Ibrk at the NTCIR-14 QA Lab-PoliInfo Classification Task

Minoru Sasaki[1] and Tetsuya Nogami[1]

Dept. of Computer and Information Sciences
Faculty of Engineering, Ibaraki University
{minoru.sasaki.01, 15t4051t}@vc.ibaraki.ac.jp

Abstract. Stance classification has been defined as automatically identifing speaker's positions about a specific discussion topic from text. Although stance classification has been active research area, there is no approach that uses external knowledge to improve the classification. In this paper, we propose stance classification system using sentiment dictionary. To evaluate the efficiency of the proposed system, we conduct some experiments to compare with the result of the baseline method using Support Vector Machine on the NTCIR-14 QA Lab-PoliInfo classification task formal run dataset. The results showed that the proposed methods using sentiment dictionary obtains higher precision compared with the baseline method using SVM for the "support" and "against" samples. However, the precision of the proposed method is decreased about 10% in comparison to the baseline system for the "neutral" samples.

Team Name. Ibrk

Subtasks. Classification Task (Japanese)

Keywords: stance classification · sentiment dictionary · topic relation · fact-checking

1 Introduction

Stance classification has been defined as automatically identifying speaker's positions about a specific discussion topic from text. The speaker's position could be one of the labels: "support" ("agree", "support", "pro", "favor", "for"), "against" ("disagree", "oppose", "con", "anti") and "neutral" ("none", "unrelated", "neither") [1] [3] [4]. For example, one can infer from Barack Obama's speeches that he is in favor of stricter gun laws in the US. Similarly, people often express stance towards various target entities through posts on online forums, blogs, Twitter, Youtube, Instagram, etc. Recently, stance classification has been active research area in opinion mining. Stance classification is the task to classifying whether a given document is "support" or "against". Recently, some researches have demonstrated some approaches to solve stance detection problem. Some researches focused on the analysis of the political polarization

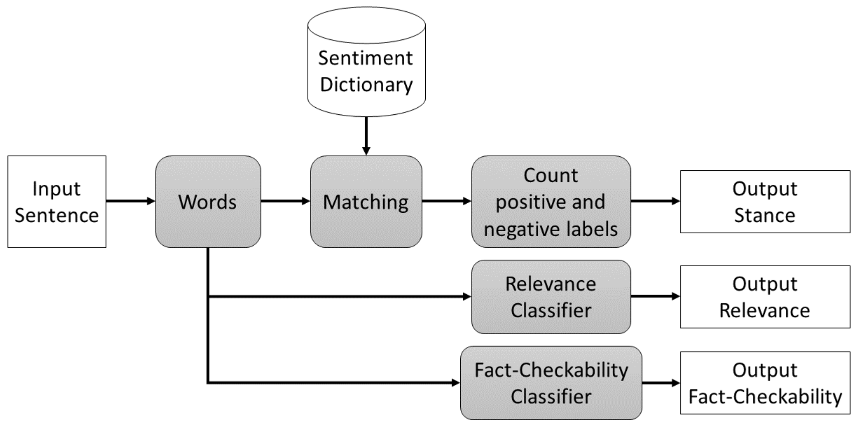

Fig. 1. The Proposed Stance Classification System

in Twitter [2] [6] and on detecting stance of debates by using semi-supervised learning in online forum [5].

In this paper, we propose stance classification system using sentiment dictionary. In this system, if the word exists in the sentiment dictionary for each word in the input sentence, then the polarity of the word is extracted to identify sentiment polarity label (positive or negative). The system counts up the number of positive and negative labels in the sentence. If the number of positive labels is greater than the number of negative labels, the system assigns "support" label to the sentence, otherwise the system assigns "against" label. To evaluate the efficiency of the proposed system, we conduct some experiments to compare with the result of the baseline method using Support Vector Machine (SVM).

2 Stance Classification System Using Sentiment Dictionary

2.1 System Description

The overall architecture of the proposed system can be seen in Figure 1. First, the input text is segmented into sentences and is represented as a sequence of sentences. Each sentence is parsed using a morphological analyzer and is divided into words. For each word in the extracted words, if the word exists in the sentiment dictionary, then the polarity of the word is extracted to identify sentiment polarity label (positive or negative).

Next, the system counts up the number of positive and negative labels in the sentence. If the number of positive labels is greater than the number of negative labels, the system assigns "support" label to the sentence, otherwise the system assigns "against" label. Finally, if the number of positive labels is equal to the number of negative labels or there is no positive/negative label, the sentence is considered to be "neutral" label.

To estimate the relevance of the topic, we use SVM to classify whether the sentence is related to a given topic. At the first step, we extract a set of features (nouns, verbs and adjectives) from the input sentence in the training data using a morphological analyzer. Then, each feature set is represented as a feature vector by calculating frequencies of the features. A classifier is constructed by SVM from labeled feature vectors. The classifier is used to predict whether a new input sentence is relevant to the topic or not. For predictive for a fact checkability of the input sentence, as well as to predict the relevance of the topic, we use SVM for classifying whether the sentence is fact checkable or not.

Table 1. Experimental results for the topic of "Integrated Resort"

System	Precision (Support)	Precision (Against)	Precision (Neutral)
Our System	7.19%	15.63%	92.10%
Baseline System	0.00%	0.00%	90.73%

Table 2. Overall precision, recall and F-measure (micro average) for the topic of "Integrated Resort"

System	Precision (All)	Recall (All)	F-measure(All)
Our System	77.80%	77.80%	77.80%
Baseline System	90.70%	90.70%	90.70%

Table 3. Experimental results on "relevance"(Rel.) classification

System	Recall (Rel.)	Precision (Rel.)	Recall (Not Rel.)	Precision (Not Rel.)
Our System	100.0%	86.5%	0.0%	Nan

3 Experimental Results

Table 1 and table 2 show the results of the experiments of applying the proposed methods in the previous section for the topic of "Integrated Resort". Note that we display missing values as "NaN" in this table when a "support" ("against") label does not exist in the test data.

As can be seen from these results, the proposed method possesses higher precision compared with the baseline method using SVM for the "support" and

Table 4. Experimental results on "fact-checkablity"(FC) classification

System	Recall (FC)	Precision (FC)	Recall (Not FC)	Precision (Not FC)
Our System	0.0%	Nan	100.0%	64.6%

Table 5. Experimental results on "fact-check-support"

System	Recall	Precision	F-measure
Our System	17.8%	6.3%	9.3%

"against" samples. When using SVM, even though this system obtains higher precision, all samples in the test data are classified as "neutral". Because there are a lot of neutral samples in the training data, all samples are classified as neutral. The proposed system obtained higher precision than the baseline system for the "support" and "against" samples. However, the precision is decreased for the "neutral" samples. F-measure score of the proposed system is also decreased about 10% in comparison to the baseline system. Therefore, we will improve our system to classify "neutral" samples effectively, while classifying the "support" and "against" examples with high precision in the future.

Table 3 and table 4 show the results of the experiments for the relevance of the topic and the fact-checkablility of the speech respectively. For the relevance label, all statements in the test data were classified as relevant to the topic. For the fact-checkability label, all statements in the test data were classified as not fact checkable. In this experiment, we construct the each classifiers using the SVM from bag-of-words representations of sentences only. However, it was not possible to detect sentences that are not related to the topic and sentences that we can conduct a fact-check. In the future, we will analyze training data in detail and will improve our system to correctly classify sentences that are misclassified.

Table 5, table 6 and table 7 show the results of the experiments for the class label (fact-check-support, fact-check-against and class-other respectively). As you can see from these tables, the proposed method obtains F-measure of 9.3% for the fact-check-support and F-measure of 7.4% for the fact-check-support. By using the proposed method, the small number of test data can be classified into "fact-check-support" and "fact-check-against" classes correctly. However, the proposed method obtains a low F-measure score.Therefore, we will improve our system to classify "class-other" samples effectively, while classifying the "fact-check-support" and "fact-check-against" examples with a high score in the future.

4 Conclusions

In this paper, we proposed a new method for stance classification using sentiment dictionary. The efficiency of the proposed method was evaluated on the NTCIR-14 QA Lab-PoliInfo classification task formal run dataset. The results showed that the proposed methods using sentiment dictionary obtains higher precision

Table 6. Experimental results on "fact-check-against"

System	Recall	Precision	F-measure
Our System	20.2%	4.5%	7.4%

Table 7. Experimental results on "class-other"

System	Recall	Precision	F-measure
Our System	77.0%	93.4%	84.4%

compared with the baseline method using SVM for the "support" and "against" samples. However, the precision of the proposed method is decreased about 10% in comparison to the baseline system for the "neutral" samples.

References

1. Anand, P., Walker, M., Abbott, R., Tree, J.E.F., Bowmani, R., Minor, M.: Cats rule and dogs drool!: Classifying stance in online debate. In: Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. pp. 1–9. WASSA '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011), http://dl.acm.org/citation.cfm?id=2107653.2107654
2. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., Menczer, F.: Political polarization on twitter. In: Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM) (2011), http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2847
3. Hasan, K.S., Ng, V.: Stance classification of ideological debates: Data, models, features, and constraints. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing. pp. 1348–1356. Asian Federation of Natural Language Processing (2013), http://aclweb.org/anthology/I13-1191
4. Kimura, Y., Shibuki, H., Ototake, H., Uchida, Y., Takamaru, K., Sakamoto, K., Ishioroshi, M., Mitamura, T., Kando, N., Mori, T., Yuasa, H., Sekine, S., Inui, K.: Overview of the ntcir-14 qa lab-poliinfo task. In: Proceedings of the 14th NTCIR Conference (2019)
5. Rajadesingan, A., Liu, H.: Identifying users with opposing opinions in twitter debates. In: Kennedy, W.G., Agarwal, N., Yang, S.J. (eds.) Social Computing, Behavioral-Cultural Modeling and Prediction. pp. 153–160. Springer International Publishing, Cham (2014)
6. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. pp. 178–185 (2010)