ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 9, September 2017

Student Performance Prediction in Higher Educational Institutes Using Multi-Swarm PSO

Dr. T. Sree Ram Kumar[1]
[1]Teaching Assistant, Dept. of Computer Science, M. K. U. College, Madurai, India.

Abstract: The heterogeneity of the student population in higher education institutes presents a big challenge to teachers. Since it is impossible to devise individual training methods for all students, the teacher would greatly benefit if he/she is in possession of a tool that can predict the performance of students. Greater attention can be paid to students predicted to show poor results. In this area, Data Mining which is an advanced area in Computer Science can be of immense utility. Data Mining techniques can be used to predict a students performance based on past data available pertaining to the students of previous batches. Evolutionary Computation techniques like Particle Swarm Optimization (PSO) can provide better results given their successful application in multiple domains. The paper presents a multi-swarm PSO based classification model that can predict a student's performance. Data pertaining to students of an higher educational institution are collected and multi-swarm PSO is applied for classification. The results obtained by the technique are compared with actual results throwing light on the soundness of the proposed technique.

Keywords: Data Mining, Particle Swarm Optimization, Multi-swarm PSO, Classification, Prediction, Student Performance Prediction

I. INTRODUCTION

Teachers of higher educational institutions are confronted with a lot of challenges. Students entering the higher education institutes arrive from heterogeneous backgrounds. There also seems to be variations in the performance of students at the higher secondary school level and college level. Teachers vested with the responsibility of educating this heterogeneous population of students find it difficult to predict the performance of students at the beginning stage. But such a prediction is vital as students likely to under perform are in a need of more attention than other students. A separate teaching-learning methodology may also be required for such students who are likely to under perform.

Higher education teachers who have a rich experience are armed with the information necessary to confront this crucial challenge. They have access to data pertaining to past students and also their performances. This paper proposes an approach to exploit this useful information in predicting the performance of the newly enrolled students.

II. BACKGROUND

Data Mining is a novel arena within the discipline of Computer Science. It is a technique that attempts to derive useful knowledge from large volumes of available data. Data Mining is extraction or mining knowledge from large amounts of data (Han & Kamber, 2006).

Data Mining techniques have been applied in various domains and have yielded very successful results in terms of the patterns uncovered in these domains. The present paper is an attempt to apply the same for student performance prediction. In particular, the paper proposed the use of multi-swarm PSO for performing the task of prediction.

A) Particle Swarm Optimization (PSO) and Multi-Swarm PSO

Particle Swarm Optimization (PSO) was conceived by observing the social behavior of birds in a flock and attempts to simulate the same in uncovering solutions to problems (Ebberhart & Kennedy, 1995; Kennedy & Ebberhart, 1995). The basic idea behind PSO is as follows: each particle flies in the search space with a velocity adjusted by its own memory together with flying experience of the companions. The fitness of each particle is determined by an objective function given as:

$$v_{id}^{t} = w \times v_{id}^{t-1} + c_1 \times r_1 \left(p_{id}^{t} - x_{id}^{t} \right) + c_2 \times r_2 \left(p_{gd}^{t} - x_{id}^{t} \right)$$

. Here i represents the ith particle, d is the dimension of the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 9, September 2017

solution space, c_1 denotes the cognition learning factor, and c_2 indicates the social learning factor, r_1 and r_2 are random numbers uniformly distributed in (0,1), p_{id}^t and p_{id}^t stand for the position with the best fitness found so far for the ith particle and the best position in neighborhood, v_{id}^t and v_{id}^{t-1} are the velocities at time t and time t – 1 respectively, and x_{id}^t is the position of ith particle at time t. Each particle moves to a potential new solution based on the equation $x_{id}^{t+1} = x_{id}^t + v_{id}^t, d = 1,2,...,D$

The technique used in this research is the modified Multi-Swarm PSO proposed by Liu et. al. Liu et.al. use the Multi-Swarm PSO for feature selection. In this approach, a number of sub-swarms are employed and a multi-swarm scheduler monitors and controls each sub-swarm using the following rules (Liu et.al, 2011):

The swarm request rule

Given:

$$S_i = \begin{cases} 1, \text{ if } d_i < \dfrac{tit_i - it_i}{tit_i} \times rand() \times fitness \\ 0, \text{ if } d_i \geq \dfrac{tit_i - it_i}{tit_i} \times rand() \times fitness \end{cases} \quad \text{Eq (1)}$$

Here, d represents a threshold, tit the maximum iteration number, it the current iteration number. rand() is a random number uniformly distributed in U(0,1). If S_i=1, the current sub swarm sends the results (it's corresponding pbest and gbest values) to the multi-swarm scheduler.

The multi-swarm scheduler request rule

The multi-swarm scheduler monitors each sub-swarm and can send a request to obtain a result from any sub-swarm if it is valuable. If the sub-swarm has sent swarm request rules more than a specified number of times, the multi-swarm scheduler will send the rule i.e. the multi-swarm scheduler sends this rule depending on the activity level of the sub-swarm.

The multi-swarm collection rule

The multi-swarm scheduler collects results from the alive sub-swarms and updates pbest and gbest values.

The multi-swarm destroying rule

The multi-swarm scheduler destroys sub-swarms in 2 cases. In the first case, if the sub-swarm sends the swarm request rule fewer than a threshold number of times and in the second case when the sub-swarm does not change the gbest value within a specified number of iterations.

B) The Multi-Swarm PSO algorithm

The Multi-Swarm PSO Algorithm used by the proposed methodology is the one used by Thangammal and Pethalakshmi (2017).

Step 1: Initialize: Load the data from the source. Initialize the size of swarms randomly. Initialize particle positions and velocities of each swarm with random values. Calculate the objective function and update pbest and gbest for each swarm.

Step 2: Parameter Selection: Specify the flowing parameters for each swarm: Lower and upper bounds on velocity, size of particles, number of iterations, cognition learning factor, social learning factor, threshold d_i in Eq 1, thresholds required for the multi-swarm destroying rule. Set iteration_number=0

Step 3: For each swarm, if iteration_number < max_iterations or gbest has changed in more than 50 iterations, continue with step 4. Otherwise, destroy the swarm and go to step 10. The main module compares the gbest of each swarm with the previous one in the module to decide if gbest needs to be updated using multi-swarm scheduler request rule. In case pbest or gbest is changed, multi-swarm collection rule is executed

Step 4: For each swarm, if current particle number < particle size, continue with step 5. Else, proceed to step 9.

Step 5: For each swarm, update the position and velocity for each particle. Lookup gbest and pbest values for this updation.

Step 6: Restrict position and velocity for each particle

Step 7: Calculate fitness for each particle and update pbest and gbest. Execute swarm request rule. If the current swarm needs to be destroyed according to multi-swarm destroying rule dispose the current swarm

Step 8: current particle number = current particle number + 1

Step 9: current iteration number = current iteration number + 1

Step 10: Execute Swarm collection rule and exit

III. DATA MINING FOR HIGHER EDUCATION

Data Mining holds great promises for multiple domains and higher education is no exception. Data Mining for example, can be used to study the characteristics of students that most likely contribute to their learning. It can also be used to derive those set of students with similar competency levels in an attempt to devise custom teaching

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 9, September 2017

methodologies suited for a cluster of students. And most important, data mining can also be used to classify students into predefined categories of learning levels. Thus data mining can be used to predict a student's performance. Such a prediction would be of great utility to the teacher who can focus on students likely to show poor performances.

El-Halees applies data mining to enhance knowledge on variables that are most likely to affect a student's learning behavior (El-Halees, 2009). Ayesha et.al. state that lack of deep knowledge in the higher educational system may be come a road block in the achievement of quality objectives and data mining can help address knowledge gaps in this area (Ayesha et.al., 2010). Pandey and Pal present a novel approach for prediction of student's performance using the classification (Pandey & Pal, 2011).

IV. Research Methodology

The objective of the research is to determine the potential of a multi-swarm PSO based classification in predicting performance of students in higher educational institutions. The motivation of the research is the work of Pandey and Pal who use the Bayesian method of classification. The research proposes to use Multi-Swarm PSO for the task and the soundness of the approach is validated by empirical results on performance of students collected from a college in Madurai district of Tamil Nadu.

The research attempts to classify the students admitted to the first year of M.Sc., degree in Computer Science and Information Technology. The department teachers are confronted with the challenge of underperformance and failure of a large number of students. But it was quite challenging to discover the characteristics that governed the performance of a student. Having an under graduation in Mathematics or any discipline with Mathematics as one of the subjects, the ambition of the student and a lot of other factors seemed to contribute to the results. The research was aimed at predicting the performance of students in providing a useful insight to teachers who can pay more attention to students predicted to under perform.

The characteristics of the students taken into account are tabulated below:

TABLE 1

STUDENT CHARACTERISTICS CONSIDERED FOR THE STUDY

Characteristic	Description
Medium of Instruction in UG	English / Tamil
Grade obtained in UG	I / II / III
UG Stream	Arts/Science
Economic Status	Poor/ Medium / Good
Educational Status of Parents	Poor / Medium / Good
Educational Status of Siblings	Yes / No
Community	GEN/OBC/SC/ST

The above tabulated data are collected from 500 students of older batches. Out of these, data pertaining to 350 students were used for training and the remaining 150 were used for testing. The performance of all these students were classified as Good/Poor based on their performances in the examinations.

The Multi-Swarm PSO is applied to build a security classification model based on the above collected data. The fitness function is obtained by the number of misclassifications produced by the model. The lesser misclassifications a model results in, the more fit it is. But, the damage incurred by misclassifying a poor performing student as Good is much more than the damage incurred by misclassifying a Good performing student as poor because the maximum damage that could stem from the later misclassification is wastage of resources. Therefore, the fitness function penalizes the former misclassifications more than the later. If a is the number of former misclassifications and b is the number of later misclassifications, the fitness function is c * a + b, where c is a parameter to be tuned for the application. For the purpose of this research c was chosen to be 2. Since the performance of all the students in the training and test data set are known apriori, there is little work in determining the number of misclassifications.

V. Results and Discussion

The classification accuracy of the model, following [Liu etal.] is defined as:

$$accuracy(N) = \frac{\sum_{i=1}^{|N|} assess(n_i)}{|N|}, n_i \in N,$$ where

$$assess(n) = \begin{cases} 1 \text{ if classify(n)} = \text{nc} \\ 0 \text{ otherwise} \end{cases}$$

Here N is the number of students to be classified (350 for training, 150 for testing), nc is the class of the item $n \in N$. The best result for the training data set yielded a classification accuracy of 83.21% while that for the test data set yielded a classification accuracy of 87.41% for students in the poor performance category. The testing data set yielded better results than the training data set implying that the model has captured the underlying relationships between

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 9, September 2017

the data items involved. The complete results of classification accuracy for students in the poor and good performing classes are presented in tables 2 and 3 for the training set and test set respectively

TABLE 2

RESULTS FOR THE TRAINING SET

Class	Percentage correctly classified
Poor	83.21%
Medium	79.94%
Good	81.12%

TABLE 3

RESULTS FOR THE TEST SET

Class	Percentage correctly classified
Poor	87.41%
Medium	82.34%
Good	85.22%

The results obtained are quite promising in that the system is able to predict the under- performance of 87.41% students. This can be a useful insight for the teacher who can utilize this information to pay more focused attention and adopt suitable teaching methodologies for these students.

To ascertain the superiority of Multi-Swarm PSO over plain PSO, the research also compares the results obtained with those obtained with plain PSO. Since, we are particularly interested in students under the poor performance category, the results for students in that category are compared.

TABLE 4

COMPARISON WITH PSO (FOR POOR CATEGORY STUDENTS)

Data Set	Multi-Swarm PSO	PSO
Training	83.21%	75.71%
Test	87.41%	78.16%

The lower classification accuracy reported by PSO can be attributed to the fact that PSO has the tendency to converge quickly at local optimal values. Multi-Swarm PSO is also superior to other techniques like Genetic Algorithms as it entainls the fine tuning of a relatively fewer number of parameters.

VI. Conclusion and Future Work

The prospects of applying a multi-swarm PSO based classification for predicting student's performance was illustrated with an empirical study. The results can throw light on the expected performance of the newly entering students in higher education institutes. This can be of immense utility to the teacher who can tailor the teaching methodologies to suit the students and focus more attention to the students predicted to under perform.

The characteristics taken into consideration for the research are quite limited. This seems to be one of the biggest hurdles in the problem. A student's performance is a complex interplay of multiple factors and it is plausible that some of the crucial characteristics are ignored in the study. Increasing the sample size and verifying the performance of the system and incorporation of more characteristics can be a good starting point for researchers inclined to research on the area.

References

[1]. El-Halees, *Mining Students Data To Analyze Learning Behavior: A Case Study*, 2009. Retrieved from: uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/30/papers/f158.pdf. Accessed on: 29 Aug 2013.

[2]. Goldberg, D. E., 2002, Design of Innovation: Lessons From and For CompetentGenetic Algorithms, Kluwer, Boston, MA.

[3]. Han, Jiawei, Kamber, Micheline, Data Mining" Concepts and Techniques, Second Edition, Morgan Kaufmann publishers, 2006.

[4]. Holland, J. H., 1975, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI.

[5]. Pandey, Umesh Kumar, Pal. S, Data Mining : A prediction of performer or underperformer using classification, International Journal of Computer Science and Information Technologies, Vol. 2 (2) , 2011

[6]. Shaeela Ayesha, Tasleem Mustafa, Ahsaan Raza star, M Inayat Khan-Data mining Model for higher education, ISSN 1450-216X Vol.43 No.1 (2010) pp24-29

[7]. Vivanco, Rodrigo, Pizzi, Nicolino, Finding Effective Software Metrics to Classify Maintainability Using a Parallel Genetic Algorithm, Genetic and Evolutionary Computation – GECCO 2004, Genetic and Evolutionary Computation Conference, Seattle, WA, USA, June 26-30, 2004. Proceedings, Part II

[8]. Liu, Yuanning, G Wang, H Chen, H Dong, X Zhu, S Wang, An Improved Particle Swarm Optimization for Feature Selection, Journal of Bionic Engineering, Vol 8, Issue 2, June 2011.

[9]. Eberhart R, Kennedy J. A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 1995, 39–43.

[10]. Kennedy J, Eberhart R. Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Network, Perth, Australia, 1995, 1942–1948.

[11]. Thangammal N, Pethalakshmi A, A Multi=Swarm Particle Swarm Optimization Model for Web Page Security Classification, Advances in Natural and Applied Sciences, AENSI Publications, May 2017.