Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 13 (2017) 956–963

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

Manufacturing Engineering Society International Conference 2017, MESIC 2017, 28-30 June 2017, Vigo (Pontevedra), Spain

Enhanced manufacturing storage management using data mining prediction techniques

A. Luque, F. Aguayo, J.R. Lama, E. González-Regalado

Escuela Politécnica Superior. Universidad de Sevilla. Virgen de África, 7, Sevilla 41011, Spain

Abstract

Performing an efficient storage management is a key issue for reducing costs in the manufacturing process. And the first step to accomplish this task is to have good estimations of the consumption of every storage component.
For making accurate consumption estimations two main approaches are possible: using past utilization values (time series); and/or considering other external factors affecting the spending rates.
Time series forecasting is the most common approach due to the fact that not always is clear the causes affecting consumption. Several classical methods have extensively been used, mainly ARIMA models.
As an alternative, in this paper it is proposed to use prediction techniques based on the data mining realm.
The use of consumption prediction algorithms clearly increases the storage management efficiency. The predictors based on data mining can offer enhanced solutions in many cases.

Peer-review under responsibility of the scientific committee of the Manufacturing Engineering Society International Conference 2017.

Keywords: manufacturing storage management, storage efficiency, time series forecasting, consumption prediction, data mining predictors

1. Introduction

Performing an efficient storage management is a key issue for reducing costs in the manufacturing process. And the first step to accomplish this task is to have good estimations of the consumption of every storage component [1].

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: amalialuque@us.es, faguayo@us.es, jrlama@us.es, eregalado@telefonica.net

2351-9789
Peer-review under responsibility of the scientific committee of the Manufacturing Engineering Society International Conference 2017.
10.1016/j.promfg.2017.09.166

Efficient storage control is a critical factor in the overall management of the company. Out of stock and overstock represent concealed losses. Lack of stock may hinder the completion of profitable sales. Stock surplus would not only imply a physical or technological impairment loss but it would also entail a capital lockup that could be advantageously spent instead. [2]

For making accurate consumption estimations two main approaches are possible: using past utilization values (time series); and/or considering other external factors affecting the spending rates.

Technologies applied to storage management contribute to the simplification of operations, cost reduction and to the improvement of information flows; while the main obstacles for its implementation are the high costs, the organizational culture and the inadequate process structuring. In relation to the use of ICT, a low level of implementation occurs within small and medium-sized enterprises (SMEs) and at a medium level in large companies. [3]

Time series forecasting is the most common approach due to the fact that not always is clear the causes affecting consumption. Several classical methods have extensively been used, mainly ARIMA models [4].

As an alternative, in this paper it is proposed to use prediction techniques based on the data mining realm [5].

According to [6], Data Mining refers to the process of extracting knowledge of databases. Its aim is to discover anomalous and/or interesting situations, trends, patterns and sequences in the data.

It is a stage inside the entire knowledge discovery process, which attempts to obtain patterns or models from collected data. Deciding whether the obtained models are useful normally requires a subjective user assessment. [7]

The aim, in the first place, is to demonstrate the viability of using time series analysis techniques for consumption estimations. In addition, techniques based on the data mining will be used in order to compare and elaborate on the previous results.

The works that has been carried out cover initial aspects of data preparation; processing tools choice; and analysis using time series techniques.

2. Methodology

In the first stage of data preparation, values that show an anomalous behaviour (null value or abnormally low) are eliminated, obtaining a set of values.

The expressed consumption values adopt the form of a time series. In the following sections, different prediction techniques on future series values will be tested. Many of the established prediction methods are based on the obtaining of a time series model. For that reason, and in order to be able to operate with each of the prediction techniques, data from each series will be split into three sets [8]:

Training sets. They constitute 70% of the total available data. A time series model normally depends on two or more parameters estimated using these sets.

Validation sets. They constitute 15% of the total available data. Occasionally, instead of using a single model of each series, several models will be obtained from a single family. The choice of the specific family element will be made according to these validation sets.

Test sets. They constitute 15% of the total available data. They are used to prove the effectivity of the chosen prediction technique.

Thus, for example, in a ARIMA model based forecasting, whose description will be specified further on, training data allow the estimation of parameters of several ARIMA models of different order; validation data to decide the optimal order of the ARIMA model (and its parameters); and test data to measure the efficacy of the forecasting.

In order to obtain quantitative indicators on the accuracy of certain forecasting methods, and to be able to compare models and methods, it is convenient to define certain accurate metrics [9]

While the error in some of its metrics represents a proper forecast accuracy indicator, it should not be overlooked that the ultimate objective is the reduction of stock levels. In order to determine the relationship between the forecast error and the stock it must be considered which supply policy has been adopted. Although this aspect would require significant efforts, outside the scope of this work [10, 11], a simplified method will be proposed, allowing us to assess the efficacy of forecasting stock methods.

The proposed supply policy draws from a stock (S) equals to the weekly consumption forecasting. This stock decreases throughout the week in a way that at the end it will be ideally 0. Nevertheless, forecasting errors may cause it to be slightly above or below this value.

At that time, the supply of a sufficient amount is produced so that the stock matches the following week forecasting. Once again the level decreases over the week until it ideally reaches the level 0. Once again forecasting errors may cause the final value of the week to be slightly above or below the null value.

If the process is repeated throughout all the weeks, it can be observed how there are weeks in which the stock is above zero, while in others it is below.

As it can be easily understood, a break of stock may occur with this policy and therefore supply shortage. This is highly undesirable, consequently, it is added to the previous procedure, the condition of adding to the forecasting a stock that is equal to the highest break of stock so that supply shortage will never happen.

In optimal conditions the normalize mean stock (NMS_o) will be:

$$NMS_o \equiv \frac{MS_o}{MC} = \frac{\frac{MC}{2}}{MC} = \frac{1}{2} \tag{1}$$

Hence it is deduced that the closer the NMS is to 50% the better will be the forecasting system. The forecasting efficiency (η) is defined

$$\eta \equiv \frac{NMS_o}{NMS} = \frac{\frac{1}{2}}{\frac{MS}{MC}} = \frac{MC}{2\,MS} \tag{2}$$

this metric will reach the 100% value in perfect forecasting conditions.

The consumptions observed in a case study considering 10 selected elements during 86 weeks have been used as the test set. On these data 4 classical forecasting methods are firstly considered [12]: one-year average, optimal moving average, exponential smoothing and AutoRegressive Integrated Moving Average (ARIMA). All of them are compared to the simplest prediction model: the persistence, i.e., the next value will be the current value.

It consists of making the forecast of a week as the average of the previous year consumption (of the last 52 weeks).

$$\hat{x}_n = \frac{1}{52}\sum_{k=1}^{52} x_{n-k} \tag{3}$$

Perhaps the simplest way of carrying out the time series value forecasting is to take the previous value, a widely used technique in many fields, especially in weather forecasting [13]. This method, named of persistence, is normally taken as a base in order to verify the enhancements of more elaborated methods. The forecast by the persistence method can be expressed as follows

$$\hat{x}_n = x_{n-1} \tag{4}$$

Moving average method [1]. According to this method the forecasting is calculated as the average of the previous values, which can be expressed as follows

$$\hat{x}_n = \frac{1}{M}\sum_{k=1}^{M} x_{n-k} \tag{5}$$

Exponential smoothing method [14], that can be expressed as follows

$$\hat{x}_n = \alpha \cdot x_{n-1} + (1-\alpha)\hat{x}_{n-1} \tag{6}$$

A method for time series value forecasting is to suppose that the aforementioned series behaves according to a certain stochastic model. A widely used model is explored [2] ARIMA: AutoRegressive Integrated Moving Average.

3. Results and discussion

A complete set of consumption estimations have been obtained for every proposed algorithm. The persistence estimation is considered the base algorithm for comparison purposes.

First of all, it is relevant to perform a comparison of techniques, so that it can be determined which of them behaves better to each product. To this effect, a quality forecasting indicator is the error committed. As the importance relies on the magnitude of error rather than on the sign, we will focus in the absolute error (AE: Absolute Error) from each prediction defined as

$$AE_k \equiv |\varepsilon_k| = |\hat{x}_k - x_k| \tag{7}$$

For the purpose of facilitating hereinafter the comparison with other products, normalized absolute error (NAE: Normalized Absolute Error) of each prediction will be defined as

$$NAE_k \equiv \frac{AE_k}{\bar{x}} = \frac{|\hat{x}_k - x_k|}{\frac{1}{N}\sum_{i=1}^{N} x_i} \tag{8}$$

With these definitions it is clear that we will have the same amount of NAE_k values as predictions are performed. Fig. 1 represents in a box plot the value distribution for the accumulated product consumption and for each forecasting method. The blue box includes the values between the percentiles 25 and 75; the red line represents the median; the black lines mark the maximum and minimum values (anomalous values not considered); and the red crosses establish the considered anomalous values (those that exceed in a 50% the interquartile range)

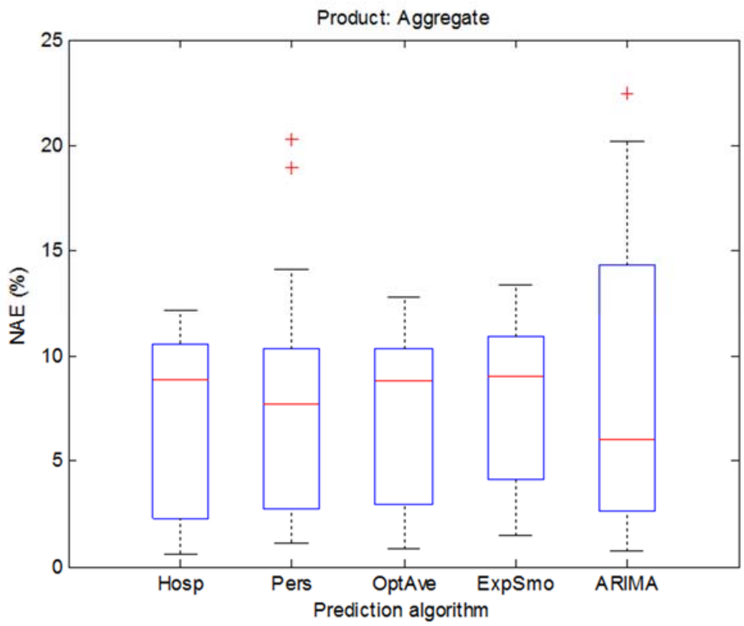

Fig. 1. NAE (%) for each prediction algortihm.

In the following graphic we can observe that related to the accumulated consumption, ARIMA method is the one presenting a lower median, although it has a higher value of error dispersion. This dispersion can cause more pronounced stock breakdowns with the consequent necessity of globally incrementing the stock in order to avoid these breakdowns. Therefore, it is convenient to have more direct and simpler metrics for techniques comparison.

One of these metrics, widely considered, is the Normalized Root Mean Square Error (NRMSE: Normalized Root Mean Square Error) previously defined. Fig. 2 includes the already stated metric for accumulated products consumption and for the analysed techniques. According to this metric the best prediction methods (those at a lower error) are the hospital method and the optimal moving average.

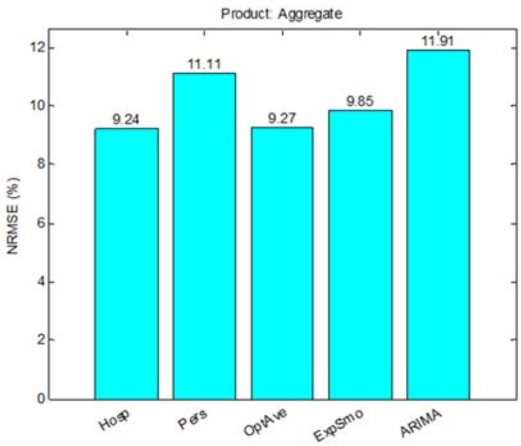

Fig. 2. NRSME (%) for each prediction algorithm.

Other metric that can be used is the forecasting efficiency, meaning the mean stock necessary in relation to the mean stock that would be required in the event of a perfect prediction. In Fig. 3 it is expressed this value for accumulated value consumption. Logically it is verified that the method providing the best results is the Exponential smoothing method, the same result as using the mean stock metric.

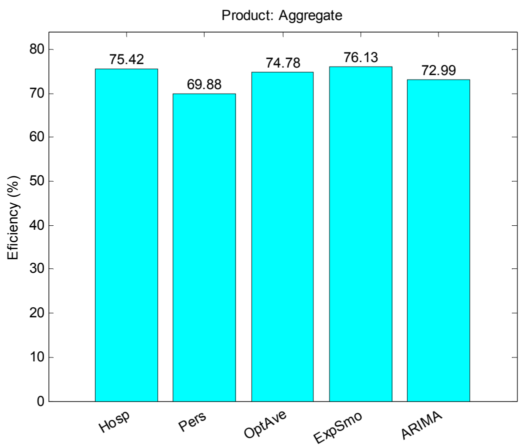

Fig. 3. Efficiency (%) for each prediction algorithm.

Summarizing, it can be concluded that for a non specified supply policy, the NRMSE is a good quality forecasting indicator. If the supply policy is known, it is preferable to use the mean stock (or the efficiency).

As it has been clearly highlighted in previous graphics, each product provides different results regarding the best prediction technique. This should not be surprising as the consumption of each of them takes the form of a time series with widely differing behaviours. It is logical that each of the techniques has a better adaptability (offer more accurate predictions) to certain series evolutions.

However, unless it the form of evolution of each product (which does not occur in this case) is previously known, it should be chosen a forecasting technique that has a good behaviour for the set of products. That is to say, a technique presenting a good compromise regardless of the consume evolution form.

In order to make this global comparison, it will be used in the first place the forecast error (RMSE) for each product and each forecasting method. The result is included in Fig. 4. In this graphic $RMSE_{ij}$ error values are reflected for each i product and each j method. Since the results are very diverse from a product to another, the value that it represents is the $rRMSE_{ij}$ meaning that the maximum $RMSE_{ij}$ value for this product, according to the following expression,

$$rRMSE_{ij} = \frac{RMSE_{ij}}{\max_j RMSE_{ij}} \quad (9)$$

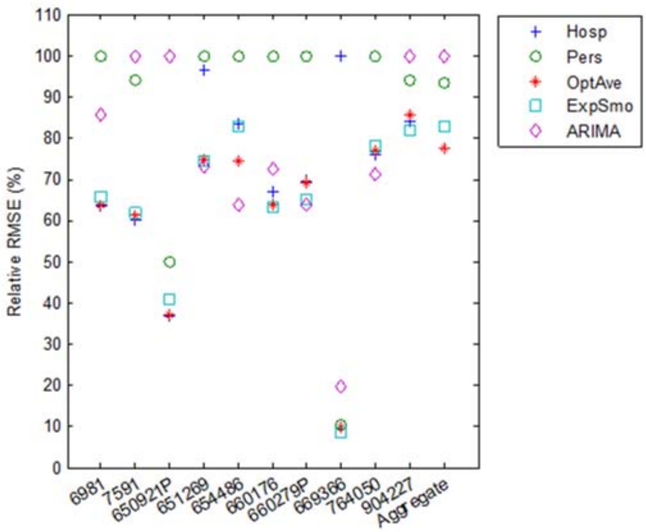

Fig. 4. Relative RMSE (%) for each product and prediction algorithm).

In the previous graphic we can distinguish the best methods for each product (with the RMSE criteria) but it is difficult to make a global comparison of the forecasting methods. In order to achieve it, an accumulated RMSE, $aRMSE_j$ will be defined for each j method, expressed as

$$aRMSE_j = \sum_i RMSE_{ij} \quad (10)$$

that is, the sum of error (RMSE) for every product. Fig. 5 contains the stated accumulated error. Following this criteria, the optimal moving average method (1820) slightly improves the hospital method (1918), and significantly improves (2502) the persistence method.

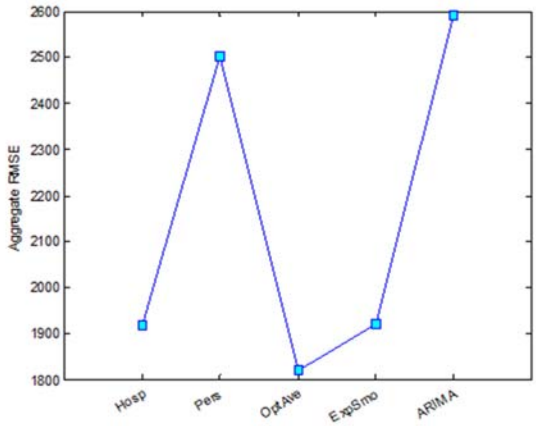

Fig. 5. Aggregate RMSE for each prediction algorithm.

If the supply policy is known instead of the error (RMSE) it can be used as a comparison metric the mean stock. The result is collected in Fig. 6. Mean stock values MS_{ij} for each i product and each j method are shown in this graphic. Since the results are very diverse from a product to another, in order to compare the results from all products, it is represented the rMS_{ij} value, relative to the maximum value of that product, according to the expression,

$$rMS_{ij} = \frac{MS_{ij}}{\max_{j} MS_{ij}} \tag{11}$$

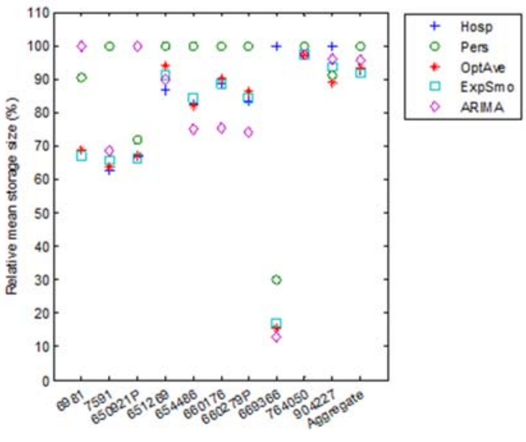

Fig. 6. Relative mean storage size for each product and prediction method.

In the previous graphic we can distinguish the best methods for each product (with the mean stock criteria) but it is difficult to make a global comparison of the forecasting methods. In order to achieve it, an accumulated MS, aMS_j will be defined for each j method, expressed as

$$aMS_j = \sum_i MS_{ij} \tag{12}$$

that is, the sum of the mean stock (MS) for every product. Following this criteria, the optimal moving average method (8816) nearly equals that of the hospital method (8775), and significantly improves the persistence method (10180).

In the performed test, every algorithm clearly supersedes the persistence base model. The optimal moving average technique is the most efficient among the classic estimators increasing a 15% the storage efficiency over the persistence method. But considering data mining predictors, the performances can be furthermore enhanced. The overall wining estimator is the linear regression technique, increasing a 20% the storage efficiency over the persistence method, an additional 5% of increasing over the classical predictors.

4. Conclusions

The use of consumption prediction algorithms clearly increases the storage management efficiency. The predictors based on data mining can offer enhanced solutions in many cases.

The main conclusions are the following:

The persistence method is, in every case, easy and largely outweighed by other tested techniques.

If a separate prediction technique is used for each product, there is no universal response about which one of them is the best forecasting method.

It should be noted that the optimal moving average method is always equal to or better than the persistence and the hospital methods, since those are particular cases of moving average.

If the same prediction technique is used for the forecasting of every product, the hospital method offers outstanding results. The optimal moving average method presents the best global results, offering a five-point increase of performance (RMSE criteria) over the previous.

Acknowledgements

This work has been supported by Telefónica, through the "Cátedra de Telefónica Inteligencia en la Red".

The authors thank Paloma Luna Garrido her availability and her support in editing of documentation, as well as language review.

References

[1] N.T. Thomopoulos, Springer Int. Publ. (2016) 1-27.
[2] R.L. Osteret, ReV. de automática. 4 (1) (1971), 103-108.
[3] A.A.C. Espinal, R.A.G. Montoya, J.A.C. Arenas, Estud. Gerenciales. 26 (117) (2010) 145-171.
[4] G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.
[5] G. Shmueli, N. R. Patel, P.C. Bruce, Data Mining for Business Analytics: Concepts, Techniques, and Applications in XLMiner, John Wiley & Sons, 2016.
[6] J. Molina, J. García, Técnicas de análisis de datos: aplicaciones prácticas utilizando Microsoft Excel y Weka. Universidad Carlos III de Madrid España, 2012.
[7] V.V. Asencios, Ind. Data, 7 (2) (2014) 083-086.
[8] I. Guyon, A scaling law for the validation-set training-set size ratio, AT&T Bell Laboratories. (1997) 1-11.
[9] R.J. Hyndman, A.B. Koehler, Int. J. Forecast. 22 (4) (2006) 679-688.
[10] N.T. Thomopoulos, Springer Int. Publ. (2015).
[11] I.C. Dima, M. Man, Modelling and Simulation in Management: Econometric Models Used in the Management of Organizations, Springer, 2015.
[12] A. Luque, F. Aguayo, J.R. Lama, E. Gonzalez-Regalado, A.M. Martín, Advanced Techniques for Stocks Management. Application to a hospital pharmacy stocks management, 20th International Congress on Project Management and Engineering (AEIPRO), Cartagena, 2016.