Hybrid Fuzzy Rule-Based Classification System for MOODLE LMS System

Qun Zhao[1], Shou-Chuan Lai[2], Jin-Long Wang[2], Li-Yu Wang[3]
[1] College of Science and Technology, Ningbo University, China
[2] Department of Information and Telecommunications Engineering, Ming Chuan University, Taiwan
[3] Division of Student Affairs, Ming Chuan University, Taiwan
zhaoqun@nbu.edu.cn, sclai@mail.mcu.edu.tw, jlwang@mail.mcu.edu.tw, lilyuwang@mail.mcu.edu.tw

Abstract

Educational data are widely applied to predict students' academic performance in educational systems. Prior research mainly used students' past learning data to predict their future performance. However, these predicted results could not provide teachers with the opportunity to remediate the students in time.

In order to achieve the effect of early warning, this study uses only the activity log of the first third of the semester to build models and prediction results. A hybrid classification decision mechanism is proposed to combine the results of different predictions based on the accumulated training cases to further improve the accuracy of prediction.

The proposed system is then applied to discover students' learning outcomes in a C programming language course in the early stage of a semester according to the log files of the MOODLE LMS system.

The results show that the transformation of learning activity data has a critical impact on prediction accuracy. Using cumulative training cases can significantly improve prediction accuracy. And the proposed hybrid fuzzy classification decision-making scheme, which combines data conversion with cumulative training cases, can produce higher prediction accuracy by using just one-third of a semester's learning activity data.

Keywords: MOODLE, Early warning, Fuzzy rule-based classification system

1 Introduction

Moodle is one of the most popular learning management systems, which can support many teaching and learning activities to facilitate information sharing and communication between course participants. Teachers can upload their lecture materials, arrange assignments and online exams, set up discussion fora, and deliver messages. If most students at an institution use this system as their major learning platform, MOODLE can store all the learning patterns of students in the log file. The accumulated data of learning activities is a very valuable treasure to which educational data mining (EDM) can be applied for further investigation [1]. The results of EDM can predict student performance in order to deepen understanding of the main factors that influence their performance in school [2].

Previous studies utilized students' learning data for the same semester, with part of the data as a training case and the remainder as a testing case to verify the predictive validity of the data mining algorithm. Such a verification method must be carried out at the end of the semester, so it can only be used to verify the validity of the algorithm, and the teacher cannot have the opportunity to counsel any student with learning difficulties accordingly. Therefore, this study attempts to use the learning pattern of a previous semester of known learning outcomes to predict student outcomes for the same course in a later semester. In order to provide teachers enough time to tutor students, the learning activity data of the first 6 weeks of previous semesters is used as a training case to build a predictive model and, in a similar manner, the first 6 weeks of later semesters is used as the testing case. In this way, the predicted results may provide the teacher enough time to offer remedial learning opportunities to the students with learning difficulties.

Although the teacher provides the same course during different semesters using the same pedagogy and arranging similar teaching activities, the students belong to two different independent groups. The distribution of learning activities as displayed by the students from these two classes will be different. Therefore, five transformation methods are used to convert the log data from different semesters to be more consistent with each other. The converted log data can then be used to establish a prediction model utilizing the fuzzy rule-based classification system (FRBCS).

Fuzzy rule-based classification systems can typically be classified into five categories according to different implementation methods, including space partition [3], genetic algorithms [4], clustering [5], neural networks

*Corresponding Author: Shou-Chuan Lai; E-mail: sclai@mail.mcu.edu.tw
DOI: 10.3966/160792642021012201008

[6], and gradient descent [7]. FRBS based on a space partition approach is one of most useful and effective FRBCSs, in which the learning method first utilizes the strategy of dividing the variable space, and then the divided partitions are used to obtain the parameters of the membership functions.

In order to improve the prediction accuracy, this study uses the weighted results of the matching data from the query data and the fuzzy rules of each class to determine the final class. That is to say, the prediction part of the proposed classification algorithm first evaluates the degree of match between the query data and each fuzzy rule. For each fuzzy rule of class, the sum of the product of the degree of match and the fuzzy rule weighting value is calculated. Then, among the sum values of all classes, the class with the largest total value will be used to classify the class of the query data [8]. Besides, this study applies the data obtained by the five data transformation methods to the weighted class decision. The hybrid decision method is proposed to calculate a weighted result from five weighted results of each test data for comprehensively determining the final class of test data.

This study proposes two novel methods to predict student learning outcomes, including the accumulative FRBCS method (A-FRBCS) and the Hybrid Cumulative FRBCS Method (HA-FRBCS). The A-FRBCS method utilizes the accumulation of activity logs over several semesters to build a predictive model, which then is employed to predict the learning outcomes for the latest semester. The HA-FRBCS synthesizes several forecast results of the A-FRBCS in advance from different data transformation methods to accurately predict the learning outcomes of the most recent semester. In predictive analysis, the performance parameters based on a confusion matrix are used to present the prediction metrics such as precision rate, recall rate, and F-measure.

The rest of this paper is organized as follows. In Section 2, the basics of the previous research methodology are introduced. The proposed method is presented in Section 3. Section 4 illustrates the results of performance evaluation of the proposed method compared with other methods. At last, Section 5 concludes the paper.

2 Research Methodology

Many modern universities use learning management systems in teaching and learning to promote a blended learning model that combines traditional classroom and online learning. MOODLE system is one of the most popular learning management systems used in higher education worldwide, as it provides a learner-centric learning environment [9]. In order to successfully carry out educational data mining, it is required first to collect the various materials related to students' academic performance and learning style. The main sources of learning-related data are instructive software and school management tools [10]. For example, the MOODLE platform is the most popular free software of this kind.

EDM typically uses the student learning activities documented in the database on the instructional platform to create a more complete picture of student learning activities. Moreover, EDM can provide teachers with some teaching strategies by establishing models and making predictions to improve students' learning outcomes to reduce the failure rate at the end of each school year. For example, EDM can be used to discover successful patterns and phenomena about student learning processes. Using the models developed by EDM, the effectiveness of teaching can be validated and evaluated in order to enhance the quality of education [11]. These results can provide practical applications for teachers and students. For example, teachers can use effective models revealed by EDM to prepare their teaching materials. Students can also get feedback on the learning process through EDM [12].

The most commonly used data mining methods are regression, clustering, classification and association rule mining [13], in which the classification method can systematically classify the category data, so it is very suitable for predicting students' learning outcomes. This study investigates student behavior patterns in four steps, comprising data retrieval, data transformation, data processing, and results reporting [14].

2.1 Data Transformation

In decision-making with multiple indicators, each indicator has different characteristics and value ranges. Therefore, data transformation must be performed before data processing to ensure the reliability of multiple index decisions. There are five commonly used data transformation methods, including min-max transformation, zero-mean transformation, square root transformation, ranking transformation, and discrete transformation [9].

(1) Min-max transformation

Min-max transformation maps the minimum value and the maximum value to 0 and 1, respectively. And, other values are mapped according the following formula:

$$y_i = f_{\min-\max}(x_i) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \tag{1}$$

where x_i is the original data, y_i is the normalized data, $x_{\max}$ is the maximum value and $x_{\min}$ is the minimum value in the sample data [15].

(2) Discrete transformation

Discrete transformation may divide the original data into n divisions. If one piece of data is located in the k-th division, the return value will be k.

$$y = f_{discrete}(x_i) = k, \tag{2}$$

where x_i is the original data, y is the normalized data, and f() is the discrete conversion function [16].

(3) Zero-mean transformation

Zero-mean transformation can transfer an original value to a scale that indicates the distance of the original value and the mean value in terms of standard deviation [17].

$$y = f_{zero-mean}(x_i) = \frac{x_i - \mu}{\sigma}, \tag{3}$$

where x_i is the original data, y_i is the transformed data, μ is the mean and σ is the standard deviation of all sample data.

(4) Ranking transformation

Ranking transformation first sorts the sample data, and then converts the minimum value to 1, the second minimum value to 2, and so on.

$$y_i = f_{ranking}(x_i) = k, \tag{4}$$

where y_i is the transformed data, x_i is the original value, which is the k-th minimum value of the sample data [18].

(5) Square root transformation

Square root transformation can convert the original data x_i to the value of $\sqrt{x_i}$.

$$y_i = f_{square-root}(x_i) = \sqrt{x_i}, \tag{5}$$

where x_i is the original data, and y_i is the transformed data [19].

2.2 Classification in Data Mining

Classification aims to build a model based on the training set and use the built model to identify the category of new observations. Depending on the task, the classification accuracy will vary. The popular classification methods are the C5.0 decision tree, the Support Vector Machines (SVM) and the Naive Bayes Classifier.

The decision tree is a machine learning modeling technique mainly used in regression and classification problems. The learning strategy of the decision tree is based on the principle of minimizing the loss function. After the decision tree is generated, the decision tree can be pruned to prevent the decision tree from being too complicated. SVM is a linear model used to perform classification and regression problems. The SVM algorithm aims to find a hyperplane in the N-dimensional space to classify a data group. Naive Bayes classifier is a simple probabilistic classifier, which is especially suitable for high input dimensions. The key to the naive Bayes classifier is based on Bayes' theorem, which directly assumes that all random variables are conditionally independent [20].

2.3 Fuzzy Classification

The Mamdani model, a typical fuzzy decision method with multiple inputs and single output (MISO), is composed of four parts, including the fuzzifier, the knowledge base, the inference engine and the defuzzifier [21-22]. The fuzzy rule classification system, based on space partition (FRCS-SP), uses the strategy of variable space partitioning in the learning method, which is employed to generate the parameters of the membership function. There are two main FRCS-SP algorithms, including FRBCS using Chi's method and FRBCS using Ishibuchi's method with weight factor [22]. The FRBCS using Chi's method (FRBCS-C) is based on Wang and Mendel's method [23], in which the fuzzy rules are derived from the techniques of Wang and Mendel's technique, but the consequent parts are replaced with the class of data. Besides, the degree of each rule is determined by the antecedent part of the rule [23]. FRBCS using the Ishibuchi method with weight factor (FRBCS-W) has certainty weights in the consequent parts of the fuzzy IF-THEN rule. The grid-type fuzzy partitioning method is used to determine linguistic values and membership functions of antecedent parts, while the consequent class is defined as the dominant class of the antecedent parts in the fuzzy space [24]. In the prediction stage, both FRBCS-C and FRBCS-W algorithms calculate the degree of match of the fuzzy rule for each test data, and then the result class of the fuzzy rule with the largest degree of match will be the result class of the corresponding test data. This way of 'winner takes all' does not represent the best solution in many applications.

The operation of FRBCS can be divided into two phases, including the learning phase and the prediction phase. In the learning phase, structural identification and parameter evaluation are generated using the training data [25], whereby fuzzy rules and membership functions are automatically generated and stored in the database and the rule base of the knowledge bank. In the prediction phase, the testing data is first converted into fuzzy data. Then, based on fuzzy rules and membership functions in the knowledge bank, the fuzzy data is inferred into prediction data. Finally, the predicted fuzzy data is transferred to crisp data through the defuzzification process.

3 The Proposed Method

This study proposes two methods to precisely predict student learning outcomes, including the accumulative FRBCS method (A-FRBCS) and the Hybrid Cumulative FRBCS Method (HA-FRBCS). The A-FRBCS method utilizes the accumulation of activity logs over several semesters to build a predictive model, which then is employed to predict the learning outcomes of a later semester. The HA-FRBCS synthesizes several forecast results of the A-FRBCS from different periods to more accurately predict the learning outcomes for a later semester. The structure of proposed data mining processes, including the data retrieval, the data transformation, the data modeling, and the predictive analysis, is as shown in Figure 1. In the data retrieval phase, the log data of learning activities is extracted from the database of MOODLE system. In the data transformation phase, either the training case or the testing case is normalized via five methods, including min-max, discrete, ranking, Z-mean, and square root (SQRT). The converted training case is used for creating the predictive model, which contains related parameters in the membership functions and IF-THEN decision rules. The converted testing case is first fuzzified and then is used to generate the fuzzy result value according to the decision rules of the predictive model. The fuzzy result values are then processed through defuzzification to become crisp result values, which are then used to determine the final decision result value via the hybrid strategy.

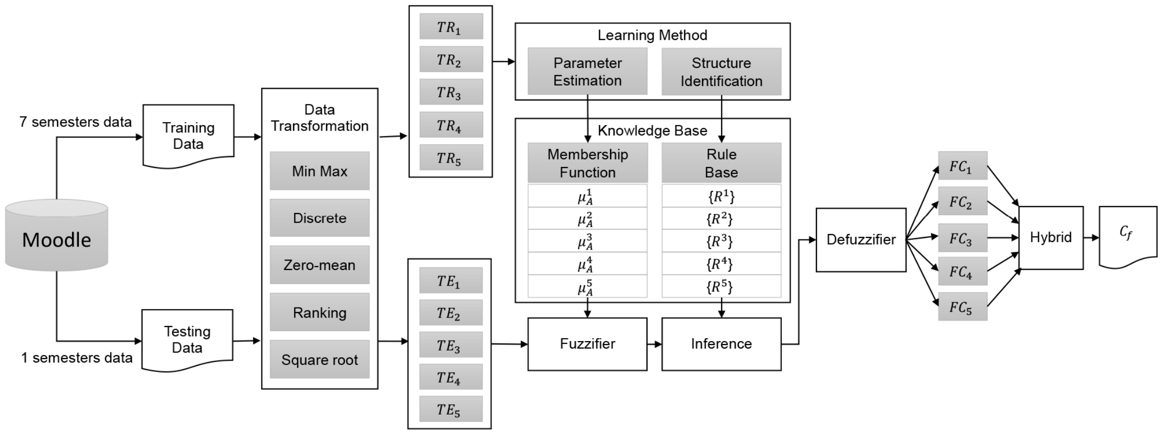


Figure 1. The proposed Hybrid FRBCS system

The A-FRBCS method consists of 4 phases, including (1) retrieval of student learning activity data from the log file in the MOODLE database, (2) accumulation, cleaning and alignment of data, (3) establishment of prediction model, and (4) establishment of a confusion matrix [9]. Of the 8 semesters of activity data selected for this study, the training cases are obtained from the data of the first seven semesters, and the test cases are acquired from the eighth semester. In the data cleaning phase, data with an average activity below the threshold is first removed. The training case is aligned with the testing case, so that the fields of both cases are consistent with each other. The training case is used to build the prediction model by using the fuzzy classification method, "FBCS.CHI" of the R frbs package. Then, the predictive results are produced by applying the testing case to the prediction model. The prediction can produce two kind of outcomes, including pass and fail. It is assumed that the corresponding value of pass is set to be 2(C_2) and the corresponding value of fail is set to be 1(C_1). Finally, the confusion matrix is constructed to illustrate the performance accuracy of prediction metrics.

1. #Data retrieval
   con = dbConnect (MySQL(), user = 'admin', password = 'xxxxx', dbname = 'MOODLE', host = 'localhost')
   data[i] = dbReadTable (con, mdl_log[i])[,c(1:8)], i = 1..8
2. #Data clean and alignment
   d1<-cleanData (data [i_1:j_2]), 1≤i_1, i_2≤7 # training data
   d2<-cleanData (data [j]), j = 8 # testing data
   sem1<-alignColumn (d1, d2)
   sem2<-alignColumn (d2, sem1)
   real<-sem2[,8]
3. #FRBSC Model establishment and Prediction
   frbcs<- frbs.learn (d1[i], .., method.type = "FRBCS.CHI", ..)
   pred<-predict (frbcs, testCase[i]), i = 1..5
4. # Establishment of a confusion matrix.
   conMat<-table (pred, real)

The HA-FRBCS method consists of 6 phases, including: (1) retrieval of learning activity data from the MOODLE database, (2) accumulation, cleaning and alignment of data, (3) data transformation, (4) establishment of prediction model, (5) hybrid classification, and (6) establishment of a confusion matrix.

1. #Data retrieval
   con = dbConnect (MySQL(), user = 'admin', password = 'xxxxx', dbname = 'MOODLE', host = 'localhost')
   data[i] = dbReadTable (con, mdl_log[i])[,c(1:8)], i = 1..8

2. #Data clean and alignment
   d1<-cleanData (data[i1:j2]), 1≤i1, i2≤7 # training data
   d2<-cleanData (data[j]), j = 8 # testing data
   sem1<-alignColumn (d1, d2)
   sem2<-alignColumn (d2, sem1)
   real<-sem2[,8]
3. #Data transformation:
   dTran<-c (rank, sqrt, minmax, zmean, discrete)
   trainCase[i]<-dTran[i](sem1)
   testCase[i]<-dTran[i](sem2), i = 1..5
4. #FRBS Model establishment and Prediction
   frbcs[i]<- frbs.learn (trainCase[i], .., method.type = "FRBCS.CHI", ..), i = 1..5
   pred[i]<-predict (frbcs[i], testCase[i]), i = 1..5
5. #Classification
   pred.HYBRID<-HYBRID (pred[i]), i = 1..5
6. threshold<-r·(C1+ C2)/2, r = 5, C1 = 1, C2 = 2
   pred<-ifelse (pred.HYBRID<threshold, C1, C2)
   conMat<-table (pred, real)

In the proposed HA-FRBCS method, the first two steps are the same as the A-FRBCS method. In data transformation, five data transformation schemes can be used to transform both training case and the testing case. Therefore, five training cases and five testing cases can be generated after the data transformation. Next, by using the fuzzy classification method, "FBCS.CHI" of the R frbs software package, one prediction model can be constructed with each training case. Following, the testing case is applied to the prediction model generated by the same transformation scheme to produce the prediction result. Then, a hybrid result can be obtained from the five prediction results according to equation (10). Finally, predictions can produce two kinds of outcomes: pass and fail. It is assumed that the corresponding value of the pass and the fail are set to 2(C_2) and 1(C1), respectively. The corresponding values of the two predictions are generated using equation (11).

Let $R = \{R_i, i = 1, ..., M\}$ be the set of M constructed fuzzy rules. The fuzzy rule R_i in the fuzzy rule base for a binary-class pattern classification problem with K features can be defined as:

R_i : If P is A_i , then the consequence is S_i with rule weight W_i , i = 1, 2, …, M,

where $P = \{p_1, p_2, ..., p_k\}$ is the linguistic variables and $A_i = (A_i, 1, ..., A_{i,k})$ is the linguistic values, $S_i \in C \triangleq \{C_1, C_2\}$ is an integer variable whose value denotes the label of the consequent class, M is the number of fuzzy rules, and W_i is the rule weight of the fuzzy rule R_i .

A query pattern $Q = (q_1, q_2, ..., q_k)$ is classified by the degree of correspondence $D_i, i = 1, ..., M,$ between the input Q and the i-th rule are

$$D_i = \mu_{A_i}(Q) \cdot W_i \quad \textbf{(6)}$$

which has the product of the matching degree $\mu_{A_i}(Q)$ and the rule weight W_i in R. The matching degree $\mu_{A_i}(Q)$ is defined as

$$\mu_{A_i}(Q) = \prod_{j=1}^{K} \mu_{A_{i,j}}(q_j), \quad \textbf{(7)}$$

where $\mu_{A_{i,j}}(\cdot)$ indicates the membership function of the fuzzy set $A_{i,j}$. Then, the maximum degree of correspondence $D_{\max}$ is

$$D_{\max} = \max_{i=1...M} D_i. \quad \textbf{(8)}$$

The resulting class C_j of j-th query pattern is the consequence value with maximum degree of correspondence $D_{\max}$, that is

$$C_j = S_j, \text{ where } \mu_{A_j}(Q) \cdot W_j = D_{\max}. \quad \textbf{(9)}$$

r consequence classes, denoted by $C_{j,u}, u = 1...r$, are produced by r data transformation methods for one query data, the combined resulting class C_w can be calculated:

$$C_w = \sum_{u=1}^{r} C_{j,u}. \quad \textbf{(10)}$$

The final consequence class C_f can be determined by checking if C_w is close to C_1 or C_2 for a binary-class pattern classification problem.

$$C_f = \begin{cases} C_1, C_w \le \dfrac{r \cdot (C_1 + C_2)}{2} \\ C_2, C_w > \dfrac{r \cdot (C_1 + C_2)}{2} \end{cases} \quad \textbf{(11)}$$

After the final consequence class is obtained, the confusion matrix is used to illustrate the prediction results of a binary-class pattern classification problem. True Positive (TP) and True Negative (TN) are used to represent the number of instances that are predicted truly as fail or pass, respectively. False Positive (FP) and False Negative (FN) are used to represent the number of instances that are predicted falsely as fail or pass, respectively.

Based on the TP, TN, FP, and FN, three performance indexes are introduced to describe the prediction performance, including precision, recall, and the F-measure. Precision is the percentage of predictions that are relevant, while recall is the percentage of total relevant instances that are predicted.

There are two classes, in which class-1 indicates student failure in this course, while class-2 indicates student success in this course. Pfail indicates the precision rate of class-1 which can be defined as the proportion of instances that truly belong to class-1(fail) divided by the total instances classified as class-1. Rfail indicates the recall rate of class-1, which can be defined as the proportion of instances classified as class-1 (fail) divided by the actual total in class-1. Ffail indicates the F-measure of the instances of fail. Ppass indicates the precision rate of class-2, which can be defined as the proportion of instances that truly belong to class-2 (pass) divided by the total instances classified as class-2. Rpass indicates the recall rate of class-2, which can be defined as the proportion of instances classified as class-2 (pass) divided by the actual total in class-2. Fpass indicates the F-measure of the instances of pass. Moreover, Acc indicates the classification accuracy of the overall performance [26].

4 Evaluation Results

This research uses the R system and its related packages to perform data retrieval, data transformation, data analysis, and results presentation. In order to achieve early predictions, this study only used student behavioral data for the first six weeks of each semester. The learning behavior data of students in each of the earlier semesters is used as a training case to build a predictive model, and the learning behavior data of students in the latest semester is used as a testing case to verify the accuracy of the prediction model. In data transformation, five methods are used to normalize the data, including rank, square root (SQRT), Z-score, min-max, and discrete. Training cases and testing cases are normalized by the same data transformation method to facilitate the establishment of the model and the prediction of student learning outcomes. In predictive analysis, based on the predictive model built from the activity logs of previous semesters, activity logs from the latest semester are used to test the effectiveness of the algorithm. Therefore, the prediction model established by the log data of the first semester is used to predict the learning outcome of the eighth semester, and the prediction model of the second semester is used to predict the learning activities of the eighth semester, and so on. Besides, this study proposes two methods to precisely predict the student learning outcomes, including the A-FRBCS method and the HA-FRBCS method. Finally, the comparisons of prediction results from the FRBCS, the A-FRBCS and the HA-FRBCS are given through the three indicators (accuracy, recall, and F1 measurement) for two kinds of results, pass and fail.

Figure 2 shows the performance accuracy of FRBCS, A-FRBCS and HA-FRBCS, through six performance metrics, including Pfail, Rfail, Ffail, Ppass, Rpass, and Fpass. From the median of each indicator, HA-FRBCS is indeed better than the other two methods, and A-FRBCS is superior to FRBCS. Besides, the performance distribution of FRBCS, A-FRBCS and HA-FRBCS are shown in the Ridgeline plot of Figure 3. The distribution of HA-FRBCS is located obviously at the right side of the Ridgeline plot, while the distribution of A-FRBCS is located in the middle and the distribution of FRBCS at the left side. This is because the performance of HA-FRBCS is the best among these three methods, the performance of A-FRBCS is second, followed by the performance of FRBCS.

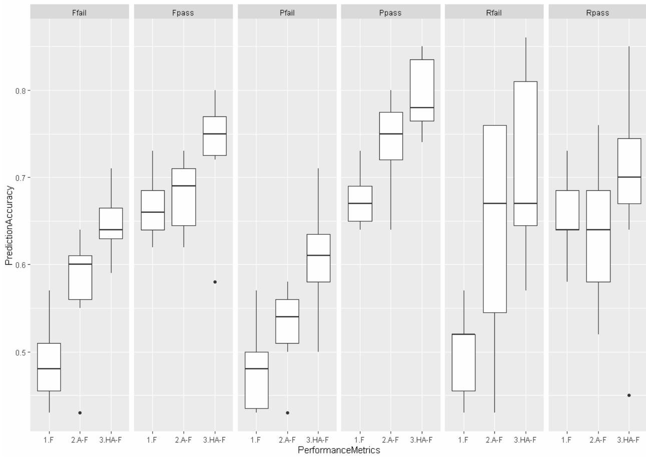


Figure 2. Comparison of prediction accuracy for FRBCS (1.F), A-FRBCS (2.A-F) and HA-FRBCS (3.HA-F)

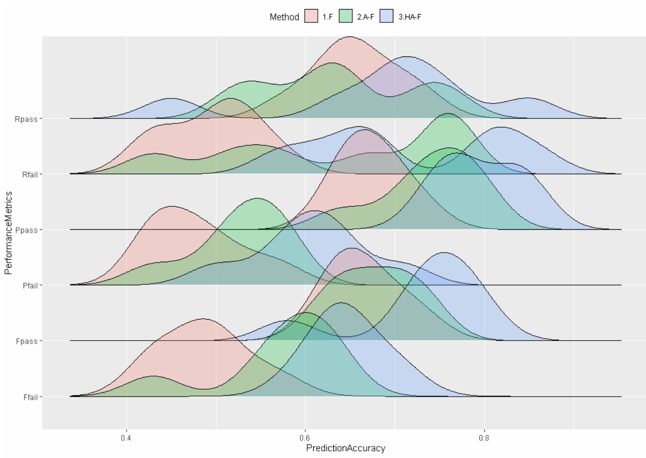


Figure 3. Distributions of prediction accuracy for FRBCS (1.F), A-FRBCS (2.A-F) and HA-FRBCS (3.HA-F)

Average prediction rates of precision rate of fail (Pfail), recall rate of fail (Rfail), F-measure of fail (Ffail), precision rate of pass (Ppass), recall rate of pass (Rpass), and F-measure of pass (Fpass) for the FRBCS, the A-FRBCS and the HA-FRBCS methods are illustrated in Figure 4. Among all the average prediction rates, HA-FRBCS is clearly better than FRBCS and A-FRBCS. The average prediction rate of FRBCS is about 58%, while A-FRBCS can increase the average prediction rate to 63%. Furthermore, HA-FRBCS improves the average prediction rate to 70% by synthesizing the prediction results of A-FRBCS.

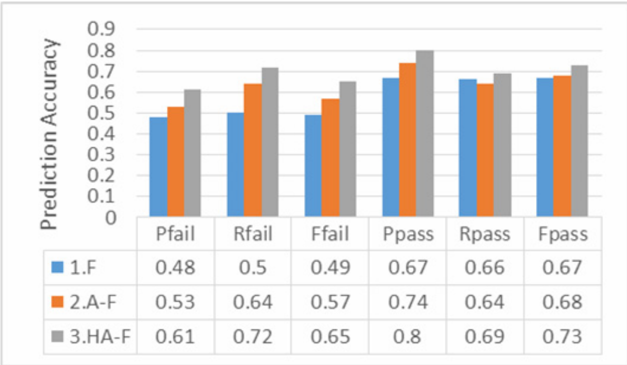


Figure 4. Comparison of average prediction rates among FRBCS (1.F), A-FRBCS (2.A-F) and HA-FRBCS (3.HA-F)

Table 1 illustrates the prediction efficiency and the performance rank of FRBCS, A-FRBCS, and HA-FRBCS. Student activity data for the first seven of eight semesters is used to build a predictive model, while student activity data for the eighth semester is used for predictions. The {X}-> Y equation in the first column of Table 1 shows that the student activity data for the Xth semester is used as a training case to predict the testing cases of the Yth semester. The equation {X1-X2}-> Y in the fifth column indicates that the learning activities from semester X1 to semester X2 are used as training cases to predict the test cases of semester Y. The data values in columns of 3, 7, and 9 represent the prediction efficiency, the values in columns of 4, 8, and 10 indicate their performance ranks in the same row, and the last row presents the average performance rank and the average performance accuracy obtained by various methods. The average performance rank of FRBCS, A-FRBCS, and HA-FRBCS are 2.6, 1.95, and 1.17, which indicates that HA-FRBCS is the top-ranked method among them, followed by A-FRBCS, and finally FRBCS.

Table 1. Comparison of prediction accuracy and performance ranking for different methods

Prediction	Performance Metrics	FRBCS	Rank (FRBCS)	Prediction	Performance Metrics	A-FRBCS	RANK (A-FRBCS)	HA FRBCS	RANK (HA-FRBCS)
{1}->8	Pfail	0.48	3	{1-7}->8	Pfail	0.52	2	0.63	1
{1}->8	Rfail	0.52	3	{1-7}->8	Rfail	0.76	2	0.81	1
{1}->8	Pfail	0.5	3	{1-7}->8	Pfail	0.62	2	0.71	1
{1}->8	Ppass	0.68	3	{1-7}->8	Ppass	0.78	2	0.85	1
{1}->8	Rpass	0.64	2	{1-7}->8	Rpass	0.55	3	0.7	1
{1}->8	Fpass	0.66	2	{1-7}->8	Fpass	0.65	3	0.77	1
{2}->8	Pfail	0.48	3	{2-7}->8	Pfail	0.55	2	0.61	1
{2}->8	Rfail	0.48	3	{2-7}->8	Rfail	0.76	1	0.67	2
{2}->8	Pfail	0.48	3	{2-7}->8	Pfail	0.64	1	0.64	1
{2}->8	Ppass	0.67	3	{2-7}->8	Ppass	0.8	1	0.77	2
{2}->8	Rpass	0.67	2	{2-7}->8	Rpass	0.61	3	0.73	1
{2}->8	Fpass	0.67	3	{3-7}->8	Fpass	0.69	2	0.75	1
{3}->8	Pfail	0.44	3	{3-7}->8	Pfail	0.58	2	0.64	1
{3}->8	Rfail	0.52	2	{3-7}->8	Rfail	0.52	2	0.67	1
{3}->8	Pfail	0.48	3	{3-7}->8	Pfail	0.55	2	0.65	1
{3}->8	Ppass	0.66	3	{3-7}->8	Ppass	0.71	1	0.78	1
{3}->8	Rpass	0.58	3	{3-7}->8	Rpass	0.76	2	0.76	1
{3}->8	Fpass	0.62	3	{3-7}->8	Fpass	0.73	2	0.77	1
{4}->8	Pfail	0.43	3	{4-7}->8	Pfail	0.54	2	0.59	1
{4}->8	Rfail	0.43	3	{4-7}->8	Rfail	0.67	2	0.81	1
{4}->8	Pfail	0.43	3	{4-7}->8	Pfail	0.6	2	0.68	1
{4}->8	Ppass	0.64	3	{4-7}->8	Ppass	0.75	2	0.84	1
{4}->8	Rpass	0.64	2	{4-7}->8	Rpass	0.64	2	0.64	1
{4}->8	Fpass	0.64	3	{4-7}->8	Fpass	0.69	2	0.73	1
{5}->8	Pfail	0.52	3	{5-7}->8	Pfail	0.57	1	0.71	1
{5}->8	Rfail	0.52	3	{5-7}->8	Rfail	0.57	2	0.57	1
{5}->8	Pfail	0.52	3	{5-7}->8	Pfail	0.57	2	0.63	1
{5}->8	Ppass	0.7	3	{5-7}->8	Ppass	0.73	2	0.76	1
{5}->8	Rpass	0.7	3	{5-7}->8	Rpass	0.73	2	0.85	1
{5}->8	Fpass	0.7	3	{5-7}->8	Fpass	0.73	2	0.8	1
{6}->8	Pfail	0.57	1	{6-7}->8	Pfail	0.5	2	0.5	2
{6}->8	Rfail	0.57	3	{6-7}->8	Rfail	0.76	2	0.86	1
{6}->8	Pfail	0.57	3	{6-7}->8	Pfail	0.6	2	0.63	1
{6}->8	Ppass	0.73	3	{6-7}->8	Ppass	0.77	2	0.83	1
{6}->8	Rpass	0.73	1	{6-7}->8	Rpass	0.52	2	0.45	3
{6}->8	Fpass	0.73	1	{6-7}->8	Fpass	0.62	2	0.58	3
{7}->8	Pfail	0.43	2	{7-7}->8	Pfail	0.43	2	0.57	1
{7}->8	Rfail	0.43	2	{7-7}->8	Rfail	0.43	2	0.62	1
{7}->8	Pfail	0.43	2	{7-7}->8	Pfail	0.43	2	0.59	1
{7}->8	Ppass	0.64	2	{7-7}->8	Ppass	0.64	2	0.74	1
{7}->8	Rpass	0.64	2	{7-7}->8	Rpass	0.64	2	0.7	1
{7}->8	Fpass	0.64	2	{7-7}->8	Fpass	0.64	2	0.72	1
Average		0.58	2.6			0.63	1.95	0.7	1.17

The number of wins can be employed to compare the performance of FRBCS, A-FRBCS and HA-FRBCS methods by using six performance parameters, including Pfail, Rfail, Ffail, Ppass, Rpass, and Fpass, as shown in Figure 4. For each performance parameter, there are totally 7 cases from the combination of different training cases. HA-FRBCS is superior to FRBCS, where the winning rate of HA-FRBCS is from 79% to 86% for different performance parameters, while A-FRBCS has a winning rate from 0 to 21% and FRBCS from 0 to 14%.

Since the samples compared are from two different student groups in different school years, in order to achieve generalization, we use five transformation methods to standardize the student learning activities, and then comprehensively predict the results in a hybrid manner. It is known from the experimental results that the prediction result can be improved by the hybrid method. As shown in Figure 5, the proposed HA-FRBCS method is more efficient than the original FRBCS on all performance metrics. As shown in Table 1, the proposed HA-FRBCS method is more effective than the original FRBCS method and can significantly improve the prediction performance.

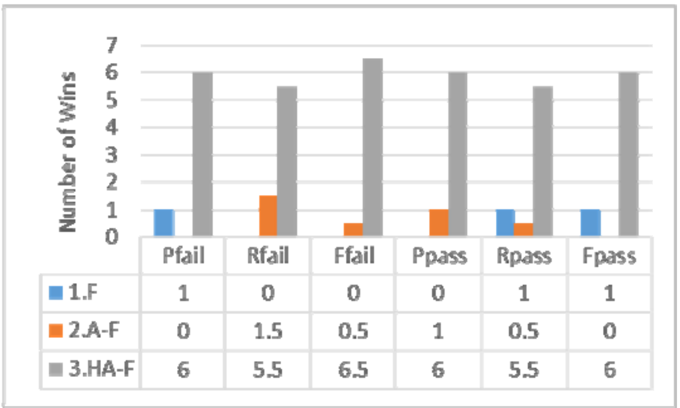


Figure 5. Comparison of number of wins in prediction among FRBCS (1.F), A-FRBCS (2.A-F) and HA-FRBCS (3.HA-F)

Moreover, the proposed hybrid FRBCS is compared with other classification methods, including Naïve Bayes (denoted as NB), SVM, and Decision tree (denoted as TREE). The accuracy (Acc) index among these methods is illustrated in Figure 6. According to the descending order of effectiveness, these four methods are Hybrid, SVM, NB, and TREE. Thus, the proposed Hybrid FRBCS is better than these other methods, as shown by higher accuracy. Classification accuracy is the overall performance, defined as the number of correct predictions divided by the total number of predictions. Moreover, Figure 7 demonstrates the more detailed performance index among the four methods, including Pfail, Rfail, Ffail, Ppass, Rpass, Fpass, and Acc. Clearly, the proposed hybrid FRBCS is superior to other methods in the index of Pfail, Ffail, Rpass, Fpass, and Acc. The NB method performs better in the Rfail index, while the SVM method performs better than other methods in the Ppass indexes. However, because the proposed hybrid FRBCS performs better in 5 out of 7 indexes, it has the best performance. In particular, the proposed hybrid FRBCS is higher than other methods in the Acc index, representing the overall performance.

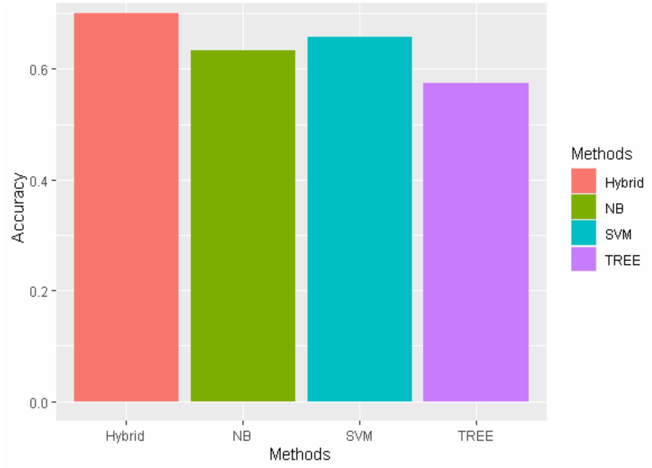


Figure 6. Accuracy among the methods of Hybrid FRBCS, Naïve Bayes (NB), SVM, and Decision tree

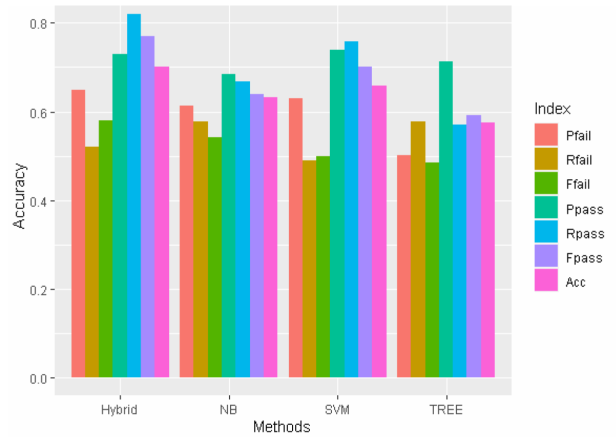


Figure 7. Index among the methods of Hybrid FRBCS, Naïve Bayes (NB), SVM, and Decision tree (TREE)

5 Conclusion

This study systematically predicts students who might fail in a C programming course based on student activity log files from eight classes over four academic years. The study found that the transformation of learning activity data is an essential step, and the hybrid FRBCS has a significant influence on the prediction accuracy because the hybrid FRBCS method has the characteristics of high stability and high prediction efficiency. The training cases and test cases, after data transformation, tend to be consistent. The results show that the proposed Hybrid FRBCS prediction data's accuracy is significantly higher than that of the original FRBCS. Compared to Naïve Bayes, SVM, and Decision tree methods, the proposed hybrid FRBCS also has the best performance. the results of this prediction can generate early warnings so teachers can identify students with poor academic performance at the beginning of the semester, and then actively conduct educational counseling to help students achieve their learning goals by the end of the semester.

References

[1] R. S. Baker and K. Yacef, The state of educational data mining in 2009: A review and future visions, *Journal of Educational Data Mining*, Vol. 1, No. 1, pp. 3-17, October, 2009.

[2] R. R. Kabra and R. S. Bichkar, Performance prediction of engineering students using decision trees, *International Journal of Computer Applications*, Vol. 36, No. 11, pp. 8-12, December, 2011.

[3] L. S. Riza, C. Bergmeir, F. Herrera and J. M. Benitez, frbs: Fuzzy Rule-Based Systems for Classification and Regression in R, *Journal of Statistical Software*, Vol. 65, No. 6, pp.1-30, May, 2015.

[4] O. Cordon, F. Herrera, F. Hoffmann and L. Magdalena, Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases, *Advances in Fuzzy Systems - Applications and Theory*, Vol. 19, World Scientific, July, 2001.

[5] N. K. Kasabov and Q. Song, DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction, *IEEE Transactions on Fuzzy Systems*, Vol. 10, No. 2, pp. 144-154, April, 2002.

[6] J. J. Buckley and Y. Hayashi, Fuzzy neural networks: A survey, *Fuzzy Sets and Systems*, Vol. 66, No. 1, pp. 1-13, August, 1994.

[7] H. Ishibuchi, H. Tanaka, and H. Okada, Interpolation of Fuzzy If-Then Rules by Neural Networks, *International Journal of Approximate Reasoning*, Vol. 10, No. 1, pp. 3-27, January, 1994.

[8] Q. Zhao, J.-L. Wang, T.-L. Pao and L.-Y. Wang, Modified Fuzzy Rule-Based Classification System for Early Warning of Student Learning, *Journal of Educational Technology Systems*, Vol. 48, No. 3, pp. 385-406, March, 2020.

[9] L.-Y. Wang, Fuzzy Rule-Based Classification Prediction Model for Student Learning Early Warning, *Master of Science Thesis*, Tatung University, Taipei, Taiwan, 2019.

[10] K. R. Koedinger, K. Cunningham, A. Skogsholm and B. Leber, An open repository and analysis tools for fine-grained, longitudinal learner data, *The 1st International Conference on Educational Data Mining*, Montreal, Québec, Canada, June 20-21, 2008, pp. 157-166.

[11] C. Romero, S. Ventura and P. D. Bra, Knowledge discovery with genetic programming for providing feedback to courseware authors, *User Modeling and User-Adapted Interaction*, Vol. 14, No. 5, pp. 425-464, January, 2004.

[12] A. Merceron and K. Yacef, Educational data mining: A case study, *The 12th International Conference on Artificial Intelligence in Education*, Amsterdam, Netherlands, July 18-22, 2005, pp. 467-474

[13] C. Romero and S. Ventura, Educational data mining: a review of the state of the art, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 40, No. 6, pp. 601-618, November, 2010.

[14] C. Romero, S. Ventura, and E. Garcia, Data mining in course management systems: MOODLE case study and tutorial, *Computers & Education*, Vol. 51, No. 1, pp. 368-384, August, 2008.

[15] C. Saranya and G. Manikandan, A Study on Normalization Techniques for Privacy Preserving Data Mining, *International Journal of Engineering and Technology*, Vol. 5, No. 3, pp. 2701-2704, June, 2013.

[16] T. Nagasawa, M. Sakamoto, and K. Takenaga, Supersymmetry and discrete transformations on S^1 with point singularities, *Physics Letters B*, Vol. 583, No. 3-4, pp. 357-363, March, 2004.

[17] T. Jayalakshmi and A. Santhakumaran, Statistical Normalization and Back Propagation for Classification, *International Journal of Computer Theory and Engineering*, Vol. 3, No. 1, pp. 89-93, February, 2011.

[18] E. W. Weisstein, Uniform Distribution, *MathWorld - A Wolfram Web Resource*, https://mathworld.wolfram.com/UniformDistribution.html (accessed March 13, 2019).

[19] T. G. Gregoire, Q. F. Lin, J. Boudreau, and R. F. Nelson, Regression Estimation Following the Square-Root Transformation of the Response, *Forest Science*, Vol. 54, No. 6, pp. 597-606, December, 2008.

[20] T. Pranckevicius and V. Marcinkevicius, Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification, *Baltic Journal of Modern Computing*, Vol. 5, No. 2, pp. 221-232, January, 2017.

[21] E. H. Mamdani, S. Assilian, An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, January, 1975.

[22] Z. Chi, H. Yan, and T. Pham, *Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition*, World Scientific Pub Co Inc: Singapore, 1996.

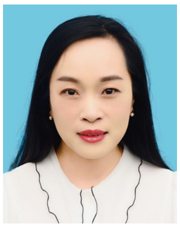
[23] H. Ishibuchi and T. Nakashima, Effect of rule weights in fuzzy rule-based classification Systems, *IEEE Transactions on Fuzzy Systems*, Vol. 9, No. 4, pp. 506-515, August, 2001.

[24] M. Sugeno and T. Yasukawa, A Fuzzy-Logic-Based Approach to Qualitative Modeling, *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 1, pp. 7-31, February, 1993.

[25] W. Pedrycz, *Fuzzy Modelling: Paradigms and Practice*, Springer Science & Business Media, 1996.

[26] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, An improved method to construct basic probability assignment based on the confusion matrix for classification problem, *Information Sciences*, Vol. 340-341, pp. 250-261, May, 2016.

Biographies



Qun Zhao is an Associate professor in the Department of Electronic Commerce at the College of Science and Technology, Ningbo University, Ningbo, China. Her research interests focus on E-commerce, Data Mining, FinTech, and Online Consumer Behaviour. Her major publications have been in the areas of e-business and management of IS, and her paper appeared in Telematics and Informatics, Sustainability and other journals.

Shou-Chuan Lai received the Ph.D. degree in computer science from the National Tsing Hua University, Hsinchu, Taiwan in 2000 and is an assistant professor in the Department of Information and Telecommunications Engineering at Ming Chuan University, Taoyuan, Taiwan. His research interests include computer networks, network security, and network management.

Jin-Long Wang is a Professor in the Department of Information and Telecommunications Engineering at Ming Chuan University, Taipei, Taiwan. He is also the vice president for academic affairs in Ming Chuan University. His current research areas include High-Speed Networks, Wireless Networks, Data Mining, Educational Technology and Electronic Commerce. Dr. Wang has about 30 years of experience in teaching and researching computers, networks, educational technology, and data mining fields.

Li-Yu Wang is a staff member in the Student Affairs Division at Ming Chuan University, Taipei, Taiwan. Ms. Wang has received master's degree from the Department of Computer and Science Engineering of Tatung University in 2019. His current research areas include Data Mining and Database Systems.