

Opinion Mining on Social Media

Sun-A Kang, Sang Hyun Choi

Abstract— In these days, there are so much conflicts in a society and we have difficulties in decision-making through continuous conflicts. In order to quickly deal with social conflict, the study is necessarily to identify public opinion on problem, conflict faced in common, and the interests of the conflict parties. This study will propose an algorithm that allows you to quickly grasp the positive/negative opinions of a particular conflict. The subject of society conflict is free school meals in Korean elementary schools. The data have been collected from Naver articles comment-keyword “free school meals”. After data collecting, Naver articles comment separates positive and negative. We are construct sentiment lexicon and develop algorithm that can automatically classify.

Keywords—Opnion Mining, Sentiment Analysis, Free Meal, Online Comment.

I. INTRODUCTION

SINCE democratization has radically evolved in South Korea from the 1980s, society has been diversified. The various types of conflict has occurred due to the conflict of another philosophy, value, interest. Recently, the large-scale public works and policies is promoted and repaired by the government. So the public conflict occurred frequently. The public conflict lead to a waste of social costs such as the occurrence of costs associated with the delay of the time in the process to resolve the conflict [1].

Therefore, it is important to understand positive/negative decision of the conflict for quick decision making about social conflict. Research to study positive/negative of social conflict have been progress through qualitative research such as most of the questions. However the communication made in conventional face to face has changed in social communication because we occur the change of how and where to communication due to coming ‘big data era’ [2]. It means that the positive/negative grasp of social issue is possible through the social communication. Moreover, the social network service, internet media will be able to send and receive data in real time. So it is useful to quickly understand the specific social problems. The data using this study is news comment contained keyword “free meals” from October 24, 2010 to January 10, 2012. In this paper, we suggest algorithm that can be classified which positive or negative about new comment.

II. RELATED STUDY

Opinion mining is the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards products, services, issues [3]. And opinion mining purpose to decide and use opinion of positive/negative in text data [4]. Recently, opinion mining has actively promoted in a variety of fields such as politics, business, culture and medical. Sentiment vocabulary is very different in each field. Therefore we can use sentiment dictionary specializing in topic and field [5].

Beomil kang at el.(2013) has been described differences of topic composition according to liberal leanings and conservative leanings news article about 18 presidential election using topic modeling algorithm. As a result, this paper can confirm that the news article was reported according to tendency by their ideology. Also, this paper can find that topic based opinion mining is able to performance about valuable function of decision analysis.

III. RAESEARCH FRAMEWORK

In this study, article comments about free meals are analyzed by sensibility analysis. Automatic classification algorithm of positive opinion and negative opinion about social conflict can be suggested by sensibility analysis. In order to grasp the major factor of social conflict, the comments are classified and based on the policy.

For the sensibility analysis of comment, the process of this study is shown below. First, data preprocessing is used to process the collected data. Second, the KoNLP of R package is used to process the morphological analysis for extracting the vocabulary. The constructing of morphological analysis and sentiment dictionary are the important component parts of automatic classification algorithm. This process is necessary to select the word. Third, presented algorithm is explained.

A. Data description

In this paper, data is collected in the Naver news ranking-main comments for getting the reliable data. The search method is that uses the keyword of “Free Meals” to search news comment data in the Naver news ranking-main comments. The comments data is collected from October 24th 2010 to January 10th 2012 which data is about the controversies of middle and primary school free meals in Seoul. The search result was that 2004 comments data was collected in Naver news ranking-main comments. The data was deleted which data was the positive opinion and negative opinion were not clearly divided or neutral comment. After cleaning the data, there was 767 comments data was used.

Sun-A. Kang was with the Department of Management Information Systems, Chungbuk National University, 1 Chungdaero Seowongu Cheongju, Chungbuk, Korea (e-mail: hdes3822@naver.com).

Sang Hyun. Choi is with the Department of Management Information Systems, Chungbuk National University, 1 Chungdaero Seowongu Cheongju, Chungbuk, Korea (corresponding author’s phone: 82-43-2613742; fax: 82-43-2732355; e-mail: chois@cbnu.ac.kr).

B. The sentiment vocabulary extraction

Sentiment Word is an important reason that positive and negative of online comment opinion is classified. In this study, there were 180 positive comments and negative comments to be extracted at random and processed morphological analysis which comments had tagging for classifying sentiment word. After extracting, the regular words were cleared up.

- These words are removed which the frequency of word is appeared less than 2.
- These words are removed which word has difficult vocabulary in English and Chinese.
- Conjunction and internet terms are removed.

C. The sentiment vocabulary extraction

In this section, sentiment dictionary was constituted by the executed sentiment word in 3.2 section. This sentiment dictionary is used to distinguish the Positive-Negative of sentence. There were 767 Positive-Negative comment was applied sentiment classification algorithm. The amount of Positive-Negative words was searched. If the amount of Positive words is more than the amount of Negative words, the words about Tweets are distinguished as Positive. And if the amount of Positive words is less than the amount of Negative words or the amount is 0, the words about Tweets are distinguished as Negative.

IV. RESULT

The presented research method was deduced and made in the section of 3. For the Positive-Negative opinion of Naver news ranking-main comments was distinguished in the subject of ‘Free Meals’. First, Positive-Negative dictionary was created by the 961 Positive words and 1197 Negative words about social issues of free meals comments. The result was got by sentiment dictionary. And the result of Positive-Negative opinion likes the Fig. 1. Precision is 70.29% and 81.17% performance was found out in F1-score.

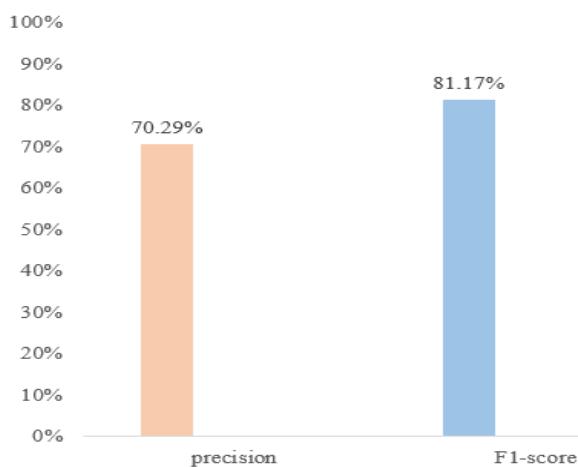


Fig. 1. Precision, F1-score of experiment

V.CONCLUSION

This paper aim to suggest positive/negative decision about social issue classification algorithm. We have performed sentiment dictionary construction after collecting news comment about social issue and building the sentiment dictionary. For analysis first we progress natural language. Second we build sentiment dictionary against 961 pieces of positive vocabulary and 1197 pieces of negative vocabulary. Finally, we carry out sentiment analysis using 767 news comments. As a result, this paper shows a performances of 70.29% accuracy and 81.17% F1-score.

This paper had been restricted social issue -“free meals”. And we don’t proceed qualitative research about contents analysis. So we will study contents analysis for qualitative research.

ACKNOWLEDGMENT

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2016-H8501-16-1013) supervised by the IITP(Institute for Information & communication Technology Promotion).

REFERENCES

- [1] D. J. Lim, “A Study on the Current Conditions of Public Conflict Management and the Factors of Conflict Resolution,” *KOREAN REPUBLIC ADMINISTRATION REVIEW*, Vol 45, No. 2, pp. 291-318, 2011.
- [2] H. J. Yun, B. H. Park, “Social TV : The Effects of Source Credibility and Repetition Type on Viewing Intention,” *Korean Society For Journalism And Communication Studies*, Vol 57, No. 1, pp. 364-391, 2013.
- [3] B. Liu, “Sentiment analysis and opinion mining,” *Synthesis Lectures on Human Language Technologies*, Vol 5, No.1, pp. 1-167, 2012.
- [4] G. E. Jang, S. H. Park, and W. J. Kim, “Automatic construction of a negative/positive corpus and emotional classification using the internet emotional sign,” *Journal of KIIS*, Vol 42, No. 4, pp. 512-521, 2015.
- [5] S. Owsley, S. Sood, and K. J. Hammond, “Domain Specific Affective Classification of Documents,” *In AAAI Spring Symposium: Computational Approaches to Analyzing*, 2006.
- [6] B. I. Kang, M. Song, and W. S. Jho, “A Study on Opinion Mining of Newspaper Texts based on Topic Modeling,” *Journal of the Korean Society for Library and Information Science*, Vol 47, No. 4, pp. 315-334, 2013.