International Journal of Engineering & Technology, *7 (4.31) (2018) 28-33*

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Comparison of Different Feature Selection Techniques in Attribute Selection of Learning Style Prediction

Haziqah Shamsudin[1]*, Umi Kalsom Yusof[2], Maziani Sabudin[3]

School of Computer Science, Universiti Sains Malaysia, 11800, Georgetown, Pulau Pinang, Malaysia
**Corresponding author E-mail: haziqah.ucom12@student.usm.my*

Abstract

Learning style of specific users on an online learning system is determined based on their interaction and behavior towards the system. The most common online learning theory used in determining the learning style is the Felder-Silverman's Theory. Many researchers have proposed machine learning algorithms to establish learning style by using the log file attributes. Due to many attributes in predicting the learning style, the performance and efficiency of the classification and prediction are still poor; so far it is only between 58%-85%. This research is conducted to determine the most relevant attributes in predicting the learning style. First, three different feature selection methods are used to select the most relevant number of attributes, which are Rank by Importance (RbI), Recursive Feature Elimination (RFE) and Correlation. Next, five different classifiers are used to evaluate those selected feature selection methods. The classifiers are Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbour (KNN), Linear Discriminant Analysis (LDA) and Classification and Regression Tree (CART). From the experiments, RbI has proven to be the most effective feature selection method, with the accuracy improvement from 87% to 91%.

Keywords: *Classification Algorithm; Feature Selection; Learning Style; Online Learning.*

1. Introduction

Learning style is known as learning way or preference by the learner on how materials are presented, how to work with it and how to internalize information [1]. Identifying a student's learning style has several benefits such as making students aware of their strength and weaknesses when it comes to learning and it is also meant to be used in determining the preferences in learning of each student either in a traditional classroom or through an online learning based system. An online learning based system can be defined as the online system where there is interaction between students and system. Initially, in an online learning based system, the learning style of the user is determined by using available learning style questionnaires based on the selected learning style model. There are many learning style mod- els available, such as the Kolb Learning Theory, Felder-Silverman's Learning Style Model (FSLSM), Dunn and others. The most com- monly used learning style model is the FSLSM. It also incorporates different elements from different learning style models such as Kolb, Pask and Myers-Briggs [2]. However, when students are asked to fill in the questionnaire, the students take a longer time to fill it as the questions are long. Apart from that, the students are commonly refused to spend their time to fill in the questionnaire. This causes them to just put random answers [3].

So with that, researchers came out with a new alternative where they determine the learning style automatically [2]. This is done by collecting the log files of the interactive behavior of the user with the system. This consists of the number of mouse clicks, the time taken to do the task, number of views towards certain materials, and others. These attributes were then matched with the learning style model that they choose. From there, the results obtained are analyzed further and the learning style of the user is revealed.

Unfortunately, there is still some weakness in the method whereby researchers are not certain on which attributes are relevant enough to determine the learning style. Current results in terms of percentage of accuracy in determining the learning style still need a lot of improvement, where the percentage of accuracy ranges from 50%-85% [4, 5]. With that, several techniques of feature selection are used to address the problem of reducing irrelevant and redundant variables which are a burden on challenging tasks. Feature selection (variable elimination) helps in understanding data, reducing computation requirement, reducing the effect of the curse of dimensionality and improving the predictor performance [6].

There are three general classes of feature selection namely filters, wrappers, and embedded [7]. As the feature selection influences the prediction accuracy of any performance model, it is essential to study elaborately the effectiveness of student performance models in connection with feature selection techniques [8].

In this paper, three different feature selection methods are being used to find features that improve the overall prediction performance. Hence, by eliminating the dependent variables, the number of attributes can be reduced which can lead to improvement in classification performance. The feature selection methods used are Rank by Importance, Correlation, and Recursive Feature Elimination. Next, the attribute selected is tested and evaluated using five different classification algorithms which are SVM, RF, KNN, CART and LDA. The percentage of accuracy of each model is determined to find the best attribute that contributes in increasing the percentage of accuracy.

To remove an irrelevant feature, a feature selection criterion which can measure the relevance of each attribute with the output class/labels is required. From a machine learning point of view, if a system uses irrelevant attributes, it will use this information for new data which lead to a poor generalization. With that, this research is conducted to identify a feature selection method that will determine the most relevant attributes that can be used to predict learning styles used by the learner on the materials presented, how to work with it and how to internalize information [9]. It is meant to be used in determining the preferences in learning of each student either in a traditional classroom or an online learning based system. Online learning is described by most authors as access to learning experiences via the use of some technology [10, 11, 12]. Nowadays, online learning systems have been improvised further to have an adaptive system where the system has the ability to improvise or change its content or appearances based on feedback obtained from the user.

2. Related Work

2.1. Learning Style

A learning style is a student's consistent way of responding to and using stimuli in the context of learning [6]. Determining an accurate learning style of the user will lead to better adaptivity of the online learning system which will then increase the user's performance. There are many learning style model available in this area as mentioned by H.M.Truong [2] in the last 30 years, where over 70 theories were developed [2]. One of the most commonly used models is the Felder-Silverman's model, which differentiates learning styles through 4 different dimensions which are Perception, Input, Processing and Understanding. This theory is by far the most widely used in adaptive learning systems (accounted for 70.6%) of all papers in the survey conducted by H.M. Truong, [2].

FSLSM can describe the student's learning style in great detail. For the Perception dimension which consists of sensing and intuitive learning style it describes a preference for processing information. In this dimension, learners with sensing learning styles prefer to learn facts and concrete materials, using their sensory experiences of particular instances as a primary source. On the other hand, intuitive learners prefer to learn abstract learning material, such as theories and their underlying meanings, with general principles rather than concrete instances being a preferred source of information. The Input dimension consists of visual and verbal learning style. Visual learners prefer materials such as graphs, charts or videos, while verbal learners prefer words either written or spoken.

The third dimension which is the processing dimension consists of active and reflective learning style. Active learners prefer to learn by doing, experimentation and collaboration while reflective learners prefer to think the information and absorb it alone or in small groups. Lastly, for the understanding dimension it consists of sequential and global learning style, where sequential learners prefer information to be provided in a linear (serial) fashion and tend to make small steps through learning material while global learners tend to make larger leaps from non-understanding to understanding to require seeing the "big picture" before understanding a topic.

2.2. Feature Selection

Feature selection methods allow researchers a way to reduce computation time, improve prediction performance, and have better un- derstanding of the data in machine learning. The focus of feature selection is to select a subset of variables and still provide good prediction results. The computation requirement and prediction ac- curacy can be improved by applying feature selection technique. A subset of features is selected from the original features without any transformation, and the physical meanings of the original features

is maintains [7]. Higher number of features will result in over fit of learning models which will result in reduction of performance [13]. With that, a feature selection method is conducted to choose relevant features from the original features according to specify the relevant criterion, which leads to better learning performance with higher learning accuracy of classification, lower computational cost, and better model interpretability [14].

In this research, feature selection is being done to compare two cases in terms of number of features before and after feature selection. The goal of this task is to observe if feature selection achieves its intended objectives with the aspects of evaluation such as number of selected features, time, scalability, and learning model's performance. The feature selection model that is used in this research is the filter model and wrapper. In wrapper approach, there are different methods available such as Ranking and recursive feature elimination. Ranking method uses variable ranking techniques as the principle criteria for variable selection by order. Ranking method is used due to its simplicity and reported success for practical applications. A suitable ranking criterion is used to score the variables, and a threshold is applied to remove unsuitable variables. Ranking methods are filtering methods since they are applied before classification to filter out the less relevant variables. A basic property of a unique feature is to contain useful information about the different classes in the data. This property can be defined as feature relevance which provides a measurement of the features' usefulness in discriminating the different classes. Here the issue of relevancy of the feature has to be raised for example on how to measure the relevancy of the feature to the data or to the output.

2.3. Prediction Algorithm

In this paper prediction is used to evaluate the effectiveness of a feature selection method in increasing the performance of accuracy. The goal of prediction is to develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of data (predictor variables). In the first type the features of the model are used for prediction. This can be mainly used for analysis of students' performance. In the second type the output values are predicted based on the context. Prediction can be classified into three types classification, regression, and density estimation. In classification there are some popular methods including logistic regression, support vector machines, decision tree, Bayesian network and also neural network. Some complex techniques are required to forecast the values using combinations of various techniques as real-world EDM problems cannot be simply predicted [15].

For this paper, five different algorithms are used to evaluate the effectiveness of feature selection through increase in performance of learning style prediction in terms of percentage of accuracy and other evaluation measures. The algorithms chosen are Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF), Classification and Regression Tree (CART), and Linear Discriminant Analysis (LDA).

3. Methodology

In this research, there are two main stages involved. The first stage is to do a feature selection method to find the best attribute for each method. Then, a classification technique is used to mine data from feature selection steps. The performance of each feature selection method is compared with the classification technique based on selected performance evaluation measures.

3.1. Data Selection

The data used in this paper was taken from a research done by Renato [5]. The data is collected from the year 2012 to 2016. It contains a record of 507 students enrolled in Computer Technology courses which have successfully completed the Computer Programming 1 subject. This dataset consists of 15 different attributes. Table 1 shows the different attributes involved which is then divided into the respective dimensions of learning style according to the FSLSM theory. These attributes are then matched to a learning style model specified by the researcher. In this case, FSLSM is used which consists of four different dimensions. The dimensions involved are Input, Processing, Perception and Understanding. Even though there are four different dimensions involved (input, processing, perception and understanding) in this paper, the understanding dimension is not to be taken into consideration as it only contains 2 different attributes. In this paper, the learning style dimensions are first undergoing feature selection to select the most relevant attributes that match with it. It is then further analyzed by running the selected attributes with selected classifiers. The percentage of accuracy and other performance measure is taken into consideration and the best attribute is then selected.

3.3.2. Correlation based Feature Selection

Correlation Based Feature Selection (CFS) is a simple filter algorithm that ranks feature subsets according to a correlation based heuristic evaluation function [6]. This method will ignore the irrelevant features as they have low correlation with the class. It evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Redundant features should be screened out as they will be highly correlated with one or more of the remaining features. The CFS feature subset evaluation is:

$$M_s = \frac{k \overrightarrow{r_{cf}}}{\sqrt{k + k(k-1) r_{\overrightarrow{ff}}}} \qquad (1)$$

3.2. Method

There are 3 different feature selection methods being used in this paper which are Rank by Importance (RbI), Correlation based feature (Corre) and Recursive Feature Elimination (RFE). The aim of doing a feature selection is to improve the classification performance in the prediction of learning style. The most relevant attribute in predicting the learning style also can be determined by doing a feature selection. For the classification method, five different algorithms are used namely Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN), Classification and Regression Tree (CART) and Linear Discriminant Analysis (LDA). These algorithms are used to further evaluate to increase the percentage of accuracy in predicting the learning style when evaluating with different at- tributes.
Two different stages involved in fulfilling this experiment where the first stage is the feature selection method in selecting the most rele- vant attributes while for the second stage it involved in evaluating the selected attributes in predicting the learning style by using different classification algorithm. The stages are shown in Figure 1.

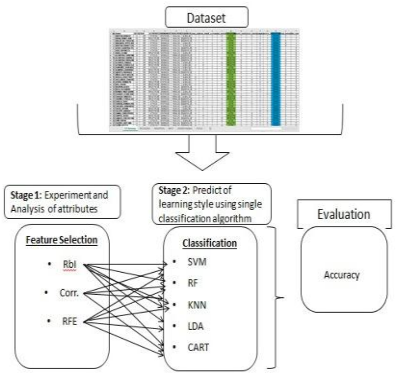


Fig. 1: Main Stages Involve in Selecting the Most Relevant Attributes

3.3. Feature Selection

3.3.1. Rank by Importance

Rank by Importance (RbI) rank features by their importance by using learning vector quantization (LVQ) to train a model and find feature importance. It is a prototype-based supervised classification and is the supervised counterpart of vector quantization systems. It uses varImp function to calculate variable importance for objects produced by the train data.
Where M_S is the heuristic merit of a feature subset S containing k attributes. $\overrightarrow{r_{cf}}$ is the main feature-class correlation and $r_{\overrightarrow{ff}}$ is the average feature inter-correlation. Corre will remove any attributes which have high correlation value.

3.3.3. Recursive Feature Elimination

Recursive Feature Elimination works by building a model repeatedly and choose either the best or worst performing feature. The model will then, repeating the process with the rest of the features while setting the feature aside. This process is applied until all features in the dataset are exhausted. The elimination process is then taking place and at the same time, the features are ranked accordingly. RF is used in the iteration to evaluate the model. The algorithm is configured to explore all possible subsets of the attributes.

3.4. Classification Algorithm

3.4.1. Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, each data is plotted as a point in n-dimensional space (where k is the number of features) with the value of each feature being the value of a particular coordinate. Then, classification is performed by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the coordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line) [16].
This method first divides the two class labels of the 3 different dimensions. Then it makes a classification by using a sampling method of 10-fold cross validation. The parameter is adjusted in terms of the kernel, the number of gamma value and the cost value. The best combination of parameters which yield highest percentage of accuracy and low misclassification error rate is selected as the best attributes for predicting learning style.

3.4.2. K-Nearest Neighbour

KNN classification is one of the most fundamental and simple classification methods and should be one of the first choices for a classification study when there is little to no prior knowledge about the distribution of the data [17]. KNN was developed from the need to perform discriminant analysis when reliable parametric estimates of probability densities are unknown or difficult to determine. NN is a non-parametric lazy learning algorithm.

3.4.3. Random Forest

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [18]. The generalization error for forests converges to a limit as the number of trees in the forest becomes large. The strength of the individual trees in the tree and the correlation within will decide the generalization error of the forest of tree classifier. Using a random selection of features to split each node yields error rates that compare favorably to Ada boost but are more robust with respect to noise.

Table 1: Attributes match to Dimension of FSLSM

No	Attribute Name	Description of Attribute	Learning Style	Dimension
1	Forum Post	Post more often in discussion forum	Active	PROCESSING
2	Forum View	Reading post but rarely posting by them- selves	Reflective	
3	Self assessment	Perform more self-assessment tests	Active	
4	Text Materials	Prefers learning material in textual form	Reflective	
5	Concrete Materials	Prefers concrete learning materials (facts, data)	Sensing	PERCEPTRON
6	Abstract Materials	Prefer abstract learning material (definition, theories, syntax, flowcharts)	Intuitive	
7	Examples	Prefer examples	Sensing	
8	Exercise rev	Prefers to review answers in graded exercise tests	Intuitive	
9	Visual Materials	Prefers learning materials supplemented with pictures, diagrams, graphs	Visual	INPUT
10	Video Materials	Prefers learning material presented in text or audio	Visual	
11	Text Materials	Prefers learning material in text or audio	Verbal	
12	Forum Post	Post more often in discussion forum	Verbal	
13	Course overview	Prefers overviews, outlines	Global	UNDERSTANDING
14	Nav euclidean distance	Prefers to go through the course step by step (linear)	Sequential	
15	Nav euclidean distance	Prefers to skipping the material (non-linear way)	Global	

3.4.4. Classification and Regression Tree

Classification and regression trees are machine-learning methods for constructing prediction models from data [19]. The models are ob- tained by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the divisioning can be represented graphically as a decision tree. Classification trees are designed for dependent variables that take a finite number of unordered values, with prediction error, measured in terms of misclassification cost.

3.4.5. Linear Discriminant Analysis

Suppose a learning set L of multivariate observations (i.e. input values in r), and suppose each observation is known to have come from one of K predefined classes having similar characteristics. These classes may be identified, for example, as number of exercise completed, time taken to complete the task, number of times post in the forum, mouse clicks, views on the visual materials or text material. To distinguish the known classes from each other, a unique class label is associated (or output value) with each class. The observations are then described as labeled observations. In each of these situations, there are two main goals:
Discrimination: Use the information in a learning set of labeled observations to construct a classifier (or classification rule) that will separate the predefined classes as much as possible.
Classification: Given a set of measurements on a new unla- belled observation, use the classifier to predict the class of that observation.

3.5. Performance Measure

3.5.1. Efficiency Measure

The performance of the proposed model is measured using a confusion matrix. It illustrates the accuracy of the solution to a classification problem. The confusion matrix contains information about actual and predicted classifications done by a classification system. Using data in the matrix, the performance is evaluated. Figure 2 shows the confusion matrix. The predicted TP and TN classifications are calculated based on the formula as follows:

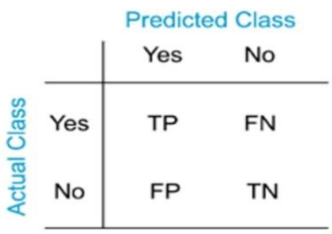


Fig. 2: Confusion Matrix

3.5.2. Objective Function

Objective function is an equation to be optimized by using cer- tain constraints and with variables that need to be minimized or maximized using nonlinear programming techniques. An objective function can be the result of an attempt to express a business goal in mathematical terms for use in decision analysis, operation research or optimization studies. In this research the objectives function is needed to maximize the accuracy value of the learning style predic- tion and user performance. The calculation is determined based on the confusion matrix as shown in Figure 2. Accuracy is needed to determine how often the classifier is correct.

4. Result and Evaluation

This subsection presents and discusses the results of each feature selection method, then continues with a discussion on the percentage of accuracy of classification algorithms in predicting the learning style. Subsequently, the results are compared with previous work from literature based on its accuracy. Table 2 shows the result of attribute selected according to its respective dimension.

Table 2: List of selected Attributes using different Feature Selection Method

Feature Selection Method	Selected Attribute	Dimension
RbI	2, 3, 1	PROCESSING
Corre	1, 3, 4	
RFE	2, 3, 1	
RbI	5, 8, 7	PERCEPTION
Corre	5, 6, 7, 8	
RFE	5, 8, 6, 7	
RbI	10, 9	INPUT
Corre	9, 10, 11, 12	
RFE	10, 9, 12	

From the table it turns out that for perception and input dimensions, Corre retains all the attributes as it is considered relevant. Rank by importance and RFE is also rearranging the attribute by choosing the highest importance of the attributes in the dataset. For the Input dimension both methods specify video material as the most important attribute in the dimension with the highest value being 0.86. The result is the same for both perception and processing dimensions.

Table 3 shows the percentage of accuracy of the five different classifiers selected against the feature selection method. From the table, it is observed that RbI gives the highest percentage of accuracy for the input dimension compared to the other dimensions. For the perception dimension, correlation based feature selection gave a higher percentage of accuracy compared to the other methods. Lastly, for the processing dimension, RFE gave a higher percentage of accuracy for all the classifiers compared to the method. But, in overall RbI has a constant range of percentage accuracy when validated using the classification accuracy which ranges from 0.84-0.94(%) compared to the other methods. For RFE, the percentage of accuracy for the classifier range from 0.82-0.93(%) and lastly, for Corre the classification accuracy ranges from 0.80-0.93(%).

Table 3: Overall value of accuracy feature selection against classifier

Method	SVM	RF	LDA	KNN	CART	Dimension
All feature	0.90 (2)	0.90 (2)	0.86 (3)	0.90 (2)	0.87 (2)	PROCESSING
RbI	0.90 (2)	0.92 (1)	0.88 (2)	0.93 (1)	0.86 (3)	
CORRE	0.89 (3)	0.80 (3)	0.80 (4)	0.81 (3)	0.87 (2)	
RFE	0.93 (1)	0.92 (1)	0.91 (1)	0.93 (1)	0.91 (1)	
All feature	0.84 (4)	0.86 (2)	0.82 (3)	0.85 (3)	0.88 (1)	PERCEPTION
RbI	0.89 (2)	0.86 (2)	0.86 (2)	0.86 (2)	0.84 (2)	
CORRE	0.91 (1)	0.93 (1)	0.88 (1)	0.87 (1)	0.82 (3)	
RFE	0.85 (3)	0.86 (2)	0.77 (4)	0.85 (3)	0.82 (3)	
All feature	0.89 (3)	0.84 (2)	0.86 (2)	0.87 (3)	0.89 (4)	INPUT
RbI	0.94 (1)	0.92 (1)	0.86 (2)	0.92 (1)	0.93 (1)	
CORRE	0.86 (4)	0.82 (3)	0.85 (3)	0.86 (4)	0.90 (3)	
RFE	0.93 (2)	0.84 (2)	0.88 (1)	0.90 (2)	0.91 (2)	

To observe the relationship between the number of attributes selected and the relevant attributes in regards to the percentage of accuracy, a graph is plotted and shown as in Figure 3a, 3b, 3c for the Processing, Perception, and Input dimension respectively.

From Figure 3a, all 3 methods select the same number of attributes which are 3 attributes out of 4 attributes. However for the RbI and RFE method, both select the same attributes, which are attributes number 1, 2 and 3 while for the Corre it selects attribute number 1, 3 and 4. The attributes selected by the RbI and RFE methods gave a higher percentage of accuracy compared to the attributes selected by Corre. This is because when using the Corre method, it shows that attribute number 2 (Forum View) is not relevant and are highly correlated, so it was removed. This shows that, the percentage of accuracy not only depends on the number of attributes but also depend on the importance level of the attribute which can give higher percentage of accuracy.

In terms of relations between number of attributes selected with the percentage of accuracy, the comparison can be seen from Figure 3c. It can be observed that when using the LDA classifier with (k = 2) where k=number of attributes the accuracy value is 0.86% compared to processing dimension with (k = 3), the accuracy value is 0.93%. On the other hand, the percentage of accuracy for the input dimension when using SVM classifier is higher with a value of 0.94% compared to processing dimension when using SVM only get 0.90% of accuracy value. This shows that an optimal number of attributes subset needed in order to obtain the best model [5]. Overall, all of the dimensions produce a higher percentage of accu- racy, when undergoes feature selection method. This shows that by doing a feature selection it can improve the accuracy of prediction of the learning style. The attributes selected after undergo feature selection fulfill the past literature discussion where the attributes used must have contribution on determining the learning style of the stuents [4, 20, 5]. But, the best feature selection method still cannot be decided because there is an inconsistency in terms of the percentage of accuracy for all the dimensions. With that, an average value for all the dimension is to be taken into consideration to select the best feature selection method. The result is tabulated in Table 4.

Table 4: Average value of accuracy Feature selection against classifier

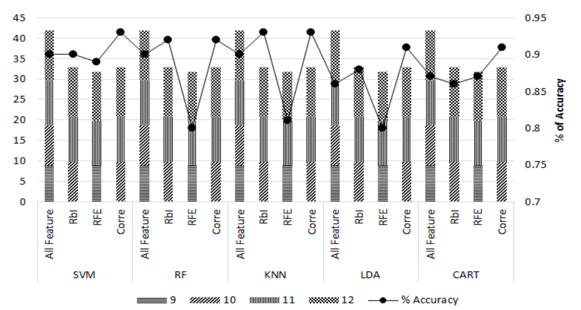
Method	SVM	RF	LDA	KNN	CART
All feature	0.87(4)	0.87(2)	0.85(2)	0.87(3)	0.87(2)
Rbl	0.91(1)	0.90(1)	0.86(1)	0.90(1)	0.88(1)
RFE	0.90(2)	0.87(2)	0.85(2)	0.89(2)	0.87(2)
CORRE	0.89(3)	0.85(3)	0.84(3)	0.85(4)	0.86(3)

From table 4, the result of the percentage of accuracy in average is calculated. It is observed that RbI produce a better result in terms of percentage of accuracy. RbI works in a more detailed manner in selecting the most relevant attributes as it works with the aid of algorithm called LVQ. LVQ select an optimal value of k and size involve in selecting the highest relevant attributes. Different dimensions have different value of k and size in determine the most relevant attributes. In RbI, after the value of k and size is determined, it proceeds in determined the most relevant attribute and ranked them accordingly. This detail process leads to a more accurate in determining the most relevant attributes of the datasets.

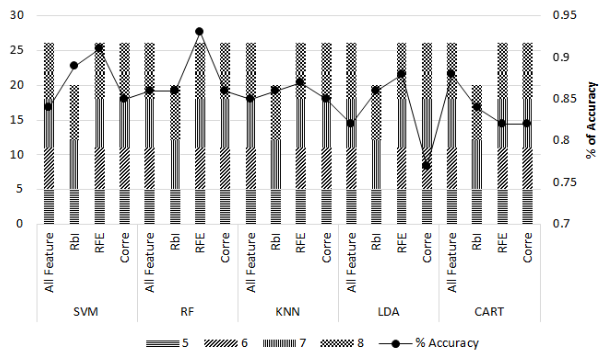
The parameter for all the five classifiers used in this paper is left at default. This is because, the main objective of this paper is to see the increment value of accuracy before and after undergoing a feature selection method. The comparison of RbI feature selection in terms of classification accuracy is tabulated in Table 5. The result is compared with previous work from literature which use the same learning theory which is the FSLSM but without any feature selection method.

Table 5: Comparison of Average accuracy results with literature

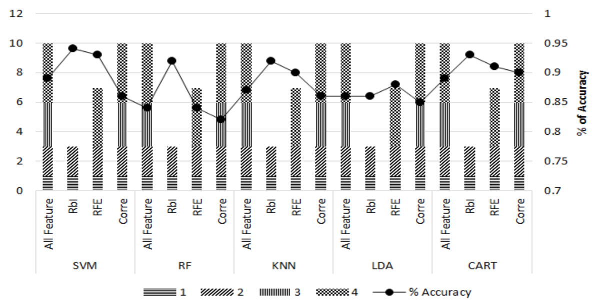
Method	Average Accuracy
RbI-SVM	0.91 (1)
RbI-RF	0.90 (2)
RbI-LDA	0.86 (5)
RbI-KNN	0.90 (2)
RbI-CART	0.88 (4)
LSID-ANN[4]	0.81 (6)
J48 [5]	0.89 (3)
DeLeS [20]	0.79 (7)



a) Processing Dimension



b) Perception Dimension



c) Input Dimension

Fig. 3: % of Accuracy based on Attributes selected for respective Dimension

Even though, the increase of accuracy seems small as can be observed from a previous research by [5], the accuracy rose by only 0.02 compare to RbI-SVM. But this small improvement in the accuracy of learning style identification can make a significant difference for students. It means a more accurate identification of students learning style and accordingly, more accurate information, intervention and advice for students [4]. This will also lead to a better adaptively of the learning systems.

5. Conclusion

This paper used three different feature selection methods which are RbI, RFE and Corre to determine the most relevant number of attributes in predicting the learning style. Different feature selection methods are used to test further and to have an effective and more accurate result on the number of attributes selected. From there, it is found that RbI is the best method for deciding the most relevant attribute.

Next, to evaluate the effectiveness of the feature selection method classification model is used. In this paper five different classifier are used (i.e. SVM, RF, KNN, LDA and CART). The result of the classifier in terms of its accuracy value is calculated for each feature selection method. Increase in accuracy value will further help in predicting the learning style of users more accurately. This will further help in enhancing the learning system of online learning.

The results of the method were compared with existing approaches using the accuracy value, which is commonly used in research on identifying learning styles [4, 5, 20]. Based on the accuracy value, the best solution on the averaging value of the FSLSM dimension always come from the classifier after undergoing the feature selection method. By identifying students' learning styles with higher accuracy value, it can provide more accurate personalization for adaptive learning systems. In conclusion, number of attributes will affect the performance of classification model. However, it is not decided on the lesser number of attributes but instead on the optimal number of attributes which is relevant in increasing the performance of the classification model.

The possible future work is outlines as follows:-

1. To use embedded feature selection method and observe any changes in terms of number of attributes selected.
2. To further test with different classification methods by including a parameter changes and observing the increase of the classification accuracy.
3. To test the model in an online learning system and prove the effectiveness in a real case study.

Acknowledgement

The authors wish to thank the Universiti Sains Malaysia (USM) for the support it has extended in the completion of the present research through the Fundamental Research Grant Scheme (203/PKOMP/6711463). The authors also wish to thanks Asst. Prof. Dr. Renato Racelis Maaliw III from Southern Luzon State University, Philippines for dataset sharing.

References

[1] R. M. Felder and L. K. Silverman, "Learning and teaching styles in engineering education", *Engineering education*, Vol. 78, no. 7, (1988) pp. 674–681.

[2] H. M. Truong, "Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities", *Computers in Human Behavior*, vol. 55, (2016), pp. 1185–1193.

[3] R. Mokhtar, N. A. M. Zin, and S. N. H. S. Abdullah, "Rule-based knowledge representation for modality learning style in aiwbes", *Knowledge Management International Conference*, 2010.

[4] J. Bernard, T.-W. Chang, E. Popescu, and S. Graf, "Learning style identifier: Improving the precision of learning style identification through computational intelligence algorithms", *Expert Systems with Applications*, vol. 75, (2017), pp. 94–108.

[5] R. R. Maaliw III, "Classification of learning styles in virtual learning environment using data mining: A basis for adaptive course design," 2016.

[6] Y. Kim, W. N. Street, and F. Menczer, "Feature selection in data mining", *Data mining: opportunities and challenges*, vol. 3, no. 9, (2003),pp. 80–105.

[7] J. Miao and L. Niu, "A survey on feature selection", *Procedia Computer Science*, vol. 91, (2016), pp. 919–926.

[8] G. T. Hilda and R. Rajalaxmi, "Effective feature selection for supervised learning using genetic algorithm", *in Electronics and Communication Systems (ICECS)*, (2015), pp. 909–914.

[9] T. A. Litzinger, S. H. Lee, and J. C. Wise, "A study of the reliability and validity of the felder-soloman index of learning styles," 2005.

A. D. Benson, "Using online learning to meet workforce demand:A case study of stakeholder influence", *Quarterly review of distance education*, vol. 3, no. 4, (2002), pp. 443–452.

[10] J. L. Moore, C. Dickson-Deane, and K. Galyen, "e-learning, online learning, and distance learning environments: Are they the same?", *The Internet and Higher Education*, vol. 14, no. 2, (2011), pp. 129–135.

[11] D.Conrad, "Deep in the hearts of learners: Insights into the nature of online community," *International Journal of E-Learning & Distance Education*, vol. 17, no. 1, (2002), pp. 1–19.

[12] S. Gao, G. Ver Steeg, and A. Galstyan, "Variational information maximization for feature selection", *Advances in Neural Information Processing Systems*, (2016), pp. 487–495.

[13] N. S´anchez-Maro˜no, A. Alonso-Betanzos, and M. Tombilla-Sanrom´an, "Filter methods for feature selection–a comparative study," in International Conference on Intelligent Data Engineering and Automated Learning. Springer, 2007, pp. 178–187.

[14] D. J. K. L. Selvapriya M, "Data mining in education", *International Journal of Multidisciplinary Research and Development*, vol. 3, no. 3, (2016), pp. 43–48.

[15] B. Pradhan, "A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy mod- els in landslide susceptibility mapping using gis", *Computers & Geosciences*, vol. 51, (2013), pp. 350–365.

[16] H. Parvin, H. Alizadeh, and B. Minaei-Bidgoli, "Mknn: Modified knearest neighbor", *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1. Citeseer, 2008.

[17] L.Breiman, "Random forests,"*Machine lerning*, vol. 45, no. 1, (2001), pp. 5-32.

[18] J. –J. Lo, Y.-C. Chan, and S.-W. Yeh, "Designing an adaptive web-based learning system based on students coginitive styles identified online", *Computers & Education*, vol. 58, no.1, (2012),pp.209-222.

[19] L. T.-C. Graf, Sabine et al., "Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach", Journal of Educational Technology Society, vol.12, no. 4, (2009), p.3.