Revealing at-risk learning patterns and corresponding self-regulated strategies via LSTM encoder and time-series clustering

Mingyan Zhang and Xu Du
National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China
Kerry Rice and Jui-Long Hung
Department of Educational Technology, College of Education, Boise State University, Boise, Idaho, USA, and
Hao Li
National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China

Abstract

Purpose – This study aims to propose a learning pattern analysis method which can improve a predictive model's performance, as well as discover hidden insights into micro-level learning pattern. Analyzing student's learning patterns can help instructors understand how their course design or activities shape learning behaviors; depict students' beliefs about learning and their motivation; and predict learning performance by analyzing individual students' learning patterns. Although time-series analysis is one of the most feasible predictive methods for learning pattern analysis, literature-indicated current approaches cannot provide holistic insights about learning patterns for personalized intervention. This study identified at-risk students by micro-level learning pattern analysis and detected pattern types, especially at-risk patterns that existed in the case study. The connections among students' learning patterns, corresponding self-regulated learning (SRL) strategies and learning performance were finally revealed.

Design/methodology/approach – The method used long short-term memory (LSTM)-encoder to process micro-level behavioral patterns for feature extraction and compression, thus the students' behavior pattern information were saved into encoded series. The encoded time-series data were then used for pattern analysis and performance prediction. Time series clustering were performed to interpret the unique strength of proposed method.

Findings – Successful students showed consistent participation levels and balanced behavioral frequency distributions. The successful students also adjusted learning behaviors to meet with course requirements accordingly. The three at-risk patten types showed the low-engagement (R1) the low-interaction (R2) and the non-persistent characteristics (R3). Successful students showed more complete SRL strategies than failed students. Political Science had higher at-risk chances in all three at-risk types. Computer Science, Earth Science and Economics showed higher chances of having R3 students.

Research limitations/implications – The study identified multiple learning patterns which can lead to the at-risk situation. However, more studies are needed to validate whether the same at-risk types can be found in other educational settings. In addition, this case study found the distributions of at-risk types were vary in different subjects. The relationship between subjects and at-risk types is worth further investigation.

Originality/value – This study found the proposed method can effectively extract micro-level behavioral information to generate better prediction outcomes and depict student's SRL learning strategies in online learning. The authors confirm that the research in their work is original, and that all the data given in the paper are real and authentic. The study has not been submitted to peer review and not has been accepted for publishing in another journal.

Keywords Learning pattern analysis, Early warning in online education, Learning performance prediction, Long short-term memory encoder

Paper type Research paper

1. Introduction

Online learners need to determine when and how to engage with course content. For online learning learners, regulating one's learning process is a critical skill to achieve personal learning objectives. Self-regulated learning (SRL) refers to a learner's learning strategy to "monitors, directs, and regulates actions toward goals of information acquisition, expanding expertise, and self-improvement" (Li, 2019). Self-regulated learners are characterized by their ability to initiate cognitive, metacognitive, affective and motivational processes (Boekaerts, 1997; Lehmann *et al.*, 2014). Literature indicates that the ability to effectively use SRL learning strategies is an essential skill to succeed in online learning. When viewed as a process, SRL can be conceptualized as a series of events or actions

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/2398-6247.htm**

Information Discovery and Delivery
50/2 (2022) 206–216

[DOI 10.1108/IDD-12-2020-0160]

This work was supported by the National Natural Science Foundation of China under Grant 61807013 and 61877027.

Received 27 December 2020
Revised 4 March 2021
6 April 2021
Accepted 7 April 2021

generated during the learning process (Bannert *et al.*, 2014). In observable format, the process is called as learning pattern.

Vermunt and Donche (2017) noted that "learning pattern" can be "conceptualized as a coherent whole of learning activities that learners usually employ, their beliefs about learning and their learning motivation, a whole that is characteristic of them in a certain period of time." As a student's learning strategies are reflected by his/her learning patterns, analyzing learning patterns:

- can help instructors understand how their course design or activities shape students' learning behaviors;
- can depict individual students' learning strategies; and
- can predict learning performance by analyzing individual students' learning patterns.

Through the understanding of students' learning patterns, learning strategies and key performance predictors, instructors will be able to provide personalized instructions to individuals (Fratamico *et al.*, 2017).

Learning patterns can be considered as a type of time-series data. For predictive research in learning patterns, time-series analysis is one of the most popular predictive methods (Brooks *et al.*, 2015). However, the literature shows that predictive research in learning patterns still has some noted research gaps. Although various learning behaviors can be aggregated weekly or bi-weekly, total frequency is still the most popular variable to represent a student's learning patterns in a learning management system (LMS; Yang *et al.*, 2018). However, Hung *et al.* (2015) explained that total frequency could not result in satisfactory prediction accuracy. For example, two students might have identical aggregated total frequencies per week with totally different patterns. One student might evenly allocate his/her learning time across weekdays, but the other student might spend a longer time studying only on the weekends. A predictive model constructed with aggregated frequencies cannot recognize their pattern differences and would predict that these two students would have the same performance level. These limitations might result in studies identifying one type of at-risk students only: those who showed low participation in course activities (Fratamico *et al.*, 2017). Yang *et al.* (2018) found some exceptional cases, such as students who exhibited low participation but high performance or high participation but low performance. Therefore, relying only on students' total frequency patterns may not provide enough insight into the variety of learning patterns. Micro-level pattern analyses may provide an approach that further improves analysis results. Some attempts can be found in descriptive research (Gamulin *et al.*, 2016) or predictive research (Brooks *et al.*, 2015) on pattern analyses. However, these studies observed patterns of individual behaviors separately, without examining behaviors holistically.

The development of deep learning algorithms has opened new possibilities to use more advanced neural network infrastructures (Pouyanfar *et al.*, 2018) for better analysis outcomes. Across all deep learning structures, autoencoder has attracted attention for its applications in feature extraction and information compression. Autoencoder is an unsupervised deep learning model that can perform dimension compression and can then restore the information from the input data and remove the noise from the input data at the same time (Masci *et al.*, 2011). In time-series analysis, Recurrent Neural Networks, with their capability for both short-term and long-term memory, have shown impressive progress in various research areas (Cui *et al.*, 2020). The so-called LSTM method can analyze time-series patterns and then generate a future pattern prediction (Ma *et al.*, 2015) or a categorical target prediction (Gamboa, 2017).

Therefore, this study proposes a customized LSTM-autoencoder for student's learning pattern analyses. We assume that this approach can improve a predictive model's performance and discover hidden insights about the connection among a student's learning patterns, his or her learning strategies and performance. We are especially interested in learning patterns which can lead to at-risk situations. A case study is conducted to validate the model's effectiveness, discover hidden knowledge and develop corresponding interventions for different types of at-risk students. This study includes the following key steps:

- Apply the LSTM-autoencoder to process micro-level behavioral patterns for feature extraction and compression. Then use the encoded time-series as inputs for performance prediction. The goal is to test whether the proposed method can generate more accurate at-risk predictions than baseline approaches.
- If the proposed method is better, follow up analyses are conducted to identify strengths of the proposed method in pattern analysis.
- Finally, a series of analysis focus on extracted knowledge and connect them with SRL strategies.

Q1. Do the proposed methods result in better prediction results?

Q2. What are the unique strengths of the proposed methods in the micro-level learning pattern analysis?

Q3. What hidden knowledge and implications can be extracted via micro-level learning pattern analysis with the proposed methods?

The structure of this paper is as follows:

- the related work section reviews research efforts in SRL and learning pattern analysis, time-series analysis in education and autoencoder in education;
- the methodology section describes key steps of proposed method;
- the case study section presents the process of model training, validation and results;
- the discussion section discusses key findings and connects them with SRL strategies to extract implications; and
- the conclusion section summarizes the unique contributions of this study.

2. Related work

2.1 Self-regulated learning and learning pattern analysis

SRL represents a learner's ability to spend time and energy in the most productive ways, to obtain a more effective and rewarding learning experience (Pintrich, 2004). The ability is even more important in online learning environments. As the instructor's presence is low, learners must make decisions regarding when to study or how to approach the study materials. The framework of SRL is a core conceptual framework to understand the cognitive, motivational and emotional aspects of learning (Viberg *et al.*, 2020). SRL processes are seen as "the processes whereby students activate and sustain cognitions, affects, and behaviours that are systematically oriented toward the attainment of personal goals" (Zimmerman and Schunk, 2011).

Zimmerman and Moylan's Cyclical phases model (Zimmerman and Moylan, 2009) was the most popular SRL model. The framework contains three phases – forethought, performance and self-reflection.

- The forethought phase replies to students to analyze the task, to set goals, to plan how to reach them. The process involves several motivational beliefs and influences the activation of learning strategies.
- In the performance phase, the students execute the task, while they monitor how they are progressing, and use several self-control strategies to keep themselves cognitively engaged and motivated to finish the task.
- In the self-reflection phase, students assess how they have performed the task, making attributions about their success or failure.

These attributions generate self-reactions that can positively or negatively influence how the students approach the task in later performances.

As students can select and adjust their learning strategies according to the requirements of the learning context (Du *et al.*, 2019), studies have demonstrated a positive relationship between the use of SRL strategies in online environments and academic achievement (Broadbent, 2017). In other words, a student's SRL capability is the critical success factor in online courses. However, Viberg *et al.* (2020) found many students possess poor SRL practices. Students might fail to accurately calibrate their learning processes and then result in at-risk status. As a student's learning strategy can be observed via online behaviors performed in the online course, research efforts have been invested in understanding the relationship between SRL strategies (or online behaviors) and performance (Broadbent and Poon, 2015). For example, Maldonado-Mahauad *et al.* (2018) aimed to reveal the relationships between student's behavioral patterns and their SRL strategies in MOOCs. The authors identified three behavioral patterns and their corresponding SRL strategies. The comprehensive learners followed the sequential structure of the course materials, which sets them up for gaining a deeper understanding of the content. The targeting learners who strategically engaged with specific course content will help them pass the assessments. The sampling learners who exhibited more erratic and less goal-oriented behavior. These learners had lower SRL and underperformed.

Several studies have demonstrated a positive relationship between SRL strategies and academic achievement in online learning environments. The literature showed SRL strategies were crucial to be a successful student in online learning. Connecting online learning patterns with SRL strategies can improve course design and offer personalized learning.

As learning patterns reflect a student's learning strategies, efforts also can be found in learning pattern analysis. Possible directions include the following:

- identifying and comparing different types of learning patterns (Kahan *et al.*, 2017);
- analyzing patterns with frameworks for personalized interventions (Chen *et al.*, 2012);
- discovering connections between learning pattern and learning performance (Goda *et al.*, 2015);
- exploring the influencing factors of learning pattern (Sun *et al.*, 2018);
- shaping learning patterns via instructional strategies (Liu, 2016); and
- predicting student's performance via predictive modeling (Derntl and Motschnig-Pitrik, 2004).

The literature review showed that topic (3) attracted lots of research. Learning patterns research can be classified into descriptive and predictive research. Descriptive research aims to depict student's behavior patterns across time to reveal the impact of course design and personal learning preferences and predictive research aims to predict learning performance or to identify at-risk students via learning pattern analyses. However, only few of research were predictive, the work of this paper was a supplement to the existed research.

2.2 Time-series analysis in educational research

A student's learning patterns consist of his or her behaviors performed in the LMS. Therefore, a student's learning patterns can be considered as time-series data. Time-series analysis includes a set of analytic methods to analyze time-series data. Related techniques have been applied in many areas, such as traffic flow forecasting, water quality monitoring and natural language processing.

Based on the literature, a student's final grade is positively correlated with his or her engagement level in the course. Most of the studies used a student's total frequency to represent student engagement (Yang *et al.*, 2016) for depicting student's learning patterns or for predicting learning performance. However, there are studies with exceptional cases in which engagement level was not correlated with expected outcomes (Yang *et al.*, 2018) and resulted in low prediction accuracy. Hung *et al.* (2015) explained that total frequency is not the best approach for satisfactory prediction accuracy for the following reasons:

- total frequency data failed to consider variances in learning pattern; different learning patterns might result in the same total frequencies;
- total frequency failed to consider differences in student learning preferences;
- total frequency failed to consider variances in course activity requirements across different course modules; and
- total frequency failed to consider that more than one pattern might lead to success or failure.

The authors proposed a time-series clustering method to classify students by their micro-level behavior similarities, which resulted in a significant improvement in the model's accuracy and recall rates, along with the ability to identify at-risk students in the middle of the semester. Brooks *et al.* (2015) proposed a method for modeling learner interaction features based on time-series data to predict learning performance and to find patterns for successful and failed students. Yang *et al.* (2017) used the students' click record of watching videos on the MOOC platform to build time-series data and to predict the student's performance through artificial neural networks. Based on the literature review, only limited studies adopted deep learning models for early warning prediction.

2.3 Autoencoder in educational research

As it uses deep learning for feature extraction, autoencoder has attracted a lot of research effort in recent years. Autoencoder is an unsupervised deep learning model that can perform dimension compression and then restore the information from the input data, while removing the noise from the input data at the same time (Masci *et al.*, 2011). The model consists of encoder and decoder layers. The encoder compresses information through several layers of deep neural networks, and the decoder restores the information of the initial input data from the output data of the encoder (Wang *et al.*, 2016). The training process is to ensure that the model output data does not lose the critical information contained in the original data. Depending on the data type and the encoder-decoder structure, autoencoder can exist in many different forms, such as deep neural network (DNN)-autoencoder, convolutional neural networks autoencoder and LSTM-autoencoder (Principi *et al.*, 2019). Autoencoder is one of the most popular models for feature processing. It has good effectiveness at feature extraction, dimension reduction, data denoise and information recovery (Gondara and Wang, 2017). The process of feature extraction can be accompanied by denoising and by dimensionality reduction. Several studies indicate that using an autoencoder can improve a model's performance via feature compression and noise removal (Zhao *et al.*, 2017). The autoencoder has also been adopted in educational research. For example, Du *et al.* (2020) proposed an integrated framework (LVAEPre) based on a latent variational autoencoder (LVAE) with a DNN to learn the latent distribution of at-risk students and to generate at-risk samples for solving the highly imbalanced data issue. Extensive experiments show that LVAEPre can effectively handle the imbalanced education dataset and provide better and more stable prediction results. Li *et al.* (2017) proposed different composite models that incorporate multiple features to infer the behaviors of the next two weeks based on features extracted from a weekly history of learning data. The stacked sparse autoencoder (SSAE) was used to discover the high-level representation of input features and the correlations among them. The results showed that the SSAE+Softmax model achieved a higher AUC score consistently, and that it was superior to the baseline support vector machine model.

Most of the existing studies used the data recovered by the decoder with the same number of dimensions as the input data. In this paper, the customized method utilizes the outcomes from the encoder, rather than the decoder. Besides, the encoder compresses multiple behavioral behaviors into one encoded time-series. Such processing provides a convenient format for later analyses.

3. Methodology

Figure 1 shows the framework of the proposed method. It contains four steps:

1. Encoding behavioral patterns: students' time-series behaviors are encoded with LSTM-autoencoder.
2. Predicting learning performance: the encoded time-series data are used for predictive modeling.
3. Deciphering learning patterns. successful and failed patterns are classified with time-series clustering.
4. Integrating and interpreting results: integrate hidden knowledge discovered from different types of successful and at-risk students for decision-making.

Figure 2 shows the structure of LSTM-autoencoder. Different from other studies to utilize the outcomes generated by the decoder. This study needs the output from the encoder only and decreases the data dimension into one for later analyses. The input variables are the students' behavioral sequences performed in the LMS. Because a student can choose when and how to interact with course components, the collected time-series data can contain excessive noise data. The customized encoder aims to perform two actions:

1. extract key features from the time-series data (N behaviors * M weeks) and remove data noise; and
2. compress multiple micro-level behavioral time-series data into one encoded time-series data (1 encoded behavior * M weeks).

The one encoded time-series is a friendly format for predictive modeling and time-series clustering.

After the LSTM-autoencoder detected the best long-short-term series to extract key features and encode multiple patterns into one behavioral pattern, a predictive model was constructed using the encoded behavioral pattern on performance prediction. To discover hidden insights about the connection between a student's learning pattern and his/her learning performance, time-series clustering was conducted to group at-risk patterns by encoded pattern similarity. All of the analytic results were integrated, to interpret the results and to develop personalized interventions.

Figure 1 Framework of proposed method

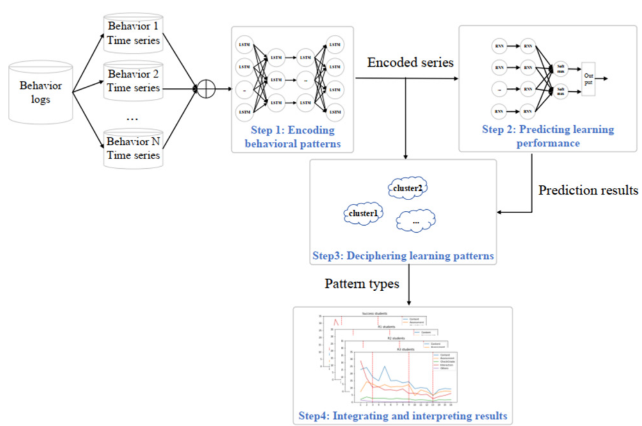
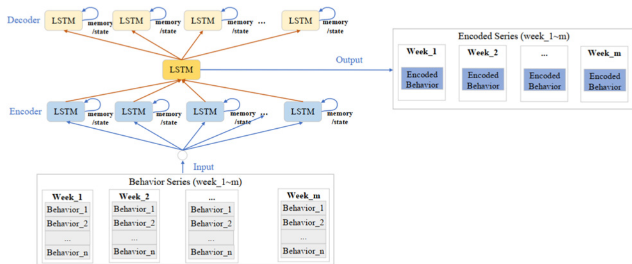


Figure 2 Structure of LSTM-autoencoder



4. Case study

A series of experiments were conducted to validate the effectiveness of the proposed analytic flow. In the predictive modeling, we compared the models constructed by total frequency and the proposed method with the data collected from the whole semester. Second, we compared both models' early warning capabilities. Finally, time-series clustering was applied to examine why the proposed method was better than the traditional approach. Data were collected from a fully online K-12 school in the USA in the academic years 2014–2016. All of the courses were hosted in the Blackboard LMS. The dataset contained 4,706 students in 476 courses, with 3,625,619 logs in total. Blackboard tracked all student actions performed on the platform. However, depending on the tool adopted in the course, the same behaviors might have been recorded as different behavior descriptions. In addition, certain behaviors were only exhibited in specific courses. To avoid data sparsity and to be able to generalize the model, all of the learning behaviors were grouped into the following five categories:

1 *Content*: all behaviors related to the course material access.
2 *Assessment*: all behaviors related course assessment access.
3 *Grade Check*: behaviors to check course grade (this behavior had been identified as a key early indicator in the literature).
4 *Interaction*: all behaviors related to in-person interactions, including peer interactions or teacher-student interactions.
5 *Others*: other minor types of behaviors that do not belong in the above categories.

One key decision in the time-series analysis is to determine the optimal duration for forming time-series data. Hung *et al.* (2015) compared a daily, weekly and bi-weekly duration and found that the weekly duration generated the most accurate results. Therefore, all of the behavior data in this study were aggregated weekly. After initial data processing, each student had five categories of time-series data across 16 weeks (content, assessment, grade check, interaction and others). The sum of the five behaviors was equal to the total frequency. Individual behaviors represent time-series data at the micro-level.

To perform the early warming prediction, the dataset needed to label a student's at-risk status as the target dataset. The final grade was used as the target, with 60 as the passed/failed threshold. Based on the threshold, 3,673 students were labeled as successful and 934 were labeled as at-risk. All of the time-series data went through two transformation methods – log and Softmax – to improve the model's performance. These two transformations were also compared to identify the best approach for predictive modeling.

Grid search and five-fold cross-validation were applied to identify the best parameters and to ensure the result's stability. The outcomes were evaluated by the overall accuracy and recall rates. In early warning prediction, the recall rate (i.e. the percentage of at-risk students that can be captured by the model) is more important than the accuracy rate (Shelton *et al.*, 2018).

The baseline methods included for the model performance comparisons were as follows:

- Total Frequency: each of the students was represented with 16 (weeks) * 1 (total frequency) data.
- Five Behaviors: each of the students was represented with 16 (weeks) * 5 (behaviors) data.

The proposed method compressed five learning behavioral patterns into one encoded time-series (16*5 => 16*1).

4.1 Data security and de-identification

The authors obtained the institutional review board panel's approval of the institution before conducting any data collection and analysis. All data files were password protected and stored in a secured server. Personal variables such as name, residential address and email were deleted from data files. The student ID was transformed via SHA-256. This study only reports results at the group level. There is no individual who can be identified from the results.

4.2 Behavioral pattern encoding

All five behavioral patterns were extracted and compressed into one encoded time-series via the LSTM-autoencoder. Figure 3 compares the encoded time-series between the actual successful students (in green) and the actual at-risk students (in yellow). The results show that at-risk students' encoded series seem to be gathered into several different groups; this is worth further investigation. The results validated the assumptions of van Leeuwen *et al.* (2019) and Hung *et al.* (2015) about multiple successful and at-risk patterns.

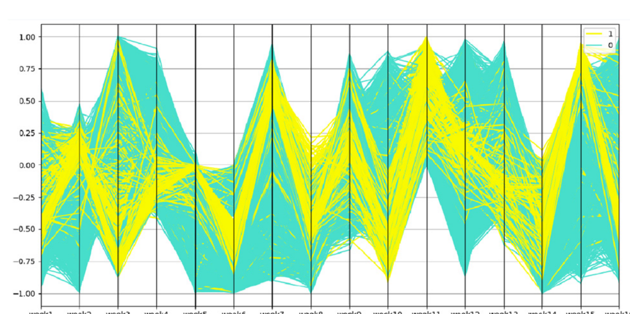
4.3 Predicting learning performance via learning patterns

4.3.1 Performance comparisons

Two approaches were adopted as baseline models:

1 *Total Frequency*, which, based on the literature review, is the most common input used in predictive modeling.

Figure 3 Distribution of students' encoded series



2 *Five Behaviors*, which included the five total behavioral patterns without autoencoder processing.

The first step was to compare the LSTM-autoencoder with the two baseline approaches by analyzing the entire semester's 16-week data. All three input datasets (total frequency behavioral patterns, five behavioral patterns and the LSTM-autoencoder pattern) connected with the RNN for predictive modeling.

Table 1 lists the five-fold cross-validation results of the three methods with the best parameters identified by grid search, individually. The results show that the log transformation is a better data processing method, with higher recall and precision rates in all three methods. The best model is the encoded series, with the best recall and accuracy rates. The LSTM-autoencoder + RNN model can capture about 71.10% of at-risk students at the end of the semester. The LSTM-autoencoder + RNN model captured 2.9% more students than the Five Behaviors + RNN model and 3% more students than the Total Frequency + model. As there were 934 failed students in our data set, that means the LSTM-autoencoder + RNN model identified about 28 more students than the other models.

4.3.2 Early warning prediction

The second comparison aimed to identify which approach had the best early warning capability. Therefore, different lengths of shorter time-series data were extracted to compare prediction accuracies and recall rates (from 2 weeks to 16 weeks).

Table 2 shows five-fold cross-validation with the best parameters from the grid search for individual models. The results show that the longer lengths provided richer information, so accuracy and recall rates increased with the longer time-series. However, the LSTM-autoencoder did not have better accuracy rates than Total Frequency or Five Behaviors. The strength of LSTM-autoencoder showed in the recall rates. As seen in Table 1, the three methods showed similar accuracy rates across different time lengths, but the LSTM-autoencoder had higher recall rates in each of the comparisons than either of the baseline methods.

To further confirm the effectiveness of the proposed approach, these models were deployed to generate early warning predictions. This data set contained 1,422 fully online K-12 students in the United States in Fall 2018. All three models performed early-warning predictions in the middle of the semester and compared the predictions with students' final performance. The LSTM-autoencoder showed a higher recall rate (64.54%) than the Five Behaviors + RNN (59.64%), or the Total Frequency + RNN (53.54%).

4.4 Deciphering learning patterns

Because the encoded time-series contains synthesized information from students with the mechanism of long-short memory, it is hard to reveal how the information was weighted and condensed. Therefore, we conducted time-series clustering to group at-risk students based on their encoded time-series similarity. The outcomes classified all at-risk students into three groups: R1(402 students, 43.04%), R2 (425 students, 45.50%) or R3 (107 students, 11.46%).

To further understand what caused the prediction differences, we chunked the entire semester into different sections based on the turning points in the behavioral series via visual observation. In time-series analysis, each of the turning points represents a pattern change for students (Wecker, 1979). These turning points resulted in four sections (chunks) over the entire semester: S1 (weeks 1–3), S2 (weeks 4–9), S3 (weeks 10–12) and S4 (weeks 13–16). To understand how students allocated their efforts in the course, the average frequencies of individual behaviors were calculated. Figure 4 compares the behavioral patterns of successful, R1, *R2* and R3 students in four sections (S1-S4). Patterns of successful students (average total frequency: 895.3). Successful students had higher engagement frequencies throughout the semester. The smoother patterns also indicate their consistent participation levels in different behaviors (content access, assessment and discussion interaction). Week 13 was a holiday break and showed lower participation levels, on average.

Patterns of R1 students (average total frequency: 145.6). R1 students had the lowest frequency throughout the semester. Their interaction had higher average frequencies than other behaviors. Overall, R1 students had the lowest participation levels compared with other groups. Therefore, these students were named *Low-Engaged Students*.

Patterns of *R2* students (average total frequency: 515.8). *R2* students showed patterns similar to successful students, except they had the lowest average participation level in discussion interaction. Therefore, these students were named *Low-Interaction Students*.

Patterns of R3 students (average total frequency: 554.4). Unlike the other two at-risk groups, R3 students showed the highest participation levels at the beginning of the semester and then decreased gradually toward the end of the semester. It seems that R3 students gradually lost their motivation in the courses. Therefore, these students were named *Non-Persistent Students*.

4.5 Integrating and interpreting results

4.5.1 Unique strength of the proposed method in identifying at-risk students

Table 3 lists students in individual clusters and shows that the percentages of at-risk students identified by individual approaches, Total Frequency, Five Behaviors and LSTM-

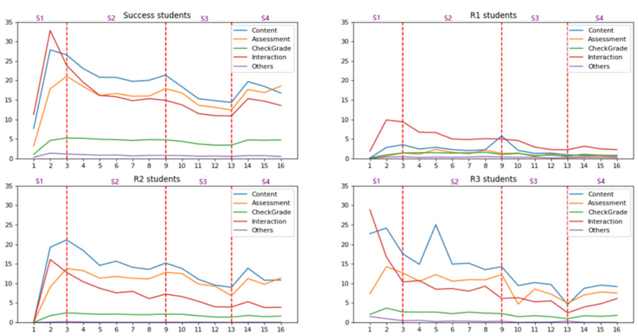
Table 1 Learning performance prediction comparison

Series Type	Prediction Parameter	5-Fold Average (log)	5-Fold Average (Softmax)
Total Frequency	Accuracy	92.16	84.44
	Recall	68.10	47.97
Five Behaviors	accuracy	91.49	88.69
	recall	68.20	61.14
LSTM-autoencoder	accuracy	92.23	89.23
	recall	71.10	62.20

Table 2 Early weeks prediction comparison

Average Accuracy of Five-Fold Cross-Validation								
Series type	2 weeks	4 weeks	6 weeks	8 weeks	10 weeks	12 weeks	14 weeks	16 weeks
Total frequency	81.88	86.59	87.71	88.19	89.69	90.32	91.49	92.16
Five behaviors	81.72	86.22	87.48	88.02	89.00	89.19	90.08	91.49
LSTM-autoencoder	81.44	86.32	87.56	87.56	89.82	90.28	91.49	92.23
Average Recall of 5-Fold Cross-Validation								
Series type	2 weeks	4 weeks	6 weeks	8 weeks	10 weeks	12 weeks	14 weeks	16 weeks
Total frequency	40.79	44.01	50.65	52.46	59.00	60.92	66.06	68.10
Five behaviors	42.30	44.32	51.71	54.18	57.50	60.93	64.88	68.20
LSTM-autoencoder	43.26	46.57	52.78	56.28	59.53	61.67	66.59	71.10

Figure 4 Comparisons of successful and failed students in behavioral patterns



autoencoder. The results indicate that total frequency has the highest recall rate in R1. LSTM-autoencoder has the highest recall rates in *R*2 and R3 (with R3's recall rate the same as Five Behaviors). In total, the LSTM-autoencoder can capture around 4% more at-risk students than the baseline methods. The differences mainly come from *R*2 and R3 students (compared with Total Frequency) and *R*2 students (compared with Five Behaviors).

4.5.2 Distributions of at-risk types by subjects

Table 4 compares the relationships between subject areas and at-risk types with lift values. The lift value is an indicator that shows the probabilities of individual risk types in specific subjects over the probabilities of individual risk types in all students. For example, if R1's lift value in Computer Science is equal to 1, then R1's probability in Computer Science is equal to R1's probability in all students. If the value is larger than 1, then R1's probability in Computer Science is larger than the one in all students. Therefore, if the value is smaller than 1, then R1's probability in Computer Science is smaller than the overall occurrence. Political Science showed higher probabilities in all three types of at-risk students, while Social Science showed all lower probabilities.

Excluding these two subjects, Table 4 shows that Computer Science, Geography and History had a higher probability of low-engaged students (R1). Earth Science, History and Life Science each had a higher probability of Low-interaction students (*R*2). Computer Science, Earth Science and Economics each had a higher probability of non-persistent students (R3). This information could be helpful in better understanding the motivations that influence student outcomes and could lead to a more careful design of curriculum and interactions, to better address the needs of these types of students.

5. Discussion

5.1 Effectiveness of learning performance prediction via *learning patterns*

The effectiveness of the LSTM-autoencoder was validated by the results of the entire semester (16 weeks) and the predictions at earlier time points, along with another dataset from a different academic year. All three approaches show similar overall accuracy rates. The results indicate that LSTM-autoencoder captured more at-risk students than the other two baseline methods. The results indicate that the proposed method can extract and compress key features that can lead to the target situation (i.e. at-risk). Besides, because the findings were validated from more than four hundred K-12 online courses, the approach should have practical value for learning pattern analysis.

For early warning, the most important task is to identify as many at-risk students as possible for early interventions. Therefore, the recall rate is more meaningful than the accuracy in model evaluation (Shelton *et al.*, 2018). For educational purposes, a small increase in the recall rate is still meaningful to prevent more students from possible academic failure. In the case study, the LSTM Autoencoder captured about 2.1% and 3.8% (i.e. 21 and 38 students) more at-risk students than the

Table 3 Time-series cluster results for failed students

Cluster	Student number	(%)	Total frequency	Five bBehaviors	LSTM-autoencoder
R1	402	43.04%	350(87.06%)	335(83.33%)	336(83.58%)
R2	425	45.50%	241(56.71%)	248(58.35%)	273(64.24%)
R3	107	11.46%	45(42.06%)	55(51.40%)	55(51.40%)
Total	934	100%	636(68.10%)	638(68.2%)	664(71.10%)

Table 4 Relationship between of course category and failed learning pattern

Cluster	R1	R2	R3
Computer Science	1.29	0.98	1.29
Earth Science	0.80	1.10	2.10
Economy	0.99	0.63	1.49
Geography	1.59	0.90	0.00
History	1.41	1.34	0.99
Life Science	0.81	1.17	0.56
Physical Science	0.99	1.00	0.69
Political Science	1.45	1.11	1.92
Social Science	0.77	0.60	0.68

Five Behaviors and Total Frequency models respectively in the middle of the semester. The increments are meaningful for instructors and school administrators (Pelaez, 2019).

5.2 Unique strengths of the proposed method

R1 (low-engaged) students represent the classic at-risk students identified in the literature (Hussain *et al.*, 2018). Pelaez (2019) also found other types of at-risk students who are hard to be identified via predictive modeling. However, in the case study, we found that the LSTM-autoencoder performed better in identifying low-interaction (*R*2) and non-persistent (R3) students than the baseline models, as R1 students are easier to be identified through the instructor's observations or via weekly activity reports. The ability to capture *R*2 and R3 students is more beneficial for in-time interventions.

In addition, the proposed approach allows researchers and teachers to observe at-risk students at the micro-level. For example, Political Science had higher at-risk chances in all three at-risk types. This could be due to something related to the course design or the grading policy, so we excluded Political Science for the ensuing discussions. The patterns of *R*2 students were similar to those of successful students, except for the lower frequencies of discussion interactions. When further comparing the lift values by subject area, the results showed that Earth Science, Life Science and History had higher chances of having *R*2 students. Our interpretation is that these subjects require more online discussions to demonstrate both students' knowledge integration and higher-order thinking (Blasco-Arcas *et al.*, 2013).

On the other hand, Computer Science, Earth Science and Economics showed higher chances of having R3 students. These subjects might be difficult for students. Therefore, R1, *R*2 and R3 students showed higher or lower occurrence probabilities in different subjects. The results might be related to the difficulty level of the course, the course design, or the instructional strategies in these subjects. Overall, LSTM-autoencoder provided valuable insights and identified additional types of at-risk students with high research and practical values.

5.3 Hidden knowledge and implications extracted from the case study

In the case study, the proposed method identified one successful pattern with the following characteristics:

- Showing consistent participation levels (Anouschka *et al.*, 2018). The pattern reflects SRL strategies of goal orientation and time management strategy, respectively, at the forethought and performance phases (Zimmerman and Moylan, 2009).
- Showing balanced behavioral frequency distributions (Kahan *et al.*, 2017). The pattern reflects SRL strategies of studying, help-seeking and self-evaluation, respectively, at the performance and reflection phases (Zimmerman and Moylan, 2009).
- Adjusting learning behaviors to meet with course requirements accordingly (Winne, 2006). The pattern reflects the SRL strategy of adaptive/defensive at the reflection phase (Zimmerman and Moylan, 2009).

Below we discuss possible interventions for each of the at-risk types based on the "hidden knowledge" extracted from the analysis results.

The R1 students were low-engaged throughout the semester. The higher interaction frequency might represent help-seeking behaviors. They were learners with low SRL skills. As they were K-12 students, instructors might consider guiding them to practice SRL skills in all three phases (Panadero, 2017). Another strategy is to cultivate collaborations among learners at the beginning of the semester. Studies have shown that such collaborations can foster students' tendency of active participation (Blasco-Arcas *et al.*, 2013).

The *R*2 students SRL strategies of studying, repeating and self-evaluation with higher course material accessing and assessment taking frequencies. However, these students showed a low interaction level with their peers. They missed opportunities to obtain peer feedback at the reflection phase. Earth Science, Life Science and History showed higher chances of having *R*2 students. These subjects might require intensive discussions, which are beneficial to the students' understanding. One possible intervention for *R*2 students would be the use of small group discussions (Stokoe, 2000). Jacobs and Seow (2015) suggested that the best group size is two members with more opportunities to be active. The instructor's social presence is another strategy to foster online discussions.

The R3 students showed a high participation level at the beginning of the semester, but they lacked persistence engagement. That indicates these students might lose their self-efficacy and motivation during the semester. These students might have issues with the goal-setting or maintaining motivation during the learning process. Computer Science, Earth Science and Economics had higher percentages of non-persistent students. Lepper *et al.* (2005) found positive reinforcement is an effective strategy to boost students' intrinsic motivation. Positive reinforcement should occur right after students performed the desired behavior. Provide frequent praise when students are learning a new skill and decrease the amount once the skill has been mastered.

6. Conclusion

This article analyzed learning pattern from the micro-level based on LSTM-encoder, identified the at-risk pattern types and found significant educational indication from the case study. We found that, successful students:

- showing consistent participation levels;
- showing balanced behavioral frequency distributions; and
- adjusting learning behaviors to meet with course requirements accordingly.

The at-risk students:
- were low-engaged throughout the semester;
- showed a low interaction level with their peers; and
- lacked persistence engagement.

By combining learning patterns with SRL strategies, we found that successful students showed more reasonable or complete SRL strategies than at-risk students.

The above results show at-risk status might happen at different time points during the semester. The contribution of this article was as follows:
- Proposed a method that can identify different at-risk types in time and suggested personalized interventions.
- Identified one successful and three at-risk patterns with corresponding SRL strategies [consistent and balanced (S), low engaged (R1), low-interaction (R2) and non-persistent (R3)] via LSTM encoder and time-series clustering. R1 students were commonly found in the literature, but R2 and R3 were rarely discussed in the literature. Therefore, more studies are needed to confirm whether R2 and R3 students also exist in other online educational settings.
- Found that the distributions of R1, R2 and R3 are different across subjects. In addition, this study discussed the unique strength of the LSTM encoder in identifying more non-traditional at-risk students than baseline models.

6.1 Limitation and future research

The limitations of this article were as follows:
- The study identified multiple learning patterns which can lead to the at-risk situation. However, more studies are needed to validate whether the same at-risk types can be found in other educational settings.
- Future studies might further investigate the relationships between specific at-risk types and the unique characteristics of different subject areas.
- This study discussed the unique strength of the LSTM encoder in identifying more non-traditional at-risk students than baseline models.

As students' learning patterns are highly influenced by course design, the analytic approach might need adjustments in other course design settings. More studies are needed to validate the approach's generalizability.

This study demonstrates how the proposed method can be used to analyze students' learning patterns and interpret them with the SRL model. As many educational frameworks or theories lack supports from practical data, future research might focus on adopting the analytic method for theory validation. In addition, more studies are needed to validate whether the same at-risk types can be found in other educational settings. Finally, the relationship between subjects and at-risk types is worth further investigation.

References

Anouschka, V.L., Nynke, B., Heleen, V.R. and Jurgen, V.O. (2018), "The role of temporal patterns in students' behavior for predicting course performance: a comparison of two blended learning courses", *British Journal of Educational Technology*.

Bannert, M., Reimann, P. and Sonnenberg, C.J.M. (2014), "Process mining techniques for analysing patterns and strategies in students' self-regulated learning", *Metacognition and Learning*, Vol. 9 No. 2, pp. 161-185.

Blasco-Arcas, L., Buil, I., Hernández-Ortega, B. and Sese, F.J. (2013), "Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance", *Computers & Education*, Vol. 62, pp. 102-110.

Boekaerts, M.J.L. (1997), "Self-regulated learning: a new concept embraced by researchers, policy makers, educators, teachers, and students", *Learning and Instruction*, Vol. 7 No. 2, pp. 161-186.

Broadbent, J. (2017), "Comparing online and blended learner's self-regulated learning strategies and academic performance", *The Internet and Higher Education*, Vol. 33, pp. 24-32.

Broadbent, J. and Poon, W.L. (2015), "Self-regulated learning strategies & academic achievement in online higher education learning environments: a systematic review", *The Internet and Higher Education*, Vol. 27, pp. 1-13.

Brooks, C., Thompson, C. and Teasley, S. (2015), "A time-series interaction analysis method for building predictive models of learners using log data", *In Proceedings of the fifth international conference on learning analytics and knowledge*, pp. 126-135.

Chen, J., Jin, Q. and Huang, R. (2012), "Goal-driven process navigation for individualized learning activities in ubiquitous networking and IoT environments", *J. Ucs*, Vol. 18 No. 9, pp. 1132-1151.

Cui, Z., Ke, R., Pu, Z. and Wang, Y. (2020), "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values", *Transportation Research Part C: Emerging Technologies*, Vol. 118, p. 102674.

Derntl, M. and Motschnig-Pitrik, R. (2004), "A pattern approach to person-centered e-learning based on theory-guided action research", *In Networked Learning Conference*.

Du, X., Yang, J. and Hung, J.L. (2020), "An integrated framework based on latent variational autoencoder for providing early warning of at-risk students", *IEEE Access*, Vol. 8, pp. 10110-10122.

Du, X., Yang, J., Shelton, B. and Hung, J.L. (2019), "Is learning anytime, anywhere a good strategy for success? Identifying successful spatial-temporal patterns of on-the-job and full-time students", *Information Discovery and Delivery*, Vol. 47 No. 4, doi: 10.1108/IDD-09-2019-0060.

Fratamico, L., Conati, C., Kardan, S. and Roll, I. (2017), "Applying a framework for student modeling in exploratory learning environments: comparing data representation granularity to handle environment complexity", *International Journal of Artificial Intelligence in Education*, Vol. 27 No. 2, pp. 320-352.

Gamboa, J.C.B. (2017), "Deep learning for time-series analysis", arXiv preprint arXiv:1701.01887.

Gamulin, J., Gamulin, O. and Kermek, D. (2016), "Using Fourier coefficients in time-series analysis for student performance prediction in blended learning environments", *Expert Systems*, Vol. 33 No. 2, pp. 189-200.

Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y. and Miyagawa, H. (2015), "Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes", *Learning and Individual Differences*, Vol. 37, pp. 72-80.

Gondara, L. and Wang, K. (2017), "Recovering loss to followup information using denoising autoencoders", *In 2017 IEEE International Conference on Big Data (Big Data)*, pp. 1936-1945, IEEE.

Hung, J.L., Wang, M.C., Wang, S., Abdelrasoul, M., Li, Y. and He, W. (2015), "Identifying at-risk students for early interventions – a time-series clustering approach", *IEEE Transactions on Emerging Topics in Computing*, Vol. 5 No. 1, pp. 45-55.

Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R. (2018), "Student engagement predictions in an e-learning system and their impact on student course assessment scores", *Computational Intelligence and Neuroscience*, Vol. 2018

Jacobs, G. and Seow, P. (2015), "Cooperative learning principles enhance online interaction", *Journal of International and Comparative Education (JICE)*, Vol. 4 No. 1, pp. 28-38.

Kahan, T., Soffer, T. and Nachmias, R. (2017), "Types of participant behavior in a massive open online course", *The International Review of Research in Open and Distributed Learning*, Vol. 18 No. 6, pp. 1-18.

Lehmann, T., Hähnlein, I. and Ifenthaler, D. (2014), "Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning", *Computers in Human Behavior*, Vol. 32, pp. 313-323.

Lepper, M.R., Corpus, J.H. and Iyengar, S.S. (2005), "Intrinsic and extrinsic motivational orientations in the classroom: age differences and academic correlates", *Journal of Educational Psychology*, Vol. 97 No. 2, p. 184.

Li, K. (2019), "MOOC learners' demographics, self-regulated learning strategy, perceived learning and satisfaction: a structural equation modeling approach", *Computers & Education*, Vol. 132, pp. 16-30.

Li, Y., Fu, C. and Zhang, Y. (2017), "When and who at risk? Call back at these critical points", *International Educational Data Mining Society*.

Liu, H. (2016), "An analysis on blended learning pattern based on blackboard network platform: a case study on the course of recruitment and employment management", *International Journal of Emerging Technologies in Learning (iJET)*, Vol. 11 No. 9, pp. 4-8.

Ma, X., Tao, Z., Wang, Y., Yu, H. and Wang, Y. (2015), "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data", *Transportation Research Part C: Emerging Technologies*, Vol. 54, pp. 187-197.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R.F., Morales, N. and Munoz-Gama, J. (2018), "Mining theory-based patterns from big data: identifying self-regulated learning strategies in massive open online courses", *Computers in Human Behavior*, Vol. 80, pp. 179-196.

Masci, J., Meier, U., Cireşan, D. and Schmidhuber, J. (2011), "Stacked convolutional auto-encoders for hierarchical feature extraction", in *International Conference on Artificial Neural Networks*, Springer, Berlin, Heidelberg, pp. 52-59.

Panadero, E. (2017), "A review of self-regulated learning: six models and four directions for research", *Frontiers in Psychology*, Vol. 8, p. 422.

Pelaez, K. (2019), "Using a latent class forest to identify at-risk students in higher education", *Journal of Educational Data Mining*, Vol. 11 No. 1, pp. 18-46.

Pintrich, P.R. (2004), "A conceptual framework for assessing motivation and self-regulated learning in college students", *Educational Psychology Review*, Vol. 16 No. 4, pp. 385-407.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S.S. (2018), "A survey on deep learning: algorithms, techniques, and applications", *ACM Computing Surveys*, Vol. 51 No. 1, pp. 1-36.

Principi, E., Rossetti, D., Squartini, S. and Piazza, F. (2019), "Unsupervised electric motor fault detection by using deep autoencoders", *IEEE/CAA Journal of Automatica Sinica*, Vol. 6 No. 2, pp. 441-451.

Shelton, B.E., Yang, J., Hung, J.L. and Du, X. (2018), Two-stage predictive modeling for identifying at-risk students, *In International Conference on Innovative Technologies and Learning*, pp. 578-583, Springer, Cham.

Stokoe, E.H. (2000), "Constructing topicality in university students' small-group discussion: a conversation analytic approach", *Language and Education*, Vol. 14 No. 3, pp. 184-203.

Sun, J.C.Y., Lin, C.T. and Chou, C. (2018), "Applying learning analytics to explore the effects of motivation on online students' reading behavioral patterns", *International Review of Research in Open and Distributed Learning*, Vol. 19 No. 2.

van Leeuwen, A., Bos, N., van Ravenswaaij, H. and van Oostenrijk, J. (2019), "The role of temporal patterns in students' behavior for predicting course performance: a comparison of two blended learning courses", *British Journal of Educational Technology*, Vol. 50 No. 2, pp. 921-933.

Vermunt, J.D. and Donche, V. (2017), "A learning patterns perspective on student learning in higher education: state of the art and moving forward", *Educational Psychology Review*, Vol. 29 No. 2, pp. 269-299.

Viberg, O., Khalil, M. and Baars, M. (2020), "Self-regulated learning and learning analytics in online learning environments: a review of empirical research", *In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pp. 524-533.

Wang, Y., Yao, H. and Zhao, S. (2016), "Auto-encoder based dimensionality reduction", *Neurocomputing*, Vol. 184, pp. 232-242.

Wecker, W.E. (1979), "Predicting the turning points of a time-series", *The Journal of Business*, Vol. 52 No. 1, pp. 35-50.

Winne, P.H. (2006), "How software technologies can improve research on learning and bolster school reform", *Educational Psychologist*, Vol. 41 No. 1, pp. 5-17.

Yang, X., Guo, X. and Yu, S. (2016), "Effects of cooperative translation on Chinese EFL student levels of interest and self-efficacy in specialized English translation", *Computer Assisted Language Learning*, Vol. 29 No. 3, pp. 477-493.

Yang, X., Li, J. and Xing, B. (2018), "Behavioral patterns of knowledge construction in online cooperative translation activities", *The Internet and Higher Education*, Vol. 36, pp. 13-21.

Yang, T.Y., Brinton, C.G., Joe-Wong, C. and Chiang, M. (2017), "Behavior-based grade prediction for MOOCs via time-series neural networks", *IEEE Journal of Selected Topics in Signal Processing*, Vol. 11 No. 5, pp. 716-728.
Zhao, F., Feng, J., Zhao, J., Yang, W. and Yan, S. (2017), "Robust LSTM-autoencoders for face de-occlusion in the wild", *IEEE Transactions on Image Processing*, Vol. 27 No. 2, pp. 778-790.
Zimmerman, B.J. and Moylan, A.R. (2009), "Self-regulation: where metacognition and motivation intersect", *In Handbook of Metacognition in Education*, pp. 311-328.Routledge.
Zimmerman, B.J. and Schunk, D.H. (2011), *Handbook of Self-Regulation of Learning and Performance*, Routledge/Taylor & Francis Group.

Further reading

Hung, J.L. and Zhang, K. (2008), "Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching", *MERLOT Journal of Online Learning and Teaching*.
Lu, O.H., Huang, A.Y., Huang, J.C., Lin, A.J., Ogata, H. and Yang, S.J. (2018), "Applying learning analytics for the early prediction of students' academic performance in blended learning", *Journal of Educational Technology & Society*, Vol. 21 No. 2, pp. 220-232.

Corresponding author

Jui-Long Hung can be contacted at: andyhung@boisestate.edu