Modeling and Predicting Learning Behavior in MOOCs

Jiezhong Qiu[†], Jie Tang[†♯], Tracy Xiao Liu[‡], Jie Gong[⋆],
Chenhui Zhang[†], Qian Zhang[†], and Yufei Xue[†]
[†]Department of Computer Science and Technology, Tsinghua University
[♯]Tsinghua National Laboratory for Information Science and Technology (TNList)
[‡]Department of Economics, School of Economics and Management, Tsinghua University
[⋆]NUS Business School, National University of Singapore
qjz12@mails.tsinghua.edu.cn, jietang@tsinghua.edu.cn, liuxiao@sem.tsinghua.edu.cn, gong@nus.edu.sg

ABSTRACT

Massive Open Online Courses (MOOCs), which collect complete records of all student interactions in an online learning environment, offer us an unprecedented opportunity to analyze students' learning behavior at a very fine granularity than ever before.

Using dataset from xuetangX, one of the largest MOOCs from China, we analyze key factors that influence students' engagement in MOOCs and study to what extent we could infer a student's learning effectiveness. We observe significant behavioral heterogeneity in students' course selection as well as their learning patterns. For example, students who exert higher effort and ask more questions are not necessarily more likely to get certificates. Additionally, the probability that a student obtains the course certificate increases dramatically (3× higher) when she has one or more "certificate friends".

Moreover, we develop a unified model to predict students' learning effectiveness, by incorporating user demographics, forum activities, and learning behavior. We demonstrate that the proposed model significantly outperforms (+2.03-9.03% by F1-score) several alternative methods in predicting students' performance on assignments and course certificates. The model is flexible and can be applied to various settings. For example, we are deploying a new feature into xuetangX to help teachers dynamically optimize the teaching process.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Sociology; H.2.8 [**Database Applications**]: Data Mining

Keywords

MOOCs, Predictive model, User behavior, Online engagement

1. INTRODUCTION

Massive open online courses (MOOCs) have become increasingly popular and offered students around the world the opportunity to take online courses from prestigious universities. Three pioneer MOOC platforms—Coursera, edX, and Udacity—offer hundreds of courses and draw more than 100,000 registrants per course [25]. 2012 was called "The Year of the MOOC" by New York Times [23]. Following the three pioneers, many other platforms have also been developed quickly around the world, such as Khan Academy in North America, Miriada and Spanish MOOC in Spain, Iversity in German, FutureLearn in England, Open2Study in Australia, Fun in France, Veduca in Brazil, Schoo in Japan, and xuetangX in China. MOOCs are without doubt leading a revolution of education, by gathering global education resources and restructuring the learning environment, e.g., providing online forums to geographically-dispersed students. Furthermore, as all students' learning behavior occurs online, it enables us to evaluate students' performance in a more objective and quantitative way. For example, KDD CUP 2015 used the MOOC data to offer a challenge of predicting students' dropout rate among a number of courses.[1]

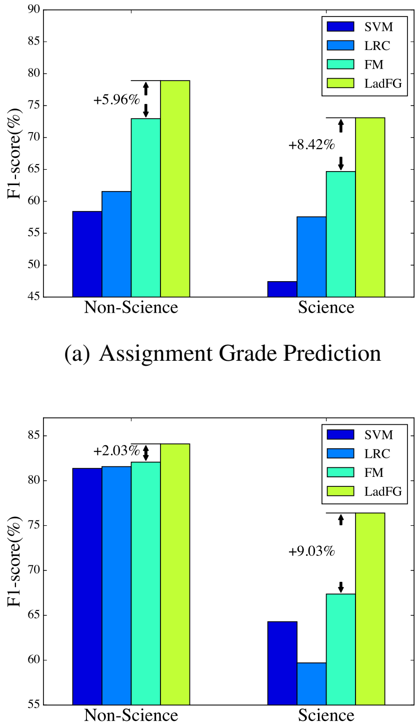
MOOC platforms collect complete records of students' online activities, which enables us to take a closer look at students' learning behavior than ever before. However, new challenges have also emerged. For example, the low completion rates of MOOC participants—preliminary statistics show that less than 5% of the participants have completed a course—has been a central criticism [16, 25]. Moreover, MOOC is not just a place for teaching or learning. It also provides an interactive platform to support group-level interactions among students, lecturers, and teaching assistants. Students with various motivations to study MOOCs [2] have very different learning behavioral patterns, and these online learning behavior may also deviate from those in traditional learning environments. Therefore, there is a clear need to understand user behavior in MOOCs, and more importantly, to design effective mechanisms to motivate more participation in both courses learning and social interaction.

In this paper, we focus on studying how students engage in MOOCs and to what extent we can predict their learning behavior. Understanding the complex and subtle forces underlying the learning process can significantly help design better courses and improve the learning effectiveness. More specifically, how to retain students in a course? How to estimate the completion rate of a course? How to evaluate the learning performance of different students? Despite several relevant studies, such as course completion analysis [16], learning behavior analysis [25], and student classification and engagement analysis [2], there are few systematical studies on modeling students' learning behavior for different categories of courses.

Employing xuetangX,[2] one of the largest MOOC platform from China, as the basis of our study, we systematically investigate the aforementioned problem. We first conduct an in-depth analysis to

WSDM'16, February 22–25, 2016, San Francisco, CA, USA.
© 2016 ACM. ISBN 978-1-4503-3716-8/16/02 ...$15.00.
DOI: http://dx.doi.org/10.1145/2835776.2835842.

[1] http://kddcup2015.com
[2] http://www.xuetangx.com



(a) Assignment Grade Prediction

(b) Certificate Earner Prediction

Figure 1: Performance of Learning Behavior Prediction by Different Methods. (a) Assignment grade prediction, and (b) Certificate earner prediction. Please refer to § 4 for definitions of the comparative methods.

understand the degree to which user demographics (gender, age, and education background) influence students' forum activities and learning behavior including watching videos and working on assignments. Moreover, we estimate each of the three factors' importance on students' learning performance, e.g., the likelihood of getting certificates. We have several intriguing discoveries. First, hard working and frequent questioning do not necessarily imply high learning performance. Second, engaging in the course forum is a significant indicator for students' learning performance. Third, homophily [18] — the idea that similarity and connection tend to co-occur, or "birds of a feather flock together" [20] — plays an important role for predicting students' learning behavior. The probability that a student obtains the course certificate increases dramatically (3× higher) when she has one or more "certificate friends".

We then propose a latent dynamic factor graph (LadFG) to model and predict learning behavior in MOOCs. The model incorporates students' demographics, forum activities, and learning behavior into a unified framework. It captures the dynamic information and homophily correlations between students, and projects students' learning behavior into a latent continuous space. Although the model is general enough to be applied to various settings, we particularly consider two prediction tasks: assignment grade prediction and certificate earner prediction. The former predicts students' performance on assignment, while the latter predicts which students could get the certificates after a course ends. Figure 1 shows the performance of different comparison methods for the two prediction tasks on both science courses and non-science courses. Clearly, the proposed method improves the prediction accuracy by up to 9.03% over the alternative methods. Based on the results of the proposed model, we are deploying a new feature into the xuetangX system to help lecturers optimize their teaching process.

Table 1: The Description of the Dataset.

Category	Type	Number
Course	All	11
Science	CS	3
	EE	2
Non-Science	Economics	2
	History	3
	Sports	1
User	Total #	88,112
	Max #students/course	31,120
	Min #students/course	2,631
Forum	# new posts	13,021
	# replies	16,367
Activity	Activity types	21
	# Activity logs	56,800,000

Note: the data spans from Oct 10, 2013 to Jun 26, 2014.

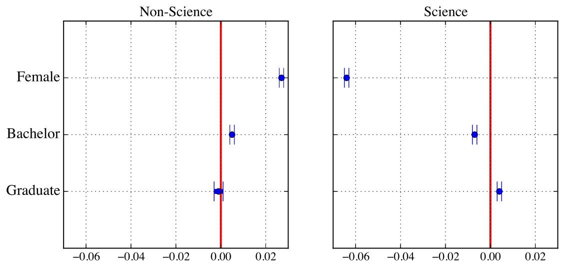


Figure 2: Regression Analysis for Course Selection by the Linear Probability Model.

Data. Our data comes from xuetangX, a partner of edX. The system was launched in October 2013 and up to November 2015, it has offered 670 courses (including courses from Tsinghua University, Peking University, and edX courses from MIT, Stanford, UC Berkeley, etc.) and attracted 1,700,000 registered users. The dataset used in this paper consists of 11 completed courses in the Fall 2013 and Spring 2014 semesters. We categorize the courses into two types: science (Computer Science and Electronic Engineering) vs. non-science (Economics, History, and Sports) courses. The average number of registered users for the science courses is lower than that of the non-science courses, though the difference is not statistically significant (5,911 vs. 12,959, $p = 0.247$, two-sided t-tests). Each course has a discussion forum for students to post/reply questions and to interact with each other. The dataset also consists of multiple types of students' activities such as watching videos, working on assignments, downloading resources and etc. In total, there are 56,800,000 time-stamped activity logs. Table 1 reports summary statistics of the dataset.

We first conduct a regression analysis to examine correlation between student demographics and course selection. Regarding demographics, we consider gender, education (graduate degree including master and PhD, bachelor, and those with degree below bachelor), and age. As age and education are significantly and positively correlated ($p = 0.000$, chi-square test), we focus on education here. Figure 2 shows regression results from a linear probability model of course enrollment.[3] The dependent variable is whether user v_i registers a course. The independent variables are (1) female dummy and (2) education level dummies including bachelor, graduate, and the omitted variable is bachelor below. Additionally, we control for age and course characteristics by including course category and estimated weekly hours in learning. From the results, we observe that compared to male students, females are significantly more likely to take non-science courses ($p < 0.01$), and less likely to choose science courses ($p < 0.01$). Moreover, compared to users with low education (<bachelor), bachelors are significantly more likely to take non-science courses ($p < 0.01$), and significantly less likely to choose science courses ($p < 0.01$). In contrast, graduate students are significantly more likely to take science courses ($p < 0.05$).

Organization. Section 2 presents pattern analysis for student activities; Section 3 formulates the problem and presents the proposed model; Section 4 presents the experimental results; Section 5 discusses related work and Section 6 concludes the work.

2. PATTERNS OF STUDENT ACTIVITY

In this section, we investigate students' learning activities including forum activities and time spent on videos and assignments. Moreover, we examine how each of these factors affects a user's likelihood of getting certificates, i.e., a student's final grade in a course is at least 60 out of 100.

Several common features apply throughout our regression analysis. First, robust standard errors are reported in parentheses. Second, two-sided p-values are reported and significant at: * 10%; ** 5%; *** 1%. Third, age and course type dummies are controlled in all regressions. Finally, in all regression analyses, we consider ordinary least squares (OLS) models.

2.1 Learning Activity Patterns

We first study users' participation pattern of forum activities including posting new threads and replying to questions. Second, we study their learning behavior for watching videos and doing assignments.

Engagement patterns of forum activity. In course forums, students can post new questions and answer existing questions. Overall, the level of forum activities is very low: 94% users in our sample never participate in posting or replying to questions. Additionally, among active users, their forum activities decrease with time ($p < 0.05$) for all but new post in science courses ($p > 0.1$), suggesting that users' participation enthusiasm in courses decays over time. This observation is consistent with that in previous studies [2, 25].

We present ordinary least squares (OLS) estimates of the relationship between student characteristics and the number of forum activities in Table 2. The dependent variable is the number of new posts (Columns 1 and 3) or replies (Columns 2 and 4) per student, and the independent variables include gender, education level, and the required effort level for a course (specified by the teacher). It is interesting that women are only more active in asking questions in non-science courses, while they are much more quiet in science courses. Within non-science courses (Columns 1 and 2), women post significantly more questions than men ($p < 0.01$), though they reply marginally significantly fewer questions ($p < 0.1$). In contrast, their amount of new post in science courses is significantly less than male students (Column 3, $p < 0.05$).

Table 2: Regression Analysis for Forum Activities.

	Non-Science		Science	
	New Post	Reply	New Post	Reply
	(1)	(2)	(3)	(4)
Female	0.089***	-0.024*	-0.026**	-0.053
	(0.013)	(0.013)	(0.011)	(0.145)
Bachelor	0.029***	0.007	0.015	0.074*
	(0.011)	(0.010)	(0.010)	(0.045)
Graduate	0.001	0.007	0.016	0.306*
	(0.016)	(0.030)	(0.013)	(0.156)
Effort	0.277***	-0.053***		
	(0.058)	(0.017)		
Constant	-0.534***	0.228***	0.050**	0.092
	(0.122)	(0.054)	(0.022)	(0.059)
Obs.	74,480	74,480	19,269	19,269
R^2	0.013	0.001	0.001	0.002

Note: Constant—the learned offset by the regression model; Obs.—the number of observations in each category; and R^2 — the proportion of variance in the criterion that is explained by the estimated regression model.

Table 3: Regression Analysis for Effective Learning Time.

	Non-Science		Science	
	Video	Assignment	Video	Assignment
	(1)	(2)	(3)	(4)
Female	8.588***	3.985***	-7.890***	-4.793**
	(1.181)	(0.400)	(2.281)	(2.277)
Bachelor	1.019	1.123**	8.032***	8.946***
	(1.389)	(0.454)	(1.876)	(1.898)
Graduate	-5.618***	-1.918***	6.945***	4.817**
	(1.774)	(0.567)	(2.585)	(2.247)
Effort	-29.489***	1.895***		
	(1.345)	(0.521)		
Constant	101.258***	-4.246***	26.746***	22.566***
	(4.970)	(1.538)	(4.241)	(6.293)
Obs.	74,480	74,480	19,269	19,269
R^2	0.035	0.021	0.002	0.002

Pertaining to forum activities between students with different education levels, we find that bachelors post significantly more questions in non-science courses (Column 1). In contrast, graduate students do not ask many questions while their amount of replies in science courses is marginally significantly higher than those with degrees below bachelor (Column 4, $p < 0.1$). The effect size is also higher than that for bachelors (0.306 vs. 0.074, $p = 0.105$), which indicates that people with different education levels may play different roles in forums. In addition, within non-science courses, users post more questions when a course requires more working hours (Column 1, $p < 0.01$). However, the amount of replies is significantly negatively correlated with course effort requirements (Column 2, $p < 0.01$). One explanation is that more difficult courses may induce more questions, while these questions may be more challenging for users to answer.

Engagement patterns of videos and assignments. First, we define user v_i's effective learning time below.

[3] We present the linear probability model as it is easier for results interpretation. All results are robust when probit or logit models are used.

Definition 1. **Effective learning time.** It represents the actual (or valid) time that a student spends on watching videos and working on assignments.

In practice, it is difficult to accurately measure students' study time [8]. For example, after a student clicks a video, she may leave and work on something else. Therefore, we design an algorithm based on deterministic finite automaton to approximate effective learning time. Specifically, we define three states: *idle*, *video*, and *assignment*, in the state automaton. The automaton starts with idle, and changes states when it receives certain activities triggered by students. For example, when the student triggers a "play video" activity, the state changes from idle to video. While a "pause" activity changes the state back to idle. Altogether, the duration between two activities will be counted as the effective learning time. In addition, there is threshold-triggered transition from any states to idle. The idea is that if a student stays in a state, e.g., video, for a long time (longer than a threshold),[4] the automaton moves back to idle.

Applying the algorithm described above, we estimate each user's effective learning time on a certain course. Similar to forum activities, time spent on videos and assignments is extremely few. Specifically, the median time for watching videos is 4.53 minutes per course and 0 for working on assignments. In fact, 36% users never watch videos, and 52% of them never do assignments, suggesting that doing assignments requires more effort from users. In addition, for users who have active learning records, the amount of video and assignment related activities increases in time ($p < 0.01$) for all but assignments in non-science courses ($p > 0.1$). We suspect that students may only spend time on learning when the deadline approaches.

We report regression results of users' effective learning time in Table 3. First, compared to male students, female students spend significantly more time on both videos and assignments in non-science courses (Columns 1-2), while they spend significantly less time on both activities in science courses (Columns 3-4). Again, this echoes our prior findings on gender differences in both course selection and forum activities. Second, for both science and non-science courses, bachelors work hardest among all education groups, and the effect size is significantly stronger for science courses than non-science course (video: 8.032 vs. 1.019, $p = 0.003$; assignment: 8.946 vs. 1.123, $p = 0.000$). Graduate students spend least time on study in non-science courses (Columns 1-2). In science courses, though their effort is higher than those with degrees below bachelors, it is still lower than that for bachelors (video: 6.945 vs. 8.032, $p = 0.648$; assignment: 4.817 vs. 8.946, $p = 0.076$). Additionally, when the course requires higher effort, students spend significantly more time on assignments (Column 2). Surprisingly, they spend significantly less time on videos (Column 1). One possibility is that the time spent on videos is also correlated with unobserved course characteristics, e.g., the length of videos.

2.2 Certificate Rate

Now, we examine whether the likelihood of getting the certificate is correlated with user demographics, forum activities and effective learning time. Overall, among 11 courses in our data, the certification rate lies between 0.84% and 14.95%. The average certification rate for science courses is lower than that in non-science courses, though the difference is not statistically significant (1.11% vs. 4.68%, $p = 0.178$, two-sided t-tests).

Table 4: Regression Analysis for Certificate Rate: All Users.

	Model 1		Model 2	
	Non-Science (1)	Science (2)	Non-Science (3)	Science (4)
Female	0.014*** (0.002)	-0.003 (0.002)	0.002* (0.001)	0.001 (0.002)
New Post	—	—	0.004*** (0.001)	0.038*** (0.008)
Reply	—	—	0.004** (0.002)	0.001* (0.001)
Video	—	—	0.000*** (0.000)	-0.000 (0.000)
Assignment	—	—	0.003*** (0.000)	0.000*** (0.000)
Bachelor	0.014*** (0.002)	0.003* (0.002)	0.011*** (0.001)	-0.001 (0.001)
Graduate	0.007*** (0.002)	0.004 (0.002)	0.013*** (0.002)	0.001 (0.002)
Effort	-0.072*** (0.003)		-0.072*** (0.003)	
Constant	0.286*** (0.013)	0.018*** (0.006)	0.280*** (0.011)	0.006 (0.004)
Obs.	74,480	19,269	74,480	19,269
R^2	0.024	0.001	0.462	0.363

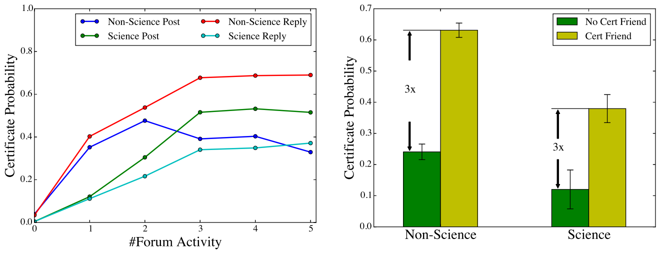
Table 4 summarizes results from linear probability models. In the first model specification (Columns 1-2), we only include demographic information used in prior regressions, and we control for forum activities and effective learning time in the second model (Columns 3-4). First, we find that compared to male students, females are significantly more likely to get the certificate in non-science courses (Column 1 of Model 1); however, the size of the gender difference decreases significantly after we control for forum activities and effective learning time in Model 2 (0.014 vs. 0.002, $p = 0.000$, chi-square test). This suggests that the superiority of women's performance in non-science courses is mainly driven by their effort in forum and learning activities.

Second, compared to students with degrees below bachelors, bachelors are significantly more likely to get the certificate in non-science courses and the result is robust in both models (Columns 1 and 3, $p < 0.01$), while their high effort on science courses does not transform to significantly higher certificate rate.[5] Surprisingly, graduate students are also significantly more likely to get the certificate in non-science courses (Column 1: 0.007, $p < 0.01$), although their effective learning time is significantly lower than others. Moreover, after controlling for effort, the size of the effect is almost doubled (Column 3: 0.013, $p < 0.01$). This implies that the certificate rate is not only related to effort, but also ability and existing knowledge level. In particular, for graduate students who may have higher learning ability, once they exert effort, they have a higher chance to get certificates. Additionally, the required effort level for a course is significantly and negatively correlated with the likelihood of getting certificates (Columns 1 and 3).

Last but not least, we discuss effect of forum activities and effective learning time on certificate rate. Specifically, both forum activities are good predictors for getting certificates. In non-science courses, the size of the effect between posting and replying questions is about the same (0.004 vs. 0.004, $p = 0.837$, F-test). In

[4] In our implementation, we set the threshold to 20 minutes based on human feedback.

[5] The effect is positive and marginally significant in Model 1 and it becomes negative and insignificant in Model 2.



(a) Forum Activity

(b) Certificate Rate

Figure 3: Forum Activity and Certificate Rate for Active Users.

contrast, in science courses, posting questions matters more than answering them (0.038 vs. 0.001, $p = 0.000$, F-test), suggesting that students who ask questions are more likely to get certificates than those who answer questions. Moreover, in both two types of courses, the amount of time spent on videos and assignments are significantly positively correlated with certificate rates,[6] although the magnitude of the effect is very small.

As the distribution of forum activities is highly skewed, our findings might be contaminated by the large number of inactive users. As a robustness check, we focus on the active users, e.g., those who have posted at least one thread in the forum. Overall, the relationship between forum activities and certificate rate is still positive and significant, with a slight decline in magnitude. It implies that to improve the likelihood of getting certificates, it is more important to be present and participate, while the intensity of participation, for instance posting another thread, matters less. This is supported by Figure 3(a), in which we plot the percentage of students obtaining a certificate by the number of forum threads they post. Clearly, the certificate rate increases sharply with the first few posts, but the benefit from further incremental activities becomes smaller. Another empirical finding is the spillover effect. We define a "certificate friends" dummy which is coded as 1 if a user interacts with other students who finally get certificates, and it is significantly positively correlated with users' likelihood of getting certificates, suggesting the importance of homophily correlation and positive learning spillover from well-performed students.

Summary. In summary, we have the following observations:

- Female students are more likely to ask questions in non-science course forums, though not necessarily reply more questions. They also spend more time on watching videos and working on assignments for non-science courses.
- Bachelors ask more questions, especially in non-science courses. Interestingly, graduate students are not as active as bachelors in terms of asking questions, but they are quite active in answering questions, especially in science courses.
- Both forum activities and effective learning are significant predictors for certificate rate, suggesting the importance of encouraging students to participate in forum discussions.

3. MODEL FRAMEWORK

Now, we turn to discuss how to model the learning behavioral data. We propose a latent dynamic factor graph (LadFG) model to address the problem. The model incorporates observed learning activities, including forum activities, watching videos, and doing assignments, into a unified framework. Different from previous research on factor graph model [26], in LadFG, we use a latent learning state to model students' learning state. Based on the modeling results, we will introduce how to apply the model to predict students' learning behavior.

3.1 Formulation

To present the model precisely, we introduce some necessary notations. Let V denote a set of $|V| = N$ students. We first define students' observed learning activities as follows:

Definition 2. **Learning Activity**: Let $Y \in \{0,1\}^{T\times N\times n}$ be a tensor, with each element $Y_{t,i,j}$ representing the j^{th} activity by student $v_i \in V$ at time t, where T is the number of time stamps and n is the number of activities.

The activity space includes all activities (e.g., doing assignments and getting certificates) that we are interested in. We use a $n-$dimensional vector $Y^t(i) = [Y_{t,i,0}, Y_{t,i,1}, \ldots, Y_{t,i,n-1}]^{\mathrm{T}}$ to represent all activities performed by student v_i at time t. Moreover, we introduce the definition of latent learning state.

Definition 3. **Latent Learning State**: For each student v_i at time t, we define a $m-$dimensional vector of continuous latent states $Z^t(i) = [Z_{t,i,0}, Z_{t,i,1}, \ldots, Z_{t,i,m-1}]^{\mathrm{T}}$, with $Z_{t,i,j} \in [0,1]$. Latent states of all students at all time stamps are recorded in a tensor $Z \in [0,1]^{T\times N\times m}$.

To facilitate model description, we define $\mathbf{Z}_{t-p}^{t-1}(i) = \left[Z^{t-1}(i)^{\mathrm{T}}, Z^{t-2}(i)^{\mathrm{T}}, \ldots, Z^{t-p}(i)^{\mathrm{T}}\right]^{\mathrm{T}} \in [0,1]^{m\times p}$ as the vector of latent states of student v_i in the previous p time stamps. Finally, for modeling student behavior, we define attributes (features) for students in different time stamps as follows.

Definition 4. **Time-varying Attribute Tensor**: Let $X \in \mathbb{R}^{T\times N\times d}$ be a tensor in which $X_{t,i,j}$ represents the j^{th} attribute of student v_i at time t, and d is the number of defined attributes (features). We use a d-dimensional vector $X^t(i) = [X_{t,i,0}, X_{t,i,1}, \ldots, X_{t,i,d-1}]^{\mathrm{T}}$ to represent the attribute values of student v_i at time t.

The attribute tensor includes all possible attributes associated to students at different time stamps, e.g., demographics and all behavior except the defined learning activities.[7] The attributes may not change over time, e.g., gender. In this case, we have $X^0(i) = X^1(i) = \cdots = X^{(T-1)}(i)$.

Our goal for modeling learning behavior is to find a mapping from students' attribute tensor to the observed learning activities. Instead of directly learning the mapping, we use latent learning states as the bridge to connect the two sets of observation variables. Based on this idea, we propose a latent dynamic factor graph (LadFG) model.

3.2 Latent Dynamic Factor Graph Model

Figure 4 shows the graphical representation of the LadFG model. Each student is associated with a feature vector $X^t(i)$ and a set of activities $Y^t(i)$ for time t. We use latent states $Z^t(i)$ to model students' activities and features. In Figure 4, each group of circles stands for a student's latent learning states $Z^t(i) = [Z_{t,i,0}, \cdots, Z_{t,i,m-1}]$ at time t, which is used to characterize the

[6] The only insignificance is the effect of video time in science courses.

[7] In different prediction applications, the defined learning activities and attribute tensor may be different.

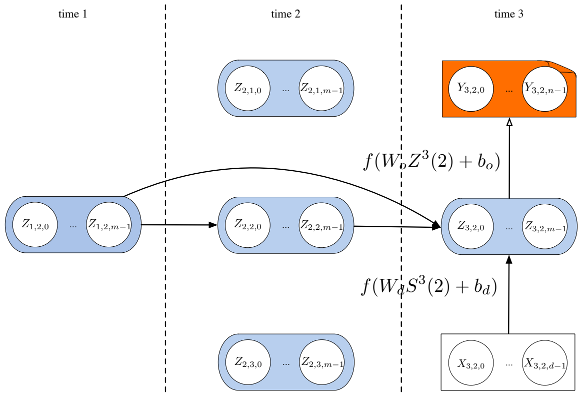


Figure 4: Graphical Representation of the LadFG Model. Each group of circles stands for a student's latent learning states $Z^t(i)$ at time t, which is used to characterize the intention of the student to perform activities $Y^t(i)$; each group of latent states is associated with an activity vector $Y^t(i)$, a vector of attributes $X^t(i)$, and depends on the student's historic learning states $\mathbf{Z}_{t-p}^{t-1}(i)$ in previous p time stamps; $f(W_dS^t(i)+b_d)$ and $f(W_oZ^t(i)+b_o)$ denote two factors to capture the latent learning states and dynamic dependency, respectively.

intention of the student to perform different activities. Each group of latent states is associated with an activity vector $Y^t(i)$, a vector of attributes $X^t(i)$, and it depends on the student's historic learning states $\mathbf{Z}_{t-p}^{t-1}(i)$ in previous p time stamps. Furthermore, $f(W_dS^t(i)+b_d)$ denotes a factor function to capture the dynamic dependency. The method of modeling students' activities using latent states is similar to the assumption in Dynamic Factor Graph (DFG) [21], Markov Decision Process (MDP) [14], and Deep Learning [13]. Activity correlations between students are also modeled in the latent state space using function $f(.)$. Different from traditional models, we also model a long-distance dependency—a student's latent state can depend on her states in previous p time stamps.

In the LadFG model, we use function $f(.)$ to capture correlations between different (observed or latent) variables. For simplicity, we use logistic regressions to model the dependency, although it can be replaced by any other functions. Specifically, we define f as a sigmoid function

$$f(z)=\frac{1}{1+e^{-z}}.$$

Moreover, for latent state $Z^t(i)$, LadFG models the observed learning activities $Y^t(i)$ by:

$$Y^t(i)^*=f(W_oZ^t(i)+b_o) \tag{1}$$

where $W_o\in\mathbb{R}^{n\times m}, b_o\in\mathbb{R}^n$ are observation model parameters.

The dynamic model establishes a time-dependent correlation between a sequence of p past latent state $\mathbf{Z}_{t-p}^{t-1}(i)$, attributes $X^t(i)$ and latent state $Z^t(i)$. For easy explanation, we define inputs of dynamic model as:

$$S^t(i)=\left[\mathbf{Z}_{t-p}^{t-1}(i)^{\mathrm{T}},X^t(i)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2}$$

Then the latent states can be obtained by :

$$Z^t(i)^*=f(W_dS^t(i)+b_d) \tag{3}$$

```
Input: number of epochs $L$;
number of batches in each epoch in M Step $B$;
latent states dimension $m$;
E Step learning rate $\eta_z$ and M Step learning rate $\eta_\Theta$;
weight decay parameters $\lambda_z$ and $\lambda_w$
Output: learned parameters $\Theta=\{W_d,W_o,b_d,b_o\}$, latent states $Z$;
Initialize model parameters $\Theta$;
Initialize latent states $Z$;
for $l=1$ to $L$ do
    E Step:% fix $\Theta$, update $Z$
    Compute gradient $\nabla Z_{t,i,j}$;
    Update $Z_{t,i,j}\leftarrow Z_{t,i,j}+\eta_z\nabla Z_{t,i,j}$;
    M Step:% fix $Z$, update $\Theta$
    for $b=1$ to $B$ do
        Generate random integer $t_\alpha<t_\beta$ from $[0,T)$
        Calculate the gradient of all parameters and update according
        to $X,Y,Z$ in time span $[t_\alpha,t_\beta]$:
        Update $W_o\leftarrow W_o+\eta_\Theta\nabla W_o$;
        Update $b_o\leftarrow b_o+\eta_\Theta\nabla b_o$;
        Update $W_d\leftarrow W_d+\eta_\Theta\nabla W_d$;
        Update $b_d\leftarrow b_d+\eta_\Theta\nabla b_d$;
    end
end
return $\Theta$ and $Z$;
```

Algorithm 1: Learning and Inference by LadFG.

where $W_d\in\mathbb{R}^{m\times(mp+d)}$ and $b_d\in\mathbb{R}^m$ are model parameters.

Based on all parameters above, we define the objective function of the LadFG model as follows.

$$\begin{aligned}\mathcal{O}(\Theta)=&\sum_{i=0}^{N-1}\sum_{t=0}^{T-1}\left\|Z^t(i)-Z^t(i)^*\right\|^2+\sum_{i=0}^{N-1}\sum_{t=0}^{T-1}\left\|Y^t(i)-Y^t(i)^*\right\|^2\\&+\lambda_w\left(\|W_o\|^2+\|W_d\|^2\right)+\lambda_z\sum_{i=0}^{N-1}\sum_{t=0}^{T-2}\|Z^t(i)-Z^{t+1}(i)\|^2\end{aligned} \tag{4}$$

where λ_w is a parameter that controls the regularization value; and λ_z controls the degree of smoothness between learning states of consecutive time stamps. Here, we use square loss as the loss function and apply L-2 regularization on model parameters to overcome overfitting problems.

Learning an LadFG model is to estimate a configuration of parameters $\Theta=\{W_o,b_o,W_d,b_d\}$ and latent states Z from a given historic behavior log that minimizes the objective function Eq.(4). As the objective function has a closed form solution, it can be learned efficiently by the gradient descent method, e.g. stochastic gradient descent (SGD).

3.3 Model Learning

We have two sets of learning variables: model parameters Θ and latent states Z. To learn these parameters, we use an EM-style algorithm to achieve the minimization of the objective function (Eq. 4) iteratively. The learning algorithm is summarized in Algorithm 1. Specifically, there are two major steps:

1. **E-step:** fix all model parameters Θ and update Z, by using a gradient descent method.

2. **M-step:** fix all latent states Z and update each model parameter in Θ.

Each parameter is updated by $\theta_i\leftarrow\theta_i+\eta\frac{\partial\mathcal{O}}{\partial\theta_i}$, where η is the learning step. The gradient of each parameter w.r.t the objective function can be derived in the following ways. We use $Z_{t,i,j}$ and

$W_{du,v}$ as the example to explain the gradient, and omitted others due to space limitation.

$$
\begin{aligned}
\frac{\partial \mathcal{O}}{\partial Z_{t,i,j}} &= 2(Z_{t,i,j} - Z^*_{t,i,j}) \\
&- 2\sum_{k=1}^{p}\sum_{l=0}^{m-1}(Z_{t+k,i,l} - Z^*_{t+k,i,l})Z^*_{t+k,i,l}(1 - Z^*_{t+k,i,l})W_{dl,(k-1)m+j} \\
&- 2\sum_{l=0}^{n-1}(Y_{t,i,l} - Y^*_{t,i,l})Y^*_{t,i,l}(1 - Y^*_{t,i,l})W_{ol,j} \\
&+ 2\lambda_z(Z_{t,i,j} - Z_{t+1,i,j}) - 2\lambda_z(Z_{t-1,i,j} - Z_{t,i,j})
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{O}}{\partial W_{du,v}} = &- 2\sum_{i=0}^{N-1}\sum_{t=0}^{T-1}(Z_{t,i,u} - Z^*_{t,i,u})Z^*_{t,i,u}(1 - Z^*_{t,i,u})S_{t,i,v} \\
&+ 2\lambda_w W_{du,v}
\end{aligned}
$$

Finally, according to our definition, we have:

$$
S_{t,i,v} = \begin{cases} Z_{t-\lfloor \frac{v}{m} \rfloor -1, i, v \bmod m}, & \text{if } v < m \times p \\ X_{t,i,v-mp}, & \text{if } v \geq m \times p \end{cases} \tag{5}
$$

Feature Definition and Implementations. Our implementation of learning algorithm is based on the machine learning framework Theano [3, 5]. To train the LadFG model, we mainly define three categories of features.

- **Demographics:** These features are derived from students attributes including: gender, age, education. We consider binary features for each discrete value of every attribute, and in total we have eighteen demographics-based features.
- **Forum:** These features are derived from forum activities, respectively representing the number of different forum activities, the number of replies received from other students, etc. We also consider some homophily correlation features such as the number of replies received from well-performed students. There are six forum activity-based features.
- **Learning Behavior:** These features include a list of statistical features such as the number of chapters a student browses, the number of deadlines she completes and the total time she spends on watching videos and doing assignments. In total, there are ten features defined in this category.

4. EXPERIMENTAL RESULTS

The proposed method for modeling students' learning behavior is very general and can be applied to different settings in MOOCs. Furthermore, using data from xuetangX, we present various experiments in this section to evaluate the effectiveness and efficiency of the proposed method. All datasets and codes will be publicly available.

4.1 Experimental Setup

Evaluation Aspects. To quantitatively evaluate the proposed model, we consider the following performance measurements:

- **Assignment Grade Prediction.** We apply our methods to the MOOC data to predict students' grades on assignments and compare different methods. Specifically, each course has multiple assignments and a student will get a grade after completing an assignment. Our goal is to predict students' grade for each assignment. We cast the prediction task as a binary classification problem, e.g., for non-science courses, "Yes", if a student's grade is ranked top 30% of all students. Otherwise, it is "No".[8]
- **Certificate Earner Prediction.** We use our methods to model and predict whether a student will get the certificate after completing a course.
- **Parameter Sensitivity Analysis.** We analyze the sensitivity of different parameters in our methods: latent dimension m, the parameter p, and model convergence.

Finally, we present results of error analysis for our method. In all experiments, we remove the "register-only" students in our experiments.[9]

Comparison Methods and Evaluation Metrics. We compare our model with several alternative predictive models:

- **Logistic Regression (LRC) [19]:** It uses logistic regressions to train a classification model and employs the classification model to make the prediction.
- **SVM [10]:** It uses SVM to train the classification model and employs it to make the prediction. For SVM, we employ LIBLINEAR [10].
- **Factorization Machines [1, 17]:** Factorization models have been proposed and successfully applied to recommendation and prediction tasks. As the factorization model projects the input feature space onto a latent space, it enables us to learn more complex interactions between features.

Altogether, we try to use the same set of features in all comparison methods. Regarding tunable parameters m, p, η_Θ, η_z, λ_w and λ_z in LadFG, we find the best configuration using cross validation (i.e., $m = 5$, $p = 2$, $\eta_z = 0.5$, $\eta_\Theta = 0.1$, $\lambda_w = 0.01$ and $\lambda_z = 0.01$). Moreover, we evaluate the performance of comparison methods in terms of Area Under Curve (AUC), Precision (Prec.), Recall (Rec.), and F1-Measure (F1) [7]. Additionally, all algorithms are implemented in Python, and all experiments are performed on an x64 machine with 2.9GHz intel Core i7 CPU and 8GB RAM.

4.2 Assignment Grade Prediction

We conduct assignment grade prediction for the 11 courses in our dataset. In each course, we use the first half (before the mid-term) of the data for training and the second half (after the mid-term) for testing the prediction performance. Table 5 presents the average prediction performance of different methods and the largest performance numbers under each index are bolded. Overall, the proposed LadFG model clearly outperforms all alternative methods. In terms of F1-score, LadFG achieves a 16.0-25.6% improvement compared to SVM and LRC, neither of which considers the latent interactions between variables. Although FM which also considers the interactions between variables outperforms SVM and LRC, it cannot effectively leverage the temporal information and correlations between students. Consequently, it still underperforms LadFG. We perform two sided t-tests and all the p-values are < 0.01, which

[8] We use 30% as the cutoff for non-science courses because on average, 30% of students who completed at least one assignment eventually earned the course certificate. The cutoff for science courses is 10%.

[9] It refers to students who only register a course, but never actually engage in any course activities.

Table 5: Performance of Assignment Grade Prediction with Different Methods (%).

Category	Method	AUC	Precision	Recall	F1-score
Science	LRC	80.96	50.65	68.02	57.57
	SVM	70.99	50.44	45.93	47.42
	FM	90.39	61.48	70.35	64.67
	LadFG	**96.26**	**66.52**	**81.40**	**73.09**
Non-Science	LRC	73.10	75.59	53.45	61.54
	SVM	71.45	**77.01**	48.94	58.41
	FM	85.73	67.36	81.69	72.95
	LadFG	**90.47**	73.28	**85.68**	**78.91**

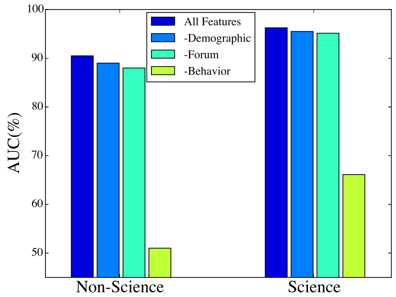


Figure 5: Feature Contribution Analysis for Assignment Grade Prediction.

indicates that the improvements of our proposed models over the comparison methods are statistically significant.

Feature Contribution Analysis. We study how different categories of features (demographics, forum activities, and learning behavior) help the prediction task. The learning behavior includes all activities related to watching videos and doing assignments. Specifically, each time, we respectively remove demographics, forum activities, and learning behavior when training our proposed model, and compare performance of assignment grade prediction based on the trained models. Figure 5 shows the AUC performances for each type of courses. "All Features" stands for the LadFG model considering all features defined in our method. "-Demographic", "-Forum", and "-Behavior" indicate results from removing demographics, forum activities, and learning behavior, respectively. Apparently, the learning behavior features contribute significantly to the results for both types of courses. It is also worthwhile to note that learning behavior seems to be more important for modeling the non-science courses. Without learning behavior-based features, the performance of assignment grade prediction in non-science courses is worse than that in science courses.

Effect of the percentage of the training data. We investigate the performance of assignment grade prediction by varying the percentage of the training data. Figure 6 shows the average prediction performance. The x-axis ($k = 20 - 90$) indicates the percentage of data we use for training. It is very interesting that non-science courses seem to be more *predictable*. LadFG achieves a stable performance for non-science courses using only 30% of data for training. For science courses, the prediction performances vary a lot and they depend on the amount of data used for training. Additionally, it seems difficult to predict grades of the last assignments in science courses.

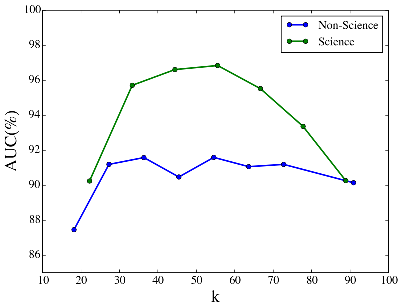


Figure 6: Average Prediction Performance by Varying the Percentage of Data for Training.

Table 6: Performance of Certificate Earner Prediction with Different Methods (%).

Category	Method	AUC	Precision	Recall	F1-score
Science	LRC	92.13	**83.33**	46.51	59.70
	SVM	92.67	52.17	83.72	64.29
	FM	94.48	61.54	74.42	67.37
	LadFG	**95.73**	73.91	**79.07**	**76.40**
Non-Science	LRC	94.16	76.93	89.20	82.57
	SVM	93.94	76.96	88.60	82.37
	FM	94.87	**80.22**	86.23	83.07
	LadFG	**95.54**	79.76	**89.01**	**84.10**

4.3 Certificate Earner Prediction

Moreover, we use certificate earner prediction to evaluate the effectiveness of the proposed model. Again, we use the first half of the data (before the mid-term) for training each method, and apply the learned model to predict whether a student would get the course certificate. Table 6 summarizes the performance of different methods for predicting certificate earners. Consistently, LadFG performs better than alternative methods. In addition, we conduct a similar feature contribution analysis for certificate earner prediction. Figure 7 shows the prediction performance of the proposed model by considering different categories of features and we have similar results as those for assignment grade prediction.

4.4 Parameter Sensitivity Analysis

We now discuss how different parameters influence the performance of our methods, as well as presenting the efficiency performance. In all the analyses, we use the assignment grade prediction

Table 7: Efficiency Performance of Different Methods.

Category	LRC	SVM	FM	LadFG
Science	0.07sec	2.35sec	2.7min	1.4min
Non-Science	1.46sec	9.8min	4.1min	8.6min

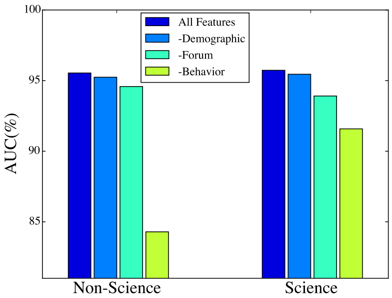


Figure 7: Feature Contribution Analysis for Certificate Earner Prediction.

as the example, and report the average performance on non-science and science courses respectively.

Effect of Latent Dimension m. We evaluate how the latent dimension m affects the quality of the models learned by LadFG. We perform an analysis by varying the dimension of latent space in the proposed LadFG method. Figure 8(a) shows its AUC performance with different numbers of topics. We find that LadFG results in a better performance when the latent dimension is smaller than 20. In addition, increasing the dimension results in a performance drop and this might be due to the data sparsity.

Effect of Parameter p. Pertaining to the parameter p, Figure 8(b) shows the performance of LadFG with different p (other parameters are fixed, e.g., $m = 5$). Although the performance changes when we vary the value of p, the best performance is obtained when $p = 2$ (by considering latent states of two previous time stamps).

Convergence analysis. We further investigate the convergence of the learning algorithm for LadFG. Figure 8(c) presents the convergence analysis of the algorithm. The algorithm converges within 500 iterations. Furthermore, this rapid convergence enables us to do efficient training of the model on large scale datasets.

Efficiency. We compare the efficiency of the different methods. Table 7 lists the average running time of the comparison algorithms for training the prediction models. Overall, all methods have good efficiency performance, and the running time of different methods ranges from seconds to minutes. LadFG results in a slightly lower efficiency compared with FM.

4.5 Error Analysis

In the end, we conduct an error analysis on the results of our approach and observe three major types of errors.

(1) *Unpredictable negative cases.* The proposed LadFG model fails in the following scenario in which a substantial proportion is credited to final exams in certain courses. Some active learners, who engage in forums and lectures enthusiastically, may not take tests due to personal reasons, e.g., scheduling conflicts, and they would be misclassified into certificated group.

(2) *Unpredictable positive cases.* In contrast, some inactive students never participate in forums and have very few learning activities. However, they still complete assignments (or pass the final exams) with extremely high scores, and they are incorrectly grouped to be low grade in the assignment grade prediction (or uncertificated in the certificate earner prediction). We randomly choose several of those students and conduct interviews with them. We find that a substantial proportion of them are very skillful undergraduate or graduate students who have taken similar courses offline. Therefore, they are capable of taking online tests without spending much effort.

(3) *Swing cases.* LadFG would become ineffective for certificate earner prediction on "swing" cases as well. Each course has a minimum passing score, above which the student would be certificated. Swing cases are those students whose scores are hovering around the minimum score. Our model would misclassify these swing cases and draw a wrong conclusion.

5. RELATED WORK

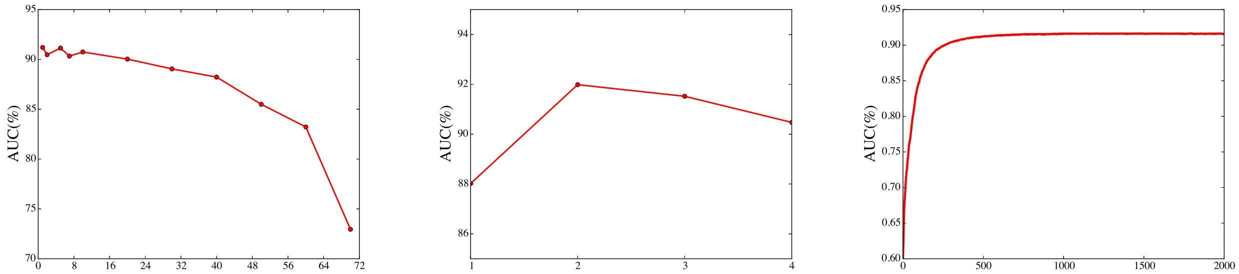
MOOCs boom swiftly in recent years and have attracted millions of users worldwide. Experiences in offline education which predict dropouts and school failures do not fully satisfy the need in MOOCs. Analyzing and mining the big data from online courses becomes an important topic to understand students' behavior. We review related literature in three topics: attribute analysis, engagement analysis and time-related feature analysis.

Attribute analysis: This line of researches mainly focuses on studying the relationship between user demographic attributes and their behavioral patterns in online courses. For example, Guo et al.[11] investigated how navigation strategies vary by demographics. Based on demographic results, Wilkowski et al.[27] found no correlation between prior skills and course completion rates. Furthermore, their study showed that students who completed course activities were more likely to earn certificates than those who did not. Moreover, Seaton et al.[25] examined activities that help students get certificates.

Engagement analysis: More and more researchers start to analyze students' engagement in courses. For online open courses, Anderson et al.[2] developed a taxonomy of individual engagement style, and there was a further discussion between student engagement and their grades. Since student dropouts in MOOCs have gathered widespread attention, Ramesh et al.[24] proposed a latent representation model which could be applied to abstract student engagement types and to predict dropouts. Additionally, for traditional courses, Bayer et al.[4] predicted dropouts and school failures when student data has been enriched with data derived from students' social behavior. Related studies can be also found in [12, 22]. However, most existing research on engagement shared similarities with Champaign et al.'s work[8]. They estimated time spent on different resources and examined correlations between time and students' performance.

Forum analysis: Several interesting work examines MOOC forums, which plays an important role in online learning. Chaturvedi et al.[9] predicted instructor's intervention in forums. Brinton et al.[6] investigated factors correlated with the decline of forum activities, and found strategies to classify and rank thread relevance. Additionally, Huang[15] explored super-posters on forums and studied their engagement patterns. They found that super-posters display above-average engagement, enroll in more courses, and obtain better grades than the average forum participants. However, these studies might isolate and overstate the importance of forum performance.

To the best of our knowledge, there was little work utilizing temporal correlations between demographics, forum behaviors, and learning activity patterns to model and predict assignment grade performance and certificate earners.



(a) Latent dimension m

(b) Parameter p

(c) Convergence analysis

Figure 8: Parameter Analysis. (a) Performance of LadFG model by varying the latent dimension m; (b) Performance of LadFG model by varying the parameter p; (c) Convergence analysis of LadFG.

6. CONCLUSION

In this paper, we study a novel problem of modeling and predicting learning behavior in MOOCs. We conduct in-depth analysis for student demographics, and learning activity patterns in course forums, videos and assignments. We propose a latent dynamic factor graph (LadFG) to incorporates students' demographics, forum activities, and learning behavior into a unified framework. Our experimental results on two prediction tasks: assignment performance prediction and certificate earner prediction, validate the effectiveness of the proposed model.

The general idea in this paper, to model and predict learning behavior in MOOCs, represents an interesting and new research direction. There are many potential future directions for this work. A straightforward task would be to incorporate human feedback into the proposed model. Other courses and more information on users would also be worth exploring. Looking further ahead, we believe that different models and semi-supervised learning algorithms for exploring social network structures should be beneficial. Finally, building a theory of why and how students join and quit different courses is an intriguing direction for future research.

Acknowledgements. The work is supported by the National High-tech R&D Program (2014AA015103), Natural Science Foundation of China (61222212), National Social Science Foundation of China (13&ZD190), a research fund supported by Huawei Inc.

7. REFERENCES

[1] D. Agarwal and B.-C. Chen. Regression-based latent factor models. In *KDD'09*, pages 19–28, 2009.

[2] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec. Engaging with massive online courses. In *WWW'14*, pages 687–698, 2014.

[3] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio. Theano: new features and speed improvements. In *Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop*, 2012.

[4] J. Bayer, H. Bydzovská, J. Géryk, T. Obsivac, and L. Popelinsky. Predicting drop-out from social behaviour of students. In *EDM'12*, pages 103–109, 2012.

[5] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio. Theano: a cpu and gpu math expression compiler. In *SciPy'10*, volume 4, page 3, 2010.

[6] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong. Learning about social learning in moocs: From statistical analysis to generative model. *IEEE Transactions on Learning Technologies*, 7(4):346–359, 2014.

[7] C. Buckley and E. M. Voorhees. Retrieval evaluation with incomplete information. In *SIGIR'04*, pages 25–32, 2004.

[8] J. Champaign, K. F. Colvin, A. Liu, C. Fredericks, D. Seaton, and D. E. Pritchard. Correlating skill and improvement in 2 MOOCs with a student's time on tasks. In *L@S'14*, pages 11–20, 2014.

[9] S. Chaturvedi, D. Goldwasser, and H. Daumé III. Predicting instructor's intervention in MOOC forums. *ACL'14*, pages 1501–1511, 2014.

[10] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *JMLR*, 9:1871–1874, 2008.

[11] P. J. Guo and K. Reinecke. Demographic differences in how students navigate through MOOCs. In *L@S'14*, pages 21–30, 2014.

[12] A. Hershkovitz, R. S. Baker, S. M. Gowda, and A. T. Corbett. Predicting future learning better using quantitative analysis of moment-by-moment learning. In *EDM'13*, pages 74–81, 2013.

[13] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

[14] R. A. Howard. *Dynamic Programming and Markov Processes*. The M.I.T. Press, 1960.

[15] J. Huang, A. Dasgupta, A. Ghosh, J. Manning, and M. Sanders. Superposter behavior in MOOC forums. In *L@S'14*, pages 117–126, 2014.

[16] R. F. Kizilcec, C. Piech, and E. Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *LAK'13*, pages 170–179, 2013.

[17] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *KDD'08*, pages 426–434, 2008.

[18] P. F. Lazarsfeld and R. K. Merton. Friendship as a social process: A substantive and methodological analysis. *M. Berger, T. Abel, and C. H. Page, editors, Freedom and control in modern society, New York: Van Nostrand*, pages 18–66, 1954.

[19] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW'10*, pages 641–650, 2010.

[20] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[21] P. Mirowski and Y. LeCun. Dynamic factor graphs for time series modeling. In *ECML/PKDD'09*, pages 128–143, 2009.

[22] S. O. Nesterko, D. Seaton, J. Reich, J. McIntyre, Q. Han, I. Chuang, and A. Ho. Due dates in MOOCs: does stricter mean better? In *L@S'14*, pages 193–194, 2014.

[23] L. Pappano. The year of the MOOC. *The New York Times*, 2(12):2012, 2012.

[24] A. Ramesh, D. Goldwasser, B. Huang, H. Daume III, and L. Getoor. Learning latent engagement patterns of students in online courses. In *AAAI'14*, pages 1272–1278, 2014.

[25] D. T. Seaton, Y. Bergner, I. Chuang, P. Mitros, and D. E. Pritchard. Who does what in a massive open online course? *Communications of the ACM*, 57(4):58–65, 2014.

[26] W. Tang, H. Zhuang, and J. Tang. Learning to infer social ties in large networks. In *ECML/PKDD'11*, pages 381–397, 2011.

[27] J. Wilkowski, A. Deutsch, and D. M. Russell. Student skill and goal achievement in the mapping with google MOOC. In *L@S'14*, pages 3–10, 2014.