

LIPN-UAM at EmoInt-2017: Combination of Lexicon-based features and Sentence-level Vector Representations for Emotion Intensity Determination

Davide Buscaldi

LIPN, Université Paris 13,
Villetaneuse,
France

buscaldi@lipn.univ-paris13.fr

Belem Priego

UAM
Atzacapotzalco,
Mexico

belemps@gmail.com

Abstract

This paper presents the combined LIPN-UAM participation in the WASSA 2017 Shared Task on Emotion Intensity. In particular, the paper provides some highlights on the system that was presented to the shared task, partly based on the Tweetaneuse system used to participate in a French Sentiment Analysis task (DEFT2017). We combined lexicon-based features with sentence-level vector representations to obtain a random forest model.

1 Introduction

Nowadays, an important quantity of the textual information that is produced everyday on the Web originates from social media and commercial sites with crowd-sourced reviews. These data include beliefs, opinions and judgments, expressed in various forms, sometimes resorting to the use of figurative language, such as irony, which makes an automated analysis of these texts even more difficult. Therefore, there is an increased interest by academia and industry towards the field of Sentiment Analysis (SA). This research activity has been mainly focused to extract and characterize opinions by recognizing the attitude (positive, negative or objective) of an opinion holder on a certain topic, or determine the global polarity of a given text.

A more recent and emerging field consists of studying the opinions in a more detailed way, revealing the underlying emotions, such as anger, fear, joy and disgust. One of the pioneer works in this sense is the one by (Strapparava and Mihalcea, 2008), in which they proposed for the first time a dataset dedicated to emotion analysis and some knowledge and corpus-based approach.

Their proposal included texts annotated with six emotions: anger, disgust, fear, joy, sadness and surprise. More recently, (Cambria et al., 2014) proposed Sentic.net¹, a resource for concept-level sentiment analysis, containing word senses annotated with weighted emotions.

The Shared Task proposed at WASSA2017 (Mohammad and Bravo-Marquez, 2017) aims to steer research about sentiments and emotions in text towards the intensity of the expressed emotions, and not only on binary polarity values or assigning an emotion to the texts. This paper describes the system submitted to the WASSA 2017 shared task by the joint LIPN-UAM team, in part based on the “Tweetaneuse” system that participated to the French Sentiment Analysis task DEFT 2017 (Benamara et al., 2017). The rest of the paper is structured as follows: in Section 2 we describe the features used and the machine learning approach; in Section 3 we show the results obtained on the official data together with some experiments to verify the effectiveness of the proposed features. Finally, in Section 4 we draw some conclusion about our participation.

2 System Description

The system that we built for our participation in the Shared Task at WASSA2017 is based on a set of 8 features derived from lexicons and various textual clues, and 600 features derived from word embeddings. These features are used to train a random forest regressor. These features are inspired by those previously used for our participation in the French sentiment analysis task at DEFT2017. The basic textual clues were the following ones:

- *smi*: presence of a smiley;

¹<http://sentic.net/>

- *shout*: number of uppercase words (to detect the fact that the writer is shouting);
- *excl*: number of exclamation marks;
- *int*: number of interrogation marks.

We used 4 different lexicons: sentic.net (Cambria et al., 2014), labMT (Dodds et al., 2011), the NRC Affect Intensity lexicon (Mohammad, 2017), and the emojis sentiment ranking by (Novak et al., 2015). We already talked about sentic.net in Section 1. We limited the use of sentic.net to the polarity values since the shared task did not involve determining which emotion was contained in the sentence but only its intensity. LabMT is a lexicon obtained via Mechanical Turk that is currently used in the hedonometer.org project to measure average happiness in Twitter. We thought that this lexicon would be particularly useful for the joy and sadness categories. The emojis sentiment ranking is a lexicon obtained from a set of 1.6 million tweets manually annotated with their polarity strength, and is currently, to our knowledge, the only available resource providing the polarity and the intensity for emojis. The features extracted from these lexicons were the following ones:

- *pol*: average of sentic.net polarity values in the sentence;
- *happiness*: average of happiness values according to labMT;
- *nrc_ai*: average of scores from the NRC affect intensity lexicon (according to the emotion being tested);
- *emoji*: sum of scores from the emojis sentiment ranking.

The scores for all dictionaries have been modified to take into account the position where the score is detected. This modification reflects the idea that affective words towards the end of the sentence are more important than those at the beginning or the middle of the sentence. This is particularly true in the case of tweets where there may be affective hashtags at the end of the message, such as in the case “*All I want to do is watch some netflix but I am stuck here in class. #depressing*” (we normalized hashtags by removing the leading #). The formula used is the following one:

$$\hat{s}(w) = s(w) * (1 + 0.15 * rpos(w))$$

Where $rpos(w)$ is the relative position of word w within the sentence (i.e. $pos(w)/len(sentence)$) and $s(w)$ is the original score from the lexicons. The 0.15 weight was arbitrarily chosen.

These features are completed with sentence-level vector representations based on word embeddings. Word embeddings, as introduced by (Bengio et al., 2006), are vector representations of words that capture a certain number of syntactic and semantic relationships, generated with neural networks. One of the problems with word embeddings is how to compose them to obtain a representation of a sentence, knowing that sentences may have variable sizes. (De Boom et al., 2016) showed that it’s possible to exploit the properties of embeddings to represent sentences with the average or a combination of the max and the min (per dimension) of the vectors of the composing words. We chose to use the second method since it is the one that achieved the best results in their experiments.

In our work, we used the pre-trained vectors trained on 100 billion words from the Google News dataset used for word2vec (Mikolov et al., 2013). The vocabulary size is 3 million words and the vector length is 300. Therefore, in our system each sentence is represented by a vector of size 600.

The advantages of this representation are two: on one hand, it is more concise than the bag-of-words representation (600 dimensions while a typical BOW vector has thousands of components); on the other, it compensates for the words that are not observed in the training set (since the vocabulary size for embeddings is \gg than the vocabulary size for the task training corpora).

3 Results

The official results are listed in Table 1. The system ranked slightly below the baseline system, except on the ‘sadness’ test set, where our system was better. The results obtained for the emotion intensities in the range 0.5 – 1 (shown in Table 2) are also very close to the baseline system, with the exception of the results obtained on the ‘sadness’ test set. This evaluation scenario highlights some problems that our system had on the ‘joy’ dataset.

Test Set	Pearson	Spearman
anger	0.580	0.575
fear	0.639	0.630
joy	0.583	0.601
sadness	0.676	0.686
average	0.619	0.623
baseline	0.648	0.641

Table 1: Results obtained at the WASSA2017 Shared Task.

Test Set	Pearson	Spearman
anger	0.435	0.439
fear	0.496	0.463
joy	0.366	0.347
sadness	0.489	0.503
average	0.446	0.438
baseline	0.477	0.442

Table 2: Results obtained at the WASSA2017 Shared Task, for intensity values in the range [0.5, 1.0].

We already observed during the development phase that the system was quite ‘cautious’ in the output scores, providing scores in the range (0.3, 0.7), with some exceptions. We impute this behaviour to two factors: the scarcity of extreme examples in the training set, and the use of random forests. However, we tried to use a Support Vector Regressor but the results were significantly worse (from 5 to 10% less depending on the test set).

Table 3 shows the results we obtained with different configurations of the system, in particular using only vectors, or only dictionary and text-based features. This experiment highlights the fact that on the ‘joy’ dataset, lexicons and text clues alone were able to beat the vector representations. On the other hand, we can observe that when the vector representations worked, the system was able to perform well. This is difficult to explain, but we suspect it to be related to the data used to train the vectors. We expect newswire data to contain more details about negative events, such as wars, natural disasters or accidents, which contains more words related to fear and sadness. This bias may result in modelling negative words better than positive ones.

Finally, we carried out Correlation Feature Selection (CFS) to test which features were most related to the intensity values. The CFS showed that *nrc_ai* and *emoji* were among the best features for all datasets. Among the base features, the CFS indicates that *excl* was important for ‘joy’ and ‘anger’, while *shout* was one of the best features for ‘joy’ and ‘fear’.

Test Set	All features	Vectors	Lexicons+base
anger	0.580	0.547	0.388
fear	0.639	0.632	0.439
joy	0.583	0.524	0.555
sadness	0.676	0.651	0.615
average	0.619	0.588	0.499

Table 3: Pearson correlation obtained with different configurations of the system: *Vectors* - only the max/min of word embeddings are used; *Lexicons+base* - only text-based clues and lexicon-based features are used.

4 Conclusions

In this participation we combined the use of word embeddings with lexicon-based features and simple text clues. According to the low complexity of the system created, the obtained results were close to the baseline system. Further analysis of the results allowed us to detect a possible problem with the news corpus used to train the word embeddings: news language does not necessarily use emotions, and when it does, the emotions are often related to negative events such as wars, natural disasters, etc. We plan to carry out the experiments with a different set of pre-trained vectors, in particular those extracted from Twitter by (Godin et al., 2013).

Feature analysis indicates that the NRC affective intensity dictionary (Mohammad and Turney, 2013) and the Emojis dictionary by (Novak et al., 2015) were particularly useful. As a future work, we plan to add a classification layer to the system to detect whether the emotion expressed is extreme or not, in order to improve the results on the most polarizing messages. Finally, we would like to test the effectiveness of the positional weighting for lexicon scores.

Acknowledgements

This work is part of the program ‘‘Investissements d’Avenir’’ overseen by the French National Research Agency, ANR-10-LABX-0083 (Labex EFL).

References

Farah Benamara, Cyril Grouin, Jihen Karoui, Véronique Moriceau, and Isabelle Robba. 2017. Analyse d’opinion et langage figuratif dans des tweets : présentation et résultats du Défi Fouille de Textes DEFT2017 (In French). In *Actes de l’atelier DEFT2017 associé la conférence TALN*. Orleans, France. To appear.

- Yoshua Bengio, Holger Schwenk, Jean-Sébastien Senécal, Frédéric Morin, and Jean-Luc Gauvain. 2006. *Neural Probabilistic Language Models*, Springer Berlin Heidelberg, Berlin, Heidelberg, pages 137–186. https://doi.org/10.1007/3-540-33486-6_6.
- Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. 2014. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Twenty-eighth AAAI conference on artificial intelligence*.
- Cedric De Boom, Steven Van Canneyt, Thomas Demeester, and Bart Dhoedt. 2016. Representation learning for very short texts using weighted word embedding aggregation. *Pattern Recogn. Lett.* 80(C):150–156. <https://doi.org/10.1016/j.patrec.2016.06.012>.
- Peter Sheridan Dodds, Kameron Decker Harris, Isabel M Kloumann, Catherine A Bliss, and Christopher M Danforth. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one* 6(12):e26752.
- Frédéric Godin, Viktor Slavkovikj, Wesley De Neve, Benjamin Schrauwen, and Rik Van de Walle. 2013. Using topic models for twitter hashtag recommendation. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, pages 593–596.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pages 3111–3119.
- Saif M Mohammad. 2017. Word Affect Intensities. *arXiv preprint arXiv:1704.08798*.
- Saif M. Mohammad and Felipe Bravo-Marquez. 2017. WASSA-2017 shared task on emotion intensity. In *Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*. Copenhagen, Denmark.
- Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence* 29(3):436–465.
- Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of emojis. *PloS one* 10(12):e0144296.
- Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM Symposium on Applied Computing*. ACM, New York, NY, USA, SAC '08, pages 1556–1560. <https://doi.org/10.1145/1363686.1364052>.