

Video-based Big Data Analytics in Cyberlearning

Shuangbao (Paul) Wang
Metonymy Labs
paul.wang@computer.org

William Kelly
Metonymy Labs

ABSTRACT. In this paper, we present a novel system, inVideo, for video data analytics, and its use in transforming linear videos into interactive learning objects. InVideo is able to analyze video content automatically without the need for initial viewing by a human. Using a highly efficient video indexing engine we developed, the system is able to analyze both language and video frames. The time-stamped commenting and tagging features make it an effective tool for increasing interactions between students and online learning systems. Our research shows that inVideo presents an efficient tool for learning technology research and increasing interactions in an online learning environment. Data from a cybersecurity program at the University of Maryland show that using inVideo as an adaptive assessment tool, interactions between student–student and student–faculty in online classrooms increased significantly across 24 sections program-wide.

Keywords: e-learning, video index, big data, learning analytics, assessment

1 INTRODUCTION

Learning technology practitioners and researchers face the challenges of analyzing enormous amounts of digital data in the new cultures of learning that are emerging, especially videos on Massive Open Online Courses (MOOCs). Attempting to address the challenge of initial analysis and selection of digital video data can provide significant benefit for learning and research in the field.

Big data analytics are used to collect, curate, search, analyze, and visualize large data sets generated from sources such as texts (including blogs and chats), images, videos, logs, and sensors (Bakshi, 2012). Video data is a major format of unstructured data, and should be an indispensable area of big data analytics. However, most analytics tools are only able to use structured data. Due to the nature of the special file format, traditional search engines cannot penetrate into videos, and therefore video indexing becomes a problem.

Videos contain both audio and visual components, and neither of these components is text based. To understand a video, viewers must actually play it and use their eyes and ears to analyze the sounds and visuals being presented to them. Without watching a video, it is hard to glean information from its content or even know whether there is information to be found within. Existing search engines and data

analytics tools such as Google, SAS, SPSS, and Hadoop are effective only in analyzing text and image data. Video data, however, are difficult to index and therefore difficult to analyze.

In education, video presents a large opportunity for both classroom and online education (Rideout, Foehr, & Roberts, 2010). In addition, video is a great teaching format because it can both be more enjoyable and more memorable than other instruction formats (Choi & Johnson, 2005). Furthermore, video instruction allows for students to work at their own pace, for teachers to be able to teach more students, and for more reusable teaching materials to be available when compared to an in-person lecture. MOOC creators realize the many benefits of video, as evidenced by the prevalence of video in MOOCs. Many MOOCs focus on video files for the bulk of their instructional material so it is clear that the MOOCs of the future must also focus on videos.

Structuring video data and accurately modelling relevant metrics has value for future courseware applications, especially MOOCs. A MOOC can be enhanced or limited by the design of the user interface, as well as the analytics by which the MOOC is assessed. InVideo provides rich data on videos, turns them into more effective and interactive learning objects, and improves MOOCs. One teacher cannot possibly personally oversee the development of thousands of students with any effectiveness, and therefore, effectively leveraging technology to develop effective analytics is essential to the effectiveness of MOOCs. Yet, the problem of education at such a large scale is a relatively novel problem. The wide use of applications of inVideo could potentially address this novel and growing problem.

InVideo (Wang & Behrmann, 2010), developed under a US Department of Education grant, is able to analyze video content (language and video frames) prior to initial close researcher review of the video. A highly efficient video indexing engine can analyze both language and video frames based on natural language and referent objects. Once a video is indexed, its content becomes searchable and statistical analysis as well as qualitative analysis are possible. Commenting and tagging add a layer of interaction between students and online courses. They also increase the accuracy of the transcript, which was automatically extracted from the video by the inVideo tool. The indexing technology is especially useful in mining video data for learning events. InVideo can promote initial data selection, analyzing video data in two different ways: one is to find keywords that were spoken in the video; another way is to identify an object in the video if a reference picture is provided. inVideo also has an automatic caption system that can transcribe the words spoken in the video. Instructors can use the tool to construct in-place video quizzes for assessments.

Learning is an integration of interaction. The interaction might exist between learners and instructors or between learners and computers. While the traditional approach would be to analyze grades at the end of the semester, this lacks the benefits that come from interactions that occur during the course (Elias, 2011). As an increasingly large number of educational resources move online, analyzing interactions between students and online course material is becoming more important. Many learning management systems (LMSs) have built-in learning analytics tools to look into the data. Due to the limitation of the

data gathering and indexing, the built-in tools are generally not sufficient for assessing study outcomes, especially for video content.

2 RELATED WORK

Ángel, Javier, Pablo, and Baltasar (2012) proposed a way to track student interactivity by logging their interactions with educational video games. Their conclusion was that even simple semi-automatic tracking, which uses both computer and human, shows advantage compared to most video-related systems that are fed lower quality data.

Big data and learning analytics can become part of the solutions integrated into administrative and instructional functions of higher education (Picciano, 2012). Traditional face-to-face instruction supports traditional data-driven decision-making process. Videos as a form of big data are more extensive and especially time-sensitive learning analytics applications. It is important that instructional transactions are collected as they occur.

Learning analytics can provide powerful tools to support teachers in the iterative process of improving the effectiveness of their course and to collaterally enhance their students' performance (Dyckhoff, Zielke, Bültmann, Chatti, & Schroeder, 2012). Dyckhoff developed a toolkit to enable teachers to explore and correlate learning object usage, user behaviour, as well as assessment results based on graphical indicators. This learning analytics system can analyze data such as time spent, areas of interest, usage of resources, participation rates and correlation with grades data and visualize them using a dashboard. However, the system is unable analyze the interactions between students and the online learning systems on videos.

Haubold and Kender (2004) introduced techniques for extracting, analyzing, and visualizing textual contents from instructional videos. They obtained transcripts from videos of university courses. Using the information, they indexed the transcripts and displays in graphs that help in understanding the overall course structure. Unfortunately, the index process only takes static data, no interactive data are included.

In order to improve interactions between students and online course material, especially videos, we have developed a video index engine to look at every word spoken in the video and categorize it using our custom index algorithm. In addition, a content-based pattern recognition engine can search individual frames of the video to recognize objects and individuals being displayed. The collaborative commenting, tagging, and in-place quizzes make videos more accessible and also increase the accuracy of the search engine.

3 VIDEO INDEXING AND SEARCH

Videos are a different data type than text and images, in that they are unstructured data. Traditional search engines are mostly text based, with a few tools that allow for searching of images. In order to index a video, a search engine needs to extract meaningful language from the audio and convert it to

text, while simultaneously converting the visual frames into a series of images that can be used to recognize persons and objects in the video. This is an extremely difficult task, given that videos are a compound format. Not only are the audio and visual components integrated, but also within each of these components there is a blend of information being presented in a manner that cannot be distinguished as easily by a computer as by a human brain. For example, the audio of the file may contain speech, music, and background noises that a computer will have a hard time recognizing and analyzing.

3.1 Automatic Indexing Algorithm

The video indexing engine uses the vector space model to represent the document by a set of possible weighted content terms. The weight of the term reflects its importance in relation to the meaning of the document (Wang, Chen, & Behrmann, 2004). After calculating the normalized frequency of a term in the document, the weight to measure the relative importance of each concept or single term is obtained. The automatic index algorithm then calculates the final position in n-dimensional space. The result is to be used for generating search results or visualization.

3.2 Searching Videos by Keywords

Video search involves two steps: analyzing by keywords and analyzing by image references. When a keyword is entered, the system looks through the indexed audio transcript to see if there is a match. An image reference may refer to either a picture or keywords that describe an object in the video using an appropriate semantic space. Video clips whose language contains the keywords will be retrieved. Figure 1 shows how indexed videos can be searched using keywords in the spoken language.

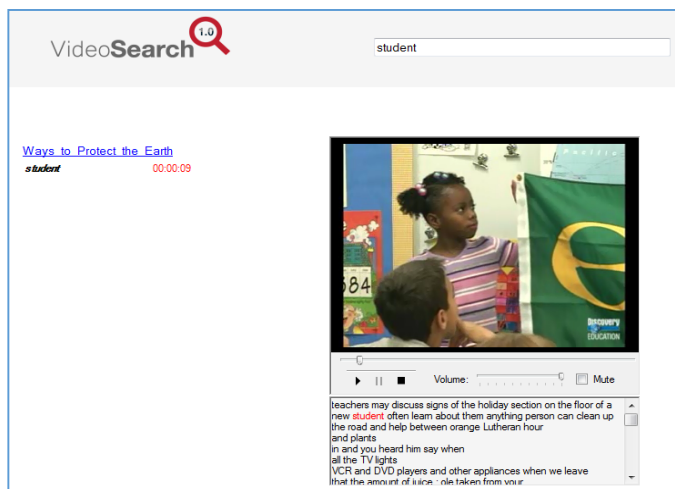


Figure 1: Analyzing videos by keywords.

3.3 Searching Videos by References

Searching videos by references examines the frames of the video to see if the given picture or keyword is found. If the reference is a picture, then the system uses a Content Based Image Retrieval (CBIR)

(2017). Video-based big data analytics in cyberlearning. *Journal of Learning Analytics*, 4(2), 36–46. <http://dx.doi.org/10.18608/jla.2017.42.5>

algorithm to find the match frames and return the video clips that contain the reference picture. Figure 2 shows the image-based CBIR algorithm that retrieves the video frames corresponding to the reference picture (at the bottom).



Figure 2: Analyzing videos by image references using the CBIR algorithm.

If the reference is a keyword (e.g., “credit card”) then the system uses a knowledge tree to find matches in the video. If one video frame contains an object matching the features associated with the keyword, the section of the video is returned. Figure 3 shows how a search for the keyword “credit card” will retrieve the video frames that contain objects as credit cards.

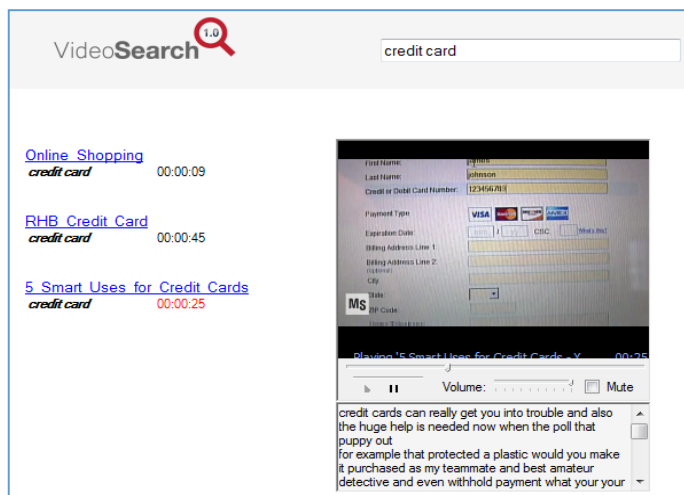


Figure 3: Analyzing videos by keyword references using a knowledge tree.

3.4 Searching Videos with Multiple Languages

Sometimes multiple languages may be found in videos. Transcribe engines normally only work in one language or in closely related languages. For other languages, a different transcribe engine may be required. InVideo addresses this problem by allowing videos with different languages to be searched from a single user interface. The inVideo system does not translate between languages. It only transcribes based on the language of original videos. For example, a Chinese video will result in a transcript in Chinese.

Figure 4 shows the indexing engine properly analyzing the Chinese language. When entering the word “student” in Chinese, the video search engine will locate that term in the transcript and return the corresponding frames. Currently there are multiple languages that can be analyzed by the inVideo system, with more to be added.



Figure 4: Analyzing videos with different languages.

4 COLLABORATIVE AND INTERACTIVE LEARNING

The ability to take unstructured video files and bring structure to the data embedded in them greatly enhances the value of any MOOC. Because MOOCs make heavy use of video as a medium for providing instruction, the ability to search video content to create, access, and organize the data contained within the videos is paramount. The information provided by instructional materials holds less value if students are unable to access it properly, and thus inVideo’s ability to index and structure video-based instruction will provide great value to a MOOC’s effectiveness.

Automatically generated video transcripts may have accuracy problems. Besides, the vast numbers of videos in MOOCs make them impossible to be retrieved correctly with just one or a few simple keywords. To solve these problems, we have implemented a collaborative filtering mechanism including commenting, tagging, and in-place quizzes. These features improve accuracy and increase interaction

between students and the online learning system. With collaborative filtering, learning resource retrieval on MOOC systems is greatly improved, and better student achievement is therefore expected.

4.1 Collaborative Filtering

Collaborative filtering is a process of improving the accuracy of the automatic indexing algorithm by leveraging user feedback. This is popular on websites that have millions of users and user-generated content. Users are able to create time-stamped comments on videos. These comments can be hidden or made public so that someone else who views the video can see the comment at a specific time. These comments help increase accuracy of the search tool and transcript and enhance interactions in online learning.

Tagging on videos is another implementation in the inVideo system, attaching keyword descriptions to identify video frames by category or topic. Videos with identical tags can then be linked together, allowing students to search for similar or related content. Tags can be created using words, acronyms, or numbers. This is also called social bookmarking.

A search term usually yields many related results, which in many cases are hard to differentiate. Commenting and tagging add additional information, refine the knowledge, and increase the video search accuracy. Since information is growing exponentially, these features are extremely helpful for students to obtain the knowledge in the least amount of time.

4.2 In-place Assessment with iQuiz

Internet computing has the advantage of employing powerful CPUs on remote servers to provide applications across the network. inVideo comes with an internet computing-based video quiz system (iQuiz) to utilize the computational power of remote servers to provide video quiz services to users across the internet.

Currently, videos are mostly non-interactive, therefore there are no interactions between students and the learning content. Students view videos either online or download them to their personal devices. There is no way for educators to know whether a student has understood the content or even to know whether the student has viewed the video.

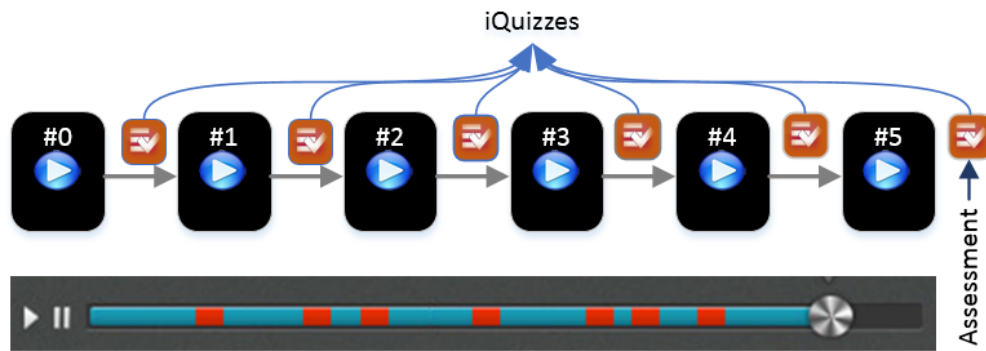
iQuiz can be used to assess learning outcomes associated with video study. Quizzes can be embedded into videos at any place where an instructor wants to assess the outcome of the student's study. iQuiz runs as a service on servers. This enables users to execute this resource-intensive application with personal computers or iPads, which would not be possible otherwise.

Instructors can enter the authoring mode where they can write quizzes by indicating the start and stop positions on the video and adding questions. Video quizzes are stored in XML format, and are automatically loaded while students are watching the video in the learning mode. Answers to the quizzes, either correct or incorrect, are also stored in the XML database for immediate assessments.

Assessment of adaptive learning on videos provides better outcomes for students than the traditional video content study with little or no feedback (Wang & Behrmann, 2009).

4.3 Transform Linear Videos into Interactive Learning Objects

Video is linear in nature. It is hardly interactive nor does it contain branches. Using the inVideo tool, classical videos can be transformed into a series of video clips with assessments in between and at the end. So the video-based learning material becomes interactive. Figure 5 shows a test we conducted that turned a 46-minute video into six selected 2–3 minute video clips. The red segments on the stage bar are the samples. So it is clear that not all video content was used in the samples.



A long video is “cropped” into 2-3 min video clips with assessments in-between

Figure 5: Transforming linear videos into interactive learning objects.

5 RESEARCH RESULTS

To test whether the inVideo system improves learning, we selected the 20 most recent videos from the National Science Digital Library (NSDL) in cybersecurity and used the inVideo tool to extract keywords that appeared in the transcripts. From this set, we selected the top two ranked keywords: “Target” (data breach) and “encryption” (using encryption to secure data). We were confident that those two keywords made good discussion topics that could increase classroom interaction.

As a result, we added two discussion topics to the spring 2014 Masters of Science in Cybersecurity program (24 class sections with an average of 25 students in each section).

Videos lack interactions between learners and the online learning environment. Even worse, videos above a certain length will likely never be watched at all because students cannot easily determine what content is within it or how to locate that content. To address this issue, we used the inVideo tool to index the content and break the large videos into a series of small video clips. By doing so, we made it possible for students to watch short video clips covering individual key concepts directly, while retaining the ability to view the whole video if necessary. This served not only to increase student interest and engagement in the lesson but also more importantly to improve their ability to comprehend and retain information.

Student responses and interactions can be used as a proxy for their degree of engagement with any particular part of the course. As one example of how the inVideo indexing served to increase this measure, consider Week 2 of the class. In our assessment of past offerings (pre-inVideo) we discovered that this part of the course is a “quiet” week, because the individual assignment starting in the week will not be due until Week 8. This meant that the interactions in the classrooms dropped significantly from Week 1. Based on this assessment, we decided to use the inVideo intervention in an attempt to generate more interaction during Week 2 of the course.

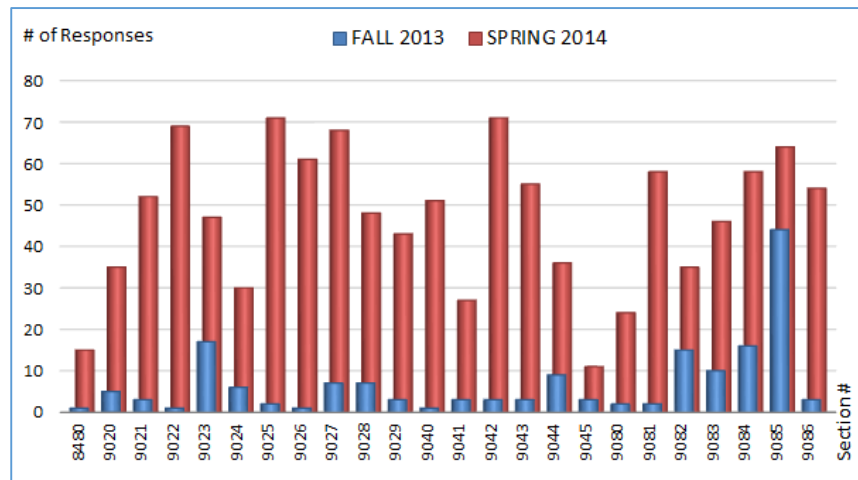


Figure 6: Number of responses for 24 sections — Week 2 discussions.

Our initial observation of one class was very promising; the total number of responses — defined as a posting after viewing a video clip — for Week 2 reached 68, as compared to only 2 for the same week in the previous semester. This initial finding encouraged us to investigate the results for all 24 sections program-wide. Figure 6 shows the number of responses for the 24 sections comparing Fall 2013 to Spring 2014 during Week 2. For the research we conducted, Week 2 student responses across the 24 sections were almost seven times higher during Spring 2014 (1,129 responses) than during Spring 2014 (164 responses).

For the cybersecurity online/hybrid class, we have five graded discussions, one individual assignment, one team assignment, and two lab assignments. Two more hands-on exercises (labs) have been added since Spring 2014. Data from the team projects, using the same intervention method, show that student–student and student–faculty interactions were 6.5 times greater for the courses with the inVideo intervention (104 responses compared to 16 responses). We also measured student performance against desired learning outcomes. The average grades on both team projects and final grades was higher in Spring 2014 than in Fall 2013. Here we see that the index and data analytics tool inVideo, in combination with just-in-time assessment and intervention, improved learning outcomes.

Based on our finding, we are in the process of breaking up every large learning module into several learning objects using inVideo. The new competency-based learning objects will be used to construct the

knowledge cloud. These new learning modules will consist of many competency-based learning objects, and will be more interactive, rational, and accessible.

We will use inVideo to expand the scope of this research to other activities in courses within the cybersecurity program. This tool could also be useful to courses in other disciplines. Using the inVideo tool, linear videos are transformed into a series of interactive learning objects. This is vital in an online learning environment where interactions and learning outcomes are valued the most.

6 SUMMARY

This paper discussed a novel learning analytics tool to analyze video data in an online learning environment and use of the tool to analyze data generated from classrooms. Video indexing engines analyze both audio and visual components of a video, and the results of this analysis provide novel opportunities for search. Indexed videos can be further used in assessing learning outcomes to collaboratively comment, tag, and create video quizzes.

Learning analytics based on indexed video can be generated in a number of ways: first by analyzing keywords that appear in the audio track; second, by analyzing people or objects that appear in the video frames; third, by analyzing these objects based on descriptive keywords; and finally by analyzing with different languages. This technology is especially useful when it comes to mining video data in a learning environment.

To improve accuracy, we can either improve the transcribe engine, analyze video frames better where there is no audio, or crowd-source accuracy through collaborative filtering. For transcription accuracy, one potential accuracy improvement can come from using a self-learning artificial intelligence (AI) system that could be taught to recognize certain accents or languages. The process or requirements for instituting such a system and the magnitude of the improvement in accuracy are yet to be studied.

Further research in profiling will increase accuracy. Making inVideo a web API by allowing commenting, tagging, and using cloud computing technology to add more user interactions will make the application more applicable to various users.

At present, the inVideo tool is only limited to analyze native (non-streaming) videos. Since we are using many streaming videos from various sources in the courses, adding a streaming video analysis feature would be very helpful for the online classroom data analytics and assessment.

The initial assessment and intervention yielded significant improvements in student interaction in the cybersecurity classrooms. The activities and responses in classrooms increased, student–student and student–faculty interactions enhanced, and the grades for team projects and exams both improved.

ACKNOWLEDGEMENT

This research is funded in part by a grant from the National Science Foundation, NSF-1439570.

REFERENCES

- Ángel, S., Javier, T., Pablo, M., & Baltasar, F. (2012). Tracing a little for big improvements: Application of learning analytics and videogames for student assessment. *Proceedings of the 4th IEEE International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES '12)*, 28 October 2012, Genoa, Italy (pp. 203–209). <http://dx.doi.org/10.1016/j.procs.2012.10.072>
- Bakshi, K. (2012). Considerations for big data: Architecture and approach. *Proceedings of the IEEE Aerospace Conference*, 3–10 March 2012, Big Sky, MT, USA (pp. 1–7).
- Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *The American Journal of Distance Education*, 19(4), 215–227. http://dx.doi.org/10.1207/s15389286ajde1904_3
- Dyckhoff, A., Zielke, D., Bültmann, M., Chatti, M., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society*, 15(3), 58–76.
- Elias, T. (2011). Learning analytics: Definitions, processes and potential. Retrieved from <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
- Haubold, A., & Kender, J. R. (2004). Analysis and visualization of index words from audio transcripts of instructional videos. *Proceedings of the IEEE Symposium on Multimedia Software Engineering*, 1–3 March 2004, Norfolk, VA, USA (pp. 570–573).
- Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, 16(3), 9–20.
- Rideout, V., Foehr, U., & Roberts, D. (2010). Generation M2: Media in the lives of 8- to 18-year-olds. Kaiser Family Foundation. Retrieved from <http://kaiserfamilyfoundation.files.wordpress.com/2013/01/8010.pdf>
- Wang, S., & Behrmann, M. (2009). Automatic adaptive assessment in mLearning. *Proceedings of the International Conference on Cognition and Exploratory Learning in Digital Age (IADIS CELDA 2009)*, 20–22 November 2009, Rome, Italy (pp. 435–438).
- Wang, S., & Behrmann, M. (2010). Video indexing and automatic transcript creation. *Proceedings of the 2nd International Conference on Education Research*. Retrieved from http://kihd.gmu.edu/assets/docs/kihd/activ/new/Wang_Behrman_press.pdf
- Wang, S., Chen, J., & Behrmann, M. (2004). Visualizing search engine results of data-driven web content. Retrieved from <http://www.w3.org/WAI/RD/2003/12/Visualization/VisSearch/VisualSearchEngine.htm>