A Hybrid Model On Data Clustering And Computational Intelligence For Bank Crisis Classification And Prediction

Chiung-Hua Huang [1,3], Pei-Chann Chang[2], Chi-Yang Tsai[1]

[1,] *Department of Industrial Engineering and Management, Yuan-Ze University,Taoyuan, Taiwan*
[2,] *Department of Information Management, Yuan-Ze University, Taoyuan ,Taiwan*
[3,] *Department of Industrial Engineering and Management, Ta-Hwa University of Science and Technology, Hsinchu , Taiwan*

Abstract : *The prediction of bank failures is an important academic topic of* which many have used *artificial intelligence methods to build an early warning system for this purpose. The objective of this study is to enhance the accuracy in predicting bank failures by proposing two hybrid models. In this research, two hybrid models are developed by integrating a K-means cluster method to integrate K-means with a Back-Propagation Neural Network (BPN) and a Support Vector Machines (SVM) technique for financial data classification. Datasets from the website of Federal Reserve Bank of Chicago are employed for benchmark test. Initially a K-mean clustering method is applied to preprocess the dataset thus a more homogeneous data within each cluster will be attainted. A clustering method is employed to separate the case library into smaller clusters; and lastly, a Back-Propagation Neural Network (BPN) and a Support Vector Machine (SVM) model are established and prediction results are being generated. The average forecasting accuracy for bank failures of Kmeans-BPN model is 92.79% and Kmeans-SVM model is 92.43%. In comparison to other methods, the proposed model outperforms other prediction models as the prediction accuracy of bank failures are being enhanced while it simultaneously produces valuable information for business owners and investors.*
Keywords - *Data Clustering, K-means, Artificial Intelligence, Back-Propagation Neural Network, Support Vector Machine*

I. INTRODUCTION

In recent years, there were many cases for bank failure and financial crisis, the US Subprime Lending makes many banks into bankruptcy and enterprises reduce the staff. Therefore, describing to have an early warning system has already become a hot issue. Many academic researchers have used a wide variety of approaches, from adjusting financial reports and statistical methods, to searching for effective financial variables and ratios in order to predict the financial crisis[1,2]. Previous researches were based on statistical theories [3-6] to develop warning systems for financial crisis. However, previous prediction methods that were usually based on statistics analysis need to establish relationships between numerous variables and use complex mathematics and regression models, which could be time-consuming. The field of Artificial Intelligence (AI) includes artificial neural network [7], fuzzy logic [8-13], fuzzy support vector machine [14], and data mining [15] and so on. Neural networks (NN) can be alternative methodologies for classification problems to which traditional statistical methods have long been applied. NN have shown to have better predictive capabilities than other statistical methods or rule-based systems to predict the probability of business failure [16-18].Shin, Lee and Kim respectively studied the prediction ability and stability of SVM and Back-Propagation Neural Network (BPN) [19] models. The sample of this research was taken from 2320 enterprises of Korea Credit Guarantee Fund during 1996 to 1999, including 1160 healthy firms and 1160 crisis firms. In comparing SVM with BPN, the ratio design of research samples was divided into 5 training samples: 80%, 20%, 10%, 5% and 2.5%, with a test sample set as 20% of the full-size sample. To inspect the stability of the two models, the prediction ability of the training sample, while its ratio decreasing, was used to verify the prediction ability of the test sample.

Clustering is an important unsupervised classification technique. When used on a set of objects, it helps identify some inherent structures present in the objects by classifying them into subsets that have some meaning in the context of a particular problem. Less information is obtained but the ability to deal with large number of instances is improved. There have also been numerous hierarchical clustering algorithms proposed, that shows the relationship between all the instances [20]. For hierarchical clustering, there has no single clustering that is optimal, although one may ask the question if it is possible to find a hierarchy that is optimal at every level and extend to known the partitioned clustering results [21]. K-means is a representative method of the partitioning algorithms. K-means clustering methods especially the most used iterative partitioning clustering algorithm [22].

The K-means algorithm, starting with k arbitrary cluster centers, partitions a set of objects into k subsets. It is the most popular and widely used clustering techniques because it is easy to implement and very efficient, with linear time complexity [23]. According to the motivation discussed above, the objective of this study is to use financial variables with a proposed novel model to integrate K-means clustering with Back-Propagation Neural Network (BPN) and Support Vector Machines (SVM) technique to increase the accuracy of the prediction of bank failure. Forecasting researchers have taken combine clustering and other forecasting method to improve classification efficiency.

II. RESEARCH METHODOLOGY

This section reviews the methods in the areas of Clustering methods and Classification Models, respectively.

2.1 K-means Algorithm

The K-means method is a non-hierarchical clustering technique in which the dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. The detailed procedures of K-means Algorithm are described as follows：

SE（Squared-Error）is defined as:

$$SE = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \qquad (1)$$

where p is the data point; C_i is the i-th cluster; m_i is the mean of all data points in C_i; k is the total clusters number, and its mean value is defined as:

$$m_i = \sum_{j=1}^{s_i} t_{ij} \Big/ s_i \qquad (2)$$

Where t_{ij} is the j-th data point in cluster i; s_i is the total number of data points

2.2 Back-Propagation Neural Network in predicting model

The *Back-Propagation Neural Network* (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 Fig1. 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., each of which represents w_{jk} a numerical weight. The w_{ij} is denoted as numerical weights between input and hidden layers and so is between hidden and output layers as also shown in Fig. 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:
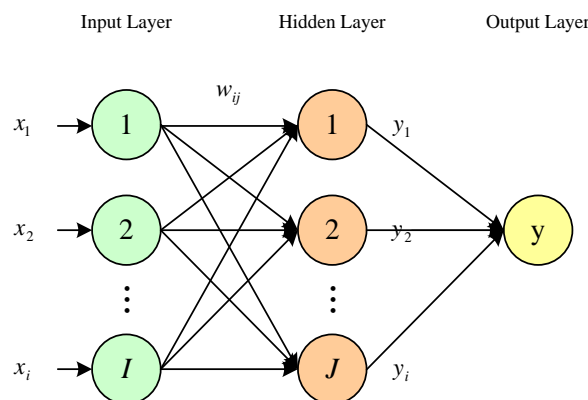

Figure. 1 The structure of BPN

2.2.1 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:

$$y_j = f(X_j) = f\left(w_{oj} + \sum_{i=1}^{I} w_{ij} x_i\right) \tag{3}$$

and

$$z_k = f(Y_k) = f\left(w_{ok} + \sum_{j=1}^{J} w_{jk} y_j\right) \tag{4}$$

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:

$$y_j = f(X_j) = \frac{1}{1 + e^{-X_j}} \tag{5}$$

and

$$z_k = f(Y_k) = \frac{1}{1 + e^{-Y_k}} \tag{6}$$

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 Equation (3, 4) can be easily derived as:

$$f' = f(1 - f) \tag{7}$$

2.2.2 Back Propagation Weight Training

The error function is defined as:

$$E = \frac{1}{2} \sum_{k=1}^{K} e_k^2 = \frac{1}{2} \sum_{k=1}^{K} (t_k - z_k)^2 \tag{8}$$

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

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial Y_k} \frac{\partial Y_k}{\partial w_{jk}} = -e_k \frac{\partial f(Y_k)}{\partial Y_k} y_j = -e_k f'(Y_k) y_j = -\delta_k y_j \tag{9}$$

where

$$\delta_k = e_k f'(Y_k) = (t_k - z_k) f'(Y_k) \tag{10}$$

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

$$\Delta w_{jk} = \alpha \cdot y_j \cdot \delta_k \tag{11}$$

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

$$w_{jk}(n+1) = w_{jk}(n) + \Delta w_{jk}(n) \tag{12}$$

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}

$$\frac{\partial E}{\partial w_{ij}} = \left[\sum_{k=1}^{K} \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial Y_k} \frac{\partial Y_k}{\partial y_j}\right] \cdot \frac{\partial y_j}{\partial X_j} \cdot \frac{\partial X_j}{\partial w_{ij}} = -\Delta_j x_i \tag{13}$$

where

$$\Delta_j = f'(X_j) = \sum_{k=1}^{K} \delta_k w_{jk} \tag{14}$$

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

$$\Delta w_{ij} = \alpha \cdot x_i \cdot \Delta_j \tag{15}$$

and

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n) \tag{16}$$

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

$$\Delta w(n+1) = \eta \times \Delta w(n) + \alpha \times \frac{\partial E}{\partial w(n)} \tag{17}$$

2.3 Support vector Machine in predicting model

This research has obtained the classified results of each given datum: output value of 1 represents crisis companies, and 0 represents healthy companies. Apply the clustered information obtained from CBR along with corresponding properties of each datum to set in SVM to establish the prediction model for each classified group respectively. Finally, examine the model for its accuracy and the prediction ability. The purpose of SVM is to find the optimal margin hyper-plane separating the two different-distance classes as the prediction model from the clustered training data. Although in the high dimension feature space, it can diminish the precision errors caused by computation and additionally differentiate different clusters to classify the data which have not yet been grouped. When the SVM classification technique is used to predict enterprise financial crises, non-liner classification problems often occur because more information on financial ratios comes with more input variables as well. Therefore, the kernel function mapping is needed for data conversion to make SVM classification more precise. Radial basis function (RBF) kernels are popular for classifying non-linear and high dimension data. It is easier to operate while possessing higher prediction ability. The data is limited between [0, 1], which can minimize the complexity of computation and time-consuming problems. Consequently, RBF kernels are applied in this research to set the initial state for non-linear data conversion to establish the prediction model of SVM.

$$K(x_i \cdot x_j) = e^{-\|x-y\|^2/2\sigma^2}, \text{ where } \sigma^2 \text{ is user defined.} \tag{18}$$

Due to the overlapping data problems on the boundary, when non-linear data are mapping converted by the kernel function, it is possible not to find the best Separating Hyperplane. Therefore, a deviation value ξ is needed to Support Hyperplane equations.

$$\begin{aligned} (w^T \cdot x) + b > 1 - \xi_i & \text{ if } y_i = +1 \\ (w^T \cdot x) + b > -1 + \xi_i & \text{ if } y_i = -1 \end{aligned} \tag{19}$$

Parameter $\xi_i \geq 0$ is the deviation value for the training data. The smaller the range of ξ_i the better. Based on the deviation of the whole mechanism, the cost of punishment for this mechanism model is defined as follows:

$$\cos t = C \sum_i \xi_i \tag{20}$$

C：The weight set for the cost of punishment in this mechanism model.

The larger weight C indicates the fewer errors made in classifying the training data; the smaller weight C indicates the larger of the maximum boarder M. Therefore, new equations are obtained after the linear SVM model is modified as follows:

$$\text{minimize} \quad \frac{1}{2}\|w\|^2 + C\sum_i \xi_i \tag{21}$$

$$y_i\left[(w^T \cdot x) + b\right] - 1 + \xi_i \geq 0, \ \forall i, \xi_i \geq 0 \tag{22}$$

Use the Lagrange multiplier approach, add constraints to the previous equations and reduce them into a quadratic equation：

$$L(w,b,\xi,\alpha,\mu) = \frac{1}{2}\|w\|^2 + C\sum_i \xi_i - \sum_{i=1}^{N} \alpha_i\left[y_i(w^T x_i + b) - 1 + \xi\right] - \sum_{i=1}^{N} \mu_i \xi_i \tag{23}$$

By partial differentiation of w, b and ξ, the parameter that minimizes L is found, then the parameter is brought into equation 3.13 to obtain new Karush-Kuhn-Tucker (KKT) Conditions to modify the errors of conversion:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{24}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{25}$$

$$\frac{\partial L}{\partial \xi} = 0 \rightarrow C - \alpha_i - \mu_i = 0 \tag{26}$$

Constraint: $y_i\left[\left(w^T \cdot x\right)+b\right]-1+\xi_i \geq 0$ (27)

Lagrange multiplier condition: $\alpha_i \geq 0, \mu_i \geq 0$ (28)

Complementary Slackness: $\alpha_i\left[y_i\left(w^T x_i + b\right)-1+\xi_i\right]=0, \mu_i\xi_i = 0$ (29)

Bring new KKT Conditions into training data to search for Support Hyperplane. Find value b through collecting these support vectors, and get the best Separating Hyperplane with minimum error as the base of SVM. Finally, classify the data that have not been classified in order to assist investors in this fluctuating stock market to reach an accurate prediction.

III.   RESEARCH PROCEDURE & DATA

By according to the discussion in section 2.1, the k-means cluster technique is adopted in this study. In this section, the data set will be clustered into k subgroups; then each subgroup will generate to BPN and SVM.

3.1 K-means Algorithm Process

a) Input Value: K: clusters number and n: dataset amount

b) Output Value: K clusters dataset obtained by SE

c) Clustering process:

(1) Choose randomly from the dataset a point K as the initial data point.

(2) Calculate the distance from the data point to each cluster. If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster. Then recalculate the data point of each cluster.

(3) Repeat the above step until a complete pass through all the data points' results in no data point moving from one cluster to another or until the termination condition is fulfilled.

3.2 Performance evaluation of prediction accuracy

After the prediction model is established, the test data are brought into BPN and SVM to calculate the Hit Ratio of this prediction model. Equations are defined as follows:

$$\textbf{Hit Ratio} = \frac{\sum_{i=1}^{m} x_i}{n} \times \textbf{100\%}$$

where, m ：Clusters

x_i ：Number of accurate prediction for the ith cluster (21)

n ：Number of test data

T he higher value of the ratio indicates the better the accuracy of the model. If it is the other way around, it means the model needs to be improved and the factors for the model are re-pondered, or, other prediction models or financial ratios that are more feasible to apply should be considered for replacement.

3.3   Data source and Research samples

This research in predicting financial crisis also drew samples from Federal Reserve Bank of Chicago. The sample ratio between distressed and healthy companies was also 1:6, and the scope of the research concerned with data within one year and two year prior to the occurrence of the financial crises. The research samples are distinguished by dichotomy where 1 represents the crisis company while 0 represents the healthy company. In this study, the crisis company is regarded as an altered trading method category to full-cash delivery method category. The data set is arbitrarily split into two subsets for a training set and testing set. The numbers of training and test samples are shown in Table 1.

Table 1. Numbers of training and testing samples

One year prior period to the financial crisis			
Samples ：Period	Crises	Non-crises	Sum
Training：1987～1992	180	1094	1274
Testing：1992～2008	58	373	431
Total firms	238	1467	1705
Two years prior period to the financial crisis			
Samples ：Period	Crises	Non-crises	Sum
Training：1986～1991	209	1227	1436
Testing：1991～2007	69	389	458
Total firms	278	1616	1894

3.4 Research variables

The financial ratios used in the bank crisis prediction application are extracted from the call Reports, which are downloaded from the website of Federal Reserve Bank of Chicago. There are nine financial ratios selected in this research are based on the six properties of the prospectus are shown in Table 2.

Table 2. Nine financial ratios of research variables

| Category	Name	Variables	Financial ratios
Capital adequacy	CAPADE	X1	Average total equity capital(3210)/average total assets(2170)
Asset (loan) quality	OLAQLY	X2	Average (accumulated)loan loss allowane(3123)/average total loans and leases,gross(1400)
Asset (loan) quality	PROBLO	X3	Average(accumulated)loans 90+days late(1407)/average total loans & leases,gross(1400)
Asset (loan) quality	PLAQLY,	X4	(annual) loan loss provision(4230)/average total loans and leases, gross(1400)
Management	NIEOIN	X5	Non-interest expense(4093)/operating income(4000)
Earning	NINMAR	X6	Total interest income(4107)-interest expense(4073)/average total asset(2170)
Earning	ROE	X7	((net incomeafer tax)(4340)+applicable income taxes(4302))/average total equity capital(3210)
Liquidity	LIQUID	X8	(average cash(0010)+average federal funds sold(1350))/(average total deposit(2200)+average fed funds purchased(2800)+average banks' liability on acceptance(2930)+average other liabilities(2930))
Miscellaneous	GROWLA	X9	Total loans & leases,gross(1400)$_t$ –total loans and leases, gross(1400)$_{t-1}$/total loans and leases,gross(1400)$_{t-1}$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

With regard to the prediction accuracy of financial crises for Kmeans-BPN, Kmeans-SVM, BPN, SVM and Bayes, the proposed Kmeans-BPN and Kmeans-SVM models outperforms other prediction methods with more detailed information provided for one, and two years before the financial crisis as follows:

4.1 One year prior to the financial crisis

The Kmeans-BPN method in this research clusters the data into 2 groups with the prediction accuracy of 92.79%. Whereas in the Kmeans-SVM method, Kmeans clusters the data into 6 groups with the prediction accuracy of 92.43%. As for the BPN method without clustering, the prediction accuracy is 91.55%. As for the SVM method without clustering, the prediction accuracy is 91.14%. Bayes is based on statistical theories with the prediction accuracy of 89.33%. The comparison and the prediction accuracy of these five methods are shown in Figure 2.

Figure 2. Comparison for the prediction accuracy of one year prior to the financial crisis
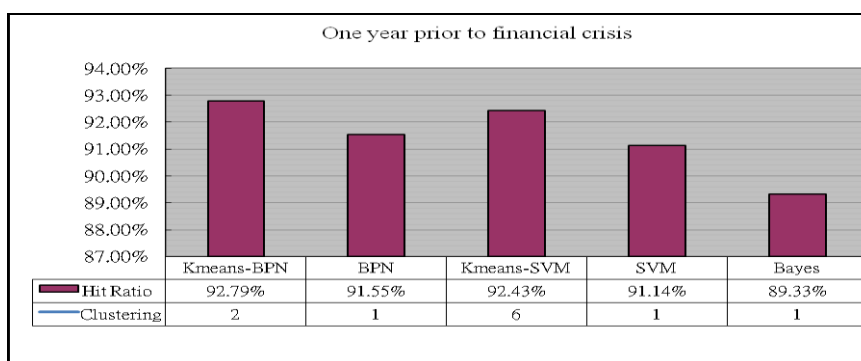
4.2 Two year prior to the financial crisis

The Kmeans-BPN method in this research clusters the data into 4 groups with the prediction accuracy of 90.21%. Whereas in the Kmeans-SVM method, Kmeans clusters the data into 3 groups with the prediction accuracy of 89.68%. As for the BPN method without clustering, the prediction accuracy is 89.09%. As for the SVM method without clustering, the prediction accuracy is 87.21%. Bayes is based on statistical theories with the prediction accuracy of 86.28%. The comparison and the prediction accuracy of these five methods are shown in Figure 3.

Figure 3. Comparison for the prediction accuracy of two year prior to the financial crises
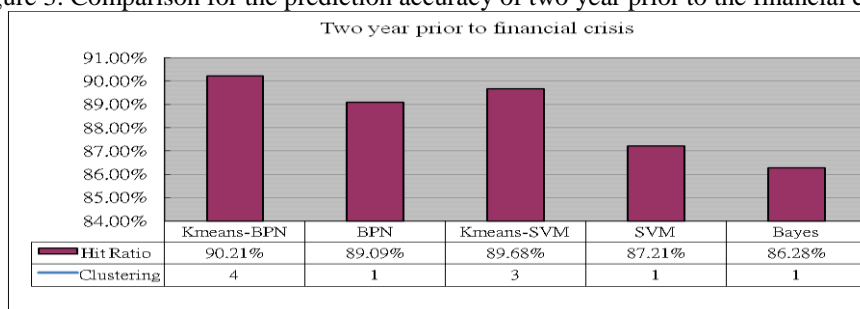
The above results indicates that the Kmeans-BPN and Kmeans-SVM models have better prediction accuracy than BPN and SVM, and the accuracy would increase when the prediction is made closer to the time the financial crises occur. It is also obvious that clustering would increase the prediction accuracy.

4.3   Comparison between the Kmeans-BPN, Kmeans-SVM and Benchmark Model
Through comparison between our research results from the Kmeans-BPN model and the Kmeans-SVM model with the result from the benchmark research model presented by Ng G. S. [2], which were Fuzzy-NN model, we have organized the prediction accuracies in Table 3.

Table 3. Comparison of hit rates for the proposed models with benchmark model

Performance benchmark	Kmeans-BPN	Kmeans-SVM	Fuzzy-NN [2]
Hit-Ratio of One year prior	92.79%	92.43%	90.21%
Hit-Ratio of Two year prior	90.21%	89.68%	83.76%

Comparison for the prediction accuracy of one year prior to the financial crisis of our Kmeans-BPN model stands at 92.79% while the accuracy for Ng's Fuzzy-NN model stands 2.58% less at 90.21%. The prediction accuracy of two year prior to the financial crisis of our Kmeans-BPN model stands at 90.21% while the accuracy for Ng's Fuzzy-NN model stands 6.45% less at 83.76%.

V.   CONCLUSIONS
This study is a combination between the Kmeans-BPN and Kmeans-SVM methods in establishing a model for the prediction of enterprise financial crisis. The research data are provided by Federal Reserve Bank of Chicago. The data set is arbitrarily split into two subsets: about five sixth of the data is used for a training set (1987 to 1992) and one sixth for a validation set (from 1992 to 2008). The K-means algorithm was used for clustering. Finally, the BPN and SVM methods was adopted to determine if the companies were healthy or not, and the prediction accuracy was computed. Comparing to other methods, the proposed K-means BPN model outperforms other forecasting methods, it not only can increase the accuracy of the prediction of bank failure, but also provides great information for business owners and investors.

References
[1]   C. C. Yeh, D. J. Chi and M. F. Hsu, A hybrid approach of DEA, rough set and support vector machines for business failure prediction, Expert Systems with Applications, vol. 37, 2010 , 1535-1541.
[2]   G. S. Ng, C. Quek and H. Jiang, FCMAC-EWS: A bank failure early warning system based on a novel localized pattern learning and semantically associative fuzzy neural network, Expert Systems with Applications, vol. 34, 2008, 989-1003.
[3]   W. H. Beaver, Financial ratios as predictors of failure, Journal of Accounting Research, vol. 4, 71-111, 1966.
[4]   E. I. Altman, Financial ratios, discriminated analysis, and the prediction of corporate bankruptcy, Journal of Finance, vol. 23, 1968, 589-609.
[5]   D. Reagle and D. Salvatore, Forecasting financial crises in emerging market economies, Open Economies Review, vol. 11, 2000, 247-259.
[6]   E. P. Davis and D. Karim, Comparing early warning systems for banking crises, Journal of Financial Stability, vol. 4, 2008, 89-120.
[7]   SPC chart signal using artificial neural networks and support vector machine, International Journal of In Y. E. Shao and B. S. Hsu, Determining the contributors for a multivariate novative Computing, Information and Control, vol. 5, 2009, 4899-4906.

[8]   N. A. Achsani and H. Siregar, Classification of the ASEAN+3 economies using fuzzy clustering approach, European Journal of Scientific Research, vol. 39, pp. 489-497, 2010.

[9]   S. C. K. Shiu, C. H. Sun, X. Z. Wang and D. S. Yeung, Maintaining case-based reasoning systems using fuzzy decision trees, LNAI, 2000, 285-296.

[10]   P. C. Chang, C. H. Liu, Y. W. Wang, A hybrid model by clustering and evolving fuzzy rules for sale forecasting in printed circuit board industry, Decis. Support Syst. 42(3) , 2005, 1254-1269.

[11]   P. C. Chang, T. W. Liao, Combing SOM and fuzzy rule base for flow time prediction in semiconductor manufacturing factory, Appl Soft Computing. 6(2), 2006, 98-206.

[12   P. C. Chang, C. H. Liu, A TSK type fuzzy rule based system for stock price prediction, Expert Syst. Appl. 34(1), 2008, 135-144.

[13]   H. L. Wong, C. C. Wang and Y. H. Tu, Optimal selection of multivariate fuzzy time series models to non-stationary series data forecasting, International Journal of Innovative Computing, Information and Control, vol. 6, 2010, 5321-5332.

[14]   Y. Wang, S. Wang and K. K. Lai, A new fuzzy support vector machine to evaluate credit risk, IEEE Transactions on Fuzzy Systems, vol. 13, 2005, 820-831.

[15]   S. Olafsson, X. Li and S. Wu, Operations research and data mining, Eur. J. Oper. Res. 187,2008,1429-1448.

[16]   G. Zhang, M. Y. Hu, B. E. Patuwo and D. C. Indro, Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis, European Journal of Operational Research, vol.116, 1999, 16-32.

[17]   A. F. Atiya, Bankruptcy prediction for credit risk using neural networks: A survey and new results, IEEE Transactions on Neural Networks, vol. 12, 2001, 929-935.

[18]   P. Ravi Kumar and V. Ravi, Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review, European Journal of Operational Research, vol. 180, 2007, 1-28.

[19]   K. S. Shin, T. S. Lee and H. J. Kim, An application of support vector machines in bankruptcy prediction model, Expert Systems with Applications, vol. 28, 2005, 127-135.

[20]   L. Kaufman, P. J. Rousseeuw, Finding groups in data: An introduction to cluster analysis (John Wiley & Sons, New York, 1990)

[21]   S. Dasgupta, Performance guarantees for hierarchical clustering, Proc. Int. Conference on Computational Learning Theory ,Sydney, Australia,351–363, 2002.

[22]   S. Z. Selim and M. A. Ismail, K-means type algorithms: a generalized convergence theorem and characterization of local optimality, IEEE Trans. Pattern Anal. Mach. Intell. 6,1984, 81–87.

[23]   C. Y. Chen, and F. Ye, Particle swarm optimization algorithm and its application to clustering analysis, Proc. IEEE Inter. Conf. on Networking, Sensing and Control ,Taipei, Taiwan, 2004, 789–794.