Journal of Business Research 74 (2017) 90–100

Contents lists available at ScienceDirect

Journal of Business Research

Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis

Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen *

Department of Information Management and Decision Sciences, School of Business Administration, Northeastern University, Shenyang 110167, China

ARTICLE INFO

Article history:
Received 14 May 2016
Received in revised form 19 January 2017
Accepted 21 January 2017
Available online xxxx

Keywords:
Product sales forecasting
Online reviews
Sentiment analysis
Bass model
Norton model

ABSTRACT

Online reviews provide consumers with rich information that may reduce their uncertainty regarding purchases. As such, these reviews have a significant influence on product sales. In this paper, a novel method that combines the Bass/Norton model and sentiment analysis while using historical sales data and online review data is developed for product sales forecasting. A sentiment analysis method, the Naive Bayes algorithm, is used to extract the sentiment index from the content of each online review and integrate it into the imitation coefficient of the Bass/Norton model to improve the forecasting accuracy. We collected real-world automotive industry data and related online reviews. The computational results indicate that the combination of the Bass/Norton model and sentiment analysis has higher forecasting accuracy than the standard Bass/Norton model and some other sales forecasting models.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Firms use product sales forecasting as a foundation to estimate sales revenue and make decisions regarding production, operation and marketing strategies (Marshall, Dockendorff, & Ibáñez, 2013; Shi, Bigdeli, & Li, 2015). Through product sales forecasting, firms can create a plan for marketing, sales management, production, procurement, logistics and so on to improve their economic benefits and reduce losses caused by weaknesses in the production plan (Mentzer & Bienstock, 1998). According to the extant research, two primary factors influence consumers' purchasing decisions. One is the influence of other consumers who have bought the product and recommended it through verbal communication. The other is the influence of advertisements and the mass media, among other factors. A number of researchers have studied product sales forecasting and developed effective forecasting models that take relevant factors into account. Among them, the Bass model (Bass, 1969) simultaneously considers these factors as external and internal coefficients. Thus, the Bass model along with its extensions, such as the Norton model (Norton & Bass, 1987) and the contingent diffusion model (Peterson & Mahajan, 1978) is commonly used for new products, technology diffusion and product sales forecasting (Hyman & Michael, 1988), and it has been successfully applied in many fields, particularly in the durable consumer goods (Bass, 2004; Wang, Chang, & Hsiao, 2013), equipment and IT technology (Speece & Maclachlan, 1995; Barnes, Southwell, Bruce, et al., 2014; Wu & Chu, 2010), telecommunication services and retail (Seol, Park, Lee, & Yoon, 2012; Song, Lee, Zo, & Lee, 2015; Guo, 2014; Turk & Trkman, 2012) industries.

Word of mouth (WOM) is considered one of the most important factors influencing the purchasing decisions of consumers, especially with regard to imitators (Herr, Kardes, & Kim, 1991; Taylor, 2003). Online WOM, such as online reviews and microblogs, have become popular with the development of Internet technologies. A number of e-commerce websites such as Amazon and Taobao have established online review systems to encourage consumers to post product reviews and, as a result, have gradually changed consumer behavior patterns and affected consumer purchasing decisions. For example, consumers are paying increasingly more attention to online opinions when deciding which movies to watch, in which stocks they should invest, etc. (Wysocki, 2000; Ryu & Han, 2010). In addition, many online communities, such as Facebook and Douban, provide platforms for consumer discussions. These reviews often reveal personal emotions, such as happiness, anger, sorrow, criticism and praise, and potential consumers can browse the public opinions on a product to inform their purchase decisions. Accordingly, in the last decade, sentiment analysis techniques have been used to measure the sentiments conveyed through the content of online reviews (Pang & Lee, 2005; Prabowo & Thelwall, 2009). As indicated by Yu, Liu, Huang, and An (2012), the sentiment index extracted from the content of online reviews by sentiment analysis techniques can be used to forecast many social economic phenomena, including product market shares, box office attendance, transmissions of information or diseases (Culotta, 2010) and the results of political elections (Lee, 2009). The sentiment index can also be used to analyze macroeconomic

* Corresponding author.
E-mail addresses: zpfan@mail.neu.edu.cn (Z.-P. Fan), 351560460@qq.com (Y.-J. Che), zychen@mail.neu.edu.cn (Z.-Y. Chen).

http://dx.doi.org/10.1016/j.jbusres.2017.01.010
0148-2963/© 2017 Elsevier Inc. All rights reserved.

conditions (Bollen, Mao, & Zeng, 2011) and warn the public of emergencies (Yu & Kak, 2012).

Many studies indicate that online WOM influences consumer behavior and product sales (Liu, 2006; Godes & Mayzlin, 2004; Chevalier & Mayzlin, 2006). They find that the attributes of online reviews, such as the number of online reviews (Duan, Gu, & Whinston, 2008; Ye, Law, & Gu, 2009; Liu, 2006), ratings (Chevalier & Mayzlin, 2006; Segal et al., 2012), and sentiments shared in the reviews (Ye et al., 2009), have effect on product sales. Several researchers have explored the relationship between online reviews and product sales (Chevalier & Mayzlin, 2006; Dellarocas, Awad, & Zhang, 2007) and have developed forecasting methods that combine the Bass model and the ratings from online WOM (Wu, Wang, & Li, 2015; Dellarocas et al., 2007). However, these models have used only the historical sales and rating data to forecast product sales. Few studies have developed improved versions of the Bass models to consider the sentiments expressed in the content of online reviews. As discussed by Dellarocas et al. (2007), combining the Bass model and sentiment analysis has the potential to improve the forecasting performance of the standard Bass model.

In this study, a method that combines the Bass/Norton model and sentiment analysis is proposed to forecast product sales using product review data. This method incorporates the Naive Bayes (NB) algorithm to compute the sentiment index of online reviews and then employs the sentiment index to extend the imitation coefficient in the Bass/Norton model. To the best of our knowledge, few studies have taken into account the content of online reviews when extending the Bass/Norton model to improve product sales. Moreover, in this study, real-world automotive industry data are used to evaluate the forecasting performance of the proposed method.

The remainder of this paper is structured as follows. Section 2 provides a comprehensive literature review, whereas Section 3 describes the research framework, including data collection, online review data processing, forecasting models and performance criteria. Section 4 then provides the forecasting results and comparisons with the standard Bass/Norton model and some other sales forecasting models, and Section 5 discusses the conclusions and limitations of this study and suggests future research directions.

2. Literature review

The extant literature regarding product sales forecasting using online review data and using the Bass model are discussed herein.

2.1. Product sales forecasting based on online review data

A number of scholars have developed sentiment analysis techniques for predicting sales performance using online product review and blog data mining (Asur & Humberman, 2010). In the existing literature, three types of information are extracted from online reviews in the forecasting models. The first type of information is volume, which refers to the number of online reviews. As the number of reviews a product has increases, consumers' knowledge about the product increases (Liu, 2006). The second type is valence, which refers to the degree of consumer satisfaction with the product, e.g., the number of positive and negative reviews (Liu, 2006; Godes & Mayzlin, 2004). The third type of information is dispersion. As the distribution of product review information becomes more dispersive, consumers' knowledge about the corresponding product increases (Godes & Mayzlin, 2004).

Yu et al. (2012) trained a sentiment-based probabilistic latent semantic analysis model to obtain sentiment information from online reviews and then proposed an auto-regressive sentiment-aware model for sales forecasting. Using movie reviews and box office data, they found that sentiment information and the quality of online reviews have a substantial effect on box office forecasting. Asur and Humberman (2010) adopted the chatter from Twitter.com to forecast box office sales. They used the LingPipe linguistic analysis package to construct a sentiment analysis classifier and measured the ratio of positive to negative tweets to quantify the sentiments about a movie, and then constructed a linear regression model of the rate of positive and negative online film reviews. They found that the sentiments extracted from Twitter improve forecasting power. Liu, Huang, An, and Yu (2007) collected blogs using Google's blog search engine and the box office revenue data from the IMDB website to explore the forecasting power of blogs. They forecasted product sales utilizing an auto-regressive sentiment-aware model and the sentiment information obtained from a sentiment-based probabilistic latent semantic analysis. Archak, Ghose, and Ipeirotis (2011) used the programming interface provided by Amazon Web Services to collect daily product prices and product ratings from consumer reviews on Amazon and combined natural language processing and crowdsourcing on Amazon Mechanical Turk to extract opinions from online reviews and to model a linear equation with product reviews. They demonstrated that textual data in product reviews could be used to determine consumers' relative preferences for different product features and thereby forecast future changes in sales. Different from most existing research which uses online reviews to forecast box office sales and sales of digital products, our research focuses on the automotive industry and uses the sentiment index to extend the Bass model to forecast product sales.

2.2. Product sales forecasting using the Bass model

Recently, many researchers have modified the Bass model to improve the forecasting accuracy and have provided explicit guidance (Wang et al., 2013; Speece & Maclachlan, 1995; Barnes et al., 2014; Seol et al., 2012; Song et al., 2015; Guo, 2014; Turk & Trkman, 2012). Speece and Maclachlan (1995) extended the Bass and Norton models by adding pricing and market growth factors to forecast the use of packaging technology. Wang et al. (2013) used a modified Bass model to forecast the notebook shipments from Taiwanese firms and used a hybrid evolutionary algorithm for the parameter estimates to improve forecasting accuracy. Barnes et al. (2014) used the Bass model to explore the effects of incentive schemes on carbon-reducing technologies and provided a general quantitative measure of the effect of an incentive scheme on technology adoption. Seol et al. (2012) proposed a competitive Bass model to forecast the demand for new services while considering competitive relationships with existing services. Song et al. (2015) used an improved Bass model, the hybrid Bass-Markov model, to forecast the competitive service diffusion process. Turk and Trkman (2012) forecasted broadband diffusion in European countries using the Bass model and analyzed the future of broadband services. Lee, Kim, Park, and Kang (2014) used a statistical and machine learning-based approach based on the Bass model for the pre-launch forecasting of new product demand. Fernández-Durán (2014) defined a seasonal Bass model that took into account the seasonal effects of products and used a family of distributions for circular random variables to estimate seasonal effects. For the automobile industry, historical product sales data are incorporated into Bass model to forecast the sales of Alternative Fuel Vehicle (Shoemaker, 2012) and future automobile products (Zhu, Jiang, & Chen, 2008) and to explore the maturity of hybrid power technology (Gao, Chai, & Tang, 2013).

In a different view, few studies have emphasized combining the Bass model with online review data for forecasting models. Dellarocas et al. (2007) developed a Bass model based on the revenue forecasting model. To test the innovation and imitation coefficients of the Bass model, they used online ratings, the number of posted reviews and information about the reviewers obtained from Yahoo movies. They found that the arithmetic mean of ratings is a useful proxy for WOM when forecasting box office sales. In the extant literature related to the Bass model, only Dellarocas et al. (2007) investigated the relationship between online reviews and product sales and used online review data to forecast product sales. This paper differs from this previous research in that we extract the sentiment index from the content of online reviews, rather than ratings, and use it to extend the Bass and Norton model.

3. Methodology

3.1. Research framework

The research framework is illustrated in Fig. 1, which shows that product sales forecasting using online review data includes the following three steps:

(1) Data collection and preprocessing. In this step, products with multiple technological generations are selected and sales, along with online review data, are collected. Word segmentation methods and word frequency statistics are then used for data preprocessing.

(2) Sentiment index extraction and forecasting model building. In this step, a sentiment analysis method, i.e., the NB method (Yu, Duan, & Cao, 2013), is used to analyze the review data and to calculate the sentiment index based on the time point. A new forecasting model that combines the Bass and Norton model with the extracted sentiment index is developed to improve the forecasting accuracy.

(3) Performance validation. In this step, we fit the proposed model and evaluate the forecasting performance using specific measures. At the same time, the results of the proposed method are compared with those for the standard Bass and Norton model.

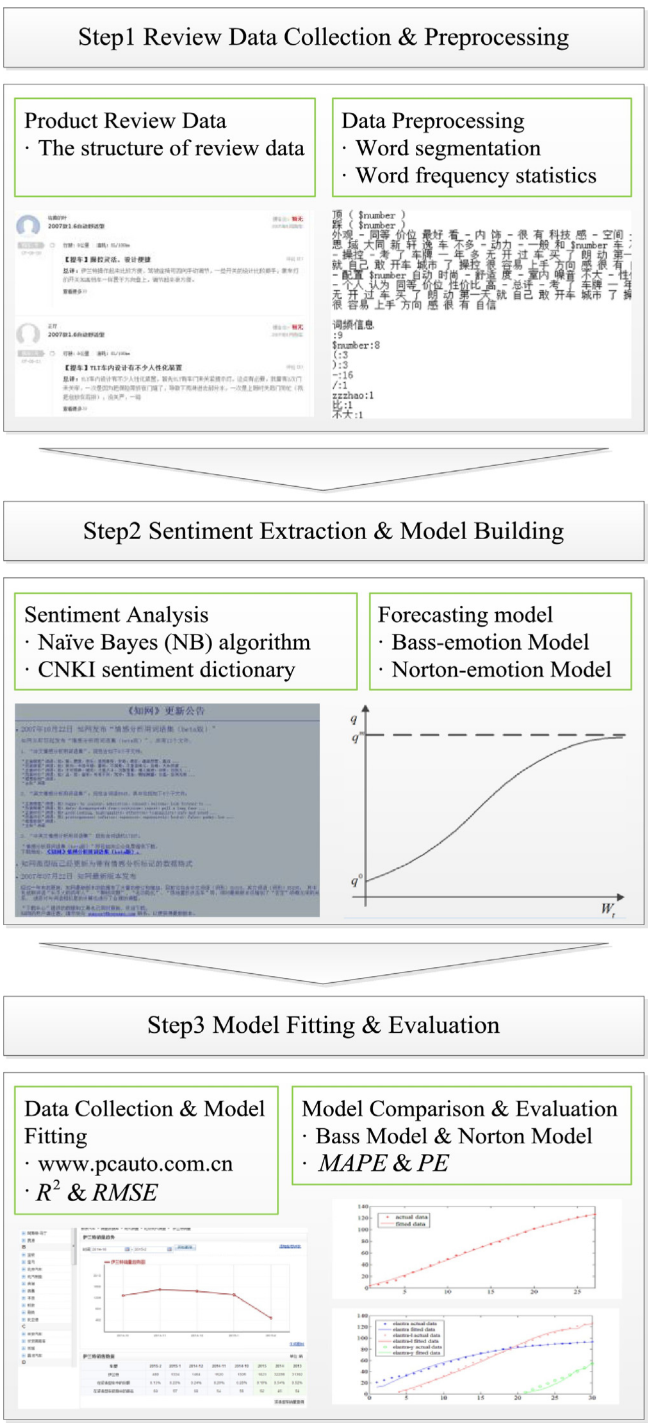

Fig. 1. Research framework.

3.2. Data collection and preprocessing

In this study, two types of data, i.e., historical sales and online reviews data, were collected. We extracted the following attributes for each online review: reviewer ID, title, rating and time, the number of times other users have browsed the review, the number of users who agree with (oppose) the review, and content. These attributes are presented in Table 1.

As discussed in Section 2.2, information extracted from online reviews can be classified in terms of volume, valence and dispersion, as presented in Table 1. The number of reviews made during a particular time period is regarded as volume. The ratings and content of the reviews denote the degree of consumer satisfaction with the product, which is represented by valence. The number of views is regarded as dispersion, which means that as more users see the reviews, the effect on others increases.

Selecting useful attributes when adopting a forecasting model is a critical task. Some scholars have focused on the influence of online reviews on different types of products. For the prediction of product sales, different kinds of products can be classified as experiential and search products (Nelson, 1974). For experiential products such as movies and books, the number of online reviews and the contents of such reviews have a higher impact on this type of product (Chevalier & Mayzlin, 2006; Dellarocas et al., 2007; Duan et al., 2008; Liu, 2006; Ye et al., 2009; Segal et al., 2012). For search products such as digital camera, video camera and notebook computer, the content of online reviews has a higher impact on this kind of product (Li, 2012; Cabral & Hortacsu, 2003; Archak et al., 2011). Regardless of whether it is an experiential products or a search product, the content of online reviews acts as an important part of sales forecasting.

In this study, the second type of information, i.e., valence, explains the imitation coefficient in the Bass and Norton model. Dellarocas et al. (2007) regard the rating of online reviews as the valence of online reviews to measure the product customer satisfaction. Compared with the work of Dellarocas et al. (2007), the sentiment index extracted from the content of online reviews, rather than ratings, is used to reveal individual preferences reflected in the imitation coefficient of the Bass model. Thus, consumer preferences can be measured more accurately by our model because consumers who share equal degrees of satisfaction with a product may assign different ratings to that product.

Table 1
Attributes extracted from online reviews.

Attribute	Description
ID	Reviewer ID
Title	Main content of the review
Rating	Product rating by reviewer
Time	Time reviewer conducted the review
Browse	Number of times users browsed the review
Agree	Number of users who agreed with the review
Oppose	Number of users who opposed to the review
Content	Content of the review

To calculate the sentiment index from the content of online reviews, a sentiment dictionary is needed. Sentiment terms, according to the CNKI sentiment dictionary (http://www.keenage.com/html/c_bulletin_2007.htm), are identified to calculate the sentiment index discussed.

3.3. Sentiment index extraction

In this study, the NB method (Yu et al., 2013) is used for polarity classification with the aim of obtaining a sentiment index for each online review. We suppose two categories of sentiment, each of which is represented by C_i with $i \in \{+,-\}$. That is, C_+ and C_- represent positive categories and negative categories, respectively. The set of emotional words is represented by $D \in R^{\eta}$, where η is the number of emotional words. The set of emotional words in a review is represented by $D_k \in R^{\eta}{}_k (k=1, \cdots, n)$, where n is the number of reviews and m_k is the number of emotional words in the kth review. An emotional word appearing in D_k is represented by w_{jk}.

The probability that D_k is in category C_i, which means that $i \in \{+,-\}$, is calculated as follows:

$$\mathrm{argmax}_{c_i}(P(C_i|D_k)) = \mathrm{argmax}_{c_i}\left(\frac{P(D_k|C_i) \times P(C_i)}{P(D_k)}\right), \tag{1}$$

where $P(C_i)$ is the probability of the ith category, which can be estimated using the number of positive and negative categories in the training set, $P(D_k)$is the probability of that specific set of emotional words occurring, and $P(D_k|C_i)$ is the probability that the terms in D_k appear in category C_i.

The probability $P(D_k|C_i)$ is calculated as follows:

$$P(D_k|C_i) = P(w_{1k}, w_{2k}, \ \ldots, w_{nk}|C_i). \tag{2}$$

We assume that emotional words are independent of one another. Thus, Eq. (2) is simplified as follows:

$$P(D_k|C_i) = \prod_j P(w_{jk}|C_i), \tag{3}$$

where $P(w_{jk}|C_i)$ is the number of w_{jk} that appear in category C_i divided by the total number of terms in category C_i. The probability $P(w_{jk}|C_i)$ is calculated as:

$$P(w_{jk}|C_i) = \frac{t_{ij}+1}{\sum_{w'_j \in V} t_{i'j} + K}, \tag{4}$$

where K is the number of terms in the sentiment dictionary, t_{ij} is the number of times w_{jk} appears in the training set belonging to category C_i, $t_{i'j}$ is the number of times $w_{j'k}$ appears in the training set belonging to category C_i.

Because the NB method determines the independent assumption of terms which is not true for most languages, it is less accurate than more complex models, such as support vector machine (SVM) and k-nearest neighbors (KNN). However, the NB method often works well to classify sentiment polarity (Cao, Thompson, & Yu, 2013; Yu et al., 2013). Using the NB sentiment classification algorithm, we are able to not only classify the online reviews as positive or negative category but also calculate the sentiment index of each review.

Here, we let W_{tk} equal the value of $P(C_i|D_k)$ and represent the sentiment index of review k in time period t. The value of W_{tk} is calculated using the NB method. The sentiment index in time period t, W_t, is calculated by

$$W_t = \sum_h (W_{tk} \times c), \tag{5}$$

where h is the number of reviews in the time period and c is a constant whose value is 1 or −1. The value of c depends on the category of W_{tk}. If W_{tk} belongs to a positive category, $c=1$; otherwise, $c=-1$. By using Eqs. (1) to (5), we can calculate the sentiment index for every generation product.

3.4. Forecasting model

In this study, the Bass model is extended to consider the sentiment index of online reviews. The Bass model assumes that the potential adopters can be divided into two categories, namely, innovators and imitators, and that the general form of the Bass model is as follows:

$$S(t) = m\frac{1-e^{-(p+q)t}}{1+\frac{q}{p}\times e^{-(p+q)t}}, \tag{6}$$

where $S(t)$ is the cumulative sales by the end of time period t, p refers to the coefficient of innovation, q refers to the coefficient of imitation, and m refers to the total number of potential adopters. The Bass model calculates the number of buyers or users, rather than the product sales, even though product sales can be estimated indirectly according to the frequency of consumer utilization. In such a situation, when the product is a durable consumer good, the number of buyers or users can be considered product sales.

In the existing research, studies often estimate the values of p, q, and m using the data of product sales directly. However, in this study, we use the data on product sales and online reviews to estimate parameters.

As discussed in section 3.2, q is related to the sentiment index and can be perceived as a function of the online review sentiment index $q=f(W_t)$. Given an increase in the sentiment index, if a product has received more praise from consumers, these reviews will influence more potential adopters to purchase the product. However, regarding a high sentiment index, the effect of the sentiment index on potential adopters is relatively small. For example, when the increase in the sentiment index is between 10,000 and 20,000, the increase in q is less significant than when the increase in the sentiment index is between 100 and 200. Thus, the function $q=f(W_t)$ forms an S-shaped curve. Based on the logistic-S model (Verhulst, 1938), the function $q=f(W_t)$ is an S-curve, as shown in Fig. 2. Based on Fig. 2, it is evident that q increases with the increase in W_t, and the second derivative of q monotone decreases as W_t increases. Therefore, the function $q=f(W_t)$ is described as

$$q = \frac{q^m q^0}{q^0 + (q^m - q^0)e^{-\gamma W_t}}, \tag{7}$$

where q denotes the effect of WOM via oral communications between people and online reviews, q^0 refers to the minimum of q, q^m refers to the maximum of q, and γ is a constant that controls the steepness of

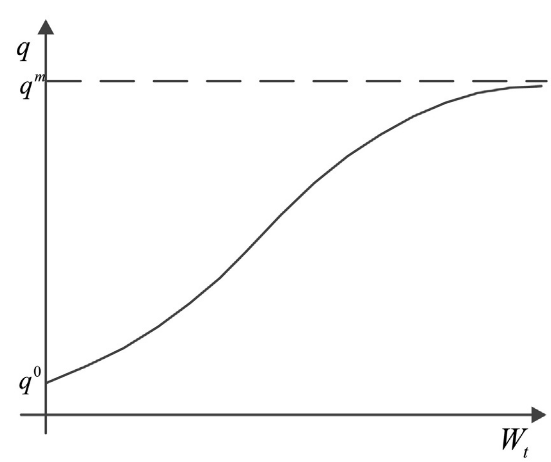

Fig. 2. The relationship between W_t and q.

Table 2
Cumulative sales of three generations of cars.

	Time period	Elantra	Elantra-l	Elantra-y
2007-Q3	1	208,487	0	0
2007-Q4	2	245,239	0	0
2008-Q1	3	288,619	0	0
2008-Q2	4	316,105	37,248	0
2008-Q3	5	334,164	58,649	0
2008-Q4	6	362,999	85,974	0
2009-Q1	7	400,655	132,620	0
2009-Q2	8	449,096	200,698	0
2009-Q3	9	492,761	266,406	0
2009-Q4	10	534,604	325,423	0
2010-Q1	11	580,675	383,975	0
2010-Q2	12	616,982	443,234	0
2010-Q3	13	653,678	498,503	0
2010-Q4	14	687,345	558,777	0
2011-Q1	15	722,173	612,881	0
2011-Q2	16	750,802	659,367	0
2011-Q3	17	777,631	710,580	0
2011-Q4	18	800,713	749,772	0
2012-Q1	19	827,039	794,515	0
2012-Q2	20	845,582	847,288	0
2012-Q3	21	855,965	910,312	26,856
2012-Q4	22	865,076	963,746	80,460
2013-Q1	23	871,629	1,016,473	132,341
2013-Q2	24	882,711	1,060,359	183,700
2013-Q3	25	891,458	1,097,173	239,631
2013-Q4	26	899,530	1,135,293	286,808
2014-Q1	27	907,671	1,176,867	342,190
2014-Q2	28	919,771	1,212,260	400,893
2014-Q3	29	927,477	1,239,208	465,947
2014-Q4	30	931,766	1,267,657	539,146

the S-curve. This study refers to the extended Bass model as the Bass-emotion model.

There are many ways to extend the Bass model, such as considering price factors (Robinson & Lakhani, 1975), the Norton model (Norton & Bass, 1987), combining with the marketing variable (Bass, Krishnan, & Jain, 1994), and the contingent diffusion model (Peterson & Mahajan, 1978). The Norton model is a typical model for multiple-generation products. The same method used for the Bass-emotion model can create the Norton-emotion model. When three-generation products are considered, the standard Norton model is

$$S_1(t) = F_1(t)m_1[1-F_2(t-\tau_2)] \quad \text{for } t>\tau_2, \tag{8a}$$

$$S_2(t) = F_2(t-\tau_2)[m_2 + F_1(t)m_1][1-F_3(t-\tau_3)] \quad \text{for } t>\tau_2, \tag{8b}$$

$$S_3(t) = F_3(t-\tau_3)[m_3 + F_2(t-\tau_2)[m_2 + F_1(t)m_1]] \quad \text{for } t>\tau_3, \tag{8c}$$

where $S_i(t)$ refers to the cumulative sales of the ith generation by the end of time period t, m_i refers to potential adopters to the ith generation, τ_i refers to the time when the ith generation is introduced, $F_i(t-\tau_i)=0$ if $t<\tau_i$, and $F_i(t)$ refers to the ith generation's cumulative fraction of adopters in time period t. $F_i(t)$ is calculated as follows:

$$F_i(t) = \frac{1-e^{-(p_i+q_i)t}}{1+\dfrac{q_i}{p_i}e^{-(p_i+q_i)t}}, \tag{9}$$

where p_i refers to the ith generation's coefficient of innovation and q_i is the imitation coefficient for the ith generation.

In the Norton model, q_i is also a function of the online review sentiment index. The sentiment index in time period t for the ith-generation product is computed as follows:

$$W_{it} = \sum_h (W_{itk} \times c), \tag{10}$$

where W_{itk} represents the sentiment index of reviews k for the ith-generation product in time period t, and h is the number of reviews.

The coefficient q_i can be calculated as.

$$q_i = f(W_{it}) = \frac{q_i^m q_i^0}{q_i^0 + (q_i^m - q_i^0)e^{-\gamma_i W_{it}}}, \tag{11}$$

where q_i^0 refers to the minimum of q_i for the ith generation, q_i^m refers to the maximum of q_i, and γ_i is a constant that controls the steepness of the curve to the ith generation.

3.5. Validation method and performance measure

To verify the fit between the forecasting model and the actual data, specific criteria are used to evaluate performance (Marshall et al., 2013; Dellarocas et al., 2007). R^2[1] and the root mean squared error (RMSE)[2] are used to measure the fit precision. To verify the effectiveness of the forecasting model, we use the mean absolute percentage error (*MAPE*) and percentage error (*PE*) to evaluate the performance of the model:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \dot{y}_i}{y_i}\right|, \tag{12a}$$

$$PE = \frac{\dot{y}_i - y_i}{y_i}. \tag{12b}$$

[1] $R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i-\overline{y}_i)^2}{\sum_{i=1}^{n}(y_i-\overline{y}_i)^2}$. where $\hat{y}_i$ refers to the fit value in the ith time period, y_i refers to the actual value in the ith time period, $\overline{y}_i$ refers to the average value in the ith time period, and n refers to the number of time periods.

[2] $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\dot{y}_i)^2}$. Where $\dot{y}_i$ refers to the forecasting value in the ith time periods.

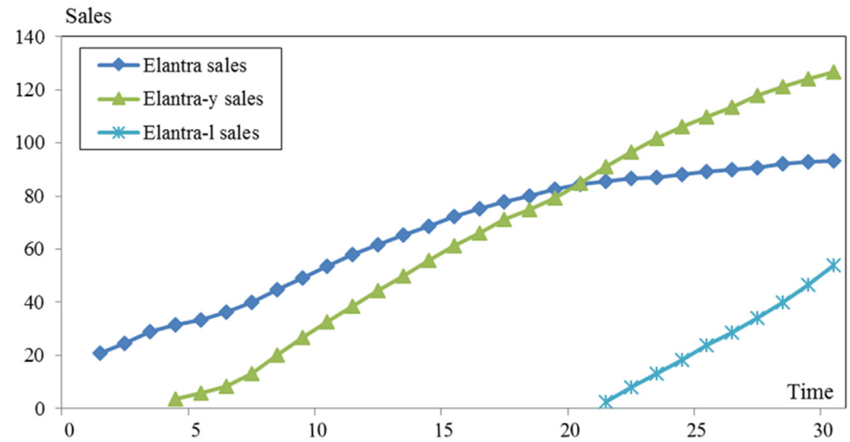

Fig. 3. The relationships between product sales and time for the three generations of cars.

To increase the intuition of the results of *PE*, we define $1-|PE|$ as the accuracy of forecasting:

$$1-|PE| = 1-\left|\frac{\dot{y}_i-y_i}{y_i}\right|. \tag{13}$$

4. Data and results

4.1. Data, experimental design and performance validation

Given its assumptions and concepts, the Bass model is suitable to forecast the sales of the product for which the number of users are approximately equal to the product sales (Wang et al., 2013; Speece & Maclachlan, 1995; Barnes et al., 2014; Seol et al., 2012; Song et al., 2015; Guo, 2014; Turk & Trkman, 2012). Regarding the Norton model, however, a product with multiple generations should be selected. Therefore, we chose automobiles as our research object.

As shown in Section 3.2, automobiles is a kind of search product, and thus the historical sales data and online reviews of each generation were collected. There are three generations of the Beijing Hyundai Elantra. Therefore, in this study, we referred to these three generations as the Elantra, Elantra-y and Elantra-l, and we collected sales data and online product reviews for these three generations from the automotive website Bitauto, the largest auto-trading platform in China.

The online review data collected for the Elantra are for the period from July 2007 to February 2015, for the Elantra-y from April 2008 to February 2015, and for the Elantra-l from April 2012 to March 2015. Similarly, the sales data for the Elantra cover the period from April 2006 to December 2014, for the Elantra-y from April 2008 to December 2014, and for the Elantra-l from August 2012 to December 2014. To ensure their consistency, sales and review data were chosen from the same time periods: data for the Elantra from July 2007 to December 2014, for the Elantra-y from April 2008 to December 2014, and for the Elantra-l from August 2012 to December 2014. We collected 1407, 2524 and 368 reviews for the Elantra, Elantra-y, Elantra-l, respectively. According to the existing literature (Marshall et al., 2013; Turk & Trkman, 2012), we defined the time period of the sales data as three months. The cumulative sales data are presented in Table 2. The relationship that product cumulative sales change over time is shown in Fig. 3.

4.2. Sentiment classification

Before demand forecasting using the Bass-emotion and Norton-emotion models, three commonly used sentiment classification models, i.e., NB, SVM and KNN, are used to classify online reviews of the three generations of cars. The toolboxes of Matlab R2014 were used to implement these three models. A three-fold cross-validation technique is used to obtain the final results. In each fold, the training set is further divided into two parts to select the free parameters of the three models. The results on the independent testing set is evaluated by three criteria, i.e., the overall correct classification ratio (Accuracy), the correct classification ratio of reviews with positive sentiment (Sensitivity) and the correct classification ratio of reviews with negative sentiment (Specificity).

The average results on the independent testing set by cross-validation are shown in Table 3. As shown in Table 3, the NB obtains slightly better results than or approximately the same results as the SVM and KNN on the testing data of the three generations. Because the NB is easily to be implemented and has good performance on sentiment classification, it is chosen as the classification model in this study.

4.3. Parameter estimation and forecasting accuracy

In this study, we use the First Optimization software package to estimate the Bass-emotion and Norton-emotion models.

With respect to the Bass-emotion model, we must estimate five parameters, m, p, q^0, q^m, and γ in Eqs. (4) and (5). Regarding the three generations of products, we use data on the 2nd-generation product to estimate the Bass-emotion model. The known conditions of this model are presented in Table 4. According to the actual situation and the definitions of the parameters in the Bass and logistic-S models, we define the value range of parameters in Table 5. According to the known conditions and the value range of parameters, the results of the parameter estimations of the Bass-emotion model are presented

Table 3
Sentiment classification results of the three generations of cars using the NB, SVM and KNN.

Generation	Model	Accuracy	Sensitivity	Specificity
Elantra	NB	59.92	66.71	51.94
	SVM	55.24	41.70	69.15
	KNN	57.62	65.55	49.47
Elantra-l	NB	63.41	82.76	31.63
	SVM	55.16	77.89	32.66
	KNN	64.60	85.72	28.08
Elantra-y	NB	75.87	89.65	44.34
	SVM	65.16	69.15	55.90
	KNN	66.90	82.57	34.85

Table 4
Known conditions of the Bass-emotion model.

No.	Calculation parameter	Value
1	Number of parameter estimates	5
2	Number of known data	27
3	Number of used data	27

in Table 6. The R^2 and *RMSE* values of the Bass-emotion model are 0.9987 and 1.4910, respectively.

For the Norton-emotion model, we must estimate 15 parameters, i.e., m_1, m_2, m_3, p_1, p_2, p_3, q_1^0, q_2^0, q_3^0, q_1^m, q_2^m, q_3^m, γ_1, γ_2 and γ_3. The known conditions of the Norton-emotion model are detailed in Table 7, and the value ranges of the parameters are presented in Table 8. The estimations of the parameters of the Norton-emotion model are presented in Table 9. In addition, the values of R^2 and the RMSE statistics for three generations are presented in Table 10.

To visually reflect the model's simulation results, the fitted values and the actual values of the Bass-emotion and Norton-emotion models are presented in Figs. 4 and 5, respectively.

To further illustrate the effectiveness of the model, as the forecasting targets, we use the sales values for time periods 23 to 27 for the Bass-emotion model and for time periods 26 to 30 for the Norton-emotion model as presented in Table 2. Furthermore, we use data containing 25 time periods before the targeted time period to train the Bass-emotion and Norton-emotion models. The comparisons between forecasting and actual values of the Bass-emotion and Norton-emotion models are presented in Tables 11 and 12, respectively. We also calculate the forecasting precision for every forecasting period.

4.4. Model comparison

In extant studies, few methods for product sales forecasting use both historical sales data and online review data. The existing methods differ from those used in this study, including the form of the data and the structure of the model. The proposed method was compared with the original Bass model, which used only historical sales data. Moreover, in order to compare with the study of Dellarocas et al. (2007), average ratings of online reviews are used instead of sentiment index in Eqs. (7) and (11) of the proposed model to forecast product sales. The Bass/Norton model using the ratings is called the Bass-rating/Norton-rating model. In addition, we make a comparison with another forecasting model, i.e., log-linear model (Ye et al., 2009), that considers sentiment showing in online reviews data. Here, the online review and historical sales data for the Elantra-l were used to fit this model. The forecasting results of the proposed models are compared with those of the standard Bass and Norton models, those of the Bass-rating and Norton-rating models and the log-linear model.

The forecasting values of the Bass and Norton models using the same experimental design as shown in Section 4.3 are presented in Tables 13 and 14. The forecasting values of Bass-rating, Norton-rating and log-linear models are presented in Tables 15, 16 and 17. The average accuracy of the Bass, Bass-rating, Bass-emotion and log-linear models is 0.9933, 0.9936, 0.9946 and 0.9413. The average accuracy of the Norton, Norton-rating and Norton-emotion models is 0.9071, 0.9423 and 0.9647. Compared with the forecasting values presented in Tables 11 and 12, the forecasting results of the Bass-emotion (Norton-emotion) model are more accurate than those of the Bass (Norton) and Bass-rating (Norton-rating) models.

Finally, we choose the last five data points as the forecasting points and the rest of the data points as the parameter estimation points. The values of *MAPE*for each model are presented in Table 18. From the

Table 5
Value range of the parameters.

No.	Parameter	Value range
1	m	[20, +∞]
2	p	[0, 1]
3	q^0	[0, 1]
4	q^m	[0, 1]
5	γ	[0, +∞]

Table 6
Results of the parameter estimations in the Bass-emotion model.

No.	Parameter	Results
1	m	156.0306
2	p	0.023777
3	q^0	0.090407
4	q^m	0.093113
5	γ	0.170784

Table 7
Known conditions of the Norton-emotion model.

No.	Calculation parameter	Value
1	Number of parameter estimates	15
2	Number of known data	30
3	Number of used data	30
4	Time 2nd generation is introduced	$\tau_2=4$
5	Time 3rd generation is introduced	$\tau_3=21$

Table 8
Parameter value range in the Norton-emotion model.

No.	Parameter	Value range
1	m_i	[20, +∞]
2	p_i	[0, 1]
3	q_i^0	[0, 1]
4	q_i^m	[0, 1]
5	γ_i	[0, +∞]

Table 9
Results of the parameter estimations in the Norton-emotion model.

No.	Parameter	Results
1	m_1	119.92702
2	p_1	0.0647021
3	q_1^0	1.29E-12
4	q_1^m	0.9999994
5	γ_1	2.2371958
6	m_2	1735.3944
7	p_2	0.0019562
8	q_2^0	0.0441395
9	q_2^m	0.6767519
10	γ_2	1.83E-13
11	m_3	20.000011
12	p_3	0.0146138
13	q_3^0	0.1614071
14	q_3^m	0.2093965
15	γ_3	2851.857243

Table 10
Statistics of the Norton-emotion model.

Endogenous variable	R^2	*RMSE*
$S_1(t)$	0.9931	3.1873
$S_2(t)$	0.9980	2.7580
$S_3(t)$	0.9932	3.3278

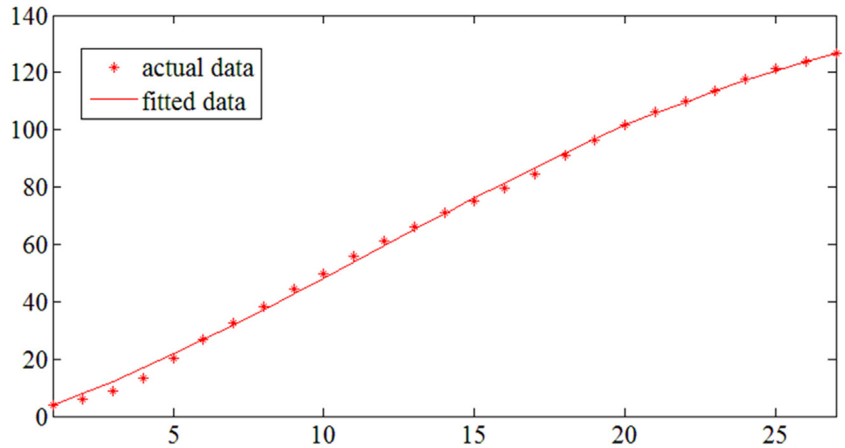

Fig. 4. Forecasting results from the Bass-emotion model.

values presented in Table 18, we find that the *MAPE* of the Bass-emotion (Norton-emotion) is smaller than that of other comparative models. As the forecasting accuracy of the Bass-emotion (Norton-emotion) models using online review data is stronger, using the content of online review data can improve prediction accuracy.

4.5. robustness examination using another dataset

To explore the robustness of this paper's research methods, the historical sales data and online reviews of Volkswagen's three generations (Jetta, Sagitar and Bora) are used. The online review data and sales data collected for the Jetta are for the period from January 2003 to March 2016, for the Sagitar from April 2006 to March 2016, and for the Bora from April 2008 to March 2016. We collected 254, 411 and 2485 reviews for the Jetta, Sagitar and Bora, respectively.

The forecasting results are presented in Tables 19–22. The average accuracy of the Bass and Bass-emotion models are 0.989 and 0.991, and that of the Norton and the Norton-emotion model is 0.969 and 0.984. From the results, it can be found that the proposed model obtains robust results on the two automobile, i.e., Beijing Hyundai Elantra and Volkswagen, datasets. The values of *MAPE*for each model are presented in Table 23.

5. Conclusion

In this paper, a forecasting model that combines the Bass/Norton model and sentiment analysis techniques is proposed. In contrast to the extant literature that uses online ratings, this paper extends the Bass model by analyzing sentiments expressed in online reviews. In contrast to the original Bass model, both historical sales and online review data are directly used in the extended model. The NB method is adopted to calculate the sentiment index and conduct polarity classifications for each online review, and the extracted sentiment index is used to expand the imitation coefficient in the Bass model. The same method is used to expand the Norton model. Sentiment information is rarely used to extend the Bass model in existing studies.

We use actual sales and online review data of automobiles to evaluate forecasting accuracy. The forecasting accuracy of the proposed models, i.e., the Bass-emotion and Norton-emotion models, is then compared with the standard Bass and Norton models, the Bass-rating and Norton-rating models and the log-linear model. The results indicate that the proposed models exhibit lower forecasting errors than the comparative models mentioned above. Moreover, we collect the online review and sales data of other types of automobiles, i.e., Volkswagen's three generations, and verify the robustness of the proposed models by computations. In addition, we compare the results of three

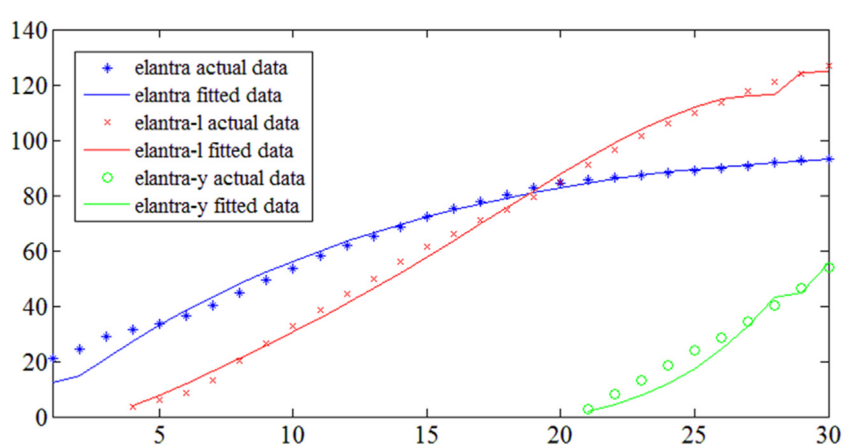

Fig. 5. Forecasting results from the Norton-emotion model.

Table 11
Forecasting data from the Bass-emotion model.

Forecasting period	23	24	25	26	27
Actual value	113.5293	117.6867	121.226	123.9208	126.7657
Forecasting value	112.6674	116.5294	120.1996	123.7927	126.8046
1-|PE|	0.9924	0.9901	0.9915	0.9990	0.9997

Table 12
Forecasting data from the Norton-emotion model.

Forecasting period	26	27	28	29	30
Elantra actual value	89.948	90.763	91.973	92.743	93.172
Elantra forecasting value	90.213	91.046	91.128	92.919	93.56
Elantra 1-|PE|	0.997	0.997	0.991	0.998	0.996
Elantra-l actual value	113.529	117.687	121.226	123.921	126.766
Elantra-l forecasting value	114.023	111.401	118.797	120.962	121.162
Elantra-l 1-|PE|	0.996	0.944	0.980	0.976	0.954
Elantra-y actual value	28.681	34.219	40.089	46.595	53.915
Elantra-y forecasting value	28.009	37.809	42.088	49.407	60.552
Elantra-y 1-|PE|	0.976	0.905	0.953	0.943	0.890

Table 13
Forecasting data from the Bass model.

Forecasting period	23	24	25	26	27
Actual value	113.5293	117.6867	121.226	123.9208	126.7657
Forecasting value	112.4249	116.311	120.1491	123.6808	126.6576
1-|PE|	0.990	0.988	0.991	0.998	0.999

Table 14
Forecasting data from the Norton model.

Forecasting period	26	27	28	29	30
Elantra actual value	89.948	90.763	91.973	92.743	93.172
Elantra forecasting value	90.636	91.439	92.124	92.804	93.414
Elantra 1-|PE|	0.992	0.993	0.998	0.999	0.997
Elantra-l actual value	113.529	117.687	121.226	123.921	126.766
Elantra-l forecasting value	103.494	106.995	111.261	115.252	118.653
Elantra-l 1-|PE|	0.903	0.900	0.910	0.925	0.932
Elantra-y actual value	28.681	34.219	40.089	46.595	53.915
Elantra-y forecasting value	33.857	42.174	48.559	55.391	62.835
Elantra-y 1-|PE|	0.847	0.811	0.826	0.841	0.858

Table 15
Forecasting data from the Bass-rating model.

Forecasting period	23	24	25	26	27
Actual value	113.5293	117.6867	121.2260	123.9208	126.7657
Forecasting value	112.4337	116.3928	120.1446	123.7248	126.7046
1-|PE|	0.9903	0.9889	0.9910	0.9984	0.9995

Table 16
Forecasting data from the Norton-rating model.

Forecasting period	26	27	28	29	30
Elantra actual value	89.948	90.763	91.973	92.743	93.172
Elantra forecasting value	90.498	91.399	92.174	92.813	93.632
Elantra 1-|PE|	0.9939	0.9930	0.9978	0.9992	0.9951
Elantra-l actual value	113.529	117.687	121.226	123.921	126.766
Elantra-l forecasting value	107.545	108.884	117.334	118.997	120.223
Elantra-l 1-|PE|	0.9444	0.9192	0.9668	0.9586	0.9456
Elantra-y actual value	28.681	34.219	40.089	46.595	53.915
Elantra-y forecasting value	30.402	41.874	46.561	51.267	60.182
Elantra-y 1-|PE|	0.9434	0.8172	0.8610	0.9089	0.8959

Table 17
Forecasting data from the log-linear model.

Forecasting period	23	24	25	26	27
Actual value	113.5293	117.6867	121.226	123.9208	126.7657
Forecasting value	100.6333	113.0830	115.0831	119.4967	122.5828
1-|PE|	0.8719	0.9593	0.9466	0.9630	0.9659

Table 18
Comparison of MAPE values.

Model	Generations	MAPE
Norton model	Elantra	0.0040
	Elantra-l	0.0791
	Elantra-y	0.1957
Norton-rating model	Elantra	0.0042
	Elantra-l	0.0502
	Elantra-y	0.1323
Norton-emotion model	Elantra	0.0043
	Elantra-l	0.0292
	Elantra-y	0.0723
Bass model	Elantra-l	0.0066
Bass-rating model	Elantra-l	0.0063
Bass-emotion model	Elantra-l	0.0054
Log-linear model	Elantra-l	0.0544

Table 19
Forecasting data from the Bass model.

Forecasting period	49	50	51	52	53
Actual value	136.5508	142.7008	148.9708	157.2308	165.8508
Forecasting value	134.2138	142.2506	150.5591	159.1318	167.9592
1-|PE|	0.983	0.997	0.989	0.988	0.987

Table 20
Forecasting data from the Bass-e model.

Forecasting period	49	50	51	52	53
Actual value	136.5508	142.7008	148.9708	157.2308	165.8508
Forecasting value	134.5258	142.0258	150.1621	158.8549	167.2136
1-|PE|	0.985	0.995	0.992	0.990	0.992

Table 21
Forecasting data from the Norton model.

Forecasting period	49	50	51	52	53
Jetta actual value	258.2172	264.9872	270.0000	277.5572	285.6372
Jetta forecasting value	255.7322	262.2331	268.7070	275.1464	281.5438
Jetta 1-|PE|	0.990	0.989	0.995	0.991	0.985
Sagitar actual value	136.5508	142.7008	148.9708	157.2308	165.8508
Sagitar forecasting value	133.4939	141.4315	149.6437	158.1253	166.8695
Sagitar 1-|PE|	0.977	0.991	0.996	0.994	0.994
Bora actual value	125.7851	129.7751	134.3551	140.6451	144.8551
Bora forecasting value	125.3101	134.7505	144.7672	155.3923	166.6593
Bora 1-|PE|	0.996	0.963	0.928	0.905	0.869

Table 22
Forecasting data from the Norton-e model.

Forecasting period	49	50	51	52	53
Jetta actual value	258.2172	264.9872	270.0000	277.5572	285.6372
Jetta forecasting value	256.0276	262.5468	269.0735	275.5362	281.9613
Jetta 1-|PE|	0.991	0.991	0.997	0.993	0.987
Sagitar actual value	136.551	142.701	148.971	157.231	165.851
Sagitar forecasting value	133.715	141.937	150.520	159.443	168.689
Sagitar 1-|PE|	0.979	0.995	0.990	0.986	0.983
Bora actual value	125.7851	129.7751	134.3551	140.6451	144.8551
Bora forecasting value	125.1989	131.6144	138.2410	144.9909	152.2655
Bora 1-|PE|	0.995	0.986	0.972	0.970	0.951

Table 23
Comparison of MAPE values.

Model	Generations	MAPE
Norton model	Jetta	0.0096
	Sagitar	0.0095
	Bora	0.0750
Norton-emotion model	Jetta	0.0083
	Sagitar	0.0135
	Bora	0.0260
Bass model	Sagitar	0.0111
Bass-emotion model	Sagitar	0.0092

commonly used sentimental classification methods, i.e., the NB, SVM and KNN, and find that the NB has good performance on sentiment classification.

However, the approach proposed in this paper has limitations that suggest further research is needed. As suggested by the existing literature (Yu et al., 2012; Dellarocas et al., 2007), we use only the sentiment index extracted from the content of the online reviews. For further research, additional attributes, such as the number of users who agree or disagree with the reviews and the number of times other users browsed the reviews, can be used to calculate the imitation coefficient in the Bass model and thereby improve forecasting accuracy.

Acknowledgements

This work was partly supported by the National Natural Science Foundation of China (Project Nos. 71471035 and 71571039) and the 111 Project (B16009).

References

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509.
Asur, S., & Humberman, B. A. (2010). *Predicting the future with social media* (pp. 492–499)-Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) Toronto, ON.
Barnes, B., Southwell, D., Bruce, S., et al. (2014). Additionality, common practice and incentive schemes for the uptake of innovations. *Technological Forcasting & Social Change*, *89*(1), 43–61.
Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, *15*(5), 215–227.
Bass, F. M. (2004). Comments on "a new product growth model for consumer durables". *Management Science*, *50*(12), 1833–1840.
Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass model fits without decision variables. *Management Science*, *13*(3), 203–223.
Bollen, J., Mao, H., & Zeng, X. J. (2011). Twitter mood predicts the stock market. *Journal of Computer Science*, *2*(1), 1–8.
Cabral, L., & Hortacsu, A. (2003). *The dynamics of seller reputation: Theory and evidence from e-Bay. Working paper.* University of Chicago https://ssrn.com/abstract=541161.
Cao, Q., Thompson, M., & Yu, Y. (2013). RETRACTED: Sentiment analysis in decision sciences research: An illustration to IT governance. *Decision Support Systems*, *54*(2), 1010–1015.
Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354.
Culotta, A. (2010). Towards detecting influenza epidemics by analyzing twitter messages. *Proceedings of the first workshop on social media analytics (SOMA'10)* (pp. 115–122). New York: ACM.
Dellarocas, C., Awad, N. F., & Zhang, X. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, *21*(4), 23–45.
Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007–1016.
Fernández-Durán, J. J. (2014). Modeling seasonal effects in the bass forecasting diffusion model. *Technological Forcasting & Social Change*, *88*, 251–264.
Gao, X., Chai, H., & Tang, S. (2013). Study on innovation diffusion of hybrid electric vehicle based on patent citation data. *World Sci-tech R & D*, *35*(6), 784–787 (in Chinese).
Godes, D., & Mayzlin, D. (2004). Using online conversations to study word of mouth communication. *Marketing Science*, *23*(4), 545–560.
Guo, X. (2014). A novel Bass-type model for product life cycle quantification using aggregate market data. *International Journal of Production Economics*, *158*, 208–216.
Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research*, *17*(4), 454–462.
Hyman, & Michael, R. (1988). The timeliness problem in the application of Bass-type new product-growth models to durable sales forecasting. *Journal of Business Research*, *16*(1), 31–47.
Lee, Y. (2009). Internet election 2.0? Culture, institutions, and technology in the Korean presidential elections of 2002 and 2007. *Journal of Information Technology & Politics*, *6*(3), 312–325.
Lee, H., Kim, S. K., Park, H., & Kang, P. (2014). Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forcasting & Social Change*, *86*, 49–64.
Li, J. (2012). An empirical study of the impact of online reviews on product sales. *Journal of Modern Information*, *32*(1), 164–167 (in Chinese).
Liu, Y. (2006). Word-of-mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*(3), 74–89.
Liu, Y., Huang, X., An, A., & Yu, X. (2007). *ARSA: A sentiment-aware model for predicting sales performance using blogs* (pp. 607–614)Proceedings of 30th annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM.
Marshall, P., Dockendorff, M., & Ibáñez, S. (2013). A forecasting system for movie attendance. *Journal of Business Research*, *66*(10), 1800–1806.
Mentzer, J. T., & Bienstock, C. C. (1998). *Sales forecasting management: Understanding the techniques, systems, and management of the sales forecasting process.* Sage Publications.
Nelson, P. J. (1974). Advertising as information. *Journal of Political Economy*, *82*(4), 729–754.
Norton, J. A., & Bass, F. M. (1987). A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science*, *33*(9), 1069–1086.
Pang, B., & Lee, L. (2005). *Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales* (pp. 115–124)Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics.
Peterson, R. A., & Mahajan, V. (1978). Multi-product growth models. In J. Sheth (Ed.), *Research in Marketing* (pp. 201–231).
Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), 143–157.
Robinson, B., & Lakhani, C. (1975). Dynamic price models for new product planning. *Management Science*, *21*(10), 1113–1122.
Ryu, K., & Han, H. (2010). Influence of the quality of food, service, and physical environment on customer satisfaction and behavioral intention in quick-casual restaurants: Moderating role of perceived price. *Journal of Hospitality and Tourism Research*, *34*(3), 310–329.
Segal, J., Sacopulos, M., Sheets, V., Thurston, I., Brooks, K., & Puccia, R. (2012). Online doctor reviews: Do they track surgeon volume, a proxy for quality of care? *Journal of Medical Internet Research*, *14*(2), e50.
Seol, H., Park, G., Lee, H., & Yoon, B. (2012). Demand forecasting for new media services with consideration of competitive relationships using the competitive Bass model and the theory of the niche. *Technological Forcasting & Social Change*, *79*(7), 1217–1228.
Shi, X., Bigdeli, A. Z., & Li, F. (2015). An examination of NPD models in the context of business models. *Journal of Business Research*, *69*(7), 2541–2550.
Shoemaker, M. H. (2012). *A Bass diffusion model analysis: Understanding alternative fuel vehicle sales.* CMC Senior Theses. http://scholarship.claremont.edu/cmc_theses/399.
Song, Y., Lee, S., Zo, H., & Lee, H. (2015). A hybrid Bass-Markov model for the diffusion of a dual-type device-based telecommunication service: The case of WiBro service in Korea. *Computers & Industrial Engineering*, *79*, 85–94.
Speece, M. W., & Maclachlan, D. L. (1995). Application of a multi-generation diffusion model to milk container technology. *Technological Forcasting & Social Change*, *49*(3), 281–295.
Taylor, J. (2003). Word of mouth is where it's at. *Brandweek*, *44*(22), 26–27.
Turk, T., & Trkman, P. (2012). Bass model estimates for broadband diffusion in European countries. *Technological Forcasting & Social Change*, *79*(1), 85–96.
Verhulst, P. F. (1938). Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique*, *10*, 113–121.
Wang, F. K., Chang, K. K., & Hsiao, Y. Y. (2013). Implementing a diffusion model optimized by a hybrid evolutionary algorithm to forecast notebook shipments. *Applied Soft Computing*, *13*(2), 1147–1151.
Wu, F. S., & Chu, W. L. (2010). Diffusion models of mobile telephony. *Journal of Business Research*, *63*(5), 497–501.
Wu, M., Wang, L., & Li, M. (2015). An approach based on the bass model for analyzing the effects of feature fatigue on customer equity. *Computational & Mathematical Organization Theory*, *21*(1), 69–89.
Wysocki, P. D. (2000). *Private information, earnings announcements and trading volume, or stock chat on the internet: A public debate about private information. Working paper.* University of Michigan Business School https://www.researchgate.net/publication/228431130_Private_Information_Earnings_Announcements_and_Trading_Volume_or_Stock_Chat_on_the_Internet_A_public_Debate_about_Private_Information_Working_Paper.
Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180–182.
Yu, S., & Kak, S. (2012). A survey of prediction using social media. *Computer science* http://arxiv.org/abs/1203.1647v1.
Yu, X., Liu, Y., Huang, J. X., & An, A. (2012). Mining online reviews for predicting sales performance: A case study in the movie domain. *IEEE Transactions on Knowledge and Data Engineering*, *24*(4), 720–734.
Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, *55*(4), 919–926.

Zhu, Y., Jiang, Z. B., & Chen, L. (2008). Analysis on 3M new product diffusion and regression based on Bass model. *Logistics Technology*, *27*(11), 77–79.

Zhi-Ping Fan received the Ph.D. degree in control theory and applications from Northeastern University (NEU), Shenyang, China, in 1996. He is currently a Professor in the Department of Information Management and Decision Sciences, School of Business Administration, NEU. He was Research Fellow at City University of Hong Kong, Kowloon, Hong Kong, in 2001, 2003, 2004 and 2005, respectively. He is the author or coauthor of over 80 refereed articles published in international journals including the IEEE Transactions on Systems, Men and Cybernetics, the IEEE Transactions on Engineering Management, the Information Sciences, and the European Journal of Operational Research. His current research interests include decision analysis and operations research.

Yu-Jie Che is a M.Sc. candidate from the Northeastern University, China. His research interests include text mining, demand and technological forecasting.

Zhen-Yu Chen received the Ph.D. degree from the University of Chinese Academy of Sciences in 2008. He is an associate professor in the Department of Information Management and Decision Sciences, School of Business Administration, Northeastern University, Shenyang, China. His research interests include data mining, electronic commerce and customer relationship management. He has coauthored some journal publications including the European Journal of Operational Research, the Knowledge and Information Systems, the Knowledge-based Systems, the Expert Systems with Applications, the Information Sciences, etc.