Taking advantage of images and texts in recommender systems: semantics and explainability

PABLO PÉREZ-NÚÑEZ, Artificial Intelligence Center, Universidad de Oviedo en Gijón

Traditionally, recommender systems are built on the users' item consumption history. Many times, the users also make items reviews, giving us additional information in the form of text and images that are generally, not fully exploited. In this research we propose different approaches in the context of restaurants recommendation, where we take advantage of this information, by extracting a semantic meaning, in order to improve the traditional RS in terms of personalization and explainability.

CCS Concepts: • **Information systems** → **Recommender systems**.

Additional Key Words and Phrases: semantics, photos, collaborative platform, personalize, restaurants

ACM Reference Format:
Pablo Pérez-Núñez. 2020. Taking advantage of images and texts in recommender systems: semantics and explainability. 1, 1 (June 2020), 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Recommender systems (RS) are present nowadays in the vast majority of websites dedicated to offering products or content to users (selling goods, services, movie streaming, etc.). The use of RS is increasingly popular because of the benefits they provide, both for customers and suppliers. The former benefit from an improved experience by finding out new products or services they like, while the latter benefit from having their customers satisfied and, thus, increasing their sales volume. The recommendations given by an RS can be generic (the same for every user) or personalized, but traditionally they are made using nothing but the consumption history of the users. In some cases this is the only available information. However, it is increasingly common to find websites (like Amazon or TripAdvisor) where the users can not only assess the items/services, but also upload reviews including texts and images. These unstructured information can be crucial to achieve better recommendation results or solve complex recommendation tasks. Our goal is to grasp the users' tastes implicit in those pieces of information. For this purpose we have to take into account the *semantics* of the images and texts, not in a pictorial or linguistic sense, but in the sense of what they mean for the user's opinion about the item being assessed. We have to properly encode the semantics of this unstructured information to achieve a better personalization of the recommendations.

In order to accomplish our research, we have focused in restaurant recommendation. For this purpose we created a dataset from restaurant reviews including both structured and unstructured (text and images) information. The findings in this particular recommendation task are directly generalizable to any kind of product or service recommendation, such as consumer goods or touristic points of interest.

Author's address: Pablo Pérez-Núñez, pabloperez@uniovi.es, Artificial Intelligence Center, Universidad de Oviedo en Gijón, Campus de Viesques, 33204, Gijón, Spain.

Manuscript submitted to ACM

A first use of images in the context of RS can be to provide a visual explanation of the recommendation. For the case of restaurant recommendation, providing an image together with the recommendation can be understood as "we recommend you this restaurant because this (the image) is what you can find there". Obviously, not every image of the restaurant is suitable for every user, we have to select the image the user would most likely take. This image would be an explanation of the recommendation and would increase the user's confidence towards the RS.

We can also use the images as a way to represent the user's profile, allowing us to recommend her a personalized list of restaurants with nothing more than her taken pictures. In this case we have to take into account not only the image content but the semantics too, as can be seen on Figure 1.

Fig. 1. Two images with pizza as the main content, but causing different reactions on users

A similar approach can be used to process text data, not only the meaning of the words is relevant but the context also matters. We can learn to encode information about how the user experience was in the restaurant, ranging from food taste to atmosphere or staff descriptions.

In this research work our goal is to study and propose different ways of utilizing unstructured information in the form of texts and images, either separate or in combination, with the aim of improving recommendation tasks by adding explainability or by grasping the semantics to increase performance.

The paper is organized as follows: In Section 2, we provide an overview of some related work. In Section 3 we introduce the source of data from which we built the datasets to be used in the experimental tests. Section 4 presents our proposed methods, and finally, in Section 5 we point out some conclusions and future work.

2 RELATED WORK

In this section we include a brief review of some works related to our research.

2.1 Recommender systems and images

Almost every big multimedia company have their custom built RS nowadays. YouTube, the largest video platform, describes its RS in [5]; in here two different networks are capable of doing personalized video recommendations in a relatively easy way. One of the nets generate a personalized video candidates list, whereas the second sorts that list in the most appealing way for the user.

Another very relevant company in the sector, Netflix, have several recommendation tools that are deeply explained in [11]. One remarkable technique that they use, is the one in charge of selecting movie covers [24]. The idea lies in exploiting movies' and users' information in order to select the picture that best fits a movie across all users. Amat et al. [1] took this concept further, selecting personalized covers for each user with the objective of convincing users to watch a movie by showing them some visual evidence that supports the recommendation.

As far as we know, the fist try to use images as a way to provide recommendations is defined in [15]. In this work, the authors presented a matrix factorization system called VBPR (Visual Bayesian Personalized Ranking) where the items' (restaurants) description are the features learned by a convolutional neural network (CNN) using one single image from the item. This work was later extended by Kang et al. [19], who improved the performance of the system by training the image representation and the RS jointly.

2.2 Semantics extraction

Semantics has been widely studied, mainly with the aim of capturing meanings in texts. The way of encoding texts has evolved greatly towards learning a semantic meaning. Initially, more traditional methods such as bag of words [12] did not take semantics into account, but today there are numerous methods, such as Word2Vec [21], Doc2Vec [20], LSTM [17] or BERT [8] where semantics is the main focus of the study. Many research works use these encoding methods, for example the one described in [10], where semantic methods are used to recommend adverts in social networks or the work of Hassan et al. [13], who present an RS to recommend scientific articles based on the codification of titles and abstracts. Other examples like the works that proposed the use of semantically annotated query logs to recommend queries in search engines [3] or the product size RS based on customers' feedback [22] can be used as an example.

Contrary to what happens with texts, the extraction of a semantic meaning from images has not been explored so much in the field of RS. Commonly, pre-trained CNNs are used as generic feature extractors [23], obtaining the so-called deep features which are then used to encode images. These features do not represent basic image properties, like the traditional handcrafted ones used in the early days of research in computer vision. Instead, they represent semantic information for a given task (usually an object classification task). However, the deep features yielded by these networks are not devised to be used in the context of RS, so they have to be fine-tuned for that purpose.

We can also find multiple research works where images are included in RS. In addition to those cited in Section 2.1, stands out the approach used on TripAdvisor, one of the most popular platforms in the hospitality sector, where the most appealing pictures for each restaurant or hotel are selected [2]. In order to achieve that, they encoded the images by using the convolutional base of a ResNet50 [14]. This codification method based on pre-trained CNNs is also used in [4] with the aim of providing restaurant recommendations based on visual features.

3 DATASETS

In order to perform some empirical tests, we have created a collection of six datasets from users's reviews of restaurants uploaded to the TripAdvisor[1] platform throughout the years 2018 and 2019. Each review is made up by the user's textual opinion together with a rating, which ranges from one to five stars. Optionally, the user can upload images in order to provide some visual information of the reviewed restaurant.

Each dataset corresponds to a different city of the world. We have selected them trying to achieve some diversity regarding population, culture, touristic influence and countries of origin.

These differences are reflected in the figures of Table 1 , which depicts the basic statistics of each dataset. Thus, the number of users in these dataset ranges from a value slightly above 26000 (Gijón) up to more than 1 million (London). This variation is also reflected in the rest of elements of the datasets, i.e., reviews, images and restaurants. Worth of mention is the ratio of images per review, which is rather similar among all the cities (it ranges between 0.22 in London up to 0.35 in Gijón) and very low, indicating that most of the reviews have no visual information attached. It is also

[1]https://www.tripadvisor.com

City	Reviews	Images	Users	Rest.	Rev./Rest.	Img./Rest.	Us./Rest.	Rev./Us.	Img./Us.	Rest./Us.
Gijón	54787	19362	26450	716	76.518	27.042	36.941	2.071	0.732	0.027
Barcelona	466964	153707	184307	7602	61.426	20.219	24.245	2.534	0.834	0.041
Madrid	641561	208430	246618	8706	73.692	23.941	28.327	2.601	0.845	0.035
New York	1008761	234892	430082	10271	98.214	22.869	41.873	2.346	0.546	0.024
Paris	1135192	257447	463098	15410	73.666	16.706	30.052	2.451	0.556	0.033
London	2271164	489064	1037845	17976	126.344	27.206	57.735	2.188	0.471	0.017

Table 1. Some basic statistics of the data gathered from TripAdvisor.

noticeable the increase in the amount of reviews, images and users in London with respect to Paris, especially when compared to the increase in the number of restaurants. Thus, the London dataset has 16% more restaurants than Paris but 100% more reviews and 124% more users.

4 PROPOSED METHODS

4.1 Explaining with images

Explainable artificial intelligence (XAI) is becoming an important area of interest since explanations may help increase the trust of users in AI algorithms. In particular, an explained recommendation may result more convincing that a

Taking the above into account, and following the popular saying that a picture is worth a thousand words, we propose a method able to find the most appealing image of a restaurant for a particular user. Then, this image can be used to explain such restaurant recommendation. Notice that our method is *not* a RS, but a complement that enriches the recommendation output (obtained from any given method) by adding a visual explanation.

In order to do that, we need to learn a model capable of estimating the probability of an image $\boldsymbol{i}$, taken in a restaurant $\boldsymbol{r}$, being representative for the tastes of user $\boldsymbol{u}$. We assume that a user takes photos of an item trying to reflect the most important characteristics that make her like/dislike the item. Therefore, we want to learn how to select the photo that a user would eventually take.

For this purpose we propose to train a model for detecting the authorship of images. The rationale is that, if the model is able to gather the common essence from the images taken by a user, then we can use it to select the most engaging image from a bunch of images of the recommended restaurant. In symbols,

$$\boldsymbol{i}^* = \underset{\boldsymbol{i} \in \text{photos}(\boldsymbol{r})}{\text{argmax}} \ \Pr(\boldsymbol{u}, \boldsymbol{i}). \tag{1}$$

Thus, we aim at solving a binary classification task to estimate $Pr(\boldsymbol{u}, \boldsymbol{i})$. The reason is that once we learn to predict the probability of a photo to be taken by a user $\boldsymbol{u}$, we can apply that model to select photos taken by other users, but with a high probability of being also taken by $\boldsymbol{u}$.

We presented our findings regarding the photo authorship in a recent publication [9], where we detailed and evaluated the performance of our approach, ELVis, which stands for Explaining Likings Visually. The evaluation consisted of measuring the quality of a ranking, given that our model sorts the available photos of a restaurant according to the authorship probability (for any given user). Therefore, traditional measures such as top-n [6] can be applied.

We built test sets (one for each city) from the TripAdvisor collection such that, for each restaurant, there was a pair $(\boldsymbol{u}, \boldsymbol{f})$ with a positive label, indicating that $\boldsymbol{u}$ took the photo $\boldsymbol{f}$ in that restaurant. We included as negative test examples pairs $(\boldsymbol{u}, \boldsymbol{f}')$ for all the photos taken by users other than $\boldsymbol{u}$ in the same restaurant of $\boldsymbol{f}$. The optimal result happens

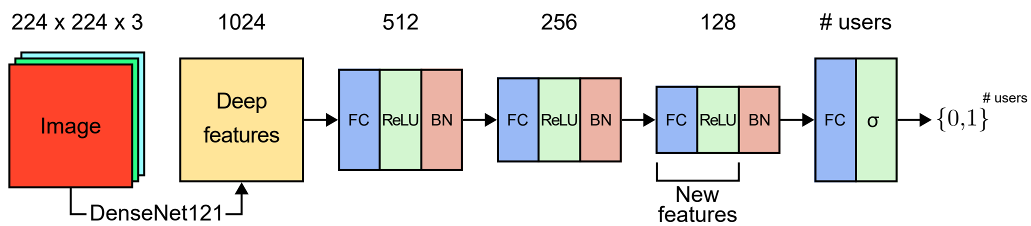

Fig. 2. Proposed architecture. FC stands for fully connected layer, BN for bacth normalization and σ represents the sigmoid activation function.

whenever the method places the user's picture, $\boldsymbol{f}$, in the first position of the ranking of all these photos when ordered by the joint (authorship) probability, $\Pr(\boldsymbol{u}, \boldsymbol{f})$.

Provided there is only one correct photo in our tests to be ranked among others in the highest possible position, this top-n measure coincides with *Recall@n*, and it is equivalent to $n \times$ *Precision@n*, expressed as a percentage. Herlocker et al. [16] claim that these measures are adequate for tasks of the type *Find Good Items*, as is our case.

Obviously, any ranking with less than n photos will be considered as a successful top-n prediction, provided it will always contain the correct photo. Thus, we will also analyzed the quality of the rankings with another measure, the percentile position of the correct photo in the ranking, computed as

$$\text{percentile}(\boldsymbol{f}, \boldsymbol{R}) = 100 \cdot \frac{\text{index}(\boldsymbol{f}, \boldsymbol{R}) - 1}{|\boldsymbol{R}|}, \tag{2}$$

where $\boldsymbol{R} = \{\boldsymbol{f}\} \cup (\text{photos}(\boldsymbol{r}) \setminus \text{photos}(\boldsymbol{u}))$ is the ranking containing the photo $\boldsymbol{f}$ taken by user $\boldsymbol{u}$ together with all the photos of the same restaurant but taken by other users. Here, $\text{index}(\boldsymbol{f}, \boldsymbol{r})$ is the position of $\boldsymbol{f}$ in the ranking.

The ranking is in descendant order of $\Pr(\boldsymbol{u}, \boldsymbol{f})$; therefore, the lower the percentile, the better the ranking.

4.2 Image semantics

Semantics studies the meaning of the expressions that we use in our language. We can extrapolate this concept to the computer vision field, assuming that the images have also a meaning and that it can be learnt in some way. Considering that, we propose a method where images can be given meaning in the context of a RS dealing with restaurants. Extracting the semantics of a group of images will help us to make recommendations starting just from a bunch of images previously taken by a user.

Images are typically encoded by the convolutional base of a general-purpose CNN, for example DenseNet [18] pre-trained on ImageNet [7]. In this case, the meaning of an image is linked to the visual features that allow to detect its contents. Unfortunately, this is not very useful for our purpose because it does not take into account the users' tastes, so we have to build a custom codification.

This codification must ensure a similar representation for the images of two strongly related restaurants. We will define the strength of this relation in function of the amount of users who visited both restaurants (i.e., the intersection of visitors). We propose to gather the semantics of an image by means of a multi-label classifier trained to label each image with a group of users, those who visited the restaurant where the image was taken.

This classifier can be obtained using a network architecture like the one depicted in Figure 2. The input of this network will be fed with images encoded using the convolutional base of a DenseNet121. For each input image, the output will be a vector of probabilities, where the i-th component will be the probability that user i has visited the restaurant where that image was taken. By removing the last layers of the trained model, we can take the 128-feature vector to be used as a new representation for each image in our new semantic space.

In order to evaluate the performance of this approach we propose to use a test set with users not included in the training set. Our trained model will encode each of their images into the 128-dimensional space mentioned above. Then, we will compute the centroid of each user's images, which will represent the user's profile in the semantic space.

Once we have the user profile, we can obtain the similarity between this 128 dimensional vector and all the train images, to, finally recommend the restaurant where the most similar image was taken. If the user has visited the restaurant we account it as a hit, otherwise, as a failure. We can compute the average hit percentage or the average median position for a whole test set.

The described approach has been successfully compared with a baseline representation. The results obtained on the datasets described in Section 3 are discussed in a recently submitted paper, which is still under review.

4.3 Text semantics

Talking about semantics sounds more natural when dealing with text than with images. In fact, there is much more research in the former case, which has been conducted since the early days of natural language processing (NLP). But we want to link the texts written by users in their reviews with their tastes. Thus, the encoding of texts should be tuned in the same sense of the tuning proposed for the generic representation of images, by means of convolutional base features. In other words, we need to go beyond the meaning of the texts in a linguistic sense, and try to capture some latent semantic features related to the users' tastes.

We expect to successfully develop an approach similar to the one described and already tested with images. Thus, our intention is to build a model able to re-encode the texts, originally represented by feature vectors obtained with generic procedures, such as Doc2vec, LSTM or BERT networks. These well-known methods will play the role of the DenseNet in the image-based approach.

We are currently in an early stage of the development of this approach, so we do not have any results yet.

4.4 Combination of semantic information

Once we successfully find a way to recover semantic information from images (some advances achieved) and from text (work in progress), the natural step will be to devise a combination method which hopefully will achieve the best performance in recommendation-related tasks.

Therefore, we expect to be able to obtain user profiles using both images and texts which, to some extent, define the tastes of the user. Moreover, following the same principle, we also plan to obtain restaurant profiles. A restaurant profile will be obtained from the combination of the semantic information provided by the text reviews and by the images taken in it.

Our idea is to encode the user (respectively, restaurant) images using the approach presented in Section 4.2 and the user (respectively, restaurant) text reviews with the approach mentioned in Section 4.3 (still in development). As a result, we will end up with two 128-dimensional vectors for each user (respectively, restaurant) that can be used as a unique profiles. All this proposed process is represented in Figure 3, where the block named Recommender System has to be defined. For instance, we can create an RS that use these profiles as an input of a binary classification model, where the objective is to learn the compatibility of both inputs, i.e., if the user likes the restaurant. However, other RS can also be explored.

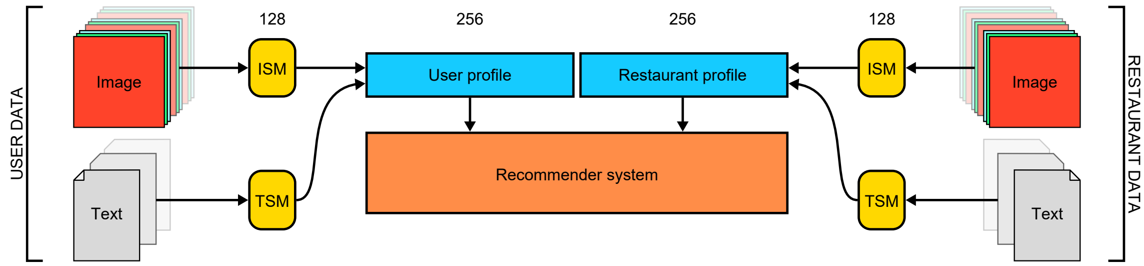

Fig. 3. Proposed combination architecture. ISM stands for image semantic model and TSM represents the text semantic model, both learned in previous tasks.

5 CONCLUSIONS AND FUTURE WORK

In this research we focus on the improvement of recommendations by means of incorporating semantics extracted from unstructured information like texts and images. We have already presented some advances when dealing with images. Our proposals have been published in a journal paper, and we also have a second paper under review. In these works we show that the images taken by users in restaurants have some sort of semantic meaning related to the users' tastes. It is possible to take advantage of that meaning to explain the recommendations to the users.

We are currently starting to implement the necessary modifications to also handle texts in a similar way. We plan to develop a method able to combine the semantic information from both images and texts, in order to improve recommendation tasks.

Additionally, we are also exploring the possibility of making a more realistic semantic evaluation, which will require to label datasets accordingly. For this purpose we need to use human intervention, so we plan to use services like Amazon Mechanical Turk[2].

ACKNOWLEDGMENTS

This doctoral project is supervised by Assoc. Prof. Oscar Luaces and Assoc. Prof. Jorge Díez. This work was funded under grants TIN2015-65069-C2-2-R and PID2019-109238GB-C21 from the Spanish Ministry of the Economy and Competitiveness, and IDI-2018-000176 from the Principado de Asturias Regional Government, partially supported with ERDF funds. Pablo Pérez-Núñez acknowledges the support of the Principado de Asturias Regional Government under *Severo Ochoa* Program (ref. BP19-012).

We are grateful to NVIDIA Corporation for the donation of the Titan Xp GPU used in this research.

REFERENCES

[1] Fernando Amat, Ashok Chandrashekar, Tony Jebara, and Justin Basilico. 2018. Artwork personalization at Netflix. In *Proceedings of the 12th ACM conference on recommender systems*. 487–488.

[2] Gregory Amis. 2017. Improving TripAdvisor Photo Selection With Deep Learning. https://www.tripadvisor.com/engineering/improving-tripadvisor-photo-selection-deep-learning/.

[3] Arthur Câmara and Rodrygo LT Santos. 2019. Traversing semantically annotated queries for task-oriented query recommendation. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 511–515.

[4] Wei-Ta Chu and Ya-Lun Tsai. 2017. A hybrid recommendation system considering visual information for predicting favorite restaurants. *World Wide Web* 20, 6 (2017), 1313–1331.

[2] https://www.mturk.com/

[5] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*. 191–198.

[6] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*. 39–46.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 248–255.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805* (2018).

[9] Jorge Díez, Pablo Pérez-Núnez, Oscar Luaces, Beatriz Remeseiro, and Antonio Bahamonde. 2020. Towards explainable personalized recommendations by learning from users' photos. *Information Sciences* 520 (2020), 416–430.

[10] Francisco García-Sánchez, Ricardo Colomo-Palacios, and Rafael Valencia-García. 2020. A social-semantic recommender system for advertisements. *Information Processing & Management* 57, 2 (2020), 102153.

[11] Carlos A Gomez-Uribe and Neil Hunt. 2015. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6, 4 (2015), 1–19.

[12] Zellig S Harris. 1954. Distributional structure. *Word* 10, 2-3 (1954), 146–162.

[13] Hebatallah A Mohamed Hassan, Giuseppe Sansonetti, Fabio Gasparetti, and Alessandro Micarelli. 2018. Semantic-based tag recommendation in scientific bookmarking systems. In *Proceedings of the 12th ACM Conference on Recommender Systems*. 465–469.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

[15] Ruining He and Julian McAuley. 2016. VBPR: visual bayesian personalized ranking from implicit feedback. In *30th AAAI Conference on Artificial Intelligence*. 144–150.

[16] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22, 1 (jan 2004), 5–53.

[17] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708.

[19] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. 2017. Visually-aware fashion recommendation and design with generative image models. In *IEEE International Conference on Data Mining*. 207–216.

[20] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*. 1188–1196.

[21] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[22] Rishabh Misra, Mengting Wan, and Julian McAuley. 2018. Decomposing fit semantics for product size recommendation in metric spaces. In *Proceedings of the 12th ACM Conference on Recommender Systems*. 422–426.

[23] Loris Nanni, Stefano Ghidoni, and Sheryl Brahnam. 2017. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition* 71 (2017), 158–172.

[24] Netflix. 2016. Selecting the best artwork for videos through A/B testing. Netflix Technology Blog: https://medium.com/netflix-techblog/selecting-the-best-artwork-for-videos-through-a-b-testing-f6155c4595f6.

Date of the CVA	16/06/2020

Section A. PERSONAL DATA

Name and Surname	Pablo Pérez Núñez		
DNI	15485080P	Age	25
Researcher's identification number	Researcher ID		
	Scopus Author ID		
	ORCID		

A.1. Current professional situation

Institution	University of Oviedo				
Dpt. / Centre	Artificial Intelligence Center / EPI Gijón				
Address					
Phone	(+34) 666614694	Email	pabloperez@uniovi.es		
Professional category	Pre-doctoral student (Severo Ochoa scholarship)			Start date	2020
UNESCO spec. code					
Keywords					

A.2. Academic education (Degrees, institutions, dates)

Bachelor/Master/PhD	University	Year
Master's degree in Computer Science	University of Oviedo	2018
Bachelor's degree in Computer Science	University of Oviedo	2016

A.3. General quality indicators of scientific production

Section B. SUMMARY OF THE CURRICULUM

Since I finished my bachelor studies in 2016, I started to work in projects related with the Machine Learning.

That same year, I joined TreeLogic (treelogic.com), where I participated in a European project as a data scientist collaborating alongside other enterprises. After spending one year in this company, the opportunity arose to join the Artificial Intelligence Center in the University of Oviedo with a research contract in collaboration with the electric company EdP (edp.com). The objective was to study to what extent the use of Smart-meters culd be useful to anticipating the appearance of faults. The research carried out in this project gave rise to a conference paper in which it was shown how to use preference learning to obtain a ranking with the power lines ordered according to their failure probability.

This contact with preference learning drove me to read articles on recommendation systems and there my interest in this topic began. This led me to propose as a final master project a study on how to generate user profiles from their lists of listened songs. For this, we obtained a projection of all the songs using the philosophy of word2vec and then, based on that representation, we obtained the user's profiles with techniques similar to doc2vec. An article for a conference and another for a journal resulted from this study.

Continuing in the line of work with user profiles, I started a job about clustering items and users based on user preferences. We obtained very good results and we have an article on this topic under review.

I am currently working on my Ph.D. thesis in the field of recommendation systems. The ultimate goal is to take advantage of the texts and images that users incorporate into their reviews. At the moment we have advanced in the use of images as an explanation of the recommendation (an image is worth a thousand words) and we have published a journal paper. We have also advanced in obtaining the semantics of images for use in recommendation systems,with a paper under review showing this approach.

At the end of 2019 I was awarded with one of the 65 "Severo Ochoa" competitive pre-doctoral scholarships given by the Principality of Asturias Regional Government which has a duration of four years.

Section C. MOST RELEVANT MERITS (ordered by typology)

C.1. Publications

1 **Scientific paper**. Jorge Díez; et al. 2020. Towards explainable personalized recommendations by learning from users' photos Information Sciences. Springer.
2 **Scientific paper**. Pablo Pérez Núñez; et al. 2019. Improving recommender systems by encoding items and user profiles considering the order in their consumption history Progress in Artificial Intelligence. Springer.
3 **Scientific paper**. Tania Maria Vázquez Sánchez; et al. 2019. Fault detection in low voltage networks with smart meters and machine learning techniques CIRED. AIM.
4 **Scientific paper**. Pablo Pérez Núñez; et al. 2018. Influencia de los perfiles de usuario recientes frente a los consolidados en tareas de recomendación CAEPIA. AEPIA.
5 **Scientific paper**. Pablo Pérez Núñez; et al. Preference-Based User Clustering in Very Sparse Recommender Systems Tasks Under review.
6 **Scientific paper**. Pablo Pérez Núñez; et al. SemPic: Towards restaurant recommendations by means of the semantics of users' pictures Under review.
7 **Scientific paper**. Carlos Alonso Huerta; Pablo Pérez Núñez. Visualizing Recommendations for Points of Interest Under review.

C.2. Participation in R&D and Innovation projects

1 PID2019-109238GB-C21. Explainable learning for Dyadic Data (xLearn) Ministerio de Ciencia e Innovación. Antonio Bahamonde Rionda. (University of Oviedo). 01/06/2020-31/05/2024. 82.764 €.
2 Algoritmos escalables de Aprendizaje Computacional: Más allá de la clasificación y la regresión (ALEC+, TIN2015-65069, proyecto coordinado) Ministerio de Economía y Hacienda. Amparo Alonso Betanzos. 01/01/2016-31/12/2019. 112.000 €.

C.3. Participation in R&D and Innovation contracts

Proyecto GAMMA para aprendizaje automático en eventos de sobre y sub-tension de contador FUNDACION UNIVERSIDAD DE OVIEDO; HIDROCANTABRICO DISTRIBUCION ELECTRICA SA. Jorge Díez Peláez. 01/05/2017-01/05/2018. 22.000 €.

University of Oviedo at Gijón
Principality of Asturias, Spain

June 15, 2020

To whom it may concern,

Since my PhD topic focuses on recommendation systems, I think there is no better place than the RecSys Doctoral Symposium to obtain feedback from my mentors. I would expect them to answer questions about my work, such as:

- Could you think in any other way of evaluate the semantics extracted in our proposals?
- We plan to use for texts the same semantic extraction process applied to images, does it make sense to use this method or is there some other?

Sincerely,

Pablo Pérez Núñez

Artificial Intelligence Center
University of Oviedo at Gijón
33204 - Spain

June 15, 2020

Doctoral Symposium Chairs
RecSys - 2020

To whom it may concern:

We are writing this letter to support the participation of our student Pablo Pérez-Núñez in the Doctoral Symposium of RecSys 2020.

Pablo started working with us on research projects as soon as he graduated and some time later, after completing a Master's degree in Computer Science, he began the development of what will be his doctoral thesis, which is totally linked to the field of recommendation systems.

At this moment, his thesis is at its midpoint and there are some branches than can be taken towards the end of the work. In this sense, Pablo has been thoroughly reviewing the related bibliography and proposes many ideas that could be applied. Some of them have been reflected in the abstract submitted to the Doctoral Symposium. In the last months we have seen how Pablo has matured scientifically and how he is able to bring creativity and, of course, ethics to his work.

Pablo's work is at that point where the advice from experts in the field can make a difference, and we can't think of a better place to receive such help other than this Doctoral Symposium.

Pablo's participation and collaboration in the meeting are guaranteed. He is very eager to receive such invaluable help, so we highly recommend Pablo to be accepted as a participant in the symposium.

Yours Faithfully,

Jorge Díez Oscar Luaces
(Ph.D Advisors)