Sales Forecast Based on Fine-grained Sentiment Analysis of Pro duct Reviews

Tao Zhang[1,a], HuaPei Liu[1,b] and XianKun Zhu[1,c]

[1]School of Software Beijing University of Technology Beijing, China
a. 728720952@qq.com, b. 15733809676@163.com, c. iizxkii@qq.com

Keywords: Product review, fine-grained sentiment analysis, sentiment index, sales forecast.

Abstract: With the rapid development of Internet and e-commerce, online shopping has become the mainstream way of shopping, and product reviews have also become the reference basis for potential consumers' purchase decisions, affecting the sales of products. Product sales forecast is an important basis for enterprise procurement. Taking restaurant reviews as an example, this paper uses fine-grained sentiment analysis model to identify the sentiment tendency of each aspect of products in reviews, constructs fine-grained sentiment index, and proposes a sales forecasting model based on sentiment index. Through the experimental verification, the sales forecast model with sentiment index improves the accuracy of sales prediction.

1. Introduction

Product sales forecast is an important basis for enterprise procurement. If the sales forecast is less than the consumer demand, the consumer experience will be affected, and the consumer may refuse to consume again; if the sales forecast is too much, the product backlog will be caused, and the cost of the enterprise will be increased. In addition, the accuracy of sales prediction is more important for enterprises with higher requirements for product preservation. It can be seen that the accurate prediction of product sales is the basis of good development of the enterprise.

With the development of Internet and e-commerce, online ordering has become a very popular way of shopping, online shopping platform is not limited by time and space, ordering is more convenient and efficient, and product reviews have become an effective way for consumers to publish information. Potential consumers can get more comprehensive and objective information about product quality, usage and service attitude from product reviews. Compared with the product information released by enterprises, product reviews are more reliable and can affect consumers' purchase decisions, and then affect product sales. Product reviews have been recognized by academia and industry as an important factor influencing consumers' purchase decisions and product sales[1].

According to the word-of-mouth theory in the field of marketing, product reviews play a persuasive and knowing role in sales[2]. Research shows that the emotional factors in product reviews have a certain impact on the box office, sales and stock trend. However, most of these studies are about coarse-grained sentiment analysis of comment texts. Only by mining the emotional polarity of different aspects of products in the reviews, can the real emotional expression of consumers be more closer and the ability of sales prediction be improved. Therefore, this paper proposes a sales forecasting model based on fine-grained sentiment analysis of product reviews. The fine-grained sentiment analysis of product reviews is carried out to obtain the emotional tendency of all aspects of the product, construct the fine-grained sentiment index, and use the sentiment index in the sales forecast to improve the accuracy of sales forecast.

2. Related Work

At present, scholars at home and abroad mainly focus on product reviews in movie, book, hotel and other fields, and analyze the impact of product reviews on sales. Dellarocas et al. added the number of movie reviews to the box office prediction model, and found that the prediction accuracy of the model has improved, proving that the number of reviews is related to the box office[3]. Yoo et al. Take the intensity of emotion words defined in emotion dictionary as emotion score, and use it to predict the box office of movies. The experimental results show that emotion can predict the box office[4]. Yu et al. added comments and other related factors to the autoregressive model of box office prediction, and found that it can effectively improve the box office prediction effect[5]. Lu and Ye found that the number of hotel reviews and the emotional tendency of reviews can affect the hotel network bookings[6]. Chevalier and Mayzlin studies have shown that the number of ratings and reviews has a certain impact on book sales[7]. Shiyang Gong studied the influence of the endogeneity of online word-of-mouth, and the experiment shows that the number of comments is the main factor affecting the book sales[8]. Shaowu Zhang conducted opinion mining on book reviews, integrated emotional factors into the autoregressive model, established an autoregressive model based on emotional perception, and proved that the model has better sales forecasting performance[9]. Hua Wei et al. Found that the number of reviews had a significant impact on purchase decision[10]. Nan Shang mining emotional information from the car review text and applying it to the car sales prediction model, the experiment shows that it has better effect on the sales prediction of a single car model[11].

From the research of domestic and foreign scholars on the influence of product reviews on sales, it is found that most scholars take the structured data of product reviews, such as the number of reviews and scores, as the factors that affect sales. Some scholars have studied the influence of unstructured data such as review texts on sales volume, but most of these scholars conduct coarse-grained sentiment analysis on reviews, and few researchers take the fine-grained sentiment index of reviews as a factor affecting sales. In addition, most of the researches are focused on film, books, hotels, with little research on restaurant reviews. Therefore, this paper takes restaurant reviews as the research object, conducts fine-grained sentiment analysis on reviews, constructs sentiment index, and studies the prediction ability of sentiment index on sales.

3. Implementation Detail

3.1. Experimental Data

We have crawled through the product reviews, sales, review time and other information of 100 buffet stores. The data from January 2015 to December 2019 are selected as the experimental dataset. Taking month as the cycle, the emotional tendency and store sales of the comment text were counted.

The original time series of sales is relatively large, and the value range of emotion index is [-1,1]. Different dimensions and orders of magnitude of each data will affect the convergence speed and training effect of the sales prediction model. Therefore, it is necessary to standardize the sales so that the sales and sentiment index are in the same order of magnitude. In this paper, min-max is used to standardize the sales.

In order to expand the dataset, each store builds the data set in the way of sliding window, and divides the data set of each store into training set and test set in the proportion of 8:2. Then the training sets of each store are combined and disordered to form the training set of sales prediction model. Similarly, the test set is obtained. There are 3500 pieces of data in the expanded data set, 2800 of which are training sets and 700 are test sets.

3.2. Calculation of Sentiment Index

For consumers, important product features will influence consumers' purchase decisions. For example, consumers' evaluation of the catering industry involves many aspects such as "geographical location, service attitude and taste of dishes". For consumers, they may pay more attention to "service attitude" and "taste of dishes", while "geographical location" is secondary. Therefore, according to the importance of all aspects of the product, the weight of aspect words is allocated. Equation (1) is the calculation method of weight.

$$W_a=\theta S_a+(1-\theta)F_a \tag{1}$$

Where W_a represents the weight of product feature a, F_a represents the frequency of feature a, and S_a represents the ratio of samples whose emotional polarity is consistent with the comprehensive emotional polarity to the total samples. θ is set to 0.5.

In this paper, we use the month as the cycle to calculate the coarse-grained and fine-grained sentiment index. Coarse-grained sentiment index shows the overall emotion tendency of the product, without considering the emotion of all aspects of words.

$$month\_sentiment=\left|1-\left|\frac{N_{pos}-N_{neg}}{N_{pos}+N_{neg}}\right|\right| \tag{2}$$

Fine-grained emotion index shows the consumers' emotional tendency to all aspects of the product and considers the weight of all aspects of words.

$$sentiment_a=\left|1-\left|\frac{N_{pos}^a-N_{neg}^a}{N_{pos}^a+N_{neg}^a}\right|\right| \tag{3}$$

Where sentimenta represents the sentiment index of aspect a, N_{pos}^a represents the number of positive comments on the emotional polarity of aspect a, and N_{neg}^a represents the number of negative comments on the emotional polarity of aspect a.

Use (1) and (3) to calculate the fine-grained sentiment index of the product every month. A represents aspect word set.

$$month\_sentiment_{aspect}=\sum_{a\in A} W_a*sentiment_a \tag{4}$$

3.3. Sales Forecast Model

The influence of sentiment index on the monthly sales is lagged, and the lag period is one month. This paper uses the historical sales of the last few months of the product and the sentiment index of the last month of the product as the input characteristics of the model to predict the sales of the store in the next month. In this paper, the LSTM model with sentiment index is called sentiment-LSTM (SLSTM model). Figure 1 shows the structure of SLSTM model.

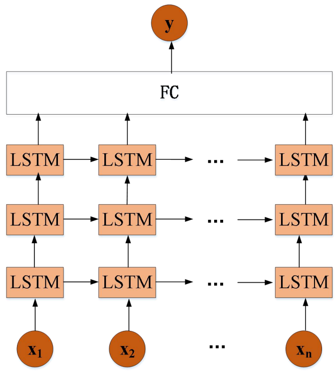

Figure 1: SLSTM model structure.

If SLSTM model is to predict the sales of a store in the ith month, the expression of SLSTM model is as follows.

$$pred_{sale_i} = f(w^s sentiment_{i-1} + \sum_{j=1}^{i-1} w_{i-j}\, sale_{i-j}) \tag{5}$$

Where salei-j represents the sale of the previous j months, wi-j represents the weight of the historical sales, sentimenti-1 represents the sentiment index of the previous month, ws represents the weight of the sentiment index. f is a nonlinear activation function.

Input the training set into the SLSTM model to train model. The SLSTM model uses RMSE as the loss function, and the optimizer uses Adam algorithm. The parameters of the model are updated by back propagation algorithm and small batch gradient descent algorithm.

4. Experimental Results

In order to highlight the advantages of SLSTM, a sales forecasting model based on sentiment index, the following three experimental results are used to compare the effects of different input characteristics on sales forecasting.

Experiment 1: only the first five months' sales is used as the input feature to train LSTM model. In order to clearly show the prediction effect of the model, only the first 45 test sets are selected to draw the fitting graph of the real value and the predicted value. Figure 2 shows the fit graph of the test set.

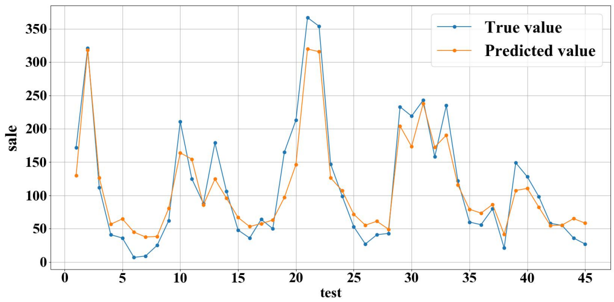

Figure 2: Line chart based on historical sales.

Experiment 2: The SLSTM model is trained by using the first five months' sales and the last month's coarse-grained sentiment index as input features. Figure 3 shows the fit graph of the test set.

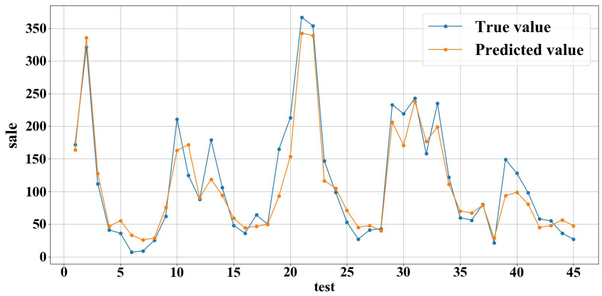

Figure 3: Line chart based on coarse-grained sentiment index.

Experiment 3: The SLSTM model is trained by taking the first five months' sales and the last month's fine-grained sentiment index of products as input characteristics. Figure 4 shows the fit graph of the test set.

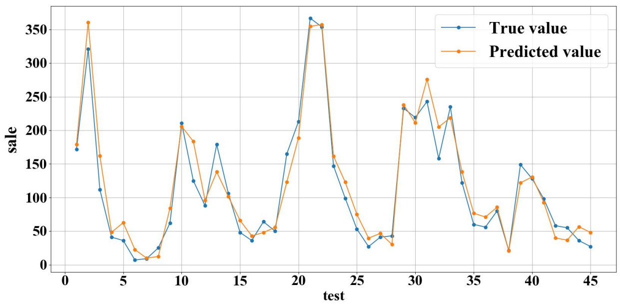

Figure 4: Line chart based on fine-grained sentiment index.

The effects of the three models are evaluated by MAE, RMSE and MAPE. The results are shown in Table 1.

Table 1: Evaluation results of each model.

Model	Evaluation Standard		
	MAE	RMSE	MAPE
Based on historical sales	23.8182	29.1346	0.4738
Based on coarse-grained sentiment index	20.4543	26.8205	0.3311
Based on fine-grained sentimet index	17.6634	22.3892	0.2694

It can be seen from TABLE I that the effect of the model based on sentiment index is better than that of the model based on only historical sales, while the effect of the model based on fine-grained sentiment index is better than that based on coarse-grained sentiment index. Therefore, using fine-grained sentiment index and historical sales as input features to train the sales forecast model improves the ability of data fitting. The fine-grained sentiment index of product reviews improves the accuracy of sales forecast.

5. Conclusions

This paper analyzes the correlation between product sales and product reviews, uses the sentiment analysis model to identify the sentiment polarity of all aspects of the product, constructs the sentiment index, and proposes a sales forecasting model based on the sentiment index. Experiments show that the model with sentiment index can better fit the data and improve the accuracy of sales forecast than the traditional sales forecast model. At present, this paper only uses historical sales and sentiment index as the factors that affect sales. In the future, we can add many factors, such as product price and comment quantity, to further improve the prediction accuracy of the model.

Acknowledgments

We thank my tutor, Professor Tao Zhang, who gave me a lot of constructive suggestions. Under his guidance, I successfully completed the writing of the paper. In addition, we thank the students in the lab, who gave me help and support.

References

[1] Hu N, Liu L, Zhang J, Do online reviews affect product sales? The role of reviewer characteristics and temporal effects, 9rd ed., vol. 3. Information Technology and Management, 2008, pp.201–214.

[2] Chen Y, Xie J, Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix, 54rd ed., vol. 3.Social Science Electronic Publishing, 2008, pp.477–491.

[3] Dellarocas C, Zhang X, Awad N F, Exploring the value of online product reviews in forecasting sales:The case of motion pictures, 21rd ed., vol. 4.Journal of Interactive Marketing, 2007, pp.23–45.

[4] Yoo S, Kanter R, Cummings D, Predicting Movie Revenue from IMDB Data, 2013.

[5] Yu X, Liu Y, Huang X, et al, Mining Online Reviews for Predicting Sales Performance:A Case Study in the hfovie Domain, 24rd ed., vol. 4.IEEE Transactions on Knowledge &Data Engineering, 2012, pp.720–734..

[6] Lu Q, Ye Q, How hotel star rating moderates online word-of-mouth effect:A difference-in-difference approach.Management Science and Engineering (ICMSE),2013 International Conference on, 2013, pp.3–8.

[7] Chevalier J A, Mayzlin D, The Effect of Word of Mouth on Sales:Online Book Reviews,43rd ed., vol. 3.Social Science Electronic Publishing, 2004, pp.345–354.

[8] Shiyang Gong, Xia Liu, Yang Liu, et al, Does online word of mouth decide the fate of products?An empirical analysis of online book reviews, 15rd ed., vol.4.Nankai management review, 2012, pp.118–128.

[9] Shaowu Zhang, Research on sales forecasting based on critical emotion and autoregressive model, Dalian University of technology, 2013.

[10] Hua Wei, Hong Jin, The influence of online reviews on consumers' purchase decision-making: regulation of product involvement and professional ability Role, 31rd ed., vol. 11.China's circulation economy, 2017, pp.78–84.

[11] Nan Shang, Research on automobile sales forecast based on comment sentiment analysis, Chongqing University of Posts and telecommunications, 2018.