Murthy, S. et al. (Eds.) (2018). Doctoral Student Consortium Proceedings. 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education.

# Predicting Engagement Using Machine Learning Techniques

**Vandana Naik[a*] & Venkatesh Kamat[b]**
[a]*Research Scholar, DCST, Goa University, India*
[b]*Professor, DCST, Goa University, India*
*dcst.vandana@unigoa.ac.in

**Abstract:** A learner who is sufficiently involved in the learning process will surely show higher performance with respect to the objectives set for the process. With detailed planning of every miniscule unit of learning in place, focus is entirely on the learner's engagement for successful learning to happen. Engagement prediction has been investigated in various ways not only in education but in other domains also. Most of them that involve machine learning techniques either use invasive methods or narrowly focus on only one aspect of the learning environment particularly if they are using log based measurements. The aim of this work is to propose and validate a framework that will be suitable for detecting engagement at a unit level of sessions in an online environment. Using machine learning techniques, an engagement detector will then be created that will track the tasks under each of the component of this framework and depending upon depth of involvement would predict the learner's engagement level. It will not require any special wearable or external instruments like cameras to capture the data used for the detection. It will use only the logs from the Moodle platform for the activities set up by the faculty for the session and track the activities of the learner for them to calculate the engagement level. This calculated index of engagement can then help the instructors to take corrective actions so that the learner's involvement can be increased.

**Keywords:** Engagement, Engagement Parameter Framework

## 1. Introduction

Engagement is one of the most fundamental aspects that determine the performance of an individual in the learning process. It is considered as a major factor that can control high levels of student boredom and disaffection, and high dropout rates (Fredricks, Blumenfeld, & Paris, 2004). Student engagement refers to the degree of attention, curiosity, interest, optimism, and passion that students show when they are learning or being taught, which extends to the level of motivation they have to learn and their progress (Great Schools Partnership, n.d.). Student engagement is important to study for three reasons (Skinner & Pitzer, 2012): First, it is a necessary condition for students to learn. Second, engagement decides student's everyday experiences in school, both psychologically and socially. Third, engagement is a critical contributor to student's long term academic development. The crux of this study is therefore to perform learning analytics on the learner's trail over the Moodle course to track the level of engagement. Engagement measurement has been attempted in several ways. Self-reporting is the most widely used, wherein questionnaires are administered and the students themselves have to choose from options provided or answer to the questions that try to judge the engagement level. Some other quite common methods similar to these include teacher rating or field observations etc. (Fredricks & McColskey, 2012). These methods suffer from restrictions like biasing, false reporting and being un-scalable to use in every learning environment. The other type of detection technique involves tracking through external devices like webcam that can be used for eye tracking or for capturing face, body posture and hand gestures (Grafsgaard, Wiggins, Boyer, Wiebe, & Lester, 2013). These require hardware that needs to continuously run to track the activities and need the learner to be screen tied use a number of assumptions that may not be always true (Miller, 2015). Another type of external device controlled technique includes physiological sensor that capture features like EEG, blood pressure, heart rate for prediction (Chaouachi, Chalfoun, Jraidi, & Frasson, 2010). A more recent method is of automatic inference of the engagement level through the logs of the learner created in online environment (Gobert, Baker, & Wixon, 2015; Ramesh, Goldwasser, Huang, Daume, & Getoor, 2014). The current work also falls in this category of research. The first problem that is seen during the investigation in the literature of these studies is that most of them lack theoretical background and are either based on intuition or on patterns that emerge from data. Next these are found to be restrictive in terms of either the features used for detection or the domain in which they are used. In some of them, though they claim to capture wide range of activities, finally only the time aspect of these activities is used (Gobert et al., 2015). Keeping all these issues in mind, first a conceptual framework is derived based on the theories of engagement. Next features under each of the components of this framework that can help to identify the engagement level are developed. These features are not restrictive and would be coming from all the activities setup for the session. It is proposed that this engagement level is tested by comparing with human observers that would code the logs of each learner to establish ground labels and then used for supervised learning.

## 2. Related Work

Engagement detection using logs of Moodle platform has not been studied widely. One of the studies, which this work relates to, is described in (Liu, Richards, Dawson, Froissard, & Atif, 2016). In this the authors have used an algorithm to first determine optimal parameters of engagement, restricted to only three indicators namely assessment, forum, and login. The algorithm calculates the parameters and weightings, starting with some initial guess that is improved by maximizing the inverse correlation between total risk rating and final course grade for each student. These algorithm derived ratings were compared with the ones calculated by teachers based on their experience by conceptualizing what they expected of a good student. However the current work differs with this work in two aspects. Firstly, there is no constraint on the indicators; all activities set up by the instructor will be used. Secondly, there will be no human intervention in these predictions as the job of the instructor will be only to take corrective action after the prediction is over. Another recent related work is that of Analytics Moodle API (Dalton & Monllaó, n.d.) that comes bundled in Moodle 3.5. This is very close to the proposed research work as this API also allows selecting indicators for required target and uses analytics to display the results for the course for each learner. It then suggests a few corrective actions to the instructor to help the learner to overcome their specific difficulty. One of the limitations that has not yet been taken care of is that engagement during session is not targeted in this version although it is flexible enough and can be set by the Moodle site manager but requires good programming skills for the same. Other limitation is that it allows only binary classification. Although the current work like this API proposes to have a specific target that is engagement level, it will allow a broader range of indicators to be included. Also it will be using multiclass classification. Another difference that arises at this moment of time is that the Analytics API currently defines target for overall course, however the proposed research is aimed at lower unit level of learning i.e. engagement in a session. Nonetheless this work can be viewed as extension over the Analytics API as it will serve as the basis by providing the interface for data collection and analysis to some extent.

## 3. Proposed Research

### 3.1 *Engagement Parameter Framework*

The current research work aims to propose and validate a framework referred to as Engagement Parameter Framework (EPF) shown in fig 1. The sessions will constitute of tasks that

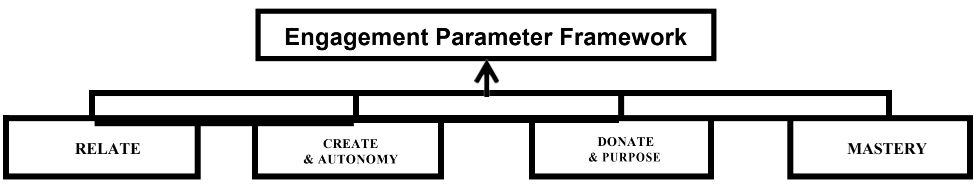


*Figure 1.* Engagement Parameter Framework

will incorporate the features from each of the components of the EPF that will play the role of indicators for the detection. These indicators revolve around the four main components, namely a) Relate, b) Create and Autonomy, c) Donate and Purpose, d) Mastery. This proposed framework takes into consideration three principles from engagement theory (Kearsley & Shneiderman, 1998) and the three motivational factors (Pink, 2011). Its components are then formed by merging those that show significant overlap whilst retaining intact those that do not have anything in common. First component combines create and autonomy. Create as defined in the engagement theory is defined as the freedom given to learner to come up with some realistic problem and apply their ideas to solve the problem. The autonomy as a motivational factor also indirectly implies the same, where the onus of shaping the learning activity to fulfill the learning objective is passed onto the learner and refers to higher level of Bloom's taxonomy (Bloom & Others, 1956). Several tools of Moodle like simulators, virtual labs, choice, and database help to achieve these. The next component that is formed from amalgamation of two concepts is donate and purpose. Donate principle of the engagement theory focuses on the value of making a useful contribution while learning. The purpose motivational factor can be extended to learning context in a similar way and would mean framing activities that would involve some pro-social tasks which can motivate the learner to contribute towards greater cause (David S. Yeager et al., 2014). The features like Forum, Q&A, chat etc. can be used for this constituent. The Relate component added to the framework is derived from the engagement theory. It is expected that tasks that demand collaboration can inspire the learners to participate with higher interest (Ryu & Lombardi, 2015; Vygotsky, 1980). Some of the instructional tools from Moodle that allow collaborations may include forums, emails, chat etc. Lastly the fourth component, Mastery is the motivational factor that will help to increase the engagement level by devising activities that will push the learner towards the learning objective. Such activities need to provide space for the learner to practice several times, play around the system, review their technique and then improve in each round. A learner who is passionate about improving his skills will unquestionably be engaged in the learning environment. To provide such opportunities, progress bar, badges, leaderboard etc. can be used. The current research work has reached till the stage of devising of this framework and the next steps, yet to be taken are discussed in the sections.

### *3.2 Experiment and validations:*

The process of designing an engagement detector begins with categorizing the activities that can help detect engagement into the four components. This will be done through literature study and by asking experienced faculty through questionnaire. Next for each indicator, value assignment will be designed. The value will be assigned in two ways. For each learner each time any activity is accessed will result in assignment of +1 value for that activity for that learner. Next within that activity as the learner accesses further features this value will get incremented. For example, accessing a forum by learner will result in first value assignment while in forum if the learner continues to use other features like posting query, replying, up voting or down voting will result in increasing the initial value each time. After this phase, planning of sessions on Moodle by picking up appropriate activities will be done. The sessions will be conducted and the first round of dataset consisting of values for each indicator for each learner will be collected. This dataset will be subjected to unsupervised algorithms like Expectation–Maximization (EM) Clustering using Gaussian Mixture Models (GMM), Agglomerative Hierarchical Clustering, etc. to identify strong indicators. The data clips from the same dataset will be used to code the strong indicators by human observers. The comparison between the human coded and clustering algorithm indicators will help to identify the best features that can help to detect the engagement level. The model of indicators and engagement as target will be embedded in the Analytics API and checked for accuracy. The important aspects that will be investigated are whether adding of this framework to categorize the activities is helpful or not by comparing Moodle sessions with and without this framework and checking the relation between engagement level detected and performance of the students over the session objectives. If strong correlations are established between the indicators and the engagement levels then the model would be promoted using supervised algorithms for future predictions. Response time within the lowest level of activities that serve as indicator could be used for further analysis of engagement detection. It is also intended to identify the best machine learning algorithms for such learning analytics.

## 4. Conclusion

This research work offers to establish engagement detection techniques based on sound theory that could be used by the instructors to monitor the engagement of learners and provide them with necessary help during the session. It recommends using well established factors that help to increase the involvement of the learners and thereby improve their performance. The aim is to investigate the construct and identify the best indicators to help detect it. The study will help to identify relation between the logs of the learner's activities and their engagement. It will also bring out the best suited machine learning technique for learning analytics.


## Acknowledgements

This publication is an outcome of the R&D work undertaken project under the Visvesvaraya PhD Scheme of Ministry of Electronics & Information Technology, Government of India, being implemented by Digital India Corporation.


## References


Bloom, B. S., & Others. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay*, 20–24.

Chaouachi, M., Chalfoun, P., Jraidi, I., & Frasson, C. (2010). Affect and mental engagement: towards adaptability for intelligent systems. In *23rd International FLAIRS Conference*. Retrieved from https://www.iro.umontreal.ca/~frasson/FrassonPub/FLAIRS-2010-Chaouachi-et-al.pdf

Dalton, E., & Monllaó, D. (n.d.). Analytics - MoodleDocs. Retrieved August 16, 2018, from https://docs.moodle.org/35/en/Analytics

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109.

Fredricks, J. A., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 763–782). Boston, MA: Springer US.

Gobert, J. D., Baker, R. S., & Wixon, M. B. (2015). Operationalizing and Detecting Disengagement Within Online Science Microworlds. *Educational Psychologist*, *50*(1), 43–57.

Grafsgaard, J., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. (2013). Automatically recognizing facial expression: Predicting engagement and frustration. In *Educational Data Mining 2013*. Retrieved from https://cise.ufl.edu/research/learndialogue/pdf/LearnDialogue-Grafsgaard-EDM-2013.pdf

Great Schools Partnership. (n.d.). The Glossary of Education Reform -. Retrieved August 16, 2018, from https://www.edglossary.org/

Kearsley, G., & Shneiderman, B. (1998). Engagement Theory: A Framework for Technology-Based Teaching and Learning. *Educational Technology Research and Development: ETR & D*, *38*(5), 20–23.

Liu, D. Y. T., Richards, D., Dawson, P., Froissard, J.-C., & Atif, A. (2016). Knowledge Acquisition for Learning Analytics: Comparing Teacher-Derived, Algorithm-Derived, and Hybrid Models in the Moodle Engagement Analytics Plugin. In *Knowledge Management and Acquisition for Intelligent Systems* (pp. 183–197). Springer International Publishing.

Miller, B. W. (2015). Using Reading Times and Eye-Movements to Measure Cognitive Engagement. *Educational Psychologist*, *50*(1), 31–42.

Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*. Penguin.

Ramesh, A., Goldwasser, D., Huang, B., Daume, H., III, & Getoor, L. (2014). Uncovering hidden engagement patterns for predicting learner performance in MOOCs. In *Proceedings of the first ACM conference on Learning @ scale conference* (pp. 157–158). ACM.

Ryu, S., & Lombardi, D. (2015). Coding Classroom Interactions for Collective and Individual Engagement. *Educational Psychologist*, *50*(1), 70–83.

Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In *Handbook of research on student engagement* (pp. 21–44). Springer.

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Yeager, D. S., Henderson, M. D., Paunesku, D., Walton, G. M., D'Mello, S., Spitzer, B. J., & Duckworth, A. L. (2014). Boring but important: a self-transcendent purpose for learning fosters academic self-regulation. *Journal of Personality and Social Psychology*, *107*(4), 559–580.