Computers & Education 75 (2014) 282–292

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Modelling and quantifying the behaviours of students in lecture capture environments

Christopher Brooks*, Graham Erickson, Jim Greer, Carl Gutwin

Laboratory for Advanced Research in Intelligent Educational Systems (ARIES), Department of Computer Science, University of Saskatoon, 110 Science Place, Saskatoon, SK, Canada

A R T I C L E I N F O

Article history:
Received 29 July 2013
Received in revised form
4 March 2014
Accepted 6 March 2014
Available online 15 March 2014

Keywords:
Lecture recording
Lecture capture
Podcasting
Media in education
Learning analytics

A B S T R A C T

The literature is mixed as to whether the addition of lecture capture technologies provide for better student success. In this work, we consider not just the broad effect of lecture capture technology on academic achievement between cohorts, but whether this effect is related to patterns of viewership among learners. At the centre of our interest is determining whether there are strategies learners take in their reviewing of content week-to-week that may result in better achievement. To investigate this, we describe a method for modelling learners based on their interactions with lecture capture systems. Unlike investigations done by others, our models emerge from the activities of the learners themselves, and are based on the results of applying unsupervised machine learning (clustering) techniques to student viewership data. These models describe five different classifications of learner interactions, and we show that one of these is positively correlated with academic achievement. We further validate our results through repeated experimentation, and describe how such models might be used by early-alert systems.

1. Introduction

Over the last ten years there has been a shift in classroom recording away from expensive, studio-based televised solutions primarily used for distance education, to inexpensive automated solutions used to augment the traditional on-campus student experience. This shift has come both because of increased expectation of video resources from students (e.g. who are used to social video technologies such as YouTube) as well as the increased ability to scale recording technologies throughout the institution (through automation technologies and cost reduction of video recording devices). Instead of small courses being offered with professional television crews and videographers supporting them, dozens of courses in different disciplines and programs can be simultaneously recorded, processed, and made available to learners using inexpensive automated systems.

This change in the use of video technology brings with it both challenges and opportunities. One of the challenges that occurs with low cost solutions is that pedagogical approach and effectiveness of the technology is often poorly understood. Since the costs are low there is minimal motivation to determine how well the technology helps students. Many deployments in the field are led by technicians as opposed to instructional designers, who are often more involved in higher-cost educational support activities. At the same time, lecture capture learning environments, like learning content management systems, bring with them unprecedented opportunities to gather information on learner behaviours. Being able to log student actions in the learning environment allows educational researchers to compare cohorts of learners to one another based on learning activity and educational outcome. Such activity has immediate benefits to instructors and students who use these technologies, as it provides a deeper understanding of whether a technology is valuable to the teaching and learning process.

While there are a number of interesting educational issues with respect to lecture capture technology, such as student motivation, classroom attendance, and pedagogical techniques, this work focuses on the issue of whether there are patterns of viewership of

* Corresponding author.
E-mail addresses: cab938@mail.usask.ca, brooksch@umich.edu (C. Brooks), gke703@mail.usask.ca (G. Erickson), greer@cs.usask.ca (J. Greer), gutwin@cs.usask.ca (C. Gutwin).

http://dx.doi.org/10.1016/j.compedu.2014.03.002
0360-1315/

recorded lectures and, if so, how these patterns correlate with academic performance. The end goal of this line of research is to determine both good and bad uses of lecture capture technology, in order to encourage students to use it appropriately. The key data we examine in this work is usage data. Created as a side effect of student interaction with learning systems, usage data describes the actions and choices students have made. We look at this data through the lens of machine learning, a computational and statistical process, which can be effective at summarizing and revealing patterns in large or complex datasets. Specifically, through the application of unsupervised machine learning, we demonstrate that there are at least five distinct patterns of lecture capture use, and that one of these (the consistent weekly viewing of lecture video) is positively correlated with higher academic performance. We go further and replicate these results on two related cohorts, showing that the effect persists despite no significant difference between the incoming grades of learners in different clusters. We go on to demonstrate how this technique can be used to form a descriptive abstract usage model. This model provides insights into how learners use lecture capture, and can be used by educational researchers and instructional designers for scaffolding of the learning process. We finally address the implications such a model may have for predicting student success for early-alert student support systems.

1.1. Related work

There has been a variety of conflicting evidence with respect to whether the use of lecture capture technologies affects academic performance. In a broad survey of the area, Heilesen (2010) cites a number studies that show no effect on performance (Abt & Barry, 2007; Baker, Harrison, Thornton, & Yates, 2008; Hodges, Stackpole-Hodges, & Cox, 2008) as well as a number that show a positive effect (up to 9.5%) depending on the kind of evaluation being given (Carle, Jaffee, & Miller, 2009; Kurtz, Fenwick, & Ellsworth, 2007; McCombs & Liu, 2007; McKinney, Dyck, & Luber, 2009; Smith & Fidge, 2008). Heilesen ends his survey considering the lack of consensus of the efficacy of lecture capture (which he refers to as podcasting) with an insightful point:

> "As matters stand, the answer to whether or not engaging in podcasting is worthwhile for purely academic reasons is not entirely clearcut. Evidence that students score better at exams after having listened to podcasts is inconclusive, and most likely the positive effects claimed should be attributed to the uses made of the technology rather than the technology per se." (pg. 1066)

It is the way lecture capture technology is used that Heilesen attributes to the change in student achievement, and this usage is the data we are interested in better understanding through the modelling of learners. Leadbeater, Shuttleworth, Couperthwaite, and Nightingale (2013) have similar goals, and clustered learners into one of two groups depending on the frequency with which they watched lectures. These groups, denoted as *high usage* and *low usage*, did not differ with respect to academic performance; however other demographic information (dyslexia and English as a second language status) did differentiate these two groups. Our study on learner usage of lecture capture technology is similar in goal but our methodology allows us to ask slightly more nuanced questions, and thus reach different conclusions. We revisit the differences between our work and Leadbeater et al. (2013) in more detail in §4.

Finally, it is useful to note that there are many studies on the use of lecture capture that approach the issue using qualitative methods. Studies that use questionnaires to elicit self-reports of increased learner satisfaction, increased flexibility of learning, and positive performance are numerous within the literature (see Pursel & Fang, 2012; for an overview of some of the approaches). We do not explicitly consider these issues here, but note that there may be interesting correlations with the way lecture capture tools are used (e.g., the clusters we identify) and the responses that learners give to various survey instruments.

2. Methodology

The lecture capture environment that we used was a predecessor of the freely available Opencast Matterhorn system (Brooks, McKenzie, et al., 2011) and, like Matterhorn, it provides the ability to track interactions learners have with the system. As learners interact with the user interface through pressing buttons, selecting list items, hovering the mouse over images, and seeking on the timeline, logs of this data are sent back to a central server. Most importantly for this work is that *heartbeats* are also logged – automatic events every 30 s that indicate the video the learner has loaded, the position the learner is at in that video, and whether that video is playing or not. From this, a rough estimate of the amount of video that has been watched can be calculated.[1]

Our experiments involved three cohorts of second year (sophomore) students in science courses at a research-intensive university. Each cohort was taught through traditional face-to-face delivery (i.e., not distance education) on a university campus. Students could attend three fifty minute lectures per week for thirteen weeks, all of which were recorded and made available online 24–48 h of the live lecture. Students also had access to various other study aids such as textbooks, online resources, laboratories, and tutorials, and each class included a midterm and a final examination, and students were notified via e-mail when lectures were posted online. The cohorts used were:

1. *Chemistry 2010 Spring*: This course was an introduction to organic chemistry and was taught in three parallel sections by different faculty members, but used a single curriculum and a shared set of assessments (common examinations and assignments). A total of 636 learners were given access to the lecture capture environment over the spring 2010 term.
2. *Chemistry 2011 Spring*: This course was an introduction to organic chemistry and was taught similarly to the 2010 offering, with a slightly modified curriculum. Multiple sections of the course were taught and used a shared set of assignments and examinations. A total of 546 learners were given access to the lecture capture environment over the spring 2011 term.

[1] There are multiple methods for creating this measure of "minutes of video watched", all of which are susceptible to different forms of error. For this work, heartbeats were considered to be indicative of 30 s of video watched if the heartbeat indicated the player was in a playing state, and all other events were discarded. Thus the total amount of time a learner spent watching video in some time period was the number of heartbeats in that time period times 30 s.