Using Mouse Cursor Coordinates to Predict the Learners' Usefulness Rankings of Online Resources with an Intelligent Web Browser

Applicant:
Vai Suliafu
27 June, 2016

Faculty Sponsor:
Eric G. Poitras
Assistant Professor of Instructional Design and Educational Technology
Department of Educational Psychology
University of Utah

Topic of Research – UROP Proposal

Statement of the Problem/Topic of the Research or Creative Work

The last two decades have seen significant advances in technology that empower educators to take advantage of innovative new methods that address the limitations of traditional instruction (Collins & Halverson, 2009). In particular, artificial intelligence and user-adaptive learning technologies stand to facilitate the time-consuming process of searching and gathering high-quality educational resources and materials from the web. Research suggests that teachers should improve their own information-seeking skills because it allows them to better plan and adapt their own teaching practices to changes in the curriculum, and in turn contributes to increased student performance (Moos & Ringdal, 2012). The proposed project aims to improve the nBrowser (Poitras & Fazeli, in press), an intelligent web browser designed to recommend online resources crawled from popular teaching blog sites as well as the National Science Digital Library (nsdl.oercommons.org) to support them in designing their own lesson plans.

The issue addressed in the current project is the reliance of nBrowser on self-report data obtained from pre-service teachers to determine the usefulness of online resources in building a lesson plan. Recent empirical research has shown that teachers seldom report the usefulness of online resources as only 57% of resources in the nBrowser database were rated by users. Moreover, of the total amount of sites visited by users, regardless of whether these are recommended by nBrowser, a mere 8% were rated by the users (Poitras & Fazeli, in preparation). The aim of this project is to build on these findings by exploring alternative sources of data to inform the nBrowser system recommendations, namely, the mouse cursor position and movement on the screen. It is hypothesized that patterns in mouse cursor movements are predictive of the ratings of the utility of online resources submitted by teachers, thereby serving as a complementary source of information that is logged by the system.

Relevant Background/Literature Review

In the literature on intelligent tutoring systems, the importance of the learner model has been subject to considerable discussion (Desmarais & Baker, 2012). A learner model consists of a representation that is updated by a tutoring system that captures the progress made by a learner in achieving an instructional objective or performing a task (Shute & Sapta-Rivera, 2012). The particular application which serves as the focus of this proposal, nBrowser, is "an intelligent web browser that implements a spreading activation algorithm to converge a network-based model towards an optimal arrangement based on learners' self-reported usefulness ratings of online resources" (Poitras & Fazeli, in press). During nBrowser's earlier stages of development, Poitras and Fazeli (2016) sought to construct a database of useful resources for teachers who desired to implement technology into their instructive methods. With this in mind, Poitras and Fazeli (2016) crawled a total of forty educational blog sites that focuses on the implementation of technology into the classroom. From here, blogs which included URL's that stated the date of publication for the site were entered into the database. Contrastingly, blogs which "featured pre-determined extensions (e.g., .png, .jpeg) as well as statements (e.g., tag, feed, share)" were excluded. The content of included blog posts were then transformed into vectors comprised of term occurrences so that similarities between blog content could be established by the system.

As aforementioned, nBrowser's original algorithm relied upon user-determined usefulness ratings of the various sited visited while using nBrowser. These usefulness ratings then allowed nBrowser to converge toward an optimal arrangement of networks to suggest to users. However, while assessing the rate of nBrowser's network convergence, Poitras and Fazeli (in preparation) learned that of the resources featured on nBrowser's network, only 57% were

given a usefulness rating by a user. Moreover, of the total resources visited by users, a mere 8% were rated by nBrowser's users (Poitras & Fazeli, in preparation).

Henceforth, it stands to reason that in order for nBrowser to converge towards an optimal state, the system required more feedback from the users. Going further, nBrowser must interpret alternative metrics that measure the usefulness of online resources. In order to accomplish this, log-file data on several user behaviors was obtained from nBrowser, before being filtered only for the sites that were given a usefulness rating (Poitras & Fazeli, in preparation). The behaviors included elapsed time on a webpage, nBrowser panel selections, total amount of pages visited during the session and lesson plan edits made while viewing a page in nBrowser. A prediction model was trained on that dataset and achieved a 70% rate of accuracy in distinguishing between positive and negative ratings of online resources. The proposed project builds on these previous efforts through the use of a popular technique in the field of economics, namely, the use of time-series models for forecasting events. To the best of our knowledge, this is the first application of such an approach in the contest of intelligent tutoring systems and the logged mouse cursor positions and movements made by the users, which stands to significantly improve the predictions made by the model and address the limitations of the self-report measure.

Specific Activities to be Undertaken and a Timetable Allotted for Each Activity

1. **Loading and cleaning the log-file data in R.** I will identify all of the various pieces of information within the different types of log-file data, including but not limited pixel coordinates, descriptors of the observations, timestamps, participant IDs, row ID's, categories of events, and usefulness ratings. Additionally, I will trim all mouse cursor coordinates that were captured while no user rated resource was active in nBrowser. I will then read this data into the statistical software program R, while filtering for the desired variables so that the

cleaned data is organized and ready for analysis. I will complete the "Importing Data into R" and the "Cleaning Data in R" courses offered by Datacamp to learn the skills required for this. The log-file data will be completed and ready for analysis no later than July of 2016.

2. **Plot mouse cursor positions over time for each user rated resource.** I will construct plots of mouse cursor positions over time for each occasion a user viewed a resource and then gave it a usefulness rating. Completed during August of 2016.
3. **Regression analyses on mouse cursor positions over time per occasion.** I will run a regression analysis, detecting any causal effect mouse cursor positions may have on subsequent ratings of usefulness. This will allow me to detect any broad trends. Completed during September of 2016.
4. **Compute expected values of mouse cursor position within time intervals.** Considering both the total number of observations and the desired number of observations, I will construct bins of data, segmented by elapsed time. I will then calculate the average location of the mouse cursor for each given interval. Completed during October and November of 2016.
5. **Plot average locations of mouse cursors for each interval.** While plotting these average locations, I would expect to see small emergent trends between users of the same module. Completed during December of 2016.
6. **Analyze results.** At this point I may identify common areas users dwelled upon with their cursor, as well as those locations relations to usefulness ratings. Additionally, I might identify outliers, or irrelevant mouse cursor measurements at this time. I could run multiple regression analyses on different locations in order to determine any causal effect between the cursor location while viewing the resource on usefulness ratings. This causal effect may allow me to

identify the probability that a user would give a resource a high reading, given they scrolled upon a certain location within a given resource. Completed during January of 2017.

7. **Test and weight mouse cursor behavior as predictor variable.** Using the results from the different regression analysis, we could test mouse cursor behavior as a predictor variable within the data sets which contain usefulness ratings. Completed during February of 2017.
8. **Prepare findings for academic events.** At the beginning of March 2017, I will prepare an oral/poster presentation for the 14th annual Undergraduate Research Symposium hosted by the Office of Undergraduate Research on Tuesday April 14, 2017 in the Olpin Union Building.

Relationship of the Proposed Work to the Expertise of the Faculty Mentor

The proposed work builds on the research led by Dr. Eric G. Poitras and graduate research assistants affiliated with the Advanced Instructional Systems and Technologies laboratory in the Department of Educational Psychology at the College of Education. The program of research on network-based learner models leverages the Novessa suite of software applications, including nBrowser which serves as both a research and training platform with pre-service teachers enrolled in the College of Education. This research is supported through the Learning Environments Across Disciplines research partnership grant of the Social Sciences and humanities Research Council of Canada.

The issue addressed by this research is key to improving the adaptive capabilities of the system, and has been the subject of several lines of inquiry aimed at discovering novels ways to leverage the data logged by the system and utilizing new features to converge the networks. The broader implications is to optimize the processes that allow nBrowser to make better recommendations, thus resulting in improved learning gains and user satisfaction. These preliminary findings will serve to pursue funding from external agencies during the summer

2017 period (i.e., NSF and IES) to scale-up the research with a larger sample of participants and utilize simulated learner scenarios to improve the algorithms that underlie the recommender system based on patterns detected in the mouse cursor movements and position on the screen.

Relationship of the Proposed Work to the Student's Future Goals

Acceptance into the Undergraduate Research Opportunities Program would provide for me intellectual fulfillment, a source of income that doesn't detract my time and focus from academic achievement, and an invaluable credential to list on my applications to PhD programs after graduation at the U. Throughout my undergraduate experience, my fascination with economics and mathematics as they relate to economics has flourished. Independent research that applies an economist's approach to statistical modeling would consume my intellectual curiosities in the best way imaginable! Additionally, the funds from UROP would provide me with a much needed source of income that, unlike other forms of part-time employment, would not divert my focus away from my education as a budding economist. By spring of 2017, I will have completed all of the major requirements for a B.S. in Economics with a Statistical Analysis Emphasis, as well as all of the requirements for a minor in Mathematics. In the subsequent fall and summer semester, I will complete my remaining general requirements and some additional math courses that better prepare me for Economics PhD programs. Plainly put, the Undergraduate Research Opportunities Program would greatly bolster the strength of my applications to PhD programs, placing me in an entirely new tier of applicants. Thus, when considering my intellectual fulfillment, my need for income as a student, and my future applications to Economics PhD programs, the proposed research advances the cause of each.

References

Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America.* New York: Teachers College Press.

Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction, 22*(1-2), 9-38.

Moos, D. C., & Ringdal, A. (2012). Self-regulated learning in the classroom: A literature review on the teacher's role. *Education Research International, 423284,* 1-15.

Poitras, E. G., & Fazeli, N. (2016). Educational web mining to support teachers' self-regulated learning in planning technological integration in the classroom. *Paper presented at the 2016 American Educational Research Association Annual Meeting*. Washington, DC.

Poitras, E. G., & Fazeli, N. (in preparation). Modeling information-seeking behaviors during learning from hypermedia with nBrowser. *Paper to be submitted to the 2017 American Educational Research Association Annual Meeting.* San Antonio, TX.

Poitras, E. G., & Fazeli, N. (in press). Mining the edublogosphere to enhance teacher professional development. *Social Media Data Extraction and Content Analysis.* IGI Global.

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach (Ed.), *Adaptive technologies for training and education* (pp. 7-27). New York, NY: Cambridge University Press.