Middle-East Journal of Scientific Research 19 (6): 795-798, 2014
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2014.19.6.1470

Mining the Financial Multi-Relationship with Accurate Models

D. Kerana Hanirex and K.P. Kaliyamurthie

Bharath University, Chennai, India

Abstract: In order to overcome the difficulty in mining the large-scale multi-relationship system, a classification method or algorithm is proposed. The development of data-mining applications such as classification or clustering has shown the need for machine learning algorithms to be applied to large scale data. In this paper we present the comparison of different classification techniques using WEKA tool. The algorithm such as are Bayes Net, Logistic, Decision Table, Random Tree, JRip, Decision Stump and J48. Finally, the experiments in financial multi-relationship dataset demonstrate the efficiency of the methods.

Key words: Classification · Bayes Net · Logistic · Decision Table · Random Tree · JRip and J48

INTRODUCTION

FCM simulates the dynamic behaviour of system through the causal link between nodes of the entire map and plays a significant role in qualitative reasoning compared with other models. FCM model is inadequate in dealing with multi-relationship system, because there are a large collection of objects or concepts and links between them [1].

The aim of our work is to measure the performance of different classification methods using WEKA for financial dataset. WEKA is a collection of open source software for classification, clustering, association rule extraction [2].

A major problem in financial data analysis is achieving the correct accuracy of certain important information. For the function of multi- relationship data set, normally, many tests generally involve the clustering or classification of large scale data.Multi –relationship may be complicated if we conduct many tests. This kind of difficulty can be solved with the help of of machine learning which could be used directly to obtain the end result with the aid of several artificial intelligent algorithms which perform the role as classifiers [3].

There is considerable amount of research with machine learning algorithm such as Bayes Network, Logistic, Decision Table, Random Tree, JRip and J48 [4].

METHODS A. Bayes Network Classifier: A sequence Bayesian networks are a powerful probabilistic representation classifer. This classifier learns from training data the conditional probability of each attribute Ai given the class label C [5]. Classification is then done by applying Bayes rule to compute the probability of C given the particular instances of A1…..An and then predicting the class with the highest posterior probability. The goal of classification is to correctly predict the value of a class variable given a vector of predictors or attributes. In particular, the naive Bayes classifier is a Bayesian network where the class has no parents and each attribute has the class as its sole parent [6].

Logistic Regression: The term regression is defined as an analyzing the relation between a dependent variable and one or more independent variable. Regression may be of 2 types:Linear regression and logistic regression. Logistic regression is a generalization of linear regression [5], it is basically used for estimating binary or multi-class dependent variables. Logistic regression basically is used to classify the low dimensional data having non- linear boundaries. It also provides the difference in the percentage of dependent variable and provides the rank of individual variable according to its importance. So, the main motto of Logistic regression is to determine the result of each variable correctly. Logistic regression is also known as logistic model that provide categorical variable for target variable with status.

Decision Table: Decision Tree and Pruning a decision tree partitions the input data set into mutually exclusive regions, each of which is assigned a label, a value or an

Corresponding Author: D. Kerana Hanirex, Bharath University, Chennai, India.

action [6], the decision tree mechanism is transparent and we can follow a tree structure easily to identify how the decision is made. A decision tree is a tree structure consisting of internal and external nodes. An internal node is a decision making unit that evaluates a decision function and determine which child node to visit next. The external node has no child nodes and is associated with a label or value. However, many decision tree construction algorithms involve a two - step process. First, a very large decision tree is grown. Then, to reduce large size and overfiting the data, in the second step, the given tree is pruned. The pruned decision tree is used for classification [7].

Random Tree: The Random Forest method is based on bagging models built using the Random Tree method, in which classification trees are grown on a random subset of descriptors. The Random Tree method can be viewed as an implementation of the Random Subspace method for the case of classification trees. Combining two ensemble learning approaches, bagging and random space method, makes the Random Forest method very effective approach to build highly predictive classification models [8].

Jrip: JRip (RIPPER) is one of the popular classification algorithms. Classes are examined in increasing size and an initial set of rules for the class is generated.

J48: The C4.5 technique is one of the decision tree families that can produce both decision tree and rule-sets; and construct a tree. C4.5 models are easy to understand because the rules that are derived from the technique have a very straightforward interpretation. J48 is thepowerful decision tree classifiers. C5.0 and J48 are the improved versions of C4.5 algorithms.

Experimental Results and Analysis: In this section, we test the implementation accuracy of algorithm and compare with multi- relationship financial data set. Weka tool is used to select the attributes from the dataset.

- Dataset
- Dataset Information

The experiment is established on financial data set of PKDD CUP 1999.

Attribute Information: The data about the clients and their accounts consist of following relations:[1].

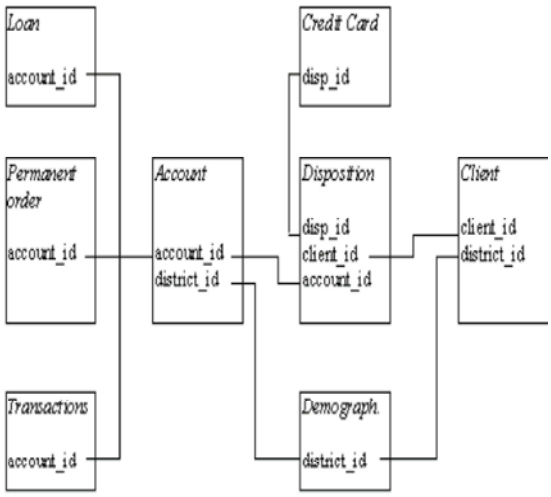

Fig. 1: Financial multi-relationship

- Relation account each record describes static characteristics of an account,
- Relation client each record describes characteristics of a client,
- Relation disposition
  - Each record relates together a client with an account i.e. this relation describes the rights of clients to operate accounts,
- Relation permanent order each record describes characteristics of a payment order
- Relation transaction- each record describes one transaction on an account,
- Relation - each record describes a loan granted for a given account,
- Relation credit card each record describes a credit card issued to an account,
- Relation demographic - each record describes demographic characteristics of a district.

RESULTS

The experiment was conducted using selected classification methods or algorithms namely Bayes Network classifier, Logistic Regression, Decision Table, Random Tree, JRip and J48. we use the same experiment procedure as suggested by WEKA. The 75% data is used for training and the remaining is for testing purposes. In WEKA, all data is considered as instances and features in the data are known as attributes.

The results of the simulation are shown in Tables 2 and3 below. Table 2 mainly summarizes the result based on accuracy and table 3 describes the time taken for each simulation.

Table 1: Background of Financial multi- relationship dataset

Table	Attributes
Loan	includes: amount,
Account	duration,
Order	payments
Trans	includes:
Disp	account_id,
District	district_id
Client	includes:
	amount
	includes:
	account_id,
	amount
	includes:
	account_id,
	client_id
	includes: A3
	includes:
	client_id,
	account_id

Table 3: Comparison of accuracy

	After normalize	
Method	Correctly classified	IN correctly classified
bayes. Bayes Net	95.98%	4.02%
bayes. Naive Bayes	95.18%	4.82%
functions. logistic	96.12%	3.88%
rules. PART	98.78%	1.22%
rules. Decision Table	96%	4%
rules. Jrip	96.24%	3.76%
rules. ZeroR	86.36%	13.64%
trees. Decision Stump	94.42%	5.52%
trees. J48	97.18%	2.82%

Table 4: Comparison of execution time

Algorithm	Runtime (s)
Bayes Net	0.02
Naïve Bayes	0.02
Logistic	0.02
Decision Table	0.01
Jrip	0
Decision Stump	0
J48	0.04

Based on the above and Table 1 and 2, we can clearly see that the highest accuracy is 98.78% and the lowest is 86.36%. The other algorithm yields an average accuracy of around 96%. In fact, the highest accuracy belongs to the PART, followed by J48 with a percentage of 97.18% and subsequently decision tree with pruning and single conjunctive rule learner. The total time required to build the model is also a crucial parameter in comparing the classification algorithm.

CONCLUSIONS

As a conclusion, we have met our objective which is to evaluate and investigate five selected classification algorithms based on Weka. The best algorithm based on the financial data set is PART with an accuracy of 96.24% and the total time taken to build the model is at 0.02 seconds. Logistic has the lowest average error compared to others. These results suggest that among the machine learning algorithm tested, Bayes network Logistic classifier has the potential to significantly improve the conventional classification methods for use in financial multi-relationship data set [9-11].

REFERENCES

1. Comparison of Different Classification Techniques Using WEKA for Breast Cancer Mohd Fauzi bin Othman, Thomas Moh Shan Yau Control and Instrumentation Department, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia.
2. International Conference & Workshop on Recent Trends in Technology, (TCET) 2012. Proceedings published in International Journal of Computer Applications® (IJCA) 27 Analysis of Machine Learning Algorithms using WEKA. Aaditya Desai Ph.D. Scholar, NMIMS University. TCET, Mumbai and Dr. Sunil Rai Ph.D. Guide, NMIMS University.
3. WEKA at http://www.cs.waikato.ac.nz/~ml/weka.
4. Lee, C.H. and D.G. Shin, 1999. A multi strategy approach to classification learning in database, Data Knowledge Engg., 31: 67-9.
5. WEKA Tutorial http:// www.cs.utexas.edu/ users/ ml/ tutorials/Weka-tut/
5. Kumaravel, B. Anatha Barathi, 2013. Personalized image search using query expansion, Middle-East Journal of Scientific Research, ISSN: 1990-9233, 15(12): 1736-1739.
6. Kumaravel, A. and R. Udayakumar, 2013. Web Portal Visits Patterns Predicted by Intuitionistic Fuzzy Approach, Indian Journal of Science and Technology, ISSN: 0974-6846, 6(5S): 4549-4553.
7. Kumaravel, A. and K. Rangarajan, 2013. Algorithm for Automation Specification for Exploring Dynamic Labyrinths, Indian Journal of Science and Technology, ISSN: 0974-6846, 6(5S): 4554-4559.
8. Kumaravel, A. and Oinam Nickson Meetei, 2013. An Application of Non-uniform Cellular Automata for Efficient Crytography, Indian Journal of Science and Technology, ISSN: 0974-6846, 6(5S): 4560-4566.

9. Pattanayak, Monalisa and P.L. Nayak, 2013. Green Synthesis of Gold Nanoparticles Using Elettaria cardamomum (ELAICHI) Aqueous Extract World Journal of Nano Science & Technology, 2(1): 01-05.
10. Chahataray, Rajashree. and P.L. Nayak, 2013. Synthesis and Characterization of Conducting Polymers Multi Walled Carbon Nanotube-Chitosan Composites Coupled with Poly (P-Aminophenol) World Journal of Nano Science & Technology, 2(1): 18-25.
11. Parida, Umesh Kumar, S.K. Biswal, P.L. Nayak and B.K. Bindhani, 2013. Gold Nano Particles for Biomedical Applications World Journal of Nano Science & Technology, 2(1): 47-57.