

Applications of Text Data Retrieval and Analysis in Augmenting the Learnability of Teaching Materials

Kaung Yang

kaungky2@illinois.edu

CS 410: University of Illinois, Urbana-Champaign

Nov 7th 2021

Introduction

With the unprecedented sophistication and ease of access to digital technologies, online learning platforms have become crucial for learners to attain their knowledge. Today, there are a plethora of such platforms to choose from, including Massively Open Online Courses (MOOCs) (i.e. Coursera, edX, Udacity, and Udemy), and Youtube tutorials. Despite the availability of these platforms, learners often struggle to complete courses online [1], and web-based learning also has its disadvantages, including de-individualized instruction and poor instructional design [2]. This leads to a less engaging and less effective learning through the online space. However, the recent developments in *the use of* text mining and machine learning have helped researchers address these issues in different ways. In this review article, we will explore the different approaches that researchers take to improve, augment and extend the *learnability of teaching materials* (i.e. Youtube tutorials, annotation of texts, etc.) by means of data mining, machine learning and crowdsourcing techniques.

ToolScape [3]

The Youtube platform is glittered with how-to videos. Often, learners are eager to master a new skill, but are faced with inefficient browsing and video playing interfaces. A few ways to solve this involves asking instructors to include step-by-step information or using computer vision; however, these are rather expensive. ToolScape is an interactive tool that augments traditional video players with timelines and *step-by-step annotations* to help learners navigate 'how-to' videos with more interactivity. In order to achieve this, they establish a crowdsourcing workflow, which includes the Find-Verify-Expand design pattern, to (1) find candidate steps with timestamps and text descriptions, (2) verify time and description for all steps and (3) expand a verified step with before and after images. During the Find step, the task is outsourced to a crowd, where it collects timestamps and text descriptions for possible steps; multiple mechanical turkers are assigned the same task whose results are later merged. Then DBSCAN [6] clustering algorithm is performed, with a timestamp difference as the

distance metric -- this is to ensure that the timestamps are clustered together. At the Verify step, the results from the previous step are verified so that the timestamps and descriptions are finalized. The Expand step collects the before and after images of a step which showcases the effect of a 'step'. Much of this tool was built with knowledge gained from formative studies that indicated the different strengths of using annotations and interface features. For example, results showed that the ability to jump back to a particular segment of a how-to video greatly enhanced learning.

AXIS [4]

Learnersourcing [5] is a type of crowdsourcing technique where the learners contribute new content for future learners. AXIS (**A**daptive **eX**planation **I**mprovement **S**ystem) is a system built around that idea. Its capabilities include providing quality explanations for problems at a similar level than that of an instructional designer. The way it achieves this is by reframing the problem to a reinforcement learning problem called multi armed bandits. In this type of problem, the system is faced with a choice of which action to take and tries to maximize the total cumulative reward over many repetitions. In essence, AXIS asks learners to generate, revise and evaluate explanations as they solve a problem. In this case, the actions to choose from are explanations generated by learners, and the reward for taking the action of presenting an explanation is the learner's rating of its helpfulness. Overtime, this improves the explanation generated by the system and helps learners understand how to solve problems.

Conclusion

Text mining and crowdsourcing techniques have long been applied in information retrieval space (i.e. search engines, recommender systems). It has, however, only been recently integrated with online learning platforms to extend the *learnability of teaching materials*. As we saw in this review paper, tools like ToolScape and AXIS improves the learning experience by integrating different crowdsourcing, data mining and machine learning techniques. Another aspect of these tools is that they were designed with human learners in mind. With ToolScape, we saw the ease of use in jumping between the different segments via the interactive video player. With AXIS, explanations are improved over multiple repetitions and engaged learners to provide feedback and evaluations. Although these applications have given us glimpses of how we could best utilize text mining techniques in the space of education, it has its own limitations. For example, not all the how-to videos are in the same domain nor are AXIS's explanations universally applicable to all subjects; more research needs to be done. However, more and more opportunities are opening up as we infinitely connect ourselves with the Internet. With the rise of educational and commercial platforms like Coursera, and edX,

applications of text mining will undoubtedly be incorporated to improve the learning experience.

References

- [1] Jordan, K. 2015. Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning*. 16, 3 (Jun. 2015). DOI:<https://doi.org/10.19173/irrodl.v16i3.2112>.
- [2] Cook, D. A. (2007). Web-based learning: pros, cons and controversies. *Clinical Medicine*, 7(1), 37–42. doi:10.7861/clinmedicine.7-1-37
- [3] Kim, J., Nguyen, P. T., Weir, S., Guo, P. J., Miller, R. C., & Gajos, K. Z. (2014). Crowdsourcing step-by-step information extraction to enhance existing how-to videos. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*. doi:10.1145/2556288.2556986
- [4] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z. Gajos, Walter S. Lasecki, and Neil Heffernan. 2016. AXIS: Generating Explanations at Scale with Learnersourcing and Machine Learning. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale (L@S '16)*. Association for Computing Machinery, New York, NY, USA, 379–388. DOI:<https://doi.org/10.1145/2876034.2876042>
- [5] J. Kim, “Improving learning with collective learner activity”, Massachusetts Institute of Technology, Cambridge, MA, USA, 2015.
- [6] Ester, M., Kriegel, H.P., Sander, J. and Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).