XuetangX
—The Platform for Chinese MOOC

Jie Tang
Tsinghua University

www.xuetangX.com
2012: The Year of MOOC

- **108 partners**
- **633 courses**
- **7.1 million users**

- **20+ courses**
- **100,000 users**
- **Chinese EDU association**

- **50+ partners**
- **160+ courses**
- **2.1 million users**

- **~10 partners**
- **40+ courses**
- **1.6 million users**

- **host >900 courses**
- **millions of users**
Outline

ี้ The platform—xuetangX.com
ี้ Our courses
ี้ Some interesting findings
## Milestones

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 May</td>
<td>Tsinghua University joined edX</td>
</tr>
<tr>
<td>24 May</td>
<td>Tsinghua Center for Massive Online Education</td>
</tr>
<tr>
<td>06 June</td>
<td>edX releases OpenEdX framework</td>
</tr>
<tr>
<td>08 June</td>
<td>Tsinghua sets up XuetangX Team</td>
</tr>
<tr>
<td>10 Oct</td>
<td>Tsinghua releases XuetangX platform</td>
</tr>
</tbody>
</table>
Our Team

📚 Tsinghua Center for Massive Online Education
  ➢ Jointly set up by Dept. Computer Science, Institute for Interdisciplinary Information Sciences, School of Social Sciences, and Institute of Education

📚 XuetangX Team
  ➢ 30 members from different departments
  ➢ Most from Dept. Computer Science
  ➢ Some from Tsinghua Press and Dept. Electronic Engineering
Brief Introduction

- Develop based on OpenEdX
- XuetangX Team adds and modifies source codes of ~100,000 lines
Key Tech 1: i18n and l10n

Help edX in i18n

Revise 2890 lines of edX to support i18n and already merged in OpenEdX
Key Tech 1: i18n and l10n

Translate about 30 thousand English words into Chinese and finish l10n

The work speeds up i18n of edX. Currently. There are 30+ languages are performing translation for l10n. Chinese, French, Korean, Portuguese and Spanish have finished translation.
Key Tech 2: Video Player

- Develop video player based on HTML 5 / Flash
  - Independent of YouTube. Support multiple sources
  - Support a unified interface to upload videos in CMS
Key Tech 3: Course Search

- Support search of video captions in Chinese/English
- Click search results to play the video from the expected point
Key Tech 4:
WYSIWYG Editor of Math Formulas

- WYSIWYG editor of math formulas can facilitate teachers / students for math input
  - Support rich math symbols for input
  - Comfortable input experience in Web browsers supported by HTML + CSS
  - Integrate with rich text editor

\begin{equation}
\int x^\mu \, dx = \frac{x^{\mu + 1}}{\mu + 1}
\end{equation}

LaTeX editor by EdX

WYSIWYG editor by XuetangX
Key Tech 5: Auto Grading for Programming Homework

Student can upload their source codes and get auto grading results
After several months of operation, xuetangX has attracted ~100K users and the number is continuously increasing…
In Service

- Support ~20 Tsinghua MOOCs simultaneously with edX
  - Principles of Electric Circuits; History of Chinese Architecture; Data Structure; Historical Relic Treasures and Cultural China; Financial Analysis and Decision Making

- Partners’ courses
  - MIT: Circuits and Electronics
  - UC Berkeley: Cloud Computing and Software Engineering
  - Peking University: Principles and Practice of Computer Aided Translation

- Support 2 Tsinghua SPOCs
  - C++ Programming by Prof. ZHENG Li for 93 students
  - Cloud Computing and Soft Engineering by Prof. XU Wei for 35 students
User enrolment in the past months
Outline

- The platform—xuetangX.com
- Our courses
- Some interesting findings
Courses

~20 courses offered by 4 partners (MIT, UC Berkeley, Peking U, Tsinghua)

- Data Structure
- Principles and Practice of Computer Aided Translation
- Financial Analysis and Decision Making
- Historical Relic Treasures and Cultural China
- Cloud Computing and Soft Engineering
- Circuits and Electronics
- Principles of Electric Circuits
- Introduction to Psychology
- Combinatorics
- Football Science
- …

7 Tsinghua courses also on edX

- History of Chinese Architecture; Data Structure; Historical Relic Treasures and Cultural China; Financial Analysis and Decision Making
# Enrolment on edX

- Financial: 27,000
- Chinese Culture: 4,000
- Chinese Architecture: 1,000
- Circuits and Electronics: 4,000
- Data Structure: 7,000
Enrolment on XuetangX

A survey conducted on 6,116 randomly chosen Internet users before we released xuetangX…

Favorite courses

- Computer
- Business
- Culture
- Arts
- Energy
- Engineering
- Biology
- Education

Percentage
Enrolment on XuetangX

Favorite courses by survey

- Computer: 60.00%
- Business: 40.00%
- Culture: 50.00%
- Arts: 30.00%
- Energy: 10.00%
- Engineering: 20.00%
- Biology: 10.00%
- Education: 5.00%

#enrolment (2013)

- Data Structure
- Computer Aided Translation
- Financial
- Chinese Culture
- Chinese Architecture
- Circuits and Electronics
- Principle of Electric
Enrolment on XuetangX

Favorite courses by survey

#enrolment (2014)
Two SPOC courses in Tsinghua

- Two Tsinghua SPOCs
  - C++ Programming for 93 students
  - Cloud Computing and Soft Engineering for 35 students

- Deploying multiple SPOC systems in China
  - Wenzhou University
  - Jiangxi Finance Institute
  - …
Outline

- The platform—xuetangX.com
- Our courses
- Some interesting findings
Rich tracking logs of student behaviors

- The huge amount of data available in MOOC offers a unique opportunity for understanding student behavior.
- Such logs include: watch video, homework, forum, etc.

<table>
<thead>
<tr>
<th>Item</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>76,215</td>
</tr>
<tr>
<td>Courses</td>
<td>9</td>
</tr>
<tr>
<td>Logs</td>
<td>~7.5M</td>
</tr>
<tr>
<td>Date span</td>
<td>2013/09/28-2014/01/15</td>
</tr>
</tbody>
</table>
Recent Research

- Personalized service based on education data analysis
- Learning behavior analysis
- Peer assessment based on social networks
- Intelligent course QA systems
- Auto grading for student essays
- Virtual labs
- Mobile apps
- …
One particular question

One fact: 76,215 users and only 3%-6% received the certificates

An interesting question is: Who finally received the certificates?
Age+Education vs. Certificate
Gender+Age vs. Certificate
Gender + Location vs. Certificate

The diagram illustrates the probability of obtaining different certificates based on gender and location. The x-axis represents the gender group (male, female), and the y-axis represents the probability. Different locations are represented by various colors, as indicated in the legend.
Forum vs. Certificate

![Graph showing activity vs. participation times in forum](image-url)
#friends vs. Certificate

![Graph showing the relationship between the number of friends and the probability of obtaining a certificate.]
Influence vs. Certificate

![Bar chart showing influence vs. certificate]

- **weak tie**
- **strong tie 3**
- **strong tie 5**
Can we predict who will/could receive the certificate

Given behavior log data by all users in the MOOC system,

Predict whether a user will finally graduate and receive the certificate of a specific course.
Factorization Machines

The prediction of feature vector $x_i$:

$$y(x_i) = w_0 + \sum_{j=1}^{d} w_j x_{ij} + \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} x_{ij} x_{i'j} \langle p_j, p_{j'} \rangle$$

and the corresponding objective function:

$$O = \sum_{x_i} (y(x_i) - y_i)^2 + \lambda \sum_{i=1}^{d} \|p_i\|^2$$
### Preliminary Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Features</th>
<th>AUC</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factorization Machines</td>
<td>Demographics</td>
<td>90.80</td>
<td>5.91</td>
<td>45.24</td>
<td>9.89</td>
</tr>
<tr>
<td></td>
<td>+ Social Influence</td>
<td>96.28</td>
<td>27.90</td>
<td>51.89</td>
<td>36.53</td>
</tr>
<tr>
<td>SVM</td>
<td>Demographics</td>
<td>84.36</td>
<td>5.54</td>
<td>42.31</td>
<td>9.81</td>
</tr>
<tr>
<td></td>
<td>+ Social influence</td>
<td>96.49</td>
<td>25.90</td>
<td>50.85</td>
<td>34.27</td>
</tr>
</tbody>
</table>

* SVM is a state-of-the-art algorithm for classification/prediction. We use it as the baseline method in our experiments.*
Outline

- The platform—xuetangX.com
- Our courses
- Some interesting findings
- Summaries
Summaries

- We have developed and deployed one of the largest MOOC platforms in China
  - attracted hundreds of thousands of users

- Interesting findings
  - “Financial analysis” is the most popular course
  - Most users are browsers, and 3-6% are certificate earners
  - There is a strong influence between users’ learning behavior
Future Research Directions

- Personalized learning path recommendation
- Learning pattern discovery
- Peer assessment based on social networks
- Intelligent course QA systems
- …
Thanks!

http://xuetangx.com
http://keg.cs.tsinghua.edu.cn/jietang
Age+Education vs. Certificate

The graph shows the probability distribution of various age groups and education levels. The x-axis represents different age groups, and the y-axis represents the probability. The bars in different colors correspond to different education levels:

- PhD
- MS
- BS
- Junior colleague
- High school
- Junior high school
- Elementary school

Each bar group is divided into sub-groups based on age ranges: <18, [18,22], [23,27], [28,35], [36,50], >50. The graph visually compares the probability distributions across these categories.