Steel Consumption Forecasting Using Nonlinear Pattern Recognition Model Based on Self-Organizing Maps

S. Torbat [1*] , M. Khashei [2], M. Bijari [3]

Department of Industrial and Systems Engineering, Isfahan University of Technology (IUT), Isfahan, Iran

Abstract

Steel consumption is a critical factor affecting pricing decisions and a key element to achieve sustainable industrial development. Forecasting future trends of steel consumption based on analysis of nonlinear patterns using artificial intelligence (AI) techniques is the main purpose of this paper. Because there are several features affecting target variable which make the analysis of relationships between variables difficult, so it is needed forintelligence tools in order to recognize the nonlinear patterns and select the most effective features as final input variables. Generally, in the real world problems, a nonlinear pattern recognition and feature clustering usually involve high-dimensional features that make the clustering problem complex. In fact without feature subset selection and dimensionality reduction, both training accuracy and generalization capability of forecasting models will be significantly reduced. In this paper, artificial neural networks are applied in order to profit from unique advantages of both forecasting and clustering power of the artificial neural networks to create an efficient and accurate model in high-dimensional situations. This way, at first, self-organizing map (SOM) is applied as an unsupervised clustering technique to detect nonlinear patterns between explanatory variables and to determine final effective input variables. Then the best multilayer perceptron is designed by these variables in order to forecast future trends of steel consumption. The empirical results of Iran's steel consumption forecasting confirm that the proposed model exhibits effectively improved forecasting accuracy in comparison to traditional feature selection methods such as forward and backward.

Keywords: Self-organizing maps (SOMs); Nonlinear Pattern recognition; Feature clustering; Dimensionality reduction; Artificial neural networks (ANNs); Steel consumption forecasting.

1. Introduction

In today's world, fast and accurate decisions is one of the effective factors that increase organizations efficiency. Thus, management is a key element to make effective decisions and achieve organizational goals in each industry. In fact, to make smarter decisions provide an appropriate framework to improve upon the staff's performance, and steel industry is the same. The steel industry is among the strategic industries and it plays an important role in the persistent economic growth of each country, in a way that, managers usually seek to find logical approaches in order to improve the quality of their economic decisions. In this regard, awareness of the current and future conditions of this industry and identification of the factors affecting them are important to economic analysis.

The main purpose of the economic analysis is to identify the effective factors in order to provide accurate forecasts of them.

The production and consumption are critical factors affecting pricing decisions. Therefore, fast and accurate detection of variables affecting them can be used to help improve forecasting accuracy. In recent years, the intelligent methods are crucial for forecasting purpose in financial markets as well as for improved decisions and investments.

Artificial Neural Networks (ANNs) is used in financial forecasting as an effective technique to yield the more accurate results.

* *Corresponding author*
Tell: +98 31 3911482
Email: s.torbat@in.iut.ac.ir
Address: Department of Industrial and Systems Engineering, Isfahan University of Technology, Isfahan, Iran
1. M.Sc. Student
2. Assistant Professor
3. Professor

ANNs are considered as logical approaches in order to model complexity of real world problems and exclude inherent limitations of traditional statistical techniques. One of the most important advantages of artificial neural networks is their flexible nonlinear modeling capability. With ANNs, there is no need to specify a particular model form. Rather, the model is adaptively formed base on the features presented in the data. This data-driven approach is suitable for many empirical data sets where no theoretical guidance is available to suggest an appropriate data generating process. Despite the advantages cited for theme, ANNs have weaknesses that one of the most important of them is the number of input variables. Theoretically, there is no defined limitation for number of input variables in ANNs. However, the performance of ANNs will be generally decreased, if the number of input variables is too high. This is a critical disadvantage to forecasting purposes, especially for causal forecasting in complex environments.

In the literature review, several methods were provided in order to analyze the relationships between features and identification the most effective features as the final input variables. These methods are generally divided into two categories: statistical and intelligent methods. Traditional statistical methods have only the ability to analyze linear structures, while most structures in the real world systems, particularly in the financial systems, are complex and nonlinear. Self-organizing map (SOM) is a specific category of the artificial neural networks which is proposed by Kohonen [1, 2]. These types of networks are one of the most popular unsupervised learning techniques, which they often use for analyzing complex environments. Therefore, SOM is probably the best-known clustering method. There are several investigations in the literature that they used SOM for clustering and feature selection tasks. In this regard, Yang et al [3]. proposed a novel constrained self-organizing map in order to combine semi-supervised clustering methods for enhancing the quality of complex decisions in intelligent decision support systems.

Shieh and Liao [4] proposed a new approach for clustering using self-organizing maps which was focused on developing a hybrid data clustering and visualization strategy that the inter-neuron distances are visible using a coloring scheme and the probability of each neuron is measurable. Chen et al. [5] used self-organizing map to analyze and visualize the bankruptcy trajectory of companies over several years. Hajjar and Hamdan [6] presented a self-organizing map for interval-valued data based on an adaptive Mahalanobis distance in order to do clustering of the interval data with topology preservation. Ghasemzadeh and Karami [7] proposed a novel SOM-based algorithm that can automatically cluster discrete groups of the data using an unsupervised method.

Liu and Ban [8] proposed a new clustering algorithm that detects clusters by learning data distribution of each cluster. The proposed model is able to generate a dynamic two-dimensional topological graph, which is used to explore both partitional information detailed data relationship in each cluster. This method is able to work incrementally and detect arbitrary-shaped clusters. Xu et al. [9] proposed a self-organizing map based on hashing framework (SOMH) that can keep similarity relationship, and also it can preserve topology of data. Qiu et al. [10] proposed a multi-stage design space reduction and meta-modeling optimization methodology based on self-organizing maps and fuzzy clustering. Haga et al. [11] applied self-organizing map as a clustering approach that rely on local linear regression-based estimation models. Wang and Wu [12] adopted manifold learning algorithm to feature subset selection and employed the kernel-based fuzzy self-organizing map (KFSOM) to compose base classifiers for business failure prediction.

In this paper, self-organizing map is applied as one of the most accurate unsupervised clustering methods in order to feature subset selection based on nonlinear pattern recognition to improve accuracy of forecasts. In this way, at first, primary features are clustered using SOM, in a way that, the variables in each cluster have the highest correlation with each other and the lowest correlation with the other clusters. Next, in each cluster a variable is selected as the final effective input variables, which has the highest correlation with the target variable. In the end, an ANN model is fitted to forecast steel consumption and its performance is compared with classical linear feature selection techniques such as forward and backward methods. The empirical results indicate that the self-organizing map is an effective method to feature subset selection and improve forecasting accuracy.

The rest of the paper is organized as follows: The self-organizing map (SOM) is reviewed in section 2. Basic concepts of artificial neural networks are explained in section 3. In section 4, the proposed feature selection model is applied to steel consumption forecasting and its performance is compared with classical linear feature selection techniques. In the end, comes a section on conclusions.

2. Self-Organizing Maps

The self-organizing map (SOM) [1, 2] is one of the most widely used unsupervised artificial neural networks which has drawn especial attention in recent years. The basic structure of SOM has two layers, input layer and output layer. The input layer neurons are responsible for transmitting data to the network and generally, the number of input neurons is equal to the dimension of input vectors. The output layer, so called the competition layer, consists of a set of neurons that are located in a flat plan. The outputs of the network are created by the output layer neurons based on the neighborhood relations. The number of neurons in the output layers depends on the under study problem and is determined by the user. The structure of SOM (4*5) is shown graphically in Fig. 1.

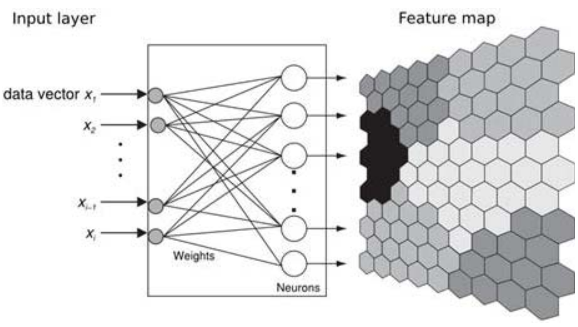

*Fig. 1. Structure of a self-organizing map (4*5).*

As mentioned previously, the self-organizing map is an unsupervised learning algorithm, which is essentially characterized by first-order differential equations. The SOM algorithm is based on the selection of winning neuron and the movement of this neuron and some of its neighbors toward the input data. Learning algorithm of SOM can be summarized in the following steps:

Step I: In this step, the weight of each neuron is $X = (x_1, x_2, ..., x_d)$ considered as a random number. Then, an input vector is presented to the networks at each training iteration.

Step II: In the second step of the learning algorithm, winning neuron is determined based on similarity criterion. There are different criteria, which can be applied in the learning process of the self-organizing map, but Euclidean distance is the most widely used similarity criterion. Euclidean distance between input vector$\left(X = (x_1, x_2, ..., x_d) \right)$ and all the weight vectors are calculated as follows:

$$\|X - W\| = \left(\sum_{i=1}^{d} \left(x_i - w_i \right)^2 \right)^{1/2} \qquad \text{Eq. (1)}$$

Then, the input vector $\left(X = (x_1, x_2, ..., x_d) \right)$ is simultaneously compared with all elements in the network. Therefore, the neuron which weight vector lies the closest to the input vector is considered as winning neuron.

$$\|X - m_c\| = min \left\{ \|X - m_r\| \right\} \qquad \text{Eq. (2)}$$

, where m_c is the winning neuron and m_r are the references vectors.

Step III: After determining the winning neuron, a set of neighboring neurons are specified that their values should be changed.

Step IV: In the end, the weights of winning neuron and its neighbors should be reformed based on the network input as follows:

$$m_r(t+1) = m_r(t) + \alpha(t). \\ h_{cr}(t) \left[x(t) - m_r(t) \right] \qquad \text{Eq. (3)}$$

, where $x(t)$ is the input vector at time t, $m_r(t)$ is rth basic structure at time t,   is the learning rate at time t, and $h_{cr}(t)$ is the neighborhood function that is defined based on kernel function as follows:

$$h_{cr}(t) = exp\left(-\frac{\|k_c - k_r\|^2}{2\sigma(t)^2} \right) \qquad \text{Eq. (4)}$$

, where $k_c, k_r \in \Re^d$ are the winning neuron and its neighbor basic structure respectively and $\sigma(t)$is the kernel radius at time t. $\alpha(t)$ is the exponential decay learning factor which is applied in order to control convergence of algorithm and it depends on the number of iterations.

3. The ANN Approach to Causal Modeling

Artificial neural networks have powerful pattern classification and pattern recognition capabilities and they have several distinguishes features that make them attractive and practical for a forecasting task [13]. One of the most advantages of the ANN model over other classes of nonlinear models is that ANNs are universal functional approximators. It has been shown that a network can approximate any continuous function to any desired accuracy. Their power stems from the parallel processing of the information from the data that this characteristic makes them a powerful computational device and able to learn from examples and then to generalize to examples never before seen. The artificial neural networks used in widely problems especially in forecasting because of their inherent capability of arbitrary input-output mapping [13]. The model is characterized by a network of three layers of simple processing units connected by acyclic links (Fig. 2).

For causal forecasting the relationship between the output (y_t) and inputs ($x_1, x_2, ..., x_p$) there is the following mathematical representation:

$$y_t = w_0 + \sum_{i=1}^{q} w_j . g\left(w_{0j} + \sum_{i=1}^{p} w_{i,j} . x_{t,i} \right) + \varepsilon_t \qquad \text{Eq. (5)}$$

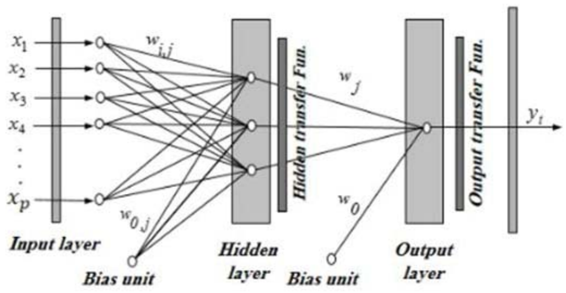

Fig. 2. Neural network structure($N^{(p-q-1)}$).

, where $w_{i,j}(i=0,1,2,...,p, j=1,2,...,q)$ .and $w_j(j=0,1,2,...,q)$ are model parameters often called connection weights; p is the number of input nodes; and q is the number of hidden nodes. Data enter the network through the input layer, moves through hidden layer, and exits through the output layer. Each hidden layer and output layer nodes collects data from the nodes above it (either the input layer or hidden layer) and applies an activation function. Activation functions can take several forms. The type of activation function is indicated by the situation of the neuron within the network. In the majority of cases input layer neurons do not an activation function, as their role is to transfer the inputs to the hidden layer. The logistic and purelin functions are often used as hidden layer and output transfer function for forecasting problems that are shown in Eq. (6) and Eq. (7), respectively.

$$Sig\ (x) = \frac{1}{1+exp\ (-x\ )} \qquad \text{Eq. (6)}$$

$$f\ (x\ ) = x \qquad \text{Eq. (7)}$$

Hence, the ANN model of (5), in fact, performs a nonlinear functional mapping from explanatory variables to the value (y_t), i.e.,

$$y_t = f\ (x_1, x_2, ..., x_p, w) + \varepsilon_t \qquad \text{Eq. (8)}$$

, where w is a vector of all parameters and f(.) is a function determined by the network structure and connection weights. Thus, the neural networks are equivalent to a nonlinear regression model. Note that expression (5) implies one output node in the output layer, which is typically used for one-step-ahead forecasting. The simple network given by (5) is surprisingly powerful in that it is able to approximate the arbitrary function as the number of hidden nodes when q is sufficiently large [14)]. In practice, simple network structure that has a small number of hidden nodes often works well in out-of-sample forecasting. This may be due to the over fitting effect typically found in the neural network modeling process. An over fitted model has a good fit to the sample using for model building but has poor generalizability to data out of the sample [15)].

The choice of q is data-dependent and there is no systematic rule in deciding this parameter. In addition to choosing an appropriate number of hidden nodes, another important task of ANN modeling of a causal problem is the selection of the number of input variables, p, and the dimension of the input vector [16)]. This is perhaps the most important parameter to be estimated in an ANN model because it plays a major role in determining the (nonlinear) autocorrelation structure of the causal model.

There exists many different approaches such as the pruning algorithm, the polynomial time algorithm, the canonical decomposition technique, and the network information criterion for finding the optimal architecture of an ANN [17)]. These approaches can be generally categorized as follows: (i) Empirical or statistical methods that are used to study the effect of an ANN's internal parameters and choose appropriate values for them based on the model's performance [18, 19)]. The most systematic and general of these methods utilize the principles from Taguchi's design of experiments [20)]. (ii) Hybrid methods such as fuzzy inference [21)] where the ANN can be interpreted as an adaptive fuzzy system or it can operate on fuzzy instead of real numbers. (iii) Constructive and/or pruning algorithms that, respectively, add and/or remove neurons from an initial architecture using a previously specified criterion to indicate how ANN performance is affected by the changes [(19, 22- 24)]. The basic rules are that neurons are added when training is slow or when the mean squared error is larger than a specified value, and that neurons are removed when a change in a neuron's value does not correspond to a change in the network's response or when the weight values that are associated with this neuron remain constant for a large number of training epochs [25)]. (iv) Evolutionary strategies that search over topology space by varying the number of hidden layers and hidden neurons through application of genetic operators [26, 27)] and evaluation of the different architectures according to an objective function [28, 18)].

4. Iran's Steel Consumption Forecasting
4.1. Feature subset selection

The information used in this investigation consists of 37 annual observations of the crude steel consumption of Iran and the other explanatory variables from 1357 to 1393 (based on the Solar Hijri calendar). In this way, 30 observations are first used to formulate models and seven observations are randomly used to evaluate the performances of the proposed models. In this paper, primary variables are determined using the literature review [30-33)] and survey of experts in Mobarakeh Steel Company (MSC), which are shown in Table 1.

Then, these variables are clustered using self-organizing maps based on the correlations between explanatory variables with the target variable. Clustering process is carried out by analysis of the maps associated with the independent variables and steel consumption as shown in Fig. 3.

According to analysis of nonlinear correlations between the independent variables and the target variable, primary explanatory variables are clustered into four categories as shown in Table 2, in a way that, the variables in each cluster have the highest correlation with each other and the lowest correlation with variables of the other clusters. Hence, in order to increase the accuracy and performance of steel consumption model, only one of the variables in each category should be used as a final effective input variable, unless the number of final input variables limitation is satisfied or the explanatory of the new input variable is more than its correlation with other variables. As a result, final effective input variables are shown in Table 3.

Table 1. Primary Variables.

Symbol	Variable	Symbol	Variable
X01	Crude steel consumption	X14	Oil value added
X02	Crude steel production	X15	Inflation
X03	Gross domestic production	X16	Liquidity
X04	Net indirect taxes	X17	Exchange rate
X05	National income	X18	Investment (Total)
X06	Gross national production	X19	Investment (public sector)
X07	Economic growth rate	X20	Investment (private sector)
X08	Oil price	X21	Interest rate
X09	Industries and mines value added	X22	Electricity prices
X10	Industry value added	X23	Wage rates
X11	Building value added	X24	Crude steel price
X12	Services value added	X25	Steel import
X13	Agriculture value added		

Table 2. Clustering of the primary variables.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Gross domestic production	Steel import	Steel price	Investment (Total)
Oil price	Inflation	Crude steel	Investment (public sector)
National income	Liquidity	production	Investment (private sector)
Value added (all sectors)	Costs	Exchange rate	Interest rate

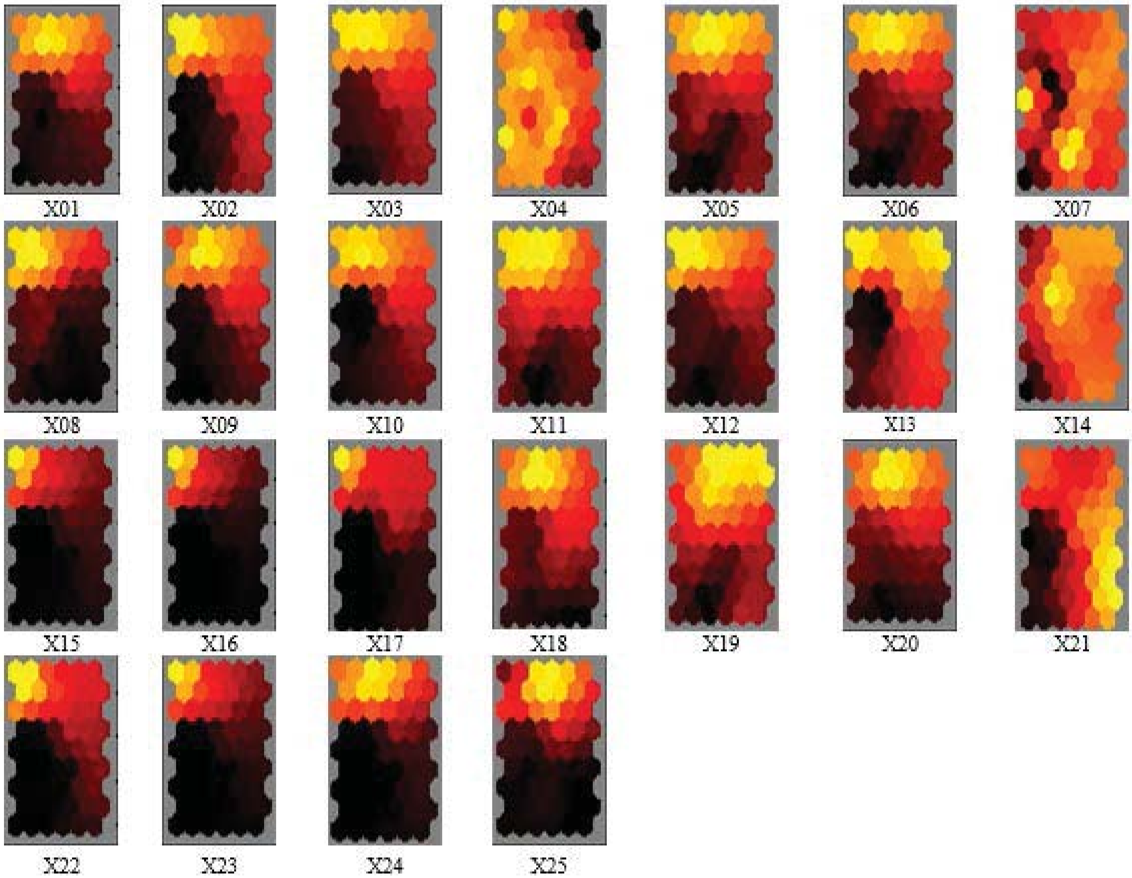

Fig. 3. Self-organizing maps.

In the end, after determination the final effective input variables, in order to obtain the optimum network architecture, based on the concepts of artificial neural network design and using Constructive algorithm in MATLAB 2014 package software, different network architectures were evaluated to compare the ANNs performance. The best fitted network which was selected, and therefore, the architecture which presented the best forecasting accuracy with the test data, was composed of five inputs, three hidden and one output neurons ($N^{(5-3-1)}$). The structure of the best-fitted network is shown in Fig. (4).

4.2. Comparison the proposed model with other feature selection models

In this section, the pattern recognition capability of the self-organizing map (as the nonlinear feature selection method) is compared with the forward and backward models (as the linear feature selection methods), using Iran's steel consumption data set. According to this, the MAE (Mean Absolute Error) and MSE (Mean Squared Error) are considered as performance indicator in order to measure forecasting accuracy of intelligent model in comparison with classical models.

Table 3. Final effective input variables.

Symbol	Variable
X1	Crude steel production
X2	Gross domestic production
X3	Steel price
X4	Investment (Total)
X5	Steel import

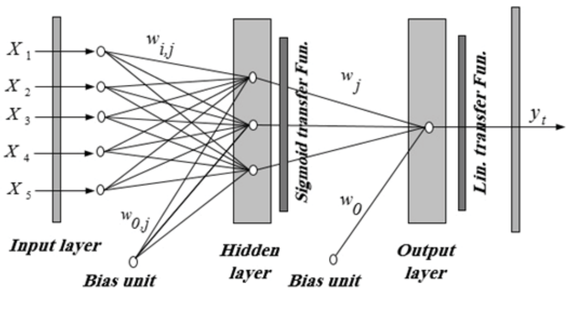

Fig. 4. Structure of the best-fitted network, $N^{(5-3-1)}$.

These performance indicators are respectively calculated from the following equations:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |e_i| \qquad \text{Eq. (9)}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (e_i)^2 \qquad \text{Eq. (10)}$$

As mentioned above, Iran's crude steel consumption data set is applied to compare the effect of feature selection procedures on forecasting capability. According to the obtained results from Table 4, self-organizing neural network has the lowest error on the test portions of the Iran's steel consumption data in comparison to other those feature selection models, in a way that, the MAE and MSE of the SOM model are significantly lower than the traditional linear feature selection models for steel consumption case. The numerical results show that the intelligent feature selection model gives significantly better forecast than classical models. For example based on obtained results from Table (5) in terms of MSE (when the ANN is considered as forecasting method), SOM indicates 9.61% and 9.15% improvement over the forward and backward models, respectively.

It is important to mention that the forecasting performance of the artificial neural network model is rather satisfactory than the classical linear regression model, in a way that, the MAE and MSE of the ANN model are significantly lower than the traditional regression model for all of the feature selection methods. The reason is that the artificial neural networks can properly model both linear and nonlinear structures.

5. Conclusions

This paper revealed the advantages of using self-organizing map as a nonlinear feature selection method to improve the clustering accuracy of neural networks in the context of a nonlinear pattern recognition method. In this way, primary features have been clustered using self-organizing maps. Then, the final effective variables are selected based on the maximum nonlinear correlations between the variables of each cluster and the target variable. In the end, the performance of this intelligent feature selection algorithm is compared with classical linear feature selection algorithm such as forward and backward methods.

Table 4. Comparison of the performance of the linear and nonlinear feature selection methods.

Feature Selection Method / **Forecasting Model**	**Forward**		**Backward**		**Self-Organizing Maps**	
	MSE	MAE	MSE	MAE	MSE	MAE
Classical Linear Regression	949981.6	695.49	882115.4	683.07	1780650	905.81
Artificial Neural Network	676489.8	673.26	673023.3	661.25	611445.8	648.29

Table 5. Comparison of the performance of feature selection methods in the ANN model.

Model	MSE	Improvement Percentage		
		Forward	Backward	SOM
Forward	676489.8	0.0%	…	….
Backward	673023.3	0.51%	0.0%	….
SOM	611445.8	9.61%	9.15%	0.0%

The empirical results of this research showed that SOM performed the best as intelligent feature selection method for nonlinear pattern recognition and improving clustering accuracy in complex and high-dimensional situations. According to the results of forecasting by the classical linear regression model, using the linear feature selection methods were provided of a higher quality results than the nonlinear feature selection method. Because the traditional forecasting model only had the ability to detect linear structures between the variables. For the same reason, for intelligent forecasting model, nonlinear feature selection technique was provided more satisfactory results than the linear feature selection methods. In addition, since the artificial neural networks were able to model pure linear and nonlinear structures, forecasting by the ANNs produced more accurate results than the classical linear regression model when both linear and nonlinear feature selection methods were applied.

Acknowledgments

The authors would like to express their deepest gratitude to Mr. Gholamreza Taheri, Director of the sales department of Mobarakeh Steel Company (MSC), and Mr. Majid Fakhari, Expert of the sales department of Mobarakeh Steel Company (MSC), for their helpful comments and support, which considerably help us improve our paper.

References

[1] T. Kohonen: Proc. of 2SCIA. Conf. on Image Analysis, Helsinki, Finland, (1981), 214.
[2] T. Kohonen: Springer Series in Information Sciences, 30(2001), Springer, Berlin, New York.
[3] Y. Yang, W. Tan, T. Li and D. Ruan: Knowl. Based. Syst., 32(2012), 101.
[4] S. L. Shieh, L. E. Liao: Expert. Syst. Appl., 39(2012), 11924.
[5] N. Chen, B. Rebeiro, A. Vieira and A. Chen: Expert. Syst. Appl., 40(2013), 385.
[6] C. Hajjar and H. Hamdan: Neural Networks., 46(2013), 124.
[7] M. H. Ghasemzadeh and A. Karami: Appl. Soft. Comput., 11(2011), 3771.
[8] H. Liu and X. J. Ban: Expert. Syst. Appl., 42(2015), 4965.
[9] X. S. Xu, X. L. Liamg, G. Q. Yang, X. L. Wang, S. Guo and Y. Shi: Neurocomputing., 236(2017), 56.
[10] H. Qui, Y Xu, L. Gao, X. Li, L. Chi: Expert. Syst. Appl., 46(2016), 180.
[11] J. Haga, J. Siekkienen and D. Sundvik: Expert. Syst. Appl., 42(2015), 8327.
[12] L. Wang and C. Wu: Knowl. Based. Syst., 121(2017), 99.
[13] P. Zhang, B. E. Patuwo and M. Y. Hu: INT. J. Forecasting., 14(1998), 35.
[14] K. Smith, N. D. Jatinder: Comput. Oper. Res., 27(2000), 1045.
[15] H. Demuth and B. Beale: The Math Work Inc., Natick, 2004.
[16] S. Thawornwong and D. Enke: Neurocomputing., 31(2000), 1.
[17] M. Khashei, Master of Science Thesis, Isfahan University of Technology, 2005.
[18] P. G. Benardos and G. C. Vosniakos: Eng. App. Of Artificial Intelligence., 20(2007), 365.
[19] L. Ma and K. Khorasani, Neurocomputing., 51(2003), 361.
[20] J. P. Rpss: McGrow-Hill, New York, 1996.
[21] J. Leski and E. Czogala: FUZZY SET SYST., 108(1999), 289.
[22] S. D. Balkin and J. K. Ord: Int. J. Forecasting., 16(2000), 509.
[23] M. M. Islam and K. Murase: Neural Networks, 14(2001), 1265.
[24] X. Jiang and A. H. K. S. Wah: Pattern Recognition., 36(2003),853.
[25] D. Marin, A. Varo and J. E. Guerrero: Talanta., 72(2007), 28.
[26] P. A. Castillo, J. J. Merelo, A. Prieto, V. Rivaas, G. Romero and G. Prop: Neurocomputing., 35(2000), 149.
[27] J. Lee, S. Kang: Int. J. Solids. Struct., 44(2007), 5980.
[28] J. Arifovic and R. Gencey: Phys. A., 289(2001), 574.
[29] G. P. Zhang and B. E. Hu: Neurocomputing, 56(2004), 205.
[30] M. Evans: Resources Policy, 36(2001), 97.
[31] M. C. Roberts: Resources Policy., 16(1990), 56.
[32] M. C. Roberts:Resources Policy., 22(1996), 183.
[33] P. Crompton: Resources Policy., 26(2000), 26.