IIAI Open Conference Publication Series
IIAI Letters on Business and Decision Science
Vol. 005, LBDS433
DOI: https://doi.org/10.52731/lbds.v005.433

Mining Online Reviews with Fake Review Detection to Improve Logistics Service Quality

Mu-Chen Chen *, Hsien-Hua Wu *,
Yu-Hsiang Hsiao †

Abstract

This study proposes an analytical framework to identify important information in online reviews and uses the service quality models and relevant literature on logistics service quality to help construct and analyze the logistics service quality topics. First, this study establishes a fake review detection model based on the convolution neural network (CNN) model to remove potential fake reviews. Then, the SERVQUAL and E-S-QUAL models are used to assist in constructing a Latent Dirichlet Allocation (LDA) topic model to identify relevant topics mentioned in the online reviews about logistics service quality. In further data analysis, this study segments the comments, calculates the sentiment scores of the sentences that mention the logistics service quality in the online reviews, and uses the number of helpfulness votes of each online review to calculate the weight of the sentiment score.

Keywords: Online review analytics, fake review detection, topic modeling, logistics service quality.

1 Introduction

Previous studies analyzing online reviews mostly analyzed reviews as a unit to find overall improvement suggestions for product or logistics service topics [1][2], and assumed that all online reviews used for analysis were real reviews [2][3][4]. Relatively few studies mentioned the possible impact of fake reviews and how to accurately remove them. These studies used mathematical models or machine learning methods to identify potential fake review features and further remove fake reviews [5][6][7]. Lee et al. [8] pointed out that online reviews that receive a high number of helpful votes on online platforms are important reviews and can be called highly helpful reviews. However, previous studies mainly aimed to predict the helpfulness of reviews by taking extracted text features in reviews, review readability indicators, or profile features of reviewers as input attributes to train machine learning models to predict the helpfulness of reviews [9][10]. In the past, relatively few studies have considered the helpfulness of reviews as an influence on sentiment analysis results, and amplified the sentiment score of highly helpful reviews. Additionally, relatively few studies explored the logistics service quality topics that companies are competitive in and that need to be improved in terms of logistics service quality constructs. Previous research has also used association rule mining to analyze online reviews and explored the important issues mentioned in the reviews and the word combinations that can be used to

* National Yang Ming Chiao Tung University, Taipei, Taiwan
† National Taipei University, New Taipei, Taiwan

provide advice to management [11][12]. However, relatively few studies have further combined topic models with the construct of logistics service quality to conduct a more in-depth discussion on the topic and find out the reasons behind consumers' specific sentiment polarity towards the topic of logistics service quality.

This study intends to combine logistics service quality, sentiment analysis and topic modeling in text mining, fake review detection, and weighted review helpfulness scores to find out the topic construct of logistics service quality that consumers feel satisfied or dissatisfied with, and analyze the review keyword association rules of this topic to understand the reasons why consumers have positive or negative sentiment towards specific topics, and then draw suggestions for improving logistics service quality and enhancing corporate competitiveness.

2 Method

The method framework of this study is divided into twelve stages as follows. Stage 1 collects consumer online review data and selects the dataset for training and the dataset for testing the fake review detection model. Stage 2 is data preprocessing. Stage 3 identifies fake reviews by using CNN [6]. Stage 4 extracts logistics service quality topics by using LDA [2]. Stage 5 is review sentence segmentation. Stage 6 identifies sentences on logistics service quality. Stage 7 is sentiment analysis. Stage 8 is the weighting of sentiment scores with review helpfulness. Stage 9 merges the sentences with the same topic and discusses the logistics service quality based on the topic. Stage 10 calculates the average sentiment score of each logistics service quality topic. Stage 11 is the diagnosis of logistics service quality by using association rule mining with the Apriori algorithm [13]. Stage 12 proposes suggestions for improving logistics service quality.

In this study, the following three datasets used in previous studies are used to build a fake review detection model: Amazon generated review dataset [14], Fake news dataset [15], and TripAdvisor + Amazon Mechanical Turk dataset [16]. For the fake review detection model architecture, this study adapts the fake review detection model architecture proposed by Hajek et al. [5] to meet the analysis requirements. This study uses a CNN model [5] to classify reviews.

In the review topic analysis stage, this study uses the LDA model [2] to extract topics from reviews and pre-sets a larger number of topics and keywords to be extracted so that topics related to logistics service quality can be found through manual screening later. After the topic extraction is completed, this study uses the SERVQUAL [17] model constructs to explore the service quality of physical logistics, and uses the E-S-QUAL [18] model constructs to explore the service quality of virtual environments. The lexicon is expanded by referring to previous studies that analyzed online reviews to explore logistics service quality [4][19][20] to find out the topics related to logistics service quality in the reviews.

This study uses the Afinn sentiment dictionary Neilsen [21] to calculate the sentiment scores of each sentence related to logistics service quality. Since this study incorporates the impact of review helpfulness on sentiment scores into the analysis, the sentiment scores are weighted based on review helpfulness percentiles. In the final stage of data analysis, this study uses the Apriori algorithm [13] in association rule mining to conduct review word association analysis on the logistics service quality topic in the dataset to understand the reasons why consumers have high or low sentiment scores for logistics service quality.

3 Computational Results

This study uses one of the Amazon consumer online review datasets collected and organized by Ni et al. [22], which includes four datasets: office products, musical instruments, Prime pantry, and beauty. The dataset of office products, including 8004 reviews, is used for analysis in this study. By using the built fake review detection model, 7489 reviews in the Amazon review dataset of office products are identified as non-fake reviews, and 515 reviews are identified as fake reviews, accounting for 6.43%. Online reviews detected as fake reviews by the built model are directly removed and not included in subsequent analysis.

This study applies the LDA model proposed by Blei et al. [23] to conduct the topic modeling analysis. At this stage, the topic modeling module LatentDirichletAllocation in the scikit-learn package in Python (https://scikit-learn.org) is used for analysis. A total of 38 topics related to logistics service quality are found in the dataset of office products, accounting for 12.67% of all topics. The sentiment score of review is weighted by review helpfulness. After determining the weight multiples of review helpfulness, the sentiment scores of sentences in the review dataset are calculated, then the sentiment scores of logistics service quality topic constructs are calculated, and finally, they are organized for subsequent managerial implications discussion. Table 1 presents the average sentiment scores of a part of topic constructs of office products dataset.

Table 1: Average sentiment score of topic construct of office products dataset

Topic No.	Average sentiment score	Construct	Topic No.	Average sentiment score	Construct
230	5.88	Reliability	20	2.07	Responsiveness
167	5.35	Responsiveness	145	2.07	System usefulness
53	4.06	Responsiveness	57	1.86	Tangibles
107	3.71	Reliability	226	1.79	Reliability
130	3.55	Reliability	245	1.74	Responsiveness
179	3.49	Responsiveness	283	1.74	Tangibles
96	3.44	Reliability	155	1.73	Assurance
280	3.41	Responsiveness	1	1.7	Empathy
275	3.36	Empathy	135	1.68	Responsiveness
95	3.09	Responsiveness	113	1.59	Responsiveness
11	2.51	Reliability	100	1.06	Tangibles
29	2.37	Responsiveness	209	0.95	Empathy
252	2.31	Responsiveness	158	0.53	Empathy
140	2.24	Reliability	48	-0.98	Service personnel

This study divides the logistics service quality topics into "competitive advantage topic", "maintaining quality topic", and "immediate improvement topic" based on the sentiment scores. If the average sentiment score of a logistics service quality topic is higher than 5 points, the topic is classified as a competitive advantage topic and can be used as a reference for the company to maintain its competitive advantage in the future. If the average sentiment score of a topic is between 0 and 5 points, the topic is classified as a maintaining quality topic, and the company can continue to maintain the service level of this logistics service quality topic, and it can also be included in the medium- and long-term improvement plan. If the average sentiment score of a topic is lower than 0 points, the topic is classified as an immediate improvement topic, and the company needs to improve the service quality of the topic as soon as possible.

According to the average sentiment scores of the topics in the office products dataset, topic 230, belonging to the reliability dimension, and topic 167, belonging to the responsiveness dimension, both have sentiment scores higher than 5 points, belonging to the competitive advantage topic. Topic 48, belonging to the service personnel dimension, has a sentiment score lower than 0 points, belonging to the immediate improvement topic. The remaining topics belong to the maintaining quality topic.

This study uses apriori and association_rules packages in mlxtend (https://rasbt.github.io/mlxtend) for association rule mining. A total of 16 topics are found in the office products dataset, having logistics service quality words with relatively high association. Take topic 230 as an example, it includes “great”, “excellent” and “high”, and the consequent word is “quality”. The corresponding sentiment score of topic 230 is 5.88 points, which belongs to the competitive advantage topic. It can be seen that consumers think that the logistics service provided by Amazon’s office products is of good quality.

4 Conclusions

The research objectives of this study are to analyze online reviews to identify topics of logistics service quality that need to be improved and that can serve as the competitive advantage for enterprises, and to propose managerial implications and suggestions based on the analysis results. In screening topics related to logistics service quality, this study spent a lot of time and effort manually screening relevant topics and naming topic constructs in the topic modeling results. Future research can develop automated screening methods to perform these tasks more accurately and efficiently.

Acknowledgement

This work is partially supported by National Science and Technology Council, Taiwan under grant NSTC 111-2221-E-A49-084-MY3.

References

[1] Chatterjee, S., Goyal, D., Prakash, A. and Sharma, J. (2021). Exploring healthcare/health-

product ecommerce satisfaction: A text mining and machine learning application. Journal of Business Research, 131, 815-825, doi: 10.1016/j.jbusres.2020.10.043.

[2] Lim, M. K., Li, Y. and Song, X. (2021). Exploring customer satisfaction in cold chain logistics using a text mining approach. Industrial Management & Data Systems, 121(12), 2426-2449, doi: 10.1108/IMDS-05-2021-0283.

[3] Leem, B. H. and Eum, S. W. (2021). Using text mining to measure mobile banking service quality. Industrial Management & Data Systems, 121(5), 993–1007, doi: 10.1108/IMDS-09-2020-0545.

[4] Zhao, J., Gui. X., and Zhang. X. (2018). Deep Convolution Neural Networks for Twitter Sentiment Analysis. IEEE Access, 6, 23253-23260, doi: 10.1109/ACCESS.2017.2776930.

[5] Hajek, P., Barushka, A. and Munk, M. (2020). Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. Neural Computing and Applications, 32, 17259–17274, doi: 10.1007/s00521-020-04757-2.

[6] Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R. and Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. Industrial Marketing Management, 90, 523-537, doi: 10.1016/j.indmarman.2019.08.003.

[7] Paul, H. and Nikolaev, A. (2021). Fake review detection on online E-commerce platforms: a systematic literature review. Data Mining and Knowledge Discovery, 25, 1830–1881, doi: 10.1007/s10618-021-00772-6.

[8] Lee, P. J., Hu, Y. H. and Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. Telematics and Informatics, 35(2), 436–445, doi: 10.1016/j.tele.2018.01.001.

[9] Cao, Q., Duan, W. and Gan, Q. (2011). Exploring determinants of voting for the ‘helpfulness’ of online user reviews: A text mining approach. Decision Support Systems, 50(2), 511–521, doi: 10.1016/j.dss.2010.11.009.

[10] Yang, S., Zhou, Y., Yao, J., Chen, Y. and Wei, J. (2019). Understanding online review helpfulness in omnichannel retailing. Industrial Management & Data Systems, 119(8), 1565–1580, doi: 10.1108/IMDS-10-2018-0450.

[11] Wu, P. J. and Lin, K. C. (2018). Unstructured big data analytics for retrieving e-commerce logistics knowledge. Telematics and Informatics, 35(1), 237–244, doi: 10.1016/j.tele.2017.11.004.

[12] Wu, P. J. and Yang, C. K. (2018). Unstructured big data analytics for air cargo logistics management. 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Singapore, 2018, 274-278, doi: 10.1109/SOLI.2018.8476741.

[13] Agrawal, R., Imielinski, T. and Swami, A. (1993) Mining Association Rules between Sets of Items in Large Databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington DC, May 1993, 207-216, doi:

10.1145/170036.170072.

[14] Salminen, J., Kandpal, C., Kamel, A. M., Jung, S. and Jansen, B. J. (2022). Creating and detecting fake reviews of online products. Journal of Retailing and Consumer Services, 64, 102771, doi: 10.1016/j.jretconser.2021.102771.

[15] Ahmed, H., Traore, I. and Saad, S. (2018). Detecting opinion spams and fake news using text classification. Security and Privacy, 1(1), e9, doi: doi.org/10.1002/spy2.9.

[16] Ott, M., Choi, Y., Cardie, C. and Hancock, J. T. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 309–319, Portland, Oregon, USA. Association for Computational Linguistics, https://arxiv.org/abs/1107.4557.

[17] Parasuraman, A. Zeithaml, V. A. and Berry, L. L. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality, Journal of Retailing, 64(1), 12–40.

[18] Parasuraman, A., Zeithaml, V. A. and Malhotra, A. (2005). E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality. Journal of Service Research, 7(3), 213–233, doi: 10.1177/1094670504271156.

[19] Chakrabarti, S., Trehan, D. and Makhija, M. (2018). Assessment of service quality using text mining – evidence from private sector banks in India. International Journal of Bank Marketing, 36(4), 594-615, doi: 10.1108/IJBM-04-2017-0070.

[20] Hong, W., Zheng, C., Wu, L. and Pu, X. (2019). Analyzing the Relationship between Consumer Satisfaction and Fresh E-Commerce Logistics Service Using Text Mining Techniques. Sustainability, 11(13), 3570, doi: 10.3390/su11133570.

[21] Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. [online] Available at: https://arxiv.org/abs/1103.2903 [Accessed 19 May 2022].

[22] Ni, J., Li, J. and McAuley, J. (2019). Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 188–197, Hong Kong, China. Association for Computational Linguistics, doi: 10.18653/v1/D19-1018.

[23] Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation. The Journal of Machine Learning Research, 3, 993-1022, doi: 10.5555/944919.944937.