Research Article

Intelligent Sales Prediction for Pharmaceutical Distribution Companies: A Data Mining Based Approach

Neda Khalil Zadeh,[1] Mohammad Mehdi Sepehri,[1] and Hamid Farvaresh[2]

[1] *Department of Industrial Engineering, Tarbiat Modares University, Tehran 14117 13114, Iran*
[2] *Department of Industrial Engineering, University of Kurdistan, Sanandaj 66177 15177, Iran*

Correspondence should be addressed to Mohammad Mehdi Sepehri; mehdi.sepehri@gmail.com

One of the problems of pharmaceutical distribution companies (PDCs) is how to control inventory levels in order to prevent costs of excessive inventory and to prevent losing customers due to drug shortage. Consequently, the purpose of this study is to propose a novel method to forecast sales of PDCs. The presented method is a combination of network analysis tools and time series forecasting methods. Due to lack of enough past sales records of each drug, an explorative network based analysis is conducted to find clique sets and group members and to use comembers' sales data in their sales prediction. Afterwards, time series sales forecasting models were built with three different approaches including ARIMA methodology, neural network, and an advanced hybrid neural network approach. The offered hybrid method by applying each drug and its comembers past records facilitates capturing both linear and nonlinear patterns of sales accurately. The performance of the proposed method was evaluated by a real dataset provided by one of the leading PDCs in Iran. The results indicated that the proposed method is able to cope with low number of past records while it forecasts medicines sales accurately.

1. Introduction

Precise sales prediction is an essential and inexpensive way for each company to augment their profits, decrease their costs, and achieve greater flexibility to changes. In other words, exact sales forecasting is utilized for capturing the tradeoff between customer demand satisfaction and inventory costs [1]. Especially, for the pharmaceutical industry, successful sales forecasting systems can be very beneficial, due to the short shelf-life of many pharmaceutical products and the importance of the product quality which is closely related to the human health [2].

Actually, PDCs are facing several challenges, including huge amount of inventory, increased competition, and tough regulations that limit advertising. They have to meet their customers' needs by delivering the right amount of medicines to the right place and at the right time. In PDCs, both shortage and surplus of goods can lead to loss of income for these companies. Accordingly, one of the problems in PDCs is how much quantity of each drug should be kept in the inventory.

Distribution of pharmaceutical products in Iran is mainly by independent wholesalers, who buy stock for their own account from manufacturers and sell to their customers (mainly hospitals and pharmacies). Distribution of pharmaceutical products is done either by full-line or short-line wholesalers. Full-line wholesalers distribute the full range of available pharmaceutical products and they have some central depots in big provinces, while short-line wholesalers trade in a selection of products only. In Figure 1, a schematic of pharmaceutical distribution channel in Iran is presented. As it is demonstrated in Figure 1, pharmaceutical manufacturers regularly purchase required raw material and product packages from foreign suppliers. After producing, they sell their finished products to distributors through face-to-face visit, web ordering, or teleordering. However, in some cases, distribution companies belong to the manufacturers; accordingly, they directly distribute their own products, which are shown by a direct link from manufacturer to distributors in Figure 1. Furthermore, importers typically sell the imported products to distributors through face-to-face visits, web
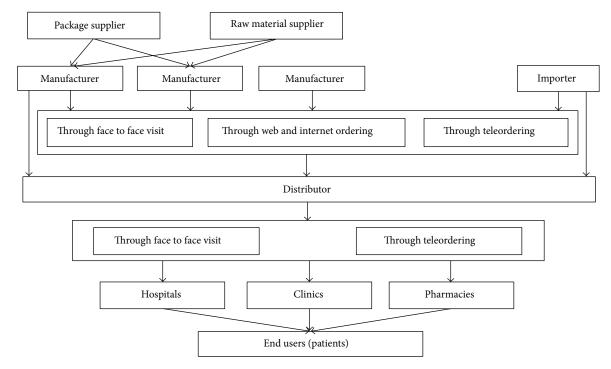
Figure 1: Pharmaceutical distribution channel.

ordering, or teleordering. However, sometimes, importing companies belong to distributors. Consequently, distributors sell their own imported medicines. That is why there are two links between importer and distributor in Figure 1. Afterward, pharmaceutical distributors sell the medicines to hospitals, clinics, and pharmacies where patients obtain their required drugs, through teleordering or face-to-face visit.

Most PDCs in Iran still use heuristic or simple statistical method for their sales forecasting. With the access to past sales data and by use of data mining techniques, almost all companies and especially pharmaceutical distribution centers can make accurate and reliable prediction for the future sales. Since sales prediction must be performed with high accuracy and in a short time, it is impossible to do it with manual or traditional methods. Consequently, it is really preferred to apply one of the data mining techniques to enhance the accuracy of sales prediction and also speed up its process.

In this research, we have collected the required data from a large PDC, which dispenses medicines to customers in a number of provinces in Iran. After receiving the orders, this company, likewise other pharmaceutical distributors, is committed to supplying the required drugs to provinces within 24 hours, to cities within 48 hours, and to remote areas within 72 hours. In keeping with its market-leading position, this company needs to have large products' inventories in order to meet customers' demand, as shortage of drugs is not acceptable in this industry. This company like other PDCs in Iran usually keeps inventories for the needs of next 1.5–2 months in advance. This fact causes many excessive

costs and investments for Iranian PDCs. Inventory control, transportation, and financial costs contain a high percentage of total expenses in PDCs. Generally, PDCs buy products from manufacturers and pay for them at once, but they sell products and receive the related money gradually. Thus, this gap causes them undesired expenses. Monthly and precise sale prediction would shorten or even eliminate this gap. Therefore, most distributors in this industry are looking for modern and accurate methods to predict the future sales in order to reduce undesired inventory costs and increase profits while keeping their customers satisfied.

According to the restrictions on sales of medicines such as existing new items with short numbers of past sales records and having great diversity of medicines, the common existing forecasting methods are not effective for pharmaceutical companies since these methods require many past sales records of each item to predict accurately. Consequently, the objective of this research is concerned with the development of a novel and accurate sales forecasting model for pharmaceutical products by means of one of the related data mining approaches to overcome the problem of having numerous kinds of medicines and not having enough past sales records of each medicine.

The organization of the remainder of this paper is as follows. Section 2 reviews the previous literatures. In Section 3, research methodology, including overall research process, methods were applied to each step of graph based analysis and sales prediction, and related results obtained by each step are presented. Finally, conclusion and contributions of the study are discussed.

2. Forecasting on Time Series Data

Studies on forecasting encompass many researches that have examined alternative methods to find out which ones are the most efficient. From 1970, the methodologies that have been applied in sales forecasting have normally been time series methods that can be categorized as linear or nonlinear, according to the nature of the model they are dependent on [2]. Linear models, such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) [3], are successful and famous linear methodologies [4], but their prediction ability is restricted by their supposition of a linear behavior. Consequently, they are not always acceptable [5], specially for applying in non-linear time series forecasting cases [6]. Accordingly, several forecasting techniques have been evolved through last decades; each technique has its specific advantages and drawbacks compared to other approaches. During the 1980s, two critical progresses occurred that influenced the evolution of time series study. On the one hand, the advancements of computer science resulted in using more complex algorithms. On the other hand, the evolution of machine learning techniques, such as artificial neural networks, leaded to significant progress in time series domain [7].

As a matter of fact, neural networks are the most famous models in time series forecasting domain since they have been confirmed as an efficient method to make prediction by dealing with nonlinear input and output variables, being able to estimate any function in definite circumstances [8]. In other words, if the data are influenced by nonlinear behavior, linear methods are not capable of modeling underlying nonlinear relationships. Accordingly, nonlinear methodologies like artificial neural networks are more appropriate than conventional statistical methods [6, 9, 10]. Recently, artificial neural networks have been extensively applied in forecasting and specifically in time series forecasting with considerably good prediction performance [11–14]. The popularity of the neural networks is attributed to their distinctive capability to simulate a wide range of underlying nonlinear behaviors and to cover a broad variety of fields [15–17]. Therefore, numerous studies have tried to implement neural networks to time series forecasting and compared them with conventional methods. However, the outcomes are not distinctly in favor of one specific methodology [2, 18].

2.1. Comparisons of Different Time Series Forecasting Models. Various comparative studies between traditional models and neural networks have been carried out, but outcomes are different with regard to whether ANNs are better than the linear methods in forecasting or not [19, 20]. Numerous researchers have offered empirical confirmation on the comparative preference of one method over the other in different forecasting cases [8, 16, 18, 21]. The majority of these studies illustrate varied results regarding the effectiveness of the ANNs compared with the traditional models like ARIMA [4]. The following survey displays comparisons of diverse time series forecasting methods that developed during 1991–2008.

Since generation of neural networks, only few studies have concluded that traditional methods have superior performance to or at least the same performance as neural networks. For instance, Tang et al. [22] reported the results of a study that compared the performance of neural networks and traditional methods in time series forecasting. Their research demonstrated that for time series with long memory, both ANNs and Box-Jenkins models had the same performance. Furthermore, Foster et al. [23] noticed that exponential smoothing is better than neural networks in forecasting yearly data, and comparable in forecasting quarterly data. Although it appears that neural networks may outperform conventional statistical methods in forecasting time series with trend and seasonal patterns, Nelson et al. [21] found that neural networks cannot properly model the seasonal patterns in their data. Also, Callen et al. [24] showed that forecasting performance of linear time series models was better than that of neural networks even if the data were nonlinear. Moreover, Church and Curram [25] and Ntungo and Boyd [26] demonstrated that neural networks performed nearly the same as econometric and ARIMA approaches. Afterwrads, Heravi et al. [27] showed that linear models generate more reliable prediction results than neural networks for European industrial production series.

In contrast, many researchers concluded that neural networks had better forecasting performance than linear methods. As an illustration, in a comparative study of the performance of neural networks and conventional methods in forecasting time series, Tang et al. [22] found that ANNs surpassed the Box-Jenkins in short term forecasting. Then, Chakraborty et al. [28] applied neural networks approach to multivariate time series analysis. They precisely forecasted the flour prices in three cities in USA. According to their outcomes, neural networks approach was better than classic methods. In a succeeding study, Ansuj et al. [29] compared the ARIMA model with interventions and ANN model in examining the behavior of sales in a medium size corporation. The outcomes proved that ANN model is more precise. Hill et al. [8] also demonstrated that neural networks were considerably superior to traditional techniques when predicting quarterly and monthly data. Agrawal and Schorling [30] compared the test results of neural networks with multinomial logit model of forecasting and they proved that neural networks could predict brand shares more accurately. Furthermore, Yip et al. [31] examined application of neural networks in sales forecasting. Applying several measures of accuracy, the result of the evaluation confirmed that the neural networks predict better than the time series smoothing methods of forecasting.

Afterwards, Zhang et al. [20] did an inclusive review of the literature regarding the employment of ANNs in various forecasting domains. In less than one-thirds of cases, ANNs were equivalent to linear methods and in about two-thirds of the cases, ANNs were better. In 1998, Kuo and Xue [32] proved that their developed artificial intelligence models could find nonlinear relationships better than conventional time series methods. Elkateb et al. [33] also compared neural networks with ARIMA models in peak load forecasting. The results of their study demonstrated that neural networks had better forecasting performance than ARIMA models. In fact, neural networks have surpassed conventional forecasting

techniques in definite conditions (Hornik et al; cited by [34]), for example, when there are nonlinearities (White and Stinchcombe; cited by [34]) and also when there are considerable interactions among inputs [35].

Subsequently, Ainscough and Aronson [36] compared neural network and regression analysis, as a linear model, in modeling and predicting the results of retailer activity on the sales of definite products applying scanner data. According to the results of their study, neural networks had better performance than regression model. Then, Qi [37] reported that ANNs are very likely to do better than other methods when the required data are saved as recent as possible. Although statistical techniques have been proven effective for a long time, they still have definite drawbacks [10, 32, 38]. For example, when the data are influenced by particular conditions, like promotion, the prediction results of conventional methods are undesirable [18, 38]. In addition, it is always logical to anticipate having a noteworthy level of nonlinearity in sales behavior [9]. However, traditional methods are not capable of modeling nonlinear relationships, so most researchers apply neural network models to cope with forecasting problems. Although the contribution of neural networks compared to the traditional methods seems degraded in some cases [39], their vigorous capability to model nonlinear relations and their adaptation are highly attractive for most forecasting subjects [40]. Therefore, advanced methods like neural networks could serve as more appropriate approximator [40] and would be more suitable for the time series sales forecasting than linear models like ARIMA.

However, Adya and Collopy [19] in their review paper presented that regardless of growing applications of neural networks in prediction and more specifically in business prediction, opinions and outcomes concerning their contribution are varied. Adya and Collopy [19] concluded that assessing studies in this area is complicated, owing to lack of obvious criteria. It means that, although neural networks have acquired growing attention in forecasting domain, resulting in competent applications in time series sales forecasting [41], some studies point out that no single approach works best in every condition, and combining diverse methods is an efficient and effectual way to advance forecasting accuracy [5]. Consequently, some recent researches offered good explanations of the hybrid ARIMA-ANN models or combination of other conventional and ANN techniques [5, 42–45]. As an illustration, Zhang [5] recommended a hybrid ARIMA-ANN approach in which ARIMA was applied to model the linear part and ANNs were used to model the prediction errors. He showed that the hybrid method outperformed both separate methodologies. Furthermore, in a research by Kuo et al. [46] a hybrid algorithm founded on radial basis function neural network for sales prediction. They presented hybrid of "particle swarm and genetic algorithm based optimization (HPSGO) algorithm" congregated advantages of "particle swarm optimization (PSO) and genetic algorithm" to advance the learning performance of RBF neural network. Outcomes of the research showed that the presented HPSGO algorithm had better performance than PSO, genetic algorithm,

and Box-Jenkins model [46]. Also, Khashei and Bijari [47] proposed a new hybrid neural networks model. They applied an artificial neural network (p, d, q) model for time series forecasting to have a more precise forecasting model than neural networks [47]. The results of three real data sets pointed out that the proposed model was an appropriate way to advance forecasting accuracy accomplished by neural networks. Both theoretical and empirical results have confirmed that combination of various models is a successful way of enhancing the performance of forecasting models [47]. Subsequently, a novel hybrid model merging autoregressive fractionally integrated moving average (ARFIMA) and feed forward neural networks (FNN) was offered by Aladag et al. [48] to examine time series tourism data. They compared their hybrid approach with other methods and it was revealed that the offered hybrid methodology had the great superiority over other approaches in terms of forecasting precision [48]. Furthermore, Wang et al. [4] suggested a hybrid model, integrating the benefits of ARIMA and ANNs in modeling both the linear and nonlinear behaviors in the data set. Their hybrid model was examined on three sets of real data, that is, the Wolf's sunspot data, the Canadian lynx data, and the IBM stock price data. The results pointed out the superiority of their proposed combinatorial methodology in acquiring more precise forecasting in comparison with existing approaches [4]. Accordingly, it appears to be logical to apply hybrid models in most sales forecasting domains.

According to the presented survey, numerous prediction methods have been offered and each method has its specific advantages and disadvantages in comparison with other techniques. However, none of the accomplished studies described the applications of hybrid linear and nonlinear neural networks in forecasting. They also did not offer a novel technique for handling the problem of not having enough past records for forecasting. However, owing to the specific constraints of pharmaceutical products that is having numerous new items with few historical data, existent forecasting methods are generally inappropriate. This motivates the evolution of a novel hybrid approach, which combines both linear and nonlinear methods and their relevant strengths. Accordingly, in this research, the use of hybrid neural network by using each medicine's past records and its group members' past records to make precise sales prediction for PDCs was examined.

3. Research Methodology

As demonstrated in Figure 2, the overall procedure of this research consists of data collection and preparation, exploratory analysis, graph based analysis, data sampling for model fitting and testing, model building or sales prediction, model evaluation, and finally, conclusion. In this section, each step and different methods associated with each stage are explained.

3.1. Data Collection and Preparation. To predict the sales of a company's products in an appropriate time horizon, past sales records of a preferred PDC (Pakhsh Hejrat Co.) were

```
                                          ┌─────────────────────┐
                                          │   Data collection   │
                                          └──────────┬──────────┘
                                                     ↓
┌─────────────────────┐ ............▷  ┌─────────────────────┐
│  Data preparation   │                  │  Data preprocessing │
└─────────────────────┘                  └──────────┬──────────┘
                                                     ↓
┌─────────────────────┐ ............▷  ┌─────────────────────┐
│ Exploratory analysis│                  │  Data visualization │
└─────────────────────┘                  └──────────┬──────────┘
```

Cross-correlation matrix	Adjacency matrix	Clique sets	Graph visualization

Graph based analysis

Training data (fitting)	Validation data (fitting)	Testing data (testing)

Data sampling for model fitting and testing

ARIMA	Hybrid neural networks using past records of each drug	Hybrid neural networks using past records of each drug and its co-members'

Model building and fitting

Mean squares error	Mean absolute error

Performance evaluations on test data

Results and prediction models' ranks
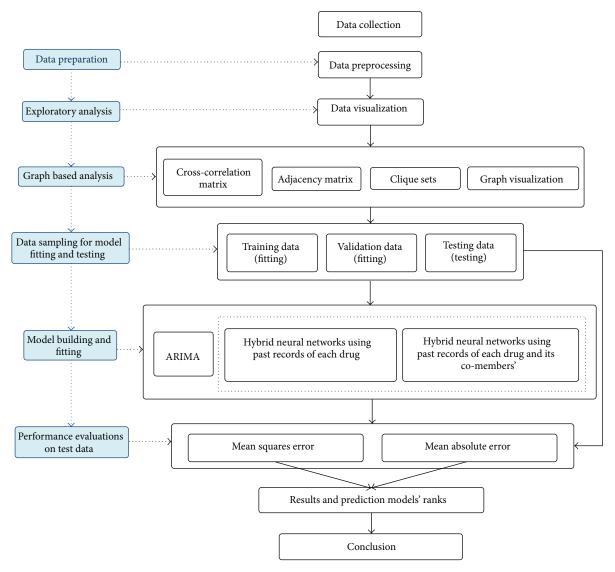
Conclusion

Figure 2: Research Process.

collected. This company provided us with monthly sales data of near 1200 kinds of medicines which were sold to different provinces or centers of Iran during three years. Database of Pakhsh Hejrat Co. includes name and code of medicines, sales number, name and code of centers, name of manufacturers, and price and monthly date of sales. To approach the objective of the research, code, date, and number of sold products were selected from database. Subsequently, 217 kinds of drugs which have been sold in all 36 months were extracted. Code of medicines was changed to 1-217 for more easiness. Consequently, these 217 kinds of drugs were employed in grouping and model building phases.

3.2. Exploratory Analysis. After data preprocessing, in order to better understand the nature of our data, an exploratory analysis was conducted. The exploratory analysis consists of the following steps.

3.2.1. Data Visualization. In this part, sales plots and surface plots of time series data for all medicines were drawn. For example, Figure 3 shows the time series plot of drug 24. In this plot, we can see considerable downward trend with nonlinear fluctuations around it. The plot also shows a weak seasonality, and its mean increases as the number of months increases. The variance is also altering. Sales plots of drugs represented that (1) some sales series had upward trend, some had downward trend, and some did not have any trend, (2) some series were stationary, but the others were not, (3) some series were seasonal, yet the others were nonseasonal, and (4) the variance was also altering through time for most products. As a result, it is not possible to apply a unique model for all drugs. It means that these medicines were different products and have diverse characteristics. However, it may be possible to make use of similar sales behavior of some of them in their sales prediction.
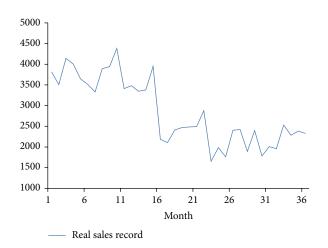
FIGURE 3: Time series plot of sales data for drug 24.

As an informal way to explore the relationships between variables, the surface plots for all medicines were also depicted in order to see whether they have linear or nonlinear relationships. Consequently, after examining the relationships of all drugs we can say that although in some cases there are some linear relationships, in most cases the relationships among medicines are mainly nonlinear.

To summarize, according to the exploratory analysis it was concluded that (1) most medicines had different and specific characteristics and sales behavior, (2) it was impractical to make a single prediction model for all medicines, and (3) there were both linear and nonlinear relationships among sales records, but there were mostly nonlinear. Thus, it is logical to mainly apply nonlinear or even hybrid models in this research and consider linear models as the second priority.

3.3. Graph Based Analysis of Medicines' Networks. In the current study, forecasting problem has specific features that distinguish it from similar cases. These features are that (1) there are relatively many different kinds of products, (2) there are low numbers of past sales records for each product, (3) there are complementary and substitutable relationships in consumption of products, and (4) there may be irregular and nonlinear patterns in consumption behavior of most products. These conditions exist in a PDC that has to provide many and various kinds of medicines for its customers. In these situations, ARIMA and most of linear statistical time series forecasting methods are not recommended.

In forecasting problems (including sales forecasting), it is basically needed to identify effective variables and model their behavior. In ARIMA methodology, identifying and modeling of change patterns in target variables are emphasized. In fact, it is assumed that target variable trends and fluctuations are reflection of the outcome effect of all effective but absent variables. Accordingly, it is logical to use past records of the target variable in order to predict its future values. However, this question has always been propounded that whether change patterns of a single variable are able to reflect the resultant effects of all effective variables properly

or it would be better to use a set of most explanatory variables instead of a single variable. Subsequently, vector autoregression (VAR) methodology had been presented to find the evolution and the interdependencies among several time series, generalizing the univariate autoregressive models. All variables in a VAR model are considered symmetrically by containing an equation for each variable clarifying its development based on its own lags and lags of all other variables in the model. However, excessive necessities, strong assumptions, and conditions of this method limit its usage in most cases.

In fact, in this research, it is experimentally showed that in cases with aforementioned conditions (1 to 4), using past records of the target variables, as well as past records of some potentially explanatory variables, could not give us worse prediction results than those of single variable. However, the key question is that for sales prediction of some products like medicines that there are many various items, past sales records of which medicines should also be used to predict future sales of other drugs. To answer the mentioned question, we introduced a novel approach, which was to conduct a graph based analysis in order to find groups of medicines that had similar sales behavior.

3.3.1. Graph Based Analysis. In this research, a graph based analysis was performed to find groups of medicines which have similar sales behavior or high sales cross-correlation, then using past sales records of comembers for each other in their sale prediction. The proposed graph based analysis permits one to visualize a dataset through signifying its components as vertices and monitor definite relationships among them. One can simply visualize a graph as a set of vertices and edges connecting those [49]. In order to explain the network analysis part, it is needed to present the following symbols and notations:

$$I = \{i_1, i_2, \ldots, i_{|I|}\}: \text{Set of products that their sales}$$
$$\text{must be predicted.}$$
$$S_i = \{s_1, s_2, \ldots, s_T\}: \text{Sales records of item } i \text{ in the period}$$
$$\text{between 1 and } T.$$
$$(1)$$

A graph representation of medicines is based on the cross-correlations of their sales. The graph is constructed as follows: a vertex is associated with each medicine, and two vertices are connected by an edge if the correlation coefficient of the corresponding pair of drugs (calculated over a certain period of time) exceeds a specified threshold. Let $G = (V, E)$ be an undirected graph with the set of n vertices V and the set of edges $E = \{(i, j) \mid i, j \in V \text{ [50] and } r_{ij} > \theta\}$, where r_{ij} and θ are the correlation between drugs i and j and the adopted threshold, respectively. It is said that the graph $G = (V, E)$ is connected if there is a path from any vertex to any other vertex in the graph. The degree of a vertex is the number of edges emerging from it [50]. A clique within a graph is a set of totally interrelated vertices of the graph [50]. In this research, it is proposed to find clique sets of medicines in order to identify

sets of items that show harmonious changes. Accordingly, the following steps were followed.

Step 1. Firstly, cross-correlation matrix $(R = [r_{ij}]_{|I| \times |I|})$ of variables was built in related period. The elements of this matrix consist of correlation coefficient $r_{ij} =$ correlation (s_i, s_j) between past sales records of i and j products.

Step 2. Then, adjacency matrixes $(A = [a_{ij}]_{|I| \times |I|})$, that contains of 0, 1 values, were built so that if the Correlation r_{ij} between past sales records of products i and j was greater than θ, the element a_{ij} of the adjacency matrix would be 1, otherwise 0. In fact, this matrix is an adjacency matrix of a graph so that each node of the graph presents a product. Values 1 in adjacency matrix corresponding to two products indicate relatively strong relationship among their past sales records and 0 shows weak relationship.

Step 3. The graph, which was equivalent to matrix A, was built with an appropriate threshold for θ. In fact, it was desired to choose the best threshold that gave an appropriate number of cliques and the best combination of cliques (medicines' groups). Therefore, as a heuristic approach, it was desirable to have all the following criteria simultaneously:

(1) a large number of cliques,

(2) modest variance or standard deviation of the cliques' size,

(3) high mean of degree for the nodes,

(4) nearly identical distribution of nodes' degrees (low variance or standard deviation of degrees).

Accordingly, it was preferred to have the maximum of the following index:

$$\begin{aligned} \text{Index} = &[(\# \text{ of Cliques}) \times (\text{Mean of Degree})] \\ &\times ([[(\text{Standard Deviation of cliques' size}) \\ &\times (\text{Standard Deviation of Degrees})]]^{-1}. \end{aligned} \quad (2)$$

In consequence, above index for different thresholds from $\theta = 0.3, 0.35, 0.4, \ldots, 0.9$ was calculated and it is presented in Figure 4. Heuristically, the best threshold was $\theta = 0.65$. Consequently, the adjacency matrix for $\theta = 0.65$ was the input of the graph based analysis. Then, it was feasible to group medicines according to their cross-correlation or their mutual relationship. This grouping would help to perform improved and accurate sales forecasting.

Clique Sets. Clique sets (groups) of medicines were detected according to threshold of 0.65 and by means of Gephi software. In analysis of the medicines' graph, a relatively high correlation was chosen to ensure that only significantly correlated edges corresponding to the pairs of drugs were considered. The results, demonstrated in Table 1, stated that which members (drugs) are in each clique. As it is evident, 63
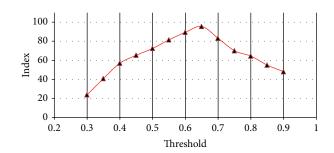

FIGURE 4: Heuristic indices changes for different values of θ.

cliques or groups which contained medicines that have cross-correlations more than 0.65 with each other were found. In fact, in this phase, similar and harmonious changes in sales behavior of medicines were used in order to find clique sets of drugs.

After detecting clique sets, the network of medicines was visualized to summarize the findings of previous section (Figure 5). The drugs that are connected to each other are obvious in Figure 5. For instance, some drugs, such as drug 115, were alone and were not connected to any other drugs. These medicines had distinct consumption or sales behavior; accordingly, we could not predict their sales by using sales records of other medicines. However, some medicines like 24, 174, and 103 were pivotal drug as they were connected to many other drugs. These key drugs had similar consumption or sales behavior to many other drugs. Thus, sales records of various drugs could be used in sales prediction of these critical products. In addition, their sales records could be employed in prediction of many other medicines, which were connected to these critical products. Consequently, output of this section would help to overcome the problem of not having enough past sales records of each medicine since it would be possible to make prediction by using past records of each drug and its group members as input variables.

Actually, finding clique sets (groups) of medicines, which have similar changes in their past sales records, would have substantial applications in pharmaceutical sales management and warehousing. Finding groups of products that have similar sale behavior would make it possible to apply sales records of group members as indicators or input variables for sales prediction of each other. In addition, finding group members would have significant application in sales management of these products. As an illustration, managers may decide which items would have similar sales trend or sales behavior in future in order to set up similar sales projection for them.

This grouping of products, based on their sales behavior, would also assist managers to decide which items should be placed adjacent to each other in transportations and warehouses in order to optimize storage and transport of products.

3.4. Building Sales Forecasting Models. In this research, both linear models (ARIMA) and nonlinear models (hybrid neural networks) were applied. Their performance was also compared based on various evaluation criteria.
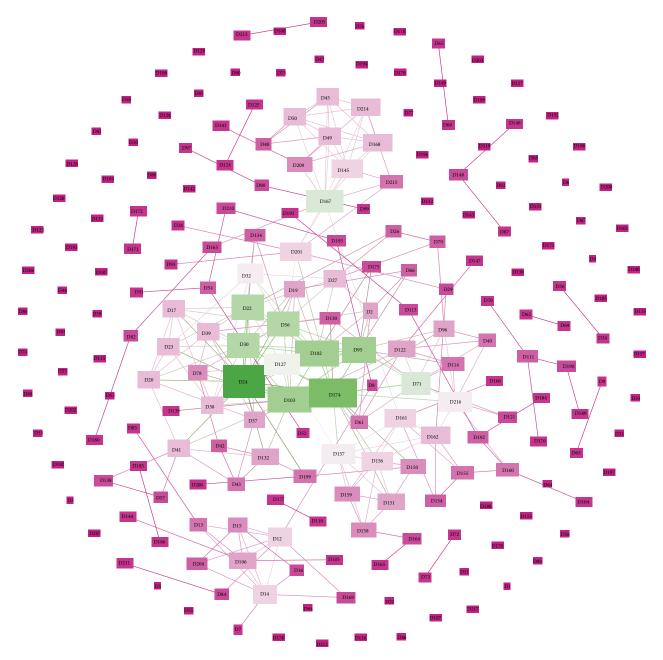
FIGURE 5: Networks of Drugs.

3.4.1. ARIMA Methodology for Time Series Forecasting. ARIMA models (p, d, q), developed by Box and Jenkins [3], have been widely used for time series forecasting. The three kinds of parameters in the model are the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q). In this research, ARIMA methodology was conducted to compare its performance with our proposed methodology. To fit ARIMA to the available time series, the following steps were executed.

(1) Transforming the series to make it stationary (finding d parameter): in order to determine the necessary level of differencing, time series plot, autocorrelogram, and partial autocorrelogram diagrams were examined. For instance, for a sample drug, named drug 24, after two times differencing and log transforming the series became approximately stationary.

(2) Identification of the initial p and q parameters: based on the visual examination of correlograms of autocorrelation, and partial autocorrelation plots, the best model for sales data is recognizable. For example, for drug 24, the best model was ARIMA (2, 2, 0). To be sure of the results of this step, the following steps were followed:

TABLE 1: Cliques sets.

Clique #	Members
1	{24, 30, 56, 95, 102, 103}
2	{19, 24, 30, 56, 102, 103}
3	{22, 24, 30, 32, 56, 102}
4	{24, 30, 32, 38, 102,}
5	{24, 30, 38, 102, 103,}
6	{24, 30, 95, 102, 103, 174}
7	{17, 20, 23, 24}
8	{17, 22, 23, 24, 30}
9	{22, 23, 24, 30, 32}
10	{23, 24, 30, 32, 38}
11	{17, 23, 24, 30, 38}
12	{17, 22, 24, 30, 56}
13	{22, 24, 37, 39}
14	{20, 24, 37, 39}
15	{22, 24, 39, 127}
16	{24, 41, 103, 132}
17	{20, 24, 41}
18	{24, 95, 102, 103, 127, 174}
19	{22, 24, 102, 127}
20	{24, 132, 199}
21	{2, 56, 95, 102}
22	{2, 27, 56, 102}
23	{2, 66, 95}
24	{12, 14, 15, 106}
25	{12, 14, 15, 204}
26	{12, 13, 14, 106}
27	{12, 14, 169}
28	{14, 15, 16}
29	{22, 26, 27}
30	{26, 27, 79}
31	{22, 27, 32, 56, 102}
32	{40, 71, 96, 216}
33	{40, 71, 122}
34	{42, 43, 132}
35	{41, 43, 132}
36	{45, 49, 50, 145, 167, 168}
37	{45, 48, 49, 50}
38	{45, 50, 145, 167, 168, 214}
39	{49, 167, 168, 209}
40	{41, 57, 138}
41	{61, 71, 122}
42	{61, 71, 174}
43	{71, 95, 96, 174, 216}
44	{71, 95, 103, 174}
45	{71, 95, 114, 216}
46	{71, 121, 216}
47	{71, 114, 122}
48	{71, 130, 174}
49	{56, 78, 103}

TABLE 1: Continued.

Clique #	Members
50	{78, 103, 127}
51	{39, 78, 127}
52	{22, 32, 56, 134}
53	{145, 167, 214, 215}
54	{145, 167, 201, 215}
55	{150, 151, 156, 157, 174}
56	{151, 156, 157, 158, 159}
57	{154, 155, 161, 162}
58	{156, 157, 161, 162}
59	{156, 157, 159, 161}
60	{127, 157, 174}
61	{167, 168, 209, 214}
62	{95, 102, 175}
63	{127, 174, 201}

TABLE 2: Four months' forecasting results for the original data.

Months	Forecast	Observed	Residual	MSE	MAE
33	2054.746	2531.000	476.254		
34	1968.661	2286.000	317.339	140793.3	369.842
35	2057.058	2384.000	326.942		
36	1969.168	2328.000	358.832		

(a) estimation of the p (autoregressive) and q (moving average) components to see if they contribute significantly to the model or they do not; (b) both the approximate maximum likelihood method according to McLeod and Sales [51] and the exact maximum likelihood method according to Melard [52] were examined and compared with all products to compute the SS (sum of squares) of residuals. In some cases, the first method had better results, in some cases the second one. For drug 24, the exact maximum likelihood had better performance. The result of this step shows that the best model for time series sales data of drug 24 was (1, 2, 0).

(3) Forecasting: to avoid overfitting and to do model evaluation, the dataset was divided into two subsets: one for training (32 months) and one for testing and performance evaluation (4 months). To evaluate the performance of the model, the values of residuals, mean squared error (MSE), and mean absolute error (MAE) of the test data were calculated. The result of forecasting for the test data is presented in Table 2.

(4) Diagnosis of the residuals: a proper assessment of the model is to plot the residuals, to check them regarding any organized trends, and to inspect the autocorrelogram of residuals. Normal probability plot of the residuals approved the normality of them. If the goodness of the model is not approved in this step, it is necessary to redo all previous steps. For time series sales data of drug 24, autocorrelation function and partial autocorrelation function of residuals showed

TABLE 3: Comembers.

Drugs	Group comembers																			
24	17	19	20	22	23	**24**	30	32	37	38	39	41	56	95	102	103	127	132	174	199
174	24	30	61	71	95	96	102	103	127	130	150	151	156	157	**174**	201	216			
95	2	24	30	56	66	71	**95**	96	102	103	114	127	174	175	216					
103	19	24	30	38	41	56	71	78	95	102	**103**	127	132	139	174					
22	17	**22**	23	24	26	27	30	32	37	39	56	102	127	134						
30	17	19	20	22	23	24	**30**	32	38	56	95	102	103	174						
71	40	61	**71**	95	96	103	114	121	122	130	174	216								
167	45	49	50	145	161	163	**167**	168	201	209	214	215								
127	22	24	39	78	95	102	103	**127**	157	174	201									
132	22	23	24	27	30	32	38	56	102	**132**	134									
12	8	**12**	13	14	15	84	106	169	204											
168	45	49	50	145	167	**168**	209	214												
37	20	22	24	**37**	39	43	122													
78	39	42	56	**78**	103	127														
61	2	**61**	71	122	174															
66	2	19	**66**	95																
84	12	**84**	211																	
124	**124**	125																		
51	**51**	54																		
31	**31**																			
86	**86**																			

that there was no serial dependency between residuals. This step confirmed the goodness of the model obtained in previous step. Although the best ARIMA model for sales data of drug 24 was chosen, the prediction accuracy was not satisfactory enough. In next section, our proposed sales forecasting methodology is explained.

3.4.2. A Hybrid Neural Network Approach for Time Series Forecasting. In this research, hybrid neural networks were carried out since it is not acceptable to apply a fully linear or fully nonlinear model on sales data. The origin of this approach was related to Zhang [40] that combined ARIMA methodology with ANNs in order to model linear parts with ARIMA and nonlinear parts with ANNs. However, the basic idea of the applied hybrid approach in this research is to let linear ANN model the linear parts and let nonlinear ANN model the nonlinear parts and then combine the results from both linear and nonlinear models. The model and its forecasting process are explained in more detail below.

The Hybrid Model. The hybrid model can be written as [40]

$$Y_t = L_t + NL_t + u_t, \tag{3}$$

where Y_t is the time series observation at time period t; L_t and NL_t are the linear and nonlinear components of the time series, respectively, and u_t is the random error term.

The model building process involved three main steps: (1) fitting a linear neural network model to the time series

under study, (2) building a nonlinear neural network model based on the residuals from the linear neural network model, and (3) combining the linear neural network prediction and the nonlinear neural network result to form the final forecast. The purposes of combining mentioned models were to take advantage of the unique modeling capability of each individual model and improve forecasting performance considerably. If the estimated component from the linear model is L_t^e and the estimated nonlinear component is NL_t^e, then the combined forecast Y_t^e will be [40]

$$Y_t^e = L_t^e + NL_t^e. \tag{4}$$

Step 1 (data selection). As stated, 217 selective medicines were grouped in 63 cliques. It should be noted that evolved cliques have some overlapping members such that a drug may be a member of two or more cliques. To build forecasting model it was needed to examine the validity of proposed methodology on different products. Thus, different drugs were extracted from various degrees and Table 3, that explicitly presents each drug with its group members, was built. This table was one of the main tools of sales prediction phase. For instance, to predict the future sales of drug 24 in addition to its own past sales records, sales records of drugs 17, 19, 20, 22, 23, 30, 32, 37, 38, 39, 41, 56, 95, 102, 103, 127, 132, 174, and 199 could be used. In contrast, drugs 31 and 86 have no comember and their sale forecasting is inevitably done by solely their past sale. However, for making a comparison, hybrid neural network models were built with two different approaches: (1) just using

TABLE 4: Forecasting results and accuracy measures of test data for drug 24 (with its own past records).

Observed test data	Predicted test data	Residual (R)	R^2	$|R|$	MSE	MAE
2531	2267.458	−263.542	69454	263.542		
2286	2318.301	32.301	1043	32.301	18695	90.568
2384	2318.552	−65.448	4283	65.448		
2328	2328.98	0.980	1	0.980		

TABLE 5: Forecasting results and accuracy measures of test data for drug 24 (with its own and its comembers' past records).

Observed test data	Predicted test data	Residual (R)	R^2	$|R|$	MSE	MAE
2531	2531.000	0.0000	0.00	0.0000		
2286	2243.943	−42.057	1768.80	42.057	442.2	10.514
2384	2384.000	0.0000	0.00	0.0000		
2328	2328.000	0.0000	0.00	0.0000		

each drug's own past records and (2) using each drug's own past records and its group members' past records as well.

Step 2 (split data into train and test set). The data set was divided into two data sets. The last 4 months' data was reserved as the holdout sample (test data) in order to be used for forecasting evaluation. The rest of the data (in-sample) was used for model selection and estimation. The in-sample data were further divided into two portions of a training sample and a validation sample. The validation sample consists of the last 4-month observations while the training sample consists of remaining observations.

Step 3 (model building (hybrid neural networks)). This step will be illustrated as follows.

Building Linear ANN. In this step, for all 21 drugs and for our two approaches, first a linear neural network model was fitted to the time series under study and the residuals were found.

Building Nonlinear ANN. In this part, ANNs were applied to the residuals (that contained some nonlinearity) from linear model to estimate the probable nonlinear components. In this step, different types of network such as linear, PNN or GRNN, radial basis function, and one-, two-, and three-layer perceptron were examined. In the experiment, various numbers of hidden units were chosen and examined to find the best network architecture. For activation functions, linear and sigmoid logistic functions were tried. Finally, for training, various learning algorithms including BP, KM, KN, CG, and PI were examined automatically.

Step 4 (combining both models). Finally, two models were combined together to make a forecast. The previous steps were carried out through 25 different neural network models with different architectures and learning algorithms. Then, the best-fitted model for each drug was extracted.

3.5. Model Evaluation. Once a model has been generated and tested, its performance should be evaluated. In this study, two forecast error measures, namely, mean squared error (MSE $= \sum_i (d_i - y_i)^2/n$) and mean absolute error

(MAE $= \sum_i |d_i - y_i|/n$) were employed for model evaluation and model comparison. Tables 4 and 5 show the best-fitted model, training algorithms, and predicted values in comparison with observed values, residuals (difference between real and predicted values), R^2, $|R|$, MSE, and MAE for drug 24. As it is obvious from these tables, the prediction performance was improved greatly when partners' past records were used as well. For instance, Table 4, which is related to sales prediction of drug 24 with its own records, shows MSE = 4094.8 and MAE = 51.5. However, Table 5, which is related to sales prediction of drug 24 with its own and comembers records, shows MSE = 1973.5 and MAE = 33 that indicates a great improvement in accuracy of the prediction.

By comparing these results with results of ARIMA model in previous section (MSE = 140793.3 and MAE = 369.8417), the differences among three methods were clearly observed. The forecasting results for drug 24 show that the novel method outperforms both examined methods considerably.

Accuracy measures of ARIMA and both ANN approaches were calculated and compared for 21 selective drugs (see Table 6). By comparing MSE measures for all three approaches, it can be observed that in 18 cases out of 21 cases, MSE measures for the second approach were smaller than those for the first method. In 17 cases out of 19 cases, MSE measures for the third approach were smaller than those for first one. In all 19 cases, the MSE measures of the third approach were smaller than those for the second approach. Moreover, by comparing MAE measures for all three approaches, in 18 cases out of 21 cases, MAE measures for the second approach were smaller than those for the first one. In 17 cases out of 19 cases, MAE measures for the third approach were smaller than those for the first one. In 17 cases out of 19 cases, the MAE measures of the third approach were smaller than those for the second approach. Furthermore, by comparison of $\overline{\text{MSE}}$ and $\overline{\text{MAE}}$ for all three approaches, it is concluded that both $\overline{\text{MSE}}$ and $\overline{\text{MAE}}$ were smaller for the third approach than those for other two approaches. It is also observed that both $\overline{\text{MSE}}$ and $\overline{\text{MAE}}$ were smaller for the second approach than those for the first one.

This research verified that by applying network analysis tools to group medicines, using records of group members

TABLE 6: Comparisons of the conducted methods.

Drugs	Number of group members	ARIMA (MSE)	ARIMA (MAE)	Hybrid neural networks by inputting past records of each drug (MSE)	Hybrid neural networks by inputting past records of each drug (MAE)	Hybrid neural networks by inputting each drug's past records and its group members' past records (MSE)	Hybrid neural networks by inputting each drug's past records and its group members' past records (MAE)
24	19	140,793.30	369.84	18,695.00	90.57	442.20	10.51
174	16	3,614,887,731.00	53,514.70	274,513,013.30	12,195.90	62,575,380.00	5,803.30
95	14	3,429,391.38	1,340.07	1,564,831.50	955.30	143,032.10	362.50
103	14	654,331.20	726.33	109,700.60	316.70	56,819.20	237.20
22	13	479,018.80	681.71	291,633.20	353.60	51,155.50	171.50
30	13	1,465,926.00	676.40	2,477,276.60	1,186.20	2,032,221.60	1,197.70
71	11	3,316.56	51.26	4,094.80	51.50	1,973.50	33.00
167	11	176,244,064.00	9,964.42	20,348,171.40	4,013.20	10,173,585.60	2,223.20
127	10	24,960.33	92.06	34,114.30	142.70	34,091.50	136.00
132	10	2,191,071.40	1,111.01	621,768.90	399.40	512.80	14.50
12	8	1,732,201.00	1,047.98	1,624,244.30	1,032.10	18,390.40	125.00
168	7	239,968,284.00	11,398.75	71,612,145.90	7,270.30	13,221,096.40	3,267.20
37	6	1,197,148,086.00	34,048.78	40,463,339.40	5,444.40	87,854.60	150.00
78	5	6,791,864.39	2,097.51	5,421,172.90	1,638.20	2,339,520.90	1,169.00
61	4	7,332,695.00	2,500.56	4,351,902.60	1,517.10	2,272,363.40	866.90
66	3	445,472,962.00	19,322.10	519,776.20	508.00	13,299.30	103.20
84	2	2,829,442.90	1,389.77	968,377.60	618.00	686,472.10	540.10
124	1	43,310.84	206.80	7,468.10	59.40	861.10	20.10
51	1	1,651,518.00	1,086.98	679,480.70	601.60	547,755.70	633.90
31	0	451,450.00	648.47	343,539.50	426.60	—	—
86	0	133,612,901.00	10,251.71	52,155,027.10	3,610.90	—	—
Mean		277,931,205.67	7,263.20	22,768,084.47	2,020.56	4,960,885.68	898.15

as input variables, and combining both linear and nonlinear models, forecasting performance can be considerably improved for two reasons: (1) grouping of products and using group members' past sales records solved the problem of not having enough past sales records for each product, (2) the hybrid approach studied in this research overcame the limitation of a pure linear or nonlinear modeling approach while at the same time took advantage of their unique modeling capability to capture different patterns in the data. In addition, it has been stated repeatedly in related literature that increasing number of past data would improve accuracy of time series forecasting results. In fact, the experimental observations in this research are a confirmation to the abovementioned theory. Consequently, evidences confirmed that the offered forecasting method (hybrid neural networks by inputting both each drug's past records and its partners' past records) performed noticeably better than both hybrid ANNs without use of only each drug's past recodes and ARIMA methodology.

4. Conclusion

According to the problem of having many new items with short number of past records, and having great diversity of medicines, common prediction methods are mainly inappropriate or ineffectual for PDCs. The basic objective of this research was to offer a novel and precise sales prediction method to help companies, especially PDC, to forecast product sale and tuning inventory management policies in order to prevent costs of excessive inventory and prevent losing customers due to drug shortage. In order to validate the proposed method, three-year monthly sales data was gathered from Pakhsh Hejrat Co. In data preprocessing phase, raw data was prepared to suit the research objectives. Subsequently, an exploratory analysis was conducted to better specify the nature of data. Next, a comprehensive graph based analysis was performed to find clique sets and group members and visualize the network of drugs.

Afterwards, sales forecasting models were built with three different approaches: (1) ARIMA methodology for time series forecasting, (2) hybrid neural network approach for time Series forecasting by means of each drug's past recodes, and (3) hybrid neural network approach for time. Series forecasting by means of each drug's past records and its group members' past records. This research revealed that by grouping products, using records of group members as input variables, and combining both linear and nonlinear models, forecasting performance can be significantly improved for two reasons: (1) grouping of products and using group members' past sales records solved the problem of not having enough past sales records for each product and (2) the hybrid approach studied in this research overcame the limitation of a pure linear or nonlinear modeling approach while at the same time took advantage of their unique modeling capability to capture different patterns in the data.

4.1. Contributions. If different contributions of researches are classified into three categories of theoretical, methodological, and empirical contribution, it could be stated that this research has methodological and empirical contributions.

As a methodological contribution, in this research, hybrid neural networks were carried out to let linear ANN model the linear components and let nonlinear ANN model the nonlinear components and then merge the results from both linear and nonlinear models. In addition, this research introduced a novel method of grouping products to make use of group members' past sales data for each other in sales prediction and increase the accuracy of the prediction. The introduced scheme outperformed two previously known methods of (1) ARIMA modeling and (2) building ANNs by just using each drug's own past records. The empirical contribution of this research can be considered from two different points of view. Firstly, a real problem and an actual company (Pakhsh Hejrat Co.) with its genuine sales data were chosen, and sales prediction results were compared with unseen sales data of this company, and acceptable outcomes were achieved. Secondly, 8 managers and experts of 4 main and famous PDCs (Pakhsh Hejrat, Armaghan Daroo, Pakhsh Ferdows, and Daroo Pakhsh) that stand for more than 70% of total sales of medicines to drug stores and hospitals in Iran were interviewed. All managers and experts were attracted by the introduced method of finding group members and sales forecasting. They strongly believed that finding groups of products with similar changes in their past sales records would also help managers to optimize ordering, storage, transport, and delivery of products. Accordingly, empirical supports were attained for the proposed methodology.

4.2. Further Research. In this research, clique sets were found based on sales cross-correlation, which only analyzed linear covariations. However, it is probable that medicines have also nonlinear covariations with each other. Thus, it is proposed to find cliques sets based on other similarities or harmonies in changes of sales behaviors to discover nonlinear relationships as well. It means that it is logical to test nonlinear relationships or even combination of linear and nonlinear relationships to have more efficient cliques and more accurate forecasting. Moreover, networks of medicines can be a good source for exploring the relationships between medicines. The roles of some medicines in the network and various types of relationships between them can be analyzed from the network analysis point of view and it will be helpful to consider therapeutic attributes of medicines in order to find groups of similar medicine. Furthermore, running and comparing other approaches such as evolving neural networks and evolving fuzzy rules will be very valuable but could not be included in the scope of this research and can be considered as a further research.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] A. Gupta, C. D. Maranas, and C. M. McDonald, "Mid-term supply chain planning under demand uncertainty: customer demand satisfaction and inventory management," *Computers and Chemical Engineering*, vol. 24, no. 12, pp. 2613–2621, 2000.

[2] P. Doganis, A. Alexandridis, P. Patrinos, and H. Sarimveis, "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing," *Journal of Food Engineering*, vol. 75, no. 2, pp. 196–204, 2006.

[3] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 2nd edition, 1976.

[4] L. Wang, H. Zou, J. Su, L. Li, and S. Chaudhry, "An ARIMA-ANN hybrid model for time series forecasting," *Systems Research and Behavioral Science*, vol. 30, no. 3, pp. 244–259, 2013.

[5] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[6] F. Alpaslan, E. Eğrioğlu, C. H. Aladağ, and E. Tiring, "An statistical research on feed forward neural networks for forecasting time series," *American Journal of Intelligent Systems*, vol. 2, no. 3, pp. 21–25, 2012.

[7] R. P. Diez and J. P. Fernández, *Forecasting the Price of Energy in Spain's Electricity Production Market*, 9th Congress of Industrial Engineering, Gijón, Spain, 2005.

[8] T. Hill, L. Marquez, M. O'Connor, and W. Remus, "Neural networks models for time series forecasts," *Management Science*, vol. 42, no. 7, pp. 1082–1092, 1996.

[9] R. Carbonneau, K. Laframboise, and R. Vahidov, "Application of machine learning techniques for supply chain demand forecasting," *European Journal of Operational Research*, vol. 184, no. 3, pp. 1140–1154, 2008.

[10] F. L. Chen and T. Y. Ou, "A neural-network-based forecasting method for ordering perishable food in convenience stores," in *Proceedings of the 4th International Conference on Natural Computation (ICNC '08)*, pp. 250–254, IEEE Computer Society, October 2008.

[11] C. Frank, A. Garg, A. Raheja, and L. Sztandera, "Forecasting women's apparel sales using mathematical modeling," *International Journal of Clothing Science and Technology*, vol. 15, no. 2, pp. 107–125, 2003.

[12] L. Yu, S. Wang, and K. K. Lai, "A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates," *Computers and Operations Research*, vol. 32, no. 10, pp. 2523–2541, 2005.

[13] 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.

[14] P.-C. Chang and Y.-W. Wang, "Fuzzy Delphi and back-propagation model for sales forecasting in PCB industry," *Expert Systems with Applications*, vol. 30, no. 4, pp. 715–726, 2006.

[15] C.-W. Chu and G. P. Zhang, "A comparative study of linear and nonlinear models for aggregate retail sales forecasting," *International Journal of Production Economics*, vol. 86, no. 3, pp. 217–231, 2003.

[16] M. Paliwal and U. A. Kumar, "Neural networks and statistical techniques: a review of applications," *Expert Systems with Applications*, vol. 36, no. 1, pp. 2–17, 2009.

[17] S. Goyal, "Artificial neural networks (ANNs) in food science—a review," *International Journal of Scientific World*, vol. 1, no. 2, pp. 19–28, 2013.

[18] R. J. Kuo, P. Wu, and C. P. Wang, "An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination," *Neural Networks*, vol. 15, no. 7, pp. 909–925, 2002.

[19] M. Adya and F. Collopy, "How effective are neural networks at forecasting and prediction? A review and evaluation," *Journal of Forecasting*, vol. 17, no. 5-6, pp. 481–495, 1998.

[20] G. Zhang, E. B. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: the state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[21] M. Nelson, T. Hill, B. Remus, and M. O'connor, "Can neural networks applied to time series forecasting learn seasonal patterns: an empirical investigation," in *Proceedings of the 27th Annual Hawaii International Conference on System Sciences*, pp. 649–655, Wailea, Hawaii, USA, January 1994.

[22] Z. Tang, C. de Almeida, and P. A. Fishwick, "Time series forecasting using neural networks vs. Box-Jenkins methodology," *Simulation*, vol. 57, no. 5, pp. 303–310, 1991.

[23] W. R. Foster, F. Collopy, and L. H. Ungar, "Neural network forecasting of short, noisy time series," *Computers and Chemical Engineering*, vol. 16, no. 4, pp. 293–297, 1992.

[24] J. L. Callen, C. Y. Kwan, C. Y. Yip, and Y. Yuan, "Neural network forecasting of quarterly accounting earnings," *Journal of Forecasting*, vol. 12, no. 4, pp. 475–482, 1996.

[25] K. B. Church and S. P. Curram, "Forecasting consumers' expenditure: a comparison between econometric and neural network models," *International Journal of Forecasting*, vol. 12, no. 2, pp. 255–267, 1996.

[26] C. Ntungo and M. Boyd, "Commodity futures trading performance using neural network models versus ARIMA models," *Journal of Futures Markets*, vol. 18, no. 8, pp. 965–983, 1998.

[27] S. Heravi, D. R. Osborn, and C. R. Birchenhall, "Linear versus neural network forecasts for European industrial production series," *International Journal of Forecasting*, vol. 20, no. 3, pp. 435–446, 2004.

[28] K. Chakraborty, K. Mehrotra, C. K. Mohan, and S. Ranka, "Forecasting the behavior of multivariate time series using neural networks," *Neural Networks*, vol. 5, no. 6, pp. 961–970, 1992.

[29] A. P. Ansuj, M. E. Camargo, R. Radharamanan, and D. G. Petty, "Sales forecasting using time series and neural networks," *Computers and Industry Engineering*, vol. 31, no. 1-2, pp. 421–424, 1996.

[30] D. Agrawal and C. Schorling, "Market share forecasting: an empirical comparison of artificial neural networks and multinomial logit model," *Journal of Retailing*, vol. 72, no. 4, pp. 383–408, 1996.

[31] D. H. F. Yip, E. L. Hines, and W. W. H. Yu, "Application of artificial neural networks in sales forecasting," in *IEEE International Conference on Neural Networks*, vol. 4, pp. 2121–2124, Houston, Tex, USA, 1997.

[32] R. J. Kuo and K. C. Xue, "An intelligent sales forecasting system through integration of artificial neural network and fuzzy neural network," *Computers in Industry*, vol. 37, no. 1, pp. 1–15, 1998.

[33] M. M. Elkateb, K. Solaiman, and Y. Al-Turki, "A comparative study of medium-weather-dependent load forecasting using enhanced artificial/fuzzy neural network and statistical techniques," *Neurocomputing*, vol. 23, no. 1–3, pp. 3–13, 1998.

[34] T. S. Gruca, R. K. Bruce, and E. A. Furr Petersen, "Mining sales data using a neural network model of market response," *SIGKDD Explorations*, vol. 1, no. 1, pp. 39–43, 1999.

[35] D. E. Rumelhart and J. McClelland, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, MIT Press, Cambridge, Mass, USA, 1986.

[36] T. L. Ainscough and J. E. Aronson, "An empirical investigation and comparison of neural networks and regression for scanner data analysis," *Journal of Retailing and Consumer Services*, vol. 6, no. 4, pp. 205–217, 1999.

[37] M. Qi, "Predicting US recessions with leading indicators via neural network models," *International Journal of Forecasting*, vol. 17, no. 3, pp. 383–401, 2001.

[38] R. J. Kuo and P. H. Cohen, "Intelligent tool wear estimation system through artificial neural networks and fuzzy modeling," *Artificial Intelligence in Engineering*, vol. 12, no. 3, pp. 229–242, 1998.

[39] T. Hill, L. Marquez, M. O'Connor, and W. Remus, "Artificial neural network models for forecasting and decision making," *International Journal of Forecasting*, vol. 10, no. 1, pp. 1–15, 1994.

[40] G. P. Zhang, *Neural Networks in Business Forecasting*, Idea Group Publishing (IGP), Hershey, Pa, USA, 2004.

[41] S. F. Crone, "Prediction of white noise time series using artificial neural networks and asymmetric cost functions," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 2460–2465, IEEE Computer Society, July 2003.

[42] L. Aburto and R. Weber, "Improved supply chain management based on hybrid demand forecasts," *Applied Soft Computing Journal*, vol. 7, no. 1, pp. 136–144, 2007.

[43] A. Aslanargun, M. Mammadov, B. Yazici, and S. Yolacan, "Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting," *Journal of Statistical Computation and Simulation*, vol. 77, no. 1, pp. 29–53, 2007.

[44] A. Jain and A. M. Kumar, "Hybrid neural network models for hydrologic time series forecasting," *Applied Soft Computing Journal*, vol. 7, no. 2, pp. 585–592, 2007.

[45] L. A. Díaz-Robles, J. C. Ortega, J. S. Fu et al., "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile," *Atmospheric Environment*, vol. 42, no. 35, pp. 8331–8340, 2008.

[46] R. J. Kuo, T.-L. Hu, and Z.-Y. Chen, "Application of radial basis function neural network for sales forecasting," in *Proceedings of the International Asia Conference on Informatics in Control, Automation, and Robotics (CAR '09)*, pp. 325–328, IEEE Computer Society, September 2009.

[47] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Systems with Applications*, vol. 37, no. 1, pp. 479–489, 2010.

[48] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Improvement in forecasting accuracy using the hybrid model of ARFIMA and feed forward neural network," *American Journal of Intelligent Systems*, vol. 2, no. 2, pp. 12–17, 2012.

[49] V. Boginski, S. Butenko, and P. M. Pardalos, "Mining market data: a network approach," *Computers and Operations Research*, vol. 33, no. 11, pp. 3171–3184, 2006.

[50] V. Boginski, S. Butenko, and P. M. Pardalos, "Statistical analysis of financial networks," *Computational Statistics and Data Analysis*, vol. 48, no. 2, pp. 431–443, 2005.

[51] A. I. McLeod and P. R. H. Sales, "An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models," *Applied Statistics*, vol. 32, no. 2, pp. 211–223, 1983.

[52] G. Melard, "A fast algorithm for the exact likelihood of autoregressive-moving average models," *Applied Statistics*, vol. 33, no. 1, pp. 104–114, 1984.