



Determining the Co-relation between Behavioral Engagement and Performance of e-Learners

Dias B.A.M.T., Malida K.K.D.S., Sahani M.A., Jayathilaka J.M.S.C., T. C Sandanayaka, G.T.I Karunarathne
Faculty of Information Technology, University of Moratuwa,
Moratuwa, Sri Lanka

Abstract: The main focus of the research is to identify the co-relation between performance and behavioral engagement of learners in the e-learning environment. Two experiments were conducted in controlled environment to gather data using behavioral tracking software. Data analysis is done by using both statistical approach and data mining approach to develop the final model. Set of specialized software was used to preprocess the dataset. The developed model can be used to predict the learner's performance level based on identified independent parameters.

Key words: Co-relation, Data-mining, Behavioral Engagement

I. Introduction

E-Learning is defined as all forms of electronic supported learning and teaching, which are procedural in character and aim to effect the construction of knowledge with reference to individual experience, practice and knowledge of the learner. Information and communication systems, whether networked or not, serve as specific media to implement the learning process. [1] As a highly dynamic application area in modern computing, e-Learning promises to play a crucial role in transforming traditional education.[2] The knowledge transfer process within the context of technology-based teaching and learning environments can be interpreted as a holistic phenomenon composed of two related streams: the teaching process and the learning process. [3] Furthermore, e-learning embraces more than simply reading online lessons. It is a large and complex field of research encompassing a variety of learning and teaching paradigms, for example: constructivistic, serial, symmetric, cognitive, face-to-face, discovery, managed learning [4]. Though early approaches merely focused on the student's cognitive processes, this has changed with new research emphasizing the role of user behavioral engagement in creating a productive and enjoyable learning process.[5] The goal of the project is to identify the co-relation between learner performance and learner behavioral engagement by tracking e-learner's different behaviors together with performance. E-learner performance level has been determined by using the grade the e-learner has achieved at the end of the learning activity and that is used as the dependent variable in the relationship identification. Selected learner behavioral engagement parameters are eye focus area on the screen, cursor points with time, time spent on the lesson material, time spent on other websites (social media), keyboard input of learner. The developed background software applications were used to record data relevant to the learner interactions with the lesson. Statistical analysis and data mining techniques were used to analyze the dataset and develop the model. As the final outcome, a model is developed to automatically predict the performance level of the learner by capturing, processing and analyzing data relevant to learner's behavioral engagement.

II. Research Methodology

The research project is conducted under the assumption that there is a relationship between the e-learner behavioral engagement and the e-learner performance during an e-learning activity.

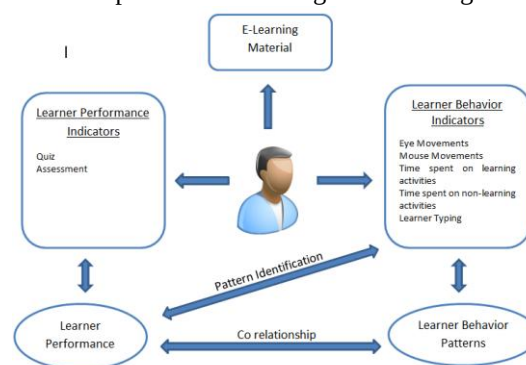


Figure 2.1: High Level Architecture of the System

The research was conducted under four main phases. First phase is developing data capturing methodologies and software applications. Four basic software applications were developed to track behavioral engagement of the learner and finally all those separate software applications were integrated in to one system. Second phase is gathering primary data on learner behavior and performance by conducting experiment. Two experiments were conducted for data gathering. Round 01 data gathering was conducted to 177 students in a controlled lab environment.



Figure 2.2: Round 01 Data Collection

The experiment of round two data gathering was conducted for 53 undergraduate students at the same lab environment as round one data gathering. Phase three is data analysis and co-relationship identification. Data analysis is done by using statistical approach and data mining techniques. Statistical methodologies such as ‘Correlation Calculation’ and ‘Linear Regression Line Calculation’ were applied to the data set to identify the relationship between the independent variables and dependant variables. Data mining techniques also used to map the behavioral data with performance data. Evaluation is the fourth phase of the project. The evaluation was conducted in two main ways such as evaluation of the tools developed for capturing data and evaluation of the final model developed by analyzing learner behavior and performance data.

III. Implementation

(A) Eye Movement Capturing Software

OpenCv library and C++ was used to develop the application. The face of the user is detected by the developed eye tracking application and eyes are detected after that. After detecting the eyes, the area of the each eye is divided in to two parts vertically. The sclera percentage of each eye is detected and saved to database at each frame in which the eye is detected.

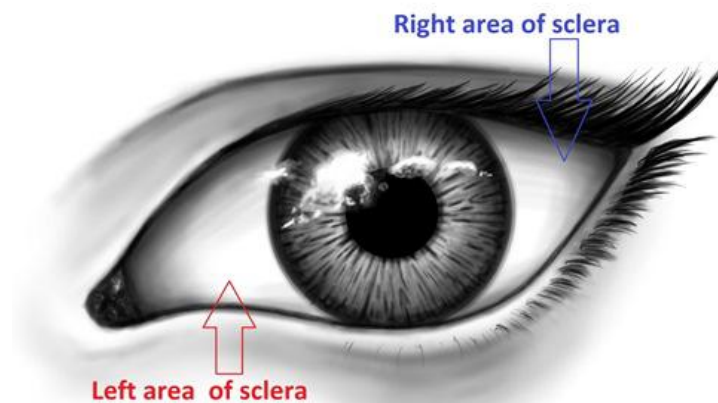


Figure 3.1: Areas used for Image Processing

At the beginning, a calibration page is presented to the user and the user is asked to focus eyes on the left corner and the right corner of the screen. By that the maximum and minimum values for sclera area are identified. The computer screen is divided into three equal vertical areas (A, B, C) and using the minimum and maximum calibration values, the area on which the user is focusing his/her eyes at each frame is predicted using the following algorithm:

$$Area = \frac{w}{(a_2 - a_1)} A$$

Where:

w = screen width (in pixels)

a_1 = minimum area of sclera at calibration
(at left side)

a_2 = Maximum area of sclera at calibration
(at left side)

A = area of the sclera at a movement

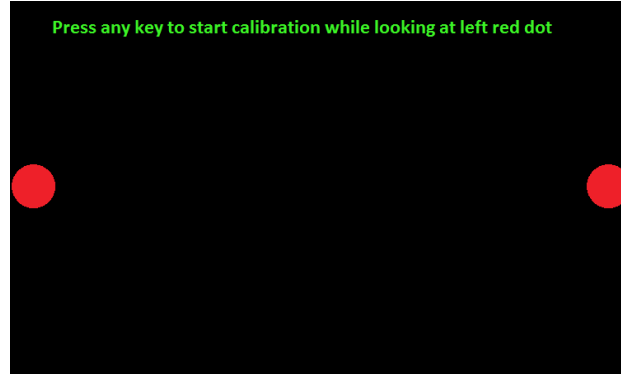


Figure 3.2: Eye Calibration Points

(B) Mouse Movement Capturing Software and Data Preprocessing

The X, Y coordination of the cursor in the activity window with time (during each hundred millisecond) is captured by a Java program. The data captured is written to a text file automatically and the text file is created in C drive of the user's computer. The data set cannot be directly applied for analysis process since it is not compatible with the type of other data. Hence the variance of mouse movements with respect to time was used for the analysis process. Mouse movement variance was calculated by using following algorithm.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n (z_i - \bar{z})$$

Where:

$$\bar{z} = \frac{1}{N} \sum_{i=1}^n \sqrt{(x_{i2} - x_{i1})^2 + (y_{i2} - y_{i1})^2} \quad z_i = \sqrt{(x_{i2} - x_{i1})^2 + (y_{i2} - y_{i1})^2}$$

Where:

Z = Length of one movement

X2 = Second x coordination of the screen

X1 = First x coordination of the screen

Y2 = Second y coordination of the screen

Y1 = First y coordination of the screen

N = number of movements

(C) Keyboard Tracking Software and Data Preprocessing

In this project, a key logger was developed which quietly monitors keyboard actions to log any key press the user makes. The program stores the key board input in a log file and sends the input to the database. The program is written operating system-specific hence at this stage the program for keyboard capturing is written in c# for windows OS. The input keyboard data is saved in a text file. The text file is then read by a java program and checked against a string of pre-defined keywords from the lesson. Keyboard input data set was preprocessed to get the number of keyword searched by the learner. The count of the keywords typed by the user is calculated and sent to the database as the input parameter.

(D) Web Extension to track the sites visited and respective time on each site and Data Preprocessing

An extension for Google Chrome browser is developed using Javascript. Extensions are extra features and functionality that can be added to Google Chrome. By using the extension, Google Chrome is customized to capture the switching time and the URL between switching materials while engaging with the e-learning activity. Therefore the software records all the URLs visited by the learner and the time spent on each URL. Based on the URLs and the respective time spent on each URL; the time spent on the learning activity was taken as a percentage of the total time spent as an input parameter. In the same way, the time spent on non-learning

URLs was taken as a percentage of the total time as an input parameter. The ten most visited, social media and entertainment related sites by Sri Lankans were defined as the non-learning URLs in the calculation process.

(E) Student marks prediction software

Finally a software application was developed to predict the learner performances by getting the details of independent parameters based on the statistical model which is the final outcome of the statistical data analysis. The application was developed by using java programming language.



Site	Minutes	Percent
Total	1.65	100
☐ http://moodle.itfac.mrt.ac.lk	0.95	57.96
☐ chrome://newtab	0.31	19.00
☐ https://www.facebook.com	0.20	11.98
☐ http://stackoverflow.com	0.15	8.91
☐ chrome://extensions	0.02	1.38
☐ http://www.espninfo.com	0.01	0.77

Figure 3.3: Web Extension

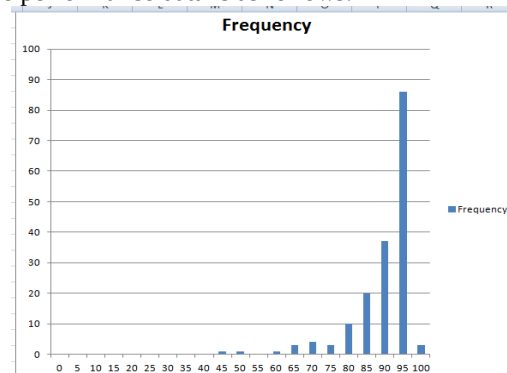
IV. Data Analysis

(A) Statistical Data Analysis

Statistic analysis is done to determine whether there is a relationship among the obtained data. Correlation calculation and Multiple Regression Line calculation was applied to the data set during statistical data analysis using “Data Analysis Plug-in” in Microsoft Excel

(B) Round one data collection

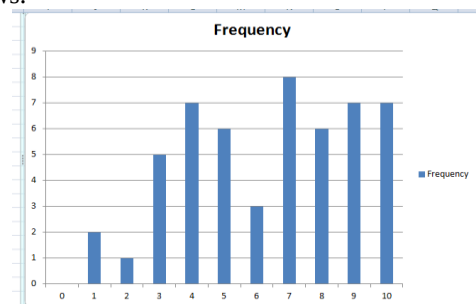
As the majority of marks are ranked in the 80-100 range, the marks do not have a normal distribution and the frequency of marks of round one performance data is as follows:



X axis denotes the marks and Y axis denotes the Number of students who obtained the relevant marks.

(C) Round two data collection

The e-learning experiment to track behavioral and performance data was conducted on 50 level 04 undergraduates. This experiment was conducted avoiding the practical and technical issues in round one data collection and was based on a more common, historical subject. The performance data are distributed more naturally in second round data set. Hence the data collected in the round two is used for advance data analysis to develop a model which is used to predict the learner performance level. Frequency of the student marks of second round data set is as follows:



X axis denotes the marks and Y axis denotes the Number of students who obtained the relevant marks.

The following co-relationships were identified during correlation analysis.

1. Co-relation Coefficient between Eye Movements and Marks is 0.2400
2. Co-relation Coefficient between Mouse Movements (variance) and Marks is 0.3436
3. Co-relation Coefficient between Social Media Engagement (as a percentage from time spent on the whole activity) and Marks is 0.6713
4. Co-relation Coefficient between number of keywords typed and Marks is 0.4852

Multiple linear regression methodology was used to model the relationship between dependent parameter and several independent variables. y denoted Student Marks and x denoted independent variables such as eye focus points on the screen, cursor points with time, time spent on the lesson material, time spent on other websites(social media), keyboard input of the learner.

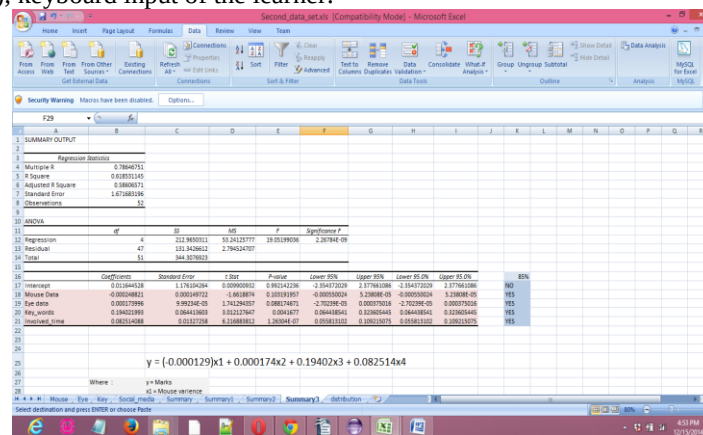


Figure 4.1: Multiple Linear Regression Model Calculation

The Multiple Linear Regression line was calculated using Microsoft Excel for the round 02 data by considering 85% of accurate level. The final model developed based on statistical analysis is as follows:

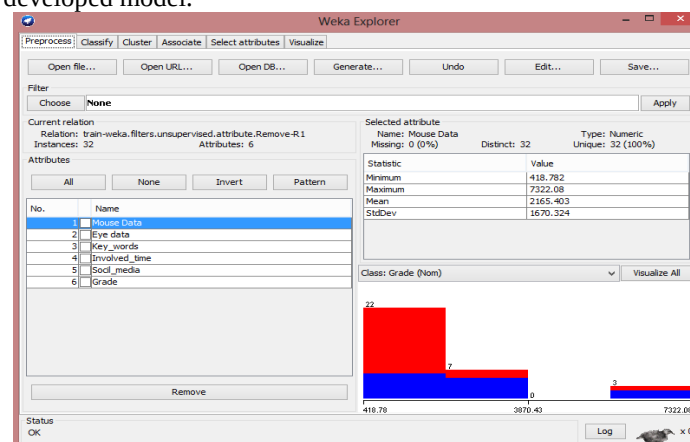
$$y = (-0.000129) x_1 + 0.000174x_2 + 0.19402x_3 + 0.082514x_4$$

Where:

- y = Marks
- x_1 = Mouse variance
- x_2 = Eye focus area
- x_3 = No of key words
- x_4 = Involve Time

(D) Data Mining

Data mining techniques were used for further analysis of the data set and validate the model. Weka 3.6 was used for mining purposes. The collected data are used as the training set on which the model is built using J48 classifier. A new dataset is used to validate the model and the grade of the learners of the new dataset is predicted based on the developed model.



The developed system is capable to predict the learner performance level based on the developed model by tracking the behavioral engagement of the student in the e-learning environment.

(E) Evaluation

Evaluation methods were chosen based on the functionality and the nature of the modules of the system. The developed software was evaluated using actual e-learners and by doing cross validation and backward validations.

E1. Evaluation of the model

The developed model is tested by using ten unused data elements from the round two data collection. To evaluate the accuracy of the model, the data regarding eye focus point, mouse movement variation, searched text and visited sites of e-learners during the e-learning activity were applied to the model developed, and thereby based on those behavioral data the marks of each learner was predicted by using the model. Accuracy level of the model was 85%.

E2. Evaluation of the software**E2.1. Evaluation of the eye tracker**

If the eye tracker works accurately, it should provide exact area (A,B,C) when a learner focuses his/her eyes at some area of the computer screen. Hence a set of ten eye focus instances were used as the inputs for the evaluation. The test users were asked to focus their eyes within the A,B,C areas and the values recorded as the output by the eye tracker were checked against the actual eye focus areas of the three points on the screen.

No. of Instances	Actual eye focused area on the screen	Correctly predicted no. of instances	Incorrectly predicted no. of instances
10	A	6	4
10	B	8	2
10	C	5	5

Table 5.1: Test case for the “eye tracker”**E2.2 Evaluation of the Mouse Tracker**

If the mouse tracker works accurately, it should provide exact x, y coordination when a learner points the cursor at some point of the computer screen. Hence a set of x, y coordination and user cursor points were used as the inputs for the evaluation. Therefore four corner points on the screen were used of which the x, y coordination is known and pre-defined. The test users were asked to point the cursor upon the four corners of the screen (upper-left, upper right, lower right, and lower left) and the x, y coordination recorded as the output by the mouse tracker were checked against the actual x, y coordination of the four corners of the screen.

Cursor Point on the screen (Input)	Expected x, y coordination	Output by the mouse tracker
upper- left	0,0	0,0
upper right	1366,0	1366,0
lower right	1366,768	1366,768
lower left	0,768	0,768

Table 5.2: Test case for the “mouse tracker”**E2.3 Evaluation of the Keyboard Tracker**

If the keyboard tracker works accurately, it should record exact words when a learner types the particular words using the keyboard. Hence a pre-defined words were used as the inputs for the evaluation. The test users were asked to type those words while the software runs as a background application and the set of words recorded as the output by the keyboard tracker were checked against the actual words.

Expected Outcome	Actual Outcome
Sahani Matharaarachchi	Sahani Matharaarachchi
Madushi Dias	Madushi Dias
Malinda Kandalama	Malinda Kandalama
Sadun Jayasekara	Sadun Jayasekara

Table 5.3. Test case for the “keyboard tracker”**E2.4 Evaluation of the web extension**

If the web extension works accurately, it should record the visited web URLs and time spent on each URL accurately when a learner visits various websites during the e-learning activity. Therefore four most commonly used websites were used for the evaluation. The test users were asked to visit these websites and stay on each website for two minute duration while the web extension runs as a background application and the URLs and time spent on each URL recorded as the output by the web extension were checked against the actual URLs and the time spent on them.

The URLs visited	The Approximate Time Spent	Recorded URLS by the web extension	Recorded Time Duration
https://www.facebook.com/	2 minutes	https://www.facebook.com/	2.02 minutes
http://en.wikipedia.org/	2 minutes	http://en.wikipedia.org/	2.00 minutes
https://www.youtube.com/	2 minutes	https://www.youtube.com/	1.99 minutes
http://moodle.itfac.mrt.ac.lk/	2 minutes	http://moodle.itfac.mrt.ac.lk/	2.03 minutes

Table 5.4 Test case for the “web extension”**E2.5 Evaluation of the Developed Model**

After determining the co-relationships and developing the relationship model, 10 unused datasets from the Round 02 Data Collection were used to validate the relationships identified by applying to the model. To evaluate the accuracy of the model, the data regarding eye movement variation, mouse movement variation, searched text and visited sites of e-learners during the e-learning activity were applied to the model developed, and thereby based on those behavioral data the marks of each learner was predicted by using the model.

VI. Further Work

This knowledge obtained from this research project and the relationships and sub relationships identified by pattern recognition process can be used in the future researches to enhance learner performances, enhance the quality and design of e-learning materials, decision making on student's overall performance and knowledge gathering and also this research results will be able to be used for future researches and implementations in e-learning applications.

VII. Conclusion

By considering the developed model, it can be concluded that the mouse movement variance negatively affects to the marks achieved by the learner, Eye focus pattern of the learner positively affects the marks achieved, number of keywords searched is positively affects the marks achieved, time involved with the learning activity is positively and strongly affects to the marks achieved by the e-learner. Compared to other attributes; the time spent on non-learning activities have very weak co-relation with the marks achieved. Hence it does not have considerable effect for the marks of the student, it was rejected from the Multiple Linear Regression Model.

References

1. Dr.P.Nagarajan and Dr.G.Wiselin Jiji, "ONLINE EDUCATIONAL SYSTEM (e- learning)" in *International Journal of u- and e- Service, Science and Technology*, Vol. 3, No. 4, December, 2010
2. Johan Ismail, "The design of an e-learning system Beyond the hype", in *Internet and Higher Education* 4 (2002) 329–336
3. Victor Manuel García Barrios, Christian Gütl, Alexandra M. Preis, Keith Andrews, Maja Pivec, Felix Mödritscher and Christian Trummer, "AdELE: A Framework for Adaptive E-Learning through Eye Tracking", *Proceedings of I-KNOW '04 Graz*, Austria, June 30 - July 2, 2004
4. Jennifer Lennon and Hermann Maurer, "Why it is Difficult to Introduce e-Learning into Schools And Some New Solutions" in *Journal of Universal Computer Science*, vol. 9, no. 10 (2003), 1244-1257
5. Oye N.D., Mazleena Salleh and N. A. Iahad, "E-Learning Methodologies and Tools" in *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 3, No.2, 2012