JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 22, No. 9: 1153~1158, Sep. 2018

엔트로피 점수를 이용한 감성분석 분류알고리즘의 수행도 평가

박만희*

Evaluation of Classification Algorithm Performance of Sentiment Analysis Using Entropy Score

Man-Hee Park*

Department of Business Administration, Catholic University of Pusan, Busan 46252, Korea

요 약

다양한 온라인 고객 평가 및 소셜 미디어 정보는 고객의 의사결정에 영향을 미치기 때문에 기업에게 매우 중요한 정보 출처라고 할 수 있다. 설문 조사를 통해 고객의 다양한 요구와 불만 사항을 파악하는 데는 많은 비용과 시간적인 제약이 발생하고 있다. 온라인 쇼핑몰의 고객 후기 데이터는 제품에 대한 고객들의 감성을 분석할 수 있는 이상적인 자료를 제공하고 있다. 본 연구에서는 삼성과 애플 스마폰에 대한 감성분석을 위해 아마존 쇼핑몰로부터 고객 리뷰 데이터를 수집하였다. 선행 연구에서 대표적인 감성분석 기법으로 사용된 5가지 분류 알고리즘을 적용하였다. 5가지 분류알고리즘은 support vector machines, bagging, random forest, classification or regression tree, maximum entropy 등이다. 본 연구에서는 분류 알고리즘의 수행도를 종합적으로 평가할 수 있는 entropy score를 제안하였다. Entropy score를 이용하여 5가지 알고리즘을 평가한 결과에 따르면 support vector machines 알고리즘의 entropy score가 가장 높은 것으로 분석되었다.

ABSTRACT

Online customer evaluations and social media information among a variety of information sources are critical for businesses as it influences the customer's decision making. There are limitations on the time and money that the survey will ask to identify a variety of customers' needs and complaints. The customer review data at online shopping malls provide the ideal data sources for analyzing customer sentiment about their products. In this study, we collected product reviews data on the smartphone of Samsung and Apple from Amazon. We applied five classification algorithms which are used as representative sentiment analysis techniques in previous studies. The five algorithms are based on support vector machines, bagging, random forest, classification or regression tree and maximum entropy. In this study, we proposed entropy score which can comprehensively evaluate the performance of classification algorithm. As a result of evaluating five algorithms using an entropy score, the SVMs algorithm's entropy score was ranked highest.

키워드 : 엔트로피 점수, 감성분석, 분류알고리즘, 수행도 평가

Key word : Entropy score, sentiment analysis, classification algorithm, performance evaluation

Received 2 April 2018, **Revised** 20 April 2018, **Accepted** 5 May 2018

*** Corresponding Author** Man-Hee Park(E-mail:mhpark@cup.ac.kr, Tel:+82-51-510-0875)
Department of Business Administration, Catholic University of Pusan, Busan 46252, Korea

Open Access http://doi.org/10.6109/jkiice.2018.22.9.1153 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. INTRODUCTION

Sentiment Analysis is a technique for analyzing subjective data such as attitudes, opinions, and tendencies of people in text. Social media data is raw data that consumers express voluntarily, so it is a marketing indicator that can best grasp consumer's opinion[1]. Sentiment classification techniques can provide critical information to business decision makers because they better understand the customer's feelings, opinions, and satisfaction. In this study, we collected customer review data written for a particular smart phone from Amazon to analyze the sentiment of the product. Five typical classification algorithms were used to categorize the sentiment of customers and to compare classification performance. There are various evaluation scales such as accuracy, precision and recall as the performance evaluation scale of the classification algorithm, but there is no evaluation scale that can comprehensively evaluate these evaluation scales. Therefore, in this study, we proposed entropy score which can comprehensively evaluate various evaluation scales by applying shannon's entropy concept. Based on the entropy score proposed in this study, we evaluated the entropy result of five classification algorithms.

Ⅱ. SYSTEM MODEL AND METHODS

In this study, we used the research framework as shown in Fig. 1 to conduct sentiment analysis of product review data created by customers who purchased smartphone at Amazon shopping mall.

We performed web crawling to collect product's review data from Amazon shopping mall. We have taken a special care in extracting the data from web pages smallest necessary data is extracted for processing. The list of items that we have extracted is review title, user review, rating and date of review. The data extracted need to be cleaned so that we get proper text review on which analysis can be performed. The product review's polarity is the rating the user provides for that review. The positive reviews are those with rating 5 stars and 4 stars. Neutral reviews are those with rating 3 stars and negative reviews are those with rating 2 stars and 1 star. The review data was classified into test set and training set, and five classification algorithms were applied to categorize sentiment. The entropy score was calculated by using the entropy score which can take into account various evaluation scales and the classification performance of the individual algorithms was evaluated.

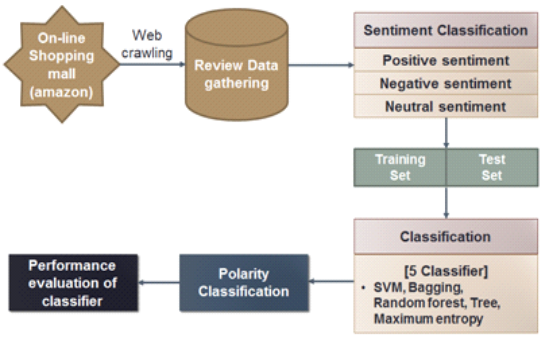

Fig. 1 Research framework

The outline of the five algorithms used in sentiment classification in this study is as follows.

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest[2]. The strength of random forests method is that it is useful for detecting outlier, identifying important factors, or discovering patterns in data.

Maximum Entropy Classifier always tries to maximize the entropy of the system by estimating the conditional distribution of the class label[3]. This method provides a probability density function that maximizes the uncertainty based on what information is given, so that the uncertainty can be estimated quantitatively.

Decision tree induction has been extensively used in many areas such as statistics and machine learning for

the purposes of classification and prediction[4]. Decision trees can be used in any case for classification or prediction purposes, but they are more useful when an description of the process is needed rather than the accuracy of the analysis.

Support Vector Machines are pattern classifiers that can be expressed in the form of hyper-planes to discriminate positive instances from negative instances[5]. The classification and regression technique using SVM is widely used in various fields because it has a merit that it can perform classification work with a relatively small amount of learning data quickly.

Bootstrap aggregation, or bagging, is a technique proposed by Breiman[6] that can be used with many classification methods and regression methods to reduce the variance associated with prediction, and thereby improve the prediction process[7]. It is a method to improve prediction ability by reducing variance of prediction model through sampling with replacement from original data. Bagging is generally suitable for over-fit models, small-sized, large-dispersion models.

A confusion matrix[8] contains information about actual and predicted classifications done by a classification system. Table 1 shows the confusion matrix for a two class classifier.

Table. 1 Confusion matrix[8]

Data class	Predicted Positives	Predicted Negatives
Actual Positives	true positive (tp)	false negative (fn)
Actual Negatives	false positive (fp)	true negative (tn)

The performance of sentiment classification can be evaluated by using four scales calculated as the following Table 2.

Table. 2 Evaluation scales[8]

Scale	Formula	Evaluation focus
Accuracy	(tp+tn)/ (tp+fn+fp+tn)	Overall effectiveness of a classifier
Precision	(tp)/ (tp+fp)	Class agreement of the data labels with the positive labels given by the classifier
Recall (Sensitivity)	(tp)/ (tp+fn)	Effectiveness of a classifier to identify positive labels
F-measure	2*(Precision* Recall)/ (Precision+ Recall)	Relations between data's positive labels and those given by a classifier (harmonic mean of precision and recall)

There are different evaluation scales to assess the classification results after applying various classification algorithms. However, there are no scales to comprehensively evaluate various evaluation scales. Therefore, in this study, an entropy based evaluation scales that can comprehensively evaluate various evaluation scales was presented and evaluation results were evaluated using empirical data.

The process of calculating the entropy score based on various evaluation scales is as follows.

Assume that n classification algorithms and evaluation scales are K, and let S_{jk} be the evaluation scale score matrix for each classification algorithm. The vector consists of n row vectors A_j and K column.

Step 1: The evaluation score matrix(S_{jk}) is calculated based on the evaluation algorithm and the evaluation scales.

Step 2: Normalize the score matrix S_{jk} and set $e_{jk} = \frac{S_{jk}}{\sum_{j=1}^{n} S_{jk}}$, $k = 1,2,\cdots,K$.

Step 3: Calculate entropy E_k[9] as

$$E_k = -(\ln n)^{-1} \sum_{j=1}^{n} e_{jk} \ln(e_{jk}),\ k = 1,2,\cdots,K.$$

Step 4: Compute the diversity level as $D_k = 1 - E_k$, $k = 1,2,\cdots,K$.

Step 5: Calculate the normalized value of D_k as

$$W_k = \frac{D_k}{\sum_{k=1}^{K} D_k},\ k = 1,2,\cdots,K,\ \text{Such that}\ \sum_{k=1}^{K} W_k = 1.$$

Step 6: Compute the entropy score as

$$ES_j = \sum_{k=1}^{K} W_k S_{jk},\ j = 1,2,\cdots,n.$$

The process of calculating the entropy score when the evaluation result by algorithm is same as the table 3 is as follows.

Table. 3 Evaluation score

Algorithms	Accuracy	Precision	Recall
SVM	0.8441	0.84	0.97
Bagging	0.8389	0.86	0.94
Random Forest	0.8354	0.84	0.97
Tree	0.8214	0.85	0.92
MaxEntropy	0.8546	0.89	0.92

Table 4 shows the result of the normalization of the score matrix (S_{jk}).

Table. 4 Normalized score

e_{jk}	Accuracy	Precision	Recall
SVM	0.2012	0.1963	0.2055
Bagging	0.2000	0.2009	0.1992
Random Forest	0.1992	0.1963	0.2055
Tree	0.1958	0.1986	0.1949
MaxEntropy	0.2037	0.2079	0.1949

Based on table 4, entropy and diversity level are calculated and the normalized value (W_k) is calculated as shown in table 5 below.

Table. 5 Result of various parameter

Parameter	Accuracy	Precision	Recall
E_k	0.9999	0.9999	0.9998
D_k	0.0001	0.0001	0.0002
W_k	0.1404	0.3886	0.4710

Based on Table 5, the entropy score is calculated as shown in Table 6.

Table. 6 Entropy score

Algorithms	Entropy score	Rank
SVM	0.9018	1
Bagging	0.8947	4
Random Forest	0.9006	2
Tree	0.8790	5
MaxEntropy	0.8992	3

Ⅲ. RESULTS

In this study, we collected 2,400 product reviews and star rating of customers' data about Samsung Galaxy S7 and Apple iPhone 7 from the Amazon online site using R. Five classification algorithms were applied for collected data. The five algorithms are based on support vector machines[10], bagging[11], random forest[12], classification or regression tree[13][14] and maximum entropy[15].

Based on the star rating of the collected data, the sentiment classification is classified as positive if 4 or 5 points, neutral if 3 points, and negative if 1 or 2 points. The frequency and the ratio of each sentiment frequency by product model are shown in Table 7. The distribution of the star rating evaluated by the customers are shown in Fig. 2.

Table. 7 Class distribution

Sentiment	Galaxy S7		iphone 7	
	Frequency	Frequency Ratio	Frequency	Frequency Ratio
negative	496	27.4%	143	24.1%
positive	1241	68.7%	428	72.2%
neutral	70	3.9%	22	3.7%
Total	1,807	100%	593	100%

Accuracy, precision and recall were selected as evaluation scales to evaluate the analysis results of the five classification algorithms considered in this study. Table 8 shows the entropy scores that can be evaluated by considering the three evaluation scales together.

According to the analysis results shown in Fig 3., the entropy score of the SVM algorithm is the highest and the classification performance is the best. The results of this study are consistent with previous studies[16] that SVM has high efficiency and high accuracy.

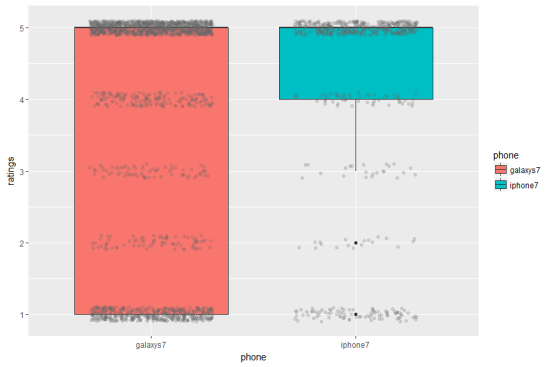

Fig. 2 Distribution of star rating

Table. 8 Entropy score

Algorithms	Accuracy	Precision	Recall	Entropy Score
SVM	0.8441	0.84	0.97	0.9018
Bagging	0.8389	0.86	0.94	0.8947
Random Forest	0.8354	0.84	0.97	0.9006
Tree	0.8214	0.85	0.92	0.8790
MaxEntropy	0.8546	0.89	0.92	0.8992

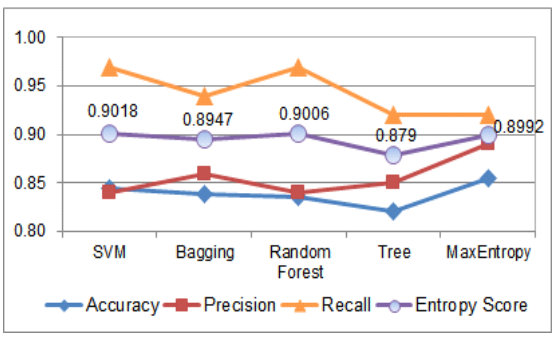

Fig. 3 Result of analysis

Table. 9 Correlation analysis between evaluation scales

	Accuracy	Precision	Recall	Entropy score
Accuracy	1	-	-	-
Precision	0.6061	1	-	-
Recall	0.0716	-0.7301	1	-
Entropy score	0.7931	0.0504	0.6449	1

Table 9 shows the correlation coefficient between the three evaluation scales and the entropy score used in the performance evaluation of the classification algorithm. The results showed that entropy score had a strong positive correlation with the accuracy scale (0.7931) and a significant positive correlation with the recall scale (0.6449).

Ⅳ. CONCLUSION

In this study, the entropy score which can be applied to the evaluation of classification algorithm performance of sentiment analysis is proposed and empirical analysis is performed. To evaluate the classification result of classification algorithms, five classification algorithms were selected as the representative algorithms in previous studies. Three scales of accuracy, precision and recall were selected for the performance assessment of the classification algorithm. We collected 2,400 smartphone customer review data from amazon.com for empirical analysis.

In this study, entropy scores of support vector machines, bagging, random forest, classification tree, and maximum entropy algorithm for accuracy, precision, and recall were calculated as 0.9018, 0.8947, 0.9006, 0.8790, 0.8992. According to the results of applying the entropy score proposed, the entropy score of the support vector machines algorithm is the highest. Also, according to the correlation analysis results, the entropy score has a strong positive correlation with the accuracy scale and has a significant positive correlation with the recall scale. In this paper, we present a methodology for

evaluating various evaluation scales of classification algorithms and demonstrate their applicability through empirical analysis. The sentiment analysis of the customer review data can save a lot of cost and time if the requirements and improvements to the product or service are applied to the improvement of existing products or the development of new products.

ACKNOWLEDGEMENT

This study was supported by 2016 research fund of catholic university of pusan.

REFERENCES

[1] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2015.

[2] B. Gregorutti, B. Michel and P. Saint-Pierre, "Correlation and variable importance in random forests," *Statistics and Computing*, vol. 27, no. 3, pp. 659 - 678, Apr. 2017.

[3] V. A. Kharde and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," *International Journal of Computer Applications*, vol. 139, no. 11, pp. 5-15, Apr. 2016.

[4] G. Vinodhini and RM. Chandrasekaran, "Performance Evaluation of Machine Learning Classifiers in Sentiment Mining," *International Journal of Computer Trends and Technology*, vol. 4, no. 6, pp. 1783-1786, Jun. 2013.

[5] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21-45, Sep. 2006.

[6] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, Jan. 1996.

[7] C. D. Sutton, "Classification and Regression Trees, Bagging, and Boosting," *Handbook of Statistics*, vol. 24, pp. 303-329, Apr. 2005.

[8] R. Kohavi and F. Provost, "Glossary of terms. Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process," *Machine Learning*, vol. 30, pp. 271-274, Feb. 1998.

[9] Q. Xie, Q. Dai, Y. Li and A. Jiang, "Increasing the Discriminatory Power of DEA Using Shannon's Entropy," *Entropy*, vol. 16. pp. 1571-1585, Mar. 2014.

[10] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071) [Internet]. Available: http://CRAN.R-project.org/package=e1071.

[11] A. Peters and T. Hothorn, ipred: Improved Predictors [Internet]. Available: http://CRAN.R-project.org/package=ipred.

[12] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18-22, Dec. 2002.

[13] B. Ripley, tree: Classification and Regression Trees [Internet]. Available: http://CRAN.R-project.org/package=tree.

[14] T. P. Jurka, L. Collingwood, A. E. Boydstun, Grossman, and W. E. Atteveldt, "RTextTools: A Supervised Learning Package for Text Classification," *The R Journal*, vol. 5, no. 1, pp. 6-12, Jun. 2013.

[15] T. P. Jurka, "maxent: An R package for low-memory multinomial logistic regression with support for semi-automated text classification," *The R Journal*, vol. 4, no. 1, pp. 56-59, Jun. 2012.

[16] Y. Wan and Q. Gao, "An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis," in *Proceedings of 15th IEEE International Conference on Data Mining Workshop*, pp. 1318-1325, 2015.

박만희(Man-Hee Park)

2002년 성균관대학교 산업공학과(공학박사)
2003년~2004년 KAIST 연구교수
2006년~2010년 KISTI 선임연구원
2013년~현재 부산가톨릭대학교 경영학과
※관심분야 : 품질경영, 데이터마이닝, 빅데이터, 계량정보분석