Sequential Pattern Mining Model of Performing Video Learning History Data to Extract the Most Difficult Learning Subjects

Edona Doko
SEE University, Macedonia
Lejla Abazi Bexheti
SEE University, Macedonia
Visar Shehu
SEE University, Macedonia

Abstract

The paper aim is to define a method for performing video learning data history of learner's video watching logs, video segments or time series data in consistency with learning processes. To achieve this aim, a theoretical method is introduced. Sequential pattern mining with learning histories are used to extract the most difficult learning subjects. Based on this method, it is designed a model for understanding and learning the most difficult topics of students. The performed video learning history data of learner's video watching logs makeup of stop/replay/backward data activities functions. They correspond as output of sequence of the learning histories, extraction of significant patterns by a set of sequences, and findings of learner's most difficult/important topic from the extracted patterns. The paper mostly aim to devise the model for understanding and learning the most difficult topics through method of mining sequential pattern.

Introduction

Learning Analytics and Education Data Mining look into the usage of data to rise the penetration of learning environments and elevate the general quality of experience for students. Learning systems are affected by certain domains to research and build models of educational data mining and learning analytics (Prakash et al., 2014). The research on this paper is focused on EDM and LA analyzes of papers to stand uot with improved learning methods. First, it is presented the methodology and the research question is defined. 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 extract the most difficult learning subject. In the discussion part are discussed and compared some research solutions associated to the research field. The educational data-mining model in general 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 deliver the responsible "data". The flow of working of data mining includes data collection, preprocessing of data, pattern interpretation, application, data mining and so on.

Methodology

This paper aim to find learning method and define model to extract the most difficult learning subjects for students. There are selected 45 papers from 151 searching results. The papers include EDM and LA topics containing different methods, models and algorithms in order to answer the research question:

- What technical method should use to define a model for understanding and learning the most difficult topics of students and defining a model according to it?

All publications related to the field are gathered. Secondly is performed the behavior of the research for first studies. Some studies that were not considered as important for the related field are left out and were not answered. To ride the structure of the paper the research question is provided. To make sure that the method consists of the current studies and provide improvements of the results, it is considered as the third step. After analyzing solutions and papers of the selected research papers, the aim of the research is defined and answered. In the selected paper the researches uses five different categories of technical methods:

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

The sequential pattern mining method is introduced as the answer of the research question as the most reasonable topic to work on. The results are divided in two main topics to fulfil the goal:

- Sequential pattern mining model of performing video learning data history
- Results to extract the most difficult learning subjects

Results

Sequential pattern mining model of performing video learning data history

We introduce the method of data mining to analyze relations among learning processes and learning situations. We also present a the system of sequential pattern mining adjusted to perform the data of learning history in video watching logs, video segments or time series data. To verify the hypothesis of this paper need to make it feasible supporting the meaning of learning situations of learner's video watching logs through learning history data of activities.

Continuous sub sequences are detected by sequential pattern mining like patterns in a sequence database. Records are stored by a sequence database. With or without real notions of time the records are sequences of ordered events, An sequence database sample 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>], each sequence event is a set of store keyword of topic like XHTML, XML, JavaScript, JSON, etc.

An sample video watching sequential database for two students 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/more videos throughout every login. Diverse lengths of records can be in a sequence database. Every event in a sequence can have one/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.

Time constraint lack – Maximum/minimum time gaps among connected elements of the sequential pattern are usually specified by users. An sample is when a teacher apparently does not take attention if someone watched the video of keyword of topic "KT1", followed by "KT3" one year before. The teacher may have a desire to describe that a student should support a sequential pattern just when close elements happen inside a appointed time interval, let's say for four months. (Hence, supporting this pattern of a student, he/she should have watched the videos of the "KT3" within four months of the watched previously "KT1".)

A *user watching logs* solid definition– For most applications, it is not important if keywords of topics in an element of a sequential pattern were given in two different user-watching logs. This is valid as long as the user watching times of those user-watching logs are inside some small time window. The union of the keywords bought in a set of user watching logs includes every element of the pattern. It is like that as long as the distinction among the maximum/minimum user watching log-times is less than the size of a sliding time window (Nakamura et al., 2015). An sample can be if the teacher describes a time window of four months, a student who watched the video "KT1" on Monday, "KT2" on Saturday, and then "KT3" and "KT4" in a single day six month afterwards can still support the pattern "KT1" and "KT2", pursued by "KT3" and "KT4".

Lack of taxonomies – Users are searching topics through different levels of the taxonomy to find patterns. A lot of datasets have a user-defined taxonomy above data of the keyword topics (Srikant et al., 1996). A taxonomy sample is given in Figure above. Student that watched "KT1" pursued by "KT7" would support the patterns "KT1" pursued by "KT7", "Sara" pursued by "KT7", "Programing I" pursued by "Diell", etc.

The GSP (Generalized Sequential Patterns), algorithm (Srikant et al., 1996) discovers new algorithm considering the constraints of time, taxonomies and sliding time windows in sequential patterns. Numerous data-sequences in video watching keyword topics together with GSP scales are indicated by empirical evaluation. In addition there are good scale-up properties regard to the number of watching logs per data-sequence and number of keyword topic stop/replay/backward per student.

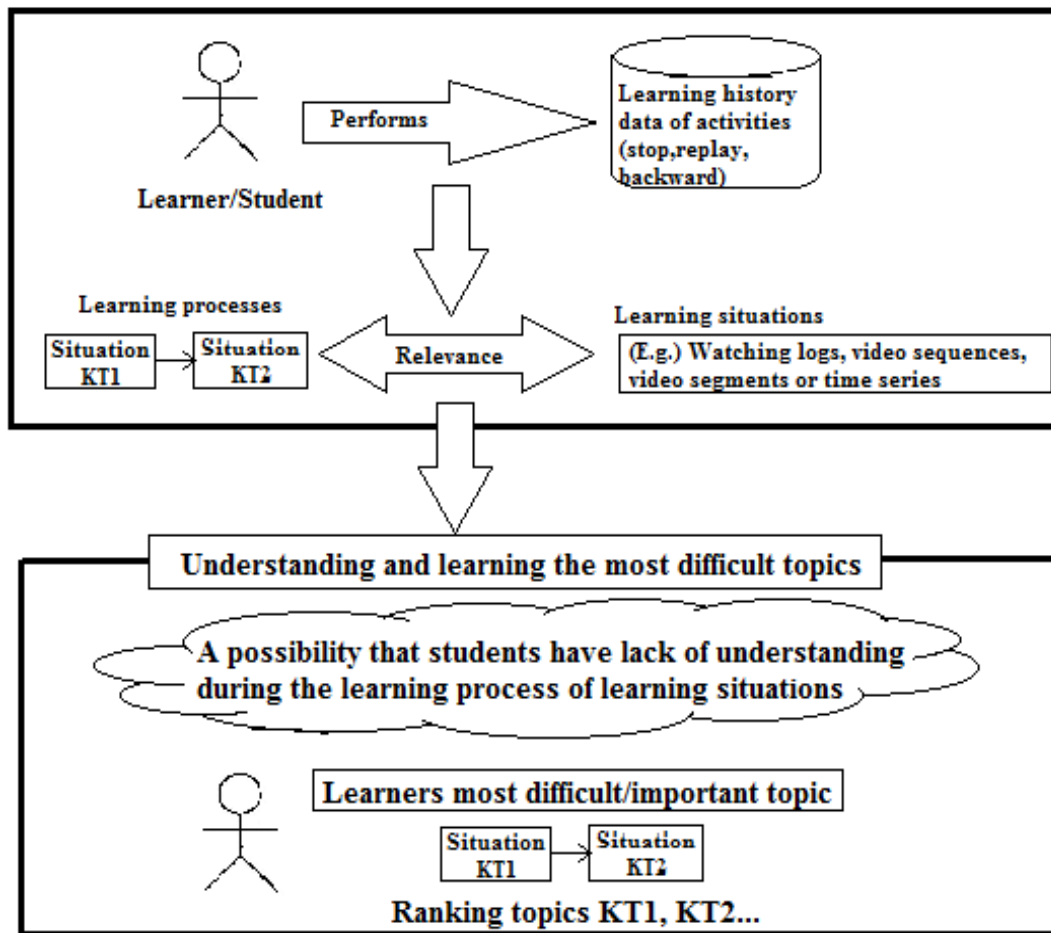*Figure 1*
Example of a Taxonomy
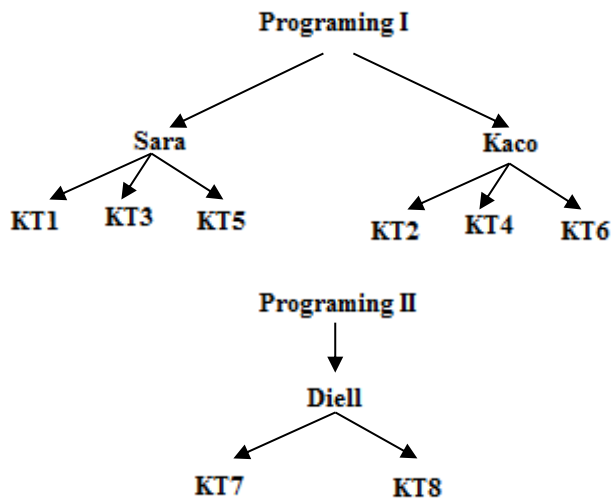
Source: Authors' work

Results to extract the most difficult learning subjects

To extract the most difficult learning subject it is focused on the contents of video learning keyword topics. By the time that students complete watching the video KT1, the "Learning Situation KT1" are defined as the conditions of activities. Situation KT1 includes keywords of the topics that represents to the video sequence KT1. "Learning process" are defined as transitions of situations that can slowly change according to activities such as stop, pause, replay, backward in a learning situations (watching logs, video segments) (middle part of Fig.2).

Learning situations of the student KT of watching logs/video segments are based on learning processes and activities such as stop, pause, replay or backward. Understanding them it is required to clear up the relations among learning processes and learning situations. For example, keyword topics KT1, KT2 or other can fall to specific situation. This is the case for students who have experienced a distinct learning process. The learning situations of new students are easily evaluated with the explanation of such relationships. In addition, the most significant is that the learner will extract the most difficult learning subjects in order to improve, understand and learn the most difficult ranking topics. The method of the relations among situations and learning processes is determined in the model of Fig.2. Hence, the methods to analyse learning history data are considered for learning processes to make learning easier for learners (Top of Fig.2). Improving learning process of students (bottom of Fig.2) facilitates the learning situations derived on learning process. In this way, students will seek to improve their self in difficult topics by self-learning in order to provide learning enhancements.

Figure 2
Model Understanding and Learning the Most Difficult Topics (extract the most difficult learning subject)

Source: Authors' work

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 extract the most difficult learning subjects 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 significant problem under wide applications. Here are involved the student behavior analysis, scientific proofs, web access patterns etc. (Mabroukeh et al., 2010). A special case of association rule mining are analyzed to find temporal associations among events (Srikant et al., 1996). Association rule mining method is used for a diversity of applications that involves the studding of what paths in student co-operation of conduct leads to a much more successful group project (Perera et al, 2009).

According to behavioral patterns of learning in Al-Shabandar et al. (2017) paper are presented two sets of characteristics. Foreseeing adaptability of outcome course learners that takes part in MOOCs are compared. The relation among behavioral data and learners' certification are applied in correlation matrix. Rough sets methods with different mining techniques is used by Bell et al. (2007). The web log data is presented by Lodhi (2014) as conventional algorithm of mining sequential pattern

named as "Sequential ID3". It is experimental validation on web log data. (Nakamura et al., 2014).

The technique of clustering can be used to discover the new categories, which share the similar interest. The behavior of K-Means algorithm is used to separate learners at natural groups founded on its behavior for a wide dataset (Ratnapala et al., 2014).

E-learning can be more operative by utilization of Using Web Usage Mining techniques. An effective E-learning environment/implementation of the system reflects on learning style among pattern discovery and pattern analysis (Yadav et al., 2011).

Conclusion

According to the analyzed literature, it is suggested a model for learning history analysis founded on learning processes to extract the most difficult learning subjects of students. General look of a sequential pattern mining method is introduced first. It contains video learning data histories of programming sample classes. Afterwards it is defined a model for analyzing the programming learning history. It is founded on the suggested method.

To propose recommendation system based on rule-space model is considered as future work. It involves updated/improved across different inspections and practical learning histories to be used.

References

1. Al-Shabandar, R., Hussain, A., Laws, A., Keight, R., Lunn, J., Radi, N. (2017), "Machine learning approaches to predict learning outcomes in massive open online courses", in proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), 14-19 May 2017, Anchorage, AK, USA, IEEE, pp. 713-720.
2. Bell, D. A., Beck, A., Miller, P., Wu, Q. X., Herrera, A. (2007), "Video Mining-Learning Patterns of Behaviour via an Intelligent Image Analysis System", in proceedings of the 7th International Conference on Intelligent Systems Design and Applications (ISDA 2007), 20-24 October 2007, Rio de Janiero, Brazil, IEEE, pp. 460-464.
3. Lodhi, S. S. (2014), "Development of Sequential ID3:"An advance Sequential mining Algorithm"". American Journal of Software Engineering, Vol. 2, No. 2, pp. 16-21.
4. Mabroukeh, N. R., Ezeife, C. I. (2010), "A taxonomy of sequential pattern mining algorithms", ACM Computing Surveys (CSUR), Vol. 43, No. 1, p. 3.
5. Nakamura, S., Nozaki, K., Morimoto, Y., Miyadera, Y. (2014), "Sequential pattern mining method for analysis of programming learning history based on the learning processs", in proceedings of the 2014 International Conference on Education Technologies and Computers (ICETC), 22-24 September 2014, Lodz, Poland, IEEE, pp. 55-60.
6. Nakamura, S., Nozaki, K., Nakayama, H., Morimoto, Y., Miyadera, Y. (2015), "Sequential Pattern Mining System for Analysis of Programming Learning History", in proceedings of 2015 IEEE International Conference on Data Science and Data Intensive Systems (DSDIS), 11-13 December 2015, Sydney, NSW, Australia, IEEE, pp. 69-74.
7. Perera, D., Kay, J., Koprinska, I., Yacef, K., Zaïane, O. R. (2009), "Clustering and sequential pattern mining of online collaborative learning data", IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 6, pp. 759-772.
8. Prakash, B. R., Hanumanthappa, M., Kavitha, V. (2014), "Big data in educational data mining and learning analytics", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, No. 12, pp. 7515-7520.
9. Ratnapala, I. P., Ragel, R. G., Deegalla, S. (2014), "Students behavioural analysis in an online learning environment using data mining", in 7th International Conference on Information and Automation for Sustainability, 22-24 December 2014, Colombo, Sri

Lanka, IEEE, pp. 1-7.

10. Srikant R., Agrawal R. (1996), "Mining sequential patterns: Generalizations and performance improvements", in Apers P., Bouzeghoub M., Gardarin G. (Eds.), Advances in Database Technology — EDBT '96, EDBT 1996, Lecture Notes in Computer Science, Vol. 1057, Springer, Berlin, Heidelberg.

11. Shanabrook, D. H., Cooper, D. G., Woolf, B. P., Arroyo, I. (2010), "Identifying high-level student behavior using sequence-based motif discovery", in proceedings of the 3rd International Conference on Educational Data Mining, 11-13 June 2010, Pittsburgh, PA, USA, pp. 191-200.

12. Yadav, A., Jain, S. (2011), "Analyses of web usage mining techniques to enhance the capabilities of E-learning environment", in proceedings of the 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), 22-24 April 2011, Udaipur, India, IEEE, pp. 223-225.

About the 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.**

Visar Shehu has a PhD in computer science and is employed as an Assistant Professor at the CST department of South East European University. His main research interests are related to data mining and machine learning technologies with a focus on text data and natural language processing. He is also involved in applying his research in the industry, being the CTO of a software development company in Macedonia that develops HR solutions for the EU market. The author can be contacted at **v.shehu@seeu.edu.mk.**