

Task Definition in Big Sets of Heterogeneously Structured Moodle LMS Courses

Teodora Dogaru¹, Nora Götzke¹, Daniela Rotelli ², Yoel Berendsohn¹, Agathe Merceron ¹ and Petra Sauer¹

Abstract: Analysing Learning Management System (LMS) log data gives insight into student learning behaviour that can help to predict performance, and as a consequence to avoid drop-out. This contribution provides an application and an adaptation of Rotelli and Monreale’s methodology [RM22] for defining tasks in a set of 10,532 online courses collected from seven universities. Unlike [RM22], we access the log data directly from the Moodle database. Even though our data set is much bigger and more heterogeneous than the one described in [RM22], we could adapt the data selection and filtering, as well as the components’ redefinition and alignment and employ their methodology to define tasks. This work is a contribution to make log data preprocessing open, replicable and more transparent.

Keywords: LMS, Moodle log data analysis, Data preprocessing

1 Introduction

The DiSEA project is concerned with investigating factors leading to students’ success or dropout in online degree programs. To provide students with feedback on their learning behaviours in the Learning Management System (LMS) and predict their performance, one aim of the DiSEA project is to analyse their interaction in the LMS. This is done by processing and analysing the log data that the LMS, in our case Moodle, stores. Unfortunately, there is little publication on how scientists preprocess log data, even though the decisions made in the preprocessing step influence the results. In this paper, we apply and adapt the methodology proposed in [RM22] to transform Moodle log data into tasks to better grasp students’ learning behaviour, such as content or duration of their interaction with the LMS.

¹ Berliner Hochschule für Technik, Fachbereich VI, Luxemburger Straße 10, 133535 Berlin, Germany, {teodora.dogaru,nora.goetze,yoel.berendsohn,merceron,sauer}@bht-berlin.de, <https://orcid.org/0000-0003-1015-5359>

² University of Pisa, Department of Computer Science. Largo B. Pontecorvo 3, 56127, Pisa (Italy), daniela.rotelli@phd.unipi.it, <https://orcid.org/0000-0002-0943-6922>

2 From Data to Tasks

In the DiSEA project, we applied and adapted the methodology described in [RM22] to data collected via the Moodle platform that stores courses from seven universities that are part of a network of universities of applied sciences, Virtuelle Fachhochschule (VFH). Our data is vastly different with regard to data described in [RM22]. We have a high number of heterogeneous Moodle courses (>10,000) and we cannot use the log generation interface for log data extraction as done in [RM22], resulting in a more generic log that does not lend itself to easy readability. We only collect data from Moodle users that have explicitly stated their consent via a consent plugin that we specially developed. Each week, either new users consent to the use of their log data or existing participants withdraw their consent, resulting in a constantly changing data set. Because of the ever increasing, massive amount of log data, we must reduce our dataset to serve our purpose. Hence, we included only a) logs related to student activity, b) valuable Moodle components (at least 600 log records), and c) only courses with at least 75 records. After data filtering, our log contains 127 different events, categorised in 31 components. From the original 10,532 courses, only 982 are now left.

In the Moodle logging system every event is described by an *action* performed on a *component*, at a specific *time*, in a specific *course* or area of the platform. Components correspond to modules of the platform, like *dashboard*, *quiz*, *assignment* or *core*. Although they represent actions that the user performed on specific modules, some events are assigned the component *core*. To prevent information loss and a potential misunderstanding of users' behaviour, this component must be redefined. For every event assigned the *core* component that [RM22] describes and that we found in our data, we followed their example in redefining the component. Because the LMS Moodle is very flexible and feature-rich, allowing for several plugins and e-portfolios to be installed, we found events that are not described in [RM22] like (*mod_bigbluebuttonbn*, *activity*, *viewed*). These events can still be handled after the methodology described in [RM22].

Following [RM22], after component redefinition, each single record can now be grouped into so-called *tasks*. Roughly, a task is a sequence of log records with the same component for one user, that together can be construed as a singular, purposeful activity in one course or area of the platform. A task will have a number of actions, a course, a starting time and a duration. Once the list of components is established, the methodology of [RM22] to define tasks can be applied as is.

This work complements the study of [RM22], contributes to make log data preprocessing open, replicable and more transparent.

Bibliography

- [RM22] Rotelli, D.; Monreale, A.: Time-on-Task Estimation by data-driven Outlier Detection based on Learning Activities. In: LAK22, *doi*: 10.1145/3506860.3506913