Discovering Process in Curriculum Data to Provide Recommendation

Ren Wang
Department of Computing Science
University of Alberta
ren5@ualberta.ca

Osmar R. Zaïane
Department of Computing Science
University of Alberta
zaiane@cs.ualberta.ca

ABSTRACT

Process mining is an emerging technique that can discover the real sequence of various activities from an event log, compare different processes and ultimately find the bottleneck of an existing process and hence improve it. Curriculum data is the history of the courses effectively taken by students. It is essentially process-centric. Applying process mining on curriculum data provides a means to compare cohorts of students, successful and less successful, and presents an opportunity to adjust the requirements for the curriculum by applying enhancement of process mining. This can lead to building recommenders for courses to students based on expected outcome. In this paper we first discover a process model of students taking courses, then, compare the paths that successful and less successful students tend to take and highlight discrepancies between them. The conclusion we reached is that process mining indeed has a great potential to assist teachers and administrators to understand students behavior, to recommend the correct path to students, and at last to enhance the design of a curriculum.

1. INTRODUCTION

The term curriculum often refers to a predefined recommended or mandatory sequence of actions including courses or resources for students. It is designed by a school or a university in order to achieve some educational goals. To maximize this goal, some constraints are frequently imposed, e.g., students must take some specified courses before taking others. Given the liberal approach for selecting courses and taking into account these prerequisites for the courses and the requirements for the programs, students can follow different paths from start to finish. Discovering and understanding the process students follow, or some cohort, such as the most successful learners, can be very indicative to curriculum administrators and can also be the basis for a recommender system to recommend appropriate paths to students in terms of courses to take and in terms of prioritizing the sequence of courses. The common way to analyze educational data is using simple statistics and traditional data mining. However statistics and conventional data mining techniques do not focus on the process as a whole, and do not aim at discovering, analyzing, nor providing a visual representation of the complete educational process [3]. Process mining consists of extracting knowledge from event logs recorded by an information system and is inherent in discovering business process from these event logs, comparing and conforming processes, and providing mechanisms for improvements in these processes[4]. Process mining techniques are often used in the absence of formal description of the process and can provide a visualization with a flowchart as a sequence of activities with interleaving decision points or a sequence of activities with relevance rules based on data in the process.

Some attempts have already been made to exploit the power of process mining in curriculum data, historical data encompassing the sequence of courses taken by students. For instance, the authors of one chapter in [2] give a broad introduction of process mining and indicate that it can be used in educational data. The first paper that proposes to utilize process mining on curriculum data is [3]. The main idea is to model a curriculum as a Colored Petri net using some standard patterns. [1] directly targets curriculum data and brings up a notion called curriculum mining. Similar to the three components of process mining, it clearly defines three main tasks of curriculum mining, which are curriculum model discovery, curriculum model conformance checking and curriculum model extensions.

The application of process mining on curriculum data offers a wide range of possibilities. First it can help the educators understand and make better decisions with regard to the offered curriculum. For example, what is the real academic curriculum? Are there paths seldom used and others more popular? Do current prerequisites make sense? Are the particular curriculum constraints obeyed? How likely is it that a student will finish the studies successfully or will drop out? It can also assist students to choose among different options and even make recommendations to students. For instance, How can I finish my study as soon as possible? Is it more advantageous to take course A before B or B before A? Should I take courses A and B or courses B and D this semester in order to maximize my GPA? Answering such questions to both educators and students can greatly enhance the educational experience and improve the education process. We show in this paper how some of these questions can be answered using the history of courses taken.

2. CURRICULUM DATA

Although the data about courses have already been collected by the Computing Science department of the University of Alberta, we cannot publish any result related to such data due to lack of ethical approval. However, we wrote a curriculum simulator to mimic the behaviors of different kinds of students from the department and be as close as possible to the real data. First, we predefined a set of rules or requirements similar to those in the offered programs in the department. For example, prerequisites, i.e. some specified courses must be taken before the student takes another one. Other requirements include the first and the last course a student must take, mandatory courses, and non-coexisting courses, i.e. if the student takes one course in the group then they cannot take any other course belonging to the same group. Then, we generated students in three categories: the responsible students who always satisfy the course constraints; the typical students who seldom violate course constraint rules; and the careless students who often do not follow the set rules. Moreover, we differentiated the students based on the range of marks they are assigned in courses they take creating clusters of successful and less successful students. We generated the historic courses data for each student adding some probability that a student withdraws from a course giving the course load and previous withdrawing behavior.

3. DATA ANALYSIS

The final goal is to examine what kind of paths successful students tend to take and what is the discrepancy between successful students and less successful students so that we can make recommendations to steer the students to the successful paths. Since we have predefined rules for different types of students in our simulations, the goal is to verify whether we can discern these rules purely from the model we discover by process mining. If we can find the rules from the model, then we are safe to say it is possible to distinguish the ”correct path” that can yield the best result by means of process mining, thus a recommendation, that closes the gap among students, can be achieved.

The several process models that were discovered from the curriculum log are close portrayal of real curriculum models in our computing science department. Each model covers the most frequent activity paths, given some thresholds. This is because the model map would be too dense and cluttered to recognize patterns if we present all of them. We added an additional activity at the end of each case to indicate the type of the student. In practice, this type can be any cohort of students such as based on the GPA ranges, based on graduation distinction, withdrawal, or other criteria. To inspect students' behavior patterns in more detail, we further filtered the model with their last activity, i.e., partitioning students based on their type so that we can compare them. The rules we imposed while generating the curriculum data can indeed be easily verified. For the students who seldom violate course constraint rules, the frequent paths appear very similar to those of the first group. However, contrasting the complete graphs of these groups reveals peculiar paths specific to one or the other group. The contrast is even more pronounced when comparing the responsible students and the careless students, as defined in the data. This grouping can be a placement test in some other cases. The categorization can also be done at the end of the paths based on the outcome at the end of the program or the end result for a given course. This allows contrasting the paths taken by successful students with other paths at the end of a program, or comparing the initial paths of students who dropped out of a course to paths leading to the same course taken by those who finish that course. The result of contrasting paths of different cohorts of students stresses out desired and undesired paths specific to some groups, the analysis of which can highlight recommendations for new prerequisites to align new students from a potentially undesired path to the desired one. In the case of drop-outs from courses, this analysis provides insights on the potentially faulty sequence of courses or lack of certain courses in the sequence that lead to higher risk of dropping out. In addition to providing better understanding of the curriculum data and a way to discern between behaviors of different cohorts of students, contrasting between process models from different groups of students presents an opportunity for a course recommender system. By contrasting between the processes followed by students grouped based on their course outcome or based on final GPA, we can find and visualize the sequences of courses that lead to the highest probability of success for a given course. Based on the courses already taken by a student, the system can indicate the options to take that have the highest chance to improve the GPA. Similarly, the system can recommend to take a course before another to maximize outcome. The same data can also be used by administrators to define new prerequisites for courses and thus improve the chance for the adoption of better paths. We are currently building such a recommender system for students. The system would use evidence from historical data to provide comparison of average ranges of prospective marks if a student follows one path or the other when selecting courses.

4. CONCLUSION

Process mining, to discover sequences of courses taken by students, is indeed a powerful tool to analyze curriculum data. By this means, we can visualize and formalize the real paths students actually take, and reveal the underlying patterns such as prerequisites and other constraints. Moreover, conformance in process mining can reveal paths that are unexpectedly not followed by students. Furthermore, contrasting processes from different cohorts of students discloses hidden specificity that we can act upon. Most importantly, contrasting processes provides means to recommend more appropriate sequences of courses to students personalized to their own cases and exposes new insights to administrators.

5. REFERENCES

[1] M. Pechenizkiy, N. Trcka, P. De Bra, and P. Toledo. Currim: Curriculum mining. In *Intl. Conf. on Educational data Mining*, pages 216–217, 2012.

[2] C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker. *Handbook of educational data mining*. CRC Press, 2010.

[3] N. Trcka and M. Pechenizkiy. From local patterns to global models: Towards domain driven educational process mining. In *Intl. Conf. on Intelligent Systems Design and Applications*, pages 1114–1119, 2009.

[4] W. Van Der Aalst. *Process mining: discovery, conformance and enhancement of business processes.* Springer Science & Business Media, 2011.