Developing teaching material using a combination of data mining, machine learning, and crowdsourcing techniques

Author: Diego Carreno Garcia
diegoac3@illinois.edu

Introduction

The purpose of this technological review is to compare current data mining, ML, crowdsourcing techniques used to develop new teaching material for online courses, with their pros and cons, provide an overview of their current applications, and present possibilities to further develop these techniques and applications in the future. There were 3 technologies reviewed in this document, which to my knowledge are some of the most cutting edge approaches to generate teaching material for online courses today and offer a broad overview of the field. The reason for a review in this area is because there is an explosion of content that can be found online but not enough detailed and personalized material so new learners can truly grasp underlying concepts. Instructors that create the content are scarce and so it is difficult for them to keep the content fresh or to answer every question a student might have.

Comparison and Overview

1. Axis (Williams et al.)

Axis is a system that provides explanations to problem sets for online classes through a combination of crowdsourced content from users and machine learning to predict the most helpful explanation for a given question. The platform prompts users to generate and revise explanations, and uses ML to predict which explanations are useful. Through this approach the system has shown to objectively enhance learning for students when compared with the standard method of reviewing answers without explanations (students have indicated they feel more confident in their ability to answer similar questions). Axis Uses a bayesian modeling approach called Thomson sampling to address the problem of exploitation vs exploration (ie. finding the right balance between offering an already helpful explanation vs offering explanations that might be more, or less helpful), which has shown to perform well on practical problems.

Pros

- Offloads explanation generation from instructors to learners and has shown to be as effective as instructor-generated explanations and can help both learners and instructors. Technology stack uses systems easy to consume and adapt to most end-users, Qualtrics and App Script for Google Spreadsheets.

Cons

- Built with a short term focus. It might be better to set grades or accuracy in subsequent problems as the reward for the system. Although this is more complex, current evaluation was done on a short term reward (ie. rating provided by the user to an explanation). No individual-level personalization. The system has no regard for level of knowledge and preference of learners.

Future work

- Immediate explanation helpfulness is not necessarily a good measure of long term knowledge and it might be recommended to focus on a longer term performance metric.
- Personalize content to study habits of students.

2. ConceptScape (Liu, Kim, and Wang)

ConceptScape aims to improve the experience of learning from instructional videos. Videos offer a linear representation of concepts, which is challenging to grasp to novice students. ConceptScape offers a concept-based navigation approach in which users are prompted to reflect on the content they've consumed which creates a crowd-sourced concept map of concepts included in a video, which has shown to improve retention for users with limited metacognitive skills. It follows a three step workflow that users must follow to crowdsource the information needed to generate the map: list concepts, link concepts, explain relationships. Through this approach, researchers have managed to produce concept maps from novice users that are of comparable quality to those produced by experts in the field.

Pros

- Can offer a structured way map so learners can more easily navigate and digest concepts in instructional videos (concept-driven navigation).
- Can help with content retention by asking students to externalize reflections on the content they're consuming.
- Well-defined concept map generation workflow that yields concept maps from novice users that are comparable as those generated by experts.

Cons

- No assessment of long term concept retention, but rather focused on personal perception of understanding of the content.
- It is unclear how much a concept map help with long term learning.
- Linking of concepts only happens within the same video, and concept tagging is a very manual task.

Future work

- Further exploration is needed on the interaction between learners, video content, and concept maps.
- Have the system produce timestamps and potential concepts that learners would prune and adjust to further promote scalability of the system using NLP.

3. **Tutorons (Head et al.)**

Tourons is a set of routines that generate Context-relevant and on-demand micro-explanations for CSS, regular expressions, and Unix command “wget”. This is particularly needed for new programmers learning a new language, in which they find examples of code written by instructors who wrote for a certain audience in mind but which might leave out explanations on underlying languages used for a novice audience. Sites like stackoverflow help but still can have insufficient explanations for all backgrounds.

Pros

- Can produce effective micro-explanations so novice programmers don’t need to reference external documentation.
- Tutorons rules can be modified and expanded to include more languages as it already offers a framework to build a routine for any language in three steps: detection, parsin, and explanation.

Cons

- Needs to be more broadly expanded to other applications. Also, not all explanations offer visual affordances, and some explanations were incomplete.
- Further testing is needed. Study might be biased since they knew the micro-explanations were part of the study.
- Needs to be augmented before they can provide insights for additive tasks.
- Explanations are rather deterministic and might lack input from advanced users that might offer additional context to explain why a given piece of code performs the way it does.

Future work

- General expansion and disambiguation across languages is needed since current rules only apply to CSS, regex and wget commands.

Discussion and future applications

Although developing teaching material through data mining, machine learning, and crowdsourcing techniques is a very promising area, it is still incipient in current scope and applications. The information systems discussed have a limited scope in that they only apply to very particular scenarios, and also have challenges in scalability since they do not personalize content and approach to individual learner preferences. In the future, an expansion area for this work can include developing more holistic systems that can offer a personalized learning experience to individuals regardless of the topic they are learning about. It might also be beneficial to link concepts across fields in ways similar to what Google Knowledge Graph has started doing in text-based information to offer a more holistic learning experience.

Nevertheless, even with limitations found for current techniques presented in this review, a clear application that might be of value in today’s industry is the rapid acquisition of new knowledge for employees. For example, companies might not have a machine learning expert to guide new AI projects or product development, but companies might be able to upskill current employees leveraging the tools presented here (because these tools can replace the role of an expert in the company).

Conclusion

Developing teaching material is a key need in today’s world given the explosion of MOOCs and general instructional content on the internet. It is very promising to see that there are techniques that can further scale online learning to better grasp underlying concepts even in the absence of an expert instructor. These are perhaps the first steps in creating a fully machine-powered instructor or information system that can help novice students learn concepts at a large scale in a rapid way.

References

Head, Appachu, Hearst, and Hartman. n.d. “Tutorons: Generating Context-Relevant, On-Demand Explanations and Demonstrations of Online Code.” https://andrewhead.info/assets/pdf/context-relevant-explanations.pdf.

Liu, Kim, and Wang. n.d. “ConceptScape: Collaborative Concept Mapping for Video Learning.” https://dl.acm.org/doi/abs/10.1145/3173574.3173961.

Williams, Kim, Rafferty, Maldonado, Gajos, Lasecki, and Hefferman. n.d. “AXIS: Generating Explanations at Scale with Learnersourcing and Machine Learning.” https://juhokim.com/files/LAS2016-AXIS.pdf.