Intelligent system for time series classification using support vector machines applied to supply-chain

Fernando Turrado García [a], Luis Javier García Villalba [a,*], Javier Portela [b]

[a] Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), School of Computer Science, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain
[b] Group of Analysis, Security and Systems (GASS), Department of Statistics and Operational Research III, School of Statistics, Universidad Complutense de Madrid (UCM), Avenida Puerta de Hierro s/n, Ciudad Universitaria, 28040 Madrid, Spain

ARTICLE INFO

ABSTRACT

To be able of anticipate demand is a key factor for commercial success in the supply-chain sector. The benefits can be grouped around two main concepts: firstly the optimization of operations through the development of optimal strategies for procurement and secondly the stock reduction that reduces storage costs, handling, etc. There is currently a variety of methods for making predictions, these methods vary from pure statistical methods such as exponential smoothing Holt-Winters or ARIMA models, to those based on artificial intelligence techniques like neural networks or fuzzy systems. However, despite being able to build accurate models, in managing the supply chain based on forecasts there is a problem known as "Forrester effect" irrespective of the model chosen. To monitor the impact of this effect, given the volume of information handled in large corporations, is a very expensive task (often manual) for such corporations because it requires investigating issues such as the adequacy of the model, allocation of known models to the sales time series, discovery of new patterns of behavior, etc. This article proposes an intelligent system based on support vector machines to solve problems concerning the allocation and discovery of new models. With this focus in mind, the system objective is to build groups of time series that share the same forecasting model. For the identification of new models, the system will assign "virtual models" for those groups that do not have a predefined pattern. Using the proposed method, it has been possible to group a sample of more than 14,000 time series (real data taken from a store) in around 70 categories, of which only 12 of them already grouped over 98% of the total.

1. Introduction

In the retail area, specifically within the order management in the supply chain, providing a precise mechanism for forecasting demand is a key factor in the success of large corporations. The focus of most of its operations is based on the demand: purchasing to suppliers, inventory management, etc. All those tasks are required to ensure quality service to its customers.

Efficient management of the supply chain (based on forecasts of demand) is a complex issue, with hard-resolution problems as the "Forrester effect" (also known as bullwhip effect (Lee, Padmanabhan, & Whang, 1997)). This problem, in which adding steps in the chain provokes an increase in the forecasts variance, was described in 1961 by Forrester (1999) and has been "reformulated" in 2000 by Chen, Drezner, Ryan, and Simchi-Levi (2000). Today, work continues on the analysis of relationships between customers and suppliers (as seen in the work of Danese & Romano (2011)).

In this study, we will analyze a case of a worldwide logistical distributor, which maintains thousands of stores. The complexity of the logistical work that takes place in a company of this type can be outlined through the presentation of the volume of operations performed. From here on in, we will refer to each item available for sale in one of the company stores by SKU (Stock Keeping Unit). For each SKU, studies regarding the evolution of sales, stock, purchase ordering, receipt of goods (measuring the quality of delivery), etc. are done by a statistical department. In the current case, we focus on the time series formed by the daily sale of one SKU at a single shop. The main objective of that analysis is to elaborate accurate sales forecasts, which will be used in the stock replenishment at the company warehouses. Although obvious, the quality of these forecasts will have a strong impact on the income account of the company.

* Corresponding author. Address: Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence, School of Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain. Tel.: +34 91 394 7638; fax: +34 91 394 7547.

E-mail address: javiergv@fdi.ucm.es (L.J. García Villalba).

Thus, in each store (large stores such as hypermarkets), more than 10,000 different SKUs are available to customers. In the current situation, an international distributor that currently has more than 9500 stores in 32 countries, over 4500 of which can be located in a single country. To resolve the problem only in that territory, it would be necessary to compare/analyze more than 45,000,000 time series.

It is in these circumstances where the existence of an autonomous system in charge of analyzing this set of time series represents a competitive advantage. This system will increase quality of forecasts (via a better model allocation) and the efficiency in the organization because it allows the company to make its daily operations with a savings in personnel.

The study of sales time series, in order to make forecasts, can be addressed using explicit models in which internal factors to the SKU (seasonality, trend, promotions, price changes etc.) and external (cross-selling, cannibalization effect, competitor's promotions etc.) are taken into consideration. Nowadays there are different techniques for calculating forecasts, starting from pure statistical models such as exponential smoothing Holt-Winters (you can check its definition and some implementation details in the work of Chatfield (1978); (Chatfield & Yar, 1988)) or ARIMA models (defined by (Box, Jenkins, & Reinsel, 1994)) or methods based on artificial intelligence techniques like support vector machines (may be taken as an example the work of Shahrabi, in Shahrabi, Mousavi, and Heydar (2009) or that of Cai, Chen, and Zhao (2011)), genetic algorithms (consult the work of Kuo and Han (2011) or that of Min, Lee, and Han (2006)) or fuzzy logic based systems (Wang, 2011).

In the whole set of time series, there are different elements of behavior, such as sporadic selling SKUs (high end appliances) or SKUs with high customer demand (food products like bread). Due to the nature of the retail sector cyclical behaviors are observed at different levels (weekly, monthly, annually, etc.). Also, trend and calendar effects (holidays on which stores do not open) are detected in the series, which makes them harder to process.

We will use ARIMA models defined by Box et al. (1994) as reference, since they are well suited for time series with trend and seasonality. As can be seen in Shukla and Jharkharia (2011), these models have been successfully applied to the generation of
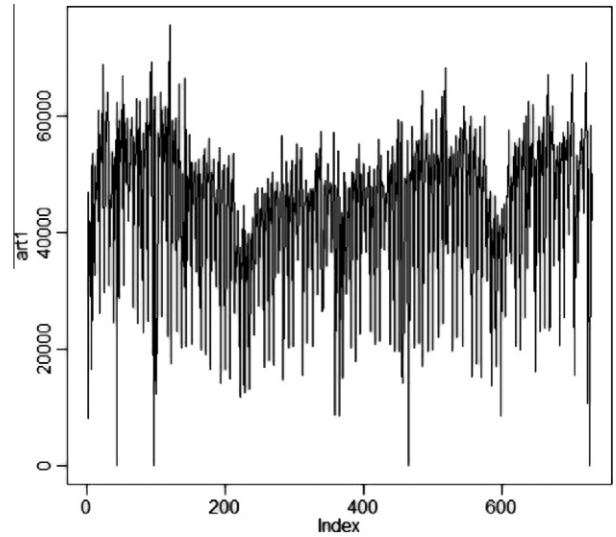

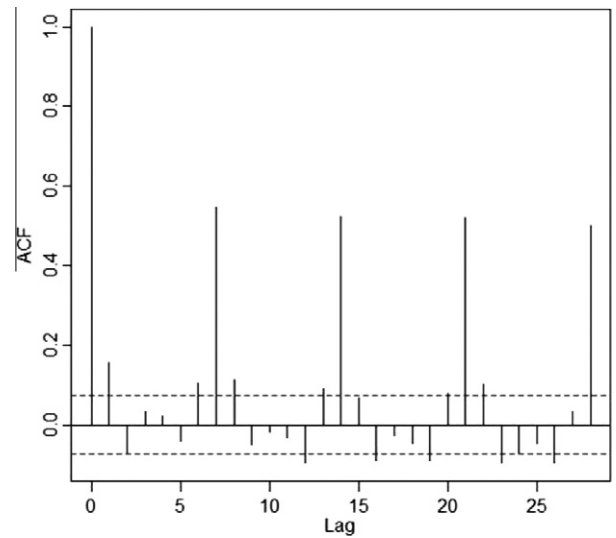Fig. 2. 2-year series of sales of the same SKU.

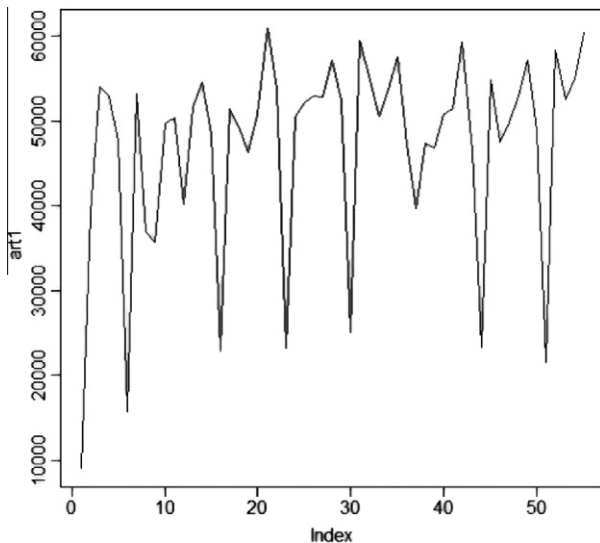Fig. 3. Graphs of the functions ACF applied to the series.

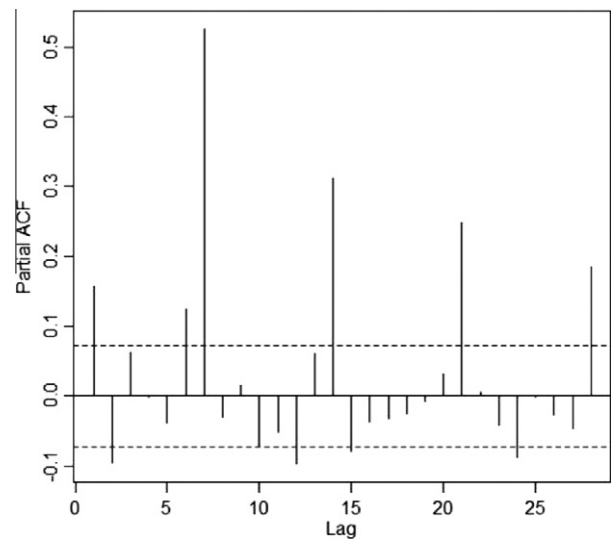Fig. 1. 56 days of the time series of sales for a SKU (in units).

Fig. 4. Graphs of the functions PACF applied to the series.