



Adaption of Genetic Algorithm Technique in Optimization of Decision Tree In Order To Validate Customer 'S Banking System. (Melat Bank Branches of Mahabad City)

Ayub Nahardah, Kamal Khalilpour*

Department of Public Administration, Mahabad Branch, Islamic Azad University, Mahabad, Iran

* Corresponding Author email: kamal.khalilpour@yahoo.com

Abstract: Decision tree as one of data mining technique can validate bank customers. Main issue is the construction of a decision tree that can optimally classify customers. In thesis, presented a suitable model validation bank customers for providing credit facilities proportionate with each class based on genetic algorithm. Genetic algorithms by choosing the right features and making the optimal decision tree can help validation of customers. In pattern recognition and used process for CRISP validation of customers. The proposed classification model is based clustering techniques, characters selection, decision trees and genetic algorithm. This model tries to select and combine the best decision trees based on optimal standards and making the final decision tree to validate customers. The results show that accuracy of classification of model proposed classification is higher than all the decision tree models in this thesis. Also number of leaves and the size of decision tree and also the complexity of it are lower than all the other aspects of it.

Keywords: validation, genetic algorithm, characters selection, decision tree, clustering

Introduction

Banks in order to awareness of the needs and the customer behavior in providing credit must identify their characteristics. This leads to decrease in bank risks including credit risk. Research and different application in the validation field is done to identify good and bad bank customers judgment method in validation because of the error and longtime gradually change its place to parametric and nonparametric method. Parametric methods like probit, lojic, differential analysis and logistic regression used at the beginning of validation, and nonparametric method and data mining like decision tree, neural network, and experienced systems. Decision tree is one of the data mining technique with high intelligibility and appropriate speed in learning pattern can be useful in validation of classifies customers. Now days the issue of validation has become an important issue of managers and bank experts. In a bank we can build different decision trees to classification and validation of customers.

Increased accuracy in structure and content of decision tree by genetic algorithm, improving decision tree for easy understanding of taking wrong probable decision of bank experts in validation, reduction of need to costly and time consuming analysis in classification of customers. Optimal selection in characteristics of customers validation and finally their satisfaction in presenting proportional credit facilities with each class cussing genetic algorithm techniques in optimizing the decision tree and selection of characteristics in addition to resolve the issues raised has this advantages: genetic algorithm at the same time works with a set of solutions but inference algorithm greedily examines partial solution at every stage.

The main issue is building decision tree that can classify good and bad bank customers optimally. It seems using optimization techniques like genetic algorithm is useful in character selection and making optimal decision trees for validation of bank customers. Character selection algorithms may placed in local optimum. On the other hand don't consider interaction between features and imagine, relations between features are liner and the characters are independent algorithm of decision trees are greedy in growth

process. In some kind of algorithms there is more sensitive toward some characters for dismantling. They build large and complicated trees that the result has low accuracy in classification of test data collection and excessive evaluation. Decision trees compared to the large amount of data show interest in features and in lower parts the tree doesn't like to increase accuracy, and reducing so much proportion. This algorithm in its process just use some of criteria and functions to make decision tree.

In this essay a suitable model for validation of banks customer is presented like Melat bank for providing suitable credit with each class based of genetic algorithm.

Review of literature

Various studies has been done on the parametric and nonparametric in validation. The following are some examples that has been done abroad.

Two examples of research on validation methods abroad:

Installation differentiation analysis and algorithm after publication in neural network by Lee, Chiu, Lu and Chen in 2002.

Incorrect classification scheme Kim and Sohn in 2004

Two examples of researches in applying genetic algorithm in characters selection:

Using genetic algorithm for choosing entrance variables by D'heygere, Goethals and Pauw int 2003.

Genetic algorithm in choosing variables by using customers clustering by Liu and Ong in 2008.

Two examples of research in applying genetic algorithm in making decision trees.

Investigation of decision tree technique profits based on genetic algorithm by Huang, Gong, Shi, Liu and Zhang in 2007.

Optimization predictions model based on decision trees and neural networks by D'heygere, Goethals and Pauw in 2006.

Strengths of presented examples, in contrast with some internal and external similar examples and external similar examples and in case includes the following:

Using preparation and data preprocessing methods.

Using clustering in preprocessing of data in order to possible increase accuracy and decreasing complexity in customers classification.

Combination of several character selection algorithm based of filtering approach Wrapper and embedded design for increasing flexibility and classification accuracy in building decision tree instead of using just one class.

Production and comparison various decision tree and using them in different condition of case study against other research that uses just one kind of algorithm or model special study.

Optimization of decision tree by genetic algorithm unlike previous research that build the only classification model and don't try to improvement.

Using of complexity pattern in optimal decision tree other than accuracy classification pattern.

Using artificial intelligence methods and pattern recognition in customers validation bank system to deal with complicated conditions and in terms of non- linear relationships in customers classification and choosing character of validation.

Existence process attitude in pattern recognition and data mining in customers validation of bank system against some other research about this field.

Customer credit characteristics in the model

Total customer's credit features receiving facilities. "code type of record" character is the target feature .

Date of contract (nominal), core code (nominal), number of branch (nominal), number of request (nominal), genre apparent (nominal), local code (nominal), customer name (nominal), ID number (nominal), date of birth (nominal), location register cod (nominal), area cod (nominal), type of request code (nominal), the first due date (nominal), the last due date (nominal), number of installments (numerical), percent of bank share (numerical), contract cost (numerical), how to use the facilities (nominal), amount used (numerical), type security code (nominal), value of collateral (numerical), last remind debt (numerical), type of part code (nominal), activity cores (nominal), type of contract code with bank (nominal), the purpose of receiving facility (nominal), the place of using facility (nominal), the case of using facility (nominal), record type assignment or nonassignment (nominal), date of record create (nominal), date (month, year) (nominal), name and surname of the reversion (nominal), credit code

(nominal), type of record code (nominal), the season account code (nominal), and the last turnover of account (nominal).

Operational definition of concept model

Decision tree: decision tree includes some parts:

Nodes with the specified name or object properties have been labeling.

Leaves that are equal to different classes. A decision tree includes several inner nodes and several leaf. All the inner nodes includes two or several child each decision place in one node. The last node shows the final exit that in decision tree has discrete value.

Optimal decision tree: in the study of decision tree it is considered to be three factors:

Sample rate of correct classification (percentage of correctly classified)

Sample rate of correct classification

$$CCI = \frac{(TP + TN)}{(TP + FP + FN + TN)} \times 100$$

Complexity of decision tree that includes number of leaves and size of tree.

Number of predictable features available in decision tree.

Size of the tree: relate to the numbers of branches to reach the leaves in decision tree and equal to the sum of number of leaves and nodes in decision tree.

Customers class in target characteristics

Includes current customers, past due date and outstanding claims. Current customers are customers that haven't the failure to pay behavior of received facilities. The past three months of the repayment period of credit facilities. Claim outstanding customers are those past more than six months of repayment period.

Case study

Because of confidentiality of bank information it is refused to offer descriptive statistics cast from the bank.

The population in this thesis are 675 credit customers of Melat bank.

Research operational model shows the proposed model that refers to building process and decision tree's test in validation of bank customers.

Hence the first three stages performed in pattern recognition, preprocess of data done. We can use clustering method for preprocess of data. Due to the operational research model Flowchart.

Total steps in building the proposed classification model is like the analysis research model.

In choosing characters, searching method, is haphazard and based of genetic algorithm. Feature selection algorithm based on approach of filter assess the following in set of features: integrity between features together and with target feature, compatibility of subset features with supplies in target feature, C4.5 accuracy classifier selection feature algorithm with the evaluation function Wrapper with the C4.5 classifier based on Wrapper decision tree: because this algorithm uses a classified algorithm like genetic decision tree for selecting suitable features, based on embedded design approach in selection features.

In selecting features based on filter approach and Wrapper by genetic algorithm we are using coded pattern zero and one for coded chromosomes (set of feature). One and zero respectively show presence and nonpresence of one feature in the set of features. Set of chromosomes are produced haphazardly. After this phase it is time to assess all chromosomes by assessor function.

Premier chromosomes selection is based on merry – go- round way from the petals and for reproduce, intersection operations and random mutation take place on them. At first determined a probable number, after that the algorithm allocates a haphazard number to each couple of chromosomes, if this number be more than probable number predetermined, single- point crossover zero and one take place from the petals. Then single – point mutation act zero and one applies on the new chromosomes.

In the following new chromosomes have made and again by assessor function have assessed and best of the replaced by weaker chromosome from previous generation. Kind of assignment condition of this algorithm is reaching to certain number of times of algorithm with maximum number of generation.

In the feature selection methods parameter values are as follow: number of cross validation is equal to 10, intersection rate 0.9, mutation rate 0.01, number of generation and initial population 20 and random number Seed equal to 1 is considered. Number of categories and number and threshold in selecting feature algorithm with evaluation function Wrapper with C4.5 classifier respectively are equal to 10 and 1 and 0.01. parameter values of selecting feature algorithm based on genetic decision tree are as follow:

Using of cross-validation approach with number 10 in education and genetic decision tree test, crossover operator: random standard, mutation act: random standard genome replacement percent equal to 0.25 error rate equal to 0.95, Capability of preference in decision tree with higher accuracy on smaller decision tree, lack of dynamic change in preference of decision tree with higher accuracy on smaller decision tree at the start and end of evolution, intersection rate number 0.99, mutation rate number 0.01, number of generation number 100, initial population number 100, random number Seed equal to 123456789.

Methods of manufacturing decision trees in proposed classification model are as follow:

Manufacturing five C4.5 decision tree by five way of selecting feature.

Manufacturing four C4.5 decision tree by meta (synthesis selecting feature algorithms and tree. C4.5 decision) with the use four selecting feature algorithm based on filter approach and Wrapper.

Using of C4.5 decision tree in selection feature and manufacturing genetic decision tree with selective features (the approach based on embedded model).

Genetic decision tree

Using of genetic decision tree in selecting features and manufacturing genetic decision tree with selective features (the approach based on embedded model).

Using of hierarchy decision trees: manufacturing genetic decision tree and then manufacturing C4.5 decision tree in two branches mode from second floor above genetic decision tree.

The proposed classified model in Melat bank were studied. Collection of data

Validation required in this model received like an electronic version from research and planning of Melat bank. In order to analyze the data and Melat bank validation information and receiving the necessary results of descriptive statistical tools, we used learning machine and data mining. In order to information analysis and recording the results we used learning machine WEKA soft wares and GATree software and Excel software. For manufacturing and validation of model from the credit customers data of Melat bank of Mahabad city were used that in the first six months of 2004 received facilities in the form of contract.

Results of propose classified decision tree is like as image 1.

Table 1. results of proposed classified model of validation of customers in credit data collection of Melat bank

Class accuracy of customers 3	Class accuracy of customers 2	Class accuracy of customers 1	Size of tree	Number of leaves	Percentage of true classified views	True classified views	Number of selective predictable feature	Total views	Classified algorithm	Row
0.8942	0.9321	0.9873	30	26	97.03	655	10	675	Proposed modeling	1

In the following proposed classified model of decision tree is given to comparison with other C4.5 decision trees in credit data collection of Melat bank.

Table 2. results of C4.5 performance without selection features and clustering credit collection of Melat bank

Class accuracy of customers 3	Class accuracy of customers 2	Class accuracy of customers 1	Size of tree	Number of leaves	Percentage of true classified views	Number of true classified investigation	Total views	Row
0.8746	0.9233	0.9654	83	33	96.44	651	675	1

In the following proposed classified model of decision tree is given to comparison with other C4.5 decision trees in credit data collection of Melat bank.

Table 3. results of C4.5 performance with selection features based on genetic algorithm and without clustering credit data collection of Melat bank

Class accuracy of customers 3	Class accuracy of customers 2	Class accuracy of customers 1	Size of tree	Number of leaves	Percentage of true classified views	Number of true classified investigation	Total views	Evaluation function of selecting feature based on genetic algorithm	Row
0.9021	0.928	0.9812	35	31	97.33	657	675	Wrapper with C4.5 classifier	1
0.8863	0.9172	0.9731	32	29	96.59	652	675	Correlation between features together and with target feature	2
0.8726	0.9245	0.9707	33	30	96.29	650	675	Subset compatible features with attribute values	3
0.8973	0.9216	0.9784	34	30	97.18	656	675	C4.5 classifier	4

Table 4. results of C4.5 performance with selection features based on best and first search and without clustering credit data collection of Melat bank

Class accuracy of customers 3	Class accuracy of customers 2	Class accuracy of customers 1	Size of tree	Number of leaves	Percentage of true classified views	Number of true classified investigation	Total views	Evaluation function of selecting feature based on genetic algorithm	Row
0.8972	0.9321	0.9826	38	33	97.77	660	675	Wrapper with C4.5 classifier	1
0.8874	0.9145	0.9712	31	28	96.88	654	675	Correlation between features together and with target feature	2
0.8812	0.9321	0.9781	37	33	97.03	655	675	Subset compatible features with attribute values	3
0.8918	0.9333	0.9792	35	31	97.48	658	675	C4.5 classifier	4

Table5. results of C4.5 performance with selection features based on genetic algorithm with considering feature of clustering credit type in data collection of Melat bank

Class accuracy of customers 3	Class accuracy of customers 2	Class accuracy of customers 1	Size of tree	Number of leaves	Percentage of true classified views	Number of true classified investigation	Total views	Evaluation function of selecting feature based on genetic algorithm	Row
0.8994	0.9253	0.9815	40	35	97.18	656	675	Wrapper with C4.5 classifier	1
0.8874	0.9145	0.9712	31	28	96.88	654	675	Correlation between features together and with target feature	2
0.8852	0.9307	0.9735	35	31	96.74	653	675	Subset compatible features with attribute values	3
0.8875	0.9385	0.9745	34	30	97.77	660	675	C4.5 classifier	4

Decision tree of proposed classified model has high accuracy of classifying and low complexity rather than compared decision trees. Just C4.5 decision tree from selecting feature with the first and best search and Wrapper evaluation function with C4.5 classifier has higher accuracy of classifying just to %0.1 but number of leaves and size of tree in decision tree is more than number of leaves and size of decision tree proposed classified model.

Based on proposed classified model, of genetic algorithm used in manufacturing decision trees like C4.5. Also classifying and complexity of accuracy decision tree proposed classified model is improved for customers validation of Melat bank rather than compared decision trees.

Conclusion and Recommends

Banks in granting credit facilities to customers need to validate them. Decision trees can classified customers in this field. Main issue in making decision trees that can optimally classify the customers. Target is offering suitable validation model of bank customers like Melat bank for granting credit facilities fit with each class.

This model in the form of development process in pattern recognition and CRISP process do the final manufacturing of decision tree for validation of bank customers. Clustering techniques and selecting features are used based on genetic algorithm in making decision trees. Decision tree made from proposed classifying model has higher accuracy in classifying, number of leaves and size of decision tree and as a result had lower complexity due to all the decision trees in this thesis.

According to what said we can use proposed classified model for manufacture and decision tree test in order to validation of bank customers due to searching background and proposed model these cases are recommended:

In terms of costs of false classification in decision tree algorithm and cost of false selecting features in selecting feature algorithm in proposed model.

Development of the proposed model with using other classifying way of decision tree like as ID3 in making proposed classifying model.

Also applied recommends for banks are as follows:

Using proposed classifying model in validation of bank customers for allocation of credit facilities.

Using development process in recognition pattern for manufacturing classified model for better validation of banks customers.

Design and manufacturing protection system of decision and applied software for validation of bank customers based on proposal model.

References

- Abdou H. Pointon J. Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking. *Expert Systems with Applications*; 2008. Available at: www.sciencedirect.com.
- Akhbari, Akhbari. Makhatab Rafiee, Farimah (2009). Using Neuro- Fuzzy logic systems in credit ranking of legal bank customers economic research magazine. N92, P 1-21. (in Persian).
- Alborzi, Mahmoud & et al (2010). Using genetic algorithm in optimal decision tree for validation of bank customer. *Journal of information technology management*, second period, N4, P 23-38. (In Persian).
- Arab Mazar, Abbas (2006). Effective factors on credit risk of bank customers, case study keshavarzi bank, economic gesture research quarter, third year, N6. (In Persian).
- Carvalho D. R., Freitas A. A. A Hybrid Decision Tree/Genetic Algorithm Method for Data Mining. *Information Sciences* 2004; 163.
- D'heygere T., Goethals P. L., Pauw N. D. Use of Genetic Algorithms to Select Input Variables in Decision Tree Models for the Prediction of Benthic Macroinvertebrates. *Ecological Modelling* 2003; 160. Available at: www.elsevier.com.
- D'heygere, T., Goethals, P. L., Pauw N. D. Genetic Algorithms for Optimisation of Predictive ecosystems Models Based on Decision Trees and Neural Networks. *Ecological Modelling* 2006; 195
- Dehuri S., Patnaik S., Ghosh A., Mall R., Application of Elitist Multiobjective Genetic Algorithm for Classification Rule Generation. *Applied Soft Computing* 2008; 8.
- Hsu P. L., Lai, R., Chiu, C. C., & Hsu, C. I. The hybrid of Association Rule Algorithms and Genetic algorithms for Tree Induction: An Example of Predicting the Student Course Performance. *Expert Systems with Applications* 2003; 25
- Jamshid, Saeed (2009). Ways of validating customers, money and bank research, Central Bank of Islamic Republic of Iran. (In Persian).
- Keshavarz, Gholamreza. Ayati, Hosein (2007). Comparing the work of Lajit model and way of classified trees and regression in validation of process in legal applicants for using bank facilities, economic research quarter, seventh year, N4, P71-79. (In Persian).
- Kim E., Kim W., Lee Y. Combination of Multiple Classifiers for the Customer's Purchase Behavior Prediction. *Decision Support Systems* 2002; 34.
- Kim Y. S., Sohn S. Y. Managing Loan Customers Using Misclassification Patterns of Credit Scoring Model. *Expert Systems with Applications* 2004; 26.
- Lee T. S., Chiu C. C., Lu C. J., Chen I. F. Credit Scoring Using the Hybrid Neural Discriminant Technique. *Expert Systems with Applications* 2002; 23.
- Mahmoud Abadi, Hamid. Ghayouri Moghadam, Ali (2011) credit ranking in terms of financial power payment of principal and interest debet with using covering analyses of data way, accounting knowledge magazine, second year, N4, P 125-145. (In Persian).
- Mehr Ara & et al (2011). Ranking of corporate credit clients of Parisan bank, quarterly economic modeling, third year, N3. P. 1121-150. (In Persian).
- Mousavian, Aba. Mousavi, Mahammad Mehdi (2009). Credit risk management in Islamic banking by credit su up necoul, economic Islamic research quarter, ninth year, N33, P 95-125. (in Persian).
- Nadali. A. khanbabaye. M. (2008).using techniques of decision tree and genetic tree to validation of banks customers in a protected decision system. Tehran: Industry of Amir Kabir. (In Persian).
- Sharifi. K. (2009). Authentication method in Melat bank (M. Khan Babaye, interviewer). (In Persian).
- Studies and control of risk office of Tejarat bank (2007). "risk management in Bankikng". First edition, Teharn: Press Center of Tehran University. (In Persian).