CurriM: Curriculum Mining

Project Summary

CurriM project is aimed at investigating the challenges of the educational process mining, of mining an academic curriculum in particular. Curriculum mining includes three main kinds of tasks: (i) constructing complete and compact academic curriculum models that are able to reproduce the observed behavior of students (curriculum model discovery), (ii) checking whether the modeled and expected behavior (either pre-authored or discovered from data) matches the observed behavior of students (curriculum model conformance checking), and (iii) projecting information extracted from the logs onto the model, to make the tacit knowledge explicit and facilitate better understanding of the particular academic processes (curriculum model extension).

The main expected results:
- the first software prototype for academic curriculum modeling;
- examples of academic curriculum modeling at the Department of Computer Science, TU Eindhoven;
- guidelines for management of curriculum related data;
- a roadmap for further study and development of the academic curriculum modeling software tools.

1. Objective

The main objective of CurriM project is to help different target groups, i.e. the directors of education, study advisors, study program coordinators and students, to get a holistic view on the study programs and to better understand the envisioned and actual educational processes.

Particular kinds of questions, for which we want to provide an automated support, include but are not limited to:

- What is the real academic curriculum (study program)?
- How do students really study?
- Is there a typical (or the best) way to 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?
- What is my expected time to finish?
- Should I take course A or course B now?

The goal of this project is to make the first practical steps towards having user friendly, intuitive and interactive educational software tools, built on (but hidden from the end users) solid foundations in process mining and data mining technology.

2. Positioning of the project

Background. In modern education various information systems are used to support educational processes. In the majority of cases these systems have logging capabilities to audit and monitor the processes they support. At the level of a university, administrative information systems collect information about students, their enrolment in particular programs and courses, and examination grades. These data can be analyzed from various levels and perspectives, showing different aspects of the educational process.

Traditional data mining techniques have been extensively applied to find interesting patterns, build descriptive and predictive models from large volumes of data accumulated through the use of different information systems. The results of data mining can also be used for getting a better understanding of educational processes, for generating recommendations and advices to students.
Process mining tools and techniques focus on extracting process-related knowledge from event logs recorded by an information system. The general ideas behind our framework can be illustrated with Figure 1 below. An information system that supports 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. If we can induce the process model, we can next perform standard tasks of process mining, like for example: (i) determining popular paths or “narrow” places 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. Note that we do not exclude the possibility of a manual design of the complete process model by the domain experts. From the point of view of the analysis, the same tasks can be performed on the manually designed model as on the mined one. However, a benefit of this is that we can, using the standard conformance checking techniques, automatically check whether any of the required constraints (now captured in the designed model) have been violated in the past.

**Project focus.** In this project we plan to develop curriculum mining software based on the domain-driven Educational Process Mining (EPM) framework (see Figure 2) that facilitates more focused search for local patterns and their further assembling into a global model.

---

**Figure 1 Educational process mining framework**

**Figure 2 Curriculum mining framework**
Our EPM framework assumes that, based on domain knowledge, we can predefine a set of pattern templates that are crucial for the problem at consideration. This pattern set can always be modified and extended, but we assume that the templates (and possibly concrete patterns too) are pre-authored at the moment of process mining. Thus, we can reduce the search space and direct the process mining in the desired way. Conformance checking and model extension can also be focused on the patterns of our prior interest. Moreover, many other activities can be performed in exactly the same way as it is currently done in the existing process mining framework.

In this project educators, responsible for curriculum development and monitoring its effectiveness, can help us 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 (represented as a colored Petri net graph in our case) can be assembled into a graph structure (or first abstracted to a more comprehensive representation, so that domain experts are not puzzled with the lower-level primitives).

3. Project planning

The project can be divided into the following major tasks:

**T1. Developing the first software prototype for academic curriculum modeling.**
This task consists of a few important subtasks, which being tightly combined together, constitute mini R&D cycles:
- identifying types of curriculum specific patterns we need to mined from the event logs (in collaboration with the domain experts) and include in the modeling and developing corresponding pattern mining and pattern assembling techniques;
- implementing techniques and integrating it with ProM software that provides an important process mining foundation framework and many of the building blocks for curriculum modeling software;
- testing a particular piece of software.

Because of the limited resources for this initial project our plan is to concentrate primarily on the “easygoing” pieces and leave the more challenging parts for further work, to be included in T3.

Depending on the speed of the progress we will or will not focus on usability issues, in particular “hiding” the technical details by introducing the higher-level abstractions for curriculum-related patterns.

T1 will be executed in parallel with T2.

**T2. Case study: modeling the curriculum of the Department of Computer Science, TU Eindhoven;**
The ultimate goal of this task is to validate the usefulness of the developed mining curriculum mining techniques and their implementations in the software. In data mining and process mining it is often not enough to build the correct or sound algorithms and to implement them in a software toolkit. We want the resulting models, which are constructed with these techniques, to provide certain utility, i.e. be useful for the end users. Given the timeline of this project it is important that we have a few of such cycles during the project execution to receive a timely feedback from the analysis of the resulting models. Experimenting with the real historical data will hint us what issues have been omitted in the initial R&D sprints.

Working with real data also gives an understanding how good or bad it is wrt organization, noise, redundancy, consistency, completeness etc. Obviously through the hands on experience with the data that has been collected already in the past we can developing guidelines for management of the curriculum related data to avoid the problems we will encounter or envision during the case study.

**T3. Creating a roadmap for further study and development of the academic curriculum modeling software tools.**
The main goal of this task is to develop R&D agenda for the coming years. This includes identification of not only research challenges (i.e. answering the question “what kind of new data mining and process mining techniques are needed to address the peculiarities of the curriculum mining domain?”) but also the strategy of the smooth technology transfer.
to the prospective end users, i.e. early adopters (e.g. TUE or 3TU departments) that would help to validate the usability and usefulness of the curriculum mining software “in the wild”.

**Tentative timeline:**

<table>
<thead>
<tr>
<th>Time period</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.04. 2012 – 15.06. 2012</td>
<td>T1. Software prototype development, D1</td>
</tr>
<tr>
<td>01.04.2012 – 15.06.2012</td>
<td>T2. Bachelor/master curriculum modeling, D2</td>
</tr>
<tr>
<td>15.06.2012 – 30.06.2012</td>
<td>Experience report on T1 and T2, including also guidelines for management of curriculum related data D3</td>
</tr>
<tr>
<td>30.06.2012 – 14.07.2012</td>
<td>Holiday break</td>
</tr>
<tr>
<td>15.09.2012 – 30.09.2012</td>
<td>T3. A roadmap for further study and development of the academic curriculum modeling software tools, D4</td>
</tr>
</tbody>
</table>

T1 and T2 will be executed in parallel, as T2 provides an immediate feedback to the development of T1, T1 and T2 will be also continued as much as time permits while working on D3, D4 and D5.

*Available resources:* 10 years of student examination data at TUE (date of the exam, exam code, grade + static data per student). Basic preprocessing tools for using the data in data mining and process mining software are available

*Continuation:* The goal of this project is two-fold – to make first practical steps towards development of the curriculum mining software tools and to prepare the roadmap for its further development, testing with prospective end users and deployment into the educational ICT.

Already at this stage we can envision that some of the curriculum mining tasks will be more challenging than the others and will likely require developing new mining techniques tailored for this educational domain. Depending on the availability of particular further funding opportunities, we will plan how to balance appropriately research, development, testing and deployment components of the project.

**Related work**


