

# Learning Analytics for Monitoring Students Participation Online: Visualizing Navigational Patterns on Learning Management System

Leonard K.M. Poon<sup>(✉)</sup>, Siu-Cheung Kong, Thomas S.H. Yau,  
Michael Wong, and Man Ho Ling

Department of Mathematics and Information Technology,  
The Education University of Hong Kong, Tai Po, Hong Kong SAR, China  
{kmpoon, sckong, shyau, mywwong}@eduhk.hk, alphaisfishing@gmail.com

**Abstract.** With the increasing use of blended learning approaches in classroom, various kinds of technologies are incorporated to provide digital teaching and learning resources to support students. These resources are often centralized in learning management systems (LMSs), which also store valuable learning data of students. The data could assist teachers in their pedagogical decision making but they are often not well utilized. This paper proposes the use of data mining and visualization techniques as learning analytics to provide a more comprehensive overview of students' learning online based on log data from LMSs. The focus of this study is the discovery of frequent navigational patterns by sequential pattern mining techniques and the demonstration of how presentation of patterns through hierarchical clustering and sunburst visualization could facilitate the interpretation of patterns. The data in this paper were collected from a blended statistics course for undergraduate students.

**Keywords:** Learning analytics · Blended learning · Sequential pattern mining · Hierarchical clustering · Navigational pattern · LMS · Moodle

## 1 Introduction

In Hong Kong, the integration of digital technologies to teaching and learning in a wide range of subjects is growing rapidly. Most higher education institutions in Hong Kong adopt Learning Management Systems (LMSs) as the digital learning environment to support the development of novel e-Learning practices such as blended learning. Despite successful cases showing the new pedagogical approaches are effective in enhancing students learning (e.g. [9]), not many previous studies have fully exploited the learning data of students from digital environment such as LMSs, which reflect the limited usage of these data to inform teaching and learning.

Reimann et al. [18] recently summarized three reasons contributing to the limited usage of learning data. First, large-scale assessment data available to teachers and students are not linked to the learning processes and outcomes of students. Another reason is that even the data reporting students learning in the system are available, the focus of the data is often on activity tracking rather than knowledge tracking, which

provide limited insights for teachers to identify learning problems of student. Lastly, teachers are reported that neither do they feel themselves sufficiently qualified nor do they possess enough time to make good use of the learning data for teaching. These points corresponded to some major problems mentioned by Gómez-Aguilar et al. [4] regarding the information provided by learning platforms (e.g. LMSs), which is only available in text files format or with some basic charts in general. They are not effective and interactive for users to understand the actual usage situation in the system. Also, the massive amount of learning data without effective processing could lead to information overload to users, which drives them away to make use of these valuable data. Furthermore, advanced and pedagogically useful information such as those highlighting various patterns of usage on the system is often not available. In sum, the high complexity of available data and the lack of means to translate these data into comprehensible output on current learning platforms are also part of the reasons contributing to the limited usage reported by Reimann et al. [18].

In view of this, researchers have proposed using learning analytics based on educational data to strengthen teaching and learning in a digital environment. Learning analytics refer to the collection, organization, analysis and reporting of a wide range of data produced by students in the learning context in order to generate information and identify potential issues for prediction and pedagogical decision-making [2]. There is a close relationship between learning analytics and data mining as different data mining methods are also applied in learning analytics. For instance, *sequential pattern mining* (SPM) is one of the most popular methods which study the temporal association between variables, and one of its applications in education is the discovery of common action sequences conducted by users in the online learning environment such as LMSs. The use of data mining methods in learning analytics could better inform teachers about the overall learning processes of students online for them to determine if any interventions are needed, especially in the context of blended learning where a significant portion of time is allocated for students to study online.

This paper attempts to show the use of online learning data to better inform teaching and learning in a blended learning context. Essentially, we will demonstrate the way to effectively discover the frequent navigational patterns through SPM and present the results with hierarchical clustering and sunburst visualization, to show that how learning data could be better used for learning analytics, which still sees a lack of literature support. Results will be illustrated based on the data from a blended statistics course.

## 2 Previous Work

In common blended learning strategies, the provision of effective learning materials is an indispensable component for successful learning of students. The delivery of these materials is mostly by LMSs in current higher education institutions where students could perform different learning tasks or activities online during their time outside classroom. However, the behaviors of students online and their interactions with the learning materials, peers or teachers could also contribute to our understanding of the

learning processes involved. This is the area where learning analytics could best help in revealing different sorts of relations in the online learning environment.

To study the navigational behavior of learners or access patterns of the resources online, analysis could be done by statistical methods (e.g. [5, 10]). However, more often data mining methods are adopted due to its ability to handle large amount of learning data, such as the log files from LMSs. *Sequential pattern mining* (SPM), which is a process of discovering and showing the hidden interrelationships, clusters and data patterns within data for supporting better decision making [23], would be a suitable approach for the investigation of access patterns. Although it is not initially designed for educational purposes, different researches have applied it for pattern discovery in learning data with some valuable insights offered (e.g. [7, 8, 17]). In a recent work by Ziebarth et al. [24], their analysis of the resource usage pattern of students during the exam preparation period in a blended learning course with SPM has discovered the association between the patterns and grades of students in the course. Some other studies have also used both the clustering method and sequential pattern mining on learning data to analyze the behaviors and performance of students online. For instance, Perera et al. [16] integrated these two methods in studying online collaborative activities to distinguish student groups in terms of performance based on the activity patterns of students.

On the other hand, the application of data mining in education has been criticized for the difficulty in the interpretation of results for users due to the complexity. Therefore, developing an effective mean to deliver the results is also vital. Visualization techniques were commonly adopted in the delivery of data mining results to highlight the overall trends and patterns [20]. Zhou et al. [23] further suggested that the implementation of intuitive graphic charts, tables or diagrams could facilitate users' understanding of the pattern based results generated from data mining algorithms, while a possible incorporation of interactive elements for user to explore the results is also desirable [21]. The visual representation of analytics results is sometimes referred as visual analytics, which uses interactive visual interfaces for systematic and analytical reasoning [4].

### 3 Sequential Pattern Mining

In this section, we review some technical details and introduce algorithms for SPM.

Consider a set of items such as resources on an LMS. A *sequence* is an ordered list of items. It may indicate the order of which resources are accessed by a student. An item can occur multiple times in a sequence. The number of items in a sequence is known as its *length*. A sequence  $\alpha = \langle a_1 a_2 \dots a_n \rangle$  is called a *subsequence* of another sequence  $\beta = \langle b_1 b_2 \dots b_m \rangle$ , or  $\alpha$  is *contained* in  $\beta$ , if there exists integers  $1 \leq j_1 \leq j_2 \leq \dots \leq j_n \leq m$  such that  $a_1 = b_{j_1}, a_2 = b_{j_2}, \dots, a_n = b_{j_n}$ . In other words, the subsequence  $\alpha$  can be formed by removing some items from  $\beta$  with the order of the remaining items preserved. For example, the sequence  $\langle a, c, d \rangle$  is a subsequence of the sequence  $\langle a, b, c, d, e \rangle$ , where  $a, b, c, d$ , and  $e$  are items. Due to brevity, we do not consider the general case where an element of a sequence can be a set of items.

Let  $S$  be a set of sequences. We refer to the number or proportion of sequences in  $S$  that contain a sequence  $\alpha$  as the *support* of  $\alpha$  in  $S$ . We denote the support as  $support_s(\alpha)$  and omit the subscript when it is clear from context.

*Sequential pattern mining* (SPM) aims to discover frequent subsequences among a set of sequences. The frequent subsequences are also called *sequential patterns*. Specifically, given a set of sequences  $S$  and a real number  $\xi \in [0, 1]$  as threshold, the problem of SPM is to find all the sequences  $\alpha$  such that  $support_s(\alpha) \geq \xi$ . The threshold  $\xi$  is also known as the *minimum support*.

Many algorithms have been developed to mine sequential patterns efficiently. Some of the earliest attempts, e.g. [1], are based on the well-known Apriori algorithm for association rule mining. However, Apriori-like algorithms needs to consider an exponential number of candidate sequences in the worst case and they require repeated scanning of the data set to check the support of candidate sequences [12]. Those problems make Apriori-like algorithms infeasible for large data sets or when the sequential patterns are expected to be long.

PrefixSpan [14], on the other hand, avoids those problems by taking another approach. Its main idea is to project the database of sequences into a set of smaller databases based on a set of frequent subsequences. The frequent subsequences are then grown and checked in the smaller projected databases separately. The projection step and the growth step are done recursively until no more longer frequent sequences are found. PrefixSpan has been shown empirically to be considerably faster than another Apriori-like algorithm.

SPM may result in a large number of sequential patterns. This may lead to a long processing time and make the mining results hard to understand. Therefore, constraints have been imposed to limit the mining results to those more interesting to users [15]. One constraint used in our work is the gap constraint. It requires the time difference between two items in sequential patterns to be shorter or longer than a given gap. To support the gap constraint, along with other constraints, Hirate and Yamana [6] proposed an algorithm based on PrefixSpan. Their algorithm was further extended by Fournier-Viger et al. [3] to give only closed sequential patterns, meaning that those patterns contained by another patterns with at least the same support are not included in the output.

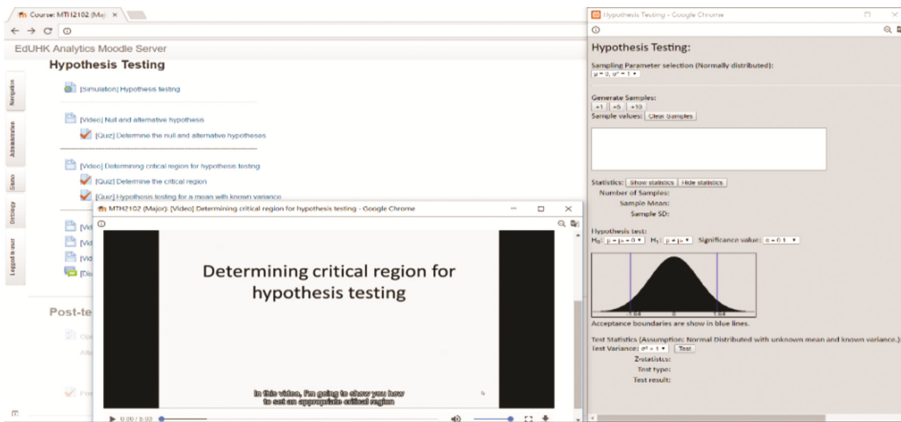
In the current study, we used a SPM algorithm adapted from that by Fournier-Viger et al. [3]. The adaptations were described in [13]. The adapted algorithm allows gap constraint to be specified and produces only closed sequential patterns.

## 4 Preliminary Study

In this section, the proposed methods are demonstrated through a blended statistics course in a university. The study context is first described followed by the overall trend in the course. Next, the discovered patterns from SPM are discussed and more detailed explanations with regard to the hierarchical and sunburst visualization of patterns are provided in the last two parts.

## 4.1 Study Context

The blended statistics course used Moodle as the LMS and the available digital resources on it were statistical simulations, online videos and online quizzes. These resources focused on four fundamental topics in statistics, which were sampling distribution, the central limit theorem, confidence interval and hypothesis testing. Two groups of undergraduate students enrolled in this course, with one group of 70 students who major in Mathematics (major group) and another group of 41 students who minor in Mathematics (minor group). The course length was 8-week long for the major group and 14-week long for the minor group. The major and minor groups had a double and single 3-hour lecture per week respectively, with an examination scheduled at the last week of the course. Students were encouraged to access the resources before and after the lessons for self-learning and knowledge consolidation. Although the course also covered other topics in statistics in addition to the focused topics by the digital resources, students were free to visit the platform in any period during the course. Some additional learning activities were given by teachers on the platform with additional bonus points as incentive. The interface of the platform and the digital resources could be seen in Fig. 1.

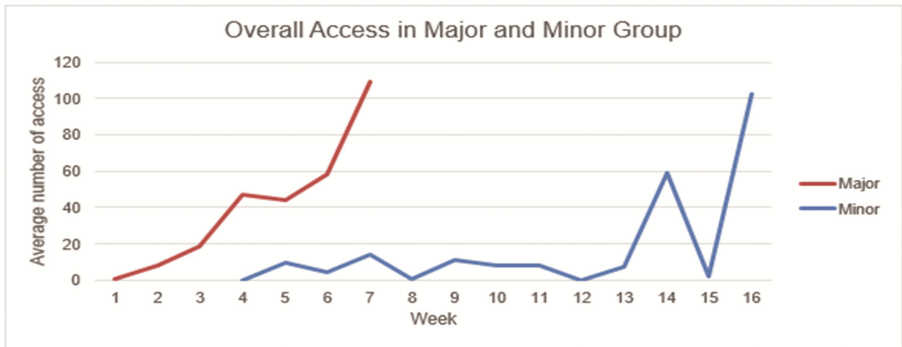


**Fig. 1.** Screenshot of the learning platform and the associated resources. The platform provides three types of digital resources, which are statistical simulations, online videos and online quizzes, for students to learn.

## 4.2 Overall Trend

The general usage situation on the platform could be illustrated by the average number of access across different time periods. Figure 2 provides an example of the overall trend based on the major and minor group. The class for major group started 3 weeks earlier than the minor group. The overall trend indicates that students from the major group were more engaged than the minor group in using the resources available on the platform outside lecture time to support their learning. The access of the platform surged rapidly during the last few weeks of the course in both groups, mainly because these periods

were the preparation time for the final exam during near the end of course, so students accessed the platform more frequently for revision. Through the result in Fig. 2, we could identify the general students' behavior pattern over time. However, more advanced information would be needed to further understand how students make use of the resources online. Results from SPM could provide such information by finding and displaying the frequent navigational patterns of students on Moodle.



**Fig. 2.** Overall access to the digital resources on Moodle over time. The red curve represents the major group while the blue curve represents the minor group. The major group had a generally higher number of accesses to the platform. Both groups demonstrated a higher number of accesses near the end of the course.

### 4.3 Pattern Discovery

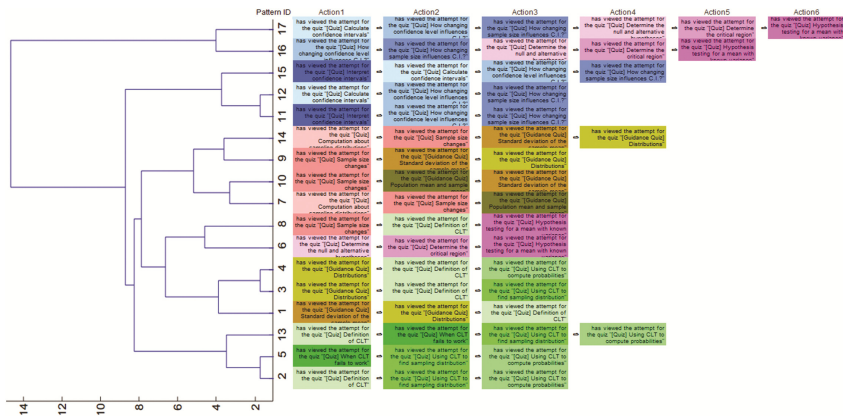
Before the pattern discovery process with our modified SPM, one further preprocessing step was conducted concerning the Moodle log data. Moodle contains fine-grained information about the behaviors of student on the system. In the case of quiz activity, there are different descriptions used by Moodle to record each action of students when they complete the quizzes. If all of them are included, many trivial patterns would be generated. In view of this, those consecutive quizzes related actions (or items) are aggregated into a single quiz action to improve the quality of patterns for better interpretation. The log data on Moodle of both major and minor group students were processed by the algorithm to discover the common access patterns of resources on the platform. We used the adapted method described in Sect. 3 for SPM. The algorithm returned 83 patterns for the major group on the support level of 0.5, and 52 for the minor group on the same support level. These patterns could help teachers to identify the most common patterns adopted by students when they learned online. The results could also show what kind of resources students prefer and the order of which these resources were used.

### 4.4 Visualization Based on Hierarchical Clustering

Results from SPM could be ordered in different ways, for instance according to the support level of the mined sequences or a specific position within the sequences such as

the first or second item. However, we propose that further grouping and visualization of patterns with clustering technique could provide a more coherent presentation of results. This section will first briefly explain the clustering techniques used and explain the way to interpret the associated results, then comparison would be made with the first and second action alphabetical sorting. The mined patterns from minor group are used as an example for the following discussion due to space limitation.

Hierarchical clustering is chosen in the current study for the pattern grouping process. Hierarchical clustering assumes that structure of cluster appears as nested structural level and the results are usually visualized through a graph called dendrogram. In the consideration of distance measure, Levenshethin distance is adopted to calculate the similarity index between pairs of sequences. The definition of Levenshethin distance is the minimum number of operation required to transform a single sequence into the other using insertion, deletion, or substitution of an action in a sequence [22]. Figure 3 presents a row dendrogram based on SPM results from the minor group. Each pattern in the vertical axis is represented by the leaf nodes, which are the right most nodes in the diagram. Each node also represents a corresponding cluster of patterns located at the right of it. The horizontal axis represents the distance or dissimilarity between clusters and the horizontal position of the node shows the dissimilarity between two clusters, as illustrated by the vertical bar. For instance, pattern 2 and pattern 5 share a high similarity, indicated by the short distance between the two leaf nodes and their parent node.



**Fig. 3.** Arrangement of patterns based on hierarchical clustering visualized by a row dendrogram. By the end of each leaf node, the pattern ID is indicated to represent the corresponding pattern on the right. All identical actions in the results are highlighted with the same colors.

Patterns from the minor group are compared under the arrangement of hierarchical clustering visualized by a dendrogram (Fig. 3) and first and second action alphabetical sorting in the sequence (Fig. 4). To aid readers’ interpretation, all actions within the patterns are colorized and the identical actions are highlighted with the same colors across patterns. Those actions which accessed the resources within the same topic in the course were represented by different depth levels of the same color. The result shows

that hierarchical clustering could effectively group similar patterns together in the presentation, especially for those patterns with a long length.



**Fig. 4.** Arrangement of patterns based on first and second action alphabetical sort. Using the pattern pair 16 & 17 and pattern 2 & 5 as comparison with Fig. 3, similar patterns are not grouped together under this sorting criterion.

For example in Fig. 3, pattern 17 with length 6 and pattern 16 with length 7 share a high similarity with only one different action between them. The processing of hierarchical clustering successfully arranges these two patterns together. In contrast, they are not put together under alphabetical sort in Fig. 4. This is because these two patterns contain a different first action and hence even though they have a similar composition of actions, they are not ordered close to each other. Similar situation happens to the pattern with identical and relatively short length such as pattern 2 and pattern 5 with length 3. They are arranged closely by hierarchical clustering but not by alphabetical sort.

In addition, judging from the distribution of colors in the results, the arrangement under hierarchical clustering could effectively group those patterns which focus on a similar topic together, even though the topic information is not used in the clustering process. For instance, actions highlighted with different color depth of blue showed they were actions within the topic of “Confidence Interval”. Figure 3 showed that these patterns were concentrated in the upper part of the results while Fig. 4 showed that patterns highlighted with blue were separated under alphabetical sort.

### 4.5 Sunburst Visualization

The patterns discovered by the SPM algorithm have a text format output. The lack of visual element would cause difficulties during the interpretation process of patterns and this would not be favorable to users such as teachers, who are technically less



sophisticated and could not afford to spend a large amount of time for analysis of the patterns. Visual analysis was found to be effective in this situation since it allows for an interactive exploration of the results and satisfaction of curiosity, which were both essential to exploratory analysis [4]. Furthermore, interactivity is an important element to the interpretation as users might spot interesting facts when the visual display changes which drive them to manipulate the visualization or the underlying data for further exploration of such changes [11]. Therefore, an effective visual representation of the discovered patterns would be necessary to facilitate the pattern interpretation process.

This study adopted the D3 sunburst visualization by Rodden [19] out of different available visualization packages for sequential data, mainly due to the interactive elements in it and the effectiveness in presenting the results from SPM. Similar visualizations were mostly designed for sites from industries such as e-commerce to understand the discovery of products by customers. The sunburst visualization was originally designed to display the webpage navigation sequences on YouTube to study the video discovery process of users and only focused on those sequences occurred at the start of a user’s visit to the website. However, the navigation patterns in the current study context are frequent sequences of visit, which may occur at different time points during user’s visit under the specified time gap. This major definitional difference to sequence leads to modifications in the package.

The visualization can display all sequences in a coherent manner which was concrete and easy to understand the patterns as reported by Rodden [19]. When the pattern discovery process was completed by the algorithm, the results could be visualized automatically. Figure 5 presents the sunburst visualization based on the results from the major group in the current study as an example. The legend is placed on the right hand

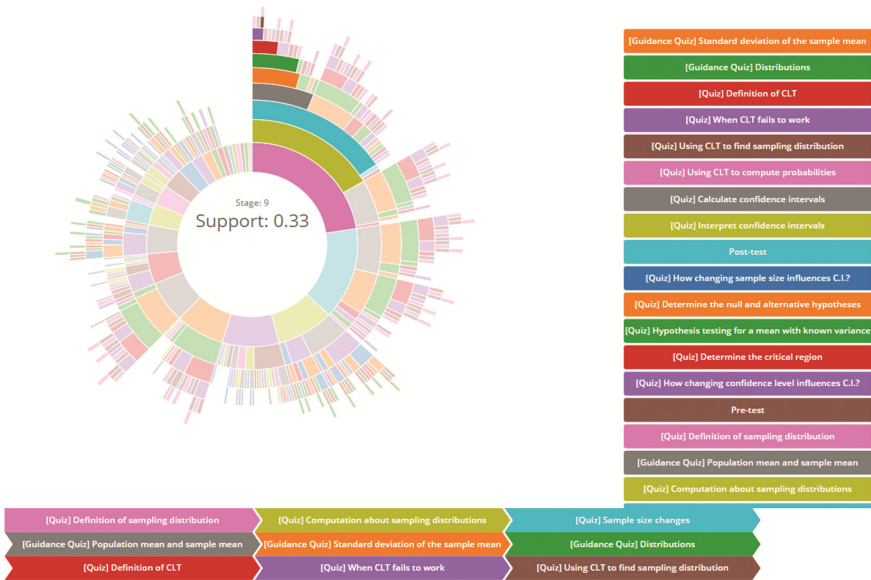


Fig. 5. Sunburst visualization of patterns.

side of interface and each action has its own corresponding color for easier differentiation. Using a mouse to hover any segment will highlight the entire associated pattern and desaturate all other segment in the visualization, with the display of support value at the center of the graph and detailed information about each action in the sequence below the graph. The size of the color block represents the proportion of total number of visit on that level of action. Comparing to the text-based output, the visualization could increase the effectiveness in delivering the results of discovered patterns from the mining process. For instance, teachers could easily find out which digital resources is the most popular among students and all other connected resources as revealed by the mined sequences. The graphical and interactive output also allows them to explore the results with higher flexibility and reduces the possible anxiety induced when facing a large quantity of information regarding the patterns. The sunburst visualization could make the analysis of navigational patterns easier to approach and use. More teachers could therefore be benefited to gain a more concrete understanding on the learning processes of students online.

## 5 Conclusion

This paper uses SPM to discover the navigational patterns of students on LMSs and presents the results through hierarchical clustering and sunburst visualization. The results are transformed into a more interactive and interpretable format for learning analytics, which aims to lower the difficulties in understanding the patterns for end users such as teachers. We hope that this paper could address the issues raised by Reimann et al. [18] and Gómez-Aguilar et al. [4] and promote further usage of learning data in a digital environment such as on LMSs, which contains valuable information of students which could facilitate the pedagogical decision making of teachers, especially with the prevalence of blended learning approach in instruction.

**Acknowledgement.** The study was funded by Teaching Development Grant (Ref: HKIED7/T&L/12-15) under the Hong Kong University Grants Committee.

## References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: ICDE (1995)
2. Ferguson, R.: Learning analytics: drivers, developments and challenges. *Int. J. Technol. Enhanced Learn.* **4**(5–6), 304–317 (2012)
3. Fournier-Viger, P., Nkambou, R., Nguifo, E.M.: A knowledge discovery framework for learning task models from user interactions in intelligent tutoring systems. In: MICAI 2008, pp. 765–778 (2008)
4. Gómez-Aguilar, D.A., Hernández-García, A., García-Peñalvo, F.J., Therón, R.: Tap into visual analysis of customization of grouping of activities in elearning. *Comput. Hum. Behav.* **47**, 60–67 (2015)
5. Graf, S., Liu, T.C.: Analysis of learners' navigational behaviour and their learning styles in an online course. *J. Comput. Assist. Learn.* **26**(2), 116–131 (2010)

6. Hirate, Y., Yamana, H.: Generalized sequential pattern mining with item intervals. *J. Comput.* **1**(3), 51–60 (2006)
7. Kang, J., Liu, M., Qu, W.: Using gameplay data to examine learning behavior patterns in a serious game. *Comput. Hum. Behav.* **72**, 757–770 (2016)
8. Kerr, D.: Visualizing changes in strategy use across attempts via state diagrams: a case study. *Int. J. Comput. Games Technol.* **2016**, 4 (2016)
9. Kong, S.C., Song, Y.: An experience of personalized learning hub initiative embedding BYOD for reflective engagement in higher education. *Comput. Educ.* **88**, 227–240 (2015)
10. Mahajan, R., Sodhi, J., Mahajan, V.: Usage patterns discovery from a web log in an Indian e-learning site: a case study. *Educ. Inf. Technol.* **21**(1), 123–148 (2016)
11. Mazza, R.: Visualization in educational environments. In: Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.D. (eds.) *Handbook of Educational Data Mining*, pp. 9–26 (2010)
12. Mooney, C.H., Roddick, J.F.: Sequential pattern mining — approaches and algorithms. *ACM Comput. Surv.* **45**(2), 19:1–19:39 (2013)
13. Poon, L.K.M., Kong, S.C., Wong, M., Yau, T.: Mining sequential patterns of students' access on learning management system. In: *The Second International Conference on Data Mining and Big Data (DMBD 2017)* (2017)
14. Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H.: Prefixspan: mining sequential patterns efficiently by prefix-projected pattern growth. In: *ICDE* (2001)
15. Pei, J., Han, J., Wang, W.: Constraint-based sequential pattern mining: the pattern-growth methods. *J. Intell. Inf. Syst.* **28**, 133–160 (2007)
16. Perera, D., Kay, J., Koprinska, I., Yacef, K., Zaiane, O.R.: Clustering and sequential pattern mining of online collaborative learning data. *IEEE Trans. Knowl. Data Eng.* **21**(6), 759–772 (2009)
17. Psaromiligkos, Y., Orfanidou, M., Kytagiass, C., Zafiri, E.: Mining log data for the analysis of learners behaviour in web-based learning management systems. *Oper. Res.* **11**(2), 187–200 (2011)
18. Reimann, P., Hesse, F., Freebody, P., Cierniak, G., Imhof, B., Wasson, B., Hansen, C., Utz, W., Vatraru, R., Bull, S., et al.: Supporting teachers in capturing and analyzing data in the technology-rich classroom. In: *Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012)*, pp. 33–40. International Society of the Learning Sciences (2012)
19. Rodden, K.: Applying a sunburst visualization to summarize user navigation sequences. *IEEE Comput. Graph. Appl.* **34**(5), 36–40 (2014)
20. Srivastava, J., Cooley, R., Deshpande, M., Tan, P.N.: Web usage mining: discovery and applications of usage patterns from web data. *ACM SIGKDD Explor. Newsl.* **1**(2), 12–23 (2000)
21. Van Barneveld, A., Arnold, K.E., Campbell, J.P.: Analytics in higher education: establishing a common language. *EDUCAUSE Learn. Initiative* **1**(1), 1–11 (2012)
22. Zhou, M.: Using traces to investigate self-regulatory activities: a study of self-regulation and achievement goal profiles in the context of web search for academic tasks. *J. Cogn. Educ. Psychol.* **12**(3), 287–305 (2013)
23. Zhou, M., Xu, Y., Nesbit, J.C., Winne, P.H.: Sequential pattern analysis of learning logs: methodology and applications. In: *Handbook of Educational Data Mining*, pp. 107–121 (2010)
24. Ziebarth, S., Chounta, I.A., Hoppe, H.U.: Resource access patterns in exam preparation activities. In: *Design for Teaching and Learning in a Networked World*, pp. 497–502. Springer (2015)