Archive of SID

The Development of a Hybrid Error Feedback Model for Sales Forecasting

Mehdi Farrokhbakht Foumani*
Department of Computer Engineering, Islamic Azad University, Fouman and Shaft Branch, Fouman, Iran
M.Farrokhbakht@fshiau.ac.ir
Sajad Moazami Goudarzi
Department of Computer Engineering, Islamic Azad University, Tehran North Branch, Tehran, Iran
SajadMoazami@yahoo.com

Received: 08/Oct/2020 Revised: 17/Apr/2021 Accepted: 11/May/2021

Abstract

Sales forecasting is one of the significant issues in the industrial and service sector which can lead to facilitated management decisions and reduce the lost values in case of being dealt with properly. Also sales forecasting is one of the complicated problems in analyzing time series and data mining due to the number of intervening parameters. Various models were presented on this issue and each one found acceptable results. However, developing the methods in this study is still considered by researchers. In this regard, the present study provided a hybrid model with error feedback for sales forecasting. In this study, forecasting was conducted using a supervised learning method. Then, the remaining values (model error) were specified and the error values were forecasted using another learning method. Finally, two trained models were combined together and consecutively used for sales forecasting. In other words, first the forecasting was conducted and then the error rate was determined by the second model. The total forecasting and model error indicated the final forecasting. The computational results obtained from numerical experiments indicated the superiority of the proposed hybrid method performance over the common models in the available literature and reduced the indicators related to forecasting error.

Keywords: Data mining; Machine learning theory; Supervised learning; Sales forecasting.

1- Introduction

Data processing and analysis has been recognized as an efficient and lucrative process for organizations over the recent years. Thus the data mining methods that discovers and extracts useful patterns from such large data sets to find hidden and worthy patterns for the decision-making as well as machine learning theory methods have turned out to be very useful[1]. Sales forecasting is known as one of the most important applications of these methods. Sales forecasting is the process of determining the future demand of customers which can be conducted in the short term or long term. In sales forecasting, it is determined how much sales the sales team or participation at a certain time period (weekly, monthly, quarterly, or annually) achieves. Managers use sales forecasting in all of their agencies to estimate the amount of conducted transactions by their whole team. In addition, they use the forecasting conducted by the sales team to forecast the total amount of sales in that part of the organization[2].

A variety of methods have already been proposed for sales forecasting. These methods have generally been developed using two general approaches[3]:

1- Approaches related to time series forecasting. In this approach, the existing models focus on the demands that contain seasonality patterns. In the absence of a regular seasonality pattern, time series forecasting methods will not be very successful.

2- Calendar information that influences demands and consequently the forecasting procedures using the approaches derived from machine learning theory. These approaches have been successful in some businesses such as retailers, airlines, car sales, etc., and are generally preferred over time series approaches.

Researchers believe that sales forecasting is indispensable from the following perspectives[4]:

A) Guiding the chain towards the point where the inventory level is optimized and the products needed to meet customer demands and needs are not in short supply.

B) Facilitation of capacity planning, warehouse placement, production planning, procurement and prediction of raw materials required in the production stage.

C) Facilitation of matters related to sales and operations planning

According to the literature, companies with accurate sales forecasts are characterized by smaller warehouse space (by 15 %), higher OTIF index (by 17%) and lower

www.SID.ir

Archive of SID

cash-to-cash cycle time (by 35%)compared to other companies[5].

Taking into account the benefits of implementing sales forecasting programs, the need to design algorithms for this purpose is highlighted. Thus, the current research proposal seeks to develop an algorithm for short-term sales forecasting.

Accordingly, achieving a sales forecasting method with acceptable accuracy requires to consider all the effective factors and is regarded as a complicated process. In this regard, this study uses the two-step error feedback algorithm to present a new method for sales forecasting where the forecasted error values were corrected by a forecasting algorithm and thus the model error was reduced.

2- Review of literature

Sales forecasting can be largely significant in the success and performance of organizations. Forecasting with accuracy may lead to the lost demand or increase the product durability. Sales forecasting is one of the complicated problems in analyzing time series and data mining due to the number of intervening parameters. The factors such as population, purchase behavior, customers' tastes, and competitors' performance affect the product sales. In addition, the factors such as holidays, climatic conditions, daily events in society, economic status, and some unpredictable factors affect the customer purchase [4]. Another challenge is originated from the fact that fashion products are mostly saturated at a relatively short time and replaced by another fashion product leading to the lack of historical data for sales [6].

Sohrabi et al.[7] in their study, attempted to develop accurate classification algorithms (e.g., decision tree, art neural networks, support machines, and logistic regression).

For sales forecasting, the statistical methods such as exponential smoothing, ARIMA, Box and Jenkins, Regression, and Holt-Winters models were used. However, some advanced models such as neural network, support vector regression, and other hybrid models of data mining were used in recent years [8].

As the introduction and brief review of the related studies mentioned, Sun et al.[9] used extreme learning machine (ELM) for clothing sales forecasting. Their study followed up the study by Zhu et al.[10] using extreme learning machine as a fast method and introducing it as appropriate for bulk data.

Wong and Guo [11] combined extreme learning machine to harmony search algorithm and indicated that such a combination improves the accuracy of extreme machine in sales forecasting in the retail industry. In the proposed method, the harmony search algorithm has the task of training the extreme machine. Parameter turning is conducted by using a heuristic method.

Yu et al. [12] used the support vector machine regression to forecast the number of newspaper daily sales for a newspaper and magazine publication company. In this study, the support vector machine linear model and its nonlinear model were used by combining the rapid basis kernel function where the rapid basis model has a higher accuracy. Di pillo et al.[13] in another study explained the use of support vector regression for sales forecasting and used this method in another case study. Table 1 indicates some areas where machine learning methods and data mining are used for sales forecasting.

Table 1: Different areas in sales forecasting

Researcher(s)	***year***	***Area of study***
Aramaki et al. [14]	2011	Epidemiology
Asur&Haberman[15]	2010	Mailbox sales profits
Bollen et al. [16]	2011	Supermarket sales
Choi & Varian [17]	2012	Ticket sales on a tour
Dhar& Chang [18]	2009	Online music sales

The models in sales forecasting are divided into three main groups based on the focus and objective of the developed model.

1. The models related to sales variable: This type of models makes predicting models by considering all sales factors such as calendar data such as calendar information, special discounts, the plans related to customers' club, etc. to the high number of variables. In this regard, the methods of feature selection are widely used and improve the model speed.

2. The models related to the sales dependency of the product group: In such models, the method used for forecasting each product considers the sales of other products and discovers the dependency between the sales of different products. In the forecasting phase, the future sales value for a specific product group is determined according to the sales of other products. For example, the sales of macaroni and tomato paste have an effect on each other due to high dependency. For this reason, the forecasted sale value is considered for the other product group while forecasting the sale of one of them and finally will be effective on the final value of the forecasting.

3. Stock-keeping unit models: The models related to this area forecast the sales value of each stock-keeping unit (a specific barcode) instead of forecasting the sales of the product group. The sales fluctuations of a specific barcode are more than a product group and thus the sales forecasting for that is more difficult than the forecasting of the product group. However, the efficiency of such models is more useful to the organization in terms of implementation.

www.SID.ir

Archive of SID

Here is the study of the sales of forecasting models.

2-1- The Models Related to Sales Variables

The models related to the variables affecting the sales of a product examine the effective factors in the product sales and perform sales forecasting by participating them in modeling[19]. The statistical models such as regression method are very useful in this regard. Since the methods related to this area usually involve many variables and lead to the reduction of algorithm processing speed, the feature selection techniques are required to remove the variables with less degree of significance or select the best subset of the current variables. Based on the study of John et al. [20], selecting a subset of the best variables is placed on in the class of Np-hard problems. For this reason, feature selection is conducted heuristically through two approaches of forward selection or backward elimination. These two approaches were developed in form of the algorithms such as genetic algorithm and simulated annealing[21]. Using such algorithms can converge to optimal subsets at a short time.

2-2- The Models Related to The Sales Dependency of The Product Group

Since buying some products may depend on some other products, the models related to the dependency of groups can be used as the efficient methods. The issue of demand dependency of different products was created in economics where the supplement and alternative products affect each other. Thus, the information related to changing one of them can be used in forecasting the demand of other products. In a product group, different products with different weights and packaging play the role of alternative products[22], [23]. The findings of researchers indicated that a large part of purchase behavior changes at the discount time is due to the presence of dependency in buying different products [24].

In most cases, buying a product available in the store or a product with discount, leads to the purchase of other products.

For example, the presence of discount on a specific kind of cake will lead to the increase of purchase on the complete package of different cakes. This type of dependency is known as intra group. As another example, the presence of discount on a type of macaroni may increase the sales of tomato paste. This is an inter group dependency. The effect of intra group and inter group dependencies leads to the increase of selling the products without discount.

Considering the inter group and intra group dependencies will lead to the increase of complexity in the final model. For example, in the VAR method, the number of variables increases as a quadratic function leading to the final complexity of the model.

In such cases, even simple linear models cannot be used due to the multiple variables. Figure 1 indicates the multiplicity of dependency among different product groups [25].

2-3- The Forecasting Models of The Stock-keeping Unit

The models related to the stock-keeping unit are usually referred to the single variable models which forecast the future sales in the short term using time series analysis [26].

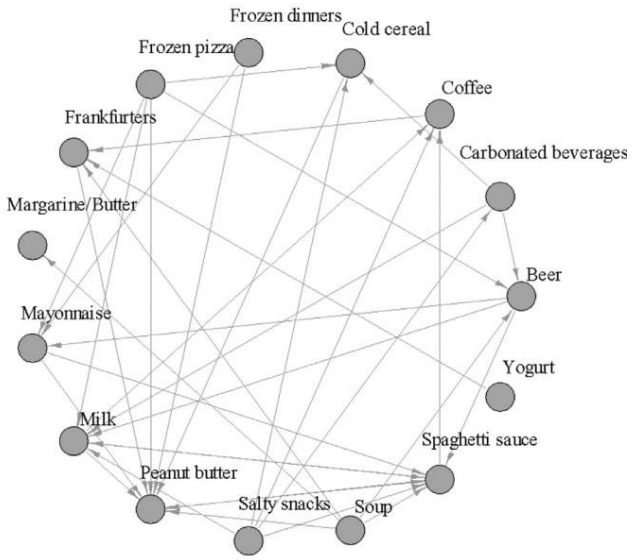


Fig. 1 The inter group dependency in making the forecasting models

Such methods do not consider the external variables such as the changes of price or discount plans. Goorali et al. [27] stated that the methods related to time series for the periods without discount work better than the other methods and show less error. However, for the periods where the policy of discount is considered, the methods such as regression entering external factors in the model will have superiority. Baeke et al. [28] presented a model where judgement and opinion of managers are combined to statistical models and finally the future demand value is forecasted.

Two approaches were presented for considering the managers in the used statistical model. The restrictive model where the forecasting of managers is presented as a range and used as the upper and lower limits of future forecasting in the model as well as the integrative model where the forecasting of managers is entered into the model as a new variable.

In this regard, Li and Lim [29] introduced a decomposition- aggregation algorithm and used it as the sales forecasting of clothing in Singapore. Their proposed

www.SID.ir

Archive of SID

method used the analysis of previous process of each item and the total behavior of sales in a store to conduct the forecasting. The comparison to previous methods confirmed the superiority of decomposition- aggregation algorithm.
The other related methods can be searched in the studies by Kourentzes and Petropoulos [30] and Babai et al. [31].

3- The Proposed Model

Using the Sine models in modeling the time series was introduced by Simons [32] and its general relation is as Equation (1).

$$\{Y_t\} = a_0 + \sum_{j=1}^{n} \left(a_j \cos\left(\lambda_j t\right) + b_j \sin\left(\lambda_j t\right) \right) \quad (1)$$

Where a_j and λ_j are the parameters related to the model and should be adjusted accurately for the access to the acceptable model. Using this model can be an alternative to the time series classic decomposition model.

Ord and Fildes [26] use the Sine model and multiple regression model to develop an algorithm for forecasting the time series being associated with the improved accuracy of process forecasting and seasonal model of time series. However, the main problem of the proposed method is the lack of paying attention to the external variables which may affect time series at specific times. For example, the presence of lunar occasions or official holidays in case of product sales are among these variables. Accordingly, the present study combined the multiple regression method (considering the external variables based on Table 2) to a Sine feedback model to forecast the sales. In other words, the proposed sales forecasting model is a hybrid algorithm correcting the conducted forecasting based on the feedback of error values. In this regard, after the initial training, the error values are determined and entered into another algorithm as a time series. Then, the second algorithm is trained based on the error values. For forecasting the future sales, first the time series related to the sales on previous days are entered into the model and an initial forecasting is obtained. Then, the initial forecasting is entered into the second algorithm and the final correction is made based on the correction factor obtained in the training group. Figure 2 shows a general view of the proposed hybrid algorithm.

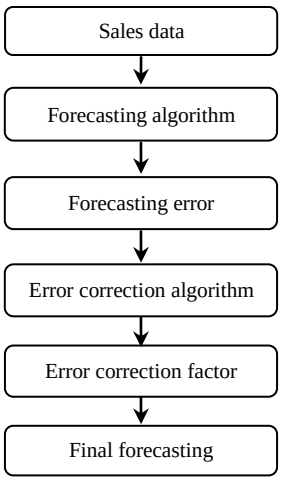


Fig. 2 The hybrid error feedback algorithm

3-1- Problem Data

The data used for constructing the proposed hybrid model and measuring its performance were related to the sale of 1850 items in one of the chain stores of Iran collected during 400 days. Since the information about the inventory is not available and the zero inventory may result in the non-sale of the product in relation to zero sale in the store, the proposed model can be only implemented for the products with at least 300 days of positive sale (At least 75% of the total days). The variables used for determining the amount of sale and extracting the related pattern were only of calendar variable type. In this regard, Table 2 indicates the variables used for constructing the sale forecasting model.

Table 2:The variables used in problem modeling.

Variable name	***Variable type***	***Description***
Season	Categorical	It represents the season of the year.
Month	Categorical	It represents the month of the year.
Week	Categorical	It indicates the number of the week including the numbers from 1 to 4.
Day	Categorical	The numbers 1 to 31 indicate the date of the day. Since the sale increase at the end of the month due to discounts, this variable can control some parts of the sale fluctuations.
Off day	Binary	Zero-one variable indicates if a particular day is off or not.

www.SID.ir

Variables in the present study can be divided into two groups: independent variables and dependent variables. According to the literature on sales forecasting, the independent variables that can affect sales are as follows:
1- Year: This variable can be used to discover the pattern in the sales trends.
2- Season: Some seasonal patterns can be extracted based on the level of sales in different seasons of the year.
3- Month: Since product sales may vary from one month to another, this variable will also be taken into account in the model.
4- Week days: Product Sales are more dramatic on busy shopping days (usually weekends) than week days. Thus, the number of week days will also be taken into account as a variable.
5- Holiday: Variable 0 and 1 indicate whether or not a particular day is holiday.
The dependent variable in this study is the quantity of sales (number of products sold)on a particular day.

3-2- Outlier Detection

Among the above-mentioned data, sometimes the sale values have an emotional growth which should be detected and eliminated at the corresponding time series. For example, some products such as milk face an emotional growth during the days when the air pollution increases due to inversion. Since the weather condition is not available because of the current limitations and cannot be entered into the model as a variable, the presence of the data related to emotional growth will lead to an increase in the model error. As another example, the products such as salt or cheap biscuits are sometimes offered to customers as free which increases the sale and the effect of free supply cannot be included in the forecasting model. For this reason, the outlier should be detected and eliminated at the time series. In this regard, the methods of outlier detection has been checked out and the most appropriate method for use in the proposed model is described.
In general, outliers to data that are much smaller or larger than other members of a given dataset. Outliers can affect the results and in fact impair their accuracy. Therefore, such data need to be analyzed and deleted from the dataset(if any)[33].
Depending on the number of variables, outlier detection techniques are categorized in two groups: univariate and multivariate detection techniques. In former case, only one variable is taken into account and a value is assigned to them based on their placement in the outlier region (univariate techniques) in the latter case, however, the variables in the data set are taken into account and addressed simultaneously. (multivariate techniques).
A method called Hempel technique is used for filtering the wave in the topics related to signal analysis. Hempel technique is appropriate for the cases where the value of a variable is collected over time and time intervals can be continuous (momentary) or discrete. Since the time series related to the sale of products is a wave with discrete time intervals (a quantity for sale every day). Thus, the present study will use Hempel technique.
For more explanation, consider a time series as Equation (2).

$$\omega_k^K = \{ x_{k-K}, ..., x_k, ..., x_{k+K} \} \quad (2)$$

The mean value of this time series is indicated as m_k and defined as follows:

$$m_k = median \{ x_{k-K}, ..., x_k, ..., x_{k+K} \} \quad (3)$$

Accordingly, the Hempel detector variable $y_k \in \{ x_k, m_k \}$ will be a binary variable for which m_k value means the k -th in the time series introduced time series in Equation (3) is an outlier. Thus, its median value will be replaced, otherwise the above-mentioned data cannot be considered as outlier. The detector variable $y_k \in \{ x_k, m_k \}$ is defined as Equation (4).

$$y_k = \begin{cases} x_k & |x_k - m_k| \le t.s_k \\ m_k & |x_k - m_k| > t.s_k \end{cases} \quad (4)$$

Where the estimator is the Hampel median coefficient following Equation 5.

$$s_k = 1.4826 \times \underset{j \in [-k,k]}{median} \{ |x_{k-j} - m_k| \} \quad (5)$$

In addition, parameter t is known as the decision boundary and is typically considered in such a way that the following equation is true:

$$t \ge \frac{\max_k |x_k - m_k|}{\min_k s_k} \quad (6)$$

3-3- Hybrid error Feedback Algorithm

As mentioned in the current section, the proposed error feedback model includes two algorithms one of which is used for the initial sales forecasting and the other one for correcting the error values. In this procedure, first a clustering algorithm is used on the time series related to the sales of used items and a forecasting is obtained. An initial forecasting is obtained using the created model and the forecasted error values are stored. For more explanation, assume that there is a time series as $\Xi = \{ x_1, x_2, ..., x_n \}$ for one of the items in the store where x_i represents the sales amount on the i-th day. By considering the above-mentioned time series, the first clustering algorithm considers the calendar data introduced in Table 3 as the input and the sales amount as the output

www.SID.ir

Archive of SID

and then is trained. After training, an initial sales forecasting is obtained using the same inputs. The initial forecasting value for sales is named $\hat{y}_i$. Since the model may have errors, the error value is calculated using the initial forecasting through Equation (7).

$$e_i = \hat{y}_i - x_i \quad (7)$$

Where e_i represents the error related to the model.

In the second group, the inputs related to the calendar data based on table 2. (similar to the first algorithm) are entered into the second algorithm (error corrective algorithm) and then the values related to error are considered as the output.

In fact, the second algorithm attempts to determine the error related to the daily sales. At the final step, the calendar data related to the future days are entered into the first algorithm. Then, an initial sales forecasting value $\hat{y}_i$ with the calendar data are entered into the second algorithm and the forecasting error value is specified $\hat{e}_i$. Finally, the sales forecasting value is calculated using Equation (8).

$$y_i = \hat{y}_i + \hat{e}_i \quad (8)$$

The flowchart related to the hybrid error feedback algorithm is shown as Figure 3.

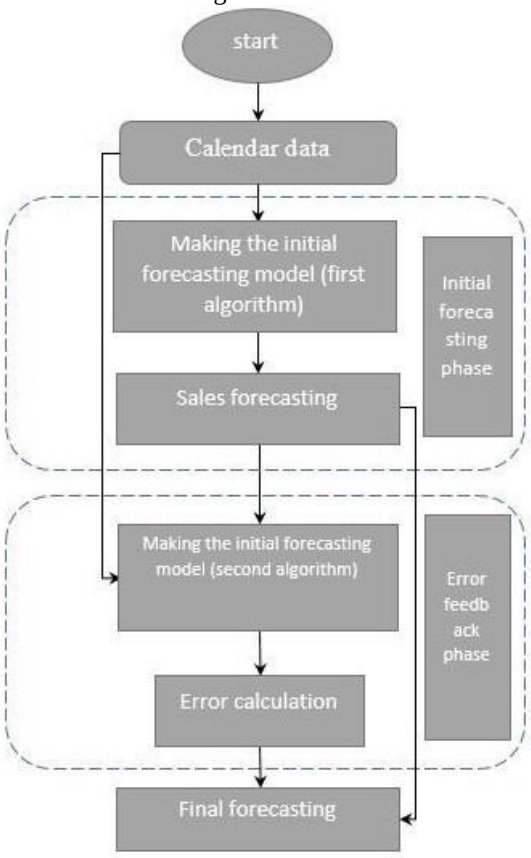


Fig. 3 The flowchart related to the hybrid error feedback algorithm

In the error feedback phase, the model error is estimated using a general relation according to Equation (9) having fewer parameters than the Simon's model [32].

$$\hat{e}_i = a_1 \sin\left(a_2 i + a_3\right) + a_4 \quad (9)$$

In other words, the values related to the error are calculated after making the initial model and then considered as a dependent variable for Equation 3.10. Then, the best values for $\{a_1, a_2, a_3, a_4\}$ are calculated using model (10) called least square method.

$$\hat{e}_i = \sum_{i=1}^{n} \left(a_1 \sin\left(a_2 i + a_3\right) + a_4 - e_i\right)^2 \quad (10)$$

Where n represents the number of train data and i indicates the number of data.

By specifying the values $\{a_1, a_2, a_3, a_4\}$, Equation (9) is used and error estimate $\hat{e}_i$ is calculated. Finally, the final forecast is obtained by Equation $\hat{y}_f = \hat{y}_i + \hat{e}_i$.

4- Computational Results

All computations related to the proposed model were conducted in MATLAB R2017b software using a system with Core i3 processor and internal memory of 4Gb. Since the direct forecasting for each product leads to the creation of a lot of chaos in the models and cannot have enough accuracy, the proposed model is matched on the sales data of the product group. In this regard, the sales of each product in one group is added to each other and finally the sales of the product group are examined as the studied variable (e.g. the low fact milk group includes different products such as 500 CC and 1000 CC from different companies). Accordingly, 371 product groups are extracted from 1850 products in the dataset. In order to measure the performance of the proposed model and compare it to the previous models in the literature, the following parameters are used:

Root mean square of errors: If the real demand value on day t is equal to y_t and its forecasting value is equal to $\hat{y}_t$, the value of this index for n periods will be calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{n}} \quad (11)$$

Mean Absolute Errors:

$$MAE = \frac{\sum_{i=1}^{n} \left|y_i - \hat{y}_i\right|}{n} \quad (12)$$

Mean Absolute scaled Errors:

www.SID.ir

Archive of SID

$$MASE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{\frac{1}{n-1}\sum_{k=2}^{n-1}|y_k - y_{k-1}|}\right) \quad (13)$$

The variables related to the dataset include the calendar data such as year, month, day, week, and holidays. Accordingly, the sales amount is considered as the dependent variable and the forecast model is matched to it. In addition, other data of product are considered for reaching the sales forecasting of some. In this regard, the sales of similar products are classified in a variable called product group. For example, the sales of cola, beer, etc. will be classified under the name "carbonated beverage". The sales of mineral water and other non-carbonated beverages is classified under the name "non-carbonated beverages" and forecasting is conducted for the level-4 sales of the product group. Finally, the share of each product from the total sales of product group is determined and accordingly, the final sales will be specified in terms of products.

In order to achieve an accurate dataset, the preprocessing should be conducted and some outliers should be eliminated from the studied dataset. For this reason, the Hampelfilteris used where the time series related to the sales is studied and the related outliers are eliminated. In this regard, Table 3 shows the data in relation to the number of eliminated data in each product group.

In addition, Figure 4 shows the frequency of outliers. This figure indicates that the maximum repetition of outliers is between 60 to 80 outliers being repeated for 113 product groups.

Table 3: The summary of data related to outliers

Title	***value***	**Product group**
the maximum number of outliers	206	Hand soap powder
the minimum number of outliers	2	Shaving foam
the average number of outliers in each group	73.29	-
The total number of outliers	19571	-

The proposed hybrid model includes two phases. In the first phase, the initial forecasting is conducted by using a forecasting model and then the errors related to the forecasting are determined, in the second phase, the values related to forecasting error are determined using an error corrective algorithm. Finally, the final output is determined by adding the initial forecasting and predetermined error.

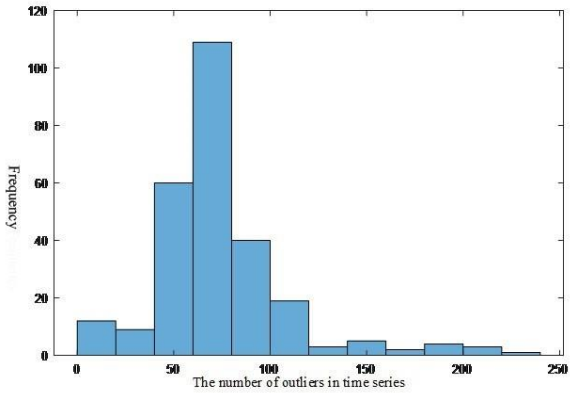


Fig. 4 The histogram related to the frequency of outliers

In this regard, for example the diagram related to the daily demand of minced meat is shown in Figure 5.

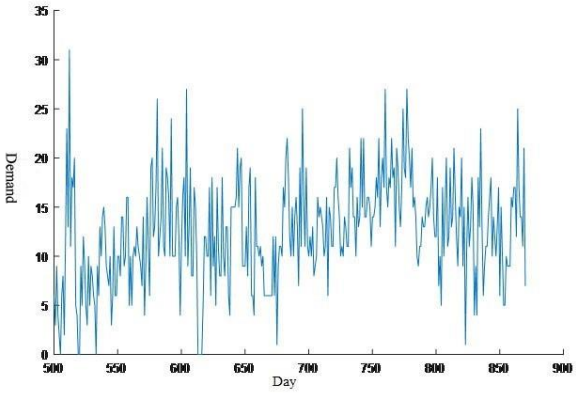


Fig. 5 The demand related to minced meat product group

First, the linear regression model is calculated using the calendar data in Table 2 based on the train data and then is used for test data. The output of the above-mentioned regression model will be the initial forecasting value $\hat{y}_i$ .Figure 6 shows the related algorithm.

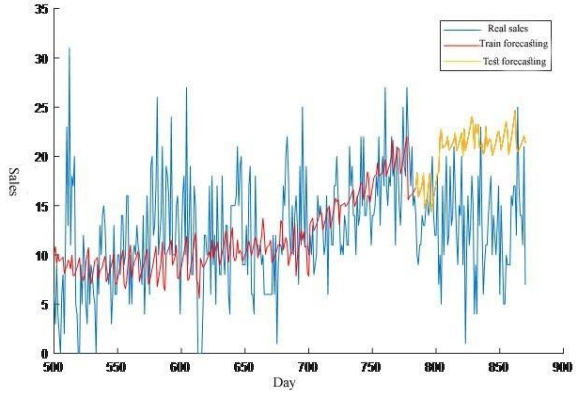


Fig. 6 The output related to the regression model as the initial forecast of minced meat product group

www.SID.ir

Archive of SID

Accordingly, it is clear that the regression model is ascendingly trained due to the presence of an ascending trend on the sales data and caused some errors while forecasting the days related to the test set.
The related indices for this case are RMSE=9.19, MAE=7.96, and MASE=1.69.
Thus, due to the presence of error in discovering the models related to the time series of studied dataset, the error in the forecasting should be detected and simulated for correction by the Sine model introduced in the third chapter. After performing the simulation and training of the Sine model, Figure 7 is considerable.

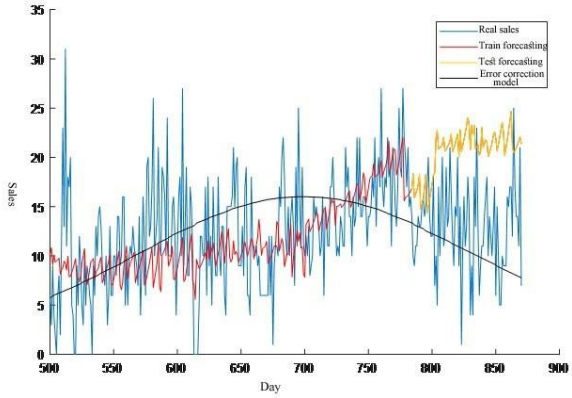


Fig. 7 The output related to the corrective model of error in minced meat product group

Obviously, the model shown in Figure 7 has the ability to discover the sales reduction than the trend related to the regression model.

Accordingly, the final value can be determined based on the initial forecast and output of the Sine model. Figure 8 shows the final forecasting.
It is graphically obvious that the final forecasting value is superior to the regression model.
In this relation, the correlation indicators are as RMSE=4.63, MAE=3.62, and MASE= 0.77.

Similar calculations are also conducted for other product groups and finally the obtained values are compared to the indicators related to the common regression model in the literature.
Accordingly, the results related to the calculations from the perspective of three indicators mentioned above are as Table 4.

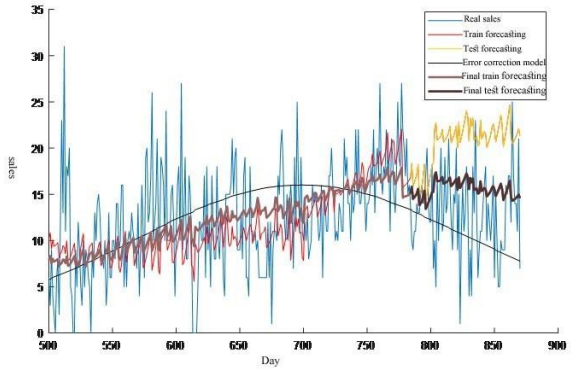


Fig. 8 The final forecasting related to the sales of minced meat product group

Based on Figure 9 and Table 4, it is clear that the proposed method in leads to the improvement of 13 product groups. The average value of improvement observed on all product groups is equal to 4.13 in terms of MAE.

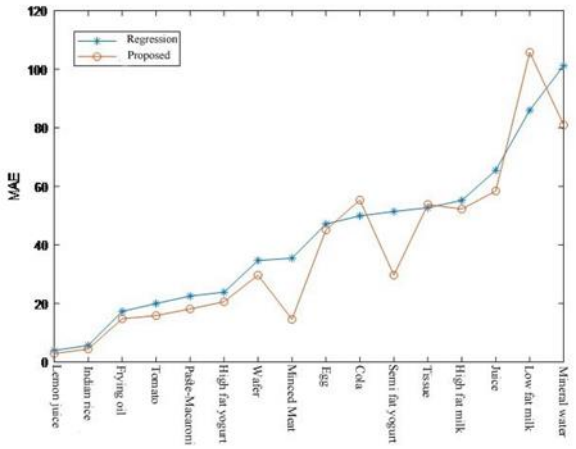


Fig. 9 Comparing the MAE performance indicator of the regression model and the proposed model

www.SID.ir

Table 4. The obtained results for the indicators of regression model performance, Ord and Fildes' model, and the proposed model

Product group	*Regression model*			*Ord and Fildes model (2003)*			*The proposed method*		
	MAE	RMSE	MASE	MAE	RMSE	MASE	MAE	RMSE	MASE
Lemon juice	3.99	4.85	1.06	3.03	4.29	0.83	2.94	3.89	0.78
Indian rice	5.71	8.10	1.17	5.35	7.17	0.96	4.48	7.01	0.92
Frying oil	17.32	21.46	1.01	16.36	20.78	1.02	14.81	19.53	0.86
Tomato paste	20.02	24.65	1.06	16.71	23.59	0.99	15.90	20.44	0.84
macaroni	22.56	26.68	0.96	14.92	22.29	0.73	18.16	22.53	0.77
High fat yogurt	23.89	27.75	1.34	23.78	27.85	1.22	20.63	24.36	1.16
Wafer	34.73	44.86	1.17	31.48	38.86	1.16	29.72	38.77	1.00
Minced meat	35.49	39.90	2.89	13.74	14.16	0.97	14.63	17.28	1.19
egg	47.18	52.83	2.32	50.34	54.80	2.56	45.18	50.66	2.22
soda	49.96	61.74	0.88	63.54	73.23	1.06	55.38	65.65	0.97
Half-fat milk	51.45	54.46	3.73	26.92	31.13	1.84	29.71	33.87	2.16
Tissue	52.73	69.82	0.71	56.28	78.77	0.80	53.85	73.27	0.72
High fat milk	55.24	76.59	0.91	61.63	81.56	0.99	52.29	75.25	0.86
Juice	65.51	88.79	0.83	65.53	84.39	0.80	58.44	84.25	0.74
Low fat milk	86.06	102.91	1.44	118.40	128.92	2.01	105.81	123.67	1.78
Mineral water	101.26	142.05	0.99	85.37	121.76	0.85	81.00	118.97	0.79
Average	42.07	52.97	1.40	40.84	50.85	1.17	37.68	48.72	1.11

5- Conclusion

The present study provided a new model by combining the regression forecasting algorithm and error feedback model

to forecast the sales of products in the stores where the initial forecasting was conducted using the regression model and then the remaining values (error) in the forecasting were corrected using an error feedback model to correct the initial forecasting values. In order to review the activities conducted in this study and based on the computational results, the proposed hybrid method had superiority over the common regression model in the available literature and reduced the indicators related to forecasting error. The improvement rate observed on 16 product groups was explained in detail and it was specified that the proposed method created 4.13 improvement units in MAE index. Other performance indicators of the model also indicated the improvement.

In order to improve the results and develop the introduced model, the following issues can be discussed in this chapter.

- Using the Gaussian kernel function due to high flexibility in modeling the complicated structures can lead to the improvement of the model performance. In this regard, the computation time may increase slightly. However, the increase of computational time can be ignored in case of improving the accuracy.
- Using other forecasting methods instead of regression method to improve the initial forecasting which can improve the final forecasting and increase the final accuracy.
- Using a nested approach which means combining the clustering and classification and using a forecasting model inside each cluster, by doing this, each classifier is used on the similar data and can increase the accuracy.

References

[1] B. Sohrabi, I. RaeesiVanani, N. Nikaein and S. Kakavand, "A predictive analytics of physiciansprescription and pharmacies sales correlation using data mining", International Journal of Pharmaceutical and Healthcare Marketing, vol.13, No.3, pp. 346-363, 2019.

[2] Y. Kaneko and K. Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales", In Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference, 2016, December, pp. 531-537.

[3] M. Bohanec, M.K. Borštnar and M. Robnik-Šikonja, "Explaining machine learning models in sales predictions", Expert Systems with Applications, vol. 71, pp. 416-428, 2017.

[4] S. Thomassey, "Sales forecasts in clothing industry: The key success factor of the supply chain management", International Journal of Production Economics, Vol. 128, No. 2, pp. 470-483, 2010.

[5] A. Demiriz, "Demand Forecasting based on Pairwise Item Associations",Procedia Computer Science, vol. 36, pp. 261-268, 2014.

[6] S. Thomassey,"Sales forecasting in apparel and fashion industry: A review", Intelligent fashion forecasting systems:

www.SID.ir

Archive of SID

Models and applications,Springer, Berlin, Heidelberg, pp. 9-27, 2014.

[7] B. Sohrabi, I. RaeesiVanani, A. Gooyavar and N. Naderi, " Predicting the Readmission of Heart Failure Patients through Data Analytics ", Journal of Information & Knowledge Management, Vol.18, No.1, pp.1950012-1, 1950012-20, 2019.

[8] N. Liu, S. Ren, T.M. Chio, C.L.Hui,andS.F.Ng,"Sales forecasting for fashion retailing service industry: a review", Mathematical Problems in Engineering, Vol.20, No. 2, pp. 22-29, 2013.

[9] Z.L. Sun, T.M. Choi, K.F. Au, and Y. Yu,"Sales forecasting using extreme learning machine with applications in fashion retailing", Decision Support Systems, Vol.46, No. 1, pp. 411-419, 2008.

[10] Q.Y.Zhu, A.K.Qin, P.N.Suganthan, and G.B.Huang, "Evolutionary extreme learning machine", Pattern recognition, Vol.38, No. 10, pp.1759-1763, 2005.

[11] W. K. Wong, and Z. X. Guo,"A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm", International Journal of Production Economics, Vol.128, No. 2, pp.614-624, 2010.

[12] X. Yu, Z. Qi, and Y. Zhao,"Support vector regression for newspaper/magazine sales forecasting", Procedia Computer Science,vol. 17, pp. 1055-1062, 2013.

[13] G. Di Pillo, V.Latorre, S. Lucidi, and E. Procacci,"An application of support vector machines to sales forecasting under promotions", 4OR, Vol.14, No. 3, pp. 309-325, 2016.

[14] E. Aramaki, S. Maskawa, and M. Morita,"Twitter catches the flu: detecting influenza epidemics using Twitter", In Proceedings of the conference on empirical methods in natural language processing, Association for Computational Linguistics, N. Eight Street, Stroudsburg, PA, 18360 United States, 2011, pp. 1568-1576.

[15] S. Asur and B.A. Huberman,"Predicting the future with social media", In 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010, IEEE, Toronto, Canada,Vol.1, pp. 492-499.

[16] J. Bollen, H. Mao, and X. Zeng,"Twitter mood predicts the stock market", Journal of computational science, Vol. 2, No. 2, pp. 1-8, 2011.

[17] H. Choi, and H. Varian,"Predicting the present with Google Trends", Economic Record, Vol.88, No. S1, pp.2-9, 2012.

[18] V. Dhar, and E.A. Chang,"Does chatter matter? The impact of user-generated content on music sales", Journal of Interactive Marketing, Vol.23, No. 4, pp.300-307, 2009.

[19] D. L. Donoho,"High-dimensional data analysis: The curses and blessings of dimensionality", AMS math challenges lecture, pp. 1-32, 2000.

[20] G. H. John, R. Kohavi, and k. Pfleger,"Irrelevant features and the subset selection problem", In Machine Learning Proceedings, Proceedings of the Eleventh International Conference,1994,Elsevier, Rutgers University, New Brunswick, NJ, pp. 121-129.

[21] R. Meiri, and J. Zahavi,"Using simulated annealing to optimize the feature selection problem in marketing applications", European Journal of Operational Research, Vol.171. No. 3, pp.842-858, 2006.

[22] V. Kumar, and R.P. Leone,"Measuring the effect of retail store promotions on brand and store substitution", Journal of Marketing Research, Vol.25, No. 2, pp. 178-185,1988.

[23] R.G. Walters,"Assessing the impact of retail price promotions on product substitution, complementary purchase, andinterstore sales displacement", The Journal of Marketing, Vol.55, No. 2, pp. 17-28, 1991.

[24] H.J. Heerde, S. Gupta, and D.R. Wittink,"Is 75% of the sales promotion bump due to brand switching? No, only 33% is", Journal of Marketing Research, Vol.40, No. 4, pp.481–491, 2003.

[25] S. Ma, R. Fildes, and T. Huang. "Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra-and inter-category promotional information", European Journal of Operational Research, Vol.249, No. 1, pp.245-2572, 2016.

[26] K. Ord, and R. Fildes,Principles of business forecasting, Mason, 1st ed, OH: South Western Cengage Learning, 2013.

[27] Ö. GürAli, S. Sayın, T. Van Woensel, and J. Fransoo,"SKU demand forecasting in the presence of promotions", Expert Systems with Applications, Vol.36, No. 10, pp.12340-12348, 2009.

[28] P. Baecke, S. De Baets, and K. Vanderheyden,"Investigating the added value of integrating human judgement into statistical demand forecasting systems", International Journal of Production Economics, Vol.191, pp.85-96, 2017.

[29] C. Li, and A. Lim,"greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing", European Journal of Operational Research, Vol.269, No. 3, pp.860-869, 2018.

[30] N. Kourentzes, and F. Petropoulos,"Forecasting with multivariate temporal aggregation: The case of promotional modelling", International Journal of Production Economics, Vol.181, part A,pp.145-153, 2016.

[31] M.Z. Babai, Y. Dallery, S. Boubaker, andR.Kalai,"A new method to forecast intermittent demand in the presence of inventory obsolescence", International Journal of Production Economics, Vol.209, pp.30-41, 2019.

[32] L. F. Simmons,"Time-series decomposition using the sinusoidal model", International Journal of Forecasting, Vol. 6, No. 4, pp. 485-495, 1990.

[33] L.I. Zheng-Feng, X.U. Guang-Jin, W.A. Jia-Jun, D.U. Guo-Rong, C.A. Wen-Sheng and SH. Xue-Guang, "Outlier Detection for Multivariate Calibration in Near Infrared Spectroscopic Analysis by Model Diagnostics",Chinese Journal of Analytical Chemistry, vol.44, No.2, pp.305-309, 2016 Feb.

Mehdi Farrokhbakht Foumani is an Assistant Professor of Computer Engineering at Islamic Azad University, Fouman and Shaft branch, Iran. He received his Ph.D. in Computer Engineering from Saint Petersburg Electrotechnical University (LETI) in 2013. His research interests include Artificial Intelligence, Machine Learning and Data Mining.

Sajad Moazami Goodarzi Received his M.Sc. in Computer Engineering from the Faculty of Electrical and Computer Engineering at Islamic Azad University, Tehran North Branch in 2018. His research interests include Data Mining and Machine Learning.

www.SID.ir