Full length article

Using process mining to analyze students' quiz-taking behavior patterns in a learning management system

Libor Juhaňák [a], Jiří Zounek [a], Lucie Rohlíková [b, *]

[a] Department of Educational Sciences, Faculty of Arts, Masaryk University, Brno, Czech Republic
[b] Department of Computer Science and Educational Technology, University of West Bohemia, Pilsen, Czech Republic

ARTICLE INFO

ABSTRACT

The aim of this paper is to explore students' behavior and interaction patterns in different types of online quiz-based activities within learning management systems (LMS). Analyzing students' behavior in online learning activities and detecting specific patterns of interaction in LMS is a topic of great interest for the educational data mining (EDM) and learning analytics (LA) research communities. Previous studies have focused primarily on frequency analysis without addressing the temporal aspects of students' learning behavior. Therefore, we apply a process-oriented approach, investigating perspectives on using process mining methods in the context of online learning and assessment. To explore a broad range of possible student behavior patterns, we analyze students' interactions in several online quizzes from different courses and with different settings. Using process mining methods, we identify specific types of inter-action sequences that shed new light on students' quiz-taking strategies in LMS. We believe that these findings bring important implications for researchers studying student behavior in online environments as well as practitioners using online quizzes for learning and assessment.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Information and communication technologies (ICT) have become one of the pillars of contemporary society and a natural part of the everyday lives of most people. The realm of education is no exception. Nowadays, ICT use in education varies across countries and types of educational institutions, as well as among individual institutions. Using technology is nevertheless a standard in most institutions of higher education. Examples include learning management systems (LMS) and virtual learning environments (VLE). Ferguson (2012), for instance, mentions that over 85% of higher education institutions in Great Britain were using VLE of some kind as early as 2003. Macfadyen and Dawson (2012) point out similar numbers for USA, where more than 90% of institutions of higher education have made significant investments into campus-wide implementation of web-based LMS since the late 1990s. In the Czech Republic, all institutions of higher education except art-oriented ones have been using VLE (Poulová).

Until recently, these learning systems (LMS, VLE) have represented a kind of "black box." Although in many cases, these systems collect and store information on their use in the form of logs generated automatically on a server or in the system's database, such data are rarely employed in research or systematically used by the institution. A significant change has occurred only recently in connection with the development of research areas such as educational data mining (EDM) and learning analytics (LA). EDM started to garner more attention around 2008–2009 (Romero & Ventura, 2013), although its roots can be traced back to as early as 1995 (Romero & Ventura, 2007). LA, with roots in 2010–2011 (Ferguson, 2012; Juhaňák & Zounek, 2016; Siemens, 2013), is somewhat younger. Although the two areas are distinct, they share an interest in fully exploiting the potential of the unique types of data generated and stored by the learning environment itself. To achieve this end, they use a variety of analytical and data mining methods and techniques, enabling them to acquire important information and knowledge about how students behave and learn in these environments.

The considerable potential of these methods consists of, among other things, the ability to open the black box and "see" what learning in online learning environments actually looks like. This may bring benefits not only in terms of learning research but also in terms of didactics and actual teaching practice. Although using LMS

* Corresponding author. Sportovní 21, Plzeň, 30100, Czech Republic.
E-mail address: rohlikova@gmail.com (L. Rohlíková).

in teaching is relatively common in many institutions of higher education, teachers using LMS in their classes often do not have many opportunities to follow what exactly is occurring in their online courses, how students behave in them, how they approach studying online materials, or how they proceed when engaging in learning activities. LMS generally do not automatically contain advanced data mining tools, while external data mining tools are too complex for teachers and their features go well beyond the scope of what an educator might require (Romero & Ventura, 2010; Romero, Ventura, & García, 2008; Romero, Cerezo, Bogarín, & Sánchez-Santillán, 2016). Thus, it is necessary to develop new analytical and data mining approaches and tools for teachers to be able to observe the behaviors and interactions of students engaging in online learning activities in simple ways and in detail.

The impossibility to follow in detail how students behave while learning in LMS may have rather crucial consequences, especially for evaluating students' online activities (online assessment). And although in EDM, student behavior modeling is receiving relatively a lot of attention (Papamitsiou & Economides, 2014; Peña-Ayala, 2014), a great deal of research focuses on data from intelligent tutoring systems (ITS) rather than LMS (e.g. Baker & Gowda, 2010; Baker, Goldstein, & Heffernan, 2011; Muldner, Burleson, Van de Sande, & VanLehn, 2011). Studies focusing on learning systems of the LMS type often pay attention especially to academic success prediction or at-risk student identification (e.g. Baker, Lindrum, Lindrum, & Perkowski, 2015; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014; Smith, Lange, & Huston, 2012). Moreover, research often treats a more general level, e.g. by following student behavior throughout a course (e.g. Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Hershkovitz & Nachmias, 2011). Interactions of students in specific contexts of specific learning activities receive less research attention (Phillips et al., 2011), even though ongoing research seems to indicate that the context of online learning is crucial. Not considering context (the technologies used and the ways they are used — i.e. instructional conditions) enough likely prevents the development of well-functioning and yet sufficiently general predictive models (Gašević, Dawson, Rogers, & Gasevic, 2016).

We can therefore see that although using LMS has been widespread in the context of higher education and increasing numbers of researchers active in areas such as learning analytics and educational data mining focus on analyzing data from these systems, relatively little attention has been paid to developing analytical techniques to explore unique behaviors of students in online learning environments. The main motivation behind this study therefore is the wish to contribute to developing novel approaches to analyze interactions by individual students involved in specific online activities, which seems to be one of the important tasks of current learning analytics research.

Accordingly, this study focuses on student behavior in LMS (in this case, the widely used open-source system Moodle), specifically on student interactions while engaging in specific quiz-based learning activities. Our analysis of student interactions uses process mining methods, allowing for mapping and modeling the process of completing quizzes by students. We seek to answer the following research questions: Can process mining methods be used to detect specific student behavior patterns during various types of quizzes in the learning management system Moodle? If so, what types of student behavior patterns can be identified using these methods? And how frequently do the individual types of student behavior occur? We are trying to open the "black box" and understand how one of the aspects of learning in online environments works. Our study targets research methodology process mining approach; the results can be used not only in research but

also in real-life teaching supported by LMS, or in test development.

Another part of our study therefore maps the existing lines of research on student quiz-taking behavior detection and modeling. We also describe our research methodology and the data we have analyzed. The Results section presents the main findings of our research, followed by a discussion of the main findings and their application in practice and in further research.

2. Previous research

In the Introduction, we have noted that student behavior modeling in various learning systems (LMS, ITS, etc.) has been connected especially to EDM and LA, where it currently receives much attention. If, however, one focuses on student behavior in quiz-based learning activities or online tests, one discovers that the topic is studied by many researchers even in contexts other than EDM or LA. Quiz-taking behavior may have crucial consequences for student evaluation; therefore, the identification of various types of quiz-taking behaviors (especially unusual or non-standard ones) is the subject of much attention, mainly in the context of student assessment. In connection with learning in LMS or VLE, online assessment, computer-based assessment, or electronic assessment is discussed (Buchanan, 2002; Gaytan & McEwen, 2007; Terzis & Economides, 2011; Whitelock, 2006).

One of the most frequently employed approaches to formative and summative online assessment in LMS includes various forms of quizzes and tests, depending on the specific technical features available in the system in question. In the learning management system Moodle, a tool referred to as Quiz (or Quiz activity or Quiz module) is used. While tests are usually discussed in the context of summative evaluation, quizzes tend to be used for the purposes of formative evaluation (Kibble, 2007; Peat & Franklin, 2002). In this study, the term "quiz-based activity" is used to denote activities created within Moodle through the Quiz tool, which may be used for formative or summative evaluation.

Another important classification is usually discussed in connection with tests, which can, however, be associated with quizzes as well. It consists of distinguishing between low-stakes and high-stakes tests (or quizzes). This distinction is based on the consequences the performance in the test has for the given student. Performance in low-stakes tests usually has no or minimal consequences for the test taker, while performance in high-stakes tests is very important or crucial for the student. One example of a high-stakes test can be a qualifying exam, which must be passed for students to be able to continue their studies. An example of a low-stakes test is a common semester test, which does not contribute to the final course grade. Low-stakes tests are usually associated with formative evaluation, while high-stakes tests tend to be used for summative evaluation (Davis, 2013; Zhang & Henderson, 2015).

The above distinctions are important firstly because various types of tests and quizzes may attract different forms of student behavior and secondly, because various types of quiz-taking behavior detection and modeling may occur in the context of various types of quizzes and tests. One of the approaches used to study summative evaluation and high-stakes tests is the approach based on so-called person-fit statistics.

2.1. Person-fit statistics

Analyses based on person-fit statistics have a relatively long tradition and are used not only for educational assessment but also (perhaps even primarily) for psychological assessment, personality assessment, or attitudinal assessment (Meijer & Sijtsma, 2001; Rupp, 2013). Person-fit methods are usually developed within the so-called item response theory (IRT), and they are generally used to