Students’ Modeling based on their Problem Solving Behavior

Rabia Maqsood

Department of Computer Science, Università degli Studi di Milano, Italy

rabia.maqsood@unimi.it

Abstract. This research work aims at designing a framework to process the students’ logged traces and identifying different learning models based on their problem solving behavior specifically through trace-based exercises. Students depict different behaviors during problem solving including learning sequences and engagement level; thus yielding less structured and more complex interaction traces. It is therefore proposed to use Fuzzy Logic for pattern classification.

Keywords: Educational Process Mining, Pattern Classification, Problem Solving, Learning Traces.

1. Research Question

Teaching programming to freshers is a challenging job as most of the students fail to build conceptual models of complex programming concepts. The lack of ability to understand basic programming constructs at earlier stages results in a high failure or course drop out ratio in preliminary programming courses [8], [9]. Methods used to build viable mental models are commonly referred as: trace table, memory diagram and code visualization. These methods proved to be very effective as it helps novice programmers to better understand the flow of a program [9] and errors.

The core challenge is helping students in understanding the basic programming constructs and developing an accurate problem solving model, which is desirable by every teacher. Using trace tables for code dry-run exercises is amongst the traditional approaches used for teaching programming to the beginners [9]. Given a code snippet, students are required to fill a table showing each line of code to be executed in an expected manner and current state of the memory constructs (i.e. values of variables, arrays, etc.). However usually an instructor gets the final answer to the question(s) as in [13] and cannot keep track of the complete process followed by every individual. Consequently an instructor is not able to analyze the underlying process followed by each student [2].

Existing systems developed to support trace-based teaching keeps record of basic information only, for example: correctness/incorrectness of a solution and time spent on an exercise, as in [13] and [14]. However students usually perform several other operations during problem solving such as: adding/removing fields, changing values, submitting incomplete solution before the time expires, etc. Therefore analyzing the behavior of a student during problem solving can reveal the underlying process through interaction logs, for example: writing correct values of memory shows clarity of concepts, repeating a sequence of add/delete operations shows confusion or uncertainty about the answer, submitting incomplete solution earlier shows disengagement of a student during problem solving. Thus the problem solving process can vary to a large extent and students' behavior should be analyzed carefully to construct better students' models.

This research work aims at modeling students’ profile based on their problem solving behavior depicted through logged interactions in an e-learning tool. More specifically, this research study is intended to :

a. Use process mining (PM) techniques to analyze the learning process(es) followed by the students during solving trace-based exercises.
b. Classification of patterns discovered by process mining techniques.

c. And investigating the correlation of performance in trace-based exercises containing elementary concepts and exercises containing combination of concepts and thus yielding more complex concepts.

2. **Background**

Use of the technology to improve learning outcome has proved to be effective and is in use for almost more than two decades. With the immense use of technology in teaching, interactive e-learning tools are largely integrated into traditional classrooms [3]. Such an approach is referred to as 'blended learning' [6]. Intelligent e-learning tools are developed to support existing teaching practices as well as enhancing students' learning by providing personalized learning environment to each student based on his/her learning profile. Thus creating accurate learner's model is a critical issue in the development of intelligent e-learning tools and Intelligent Tutoring Systems [1]. A complete overview of existing approaches to create learner's profile is given in [1] which includes: Item Response Theory (IRT), Bayesian Networks and its variants, Psychometric models and Knowledge space theory models.

Moreover researchers focused on Educational Data Mining (EDM) techniques to create better learners' model. A growing interests in Educational Data Mining (EDM) has opened new challenges for the researchers to discover learning sequences [5] from interaction data recorded in e-learning tools (e.g. Intelligent Tutoring Systems (ITSs), MOOCs, etc.). Educational Data Mining (EDM) has been used extensively in the past to answer data-centric research questions such as: students' retention, performance prediction, etc. [5], [6]. In general, EDM techniques deal with demographic data, academic history and current performance attributes only and thus are not able to identifying the underlying process(es) [6] followed by a student during problem solving. However it is evident from the work conducted in [5], [10], [12] that techniques developed for Process Mining (PM) seemed promising to discover learning sequences followed by extracting information hidden in event logs recorded for each user in the e-learning tools [6], [11].

Educational Process Mining (EPM) has evolved as a new research field recently, which focuses on extracting process-related information from event logs maintained in e-learning tools [4], [5], [6]. Computer based educational tools can record extensive information from student's interaction with the system. The logged information in e-learning systems usually contains information about usage of learning material, assessment data [4], [7], system interaction history based on mouse-clicks, time spent on activities, etc. [10].

Prior research studies conducted in the context of EPM used different techniques to gain insights into the underlying learning process followed by the students during problem solving. For example, process mining techniques are used in [5] to identify learning sequences from students' behavior logged during a scientific inquiry problem solving process and classify students based on their skills. Another study conducted in [6] discovered actual process followed by the students while solving Multiple-Choice Questions (MCQs). In [4], social mining techniques were used to analyze the interactions between training providers and courses involved in students' training paths. Two groups of students were identified using Fuzzy miner in [12]: Surface and Deep learners, depicting different learning strategies followed by the students during writing a research project report. Authors in [10] also provide evidence of using logged traces to analyze different patterns adopted by the students during problem solving. A tree based algorithm is proposed in the reported work to predict with high accuracy a student's likelihood of attempting problems of unfamiliar topics. In [11] a technique is presented that transforms sequences of logged events into more meaningful actions and activities, which are then classified into activities using heuristic miner. The results demonstrated the possibility of constructing more accurate students' models from inquiry patterns using abstractions at various levels.

3. **Significance**

With the immense use of e-learning tools in traditional classrooms, it is becoming necessary to make better use of logged data to understand students' behaviors. Students depict different behaviors during problem solving including learning sequences (selection of problem sets), different order of performing tasks and engagement level; thus yielding less structured and more complex interaction traces. Prior work on investigating problem solving behavior involves writing skills [12], process inquiry skills [11], solving mathematical word problems [10], and very little or no attention is given on analyzing problem solving skills exercised by the programmers.

Earlier studies used process mining techniques to primarily deal with operational process(es) or structured data, considering students' interaction logged data as input to the process mining algorithm [11], for example as in [4], [6] and [7]. However this is not the case when we have to investigate more complex human behaviors through their logged interactions. This data tends to be less structured [11], [12] as the order of performing task is not restricted and students may perform different intermediate actions before reaching to a final solution. Therefore it is proposed to use Fuzzy Logic for pattern classification which will be described in terms of linguistic variables and if-then rules.

4. Research design and methods

It is proposed to design a tool allowing students to solve trace-based programming exercises and automatically checking their solutions as described in [13]. Moreover the tool will maintain students' profiles which will be determined by processing the logged data. The tool will record data from students' interaction including mouse-clicks, entries in the text box(es), adding/deleting rows in the solution trace table, etc. Students' profile will be updated based on their performance and output class determined by the mining algorithm using their problem solving behavior.

5. Research stage

This is an initial proposal of the research study and will be conducted along the following phases:

a. **Phase I:** Initially a tool will be designed which will generate a solution trace table for each exercise by parsing the code into tokens and recording each new identifier and its attributes in a data structure, similar to how it is being done in a compiler. The tool will also maintain students' profile by recording their interactions and performance attributes. In parallel a careful study of state-of-the-art techniques will be done and taking relevant courses in the first year.
b. **Phase II:** The tool will be used by the undergraduate students of Computer Science at the University of Milan, Italy, for experimental studies in several phases. The tool will maintain students' profiles which will be updated based on their performance and output class determined by the mining algorithm using their problem solving behavior.
c. **Phase III:** Collected data will be studied carefully and appropriate process mining algorithm(s) will be used for pattern classification. Results will be analyzed and compared with existing approaches to model students' profile.

References

1. Desmarais, M. C. and Baker R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1-2), 9-38.
2. Cambranes-Martinez, E. and Good, J. (2011). Studying problem-solving strategies in the early stages of learning programming. *In Proceedings of the 4th International Conference on Educational Data Mining*, EDM'2011.
3. Qi, B., Liu, L. and Wang, C. (2009). E-learning Tools to Improve Students' Learning Experience: a case study. *International Journal of Modern Education and Computer Science (IJMECS)*, 1(1), pp. 66.
4. Cairns, A.H., Gueni, B., Fahima, M., Cairns, A., David, S. and Khelifa, N. (2015). Process Mining in the Education Domain. *International Journal on Advances in Intelligent Systems*, 8(1 & 2).
5. Emond, B. and Buffett, S. (2015). Analyzing student inquiry data using process discovery and sequence classification. *In Proceedings of the 8th International Conference on Data Mining (ICDM'15)*, 412-415.
6. Pechenizkiy, M., Trcka, N., Vasilyeva, E., Aalst, W. and Bra, P. D. (2009). Process mining online assessment data. *In proceedings of International Conference on Educational Data Mining*, 279-288.

7. Azzini, A., Ceravolo, P., Scarabottolo, N. and Damiani, E. (2016). On the predictive power of university curricula. *2016 IEEE Global Engineering Education Conference (EDUCON)*, 929-932.
8. Hertz, M. and Jump, M. Trace-based teaching in early programming courses. (2013). *In Proceeding of the 44th ACM technical symposium on Computer science education (SIGCSE '13)*, March 06-09, 561–566.
9. Mselle, L. J. (2014). Article: Using Memory Transfer Language (MTL) as a Tool for Program Dry-running. *International Journal of Computer Applications*, 85(9), 45–51.
10. Ng, K. H. R., Hartman, K., Liu, K. and Khong, A. W. H. (2016). Modelling the way: Using action sequence archetypes to differentiate learning pathways from learning outcomes. *In Proceedings of the 9th International Conference on Educational Data Mining, EDM'16*, 167-174.
11. Southavilay, V., Markauskaite, L. and Jacobson, M. J. (2013). From "events" to "activities": Creating abstraction techniques for mining students' model-based inquiry processes. *In Proceedings of the 6th International Conference on Educational Data Mining, EDM'13*, 280-283.
12. Beheshitha, S. S., Gašević, D. and Hatala, M. (2015). A process mining approach to linking the study of aptitude and event facets of self-regulated learning. *In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK '15)*, ACM, New York, NY, USA, 265-269.
13. Wahab, A., Sadiq, M. A. and Ajmal, S. (2016). CodeMem: A tool for evaluating programming skills. *Final Year Project Report, National University of Computer and Emerging Sciences (NUCES)*, Pakistan.
14. Michael Striewe and Michael Goedicke (2014). Code reading exercises using run time traces. *In Proceedings of the 2014 conference on Innovation & technology in computer science education (ITiCSE '14)*. ACM, New York, NY, USA, 346-346.