Tracer: A tool to measure and visualize student engagement in writing activities

Ming Liu, Rafael A. Calvo and Abelardo Pardo
School of Electrical and Information Engineering
The University of Sydney, Sydney, NSW, Australia
E-mail: [1]@sydney.edu.au

Abstract— Learning analytic techniques are allowing the observation of complex learning activities that were hidden until now. Writing is a task in which behavioral patterns can be observed to measure the level of engagement. Previous studies relied mostly on data collected by observers. In this paper Tracer, a novel learning analytic system to visualize behavioral patterns of students while writing and measuring engagement is described. The tool combines and analyzes the information obtained from document revisions and Website logs while students work in a writing assignment and provides visualizations and measurements for the level of engagement. A user study was conducted in a software engineering course where students wrote and submitted a project proposal using Google Docs. Tracer generated a graphical view of the gauged engagement, and an engagement time for each student. The obtained results show that the engagement time gauged by Tracer was moderately correlated to those reported by the students.

Keywords: Visualization, Engagement, Behavioral Analytics.

I. INTRODUCTION

Student engagement is essential in any learning activity [2]. A student who is engaged and intrinsically motivated is more likely to learn from an activity. Fredricks and colleagues [3] synthesized the research in school engagement encompassing three aspects: behavioral, cognitive and emotional engagement. Behavioral engagement is defined as participation in school related activities and involvement in academic and learning tasks. Cognitive engagement refers to motivation, effort and strategy use. Emotional engagement includes emotions and interests, such as affective reactions in the classroom or attitudes towards teachers. These three aspects are interrelated and helpful to understand engagement as a whole.

Behavioral engagement has been the focus of some researchers [4, 5]. But studies typically use evidence collected from human observers, such as teachers or students. For example, Lane and Harris [4] arranged human observers to sit among students to get engagement information including listening, writing, reading, computer use and student interaction.

In online learning activities mediated by technology, such as writing, new techniques can be used to collect detailed observations. Learning analytics (LA) use this data to make inferences about student’s behaviors. LA is defined as “the measurement, collection, analysis and reporting data about the learners and contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” [6]. On the technology side, a LA system typically contains a set of predictors (e.g. a student’s GPA [7]), indicators (e.g. the number and frequency of learning management system logins [8]), visualizations (e.g. traffic light-like display of overall student status and risks [9]) and either suggestions on how to modify a learning environments, or automatic actions to perform such changes in what is known generically as “interventions”.

Studies on writing process have been investigated over the past 30 years [10, 11]. Various models have been proposed that attempt to explain how a text is composed cognitively, which involves the continuous interaction of outlining, drafting and revising tasks. With the advance of information technology, applications such as computer keystroke-logging [12], or screen capturing [13] allow to record a detailed account of the behavior of a writer including actions such as starting a new paragraph or deleting a text portion. The writer’s reflection process can then be deduced from these events stored in the document history. More recently, the combination of writing tools with revision systems allows a deeper understanding of writing behaviors [14]. Some tools have been developed to visualize the collected writing revision. For example, LS Graph [15] focused on displaying the total number of characters inserted and deleted in a revision. Because it uses the keystroke logs to detect text changes, the revision detection process is complex. Moreover, the graph ignored other important elements, such as total engagement time and engagement intensity.

In this paper, we present a LA system called Tracer that collects events from a writing tool while students work in a writing assignment, offers visualizations to show the level of behavioral engagement and gauges engagement time. The system has two main components. The first one is the Data Collection Module, which currently relies on the Google Doc API1 and the writing tool iWrite [14]. iWrite is a web-based assignment management system, which provides a platform for students to write and submit a written assignment on Google Doc. Tracer uses Google Doc API to access the revision history of the documents managed by iWrite.

1 https://developers.google.com/google-apps/documents-list/
These two tools are combined in Tracer to record activity and provide feedback to the author about the assignment. The contribution described in this paper focuses on how to obtain a measure of the level of engagement of the writer during the assignment and show this information to teachers and students. Two visualization models derived from the collected data were computed: Line-based Visualization Model (LbVM) and Point-based Visualization Model (PbVM). Both models show an engagement score derived from the time the writer spent on the activity. The plots are created with the JqPlot graphical API. It follows a detailed description of these visualizations.

II. VISUALIZATION AND ENGAGEMENT MEASUREMENT FUNCTION

A. Point-based Visualization Model

In PbVM, each point represents an action at a particular time, such as drafting an initial version, reviewing a document submitted by a peer, or reading given feedback. Figure 1 shows an example of the PbVM showing the writer’s behavioral pattern. The activity is divided into 4 sequential sub-activities described in the activity’s timeline: initial draft, peer review of drafts, read feedback given by peers, and write the final version. The vertical lines indicate submission deadlines for each sub-activity. Each point in the line indicates an event such as initial writing or reading feedback. Different points and shapes were used according to the sub-activity. The distinction among the different activities was computed using conventional Learning Analytics techniques [14, 16, 17].

The algorithm to calculate the engagement in this visualization is based on the number of data point clusters, where the data points in each cluster are in chronological order and the length of a cluster is within a certain threshold. According to the intensity of the writing task, an adjustable threshold is used. For example, if the writing task is intensive, we need a smaller threshold to detect the clusters. The final engagement score is calculated with the following equation and algorithm.

\[ \text{Engagement} = \text{ClusterCount} \times \text{Weight} \]

Where the \text{ClusterCount} is the number of clusters and the \text{Weight} could be a minute and hour.

\[
\text{calculateEngagementForPoint}(\text{List events}, \text{float threshold}, \text{float weight})
\]

\[
\begin{align*}
&\text{startTime} = \text{currentTime} \\
&\text{beginCluster} = \text{true}; \\
&\text{clusterCount} = 1; \\
&\text{for each event in the list} \\
&\quad \text{beginCluster} = \text{false}; \\
&\quad \text{if (duration > threshold)} \\
&\qquad \text{clusterCount}++; \\
&\qquad \text{beginCluster} = \text{true}; \\
&\quad \text{duration = Duration(startTime, currentTime);} \\
&\qquad \text{if (duration > threshold)} \\
&\qquad \quad \text{clusterCount}++; \\
&\quad \text{startTime} = \text{currentTime}; \\
&\text{return clusterCount \times weight;}
\end{align*}
\]

B. Line-based Visualization Model

In the second proposed visualization LbVM, the points are connected by lines and the thickness of a line indicates the intensity of the user’s behavior during a period of time. Figure 2 shows an example of graphs generated using this visualization. Figure 2a shows the whole activity, where the points are connected into lines in each sub-activity. We defined each line as a series, which has different color and thickness. Figure 2b is a supplementary figure (zoom in Figure 2a), which gives more detail about the student’s behavior on writing draft proposal activity.

The engagement algorithm for LbVM is based on counting the number of “series” in a session. A series is defined as group of events represented by a line. Each line is associated with a certain appearance, including the thickness and color. So, the whole graph is made of lines (series). Moreover, we assign a weight to a line, which indicates the intensity of the line. The weighting process is defined as follows:

1. We defined a hashmap, where each entry contains a time threshold and a corresponding weight value. For example, \((0.5h, 0.8)\) indicates that the time threshold is 0.5h and its corresponding weight is 0.8.

2. If the duration between neighboring segments is less than the smallest time threshold, we assign that corresponding weight to the series, which that segment belongs to. In the experiment, the writing draft proposal activity is within one week, which is not intensive. We defined the hashmap including following items \((0.5h, 1),\)
III. USER STUDY

A. Participants and Procedure

A total of 38 university students participated in the study. Participants were all volunteers, signed an informed consent form approved by Human Research Ethics Committee.

The participants were enrolled in an advanced software engineering course at a university. They were required to write an individual project proposal as one of the assignments in that course. The writing activity was managed using the iWrite System [14], which assigned a Google document to each student, logged the web pages each student visited and set up the deadline for the assignment submission. The writing activity contains four sub-activities in this course, including writing a draft proposal, peer reviewing, receiving feedback from peers and tutors, and writing the final proposal. Combining the revision history available in Google Docs and the logs recorded by iWrite, Tracer generated two visualizations each of them comprised of a main figure and a supplementary figure (E.g. Figure 2a and 2b).

Each participant was asked to rate the generated figures from each visualization model using a Likert scale, where 1 was “strongly disagree” and 5 was “strongly agree”, according to following quality measures (QM):

QM1: I understand what the visualization is trying to convey.

QM2: I agree with what the visualization is showing.

QM3: The visualization is useful for me to reflect on what I did.

QM4: Could you give us your estimate of the number of hours worked on the draft proposal?

Besides these questions, we also asked each participant to indicate if they wrote the proposal entirely in an editor outside Google Docs (for example Microsoft Word) and then copied and pasted its content. In this situation Tracer did not track enough revision history because the work...
was done mostly in a local application. This “copy-and-paste” behavior is common because students are used to work with MS Word and they can work at any place without internet access restriction. This behavior influences the quality of visualizations and engagement time measurements. If participants used a “copy-and-paste” strategy, less behavior data is tracked.

IV. RESULT AND DISCUSSION

Out of the 38 participants, 17 of them used the “copy-and-paste” method, whereas the remaining 21 used Google Docs for the entire task. Students using the “copy-and-paste” method found that the alternative editor (Microsoft Word) was more convenient than Google Docs.

Table I shows the average scores given by participants to both visualization models under three quality measures. In general, the average scores for both models across all quality measure are above 3, particularly in QM1 (PbVM: 3.84, LbVM: 3.74). PbVM received higher scores than LbVM in three QMs, particularly in QM2 (PbVM: 3.21, LbVM: 3.05) and QM3 (PbVM: 3.26, LbVM: 3.05). We also performed ANOVA test and no statistical differences were founded. It indicates that students almost agree to understanding of the visualization and keep neutral to usefulness and agreement of what the visualizations show. Some students commented that the circle and diamond looks similar and they did not know how thick of the line is for how much work.

If we only analyze the scores given by the participants using “Copy-and-paste” method, the performance differences between PbVM and LbVM become more obvious. On the other hand, if we only consider the scores given by the participants who do not use “copy-and-paste” method, the scores are increased for both models and their differences become smaller. Particularly, LbVM (3.76) gets slightly higher score than PbVM (3.71) in QM1. It indicates that the performance of both models is improved when more writing behaviors are tracked, particularly the LbVM.

Both visualization models predict engagement scores by using their engagement functions. Participants were asked to give a score to indicate the number of hours they spent for the first draft proposal. Pearson correlation coefficient is used as an evaluation measurement. Table II illustrates the correlation among PbVM, LbVM and Participants. In general, the correlation between our estimation and the values reported by the subjects are low since the behavior data we obtained is insufficient, which would influence the engagement function. However, when we only analyzed the engagement score given by the participants who do not used “copy-and-paste”, the correlation is moderate since r=0.52 between PbVM and Human, and r=0.42 between LbVM and Human. This indicates that the performance of engagement predication functions increased when more behavior data were tracked.

The current Google Document API (GData API) offers significant limitations to obtain an accurate recording of the writing behavior. Only a small portion of the document revision history can be retrieved, so a more effective capturing mechanism is needed. Requesting revision history at a higher frequency from Google Docs would have an important impact on the performance of the visualizations.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed two novel visualization models, which show the behavioral pattern of a student’s writing activity, and gauged the engagement of the student in terms of the amount of time they spent on the activity. These visualization models are based on the document revision history obtained from Google Doc API, and iWrite’s logs. One of the important finding showed that these models would have better performance if the students did not use “copy-and-paste” because more

<table>
<thead>
<tr>
<th>Quality Measure</th>
<th>All</th>
<th>Copy-and-paste==true</th>
<th>Copy-and-paste==false</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PbVM</td>
<td>LbVM</td>
<td>Df</td>
</tr>
<tr>
<td>QM1: understand what the visualization is trying to convey.</td>
<td>3.84</td>
<td>3.74</td>
<td>0.10</td>
</tr>
<tr>
<td>QM2: agree with what the visualization is showing.</td>
<td>3.21</td>
<td>3.05</td>
<td>0.16</td>
</tr>
<tr>
<td>QM3: useful to reflect what I did</td>
<td>3.26</td>
<td>3.05</td>
<td>0.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PBVM</th>
<th>LbVM</th>
<th>human</th>
<th>PBVM</th>
<th>LbVM</th>
<th>human</th>
<th>PBVM</th>
<th>LbVM</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>PbVM</td>
<td>1</td>
<td>1</td>
<td>PbVM</td>
<td>1</td>
<td>1</td>
<td>PbVM</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>LbVM</td>
<td>0.92</td>
<td>1</td>
<td>LbVM</td>
<td>0.56</td>
<td>1</td>
<td>LbVM</td>
<td>0.93</td>
<td>1</td>
</tr>
<tr>
<td>human</td>
<td>0.19</td>
<td>0.08</td>
<td>human</td>
<td>0.32</td>
<td>0.15</td>
<td>human</td>
<td>0.52</td>
<td>0.42</td>
</tr>
</tbody>
</table>
document revisions would be tracked. Another important finding illustrated that the engagement time predicated by our model is moderately correlated to the actual time spent by the student. However, current visualization models only consider the time the writer spent rather than the content of document, such as the length and topics of the document. Thus, our future work would focus on extracting information from the content and improving the graphical interface and engagement function. Moreover, we will develop visualizations for teachers, which can help them to easily track the progress of the whole class.

ACKNOWLEDGEMENT

This project was partially supported by Australian Research Council Discovery Project DP0986873

REFERENCES


