ARTICLE IN PRESS

Computers in Human Behavior xxx (2016) 1–12

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Enriching programming content semantics: An evaluation of visual analytics approach

I-Han Hsiao [a, *], Yi-Ling Lin [b]

[a] School of Computing, Informatics & Decision Systems Engineering, Arizona State University, 699 S. Mill Ave., Tempe, AZ 85281, USA
[b] Department of Information Management, National Sun Yat-Sen University, No. 70, Lienhai Rd., Kaohsiung 80424, Taiwan

ARTICLE INFO

Article history:
Received 1 May 2016
Received in revised form
9 October 2016
Accepted 12 October 2016
Available online xxx

Keywords:
Semantic indexing
Automatic indexing
Visual analytics
Interactive visualization
Learning analytics
Programming learning

ABSTRACT

In this work, we present an intelligent classroom orchestration technology to capture semantic learning analytics from paper-based programming exams. We design and study an innovative visual analytics system, EduAnalysis, to support programming content semantics extraction and analysis. EduAnalysis indexes each programming exam question to a set of concepts based on the ontology. It utilizes automatic indexing algorithm and interactive visualization interfaces to establish the concepts and questions associations. We collect the indexing ground truths of the targeted set from teachers and experts from the crowd. We found that the system significantly extracted more and diverse concepts from exams and achieved high coherence within exam. We also discovered that indexing effectiveness was especially prevalent for complex content. Overall, the semantic enriching approach for programming problems reveals systematic learning analytics from the paper exams.

1. Introduction

Paper-based exams are one of the main assessment methods in today's majority of classrooms. Such delivery method is especially beneficial for the sake of easiness in exam-proctoring and preventing academic dishonesty. However, they are in fact very time-consuming to grade, hard to maintain consistency among graders, and normally contain only very limited feedback to a student. Furthermore, it is impractical for an instructor to track detailed performance of a student (e.g., how s/he received partial credits in different exam questions), instead, teachers discuss on the returned exam in the class (hopefully thorough and detailed enough to cover all the students' misconceptions). Although teachers may still point out the common mistakes and try to pinpoint the key concepts related to the such mistakes during instruction, many desired detailed learning analytics are unavailable, such as how did s/he receive partial credits, was it a single concept or multiple concepts mistake, a careless mistake or a long-term misconception etc. As a result, students often focus solely on the scores they earned on the returned exams, but miss several learning opportunities (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010) such as *identification of strength and weakness, characterization of the nature of their errors or any recurring patterns if any, assessment of appropriateness of their study strategies and preparation*, etc. Hence, making it impossible to apply learning analytics for delivering personalized feedback to the student. Therefore, unlike most of the orchestration technologies, which mainly address digital form of educational data (Dillenbourg, 2013), in this work, we propose an educational technology solution that permits traditional paper delivery method to be able to utilize advanced learning analytics by analyzing the textual content and supplying semantic information.

In order to provide additional learning analytics for traditional paper-based exams in facilitating today's majority classes, we focus on a targeted domain, programming language learning, and a targeted paper delivery content, paper-based programming exams. We create an innovative visual analytics system, EduAnalysis, to analyze the content and to index it to a set of concepts based on the ontology. EduAnalysis implements an automatic indexing algorithm and interactive visualization interfaces to establish the concepts and exam questions semantic associations. Our core research question is whether the proposed approach can effectively capture advanced programming learning analytics to enhance paper-based programming assessments. Specifically, we hypothesize that the indexing method can provide richer information to the content and the indexing approach can facilitate content analysis in traditional

* Corresponding author.
E-mail addresses: Sharon.Hsiao@asu.edu (I.-H. Hsiao), yllin@mis.nsysu.edu.tw (Y.-L. Lin).

http://dx.doi.org/10.1016/j.chb.2016.10.012
0747-5632/

paper-based programming assessments. To verify these hypotheses, we collect the indexing ground truths from teachers and experts from the crowd and compare the results with the proposed algorithmic method.

The main contributions of this work are outlined as following:

- Provide immediate technology support for today's majority programming classes, particularly (large) blended instruction classrooms that are instrumenting paper-based formal assessments;
- Introduce novel intelligent semantic parser to automatically associate concepts and programming problems;
- Present visual authoring, delivery and presentation interfaces via semantic analytics visualizations in the targeted context;
- Conduct controlled crowdsourcing experiment to harness educational ground truth;
- Empirically evaluate the proposed intelligent semantic indexing method to address a real world problem.

The rest of the paper is structured with literature review on topics *classroom orchestration & learning analytics, semantic enrichment to enhance learning and visual learning analytics & student modeling*. In section 3, we present the visual analytics system, EduAnalysis. In section 4, we lay out our study methodology with our underlying assumptions and evaluation measures. Finally we present the evaluation results and discussed study implications, limitations and future work.

2. Literature review

2.1. Orchestration & learning analytics

In the field of Computer Supported Collaborative Learning (CSCL), researchers describe course-delivery as a field in transition for classroom orchestration, which defines how a teacher manages multilayered activities in real time and in a multi-constraints context (Dillenbourg, 2013). Orchestration emphasizes attention to the challenges of classroom use and adoption of research-based technologies (Roschelle, Dimitriadis, & Hoppe, 2013). It discusses how and what research-based technologies have been adopted and should be done in classrooms (Dillenbourg, 2013). We have begun to see more tabletops, smart classrooms or interactive tools such as Classroom Response Systems (AKA: Clickers) etc. provide dynamic feedback and integrative students knowledge updates (Martinez-Maldonado, 2014; Martinez-Maldonado, Dimitriadis, Martinez-Monés, Kay, & Yacef, 2013; Roschelle, Penuel, & Abrahamson, 2004; Slotta, Tissenbaum, & Lui, 2013). One of the biggest criticisms of introducing orchestration technology in class is that it might potentially add more complexity and time demands of technology and introduces new and unnecessary complications (Sharples, 2013). Thus, it motivates us to research a less intrusive technological solution that taps into blended classes allowing to manage physical and digital content and to jointly discuss learning analytics.

Vatrapu, Teplovs, Fujita, and Bull (2011) describe a preliminary framework, Triadic Model of Teaching Analytics (TMTA), discussing the importance involving three stakeholders in learning analytics: teaching expert, visual analytics expert and design-based research expert. The focus of learning analytics has been on the integration of computational and methodological support for teachers to properly design, deploy and assess learning activities. In addition, the focus is also to immerse students in rich, personalized and varied learning activities in information ecologies and data-rich classrooms (Vatrapu et al., 2011). One of the pioneer systems that align with TMTA framework is eLab (exploratory Learning Analytics Toolkit). It was designed to enable teachers to explore and correlate content usage, to help teachers reflect on their teaching according to their own interests (Dyckhoff, Zielke, Bültmann, Chatti, & Schroeder, 2012). ASSISTments (Heffernan & Heffernan, 2014) an integrative tutoring system includes assistance and assessment components for students and teachers. The system is built on a mantra - put the teacher in charge, not the computer, which creates flexibility to allow teachers to use the tool in organizing the classroom routines. However, such intelligent tutors or newly invested orchestration technologies are typically highly customized to the content or require a large collection of content for teachers to start using the tools. In this work, we propose and evaluate an automatic method to enrich content semantics in bridging physical and digital via visual learning analytics.

2.2. Semantic enrichment to enhance learning support

Semantic approaches have been widely discussed in current computer-based education. There is a line of ontology related studies being pursued by a number of researchers in different aspects of learning, such as learning content authoring and management, contextual annotation and support, personalized search and content composition, learning resource and metadata repositories (Tiropanis, Davis, Millard, & Weal, 2009), etc.

AIMS (Aroyo & Dicheva, 2001) and TM4L (Topic Maps for Learning) (Dicheva, Dichev, Sun, & Nao, 2004) are two good examples for contextual annotation and support. They both enable learners to identify related information resources for different tasks such as course assignments. They provide the complementary support for learning tasks through subject domain conceptualization methods. The project of iHelp Presentation (Bateman, Brooks, Mccalla, & Brusilovsky, 2007) helps learners to highlight important parts of the recorded lectures' slides and support them tagging, annotation, and collaboration features around the recordings. Research conducted by the LORNET network (Paquette, 2007) offers a semantic framework to manage the learners' competency portfolios and models in e-learning and knowledge management environments. The work presented in (Jovanovic et al., 2007) demonstrates the semantic technologies enable a generic implementation of feedback for content authors and teachers to aware about the quality of the learning process based on students' activities in online learning environments. ArnetMiner team (2008) developed the system at extracting and mining academic social networks to expertise search and people association search. Alomari, Hussain, Turki, and Masud (2015) developed a semantic model for collaborative learning by graphically representing course content with semantic meaning.

Assessment in learning can be characterized as an index of learning guidance or a summary of learners' performance (Basu, Jacobs, & Vanderwende, 2013). In the context of automatic evaluation, there is a stream of research focuses on the correctness of syntactical references by using pattern-matching techniques to verify solutions (i.e. WEB-CAT auto-grading (Edwards & Perez-Quinones, 2008). There are other streams of work that emphasizes on semantic relations, such as TagAssessment (Kardan, Sani, & Modaberi, 2016), which has been proposed for assessing leaners by computing the semantic relationship between educational contents and learner's tags on multiple choice questions (MCQ). Mohler and Mihalcea (2009) applied various measures of lexical similarity based on WordNeT and Latent Semantic Analysis(LSA) to automatic short answer grading. Basu et al. (2013) introduce a semi-automatic grading approach to allow teachers to grade easily with fewer actions, provide feedback to groups of similar answers, and discover modalities of students' misunderstanding. Including our current work, we design a visual analytics system that utilizes