Sequential Pattern Mining Model to Identify the Most Important or Difficult Learning Topics via Mobile Technologies

Edona Doko(✉), Lejla Abazi Bexheti, Mentor Hamiti, Blerta Prevalla Etemi
SEE University, Tetovo, Macedonia
`ed15197@seeu.edu.mk`

Abstract—The paper aim is to come up with methodology for performing video learning data history of learner's video watching logs, video segments or time series data in accordance with learning processes via mobile technologies. To reach this goal, it is introduced a theoretical method of sequential pattern mining specialized for learning histories in identifying the most important or difficult learning. Based on this method, it is designed a model for understanding and learning the most difficult topics of students topics. The user will be able to use and access the model through mobile technologies when and where he/she wants. The performed video learning history data of learner's video watching logs consists of functions that are responsible for collection of stop/replay/backward data activities, generation of sequence from the collected learning histories, extraction of important patterns from a set of sequences, and findings of learner's most difficult/important topic from the extracted patterns. The paper mainly describes the model for understanding and learning the most difficult topics through the sequential pattern mining method. Implementing the method to use in mobile phones is considered as future aim.

Keywords—Sequential Pattern Mining (SPM); Video; Learning; Most Important/Difficult Learning Topics (MIDLT); Mobile

1   Introduction

Providing students high quality, stable technology and rich environments is the challenge in improved accessibility and enhanced applications embedded in emerging mobile technologies. The late progress in technology and ideology has opened a completely new direction for education research. The areas of Learning Analytics and Education Data Mining explore the use of data to increase insight about learning environments and improve the overall quality of experience for students. Mobile technologies in learning systems are influenced from several areas to research and build models of educational data mining and learning analytics [1, 21, 24]. Smart mobile technologies, such as tablet computers and smartphones, offer advanced computing abilities as well as access to internet-based resources without the constraints of time or place. This has resulted in devices that enable the provision of ubiquitous learning

environments that combine real-world and digital world resources [19, 20, 22]. The research on this paper is focused on analyzing EDM and LA articles to come up with better learning methods. Technical methods used in EDM technologies have big influence in the realization of better learning strategies. The methodology used and the defined research question is introduced first. Secondly, the proposed method and model is introduced in the result section. As method to reach the goal is proposed sequential pattern mining approach of performing video data history. In addition, is defined the method to identify the most important/difficult learning topics (MIDLT). In the discussion part are discussed and compared some research solutions related to the research field. The EDM and LA methods are illustrated with examples.

2   Methodology

The main goal of the paper is to enhance student learning through learning methods and define model to identify the most important/difficult learning topics (MIDLT) for students. There are selected 122 papers to analyze from 285 searching string results.
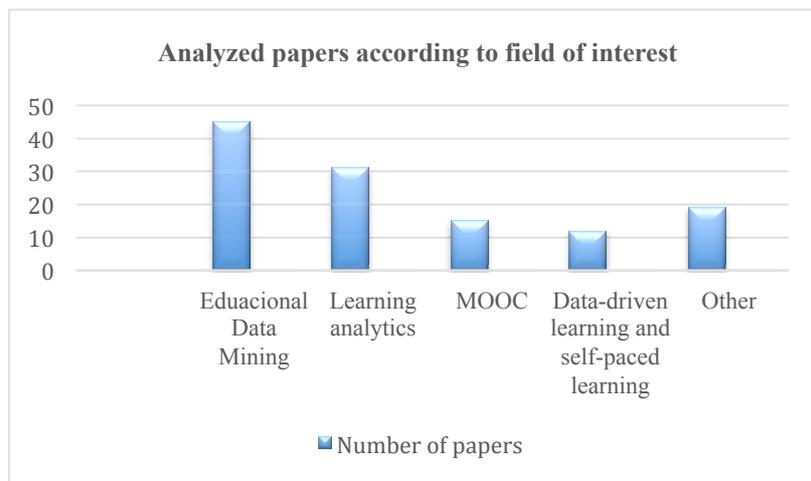

Fig. 1. Number of paper analyzed according to the video field of interest

First are collected all publications needed for the interested field. Secondly is performed the conduct search for primary studies, and are excluded the studies that are not relevant to answer. Here is defined the hypothesis to drive the structure. The third step is ensuring that the method takes the existing studies into account and providing better results. After analyzing solutions and papers of the selected research papers, the aim of the research is proposed in solution.

The papers includes EDM and LA topics containing different methods, models and algorithms in order to define the aim for enhancing student learning interactivity.

The paper aim to propose a model and system architecture of learners video watching logs, video segments or time series data that identifies the most important/difficult learning topics (MIDLT).
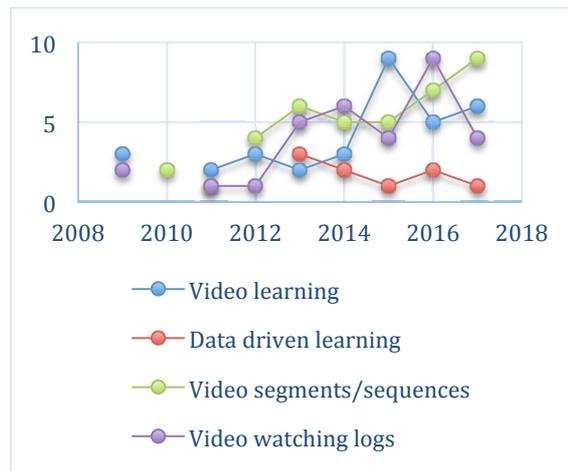

Fig. 2. Time series of papers according to video field of interest

A lot of the selected papers elaborate different video field of interest over years as showed in Fig. 2. [18] In the selected papers the researches uses five different categories of technical methods:

- Prediction
- Clustering
- Relationship mining
    - Association rule mining
    - Correlation matrix
    - Sequential pattern mining
    - Causal data mining
- Distillation for human judgment
- Discovery with models

The defined goal of this paper provides method extraction and complete the aim of the study, by identifying, analyzing and interpreting the relevant evidence. The sequential pattern mining method is introduced and the results are divided in two main topics to fulfil the goal:

- Sequential pattern mining model of performing video learning data history
- Results and System Architecture to identify most important/difficult learning topics (MIDLT).

3 Results

By using Web Usage Mining techniques, the E-learning environment can become more effective. Based on Web Usage Mining technique can designed effective E-learning environment and implementation of the model builds more intelligent system and can promote interest of learners or learning environment is modifying its behavior to reflect learning style by using mobile technologies. [15]
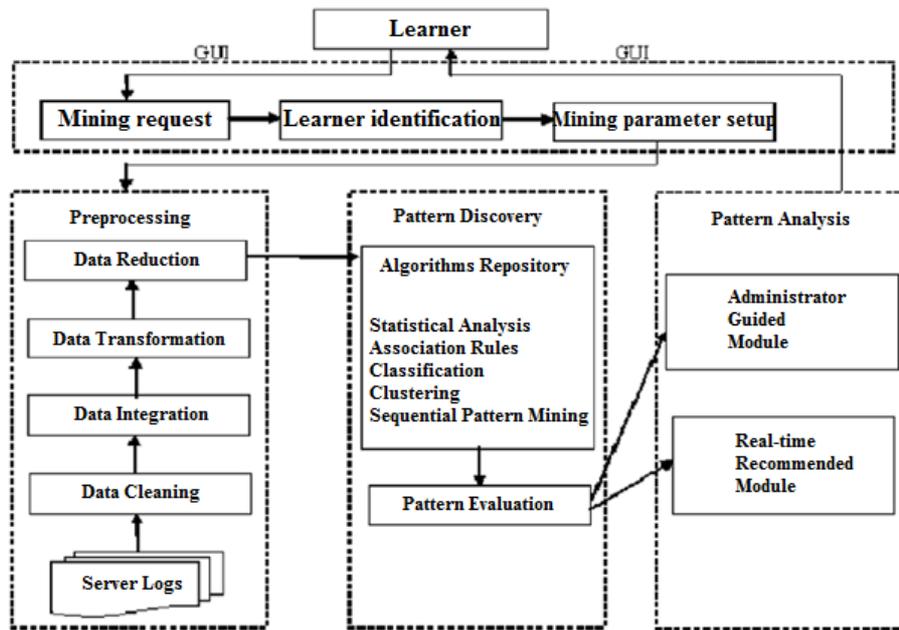

Fig. 3. Implementation of E-Learning model

3.1 Sequential pattern mining model of performing video learning data history

It is introduced a data mining method for analyzing relevancies between learning processes and learning situations used via mobile technologies is introduced. In addition, a sequential pattern mining system specialized of performing learning history data of video watching logs, video segments or time series data is presented. Understanding the learning situations of learner's video watching logs through learning history data of activities is important to fulfill the aims of the paper.

Sequential pattern mining discovers frequent sub sequences as patterns in a sequence database. A sequence database stores a number of records, where all records are sequences of ordered events, with or without concrete notions of time. An example sequence database is the watching sequence in learner's video, for each student, the collection of store topic/keyword of the video that they behave through

stop/replay/ backward every time that they watch it. These sequences of the student video watching can be represented as records with schema [Keyword of the Topic/Student ID, <Ordered Sequence Events>], where each sequence event is a set of store keyword of topic like XHTML, XML, JavaScript, JSON, and so on.

For only two such student, an example video watching sequential database is [KT1,<(XHTML, JSON), (XHTML, JSON, XML), (JSON), (JavaScript, XML)>]; [KT2, <(XHTML), (XML, JavaScript)>]. While watching the video the first student will stop/replay/backward more often in topics, with a Keyword of the topic ID shown as KT1 in the example, the second student during the watching of the video will stop/replay/backward more often on topics represented by KT2. In addition, a student can stop/replay/backward on one or more videos during each login. Thus, records in a sequence database can have different lengths, and each event in a sequence can have one or more keyword of topics in its set. Storing this data and track the logs of the student will lead to counting and ranking the most stop/replay/backward sequences containing the keyword topics.

Each element of the pattern can be contained in the union of the keywords bought in a set of user watching logs, as long as the difference between the maximum and minimum user watching log-times is less than the size of a sliding time window. [2]

The GSP (Generalized Sequential Patterns), algorithm [3] discovers new algorithm considering time constraints, sliding time windows, and taxonomies in sequential patterns. Empirical evaluation shows that GSP scales linearly with the number of data-sequences in video watching keyword topics, and has very good scale-up properties with respect to the number of watching logs per data-sequence and number of keyword topic stop/replay/backward per student.
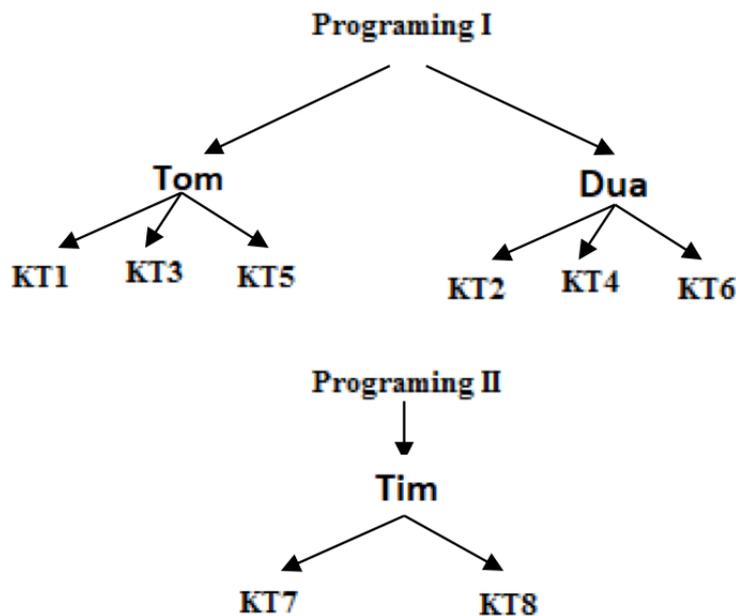

Fig. 4. Example of a Taxonomy

3.2 Results and System Architecture to identify the most important/difficult learning topics (MIDLT) via mobile technologies

To identify the most important or difficult learning topic it is focused on the contents of video learning keyword topics and define "Learning Situation KT1" as the circumstances of activities from the time students complete watching the video KT1. Situation KT1 includes keywords of the topics that represents to the video sequence KT1. The "Learning process" is defined as the transition of situations that can gradually change according to activities such as stop, pause, replay, backward in a learning situations (watching logs, video segments) (middle part of Fig.5)

To understand learning situations of the student KT of watching logs and video segments based on learning processes based on activities such as stop, pause, replay or backward it is required to clarify the relationships between learning processes and learning situations via mobile technologies. For instance, students who have experienced a specific learning process are likely to have fallen to a particular situation as are keyword topics KT1, KT2 etc. Clarification of such relationships enables teachers to easily estimate new students' learning situations from their learning processes in mobile technologies. In addition, the most significant is that the learner will identify the most important/difficult learning topics (MIDLT) in order to improve, understand and learn the most difficult ranking topics.
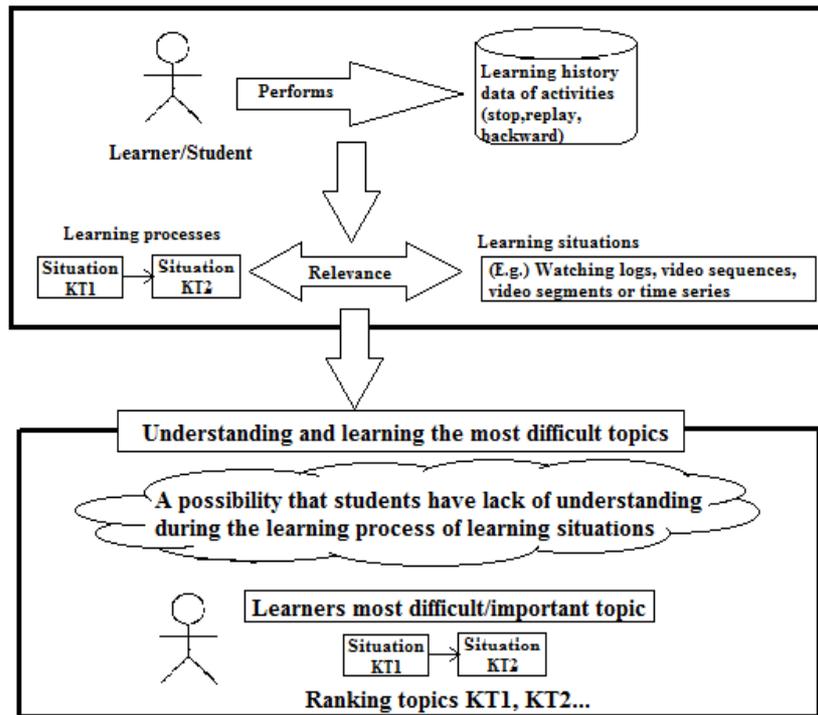

Fig. 5. Model understanding and learning the most difficult topics

The method of the relationships between learning processes and situations via mobile technologies is determined in the model of Fig.5. Therefore, are defined methods for analyzing learning history data considering learning processes in order to make learning easier for learners (upper part of Fig.5). Our approach enables to easily understand learning situations based on learning process as a key to improve learning process of students accessing them everywhere at any time via mobile technologies (lower part of Fig.5). In this way, students will seek to improve their self in difficult topics by self-learning in order to provide learning enhancements.

According to the analyzed papers for system stability, flexibility and enhancing of student learning interactivity is proposed the MIDLT system architecture. The overall system architecture is illustrated in Figure 6; where learning system provides a Web-based user interface together with the main database, and the MIDLT is responsible for collecting the log data of all learning activities.
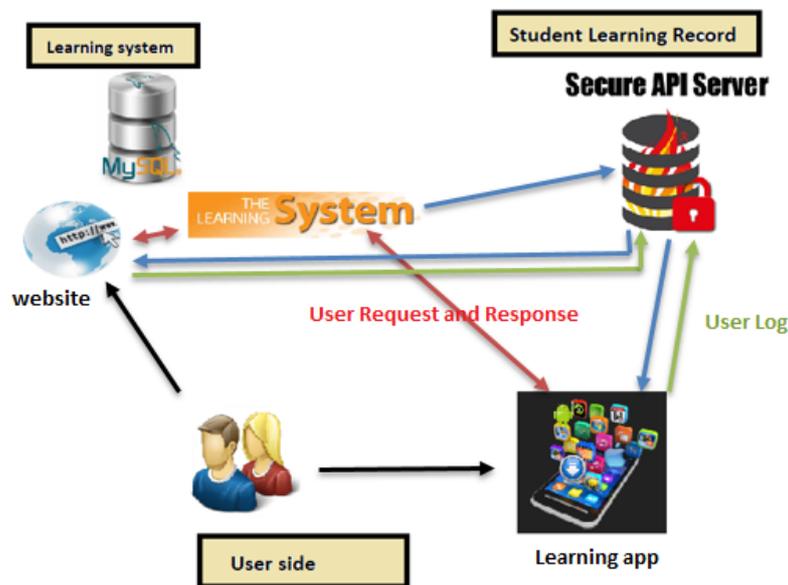

Fig. 6. System Architecture of the MIDLT as part of the learning system to enhance student learning interactivity

The system should be implemented as Web-based system to ensure the OS independence since should consider virtual machine environments on which learners conducts the exercises and educators' PCs. All the interactions can be performed by using learning applications via mobile technologies.

4   Discussion

After analyzing solutions and papers of the selected research papers the aim of the research is defined and answered. There are many researches that uses different categories of technical methods. To identify the most important/difficult learning topics (MIDLT) for students via videos it is used the sequential pattern mining method and model as solution to answer the defined research goals. The sequential pattern mining is proposed as an important problem with broad applications, including the analysis of student behavior, web access patterns, scientific experiments etc. Furthermore, researchers are using the method for customer purchase behavior, disease treatment, natural disasters and protein formations. A sequential pattern-mining algorithm mines the sequence database looking for repeating patterns (known as frequent sequences) that can be used by end users or students to find associations between the different topics or keywords in their data for purposes such as reports of the most important/difficult learning topics (MIDLT), learning enhancement, prediction and planning. With the increase in the use of the World Wide Web for different approaches, web usage mining is one of the most dominant application areas of sequential pattern mining [4, 23].

In sequential pattern mining, the goal is to find temporal associations between events. Two paradigms are seen that find sequential patterns -- classical sequential pattern mining [3], which is a special case of association rule mining. The methods, like association rule mining, have been used for a variety of applications including to study what paths in student collaboration behaviors leads to a more successful eventual group project [5], the patterns in help-seeking behaviour over time, and studying which patterns in the use of concept maps are associated with better overall learning [6][7]. Sequential pattern mining algorithms, like association rule mining algorithms, depend on a number of parameters to select which rules are worth outputting [8].

Searching forward in models and methods there were different solutions for different methods.

The areas of LA and EDM explore the use of data to increase insight about learning environments and improve the overall quality of student's experience. Learning systems are influenced from several areas to research and build models of educational data mining and learning analytics. Higher education institutions are beginning to use analytics for improving the services they provide and for increasing student grades and retention. Educational data mining tends to focus on developing new tools for discovering patterns in data [1]. Data mining is the process of automatically discovering useful information in large data repositories. It is an integral part of Knowledge Discovery in Databases (KDD), which is the overall process of converting a series of transformation steps, from data pre-processing to post-processing of data mining [16]. A key application of learning analytics is monitoring and predicting students' learning performance and spotting potential issues early so that interventions can be provided to identify students at risk of failing a course or program of study [1].

Learning analytics has a relatively greater focus on human interpretation of data and visualization. Discovery with models, and data retrieval for human judgment are useful in the field of Educational Data Mining. It provides novel possibilities for

gathering, analyzing, and presenting data. These new data sources can be utilized as guides for course redesign or as an evidence for implementing new assessment approaches [9].

The educational data mining model is built for the completion of specific mining tasks and is an integrated application of a variety of data mining tools and algorithms. It consists of "data mining work", "tools and algorithms" and "data" three elements. "Tools and Algorithms" mainly support the work of data mining, and produce the corresponding "data". The expansion of these three elements in time will respectively form data mining workflows, tools, algorithm flows, and data streams. The data mining workflow includes data collection, data preprocessing, data mining, pattern interpretation and application and so on.

Regardless of the type and subject of analysis, the ultimate objective of learning analytics must be to enable data-driven educational decision making at all levels [17]. Kim et al explored the design space of data-driven interaction techniques for educational video navigation.

4.1 LA and EDM methods

The most popular methods of LA and EDM are:

1. Classification
2. Clustering
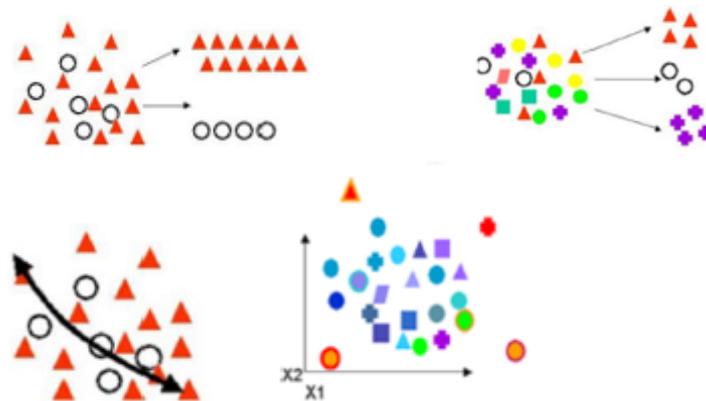3. Regression (logistic/multiple)
4. Discovery with models

Fig. 7. LA and EDM methods

Educational data mining research uses the five categories of technical methods:

Prediction entails developing a model that can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables). Examples of using prediction include detecting such student behaviors as when they are gaming the system, engaging in off-task behavior, or failing to answer

a question correctly despite having a skill. Predictive models have been used for understanding what behaviors in an online learning environment—participation in discussion forums, taking practice tests and the like—will predict which students might fail a class. Prediction shows promise in developing domain models, such as connecting procedures or facts with the specific sequence and amount of practice items that best teach them, and forecasting and understanding student educational outcomes, such as success on posttests after tutoring.

Clustering refers to finding data points that naturally group together and can be used to split a full dataset into categories. Examples of clustering applications are grouping students based on their learning difficulties and interaction patterns, such as how and how much they use tools in a learning management system and grouping users for purposes of recommending actions and resources to similar users. Data as varied as online learning resources, student cognitive interviews, and postings in discussion forums can be analyzed using techniques for working with unstructured data to extract characteristics of the data and then clustering the results. Clustering can be used in any domain that involves classifying, even to determine how much collaboration users exhibit based on postings in discussion forums.

Relationship mining involves discovering relationships between variables in a dataset and encoding them as rules for later use. For example, relationship mining can identify the relationships among products purchased in online shopping.

- Association rule mining can be used for finding student mistakes that co-occur, associating content with user types to build recommendations for content that is likely to be interesting, or for making changes to teaching approaches. These techniques can be used to associate student activity, in a learning management system or discussion forums, with student grades or to investigate such questions as why students' use of practice tests decreases over a semester of study.
- Correlation matrix
- Sequential pattern mining builds rules that capture the connections between occurrences of sequential events, for example, finding temporal sequences, such as student mistakes followed by help seeking. This could be used to detect events, such as students regressing to making errors in mechanics when they are writing with more complex and critical thinking techniques, and to analyze interactions in online discussion forums.
- Causal data mining

Key educational applications of relationship mining include discovery of associations between student performance and course sequences and discovering which pedagogical strategies lead to more effective or robust learning. This latter area—called teaching analytics—is of growing importance and is intended to help researchers build automated systems that model how effective teachers operate by mining their use of educational systems.

Distillation for human judgment is a technique that involves depicting data in a way that enables a human to quickly identify or classify features of the data. This area of educational data mining improves machine-learning models because humans can identify patterns in, or features of, student learning actions, student behaviors, or data

involving collaboration among students. This approach overlaps with visual data analytics (described in the third part of this section).

Discovery with models is a technique that involves using a validated model of a phenomenon (developed through prediction, clustering, or manual knowledge engineering) as a component in further analysis.

Technical methods used in learning analytics are varied and draw from those used in educational data mining. Additionally, learning analytics may employ social network analysis and social or "attention" metadata to determine what a user is engaged with.

El-Halees in his case study in educational data mining illustrated how useful data mining can be in higher education in predominantly to progress student routine. They used students' data from database course. They collected all accessible data together with their usage of Moodle e-learning facility and discovered association rules and arranged the rules using lift metric then they envisage the rules. Also there was exposed classification rules using decision tree and clustered the student into group using EM clustering. Finally, using outlier examination were detected all outliers in the data. Each one of this acquaintance can be used to recover the performance of student. Also, experiments could be done using more data mining techniques and data mining algorithms could be surrounded into e-learning system so that anyone using the scheme can benefit from the data mining techniques.

Based on learner behavioral patterns in Shabandar et al. 2017 paper are presented two sets of features [10]. There were compared in terms of their suitability for predicting the course outcome of learners participating in MOOCs. The Exploratory Data Analysis demonstrates that there is strong correlation between click stream actions and successful learner outcomes. The correlation matrix was applied to measure the dependency between the behavioral data and learners' certification. Figure - shows the plot of correlation matrix. It indicates a positive relationship between three behavioral attributes and the target variable.

Bell et al has describe a system for capturing time-varying sequences of behavioural patterns of agents or robots from a video clips. The detailed time slices of action are 'coarsened' to provide gross, molecular units of behaviour. Combinations of these are then represented in the form of a table, which is mined using Rough sets methods here, although various other mining techniques can also be used [11].

The traditional algorithm for mining sequential pattern from web log data is used by Swati Singh Lodhi in 2014. [12] His aim of the present work is to bridge these gaps by developing and proposing a new algorithm "Sequential ID3" for sequential pattern mining and their experimental validation on web log data. Quality data provide us with the transparency needed to understand what's happening in classrooms, schools, and entire education system. [13] In Moodle system logs are maintained by "log", "log_display" and log_queries. The log data are stored in the "log" table and two other tables are used to display the data through the system reports. The "log" table in a Moodle database records entries of each user interaction with the system. It connects with the user table, course table, course module table, resource and the activities tables such as quiz, forum, assignment, etc. to store all the log details of the system for each user interaction. [14]
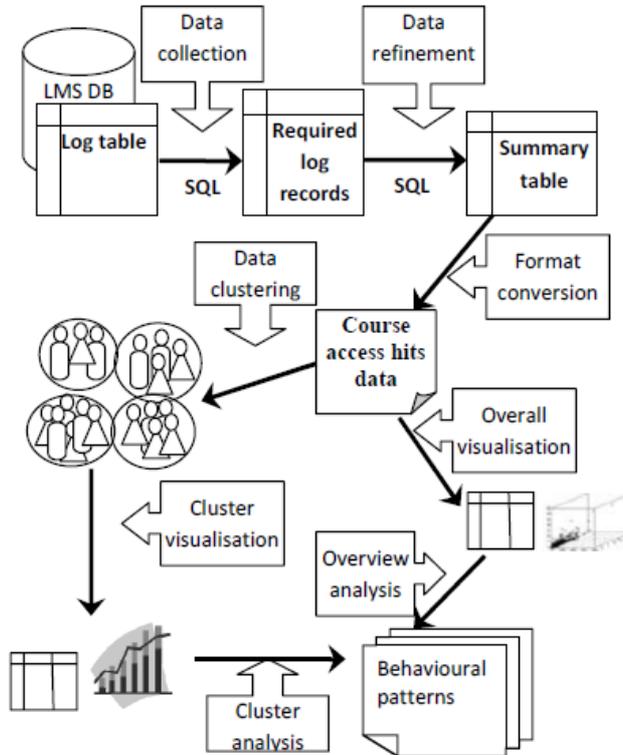
Fig. 8. The process of data collection, preparation and analysis

Educational Data Mining is capable of detecting system usage behaviors using data mining techniques. The clustering technique can be used to characterize the learner's behavior and group them based on the behavioral similarities. Clustering is one of the data mining techniques, which can be used to discover the new categories, which share the similar interest. In clustering the similar instances are grouped together. Different clustering algorithms can be used to separate the instances into given number of clusters. K-Means algorithm is generally used to divide learners into natural groups based on their behavior for a larger dataset. [14]

5   Conclusion

According to the analyzed literature, it is proposed a model and system architecture for learning history analysis based on learning processes to identify the most important/difficult learning topics (MIDLT) of students via mobile technologies. First, it is introduced an overview of a sequential pattern mining method for video learning data histories in programming exercise classes. Then it is defined a model and system

architecture for analyzing the programming learning history based on the mobile technologies proposed method.

Future work includes improving the proposed method and propose recommendation system based on rule-space model through further examinations using practical learning histories.

6   References

[1] Prakash, Hanumanthappa and Kavitha (2014), Big Data in Educational Data Mining and Learning Analytics, IJIRCCE.

[2] Nakamura, Nozaki, Nakayama, Morimoto and Miyadera (2015), Sequential Pattern Mining System for Analysis of Programming Learning History, IEEE.

[3] Ramakrishnan Srikant and Rakesh Agrawal, Mining Sequential Patterns: Generalizations and Performance Improvements, IBM Almaden Research Center, CA 95120.

[4] Nizar R. Mabroukeh and C. I. Ezeife (2010), A Taxonomy of Sequential Pattern Mining Algorithms, ACM.

[5] Perera D, Kay J, Koprinska I, Yacef K, Zaïane, Clustering and sequential pattern mining of online collaborative learning data (2009), Knowledge and data engineering, IEEE.

[6] Shanabrook DH, Cooper DG, Woolf BP, Arroyo I (2010), Identifying high-level student behavior using sequence-based motif discovery, In: Proceedings of the 3rd international conference on educational data mining.

[7] Kinnebrew J, Biswas G (2012), Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution, In: Proceedings of the 5th international conference on educational data mining.

[8] Baker and Inventado (2014), Educational Data Mining and Learning Analytics, SPRINGER.

[9] D. Gandhimathi and S. Gomathi (2015), "A Survey of Approaches and Tools Used in Educational Data Mining", IJIRCCE.

[10] Shabandar, Hussain, Laws, Keight, Lunn and Radi (2017), Machine Learning Approaches to Predict Learning Outcomes in Massive Open Online Courses, IEEE.

[11] Bell, Beck, Miller and Herrera (2007), Video Mining - Learning Patterns of Behaviour via an Intelligent Image Analysis System, IEEE.

[12] Swati Singh Lodhi (2014), Development of Sequential ID3: "An advance Sequential mining Algorithm", AJSE.

[13] Nakamura, Nozaki, Morimoto and Miyader (2014), Sequential Pattern Mining Method for Analysis of Programming Learning History Based on the Learning Process, IEEE.

[14] Ratnapala and Deegalla (2014), Students Behavioural Analysis in an Online Learning Environment Using Data Mining, IEEE.

[15] Yadav and Jain (2011), Analyses of Web Usage Mining Techniques To Enhance the Capabilities of E-Learning Environment, IEEE.

[16] Banu and Ravanan, Analysis of E-learning in Data Mining – A Dreamed Vision for Empowering Rural Students in India, IEEE, 2011.

[17] Conde and García, Learning analytics for educational decision making, ELSEVIER, 2015.

[18] Edona Doko and Lejla Abazi Bexheti,, Systematic mapping study of educational technologies based on Educational Data Mining and Learning analytics, MECO, 2018.

[19] Nejati, E., Jahangiri, A. & Salehi, M. R. (2018). The effect of using computer-assisted language learning (CALL) on Iranian EFL learners' vocabulary learning: An experimental study. *Cypriot Journal of Educational Sciences, 13*(2), 351-362. https://doi.org/10.18844/cjes.v13i2.752

[20] Soykan, E. & Ozdamli, F. (2016). The Impact of M-Learning Activities on the IT Success and M-Learning Capabilities of the Special Education Teacher Candidates. *World Journal on Educational Technology: Current Issues, 8*(3), 267-276. https://doi.org/10.18844/wjet.v8i3.1019

[21] Bahadir, C. & Karahoca, A. (2017). Airline revenue management via data mining. *Global Journal of Information Technology: Emerging Technologies, 7*(3), 128-148.

[22] Stosic, L. (2017). Does the use of ICT enable easier, faster and better acquiring of knowledge? *International Journal of Innovative Research in Education. 4*(4), 179–185.

[23] Abdugulova, Z. (2017). Allowing schools access to affordable computers: How schools can benefit from switching to inexpensive, cloud-based computing technologies. *International Journal of Learning and Teaching. 9*(3), 326-331. https://doi.org/10.18844/ijlt.v9i3.507

[24] Uzunboylu, H., & Tugun, V. (2016). Validity and Reliability of Tablet Supported Education Attitude and Usability Scale. *Journal of Universal Computer Science, 22*(1), 82-93.

7 Authors

Edona Doko is PhD candidate in Computer Sciences at South East European University in Macedonia. Her main research fields are in Educational Data Mining, Learning Analytics and Flipped Classroom. Since 2016, she is Software QA testing lead in Support&Solution Centre in Skopje, Macedonia part of Katoen Natie, Belgium. From 2013, she has experience as teaching professor at AAB University in Kosovo, FON University in Skopje and other training activities in QA testing. The author can be contacted at ed15197@seeu.edu.mk.

Lejla Abazi Bexheti is Associate Professor at the Faculty of Contemporary Sciences and Technologies at South East European University in Macedonia. She holds a PhD Degree in Computer Science and has been part of the CST teaching staff since 2002. Her main research activity is in the area of Learning Systems and eLearning and she has been involved in many international projects and research activities from this area. At SEE University she was involved on resolving issue of the Learning Management System. Currently she is Pro-rector for academic issues at SEEU. The author can be contacted at l.abazi@seeu.edu.mk.

Mentor Hamiti, PhD, received a PhD degree in Computer Sciences from South East European University, Tetovo, Macedonia in 2010. He is currently associate professor in Computer Sciences at the South East European University, Faculty of Contemporary Sciences and Technologies. His research focuses on Programming Languages and Technologies. The author can be contacted at m.hamiti@seeu.edu.mk.

Blerta Prevalla is a PHD candidate at Near East University, Cyprus on Computer Education and Instructional Technology. She is teaching at AAB University, Prishtina, Kosovo the following subjects: Software Engineering, Programming Fundamentals, Object Oriented Programming etc. The author can be contacted at blerta.prevalla@universitetiaab.com.