Evaluating Attitude Shift From Tourism Online Reviews Using Word Embedding-Based Approaches

Binyan Chen[1, a,] Yishi Zhang[2, b,] Wenjie Chen[3, c]
[a] email: 295289@whut.edu.cn, [b] email: yszhang@whut.edu.cn, [c] email: 295303@whut.edu.cn

Wuhan University of Technology, Wuhan, China

Abstract-The shift of consumers' attitudes caused by public health emergencies can be reflected in online reviews. This study applies the word embedding-based methods and sentiment propensity analysis to mine tourism online reviews. Specifically, we obtain the comprehensive vector of each review, which consists of the word vectors in the review word embedding, and evaluate the changes in tourists' emotional tendencies in the context of public health emergencies (taking COVID-19 as an example). The empirical results verify the effectiveness of the word embedding-based methods in identifying attitude shift and confirm the divergent effect of emotional valence in predicting the usefulness of tourism online reviews.

Keywords: online reviews; review usefulness; sentiment analysis; word embedding; deep learning

1 INTRODUCTION

With the rapid development of information technology, online review data provides a vast amount of product and service information that has a profound impact on consumers' attitudes and purchasing behavior, as well as business outcomes for companies. Numerous studies have confirmed that online reviews significantly influence consumers' purchase decisions and service providers' online sales. For example, based on secondary data of hotels in Ctrip.com, Ye et al. earlier confirmed the contribution of online hotel ratings to their online sales.[2] Subsequent studies have also successively verified online reviews' importance in terms of both their number and sentiment.[3-5]

However, despite the convenience of online reviews, it still suffers from information overload[1] Therefore, customers need to judge the usefulness of information from the vast amount of information and achieve reasonable screening. Therefore, customers need to find useful information and achieve reasonable screening in the context of massive information. The concept of "usefulness of reviews" is relatively consistent in the current academic community, i.e., reviews that are recognized as valuable by consumers are considered useful.[7] Much of the research on the usefulness of reviews in recent years has focused on the measurement and influencing factors. The main existing measurement methods include the number of useful votes for comments and their percentage of the total number of votes, as well as the Likert scale and so on. The domestic and international studies on the influencing factors of reviews' usefulness can be roughly divided into two parts: review content characteristics and review form characteristics. Overall, although current research on the usefulness of reviews is relatively mature, consideration of affective features and their interactions is lacking, and little

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2327170

consideration has been given to the new context of public health emergencies. Public health emergencies have an important impact on the physical and mental health of the population and the mood of society because of their suddenness, unknowability, and harmfulness. Negative emotions such as anxiety and sadness are often present along with positive emotions[8] Existing research shows that the more negative the sentiment, the higher the usefulness of the review.[6] It is worth exploring whether public sentiment has shifted in the face of the new context of public health emergencies and is thus reflected in online commenting behavior. Based on the travel online review data extracted from the Ctrip website, this study combined text mining and word vector techniques to explore the sentiment tendency of users' online comments before and after the epidemic. We also cross-analyzed the potential changes between visitors' affective tendencies and review usefulness from the perspective of review usefulness to provide visitors, businesses, and governments with more practical decision support.

2 MATERIALS AND METHODS

2.1 Data Source and Data Pre-processing

Taking the timeliness of the reviews and the geographical nature of tourist attractions into account, this paper obtained the tourism data of scenic spots in Hunan Province and Hubei Province from January 1, 2019, to December 31, 2020, on Ctrip.com, which includes username, amount of user likes, user comment text, amount r of images, and comment time. Firstly, this paper carried out a simple pre-processing of the data, eliminating invalid comments and finally retaining 11743 travel data. In addition, we supplemented the split word lexicon and deactivation word document to analyze the review text, remove deactivation words, and use the time of Wuhan city closure as the dividing line to start the analysis.

2.2 Usefulness of Reviews

Evaluating the usefulness of reviews through a singularized number of comment likes or treating perceived usefulness as reviews' usefulness has certain limitations. Therefore, this paper uses a weighted similarity algorithm based on Word2vec to obtain the usefulness of user reviews. After data cleaning, the corpus and the word vector model are first built by Word2vec:

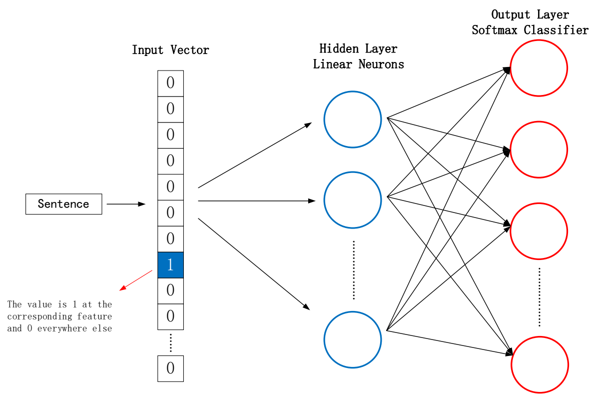

Figure1 skip-gram model

The output layer converts the result into a probability distribution by softmax, and the word vector can be obtained by using the maximum likelihood estimation and gradient descent. The probability distribution can be expressed as follows:

$$P(\omega_o|\omega_i) = \frac{exp(\mu_o^T v_i)}{\Sigma_{k\in V} exp(\mu_k^T v_i)} \quad , \tag{1}$$

where v_i denotes word vector for target words, μ_o denotes the word vector for the oth word in addition to the target word and V denotes the number of words.

And then the sentence vector is synthesized into the comment vector by averaging word vector:

$$sen\_vec_j = \frac{\sum_{i=1}^{m} v_i}{m} \quad , \tag{2}$$

where m denotes the number of words in the jth sentence.

The text similarity between each comment and all useful comments is calculated by the cosine similarity formula:

$$cos\,\theta = \frac{A \cdot B}{|A| \cdot |B|} = \frac{\sum_{n}^{i=1}(A_i * B_i)}{\sqrt{\sum_{n}^{i=1}(A_i)^2} * \sqrt{\sum_{n}^{i=1}(B_i)^2}} \quad , \tag{3}$$

where $\cos\theta$ denotes text similarity, A denotes the sentence vector for sentence A and B denotes the sentence vector for sentence B.

Finally, the similarity is normalized to the interval of [0,1], while the amount of votes of useful comments is used as the weight for weighting, and the usefulness score of each comment is obtained.

2.3 Analysis of Emotional Disposition

To tap into the changes in tourists' affective tendencies before and after the epidemic, this paper further explores the correlation between tourists' tendency scores and reviews' usefulness. Based on the observation of the data, it was found that there were comments with a negative sentiment but a rating of 5 as well as comments with a positive sentiment but a rating of 0, indicating that there is some difference in sentiment and rating. At the same time, because both positive and negative sentiments exist in a single review, it is difficult to accurately measure the overall sentiment of the reviewer through other numerical rating methods that reflect the review on a single scale. Therefore, this paper trains the sentiment propensity analysis model through Baidu AI open platform, and the training dataset includes 4426 negative sentiment data as well as 4781 positive sentiment data. The dataset was then tested by calling the API interface via python to derive the sentiment classification results, the confidence level of the classification, the positive category probability, and the negative category probability of the user comments.

2.4 Linear Regression

To examine the changes in the relationship between affective tendency scores and comment usefulness before and after the epidemic, and control those unrelated variables affecting comment usefulness at the same time, this paper conducted linear regressions on the processed data through SPSS and established the following linear regression equations.

$$Usefulness = \beta_0 + \beta_1 Length + \beta_2 Time + \beta_3 Semantic + \beta_4 Similarity + \varepsilon \quad (4)$$

3 RESULTS & DISCUSSION

3.1 Calculating Eigenvalues

Therefore, in this study, positive and negative category probabilities were obtained through natural language processing to measure the sentiment tendency scores of the reviews. The sentiment classification results, classification confidence, positive category probability, and negative category probability for each comment were derived by model training and calling the model, as shown in Table 1. The results showed that the pre-epidemic sentiment analysis predicted negative evaluations (1328 items) accounted for 21% and the post-epidemic sentiment analysis predicted negative evaluations (900 items) accounted for 14%, which shows that the overall sentiment of visitors tended to be positive after the epidemic.

Table 1 Results of emotional disposition scores

Content	Emotional classification results	The confidence level of classification	Positive Category Probability	Negative category probability
It wasn't as good as I thought it would be.	Negative	0.995049	0.002228	0.997772
The hotel's capacity is limited, but overall it is still very good！	Positive	0.995843	0.998129	0.00187068
Red Cliff Ancient Battlefield Scenic Area is the least worthwhile scenic spot I've ever been to.	Negative	0.998808	0.000536453	0.999464

Besides, the number of likes for comments is mostly 0 or 1, and the perceived usefulness does not fully reflect the usefulness of the comments. Therefore, this paper obtained the comment usefulness score by sentence vector and cosine similarity, and the comment usefulness score is between 0 and 1. Before the epidemic, The highest review usefulness score was 0.57, the lowest score was 0.47, and the average score was 0.52, with 76.94% of reviews scoring between 0.5 and 0.6 in usefulness; After the epidemic, the highest review usefulness score was 0.55, the lowest score was 0.48, and the average score was 0.51, 77.42% of the review usefulness scores were between 0.5 and 0.6.

3.2 Regression Results

This study used SPSS for linear regression. Model 1 and model 3 were constructed and control variables were added to them. Then the independent variable affective propensity score was added into model 2 and Model 4, The analysis results are shown in Table 2 and Table 3.

Table 2 Regression results Before Wuhan lockdown

	Variable	**Dependent variable: Usefulness of comments**	
		Model 1	***Model 2***
Control variables	Time	-0.001***	-0.003***
	Length	0.050	0.050
	Similarity	0.953***	0.951***
Independent variable	Semantic		-0.060***
	R^2	0.927	0.931
	Adjusted R^2	0.927	0.931

Note: *** means significant at 0.01 level

Table 3 Regression results after Wuhan lockdown

	Variable	**Dependent variable: Usefulness of comments**	
		Model 3	***Model 4***
Control variables	Time	0.126***	0.119***
	Length	0.163***	0.165***
	Similarity	0.751***	0.741***
Independent variable	Semantic		0.048***
	R^2	0.564	0.567
	Adjusted R^2	0.564	0.566

Note: *** means significant at 0.01 level

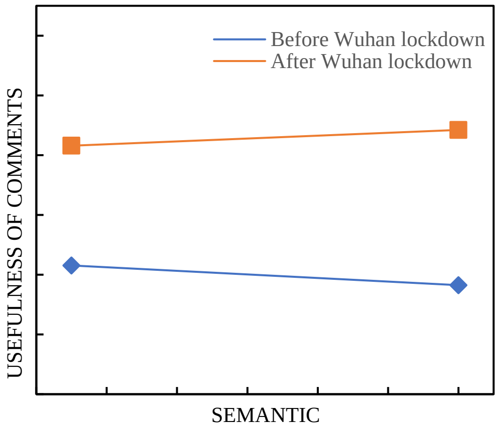

Figure 2 Moderating effect of the epidemic

As can be seen from Table 2, the usefulness of comments is negatively correlated with the affective tendency score (β=-0.060), that is, the more negative emotions are, the higher the usefulness of evaluation is. As can be seen from Table 3, the evaluation usefulness score is positively correlated with the affective tendency score (β=0.048), that is, the more positive the emotion is, the higher the evaluation usefulness is. And Figure 2 shows the moderating effect of the epidemic.

4 CONCLUSION & SUGGESTIONS

4.1 Research Conclusion

This study analyzed the comment usefulness as well as the sentiment tendency scores and found that the higher the probability of negative sentiment tendency of user comments, the higher the usefulness. But after the COVID-19, this phenomenon changed, the higher the probability of positive sentiment tendency of user comments, the higher the usefulness.

This is explained in this study as follows:

I) Shortly after a public health emergency, people's feelings of uselessness increased significantly, and negative information is more likely to elicit high-risk perceptions in individuals with high feelings of uselessness. Tourists crave positive attitudes when searching for travel information, so emotionally positive reviews can reduce the level of individual risk perceptions and cater to tourists' emotional tendencies.

II) As public health emergencies are mitigated, the national pride and self-confidence of the public is increased, as well as the personal happiness of tourists. And the attitude of tourists is more positive, so they will also make and support more positive statements.

III) Meanwhile, Wuhan is rewarded as a hero city, and under the effect of belief bias, tourists are more inclined to accept positive views that are consistent with a priori knowledge and refuse to refute them. And under the effect of spillover, tourists have more positive a priori attitudes toward Hubei Province.

4.2 Research Significance

The study has certain theoretical and practical significance. This study introduces public health emergencies as a background to analyze visitor sentiment and review usefulness in special situations, and to identify differentiation of review sentiment tendencies timely. For relevant researchers, in the process of analyzing the usefulness of reviews, they need to pay attention to the changes in the external environment, combine the changes of tourists, viewers, scenic spots, and online tourism platforms in the new context, and explore them based on a comprehensive consideration of the influence of multiple factors. The combination of social context and data results is what makes data analysis informative and relevant. In the context of the Internet era, traveler emotions not only directly affect traveler satisfaction and loyalty, but can also infect subsequent travelers by way of electronic word-of-mouth, which in turn has a direct impact on their decisions. Therefore, many brands have also started to attract customers' attention by way of emotion marketing, such as creating a brand culture with specific emotions. The findings of this paper can help scenic spot managers gain insight into the correlation between online ratings and tourists' sentimental tendencies, and can provide management suggestions for online tourism platforms and scenic spots. Online tourism platforms should recommend reviews with more positive emotional attitudes and display more emotionally positive prompt words when visitors write reviews. Because different tourists focus on different priorities, scenic spots should analyze reviews in a targeted manner, seeing both emotionally positive and emotionally negative reviews, as a way to better improve scenic spot management, enhance service quality and gain sustainable development.

4.3 Research Gaps and Outlook

Although this paper combines a variety of research methods to analyze the relationship between tourists' affective tendencies and review usefulness in the context of public health emergencies and achieves certain results, there are still some shortcomings that we hope can be improved in future research.

In this paper, when measuring comment usefulness, the sentence vectors were synthesized by averaging word vectors, ignoring the weights of different words, and neglecting the words that are decisive for classification, resulting in a certain error between the final obtained comment vectors and the real situation, which affects the accuracy of the research results. In future studies on the usefulness of comments, attention can be paid to the weight differences between different words, and sentence vectors can be obtained using sentence2vec or doc2vec for more accurate measurement.

This study uses the time of Wuhan city closure as the dividing line for public health emergencies, ignoring the lag of user reviews and the lag of the impact of the epidemic on tourists' travel, resulting in the blurring of the social context of public health emergencies, and thus the study's conclusions are not rigorous enough. Future research on online reviews in the context of public health emergencies should take into account the influence of lag and choose the temporal demarcation line more rigorously.

REFERENCES

[1] Frias, D M., Rodriguez, M A. and Castaneda, J A. (2008) Internet vs. travel agencies on pre-visit destination image formation: An information processing view Tourism Management, 29, 163-179

[2] Qiang, Y., Law, R. and Gu, B. (2009) The impact of online user reviews on hotel room sales, International Journal of Hospitality Management, 28 , 180-182.

[3] Yang, Y., Park, S. and Hu, X. (2018) Electronic word of mouth and hotel performance: A meta-analysis, Tourism Management, 67, 248-260.

[4] Zhang, Z., Liang, S., Li, H. and Zhang, Z. (2019) Booking now or later: Do online peer reviews matter? International Journal of Hospitality Management,.77, 147-158.

[5] Zhao X, Wang L, Guo X and Law R (2015) The influence of online reviews to online hotel booking intentions International Journal of Hospitality Management, 27, 1343-1364

[6] Shi D, Wang L and Yi B (2020) Deep data mining of online review usefulness - hotel review data based on TripAdvisor Nan kai Management Review, *23, 66-77*

[7] Wang Z, Li H and Sun R (2016) A study of factors influencing online review usefulness voting - moderating effect based on commodity type Management Review , 28, 143-153

[8] Zhang Y and Ma J (2020) Research on audience sentiment analysis and communication

characteristics of misinformation about sudden public health events Data Analysis and Knowledge Discovery, 4, 45-54