Stance detection at IberEval 2017: A Biased Representation for a Biased Problem

Diego Aineto García and Antonio Manuel Larriba Flor

Universitat Politècnica de València
{dieaigar,anlarflo}@inf.upv.es

Abstract. In this paper, we explain our approach to the task Stance and Gender Detection in Tweets on Catalan Independence whose objective is to detect the author's posture towards the topic of Catalan independence. An in-depth description of our system submitted to the task is presented as well as an overview of the experiments which leads to this solution during the development of the system. Our best solutions are based on a biased representation which allows us to train with the whole dataset.

1 Introduction

Stance detection is the task of identifying the favorability of a given piece of text towards a particular target. This task shares many similarities with sentiment analysis, which aims to detect whether a text is objective or subjective and, for the latter case, evaluate its positiveness or negativeness. The main difference lies in that, in stance detection, the system needs to be able to determine if the texts is in favor, against or neutral towards a topic that may not even explicitly appear in the text. However some differences exist, an opinion might be positive and against the topic.

The difficulty of this task is aggravated for the case of microblogs, such as *Twitter*, where the texts are very short. Moreover, in these kinds of platforms it is common for people to use informal language or even shortened words and emotes. Nonetheless, stance detection has many real-world applications, such as opinion mining and author profiling, and many companies and organizations are showing increasing interest in the area.

In the following section we will describe the task at hand (Section 2), explain the approach followed in the system we submitted (Section 3), and show the results obtained for the different system configurations tried out during our experiments (Section 4). Finally, we will present our conclusions in Section 5.

2 Task Description

The Stance and Gender Detection in Tweets on Catalan Independence[1] task is one of the task proposed at the IberEval 2017 workshop. The goal of this

---

[1] http://stel.ub.edu/Stance-IberEval2017/

task is to detect the author's gender and the stance with respect to the topic "Independence of Catalonia" in a collection of tweets written in Spanish and Catalan. Teams are allowed to participate in either stance detection or gender and stance detection. Our team participated only in stance detection so we will omit anything regarding gender detection from this point on.

The distribution for stances for both Spanish and Catalan tweets is shown in Table 1. As can be seen in the table, the distribution along the three possible stances is not uniform, a situation that usually arises in this kind of problem. However, the bias towards some stances is very accentuated in this particular problem with 'Favor' and 'Neutral' comprising 97% of the Catalan tweets, and 'Against' and 'Neutral' comprising the 92% of the Spanish tweets. The reduced size of the corpus and the skewed distribution makes training an accurate system a challenging class.

Table 1. Training dataset statistics

Language	Favor	Against	Neutral	Total
Spanish	335 (7.8%)	1446 (33.5%)	2538 (58.7%)	4319
Catalan	2648 (61.3%)	131 (3%)	1540 (35.7%)	4319

The evaluation of the systems is performed following the metric proposed at SemEval 2016 - Task 6 [3]. The metric is a macro average of the F1-score for 'Favor' and the F1-score for 'Against'. This metric does not disregard the 'Neutral' class and will still be negatively affected by falsely labeling 'Neutral' instances.

$$F_{avg} = \frac{F_{favor} + F_{against}}{2} \tag{1}$$

3 System Description

This section explains our approach to the task and describes the system we submitted to the shared task. The system we developed tries to take advantage of the particularities of the problem by making use of a problem-oriented preprocessing and representation.

3.1 Preprocessing

Our method starts with a preprocessing that follows an approach similar to the one described in [2], but adding a few more steps to take into account some problem-specific particularities. Following, the list of modifications performed on the tweets:

– Convert the text to lowercase

- Remove stopwords

- All URLs are replaced by keyword *URL*

- All user mentions are replaced by keyword *MENTION*

- Certain hashtags which could be found mostly in tweets labeled as 'Against' like '#siqueespot' and '#iceta27s' are replaced by a keyword *AGAINST_STANCE*

- Certain hashtags which could be found mostly in tweets labeled as 'Favor' like '#cup' and '#guanyemjunts' are replaced by a keyword *FAVOR_STANCE*

- All other hashtags are replaced by a keyword *HASHTAG*

- When multiple question marks are found together they are replaced by a token *MULTIPLEQUESTIONMARKS*

- When multiple exclamation marks are found together they are replaced by a token *MULTIPLEEXCLAMATIONMARKS*

- When mixed question and exclamantion marks are found together like '!?', they are replaced by a token *MIXEDMARKS*

- Given that the collection of tweets takes place during the regional elections, many percentages can be found in the tweets which may indicate objectivity. As such, we decided to replace such percentages with a keyword *PERCENT-AGE*

- All punctuation marks save for exclamation and questions marks are removed

Following these steps a tweet like 'Así están las cosas con el 60% escrutado. #eleccionescatalanas #27S http://t.co/XlLf8RqRZ8' is converted to 'asi estan cosas PERCENTAGE escrutado HASHTAG HASHTAG URL'

3.2 Features and Representation

The system uses a model based on TF-IDF vectors and two separate sets of features. The first set of features is built from unigrams extracted from the preprocessed tweet text. For this set of features only the top 2500 most frequent unigrams from the vocabulary extracted from the tweet texts are used. A second set of features built from hashtags is also used, in this case using only the top 500 most frequent ones. Both sets of features are extracted separately for tweets in Catalan and tweets in Spanish.

Although the sets of features is extracted separately, the feature vector for each tweet is built by concatenating the vectors from both Catalan and Spanish features as shown in Figure 1. This representation aims to depict the heavy bias found in the distribution of stances in two ways. First, a tweet in Catalan would be less likely to hold an against stance while a tweet in Spanish would rarely

hold an in favor stance. By clearly differentiating two parts in the feature vector we want to transmit this bias to the classifier. Second, because Catalan and Spanish have many words in common, the same word may be used to transmit different postures towards the topic. This also applies to hashtags, and thus, it is important to differentiate if a word or hashtag was used in a language or another.
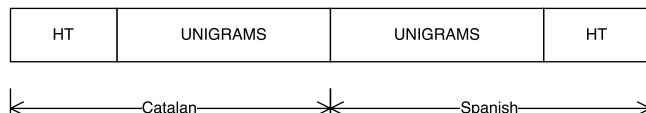
HT	UNIGRAMS	UNIGRAMS	HT

$\longleftarrow$————Catalan————$\longleftarrow\!\!\!\longrightarrow$————————Spanish————————$\longrightarrow$

Fig. 1. Feature vector. HT = hashtags and UNIGRAMS = unigrams extracted from the preprocessed text

3.3 Classifiers

Two different systems were submitted to the workshop, one based on support vector machine (SVM) and one based on artificial neural networks (ANN). The system using the SVM classifier implements a one-vs-rest multiclass strategy and uses a linear kernel. On the other hand, the ANN-based system implements a 3 hidden layer fully-connected network with 2000 neurons on the first layer, 1000 neurons on the second, and 500 on the last. All layers use ReLu as activation function (save for the output layer which uses SoftMax), and dropout of 40%.

4 Experiments

During the development phase, we tried out different sets of features, representations and parameters and evaluated them using 5-fold cross-validation. Table 2 displays the results obtained for each experiment, measuring the F1-score for each stance and the official metric used in the task evaluation (F_{avg}) obtained through equation 1.

S0 is a system trained using only unigrams as features and basic preprocessing but implementing the representation depicted in Figure 1. We used this as base for the other systems developed and as baseline to evaluate them. Next, we present a brief explanation of each of the systems evaluated:

– S1: In an attempt to evaluate whether our 2-part feature vector added any value over a common representation, we decided to treat the Catalan and Spanish datasets as two different subtasks. As such, we trained two separate systems using only the dataset corresponding to the language in question. Results show that our representation achieves better results, also because a common feature vector allows us to train using double the data.

Table 2. Results obtained for the different configurations used during the development

Configuration	Language	F_{favor}	$F_{against}$	$F_{neutral}$	F_{avg}
S0: Unigrams	Catalan	0.8438	0.4163	0.7512	0.6300
	Spanish	0.3574	0.5983	0.7881	0.4896
S1: Sep. models	Catalan	0.8332	0.4149	0.7165	0.6241
	Spanish	0.3100	0.6217	0.7955	0.4541
S2: S0 + St. HT	Catalan	0.8267	0.1612	0.7480	0.4939
	Spanish	0.2252	0.6061	0.8041	0.4156
S3: Final SVM	Catalan	0.8483	0.5288	0.7578	**0.6885**
	Spanish	0.4040	0.6291	0.8159	0.5165
S4: Final ANN	Catalan	0.8326	0.5175	0.7503	0.6751
	Spanish	0.4228	0.6318	0.8217	**0.5273**
S5: S3 + length + init	Catalan	0.8475	0.5142	0.7594	0.6809
	Spanish	0.3990	0.6356	0.8174	0.5173

- S2: This system uses an extra feature indicating if the text contains any hashtag associated to a stance, i.e., hashtags that can only be found in tweets labeled with one particular stance. This approach didn't render good results, but replacing these hashtags with a keyword during the preprocessing provided a slight improvement.

- S3 and S4: These are the systems described in Section 3, and the ones submitted to the shared task. Both use the same feature vector but S3 employs a SVM while S4 an ANN.

- S5: This last system is built upon S3, but adds some new features in the form of length of tweet in words and initial unigrams [1].

As can be seen in table 2 (bolded numbers) we got the best results using support vector machine (SVM) for Catalan and artificial neural networks (ANN) for Spanish.
These systems allowed us to classify as the third team for Catalan and the fifth for Spanish in the shared task [4]. Results obtained over the test set can be seen in table 3. S1 systems are based on SVM and s2 on ANN. We provide the system and team ranking for each task.

Table 3. Results over the test set

Language	System	F_{avg}	System Ranking	Team Ranking
Spanish	ARA1337.s1	0.453	11	5
	ARA1337.s2	0.431	16	
Catalan	ARA1337.s1	0.465	5	3
	ARA1337.s2	0.451	6	

5 Conclusion

In this paper we describe our submission to the task Stance and Gender Detection in Tweets on Catalan Independence at IberEval 2017. The problem considered in this task has many particularities which we tried to take advantage of by developing a system that relies heavily in preprocessing and representation. SVM and ANN have offered similar results, our guess it is that with more data to train those differences will grow and we will have the possibility to try new architectures.

References

1. Anand, P., Walker, M., Abbott, R., Tree, J.E.F., Bowmani, R., Minor, M.: Cats rule and dogs drool!: Classifying stance in online debate. In: Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. pp. 1–9 (2011)
2. Krejzl, P., Steinberger, J.: UWB at semeval-2016 task 6: Stance detection. In: Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016. pp. 408–412 (2016)
3. Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: Semeval-2016 task 6: Detecting stance in tweets. In: Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016. pp. 31–41 (2016)
4. Taulé, M., Martí, M.A., Rangel, F., Rosso, P., Bosco, C., Patti, V.: Overview of the task of stance and gender detection in tweets on catalan independence at ibereval 2017. In: Notebook Papers of 2nd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IBEREVAL), Murcia, Spain, September 19, CEUR Workshop Proceedings (2017)