Decision Support Systems 55 (2013) 247–255

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

A multivariate intelligent decision-making model for retail sales forecasting

Z.X. Guo [a,*], W.K. Wong [b], Min Li [b]

[a] Business School, Sichuan University, Chengdu 610065, PR China
[b] Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong, China

ARTICLE INFO

Article history:
Received 11 August 2011
Received in revised form 28 August 2012
Accepted 21 January 2013
Available online 12 February 2013

Keywords:
Retail industry
Early sales
Sales forecasting
Multivariate forecasting

ABSTRACT

A sales forecasting problem in the retail industry is addressed based on early sales. An effective multivariate intelligent decision-making (MID) model is developed to provide effective forecasts for this problem by integrating a data preparation and preprocessing module, a harmony search-wrapper-based variable selection (HWVS) module and a multivariate intelligent forecaster (MIF) module. The HWVS module selects out the optimal input variable subset from given candidate inputs as the inputs of MIF. The MIF is established to model the relationship between the selected input variables and the sales volumes of retail products, and then utilized to forecast the sales volumes of retail products. Extensive experiments were conducted to validate the proposed MID model in terms of extensive typical sales datasets from real-world retail industry. Experimental results show that it is statistically significant that the proposed MID model can generate much better forecasts than extreme learning machine-based model and generalized linear model do.

1. Introduction

Retail sales forecasting is estimating the future demand of a retail product, which is essential to sound business planning [1] and is playing a more and more prominent role in dynamic supply chain facing the ever-intensifying global competition [33]. Sales forecasting is the foundation on which company plans are built [20], which is essential to improve a retail company's competition capacity [11]. This paper addresses a multivariate sales forecasting problem, which forecasts the overall sales of a retail product based on its early sales volume.

1.1. Sales forecasting

The history of sales forecasting can be traced back to more than 50 years ago [1,32]. Since then a large number of sales forecasting papers have been published [4,5,21,24,32,33], which involves a wide variety of applications in real-world industries, such as print circuit board industry [4], food industry [5] and apparel industry [14].

Some researchers investigated the sales forecasting problem by considering it as univariate time series forecasting problems [23,33]. Their studies utilized input data directly from the historical sales data of time series being forecasted, which are usually based on a basic assumption that the underlying data-generating process of the time series is constant. This assumption is usually invalid in real world since a variety of factors influencing product sales, called influencing factors, may cause the uncertain change of data pattern, particularly in a dynamic and quick response business environment such as apparel and footwear industry. As a result, the univariate forecasting model cannot handle sudden changes caused by various influencing factors such as product attributes and economic environment.

To handle this, some researchers integrated the sales forecasting problem by using multivariate inputs including historical sales and related influencing factors [4,5]. Chang et al. [4] addressed the monthly sales forecasting problem to help printed circuit board companies generate effective customer demand forecasts by considering indexes from 4 different domains such as macroeconomic and industry production ones. Chen and Ou [5] investigated a perishable food forecasting problem with the consideration of sales data of target store and neighboring stores as well as weather data. In the multivariate sales forecasting literature, a limited number of influencing factors are considered due to reasons such as data unavailability, which perhaps omits important factors and thus weakens forecasting performance.

Although the multivariate forecasting model has potential and advantage to effectively model the relation between sales data and various influencing factors, its real-world application is constrained by some issues: (1) a large number of influencing factors can influence sales, however it is hard to identify the relationships between these factors and final sales and (2) it is usual that there are not sufficient historical data especially when lots of influencing factors are considered or the sales of a new product is forecasted.

In the existing literature, most studies dealt with sales forecasting on the basis of historical sales data of products to be forecasted. To forecast the overall sales of new retail products without directly available historical sales data, previous studies usually considered the historical sales of their similar products [7,30,31]. Thomassey and Fiordaliso [30] also

* Corresponding author. Tel.: +86 28 85417867; fax: +86 28 85415628.
E-mail addresses: zx.guo@alumni.polyu.edu.hk (Z.X. Guo), tcwongca@inet.polyu.edu.hk (W.K. Wong), 11900185r@connect.polyu.hk (M. Li).

0167-9236/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dss.2013.01.026

investigated the effects of 3 variables (item price, the starting time of the sales and the life span of items) on final sales. However, it is hard to identify which past products are similar in forecasting process in industries with large product variety and frequent product changes. These studies also cannot handle future sudden changes caused by influencing factors.

It is commonly accepted among retailers that the early sales of a retail product is an excellent indicator of its overall sales [12]. However, little attention has been paid in utilizing early sales to forecast overall sales of retail products so far. The only one work found was conducted by Tanaka [27], which performed long-term sales forecasting based on the early sales and the correlations between short- and long-term accumulated sales within similar products groups. In his work, the forecasting accuracy largely relied on the selection of reference group, which were determined by expert knowledge and thus subjective and probably unreliable. In addition, Tanaka's work has not considered the effects of various influencing factors on overall sales, which thus could not handle sales changes caused by these factors such as production attributes and promotion strategy.

In this research, we will aim at addressing a sales forecasting problem for retail products based on their early sales, called early sales-based sales forecasting problem, which forecast the overall sales volume of a retail product in a selling season in terms of its early sales with the consideration of a variety of influencing factors instead of using the historical sales data of its similar products. The influencing factors include promotion strategy, product attributes (such as selling price, product type and material type), early sales volumes, various economic indices, climate index, etc. The effects of economic factors can be reflected by various economic indices such as consumer confidence index (CCI), consumer price index (CPI) and gross domestic product (GDP). However, it keeps unknown if and how these factors influence the overall sales volumes of retail products. Among these influencing factors, some may be redundant or even irrelevant to the overall sales, which will detract the accuracy of forecasting model and increase its complexity and computational burden. To the best of our knowledge, it is the first paper that investigates the effects of various influencing factors on overall sales of retail products, and the relations between these factors and overall sales. This research can forecast overall sales of a retail product no matter whether the product has available historical sales data of its similar products.

1.2. Techniques for sales forecasting

Techniques for time series forecasting have been widely applied to sales forecasting since most of existing sales forecasting studies were based on the concept of time series. Existing techniques for time series forecasting are generally divided into two groups: classical techniques based on mathematical and statistical models and artificial intelligence techniques. Classical techniques used include exponential smoothing [13,29], autoregressive moving average (ARIMA) methods [8,23], Kalman filter methods [17,35], and so on. Artificial intelligence techniques used include expert systems [21,25], fuzzy systems [24], neural network (NN) models [3–5], and hybrid models integrating multiple intelligent techniques [30,33].

These classical techniques cannot be used for the modeling of non-time series. However, the sales forecasting problem investigated in this paper does not use time series of historical sales. The classical techniques thus cannot be used for the investigated problem. Among artificial intelligence techniques, the NN model was the most commonly used one, which has been proved to be universal approximators and can effectively model various time series and non-time series. A number of studies also demonstrate that the NN approach outperforms the classical models due to its capacity of nonlinearity, generalization and universal function approximation [2,8,28].

In recent years, a novel NN, extreme learning machine (ELM), has attracted more and more attentions from forecasting researchers due to its feature of easy to use and fast learning speed [6,26]. The ELM can avoid many difficulties faced by traditional NN learning algorithms, such as the selections of stopping criteria, learning rate, and learning epochs because of its distinct learning mechanism. On the basis of ELM, Wong and Guo [33] recently developed an effective hybrid intelligence model to provide effective and reliable forecasts for fashion sales series characterized by nonlinearity, seasonality or irregularity, which utilized harmony search (HS) to improve the generalization and forecasting performance of the ELM, and utilized a fine-tuning process to further improve forecasting accuracy. Their model can outperform ARIMA models and an evolving NN-based forecasting model and an ELM-based model proposed by Sun et al. [26].

In this research, a multivariate intelligent decision-making (MID) model is developed to deal with the investigated sales forecasting problem based on the hybrid intelligence model proposed by Wong and Guo [33]. To decrease the side effects of irrelevant and redundant inputs and the large computational complexity caused by too many input variables, a novel variable selection method, HS-wrapper-based variable selection, is developed in the MID model to select out an appropriate input variable subset from a large number of input variables and explore the relations between overall sales of a retail product and its early sales as well as various influencing factors.

The remainder of this paper is organized as follows. In Section 2, the proposed MID model for the investigated sales forecasting problem is presented. Experimental design and results are presented in Section 3. Section 4 further analyzes and discusses the experimental results and the performance and effectiveness of the proposed model. Finally, conclusions and future work are described in Section 5.

2. Multivariate intelligent decision-making model for sales forecasting

This research considers the early sales of a retail product and various influencing factors as candidate input variables and assumes that there are m input variables in total and n pairs of multi-input–single-output (MISO) samples data are given. Let (X_i, y_i) denotes the ith input/output data pair $(1 \le i \le n)$.

To forecast the overall sales of a retail product in a selling season on the basis of its early sales, this research firstly identifies the relationships between m input variables $(x_{i1}, x_{i2}, \cdots, x_{im})$ of X_i and the overall sales volume y_i and selecting out an appropriate subset of input variables from X_i as the valid inputs, and then establishes an effective model to approximate these relationships in terms of n given samples data. Lastly, the established model is utilized to forecast the sales volumes of other retail products on the basis of their early sales and related input variables. A multivariate intelligent decision-making (MID) model is developed to implement the said processes.

Fig. 1 shows the architecture of the proposed MID model, which is composed of 3 modules, including a data preparation and preprocessing (DPP) module, a HS-wrapper-based variable selection (HWVS) module and a multivariate intelligent forecaster (MIF) module. The DPP module extracts sales data of past products from the point-of-sales database of retailers, and then preprocesses these data for the development and validation of the MID model. The HWVS module selects out the optimal

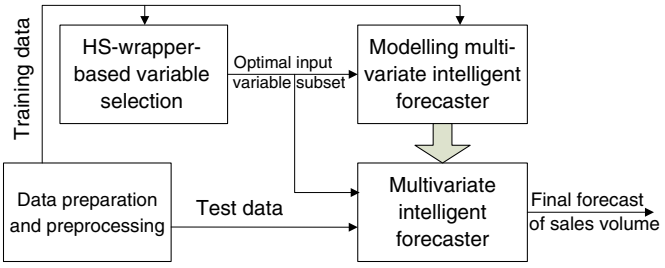

Fig. 1. Architecture of the MID model.