Research Overview

Youngjin Lee

April 10, 2020

My research interests lie at the intersection of advanced learning technologies, such as computer simulations and games, and Educational Data Mining (EDM)/Learning Analytics (LA). Although these two areas may look unrelated at first glance, they are closely linked because advanced learning technologies, such as computer simulations and games, often create a log file of user actions that can be analyzed only by EDM or LA. The following shows a few examples of advanced learning technologies I developed, and EDM/LA projects I conducted in the past.

Developing Advanced Learning Technologies

Virtual Microscope Project

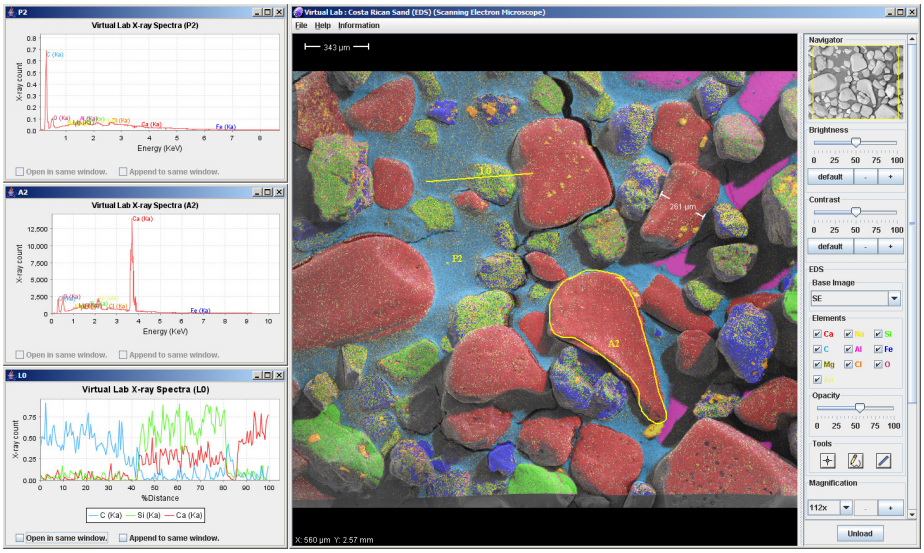

Figure 1: Virtual Microscope

While working at the National Center for Supercomputing Applications (NCSA, `http://www.ncsa.illinois.edu`) as a senior research programmer, I developed an educational computer simulation called Virtual Microscope (see Figure 1). Virtual Microscope was a NASA-funded project that provides simulated scientific instrumentations for students around the world as part of NASA's Virtual Laboratory initiative. Using Java programming language, I developed software that can not only enable students to examine various specimens using a virtual electron microscope, but also can replay how students used the software based on user actions captured in the log file, allowing teachers to better understand the ability of their students.

Learning Weight Maintenance Skills in Second Life

During 2012–2016, I collaborated with researchers in the University of Kansas (KU) Medical Center in a project funded by the National Institute of Health (NIH). The goal of this project was to develop a virtual community in Second Life (`https://secondlife.com`) that allows people to learn and practice weight maintenance skills without going to a face-to-face clinic (see Figure 2).

Figure 2: Virtual Gym in Second Life

The virtual community includes gyms, grocery stores, restaurants and more in which they can practice various weight maintenance skills such as purchasing healthy foods and learning how to use a gym facility. As part of this project, I developed an automated data collection system in Second Life that captures participants' action without interrupting their learning activities, allowing clinical researchers to examine the effectiveness of virtual learning activities taking place in Second Life. The result of the clinical test indicates that the virtual

learning experience is as effective as much more expensive face-to-face learning activities.

EDM/LA projects

Log File Analysis of Computer-Based Physics Tutor

I started working on the EDM/LA research projects when I joined MIT's RESearch in Learning, Assessing and Turing Effectively (RELATE) research group (`http://relate.mit.edu`). When I joined, the RELATE group developed a Web-based physics tutoring environment named Cyber Tutor[1] which was being used by MIT freshmen. I analyzed the clickstream log files of Cyber Tutor to investigate how students used instructional hints provided by the tutoring system, and when those instructional hints became most helpful to students. In another EDM/LA project, we compared MIT freshmen's self-reported cheating on physics homework to their behaviors captured in the log files. Overall, students under-reported their homework cheating behaviors, compared to their suspicious activities recorded in the Cyber Tutor log file. The largest discrepancy between the self-report and logged behaviors was observed from the students who choose "business" as their major in the following year. To the best of my knowledge, it was the first study adopting a quantitative approach based on the log file analysis that goes beyond the self-reported survey. The findings from these EDM/LA projects were published in the *Physical Review Physics Education Research*, the most respected journal in physics education, and favorably reviewed by MIT, KU and *US News & World Report*.

Predicting Problem Solving Performance of MOOC Students

With the support from MIT's RELATE research group, I applied various EDM/LA algorithms and techniques to the log files of an edX MOOC in order to quantitatively study student activities, behaviors and learning performance. I used Support Vector Machine (SVM) and regularized regression to build a predictive model of student learning that can estimate the probability of correctly solving mathematics or physics problems, based on how well students solved other problems in the past. In most of computer-based learning environments, instructional scaffolding, such as hint or feedback, is provided when students are unable to resolve a learning task for a certain number of times or in response to the request of students. However, these approaches would not be able to maximize student learning outcomes because they do not take into account the ability of students and the difficulty of learning tasks. A predictive model that I developed in this study can be used as a quantitative guideline for when

[1]Cyber Tutor was later purchased by Pearson Education, and rebranded as MasteringPhysics (`https://masteringphysics.com`).

to provide instructional supports and guidance in the computer-based learning environment.

Self-Regulated Learning (SRL) in MOOC

It is a conventional wisdom that the more SRL skills students have, the better they can perform the learning tasks. Instead of relying on self-reported questionnaires, I used EDM and LA techniques, such as Self-Organizing Map (SOM) and hierarchical clustering algorithms, when investigating the relationship between students' SRL skill and their academic performance in the MOOC. The outcome of this research indicates that MOOC students were able to achieve higher scores on the weekly homework and quiz problems, and had a better chance to pass the course when they started working on their weekly homework and quiz problems early during the one-week assignment cycle, and when they spent more time using learning resources without getting distracted by other tasks not directly related to the course. The EDM and LA techniques developed in this research can help MOOC instructors easily identify similar students based on a large number of variables and examine their characteristics from multiple perspectives, allowing them to provide instructional supports and guidance tailored to the identified groups of students showing similar characteristics.

Research Quantity and Quality

I have published 24 SCI or SSCI indexed peer-reviewed journal articles, 10 of which were after tenure. In addition, I published 2 articles and have 3 manuscripts currently under review or in revision after joining UNT. I have presented another 33 conference papers, 14 of which are after tenure. I published my works in top tier, highly regarded, SCI or SSCI indexed journals in the field such as:

- *Information Discovery and Delivery* (5-year impact factor in 2018: 7.462)
- *Computers & Education* (5-year impact factor in 2018: 5.902)
- *Computers in Human Behavior* (5-year impact factor: 4.964)
- *Educational Technology & Society* (5-year impact factor in 2018: 2.682)
- *Contemporary Clinical Trials* (5-year impact factor in 2018: 2.660)
- *Journal of Computer Assisted Learning* (impact factor in 2018: 2.451)
- *Physical Review Physics Education Research* (impact factor in 2018: 1.964)
- *Interactive Learning Environments* (impact factor in 2018: 1.929)
- *Journal of Educational Computing Research* (5-year impact factor in 2018: 1.542)

- *International Journal of Information and Learning Technology* (impact factor in 2018: 1.450)

Grants

I served as a Co-PI in two federally funded grants. In the *Center on online learning and students with disabilities* project ($7,484,978), which was funded by U.S. Department of Education, I led a research team developing a suite of online learning contents for students with learning disabilities that are designed to reduce cognitive load by adopting Mayer's Cognitive Theory of Multimedia Learning principles.

In another project, *A virtual reality intervention (Second Life) to improve weight management* ($3,676,028), which was funded by the National Institute of Health (NIH), we developed a virtual community in Second Life where people can learn and practice weight management skills, and researchers can conduct rigorous clinical trials to investigate the effectiveness of virtual learning activities, compared to more expensive face-to-face weight management clinics. I conducted a usability test on the virtual learning activities, and developed an automated data collection module that enabled clinical researchers to collect data while research participants were using the developed virtual community.

In addition, I was able to secure 5 internal grants (totaling $37,324) to support my research activities. Also, I have applied for 6 external grants as PI or Co-PI. I have learned greatly from all the funding application efforts, successful or unsuccessful. In the process, I have gained great insights as well as solid collaborative networks.

Recently, I started collaborating with Dr. Tania Heap at the Center for Learning Enhancement, Assessment and Redesign (CLEAR). By utilizing EDM and LA, we plan to examine the learning performance and satisfaction of students enrolled in face-to-face courses that had to be switched to online classes in the middle of the semester due to COVID-19 pandemic, compared to students taking an online course from the beginning of the semester. Also, I am working with Dr. Mary Barton and Dr. Dan Hubbard at the Data, Analytics and Institutional Research (DAIR) in the University to identify gateway courses for Science, Technology, Engineering and Mathematics (STEM) degrees. By using EDM/LA, we will identify the courses that have a strong detrimental effect on the completion of STEM degrees, allowing the University to allocate more resources in those gateway courses to improve student success. When successfully completed, these projects can be extended into a larger, externally funded grant proposal.

Conclusion

In summary, I am well prepared for conducting state-of-the-art research in advanced learning technologies and EDM/LA with scholars in education, STEM,

data science and computer science. While reflecting on my tenure-track experience at KU and UNT, I believe I am ready to take my research to the next level. I am a productive and proud member of the UNT community, grounded in its vision, mission and goals. I will continue to publish and present research, and seek grants, contributing my expertise to UNT and the global community.