Applied Computer Science, vol. 14, no. 2, pp. 38–47
doi:10.23743/acs-2018-11

Submitted: 2018-05-14
Revised: 2018-06-11
Accepted: 2018-06-20

ANN, support vector regressions, Intermittent Demand Forecasting

Gamze Ogcu KAYA[*], *Ali TURKYILMAZ*[**]

INTERMITTENT DEMAND FORECASTING USING DATA MINING TECHNIQUES

Abstract

Intermittent demand occurs randomly with changing values and a lot of periods having zero demand. Ad hoc intermittent demand forecasting techniques have been developed which take special intermittent demand characteristics into account. Besides traditional techniques and specialized methods, data mining offers a better alternative for intermittent demand forecasting since data mining methods are powerful techniques. This study contributes to the current literature by showing the benefit of using data mining methods for intermittent demand forecasting purpose by comprising mostly used data mining methods.

1. INTRODUCTION

Companies need forecast values regarding different variables related to their business which may be a general variable like population forecast of the country or more specific variables like machine availability, costs, profit, inventory, and lead times. Demand forecasting is a specific type of forecasting which a very significant forecast needed by companies. All departments of the company plan themselves according to demand forecast since demand forecast is the prediction of sales values which is the primary revenue source for the company.

If the demand of item to be forecasted is smooth and continuous, then it is successfully predictable with traditional forecasting methods. On the other hand, when the demand does not occur at every forecasting period and has changing values, then it is hardly predictable with traditional forecasting methods. This kind of items are called intermittent demand items.

[*] Sampoerna University, Department of Industrial Engineering, Jakarta, Indonesia, gamzeogcu@gmail.com
[**] Nazarbayev University, School of Engineering, Astana, Kazakhstan, ali.turkyilmaz@nu.edu.kz

Data mining methods are powerful techniques which can solve complicated practical problems in various areas and are becoming more and more widespread day by day. They can be applied to time series forecasting problems effectively due to their distinguished properties. Data mining methods can handle many difficulties in the modeling of the time series like non-stationary. Moreover applying data mining methods is quite easy since there is no need for any complex mathematical formulations.

There is a lack of effective forecasting methods in intermittent demand forecasting area. This study shows the ability of data mining methods for accurate intermittent demand forecasting. Famous data mining methods which are artificial neural networks, support vector regressions and decision tree techniques are applied to simulated intermittent demand data as well as Croston's method which is the main ad-hoc intermittent demand forecasting method. The study is organized as follows: Section 2 includes the literature review, the following section contains experimental results, conclusions, and final comments are made in Section 4.

2. LITERATURE REVIEW

Intermittent demand is common in spare parts like in the automotive and aircraft industry, military maintenance and high-priced capital goods. Intermittent demand is commonly stated as spare parts demand in the literature (Hua & Zhang, 2006; Syntetos, 2007; Regattieriet, Gamberi, Gamberini & Manzini, 2005, Bacchettive & Saccani, 2012). The reason for this statement comes from the intermittent manner of spare parts items since they are only demanded when a maintenance problem occurs. It is clear that management of spare parts demand is crucial from technical and economic perspectives which make intermittent demand forecasting a very significant need. In case of a poor demand forecast, there might be difficulties in supplying the proper spare part item which leads to production interruptions and also technical difficulties. Kennedy stated that production interruptions result with direct economic losses since customer demand cannot be satisfied due to lost production and quality costs and also indirect economic losses due to customer dissatisfaction. So the aim of inventory planners is to reduce inventory costs while achieving high customer service levels (Kennedy, Patterson & Fredendall, 2002).

Although the availability of spare parts in inventory is very important due to economic and technical handicaps, two properties of spare parts items make it difficult to store high amounts of the items in inventory. The first property of spare parts is that: they aren't of the type which is called "general purpose" which makes them needed only if their specific requirement is realized. So it is not reasonable to store a high amount of spare parts in inventory rather than general purpose items with limited inventory storage capacity. The second property of spare parts items is their high unit values. Since spare parts have technical characteristics, their acquisition and also storage are costly for companies which store them in low quantities.

Intermittent demand forecasting is not an easy task since there are a lot of time periods with zero demand. Since intermittent demand is not stationary, traditional forecasting methods not applicable for intermittent demand forecasting. Traditional forecasting techniques are not capable of forecasting intermittent demand successfully since they assume stationary which is not the case in items having intermittent demand structure. This situation reveals the need for specialized techniques for intermittent demand forecasting.

Simple exponential smoothing method is the most common traditional method which is applied by companies dealing with intermittent demand. But in the work of Croston, it was first shown that exponential smoothing generally results with inappropriate stock levels (Croston, 1972). Croston's method corrected by Rao is a widely used approach for intermittent demand forecasting and inspires a lot of researchers (Rao, 1973).

In his work, Croston handles intermittent demand data as a combination of two elements which are the non-zero demand size Z_t and the inter-demand intervals P_t. He used two separate forecasts for the size of demand and for the demand arrival rate. Croston's forecasting method is superior to traditional single exponential smoothing forecast when applied to intermittent demand. It is pointed that single exponential smoothing predicts average demand higher than the real average which results in inappropriate stock levels, the reason was intermittent and zero demand occurrences at many time periods.

Nomenclature:

Z_t – demand in period t,

$\hat{Z}_t$ – the estimate of the mean size of a nonzero demand,

α – smoothing parameter, the value is between 0 to1,

$\hat{P}_t$ – the estimate of the mean interval between nonzero demands,

D_t – forecast value at time t.

Croston forecasts $\hat{P}_t$ separately the time between consecutive transaction.

P_t and the magnitude of the individual transactions Z_t.

If $Z_t = 0$,

$$\hat{Z}_t = \hat{Z}_{t-1} \text{ and } \hat{P}_t = \hat{P}_{t-1} \text{, forecast values remain unchanged} \tag{1}$$

If $Z_t \neq 0$,

$$\hat{Z}_t = \alpha Z_t + (1-\alpha)\hat{Z}_{t-1} \tag{2}$$

$$\hat{P}_t = \alpha P_t + (1-\alpha)\hat{P}_{t-1} \tag{3}$$

The forecast of demand per period at time t is found as a division of demand value by inter-arrival time:

$$D_t = \frac{\hat{Z}_t}{\hat{P}t} \tag{4}$$

where: Z_t – demand in period t, $\hat{Z}_t$ – demand forecast in period t for the forecast of next demand size, α – smoothing parameter taking a value between zero and one, P_t – time between consecutive transactions, $\hat{P}_t$ – forecast of the demand interval, D_t – the final forecast value.

Data mining methods are powerful techniques presenting an alternative approach to conventional techniques. They are good at solving complicated practical problems in various areas, thus increasing their popularity day by day. Data mining is the non-trivial process of extracting implicit, previously unknown and potentially useful information from data which can also be referred to as knowledge discovery in the database. Data mining techniques can be used for a number of objectives such as clustering, classification, data summarization, discovering dependency networks, and detecting anomalies as stated by Punjari (2006).

Data mining techniques differ from other methods with their learning ability. Once they are trained they can perform prediction with high accuracy and high speed. Moreover, they have the ability to handle noisy data and non-linearity which make them applicable to most of the problems. Data mining methods consist of many models like artificial neural networks, expert systems, genetic algorithms, support vector regressions, fuzzy logic, decision tree and hybrid systems which are formed as a combination of two or more methods. In this study, we employed artificial neural networks, support vector regressions and decision tree techniques since they are commonly used data mining methods.

Artificial neural networks are formed as an imitation of the brain and its associated nervous systems. The model mainly contains an input layer, an output layer, and one or more hidden layers. The layers include interconnected neurons each of which is connected to other neurons in the subsequent layer. The neurons at hidden layer and output layer use a non-linear transfer function for processing their layer inputs by summing the product of each input by its weight in order to get a result (Chau, 2006). Figure 1 demonstrates neural network architecture with an input layer of three neurons, a hidden layer of two neurons and an output layer of one neuron.

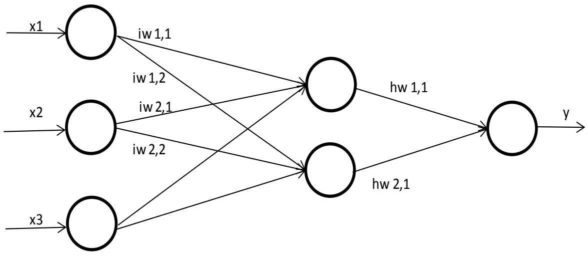

Fig. 1. Illustration of artificial neural network architecture

The second data mining method used in this study is Support vector regression which is the application of Support vector machines to time series forecasting. Support vector machine was introduced by (Vapnik, 1995). It is a powerful classification method that separates two classes by a hyperplane. Support vector machines are based on structured risk minimization principle since the method tries to minimize upper bound of error by obtaining a separating hyperplane as wider as possible. This property of the method is the source of its strength over traditional methods. Figure 2 illustrates the simplest case of two linearly separable classes where empty circles represent –1 and full circles represent +1. The hyper-plane separating two classes has the maximum margin between the samples of the two classes.

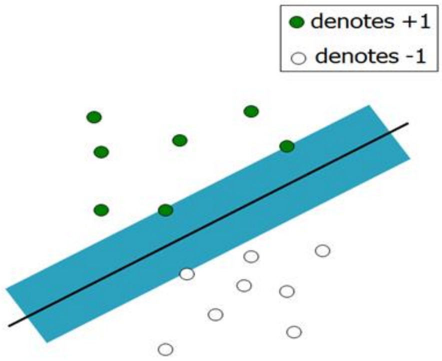

Fig. 2. The simplest illustration of support vector machines

The third data mining method used in this study is decision tree which aims to predict the value of a target variable based on several input variables. Decision tree models can also be learned like artificial neural network and support vector regression by splitting the source set. Mainly decision tree method uses a tree structure to build the classification models. Leaf nodes represent a decision and groups of instances that receive the same class label. Based on feature values of instances, they are classified in the decision tree according to labels of leaves. Each node in the tree represents a feature of an instance and each branch represents a value.

Figure 3 is a representation of a simple decision tree (Mitchell, 1997). There are three nodes in the decision tree which are outlook, humidity, and wind. The branches of the attributes show the possible values that the node can take. For this example, wind node can take on the values strong and weak. The leaf nodes at the bottom of the tree are the possible labels that can be made by classifying an instance using the tree. In this decision tree, an instance can be labeled as either a yes or no.

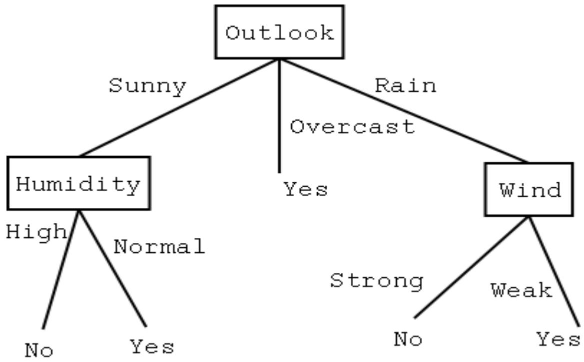

Fig. 3. Simplest illustration of decision tree

There are many variances of decision trees such as ID3, C4.5, and C5. But in this study, C5 is used since it has superior performance over others. For instance, C5 is faster than C4.5, it gets smaller decision trees than C4.5. Moreover, C5 is a better alternative than ID3 for handling the training data with missing attribute values (Pandya & Pandya, 2015).

3. EXPERIMENTAL RESULTS

This section includes the application of forecasting methods on simulated intermittent data and the performance results and comparisons. Common data mining methods which are artificial neural networks, support vector regressions and decision tree techniques are employed. Moreover, Croston's method with a smoothing parameter value of 0.2 is applied to the same data as the benchmark method since it is a very significant method in intermittent demand forecasting area since it is the first and main ad-hoc intermittent demand forecasting method.

3.1. Intermittent data

Since intermittent demand characteristics are not common for most of the fast moving products, there is limited real data. This situation leads us to generate artificial intermittent data in order to assess the performance of methods. In order to generate artificial demand data, the intermittent package of R software is used. 400 intermittent demand series are generated each of which is composed of 300 data points. Statistical details regarding the simulated intermittent data are presented in Table 1.

Tab. 1. Descriptive statistics of the simulated data set

Statistics	Demand Size	Demand Interval
Mean	4.98	3.01
Std. Dev.	0.17	1.98
Minimum	4.36	1.03
1. Quartile	4.88	1.05
Median	4.98	2.65
3. Quartile	5.08	4.97
Maximum	5.66	6.05

In Figure 4, a simulated intermittent demand time series data is demonstrated. As can be seen from the figure, there is an interval between demand occurrences with changing values and demand values are also changing in value.

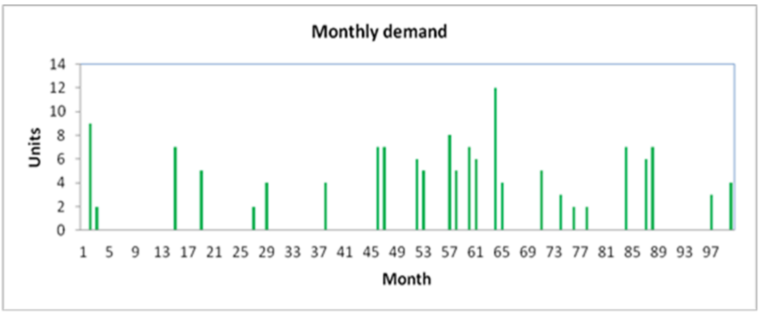

Fig. 4. An example of intermittent data

3.2. Accuracy evaluation and comparison

The prediction performance of the methods are evaluated using the following statistical metrics, namely, the root mean squared error (RMSE), mean absolute deviation (MAD), and alternative mean absolute percentage error (A-MAPE). For intermittent demand forecasting performance evaluation, the original MAPE cannot be used since actual demand values take zero value for most of the time periods which results in division by zero error. So, we used A-MAPE introduced by Hoover which is the ratio of mean absolute deviation to mean value. The employed metrics are calculated as follows (Hoover, 2006):

Let Y_t denote the observation at time t and F_t denote the corresponding forecast value Then define the forecast error as $e_t = Y_t - F_t$. Then:

$$\mathrm{RMSE} = \left(\frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2 \right)^{1/2} \quad (5)$$

$$MAD = \frac{1}{n} \sum_{t=1}^{n} | Y_t - F_t | \quad (6)$$

$$A - MAPE = \frac{\frac{1}{n} \sum_{t=1}^{n} | Y_t - F_t |}{\frac{1}{n} \sum_{t=1}^{n} Y_t} \quad (7)$$

3.3. Forecast results

All four methods are applied to intermittent demand data. Since data mining methods have a learning process, data is divided into two as the training set and testing set where the method learns the structure in training stage. Training set includes the first 75% time periods for each intermittent time series (215 data points) and the last 25% time periods form the testing set (75 data points). In the following table, performance results for the training set are presented.

Tab. 2. Results for the training set

Performance	ANN	SVM	C5	Croston
RMSE	3.11	1.76	4.61	5.26
MAD	2.27	0.83	3.11	4.10
A-MAPE	0.76	0.28	1.04	1.37

As can be seen from the table, use of data mining method is very promising for intermittent demand forecasting since they have superior performance over Croston's method based on all performance metrics. This is an important result since Croston's method is a specialized intermittent demand method which has shown better performance over traditional methods. Moreover, Support vector machine is the best performing method among the artificial methods. Performance results for the testing set are presented in Table 3.

Tab. 3. Results for the testing set

Performance	ANN	SVM	C5	Croston
RMSE	2.66	1.35	2.93	4.38
MAD	2.17	0.67	2.34	3.16
A-MAPE	0.69	0.21	0.75	1.01

The performance results for the testing set are in line with the results for the training set. All data mining method has lower error values than the Croston's method with respect to all performance metrics.

4. CONCLUSIONS AND PERSPECTIVES

Due to peculiar characteristics of intermittent demand, traditional forecasting methods perform poorly. Traditional methods assume demand to be stationary where intermittent demand data is not stationary. More complicated methods like data mining methods are more appropriate to be used for intermittent demand forecasting since they have been shown to perform successfully in most of the research.

Considering both the importance and the difficulty of accurate demand forecasting for intermittent demand items, this study focuses on applying data mining methods since they are powerful methods which are able to deal with non-stationary data. Since intermittent demand time series have many time periods with zero demand and have variable demand sizes, they possess non-stationary pattern.

Common data mining methods which are artificial neural networks, support vector regressions and decision tree techniques and Croston's method which is developed specifically for intermittent demand forecasting are applied to simulated intermittent demand data. Results showed that use of data mining methods are beneficial with their high-performance values and support vector machine is the most appropriate one since it resulted with the lowest error value among all techniques.

Finally, considering the lack of effective forecasting methods for intermittent demand, it would be interesting to compare the employed data mining methods with other methods such as genetic programming, fuzzy logic and hybrid methods. These perspectives are let as future works.

REFERENCES

Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, *40,* 722–737.

Chau, K. W. (2006). A review on the integration of data mining into coastal modeling. *Journal of Environmental Management*, *80,* 47–57.

Croston, J. F. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, *23*, 289–304.

Hoover, J. (2006). Measuring Forecast Accuracy: Omissions in Today's Forecasting Engines and Demand-Planning Software. *International Journal of Applied Forecasting*, *1*, 32–35.

Hua, Z., & Zhang, B. (2006). A hybrid support vector machines and logistic regression approach for forecasting intermittent demand of spare parts. *Applied Mathematics and Computation*, *181*, 1035–1048.

Kennedy, W. J., Patterson, J. W., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *Int. J. of Production Economics, 76*, 201–215.

Mitchell, T. (1997). *Machine Learning.* Boston: McGraw-Hill.

Pandya, R., & Pandya, J. (2015). C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. *International Journal of Computer Applications*, *11*, 718–21.

Punjari, A. K. (2006). *Data Mining Techniques*. Universities Press Private Limited.

Rao, A. V. (1973). A comment on: forecasting and stock control for intermittent demands. *Operational Research Society*, *24*, 639–640.

Regattieri, A., Gamberi, M., Gamberini, R., & Manzini, R. (2005). Managing Lumpy Demand for Aircraft Spare Parts. *Journal of Air Transport Management*, *11*, 426–431.

Syntetos, A. A. (2007). A note on managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*, *13*, 166–167.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory.* New York: Springer.