



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

Fernando Turrado García<sup>a</sup>, Luis Javier García Villalba<sup>a,\*</sup>, Javier Portela<sup>b</sup>

<sup>a</sup> 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

<sup>b</sup> 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

**Keywords:**  
Support vector machines  
ARIMA  
Supply chain  
Time series

## ABSTRACT

To be able to 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.

© 2012 Elsevier Ltd. All rights reserved.

## 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 de-

scribed 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: [javiervg@fdi.ucm.es](mailto:javiervg@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

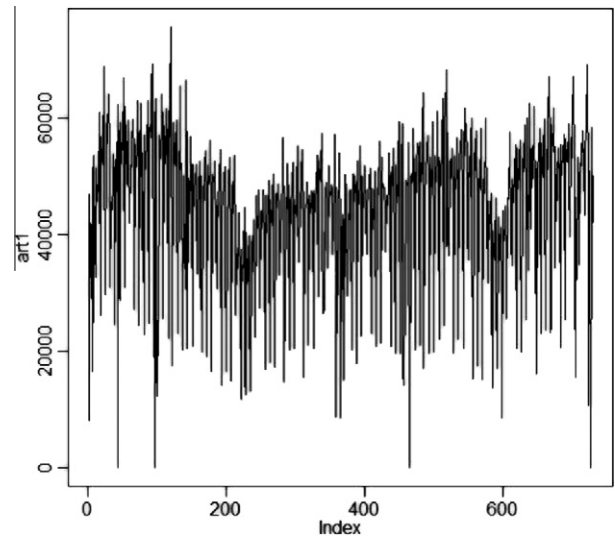


Fig. 2. 2-year series of sales of the same SKU.

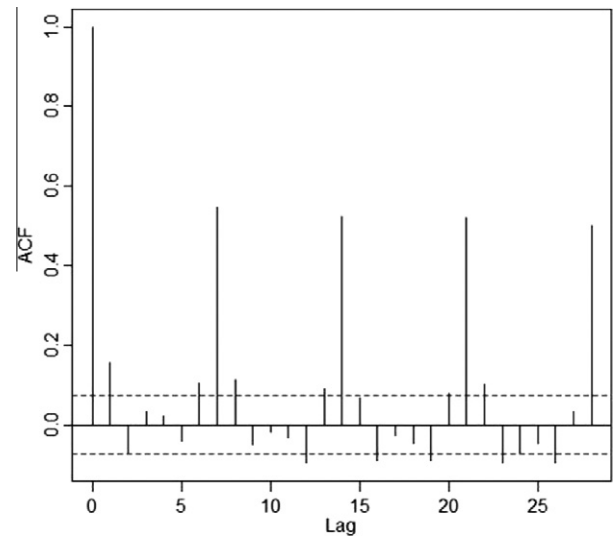


Fig. 3. Graphs of the functions ACF applied to the series.

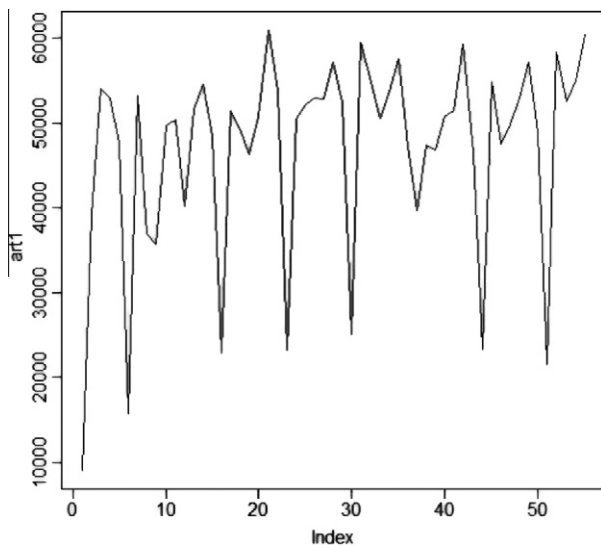


Fig. 1. 56 days of the time series of sales for a SKU (in units).

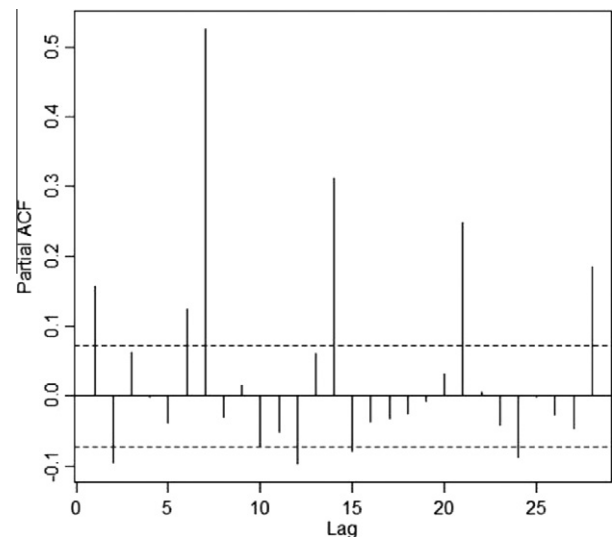


Fig. 4. Graphs of the functions PACF applied to the series.

**Table 1**  
Clusters built by the system.

Cluster	N. Elem.	%	Cluster	N. Elem.	%
7073790000	6240	43.75	583360000	4	0.03
7964670000	3628	25.44	6368510000	4	0.03
922670000	2474	17.35	5821040000	3	0.02
1235480102	516	3.62	7962570000	3	0.02
864550000	433	3.04	2925740000	3	0.02
7346610000	316	2.22	1246820000	3	0.02
1425630000	131	0.92	2115210000	2	0.01
2896410000	78	0.55	2345710000	2	0.01
1308990000	67	0.47	1418860000	2	0.01
2354850000	55	0.39	7951110000	2	0.01
2448570000	31	0.22	2006730000	2	0.01
5606700000	26	0.18	2214730000	2	0.01
6058530000	24	0.17	7160420000	2	0.01
1282150000	23	0.16	5276720000	2	0.01
1282220000	20	0.14	2971340000	2	0.01
7928170000	12	0.08	6296260000	2	0.01
3772150000	11	0.08	1033050000	2	0.01
7610730000	10	0.07	1560690000	2	0.01
5951890000	10	0.07	1418870000	2	0.01
825790000	9	0.06	5797540000	2	0.01
7936600000	8	0.06	503550000	2	0.01
1015370000	7	0.05	2890330000	2	0.01
266300000	7	0.05	1145640000	1	0.01
1078810000	7	0.05	29120000	1	0.01
2910720000	7	0.05	4116740000	1	0.01
6657420000	6	0.04	1356910000	1	0.01
7127200000	6	0.04	88770000	1	0.01
1282260000	5	0.04	6952890000	1	0.01
5606790000	5	0.04	1130420000	1	0.01
5568570000	5	0.04	5621990000	1	0.01
2923480000	5	0.04	2827900000	1	0.01
1186590000	5	0.04	967610000	1	0.01
7912370000	4	0.03	2352970000	1	0.01
1488590000	4	0.03	460000	1	0.01
6733000000	4	0.03			

forecasts in the supply of fresh foods. However, the development of such models is out of the scope of this research. Fig. 1 shows

56 days of the time series of sales for a SKU (in units). Fig. 2 shows 2-year series of sales of the same SKU.

In the best case, the ARIMA statistical models were clearly established, but there would still be another big problem with the huge volume of data to be processed, which makes any approach to individualized treatment of the series is impracticable because of the economic cost associated.

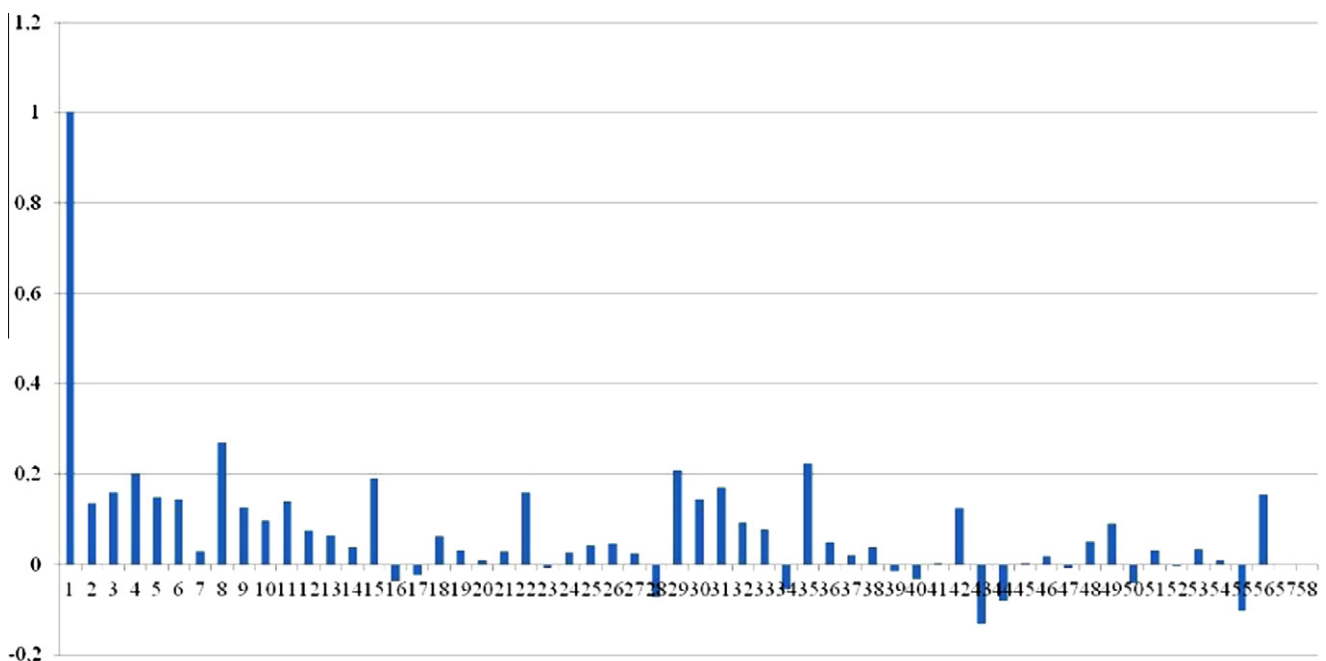
At this point, someone might think that the behavior of a SKU is similar in all the stores, but nothing is further from reality, there are a lot more factors to be considered. As an example, the number of sales of the same SKU can be completely different depending on stores geographic location. This is obvious if we take as an ice pop as an example and analyze their series at two stores, one in a mountain region and another located in a tourist resort near a beach.

Finally, the aim of this study is to define an intelligent system of classification series based on support vector machines. The categories or clusters calculated will allow the allocation of statistical models (ARIMA models as the preferred option) to time series groups in order to make predictions on them.

## 2. Material and methods

Following the guidelines given by Box et al. (1994), the identification of ARIMA models is based on two functions defined as autocorrelation (ACF) and partial autocorrelation (PACF). The values of these functions are calculated on the number of “regression terms” used in the model definition (from here on in, we will refer to this number as  $N$ ). With that work in mind, two series will have the same ARIMA model associated if and only if the autocorrelation and partial autocorrelation functions give similar results in their  $N$  first positions.

So, if we define the characteristics of the series (from a classification point of view) as the results of the above functions, we can represent them using a vector of  $2N + 1$  positions. The first  $N + 1$  positions represent the results of the autocorrelation function



**Fig. 5.** Cluster 7073790000.

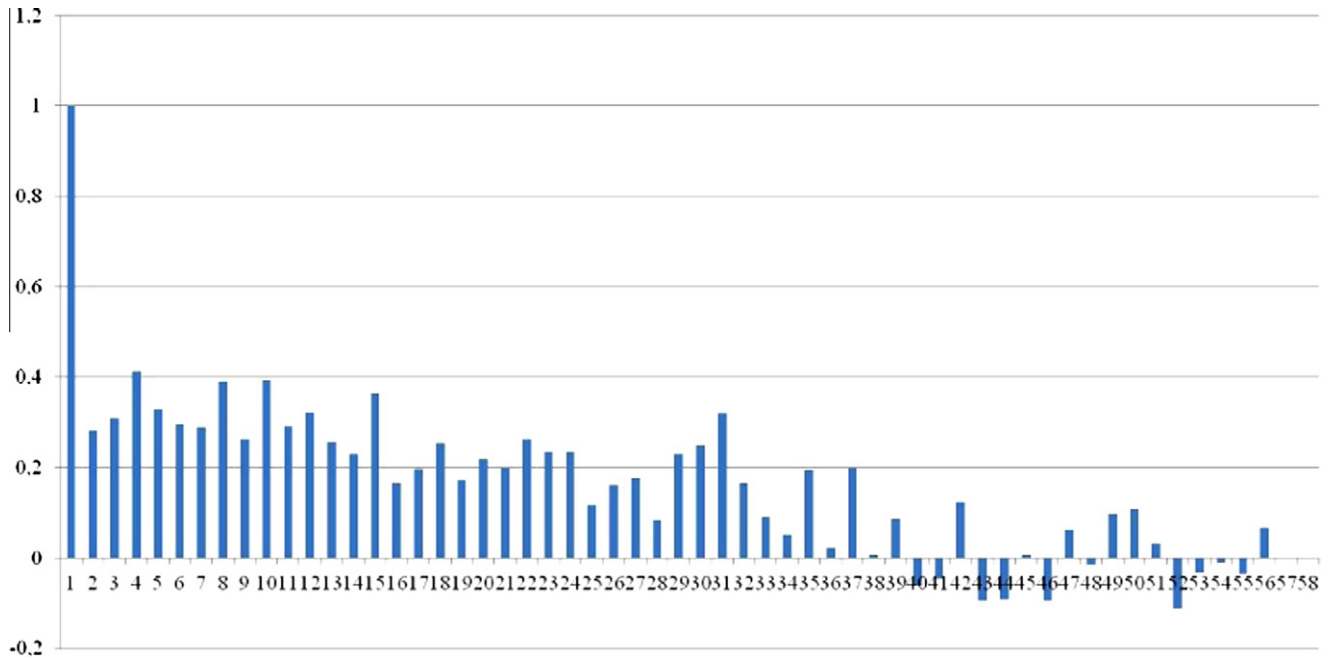


Fig. 6. Cluster 7964670000.

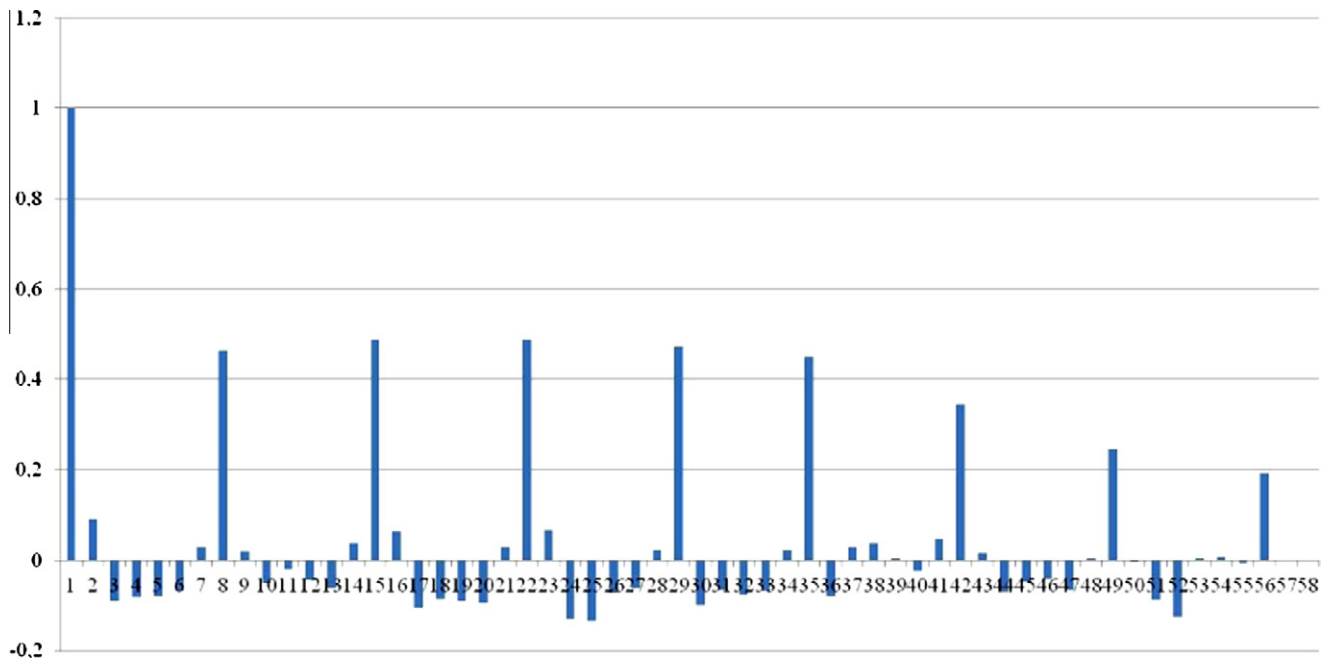


Fig. 7. Cluster 922670000.

(position 0 always takes the value 1 and then  $N$  results) and the last  $N$  position belongs to the results of partial autocorrelation function.

An important fact to note is that the values calculated by the above functions take their values in the interval  $[-1,1]$ . This facilitates treatment of the series, because no conversions or adjustments are needed in the input data for the support vector machines (SVMs).

Figs. 3 and 4 are the graphs of the functions ACF and PACF applied to the series used as an example in the previous section.

The classifier algorithm is as follows:

Let  $C$  be the set of clusters;

Let triad  $\langle R, SVM(R), S(R) \rangle$  be a cluster composed of a vector of reference, a SVM that recognizes the elements that will belong to the cluster and the list of items currently belonging to it;

Let  $D$  be the set of vectors to classify;

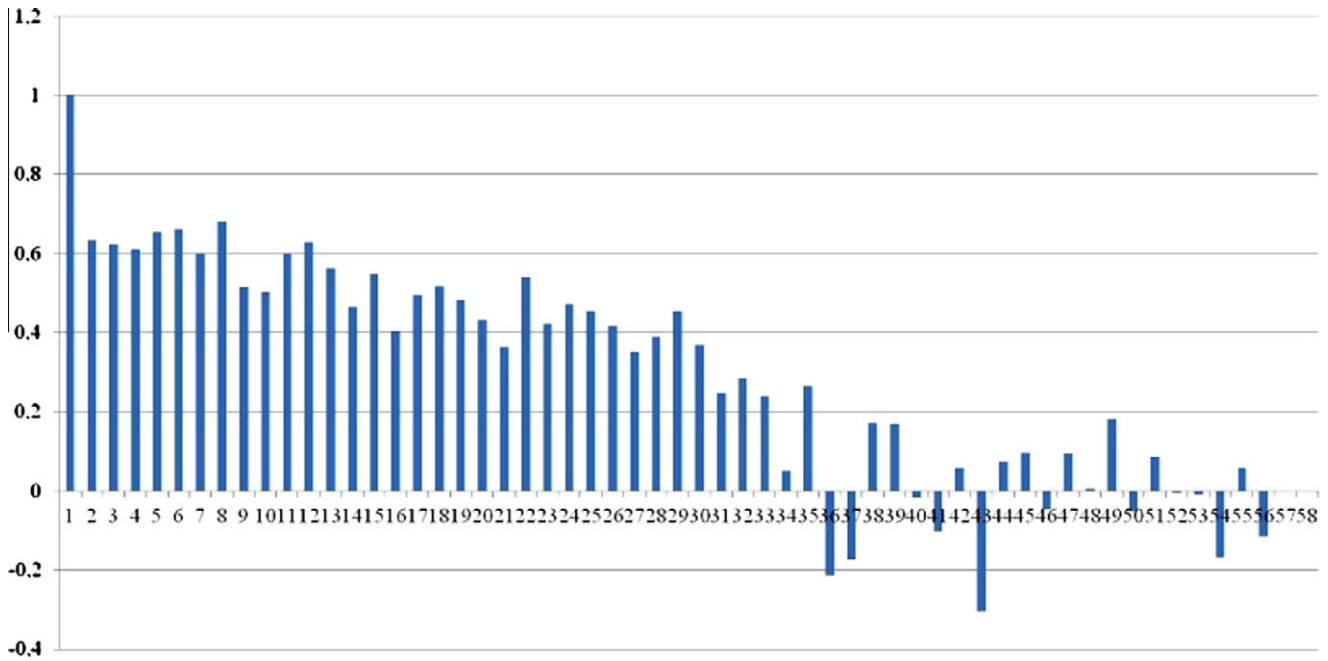


Fig. 8. Cluster 1235480102.

```

foreach vector  $d \in D$  do
  foreach element  $\langle X, SVM(X), S(X) \rangle \in C$  do
    if  $SVM(X)$  accepts  $d$  then
      Add  $d$  to  $S(X)$ ;
      Found = True;
      Exit foreach;
  if Not Found then
    Make new triad  $\langle d, SVM(d), \{d\} \rangle$  and add
    it to  $C$  set;

```

Since there is no training data sets for the support vector machines, we need an algorithm that builds those data sets. In this case, we used the following algorithm:

#### Definition of input parameters:

$R$  is defined as the reference vector of the cluster to identify;  
 $\alpha$  is defined as the maximum acceptable difference in each vector coordinate generated comparing it with the reference vector;  
 $\beta$  is defined as the maximum number of coordinates in which the difference is greater than allowed by  $\alpha$  parameter;  
 $\gamma$  is defined as the number of training cases to build that belong to the cluster;  
 $\theta$  is defined as the numbers of training cases to build that do not belong to the cluster;

#### Construction of cases:

$I$  is defined as the number of cases built outside the cluster

$J$  is defined as the number of cases built inside the cluster

```

 $I = 0, J = 0;$ 
while  $I < \theta$  do
  Make a random vector  $X/X(r) \in [-1, 1], r \in [1, 2N + 1];$ 
  Let  $D =$  coordinate count that  $|X(k) - R(k)| > \alpha,$ 
   $k \in [1, 2N + 1];$ 
  if  $D > \beta$  then
     $I = I + 1;$ 
    Mark  $X$  as outside;
  else
     $J = J + 1;$ 
    Mark  $X$  as inside;
while  $J < \gamma$  do
  Make a vector  $X/X(r) = 0, r \in [1, 2N + 1];$ 
  foreach  $k \in [1, 2N + 1]$  do
    Make  $Y$  in  $[-1, 1]$  randomly;
     $X(k) = R(k) * (1 + Y * \alpha);$ 
   $J = J + 1;$ 
  Mark  $X$  as inside;

```

For each group or cluster obtained by the algorithm, is necessary to define an ARIMA statistical model where possible. Examples of its use in the field of supply-chain can be found in works like the made by Shukla and Jharkharia (2011) in which he develops an ARIMA (3,0,3) for forecasting demand for fresh produce. Another example can be found in the work of VS (Ediger & Akar, 2007) in using another ARIMA model for the treatment of fossil fuel demand. Another factor to consider in the choice of ARIMA models, is that the fitness test defined by Ljung and Box (1978) can be a first step in the automation of the verification of the goodness of fit of the ARIMA model to the time series. In some

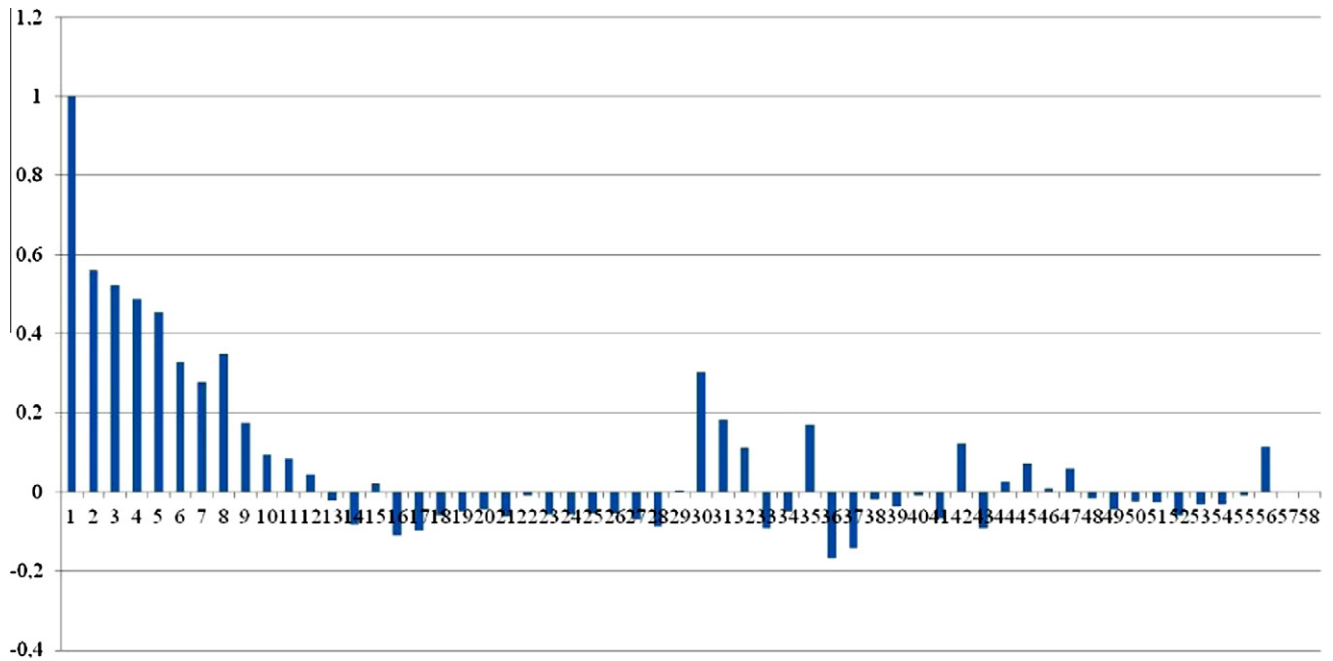


Fig. 9. Cluster 864550000.

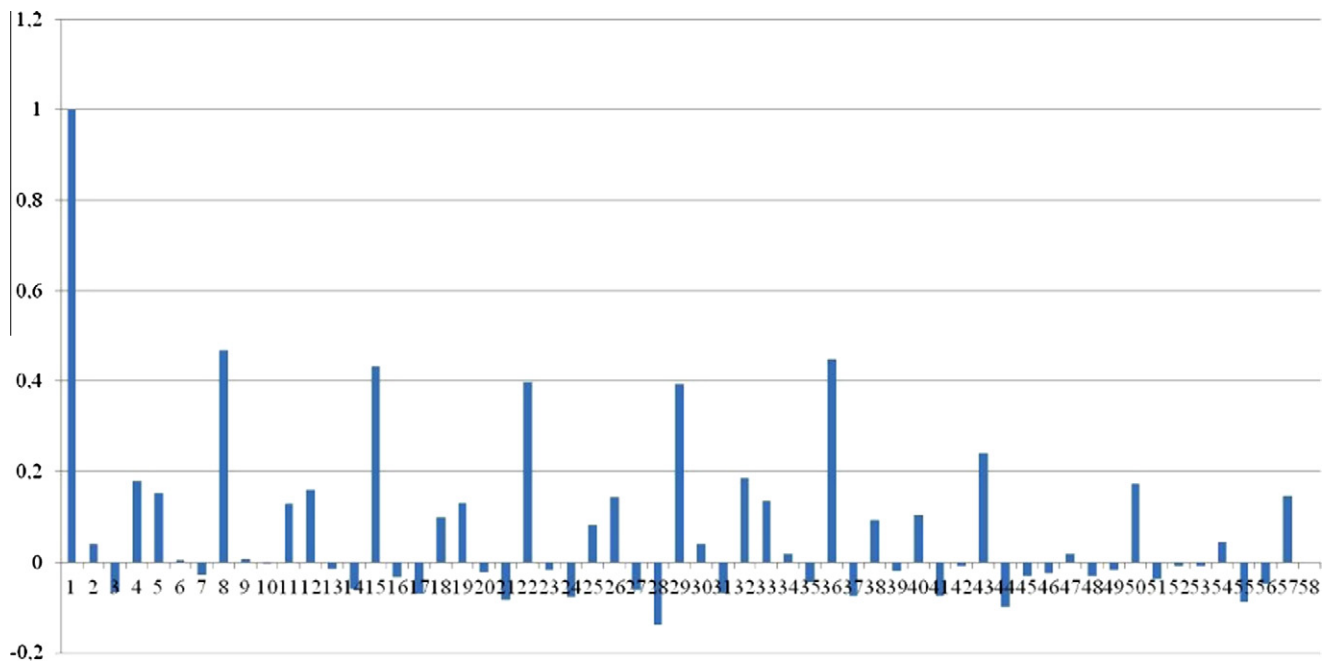


Fig. 10. Cluster 7346610000.

cases, where no ARIMA models can be applied, other estimators may be defined as, for example, the method of exponential smoothing Holt-Winters or basic estimators as the arithmetic mean, the last known value, etc.

### 3. Calculation

To calculate the vectors from the data set, we have used the open source R statistical library (R Development Core Team, 2008), executing the following script for each time series:

```
require(graphics);
data = read.csv("/TMP/inputData.csv");
attach(data);
objAcf <- acf(data, plot = FALSE, lag = 28);
objPacf <- pacf(data, plot = FALSE, lag = 28);
list <- union(objAcf$acf, objPacf$acf);
write(list, "/TMP/resultData.txt", ncolumns = 91, append = FALSE,
      sep = ";");
quit("no", 0, FALSE);
```

For the construction of support vector machines, we have used the LIBSVM library (Chang & Lin, 2011) taking into account the results obtained by Schölkopf, Smola, Williamson, and Bartlett

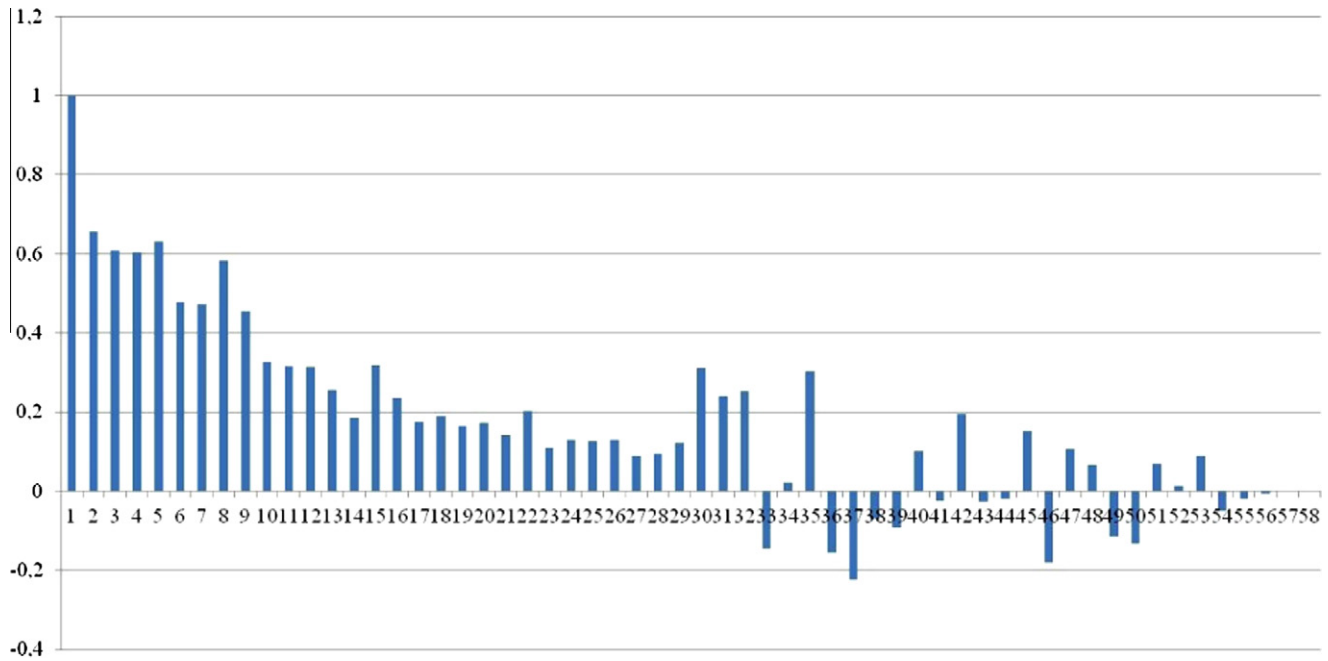


Fig. 11. Cluster 1425630000.

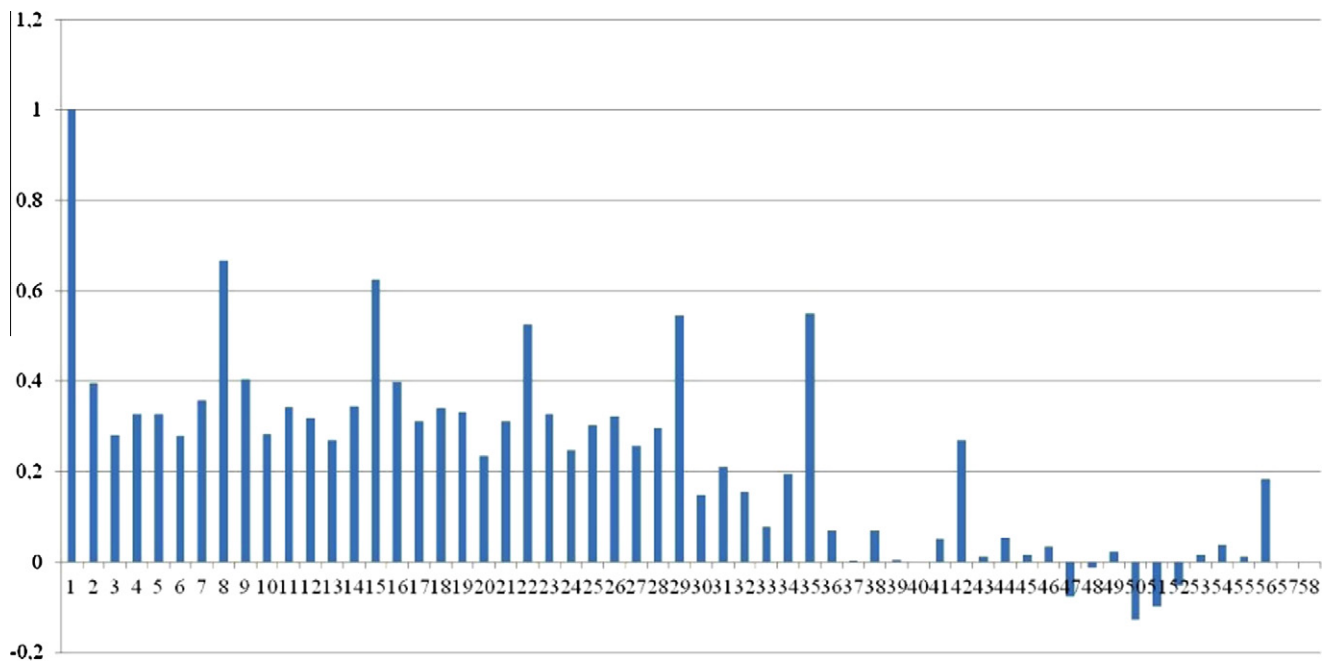


Fig. 12. Cluster 2896410000.

(2000), so we have used  $\nu$ -SVC machines with kernel based on radial functions (the theoretical model was defined by Schölkopf et al. (1997)).

The training files were generated by a Java program, with parameters  $\alpha = 0.2$ ,  $\beta = 3$ ,  $\gamma = 2000$ ,  $\theta = 2000$ .

#### 4. Results

Once the system is defined, it proceeds to perform a simulation on a sample of actual data, consisting of more than 14,000 series. To calculate the vectors of autocorrelation and partial

autocorrelation, 57 features have been used (28 + 1 for autocorrelation and partial autocorrelation for 28). This is justified by the fact that 28 days is the monthly sales cycle (4 weeks) in a distributor.

The Table 1 shows the clusters built by the system.

From the results, it is worth noting that with the method described, the 14,000 series are grouped only around 68 clusters. In addition, the first 12 grouped 98% of the sample series. Although it is obvious to mention, this would only be 12 different ARIMA defined models to make forecasts on 98% of all series.

Another highlight is the case regarding the cluster with the largest number of elements, represented by the series called



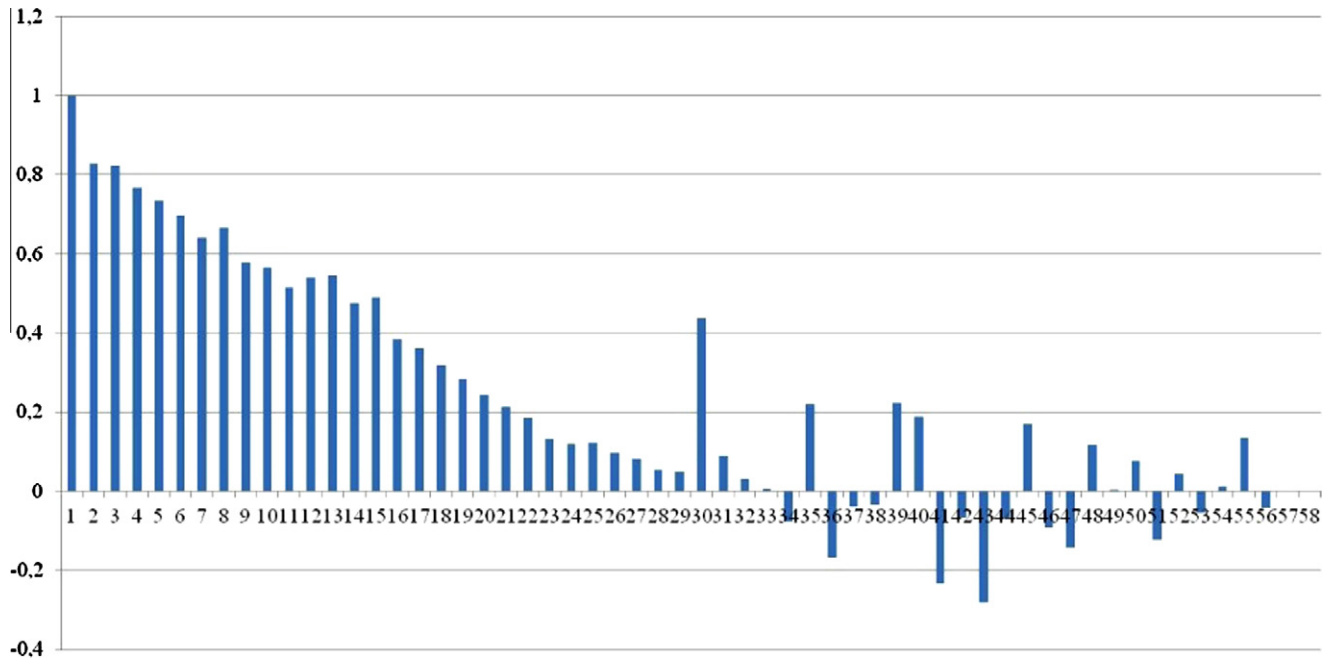


Fig. 13. Cluster 1308990000.

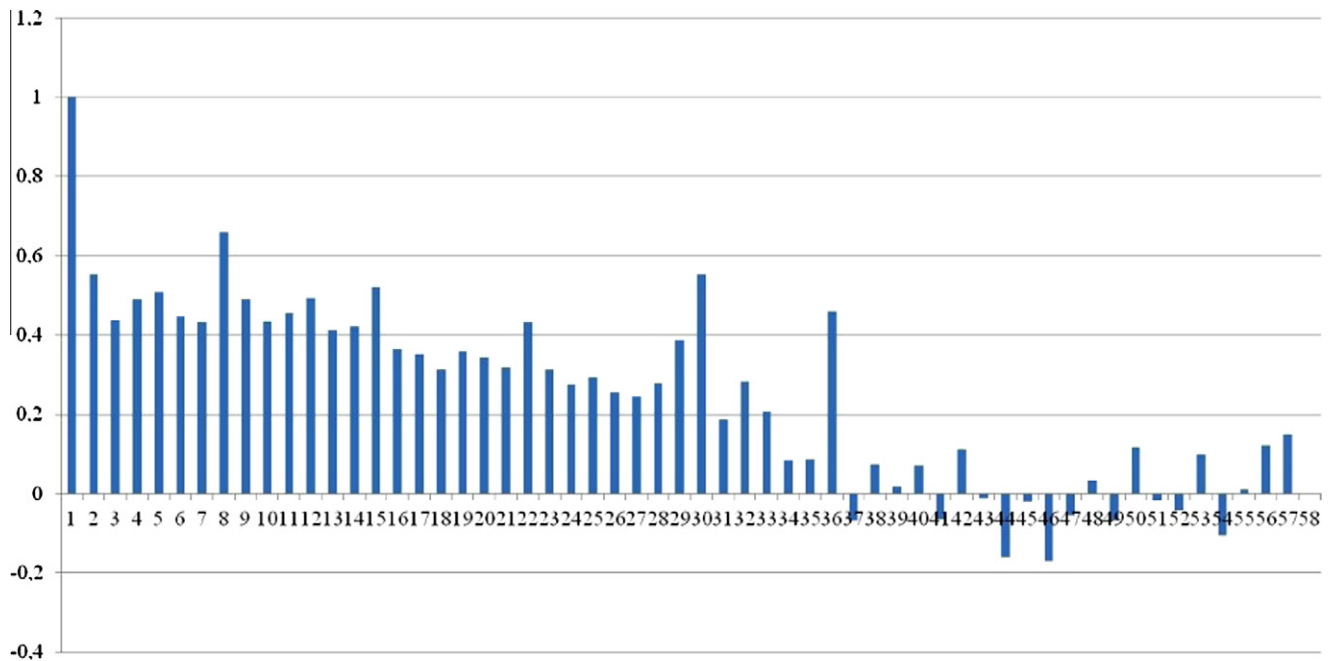


Fig. 14. Cluster 2354850000.

7073790000 (accounting for 43.75% of the series) because it is the cluster of the series that doesn't support an ARIMA model (the vector values are very close to zero). On the other hand, clusters represented by the series called 922670000 and 7346610000 (the two together represents about 20% of the series) show a strong dependence on the values of the series with period 7. This, from a business standpoint, it means that in these cases a good estimate can be constructed using the same days of previous weeks. That is, for making predictions for Monday of next week, the most reliable values are from the Monday of the previous weeks.

Figs. 5–16 are the vectors of reference of the main clusters.

## 5. Conclusion

The distribution sector is characterized by the rapidity with which changes occur; the volatility of customer buying trends is a well known concept (which in many cases becomes fashions). Because of this, big corporations must adapt to these changes as quickly as possible. This fact, added to the large volume of data they handle, makes the advantage of automating the discovery and classification of customer buying patterns almost an obligation.

In this paper, having a representative sample of actual data (less than 1% of the total, comprising all time series belonging to a store)



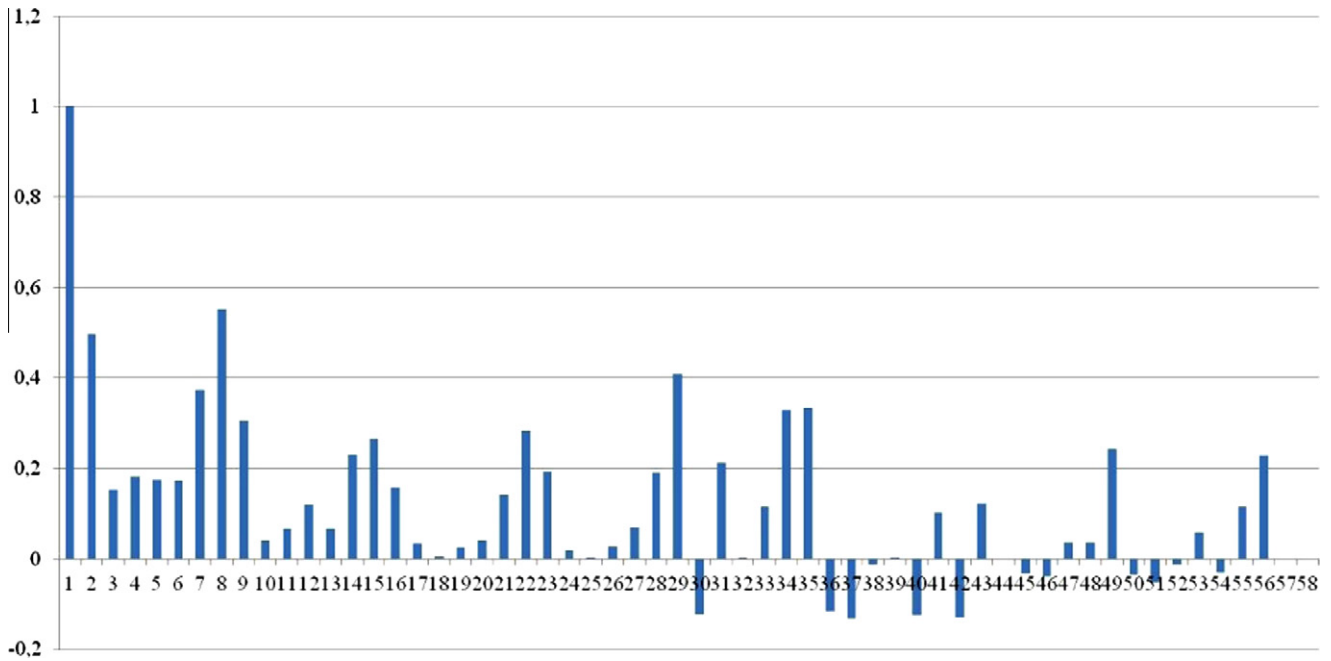


Fig. 15. Cluster 2448570000.

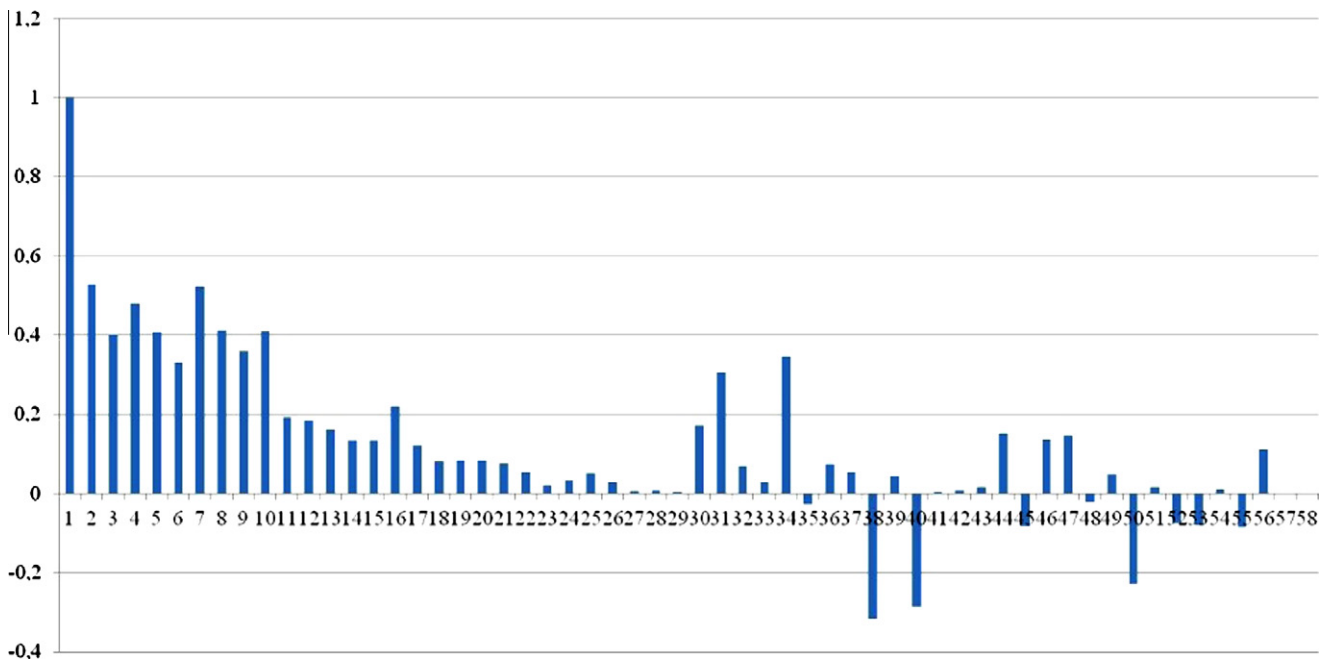


Fig. 16. Cluster 5606700000.

the results are more than satisfactory. This is because the time series classifier we have built using support vector machines, generates a very small list of clusters. This has a number of advantages in its application to all data.

On one hand, the task of finding statistical models is simplified (with only 12 models the 98% of all series are represented). This implies that the task of model creation and its maintenance can be done with a relatively small group of people. Although from

an operational point of view, it is necessary to calculate the ARIMA model coefficients for each time series.

On the other hand, as we assume that the sample is sufficiently representative (contains all SKUs of a store) presumably by using the method on the full set of series, the number of clusters will not be very high.

In order to implement the method, towards the creation of an autonomous system and taking into account the training/definition

(testing/validation data) of the support vector machines is an automatic process and that the calculation ARIMA coefficients is too, we find a system that only requires user intervention in specific tasks: analysis of the new clusters and definition of the corresponding ARIMA models.

Note also that having shared calculations (the generation of vector autocorrelation and partial autocorrelation) in the classification stages and construction of the models makes the hardware requirements smaller.

## Acknowledgements

This work was supported by the Ministerio de Industria, Turismo y Comercio (MITyC, Spain) through the Projects Avanza Competitividad I + D + I TSI-020100-2011-165 and TSI-020100-2010-482 and the Ministerio de Ciencia e Innovación (MICINN, Spain) through the Project TEC2010-18894/TCM.

## References

- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: forecasting and control* (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall.
- Cai, W., Chen, X., & Zhao, X. (2011). Supply chain performance comprehensive evaluation based on support vector machine, modeling risk management for resources and environment in China. *Computational Risk Management*, 3, 225–232.
- Chang, C. C., and Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2:27:1–27:27. Software available at <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>.
- Chatfield, C. (1978). The holt-winters forecasting procedure. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 27(3), 264–279.
- Chatfield, C., & Yar, M. (1988). Holt-winters forecasting: Some practical issues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37(2), 129–140.
- Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: the impact of forecasting, lead times, and information. *Management Science*, 46(3), 436–443.
- Danese, P., & Romano, P. (2011). Supply chain integration and efficiency performance: a study on the interactions between customer and supplier integration. *Supply Chain Management: An International Journal*, 16(4), 220–230.
- Ediger, V. S., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, 35(3), 1701–1708.
- Forrester, J. W. (1999). *Industrial dynamics*. Boston, EUA: Pegasus Communications.
- Kuo, R. J., & Han, Y. S. (2011). A hybrid of genetic algorithm and particle swarm optimization for solving bi-level linear programming problem – A case study on supply chain model. *Applied Mathematical Modeling*, 35(8), 3905–3917.
- Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: the bullwhip effect. *Management Science*, 43, 546–558.
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, 65, 297–303.
- Min, S. H., Lee, J., & Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 31(3), 652–660.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org>.
- Shukla, M., & Jharkharia, S. (2011). ARIMA models to forecast demand in fresh supply chains. *International Journal of Operational Research*, 11(1), 1–18.
- Shahrabi, J., Mousavi, S. S., & Heydar, M. (2009). Supply chain demand forecasting; a comparison of machine learning techniques and traditional methods. *Journal of Applied Sciences*, 9, 521–527.
- Schölkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T., et al. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing*, 45, 2758–2765.
- Schölkopf, B., Smola, A., Williamson, R., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation*, 12, 1207–1245.
- Wang, C.-C. (2011). A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export. *Expert Systems with Applications*, 38(8), 9296–9304.