Machine Learning For Intermittent Demand Forecasting

[1]Zainab Assaghir, [2]Sara Makki, [3]Hassan Zeineddine

Lebanese University
Applied Mathematics,
Lebanese University, Beirut, Lebanon

***Abstract*—Forecasting demand is a crucial stepin inventory control; its accuracy affects the managementof storage and materials, and therefore the profitability of the company.When the demand is intermittent, infrequent, and highly variable when it occurs, itcauses problems when using traditional statistical models and forecasting techniques(constant forecasts, zero forecasts, unsuitable accuracy metrics, etc).This work is a comparison between several methods, including the Croston's method specific for intermittent demand, and other important machine learning techniquessuch as artificial neural networks, using a database of different spare part products.Our objective is to forecast the demand of these products in the most accurate way possible, and to apply methods that can outperform the conventional methods.**

***IndexTerms*— Intermittent, demand forecasting, Croston's method, artificial neural networks, support vector machines.**

I. INTRODUCTION

Demand forecasting is the prediction of sales of a certain product in a futuretime period using the historical demand; it is becoming very important in the business and economic sector. Companies nowadays make plans and decisions regarding production, taking into consideration costs and profits, based on sales forecasting. The purpose of forecasting demand is to spot the seasonal effect and long term changes in order to make strategies with the purpose to manage stocks and to satisfy the customers' needs, because stocking-out can be harmful.

Demand forecasting is usuallydone using time series methods like single exponential and weighted moving average. The problem addressed in this study is the forecast of items requested infrequently, sporadically and with high variability; known as "intermittent demand". This kind of demand can hardly be forecasted using traditional time series. Several issues should be studied first, such as the frequency of demand of the product, the inter-demand interval and the variance of demand.

Categorization of demand patterns aims to classify the items into groups or categories in order to choose the best forecasting method suitable to each category. According to Silver, intermittent demand is defined as "infrequent in the sense that the average time between consecutive transactions is considerably larger than the unit time period" [1]. Smart considered an intermittent demand series where at least 30% of the demand is zero [1].An important categorization scheme was introduced by Syntetos et al. [2]. They compared the Mean Square Error (MSE) of forecasts generated by three forecasting methods: Croston's method, Syntetos and Boylan approximation, and single exponential smoothing, and they defined four categories: erratic, smooth, intermittent and lumpy, using two parameters: the squared coefficient of variation CV^2 which is the standard deviation of demand divided by the demand average over a number of time periods; and the average inter-demand interval ADI which is the average number of time periods between two successive non-zero demand. The demand series is classed according to threshold values of these two parameters. A high CV^2value means that there is a high variation in the size of demand, and a high ADI value indicates a demand pattern with many time periods having zero demand.

The intermittency of demand makes it difficult to forecast the future sales because these products are requested infrequently, randomly, with multiple periods of zero demand and a slow-moving nature. Croston was the first to propose a new method specified to forecast intermittent demand known as the Croston's Method (CM) [3]. Syntetosand Boylan [4] identified the limitations of CM and found a bias caused when assuming that demand size and demand interval are independent. They introduced a modification to this method known as the "Syntetos and Boylan Approximation" [5]. Another method was proposed by Willemain et al. [6] known as the Willemain, Smart and Schwarz bootstrap method (WSS), they modeled the autocorrelation between the series of zero and non-zero demand with a two states, first order markov chain, and generated a forecast series of zero and non-zero demand, and gave numeric values to non-zero demand using a bootstrap method, from the non-zero values in the past.Bao et al. proposed a Support Vector Machines (SVM) to forecast intermittent demand [3]. More recently, Pour et al. [7] presented a hybrid method based on Artificial Neural Networks (ANN) to forecast lumpy demand time series.More approaches were adapted recently, Nikolopoulos et al. [8] used temporal aggregation to forecast intermittent demand, their method is known as "Aggregate - Disaggregate Intermittent Demand Approach" (ADIDA); the basic idea is to combine periods into blocks, which will reduce the zero demands, and then to forecast the demand using the aggregated series and normal forecast methods, and finally the forecast is disaggregated or decomposed into time periods as the original series. Even though these methods reduce the number of historical data, it improved the performance of the forecast methods when applied without aggregation.

The paper is organized as follows, section II presents the materials and a brief description of forecasting methods used; section III summarizes the obtained results using an example of different spare part products, and finally a conclusion.

II. Materials And Methods

Data used in this work consist of a set of spare parts demand in the iron and steel sector. Out of 10 items, we select three items with different statistical characteristicsfor the studyas shown in tab 1. Data represent the weekly demands of the products over a year. Several forecasting methods are used in order to predict consumption of each product. We use 70% of the data as a train set to develop the models, and 30% as a test set to compare the performance of the methods.

Table 1 Descriptive analysis of the selected items

Item	Percentage of zero values	Mean	Standard deviation	Minimum	Maximum	CV^2	ADI
1	35.8%	56.20	121.69	4	664	4.68	1.55
2	20.7%	15.50	13.27	1	53	0.73	1.21
3	69.8%	100	36.33	1	200	0.13	3.12

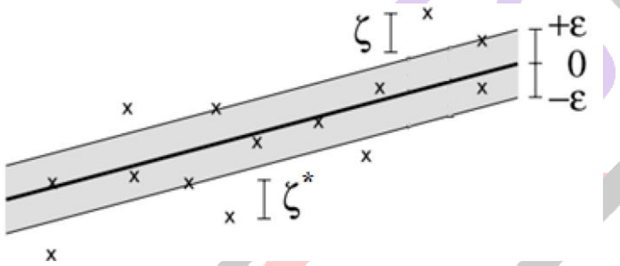
Croston's Method

First introduced by Croston in 1972, this method consists of estimating the demand size (assumed to follow a normal distribution with mean μ) and the interval between non-zero demand (geometrically distributed with mean p) separately using exponential smoothing with a common smoothing parameter α. These estimates are updated each time when demand occurs, otherwise it's unchanged [3].

Support vector machines

Support vector machines (SVM) were first developed by Vapnickand Cortes in 1995 to resolve classification problems, but recently SVM is used as a regression tool.Our aim is to find a regression function $f(x) = \langle w, x \rangle + b$ as flat as possible, which means having a small value of $\|w\|$,that estimates the relation between the inputs vector (x) and the target values (y), with at most ϵ deviation (Fig 1).

Figure 1 ϵ-regression using support vector machines



To make sure that this problem is feasible and that such a solution exists, we allow for some errors by adding slack variables ζ and ζ^* and a cost parameter C to indicate the trade-off between the flatness of f and the amount up to which deviations larger than ϵ are acceptable. The problem to be solved is then summarized as follows:

$$\text{Minimize} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \qquad (1)$$

$$\text{Subject to} \begin{cases} y_i - \langle w, x_i \rangle + b \leq \epsilon + \zeta_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \zeta_i^* \\ \zeta_i \geq 0, \zeta_i^* \geq 0 \end{cases}$$

This problem is solved using a dual problem transformation and Lagrange multipliers detailed in [9].

Most of the cases, it is difficult to find a linear function that fits the model, so it is necessary to find a non-linear SVM algorithm. This is done by mapping the inputs in another feature space $\mathcal{F}$ of higher dimension where they are linearly separable using a mapping functionϕ. Since the SVM algorithm only depends on the dot products between data points [9], it is sufficient to define a function called "Kernel function" defined as:$k(x, y) = \langle \phi(x), \phi(y) \rangle$, without the need to explicitly find $\phi(x)$because it may be too complicated. This is known as "the Kernel trick".

Artificial neural networks

Artificial Neural Networks (ANN) are quantitative models, first developed by McCulloch and Pitts in 1943who introduced the first notion of the simple neuron model [10].Their purpose was to imitate the behavior of the human brain and to create artificial systems able to do difficult and complicated computations analogous, such as pattern recognition. It is an adaptive model; it uses examples to train the network in order to connect inputs with outputs through estimated parameters, and thus creating a generalization beyond the training data that will be used for prediction.

A neural network is the connection of elementary objects called “the simple neuron”. First, inputs are fed into each simple neuron, thus each input is multiplied by a specific weight. These weights are first set randomly, then using a training set which is a combination of inputs and their matching outputs and a training algorithm, the weights are optimized. These weighted inputs are then added together with a bias and then an activation function is applied to obtain the final output of the neuron.

The multilayer feed-forward perceptron is one of the architectures of a neural network composed of several layers: input, hidden and output layers where connections are unidirectional; each layer transfers the input to the next one. An input layer reads the incoming signals, and the output layer provides the system's response. A neuron in a hidden layer is connected at input to each of the neurons in the preceding layer and output to each neuron in the next layer.

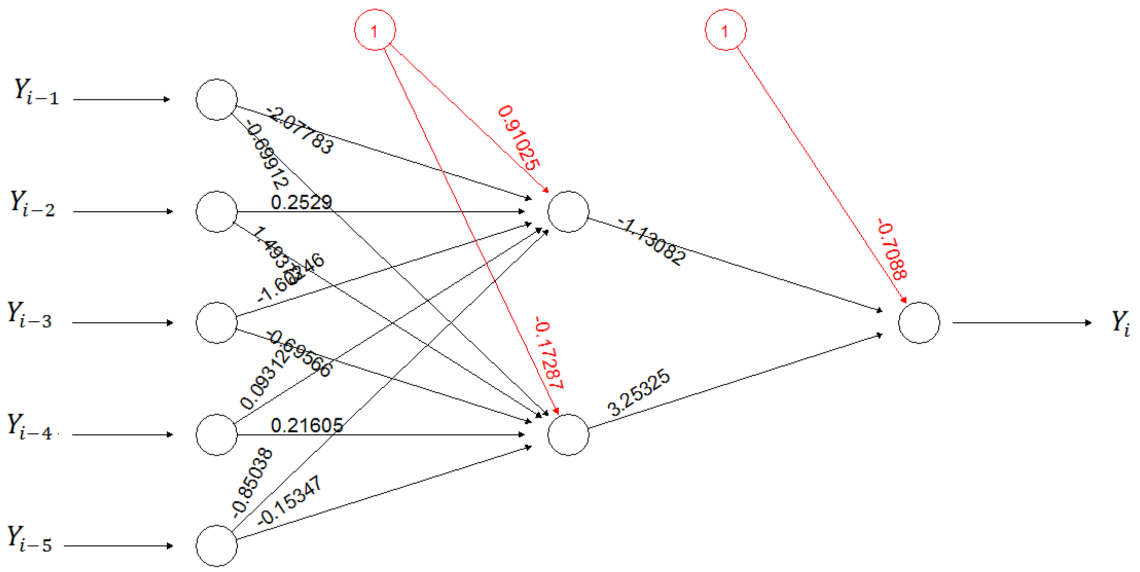
The back-propagation algorithm is the learning algorithm commonly used, it's a supervised learning technique that consists of two steps, first the feed-forward computation, generating an output using the activation function and the weights initialized randomly. Second, the computed output is compared to the expected one, creating an error signal that takes into consideration the derivative of the activation function. Here, the activation function of the network should be differentiable. The error signal is then propagated back into the network layer by layer using the weights. Finally the weights are adjusted in a way to minimize the error.

III. RESULTS AND DISCUSSION

Forecasts are created for all three items, using the methods described in the previous section. The smoothing parameter used for CM is equal to 0.2. For SVM, the five demand values preceding target period were used as regressors to predict the demand value. The kernel function used is the Gaussian Radial Basis Function (RBF): $f(x) = e^{-\gamma\|x-\nu\|^2}$, with γ = 0.2 and ν = 0.5. The cost parameter C was set to 1.

For the NN, a multilayer feed-forward network was applied; that takes five inputs corresponding to the five demand values preceding target period, one hidden layer with twonodes, and the output layer.The network's output is between 0 and 1, thus the inputs were normalized using the min-max scale (subtracting the minimum value of the input variable and dividing it by maximum-minimum), and then the predicted values were denormalized. The algorithm used to optimize the weights and biases is the resilient back-propagation algorithm (Rprop) which is similar to the original back propagation algorithm but with reducing the impact of the partial derivative of the activation function on the weights adjustment [11]. The activation function used is the sigmoid function: $f(x) = {}^{1}/_{(1+e^{-x})}$. The NN result for item 1 is shown in Fig 2.

Figure2 ANN plot for item 1



The choice of a performance measure to compare the obtained results;is not straightforward. Conventional accuracy metrics are either incorrect or gives undefined results due to the zero value in the series. We use the Mean Absolute Error(MAE) and the Mean Absolute Scaled Error (MASE) which was introduced by Hyndman and Koehler [12], and recommended it to measure the intermittent demand forecast errors. The MASE consists of scaling the error $|Y_i - F_i|$based on the in-sample MAE from the naïve method. MASE is defined as:

$$MASE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - F_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}$$

Where F_i is the predicted value of demand, and n is the dimension of the test set.A MASE less than one can be interpreted that the forecast is better than the average one-step naive forecast computed in-sample. On the contrary, if it's higher than one it means the forecast is worse than the average one-step naive forecast computed in-sample. The following tables show the MAE and MASE, calculated for each item and methods.

Table 2Methods' performance for each item

Item	MAE		
	CM	*SVM*	*ANN*
1	35.31	13.68	7.31
2	12.56	12.56	11.06
3	38.5	42.81	29.68

Item	MASE		
	CM	*SVM*	*ANN*
1	0.468	0.181	0.097
2	1.009	1.009	0.889
3	0.77	0.856	0.594

The error values show that ANN outperforms the other two methods, and that both SVM and ANN are better than CM for intermittent demand forecasting. The results show that the neural network is the better method to forecast this type of demand for all items according to both MAE and MASE. The advantage of using this method is minor when the item is not so intermittent (item 2); but it's significant for the other items (1 and 3). Regarding CM and SVM, the higher the CV^2, the better it is to use SVM, If CV^2 is very low (0.13 for item 3), CM's results are better than SVM's.

IV. Conclusion

Accurate forecasts are essential to the controlling of the ordering, storage, and components or materials that a company uses in the production of the items. The problem with intermittent demand is the high percentage of zero values, and the variability of demand when it occurs; which makes this kind of demand series very difficult to forecast.

In this study, we compare the performance of Croston's method specific for intermittent demand, support vector machines, and artificial neural networks.

The advantage of using these SVM and ANN rather than CM is their ability to describe the relation between the non-zero demand and inter-demand intervals supposed irrelevant in CM. Moreover, ANN and SVM are able to catch the dynamics of the demand series because they are not constrained to the abstract character of traditional mathematical approaches like CM that generates constant forecasts, and sometimes zero forecasts if the demand series have a higher number of zero values. These methods relied on the detailed historical demand series of each item alone, rather than unconvincing mathematical assumptions (normal and geometric distribution).

References

[1] V. Varghese and M. Rossetti, "A Meta Forecasting Methodology for Large Scale Inventory Systems with Intermittent Demand",Proceedings of the 2009 Industrial Engineering Research Conference.

[2] A. A. Syntetos, J. E. Boylan, and J. D. Croston. "On the categorization of demand patterns." Journal of the Operational Research Society, 56(5):495-503, 2005.

[3] E. Babiloni, M. Cardos, J. M. Albarracin, and M. E. Palmer. "Demand categorization,forecasting, and inventory control for intermittent demand items." SouthAfrican Journal of Industrial Engineering, 21(2):115-130, 2010.

[4] A. A. Syntetos and J. E. Boylan. "On the bias of intermittent demand estimates." International Journal of Production Economics, 71(1-3):457-466, 2001.

[5] A. A. Syntetos and J. E. Boylan. "The accuracy of intermittent demand estimates." International Journal of Forecasting, 21(2):303-314, 2005.

[6] T. R. Willemain, C. N. Smart, and H. F. Schwarz. "A new approach to forecasting intermittent demand for service part inventories." International Journal of Forecasting, 20(3):375-387, 2004.

[7] A. N. Pour, B. R. Tabar, and A. Rahimzadeh. "A hybrid neural network and traditional approach for forecasting lumpy demand." International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2(4):1028-1034, 2008.

[8] K. Nikolopoulos, A. A. Syntetos, J. E. Boylan, F. Petropoulos, and V. Assimakopoulos. "An aggregate-disaggregate intermittent demand approach to forecasting: An empirical proposition and analysis." Journal of the Operational Research Society, pages 544-554, 2011.

[9] A. J. Smola and B. Scholkopf. "A tutorial on support vector regression." Technical Report - ESPRIT Working Group in Neural and Computational Learning II, 1998.

[10] M.Cilimkovic, "Neural Networks and Back Propagation Algorithm." Institute of Technology Blanchardstown.

[11] M. Riedmiller. "Rprop - description and implementations details." Technical Report University of Karlsruhe, 1994.

[12] R. Hyndman. "Another look at forecast accuracy metrics for intermittent demand." Foresight: the International Journal of Applied Forecasting, 4(4):43-46, 2006.