The Winning Approach for Author Profiling of Mexican Users in Twitter at MEX.A3T@IBEREVAL-2018

Rosa María Ortega-Mendoza[1] and A. Pastor López-Monroy[2]

[1] Instituto Tecnológico Superior del Oriente del Estado de Hidalgo
Apan, Hidalgo, México
mortega@itesa.edu.mx
[2] Department of Computer Science, University of Houston, Texas USA
alopezmonroy@uh.edu

Abstract. This paper describes the best performing system for Author Profiling of Mexican Twitter-Users presented at MEX.A3T@IBEREVAL-2018. This competition addresses two tasks: author profiling and aggressiveness detection. The first one, is aimed to predict two traits from authors profile: occupation and place of residence. Our proposed approach exploits the relevance of personal information by means of feature selection and term weighting methods. The aforementioned approach was previously studied in English texts; in this paper we adapt and evaluate it for Spanish texts. The approach considers that sentences where users talk about themselves expose highly valuable information like interests, habits and fears. Our hypothesis is that such personal personal information can reveal characteristics of their profile or aggressive behavior. In the test stage our system obtained the first place in author profiling task and the sixth position in aggressiveness detection. The results showed evidence of the usefulness of the proposed approach in Spanish language, particularly to determine occupation and place of residence profiles. On the other hand, for aggressiveness detection we lay the ground for future research in emphasizing personal information.

Keywords: Aggressiveness detection· Author profiling· Personal Information.

1 Introduction

The Internet provides a number of services where users easily share information in social networks. The textual analysis of this information has been a popular research topic for the scientific community. This is especially true in applications that prevent risks or study the way the language is used by people. Recently, some academic competitions have emerged as forums where researches evaluate their approaches for a particular task. For example in the PAN forum [18–22] author profiling systems are submitted for predicting specific author traits such as personality, age and gender.

The MEX.A3T 2018 forum tackles two tracks focused on digital text forensics over tweets in Mexican Spanish [1]: a track on author profiling (AP) and a track on aggressiveness detection (AD). The first one is aimed to predict occupation and place of residence dimensions of the user profiles, which have been scarcely tackled by the

community and never approached for Spanish. The second one is focused in detecting aggressive comments in Twitter.

The AP is aimed to predict general or demographic attributes that integrate authors' profiles such as: personality [2, 23], native language [3], political orientation [16] and predominantly has been most focused on prediction of age [9, 13, 15] and gender [3, 8, 9, 15]. Traditionally, the AP task has been tackled from a text classification perspective [24] based on a Bag of Words (BoW) models. The research has been mainly focused on the selection of the best set of features for modeling the authors' writing profile. Recently, some works have used more sophisticated representations as second order attributes [10] or different deep learning models and strategies to learn representations for AP [7, 11, 12]. However, traditional approaches such as the simple BoW and word n-grams have outperformed those elaborated methods in many scenarios for AP [18].

The AD task consists in determining which comments attempt to insult, offend, attack, or hurt others putting the integrity of people in risk. It can be seen as the first step towards cyberbullying automatic identification [5]. Therefore, AD can prevent damages and harmful patterns or even suicide. The AD on texts is a task less tackled, it is commonly approached as a classification problem. For example, in [5] authors approached aggressive text detection as a regression problem that consists in mapping a document to an aggressiveness score. They used a dataset extracted from Twitter. Regarding to classification approaches, some works are aimed to discover a set of characteristics for modeling aggressive behavior as in [4] the authors explored text, user and network-based attributes. Finally, in [6] the authors successfully used profile-based representations for the early aggressive text identification where using a minimum amount of information is critical for prevention.

In both tasks, the language used in tweets of Mexican users has challenging cultural traits. For this reason, we proposed an approach based in physiological findings [17] that aims to expose personal information for discriminating user profiles. In a previous work [15] we analyzed English texts for detecting age and gender of users and demonstrated that terms inside personal phrases help to characterize and discriminate profiles. This is because they expose interests, preferences and habits among such personal information. In this work we adapt and evaluate the work in [15] for tweets written in Spanish by Mexican users, where the variation of language is a real challenge. For this purpose we have adapted the feature selection and term weighting stages in order to emphasize the value of these personal terms. Experimental evaluation shows that this is in fact useful for modeling the Mexican writing and discriminating among author classes.

The remainder of this document is organized as follows: Section 2 presents the proposed method. Section 3 describes experimental settings. The experiments and results are presented in Section 4. Finally, Section 5 outlines the final conclusions and future work.

2 Method proposed

Based on psychological findings, we studied the role of personal phrases (i.e., phrases containing singular first-person pronouns) for the AP problem [14]. In a previous

work we also proposed the approach DPP-EXPEI [15] for emphasizing the value of personal phrases in the AP task. In that work, we studied age and gender prediction on social media documents written in English. In this work, we bring the approach to work with tweets written in Spanish from Mexican users in order to predict two new dimensions: *location* and *occupation*. The underlaying hypothesis is that terms inside personal phrases help to characterize and discriminate these profiles in the Spanish language. For example, in the phrase: "*jajaja! pues **yo** soy estudiante, dejen dormir.*", which in English would mean something like *"Hahaha! I'm just a student, don't mess with me"* the author is declaring his occupation. Other example is the text "*fav por **mis** compañeros que **me** van a pasar la tarea*", which in English is close in meaning to *"fav for my classmates who will pass me the homework"* where there are keywords suggesting that the author is a student.

Specifically, the pronouns considered to define a sentence as personal were: *yo, me, mí, conmigo* (I, me, my, with me), as well the possessive case: *mío* (mine); also, their respective variants in number and gender[3] were considered as: *mío, mía, míos, mías* and the possessive adjectives: *mi* and *mis* (*my* in English language). In addition, the pronouns without accents were considered due to their popularity in social media.

Our method corresponds to a supervised classification approach. The aim is to use feature selection and term weighting strategies (DPP and EXPEI, respectively) that emphasize the value of personal information in the representation before feeding the classifier. The base of these strategies is a measure called *Personal Expression Intensity (PEI)* [15], which determines the amount of personal information revealed by each term.

PEI. This is a weight for boosting the relevance of terms that are more associated to the interests of the authors, therefore, it indicates that the more frequent is a term in the personal phrases of a document, and the less frequent it is in non-personal phrases, the more revealing is the term about the characteristics of the document's author. The PEI measure is defined for a term t_i occurring in a document d_j according to Formula 1. It is a combination of personal precision ρ and personal coverage τ. The former indicates the concentration of personal information revealed in the context of a term, it is estimated as the percentage of personal phrases in the subset of phrases containing the term. On the other hand, τ indicates the portion of the personal phrases from a document covered by the term.

$$PEI(t_i, d_j) = 2\frac{\rho(t_i, d_j) \cdot \tau(t_i, d_j)}{\rho(t_i, d_j) + \tau(t_i, d_j)} \quad (1)$$

Feature Selection. We consider that personal information is common among people sharing a trait of their profile (e.g. females), at the same time this can be discriminative for the others (e.g. men). Particularly, we used a novel technique called *discriminative personal purity* (DPP), which is an approach especially suited for AP in social media proposed in [15]. It not only considers the distribution of terms across the categories, as most traditional measures such as information gain do, but it also considers the kind

[3] In Spanish language, the number and gender of the words can modify the particles in the sentences

of phrases where they appear in according the Formula 2, which can be used to select a number of more relevant terms. DPP consists of two components: first, a descriptive factor, defined as the maximum value of the function *categorical personal purity*, PP_k (Eq. 3), that captures the capability of a term to describe personal information of authors belonging to the category (c_k); and second, a discriminative factor, based on the *gini* coefficient for scoring the ability of the term to discriminate among the different categories (profiles) of authors[4].

$$DPP(t_i) = \max_{k=1}^{|C|} \{PP_k(t_i)\} \cdot gini(t_i) \tag{2}$$

where,

$$PP_k(t_i) = \log_2 \left(2 + \frac{1}{2} \sum_{d_j \in c_k} \frac{PEI(t_i, d_j) + 1}{NEI(t_i, d_j) + 1} \right) \tag{3}$$

Term weighting. The terms existing in a text are used in different ways, most of them are informative and their value in the text representation should analyze their precedence context, for example personal or non-personal phrases. We considered that terms used in personal phrases are more descriptive for the profiles of author, therefore their value should be emphasized. We used the EXPEI scheme, an exponential rewarding to the weight of terms occurring in personal phrases. This scheme considers all the terms from the documents, from both personal and non-personal phrases, but it seeks to emphasize the personal information. It was also proposed in [15] and it is showed in the Formula 4.

$$w_{ij} = \left(\sqrt{TF(t_i, d_j)} \right)^{1 - PEI(t_i, d_j)} \tag{4}$$

where $TF(t_i, d_j)$ represents the normalized frequency of t_i in d_j, computed as $\frac{\#(t_i, d_j)}{\#(d_j)}$. The reward is based on the PEI measure causing the weight of a term to increase according to its personal intensity.

3 Experimental settings

For both tracks, data sets of the competition were divided into training and testing partitions whose characteristics are described in [1]. The dataset for AP is conformed by 3,500 and 1,500 tweets respectively. There are six *location* classes: *center*, *southeast*, *northwest*, *north*, *northeast*, and *west*. About the *occupation* problem: *arts*, *student*, *social*, *sciences*, *sports*, *administrative*, *health*, and *others*. On the other hand, the dataset for AD contains 7,700 tweets for training and 3,156 for test which were tagged by two classes: *aggressive* comment or *no aggressive*. The experimental framework is described below:

[4] The full description and details of this strategy can be found in [15]

- Features. In general, we used a combination of content and style attributes, which include unigrams of content words, punctuation marks, slang words and out-of-dictionary terms like emoticons. We also considered the occurrences of function words. By means of the n top terms according DPP we built a standard BoW representation where the weights of the terms are estimated with the EXPEI scheme.
- Classification. For all the experiments we considered a Support Vector Machine as learning algorithm with L2 normalization. To train our models we used 70% of the training data and the remaining 30% was used for validation.
- Evaluation. The performance of the author profiling systems is ranked by the macro average F1 measure for both dimensions: *location* and *occupation*. On the other hand, the performance for aggressive systems are ranked by the F1 measure on the positive class.
- Baseline. To compare our approach we built a very strong baseline based in word and char n-grams of different sizes. Then, for each space of word and char n-grams, we separately selected a subset of the best features[5]. For the profiling task, we used subsets of the best 1-3 word n-grams and 3-5 char n-grams. For the Agressiveness task, we also extracted subsets of the best word n-grams of sizes 1 to 5 and char n-grams from 2 to 6.

4 Experiments and results

4.1 Author profiling

In this section we evaluate two approaches: i) DPP-EXPEI and ii) the baseline. Regarding to DPP-EXPEI, for *location* we used 1000 top terms, function words (*fw*) and content attributes because we observed the use of *fw* is a discriminative factor of location. For *occupation* we used only the top 2000 content terms.

The general results in train and test phase are showed in Table 1. The approach based on DPP-EXPEI outperformed the baseline in both AP dimensions and also in the average value of F1 measure. This confirms the feasibility of the approach showing that the personal phrases have a special role for author profiling, even when texts are written in Spanish. Moreover, the approach obtained the first place in the competition in the rank of the six proposed methods from four teams.

Table 1: Results on AP at the MEX.A3T competition

		F-measure (macro)		Avg. F-measures
Dataset	Representation	Location	Occupation	
Train	DPP-EXPEI	**0.8275**	**0.4913**	**0.6594**
	baseline	0.8272	0.4367	0.6319
Test	DPP-EXPEI	0.5122	0.8301	0.67115

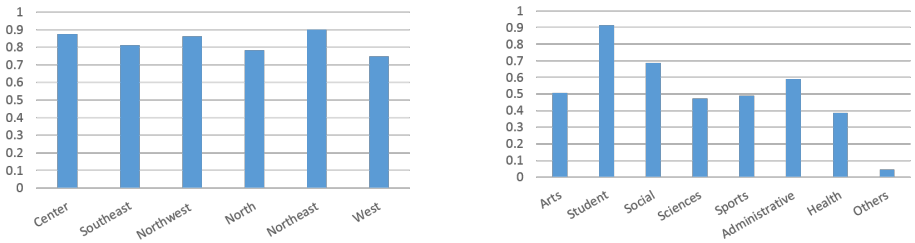
[5] We empirically chose the size and number of n-grams using a validation set of 30% out of the training.

In order to deepen the analysis of results of AP tasks, we showed the top words according to DPP in Table 2. In general, for *location* dimension the method is extracting terms frequently used, idioms, in the regions of México (e.g. *cuera* in the northeast region) and names of places in regions of the country. On the other hand, for *occupation* dimension it is associating words commonly used in the occupations.

Table 2: Examples of top words selected by DPP in the AP task of the competition. The words are shown by rank

ubication	occupation
gpi, cuera, kino, alchile, facu, poblanos, Cholula, Atlixco, Villahermosa, Tepeaca, Tlalpan, UDEM, UJAT, Nacajuca, achis, poblana, duranguenses, BUAP, curiouscat, Hermosillo, DGO, lit, poblano, Tenosique, Durango, calo, twittab, Polanco, Tehuacán, ansia, Lerdo, Tuxtla, Cuajimalpa, Mérida...	*girlposts, cawnas, nmms, lit, borrachosvip, aprobamos, yase, snaps, tlj, legislativa, alchile, cuera, srio, dirigente, sisi, legislativo, dlv, jurídica, diriegencia, bai, militancia, estatales, legislatura, okay, senadora, anatomy, beneficiarios, facu, apoyos, comparecencia, feelings...*

The Figure 1 shows the results of our system by class in the test phase. It is observed the approach is detecting with the similar performance the classes of the *location* dimension. Regard *occupation* dimension the approach is recognizing better the student class than the others.



(a) Location (b) Occupation

Fig. 1: AP results by class on the test dataset. Results show the F-measure performance using DPP-EXPEI approach.

4.2 Aggressiveness dectection.

We trained with two approaches: i) DPP-EXPEI using the top 2000 content attributes according to DPP measure and ii) baseline. We only submitted decisions of the baseline because it performed better. Results for AD in the test phase are showed in Table 3. The baseline obtained the sixth place in the rank of 12 proposed methods of eight teams.

Table 3: Results on aggressiveness class at the MEX.A3T competition

Dataset	Representation	F-measure	Precision	Recall
Train	baseline	**0.74**	**0.75**	**0.73**
	DPP-EXPEI	0.66	0.65	0.67
Test	baseline	0.4198	0.4225	0.4171

We believe DPP-EXPEI, in train stage, has a minor performance than the baseline because the attributes that we used are words unigrams; but the Spanish language has many aggressive phrases composed by word sequences. For example, the Table 4 shows the top words selected by DPP, there are some words that they are part of aggressive phrases. Hence, we think that baseline is better because it considers n-gram of words capturing aggressive phrases. This lays the ground for a future research path that could consider n-grams in personal phrases by means of our proposed approach.

Table 4: Examples of top words selected by DPP in the AD task of the competition. The words are shown by rank

loca, hdp, estoy, mamar, chinguen, escuela, pendejo, pendejos, fui, amo, volviendo, hijos, semana, vuelvo, hondureños, chingas, dije, mierda, valer, valiendo, puto, llevo, caliente, vergazos, salgo, vida, vuelve, soy, gringos, alma, tengo, noche, ando, siento

5 Conclusions and future work

This paper describes the proposed approach designed for the MEX.A3T 2018 tasks where two tasks are tackling: Author Profiling on *location* and *ocuppation* traits and Agressiveness Detection. The approach exploits personal information for determining traits from profile author and behavior. We consider that personal phrases expose interests, preferences, habits and routines, which help to highlight terms revealing personal traits including in Spanish texts.

The proposed approach emphasizes, in the text representation, the value of terms inside personal phrases by using DPP and EXPEI, which are schemes for feature selection and term weighting respectively. We adapted it for evaluating them using tweets written in Spanish by Mexican users. The results indicate that the approach appears to be effective in AP for Spanish, supporting the idea that personal phrases (sentences having a first-person pronoun) integrate the essence of texts for the AP task. We consider that aspect as the cornerstone to obtain the first place of the competition in this task.

On the other hand, for the AD task the results showed that the proposed approach, configured with word unigrams, has lower performance than the baseline which considers word sequences. In a future, we plan to enrich our approach by using words n-grams since the use of personal phrases in AD task is helpful to study aggressive behavior linked to personal information.

References

1. Álvarez-Carmona, M.Á., Guzmán-Falcón, E., Montes-y-Gómez, M., Escalante, H.J., Villaseñor-Pineda, L., Reyes-Meza, V., Rico-Sulayes, A.: Overview of MEX-A3T at IberEval 2018: Authorship and aggressiveness analysis in Mexican Spanish tweets. In: Notebook Papers of 3rd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IBEREVAL), Seville, Spain, September (2018)
2. Argamon, S., Dhawle, S., Koppel, M., Pennebaker, J.W.: Lexical Predictors of Personality Type. In: Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America (2005)
3. Argamon, S., Koppel, M., Pennebaker, J.W., Schler, J.: Automatically Profiling the Author of an Anonymous Text. Commun. ACM **52**(2), 119–123 (2009)
4. Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., Vakali, A.: Detecting Aggressors and Bullies on Twitter. In: Proceedings of the 26th International Conference on World Wide Web Companion. pp. 767–768. WWW '17 Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2017)
5. Del Bosque, L.P., Garza, S.E.: Aggressive Text Detection for Cyberbullying. In: Gelbukh, A., Espinoza, F.C., Galicia-Haro, S.N. (eds.) Human-Inspired Computing and Its Applications. pp. 221–232. Springer International Publishing, Cham (2014)
6. Escalante, H.J., Villatoro-Tello, E., Garza, S.E., López-Monroy, A.P., Montes-y-Gómez, M., Villaseñor-Pineda, L.: Early detection of deception and aggressiveness using profile-based representations. Expert Systems with Applications **89**, 99 – 111 (2017)
7. Franco-Salvador, M., Plotnikova, N., Pawar, N., Benajiba, Y.: Subword-based Deep Averaging Networks for Author Profiling in Social Media—Notebook for PAN at CLEF 2017. In: Cappellato, L., Ferro, N., Goeuriot, L., Mandl, T. (eds.) CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, 11-14 September, Dublin, Ireland. CEUR-WS.org (2017)
8. Koppel, M., Argamon, S., Shimoni, A.R.: Automatically categorizing written texts by author gender. Literary and Linguistic Computing **17**(4), 401–412 (2002)
9. López-Monroy, A.P., Montes-y-Gómez, M., Escalante, H.J., Villaseñor-Pineda, L.: Using Intra-Profile Information for Author Profiling. In: Cappellato, L., Ferro, N., Halvey, M., Kraaij, W. (eds.) CLEF 2014 Evaluation Labs and Workshop – Working Notes Papers, 15-18 September, Sheffield, UK. CEUR Workshop Proceedings, vol. 1180, pp. 1116–1120. CEUR-WS.org (2014)
10. López-Monroy, A.P., Montes-y-Gómez, M., Escalante, H.J., Villaseñor-Pineda, L., Stamatatos, E.: Discriminative subprofile-specific representations for author profiling in social media. Knowledge-Based Systems **89**, 134 – 147 (2015)
11. Meina, M., Brodzinska, K., Celmer, B., Czoków, M., Patera, M., Pezacki, J., Wilk, M.: Ensemble-based Classification for Author Profiling using Various Features. In: Forner, P., Navigli, R., Tufis, D., Ferro, N. (eds.) Working Notes for CLEF 2013 Conference, Valencia, Spain. CEUR Workshop Proceedings, CEUR-WS.org (2013)
12. Miura, Y., Taniguchi, T., Taniguchi, M., Ohkuma, T.: Author Profiling with Word+Character Neural Attention Network—Notebook for PAN at CLEF 2017. In: Cappellato, L., Ferro, N., Goeuriot, L., Mandl, T. (eds.) CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, Dublin, Ireland. CEUR-WS.org (2017)
13. Nguyen, D., Smith, N.A., Rosé, C.P.: Author Age Prediction from Text Using Linear Regression. In: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. pp. 115–123. LaTeCH '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011)

14. Ortega-Mendoza, R.M., Franco-Arcega, A., López-Monroy, A.P., Montes-y-Gómez, M.: I, Me, Mine: The Role of Personal Phrases in Author Profiling. In: Fuhr, N., Quaresma, P., Gonçalves, T., Larsen, B., Balog, K., Macdonald, C., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction: 7th International Conference of the CLEF Association, CLEF 2016, Évora, Portugal, September 5-8, 2016, Proceedings, pp. 110–122. Springer International Publishing, Cham (2016)
15. Ortega-Mendoza, R.M., López-Monroy, A.P., Franco-Arcega, A., Montes-y-Gómez, M.: Emphasizing personal information for Author Profiling: New approaches for term selection and weighting. Knowledge-Based Systems **145**, 169 – 181 (2018)
16. Pennacchiotti, M., Popescu, A.M.: Democrats, Republicans and Starbucks Afficionados: User Classification in Twitter. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 430–438. KDD '11, ACM, New York, NY, USA (2011)
17. Pennebaker, J.: The Secret Life of Pronouns: What Our Words Say About Us. Bloomsbury USA (2011)
18. Potthast, M., Rangel, F., Tschuggnall, M., Stamatatos, E., Rosso, P., Stein, B.: Overview of PAN'17. In: Jones, G.J., Lawless, S., Gonzalo, J., Kelly, L., Goeuriot, L., Mandl, T., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. pp. 275–290. Springer International Publishing, Cham (2017)
19. Rangel, F., Celli, F., Rosso, P., Potthast, M., Stein, B., Daelemans, W.: Overview of the 3rd Author Profiling Task at PAN 2015. In: Cappellato, L., Ferro, N., Jones, G., San Juan, E. (eds.) CLEF 2015 Evaluation Labs and Workshop – Working Notes Papers, Toulouse, France. CEUR-WS.org (2015)
20. Rangel, F., Rosso, P., Chugur, I., Potthast, M., Trenkmann, M., Stein, B., Verhoeven, B., Daelemans, W.: Overview of the 2nd Author Profiling Task at PAN 2014. In: Cappellato, L., Ferro, N., Halvey, M., Kraaij, W. (eds.) CLEF 2014 Evaluation Labs and Workshop – Working Notes Papers, Sheffield, UK. CEUR-WS.org (2014)
21. Rangel, F., Rosso, P., Koppel, M., Stamatatos, E., Inches, G.: Overview of the Author Profiling Task at PAN 2013 (2013)
22. Rangel, F., Rosso, P., Verhoeven, B., Daelemans, W., Potthast, M., Stein, B.: Overview of the 4th Author Profiling Task at PAN 2016: Cross-Genre Evaluations. In: Working Notes Papers of the CLEF 2016 Evaluation Labs. CEUR Workshop Proceedings, vol. 1609. CLEF and CEUR-WS.org (2016)
23. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E., et al.: Personality, Gender, and Age in the Language of Social Media: The Open-vocabulary Approach. PloS one **8**(9), e73791 (2013)
24. Sebastiani, F.: Machine Learning in Automated Text Categorization. ACM Computing Surveys **34**(1), 1–47 (2002)