IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON HEALTHCARE INFORMATION TECHNOLOGY FOR THE EXTREME AND REMOTE ENVIRONMENTS

Received May 5, 2019, accepted May 25, 2019, date of publication May 30, 2019, date of current version July 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920091

Textual Analysis for Online Reviews: A Polymerization Topic Sentiment Model

LIJUAN HUANG[1], ZIXIN DOU[1], YONGJUN HU[1], AND RAOYI HUANG[2]

[1]School of Management, Guangzhou University, Guangzhou 510006, China
[2]Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong

Corresponding authors: Lijuan Huang (somhuanglj@gzhu.edu.cn) and Zixin Dou (13826064897@163.com)

This work was supported in part by the 2015 National Social Science Fund of China under Grant 15BGL201, in part by the 2018 National Social Science Fund of China under Grant 18BGL236, in part by the 13th Five-Year Plan Thinktank Project of Social Sciences of Guangzhou under Grant 2018BGZZK23, and in part by the Research Plan for a Graduate Joint-Training Program of Guangzhou University.

ABSTRACT More and more e-commerce companies realize the importance of analyzing the online reviews of their products. It is believed that online review has a significant impact on the shaping product brand and sales promotion. In this paper, we proposed a polymerization topic sentiment model (PTSM) to conduct textual analysis for online reviews. We applied this model to extract and filter the sentiment information from online reviews. Through integrating this model with machine learning methods, the results showed that the prediction accuracy had improved. Also, the experimental results showed that filtering sentiment topics hidden in the reviews are more important in influencing sales prediction, and the PTSM is more precise than other methods. The findings of this paper contribute to the knowledge that filtering the sentiment topics of online reviews could improve the prediction accuracy. Also, it could be applied by e-commerce practitioners as a new technique to conduct analyses of online reviews.

INDEX TERMS Textual analysis, sentiment model, polymerization computing, online reviews.

I. INTRODUCTION

ith the development of World Wide Web, posting online reviews has become a popular way for people to share their opinions and sentiments. It has become a common practice for e-commerce websites to provide people with the function to publish their reviews. From this perspective, online reviews can be a valuable resource for researchers to observe and even explore the real world.

Various studies have been conducted to examine the relationship between online reviews and product sales [1]–[5]. Findings of these studies proved that online reviews are a substantial information source for consumers. On the other hand, although most of the studies suggest that online reviews have an impact on future sales, these findings are not always consistent. For example, Duan *et al.* [6] and Ye *et al.* [7] found that the volume of online reviews had a positive effect on future movie revenues, while Chintagunta *et al.* [8] and Segal *et al.* [9] showed that the star ratings of reviews had a positive effect on future movie revenues. However, Hu *et al.* [10] pointed out that consumers pay more attention to the content of reviews rather than the simple statistics. Accordingly, a growing number of researchers study the sentiment embedded in the reviews to forecast product's sales. In this paper, we focus the sentiment index of online reviews by developing some heuristic algorithms.

Some studies have demonstrated the influence of textual sentiments on product sales, particularly in the automotive industry [11], stock market [12], and box office domain [13]. Fan *et al.* [11] used sales data and the sentiment score to predict sales performance. Batra and Daudpota [12] used the sentiment score and market data to develop an SVM model to predict stock price trends. Yu *et al.* [13] used the sales data and the sentiment factors to employ ARSA model to predict sales performance. It has been reported that combining the sentiment embedded in the reviews can improve forecasting performance. As indicated by Yu *et al.* [13], with each hidden factor focused on a specific aspect of the sentiments, the sentiment topics of online reviews allow us to understand the intricate nature of sentiments. However, few studies have considered filtering the sentiment topics of online reviews. If the number of the hidden topic is large, it may cause the problem of overfitting [14]. Also, the sentiment topics

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed.

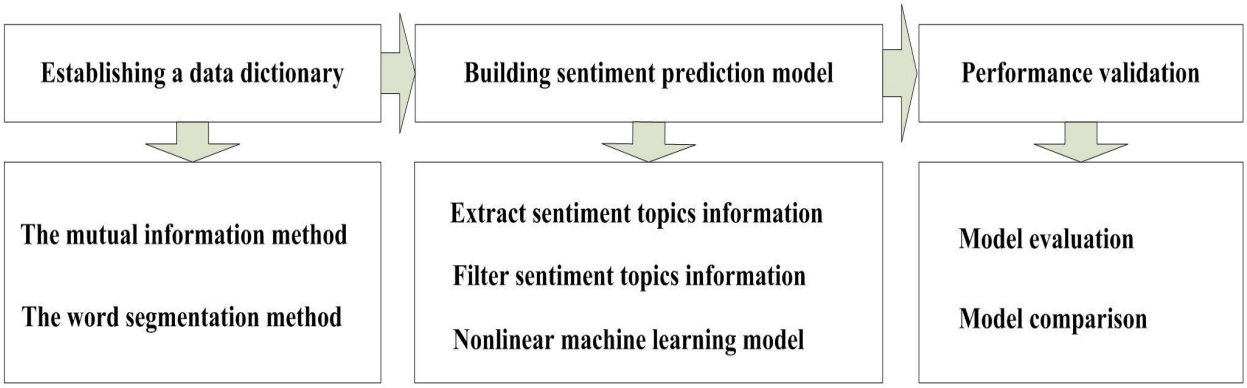


FIGURE 1. Research framework.

information contains invalid data, and it could influence the accuracy of the sales prediction.

In this paper, a Polymerization Topic Sentiment model (PTSM) is proposed to address the hidden sentiment topics problem in document-level sentiment. Not only can it overcome the shortcoming of the over-fitting problem, but also can filter the unnecessary information of sentiment topics. Furthermore, some machine learning models have been chosen.

The rest of the paper is organized as follows. Section 2 introduced the methodology including the establishment of a data dictionary, the development of a sentiment prediction model, and performance validation. Section 3 presented the experimental procedures and results. Finally, we present the conclusion of this study.

II. METHODOLOGY

Based on the discussion above, the research framework is illustrated in Figure 1, which mainly includes the following three steps:

Step 1: Establishing a data dictionary.

A corpus-based method is developed to construct a data dictionary. This dictionary uses mutual information to identify the most positive-relevant and the most negative-relevant features, rank them in two separate groups, and make the features that have a high level of sentiment strength as sentiment words.

Step 2: Developing a sentiment prediction model.

Not only can it overcome the shortcoming of the over-fitting problem, but also can filter the unnecessary information of sentiment topics. Also, the machine learning model has been chosen.

Step 3: Validating performance.

The forecasting performance is evaluated by using specific measures. At the same time, these results of the proposed method are compared with those alternative methods.

A. DATA DICTIONARY

A sentiment dictionary is necessary to calculate the sentiment factor from the content of online reviews. There are two methods: the lexicon-based approach and the corpus-based method [15]. The lexical resources are abundant, and a direct way is to get the sentiment dictionary from the well-defined lexicons, such as the dictionary of HowNet in English [16]. However, it is still hard to guarantee that the vocabularies in the dictionary are domain-consistent with our tasks. This paper develops a corpus-based method to construct a domain sentiment dictionary to solve this problem. In this paper, we choose the movie dataset as the corpus [17]. It provides a set of 50,000 highly polar movie reviews for training and testing.

The mutual information (MI) is widely used as a featured selection method [18].

First, all words in the neutral set in the training corpus are counted as filter words or filtering movie reviews of negative and positive sets in the training corpus.

Second, all words in the positive and negative sets in the training corpus are chosen as positive and negative features respectively and use word segmentation (WS) method to calculate the frequency of occurrences of each feature. Only these features which have the frequency number exceeding 10 are retained. These sentiment features are added to the generally lexicon-basic dictionary.

Third, all words in the generally lexicon-basic dictionary are chosen as candidate features. The MI metric be used to calculate the relevance of each candidate feature pw_i to the positive (+) class in the positive reviews group in the testing corpus by using equation (1), and the relevance of each negative candidate feature nw_i to the negative (-) class in the negative group by using equation (2), respectively:

$$MI(pw_i, +) = \log \frac{p(pw_i, +)}{p(pw_i)p(+)} \tag{1}$$

$$MI(nw_i, -) = \log \frac{p(nw_i, -)}{p(nw_i)p(-)} \tag{2}$$

Then, two groups of features in decreasing order of $MI(pw_i, +)$ and $MI(nw_i,-)$ are ranked by using equation (3) and equation (4) respectively:

$$W^+ = [pw_1^+, pw_2^+, \ldots, pw_n^+] \tag{3}$$

$$W^- = [nw_1^-, nw_2^-, \ldots, nw_n^-] \tag{4}$$

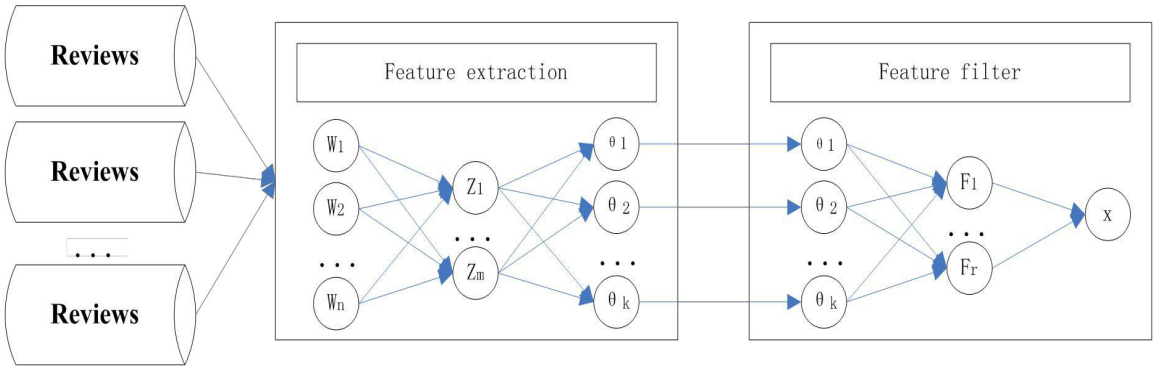


FIGURE 2. Working principle of PTSM.

Finally, the corpus-based dictionary is obtained by zipping pw_i and nw_i. Precisely, only polar words that have the values beyond 0 are retained.

It is vital to notice that this dictionary has more strength polarity word. The condition of the sentiment strength can be determined by requiring the positive-relevant and negative-relevant words to have the values beyond 0.

B. SENTIMENT EXTRACTION

To obtain a more accurate and useful sales prediction result, a model that can extract and filter the sentiments information from textual reviews is necessary. The basic idea of our model is shown in Figure 2. It not only can analyze the review data and calculate the sentiment factors but also can filter the sentiment information.

The PTSM model is formally presented. The variable is the sentiment distribution for document D. The same situation is the variable z, which is the sentiment word distribution. Supposing that a set of reviews are given $D = \{d_1, d_2, \ldots, d_A\}$ and a set of sentiment words from our corpus-based dictionary $W = \{w_1, w_2, \ldots, w_N\}$. The website review entry is associated with many unobserved hidden factors, $K = \{1, 2, \ldots, K\}$. Even if those sentiments are not directly observable in documents, it could correspond to the combinations of sentiment words from our corpus-based dictionary. Hence, PTSM is described as the following generative model.

A k-dimensional Dirichlet variable $\boldsymbol{\theta}$ has the following probability density on equation (5):

$$P(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \ldots \theta_k^{\alpha_k - 1}, \tag{5}$$

where the parameter $\boldsymbol{\alpha}$ is a k-vector with components $\boldsymbol{\alpha}_i > 0$, and where $\Gamma(x)$ is the Gamma function.

Given the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the joint distribution of all the variables in our model can be factorized as equation (6):

$$P(\theta, K, d|\alpha, \beta) = P(\theta|\alpha) \prod_{i=1}^{K} P(k_i|\theta_j) P(w_n|k_i, \beta), \tag{6}$$

where $P(k_i|\boldsymbol{\theta})$ is simply $\boldsymbol{\theta}_k$ for the k such that $\sum_{i=1}^{K} k_i = 1$.

Integrating over $\boldsymbol{\theta}$ and summing over k, the marginal distribution of a review is the equation (7):

$$P(d|\alpha, \beta) = \int p(\theta|\alpha) (\prod_{i=1}^{k} \sum_{k_i} P(k_i|\theta) P(w_n|k_i, \beta)) d\theta. \tag{7}$$

Finally, taking the product of the marginal probabilities of single reviews, the probability of a corpus is as in equation (8):

$$P(D|\alpha, \beta) = \prod_{d=1}^{D} \int p(\theta_d|\alpha) (\prod_{i=1}^{k_d} \sum_{k_{d,i}} P(k_{d,i}|\theta_d) P(w_{n,d}|k_{d,i}, \beta)) d\theta_d \tag{8}$$

According to the Dirichlet distribution expectation, the probability distribution parameters are shown as equation (9):

$$\theta_{k,d} = \frac{m_d^k + \alpha_k}{\sum_{k=1}^{K} m_d^k + \alpha_k}. \tag{9}$$

Let μ_t be the number of reviews which contain sentiment words for a given movie posted on day t, and $\boldsymbol{\theta}_{t,d,k}$ be the probability of the kth sentiment factors conditional on the dth review posted on day t. Then these sentiment factors of reviews are the Equation (10):

$$\pi_{t,k} = \frac{1}{\boldsymbol{\mu}_t} \sum_{j=1}^{\mu_t} \boldsymbol{\theta}_{t,k}. \tag{10}$$

Then the k-dimensional emotional data into a one-dimensional is polymerized by using Equation (11). It can filter the redundant sentiment information.

$$x_t = F_{t,r} * \{\pi_{t,k} * A_{t,j}\}^T, \tag{11}$$

where $A_{t,j}$ is the variance-covariance matrix, $F_{t,r}$ is the common factor, and r is the number of factors.

It is worthy of noting that x_t is the generalization of documents about emotions on day t. This generalization is used in the prediction of future product sales.

C. SENTIMENT PREDICTION MODEL

In this section, we choose an appropriate prediction model for our problem. The input of our model is in Equation (12):

$$y_t = (y_1, \ldots, y_p, x_1, \ldots, x_q) \tag{12}$$

Here, x_q is the feature which extracts from reviews in the t-qth day and y_p is the box office of the movie in the (t-p)th day. The output of the prediction model is the (t + 1)th day. In this paper, three existing regression models are compared, and their performances are tested in our experiments.

The linear model is represented as Equation (13):

$$y_t = g(i) + \varepsilon = W^T + \varepsilon \tag{13}$$

where $W = (w_1, w_2, \ldots, w_{p+q})^T$ is the vector of the weights and ε is a constant here.

To make the model simpler, it is modified as Equation (14), where $W' = (w_{p+q}, \varepsilon)^T$, $I' = (y_1, \ldots, y_p, x_1, \ldots, x_q, 1)^T$.

$$y_t = W'^T I' \tag{14}$$

However, a more expressive model is needed to represent. Therefore, in this paper, some powerful nonlinear models are applied to predict the box office. Here, the neural network (NN) model and support vector machine (SVM) model are selected to be our prediction models. In this aspect, the prediction results will be accurate enough by using these models.

In the next section, these models will be compared to demonstrate this speculation through our experiments.

D. VALIDATION METHOD

In this paper, the Mean Absolute Percentage Error (MAPE), mean squared error (MSE) and mean absolute error (MAE) are utilized to measure the prediction accuracy as equation (15), equation (16) and equation (17):

$$MAPE = \frac{1}{total} \sum_{i-1}^{Total} \frac{|\mathrm{Pred}_i - True_i|}{True_i}, \tag{15}$$

$$MSE = \frac{1}{total} \sum_{i-1}^{Total} |\mathrm{Pred}_i - True_i|^2, \tag{16}$$

$$MAE = \frac{1}{total} \sum_{i-1}^{Total} |\mathrm{Pred}_i - True_i|, \tag{17}$$

where *total* is the total number of predictions made on the testing data, $Pred_i$ is the predicted value, and $True_i$ represents the actual value of the box office revenues.

All the accuracy results reported herein are averages of 30 independent runs.

III. EXPERIMENTS ANALYSIS

A. DATA COLLECTION

Daily movie online reviews were collected for movies released in the United States. For a particular movie, reviews were collected from one day before the release to 30 days after. In total, 13,417 reviews on different movies were collected. We also collected the gross box office revenue data for the 30 movies from the same website.

B. PROCEDURE

In each run of the experiment, the following procedure was processed:

- 80% of movies dataset are selected for training and other 20% for testing; the online reviews and box office revenue data are correspondingly partitioned into training and testing data sets.
- A polymerization sentiment analysis model is trained using online reviews. For each review d, the sentiments toward a movie are summarized using a vector of the posterior probabilities of the sentiment factor.
- After the prediction model being fed with the probability comprehensive sentiment factor, the prediction performance of the prediction model is evaluated by experimenting with the testing data set.

C. RESULTS

In this section, these results obtained from a set of experiments conducted on a movie data set are reported to validate the effectiveness of the PTSM model.

1) PARAMETER ESTIMATION

Assume x_t to be the information, which is extracted from reviews. These are published during the days from *1* to q. Moreover, assume y_t is the box office of a certain movie during days from *1* to p. In other words, the prediction of equation (18):

$$y_t = g(x_1, x_2, \ldots, x_q, y_1, y_2, \ldots, y_p) \tag{18}$$

This paper now studies how the choice of these parameter' values (*K, p, q*) affect the prediction accuracy.

First, with fixed $p = 7$ and $q = 1$, K is varied. As shown in Figure 3(a), different K of our method have different MAPE. PTSM model achieves its best prediction accuracy when $K = 3$. This implies that it can not only fully capture the sentiment information but also can effectively filter redundant sentiment information and conserve critical information.

Then, with fixed q and K values, p is varied. From Figure 3(b), it can show that the model achieves its best prediction accuracy when $p = 7$. This suggests that p should be large enough to be factored in the important influence of the previous days' and not so large that irrelevant information in the farther past can affect the prediction accuracy.

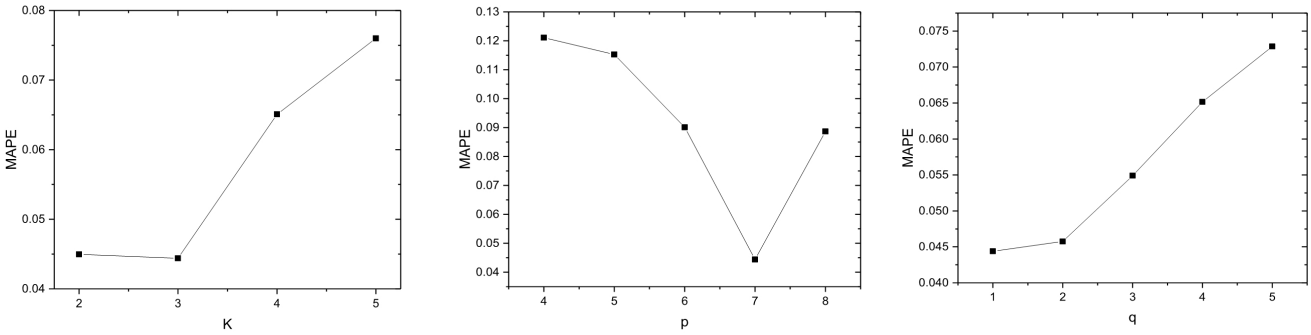
Finally, with fixed p and K values, q is varied. As shown in Figure 3(c), the best prediction accuracy is achieved at $q = 1$, which means that the sentiment information is most strongly related to the current sales.

Here, it is confirmed that $p = 7$ and $q = 1$, so we need to predict as equation (19):

$$y_t = g(x_{t-1}, y_1, y_2, \ldots, y_7) \tag{19}$$

2) COMPARISON AMONG PREDICTION MODELS

In this section, this paper needs to select a regression model. There are three models evaluated. The sentiment features of a PTSM are used as a part of the inputs.



(a) Effect of K (b) Effect of p (c) Effect of q

FIGURE 3. The effects of parameters on the prediction error.

TABLE 1. Average result of different prediction models.

MODEL	NN	SVM	AR
MAPE	4.44%	8.78%	11.47%
MSE	0.49%	1.30%	2.41%
MAE	4.08%	8.21%	12.16%

TABLE 2. The Average result of different sentiment methods in NN.

MODEL	PTSM	S-PLSA	S-LDA
MAPE	4.44%	5.88%	5.10%
MSE	0.49%	1.54%	0.49%
MAE	4.08%	5.25%	4.79%

For the machine learning models, we used *Lib-SVM* to implement the SVM model and *Nt-stool* to implement the NN model. The *RBF-kernel* of SVM is chosen, and the parameters applied in the RBF-kernel are set as defaults. As for the NN model, we set the number of hidden layers as 1 and the number of neurons per layer as 3. Our evaluation criteria are MAPE, MSE, and MAE.

As shown in TABLE 1, these nonlinear models are more accurate than the linear model (AR). Meanwhile, NN is better than SVM model.

3) COMPARISON AMONG SENTIMENT FEATURE METHODS

To verify the effectiveness of using our factors as the input vectors, we experimentally compared NN with a method using the other features as inputs. That is, instead of using PTSM, we trained an S-PLSA model and an S-LDA model and fed their features into the NN model for training and prediction. The number of factors is set to 3.

As shown in TABLE 2, the prediction model is more accurate when using the PTSM. This demonstrates that filtering sentiment feature is necessary to improve the accuracy of prediction results.

4) COMPARISON WITH LEXICON-BASED DICTIONARY

To test the effectiveness of using a corpus-based dictionary as the feature set, we experimentally compared our method with a model that uses the lexicon-based dictionary method for sentiment feature selection.

TABLE 3. The average result of different dictionaries.

MODEL	OUR DICTIONARY	LEXICON-BASED DICTIONARY
MAPE	4.44%	5.03%
MSE	0.49%	0.69%
MAE	4.08%	4.69%

TABLE 4. The average result of different prediction methods.

MODEL	PTSM	TRADITIONAL METHOD
MAPE	4.44%	6.06%
MSE	0.49%	0.80%
MAE	4.08%	5.54%

As shown in TABLE 3, using corpus-based dictionary words outperforms the lexicon-based dictionary words approach.

5) COMPARISON WITH TRADITIONAL PREDICTION METHODS

In this paper, our goal is to demonstrate the predictive power of sentiment of online reviews. Therefore, we chose one of the traditional methods which only removes the sentiment information from our features. The output is still y_t, and input is modified as equation (20):

$$I = (y_{t-1}, \ldots, y_{t-7})^T \tag{20}$$

The same train data and the NN model are used here. The evaluation criteria for the two models are compared.

As shown in TABLE 4, the gap on the accuracy is evident between the two. This phenomenon shows that adding sentiment information of reviews can improve the prediction performance.

IV. CONCLUSIONS

While prior studies have recognized the sentiment embedded in the reviews to forecast product sales more accurately, few studies consider filtering the sentiment from online reviews, and our study is to fill this gap.

In this paper, we conduct a case study in the movie domain and solve the problem of invalid information of reviews for predicting product sales performance. This paper firstly expounds the gaps in the existing literature; Secondly, a data

dictionary is established based on online movie reviews; Then develop a PTSM model, which is a novel probabilistic modeling framework. Based on this framework, the sentiment information is extracted and filtered from online textual reviews; Finally, our experiment shows that by integrating PTSM into the machine learning method can improve prediction accuracy.

The experiment results also show that filtering sentiment topics hidden in the reviews play a more important role in sales prediction, and the PTSM is more precise than alternative methods. With the proposed method, e-commerce companies can better harness the predictive power of reviews and conduct their business more effectively. For example, if an online retailer finds out that a movie is expected to generate more revenues, it could allocate larger rooms for that movie to accommodate more audiences.

Like other studies, this paper has its limitations. Our experimental language field could be extended to other areas other than English. In the future, we can even work with scholars from other countries to get sentiment reviews in different languages to conduct cross-cultural studies in this field.

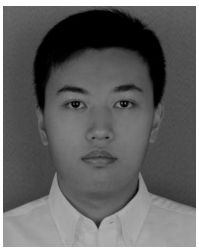
ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable remarks and comments.

REFERENCES

[1] K. Park and S. H. Ha, "Mining user-generated opinions online with LDA model to discover service complaints," *Information*, vol. 21, no. 3, pp. 875–884, 2018.
[2] Z. Li and A. Shimizu, "Impact of online customer reviews on sales outcomes: An empirical study based on prospect theory," *Rev. Socionetw. Strategies*, vol. 12, no. 2, pp. 135–151, 2018.
[3] D. Suryadi and H. Kim, "A systematic methodology based on word embedding for identifying the relation between online customer reviews and sales rank," *J. Mech. Des.*, vol. 104, no. 12, 2018, Art. no. 121403.
[4] H. Yuan, W. Xu, Q. Li, and R. Lau, "Topic sentiment mining for sales performance prediction in e-commerce," *Ann. Oper. Res.*, vol. 9, no. 1, pp. 1–24, 2017.
[5] R. Y. K. Lau, W. Zhang, and W. Xu, "Parallel aspect-oriented sentiment analysis for sales forecasting with big data," *Prod. Oper. Manage.*, vol. 27, no. 10, pp. 1775–1794, 2018.
[6] 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, 2008.
[7] Q. Ye, R. Law, and B. Gu, "The impact of online user reviews on hotel room sales," *Int. J. Hospitality Manage.*, vol. 28, no. 1, pp. 180–182, 2009.
[8] P. K. Chintagunta, S. Gopinath, and S. Venkataraman, "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets," *Marketing Sci.*, vol. 29, no. 5, pp. 944–957, 2010.
[9] J. Segal, M. Sacopulos, V. Sheets, I. Thurston, K. Brooks, and R. Puccia, "Online doctor reviews: Do they track surgeon volume, a proxy for quality of care?" *J. Med. Internet Res.*, vol. 14, no. 2, p. e50, 2012.
[10] N. Hu, L. Liu, and J. Zhang, "Do online reviews affect product sales? The role of reviewer characteristics and temporal effects," *Inf. Technol. Manage.*, vol. 9, no. 3, pp. 201–214, 2008.
[11] Z.-P. Fan, Y.-J. Che, and Z.-Y. Chen, "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis," *J. Bus. Res.*, vol. 74, pp. 90–100, May 2017.
[12] R. Batra and S. M. Daudpota, "Integrating stocktwits with sentiment analysis for better prediction of stock price movement," in *Proc. IEEE Comput., Math. Eng. Technol.*, Mar. 2018, pp. 1–5.
[13] X. Yu, Y. Liu, X. Huang, and A. An, "Mining online reviews for predicting sales performance: A case study in the movie domain," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 4, pp. 720–734, Dec. 2012.
[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[15] Q. Miao, Q. Li, and D. Zeng, "Fine-grained opinion mining by integrating multiple review sources," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 11, pp. 2288–2299, 2010.
[16] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM 10th SIGKDD Int. Conf. Knowl. Discovery Data Mining.*, 2004, pp. 168–177.
[17] S. Li, R. Xia, C. Zong, and C.-R. Huang, "A framework of feature selection methods for text categorization," in *Proc. Joint Conf. 47th Annu. Meeting ACL 4th Int. Joint Conf. Natural Lang. Process. (AFNLP)*, 2009, pp. 692–700.
[18] M. Hur, P. Kang, and S. Cho, "Box-office forecasting based on sentiments of movie reviews and Independent subspace method," *Inf. Sci.*, vol. 372, pp. 608–624, Dec. 2016.

LIJUAN HUANG received the M.A. and Ph.D. degrees from Nanchang University, in 1991 and 2006, respectively. She is currently the Director of the Academy of E-Commerce Research, Guangzhou University. She also holds a postdoctoral position at the Jiangxi University of Finance and Economics. Her current research interests include E-commerce and logistics, and supply chain management.



ZIXIN DOU received the B.S. degree in mathematics and applied mathematics from Guangzhou University, Guangdong, China, in 2016, where he is currently pursuing the M.S. degree in technology economy and management. From 2017 to 2018, he was a Visiting Research Student with The University of Sydney, Australia. His current research interests include E-commerce and sentiment analysis.

YONGJUN HU received the B.S. degree from Shandong University, in 2000, and the M.A. degree in electronic and communication engineering and the Ph.D. degree in management science and engineering from Sun Yat-sen University, in 2005 and 2013, respectively. From 2016 to 2018, he was a Visiting Research Fellow with The University of Sydney, Australia. He is currently the Director of the Institute of Business Intelligence and Data Science, Guangzhou University. His current research interests include machine learning and sentiment analysis.

RAOYI HUANG is currently pursuing the bachelor's degree with The Hong Kong Polytechnic University. She has a strong interest in E-commerce, data analytics, modeling, and statistics. She has been to the USA, U.K., and Australia for further study.