Analytics procedures and tool on Moodle:

for Predicting Student Learning Performances

1. Introduction

In this TDG project, Moodle log data of previous deliveries of participating courses are collected and analyzed, with the goal of building prediction models based on student behaviors on Moodle. Multiple prediction models were tried and compared, including Linear Regression, Support Vector Machines, Logistic Regression, and Lasso Regression. A finalized prediction model was based on Linear Regression with Step-wise backward feature selection. This model can partially predict student course performances (e.g., total scores), where a rudimentary analytic tool in the form of a dashboard is developed for Moodle.

2. Prediction Model

To build the prediction model, we used the dataset of Moodle logs from the course CCST9003, delivered in the year of 2014 - 2015. The following model achieved an R^2 score of 0.695, which means about 70% variance of the data can be explained by this model. According to the literature (Hu et al., 2017), better prediction performances are possible but mostly for the contexts of online learning (e.g., online courses, MOOC[1]s) where learning activities mostly occur in the online learning systems (e.g. LMSs or MOOC platforms). Given that our Common Core courses are all in the face-to-face instruction mode, this and other similar prediction models would be much more useful for HKU MOOCs or SPOCs[2].

The finalized prediction model can be presented as the following linear equation:

overall score = 0.016 * course_viewsection +
(-0.524) * turnitintool_viewall +
7.478 * forum_viewforums +
(-0.194) * url_view +
4.053 * assign_submit +
0.751 * quiz_closeattempt +
0.771 * quiz_attempt +
(-16.014) * blog_view +
(-1.507) * course_recent +
0.152 * quiz_continueattempt +
0.062 * wiki_edit +
(-3.848) * assign_viewsubmitassignmentfor +
45.812

Each item, as shown in the equation, is the count of logs with the combination of course Component (i.e., course page, Quiz) and Action on the component (e.g., view, submit, etc.) on Moodle. For example, "course_viewsection" means the count of logs from Moodle databases, with Component = *course*, while Action = *view section*. In those cases where students do not have such action on this component, the model will ignore that item. The following table shows what each item is corresponding to.

[1] Massive Open Online Courses
[2] Small Private Online Courses

Table 1. Components and actions in the predictive model

Component	Action	Item in the Prediction Model
Course Moodle page	View Section	courseviewsection
URL*	View	urlview
Blog	View	blogview
Recent Activity	View	courserecent
Assignment	Submit	assignsubmit
	View submission	assignviewsubmitassignmentfor
Turnitin	View	turnitintoolviewall
Quiz	Start an attempt	quizattempt
	Continue an attempt	quizcontinueattempt
	Finish an attempt	quizcloseattempt
Forum	View	forumviewforums
Wiki	Edit	wikiedit

*It is noteworthy that the URL component means any direct external links (e.g., to a YouTube video, etc.) created by the instructor in charge of the course Moodle.

Besides predicting the overall scores, models were built for predicting each sub-score included in the course assessment, for the same course delivered in 2013 and 2014. Table 2 shows the results.

Table 2. Prediction Performance (R^2) on sub-scores of CCST9003 in year 2013 and 2014

	Homework	Quiz	Tutorial	Group Wiki	Group Presentation	Individual Essay	Individual Presentation
2013 N=104	0.40	0.37	0.15	0.25	0.21	0.09	0.12
2014 N= 152	0.54	0.59	0.33	0.40	0.36	0.23	0.24

A general trend can be observed from Table 2 that an assessment component would have a higher prediction performance if it involved more online activities. For example, Quiz was conducted directly on Moodle, and thus nearly all student behaviors with Quiz were recorded (e.g., view, start an attempt, finish an attempt, etc.). In contrast, assessment components such as individual essay and presentation involved little online activities, and they were harder to predict from Moodle logs. This is reasonable as learning activities of these components were hardly captured by Moodle logs. Future work could explore the connection between digital and physical learning environments (Martinez-Maldonado & Hernandez-Leo, 2016).

3. System Architecture

This analytic tool adopts the Model-View-Controller (MVC) paradigm. The view is responsible for all visualizations while the controller requests data from the server using AJAX and feeds the data to the view. The server side is mainly responsible for calculation of the

model, using logs recorded in Moodle. The predicted result of a students' performances is stored in the database at the server side for efficient retrieval.

The User Interface of the dashboard is shown in Figure 1. The tool is visualized as a circled block titled "Grade_Prediction" on a course Moodle page. It is like other Moodle blocks whose location on the page can be adjusted by the instructor. The block shows an estimated score given by the predictive model.

Figure 1. Screenshot of the analytic tool

4. Reference

Hu, X., Cheong, C. W. L., Ding, W., & Woo, M. (2017). A Systematic Review of Studies on Predicting Student Learning Outcomes Using Learning Analytics, In Proceedings of the 7th International Conference on Learning Analytics and Knowledge, LAK '17.

Martinez-Maldonado, R. and Hernandez-Leo, D. (2016). Learning Analytics Across Physical and Digital Spaces 2016, *CEUR-WS*, Vol-1601 urn:nbn:de:0074-1601-3 ISSN 1613-0073