INTERNATIONAL RESEARCH JOURNAL ON ADVANCED SCIENCE HUB

e-ISSN : 2582 - 4376
Open Access

RSP SCIENCE HUB

(The Hub of Research Ideas)
Available online at www.rspsciencehub.com

Comparative Study of Product Sales Forecasting Methods

M.Priya Alagu Dharshini[1], Dr. S. Antelin Vijila[2]
[1]*Research Scholar, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu.*
[2]*Assistant Professor, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu.*
preethi.murugiah10@gmail.com[1], antelinvijila@gmail.com[2]

Abstract

Sales forecasting plays a significant role in the development and success of consumer-oriented companies. Sales forecasting without high accuracy generates massive losses to the companies. To avoid losses, the company should focus on the factors which are affecting the sales forecasting. Nowadays people prefer e-commerce websites for purchasing products and they give online reviews and ratings about the products. These online reviews are used for computing the sentiment index which is necessary for sales forecasting. This paper surveys the different state-of-the-art sales forecasting techniques with different approaches. This survey also focused the sentiment analysis to predict sales forecasting.
Keywords: Sales forecasting, sentiment analysis, regression, bass model

1. Introduction

In this competitive digital world, organizations are competing with each other in terms of their dynamic business activities. The organizations make an effort to satisfy the expectations of the customers in terms of its quality and cost. To satisfy the customer demands, the manufacturer exploited effective supply chain management techniques [1]. In this digital world, Information Technology guides the manufacturers of the company to enhance the management techniques. Some of the effective supply chain management techniques are provided with the assist of Radio Frequency Identification (RFID), B2B websites, Enterprise Resource Planning, etc [2]. The best product sales forecasting methods are required for the effective supply chain management. Nowadays, big data and user contents are the growing areas in product sales forecasting. Different researches are carried out by different

researchers to study the impacts generated on the product of sales. From 2012 onwards most of the customers trusted the reviews given by the users. 30% of the customers haven't any trust about the reviews and 70% of customers fully depend on the ratings and reviews. These reviews are one of the most important factors for e-commerce business. Some of the e-commerce companies are amazon.com, tobacco.com, etc are succeeded because of the better reviews from their effective customers.In e-commerce, user contents have a significant role in manipulating the purchasing decision of customers. It also supports the organizations to know about the demands of the customers. The information required for the customers such as price, offers, discounts, varieties of the products, online reviews, etc is available on e-commerce websites. Nowadays e-commerce is preferred as the most important purchasing

channel by customers. For effective supply chain management, the organizations need to predict the purchasing decision of the customers. Historically obtained sales data, data get from the market are the existing product sales forecasting methods [3]. The information given by the potential customers in the e-commerce environment is used for the prediction of customer's needs and product sales. Product sales forecasting is one of the important requirements in the recent modern business environment. Due to the product sales forecasting, the companies can mitigate the losses and enhance their economic level [4]. The advancements in technology allow the customers to post their judgement about the products on social media and websites. The reviews are in real-time and these can be posted by different categories of persons from different locations [5]. With the assist of customer reviews, the companies can rectify the mistakes and find the steps to enhance their profits. The previous studies showed the necessity of online reviews for product sales. Equation-based approaches are preferred to find the association between box office revenues and online reviews [6]. It showed that the online review counts play a significant role in the sale of the box office. During the business plan formation, the forecasting methods assist the managers to make a better decision. This planning process should utilize the resources effectively to achieve a better profit over the years. These forecasting methods can be categorized into three as a short-term forecast, medium-term forecast, and long-term forecast [7]. Only three months are allocated for a short-term forecast where the production plan is prepared. In the medium-term forecast, the budget is prepared which is necessary for the business environment. For long-term forecasting, three years is required and the computer industry, steel factories, etc come under this category [8]. In this paper, different sales forecasting approaches and sentiment analysis are discussed to evaluate the performance. These approaches fully depend on the online reviews and previous sales data. The remainder of the survey paper is as follows. Section 2 briefly explained the forecasting techniques and its classification. Different forecasting models are discussed in Section 3 and these forecasting approaches are compared in Section 4. Finally, the conclusion is given in Section 5.
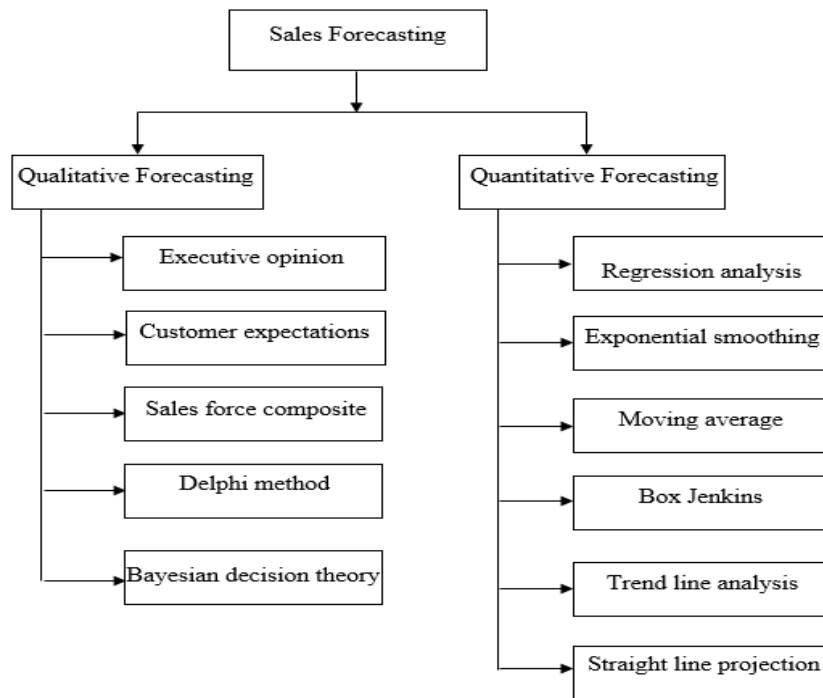
2. Forecasting Techniques

Forecasting techniques can be categorized into quantitative forecasting and qualitative forecasting. Most of the managers of the organizations preferred the quantitative forecasting technique if they have previous data about the product sales and have the capability to predict the situation. Due to the better prediction and mathematical techniques quantitative forecasting is fruitful to the users. Qualitative forecasting technique is preferred by the managers if they have not the capability to predict the situation and have not the past data for better analysis. This technique is mostly exploited by managers during the introduction of innovative products and technologies. Some of the quantitative and qualitative forecasting techniques are shown in Figure 1. The qualitative forecasting technique is also termed a judgemental technique because it mostly depends on the opinions of the executives. In this method, the product of sales depends on the survey, customer's expectations, opinion, etc. The forecasting technique based on the executive's opinion is a top-down approach that fully focused the sales on the future. In sales forecasting, the expectations of the customers are very important because customer satisfaction leads to more profit. The survey is taken among the customers or sales force applied to collect the expectation of the customers. Sales force composite is the bottom-up approach that allotted the sales persons to forecast the product in their respective areas. Delphi method is the same as the executive opinion based forecasting method. Delphi method differs from the executive opinion based forecasting method is that it does not need to gather all the members of the committee. In this approach, the questionnaire based on behavioural nature is prepared by the team lead and it is given to every member of the team. The main objective of this forecasting technique is to convert the opinion into one of the forms of the forecast. Bayesian decision theory is the combination of objective questions as well as subjective questions. This approach exploited the network analysis diagram for analyzing the critical path. The above discussed qualitative forecasting approaches are preferred by the managers when they have fewer amount of information. The

introduction of the new products has less amount of information, so it is the best example for qualitative forecasting methods and it is well suited for the prediction of sales revenue. These qualitative approaches are preferred when the market is affected by natural disasters, strikes, war, inflation, etc. During these situations, the collected past data will not use; only judgemental analysis gives the parameters which affect the market stocks accurately. The quantitative technique is also termed a mathematical technique because it mostly depends on the mathematical equations. Due to computerization, these forecasting techniques are mostly preferred. Regression analysis finds out the relationship between the independent variables and sales. Independent variables are related to the sales parameters such as cost, economic information, competitive information, and some other decision related with

products. In the exponential smoothing approach, the weighted average of the historical past data is computed to obtain the forecast. In the moving average approach, the forecasting is obtained by taking the average for the past historical data. This approach eliminates the unwanted oldest data and adds new information to get the recent forecast. In box Jenkins quantitative forecasting approach, autocorrelation is applied to the sales data to get the autoregressive based forecasting. It can be obtained from the collection of forecasting errors and past historical data. In the trend line analysis forecasting approach, the squared error between the actual and expected sales data is minimized to make the forecasting for future purposes. In a straight line projection approach, visual estimation of historical past data is collected and it is used as a future forecast.

Fig.1. Sales Forecasting Techniques

3. Forecasting models for product sales

In this survey different forecasting models such as linear regression model, sentiment analysis, bass model and econometric model are focused to know which forecasting model is better in terms of its product sales.

3.1 Linear Regression Model

In recent days, social media plays an important role in the box office revenues of movies. Sitaram

Asur et al analyzed how future sales forecasting is predicted with the help of Twitter as a social media [9]. The most popular regression model is preferred to predict the box office revenues for movies. For that prediction 24 movies are focused and the correlation is computed for the tweet rate. It attained a better correlation with the correlation coefficient value of 0.90. It recommends to find the linear relationship between the variables which was computed with the assist of the regression

model. It computes the minimum squared value for the 24 movies average tweet rate. The obtained square value is 0.80 showed that the predictive relationship is better. From the regression analysis, it is cleared that social media gives better predictions for movies. The linear regression model can be computed by the following expression

$$y = (\beta_a \times A) + (\beta_p \times P) + (\beta_d \times D) + \epsilon \quad (1)$$

Here y denotes the predicted revenues, A denotes the attention rate, P denotes the polarity, D denotes the factor distribution and ϵ denotes the error. Giang H. Nguyen et al analyzed the sales forecasting in terms of Artificial Neural Network (ANN) and regression [10]. This sales forecasting is evaluated based on economic indicators and past sales data. In this technique, short and long-term predictive approaches are preferred to compute the twenty-quarter predictions. They focused on the sales in the industry and exploited the regression analysis for computing the economic indicators. The regression finds the relationship between the economic indicators and industry sales by regarding economic indicators as an independent variable and industry sales as a dependent variable. Correlation computation, economic indicator selection, and prediction using ANN are the operations followed by regression. To fit the regression process, adjusted R^2 [11] is used and it can be computed by the following expression

$$\overline{R^2} = 1 - \frac{\sum e^2/(N-k-1)}{\sum(y-\overline{y^2})/(N-1)} \quad (2)$$

ANN performed the prediction by utilizing the training phase, validation phase, and testing phase. The specified patterns are categorized in the training phase, errors are minimized in the validation phase and finally, the sales are predicted in the testing phase. Thus this regression with the ANN approach effectively computes the factors which generate the impacts on future sales.

3.2   Sentiment based Approaches

In this modern digital world, e-commerce is increased rapidly, because most people prefer e-commerce websites to get the desired products. This process is sentimental-oriented because the customers verify the online reviews before buying the products. Yang Liu et al, categorize the products ranking based on sentiment analysis and technique for order preference by similarity to an ideal solution (TOPSIS). The major processes involved in this approach are sentiment classification and ranking using Fuzzy TOPSIS [12]. In sentiment classification, initially pre-processing is performed and it contains Part of Speech (POS) tagging and elimination of stop word. Then Support Vector Machine (SVM) and One Vs One (OVO) approach is used to classify the sentiments based on the online reviews as positive (supported vote), negative (opposed vote) and neutral (hesitated vote). Then product ranking is performed with the assist of determination of fuzzy members and computation of weights. Yu Mon Aye and Sint Sint Aung analyzed the sentiments of Myanmar people by considering the online reviews in Myanmar text [13]. They took the reviews from food and restaurants for sentiment analysis which is based on a dictionary-based approach and comes under the category of lexicon-based approach. This approach is one of the unsupervised learning methods and so there is no need for training data. In addition to that, it computes the summation of sentimental words. In dictionary-based techniques, initially, a list of sentiment words is prepared by human beings. The prepared list is termed as a sentiment lexicon that may be prepared automatically or manually. Manually generated sentiment lexicon is called as opinion lexicon which has less complexity and it consumes more time is the major issue of this approach. Here, the corpus is created by collecting the positive, negative and neutral restaurant reviews from social media such as Facebook, Twitter, etc. The researchers gathered 800 reviews for sentiment analysis. Then they preferred the senti-lexicon for classifying the sentiments in the Myanmar language. The sentiment lexicon can be generated with the assist of the following factors

L = {Target, Sentiment word, POS, Polarity}   (3)

In the sentiment lexicon, first pre-processing is applied to the formal and informal Myanmar reviews. Myanmar text comprises syllables that can be segmented, merged and POS tagging is applied. Then the sentiment words are extracted and matched with the dictionary which contains sentiments.

3.3 Econometric Model

Automobile industries require accurate sales forecasting due to the competition between the companies. Junjie Gao et al developed the econometric model to compute the sales forecasting of Chinese automobile companies [14]. The data taken for the econometric model are automobile sales in china, branded automobile sales and economic variables. In this econometric model time series is computed as an initial step. Then unit root test is taken with the assist of the univariate autoregression (VAR) model. An exogenous test is carried out to find the relationship between the endogenous variables. Further cointegration test is applied to find whether the duration of the relationship is long or short. In addition to that, the causality of the relationship is checked by Granger Causality test. Finally, Vector Error Correction Model (VECM) is carried out for predicting the sales. Based on VAR and VECM, the stability of automobile sales is determined. If the rooted values of both the factors are not exceeding one, then it is considered stationary. The accuracy of the forecasting methods is evaluated by computing the absolute percentage error and mean square error. Yves R. Sagaert et al analyzed the sales forecasting in terms of macroeconomic indicators [15]. They evaluate the sales forecasting using different approaches based on economic indicators. In this approach macroeconomic indicators are considered every month which contains information such as competition activities, promotion, price, etc. The presented forecasting approach easily chooses the indicators which are in lead and then arranges the indicators based on its utilization efficiency. In this forecasting approach, three different classes of information based on seasonality, autoregressive and indicators are preferred. This forecasting framework compared the conditional forecasting approaches with the unconditional forecasting approaches to evaluate the performances. The accuracy of this leading indicators based forecasting approach is better than other existing forecasting approaches. Chuan Zhang et al used the online reviews and macroeconomic indicators to evaluate the forecasting of product sales [16]. The major operations involved in this approach are macroeconomic indicators selection, conversion of online reviews into numerals and computation of word of mouth (WOM) outcome, construction of logarithmic autoregressive model and adoption of Adam optimizer. The selection of the macroeconomic indicators should satisfy the following conditions. The first one is the correlation between the sales volume and macroeconomic indicators should be strong and the second one is there is no multicollinearity between the selected macroeconomic indicators. The multicollinearity can be computed using the variance inflation factor (VIF) approach and it can be computed by the following expression.

$$VIF(y1, y2) = \frac{1}{1 - r^2(y1, y2)} \qquad (3)$$

Here y1, y2 denotes the selected macroeconomic indicators and r^2 denotes the linear correlation coefficient. After the macroeconomic selection, the sentiment indexes are computed by performing crawling, pre-processing and WOM outcomes. The crawling is used to get the related contents such as online reviews, browsing information, etc. In pre-processing approach, Jiewa technique is preferred for segmenting the words. Then the sentiments are analyzed based on the online reviews which are in text format. The WOM effect can be computed by the following expression

$$M_{ti} = \frac{C_{ti}(1 + B_{ti})}{1 + \sum_{j=1}^{n} C_{tj}} + \left(V_{ti} - \frac{V_{max}}{2} \right) \qquad (4)$$

Here C_{ti} denotes the ith review browsing number at time t, B_{ti} denotes the number of users, V^{ti} denotes the score given by the customers, V_{max} denotes the upper limit and n denotes the number of online reviews. In this approach, prospect theory is exploited to evaluate the final sentiment index. Then logarithmic autoregression model is applied to predict the sentiment index well suited for increasing the product sales.

3.4. Bass Model

Johan Grasman et al used the stochastic bass approach to estimate sales forecasting with high accuracy. In this method, initially, the required sales data are collected and then these data are fitted in the bass diffusion method [17]. Based on these data the sample size is determined for consistent forecasting. The obtained product samples are compared with the historical samples. The authors introduced the extension of stochastic

in the bass approach. The stochastic inserted in this framework is white noise. Then the variance is computed to find the upper and lower limit of the product sales on a yearly basis. Finally, the forecasting method is applied to derive point forecasts. The bass model is in the form of a differential equation and it can be computed by the following expression.

Table. 1. Comparison

S.No	Author	Sales forecasting methods	Purposes
1	Sitaram Asur et al [9]	Linear regression model, Sentiment analysis	It selects socialmedia as a Twitter for forecasting the revenues for box office movies in the future It also analyzed the sentiments obtained from tweets.
2	Giang H. Nguyen et al [10]	Regression model, Artificial Neural Network (ANN)	It exploits the regression method to identify the economic indicators. It uses ANN for predicting sales in the future.
3	Yang Liu et al [12]	Sentiment analysis, Fuzzy TOPSIS	Sentiment analysis for classifying the sentiments based on the online reviews. Fuzzy TOPSIS is used for ranking the products.
4	Yu Mon Aye et al [13]	Lexicon based approach,senti-lexicon	Lexicon-based approach for computing the sum of sentimental words. Senti lexicon for classifying the sentiment words.
5	Junjie Gao et al [14]	Econometric model	It predicts the Chinese automobile sales with the aid of unit root, exogenous, cointegration and ganger causality test.
6	Yves R. Sagaert et al [15]	Macroeconomic indicators	It evaluates sales forecasting by computing and ordering the leading indicators.
7	Chuan Zhang et al [16]	Macroeconomic indicators, nonlinear logarithmic autoregression model, prospect theory	It uses the prospect theory to compute the sentiment index. It uses the autoregression model to construct the sales forecasting
8	Johan Grasman et al [17]	Stochastic Bass model (stochastic approach + Bass diffusion model)	The stochastic information content is introduced in the form of white noise and it is fit in a bass diffusion model to estimate the sales parameters with high accuracy.
9	Zhi-Ping Fan et al [18]	Bass model, Sentiment analysis	It predicts sales forecasting with the help of historical sales data and online reviews. The Naive Bayes method is preferred for extracting sentiment-related content from online reviews.

$$x'(t) = f(x(t)) \quad (4)$$

$$f(x) = (n - x)\left\{u + \left(\frac{v}{n}\right)x\right\} \qquad (5)$$

Here, u denotes the innovation coefficient, v denotes the imitation coefficient and n denotes the number of buyers in the specified year. In this framework, the linear regression model has preferred for estimating the parameters. If the size

of the dataset is very small, then the estimated parameters are not attained high accuracy. But with the exploitation of a large dataset, the parameters are estimated with high accuracy. Zhi-Ping Fan et al analyzed the sales forecasting using the bass model and sentiment analysis with the assist of past data and online reviews [18]. The general form of the bass model consists of imitation and innovation and it can be computed by the following expression

$$C_s(t) = n \frac{1-e^{-(u+v)t}}{1+\frac{v}{u}(e^{-(u+v)t})} \qquad (6)$$

Here $C_s(t)$ denotes the sales, n denotes the users count, u denotes the innovation coefficient and v denotes the imitation coefficient. In the previous studies, the parameters n, u and v are used to predict the product sales forecast. But in this research, the abovementioned parameters, as well as online reviews are carried out to predict the sales forecast. This is the extended version of the bass emotion approach which has a similar process as the Norton model. It used the root mean squared error to compute the fitness between the actual data and predicted data. The performance of this approach is evaluated in terms of percentage error and mean absolute percentage error. The forecasting accuracy is determined by computing the difference between one and percentage error. This proposed approach has fewer prediction errors than existing approaches.

4. Comparison of sales forecasting methods

Multiple kinds of research related to sales forecasting methods are carried out by different organizations to predict sales forecasting in the future. Different sales forecasting methods and its multiple purposes are shown in table 1. From the table.1, it is observed that different hybrid sales forecasting methods are introduced by the researchers to enhance the sales forecasting efficiency. The lexicon-based approach collected 85% accurate sentimental words. Fuzzy TOPSIS performed the product ranking based on the online reviews. The econometric model utilized fewer economic indicators due to the difficulty presented in the monthly basis data collection. In the bass model, the number of forecasting errors is lesser than existing approaches. This study showed that

the combination of prospect theory with sentiment analysis performed better than other approaches.

Conclusion

This study surveyed the state-of-the-art techniques in the growing area of sales forecasting. This study mostly concentrated the sales forecasting and sentiment analysis with the exploitation of historical sales data and online reviews. Most of the approaches had high complexity because it needs to attain accurate sales forecasting. This review paper also focused on the macroeconomic indicators to predict product sales forecasting in the future. Different approaches are discussed to examine which one generated high forecasting accuracy. Finally, from this survey, it is cleared that most of the approaches used online reviews to evaluate sales forecasting.

References

[1]. Chong, A.Y.L and Zhou, L, 'Demand Chain Management: Relationships Between External Antecedents, Web-Based Integration And Service Innovation Performance', International Journal of Production Economics, Elsevier, vol. 154, pp. 48–58, 2014

[2]. Croom, S.R, 'The Impact of E-Business on Supply Chain Management: An Empirical Study of Key Developments', International Journal of Operations & Production Management, 2005.

[3]. Duan, W, Gu, B. and Whinston, A, 'The Dynamics of Online Word-Of-Mouth and Product Sales - An Empirical Investigation of The Movie Industry', Journal of retailing, 2008.

[4]. Marshall P, Dockendorff M and Ibanez S, 'A Forecasting System for Movie Attendance', Journal of Business Research, vol. 66, no. 10, pp. 1800–1806, 2013.

[5]. Chen Y and Xie J, 'Online Consumer Review: Word-Of-Mouth as a New Element of Marketing Communication Mix', Management Science, vol. 54, no. 3, pp. 477–491, 2008.

[6]. Duan W, Gu B and Whinston AB, 'Do Online Reviews Matter? An Empirical Investigation of Panel Data', Decision Support Systems, vol. 45, no. 4, pp. 1007–1016, 2008.

[7]. Martinovic Jelena and Damnjanovic Vesna, 'The Sales Forecasting Techniques', International Scientific Days, pp. 1-6, 2000.

[8]. Jobber, D. and Lancaster, G, 'Selling and Sales Management', London, Pearson Education, pp. 412-413, 2003.

[9]. Asur S and Huberman BA, 'Predicting the Future with Social Media', In: Proceedings of the international conference on web intelligence and intelligent agent technology, pp 492–499, 2010.

[10]. Giang H. Nguyen, Jai Kedia and Ryan Snyder, 'Sales Forecasting using Regression and Artificial Neural Networks', Research Gate, pp. 1-11, 2013.

[11]. Ohtani, K, 'Bootstrapping R^2 and Adjusted R^2 in Regression Analysis', Economic Modelling, vol. 17, no. 4, pp. 473–483, 2000.

[12]. Yang Liu Jian-Wu Bi and Zhi-Ping Fan, 'A Method for Ranking Products Through Online Reviews Based on Sentiment Classification and Interval-Valued Intuitionistic Fuzzy TOPSIS', International Journal of Information Technology and Decision Making, vol. 16, pp. 1-27, 2017.

[13]. Yu Mon Aye and Sint Sint Aung, 'Senti-Lexicon and Analysis for Restaurant Reviews of Myanmar Text', International Journal of Advanced Engineering, Management and Science (IJAEMS), vol. 4, no. 5, pp. 380 – 385, 2018.

[14]. Junjie Gao, Yanan Xie, Xiaomin Cui, Han Yu and Feng Gu, 'Chinese Automobile Sales Forecasting Using Economic Indicators And Typical Domestic Brand Automobile Sales Data: A Method Based on Econometric Model', Adances in Mechanical Engineering, vol. 10, no. 2, pp. 1-11, 2018.

[15]. Yves R. Sagaert, El-Houssaine Aghezzaf, Nikolaos Kourentzes and Bram Desmet, 'Tactical Sales Forecasting Using A Very Large Set of Macroeconomic Indicators', European Journal of Operational Research, pp. 1-35, 2017.

[16]. Chuan Zhang, Yu-Xin Tian, Zhi-Ping Fan, Yang Liu and Ling-Wei Fan, 'Product Sales Forecasting Using Macroeconomic Indicators and Online Reviews: A Method Combining Prospect Theory and Sentiment Analysis', Soft Computing, pp. 1-14, 2019.

[17]. Johan Grasman and Marcel Kornelis, 'Forecasting Product Sales with A Stochastic Bass Model', Journal of Mathematics in Industry, pp. 1-10, 2019.

[18]. Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen, 'Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis', Journal of Business Research, vol. 74, pp. 90-100, 2017.