Probabilistic Modeling Towards Understanding the Power Law Distribution of Video Viewing Behavior in Large-scale e-Learning

Ni Xue, Huan He, Jun Liu, Qinghua Zheng, Tian Ma, Jianfei Ruan, Bo Dong
SPKLSTN Lab, Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China
xn_shelly@qq.com; hehuan@mail.xjtu.edu.cn; liukeen@mail.xjtu.edu.cn; qhzheng@mail.xjtu.edu.cn
sky.ma@stu.xjtu.edu.cn; xjtu_jfruan@163.com; dong.bo@mail.xjtu.edu.cn

***Abstract*—In the era of internet, e-Learning has become vastly widespread and generated huge amount of log data of video viewing behavior. Through analyzing and mining these log data, significant Power Law Distribution (PLD) of viewing behavior is observed, which is different from small-scale e-Learning or traditional classroom environment. In this paper, we apply the mechanisms for generating the PLDs in analyzing log data of a large-scale e-Learning platform to discover the factors influencing the video viewing behavior. Firstly, four factors correlated to the video viewing behavior are discovered from log data, including the number of videos viewed, the start date of viewing videos, the date of final exam, and the duration of enrollment. Furthermore, we present a probabilistic model of viewing behavior based on the four factors. Finally, the accuracy of the model is validated with nine online courses in which each course enrolled more than 1,000 students. In addition, we analyze the application of the proposed model and provide some valuable suggestions for teachers to improve the performance of students.**

Keywords-probabilistic modeling of viewing behavior; large-scale e-Learning; log data; power law distribution; factors influcing video viewing behavior.

I. INTRODUCTION

Nowadays, with spread usage of internet and web technology, e-Learning has become vastly widespread and attracted many students to start learning online [1][2][3]. In 2013, there were more than 6.1 million enrolled students in web-based undergraduate colleges which was 7% greater than previous year according to the Ministry of Education of the People's Republic of China [4]. In order to improve the learning performance of these online students, a great deal of models are proposed based on students' learning behaviors [5][6]. Nevertheless, the majority of these models were designed and validated in relatively small-scale e-Learning environments, and hence they are not suitable for analyzing the power law distribution (PLD) of learning behaviors in large-scale e-Learning, such as discussion in online forum [7], attempts submitted in online quizzes [8], edit in course wiki [9], etc.

The online courses are mostly organized as sequences of instructor-produced videos interspersed with other resources such as assessment problems and interactive demos. A study of the first edX course (6.002x, Circuits and Electronics) found that students spent the majority of their time watching videos which makes video lecture one of the premier educational resources of e-Learning platforms [7]. Hence, it has been critical to conduct research to find out how students learn via video lectures. As the amount of videos and students increases rapidly, the data of viewing behavior grow explosively. Therefore, a natural question arises: How can we leverage the large-scale, extensive data that has emerged in recent months to better understand e-Learning viewing behavior of students?

In our research, we study the viewing behavior based on the data set collected from the log data of Xi'an Jiaotong University Distance Learning College (XJTUDLC) platform during the autumn semester of 2014. We find that the distribution of Video Views (VV, which we define as the views of specified course videos for students who registered the course) has a pronounced long tail as is shown in Fig 1. Furthermore, these distributions are plotted on the log-log scale, shown in Fig. 2, and they are all nearly linear with negative slopes which indicates they all follow the PLD.

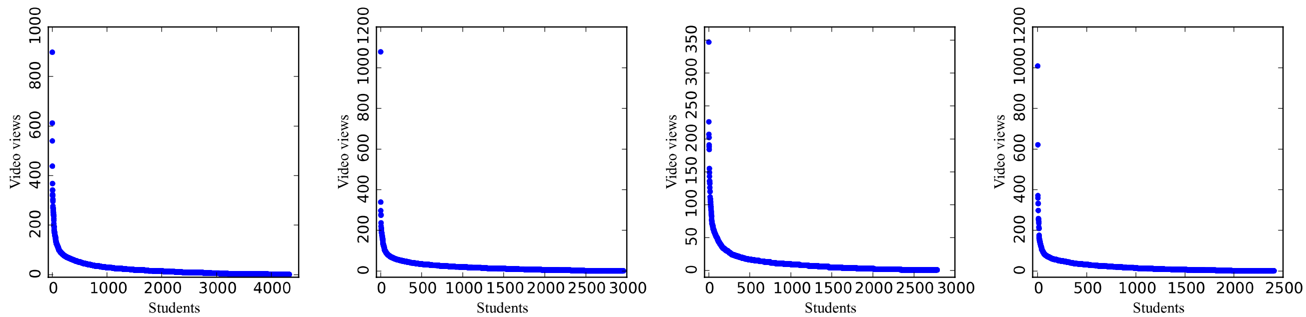
According to the course requirement of XJTUDLC, students who registered the course should finish watching more than 60% videos. Hence, the distribution of VV following the PLD indicates that only very few students meet or even exceed the course requirement and for most students there is still a considerable gap between their completion of course videos and the course requirement. Thus, we conclude that a lot of students have poor performance on their learning process and they fail to take full advantage of the video resources, which may cause negative influence on their understanding of the knowledge and final grade. Thus, we naturally raise the following questions:

Question 1 (Q1). What factors could be directly relevant to the viewing behavior of students?

Question 2 (Q2). Is there a way to codify the viewing behavior of students into a simple model, and if so, can we leverage such a model to facilitate teachers to improve the performance of students?

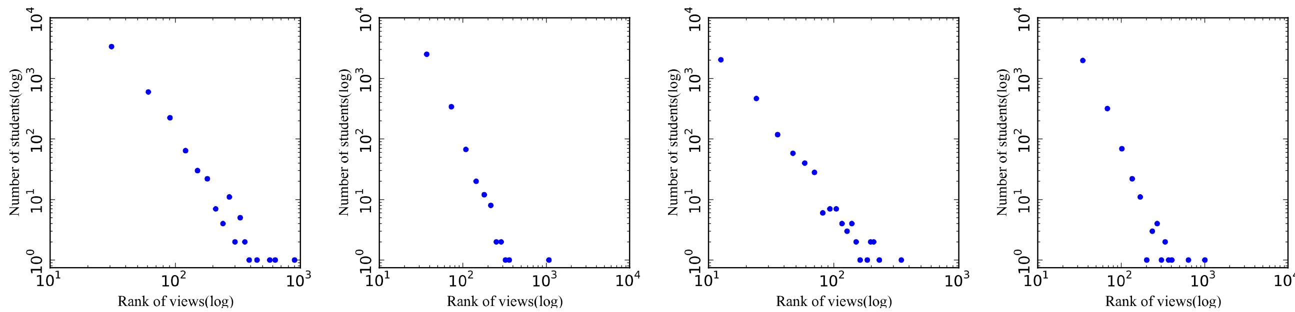
Motivation. The motivation behind studying Q1 is as follows. Video lecture is a most widely used educational tool of online learning. Despite the increasing number and variety of video lectures available, the understanding of how students use and learn from video lectures is still limited. To explore the laws of the viewing behavior and their causes, we analyze the factors influencing the VV directly. Note, however, that some factors contribute to the VV indirectly and are hard to be quantified, such as course content difficulty, student's cognition ability, knowledge background, etc. The question of how these indirect factors contribute to the VV is beyond our scope here.

As for Q2, crystallizing the viewing behavior of students into a simple model will help us better understand the PLD of the VV and the reason why it would happen which in turn makes it available for us to present valuable suggestions for



(a) Politics (b) English II (c) Distance Learning Method (d) English I

Figure 1. Distribution of VV. We choose four of the courses that enrolled more than 1,000 students.



(a) Politics (b) English II (c) Distance Learning Method (d) English I

Figure 2. Distribution of VV (log-log scale).

teachers to improve the performance of students. Differing from most existing research which focus on correlativity analysis of the viewing behavior [10][11][12], the primary goal here is to study the causality, which in our research means how the PLD of VV emerged.

Our methodology. Our analysis consists of the following components.

(1)Statistical analysis. Since the distribution of VV following the PLD, we present some of the mechanisms for generating PLDs, including preferential attachment [13],self-organized criticality [14] and random walks [15] (section 4 presents more details). Based on the mechanisms, we carry out an analysis to understand the factors that are associated with the viewing behavior. As an example, one of the interesting discoveries is that students who have viewed more videos may have higher viewing frequency.

(2)Generative model. To address Q2, we propose a Probabilistic Viewing Behavior Model (PVBM) to simulate the viewing behavior of online students in a specified course. In order to check the validity of the model, we compare the real viewing data of nine online courses and the simulate data obtained from the model, and the results prove that the model works well. In addition, we provide some suggestions for teachersto improve the online learning experience.

II. PRELIMINARIES

This section gives an overview of XJTUDLC platform and presents how we collected our data set.

Our study is based on the learning behaviors on BlueSky platform [16], which is an e-Learning system built by XJTUDLC. After the payment of students, the system will choose courses for them automatically according to their majors. The major teaching method in XJTUDLC is video lecture, and the course materials and the forum are also available as teaching aids. XJTUDLC provides abundant video lecture resources for students according to their specialties. Furthermore, all the students are required to view all specified video lectures, and their final grade are partly based on the completion of their video participation, which makes the video viewing behavior significantly important. The students are required to make an appointment to get the online tutoring according to the notification before the final exam and attend the tutoring on time. Only students who have finished the course assignments and exercises can attend the final examination.

The coursewares of XJTUDLC are the three-part-separated screen coursewares which separate the screen into the video playing area, the directory selection area and the document presentation area. The video playing area shows the instructor's talking head while the directory selection area and the document presentation area show the directory of the coursewares and the usual slides or code , respectively. Each course is divided into several sections, each of which contains a knowledge map and a video list. Students are allowed to choose sections to learn about according to the course schedule and individual arrangement. Each knowledge point in the knowledge map corresponds to a

TABLE I. THE MOST IMPORTANT FIELDS OF THE LOG RECORD

Field Name	Description	Example Value
Logid	The ID of the log	4764191
IP	The IP of the student	121.9.200.46
Platform	The platform number	2
Oper	The operation number	76
Courseid	The course code	162
Time	The time when the log record was written	2014-09-21 09:01:00
Stuno	The number of the student	1069809112******
Tlen	The length of effective viewing time in seconds	1522
Tid	The ID of the title	10045002
Title	The title of the course video	Text A: Friendship

video lecture, and the videos in the video list are sorted in the sequence of knowledge points. Students are able to view the videos by clicking the knowledge points in the knowledge map or choosing videos from the video list directly.

In 2014, there were more than 70 thousands enrolled students in XJTUDLC, in order to analyze their learning behaviors, we developed the Big Log Analysis System (BLAS) [17] for collecting, storing, and analyzing the log data. We store the log records in the form of event data files. One event data file contains all event data of the learning system within one-day period in text format. Each complete log record consists of 18 fields, and the field names, descriptions and example values of the most important ones are shown in Table 1.

Since we study the viewing behavior of students, we only care about the events related to viewing videos. When a student clicks the play button of a video lecture, the learning system will record current time as the time he or she starts viewing the video and start a timer background. The timer will pause once the student clicks the pause button and get back to work until the play button is clicked again. When the student clicks the stop button, the learning system will write a log record with the operation number 76 and write the timer value, which is the length of the effective viewing time, into the Tlen field. Considering that a short-time viewing is always of no value, in our research, we consider that the student views the video once only when the length of the effective viewing time is longer than 5 minutes. Therefore, only log records the Tlen field value of which are longer than 5 minutes are used when we calculating the VV.

In this work, we use the dataset collected from the log data of XJTUDLC between 2014.09 and 2015.01, which contains complete viewing behavior of all new enrolled students in autumn of 2014. In total, our data set consists of 5,028,459 log records, including 268 courses, 52,340 videos and 13,238 students.

III. ANALYSIS OF FACTORS INFLUENCING THE VV

The VV is an important indicator to measure students' completion of courses, and as previously mentioned, a large number of students have significantly fewer video viewing behavior during the course duration. The VV is affected by many factors, including course content difficulty, student's cognition ability, knowledge background, etc. Nevertheless,

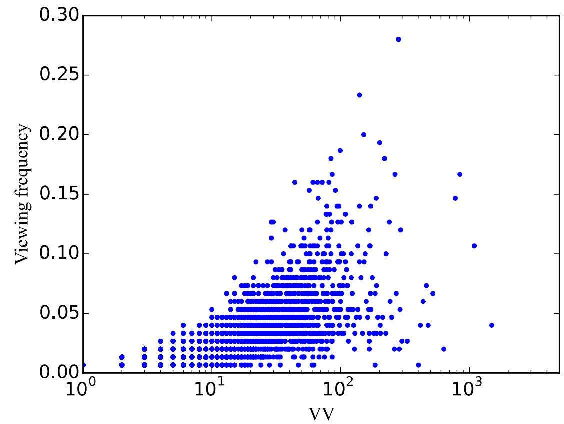


Figure 3. Viewing frequency VS VV

these factors contribute to the VV indirectly and are hard to be quantified. Therefore, we seek for the factors directly influencing the VV.

Since the distribution of the VV follows the PLD, we believe the mechanisms for generating PLDs are of great value for analyzing the factors influencing the VV. Thus, we look at possible candidate mechanisms by which power-law distributions might arise in natural and man-made systems. Barabási and Albert proposed the BA model for the PLDs in complex networks and they explained the PLD is a consequence of two generic mechanisms that networks expand continuously by the addition of new vertices, and new vertices attach preferentially to already well connected sites [13]. The "preferential attachment" is so-called "rich get richer" rule. Self-organized criticality has been taken as the dynamic factor for generating the PLDs, and it argued that the system composed of massive interacting components will naturally move to self-organized criticality and even slight interruptions may cause a series of disasters when the system reaches the state [14]. Another most convincing and widely applicable mechanisms for generating PLDs is random walks. Many properties of random walks are distributed according to PLDs, and it could explain some PLDs observed in nature [15], such as the PLD of the lifetime of biological taxa [18].

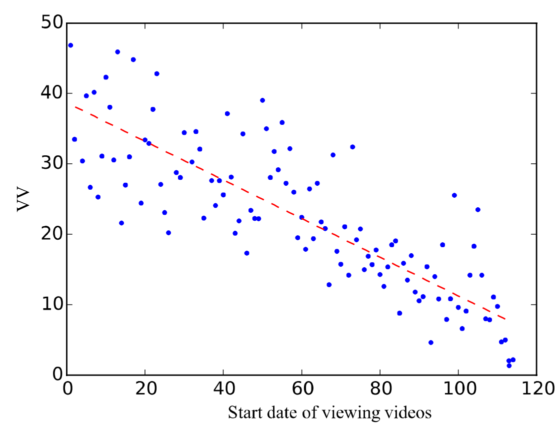


Figure 4. VV VS Start date

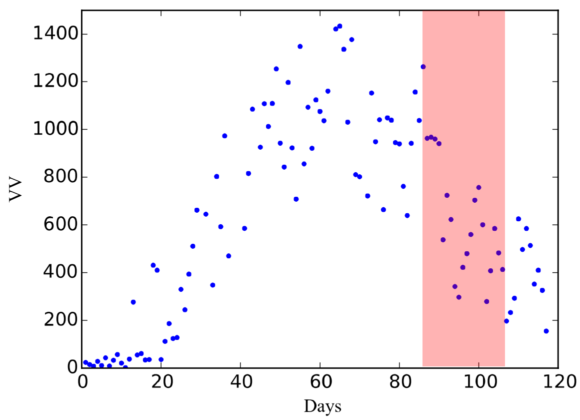


Figure 5. VV VS Days

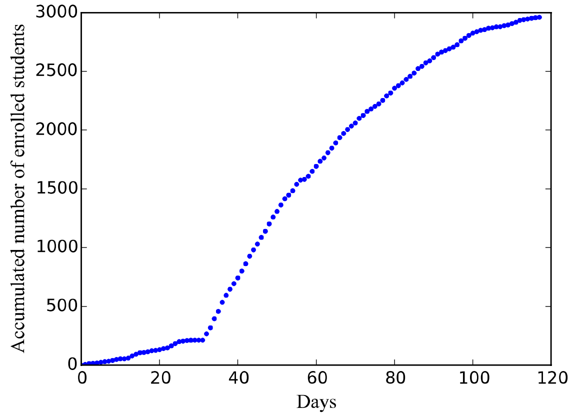


Figure 6. Accumulated number of enrolled students VS Days

Through analyzing the log data of the viewing behavior of Course No.162 (an English basis course) based on aforementioned mechanisms, four factors which are correlated to the VV directly are discovered as follows.

Factor A: The number of videos viewed

In this paper, we define the viewing frequency of a student as the ratio between the actual days he or she viewed and the course duration. The higher viewing frequency indicates the higher probability that the student will view videos next time. Fig. 3 shows the relationship between the VV and the viewing frequency. The Spearman correlation coefficient between the viewing frequency and the VV is 0.838 ($p<0.01$), which indicates that the viewing frequency of students has a strong positive correlation with the VV. In other words, students who have viewed more videos may have higher viewing frequency which is similar with the "rich get richer" rule.

Factor B: The start date of viewing videos

Fig. 4 shows how VV changes over the start date of viewing videos, and each blue dot represents the median of all students' VV who start viewing videos at that day. Obviously, the VV has a negative correlation with the start date of viewing videos (the Spearman correlation coefficient between the start date of viewing videos and the VV is -0.727, $p<0.01$), which means that students who start viewing videos earlier may have higher probability to view videos.

Factor C: The date of final exam

Fig. 5 shows how the total VV of students changes over time, and the red area represents the period of examinations (The exam date differs in different learning centers). It can be seen that the total VV of students shows a significant increasing trend close to the period of examination and declines after it, which shows that the VV is related to the examination date, and the closer to the examination, the higher probability the students will view videos.

Factor D: The duration of enrollment

As described in Fig. 4, the start dates of each student are different, and it is obvious that the enrollment lasts a long time. Therefore, to understand the distribution of the number of enrolled students by the start date, we analyze the accumulated number of enrolled students over time as shown in Fig. 6. During the whole enrollment process, the accumulated number of enrolled students increases slowly at beginning and almost linearly over time after about 4 weeks until all students have enrolled.

IV. MODELING OF VIEWING BEHAVIOR

Based on the four factors and their patterns analyzed above, we propose the following model to simulate the viewing behavior of online students in a specified course.

A. The Probabilistic Viewing Behavior Model

Firstly, we propose the following assumptions to simplify the model:

1. The number of newly enrolled students of a course in each day is a constant value, which is determined by the total number of students who have elected the course and the duration of enrollment.

TABLE II. VARIABLE DEFINITION

Variable Symbols	Definition
N	Total number of students who should study the course
L	List of enrolled students
$w_{i,t}$	Accumulated VV of student i on the tth day
s_i	The number of days from the course start date to the date
$p_i(t)$	The probability for student i to view videos on the tth day
Δw	The incremental VV when a student views videos
u	The number of newly enrolled students per day
E	The number of days from the course start date to the exam
C	The number of course videos
D	The number of days of the course duration

TABLE III. PARAMETERS OF COURSE NO.162

Parameters	***N***	**Δw**	***u***	***E***	***C***	***D***
Value	2,960	4	54	95	130	117

2. Student viewing videos is a random event in each day of the course duration, and the probability of the event is determined only by the factor A, B, and C we discovered in section III.

3. The incremental VV for any student is a constant value when the video viewing event occurs.

We define some symbols shown in Table 2 which are used in the design and calculation of the PVBM.

The total VV of students are accumulated with the daily VV of students during the whole course duration. By simulating the viewing behavior of each student and iterating the process day by day, we obtain the collections of VV of every student in a course. Then, the distribution of the simulated VV of the course can be calculated. The increase of a student's VV is depend on whether he or she will view videos on each day during the course duration. Therefore, based on the independent factors A, B and C in section III, we calculate the probability of student *i* to view videos in the *t*-th day by the following equation:

$$p_i(t) = \frac{w_{i,t}}{C} \times (1 - \frac{s_i}{D}) \times \frac{t}{E}$$

In the above equation, $p_i(t)$ consists of three parts:

1. $\frac{w_{i,t}}{C}$ represents the completion of student *i* in viewing videos, and it corresponds to the factor A in section III, where *C* is a constant. Therefore the bigger $w_{i,t}$ is, the higher probability student *i* will view videos.

2. $(1 - \frac{s_i}{D})$ represents how early student *i* starts viewing videos, and it corresponds to the factor B in section III, where *D* is a constant. Therefore the smaller s_i is, the higher probability student *i* will view videos.

3. $\frac{t}{E}$ represents how close the current date is to the exam date, and it corresponds to factor C in section III , where *E* is a fixed value. Therefore the bigger *t* is, the higher probability student *i* will view videos.

Based on the aforementioned equation, we use the following algorithm to simulate the video viewing behavior of all students, which includes:

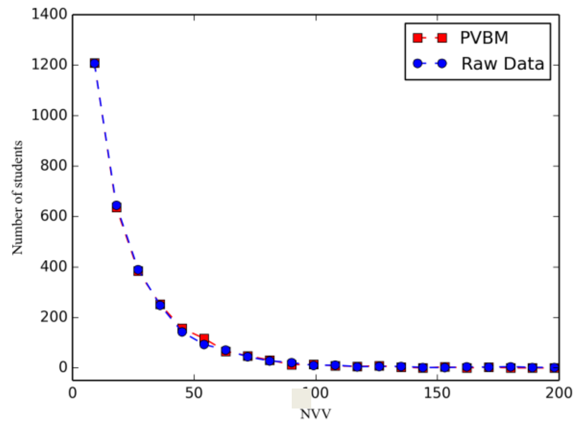
Step 1 If the current number of students equals to the total number *N*, then go to step 2, otherwise add *u* students to *L*, and initialize $w_{i,t}=\Delta w$, $s_i=t$.

Step 2 Calculate $p_i(t)$ for student *i* in *L*, and update $w_{i,t}$ as follows:

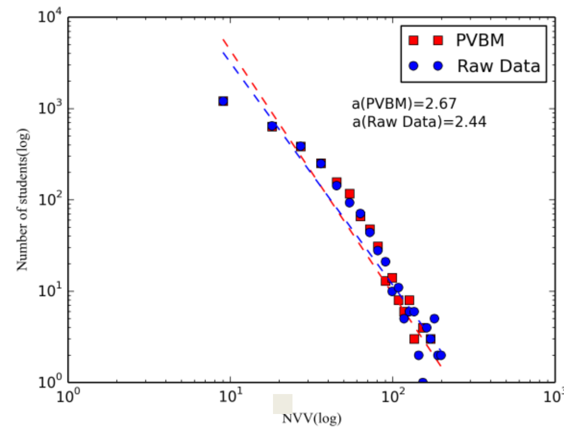
$$w_{i,t} = \begin{cases} w_{i,t-1} + \Delta w, & \text{with } p_i(t) \text{ probability} \\ w_{i,t-1}, & \text{with 1-} p_i(t) \text{ probability} \end{cases}$$

Step 3 If *t* equals to *D*, stop, otherwise increase *t* as *t*+1, and turn to step1.

We validate the PVBM as follows.



(a) Distribution of VV



(b) Distribution of VV (log-log scale)

Figure 7. Distribution of VV of Course No.162.

B. *Validation*

We use Course No.162 as the analysis object, and simulations are performed using the related parameters of the course which are shown in Table 3. We perform 100 times simulations to obtain the distribution of VV shown in Fig.7.

In Fig. 7, blue dots represent the raw data, and red dots represent the simulation data. It can be seen that the simulation result is fairly approximate to the actual distribution. Furthermore, Fig. 7 (b) shows the log-log scale plot of Fig. 7 (a). It is obvious that the simulated VV distribution yields to PLD. According to raw data and simulated data, the slopes of their fitting straight-lines are 2.44 and 2.67, respectively. The reason why the slope of simulation result is slightly larger than the raw data is that PVBM achieves high accuracy for students who have low VV and low accuracy for students who have high VV, shown in Fig. 7(b). The difference between the slopes may because that Δw is supposed to be constant. Actually, Δw may be affected by multiple factors and changes dynamically which will be studied in the future work.

In addition to the Course No.162, we validate other eight courses, each of which enrolled more than 1,000 students, and the simulation results are shown in Table 4. Observed

TABLE IV. STATISTICAL RESULTS OF 8 COURSES

Course No	Course Name	a (raw data)	a (PVBM)
8	Fundamentals of Computer Application	2.41	2.64
59099	Politics	2.30	2.54
1024	Distance Learning Methods	2.62	2.37
161	English I	2.50	2.74
193	Political Economics	2.52	2.26
65	Advanced Mathematics I	2.30	2.11
185	Introduction to Sociology	2.15	1.80
84	Management Science	2.27	1.91

from Table 4, there is only small error between a (PVBM) and a (Raw Data) of the eight courses which proves that the model works well. The error between a(PVBM) and a(Raw Data) of last two courses are bigger than the other courses due to the less number of students, for the reason that the more the students, the better the distribution of VV will follow a power law which makes the model work better.

V. APPLICATION OF THE PVBM

Based on the aforementioned model and course related parameters, the distribution of VV at the end of a semester can be estimated. By modifying the parameters in PVBM, the effect in results of VV distribution can be observed, which indicate different performance of students. Therefore, we can apply the PVBM in two crucial stages of teaching process to assist teachers in improving the performance of students.

STAGE I: Syllabus design. In XJTUDLC, all videos of a course are released at the beginning of a semester for students who registered that course, and they can learn at any time on their own schedule. Therefore, the structure and quantity of videos should be designed carefully before the semester starts so that students can achieve better performance. With the assistance of the PVBM, teachers can obtain the prediction of final video usage statistics with specified course parameters, which provide some guidance in designing the syllabus.

STAGE II: Syllabus implementation. During the semester, the prediction of video usage of each course can be calculated at any time based on the log data of current learning state. When the prediction result indicates poor learning performance at the end of semester, teachers can modify syllabus or notify students to achieve better performance.

In order to quantify the effect of applying PVBM in the two stages above, we define Course Completion Rate (CCR) to represent the video usage of students. For a specified course, the CCR for student i is computed as:

$$CCR_i = \frac{VV_i}{C}$$

TABLE V. VALUES EMPLOYED FOR EACH SCENARIO

Scenario	u	D	C	E	Δw	A/%	B/%	C/%
1	54	117	130	95	4	0.27	2.70	97.03
2	**74**	117	130	95	4	0.47	3.31	96.22
3	54	**137**	130	95	4	1.59	6.79	91.62
4	54	117	**150**	95	4	0.03	0.51	99.46
5	54	117	130	**75**	4	0.84	3.89	95.27
6	54	117	130	95	**5**	4.32	8.72	86.96

where VV_i is the VV of student i, and C is the number of course videos. According to the curriculum of XJTUDLC, in each course, students are required to view more than 60% of course videos, thus, students can be divided into three classes by their CCR:

CLASS A: CCR>=1. The students of this class have viewed all videos, and may viewed some videos more than once.

CLASS B: 0.6<=CCR<1. They viewed more than 60% of videos, but not all. In this case, the curriculum is accomplished.

CLASS C: CCR<0.6. They haven't accomplished the curriculum.

Based on the proposed CCR and three classes, we will take course No.162 as example to demonstrate the application of PVBM. The original course parameters of course No.162 are illustrated in Table 3.

During the STAGE I, teachers can apply the PVBM for prediction using the six parameters in Table 3. Among these parameters, the N (total number of students who should study a specified course, defined in Table 2) is confirmed and seldom changed during the whole process of learning, and the Δw can be determined by the calculation from course history logs. The other four parameters can be modified according to the curriculum and the expected distribution of CCR.

- The u. As analyzed in section III, the accumulated number of enrolled students increases almost linearly and the VV is negative correlation with the start date of study. Therefore, in order to increase ratio of CLASS A+B, the u can be set to bigger value, as shown in Table 5. According to above analysis, teachers can set an earlier "deadline" for enrollment to boost the number of enrolled students.
- The C. The number of course videos shows negative correlation with the ratio of CLASS A+B, as shown in Table 5, which indicates that C affect the viewing behavior obviously. When C increases, students must spend more time in selecting and viewing videos, thus the CCR decreases oppositely. As a result, teachers can adjust the course structure and chapter arrangement, such as reducing the number of course videos, centralizing important content in beginning chapters, etc.

- The *D*. The duration is one of the most important aspects of a course, which is determined by many factors such as course difficulty, amount of contents, etc. In order to achieve higher ratio of CLASS A+B, the proposed PVBM suggests a bigger *D*, as shown in Table 5.
- The *E*. In the duration of a course, after the exam, there are several days or weeks left for students to finish their homework or project report. According to the analysis in section III, students view more videos when close to exam date. When *E* is set to a smaller value, the ratio of CLASS A+B will increase as shown in Table 5.

Since the STAGE II begins, the *N*, *E*, *C* and *D* are confirmed and hardly modified. Meanwhile, the Δw, *u* can be modified to affect the final distribution of CCR and provide guidance to teachers.

- The *u*. As analyzed above, the PVBM suggests a bigger *u* to boost the enrollment. However, if the number of enrolled students increases slowly, teachers can take advantage of PVBM to calculate a new proper *u* on the basis of current state, and urge those students who haven't enrolled to start learning.
- The Δw. It is not required to view certain amount of videos for each student on one day. Nevertheless, we find this parameter affect the ratio of CLASS A+B significantly, as shown in Table 5. Obviously, the more videos a student views in one day, the higher CCR he/she will achieve. With the prediction of PVBM based on the current state, teachers can advise students to view more videos in one day to achieve higher CCR.

VI. CONCLUSION AND FUTURE WORK

Along with the increasing number of students on e-Learning, the PLD of the viewing behavior emerged in large-scale e-Learning environment, which is obviously different from those statistical results in small-scale e-Learning or traditional classroom environment. In this paper, we present a statistical analysis of log data of a large-scale e-Learning platform (XJTUDLC), in which we discover four factors related to the video viewing behavior based on the mechanisms for generating PLDs. Furthermore, we propose the PVBM based on the four factors and validate its accuracy with nine online courses. We finally analyze the application of PVBM and provide some suggestions for teachers to improve the viewing performance of students.

Future work will involve more studies on the distribution law of Δw and related factors to improve the accuracy of the proposed model. In addition, we will apply this model to analyze other learning behaviors and provide practical suggestions to faculty.

ACKNOWLEDGMENT

This research was partially supported by the National Science Foundation of China under Grant Nos. 91118005,91218301, 61472317 and 61428206, the Ministry of Education Innovation Research Team No. IRT13035, and the National Key Technologies R&D Program of China under Grant No. 2013BAK09B01.

REFERENCES

[1] P. C. Sun, R. J. Tasi, G.Finger, Y. Y. Chen, and D. Yeh. "What drives a successful e-Learning? An empirical investigation of the critical factors influencing student satisfaction," Computers & Education. vol. 50, 2008, pp. 1183-1202.

[2] A. Hassanzadeh, F. Kanaani, and S. Elahi. "A model for measuring e-learning systems success in universities," Expert System with Applications. vol. 39, no. 12, 2012, pp. 10959-10966.

[3] L. P. Deng, and N. J. Tavares. "From Moodle to Facebook: Exploring students' motivation and experiences in online communities,". Computers & Education. vol. 69, 2013, pp. 167–176.

[4] Number of Students of Formal Education by Type and Level in 2013. "http://www.moe.gov.cn/". Accessed: 2015-02-01.

[5] C. Romero, and S. Ventura. "Educational Data Mining: A Review of the State of the Art, " IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews. vol. 40, no. 6, 2010, pp. 601-618.

[6] A. P. Ayala. "Educational data mining: A survey and a data mining-based analysis of recent works, " Expert Systems with Applications. vol. 41, no. 4, 2014, pp. 1432-1462.

[7] L. Breslow, D. E. Pritchard, J. DeBoer, G. S. Stump, A. D. Ho, and D. T. Seaton. "Studying Learning in the Worldwide Classroom Research into edX's First MOOC," Journal of Research & Practice in Assessment, vol. 8, 2013, pp. 13-25.

[8] H. Miller, P. Haller, and L.Rytz. "Functional Programming For All! Scaling a MOOC for Students and Professionals Alike," Proc. 36th Int. Conf. on Software Engineering (ICSE 2014), New York: ACM, 2014, pp. 256-263, doi: 10.1145/2591062.2591161.

[9] P. Mitros, and F. Sun. "Create Educational Resources at Scale," Proc. 14th Int. Conf. on Advanced Learning Technologies (ICALT 2014), IEEE Press, 2014, pp. 16-18, doi: 10.1109/ICALT.2014.15.

[10] Kleftodimos A, Evangelidis G. "Exploring Student Viewing behavior in Online Educational Videos," IEEE International Conference on Advanced Learning Technologies, IEEE Computer Society, 2014, pp. 367-369.

[11] Guo P J, Kim J, Rubin R. "How video production affects student engagement: an empirical study of MOOC videos," First Acm Conference on Learning, 2014, pp. 41-50.

[12] Mullaney T, Reich J. "Staggered Versus All-at-Once Content Release in Massive Open Online Courses: Evaluating a Natural Experiment," Learning at Scale, Vancouver: ACM, 2015, pp. 185-194.

[13] Barabasi A, Albert R. "Emergence of scaling in random networks," Science, 1999, 286(5439), pp. 509-512.

[14] Bak P. "How nature works: The Science of Self-Organized Criticality," American Journal of Physics, 1997, 65(6).

[15] M. E. J. Newman. "Power laws, pareto distributions and zipf's law," Contemporary Physics. vol.46, no. 5, 2005, pp. 323-351.

[16] B. Dong, Q. H. Zheng, M. Qiao, J. Shu, and J. Yang. "BlueSky cloud framework: an e-Learning framework embracing cloud computing, " In Cloud Computing, Springer Berlin Heidelberg, 2009, pp. 577-582.

[17] Q. H. Zheng, H. He, T. Ma, N. Xue, B. Li, and B. Dong. "Big Log Analysis for e-Learning Ecosystem," Proc. IEEE 11th Int. Conf. on e-Business Engineering (ICEBE 2014), IEEE Press, Nov. 2014, pp. 258-263, doi: 10.1109/ICEBE.2014.51.

[18] F. Rozenfeld A, Laneri K, V. Albano E. "Critical dynamic approach to stationary states in complex systems," European Physical Journal Special Topics. vol. 143, no. 1, 2007, pp. 3-8.