Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

Data Clustering and Fuzzy Neural Network for Sales Forecasting in Printed Circuit Board Industry

Pei-Chann Chang[1,2*], Chen-Hao Liu[2], Chin-Yuan Fan[2], Hsiao-Ching Chang[2]

[1]Department of Information Management, Yuan-Ze University,
[2]Department of Industrial Engineering and Management, Yuan-Ze University,
135 Yuan-Dong Rd., Taoyuan 32026, Taiwan, R.O.C.
iepchang@saturn.yzu.edu.tw

***Abstract*-Reliable prediction of sales can improve the quality of business strategy. This research develops a hybrid model by integrating K-mean cluster and Fuzzy Back Propagation Network (KFBPN) to forecast the future sales of a printed circuit board factory. Based on the K-mean clustering technique, the historic data can be classified into different clusters, thus the noise of the original data can be reduced and a more homogeneous region can be established for a more accurate prediction. Numerical data of various affecting factors and actual demand of the past 5 years of the printed circuit board (PCB) factory are collected and input into the hybrid model for future monthly sales forecasting. Experimental results show the effectiveness of the hybrid model when compared with other approaches.**

1. INTRODUCTION

Printed Circuit Board (PCB) production in Taiwan increased from \$2.8Bn in 1997 and peaked at \$4.5Bn in 2000. The recession in world economy affected Taiwan's production which fell to \$3.7Bn in 2001, but it rebounded to \$4.1Bn in 2003. While local production suffered from the global recession and unfavorable environment, the growth of overseas production continued, especially in China. Today, Taiwan has established one of the best PCB production supply chain and infrastructures in the world with advanced technical capabilities and an extended customer base. The emergence of China however highlights some of Taiwan's intrinsic difficulties, such as the lack of local market and rising labor and land costs, which make the future growth more challenging.

Under the reality situation of the short lifespan and high circulating rate of related electronic products of PCB, general production models cannot fulfill customers' demands effectively. Thus, how to predict customer's demand and prepare material flows in advance to reduce the cycle time has become a pressing issue to be dealt with. Furthermore, an efficient sales forecasting tool can be the key to strengthen the company's survival ability in the competitive environment. Therefore, it becomes indispensable to build a forecasting model to predict the monthly sales in PCB industry through an efficient and effective manner.

Sales forecasting is a very general topic of research. When dealing with the problems of sales forecasting, many researchers have used hybrid artificial intelligent algorithms to forecast, and the most rewarding method is the application integrating artificial neural networks (ANNs) and fuzzy theory. This method is applied by incorporating the experience-based principal and logic-explanation capacity of fuzzy theory and the capacity of memory and error-allowance of ANNs, as well as self learning by numeral data.

This research focuses on the sales forecasting of printed circuit board (PCB) and modifies the back-propagation network system (BPN), to select variables with a better and more systematic way from expert experience, with the purpose of improving the forecasting accuracy and using this information to help managers make decisions.

2. LITERATURE REVIEW

Although the traditional sales forecasting methods have been proved effective, they still have certain shortcomings. As referred in ref. [15], the new developed Artificial Intelligent (AI) models have more flexibility and can be applied to estimate the non-linear relationship, without the limits of traditional Time Series models. Therefore, more and more researchers tend to use AI forecasting models to deal with various manufacturing planning and control problems.

ANN appear to be particularly suited for financial time series forecasting, as they can learn highly non-linear models, have effective learning algorithms, can handle noisy data, and can use inputs of different kinds (see ref. [4], [5], [14], [16], [20] for a survey). Furthermore, complex non-linear models based on exponential GARCH processes [1] show similar results (in terms of out-of-sample prediction performance) to those obtained by much simpler ANN based on multi-layer perception (MLP) architectures [2].

Ref. [13] applied Back-Propagation Neural (BPN) Network to predict the stock price then determine buying and selling time for Tokyo Stock. They used six input indexes, vector curve, interest rate, New York Dow-Jones average, turnover, foreign exchange rate and a teaching data, to successfully predict the stock price. Due to the high accuracy and quick solving effect, lot of researchers adopted BPN to be a forecasting method.

Fuzzy theory has been broadly applied in forecasting. (See ref. [7], [10], [11]) Fuzzy theory is first combined with ANNs in ref.

[17], who incorporated the traditional fuzzy controller and ANNs to a network structure to proceed appropriate non-linear planning of unplanned control systems based on the relationship of input and output through the learning capacity of ANNs. Following that, many researchers started a new era of researches based on the combination of fuzzy theory and ANNs. In recent years, fuzzy theory combining with ANNs is widely applied in different areas and has many positive performances. (See ref. [6], [8], [15])

3. METHODOLOGY

There are three main stages in this research as shown in Fig.1. The first stage is the variables selection stage and this stage is to select many possible variables, which may influence PCB product monthly sales amount. In order to eliminate the unrelated variables, Stepwise Regression Analysis (SRA) was applied to choose the key variables to be considered in the forecasting model. The second stage is the data preprocessing stage; and the K-mean clustering technique would be adopted. The parameter of the clustering number k need to be determined first and this research will apply design of experiment to decide the best clustering number. The last stage is the Fuzzy BPN (FBPN) forecasting stage, which is developed to forecast the sales demand of PCB. FBPN will be described in details in the following section. Finally, FBPN will be compared with other three forecasting models, the superior model will be recommended to the decision makers. The details of each stage are described as follows.

3.1 Variable Selection Stage

In this stage, there are a set of factors which have been provided according to the recommendations from the shop managers and sales representatives. However, various factors need to be further selected in order to increase the efficiency of FBPN learning. Many researchers have applied different methods to select key factors in their forecasting system (Ref. [3], [8], [17]). In this research, the SRA method is used to determine the main factors that would influence the PCB sales amount.

3.1.1 Stepwise Regression Analysis (SRA)

Stepwise regression procedure determines the set of independent variables that most closely determine the dependent variable. This is accomplished by the repetition of a variable selection. At each of these steps, a single variable is either entered or removed from the model. For each step, simple regression is performed using the previously included independent variables and one of the excluded variables. Each of these regressions is subjected to an 'F-test'. If the variable small F value is greater than a user defined threshold (0.05), it is added to the model. When the variable large F value, is smaller than a user defined threshold (0.1), it is removed from the model. This general procedure is easily applied to polynomials by using powers of the independent variable as pseudo-independent variables. The statistical software SPSS for Windows 10.0 was used for stepwise regression analysis in this research.

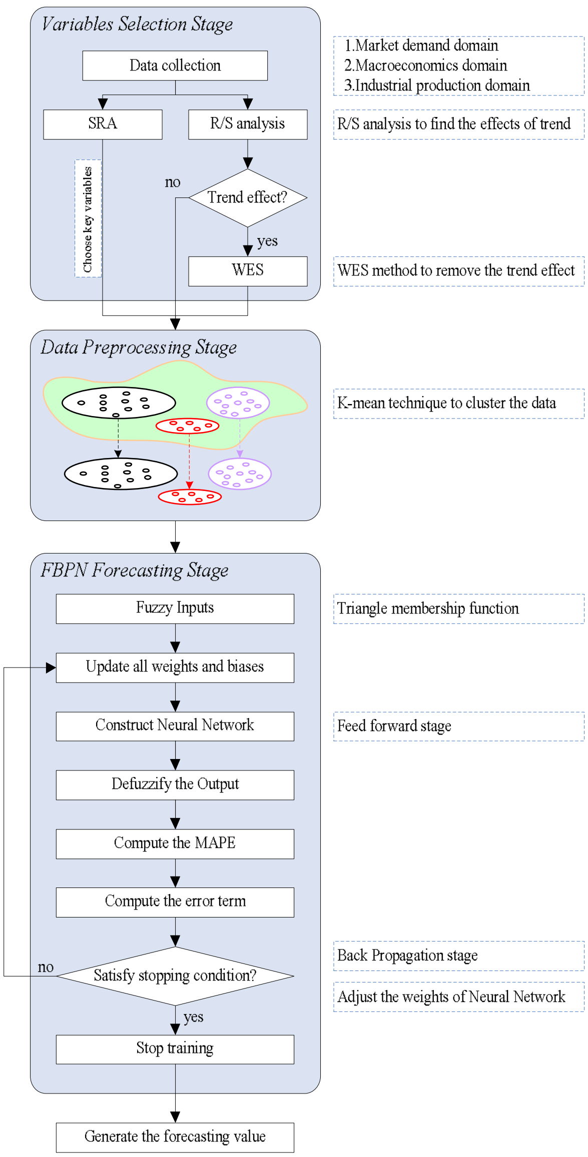

Fig. 1. Architecture of This Research

3.1.2 Winter's Exponential Smoothing (WES)

In order to take the effects of seasonality and trend into consideration, Winter's Exponential Smoothing (WES) is used to preliminarily forecast the quantity of PCB production. According to this method, three key components in the model are identified: a permanent constant component, a trend, and a seasonal component. Each component is continuously updated using a smoothing constant applied to the most recent observations and the last estimate. Ref [18], [19] compared WES with other forecasting methods, like ARIMA, and all showed that the Winter's method had a superior performance. In this research, we assume $\alpha = 0.1$, $\beta = 0.1$ and $\gamma = 0.9$.

3.2 Data Preprocessing Stage (a K-means Clustering)

The K-means clustering method [12] is probably the most well known method among researchers. The algorithm starts with k initial seeds of clustering, one for each cluster. All the n objects are then compared with each seed by means of the Euclidean distance and assigned to the closest cluster seed. The procedure is then repeated over and over again. In each stage the seed of each cluster is recalculated by using the average vector of the objects assigned to the cluster. The algorithm stops when the changes in the cluster seeds from one stage to the next are close to zero or smaller than a pre-specified value. Every object is then assigned to only one cluster.

The accuracy of the K-means procedure is very dependent upon the choice of the initial seeds. To obtain a better performance, the initial seeds should be very different among themselves. One efficient strategy to improve the K-means performance is to use, for example, the Ward's procedure first to divide the n objects into k groups and then use the average vector of each of the k groups as the initial seeds to start the K-means. As all the agglomerative clustering procedures, this method is available in a majority of statistical software.

3.3 A Fuzzy Back Propagation Network Forecasting Stage

An Artificial Neural Network (ANN) is a simplified simulation of biological neural networks in human brains. ANN is capable of "learning"; that is, it can be trained to improve its performance by either supervised or unsupervised learning. In this paper, the fuzzy sets will be adopted to transform the raw data before it is fed into the BPN. That is because each factor in the PCB industry can be further fuzzified into different terms and these terms are very often used in the factory such as "small", "medium" and "large". In addition, the triangle membership function is used to fuzzify the raw data.

The BPN and the supervised learning, i.e., learned by samples, are chosen to train the forecasting process. After learning (or training), the trained weight can be used for the prediction of future occurrences. The BPN is an ANN using back-propagation algorithm and is one of the popular ANNs, which has been widely applied to many scientific and commercial fields for non-linear analysis and prediction. The structure of BPN contains three layers: input, hidden, and output layers as shown in fig. 2.

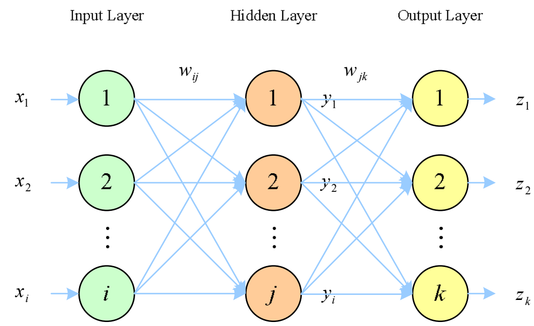

Fig. 2. The Structure of Back-Propagation Neural Network

Each layer contains I, J and K nodes denoted respectively by circles. The node is also called neuron or unit. The circles are connected by links, denoted by arrows in fig. 2, each of which represents a numerical weight. The w_{ij} is denoted as numerical weights between input and hidden layers and so is w_{jk} between hidden and output layers as also shown in fig. 2. The processing or the computation is performed in each node in the hidden and output layers. As for the number of layers and number of nodes, they will be further decided using design of experiment.

The back-propagation learning algorithm is composed of two procedures: (a) a feed forward step and (b) a back propagation weight training step. These two separate procedures will be explained in detailed as follows:

(a) Feed Forward

Assume that each input factor in the input layer is denoted by x_i, y_j and z_k represent the output in the hidden layer and the output layer, respectively. And, y_j and z_k can be expressed as follows:

$$y_j = f(X_j) = f(w_{oj} + \sum_{i=1}^{I} w_{ij} x_i) \quad (1)$$

and

$$z_k = f(Y_k) = f(w_{ok} + \sum_{j=1}^{J} w_{jk} y_j) \quad (2)$$

where the w_{oj} and w_{ok} are the bias weights for setting threshold values, f is the activation function used in both hidden and output layers, and X_j and Y_k are the temporarily computing results before applying activation function f. In this study, a sigmoid function (or logistic function) is selected as the activation function. Therefore, the actual outputs y_j and z_k in hidden and output layers, respectively, can be also written as:

$$y_j = f(X_j) = \frac{1}{1+e^{-X_j}} \tag{3}$$

and

$$z_k = f(Y_k) = \frac{1}{1+e^{-Y_k}} \tag{4}$$

The activation function f introduces the non-linear effect to the network and maps the result of computation to a domain (0, 1). This sigmoid function is differentiable. The derivative of the sigmoid function in Eq. (3, 4) can be easily derived as:

$$f' = f(1-f) \tag{5}$$

(b) Back Propagation Weight Training

The error function is defined as:

$$E = \frac{1}{2}\sum_{k=1}^{K} e_k^2 = \frac{1}{2}\sum_{k=1}^{K}(t_k - z_k)^2 \tag{6}$$

where t_k is a predefined network output (or desired output or target value) and e_k is the error in each output node. The goal is to minimize E so that the weight in each link is accordingly adjusted and the final output can match the desired output. To get the weight adjustment, the gradient descent strategy is employed. In the link between hidden and output layers, computing the partial derivative of E with respect to the weight w_{jk} produces

$$\begin{aligned}\frac{\partial E}{\partial w_{jk}} &= \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial Y_k}\frac{\partial Y_k}{\partial w_{jk}} = -e_k\frac{\partial f(Y_k)}{\partial Y_k}y_j \\ &= -e_k f'(Y_k)y_j = -\delta_k y_j\end{aligned} \tag{7}$$

where

$$\delta_k = e_k f'(Y_k) = (t_k - z_k) f'(Y_k) \tag{8}$$

The weight adjustment in the link between hidden and output layers is computed by

$$\Delta w_{jk} = \alpha \cdot y_j \cdot \delta_k \tag{9}$$

where α is the learning rate, a positive constant between 0 and 1. The new weight herein can be updated by the following

$$w_{jk}(n+1) = w_{jk}(n) + \Delta w_{jk}(n) \tag{10}$$

where n is the number of iteration. Similarly, the error gradient in links between input and hidden layers can be obtained by taking the partial derivative with respect to w_{ij}

$$\frac{\partial E}{\partial w_{ij}} = \left[\sum_{k=1}^{K}\frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial Y_k}\frac{\partial Y_k}{\partial y_j}\right]\cdot\frac{\partial y_j}{\partial X_j}\cdot\frac{\partial X_j}{\partial w_{ij}} = -\Delta_j x_i \tag{11}$$

where

$$\Delta_j = f'(X_j) = \sum_{k=1}^{K}\delta_k w_{jk} \tag{12}$$

The new weight in the hidden-input links can be now corrected as:

$$\Delta w_{ij} = \alpha \cdot x_i \cdot \Delta_j \tag{13}$$

and

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n) \tag{14}$$

Training the BPN with many samples is a very time-consuming task. The learning speed can be improved by introducing the momentum term η. Usually, η falls in the range [0, 1]. For the iteration n, the weight change Δw can be expressed as

$$\Delta w(n+1) = \eta \times \Delta w(n) + \alpha \times \frac{\partial E}{\partial w(n)} \tag{15}$$

4. Experimental Results

The data in this research are taken from an electronic company in Taiwan from 1999/1 to 2003/12. Monthly sales amount is considered as an objective of the forecasting model. The variations of the historical monthly sales data from the subject PCB Company are shown in Fig.3.

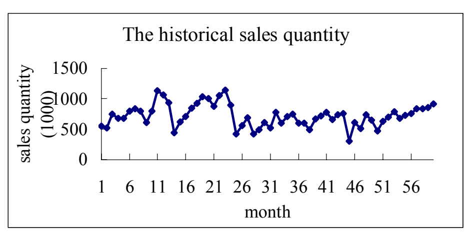

Fig. 3. Variations of the Historical Monthly Sales in Taiwan PCB Company

The main purpose of the data clustering is to reduce the effect of data noise. Therefore, the data set can be separated into a more homogeneous sub data, a new data point to be forecasted then can be clustered into a specific group according to these input factors. Thus, the FBPN model can have a better forecasting accuracy after data clustering.

K-means is applied to cluster the set of sales data from the PCB Company and the clustering results are illustrated in Figure 4. Although the number of clusters is a key factor to be decided, there is no any theory which can exactly decide and solve the data clustering problem. In the preliminary test, different clusters are tested to see the effect of the data clustering after combining with other forecasting models. Therefore, two, three and four clusters for these 60 historic data are presented. The more the number of clusters is, the better the forecasting accuracy will be. However, the total number of historic data is only 60 so it is not necessary to cluster these data into too many clusters. Therefore, to keep the required accuracy, the number of clusters is divided between 2 to 4.

The final number of clusters will influence the forecasting results; therefore, the decision of the number of clusters will be further discussed in the next section.

The proposed KFBPN (with different clustering numbers) for sales forecasting in PCB industry will be compared with other traditional methods such as Winter's Exponential Smoothing (WES), BPN, and FBPN. Furthermore, the configuration of the KFBPN is established after certain experimental runs and they are listed as follows:

- number of neurons in the input layer: 5
- number of neurons in the output layer: 1
- single hidden layer
- number of neurons in the hidden layer: 5
- network-learning rule: delta rule
- transformation function: sigmoid function
- learning rate: 0.1
- learning times: 30000

Mean average percentage error (MAPE) is applied as a standard performance measure for all four different models. After the intensive experimental tests, the MAPEs of four different models are 9.18%, 6.03%, 3.51%, and 2.19% (as shown in Table 1). Among that, the WES has the largest errors, and then BPN, FBPN, and the least is KBPN with 4 clusters.

TABLE 1
Comparisons among Four Different Forecasting Models

Method	MAPE	Improvement Rate
WES	9.18%	-
BPN	6.03%	34.31%
FBPN	3.51%	61.76%
KFBPN(k=2)	3.05%	66.78%
KFBPN(k=3)	2.51%	72.66%
KFBPN(k=4)	2.19%	76.14%

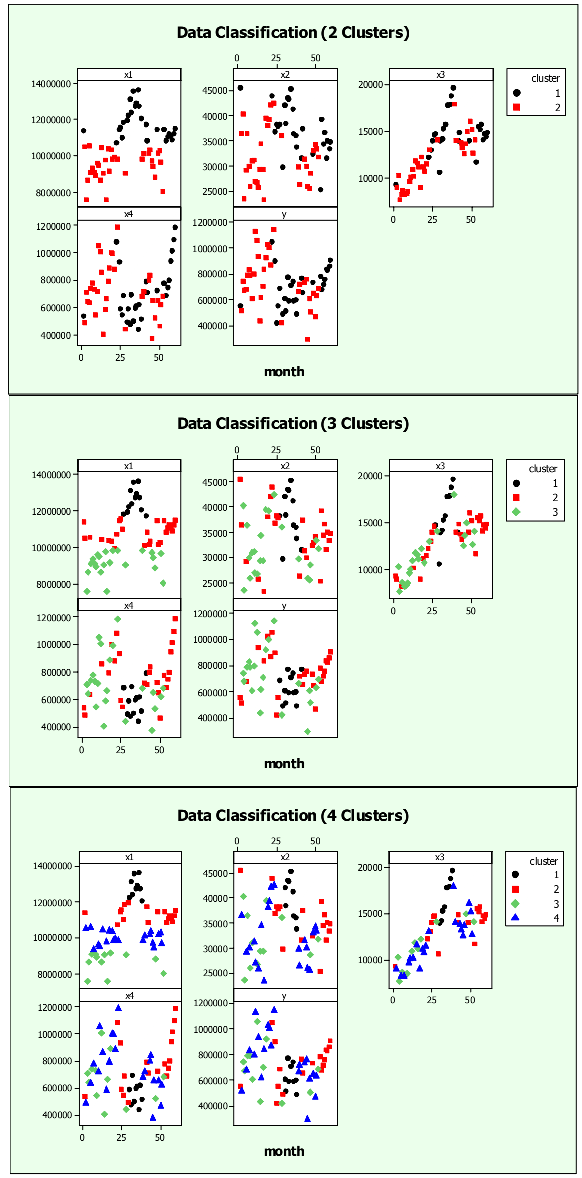

Fig. 4. The different clustered results by using K-means

As can be seen in Fig 5, the forecasting from WES has a very significant up and down tendency in the end segment of the graph. Thus the overall MAPE is around 9%. Traditional BPN model is in a stable situation and the overall MAPE is smaller than WES and it is around 6%. The same situation exist for FBPN, it has a small MAPE around 3.5%. The proposed models by integrating data clustering technique and FBPN performs very well in the end since it is very close to the real data. The MAPE of KFBPN is around 2.19%. Therefore, KFBPN performs the best among others.

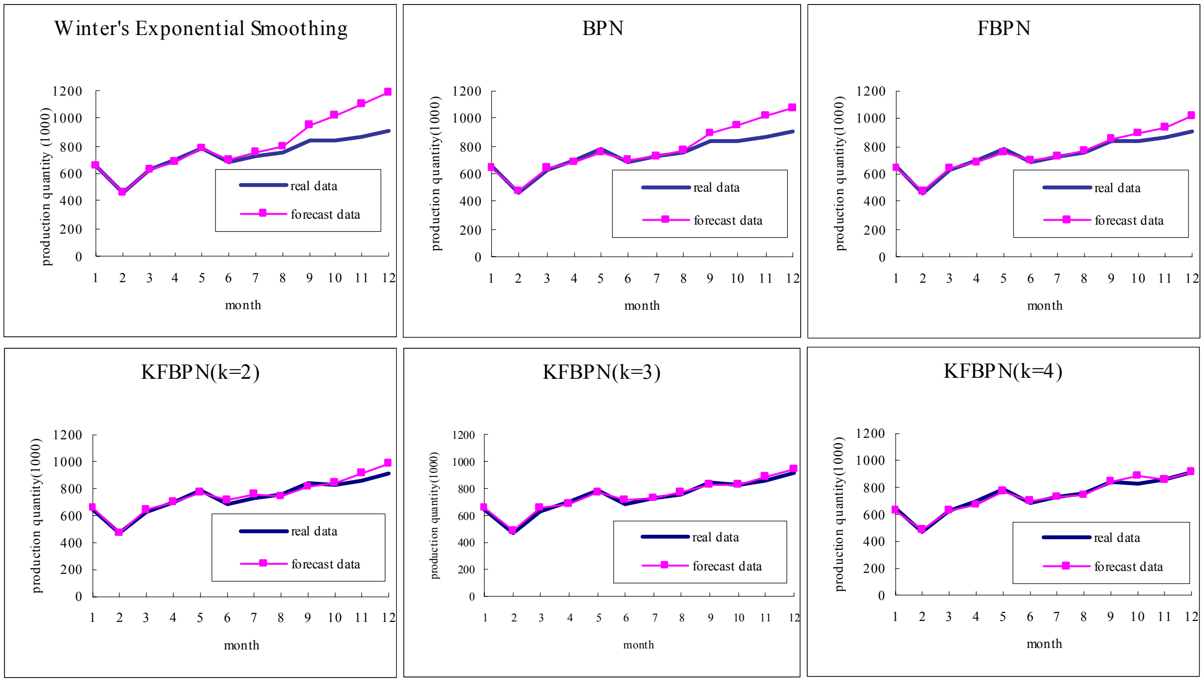

Fig. 5. The MAPE of each Different Forecasting Model

According to the various criteria, i.e., encompassing test, MAPE, and forecasting accuracy, the best model among these four different models is KFBPN with a MAPE of 2.19%. Therefore, we can claim that by combining the k-mean clustering and FBPN, the hybrid model can be applied in forecasting the sales of PCB industry and the result is very convincing and deserve further investigation in the future for other application areas.

Although, WES is very powerful as mentioned by many researchers when the data is very scarce, they even claimed that with only four data points the model can be applied to forecast the future result. However, after intensive experimental tests, the methods did not perform very well especially for those non-linear and highly dynamic data. K-mean fuzzy back-propagation model can reduce the noisy effect from the original data; therefore, this hybrid model can really reduce the forecasting errors and perform much better than other models.

5. CONCLUSIONS

The experimental results in section 4 demonstrated the effectiveness of the KFBPN that is superior to other traditional approaches. The KFBPN approach also provides another informing tool to the decision maker in PCB industries. In summary, this research has the following important contribution in the sales forecasting area and these contributions might be interested to other academic researchers and industrial practitioners:

- Clustering data into small sub groups tends to reduce the noise and forms a more homogeneous sub data. Therefore, a better forecasting accuracy can be reached when integrating these clustered data with other forecasting models.
- When taking tendency effect into consideration, the overall errors of a forecasting model are decreased. Tendency and seasonality are included in the time series data and these two factors will affect the accuracy of the forecasting method dramatically. This research applies the "Winter's trend and seasonality exponential smoothing model" to forecast the sales and then classify this data as an input to the FBPN model.
- By observing the numerical results, data preprocessing and data clustering have significant effect on performance improving.

This research applies three different performance measures, i.e., encompassing test, forecasting errors and accuracy of forecasting to compare the KFBPN with other methods, i.e., WES, BPN and FBPN. The intensive experimental results show the following: 1. In encompassing test, KFBPN and FBPN models are superior to WES and BPN. 2. As for MAPE, KFBPN has the smallest MAPE and it is only 2.19%. Therefore, KFBPN model by combining K-mean cluster and fuzzy BPN model is a very powerful and effective forecasting tool.

REFERENCES

[1] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedaticicty," *Journal of Econometrics,* vol. 52, pp. 307-327, 1986.

[2] J.Y. Campbell, A.W. Lo, A.C. MacKinlay, "The Econometrics of Financial Markets," Princeton University Press, Princeton New Jersey, 1997.

[3] P.T. Chang, L.C. Huang, H.J. Lin, "The Fuzzy Delphi Method via Fuzzy Statistics and Membership Function Fitting and An Application to The Human Resources," *Fuzzy Sets and Systems*, vol. 112, pp. 511-520, 2000.

[4] P.C. Chang, Y.W. Wang, C.H. Liu, "Fuzzy Back-Propagation Network for PCB Sales Forecasting," *Lecture Notes in Computer Science*, vol. 3610, pp. 364-373, 2005.

[5] P.C. Chang, Y.W. Wang, "Fuzzy Delphi and Back-Propagation Model for Sales Forecasting in PCB Industry," *Expert Systems with Applications*, vol. 30, pp. 715-726, 2006.

[6] T. Chen, "A Fuzzy Back Propagation Network for Output Time Prediction in a Wafer Fab," *Applied Soft Computing,* pp. 211-222, 2003.

[7] T. Chen, M.J.J. Wang, "Forecasting Methods using Fuzzy Concepts," *Fuzzy Sets and Systems*, vol. 105, pp. 339-352, 1999.

[8] P.K. Dash, A.C. Liew, S. Rahman, "Peak load forecasting using a fuzzy neural network," *Electric Power Systems Research*, vol. 32, pp. 19-23, 1995.

[9] C.C. Hsu, C.Y. Chen, "Applications of Improved Grey Prediction Model for Power Demand Forecasting," *Energy Conversion and Management*, vol. 44. pp. 2241-2249, 2003.

[10] K. Huarng, "Heuristic models of fuzzy time series for forecasting," *Fuzzy Sets and Systems*, vol. 123, pp. 369-386, 2001.

[11] J.R. Hwang, S.M. Chen, C.H. Lee, "Handling forecasting problems using fuzzy time series," *Fuzzy Sets and Systems*, vol. 100, pp. 217-228, 1998.

[12] R.A. Johnson, D.W. Wichern, "Applied Multivariate Statistical Analysis," Prentice-Hall, New Jersey, 2002.

[13] T. Kimoto, K. Asakawa, "Stock Market Prediction System with Modular Neural Network," *IEEE International Joint Conference on Neural Network*, pp. 1-6, 1999.

[14] H.M. Krolzig, J. Toro, "Multiperiod Forecasting in Stock Markets: A Paradox Solved," *Decision Support Systems*, vol. 37, pp. 531-542, 2004.

[15] R.J. Kuo, K.C. Xue, "A Decision Support System for Sales Forecasting through Fuzzy Neural Networks with Asymmetric Fuzzy Weights," *Decisions Support Systems*, vol. 24, pp. 105-126, 1998.

[16] W. Leigh, R. Purvis, J.M. Ragusa, "Forecasting the NYSE Composite Index with Technical Analysis, Pattern Recognizer, Neural Network, and Genetic Algorithm: A Case Study in Romantic Decision Support," *Decision Support Systems*, vol. 32, pp. 361-377, 2002.

[17] C.T. Lin, C.-S.G. Lee, "Neural-Network-Based Fuzzy Inference Systems," *IEEE Transaction on Computer*, vol. 40, pp. 1320-1336, 1991.

[18] J.T. Luxhoj, J.O. Riis, B. Stensballe, "A Hybrid Econometric Neural Network Modeling Approach for Sales Forecasting," *The International Journal of Production Economics*, vol. 43, pp. 175-192, 1996.

[19] T.C. Mills, "Time Series Techniques for Economists," Cambridge University Press, 1990.

[20] A.P.N. Refenes, A.N. Burgess, Y. Bentz, "Neural Networks in Financial Engineering: A Study in Methodology," *IEEE Transactions on Neural Networks*, vol. 8, pp. 1222-1267, 1997.