

Using Decision Tree C4.5 Algorithm to Predict VARK Learning Styles

Oranuch Pantho¹
and Monchai Tiantong²

King Mongkut's University of Technology North Bangkok, Thailand

¹ammubon@hotmail.com

²monchai@kmutnb.ac.th, www.kmutnb.ac.th

Abstract - This research has an objective to classify VARK learning styles of learners by using Decision Tree C4.5 algorithm. Data concerning learning styles of learners were collected via a questionnaire responded by 1,205 students. The collected data, which included sex, age, major, year of education, GPA, previous educational background, and VARK learning styles of learners, were then classified by using Decision Tree C4.5 algorithm run on Weka software with 10-fold cross validation technique. The study results revealed that the learners' VARK learning style classification based on Decision Tree C4.5 algorithm yielded an accuracy at 83.40%, and a total of 108 rules were obtained. It can be concluded that Decision Tree C4.5 algorithm can be used to classify VARK learning styles of learners efficiently.

Keywords - Decision Tree C4.5 Algorithm, VARK Learning Styles

I. INTRODUCTION

Currently, instructors have increasingly used e-Learning systems for their teaching. E-Learning allows learners to learn at any time without any limitations in terms of time and place. However, learning with e-Learning seems to be not very successful. This is because, in e-Learning, learners have to learn by themselves. Learners may sometimes feel bored when e-Learning presents content that is not in agreement with their proficiency or learning styles. A learning style is a unique characteristic of each individual [1]. By definition, a learning style refers to a physical

characteristic, a thought, and a feeling that a person uses for perceiving, responding, and interacting with the environment consistently. Therefore, learning styles are important and have influence toward learning motivation and efficiency of learners [2]. Instructors should investigate learning styles of learners in order to plan appropriate learning activities for more efficient perception of learners. There are various questionnaires available for examining learning styles of learners, such as Kolb's Learning Style Inventory (LSI), Myers-Briggs Type Indicator, Felder-Silverman Learning Style Model, and VARK Learning Style. This research chose to study learning styles of learners based on the VARK (Visual, Aural, Read/Write, Kinesthetic) model, which is a concept developed by Fleming [3]. VARK model is very popular for examining learners' learning styles because it uses a simple and practical technique for classifying learners based on their most used senses of learning namely vision, listening, reading/writing, and practicing.

Before starting to teach, instructors should know learning styles of learners who may have different fundamental characteristics and skills. Surveying learners' learning styles via a questionnaire prior to taking a class may take a long time, and the results might contain some errors due to negligence of learners in answering the questionnaire. Therefore, this research will use a technique called data mining, which can help analyze data correctly and quickly. Several researchers have applied data mining for analyzing learning styles of learners. For example, Ba-Omar et al. [4] let learners answer an online VARK questionnaire

before studying the lessons via e-Learning. Then they took the data concerning study behavior and web access of the learners from the web log to be analyzed with a data mining technique in order to find learning courses of learners that reflected their different learning styles. A research of Norwawi et al. [1] classified learners in the Computer course according to their learning styles by applying a data mining technique called decision tree and a k-means grouping technique. They found that the learners had different learning styles. In this research, the applied classification technique was Decision Tree C4.5 algorithm. This technique was chosen due to its popularity and its capability in classifying data with the greatest accuracy. For instance, the study of Ahmad and Shamsuddin [5] compared the efficiency results from determining learning styles of learners with 3 classification techniques, namely Naïve Bayes, Decision Tree, and Rule-based techniques. They found that the decision tree algorithm yielded the highest accuracy. A study of Petchboonmee [6] applied decision tree algorithm and Naïve Bayes for analyzing Kolb's learning styles, and found that the decision tree technique was more accurate. A research of Kaewkiriya [7] classified learning styles of learners by using ID3 algorithm, C4.5 algorithm, NBTree, Naïve Bayes, and Bayes Net techniques, and found that Decision Tree C4.5 algorithm gave the highest accuracy. Therefore, this research chose to use Decision Tree C4.5 algorithm for predicting VARK learning styles of learners with an aim to allow instructors to know learning styles of learners quickly. In addition, this research is expected to provide a guideline for organizing teaching and learning activities in e-Learning systems to be more appropriate to each learner. This approach would eventually motivate learners to learn with higher intention and have greater efficiency in their learning.

II. LITERATURE REVIEW

A. VARK Learning Styles

Learning styles of learners can be classified into several styles depending on the concept used in analyzing and studying. This research

used the VARK model, developed by Neil Fleming, for classifying learning styles, which is hereby referred to as VARK learning styles. The learners were presented with a questionnaire containing 16 items, each of which has 4 choices. The questions will help classify individuals according to their 4 channels of perception, which imply 4 learning styles namely: 1) Visual learning style (V), which is a style of learners who prefers to learn via images or visual media; 2) Aural learning style (A), which is used by learners who prefers to listen rather than reading texts or watching images; 3) Read/Write learning style (R), which is a style of learning via reading or writing; and 4) Kinesthetic learning style (K), which is used most by learners who prefer to learn via experiments, tests, activities, or experience [8].

B. Decision Tree C4.5

Data classification is a process to figure out a model to be used for explaining each type of data. The main objective of classification is to predict a group of something. Such group is determined by relying on a training set of data, which has already been assigned groups or classes prior to the classification. Prediction is then conducted with a testing set of data, which is similar to the training set but has not been grouped. The result from the process is a model for classifying data. There are various techniques or algorithms that can be used for classifying data, such as Decision Tree, Bayes, Support Vector Machines, K-nearest Neighbors, Gaussian Mixture Model, and RBF [9, 10].

Decision Tree C4.5 algorithm is a mathematical model for classifying data by making decisions in a form of a tree. Decision Tree algorithm requires a supervised or trained data set. Classification results from Decision Tree algorithm have a characteristic of having structure like a tree. In this tree, there are nodes that represent attribute of data. The branches show the test results, and the leaf nodes show class names. Decision Tree algorithm predicts classes of the data by starting from the root nodes, and then continues through branches according to

results from the test toward the next fruit node. This testing is repeated until arriving at a leaf node, and the prediction result will be shown.

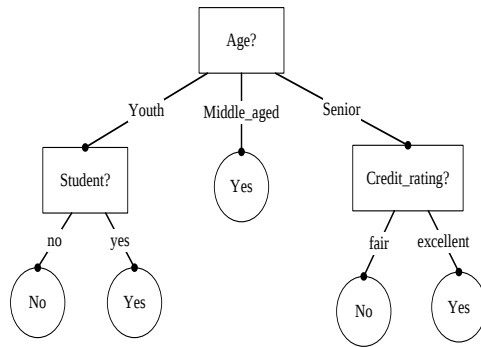


Fig. 1 Structure of a Decision Tree [10]

The use of decision tree algorithm makes it easy to convert the results into rules for classifying data. An advantage of data classification with decision tree algorithm is that it shows relationships of classes clearly by using graphic for better visualization of the data classification process. A decision tree diagram can indicate outcomes that may happen from each alternative. In addition, it helps analyze the order of decisions and the results from making decision with each approach. Decision tree algorithm is suitable for classifying data that do not have too many classes. A disadvantage of this technique is that it is complicated to make decisions when being used with data having a lot of classes.

C. WEKA

Weka software has been developed since 1997 by Waikato University in New Zealand. It is freeware application software developed under GPL License. Its source code has been written entirely with JAVA language. Weka operates by emphasizing on machine learning

and data mining. The software consists of sub-modules to be used for managing data. It can be controlled by using either GUI or command, and can be run under several operating systems. Developers can extend capabilities of this software. It is a tool that contains various algorithms to be used for data mining purpose. These algorithms can be chosen via two channels: one is from the provided algorithm tool sets and the other is from additional algorithms that have been written as new tool sets. To work on the data, these tool sets have multiple functions including Classification, Regression, Clustering, Association rules, Selection, and Visualization.

III. METHODOLOGY

This research was conducted by collecting data from a questionnaire distributed to learners who were studying at the undergraduate level of Chiang Mai Rajabhat University during the first semester of 2013 academic year. The obtained data included sex, age, major, year of education, GPA, previous educational background, and VARK learning styles derived from the VARK questionnaire.

The total of 1,205 responders were then classified according to their given data. It appeared the numbers of learners who were classified into Visual, Aural, Read/Write, and Kinesthetic learning styles were 167, 190, 567, and 281 students, respectively, as shown in Fig. 2. The data were saved in *.CSV file format so that it is appropriate to be processed with Weka software. An example of the collected data is shown in Table I.

Current relation		Selected attribute	
Relation: Data (5-weka.filters.unsupervised.attribute.Remove-R7-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-test)		Name: Class	
Instances: 1205		Meaning: 0 (0%)	
Attributes: 7		Distinct: 4	
Type: Nominal		Unique: 0 (0%)	
Attributes			
All None Invert Pattern			
No.	Name	No.	Label
1	Sex	1	A
2	Age	2	K
3	Branch	3	R
4	Year	4	V
5	AvgGrade		
6	OldEdu		
7	Class		
		Count	
			190
			281
			567
			167

Fig. 2 Variables Used for Data Classification

TABLE I
EXAMPLE OF DATA FOR THE VARIABLES
REPRESENTATION WITH A .CSV FILE

Sex	Age	Branch	Year	AvgGrade	OldEdu	Class
1	3	ComEd	4	2	1	A
1	2	ComSci	2	2	1	A
2	3	IT	1	2	2	A
1	2	IT	2	1	1	V
2	2	Math	3	2	1	K
2	3	Math	4	2	1	R
2	1	Other	1	3	2	R
1	2	Other	3	2	2	R
1	2	Web	3	2	1	V
1	2	Web	2	2	1	K
...	...	...	...	...	...	...

Data classification applied to VARK learning styles is a way to predict and assign data to be in a specified group. This research set 4 groups of learning styles namely Group V (Visual), Group A (Aural), Group R (Read/Write), and Group K (Kinesthetic). In addition, the data were divided into 2 parts. Part 1 is the training set data, which is used to train the system to know about groups that the data records should be assigned to so that a classification model can be created. Part 2 is

the testing set data, which consists of data in the same format as the training set data. However, these data in Part 2 have not been given group names, as shown in Fig. 3.

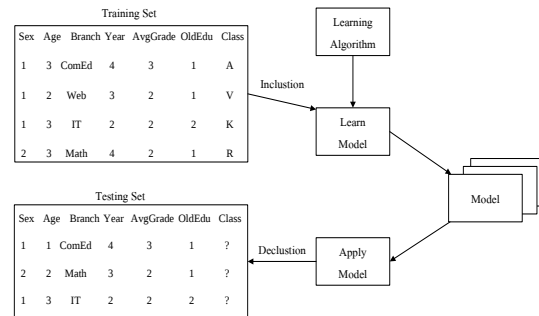


Fig. 3 Methodology for Classifying VARK Learning Styles

IV. RESULTS OF THE RESEARCH

This research classified learners into 4 groups of VARK learning styles by using Weka software to run Decision Tree C4.5 algorithm with a Cross Validation technique. Cross validation is a process to verify the results by randomly selecting sample data in the two data sets equally. The number of Folds in the Cross Validation process was set as 10. The classification result yielded an accuracy at 83.40%, as shown in Fig. 4, The resulted decision tree obtained from classification with Decision Tree C4.5 algorithm is shown in Fig. 5.

```

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1005          83.4025 %
Incorrectly Classified Instances    200          16.5975 %
Kappa statistic                    0.7561
Mean absolute error                 0.1113
Root mean squared error             0.2566
Relative absolute error              32.7202 %
Root relative squared error         62.2201 %
Total Number of Instances          1205

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.7	0.042	0.756	0.7	0.727	0.903	A
	0.918	0.08	0.777	0.918	0.842	0.949	K
	0.901	0.064	0.926	0.901	0.913	0.953	R
	0.617	0.04	0.71	0.617	0.66	0.879	V
Weighted Avg.	0.834	0.061	0.834	0.834	0.832	0.934	

```

=== Confusion Matrix ===
 a  b  c  d  <-- classified as
133 22 21 14 | a = A
 7 258 5 11 | b = K
18 21 511 17 | c = R
18 31 15 103 | d = V

```

Fig. 4 Accuracy of the Classification Process with Decision Tree C4.5 Algorithm

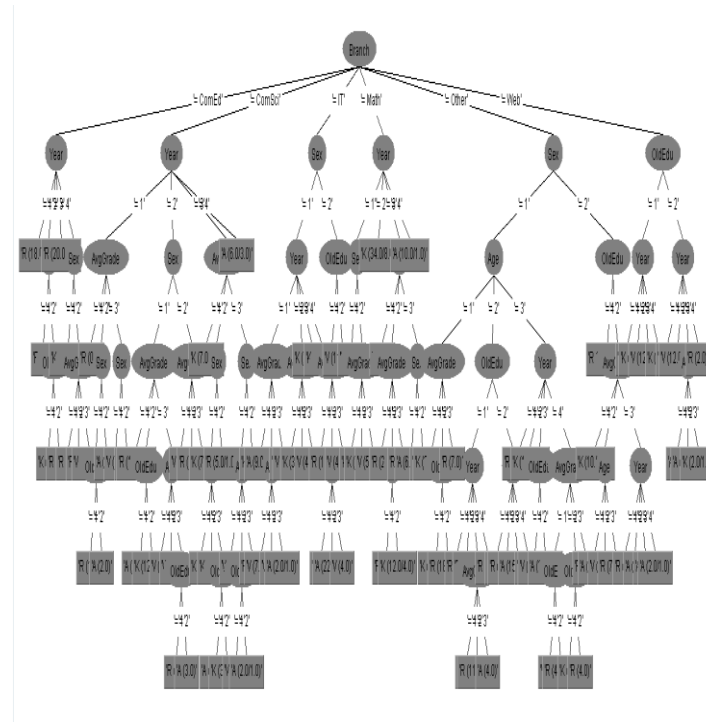


Fig. 5 Model for Predicting VARK Learning Styles Using Classification Rules with Decision Tree C4.5 Algorithm

Fig. 5, shows characteristic of the decision tree obtained from classifying VARK learning styles of the learners. According to the decision tree, 108 rules were obtained. Among them, 22, 24, 37, and 25 rules are for Visual learning style, Aural learning style, Read/Write learning style, and Kinesthetic learning style, respectively. These rules were made by taking the Decision Tree C4.5 to write into rules with IF...Then... clause, as shown for example in Table II. The obtained rules will be used for analyzing learners in the next procedure.

5	IF Branch=IT and Sex=1 and Year=1 and AvgGrade=3	Read/Write
6	IF Branch=IT and Sex=1 and Year=2 and AvgGrade=1	Visual
7	IF Branch=Math and Year=1 and Sex=2 and AvgGrade=3	Visual
8	IF Branch=Math and Year=2	Kinesthetic
9	IF Branch=Web and OldEdu=1 and Year=2	Kinesthetic
10	IF Branch=Web and OldEdu=1 and Year=3	Visual

V. CONCLUSIONS

TABLE II
SOME EXAMPLES OF RULES OBTAINED
FROM CLASSIFYING DATA WITH DECISION
TREE C4.5 ALGORITHM

No.	Rules from Decision Tree C4.5	VARK Learning Style
1	IF Branch=ComEd and Year=2 and Sex=1	Read/Write
2	IF Branch=ComEd and Year=2 and Sex=2 and OldEdu=1 and AvgGrade=1	Kinesthetic
3	IF Branch=ComSci and Year=2 and Sex=1 and AvgGrade=2 and OldEdu=1	Aural
4	IF Branch=ComSci and Year=2 and Sex=1 and AvgGrade=3 and Age=1	Visual

This research has an objective to classify VARK learning styles of learners by using Decision Tree C4.5 algorithm. The classification process yielded a result of learners' VARK learning styles with an accuracy at 83.40%. The decision tree obtained from the VARK learning style classification consists of 108 rules altogether. These rules can be divided for each class of learning style as follows: 22 rules are for Visual learning style, 24 rules are for Aural learning style, 37 rules are for Read/Write learning style, and 25 rules are for Kinesthetic learning style. The obtained rules will be

highly beneficial for instructors to organize learning/teaching that is in accord with learning styles of learners in order to motivate them to learn and allow them to understand the contents better.

In the next research, efficiencies of VARK learning style classification by using other algorithms, such as Naïve Bayes, Bayesian Network, and Rule-based technique, will be compared. This comparison is expected to reveal the best technique that can distinguish characteristics of learners with the highest efficiency. Furthermore, the derived model is intended to be used for analyzing learners before presenting them with e-Learning lessons of different courses that match their learning styles.

VI. ACKNOWLEDGMENT

The author would like to thank Misnistry Science and Technology, Thailand. That support scholarship.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

- [1] Norwawi, N., Abdusalam, S., Hibadullah, C., and Shuaibu, B. (2009). "Classification of students' performance in computer programming course according to learning style". Data Mining and Optimization, 2009 DMO'09 2nd Conference on IEEE, 2009, pp. 37-41.
- [2] Koorse, M., Cilliers, C.B., and Calitz, A.P. (2010). "Motivation and learning preferences of information technology learners in South African secondary schools". Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists. ACM, pp. 144-152.
- [3] Fleming, N.D. and Baume, D. (2006). "Learning Styles Again: VARKing up the right tree!". Educational Developments. Issue 7.4: pp. 4-7.
- [4] Ba-Omar, H., Petrounias, I., and Anwar, F. (2007). "A framework for using web usage mining to personalise e-learning". Advanced Learning Technologies, (2007). ICALT 2007 Seventh IEEE International Conference on IEEE, pp. 937-938.
- [5] Ahmad, N.B.H. and Shamsuddin, S.M. (2010). "A comparative analysis of mining techniques for automatic detection of student's learning style". Paper presented at the Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on.
- [6] Petchboonmee, P., Phonak, D., and Tiantong, M. (2015). "A Comparative Data Mining Technique for David Kolb's Experiential Learning Style Classification". International Journal of Information and Education Technology. Vol.5 No.9: pp. 672-675.
- [7] Kaewkiriya, T., Utakrit, N., and Tiantong, M. (2016). "The Design of a Rule Base for an e-Learning Recommendation System Base on Multiple Intelligences". International Journal of Information and Education Technology Vol.6 No.3: pp. 206-210.
- [8] Fleming, N.D. (1995). "I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom". Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA), 1995, pp. 308-313.
- [9] Silla, C.N., Koerich, A.L., and Kaestner, C.A. (2008). "Feature selection in Automatic Music Genre Classification". Paper presented at the Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on.
- [10] Han, J., Kamber, M., and Pei, J. (2011). "Data Mining Concepts and Techniques". USA: Working together to grow libraries in developing countries.