CrossMark

Modelling adult learners' online engagement behaviour: proxy measures and its application

Adam Wong[1] · **Sylvia Chong**[1]

Abstract With teaching and learning taking place in an increasingly networked environment over e-learning platforms, voluminous data are logged in databases. These can be mined and processed to support teaching and learning practices. To determine the impact of these practices, meaningful measures of learners' online engagement are needed. In particular, this research focuses on discovering useful and meaningful data features of online engagement in learning activities from traces of adult learners' online behavioural data. Whilst vast data from learning activities may be available over a technology-enhanced learning environment, it becomes significantly important to identify a methodical approach to transform data features in a modus that is useful for analysis. Hence, the purpose of this study is twofold: (1) to adapt and reconstruct the RFM model (a marketing segmentation technique on customers' recency, frequency and monetary purchasing behaviour), as a common framework to codify and quantify learners' online study behaviour, in the learning analytics context; and (2) to explore the online engagement patterns of adult learners using data-mining techniques. We show examples of its applications using real-world data from an online course—by modelling adult learners' online engagement patterns and discovering learners' segments based on immediacy, frequency and duration.

Keywords Data features · Online engagement · RFM model · Learning analytics

---

✉ Sylvia Chong
  sylviachong@suss.edu.sg

[1] Singapore University of Social Sciences, Singapore, Singapore

Introduction

Society, as a whole, has become more entwined with the use of technology and computers in everyday life and work. This has increasingly influenced online learning environments to assist students in acquiring knowledge and skills. The use of technology-based training and the growing adoption of learning management systems have more than doubled in the past decade (Brown et al. 2012). Higher education has seen paradigm shifts from the traditional classrooms and tangible learning resources to asynchronous e-learning environments. This paradigm shift has fundamentally changed how learners are engaged. Technology-enhanced learning has many challenges, one of which is the learner online engagement. Learner engagement is primary factor to the effective teaching and learning in an e-learning environment. Bruner (2013) postulated that "engagement is the ultimate test" (p. 34) to successful learning in an online environment.

As economies of the world continue to evolve, there is a continuum of educational and training needs for adult learners in lifelong learning. According to OECD statistics, 57% of the population aged 25–64 years old in Singapore, participated in formal and/or non-formal education in 2015 (Kuczera 2017). From an adult education survey, this statistic saw an upward trend in EU-28, from 2011 to 2015. Increasingly, course developers and instructional designers have to grapple with the learning needs of this growing group of adult learners. As formal learning take place in an increasingly networked environment over e-learning platforms and learning management systems, in order to cater to the changing demographics of learners. The shift towards digital environments makes it possible to retrieve, store, manage and analyse increasingly large amounts of data over the cloud in the digital earth and relevant science domains (Yang et al. 2017). In particular, in the online learning context, the abundance of clickstream data provides an opportunity for educational practitioners and researchers to harvest data related to learner engagement. Aided by data-mining methods, the analysis and sense-making of the interaction data between the learner, learning environment and learning activities has become less cumbersome than before, and this can support a better understanding of the online engagement behaviour (Gašević et al. 2015). The importance of understanding these interactions and what might increase effectiveness of such interactions in online education is paramount for meaningful adult learning. In particular, this research focuses on discovering meaningful patterns of engagement and disengagement in learning activities, observed from traces of the adult learners' online engagement behavioural data.

Measuring learner engagement and its influence on learning is challenging. While the definition of learner engagement should stay consistent with more traditional learning environments, the measurement of learner engagement should be unique to the data availability of the online learning environment (Anderson 2017). Identifying proxies of online learner engagement can provide a degree of measurability that can be used to inform and improve upon existing teaching and learning practices (Beer et al. 2010, p. 75). Hence, the aims of this study is twofold: (1) to explore the potential of reconstructing a variation of the RFM (a marketing

segmentation technique based on customers' recency, frequency and monetary purchasing behaviour) analysis, as a framework to codify and quantify the adult learners' online engagement; and (2) to explore the online engagement patterns of adult learners using data-mining techniques (i.e. unsupervised learning method: non-fuzzy clustering).

Literature review

The literature review comprises four parts. The first part of the section describes the online learner engagement. The second part presents reviews of some studies on the RFM model as a measurement variable. The third part focuses on the log data with sub-parts on the engagement metrics. The last part of the section presents some studies on applications of learning management system data analysis.

Online learner engagement

There are several different definitions of learner engagement. A literature review of the learner engagement by Trowler in 2010 defined learner engagement as "the investment of time, effort and other relevant resources by both students and their institutions intended to optimise the student experience and enhance the learning outcomes and development of students, and the performance and reputation of the institution" (Trowler 2010, p. 3). The definition is learner centred while also acknowledging the role of the institution in learning engagement. It is possible to argue that activity within a learning management system (LMS) might serve as a measure of learner engagement because LMS log data contain a record of learner behaviours as well as institutional practices. Learner behaviours such as logging in, accessing course content, completing quizzes, submitting assignments and taking part in discussions related to coursework are captured.

According to Fredricks and MsColskey (2012), "researchers, educators, and policymakers are increasingly focused on student engagement as the key to address problems of low achievement, high levels of student boredom, alienation, and high dropout rates" (p. 763). The definition and measurement of learner engagement become more complex in the case of online learning environments. In 2009, Allen and Seaman (2010) conducted a survey on online teaching and learning and concluded that learner engagement online was a key challenge in higher education. A National Survey of Student Engagement (2012) showed that the online learners while faced with more challenges are less engaged in their online learning activities. This has led to their dissatisfaction with the overall learning experience.

Learner engagement has been defined in research in terms of effort (Jung and Lee 2018), time-on-task (Lee 2018), and motivation (Jin 2017). Research has also linked learner engagement with involvement time in learning activities and practices that lead to increased academic achievement (Axelson and Flick 2011; Leach and Zepke 2011; You 2016).

Learner engagement is made up of the interaction that a learner has with their instructor, course content, and other learners (Siemens 2013). Online student

engagement has been measured by single observed variables such as independent time in course (You 2016), the number of on-task as well as off-task internet activities (Lee 2018), and self-esteem (Harbaugh and Cavanagh 2012). Students engage instructors and other learners by posting on discussion boards, attending faculty live-chat sessions, instantly messaging their instructor during faculty office hours, viewing faculty-recorded chat sessions, viewing faculty comments and announcement postings, and e-mailing their instructor and classmates (Anderson and Dron 2011). In addition, students receive online instruction using streaming media. Most campuses offer students a mix of both asynchronous and synchronous communication technology in the delivery of online instructions (Cleveland-Innes and Ally 2013). Online interaction and collaboration occur in virtual classrooms that allow learners to decide which e-learning technologies they will employ to engage in online instructional activities (Anderson and Dron 2011).

Given the above, there does not appear to be a distinct definition for online engagement; however, the following definition characterizes an amalgamation of existing literature. Online engagement is seen to comprise active and collaborative learning, participation in challenging academic activities, formative communication with academic staff, involvement in enriching educational experiences, and feeling legitimated and supported by university teaching communities (Coates 2007, p. 122). At its basic tenet, online engagements are manifestations of a range of interactions such as interactions of learner-to-content, learner-to-learner, and learner-to-teachers. Closely related to this, this paper proposes a study that looks at online engagement as a construct that is composed of three key indicators (i.e. Immediacy, frequency and duration); the inclusion of these indicators is expected to be adequate in providing a degree of measurability for online engagement.

RFM and its application in learning analytics

In this study, the RFM (or recency, frequency, monetary) model is used as a reference to derive the proposed online engagement indicators. The RFM model is most widely used in marketing research and practice to quantify customer loyalty and purchase behaviour (Chang 2010). Based on this, the three variables are harvested individually from marketing databases, combined linearly and then scored as a mean to quantifying a customer's overall purchasing intent (Hu et al. 2012). Following and related to this, the concepts of RFM have also been proposed for different domains such as for the electronics industry (Chiu et al. 2009) and for the automobile industry (Chan 2008).

The RFM Model's fluid and yet simplistic approach is perhaps the fundamental reason to its effectiveness and reliability when it is replicated in different domain areas. Other studies have extended its use to the learning context, Chang (2010) has adopted and re-conceptualized the RFM model to an eLearning-recency, frequency, monetary (EL-RFM) model to score learners' participation in an e-learning environment. The EL-RFM model was further customized to quantify learners' learning outcome online. The model, a reclassification of the RFM model, has been redefined to EL-R (recency of eLearning), frequency of eLearning (EL-F), and investment of time in eLearning (EL-M). The EL-RFM (eLearning-recency,

frequency, monetary) model was designed to measure learners' motivation, and for the instructor to analyse learners' behaviours online (Chang 2010).

Similarly, Kim and his colleagues (Kim et al. 2015) proposed a LS-RFD model (learning style-recency-frequency-durability) to score the level of activities of learners for analysis and modelling. The study extracted the variables of teaching–learning activities and mapped them according to the learning style of the learner. This is to provide a measurement of the learners' preferences according to teaching–learning activity. The measurement is based on the recency, frequency and durability scores of each learning activity. From the results, user characteristics were extracted and grouped according teaching–learning activities. This is further categorized by the levels of preference and activity of each learner. In another study conducted in Hungary, the RFM model was re-classified to score online learning behaviour (Toth 2013). The classification streams of the variables of recency and frequency were applied to identify online learning behaviour. The segmentation of the two variables were done in five parts: ranging from high (80%, top 20%), to medium–high, to medium, medium–low and to low (20%, bottom 20%).

Log data as a data source

Online learning invariably revolves around the learning management system (LMS) and its online access by learners. The measurement of learner engagement is unique to the data availability of the online learning environment. In such an environment, online materials (e.g. recorded lectures, quizzes, course content and readings) are provided to the learners, and access to these online materials typically happens between the start and end of a course. An access (i.e. the action) to the online materials (i.e. the learning object) by a learner (i.e. the actor) triggers a learning event, and this is logged by the LMS in the form of an interaction, with a timestamp. Each interaction (e.g. a learner accessing a certain content or watching a video) is recorded as an instance (i.e. a row of data) (refer to Fig. 1 for an illustrative example).

In this study, however, the recency, frequency and monetary (RFM) segmentation is adapted to immediacy, frequency and duration (IFD) where recency is replaced with learner's immediacy (of access; which reflects the learner's sense of urgency or excitement to learn in an online learning context) and monetary is replaced with duration (or better known as time-on-task). Besides using the RFM

eid	username	course_id	time	action	event
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:38:29	navigate	Oj6eQgzrdqBMIaCtaq1IkY6zruSrb71b
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:38:39	access	3T6XwoiMKgol57cm29Rjy8FXVFclomxl
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:38:39	access	qxvBNYTfiRkNcCvM0hcGwG6hvHdQwnd4
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:38:48	access	2cmZrZW2h6I91itO3e89FGcABLWhf3W
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:41:49	video	N2XbfvONHjUbc4RrG6hwhXUpHhuVD1I
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:41:50	page_close	3T6XwoiMKgol57cm29Rjy8FXVFclomxl
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:42:28	video	ZzTRvqjvsdEVevZ8phOhFpLO5lyA7ruU
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:42:30	access	ep57uq0N5Fxg1Irgy68ROj8HOuD4xbsi
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:43:20	access	1Q5vBuafwY9UtwES85wN8NM0ARi2qQFS
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:43:25	access	qxvBNYTfiRkNcCvM0hcGwG6hvHdQwnd4
1	9Uee7oEuuMmgPx2IzPfFkWgkHZyPbWr0	DPnLzkJJqOOPRJfBxlHbQEERiYHu5iIa	2014-06-14T09:43:25	access	2cmZrZW2h6I91itO3e89FGcABLWhf3W

Fig. 1 Illustrative example of a source dataset (anonymized)

model as a reference to derive online engagement metrics, the RFM approach of discovering and segmenting the profiles of customers' purchasing behavioural intent (e.g. Birant 2011) is also adapted in part to describe learners' engagement level.

Associated with the discussion above, there is typically a designated starting point/date for online learning (e.g. the start of a course), a designated ending point/date (e.g. the end of all assessments or a course). These are necessary reference dates for the extraction of indicators of online engagement proposed in this study.

Immediacy (of access)

The role of immediacy behaviour in the learning context was mentioned as early as 1970s and 1980s, and is widely documented in the area of instructional communication (e.g. Mehrabian 1981). In those times, the concept of immediacy construct is characterized by teacher–student interaction (Hosek et al. 2017), where the focus of instructional methods are very much teacher-centric. However, the adoption of technology has also in part reversed the role of a learner. Teaching and learning activities have become increasingly learner-centric, where a learner is expected to play a more active role in achieving their learning outcomes. Active learning is the interactive role learners assume with course materials, instructors and peer learners in constructing new and more complex knowledge and understanding to make sense of information, interpret and solve problem (Keengwe et al. 2014). In an online learning environment, learners' active learning can be observed by their accessing log data, and by evaluating learners' interactions with the content, his peers and the instructors. And so in this study, we relate a learner's immediacy to the time lapse between the online materials' access start time and the start of his first online access. In this context, it can be expected that an increased level of online engagement is associated with a shorter Immediacy (i.e. an engaged learner is expected to be faster in accessing online materials once they become available).

Frequency (of access; or login frequency)

Research has found the learners' login frequency in an online course to be a key variable in predicting learning performance. An early study by Piccoli et al. (2001) found LMS' login frequency of learners to be highly correlated with course satisfaction. Kang et al. (2009)'s study examined the learners' online participation and frequency and their effects on academic achievement in an e-learning course. They found that the access frequency online is correlated with academic grades as well as attendance rate.

Similarly, Jo et al. (2015) highlighted the significance of frequency of learners' login as a predictor of learner's academic performance. They studied how frequently learners logged into the LMS by totalling the learner's login time and frequency in the LMS. The study showed that the (ir)regularity of the learning interval is correlated with and can predict learning performance. The regularity of the learning interval represents the learners' engagement, active participation and awareness of their learning level during the e-course.

In this study, a learner's frequency is defined as the number of episodes of online access in the maximum relevant time period (i.e. from access start time to the end of course). In the current context, it can be expected that a higher level of online engagement is associated with higher levels of frequency (i.e. an engaged learner can be expected to access the online materials more frequently; maximizing more of his learning opportunities).

Duration (of access; or total time-on-task)

Time-on-task has long been recognized to be a significant variable that is correlated with learner engagement as well as a predictor of learners' achievement. Time-on-task becomes a major component of how well learners learn and retain knowledge (Albert and Kussmaul 2008). Learners' participation time during the online course is calculated based on "login time", as it is deemed that in an eLearning course, learning usually takes place while the learners are logged on. Gunn et al. (2007) also posit that "learning takes time because it requires growing new dendrites, synapses, and neural networks" (p. 63). As the learners spent time manipulating what they are learning, it became integrated into their memories. In a research project that studied the online learning behaviours of 824 post-secondary learners, Wagner et al. (2008) found a positive co-relation between academic grades and the amount of time invested. The quantity of online time invested was calculated based on "login time". The research has noted the increased learning time with improved learning outcomes. In this study, duration represents a learner's total online access time.

Applications of LMS analysis

The analysis of the learners' log data and patterns is often performed using educational data mining (EDM). Delavari et al. (2008) suggested that the use of data mining can enable educators and their learning institutions to improve decision making within administration, to have greater accuracy in predicting learners' academic performance, to be effective in developing strategies, and to have a more efficient resource allocation. In particular, analysing the log data in the LMS can identify the at-risk learners during the course to provide the just-in-time support and guidance (van Barneveld et al. 2012). The data trails obtained from the LMS are significant and can provide key inputs with regard to the online teaching and learning processes.

The Blackboard, a widely used LMS, provides for the early prediction of learners' academic performance by analysing the learners' log data (Fountain 2016). Similarly, Purdue University implemented Course Signal, a mid-course warning system. Course Signal provides different colour signals that are the same as those of a traffic light to alert both the instructor as well as the learners of their learning status. Findings from a research study conducted by Purdue University indicated that the Course Signals assisted the at-risk learners to improve their academic performance and helped decrease attrition rates (Purdue University 2013). The study mined data from the LMS and employed learning analytics to predict key teaching and learning challenges of the at-risk learners. The application of educational and

learning analytics enabled the educators to provide timely interventions to improve learning outcomes (van Barneveld et al. 2012).

More importantly, besides harvesting useful indicators to inform educational practice and future research, this study takes a more reasonable approach of examining the deplorability of proposed indicators, and of equal importance, its extensibility to analyse learner's behaviour in other online learning courses. In using the RFM approach to learn learners' segments based on the three indicators as proposed in this study, the ease of the interpretation of findings by end-users (e.g. instructors and administrators of courses), who may not be so technically conversant, carries more weight in our consideration of our proposed methodology.

Methodology

Dataset

The data for this study were extracted from a six-week-long undergraduate course offered at a University in Singapore. The course was offered primarily to adult learners. The delivery of the course was conducted predominantly online using interactive study materials. Adult learners received guidance and support largely from an online learning environment, supplemented with some face-to-face sessions. Learning resources, pre-class quizzes, and online discussions are hosted on the University's Canvas LMS platform. Data are captured by a learner's click-action, such as, when an adult learner accesses the content or interacts with his peers in discussions, triggering an interaction log with the LMS. The dataset contained such behavioural data logged from 418 adult learners, comprising more than 100,000 instances of online access.

Data preparation

Several data processing steps will be discussed in the following sections to prepare the data for modelling to analyse learners' engagement patterns. Table 1 recalls the three indicators of online engagement proposed in this study.

Table 1 Description of online engagement metrics—immediacy, frequency and duration

Online engagement metrics	Variable name	Description
Immediacy	I_I	Measures the time lapse between the online activity access start time and the start of the first online access
Frequency	F_I	Measures the number of episodes of online access in a relevant given period
Duration	D_I	Measures the total online access time

Accordingly, the following online engagement metrics can be harvested for a particular learner based on Fig. 2.

Calculating and normalizing immediacy, frequency and duration

It is important to note that different forms of analysis can have different implications for how the data should be prepared. For the purpose of this analysis, we address some general considerations for data preparation.

It is given that the calculated values of immediacy, frequency and duration will vary according to their characteristics. In the original form, the metrics can take on different ranges and scale of measurements. For example, while Immediacy may be bounded within a range from 0 to 45, and measured in days, the values of frequency and duration can range from 0 to relatively larger numbers, and measure in different scales (i.e. counts and in minutes, respectively).

In general, learning algorithms benefited from data normalization. Furthermore, for comparability purposes, there is a need to scale individual samples to unit norm. In this regard, a min–max normalization procedure is adapted to transform the original values to within a range of 0–1. The equation is given as

$$\text{Normalized metric value} = (\text{original metric value} - \text{min}) / (\text{max} - \text{min})$$

As an illustration, suppose the minimum and maximum Frequency values are 1 and 13 for the entire class of learners. Then for learner I (with a Frequency of 5) we have,

$$\text{Normalized } F_{\text{I}} = (\text{Original } F_{\text{I}} - \text{Min } F) / (\text{Max } F - \text{Min } F)$$
$$= (5 - 1) / (13 - 1) = 4 / 12 = 0.25$$
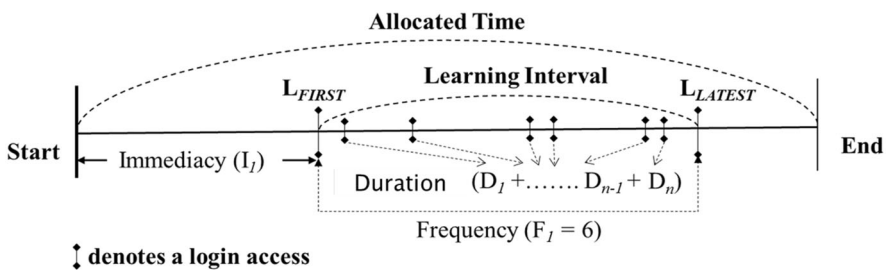
i. Instructor publishes a set of online materials ('**Start**' denotes the online materials access start time)
ii. '**End**' denotes the end of the allocated time to access this set of online activity
iii. L_{FIRST} denotes a learner's first access to this set of online activity
iv. L_{LATEST} denotes a learner's last access to this set of online activity
v. **Immediacy (I)** measures the time lapse between the a learner's first access and the '**Start**'
vi. (L_{FIRST}) to (L_{LATEST}) represents the learning interval
vii. **Frequency (F)** measures the sum of the online access count within the learning interval
viii. **Duration (D)** measures the aggregation of the time-on-task at each online access

Fig. 2 Schematic diagram—harvesting online engagement metrics (immediacy, frequency, duration) for a learner

Normalized F_I will range from 0 (when original F_I = Min F) and 1 (when original F_I = Max F). The same applies to the other online engagement metrics.

Computing an aggregate online engagement index

As discussed in the preceding section, a shorter immediacy is expected to correspond with a higher level of engagement. In deriving an aggregate online engagement index, all of its included component metrics should be measured in the same and "right" direction (i.e. a higher component metric should be associated with a higher level of online engagement). Otherwise, the component metrics can net-off each other, and the aggregate index will not be meaningful.

In the proposed framework, normalized frequency and normalized duration are in the right direction in that higher values indicate a higher level of online engagement. To set the normalized Immediacy to the right direction, the following reverse-scoring needs to be applied:

$$\text{Reversed}-\text{scored immediacy} = 1 - \text{Normalized } I_I$$

Then, taking reference from the RFM Model, it is possible to aggregate I_I, F_I and D_I into a single online engagement index (such that a higher index reflects a higher level of online engagement). The approach is to first reclassify each of I_I, F_I and D_I values into bins (i.e. in this case, into terciles). In the study, the values are split into three bins such that Bin 1 contains the lowest values and Bin 3 the highest. At the end of the reclassification, each learner will belong to three unique bins corresponding to their I_I, F_I and D_I values.

Next, each bin membership is quantified by its corresponding bin number (i.e. Bin I = I). In other words, Bin 1 is given a value of 1 (reflecting a low level of online engagement as described by the variable), Bin 2 a value of 2 (reflecting a moderate level of online engagement as described by the variable), and so on. Finally, the index is computed by summing the three bin numbers. Weights can be incorporated, if desired, to reflect the importance of the component metrics (i.e. $\text{Index}_I = \text{IFDScore}_I = W_I*\text{Bin}[I_I] + W_F*\text{Bin}[F_I] + W_D*\text{Bin}[D_I]$, where W_I, W_F and W_D are the weights of I, F, and D, respectively.

As an illustration (see Fig. 3), suppose Student A is placed in Bin 3 for each of I, F, and D. He obtains an index score of nine (given that $\text{Index}_A = \text{IFDScore}_A = 1*3 + 1*3 + 1*3 = 9$, where $W_I = W_F = W_D = 1$).

Performing clustering to identify learners' clusters

It is possible to carry out clustering to identify the natural grouping of learners according to behavioural characteristics described by their I_I, F_I and D_I values. The challenge of clustering, that is, the number of clusters is not known a priori. In this study, clustering was performed using a two-stage approach. Firstly, the data are passed through a TwoStep clustering analysis algorithm (available in IBM SPSS Modeler Version 18.0) to determine an optimal number of k clusters. Then, k-means is applied to discover the groupings. The hard clustering algorithm will group each
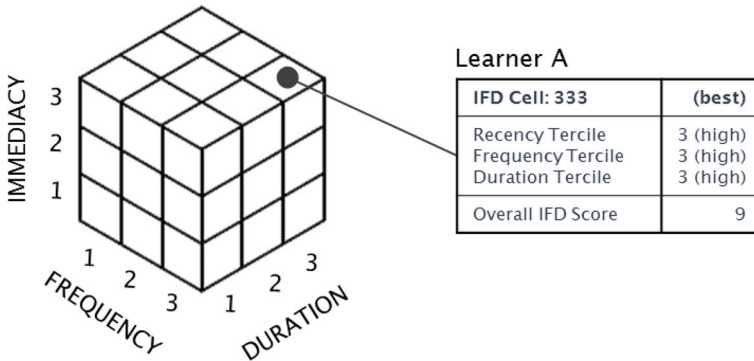
Fig. 3 Illustration of online engagement level based on IFD model

instance (i.e. in this case, each learner) into one of the clusters based on their IFD values, giving us homogenous groupings.

Scoring IFD

To illustrate the operationalization (i.e. measurement) of online engagement, the focus was on one learning activity (i.e. pre-class quiz) identified from the course. Table 2 presents samples of the original and normalized I_I, F_I and D_I values for four adult learners, in a tabular format (note that, the normalized Immediacy (normalized I) values have been reversed-scored).

As discussed earlier, en-route to computing an aggregate index, each learner will be "segmented" into three unique bins corresponding to their I_I, F_I and D_I values. It is possible to interpret an individual learner's online engagement level, relative to the entire class, as described by their I_I, F_I and D_I values.

To illustrate its application (see Table 3), suppose that Learner ID130 has been grouped into $Bin_I = 3$, $Bin_F = 3$ and $Bin_D = 1$, which indicates that this learner has demonstrated a higher level of engagement (i.e. in terms of a shorter immediacy and a higher level of frequency), as compared to Learner ID241. Learner ID130 is also associated with a higher index score, relative to Learner ID241.

In addition, it is possible to further analyse the output and interpret online engagement patterns, at a whole-of-class level (see Fig. 4). From it, we can see that close to 50% of the learners who have exhibited high immediacy, have also

Table 2 Example of learners' I, F, and D values for the learning activity

ID	Original I	Original F	Original D	Normalized I	Normalized F	Normalized D
1	7	11	14	0.649	0.058	0.002
2	17	3	7	0.059	0.012	0.001
418	7	4	81	0.649	0.017	0.010
419	2	53	8	0.941	0.302	0.001

Table 3 Example of results from IFD analysis

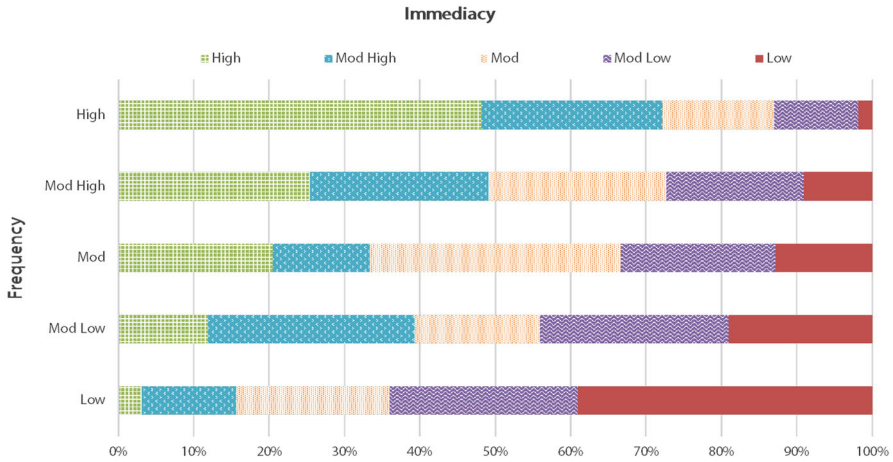| ID	Bin_I	Bin_F	Bin_D	IFD score$_i$
130	3	3	1	7
131	3	2	3	8
240	1	2	1	4
241	1	1	1	3

Fig. 4 Visualization of engagement patterns (based on immediacy and frequency)

Table 4 Learner segments, centroid values and engagement level

Centre	C_1	C_2	C_3	C_4	C_5
I	1.92	2.60	1.73	2.17	1.00
F	3.00	2.00	3.00	1.00	1.71
D	2.11	1.79	1.14	2.43	2.08
IFD score	7.03	6.40	5.87	5.60	4.79
Cluster size	37	43	85	166	87
Engagement level	High	Mod high	Moderate	Mod low	Low

demonstrated high level of online activity. Conversely, close to 40% who have exhibited low immediacy, have also demonstrated a low level of online activity.

The application of RFM segmentation (re-adapted here as IFD) results in 27 possible groupings, from a $3 \times 3 \times 3$ RFM matrix. While this application has its benefits in our proposed framework, in that, it becomes easier to interpret the learners' level of online engagement, in a relative manner. The groupings may still appear divergent, and it may be become daunting to administer or tailor targeted actions to engage 27 distinct groups.

It is possible to perform clustering to further group learners into a smaller number of groups. As discussed earlier, clustering is applied via a two-stage process. In doing so, the TwoStep clustering algorithm recommends that $k = 5$ (k = the number of clusters).

To illustrate the utility of clustering and its output, please refer Table 4. The application of clustering gives us five natural groupings of learners, explained by the centroid value of I, F, D, and IFD score. The centroid value is the average of the data points in a cluster. In the context of this application, the centroid of I, F, and D is any value ranging from 1 to 3, whereas, the centroid of IFD score is anything from 3 to 9. The group's average IFD score is used to estimate their level of online engagement.

It is possible to further analyse the output and interpret online engagement patterns. For example, the results demonstrate that learners from cluster C_1 are more engaged compared to learners from the other clusters. C_1 consists of learners who have accessed the online materials more often. They belong to the group who have demonstrated moderate urgency to engage with the online materials, and also spent a fair amount of time on the materials like an average learner in class. On the contrary, learners in cluster C_5 are somewhat more dormant. They show a lack of enthusiasm (i.e. unhurried) in learning, and appear to be disinclined to exert more effort to engage with the online materials. Thus, they are our learners with low level of online engagement.

Discussion

The silhouette function is used to interpret and validate cluster validity (e.g. Battaglia et al. 2015; Jun and Lee 2010). The average silhouette coefficient of 0.6 has been shown empirically to be effective given a clear and distinct cluster structure. The study has also taken considerable steps in ensuring the validity of the measurements. Past works have been consulted to support the derivation of the proxy measures of online engagement. Both the methodology and its results are expected to be replicable in other studies of the online engagement behaviour of adult learners.

The outputs from this analytics procedure is expected to advance analytical procedures for decision-making in the area of teaching and learning, instructional designs, and student support. For instance, online activities that have fallen short of the desired engagement level can be re-examined for its effectiveness at the whole-of-class level. Related to this, online activities that have not helped students achieve their learning outcomes may require reinforcement of instructional methods.

The measurement of an adult learners' online engagement behaviour is an important precursor for deeper analysis of student learning at the individual level: from the perspective of adult learning, and especially in the shoes of an adult learner who may have to juggle the demands of work, life and academic progression. Following this, it will be interesting to examine engagement levels with antecedents and consequents of learning, that is, the degree of association of an adult learner's engagement with the course and his course performance. In addition to this, it will

be interesting to uncover further unknown patterns or trends from learner engagement to aid educational practice and research at the university level.

Conclusion

The study presented findings from an online learning activity that is targeted at adult learners. Firstly, this study, while being simplistic in its approach, explored possible data features as proxy measures of online engagement based on the LMS log data presented. Although the use of similar count-based and time-based measures have been frequently discussed in the existing literature, the potential of using the "Immediacy" variable as one of component metrics for estimating online engagement was discussed in this study. The study also looks at re-adapting a well-proven approach in behavioural segmentation in the RFM Model, and hopefully gives fresh impetus to it in the learning analytics context for modelling learning engagement behaviour. Extending this approach, the use of clustering was also examined to discover distinct learner segments from multidimensional data.

Secondly, good conceptualization and operationalization are expected to facilitate productive utility of the IFD approach and its metrics. Very importantly, future research can be conducted to examine and test the reliability, validity and utility of the proposed online engagement metrics. In line with this, future research can be conducted to include the antecedents and consequents of online engagement. (That is, to study if certain learner, course or other attributes are associated with certain patterns of online engagement.) In addition, it will be useful to know if particular patterns of online engagement, together with other potential determinants of academic performance, are associated with particular levels of academic performance, for example. Such findings can aid teaching and learning and optimize their outcomes.

Thirdly, there are several limitations arising from the present study. The limitations affect both the results and the directions for future research. Some of the limitations are embedded within an overarching limitation of how learner engagement is defined in an online learning environment. Whilst the primary focus of our study is on discovering useful and meaningful data features of online engagement from traces of learners' online behavioural data. It is important to note that these measures are proxy indicators of online engagement, and may also not represent an all-encompassing measure of the extent of online engagement. The study is further limited by the Canvas LMS data source that is accessible and available to us at present. Addressing the limitations highlighted above can provide some future directions for the research. More generally, future research can derive new online engagement metrics to aid educational research and practice in the technology-enhanced learning environment.

Finally, it is the spirit of producing open, replicable and extensible research, which drives this study. It is hoped that this study can provide the groundwork, from a practical perspective, to better model online engagement behaviour.

References

Albert, S., & Kussmaul, C. (2008). Why wikis are wonderful for writing. In T. Carter & M. Clayton (Eds.), *Writing and the iGeneration: Composition in the computer-mediated classroom* (pp. 49–67). Southlake, TX: Fountainhead Press.

Allen, I. E., & Seaman, J. (2010). *Learning on demand: Online education in the United States, 2009.* Newburyport, MA: Sloan Consortium. Retrieved from http://www.eric.ed.gov/PDFS/ED529931.pdf

Anderson, E. (2017). *Measurement of online student engagement: Utilization of continuous online student behaviors as items in a partial credit rasch model* (Doctoral dissertation). Retrieved from https://digitalcommons.du.edu/etd/1248

Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *The International Review of Research in Open and Distributed Learning, 12*(3), 80–97.

Axelson, R. D., & Flick, A. (2011). Defining student engagement. *Change: The Magazine of Higher Learning, 43*(1), 38–43.

Battaglia, O. R., Di Paola, B., & Fazio, C. (2015). Cluster analysis of educational data: An example of quantitative study on the answers to an open-ended questionnaire. *arXiv preprint* arXiv:1512.08998.

Birant, D. (2011). In K. Funatsu (Ed.), Data mining using RFM analysis, knowledge-oriented applications in data mining. ISBN: 978-953-307-154-1, InTech. Retrieved from: http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-using-rfmanalysis

Brown, K. G., Charlier, S. D., & Pierotti, A. (2012). E-learning at work: Contributions of past research and suggestions for the future. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology* (Vol. 27, pp. 89–114). Oxford: Wiley-Blackwell.

Bruner, E. (2013). Mobility: It's about user adoption. *Chief Learning Officer, 13*(3), 34–37.

Beer, C. Clark, K., & Jones, D. (2010). Indicators of engagement. In C. H. Steel, M. J. Keppell, P. Gerbic & S. Housego (Eds.), *Curriculum, technology & transformation for an unknown future.* In *Proceedings ascilite Sydney 2010* (pp. 75–86)

Chan, C. C. H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert Systems with Applications, 34*(4), 2754–2762.

Chang, H. C. (2010). Developing EL-RFM model for quantification learner's learning behavior in distance learning. In *Education Technology and Computer (ICETC), 2010 2nd International Conference on* (Vol. 4, pp. V4–452). IEEE.

Chiu, C.-Y., Kuo, I.-T., & Chen, P.-C. (2009). A market segmentation system for consumer electronics industry using particle swarm optimization and honey bee mating optimization. In S.-Y. Chou (Ed.), *Global perspective for competitive enterprise, economy and ecology* (pp. 681–689). London: Springer.

Cleveland-Innes, M., & Ally, M. (2013). Affective learning outcomes in workplace training: A test of synchronous vs. asynchronous online learning environments. *Canadian Journal of University Continuing Education, 30*(1), 15–35.

Coates, H. (2007). A model of online and general campus-based student engagement. *Assessment & Evaluation in Higher Education, 32*(2), 121–141. https://doi.org/10.1080/02602930600801878.

Delavari, N., Beikzadeh, M. R., & Phon-Amnuaisuk, S. (2008). Data mining application in higher learning institutions. *Informatics in Education, 7*(1), 31–54.

Fountain, R. S. (2016). *Searching for Predictors of Success in Community College Online Courses* (Doctoral dissertation, Appalachian State University).

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson (Ed.), *Handbook of research on student engagement* (pp. 763–782). New York: Springer.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71.

Gunn, A. M., Richburg, R. W., & Smilkstein, R. (2007). *Igniting student potential: Teaching with the brain's natural learning process.* Thousand Oaks: Corwin Press.

Harbaugh, A. G., & Cavanagh, R. F. (2012). *Associations between the classroom learning environment and student engagement in learning 2: A structural equation modelling approach*. Retrieved from https://files.eric.ed.gov/fulltext/ED542256.pdf.

Hosek, A. M., Houser, M. L., & Richmond, V. P. (2017). Immediacy and the teacher–student relationship. In M. L. Houser (Ed.), *Handbook of instructional communication* (pp. 115–129). Abington: Routledge.

Hu, Y. H., Lo, C. L., & Chung, F. I. (2012). Applying RFM analysis to library-book recommender systems. *Commun. ICISA, 13,* 79–95.

Jin, S. H. (2017). Using visualization to motivate student participation in collaborative online learning environments. *Journal of Educational Technology & Society, 20*(2), 51–62.

Jo, I. H., Kim, D., & Yoon, M. (2015). Constructing proxy variables to measure adult learners' time management strategies in LMS. *Educational Technology & Society, 18*(3), 214–225.

Jun, S., & Lee, S. (2010). Empirical comparisons of clustering algorithms using silhouette information. *International Journal of Fuzzy Logic and Intelligent Systems, 10*(1), 31–36. https://doi.org/10.5391/ijfis.2010.10.1.031.

Jung, Y., & Lee, J. (2018). Learning engagement and persistence in massive open online courses (MOOCS). *Computers & Education, 122,* 9–22.

Kang, M., Kim, J., & Park, I. (2009). The Examination of the variables related to the students' e-learning participation that have an effect on learning achievement in e-learning environment of cyber university. *Journal of Korean Society for Internet Information, 10*(5), 135–143.

Keengwe, J., Onchwari, G., & Agamba, J. (2014). Promoting effective e-learning practices through the constructivist pedagogy. *Education and Information Technologies, 19*(4), 887–898.

Kim, K., Choi, Y. J., Kim, M., Lee, J. W., Park, D. S., & Moon, N. (2015). Teaching-learning activity modeling based on data analysis. *Symmetry, 7*(1), 206–219.

Kuczera, M. (2017). Striking the right balance: Costs and benefits of apprenticeship. *OECD Education Working Papers*, (153), 0_1.

Leach, L., & Zepke, N. (2011). Engaging students in learning: a review of a conceptual organiser. *Higher Education Research & Development, 30*(2), 193–204. https://doi.org/10.1080/07294360.2010.509761.

Lee, Y. (2018). Effect of uninterrupted time-on-task on students' success in Massive Open Online Courses (MOOCs). *Computers in Human Behavior, 86,* 174–180.

Mehrabian, A. (1981). *Silent messages: Implicit communication of emotion and attitude*. Belmont, CA: Wadsworth.

National Survey of Student Engagement. (2012). *Promoting student learning and institutional improvement: Lessons from NSSE at 13*. Bloomington, IN: Indiana.

Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A Research framework and a preliminary assessment of effectiveness in basic IT skill training. *MIS Quarterly, 25*(4), 401–426.

Purdue University. (2013). *Course signals*. Retrieved August 14, 2017, from http://www.itap.purdue.edu/learning/tools/signals/.

Siemens, G. (2013). Massive open online courses: Innovation in education? In R. McGreal, et. al. (Eds.), *Open educational resources: Innovation, research and practice* (chap. 1, pp. 5–15). Athabasca: Athabasca University Press.

Toth, P. (2013). Online learning behavior and web usage mining. *WSEAS Transactions on Advances in Engineering Education, 10,* 71–81.

Trowler, V. (2010). Student engagement literature review. *The Higher Education Academy, 11,* 1–15.

van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). *Analytics in higher education: Establishing a common language (white paper)*. Boulder, CO: EDUCAUSE Learning Initiative.

Wagner, P., Schober, B., & Spiel, C. (2008). Time students spend working at home for school. *Learning and Instruction, 18,* 309–320.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth, 10*(1), 13–53. https://doi.org/10.1080/17538947.2016.1239771.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education, 29,* 23–30.

Adam Wong is a Research Analyst with the Institutional Research and Analytics Unit at the Singapore University of Social Sciences (SUSS). His current research interests are primarily in the areas of data and text mining, along with their applications in the learning analytics space. At the SUSS, he assists the principal investigator in internally funded analytics projects.

Sylvia Chong is an associate professor with the Institutional Research and Analytics Unit at the Singapore University of Social Sciences (SUSS). Her research interests are interdisciplinary in nature, and include substantive and methodological areas of higher education research. At the SUSS, she is the principal investigator for several institutional analytics projects as well as a couple of externally funded projects.