TASS 2017: Workshop on Semantic Analysis at SEPLN, septiembre 2017, págs. 85-90

C100TPUCP at TASS 2017: Word Embedding Experiments for Aspect-Based Sentiment Analysis in Spanish Tweets

C100TPUCP en TASS 2017: Experimentos con Word Embeddings para Análisis de Sentimiento basado en Aspectos sobre Tweets en Español

Franco Tume Fiestas
Grupo de Reconocimiento de Patrones e Inteligencia Artificial Aplicada
Pontificia Universidad Católica del Perú
Lima, Perú
a20110060@pucp.pe

Marco A. Sobrevilla Cabezudo
Grupo de Reconocimiento de Patrones e Inteligencia Artificial Aplicada
Pontificia Universidad Católica del Perú
Lima, Perú
msobrevilla@pucp.edu.pe

Abstract: Aspect-Based Sentiment Analysis is in charge of study the opinion of people about different aspects from a certain entity. This task is challenging and highly relevant for the Natural Language Processing community. In this paper, we report the participation of C100T-PUCP team in the TASS 2017 for the second task about sentiment analysis. In this edition, we used word embeddings to get the similarity between words selected from a training set that had tweets about political parties from Spain and made a model to classify each polarity of each aspect for each tweet. The results showed that using more examples to training the model with this approach is more convenient. Moreover, the proposed approach avoids the problem of the classical methods that are oriented to a specific training data set.
Keywords: Word Embeddings, Sentiment Analysis, Twitter

Resumen: El Análisis de Sentimientos basado en Aspectos está encargado del estudio de las opiniones de las personas sobre diferentes aspectos de cierta entidad. Esta tarea es desafiante y muy importante para la comunidad de Procesamiento de Lenguaje Natural. En este trabajo se describe la participacion del equipo C100T-PUCP en el TASS 2017 para la segunda tarea sobre analisis de sentimientos. En esta edición, usamos word embeddings para conseguir la similitud entre diferentes palabras seleccionadas del conjunto de datos de entrenamiento, la cual contiene tweets sobre grupos políticos de España y construimos un modelo para clasificar la polaridad de los sentimientos expresados sobre cada aspecto en cada tweet. Los resultados indicaron que el utilizar mayor cantidad de ejemplos para entrenar el modelo con este método es conveniente. Además, el enfoque propuesto evita los problemas de metodos clásicos que están orientados a un set de datos de entranamiento específico.
Palabras clave: Word Embeddings, Análisis de Sentimiento, Twitter

1 Introduction

The 6th edition of the TASS workshop consists of two task in sentiment analysis focusing of Spanish tweets: (1) polarity classification at global level and (2) aspect-based sentiment analysis in which the goal is to predict the polarity of tweets in relation to a set of identified aspects (Martínez-Cámara et al., 2017).

The task of polarity classification has been taken with many different approaches. One of them consists in representing a word as a vector and using it to get a similarity with other words. This method is called word embeddings. Word Embeddings is a well-know technique to get a vector of a word in natural language processing. Although this method is widely used in English, there are few implementations of this approach for Spanish.

In this sense, many studies use deep learn-

ISSN 1613-0073

ing as a main approach to tackle sentiment analysis so they can get a similarity between a bag of words that represent each sentiment (Alvarez-López et al., 2016). With this comparison we can get a vector features and use it with classical machine learning algorithms. Our system uses this kind of approach to classify the sentiment of each aspect of each tweet presented in the task.

This paper summarizes the participation of the C100T-PUCP team from Pontificia Universidad Católica del Perú in the second task of the workshop. In this edition, we propose a word embedding-based approach to tackle the problem of aspect-level polarity classification. Firstly, we obtain a word embeddings set from politics corpus to get similarity between tweets. Then, we explored a feature selection method for unbalanced data. Finally, we built experiments with some classifiers using word embeddings and the obtained features.

This paper is organized as follows: an overview of related works is shown in Section 2. Section 3 presents the analyzed corpus and its class distribution. The system description is described in Section 4. Section 5 shows the experimentation and results, and finally, Section 6 presents some conclusions and future works.

2 Related Work

In TASS 2014, an Aspect Detection and Aspect-based Sentiment Analysis Task were proposed (Román et al., 2015). The corpus was composed by tweets related to the final game of the "Copa del Rey" in Spanish called Social-TV. In general, there were two works submitted to these two tasks. The first one proposed a method to detect aspects based on the match of a tweet content with a pre-specified set of features related to the football domain (Vilares et al., 2014). This method obtained a F1-measure of 0.854. To identify the polarity on each aspect, the authors used a supervised method with syntactic-based features. The method obtained a F1-measure of 0.546. The second one proposed an aspect detection method based on a list of features and a set of regular expressions (Hurtado and Pla, 2014). This method obtained a F1-measure of 0.909. In the polarity detection, the authors proposed a supervised method which used as features a list of positive and negative terms and a list of words obtained from the training corpus sorting by TF-IDF. The results obtained showed a F1-measure of 0.587.

In TASS 2015, only the Aspect-based Sentiment Analysis Task was proposed, but a new corpus was added in the evaluation (Villena-Román et al., 2015). This corpus was composed by tweets related to Politics in Spanish, called STOMPOL. In this edition, a method similar to that presented in TASS 2014 was proposed in (Hurtado, Pla, and Buscaldi, 2015). This method included an additional dictionary and SVM algorithm. This method showed an accuracy of 65.50% in Social-TV corpus and 63.3% in STOMPOL corpus and the F-measure was not shown. Another method was presented which used a set of lexical and morphosyntactic features in a supervised learning algorithm (Araque et al., 2015). The way to tackle the problem was divided into three steps: (1) identifying entities, (2) getting the context (using a graph-based algorithm) and (3) executing the supervised learning algorithm. Their method obtained an accuracy of 63.5% and a F-measure of 0.606 in Social-TV corpus. Finally, The third work proposed a deep learning-based approach (Vilares et al., 2015). These authors used a LSTM Neural Network to tackle the problem of polarity detection. Their method obtained an accuracy of 61.00% in Social-TV corpus and 59.9% in STOMPOL corpus. The F-measure was not shown in this work.

In TASS 2016, two proposals were submitted to the Aspect-based Sentiment Analysis task (Villena-Román et al., 2016) on STOMPOL corpus. The first one applied a supervised algorithm using features as Aspect, Lemma, POS-Tag, Negation and Word Tokens from the training corpus (Alvarez-López et al., 2016). The result obtained was a F1-measure of 0.463. Finally, the other method proposed an experimentation of different supervised algorithms using the same features as TASS 2015 (Hurtado and Pla, 2016). Their best method obtained a F1-measure of 0.526.

3 STOMPOL Corpus

The STOMPOL corpus is composed of tweets in Spanish about Spanish elections of 2015. This corpus was presented in TASS 2014 (Román et al., 2015). Each tweet is related to one of the following aspects: Economics, Health System, Education, Political party and other aspects. Also, each aspect is related to one of these sentiments: positive, negative and neutral. The distribution of each aspect in the training data and the distribution for each sentiment per aspect is shown in Table 1. As shown, this dataset presents unbalanced classes and aspects. For example, there are too many samples about the Political Party aspect and also about negative sentiment.

4 System Description

The system presented in this edition of the TASS uses a preprocessing removing stopwords using NLTK tool. Also, words like URL's and special characters are removed from the tweets. After this, all the words are passed through Freeling lemmatizer, version 4.0. Furthermore, hashtags and labels to users are kept in this processing. As using this preprocessing the tokenization for all the tweets is completed.

The Aspect-based Sentiment Analysis was tackled as a classification problem. For this, support vector machines (SVM) and adaptive boosting (AdaBoost) classifiers were used because of the precedence in previous works that showed that they behave well in classifying long vector features. For these models, scikit-learn implementations are used from the toolkit. These, are sklearn.ensemble.AdaBoostClassifier for the adaptive boosting and sklearn.svm.SVC for the SVM implementation with a polynomial kernel.

Also each vector was filled with the cosine similarity between each feature and the top 50 most important words for each sentiment using a probabilistic appearance metric (Liu, Loh, and Sun, 2009) to give context to the aspect. This similarity was calculated using the vector representation of the words. For this, Mikolov Word2Vec model was used. Finally, each model was verified using cross fold validation with 10 iterations and using a learning curve to verify that our models are not over-fitted.

4.1 Word Embeddings Generation

A word embeddings model was created using Mikolov Word2Vec model (Mikolov et al., 2013) of GenSim implementation. The sentences selected by the model were from the same domain of the corpus, i. e., politics.

Tweets were selected using search queries related to political parties from South America and Spain in a range from 2012 to 2015. Even though many tweets were selected for this approach much more data was needed, In that sense, online news websites were also scrapped. Specifically, we obtained texts from "El país"[1], "ABC" [2] and "20 Minutos"[3] Spanish newspapers. From these sites only political news were selected and no range in time was used.

After this, we got 400MB of data which included about 5 million sentences, 30 millions words and 1.5 million unique words. The corpus was preprocessed following the next steps:

- removing stopwords and special characters that are common in tweets as "..." and URLs
- removing words that have only one character
- removing numbers
- lemmatizing words using Freeling[4], keeping all mentions to users and hashtags

Additionally, mentions to particular entities in news data were replaced with their user ids from Twitter. This was to make the news data the most similar to the twitter data so the embeddings get the same relations between words. For example, if we have a new that mentions Pablo Iglesias we replace it with his Twitterś id Pablo_Iglesias_ and we do the same with other common political persons.

To create the word embeddings from the corpus, we used the Word2vec model(Mikolov et al., 2013). This model use two type of neural network to generate the embeddings. In this case, Skip-gram model was used because it is recommended using it when the training data for the model is small. After this, we tested different values for the model parameters. The hyperparameters that were tested were: minimum word count, context window and size of the

[1] Available in https://elpais.com/
[2] Available in http://www.abc.es/
[3] Available in http://www.20minutos.es/
[4] Available in http://nlp.lsi.upc.edu/freeling/

Aspects	Tweets	Negative	Neutral	Positive
Economics	117	78	20	19
Health	21	7	10	4
Education	30	16	12	2
Political Party	777	431	176	170
Others	98	54	22	22
Total	1043	586	240	217

Table 1: STOMPOL Corpus distribution

vector. The values were selected based on experiments and those selected are in Table 2.

Hyperparameter	Value
Minimum Word Count	100
Vector Size	300
Context Window	6

Table 2: Hyperparameter Values

4.2 Aspect-based Sentiment Analysis

For this task, a window of three words were selected from the previously identified aspect in the training corpus to extract the features. Each tweet in the training corpus was also preprocessed in the same way as the word embeddings model, removing stopwords, special characters and URLs. After this, the corpus was lemmatized keeping user tags and hashtags. The analysis was based on the cosine similarity of these words with the words selected as a feature for detecting the sentiment.

In this step, firstly, we needed to create a dictionary with words that represent in a better way each sentiment that was classified using the labels for the sentiments in the training data. Thus, we used two metrics, TF-IDF and probabilistic occurrence (Liu, Loh, and Sun, 2009), to see how relevant is a word for a sentiment so these words could be used as a dictionary for each sentiment. Thus, there were more words to compare the selected words in the context of the detected aspect that represents the sentiment. Also, these metrics were applied for a new dictionary but based for each aspect relating them with each sentiments, so, a more versatile feature set was extracted.

After selecting the features, two vectors were created: (1) based on probabilistic occurrence and (2) based on probabilistic occurrence per aspect. Each vector was filled in two ways, one was only using the traditional approach, i. e., the vector a bag-of-words and filling the vector with 1 if the word feature occurs in the window and the other way is to fill the feature with the most similarity value to the feature. Then each vector was used to train a SVM and AdaBoost models. In general, we sent three runs which are shown as below:

- Run 1: The first run consisted of using vector filled with the cosine similarity between each word and the words in the dictionary of polarities. This vector was used to train a SVM model with gamma value of 0.31622776601683794, C 1 and degree 2.
- Run 2: The second run was using the same vector were the hyper parameters were: gamma was 0.0316, C was 1122.018 and degree was 2.
- Run 3: The last run was using AdaBoost Classifier with a modified vector that have the most representative word of each sentiment but for each aspect. Also this vector has one hot encoding based on each aspect. This model was trained with a Naive Bayes classifier as the weak classifier. The AdaBoost model was created with the following hyperparemeters selected by experimentation: learning rate was 0.000001, number of estimators was 100.

5 *Results and Discussions*

For the experiments a SVM classifier and an AdaBoost classifier were tested using each set of features prepared using the dictionaries previously created. In order to test these approaches F1-score (F1), Precision (P) and (R) recall were used. These metrics evaluate how well the model predicts based on how many true-positive it predicts and the inverse how many false-positive it predicts.

In the other hand, accuracy just measured how many hits it does in the prediction. Furthermore, to test the model to detect the sentiment polarity the TASS experiment page was used. This page measure a macro F1-score, recall, precision and accuracy based on the combination of how well it predicts each aspect in an specific entity.

The results for the polarity classification are presented in Table 3. In this case a SVM classifier and an AdaBoost classifier using a Naive Bayes model were tested. Also, as discussed earlier, the data is not well balanced, so, for these methods a SMOTE oversampling was applied. For these results, the SVM-run2 model was created by using the features using probabilistic weights of the words. In this sense, this model predicts the sentiment based on the similarity of the word selected as the sentiment representatives with each word that represent a feature in the vector, keeping the best score of these words. This vector was also used for the other models except for the ADA-run3 model.

The SVM-run1 considers that every word may have different meanings for each context so it uses a vector that has words that represents all the sentiments for each aspect and a feature that represents from which element the tweet came. This model has better results in comparison to its similar SVM that only use top words in all the training set and have a better accuracy compared with the ADA-run3 but not a better F1-Score.

Execution	P	R	F1	Ac.
run1	0.418	0.413	0.415	0.563
run2	0.412	0.426	0.416	0.517
run3	0.452	0.438	0.445	0.528

Table 3: Results of the different experiments

As seen, the model was not trained with a big amount of data but it still had almost all the words inside it's vocabulary. This pointed us that the errors for the sentiment polarity were caused probably by the unbalance data for each sentiment per aspect. The sentiments were in it's majority negative and few were neutral or positive. Although, using this approach gave us results that were close to other participants that uses domain specific systems.

6 Conclusions and Future Works

In the 6th edition of the TASS 2017, we have tried a word embedding approach to classify sentiment of an aspect. With this approach we generate a vector feature of how similar each selected word is to the features in the vector. The performance of the models has been compared with the models of the other participants of the TASS.

Although the model was not trained with a big amount of data that it was required, this got the similarity for most of the words in the training set. Also, the results were close to the other participants. The obtained results may give us an idea about how well word embeddings could perform in sentiment analysis due to word embeddings are not adjusted to a specific set of words like traditional methods (using Bag-Of-Words).

With these results, we propose to test this approach using a more balanced data for aspects and sentiments for training. Also, getting more sentences to train the word2vec model could improve the differentiation of the words that are used in the features. Using this improvement could get better results because as we saw the models can learn to predict a class very well if they have a lot of information about it.

References

Alvarez-López, T., M. F. Gavilanes, S. García-Méndez, J. Juncal-Martínez, and F. J. González-Castano. 2016. Gti at TASS 2016: Supervised approach for aspect based sentiment analysis in Twitter. In *Proceedings of TASS 2016: Workshop on Semantic Analysis at SEPLN (TASS 2016)*, pages 53–57.

Araque, O., I. Corcuera, C. Román, C. A. Iglesias, and J. F. Sánchez-Rada. 2015. Aspect based sentiment analysis of Spanish tweets. In *Proceedings of TASS 2015: Workshop on Semantic Analysis at SEPLN (TASS 2015)*, pages 29–34.

Hurtado, L.-F. and F. Pla. 2014. Elirf-upv en TASS 2014: Análisis de sentimientos, detección de tópicos y análisis de sentimientos de aspectos en Twitter. *Procesamiento del Lenguaje Natural.*

Hurtado, L.-F. and F. Pla. 2016. Elirf-upv en TASS 2016: Análisis de sentimientos en Twitter. In *Proceedings of TASS 2016:*

Workshop on Semantic Analysis at SEPLN (TASS 2016), pages 47–51.

Hurtado, L.-F., F. Pla, and D. Buscaldi. 2015. Elirf-upv en TASS 2015: Análisis de sentimientos en Twitter. In *Proceedings of TASS 2015: Workshop on Semantic Analysis at SEPLN (TASS 2015)*, pages 75–79.

Liu, Y., H. T. Loh, and A. Sun. 2009. Imbalanced text classification: A term weighting approach. *Expert systems with Applications*, 36(1):690–701.

Martínez-Cámara, E., M. C. Díaz-Galiano, M. A. García-Cumbreras, M. García-Vega, and J. Villena-Román. 2017. Overview of tass 2017. In J. Villena Román, M. A. García Cumbreras, D. G. M. C. Martínez-Cámara, Eugenio, and M. García Vega, editors, *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN (TASS 2017)*, volume 1896 of *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Román, J. V., E. M. Cámara, J. G. Morera, and S. M. J. Zafra. 2015. TASS 2014-the challenge of aspect-based sentiment analysis. *Procesamiento del Lenguaje Natural*, 54:61–68.

Vilares, D., Y. Doval, M. A. Alonso, and C. Gómez-Rodrıguez. 2014. Lys at TASS 2014: A prototype for extracting and analysing aspects from Spanish tweets. In *Proceedings of TASS 2014: Workshop on Semantic Analysis at SEPLN (TASS 2014)*.

Vilares, D., Y. Doval, M. A. Alonso, and C. Gómez-Rodríguez. 2015. Lys at TASS 2015: Deep learning experiments for sentiment analysis on Spanish tweets. In *Proceedings of TASS 2015: Workshop on Semantic Analysis at SEPLN (TASS 2015)*, pages 47–52.

Villena-Román, J., M. Á. G. Cumbreras, E. M. Cámara, M. C. Díaz-Galiano, M. T. Martín-Valdivia, and L. A. U. López. 2016. Overview of TASS 2016. In *Proceedings of TASS 2016: Workshop on Semantic Analysis at SEPLN (TASS 2016)*, volume 1702 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Villena-Román, J., J. García-Morera, M. A. G. Cumbreras, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. U. López. 2015. Overview of TASS 2015. In *Proceedings of TASS 2015: Workshop on Semantic Analysis at SEPLN (TASS 2015)*, volume 1397 of *CEUR Workshop Proceedings*, pages 13–21. CEUR-WS.org.