2009

Credit Scoring Based on Hybrid Data Mining Classification

Feng-Chia Li
Jen Teh Junior College, lear999@gmail.com

Peng-Kai Wang
Hwa Hsia College, wangpk@cc.hwh.edu.tw

Ying Chieh Liu
Chao Yang University, allanliu@cyt.edu.tw

Credit Scoring based on Hybrid Data Mining Classification

Feng-Chia Li
Lecturer, Department of Information Management at Jen Teh junior College, Miaoli, Taiwan
lear999@gmail.com

Peng-Kai Wang
Lecturer, Department of Information Managementat Hwa Hsia College, Taipei, Taiwan
wangpk@cc.hwh.edu.tw

Ying Chieh Liu
Associate Professor, Department of Information Management at ChaoYang University of Technology
allanliu@cyut.edu.tw

ABSTRACT

The credit scoring has been regarded as a critical topic. This study proposed four approaches combining with the NN (Neural Network) classifier for features selection that retains sufficient information for classification purpose. Two UCI data sets and different approaches combined with NN classifier were constructed by selecting features. NN classifier combines with conventional statistical LDA, Decision tree, Rough set and F-score approaches as features preprocessing step to optimize feature space by removing both irrelevant and redundant features. The procedure of the proposed algorithm is described first and then evaluated by their performances. The results are compared in combination with NN classifier and nonparametric Wilcoxon signed rank test will be held to show if there has any significant difference between these approaches. Our results suggest that hybrid credit scoring models are robust and effective in finding optimal subsets and the compound procedure is a promising method to the fields of data mining.

Keywords

Neural network, Linear discriminate analysis, Decision tree, Rough set, F-score

INTRODUCTION

Consumer credit prediction is a very important issue in the credit industry. With the rapid growth in this field, credit scoring models have been widely used for the credit admission evaluation. The credit scoring models are developed to distinguish which customers are belong to good or bad class with their related attributes such as income, marital status, age or based on the past records. Most credit scoring models have been widely developed by reducing redundant features to improve the accuracy of credit scoring models during the past few years. Dash (1997) provided a detailed survey and overview of the existing methods for feature selection and suggested a feature selection process that consists of four parts including feature generation, feature evaluation, stopping criteria and testing. The classic

evaluation measures such as distance and dependence were used for removing irrelevant features, however, artificial intelligence and machine learning techniques have been used to solve these decision-making problems. The modern data mining techniques have been adopted to build the credit scoring models (Huang et al., 2007). In addition to expert systems, numerous classification techniques have been developed and widely used in credit scoring applications (Baesens et al., 2003). Researchers have developed a variety of conventional statistics models which involve linear discriminate model (Thomas, 2000), decision tree model (Huang et al., 2006), rough set theory model (Caballero et al., 2007), F-score model and genetic programming model (Ong et al., 2005). Recently, many researchers have proposed the hybrid data mining approach in the design of an effective credit scoring model. Lee integrated neural network with traditional discriminate analysis approach (Lee et al., 2002) and Chou applied machine learning techniques such as ANN and DT to solve decision-making problems (Chou et al., 2006). Thus, credit scoring can be regarded as the binary classification problem of classifying an observation into pre-defined groups. Previous studies focused on increasing the accuracy rate of credit score modeling, however, even though a little bit improvement will cause noteworthy cost savings. According to previous studies, machine learning techniques are superior to that of traditional methods in dealing with credit scoring problems, especially in nonlinear pattern classification. For conventional statistical classification, an underlying probability model should be assumed. The more recently developed data mining techniques can perform the classification task without this limitation and achieved better performances than traditional statistical methods (Huang et al., 2007).

Feature subset selection algorithms can be classified into two categories: the filter approach and the wrapper approach (Liu, 1998). The filter approach first selects important features, and then learning algorithms are applied for classification. The wrapper approach either modifies learning algorithms to choose important features as well as conducts training/testing or combines learning algorithms with other optimization tools to perform feature selection. Filter approach usually selects the most relevant variables, but not necessarily the optimal ones for the construction of a good predictor as the selected ones may be redundant. On the other hand, although computationally expensive for larger data set, wrapper approach may perform better in finding useful subsets of relevant variables.

In this study, NN classifier is combined with four features selection approaches to perform better classification. This hybrid credit scoring model is effective in finding optimal subsets and the compound procedure is a promising method to the fields of data mining. The paper is organized as follows. The model development section describes four features selection models, basic NN concepts and the four approaches combined with NN in this research. Next section presents the experimental results from the proposed approaches to classify two real world data sets. Final section gives remarks and provides a conclusion.

MODEL DEVELOPMENT

Linear discriminate analysis (LDA) is a well-known technique which was first proposed by Fisher as a classification technique (Fisher, 1936). LDA has been regarded as a data mining technique in handling classification problems which reduces the observed variables into a smaller number of dimensions that

would result in decreasing the number of features to be considered by the classifiers. Rather than directly eliminating irrelevant or redundant variables from the original feature space, LDA merely transforms the original variables through linear combination into a new subset of variables. Thus, the linear methods provide a way of understanding the data, but they are not able to reduce the number of original features (Li, 2006). The LDA can be expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \qquad (1)$$

Where y represents the discriminate score, β_0 is the intercept term. β_i ($i=1,...,n$) represents the β coefficient associated with the explanatory variable x_i ($i=1,...,n$). LDA is a traditional statistical method and the credit scoring classification accuracy of LDA has been treated as the benchmark to other modern classification approaches. The simple parametric model was the first model employed for credit scoring. However, the covariance matrices of the good and bad credit classes may be unequal for the nature of the credit data. Researchers are investigating hybrid models to overcome the deficiencies of the LDA model. One of the efforts is combined with NN for credit scoring applications.

Rough sets theory (RST) is a mathematical tool that had been used successfully to discover data dependencies and reduce the number of attributes contained in a data set by purely structural methods. RST was first proposed by Pawlak (1984) to deal with vagueness or uncertainty. Rough sets do not need any pre-assumptions or preliminary information about the data. One attribute is chosen as the decision variable and the rest of them are the condition attributes. Two partitions are formed in the mining process. The approach is based on the refusing certain set boundaries, implying that every set will be defined using a lower and an upper approximation. Decision rules derived from lower approximation represents certain rules as well as extracted from upper approximation corresponds to possible rules. An important issue in the RST is about feature reduction based on reduct concept. A reduct is a minimal set of attributes $B \subseteq A$ such that IND(B)=IND(A), where IND(X) is called the X-indiscernibility relation. In other words, a reduct is a minimal set of attributes from A that preserves the partitioning of universe and hence the ability to perform classifications. RST has been successfully applied to real-world classification problems in a variety of areas, such as pattern recognition. Wang proposed a new feature selection strategy based on rough sets and particle swarm optimization (Wang et al., 2007). Zhao also made an empirical experiment for letter recognition for demonstrating the usefulness of the discussed relations and reducts (Zhao et al., 2007). There are many rough set algorithms for feature selection. The basic solution to finding minimal reducts is to generate all possible reducts and choose any with minimal cardinality, which can be done by constructing a kind of discernibility function from the dataset and simplifying it. However, this is time-consuming and therefore is only practical for simple datasets. Finding minimal reducts or all reducts has been shown as NP-hard problems (Skowron, 1992).

Decision tree models are able to represent knowledge in a flexible and easy form. Their popularity is as a result of interpretability and implementation easily. The first decision tree generating algorithm is introduced by Quinlan (1979). Selecting an attribute to place at the root node is the first step to construct a decision tree, and then make one branch based on an attribute value test. This process is repeated recursively on each branch and only those instances that actually reach the branch. Once all cases at a

node have satisfied a certain criterion, stop developing the part of the tree. Calculate the information gain for each attribute and choose the one that gains the most information to split on. The first measure called entropy that characterizes the purity of a arbitrary collection of instances is defined as

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \qquad (2)$$

Pi is the proportion of S belongings to class i. The information gain, Gain (S, A) of an attribute A, the expected reduction in entropy caused by partitioning the examples according to this attribute relative to S, is defined as

$$Gain(S, A) = Entropy(S) - \sum_{value(A)} \frac{S_V}{S} * Entropy(S_v) \qquad (3)$$

Where value (A) is the set of all possible values for attributes A, and S_v is the subset of S for which attribute A has value v. Decision tree model is a popular technique for classification and has been widely used in the community of data mining. Classification trees are constructed to try and maximize their mean classification accuracy. Decision tree model is composed of three basic elements, decision nodes corresponding to attributes, edges or branches corresponding to the different possible attributes and leaves including objects that typically belong to the same class. Several algorithms for building decision trees have been developed such as ID3, C5.0 and CART (Breiman et al., 1984). Classification and regression trees (CART) is a classification method that has been successfully used in many classification applications including cancer survival groups (Garson, 1991) and credit scoring (West, 2000). Besides, CART is a non-parametric statistical method via both categorical and continuous variables. When the dependent variable is categorical, CART produces a classification tree, when it is continuous it will lead to a regression tree.

F-score is a simple technique which measures the discrimination of two sets of real numbers. Given training vectors X_k, k = 1.2,.m, if the number of positive and negative instances are n+ and n-, respectively, then the *F*-score of the i[th] feature is defined as follows (Chen and Lin, 2005).

$$F(i) \equiv \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \qquad (4)$$

Where $\bar{x}_i$, $\bar{x}_i^{(+)}$, $\bar{x}_i^{(-)}$ are the averages of the i[th] feature of the whole, positive, and negative data sets respectively. The numerator indicates the discrimination between the positive and negative sets, and the denominator indicates the one within each of the two sets. The larger the *F*-score is, the more likely this feature is more discriminative (Chen and Lin, 2005).

Neural Network (NN) techniques have long been applied to classification field and have gained widely acceptance beginning from the 1990s (Razi and Athappilly, 2005). NN requires desired outputs to learn like the brain to process information. The key element of NN is the structure of the information processing system which is composed of a large number of highly interconnected processing elements to solve specific problems. The goal of NN is to create a model that correctly maps the input to the output.

NN learns by examples and historical data so that the model can then be used to produce the output when the desired output is unknown. Fig. 1 provides an example of NN with one hidden layer and output neuron.

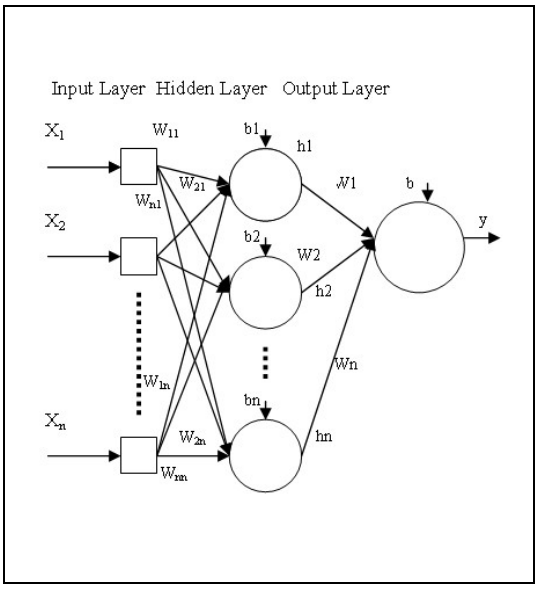

Figure 1. NN algorithm

The output of i[th] hidden neuron is then computed by processing the weighted inputs and its bias term b_i as follows:

$$h_i = f^h(b_i + \sum_{j=1}^{n} w_{ij} x_j) \qquad (5)$$

Where w_{ij} represents the weight connecting input x_j to hidden unit h_i.

The output of the output layer is computed as follows:

$$y = f^{output}(b + \sum_{j=1}^{n} w_i x_j) \qquad (6)$$

With n being the number of hidden neurons and w_j represents the weight connecting hidden unit j to the output neuron. A transfer function is then applied to map the network output y to a classification label. The transfer functions allow the network to model nonlinear relationships in the data and the number of hidden layer nodes does not need to be the same as the number of input nodes. This research will take NN approach as the classifier. Three parameters, learning rate (η), momentum (β) and epoch (t), should be tested in the NN model. To clearly establish a NN based feature selection and parameter optimization system, the system architectures shows in Fig.2. Main steps will be proceeded by two phrases and the detailed explanation is as follows:

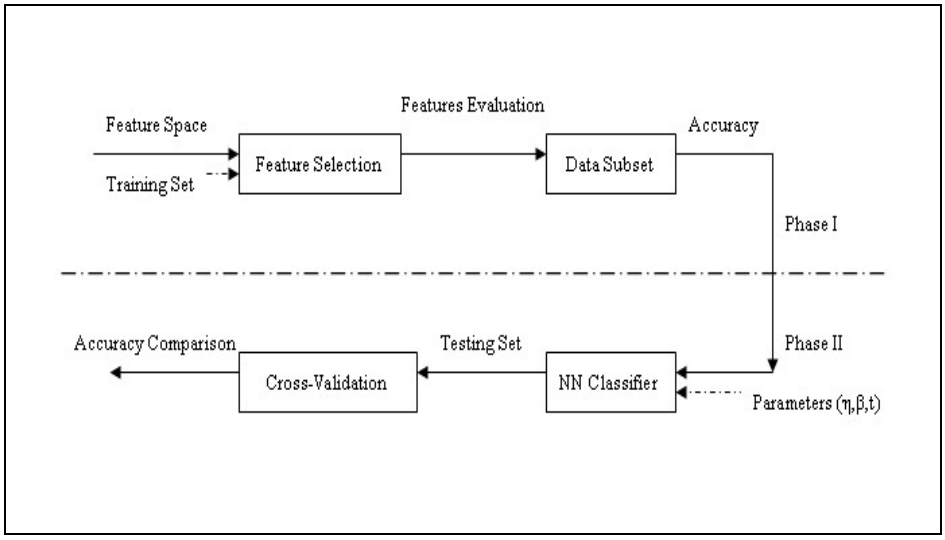

Figure 2. Architecture for the Hybrid Approach

Phase I: Selection the feature subset.

Step1: Collect a set of observed data.

Step2: Calculate the optimal feature subsets such as split decision tree node, effective discriminate variables, rough set indispensable core and F-score of each attributes to decide the initial input variables.

Step3: Generate the optimal subsets according to each algorithm and choose the same number of features as the comparison base.

To guarantee that the present results are valid and can be generalized for making predictions regarding new data, the data set is further randomly partitioned into training and independent testing sets via a k-fold cross validation. Each of the k subsets acts as an independent holdout test set for the model trained with the rest of k-1 subsets. The influence of data dependency is minimized and the reliability of the results can be improved via cross validation (Kudo and Sklansky, 2000). For each of the k subsets of the data set D, create a training set T = D-k, then run a cross-validation process (Chen and Lin, 2005).

Phase II: NN parameters optimization

Step1: Data preprocess of scaling to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Also, avoid numerical difficulties during the calculation and help to increase accuracy. Each variable can be linearly scaled to the range [-1, +1] or [0, 1] by formula (7), where v is original value, vnew is scaled value, max v is upper bound of the feature value, and min v is low bound of the feature value.

$$\text{vnew} = \frac{v - \max^{v}}{\max^{v} - \min^{v}} \qquad (7)$$

Step2: Consider different learning rate (η), momentum(β) and epoch(t)

Step3: For each parameter (η, β, t) in the testing space, conduct cross validation on the training set.

Step4: Choose the parameter (η, β, t) that leads to the lowest CV error classification rate.

Step5: Use the best parameter to create a model as the predictor.

Overall accuracy is averaged across all k partitions. These k accuracy values also give an estimate of the accuracy variance of the algorithms. This study used k = 10, meaning that all of the data will be divided into 10 parts, each of which will take turns at being the testing data set. The other 9 data parts serve as the training data set for adjusting the model prediction parameters. The empirical evaluation was performed on Intel Pentium 4 CPU running at 3.4 GHz and 1G RAM.

RESULT

1. Experimental results

Credit data sets in the real world including various attributes. Two real world data sets presented in Table 1, the Australian and German credit data sets, are derived from the UCI Repository of Machine Learning Databases. The Australian data set consists of 307 "good" applicants and 383 "bad" ones. For each applicant contains 15 features, including 6 nominal, 8 numeric attributes and the final one is class label (good or bad credit). These attributes names have been changed to meaningless symbolic data for the confidential reason. The second accuracy evaluation data set is the German credit scoring data set which is composed of 24 numeric features, including credit history, account balance, loan purpose, loan amount, employment status, personal information, age, housing and job. Additionally, 700 cases are creditworthy and the rest of 300 applicants are not. Four approaches were used in this study, namely "LDA+NN," "DT+NN," "Rough set +NN," "F-score+ NN." The results for the two data sets were obtained by using the four approaches which are summarized in Table 2 and Table 3, respectively.

Names	Total instances	Nominal features	Numeric features	Total features	Number of classes
Australian	690	6	8	14	2
German	1000	0	24	24	2

Table 1. Two Adopted UCI Repository Data Sets

In Table 2, we select 7 different attributes among these four approaches to be the benchmark for the same numbers of features. For the Australian data set, the accuracy rate of the original feature space without selection is 84.78% as well as the accuracy rate of the four approaches achieved 85.64%, 87.10%, 85.63% and 85.49%. DT+NN is slightly superior to that of other three approaches and significant better than the original feature space. In Table 3, we select 12 different attributes among these four approaches to be the benchmark for the same numbers of features. For the German data set, the accuracy rate of the original feature space without selection is 71.90% as well as the accuracy rate of the four approaches achieved 69.80%, 69.81%, 71.66% and 73.34%. It clearly revealed that F-score+ NN is superior to all the other ways.

Combined	Features	Accuracy rate	Accuracy rate

approaches	selected	Avg. (%)	Std. (%)
NN (without selection)	14	84.78	5.37
LDA+NN	7	85.64	5.82
DT+NN	7	87.10	6.03
Rough Set+ NN	7	85.63	4.80
F-score+ NN	7	85.49	5.68

Table 2. Results of Four NN-Based Approaches (Australian Data Set)

Combined approaches	Features selected	Accuracy rate Avg. (%)	Accuracy rate Std. (%)
NN (without selection)	24	71.90	6.12
LDA+NN	12	69.80	4.98
DT+NN	12	69.81	4.73
Rough Set+ NN	12	71.66	5.28
F-score + NN	12	73.34	5.33

Table 3. Results of Four NN-Based Approaches (German Data Set)

To compare classification accuracy of the testing set, a nonparametric Wilcoxon signed rank test was performed for the 10 folds and showed as upper triangle of Table 4 (α=0.05). We found there is a significant difference among LDA+NN approach and Original NN (P-value = 0.046 < 0.05). The result means that LDA in the four approaches associated with NN classifier using only 7 features and can achieve better classificatory accuracy. On the other hand, the lower triangle of Table 4 also illustrates that there are significant differences between Decision Tree and F-score approaches (with α=0.05). F-score+ NN approach is significant better than the approach of DT+NN.

		Australian				
G		Original	Dtree	Rough set	F-score	LDA
e	Original		0.065	0.114	0.107	***0.046***
r	Dtree	0.475		0.169	0.263	0.508
m	Rough set	0.798	0.147		0.799	0.674
a	F-score	0.286	***0.015***	0.125		0.959
n	LDA	0.442	0.888	0.202	0.058	

Table 4. Wilcoxon Signed Rank Test (Australian and German Data Set)

2. Comparison of the accuracy

The ability of these approaches to discriminate between 'good' and 'bad' cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis. ROC curves can also be used to compare the separated performance of two or more classifiers (DeLeo and Rosenfeld, 2001). Every possible point or value on this curve can be selected to discriminate between the two populations with good or bad credit class. Each individual approach will generate a pair of sensitivity and specificity. The ROC curve shows the trade-off between sensitivity and specificity. The closer the curve follows the left and the top borders of the ROC space, the more accurate the model is. The area under the curve (AUC) is the evaluation criteria for the approaches. Taking 10 fold of Australian and German dataset, for example, ROC curve of four approaches are shown in Fig. 3 and Fig. 4 respectively. The average AUC are summarized in Table 5 where DT of Australian data set is the largest one and each approach is better than original NN. On the other hand, F-score of German data set outperforms the other three approaches. All the three approaches (LDA+NN, DT+NN and RST+NN) that they are smaller than original NN classifier. Compared with other approaches, the proposed approaches only improve the classification accuracy by F-score approach and have fewer input features before entering NN classifier.

Datasets	NN	LDA+NN	DT+NN	F-score+ NN	RST+NN
Australian	0.848	0.857	0.871	0.855	0.857
German	0.719	0.698	0.699	0.733	0.717

Table 5. Average AUC of the Four Approaches

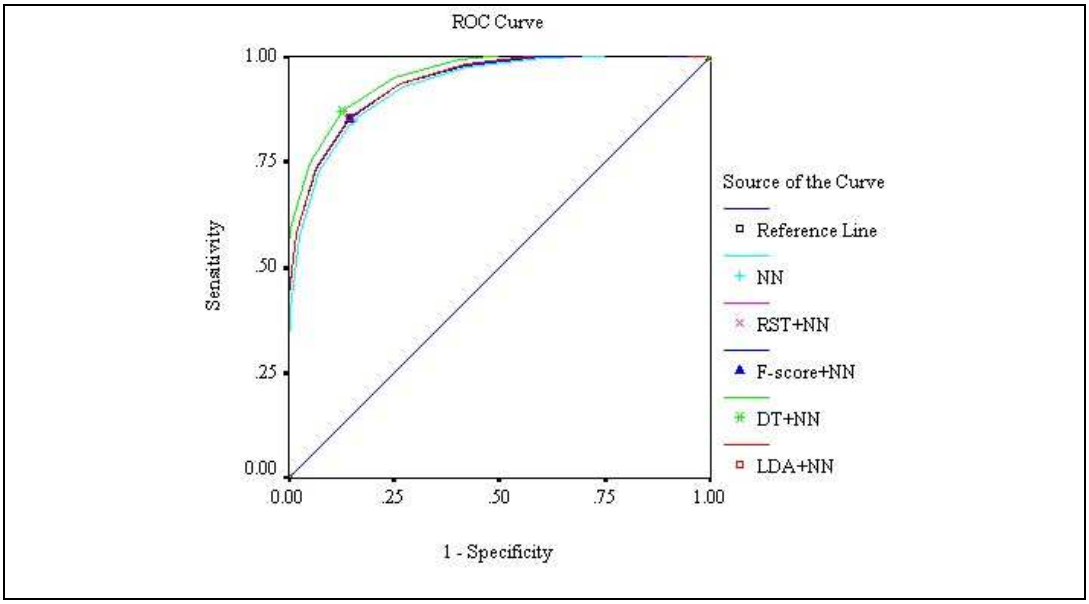

Figure 3. ROC curve for 4 approaches of Australian Dataset

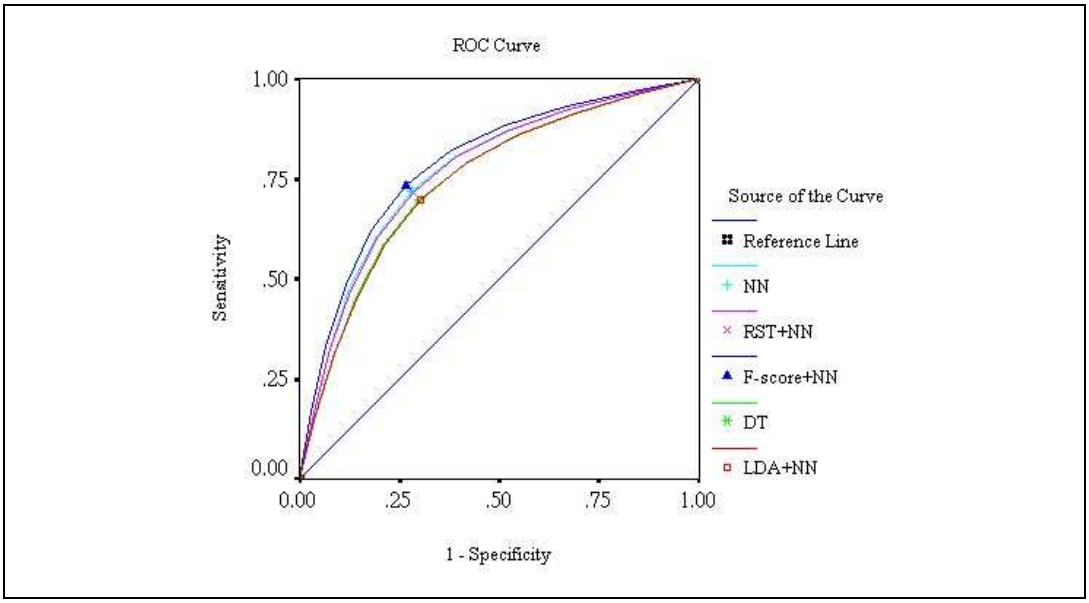

Figure 4. ROC curve for 4 approaches of German Dataset

CONCLUSION

Feature selection involves determining the highest classifier accuracy of a subset or seeking acceptable accuracy of smallest features. This study compromises between accuracy and feature numbers by the same amount of features. From this study, application of data dimensionality reduction pre-processing step is prior to the classification procedures which really improve the overall classification performance. Through four feature selection approaches, it also provides the process that uncovering the essential features and how these features affect the credit scoring model. Fewer features mean that credit department can concentrate on collecting relevant and essential variables. Loading of credit evaluation personnel can be reduced as they do not have to take into account a large number of features during the evaluating procedure and is somewhat less computational intensive. Inside machine learning, feature selection is an important task. It consists of focusing on the most relevant features for use in representing data in order to delete those features considered as irrelevant and that make more difficult a knowledge discovery process inside a database. Feature selection is an important task in the field of classification. This research presents the comparison of hybrid methods based on NN classifier to address feature selection. It consists of focusing on the most relevant features for use in representing data in order to delete those features considered as irrelevant. Meanwhile, it concerns with not only reducing the number of variables but also eliminating noise inputs. The results of the study show that the hit rates of hybrid feature selection methods are higher than those of single methods, especially when the instances equal to both parts (Australian data set). On the other hand, the training result will be dominated by the numbers. Different classes of data sets can be split into different ratios (1:1, 1:2, 2:1) to gain better accuracy in the future study. Additionally, a lot of effective feature selection approaches such as Genetic Algorithms (GA), Simulated Annealing (SA), Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) are developed and are worth experimenting.

ACKNOWLEDGEMENTS

The authors would like to thank the editor, the associate editor, and two anonymous referees for their valuable comments that greatly improved the quality of the paper. Special thanks to the credit department of anonymous bank for providing the scoring process that made this research possible.

REFERENCES

[1] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," *Journal of the Operational Research Society,* vol. 54, pp. 627-635, Jun 2003.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees," Monterey: Wadsworth & Brooks, 1984.

[3] Y. Caballero, D. Alvarez, R. Bel, and M. M. Garcia, "Feature selection algorithms using Rough Set Theory," in *Proceedings of the 7th International Conference on Intelligent Systems Design and Applications*, L. D. Mourelle, N. Nedjah, J. Kacprzyk, and A. Abraham, Eds., 2007, pp. 407-411.

[4] C. H. Chou, C. C. Lin, Y. H. Liu, and F. Chang, "A prototype classification method and its use in a hybrid solution for multiclass pattern recognition," *Pattern Recognition,* vol. 39, pp. 624-634, Apr 2006.

[5] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis,* vol. 1, pp. 131-156, 1997.

[6] J. M. DeLeo and S. J. Rosenfeld, "Essential roles for receiver operating characteristic (ROC) methodology in classifier neural network applications.," in *Proceedings of the international joint conference on neural networks (IJCNN2001)* Washington, DC, USA., 2001, pp. 2730 - 2731.

[7] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics,* vol. 7, pp. 179-188, 1936.

[8] G. D. Garson, "Interpreting neural-network connection weights," *AI Expert Systems in the Micro-electronic Age,* vol. 6, pp. 47-51, 1991.

[9] C. L. Huang, M. C. Chen, and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications,* vol. 33, pp. 847-856, Nov 2007.

[10] J.-J. Huang, G.-H. Tzeng, and C.-S. Ong, "Two-stage genetic programming (2SGP) for the credit scoring model," *Applied Mathematics and Computation,* vol. 174, pp. 1039-1053, 2006.

[11] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition,* vol. 33, pp. 25-41, Jan 2000.

[12] T.-S. Lee, C.-C. Chiu, C.-J. Lu, and I. F. Chen, "Credit scoring using the hybrid neural discriminant technique," *Expert Systems with Applications,* vol. 23, pp. 245-254, 2002.

[13] T.-S. Li, "Feature Selection for Classification by Using a Ga-Based Neural Network Approach," *Journal of the Chinese Institute of Industrial Engineers,* vol. 23, pp. 55-64, 2006.

[14] H. a. M. Liu, H, *Feature Selection for Knowledge Discovery and Data Mining*. Boston: Kluwer Academic Publishers, 1998.

[15] C.-S. Ong, J.-J. Huang, and G.-H. Tzeng, "Building credit scoring models using genetic programming," *Expert Systems with Applications,* vol. 29, pp. 41-47, 2005.

[16] Pawlak, "Rough classification." vol. 20: Academic Press Ltd., 1984, pp. 469-483.

[17] J. R. Quinlan, "Discovering rules from large collections of examples: a case study.," *Expert Systems in the Micro-electronic Age,* pp. 168--201, 1979.

[18] M. A. Razi and K. Athappilly, "A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models," *Expert Systems with Applications,* vol. 29, pp. 65-74, Jul 2005.

[19] A. Skowron, Rauszer, C., "The discernibility matrices and functions in information systems. In: Slowinski, R. (Eds.):," in *Intelligent Decision Support--Handbook of Applications and Advances of the Rough Sets Theory*: Kluwer Academic Publishers, 1992, pp. 311-362.

[20] L. C. Thomas, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers," *International Journal of Forecasting,* vol. 16, pp. 149-172, 2000.

[21] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters,* vol. 28, pp. 459-471, 2007.

[22] D. West, "Neural network credit scoring models," *Computers & Operations Research,* vol. 27, pp. 1131-1152, 2000.

[23] Y. Zhao, Y. Yao, and F. Luo, "Data analysis based on discernibility and indiscernibility," *Information Sciences,* vol. 177, pp. 4959-4976, 2007.