I.J. Intelligent Systems and Applications, **2014, 08, 60-68**
Published Online July 2014 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2014.08.07

Nearest Neighbor Classifier Method for Making Loan Decision in Commercial Bank

Md.Mahbubur Rahman, Samsuddin Ahmed, Md. Hossain Shuvo
Dept. of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh
Email: mahabub.cse.buet@gmail.com, sambd86@gmail.com, m2hossain.shuvo108590@gmail.com

Abstract— Bank plays the central role for the economic development world-wide. The failure and success of the banking sector depends upon the ability to proper evaluation of credit risk. Credit risk evaluation of any potential credit application has remained a challenge for banks all over the world till today. Artificial neural network plays a tremendous role in the field of finance for making critical, enigmatic and sensitive decisions those are sometimes impossible for human being. Like other critical decision in the finance, the decision of sanctioning loan to the customer is also an enigmatic problem. The objective of this paper is to design such a Neural Network that can facilitate loan officers to make correct decision for providing loan to the proper client. This paper checks the applicability of one of the new integrated model with nearest neighbor classifier on a sample data taken from a Bangladeshi Bank named Brac Bank. The Neural network will consider several factors of the client of the bank and make the loan officer informed about client's eligibility of getting a loan. Several effective methods of neural network can be used for making this bank decision such as back propagation learning, regression model, gradient descent algorithm, nearest neighbor classifier etc.

Index Terms — Credit Evaluation, Decision Process, Backpropagation, Nearest Neighbor Rule, Gradient Descent Algorithm

I. INTRODUCTION

The use of Artificial Neural Network has become prominent in different sectors and field of finance. In predicting future stock, bankruptcy, exchange rate, and detecting credit card fraudulence neural network has shown its superiority over other systems. This is because, the highly distributed parallel structure of neural network making it possible to make crucial decision solved. Now with the rapid growth of economic and money flow, making loan decision has become great concern of bank authority of any country, because fraud in providing loan is spreading rapidly. The granting of loans by a financial institution (bank or home loan business) is one of the important decision problems that require delicate care. Loan applications can be categorized into good applications and bad applications. Good applications are the applications that are worthy of giving the loan. Bad applications are those ones that should be rejected due to the small probability of the applicants ever returning the loan. The institution usually employs loan officers to make credit decisions or recommendations for that institution. These officers are given some hard rules to guide them in evaluating the worthiness of loan applications. After some period of time, the officers also gain their own experiential knowledge or intuition (other than those guidelines given from their institution) in deciding whether an application is loan worthy or not.

Recently biggest amounts of loan scandal (around to 6 billion dollar) in the history of Bangladesh was happened by Sonali Bank who sanctioned loan to Hallmark Group without proper and strong judgment that makes the bank authority more anxious. Inability of detecting fake documents and client attribute caused this massive loan scandal[1].

Generally, there is widespread recognition that the capability of humans to judge the worthiness of a loan is rather poor (Glorfeld, 1996). Some of the reasons are: (1) There is a large gray area where the decision is up to the officers, and there are cases which are not immediately obvious for decision making; (2) Humans are prone to bias, for instance the presence of a physical or emotional condition can affect the decision making process. Also personal acquaintances with the applicants might distort the judgmental capability; (3) Business data warehouses store historical data from the previous applications. It is likely that there is knowledge hidden in this data, which may be useful for assisting the decision making. Unfortunately, the task of discovering useful relationships or patterns from data is difficult for humans. The reasons for such difficulties are the large volume of the data to be examined, and the nature of the relationships themselves that are not obvious.

Given the fact that humans are not good at evaluating loan applications, a knowledge discovery tool thus is needed to assist the decision maker to make decisions regarding loan applications. Knowledge discovery provides a variety of useful tools for discovering the non-obvious relationships in historical data, while ensuring those relationships discovered will generalize to the new/future data. This knowledge in the end can be used by the loan officers to assist them in rejecting or accepting applications. Past studies show that even the application of a simplistic linear discriminant technique in place of human judgment yields a significant, although still unsatisfactory increase in performance (Glorfeld, 1996).Treating the nature of the loan application evaluation as a classification (Smith, 1999) and forecasting problem(Thomas, 1998), it is argued here that neural networks may be suitable as knowledge discovery tools for the task. Solving of this problem requires very subtle calculation and observation. Classifying banks clients depending on several attributes who seek for loan into categories can be a solution to this problem. That is, to observe different attributes of clients including their professions, their credit ratings, transactions, past history of loan, loan transactions with other bank etc and then classify them into two categories like eligible and ineligible.

As many banks provide loan to its clients through a strong assessment, using of neural network and this will make it easier to make the decision more precise.

Therefore, the main objectives of this study are: (i) To develop a robust knowledge discovery tool using Nearest Neighbor classifier neural network models that is both reliable and easy to build, and (ii) To compare the performance with the basic neural network model called Multi Layer Perceptron (MLP) with Committee Machine models (namely Ensemble Averaging, MLP with Backpropagation and Boosting by Filtering) in scoring credit loan applications.

This paper is structured as follows: section 1 discuss about introductory concept, section 2 describes the background of the study, section 3 reviewed the previous works and concepts, section 4 proposed Nearest Neighbor Classifier as a decision methodology, section 5 defines the simulation parameters and concepts, section 6 analyzes the performances and results, section 7 identifies about the limitations and set the future goals and section 8 concludes the paper.

II. BACKGROUND

A. Credit Evaluation

When it is required to obtain credit scoring, one has to undergo a process of evaluation before the credit score is sanctioned. This process is called as credit evaluation, which may take time, but concludes in either an approval or a rejection. Credit approval for financial sanctions is a vital component of any evaluation process as it is related to the economy of a country. Before a potential debtor wants to obtain credit, he must be evaluated on certain areas. There are five C's involved in credit evaluation. As discussed in [14] they are: character, credit report, capacity, cash flow, and collateral. The character of a person applying for a credit is a big factor to the decision for credit approval. A person with a sound financial objective is likely to be granted a credit approval quickly and are possibly than an individual who is in bad shape, not just on the financial facet, but also on other aspect Credit history is another important factor considered by lenders in their decision to grant and approve credit applications. The credit report is a record of an individual's past borrowing and reimbursing transactions. It also includes information about late payments and bankruptcy. A credit report can be tarnished. A credit score can be at its low. Under these circumstances it is unlikely for you to earn the confidence of the lender for a credit approval. If your cash flow is good, there is a possibility of getting the credit approval. Lenders may also have to check the liquidity of an individual. This can be done by checking the bank statements of an individual borrower. In the case of businesses, lenders may have to obtain a copy of the audited financial statements. The financial statements of businesses and bank statements can be utilized to show the capacity of a borrower to settle and repay a line of credit. The capacity of the borrower to pay a credit is determined during credit evaluation and approval. Credit evaluation is a process taken by the lender with the participation of the credit applicant. If you want to undergo this process, it is important to make substantial preparation so you are more likely to obtain a credit quickly and less expensively.

B. Decision Process for Credit Evaluation

Credit managers rely heavily upon external data sources to guide them in the credit decision process. To approve or reject a credit request is a delicate task. A credit manager must evaluate the risk associated with extending credit and declining an applicant based on numerous factors [15]. The need for sufficient and reliable information is the foundation of a successful credit decision. A credit manager may call on references, run background checks, pull a credit report, verify bank accounts or ask questions of the applicant to validate the information on the credit application. Credit managers are challenged with the task of obtaining readily available information to support their decision while sending a timely response to the applicant. A major obstacle in achieving this task is the turnaround time associated with checking references. The process varies from business to business and may include a background check, a verification of a bank deposit or credit references with existing suppliers.

C. Motivation for Designing Different Credit Functions Using Different Models

Some businesses require written requests, while others may offer to do a phone interview at their convenience.

The credit function is the heart of banking, under the ever changing market conditions. The lack of general credit review system in many banks and the lack of precise methods for measuring credit risk are two important reasons why an expert support system is necessary [15]. Such a system can be implemented by using advantages of Radial Basis Neural Network techniques, Multilayer perceptron Model, Regression techniques, SVM and Decision trees. Decision trees and Logistic Regression are well-established traditional

Credit Evaluation Model of Loan Proposals for Bangladeshi Banks statistical techniques, whereas Radial Basis Neural Networks are relatively new data mining tools that have been successfully used for classification and prediction. Multilayer perceptrons using a backpropagation algorithm are the standard algorithm for any supervised-learning pattern recognition process. Decision trees are particularly useful for classification tasks. Like Radial Basis Neural Networks, decision trees learn from data. Using search heuristics, decision trees are able to find explicit and understandable rules-like relationships among independent and dependent variables.

The purpose of the logistic regression model is to obtain a regression equation that could predict in which of two or more groups an object could be placed (i.e. whether a credit should be classified as approved or rejected). SVM is a class of data driven and non linear methods that do not require specific assumptions on the underlying data.

This feature is suitable for practical business problem where there are massive data. The feature of different model that can be used for designing credit function can be given as follows.

1. Decision tree model

Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance and each branch descending from that node corresponds to one of the possible values for this attribute [16]. Advantages of using decision learning tree algorithms are:

1) They generalize in a better way for unobserved instances, once examined the attribute value pair in the training data.

2) They are efficient in computation as it is proportional to the number of training instances observed.

3) The tree interpretation gives a good understanding of how to classify instances based on attributes arranged on the basis of information they provide and makes the classification process self-evident.

The operation of decision tree is based on ID3 or C4.5 algorithms. Decision tree are self explanatory and can be easily converted to set of rules so they are used in credit evaluation process.

2. Radial basis neural network model

Artificial neural networks are one of the most common data mining tools. Neural networks are particularly useful for the tasks of classification, prediction, and clustering in business applications. Neural network models are characterized by three properties: the computational property, the architecture of the network, and the learning property [17]. Computational Property: - Neural networks are made up of neurons or nodes, which are simple processing elements. Each neuron contains a summation node and often a nonlinear sigmoidal activation function of the form

$$F(n)=\frac{1}{\left(1+\exp\left(-2n\right)\right)} \tag{1}$$

Where n = WP is the output from a summation node; λ is the steepness of the activation function; W is a weight matrix and P is an input vector. Because a single neuron has a limited capability, neurons (sometimes hundreds) are organized in layers and are interconnected between layers using connections called weights. Each weight carries a numerical value that represents the strength of connection or expresses the relative importance of each input to the neuron.

Architecture: - Radial basis neural network is most commonly used architectures used in financial applications is radial basis neural network. Radial Basis Functions are powerful techniques for interpolation in multidimensional space. They can model non linear function using a single hidden layer which removes some design decisions about number of layers. The simple linear transformation in the output layer can be optimized fully using traditional linear modeling techniques which are fast and do not suffer from the problems such as local minima. RBF networks can therefore be trained extremely quickly.

Learning: - Neural networks use a three types of learning modes supervised learning, unsupervised learning, and reinforcement learning. During supervised learning, which is the most common for the mentioned feed-forward networks, weights are initialized at small random values and training patterns are presented to the network one pattern at a time. The output produced by the training pattern is compared with the actual response provided by a teacher. The differences modify the weights of the network to make them closer to the actual output. This process is repeated for all training patterns contained in a training set until the cumulative error between the actual outputs and the network's output is reduced to a small value.

Weights are crucial to the operation of the neural network because through their repeated adjustment the neuron (or network) learns. Knowledge of the network is encoded in its weights. The most attractive features of these networks are their ability to adapt, generalize, and learn from training patterns. Due to this feature it is used in credit evaluation process.

3. Logit Regression Model

The purpose of the logistic regression model is to obtain a regression equation that could predict in which of two or more groups an object could be placed (i.e. whether a credit should be classified as a good credit or a bad credit) [17]. The logistic regression also attempts to predict the probability that a binary or ordinal target will acquire the event of interest (e.g. credit payoff or credit default) as a function of one or more independent variables (i.e. amount of credit, borrower job category, reason of credit). The logit model is represented by the logistic response function P(y) of the form:

$$P\left(y\right)=\frac{1}{\left(1+\exp\left(-z\right)\right)}, \tag{2}$$
$$where\ z=b_0+\sum b_i x_i,\ \forall_i=1\,to\ m$$

The function P(y) describes a dependent variable y containing two or more qualitative outcomes. Z is the function of m independent variables x called predictors, and b represents the parameters. The x variables can be categorical or continuous variables of any distribution. The value of P(y) that varies from 0 to 1 denotes the probability that a dependent variable y belongs to one of two or more groups. The principal of maximum likelihood can commonly be used to compute estimates of the b parameters [17]. This means that the calculations involve an iterative process of improving approximations for the estimates until no further changes can be made. Unlike radial basis neural networks, logistic regression models are designed to predict one dependent variable at a time. Logistic regression output provides statistics on each variable included in the model. Researchers then can analyze the applications with these statistics to test the usefulness of specific information. This model is easy to use and has good flexibility so it is used in credit scoring application.

4. Multilayer Perceptron Neural Network Model

A Multilayer Perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network MLP is suitable for credit evaluation process because of its classification accuracy [18]. If a multilayer perceptron has a linear activation function in all neurons, that is, a simple on-off mechanism to determine whether or not a neuron fires, then it is easily proved with linear algebra that any number of layers can be reduced to the standard two-layer input-output model Each neuron uses a nonlinear activation function which was developed to model the frequency of action potentials, or firing, of biological neurons in the brain. This function is modeled in several ways, but must always be normalizable and differentiable. The two main activation functions used in current applications are both sigmoid, and are described by

$$\begin{aligned}\Phi(y_i) &= \tanh(v_i)\\ \Phi(y_i) &= (1+\exp(-v_i))^{-1}\end{aligned} \quad (3)$$

in which the former function is a hyperbolic tangent, which ranges from -1 to 1, and the latter is equivalent in shape but ranges from 0 to 1. Here y_i is the output of the ith node (neuron) and v_i is the weighted sum of the input synapses. More specialized activation functions include radial basis functions which are used in another class of supervised neural network models. The multilayer perceptron consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes. Each node in one layer connects with a certain weight W_{ij} to every node in the following layer. Multilayer Perceptron most commonly seen in speech recognition, image recognition, and machine translation software, but they have also seen applications in other fields such as cyber security.

5. Support Vector Machine Model

A support vector machine (SVM) is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. Support vector machine was first proposed by Vapnik (1998). Its main idea is to minimize upper bound of the generalization error and it maps the input vector into high dimensional feature space through some nonlinear mapping. In this space, the optimal separating hyper plane, which separates the two classes of data with maximal margins, is constructed by solving constrained quadratic optimization problem whose solution has an expansion in terms of a subset of training patterns that lie closest to the boundary. It is been discussed in [19] how SVM has a best method for classification in terms of credit approval process. Sequential minimal optimization (SMO) is an algorithm for efficiently solving the optimization problem which arises during the training of support vector machines. It was invented by John Platt in 1998 at Microsoft Research. SMO is widely used for training support vector machines and is implemented by the popular libsvm tool. The publication of the SMO algorithm in 1998 has generated a lot of excitement in the SVM community, as previously available methods for SVM training were much more complex and required expensive third-party QP solvers.

III. Literature Review

Artificial Neural Network is motivated by human brain which is massively distributed parallel processor. That's why, researchers is trying to find out optimal solution to find solution to the critical problems like making loan decision using ANN. There are many method of neural network that have been applied and being developed for making loan decisions.

The combination of architecture, learning paradigm and learning rules define a particular neural network model [2]. The architecture can be in the form of feed forward, limited recurrent and fully recurrent networks. The learning paradigm can be classified as one of these: supervised, unsupervised and reinforcement. There are different learning algorithms for neural networks, for example: error correction learning, Hebbian learning, competitive learning and Boltzmann learning. So combining the architecture, learning paradigm and learning rules of ANN various method have been developed to solve Bank Loan Decision Problem.

The widespread method for making loan decision is the use of Backpropagation method of Neural Network that represents an adjustment of weight based on smaller decreasing method which is known as Gradient Descent. Back propagation was created by generalizing the Widrow-Hoff learning rule to multiple-layer perception and nonlinear differentiable transfer functions. In Back Propagation the network is first trained with known input output data set. And if the Backpropagation network can be trained properly then it can provide reasonable answer when presented to inputs that were not encountered in the training.

The major limitations of back propagation method are.

a) The Backpropagation is generally very slow because it requires small learning rates for stable learning.

b) The step-size problem occurs because the standard back-propagation method computed only $\partial E/\partial w$, the partial first derivative of the overall error function with respect to each weight in the network.

c) Another problem of back-propagation learning is what we call the moving target problem.

Kohenen Feature Map is another method used for this purpose. The Kohonen Feature Map was first introduced by Finnish professor Teuvo Kohonen (University of Helsinki) in 1982. It is probably the most useful neural net type. The "heart" of this type is the feature map, also called Self organizing Map (SOM) which is a feed forward / feedback type neural network which is build of an input layer. The feature map can be one or two-dimensional and each of its neurons is connected to all other neurons on the map. It is mainly used for classification. The major disadvantage of a Kohenen Feature Map is that it requires necessary and sufficient data in order to develop meaningful clusters. The weight vectors must be based on data that can successfully group and distinguish inputs. Lack of data or extraneous data in the weight vectors will add randomness to the groupings [4].

Glorfeld and Hardgrave (1996) presented a comprehensive and systematic approach to developing an optimal architecture of a neural network model for evaluating the creditworthiness of commercial loan applications. The neural network developed using their architecture was capable of correctly classifying 75% of loan applicants and was superior to neural networks developed using simple heuristics. Tessmer (1997) examined credits granted to small Belgian businesses using a decision tree-based learning approach. Tessmer focused on the impact of Type I credit errors (classifying good loans as bad loans), and Type II credit errors (classifying bad loans as good loans), on the accuracy, stability and conceptual validity of the learning process. Subsequent authors built on the existing research by comparing the performance of various data mining techniques in various credit risk assessment contexts. Desai et al. (1996) analyzed the usefulness of neural networks and traditional techniques, such as discriminant analysis and logistic regression, in building credit scoring models for credit unions. Desai studied data samples containing 18 variables collected from three credit union sand showed that neural networks were particularly useful in detecting bad loans, whereas logistic regression outperformed neural networks in the overall (bad and good loans) classification accuracy. Barney et al. (1999) compared the performance of neural networks and regression analyses in identifying the farmers who had defaulted on their Home Administration Loans and those farmers who paid off the loans as scheduled. Using an unbalanced data, Barney found that neural networks outperform logistic regression in correctly classifying farmers into those who made timely payments and those who did not. Jagielska et al. (1999) investigated credit risk classification abilities of neural networks, fuzzy logic, genetic algorithms, rule induction software, and rough sets and concluded that the genetic/fuzzy approach compared more favorably with the neurofuzzy and rough set approaches.

IV. Nearest Neighbor Classifier Method

In this section we will describe about our solutions and techniques used to solve this problem. We will successively discuss about our Methodology, The technique, Logic behind choosing this technique and its architecture.

A. Methodology

Decision of bank loan allotment requires strong analysis and productive methodology. In the previous section we have discussed a little about our solution that we are going to make bank loan decision using classification method of neural network. That is to classify loan seekers into two categories-

a) Eligible
b) Ineligible

To be categorized the loan seeker among the bank clients; information including personal and financial information is required and all the information is collected from Brac Bank of Bangladesh by special request for use in academic research. After the completion of the collection of information, collected information will be made prepared for applying through neural network is norm of error vector.

B. The technique

Though back propagation method is widely used for classification and problems such as Bank loan decision, but it has some disadvantages that were described previously that can make the whole process quite slow. So we are not going to use this method rather we will show how we can make the decision making process faster finding out the classification using Nearest Neighbor classifier.

Theoretically, in this method, first of all the neural network will be trained with given dataset and known output. Then, after finishing the training session, the neural network will be taken into its application session. And here new unknown data will be provided that was totally different from the data given in the training session. After getting the data neural network will try to find out the distance between the training result and present result. The result which will be minimum will considered as the closest neighbor. And eventually we can consider that that the client who is eligible for granting a loan.

Mathematically, it works based on the Euclidian Distance —

The vector $X^{'}_{n} \in_{,} \{ X_1, X_2, \ldots, X_N \}$ is the nearest neighbor of X_{test} if $X^{'}_{n}$ is the class of $X_{test.}$

As learning and training is essential for every experiment of neural network, we will use supervised learning method which will be conducted with the help of a teacher or supervisor to make the Neural Network learned. In supervised learning the classes are predetermined that facilitates the application of this method.

C. *Why Nearest Neighbor classifier?*

Nearest Neighbor classifier has tremendous advantage in classification. Specially, it has superiority over BackPropagation Method. Also Nearest neighbor methods have the advantage that they are easy to implement. They can also provide quite good results if the features are chosen carefully.

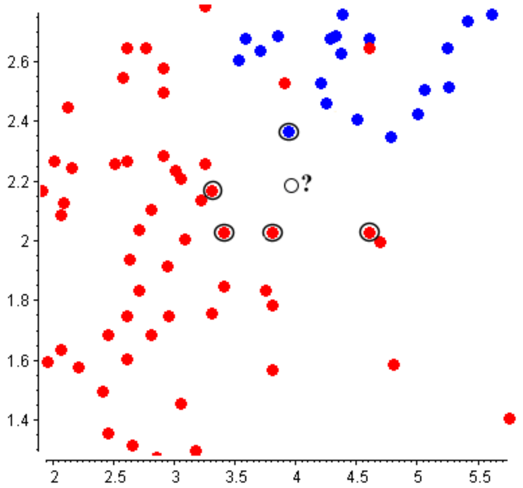

Fig. 1. Nearest Neighbor Classifier concept

The above figure is showing the method of computing class probabilities for the example marked with question mark ("?") using nearest neighbor rule .Since four of its nearest neighbors belong to the red class and only one to the blue, the Neural Network will predict the probability of the example belonging to the red class is 80%.

V. SIMULATION

The data set used in this research is divided into training and testing data sets. All training cases are set by default taking into account the banks' guidelines for personal credit approval in the banks. Data used is of 1000 customer's data. To train or make the Neural Network learned we must collect data. Another interesting thing of Neural Network is that it can learn from noisy, incomplete or distorted sample of data (Glorefeld, 1996). However the train Dataset will consist of the data taken from Brac Bank customers as described before. For fitting the data with Nearest Neighbor Classifier, we have to collect the data in matrix form.

It consists of different independent variables and one dependent variable [23]. Variables are the conditions or characteristics that the investigator manipulates, controls or observes. It is necessary to optimize variables by using SVM as mentioned in [20], [21], [22]. Variables are classified as dependent and independent variables. An independent variable is the condition or characteristic that affects one or more dependent variables: its size, number, length or whatever exists independently and is not affected by the other variable. A dependent variable changes as a result of changes to the independent variable.

Independent Variables

1) Identification number
2) Amount of loan
3) Amount on purchase invoice
4) Percentage of financial burden
5) Term
6) Personal loan
7) Purpose
8) Private or professional loan
9) Monthly payment
10) Savings account
11) Other loan expenses
12) Income
13) Profession
14) Number of years employed
15) Number of years in Bangladesh
16) Age
17) Sex
18) Payment history
19) Home ownership
20) Applicant type
21) Nationality
22) Marital status
23) Number of years since last house move
24) Code of regular saver
25) Property
26) Existing credit info
27) Number of years client
28) Number of years since last loan
29) Number of checking accounts
30) Number of term accounts
31) Number of mortgages
32) Number of dependents
33) Pawn
34) Economical sector
35) Employment status
36) Title/salutation

Dependent variable:

1) Credit (Eligible or Ineligible)

Using the neural network node, we can construct, train, and validate a network. In this method we can directly use the credit information for example the customer is 90% eligible for granting loan ,otherwise we can set a threshold to credit for eligibility or ineligibility to grant a loan and then we can get binary output eligible or ineligible.

For our convenience we used use Matlab, one of the most leading and broadly used software to simulate Neural Network experiments. There is a class of Matlab which is known as knnclassify that we used to implement and make classification of the data.

Knnclassify helps efficiently classify data after being trained with the given data.

The formal structure of this class is:

Knnclassify (Sample, Training, Group, k, distance, rule)

It has predefined parameters those are used for taking input and providing desired output.

Sample: Matrix whose rows will be classified into groups. Sample must have the same number of columns as Training.

Training: Matrix used to group the rows in the matrix Sample. Training must have the same number of columns as Sample. Each row of training belongs to the group whose value is the corresponding entry of group. Group Vector whose distinct values define the grouping of the rows in Training. Kth number of nearest neighbors used in the classification. Default is 1.

Distance: String specifying the distance metric. Choices are:

- 'Euclidean' — Euclidean distance (default)
- 'city block' — Sum of absolute differences
- 'cosine' — One minus the cosine of the included angle between points (treated as vectors)
- 'correlation' — One minus the sample correlation between points (treated as sequences of values)
- 'hamming' — Percentage of bits that differ (suitable only for binary data)

Rule: String to specify the rule used to decide how to classify the sample. Choices are:

- 'nearest' — Majority rule with nearest point tie-break (default)
- 'random' — Majority rule with random point tie-break
- 'consensus' — Consensus rule

The dataset we use from the Brac Bank that is like following:

Classifying Rows:

The following example classifies the rows of the matrix sample:

```
sample = [.9 .8;.1 .3;.2 .6]

sample =     0.9000     0.8000
             0.1000     0.3000
             0.2000     0.6000

training = [0 0;.5 .5;1 1]

training =          0           0
               0.5000      0.5000
               1.0000      1.0000

group = [1;2;3]

group =      1
             2
             3
class = knnclassify(sample, training, group)

class =     3
            1
            2
```

Row 1 of sample is closest to row 3 of training, so class (1) = 3. Row 2 of sample is closest to row 1 of training, so class (2) = 1. Row 3 of sample is closest to row 2 of training, so class (3) = 2.

Classifying Rows into One of Two Groups (Eligible & Ineligible)

The following example classifies each row of the data in sample into one of the two groups in training. The following commands create the matrix training and the grouping variable group, and plot the rows of training in two groups.

```
training = [mvnrnd([ 1  1],   eye(2), 100); ...
            mvnrnd ([-1 -1], 2*eye (2), 100)];
group = [repmat(1,100,1);  repmat(2,100,1)];
gscatter(training(:,1),training(:,2),group,'rb','+x');
legend('Training group 1', 'Training group 2');
```

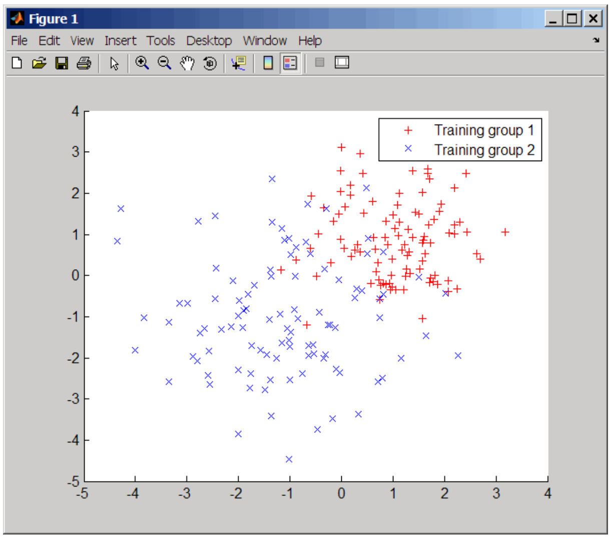

Fig. 2. Classifying into two groups

The following commands create the matrix sample, classify its rows into two groups, and plot the result.

```
sample = unifrnd(-5, 5, 100, 2);
```

% Classify the sample using the nearest neighbor classification

```
c = knnclassify(sample, training, group);

gscatter(sample(:,1),sample(:,2),c,'mc');  hold on;
legend('Training group 1','Training group 2', ...
     'Data in group 1','Data in group 2');
```

VI. Performance Analysis

The overall results of the study are presented in Table 1 and clearly suggest that MLP performance in classifying loan applications can be further improved by nearest neighbour classifier models. In particular, Ensemble Averaging has been shown to be able to reduce the percentage error due to the bias-variance problem inherited by MLP model. However, on average, the improvement ensemble averaging brings on negative data classification is not great. This is evidenced by marginal 0.09% improvement on negative data. The performance on all data was more convincing with 0.32% improvement. While ensemble averaging was able to produce lower percentage errors, it did so at the cost of training time, as evident in the training time for this model compared to other models. Furthermore, the results indicate that Boosting by filtering outperformed other models in this study. It was able to improve the performance of MLP model by 0.49% on negative data and 0.73% on all data. The boosting by filtering committee machine was the best performer in this experiment, in that it was able to produce the least percentage error. This was done at a comparatively low training time cost.

Table 1. Comparative Model Performance

Neural Network Model	Number of Epochs	Percentage Error on Negative Data	Percentage Error on All Data	Average Training Time
Multi Layer Perceptron (MLP)	600	1.81 %	2.38 %	13 sec
Ensemble Averaging (10 MLPs)	6800	1.73 %	2.06 %	245 sec
Boosting by Filtering (3 MLPs)	1500	1.32 %	1.65 %	32 sec
MPL with Bacpropagation	5000	0.85%	0.45%	300sec
MLP with Nearest Neighbour Classifier	800	0.90%	0.30%	15 sec

MLP with backpropagation has percentage error produced by neural network models in this experiment (0.85% average percentage error on negative data and 0.45% average percentage error on all data but it has higher learning time as 300sec and too slow.

The MLP with nearest neighbour classifier has percentage error produced by neural network models in this experiment (0.90% average percentage error on negative data and 0.30% average percentage error on all data but it has higher learning time as 15sec, shows that training different experts on hard to classify applications brings a significant performance improvement.

Nearest neighbour classifier also shows that these performance improvements can be achieved at a low costs(less training time and computational cost).

VII. Discussion And Limitation

This technique and implementation can dynamically response with the change in the characteristics and behavior of the customers and can provide best solution to the bank officers to make the best choice for granting loan. The major limitation of the nearest neighbor method is that, it suffers from curse of dimensionality [6]. It is the problem where the accuracy of Nearest Neighbor method deteriorates due to high dimension of space. This is because; in high or large dimensional space all points are located at a great distance from each other, so the nearest neighbors are not exposed as similar rather dissimilar.

Another problem is that, if the distance function is not chosen perfectly, the systemmay provide random result.

Another disadvantage of this method is that we need lots of training samples to ensure lots of vectors are within the ‘k’ sphere. And we need to have all training vectors in memory at all times.

VIII. Conclusion

This proposed system can ensure reliability to the officer of the commercial bank all over the world. Indeed this method of neural network works well for finding who are eligible and ineligible. As this method has some problems and limitations like curse of dimensionality and some other minor problems those were described, these problems will be solved gradually. There are hundreds of neural network methods. Using the brute force approach we can reduce the problem of nearest neighbor. Moreover if the data can be represented as points in a high dimensional space, then the points will be located close to a subspace of low dimensionality and in this case Nearest Neighbor method will work well.

References

[1] Daniel Sabet and Ahmed S. Ishtiaque,”Understanding The Hallmark-Sonali Bank Loan Scandal , pp.2”, “University of Liberal Arts of Bangladesh”, January 2013.

[2] Meliha Handzic, Felix Tjandrawibawa and Julia Yeo,” How Neural Networks Can Help Loan Officers to Make Better Informed Application Decisions, pp.2”,” The University of New South Wales, Sydney, Australia”, ”June 2003”.

[3] C. N. Dragotă,” The Prediction of Bankruptcy Using Backpropagation Algorithm for “IO” Model Analysis, pp.11”,” Babeş – Bolyai” University, Cluj – NapocaRomania”,”January 2007”,”section 2”.

[4] Kevin Pang,” Self-organizing Maps”,” Neural Networks”,”Fall-2003”,.

[5] Meliha Handzic, Felix Tjandrawibawa and Julia Yeo”, How Neural Networks Can Help Loan Officers to Make Better Informed Application Decisions”,” The University of New South Wales, Sydney, Australia, pp.3”,”June-2003”.

[6] Charles Elkan” Nearest Neighbor Classification, p.p. 3”,elkan@cs.ucsd.edu”,” January 11, 2011.

[7] Bishop, C. M. 1995. Neural Networks for PatternRecognition. Oxford University Press, Oxford, U.K.

[8] Desai, V. S., J. N. Crook, G. A. Overstreet Jr. 1996.comparison of neural networks and linear scoring models in the credit union environment. Eur. J. Oper. Res. 95(1).

[9] Nauck, D. 2000. Data analysis withneuro-fuzzymethods. Habilitation thesis, University of Magdeburg, Germany.

[10] Capon, N. 1982. Credit scoring systems: A critical analysis. J. Marketing 46 82–91.

[11] West, D. 2000. Neural network credit scoring models. Comput. Oper. Res. 27 1131–1152.

[12] Gately E. J., Neural Networks for Financial Forecasting, WIG-Press, Warszawa, 1999 (in Polish).

[13] Rahimian E., Singh S., Thammachote T., Virmani R, "Bankruptcy Prediction by Neural Network",Probus Publishing Company, Chicago- London, 1993, pp. 159 – 176.

[14] M. Handzic, F. Tjandrawibawa , and J. Yeo, “How/neuralnetworks can help loan officers to make better informed applications decisions,” in Proc. 2003 Informing Science+ IT Education Conference 2003, pp. 97-108.

[15] J. E. Boritz and D. B. Kennedy, “Effectiveness of neural Network Types for prediction of business failure,” Expert Systems with Applications, vol. 9, no. 4, pp. 503-512, 1995..

[16] M. Bensic, N. Sarlija, and M. Zekic-Susac, “Modeling small-business credit scoring by using logistic regression, neural networks and decision trees,” international Journal of Intelligent Systems In Accounting, Finance And Management, vol. 13, no. 3, pp.133-150, July2005.

[17] J. Zurada and M, Zurada., “How Secure Are “Good loans”: validating loan-granting decisions and predicting default rates on consumer loans,” The review of business information systems, vol. 6, no.3, pp. 65-83, 2002.

[18] H. G. Nguyen, “Using neural network in predicting corporate failure,” Journal of Social Sciences, vol.1, no. 4, pp.199-202, 2005.

[19] C. L. Huang, M. C. Chen, and C. J. Wang, “Credit scoring with data mining approach based on support vector machines,” Expert systems with applications, vol. 33, no.4, pp. 847-856, November 2007.

[20] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies,” in Feature extraction, foundations and applications, New York: Springer, 2005, pp.319-328.

[21] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” Journal of machine learning research, vol. 3, pp. 1157-1182, 2003.

[22] Y. Q. Wang, “Building credit scoring systems based on support-based support vector machine,” in Proc. Fourth international conference on natural computation October 2008, pp. 323-327.

[23] Bart Baesens,Rudy Setiono,Christophe Mues and Jan Vanthienen,"Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation", Management Science , Vol. 49, No. 3, March2003 pp. 312–329.

Authors’ Profiles

Md.Mahbubur Rahman has been lecturing in CSE since mid of 2011, he received his B.Sc.Engg. in CSE from Patuakhali Science and Technology University in 2011and continuing his M.Sc. Engg. in CSE at Bangladesh University of Engineering and Technology(BUET), Bangladesh. He is now serving one of the top most private Universities in Bangladesh named Bangladesh University of Business and Technology (BUBT). His research interests are Digital Forensics,Secure and Trustworthy computing, Data mining, Graph theory, Biometric system.

Samsuddin Ahmed has been lecturing in CSE since mid of 2010. He is in Computer Science and Engineering from University of Chittagong with highest CGPA till date. His under-grade Thesis was on “Handling Uncertainties in Spatial Feature Extraction”. His hobbies include thinking about underlying mathematical formulations in natural phenomena. His research interests include data and image mining, Semantic Web, Business Intelligence, Spatial Feature Extraction etc. He is now serving one of the top most private Universities in Bangladesh named Bangladesh University of Business and Technology (BUBT).

MD. Hossain Shuvo is an undergraduate student, currently cintinuing his B.Sc in Computer Science & Engineering (CSE) from Bangladesh University of Business & Technology (BUBT), a reputed university of Bangladesh. He has recently involved in research. His research interests are Neural Network, Data Mining, Algorithms and Theory of Computation and Bioinformatics. He is now working with estimation and decision making model of Neural Network.

How to cite this paper: Md.Mahbubur Rahman, Samsuddin Ahmed, Md. Hossain Shuvo,"Nearest Neighbor Classifier Method for Making Loan Decision in Commercial Bank", International Journal of Intelligent Systems and Applications(IJISA), vol.6, no.8, pp.60-68, 2014. DOI: 10.5815/ijisa.2014.08.07