Discovering Relevancies in Very Difficult Regression Problems: Applications to Sensory Data Analysis.

CONFERENCE PAPER · JANUARY 2004

Source: DBLP

Discovering relevancies in very difficult regression problems: applications to sensory data analysis

Jorge Díez and **Gustavo F. Bayón** and **José R. Quevedo** and **Juan José del Coz** and **Oscar Luaces** and **Jaime Alonso** and **Antonio Bahamonde**[1]

Abstract. Learning preferences is a useful tool in application fields like information retrieval, or system configuration. In this paper we show a new application of this Machine Learning tool, the analysis of sensory data provided by consumer panels. These data sets collect the ratings given by a set of consumers to the quality or the acceptability of market products that are principally appreciated through sensory impressions. The aim is to improve the production processes of food industries. We show how these data sets can not be processed in a useful way by regression methods, since these methods can not deal with some subtleties implicit in the available knowledge. Using a collection of real world data sets, we illustrate the benefits of our approach, showing that it is possible to obtain useful models to explain the behavior of consumers where regression methods only predict a constant reaction in all consumers, what is useless and unacceptable.

1 INTRODUCTION

An important part of the success of food industries relies on their ability to produce their specialties satisfying the consumers' sensory requirements. Then, it is necessary to organize polls to discover the opinions of potential consumers about the quality or the acceptability of market products that are principally appreciated through sensory impressions. The aim of the analysis of sensory data is to process consumers' answers that can be represented as in regression problems: the description of each object x in a set E is endowed with a number $r(x)$ that assesses the degree of satisfaction for each consumer or the average value for a group of consumers.

Traditionally the process given to these data sets includes testing some statistical hypothesis [13, 12, 1]. On the other hand, the approach followed in [3] is based on the use of Bayesian belief networks. In all cases these previous approaches demand that all available food products (the objects x) must be rated by all consumers; in practice, this is an impossible assumption most of the times. In general, we will have sets of ratings $(r(x) : x \in E_i)$ for each consumer or group of consumers i, where $\cup(E_i : i \in I) = E$.

A straightforward alternative approach can be based on regression. In this way, we can try to induce a function that maps object descriptions into ratings. However, this is not a faithful capturing of people's preferences. In fact, frequently, regression algorithms obtain errors near those achieved by the trivial *mean predictor*; that is, the predictor that suggests the mean of $\cup((r(x) : x \in E_i) : i \in I)$ as the rating for all possible objects, what is clearly unrealistic when we are trying to discover consumer preferences.

[1] Artificial Intelligence Center, Gijón (Asturias), Spain. E-mail: {jdiez,gbayon,quevedo,juanjo,oluaces,jalonso,antonio}@aic.uniovi.es

We will discuss the reasons for this behavior of regressors. For this purpose, the next section is devoted to present with some detail the peculiarities involved in sensory data. Thus, we find that sensory data expressed as a regression problem do not represent all available knowledge. In special, we would like to emphasize that consumers' rating are just a way for expressing a relative ordering. There is a kind of *batch effect* that often biases the ratings. Thus an object presented in a batch with clearly worse objects will probably obtain a higher rating than if it were surrounded by evidently preferable objects. Therefore, we must consider as a very important issue the information about the batches tested by consumers in a rating session.

In this paper we discuss how to tackle sensory data analysis in Machine Learning with a new point of view. Our approach postulates to learn *consumer preferences*, see [8, 11, 4]. In this way, training examples can be represented by preference judgments: pairs of vectors (v, u) where someone expresses the fact that he or she prefers the object represented by v to the object represented by u. In other words, training sets are samples of binary relations between objects described by the components of vectors of real numbers. As pointed out in [2, 5], obtaining preference information may be easier and more natural than obtaining the ratings needed for a regression approach. Moreover, we will show that the adequate processing of this type of information gives rise to explanations of consumer tastes reasonably accurate that could not be reached at all by other approaches based on regression.

We conclude the paper with a report of the experiments conducted to illustrate our approach with two families of data sets; they have arisen from the analysis of sensory data of beef meat and traditional Asturian cider.

2 ANALYSIS OF SENSORY DATA

An excellent survey of the use of sensory data in the food industry can be found in [12, 1]; for a Machine Learning perspective, see [3] and it is closely related [7].

From a conceptual point of view, what is relevant for a Machine Learning approach, sensory data include the assessment of food products provided by two different kinds of groups of people usually called *panels*. The first one is made up of a small group of expert, trained judges; these will describe each product by attribute-value pairs. Expert panelists are thus required to have enough sensory accuracy so as to discriminate between similar products; note that experts are not necessarily asked to rate the overall quality or acceptability of products; their opinions may be quite different from untrained consumers ideas. This panel will play the role of a bundle of sophisticated sensors, probably acting in addition to some chem-

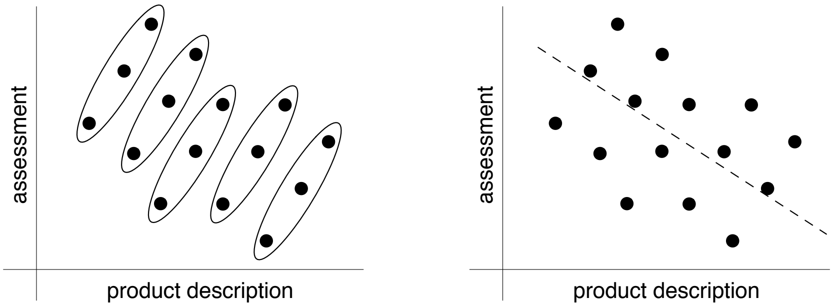

Figure 1. The left hand side picture represents, into each ellipse, the assessments of different sessions. If we do not take into account sessions, the cloud of points of the right side represents the whole data set to be handled by regression methods. These interpretations of data suggest quite different assessment policies

ical or physical devices. To achieve this performance, 2-3 times as many panelists as those required are screened through a selection or casting process that may take several months.

Notice that expert descriptions are ratings in a scale of different aspects of products related to their taste, odor, color, etc.. Here we must assume that a rating of "7" (in say, texture) means the same for a given expert in every product; though not necessarily for every expert. In other words, the most essential property of expert panelists, in addition to their discriminatory capacity, is their own coherence, not the uniformity of the group. Therefore, the selection of expert panelists must check this capacity of candidates throughout a number of experiments.

The second kind of panel is made up of untrained consumers; these are asked to rate their degree of acceptance or satisfaction about the tested products on a scale. The aim is to be able to relate sensory descriptions (human and mechanical) with consumer preferences in order to improve production decisions.

If we consider the whole data collected in a sensory study, we have to take into account that these data sets have some important properties that must be considered. First, we observe that the assessments come from a set of different consumers. This implies that we will have different scales in the available ratings. In other words, "7" does not mean the same for every body. Second, consumers use the assessments to express a relative ordering of the samples presented, during a testing session, in the same batch, but their ratings can not be considered as a general value. This is the phenomenon alluded to in Section 1 as the batch effect. Finally, there is an important peculiarity of food products to be contemplated here: consumers do not test all available samples; otherwise it would be physically impossible in some cases, or the number of tests performed would damage the sensory capacity of the consumer. Typically, each consumer only participates in one or a small number of testing sessions, usually in the same day.

Let us emphasize the importance of sessions with a graphical example depicted at Figure 1. Here there are a collection of consumers assessments (represented in the vertical axis) about some products whose descriptions are given by a single number x represented in the horizontal axis. If we observe the left hand side, where the assessments of the same session are drawn inside ellipses, we can say that in each session the message of the consumers is the same: the more x the better. However, there are discrepancies about how this knowledge is expressed in different sessions. Probably because there are different consumers in each session; or perhaps because the same consumer forgets the exact number used to assess a given degree of satisfaction; or the sensory reactions were forgotten from one session to another.

If we do not consider sessions, the data collected become the cloud of points represented in the right hand side of Figure 1. Then, it will be difficult for a regression method to discover the unanimous opinion of consumers. In fact, in this case, regression methods will conclude that the more x the worse, since that seems to be the general orientation of those points in the space. Nevertheless, notice that any other conclusion could have been drawn just if the relative positions of the sessions would change.

Therefore, the information about the sessions must be integrated in the data to be processed with the rest of sensory opinions and descriptions of the products tested by consumers. Thus, market sensory studies should arrange data in tables such as Table 1. Each row represents a product rated by a consumer in a given session.

Table 1. Sensory data collected from panels of experts and consumers. Each product is described by expert assessments (Att_j) in addition to other ($\mathrm{O\text{-}Att}_i$) chemical or physical analysis outputs

Expert sensory appreciations: Expert-1				Expert-k			Other descriptive attributes			Consumer preferences		
Att_1	…	Att_m	…	Att_1	…	Att_m	$\mathrm{O\text{-}Att}_1$	…	$\mathrm{O\text{-}Att}_n$	Ses.	Con.	Rating
<num>	…	<num>	…	<num>	…	<num>	<num>	…	<num>	<i>	<Id>	<num>
⋮	…	⋮	…	⋮	…	⋮	⋮	…	⋮	⋮	⋮	⋮
<num>	…	<num>	…	<num>	…	<num>	<num>	…	<num>	<i>	<Id>	<num>

In the next section we will present our approach to deal with sessions explicitly. The overall idea is avoid trying to predict the exact value of consumer ratings; instead we will look for a function that returns higher values to those products with higher ratings. Such functions are called preference or ranking functions.

3 BINARY SEPARATION AND PREFERENCES

Although there are other approaches to learn preferences, following [9, 8, 11, 4] we will try to induce a real preference or ranking function f from the space of objects considered, say $\mathfrak{R}^m$, in such a way that it maximizes the probability of having $f(v) > f(u)$ whenever v is preferable to u. This functional approach can start from a set of

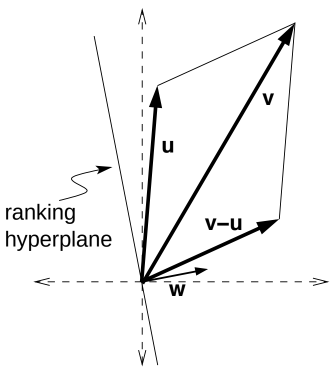

Figure 2. The difference vector $v - u$ is on the positive side of the hyperplane with normal vector w. Therefore, $f_w(x) = w \cdot x$ coherently ranks the preference judgment $v > u$

objects endowed with a (usually ordinal) rating, as in regression, but essentially, we need a collection of preference judgments

$$PJ = \{v_j > u_j : j = 1, \ldots, m\} \tag{1}$$

When we have a family of ratings $(r(x) : x \in E_i)$ for a set $i \in I$, we transform them into a preference judgments PJ set considering all pairs (v, u) such that objects v and u were presented in the same session to a given consumer i, and $r_i(v) > r_i(u)$. Hence, without any lost of generality, we can assume a set PJ as in formula (1). Then, it is possible to reduce the induction of a ranking function to a problem of binary classification.

If we assume that f must be a linear function; then it will have the form $f_w(x) = w \cdot x$ for a given w. From a geometrical point of view, the output of this map is proportional to the distance to the hyperplane of vectors perpendicular to w; see Figure 2. Thus, given $v > u$, we need w and $(v - u)$ to be vectors with a positive cosine, i.e. with a positive scalar product; equivalently $w \cdot (v - u) > 0$. In symbols,

$$\begin{aligned} f_w(v) > f_w(u) \Leftrightarrow 0 < f_w(v) - f_w(u) = \\ = f_w(v - u) = w \cdot (v - u) \end{aligned} \tag{2}$$

Thus we are searching for a hyperplane able to separate increasing or positive differences (like $v_j - u_j$ when $v_j > u_j \in PJ$) from decreasing or negative differences (like $u_j - v_j$). We will employ an SVM classifier [16] to find this w; the implementation used is Joachims' SVM$^{\text{light}}$ [10].

If we want (or need) nonlinear ranking functions, we can use the approach of [9, 8]. Thus we can look for a function $F : \mathfrak{R}^m \times \mathfrak{R}^m \to \mathfrak{R}$ such that

$$\forall x \in \mathfrak{R}^m, F(x, y) > 0 \Leftrightarrow F(x, 0) > F(y, 0). \tag{3}$$

Then, the ranking function $f : \mathfrak{R}^m \to \mathfrak{R}$ can be simply defined by

$$\forall x \in \mathfrak{R}^m, f(x) = F(x, 0). \tag{4}$$

Given the set of preference judgments PJ (equation (1)), we can specify F by means of the restrictions

$$\forall j = 1, \ldots, m, F(v_j, u_j) > 0 \text{ and } F(u_j, v_j) < 0 \tag{5}$$

Therefore, we have another binary classification problem that can be solved by a SVM obtaining a function of the form:

$$F(x, y) = \sum_{i=1}^{n} \alpha_i z_i \mathbb{K}(x_i^{(1)}, x_i^{(2)}, x, y) \tag{6}$$

where the pairs $(x_i^{(1)}, x_i^{(2)})$ are the support vectors, and $\mathbb{K}$ is the kernel used by SVM. The key idea is the definition of the kernel $\mathbb{K}$ as follows

$$\begin{aligned} \mathbb{K}(x_1, x_2, x_3, x_4) &= k_p(x_1, x_3) - k_p(x_1, x_4) - \\ &- k_p(x_2, x_3) + k_p(x_2, x_4) \end{aligned} \tag{7}$$

where now k_p is a kernel function defined as the inner product of the representation of two objects. In this case, it is easy to proof that F fulfils conditions (3) as we needed to learn a ranking function.

In the experiments reported in the next section, we will employ a polynomial kernel, defining

$$k_p(x, y) = (x \cdot y + c)^d, \tag{8}$$

with $c = 1$ and $d = 2$. Notice that, in general, according to the previous definitions,

$$f(x) = \sum_{i=1}^{n} \alpha_i z_i (k_p(x_i^{(1)}, x) - k_p(x_i^{(2)}, x)) \tag{9}$$

Hence, for the polynomial kernel we will obtain a nonlinear ranking function that assesses the ranking for each object x.

4 EXPERIMENTAL RESULTS

To illustrate the benefits of our approach, we have conducted some experiments with a couple of sensory data bases. In both cases we performed a comparison between the scores achieved by preference approaches and those obtained by regression methods. In all cases, to estimate the errors, we used 10-folds cross validation repeated 5 times.

As was explained above, the core point is the concept of testing session. Thus, for each session, to summarize the opinions of consumers, we computed the mean of the ratings obtained by each food product; notice that in this context all consumers have tested all products at the same time. This gives rise to some entries in a regression training set; that is, object descriptions endowed with a continuous class. And additionally, we can obtain some preference judgments. The regression training set so formed can be used to induce a function that predicts the exact ratings of consumers. We made this experiment with a simple linear regression and with a well reputed regression algorithm: Cubist, a commercial product from RuleQuest Research [15].

To interpret the results we used the relative mean absolute deviation ($rmad$). This amount is computed from the mean absolute distance or deviation, mad of the function f learned by the regression method, that is

$$mad = \frac{1}{|E'|} \sum (|f(x) - x_{class}| : x \in E') \tag{10}$$

where E' is a test set. Then the $rmad$ is computed as 100 times the quotient of mad and the mad of the unconditional constant predictor that returns the mean value in all cases. In symbols,

$$rmad = 100 \cdot \frac{mad(f)}{mad(\text{mean})} \tag{11}$$

Notice that a $rmad$ of 100% means that the regression method has the same mean absolute deviation as the constant or mean predictor.

On the other hand, when we deal with preference judgments, the errors have a straightforward meaning as misclassifications. To handle those data sets, we will use linear and nonlinear SVM, in this case with a polynomial kernel of degree 2. But, additionally, we have given another opportunity to regression methods. Thus, we trained them with regression entries, and tested with a set of preference judgments formed with ordered pairs of test examples according to their class: for each example we randomly selected other 10 examples to form, in this way, 10 preference judgments pairs.

The first data base comes from a study carried out to determine the attributes that entail consumer acceptance of beef meat [6]; the aim of the study was to test the influence of the beef cattle breed and the time of maturing of meat pieces after slaughter. For this purpose, a set of animals from seven Spanish breeds were used. A set of consumers rated 4 or 5 pieces of meat at the same testing session; the pieces of meat were described by: 12 features rated by 11 different experts, the weight of the animal, the maturing time, the breed, and 6 physical attributes describing the texture. Given that the breed was represented by means of 7 Boolean attributes, the whole description of each piece of meat uses 147 attributes. In Table 4 we show the cross validation scores achieved with these data sets.

Table 2. For each real-world problem used, this table shows the number of attributes as well as the number of examples, which depends on the approach followed, regression or preference learning

		# Examples	
	# Atts.	Regression	Pref. judgements
tenderness	147	468	2443
flavor	147	468	2411
acceptance	147	468	2412
acidity	64	98	238
bitterness	64	98	231
flavor-1	64	98	239
flavor-2	64	98	225
flavor-3	64	98	239
bouquet	64	97	233
color	64	98	241
visual-1	64	98	226
visual-2	64	98	226
visual-3	64	98	224
visual-4	64	98	205

Table 3. Beef meat error results. In regression we report the relative mean absolute deviation; in preferences, the percentage of preference judgments pairs misclassified is shown. In all cases the errors have been estimated by a 10-folds cross validation repeated 5 times

	Regression		Preferences			
	Linear	Cubist	SVM linear	**SVM Poly**	Linear	Cubist
tenderness	96.3%	97.8%	29.6%	19.4%	41.5%	43.1%
flavor	99.3%	103.4%	32.7%	23.8%	43.8%	46.5%
acceptance	94.0%	97.2%	31.9%	22.1%	38.4%	40.2%
Average	96.51%	99.49%	31.39%	**21.79%**	41.24%	43.27%

The second data base deals with sensory data about traditional Asturian cider. In this case, the description of each cider was given by 64 chemical and physical features see ([14]. So there were no expert descriptions. In fact, the consumers here were a set of 14 candidates to become experts, and the rating sessions were taken during the training and selection stage of these future experts. The experiment took place during several days, and there were samples of 91 different ciders that were presented in testing sessions of 3, 4 or 5 ciders. The number of ciders rated per person varied from 8 to 78, with an average of around 40. Therefore, this group of 14 people has the typical properties of consumer panels, as they were explained in previous sections.

Table 4. Cider error results. See caption of Table 4 for details

	Regression		Preferences			
	Linear	Cubist	SVM linear	**SVM Poly**	Linear	Cubist
acidity	103.0%	109.4%	29.9%	18.0%	40.0%	42.4%
bitterness	105.8%	111.9%	30.5%	23.1%	56.0%	47.4%
flavor-1	105.3%	111.7%	27.2%	17.1%	42.4%	44.3%
flavor-2	107.2%	116.0%	28.6%	17.9%	45.6%	45.0%
flavor-3	110.3%	107.7%	33.6%	17.7%	43.8%	41.8%
bouquet	104.0%	110.2%	26.4%	21.0%	43.5%	42.7%
color	98.4%	109.9%	26.1%	17.8%	41.3%	43.4%
visual-1	103.2%	113.0%	25.9%	13.4%	41.7%	43.1%
visual-2	102.3%	112.0%	34.0%	20.0%	43.8%	45.7%
visual-3	107.2%	120.5%	25.3%	20.6%	45.6%	49.3%
visual-4	98.7%	97.2%	23.0%	14.0%	36.5%	38.2%
Average	104.12%	110.87%	28.24%	**18.23%**	43.65%	43.92%

Additionally, given that this group was trained to become experts, they were asked to rate a high number of qualities of ciders: color and a group of 4 additional visual aspects; acidity, bitterness, and 3 more flavor related aspects; and the bouquet. Thus, we have 12 qualities of cider considered. Table 4 reports the scores achieved, both with regression and preference methods.

The results showed in Tables 4 and 4 exhibit quite similar behaviour of the computational tools used to process each data set. First, let us observe that regression methods are unable to learn any useful knowledge: their relative mean absolute deviation ($rmad$) is near 100% in all cases, what means that the mean absolute deviation is more or less the same as that of the mean predictor.

On the other hand, when we use the point of view of preferences, the usefulness of the models so obtained can be considerable increased. However, we appreciate important differences between regression based methods and those based on finding a separating function in a binary classification. So, if we try to use what was learned with Cubist or a simple linear regression in order to discriminate what was preferred, then the scores are very poor; a cross validation shows that, in average, in this way the errors are over 40%. Nevertheless, separating methods based on SVM as described in Section 3 can reduce these errors to reach around 30% when we use a linear kernel, but we obtain errors near 20% if the kernel is a polynomial of degree 2. The rationale behind this improvement of nonlinear kernels results can be explained taking into account that the positive appreciation of food products usually requires a precise equilibrium of its components, and the increase or decrease of any value from that point is frequently rejected.

5 CONCLUSIONS

We have presented a new approach to the analysis of sensory data supplied by consumer panels. This is a very interesting issue for food industries, since it provides the knowledge that allows leading production systems in order to satisfy the consumers' sensory requirements. Previous methods are frequently difficult or impossible to use in practice. In this paper we have discussed why regression algorithms can not be successfully applied. The main reason is that these methods do not take into account that consumers do not rate all available products; they only assess groups or batches of products presented in a small number of sessions; and consumers give numerical assessments only as a way to express a relative preference, not to be considered as a general category.

Our proposal is to learn functional models able to explain consumer preferences, instead of the exact ratings. In a very practical sense, we can conclude that consumer panels should be asked to concentrate in providing preference judgments pairs instead of lists of ratings. As we have shown, from these data sets it is possible to induce useful models. In other words, it is possible to summarize the opinion of consumers about a kind of food products in such a way that the appreciation of each sample is functionally related with its objective description. To illustrate our approach, we have presented a set of experimental results obtained from real world data sets that collects sensory data about beef meat and traditional Asturian cider.

ACKNOWLEDGEMENTS

The research reported in this paper has been supported in part under Spanish Ministerio de Ciencia y Tecnología (MCyT) and Feder grant TIC2001-3579. The authors would additionally like to thank: Carlos Sañudo and his team of Animal Production in the Universidad de Zaragoza, for providing us with the data set of sensory quality of beef meat [6]; Anna Picinelli and her colleagues at Cider group on SERIDA, Asturias, for allowing us to use the cider data sets [14]; and Thorsten Joachims for his SVM$^{\text{light}}$ [10] used in the experiments.

References

[1] D. Buck, I. Wakeling, K. Greenhoff, and A. Hasted, 'Predicting paired preferences from sensory data', *Food quality and preference*, **12**, 481–487, (2001).

[2] W.W. Cohen, R.E. Shapire, and Y. Singer, 'Learning to order things', *Journal of Artificial Intelligence Research*, **10**, 243–270, (1999).

[3] D Corney, 'Designing food with bayesian belief networks', in *Proceedings of the International Conference on Adaptive Computing in engineering Design and Manufacture*, pp. 83–94, (2002).

[4] J. Díez, J.J. del Coz, O. Luaces, F. Goyache, J. Alonso, A.M. Peña, and A. Bahamonde, 'Learning to assess from pair-wise comparisons', in *Proceedings of the 8$^{\text{th}}$ Iberoamerican Conference on Artificial Intelligence (IBERAMIA)*, volume 2527 of *Lecture Notes in Artificial Intelligence*, pp. 481–490, Sevilla, Spain, (2002). Springer-Verlag.

[5] S. Dumais, K. Bharat, T. Joachims, and A. Weigend, eds. *Workshop on implicit measures of user interests and preferences. In ACM SIGIR Conference*, Toronto, Canada, 2003.

[6] M. Gil, X. Serra, M. Gispert, M.A. Oliver, C. Sañudo, B. Panea, J.L. Olleta, M.M. Campo, M. Oliván, K. Osoro, M.D. Garcia-Cachan, R. Cruz-Sagredo, M. Izquierdo, M. Espejo, M. Martín, and J. Piedrafita, 'The effect of breed-production systems on the myosin heavy chain 1, the biochemical characteristics and the colour variables of longissimus thoracis from seven spanish beef cattle breeds', *Meat Science*, **58**, 181–188, (2001).

[7] F. Goyache, A. Bahamonde, J. Alonso, S. López, del Coz J.J., J.R. Quevedo, J. Ranilla, O. Luaces, I. Alvarez, L.J. Royo, and J. Díez, 'The usefulness of artificial intelligence techniques to assess subjective quality of products in the food industry', *Trends in Food Science and Technology*, **12**(10), 370–381, (2001).

[8] R. Herbrich, T. Graepel, and K. Obermayer, 'Support vector learning for ordinal regression', in *Proceedings of the Ninth International Conference on Artificial Neural Networks*, pp. 97–102, Edinburgh, UK, (1999).

[9] R. Herbrich, T. Graepel, and K. Obermayer, *Large Margin Rank Boundaries for Ordinal Regression.*, chapter 7, 97–102, MIT Press, Cambridge, MA, 2000.

[10] T. Joachims, 'Making large-scale support vector machine learning practical', in *Advances in Kernel Methods: Support Vector Machines*, ed., A. Smola B. Schölkopf, C. Burges, MIT Press, Cambridge, MA, (1998).

[11] T. Joachims, 'Optimizing search engines using clickthrough data', in *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, (2002).

[12] J.M. Murray, C.M. Delahunty, and I.A. Baxter, 'Descriptive sensory analysis: past, present and future', *Food Research International*, **36**, 461–471, (2001).

[13] T. Næs and E. Risvik, *Multivariate analysis of data in sensory science*, Elsevier, 1996.

[14] A. Picinelli, B. Suárez, J. Moreno, R. Rodríguez, L. Caso-García, and J.J. Mangas, 'Chemical characterization of Asturian cider', *Journal of Agricultural and Food Chemistry*, **48**, 3997–4002, (2000).

[15] RuleQuest. Computer software for regression, distributed by RuleQuest Research, 2001.

[16] V. Vapnik, *Statistical Learning Theory*, John Wiley, New York, NY, 1998.