Comparison of Classification algorithms: Decision Tree and Bayesian Network for prediction of Student Graduation

Pathom Pumpuang

Department of Computer Science, Faculty of Science, Kasetsart University, 50 Phaholyothin Rd, Jatujak Bangkok 10903 Thailand

g4764231@ku.ac.th

Anongnart Srivihok

Department of Computer Science, Faculty of Science, Kasetsart University, 50 Phaholyothin Rd, Jatujak Bangkok 10903 Thailand

anongnart.s@ku.ac.th

Prasong Praneetpolgrang

Master of Science Program in Information Technology Graduate School, Sripatum University, 61 Phaholyothin Rd, Jatujak Bangkok 10900 Thailand

prasong.pr@spu.ac.th

Abstract

Each year many undergraduate students finish their studies smoothly, however some students have difficulties in studying. They are retired, failed or dropped out of the university. It is worth to find the model for predicting the success of undergraduate students from their characteristics and educational backgrounds. This study compares the performances of two classifiers: Decision Tree (C4.5) and Bayesian Network in predicting the success of undergraduate students. Cross Validation and Hold-Out Method are applied for the model of evaluation. Further, Correct Percentage, True Positive Rate, False Positive Rate, Precision, Recall and F-Measure are used for measuring the prediction accuracy. Results show that Decision Tree (C4.5) has better precision and lower false positive rate than Bayesian Network in predicting student class. The model from this study can be applied in predicting student status and further used in helping students with tendency to dropped out.

Key Words: Data Mining, Bayesian Network, Decision Tree (C4.5), Student Graduation

1. Introduction

At present there are more educational competitions than before and the students have chances to choose their own ways of studying. Unfortunately, some students face the obstacle and this makes them be retired, failed or dropped out of the university. Thus, it is necessary to help them overcome the problems occurred.

As a result of the difficulties mentioned, the solution for the students who would like to graduate from the universities is to look for the methods of predicting the success from their characteristics and educational backgrounds.

Data mining technique is based on statistical analysis, it has been used in finding and describing structural patterns in data segmentatoin and predictions. This technique has been applied extensively in many industries including banking and finances, education, medical sciences and manufacturing.

Xenos [2] proposed Bayesian Network for modeling student behaviours in order to enable prediction, status assessment and decision-making. Accuracate and useful results can be obtained.

Garcia and et al. [3] evaluated Bayesian Network precision for representing and detecting students' learning styles in a Web-based education system. The Bayesian Network could be estimated with high precision the categorizing students to pre defined dimensions.

Mukoolskunpibal and Kitisin [4] compared the efficiency of C4.5, ADTree and Naïve Bayes algorithms on international postal mail and packages on the prediction of concealed narcotics. Performance comparisons used Hold-Out and k-fold cross-validation methods. Correct rate of ADTree algorithm is the best.

Yingkuachat and et al. [5] proposed the prediction of education accomplishment by using data mining technique, the Bayesian Network. Result shows that Bayesian Network is able to determine important variables for the prediction of the result of education accomplishment and high prediction accuracy.

Yamansabideen and et al. [6] used data mining to develop of Customer Relationship Management for the student, by using Decision Tree. Result can be used as decision supporting data to solve the problems concerning students who would be nearly eliminated.

Sun and Shenoy [7] used Bayesian Network for bankruptcy prediction based on a 10-fold cross-validation. Result shown that the model's performance is the best.

Hidekazu and et al. [8] used Decision Tree for estimating sentence types. The representative Decision Tree algorithm C4.5 was revised. The gain ratio criterion was changed, and the hill climbing method was replaced with a genetic algorithm. Result shown high accuracy performance.

Lee and et al. [9] used Decision Tree to develop a prediction model for success based on customer recognitions of service offerings in e-commerce. Result shown superior prediction accuracy.

Both Decision Tree and Bayesian Network seems to be good classifiers for the prediction of education industries especially students success. This paper presents a comparison of Decision Tree and Bayesian Network for prediction of undergraduate student's graduation. The objective of this study is to identify the best fit classification algorithm for prediction of undergraduate students by comparing the two algorithms.

This paper is organized as follows. Section 2 is the theory of classification algorithms. Section 3 is the Cross-Validation. Section 4 is the Evaluate technique. Section 5 is the study framework. The experimental results are revealed in Section 6. The last section is about Conclusions and Future work.

2. Classification algorithm

2.1 Bayesian Network

A Bayesian Network is a specific type of graphical model which is a directed acyclic graph. That is, all of the edges in the graph are directed and there are no cycles. A Bayesian Network can be used to compute the conditional probability of one node, given values assigned to the other nodes. A Bayesian Network can be used as a classifier that gives the posterior probability distribution of the class node given the values of other attributes.

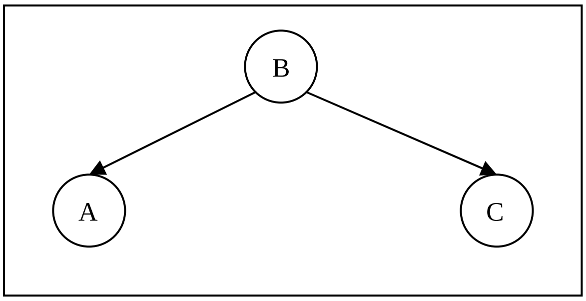

Figure 1. Example of Bayesian Network.

Figure 1. illustrates a Bayesian Network. Its set of edges is E={(B,A),(B,C)}. The edges in the Bayesian Network encode a particular factorization of the joint distribution. In this example, the joint distribution of all the variables, as factorized by this Bayesian Network, is

$$P(A,B,C) = P(A \mid B) \bullet P(B) \bullet P(C \mid B) \quad (1)$$

A Bayesian Network is a carrier of the conditional independencies of a set of variables, not of their causal connections. However, causal relations can be modeled by the closely related causal Bayesian Network.

```
Assume for simplicity that no nodes are evidence
nodes:
1 Generate initial values for all nodes in the net
2 Repeat until have enough data {
    2.1 For each node X {
      2.1.1 Throw away the current value for X
      2.1.2 Generate a new value for X according
            to the probability distribution over X
            conditional on the values currently
            given to the other nodes
    }
    2.2 record current values for all nodes of
        interest
}
```

Figure 2. Bayesian Network in pseudo-code.

2.2 Decision Tree

A Decision Tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a classification or decision. C4.5 is extension of the basic ID3 algorithm in figure 3. designed by Quinlan.

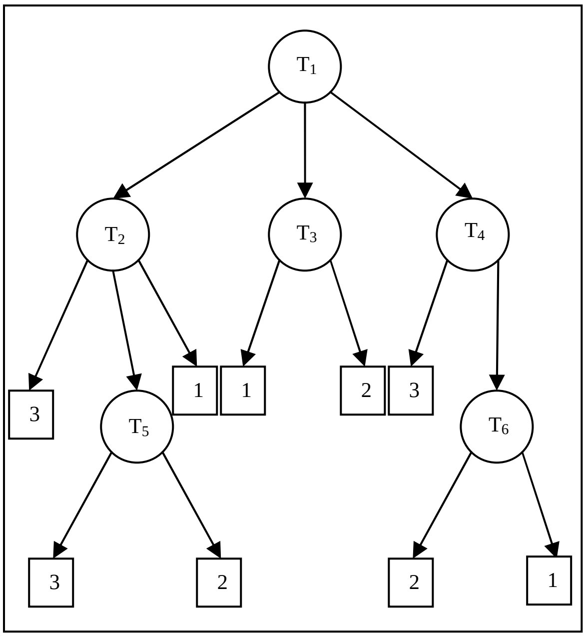

Figure 3. Example of Decision Tree.

Fig. 3. illustrates a Decision Tree. It's a tree whose internal nodes are tests: T_1...T_6, whose leaf nodes are categories. A Decision Tree assigns a class number to an input pattern by filtering the pattern down through the tests in the tree.

Various Decision Tree algorithms such as CHAID, C4.5/C5.0, CART, etc., produce trees that are different from one another in the following ways: how many splits are allowed at each level of the tree, how those splits are chosen when the tree is built, and how the tree growth is limited to prevent over-fitting.

Input: A data set, S
Output: A Decision Tree
If all the instances have the same value for the target attribute then return a decision tree that is simply this value
Else
1. Compute Gain values for all attributes and select an attribute with the lowest value and create a node for that attribute.
2. make a branch from this node for every value of the attribute.
3. assign all possible values of the attribute to branches.
4. follow each branch by partitioning the dataset to be only instances whereby the value of the branch is present and then go back to 1.

Figure 4. ID3 Decision Tree in pseudo-code.

C4.5 builds Decision Tree from a set of training data in the same way as ID3, using the concept of Information Entropy. C4.5 uses the fact that each attribute of the data can be used to make a decision that splits the data into smaller subsets. C4.5 examines the normalized Information Gain that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is the one used to make the decision.

In order to define information gain precisely, we need to define a measure commonly used in information theory, called entropy, that characterizes the purity of an arbitrary collection of examples. Given a set S, containing only positive and negative examples of some target concept: two class problem, the entropy of set S relative to this simple, binary classification is defined as:

$$Entropy(S) = -p_p \log_2 p_p - p_n \log_2 p_n \quad (2)$$

Where p_p is the proportion of positive examples
p_n is the proportion of negative examples

Given entropy as a measure of the impurity in a collection of training examples, we can now define a measure of the effectiveness of an attribute in classifying the training data. The measure we will use, called information gain, is simply the expected reduction in entropy caused by partitioning the examples according to this attribute. More precisely, the information gain, Gain(S, A) of and attribute A, relative to a collection of examples S, is defined as

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (3)$$

Where A is the set of all possible values
S_v is the subset of S for which attribute A

3. Cross-validation

Cross validation is a model evaluation method that is better. Some of the data is removed before training begins. Then when training is done, the data that was removed can be used to test the performance of the learned model.

The holdout method is the simplest kind of cross validation. The data set is separated into two sets, called the training set and the testing set. The function approximator fits a function using the training set only. Then the function approximator is asked to predict the output values for the data in the testing set.

4. Evaluation

To evaluate classifiers used in this work, we apply a range of standard reference metrics defined as follow:

Table 1. Different outcomes of a two-class prediction.

		Predicted class	
		Yes	No
Actual class	Yes	True positive	False negative
	No	False positive	True negative

In table 1. True Positive (TP) and True Negative (TN) are correct classification. A False Positive (FP) occurs when the outcomes is incorrectly predicted as yes (or positive) when it is actually no (negative). A False Negative (FN) occurs when the outcomes is incorrectly predicted as negative where it is actually positive.

Correct Percentage is the number of correct classifications divided by the total number of classifications:

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

True Positive Rate is TP divided by the total number of positives, which is TP+TN.

False Positive Rate is FP divided by the total numbers of negative, FP+TN.

Precision is number of documents relevant and retrieved divided by total number of documents that are retrieved.

Recall is number of documents relevant and retrieved divided by total number of documents that are relevant.

F-Measure is evaluated classification performance based on precision and recall.

$$F = \frac{2*TP}{2*TP + FP + FN} \tag{5}$$

5. Study framework

5.1 Data pre-processing

Data pre-processing is an important process because in real world data are generally: incomplete, noisy and inconsistent.

Tasks in data preprocessing process consists of two step

1. Data cleaning: Fill in missing values and identify outliers and smooth out noisy data.
2. Data Transformation: Normalization, aggregation, generalization and attribute construction.

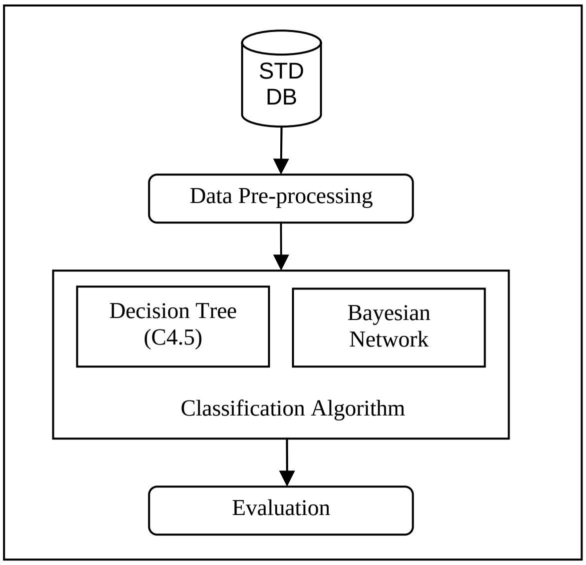

Figure 5. Study framework.

5.2 Classification

Classification is the prediction of nominal values. It was decided to concentrate on an algorithm for generating two classification algorithms: Decision Tree (C4.5 algorithm) and Bayesian Network.

5.3 Evaluation

There are six measurements used in this study: Correct Percentage, True positive rate, False positive rate, Precision, Recall and F-measure.

6. Experimental

In this section, we compare two classification algorithms such as Decision Tree (C4.5) and Bayesian Network by using Hold-out cross-validation. In the results of experiment of each algorithm are showed in Fig. 6. to 11. The performance measurements are Correct Percentage, True Positive Rate, False Positive Rate, Precision, Recall and F-Measure.

6.1 The dataset

All data used in this experiment are collected from undergraduate students is one private university in Thailand. The status of students can be classified by status learning. The dataset are grouped into six classes. The input data set used in the Waikato Environment for Knowledge Analysis (WEKA) program, it has format extension ".arff" file. In Table 2. shows six classes of students status. The dataset has 35 nominal attributes as shown in Table 3, There are 20,914 instances, and as indicated in six classes.

Table 2. Classes of student status.

Class	Description
S	Normal
G	Graduated
T	Retired
R	Resign
Q	Lost Contact
L	Take Leave

Table 3. Attribute of dataset.

Attribute Names	Description
Sex	Sex
Zone_id	Curriculum
Round_id	Type of study
Spc_id	Scholarship
Nationality_id	Student's nationality
Religion_id	Student's religion
Region_sch	Previous school region
Region_address	Student address region
Fac_id	Faculty
Edulevel_id	Qualification for admission
Ent_id	Type of enrollment
Cert_grade_id	Grade for admission
Fth_income_id	Father's income
Mth_income_id	Mother's income
Fth_occ_id	Father's occupation
Mth_occ_id	Mother's occupation
Dept_id	Student's department
F_gpa_id	First semester's GPA
S_gpa_id	Second semester's GPA
Age_id	Student's age
Str_yr	Year of admission
Str_sem	Semester of admission
Extrac_id	Student's extraction
Total_grade_F	Total grade F
Total_grade_A	Total grade A
Total_grade_B_P	Total grade B Plus
Total_grade_B	Total grade B
Total_grade_C_P	Total grade C Plus
Total_grade_C	Total grade C
Total_grade_D_P	Total grade D Plus
Total_grade_D	Total grade D
Total_course_regis	Total course registration
Total_credit_all	Total credit registration
Total_credit_get	Total credit pass
Std_status	Student's status

6.2 Result

Fig. 6. shows correct percentage of two algorithm, C4.5 has the highest correct percentage value (85%) which can be implied that it is the most accuracy.

Correct percentage in C4.5 algorithm of Hold-out method Cross-Validation in range 40 – 90 % is highest correct percentage value which can be implied that more training set make enhance highest accuracy.

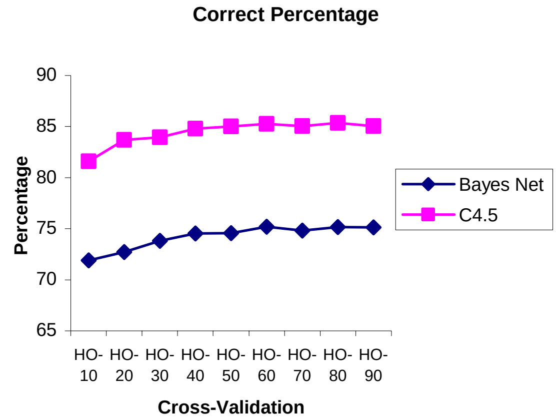

Figure 6. Comparison percentage of Bayesian Network and Decision Tree (C4.5) in correct percentage.

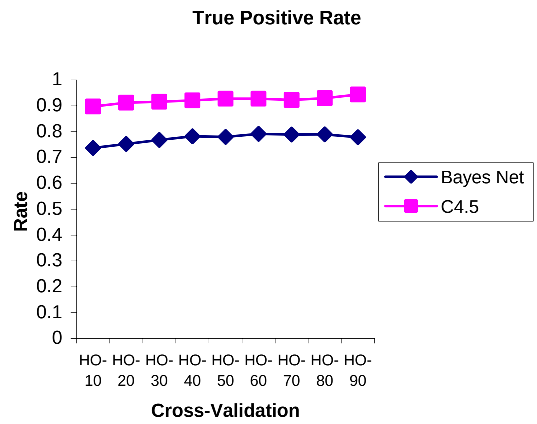

Figure 7. Comparison rate of Bayesian Network and Decision Tree (C4.5) in true positive rate.

Fig. 7. shows true positive rate of two algorithm, C4.5 has highest true positive rate (0.94) which can be implied that it is the most prediction accuracy. True positive rate in Hold-out method Cross-Validation in C4.5 and Bayesian Network do not have more effect on true positive rate accuracy.

Fig. 8. shows false positive rate of two algorithm, C4.5 has lowest false prediction value (0.05) which can be implied that it is the most prediction accuracy.

In 10 % of Hold-out method Cross-Validation C4.5 has highest false prediction value (0.083) and Bayesian Network has lowest false prediction value (0.073).

More than 10 % of Hold-out method Cross-Validation C4.5 has lowest false prediction value and Bayesian Network has highest false prediction value which can be implied that more training set make enhance accuracy.

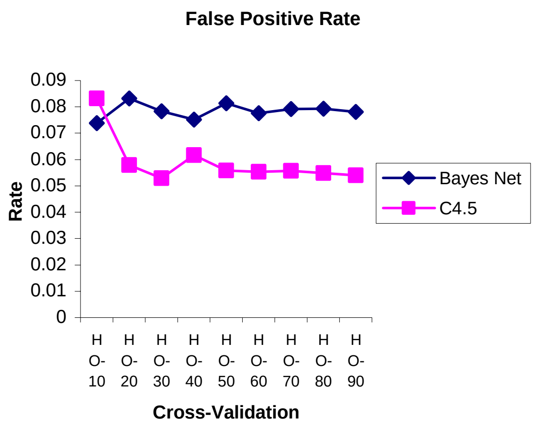

Figure 8. Comparison rate of Bayesian Network and Decision Tree (C4.5) in false positive rate.

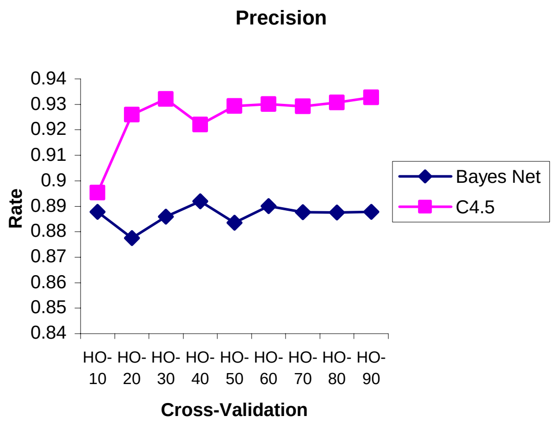

Figure 9. Comparison rate of Bayesian Network and Decision Tree (C4.5) in precision.

Fig. 9. shows precision of two algorithm, C4.5 has highest precision (0.93) which can be implied that it is the most true prediction accuracy.

In 10 % of Hold-out method Cross-Validation C4.5 and Bayesian Network has lowest precision. In more than 10 % C4.5 has highest precision and Bayesian Network has still lowest precision.

Fig. 10. shows recall of two algorithm, C4.5 has highest recall (0.9) which can be implied that it is the most prediction accuracy.

Hold-out method Cross-Validation do not have more effect on recall accuracy in C4.5 and Bayesian Network.

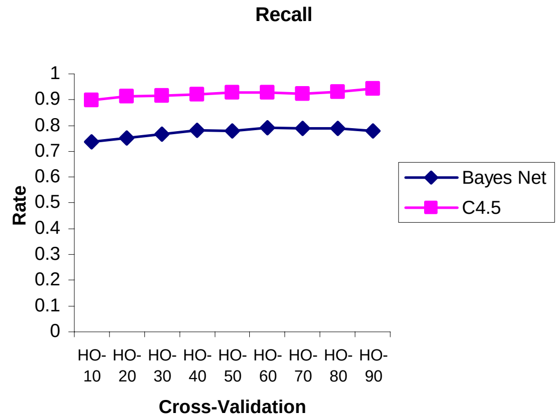

Figure 10. Comparison rate of Bayesian Network and Decision Tree (C4.5) in recall.

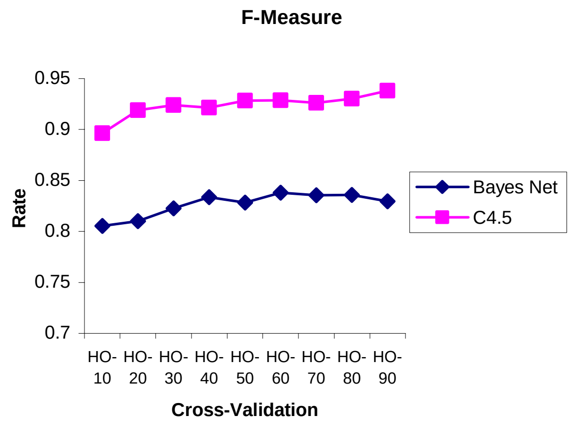

Figure 11. Comparison rate of Bayesian Network and Decision Tree (C4.5) in F-Measure.

Fig. 11. shows F-Measure of two algorithm, C4.5 has highest F-Measure value (0.95) which can be implied that it is the most accuracy.

In C4.5 and Bayesian Network Hold-out method Cross-Validation at 10 % has lowest F-Measure. More than 10 % in C4.5 has highest F-Measure and more than 30 % in Bayesian Network has highest F-Measure.

7. Conclusions and future work

This paper proposes a comparison of two algorithms for prediction undergraduate student success. In this study, we use Decision Tree (C4.5) and Bayesian Network in practical experiments with Hold-out cross-validation. The prediction performances of two classifiers are measured by six indices used for evaluating the efficiency of classification. The indices include Correct Percentage, True Positive rate, False Positive rate, Precision, Recall and F-Measure. In the prediction accuracy, Decision Tree (C4.5) is higher than Bayesian Network, and error rate prediction Decision Tree (C4.5) has the smaller numbers. The university office can use Decision Tree to build the model in order to predict the success of undergraduate students. This model will be helpful in guiding problematic students to overcome the difficulties in their studies.

Our future work is applying data mining technique for prediction. In order to increase the prediction power of classification, alternative feature selection such as Genetic Algorithm might be apply to select importance attributes before classification.

References

[1] I. H. Witten and E. Frank, “Data Mining: Practical Machine Learning Tools and Techniques,” Second Edition, Morgan Kaufmann, San Francisco, 2005.

[2] M. Xenos, “Prediction and assessment of student behavior in open and distance education in computers using Bayesian networks,” Computer & Education Volume 43, Issue 4, pp. 345-359, 2004.

[3] P. Garcia, A. Amandi, S. Schiaffino and M. Campo, ”Evaluating Bayesian networks’ precision for detecting students’ learning styles,” Computer & Education Volume 49, Issue 3, pp.794-808, 2007.

[4] S. Mukoolskunpibal and S. Kitisin, “Performance comparison of C4.5, ADTree and Naïve Bayes classification algorithms on international postages with narcotics,” National Joint Conference on Computer Science & Software Engineering, 2007.

[5] J. Yingkuachat, B. Kijsirikul and P. Praneetpolgrang, ”A Prediction of Higher Education Students’ Graduation with Bayesian Learning and Data Mining,” Research and Innovations for Sustainable Development Conference, 2006.

[6] P. Yamansabideen, B. Kijsirikul, P. Jirapongsuwan and P. Praneetpolgrang, ” An Application of Data Mining in Customer Relationship Management for Higher Education Students,” Research and Innovations for Sustainable Development Conference, 2006.

[7] L. Sun and P.P. Shenoy, ”Using Bayesian networks for bankruptcy prediction: Some methodological issues,” European Journal of Operational Research,,Volume 180, Issue 2, pp.738-753, 2007.

[8] T. Hidekazu, A. El-Sayed, F. Masao, M. Kazuhiro, T. Kazuhiko and J. Aoe, ” Estimating sentence types in computer related new product bulletins using a decision tree,” Information Sciences, Volume 168, Issues 1-4, pp.185-200, 2004.

[9] S. Lee, S. Lee and Y. Park, ”A prediction model for success of services in e-commerce using decision tree: E-customer’s attitude towards online service,” Expert Systems with Applications, Volume 33, pp.572-581, 2007.

[10] R. R. Bouckaert, Bayesian Network Classifiers in Weka, Department of Computing Science, University of Waikato, New Zealand, 2005.

[11] B. Kijsirikul, Artificial Intelligence, Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, 2003 (in Thai).

[12] K. Waiyamai, T. Rakthanmanon and C. Songsiri, “Data Mining Techniques for Developing Education in Engineering Faculty,” NECTEC Technical Journal, vol. III, no.11., pp. 134-142, 2001.

[13] B. Paulo and J. M. Silva, Mining On-line Newspaper Web Access logs, Departamento de informatica Faculdade de Ciencias, Universidade de Lisboa, 2001.

[14] R. Kirkby and E. Frank, Weka Explorer User Guide, University of Waikato, New Zealand, 2005.

[15] H. Almuallin and T. G. Dietterich, “Efficient Algorithms for Identifying Relevant Features,” In Proceedings of the Ninth Canadian Conference on Artificial Intelligence, Vancouver, BC. San Francisco: Morgan Kaufmann, pp. 38-45. 1992.

[16] R. R. Bouckaert, “Bayesian Network Classifiers in Weka,” Working Paper 14/2004, Department of Computer Science, University of Waikato, New Zealand, 2004.

[17] R. Kohavi, “A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection,” Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Canada, San Francisco: Morgan Kaufmann, pp. 1137-1143, 1995a.

[18] P. Langley, W. Iba and K. Thompson, “An Analysis of Bayesian Classifiers,” In W. Swartout, editor, Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA. Menlo Park, CA: AAAI Press, pp.223-228, 1992.

[19] Y. Wang and I. H. Witten, “Modeling for Optimal Probability Prediction,” In C. Sammut and A. Hoffmann, editors, Proceedings of the Nineteenth International Conference on Machine Learning, Sydney, Australia. San Francisco: Morgan Kaufmann, pp. 650-657, 2002.

[20] Y. Yang and G. I. Webb, “Proportional k-interval Discretization for Naïve Bayes Classifiers,” In L. de Raedt and P. Flach, editors, Proceedings of the Twelfth European Conference on Machine Learning, Freiburg, Germany, Berlin: Springer-Verlag, pp. 564-575, 2001.

[21] D. Heckerman, D. Geiger, and D. M. “Chickering, Learning Bayesian Networks: The Combination of Knowledge and Statistical Data,” Machine Learning 20 (3):197-243, 1995.

[22] D. Grossman, and P. Domingos, “Learning Bayesian Network Classifiers by Maximizing Conditional Likelihood,” In R. Greiner and D. Schuurmans, editors, Proceedings of the Twenty-First International Conference on Machine Learning, Banff, Alberta, Canada. New York: ACM, pp.361-368, 2004.

[23] G. John and P. Langley, “Estimating Continuous Distributions in Bayesian Classifiers,” In P. Besnard and S. Hanks, editors, Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence,

Montreal, Canada. San Francisco: Morgan Kaufmann, pp. 338-345. 1995.
[24] D. J. Hand, H. Mannila and P. Smyth, Principles of Data Mining, Cambridge, MA: MIT Press, 2001.
[25] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, New York: Springer-Verlag, 2001.
[26] G. H. John, Enhancements to the Data Mining Process, Ph.D Dissertation, Computer Science Department, Stanford University, Stanford, CA. 1997.
[27] Z. Zheng and G. Webb, “Lazy learning of Bayesian rules,” Machine Learning 41(1): 53-84, 2000.
[28] J. Swets, “Measuring the accuracy of Diagnostic Systems,” Science, 240: 1285-1293. 1998.
[29] M. Sahami, S. Dumais, D. Heckerman and E. Horvitz, “A Bayesian Approach to Filtering Junk e-Mail,” Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, Madison, WI. Menlo Park, CA: AAAI Press, pp. 55-62, 1998.
[30] R. Kass and L. Wasserman, “A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion,” Journal of the American Statistical Association, 90:928-934. 1995.
[31] M. V. Johns, An Empirical Bayes Approach to Nonparametric Two-way Classification, H. Solomon, editor, Studies in item analysis and prediction, Palo Alto, CA: Stanford University Press, 1961.
[32] G. H. John and P. Langley, “Estimating Continuous Distributions in Bayesian Classifiers,” In P. Besnard and S. Hanks, Editors, Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montreal, Canada. San Francisco: Morgan Kaufmann, pp.338-345. 1995.
[33] D. E. Appelt, “Introduction to Information Extraction Technology,” Tutorial, Int. Joint Conf. on Artificial Intelligence IJCAI’99. Morgan Kaufmann, San Mateo, 1999. Tutorial notes available at www.ai.sri.com/~appelt/ie-tutorial.
[34] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules in Large Databases,” In J. Bocca, M. Jarke, and C. Zaniolo, editors, Proceedings of the International Conference on Very Large Databases, Santiago, Chile. San Francisco: Morgan Kaufmann, pp. 478-499. 1994.
[35] R. Agrawal, T.Imielinski and A. Swami, “Database Mining: A Performance Perspective,” IEEE Transactions on Knowlegde and Data Engineering 5(6): 914-925. 1993.
[36] D. Pyle, Data preparation for Data Mining, San Fanscisco: Morgan Kaufmann.
[37] Quinlan, J. R. 1986. Induction of Decision Trees, Machine Learning 1(1):81-106. 1999.
[38] S. M. Weiss and N. Indurkhya, Predictive Data Mining: A Practical Guide, San Francisco: Morgan Kaufmann, 1998.
[39] J. P. Bigus, Data Mining with Neural Networks, New York: McGraw Hill, 1996.
[40] P. Adriaans and D. Zantige, Data Mining, Harlow, England: Addison Wesley, 1996.
[41] A. Berson and J. S. Smith, Data Warehousing, Data Mining, and OLAP, McGraw-Hill, Inc., 1997.
[41] S. Mitra and T. Acharya, Data Mining Multimedia, Soft Computing, and Bioinformatics, John Wiley & sons, Inc., New Jersey. 2003.
[42] D. Olson and Y. Shi, Introduction to Business Data Mining, McGraw-Hill, International Edition, 2007.
[43] J. R. Roiger and W. M. Geatz, Data Mining: A Tutorial - based Primer, Addison Wesley Publishing Company, 2003.
[44] G. Piatetsky-Shapiro and W. J. Frawley, Knowledge Discovery in Databases, MTT Press, 1991.
[45] J. Han and M. Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 2001.
[46] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, Advances in Knowledge Discover and Data Mining, AAAI/MTT Press, 1996.
[47] L. Talavera and E. Gaudioso, “Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces,” Proceedings of the Sixteenth European Conference on Artificial Intelligence, 2004.
[48] C. Ordonez, “Comparing association rules and decision trees for disease prediction,” Proceedings of the international workshop on Healthcare information and knowledge management, pp. 17-24, 2006.
[49] C. X. Ling, Q. Yang, J. Wang and S. Zhang, “Decision trees with minimal costs,” Proceedings of the twenty-first international conference on Machine learning, 2004.
[50] S. Esmeir and S. Markovitch, “Anytime Learning of Decision Trees,” The Journal of Machine Learning Research, Volume 8, pp. 891-933, 2007.
[51] R. Potharst and J. C. Bioch, “Decision trees for ordinal classification,” Intelligent Data Analysis, Volume 4, Issue 2, pp. 97-111, 2000.
[52] D. Kalles, A. Papagelis and E. Ntoutsi, “Induction of decision trees in numeric domains using set-valued attributes,” Intelligent Data Analysis, Volume 4, Issue 3,4, pp. 323-347. 2000.
[53] M. Danielson, L. Ekenberg and A. Larsson, “Distribution of expected utility in decision trees,” International Journal of Approximate Reasoning, Volume 46, Issue 2, pp. 387-407, 2007.
[54] J. W. Kim, B. H. Lee, M. J. Shaw, H. Chang and M. Nelson, “Application of Decision-Tree Induction Techniques to Personalized Advertisements on Internet Storefronts,” International Journal of Electronic Commerce, Volume 5, Issue 3, pp. 45-62, 2001.
[55] S. Nijssen and E. Fromont, “Mining optimal decision trees from itemset lattices,” Proceeding of the thirteenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 530-539, 2007.
[56] V. T. Chakaravarthy, V. Pandit, S. Roy, P. Awasthi and M. Mohania, “Decision trees for entity identification: approximation algorithms and hardness results,” Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 53-62, 2007.