IJCTA, 9(27), 2016, pp. 295-301 ISSN: 0974-5572

Disengagement Detection in Online Learning Using Log File Analysis

P.V. Praveen Sundar* and A.V. Senthil Kumar**

ABSTRACT

Nowadays, most of the educational institutions/universities prefer online learning due to its flexibility to learners. The learners can learn at their flexible place and time. The main disadvantage of the online learning is;it is difficult to identify whether the learners are motivated during the learning process or not. The successful online learning environments will automatically identify the disengaged learners and motivates them as quick as possible. Thus disengagement detection is considered as an important factor in online learning environments. There are so many researchers are already undergoing for detecting the disengaging learners. Through this proposed work, we detect the disengaged learners using log file analysis. Generally, in log file analysis, each and every log file is monitored. As the number of users and log file sequences are increased, it takes more time to identify the unmotivated learners. Instead of taking all log file sequences, the proposed work takes the overall values and predict the disengagement. In addition to this disengagement can be predicted using their academic performance. The results show that our proposed methodologies will give better results than the traditional log file based approaches.

Index terms: Online Learning, Log-file analysis, Disengagement Detection, Quasi Framework.

1. INTRODUCTION

Motivation is recognized as an important prerequisite of learning. While in a classroom setting motivation can be addressed by teachers, but in e-Learning environments new ways for motivating learners are required. Several approaches addressing motivational issues have been proposed, including the design of attractive e-Learning systems [1], using game features to motivate learners [2] and clickers [3] as well as animated agents. Nevertheless, learners are not getting the full benefit of these features if they do not engage in the first place. These approaches focus on making the interaction attractive rather that addressing motivation in a personalized manner. Motivational issues often go beyond the facilities of a system and its engaging character to personal characteristics like the learners' attitudes to the subject matter, their attitudes toward the tutor [4] and their current mood [5]. Therefore, knowledge about the engagement status and the motivational characteristics of learners could enhance the educational systems with detection capabilities and, ultimately, with personalized intervention strategies targeting the motivational status and characteristics of the learners.There has been a surge of interest and research on the topic of engagement in the last twenty years [6].

Student engagement is an important topic for teachers, parents, and other stakeholders. Student engagement is critical to study for three reasons [7]. First, it is a necessary condition for students' learning because engagement is a critical component of long-term achievement and academic success [8]. Second, engagement shapes students' school experiences in school, both psychologically and socially. Last, engagement plays a role in students' academic resilience, and the development of resources for coping adaptively with stressors, which in turn, may affect the development of long-term academic mindsets[9]. To do this efficiently, automatic analysis is necessary.

* Research Scholar, PG and Research Department of Computer Applications, Hindusthan College of Arts and Science, Coimbatore, India, *Email: praveensundarpv@gmail.com*

** Director, PG and Research Department of Computer Applications, Hindusthan College of Arts and Science, Coimbatore, India, *Email: avsenthilkumar@yahoo.com*

The learner's actions preserved in log files have been relatively recently discovered as a valuable source of information and several approaches to increase the motivational strategies have been used in log-file analysis. An important advantage of log-file analysis over self-assessment approaches is the unobtrusiveness of the assessment process, similar to the classroom situation where a teacher observes that a learner is not motivated without interrupting his/her activities.

Several efforts to detect motivational aspects from learners' actions are reported in the literature [10-12]. However, all these efforts are concentrated on Intelligent Tutoring Systems or problem solving environments. As online content-delivery systems are increasingly used in formal education, there is a need to extend this research to encompass this type of systems as well. The interaction in these systems is less constrained and structured compared with problem-solving environments, posing several difficulties to an automatic analysis of learners' activity.To address this challenge, restricted proposed research to one motivational aspect, disengagement, and looked at identifying the relevant information from learners' actions to be used for its prediction. Being able to automatically detect disengaged learners would offer the opportunity to make online learning more efficient, enabling tutors and systems to target disengaged learners, to reengage them, and thus, to reduce attrition. In this research work investigate the extendibility of classifier approach to other systems by studying the relevance of these attributes for predicting disengagement in a different e-learning system.

2. RELATED WORKS

In the past, researchers generally tended to reflect motivation and engagement within a single theoretical framework and conceptualized disengagement as emerging, at least in part, from variables including student attributes and presuppositions.Studying disengaged behaviors in the context of online learning is a recent field of study. Research has shown that many students engage in haphazard and non-goal directed behaviors during inquiry and problem solving [13]; one possible explanation for this is disengagement. Some forms of disengaged behavior in online learning have been shown to not only have immediate effects on domain learning [14], but also have shown to result in lower achievement on standardized exams [15], and even to lead to lower probability of attending college [9].

While some researchers refer to a single behavior pattern as "disengagement" [16] work over the last several years has suggested that learners can disengage from learning in several ways. Instead of engaging deeply inlearning, many students disengage by (a) gaming the system; (b) by engaging in off-task behaviour[17] or haphazard learning [13] (c) by becoming careless and giving wrong answers due to lack of effort rather than lack of knowledge. All of these behaviors can occur within traditional learning settings as well as in online environments.

As noted in [18], what are needed are better theoretical models that can account for student engagement (and disengagement) in science. Specifically, seek to shed further light on how to operationalize engagement, a fuzzy construct. Defining and identifying behaviors that are associated with disengagement, turning the construct "on its head" to define, operationalize, and detect (i.e., identify using a computational technique) engagement by identifying its opposite, disengagement. Research into detection of disengagement using time, performance and mouse movement features has also been investigated. This researched showed that the approach utilizing the mouse movement information outperforms a detection model that only utilizes time features and also outperforms a detection model that uses time and performance features together [19]. Research into detecting student disengagement using eye tracking methods has also been conducted in Gaze Tutor [20] which detected disengagement and boredom and developed interventions that used dialogs which responded to eye movement in an attempt to reorient student's attention to the tutor.

3. PROPOSED METHODOLOGY

The Proposed work identifies the unmotivated learners through log file analysis. The main advantage of using log file analysis is without interrupting the learner, the instructor can observe the learner's activities. The log files generally have more information about the learner's attitude. Meanwhile, log file analysis alone could not have enough information to identify the unmotivated learners. Thus in the proposed work we integrate the log file information with database values for better prediction results.

The researchers [21], [22], [23], [25] identifies the unmotivated learners using log file analysis. The disengagement can be identified based on their learning attitudes. [24] mentions that the amount of time spent on learning alone is not enough to decide the engagement level. Thus they incorporate the log file values with their academic performance values for identifying the unmotivated learners. The proposed work is an enhanced version of the research [24].

The methodologies followed by the [24] research is they classify the log file sequences into three categories namely learning, assessment and other sequences. Then minimum and maximum thresholding is calculated based on the data collected from [26]. If the time spent on a log file sequence is between minimum and maximum threshold means then that sequence is assigned as engaged, otherwise it has been assigned as disengaged. Likewise, each and every log file sequences are monitored. Finally, motivational level of the learner is decided based on the 2/3 status of the engagement status of the log file sequences. Although it is an automatic process of identifying the motivation level of the learners, it is considered as a highly time consuming. Instead of checking the entire log file sequences, in the proposed work we check only the overall values, so that system is faster and efficient than the previous approaches. For this we introduce two meta attributes namely Gindex and index. Average time spent on page is considered as Gindex. Average time spent by an individual learner is considered as index.

$$\sum_{i=m}^{n} ai = a_m + a_m + 1 + \ldots a_{m+n-1} + a_n$$

$$\sum_{i=1}^{n} ai = 66{,}13{,}426 \tag{1}$$

Where a = time spent on a page, n = total no of pages read,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} ai$$

$$Gindex(\mu = \frac{1}{311660} * 66{,}13{,}426 = 21.22 \tag{2}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu^{2} = 3.74} \tag{3}$$

$$\text{Minimum\_threshold} = \mu - \sigma = 17.48 \tag{4}$$

$$\text{Maximum\_threshold} = = \mu + \sigma = 24.96 \tag{5}$$

$$index = \frac{Total\ time\ spent\ on\ Learning}{Total\ no\ of\ Pages\ read} \tag{6}$$

Disengagement Construction and Prediction Algorithm (DCP) is used to predict the disengaged learners. DCP Algorithm suggests a new model when no benchmark index exists. Further if and index is found with higher level deviation then it is suggested to re-compute the index. The disengagement can be confirmed when the value of an index does not lie between minimum and maximum threshold values. Similarly, the disengagement can be predicted in connection with their academic performance. This shows more quality of prediction than the existing approaches. Through this proposed work, we suggest that dynamic evaluation of learner's assessment will helps to increase the motivational level of the learners.

Disengagement Construction and Prediction Algorithm(DCP)

Initialize log file sequences lf_1, lf_2, lf_3…… lf_n
Output: Preprocessed log file with engagement status
Step 1: Begin
Step 2: If Gindex = NULL Then
Step 3: Begin
Step 4: Calculate Gindex using Equ (2)
Step 5: Calculate Using Equ (3)
Step 6: Calculate Min_Threshold using Equ (4)
Step 7: Calculate Max_Threshold using Equ (5)
Step 8: End if
Step 9: For Each item in Student Database do
Step 10: Begin For
Step 11: Calculate index using Equ (6)
Step 12: If index >Min_threshold and index <Max_threshold and cp>app
Step 13: Assign Eng_Status='Engaged'
Step 14: Else
Step 15: Assign Eng_Status='Disengaged'
Step 16: End If
Step 17: End For
Step 18: End

Usually educational systems restrict some minimum marks to qualify the examination. If a learner gets the lower than the minimum marks means, then the learner is disqualifying in the examination and declared as fail. By disqualifying the learners in an examination will not motivates the learners at any circumstances. Instead of such mechanism, our proposed system evaluates individual criteria for every candidate. That is, if a learner's current performance is greater than his average previous performance then the learner motivational level is increased when compared to previous performances, so he will be considered as an engaged learner. Similarly, if the learner's current performance is lesser than the average previous performances then his concentration on learning and assessment is decreased, so the learner is considered as a disengaged learner.

4. EXPERIMENTAL RESULTS

For this experiment, we have considered the real time dataset, collected from Quasi framework [26]. 247 users are participated in this research. These users spent ten sessions for learning and attended ten assessments, where the activities between login and logout are considered as session. From this process, 7,90,859 instances have been obtained. The attributes used for this analysis are listed in Table-I

Table 1
Attributes used for Analysis

Code	*Attribute Description*
NoP	No of Pages read
Tot_TS	Total Time Spent for Learning
AvgTL	Average Time spent for learning
Index	Average Time Taken for a page
Cp	Current Performance
App	Average Past Performance

As mentioned in previous chapter, the sequencing approach followed by the researchers [21], [22], [23], [24], [25] is highly time consuming and the proposed indexing mechanism on log file analysis will evaluate faster than the previous approaches. DCP approach will evaluates 7,90,859 instances and identifies the engagement status of 247 learners in 1.08 seconds. The time complexity of the log file analysis is presented in Table-2.

The rest of the paper we discuss about the need for Dynamic evaluation method on assessment and finally we check classificationaccuracy results with WEKA tool.

Out of ten assessments attended by the learners. First 5 assessments are based on the traditional assessment methods and remaining 5 are based on our proposed methodology. Our assessment database contains 6000 question and answers. The learners do not get the same set of questions. Questions are generated on randomized manner. If a learner attends one question means, then it is marked as visited and the same question will not repeat until he attends the remaining 5999 questions.In a traditional assessment method, the learner has to get minimum 50% of correct answers or else the learner is considered as disengaged. In a proposed method, if the learner's current performance is lesser than his average previous performance then the learner is considered as disengaged.

Table-3 shows the comparison analysis of two assessment methods. While following the traditional based assessment method 149 learners are failed in assessments,34 learners are failed in learning and 64 learners are passed in both learning and assessment. While in our proposed assessment method we found only 107 learners are failed in assessments and 34 are failed in learning and 106 learners got passed in both learning and assessment. When compared to the traditional assessment method 42 learners are motivated and get more marks than their previous assessments. Similarly, Figure-1 shows the visual representation of Comparison Analysis.

Table 2
Time Complexity

	Execution time in minutes
Sequence approach on log file analysis	4.31
Proposed indexing mechanism.	1.08

Table 3
Comparison Analysis of Assessment Methods

	Traditional Approach	*Dynamic evaluation*
No of Participants	247	247
No of Assessments	5	5
No of Passed Candidates	64	106
No of Failed Candidates	149	107

Figure 1: Comparison Analysis of Assessment Methods

Table 4
Experimental Results

Dataset	*DCP Algorithm*
%correct	94.33
TP Rate	0.936
FP Rate	0.05
Precision	0.964
Error	0.06

Table 5
Confusion Matrix

Engagement Status	*Disengaged*	*Engaged*
Disengaged	132	9
Engaged	5	101

To check the accuracy level of our prediction values, two indicators are mainly considered: Accuracy and True positive rate, where accuracy is calculated as the percentage of correctly classified instances divided by total number of instances. True positive for disengaged class is an indication of the correct identification of disengaged learners, similarly other indicators such as false positive rate, precision, error are calculated. The experimental results obtained based on the DCP Algorithm is presented in Table-4.

The confusion matrix of DCP Algorithm is presented in Table-5.

Hence the proposed work proves that indexing mechanism on log file analysis will executes faster than the previous sequence approach on log file analysis. in addition to this classification accuracy is also effective than the previous approaches.

5. CONCLUSION AND FUTURE WORK

Through this study, we deeply discuss about two issues on detecting the disengaged learners. The first issue is related to traditional approach on log file analysis. Instead of taking all log file sequences, DCP algorithm creates an indexing attributes based on the overall values. Our experimental results show that the DCP algorithm will executes faster than the traditional log file approaches. The second issue is related to the drawbacks in traditional assessment methodology. Through this study, we suggest the dynamic evaluation method for learners in their assessments. Due to this approach, the more disengaged learners are motivated and steps towards engagement. In future perspective, the learners may classify based on the new factors like area of residence, economic status, previous educational qualification, etc., etc., thus finding the individual threshold values based on their classification may detects the disengaged learners more accurately.

REFERENCES

[1] Ishii, T., Saitou, M., Hiramoto, S.: An instructional design for developing learning contents by subject matter experts. In: *Proceedings of the 9th World Conference on Continuing Engineering Education,* pp. 469–474, 2004.

[2] Connolly, T., Stansfield, M.: Using games-based e-Learning technologies in overcoming difficulties in teaching information systems. *J InfTechnolEduc***5**, 459–476 (2006)

[3] Martyn, M.: Clickers in the classroom: an active learning approach. *EDUCAUSE Q***30**(2), 71–74 (2007)

[4] Beal, C.R., Lee, H.: Creating a pedagogical model that uses student self reports of motivation and mood to adapt ITS instruction. In: *Proceedings of theWorkshop on Emotion and Motivation in Educational Software (EMES).* IOS Press, Amsterdam. Retrieved on 10 June 2006 from http://k12.usc.edu/Pubs/ (2005)

[5] Beal, C.R., Qu, L., Lee, H.: Classifying learner engagement through integration of multiple data sources. In: *Proceedings of the 21st National Conference on Artificial Intelligence*. AAAI Press, Menlo Park. Retrieved on 10 June 2006 from http://k12.usc.edu/Pubs/ (2006)

[6] G.D. Chen, G.Y. Shen, K.L. Ou, and B. Liu, "Promoting Motivation and Eliminating Disorientation for Web Based Courses by a Multi- User Game," *Proc. World Conf. Educational Multimedia and Hypermedia and World Conf. Educational Telecomm.*, June 1998

[7] Christenson, S., Reschly, A. L., & Wylie, C. *Handbook of research on student engagement.* New York, NY: Springer.2012.

[8] Skinner, E. A., Kindermann, T. A., &Furrer, C. A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, **69**, 493-525,2009.

[9] San Pedro, M. O. Z., Baker, R. S., Bowers, A. J., & Heffernan, N. T.. Predicting college enrollment from student interaction with an intelligent tutoring system inRunning head: middle school. *In Proceedings of the 6th international conference on educational data mining* (pp. 177-184). 2013.

[10] J. Johns and B. Woolf, "A Dynamic Mixture Model to Detect Student Motivation and Proficiency," *Proc. 21st Nat'l Conf. Artificial Intelligence* (AAAI-06), 2006.

[11] J. Walonoski and N.T. Heffernan, "Detection and Analysis of Off- Task Gaming Behavior in Intelligent Tutoring Systems," *Proc. Eighth Int'l Conf. Intelligent Tutoring Systems*, M. Ikeda, K. Ashley, and T.-W. Chan, eds., pp. 382-391, 2006.

[12] J. Walonoski and N.T. Heffernan, "Prevention of Off-Task Gaming Behaviour within Intelligent Tutoring Systems," *Proc. Eighth Int'l Conf. Intelligent Tutoring Systems*, M. Ikeda, K. Ashley, and T.-W. Chan, eds., pp. 722-724, 2006.

[13] Buckley, B.C., Gobert, J., Horwitz, P. &O'Dwyer, L. Looking inside the black box: Assessments and decision-making in BioLogica. *International Journal of Learning Technologies,* **5(2),** 166 – 190, 2010.

[14] Baker, R. S., Corbett, A., Koedinger, K., & Wagner, A. Off-Task Behaviour in the Cognitive Tutor Classroom: When Students "Game the System." *Proceedings of ACM CHI 2004: Computer-Human Interaction,* 383-390, 2004.

[15] Pardos, Z.A., Baker, R.S.J.d., San Pedro, M.O.C.Z., Gowda, S.M., Gowda, S.M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge,* 117-124, 2013.

[16] Cocea, M., Hershkovitz, A., & Baker, R.S. The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, 507-514,2009.

[17] Baker, R. S.Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems. *Proceedings of ACM CHI*: Computer-Human Interaction, 1059-1068, 2007.

[18] Tytler, R., & Osborne, J. Student attitudes and aspirations towards science. In *Second International Handbook of Science Education* (pp. 597-625). Springer Netherlands,2012.

[19] Cetintas S., Luo Si, Yan Ping Xin, and C. Hord. Automatic detection of o-task behaviors in intelligent tutoring systems with machine learning techniques. *Learning Technologies, IEEE Transactions* on, **3(3):**228-236, July 2010.

[20] Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, **70(5):**377 - 398, 2012.

[21] Sundar PVP. Quasi Framework: A new student disengagement detection in online learning. *International Journal of Engineering Research and Technology (IJERT).* Dec;**1(10):**1–6. 2012

[22] Sundar PVP, Kumar AVS. Disengagement detection in online learning using Quasi framework. *IOSR Journal of Engineering (IOSRJEN)*.**5(1):**4–9.2015.

[23] Sundar PVP, Kumar AVS. An enhanced disengagement detection in online learning using Quasi framework. *Special Edition of International Journal of Applied Engineering Research (IJAER).* Jun; **10(55):**1298–302.2015

[24] Sundar PVP, Kumar AVS. A Systematic approach to identify the unmotivated learners in online learning. *Indian journal of science and technology*. April; Vol **9(14):** 1-6.2016.

[25] Sundar PVP,Kumar AVS. A Hybrid classification method for disengagement detection in online learning, *International Journal of Education and Information Studies.* **5(1)**:67–74.2015

[26] Quasi Framework. Available from: www.quasiframework.com