

# ej-sa-2017 at SemEval-2017 Task 4: Experiments for Target oriented Sentiment Analysis in Twitter

Enkhzol Dovdon and José Saias

DI - ECT - Universidade de Évora

Rua Romão Ramalho, 59

7000-671 Évora, Portugal

d36506@alunos.uevora.pt, jsaias@uevora.pt

## Abstract

This paper describes the system we have used for participating in Subtasks A (Message Polarity Classification) and B (Topic-Based Message Polarity Classification according to a two-point scale) of SemEval-2017 Task 4 Sentiment Analysis in Twitter. We used several features with a sentiment lexicon and NLP techniques, Maximum Entropy as a classifier for our system.

## 1 Introduction

Text data has been growing dramatically. We have demands to process and mine from Social networks and online platforms. Opinions in user-generated content, are valuable for market and trend analysis. Processing of sentiment analysis helps us to automatically distinguish from these written opinions.

This paper describes a participation in SemEval-2017 Task 4 with the *ej-sa-2017* system. We have participated in SemEval-2017 Task 4 on Sentiment Analysis in Twitter, subtasks A (Message Polarity Classification), B (Topic-Based Message Polarity Classification) (Rosenthal et al., 2017). Subtask A is to classify message polarity from given a message that is of positive, negative, or neutral sentiment. Subtask B is to classify positive or negative sentiment of a tweet towards that topic on a two-point scale.

We utilized a supervised machine learning classifier, having bag-of-word (BoW), lemmas, bigrams of adjective, punctuation based features, and lexicon-based features. The rest of the paper is structured as follows: In Section 2, we present some related work in features and approaches with a lexicon. In Section 3, this section describes the algorithm and feature representation used to detect

sentiment of text. In Section 4, the experimental results are introduced. Finally, the conclusions as well as further work are described in Section 5.

## 2 Related Work

There are many works associated with the target-oriented sentiment analysis. Some of these works have focused on probability distribution model of particular features and approach. The system of *Sentiue* (Saias, 2015) used a separate MaxEnt classifier of MALLETT (MACHINE Learning for Language Toolkit) (McCallum, 2002) with bag-of-word like features (lemmas, bigram, presences, etc.) for Aspect based Sentiment Analysis in SemEval-2015 Task 12 and accuracy was approximately 79%. Kamps (Kamps et al., 2004) developed a simple distance measure, that focuses almost exclusively on taxonomic relations and WordNet and determined usage of the semantic orientation of adjectives. Pak (Pak and Paroubek, 2010) utilized the presence of n-grams, for  $n \in \{1, 2, 3\}$ , as a binary feature of a BoW representation using TreeTagger. They collected a corpus of 300000 text posts from Twitter. Fong (Fong et al., 2013) focused on news articles, which tend to use a more neutral vocabulary using MALLETT to implement and train six classifiers for sentiment analysis and compared them. Their experimental results show that the Naive Bayes classifier performs the best of six algorithms. Singh (Singh et al., 2013) have been implemented double Machine Learning based classifiers (Naive Bayes as a 2-class text classification problem and SVM with tf.idf vectors), the Unsupervised Semantic Orientation approach with POS tagging and the SentiWordNet approaches for sentiment classification of a huge amount of movie reviews. Their used priority scoring Adjective + Adverb combine scheme of SentiWordNet approach was performed

0.811 F1-score in their experiments.

### 3 Method

This section describes feature extraction and a classifier of the sentiment analysis for our system. We used the tool MALLET that supports a variety of supervised classifiers, which makes it ideal for the comparative study of our experiences. We developed the current system using several valuable ideas from previous work (Saias, 2015) for Target and Aspect based Sentiment Analysis.

#### 3.1 Feature extraction

We have performed standard data preprocessing steps on the system of tweets prior to classification. Text preprocessing consists of tokenization, removing all capitalization, stop word removal, POS tagging, and lemmatization with Stanford CoreNLP (Manning et al., 2014) and MALLET. An instance was created for each tweet text which includes extracted features. Some features are used additional lexicon resources such as SentiWordNet lexicon (Baccianella et al., 2010).

##### Subtask A (Message Polarity Classification).

The below features to represent each instance in Subtask A were:

- BoW with a feature for each token text;
- lemmas for nouns, verbs, adjectives and adverbs;
- a polarized term for each word;
- average polarized term for each instance;
- presence of negation terms.

The polarized terms based on SentiWordNet and used count of positive or negative polarity words using polarity scores. Some words appear more than once in this lexicon. For an example: "easy", this word is used in 28 different sentences on SentiWordNet. In other words, there are 28 use cases of the word and diverse polarity scores (positive or negative score). Thus, we have chosen an approximate use case of the word from the lexicon using BoW.

**Subtask B (Topic-Based Message Polarity Classification).** The below features to extract from each instance in Subtask B were:

- BoW with a feature for each token text after target position;
- lemmas for nouns, verbs, adjectives and adverbs with next to target position;
- polarized term for unigram and bigram words after given-target (topic) position in a text;
- presence of negation terms;
- presence of exclamation/question mark.

In this case, a polarized term was based on average polarity score which was created using all used cases of a word in SentiWordNet records. If any of an adjective appears next to target in a text, it will be chosen as the polarized term feature and set a tag as a positive, negative or neutral. Some features of an example tweet presented are:

Target: "denzel"; Tweet: "Gotta go see Flight tomorrow Denzel is the greatest actor ever!";  
Extracted features: (1) #AFTER.VBZ.positive for "is", (2) #AFTER.JJS.positive for "greatest", (3) #AFTER.NN.neutral for "actor" (4) #AFTER.RB.neutral for "ever" (5) #polExclMark.positive for "!".

After this step, each text document in the system will be represented by a feature vector using MALLET.

#### 3.2 Classifier training

The classifier algorithm was Maximum Entropy and the classifier model features were previously mentioned features. MaxEnt seeks the probability distribution model that best fits the features observed in the text. We have trained a classifier with instance list where each tweet text had been created as an instance with feature vectors using MALLET pipeline. A single label multi-class classification is used for the training in subtask A. Each tweet must be classified into exactly one of the following three classes (positive, neutral and negative). We also used a binary classification (positive or negative) for the training in subtask B. A single sentence in a tweet may have several sentiment polarities about different aspects. Thus, we tried to consider it in feature selection phase that has to choose correct sensitive words as a feature depends on a target.

### 4 Results

In this section, the results obtained with the proposed system and datasets are written. The prelim-

inary experiments, we performed for the system were carried out by training and testing our models on datasets generated in editions of previous years of the tasks (see Table 1 and 2). All tweets are annotated for polarity by the organizers. Unbalanced training corpus is used where there are more positive tweets than others.

Dataset	All	Pos.	Neg.	Neut.
twitter-2013	9684	3640	1458	4586
-train-A twitter-2013	1654	575	340	739
-dev-A twitter-2014	86	33	40	13
-sarcasm-A twitter-2015	489	170	66	253
-train-A twitter-2016	6000	3094	863	2043
-train-A twitter-2016	1999	843	391	765
-dev-A twitter-2016	2000	994	325	681
-devtest-A				
Total (no duplication)	21403	9171	3412	8820

Table 1: Trainset for our system in Task 4-A.

Dataset	All	Pos.	Neg.
twitter-2015	198	142	56
-train-BD twitter-2015	1127	867	260
-testBD twitter-2016	4346	3591	755
-train-BD.txt twitter-2016	1325	986	339
-dev-BD twitter-2016	1417	1153	264
-devtest-BD twitter-2016	10551	8212	2339
-test-BD			
Total	18964	14951	4013

Table 2: Trainset for our system in Task 4-B.

The classification results are presented in Table 3. In Subtask A, 37 submissions evaluated, the best F1-score value was 0.685, while our result F1-score was 0.539. There are 24 submissions in Subtask B, the best F1-score was 0.89 and our F1-score was 0.486.

## 5 Conclusions

We have presented an approach that incorporates the MaxEnt with various features to solve the over-

Subtask	F1	Recall	Acc
A	0.539	0.571	0.582
B	0.486	0.594	0.518

Table 3: Results achieved by our system

all polarity and topic-based message polarity. Our system is part of first author’s work on text classification, included in PhD ongoing work. From the results, we noticed that our system was unsatisfactory compared to other teams. However, this evaluation became a good experience for us. Many people usually use an entirely different language on social media sites such as Twitter and Facebook. Thus, we will focus on social media and informal language learning. As further work we propose the following:

- compare the classical approaches with common features
- investigate the usage of a combination of classical approaches
- explore different techniques that can be used in target-oriented sentiment analysis
- investigate efficient features and new feature
- use more lexicons such as AFINN (Nielsen, 2011) and NRC Emoticon (Mohammad and Turney, 2010)
- develop the possibility of the system on multilingual

## 6 Acknowledgments

We would like to thank the gLINK project of "Erasmus Mundus Programme, Action 2 - STRAND 1, Lot 5, Asia (East)". We would also like to thank the LabInterop project, for providing the infrastructure. LabInterop is funded by *Programa Operacional Regional do Alentejo* (INALLENTEJO).

## References

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*. volume 10, pages 2200–2204.
- Simon Fong, Yan Zhuang, Jinyan Li, and Richard Khoury. 2013. Sentiment analysis of online news

- using mallet. In *Computational and Business Intelligence (ISCBI), 2013 International Symposium on*. IEEE, pages 301–304.
- Jaap Kamps, Maarten Marx, Robert J Mokken, Maarten De Rijke, et al. 2004. Using wordnet to measure semantic orientations of adjectives. In *LREC*. Citeseer, volume 4, pages 1115–1118.
- Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*. pages 55–60.
- Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit .
- Saif M Mohammad and Peter D Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. Association for Computational Linguistics, pages 26–34.
- Finn Årup Nielsen. 2011. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903* .
- Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*. volume 10.
- Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, Vancouver, Canada, SemEval '17.
- José Saias. 2015. Sentiue: Target and aspect based sentiment analysis in semeval-2015 task 12. Association for Computational Linguistics.
- VK Singh, R Piryani, A Uddin, P Waila, et al. 2013. Sentiment analysis of textual reviews; evaluating machine learning, unsupervised and sentiwordnet approaches. In *Knowledge and Smart Technology (KST), 2013 5th International Conference on*. IEEE, pages 122–127.