International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-3, Issue-4, September 2013

A Comparison of Different Classification Techniques for Bank Direct Marketing

K. Wisaeng

Abstract— ***Classification is a data mining techniques used to predict group membership for data instance. In this paper, we present the comparison of different classification techniques in open source data mining software which consists of a decision tree methods and machine learning for a set of bank direct marketing dataset. All decision tree methods tested are J48-graft and LAD tree while machine learning tested are radial basis function network and support vector machine. The experiment results show are a bout classification sensitivity, specificity, accuracy, mean absolute error and root mean squared error. The results on bank direct marketing data also the efficiency of machine learning methods by using support vector machine is better than that of all algorithms.***

Index Terms—***Data mining, bank direct marketing, J48-graft, LAD tree, radial basis function network, support vector machine.***

I. INTRODUCTION

Data mining is the process of extracting previously unknown information from a large datasets. Today, data mining is being used by several industries including finance and banking. The bank is marketing department can use data mining to analyze customer datasets and develop statistically profiles of individual customer preference for product and service. In bank direct marketing domain, there are several data mining techniques can be used for classifying marketing service such as decision tree, naive Bayes classifier, support vector machine, classification and association rule mining and six-sigma methodology.

Zhixin et al. [1] improved classification method base on association rules. Qiang et al. [2] applied association classification method based on compactness of rules. The experimental shown that proposed method has better classification in comparison with classification and association rule mining technique. Barlik [3] used association rule mining classification for relational data and its use in web mining. Sumithra et al. [4] applied a distributed apriori association rule techniques for grid based knowledge discovery. Trnka [5] used six-sigma technique for market basket analysis. Xie et al. [6] applied association rule mining, Chiu et al. [7] used principal component for market basket analysis. Wang et al. [8] proposed association rules mining in e-commerce. As the number of available methods becomes increasingly difficult, each with their own advantages and disadvantages.

In this paper, I presents the performance analysis of different classification methods by between decision tree methods and machine learning for bank direct marketing data set. A major problem in bank direct marketing is in attaining the accuracy of certain important information.

Manuscript received September, 2013.

K. Wisaeng, Mahasarakham Business School (MBS), Mahasarakham University, Kantharawichai, khamriang, Maha sarakham, 44150, Thailand.

However, too many tests could complicate the main in evaluation process and lead to the difficulty in obtaining the end result, particularly in the case where many tests are performed. This kind of difficulty could be resolved with the aid of decision tree methods and machine learning are J48-graft, LAD tree, radial basis function network, and support vector machine used directly to obtain the end result.

II. METHODS

We analyzed the performance of different classification techniques to select the one with the most accurate results for classification of bank direct marketing dataset. We choose four very commonly used techniques from different classification techniques, two techniques from decision tree and the rest from machine learning. Decision tree consist J48-graft algorithm and LAD tree algorithm while machine learning consist radial basis function network and support vector machine. The overall procedure of the classification techniques for bank direct marketing is shown in Fig. 1.

A. J48-Graft Algorithm

J48-graft algorithm generates a grafted decision tree from a J48 tree algorithm. The grafting technique is an inductive process that adds nodes to inferred decision trees. The grafting technique is an inductive process that adds nodes to inferred decision trees with the purpose of reducing pre-diction errors. The J48-graft algorithm classify region of the multidimensional space of attributes not occupied by the training examples [9]. This process is demonstrated to frequently improve predictive accuracy. Special analysis might suggest that decision tree grafting is the direct reverse of pruning. To the contrary, it is argued that the two processes are complementary. This is because, like standard tree growing techniques, pruning uses only local information, whereas grafting uses non-local information. The use of both pruning and grafting in conjunction is demonstrated to provide the best general predictive accuracy over a representative selection of learning tasks [10].

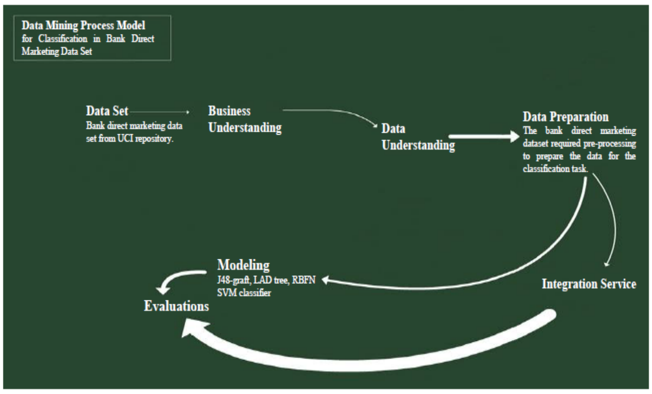


Figure 1. The data mining process model (adapted from [11])

B. LAD Tree Algorithm

Logical Analysis of Data (LAD) tree is the classifier for binary target variable based on learning a logical expression that can distinguish between positive and negative samples in a data set. The central concept in LAD tree algorithm is that of classification, clustering, and other problems. The construction of LAD model for a given data set typically involves the generation of large set patterns and the selection of a subset of them that satisfies the above assumption such that each pattern in the model satisfies certain requirements in terms of prevalence and homogeneity [12].

C. Radial Basis Function Network

Radial basis function network have a static Gaussian function as nonlinearity for hidden layer processing elements. The Gaussian function responds only to small region of the input space where the Gaussian is centered [13]. The key to a successful implementation of these networks is to find suitable centers for the Gaussian functions [14]. Radial basis function networks typically have three layers are input layer, hidden layer, and linear output layer. The input can be modeled as a vector of real numbers $X \in R^n$. The output of the networks is then a scalar function of the input vector, $\varphi : R^n \to R$, and is given by Eq. (1).

$$\varphi(x) = \sum_{i=1}^{N} a_i p \left\| x - c_i \right\| \quad (1)$$

Where **N** is the number of neurons in the hidden layer, $\mathbf{c_i}$ is the center vector for neuron **i**, and $\mathbf{a_i}$ is the weight of neuron **i** in the linear output neuron. Functions that depend only on the distance from a center vector are radially symmetric about that vector, hence the name radial basis function. In the basic form all inputs are connected to each hidden neuron.

The norm is typically taken to be the Euclidean distance and the radial basis function is commonly taken to be Gaussian is defined as Eq. (2).

$$p \left\| x - c_i \right\| = \exp\left[-\beta \left\| x - c_i \right\|^2 \right] \quad (2)$$

The Gaussian basis functions are local to the center vector is shown as Eq. (3).

$$\lim_{\|x\| \to \infty} p \; x - c_i = 0 \quad (3)$$

Given certain mild conditions on the shape of the activation function, radial basis function networks are universal approximates on a compact subset of $\mathbf{R^n}$ [15]. This means that a radial basis function network with enough hidden neurons can approximate any continuous function with arbitrary precision.

D. Support Vector Machine

Support vector machine are basically binary classification algorithms. The basic support vector machine takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, a support vector machine training algorithm builds a model that assigns new examples into one category or the other. The formulations of support vector machine algorithm are described below.

Given some training sample $\{y_i, x_i\}_{i=1}^{N}$, with the label $y_i \in \{-1, 1\}$ indicating the class to which the feature vector $x_i \in R^d$ belongs.

Support vector machine finds linear separating hyper plane with a maximum-margin in the higher feature space induced by kernel function $K(\cdot,\cdot)$. In summary, given an input vector **x**, a support vector machine according to Eq. (4) - Eq. (6).

$$\hat{y} = \text{sign}\,\{f(x)\} \quad (4)$$

where $\hat{y}$ is the estimate to the classification, and

$$f(x) = \sum_{i \in S} \alpha_i y_i \Phi(x_i) \times \Phi(x) + b, \quad (5)$$

$$f(x) = \sum_{i \in S} \alpha_i y_i K(x_i, x) + b \quad (6)$$

where **x** is the feature vector to be classified, **i** represents an indexes the training example, **S** is a set of indices for which $\mathbf{x_i}$ is a support vector, i.e., a vector for which $\alpha_i \neq 0$ after optimization α_i, and **b** are fit to the data to maximize the margin, $\mathbf{y_i}$ is the label {-1, 1} of training example **i**, $\{x_i\}_{i=1}^{n}$ are referred to as support vectors which are a small set of training data near the separating hyper plane and **K(.,.)** is the kernel function. A serious problem with nonlinear kernel support vector machine is their complexities of classification which are high when a large number of support vectors is needed.

E. Data Set Description

We have extracted the datasets of bank direct marketing from UCI repository. It has a dimensions of 16 attribute and 45,211 instances. For proposes of training and testing, only 70% of the overall data is used for training and the rest is used for testing the accuracy of the selected classification algorithms. The descriptions of the data sets are summarized in Table 1.

Table 1. Attribute from the bank direct marketing data set for classification algorithm

ID	Attributes	Type	Values	Descriptions
1	Age	Numeric	Real	Age at the contact date (≥18)
2	Job	Categorical	Admin, Unknown, Unemployed, Management, Housemaid, Entrepreneur, Student, Blue-collar, Self-employed, Retired, Technician, Services	
3	Marital	Categorical	Married, Divorced, Single, widowed	
4	Education	Categorical	Unknown, Secondary, Primary, Tertiary	
5	Default	Binary	Yes, No	Yes or No
6	Balance	Numeric	Real	In euro currency
7	Housing	Binary	Yes, No	Yes or No
8	Loan	Binary	Yes, No	Yes or No
9	Contact	Categorical	Unknown, Telephone, Cellular	
10	Day	Numeric	Real	Referring to when the contact was made
11	Month	Categorical	Jan, Feb, mar,…, Nov, Dec	
12	Duration	Numeric	Real	Of the contact (in seconds)
13	Campaign	Numeric	Real	
14	Pday	Numeric	Real	
15	Previous	Numeric	Real	
16	Poutcome	Categorical	Unknown, Failure, Success	

III. PERFORMANCE MEASUREMENT

The algorithms performance are partitioned into several sub item for easier analysis and evaluation. In first part, the sensitivity, specificity and accuracy are used [16]. All measures can be calculated based on four values, namely True Positive (TP, a number of correctly classified that an instances positive), False Positive (FP, a number of incorrectly classified that an instance is positive), False Negative (FN, a number of incorrectly classified that an instance is negative), and True Negative (TN, a number of correctly classified that an instance is negative). These values are defined in Table 2.

Table 2. Predicted Class

True Class	Yes	No	Total
Yes	TP	FN	TP+FN
No	FP	TN	FP+TN
Total	**TP+FP**	**FN+TN**	**TP+FN+FP+TN**

From these quantities, the sensitivity and specificity computed by using Eq. (7) and (8) respectively.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{Specificity} = \frac{TP}{TP + FP} \tag{8}$$

Thus "Sensitivity" was defined as percentage of correctly classified instances, and "Specificity" was defined as percentage of incorrectly classified instances. Also, "Accuracy" was defined as the overall success rate of the classifier and computed by using Eq. (9).

$$\text{Acurracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

In the second part, we also show the relative MAE, RMSE, and RAE for reference and evaluation.

IV. EXPERIMENTAL RESULTS

This section presents the results obtained after running the four classification techniques for bank direct marketing data set. To construct the algorithms, we use Waikato Environment for Knowledge Analysis (WEKA version 3.6.10), an open source data mining too [17]. Which was developed at University of Waikato New Zealand. WEKA is an open source application that is freely available under the GNU general public license agreement. All experiment were performed on Duo Core with 1.8GHz CPU and 2G RAM. The result for each classification algorithms are shown and described below.

A. Results for classification using J48-Graft Algorithm

Class for generating a grafted (pruned or unpruned) decision tree. J48-graft algorithm in WEKA is done with parameter are: binary splits on nominal attributes when building the trees = false, the confidence factor used for pruning = 0.25, debug = false, the minimum number of instances per leaf = 2, relabeling is allowed during grafting = false, to save the training data for visualization = false, to consider the sub tree raising operation when pruning = true, unpruned = false, and use Laplace = false. For J48-graft algorithm, achieves a sensitivity of 76.50%, specificity of 78.60%, accuracy of 76.52%, mean absolute error of 0.32+, root mean squared error of 0.42+, and relative absolute error of 71.33+, respectively.

B. Results for classification using LAD tree algorithm

Class for generating a multi-class alternating decision tree using the LogitBoost strategy. LAD tree in WEKA is selected parameter are: debug = false, and the number of boosting iterations to use, which determines the size of the tree = 10. For LAD tree algorithm, achieves a sensitivity of 76.10%, specificity of 75.00%, accuracy of 76.08%, mean absolute error of 0.31+, root mean squared error of 0.40+, and relative absolute error of 70.08+, respectively.

C. Results for classification using radial basis function network

Class that implements a normalized Gaussian radial basis function network. It uses the k-means clustering algorithm to provide the basis functions and learns either a logistic regression (discrete class problems) or linear regression (numeric class problems) on top of that. Symmetric multivariate Gaussians are fit to the data from each cluster. If the class is nominal it uses the given number of clusters per class. It standardizes all numeric attributes to zero mean and unit variance. Radial basis function network in WEKA is done with parameter are: the random seed to pass on to K-means =1, debug = false, maximum number of iterations for the logistic regression to perform = -1, sets the minimum standard deviation for the clusters = 0.1, the number of clusters for K-Means to generate = 2, and set the Ridge value for the logistic or linear regression = 1.0E-8, respectively. For radial basis function network, achieves a sensitivity of 74.30%, specificity of 73.50%, accuracy of 74.34%, mean absolute error of 0.35+, root mean squared error of 0.42+, and relative absolute error of 79.49+, respectively.

D. Results for classification using support vector machine

A wrapper class for the libsvm tools (the libsvm classes, typically the jar file, need to be in the class path to use this classifier). LibSVM runs faster than SMO since it uses LibSVM to build the SVM classifier. LibSVM allows users to experiment with one-class SVM, Regressing SVM, and nu-SVM supported by LibSVM tool. LibSVM reports many useful statistics about LibSVM classifier (e.g., sensitivity, specificity, accuracy, etc.). Support vector machine in WEKA is selected parameter are: the type of SVM to use = C-SVC, the cache size = 40, the coefficient to use = 0.0, the cost parameter C for C-SVC, epsilon-SVR and nu-SVR = 1.0, debug = false, the degree of the kernel = 3, to turn off automatic replacement of missing values = false, the tolerance of the termination criterion = 0.001, the gamma to use = 0, the type of kernel to use = RBF, the epsilon for the loss function in epsilon-SVR = 0.1, normalize the data = false, the value of nu for nu-SVC, one-class SVM and nu-SVR = 0.5, probability estimates = false, the random number seed to be used = 1, and the shrinking heuristic = false. For support vector machine, achieves a sensitivity of 87.00%, specificity of 86.70%, accuracy of 86.95%, mean absolute error of 0.26+, root mean squared error of 0.38+, and relative absolute error of 82.73+, respectively.

V. CONCLUSIONS

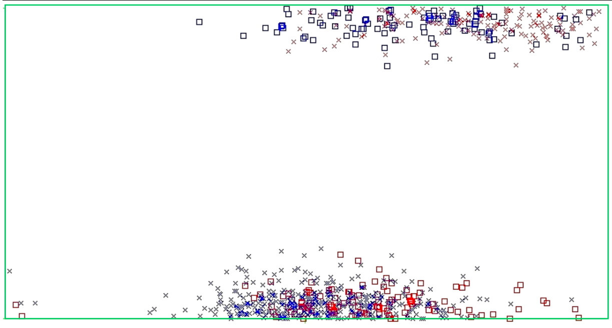
Many algorithms have been proposed for classification in bank direct marketing data set. We choose four very commonly used algorithms such as J48-graft algorithm, LAD tree (LADT) algorithm, radial basis function network (RBFN, and support vector machine (SVM) to towards our classification of bank direct marketing. Among all classifier, our experimental results show that the support vector machine achieves highest sensitivity, specificity and accuracy. In other hand, the worst classification was performed by radial basis function network, the algorithm achieves highest sensitivity but lowest specificity and accuracy. The summary results from all algorithms are compared show in Table 2 and Table 3. The blue squares represent the tested negative while red squares represent the tested positive by the algorithm (presented in Figure 2).

Table 2. Comparison results of classification for bank direct marketing from each algorithm (values as %).

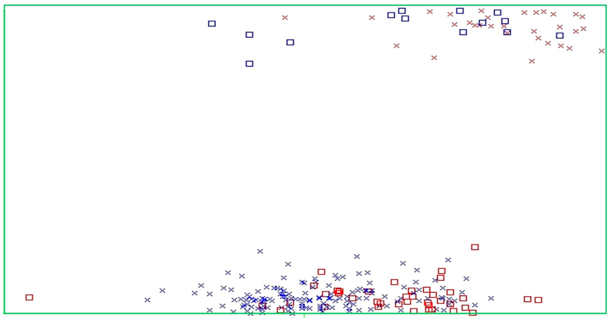
Algorithm	Sensitivity	Specificity	Accuracy
J48	76.50	78.60	76.52
LADT	76.10	75.00	76.08
RBFN	74.30	73.50	74.34
SVM	87.00	86.70	86.95

Table 3. Comparison error results of classification for bank direct marketing from each algorithm.

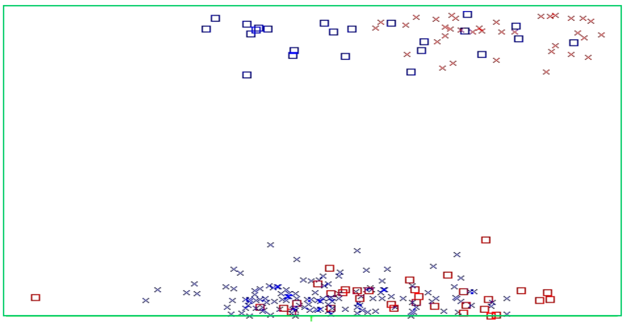
Algorithm	MAE	RMSE	RAE
J48	0.32+	0.42+	71.33+
LADT	0.31+	0.40+	70.08+
RBFN	0.35+	0.42+	79.49+
SVM	0.26+	0.38+	82.73+



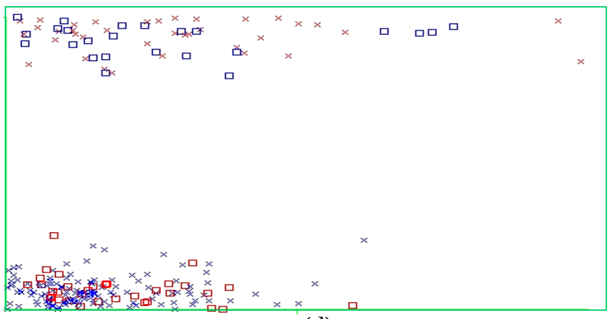
(a)



(b)



(c)



(d)

Figure 2. The visualize classifier error from each algorithm (a) J48-graft, (b) LADT, (c) RFBN, (d) SVM.

ACKNOWLEDGMENT

This paper was supported by the Mahasakham Business School, Mahasahakham University, Thailand. We also would like to thank UCI repository for the data set used in this experiment.

REFERENCES

[1] Z. Hao, X. Wang, L. Yao, and Y. Zhang, “Improved classification based on predictive associative rules”, *IEEE International Conference on System*, Man and Cybernetics, 2009, pp. 1165–1170.

[2] Q. Niu, S.X. Xia, and L.Zhang, “Association classification based on compactness of rules”, *International workshop on knowledge discovery and data mining*, WKDD, 2009, pp. 245–247.

[3] V. Bartik, “Association based classification for relational data and its use in web mining”, *IEEE Symposium on Computational Intelligence and Data Mining*, 2009, pp. 252-258.

[4] R. Sumitha, and S. Paul, “Using Distributed Apriori Mining Algorithms for Grid Based Knowledge Discovery,” *International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2010, pp.1-5.

[5] A. Trnka, “Market Basket Analysis with Data Mining Methods,” *International Conference on Networking and Information Technology (ICNIT)*, 2010, pp. 446-450.

[6] W.X. Xie, H.N. Qi, and M.I. Huang, “Market Basket Analysis Based on Text Segmentation and Association Rule Mining,” *International Conference on Neural Networking and Distributed Computing (ICNDC)*, 2010, pp. 309-313.

[7] K.S.Y. Chiu, R.W.P. Luk, K.C.C. Chan, and K.F.L. Chung, “Market-basket Analysis with Principal Component Analysis: An Exploration,” *IEEE International Conference on Systems*, 2002.

[8] X. Wang, Z. Ni, and H. Cao, “Research on Association Rules Mining Based-On Ontology in E-Commerce,” *International Conference on Wireless Communications*, Networking and Mobile Computing, 2007, pp. 3549-3552.

[9] D. Jachyra, K. Pancerz, and J. Gomula, “Classification of MMPI Profiles using Decision Trees” Concurrency, Specification and Programing, Poland, 2011, pp. 397.407.

[10] C.S. Trilok, J. Manoj, “WEKA Approach for Comparative Study of Classification Algorithm,” International Journal of Advanced Research in Computer and Communication Engineering, 2013, pp. 1925-1931.

[11] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, “CRISP-DM 1.0 Step by Step data mining guide,” CRISP-DM Consortium, 2000. Gabriela Alexe, Sorin Alexe, Tibérius O. Bonates, Alexander Kogan, “Logical analysis of data --- the vision of Peter L. Hammer,” Annals of Mathematics and Artificial Intelligence archive Volume 49 Issue 1-4, April 2007 Pages 265 - 312

[12] D. Buhmann, “Radial Basis Functions: Theory and Imple-mentations,” *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge, 2003.

[13] S. V. Chakravarthy, and J. Ghosh, “Scale Based Clustering using Radial Basis Function Networks,” *In Press of Proceeding of IEEE International Conference on Neural Networks*, Florida, 1994, pp.

[14] J. Park, and I. W. Sandberg, “Universal Approximation Using Radial-Basis-Function Networks,” *Neural Computation*, 2013, pp.246-257.

[15] A.M., Samuel, M.D., Daniel, K., James, and M.W. Gray, “Are decision trees always greener on the open (source) side of the fence?,” *In Proc. of the International Conference on Data Mining*, 2009, pp. 185–188

[16] M., Hall, E., Frank, G., Holmes, B., Pfahringer, P., Reutemann, and I.H Witten, “The WEKA Data Mining Software: An Update,” *SIGKDD Explorations*, Volume 11, 2009.