

Comparing Machine Learning Classifiers for Movie WOM Opinion Mining

Yosin Kim¹, Do Young Kwon², Seung Ryul Jeong^{2*}

¹ Department of Management Information System, Chungbuk National University, Cheongju, Korea
e-mail: trust@chungbuk.ac.kr

² Graduate School of Business IT, Kookmin University, Seoul, South Korea
e-mail: doyoung, srjeong@kookmin.ac.kr

*Corresponding author: Seung Ryul Jeong

Received March 30, 2015; accepted July 1, 2015; published August 31, 2015

Abstract

Nowadays, online word-of-mouth has become a powerful influencer to marketing and sales in business. Opinion mining and sentiment analysis is frequently adopted at market research and business analytics field for analyzing word-of-mouth content. However, there still remain several challengeable areas for 1) sentiment analysis aiming for Korean word-of-mouth content in film market, 2) availability of machine learning models only using linguistic features, 3) effect of the size of the feature set. This study took a sample of 10,000 movie reviews which had posted extremely negative/positive rating in a movie portal site, and conducted sentiment analysis with four machine learning algorithms: naïve Bayesian, decision tree, neural network, and support vector machines. We found neural network and support vector machine produced better accuracy than naïve Bayesian and decision tree on every size of the feature set. Besides, the performance of them was boosting with increasing of the feature set size.

Keywords: WOM, Opinion Mining, Machine Learning, NLP, Film Market

1. Introduction

With emergence of social media and online communication, consumer's electronic word-of-mouth (eWOM), often called user-generated content, directly influenced to product sales and marketing, so that the firms aggressively participated in online communication and used it for marketing strategies and public relations [1]–[3]. As growing eWOM impact in business, requirement of analyzing eWOM content also increased. Opinion mining and sentiment analysis has been frequently adopted into market research and business analytics, since they emphasize extracting sentiment polarity and author's opinions such as positive, negative, and neutral sentiment [4]. Opinion mining involving sentiment analysis is regarding as a set of processes used to identify sentiment, opinion, and the author's attitude shown in texts, turn them into meaningful information, and use it in making business decisions.

Opinion mining is usually conducted of two ways, lexicon based approach and machine learning approach. Lexicon based approach uses the linguistic resource, called sentiment dictionary or sentiment lexicon, such as SentiWordNet, SenticNet, and OpinionFinder. Machine learning approach generates the classifying algorithm through learning with the set of linguistic features and applies it.

Many opinion mining studies have been conducted to analyze unstructured text data including customer reviews, blogs, tweets, news, and online contents [2], [3], [5]–[9]. However, there are still many challengeable areas left for more effective and accurate sentiment analysis of customer opinion. First, there were few sentiment analysis studies aiming at Korean contents. Korean, an agglutinative language, has much complicate structure to analyze corpora and syllables from text [10]. Next, there are not enough references of tools, techniques, and methods for Korean natural language processing (NLP) and machine learning models. Machine learning based sentiment analysis is able to analyze text data without the sentiment lexicon, but it requires well-defined data and much complicate classification algorithms for analytics. Last, even though few studies examined different sizes of the feature set to achieve higher performance of text categorization [11], [12], it was very rare to consider the impact of the linguistic features set.

Therefore, we tried to conduct sentiment analysis of Korean eWOM using machine learning models and see usability through performance comparison. We selected 10,000 movie reviews which had extremely negative/positive rates of popularity in a movie portal site, extracted linguistic features from them using NLP, and compared their performances within accuracy using machine learning models: naïve Bayesian (NB), decision tree (DT), artificial neural network (NN), and support vector machines (SVM). Furthermore, to see the role of the linguistic feature set in the algorithms, we tried to examine how much the performance of each classifier changes by the feature set size.

This paper thus has three objectives. First, we will try to provide the process of machine learning approach for sentiment analysis from NLP of unstructured text data to algorithm validation. It should be a reference of machine learning approach to mine consumer opinions from Korean eWOM. Secondly, we will examine classification algorithms of four machine learning models and compare their performances with accuracy measures. The result should provide helpful guides to business analysts and researchers. Last, we will test classifiers' performances on different sizes of the linguistic feature set. We expect it can present that

sentiment classifiers can maximize their performance through moderating the size of the feature. Additionally, we will employ open source packages of the R project for demonstrating Korean NLP and machine learning models, thus potential users can consider these tools and techniques for immediate adoption.

2. Related Works

2.1 Electronic Word of Mouth

Word-of-mouth, meaning personal communication among people, has been recognized as significant source of information to understand customers' interest, sentiment, and opinion about companies' products and service such as movies, books, music albums, hotels, and restaurants [2], [13]–[18]. Since many consumers have received word-of-mouth as useful and credible information generated by pre-experienced consumer not company's advertisements when they buy products or service. In particular, advancement of Internet technology has enabled people to easily generate electronic word-of-mouth(eWOM), share it with other people, and exchange opinions. This eWOM is including customer reviews, online comments, and population rankings, and spreading real-time through online channels such as e-commerce sites, online forums, blogosphere and social networking sites. Thus, eWOM is recognized as not only convenience for consumer but also new challenges and opportunities for business analysts to analyze consumer's interests and opinions.

Many researchers studied the influence of eWOM on marketing and sales. A study with the two leading online booksellers, Amazon.com and Barnesandnoble.com, characterized the behavior pattern of reviewers and examined the positive effect of customer reviews on product sales [16]. This research revealed that poor reviews with one-star reviews have the greater impact than complimentary reviews ranked with five-star and consumers prefer to read review text rather than just seeing summary statistics. Another eWOM research analyzed 40 movies' reviews collected from the Yahoo Movie, and stated movie eWOM caused significant effect on the box office revenue [2]. The impact of eWOM can be shown to travel business, one of industries with very active online communication. A research of eWOM impact on travel business analyzed 1,639 hotels' 40,424 review rating data of a major online travel agency in China and showed a very positive result that increasing 10 percent in customer review ratings enables online reservation to boost more than five percent [19]. In music album market, blog chatter about music albums has strong impact on music album sales [20]. In addition, eWOM of restaurants showed that consumer-generated reviews have the positive effect as increasing the online popularity, but editors' reviews and ratings are negative to visiting the restaurants websites [17]. Consequently, many researchers and their studies showed the influence of eWOM on marketing and sales in various retail business, e-commerce, and even stock market.

2.2 eWOM Analysis in Film Market

Consumer voice of film market has been a hot item for finding eWOM impact to marketing and sales for a while. A researcher analyzed review volume and ratings of 40 movies collected from the message board of Yahoo Movies, showed that eWOM before release of the movie is very active and has significant effect on the box office revenue [2]. On the other hand, another research produced different results with previous studies in their research investigating the relationship between movie sales and their online reviews and ratings [18].

According to this study, there are true effects between two areas but higher ratings do not ensure higher sales. It also additionally explained that the awareness effect works but the persuasive effect does not influence on consumers. A study in Korean film market examining the role of eWOM compared movie reviews in Internet portal website and tweets in Twitter, and showed that both tweets and movie reviews have the significant effect of eWOM [21]. The study used several variables such as released screens, average ratings, number of ratings, number of tweets, and revenues for analyzing eWOM effect, but sentiment of reviews and tweets did not considered to the analysis model.

Early stage research of eWOM tended to use only volume and ratings without analyzing syntax of content, thus it was increasingly seemed to be insufficient to explain dynamics of online communication. Since eWOM was generated as the type of text contents including consumers' interests and opinion, sentiments of user-generated contents should be applied as a parameter in eWOM analysis [22]. A study about movie eWOM selected online contents of 257 movies from Yahoo Movies during 2005-2006, extracted five measures with SentiWordNet and OpinionFinder: volume, valence, subjectivity, number of sentences, and number of valence words [23]. This paper stated that the volume of messages has significant correlation to forecast movie box office sales but the valence of eWOM did not show the meaningful relation between WOM and movie sales. However, another research analyzing movie eWOM suggested that consumer's willingness to watch a movie is significantly influenced by valence of eWOM [3]. In addition, a meta-analysis research about eWOM stated that review valence has more significant influence than review volume on sales elasticity [24].

2.3 Opinion Mining & Sentiment Analysis

Opinion mining and sentiment analysis, which emphasizes extracting author's opinions and sentiment polarity like positive and negative, is regarding as a set of processes used to identify the author's sentiment, opinion, and attitude shown in texts, and turn them into meaningful information. In accordance with needs of unstructured text documents mining, opinion mining and sentiment analysis is frequently applied into analyzing sentiment of eWOM [4]. According to Chen & Zimbra [9], "Opinion mining, as a sub-discipline within data mining and computational linguistics, is referred to as the computational techniques used to extract, classify, understand, and assess the opinions expressed in various online news sources, social media comments, and other user-generated content" (p. 74). Sentiment analysis is often used in opinion mining to identify sentiments, affect, subjectivity, and other emotional state in the text content [25].

Opinion mining is basically in charge of classifying sentiment such as polarity or emotion of a piece of text [4], [9], [25]. Classifying sentiment of text content with automatic way is usually conducted of two ways, lexicon based approach and machine learning approach [10]. Lexicon based approach uses the linguistic resource, called sentiment dictionary or sentiment lexicon, such as SentiWordNet, SenticNet, and OpinionFinder. Machine learning approach generates the classifying algorithm through learning with the set of linguistic features and applies it.

Lexicon Based Approach

In lexicon based approach, public lexicons like SentiWordNet is frequently applied into many studies due to high coverage of them and reliability problem of a manual sentiment

dictionary [26]. For instance, a study identified sentiment with SentiWordNet in a market intelligence framework for evaluating business constituents and conducted a case study of Wal-Mart [6]. This paper showed several opinion analysis results including traffic dynamics, topic and sentiment evolution, active topics and sentiments, and opinion leader analysis, and also stated that those results can explain that Wal-Mart has to recognize marketing intelligence where, which items, and who need to pay special attention. A study analyzing eWOM of the Hollywood film market extracted opinion measures like valence and number of valence words using SentiWordNet and OpinionFinder [23]. Another study measuring the correlation between Dow Jones Industrial Average (DJIA) and online mood states on tweets collected over nine million tweets and tagged the mood of each tweet with OpinionFinder and Google-profile of mood states [27].

On the other hand, a study proposed probability model to measure sentiment of financial news article and predicted movement of the Korea composite stock price index [5]. It suggested an opinion mining method generating a domain-specific sentiment lexicon, and eventually accomplished 63.0% F1 score on the threshold of validation data. Another paper compared stock domain specific sentiment dictionary with a general dictionary and gained higher accuracy [28]. One of the contextual meaning study also proposed an algorithm to automatically build a word-level emotional dictionary for social emotion detection and stated the dictionary generated for certain purpose is more efficient to predict the emotional distribution of news articles [29]. In addition, another study took 4.10% improvement of accuracy from revised SentiWordNet using objective words in SentiWordNet rather than original lexicon [26].

Machine Learning Based Approach

Machine learning methods are also frequently used in sentiment analysis and opinion mining. This approach generates classification algorithms with linguistic features in training data set and tests the performance of algorithms in validation data set. Various machine learning algorithms such as NB, SVM, NN, genetic algorithm (GA) and k-Nearest Neighbors (kNN), have been applied into classification, optimizations and predictions from text document [3], [10], [11], [30], [31].

Pang et. al, applied three machine learning methods such as NB, SVM, and maximum entropy classification into determine whether a movie review is positive or negative[32]. They examined several conditions such as n-gram, tense, and linguistic set size, therefore achieved the best performance of 82.9% accuracy in SVM method combining with unigrams, presence, and a 16,165 features. Interestingly, their experiment showed using top frequent 2633 unigrams reached 81.4% accuracy as much similar as the best. It means a small size feature set can be considered as an efficient method for sentiment analysis of big data. Another study employed SVM and NB classifier to test eWOM impact of consumer's willingness[3]. It categorized 4,166,623 movie tweets into four mutually exclusive categories: intention, positive, negative, and neutral. The researchers trained NB for intention and SVM for sentiment, and validated eWOM impact of tweets with precision and recall as performances measures.

In stock market prediction, a study applied a SVM method to analyze financial news articles [31]. It examined article terms, the stock price, and three textual representations: bag of words, noun phrases, and named entities, for investigating the prediction performance, and used closeness, directional accuracy, and a simulated trading engine for evaluation.

Eventually it showed that both article terms and stock price at the time article release had the best performance and proper nouns had the better textual representation performance. Another study, which predicted daily stock price trend with company-specific news articles, proposed a prediction model, NewsCATS[11]. The researchers used an automated text categorization technique with a manual thesaurus and employed variables such as collection term frequency (CTF), inverse document frequency (IDF), CTFxIDF, and used like the document processing like within-document frequency (WDF), IDF, and WDFxIDF. Furthermore, they examined different sizes of the feature set, and trained classifiers with Rocchio, kNN, and non-linear SVM (nLSVM). The results were reported with harmonic mean F1 and overall accuracy as measures, and achieved best performance when conditions were CDF, 15% size of feature set, WDFxIDF, and the nLSVM classifier.

On the other hand, the research with Korean contents has been very rare yet till now. A recent study tried machine learning methods to classify emotions of micro-blog, tweet, using features of morphemes and n-grams [10]. They also employed NB and SVM as classification algorithms and used F1 score as the measure of classification performance. According to the experiment results, SVM showed higher performance than NB and especially nLSVM with a polynomial kernel in tri-gram produced almost 84% accuracy.

3. Approach

3.1 Overview

Our sentiment classification model is composed of NLP phase and machine learning phase. NLP phase consists of gathering data, cleansing data, manipulating data, extracting linguistic resources, selecting features and so on. Machine learning phase contains includes combining features and documents, learning algorithms in training data set, and testing the algorithms in validation data set. During a learning phase, algorithms capture the polarities inherent in training data categorized as negative or positive. These algorithms are examined in validation data set to see their performance and availability. Fig. 1 presents process and functions of our sentiment classification model in more detail.

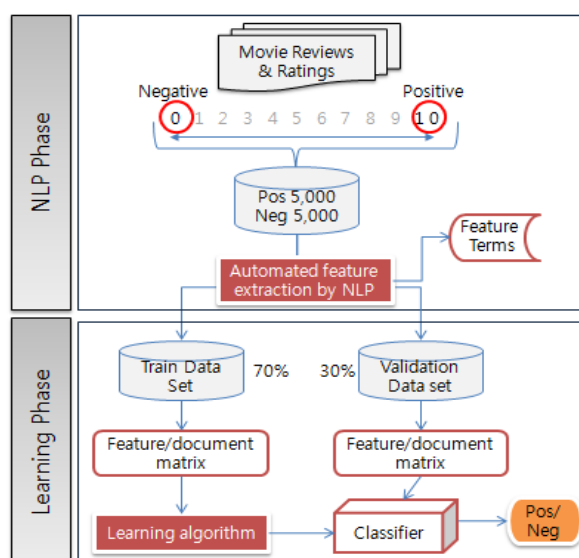


Fig. 1. Overview of the Sentiment Classification Model

3.2 Data Aggregation

The data of movie reviews and ratings was collected from movie section of portal websites in Korea (<http://nate.com> and <http://movie.daum.net>). In these sites, online users can write reviews and ratings about movies. A review is allowed up to 140 letters and a rating score can be evaluated between zero meaning extremely negative and ten as the highest score. The collecting data was generated from December 2008 to July 2013, and we selected both 5,000 negative reviews rated zero (0) and 5,000 positive reviews with ten (10) for clear polarity division.

3.3 NLP and Generating Feature Sets

In order to optimally utilize the amount of movie reviews available in the form of unstructured text data, it is necessary to employ tools and technologies that extract relevant information and analyze efficiently them. After movie WOM data were collected, we conducted text data processing to moderate unstructured text data and generate the feature set which are applied into machine learning models. In normal WOM text with natural language, there are many kinds of useless characters such as single letters (e.g., kkk and hhh), emoticons (e.g., :), -.-, and ^^), numbers, and punctuations, which have often been barriers to extract the feature from contents. Thus, we first removed these obstacles from experiment data. We also eliminated English stop-words (e.g., I, me, my, and mine) and too much common words like “Korean” and “movie” for the efficient processing. We next extracted nouns from the sentence of reviews and filtered them with a rule which remains characters between eight letters and two letters in lengths.

As the result, the extracted nouns were totally 22,652 terms, and we transformed them as a document-term matrix. Even though terms were extracted for machine learning methods, it is impossible to use all terms as the features in machine learning methods. The size of the features is one of important factors in machine learning algorithm since it can be influence of performance both accuracy and processing. We thus selected more frequent terms in the data using a function of NLP which removes sparse terms, and applied them into training and testing phase. That selected feature sets, as shown **Table 1**, were 5 groups as their terms size from a set with 52 terms to a group with 434 terms. We used the Korean NLP package (KoNLP) in R project for this pre-processing works.

Table 1. Feature Sets for Machine Learning

Group	Sparsity	No. of Terms
Set 1	.9966	52
Set 2	.9980	108
Set 3	.9988	215
Set 4	.9992	312
Set 5	.9994	434

3.4 Machine Learning

Machine learning is a subfield of artificial intelligence and computer science, and deals with algorithms and techniques that improve automatically through experience rather than follow programmed instructions. Machine learning is applied into various tasks from data mining programs that discover general rules in large data sets, to information filtering systems.

There are various machine learning methods, but we employed four methods which are frequently adopted in previous research: NB, DT, NN, and SVM. NB is one of frequently employed classifiers for text mining because of their simplicity. DT learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. Artificial neural network method is a learning algorithm that is inspired by the structure and functional aspects of biological neural networks.

4. Experiment and Result

4.1 Evaluation

The performance of algorithms and feature sets in text classification can be evaluated by several statistical measures such as recall, precision, accuracy α and F1-score [3], [5], [10], [11]. In this study, we used the overall accuracy α and F1-score. Total accuracy is defined as the percentage of sentiments correctly predicted of total instances:

$$\text{Accuracy } \alpha = \frac{\text{instances correctly predicted}}{\text{Total Instances}}$$

F1 score is the harmonic mean of macro-averaged precision and recall. Precision is the proportion of right instances of total instances predicted, and recall is the proportion of instances correctly predicted of real instances. It seems to be good to take higher results on both of them, but they generally have inverse relationship. For that reason, many researchers have used F1 score combined two measures as a harmonic mean of them:

$$\text{F1 Score} = \frac{2 * \text{recall} * \text{precision}}{\text{recall} + \text{precision}}$$

4.2 Classifying Sentiment

We conducted sentiment classification with four algorithms and five groups with different feature set size. The aim of this analysis is to see 1) if machine learning algorithms for sentiment classification work with only linguistic features extracted by NLP without manual handling, 2) if the size of the feature set influence the performance of classification. **Table 2** is our experiment results. It presents five groups of the feature set size, and there are two performance measures (accuracy and F1 score) of four algorithms in two data sets (training and validation set). In train data set, we can see that algorithms within NN and SVM produced significant classifying accuracy around 75 % in Group 4 and Group 5. SVM algorithm reached the highest accuracy at Group 5 (accuracy=.7434, F1=.6307), NN classifier also achieved the maximized value of measure at Group 4 (accuracy=.7534, F1=.6361). On the other hand, NB method remained under 60% accuracy and showed a much low F1 score even at the highest accuracy (accuracy=.5752, F1=.3133 in Group 2). In text classification, it is frequently founded in many studies that SVM produced higher accuracy than NB method [33]. As the result in train data set, we can consider the results that NN and SVM lead to higher performances than NB and DT, as well as the accuracy is increasing along with the size of the feature in train data set.

However, in validation data set, the highest accuracy keeps staying around 70%, as shown Group 4 (accuracy=.7072, F1=.5984 with NN) and Group 5 (accuracy=.7063, F1=.6033 with SVM) despite the size of the feature set is increasing. It can be interpreted that the size of the feature influences the accuracy of classifying but the effect of it does not rise after reaching a certain threshold for the performance.

Table 2. Classifying Performance of Classifiers on Train Data vs. Validation Data

Group	Classifier	Train Data Set		Validation Data Set	
		Accuracy	F1 Score	Accuracy	F1 Score
<i>Group1: Feature Terms=52</i>	NB	0.5513	0.2650	0.5605	0.2833
	DT	0.5819	0.3012	0.5789	0.2986
	NN	0.6520	0.4871	0.6448	0.4841
	SVM	0.6445	0.4761	0.6455	0.4806
<i>Group2: Feature Terms=108</i>	NB	0.5752	0.3133	0.5755	0.3233
	DT	0.5870	0.3166	0.5865	0.3217
	NN	0.6867	0.5492	0.6583	0.6523
	SVM	0.6725	0.5244	0.6679	0.5167
<i>Group3: Feature Terms=215</i>	NB	0.5546	0.2555	0.5724	0.2795
	DT	0.5807	0.3145	0.6007	0.3262
	NN	0.7232	0.6024	0.6962	0.5772
	SVM	0.7002	0.5695	0.7025	0.5717
<i>Group4: Feature Terms=311</i>	NB	0.5197	0.1302	0.5057	0.1348
	DT	0.5816	0.2977	0.5796	0.3067
	NN	0.7381	0.6305	0.7072	0.5984
	SVM	0.7212	0.5987	0.7038	0.5822
<i>Group5: Feature Terms=434</i>	NB	0.4986	0.0397	0.4997	0.0536
	DT	0.5839	0.3027	0.5743	0.2952
	NN	0.7534	0.6361	0.7013	0.6013
	SVM	0.7434	0.6307	0.7063	0.6033

In below Fig 2, the change of classification performance is more clearly shown. In training data set, increasing of the feature set is continuously boosting the accuracy of NN and SVM, but NB and DT shows less performance on increasing figure terms. This pattern is almost same within validation data set, but we can see that the accuracy of NN and SVM remains around 70% after reaching it.

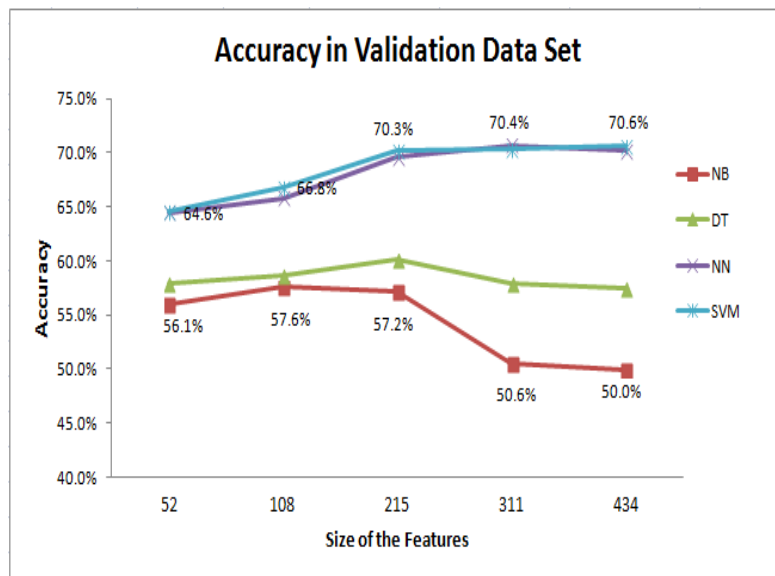


Fig. 2. Accuracy in Validation Data

In DT model for classifying sentiment, we can see which feature terms are located in the root of the tree to decide a review's polarity as positive or negative. As shown Fig. 3, several words like “trash”, “story”, “people”, “fun”, and “scenario” are presented in there. In particular, we can find out that “trash” plays the critical role in classifying reviews into the negative area and “fun” works as a major positive classifying feature.

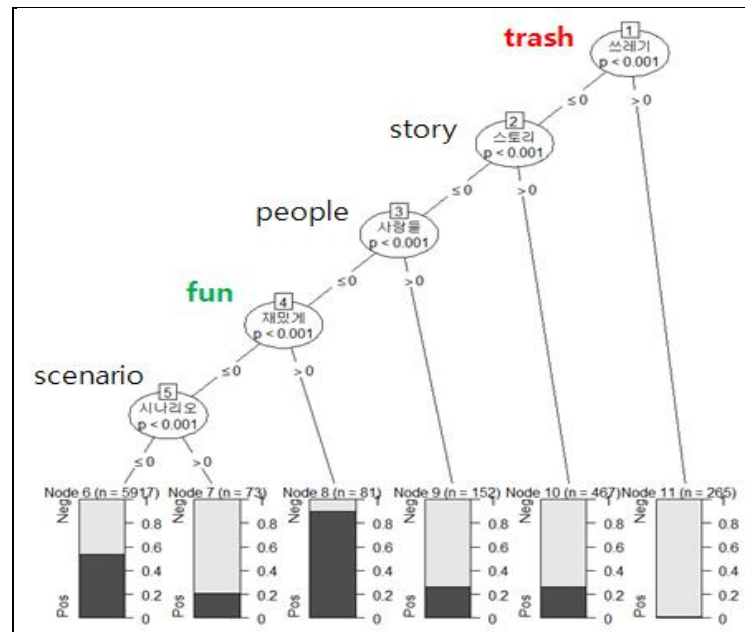


Fig. 3. Keywords in the Decision Tree Classifier

5. Conclusion

In this paper, we aim to 1) conduct sentiment analysis aiming for Korean eWOM in the film market, 2) detect the availability of machine learning models only using linguistic features, and 3) identify influence of the size of the linguistic feature set. We selected 10,000 movie reviews which were rated of extremely negative/positive popularity in the movie portal sites, and parsed words through NLP. Four learn machine methods (NB, DT, NN, and SVM) were demonstrated with the linguistic features and their performances were compared by accuracy and F1 score. In addition, we tested five groups whether the feature set size influences on the performance of classification. As a result, NN and SVM classification showed the outstanding performance. NN classifier with 311 terms reached accuracy 70.72% (F1 = .5984) and SVM classifier consisting 434 terms drove accuracy 70.63 % (F1 = .6033).

Through the experiments, we showed how machine learning algorithms are applied as sentiment classifiers for movie eWOM analytics, and the performance gap might be occurred by the method as well as the feature set size. In addition, NN and SVM generally have better performance than NB and DT under every condition. Besides, the performance of them might increase according to the feature set size.

We expect our experiment to provide several implications and contributions under below. The first is providing the process of machine learning approach for sentiment analysis from data aggregation to algorithm validation. It should be useful to discover consumer opinion of

Korean eWOM. Another is comparing various machine learning methods for sentiment analysis and revealing how different they are in performance. It should provide researchers how to make comparison research in opinion mining and business analysts which classification method is more accurate for sentiment analysis. Last, our testing with feature set size implies that the researchers and analysts should consider how many features is optimum for classification accuracy as well as analytics performance in big data opinion mining. In addition, we used open source packages of the R project for demonstrating Korean NLP and machine learning methods, thus potential users can consider these tools and techniques for immediate adoption. We believe this article can support practical and reliable reference to eWOM analysis, not just in movie review but in other eWOM data as well.

This research also has several challenges to be improved. Our research attempted sentiment classification of movie reviews within positive and negative however more various sentiment spectrums such as emotions, willingness, and complaints, will be analyzed in the future research. In addition, if linguistic feature words are handled in a more elaborate way like the expert's verification, the performance of classification would be improved. Another challenge is to employ more features in huge volume of eWOM data as real Big-data. Last, future research needs to expand to various industries such as entertainment, e-commerce, and health care service within expanding social networking sites like Twitters and Facebook.

References

- [1] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media," *Bus. Horiz.*, vol. 54, no. 3, pp. 241–251, May 2011. [Article \(CrossRef Link\)](#).
- [2] Y. Liu, "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *J. Mark.*, vol. 70, no. 3, pp. 74–89, Jul 2006. [Article \(CrossRef Link\)](#).
- [3] H. Rui, Y. Liu, and A. Whinston, "Whose and what chatter matters? The effect of tweets on movie sales," *Decis. Support Syst.*, vol. 55, no. 4, pp. 863–870, Nov 2013. [Article \(CrossRef Link\)](#).
- [4] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 15–21, Mar 2013. [Article \(CrossRef Link\)](#).
- [5] Y. Kim, S. R. Jeong, and I. Ghani, "Text Opinion Mining to Analyze News for Stock Market Prediction," *Int. J. Adv. Soft Comput. Its Appl.*, vol. 6, no. 1, 2014. [Article \(CrossRef Link\)](#).
- [6] H. Chen, "Business and Market Intelligence 2.0, Part 2," *IEEE Intell. Syst.*, vol. 25, no. 2, pp. 2–5, Mar 2010. [Article \(CrossRef Link\)](#).
- [7] L. Garcia-Moya, H. Anaya-Sanchez, and R. Berlanga-Llavori, "Retrieving Product Features and Opinions from Customer Reviews," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 19–27, May 2013. [Article \(CrossRef Link\)](#).
- [8] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu, "OpinionSeer: interactive visualization of hotel customer feedback," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1109–18, 2010. [Article \(CrossRef Link\)](#).
- [9] H. Chen and D. Zimbra, "AI and Opinion Mining," *IEEE Intell. Syst.*, vol. 25, no. 3, pp. 74–76, May 2010. [Article \(CrossRef Link\)](#).
- [10] J. Lim and J. Kim, "An Empirical Comparison of Machine Learning Models for Classifying Emotions in Korean Twitter," *J. Korea Multimed. Soc.*, vol. 17, no. 2, pp. 232–239, 2014. [Article \(CrossRef Link\)](#).
- [11] M. Mittermayer and G. Knolmayer, "NewsCATS: A News Categorization and Trading System," *Sixth Int. Conf. Data Min.*, pp. 1002–1007, Dec 2006. [Article \(CrossRef Link\)](#).
- [12] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Proc. of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79 – 86, 2002. [Article \(CrossRef Link\)](#).

- [13] S. Melian-Gonzalez, J. Bulchand-Gidumal, and B. Gonzalez Lopez-Valcarcel, "Online Customer Reviews of Hotels: As Participation Increases, Better Evaluation Is Obtained," *Cornell Hosp. Q.*, vol. 54, no. 3, pp. 274–283, Mar 2013. [Article \(CrossRef Link\)](#).
- [14] X. Lu, S. Ba, L. Huang, and Y. Feng, "Promotional Marketing or Word-of-Mouth? Evidence from Online Restaurant Reviews," *Inf. Syst. Res.*, vol. 24, no.3, September 2013, pp. 596–612, Jan 2013. [Article \(CrossRef Link\)](#).
- [15] F. V. Wangenheim and T. Bayón, "The effect of word of mouth on services switching: Measurement and moderating variables," *Eur. J. Mark.*, vol. 38, no. 9/10, pp. 1173–1185, 2004. [Article \(CrossRef Link\)](#).
- [16] J. A. Chevalier and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *J. Mark. Res.*, vol. 43, no. 3, pp. 345–354, Aug 2006. [Article \(CrossRef Link\)](#).
- [17] Z. Zhang, Q. Ye, R. Law, and Y. Li, "The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews," *Int. J. Hosp. Manag.*, vol. 29, no. 4, pp. 694–700, Dec 2010. [Article \(CrossRef Link\)](#).
- [18] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter? — An empirical investigation of panel data," *Decis. Support Syst.*, vol. 45, no. 4, pp. 1007–1016, Nov 2008. [Article \(CrossRef Link\)](#).
- [19] Q. Ye, R. Law, B. Gu, and W. Chen, "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings," *Comput. Human Behav.*, vol. 27, no. 2, pp. 634–639, Mar 2011. [Article \(CrossRef Link\)](#).
- [20] V. Dhar and E. a. Chang, "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *J. Interact. Mark.*, vol. 23, no. 4, pp. 300–307, Nov 2009. [Article \(CrossRef Link\)](#).
- [21] J. Lee, I. Son, and D. Lee, "Does Online Social Network Contribute to WOM Effect on Product Sales?," *J. Intell. Inforamtion Syst.*, vol. 18, no. 2, pp. 85–105, 2012. [Article \(CrossRef Link\)](#).
- [22] G. P. Sonnier, L. McAlister, and O. J. Rutz, "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Mark. Sci.*, vol. 30, no. 4, pp. 702–716, Jul 2011. [Article \(CrossRef Link\)](#).
- [23] Y. Liu, Y. Chen, R. F. Lusch, H. Chen, D. Zimbra, and S. Zeng, "User-Generated Content on Social Media: Predicting Market Success with Online Word-on-Mouth," *IEEE Intell. Syst.*, vol. 25, no. 1, pp. 8–12, 2010. [Article \(CrossRef Link\)](#).
- [24] K. Floyd, R. Freling, S. Alhoqail, H. Y. Cho, and T. Freling, "How Online Product Reviews Affect Retail Sales: A Meta-analysis," *J. Retail.*, vol. 90, no. 2, pp. 217–232, Jun 2014. [Article \(CrossRef Link\)](#).
- [25] B. Pang and L. Lee, *Opinion Mining and Sentiment Analysis*. 2008. [Article \(CrossRef Link\)](#).
- [26] C. Hung and H.-K. Lin, "Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 47–54, Mar, 2013. [Article \(CrossRef Link\)](#).
- [27] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011. [Article \(CrossRef Link\)](#).
- [28] Y. Yu, Y. Kim, N. Kim, and S. R. Jeong, "Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary," *J. Intell. Inforamtion Syst.*, vol. 19, no. 1, pp. 95–110, 2013. [Article \(CrossRef Link\)](#).
- [29] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," in *World Wide Web*, 2013. [Article \(CrossRef Link\)](#).
- [30] R. Schumaker and H. Chen, "A discrete stock price prediction engine based on financial news," *Computer (Long. Beach. Calif)*, no. January, pp. 51–56, 2010. [Article \(CrossRef Link\)](#).
- [31] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news," *ACM Trans. Inf. Syst.*, vol. 27, no. 2, pp. 1–19, Feb, 2009. [Article \(CrossRef Link\)](#).
- [32] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proc. of 2002 Conf. Empir. Methods Nat. Lang. Process.*, pp. 79 – 86, 2002. [Article \(CrossRef Link\)](#).

- [33] W. Paik, M. H. Kyung, K. S. Min, H. R. Oh, C. Lim, and M. S. Shin, "Multi-stage News Classification System for Predicting Stock Price Changes," *J. Korean Soc. Inf. Manag.*, vol. 24, no. 2, pp. 123–141, 2007. [Article \(CrossRef Link\)](#).



Yoosin Kim is a Research Professor in the Department of MIS at Chungbuk National University. He received a Ph.D. in MIS from Kookmin University in Seoul, Korea and was a post-doctoral research fellow at the College of Business, the University of Texas at Arlington. He had worked as a Data Scientist and a Business Analyst for a number of firms, including Accenture and SK C&C, in financial, medical, and e-commerce fields. His research interests include business analytics for consumer opinion mining, market sensing, recommender systems, and business intelligence using Big Data.



Do Young Kwon pursues a Master's degree at Kookmin University, Korea. Her research interests include opinion mining and business analytics. She has published several papers in various conference proceedings and journals such as *Journal of Intelligence and Information Systems*



Seung Ryul Jeong is a Professor in the Graduate School of Business IT at Kookmin University, Korea. He holds a B.A. in Economics from Sogang University, Korea, an M.S. in MIS from University of Wisconsin, and a Ph.D. in MIS from the University of South Carolina, U.S.A. Professor Jeong has published extensively in the information systems field, with over 60 publications in refereed journals like *Journal of MIS*, *Communications of the ACM*, *Information and Management*, *Journal of Systems and Software*, among others.