

iTACOS at IberEval2017: Detecting Stance in Catalan and Spanish Tweets

Mirko Lai^{1,2}, Alessandra Teresa Cignarella¹, Delia Irazú Hernández Farías^{1,2}

¹ Dipartimento di Informatica, Università degli Studi di Torino

² PRHLT Research Center, Universitat Politècnica de València

Abstract In this paper we describe the iTACOS submission for the *Stance and Gender Detection in Tweets on Catalan Independence* shared task. Concerning the detection of stance, we ranked as the first position in both languages outperforming the baselines; while in gender detection we ranked as fourth and third for Catalan and Spanish. Our approach is based on three diverse groups of features: stylistic, structural and context-based. We introduced two novel features that exploit significant characteristics conveyed by the presence of Twitter marks and URLs. The results of our experiments are promising and will lead to future tailoring of these two features in a finer grained manner.

1 Introduction

Recently, there is a special interest in the task of monitoring people's stance towards particular targets; thus leading to the creation of a novel area of investigation named *Stance Detection* (SD). Research on this topic could have a positive impact on different aspects such as public administration, policy-making, and security. In fact, through the constant monitoring of people's opinion, desires, complaints and beliefs on political agenda or public services, administrators could better meet population's needs. For example, a practical application of SD could improve the automatic identification of people's extremist tendencies (i.e. religious extremism [1]).

In 2016, for the first time a shared task on SD has been held at SemEval-2016, namely the task 6: Detecting Stance in Tweets³ was organized in the framework of SemEval. The participating teams were required to determine stance towards six different targets: "Atheism", "Climate Change is a Real Concern", "Donald Trump", "Feminist Movement", "Hillary Clinton", and "Legalization of Abortion". Most of the proposed approaches exploited standard text classification features such as n-grams as well as word embeddings. More details about the participating systems can be found in [2]. In general, related work on SD is scarce, only few works have been published on this novel task. Mohammad et al. [3] took advantage of word-based and sentiment-based features to perform SD on the SemEval-2016 Task 6 dataset. Lai et al. [4], instead, proposed an approach using context features to detect stance towards two targets related to

³<http://alt.qcri.org/semeval2016/task6/>

politics in the U.S. presidential elections: Hillary Clinton and Donald Trump. The obtained results outperformed those from the shared task.

In this paper we present our participation to the *Stance and Gender Detection in Tweets on Catalan Independence* task [5] at IberEval-2017⁴. The task is articulated into two subtasks about information contained in Twitter messages written both in Catalan and Spanish: the first subtask is related to detecting author’s stance towards the independence of Catalonia, while the second one aims at identifying their gender.

Inferring people’s traits such as gender, age or native language on the basis of their written texts is investigated by a field named *Author Profiling* (AP). From 2013 onwards a shared task on AP has been organized at PAN [6,7,8,9] in the framework of CLEF⁵. The intuition behind the task of gender recognition is that of studying how language is used by people and trying to identify features, devices or patterns that are more likely exploited by one gender or the other. More details on the state-of-the-art approaches on this task can be found in [9,10].

2 Our proposal

The starting point of our proposal is to be found in the method proposed in Lai et al. [4] in which the authors exploited three diverse groups of features: *Structural* such as punctuation and other Twitter marks, *Sentiment* i.e. lexica covering different facets of affect, and finally *Context-based*, which consider the relationship that exists between a given target and other entities in its domain.

Therefore, we propose a supervised approach which consists in determining stance towards the independence of Catalonia as well as the gender of the author of a given tweet. In our work, we explored some features that can be grouped in three main categories: *Stylistic*, *Structural*, and *Context*. In the present paper we were not able to explore *Sentiment* features as in [4] due to the fact that we are not aware of sentiment lexica for Spanish and Catalan. We define a set of features distributed as follows:

- **Stylistic Features**

- Bag of Words (*BoW*)⁶
- Bag of Part-of-Speech labels (*BoP*)^{6,7}
- Bag of Lemmas (*BoL*)^{6,7}
- Bag of Char-grams (*BoC*)⁸

- **Structural Features**

⁴<http://stel.ub.edu/Stance-IberEval2017/>

⁵<http://clef2017.clef-initiative.eu/>

⁶Each tweet was pre-processed for converting it to lowercase. We used unigrams, bigrams and trigrams with a binary representation.

⁷We used TreeTagger [11,12] for extracting both the part-of-speech and lemmas.

⁸We considered chargrams of 2 and 3 characteres.

- Bag of Twitter Marks (*BoTM*). We exploit a Bag of Words considering only the words extracted from multi-word Twitter Marks (hashtags and mentions) splitting them by capital letters.
- Bag of Hashtags (*BoH*). We consider the hashtags as terms for building a vector with binary representation.
- Frequency of Hashtags (*freqHash*).
- Uppercase Words (*UpW*). This feature refers to the amount of words starting with a capital letter.
- Punctuation Marks (*PM*). We take into account the frequency of dot, comma, semicolon, exclamation and question marks.
- Length (*Length*). Three different features were considered to build a vector: number of words, number of characters, and the average of the length of the words in each tweet.

- **Context Features**

- Language (*Lan*). We create a vector exploiting the labels ES for Spanish and CA for Catalan provided by the organizer.
- URL (*Url*). We observed that tweets containing a URL are common in the training dataset. We decided to take advantage of this by considering different aspects extracted from short URLs. First, we identified if the web address of reference is or not reachable. Second, we retrieved the words contained on the web address, then we build a bag-of-words using this information.

3 Experiments and Results

The organizers provided a dataset of 8,638 tweets written in Spanish and Catalan labelled with stance (AGAINST, FAVOR, and NEUTRAL) and gender (FEMALE and MALE). For what concerns gender, the distribution is balanced among FEMALE and MALE tweets. Regarding stance, the distribution is skew towards FAVOR for Catalan and skew towards NEUTRAL for Spanish (respectively 30.66% and 29.38%). Similar trends were found in Bosco et al. [13].

It appears, therefore, that language could be a useful feature for stance detection in the Catalan independence debate concerning a region characterized by a strong bilingualism and a smoldering nationalism. In fact, «Language divides and unites us. It [...] impinges upon our identity as individuals, as members of a particular ethnic or national group, and as citizens of a given polity» [14]. We therefore believe that there is a strong correlation between stance and the exploitation of language.

In order to assess the performance of the participating systems, a test set of 2,162 unlabelled tweets was provided, and the two tasks were evaluated separately. Two different evaluation metrics were used: (1) the macro-average of F-score (FAVOR and AGAINST) was used in the case of stance detection and (2) the accuracy was selected as metric to evaluate the performance in terms of gender identification.

3.1 iTACOS experiments

In our experiments, we addressed both stance and gender detection as a classification task. The code is available on github for further exploration and for allow reproducibility of our experiments⁹. We carried out several experiments¹⁰ by combining both the features introduced in Section 2 together with a set of classifiers composed by: Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LG), Decision Tree (DT), and Multinomial Naïve Bayes (MNB). Besides, we exploited a Majority Voting (MV) strategy considering the different predictions of the above mentioned classifiers as described in Liakata et al. [15]. The features we proposed in section 2 were exploited in both the tasks of stance and gender detection, but as it will be better described in the result section, they were specifically tailored for the sole purpose of detecting stance and then they were also applied to gender. For this reason, in the present paper we will focus more on the first subtask, that of stance. We analyzed the obtained results and selected the five combinations of features that showed the best performance for the stance detection task. The resulting sets of features are shown in Table 1.

We participated in the shared task with five different runs for each language and each subtask. Table 2 shows the obtained results by using both the features and the classifier used in each of the submitted runs.

Table 1. Best-ranked sets of features using the training set

Name	Features list
Set _{α}	BoW, BoL, BoC, Url, BoTM, freqHash, UpW
Set _{β}	BoW, BoL, BoP, BoC, Url, BoH, freqHash, Length
Set _{γ}	BoW, BoL, BoP, BoC, Url, freqHash, Lan, Length
Set _{δ}	BoW, BoL, BoP, BoC, Url, freqHash, PM, Length
Set _{ϵ}	BoW, BoL, BoP, BoC, Url, BoH, PM, Lan

3.2 Official results

We ranked as the first position among 10 participating teams in the subtask of stance detection in both Catalan and Spanish. Table 3 shows the official results on the test set. At a first glance, it is possible to observe that our proposed approach seems to perform slightly better in Catalan than in Spanish. Overall, our submissions performed better in Catalan, in fact our five runs ranked among the first 8 positions. In Spanish, on the other hand, our less performing run ranked as the 18th position.

As shown in the table above, the best result in each language was not achieved by the same run. iTACOS.2 performs better for Catalan, while iTACOS.1 for

⁹<https://github.com/mirkolai/iTACOS-at-IberEval2017>

¹⁰A 10-fold cross-validation setting was used.

Table 2. Results for stance detection on the training set

Run	Stance Detection			Gender Detection		
	Features and classifier	F-score		Features and classifier	Accuracy	
		<i>Catalan</i>	<i>Spanish</i>		<i>Catalan</i>	<i>Spanish</i>
iTACOS.1	Set _{α} + SVM	0.680	0.544	Set _{ϵ} + LR	0.720	0.648
iTACOS.2	Set _{ϵ} + LR	0.633	0.544	Set _{δ} + LR	0.722	0.648
iTACOS.3	Set _{β} + LR	0.625	0.548	5x5*	0.728	0.656
iTACOS.4	5x5*	0.636	0.530	Set _{α} + MV	0.719	0.646
iTACOS.5	Set _{α} + MV	0.657	0.548	All Sets** + SVM	0.709	0.636

* The final prediction is the most frequent prediction over the 25 combinations between sets of features and machine learning algorithms.

** The final prediction is the most frequent prediction over the 5 combinations between sets of features and SVM.

Table 3. Official results for stance detection

Catalan			Spanish		
Ranking	Run	F-score	Ranking	Run	F-score
1	iTACOS.2	0.4901	1	iTACOS.1	0.4888
2	iTACOS.1	0.4885	7	iTACOS.2	0.4593
4	iTACOS.3	0.4685	12	iTACOS.3	0.4528
7	iTACOS.4	0.4490	14	iTACOS.4	0.4427
8	iTACOS.5	0.4484	18	iTACOS.5	0.4293

Spanish. The poorer results in both languages were obtained by using iTACOS.4 and iTACOS.5. As expected the best performing runs (iTACOS.1 and iTACOS.2) contain both context-based features, validating the importance of considering contextual information in stance detection tasks. For example, both runs include the feature *Url*. We are interested in evaluating the impact of such feature on the performance. For this reason, we carried out experiments on the training set by applying a modified version of iTACOS.1 and iTACOS.2 removing the *Url* feature. Looking at the results, we observed a drop in the performance of -0.029% for Catalan and of -0.002% for Spanish in iTACOS.1; and of -0.004% for Catalan and of -0.002% for Spanish in iTACOS.2.

The *BoTM*, a novel feature included in the structure-based group, emerges among the relevant features in iTACOS.1 concerning Spanish, but further inquiry on its relevance is matter of future work. For what concerns classifiers, LG and SVM achieved the best performance in both languages. Surprisingly, the approach exploiting MV is not performing.

3.3 A linguistic revision

A fundamental part of our approach has been that of manually dealing with data. Being the size of the dataset very large, we were able to visualize only a small portion of tweets. Therefore, we focused on the cases of disagreement

between the results obtained with iTACOS.1 and the golden labels provided by the organizers¹¹. Below, we report some examples both in Catalan and Spanish:

1. #elecciones #catalunya #NO #27S <https://t.co/oBuTDnUEHj>
 → #elecciones #catalunya #NO #27S <https://t.co/oBuTDnUEHj>
 LANGUAGE: CATALAN
 GOLDEN LABEL: AGAINST
 iTACOS.1: FAVOR

2. Ale @JuntsPelSi, a casa, son solo unas #eleccionescatalanas autonómicas. Mañana a trabajar que es lunes. Seguíis teniendo el mismo DNI. #27S
 → @JuntsPelSi, go at home, there is only one autonomous #eleccionescatalanas. Tomorrow, go to work that it'll be Monday. You will have the same DNI (Spanish ID). #27S
 LANGUAGE: SPANISH
 GOLDEN LABEL: AGAINST
 iTACOS.1: FAVOR

3. En estas #eleccionescatalanas de decide una posible independencia y un gobierno que vele por los derechos de su pueblo, VOTA @catsiqueespot
 → In these #eleccionescatalanas we decide for a possible independence and a government that fights for the rights of its population, VOTE @catsiqueespot
 LANGUAGE: SPANISH
 GOLDEN LABEL: FAVOR
 iTACOS.1: AGAINST

Example 1, has been marked as FAVOR from our classifier in (iTACOS.1), probably because of the misleading presence of the token “catalunya”, written in Catalan. However, the explicit semantic information carried by the hashtag #NO pointing to AGAINST was ignored, thus leading to a wrong classification. Considering Spanish, example 2 has been appointed as FAVOR instead of AGAINST. The presence of the mention @JuntsPelSi (Catalan independence coalition) could have misdirected our classification. On the other hand, the tweet in example 3 was tagged as AGAINST whereas it should have been FAVOR as we clearly infer from “VOTA @catsiqueespot” and according to the golden labels.

A manual analysis of this kind helped us to shed some light on the relevance of each single feature we exploited and, after having linguistically analyzed them, to choose which features had to be included in our final sets.

4 Conclusions

In this paper we presented an overview of the iTACOS submission for the *Stance and Gender Detection in Tweets on Catalan Independence* task at IberEval-2017. We participated by submitting five different runs in the detection of author’s stance and gender both in Twitter messages in Catalan and Spanish. Our approach, chiefly based on *context* and *structural* features, proved to be highly

¹¹The tweets have been extracted from the training set.

successful concerning the task of stance in both languages, as our system ranked as the first position among ten participating teams. The results show that the addition of two particular features, namely *BoTM* and *Url*, produced a significant contribution to Stance Detection task. In the future, we plan to tailor these two features we used in an even finer grained manner.

References

1. Hogan, B.: The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. *Bulletin of Science, Technology & Society* **30** (2010) 377–386
2. Bethard, S., Cer, D.M., Carpuat, M., Jurgens, D., Nakov, P., Zesch, T., eds.: Proceedings of the 10th International Workshop on Semantic Evaluation. In Bethard, S., Cer, D.M., Carpuat, M., Jurgens, D., Nakov, P., Zesch, T., eds.: *SemEval@NAACL-HLT 2016*, San Diego, CA, USA, June 16-17, 2016, The Association for Computer Linguistics (2016)
3. Mohammad, S.M., Sobhani, P., Kiritchenko, S.: Stance and Sentiment in Tweets. *CoRR* [abs/1605.01655](https://arxiv.org/abs/1605.01655) (2016)
4. Lai, M., Hernandez Farias, D.I., Patti, V., Rosso, P.: Friends and Enemies of Clinton and Trump: Using Context for Detecting Stance in Political Tweets. In Sidorov, G., Herrera-Alcántara, O., eds.: Part I. *Lecture Notes in Artificial Intelligence. Advances in Computational Intelligence. 15th Mexican International Conference on Artificial Intelligence, MICAI 2016*. Volume 10061. (2016) 152–165
5. Taulé, M., Martí, M.A., Rangel Pardo, F.M., Rosso, P., Bosco, C., Patti, V.: Overview of the task of Stance and Gender Detection in Tweets on Catalan Independence at IBEREVAL 2017. In: *Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017)*, CEUR Workshop Proceedings. CEUR-WS.org, 2017, Murcia, Spain (2017)
6. Rangel Pardo, F.M., Rosso, P., Koppel, M., Stamatatos, E., Inches, G.: Overview of the author profiling task at PAN 2013. In: *CLEF Conference on Multilingual and Multimodal Information Access Evaluation, CELCT (2013)* 352–365
7. Rangel Pardo, F.M., Rosso, P., Potthast, M., Trenkmann, M., Stein, B., Verhoeven, B., Daeleman, W., et al.: Overview of the 2nd author profiling task at pan 2014. In: *CEUR Workshop Proceedings*. Volume 1180., CEUR Workshop Proceedings (2014) 898–927
8. Rangel Pardo, F.M., Rosso, P., Potthast, M., Stein, B., Daelemans, W.: Overview of the 3rd Author Profiling Task at PAN 2015. In: *CLEF, sn (2015)*
9. Rangel Pardo, F.M., Rosso, P., Verhoeven, B., Daelemans, W., Potthast, M., Stein, B.: Overview of the 4th author profiling task at PAN 2016: cross-genre evaluations. In Balog, K., Cappellato, L., Ferro, N., Macdonald, C., eds.: *CLEF 2016 Labs and Workshops, Notebook Papers*. CEUR Workshop Proceedings. Volume 1609., Évora, Portugal (2016) 750–784
10. Rangel Pardo, F.M., Rosso, P.: Use of language and author profiling: Identification of gender and age. *Natural Language Processing and Cognitive Science* **177** (2013)
11. Schmid, H.: Part-of-speech tagging with neural networks. In: *Proceedings of the 15th conference on Computational linguistics-Volume 1*, Association for Computational Linguistics (1994) 172–176
12. Schmid, H.: Treetagger— a language independent part-of-speech tagger. *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart* **43** (1995) 28

13. Bosco, C., Lai, M., Patti, V., Rangel Pardo, F.M., Rosso, P.: Tweeting in the Debate about Catalan Elections. In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., eds.: LREC workshop on Emotion and Sentiment Analysis Workshop (ESA), LREC-2016, Portorož, Slovenia, European Language Resources Association (ELRA) (2016) 67–70
14. Millar, R.: *Language, Nation and Power: An Introduction*. Springer (2005)
15. Liakata, M., Kim, J.H., Saha, S., Hastings, J., Rebholz-Schuhmann, D.: Three hybrid classifiers for the detection of emotions in suicide notes. *Biomedical informatics insights* **5** (2012) 175