Journal of Engineering and Applied Sciences 12 (10): 2725-2730, 2017
ISSN: 1816-949X

Opinion Mining Analyses by Online Media on the Introduction of Big Data-Based Free Semester System

Ji-Hoon Seo and Jin-Tak Choi
Department of Computer Science and Engineering,
Incheon National University, Academy-ro, 119 406-772 Yeonsu-gu, Korea

Abstract: This study collected informal data of news, blogs, twitter, social media and communities corresponding to online media and conducted Opinion Mining analysis on the free semester system currently implemented in South Korea. Past large volume big data from the time of the introduction of the free semester system to now were collected and preprocessed, the reputations of clients were analyzed through data classification, filtering and the construction of the emotion dictionary in Korean grammar and opinion mining analysis and utilization techniques in future field of education and a methodology to improve the accuracy of the emotional dictionary were presented. In addition, through the analysis in this study, the direction of development of the free semester system was searched and a solution that can be used as a measure of improvement point was derived.

Key words: Opinion mining, free-semester, big data analytics, natural language processing, semester

INTRODUCTION

At present, the South Korean education system introduced the free semester system as a solution to cultivate the creative fusion talents of learners. The free semester system was implemented for one semester in the middle school curriculum with a view to implement personal customized education to enable learners can focus on career search without the burden of the exam to undergo practice, discussion and occupation experience activities. Therefore, the introduction of the free semester system can be interpreted as a change to improve the education system into a customized education system suitable for the future career and aptitude of the learner through the implementation of creative customized education (Khan and Choi, 2015). The introduction of the free semester system has become an issue, since 2012 and the government has gradually introduced it in 2016, centering on examples of advanced education abroad. In the introduction of the free semester system as such supporting and opposing public opinions have been formed among experts, teachers, students and parents in various education classes according to the system and the resultant elements are evaluated as successful introduction. However, while the introduction of the free semester system has many positive perspectives, the formation of negative views and the public opinions which show the side effects is shown to be high.

However, since scales that quantitatively and clarify present specific problems in the opposing opinions are not sufficient, guidelines to reflect the problems on future improvement of the free semester system have not been formed. In order to solve this problem, this study collected informal data according to the free semester system written by clients on online media and conducted opinion data mining based on big data. The writings on the reputation in news, blogs, communities, social media, and Twitter shown online from the moment when the word "free semester" appeared in the online media in 2012 and until the full introduction in 2015 were analyzedto conduct quantitative analyses of supporting and opposing public opinions and keywords for opposing public opinions were presented. Therefore, the results of this study are expected to become a scale for analysis that will provide the guidelines for the improvement of the free semester system in South Korea in the future.

Literature review

Opinion mining analysis technique: Opinion mining is a classification of text mining and is also called reputation analysis. It determines the positive, negative and neutral preferences of formal and informal texts in social media and is utilized for the forecast of the market sizes of specific services and products and analyses of consumer's responses and words-of-mouth, etc., (Ghose *et al*., 2007; Pang and Lee, 2008). Opinion mining

Corresponding Author: Ji-Hoon Seo, Department of Computer Science and Engineering, Incheon National University, Incheon National University, Academy-ro, 119 406-772 Yeonsu-gu, Korea

extracts vocabulary information expressing positive and negative responses and recognizes object and sentences consisting of opinions about the object to measure positive and negative responses with the sum of patterns including opinions (Courses and Surveys, 2008). As such opinion mining can obtain more valuable information from informal data made up of opinions of many unspecified users. To interpret in another aspect, it is also called sentiment analysis and is interpreted as the broad meaning of natural language processing, computer linguistics analysis and text mining (Deok *et al*., 2015).

Reputation analysis techniques for public opinions on the free semester system: There are various techniques for the reputation analysis for public opinions on the free semester system in South Korea. "A Study of School Life Satisfaction Through the Free Semester Curriculum" announced in May 2015 utilized the ANOVA method as a method to analyze how much students are satisfied with their school life while receiving the changed flexible curriculum with first year middle school students of schools where the free semester system was implemented. "A Study on Students Need for Free Learning Semester Programs of Libraries" developed questionnaires to investigate student's demand for free semester system programs and based on the questionnaires, the perception was analyzed with first year and second year middle school students (Hee, 2016). In addition, cases where limitations were analyzed by studying the level of career recognition of middle school students and the recognition types for the free semester system were derived (Suk, 2016).

MATERIALS AND METHODS

Proposed method: In order to derive the reputation and trends of clients distributed on online media based on informal data, the Opinion Mining presented in this study collects informal data including the keyword "free semester system" from five items that is news, blogs, social media, communities and Twitter to classify words and select emotion data (Fig. 1).

The overall system configuration of the emotional dictionary development proposed in this study is divided into three models; storage server for data collection, storage and preprocessing, NLP learning model for natural language processing and stemming analysis and the emotional dictionary construction server conversion stage. In this study, data collection, data classification, data preprocessing, emotional dictionary construction, word tagging and data analysis were performed for opinion mining analysis.

Data collection: In order to perform opinion mining an emotional dictionary should be constructed. In order to improve the accuracy of the emotional dictionary, collection of informal data is suggested as an important element. The timing and scale of the collected data should be clear and the more collected data, the more accurate the sensitivity dictionary can be. In this study, informal data for five online media related to the free semester system were collected in real time.

Data classification: In this study, data of news, blog, Twitter, social media and community derived from online media were collected and categories were created and classified. The classified data was stored in the main server.

Data preprocessing: Data preprocessing is a process for selecting emotional word candidates. In this study, there are no missing values, outliers and wrong values that can appear in numerical data because preprocessing is performed based on informal data. However, it is necessary to perform filtering to extract important words which are high in weight or can be emotional words in a sentence. In the case of English grammar, SentiWordNet can be used to construct emotional dictionary. However, since the accuracy of Korean syntax may be low due the structure of the grammar in this study, word filtering was performed in order to extract meaningful words in the following rules.

Word stemming filter process:

- Application of word filtering rules for emotion dictionary construction
- Remove special characters, English and unused words
- Remove meaningless terms and one-letter texts
- Classify the essence of emotional words by separating the same words in conjunction form
- Classify homonym/synonyms
- In the case of abbreviations and newly coined words, only the terms listed in the Wikipedia and Korean dictionaries are reflected

Emotional dictionary construction: The words derived through the preprocessing of meaningful words are classified into candidates of emotional words. In order to improve the accuracy of opinion mining analysis, the top 20% of the selected words were selected as emotional words (Fig. 2). The data distributed in the lower part consist of a group of words with high weight or meaning but when they are tagged as positive, negative, neutral or other words they are composed of words close to neutral

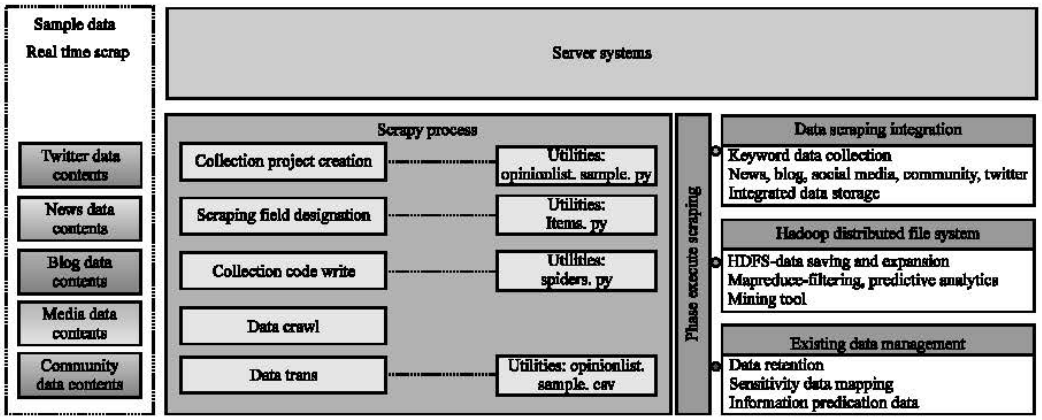

Fig. 1: Opinion analysis model

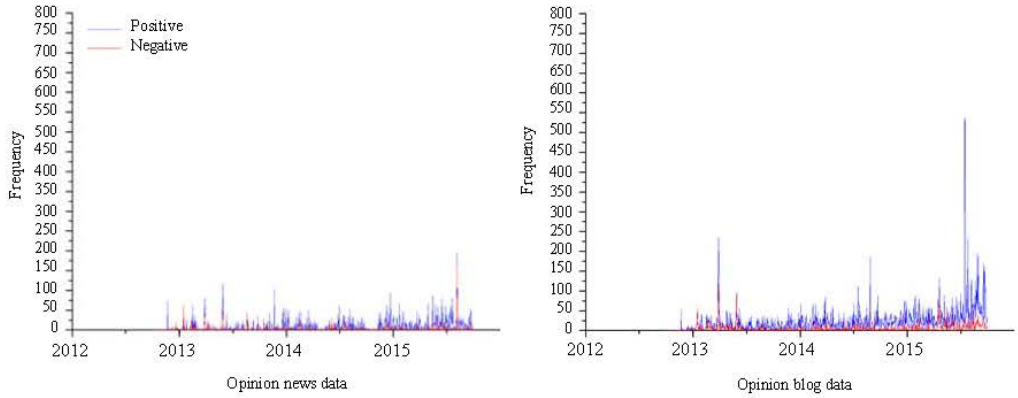

Fig. 2: Analysis of news and blog reputation for free semester system

Table 1: Emotional word selection results

Division		Sensltive word candidate			Training sensitive word
Step	Filtering data	Removal (%)	Candidate (%)	Use (%)	
Area standardp	Student: Filter by Top 20%	80	20	20	20
Filtering standard	80% / 20%	80	0	20	20
	70% / 10% / 20%	70	10	20	30
	60% / 20% / 20%	60	20	20	40
Training data standard	Remove the bottom 80%, top 20% use sensitive word	Filter Data (FD) = Total data-(Entry word+candidate word)			
	Remove the bottom 70%, top 30% use sensitive word	Sensitivity Word (SW) = $\frac{\text{Entry Word (EW)+Candidate Word (CW)}}{\text{Total Data (TD)}}$			
	Remove the bottom 60%, top 40% use sensitive word				

and other words. Therefore, the sets of data with low utility should be removed in advance. However, even among the 20% of the words selected as the emotional word, neutral and other words exist which correspond to the higher words with the high weight or the highest frequency in the document.

Tagging of emotional words and opinion mining analysis: Emotional word tagging is classified into positive, negative, neutral and other ones. The sentences recorded in one document are compared with the data in the emotional dictionary to calculate the frequencies of positive and negative words and the derived reputation data are classified into respective categories and stored monthly. An example of the frequency and type of emotional words derived from the document is shown in Table 1 and 2. Time series analysis was performed using the stored reputation data.

Table 2: Example of the tagging and frequency of emotional words derived from the document

Row No.	Word	Count	Type
1	Passion	1,191	Positive
2	True	205	Neutrality
3	Change	192	Neutrality
4	care for	164	Positive
5	Weak	156	Negative
6	Stick out	148	Neutrality
7	Difficult	137	Negative
8	Hope	104	Positive
9	Green	100	Others
10	Various	90	Neutrality
1	Various	112	Neutrality
2	New	42	Neutrality
3	Necessary	30	Neutrality
4	Good	20	Positive
5	Necessity	18	Neutrality
6	Important	17	Neutrality
7	Creative	17	Positive
8	Safety	16	Positive
9	Intensify	14	Positive
10	Funny	11	Positive

RESULTS AND DISCUSSION

According to the results of collecting informal data from online media containing the words “free semester system” in this study, the term “free semester system” was first mentioned in South Korea online from October 2012 and accordingly, analyses were conducted based on data from 2012-2015. The scope of analysis scope includes news, blog, social media, community and Twitter that are online media (Fig. 3 and 4).

In general, the news articles were not given much meaning in the analysis in this study because they are focused on the subjective view of the specialized columnist on the movement of the government and policies and people’s ideas not the objective client's reputation. However, the media view on the movements of the government generally derived positive results. Negative reactions were shown to be high at the time when the free semester system was mentioned in 2013, and positive public opinions were dominant until 2015. However, since August, 2015 as the full introduction of the free semester system was coming closer, concern about private educational expenses and problems due to side effects were derived, the contents were reflected on blogs and Twitter and news research were presented with negative research. In the case of blogs, negative public opinions were shown to be significantly higher than the reputation of news articles because clients were free to express their opinions. In other words, the client’s opinion was contrary to the high positive viewpoint reported in the media (Fig. 5).

Next, social media and community reputations for the introduction of the free semester system were analyzed. According to the results, social media showed a negative reputation in December 2012 but in 2015, positive contents increased overwhelmingly. This study did not reflect the gender and age, whether they are teachers, parents or students of writers written on blogs, Twitter, social media but the reason why social media reputation is good is expected to be the fact that students who are benefited from the free semester system showed positive responses. In the case of community sites using Internet bulletin boards, a large amount of data could not be collected because the number of client writers using the sites was small but negative views were shown to be higher compared to social media. Domestic community sites have an anonymous writing system so that clients can freely and honestly reflect their opinions (Fig. 4).

As a result, the collected data from the communities are not vast but they have some reliability and high accuracy. A review of analyses of the reputation of the free semester in the domestic Twitter shows that the negative response was high when the free semester system was mentioned before 2013 but the positive response has risen sharply since December 2014. In the process of filtering the words, the profanities, slang and abbreviations in Korean were excluded and as result, the amount of data derived was small. However, since the frequency of the contents close to negative words is high when all the profanities and slang are included, the results can be interpreted as showing high negative views in fact. Next, opinion mining analysis including all of news, blogs, Twitter, social medi and communities showed that there are high negative reputations in addition to the areas with outliers where some positive responses were shown. A. References to words with a high positive rate. B. References to words with a high negative rate (Fig. 5).

Finally, the results of visualization processing of word frequencies using word cloud when positive response rates and negative response rates increased greatly are as follows. Many words such as “freedom”, “support”, “experience”, “class”, “learning”, “career”, “field, etc. were derived when positive reputations were high. On the other hand, words such as “tuition fee”, “examination”, “learning”, “entrance examination”, “home”, “private education expenses” were mentioned when negative reputations increased. During the period of high positive reputation, the rate of positive reputations increased according to the expectation of diverse and free learning environments. During the period of high negative reputations, contents such as the burden of private educational expenses without any improvement in the education system and the lack of countermeasures against the entrance examination were presented. Therefore, in order to expand the social perspectives on the free semester system into positive perspectives, the

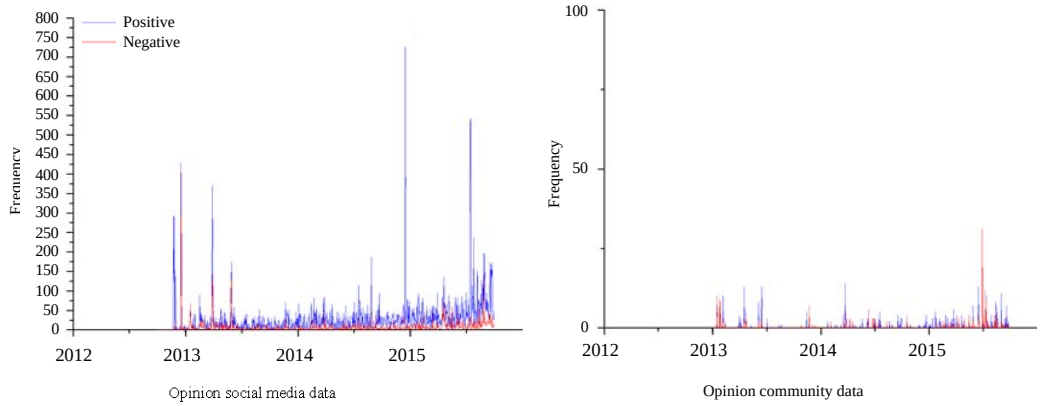

Fig. 3: Social media and community reputation analysis for the free semester system

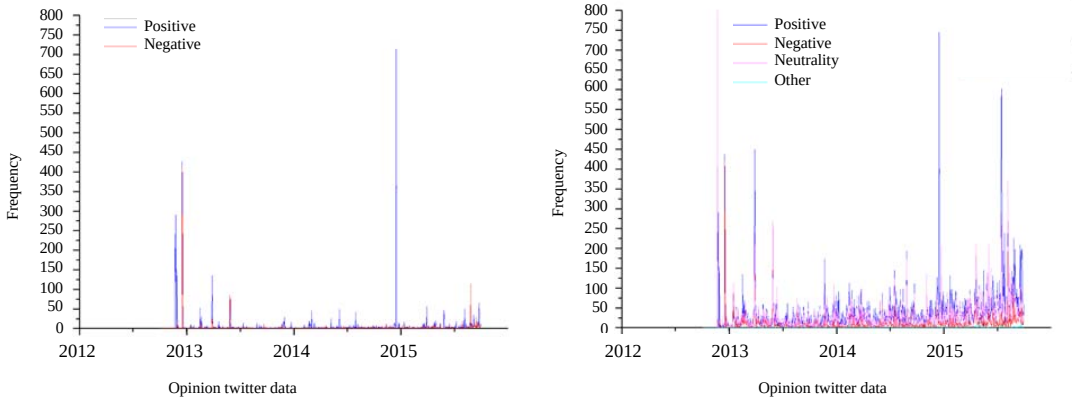

Fig. 4: Analyses of reputation of the free semester system in Twitter and all internet media

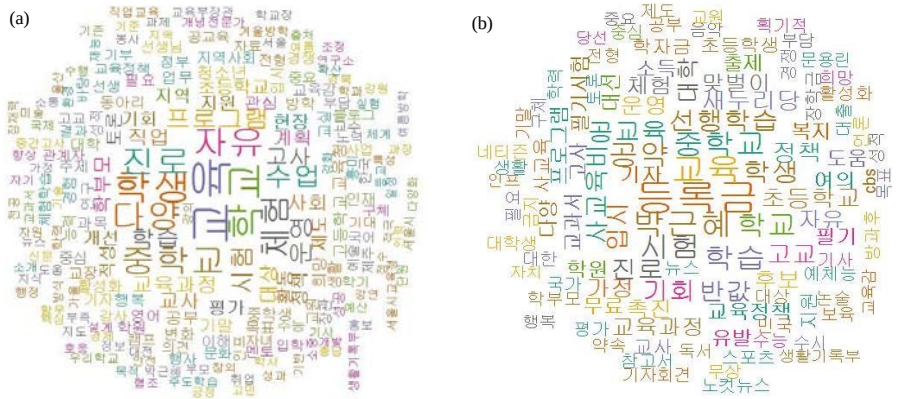

Fig. 5: Frequency and references of words for the free semester system: a) reference to words with a high positive rate and b) references to words with a high negitive rate

efficiency of the free semester system should be improved and the system should be improved to minimize the burden of private education and public education expenses. In addition, through the free semester system, customized education tailored to the aptitude of individuals should be implemented for creative education and career preparation and at the same time, the legal system should be improved so that the results of the free semester system can be reflected on the entrance examination system as negative aspects were presented at the moment when university should be entered through separate examinations.

CONCLUSION

This study collected informal data through online media to conduct analysis of client's reputation about the free semester system fully introduced into South Korea at present and presented the direction of improvement of the domestic education system in the future. According to the results of time series analyses of the data on the reputation, although high positive views were derived for the past 3 years, some blogs and communities showed high rates of contents indicating negative views.

Although, the gender, age and occupation group of the writers of data in the online media collected during the analysis process of the study cannot be identified, this study is expected to be highly utilizable for the seeking of the direction of the mid/long term development of the free semester system in South Korea and as a methodology of reputation analysis in the new education system, flip learning and Edutech.

ACKNOWLEDGEMENT

This research was supported by a grant (12-TI-C01) from Advanced Water Management Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

REFERENCES

Courses, E. and T. Surveys, 2008. Using sentiment sentiwordnet for multilingual sentiment analysis. Proceedings of the IEEE 24th International Conference on Data Engineering Workshop, April 7-12, 2008, IEEE, Cancun, Mexico, ISBN:978-1-4244-2161-9, pp: 507-512.

Deok, J.Y., Y.Y. Im and L.G. Taek, 2015. A study of school life satisfaction through free semester curriculum. J. Res. Educ., 21: 32-56.

Ghose, A., P.G. Ipeirotis and A. Sundararajan, 2007. Opinion mining using econometrics: A case study on reputation system. Proceedings of the 45th Annual Conference on Association of Computational Linguistics, June 23-30, 2007, ACL, Prague, Czech Republic, pp: 416-423.

Hee, N.Y. and K.H. In, 2016. A study on students need for free learning semester programs of libraries. J. Korean Lib. Inf. Sci. Soc., 47: 187-211.

Khan, I.A. and J.T. Choi, 2015. An application of Educational Data Mining (EDM) technique for scholarship prediction. Intl. J. Software Eng. Appl., 8: 31-42.

Pang, B. and L. Lee, 2008. Opinion mining and sentiment analysis. Found. Trends Inform. Retrieval, 2: 1-135.

Suk, Y.E., 2016. A study on the level of career awareness among middle school students and their recognition of free semester system. Master Thesis, Gongju National University, Gongju, South Korea.