Expert Systems with Applications 36 (2009) 12340–12348

Contents lists available at ScienceDirect

Expert Systems with Applications

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

SKU demand forecasting in the presence of promotions

Özden Gür Ali [a,*], Serpil Sayın [a], Tom van Woensel [b], Jan Fransoo [b]

[a] Koc University, College of Administrative Sciences and Economics, Rumeli Feneri Yolu, Sariyer 34450, Istanbul, Turkey
[b] Technische Universiteit Eindhoven, School of Industrial Engineering, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

ARTICLE INFO

Keywords:
Demand forecasting
Time series
Machine learning
Pooling
Domain knowledge
Promotions

ABSTRACT

Promotions and shorter life cycles make grocery sales forecasting more difficult, requiring more complicated models. We identify methods of increasing complexity and data preparation cost yielding increasing improvements in forecasting accuracy, by varying the forecasting technique, the input features and model scope on an extensive SKU-store level sales and promotion time series from a European grocery retailer. At the high end of data and technique complexity, we propose using regression trees with explicit features constructed from sales and promotion time series of the focal and related SKU-store combinations. We observe that data pooling almost always improves model performance. The results indicate that simple time series techniques perform very well for periods without promotions. However, for periods with promotions, regression trees with explicit features improve accuracy substantially. More sophisticated input is only beneficial when advanced techniques are used. We believe that our approach and findings shed light into certain questions that arise while building a grocery sales forecasting system.

1. Introduction

Retailers are faced by increasing assortments. In grocery retail, which is the main focus of our study, product life cycles have been decreasing (see, e.g. Bayus & Putsis, 1999). As a consequence, it is increasingly difficult to forecast sales for an individual item in a particular store for tactical decisions, as time series for those items tend to be short. Moreover, retail sales are faced with extensive promotion activities. Products are typically on promotion for a limited period of time, e.g. one week, during which demand is usually substantially higher than during periods without promotions (see, e.g. Cooper, Baron, Levy, Swisher, & Gogos, 1999), and many of the stock-outs occur during promotions due to inaccurate forecasts.

There is a clear need to take the promotions for the focal item and potentially for the related items into account when forecasting its demand. Regression of the past SKU sales with promotion variables allows for a linear (or log-linear with transformed variables) relationship, see e.g. (Hansen, McDonald, & Nelson, 2006; Mulhern & Leone, 1991). Machine learning techniques allow learning nonlinear relationships. For example, Aburto and Weber (2007) use an additive combination of SARIMA models and neural networks trained with the residuals of the time series model and the promotional inputs to forecast grocery sales. While these approaches clearly bring flexibility to the functional form of promotion impact, there are attempts to incorporate managerial knowledge of promotional impact to guide the search for more predictive models, for example (Kuo, 2001) uses fuzzy rules elicited from managers to set the initial weights of the neural network. In this paper we leverage the studies from marketing literature, usually within the context of consumer choice models, e.g. (Fader & Hardie, 1996), to formulate extensive features by summarizing past sales and promotion data of the focal and related SKUs. Considering the large number of features, we employ tree regression to forecast SKU sales.

Operationally, it is clear that the accuracy of the forecast directly contributes to higher profits by reducing stock-out situations and lowering the level of safety inventory. However, the cost due complexity of the system in terms of data preparation, setup and maintenance of sophisticated models requiring expert analysts also has an impact on the profitability of operations, giving an incentive for simpler data and models. Bucklin and Gupta (1999) report that simple models are used in practice.

In this paper we explore the complexity and data preparation cost versus forecasting accuracy tradeoff to identify models that are on the efficient frontier. We consider models to be combinations of input features, forecasting technique and scope. We construct and evaluate models that span the spectrum in complexity from the simple benchmark models used in grocery stores to the regression tree model with extensive features covering all SKUs in the category and all stores. We also evaluate the new and promising Support Vector Regression techniques with different kernels. The data for the evaluation spans 76 weeks of SKU-store level sales and promotion time series from a medium-sized European grocery

* Corresponding author. Tel.: +90 212 3381450.
E-mail addresses: oali@ku.edu.tr (Ö.G. Ali), ssayin@ku.edu.tr (S. Sayın), t.v.woensel@tue.nl (T. van Woensel), j.c.fransoo@tue.nl (J. Fransoo).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.04.052

retailer. We focus on the sales of a single category that consists of four subcategories in four stores, amounting to 168 store-SKU combinations.

Our findings indicate that for periods without promotions, it is next to impossible to beat simple time series techniques. At best, the machine learning techniques match the performance of the simple models. However, for periods with promotions, a substantial and significant improvement of up to 65% can be reached using regression trees with the explicit variables derived from the sales and promotion time series of the category SKUs. Interestingly, using more sophisticated input variables only helps when the prediction technique has the capacity to take advantage of them appropriately. Obviously, the use of more advanced techniques comes at a cost, namely the use of more extensive data (resulting in data preparation cost) and the maintenance of more complicated models. However, since the improvement is substantial, the benefits are likely to outweigh the costs involved.

Our paper is organized as follows. In the following section, we review the relevant literature. Next, we introduce the data, the modeling elements and the candidate models, and describe the experimental setup. We present and discuss the results of our experiments in Section 4 and conclude in Section 5.

2. Literature review

Sales forecasting methods can roughly be grouped into judgmental, extrapolation, and causal methods. Extrapolation methods use only the time series data of the activity itself to generate the forecast. The particular techniques range from the simpler moving averages and exponential smoothing family to the more complicated Box–Jenkins approach. While they identify and extrapolate time series patterns of trend, seasonality and autocorrelation successfully, they do not take external factors such as price changes and promotion into account (Alon, Qi, & Sadowski, 2001). Vector Auto Regression (VAR) models extend the Box–Jenkins methods to include other variables see e.g. (Enders, 2004), however their complexity makes estimation difficult. Causal forecasting involves building quantitative models using inputs representing the phenomena that are believed to be drivers of the outcome; it could be as simple as linear regression model with promotion variables. A popular commercial package, ACNielsen's SCAN*PRO tool, facilitates evolutionary model building (Van Heerde, Leeflang, & Wittink, 2002). The starting point is a regression model with promotion variables such as price cut, feature advertising, special displays in the store. The idea is that model simplicity helps managers to understand and approve or guide modification of the models, and as they become more knowledgeable about a decision aid, they may be ready to implement more sophisticated and complex models. In (Cooper et al., 1999), a promotion-event forecasting system called PromoCast is described, which uses regression analysis of SKU-store sales under a variety of promotion conditions, with store and chain specific historical performance information.

Sophisticated causal forecasting can involve concrete hypotheses about the functional form of the effect, as in the marketing literature where particular sales response models are hypothesized based on consumer behavior theory, and the data is used to estimate its parameters, see for example (Fader & Hardie, 1996). Machine learning techniques, such as trees, SVM or neural networks, on the other hand, do not assume a particular relationship between the variables and involve a search through the functional form space as well as parameter estimation.

The marketing literature found several factors to be significantly drive promotional sales. These include the size of the price decrease for a promoted item (in e.g. Blattberg, Briesch, & Fox, 1995; Christen, Gupta, Porter, Staelin, & Wittink, 1997; Cooper et al., 1999; Lattin & Bucklin, 1989; Mulhern & Leone, 1991); the frequency of the promotions for similar products (Christen et al., 1997); the advertising mode used, e.g. newspapers, TV, radio, etc. (see e.g. Sethuraman & Tellis, 2002); category or product group characteristics (see e.g. Baltas, 2005; Blattberg et al., 1995; Hughes, 1980; Narasimhan, Neslin, & Sen, 1996); the weather and bank holidays (Bunn & Vassilopoulos, 1999; Hughes, 1980); promotions of competitors (Struse, 1987), radio support of promotions (Sethuraman & Tellis, 2002), previous promotions of the same product, and the number of variants of the product in promotion. The last two aspects are closely related to cannibalizing demand of the promoted product by the same organization (see e.g. Neslin, 1990). Clearly, in order to obtain accurate forecasts in the presence of promotions, it is thus important to systematically collect and use the data about the characteristics of promotions. Note that the primary objective of the marketing studies is to determine the main effects that contribute to increased sales, rather than to forecast the specific quantity sold.

Turning to the use of machine learning techniques in sales forecasting, in (Alon et al., 2001), artificial neural networks are compared with some traditional forecasting methods including Winters exponential smoothing, Box–Jenkins ARIMA model, and multivariate regression on the aggregate US retail sales data. The authors conclude that the Winters exponential smoothing model is a viable method under relatively stable economic conditions, and that the advanced neural network models perform much better when facing dynamic nonlinear trend and seasonality patterns (Alon et al., 2001). Agrawal and Schorling (1996) compare neural networks and multinomial logit model performance in forecasting the brand *share* – rather than sales – in grocery categories and found that neural networks performed better. They use with price and promotion information for each brand as input. Zhang and Qi (2005) argue that the performance of the Neural Network improves significantly when the time series data is de-seasonalized and de-trended. Aburto and Weber (2007) use a SARIMA (seasonal ARIMA) model on the sales time series and train a neural network to predict future residuals using the residuals from the last k periods, as well as the promotional inputs and special day dummies to forecast SKU sales at a Chilean supermarket. They find that the hybrid model of SARIMA and neural network performs better than either model on their own. Kuo argues that the straightforward inclusion of promotion variables in the neural network may not be able to capture the complexity of the effect of promotions and use fuzzy rules elicited from marketing experts in a GFNN and ANN combined architecture (Kuo, 2001).

Thomassey and Fiordaliso (2006) propose a hybrid system based on clustering and classifying the items according to their sales behavior using decision tree classification to forecast sales in a textile retailing environment.

In an application of rule-based data mining approach, Cooper and Giuffrida (2000) use the regression based PromoCast (Cooper et al., 1999) to predict demand at the SKU level of a 95-store retail chain, and look for rules predicting the residuals. The study reports that the rule-based data mining approach improves forecasting errors of PromoCast by about 9% by using attributes that define the SKU such as the manufacturer, the category/subcategory it belongs to, the promotion event and the store.

Although we have not identified a grocery sales forecasting application in the literature, the more recent machine learning technique, Support Vector Machine (SVM) has been applied successfully in many different scientific disciplines (Shawe-Taylor & Cristianini, 2000). Unlike the neural networks where the training is a search that may or may not yield the best model, the SVM is based on yields the best model for the inputs. Maximizing a margin of separation, and minimizing total empirical error in a balanced way, using the implicit mapping idea induced by the Kernel functions, SVMs have performed well where complex relationships