

Academic year 2015–2016

## Hubble Project: User Profiling

Supervisor : Antoine PIGEAU

Email : [Antoine.Pigeau@univ-nantes.fr](mailto:Antoine.Pigeau@univ-nantes.fr)

Co-encadrant(s) : Yannick Prié

### Context

The Open Class Room website provides online courses in various area, from art and culture to computer science. A course is composed of text, video or ebook that users browse/read/download after a registration process. The validation of a course is carried out thanks to exercises and quizzes. A validation of a course may result in a certificate if the user chooses a premium registration.

The success of Open Class Room enables the availability of many user profiles with their associated traces on the courses:

- the personal information are provided on the registration process. Each user provides his name, his gender, his skills, his grade, his job and the courses taken;
- the traces contain all the actions carried out by the users on both the client and server sides. Accesses to each course, parts of a course or a chapter, quiz and exercise are then recorded for each user.

The goal of the project is to study the reasons for the failure or the success of the students. Providing such an answer is of much interest for both the course designers and the managers of the Open Class Room website. The objective of such a study is to improve the design of the courses and to be capable of anticipating the failure of a student.

### Goal of the internship

The objective of the internship is the modelling of user profiles based on their personal information and on the way they browse/learn a course.

Our input data are the personal information of the user and his traces, and the output is a set of user's profiles. Such a profile will summarize the background, the interests and the learning methods of a user (or a group of users).

The following research area will be studied to generate the profiles:

- process modelling [1]: a set of user traces on a same course is summarized to emphasized the main step to learn a course, from the beginning to the validation exercises;
- pattern mining [2] and sequential pattern mining [3]: search for sequence similarities on a specific course, or a set of courses. For instance, a pattern obtained from users with high marks could be pertinent to motivate a good practice;
- user clustering: search for groups of users with similar backgrounds and a similar way to browse/learn the courses. The clustering could be obtained from the two previous points. Group of users would be defined with a similar process or similar frequent patterns.

Each data mining process can be applied on a subset of users, defined with specific values of the personal user's information, the skills and grades for example. The objective here will be to study the correlations between set of users with different properties.

## Tasks

Here are the tasks to be carried out all along the internship:

1. study the raw data provided by Open Class Room. Client/server logs are provided by the Open Class Room website. A detection phase [4] will be first applied to separate each access of a user to a course;
2. define the set of user actions to retrieve from the traces;
3. generate the sequence of actions for each user on each course;
4. study and define the data mining process to apply on the sequence of actions;
5. generate the user profiles.

## Skills

- R or Python programming
- Pattern mining
- Clustering

## Practical information

- Period: from January to July
- Salary: around 550 euros per month

## How do I apply ?

Send your resume + a motivation letter + academic transcript to [antoine.pigeau@univ-nantes.fr](mailto:antoine.pigeau@univ-nantes.fr) and [yannick.prie@univ-nantes.fr](mailto:yannick.prie@univ-nantes.fr)

## References

- [1] W. V. D. Aalst. *Process Mining*. Springer, 2011.
- [2] C. C. Aggarwal and J. Han, editors. *Frequent Pattern Mining*. Springer, 2014.
- [3] C. H. Mooney and J. F. Roddick. Sequential pattern mining – approaches and algorithms. *ACM Computing Surveys*, 45(2):19:1–19:39, Mar. 2013.
- [4] M. Sadallah, B. Encelle, M. Azze-Eddine, and Y. Prié. Towards Reading Session-based indicators in Educational Reading Analytics. In *EC-TEL 2015*, Proceedings of the 10th European Conference on Technology Enhanced Learning, Toledo, Spain, Sept. 2015.