Expert Systems with Applications 39 (2012) 12784–12798

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

On the differential benchmarking of promotional efficiency with machine learning modelling (II): Practical applications

Cristina Soguero-Ruiz [a,*], Francisco-Javier Gimeno-Blanes [b], Inmaculada Mora-Jiménez [a], María Pilar Martínez-Ruiz [c], José-Luis Rojo-Álvarez [a]

[a] Signal Theory and Communications Department, University Rey Juan Carlos, Camino del Molino s/n, Fuenlabrada, 28943 Madrid, Spain
[b] Signal Theory and Communications Department, University Miguel Hernández, Av. Universidad s/n, Elche, 03202 Alicante, Spain
[c] Commercialization and Market Research Department, University of Castilla-La Mancha, Av. de los Alfares, 44, Cuenca 16071, Spain

ARTICLE INFO

Keywords:
Sales promotion
Machine learning
Bootstrap
Price indices
Marketing
Processing

ABSTRACT

The assessment of promotional sales with models constructed by machine learning techniques is arousing interest due, among other reasons, to the current economic situation leading to a more complex environment of simultaneous and concurrent promotional activities. An operative model diagnosis procedure was previously proposed in the companion paper, which can be readily used both for agile decision making on the architecture and implementation details of the machine learning algorithms, and for differential benchmarking among models. In this paper, a detailed example of model analysis is presented for two representative databases with different promotional behaviour, namely, a non-seasonal category (milk) and a heavily seasonal category (beer). The performance of four well-known machine learning techniques with increasing complexity is analyzed in detail here. In particular, *k*-Nearest Neighbours, General Regression Neural Networks, Multilayer Perceptron (MLP), and Support Vector Machines (SVM), are differentially compared. Present paper evaluates these techniques along the experiments described for both categories when applying the methodological findings obtained in the companion paper. We conclude that some elements included in the architecture are not essential for a good performance of the machine learning promotional models, such as the semiparametric nature of the kernel in SVM models, whereas other can be strongly dependent of the database, such as the convenience of multiple output models in MLP regression schemes. Additionally, the specificity of the behaviour of certain categories and product ranges determines the need to establish suitable and specific procedures for a better prediction and feature extraction.

1. Introduction

Nowadays, the context in which grocery retailers have to make their decisions is subjected to a high degree of complexity and uncertainty. These circumstances are especially difficult at this particular moment given the current economic situation, which undoubtedly influences the decision processes in relation to price positioning and sales promotions strategies. In this sense, retailers must be very careful and should evaluate different aspects related to the particular context of prices and sales promotions, such as specific features of the categories and brands carried by them (Voss & Seiders, 2003). From this perspective, and although in recent years research in this area has increased considerably (Blattberg & Neslin, 1990; Levy, Grewal, Kopalle, & Hess, 2004; Martínez-Ruiz, Gomez-Borja, Mollá-Descals, & Rojo-Álvarez, 2008; Voss & Seiders, 2003), some questions remain unsolved. For example, some studies show the difficulty to relate properly sales promotions and price discounts. In particular, authors have identified that, under certain circumstances, it is especially difficult to establish optimal pricing. This is particularly the case when simultaneous effect are taking place and hence stressing margins and profits (Tellis & Zufryden, 1995). Additionally, promotional activities may have discordant objectives among different elements within the sales chain, and this fact may eventually have a direct impact on decision making strategy (Sayman & Raju, 2004). For instance, manufacturers may be trying to raise sales, or trying to enhance brand recognition, whilst retailers are focusing on maximising efficiency, and total benefits of a whole category instead of a solely product or brand.

During decades, important efforts have been made to better understand the dynamics in sales promotion. Initially, these analyses were based on classical statistical methods, and now, more and more, important developments on machine learning and data mining techniques are being developed. The machine learning

* Corresponding author.
E-mail addresses: cristina.soguero@urjc.es (C. Soguero-Ruiz), javier.gimeno@umh.es (F.-J. Gimeno-Blanes), inmaculada.mora@urjc.es (I. Mora-Jiménez), mariapilar.martinez@uclm.es (M.P. Martínez-Ruiz), joseluis.rojo@urjc.es (J.-L. Rojo-Álvarez).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.04.035

techniques have the objective to find repetitive patterns, trends or rules, which can explain data behaviour at a given context, allowing to extract new knowledge on the consumer behaviour, to improve the performance of marketing operations, and to estimate the deal effect curve (DEC). Though a vast amount of knowledge has been obtained in this setting from machine learning techniques, there are still promotional behaviours that have not been studied in detail sufficiently. Hence, a deeper analysis on sales promotion characterization, based on empirical methods, needs to be addressed (Bell, Chiang, & Padamanabhan, 1999; Blatterg, Briesch, & Fox, 1995; Leeflang & Wittingk, 2000).

There are a number of operational issues that need to be considered when machine learning techniques are to be applied to promotional modelling (Liu, Kong, & Yang, 2004; Martínez-Ruiz, Mollá-Descals, & Rojo-Álvarez, 2006; Martínez-Ruiz, Mollá-Descals, Gómez-Borja, & Rojo-Álvarez, 2006a; Wang, Li, & Zhao, 2008; Van Heerde, Leeflang, & Wittink, 2001): (1) heavy tails and heteroscedasticity for the prediction residuals yield to Gaussianity as a residual property not always to be assumed; (2) actual risk in merit figures has to be properly taken into account throughout all the machine design process; (3) it is not always easy to set a cut-off test for results evaluation. For these reasons, the companion paper (Soguero-Ruiz, Gimeno-Blanes, Mora-Jiménez, Martínez-Ruiz, & Rojo-Álvarez, 2012) presented a simple nonparametric statistical tool, based on the paired bootstrap resampling for establishing clear statistical comparisons among methods. Additionally, the companion paper analysed the assumptions, the method, and the steps, that should be applied to properly utilise the learning-from-samples technique for promotional characterization. This study allowed us to analyze and evaluate different models in terms of averaged and scatter characterizations of merit figures for the distribution of the actual risk. As a positive result in (Soguero-Ruiz et al., 2012), it should be mentioned that the free parameter tuning procedure was strongly independent from an important number of performance measurements (*index of agreement D – D index*, Mean Absolute Error – MAE, or Relative Mean Absolute Error – RMAE), which was a demonstration of the powerfulness of the method for making statistical comparisons in this setting.

This paper presents a set of practical applications for the proposed methods and systematic benchmarking among several relevant and well-known different learning techniques. Results are shown for two representative databases with different promotional behaviour, namely, a non-seasonal stable category (milk) and a heavily seasonal category (beer). Four well-known machine learning algorithms with increasing complexity are benchmarked differentially, specifically, k-Nearest Neighbours (k-NN), General Regression Neural Networks (GRNN), Multilayer Perceptron (MLP), and Support Vector Machines (SVM). Subsequent experiments are devoted to explore the actual performance improvement obtained for machine design architecture in MLP and SVM, and the procedure is stated for feature selection analysis criteria.

The draw of the paper is as follows. Section 2 presents a description of the machine learning techniques analyzed in this work, and a short summary of the method proposed in the companion paper (Soguero-Ruiz et al., 2012). Afterward, the two databases to be used for sales promotion modelling are described (milk and beer category) in Section 3. Section 4 includes the results of the four experiments (A, B, C and D) and the analysis performed on both databases. Experiment A is devoted to a detailed comparison of the performance for the different techniques on each data set. Experiment B shows the performance of different elements in the MLP architecture design, whereas Experiment C deals with the kernel architecture in SVM estimation modelling design. Last, experiment D, gives a principled approach to feature selection using the paired bootstrap test in the merit figures. Finally, in Section 5, discussion is presented and concluding remarks are summarized.

2. Machine learning techniques for promotional sales modelling

Machine learning techniques have emerged as powerful tools to extract relevant quantitative information (Mitchell, 1997; Van Heerde, Leeflang, & Wittink, 2000). Two different types of regression methods have been mostly used in the sales promotion literature, to analyse the sales response to price promotions discounts: parametric regression and nonparametric regression. Parametric regression assumes a certain functional form underlying the data, namely linear, exponential, or logarithmic. The simplest parametric regression model is the linear model, where the parameters can be easily estimated using Ordinary Least Squares (OLS), assuming the presence of additive, uncorrelated, and Gaussian white noise. However, in the presence of heteroscedasticity, Generalized Least Squares methods are more appropriate (Hastie & Tibshirani, 1990). In addition, Maximum Likelihood models assume a given statistical distribution linking the parameters and the data (Ruppert, Wand, & Carroll, 2003).

Nonparametric regression does not assume any a priori functional form, but it rather relies on approximating the observations locally. Examples of nonparametric methods are spline regression, k-NN, and kernel estimators, among others (Ruppert et al., 2003). The main advantages of non parametric methods are flexibility and consistency, which are established under much more general conditions than for parametric modelling. In this respect, it is interesting to remark how nonlinearity, non-normal errors, and heteroscedasticity, are automatically accommodated by nonparametric methods. Nevertheless, a disadvantage of the nonparametric approach is that its convergence rate is often slower than that of parametric estimators. Therefore, precise estimation of the nonparametric multidimensional regression requires comparatively a higher number of observations.

A trade-off between parametric and nonparametric features can be found in semiparametric regression (SR), which uses a model that considers at the same time: (i) a parametric component, that provides efficiency and low variance; and (ii) a nonparametric component, that provides flexibility and small bias, whenever they are optimally combined. The use of SR to assess economic series analysis was first used by (Robinson, 1988, 1989). In this setting, a successful application of SR in deal effect curve estimation (DEC) (Martínez-Ruiz, Mollá-Descals, Gómez-Borja, & Rojo-Álvarez, 2006b; Martínez-Ruiz et al., 2008; Van Heerde et al., 2001) suggests that a comparatively high number of data is required and also over-fitting deserves attention. These aspects can limit the applicability of SR methods, where the number of observations can be only tens or some few hundreds.

2.1. General data model for promotional sales

In order to support the model architecture that is capable of learning from the relationships between inputs ($\mathbf{x}$, column vector) and outputs (y), it is required a finite number of paired observations. In sales promotion modelling, the input pattern may consist of information about price changes and promotion characteristics, whereas the output would correspond to the number of sold units for a given product. The model $f(\,)$ for the relation $y = f(\mathbf{x})$, has been mainly estimated in the marketing research literature by using three different families of regression methods. Regarding to the first of them, in parametric methods, it is assumed a previously known shape or structure for functional relation $f(\,)$. In this case, the functional is often defined by a simple relationship (linear), while the nonparametric method does not assume any prior structure in terms of data model, instead, it is built the estimated relationship based on kernels (for instance, the Gaussian kernel)