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

Ni Xue

SPKLSTN Lab,Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China

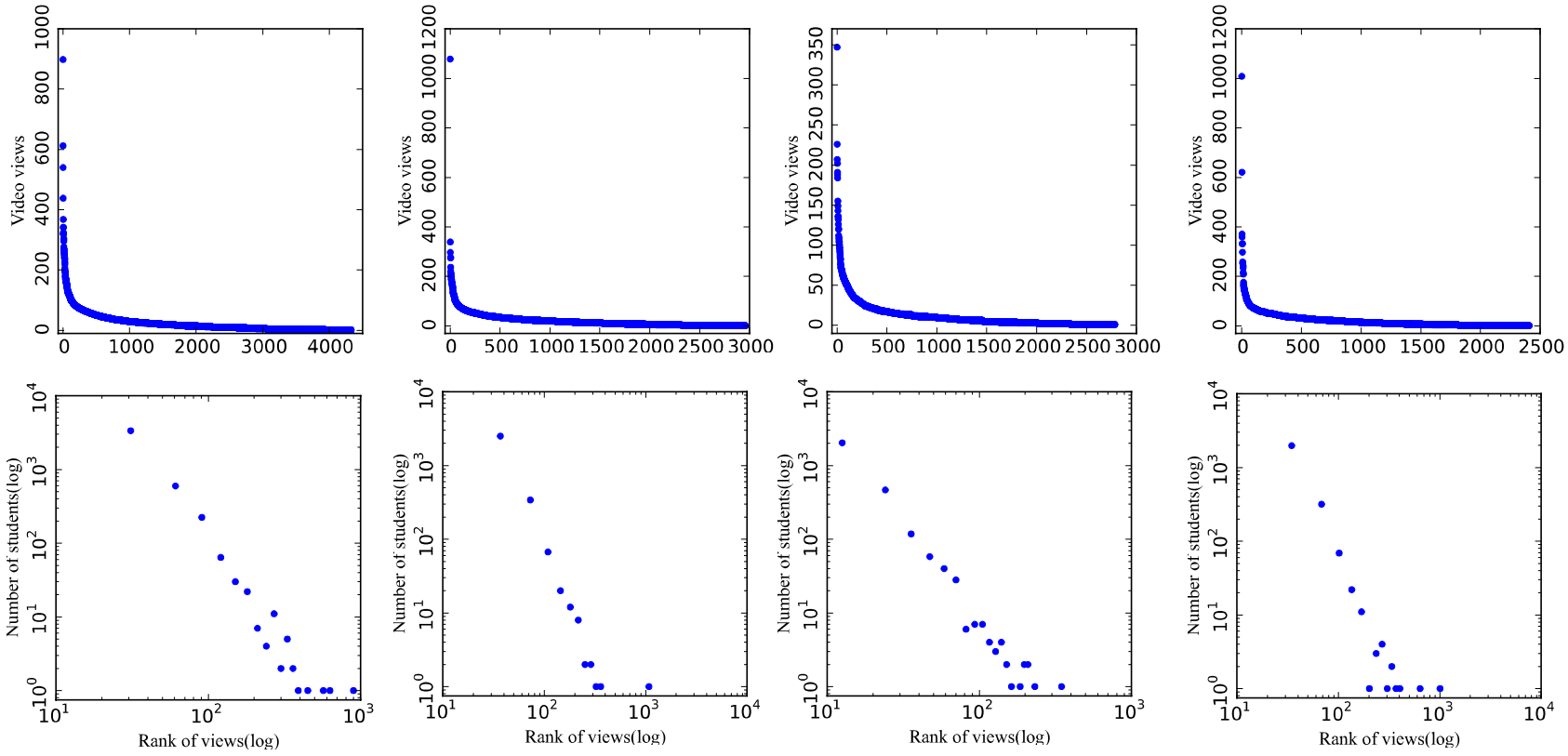
Outline

- Introduction
- Preliminaries
- Statistical analysis
- Modeling
- Application
- Conclusion and future work

Introduction

- Problem

  the distribution of Video Views (VV) follows the PLD



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

Introduction

- Questions on viewing behavior
  - Q1: Factors?
  - Q2: Modeling?
- Motivation
  - Q1: Explore the laws of the viewing behavior and their causes
  - Q2: Better understand the PLD of the VV
- Methodology
  - Statistical analysis
  - Generative model

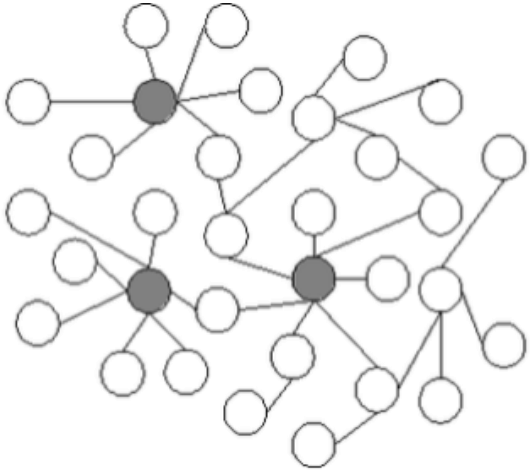
Preliminaries

- XJTUDLC platform
  - major teaching method: video lecture
  - three-part-separated screen coursewares
  - Big Log Analysis System (BLAS)

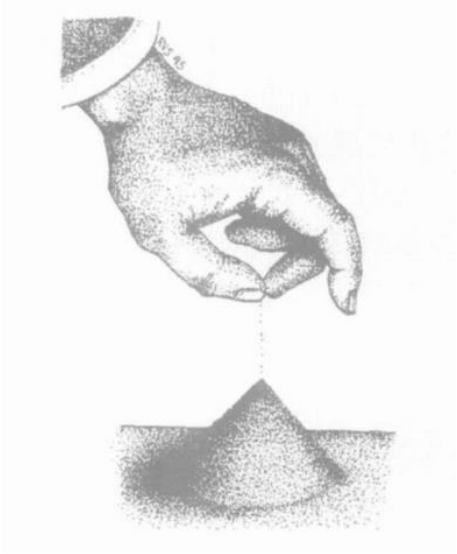
- Our dataset
  - the log data of XJTUDLC between 2014.09 and 2015.01
  - 5,028,459 log records, including 268 courses, 52,340 videos and 13,238 students.

Statistical analysis--Mechanisms for generating PLDs

- Preferential attachment
- Self-organized criticality



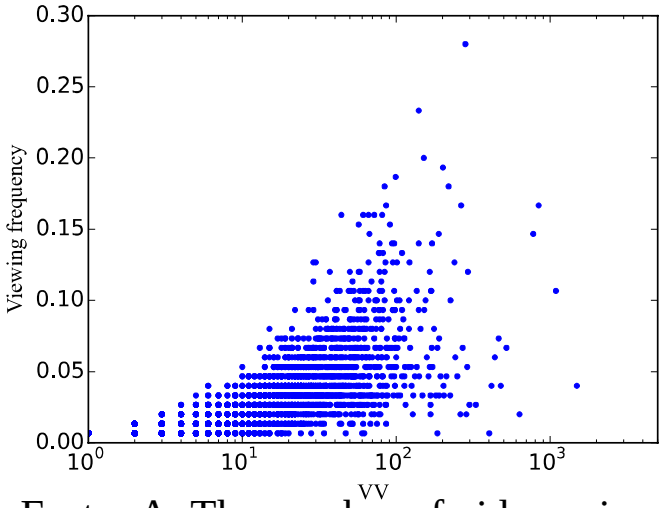
scale-free network



sandpile model

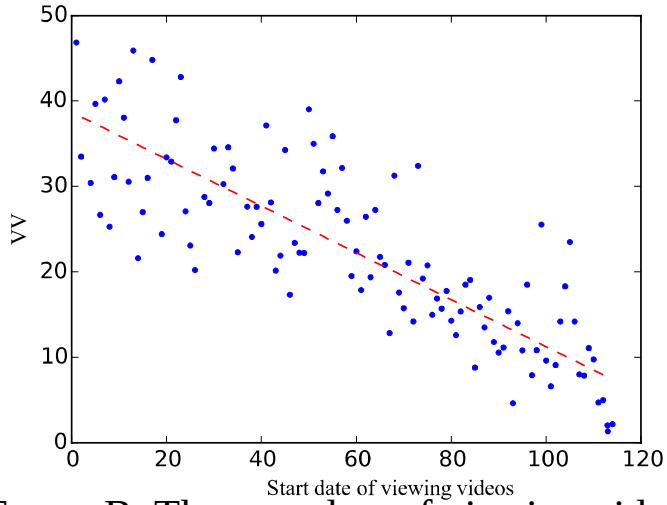
- Random walks

Statistical analysis—Factors influencing the VV



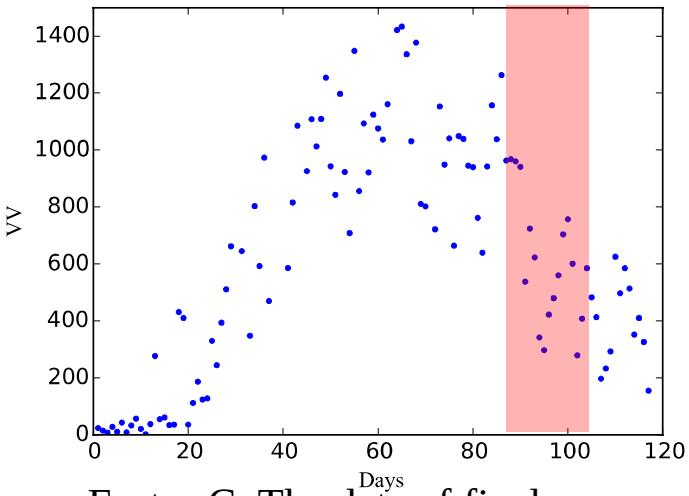
Factor A: The number of videos viewed

(a)



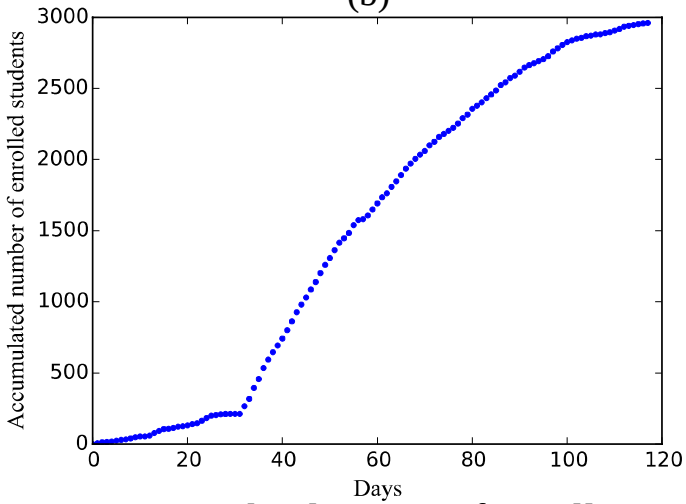
Factor B: The start date of viewing videos

(b)



Factor C: The date of final exam

(c)



Factor D: The duration of enrollment

(d)

Modeling--Assumptions

- The Probabilistic Viewing Behavior Model(PVBM)
  - Assumptions
    - The number of newly enrolled students of a course in each day is a constant value.
    - 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.
    - The incremental VV for any student is a constant value when the video viewing event occurs.

Modeling--Variable definition

- 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 *t*th day
s_i	The number of days from the course start date to the date student *i* start viewing videos
$p_i(t)$	The probability for student *i* to view videos on the *t*th 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 date
C	The number of course videos
D	The number of days of the course duration

Modeling--Probability calculation

Based on the independent factors A, B and C, we calculate the probability of student *i* to view videos in the *t*-th day by the following equation:

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

- $\frac{w_{i,t}}{C}$: the completion of student *i* in viewing videos(factor A)
- $(1 - \frac{s_i}{D})$: how early student *i* starts viewing videos(factor B)
- $\frac{t}{E}$: how close the current date is to the exam date(factor C)

Modeling--Algorithm

With the aforementioned equation, we use the following algorithm to simulate the video viewing behavior of all students:

- **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:

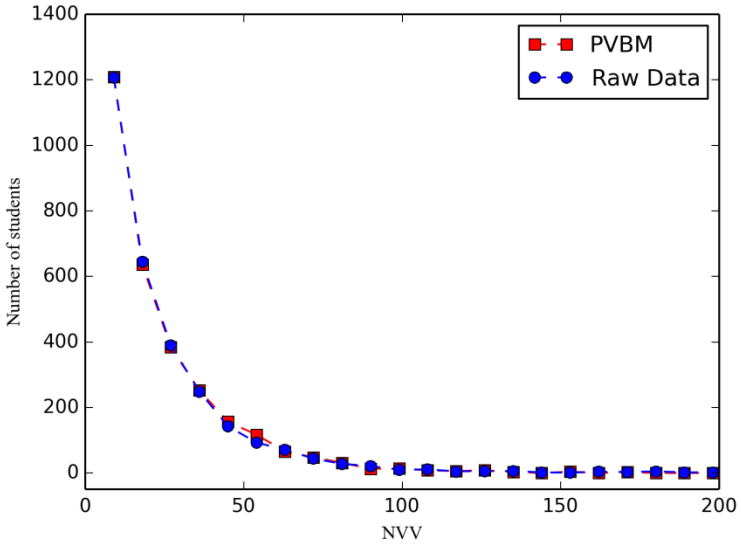
$$w_{i,t} = \begin{cases} w_{i,t-1} + \Delta w, & \text{with } p_i(t) \text{ probability} \\ w_{i,t-1}, & \text{with } 1\text{-}p_i(t) \text{ probability} \end{cases}$$

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

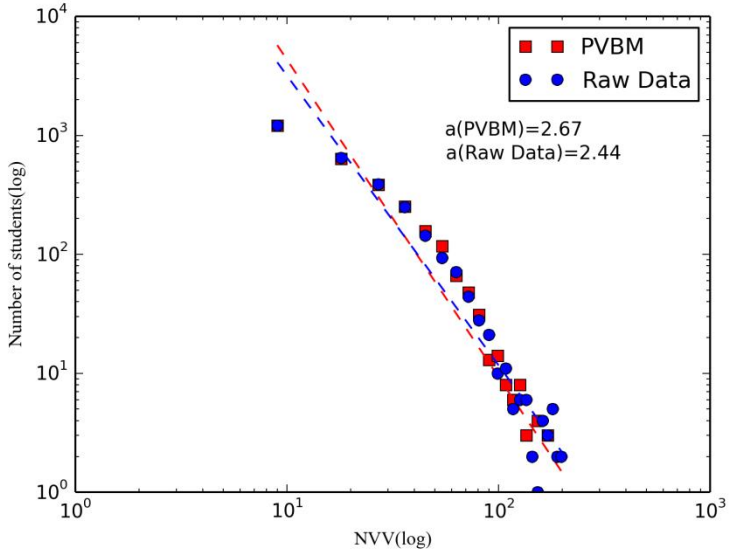
Modeling--Validation

Parameters of Course No.162

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



(a) Distribution of VV



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

PVBM achieves high accuracy for students who have low VV and low accuracy for students who have high VV

Modeling--Validation

In addition to the Course No.162, we validate other eight courses, each of which enrolled more than 1,000 students

Course No	Course Name	α (raw data)	α (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

Application--Course Completion Rate

- Crucial stages of teaching process
  - STAGE I: Syllabus design.
  - STAGE II: Syllabus implementation.
- Course Completion Rate (CCR)

For a specified course, the CCR for student *i* is computed as:

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

VV_i → the VV of student *i*

C → the number of course videos

- Three classes of students
  - CLASS A: CCR>=1
  - CLASS B: 0.6<=CCR<1
  - CLASS C: CCR<0.6

Application--Parameter adjustment strategies

Parameters of Course No.162

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

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

- STAGE I: u↑ C↓ D↑ E↑
- STAGE II: u ↑ Δw ↑

→ better viewing performance

Future work

- improve the accuracy of the proposed model
  - more studies on the distribution law of Δw
  - more factors related to viewing behabior
- analyze other learning behaviors
- provide practical suggestions to faculty

Thank you!