SENTIMENT SCORES OF SENTIMENT KEYWORDS: ANALYSIS OF HOTEL REVIEW DATA

Ye-Seul Jeon, Eun-Young Jang and Hwan-Seok Jang

*Research Engineer in Daumsoft, Graduate school of applied statistics, Yonsei-University
97, Dokseodang-ro, Yongsan-gu, Seoul, South Korea*

ABSTRACT

Utilizing hotel review data by analyzing specific keywords as either being positive or negative, it is possible to aid consumers in selecting a suitable hotel. For this purpose, we propose creating a Sentiment Score of each Sentiment Keyword. We first construct a Sentiment Lexicon by extracting Sentiment Keywords by natural language processing. Afterward, we formulate a coefficient between the Sentiment Keyword and the evaluation score through ridge regression. The coefficient is then normalized to arrive at the cumulative probability of normal distribution, ultimately employed as the Sentiment Score of the specific Sentiment Keyword. However, if the coefficient is considered to be insignificant, the Sentiment Score is replaced by the probability of the Sentiment Keyword as either being positive or negative by using label propagation with the k -Nearest Neighbor (k -NN) classifier. The data analyzed is focused on South Korean hotel reservation websites.

KEYWORDS

Sentiment Analysis, Ridge Regression, Label Propagation, Review Analysis, k -Nearest Neighbor

1. INTRODUCTION

In South Korea, many travelers utilize hotel reservation websites. Hotel reservation websites aid consumers in selecting the most suitable hotel by providing in-depth information such as their reviews. However, its shortcoming is that it requires users to spend excessive time reading the actual reviews to get a better idea of the hotel for a qualitative perspective. Previous studies endeavored to solve this issue by categorizing reviews as either positive, neutral, or negative (Kasper, W. and Vela, M., 2011). However, categorizing a single review to a single category may impair the specific ideas posed in the review. For this purpose, our study creates a Sentiment Score of each Sentiment Keyword that consumers deem important, thus complementing the conventional evaluation score.

2. METHODOLOGY

Figure 1. A framework of the proposed methodology

After collecting reviews from Korean hotel reservation websites, Sentiment Keyword candidates are extracted through natural language processing. Employing Word2Vec, a Sentiment Lexicon is then created by calculating the distance between each keyword by cosine similarity.

Figure 2. A process of calculating sentiment score

Figure 2. shows the process of extracting sentiment scores. We implemented ridge regression as the machine learning model to derive the coefficient between Sentiment Keywords (independent variable) and its evaluation score (dependent variable). The ridge regression shows better prediction when the predictor variables are highly correlated and it is stable to select variables (Donald W., 1975). We then normalize the coefficient of each positive and negative keyword. We equate the cumulative probability of each coefficient as its Sentiment Score for both positive and negative keyword categories, with the cumulative probability for the negative keyword category stated as $(-1) \times (1 - \text{Cumulative Probability})$. Meanwhile, in cases where the coefficient is considered as insignificant (i.e. Positive Keyword Category: Cumulative Probability < 40%, Negative Keyword Category: Cumulative Probability > 60%), the Sentiment Score of a Sentiment Keyword is substituted by the probability of the Sentiment Keyword as either being positive or negative. This is achieved through a semi-supervised method as according to k -NN($k=5$) using label propagation after constructing a WordNet based on the Sentiment Lexicon (Rao D., 2009). In this study, there are two labels that one is positive and the other is negative. To do label propagation, insignificant keywords are unlabeled while significant keywords are labeled as either positive or negative. As a result, insignificant keywords are labeled according to their similar meanings of the keyword's label.

3. RESULTS

We collected approximately one million reviews of more than forty thousand hotels from five different websites, including Airbnb, Booking.com, Daily Hotels, and etc. A Sentiment Lexicon was constructed with approximately thirty-nine thousand Sentiment Keywords using Word2Vec. Through the Sentiment Lexicon, we analyzed a myriad of review data posted in Booking.com, containing approximately eleven thousand Sentiment Keywords. We conduct ridge regression by assigning $\lambda = 10$ as a penalty. To calculate the accuracy of polarity resulted from ridge regression, we first assign keywords with a positive/negative coefficient as a positive/negative keyword (i.e. predict sentiment) after which we compare the accuracy between the predict sentiment and the actual sentiment from the lexicon. We conducted ridge regression and label propagation with the accuracy as shown in Table 1. Ridge regression resulted in an accuracy of 65.42% and complemented the model with label propagation, thus improving the overall level of accuracy to 81.44%. The trained Word2Vec for Label Propagation expresses each word in 256 dimensions and the size of vocabulary is 542,841.

Table 1. Model Accuracy

Ridge Regression	Ridge Regression + Label Propagation
65.42%	81.44% (Final Model)

Table 2. shows an example of Sentiment Keyword coefficients calculated by ridge regression and its cumulative probability after normalizing and the results for insignificant Sentiment Keywords using label propagation.

Table 2. Sentiment keywords and sentiment score

Polarity based on Lexicon	Sentiment Keyword	Ridge Regression Coefficient	Cumulative Probability	Label Propagation	Final Sentiment Score
Positive	Awesome View	0.0230	0.5092	Labeled	0.5092
Negative	Bad Service	-2.2054	0.0137	Labeled	-(1-0.0137) = -0.9863
Positive	Clean Bedsheets	-0.5367	Unlabeled	0.9167	0.9167
Negative	Unfriendly Staff	0.5942	Unlabeled	0.8989	-(0.8989)

Table 3. shows that Sentiment Scores of each Sentiment Keyword complement evaluation scores, thus aiding consumers to better understand the satisfaction level of the hotel. For instance, in cases where low/high evaluation scores were given despite being an overall positive/negative review, Sentiment Scores point out that the review also included a positive/negative opinion.

Table 3. Sentiment score of a sample review

Review	Sentiment Score	Evaluation Score
It had a small room. The staff was friendly.	Small Room -0.58 /Staff Friendly 0.64	2.5
... The bathroom had warm water.	Clean Room 0.61 /Quiet 0.56 /Warm Water 0.62	
The breakfast was awful.	Awful -0.77 / Clean 0.59	9.6
... The room was clean and okay.		

4. CONCLUSION & FUTURE WORK

Sentiment Scores are an effective quantifiable measurement tool that helps us grasp the essential meaning of specific Sentiment Keywords. We believe that Sentiment Scores can be a useful tool for consumers in selecting a suitable hotel that meets their needs and for hotel management in terms of selecting the right marketing strategy in ways to react to the right problems that hotel user face.

A similar study has used the product parse-paraphrase paradigm to calculate emotional scores. In this study, emotional scores were calculated using the frequency-based ratings of adjectives and adverbs in sentences (Liu, J., 2009). However, since the method using simple appearance frequency each sentiment keyword is affected by the frequency. It is unstable to calculate the exact score of sentiment keywords. In comparison, the method proposed in this study has the advantage that the emotion score can be calculated irrespective of the frequency with which the word appears in the review.

However, the methodology of Label Propagation that is used in this study also has limitations. It costs exhaustive computation to deal with such newly input samples. In order to overcome this limitation, it seems possible to continue the study to reduce the computational cost using Logistic Label Propagation (Kobayashi, T., 2012). Logistic Label Propagation is powerful when there are a lot of keywords or data to compute.

ACKNOWLEDGMENT

This research is supported by the Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency(KOCCA) in 2019 program of Integrated and Convergence Service for Supporting Tourism Development.

REFERENCES

- Kasper, W. and Vela, M., 2011.Sentiment Analysis for Hotel Reviews. Proceedings of the Computational Linguistics-Applications Conference, Jachrenka, Poland, pp. 45-52
- Kobayashi, T., Watanabe, K., & Otsu, N. 2012. Logistic label propagation. *Pattern Recognition Letters*, Vol. 33, No. 5, pp. 580-588
- Liu, J., & Seneff, S. 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Vol. 1, pp. 161-169
- Marquardt, D. W., & Snee, R. D. 1975. Ridge regression in practice. *The American Statistician*, Vol. 29, No. 3, pp. 3-20.
- Rao D. and Ravichandran D., 2009. Semi-supervised polarity lexicon induction, *12th Conference of the European Chapter of the Association for Computational Linguistics*, Athens, Greece, pp. 675-682