Internet and Higher Education 30 (2016) 30–43

Contents lists available at ScienceDirect

Internet and Higher Education

Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments

Dongho Kim [a], Yeonjeong Park [b,c], Meehyun Yoon [a], Il-Hyun Jo [b,c,*]

[a] Department of Career and Information Studies, 850 College Station Rd, University of Georgia, GA, United States
[b] Center for Teaching and Learning, 417 Eodeung Rd, Honam University, Gwangju, Republic of Korea
[c] Department of Educational Technology, 52 Ewha Rd, Ewha Womans University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 29 October 2015
Received in revised form 23 March 2016
Accepted 27 March 2016
Available online 2 April 2016

Keywords:
Learning analytics
Asynchronous online discussion
Proxy variable
Educational data mining
Learner online behavior

ABSTRACT

Although asynchronous online discussion (AOD) is increasingly used as a main activity for blended learning, many students find it difficult to engage in discussions and report low achievement. Early prediction and timely intervention can help potential low achievers get back on track as early as possible. This study presented a data mining process to construct proxy variables that reflect theoretical and empirical evidence and measured the accuracy of a prediction model that incorporated all of the variables for validation. For the empirical study, data were obtained from 105 university students who were enrolled in two blended learning courses that used AOD as their main activity. The results indicated the high accuracy of the prediction model as well as the possibility of early detection and timely interventions. In addition, we examined participants' learning behaviors in the two courses using the proxy variables and provided suggestions for practice. The implications of this study for education data mining and learning analytics are discussed.

1. Introduction

Blended learning is increasingly adopted in higher education due to its flexibility (Irvine, Code, & Richards, 2013). Blended learning has benefitted from using both synchronous and asynchronous delivery modes for its online portion to enhance students' access and engagement. Synchronous delivery modes use a two-way, real-time technologies such as videoconferencing system to support learning that takes place between two or more people at the same time (Butz & Stupnisky, 2016; Hrastinski, 2008). Synchronous learning has been lauded for its engaging nature that helps to reduce students' feeling of isolation in online learning environments. In contrast, asynchronous delivery modes are commonly facilitated by technologies that allow learners to engage in learning at any time. Asynchronous learning makes learners more flexible in terms of reflecting on learning content and refining their contributions (Hrastinski, 2008; Vignare, 2007).

Asynchronous online discussion (AOD) is a popular form of asynchronous learning used to support critical discussion and interaction among individual learners (Hew, Cheung, & Ng, 2010; Liu, Magjuka, Bonk, & Lee, 2007; Loncar, Barrett, & Liu, 2014). AOD offers learners the freedom to exhibit their own learning style without constraints of time and space (Berge & Collins, 1995). Empirical evidence of the effect of AOD has been reported in research on blended learning (Vignare, 2007). Researchers and practitioners acknowledge the potential of AOD to promote social interaction and reflection while allowing for a better understanding of contents (Andresen, 2009; Thomas, 2002). AOD is also recognized as a means to promote a sense of community in online courses in that it facilitates "information sharing, idea exchanges, and mentoring" (Liu et al., 2007, p. 12). Learners, however, often encounter challenges in deeply engaging in discussion topics (Balaji & Chakrabarti, 2010; Mason, 2011) and sustaining course-related endeavors (Hew et al., 2010; Wise, Speer, Marbouti, & Hsiao, 2013). Many learners superficially participate in AOD without contributing to the discussions (Hernández-García, González-González, Jiménez-Zarco, & Chaparro-Peláez, 2015; Mason, 2011; Nagel, Blignaut, & Cronjé, 2009). Those challenges can result in shallow interactions and fragmented communications, which impede development of a sense of online community (Liu et al., 2007).

Data-mining techniques have been used to address these issues through examining students' online learning behaviors (Lara, Lizcano, Martínez, Pazos, & Riera, 2014). Online learning behaviors are often analyzed using student log data in the discpline of educational data mining and learning analytics. Log data provide information about how students participate in various online learning activities. We can obtain, for example, information about how much time students spent on a particular online activity. Despite some concerns about translating students' log data into their actual learning behaviors, a large body of literature has provided empirical evidence of strong correlation between them (Hung & Zhang, 2008; Jo, Kim, & Yoon, 2015; Lara et al., 2014; Romero, López, Luna, & Ventura, 2013). In addition to learners'

* Corresponding author.
E-mail addresses: dongho@uga.edu (D. Kim), ypark@honam.ac.kr (Y. Park), meehyun@uga.edu (M. Yoon), ijo@ewha.ac.kr (I.-H. Jo).

http://dx.doi.org/10.1016/j.iheduc.2016.03.002
1096-7516/

visible behaviors, their psychological characteristics, such as an interest in a particular topic, may be reflected in their log data (Woolf et al., 2009).

Given the enormous amount of data that is generated in online discussions with regard to learner participation (Loncar et al., 2014; Romero et al., 2013), automated examination is of benefit to those who teach a blended course that adopts AOD (Dringus & Ellis, 2005). If instructors could track the status of students' learning, it would be possible for them to implement timely interventions (Macfadyen & Dawson, 2010; Zacharis, 2015), which would lead to students' perseverance and the successful completion of online courses (Lykourentzou, Giannoukos, Nikolopoulos, Mpardis, & Loumos, 2009).

If students are identified as potential low achievers during a course, instructors can then encourage them to participate through the use of facilitative strategies (e.g. Hew et al., 2010; Hou, 2011), incentives (Gilbert & Dabbagh, 2005), and relevant learning materials (Yeh & Van Buskirk, 2005)

The current study aimed to propose proxy variables for different AOD settings, and the variables were constructed to reflect theoretical and empirical evidence from prior research. This solid basis can contribute to the high applicability and generalizability of the generated variables to address variations in AOD context settings (e.g., discussion types and interaction patterns), instruction strategies (Dykman & Davis, 2008; Hou, 2011; Schworm & Gruber, 2012), and student attributes (Duque, Gómez-Pérez, Nieto-Reyes, & Bravo, 2015; Hernández-García et al., 2015; Topcu & Ubuz, 2008)

The objectives of this study were as follows: (a) to present a data mining process for constructing proxy variables that represent the specific behavioral and psychological characteristics of high achievers in asynchronous online discussion; (b) to empirically validate the variables in two blended courses that adopt AOD as their main activity; and (c) to provide suggestions for practice.

In this study, the learning analytics approach was used to address the following questions:

1. How accurately can the proxy variables predict low and high achievers? and
2. What learning behaviors are observed in AOD through the lens of the proxy variables?

Prior to addressing the research questions, the following section describes the process of constructing the proxy variables as indicators of success in AOD environments.

2. Constructing the proxy variables

Proxy variables are those that are alternatively used when the direct measurement of conceptual variables is difficult because of access or feasibility (Jo et al., 2015; Wickens, 1972). The concept of proxy variables is often used in the field of social science to create prediction models (Durden & Ellis, 2003). Here, we developed a method for extracting such proxy variables to represent key factors that have been identified in the literature on AOD. Three steps were taken to transform the key factors into proxy variables (see Fig. 1). First, four AOD success factors were identified through an extensive literature review: *Active participation in the course*, *Engagement with discussion topics*, *Consistent effort and awareness*, *and Interaction*. Second, specific behavioral and psychological characteristics of high achievers (e.g., regular study and many postings) were identified for each factor based on prior research that explored indicators of high academic performance in AOD. Lastly, 10 proxy variables, each of which represents one of the four factors, were calculated. To allow for automated prediction, we included only quantitative variables that did not require manual assessment from instructors. For example, we purposefully did not analyze the structure of discussion threads because they were likely confined to particular topics. The remainder of this section provides details of how the proxy variables were constructed within the four key factors.

2.1. Active participation

Active participation is a crucial measure of student engagement, and it typically leads to high academic performance in online discussions

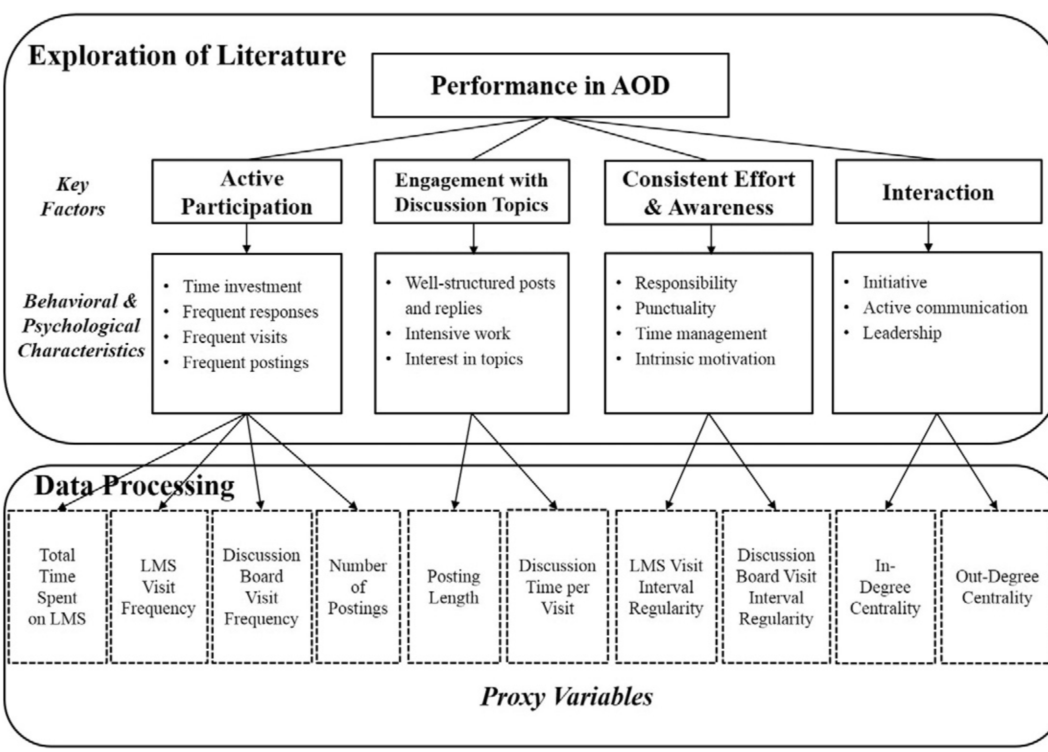

Fig. 1. Process of extracting proxy variables.