# Preliminary Analysis of MOOC Learners' Traces

Alya Itani[1], Issam Rebaï[1], Serge Garlatti[1], and Jean-Marie Gilliot[1]

[1]Institut Mines-Télécom, Télécom Bretagne
{alya.itani,issam.rebai,serge.garlatti,jm.gilliot}@telecom-bretagne.eu

**Abstract.** Despite the rapid expansion of MOOCs there is a main criticism to its discourse, which is its low completion rates. However, this expansion of MOOCs utilisation has given rise to massive raw data. This big-data is an outcome of millions of records of learner's activity on various MOOC platforms, commonly known as traces. Thus, doors grew wide open to exhaustively analysing raw learner traces. In this paper we present, a preliminary analysis applied on historical learner activity traces collected from the platform of the OpenClassrooms MOOC providers.

**Keywords:** MOOC Traces, Learning analytics, MySQL Database, Learners' Categorization.

## 1 Introduction

The development rush of educational resources in institutions and universities brought great notability to Massive Open Online Courses (MOOCs). The offered alternative education method in MOOCs changed the standards of teaching and learning forever [1]. MOOCS reformed education to become attainable for the whole public at any age, price, country, time, and mean [2]. One clear criticism of MOOCs is its high rate of attrition among learners. Authors debated the accuracy of assessing MOOCs by their high attrition rates [3]. However, it is certainly not appropriate to disregard the high attrition occurrence and consider the situation satisfactory.

The overall intention behind this research is attempting to take advantage of this marked aspect of high attrition. This advantage lies in understanding and anticipating the drop-out of learners as a way to increase their motivation and decrease their drop-out. Therefore, in this paper we briefly present the basic statistical analysis applied to 60,000 learners within four MOOCs in the domains of entrepreneurship and computer science. This research is part of the Human oBservatory Based on anaLysis of E-learning traces (HUBBLE) project. Additionally, the studied learner's traces are offered by the MOOC providers and developers, OpenClassrooms (OC), being partners of this project.

## 2 Literature Review

Several useful evaluations and case studies have been published by institutions that provide MOOCs. The evaluations shed the light on the low completion rates of MOOCs that is found to be averagely less than 13% [4]. As examples of these institutions, we name MIT, the University of Edinburgh, Duke University, UK Open University, and much more [5,6,7]. Many of the available work come to findings that might differ in measures, statistical representations, and even outcomes but not in the overall condition of high attrition.

The motivation of this research lies in analysing learners' traces of OpenClassrooms to derive trajectories, patterns, and indicators to help achieving an accurate prediction of potential success or failure, continuity or attrition of learners before occurring. The first step for achieving this goal is applying a basic analysis to better understand and assess the situation of the studied traces and here lies the ongoing research presented in this paper.

## 3 Analysis Approach

After the brief review on the objective behind this research, comes the relevance for assessing the traces understudy. Therefore, in this section, we present a basic statistical analysis of the OC traces along with the details of the used tools and software. The learning process constitutes of three phases: data collection, data cleaning, and data categorization and analysis.

**Data Collection Phase**: The data is collected through an authorized access to a secured link offered by OC. This link carries the historical traces of learners in four courses. The collected data is of two formats, CSV and JSON. The CSV files include traces related to learners' activity. Whereas, the JSON files include all course related meta-data. After collection, the data is then imported into a development environment to undergo the necessary processing of the next phase.

**Data Pre-processing Phase**: This phase mainly covers verifying the structure and efficiency of the traces, and setting the cleaning rules to perform before categorization. Basically, verifying the structure and efficiency of the traces necessitates knowledge exchange with the MOOC providers, in this case OC. An initial examination of the data is then required to point out any doubtful entries or fields within the traces. The used tool for this cleaning is RStudio, which is an open-source R-based integrated development environment.

**Categorization and Analysis Phase:** This categorization is necessary to study drop-out, completion, success, and failure on different types of learners (premium and non-premium). The goal is mainly increasing revenue or decreasing loss for OC. To achieve this goal it is important to understand the real interfering reasons of drop-out and accordingly preface an enhanced learning opportunity for learners. Figure1 illustrates the scheme of the categorization phase; the red-highlighted categories represent the controversial or the most interesting cases to examine. The expectation is that the majority of premium learners tend to complete and mostly succeed the course; whereas, few of the non-premiums are expected to complete and succeed a course. Therefore, we mainly aim at studying the accuracy of these expectations in the reality of this case-study. We found that the most efficient software for applying this categorisation would be MySQL workbench that facilitates with its query based model the applied categorisation process.

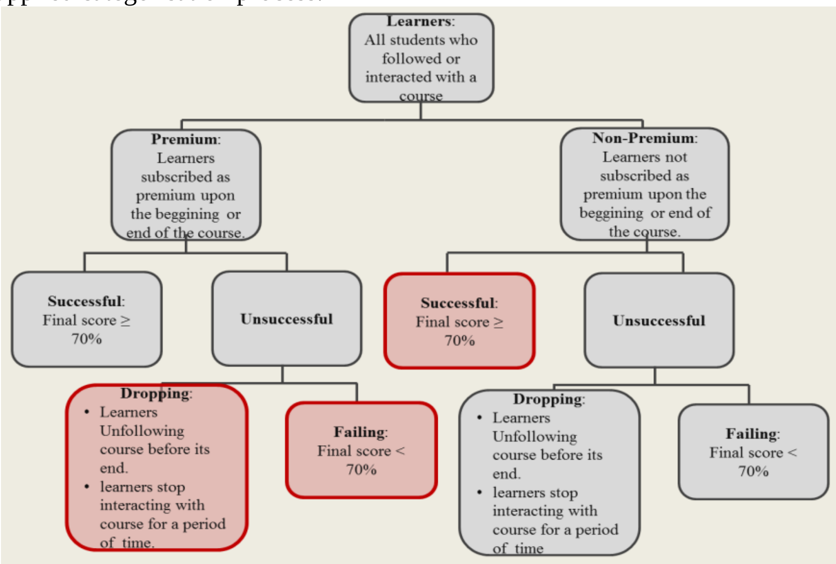


**Fig. 1.** Learner's Categorization Scheme

The studied traces were imported into a MySQL database server through MySQL workbench. This database was used then to apply the categorisation queries according to the preferences of the concerned MOOC provider and to the studied objective. Figure2, illustrates the diagram of the built database including the details of the studied traces.

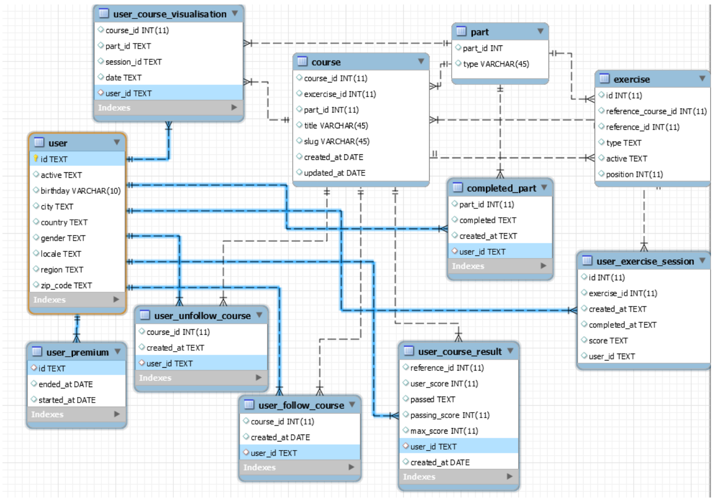

**Fig. 2.** Database Diagram of OC's Learner Traces

### 3.1 Analysis and Further Work

After applying the necessary categorisation of learners into droppers, failers and succeeders among both premium and non-premium learners, all possible indicators should be derived from the data. In this context the indicators derived from the data should correspond to the nature of data present. For example indicators on course visualisation, starting and ending time, results, exercises, etc. The next step would be choosing the most appropriate prediction algorithms to deploy on the estimates of these indicators in order to predict at-risk failing or dropping learners. The final step would be reacting efficiently to prevent the drop-out or failure of at-risk learners. Unfortunately in this paper the results of the ongoing analysis purposed cannot be presented for confidentiality reasons.

## 4 Conclusion

Mining learners' traces is proving the underlying powers of monitored on-line learning over and over. This basic assessment uncovered intriguing questions behind the manner of learners' distribution in courses. The variance of failure, success, attrition, and completion is all an influence of certain factors. It might be impossible to achieve a fully accurate forecast of attrition, but the least of prediction is considered an improvement. This grants increased revenue for MOOC providers and an improved learning experience for students.

## 5 References


1. S. Hai-Jew. IFF and other conditionals: Expert perceptions of the feasibility. Remote Workforce Training: Effective Technologies and Strategies: Effective Technologies and Strategies, page 278, 2014.
2. E. J. Emanuel. Online education: Moocs taken by educated few. Nature, 503(7476):342{342, 2013.
3. R. Rivard. Measuring the mooc dropout rate. Inside Higher Ed, 8:2013, 2013.
4. D. F. Onah, J. Sinclair, and R. Boyatt. Dropout rates of massive open online courses: behavioural patterns. EDULEARN14 Proceedings, pages 5825{5834, 2014.
5. L. Yuan, S. Powell, J. CETIS, et al. Moocs and open education: Implications for higher education, 2013.
6. Y. Belanger and J. Thornton. Bioelectricity: A quantitative approach duke university's FIrst mooc, 2013.
7. S. Cross. Evaluation of the olds mooc curriculum design course: participant perspectives, expectations and experiences. 2013.