TGB at SemEval-2016 Task 5: Multi-Lingual Constraint System for Aspect Based Sentiment Analysis

Fatih Samet Çetin
Istanbul Technical University
cetinfatih@itu.edu.tr

Ezgi Yıldırım
Istanbul Technical University
yildirimez@itu.edu.tr

Can Özbey
Turkcell Global Bilgi
canozbeycan@gmail.com

Gülşen Eryiğit
Istanbul Technical University
gulsen.cebiroglu@itu.edu.tr

Abstract

This paper gives the description of the TGB system submitted to the Aspect Based Sentiment Analysis Task of SemEval-2016 (Task 5). The system is built on linear binary classifiers for aspect category classification (Slot 1), on lexicon-based detection for opinion target expressions extraction (Slot 2), and on linear multi-class classifiers for sentiment polarity detection (Slot 3). We conducted several different approaches for feature selection to improve classification performance on both Slot 1 and Slot 3. Our proposed methods are easily adaptable to all languages and domains since they are built as constrained systems which do not use any additional resources other than the provided datasets and which uses standard preprocessing methods.

1 Introduction

Since Web 2.0 and social media platforms have become popular in recent times, the amount of accessible text data has shown rapid increase. The manual analysis of this huge amount of data is almost impossible to accomplish in a reasonable time, thus automatic sentiment analysis and opinion mining have turned into a significant requirement for companies. The most of earlier studies conducted in this area were generally focused on document level

(Yıldırım, et al., 2015; Pang, et al., 2002; Esuli & Sebastiani, 2006) until the recognition of that different opinions can exist in the same sentence or paragraphs. Another disadvantage of general sentiment analysis approaches is the disability to match the sentiment polarities to the target entities. Therefore, this type of analysis becomes insufficient for deep understanding of opinions about products and features. The most commonly referenced study (Liu, 2012) on Aspect-Based Sentiment Analysis (ABSA) discusses the problem as extracting the tuples including multiple opinions.. The need for a detailed sentiment analysis with respect to specific target entities has given birth to ABSA. In International Workshop on Semantic Evaluation (SemEval), a shared task called Aspect-based Sentiment Analysis has been actualized since 2014 (Pontiki, et al., 2014; Pontiki, et al., 2015; Pontiki, et al., 2016). In this year's task, the data is annotated at both sentential and textual levels with reference to predefined domain-dependent aspect categories. For more information and details about the aspect categories consult (Pontiki et al., 2016). The task description (SemEval, 2016) provides regulations as to how these categories should be determined.

This paper presents our system prepared for ABSA 2016. It covers 3 different languages, namely, English, Spanish and Dutch. The system has multi-lingual capabilities giving nearly con-

sistent performances for each language. Our proposed methods are easily adaptable to all languages due to their multi-lingual nature by switching language codes for stemming phase. Considering the individual tasks, we have applied specific methods for each characteristic problem. In order to accomplish the task, we used different approaches for different slots. In order to find aspect category (Slot 1), we used a multi classifier approach which uses textual and probabilistic features. A lexicon based approach is chosen to extract the opinion target expressions (OTE) (Slot 2). We used a linear classifier which utilizes aspect category, aspect attribute and OTE features as well as textual features to detect the sentiment class for an opinion tuple (Slot 3).

2 System Description

In this section we present our aspect based sentiment analysis system. The system is experimented on the Restaurant datasets. Our submission is composed of the experiments on three languages; English, Dutch and Spanish. English dataset consists of 300 reviews and 2000 sentences, Spanish dataset has 627 reviews and 2070 sentences, and Dutch dataset contains 300 reviews and 1722 sentences.

The following preprocessing steps are used in all datasets:

- Removing html codes/URLs
- Tokenization
- Stemming

Sentences are tokenized and analyzed with Apache Lucene (Foundation, 2016) Analyzers which contains different types of operations. These operations are tokenization, filtering and transforming. In this system, we have used Standard Lucene Tokenizer for tokenization. Afterwards, we applied lowercase transformation on top of the previous step. In the following stage stemming is applied to all tokens using Snowball stemmer (Porter, 2001). An applied version of Snowball stemmer is already presented in the Lucene project. Finally, Lucene Shingle Filter is used to extract unigram and bigram features. In all of the introduced methods, we used these unigrams and bigrams (of the word stems in the datasets) as textual features,.

For the classification task, we use logistic regression from LIBLINEAR (Fan, 2008-9) classification library. LIBLINEAR is an open source library for large-scale linear classification (Fan, 2008-9). It provides easy-to-use command-line tools and library calls for users and developers.

We have implemented a generic framework to make text classification on LIBLINEAR library. Our framework basically provides an infrastructure to developers for building custom preprocessing steps and classification systems. Developers only need to be aware of framework's interfaces and Lucene Analyzers.

2.1 Aspect Category Classification (Slot 1)

In order to detect aspect categories, we used a two-layered approach. First layer consists of one-vs-all binary classifiers for the detection of each different aspect entity (E) and aspect attribute (A). This first layer is used to obtain the possibilities of an instance to belong to the corresponding classes (entity or attribute). The obtained probabilities will be further used in the second layer for the ultimate classification.

In the first layer, according to the instances available in different training sets (for each different language), at most 11 distinct binary classifiers are independently trained. For instance representation, we use unigram and bigram features occurring in the training sets.

Features	Instance
Textual Features 1	1
⋮	⋮
Textual Features n	0
FOOD	0,45
DRINKS	0,33
SERVICE	0,41
AMBIENCE	0,65
LOCATION	0,17
RESTAURANT	0,75
GENERAL	0,5
PRICES	0,33
QUALITY	0,65
STYLE&OPTIONS	0,67
MISCELLANEOUS	0,85
CLASS	AMBIENCE#GENERAL

Table 1: Feature representation of ultimate classifier

In the second layer, we construct an ultimate classifier which uses additional features extracted

from the first layer's output. These are entity and attribute labels used as real-valued features (as opposed to the binary unigram and bigram textual features). This feature representation is depicted in Table 1. The probabilities from the previous layer are assigned as values for these features. The ultimate classifier is trained in order to produce aspect categories (E#A) that are composed of both entity and attribute label. Architectural model of the proposed method can be seen in Figure 1.
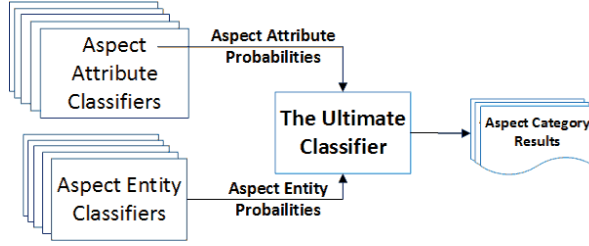

Figure 1: Slot 1 System Architecture

2.2 Opinion Target Extraction (Slot 2)

In opinion target extraction task, we basically follow the baseline opinion target extraction (OTE) procedure proposed in SemEval ABSA 2016 shared task with some changes and assumptions over it.

First, we extracted all opinion targets of each E#A pair from the training set, then used them to check if an opinion target is matched when a new sentence is examined during testing. If an opinion target is detected and the E#A pair of sentence is identical with the E#A pair of stored opinion target, this opinion target is assigned to the sentence. While we are extracting opinion targets from the training data, we applied different preprocessing steps composed of only lowercasing and asciification. However, in some cases many different opinion targets are found to be related to the same E#A pair. In these situations, we applied different approaches to choose the most suitable opinion target from all possible candidates.

Our first assumption is that if an opinion target already contains another one in itself (as a sub-unit), the longer one should be the specialized version of the same opinion target and assumed to be the most proper indication of the relevant opinion target. For instance, "*traditional dishes*" is better than only "*dishes*" while it contains the subsequent.

Even if we aimed to singularize extracted opinion targets with the first assumption, it is still possible to end up with several opinion candidates after this reduction method. Therefore, we applied some secondary approaches to singularize them. First, we experimented with utilizing the frequencies of opinion targets for each E#A pair. For instance, if "pasta" and "spaghetti" are detected as possible opinion targets of a sentence, and "pasta" is more frequent than "spaghetti" for "FOOD#PRICE" pair, "pasta" is preferred in this scenario. Second, we examined the selection of the longest opinion target from all candidates for the same E#A pair because of the similar assumption to the first one: the longer, the better. well decorated and lighted place" is a better choice than "charming place".

After we evaluated the success of our assumptions, we found the best results by selecting the first inclusive opinion targets, and later the longer ones if still necessary.

2.3 Slot 1&2

To generate the predictions for Slot 1&2, we combine Slot 1 predictions with that of Slot 2. In this phase, instead of using gold E#A pair annotations to match them to opinion target expressions as we did in the Slot 2 prediction phase, we directly utilize the predictions of Slot 1. If any of the predictions of Slot 1 cannot be linked to any opinion targets, we assigned a NULL expression to these opinion targets.

2.4 Polarity Detection (Slot 3)

For sentiment polarity detection, gold standard aspect category and aspect terms are provided in the training sets. Therefore, the problem is to find and fill the sentiment polarity of each opinion tuple. We use a single classifier approach for solving the problem. Our classifier is able to determine the polarity class of the tuple from three types of polarity. Polarity classes are "positive", "negative" and "neutral".

We use textual as well as the features obtained from the previous aspect category detection phase. In other words aspect entity, aspect attribute and aspect terms are also placed in the feature representation of the train set. Each of the aspect entity, aspect attributes and aspect terms are represented with binary features. Therefore, 6 additional features for aspect entity, 5 additional features for aspect attribute and additional OTE features of the number of

unique opinion targets in dataset are used in our feature representation. While producing feature value set for an instance, values are assigned according to the feature existence in the instance. The mentioned features can be shown in Table 2. After the feature representation, a single classifier is trained using the Logistic Regression algorithm.

Features	Instance
Textual Features 1	1
⋮	⋮
Textual Features n	0
ENTITY-FOOD	1
ENTITY-DRINKS	0
ENTITY-SERVICE	0
ENTITY-AMBIENCE	0
ENTITY-LOCATION	0
ENTITY-RESTAURANT	0
ATTRIBUTE-GENERAL	0
ATTRIBUTE-PRICES	0
ATTRIBUTE-QUALITY	1
ATTRIBUTE-STYLE&OPTIONS	0
ATTRIBUTE-MISCELLANEOUS	0
TERM-expensive	0
TERM-wine list	0
TERM-restaurant	0
TERM-delicious	1
⋮	⋮
TERM-price	0
CLASS	positive

Table 2: Slot 3 feature representation

3   Results

We submitted our constrained system predictions to all slots of Subtask 1 for the restaurant domain. Since we use a common method for each language, we expected to have similar results for each language. According to the results of the task's evaluation, this expectation seems to be satisfied. Except for the English dataset, we highly ranked for all slots and languages that we've applied. In Dutch dataset, our Slot 1, Slot 1&2 and Slot 3 results are the best one and we obtained the second rank for Slot 2.

However, we would like to emphasize that the best result for Slot 2 is obtained by an unconstrained system. So, we may say that our system provides successful results for constrained systems. We attained similar successful results also in the Spanish dataset especially among constrained systems. In the English dataset we obtained fourth best score among thirty constrained system for Slot 3 and third best score for Slot 1&2.

TEAM	C/U	F-1
TGB	C	60.153
INSIG.	C	56
IIT-T.	U	55.247
IIT-T.	C	54.98
UFAL	U	53.876
baseline	C	42.816

Table 3: Dutch dataset- Slot 1 results table

TEAM	C/U	F-1
TGB	C	77.814
IIT-T.	U	76.998
INSIG.	C	75.041
basel.	C	69.331

Table 4: Dutch dataset – Slot 3 results table

TEAM	C/U	F-1
IIT-T.	U	83.582
TGB	C	82.09
UWB	C	81.343
INSIG.	C	79.571
basel.	C	77.799

Table 5: Spanish dataset – Slot 3 results table

TEAM	C/U	F-1
NLANGP	U	52.607
XRCE	C	48.891
NLANGP	C	45.724
TGB	C	43.081
bunji	U	41.113
UWB	C	41.108
UWB	U	41.088
⋮	⋮	⋮

Table 6: English dataset - Slot1&2 results table

4 Conclusion

In this paper, we show our experiments and approaches on aspect category classification and sentiment polarity detection using supervised machine learning methods and opinion target extraction using lexicon based approaches. For these problems, we used only the supplied resources by the shared task committee. We built well-performed multi-lingual systems which do not require additional resources and special manipulations for different languages and domains except only stemmers for the preprocessing stage. By the successful data preparation and feature extraction methods, our system gives reasonable results on aspect category classification and sentiment polarity prediction of sentences and detection of related opinion expressions if in question.

Acknowledgments

References

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of LREC*, (pp. 417-422).

Fan, R. E. (2008-9). LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, 1871-1874.

Foundation, A. S. (2016). *Lucene*. Retrieved from Apache Lucene: https://lucene.apache.org/

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 1-167.

Pang, B., Lee, L., & Vaithyanathan., S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistic.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Zafra, S., Eryiğit, G. (2016). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. *In Proceedings of the 10th International Workshop on Semantic Evaluation.* San Diego: Association for Computational Linguistics.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. *In Proceedings of the 9th International Workshop on Semantic Evaluation* (pp. 486-495). Association for Computational Linguistics.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. *international workshop on semantic evaluation (SemEval 2014)*, (pp. 27-35).

Porter, M. F. (2001). Snowball: A language for stemming algorithms.

SemEval. (2016). *SemEval-2016 Task 5*. Retrieved from SemEval: http://alt.qcri.org/semeval2016/task5/

Yıldırım, E., Çetin, F. S., Eryiğit, G., & Temel, T. (2015). The impact of NLP on Turkish sentiment analysis. *TÜRKİYE BİLİŞİM VAKFI BİLGİSAYAR BİLİMLERİ ve MÜHENDİSLİĞİ DERGİSİ*.