantía para mejorar una representación estática [188], se ha demostrado que las transiciones animadas, con un diseño adecuado, pueden mejorar la percepción de cambios entre gráficos estadísticos [86].

La integración de mecanismos de interacción con métodos estadísticos para ayudar al análisis exploratorio de datos se discute en [59]. En dicho trabajo se evaluaba la interacción a nivel de observación de los datos en lugar de interacciones directas en los parámetros implicados en algunas técnicas DR. Por tanto, se postula una interacción enfocada a dotar de sentido a los propios datos presentados sin ser necesaria una interacción para la actualización de los parámetros que intervienen en el análisis.

La idea de proporcionar interacción al cálculo de proyecciones ha sido adoptada en varios trabajos previamente. Por ejemplo, el sistema denominado *iPCA* [97] propone una interfaz donde se facilita la exploración interactiva de datos multidimensionales y de su proyección por medio del método PCA. Para ello utiliza varias vistas coordinadas y un conjunto de mecanismos de interacción, que facilitan las tareas de análisis de manera intuitiva. Este sistema ayuda al usuario a un mejor entendimiento de la relación entre los datos, y el espacio calculado por la técnica utilizada (PCA). También, de forma similar, el *iVisClassifier* [40] presenta una interfaz de analítica visual interactiva para tareas de clasificación. Por medio de la técnica DR supervisada LDA y de varios mecanismos de interacción, mejora la interpretación de la técnica aplicada a los datos. Utiliza vistas de coordenadas paralelas, *scatterplots*, y *heat maps* de forma interactiva para mostrar diferentes aspectos de los datos. Permite al usuario entender las dimensiones reducidas y analizar su influencia en los datos originales. Además, ayuda al usuario en el estudio de grupos en los datos de manera eficiente y en su exploración utilizando la técnica LDA. En la Fig. 5.1 se representan capturas de pantalla de ambas interfaces.

El estudio de técnicas DR con controles interactivos, para guiar al usuario por distintas posibilidades de análisis, se presenta en el trabajo de Johansson [98]. Este enfoque combina un método de reducción de la dimensión, con métricas de calidad que pueden ser definidas por el usuario. En el artículo, las métricas utilizadas están relacionadas con el análisis de correlaciones, la detección de valores atípicos (*outliers*), y el análisis de grupos (*clusters*). La importancia de las diferentes métricas se mide por medio de funciones ponderadas que definen un valor de importancia global. A través de varias visualizaciones e interacciones en las estructuras de datos multidimensionales, la técnica ayuda al usuario en la decisión de eliminar o mantener las variables que se utilizan en la reducción, teniendo en cuenta la pérdida de información que se produce en dicho proceso. En la parte superior de la Fig. 5.2 se

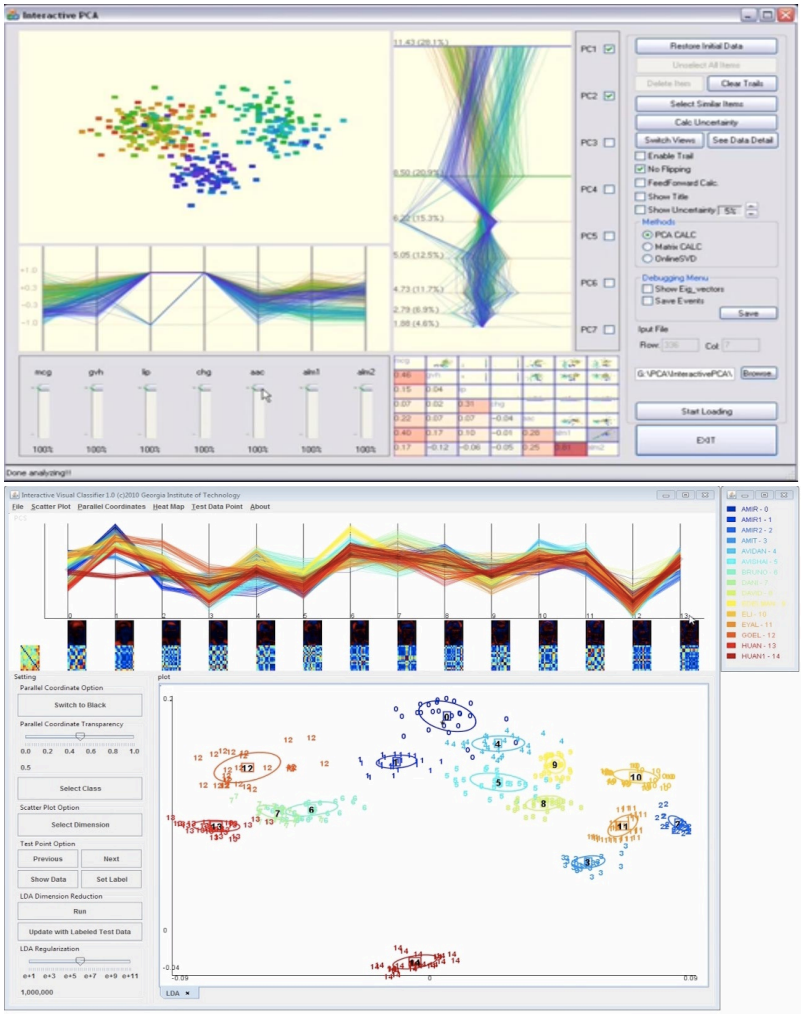


Figura 5.1.: Capturas de pantalla de las interfaces interactivas de *iPCA* [97] (parte superior) y de *iVisClassifier* [40] (parte inferior).

muestran representaciones de la interfaz que recogen varias vistas de los datos (izquierda) y la pérdida de información (derecha).

Similar a este enfoque, el sistema *DimStiller* [93] guía al usuario en el proceso de reducción de la dimensión por medio de un conjunto de abstracciones. Utiliza tablas de los datos, las visualiza y permite operaciones entre ellas. Estos operadores se pueden encadenar formando distintas expresiones definidas por dichos operadores y por el orden en el que se aplican. Con esto pueden crearse flujos de trabajo, los cuales son plantillas para la exploración que constan de una expresión con los parámetros de los operadores almacenados, por lo que pueden reutilizarse. El sistema estima inicialmente la dimensión intrínseca de los datos, y realiza un análisis por medio de controles interactivos y vistas enlazadas que permite al usuario manipular resultados en etapas intermedias. Proporciona una guía local y global para el analista, a través de la representación de las tablas de los datos durante el proceso de análisis y reducción. En la parte inferior de la Fig. 5.2 se muestra un gráfico representando una simple expresión (parte izquierda) en la que cada operador podría tener controles o vistas y la interfaz (parte derecha) que tendría esa expresión.

La modificación interactiva de puntos en las proyecciones es un enfoque que también ha sido propuesto. Varios ejemplos se muestran en [145], donde se describen técnicas de proyección en las que el usuario puede incorporar su conocimiento al proceso exploratorio. Por ejemplo, *Local Affine Multidimensional Projection* (LAMP) [99, 130] permite al usuario modificar algunos puntos (de control) en la representación visual, calcula una transformación afín y la utiliza para obtener una nueva proyección para todos los puntos a partir de los cambios realizados. También *Dis-function* [28] proporciona una interacción similar donde la modificación de los puntos en una proyección, a partir del conocimiento del usuario, se utiliza para calcular una nueva función de distancia final. Por medio de operaciones como seleccionar, arrastrar y soltar los puntos, la técnica calcula la nueva función de distancia en la que las variables son ponderadas, de manera que las variables resultantes con menor peso podrían ser incluso descartadas.

También, dada información de grupos en los datos como etiquetas de clases, el analista podría querer introducir ese conocimiento en la proyección. En [162] se estudia la transformación de ciertas características, basándose en información de clases. Con un par de estrategias de selección automática de las variables, se demuestra una mejora visual de las proyecciones resultantes, evaluadas con medidas de calidad visuales, descritas en el capítulo 3, que miden

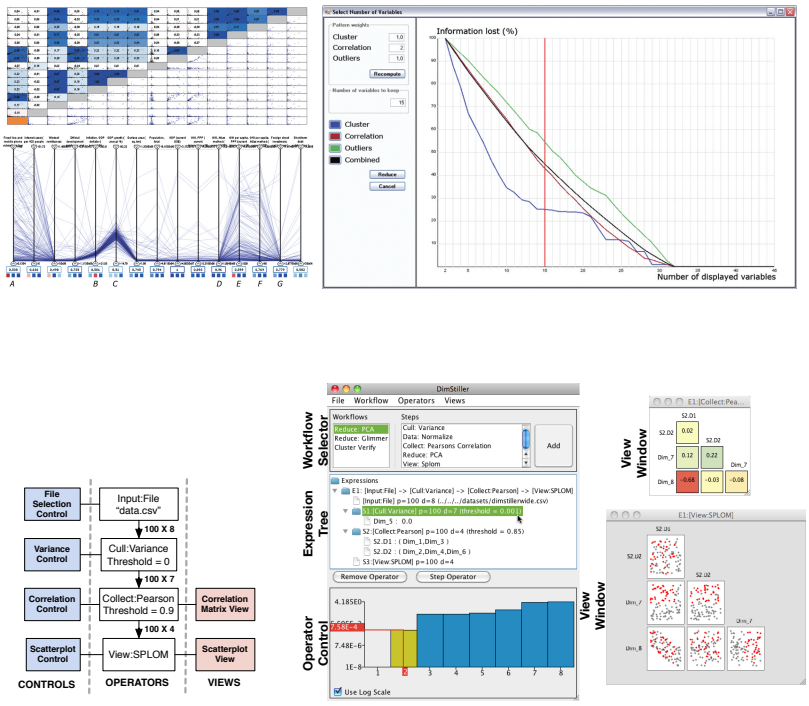


Figura 5.2.: Vistas de coordenadas paralelas y *scatterplot matrix* de un reducido grupo de datos (parte superior, izquierda) y la representación de pérdida de información (parte superior, derecha) recogidas de [98] (parte superior). Un esquema de una expresión simple (parte inferior, izquierda) y la interfaz *DimStiller* para esa expresión (parte inferior, derecha). Fuente: [93].

el área y la densidad de solapamiento entre grupos, y con medidas que miden la preservación estructural que se produce en el proceso de reducción de la dimensión. Este enfoque constituye la base para el método interactivo que se propone en los siguientes apartados.

5.3 EXTENSIÓN INTERACTIVA DE CARACTERÍSTICAS

El enfoque propuesto en esta tesis combina la transformación del espacio de características, la interacción y la visualización para la modificación de la proyección con el objeto de conseguir una mejor interpretación de los datos multidimensionales proyectados. Dado un conjunto de datos multidimensionales, el método genera un espacio de características extendido a partir de la información de grupos que se desee integrar en la proyección, que se añade al espacio original de los datos, modificando la proyección. El analista puede seleccionar ciertas variables, o el conjunto completo, para generar dicho espacio extendido y modificar la proyección original gradualmente, de manera que la dirija finalmente a una proyección satisfactoria para facilitar su análisis. Además, las proyecciones son evaluadas mediante varias medidas de calidad para una validación cuantitativa. En la Fig. 5.3 se representa un diagrama de flujo del proceso de análisis iterativo.

Básicamente, la transformación del espacio de características aplicada en este método extiende ciertas variables, basándose en información de grupos. Se considera un conjunto de datos $\mathbf{X}$ en forma de matriz, de manera que sus filas y sus columnas representan las muestras y las variables de los datos, respectivamente. Las etiquetas y_i representan la clase correspondiente de la fila i,

$$\mathbf{X} = [x_{ij}] \in \mathbb{R}^{n \times d} \qquad \mathbf{y} = [y_i] \in \mathbb{N}^n \tag{65}$$

con $i = 1, \ldots, n$ y $j = 1, \ldots, d$, siendo n el número de vectores de características y d el número de dimensiones respectivamente. Suponiendo m variables elegidas $v = v_1, \ldots, v_m$, la matriz de datos extendida $\mathbf{X'}$ se define de la siguiente manera:

$$\mathbf{X'} = [x_{ij} \mid \tilde{x}_{ij}] \in \mathbb{R}^{n \times (d+m)} \tag{66}$$

Siendo $\tilde{x}_{ij}$ el descriptor estadístico correspondiente a la clase y_i en la variable v_j. Aquí se utiliza la media aritmética dentro de las clases en una determinada dimensión. Por ejemplo, suponiendo un conjunto de datos de 4 muestras y 2 variables, que pertenecen a 2 clases distintas:

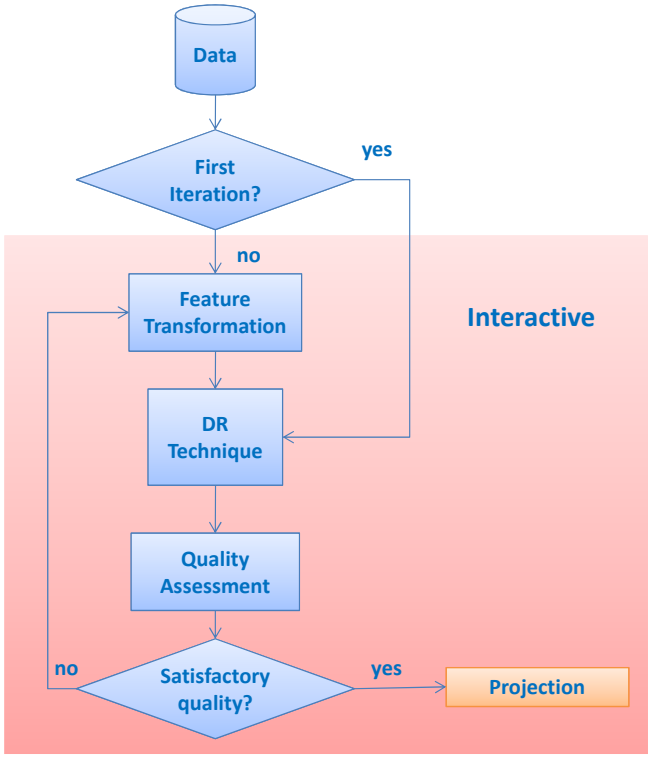


Figura 5.3.: Diagrama de flujo del método propuesto [151].

Sepal.Length	Sepal.Width	Species
5.1	3.5	*setosa*
4.9	3.0	*setosa*
7.0	3.2	*versicolor*
6.4	3.2	*versicolor*

Las etiquetas de clase (*Species*) se utilizan calculando la media de cada clase en cada dimensión:

	Sepal.Length	Sepal.Width
$media_{setosa}$	5.0	3.25
$media_{versicolor}$	6.7	3.2

La matriz $\mathbf{X}'$ se construye de forma siguiente:

Original		*Extended*		
dim1	dim2	ext1	ext2	Species
5.1	3.5	5.0	3.25	*setosa*
4.9	3.0	5.0	3.25	*setosa*
7.0	3.2	6.7	3.2	*versicolor*
6.2	3.0	6.7	3.2	*versicolor*

Aunque aquí se utiliza la media con propósitos ilustrativos, podría usarse otro descriptor estadístico distinto para la extensión. Esta matriz extendida incorpora información de clases a los datos originales. Al complementarse mediante la matriz $\mathbf{X}'$ los datos originales con la matriz extendida en el algoritmo DR, las distancias y, por tanto, la proyección resultante se modifican, reflejándose de manera combinada la estructura original de los datos y la estructura de clases. El peso de cada uno de los factores puede modularse, de forma interactiva, como se explica en el siguiente apartado.

5.3.1 *Extensión ponderada del espacio de características*

Con la matriz de datos y las etiquetas, definidas en la ecuación (65), la matriz resultante $\mathbf{X}'$ para todo el espacio de características se construye a partir de la original $\mathbf{X}$ y la parte extendida $\tilde{\mathbf{X}}$, de la forma siguiente:

$$\mathbf{X}' = \left[\mathbf{X} \mid \tilde{\mathbf{X}}\right] \tag{67}$$

Puesto que se utiliza todo el conjunto de variables para hacer la extensión, se tiene que $\mathbf{X}' \in \mathbb{R}^{n\times 2d}$. Y en este caso, utilizando las medias de cada clase, $\tilde{\mathbf{X}}$ está compuesta por los centroides de las clases descritas en las etiquetas:

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_i] \in \mathbb{R}^{n\times d} \quad \text{siendo } \tilde{\mathbf{x}}_i = \frac{1}{|C_{y_i}|} \sum_{i\in C_{y_i}} x_{ij} \tag{68}$$

con C_{y_i} representando al conjunto de índices de muestras pertenecientes a la clase y_i.

El uso de un parámetro real $\lambda \in [0, 1]$ permite una transición gradual entre el conjunto original y la parte extendida aplicada, por medio de un simple cambio en la métrica en los datos:

$$\mathbf{X}_{weight} = \mathbf{X}'\mathbf{W}_\lambda \tag{69}$$

siendo $\mathbf{W}_\lambda \in \mathbb{R}^{2d \times 2d}$ como sigue:

$$\mathbf{W}_\lambda = \left(\begin{array}{c|c} (1-\lambda)\ \mathbf{I} & \mathbf{0} \\ \hline \mathbf{0} & \lambda\ \mathbf{I} \end{array} \right),\ \lambda \in \mathbb{R} \tag{70}$$

La matriz $\mathbf{X}_{weight}$ es la matriz ponderada de datos utilizada para calcular la nueva proyección. El parámetro λ puede modificarse mediante mecanismos de interacción en la interfaz, lo que permite al usuario controlar la integración del conocimiento previo en forma de clases. Con $\lambda = 0$ la proyección representa la estructura original de los datos de partida y con $\lambda = 1$ representa la información pura de las clases proyectada, siendo en este caso los centroides de las clases.

Un buen punto de partida para el método propuesto es aplicar la extensión ponderada de todas las variables con $\lambda = 0$, lo que da lugar a la proyección original de los datos de partida. Las etiquetas de clases seleccionadas se integrarán a medida que se incrementa el valor de λ interactivamente. De esta manera, el analista puede modificar la proyección y volver al estado original en cualquier momento del análisis.

5.3.2 *Posibles mejoras computacionales*

El método propuesto es independiente de la técnica DR utilizada para calcular las proyecciones, por lo que también hereda los inconvenientes que poseen, como por ejemplo aquellos debidos a limitaciones de memoria. Los costes computacionales tienen un papel importante, puesto que el método propuesto debe calcular una proyección nueva en cada modificación interactiva que se realice.

Sin embargo, podrían aplicarse varios enfoques existentes para manejar grandes volúmenes de datos. Por ejemplo, Li et al. [124] proponen una mejora en el cálculo de la descomposición de valores singulares (SVD), que puede ser aplicada cuando se usen métodos DR que requieran una SVD en el cálculo de la proyección. El método recoge un subconjunto de columnas de la matriz de datos y realiza su cálculo por medio de algoritmos de aproximación. También Yang et al. [209] proponen un novedoso enfoque para reducir

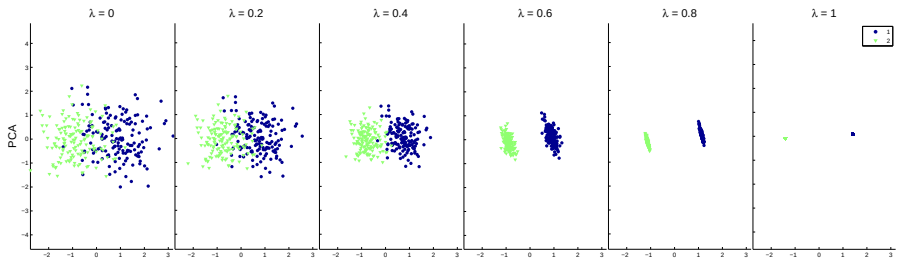


Figura 5.4.: Proyecciones de los datos de dos grupos 2D con una extensión ponderada para varios valores de λ, utilizando la técnica PCA.

el coste computacional de métodos *Neighbor Embedding*. Consideran que en estos métodos las interacciones entre muestras lejanas de datos no contribuyen mucho al gradiente en la etapa de convergencia, por tanto dicha contribución puede aproximarse para ganar eficiencia computacional.

Finalmente en [29] se estudia la incorporación de conocimiento previo de los datos basado en etiquetas mediante una modificación del enfoque de cuantización vectorial con matrices de relevancia. Con esta modificación propuesta, se consigue calcular una proyección de forma supervisada, con un menor coste computacional con respecto a anteriores técnicas y empleando un menor número de parámetros. Además este método aporta información sobre la influencia de variables en tareas de clasificación.

5.3.3 *Un ejemplo ilustrativo*

En este apartado se presenta un ejemplo muy sencillo para ilustrar el método propuesto. Los datos utilizados consisten en dos grupos *gaussianos*, en 2 dimensiones, de 150 puntos cada uno con un pequeño solapamiento. El método se aplica a los datos siguiendo la extensión ponderada de características utilizando los valores medios de cada clase correspondiente a cada grupo. La técnica DR que se utiliza en este ejemplo es el análisis de componentes principales (PCA). En la Fig. 5.4 se representan las proyecciones resultantes para varios valores del parámetro λ, en las cuales los grupos se diferencian por medio del color de cada punto.

La proyección para $\lambda = 0$ corresponde a la proyección de los datos originales, donde no se muestran los dos grupos completamente separados. A medida que el parámetro λ aumenta, la proyección se modifica revelando la información de los grupos. Puesto que la distancia dentro de cada grupo no varía, la estructura local en

Nombre	Tamaño	Dimensiones	Clases
3D clusters	500	3	5
synthetic-gaussian	500	10	5
eCons (Weekday)	338	24	7
eCons (Month)	338	24	12
hiv	78	159	6
yeast	1452	7	10

Tabla 3.: Descripción de los datos

cada grupo se conserva en las nuevas proyecciones. Para el valor más alto ($\lambda = 1$), la proyección se calcula con la información pura de los grupos de los datos, que corresponde a los valores medios de cada grupo, por lo que todos los puntos se encuentran concentrados en la proyección de los centroides correspondientes a cada grupo. Por tanto, estas proyecciones con valores altos de λ (como por ejemplo $\lambda > 0{,}6$) no resultan tan útiles para una correcta interpretación de los datos, por lo que pueden despreciarse para las tareas de análisis.

Además, se puede realizar una evaluación numérica de las proyecciones resultantes, de modo que proporcione más información al usuario para juzgar el punto óptimo de la transformación. Esto se explica en detalle en el apartado 5.5.

5.4 EXPERIMENTOS Y RESULTADOS

En esta sección se presentan los experimentos realizados y resultados obtenidos para validar el método propuesto mediante varios casos de estudio utilizando diferentes conjuntos de datos. Estos conjuntos se han seleccionado de manera que representen varias dimensiones, distinto número de clases, datos sintéticos y reales (ver tabla 3). Se presentan los siguientes casos de estudio para comprobar el método desde varias perspectivas:

1 - Datos sintéticos vs. reales: En este primer caso se aplica el método en dos conjuntos sintéticos de datos, en el resto de casos se aplicará en conjuntos reales de datos.

2 - Series temporales: En este caso se muestra un ejemplo para analizar datos dependiendo de la información temporal, como los meses o el tipo de día de la semana en un conjunto propio de datos.

3 - Extensión de variables seleccionadas vs. todo el conjunto: En este caso se muestran las posibilidades en cuanto a extender sólo unas variables determinadas, seleccionadas en base al conocimiento de los datos. En el resto de casos se utiliza el conjunto total de variables.

4 - Agrupamiento natural de los datos: Aquí se aborda el caso de que la información de etiquetas de clase no se encuentre disponible. El método se aplica utilizando los grupos resultantes de un análisis previo de agrupamiento (*clustering*) de los datos.

5 - Supervisado vs. no supervisado: En el último caso se aplica el método utilizando técnicas DR supervisadas, mientras que en el resto se usan técnicas DR no supervisadas.

Todos los experimentos expuestos comienzan con el cálculo de una proyección original de los datos empleando una técnica DR estándar, es decir, con el parámetro $\lambda = 0$. Las técnicas DR no supervisadas utilizadas son PCA [100] y t-SNE [190], muy habituales para la exploración y visualización de datos. En el caso de las técnicas DR supervisadas se eligieron NCA [75] y MCML [74]. En todos los casos, el espacio original se extiende utilizando la extensión de características mediante las medias, como se describe en la sección 5.3.

Las proyecciones son calculadas utilizando las implementaciones en Matlab de los algoritmos provistas por la *toolbox* [191]. Tanto las proyecciones originales, como las proyecciones con la extensión aplicada se calculan utilizando la misma técnica DR, después de aplicar una normalización *z-score*, es decir, con media cero y desviación típica uno. Para una mejor comparación entre las proyecciones, se han alineado empleando una transformación lineal mediante *Procrustes Analysis* [107].

A continuación, se detallan los casos estudiados, con los resultados de los experimentos. La evaluación de las proyecciones se discute en el apartado 5.5.

5.4.1 *Casos sintéticos*

Ejemplo 3D clusters

Para ilustrar la idea, se aplica el método en un conjunto sintético de datos tridimensionales. Los datos constan de 5 grupos *gaussianos* de 100 puntos cada uno de ellos. En la Fig. 5.5 se representa la estructura original del conjunto de datos, donde se puede ver que

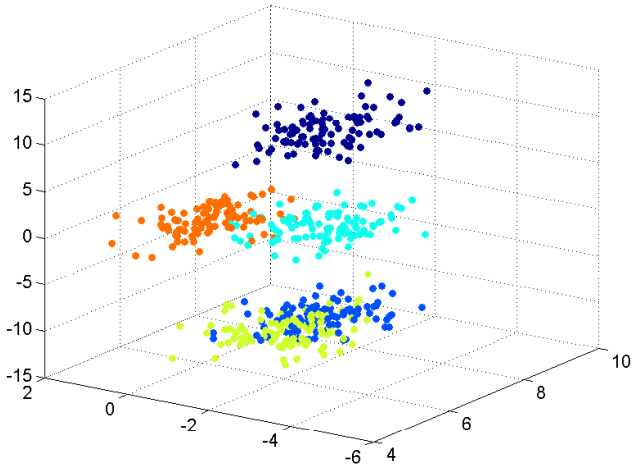


Figura 5.5.: Representación de los datos *3D clusters*

los grupos situados en la parte superior (con colores azul marino, naranja y cian) se encuentran separados, pero los dos en la parte inferior (de color azul y verde) se encuentran solapados entre sí.

Se aplica una extensión ponderada a los datos con la información de estos grupos, utilizando los valores medios correspondientes a cada clase. En la Fig. 5.6 se representan las proyecciones resultantes de la transformación para varios valores de λ, utilizando PCA (arriba) y t-SNE (abajo) con un valor de perplejidad de 20, parámetro comparable al número de vecinos en otras técnicas.

El proceso de transformación de los datos, en el cual se integra la información de los grupos, modifica los datos originales de forma que los grupos se encuentran mejor separados en las nuevas proyecciones. Para este ejemplo, con valores de λ entre 0.4 y 0.6, ambas técnicas DR producen proyecciones con una separación clara de los grupos. Para $\lambda = 0$ la proyección corresponde a los datos originales, no existe diferencia entre la que se obtendría con la correspondiente técnica DR aplicada directamente a los datos sin ninguna transformación. Esta proyección puede considerarse como referencia inicial en la cual comparar las nuevas proyecciones que se generen aplicando el método propuesto.

Ejemplo de grupos de datos gaussianos

En este ejemplo, se utiliza un conjunto de datos sintético para la validación del método en un escenario controlado. Los datos constan de 5 grupos gaussianos aleatorios con un total de 500 muestras de 10 dimensiones, estos datos provienen del trabajo [168]. Las pro-

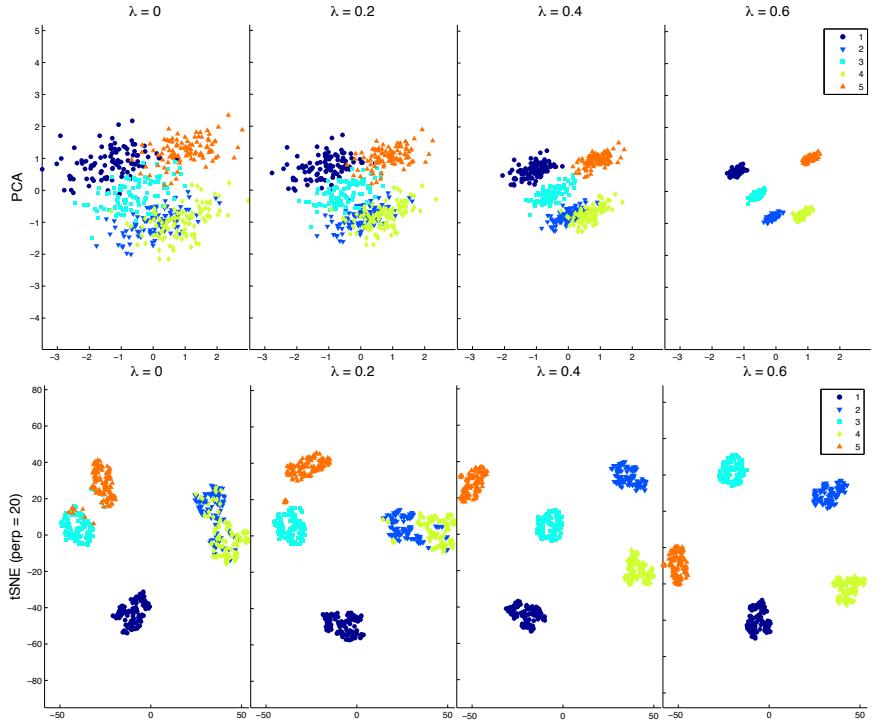


Figura 5.6.: Proyecciones de *3D clusters* con extensión ponderada, para varios valores de λ, utilizando información de grupos, calculadas mediante PCA (arriba) y t-SNE (abajo) con una perplejidad de 20.

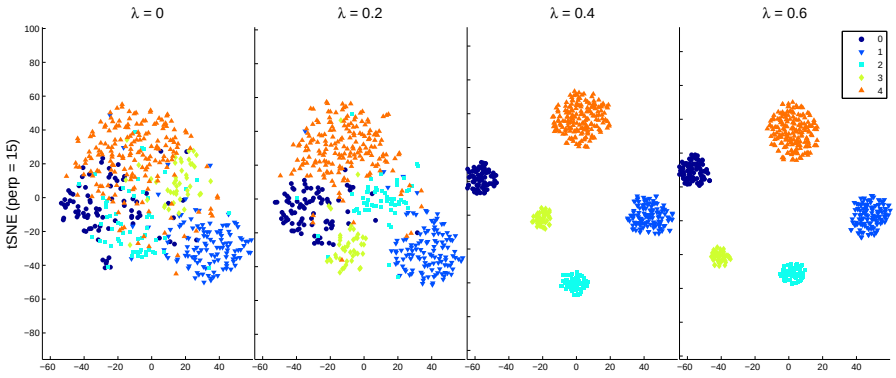


Figura 5.7.: Proyecciones t-SNE ($perp = 15$) para *synthetic-gaussian* aplicando la extensión ponderada, para varios valores de λ, utilizando las etiquetas de grupo.

yecciones se calculan por medio de la técnica t-SNE con un valor de perplejidad de 15.

En la Fig. 5.7 se muestran las proyecciones resultantes de aplicar la extensión usando la información categórica de los grupos. Las 5 clases de los grupos se representan codificadas por los diferentes colores y formas de los puntos. En primer lugar, se puede ver que la proyección de los datos originales ($\lambda = 0$) muestra los límites de los grupos mezclados entre sí. Tomando esta proyección como referencia, se puede ver que las proyecciones con la transformación aplicada (con valores de λ entre 0.2 y 0.4) proporcionan una separación más clara de los grupos, los cuales pueden ser identificados rápidamente. Incluso sin una codificación del color, no sería difícil distinguir los patrones mostrados por las nuevas proyecciones.

Obviamente, la información de grupos que se incorpora debe tener relación con los datos y ser validada por el analista para una correcta interpretación de las proyecciones que resulten.

5.4.2 *Caso de serie temporal: consumos eléctricos en un edificio universitario*

En este segundo caso, se aplica el método integrando varios grupos que definen distintos tipos de información de los datos, teniendo en cuenta su contenido temporal.

Los datos que se utilizan en este caso de estudio describen el consumo eléctrico de un edificio de la Universidad de Oviedo. Recogen el consumo de potencia activa en un edificio universitario a lo largo de un año, en concreto desde el 2 de junio de 2012 hasta el 1 de junio de 2013, con un periodo de muestreo de 15 minutos.

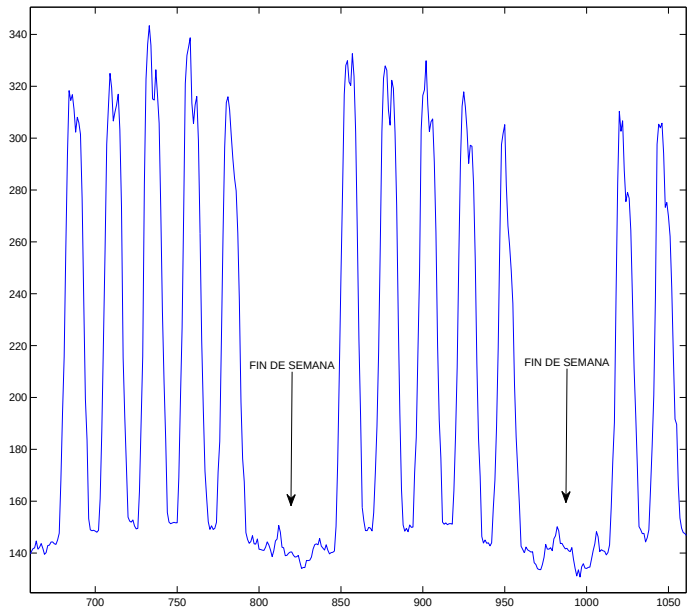


Figura 5.8.: Representación temporal de una parte de los datos *eCons*.

Los datos se han procesado para que cada muestra represente un día, el cual está descrito por el consumo medio de potencia activa por cada hora, y se han denominado como *eCons* para referirnos directamente a ellos. Constan de 338 muestras (los días con algún dato no válido fueron eliminados) y 24 dimensiones (correspondientes a cada hora del día). La tarea principal es la identificación de los distintos tipos de patrones en los consumos diarios de ese edificio universitario. En la Fig. 5.8 se puede observar la representación temporal de varias muestras de potencia activa promediadas por horas, correspondientes a unas dos semanas de consumo. Se pueden apreciar claramente las diferencias de consumos que se producen durante los fines de semana, así como los perfiles de consumo que poseen los días lectivos de la semana. En la Fig. 5.9 se representa todo el conjunto de datos de manera temporal, en la que se pueden diferenciar los consumos durante periodos festivos tradicionales como pueden ser navidades o el mes de agosto.

Este tipo de análisis ha sido estudiado previamente, por ejemplo en [194], donde se utiliza una vista de calendario combinada con información de grupos para una eficiente exploración de datos temporales. Mientras que este trabajo presenta una excelente forma

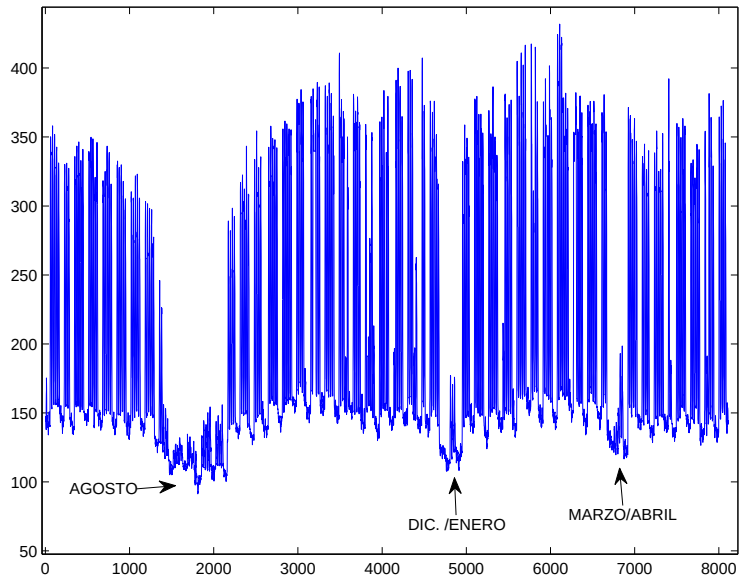


Figura 5.9.: Representación temporal del conjunto de datos *eCons*.

para el análisis univariable de este tipo de datos, el enfoque propuesto se basa en la proyección de datos multidimensionales con la flexibilidad de ajustar distintas temporalidades y/o elegir distintos tipos de información de clases. Recientemente, en [51] se presenta *Morphing Projections*, una interfaz interactiva que permite transiciones animadas entre distintos mapas 2D. Cada mapa conlleva un tipo de información relativa a los datos mediante su codificación espacial, como por ejemplo información temporal en forma de reloj. Además, combinando varios mecanismos de interacción, la herramienta permite al usuario seleccionar puntos de su interés y seguirlos a través de los diferentes mapas y de sus posiciones intermedias.

Para la aplicación del método propuesto, en este experimento se consideran dos tipos de información temporal, que se integrarán en la proyección en forma de dos tipos de etiquetas de clases: las correspondientes al día de la semana (lunes, martes, ..., domingo) y al correspondiente mes (enero, febrero, ..., diciembre).

En primer lugar, se analizan los datos con las clases del día de la semana. Se calculan las proyecciones con el método t-SNE con un valor de perplejidad de 20, las cuales se representan en la parte superior de la Fig. 5.10. Obviamente, los colores y formas de los puntos de la proyección representan el día de la semana. En la pro-

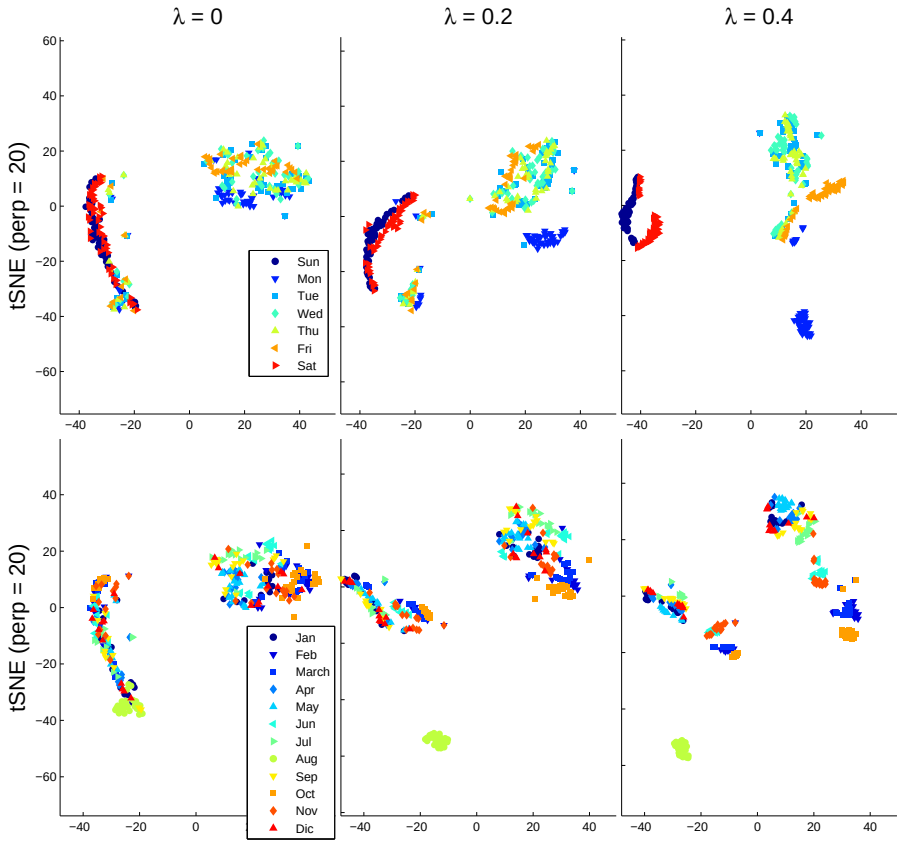


Figura 5.10.: Proyecciones t-SNE para *eCons* ($perp = 15$) (izq.) y con extensión ponderada de varios valores de λ utilizando las etiquetas del día de la semana (arriba) y de los meses (abajo).

yección con los datos originales (ver Fig. 5.10, superior izquierda) se diferencian dos grupos de consumo claramente, que se pueden interpretar como de alto y bajo nivel de consumo diario, conteniendo en su mayoría cada grupo días laborables y no laborables, respectivamente. En las nuevas proyecciones, aplicada la extensión con valor $\lambda = 0{,}2$ (ver Fig. 5.10, arriba), se distinguen sin embargo, divisiones en ambos grupos de la proyección. Por una parte, los lunes se diferencian en el grupo de alto consumo, demostrando un comportamiento diferente del resto de días. Además, el grupo de bajo consumo también se divide entre fines de semana y el resto de días, los cuales consisten principalmente en festivos que cayeron en días entre lunes y viernes.

Además, las posiciones relativas de los puntos de un grupo en concreto en la proyección inicial, se mantienen para ese mismo grupo en la nueva proyección, es decir, puntos de un día de la semana cercanos en la proyección inicial siguen estando cerca en la modificada. Dado que la estructura original dentro de cada clase se conserva en la modificación, se puede realizar una interpretación jerárquica de la proyección, eligiendo valores de λ adecuados para la extensión, en la que los grupos revelados en la proyección inicial se dividen en relación a la información de clases introducida.

Para el caso de los meses, se realiza un análisis análogo, cuyos resultados se muestran en la parte inferior de la Fig. 5.10. Los colores y formas muestran este tipo de información del mes al que corresponde cada día. Las proyecciones se calculan con t-SNE usando el mismo valor anterior de la perplejidad de 20. En la proyección original se muestran los mismos grupos anteriores de alto y bajo consumo, con todos los meses mezclados en los dos grupos de consumo diario. Sin embargo, con la extensión aplicada ($\lambda = 0{,}4$) se ve que el mes de agosto se separa claramente del resto del grupo, mostrando un comportamiento característico dentro del grupo de bajo consumo. También los meses de febrero, marzo y octubre muestran consumos diferenciados del resto de meses en el grupo de consumo alto.

Conviene apuntar que partiendo de una misma proyección, la aplicación del método la modifica de formas distintas dependiendo de la información que se introduce por parte del usuario, obteniéndose una modificación de la proyección de manera jerárquica. En este caso se han considerado dos criterios distintos, relacionados con información temporal (día de la semana y mes), los cuales dividen los dos principales grupos de consumo de la proyección original de dos maneras diferentes.

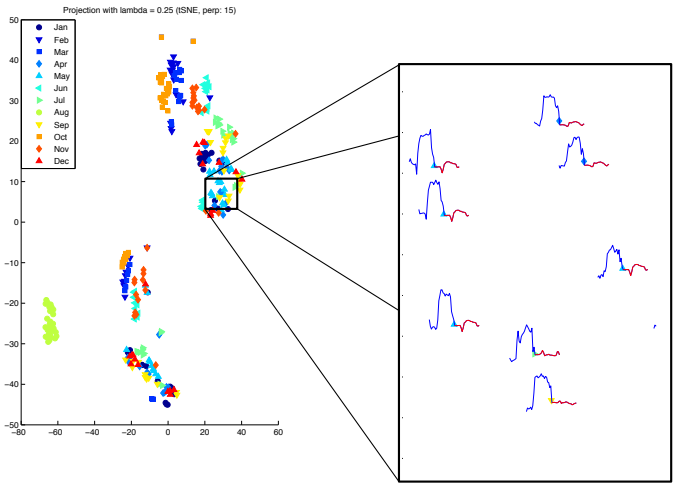


Figura 5.11.: Proyección extendida por meses y representación aumentada con valores como *sparkline* de cada muestra, original (azul) y extendida (rojo).

En la Fig. 5.11 se muestra parte de una de las proyecciones resultantes de los mismos datos, modificada con la extensión de información de grupos de los meses, con las dimensiones representadas mediante un *sparkline* en cada muestra, de los datos originales (en azul) y de la parte extendida (en rojo). De esta forma, se puede realizar una rápida comparación entre las similitudes de los puntos proyectados. Los puntos de la proyección están distribuidos en el mapa modificado por las similitudes dadas por las variables originales (trazo azul), que definen la información "intraclase" de los datos. Puntos dentro de una misma clase tienen la misma parte extendida (trazo rojo), la cual define la información "interclase" de los grupos. La organización de los puntos en la proyección varía dependiendo de la ponderación entre estas dos partes (en azul y rojo), definida por el parámetro λ, el cual para su valor extremo de 1 correspondería a la proyección de los centroides de cada clase.

5.4.3 *Extensión de una variable seleccionada*

Puede haber casos en los que un reducido grupo de variables (o incluso una sola) describen mejor la separabilidad de los grupos o clases que se quieren analizar que el conjunto total de las variables.

En estos casos, sería interesante aplicar la extensión para ese menor grupo de variables.

En este experimento se estudia el efecto de extender una variable concreta seleccionada de los datos. El conjunto de datos utilizado para ello es *yeast* [20], el cual es un ejemplo de datos reales muy utilizado para el análisis de algoritmos, cuya tarea principal es predecir la localización de una proteína. Dado el conjunto de datos y las etiquetas de clases, el analista puede estudiar la distribución de los datos y detectar las variables más relevantes para el estudio que está llevando a cabo. Esto puede hacerse por medio de una visualización multidimensional, como pueden ser las coordenadas paralelas [94].

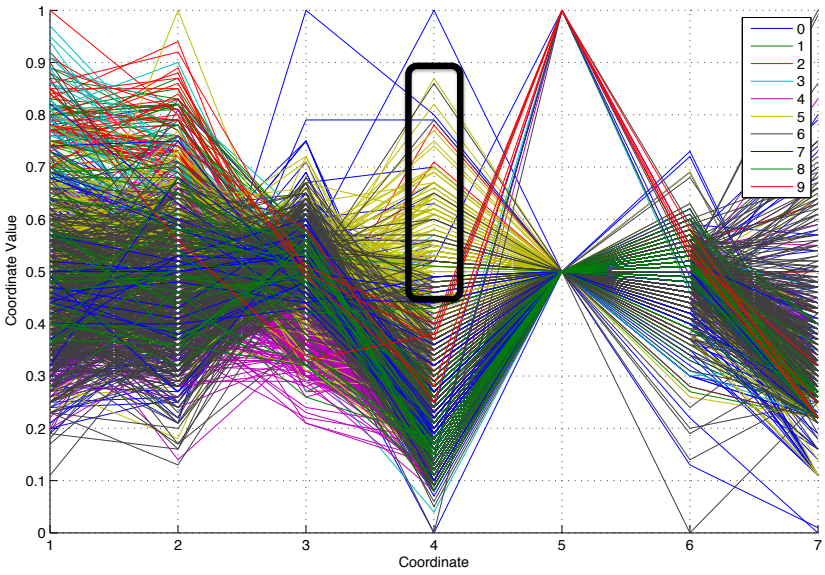


Figura 5.12.: Visualización de coordenadas paralelas del conjunto de datos *yeast*.

En la Fig. 5.12 se representa la vista de coordenadas paralelas para el conjunto de datos estudiado. Cada eje vertical representa una dimensión de los datos, y cada línea representa una muestra con los valores que va tomando en cada dimensión. Los colores representan la información de las etiquetas de clase. En esta representación se puede apreciar una distribución de las clases mezclada para las primeras y las últimas dimensiones del conjunto de datos, pero también en qué dimensiones existen diferencias entre clases. Supongamos que queremos analizar la clase 5 (amarillo), en la Fig. 5.12 se observa que posee valores separados para esa clase en la

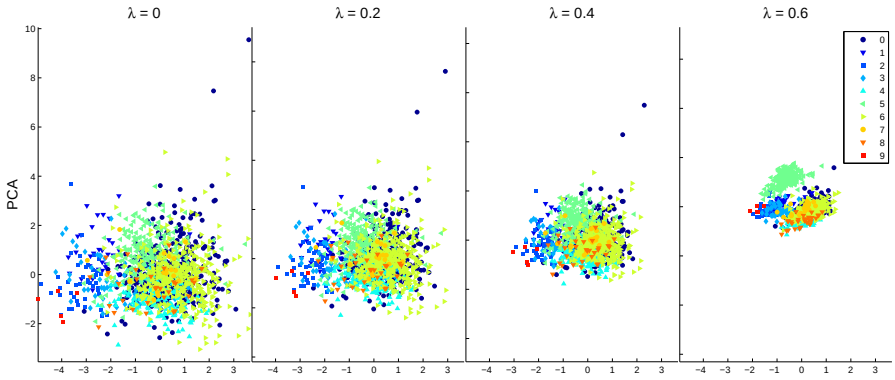


Figura 5.13.: Proyecciones de *yeast* con una extensión para varios valores de λ aplicada en la dimensión 4 utilizando la técnica PCA.

dimensión 4 (resaltado en un cuadro negro), elegiremos esta dimensión para aplicar el método, es decir, la extensión de las medias de cada clase en la dimensión 4. De manera análoga, si se quisiera estudiar por ejemplo la clase 4, se extendería la dimensión 3 dado que es la dimensión con los valores más separados para esa clase.

En la Fig. 5.13 se muestran las proyecciones resultantes de aplicar la extensión ponderada al conjunto de datos, con la información de grupos, utilizando la técnica PCA, con los colores y formas de los puntos representando cada una de las clases. En la proyección inicial de los datos originales ($\lambda = 0$) se observa una nube de puntos con todas las clases mezcladas donde resulta difícil la identificación de algún patrón en los datos. Sin embargo, en las proyecciones utilizando la misma técnica DR con la extensión ponderada (para $\lambda = 0{,}6$) aplicada sólo sobre la dimensión 4, se puede ver que en la proyección se produce una separación de los puntos de la clase 5 (en verde) con respecto al resto de las clases, como se había identificado en la selección de la variable previamente.

Para conjuntos de datos con un gran número de dimensiones (por ejemplo más de 50-100 variables), una selección manual en la vista de coordenadas paralelas resultaría tedioso. Para este tipo de casos, sería recomendable un proceso previo de selección de las variables que mejor describan la separación de las clases a analizar, lo cual puede llevarse a cabo por medio de diversos criterios automáticos de selección de características [22, 78, 46, 125]. De esta forma, se extrae un reducido grupo de variables que el usuario puede manejar fácilmente para realizar la selección final.

La elección de un valor adecuado del parámetro λ por parte del usuario no sólo se puede decidir mediante una interpretación

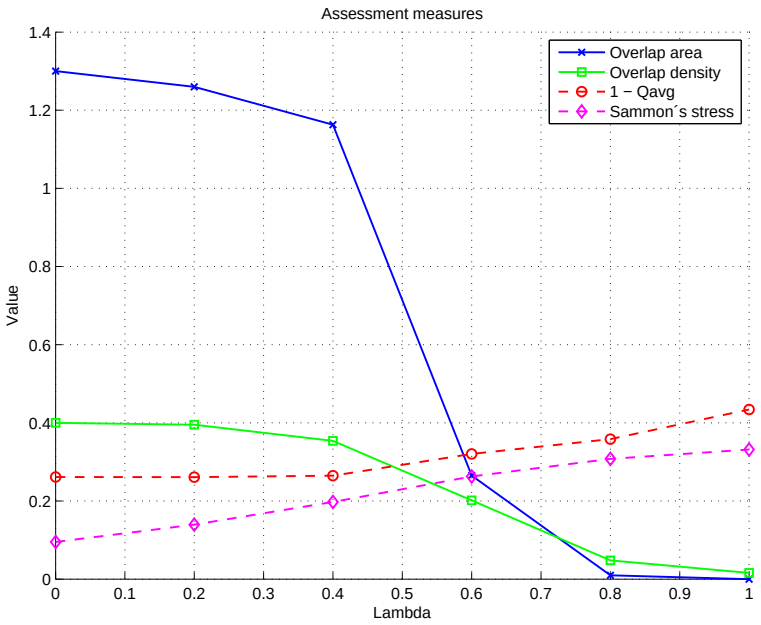


Figura 5.14.: Medidas de calidad de proyecciones PCA de *yeast* para valores de λ con una extensión de la dimensión 4.

visual, sino que también podría apoyarse en una evaluación cuantitativa de las proyecciones que se obtienen. La correcta selección de este parámetro sería aquella que proporciona una mejora visual y que preserve además la estructura original. En la Fig. 5.14 se representan unas curvas de varias medidas de calidad para diferentes valores de λ: las dos medidas de solapamiento [162], el *stress* de Sammon [160], y la media de la medida *k-ary* (Q_{avg}) [122], la cual tiene un valor de 1 para una perfecta proyección, por esto se utiliza $1 - Q_{avg}$ para que las tendencias de todas líneas tengan el mismo significado, es decir, el valor más bajo representa una mejora de la medida. En dicha figura se puede ver la mejora de las medidas visuales con un leve empeoramiento de la preservación de la estructura original de los datos. En este caso las curvas confirman la selección de $\lambda = 0{,}6$, mejorando visualmente la proyección de manera notable para ese valor sin una gran variación con respecto a la estructura original.

5.4.4 *Agrupamiento natural de los datos*

En este caso se presenta la posibilidad de aplicar el método propuesto utilizando el resultado de un análisis previo de la información de grupos (*clusters*) de los datos en lugar de las etiquetas de clases. Puede ser útil para casos en los que dichas etiquetas de clase no se encuentren disponibles. Para ello, se utiliza el conjunto de datos *yeast* [20] descrito previamente. Se aplica el algoritmo *K-means* fijando su parámetro K con el mismo valor que el número de clases (10) de sus etiquetas originales. Dicho algoritmo proporciona un resultado de la agrupación natural del conjunto de datos dividiendo el conjunto de datos en 10 grupos diferentes, esta información será utilizada en la aplicación del método.

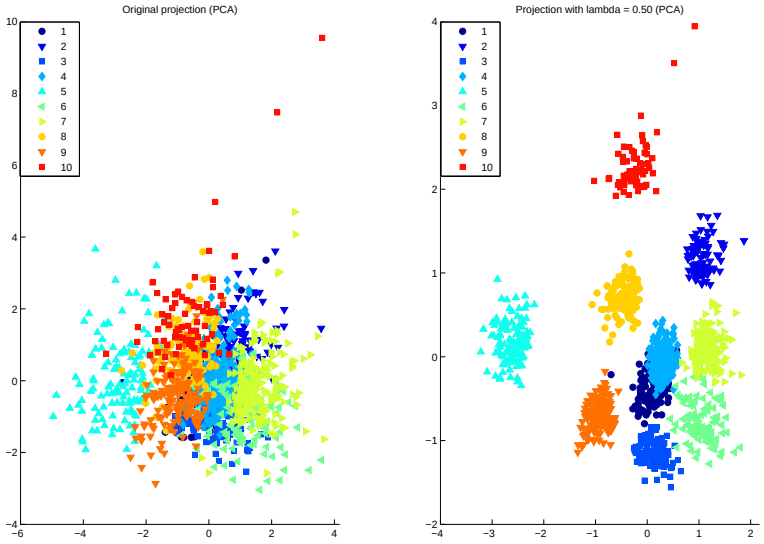


Figura 5.15.: Proyección con PCA de *yeast* (izq.) y con una extensión ($\lambda = 0{,}5$) (dcha.) utilizando el resultado del análisis de grupos.

El método se aplica de manera análoga al resto de experimentos, empleando la extensión ponderada del espacio de características, pero en este caso, basándose en los grupos obtenidos en el análisis descrito.

En la Fig. 5.15 (izquierda) se representa la proyección de los datos obtenida al aplicar PCA, siendo los colores y formas de los puntos, cada uno de los grupos resultantes del proceso previo. Se puede ver que la reducción realizada por la técnica PCA no es capaz de discriminar completamente la agrupación obtenida por

el método *K-means*, generando un mapa en el que todos los grupos se encuentran superpuestos en una nube de puntos.

En la parte derecha de la misma figura (Fig. 5.15) se puede ver la proyección que se obtiene mediante la misma técnica PCA, con un valor de $\lambda = 0{,}5$ para la extensión. En esta nueva proyección, sin embargo, se aprecian más claramente los grupos encontrados por el análisis previo facilitando su identificación visual. Esto permite un reconocimiento más rápido de los grupos estudiados dando además una idea de las similitudes existentes entre ellos, lo cual sería difícil de diferenciar en el mapa original.

5.4.5 *Aplicación mediante técnicas DR supervisadas*

En este último experimento de los expuestos para la comprobación del método propuesto, se aplica éste utilizando técnicas DR supervisadas, es decir, la información de las etiquetas de clases es utilizada también por la propia técnica de proyección. Las técnicas elegidas para llevar a cabo dicho experimento son NCA [75] y MCML [74] aplicadas en varios conjuntos de datos, utilizados previamente y en otro conjunto nuevo. Se utiliza la misma información de clases tanto para aplicar la extensión como en la propia técnica para calcular las proyecciones. A continuación se describen los resultados obtenidos para cada uno de los conjuntos de datos utilizados en el experimento. En primer lugar, se aplica el método de manera análoga a los anteriores experimentos, pero con las técnicas supervisadas, en un escenario controlado con el uso del mismo conjunto de datos sintéticos que se ha utilizado previamente en el apartado 5.4.1. Después se aplica a un nuevo conjunto de datos (*hiv*) y finalmente al conjunto *yeast*.

3D CLUSTERS El método se aplica al ejemplo sintético de *3D clusters*, las proyecciones resultantes para varios valores de λ se representan en la Fig. 5.16 para las técnicas NCA (en la primera fila desde arriba) y MCML (en la segunda). El parámetro de regularización de la técnica NCA se fija igual a 0. Se puede observar que la aplicación del método mejora y enfatiza la separación de las clases utilizando ambas técnicas.

SYNTHETIC-GAUSSIAN Las proyecciones resultantes, que se representan en la tercera fila de la Fig. 5.16, se refieren a las calculadas con la técnica MCML. Se aprecia una proyección original, con todos los grupos de las clases mezclados entre sí, pese a que la técnica DR supervisada que se emplea utiliza esta información de

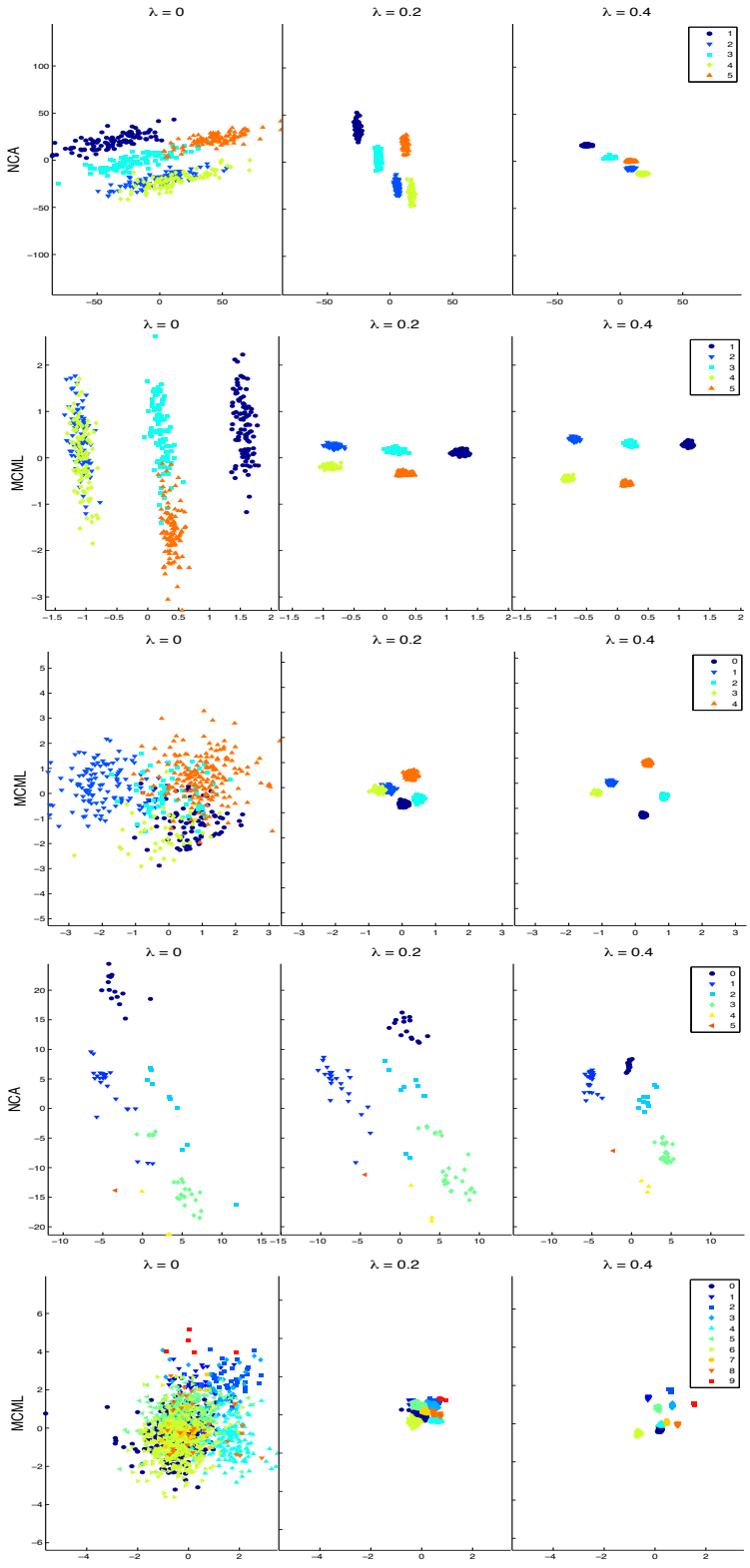


Figura 5.16.: Proyecciones para varios λ utilizando las clases de los datos. Para *3D Clusters* usando NCA (primera fila desde arriba) y MCML (segunda); *synthetic-gaussian* con MCML (tercera); *hiv* con NCA (cuarta); y *yeast* utilizando MCML (última fila).

grupos. Aplicando la extensión ponderada (en este caso $\lambda = 0{,}4$) la calidad visual de los grupos mejora sustancialmente, donde son claramente diferenciados.

HIV El conjunto de datos *hiv* ha sido utilizado previamente en el trabajo [174] y describe propiedades socio-económicas de países, que se clasifican en grupos de riesgo de VIH. Los datos contienen 159 atributos y 78 muestras que pertenecen a 6 clases distintas. Se ha aplicado el método propuesto de manera análoga proyectando con la técnica DR supervisada NCA. Las proyecciones obtenidas se representan en la cuarta fila de la Fig. 5.16, aunque los grupos están bien separados en la proyección original, la extensión aplicada con $\lambda = 0{,}2$ enfatiza la separación entre grupos.

YEAST El conjunto *yeast* ha sido utilizado previamente en el experimento del apartado 5.4.3. Aquí se aplica la extensión para todas las variables como en la mayoría de los experimentos, pero utilizando la técnica MCML. En la parte inferior de la Fig. 5.16 se muestran las proyecciones para estos datos calculadas con dicha técnica. En la proyección original de los datos las clases se muestran mezcladas entre sí, con la dificultad que conlleva para la interpretación de cualquier patrón en los datos. La modificación producida al aplicar el método, con un peso de $\lambda = 0{,}4$, proporciona una visualización más clara de la información de grupos en los datos, los cuales incluso la propia técnica de proyección tiene en cuenta pero no es capaz de discriminar correctamente.

5.5 EVALUACIÓN DE LA CALIDAD DE LAS PROYECCIONES

Las modificaciones que se producen en una proyección por la aplicación del método aquí propuesto se evalúan, en un primer momento, visualmente por parte del usuario. Para una valoración más detallada, estas proyecciones también se pueden evaluar por medio de medidas analíticas que describan su calidad matemáticamente. Se han elegido cuatro medidas para esta tarea, el *stress* de Sammon [160], la media de la medida *k-ary* (Q_{avg}) [120, 121] para la evaluación de la preservación estructural y las medidas de solapamiento descritas en [162] como mediciones de evaluación visuales.

Las medidas se calculan en las proyecciones obtenidas con la aplicación del método para varios valores del parámetro λ. Se representan gráficamente de la misma forma que en la Fig. 5.14, de

manera que valores bajos de las curvas representan mejoras en la medida de la calidad de la proyección correspondiente. Las figuras de 5.17 a 5.20 muestran estas curvas con los resultados de las medidas calculadas para los conjuntos de datos *synthetic-gaussian*, *hiv*, *eCons* (meses), y *yeast*, respectivamente. Las medidas para las técnicas PCA y *t*-SNE se muestran en la parte superior de las figuras y para NCA y MCML en la parte inferior de las mismas. Aquellos valores fuera de rango fueron escalados para facilitar la comparación entre ellas.

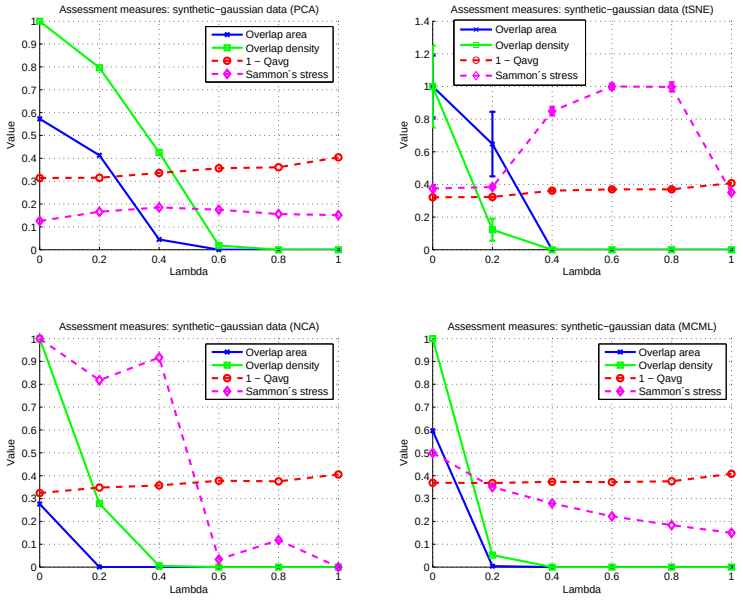


Figura 5.17.: Medidas de calidad de las proyecciones calculadas mediante las técnicas PCA, *t*-SNE (arriba), NCA, y MCML (abajo) del conjunto *synthetic-gaussian*. Valores bajos indican mejoras de calidad.

Los resultados que se aprecian en las figuras muestran una mejora de las medidas visuales de las proyecciones resultantes de aplicar el método propuesto. Se produce una reducción de los puntos superpuestos, indicada por la reducción en la medida de la densidad de solapamiento. También los grupos de las clases se encuentran menos superpuestos en las proyecciones modificadas, indicado por la reducción del área de solapamiento. La estructura original se conserva menos en las nuevas proyecciones, este hecho se muestra sobre todo en la medida global de distancias del *stress* de Sammon. Sin embargo, en muchos casos la medida de la media de *k-ary*, la

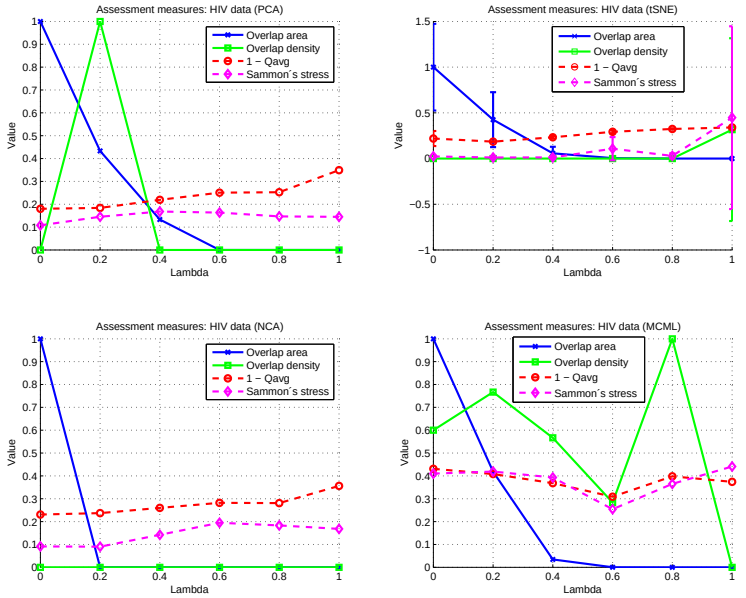


Figura 5.18.: Medidas de calidad de las proyecciones calculadas mediante las técnicas PCA, *t*-SNE (arriba), NCA, y MCML (abajo) del conjunto *hiv*. Valores bajos indican mejoras de calidad.

cual representa una medida estructural más fiable en términos de preservación local y global, permanece sin demasiadas variaciones.

La evaluación de la calidad de las proyecciones con medidas numéricas ayuda al analista a un mejor entendimiento de la distorsión que está aplicando con respecto a los datos originales. También sirve para confirmar la mejora visual que está produciendo la transformación. Con la ayuda del parámetro λ se controlan las modificaciones que se producen en la proyección gradualmente y, utilizando estas gráficas, se puede elegir una solución de compromiso entre la mejora visual y la preservación de los datos originales, para alcanzar una proyección final satisfactoria.

Además de los gráficos descritos, se pueden utilizar otros enfoques para evaluar la calidad de las proyecciones. Por ejemplo, algún método que proporcione al usuario una información visual acerca de la preservación de la estructura en la proyección. En este sentido, en [139] se propone una evaluación por puntos utilizando criterios basados en ranking que permite resaltar regiones erróneas en la visualización. Utilizando este trabajo, se puede calcular un error medio por cada punto de la proyección y codificarlo con el color. Mientras que otras medidas evalúan la proyección en su con-

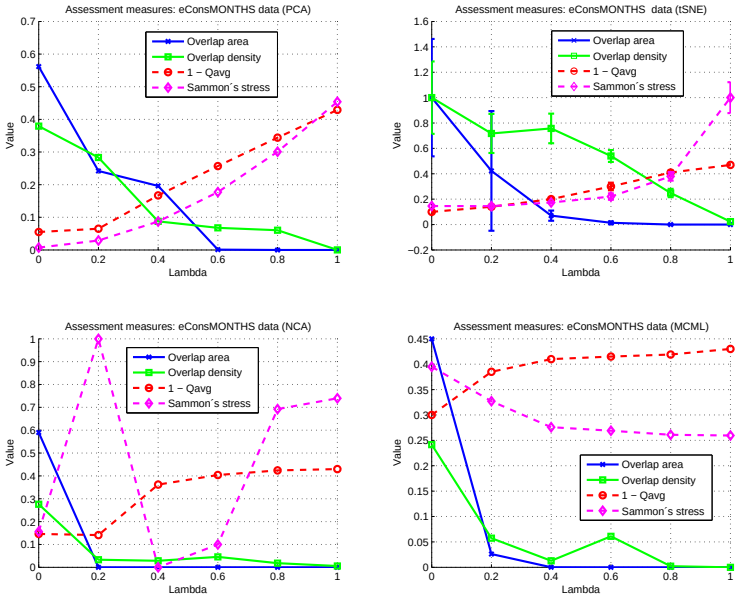


Figura 5.19.: Medidas de calidad de las proyecciones calculadas mediante las técnicas PCA, t-SNE (arriba), NCA, y MCML (abajo) del conjunto *eCons* meses. Valores bajos indican mejoras de calidad.

junto, la utilizada refleja la calidad de manera general (global y localmente) punto a punto. En la Fig. 5.21 se realiza esta visualización de la calidad para los datos eléctricos utilizando los días de la semana como información de grupos (*eCons (weekday)*), en donde se puede ver la calidad de la preservación de la estructura por medio del color de cada punto en la proyección. Con la aplicación del método se aprecia la preservación en la proyección modificada, cuyos puntos revelan la distorsión producida con respecto a los datos originales. Esta visualización ayuda al usuario a ser consciente de los errores en cuanto a la conservación de la estructura original, de forma que permite determinar un valor final del parámetro λ fácilmente, con una evaluación directa de la proyección.

5.6 DISCUSIÓN

Dada una proyección inicial, el método propuesto permite al usuario modificar la proyección con mejores calidades visuales, integrando nueva información de grupos en el proceso de reducción de la dimensión. Se supone que esta información es relevante para el usuario, proporcionando conexión con conocimiento previo

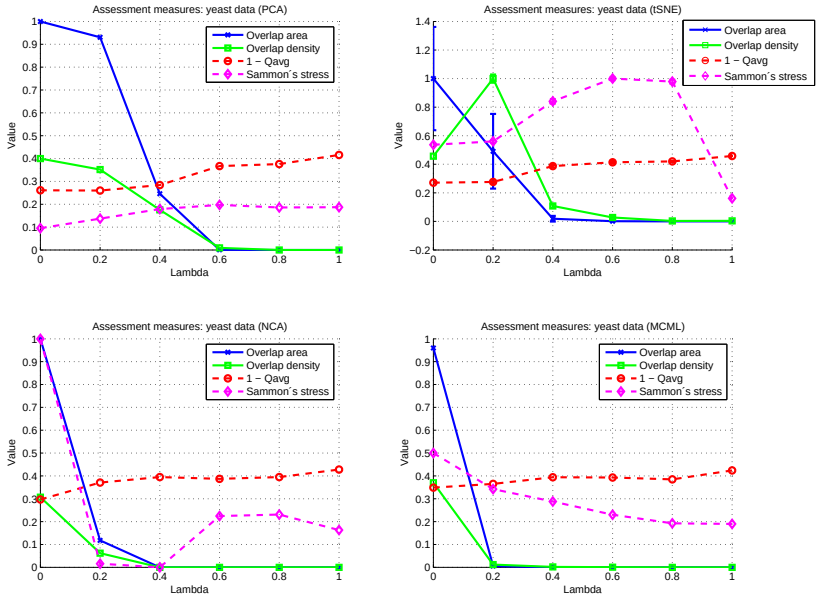


Figura 5.20.: Medidas de calidad de las proyecciones calculadas mediante las técnicas PCA, *t*-SNE (arriba), NCA, y MCML (abajo) del conjunto *yeast*. Valores bajos indican mejoras de calidad.

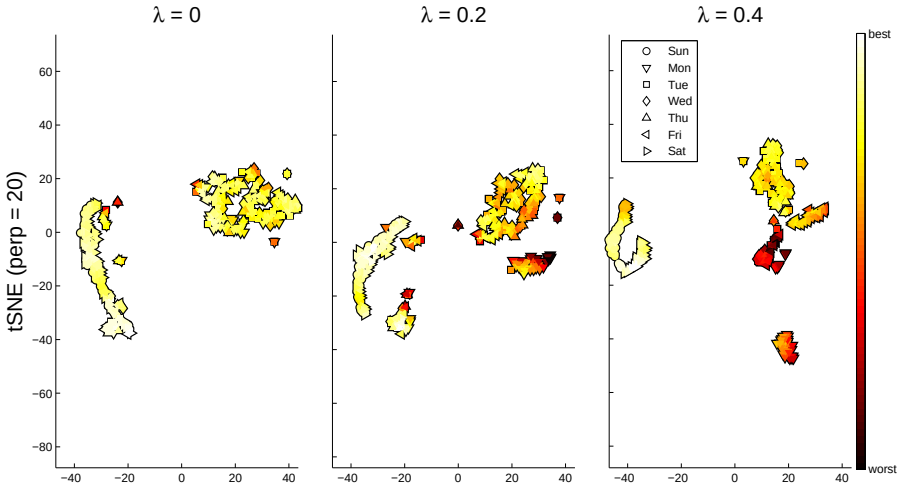


Figura 5.21.: Visualización de la calidad por cada punto para la evaluación del ejemplo *eCons (weekday)* para varios valores de λ.

relacionado con los datos analizados. El método se puede aplicar cuando la información de grupos no se encuentra revelada completamente en el cálculo estándar de una proyección de los datos, debido al ruido o porque contengan información irrelevante. Otra ventaja del método es la preservación de la estructura local. Puesto que las distancias entre puntos dentro del mismo grupo no se alteran en la transformación, la nuevas proyecciones conservan la estructura local de cada grupo en los datos originales.

Desde el punto de vista de la visualización, se podría decir que es posible visualizar la información de grupos utilizando la forma y el color. Sin embargo, esta codificación no soluciona el problema del solapamiento de puntos, lo que da lugar a una proyección desordenada. Cuando los puntos están solapados en un plano visual, la legibilidad de la proyección no se mejora. El enfoque propuesto ayuda a reducir el solapamiento de los puntos, y proporciona un canal visual más eficiente para evaluar distancias relativas entre objetos y clases. Además, utilizando la posición para separar las clases hace posible aplicar otros mecanismos de interacción como selección de puntos, información de metadatos por cada muestra, etc.

El proceso de transformación se controla mediante el parámetro λ. Cuando es 0, la proyección corresponde a los datos originales, cuando se aumenta, el método incrementa gradualmente la influencia de la información de grupos. Cuando alcanza el valor de 1, solo la información de las clases se utiliza para calcular la proyección, lo que hace colapsar los puntos a la proyección de los correspondientes centroides de cada clase. El analista puede comenzar con una proyección inicial, y aumentar λ para alcanzar una mejor separación de grupos, lo cual se puede implementar fácilmente por ejemplo mediante una barra *slider* en una interfaz interactiva.

El papel de la interacción es importante. Los cambios de λ permiten modificar la proyección de manera reversible, lo que permite mantener en la mente la proyección de referencia inicial. Las variaciones pequeñas de la proyección permiten un seguimiento continuo de los puntos, lo cual mejora su percepción gráfica por medio de la animación entre las vistas de las proyecciones. Se puede examinar la estructura original de los datos, la proyección de los centroides, y las vistas intermedias, que permiten revelar patrones no vistos con una herramienta estándar. La aplicación descrita no sólo sirve para alcanzar una separación de grupos, sino para entender qué parte del solapamiento de clases se produce en la reducción de la dimensión y cuál es característica real de los datos. Existen métodos para visualizar la calidad [7, 139, 189], que pueden ser

utilizados en combinación con los medios interactivos, una visualización de la calidad dinámica ayudaría al analista a evaluar la distorsión introducida y valorar el balance entre la distorsión y la mejora del solapamiento.

5.7 CONCLUSIONES

El método interactivo de extensión propuesto ayuda en el análisis de las proyecciones de datos multidimensionales, calculadas por medio de técnicas de reducción de la dimensión tradicionales. Dicho enfoque permite al usuario introducir conocimiento previo relativo a las clases en las que se agrupan los datos a estudiar. Esto se realiza modificando la proyección original gradualmente, lo cual proporciona al analista más control sobre el proceso exploratorio para incorporar información que puede estar oculta o no estar presente en los datos de partida, pero que es útil para una mejor interpretación de la proyección resultante. El método se lleva a cabo mediante la extensión de características usando etiquetas de clases de los datos. La integración de esta información es controlada por medio de la ponderación de las dos partes de la matriz de datos (original y extendida) con el parámetro λ. Este parámetro aporta la interacción para que el usuario controle en todo momento la transformación que se produce en la proyección. Además, varias medidas de calidad cuantitativas sirven para ayudar al usuario a decidir la adecuada proyección final, aportándole información con respecto a las mejoras visuales que se produzcan y a la preservación de la estructura original de los datos.

La comprobación del método se ha realizado por medio de distintos experimentos, cubriendo varios tipos de estudio: casos sintéticos y reales, el análisis de patrones temporales, la extensión de una determinada variable, usando tanto las etiquetas de clases de los datos como información de grupos obtenida de un análisis previo y finalmente el estudio con técnicas de proyección supervisadas. Las proyecciones resultantes de estos experimentos se han evaluado no sólo de manera visual, sino con la ayuda de medidas cuantitativas. Los resultados experimentales indican una mejora visual de las proyecciones con una preservación local dentro de cada clase.

Estos resultados validan la integración de información útil en la proyección de manera controlada por el usuario, aportando eficiencia en el proceso de análisis para el reconocimiento de patrones, la rápida identificación de los grupos y un mejor entendimiento de datos multidimensionales.

6

CONCLUSIONES Y TRABAJO FUTURO

En este capítulo se recogen las conclusiones de las investigaciones desarrolladas y se sugieren trabajos que las podrían complementar dentro de las líneas futuras de investigación que se encuentran abiertas. Además, se detallan las principales aportaciones originales derivadas de los estudios realizados y también las principales publicaciones y ponencias en congresos que dieron lugar.

6.1 CONCLUSIONES FINALES

El contexto de esta tesis se enmarca dentro del ámbito del análisis de datos, el cual emplea fundamentos de muchas áreas como la estadística, inteligencia artificial, o la visualización para extraer nuevo conocimiento útil. La gran facilidad para la adquisición de datos y la aparición de nuevos y mejores algoritmos de aprendizaje automático ha ayudado al desarrollo de este campo de la ciencia. Esto facilita la investigación de problemas difíciles de abordar en el análisis de datos procedentes de procesos complejos. Una adecuada codificación visual de la información amplifica nuestra cognición facilitando la interpretación de los resultados. Además, brinda un potencial para explotar conocimiento previo a partir de información de contexto y su combinación con mecanismos de interacción permite que el usuario pueda mejorar la exploración de los datos. Varios fundamentos relacionados con la visualización y los mecanismos de interacción se han tratado en el capítulo 2.

En los trabajos realizados, la investigación se orientó hacia las técnicas de reducción de la dimensión (DR), las cuales pueden estimar la estructura latente y obtener una proyección visual de los datos. El estudio de las numerosas técnicas existentes no sólo llevó a explorar técnicas clásicas como PCA como primera opción para proyectar datos, sino también el uso de otras no lineales, como el algoritmo *t*-SNE, el cual es adecuado para estimar la estructura latente de datos multidimensionales en dos o tres dimensiones, que se pueden visualizar en un *scatterplot*, de forma que objetos similares se representan por puntos cercanos y viceversa, siguiendo el denominado principio de espacialización. Una descripción de varias de estas técnicas se puede ver en el capítulo 3.

Mediante varias estrategias de extracción de información explicadas en el capítulo 4, como el análisis de bandas en frecuencia, y la aplicación de técnicas de proyección se caracterizó un defecto llamado *chatter*, con datos reales de un tren de laminación en frío [149, 150]. La visualización de los estados de vibración del proceso, en un mapa que describe sus comportamientos dinámicos, facilita la identificación del fallo y por tanto la supervisión de este tipo de condiciones erróneas de funcionamiento.

En el capítulo 4 también se trató de abordar alguna de las limitaciones que poseen estos algoritmos, como el número de puntos que pueden manejar a la hora de calcular la proyección. Para ello, se calculan unos prototipos que representan el conjunto total de datos, con los cuales se calcula una proyección. Con esto, mediante una versión de redes de base radial (GRNN) se puede estimar la proyección total de los puntos. Este procedimiento permite generar una proyección aproximada de conjuntos de datos prácticamente sin límite de tamaño. Se aplicó a datos eléctricos, recogidos durante un año, obteniéndose proyecciones en las que se pueden visualizar rápidamente patrones de consumo asociados también al precio de la energía correspondiente [148].

Otra de las limitaciones se encuentra a la hora de proyectar nuevos puntos en una proyección dada, para lo que son necesarias técnicas denominadas *out-of-sample*. Se encontró una solución inicial de este tipo para la técnica t-SNE, cuyo principal inconveniente fue su coste computacional, puesto que requería el cálculo por cada punto nuevo. Posteriormente, se utilizó para este propósito un enfoque similar al explicado anteriormente, con la aplicación de redes de base radial, las cuales mejoran su rendimiento [150].

También se exploraron estrategias de transformación de los datos, mediante una técnica denominada extensión de características, en la que, incorporando información de clases contenida en las etiquetas de los datos, se producen variaciones en la proyección resultante, que facilitan su interpretación [151]. Esta información contiene conocimiento previo que puede estar oculto en los datos, por ejemplo, debido al ruido, o al uso de variables irrelevantes o, simplemente, no descrito en las características de los datos, que se consigue hacer emerger en la visualización aplicando la técnica descrita en el capítulo 5.

El estudio de la combinación de técnicas de reducción de la dimensión con estrategias de transformación de datos permite al usuario modificar la proyección de forma interactiva, facilitando la visualización de los datos proyectados mediante la separación de los grupos introducidos [152]. Este proceso es reversible y ayuda a

mantener en la mente la proyección de referencia inicial. Por tanto, el usuario puede examinar la estructura original de los datos, enriquecida con información de las clases, mediante un seguimiento visual de los puntos en las vistas. Esta modificación de la proyección original, en la cual grupos de puntos que se encuentran situados por similitud se separan generando otra vista nueva, facilita la interpretación jerárquica de los datos.

Esto se ha estudiado en detalle en el capítulo 5 a través de varios experimentos utilizando distintos conjuntos de datos, sintéticos y procedentes de casos reales; y en diferentes situaciones, como por ejemplo analizando datos temporales, o utilizando técnicas de proyección supervisadas. Las proyecciones modificadas fueron, además, evaluadas mediante métricas descritas recientemente, para determinar la efectividad de cada transformación. Los resultados experimentales demuestran que mejoran la calidad visual de la proyección a través de una modificación controlada (por el usuario) de la estructura original. Esta evaluación de la calidad ayuda al analista a entender las modificaciones que está causando, y a decidir una proyección final de compromiso entre la distorsión con los datos originales y la mejora visual. Por tanto, este método no sólo puede utilizarse para lograr una separación de grupos sino también para entender el solapamiento existente en los datos.

En conclusión, podemos decir que la visualización de procesos mediante la proyección de sus datos y su exploración interactiva ayudan a analizar y tomar decisiones sobre problemas complejos. Además, las modificaciones resultantes de la exploración mejoran el entendimiento de estos datos aunque deben evaluarse para ayudar al usuario durante el análisis. Esta tesis abordó algunas cuestiones en este campo. Al mismo tiempo, se espera que las nuevas preguntas, planteadas aquí, motiven a otros investigadores a estudiarlas en el futuro.

6.2 RESUMEN DE LAS CONTRIBUCIONES

Se presentan dos direcciones principales para el análisis visual de procesos complejos: (1) el uso de varios métodos para la extracción de información de un proceso mediante sus datos, como la creación de un modelo del proceso o el estudio en frecuencia de sus estados dinámicos que posteriormente se representa mediante una proyección visual, y (2) la propuesta de un nuevo método, basado en la modificación de la proyección visual de manera interactiva que facilita la exploración en las tareas de análisis.

Las principales aportaciones originales se pueden resumir de la manera siguiente:

- Se ha demostrado la viabilidad del uso de proyecciones visuales para el análisis de datos reales procedentes de procesos complejos, incluyendo el estudio del comportamiento dinámico para la identificación de un fallo en un proceso de laminación y el análisis de consumos eléctricos en edificios.

- Se ha sugerido un procedimiento aproximado para la estimación de una proyección visual con un gran volumen de muestras en los datos, presentado en el 5^{th} *International Conference on Physics and Control (PhysCon 2011)* [148].

- Se ha desarrollado un algoritmo para la proyección de nuevos puntos en una proyección calculada previamente utilizando la técnica t-SNE [149]. También, la utilización de redes neuronales como alternativas a las existentes *out-of-sample* [150].

- Se han aplicado métodos para la extracción de información sobre el comportamiento dinámico del proceso, como la creación de un modelo del proceso de laminación utilizando la red ELM, o el estudio de las variables más relevantes del proceso mediante un análisis de bandas de frecuencia o funciones de respuesta en frecuencia.

- Se ha caracterizado el defecto de *chatter* en un tren de laminación mediante un mapa visual que permite la identificación rápida del fallo durante el proceso. Estos trabajos se publicaron en 13^{th} *International Conference on Engineering Applications of Neural Networks (EANN 2012)* [149] y en la revista *Engineering Applications of Artificial Intelligence* [150].

- Se ha estudiado un sencillo pero efectivo método interactivo que permite al usuario introducir información de grupos en una proyección de datos. Esto proporciona un nuevo control sobre una proyección estática, que gradualmente se modifica revelando conocimiento acerca de la información introducida. Este trabajo se presentó en el *workshop VAMP* en el *Eurovis 2013* [151], y posteriormente se extendió para la revista *Neurocomputing* [152].

- En el método propuesto se han integrado medidas cuantitativas, que evalúan la calidad de la proyección, tanto visual-

mente como en la preservación de la estructura, para proporcionar información adicional que ayude a la interpretación por parte del usuario.

- Basándose en la técnica Morphing Projections [51], propuesta por el grupo, se ha desarrollado un prototipo de interfaz donde se combina el uso de principios de visualización e interacción con métodos web recientes para la exploración visual de datos procedentes de un proceso, como es el análisis de consumos eléctricos en edificios. Este trabajo se presentó en el 19^{th} *IFAC World Congress, 2014* [147].

6.3 LÍNEAS FUTURAS DE INVESTIGACIÓN

La temática tratada en el presente documento, posee un gran potencial para el estudio y desarrollo de trabajo futuro. Existen varios aspectos en los que profundizar para avanzar las cuestiones estudiadas. A continuación se describen las líneas de investigación abiertas y se detallan posibles trabajos para complementar las investigaciones descritas.

- El desarrollo de un estudio de usuario del método de extensión de características para proyecciones de datos para verificar la utilidad de la técnica propuesta en un escenario real. Actualmente se está colaborando con *Middlesex University* en su realización.

- Avanzar en el estudio de la predicción del defecto de *chatter* con datos que incluyan partes antes del fallo de manera que se pueda realizar una predicción más eficiente. Por ejemplo, con un tipo de aprendizaje denominado profundo (*deep learning*) se crean abstracciones de las variables de los datos mediante modelos no lineales. Este enfoque resultaría interesante aplicarlo para el estudio de los estados dinámicos con datos reales del proceso con el objeto de supervisar dicho fallo.

- El diseño de nuevas estrategias de transformación, teniendo en cuenta la naturaleza de los datos, de manera que mejoren la efectividad de los métodos. Por ejemplo, el estudio de otro tipo de magnitudes, como la varianza, en la extensión de características propuesta y cómo modifica la proyección resultante.

- El estudio de otros métodos interactivos que aporten control al usuario en el proceso de la reducción de la dimensión de manera que la proyección se modifica revelando conocimiento nuevo, por ejemplo como la ponderación de las variables de los datos durante el proceso de cálculo de la proyección. El potencial de esta idea de visualizar resultados intermedios causados por los cambios en la métrica producidos por el usuario se presentó en el *European Symposium on Artificial Neural Networks (ESANN 2014)* [21].

- La aplicación de las técnicas propuestas a datos de otros procesos industriales, con diferentes condiciones de funcionamiento. En concreto, sería aplicable a procesos en los que, como los trenes de laminación, interesa el estudio de sus estados de vibración, mediante el análisis de señales periódicas de variables del proceso.

- Avanzar en el desarrollo del prototipo web de exploración interactiva propuesto para datos temporales, con mejoras como por ejemplo en lugar de representar todos los puntos en las vistas y en las transiciones entre ellas, representar el valor agregado correspondiente de cada punto.

7

CONCLUSIONS AND FUTURE WORK

This last chapter gathers the conclusions from the research performed and suggests works which could complement future research lines. Besides, original contributions from the works and the main publications and congress papers that they produced are detailed.

7.1 FINAL CONCLUSIONS

The context of this thesis is the field of data analysis, which uses foundations of several areas such as statistics, artificial intelligence, or visualization to extract new useful knowledge. The great ability for data acquisition and the appearance of new and improved machine learning algorithms have helped the development of this field. This makes easy the research of difficult problems for analysing data from complex processes. A proper visual encoding of information amplifies our cognition facilitating the interpretation of results. Besides, visualization provides a potential to exploit prior knowledge from context information and its combination with interaction mechanisms allows the user to improve data exploration. Several foundations related to visualization and interaction have been discussed in chapter 2.

In this work, the investigation was focused on dimensionality reduction (DR) techniques, which can estimate underlying structure and obtain a visual data projection. The study of numerous existing techniques not only led to explore classic techniques like PCA as a first option to project data, but also the use of non-linear ones, like t-SNE algorithm, which is suitable for estimating the structure of multivariate data in two or three dimensions, that can be visualized on a scatterplot so that close points represent similar objects and vice versa, following the so-called principle of spatialization. A description of several techniques can be seen in chapter 3.

Through various information extraction strategies explained in 4, such as frequency band analysis and the appliance of projection techniques, a fault called *chatter* was characterized with real data from a cold rolling mill [149, 150]. The visualization of vibrational states of the process, on a map that describes dynamical beha-

viours, facilitates identification of the fault and hence monitoring of such erroneous conditions.

In chapter 4 some of the limitations of these algorithms were addressed, like the number of points that a projection can manage. For this, prototypes that represent the whole set of data are computed, and used to calculate a projection. Using a version of radial basis function (GRNN), the projection for all the points can be estimated. This procedure allows generating an approximated data projection without practically size limit. It was applied to electric data, collected during one year, obtaining projections where consumption patterns can be displayed quickly, associating also the price of the corresponding energy [148].

Other limitation is projecting new points on a given projection, for this, the so-called *out-of-sample* techniques are usually required. An initial solution was found for t-SNE technique, whose main drawback was its computational cost, since it required computations for each new point. Later, a similar approach to the one explained previously using radial basis functions was applied for this purpose, resulting in a substantial performance improvement [150].

Data transformation strategies were also explored, using a technique called feature extension, wherein introducing class information contained in data labels, the resulting projection varies, facilitating its interpretation [151]. This information contains prior knowledge that can be hidden in the data, for example, due of noise, or the use of irrelevant variables, or simply not described in the features of the data, which it is managed to emerge in the visualization using the method described in chapter 5.

The study of the combination between dimensionality reduction techniques with data transformation strategies allow the user to modify the projection interactively, facilitating visualization of data projections by separating the introduced groups [152]. This process is reversible and helps to keep in mind the initial data projection as a reference. Therefore the user can examine original data structure, enriched with class information, using a visual tracking of the points between views. This modification of the original projection, where group of points are separated by similarities, makes easy a hierarchical data interpretation.

This has been studied in detail in chapter 5 through several experiments using different sets of data, synthetic and real cases; and different situations, such as temporal data analysis, or using supervised projection techniques. The modified projections were also evaluated using recent metrics to determine the effectiveness of

each transformation. Experimental results show an enhancement of visual quality of the projection through a controlled modification (by the user) of the original structure. This quality assessment helps the analyst to understand the produced modifications and choose a final projection deciding the trade-off between the distortion of original data and visual improvement. Therefore, this method not only can be used to achieve a grouping separation but also to understand the existing overlapping in data.

In conclusion, the visualization of process using data projections and interactive explorations help the analysis and decisions about complex problems. Moreover, the modifications resulting from the exploration process improve data understanding although projections must be evaluated to support the user during analysis. This thesis has addressed some issues in this field. At the same time, it is expected that new questions motivate other researchers to study them in the future.

7.2 SUMMARY OF CONTRIBUTIONS

Two main directions are presented for visual analysis of complex processes: (1) the use of various methods to extract information from a process using data, such as creating a model of the process or the study in the frequency domain of dynamic states that are later represented in a projection, and (2) a new method, based on the modification of a visual projection interactively that eases the exploration for analysis tasks.

The main original contributions can be summarized as follows:

- The use of visual projections have been shown for data analysis from complex processes, including the study of dynamic behavior for fault identification in a rolling mill process and analysis of electric consumptions in buildings.
- An approximated procedure has been suggested for estimating a projection with a large volume of data samples, presented at 5^{th} *International Conference on Physics and Control (PhysCon 2011)* [148].
- An algorithm has been developed to project new points on a projection, computed previously using t-SNE technique [149]. Also, the use of neural networks as alternative to existing *out-of-sample* techniques [150].
- Methods have been applied to extract information about the dynamic behavior of the process, such as creating a model

of a rolling mill process using ELM network, or the study of main variables using frequency band analysis or frequency response functions.

- The defect so-called *chatter* in a rolling mill has been characterized using a visual map that allows a quick fault identification during the process. These works were published in 13^{th} *International Conference on Engineering Applications of Neural Networks (EANN 2012)* [149] and the journal *Engineering Applications of Artificial Intelligence* [150].
- A simple but effective interactive method has been studied which allows the user to introduce grouping information into a data projection. This provides a new control of a static projection, that is gradually modified (by the user) revealing knowledge about the introduced information. This work was presented at workshop *VAMP* in *Eurovis 2013* [151], later it was extended to the journal *Neurocomputing* [152].
- Quantitative measures have been integrated in the proposed method, that assess the quality of the projection, including both visually and structure preservation, to provide additional information that helps the user for interpretation.
- Based on the Morphing Projections technique [51], a prototype interface has been developed where visualization and interaction principles are combined with recent web methods for visual data exploration from a process like the analysis of electric consumptions in buildings. This work was presented at 19^{th} *IFAC World Congress, 2014* [147].

7.3 FUTURE RESEARCH LINES

The topic herein has great potential for the development of future work. There are several research lines to advance in the questions studied. These research lines are described and possible future works are detailed to complement the research described.

- The development of a user study for feature extension method to ascertain the utility of the proposed technique in a real scenario. Currently, a collaboration with *Middlesex University* is opened to its development.
- Further study for predicting *chatter* defect using process data measured before the fault so that a more effective prediction could be performed. For example, using methods like

deep learning, where abstractions of data variables are created using non-linear models. This approach would be interesting for studying dynamic states with real data in order to monitor this fault of the process.

- New transformation strategies, taking into account the nature of the data, in order to improve the effectiveness of the methods. For example, the extension of other types of magnitudes such as variance and how it modifies the projection.

- The study of other interactive methods that provide control to the user in the dimensionality reduction process so that the projection is modified revealing new knowledge, for instance weighting data features during projection is computed. This idea of visualizing intermediate results with metric changes caused by the user was presented in *European Symposium on Artificial Neural Networks (ESANN 2014)* [21].

- The application of the proposed techniques to data from other industrial processes, with different operating conditions. In particular, this could be applied to processes, like rolling mills, where it is interesting a study of the vibration states using periodical signals from variables of the process.

- Further development of the prototype web application proposed for temporal data, with improvements such as representing all points in the transitions between views, and showing the aggregated value corresponding to each point.

BIBLIOGRAFÍA

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

[2] D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine learning*, 6(1):37–66, 1991.

[3] C. Ahlberg. Spotfire: an information exploration environment. *ACM SIGMOD Record*, 25(4):25–29, 1996.

[4] W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *Visualization and Computer Graphics, IEEE Transactions on*, 14(1):47–60, Jan 2008.

[5] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data*. Springer, 2006.

[6] M. Ankerst, D. A. Keim, and H. peter Kriegel. Circle segments: A technique for visually exploring large multidimensional data sets. 1996.

[7] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7):1304–1330, 2007.

[8] A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 1998.

[9] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural computation*, 12(10):2385–2404, 2000.

[10] H.-U. Bauer and K. R. Pawelzik. Quantifying the neighborhood preservation of self-organizing feature maps. *Neural Networks, IEEE Transactions on*, 3(4):570–579, 1992.

[11] R. A. Becker, S. G. Eick, and A. R. Wilks. Visualizing network data. *Visualization and Computer Graphics, IEEE Transactions on*, 1(1):16–28, 1995.

[12] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, volume 14, pages 585–591, 2001.

[13] M. E. H. Benbouzid. A review of induction motors signature analysis as a medium for faults detection. *IEEE Transactions on Industrial Electronics*, 47(5):984–993, Oct. 2000.

[14] Y. Bengio, A. C. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1798–1828, 2013.

[15] J. Bertin. *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin press, 1983.

[16] E. Bertini, A. Tatu, and D. Keim. Quality metrics in high-dimensional data visualization: An overview and systematization. *Proceedings of the IEEE Symposium on IEEE Information Visualization (InfoVis)*, 17:2203–2212, 2011.

[17] C. M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.

[18] C. M. Bishop. *Pattern recognition and machine learning*. Springer, New York, NY, 2009.

[19] C. M. Bishop, M. Svensén, and C. K. Williams. Gtm: The generative topographic mapping. *Neural computation*, 10(1):215–234, 1998.

[20] C. Blake and C. J. Merz. UCI repository of machine learning databases, 1998.

[21] I. D. Blanco, A. A. C. Vega, D. Pérez-López, F. J. García-Fernández, and M. Verleysen. Interactive dimensionality reduction for visual analytics. In *22th European Symposium on Artificial Neural Networks, ESANN 2014, Bruges, Belgium, April 23-25, 2014*, 2014.

[22] A. L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1):245–271, 1997.

[23] D. Borland and R. M. Taylor II. Rainbow color map (still) considered harmful. *IEEE computer graphics and applications*, 27(2):14–17, 2007.

[24] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK; New York, 2004.

[25] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[26] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[27] C. A. Brewer. Color use guidelines for data representation. In *Proceedings of the Section on Statistical Graphics, American Statistical Association*, pages 55–60, 1999.

[28] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Disfunction: Learning distance functions interactively. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 83–92, 2012.

[29] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl. Limited rank matrix learning, discriminative dimension reduction and visualization. *Neural Networks*, 26:159–173, Feb. 2012.

[30] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[31] R. Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[32] A. Cairo. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012.

[33] S. K. Card and J. Mackinlay. The structure of the information visualization design space. In *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pages 92–99. IEEE, 1997.

[34] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[35] D. B. Carr, R. J. Littlefield, W. Nicholson, and J. Littlefield. Scatterplot matrix techniques for large n. *Journal of the American Statistical Association*, 82(398):424–436, 1987.

[36] O. Chapelle, B. Schölkopf, A. Zien, et al. *Semi-supervised learning*, volume 2. MIT press Cambridge, 2006.

[37] L. Chen and A. Buja. Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *Journal of the American Statistical Association*, 104(485):209–219, 2009.

[38] M.-S. Chen, J. Han, and P. S. Yu. Data mining: an overview from a database perspective. *Knowledge and data Engineering, IEEE Transactions on*, 8(6):866–883, 1996.

[39] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973.

[40] J. Choo, H. Lee, J. Kihm, and H. Park. iVisClassifier: an interactive visual analytics system for classification based on supervised dimension reduction. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 27–34, 2010.

[41] W. S. Cleveland. *Visualizing data*. Hobart Press, 1993.

[42] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[43] M. Cottrell, B. Hammer, A. Hasenfuß, and T. Villmann. Batch and median neural gas. *Neural Networks*, 19(6):762–771, 2006.

[44] T. Cover and P. Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[45] T. F. Cox and M. A. Cox. *Multidimensional scaling*. CRC Press, 2010.

[46] M. Dash and H. Liu. Feature selection for classification. *Intelligent data analysis*, 1(3):131–156, 1997.

[47] M. C. F. De Oliveira and H. Levkowitz. From visual data exploration to visual data mining: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 9(3):378–394, 2003.

[48] P. Demartines and J. Hérault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *Neural Networks, IEEE Transactions on*, 8(1):148–154, 1997.

[49] A. P. Dempster, N. M. Laird, D. B. Rubin, et al. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal statistical Society*, 39(1):1–38, 1977.

[50] I. Díaz, A. A. Cuadrado, A. B. Diez, and M. Domínguez. Manifold learning for visualization of vibrational states of a rotating machine. In *ICANN (2)*, pages 285–292, 2011.

[51] I. Diaz-Blanco, M. Dominguez-Gonzalez, A. Cuadrado-Vega, A. Diez-Gonzalez, and J. Fuertes-Martinez. MorphingProjections: Interactive Visualization of Electric Power Demand Time Series. In *Eurographics Conference on Visualization (EuroVis)*, pages 121–125, 2012.

[52] T. G. Dietterichl. Ensemble learning. *The handbook of brain theory and neural networks*, pages 405–408, 2002.

[53] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[54] A. Dix and G. Ellis. Starting simple: adding value to static visualisation through simple interaction. In *Proceedings of the working conference on Advanced visual interfaces*, pages 124–134. ACM, 1998.

[55] P. Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.

[56] N. Elmqvist, P. Dragicevic, and J.-D. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1539–1148, 2008.

[57] N. Elmqvist, A. V. Moere, H.-C. Jetter, D. Cernea, H. Reiterer, and T. Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, page 1473871611413180, 2011.

[58] A. Endert, L. Bradel, and C. North. Beyond control panels: Direct manipulation for visual analytics. *Computer Graphics and Applications, IEEE*, 33(4):6–13, July 2013.

[59] A. Endert, C. Han, D. Maiti, L. House, S. Leman, and C. North. Observation-level interaction with statistical models for visual analytics. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 121–130, 2011.

[60] T. Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[61] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37, 1996.

[62] J.-D. Fekete, J. J. Van Wijk, J. T. Stasko, and C. North. The value of information visualization. In *Information visualization*, pages 1–18. Springer, 2008.

[63] S. Few. *Show me the numbers: Designing tables and graphs to enlighten*. Analytics Press Oakland, CA, 2004.

[64] S. Few. *Information dashboard design*. O'Reilly, 2006.

[65] S. Few. *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press, 2009.

[66] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7):179–188, 1936.

[67] R. Fletcher. *Practical methods of optimization*. John Wiley & Sons, 1987.

[68] R. W. Floyd. Algorithm 97: shortest path. *Communications of the ACM*, 5(6):345, 1962.

[69] Y. Freund and R. E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

[70] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[71] M. Friendly. A brief history of data visualization. *Handbook of Computational Statistics: Data Visualization*, 3, 2006.

[72] F. García, I. Díaz, I. Alvarez, D. Pérez, D. González, and M. Domínguez. Spectrogram analysis using manifold learning. In *PHYSCON*, León, Sept. 2011.

[73] F. García, I. Díaz, I. Álvarez, D. Pérez, D. Ordonez, and M. Domínguez. Time-frequency analysis of hot rolling using manifold learning. *IFIP Advances in Information and Communication Technology*, 363 AICT:150–155, 2011.

[74] A. Globerson and S. Roweis. Metric learning by collapsing classes. In *Nips*, volume 18, pages 451–458, 2005.

[75] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. *NIPS'04*, 2004.

[76] P. A. González and J. M. Zamarreño. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, 37(6):595 – 601, 2005.

[77] R. L. Graham. An efficient algorith for determining the convex hull of a finite planar set. *Information processing letters*, 1(4):132–133, 1972.

[78] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.

[79] D. J. Hand, H. Mannila, and P. Smyth. *Principles of data mining*. MIT press, 2001.

[80] J. R. Harger and P. J. Crossno. Comparison of open-source visual analytics toolkits. In *IS&T/SPIE Electronic Imaging*, pages 82940E–82940E. International Society for Optics and Photonics, 2012.

[81] T. Hastie, R. Tibshirani, and J. J. H. Friedman. *The elements of statistical learning*, volume 1. Springer New York, 2009.

[82] S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[83] X. He and P. Niyogi. Locality preserving projections. In *NIPS*, volume 16, pages 234–241, 2003.

[84] J. Heer, M. Bostock, and V. Ogievetsky. A tour through the visualization zoo, 2010.

[85] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 32–39. IEEE, 2005.

[86] J. Heer and G. G. Robertson. Animated transitions in statistical data graphics. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1240–1247, 2007.

[87] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Queue*, 10(2):30, 2012.

[88] G. Hinton and S. Roweis. Stochastic neighbor embedding. In *NIPS*, volume 2, pages 833–840, 2002.

[89] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, July 2006.

[90] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[91] P.-H. Hu and K. F. Ehmann. A dynamic model of the rolling process. part i: homogeneous model. *International Journal of Machine Tools and Manufacture*, 40(1):1–19, Jan. 2000.

[92] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501, Dec. 2006.

[93] S. Ingram, T. Munzner, V. Irvine, M. Tory, S. Bergner, and T. Möller. DimStiller: workflows for dimensional analysis and reduction. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 3–10, 2010.

[94] A. Inselberg and B. Dimsdale. Parallel coordinates. In *Human-Machine Interactive Systems*, pages 199–233. Springer, 1991.

[95] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[96] R. A. Jarvis. On the identification of the convex hull of a finite set of points in the plane. *Information Processing Letters*, 2(1):18–21, 1973.

[97] D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang. iPCA: an interactive system for PCA-based visual analytics. In *Computer Graphics Forum*, volume 28, pages 767–774, 2009.

[98] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):993–1000, 2009.

[99] P. Joia, F. Paulovich, D. Coimbra, J. Cuminato, and L. Nonato. Local affine multidimensional projection. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2563–2571, Dec 2011.

[100] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[101] L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *arXiv preprint cs/9605103*, 1996.

[102] K. Karhunen. *Zur spektraltheorie stochastischer prozesse*. Suomalainen tiedeakatemia, 1946.

[103] D. A. Keim. Designing pixel-oriented visualization techniques: Theory and applications. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1):59–78, 2000.

[104] D. A. Keim. Visual exploration of large data sets. *Communications of the ACM*, 44(8):38–44, 2001.

[105] D. A. Keim. Information visualization and visual data mining. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):1–8, 2002.

[106] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. *Visual analytics: Scope and challenges*. Springer, 2008.

[107] D. G. Kendall. A survey of the statistical theory of shape. *Statistical Science*, 4(2):87–99, 1989.

[108] Y. Kimura, Y. Sodani, N. Nishiura, N. Ikeuchi, and Y. Mihara. Analysis of chatter in tandem cold rolling mills. *Isij International*, 43(1):77–84, 2003. WOS:000180875000011.

[109] A. Kirk. *Data Visualization: a successful design process*. Packt Publishing Ltd, 2012.

[110] T. Kohonen. *Self-organizing maps*. Springer, 2001.

[111] M. A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal*, 37(2):233–243, 1991.

[112] J. B. Kruskal and M. Wish. *Multidimensional scaling*, volume 11. Sage, 1978.

[113] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, pages 79–86, 1951.

[114] J. Lamping, R. Rao, and P. Pirolli. A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408. ACM Press/Addison-Wesley Publishing Co., 1995.

[115] D. Laney. 3d data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6, 2001.

[116] J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11(1):65–100, 1987.

[117] J. A. Lee, A. Lendasse, and M. Verleysen. Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis. *Neurocomputing*, 57:49–76, 2004.

[118] J. A. Lee, E. Renard, G. Bernard, P. Dupont, and M. Verleysen. Type 1 and 2 mixtures of kullback-leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing*, 112:92–108, July 2013.

[119] J. A. Lee and M. Verleysen. *Nonlinear dimensionality reduction*. Springer, New York; London, 2007.

[120] J. A. Lee and M. Verleysen. Quality assessment of nonlinear dimensionality reduction based on k-ary neighborhoods. In *JMLR: workshop and conference proceedings*, volume 4, pages 21–35, 2008.

[121] J. A. Lee and M. Verleysen. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing*, 72(7–9):1431–1443, Mar. 2009.

[122] J. A. Lee and M. Verleysen. Scale-independent quality criteria for dimensionality reduction. *Pattern Recognition Letters*, 31(14):2248–2257, Oct. 2010.

[123] Y. K. Leung and M. D. Apperley. A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(2):126–160, 1994.

[124] M. Li and J. Kwok. Making large-scale nystrom approximation possible. In *ICML*, pages 631–638. Omnipress, 2010.

[125] H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *Knowledge and Data Engineering, IEEE Transactions on*, 17(4):491–502, 2005.

[126] M. Loève. *Fonctions aléatoires du second ordre*. 1965.

[127] L. W. MacDonald. Using color effectively in computer graphics. *Computer Graphics and Applications, IEEE*, 19(4):20–35, 1999.

[128] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.

[129] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. California, USA, 1967.

[130] G. M. Mamani, F. M. Fatore, L. G. Nonato, and F. V. Paulovich. User-driven feature space transformation. In *Computer Graphics Forum*, volume 32, pages 291–299. Wiley Online Library, 2013.

[131] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. Big data: The next frontier for innovation, competition, and productivity, May 2011.

[132] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963.

[133] T. Martinetz, K. Schulten, et al. *A "neural-gas" network learns topologies*. University of Illinois at Urbana-Champaign, 1991.

[134] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten. Neural-gas' network for vector quantization and its application to time-series prediction. *Neural Networks, IEEE Transactions on*, 4(4):558–569, 1993.

[135] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton. Big data. *The management revolution. Harvard Bus Rev*, 90(10):61–67, 2012.

[136] P. A. Meehan. Vibration instability in rolling mills: modeling and experimental results. *Journal of vibration and acoustics*, 124(2):221–228, 2002.

[137] I. Meirelles. *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. Rockport Publishers, 2013.

[138] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and A. Lendasse. OP-ELM: Optimally pruned extreme learning machine. *IEEE Transactions on Neural Networks*, 21(1):158–162, Jan. 2010.

[139] B. Mokbel, W. Lueks, A. Gisbrecht, and B. Hammer. Visualizing the quality of dimensionality reduction. *Neurocomputing*, 112:109–123, July 2013.

[140] A. Morán, J. J. Fuertes, M. A. Prada, S. Alonso, P. Barrientos, I. Díaz, and M. Domínguez. Analysis of electricity consumption profiles in public buildings with dimensionality reduction techniques. *Engineering Applications of Artificial Intelligence*, 26(8):1872 – 1880, 2013.

[141] A. Moreira and M. Y. Santos. Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points. In *GRAPP 2007 : proceedings of the International Conference on Computer Graphics Theory and Applications. INSTICC Press, 2007. ISBN 978-972-8865-71-9.*, pages 61–68., 2007.

[142] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[143] D. A. Norman. Things that make us smart, 1993.

[144] D. L. Paton and S. Critchley. Tandem mill vibration: Its cause and control. In *Mechanical Working; Steel Processing XXII, Proceedings of the 26th Mechanical Working; Steel Processing Conference.*, pages 247–255, Chicago, IL, USA, 1985. Iron and Steel Soc Inc.

[145] F. V. Paulovich, C. T. Silva, and L. G. Nonato. User-centered multidimensional projection techniques. *Computing in Science & Engineering*, 14(4):0074–81, 2012.

[146] K. Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[147] D. Pérez, I. Díaz, A. A. Cuadrado, F. J. García-Fernández, A. B. Diez, and M. Domínguez. Power-consumption analysis through web-based visual data exploration. *Proceedings of the 19th IFAC World Congress*, pages 11257–11262, 2014.

[148] D. Pérez, I. Díaz, F. García, and P. Barrientos. Visual analysis of electrical power consumptions using manifold learning. In *PHYSCON*, León, Sept. 2011.

[149] D. Pérez, F. García-Fernández, I. Díaz, A. Cuadrado, D. Ordonez, A. Díez, and M. Domínguez. Visual analysis of a cold rolling process using data-based modeling. *Communications in Computer and Information Science*, 311:244–253, 2012.

[150] D. Pérez, F. García-Fernández, I. Díaz, A. Cuadrado, D. Ordonez, A. Díez, and M. Domínguez. Visual analysis of a cold rolling process using a dimensionality reduction approach. *Engineering Applications of Artificial Intelligence*, 26(8):1865–1871, 2013.

[151] D. Pérez, L. Zhang, M. Schaefer, T. Schreck, D. Keim, and I. Díaz. Interactive visualization and feature transformation for multidimensional data projection. In *Proc. EuroVis Workshop on Visual Analytics Using Multidimensional Projections*, 2013.

[152] D. Pérez, L. Zhang, M. Schaefer, T. Schreck, D. Keim, and I. Díaz. Interactive feature space extension for multidimensional data projection. *Neurocomputing*, 150, Part B(0):611 – 626, 2015.

[153] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[154] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 318–322. ACM, 1994.

[155] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.

[156] W. L. Roberts. *Cold rolling of steel*. Marcel Dekker, Inc., New York, 1978.

[157] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1325–1332, 2008.

[158] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[159] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986.

[160] J. W. Sammon Jr. A nonlinear mapping for data structure analysis. *Computers, IEEE Transactions on*, 100(5):401–409, 1969.

[161] L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee. Spectral methods for dimensionality reduction. *Semi-supervised learning*, pages 293–308, 2006.

[162] M. Schäfer, L. Zhang, T. Schreck, A. Tatu, J. A. Lee, M. Verleysen, and D. A. Keim. Improving projection-based data analysis by feature space transformations. In *IS&T/SPIE Electronic Imaging*, pages 86540H–86540H. International Society for Optics and Photonics, 2013.

[163] R. R. Schoen, T. G. Habetler, F. Kamran, and R. G. Bartheld. Motor bearing damage detection using stator current monitoring. *IEEE Transactions on Industry Applications*, 31(6):1224–1279, Nov. 1995.

[164] B. Scholkopf and A. Smola. Learning with kernels, 2002.

[165] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.

[166] B. Scholkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[167] H. Schulz. Treevis. net: A tree visualization reference. *Computer Graphics and Applications, IEEE*, 31(6):11–15, 2011.

[168] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A taxonomy of visual cluster separation factors. In *Computer Graphics Forum*, volume 31, pages 1335–1344. Wiley Online Library, 2012.

[169] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.

[170] R. N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. i. *Psychometrika*, 27(2):125–140, 1962.

[171] P. Shirley, M. Ashikhmin, and S. Marschner. *Fundamentals of computer graphics*. CRC Press, 2009.

[172] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, 11(1):92–99, 1992.

[173] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[174] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Comput. Graph. Forum*, 28(3):831–838, 2009.

[175] D. F. Specht. A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6):568–576, Nov. 1991.

[176] R. Spence and L. Tweedie. The attribute explorer: information synthesis via exploration. *Interacting with Computers*, 11(2):137–146, 1998.

[177] S. S. Stevens. On the theory of scales of measurement, 1946.

[178] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *The Journal of Machine Learning Research*, 8:1027–1061, 2007.

[179] T. Tamiya, K. Furui, and H. Iida. Analysis of chattering phenomenon in cold rolling. In *International Conference on Steel Rolling,*, volume 2, pages 1191–1202, 1980.

[180] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, and D. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)*, pages 59–66, 2009.

[181] P. J. Tavner and J. Penman. *Condition Monitoring of Electrical Machines*. Research Studies Press Ltd., John Wiley and Sons Inc., 1987.

[182] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[183] J. J. Thomas and K. A. Cook. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.

[184] J. Tlusty, G. Chandra, S. Critchley, and D. Paton. Chatter in cold rolling. *CIRP Annals - Manufacturing Technology*, 31(1):195–199, 1982.

[185] E. Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, USA, 1990.

[186] E. R. Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, Cheshire, CT, USA, 1997.

[187] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.

[188] B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.

[189] A. Ultsch and H. P. Siemon. Kohonen's self organizing feature maps for exploratory data analysis. In *Proc. INNC'90, Int. Neural Network Conf.*, pages 305–308, Dordrecht, Netherlands, 1990. Kluwer.

[190] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[191] L. J. P. Van der Maaten. An introduction to DR using matlab, 2007.

[192] L. J. P. Van der Maaten, E. O. Postma, and H. J. Van Den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10:1–41, 2009.

[193] J. J. Van Wijk. The value of visualization. In *Visualization, 2005. VIS 05. IEEE*, pages 79–86. IEEE, 2005.

[194] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, pages 4–9. IEEE, 1999.

[195] J. Venna. *Dimensionality reduction for visual exploration of similarity structures*. PhD thesis, Helsinki University of Technology, Espoo, 2007.

[196] J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *The Journal of Machine Learning Research*, 11:451–490, 2010.

[197] J. Vesanto. Som-based data visualization methods. *Intelligent data analysis*, 3(2):111–126, 1999.

[198] T. Villmann, R. Der, M. Herrmann, and T. M. Martinetz. Topology preservation in self-organizing feature maps: Exact definition and measurement. *Neural Networks, IEEE Transactions on*, 8(2):256–266, 1997.

[199] M. O. Ward. Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Proceedings of the Conference on Visualization'94*, pages 326–333. IEEE Computer Society Press, 1994.

[200] M. O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4):194–210, 2002.

[201] M. O. Ward and J. Yang. Interaction spaces in data and information visualization. In *VisSym*, pages 137–145, 2004.

[202] C. Ware. *Information visualization: perception for design*. Morgan Kaufman, San Francisco, CA, 2004.

[203] C. Ware. *Visual thinking: For design*. Morgan Kaufmann, 2010.

[204] L. Wilkinson, D. Wills, D. Rope, A. Norton, and R. Dubbs. *The grammar of graphics*. Springer, 2006.

[205] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[206] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization*, pages 3–33, 1994.

[207] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, Jan. 2008.

[208] L. Yang. Sammon's nonlinear mapping using geodesic distances. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 303–306. IEEE, 2004.

[209] Z. Yang, J. Peltonen, and S. Kaski. Scalable optimization of neighbor embedding for visualization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 127–135, 2013.

[210] J. S. Yi, Y. ah Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1224–1231, 2007.

[211] J. S. Yi, R. Melton, J. Stasko, and J. A. Jacko. Dust & magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4(4):239–256, 2005.

[212] I. Yun, W. Wilson, and K. Ehmann. Review of chatter studies in cold rolling. *International Journal of Machine Tools and Manufacture*, 38(12):1499–1530, 1998.

[213] L. Zhang, A. Stoffel, M. Behrisch, S. Mittelstadt, T. Schreck, R. Pompl, S. Weber, H. Last, and D. Keim. Visual analytics for the big data era - a comparative review of state-of-the-art commercial systems. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 173–182. IEEE, 2012.

[214] M. X. Zhou and S. K. Feiner. Visual task characterization for automated visual discourse synthesis. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 392–399. ACM Press/Addison-Wesley Publishing Co., 1998.

A

PUBLICACIONES

A continuación se adjuntan los artículos más relevantes, que fueron desarrollados en el contexto de los diferentes trabajos explicados anteriormente en el documento.

En primer lugar, se encuentran los dos artículos publicados en revistas, seguidos de otros que fueron presentados en conferencias.

No se han adjuntado otros trabajos como por ejemplo [72, 73, 21] puesto que, aunque se colaboró en su realización, poseen una menor parte del contenido expuesto en esta tesis.

A.1 VISUAL ANALYSIS OF A COLD ROLLING PROCESS USING A DIMENSIONALITY REDUCTION APPROACH [150]

Engineering Applications of Artificial Intelligence 26 (2013) 1865–1871

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Visual analysis of a cold rolling process using a dimensionality reduction approach ☆

Daniel Pérez [a,*], Francisco J. García-Fernández [a], Ignacio Díaz [a], Abel A. Cuadrado [a], Daniel G. Ordonez [a], Alberto B. Díez [a], Manuel Domínguez [b]

[a] Universidad de Oviedo, Área de Ingeniería de Sistemas y Automática, Spain
[b] Universidad de León, Instituto de Automática y Fabricación, Spain

ARTICLE INFO

Article history:
Received 20 December 2012
Received in revised form
9 May 2013
Accepted 18 May 2013
Available online 12 June 2013

Keywords:
Dimensionality reduction
Rolling process
Data visualization
Dynamical systems
Fault detection
Frequency analysis

ABSTRACT

The rolling process is a strategical industrial and economical activity that has a large impact among world-wide commercial markets. Typical operating conditions during the rolling process involve extreme mechanical situations, including large values of forces and tensions. In some cases, these scenarios can lead to several kinds of faults, which might result in large economic losses. Thereby, a proper assessment of the process condition is a key aspect, not only as a fault detection mechanism, but also as an economic saving system. In the rolling process, a remarkable kind of fault is the so-called chatter, a sudden powerful vibration that affects the quality of the rolled material. In this paper, we propose a visual approach for the analysis of the rolling process. According to physical principles, we characterize the exit thickness and the rolling forces by means of a large dimensional feature vector, that contains the energies at specific frequency bands. Afterwards, we use a dimensionality reduction technique, called *t*-SNE, to project all feature vectors on a visual 2D map that describes the vibrational states of the process. The proposed methodology provides a way for an exploratory analysis of the dynamic behaviors in the rolling process and allows to find relationships between these behaviors and the chatter fault. Experimental results from real data of a cold rolling mill are described, showing the application of the proposed approach.

1. Introduction

Cold rolling is one of the most common processes in the field of metalworking, where the thickness of a steel sheet is reduced by passing it through two rolls. This reduction enlarges the strip, that is finally wound forming a coil. There are different types of variables involved, with complex control loops in very tough operating conditions. The increasing demand makes it necessary to do a continuous improvement of the efficiency of the process, and the supervision is fundamental in order to minimize faults, that may damage the material and the facilities.

One of the most serious faults in the cold rolling process of steel is the so-called *chatter* (Yun et al., 1998), an unexpected powerful vibration that takes place suddenly during the process, causing unacceptable variations of the final thickness or even the rupture of the strip. From the existing types of mill oscillations, in this work we consider the third-octave mode of vibration, a harmful mode that is recognizable by the resonant frequencies in the range between 100 and 300 Hz.

Chatter fault is a complex phenomenon resulting from a set of dynamic interactions between the mill stand and the strip being rolled. For a better understanding of these dynamics, many models of the different subsystems of the process have been developed (Orowan, 1943; Chefneux et al., 1984; Hu and Ehmann, 2000) that help to explain the nature, the causes and the stability of this kind of faults (Tamiya et al., 1980; Tlusty et al., 1982). The main results show that the chatter effect is a self-excited vibration, whose elimination is usually achieved by a speed reduction.

While sophisticated models have been developed in the literature, their use can be tedious because of their mathematical complexity. The complex nature of such models makes it difficult to grasp useful knowledge about the mill behavior, as well as to develop insightful applications that are suitable for use in plant by less specialized technicians. Besides this, such models often require tuning large numbers of parameters as well as to validate the model against real data, what makes them less practical. While models based on simplifying assumptions can be built, they often result in a loss of accuracy, and still require parameter tuning and validation against process data. This complexity of the existing models and the availability of large amounts of sensor data have

☆The material in this paper was partially presented at the 2012 EANN Conference (EANN 2012), September 20–23, 2012, London, United Kingdom.

* Corresponding author. Tel.: +34 985182543.
E-mail address: dperez@isa.uniovi.es (D. Pérez).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.05.009

suggested the use of data-based methods in previous works (Tansel et al., 1991), not only applied to cold rolling processes (Heidari and Forouzan, 2013) but also more actively to machine tools (Kuljanic et al., 2008; Quintana and Ciurana, 2011).

An alternative approach to enhance the knowledge about complex processes is visualizing their relevant information on a map (Alhoniemi et al., 1999; Díaz et al., 2008; Pérez et al., 2012). *Dimensionality reduction* (DR) techniques allow to project the underlying structure of high-dimensional data into a low-dimensional space, typically a 2D/3D for visualization purposes, improving the exploratory data analysis.

In the DR field, several techniques have been proposed (Lee and Verleysen, 2007; Kohonen, 2001). Traditional techniques, such as Principal Component Analysis (PCA) (Jollife, 1986), and Multi-dimensional Scaling (MDS) methods (Young and Householder, 1938; Torgerson, 1952) are based on a linear approach. In order to deal with complex datasets, which are more likely to be found in real applications, nonlinear approaches appeared later. In the late 1960s, Sammon proposed a nonlinear variation of the MDS algorithm (Sammon, 1969). Later, Kohonen proposed the self-organizing map (SOM) (Kohonen, 1990) that defines a topologically ordered nonlinear mapping from the data space on a low dimensional lattice for visualization. Also "bottleneck" architectures of feedforward neural networks have been proposed for dimensionality reduction as in Kramer (1991), and in the much more recent related approach called "deep autoencoder" networks (Hinton and Salakhutdinov, 2006). In the beginning of the 21st century, newer nonlinear techniques, based on neighbor embedding, were proposed, including convex techniques such as *Isomap* (Tenenbaum et al., 2000), *local linear embedding* (LLE) (Roweis and Saul, 2000) and *Laplacian eigenmaps* (LE) and non-convex techniques like *t-Stochastic Neighbor Embedding* (*t*-SNE) (Van Der Maaten and Hinton, 2008), and SNE (Hinton and Roweis, 2003). Particularly, *t*-SNE has attracted attention recently (Jamieson et al., 2010; Bushati et al., 2011), due to its performance working with high-dimensional datasets.

The contribution of this paper is the characterization of chatter faults by the visualization of vibrational states of the process in a low dimensional space. The projections of frequency-domain data are computed by a dimensionality reduction technique that allows a visual detection of this kind of faults. The paper is organized as follows: in Section 2, we describe the proposed method, that involves using frequency band analysis to characterize the vibrational state by means of a large dimensional feature vector, followed by a dimensionality reduction stage that maps all the feature vectors on a 2D visual map that allows for visual exploration; in Section 3, experiments are presented to validate the proposed methodology; in Section 4, the results are discussed; finally, Section 5 concludes the paper.

2. Data-based model analysis through manifold learning techniques

2.1. Description of the physical model

Classical cold rolling models try to calculate the necessary force and torque for a given thickness reduction. The complexity of an accurate model can be very high because of the assumptions taken (Venter and Abd-Rabbo, 1980). In order to get a simple model to work with, e.g. (Freshwater, 1996), several assumptions can be done.

A model of the rolling process makes it possible to analyze several faults arising from operating conditions, such as the chatter phenomenon. This phenomenon is a dynamic process, where variations in the roll force may lead to an unstable state. It is necessary to generate a model where the different factors that are likely to modify the force equilibrium in the rolling process are taken into account. As explained in Paton and Critchley (1985), the chatter phenomenon comes from a feedback interaction among the involved variables: the entry speed, the entry and exit tension, the force of the strip on the rolls and the exit thickness. If a dynamic model of the stand is added to this loop, a proper model to study the chatter phenomenon can be built (Meehan, 2002; Kimura et al., 2003).

The classical rolling model relates the rolling force, the tension at the entry and exit side of a rolling stand, the thickness at the entry and exit side, the width of the strip, the friction coefficient and the hardness of the material being rolled.

Based on previous theoretical studies (Roberts, 1978) and according to mechanical and physical relationships, a simplified model of the cold rolling process can be proposed as

$$y = f(F, \sigma_{en}, V_{en}, V_{ex}) \tag{1}$$

where y is the output of the model that corresponds to the exit thickness, F is the rolling force, σ_{en} is the entry tension (excluding exit tension because it is considered constant) and, finally, V_{en} and V_{ex} are the entry and exit speed of the strip respectively. On the other hand, in order to take into account the dynamic behavior of the rolling process, a linearized transfer function model is derived from the previous expression (1)

$$Y(s) = \frac{1}{A(s)}[k_1F(s) + k_2\sigma_{en}(s) + k_3V_{en}(s) + k_4V_{ex}(s)] \tag{2}$$

In this model, $A(s)$ represents the poles of the physical system and, therefore defines the transient modes of vibration, while $k_1F(s)$, $k_2\sigma_{en}(s)$, $k_3V_{en}(s)$ and $k_4V_{ex}(s)$ constitute the forced dynamics, influencing the coefficients of the transient modes of vibration. In our case, $k_1F(s)$ contains most of the vibrational information of the process, since the speeds are used to obtain the operating point and the signal σ_{en} has a sampling rate that cannot provide frequency information about chatter. Therefore the rolling force F is chosen for the analysis of the dynamics of the process.

2.2. Characterization of the dynamical behavior

In this case, to characterize the vibrational state and estimate the dynamical behavior of the cold rolling process, a *frequency-band analysis* (FBA) approach is performed. Frequency domain analysis is a widely used approach for the analysis of electrical and mechanical machinery (Díaz et al., 2011; Tavner and Penman, 1987; Benbouzid, 2000; Schoen et al., 1995). The main reason for using frequency domain approaches is based on the fact that they provide sparser representations for periodical signals than time domain representations. For such signals, the energy appears to be concentrated at certain frequencies, being zero or significantly low for the other ones. This fact, allows us to represent most of the energy of the signal using only a fraction of the frequencies, resulting in a much more compact description, that uses only a few parameters – namely, the energies at certain "harmonics" – to describe the main variations of the signal as a result of a change in the working condition.

Given a measured variable $x(t)$ at regular time intervals with a sample period T, producing a sequence $x_k = x(kT)$, consider overlapped windows of length N_w each displaced $n_d < N_w$ samples apart

$$\mathbf{x}_n = [x_{n\cdot n_d},\ x_{n\cdot n_d+1},\ \ldots, x_{n\cdot n_d+N_w-1}]$$

where $n_d = (1-(L/100)) \cdot N_w$, being L the percentage of overlapped samples, and producing N overlapped segments.

In this paper a windowed DFT transform is used to avoid Gibbs effect

$$X_i = \sum_{k=0}^{N_w-1} w(k) x_k e^{-j2\pi ik/N_w}, \quad i = 0, \ldots, N_w-1 \tag{3}$$

where x_k is a real or complex data sequence, X_i is a complex sequence describing the amplitudes and phases of the signal harmonics and $w(k)$ is a windowing function – Hanning in this paper. For the nth window $\mathbf{x}_n$, the energies in bands around m specified center frequencies $f_1, f_2, \ldots, f_m$ with predefined bandwidths $B_1, B_2, \ldots, B_m$ can be computed by summing up the squares of the harmonics inside the bands, to obtain a m-dimensional feature vector

$$\mathbf{d}_n = [d_{1n}, d_{2n}, \ldots d_{mn}]^T, \quad d_{jn} = \sqrt{\sum_{i/(N_w T) \in [f_j - (B_j/2), f_j + (B_j/2)]} \|X_i\|^2} \tag{4}$$

Feature vectors can be arranged into a *data matrix* $\mathbf{D}_x = (d_{jn}) = [\mathbf{d}_1, \mathbf{d}_2, \ldots \mathbf{d}_N]$, where d_{jn} represents the energy in the band $\{f_j, B_j\}$ – that is, with center frequency f_j and width B_j – for window n of the signal $x(t)$. The feature extraction process for the nth buffer is summarized in the block diagram of Fig. 1.

Based on the assumptions taken, namely, that the vibrational information of the rolling process is conveyed by $y(t)$ and $F(t)$, the FBA is applied to these variables, obtaining $\mathbf{D}_y$ and $\mathbf{D}_F$, respectively. Finally, an augmented matrix $\mathbf{H}$ can be constructed as

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots \mathbf{h}_N] = \begin{bmatrix} \mathbf{D}_y \\ \mathbf{D}_F \end{bmatrix} \quad \mathbf{H} \in \mathbb{R}^{2m \times N}, \ \mathbf{h}_n \in \mathcal{D} = \mathbb{R}^{2m} \tag{5}$$

where $\mathbf{h}_n$ is a vector containing the FBA of $y(t)$ and $F(t)$ for the nth window, and the input data space $\mathcal{D} = \mathbb{R}^{2m}$ contains the total vibrational states of the rolling process.

2.3. Visualization of the dynamical behavior using dimensionality reduction techniques

As a result of the previous process, N vectors of 2 m dimensions are obtained for each coil, which describe the evolution of the vibrational state in $\mathcal{D}$. Obviously, the vibrational states arising from these N behaviors are far from filling up the space $\mathcal{D}$; on the other hand, since constraints due to physical laws – complex and tightly coupled physical, mechanical, thermal and electrical phenomena – can be seen as equations that reduce degrees of freedom on the original space, the resulting feature vectors contained in $\mathbf{H}$ can be supposed to lie on a low-dimensional manifold embedded in $\mathcal{D}$.

According to this setup, a dimensionality reduction approach is suggested. DR methods can extract information about the low-dimensional structures of data and unfold this information in a visualization space.

Although there are several DR techniques – see Section 1 – we use *t-Stochastic Neighbor Embedding* (*t*-SNE) (Van Der Maaten and Hinton, 2008). According to visualization and performance requirements, one of the reasons to choose this DR algorithm is that the quality of the embeddings when working with real-world datasets is usually better with *t*-SNE (Lee and Verleysen, 2010). Aside from this, *t*-SNE is always capable of reducing the dimensionality of all the points contained in the dataset, unlike graph-based DR techniques, such as Isomap, LE or LLE, that need a fully connected graph of the complete dataset.

t-SNE is a non-convex DR technique, that is based on minimizing the differences between two probability distributions corresponding to the input and the visualization space, respectively, so that similarities of the data can be revealed in a two-dimensional projection. Initially the technique converts Euclidean distances between points into probabilities and defines a *t*-Student distribution for the projection space. In order to define these probabilities, a parameter called *perplexity*, P, which can be understood as a size of a soft K-ary neighborhood, is used. The low-dimensional map is computed by the minimization of the Kullback–Leibler divergence between the input and the output probability distributions (Van Der Maaten and Hinton, 2008).

Owing to *t*-SNE's computational and memory costs, the number of points is a key aspect to analyze: if it is too large, the execution time experiences a large increase, which is generally a bad situation, but particularly for industrial applications. To solve this problem, we compute a vector quantization of the N vectors,

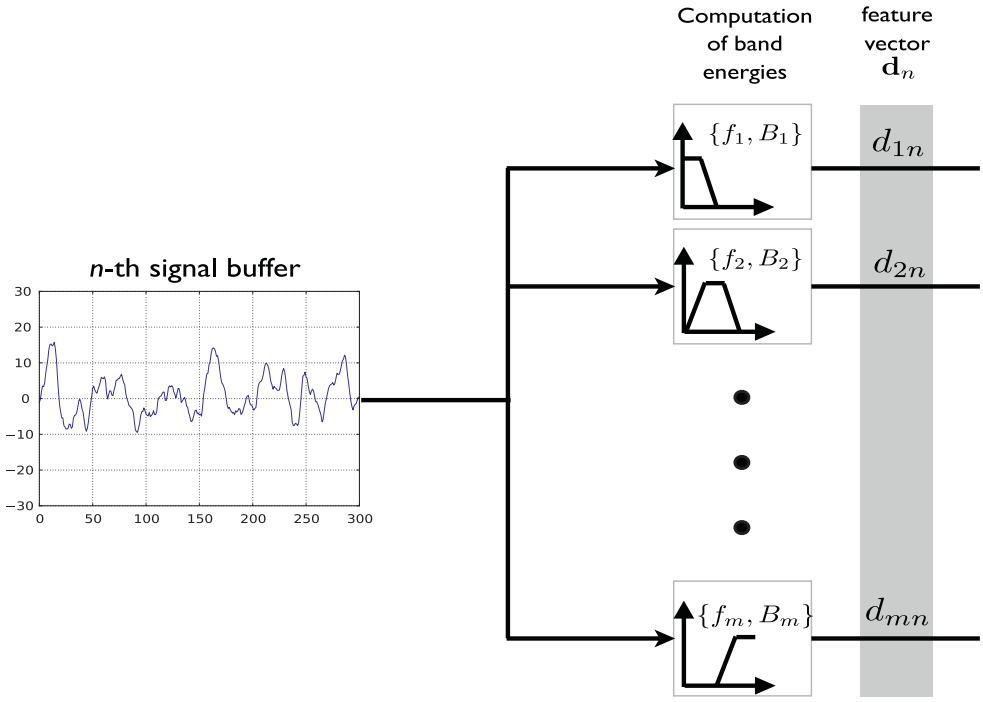


Fig. 1. Block diagram of the feature extraction for one signal buffer.

contained in $\mathbf{H}=(\mathbf{h}_n)$, using a *neural gas* algorithm (Martinetz et al., 1993; Cottrell et al., 2005). Using this approach, we obtain a set of n_r codebook vectors $\{\mathbf{w}_i\}_{1\le i\le n_r}$ that are approximately distributed following the same joint probability density function as the original input vectors. These codebook vectors are projected using t-SNE with a perplexity P, resulting in a set of points $\{\mathbf{q}_i\}_{1\le i\le n_r}$. The pair $(\mathbf{w}_i,\mathbf{q}_i)$ relates the high dimensional vibration data space $\mathcal{D}$ with the low dimensional visualization space $\mathcal{V}$.

As an alternative to the out-of-sample extension of t-SNE discussed in Pérez et al. (2012), a much faster radial basis function network (RBFN) (Moody and Darken, 1989) can be used to generalize the mapping to the N points, as well as to compute out-of-sample approximations for the test coils. Given the set of input feature vectors $\mathbf{w}_i$ and their corresponding output data points $\mathbf{q}_i$, the RBFN estimate of the resulting projection $\hat{\mathbf{y}}$ for a new feature vector $\mathbf{h}$ can be computed as

$$\hat{\mathbf{y}}=\sum_i \mathbf{a}_i\psi_i(\mathbf{h}) \tag{6}$$

where $\psi_i(\mathbf{h})$ are normalized Gaussian kernels

$$\psi_i(\mathbf{h})=\frac{\exp\left(-\frac{\|\mathbf{h}-\mathbf{w}_i\|^2}{2\sigma^2}\right)}{\sum_i \exp\left(-\frac{\|\mathbf{h}-\mathbf{w}_i\|^2}{2\sigma^2}\right)} \tag{7}$$

such that $\sum\psi_i(\mathbf{h})=1$, where σ is the kernel width and the coefficients $\mathbf{a}_i$ can be determined using a standard regularized least squares approach, $\mathbf{A}=(\mathbf{a}_i)^T=(\Psi^T\Psi+\lambda\mathbf{I})^{-1}\Psi^T\mathbf{Q}$, where Ψ is a kernel matrix such that $\Psi_{ji}=(\psi_i(\mathbf{w}_j))$, $\mathbf{Q}=(\mathbf{q}_j)^T$ and λ is the Tikhonov regularization factor.

As a summary of the methodology, a flowchart of the steps involved is shown in Fig. 2.

3. Experiments

As a validation of the proposed methodology, we apply this approach to data from a cold rolling facility. As explained in

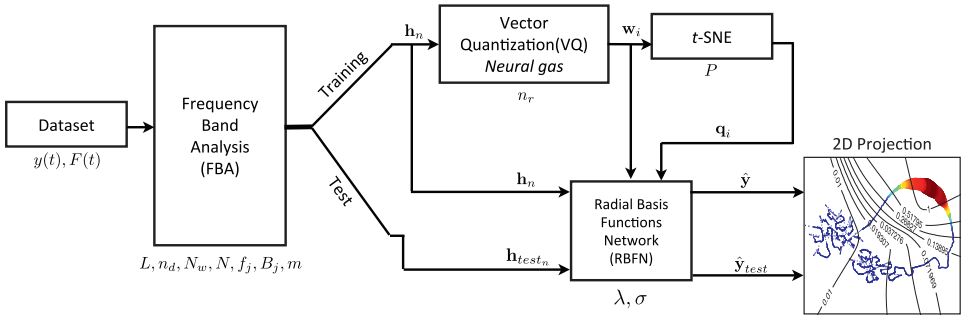


Fig. 2. Flowchart of the method.

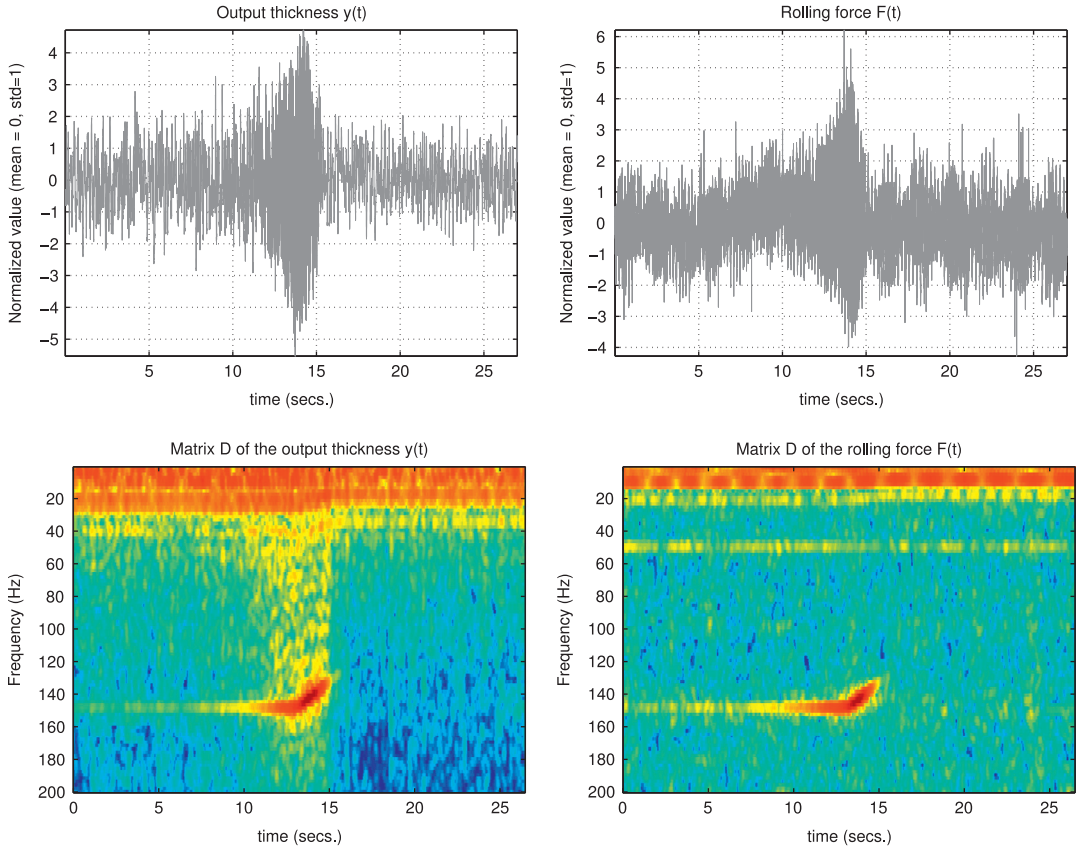


Fig. 3. Force and output signals of a chatter episode (above) and spectrograms-like visualization (below).

Section 2.2, our analysis is based on the exit thickness, $y(t)$, and the rolling force, $F(t)$, for the dynamical analysis of the rolling process. These variables were acquired using a data acquisition system with a sampling rate $F'_m = 2000$ Hz ($T' = 5 \times 10^{-4}$ s). Since the chatter phenomenon appears between 100 and 300 Hz, a decimation of ratio $r=2$ was applied to the signals in order to reduce the sample size, so the final sampling rate is $F_m = 1000$ Hz ($T = 10^{-3}$ s). In Fig. 3, an episode of chatter and the effects derived from it can be seen.

The training and testing datasets were composed of several coils, including chatter and non-chatter episodes. All the signals were windowed into segments of length $N_w=512$ with an overlapping $L = 99\%$ ($n_d = 5$). The FBA was performed over $m=200$ frequencies ($f_i = 1, 2, \ldots, 200$ Hz) and widths $B_1 = B_2 = \cdots = B_{200} =$ 5 Hz. These chosen values are aimed at obtaining a smooth transition of the frequency content of the signals. The total size of the training dataset was $N=4415$ vectors.

The vector quantization using a neural gas algorithm reduced the dataset size to $n_r=600$. For the DR stage, t-SNE was applied with $P=30$. Finally, the RBFN, tuned experimentally, was trained using a kernel width $\sigma = 100$ and a regularization parameter $\lambda = 10^{-9}$.

Here, the number of neural gas units was selected to achieve reasonable computation times. Also, since the objective of the proposed method is to provide a visual analysis, the t-SNE and RBFN parameters were tuned experimentally looking for a trade-off between the smoothness of the DR projection and the capability to separate chatter conditions from normal vibrational states in the resulting visualizations for both training and validations strips. In general, experimentation suggested to use small values of the regularizing factor λ and values of the width factor σ within the same order of magnitude than the data span in the input space.

A summary of the experiments appears in Table 1.

4. Results and discussion

According to the methodology previously described, the projection for a training dataset is computed (see Fig. 4(a)).

The different dynamical conditions are related to the position in the map. Points that are close on the map represent similar dynamic behaviors; color and size represent the average energy in the band of frequencies for the third-octave mode of vibration. The projections show up a trajectory going outside the region spanned by normal condition points, representing clearly the evolution of the chatter fault. The trajectory of the points represents the dynamical evolution of a chatter episode during rolling of a strip.

Another RBFN is used to estimate the expected value of chatter energy based on a dense regular grid on points of the 2D projection space. This allows to build isolines of chatter (contour lines of the average energy in the band of frequencies for the third-octave mode of vibration) that serve as decision boundaries for chatter prediction. As shown in the experiments – see Fig. 5 – all tested coils come out to span a given region in normal condition, then move to a chatter region, then back to normal condition. The region between normal and chatter condition is found to be accurately described by these isolines of chatter, resulting in excellent visual cues for early detection of chatter.

In Fig. 4(b), a glyph is generated from the evolution of the energies of the frequency bands between 100 and 180 Hz. Such representation gives a comprehensive view of the frequency content in which chatter fault appears.

Going into more detail about Fig. 5, several projections from different strips of the process are represented over the training projection, which is displayed in a greyscale color. The projections of the points from a new set of data are placed in zones

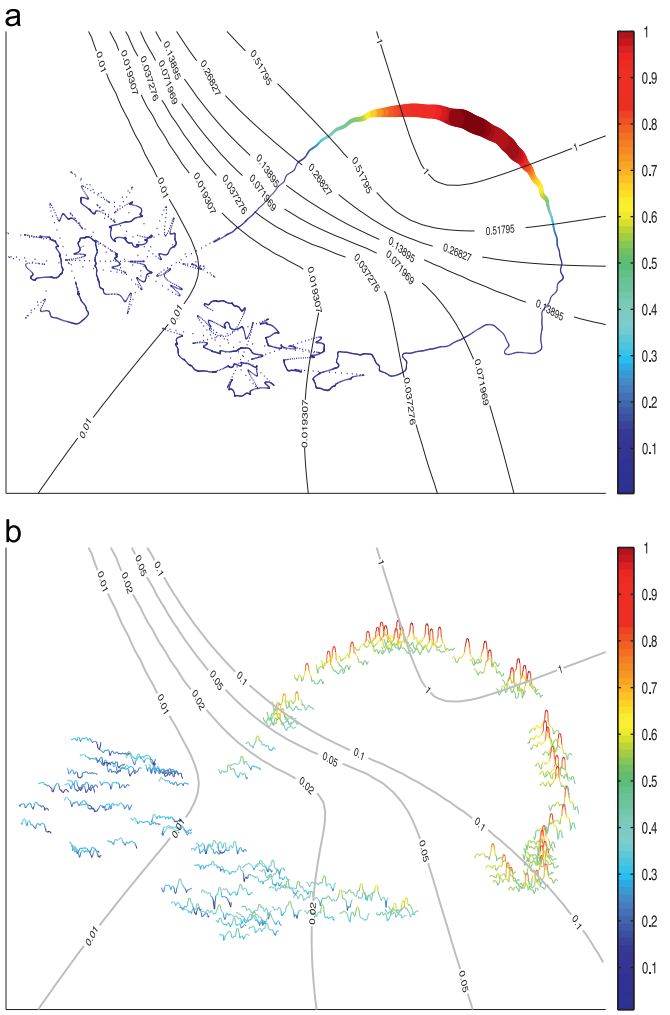


Fig. 4. Resulting map of projection from a training set. Contour lines represent the average energy in the band around the third-octave mode of vibration for each state, in (a) also encoded with color and size. In (b) small plots represent the whole frequency profiles for each state.

Table 1
Parameters of the experiments.

Name	*Symbol*	*Value*
Data acquisition		
Initial sampling rate	F'_m	2000 Hz
Initial sampling time	T'	5×10^{-4} s
Decimation ratio	r	2
Final sampling rate	F_m	1000 Hz
Final sampling time	T	10^{-3} s
Frequency-band analysis (FBA)		
Window size	N_w	512
Overlapping	L	99%
Displacement	n_d	5
# of frequencies	m	200
Center frequencies	f_i	1, 2, 3,…,200 Hz
Bandwidths	B_i	5 Hz
# of points	N	4415
Dimensionality reduction		
# of codebook vectors	n_r	600
Perplexity	P	30
RBFN kernel width	σ	100
RBFN regularization parameter	λ	10^{-9}

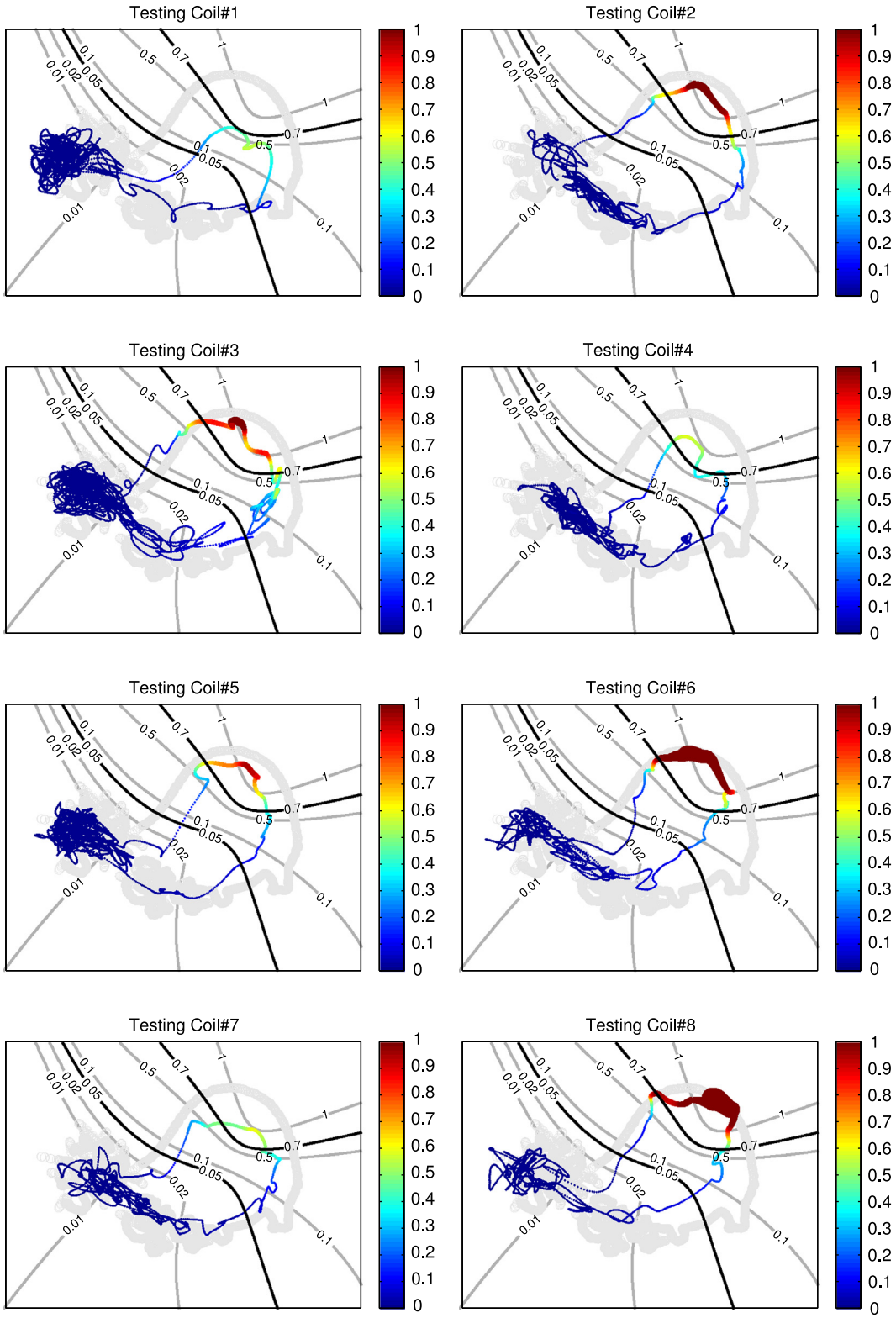


Fig. 5. Projections from several coils over training map.

corresponding to similar dynamical behaviors of the training projections. Therefore, points from a new strip with no chatter episode will stay in the area corresponding to the normal operating condition meanwhile if a chatter fault happens, it will be represented in the chatter area. The established decision boundaries on the map help to detect possible chatter episodes in the production of new strips. For all tests, boundaries between the values of 0.05 and 0.7 can be seen to correspond to a situation where chatter is about to happen – a *pre-chatter* condition – so it can be useful for an early detection of the fault.

5. Conclusions

The so-called chatter fault, which may occur during cold rolling, is studied by analyzing the dynamics of the most relevant variables of the process, i.e. the force applied and the exit thickness. The dynamical behaviors are computed by means of frequency band analysis of the signals, containing significant vibrational states of the process. Being considered as high-dimensionality vectors, a DR technique (t-SNE) that learns the intrinsic structure of the data is applied to project the vibrational states of the process on a low dimensional visualization space.

The experiments are performed using real data from a cold rolling facility. The resulting map for a steel strip reveals zones which represent the different states of the fault, distinguishing clearly between the normal and the chatter operating conditions. The estimation using radial basis functions allows the projection of different sets of data, from several strips, on the same map. In addition, such estimation is about one or two orders of magnitude faster than out-of-sample extensions of the t-SNE. Close projected points represent similar dynamical behaviors indicating if a chatter episode happens and its evolution.

The developed visualization gives insights for the third-octave mode chatter in the process, helping to improve the supervision of this type of severe faults. The displayed contour lines help to detect an early chatter condition and it can help to a possible prediction of the fault.

The proposed methodology can be applied not only to chatter problems in a cold rolling mill, but also to other industrial processes with varying operational behaviors involving (quasi) periodic signals, or even to more general cases, by changing the first feature extraction stage.

Acknowledgments

This work has been financed by a grant from the Government of Asturias, under funds of Science, Technology and Innovation Plan of Asturias (PCTI), and by the Spanish Ministry of Science and Education and FEDER funds under Grants DPI2009-13398-C02-01.

References

Alhoniemi, E., Hollmen, J., Simula, O., Vesanto, J., 1999. Process monitoring and modeling using the self-organizing map. Integrated Comput. Aided Eng. 6, 3–14.

Benbouzid, M.E.H., 2000. A review of induction motors signature analysis as a medium for faults detection. IEEE Trans. Ind. Electron. 47, 984–993.

Bushati, N., Smith, J., Briscoe, J., Watkins, C., 2011. An intuitive graphical visualization technique for the interrogation of transcriptome data. Nucleic Acids Res. 39, 7380–7389.

Chefneux, L., Fischbach, J., Gouzou, J., 1984. Study and industrial control of chatter in cold rolling. Iron Steel Eng. 61, 17–26.

Cottrell, M., Hammer, B., Hasenfuß, A., Villmann, T., 2005. Batch neural gas. In: 5th Workshop on Self-Organizing Maps, pp. 275–282.

Díaz, I., Cuadrado, A.A., Diez, A.B., Domínguez, M., 2011. Manifold learning for visualization of vibrational states of a rotating machine. In: ICANN, vol. 2, pp. 285–292.

Díaz, I., Domínguez, M., Cuadrado, A., Fuertes, J., 2008. A new approach to exploratory analysis of system dynamics using SOM: applications to industrial processes. Exp. Syst. Appl. 34, 2953–2965.

Freshwater, I., 1996. Simplified theories of flat rolling–i. The calculation of roll pressure, roll force and roll torque. Int. J. Mech. Sci. 38, 633–648.

Heidari, A., Forouzan, M.R., 2013. Optimization of cold rolling process parameters in order to increasing rolling speed limited by chatter vibrations. J. Adv. Res. 4, 27–34.

Hinton, G., Roweis, S., 2003. Stochastic neighbor embedding. In: Advances in Neural Information Processing Systems, vol. 15, MIT Press, pp. 833–840.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.

Hu, P.H., Ehmann, K.F., 2000. A dynamic model of the rolling process. Part I: homogeneous model. Int. J. Mach. Tools Manuf. 40, 1–19.

Jamieson, A.R., Giger, M.L., Drukker, K., Li, H., Yuan, Y., Bhooshan, N., 2010. Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. Med. Phys. 37, 339–351.

Jollife, I., 1986. Principal component analysis.

Kimura, Y., Sodani, Y., Nishimura, N., Ikeuchi, N., Mihara, Y., 2003. Analysis of chatter in tandem cold rolling mills. ISIJ Int. 43, 77–84.

Kohonen, T., 1990. The self-organizing map. Proc. IEEE 78, 1464–1480.

Kohonen, T., 2001. Self-Organizing Maps. Springer Series in Information Sciences. Springer, Berlin.

Kramer, M., 1991. Nonlinear principal component analysis using autoassociative neural networks. AIChE J. 37, 233–243.

Kuljanic, E., Sortino, M., Totis, G., 2008. Multisensor approaches for chatter detection in milling. J. Sound Vib. 312, 672–693.

Lee, J., Verleysen, M., 2010. Scale-independent quality criteria for dimensionality reduction. Pattern Recognition Lett. 31, 2248–2257.

Lee, J.A., Verleysen, M., 2007. Nonlinear dimensionality reduction.

Martinetz, T.M., Berkovich, S.G., Schulten, K.J., 1993. Neural-gas' network for vector quantization and its application to time-series prediction. IEEE Trans. Neural Networks 4, 558–569.

Meehan, P.A., 2002. Vibration instability in rolling mills: modeling and experimental results. J. Vib. Acoust. 124, 221–228.

Moody, J., Darken, C., 1989. Fast learning in networks of locally-tuned processing units. Neural Comput. 1, 281–294.

Orowan, E., 1943. The calculation of roll pressure in hot and cold flat rolling. Proc. Inst. Mech. Eng. 150, 140–167.

Paton, D.L., Critchley, S., 1985. Tandem mill vibration: its cause and control. In: Proceedings of the 26th Mechanical Working and Steel Processing Conference, Steel Processing XXII. Iron and Steel Society, Inc., Chicago, IL, USA, pp. 247–255.

Pérez, D., García-Fernández, F.J., Díaz, I., Cuadrado, A.A., Ordonez, D.G., Díez, A.B., Domínguez, M., 2012. Visual analysis of a cold rolling process using data-based modeling. In: Jayne, C., Yue, S., Iliadis, L. (Eds.), Engineering Applications of Neural Networks. Springer, pp. 244–253.

Quintana, G., Ciurana, J., 2011. Chatter in machining processes: a review. Int. J. Mach. Tools Manuf. 51, 363–376.

Roberts, W.L., 1978. Cold Rolling of Steel. Marcel Dekker, Inc., New York.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.

Sammon, J.W., 1969. A nonlinear mapping for data structure analysis. IEEE Trans. Comput. C-18, 401–409.

Schoen, R.R., Habetler, T.G., Kamran, F., Bartheld, R.G., 1995. Motor bearing damage detection using stator current monitoring. IEEE Trans. Ind. Appl. 31, 1224–1279.

Tamiya, T., Furui, K., Iida, H., 1980. Analysis of chattering phenomenon in cold rolling. In: International Conference on Steel Rolling, pp. 1191–1202.

Tansel, I., Wagiman, A., Tziranis, A., 1991. Recognition of chatter with neural networks. Int. J. Mach. Tools Manuf. 31, 539–552.

Tavner, P.J., Penman, J., 1987. Condition Monitoring of Electrical Machines. Research Studies Press Ltd, John Wiley & Sons, Inc.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323.

Tlusty, J., Chandra, G., Critchley, S., Paton, D., 1982. Chatter in cold rolling. CIRP Ann. Manuf. Technol. 31, 195–199.

Torgerson, W., 1952. Multidimensional scaling I: theory and method. Psychometrika 17, 401–419, http://dx.doi.org/10.1007/BF02288916.

Van Der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Venter, R., Abd-Rabbo, A., 1980. Modelling of the rolling process – I: inhomogeneous deformation model. Int. J. Mech. Sci. 22, 83–92.

Young, G., Householder, A., 1938. Discussion of a set of points in terms of their mutual distances. Psychometrika 3, 19–22, http://dx.doi.org/10.1007/BF02287916.

Yun, I.S., Wilson, W.R.D., Ehmann, K.F., 1998. Review of chatter studies in cold rolling. Int. J. Mach. Tools Manuf. 38, 1499–1530.

A.2 INTERACTIVE FEATURE SPACE EXTENSION FOR MULTIDIMENSIONAL DATA PROJECTION [152]

Neurocomputing 150 (2015) 611–626

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Interactive feature space extension for multidimensional data projection

Daniel Pérez [b,*], Leishi Zhang [c], Matthias Schaefer [a], Tobias Schreck [a], Daniel Keim [a], Ignacio Díaz [b]

[a] Data Analysis and Visualization Group, University of Konstanz, Germany
[b] Área de Ingeniería de Sistemas y Automática, University of Oviedo, Spain
[c] Interaction Design Centre, Middlesex University, United Kingdom

ARTICLE INFO

Article history:
Received 30 November 2013
Received in revised form
27 September 2014
Accepted 29 September 2014
Available online 28 October 2014

Keywords:
Feature transformation
Dimensionality reduction
Multidimensional data projection

ABSTRACT

Projecting multi-dimensional data to a lower-dimensional visual display is a commonly used approach for identifying and analyzing patterns in data. Many dimensionality reduction techniques exist for generating visual embeddings, but it is often hard to avoid cluttered projections when the data is large in size and noisy. For many application users who are not machine learning experts, it is difficult to control the process in order to improve the "readability" of the projection and at the same time to understand their quality. In this paper, we propose a simple interactive feature transformation approach that allows the analyst to de-clutter the visualization by gradually transforming the original feature space based on existing class knowledge. By changing a single parameter, the user can easily decide the desired trade-off between structural preservation and the visual quality during the transforming process. The proposed approach integrates semi-interactive feature transformation techniques as well as a variety of quality measures to help analysts generate uncluttered projections and understand their quality.

1. Introduction

Projection-based data analysis (PDA) is a widely used visual analytics approach for identifying and analyzing patterns in multi-dimensional (MD) data. The idea is to map each object in the data as a point to a two or three-dimensional visual display in such a way that similar objects are close to each other and dissimilar ones are further apart. The result is represented in a scatterplot where structures and patterns can be analyzed effectively. The mapping is usually achieved by a dimensionality reduction (DR) technique that approximates the distance (similarity) between objects in the MD data space to the lower-dimensional (LD) projection space. Fig. 1 shows an example of such projection.

A large number of DR methods exist [1,2] for generating projections that preserve the original structure and characteristic of the data. However, when the data is large and noisy, the projection can be cluttered where points and groups overlap each other. The poor visual quality can make it difficult to identify and analyze patterns in the data. This problem originates from the *curse of dimensionality problem* [3]. First of all, distances measures tend to be less meaningful while dimensionality increases, as all objects become similar and dissimilar in many ways, leading to objects being plotted to similar locations in the visual display. Secondly when there is class information involved, those features that are irrelevant to the class labels can obscure the class separation, leading to blurred group boundaries in the projection.

For PDA it is important that the projection not only preserves the data structure but also reveals patterns in the data. When class information is available, a common approach is to take a supervised DR approach that uses class labels to improve group separation in the projection. Available methods include the *Linear Discriminative Analysis* (LDA) [4] that extracts the discriminative features to the class labels and use them to generate embedding, the *Neighborhood Components Analysis* (NCA) [5] that learns a distance metric by finding a linear transformation of input data such that the average classification performance is maximized in the projection space, and the *Maximally Collapsing Metric Learning* (MCML) [6] that aims at learning a distance metric that tries to collapse all objects in the same class to a single point and push objects in other classes far away.

Supervised DR helps improve visual clarity of projections but an uncluttered projection can hardly be guaranteed. On the other hand for explorative analysis, it is important to gain an overview of the data before detailed analysis [7]. A recent work by Schaefer

* Corresponding author. Tel.: +34 985182543.
E-mail address: dperez@isa.uniovi.es (D. Pérez).

http://dx.doi.org/10.1016/j.neucom.2014.09.061
0925-2312/

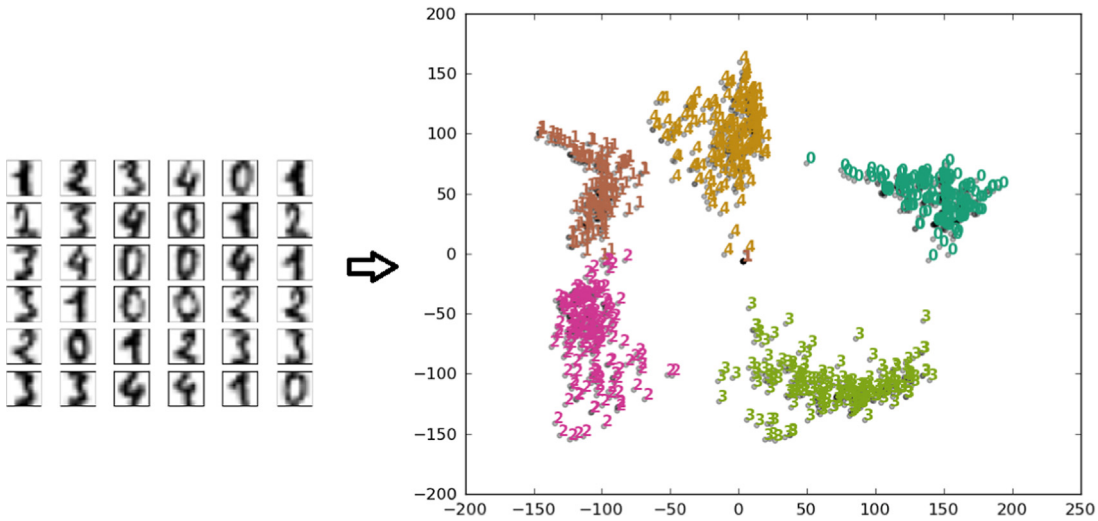


Fig. 1. Images (28×28 D) of hand-written digits projected to a 2D display.

et al. [8] proposed a novel approach that improves the visual quality of the projection by adding class-related features to the original feature space and generating projections based on the extended data. Some promising results were reported. It is not surprising that by feature extension the original structure of the data will be distorted to a certain degree. However, the paper shows that a good compromise between the structural preservation and visual quality can often be made. Moreover, when the data is large and noisy, the method often distorts the structure in a good way such that meaningful patterns obscured by the noise can be revealed especially when the class labels fit to the data structure.

Another issue of PDA is interactivity and transparency. For many application users who are not machine learning (ML) experts, the DR process is often kept in a black-box which makes it difficult to understand and control. Recent advances in solving this problem include (i) the interactive visual DR approaches that integrate the human expertise in the DR process [9,10], (ii) the interactive MD projection system that allows the user to manipulate the control points (subset of sample points) in the visual space based on their knowledge to better organize them as groups [11] and (iii) the interactive feature space transformation approach that allows the analyst to transform existing feature space using different strategies based on their knowledge and understanding about the data [8,12].

In this paper, we present a simple but effective interactive approach that allows the analyst to improve the visual quality of the projection by gradually transforming the original feature space towards clearer group separation in the projection space. This group separation helps in the exploration but it is restricted by the underlying data support. The approach is similar to supervised DR but provides additional user control over the transformation process. The user can adjust the degree of transformation, via a single weighting parameter, and stop at any point where a projection is obtained. The method can be applied on top of any existing supervised DR approach to further improve the group separation. When class labels are not available, clustering results can be used as substitutions to support explorative analysis. In such case, more uncertainty is often introduced, however a series of quality measures are provided to help understand the quality of the projection both in terms of structural preservation and visual clarity. These quality measures provide additional numerical evaluation for the decision of a final projection.

The main contributions of this paper include: (1) a novel and flexible visual analytics approach that combines interactive visualization, feature transformation, and quality evaluation for PDA; (2) a simple but effective feature transformation technique for gradually improving group separation in the projections space; (3) an interactive user interface that provides user control over the transformation process. The remainder of this paper is organized as follows. In Section 2 we discuss related work, in Section 3 we explain the details of the proposed approach, in Section 4 we demonstrate the effectiveness of the method with data by means of a set of experiment results, in Section 5 some characteristics and limitations of the method are discussed and finally, in Section 6 we draw conclusions with an outlook over future work.

2. Related work

The work presented in this paper relates to interactive MD data projection, feature transformation and quality assessment of visual embedding.

2.1. Interactive MD data projection and feature transformation

Classical DR methods estimate the structure of manifolds with a smaller intrinsic dimensionality. When used for generating visual embedding of MD data, the result can be unsatisfactory, especially when the dimensionality is high and the data contains noise. Firstly, the projection space is limited to 2D or 3D. Secondly, by its nature the reduction causes information loss and it is often difficult for the algorithms to determine which information is less relevant to the analysis tasks. In [13] the importance of integrating interactions with statistic methods (in particular, DR techniques) to support exploratory analysis of MD data is discussed. By interactive analysis, the analyst can better steer the DR process by incorporating their domain knowledge and analytical skills for generating better projections.

In recent years, the idea of interactive projection has been widely adopted. For example, a semi-supervised approach is proposed in [14] for projecting MD data. In [15] interactive projection techniques are developed to allow the analyst to integrate their knowledge about the data to the DR process. The *iPCA* [9] is proposed to provide coordinated views for interactive analysis of projections computed by PCA. In [10] the *iVisClassifier*

system that integrates supervised DR technique LDA with interactivity is developed. The analysis of DR techniques with interactive controls was also proposed in [16] and the *DimStiller* framework [17] where the user is guided during the analysis process by means of workflows.

An effective approach to improve the visual quality of the projection is feature transformation. Given grouping information such as class labels or natural groups (clusters) in the data, the analysts may want to improve the visual quality of the projection gradually so that detailed analysis can be carried out. This can be achieved by pulling group members closer to each other in the projection and pushing non-group members further apart in the projection space. In theory such a task can be fulfilled by supervised DR, however, as discussed in the previous section the fully automatic approach lacks user control and transparency. Schaefer's approach [8] improves the existing solution by allowing the analyst to extend certain features in the data based on grouping information and to add the extended features to the original feature space for generating better quality projections. The result shows that a good compromise can often be made between structural preservation and visual clarification. In [11,18] another user-driven feature transformation approach, the *Local Affine Multidimensional Projection* (*LAMP*) is proposed and implemented. *LAMP* allows the user to modify the point locations in the visual display and use the modification as feedback to update the original feature space in order to achieve better visual quality. The approach provides easy user control over the projection process and does not require much ML knowledge. However when the location of multiple points are modified in the visual display, the method may encounter heavy computation load while updating local neighborhood diagrams of multiple control points. Another interesting approach called *Dis-Function* was proposed by Brown et al. [19] which displays the projection on an interactive visual display such that the analyst can move points around to modify the distance between objects based on their own knowledge. The modification on the visual space is then used to update the distance function and recompute the distance measure. Such approach integrates new knowledge to the data which is similar to our approach, except that Dis-Function requires some prior knowledge of distance between objects, and our approach is meant for using existing grouping information.

In addition to the above mentioned work, a comparison of feature sets can be found in [20], where an interactive exploration can be made for the selection of suitable data descriptors. A related problem was addressed in [21] where dendrogram structures were extracted from alternative feature sets, and applied for interactive comparison and selection of feature sets. These interactive methods demonstrate the possibility of improving PDA by incorporating user knowledge and feedback. However interactive MD data projection remains a challenge as many of the existing methods are either dependent on a particular DR technique, or rely on a good understanding of the applied DR techniques.

2.2. Quality metrics

Despite the large number of DR techniques that have been developed, the question of quality assessment of a given projection has only been studied in several cases and systematized in recent years [22,23].

The first measures introduced to assess the quality of a projection were the so-called *stress* and *strain* measures [24,25]. These measures assess the quality of structural preservation by computing the differences of the pairwise distances between objects in the LD embedding and the corresponding distances in high-dimensional (HD) data space. They come from objective functions of a family of DR techniques such as multidimensional scaling (MDS) so that errors can be evaluated at the end of the minimization of the function. For example, one of the most commonly used *Sammon's stress* refers to the final value of the error function in Sammon's projection algorithm is as follows:

$$E = \frac{1}{\sum_{i<j} d_{ij}^{*}} \sum_{i<j} \frac{(d_{ij}^{*} - d_{ij})^2}{d_{ij}^{*}} \tag{1}$$

where d_{ij}^{*} is the distance between two points i and j in the HD data space and d_{ij} is the distance between the corresponding points in the LD projection space.

While *strain* and *stress* measures analyze the preservation of data structure based on differences of distances, several measures like *trustworthiness* and *continuity* [26] and the *K-ary neighborhoods* measure [27] assess the quality of a projection in a broader applicability, taking into consideration also neighborhood preservation using rank-based criteria.

For example, the *K-ary* measure is defined as

$$Q_{\mathrm{NX}}(K) = \sum_{i=1}^{N} \frac{|n_i^K \cap v_i^K|}{KN}, \tag{2}$$

where n_i^K and v_i^K are the K nearest neighbors of the point i in HD and LD spaces, respectively. It is usually displayed as a line for the different values of K from 0 to $N-1$, here the average of these values Q_{avg} is considered in order to summarize the overall quality in a number between 0 and 1, where the higher value indicates better projection. Beside, when the data is labelled, the classification error is a typical choice, see for instance [28] and other references in [29]. The integration of classification error measures in the DR technique leads to better group separation in the final embedding.

Apart from the structural preservation quality measures mentioned above, a set of visual quality measures has also been developed. Examples include *Histogram Density Measure* that ranks scatter plot visualizations of multidimensional data, the *Class Density Measure* that assess class separation of a given projection, both proposed in [30], and class consistency measures [31]. Moreover, the *overlap measures*, defined in [8], compute the overlap area between groups and overlap object density in a multidimensional data projection. The overlap area sums the area of all the overlap regions $intersect(i,j)$ between pairwise groups for the set g of groups:

$$ov_{reg} = \sum_{i=1}^{|g|-1} \sum_{j=i+1}^{|g|} intersect(i,j) \tag{3}$$

The overlap regions are computed from the definition of the region of each group described by using the concave hull of the objects of each group proposed in [32]. The overlap density takes into account the density of the points over-plotted in the visual display. The visual display is divided into grids units where the occupation of a specific class is determined by using Gaussian functions G in the function f defined as follows:

$$f(G_{ip}, G_{jp}) = \begin{cases} 1 & \text{if } G_{ip} > 0 \text{ and } G_{jp} > 0 \\ 0 & \text{else} \end{cases} \tag{4}$$

Thus, f is activated by 1 in the case where a grid square unit is occupied by two classes with the Gaussian model. The sum of these grids found for pairwise classes gives the overlap density measure, defined for K classes and an image of P pixels as

$$ov_{density} = \sum_{i=1}^{|K|-1} \sum_{j=i+1}^{K} \sum_{p=1}^{P} f(G_{ip}, G_{jp}) \tag{5}$$

In the next examples, the grid resolution is uniformly set to 3 pixels and σ value of the Gaussian model to 12, so that different experiment results can be compared.

3. Interactive feature extension

In this paper, we propose an analysis framework that combines the transformation of the feature space, the interactive parameter setting and visualization to help analysts achieve a better interpretation of projection results. Given a MD dataset, available grouping information is used to generate an extended feature space in such a way that the class knowledge is introduced in the extended feature space. The analyst can select certain attributes or the whole set for feature extension based on their knowledge and modify the projection gradually in order to achieve a good visual embedding. The quality of the projections will be evaluated using various quality measures. The process can be repeated iteratively until a satisfactory projection is achieved. Fig. 2 shows the flowchart of the proposed method.

3.1. Feature space extension

The basic idea of the feature space transformation is to extend certain features based on available grouping information. Consider a MD dataset as a matrix $\mathbf{X}$ where rows are data items and columns are features, and the labels $\mathbf{y}$ are given to the class corresponding to the i-th row.

$$\mathbf{X} = [x_{ij}] \in \mathbb{R}^{n\times d}, \quad \mathbf{y} = [y_i] \in \mathbb{N}^n \tag{6}$$

With $i = 1,\ldots,n$ and $j = 1,\ldots,d$, where n is the number of feature vectors and d the number of dimensions. If m features are selected $f = f_1,\ldots,f_m$, the extended data matrix $\mathbf{X}'$ is defined as follows:

$$\mathbf{X}' = [x_{ij}|\tilde{x}_{ij}] \in \mathbb{R}^{n\times(d+m)} \tag{7}$$

where $\tilde{x}_{ij}$ is the statistical value corresponding to the class label y_i in the feature f_j. Here, we use the arithmetic mean within the class members on a particular dimension. For example suppose we have a dataset with 2 attributes, 4 records that belong to 2 classes are as shown below:

Sepal.Length	Sepal.Width	Species
5.1	3.5	*setosa*
4.9	3.0	*setosa*
7.0	3.2	*versicolor*
6.4	3.2	*versicolor*

The class labels are used to compute the mean values for each class in each dimension:

	Sepal.Length	Sepal.Width
$mean_{setosa}$	5.0	3.25
$mean_{versicolor}$	6.7	3.2

The $\mathbf{X}'$ matrix can be built as an extension of the original data $\mathbf{X}$ as follows:

Original		*Extended*		Species
dim1	dim2	ext1	ext2	
5.1	3.5	5.0	3.25	*setosa*
4.9	3.0	5.0	3.25	*setosa*
7.0	3.2	6.7	3.2	*versicolor*
6.2	3.0	6.7	3.2	*versicolor*

Both original and extended space will be combined together to decide the distance metric for DR. Although we use class as an example statistical value for $\tilde{x}_{ij}$ in the above example, it should be noted that $\tilde{x}_{ij}$ can be many other statistical values such as median or other form of averages. An effective approach would be to decide which statistical values are to be used for each dimension based on the data distribution. Detailed discussion relating to this issue can be found in [8]. For all the experiments in this paper we extend mean values based on class labels for simple illustration purpose.

3.2. Weighted extension of the feature space

Having the data matrix $\mathbf{X}$ and labels $\mathbf{y}$ (see Eq. (6)), as explained above, the extended data matrix $\mathbf{X}'$ is defined by the original matrix $\mathbf{X}$ and the extended part $\tilde{\mathbf{X}}$ as follows:

$$\mathbf{X}' = [\mathbf{X}|\tilde{\mathbf{X}}] \tag{8}$$

Assuming the extension of the whole set of features and using mean values of each class labels, $\mathbf{X}' \in \mathbb{R}^{n\times 2d}$. In this case, $\tilde{\mathbf{X}}$ is composed by the centroids of the corresponding class described by the labels:

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_i] \in \mathbb{R}^{n\times d} \quad \text{being } \tilde{\mathbf{x}}_i = \frac{1}{|C_{y_i}|}\sum_{i\in C_{y_i}} x_{ij} \tag{9}$$

where C_{y_i} is the set of indices of samples belonging to class y_i.

A real parameter $\lambda \in [0,1]$ allows the gradual transition between original data ($\mathbf{X}$) and the extended part ($\tilde{\mathbf{X}}$) by applying a simple change in the metrics of the extended feature space given by $\mathbf{X}_{weight} = \mathbf{X}'\mathbf{W}_\lambda$, being the matrix $\mathbf{W}_\lambda \in \mathbb{R}^{2d\times 2d}$ as follows:

$$\mathbf{W}_\lambda = \begin{pmatrix} (1-\lambda)\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \lambda\mathbf{I} \end{pmatrix}, \quad \lambda \in \mathbb{R} \tag{10}$$

Therefore, $\mathbf{X}_{weight}$ is the weighted data matrix used for computing low-dimensional embedding. The parameter λ can be changed interactively so that the user can trade between inter-class and intra-class topological organization of data. In this way, with $\lambda = 0$ the projection reveals the structure to the original dataset and

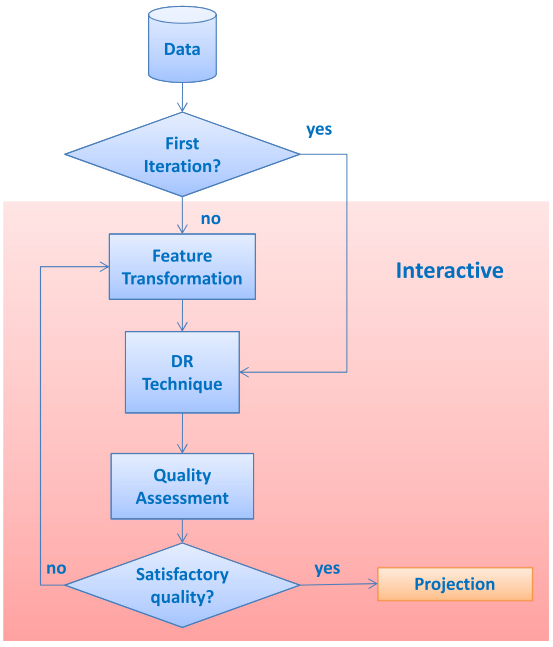


Fig. 2. Workflow of the method.

with $\lambda=1$ only to the applied extension. Thus, a good starting point for this analysis is the weighted extension of the whole feature space so that the analyst can change easily the embedding or return to the original. This is achieved only by interacting with λ parameter to obtain a more meaningful projection, assessed both visually and by quality measures.

Note that our proposed method is independent of the DR technique that computes the projection, hence it inherits the same level of computational complexity of the applied DR technique. However, various scalability approaches that involve sub-sampling and approximation have been made towards handling large data. For example, Li et al. [33] proposed a scalable scheme that improves the efficiency of the *Singular Value Decomposition* (SVD) process by first sampling a subset of columns from the input matrix and then approximate SVD on the inner sub-matrix using matrix approximation algorithms. Yang et al. [34] proposed an optimization approach that reduces the computational cost of *Neighbor Embedding* methods by computing close-by points individually but approximating far-away points by their center of mass. Bunte et al. [35] proposed a relevance learning approach that incorporates prior knowledge of the data such that the computational cost can be saved by reducing the number of adaptive parameters. Our proposed method can be used in conjunction with these methods to achieve better scalability.

3.3. An illustrative example

Here a very simple example is presented to illustrate the proposed method. The data consists of two Gaussian clusters of 150 points each with a small overlap in 2 dimensions. The method is applied to the data following the weighted extension of the feature space using the mean values for each class corresponding to each cluster. The DR technique to compute the projections is *PCA*. In Fig. 3 the resulting projections are represented for several values of the parameter λ. The projection for $\lambda=0$ corresponds to the original data where the two clusters are not fully revealed. As the λ parameter increases, the projection changes revealing the grouping information. Since the distances inside of each cluster are not modified, the local structure in each cluster is preserved for the new projections. For the highest value ($\lambda=1$) the projection is purely based on the grouping information (mean value of each cluster), therefore all the data points are pushed to the centroids of corresponding clusters. Therefore this can be considered as a representation of the classes distribution. For a correct interpretation of the original data structure the projections for high values of λ are not useful and can be neglected. The interaction by means of the λ parameter provides control to the user and improves the exploration tasks. Moreover, a numerical evaluation of the projection gives more information to the user in order to judge the optimum point of the transformation, this is explained in more detail in Section 4.5.

4. Experiments and results

In this section we evaluate the proposed method with different datasets and use cases. The datasets are selected representing data of various dimensionality, number of classes, synthetic and real (see Table 1). Four use cases are designed to test the method from different perspectives, including:

c1: Synthetic vs. real data – the first use case applies the method on two synthetic examples, the remaining use cases are applied on real data.
c2: Time series data – this use case shows an example of improving visual clarification of projections for analyzing patterns in time series data.

Table 1
Description of data.

Name	Size	Dimensions	Classes
3D clusters	500	3	5
synthetic-gaussian	500	10	5
eCons (*weekday*)	338	24	7
eCons (*month*)	338	24	12
hiv	78	159	6
yeast	1452	7	10

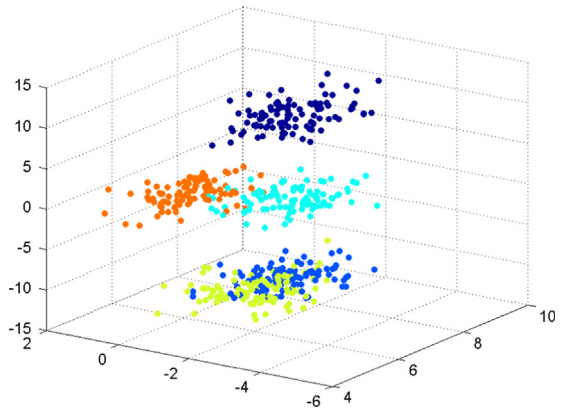


Fig. 4. Representation of *3D clusters* data example. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

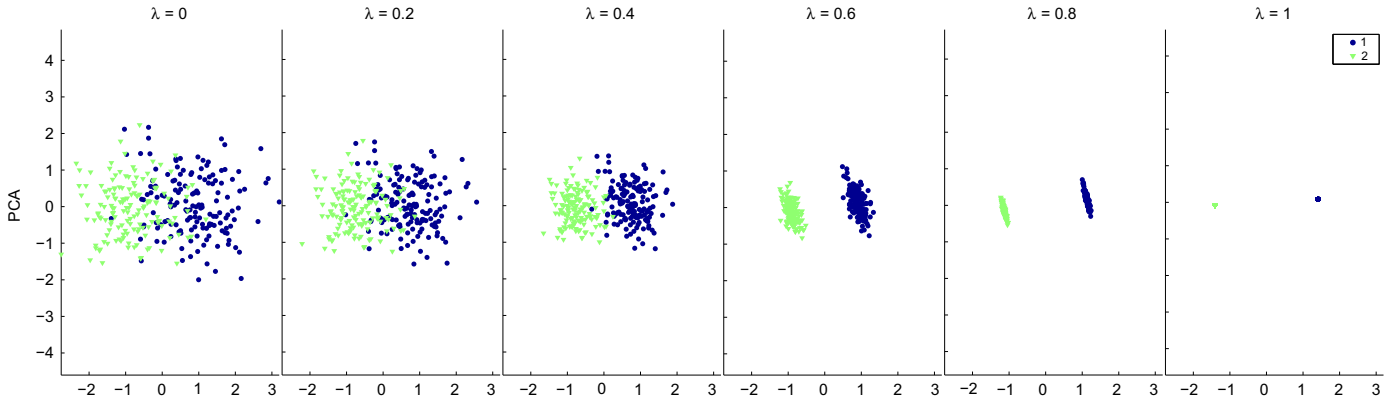


Fig. 3. Projections from the proposed method applied to 2D clusters data example using PCA.

c3: Extending full feature space vs. selected features – the third use case demonstrates the potential of improving the effectiveness of the approach by extending only a subset of the features based on knowledge and understanding about the data.

c4: Supervised vs. unsupervised DR – the last use case applies the method on two supervised DR methods, while the other use cases apply unsupervised DR techniques.

All the experiments start with an original projection generated by a standard DR technique with λ value set to 0. For unsupervised DR we apply *PCA* and *t-SNE* that are widely used by the visualization community for explorative data analysis. For supervised DR we choose two recent advances including *NCA* and *MCML* as introduced in Section 1. The original feature space is extended using the weighted extension strategy as described in Section 3. All the projections were computed using Matlab implementations of DR algorithms from the toolbox [2] or the original authors. The projections of the original and extended feature space are computed using the same DR technique with the same parameter setting, after a z-score normalization. Where the *t*-SNE technique is applied and the new projection requires the perplexity parameter to be updated, we regenerate a new projection using the new parameter setting to replace the original projection for comparison. Since the performance of *t*-SNE is quite robust in terms of variation on perplexity values, such updates does not usually change the original projection to a great extent. Mean and standard deviation of the quality measures are computed after 10 iterations for this technique. In order to make more comparable projections, a linear transformation determined by *procrustes analysis* [36] is performed between projections.

Next we illustrate the results of the experiments. The evaluation of the projections of these experiments are discussed in Section 4.5.

4.1. Synthetic examples

4.1.1. 3D clusters example

To illustrate the idea conceptually, we apply the proposed method in a synthetic dataset that contains 3 dimensions. The data consists of 5 Gaussian clusters each containing 100 samples. Fig. 4 shows the original structure of the data in a three

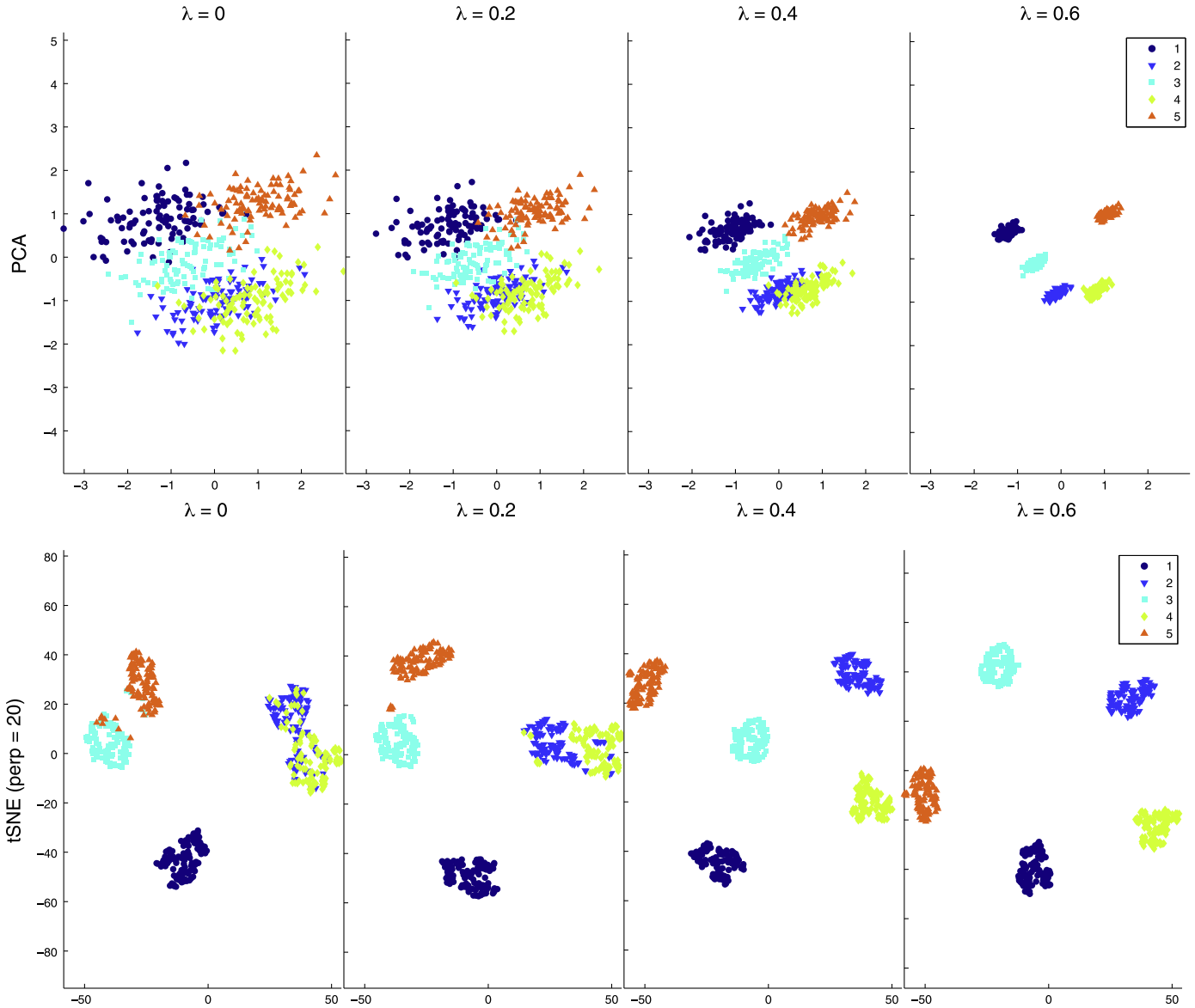


Fig. 5. Projections of *3D clusters* with weighted extension for several λ values using cluster information, computed by PCA (top) and *t*-SNE (bottom) with a perplexity of 20.

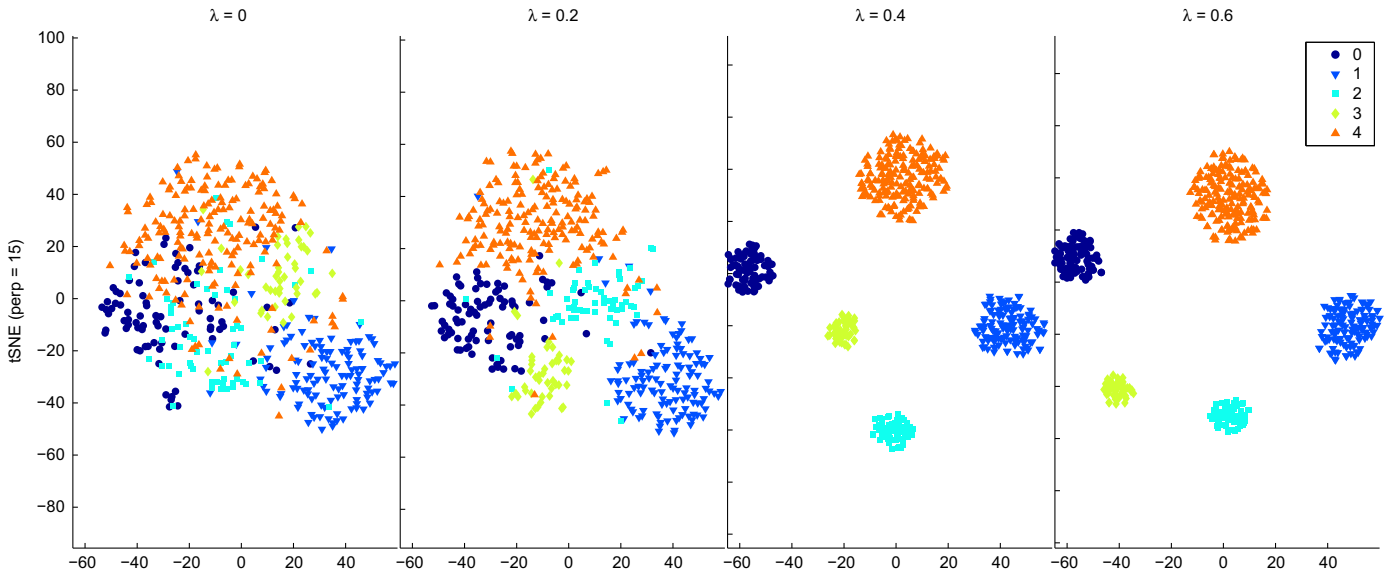


Fig. 6. *t*-SNE projections of *synthetic-gaussian* dataset with weighted extension for several λ values using cluster information with a perplexity of 15. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

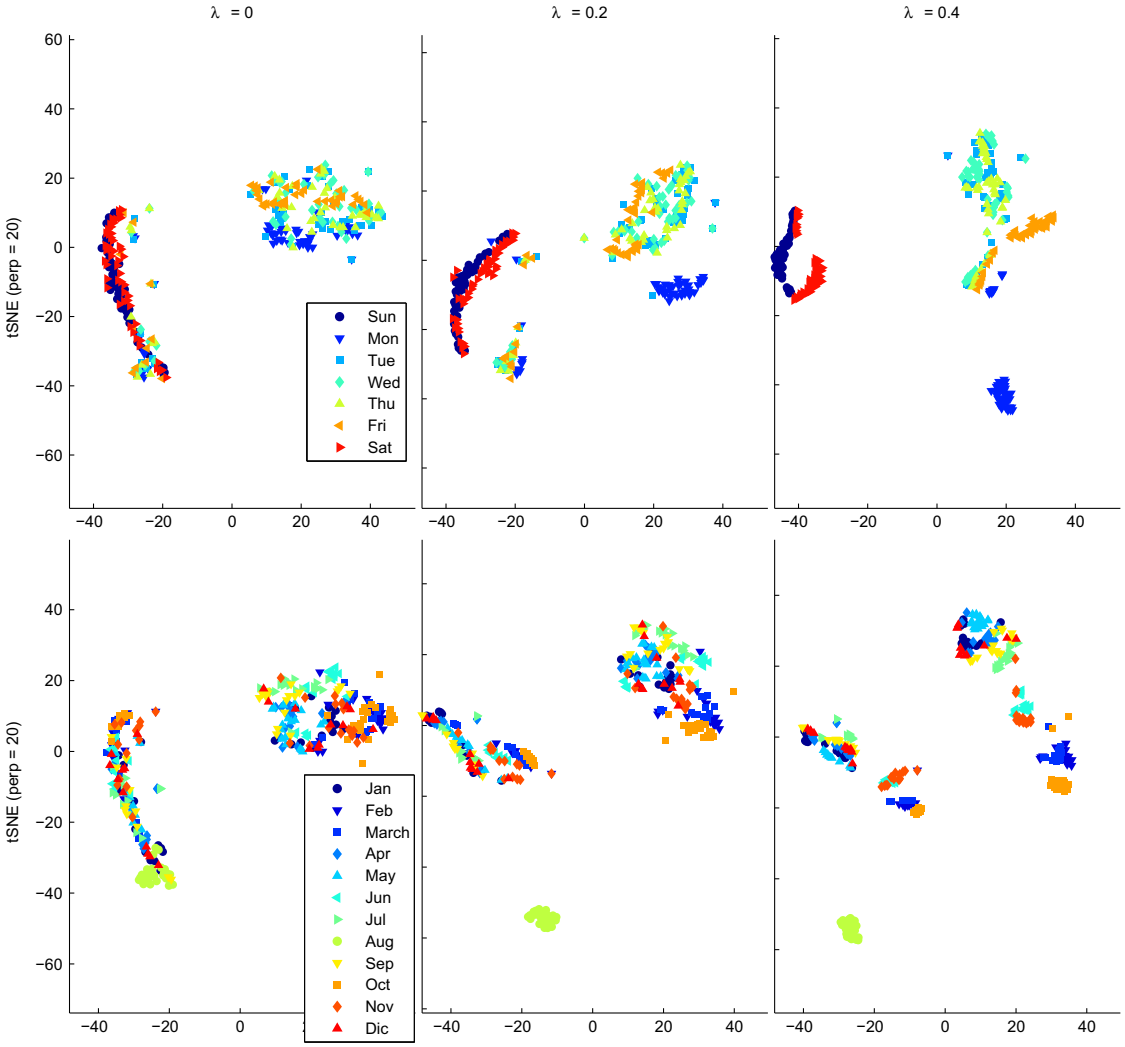


Fig. 7. *t*-SNE projections for *eCons* dataset with a weighted extension applied for several values of λ using weekday (top) and month (bottom) labels. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

dimensional coordinate system. As shown in the figure, the top three clusters (colored in navy, cyan and orange) are well separated, but the two clusters at the bottom of the display (colored in blue and green) overlap against each other.

Based on the cluster labels a series of weighted feature extension were added to the original data, with the weight (λ parameter) set to several values. Fig. 5 shows two dimensional projections of the transformed data generated using PCA (upper) and *t*-SNE (lower) technique with a perplexity value of 20.

As as one can see from the figure, the process of transforming data by integrating group information modifies the original data space in such a way that groups are better separated in the projection. In this particular example, when the values of λ is between 0.4 and 0.6, both DR techniques generate projections with clear group separation. Naturally when λ value is 0, the projection is purely based on the original data, hence in terms of the preservation of the original data structure, there is no difference between the proposed method and the standard DR technique that is applied for generating the projection. Furthermore, it can be considered as an initial reference view to compare the new projections obtained by the method.

4.1.2. *Synthetic Gaussian example*

In this experiment a synthetic dataset is used to evaluate the proposed method in a simple scenario. The data consists of 5 random Gaussian clusters of 10 dimensions and is generated from the work in [37]. Fig. 6 shows the resulting projections computed using *t*-SNE technique with a perplexity value of 15. Different colors and markers are used to distinguish the 5 different

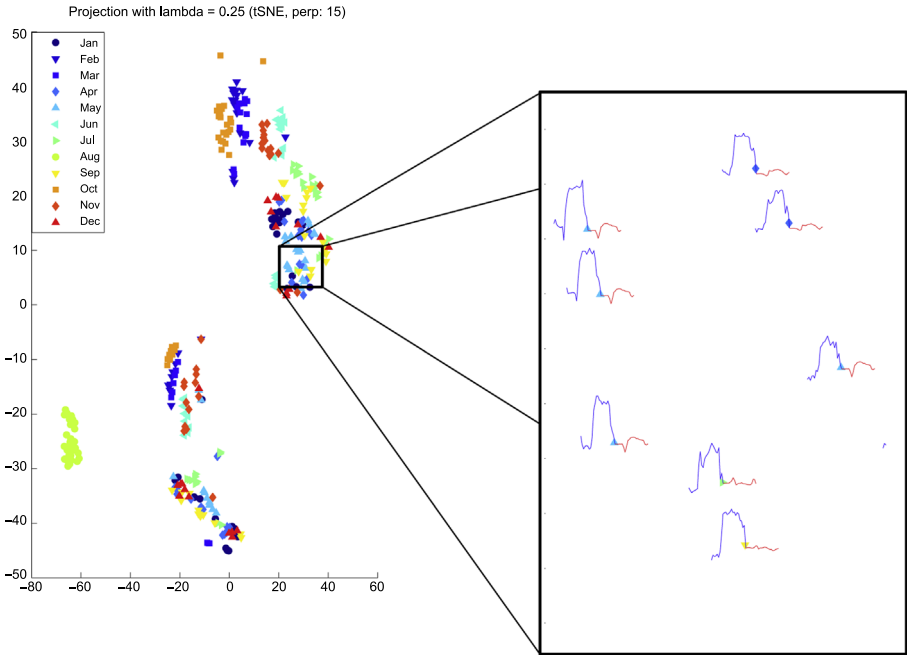


Fig. 8. Projection with extended data by month (left) and zoomed representation of the projection (right), showing the features of each item as a sparkline, original (blue) and extended (red). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

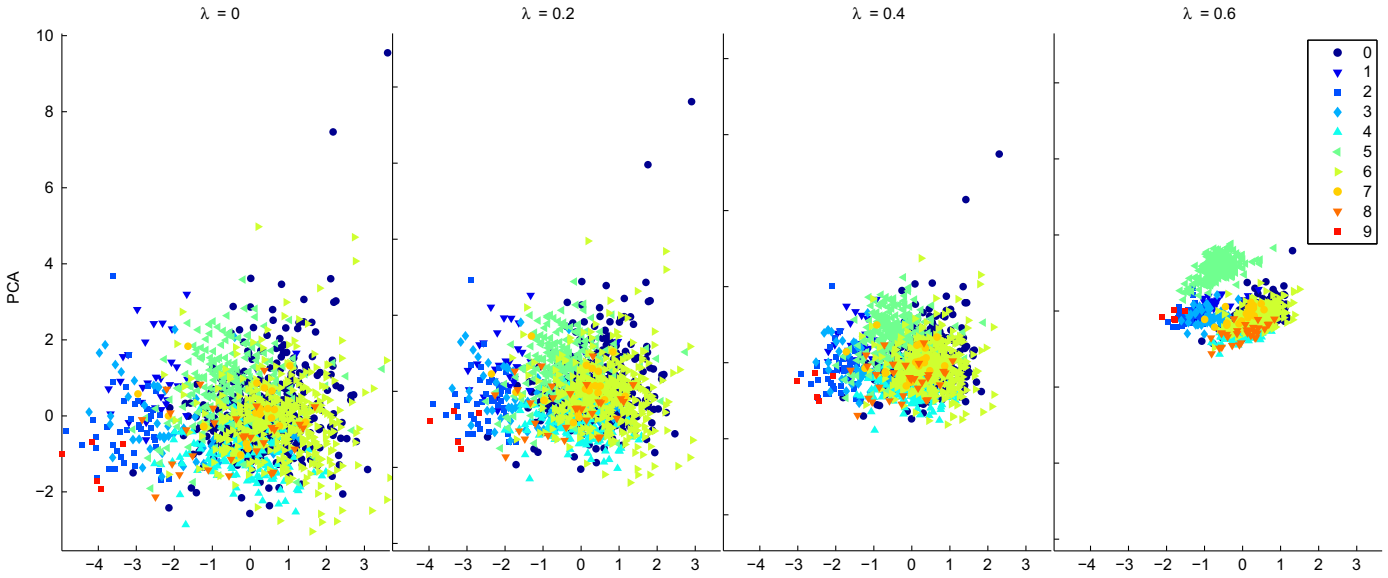


Fig. 9. PCA projections for *yeast* dataset with a weighted extension of feature 4 for several values of λ using classes information. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

clusters. As one can see, compared to the projection of the original data, the projections of transformed data (with λ values between 0.2 and 0.4) provide a clearer separation of clusters. Even without color coding, it would not be difficult for the analyst to identify the patterns revealed in the projection.

Similar projections can be obtained using *t*-SNE for high values of λ. This is due to the fact that *t*-SNE aims at preserving both the local and global structures, where the importance of modelling the separations of datapoints is almost independent of the magnitudes of those separations [38].

4.2. Time series data

In the second use case we use the *eCons* data that records the active power usage at a university building over a year. The dataset is aggregated to 338 days (samples with missing values removed) and 24 attributes (value for each hour per day). The task is to identify different types of daily consumption patterns. In this experiment different types of temporal information, such as weekday or month, are incorporated into the resulting projection. The analysis of daily consumption patterns was addressed previously, in [39], where a calendar view combined with cluster information is used for an effective exploration of time series data. While the calendar view provides a good platform for univariate time series analysis, in this case our approach is designed more towards projecting multivariate data with the flexibility of adjusting time intervals and selection of class knowledge.

Two types of class labels are considered: classes corresponding to the type of the weekday (1-Sun; 2-Mon; … 6-Fri; 7-Sat); and the corresponding month (1-January; … 12-December), respectively.

First we analyze the data based on days of the week. The projections are computed using *t*-SNE with perplexity value set to 20. Different colors are assigned to different days of the week. In the projection of the original data for $\lambda=0$, one can easily see two distinct groups (see Fig. 7, top). The result can be interpreted as "working days" and "non-working days". However in the embedding of extended data for $\lambda=0.2$ (see Fig. 7, top), we see more interesting patterns, for example, most of the Mondays (blue triangle) appear to be in a separate cluster. Furthermore, the "non-working days" cluster splits into "weekends" and "bank holidays" clusters.

At the bottom of Fig. 7 an analogous analysis of the same dataset is shown. In this figure, colors are used to differentiate which months does a date belong to. The projections are again generated using *t*-SNE and the perplexity value is set to 20. The projection of the original data (left) shows two distinct groups (high- and low-consumption clusters) as in the previous case. Each group with a mixture of dates that belong to different months. However the projection of the extended data ($\lambda=0.4$) further separates August dates from the rest of the points revealing a remarkable behavior inside the low-consumption cluster. This could be explained by the university holiday period throughout August. In addition, months such as February, March and October show different consumption patterns from the rest.

Note that the new projections modify the location of the points from original clusters taking into account the information that the user incorporates. In this case, the same initial projection, showing two main clusters of daily consumption, is modified by two different criteria (weekday and month), that divide these groups revealing the introduced information hierarchically with respect to the original.

Fig. 8 shows a part of a similar projection where the values of the features are plotted as a *sparkline* over each item, the original values (in blue) and the extended values (in red). This allows an easy comparison between similarities of the projected points. Points inside a class are topologically organized by intra-class similarities, which are given by the original features (blue). The inter-class organization between classes varies depending on the extended values (red) according to the value of λ, whose highest value (set to 1) corresponds to the pure projection of the centroids of the classes.

4.3. Extension of selected features

In this experiment we investigated the effect of simple extensions based on selected features using the *yeast* dataset [40] which is commonly used by ML and the visualization community as a benchmark dataset. The main task is to predict the localization site of proteins. Given the class labels, one thing the analyst can do is to study the distribution of the data values over different dimensions and detect discriminative features. This can often be achieved by examining visual representations of the distributions such as box plots or parallel coordinates. Furthermore other strategies of feature selection [41] can be used in cases where a multidimensional visualization cannot be performed, such as relevance learning [35] or even domain knowledge of the user. In the current example, it is observed that objects in class 5 tend to have high values in dimension 4. This leads to our next experiment to extend only one feature – class mean of dimension 4 – to see if the extended feature space leads to better visual quality.

Fig. 9 shows the resulting projections of the extended feature space using PCA. The left figure of $\lambda=0$ is the projection of the original data. The rest projections are based on the weighted extension of mean values of dimension 4 over different classes. As one can see, overall the projection for $\lambda=0.6$ is less cluttered. In particular, class 5 (in green) is much better separated from the rest of the classes.

The selection of a suitable λ value is made not only using a visual interpretation of the projection by the user, but also its evaluation performed by quality measures. The selection of this parameter takes into account the visual improvements of the projection whilst generally preserving the structure. In Fig. 10 the evolution of the measures used here can be seen for different values of λ. As the average of *k*-ary measure (Q_{avg}) equal to 1 means a perfect embedding, $1-Q_{avg}$ is taken in order to show all lines with similar trends, i.e. the lower the value the better. With the evolution of λ, a remarkable improvement of visual measures (overlap area and density) and slightly worse structural measures (stress and *k*-ary) can be seen. In this case, the analyst may want to select $\lambda=0.6$ because it is the lowest λ parameter

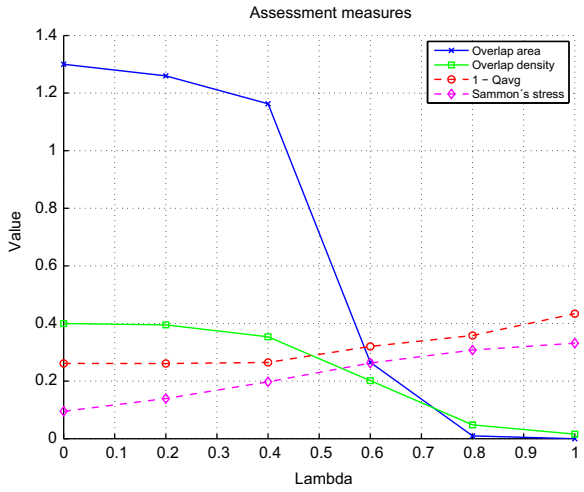


Fig. 10. Quality measures of PCA projections of *yeast* data using weighted extension of the feature 4 for several values of λ.

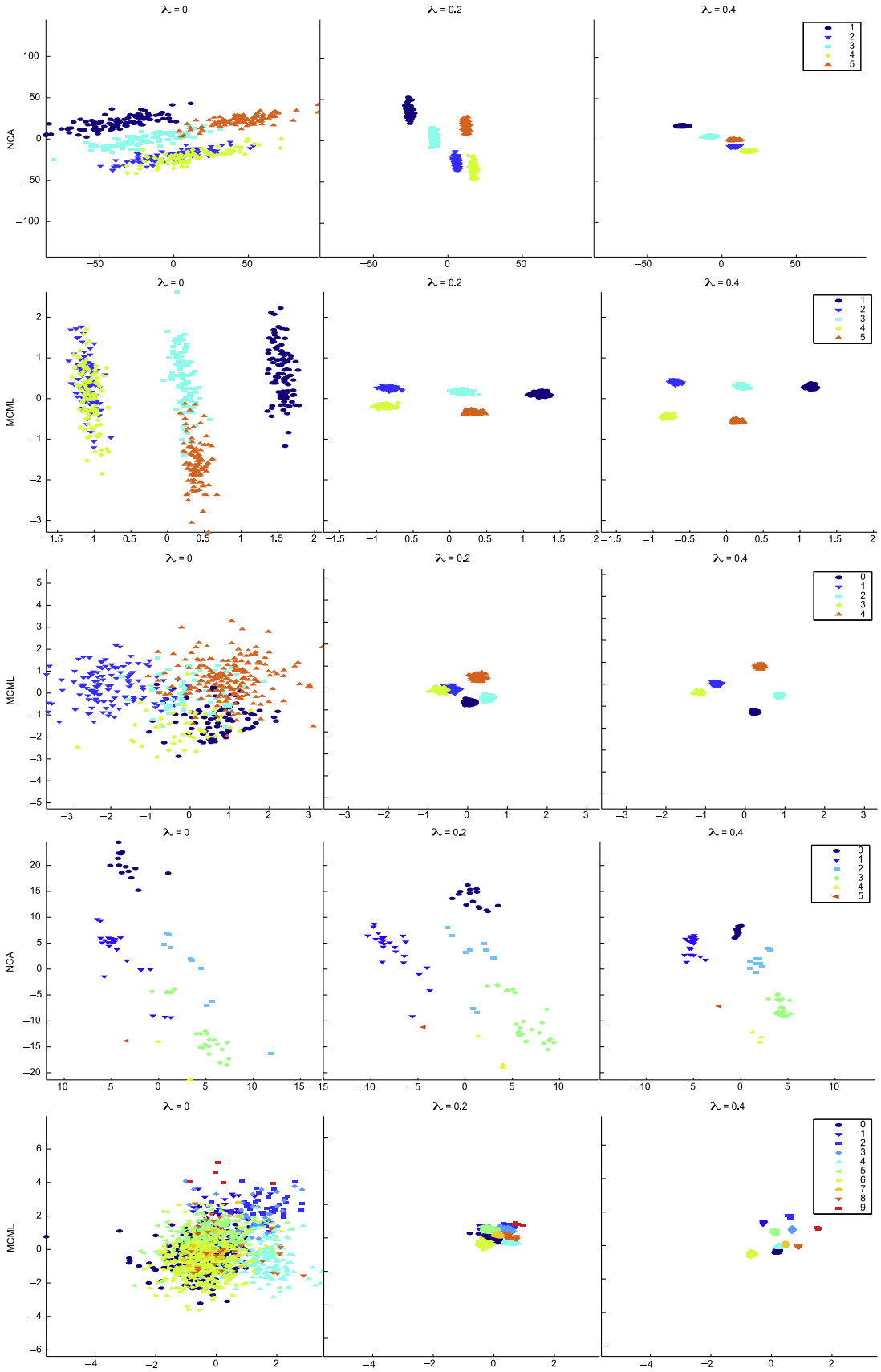


Fig. 11. Projections with weighted extension for several λ values using labels. For *3D clusters* using NCA (first from the top) and MCML (second); for *synthetic-gaussian* using MCML (third); for *hiv* using NCA (fourth); and for *yeast* dataset using MCML (bottom).

value that provides visual enhancements with small variations for the rest of the measures.

4.4. Supervised DR methods

In the last experiment we applied two supervised DR techniques, *NCA* and *MCML*, on some of the selected datasets. The same class information is used for the projection of both the original data and data with weighted extensions.

3D clusters example: The method is applied to 3D clusters example from Section 4.1.1, the resulting projections with several values of λ are displayed in Fig. 11 for both techniques *NCA* (first from the top) and *MCML* (second). The regularization parameter of *NCA* is set to 0. As it can be seen in the figure, the weighted extension emphasize the class separation using both DR techniques in a similar way.

Synthetic-gaussian dataset: Fig. 11 (third from the top) shows the projections generated using *MCML*. As one can see, even with a supervised DR method, the original data projection can be still quite cluttered ($\lambda=0$). By transforming the original feature space with weighted extension (in this case, $\lambda=0.4$), the group visual quality can be improved substantially.

hiv: The *hiv* dataset, which was used in [31], describes socio-economic properties of countries that are classified into HIV risk groups. The data has 159 attributes and contains objects that belong to 6 different classes. We project the data using *NCA*, in this case its regularization parameter is set to 0.001. The resulting projections are shown in the fourth projections of Fig. 11. Again, although the groups are well separated in the original projection, the projection with $\lambda=0.2$ enhances the inter-group separation.

Yeast: Fig. 11 (bottom) shows the projection of this dataset using *MCML*. While the projection based on the original dataset is rather crowded and one can hardly see any patterns, the projection of extended feature space ($\lambda=0.4$) provides a much clearer view of the grouping information in the data.

4.5. Evaluation of embeddings

The performance of the projections is evaluated by arithmetic measures, described in Section 2.2. In this paper we select four measures, including the *Sammon's stress* [24] and *k-ary* [27] measure for assessing the structural preservation, and the *overlapping density* and *overlapping area* measures [8] for assessing the visual clarification. Although other approaches that represent the structural preservation and distortions could also be used for analyzing the quality of the result projections [42,43].

The measures were computed for projections obtained using several values of λ. They are represented in line charts, similar to Fig. 10, where lower values imply improvements in the measures. Figs. 12–15 graphically show the measures for *synthetic-gaussian*, *hiv*, *eCons (months)*, and *yeast* datasets, respectively. The measures for *PCA* and *t-SNE* techniques are shown at the top and similarly for *NCA* and *MCML* methods at the bottom, respectively. Out of the range values were scaled in order to an effective comparison.

The result shows that in general extending feature space using class related statistical values leads to better visual quality in the final projection. Less overlapping points (reduced overlapping density measure), and group boundaries are overlapped less (reduced overlapping area measure). The data structure is less well maintained in the new projection, especially the global

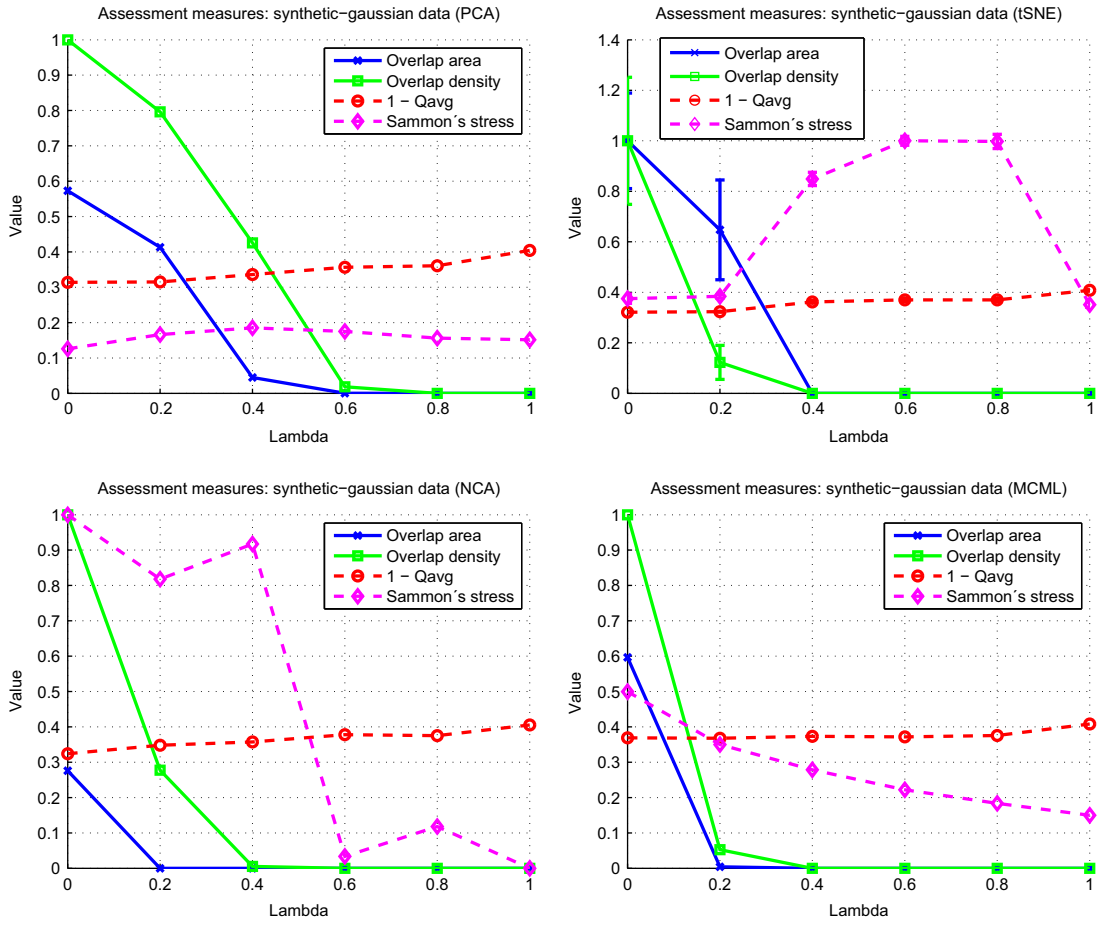


Fig. 12. Assessment measures using PCA, *t*-SNE (top), NCA, and MCML methods (bottom) for *synthetic-gaussian* dataset.

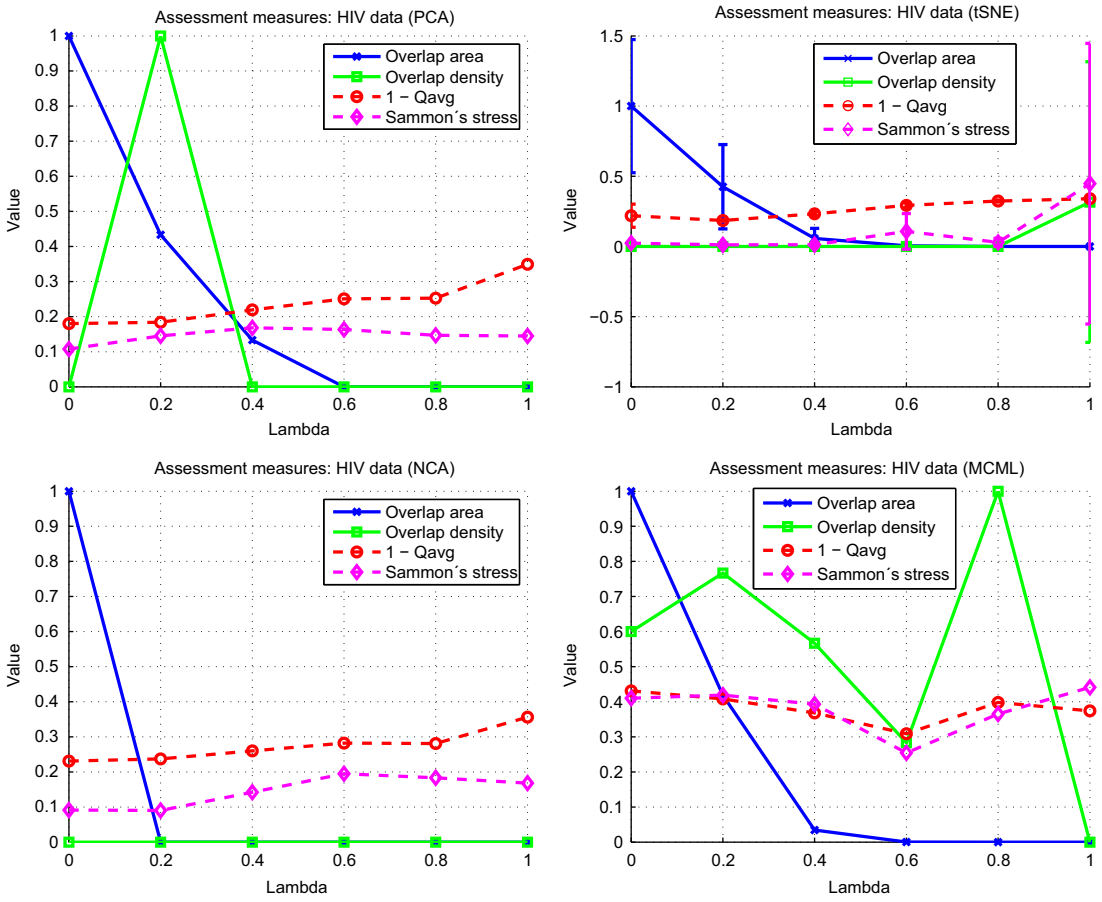


Fig. 13. Assessment measures using PCA, *t*-SNE (top), NCA, and MCML methods (bottom) for *hiv* dataset.

pairwise distance (as indicated by Sammon's stress measure). On the other hand, in many cases the *k*-ary measures stay reasonably unchanged after transformation which means overall structural preservation in terms of both global distance and local neighborhood. The quality evaluation helps the analyst better understand the distortion caused by the transformation, and evaluate the quality gain in terms of visual clarification. A trade-off can be made fairly easily if a similar "quality graph" is provided during the analysis process.

Besides, the λ value itself gives a good indication of the "degree of distortion". By modifying the λ parameter the user can gradually control the extension or come back to any previous point. This allows one to track the variations in the projections by smooth transitions, and to be aware of the trade-off between the original structure preservation and visual quality.

In addition to evaluate the projections with the quality graph explained before, there are more approaches that can be used to visualize the quality in the projection. For instance, the evaluation of a Self-Organizing Map can be visualized employing the U-Matrix [44]. This idea provides information to the user about the underlying structure preservation with respect to the original data into the embedding. In a similar way, a point-wise quality evaluation is proposed in [42] using a rank-based criteria that allows to highlight erroneous regions in the visualization. Using this approach, a mean error can be computed for each point and encoded as color in the projection. In Fig. 16 this evaluation of the quality is shown for the *eCons* data example with the labels of weekday where some points reveal worse structural quality with the appliance of the method. This useful visualization helps the user to be aware of the errors so that the control parameter can be set in a final value easily with a direct evaluation of the projection.

5. Discussion

Given an initial projection, the proposed approach allows the user to generate new projections with improved visual quality by integrating new group information into the DR process, assuming that the new information is validated by the user and provides knowledge related to the analyzed tasks. The method can be applied when the grouping information cannot be fully revealed by the distance measures that are used to compute the projection due to noise and irrelevant information. Another advantage of the proposed method is the preservation of local structure. As the distances between points within the same group are not altered by the transformation, new projections based on transformed data preserve the local structure in the original data. In addition, the DR assessment using numerical quality measures (see Section 4.5) gives an idea of the structural variations with respect to the original data and the visual improvements produced by the method.

From the visualization point of view, one may argue that given grouping information it is easier to use color or shape to differentiate groups in the projection. However, the color-coding and shape-coding approaches do not solve the cluttering problem. When points are over-plotted in a visual display, the readability of the projection is still not much improved. Our approach helps to reduce cluttering in the projection and provides a more efficient visual channel for assessing relative distances between objects and classes. Furthermore, using space to separate classes makes it possible to apply other interaction mechanisms such as hovering over points to get contextual information or area selection to calculate aggregated values.

The transformation process is controllable via a weighting parameter λ. When $\lambda=0$, the projection is purely based on the

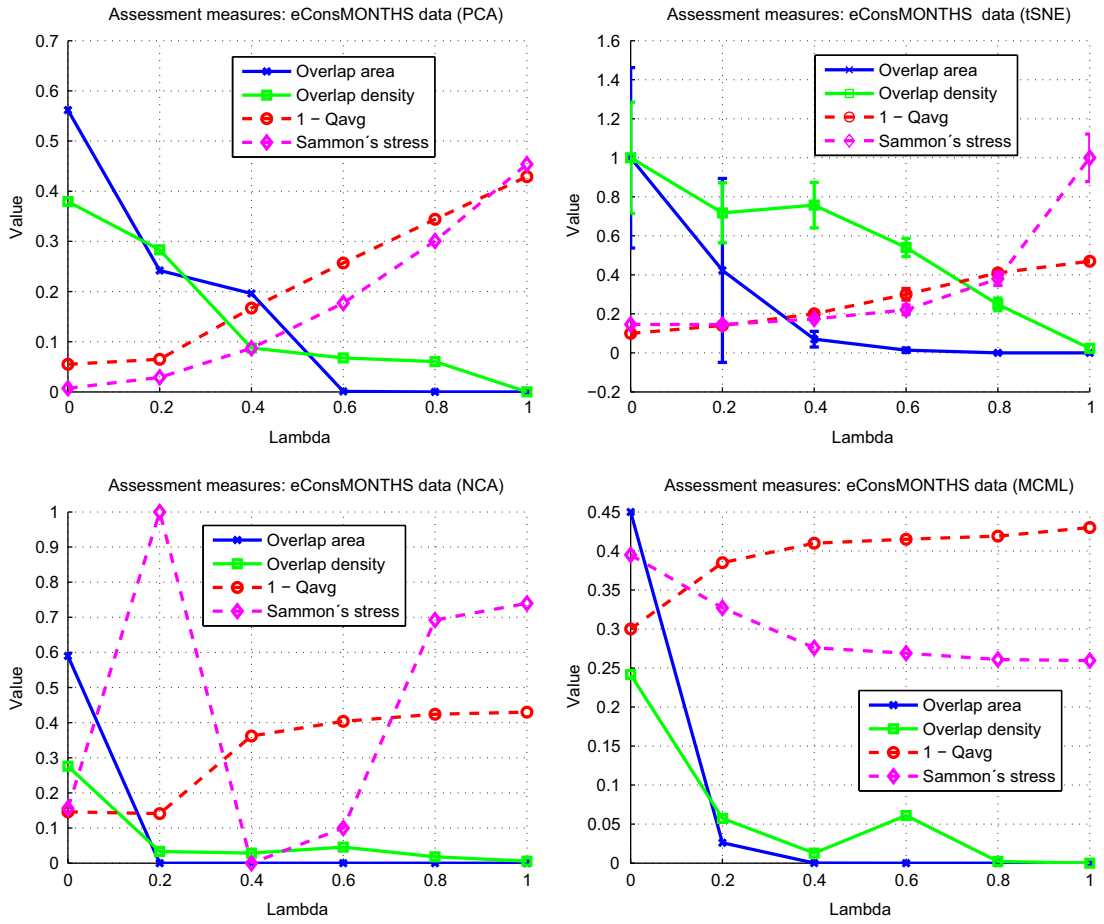


Fig. 14. Assessment measures using PCA, *t*-SNE (top), NCA, and MCML methods (bottom) for *eCons* (*months*) dataset.

original data. As λ value enlarges, the method gradually increases the influence of the grouping information. When λ value reaches 1, only class information is used for computing projection hence all points collapse to their corresponding centroids. The analyst can start with an initial projection ($\lambda=0$) and gradually increase the λ value in order to achieve clearer group separation in the projection space. Such a process can be easily facilitated by an interactive graphical interface with a sliding bar.

The role of interaction is a key aspect within this context. The interactive changes of λ values allow the user to go back and forth as much as needed. This reversible process helps to keep in mind the initial reference view at each stage. In addition, the smooth variations of the points allow a continuous object tracking that can be performed with animation improving its graphical perception [45]. The user can examine the original data structure ($\lambda=0$), the projection of class centroids ($\lambda=1$), and intermediate views ($0<\lambda<1$) that allow to get new insights not available with a single DR tool. This can be used not only to achieve an interactive grouping separation but also to understand which part of the overlap of the classes is produced by the projection or is an actual characteristic of the multidimensional data. We further note that appropriate methods for visualization of projection qualities (e.g. based on projection stress) have been developed [42–44,46,47] and can be combined with our interactive approach. Especially in combination with interactive setting of λ values, a dynamic visualization of projection quality will help the analyst to assess the distortion introduced and strike a balance between data distortion and a de-cluttered projection.

For the interactive approach, it is always desirable to have smooth transition between views when the λ value is updated. This can be challenging due to computational time required to generate new projections, especially when the data is large in size and/or dimensionality. For example, *PCA* has a complexity of $O(d^3)$ where d is the number of dimensions, so when the dimensionality of data is very high, some preprocessing stage such as feature selection may be required to reduce the dimensionality of the data. Another example is the *t*-SNE approach, the original *t*-SNE algorithm has a complexity of $O(n^2)$ where n is the number of objects in the data. Although some recent work reduced the complexity of *t*-SNE to $O(n \log n)$ [34,48], the method can still fail to support smooth transitions when n is large. In such case, sampling may be needed prior to the computation to reduce the computational load. Another possible solution to improve the scalability of the proposed approach is to pre-compute a series of projections with increasing λ values.

6. Conclusions

In this paper we propose a simple but effective approach that supports projection-based data analysis. The proposed interactive analysis framework extends traditional dimensionality reduction approaches that transform multi-dimensional data to a lower-dimensional visual display as a static view to an interactive visual display that allows the analyst to gradually modify the projection by incorporating grouping information. The proposed approach differs from traditional supervised DR methods in such a way that the user has more control over the analysis process. For example, they perform an extension of the features based on classes information and adjust the weight between original and extended

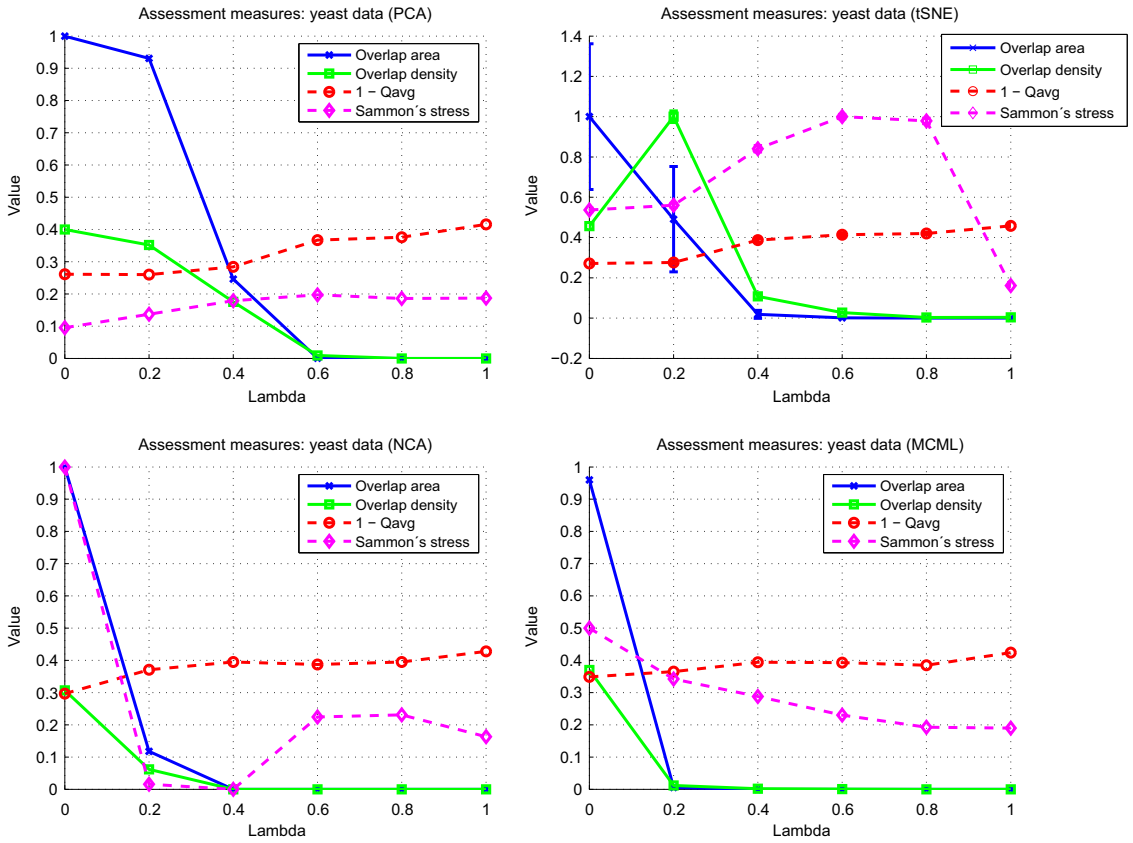


Fig. 15. Assessment measures using PCA, *t*-SNE (top), NCA, and MCML methods (bottom) for *yeast* dataset.

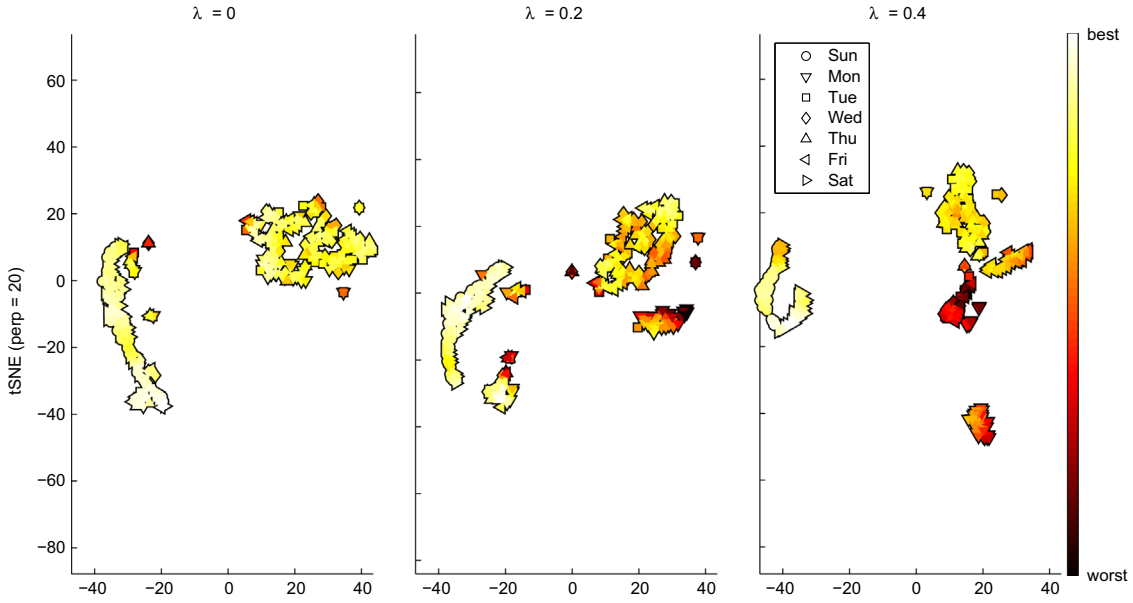


Fig. 16. Point-wise quality visualization for the evaluation of the example with *eCons* (*weekday*) dataset for several λ values.

feature space before projection so that the influence of class knowledge can be changed in the final projection. To bring more transparency to the analysis, the framework also integrates various quantitative measures to help analysts judge the quality of generated projection both in terms of structural preservation and visual clarification.

A number of experiments were carried out to evaluate the effectiveness of the proposed approach, covering different types of datasets, both supervised and unsupervised DR techniques, under different weighting conditions, and under different use case scenarios. The resulting projections are evaluated both visually and using quantitative measures that compute the structural preservation and visual quality. The experimental results indicate that the proposed methods not only lead to improved visual quality but also preserve the local neighborhood reasonably well. The resulting projections show the incorporation of meaningful

information in a transparent manner. This provides efficiency in the visual analytics process for pattern recognition, fast identification of class labels and a better understanding of the data.

Future work includes exploring more interactive visualization techniques, the design of more sophisticated extension strategies that are tailor-made to the nature of data for improving the effectiveness of the methods, and developing a wider range of quality measures for evaluating the projections. Moreover, a user study is planed to ascertain the usability of the proposed technique.

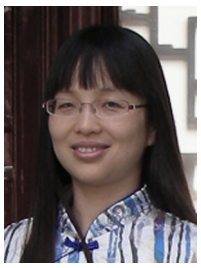
Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neucom.2014.09.061.

References

[1] J. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer, USA, 2007.
[2] L. Van der Maaten, An introduction to dimensionality reduction using matlab, Report 1201 (07) (2007) 62.
[3] D.L. Donoho, High-dimensional data analysis: the curses and blessings of dimensionality, in: Proceedings of American Mathematical Society Conference on Mathematical Challenges of the 21st Century, 2000.
[4] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (7) (1936) 179–188.
[5] J. Goldberger, S. Roweis, G. Hinton, R. Salakhutdinov, Neighbourhood components analysis, in: Advances in Neural Information Processing Systems, vol. 17, 2004.
[6] A. Globerson, S. Roweis, Metric learning by collapsing classes, in: Advances in Neural Information Processing Systems, vol. 18, 2006, p. 451.
[7] D.A. Keim, J. Kohlhammer, G. Ellis, F. Mansmann, Mastering The Information Age—Solving Problems with Visual Analytics, Eurographics, Germany, 2010.
[8] M. Schaefer, L. Zhang, T. Schreck, A. Tatu, J.A. Lee, M. Verleysen, D.A. Keim, Improving projection-based data analysis by feature space transformations, in: Proceedings of the SPIE Visualization and Data Analysis (VDA), 2013.
[9] D.H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, R. Chang, iPCA: an interactive system for PCA-based visual analytics, Comput. Graph. Forum 28 (3) (2009) 767–774.
[10] J. Choo, H. Lee, J. Kihm, H. Park, ivisclassifier: an interactive visual analytics system for classification based on supervised dimension reduction, in: 2010 IEEE Symposium on Visual Analytics Science and Technology (VAST), 2010, pp. 27–34.
[11] P. Joia, F. Paulovich, D. Coimbra, J. Cuminato, L. Nonato, Local affine multidimensional projection, IEEE Trans. Vis. Comput. Graph. 17 (12) (2011) 2563–2571.
[12] D. Perez, L. Zhang, M. Schaefer, T. Schreck, D.A. Keim, I. Diaz, Interactive visualization and feature transformation for multidimensional data projection, in: Proceedings of the EuroVis Workshop on Visual Analytics Using Multidimensional Projections, 2013.
[13] A. Endert, C. Han, D. Maiti, L. House, S. Leman, C. North, Observation-level interaction with statistical models for visual analytics, in: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), IEEE, USA, 2011, pp. 121–130.
[14] J.G.S. Paiva, W.R. Schwartz, H. Pedrini, R. Minghim, Semi-supervised dimensionality reduction based on partial least squares for visual analysis of high dimensional data, Comput. Graph. Forum 31 (3pt4) (2012) 1345–1354.
[15] F. Paulovich, C. Silva, L. Nonato, User-centered multidimensional projection techniques, Comput. Sci. Eng. 14 (4) (2012) 74–81.
[16] S. Johansson, J. Johansson, Interactive dimensionality reduction through user-defined combinations of quality metrics, IEEE Trans. Vis. Comput. Graph. 15 (6) (2009) 993–1000.
[17] S. Ingram, T. Munzner, V. Irvine, M. Tory, S. Bergner, T. Möller, Dimstiller: workflows for dimensional analysis and reduction, in: Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST), vol. 1, Citeseer, USA, 2010.
[18] G.M. Mamani, F.M. Fatore, L.G. Nonato, F.V. Paulovich, User-driven feature space transformation, in: Computer Graphics Forum, vol. 32, Wiley Online Library, USA, 2013, pp. 291–299.
[19] E. Brown, J. Liu, C. Brodley, R. Chang, Dis-function: learning distance functions interactively, in: 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), 2012, pp. 83–92.
[20] S. Bremm, T. von Landesberger, J. Bernard, T. Schreck, Assisted descriptor selection based on visual comparative data analysis, in: Computer Graphics Forum, vol. 30, Wiley Online Library, USA, 2011, pp. 891–900.
[21] S. Bremm, T. von Landesberger, M. Heß, T. Schreck, P. Weil, K. Hamacherk, Interactive visual comparison of multiple trees, in: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), IEEE, USA, 2011, pp. 31–40.
[22] E. Bertini, A. Tatu, D. Keim, Quality metrics in high-dimensional data visualization: an overview and systematization, in: Proceedings of the IEEE Symposium on IEEE Information Visualization (InfoVis), vol. 17, 2011, pp. 2203–2212.
[23] J.A. Lee, M. Verleysen, Quality assessment of dimensionality reduction: rank-based criteria, Neurocomputing 72 (7) (2009) 1431–1443.
[24] J.W. Sammon, A nonlinear mapping for data structure analysis, IEEE Trans. Comput. 18 (5) (1969) 401–409.
[25] J. Kruskal, Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new 'index of condensation', in: R. Milton, J. Nelder (Eds.), Statistical Computation, Academic Press, New York, 1969, pp. 427–440.
[26] J. Venna, S. Kaski, Neighborhood preservation in nonlinear projection methods: an experimental study, in: G. Dorffner, H. Bischof, K. Hornik (Eds.), Proceedings of ICANN 2001, Springer, Berlin, 2001, pp. 485–491.
[27] J. Lee, M. Verleysen, Quality assessment of nonlinear dimensionality reduction based on k-ary neighborhoods, in: Y. Saeys, H. Liu, I. Inza, L. Wehenkel, Y. Van de Peer (Eds.), JMLR Workshop and Conference Proceedings (New Challenges for Feature Selection in Data Mining and Knowledge Discovery), vol. 4, 2008, pp. 21–35.
[28] L. Saul, S. Roweis, Think globally, fit locally: unsupervised learning of nonlinear manifolds, J. Mach. Learn. Res. 4 (2003) 119–155.
[29] J. Venna, S. Kaski, Nonlinear dimensionality reduction as information retrieval, in: M. Meila, X. Shen (Eds.), Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (AISTATS 2007), Omnipress, San Juan, Puerto Rico, 2007, pp. 568–575.
[30] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, D. Keim, Combining automated analysis and visualization techniques for effective exploration of high-dimensional data, in: Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST), 2009, pp. 59–66.
[31] M. Sips, B. Neubert, J.P. Lewis, P. Hanrahan, Selecting good views of high-dimensional data using class consistency, Comput. Graph. Forum 28 (3) (2009) 831–838.
[32] A. Moreira, M.Y. Santos, Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points, in: GRAPP 2007: Proceedings of the International Conference on Computer Graphics Theory and Applications, INSTICC Press, Portugal, 2007, pp. 61–68, ISBN 978-972-8865-71-9.
[33] M. Li, J.T. Kwok, B.-L. Lu, Making large-scale nystrom approximation possible, in: J. Furnkranz, T. Joachims (Eds.), ICML, Omnipress, USA, 2010, pp. 631–638.
[34] Z. Yang, J. Peltonen, S. Kaski, Scalable optimization of neighbor embedding for visualization, in: ICML, vol. 2, 2013, pp. 127–135.
[35] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, M. Biehl, Limited rank matrix learning, discriminative dimension reduction and visualization, Neural Netw. 26 (2012) 159–173.
[36] D.G. Kendall, A survey of the statistical theory of shape, Stat. Sci. 4 (2) (1989) 87–99.
[37] M. Sedlmair, A. Tatu, T. Munzner, M. Tory, A taxonomy of visual cluster separation factors, in: Computer Graphics Forum, vol. 31, Wiley Online Library, USA, 2012, pp. 1335–1344.
[38] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.
[39] J.J. Van Wijk, E.R. Van Selow, Cluster and calendar based visualization of time series data, in: Proceedings of the 1999 IEEE Symposium on Information Visualization (InfoVis'99), IEEE, USA, 1999, pp. 4–9.
[40] C. Blake, C.J. Merz, {UCI} repository of machine learning databases, University of California, Irvine, School of Information and Computer Sciences, 1998.
[41] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[42] B. Mokbel, W. Lueks, A. Gisbrecht, B. Hammer, Visualizing the quality of dimensionality reduction, Neurocomputing 112 (2013) 109–123.
[43] M. Aupetit, Visualizing distortions and recovering topology in continuous projection techniques, Neurocomputing 70 (7) (2007) 1304–1330.
[44] A. Ultsch, H.P. Siemon, Kohonen's self organizing feature maps for exploratory data analysis, in: INNC Paris 90, Universitat Dortmund, Kluwer, Netherlands, 1990, pp. 305–308.
[45] J. Heer, G.G. Robertson, Animated transitions in statistical data graphics, IEEE Trans. Vis. Comput. Graph. 13 (6) (2007) 1240–1247.
[46] S. Lespinats, M. Aupetit, Checkviz: sanity check and topological clues for linear and non-linear mappings, Comput. Graph. Forum 30 (1) (2011) 113–125.
[47] T. Schreck, T. von Landesberger, S. Bremm, Techniques for precision-based visual analysis of projected data, Inf. Vis. 9 (3) (2010) 181–193.
[48] L. van der Maaten, Barnes-hut-sne, arXiv preprint arXiv:1301.3342.

Daniel Pérez is currently a Ph.D. student at the University of Oviedo, Spain, in the research group of supervision and diagnostic of industrial processes. He studied industrial engineering at the University of Oviedo, where he received his M.Eng. in 2007, later he became a research assistant in Electrical Engineering Department at the University of Oviedo. His main research interests are information visualization, machine learning, and visual analytics. He joined in his current group in 2011 and is currently working in visualization of high-dimensional data projections.



Leishi Zhang is a lecturer in visual analytics at Middlesex University, UK. She received her Ph.D. in computer science from Brunel University, UK, for her work in time series data analysis and visualization. Her research interests include time series data modelling, high-dimensional data projection and interactive visualization of subspace clusters. She has worked on various research and industrial projects in visual analytics and published her research in a number of international journals and conferences.

Matthias Schaefer is a Ph.D. student at the University of Konstanz in the Visualization and Data Analysis Group. He is working at the Department of Computer and Information Science and his current research interest are information visualization, data mining, multimedia- and multidimensional-databases and visual analytics.

Tobias Schreck is an assistant professor of visual analytics in the Department of Computer and Information Science at the University of Konstanz, Germany. His research interests include visual search and analysis in time-oriented, high-dimensional, and 3D object data, with applications in data analysis and multimedia retrieval. Schreck received a Ph.D. in Computer Science from the University of Konstanz. He is a member of IEEE.

Daniel Keim is a full professor in the Department of Computer Science at the University of Konstanz, Germany, and chair of the university's Visualization and Data Analysis Group. His research interests include visual analytics, information visualization, and data mining. Keim received a Ph.D. in computer science from the University of Munich, Germany. He is a member of the IEEE Computer Society and a coordinator of the German strategic research initiative on scalable visual analytics.

Ignacio Díaz is an associate professor of Electrical Engineering Department at the University of Oviedo since 2004. He received a M.Eng. in 1995 and his Ph.D. in industrial engineering in 2000 from the University of Oviedo. His main research interests are the application of data visualization and intelligent data analysis algorithms to industrial problems. He has led several R&D projects financed by the Spanish Government and the European Union and published his research in indexed journals, as well as numerous publications in international conferences. Professor Díaz is member of the IEEE since 1997.

A.3 VISUAL ANALYSIS OF ELECTRICAL POWER CONSUMPTION PATTERNS USING MANIFOLD LEARNING [148]

VISUAL ANALYSIS OF ELECTRICAL POWER CONSUMPTION PATTERNS USING MANIFOLD LEARNING

Daniel Pérez
DIEECS
Universidad de Oviedo
Spain
dperez@isa.uniovi.es

Ignacio Díaz
DIEECS
Universidad de Oviedo
Spain
idiaz@isa.uniovi.es

Francisco J. García
DIEECS
Universidad de Oviedo
Spain
fjgarcia@isa.uniovi.es

Manuel Domínguez
IAF
Universidad de León
Spain
diemdg@unileon.es

Pablo Barrientos
IAF
Universidad de León
Spain
pablo.barrientos@unileon.es

Abstract

Manifold learning algorithms are recent techniques that can extract useful geometric information from high-dimensional data with complex structure to obtain a projection onto a lower dimensional space. Data from power systems can show manners of spending electricity in big buildings where a straightforward management becomes difficult. Those consumptions patterns can be revised using manifold learning visualization techniques to improve their energy efficiency. In this work a dimensionality reduction stage has been applied to a main set of electrical and time variables using manifold learning algorithms, showing how behaviors can be identified and analyzed in order to decrease the electrity bill.

Key words

Manifold learning, electrical consumption, dimensionality reduction

1 Introduction

The electric power industry has experienced restructuring changes and the energy efficiency has become a priority because of enviromental and financial reasons.

The attempts of an adequate management of the electricity become more complicated with the size of the facilities. In these large equipments, the energy spent can be misused easily and this can provoke economic losses in the electric tariff.

One of the first tasks to control power consumptions is an detailed study of the most important variables which are involved in the tariff. For this analysis the dimension of data recorded can be reduced so that visualize them easily.

Manifold learning techniques have been an actively field in the last decade. These algorithms compute a manifold in a low-dimensional space from high dimensional data with an underlying structure. Some of the most known examples are isometric embedding mapping or Isomap (Tenenbaum, de Silva and Langford, 2000), Laplacian Eigenmaps (LE) (Belkin and Niyogi, 2003), locally linear embedding (LLE) (Roweis and Saul, 2000), local tangent subspace alignment (LTSA) (Zhang and Zha, 2004) and t-Distributed Stochastic Neighbour Embedding (t-SNE) (van der Maaten and Hinton, 2008).

Databases and data acquisition equipments provide an enormous amount of samples. Generally the application of these techniques have a computational complexity of $\mathcal{O}(N^3)$.

Unfortunately its performance can be non-viable for a large datasets. In this work active and reactive power and time variables were considered as a vector of high dimensionality and studied during a set time. Due to the use of a huge number of samples, a size reduction followed by a neural network based on interpolation were applied to perform the dimensionality reduction.

This paper is organized as follows: section 2 describes the applied method, section 3 explains the experiment done, section 4 shows the results of the experiment and section 5 summarizes the obtained conclusions.

2 Computation of the dimensionality reduction

In the first stage a data compression using a competitive learning algorithm has been applied to the dataset so that the size of the sample is reduced.

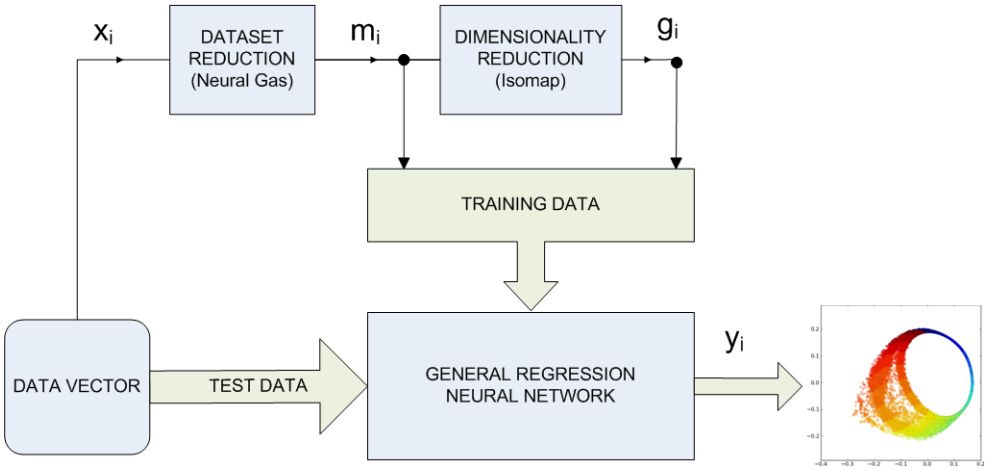


Figure 1. Description of the method

A vector quantization technique, *neural gas* algorithm (Martinetz, Berkovich and Schulten, 1993), is used for finding an optimal representation of the data, describing homogeneously the relevant geometry of input space $V \subseteq \mathbb{R}^d$ with minimal recostruction error.

A finite number of "*codebooks*" vectors $\mathbf{m}_i \in \mathbb{R}^d$ (i=1, ..., N) is obtained preserving the joint probability density function (pdf) of data vectors $p(\mathbf{x})$.

The neirborhood relationships for a given data vector $\mathbf{x} \in \mathbb{R}^n$ are determined a ranking $(\mathbf{m}_{i_0}, \mathbf{m}_{i_1}, ..., \mathbf{m}_{i_{N-1}})$ with $\mathbf{m}_{i_0}$ being the closest one to $\mathbf{x}$, $\mathbf{m}_{i_1}$ being second closest to $\mathbf{x}$, and $\mathbf{m}_{i_k}$, being the reference vector for which there are k vectors $\mathbf{m}_j$ with $\|\mathbf{x} - \mathbf{m}_j\| < \|\mathbf{x} - \mathbf{m}_{i_k}\|$.

The adaptation steps for adjusting $\mathbf{m}_i$'s is perfomed as follows:

$$\mathbf{m}_{i_k}^{t+1} = \mathbf{m}_{i_k}^{t} + \epsilon \cdot e^{-k/\lambda} \cdot (\mathbf{x} - \mathbf{m}_{i_k}^{t}) \qquad (1)$$

with $\epsilon \in [0, 1]$ as the step size and λ as the neighborhood range reduced with increasing t. The *codebooks* are a group of neural units which are considered to represent all the samples of the data.

There is a variant neural gas using *batch algorithm* (Cottrell, Hammer, Hasenfuss and Villmann, 2006) which optimizes convergence.

For a training points $(\mathbf{x}_1, ..., \mathbf{x}_p)$, the rank for the prototype i is $k_i(\mathbf{x}, \mathbf{m}) = |\{\mathbf{m}_j | \ \|\mathbf{x} - \mathbf{m}_j\|^2 < \|\mathbf{x} - \mathbf{m}_i\|^2\}|$.

Being $h_\lambda(t) = exp(-t/\lambda)$ with $\lambda > 0$, it determines assigments of prototypes as follows

$$\mathbf{m}_i = \frac{\sum_{j=1}^{p} h_\lambda(k_i(\mathbf{x}, \mathbf{m}))\mathbf{x}_j}{\sum_{j=1}^{p} h_\lambda(k_i(\mathbf{x}, \mathbf{m}))} \qquad (2)$$

In a second stage a dimensionality reduction is computed by means of the Isomap technique with these representative points in order to obtain points $(\mathbf{g}_i)$ which defines the structure of data in a lower dimensional space. This technique consists in a classical multidimensional scaling (*MDS*) which is calculated using geodesic distances instead of euclidean distances.

The steps which are followed by Isomap are composed by defining a neirborhood graph, then compute shortest paths distances, and eventually the construction of a lower-dimensional embedding of data.

Finally an estimation neural network, *General Regression Neural Network* (*GRNN*) (Specht, 1991) is applied. Although its application can lose some details it allows to calculate an embedding for whole set of data.

Assuming that $f(\mathbf{x}, \mathbf{y})$ represents the joint pdf of measures of vector randoms variables $\mathbf{x}$, and $\mathbf{y}$. The regression function is given by the conditional expectation of $\mathbf{y}$ in $\mathbf{x}$.

$$f(\mathbf{x}) = E[\mathbf{y}|\mathbf{x}] = \frac{\int_{-\infty}^{\infty} y f(\mathbf{x}, \mathbf{y}) dy}{\int_{-\infty}^{\infty} f(\mathbf{x}, \mathbf{y}) dy} \qquad (3)$$

When this density function is not known, it can be estimated with nonparametric estimators proposed by Parzen (Parzen, 1962) and applied to the multidimensional case by Cacoullos (Cacoullos, 1966).

The Specht's algorithm estimates a continuous regression $\hat{\mathbf{y}}(\mathbf{x})$ based on observed values $\mathbf{y}_i$ and $\mathbf{x}_i$ as follows

$$\hat{\mathbf{y}}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{g}_i \exp\left(-\frac{(\mathbf{x}-\mathbf{m}_i)^T(\mathbf{x}-\mathbf{m}_i)}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{(\mathbf{x}-\mathbf{m}_i)^T(\mathbf{x}-\mathbf{m}_i)}{2\sigma^2}\right)} \qquad (4)$$

The parameter σ determines the smoothness of the density estimated. When it is not possible to compute an optimal value, this is adjusted on an empirical basis.

In the Figure 1 is shown a brief description of the applied method.

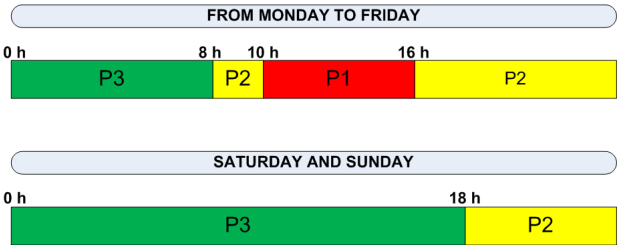


Figure 2. Distribution of the time periods of consumption

3 Experiment

The available data were recorded in the facilities of the University of León. The dataset used in the experiment corresponds with dates from 11/09/2010 to 30/10/2010 with sampling time of 2 minutes in the *Escuela de Ingeniería Agraria* building.

The experiment is made with the analysis of a group of variables, power consumptions (active power, and reactive power) previously normalized to zero mean and unit variance. Time variables, in senoidal form, are also added with a superior weigth factor over the rest.

The *batch algorithm* of neural gas is applied for a number of 300 units with 10 epochs with a final neighborhood of 0.1.

Dimensionality reduction is computed for the 300 resulting codebooks $\mathbf{m}_i$ with a number of neighbors of $k = 5$ obtaining the $\mathbf{g}_i$ points in a 2D space.

Then the *GRNN* performs the dimensionality reduction technique for all set of data from observation of $\mathbf{g}_i$'s and $\mathbf{m}_i$. The value of smoothness parameter chosen is $\sigma = 3$.

The consumption is divided by the electricity supplier into three billing periods depending on hour, day or season time.

> *Period 1 or peak* (P1): From 10 to 16 hours from Monday to Friday.
> *Period 2 or flat* (P2): From 8 to 10 hours and from 16 to 0. Weekends from 18 to 0 hours.
> *Period 3 or off-peak* (P3): From 0 to 8 hours. Weekends from 0 to 18 hours.

In Figure 2 can be seen a scheme of the periods distribution.

The demand has a different price in each period, being P1 the most expensive period and P3 the cheapest one.

The billed *power term*, P_f, is calculated from contracted and maximum value demanded one, P_c and P_m, respectively, as follows:

$$P_f = \begin{cases} 0.85\ P_c & \text{if } P_m < 0.85 P_c \\ P_m & \text{if } 0.85\ P_c \leq P_m \leq 1.05 P_c \\ P_m + 2\ (P_m - 1.05\ P_c) & \text{if } P_m > 1.05 P_c \end{cases} \quad (5)$$

The *active energy term* is billed directly from measured ones. Eventually, a *reactive energy term* is added when its value is more than 33 % of the active energy, that is, when power factor is less than 0.95.

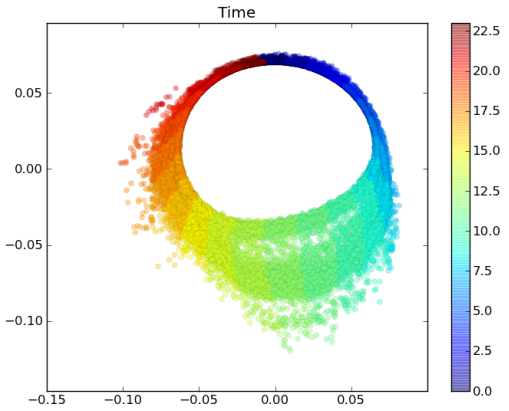


Figure 3. Results with time distribution color

4 Results

The variables visualized in Figures 3-7 are computed according to the rules described before.

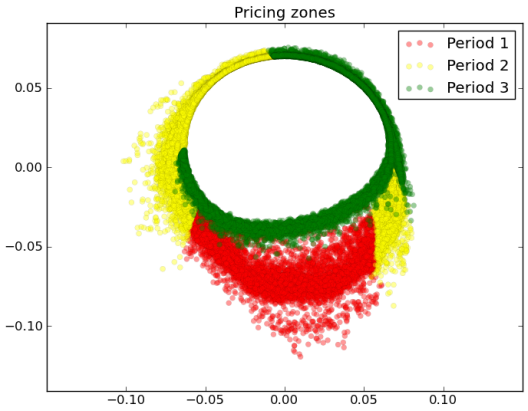


Figure 4. Billing periods distribution: P1 (red), P2 (yellow) and P3 (green)

The resulting points are rotated forming a circle resembling a clock in order to represent a 24-hour periodicity, so that 00.00 hours are at the top of the figure and the 12.00 hours at the bottom of the picture (see Figure 3). This helps to identify areas correponding with hours of the day easily.

Colors represented in Figure 4 show the distribution of the three periods where a specific price is applied for each one of them. The red area corresponds to period 1 which has the highest cost of consumption, the yellow area shows period 2 with an intermediate value of the price and finally the green area represent period 3, with the lowest price of the consumption.

It is shown in Figure 5 distribution of active power term. The contracted power has a value of $P_c = 200\ kW$, and the power billed for, P_f, is computed with equation (5). It can be seen quickly that most of

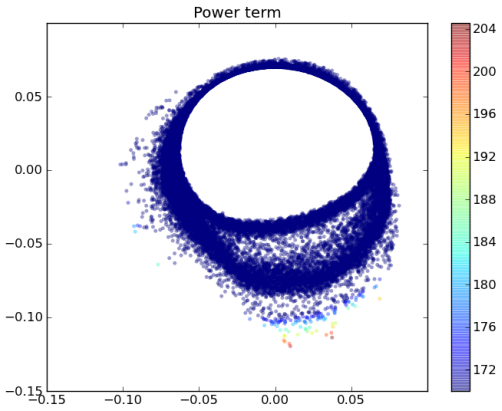


Figure 5. Results with billed power color

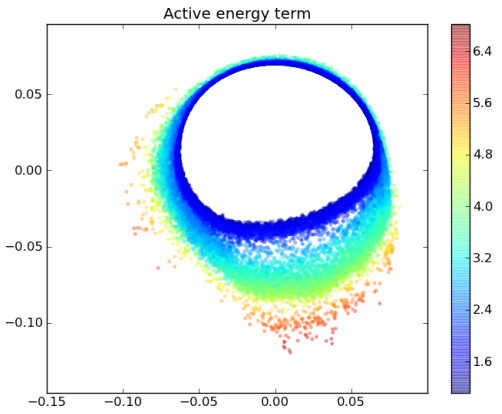


Figure 6. Results with active energy color

the time the power which is multiplied by price of the appropriate period is $Pf = 0.85 \cdot P_c = 170\ kW$.

Active energy term is directly billed from its measures. In Figure 6 the resulting projections whose color indicates the demand of active energy are represented.

Comparing with Figure 3, the points with high demand of active energy correspond to daylight hours and points at these hours with lower demand are weekends or bank holiday, as it is expected. It is also interesting to see in Figure 4 that the points with the highest active energy have the most expensive cost too (P1) and vice versa.

Figure 7 shows resulting points of the experiment with the color evaluating as 1 for reactive energy term which are charged in the bill at every time. It is seen that reactive term is not depending of billing periods.

5 Conclusions

In this paper a method has been proposed to compute a dimensionality reduction by manifold learning technique such as Isomap without limiting the size of original data.

A previous data compression using a batch algorithm

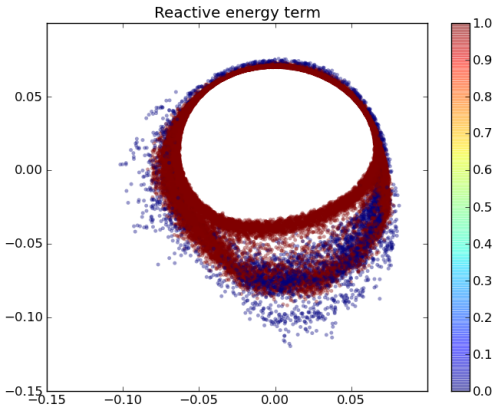


Figure 7. Results with reactive energy color

of neural gas and then a general regression neural network estimates the computed reduction for whole set of data. In the experiment, it was used to analyze power consumptions data in a period of time.

The resulting projections points from a big amount of data show consumed electrical variables in a bidimensional space depending on time. It is useful to visualize quickly variables involved in final cost of consumptions. The results show usual patterns which validate the method applied.

This method helps to analyze the visualization of related electrical variables and reach the purpose of improving consumption patterns and hence decreasing amount of the electricity bill.

Allows to visualize what hour of the day or day of the week correspond to a specific consumption behaviour, and this one can be modified and also be negociated with electricity supplier.

Changing contracted power, reconfiguring billing periods conveniently or improving compensation of power factor can be some measures in order to enhance energy efficiency. The method can also be applied to several sizes of time periods or another kind of high-dimensional data.

References

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation, *Neural computation* **15**(6): 1373–1396.

Cacoullos, T. (1966). Estimation of a multivariate density, *Annals of the Institute of Statistical Mathematics* **18**: 179–189. 10.1007/BF02869528.

Cottrell, M., Hammer, B., Hasenfuss, A. and Villmann, T. (2006). Batch and median neural gas, *NEURAL NETWORKS* **19**(6-7): 762–771. 5th Workshop on Self-Organizing Maps (WSOM 05), Paris, FRANCE, SEP 05-08, 2005.

Martinetz, T., Berkovich, S. and Schulten, K. (1993). 'neural-gas' network for vector quantization and its

application to time-series prediction, *Neural Networks, IEEE Transactions on* **4**(4): 558 –569.

Parzen, E. (1962). On estimation of a probability density function and mode, *The Annals of Mathematical Statistics* **33**(3): pp. 1065–1076.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding, *Science* **290**: 2323–2326.

Specht, D. F. (1991). A general regression neural network, *IEEE Transactions on Neural Networks* **2**(6): 568–576.

Tenenbaum, J. B., de Silva, V. and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction, *Science* **290**: 2319–2323.

van der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE, *JOURNAL OF MACHINE LEARNING RESEARCH* **9**: 2579–2605.

Zhang, Z. and Zha, H. (2004). Principal manifolds and nonlinear dimension reduction via local tangent space alignment, *SIAM Journal of Scientific Computing* **26**(1): 313–338.

A.4 VISUAL ANALYSIS OF A COLD ROLLING PROCESS USING DATA-BASED MODELING [149]

Visual Analysis of a Cold Rolling Process Using Data-Based Modeling

Daniel Pérez[1], Francisco J. García-Fernández[1], Ignacio Díaz[1], Abel A. Cuadrado[1], Daniel G. Ordonez[1], Alberto B. Díez[1], and Manuel Domínguez[2]

[1] Universidad de Oviedo. Área de Ingeniería de Sistemas y Automática
{dperez,fjgarcia,idiaz,dgonzalez,cuadrado,alberto}@isa.uniovi.es
[2] Universidad de León. Instituto de Automática y Fabricación
diemdg@unileon.es

Abstract. In this paper, a method to characterize the chatter phenomenon in a cold rolling process is proposed. This approach is based on obtaining a global nonlinear dynamical MISO model, relating four input variables and the exit strip thickness as the output variable. In a second stage, local linear models are obtained for all working points using sensitivity analysis on the nonlinear model to get input/output small signal models. Each local model is characterized by a high dimensional vector containing the frequency response functions (FRF) of the four SISO resulting models. Finally, the FRF's are projected on a 2D space, using the t-SNE algorithm, in order to visualize the dynamical changes of the process. Our results show a clear separation between chatter condition and other vibration states, allowing an early detection of chatter as well as being a visual analysis tool to study the chatter phenomenon.

Keywords: dimensionality reduction, cold rolling, data visualization, dynamical systems, data-based models.

1 Introduction

Steel production is usually considered an indicator of economic progress, as it is fairly related to infrastructures and development. After steel production from iron, the material needs to be treated and modified through several mechanical processes, such as the rolling process. The cold rolling of steel is a widely adopted process, in which a steel sheet is passed through a pair of rolls whereby the sheet thickness is reduced. Although this process has been studied for decades [1], many unsolved issues hold. The control of many different parameters is necessary, ranging from those related to the milling itself (force applied, torque,...) to those depending on different aspects, such as lubrication or refrigeration. Furthermore, there is an ever increasing demand for higher quality from costumers and, since it is a large and complex process that continuously evolves due to drifts, misadjustments and changes in working conditions, there is a need of continuous improvement in the efficiency of the process. Because of that, the supervision of this process is critical, in order to avoid faults that affect negatively to the material.

C. Jayne, S. Yue, and L. Iliadis (Eds.): EANN 2012, CCIS 311, pp. 244–251, 2012.

One of the most relevant faults in the cold rolling process of steel is called *chatter* [2], an unexpected powerful vibration that affects the quality of the rolled material by causing an unacceptable variation of the final thickness. The real problem of chatter is not only related to the bad quality of the manufactured product, but also to the economic losses suffered. Generally, when chatter appears, it is necessary to lower the rolling speed for a period of time, making the production rate decrease. A practical way to detect chatter is to compute the power spectral density in which this fault appears (normally 100-300Hz). However, although this procedure works well to show up the chatter condition, it fails as an early detector.

A way to predict chatter is to use a model of the rolling process [3]. However, the complexity of the whole process, with several tightly coupled phenomena (such as chemical, mechanical, and thermal) makes it difficult to build an accurate model and moreover to tune its parameters. An approach to enhance the knowledge about complex processes is visualizing their relevant information, using *dimensionality reduction* (DR) techniques [4,5]. DR techniques allow to project and study the structure of high-dimensional data into a low-dimensional space, typically a 2D/3D for visualization purposes, improving the exploratory data analysis [6].

In the DR field, several techniques have been proposed [7]. One of the first algorithms is Principal Component Analysis (PCA), described by Pearson [8]. After PCA, other DR techniques have been proposed, such as Multidimensional Scaling (MDS) methods, Independent Component Analysis (ICA)[9] or Self-Organizing Maps (SOM)[10]. In the beginning of 21st century, a new trend in DR based on nonlinear models appeared, inspiring a new collection of algorithms. These algorithms –known as *manifold learning*– involve a local optimization by defining local models of the k-nearest neighbours and an alignment in order to obtain the global coordinates of each model, usually implying a singular value decomposition (SVD). Some of the most known techniques are *Isomap* [11], *local linear embedding*(LLE) [12] and *laplacian eigenmaps* (LE) [13]. Similar to these techniques, but based on the probability distribution of data is *t-Stochastic Neighbor Embedding* (t-SNE) [14]. This technique, that has attracted attention recently [15,16], is capable of maintaining the local structure of the data while also revealing some important global structure (such as clusters at different scales), producing better visualizations than the rest.

In this paper, we propose a new approach for the study of chatter, using the DR principle for the analysis of the dynamical behavior of a model of the process. Using a novel feedforward neural network, called *extreme learning machine* (ELM) [17], the proposed approach computes a large feature vector composed of the frequency response functions (FRF) of a set of key physical variables and projects this vector into a 2D space by t-SNE algorithm. Thus, the changes in the dynamical behavior of the process are visualized. The paper is organized as follows: in section 2, a description of the method is shown; section 3 describes an experiment and the results of the method proposed and finally section 4 includes the conclusions obtained.

2 Data-Based Model Analysis through Manifold Learning Techniques

2.1 Description of the Physical Model

Classical cold rolling models try to calculate the force and the torque necessary for a given thickness reduction. As mentioned before, the complexity of an accurate model can be very high because of the assumptions taken [18]. In order to get a simple model to work with, e.g. [19], several assumptions can be done.

The classical form of a rolling force (F) model includes: the tension at the entry and exit side of a rolling stand (σ_{en} and σ_{ex}); the thickness at the entry and exit side (h_{en} and h_{ex}); the width of the strip (w); the friction coefficient (μ) and the hardness level of the material being rolled (S), see Eq. (1).

$$F = f(\sigma_{en}, \sigma_{ex}, h_{en}, h_{ex}, w, \mu, S) \tag{1}$$

A model of the rolling process makes it possible to analyze several defects arising from working operation, such as the chatter phenomenon. This phenomenon is a dynamic process, where variations in the roll force may lead to an unstable state. It is necessary to generate a model where the different factors likely to modify the force equilibrium in the rolling process are taken into account. As explained in [20], the chatter phenomenon comes from a feedback interaction with the variables entry speed, entry tension, the force of the strip on the rolls and exit thickness involved. If a dynamic model of the stand is added to this loop, a proper model to study the chatter phenomenon can be built [21,22].

2.2 Mathematical Estimation of the Model

As proposed in [23], data-based models are a practical way to develop a fault detection and prediction mechanism for complex processes. The development of data-based models provides a good feature for their application to industrial processes: fast responses to faults. According to the previous description of the rolling process, we propose a MISO model, defining the exit thickness y_k as the output of the system and force F_k, tension σ_k –used in the classical model Eq. (1)–, entry and exit speed of the strip (V_{en_k} and V_{ex_k} respectively) –due to their relevance in the chatter phenomenon–, as the inputs of the system. We also considered an autoregressive part of the output to account for internal dynamics, resulting in a NARX model.

$$y_k = f(y_{k-1}, \ldots, y_{k-n}, F_k, \sigma_k, V_{en_k}, V_{ex_k}) \tag{2}$$

A simple and fast learning algorithm for single hidden layer feedforward neural networks (SLFN's), called extreme learning machine (ELM) [17], is used to train

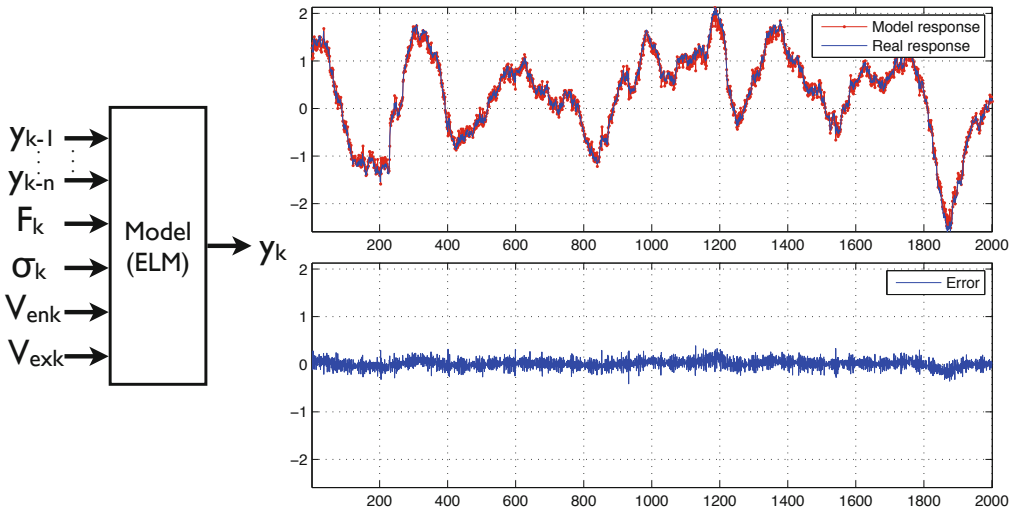


Fig. 1. Scheme of the model developed (left) and an example of its one-step-ahead prediction (right)

the NARX model. In order to obtain an optimal order of the model, we apply the Akaike Information Criterion (AIC) [24], obtaining a trade-off solution between the order and the error of the model. A scheme can be seen on the left side of Fig. 1. On the right side of Fig. 1, a comparison between the response of the real process and an example of the one-step-ahead prediction of a trained model for a testing dataset is shown.

Once the model is obtained, a local sensitivity analysis is applied to perform a system identification of the contribution of each input variable to the output. The signals are divided into M overlapped windows of size L. Let's consider $\bar{F}$, $\bar{\sigma}$, $\bar{V}_{en}$, and $\bar{V}_{ex}$, average values for the m-th window, and the delayed samples y_{k-1}, y_{k-2}, ⋯ , y_{k-n}. The m-th local sensitivity analysis ($m = 1, \ldots, M$) for the input F is performed adding to $\bar{F}$ a random value $\varepsilon_k \in N(0, \nu)$, being ν a small value.

$$\mathbf{u}_m^{\Delta F} = \left[y_{k-1}, y_{k-2}, \ldots, y_{k-n}, \bar{F} + \varepsilon_k, \bar{\sigma}, \bar{V}_{en}, \bar{V}_{ex}\right]^T \tag{3}$$

with an output $\mathbf{y}_m = [y_k]$, for $k = k_m, \ldots, k_m + L - 1$, being k_m the first sample of window m. Constructing $\mathbf{U}^{\Delta F} = [\mathbf{u}_m^{\Delta F}]$ and $\mathbf{Y} = [\mathbf{y}_m]$ resulting in an I/O pair $\left\{\mathbf{U}^{\Delta F}, \mathbf{Y}\right\}$. Similar to Eq. (3), we apply the same method to the other inputs, obtaining $\left\{\mathbf{U}^{\Delta \sigma}, \mathbf{Y}\right\}$, $\left\{\mathbf{U}^{\Delta V_{en}}, \mathbf{Y}\right\}$ and $\left\{\mathbf{U}^{\Delta V_{ex}}, \mathbf{Y}\right\}$ respectively. Each I/O pair defines a local SISO small-signal model of the process.

In order to estimate the dynamical behavior of each small-signal model, we compute FRF of the model. Let $P_y(m, j)$ and $P_{u_i}(m, j)$ be the power densities of the j-th frequency in the m-th small-signal model of the output and the i-th input, respectively. The SISO FRF of input i for all windows can be computed as $\mathbf{G}_i$ where

$$\mathbf{G}_i(m,j) = 10 \cdot log_{10}\left(\left|\frac{P_y(m,j)}{P_{u_i}(m,j)}\right|\right) \tag{4}$$

describes the gain of the j-th frequency bin for the m-th small-signal model, of the i-th input expressed in dB.

Finally, in order to project all data using a DR technique, all the FRF's of each input, $\mathbf{G}_i$, are joined into an augmented matrix $\mathbf{G}$, as expressed in Eq. (5).

$$\mathbf{G} = [\mathbf{G}_1\ \mathbf{G}_2\ \mathbf{G}_3\ \mathbf{G}_4] \qquad \mathbf{G} \in \mathbb{R}^{M \times D} \tag{5}$$

Each of the M rows of $\mathbf{G}$ is a large D-dimensional feature vector which describes the local dynamical behavior on a given window.

2.3 Dimensionality Reduction

The visualization of the dynamic behavior of the model is made by computing a dimensionality reduction using t-SNE [14]. This technique is based on similarities of the data by computing probabilities for original and projection space, defining neighborhoods with a value called *perplexity*. The computation of the two joint-probability distributions p_{ij} and q_{ij} corresponds to Gaussian and Student's t-distribution respectively, see Eq. (6) and (7).

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2/2\sigma^2)} \tag{6}$$

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{7}$$

With the aim of projecting new data points on the 2D map obtained with the training dataset, an out-of-sample extension of the t-SNE was developed. The addition of each new point was computed by ensuring that the sum of probabilities is equal to 1. The variance σ of a Gaussian centered at the new point is searched keeping the perplexity value fixed. The differences of the two probability distributions are measured by Kullback-Leibler divergences. Hence, the cost function is

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{8}$$

The coordinates of the new projections are obtained minimizing this cost function, and maintaining previous coordinates of the training dataset fixed. This minimization of the cost function is made using a gradient-descent method for optimization. Finally, as a summary of the method, a flowchart of the different stages used is shown in Fig. 2.

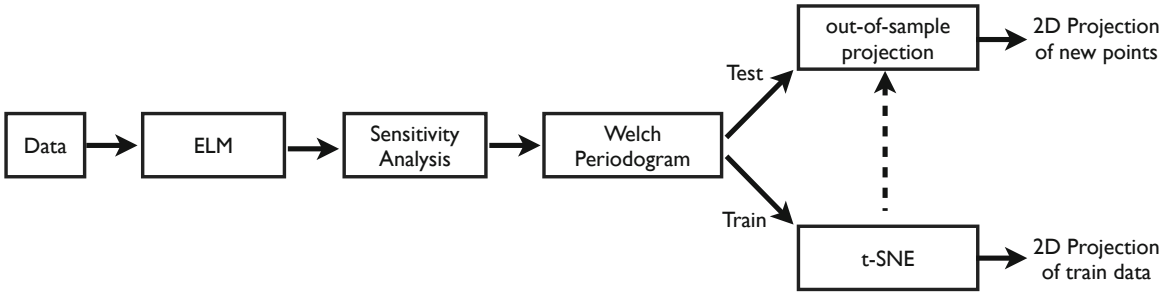


Fig. 2. Flowchart of the method

3 Experiment and Results

To validate this method, it was applied to data from a cold rolling facility. As explained in section 2.2, the variables used for modeling were: rolling force, entry tension, entry and exit speeds of the strip, and exit thickness. These variables were acquired using a data acquisition system with a sampling rate of 2000 Hz.

The training dataset was composed of signals of $N = 54300$ samples, including normal operating conditions and a chatter episode from a unique coil. In Fig. 3, the reduction of the rolling speed to avoid chatter effect can be seen, as well as the powerful variation in the other variables, especially in the thickness signal, which is closely related to the quality of the resulting product. As for testing the method, different chatter episodes of several coils were used.

To obtain the MISO model, ELM was trained using 1000 neurons in the hidden layer, with the signals normalized to zero mean and $\sigma = 1$. In Table 1, a comparison of the RMS errors of the linear regression and the EML models is shown, as well as the AIC and the Theil's Index (U) [25], for several model orders.

Table 1. Value of the RMS errors, AIC, and Theil's Index (U) for each order of the NARX model

Model Order (n)	Linear RMSE	EML RMSE	U	AIC
5	0.1856	0.1744	0.7377	$-1.5445 \cdot 10^5$
10	0.1353	0.1307	0.5529	$-1.7998 \cdot 10^5$
15	0.1252	**0.1159**	**0.4903**	$\mathbf{-1.9058 \cdot 10^5}$
20	0.1277	0.1192	0.5042	$-1.8809 \cdot 10^5$
25	**0.1251**	0.1167	0.4937	$-1.8993 \cdot 10^5$
30	0.1273	0.1194	0.5051	$-1.8787 \cdot 10^5$

According to the AIC criterion the order selected is $n = 15$. This model describes accurately the nonlinear dynamics around the working points of the process.

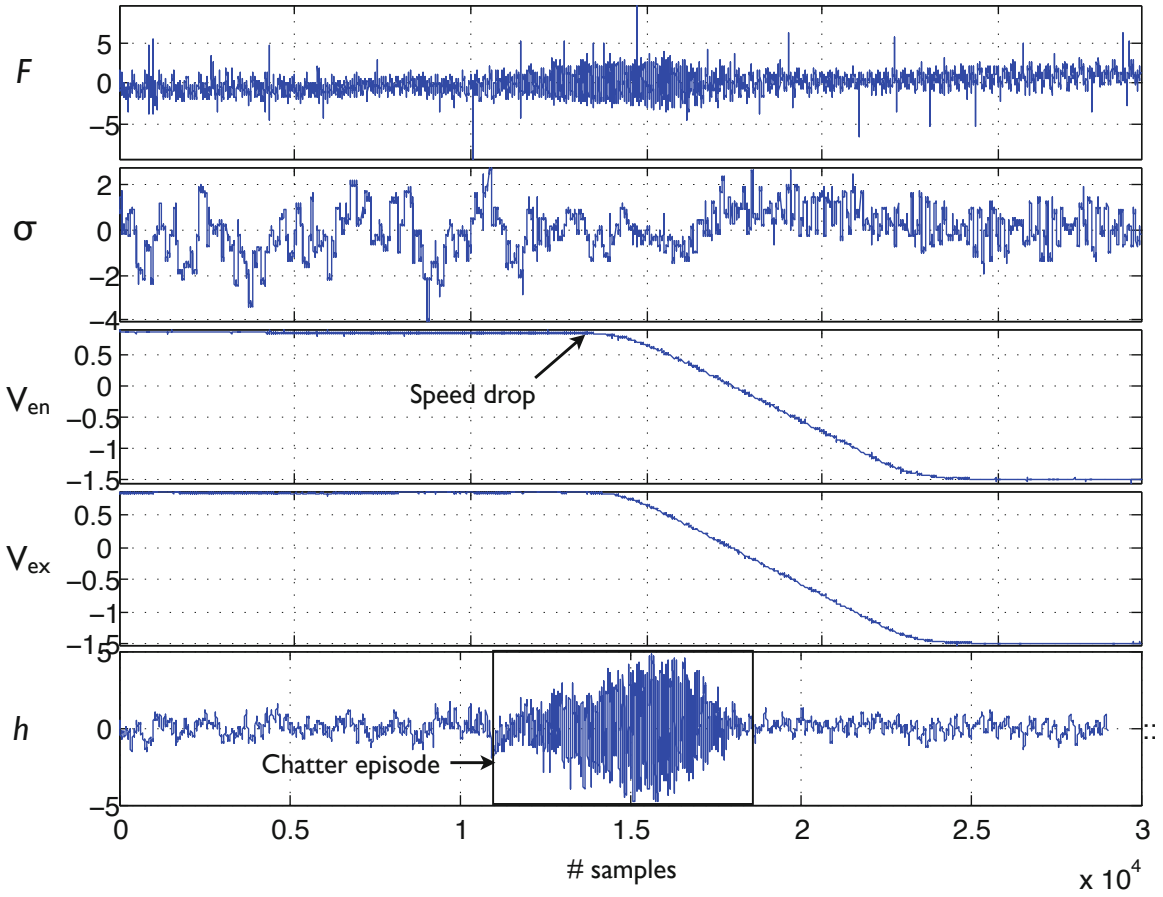


Fig. 3. Training dataset to model the process. When chatter occurs, a drop in the speed in order to reduce its effect can be seen. After achieving a proper speed level, the effect is mitigated.

For the sensitivity analysis the signals were windowed into segments of $L = 1000$ elements with an overlapping of 90%. After this analysis, the Welch Periodogram was applied to each SISO model, using a Hamming window of size 125 and an overlapping of 50%. The final size of matrix $\mathbf{G}$ is 543×516. Using this matrix, the t-SNE algorithm was applied using a value of perplexity of 30 to project data into a 2D space.

The resulting map using t-SNE technique, can be seen in Fig. 4 (a), where the entry speed is used as color and size encodings of the points. The 2D map shows two main zones of points from normal behavior of the process and a subset of points, revealing a chatter condition.

The method was applied to two subsets of data from different strips, denoted as `x` and `+`, including a chatter episode. The developed out-of-sample extension allows to project this novel test data over the trained map. These new points are placed in parts of the map that corresponds to their operating conditions (Fig. 4 (b), (c) and (d)). Their positions reveal a similar behavior to the nearest neighbors of the training points. Thus, different operating conditions are mapped in different coordinates, allowing to detect the dynamical differences in the process behavior.

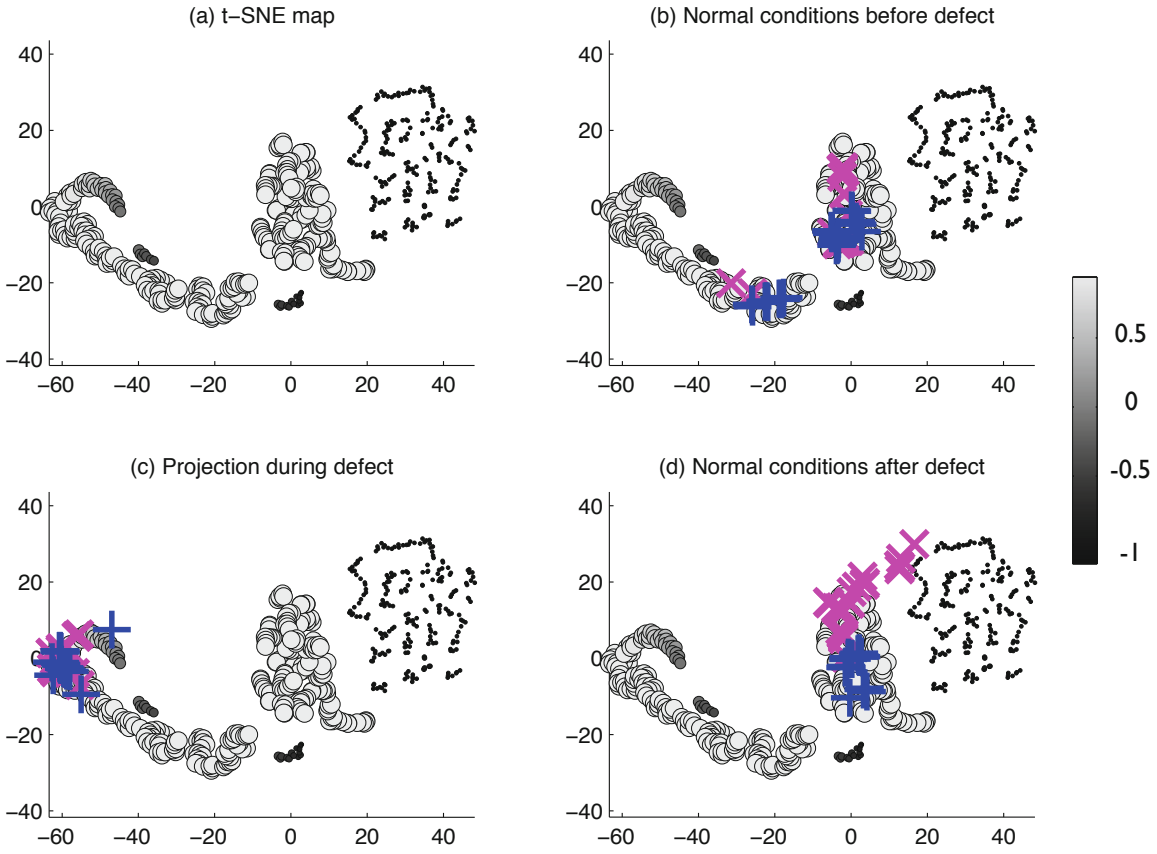


Fig. 4. (a) *t*-SNE projection, (b), (c), and (d) new points projected depending on operating condition

4 Conclusions

A spectral density estimation was applied to a model of a cold rolling process in order to analyze the chatter defect, which causes an unexpected vibration in the process. The analysis yields information about frequency response functions (FRF's), which can be considered as high-dimensionality data in order to apply a DR technique to visualize them. The algorithm applied (*t*-SNE) to a training dataset corresponding to fault situation perfectly unfolds the structure of data in the high-dimensional space, giving clear insights of this sort of faults. The method accurately defines different zones of the process, distinguishing between normal operating and chatter conditions.

The developed out-of-sample approach provides the possibility to project new data points in order to supervise the existence of problems in the process. The visualization of new projected points in the map accelerates the fault detection and helps to predict this type of faults. This procedure can be time consuming depending on the size of the dataset because once computed the map for training data the new projections are computed for each point. An alternative which reduces computational burden is to use a Nadaraya-Watson regression model [26], which estimates new projections of data based on observed values from training data.

Although it is applied to data from a cold rolling facility, this method is suitable for application to other industrial processes, whose main defects could be detected by analyzing their dynamic behavior.

Acknowledgments. This work has been financed by a grant from the Government of Asturias, under funds of Science, Technology and Innovation Plan of Asturias (PCTI), and by the spanish Ministry of Science and Education and FEDER funds under grants DPI2009-13398-C02-01.

References

1. Roberts, W.L.: Cold rolling of steel. Marcel Dekker, Inc., New York (1978)
2. Yun, I.S., Wilson, W.R.D., Ehmann, K.F.: Review of chatter studies in cold rolling. International Journal of Machine Tools and Manufacture 38(12), 1499–1530 (1998)
3. Hu, P.H., Ehmann, K.F.: A dynamic model of the rolling process. part I: homogeneous model. International Journal of Machine Tools and Manufacture 40(1), 1–19 (2000)
4. Cuadrado, A.A., Diaz, I., Diez, A.B., Obeso, F., Gonzalez, J.A.: Visual data mining and monitoring in steel processes. In: 37th IAS Annual Meeting, Conference Record of the Industry Applications Conference, vol. 1, pp. 493–500 (2002)
5. Díaz, I., Domínguez, M., Cuadrado, A., Fuertes, J.: A new approach to exploratory analysis of system dynamics using som. Applications to industrial processes. Expert Systems with Applications 34(4), 2953–2965 (2008)
6. Kourti, T., MacGregor, J.: Process analysis, monitoring and diagnosis, using multivariate projection methods. Chemometrics and Intelligent Laboratory Systems 28(1), 3–21 (1995)
7. Lee, J.A., Verleysen, M.: Nonlinear dimensionality reduction (2007)
8. Pearson, K.: LIII. On lines and planes of closest fit to systems of points in space. Philosophical Magazine Series 6 2(11), 559–572 (1901)
9. Hyvärinen, A., Karhunen, J.: Independent component analysis (2001)
10. Kohonen, T.: Self Organizing Maps. Springer (1995)
11. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323 (2000)
12. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
13. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15(6), 1373–1396 (2003)
14. Van Der Maaten, L., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research 9, 2579–2605 (2008)
15. Bushati, N., Smith, J., Briscoe, J., Watkins, C.: An intuitive graphical visualization technique for the interrogation of transcriptome data. Nucleic Acids Research 39(17), 7380–7389 (2011)
16. Jamieson, A.R., Giger, M.L., Drukker, K., Li, H., Yuan, Y., Bhooshan, N.: Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. Medical Physics 37(1), 339–351 (2010)
17. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: Theory and applications. Neurocomputing 70(1-3), 489–501 (2006)

18. Venter, R., Abd-Rabbo, A.: Modelling of the rolling process–I: Inhomogeneous deformation model. International Journal of Mechanical Sciences 22(2), 83–92 (1980)
19. Freshwater, I.: Simplified theories of flat rolling–I: the calculation of roll pressure, roll force and roll torque. International Journal of Mechanical Sciences 38(6), 633–648 (1996)
20. Paton, D.L., Critchley, S.: Tandem mill vibration: Its cause and control. In: Mechanical Working; Steel Processing XXII, Proceedings of the 26th Mechanical Working; Steel Processing Conference, pp. 247–255. Iron and Steel Soc. Inc., Chicago (1985)
21. Meehan, P.A.: Vibration instability in rolling mills: Modeling and experimental results. Journal of Vibration and Acoustics 124(2), 221–228 (2002)
22. Kimura, Y., Sodani, Y., Nishimura, N., Ikeuchi, N., Mihara, Y.: Analysis of chatter in tandem cold rolling mills. ISIJ International 43(1), 77–84 (2003)
23. Venkatasubramanian, V.: A review of process fault detection and diagnosis Part III: Process history based methods. Computers & Chemical Engineering 27(3), 327–346 (2003)
24. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723 (1974)
25. Theil, H.: Economics and information theory. North-Holland Pub. Co., Rand McNally, Amsterdam, Chicago (1967)
26. Simon, H.: Neural networks: a comprehensive foundation. Prentice Hall (1999)

A.5 INTERACTIVE VISUALIZATION AND FEATURE TRANSFORMATION FOR MULTIDIMENSIONAL DATA PROJECTION [151]

Eurographics Conference on Visualization (EuroVis) (2013)
M. Hlawitschka and T. Weinkauf (Editors)

Interactive Visualization and Feature Transformation for Multidimensional Data Projection

D. Pérez [1], L. Zhang [2], M. Schaefer[2], T. Schreck[2], D. Keim [2] and I. Díaz[1]

[1]University of Oviedo, Spain
[2]University of Konstanz, Germany

Abstract

Projecting multidimensional data to a lower-dimensional visual display as a scatter-plot-like visualization is a common approach for analyzing multidimensional data. Many dimension reduction techniques exist for performing such a task, but the quality of projections varies in terms of both preserving the original data structure and avoiding cluttered visual displays. In this paper, we propose an interactive feature transformation approach that allows the analyst to monitor and improve the projection quality by transforming feature space and assessing/comparing the quality of different projection results. The method integrates feature selection and transformation as well as a variety of projection quality measures to help analyst generate uncluttered projections that preserve the structural properties of the data. These projections enhance the visual analysis process and provide a better understanding of data.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

Projection-based data analysis and visualization is widely used for identifying patterns in multidimensional data. The idea is to project each data item (object) as a point to a two or three dimensional visual display in such a way that similar items are close to each other and dissimilar ones are far apart, result in a scatterplot-like visualization where structures and patterns can be analyzed. The projection is usually achieved by a *Dimension Reduction* (DR) technique that tries to best approximate the distance (similarity) between items in high-dimensional data space to the low dimensional visual display. A large number of DR methods exist [LV07], and one critical part of the technique is the distance measure. Multidimensional data often contains dimensions that are irrelevant to the analysis task, values in these dimensions introduce noise to the distance measure and obscure real distances between objects. Using such inaccurate distance measures may hide the real structure of the data as well as meaningful patterns. To reduce the noise in data, a number of interactive dimension selection and feature transformation techniques have been proposed [JJ09, SZS*13]. These approaches either filter out the noise by selecting relevant dimensions manually or automatically, or reduce the influence of noisy dimensions via feature transformation.

The requirements for evaluating the resulting projections lead to the definition of *quality measures* that help the analyst to understand how well the distances are approximated in the projection. Apart from measures that take into account structural preservation [Sam69, LV08], a set of *visual* quality measures has also been developed [SZS*13, BTK11]. While the techniques and measures provide means of generating meaningful embeddings of multi-dimensional data and assess their quality from different perspectives, existing projection approaches lack the flexibility of integrating interactive visualization and feature transformation mechanism to steer the projection process and improve its quality. Recent advances in the field include interactive approaches [JZF*09, CLKP10] that integrate the human expert in the analysis process and help to understand multidimensional data, as well as an improvement of class separation in projections by means of transforming feature space [SZS*13]. The work reported in this paper advances the above mentioned approaches by combining the strength of both interactive user feedback and feature transformation for generating better quality visual embeddings of multidimensional data.

The main contribution of this paper is a novel visual analytics approach that combines interactive visualization, dimension selection, feature transformation, and quality evaluation for improving the quality of multidimensional data projection. The reminder of this paper is organized as follows. In Section 2 we discuss related work, in Section 3 we explain the details of the proposed approach, in Section 4 we demonstrate the effectiveness of the method with real data, and finally, in Section 5 we draw conclusions and discuss future work.

2. Related work

2.1. Feature transformations and interactive analysis

Feature selection and transformations have been developed to improve performance of many applications in several research fields [BL97, GE03]. A recent approach [SZS*13] transforms the feature space by extending specific feature of selected dimensions. The result can be applied to improve group separation and reduce visual cluttering in the final embedding.

DR techniques estimate the underlying structure and reveal relationships in multidimensional data. However, with the increasing size and complexity of data, it becomes more difficult to generate meaningful projections in a fully automatic way. This leads to the development of *interactive multidimensional data projection* techniques that facilitate interactive analysis by integrating the analyst's knowledge about the data as well as the knowledge gained during the learning process. Examples include the iPCA approach [JZF*09] that provides coordinated views for interactive analysis of projections computed by PCA method, the iVisClassifier system [CLKP10] improves data exploration based on a supervised DR technique (LDA). Moreover, the DimStiller framework [IMI*10] analyzes dimension reduction techniques with interactive controls that guide the user during analysis process and Dis-Function [BLBC12] provides an interactive visualization to define a distance function. A comparison of features sets are determined in [BvLBS11], and an interactive exploration can be made for the selection of the suitable data descriptors.

The above mentioned techniques show that a rich body of research exists on multidimensional data visualization. However, integrating human knowledge to the analysis loop to improve the quality of visual embedding remains a challenge.

2.2. Quality Metrics

Despite the large number of DR techniques that have been developed, the question of quality assessment of a given projection has remained mostly unanswered until recent years [BTK11].

The first measures to assess the quality of a projection are the so called *stress* and *strain* measure [Sam69, Kru69]. These measures come from objective functions of nonlinear DR techniques, and assess the quality of structural preservation with the differences of the Euclidean distances between pairwise objects in a low-dimensional embedding approximate and the corresponding distances in high-dimensional data space.

While *strain* and *stress* measures analyze the preservation of global structure of data, the *trustworthiness* and *continuity* measure [VK01] and the *K-ary neighborhoods* measure [LV08] assess the quality of a projection in a broader applicability, taking into consideration also the small neighborhood preservation. In the case of labeled data, the classification error is a typical choice, see for instance [SR03] and other references in [VK07]. The integration of classification error measures in the DR technique leads to better group separation in the final embedding.

Apart from the *structural preservation quality measures* mentioned above, a set of *visual quality measures* has also been developed. Examples include *Histogram Density Measure* that ranks scatter plot visualizations, and the *Class Density Measure* that assess class separation of a given projection [TAE*09]. Moreover, the *overlap measures*, defined in [SZS*13], compute the overlap area between groups and overlap object density in a multidimensional data projection.

3. Method

In this paper, we propose a multidimensional data projection framework that combines the strength of the feature transformation approach [SZS*13], the interactive parameter setting and visualization to help analyst achieve uncluttered projections. The main workflow of the framework is shown in Figure 1. First of all, given a multidimensional dataset, with labels that define the containing classes. An initial projection is generated by a selected DR technique. The interactive

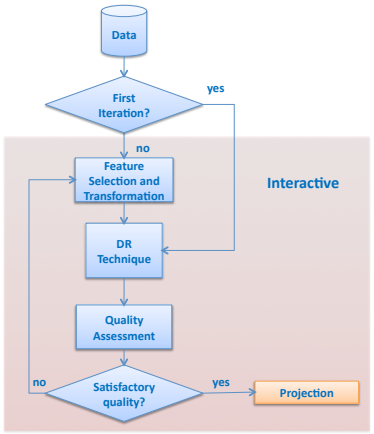


Figure 1: *Workflow of the method*

visualization panel allows the analysts to select dimensions for feature extension based on the data distribution and their knowledge about the data. After that, the system will transform the data by extending the mean values of each class for each variable selected. The DR technique is applied again to the transformed data for generating a new projection. The quality of both projections will be evaluated with quality measures and can be compared to select the one that has better quality. The analysts can iteratively repeat the process until a satisfactory projection is achieved.

3.1. Interactive Visualization for Dimension Selection

Feature selection can be performed with diverse criteria. In an automatic way, it can be used the *range* of data values over a dimension using the labels with categorical information. An interactive approach can be performed by parallel coordinates visualization which shows global data distribution over all dimensions with different color for each class. This view can help the analyst identify dimensions that provide clear distinctions between different classes. For example in Figure 2, from the parallel coordinates visualization it is not difficult to find out that in the 5th dimension, data items that belong to the same class have similar values and data items that belong to different classes are usually different. Such visual patterns often help the analyst to identify "distinctive" dimensions in multidimensional data. The result shows that transforming certain features relates to these distinctive dimensions often helps achieving better quality projection [SZS*13].

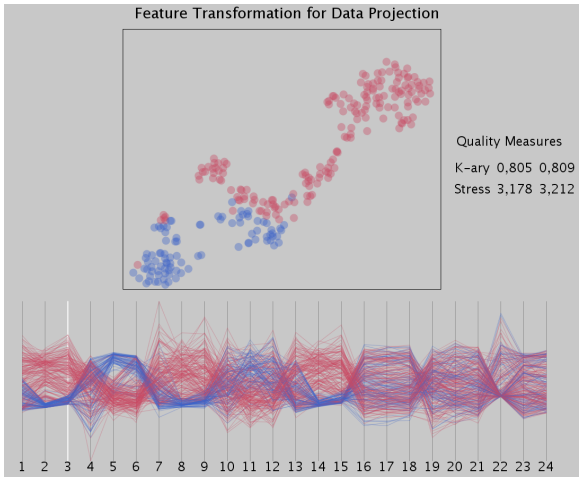


Figure 2: *Screenshot of the prototype tool*

Due to the scalability of the parallel coordinates visualization, a previous process should be considered to generate features for complex datasets.

3.2. Feature Transformation

The basic idea of the feature space transformation is to extend the selected features by adding the mean values of each class. Considering multidimensional dataset as a matrix $\mathbf{D}$ where rows are data items and columns are features, and class labels $\mathbf{c}$ are given to the class of the i-th row.

$$\mathbf{D} = \left[d_{ij}\right] \in \mathbb{R}^{m \times n} \qquad \mathbf{c} = [c_i] \in \mathbb{N}^m \tag{1}$$

With $i = 1,\ldots,m$ and $j = 1,\ldots,n$, being m the number of feature vectors and n the number of features. If one feature f is selected, the extended data table $\mathbf{D}'$ is defined as follows,

$$\mathbf{D}' = \left[d_{ij} \mid m^f_{c_i}\right] \in \mathbb{R}^{m \times (n+1)} \tag{2}$$

being $m^f_{c_i}$ the mean value of all the items corresponding to the class label c_i in the feature f.

The maximum number of extended features could be the whole set of variables. Although using this selection the result leads to a clear group separation, the similarity preservation between groups objects is damaged. Besides this simple extension strategy, a feature space can be transformed in many different ways. For example, *median* or *mode* could be applied instead of the mean value.

4. Experiments and Results

In this section the proposed approach is shown on multidimensional data with class labels from a real case. The data consists of measures of electrical and environmental variables, collected during a whole year at one university building. The task is the identification of different types of daily consumption patterns in that building. The variables that were used are: *voltage, current, apparent power, power factor, neutral current, temperature, humidity and solar radiation*. The day is divided into three shifts of eight hours each, and characterized with the average value of each shift for each variable, so that each item represents a day. Therefore the data matrix is composed by the days (items with missing values were removed) and 24 features (8 variables x 3 shifts). The used label has two classes depending on whether it is working day or holiday such as weekends.

To validate this approach, a prototype tool has been developed (see Figure 2) which displays both the projection and the parallel coordinates views with color representing labels. The parallel coordinates view helps to decide the best choices over all features. In this case, the automatic feature selection corresponds to the maximum range between mean values for each class of the whole set of attributes. Although this selection recommends using feature five, the extension of the dimension eight obtains a similar map with better quality measures.

The projections of the original and transformed data are computed with the same dimensionality reduction technique. The techniques used were t-SNE method [vdMH08], that

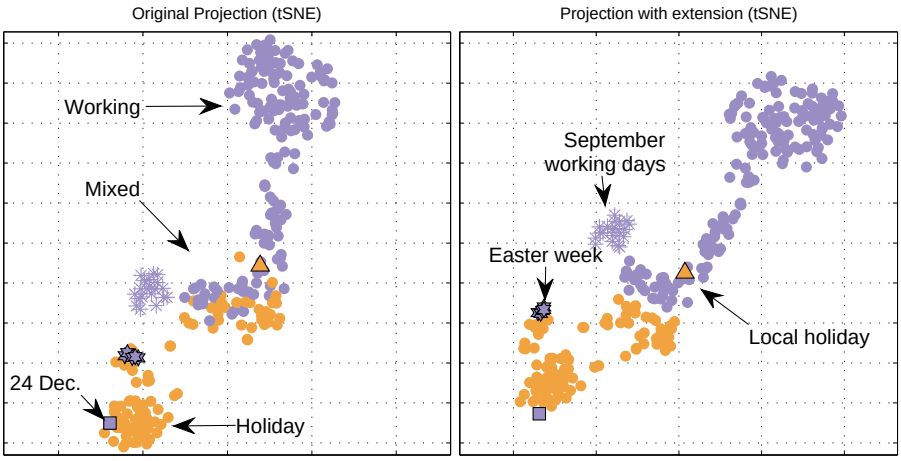


Figure 3: *Original (Left) and extended (Right) projections of daily consumption for one building with t-SNE technique. Color represents class labels (holiday/working day) and shape refers to highlighted items.*

is an effective unsupervised technique for visualizing data, and a supervised technique, *Maximally Collapsing Metric Learning* (MCML) [GR06], in order to use the label information available for computing the embedding. Notice that the transformation is independent of the DR technique chosen. The transformation performed was the extension of the selected dimension with the mean values for each class.

In the projection with the original feature vectors two daily patterns, of high and low consumption, are easily identified, clearly related to working day and holiday, respectively. But there is a third pattern in the middle, with both types of days mixed (see Figure 3, left), which is not easy to identify. The projection with the extension reveals similar daily patterns with a clearer class separation that improves the recognition of the label information in that mixed area (see Figure 3, right). For example, it is easy to distinguish, in the extended projection, a point of a local holiday, that stays close to the working days, revealing similar consumption these days in the building.

Finally the performance of the projections is evaluated by the quality measures previously described. The stress measure is referred to the Sammon's error [Sam69], k-ary neighborhood can be found in [LV08], and the overlap measures are formally defined in [SZS*13]. The values of these measures used are described in Table 1 for this example. These evaluation measures show an enhancement of the projection quality in the extended case.

Table 1: *Assessment measures for the projections*

t-SNE				
Feat. Ext.	**k-ary**	**Stress**	**Overlap area**	**Overlap density**
None	0.80	3.04	0.024	$7 \cdot 10^{-3}$
5	**0.81**	3.02	0.029	$1 \cdot 10^{-3}$
8	**0.81**	**2.95**	$\mathbf{6 \cdot 10^{-5}}$	$\mathbf{9 \cdot 10^{-4}}$
MCML				
Feat. Ext.	**k-ary**	**Stress**	**Overlap area**	**Overlap density**
None	0.6462	0.3953	0.063	$6 \cdot 10^{-4}$
5	0.6836	0.3477	**0**	**0**
8	**0.6838**	**0.3474**	**0**	**0**

5. Conclusions

In this paper we propose an interactive visualization framework for improving existing data projections. The method transforms multidimensional data by extending selected features from original data, introducing the human into the analytical loop and utilizing their perception power and domain knowledge. A case with real datasets was conducted to test the effective of the approach. With both supervised and unsupervised DR techniques, through interactive dimension selection and feature transformation, we can achieve projections with improved quality. These projections provide efficiency to pattern recognition, fast identification of class labels and understanding of data. The improvement of the projection is independent of the DR technique that are chosen to perform the projection, having the same scalability limitations that the technique itself.

As future work we would like to explore more visualization techniques for assisting feature selections, new transformation strategies for noise elimination, and wider range of quality measures for evaluating the projections.

Acknowledgments

This work has been financed by the Spanish Ministry of Science and Education and FEDER funds under grants DPI2009-13398-C02-01/02

References

[BL97] BLUM A., LANGLEY P.: Selection of relevant features and examples in machine learning. *Artificial intelligence 97*, 1 (1997), 245–271.

[BLBC12] BROWN E., LIU J., BRODLEY C., CHANG R.: Disfunction: Learning distance functions interactively. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (2012), pp. 83–92.

[BTK11] BERTINI E., TATU A., KEIM D.: Quality metrics in high-dimensional data visualization: An overview and systematization. *Proceedings of the IEEE Symposium on IEEE Information Visualization (InfoVis) 17* (2011), 2203–2212.

[BvLBS11] BREMM S., VON LANDESBERGER T., BERNARD J., SCHRECK T.: Assisted descriptor selection based on visual comparative data analysis. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 891–900.

[CLKP10] CHOO J., LEE H., KIHM J., PARK H.: ivisclassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on* (oct. 2010), pp. 27 –34.

[GE03] GUYON I., ELISSEEFF A.: An introduction to variable and feature selection. *The Journal of Machine Learning Research 3* (2003), 1157–1182.

[GR06] GLOBERSON A., ROWEIS S.: Metric learning by collapsing classes. *Advances in neural information processing systems 18* (2006), 451.

[IMI*10] INGRAM S., MUNZNER T., IRVINE V., TORY M., BERGNER S., MÖLLER T.: Dimstiller: Workflows for dimensional analysis and reduction. In *Proc. IEEE Conf. Visual Analytics Science and Technology (VAST)* (2010), vol. 1, Citeseer.

[JJ09] JOHANSSON S., JOHANSSON J.: Interactive dimensionality reduction through user-defined combinations of quality metrics. *Visualization and Computer Graphics, IEEE Transactions on 15*, 6 (2009), 993–1000.

[JZF*09] JEONG D. H., ZIEMKIEWICZ C., FISHER B., RIBARSKY W., CHANG R.: iPCA: an interactive system for PCA-based visual analytics. *Computer Graphics Forum 28*, 3 (June 2009), 767–774.

[Kru69] KRUSKAL J.: Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new 'index of condensation'. In *Statistical Computation*, Milton R., Nelder J., (Eds.). Academic Press, New York, 1969, pp. 427–440.

[LV07] LEE J., VERLEYSEN M.: *Nonlinear dimensionality reduction*. Springer, 2007.

[LV08] LEE J., VERLEYSEN M.: Quality assessment of nonlinear dimensionality reduction based on k-ary neighborhoods. In *JMLR Workshop and Conference Proceedings (New challenges for feature selection in data mining and knowledge discovery)*, Saeys Y., Liu H., Inza I., Wehenkel L., Van de Peer Y., (Eds.), vol. 4. Sept. 2008, pp. 21–35.

[Sam69] SAMMON J. W.: A nonlinear mapping for data structure analysis. *IEEE Trans. Comput. 18*, 5 (May 1969), 401–409.

[SR03] SAUL L., ROWEIS S.: Think globally, fit locally: Unsupervised learning of nonlinear manifolds. *Journal of Machine Learning Research 4* (June 2003), 119–155.

[SZS*13] SCHAEFER M., ZHANG L., SCHRECK T., TATU A., LEE J. A., VERLEYSEN M., KEIM D. A.: Improving projection-based data analysis by feature space transformations. In *Proceedings of the SPIE Visualization and Data Analysis 2013 (VDA2013)* (2013).

[TAE*09] TATU A., ALBUQUERQUE G., EISEMANN M., SCHNEIDEWIND J., THEISEL H., MAGNOR M., KEIM D.: Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)* (2009), pp. 59–66.

[vdMH08] VAN DER MAATEN L., HINTON G.: Visualizing data using t-SNE. *Journal of Machine Learning Research 9* (2008), 2579–2605.

[VK01] VENNA J., KASKI S.: Neighborhood preservation in nonlinear projection methods: An experimental study. In *Proceedings of ICANN 2001*, Dorffner G., Bischof H., Hornik K., (Eds.). Springer, Berlin, 2001, pp. 485–491.

[VK07] VENNA J., KASKI S.: Nonlinear dimensionality reduction as information retrieval. In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (AISTATS 2007)*, Meila M., Shen X., (Eds.). Omnipress, San Juan, Puerto Rico, Mar. 2007, pp. 568–575.

A.6 POWER-CONSUMPTION ANALYSIS THROUGH WEB-BASED VISUAL DATA EXPLORATION [147]

Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Power-Consumption Analysis through Web-Based Visual Data Exploration

Daniel Pérez * **Ignacio Díaz** * **Abel A. Cuadrado** *
Francisco J. García-Fernández * **Alberto B. Diez** *
Manuel Domínguez **

* *Electrical Engineering Department, University of Oviedo, Campus de Viesques s/n, Gijón, Spain, 33204, (e-mail: {dperez, idiaz, cuadrado, fjgarcia, alberto}@isa.uniovi.es).*
** *Instituto de Automática y Fabricación, University of León, Campus de Vegazana, León, 24071(e-mail: manuel.dominguez@unileon.es)*

Abstract:
In recent years, the increasing capabilities of current technologies have made the acquisition and storage of large datasets an easy task. However, most times their unmanageable size, along with their complexity makes it a challenge to handle and analyze them. The emerging field of *visual analytics* relies on visualization principles and interactive interfaces to provide ways for amplifying human cognition, so that the information processing is improved and a better understanding of these datasets is achieved. In addition, current technologies to develop visual interfaces allow to develop powerful interaction mechanisms that enable the user to be an active part in the process. In this paper, data visualization foundations and web-based methods are considered for user-driven supervision tasks in decision-support systems. A real case consisting in the analysis of electric power demand in university buildings is presented via a web application developed using the recent data visualization library called D3. Time-series visualization and similarity maps are represented along with interactive techniques, allowing a dynamic data exploration.

Keywords: smart grids, monitoring and performance assessment, visual pattern recognition, supervision, energy expenditure, electric power systems, decision support systems

1. INTRODUCTION

Energy plays a key role in our society and makes it move. A proper management of energy, involving generation, distribution and demand, has a great impact in economy, but also on environment and health. Despite many efforts are being done in sustainable, renewable and clean energy production, acting on demand through a rational use of energy is probably one of the best niches to improve the energy balance. Most of the times, energy waste is motivated by a lack of control about how, when and where it is consumed.

Recent technologies and lower costs have led to an increasing number and quality of installed sensors that measure the power demand. However, most of the times this information ends up in a database or gets plotted in a simple trend graphic that shows only aggregate information and gives no insight about inefficiencies on the demand. In many cases, important decisions to achieve an efficient use of energy are taken with tight time by politicians or staff in charge of it, and long reports or spreadsheets are often misread. Information and knowledge about the patterns of demand in a public building or facility should be presented in a clear and concise way, so that the end user can quickly assess the demand condition on a given period and be assertive to make decisions or take corrective actions.

Currently, manipulating and analyzing *big data* are challenging tasks in several fields. One interesting approach to deal with large datasets is *visual analytics* that focuses on analytical reasoning facilitated by interactive visual interfaces, as is explained in Thomas and Cook [2005]. The use of certain visualization techniques supports the powerful human cognition to perform an efficient analysis. Visualization allows humans to extract knowledge from data, and reveal patterns (to support hypotheses or even to create new ones). Moreover, it provides not only a better understanding of the data but also an excellent via for communicating ideas to the rest, see Shneiderman [1996], Card et al. [1999], Ware [2012].

In addition, by means of interaction techniques, the integration of the human in the data analysis loop allows the use of the available domain knowledge and his/her own perception; an example is shown in Ahlberg and Shneiderman [1994]. A user-driven data analysis by means of interaction techniques helps the user to focus on the interesting aspects of data and improves the efficiency of the exploration with respect to a static view.

In the last years, web-based methods have become a powerful framework for implementing interactive displays by using the *document object model* (DOM). Recently, in Bostock et al. [2011] a novel approach, consisting of a javascript library called D3, was proposed allowing the

direct manipulation of DOM, binding data to its elements, and animating them easily. It provides an efficient basis to develop dynamic visualizations on the web.

This paper aims at drawing the attention of the community of control to consider data visualization principles and techniques in applications such as supervision of complex processes or decision-support systems using web-based methods. An example built using D3 for power demand analysis in public buildings is presented here as a validation of the proposed approach. The remainder of the paper is: in Section 2 the requirements for an efficient analysis of power consumption are detailed; in Section 3 several works and visualization concepts are described; in Section 4 some tools are briefly reviewed, and the one used here is explained; in Section 5 a real case is shown for validating the previously explained techniques, and finally in Section 6 conclusions are summarized.

2. DESIGN REQUIREMENTS

Several features are desirable in order to exploit the user's knowledge for an efficient interpretation of power demand data.

A common problem in the analysis of power demand of a public building is the concurrence of different types of periodicities as well as many non-regular periods that stem from social activity, such as special events —like football matches, strikes, . . . — , holidays, etc . In addition, the power demand in a building can suffer variations due to weather or periods related to the nature of the activities carried out in it —e.g. examination periods in an educational institution. All this results in changes of the 24 hour day patterns of demand. The interface should provide an interactive, smooth and visual way to combine views showing different kinds of periodicities. For instance, it is desirable to know at a glance the hourly or yearly distribution of the demand on a given weekday. Despite this can be done using database query methods or spreadsheet operations, they imply a textual interaction that makes the exploratory process less interactive and far more ineffective.

The user should also be able to deal simultaneously with several views that tell different aspects of the problem. Existing visualizations do this mainly in two ways: a) showing all the views in the same screen, with a limited capacity to show many of them or b) allowing the user to add, remove views, or commute between them, which requires the user to locate again the interesting points or landmarks identified in the previous view. Despite methods such as *selection* and *brushing* can help to overcome this problem, changing between views remains being a "discontinuous jump" that obliges the user's mind to recompute visual landmarks that point to interesting information.

Finally, in approaches that use cluster analysis methods, like Van Wijk and Van Selow [1999], the selection of prototype day patterns is done in an automated way, with a low degree of human intervention. Since in power demand analysis many kinds of knowledge and information have to be managed –the structure of tariffs, the particular type of activity in the analyzed building, etc.– a more user-centric approach in the classification of daily demand profiles is a desirable feature.

In light of this, the interface should efficiently help the user to:

- *Identify regular temporal patterns.* The user should be able to quickly deploy a view showing the demand organized by hours of a day, by weekday or along the whole year. The user should also be able to get *combined views* as, for instance, hourly demand for each of the seven days of the week.
- *Identify special temporal patterns* such as holidays or singular events. One requirement for this is the use of specialized views such as a *calendar layout* where the spatial organization of time helps in finding special events (e.g. Easter holidays).
- *Identify groups of similar day patterns* of demand. Despite clustering methods can efficiently partition the data into groups of similar day patterns, the user may lose perspective on which clusters are similar to others and to what extent. Projections of the data on a 2D map can reveal continuous variations in the daily demand organized according to its similarity. This is performed by estimating manifold structures in data that the cluster methods would not describe properly.
- Establish mental *links* or *connections* between views. This can be achieved by multiple selection being displayed in multiple views. However, it should additionally strengthen these connections by means of a visual tracking mechanism that eliminates in a natural way the need to recognize specific points or selections of interest across the different views.

3. FOUNDATIONS

There are several works that have studied how the human perceive information. Bertin [1983], Cleveland and McGill [1984], Mackinlay [1986] are some representative examples in this field. Some books have also been written about this topic, some of them are Ware [2012] that shows a description of how human perception and cognition work; in Tufte [2001], several design principles for visual displays are proposed that are widely used by visualization practitioners; and in Few [2006], where dashboard design is exposed in detail.

There are different types of attributes inside a set of data, such as categorical, ordinal, or quantitative. Many visual channels can encode information such as position, color, or size, etc. Spatial position is the most accurate channel for the three data attributes, but the accuracy for other visual channels strongly depends on the data attribute to be represented.

3.1 Interaction and animation

Interaction is the essential part in the conception of data visualization interfaces, that provides dynamics to a static picture, allowing the user to manipulate representations in several ways and focus on the interesting aspects of data.

In Shneiderman [1996], the author proposed a taxonomy of low-level interaction techniques, including his mantra:

"*Overview first, zoom and filter, then details on demand*". There are more different works that put attention to interaction Yi et al. [2007]. Since reviewing the existing mechanisms is beyond the scope of this paper, we aim to show a brief description of the most relevant ones, used in this work:

- *Zooming* and *panning*, are simple affine transformations that allow the exploration of a large dataset by changing the scope and view in the same visual encoding.
- *Context information* allows the information related to one or several items to be checked by placing the mouse over them.
- *Brushing* relates the selection of a subset of elements in order to highlight them by a visual channel (color is usually used).
- *Linking* allows to provide a connection between the visualizations by highlighting in different views a subset of points or other visual elements –selected, for example by means of brushing in one of the views.
- *Animated transitions* is a method for perceiving changes between different visual encodings. Motion not only engages the viewer to different points of interest but also allows to track objects by means of their changes in order to communicate relationships between them. In Tversky et al. [2002] principles for effective animation are suggested and its benefits and drawbacks are discussed. In addition, in Heer and Robertson [2007] animated transitions are explored in order to improve graphical perception in statistical data graphics. Recently a framework that combines continuous transitions between 2D visual encodings with several interaction techniques was proposed in Diaz-Blanco et al. [2012]. Using this approach, a transition that depends on a single parameter $\lambda \in [0, 1]$, is possible between two views revealing meaningful intermediate states between them.

3.2 Reduction of data

Large datasets can produce visual cluttering that makes the user interpretation very difficult. One approach for reducing the size of data is *aggregation*, that consists in a single visual element representing a summary of many items. In cases of many dimensions, other strategy is *dimensionality reduction* (DR), that estimates the underlying structure of the multidimensional data into a reduced group of dimensions, that may be a combination of the original variables. For a reduction to a two dimensional space, the data can be represented in a projection of all the samples preserving similarity.

Many DR techniques have been developed, firstly the widely used principal component analysis (PCA) that linearly reduces the dimensions with maximum variance. Later, several methods have been used by non-linear approaches, such as neural networks like in Hinton and Salakhutdinov [2006] and probabilistic computations like in van der Maaten and Hinton [2008]. A detailed description of these methods can be found in Lee and Verleysen [2007].

4. METHODS

4.1 Data visualization tools

There are different types of data visualization tools, for instance software applications that produce interactive visualizations where your own set of data can be used, such as Tableau –Tableau [2013] –or sites like Many Eyes – Viegas and Wattenberg [2013]. Also Java-based graphic libraries, like the toolkit presented in Fekete [2004], Prefuse in Heer et al. [2005], or the Processing programming language, originally proposed in [Fry, 2004], can be used for supporting the implementation of advanced visualizations.

Regarding web-based tools, several toolkits have been developed. Examples are Processing.js (Processingjs [2013]), directly related to Processing and designed for the web, or Flare (Flare [2013]), that run in the Adobe Flash Player and was adapted from Prefuse. Moreover, Protovis which was proposed in Bostock and Heer [2009], provides a framework to map data attributes to visual elements. These examples show a variety of proposals for the development of web-based applications.

There are many advantages in web applications over desktop products:

- They are platform independent and no installation is needed, requiring only a modern web browser to run.
- They can combine the use of different technologies, such as widely accepted standards (HTML5, CSS3 or SVG). This allows the programmer to harness all their potential and resources, including lots of tunable controls, properties and events, to build the interface, as well as to benefit from a huge variety of open source libraries, ranging from date manipulation to numerical analysis.
- The highly optimized javascript interpreters built in today's browsers make the performance of these applications comparable to equivalent desktop applications.

On the downside, web-based approaches that run in the client side make it difficult to hide data and other resources to the end user in case they are confidential.

Similar to Protovis, in Bostock et al. [2011], the recent data visualization library called Data-Driven Documents (D3) extends this approach. D3 enables the pure manipulation and transformation of the standard Document Object Model (DOM) by mapping the data directly in it. In addition, D3 operators allow actions like modify content, select elements in correspondence with data, and the use of event listeners that enables interaction, and animated transitions can be derived by a collection of several interpolators over time.

Despite D3 can become slow because of the manipulation of large number of elements, most visualizations seldom require drawing simultaneously a huge number of visual elements on the screen. While future work is needed —e.g. a set of methods related to statistics would be worthy— D3 provides a standard representation improving the performance of previous approaches.

5. USE CASE

Here we explain an example for a data analysis using the principles and techniques explained above. A web prototype, implemented using D3, is shown for the exploration of power consumption in two buildings at the University of Oviedo. Firstly a description of data is detailed, then different data representations are explained, and finally the included interaction techniques are described.

5.1 Power demand dataset

The original dataset was retrieved from the data logging system, covering a timespan of one year, with a sample period of 15 minutes. A 4:1 sample reduction of the dataset was done by averaging the power consumption in an hourly basis, resulting in 8760 records (365×24). The data variables consisted of active, P, reactive, Q and apparent S power consumption in two university buildings. In addition, the power factor, $\cos\varphi$, and a residual, R, were computed as

$$\cos\varphi = \frac{P}{S} \tag{1}$$

$$R = S^2 - P^2 - Q^2 \tag{2}$$

Note that the residual R should be identically zero under ideal conditions ($S^2 = P^2 + Q^2$), with no harmonic distorsions. This attribute helps in revealing deviations from ideal conditions such as nonconventional loads.

5.2 Visual encodings

The main visualization presented here is performed by using different scatterplots, where relationships are revealed efficiently. Let's consider an *encoding* as a set of N points [1],

$$\mathbf{p}_A = \{\mathbf{p}_A(1), \mathbf{p}_A(2), \cdots, \mathbf{p}_A(N)\}, \quad \mathbf{p}_A(i) \in \mathbb{R}^m \tag{3}$$

whose points $\mathbf{p}_A(i)$ are descriptive of interesting information on sample i to be spatially described, such as, for instance, the 2D position in a wall calendar or in a clock-like representation of the timestamp of sample i. Thus, this position of the points encodes similarity information but also its size and color can also encode a different attribute. The encodings used in the case presented in this paper are:

"Clock" encodings We considered daily, weekly and yearly point sets, $\mathbf{p}_D$, $\mathbf{p}_W$, $\mathbf{p}_Y$, containing 2D points distributed in a circular way, with a period of one day, one week and one year, respectively:

$$\mathbf{p}_D(i) = \left[\cos\left(2\pi\frac{h(i)}{24}\right), \sin\left(2\pi\frac{h(i)}{24}\right)\right] \tag{4}$$

$$\mathbf{p}_W(i) = \left[\cos\left(2\pi\frac{d(i)}{7}\right), \sin\left(2\pi\frac{d(i)}{7}\right)\right] \tag{5}$$

$$\mathbf{p}_Y(i) = \left[\cos\left(2\pi\frac{h(i)}{365\cdot 24}\right), \sin\left(2\pi\frac{h(i)}{365\cdot 24}\right)\right] \tag{6}$$

[1] We shall typically consider *spatial encodings* consisting of 2D points in this paper ($m = 2$). However encodings of higher dimensions can be used in case the point set describes a 3D scatter or if other visual encodings, such as color and size, are under consideration.

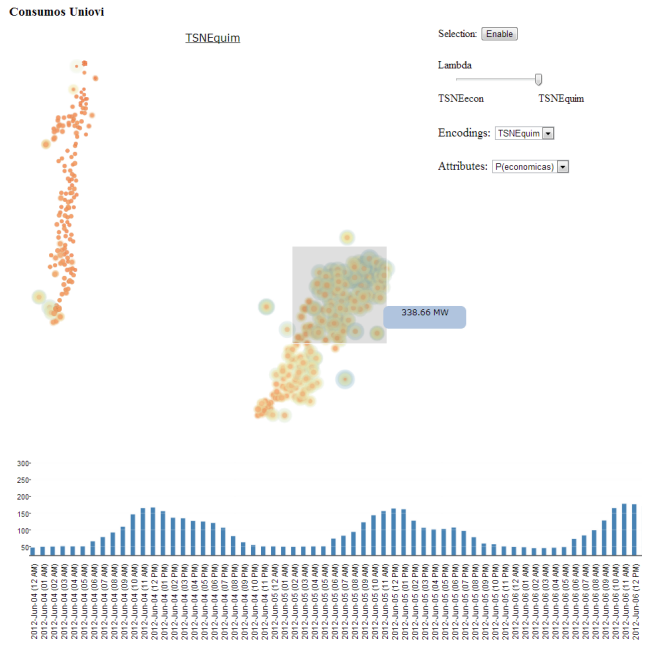


Fig. 1. Screenshot of the web-based tool showing scatterplot and barchart views

This way to encode time, already considered in related contexts in the literature —see Ohlsson et al. [1994], González and J. [2005], Shen and Ma [2008]— allows to display the data in a clock-like fashion that, on one hand, is a widely accepted convention of time representation and on the other hand provides a natural way to aggregate periodic events.

Calendar encoding A calendar encoding assigns each sample its position in a conventional wall calendar. All power demand samples corresponding to a same day will lay in the same point of the calendar in this view. The user can easily get an aggregate value of one or more days by simply selecting the points in the calendar.

The specific calendar view accounts for specific social time granularities such as the irregular number of days in a month, weekends and holiday periods. These irregular periods of time are extremely relevant to power demand analysis and cannot be properly described by classical analytical methods, such as Box-Jenkins or Fourier based methods.

365×24 *matrix encoding* Another useful encoding to represent data is a matrix representation of the items where rows show each day in the year and columns show the hours of that day. This allows the user to get, in a snapshot, variations in the 24-hour patterns of demand along the year.

Dimensionality reduction encoding The positions are the result of a 2D projection computed using a DR technique. In this case, *t-Distributed Stochastic Embedding* (t-SNE) method was used, that is an effective technique for visualizing high-dimensional data, and can be found in van der Maaten and Hinton [2008].

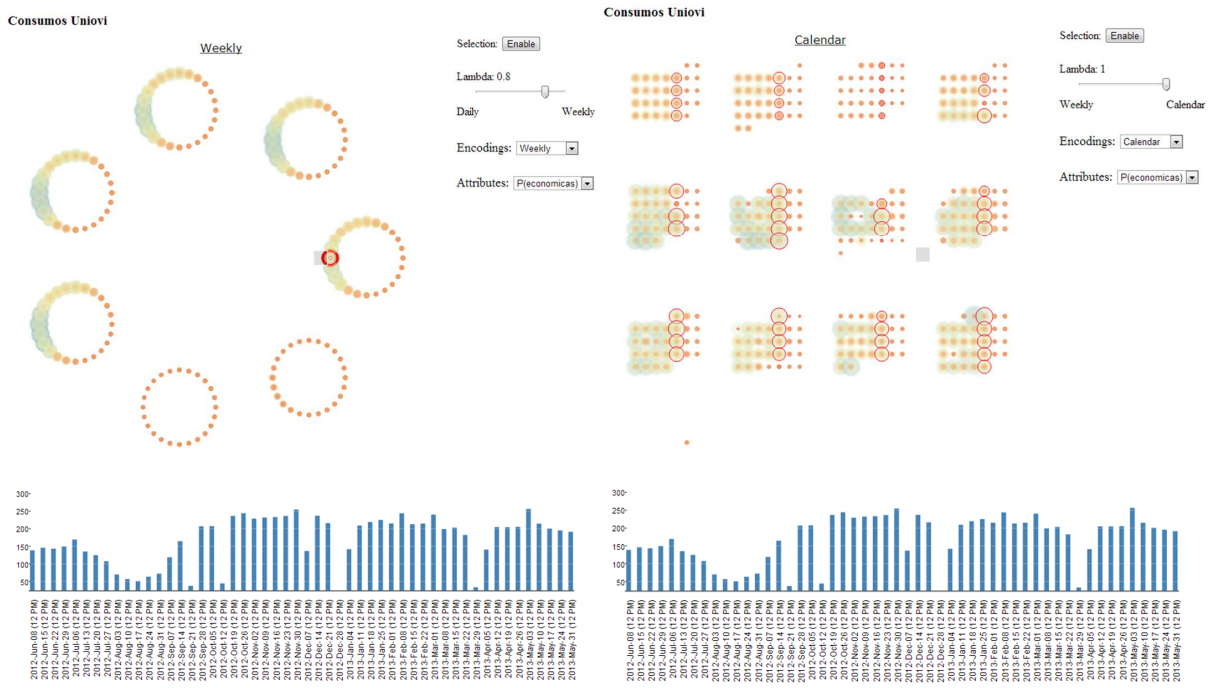


Fig. 2. Intermediate state between weekly and daily mappings (left) and calendar view with a subset of points selected (right)

5.3 Interaction techniques

D3 makes low-level interaction techniques easily available to be implemented. Transformations of the views, such as zooming and panning, improve a detailed exploration of the points. Furthermore, the user can choose several points on the map by selecting a brushing area, then a barchart visualization of the selected points is represented to perform a detailed comparison between them. A tooltip gives information about the point. If there is a selection of several points, another tooltip provides the sum of the current attribute, e.g. active power, for the selected points (see Fig. 1).

The morphing operation, explained in Diaz-Blanco et al. [2012], consists in mixing two or more encodings into a new encoding. The morphing operation between two encodings is quite straightforward. Let $\mathbf{p}_A(i)$ and and $\mathbf{p}_B(i)$ be two different encodings for $i = 1, \cdots, N$. Let also $\lambda \in [0, 1]$ be a mixing coefficient. The morphing operation between $\mathbf{p}_A$ and $\mathbf{p}_B$ would be

$$\mathbf{p}(i,t) = \lambda(t)\mathbf{p}_A(i) + (1 - \lambda(t))\mathbf{p}_B(i) \qquad (7)$$

for $i \in \{1, \cdots N\}$. Here an input control is integrated in the interface so that the user can change $\lambda(t)$ to produce a variable "mixture" of two different encodings $\mathbf{p}_A$, $\mathbf{p}_B$ by evaluating the combination for all the N points ($N = 8760$). In other words, the user can steer animated transitions by combining any two encodings with manually tunable proportions, allowing to navigate between different representations in a smooth way. This enables the user to keep visual track of the elements during the transition, establishing links between them. Moreover, a proper mixed selection results in a meaningful intermediate state. For instance, mixing a "weekday encoding" with a "daily encoding" gives rise to a new "hour of a weekday" encoding, as shown in Fig. 2 (left). This encoding, shows all 7×24 combination of days and weekdays so a given point would aggregate all power consumption along the year for a particular weekday at a particular hour.

The mix of the different visual mappings and the integration of the interaction techniques give the capability of analyzing on demand different periodical time intervals by means of visual queries. For example, in Fig. 2 (left) a visual selection (Fridays at 12 pm) can be seen with the points highlighted in red for the exploration of the consumption in these points. Furthermore, keeping this highlight in an encoding change allows linking between encodings for the selected points, see Fig. 2 (right) calendar view with the former points selected.

6. CONCLUSION

In this paper, we consider dynamic data visualizations for the exploration of complex processes and decision-support systems using web-based methods. Several visualization principles and techniques are explained for the analysis of multidimensional data. A proper visual representation of the information joined with different interaction mechanisms improve the data exploration tasks allowing the user to perform a quick analysis of large datasets. Moreover, the use of web-based tools makes the access easy for more people, and flexible for using with modern browsers. The recent javascript library D3 enables direct DOM manipulation, the native integration with different standards webs and simplifies the use of animation and interaction mechanisms.

A use case is shown using real data of power consumption acquired from two university buildings. An interactive web interface, implemented using D3, is presented where several scatterplots for the data are available. These views represent different visual encodings, such as time-series or similarities, depending on the positioning of the points whose size and color encode one certain and selectable variable of the data.

Furthermore several interaction techniques, such as zooming or brushing, are included in the application. The possibility that different information related to the selection of a subset of data can be displayed, and that animated transition between views can be controlled gradually by the user, show an efficient exploratory experience.

ACKNOWLEDGEMENTS

This work has been financed by the Spanish Ministerio de Economía y Competitividad (MINECO) and FEDER funds.

REFERENCES

Christopher Ahlberg and Ben Shneiderman. Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 313–317. ACM, 1994.

Jacques Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983. ISBN 0299090604.

Michael Bostock and Jeffrey Heer. Protovis: A graphical toolkit for visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1121–1128, 2009.

Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.

Stuart Card, Jock Mackinlay, and Ben Shneiderman. *Readings in information visualization. Using vision to think*. Morgan Kaufmann Publishers, San Francisco, 1999.

William S Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

Ignacio Diaz-Blanco, Manuel Dominguez-Gonzalez, Abel Cuadrado-Vega, Alberto Diez-Gonzalez, and Juan Fuertes-Martinez. MorphingProjections: Interactive Visualization of Electric Power Demand Time Series. Diaz-Blanco et al. [2012], pages 121–125. doi: 10.2312/PE/EuroVisShort/EuroVisShort2012/121-125.

J-D Fekete. The infovis toolkit. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 167–174. IEEE, 2004.

Stephen Few. *Information dashboard design*. O'Reilly, 2006.

Flare. `http://flare.prefuse.org/`, November 2013.

Benjamin Jotham Fry. *Computational information design*. PhD thesis, Massachusetts Institute of Technology, 2004. AAI0806331.

P. González and Zamarreño J. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, 37:595–601, 2005.

Jeffrey Heer and George G Robertson. Animated transitions in statistical data graphics. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1240–1247, 2007.

Jeffrey Heer, Stuart K Card, and James A Landay. Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM, 2005.

Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

John A. Lee and Michel Verleysen. Nonlinear dimensionality reduction, 2007.

Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.

M.B.O. Ohlsson, C.O. Peterson, H. Pi, T.S. Rognvaldsson, and B.P.W. Soderberg. Predicting system loads with artificial neural networks–methods and results from" the great energy predictor shootout". *ASHRAE Transactions-American Society of Heating Refrigerating Airconditioning Engin*, 100(2):1063–1074, 1994.

Processingjs. `http://processingjs.org/`, November 2013.

Z. Shen and K.L. Ma. Mobivis: A visualization system for exploring mobile data. In *Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific*, pages 175–182. IEEE, 2008.

Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

Tableau. `http://www.tableausoftware.com/`, November 2013.

James J Thomas and Kristin A Cook. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.

Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, second edition, 2001. ISBN 0961392142.

Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. Animation: can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

J.J. Van Wijk and E.R. Van Selow. Cluster and calendar based visualization of time series data. In *Infovis*, page 4. Published by the IEEE Computer Society, 1999.

Fernanda Viegas and Martin Wattenberg. Many eyes. `http://www.many-eyes.com/`, November 2013.

Colin Ware. *Information visualization: perception for design*. Elsevier, 2012.

Ji Soo Yi, Youn ah Kang, John T Stasko, and Julie A Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6): 1224–1231, 2007.