ASSOCIATION RULE DISCOVERY FOR STUDENT PERFORMANCE PREDICTION USING METAHEURISTIC ALGORITHMS

Roghayeh Saneifar and Mohammad Saniee Abadeh

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
r.saneifar@modares.ac.ir
saniee@modares.ac.ir

ABSTRACT

According to the increase of using data mining techniques in improving educational systems operations, Educational Data Mining has been introduced as a new and fast growing research area. Educational Data Mining aims to analyze data in educational environments in order to solve educational research problems. In this paper a new associative classification technique has been proposed to predict students final performance. Despite of several machine learning approaches such as ANNs, SVMs, etc. associative classifiers maintain interpretability along with high accuracy. In this research work, we have employed Honeybee Colony Optimization and Particle Swarm Optimization to extract association rule for student performance prediction as a multi-objective classification problem. Results indicate that the proposed swarm based algorithm outperforms well-known classification techniques on student performance prediction classification problem.

KEYWORDS

Educational data mining, bee colony optimization, continues rule extraction, classification, particle swarm optimization

1. INTRODUCTION

As the volume of archived data increases, the need for more efficient and faster data analysis techniques increases concurrently. All of the saved records in databases of organizations would be useless, if decision makers do not employ effective knowledge discovery techniques. Data mining methods analyze huge amount of databases to discover valuable and ready to use knowledge [8].

Nowadays, data mining techniques have been used in academic and educational environments and leave a remarkable effect in this domain [9]. Educational Data Mining (EDM) refers to the employment of knowledge discovery techniques and methods in education. The main goal of EDM is to enhance various educational activities such as student performance prediction, education facility improvement, etc.

As mentioned above, EDM is a domain that uses machine learning, data mining and statistical techniques, analyses educational data. Thanks to employ of these techniques, it is possible to improve the learning/teaching processes involving students or instructors.

Educational data come in many different and very complex formats. The last surveys in this scope is related to (Alejandro Pena-Ayala,2013), establishing the following EDM approaches [1]:

- Student behavior modeling
- Student performance modeling
- Student modeling
- Assessment
- Curriculum, domain knowledge, sequencing, and teachers support
- Student support and feedback

Other survey is related to Romero and Ventura [2], which is survey on educational data mining between 1995 and 2005. Using data mining techniques in higher education is a recent research domain; there are a lot of works in this area. That is because of its potentials to educational institutes.

Ayesha et al. employed the k-means data mining clustering algorithm to predict students' learning activities in an educational database including classroom quizzes, final and mid exam and other assignments. This correlated information will be conveyed to the teacher before the transfer of final exam. This study helps the teachers to improve the performance of students and reduce the failing ratio by taking appropriate steps at on time [3].

Baradwaj and Pal, in the year 2011, used the classification as data mining methods to evaluate student' performance, they applied decision tree technique for classification. The aim of their research is to extract knowledge that describes students' performance in end semester quizzes. They used students' educational data from the student' previous database including Class test , Assignment marks , Attendance, , Seminar. This study helps sooner in identifying the students who need more attention and allow the teacher to provide appropriate advising [4].

Chandra and Nandhini, applied the association rule mining method based on students courses to identifies students' break patterns. The aim of their research is to identify hidden relationship between the failed courses and suggests relevant causes of the failure to improve the low capacity students' performances. The extracted association rules lay out some hidden patterns of students' courses which could serve as a foundation stone for academic planners in making decisions and modification and an aid in the curriculum re-structuring with a view to improving students' performance and reducing break rate [5].

Shannaq et al, used the classification since data mining technique to predict the numbers of listed students by evaluating academic data from enrolled students to study the main attributes that may affect the students' truth (number of enrolled students) [6].The decision tree as a classification method to extract classification rules and the extracted classification rules are analyzed and evaluated using different evaluation methods. It allows the University management to prepare necessary resources for the new enrolled students and indicates at an early stage which type of students will potentially be enrolled and what areas to focus over in higher education systems for support and feedback.

Made a prediction model using the GP method to identify at-risk students in traditional school settings. A feature selection technique was used to reduce the attributes [7].

Wolff et al. (2013) have applied a decision-tree as data mining techniques to identify at-risk students in a virtual learning environment.

In this paper a new associative classification technique has been proposed to predict students final performance. In this research work, we have employed Honeybee Colony Optimization and Particle Swarm Optimization to extract association rule for student performance prediction as a multi-objective classification problem. Results indicate that the proposed swarm based algorithm outperforms well-known classification techniques on student performance prediction classification problem.

The rest of this paper is organized as follows: Section 2 presents the new proposed classification method for student performance prediction.

2. PROPOSED METHOD

In this section, we introduce a new approach, called Bee-RM, of multi-objective optimization based on the optimization of bee colony algorithm and particle swarm optimization.

In the following, we present the outlines of our proposed approach.

Association rule extraction is widely used data mining tasks. This is due to the interpretability feature of these rules for non-experts. The extraction of the association rules is usually performed using the meta-heuristic algorithms. In this paper, we take two major factors into consideration regarding the classification: the first one is the accuracy and the second is Interpretability.

The knowledge base used in this work is presented as a rule base. It is an important issue to select a set of optimum rules in these systems. In our Bee-RM approach, the rule extraction is performed using "pareto optimality" and considering the multi-objective factor.

Since there is rarely a unique solution which optimizes all objective functions, we look for a trade-off between objectives instead of seeking a unique solution for multi objective optimization.

2.1 RULE GENERATION BY BEE_RM

In this work, we decided to continuously extract rules as there is only few works which perform continuous rule extraction. The advantage of the continuous rule extraction is that the whole space is explored. However, the whole space exploration needs a lot of space, which demands to use more powerful algorithms.

In the following, we present how to model the association rules using the bee colony optimization and particle swarm optimization (PSO). Each member of the population is presented as an array with three rows. Then, each association rule is created by a member.

Since rules are created for each class, we use class zero as an example.

In the first array, "A" presents absence and "P" presents the specific property in the rule. In this approach, we do not need to perform bins and so the span is seen continuously.

P	A	A	P	A
0.2	5	3.5	7	0

0.9	9	5	8.5	2

The second array's values show the lower limit of each property. The third array shows the upper limit of each property. Therefore, the rule presented by these arrays is:

If (0.2<F1<0.9 and 7<F4<8.5) then class = 0

The first array contains discrete values, in the ConstructSolution function, we use the bee colony optimization in order to predicate and in the case of two other arrays which present the span, we use PSO optimization.

In the first fold of each category in the dataset, the generation is performed "MaxGeneration " times. Inside each generation, the population size is equal to the value of "Population" parameter. In each execution of the algorithm, for each class in the dataset, the generation is performed and every member of the population produces the optimized results. Then, we use the "optimized association rules extracted for all classes" as input of the classifier method in order to classify the test dataset.

Finally, the average accuracy obtained by 10-fold execution is considered as the main accuracy of the Bee_RM algorithm.

2.2 HONEYBEE HIVE OPTIMIZATION (HHO)

The "ConstructSolution" method for optimizing the first array, create a path for each bee according to the Dance Table and heuristic information. (1)

$$P_k(r,s) = \begin{cases} \frac{[\delta(r,s)]^\alpha [\eta(r,s)]^\beta}{\sum_{u \in J_{k \circledR}} [\delta(r,s)]^\alpha [\eta(r,s)]^\beta} & \text{if } s \in J_k \\ 0 \end{cases} \qquad (1)$$

The original fitness function, presented in this paper, is implemented according to the Eq (2). Below we demonstrate this function in (2).

$$H(r,s) = p_1 \times \text{support(solution)} + p_2 \times \frac{\#\text{non Don}'\text{tcare}}{\#\text{features}} \qquad (2)$$

In this formula (2) P_1 is the effectiveness and importance given to the support of produced solution, and P2 is the importance given to the "Don't-care" relative to the number of all features.

2.3 PARTICLE SWARM OPTIMIZATION (PSO)

We use the particle swarm optimization (PSO) algorithm in a continuous space and in multi objective form. The objective function in Eq (3) is used to calculate the Local-best found by each individual inside the same individual. The Global-best found in the whole population of individuals is kept in another variable called Gbest in each individual. In other word, we do not have a unique Global-best but many.

Fitness = Support Percent*Support (solution) + (1-SupportPercent)* Confidence (solution)   (3)

All Gbest are the most optimized local Non-dominated association rules obtained by Eq (4) optimization in the current population. To calculate the location of the next move of particle, we use the average of these local Non-dominated rules as demonstrated in the Eq(5) and Eq(6).

$$\alpha = \frac{\sum_{i=1}^{n} Gbest_i - x_i(t)}{n} \quad n = \text{Local Non Dominated rules} \tag{4}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{5}$$

$$v_i(t+1) = c_1 r_1 \left(Lbest - x_i(t)\right) + c_2 r_2 \, \alpha \tag{6}$$

The more general rules cover a big span of the dataset records. It reduces the interestingness of the rules. Our objective is to make a trade-off between interestingness and support value of obtained association rules. We try to extract more detailed association rules with high support value and interestingness by defining the "Interval-p" parameter.

2.4 STOPPING CONDITION

Once all rules are created by all members of the current population, local non-dominated and global non-dominated rules are determined. The most important condition to stop the training phase is a constant number of repetitions. The members continue the procedure till the stop condition is satisfied. The procedure stops if the repetition number of procedure is reached (the "Maxgeneration" number). Then, the best association rules according the Pareto-optimality optimization are selected.

3. EXPERIMENTAL RESULTS

This section shows the experimental results of the proposed method versus other classification techniques. Our proposed method will be analyses educational data generated on a Moodle platform.

Moodle's log is the baseline system used in this research. Moodle is a free virtual learning environment (VLE). Moodle is therefore evolving system and dynamic. Anyone can download and install it. An administrator is responsible for managing users (students, teachers, etc.) and course virtual classrooms. The Moodle system view differs depending on the role the user plays (teacher, student, administrator etc.).

Moodle is developed by programmers as an open source system, from all over the world. As of 2013, Moodle system has over 77,000 registered sites in over 215 countries. It prepares support to over 65 million students all over the world, trained by over 1.2 million teachers.

Moodle is only one of many support tools for virtual learning environment (VLE). There are other similar distance systems like, for example, ATutor, eCollege, Desire2Learn or Dokeos.
The information of interaction is stored as attributes in a user (student) profile. In our data set, 11 attributes and values are stored, with 357 records. These attributes include: number of interaction between student-student, student-teacher, and etc. Detail of this data set is as follows. Table 1 shows detail information about attributes of Moodle data set.

Experimentally, we have tried to set the best parameters for proposed method. The values of different user-defined parameters of Bee_RM is reported in Table 2.

Table 1.Information about features of Moodle dataset.

Category	features
Category 1	ST-ST :Student – student
Based on agent	ST-TE: student –teacher
	ST-CO :Student – content
	ST-SY : Student-system
Category 2	TC :Transmission of contents
Based on frequency of use	CI: Creating class interactions
	SA :Student assessment / evaluating students
Category 3	AC : Active
Based on participation mode	PA: Passive
Dependent variable - Academic performance	GR: Final grade

The performance of Bee_RM is evaluated using 10-fold cross-validation test (Michalski et al., 1998). In this section of research, the all obtained results are reported. Important scale to evaluate the proposed method : accuracy.

The accuracy is the number of instances correctly classified and being calculated according to Eq. (7)

$$\text{Accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN}$$

(7)

Table 2. Parameter setting of Bee_RM.

Parameter	Value
Population$_{Size}$	30
Maxgeneration	150
DefultDancers	6
C	0.5 , 0.03
P	1 , 4
SupportPercent	0.5
Interval_p	0.5
α, β	2 , 1

Figs.1 and 2 denote the effect of different population sizes of the new proposed metaheuristic algorithm on accuracy and execution time respectively. Fig. 3 shows the Influence of P_2 parameter on average length of rules.
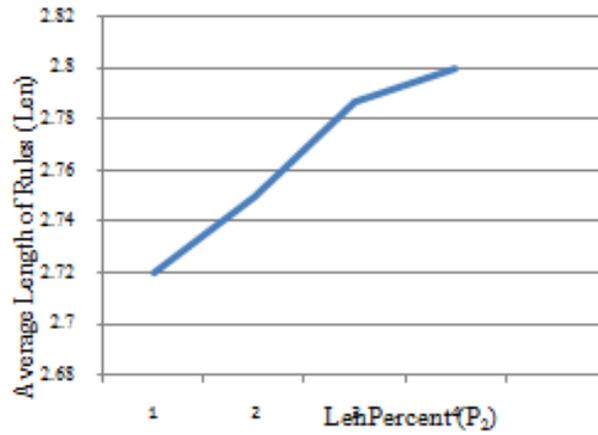

Figure 3. Infuluence of P_2 parameter on average length of rules.
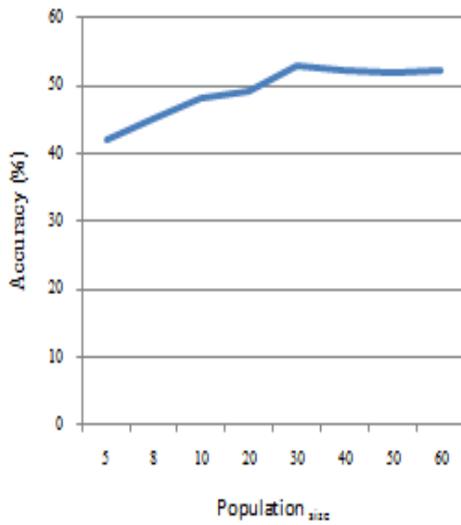

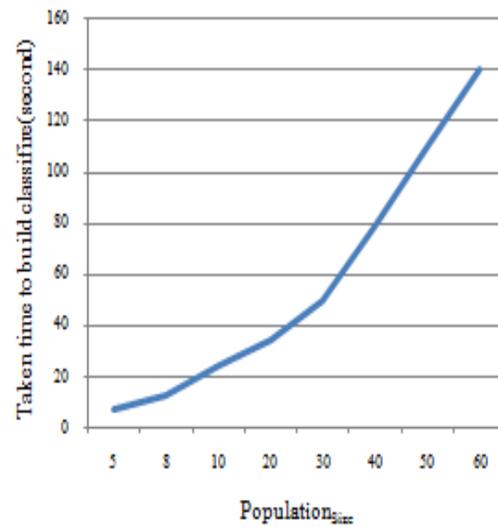Figure 1. Influence of number of individual on accuracy

Fig 2. Influence of number of individual on taken time to learn the classifier

Table 3. Classification accuracy obtained with different method for Moodle.

Method	Classification Accuracy (%)	Study
KNN	47.29% +/- 6.05%	Cover & Hart, (1967) and Rapidminer tool is available
NN	51.68% +/- 3.83%	Nsky, (1954) and Rapidminer tool is available
Baysian	43.71% +/- 8.26%	Russell, Stuart, 1995) and Rapidminer tool is available
Rule Induction	46.63% +/- 6.55%	J. Stefanowski, (1998) and Rapidminer tool is available
PART	51.26	Witten and Frank, (2005) and WEKA tool is available
OneR	45.93	Weka: http://www.cs.waikato.ac.nz/~ml/weka/
JRip	50.42	Weka: http://www.cs.waikato.ac.nz/~ml/weka/
ZeroR	41.73	Witten and Frank, (2005) and WEKA tool is available
IBK	40.33	Witten and Frank, (2005) and WEKA tool is available
Logistic	46.49	Witten and Frank, (2005) and WEKA tool is available
SimpleLogistic	51.26	Witten and Frank, (2005) and WEKA tool is available
SMO	52.10	Witten and Frank, (2005) and WEKA tool is available
NaiveBayes	36.13\	Witten and Frank, (2005) and WEKA tool is available
ClassificationVia Regression	52.66	Witten and Frank, (2005) and WEKA tool is available
Vote	41.73	Witten and Frank, (2005) and WEKA tool is available
Random Tree	45.93	Weka: http://www.cs.waikato.ac.nz/~ml/weka/
Random Forest	47.05	Witten and Frank, (2005) and WEKA tool is available
J48	46.21	J.R. Quinlan, (1993) and WEKA tool is available
CPSO-C	42	Liu et al. 2004, and KEEL tool is available
SLAVEC	51	González and Pérez 2001, and KEEL tool is available
MPLCS-C	47	Bacardit and Krasnogor 2009, KEEL tool is available
C-SVM-C	51	KEEL tool is available
XCS-c	47	Wilson 1995, and KEEL tool is available
GFS-SP-C	48	Sánchez et al. 2001, and KEEL tool is available
Bee_RM	53.46%+/-5.46%	Our study

Table 3 shows accuracy of Bee_RM versus several recent and famous classification methods. We used 3 famous tools in data mining, for comparison.

To compare our results with other studies, we have used WEKA, Rapidminer and KEEL softwares.

Six evolutionary rule learning algorithms are used in which 3 of them learn fuzzy rules and 3 of them learn crisp rules in an evolutionary way. These results reveals, our proposed method Bee_RM using 10-fold cross validation obtains the highest classification accuracy, 53.46%, reported so far. So, we can draw this conclusion that the combination of Bee Colony Optimization and particle swarm optimization with continues logic, would be very effective in predicting student final performance in educational data.

Although there is not any accurate definition for interpretability of classification methods but the number of rules (NR) and mean length of rules(Len) are often mentioned as two main factors of interpretability.

4. CONCLUSIONS

In this paper we employed the capability of swarm based techniques to extract association rules for student performance prediction as a multi-objective classification problem. The proposed algorithm had a low convergence time and it used a few number of parameters. Honeybee Colony Optimization and Particle Swarm Optimization were the two used metaheuristics to extract association rules. The fitness function in both of these algorithms considers support and length of the association rules. Results showed that using the proposed metaheuristic-based rule discovery approach enables us to extract accurate and interpretable knowledge for student performance prediction. Our future works focus on using new proposed metaheuristic algorithms such as Gravity Search and Vortex Search Algorithm instead of PSO and Honeybee Colony. Moreover, we aim to consider other measures such as confidence, correlation and interestingness along with support and rule length.

REFERENCES

[1] Peña-Ayala, Alejandro. "Educational data mining: A survey and a data mining-based analysis of recent works." Expert systems with applications 41.4 (2014): 1432-1462.
[2] Romero, Cristobal, and Sebastian Ventura. "Educational data mining: A survey from 1995 to 2005." Expert systems with applications 33.1 (2007): 135-146
[3] Baradwaj, B. and Pal, S. (2011) 'Mining Educational Data to Analyze Student s' Performance', International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, pp. 63-69.
[4] Chandra, E. and Nandhini, K. (2010) 'Knowledge Mining from Student Data', European Journal of Scientific Research, vol.
[5] Ayesha, S. , Mustafa, T. , Sattar, A. and Khan, I. (2010) 'Data Mining Model for Higher Education System', European Journal of Scientific Research, vol. 43, no. 1, pp. 24-29.
[6] Shannaq, B. , Rafael, Y. and Alexandro, V. (2010) 'Student Relationship in Higher Education Using Data Mining Techniques', Global Journal of Computer Science and Technology, vol. 10, no. 11, pp. 54-59. 47, no. 1, pp. 156-163.
[7] Marquez-Vera, C., Cano, A., Romero, C., & Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining
[8] Pieter, Adriaans. DolfZantinge, 1996. Data Mining (New York: Addison Wesley)
[9] D. T. Larose, Discovering knowledge in data: an introduction to data mining. Wiley.com, 2005.