Volume:02, Issue: 03, Pages: 714-718 (2010)

Evaluation of Feature Selection Methods for Predictive Modeling Using Neural Networks in Credits Scoring

Raghavendra B. K.
Dr. M.G.R. Educational and Research Institute, Chennai-95
Email: raghavendra_bk@rediffmail.com

Dr. Jay B. Simha
Abiba Systems, Bengaluru-50
Email: jay.b.simha@abibasystems.com

ABSTRACT

A credit-risk evaluation decision involves processing huge volumes of raw data, and hence requires powerful data mining tools. Several techniques that were developed in machine learning have been used for financial credit-risk evaluation decisions. Data mining is the process of finding patterns and relations in large databases. Neural Networks are one of the popular tools for building predictive models in data mining. The major drawback of neural network is the curse of dimensionality which requires optimal feature subset. Feature selection is an important topic of research in data mining. Feature selection is the problem of choosing a small subset of features that optimally is necessary and sufficient to describe the target concept. In this research an attempt has been made to investigate the preprocessing framework for feature selection in credit scoring using neural network. Feature selection techniques like best first search, info gain etc. methods have been evaluated for the effectiveness of the classification of the risk groups on publicly available data sets. In particular, German, Australian, and Japanese credit rating data sets have been used for evaluation. The results have been conclusive about the effectiveness of feature selection for neural networks and validate the hypothesis of the research.

Keywords – Credit Scoring, Data Mining, Feature Selection, Neural Networks.

Date of Submission: June 29, 2010 Revised: September 14, 2010 Date of Acceptance: September 23, 2010

1. Introduction

With the rapid growth in the credit industry, credit scoring models have been extensively used for credit admission evaluation [1]. Several quantitative methods have been developed for the credit admission decision. The credit scoring models are developed to categorize applicants as either accepted or rejected with respect to the applicant's characteristics such as age, income, and marital condition. Credit officers are faced with the problem of trying to increase credit volume without excessively increasing this exposure to default. The benefits of credit scoring involve reducing the credit analysis cost, enabling faster credit decisions, closer monitoring of existing accounts.

Practitioners and researchers have developed a variety of traditional statistical models and data mining tools for credit scoring which involves linear discriminant models, logistic models, k-nearest neighbor models, decision tree models, neural network models, and genetic programming models. Empirical research results from the available literature suggests that neural network models for credit scoring gives more accurate, adaptive, and robust results followed by linear discriminant analysis, logistic regression, decision trees, and k-nearest neighbor. The major drawback of a neural network approach is that neural network becomes computationally expensive with increase in number of features. In this research work an attempt has been made to evaluate various feature selection methods for credit risk classification using neural network.

The rest of the paper is organized as follows: Section 2 reviews the prior literature. Neural network technique is discussed in Section 3. In Section 4 Feature selection methods are discussed. Experimental methodology is discussed in Section 5. Section 6 includes Results and discussion followed by conclusion.

2. Literature Survey

Recent advances in computing technology in terms of speed, cost as well as access to tremendous amounts of computing power and the ability to process huge amounts of data in reasonable time has spurred increased interest in data mining applications. Machine learning has been one of the methods used in most of these data mining applications. The data used as input in any of these learning systems are the primary source of knowledge in terms of what is learned by these systems. There have been relatively few studies on preprocessing of data for use in these data mining systems as input and evaluate several feature selection methods as to their effectiveness in preprocessing input data. The real-world financial credit-risk data is used in evaluating these systems [2].

Volume:02, Issue: 03, Pages: 714-718 (2010)

In another research the optimal parameters, the comparative efficiency and accuracy of three models: Multi Layer Perception, Ensemble Averaging and Boosting by Filtering have been investigated in the light of credit loan application classification. The objective of the research was to find the best tool among the three neural network models for this kind of decision context. The experimental results indicate that Committee Machine models were superior to a single Multi Layer Perception model, and that Boosting by Filtering outperformed Ensemble Averaging [3].

The prediction of corporate bankruptcies is an important and widely studied topic since it can have significant impact on bank lending decisions and profitability [4]. Author's present two contributions: First, the review topic of bankruptcy prediction with emphasis on neural-network (NN) models. Second, the NN bankruptcy prediction model. Inspired by one of the traditional credit risk models developed by Merton, the novel indicators for the NN system has been proposed and shows that the use of these indicators in addition to traditional financial ratio indicators provides a significant improvement in the prediction accuracy (from 81.46% to 85.5%).

Huanng et al [5] used back propagation neural network (BNN) as a benchmark and obtained prediction accuracy around 80% for both BNN and artificial neural network (ANN) methods for the United States and Taiwan markets. However, only slight improvement of Support Vector Machine (SVM) was observed. Another direction of the research was to improve the interpretability of the AI-based models. Then they applied recent research results in neural network model interpretation and obtained relative importance of the input financial variables from the neural network models. Based on these results, they conducted a market comparative analysis on the differences of determining factors in the United States and Taiwan markets.

There is an approach to solve a classification problem by combining feature selection and neural networks. The main idea is to use techniques from the field of information theory to set of important attributes that can be used to classify tuples. A neural network is trained using these attributes; the neural network is then used to classify tuples [6].

3. Neural Networks

Practitioners and researchers have developed a variety of traditional statistical models and data mining tools for credit scoring which involve linear discriminant models, logistic models, k-nearest neighbor models, decision tree models, neural network models, and genetic programming models. Neural network model is outlined below.

Neural networks a form of sub symbolic computation are based on the workings of the brain. A neural network comprises a set of weighted edges and nodes. Learning is achieved by modification of these weights. Most networks contain a number of layers, the first layer being the input layer, the final layer being the output layer. Other internal layers (hidden layers) are often required to ensure sufficient computational power in the network [6].

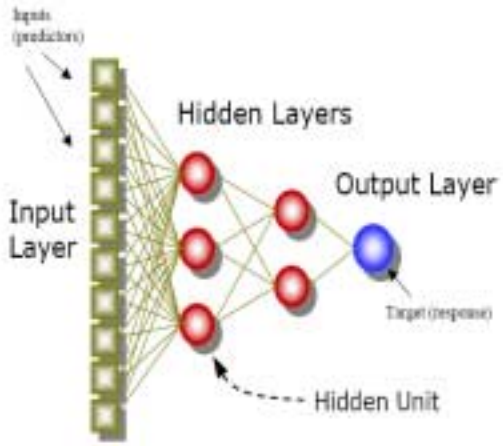

Figure 1. Multilayer perceptron (MLP)

A network can be trained to map input values to corresponding output values by providing a training set. The network is repeatedly tested and modified to produce the correct output.

The generation of output by a neural network is accomplished via firing values from nodes. An input is passed to the input layer which in turn can activate the internal layers, which in turn activate output layer, finally resulting in an output.

Given 'n' links feeding into a node, each link has an input value X_j and a weight W_j. The nodes have an associated threshold, τ. If, according to some activation function, the node has a sufficiently high activation level, the node fires a value onto the next layer. Commonly used activation functions include:

$$f(a) = \begin{cases} 1 & \text{if } a>\tau \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$f(a) = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}} \quad (2)$$

where a, the activation of a node, is $\sum_{j=1}^{n} X_j W_j$ and τ is a threshold. The initial input vector is fed into the network; sets of nodes are fired which finally results in an output vector.

To train the network any errors in the output are fed back through the network causing a modification of the weights on the nodes. Errors can be calculated at the output layer. For internal nodes, the error is a function of all nodes that use the node's output and the output from that node.

The error at the output layer is used to re-modify the weights coming to the output layer. This allows the calculation of errors at the last hidden layer etc. The error is back-propagated through the network.

The major criticism of a neural network approach include the fact that because neural networks learn the classification rules by multiple passes over the training data set, the learning time, or the training time needed for a

Volume:02, Issue: 03, Pages: 714-718 (2010)

neural network to obtain a high classification rate, is usually long [7]. In addition, there is difficulty in understanding the rules generated by neural networks as they are buried in the network architecture and the weights assigned to the links between the nodes. Also, there is difficulty in incorporating any available domain knowledge. Final and important drawback of neural network is curse of dimensionality, which will affect all the factors which are discussed above. The limitation can be minimized to some extent by proper selection of features as discussed in later sections.

4. Feature Selection Methods

Feature selection is often an essential data preprocessing step prior to applying a classification algorithm such as Multi Layer Perception (MLP). The term feature selection is taken to refer to algorithms that output a subset of the input feature set. One factor that plagues classification algorithms is the quality of the data. If information is irrelevant or redundant or the data is noisy and unreliable then knowledge discovery using training is more difficult [8]. Regardless of whether a learner attempts to select features itself or ignores the noise, feature selection prior to learning can be beneficial. Reducing the dimensionality of the data reduces the size of the hypothesis space and allows algorithm to operate faster and more effectively. In some cases accuracy on classification can be improved [8]. Feature Selection methods are outlined below.

BFS

This method searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility. Setting the number of consecutive non-improving nodes allowed controls the level of backtracking done. Best first may start with the empty set of attributes and search forward, or start with the full set of attributes and search backward, or start at any point and search in both directions (by considering all possible single attribute additions and deletions at a given point) [9].

InfoGainAttributeEval

This method evaluates the worth of an attribute by measuring the information gain with respect to the class [9]. InfoGain is calculated using the following equation, InfoGain(Class, Attribute) = H(Class) - H(Class | Attribute).

WrapperSubsetEval

This method evaluates attribute sets by using a learning scheme and cross validation is used to estimate the accuracy of the learning scheme for a set of attributes [9].

GainRatioAttributeEval

This method evaluates the worth of an attribute by measuring the gain ratio with respect to the class [9]. GainRatio is calculated using the following equation, GainR (Class, Attribute) = (H (Class) - H (Class | Attribute)) / H (Attribute).

RandomSearch

This method performs a Random search in the space of attribute subsets. If no start set is supplied, random search starts from a random point and reports the best subset found. If a start set is supplied, random searches randomly for subsets that are as good as or better than the start point with the same or fewer attributes [9].

ChiSquaredAttributeEval

This method evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class [9].

ConsistencySubsetEval

This method evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes.

Consistency of any subset can never be lower than that of the full set of attributes; hence the usual practice is to use this subset evaluator in conjunction with a Random or Exhaustive search which looks for the smallest subset with consistency equal to that of the full set of attributes [9].

PrincipalComponents

This method performs a principal components analysis and transformation of the data. Use in conjunction with a Ranker search. Dimensionality reduction is accomplished by choosing enough eigenvectors to account for some percentage of the variance in the original data-default 0.95 (95%). Attribute noise can be filtered by transforming to the PC space, eliminating some of the worst eigenvectors, and then transforming back to the original space [9].

ReliefAttributeEval

This method evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class and this can operate on both discrete and continuous class data [9].

SVMAttributeEval

This method evaluates the worth of an attribute by using an SVM classifier. Attributes are ranked by the square of the weight assigned by the SVM. Attribute selection for multiclass problems is handled by ranking attributes for each class separately using a one-vs.-all method and then "dealing" from the top of each pile to give a final ranking [9].

5. Experimental Methodology

The framework for feature selection in credit scoring is shown in Fig. 2. The process of evaluation is as follows. The feature selection algorithms are applied on the data set and the selected features from each of the algorithm are used to develop a predictive model for risk classification using a neural network. A ten-fold cross validation has been used for evaluation.

In 10-fold cross validation, the original sample is partitioned into 10 sub samples, of the 10 sub samples, a single sub sample is retained as the validation data for testing the model, and the remaining 9 sub samples are used as training data. The cross-validation process is then repeated 10 times (the folds), with each of the 10 sub samples used exactly once as the validation data. The 10 results from the folds then can be averaged (or otherwise

combined) to produce a single estimation. The advantage of this method over repeated random sub sampling is that all observations are used for both training and validation and each observation is used for validation exactly once.

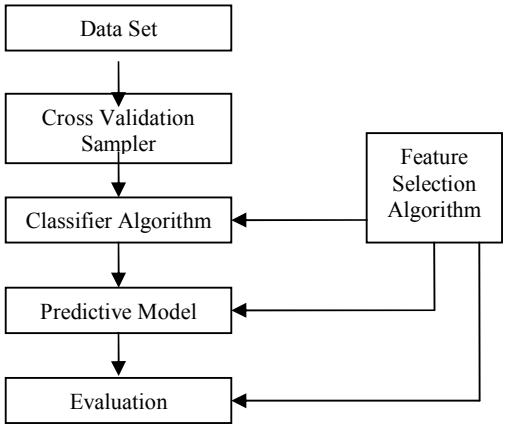

Figure 2. Feature selection framework

6. Results And Discussions

Three publicly available data sets – Australian, German, and Japanese credit data are used. In the evaluation different feature selection methods have been evaluated for the effectiveness of the classification of the risk groups using neural network.

The classification accuracy, and number of features selected are used to measure the performance of the predictive model with the neural network. Table 1 gives the specifications for the datasets and classification accuracy before pre processing. The results of the evaluation are given in the Table 2. From the results it can be observed that the classification accuracy and number of features selected by BFS, WrapperSubsetEval, and RandomSearch methods with neural network are more efficient when compared with other methods. The classification accuracy and number of features selected are almost similar. Fig. 3 gives feature dimensionality reduction, and Fig. 4 gives classification accuracy details after evaluation process.

Table 1. Specification for the data sets

Sl. No	Dataset	No of instances	Total no. of attributes	No of classes	CA
1	Australian	690	15	2	79.4
2	German	1000	21	2	71.6
3	Japanese	690	16	2	84.2

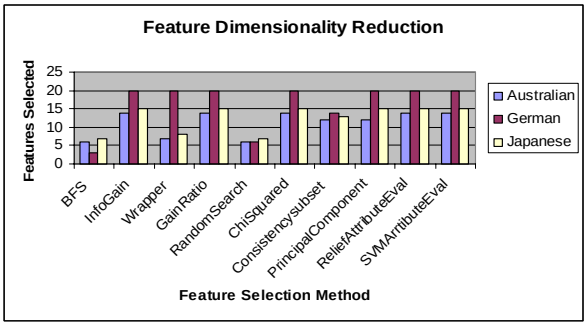

Figure 3. Feature dimensionality reduction

Table 2. Neural network specification for Australian, German, and Japanese data sets

Feature Selection Methods	Australian Data Set		German Data Set		Japanese Data Set	
	NFS	CA	NFS	CA	NFS	CA
All Features	15	79.42	21	71.6	16	84.20
BFS	6	88.08	3	72.9	7	81.27
InfoGain	14	80.42	20	73.8	15	80.85
WrapperSubsetEval	7	66.38	1	68.8	8	71.91
GainRatio	14	85.3	20	73.8	15	80.85
RandomSearch	6	79.42	6	71.5	7	84.78
ChiSquared	14	79.42	20	71.6	15	84.20
Consistencysubset	12	84.78	14	71.8	13	83.3
PrincipalComponent	12	84.78	20	71.6	15	84.20
ReliefAttributeEval	14	79.42	20	71.6	15	84.20
SVMArrtibuteEval	14	79.42	20	71.6	15	84.20

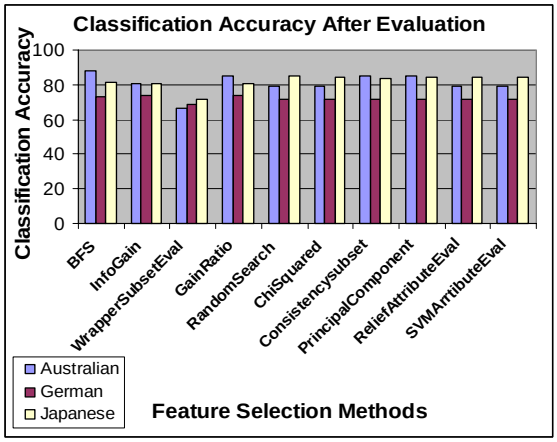

Figure 4. Classification accuracy after evaluation

7. Conclusion

In this research work an attempt has been made to evaluate feature selection algorithms on the data sets and the selected features from each of the algorithm are used to develop a predictive model for risk classification using a neural network. The results have been conclusive about the effectiveness of feature selection for neural networks and validate the hypothesis of the research.

Volume:02, Issue: 03, Pages: 714-718 (2010)

References

[1] Cheng-Lung Huang, Mu-Chen Chen, Chieh-Jen Wang, "Credit scoring with a data mining approach based on support vector machines", *Expert Systems with Applications: An International Journal*, *vol 33*, No. 4, 2007, 847-856.

[2] Piramuthu S., "Evaluation Feature Selection Methods for Learning in Data Mining Applications", *31st Hawaii International Conference on System Sciences*, vol 5, 1998, 294-301.

[3] Meliha Handzic, Felix Tjandrawibawa and Julia Yeo, "How Neural Networks Can Help Loan Officers to Make Better Informed Application Decisions", *Information Science And Information Technology Education Joint Conference*, 2003, 97-108.

[4] Amir F. Atiya, Senior member, IEEE, "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results", *IEEE Transactions on Neural Networks*, Vol. 12, 2001, No. 4.

[5] Huang Z., Chen H.C., Hsu C.J., Chen W.H., and Wu S.S., "Credit rating analysis with Support Vector Machines and neural networks: a market comparative study", *Decision Support Systems*, vol 37, no 4, 2004, pp 543-558.

[6] P.O'Dea, J.Griffith, and C.O'Riordan, "Combining Feature Selection and Neural Networks for Solving Classification Problems", *Technical Report of the Dept of IT*, NUI, Galway, 2001.

[7] H. Lu, R. Setiono, and H. Liu. Nuerorule: "A connectionist approach to data mining". *In Proceedings of the 21st VLDB Conference*, 1995, 478-489.

[8] Liu H., Sentino R, Some issues on scalable Feature Selection, *Expert Systems with Application*, vol 15, 1998, 333-339.

[9] http://www.cs.waikato.ac.nz/ml/weka/

Authors Biography

Dr. Jay B. Simha is working as a Chief Technology Officer at Abiba Systems, Bangalore, INDIA. He has received M.Tech degree from Mysore University in Maintenance Engineering (1994), M.Phil. Degree in Computer science from MS University (2003) and PhD degree from Bangalore University in data mining and decision support (2004). He worked with Alpha Systems, Siemens and other companies in software development and analysis. His research areas are Data Mining, Fuzzy Logic, and Artificial Intelligence. He has published over 25 papers in refereed journals and conferences.

Raghavendra B. K. is working as an Assistant Professor & Head, Department of Computer Science and Engineering at Ghousia College of Engineering, Ramanagaram, Karnataka, India. He has received his BE degree in Computer Science & Engineering from Bangalore University (1994) and M.Tech degree in Computer Science & Engineering from Visveswaraya Technological University, Belgaum (2004), and pursuing his PhD at Dr. MGR University, Chennai, India. His research area is Data Mining.