Unsupervised extraction of students navigation patterns on an EPFL MOOC

Thibault Asselborn¹, Victor Faramond¹, Louis Faucon¹
1. École Polytechnique Fédérale de Lausanne, Switzerland
Correspondence: firstname.lastname@epfl.ch

Introduction

How do students learn in MOOCs? This project aims at answering this question by analyzing the activities of thousands of students registered on EPFL ScalaMOOC hosted by Coursera². With the rapid growth of MOOCs, Education Science has entered the Big Data bubble, bringing new opportunities to study and improve learning technologies. We are interested in studying students navigation patterns which are the short sequences of learning activities that a student performs on the MOOC platform. In our case, the learning activities are one of watching a video lecture, reading or posting on the forum and submitting assignments. In this project we use unsupervised machine learning techniques to extract the main navigation patterns of students and gain insights on their behavior. We produce a simple and efficient visualization tool in order to provide feedback to teachers to help them understand the potential difficulties encountered by their students during the course and, if necessary, take actions accordingly.

Data processing Pipeline

Data

Our dataset contains the logs describing student’s interaction events with the MOOC platform. The events are of three type: Forum, Video and Assignment. Detailed information about the data is displayed in table 1.

<table>
<thead>
<tr>
<th>Event</th>
<th>Video</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>EventID</td>
<td>TimeStamp</td>
<td>TimeStamp</td>
</tr>
<tr>
<td>EventSubType</td>
<td>EventSubType</td>
<td>ProblemID</td>
</tr>
<tr>
<td>VideoID</td>
<td>Grade</td>
<td>HardCloseTime</td>
</tr>
</tbody>
</table>

Table 1: Schema of log data from the MOOC.

Preprocessing

- Remove events before the beginning and after the end of the course
- Remove the unnecessary data
- Remove students not working on assignments

Feature engineering

We extract students navigation patterns for each assignment of the course and transform these patterns into vectors of features describing them. Our features are designed to capture the learning behavior of students such as for example numberOfVideosBeforeFirstProblem and numberOfProblemEvent, describing respectively the number of lectures watched by a student before submitting the assignment and the number of times the student submitted the assignment. The complete set of features is displayed on figure 3.

Visualizing Learning Patterns

To visualize the navigation patterns of students, we use Sankey diagrams². The figure 2 shows two such diagrams for two different assignments of the MOOC. We can see at one sight the proportion of students skipping the videos, if a video is repeated or skipped by many students or if students fail their first attempt at the assignment.

Extracting Clusters of Navigation Patterns

We apply a K-Means clustering on the navigation patterns' vectors of features, in order to understand the different approaches of students. We chose to divide the patterns into 3 clusters as this separation gave the highest silhouette score.

Cluster 1: Typical students (40.2% of students)

These patterns show students who mostly do not watch any videos, do not use the forum and go directly to the problem. They are therefore very fast until submitting their last problem and generally obtain a very good grade. These students seem to already have the knowledge for this course and are show strong motivation in earning the certificate.

Cluster 2: Certificate seekers (28.5% of students)

These patterns show students who mostly don't do watch any videos, do not use the forum and go directly to the problem. They are therefore very fast until submitting their last problem and generally obtain a very good grade. These students seem to already have the knowledge for this course and are show strong motivation in earning the certificate.

Cluster 3: General students (31.3% of students)

These patterns show students who seem to have some difficulties with the course. It is shown by their first grade lower than the other clusters and also by the higher number of forum events. In this cluster, students submit the problem several times and have to go back to the lesson (lectures are watched after the first attempt).

Changes of Learning Strategies

Finally, we provide a visualization of students changes in navigation patterns type along the course. This can show to teachers the proportion of students in each of the three clusters for each assignment. Thus, it can reveal for example an increase in the number of struggling students or a decrease of lecture interest.

References

https://www.epfl.ch
https://www.coursera.org
²The diagrams were produced using http://sankey.cdakdev.org/