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 2 February 2017

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.

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/

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.