CS 410 Fall 2020 Technology Review: Developing Teaching Material Using a Combination of Data Mining, Machine Learning, and Crowdsourcing Techniques

By Mengyu Zhai (NetID: mzhai)

ABSTRACT

I will review some applications of data mining, machine learning, and crowdsourcing techniques in the development of teaching materials and how the applications improve the quality of instructional/learning experience design.

INTRODUCTION

Data Mining & Machine Learning

Work of Jung et al [3] focuses on improving learning experience of using slide-based lecture videos by improving effective video interactions, more specifically, emphasizing the current item in the slide that is currently being referenced by the narration, allowing slide-item-based navigation to the relevant part of the video, and supporting slide-item-based note-taking that allows copied slide parts in separate notes to preserve links to relevant parts of the video. To achieve these interaction goals, the researchers present a prototype video processing and browsing system named DynamicSlide that automatically extracts pairs of slide items and sentences in the instructor's verbal narration, which they call "*references*", for a variety of slide-based videos. The *reference* estimation algorithm pipeline includes three stages: shot boundary detection, test segmentation within slides, and text-to-script, but the algorithm is not mature or perfect yet as the authors mentioned, so there is still some distance from the research experiment to real applications.

Targeting specifically on the learning of coding or programming, Head et al [5] propose routines called Tutorons that automatically generate context-relevant, on-demand micro-explanations of code, so programmers do not need to consult external documentation to solve problems or find example code. Basically, a Tutoron detects, parses, and explains code for a programming language in HTML documents. Very attractive to programmers just by reading the idea. The study is still at a preliminary stage and there are still many areas to be improved, but it certainly sounds like a promising tool.

Research of Wang et al [4] is on software learning, which has a big market in this digital era. What they try to develop is a hierarchical approach that combines topic modeling and frequent pattern mining to classify the workflows offered by a software application and then to recommend new workflows and their associated videos to software users. Topic modeling categorizes command logs into meaningful high-level tasks, and then frequent pattern mining identifies distinct command patterns for each task.

Crowdsourcing

Working in the field of training and instructional design (ID), I highly agree with Liu et al that "linear representation of video is one of the biggest problems that limits learners from exploring learning materials effectively."[2] In ID field one common method attempting to overcome this disadvantage of using video is providing transcripts that is searchable and linked to the time stamps in the video.

Liu et al provides a new perspective. A system called ConceptScape [2] was built by the researchers to allow crowd workers (video watchers) to generate a concept map by prompting them to externalize reflections on lecture videos. The research shows that such a system can generate complex and high-quality concept maps that can be useful for concept-based video navigation and comprehension when presented as a visual tool together with the video to future video watchers. The concept map generating process also provides a learning opportunity for video watchers to verbalize and summarize concepts. The crowdsourcing technique was also evaluated with remote independent online users, making the research result more applicable in our current time where more and more training and learning are happening in a remote and online format.

Crowdsourcing & Machine Learning

Similar to Liu et al [2], a group of researchers also used crowdsourcing to obtain explanations for test problems in the Adaptive eXplanation Improvement System (AXIS) [1], which then uses machine learning to dynamically determine the best explanation to present to a future learner. Based on the research, the quality of AXIS explanations is as high as those generated by an experienced instructor, and thus using AXIS can provide the same learning benefits as having an experienced instructor to provide explanations for the test. If applied to the massive amount of online tests we have especially during this time (and in the future too), this will help greatly in solving the problem of lacking high-quality explanations on line platforms, and thus improve the learning outcome of increasing remote independent online learners.

DISCUSSION & CONCLUSION

As one currently working in the field of training and instructional design, I often need to consider how to improve learning experience and learning outcomes, and lots of tools help achieve these goals by meeting learners' needs of better interacting with the teaching materials, having easier navigation in teaching materials, becoming more independent and self-directed learners, and learning more effectively and efficiently. The techniques reviewed in this paper all have great potentials in instructional design. I can image a lot of tools we are using for instructional design went through a similar path from concept, protocol, construction, pilot testing, to the real application, and these studies give me an opportunity to learn the algorithms used in building such tools. However, at the same time, the tools/systems all have space for improvements. I also feel these studies will benefit a lot if the researchers, who are mainly from the field of computer science/machine learning, can collaborate with researchers/practitioners from education/instructional design.

REFERENCES

1. AXIS: Generating Explanations at Scale with Learnersourcing and Machine Learning. Available at: https://juhokim.com/files/LAS2016-AXIS.pdf. Accessed in October 2020.
2. ConceptScape: Collaborative Concept Mapping for Video Learning. Available at: https://dl.acm.org/doi/pdf/10.1145/3173574.3173961. Accessed in October 2020.
3. DynamicSlide: Exploring the Design Space of Reference-based Interaction Techniques for Slide-based Lecture Videos. Available at: https://dl.acm.org/doi/pdf/10.1145/3264856.3264861. Accessed in October 2020.
4. Leveraging Community-Generated Videos and Command Logs to Classify and Recommend Software Workflows. Available at: https://dl.acm.org/doi/pdf/10.1145/3173574.3173859. Accessed in October 2020.
5. Tutorons: Generating Context-Relevant, On-Demand Explanations and Demonstrations of Online Code. Available at: https://andrewhead.info/assets/pdf/context-relevant-explanations.pdf. Accessed in October 2020.