JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 38, 1109-1131 (2022)
DOI: 10.6688/JISE.202211_38(6).0002

Exploring Time-Related Micro-Behavioral Patterns in a Python Programming Online Course

HSI-MIN CHEN[1], BAO-AN NGUYEN[2,+], CHYI-REN DOW[1],
NIEN-LIN HSUEH[1] AND AN-CHI LIU[1]
[1]*Department of Information Engineering and Computer Science*
Feng Chia University
Taichung, 40724 Taiwan
E-mail: {hsiminc, acliu}@fcu.edu.tw, {crdow, nlhsueh}@mail.fcu.edu.tw
[2]*Department of Information Technology*
Tra Vinh University
Tra Vinh, 87000 Viet Nam
E-mail: annb@tvu.edu.vn[+]

Modelling learning behaviors and predicting student performance in massive open online courses (MOOCs) are vital for adaptive course planning and personalized intervention. This study proposes a new approach for discovering time-embedded behavioral patterns in micro behaviors of MOOC learners and incorporating them as features for student profiling and learning performance prediction. We embedded discretized time intervals into interaction sequences and used *n*-gram extraction to output time-related behavioral patterns. With log data from a Python programming MOOC with 591 learners, we exploited exploration data analysis, unsupervised, and supervised learning to elucidate the associations between time-related behavioral patterns and academic performance. Nine out of seventeen targeted patterns are highly correlated with the final grade, in which, three patterns related to the help-seeking, evaluation, and study activities with short or medium intervals (less than two minutes) are strong predictors of academic performance in a very early stage. The time-related behavioral patterns also serve as good features for clustering learners into three groups based on learning behaviors: Sampling learners, Comprehensive learners, and Targeting learners. Our empirical results show the usability of the proposed time-embedded behavioral patterns in immediate diagnosis learners' engagement, raising new challenges for learning analytics with time concerning to achieve precision education.

Keywords: MOOCs, learning analytics, time-embedded *n*-grams, time-related behavioral patterns, student performance prediction

1. INTRODUCTION

In recent years, MOOCs has become an important learning environment for computer science students. Courses related to computer science and programming are always in the list of most popular MOOCs of all time [1]. However, besides obvious benefits from the open online form of learning, MOOCs also encounter challenges related to their intrinsic nature, which are high dropout and low completion/success rates. In the past few

Received November 1, 2021; revised December 20, 2021; accepted January 12, 2022.
Communicated by Shin-Jie Lee.
[+]Corresponding author.

years, researchers have paid remarkable attention to modeling student behaviors, developing predictive models to finding out factors that influence academic performance, dropout, and success/failure of students in courses. The prediction of student success enhances the performance of MOOCs in a variety of tasks. First, it supports individualized timely interventions by determining actionable factors that can improve the learning experiences of students. Second, it is beneficial in adaptive content and learner pathways by presenting course content and learning experiences that were optimized for student success. Third, building prediction models help to understand the data such as learner behaviors, learner attributes, and course attributes. Explainable prediction models provide useful insights behind the outcomes that enable interventions to reach the success of both the learners and the courses [2].

Among types of data extracted from MOOCs platforms, clickstreams data and concept-oriented assignments are best predictors for student performance [3]. Clickstreams contain a variety of meta-data that allows reconstructing student behaviors by several schemes and extraction of meaningful features from discrete events under every mouse-clicks like page load, video play/stop/pause/skip, assessment submit, forum posts, *etc*. [2]. The common way is constructing counting-based activity features, *i.e.*, *number of viewed videos, number of tried assessments, number of questions in the forum*, *etc*. However, since counting features did not capture dynamic behaviors of students, the interventions would be limited in improving the frequency/quantity of learning activities associated with these summative features [4].

In a more advanced way, dynamic behaviors are modelled in form of *higher-order activity-based features*. The features are extracted from sequences of clicks/activities and named as interpretable behaviors [5–7]. For example, the help-seeking pattern might occur when a student starts playing a lecture video while doing an exercise [8]; a rewatch pattern occurs when a student is repeatedly pressing seek-back and play buttons in a lecture video [9]. Discovery of these features requires complex data analytics techniques or advanced tools like frequent sequential pattern mining [10] , process mining [8,11], motif identification [12] , *n*-gram extraction [9,13,14], *etc*. Moreover, interpreting the sequential patterns as learning behaviors required deep expertise. In return, the high-order activity-based features have shown their contribution in improving prediction performance and offering more precise interventions based on effective dynamic behaviors [3,14].

Despite their success, existing studies using high-order activity-based features in MOOCs only consider the presence of specific activity/actions in clickstreams without inspecting the *time intervals* between them, even time-gap is an intrinsic attribute of interaction sequences. Given a sequence of actions, the intervals between actions intuitively reveal how a learner is dealing with the current learning object. For example, the intervals between two consecutive submit actions in an assessment give us some implications about the difficulty of the problem *(i.e., easy problems can be solved in a shorter time or vice versa, etc.)* or the academic performance of learners *(i.e., better students can solve the problems in a shorter time or vice versa, etc.)*. It is no doubt that analysis on the time aspect of interaction sequences can gain meaningful insights for behavior steering in MOOCs.

To fulfill this research gap, we propose a novel temporal sequence mining approach to explore the time-related behavioral patterns in MOOCs. Discretized time gaps are augmented into the interaction sequence prior to *n*-gram extraction to construct *time-*

embedded n-grams. These patterns then are interpreted to identify activity at the micro-level. With the belief that temporal insights are meaningful indicators for precise interventions in MOOCs, we address the efficiency of time-related behavioral patterns in two important problems of education, *i.e.*, learner profiling and academic performance predicting. Therefore, we proposed 3 following research questions:

- **RQ1:** What is the predictive effectiveness of time-related behavioral patterns in the early prediction of learning performance?
- **RQ2:** What time-related behavioral patterns are important in the early prediction of learning performance?
- **RQ3:** What learner profiles would be detected based on the time-related behavioral patterns?

2. LITERATURE REVIEW

Among types of variables, the most commonly used are those related to interactions in videos [14, 15], exercises [16], and forum [17]. Clickstreams contain logs of client request to the platform's web server that allow reconstruction of student behaviors at multiple levels of aggregation. As clickstreams are raw, the events behind students' clicks should be aggregated and combined with the context given by the metadata of the current learning resource to construct meaningful and interpretable features. *Forum posts* is another hint of learner engagements and their interest in the lessons. The text content of forum posts reveals the problems that learners concern [17]. The connection network formed between learners in the forums can also be mined to extract social features [5]. *Quizzes and exercises* are assessment activities that are high correlated with student completion. Exercise variables are the most powerful predictors for learning performance prediction [18]. Demographics has shown its impact on interpreting learner retention [19]. However, when compared with other kinds of data, the demographics-based models have minimum predictive power [13, 18]. The most robust predictors were those variables related to exercises, and followed by clickstreams [18].

Regardless of good prediction results, analysis using counting-based feature do better on giving understanding about factors affecting learning performance rather than providing a good basis for interventions. Therefore, researchers have tried to observe learner behaviors in their contexts using temporal and sequential activity models, to find out interpretable learning patterns/strategies affecting student success. Based on theses patterns, personalized interventions can be given to learners from any stage of the course.

Hidden Markov models *(HMM)* were selected as suitable choices to extract behavioral representations of learners' activities since they can model latent characteristics of learners and their transitions in a probabilistic graphical model. Balakrishnan and Coetzee [20] captured weekly frequencies of video-watching, forum thread viewed and posted, progress page checked to construct the HMM and predict dropout. Geigle and Zhai [21] employed two-layer HMM to extract interpretable behavioral patterns from clickstreams that correlate with the learning outcomes. Ramesh *et al.* [22] proposed a statistical relational learning framework, called hinge-loss Markov random fields, to interpret learner engagement via a combination of behavioral, linguistic, structural, and temporal cues.

The model demonstrated the ability in predicting student success from the early stage as well as in understanding learners' engagement at multiple levels.

Brooks *et al.* [13] examine the incremental changes in performance with each additional day by captured student interactions in 4 different granularities of time frames *(one day, three days, one week, and one month)* and encoded the interactions between students and resources into a set of time series features. They used *n*-grams to capture the co-occurrence of features in a timeframe. With a large combination of 1221 features made up from multiple settings of time frame and *n*-gram, they demonstrate good performance over the first 3 weeks of evaluated MOOCs. Fei and Yeung [23] captured the incremental in measures activities over weeks using a sequential model. They considered seven activities on lecture, quiz, and forum of learners and fetched the weekly activity feature vectors as input to a Long Short Term Memory (LSTM) neural network to predict dropout. LSTM modeling overcomes the limitations of HMM on the assumption that the next state depends only on the current state, so that, it has been employed and delivered a high predictive performance in later studies [10].

The interaction logs contained meaningful sequential patterns which can be mapped into interpretable high-order activity-based features to build the prediction models. However, sequential pattern mining algorithms, *e.g.* cSPADE [24], often generate a high amount of patterns and require extremely high computational/memory cost when the minimum support is set too low. Fortunately, there are several alternative solutions to discover sequential patterns from interaction data that generated fruitful results. Maldonado-Mahauad *et al.* [8] used process mining to identify self-regulated learning (SRL) strategies *(Study, Rehearsal, Goal-setting, Elaboration, Help-seeking, Reviewing-record)* based on the transitions of learners over videos and assessments. The discovered SRL strategies were then used as features to segment learners into three clusters including *Targeted learners, Comprehensive learners, and Sample learners*. In a later study, they showed that event-based SRL strategies illustrated very high predictive power in dropout prediction [18]. Brinton *et al.* [12] leveraged the motif identification tool used in bioinformatics to extract behavioral motifs from video-watching clickstreams. Especially, they embedded the time intervals between events before motif extraction to obtain time-respected video-watching behavior patterns. The motifs were grouped and interpreted as four patterns *(Reflecting, Reviewing, Skimming, and Speeding)*. Some of the discovered behavioral motifs are significantly correlated with the likelihood that a student will be Correct on First Attempt *(CFA)* or give up answering a quiz question.

By adapting NLP and text mining tools to sequential mining, the *n*-grams features at coarse-grained levels can denote learner habits or frequent transitions over learning units [6]. For example, the 3-gram related to video *(V-V-V)* represents the behavior pattern of *"watchers"*, the 3-gram related to quizzes *(Q-Q-Q)* represents the behavior of *"quiz harvesters"* [25], and the longer *n*-grams can denote the learning paths. Predictive models build from *n*-grams features do not only show acceptable results but also demonstrate learning behavioral patterns that are associated with learning performance/retention.

At the fine-grained level, *n-grams* extracted from consecutive clicks can be interpreted as intended activities which can discriminate different types of learners. Sinha *et al.* [9] extracted 4-grams from students' clicks on videos, grouped them into seven behavioral actions in a self-defined taxonomy *(Rewatch, Skipping, Fast watching, Slow watching, Clear concept, Checkback reference, and Playrate transition)*, and constructed a quantitative information processing index to interpret student engagements in the cognition perspective. Yu *et al.* [14] inherited the behavioral patterns by Sinha *et al.* [9] and built prediction models for learning outcomes with the highest prediction results were given by the artificial neural network.

There are two important remarks in the aforementioned studies. First, most of them (except [9, 12, 14]) focused on macro-behaviors regarding the interactions between learners with the learning resources (extracted from the coarse-grained level of clickstreams, *e.g.* watching-video, doing-assessment, *etc.*). The patterns extracted from transitions of the learner over learning resources can be viewed as learning strategies ([3, 8], or learning paths [6, 25, 26] that can be used as predictors for learner success prediction. Second, the temporal dimension was considered in time slices in which the learning progress is observed via weekly increments whereas the time gaps between activities were mostly ignored (except in [12], regardless that the temporal aspect related to temporal perspective and self-regulation in learning [27]. Intuitively, time intervals between learners' clicks on the learning resources related to *time-on-task*, the time learners are actively engaged in learning. In the computer-based learning environment, time-on-task is positively correlated with the learning outcomes [28], positively related to learning performance, and increases with task difficulty [29]. Inspecting the time-on-tasks in a MOOC, Lee [30] stated that uninterrupted time activities with longer duration are correlated with student success. Again, these studies only focus on the duration in macro-behaviors of students.

In this paper, we aim to explore the time intervals in the micro-behaviors extracted from the finest-grained level of clickstream data. We propose a new term, call *time-related behavior pattern*, alongside its mining approach called *time-embedded n-gram.* Via discovered patterns, we aim to understand the temporal aspect of micro-behaviors of learners in the two most common activities in MOOCs which are watching videos and doing assessments. The associations between time-related behavior patterns with student success are analyzed both in student profiling and learning performance prediction.

3. METHOD

3.1 Clickstream Data and Data Extraction

Since we exploit the data from a MOOC platform developed based on Open Edx, the extracted log data includes JSON objects representing interactive events of learners. We defined two concepts for the data extraction stage which are interaction and session, as follows:

- **Interaction:** In this study, we focus on the interactions related to video watching and assessment submissions (Table 1). We observed six events recorded when learners interact with lecture videos: *load, play, pause, seek back, seek forward, and stop* and the dual-events related to assessment including two consecutive events which are *"problem-check"*, raised when the learner presses the submit button, and *"problem-graded"*, raised when the problem is graded by the platform and the result is shown on learner's browser.
- **Session:** Since learners usually open the courses and let the web browsers keeping the connection to the platform day by day without logging out, hence, a session

Table 1. Interaction events from Open edX clickstreams.

Event type	Event	Open edX event name	Event descriptions
Video	Lo	video_load	load video
	Pl	video_play	play video
	Pa	video_pause	pause video
	St	video_stop	stop video
	Sf	video_seek	seek forward
	Sb	video_seek	seek backward
Assessment	Pc	problem_check	problem submitted
	Pg	problem_graded	problem graded

Table 2. Example of an sequence database S and the time-interval sequential database Q derived from S.

SID	Sequence database S
10	$((Lo,0),(Pl,0),(Pa,30),(Pl,80),(St,180))$
20	$((Lo,0),(Pl,0),(Pa,35),(Pc,35),(Pg,36),(Pl,39),(Pc,59))$
30	$((Pc,0),(Pg,1),(Pc,31),(Pg,32),(Lo,52),(Pl,53),(St,70)$
SID	Time-interval sequence database Q
10	$((Lo,0),(Pl,30),(Pa,50),(Pl,100),(St,0))$
20	$((Lo,0),(Pl,35),(Pa,0),(Pc,1),(Pg,3),(Pl,20),(Pc,0))$
30	$((Pc,1),(Pg,30),(Pc,1),(Pg,20),(Lo,1),(Pl,17),(St,0))$

should be denoted by a sequence of interactions with the inactivity duration no longer than a threshold. This threshold was subjective to the researchers' opinion which varies around 30, 45, or 60 minutes [31].

3.2 Time-Related Behavioral Pattern Mining Using Time-Embedded n-grams

Formally, interactions can be described as a sequence database, in which each transaction presents one sequence of interactions as defined by Chen and Huang in [32] as follows:

Definition 1: A sequence s is presented as $s = ((a_1,t_1),(a_2,t_2),...,(a_n,t_n))$ where a_j is an item and t_j stands for the time at which a_j occurs, $1 \leq j \leq n$ and $t_{j-1} \leq t_j$ for $2 \leq j \leq n$.

A time-interval sequence q can be derived from the sequence s by replacing the timestamp of the item a_j with the time gaps between it and its adjacent item. It is defined as follows:

Definition 2: Given $s = ((a_1,t_1),(a_2,t_2),...,(a_n,t_n))$ as a sequence, a time-interval sequence q derived from s is presented as $q = ((a_1,i_1),(a_2,i_2),...,(a_n,0))$ where a_j is an item and $i_j = t_{j+1} - t_j$ stands for the time-interval between the timestamps of item a_j and item a_{j+1}, for $1 \leq j < n$.

Considering learner's interactions as an sequence database S, we can obtain a time-interval sequential database Q containing all interaction sequences in S, as shown in the example in Table 2.

From the perspective of data mining, the purpose of a typical algorithm for sequential pattern mining (SPM) respecting time intervals is to discover all time-interval sequential patterns whose frequencies exceed a given threshold called minimum support. In a real context, this approach encounters two following problems:

i) Time constraint: Since it is not possible to count the frequencies of continuous values, the time values should be discretized into a small number of finite values to enable summative operations (Han *et al.*, 2011). For example, Chen *et al.* (2003) embedded discretized time-interval into conventional mining algorithms and proposed I-Apriori and I-PrefixSpan to find the whole set of frequent time-interval sequential patterns. The discovered patterns are sequences of items with the time-intervals in the middle, for example, (a, I_1, b, I_2, c) with a, b, c are items and I_1, I_2 are discrete values or ranges of time-intervals. We reuse this approach in our method.

ii) Interpretability of patterns: We might not need all patterns to be discovered because of three reasons. First, finding all possible patterns is time and resource-consuming. Second, not all patterns are meaningful to the study context. Third, it is challenged to interpret or summarize the whole set of discovered patterns [33]. Hence, the mining problem can be limited by certain constraints to reduce the computation cost and enable the interpretability of patterns.

Regarding time constraints, we can discretize continuous time-intervals to enable counting in mining algorithms. Since interval discretization and quantile discretization are sensitive to outliers and do not consider the distribution of data, we employed clustering-based discretization using *k*-means to this regard. Note that *k* should be large enough to maintain the differences between time intervals in students' behaviors, as well as small enough to limit the number and enable the interpretability of generated time-embedded patterns. Based on this heuristic, $k = 5$ was adopted to obtain five ranges of duration, namely *Very-Short(1), Short(2), Medium(3), Long(4), and Very-Long(5)*. The discretized time intervals are then embedded into the interaction names and form a new format of sequential data called *time-embedded (TE) sequences*, as shown in Table 3, where discretized time-embedded intervals are represented as subscripts.

Regarding the interpretability of patterns, we can reduce the mining problem with two constraints: *i)* The patterns should present the continuous sequence of learner actions, so they contain no gap; *ii)* The patterns should be easy to be interpreted, so they should not be too long. Fortunately, subsequences with no gap can be discovered with less effort, in terms of both computation cost and memory usage, using the well-known natural language processing technique *n*-gram.

An *n*-gram is a sequence of *n* consecutive characters/words/events which can be extracted by using a sliding window on the input sequence [34]. The length of resulted patterns can be specified by the width of the sliding window. The support of each *n-gram* is counted by the frequency of that *n*-gram appearing in the sequence database. Using the support, we can select top *k n-grams*, or select all frequent *n-grams* using a specific threshold of support, like in general pattern mining algorithms.

Definition 3: *Time-embedded n-grams* are those *n-grams* extracted by applying *n-gram extraction* on time-embedded sequences.

Given a clickstream dataset *D*, the problem of time-interval SPM from *D* can be well reduced to the sub-problem called *time-embedded n-grams mining*. An example of *time-embedded-4-grams* extraction from a time-interval sequence is presented in Table 3. With the high number of time-embedded-*n*-grams discovered, they should be interpreted as learning behaviors.

Table 3. An example of time-embedded n-gram extraction.

Time-interval sequence	$((Lo,0),(Pl,30),(Pa,50),(Pl,100),(St,1),(Lo,1),(Pl,50),(St,0))$
Time-embedded sequence	$(Lo_1,Pl_2,Pa_2,Pl_3,St_1,Lo_1,Pl_2,St_1)$
Time-embedded 4-grams	(Lo_1,Pl_2,Pa_2,Pl_3); (Pl_2,Pa_2,Pl_3,St_1); (Pa_2,Pl_3,St_1,Lo_1); (Pl_3,St_2,Lo_1,Pl_2); (St_1,Lo_1,Pl_2,St_1)

Note. Time-interval ranges: $(1:[0,13];\ 2:[14,50];\ 3:[51,109];\ 4:[110,194];\ 5:[195,300])$

Table 4. Summary of time-related behavioral patterns.

Group	ID	Description	Examples of n-grams
Evaluation	EV-1	Very-short evaluation	(Pc_1,Pg_1,Pc_1,Pg_1)
	EV-2	Short evaluation	(Pc_1,Pg_2,Pc_1,Pg_2)
	EV-3	Medium evaluation	(Pc_1,Pg_3,Pc_1,Pg_2)
	EV-4	Long evaluation	(Pc_1,Pg_4,Pc_1,Pg_1)
	EV-5	Very-long evaluation	(Pc_1,Pg_5,Pc_1,Pg_2)
Study	ST-1	Very short study	(Pl_1)
	ST-2	Short study	(Pl_2)
	ST-3	Medium study	(Pl_3)
	ST-4	Long study	(Pl_4)
	ST-5	Very long study	(Pl_5)
Help-seeking	HS-1	Very short help-seeking	(Pg_1,Lo_1,Pl_1,Pc_1)
	HS-2	Short help-seeking	(Pg_1,Lo_1,Pl_2,Pc_1)
	HS-3	Medium help-seeking	(Pg_1,Pl_2,Pa_3,Pc_1)
	HS-4	Long help-seeking	(Pg_2,Lo_1,Pl_1,Pa_4)
	HS-5	Very long help-seeking	(Pl_1,Pc_1,Pg_5,Pa_1)
Video seeking	RW	Rewatch	(Pl_2,Sb_1,Sb_1,Pl_2)
	SK	Skipping	(Sf_1,Sf_1,Sb_1,Pl_2)

3.3 Mapping Time-Embedded-n-grams to Time-Related Behavioral Patterns

After time-embedded n-grams are extracted, they should be mapped into time-related behavioral patterns to explore students learning strategies that they imply. Since it is not easy to interpret too long patterns, we specified $n = 1$ and $n = 4$ for n-grams extraction as it was recommended in previous studies [9, 14, 25]. To simplify the naming scheme, a pattern is named based on the interactions and the longest time interval within it.

In the literature, there have been some pioneering works regarding mapping sequential patterns in MOOCs to learning activities or *self-regulated learning (SRL)* strategies [8, 9, 12]. For example, the pattern involving "assessment try" $\rightarrow$ "video lecture" are interpreted as *Help-seeking* strategy; the patterns involving "only assessment" related to *Elaboration* or *Evaluation* strategies; the patterns involving to "only video-lecture" related to *Study* or *Rehearsal* strategy, *etc*. Although the labels for the patterns are subjective, the results of these studies show the feasibility of interpreting learner engagements as well as predicting learning performance. In the level of micro-behavior, we observe that three out of six patterns discussed by Maldonado-Mahauad *et al.* [8] can be reused to identify learners' micro-behaviors from clickstream as follows:

- **Only assessment:** Those patterns that contain only assessment interactions (Pc and Pg) can be considered as *Evaluation(EV)* behaviors. In *n-gram* representation, it should contain at least two consecutive submitting actions. For example, the pattern (Pc_1, Pg_3, Pc_1, Pg_2) implies a *"Medium Evaluation"* when the learner continues to the second problem after a medium time interval.

- **Only videos:** As all the lectures of our MOOC are given in videos, the video-watching activity can be considered as a *Study(ST)* activity. We consider the monograms of *Pl* to inspect the single interaction of video watching. For example, the monogram (Pl_3) implies a *"Medium Study"* in which the learner watches videos in a medium time interval.
- **Mixture of assessment and video:** This pattern represents the behavior in which a student finds help from learning materials (videos) while doing assessments, so it can be labeled as *Help-seeking(HS)* behavior. For example, the pattern (Pg_1, Pl_4, Pa_2, Pc_1) implies that after finished an assessment problem, the learner needed help to solve the next problem, so he pressed playing the lecture video, watched it for a long duration to review the lesson, and presses pause before return to the assessment.
- In addition, we learn that *video seeking* is the most frequent interaction in the clickstreams, so we also target two addition patterns related to seeking behaviors in video watching: *Rewatch (RW)* and *Skipping (SK)*. Since the seeking interaction sequences often contain multiple continuous clicks of seeking within very short time intervals (short than two seconds), we do not care about the time-related information in these patterns and only consider those *4-grams* containing at least one *Pl* interaction. The patterns are named based on the seeking action appearing in the *4-grams* which are *Rewatch* for *Sb* and *Skipping* for Sf. Those *4-grams* containing both *Sb* and Sf action will be named based on the seeking action that earlier appears in the patterns. For example, the 4-grams (Pl_1, Sf_1, Sf_1, Sb_1) implies a *"Skipping"* pattern.

With 5 discretized values of time intervals and three patterns *(Evaluation, Study, and Help-Seeking)*, together with *RW* and *SK*, totally we have 17 *time-related behavioral patterns* which can be extracted from the clickstreams (Table 4).

3.4 Research Procedure

The workflow for exploring time-related behavioral patterns from clickstream data to answer the research questions is depicted in Fig. 1. It can be divided into two stages which are pattern extraction and pattern analysis. In the first stage, we used Python to reconstruct student behaviors from clickstreams in form of time-interval sequences. To enable queries and information extraction tasks, we fetched all the JSON data into MongoDB. We leverage k-means clustering to discretize time intervals and to obtain the time-embedded sequences. The time-embedded n-grams are acquired at the end of the first stage.

In the second stage, we step by step to answer the RQs. Descriptive statistics and exploratory data analysis (EDA) are employed to elucidate the associations between the patterns and the academic performance of learners. We answer RQ1 by constructing machine learning models to predict learner success. To answer the RQ2, we employ SHAP [35] to interpret the model and elucidate the effects of learning behavioral patterns on the final grades. Finally, we take hierarchical clustering analysis on the data to segment learners into groups based on their learning behaviors to answer the RQ3.

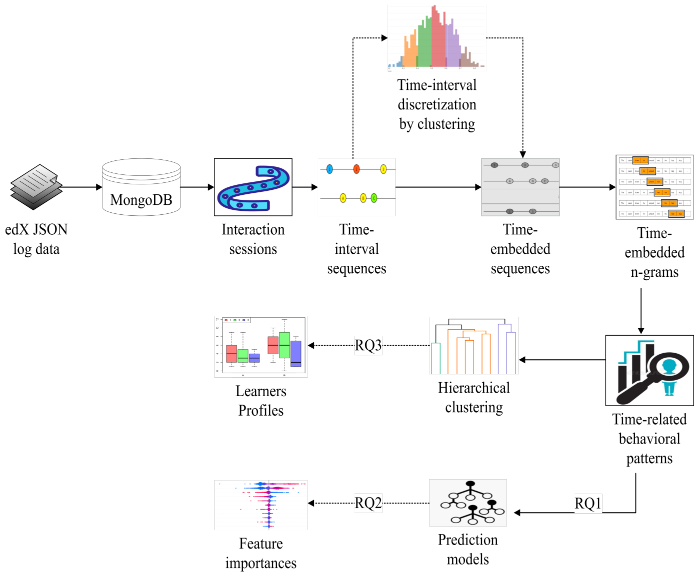


Fig. 1. Research procedure of this study.

4. EXPERIMENTAL RESULTS

4.1 The Course and Data

In this study, we encompassed a Python programming course offered by Feng Chia University on OpenEdu [36] in Spring 2020. The course was taught in Chinese with an expected duration of 16 weeks. Course contents were organized into several modules in which each module is composed of 2 to 6 lecture videos and attached in-video quizzes. Assessments were placed at the end of each module as self-tests. We used historical interactions in both in-video quizzes and self-tests in our analysis. Students who achieve over 60% in the final grade can pass the courses. There were 1201 learners enrolled in the course, however, only 591 learners had at least one interaction with the courses. At the end of the course, 71 learners passed the course (12%). The set of 591 learners leave 158,998 interactions in the log database. The distribution of all time-intervals is illustrated in Fig. 2 (frequency axis appears in log scale; time-intervals which are longer than 5 minutes were excluded).

As discussed in Section 3.2, we aim to segment the time intervals into five ranges called *Very-Short(1), Short(2), Medium(3), Long (4), and Very-Long (5)*. Since the lengths of lecture videos in the course are between 3 and 13 minutes, and needed time to answer an assessment question is often less than one minute, we considered all time-intervals longer than 5 minutes are certainly very long and thus we excluded them in the discretization procedure. We perform 1D k-means clustering on the population of time-intervals. By specifying $k = 5$, we obtained five segments: *(Very-short: [0,13], Short: [14,50], Medium: [51,109], Long: [110,194], Very-long: [195, 300])*. The distribution of five discrete values of time-intervals is also shown in Fig. 2.

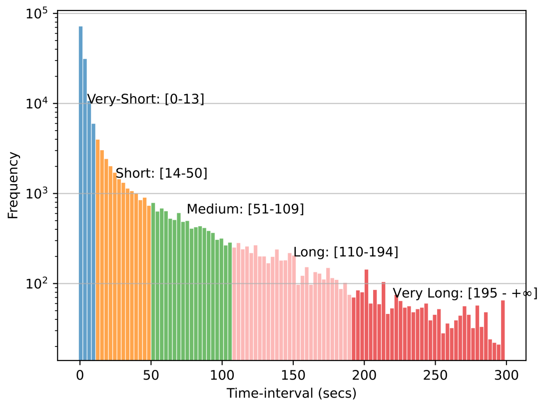


Fig. 2. Distributions of time-intervals and five segments of time-intervals produced by *k*-means.

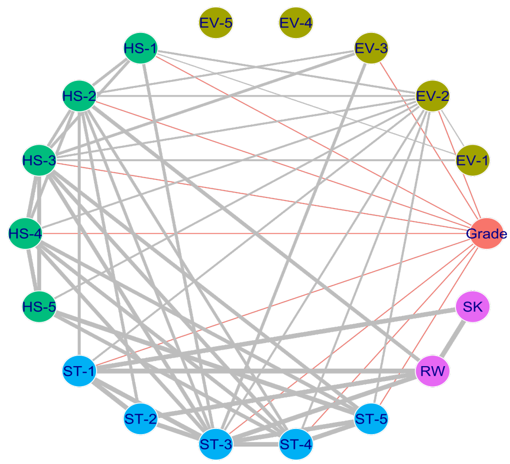


Fig. 3. Network of strongly correlated patterns.

We limited the number of patterns by specifying the minsup = 10%. 219 time-embedded *n*-grams, corresponding to 17 time-related behavioral patterns (as listed in Table 4), are obtained from the extraction phase. Consequently, a learner record is denoted by a vector containing frequencies of 17 behavioral patterns in which that learner is involved. We employed Exploration Data Analysis (EDA – a statistical approach that gain insights into the nature of the data, *e.g.*, distributions, statistical descriptions, and associations among attributes) featured by correlation matrix and correlation network to elucidate the association between behavioral patterns and their effects on the final grade. Computed Pearson coefficients told us that all behavioral patterns significantly and positively correlated with the final grade. We extracted those features whose correlations exceed 0.6 and construct the correlation networks in Fig. 3. It can interpret that academic performance associate with engagement levels. However, only intended learning patterns are strongly positively correlated with performance.

Regarding Evaluation behaviors, *EV-2* and *EV-3* are indicators of effectively continuous efforts, in which the short and medium time intervals denote appropriate durations of problem-solving. On the other hand, because of their low frequencies, EV-4 and EV-5 do not have any association with the learning result. Besides, *Help-seeking* patterns also demonstrate intended learning behaviors featured by a series of switching efforts between lecture videos and assessments. So that, all *Help-seeking* patterns are strongly correlated to performance, in which, short and medium intervals are most appropriate durations, again. With Study patterns, longer duration of video watching *(ST-3, ST-4, and ST-5)* is the implication of better performance. The *ST-1* pattern is not considered since very short video watching can result from repeated seeking interaction *(Rw and Sk)*.

4.2 Early Prediction of Learning Performance

We employ two well-known ensemble methods which are Random Forest (RF) [37] and eXtreme Gradient Boosting (XGBoost) [38] as learning algorithms. The two algorithms were selected because of their high performance as well as their ability to ignore weak predictors during the training process. With a high number of input features, the built-in feature selection of ensemble methods based on decision trees can be useful in eliminating unimportant features to boosting learning performance. The early prediction models were built based on a weekly basis. Data for the $n-th$ week is extracted from the beginning to week n. We performed two strategies of data extraction to form the datasets for early prediction: *i) Accumulated data extraction:* The features were extracted based on the accumulated interactions of learners from the course beginning to the current week. *ii) Mixed data extraction:* Since the former strategy only captures accumulated behaviors of learners as a single snapshot, weekly increments are not considered. As temporal behaviors on a weekly basis have demonstrated good improvement in prediction performance in the literature [10], we augmented the accumulated dataset with snapshots of weekly data. The resulted data of the second strategy is called mixed dataset. The experiments were implemented using Python with the libraries *scikit-learn* and *xgboost*. In each experiment, the early prediction in each week is performed on two datasets (accumulated and mixed) with two selected learning algorithms. We used random sampling to split the data into train set (67%) and test set (33%). Each experiment is repeated 10 times to generalize the evaluation results. In the prediction of binary learning result (Pass/Fail), we used the combination of Accuracy *(Acc)*, Area Under the Curve *(AUC)*, Precision *(Pre)*, Recall *(Rec)*, and *F1* to evaluate the prediction performance. Since the data is dramatically imbalanced, we employed the metrics derived from the confusion matrix *(Pre, Rec, and F1)* to obtain a better view of model performance. In this prediction, we target to predict if a learner "passes" the course, so the metrics were computed on the minority class of "pass learners". Table 5 shows the evaluation results of *Pass/Fail* prediction using the whole set of learner interactions. The early prediction performance for students' *Pass/Fail* is depicted in Fig. 4.

Results show that student success in the course can be predicted with a good *AUC* (greater than 0.9) from week 5. *AUC* values increase dramatically from week 1 to week 3 and slowly increase in later weeks and reach 0.95 around the middle week of the course. Best prediction results (*AUC* greater than 0.94) can be found after the middle point of time of the course (week 8). Mixed data can help in improving AUC in most cases. However, its contributions are not clear in improving other metrics. Concerning the algorithms,

Table 5. Evaluation results for Pass/Fail prediction.

Method	AUC	Acc	F1	Pre	Rec
Mixed + RF	**0.950**	**0.940**	**0.712**	**0.831**	**0.631**
Mixed + XGBoost	0.941	0.917	0.626	0.672	0.594
Accumulated + RF	0.940	0.939	0.709	0.814	0.629
Accumulated + XGBoost	0.926	0.917	0.619	0.674	0.583

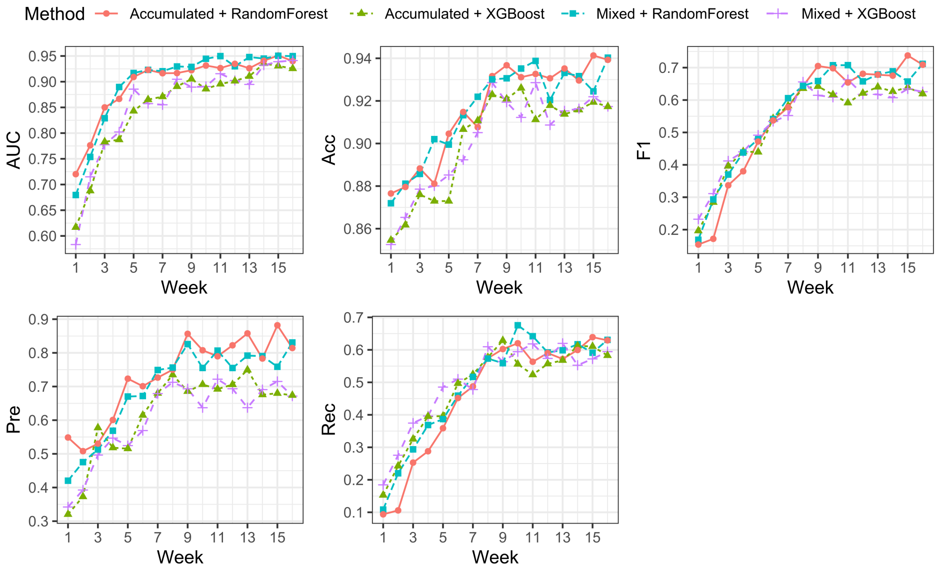


Fig. 4. Evaluation results of early Pass/Fail prediction.

RF outperformed *XGBoost* in all cases. Our results on *AUC* are comparable to what was reported by Moreno-Marcos *et al*. [3] in their dropout prediction. Using multiple col- lections of counting based features and SRL strategies (variables related to learners intentions, activity, interactions with videos and exercises, self-report SRL strategies, and event-based SRL strategies), they recommended that the time needed to achieve *AUC* of 0.9 is between 43% and 67% of the total theoretical MOOC duration.

In final grade prediction, we evaluate the prediction performance using the coefficient of determination *R2*, root mean squared errors *(RMSE)*, mean square error *(MSE)*, and mean absolute error *(MAE)*. The evaluation results using the whole set of learner interactions are shown in Table 6. The evaluation of early prediction is depicted in Fig. 5. RF with mixed data was the winner among all settings. The contributions of mixed data in reducing error values are not clear. The acceptable prediction results (R2 above 0.6 and RMSE below 0.2) can be obtained from week 5 and the best results can be found after week 7. After week 5, the prediction performance was slightly improved. Our prediction results are also comparable with prior studies [39–41], in which RMSE values varied in the range of 0.1 to 0.5, depending on the data, feature engineering, algorithms, and how early the predictions were made.

Table 6. Evaluation results for final grade prediction.

Method	R2	RMSE	MSE	MAE
Mixed + RF	**0.705**	**0.155**	**0.024**	**0.094**
Mixed + XGBoost	0.629	0.173	0.030	0.106
Accumulated + RF	0.675	0.159	0.025	0.096
Accumulated + XGBoost	0.636	0.168	0.028	0.102

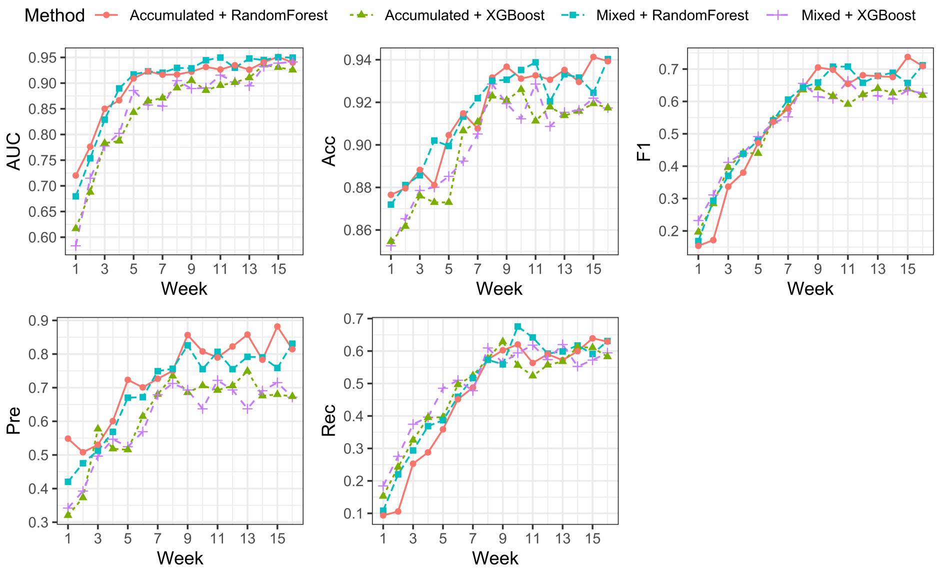


Fig. 5. Evaluation results of early prediction of final grade.

Overall, the experiments on the predictive performance of time-related behavioral patterns can respond to the *RQ1*. The learning performance of learners can be predicted effectively from the 33% theoretical duration of the MOOC (with *AUC* above 0.9 for *Pass/Fail*, *R2* above 0.6, and *RMSE* below 0.2 for final grade). The most important remark is that our models are easier to be built since the learners' features are only extracted from the clickstreams, without using any counting-based features or self-reports. In addition, as confusion-matrix-based metrics were rarely available in prior studies, the results reported in this article (regarding *Pre, Rec,* and *F1*) can be viewed as a baseline for the specificity and the sensitivity of prediction models for student success in MOOCs.

4.3 Feature Importance of Time-Related Behavior Patterns in Early Prediction

After obtaining fruitful prediction results for students' performance, we need to know the behavioral patterns affecting the learning performance which are the responses for *RQ2*. To address the question, we employed SHAP [35] as means of feature importance explanation. We performed SHAP analysis at two points of time to get a better understand of the effects of learning behaviors on learning outcomes over time. As the prediction model for the final grade reached an acceptable result at week 5, we selected it as the first inspected point. The second point of time is the final week of the course (week 16).

The SHAP plots in Fig. 6 show that the triple of *HS-3, EV-2, ST-1* are the most impact patterns on learning outcomes both in the early stages (week 5) and the final stage (week 16) of the course. Notably, in the early stage, help-seeking with medium durations *(HS-3)* seems to be the most effective; the second was evaluation with short duration *(EV-2)*; the study patterns *(ST-1)* were less important and ranked third. However, at the final stage, the *ST-1* became the most important pattern; the HS-3 became less important and ranked third; *EV-2* keeps its rank as the second. The position swap between *HS-3* and *ST-1* in the two SHAP plots implies that, in the early stage, *rewatching lectures videos is a good strategy* if a student did not prepare well for his/her quizzes. This strategy can even positively affect the final grades. However, in the whole course, engagements in studying and *self-evaluation without help-seeking* seem to be the most important. Short and very short (less than 1 minute) are most common in important patterns related to study and evaluation. Medium durations are suitable for help-seeking. Long durations (longer than 2 minutes) in the evaluation strategy are less meaningful. These findings raise new concerns for courses' instructors when specifying time-setting for assessment activities.

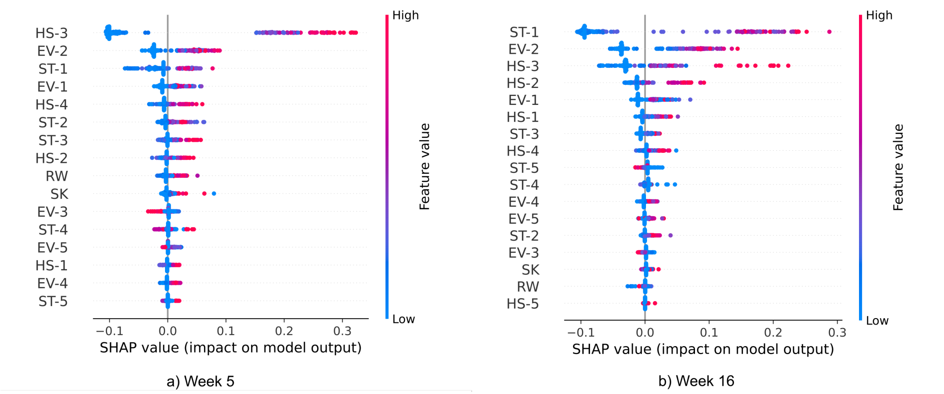


Fig. 6. Feature importance derived from final grade prediction models.

4.4 Detection of Learner Profiles based on Time-Related Behavioral Patterns

The time-related behavioral patterns not only contribute as robust predictors for learning performance using supervised learning but also can be used as meaningful features to segment students into clusters of similar students using unsupervised learning. The prototype of this task was first proposed by Maldonado-Mahauad *et al*. in [8]. In that study, they use process mining to detect six common patterns on macro-behaviors of videos and assignments, and mapped them into six *SRL strategies*; then the strategies were used to detect three *SRL learning profiles: Sampling learners, Comprehensive learners, and Targeting learners*. We repeat that procedure to determine if the same set of learner profiles can be found in our feature space and interpret the learner profiles through the lens of time-related patterns.

The agglomerative hierarchical clustering was selected as the detection algorithm. We employed the complete-link clustering to generate the dendrogram and specified the cut point to obtain three clusters. As our findings are consistent with what was discovered in prior studies [7, 8], we leverage the same set of cluster names to describe the learner profiles. With three clusters obtained, we performed cluster profile analysis to compare the difference between learner profiles. The boxplot depicted in Fig. 7 illustrates the distribution of behavioral patterns of three learner profiles. Note that the differences between clusters in our study are significant, which differ from the findings of Maldonado-Mahauad *et al.* [8], in which they did not observe the statistically significant differences between *Comprehensive learners* and *Targeting learners*. The SRL learner profiles can be described based on time-related behavioral patterns as follows:

- **Cluster 1 – Sampling learners (N=547):** This cluster includes learners with the least engagement compared with the remaining groups. They tried to skim on some lecture videos in short durations and attempt to solve a small number of assessments. They only interacted with the system at the beginning of the course to make some "sampling" actions on the learning materials. As a result, most of them are non-completers and received very low final grades.
- **Cluster 2 – Comprehensive learners (N=33):** This cluster includes those learners who tend to follow the path given by the course design to achieve the learning objectives [8]. Their level of interaction on lecture videos and the duration of sessions (on average) is higher than those of Targeting learners. Their participation in assessment is significantly lower than Targeting learners. Because they tend to obey the rules, they paid more effort in assessment without help-seeking, so they have higher frequencies of evaluation engagement. Since their learning performance is not as high as that of Targeting learners, they need more time to watch lecture videos when they encounter difficult questions, so their frequencies of long help-seeking are higher. In video-watching, they have significantly higher frequencies of long and very-long video-watching and rarely fire-seeking actions in videos. With these behaviors, most of them are completers with final grades slightly lower than those of Targeting Learners.
- **Cluster 3 – Targeting learners (N=11):** This cluster includes those learners who have specific targets when engaging with the course. They have the highest level of engagement in comparison to their counterparts. In assessment, they paid less effort than Comprehensive learners in solving the problems without reviewing lecture videos, but they seem to use the help-seeking strategy effectively. They have very high frequencies of short help-seeking *(HS-2)*, but low frequencies of long help-seeking *(HS-4)*. They can find out the solutions for the questions in shorter durations than that the Comprehensive learners do. Their targeting behaviors are also shown in video-watching behaviors in which they prefer watching lecture videos in short or medium durations and frequently skip or rewatch specific video contents. With this pragmatic strategy, all of them are completers with high academic performance.

The above analysis responds to the RQ3 by confirming that the time-related behavioral patterns found by using time-embedded *n*-grams at the micro-level of MOOC interactions can be leveraged to obtain the three learner profiles, like what SRL strategies on macro-level did [8]. The results extended the additional insights about time intervals which can be viewed as the learning velocity of different types of learners.

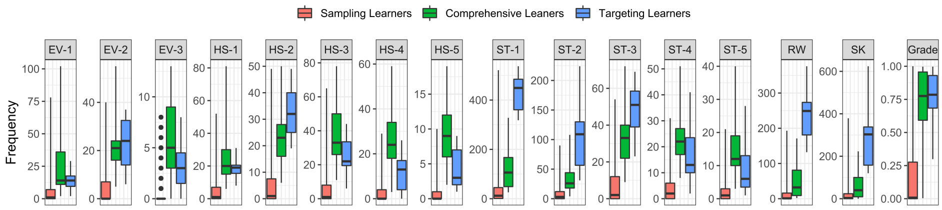


Fig. 7. Boxplots indicating time-related behavioral patterns of three learner profiles (EV-4 and EV-5 were not included).

5. DISCUSSIONS AND CONCLUSIONS

5.1 Practical Implications

The findings of this research promote improvements to the learning environment in multiple ways. Time-related behavioral patterns can be embedded into the platform as feedback for engagement tracking. For example, the idea of "referenced students" [42] can be utilized to build a personalized feedback widget indicating the immediate engagement level of learners with time-related patterns, concerning learner profiles detected in *RQ3*. Through these individualized feedbacks, the educators can diagnose struggling situations (*i.e.* very long durations of help-seeking, *etc*.) to offer precise interventions based on the learner profile that the learner belongs to; the learners can better understand their pros and cons by comparing to referenced profiles (Fig. 8). In return, the tailored prompts and targeted feedbacks would result in fostering motivation as well as reducing the dropout rate [43]. In the same way, the dropout prediction model can be early established using time-embedded patterns, like grade prediction in *RQ1*.

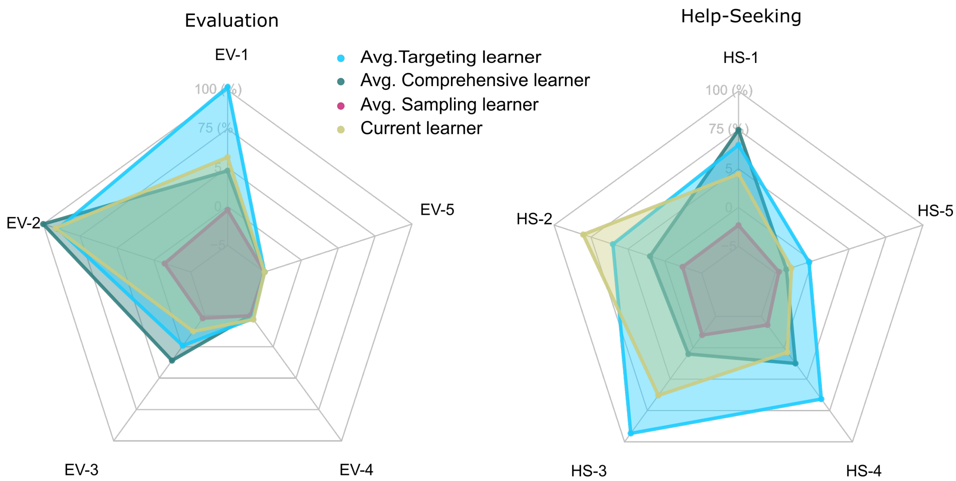


Fig. 8. Example of feedback widget regarding time-related patterns using referenced profiles.

With these predictions, potential dropout learners might receive timely interventions from the precision education platform to assist them in completing the study. Finally, time-related patterns can highlight challenge sections requiring a high cognitive load in the lectures or assessments. These insights can contribute to optimizing the lectures and course structure to reduce the negative impact of insufficient designs within the courses.

5.2 Pedagogical Implications

We argue that time intervals embedded in micro behaviors, especially in assessment, can earlier announce the competence of learners than summative assessments. They are in line with the term "time-on-task" which was recommended to be a predictor for formative assessment outcomes in blended learning (Tempelaar *et al*., 2015). In this manner, temporal insights from fine-grained levels would be shed light on the immediate performance of learners from which precise feedbacks or interventions can be offered. The strong impact of help-seeking strategies on academic performance discovered in *RQ2* is consistent with the conclusions of Corrin, De Barba, & Bakharia [44], Sun, Wu, & Lee [45], and Sands & Yadav [46]. We add to this statement two additional comments that *i)* only help-seeking interactions in relatively short intervals (less than 2 minutes) have strong effects on learners' grades; and *ii)* the shorter help-seeking interactions are, the higher grades the learners can gain. Note that the advancements of help-seeking are visible thanks to the in-video-quizzes, so that, we emphasize the importance of embedding precision education solutions to assessment activities, besides regular *SRL prompts* which are commonly provided in lecture videos.

5.3 Limitations

Despite its findings, there are a couple of problems that should be considered as the continuing of the current study. First, since the *n*-grams cannot capture patterns with gaps, *n*-gram with gaps can be considered to increase the generality of the patterns. Second, the discretization scheme for time intervals can be varied or fine-tuned depending on the research context. Since hard-clustering causes the problem of sharpening cuts in the time dimension, a fuzzy-based approach can be applied to soften the boundary of time discretized values. Third, due to the nature of the selected courses, only video, and assessment-related interactions were considered, in general cases, a more varied set of behavioral patterns could be discovered which may contain forums or textbooks related interactions. Fourth, the mapping scheme of time-embedded *n*-grams to learning behaviors can be justified by considering studies about time constraints in SRL. We leave all these problems for future works.

In summary, the contributions of this research to the body of knowledge can be summarized as follows: *i)* A proposal for *time-embedded n-gram* and its mining procedure from MOOCs clickstreams. *ii)* Identification of *time-related micro-behavioral patterns* that learners exhibit in MOOCs. *iii)* Identification of three clusters of learners based on their time-related behaviors: Sampling learners, Comprehensive Learners, and Targeting learners. *iv)* Elucidating the association between time-related learning behaviors with academic achievement: help-seeking and evaluation activities that are performed in less than two minutes have positive effects on the final grade. *v)* Come up with a potential application of time-embedded behavioral patterns in early prediction of learning performance.

Finally, we emphasize the promising directions for future research using time-related *n*-grams in MOOCs. By embedding time intervals into interaction sequences of learners, more meaningful behavioral patterns can be obtained. So that the instructors can gain more hints to build up tailored interventions for learners and achieve precision education targets.

ACKNOWLEDGMENT

Thanks to the OpenEdu team who developed the MOOCs system and provided the de-identification log data of learners' interaction events.

REFERENCES

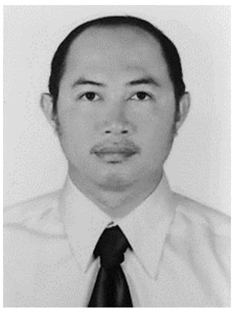
1. S. Patra, "The 250 most popular online courses of all time," https://www.classcentral.com/report/most-popular-online-courses/, 2021.
2. J. Gardner and C. Brooks, "Student success prediction in MOOCs," *User Modeling and User-Adapted Interaction*, Vol. 28, 2018, pp. 127-203.
3. P. M. Moreno-Marcos, P. J. Muñoz-Merino, J. Maldonado-Mahauad, M. Pérez-Sanagustín, C. Alario-Hoyos, and C. Delgado Kloos, "Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs," *Computers and Education*, Vol. 145, 2020, No. 103728.
4. O. H. T. Lu, J. C. H. Huang, A. Y. Q. Huang, and S. J. H. Yang, "Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course," *Interactive Learning Environments*, Vol. 25, 2017, pp. 220-234.
5. C. G. Brinton and M. Chiang, "MOOC performance prediction via clickstream data and social learning networks," in *Proceedings of IEEE InfoCom*, Vol. 26, 2015, pp. 2299-2307.
6. E. Rabin, V. Silber-Varod, Y. M. Kalman, and M. Kalz, "Identifying Learning Activity Sequences that Are Associated with High Intention-Fulfillment in MOOCs," in *Lecture Notes in Computer Science*, Vol. 11722, 2019, pp. 224-235.
7. J. Wong, M. Khalil, M. Baars, B. B. de Koning, and F. Paas, "Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course," *Computers and Education*, Vol. 140, 2019, p. 103595.
8. J. Maldonado-Mahauad, M. Pérez-Sanagustín, R. F. Kizilcec, N. Morales, and J. Munoz-Gama, "Mining theory-based patterns from big data: Identifying self-regulated learning strategies in Massive Open Online Courses," *Computers in Human Behavior*, Vol. 80, 2018, pp. 179-196.
9. T. Sinha, P. Jermann, N. Li, and P. Dillenbourg, "Your click decides your fate: Inferring information processing and attrition behavior from MOOC video clickstream interactions," in *Proceedings of Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 3-14.
10. Z. Liu, F. Xiong, K. Zou, H. Wang, Z. Liu, and H. Wang, "Predicting learning satus in MOOCs using LSTM," in *Proceedings of ACM Turing Celebration Conference*, 2018.

11. R. Conijn, A. van den Beemt, and P. Cuijpers, "Predicting student performance in a blended MOOC," *Journal of Computer Assisted Learning*, Vol. 34, 2018, pp. 615-628.
12. C. G. Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, "Mining MOOC click-streams: Video-watching behavior vs. in-video quiz performance," *IEEE Transactions on Signal Processing*, Vol. 64, 2016, pp. 3677-3692.
13. C. Brooks, C. Thompson, and S. Teasley, "A time series interaction analysis method for building predictive models of learners using log data," in *ACM International Conference Proceeding Series*, 2015, pp. 126-135.
14. C.-H. Yu, J. Wu, and A.-C. Liu, "Predicting learning outcomes with MOOC click-streams," *Education Sciences*, Vol. 9, 2019, p. 104.
15. B. Jeon and N. Park, "Dropout prediction over weeks in MOOCs by learning representations of clicks and videos," *arXiv Preprint*, 2020, arXiv:2002.01955.
16. J. A. Ruipérez-Valiente, R. Cobos, P. J. Muñoz-Merino, Á. Andujar, and C. Delgado Kloos, "Early Prediction and Variable Importance of Certificate Accomplishment in a MOOC," in *Digital Education: Out to the World and Back to the Campus*, C. Delgado Kloos, P. Jermann, M. Pérez-Sanagustín, D. T. Seaton, and S. White, (eds.), Springer, Cham, 2017, pp. 263-272.
17. S. Crossley, D. S. Mcnamara, L. Paquette, R. S. Baker, and M. Dascalu, "Combining click-Stream data with NLP tools to better understand MOOC completion," in *ACM International Conference Proceeding Series*, 2016, pp. 6-14.
18. P. M. Moreno-Marcos, T. C. Pong, P. J. Munoz-Merino, and C. D. Kloos, "Analysis of the factors influencing learners' performance prediction with learning analytics," *IEEE Access*, Vol. 8, 2020, pp. 5264-5282.
19. G. Allione and R. M. Stein, "Mass attrition: An analysis of drop out from principles of microeconomics MOOC," *Journal of Economic Education*, Vol. 47, 2016, pp. 174-186.
20. G. Balakrishnan and D. Coetzee, "Predicting student retention in massive open online courses using hidden markov models," *Electrical Engineering and Computer Sciences University of California at Berkeley*, Vol. 53, 2013, pp. 57-58.
21. C. Geigle and C. Zhai, "Modeling MOOC student behavior with two-layer hidden Markov models," *Journal of Educational Data Mining*, Vol. 9, 2017, pp. 1-24.
22. A. Ramesh, D. Goldwasser, B. Huang, H. Daume, and L. Getoor, "Interpretable engagement models for MOOCs using hinge-loss Markov random fields," *IEEE Transactions on Learning Technologies*, Vol. 13, 2020, pp. 107-122.
23. M. Fei and D. Y. Yeung, "Temporal models for predicting student dropout in massive open online courses," in *Proceedings of the 15th IEEE International Conference on Data Mining Workshop*, 2016, pp. 256-263.
24. M. J. Zaki, "Sequence mining in categorical domains: Incorporating constraints," in *Proceedings of the 9th International Conference on Information and Knowledge Management*, Vol. –, 2000, pp. – –.
25. X. Li, T. Wang, and H. Wang, "Exploring N-gram features in clickstream data for MOOC learning achievement prediction," in *Lecture Notes in Computer Science*, Vol. 10179, 2017, pp. 328-339.
26. A. van den Beemt, J. Buijs, and W. van der Aalst, "Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on pro-

cess mining and clustering," *The International Review of Research in Open and Distributed Learning*, Vol. 19, 2018.
27. M. Romero and M. Usart, "The time factor in MOOCS: Time-on-task, interaction temporal patterns, and time perspectives in a MOOC," in *Proceedings of the 6th International Conference on Computer Supported Education*, Vol. 1, 2014, pp. 53-62.
28. U. M. Krause, R. Stark, and H. Mandl, "The effects of cooperative learning and feedback on e-learning in statistics," *Learning and Instruction*, Vol. 19, 2009, pp. 158-170.
29. F. Goldhammer, J. Naumann, A. Stelter, K. Tóth, H. Rölke, and E. Klieme, "The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment," *Journal of Educational Psychology*, Vol. 106, 2014, pp. 608-626.
30. Y. Lee, "Effect of uninterrupted time-on-task on students' success in Massive Open Online Courses (MOOCs)," *Computers in Human Behavior*, Vol. 86, 2018, pp. 174-180.
31. M. Khalil and J. Wong, "The essence of time: Taking a look at learning sessions in MOOCs," in *Proceedings of International Conference on Learning With MOOCS*, 2018, pp. 131-133.
32. Y. L. Chen and T. C. K. Huang, "Discovering fuzzy time-interval sequential patterns in sequence databases," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 35, 2005, pp. 959-972.
33. J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, 2011.
34. M. Damashek, "Gauging similarity with n-grams: Language-independent categorization of text," *Science*, Vol. 267, 1995, pp. 843-848.
35. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, 2017, pp. 4765-4774.
36. "OpenEdu," https://www.openedu.tw/, 2014.
37. T. K. Ho, "Random decision forests," in *Proceedings of the 3rd IEEE International Conference on Document Analysis and Recognition*, Vol. 1, 1995, pp. 278-282.
38. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.
39. V. Almeda, J. Zuech, C. Utz, G. Higgins, R. Reynolds, and R. S. Baker, "Comparing the factors that predict completion and grades among for-credit and open/MOOC students in online learning," *Online Learning*, Vol. 22, 2018.
40. X. Li, L. Xie, and H. Wang, "Grade prediction in MOOCs," in *Proceedings of IEEE International Conference on Computational Science and Engineering*, 2016, pp. 386-392.
41. O. H. T. Lu, A. Y. Q. Huang, J. C. H. Huang, A. J. Q. Lin, H. Ogata, and S. J. H. Yang, "Applying learning analytics for the early prediction of students' academic performance in blended learning," *Journal of Educational Technology & Society*, Vol. 21, 2018, pp. 1176-3647.
42. L. De-La-Fuente-Valentín, A. Pardo, F. L. Hernández, and D. Burgos, "A visual analytics method for score estimation in learning courses," *Journal of Universal Computer Science*, Vol. 21, 2015, pp. 134-155.

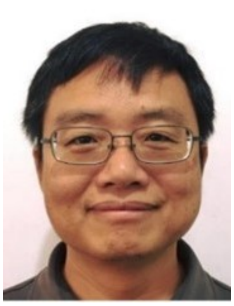
43. S. Halawa, D. Greene, and J. Mitchell, "Dropout prediction in MOOCs using learner activity features," in *Proceedings of the 2nd European MOOC Stakeholder Summit*, Vol. 37, 2014, pp. 58-65.
44. L. Corrin, P. G. De Barba, and A. Bakharia, "Using learning analytics to explore help-seeking learner profiles in MOOCs," in *ACM International Conference Proceeding Series*, Vol. 5, 2017, pp. 424-428.
45. J. C. Y. Sun, Y. T. Wu, and W. I. Lee, "The effect of the flipped classroom approach to OpenCourseWare instruction on students' self-regulation," *British Journal of Educational Technology*, Vol. 48, 2017, pp. 713-729.
46. P. Sands and A. Yadav, "Self-regulation for high school learners in a mooc computer science course," in *Proceedings of Annual Conference on Innovation and Technology in Computer Science Education*, 2020, pp. 845-851.

Hsi-Min Chen received the BS and Ph.D. degrees in Computer Science and Information Engineering from National Central University, Taiwan, in 2000 and 2010, respectively. He is currently an Associate Professor with the Department of Information Engineering and Computer Science, Feng Chia University, Taiwan. His research interests include software engineering, programming education, object-oriented technology, service computing, and distributed computing.



Bao-An Nguyen received BS in Hanoi University of Science and Technology in 2005 and the MS and the Ph.D. degrees in Information Engineering and Computer Science from Feng Chia University, Taiwan, in 2011 and 2021, respectively. He is currently a faculty of Department of Information Technology, Tra Vinh University, Viet Nam. His research interests include data mining, software engineering and education technology.

Chyi-Ren Dow received the BS and MS degrees in Information Engineering from National Chiao Tung University, Taiwan, in 1984 and 1988, respectively, and the MS and Ph.D. degrees in Computer Science from the University of Pittsburgh, PA, in 1992 and 1994, respectively. He is currently a Distinguished Professor with the Department of Information Engineering and Computer Science, Feng Chia University, Taiwan. His research interests include mobile computing, ad-hoc wireless networks, agent techniques, fault tolerance, and learning technology.



Nien-Lin Hsueh is a Professor in the Department of Information Engineering and Computer Science at Feng Chia University in Taiwan, and the Director of the Department of Information Engineering and Computer Science at FCU since 2018. He is currently the Chairman of Software Engineering Association Taiwan (SEAT; https://www.seat.org.tw) since 2017. His research interests include software engineering, object-oriented methodologies, software process improvement, and e-learning. He was the technology consultant in the CMMI-based process improvement project of Office of Information Technology at Feng Chia University. He received his Ph.D. in Computer Science from National Central University in Taiwan.

An-Chi Liu is the Chair Professor at the Department of Information Engineering and Computer Science of Feng Chia University, Taiwan. His research interests include network management, distributed system, and database system.