International Journal of Education and Information Studies.
ISSN 2277-3169 Volume 5, Number 1 (2015), pp. 67-74

http://www.ripublication.com

A Hybrid Classification Method for Disengagement Detection in Online Learning

[1]P.V. Praveen Sundar and [2]Dr. A.V. Senthil Kumar

[1]Research Scholar, PG and Research Department of Computer Applications, Hindusthan College of Arts and Science, Coimbatore, India.
[2]Director, PG and Research Department of Computer Applications, Hindusthan College of Arts and Science, Coimbatore, India.

Abstract

The main aim of the online learning system is to meet the requirements of the learners and to make efficient for learners where the aspects and complexity are taken into consideration. The learner's motivational states are undertaken by many attempts, mainly by using design. Motivations are started by using analysis of log file. Firstly, the disengaged learners are identified moderately, and then visualize the disengaged learners which includes evaluation of many motivational characteristic for learning. For improvement in learning, data mining and machine learning methods will provide us meaningful data and valuable information. The performances of Bayesian classifiers endure in the field where it involves correlated features. Naïve Bayesian classification with PSO method is already implemented in many fields, the main problem in PSO is its tendency of trapping into local optima. To overcome this problem, this research presents the hybrid algorithm by combining fast PSO and Naïve Bayesian classifier for classification to aid in the prediction of disengagement. According to the characteristics of the data, our proposed method improves the classification accuracy and avoids the loss of information. This study results showed that the method was feasible and effective.

Keywords-- Online learning, Log File Analysis, Disengagement Detection, Bayesian classifiers, Particle Swarm Optimization, Quasi Framework.

I. INTRODUCTION

An Online learning system aims to provide efficient information, but motivational assessment is not taking into consideration in the learning process. Online learning system can be improved by detecting disengagement learners which would allow personalizing the involvement at proper times of learners. For this improvement, online learning system is researched using data mining techniques in education. There are various attributes identified for this disengagement prediction using log file analysis [13, 14, 15]. Though, there have been several attempts to include motivation in Online Learning systems and currently, the influence on cognition is acknowledged and taken into consideration. A factor that offered more possibilities for the inclusion of motivation in Online Learning was the usage of logged information. To provide a high quality of education, it is important to identify the unmotivated students and motivate them periodically. Disengagement Detection is a key concept for the predicting the activities of the learners. To detect disengagement, log file analysis is used. The actions registered in log files are inspected and attributes related to them are established for the analysis.

Classification is a predictive data mining technique which makes prediction about values of data using known results found from different data. To find out the solution for our problem, we propose hybrid method to classify the dataset and then perform the independent analysis to perform the significant difference among the study. But, this Naïve Bayes algorithm suffers due to over sensitivity on unnecessary/ irrelevant attributes and PSO suffers due to local optima problems. So, the accuracy will reduce in predicting disengagement with the correlated features. This paper empowers the hybrid PSO algorithm with Naïve Bayes classifier by extends it in two behaviors. Primary, it enable the hybrid PSO to choose attribute for Naïve Bayes algorithm; which is additional innovative than the Naïve Bayes classifier. Second, it apply the hybrid PSO to a functional classification data algorithm and then compare the performance of the classifiers against the act of a hybrid PSO with classifiers on the undertaking of select attributes on those data sets. The criteria used for this comparison are (1) to maximize predictive accuracy, and (2) to finding the smallest subset of attributes.

The rest of the work is structured as follows: Section-2 provides related works of this study. Section-3 describes the prediction of the disengagement model. Section-4 describes the experimental results and finally in Section 5 conclusions are outlined.

II. LITERATURE SURVEY

Bouvier et al [16] proposes a qualitative approach for identify learners' motivation from their traces of interactions performed in the learning game. Tempelaar et al [5] contribution aims to provide a practical application of Shum and Crick's theoretical framework of a learning analytics infrastructure that combine knowledge dispositions data with data extract from computer-based, formative assessments. Based on the interaction data, Bouvier et al [17] able to identify engaged and non-engaged users and qualify their types of engaged-behaviors. Vandewaetere et al [9] aims at giving an overview of the current research that addresses advanced technologies, models, and approaches to establish personalized learning, instruction, and performance. Development of eye tracking technology brings new opportunities to enhance learning experience of students and answering questions and feedback to motivate weak students for use in computer science and e-learning [18]. Jraidi et al [6] proposes a probabilistic framework using a dynamic Bayesian network to recognize the probability of experiencing each trend as well as the emotional responses occurring subsequently. Ghergulescu et al [7] uses automatic measurement and analysis of learner's motivation in game based learning. Katuk et al [11] reports the differences in learning experience from the learner's perspectives using an adaptive e-learning system, where the learner's knowledge or skill level is used to configure the learning path. Imbalanced dataset problems have become an important research topic in data mining. Thus, Sobran et al [12] investigates the classifier performance for imbalanced dataset problems using the original Naïve Bayes classifier. Alshammari et al [8] attempts to bridges this gap by comparing a number of adaptive e-learning systems. A set of criteria is generated for each perspective, and applied to a representative sample of adaptive e-learning systems. Conati et al [4] presents a study that investigates the factor affect student attention to user-adaptive hints during interaction with an educational computer game. Alejandro [1] review pursues a twofold goal, the first is to preserve and enhance the chronicles of EDM advances development; the second is to organize, analyze, and discuss the content of the review based on the outcomes produced by a DM approach. In the present investigation Padilla et al [2] improves the features classification of a cardiac arrhythmias database using Differential Evolution and Particle Swarm Optimization and Naive Bayes classifiers. Nouaouria et al [10] offers a survey of recent work on Particle Swarm Classification (PSC). Chu et al[3] proposes a personalized e-Course Composition approach based on PSO algorithm, called PC2PSO, to compose appropriate e-learning materials into personalized e-courses for individual learners.

III. MODELING FRAMEWORK BY IMPROVING NAÏVE BAYESIAN CLASSIFIER

In this section, a technique designed to improve the Naïve Bayes classifier for predicting disengagement of learners using quasi framework log files. The PSO comprises a set of search techniques, inspired by the behavior of natural swarms, for solving optimization problems in reality. The hybrid PSO algorithm is employed as a search strategy in order to identify an optimal weighting for attributes probabilities from Naïve Bayes classifier and to solve the optimization problems.

A. Prediction Model Developments

The prediction model was developed on Quasi Framework [19]. The Structure of Quasi Framework is explained in [13, 14]. A catalog of all achievable events that are record by Quasi Framework [19] is presented in [15].

B. Naïve Bayes Classifier with Hybrid Particle Swarm Optimization Technique

Naïve Bayesian (NB) classifiers very simple classification which is composed of directed acyclic graphs with only one parent and several children with a strong assumption of independence among child nodes in the context of their parent. The basic independent, the major advantage of the Bayesian classifiers is its short computational time for training. Hybrid PSO-based classifiers require many classifications to be performed during feature selection and extraction, the use of a computationally efficient classifier such as a Bayesian classifier is indicated. The direct application of feature weighting and scaling is not effective in connection with Bayesian classifier, because the Bayes decision rule is invariant to linear scaling of the feature space.

In standard PSO initial particle distributed evenly in the searching field, PSO was effective for low dimension function optimization problem. The basic PSO can converge fast, but susceptible to fall into local minima easily. To solve this problem, this research attempt has been made with the following improvements. In hybrid PSO, the potential solutions are represented by vectors of real values i.e. the dimension of each vector corresponds to the number of constraints of the problem rather than number of items.

The inertia weight (w) is the most important parameter that moves the particle towards the optimal position. Thus, to increase the search ability the particles flight should be controlled by the objective functions. The particle which is closer to the optimal point should move slowly as compared to the other particle. This movement of the particle can be controlled using different w values according to their rank between w_{min} and w_{max}as give in following equation (1):

$$w_i = w_{min} + \frac{w_{max} - w_{min}}{T_{pop}} \times \text{rank} \tag{1}$$

Where T_{pop} is the total size of the swarm. The velocity of the particle, both for continuous and discrete space, is updated using the equation (2).

$$v_{id}(t+1) = wv_{id}(t) + c_1 rand_1()\big(p_{id}(t) - x_{id}(t)\big) + c_2 rand_2()(p_{gd}(t) - x_{id}(t)) \tag{2}$$

However, the position updating of the particle represented by continuous vector is done using equation (3) and discrete vector is updated using equation (4) respectively.

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{3}$$

$$S(v_{id}) = 1/(1 + \exp(-v_{id})) \tag{4}$$

$$x_{id} = \begin{cases} 1 & \text{if } S(v_{id}) > rand() \\ 0 & \quad \text{otherwise} \end{cases} \tag{5}$$

During the execution of the algorithm, each individuals passed through the classifier for evaluation and the cost is computed based on the classification accuracy obtained from the parameterized generalized naïve bayes formulation in classifying the known set of samples of known class. The proposed Hybrid PSO algorithm seeks to maximize the cost function. The mathematical formulation of the cost function is given in equation (5):

$$f(\vec{x}) = A_c \times CL_{accu} + \frac{A_f}{Sfr} \tag{6}$$

WhereCL_{accu} is the classification accuracy, Ac is the weight factor associated with the classification accuracy, Af is the weight factor associated with the selected number of weights and $Sfr = \sum m_i \times rank(w_i)$. The weights vectors are assigned rank based on the rank of $P(x_j|C_j)$. The value for the Ac and Af are empirically found.The designed classifier efficiency is evaluated based on the predictive accuracy of the test dataset. The algorithm for the Naïve Bayesian classifier with hybrid PSO is shown in Algorithm 1.

Algorithm 1 Pseudo code for learnable Bayesian classifier

Calculate the mean and standard deviation for continuous values and prior probability for discrete values from the training set and save it in a file.
 	
Initialize the swarm *S* with swarm size
for each particle of the *S* **do**
  Initialize the velocity v_i
  Initialize the *pbest pi*
end for
Evaluate the fitness *F* =*fitness*(*S*);
Select the *gbest* as the particle from swarm *S* with maximum
fitness value F_i
while stopping criterion is not met **do**
  for each particle **do**
    Find the weight *w* using equation (1)
    Update the velocity using equation (2)
    if particle is real valued vector **then**
        Update particle using equation(3)
else
Update particle using equation(4)
end if
end for
Evaluate the fitness $F_{new} = fitness\ (S)$;
for each particle i **do**
if $F_{new}(i) > F(i)$**then**
update the $pbest\ p_i$
end if
end for
if $\max(F_{new}) > \max(F)$**then**
update the gbest with the particle having maximum
value of F_{new}
end if
$F = F_{new}$
end while

function Fitness (S)

for each particle i **do**
find the classification accuracy and the rank
[*accuracy, rank*] =classification(training set, $particle_i$)
find the fitness based on accuracy and rank using equation(5)
end for
return fitness

The algorithm section above is updated in the testing algorithm of the proposed Naïve Bayesian classifier with hybrid PSO model. The updated algorithm is used as fitness function for the hybrid PSO algorithm. The particlei is optimized for hybrid PSO algorithm, where the fitness value here is the average precision only.

IV. EXPERIMENTAL RESULTS

For this experiment, log file of 107 users are collected from [19] where each and every user has spent up to 10sessions for learning and attended 1 to 10 assessments. The activity between login and logout is considered as session. From this process 3,67,339 instances have been obtained, from the logged events, 32 attributes were derived. The collected data from 107 learners were analyzed using Naïve Bayesian classifier and Hybrid Particle Swarm Optimization (PSO). These methods represent the most commonly used techniques for the data types of our datasets: nominal data for the predicted variable and numeric data for the predictors.

Table-1 shows the comparison of before selection process and after selection process with hybrid PSO Algorithm. The results displayed in Table-1 shows that Naïve Bayes with Hybrid PSO algorithm gets better accuracy than other classification algorithms in both before and after feature selection process.

TABLE 1: EXPERIMENTAL RESULTS

Classification	Before Feature Selection	After Hybrid with PSO
Naïve Bayes(NB)	84.11%	90.65%
J48 (DT)	83.18%	86.91%
Random Forest(RF)	85.04%	88.79%
Bayes net (BN)	87.85%	89.72%

As mentioned in the introduction, observed two patterns of behavior is proved and observed a considerable improvement of prediction. It is clear that the hybrid PSO with naïve Bayes classifier learns faster than standard classification algorithms, i.e. with small number of training data (e.g. 10%), the prediction accuracy for hybrid PSO with bayesian classifier is higher.

V. CONCLUSION

This research summarized that a simple modification to the standard PSO with naive Bayesian classifier through forward selection of attributes using estimated accuracy on an online learning system [19]. The computational results indicate that the use of variables apparently unrelated to the class attribute tends to reduce the accuracy and reliability of a classification model. However, the hybrid PSO algorithm clearly tends to find smaller subsets of attributes than the PSO with Naive Bayesian classifier. The Naïve Bayesian classifier combined with hybrid PSO feature selection method proves to be the best feature selection capability without degrading the classification accuracy. It is further proved to be an effective method for mining large structural data in much less computation time. These results were compared by utilizing log files from quasi framework [19] and the correspondence of results suggests that this classifier on disengagement prediction is very efficient, fast and successful.

REFERENCES

[1] Alejandro Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works”, Expert systems with applications, Volume 41,Issue 4, pp:1432-1462, 2014.

[2] Christian Padilla-Navarro, Rosario Baltazar-Flores, David Cuesta-Frau, Arnulfo Alanis-Garza and Victor Zamudio-Rodríguez, “Cover Cardiac Arrhythmia Classification Using KNN and Naive Bayes Classifiers Optimized with Differential Evolution (DE) and Particle Swarm Optimization (PSO)”, In Workshop Proceedings of the 9th International Conference on Intelligent Environments, Volume 17,pp:36-46,2013.

[3] Chu Chih-Ping, Chang Yi-Chun Chang and Tsai Cheng-Chang, “PC2PSO: personalized e-course composition based on Particle Swarm Optimization”, Applied Intelligence, Volume 34, Issue no. 1, pp: 141-154, 2011.

[4] Cristina Conati, Natasha Jaques and Mary Muir, “Understanding attention to adaptive hints in educational games: an eye-tracking study”, International Journal of Artificial Intelligence in Education, Volume 23, Issue 1-4,pp-136-161,2013.

[5] Dirk T Tempelaar, André Heck, Hans Cuypers, Henk van der Kooij, Evert van de Vrie, “Formative assessment and learning analytics”, In Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM, pp:205-209,2013.

[6] Imène Jraidi, Maher Chaouachi, and Claude Frasson, “A Hierarchical Probabilistic Framework for Recognizing Learners’ Interaction Experience Trends and Emotions”, Advances in Human-Computer Interaction, Volume 2014 pp:1-16,2014.

[7] Ioana Ghergulescu and Cristina Hava Muntean, “A Novel Sensor-Based Methodology for Learner's Motivation Analysis in Game-Based Learning”, Interacting with Computers, Volume 26,Issue 4, pp:305-320,2014.

[8] M. Alshammari, R. Anane, and R. J. Hendley, “Adaptivity in E-Learning Systems”, In Proceedings of The Eighth International Conference on Complex,

Intelligent, and Software Intensive Systems (CISIS 2014), Birmingham, UK,pp:79-86,2014.

[9] Mieke Vandewaetere and Geraldine Clarebout, “Advanced Technologies for Personalized Learning, Instruction, and Performance”. In Handbook of Research on Educational Communications and Technology, pp: 425-437. Springer New York, 2014.

[10] Nabila Nouaouria, Mounir Boukadoum and Robert Proulx, “Particle swarm classification: a survey and positioning”, Pattern Recognition, Volume 46, Issue 7, pp:2028-2044, 2013.

[11] Norliza Katuk, Jieun Kim and Hokyoung Ryu, “Experience beyond knowledge: Pragmatic e-learning systems design with learning experience", Computers in Human Behavior, Volume 29, Issue 3, pp:747-758, 2013.

[12] Nur Maisarah Mohd Sobran, Arfah Ahmad and Zuwairie Ibrahim, “CLASSIFICATION OF IMBALANCED DATASET USING CONVENTIONAL NAÏVE BAYES CLASSIFIER”, In Proceeding of the International Conference on Artificial Intelligence in Computer Science and ICT(AICS 2013), 25-26 November 2013, Langkawi, MALAYSIA. pp: 35-42, 2013.

[13] P.V.Praveen Sundar, “Quasi Framework: A New Student Disengagement Detection in Online Learning”, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 10, pp: 1-6, December- 2012.

[14] P.V.Praveen Sundar,A.V.Senthil Kumar,”Disengagement Detection in Online Learning using Quasi Framework”, IOSR Journal of Engineering(IOSRJEN), Volume 5 Issue-1.2015 (VI) pp:04-09.

[15] P.V.Praveen Sundar, A.V.Senthil Kumar, “An Enhanced Disengagement Detection in Online Learning using Quasi Framework”, International Journal of Applied Engineering Research(IJAER), Vol. 10 No.55 (2015), pp: 1298-1302.

[16] Patrice Bouvier, Élise Lavoué, Karim Sehaba and Sébastien George, “Identifying Learner’s Engagement in Learning Games: a Qualitative Approach based on Learner’s Traces of Interaction”, In 5th International Conference on Computer Supported Education (CSEDU 2013), pp:339-350,2013.

[17] Patrice Bouvier, Karim Sehaba, Élise Lavoué, “A trace-based approach to identifying users’ engagement and qualifying their engaged-behaviours in interactive systems: application to a social game”, User Modeling and User-Adapted Interaction, Volume 24,Issue 5,pp:413-451,2014.

[18] Peter Lach, "Intelligent Tutoring Systems Measuring Student’s Effort during Assessment." In Proceedings of Advances in Artificial Intelligence, Lecture Notes in Computer Science, Springer Berlin Heidelberg, Volume 7884, pp: 346-351, 2013.

[19] Quasi Framework, http:www.quasiframework.com, 2012.