Anqi Chen
ac89
CS 410
Technology Review

An overview of ConceptScape, its system design, and comparison to content extraction system DynamicSlide

INTRODUCTION

Videos have become an increasingly popular part of online learning experience. However, it is difficult for learners with limited prior knowledge to effectively organize information or locate parts of the video based on their needs. ConceptScape and DynamicSlide are two video processing systems that seek to enhance the video learning experience by enabling reference-based interaction and navigation. The purpose of this review is to examine and compare the two systems' design, and the effectiveness of their outcomes.

CONCEPTSCAPE

Liu et al. explores the idea to engage novice learners to generate high-quality concept maps that would help with comprehension [1]. In general, concept map generation is a complex task that requires precise domain knowledge. ConceptScape intends to solve this problem by designing a crowdsourcing workflow, which consists of three main stages: Concept and Timestamp Generation (Stage 1), where workers collaboratively produce an abundant collection of key concepts covered in a lecture video; Concept Linking (Stage 2), where workers link concepts according to their understanding of lecture structure; and Link Labelling (Stage 3), where workers add descriptive labels to the links. During each stage, there is an adjustment step where workers add/delete/modify the stage results, if necessary.

To assess of the quality of the concept maps created, evaluators are asked to score concept maps generated by individual novices, domain experts, and from ConceptScape. Statistics shows that concept maps generated by novice crowd workers match the quality of those generated by experts, while ConceptScape produce significantly higher quality concept maps than the individual novice learners.

DYNAMICSLIDE

DynamicSlide is a video processing system that targets slide-based MOOC videos. DynamicSlide extracts slides from videos, and links concepts displayed on the slides with corresponding sections of the audio transcript. Using the links it computes, the DynamicSlide player highlights the text item that is currently being explained; it also supports item-based navigation, and in turn, efficient re-watching of the videos.

DynamicSlide processing pipeline has three stages: shot boundary detection (Stage 1), text segmentation within slides (Stage 2), and text-to-script alignment (Stage 3). The goal of the first stage is to extract the set of shots used in a lecture video, by analyzing image and text difference

between consecutive frames. The text segmentation stage groups words detected on the slide into text items using position information of the words, including bounding boxes alignment, distances, and whether bullet points are detected in front of the bounding boxes. In the final stage of processing, researchers link the text items with sentences from the audio transcript, using the cosine similarity between text items and script sentences to create text item - sentence pairs.
For evaluation, the researchers compare automatically generated results against manually created ground-truth results. F1-score of Stage 3 is the highest at 0.79, while F1-score of Stage 2 is the lowest at 0.67 [2].

COMPARISON

Crowdsourcing Workflow vs. Automation

ConceptScape and DynamicSlide take different approaches to the issue of limited expert availability. ConceptScape is designed to allow crowd workers to work on the same video in parallel, and a server to aggregate these workers' output for ranking and subsequent improvement. On the contrary, DynamicSlide's processing stages are all automated. Since instructors organize their lecture videos in a number of various ways, it proves difficult to find a estimation algorithm that would work with a large collection of videos. The text-segmentation (Stage 2) shows the lowest accuracy of 67% [2]; in addition, researchers point out that videos with low performance have unusually large font sizes where the algorithm pipeline do not accommodate. In shot boundary detection (Stage 1), researchers notice that a particular video that frequently alternates between the instructor headshot, the instructor alongside the slide view, and the full slide video has a very low F1-score of 0.39. While the model can be further improved by fine tuning the parameters, we can remedy this problem by involving human effort. A human, even a novice learner, would be able to easily recognize logically valid text groupings. To minimize human workload, we can have novice learners evaluate automatically generated results and improve them as in ConceptScape.

Item-Based Navigation

ConceptScape and DynamicSlide both offer item-based navigation. In both cases, participants found the feature useful [1, 2]. For DynamicSlide, when the slides are over-simplified (i.e., when information is only present in the script), participants do not find the system helpful in finding information. Here, we see once again how the two systems relate to each other: while some video lectures offer lecture slides, sometimes they do not include information that is crucial to understanding in the slides. Since DynamicSlide uses text items that are the slides without any adjustment, its performance depends on the organization of the lecture slides. ConceptScape, on the other hand, collects concepts/terms from humans who watch the video. No matter how the slides are organized, or whether slides are provided at all, ConceptScape's crowdsourcing workflow allows a concept map to be generated from like-minded learners. ConceptScape has a broader applicability than DynamicSlide in this regard.

CONCLUSION

In summary, ConceptScape and DynamicSlide both aim to help online learners by incorporating a reference-based interactive system. ConceptScape engages novice crowd workers to collaboratively produce a concept map that reportedly reinforces learners' memory and understanding [1]. DynamicSlide builds an automatic pipeline that extracts references from slide-based lecture videos. ConceptScape has a broader applicability due to its crowdsourcing workflow yielding concept maps s of comparable quality with respect to expert-generated ones. DynamicSlide system may be improved by involving novice learners for fine adjustment.

References

1. Ching Liu, Juho Kim, and Hao-Chuan Wang. 2018. ConceptScape: Collaborative Concept Mapping for Video Learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 1-12. https://doi.org/10.1145/3173574.3173961
2. Hyeungshik Jung, Hijung Valentina Shin, and Juho Kim. 2018. DynamicSlide: Exploring the Design Space of Reference-based Interaction Techniques for Slide-based Lecture Videos. In *Proceedings of the 2018 Workshop on Multimedia for Accessible Human Computer Interface (MAHCI'18)*. ACM, New York, NY, USA, 33-41. https://doi.org/10.1145/3264856.3264861