Visualizing Lecture Capture Usage: A Learning Analytics Case Study

Christopher Brooks
Department of Computer Science
University of Saskatchewan
Saskatoon, SK, Canada
cab938@mail.usask.ca

Craig Thompson
Department of Computer Science
University of Saskatchewan
Saskatoon, SK, Canada
craig.thompson@usask.ca

Jim Greer
Department of Computer Science
University of Saskatchewan
Saskatoon, SK, Canada
jim.greer@usask.ca

Abstract
This paper outlines our initial investigations of applying information visualization techniques to lecture capture video systems. Our principal goal is to better understand how students use these systems, and what visualizations make for useful learning analytics. We apply three different methods to viewership data aimed at understanding student rewatching behaviour, temporal patterns for a single course, and how usage can be compared between groups of students.

Author Keywords
Lecture Capture; Information Visualization; Usage Data; E-Learning

ACM Classification Keywords
K.3.1 [Computers and Education]: Computer Uses in Education.; H.5.1 [Multimedia Information Systems]: Video (e.g., tape, disk, DVI).

Introduction
The activities described in this paper fall into a larger program of research focused on determining the efficacy of lecture capture on student learning. We are interested in using both statistical techniques (e.g. machine learning) and visual techniques (e.g. information visualization) to understand the patterns of interaction

learners have with lecture video. This paper introduces three activities we have undertaken to visualize learner data. The data for this paper comes from Recollect, a lecture capture environment we developed for research use in 2010 and 2011. Described more fully in [2], the data we use is generated by the Recollect *heartbeat*, a client-side event that occurs every 30 seconds and records (among other things) which video a user is playing, the location of the playhead within the video, and whether the video is paused or not. This structure has been ported to the freely available Opencast Matterhorn project[1], which is the focus of new research on lecture capture analytics at our institution.

Case Study: Chemistry 200

One course we examine in this work is a second year undergraduate Chemistry course. This course was taught using traditional lectures to 546 students in multiple sections. One instructor elected to have his lectures recorded and provided to all of the students as a study aid. Examinations and assignments between sections were the same.

Only 333 watched video content for at least five minutes, a participation rate of 61%. In total, learners watched over 77 days worth of lecture video, a remarkable number given that only 38 lectures of 50 minutes each were recorded.

Case 1: Visualizing Rewatching Behaviour

Motivated by the correlative link between re-reading discussion forum messages and academic performance in undergraduate students [1], we were in better understanding whether learners re-watch lecture capture content. Summary statistical usage of our lecture capture facilities shows that some students playback significantly more video than others (see sidebar). Our question is whether there are any meaningful patterns in how learners view lecture videos.

To visualize lecture watching activity, we plotted all of the sessions for an individual user/video pair using a scatter plot, with the x axis representing time within a video playback session in minutes, and the y axis representing the time within the lecture that was being watched. A $45°$ diagonal line represents viewing the video without pausing, while horizontal lines indicate the video was paused. By colour coding sessions and overlaying their plots on top of one another, we can identify how many times a student has watched a particular portion of video by summing the number of diagonal lines over a region of the y dimension. If learners spend a session watching some video and "pick up" from where they left off, there will be no gap in the y dimension. If they re-watch from content they have previously seen there will be some overlap of the diagonal lines, and if they skip ahead there will be space in between each diagonal in the y dimension.

[1] http://www.opencastproject.org

Figure 1 shows graphs of student interaction data using three different students and three different videos. The first image, Figure 1a, shows many long diagonal lines indicting the learner tends to navigate to the position of interest and watch for an extended period of time. These lines overlap heavily along the y dimension, indicating the student has rewatched significant portions of the video, some up to four times. There are also long horizontal lines indicating the student has paused the video for extended periods of time. Due to the strong diagonals, we have labelled this activity as *regular rewatcher*.

Figure 1b show a different kind of student, who has viewing sessions with a low slope, indicating they watch the video in less than real time. Our system didn't allow for variable speed playback, and deeper investigation shows that the learner both pauses the video frequently and slowly seeks through the video, perhaps spending time on particular segments to transcribe content or apply it to a problem they may be working on. Despite this, the learner has both rewatched one section of the video completely (given by the two overlapping lines), and ended up playing back all of the video content. Due to the perceived activity of the learner, we refer to this student as a *engaged rewatcher*.

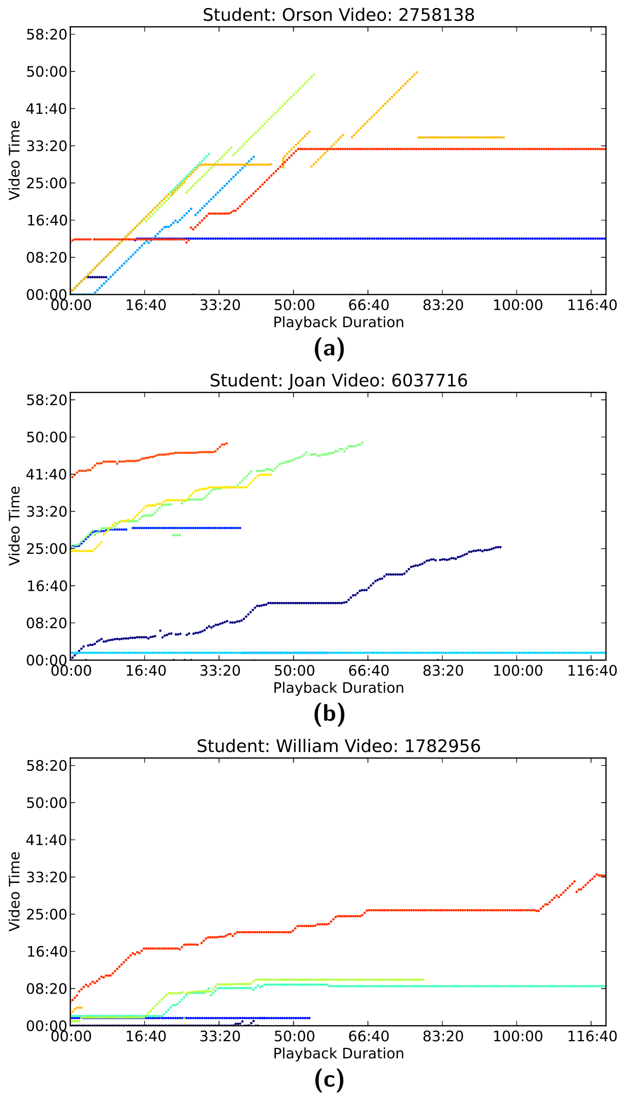

Figure 1: Rewatching graphs for three different learners.

Finally, figure 1c shows a learner who makes strong use of the pause feature of the player. While it was surprising to us, it is not uncommon to find learners who open multiple videos at once and then pause video for hours, even days, and come back to continue playing the video. This learner exemplifies this behaviour, with many sessions containing long horizontal lines, some turning into diagonals after extended periods. While our visualizations are truncated to 120 minutes in order to maintain a 1:1 ratio between axes, there are many learners who tend to follow this *pauser rewatcher*.

Case 2: Temporal Patterns of Viewership

In addition to understanding how learners watch individual videos, we are interested in understanding how learners watch videos over the span of the course. Previous work has been done by others on visualizing intravideo navigation [4], but we're interested in intervideo usage patterns, such as periods of high activity prior to examinations and assignments. We are motivated in part by our previous work, which has demonstrated that there is a statistically significant positive correlation between academic achievement and habitual weekly viewing of lecture videos [3].

To visualize temporal intervideo usage, we created three dimensional heatmaps for each course. These maps plot the time a video was made available to students (y axis), the time at which students watched that video (x axis), and the the total time that video was watched (colour axis). Data was binned to one day intervals and, while we did not hold the axes equivalent as we did in the previous example, a strong diagonal line represents learners watching lectures as soon as they become available, and the more filled in the lower right triangle of usage is, the more lectures were revisited.

While a number of different patterns were observed, here we present three distinct patterns. The first kind of course shown in Figures 2a and 2c demonstrate the most typical pattern of interaction. Here, learners rewatch lectures from earlier in the term as evaluations (midterm and final exams) approach. While there is some watching of early lecture content throughout the term, we notice large patches of blue (cool, or minimal) usage of content that was recorded before the midterm evaluation once the midterm has been delivered. While there is some viewership of this content right before the final exam, this activity is minimal.

The second kind of course is shown in Figure 2b, that has particularly strong viewership of only the most recent video right before the final examination. This suggests a very important lecture at the end of course was given (perhaps a comprehensive review of topics), or that topics from early on in the course will not be tested. As this course was an introductory programming course in Computer Science, it is quite likely that the early portion of the course focused on fundamental skills, while the latter half of the course required execution of these skills in programming assignments (leading to a reduction in watching of video content).

The last kind of course, an introductory Calculus class for non-majors, involved regular forms of evaluation spaced roughly every two weeks. Here, viewership patterns follow a diagonal band, where early content is rarely watched later on in this course. This suggests to us that the content either comes in distinct "chunks" which are unrelated, or that early content is fundamental to the later content and doesn't need to be revisited as the course progresses. This pattern is most interesting to us given our previous studies indicating regular viewership of captured lectures is correlated with academic achievement. In this course we don't observe the same amount of "cramming" when regular evaluation is applied, and we are interested in determining if it is the domain that causes this change, or the pedagogical approach.

Throughout most of these visualizations, we see a trend of high viewership (high temperature colour) early on in the course, at evaluation points, and at the end of the course.

Case 3: Comparing Usage Between Groups of Students

Our final visualization aims to shed light on how students in different identifiable groups use lecture capture facilities differently. We are particularly interested in comparing high achieving learners (those who achieved an exceptional pass of the course with a 87.5% or higher grade) to low achieving learners (those who achieved are marginal pass of the course with a 50% – 62.5% grade).

We plotted a histogram of viewership activity for each group. We concatenated the lengths of each lecture together[2], to form a continuous x axis of 28.5 hours of video. Each 15 minutes along the axis denotes a single histogram bin, viewership for that bin is equal to the number of heartbeats we would expect if every student in that group watched the whole 15 minutes (e.g., 30 heartbeats per student). This captures both initial watching and rewatching behaviour, but not behaviour where the video is paused. The two histograms were then plotted simultaneously with alpha channelling to see commonalities – in Figure 3 the red plots indicate viewership by high achieving students, the blue plots indicate viewership by low achieving students, and the shared viewership patterns are shown in purple.

[2] For data cleaning we limited the length of a lecture to 45 minutes.

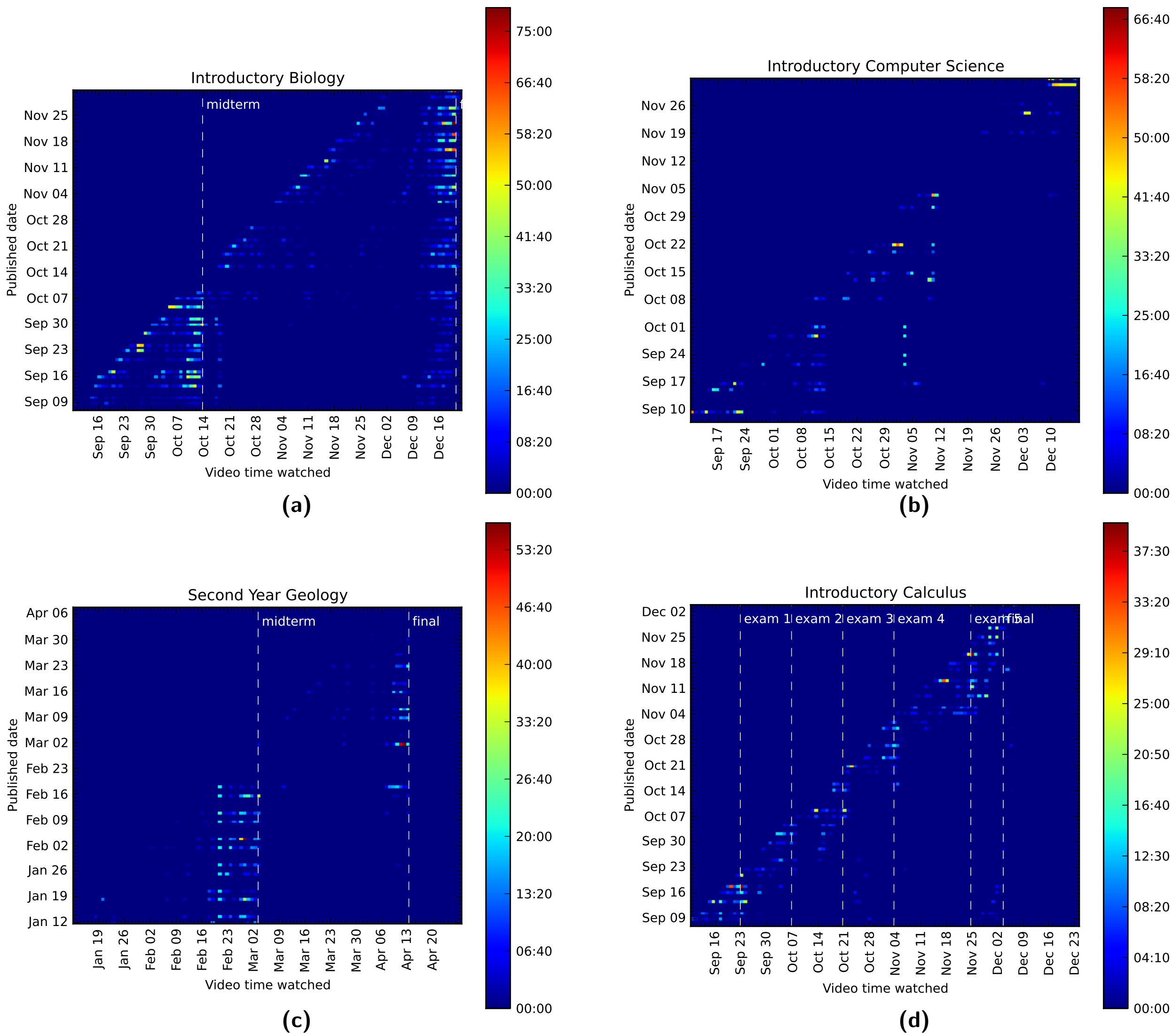

Figure 2: Heatmaps for four cohorts demonstrating the relationship between publication date of video and viewing date by students.

Conclusions

Through these visualizations, we have been able to gain insight into how learners used lecture capture, how this aligns with activities over an academic term, and how student populations differ in their use of lecture capture systems. Applying visual analytics to "big data" problems is not without caveats – the effects of parameters for charts including time offsets, resolution on heartbeat data, aggregation into bins for heatmaps and histograms, and determining the right data to process make discussion and prototyping essential steps in the process.

How to provide visual learning analytics to different stakeholders is also an issue we are carefully considering. Currently, our work is aimed at instructional designers and instructors who are deeply interested in their courses. We are interested in also showing visualizations to students and instructors to help them gain insight into their learning and teaching and how that relates to usage of technology like lecture capture.

References

[1] J. Bowes. Knowledge management through data mining in discussion forums. Master's thesis, University of Saskatchewan, 2001.

[2] C. Brooks, C. D. Epp, G. Logan, and J. Greer. The who, what, when, and why of lecture capture. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, pages 86–92, New York, New York, USA, 2011. ACM Press.

[3] Christopher. *A Data-Assisted Approach to Supporting Instructional Interventions in Technology Enhanced Learning Environments*. PhD thesis, University of Saskatchewan, 2012.

[4] P. B. Robert Mertens, Markus Ketterl. Social navigation in web lectures: a study of virtpresenter. *Interactive Technology and Smart Education*, 7(3):181–196, 2010.

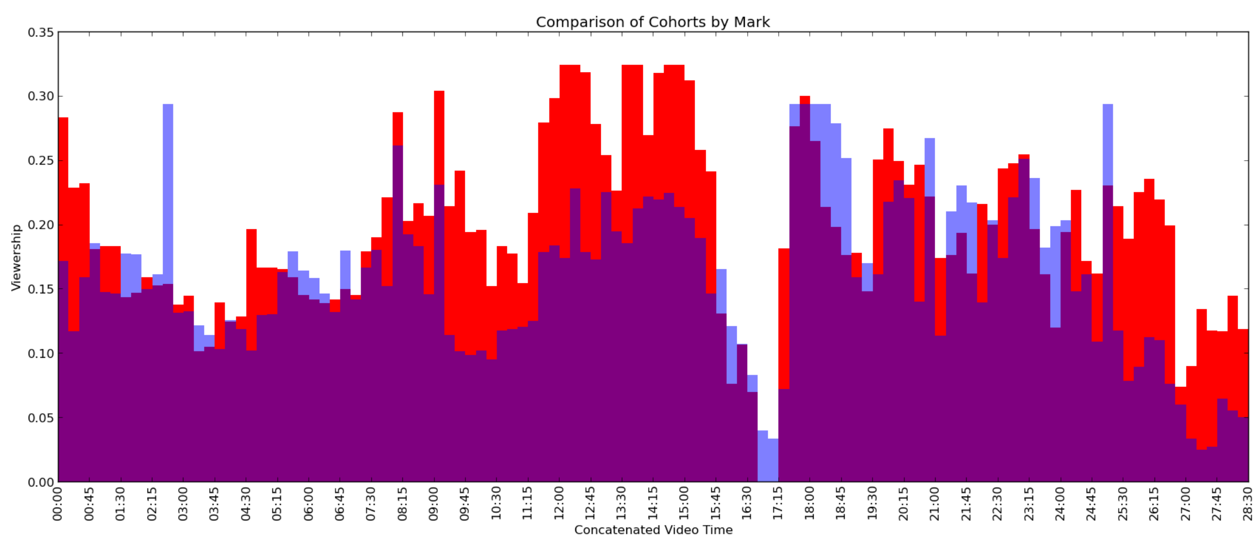

Figure 3: Overlapping histograms showing the viewership of high achieving students (red), low achieving students (blue) and that viewership common to both groups (purple). Videos of the *Chemistry 200* course were used for this plot.