The following publication Y. Yang, H. Wu and J. Cao, "Smartlearn: Predicting learning performance and discovering smart learning strategies in flipped classroom," 2016 International Conference on Orange Technologies (ICOT), 2016, pp. 92-95 is available at https://dx.doi.org/10.1109/ICOT.2016.8278987

SmartLearn: Predicting Learning Performance and Discovering Smart Learning Strategies in Flipped Classroom

Yu Yang
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
Email: csyyang@comp.polyu.edu.hk

Hanqing Wu
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
Email: cshwu@comp.polyu.edu.hk

Jiannong Cao
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
Email: csjcao@comp.polyu.edu.hk

***Abstract*—In flipped classroom, students are expected to learn new contents in online learning system before attending offline classes to reinforce their knowledge. This online and offline blended education model has become more and more popular. However, spending more time to actively engage in online learning does not result in better learning performance, so that how to wisely arrange online learning plan is a big challenge. In this paper, we build a LASSO model to accurately predict students' performance in course projects and their final grade by online learning behaviour data in flipped classroom. The LASSO selected features show that learning online between first and second flipped classes after midnight, and during the second flipped class would benefit students' project performance but studying one day before the examination and studying at night is counterproductive. Our results provide novel insight into guiding students to learn wisely and perform better in flipped classroom.**

***Index Terms*—Learning Analytics, Flipped Classroom, Online Learning, Behaviour Trends, Academic Performance.**

I. Introduction

In recent years, learning analytics has become a research hotspot, which is emerging to be an important area to improve students learning experience and progression. The foundation of learning analytics is massive available educational data[1]. Online learning system, student information system, sensors, and mobile devices now carry rich learner-produced data trails and activity streams, which provides a more complete picture on what is actually happening in the teaching and learning process than traditional measures such as grades and test scores, which only measure outcomes. Moreover, learning analytics is focused on building systems able to adjust teaching and learning content, program and other personalized services by capturing, modeling, processing and analyzing on huge volume of data[2]. Thus, learning analytics seeks to discover the valuable learning pattern from education big data source, predicting users' behaviour and performance, and optimizing teaching and learning contents and process in order to upgrade instructional quality.

The widespread popularity of the Internet and increasing volume of data are changing our life-style as well as education mode. A new model of teaching named Massive Open Online Courses (MOOCs) provides tens of thousands of classes in different disciplines taught by teachers all over the world and allows users to participate learning online at anytime and anywhere. It frees people from the classroom and the control of instructors[3]. Students could flexibly arrange their self-paced learning and reach multi-source supplementary materials like videos, readings, which is hardly provided in traditional face-to-face classroom teaching.

Although MOOCs has its advantage, some MOOCs organizers argue that most MOOCs performs well in organisation and presentation of course material but achieve poor instructional quality because it lacks offline in-class interaction[4]. One possible solution is blended learning which combines aspects of online learning and offline face-to-face instruction[5]. Students could flexibly allocate their time and energy to finish learning new knowledge online in MOOCs and then attend offline traditional classes to reinforce their knowledge by "Flipped Classroom" arrangement[6]. Offline teaching and learning process could be easily controlled and well organized by experienced teachers but how to instruct students to efficiently study online is a big challenge. In this paper, we analyse students' online learning behaviour in flipped classroom and predict their learning performance in course project as well as their final grade in order to explore insights on how to perform better in flipped classroom. The contributions of our work include 1) we predict students performance in course projects with high accuracy of 0.9471 under 5-fold cross-validation and, 2) we provide a new observation in teaching students to wisely study in the online part of flipped classroom.

II. Related Work

There are two camps of researchers who are working on education data mining and learning analytics. One camp of researchers focus on discovering new observation and confirmation of common senses in education in order to better understand teaching and learning process. EDUM system measured university students' punctuality (attendances, late arrivals, and early departures) in face-to-face teaching courses by using longitudinal WLAN data and evaluated the attractiveness of lectures through mobile phones interactive states [7]. Their results confirm that Wednesday is the most hard-working day and class attendance is the highest in the morning and drop as the day progress. Higher GPA students attended more class but went to class later than low performance students. Students were more easily to distract as the day progresses. Gloria Mark etc.[8] found that students who slept less felt higher productivity and study pressure. The more sleep debt, the more Facebook use and resulted in more serious negative mood.

In the other camp, researchers aim to explore the potential factors which positive or negative influences learning performance. The marks of quizzes or examinations are usually adopted in order to evaluate the learning performance. [9] applied web usage mining in e-learning system and adopted several well-known classifiers to predict students' final examination scores. SmartGPA[10] built a LASSO model to predict students' GPAs from StudentLife dataset which contained 10 weeks students' self-report data of workload on stress, sleep, activity, mood, sociability, mental well-being and academic performance of a single class collected by a mobile application[11]. It pointed out that students who have higher GPA spent more time on conscientiously study under positive moods and the feeling of stress declined as the term progress. He etc.[12] regarded learning performance as students' engagement in activities of watching or downloading lectures, attempting assignments or quizzes in MOOCs to identify students in risk of non-completing courses.

III. Flipped classroom design and data sources

The flipped classroom arrangement included two parts: 1) online learning in OpenEdX and, 2) face-to-face flipped classes. Students are expected to go through the materials of the online learning units and complete relevant exercises themselves before attending the offline flipped classes. Their click streams and the lecture video consumption on the OpenEdX will be captured and be used to measure their online learning behaviour.

A. Flipped Classroom Design

Two consecutive classes in Database System course for the second year undergraduate students are selected to deploy flipped classroom arrangement. These two flipped classes contain three units of contents including Entity-relationship Diagram(ERD), Normalization and B+ Tree. Each unit is separated into three to five subsections which contained a video plus some games, exercises or a discussion forum. Before starting online learning, students should finish a pre-quiz which served as a self-assessment in order to help them recall the memory of previous learning knowledge. For the wrong answers of questions in quizzes or exercises, students are able to immediately review the notes and reinforce their knowledge. Their online learning activities and performance on quizzes and exercises would not be counted into their final grade. Their final learning performance depends on the scores of final examination and their course projects. The scope of the final examination covers all contents teaching in the Datebase System course during the entire semester but the course projects only related to the contents in flipped classroom so that the project marks are more reliable to evaluate students' performance in flipped classroom.

In the offline flipped classes, a brief review of the contents of the online learning unit was provided at the beginning in order to ensure all students had recalled the unit. Students were asked to finish the individual exercises with comments from the lecturer so as to reinforce their knowledge. Finally, there was a group project for students to solve problems which are related to individual exercise. More details could be found in [13].

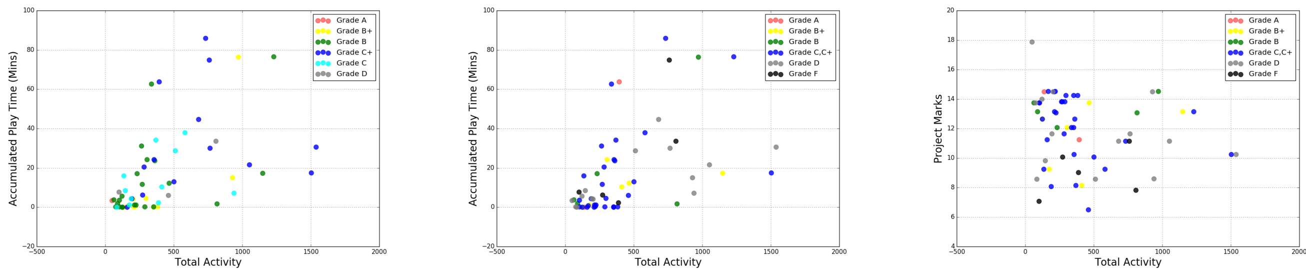
B. Online Learning Behaviour Data and Preliminary Analysis

There are 102 students enrolling in the Database System course and 79 of them voluntarily registered to the online learning system. Only 67 students have logined into the online learning system at least once but 6 of them do not have any records in watching any lecture videos, which mean that they did not indeed participate in the flipped classroom. So the remaining 61 students are considered as the participants in our flipped classroom study. Besides students' project mark and final mark, their online activity logs data including click-stream data and lecture video consumption data are captured for measuring their online learning behaviour. More specifically, click-stream data carries both video watching behaviour data and student active period data, which consists of total number of click-stream activity, total play and pause actions, total automatic end for each video, click stream in different period of two flipped classes, activity rate in three days before and after both flipped class, activity rate when 1/3/5 days before exam, and click-stream occurring in daytime, at night as well as overnight. Video consumption data covers video play time of each 13 lecture videos and the total accumulated play time for each participant.

TABLE I: Numerical Characteristic of Online Learning Behaviour Data

Data Field	Min	Max	Average
Play	1	912	121.3934
Pause	0	161	27.6066
Finish	0	12	2.3443
Before First Class	0	111	3.0656
During First Class	0	124	2.0328
Between 1st and 2nd Class	0	850	168.5738
During Second Class	0	62	3.7541
After Second Class	0	1502	243.3607
Three Days before Class	0	472	36.0984
Three Days after Class	0	82	4.1639
One Day before Exam	0	387	56.4426
Two to Three Day before Exam	0	897	66.1803
Four to Five Day before Exam	0	413	10.2787
Overnight (00:00 am - 08:00 am)	0	841	89.1639
Daytime (08:00 am - 16:00 pm)	0	520	124.0328
Night (16:00 pm - 00:00 pm)	0	1502	207.5902
Total Click Activity	49	1537	420.7869
Accumulated Video Play Time (Mins)	0.0199	85.8465	17.0262

The numerical characteristic of behaviour data is showed in Table I. Students spent an average of 17.0262 minutes in



(a) Total Activity to Accumulated Play Time (Mins) with Color Grouped by Grade of Project Mark

(b) Total Activity to Accumulated Play Time (Mins) with Color Grouped by Final Grade

(c) Total Activity to Project Marks with Color Grouped by Final Grade

Fig. 1: Students' Engagement and Learning Performance

watching lecture videos. Most student participated in online learning between first and second classes and after the second class. Students were active before the final examination which meant that they aimed to used online materials for exam preparation. However, both project marks and final grades have no direct relation with total click activity or accumulated video play time as showed in Fig. 1, which means that spending more time on watching lecture videos and engaging more in online learning do not result in better learning performance. So students should be smart and wise to learn online. Now, the question is how to arrange the online learning schedules to achieve the best performance in flipped classroom. In next section, we will build a LASSO model to predict students' learning performance and discover insight to enhance their online learning process.

IV. Performance Prediction Model

Inspired by SmartGPA in [10], we build a LASSO (Least Absolute Shrinkage and Selection Operator) model to predict student's course project marks and final scores, which has ability to indicate the valuable predictors by sparseness simultaneously. LASSO is a linear regression model with ℓ_1 regularization, which minimized the sum of squared errors. The optimization problem could be formulated as:

$$\underset{\alpha}{\operatorname{argmin}}\{\frac{1}{2}\|A\alpha - b\|_2^2 + \lambda\|\alpha\|_1\} \tag{1}$$

where $A \in \mathbb{R}^{n\times m}$ is a behaviour features matrix for n participants and each of m columns carried one behaviour feature; $b \in \mathbb{R}^n$ is a vector of students' learning performance like project marks; $\alpha \in \mathbb{R}^m$ is the discriminant vector, which will be sparse in iteration and indicated valuable features for predicting learning performance. The regularization parameter $\lambda > 0$ is selected by cross-validation. This optimization problem could be easily solved by proximal algorithm and iterative solution in $t \geq 0$ iteration is as follow:

$$\alpha^{t+1} = \mathbf{prox}_{\lambda\|\cdot\|_1}\left(\alpha^t - A^T(A\alpha^t - b)\right) \tag{2}$$

where

$$\mathbf{prox}_{\lambda\|\cdot\|_1}(x) = (x-\lambda)_+ - (-x-\lambda)_+ \tag{3}$$

and $\alpha^0 = \begin{bmatrix} 1/m & \cdots & 1/m \end{bmatrix}^T$; $(\cdot)_+$ is a vector operator setting non-positive elements to zero; T is matrix transposition; x is a column vector. LASSO model automatically selects a minimum number of valuable features and discords redundant ones to avoid overfitting.

V. Experiment and Results

In experiment, we repeat 100 times 5-fold cross-validation to verify the performance of LASSO model in predicting students' project marks and final marks. In each times of experiment, the full data set will be first randomly and evenly divided into 5-fold. Each fold of data takes turns to serve as test set to test prediction performance. The prediction accuracy Acc is calculated by one minus the mean absolute error (MAE) given by

$$\mathrm{Acc} = 1 - \frac{1}{n}\|A\hat{\alpha} - \hat{b}\|_1 \tag{4}$$

where $\hat{\alpha}$ is the optimal output discriminant vector which is trained by the other 4-fold data; $A\hat{\alpha}$ is predicted learning performance and $\hat{b}$ is groundtruth. In order to handle the data unbalance issues and make the weight regularization work properly, ℓ_2 normalization is applied to each column of the feature matrix A. Selected features coincide with the non-zero elements of optimal discriminant vector $\hat{\alpha}$. The range of project marks is from 0 to 20 but the range of final grade is $\{A, B+, B, C+, C, D, F\}$. Because the final grades are categorical text, we convert them into numerical scores range in $\{90, 85, 80, 75, 70, 60, 30\}$ before training the prediction model. The best prediction result and its corresponding average accuracy of 5-fold cross-validation are shown in Table II. Table III and table IV show selected features by LASSO to predict project marks and final scores when achieves the best accuracy of prediction.

TABLE II: Learning Performance Prediction Results

	Best Accuracy	Average Accuracy
Project Marks	0.9471	0.9029
Final Scores	0.8681	0.7824

TABLE III: LASSO Selected Features for Predicting Project Marks

Selected Activity	Nonzero Weight	r value	p value
Between 1st and 2nd Class	0.0740	0.4239	<0.001
During 2nd Class	0.0975	0.1767	0.1527
In One Day before Exam	-0.0372	-0.2840	0.0198
Over Night	0.0204	0.2058	0.0947
At Night	-0.0537	-0.0941	0.4486

TABLE IV: LASSO Selected Features for Predicting Final Scores

Selected Activity	Nonzero Weight	r value	p value
Over Night	-0.0514	-0.2622	0.0321
Watching ERD Course 5	-0.1384	-0.1426	0.2497

The highest prediction accuracy of project marks is 0.9471 with students activities between first and second flipped classes, during second flipped class, in one day before exam, activities at night and overnight. The weights indicate the strength of the activities affecting students' performance in course projects. Interestingly, students who perform better in course projects are more likely to learn online overnight between first and second flipped classes. However, using the online resource to prepare the examination in the next day is not a wise choice, which strongly negative correlates to the project marks with $r = -0.284$ and $p = 0.0198$. Those students who are far more active in online learning platform one day before examination might not solidly grasp knowledge in class of entire semester. It might be already too late for them to improve themselves and perform better in course projects which should be finished before the final examination. Another noteworthy finding is that students who learn online after midnight perform better than studying at night.

Predicting accuracy of final scores is much lower than predicting the project marks using LASSO, which is 0.8681 of the best prediction and 0.7824 on average. Meanwhile, the optimal discriminant vector does not contain any positive weight values. In other words, non-activities positively contribute to final performance according to LASSO, which does not make sense. One possible reason is that the final scores evaluate students performance comprehensively, which measures more contents in the Database System course. Students' behaviour in the online learning part of flipped classes might not be sufficient enough to evaluate their comprehensive learning performance in the entire semester.

VI. Conclusion and Future Work

By analysing students' online learning behaviour in flipped classes and their learning performance, we found that spending more time on watching lecture videos and engaging more in online learning do not give rise to better learning outcomes in flipped classroom. We build a LASSO model to predict students' learning performance in course project and their final grade by click-stream data and lecture videos consumption data, which achieves accuracy of 0.9471 in predicting their course project marks and 0.8681 in predicting final score. Meanwhile, our measurement results show that engaging the online learning between first and second flipped classes after midnight, and during the second flipped class are strongly positive correlated to students' course project performance, but studying one day before examination and studying at night are counterproductive. This insight could help students wisely arrange their learning strategies in order to achieve the best learning performance in flipped classroom.

In future, we will extend the flipped classroom arrangement to multiple disciplines in the entire semester and capture self-report emotion data to effectively measure students' learning efficiency and their emotion variation during semester.

Acknowledgment

The work described in this paper was fully supported by the funding for demonstration project on large data provided by The Hong Kong Polytechnic University (project account code: 1-9A5V).

References

[1] M. Chatti, A. Dyckhoff, U. Schroeder, and H. Thus, "A reference model for learning analytics," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5/6, pp. 318–331, 2012.

[2] T. Elias, "Learning analytics: Definitions, processes and potential," *Learning*, vol. 23, pp. 134–148, 2011.

[3] D. McConnell, *Implementing Computer Supported Cooperative Learning*. Kogan Page Limited, 2010.

[4] A. Margaryan, M. Bianco, and A. Littlejohn, "Instructional quality of massive open online courses (moocs)," *Computers and Education*, vol. 80, pp. 77–83, 2015.

[5] D. Bruff, D. Fisher, K. McEwen, and B. Smith, "Wrapping a mooc: Student perceptions of an experiment in blended learning," *MERLOT Journal of Online Learning and Teaching*, vol. 9, no. 2, pp. 187–199, 2013.

[6] F. Alonso, G. Lpez, D. Manrique, and J. M. Vines, "An instructional model for web-based e-learning education with a blended learning process approach," *British Journal of Educational Technology*, vol. 36, no. 2, pp. 217–235, 2005.

[7] M. Zhou, M. Ma, Y. Zhang, K. Suia, D. Pei, and T. Moscibroda, "Edum: classroom education measurements via large-scale wifi networks," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2016, pp. 316–327.

[8] G. Mark, Y. Wang, M. Niiya, and S. Reich, "Sleep debt in student life: Online attention focus, facebook, and mood," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI)*, 2016, pp. 5517–5528.

[9] C. Romero, P. Espejo, A. Zafra, J. Romero, and S. Ventura, "Web usage mining for predicting final marks of students that use moodle courses," *Computer Applications in Engineering Education*, vol. 21, no. 1, pp. 135–146, 2013.

[10] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell, "Smartgpa: how smartphones can assess and predict academic performance of college students," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2015, pp. 295–306.

[11] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2014, pp. 3–14.

[12] J. He, J. Bailey, B. I. P. Rubinstein, and R. Zhang, "Identifying at-risk students in massive open online courses," in *Proceedings of the 2015 AAAI Conference on Artificial Intelligence (AAAI)*, 2015, pp. 1749–1755.

[13] V. Ng, L. R. Huang, C. H. Lung, and K. Liu, "Are my students ready: A case of flipped learning in an it subject," in *Proceedings of the 2016 International Conference on e-Learning (ICEL)*, 2016, pp. 107–174.