ARTICLE IN PRESS

TFS-17819; No of Pages 16

Technological Forecasting & Social Change xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach

Hakyeon Lee [a], Sang Gook Kim [b], Hyun-woo Park [b], Pilsung Kang [a,*]

[a] Department of Industrial and Systems Engineering, Seoul National University of Science and Technology, 172 Gongreung 2-dong, Nowon-gu, Seoul 139-746, South Korea
[b] Korea Institute of Science and Technology Information (KISTI), 66 Hoegi-ro, Dongdaemun-gu, Seoul 130-741, South Korea

ARTICLE INFO

Article history:
Received 23 April 2013
Received in revised form 12 August 2013
Accepted 17 August 2013
Available online xxxx

Keywords:
Pre-launch forecasting
Bass model
Multivariate linear regression
Machine learning
Ensemble

ABSTRACT

This study proposes a novel approach to the pre-launch forecasting of new product demand based on the Bass model and statistical and machine learning algorithms. The Bass model is used to explain the diffusion process of products while statistical and machine learning algorithms are employed to predict two Bass model parameters prior to launch. Initially, two types of databases (DBs) are constructed: a product attribute DB and a product diffusion DB. Taking the former as inputs and the latter as outputs, single prediction models are developed using six regression algorithms, on the basis of which an ensemble prediction model is constructed in order to enhance predictive power. The experimental validation shows that most single prediction models outperform the conventional analogical method and that the ensemble model improves prediction accuracy further. Based on the developed models, an illustrative example of 3D TV is provided.

1. Introduction

Increasing market uncertainty is making forecasting new product demand more difficult than ever, while shorter product lifecycles are forcing managers to produce new product demand forecasts more frequently. Correct forecasts are a key determinant of the success of new products, but accurate forecasting can be challenging. Forecasting demand for a mature-stage product is unproblematic since enough historical sales data are available. However, forecasting the future sales of a new product that has a short history or no history at all is complicated. Indeed, as Bass et al. [1] note, "the most important forecast is the forecast prior to launch," as it drives important pre-launch decisions such as capital equipment purchases and capacity planning [2]. Overestimating demand can result in excess inventories, while underestimation may incur significant opportunity costs and reduce market share.

Despite its importance, however, little progress has been made in pre-launch forecasting in the literature. Product demand forecasting has mainly been examined in the context of innovation diffusion. The diffusion of an innovation is "the process by which an innovation is communicated through certain channels over time among the members of a social system" [3]. Since its introduction to marketing studies in the 1960s, a variety of models have been developed to empirically model the diffusion of innovations [4]. The main thread of the diffusion models has been based on the pioneering work of Bass [5]. The assumption underlying the Bass model is that an innovation is spread through two types of communication channels: mass media (external influence) and word-of-mouth (internal influence). Over the past 40 years, the Bass model has thus enjoyed a number of applications because it has a relatively high explanatory power despite its simple structure [6].

Pre-launch forecasting also has been investigated mainly based on the Bass model. Previous studies of pre-launch forecasting with the Bass model have struggled against how to estimate the model parameters of a new product when little or no historical sales data are available. Several methods

* Corresponding author. Tel.: +82 2 970 7279; fax: +82 2 974 2849.
E-mail address: pskang@seoultech.ac.kr (P. Kang).

0040-1625/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.techfore.2013.08.020

have been developed, which can be categorized into three types: Bayesian approaches, subjective approaches, and analogical approaches. The Bayesian approach starts with pre-launch forecasts and updates them as additional data become available. Various methods for updating the parameter estimates or the forecasts have been proposed [7–10]. However, this approach centers on how to update forecasts after launch and still calls for initial pre-launch forecasts to be updated as data become available. The subjective approach produces parameter estimates through an algebraic estimation procedure on the basis of managerial judgments of tangible information such as the time and level of peak sales [11] and the sum of the coefficients of the external and internal influences [12]. The drawback of the subjective approach is that obtaining accurate judgments is as difficult as estimating accurate parameters [13]. Finally, the analogical approach, called "guessing by analogy," has prevailed in the literature. It assumes that a new product will have a diffusion pattern similar to those of its analogous products over time [1]. Under this approach, the parameter estimates of a new product are obtained by taking a weighted sum of the parameters of analogous products, with the weights derived by establishing similarities between the new product and several analogous products [14].

Although the analogical approach has been applied widely [1,15–19], it has two main limitations. First, there are no clear guidelines for how to select benchmarks even though the estimated parameters are highly dependent on the analogous products under consideration. Second, the similarities are established by expert judgments that are naturally subjective in nature. A promising solution to these problems is using the historical empirical relationship between the parameters and attributes of analogous products. Rogers [3] emphasizes that the attributes of an innovation are important variables in explaining its rate of adoption. Once this relationship is identified, the parameters of a new product can be estimated by knowing its characteristics [20]. Although identifying the relationship between diffusion parameters and product attributes can serve a reliable basis for pre-launch forecasting of a new product, however, little research has been carried out thus far in this direction and no systematic approach has yet been developed.

The tenet of this study is that a statistical and machine learning-based approach can overcome the limitations of the conventional analogical approach to pre-launch forecasting. The goal of statistical and machine learning is to discover intrinsic, sometimes unanticipated, relationships between variables with the help of high computational power [21,22]. A typical procedure of an inductive statistical and machine learning approach is as follows. The first step is to set up the model structure by defining a learning task, configuring input–output variables, seeking appropriate algorithms, and selecting proper performance criteria. The second step is to collect sufficient real-world examples, which are then divided into training and test data sets. In the third step, the employed learning algorithms are optimized on the basis of the training data set. Finally, the best model is identified on the basis of the test data set, using the predetermined performance criteria. In this study, the learning task is defined as predicting the parameters of the Bass model prior to launch, while input and output variables are configured as product attributes and diffusion characteristics, respectively. Various regression algorithms, such as multilayer linear regression, support vector regression, and Gaussian process regression, as well as an ensemble model of these, are used in building the prediction models. Finally, their performances are evaluated using the mean absolute error (MAE) and the root mean squared error (RMSE).

Pre-launch product demand forecasting on the basis of statistical and machine learning algorithms has several advantages over conventional analogical methods. First, a reliable relationship between the attributes and diffusion characteristics of existing products can be found, which in turn enables new product demand forecasts to be based solely upon these attributes without any human manipulation. In other words, analogous products are automatically selected and their contributions to forecasting systematically determined by the prediction model; although such selection and determination processes are not easily understood by humans, they are mathematically sound and analytically tractable. Therefore, forecasting is no longer dependent on the subjective judgments of human experts, but becomes an objective outcome obtained by the combination of learning algorithms and product data. Second, since statistical and machine learning algorithms are designed for interpolation as well as extrapolation, forecasting accuracy can be improved. The parameter values predicted by conventional analogical methods are bounded by the current maximum and minimum estimates of reference products. If a new type of diffusion style occurred, whose parameter values were far from the current boundary, conventional analogical methods would not properly reflect this eventuality. Statistical and machine learning-based approaches, by contrast, would scent the change from the inside of the product and digest it into prediction. In light of the foregoing, this study proposes a new approach to the pre-launch forecasting of new product demand, which utilizes the Bass diffusion model and statistical and machine learning-based regression algorithms. In addition, we also boost prediction accuracy by constructing an ensemble of individual prediction models.

The remainder of this paper is organized as follows. Section 2 reviews previous studies of pre-launch forecasting with the Bass model. Section 3 demonstrates the proposed framework including the product and diffusion database (DB) design, single prediction model development, and ensemble model construction. Section 4 validates the single and ensemble prediction models and provides an illustrative case study on the basis of the best single and ensemble models. Finally, the conclusion and limitations of this paper are presented alongside future research directions in Section 5.

2. Bass model and pre-launch forecasting

The Bass model assumes that a technological innovation is spread by two types of influences: externally by the mass media and internally by word-of-mouth. It can be derived from a hazard function that represents the probability that an adoption occurs at time t given that it has not yet occurred:

$$h(t) = \frac{f(t)}{1-F(t)} = p + qF(t) \tag{1}$$

where $f(t)$ is the density function of time to adoption, $F(t)$ is the cumulative proportion of adopters at time t, and p and q

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020