

CurriM: Curriculum Mining

M. Pechenizkiy, N. Trčka, P. De Bra
Department of Computer Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven
the Netherlands
{m.pechenizkiy,n.trcka,debra}@tue.nl

Pedro A. Toledo
Department of ISAATC
Universidad de La Laguna
Fac. Física. Avda. Francisco Sanchez SN
38203 SC de Tenerife, Spain
pedro@isaatc.ull.es

ABSTRACT

Curriculum mining includes three main kinds of tasks: (i) *actual* curriculum model discovery, i.e. constructing complete and compact academic curriculum models that are able to reproduce the *observed* behavior of students, (ii) curriculum model conformance checking, i.e. checking whether the observed behavior of students match their expected behavior as defined by the previously discovered or pre-authored curriculum model, and (iii) curriculum model extension, i.e. projecting information extracted from the observed data onto the model, to make the tacit knowledge explicit, facilitate better understanding of the particular academic processes and enable decision making processes. We discuss student and education responsible perspectives on curriculum mining and present the achievements of the ongoing project aiming to develop curriculum mining software including process mining, data mining and visualization techniques.

1. INTRODUCTION

An academic curriculum is a (legal) document defining a specific learning program that puts certain types of constraints on how students are expected to take the courses. These constraints typically describe a set of courses and a set of relationships between these courses. The rule “*Algorithms must be passed before Advanced Algorithms can be taken*” can, for example, be seen in a computer science curriculum.

In the current practice however, the rules in a curriculum are usually stated informally, in a natural language, and are thus subject to multiple interpretations. A curriculum is considered simply as the range of courses from which students choose what courses to study. It is not uncommon that students have to approach their study advisors to ask whether they are allowed to enroll in a certain course or what would the impact be if they did. The advisors, on the other hand, are themselves often confused and must ask the board of education for clarification.

To alleviate the above problem we consider formal modeling of a generic academic curriculum. The problem of curricular modeling has been widely recognized as an important and nontrivial task, and so far there have been a few attempts to address it. The problem of authoring and formal modeling of curriculum constraints and personalized study curriculum was addressed in [1]. Modeling an academic curriculum plan as a mixed-initiative constraint satisfaction problem

was proposed in [7]. In [5] we identified some typical constraints defined in the existing curriculum and used the formal language of Colored Petri nets (CPNs) to encode these constraints in form of patterns, thus giving a precise and unambiguous semantics to the study rules.

The advantages of having a formal and executable model of a curriculum are not only in its elimination of ambiguity, but in the fact that through the use of CPN Tools and ProM (<http://www.processmining.org>), such a model almost directly offers a wide range of possibilities: students can automatically check, by themselves, whether they are allowed to do something or not; historic data stored in the log of the educational information system can be compared against the model (in this way we could see whether the curriculum was always respected in the past); the same historic data can be used to equip the model with quantitative information (probabilities, delays, etc.) enabling all kinds of performance analysis (finding, e.g., the average time to graduation, the most common paths, etc.) and casting recommendations; and online simulation of the model can facilitate real-time detection of curriculum violations.

In this poster we consider our current ongoing and further planned work on the curriculum mining tools and techniques development. We take both student and education responsible perspectives on curriculum mining that correspond to different information needs.

2. CURRICULUM MINING

Figure 1 illustrates the general ideas behind the curriculum modeling and analysis. An information system that supports an educational institution generates event logs that are stored in a database and represent e.g. student performance and enrolment into the courses and corresponding exams. Educators, responsible e.g. for curriculum development and monitoring its effectiveness, can help to identify typical constraints that should be enforced in the study process. Given the event log reflecting historical data and pattern templates we can mine all the patterns present in the database satisfying the templates. The resulting pattern set can be post-processed in a semi-automatic way, and then a unique process model can be assembled into a graph structure. On the process model we can next perform standard task of process mining, like e.g.: (i) determining popular paths and the most important parts in the curriculum, (ii) extending the model with additional information or modifying it, (iii) executing various what if scenarios to facilitate real time decision making, and (iv) real-time monitoring of the process flow as well as monitoring for the changes in the flow or the curriculum.

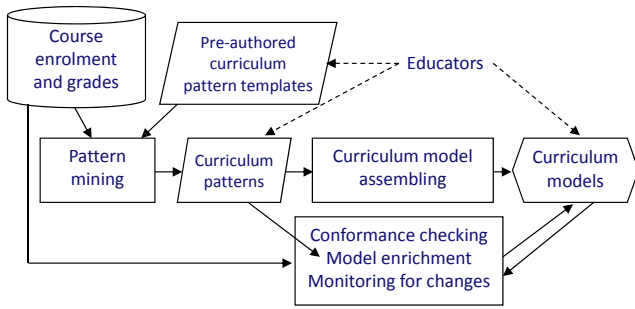


Figure 1: A conceptual view on the formal modeling of the study curriculum.

Figure 2 gives an example how the prerequisite constraints and follow up dependencies for the selected or currently recommended key course in the second year are projected to the study curriculum of a particular student. A warning for not having a high enough grade (8) for one of the prerequisite courses as recommended for taking *C* is highlighted in red.

| Year 1 | | | | Year 2 | | | | Year 3 | | | |
|------------|----|------------|----|------------|----|------------|----|------------|----|------------|----|
| Semester 1 | | Semester 2 | | Semester 1 | | Semester 2 | | Semester 1 | | Semester 2 | |
| Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |
| 9 | 8 | 7 | | | | | | | | | |
| | 8 | | | | | | | | | | |
| | 7 | | | | | | | | | | |
| | 8 | | | | | | | | | | |

Figure 2: Projecting patterns on the study curriculum.

3. ONGOING AND FUTURE WORK

Our current efforts are focused on making the *simplest* elements of curriculum mining work in practice such that students, study advisors and directors of education would find them useful. To achieve this we conduct a case study on mining the studying behavior of students at Eindhoven University of Technology over the past 10 years. A typical set of the educators’ questions to answer with CurriM include: “What is the *real* academic curriculum (study program)? “How do students *really* study?”, “Do current prerequisites make sense?”, “Is the particular curriculum constraint obeyed?”, “How likely is it that a student will finish the studies successfully or will drop out?. And students’ questions like “What are the typical (or the *best*) ways to study?”, “What is my expected time to finish?” and “Should I take courses A and B or courses B and D now?”.

Besides the relatively straightforward tasks of conformance checking for the known constraints [6], analyzing how much time and effort a particular activity takes [4], and student dropout prediction [3], we work on providing support to answering more elaborate questions. E.g. finding most common types of behavior and clustering them; finding emerging

patterns that capture significant differences in the behavior of students who graduated vs. those students who did not or significant changes in behavior of students from one generation and the other; finding frequent patterns that describe a bottleneck in the curriculum, i.e. patterns explaining for which students it is the bottleneck and why. E.g. it is possible to observe that students who take Database modeling theory before passing Logic and set theory or passing it with the minimal grade have to do many more re-examinations for several related courses. We expect that projecting such information to the students will alter their studying behavior, motivating them more strongly to focus on the critical knowledge and skills at the right periods of their studying paths.

It is not uncommon for the curriculum to evolve over time and go through major changes from time to time. Therefore our further work will also include mining and taking into account changes in the study curriculum. This is related to the recently introduced problem of concept drift in process mining [2].

4. ACKNOWLEDGEMENTS

This work has been partly funded by Surf Foundation, the Netherlands. We would like to thank STU for providing the data for curriculum mining case study.

5. REFERENCES

- [1] M. Baldoni, C. Baroglio, I. Brunkhorst, N. Henze, E. Marengo, and V. Patti. Constraint modeling for curriculum planning and validation. *Interactive Learning Environments*, 19(1):81, 123 2011.
- [2] R. P. J. C. Bose, W. M. P. van der Aalst, I. Zliobaite, and M. Pechenizkiy. Handling concept drift in process mining. In *Proc. 23rd Int. Conference on Advanced Information Systems Engineering, CAiSE 2011*, volume 6741 of *Lecture Notes in Computer Science*, pages 391–405. Springer, 2011.
- [3] G. Dekker, M. Pechenizkiy, and J. Vleeshouwers. Predicting students drop out: A case study. In *Proc. 2nd International Conference on Educational Data Mining, EDM 2009*, pages 41–50.
- [4] M. Pechenizkiy, N. Trcka, E. Vasilyeva, W. M. P. van der Aalst, and P. D. Bra. Process mining online assessment data. In *Proc. 2nd International Conference on Educational Data Mining, EDM 2009*, pages 279–288, 2009.
- [5] N. Trcka and M. Pechenizkiy. From local patterns to global models: Towards domain driven educational process mining. In *Proc. 9th International Conference on Intelligent Systems Design and Applications, ISDA 2009*, pages 1114–1119. IEEE Computer Society, 2009.
- [6] N. Trcka, M. Pechenizkiy, and W. M. P. van der Aalst. *Handbook of Educational Data Mining*, chapter Process Mining from Educational Data, pages 123–142. CRC Press, 2010.
- [7] K. Wu and W. S. Havens. Modelling an academic curriculum plan as a mixed-initiative constraint satisfaction problem. In B. Kégl and G. Lapalme, editors, *Canadian Conference on AI*, volume 3501 of *Lecture Notes in Computer Science*, pages 79–90. Springer, 2005.