PRELIMINARY PREPRINT VERSION: DO NOT CITE
The AAAI Digital Library will contain the published
version some time after the conference.

Measuring Students' Engagement with Digital Interactive Textbooks by Analyzing Clickstream Data

Breanne Crockett
University of Toledo
breanne.crockett@rockets.utoledo.edu

Abstract

This paper provides an overview of my contributions to a project to measure and predict student's mental workload when using digital interactive textbooks. The current work focuses on analysis of clickstream data from the textbook in search of viewing patterns among students. It was found that students typically fit one of three viewing patterns. These patterns can be used in further research to inform creation of new interactive texts for improved student success.

Project Goals

The goal of this project is to analyze how students use interactive digital textbooks. To do this, I examine clickstream data from a chemical engineering ZyBook collected across five different semesters. ZyBook is a web-based replacement for textbooks that contains static text passages with interactive animations and practice problems dispersed throughout (ZyBooks). Figure 1 shows two steps of an animation.

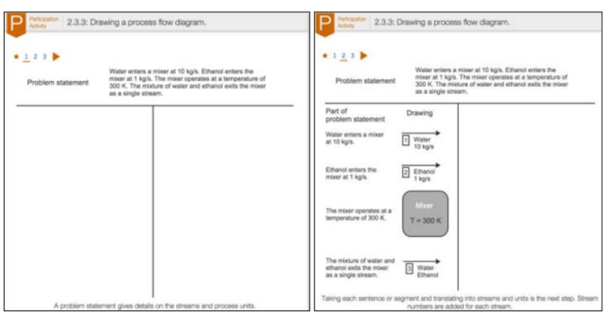

Figure 1: Example ZyBook Animation

To move through animated content, students are required to click the play button to initiate each step. Students can also click a step number to replay or go back to a previous step. This process generates a clickstream that can be used to determine how long students spend on animations and how their view time changes with animation type, animation length, point in the semester, and other factors. View time is assumed to be an indicator of the effort students spend digesting material. From clickstream data, I hope to discover patterns in students' interaction with ZyBook and use those patterns to inform feature selection for machine learning algorithms in future work.

Previous & Related Work

Historically, there has been a significant amount of research indicating that students often skip reading assignments in static textbooks. Hobson (2004) suggests that as many as 70% of students do not read the assigned passages of the textbook. With the adoption of interactive texts, more recent work, such as that by Liberatore (2017), has indicated that students complete reading assignments much more often from digital interactive texts. These works have focused largely on completion rates and correlation between completion of reading and success in the course. In interactive texts, completion rates indicate that students have at a minimum clicked through the assignment. However, completion rates cannot decipher if a student truly read/watched the animation or spent any effort to understand it. This work aims to move beyond completion rates and determine if students are mentally engaged in the reading assignments and what features of the interactive text best engage students.

Additionally, there is significant work in clustering and classification of clickstream data. Nguyen (2015) highlights several algorithms for analyzing large clickstreams, with special attention to algorithms that make a single pass through the data and require relatively small working memory. These algorithms are more complex than necessary for the small dataset examined here but would be beneficial if more data were collected from additional cohorts, textbooks, or universities.

Personal Contributions

In the initial stages of the project, my role was primarily data analytics. I performed basic statistical analysis and created figures to highlight data trends using Python and various scientific libraries. These results were critical in a master student's thesis (Stone 2021). As I became more familiar with the required tools, my role expanded into the implementation of more complex modeling techniques. To determine the number of different ways that students interact with the digital texts, clustering was performed. K-Means clustering with Euclidean distance was performed on a single cohort worth of data. Each of the 111 students was a data point and the feature vectors consisted of view times for each of the 144 animations. The clustering was performed with values of k from 2 through 10, to decide which value of k best fit the data. Additionally, clustering was performed on smaller sets of features, such as average time spent per chapter and average time spent per category of animation.

Results

It was found that student clickstream data is best grouped into three clusters for all aforementioned sets of features. Other numbers of clusters were considered, but when $k > 3$, clustering resulted in one or more groups of just one or two students with three cluster including the majority of students. This indicates that those single-student groups are outliers to the three primary clusters. The results of clustering are illustrated in the figure below, where each line represents the centroid of one cluster. Each point along the x-axis represents one animation from the interactive textbook. The animations are presented in the order they occur in the text. The y-axis shows the view time corresponding to each animation. The legend shows the number of students in that cluster and the corresponding percentage.

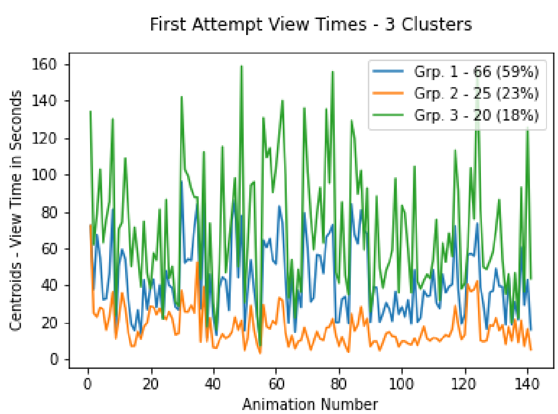

Figure 2: Results of k-Means Clustering

In Figure 2, the bottom cluster shows students with consistently shorter view times. Presumably, these are students who click through the animations just to earn the completion points. The middle cluster indicates that a group of students earnestly watch each animation. Lastly, the top cluster shows consistently longer view times. The consistently longer view times could indicate that some students are rewatching all or part of an animation, or simply spend more time digesting the information presented.

Applications & Next Steps

The results of preliminary analysis and clustering provide foundations for future work on this project. The results can be used to construct a generative model for students' view times. An application of the generative model could be forecasting student view time. Based on the view times of early chapter, it would be possible to predict view times for remaining animations.

Additional work can be done to better measure students' engagement with the interactive text. While view time may be an indicator of effort spent, it is not sufficient to determine the mental workload of the student. Ultimately, the goal for this project is to combine clickstream data with wearable sensor data to predict students' mental workload. While the pandemic has delayed the collection of sensor data, it is planned to measure students' physiological response, such as heart rate variability and electrodermal activity, while they are using the interactive text and other media. A machine learning algorithm will be trained to use student clickstream data and sensor data to predict the workload as self-reported by the student on the NASA Task Load Index. By comparing the workload of students on animations in a digital interactive textbook with traditional textbook passages, it may be possible to find a correlation between student success and mental demand. It may also be possible to determine features of an "ideal" animation that require the optimal mental workload. Those results can inform the design of digital textbooks and animations to improve student success.

Acknowledgements

This material is based upon work supported by the National Science Foundation grant EEC-2025008.

References

E. H. Hobson. (2004, December). Getting Students to Read: Fourteen Tips.

M. W. Liberatore, "High textbook reading rates when using an interactive textbook for a Material and Energy Balances course," Chemical Engineering Education, vol. 51, no. 3, pp. 109-118, 2017.

Nguyen, HL., Woon, YK. & Ng, WK. A survey on data stream clustering and classification. Knowl Inf Syst 45, 535–569(2015). https://doi.org/10.1007/s10115-014-0808-1

Sidney J Stone III. Attitudes Toward and Usage of Animations in an Interactive Textbook for Material and Energy Balances. (2021). University of Toledo.

(September). zyBooks: Interactive textbook replacements. Available: http://zybooks.com/