Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2001 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2001

Business Strategies in Credit Rating and the Control of Misclassification Costs in Neural Network Predictions

Dat-Dao Nguyen
California State University at Northridge

Dennis Kira
Concordia University

Follow this and additional works at: http://aisel.aisnet.org/amcis2001

BUSINESS STRATEGIES IN CREDIT RATING AND THE CONTROL OF MISCLASSIFICATION COSTS IN NEURAL NETWORK PREDICTIONS

Dat-Dao Nguyen
California State University at Northridge
datdao.nguyen@csun.edu

Dennis S. Kira
Concordia University
kira@vax2.concordia.ca

Abstract

This paper reports on an investigation into the control of misclassification costs in Artificial Neural Network (ANN) prediction. Given the availability of a small historical data set and without imposing any strong assumptions on the behavior of the underlying variables, this study investigates the enhancement of ANN prediction with a training set containing the replication of interested cases. With this replicated training set, the ANN will learn more about patterns of a particular category and then improve the accuracy in predicting the membership of this category. In a context of small business loan application, depending on the cost to the financial institution in the case of misclassification, one may consider replicating the accepted or rejected cases in the training set in an appropriate proportion. Consequently, one can achieve a more accurate ANN prediction on the interested category.

Introduction

Small business loans account for a large part in the commercial loans provided by financial institutions. An inaccuracy in loan evaluation, i.e., classifying an application as good instead of bad and vice versa, may lead to the rejection of a prospective good application or acceptance of a bad loan. Consequently, the institution may suffer either an opportunity loss or an actual loss. Therefore the evaluation/classification of business loan applications deserves a primary attention from the financial community. In traditional statistical methods, one may control for the probability of misclassification with an optimal sample size. In Artificial Neural Network (ANN) technique, it is desirable to have a large sample providing more information on the problem space for better approximation and prediction. However, in many situations, one is constrained by the availability of information and the unfeasibility in calculation of optimal sample size without imposing strong assumptions on the probability distribution of the data.

This paper reports on an investigation into the control of misclassification costs in ANN prediction. Given the availability of a small historical data sample and without imposing any assumptions on the behavior of the underlying variables, this study investigates the enhancement of ANN classification with a training set containing the replication of interested cases. Depending on the associated cost to the institution when the misclassification of an application is good or bad, one may consider replicating the accepted or rejected cases in the training set in an appropriate proportion in order to achieve a more accurate prediction of the interested category.

The paper is organized as follows. Section 2 reviews the evaluation/classification of small business loans with ANN is presented. Section 3 presents the investigation of using data replication for controlling misclassification costs of ANN predictions. Section 4 discusses the implications of findings. The paper concludes with some remarks on the implementation of the proposed method and its extension in future research.

Small Business Loan Evaluation and Classification with Artificial Neural Network

Many models, using qualitative as well as quantitative explanatory variables, have been developed to assist credit and loan officers in the evaluation of small business loan applications. Quantitative information is taken from standard financial statements (Altman 1983; Orgler 1970), whereas qualitative information relates to judgments on the quality of management and the prospective of the business in the market (Doreen and Farhoomand 1983). These studies use traditional statistical methods such as Discriminant Analysis and Logistic Regression, to model and predict the loan appraisal. It is noted that, in these methods, strong assumptions are imposed on the relationship and probability distribution of the underlying variables. Furthermore, the nonlinear relationship has not been taken into consideration in previous modelling.

Recently, Artificial Neural Network (ANN) (Rosenblatt 1959, 1962; Hassoun 1995) has emerged as a powerful technique in pattern recognition and function approximation. The strength of an ANN lies in its nonlinear, nonparametric approach in data modelling. An ANN consists of nodes as autonomous processing units connected by directed arcs and arranged into layers. Every node, other than input node, computes its output *S* as a function of the weighted sum of inputs directed to it from other nodes,

$$S_i = \sum_j^n w_{i,j} \, u_j \qquad\qquad (1)$$

$$u_i = f(S_i) \qquad\qquad (2)$$

where $f(.)$ is a transfer function, usually a nonlinear, bounded and piecewise differentiable function, such as the sigmoid function

$$f(x) = 1/(1 + e^{-x}) \qquad\qquad (3)$$

Such an ANN produces a response, which is the superposition of *n* sigmoid functions, where *n* is the number of hidden nodes, to map a complex function. As one adds more hidden layers, ANN will be able to map higher order functions. Therefore, the function mapping with ANN is more general than the regression of traditional methods.

It has been proved that ANN can be considered as a universal approximator for any functional relationship. Hornik et al. (1989) show that standard multi-layer feedforward networks using any arbitrary transfer function can approximate any Borel measurable function to any desired degree of accuracy. Specifically, Cybenko (1989) proves that, by using the backpropagation training algorithm (Werbos 1974) and a sigmoid transfer function, an ANN with one hidden layer can approximate any continuous function.

Many successful applications of ANN in finance and business has been reported in the literature. Kira et al. (1997) implement ANN in the qualitative evaluation of small business loans and find that ANNs perform as satisfactory as traditional statistical Logistic Regression and Discriminant Analysis. However, they suggest that the good performance of ANNs in predicting the membership of a new case to a particular category (accepted/rejected, success/failure) may be due to the abundance of information on the patterns that the network has learned from the cases belonging to this category.

The misclassification in loan evaluation may lead to costly errors in decision-making. A Type I error occurs when a good loan application is incorrectly classified as bad and being rejected, whereas a Type II error occurs when the bad loan application is incorrectly classified as good and not being rejected. If a Type I error is committed, the institution may suffer an opportunity loss since it does not earn profit from an otherwise healthy lending. But a Type II error will cause an actual loss if the institution lends money to a bad enterprise. The concern of any lending institution is to control the errors in discrimination and therefore the misclassification costs. In particular, the question is, given the available historical cases providing limited information, how one can control the degree of accuracy in classification of future cases.

Controlling Misclassification Costs of ANN Predictions

It is apparent that a large sample will provide more information on a problem space for better approximation and prediction. With an optimal sample size in traditional statistical methods, one may control for the probability of making Type I and Type II errors. However, in many actual situations, one is constrained by the availability of information. Furthermore, the calculation of optimal sample size may not be feasible unless one imposes strong assumptions on the probability distribution of the data.

Given the availability of a small data sample and without imposing strong assumptions on the probability distribution of the underlying variables, this study investigates the enhancement of ANN accurate prediction. To control the misclassification costs in ANN, this study proposes to replicate the patterns of a particular type in the training set in order to provide the ANN with the opportunity to recognize more of these peculiar patterns. This undertaking is in accordance with the desired result from any discriminant analysis, in which the goal is not to minimize the overall misclassification rate but to identify most cases of the interested category, usually the "minority" or "rare" one.

This study re-analyzes the data set on small business loan appraisals collected by Doreen and Farhoomand (1983). The sample contains 150 judgments of loan applications based on 27 qualitative criteria related to the evaluation of management, earning potential, security, and market environment. In the experiments reported herein, the data set is randomly split into three sets. The first set, referred to as the *training set*, containing 108 cases (81 accepted cases and 27 rejected cases) is used to train the network. The second set, referred to as the *test set*, containing 27 cases (18 accepted cases and 9 rejected cases) is used to measure the accuracy of the trained network. The third set, referred to as the *validation set*, containing 15 cases (10 accepted cases and 5 rejected cases) is used to measure the performance of the network in predicting out-of-sample data. One notes that, in the training set, the frequency of the accepted cases is about three times more than the one of rejected cases. Another training set, referred to as the *replicated set*, is prepared in which each rejected case is replicated 3 times so that the frequency of rejected cases is equal to the frequency of the accepted cases. Then one investigates whether ANNs predict the rejected cases more accurately once they are trained with the replicated set.

In this study, ANNs are trained in the Functional Approximation and Classification modes. In Function Approximation, the ANN approximates the functional relationship between input-output patterns and produces a single output. Depending on whether the network output is above or below a threshold value, the related case will be designated as accepted or rejected. This is similar to Logistic Regression in traditional statistics. In Logistic Regression, one estimates the parameters of a logistic function in the form

$$Y = 1/(1 + e^y) \qquad (4)$$

$$y = b_0 + \sum b_r X_i \qquad (5)$$

where Y is between [0,1] or the probability of the class outcome, y is a linear combination of X_i explanatory variables.

In Classification, the network has two output nodes, one for each related class (acceptance/ rejection). Depending on whether the network output is above or below a threshold value, the case will be assigned as belonging to the related class or not. This is similar to Discriminant Analysis in traditional statistics. In Discriminant Analysis, one constructs a linear discriminant model in the form

$$D = b_0 + \sum b_i X_i \qquad (6)$$

where D is the discriminant score, b_i is a discriminant weight, and X_i is a set of explanatory variables.

In this study, the threshold value is set at .50 . Since the ANN outputs are continuous scores from [0, 1], if the predicted score of a particular case is greater than .50, it will be classified as accepted and vice versa.

A Genetic Algorithm (GA) (Holland 1975; Goldberg 1989) is used to identify an optimal topology for a one-hidden-layered backpropagation ANN. The GA is to search for the most appropriate transfer function among the family of logistic, hyperbolic tangent and linear functions. It also searches for the optimal number of nodes in the hidden layer. This study uses all 27 input variables of the data set as it has been found that the performance with the full information of the problem space is superior to the one with parsimonious data (Kira et al. 1997). With the optimal topologies identified by GA, these networks achieve a degree of accuracy ranging from 90% to 100% in the test set. Once the ANNs predict accurately on the test set, they are used to generalize over the validation set.

In the *Function Approximation* mode, training with the original data set, the GA identifies an ANN topology having 5 hidden nodes using logistic transfer function and 1 output node using hyperbolic transfer function. This network correctly predicts 3 out of 5 rejected cases and 8 out of 10 accepted cases of the validation set.

Using the replicated training set, the ANN is trained in the same topology that has been identified with the original data set. One also builds an optimal topology to learn the patterns of the replicated set and to test for other possible enhancement. However,

the performance of this ANN learning is not improved as it attains the same degree of accuracy in correctly predicting 3 out of 5 rejected cases and 8 out of 10 accepted cases of the validation set.

In the *Classification* mode, training with the original data set, the GA identifies an ANN having 1 hidden node using hyperbolic tangent transfer function and 2 output nodes using linear transfer functions. This network correctly predicts 3 out of 5 rejected cases and 8 out of 10 accepted cases of the validation set.

Using the replicated training set, the ANN is trained in the same topology identified with the original data set. This network correctly predicts 4 out of 5 rejected cases and 8 out of 10 accepted cases. One also uses GA to search an optimal network for training the replicated set. The GA arrives at an ANN topology of 1 hidden node using linear transfer function and 2 output nodes using linear transfer function. This optimal network correctly predicts all 5 rejected cases and 9 out of 10 accepted cases of the validation set.

In order to investigate further the effect of data replication on the accuracy of ANN prediction of the rejected category, the rejected cases in the original training set are replicated 6 times such that the frequency of rejected cases is twice as many as the one of accepted cases. In this experiment, the optimal topologies of ANNs obtained from the original data are held fixed so that any change in the performance of ANNs would be due to the new replicated training data. Results of this experiment show that there is no change when ANN is trained in the Function Approximation mode. However, in Classification mode, ANN correctly predicts all 5 rejected cases and has no improvement on prediction on accepted cases, i.e., correctly predicts 8 out of 10 accepted cases.

Discussion

In a classification problem, one should not only examine the overall error rate, which is the relative magnitude of total number of errors in all categories to total number of test cases. It is more informative if one can evaluate the misclassification error for a particular category. Such an error could be assessed by cost-benefit analysis with criteria such as actual, potential costs, risk (gain/loss) and/or utility (convenient/inconvenient). The merit of a classification method -- a classifier -- is not solely based on its ability to provide an overall low error rate. In many situations, one may have a particular interest in the correct identification of the "rare" cases rather than the common cases. Therefore, one may be interest in the ability of a classifier in correct classification of the smaller group. The implication is that with an available historical data set, one should be able to build an effective classifier to support the decision-making in context.

In discriminant analysis of traditional statistics, results of different classification rules for rare cases can be identified with the ranking of all cases in the modeling set on their discriminant scores. Based on how many rare cases in various deciles, one can arrive at an appropriate rule for identifying these cases at the expense of increasing the number of misclassification of the common group (Norusis, 1990). Consequently, one increases or decreases the threshold such that a predicted score has more chance to belong to an interested group. In contrast, the ANN classification using replicated data in this study does not in any instance affect its ability in dealing with the larger group, but it does have a remarkable effect on the accurate classification of the rare cases. One notes that the data replication, in particular of the smaller group, does not have any effect in traditional statistical methods, as these methods are in the function mapping (regression) framework.

By replicating patterns of the interested group, this study observes an enhancement in the control of misclassification costs and therefore the accuracy of network prediction. The data replication is particularly effective when ANNs are trained in the *Classification* mode. In this mode, the abundance of replicated data points gives more weights to the features of an interested category. In contrast, in the *Function Approximation* mode, many points with a same coordinate in R^n are mapped into a unique coordinate in R^m. Therefore, the abundance of replicated information does not contribute to the enhancement of function mapping. This study finds that the prediction with ANN in Classification mode is superior to the Function Approximation mode. The reason is that certain features unique to each of many categories cannot be easily captured in a single function.

This study uses a threshold of .50, with a tolerance margin of .05 in favor of Type II error, to designate the related cases into the appropriate category. However, one may arbitrarily introduce a certain degree of tolerance by setting a threshold in favor of Type I or Type II error in decision-making. For instance, if the threshold value is set at .75 in favor of Type II error, all cases with predicted scores below the threshold will be classified as rejected. In a same manner, one may favor Type I error by setting a lower threshold to include more accepted cases in the prediction. There is a trade-off between cost and benefit in setting such arbitrary thresholds. Since each loan application asks for a particular amount of money, one cannot base the cost-benefit analysis solely on the number of cases that have been correctly predicted as good or bad. Depending on the decision context and business strategies, the responsibility of the management is to assess the associated monetary cost/benefit in committing a particular type

of error. If the institution is more conservative, it will seek to minimize the Type II error. As such, it would rather loose an opportunity by sending away a good prospective customer to its competitors than suffer an actual lost with the commitment to a seemingly bad enterprise. In contrast, with an aggressive lending policy, the institution might seek to minimize the Type I error by seizing as many business opportunities as possible.

Conclusion

This study has shown that one may enhance the control of misclassification costs and the accuracy of network prediction by arbitrarily replicating patterns of the interested group. With this replicated training set, the ANN will learn more about patterns of a particular category and improve the accuracy in predicting the membership of this category. Although the study is conducted in the context of small business loan applications, the proposed method can be extended to any classification problems in which one is constrained by the availability of information on a particular group.

In this study, the costs of making Type I and Type II errors are assumed to be equally important. From this point, depending on the policy of a lending institution, one can replicate proportionally the relevant patterns in order to increase correspondingly the presence of the good or the bad cases in the ANN training set. Furthermore, one can set a particular classification threshold in favor of Type I or Type II error in decision-making.

In this study, the complexity of the classification problem and the one of ANN topology are controlled to provide some insights on the effect of data replication in making correct prediction. In a future work, we shall address the effect of replicated data in relation to the complexity of the problem domain and network topology to find out the conditions in which the replication of data is fruitful in providing more accurate ANN classification.

References

Altman E.I. *Corporate Finance Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*, John Wiley, New York, 1983.

Cybenko G. "Approximation by Superimpositions of A Sigmoid Function," *Mathematics of Control, Signals, and Systems,* 2, 1989, pp. 303-314.

Doreen D. and Farhoomand F. "A Decision Model for Small Business Loans," *Journal of Small Business*, Fall-1983.

Goldberg G. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading MA, 1989.

Hassoun M.H. *Fundamentals of Artificial Neural Networks*, The MIT Press, Cambridge, MA, 1995.

Holland J. *Adaptation in Neural and Artificial Systems*, The University of Michigan Press, Ann Arbor, MI, 1975.

Hornik K., Stichcombe M. and White H. "Multi-Layer Feedforward Networks Are Universal Approximators," *Neural Networks*, 2, 1989, pp. 359-366.

Kira D., Doreen D. and Nguyen D.-D. "A Qualitative Evaluation of Small Business Loans: Using Artificial Neural Networks and Traditional Statistical Methods in Model Building and Prediction," *Proceedings of SCI 1997*, 2, 1997, pp.222-228.

Norusis M.J. *SPSS Advanced Statistics User's Guide*, SPSS Inc. Chicago, IL, 1990

Orgler Y.E. "A Credit Scoring Model for Commercial Loans," *Journal of Money, Credit and Banking*, Nov. 1970, pp. 435-445.

Rosenblatt F. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review*, 65, 1959, pp.386-408.

Rosenblatt F. *Principles of Neurodynamics*, Spartan, New York, 1962.

Werbos P. *Beyond Regression: New Tool for Prediction and Analysis in the Behavioral Sciences*. Ph.D. Thesis, Harvard University, 1974.