DSIC-ELIRF at SemEval-2016 Task 4: Message Polarity Classification in Twitter using a Support Vector Machine Approach

Víctor Martínez **Ferran Pla** **LLuís-F. Hurtado**
Universitat Politècnica de València
Camí de Vera s/n, 46022 València
{vmartinez2,fpla,lhurtado}@dsic.upv.es

Abstract

This paper contains the description of our participation at task 4 (sub-task A, Message Polarity Classification) of SemEval-2016. Our proposed system consists mainly of three steps. Firstly, the preprocessing step includes the tokenization and identification of special elements including URLs, hashtags, user mentions and emoticons. The second step aims at selecting and extracting the feature set. Finally, a supervised approach, in particular a Support Vector Machine has been applied to tackle the classification problem.

1 Introduction

In the last few years, Twitter has become a source of a huge amount of information which introduces endless possibilities of research in the field of Sentiment Analysis. Sentiment Analysis, also called Opinion Mining, is a research area within Natural Language Processing whose aim is to identify the underlying emotion of a certain document, sentence or aspect (Liu, 2012). As a case in point, Opinion Mining has been applied for recognizing reviews as recommended or not recommended (Turney, 2002) and for generating aspect-based summaries (Hu and Liu, 2004).

The goal of SemEval-2016 task 4 (Nakov et al., 2016) consists of categorizing tweets as positive, negative or neutral concerning the opinion that a user holds with regard to a certain topic. One issue to take into consiteration is that the language adopted in Social Media, especially in Twitter, needs to be treated differently than normalized language due to the use of specific characteristics such as users, hashtags, emoticons and slang as well as some linguistic phenomena including sarcasm and irony.

Our system is closely related to (Giménez et al., 2015). Section 2 describes the proposed method which consists mainly of three steps. Firstly, the preprocessing step includes the tokenization and identification of special elements including URLs, hashtags, user mentions and emoticons. The second step aims at selecting and extracting the feature set. Finally, a supervised approach such as Support Vector Machine (SVM) has been applied to tackle the classification problem. In section 3, the experiments carried out are described. Finally, section 4 discusses the results obtained for the different experiments in the tuning phase and in the official competition.

2 System Overview

In this section, we describe the steps carried out in this work to achieve the results obtained in Semeval 2016. In this approach, a matrix of ocurrences, in which tweets are represented as rows and features as columns, normalized by tf-idf was used to represent whether a certain feature appears or not in a tweet.

2.1 Preprocessing

After fetching all the data from Twitter, our corpus needs to be preprocessed. As Twitter makes an extensive use of emoticons, URLs and concrete elements such as @User mentions and #hashtags, some regex are utilized to substitute these mentioned elements of special interest by labels of the form <URL>, <HASH>, <USER>and <EMOTICON>that let us count the amount of ap-

Proceedings of SemEval-2016, pages 198–201,
San Diego, California, June 16-17, 2016.

pearances in a certain tweet. Indeed, after tokenizing the tweet, punctuation and stop words are removed.

2.2 Feature Set

In this paper, the following features have been tried out althought not all were included for the final submision: see section 4

N-grams at word-level were selected ranging from 1-grams to 6-grams. These were combined in the experimentation process.

Skip-grams at the word level with 2 words and 1 gap between them. As an example, ”What an amazing film” will generate the following list of skip-grams [(”What”,”amazing”),(”an”,”film”)]

K most frequent Skip-grams. This feature takes the k-most frequent Skip-grams and discards the other ones which are under the k threshold.

Lexicons

1. Jeffrey (Hu and Liu, 2004): This lexicon contains two sets of words: a positive and a negative word set. From this lexicon we obtain two scores coming from the addition of the positive words appering in a tweet and, likewise, from the addition of the negative words.

2. NRC Emotion Lexicon (Mohammad and Turney, 2013): This lexicon contains a set of words and a value (0 or 1) expressing whether a word is associated to a certain emotion such as anger, anticipation, disgust, fear, joy, sadness, surprise and sad.

Twitter Features.The way of expressing ideas in Twitter as in other social networks differs from the language used in formal writing. That is why we should capture the peculiarities about this language that could be useful for identifying the polarity of a tweet in certain situations.

- **Elongated Words** We count the number of elongated words. For instance, ”I love you sooooo much”.

- **ALL CAPS** We count the number of words in upper case.

- **#Hashtags**. We count the number of hashtags in a tweet.

Finally, a **tf-idf** normalization was applied in all the selected features.

2.3 Classification

In this work, we classified the tweets polarity using a SVM formalism. An implementation using regularized linear models with stochastic gradient descent (SGD) learning is provided by the scikit-learn toolkit (Pedregosa et al., 2011).

3 Experiments

In this section, we expose the experiments carried out. Every experiment applies the preprocessing explained in section 2.1. The dataset used to conduct the experimentation was the one adopted on SemEval-2013 task 2 subtask B (Nakov et al., 2013). Indeed, all the experimentation applies a linear SVM as a classifier. The following lines express the features implemented in the most successful experiments.

- **Experiment 1**
  - Unigrams and Bigrams
  - Jeffrey’s Lexicon.
- **Experiment 2**
  - 1-6 grams
  - Jeffrey’s Lexicon.
- **Experiment 3**
  - Unigrams
  - Jeffrey’s Lexicon.
  - Skip-grams
- **Experiment 4**
  - Unigrams
  - Jeffrey’s Lexicon.
  - 100-most frequent Skip-grams
- **Experiment 5**
  - Unigrams and Bigrams
  - Jeffrey’s and NRC Emotion Lexicons.
- **Experiment 6**
  - Unigrams and Bigrams
  - Jeffrey’s and NRC Emotion Lexicons.
  - All Twitter Features
- **Experiment 7**

Experiment	$F1_{pos}$	$F1_{neg}$	$(F1_{pos} + F1_{neg})$ / 2
1	0.6913	0.5593	0.6253
2	0.6383	0.5548	0.6180
3	0.6860	0.5458	0.6159
4	0.6851	0.5603	0.6227
5	**0.6973**	**0.5824**	**0.6399**
6	0.6197	0.3516	0.4857
7	0.5906	0.3262	0.4584
8	0.5807	0.3306	0.4556
9	0.6215	0.2840	0.4527

Table 1: Results. SemEval-2013 Dataset.

  - Unigrams and Bigrams
  - Jeffrey's and NRC Emotion Lexicons.
  - #Hashtags

- **Experiment 8**
  - Unigrams and Bigrams
  - Jeffrey's and NRC Emotion Lexicons.
  - ALLCAPS

- **Experiment 9**
  - Unigrams and Bigrams
  - Jeffrey's and NRC Emotion Lexicons.
  - Elongated Words

4 Results

This section summarizes the results of the tuning phase. As we can see in Table 1, the best approach is the one used in experiment 5 which uses only unigrams, bigrams and both lexicons. This fact shows the importance of unigrams and bigrams as well as the relevance of using lexicons which can improve considerably a message polarity classification model. Moreover, using n-grams larger than bigrams (6-grams in our experiments) can introduce noise in the model.

As we can see in Table 1, Twitter features decrease the performance of the classification. In experiment 6, we use all Twitter features together which leads us to a decreasing of $\mathbf{(F1_{pos}+F1_{neg})/2}$ from 0.6399 to 0.4857. Likewise, the results of experiments 7, 8 and 9 which use Twitter features individually show a diminution of similar magnitute in the evaluation measure $\mathbf{(F1_{pos}+F1_{neg})/2}$.

4.1 N-grams vs Skip-grams

In this work, we presented Skip-grams as an alternative to N-grams and we see that N-grams performed slightly better than Skip-grams. However, this difference in the performance is not statistically significant and can vary between different corpora. In addition, experiment 4 includes a variation taking only the one hundred most frequent Skip-grams. The comparison between experiment 3 and 4 shows that using the most frequent Skip-grams leads to better results than using all the Skip-grams generated.

4.2 Competition Results

For the competition, the model used in experiment 5 which outperformed the others in the tuning phase was submitted. This model consists of unigrams, bigrams and both lexicons (Jeffrey and NRC emotion lexicon). In the official rank our system achieved the 22nd out of 34 teams.

Acknowledgments

This work has been partially funded by the project ASLP-MULAN: Audio, Speech and Language Processing for Multimedia Analytics (Spanish MINECO TIN2014-54288-C4-3-R).

References

Mayte Giménez, Pla Ferran, and Lluís-F. Hurtado. 2015. Elirf: A support vector machine approach for sentiment analysis tasks in twitter at semeval-2015. In *In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 574—-581. Association for Computational Linguistics.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. Technical report, NRC Technical Report.

Preslav Nakov, Zornitsa Kozareva, Alan Ritter, Sara Rosenthal, Veselin Stoyanov, and Theresa Wilson. 2013. Semeval-2013 task 2: Sentiment analysis in twitter.

Preslav Nakov, Alan Ritter, Sara Rosenthal, Veselin Stoyanov, and Fabrizio Sebastiani. 2016. SemEval-2016 task 4: Sentiment analysis in Twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, SemEval '16, San Diego, California, June. Association for Computational Linguistics.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830.

Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.