Can AI help MOOCs?

Jie Tang
Tsinghua University

The slides can be downloaded at http://keg.cs.tsinghua.edu.cn/jietang
Big Data in MOOC

149 partners
2,000+ courses
24,000,000 users

1,000+ courses
8,000,000 users
Chinese EDU association

110 partners
1,270 courses
10,000,000 users
10+ MicroMaster

~10 partners
40+ courses
1.6 million users
"nanodegree"
Some exciting data...

- Every day, there are 5,000+ new students.
- An MOOC course can reach 100,000+ students.
- >35% of the XuetangX users are using mobile.
- traditional->flipped classroom->online degree.
Some exciting data…

• Every day, there are 5,000+ new students
• An MOOC course can reach 100,000+ students
• >35% of the XuetangX users are using mobile
• traditional->flipped classroom->online degree
• “Network+ EDU” (O2O)
  – edX launched 10+ MicroMaster degrees
  – Udacity launched NanoDegree program
  – GIT+Udacity launched the largest online master
  – Tsinghua+XuetangX will launch a MicroMaster soon
However...

- **only ~3% certificate rate**
  - The highest certificate rate is 14.95%
  - The lowest is only 0.84%

- Can **AI** help MOOC and how?
MOOC user = Student?

How to learn more effectively and more efficiently?

• Who is who? background, where from?
• Why MOOC? motivation? degree?
• What is personalization? preference?
MOOC course = University course?

How to discover the prerequisite relations between concepts and generate the concept graph automatically?

Thousands of Courses

How to leverage the external knowledge?
However to improve the engagement?
LittleMU (小木)
What is LittleMU(“小木”)

• Not a Chatbot
  – “Good morning”, “did you have the breakfast?”—NO
• Not a teacher/TA
  – “Can you explain the equation for me?” — NO

• Instead, “小木” is more like a learning peer
  – Tell you some basic knowledge in her mind
  – Tell you what the other users are thinking/learning
  – Try to understand your intention
  – Teach “小木” what you know
What is LittleMU (“小木”)

早上好

吃了吗

这个问题小木无法回答，请换一个问题～😊

问：快速排序为什么没有讲？但是出现在字幕里了，和视频不匹配
RT
我稍微看了下后续课堂的小标题，貌似没有快速排序的章节。字幕和视频不匹配，这是视频没有放上来么？

答：快速排序在第5章讲。

来自《程序设计基础》讨论区
What is LittleMU(“小木”)

小木机器人是你的智能学习助理，由学堂在线和清华大学联合研发，集知识查询、学习问答、学习管理、机器人客服于一身。目前，小木可以解答一部分课程学习疑问、平台使用疑问；小木会给你提问引导你的学习（试验中，部分课程可见）；小木会作诗逗你开心~~~更多的功能敬请期待噢~~😊

来自问答库
LittleMU (小木)

二叉树
在计算机科学中，二叉树是每个节点最多有两个子树的树结构。通常子树被称作左子树（left subtree）和右子树（right subtree）。二叉树常被用于实现二叉查找树和二叉堆。二叉树的每个结点至多只有二颗子树(不存在度大于2的结点)，二叉树的子树有左右之分，次序不能颠倒。二叉树的第i层至多有2^(i-1)个结点；深度为k的二叉树至多有2^k-1个结点；对任何一颗二叉树T，如果其终端结点数为n_0，度为2的结点数为n_2，则n_0=n_2+1。—模深度为k，且有2^k-1个节点称之为满二叉树，深度为k，有n个节点的二叉树，且当且仅当其每一个节点都与深度为k的满二叉树中，序号为1至n的节点对应时，称之为完全二叉树。
另有：
满二叉树

...同学们好，我们这一节介绍树的一种特殊，但又不失代表性的特例。
Acrostic Poem: 小木作诗
LittleMU (小木)

LittleMU: Intelligent Interaction

1. User analysis
   - Behavior modeling
   - User Profiling

2. Course analysis
   - Incentive analysis
   - Course recommendation
   - Automated video navigation
   - Question answering

3. Content analysis
   - Concept extraction
   - Prerequisite relation mining

Behavior logs

Knowledge base
LittleMU (小木)

LittleMU: Intelligent Interaction

1. User Profiling
   - Behavior modeling
   - User modeling

2. Course analysis
   - Incentive analysis
   - Course recommendation
   - Automated video navigation
   - Question answering

3. Content extraction
   - Concept extraction
   - Prerequisite relation mining

Knowledge base
MOOC user

• Who is who? background, where from?
• Why MOOC? motivation? degree?
• What is personalization? preference?
Basic Analysis

![Graphs showing basic analysis](image-url)
Observation 1 – Gender Difference

Model 1: Demographics vs Certificate
Model 2: Demographics + Forum activities vs Certificate

- Females are significantly more likely to get the certificate in non-science courses.
- The size of the gender difference decreases significantly after we control for forum activities.

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

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Science (1)</td>
<td>Science (2)</td>
</tr>
<tr>
<td>Female</td>
<td>0.014***</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>New Post</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Reply</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Video</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Assignment</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Bachelor</td>
<td>0.014***</td>
<td>0.003*</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Graduate</td>
<td>0.007***</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Effort</td>
<td>-0.072***</td>
<td>-0.072***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.286***</td>
<td>0.018***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Obs.</td>
<td>74,480</td>
<td>19,269</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.024</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Observation 2 – Ability v.s. Effort

Table 4: Regression Analysis for Certificate Rate: All Users

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Science</td>
<td>Science</td>
</tr>
<tr>
<td>Female</td>
<td>0.014***</td>
<td>0.002*</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>New Post</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Reply</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Video</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Assignment</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Bachelor</td>
<td>0.014***</td>
<td>0.003*</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Graduate</td>
<td>0.007***</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Effort</td>
<td>-0.072***</td>
<td>-0.072***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.286***</td>
<td>0.18**</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Obs.</td>
<td>74,480</td>
<td>19,269</td>
</tr>
<tr>
<td>R^2</td>
<td>0.024</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Model 1: Demographics vs Certificate
Model 2: Demographics + Forum activities vs Certificate

- **Bachelors** students are significantly more likely to get the certificate in non-science courses.
- **Graduate** students are more likely to get the certificate in science courses. After controlling for learning activities, the size of the effect is almost doubled.
- **Forum activities** are good predictors for getting certificates.
Forum activity vs. Certificate

— It is more important to be presented in forum, while the intensity matters less.

“近朱者赤” (Homophily)
— Certificate probability tripled when one is aware that she has certificate friend(s)
Dynamic Factor Graph Model

**Model:** incorporating “embedding” and factor graphs

\[
Y^t(i)^* = f(W_y Z^t(i) + b_y)
\]

\[
Z^t(i)^* = f(W_d S^t(i) + b_d)
\]

\[
S^t(i) = [Z_{t-1}^T(i), X^t(i)^T]^T
\]

**Prediction labels:**
Activities we are interested in, e.g., assignments performance and getting certificates.

\[
Y^t(i) = [Y_{t,i,0}, Y_{t,i,1}, \ldots, Y_{t,i,n-1}]^T
\]

**Latent learning states**
Every student’s status in at time \( t \) is associated with a vector representation

\[
Z^t(i) = [Z_{t,i,0}, Z_{t,i,1}, \ldots, Z_{t,i,m-1}]^T
\]

**All features:** time-varying attributes:
1. Demographics
2. Forum Activities
3. Learning Behaviors

\[
X^t(i) = [X_{t,i,0}, X_{t,i,1}, \ldots, X_{t,i,d-1}]^T
\]

Certificate Prediction

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

- LRC, SVM, and FM are different baseline models
- LadFG is our proposed model
Predicting more

• **Dropout**
  – KDDCUP 2015, 1,000+ teams worldwide

• **Demographics**
  – Gender, education, etc.

• **User interests**
  – computer science, mathematics, psychology, etc.

• ...
User Tagging

• **Observation**: With probability 43.91%, a user will enroll in a course in the same category as the last course (s)he enrolled in.

• **Method**: Use course categories to tag users who enroll in courses under this category to aid course recommendation.
Random Walk with Restart

• Use RWR on the user-tag bipartite with # of enrolled courses in the tag (category) as edge weight to generate tag preference of users.

• Offline test in course recommendation

<table>
<thead>
<tr>
<th></th>
<th>top1</th>
<th>top3</th>
<th>top5</th>
<th>top10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.0071</td>
<td>0.0247</td>
<td>0.0416</td>
<td>0.0890</td>
</tr>
<tr>
<td>+Tag</td>
<td>0.0185</td>
<td>0.0573</td>
<td>0.1022</td>
<td>0.2198</td>
</tr>
</tbody>
</table>
LittleMU (小木)

LittleMU: Intelligent Interaction

1. Behavior Modeling
   - User Profiling
   - Behavior modeling

2. Course Analysis
   - Course recommendation
   - Automated video navigation
   - Question answering

3. Incentive Analysis

Knowledge base

User Modeling
Intervention
Content Analysis
Knowledge Graph

- How to extract concepts from course scripts?
- How to recognize (prerequisite) relationships between concepts?

[1] Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang. Prerequisite Relation Learning for Concepts in MOOCs. ACL'17.
In this course, we will teach some basic knowledge about **data mining** and its application in **business intelligence**.

**Video script**

```
data mining
0.8 0.2 0.3 ... 0.0 0.0

business intelligence
0.1 0.1 0.2 ... 0.8 0.7
```

**Vector representation**
Learned via embedding or deep learning
Prerequisite Relationship

How to extract the prerequisite relationship?

[1] Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang. Prerequisite Relation Learning for Concepts in MOOCs. ACL'17.
Prerequisite Relationship Extraction

• Step 1: First extract important concepts
• Step 2: Use Word2Vec to learn representations of concepts

```
data mining
0.8 0.2 0.3 ... 0.0 0.0

business intelligence
0.1 0.1 0.2 ... 0.8 0.7
```

Vector representation
Learned via embedding or deep learning
Prerequisite Relationship Extraction

- **Step 1**: First extract important concepts
- **Step 2**: Use Word2Vec to learn representations of concepts
- **Step 3**: Distance functions
  - Semantic Relatedness
  - Video Reference Distance
  - Sentence Reference Distance
  - Wikipedia Reference Distance
  - Average Position Distance
  - Distributional Asymmetry Distance
  - Complexity Level Distance
### Result of Prerequisite Relationship

<table>
<thead>
<tr>
<th>Classifier</th>
<th>ML</th>
<th>DSA</th>
<th>CAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>SVM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>63.2</td>
<td>60.1</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>68.5</td>
<td>72.4</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>65.8</td>
<td>65.7</td>
</tr>
<tr>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>58.0</td>
<td>58.2</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>58.1</td>
<td>60.5</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>58.1</td>
<td>59.4</td>
</tr>
<tr>
<td>LR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>66.8</td>
<td>67.6</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>60.8</td>
<td>61.0</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>63.7</td>
<td>64.2</td>
</tr>
<tr>
<td>RF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>68.1</td>
<td>71.4</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>70.0</td>
<td>73.8</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>69.1</td>
<td>72.6</td>
</tr>
</tbody>
</table>

Table 2: Classification results of the proposed method(%).

---

[1] Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang. Prerequisite Relation Learning for Concepts in MOOCs. ACL'17.
System Deployed
LittleMU (小木)

LittleMU: Intelligent Interaction

1. Behavior logs
   - User Profiling
   - Behavior modeling
   - User analysis

2. Course analysis
   - Concept extraction
   - Prerequisite relation mining

3. Incentive analysis
   - Course recommendation
   - Automated video navigation
   - Question answering

Knowledge base

User Modeling

Intervention

Content Analysis
What we can do?

User modeling

Knowledge

- artificial intelligence
  - data mining
    - data clustering
  - machine learning
    - association rule
• Let start with a simple case
  – *Course recommendation* based on user interest
Course Recommendation

Course topic analysis

Low frequency
- LDA training
- User clustering
- Course prequisite modeling

High frequency
- Latent interest modeling
- Collaborative filtering

Recommendation result

Rule based adjustment

With the learned user model

[1] Xia Jing, Jie Tang, Wenguang Chen, Maosong Sun, and Zhengyang Song. Guess You Like: Course Recommendation in MOOCs. WI’17.
Course Recommendation

Guess you like
Top-k recommendation accuracy (MRR)

Comparison methods:
- HCACR – Hybrid Content-Aware Course Recommendation
- CACR – Content-Aware Course Recommendation
- IBCF – Item-Based Collaborative Filtering
- UBCF – User-Based Collaborative Filtering

Online Click-through Rate

Comparison methods:
- HCACR – Our method
- Manual strategy
Context based Recommendation

智能的测量方法

遇到疑问，小木来帮忙！点击下方知识点，查看解答

智力  记忆  问题解决

思维  问题解决  可用的启发式
More Analysis

Distribution by age

Probability vs. Age

Distribution by age

Probability vs. Age
• Let start the simplest case  
  – Course recommendation based on user interest

• What can we else?  
  – Interaction when watching video?
Smart Jump
—Automated suggestion for video navigation

Let's begin with …
The example is that …
First, we introduce …
Next … capital assets … investment property …
Average Jump

On Average: 2.6 Clicks = 5 seconds
According to what we have discussed we find that the fifth activity belongs to cash outflow of a business activity.

On Average: 2.6 Clicks = 5 seconds

$5S \times 8,000,000 \text{ users} = 1.3 \text{ years}$
Science courses contain much more frequent jump-backs than non-science courses.

Users in non-science courses jump back earlier than users in science courses.

Users in science courses are likely to rewind farther than users in non-science courses.
Observations – User Related

- 6.6% users prefer 10 seconds
- 9.2% users prefer 17 seconds
- 6.6% users prefer 20 seconds
In the next ninth economic activity
The enterprise has paid 4,000,000 yuan

What is the money used for

Of which 2,500,000 yuan is paid for the expenditure of sales department
1,500,000 for the expenditure of administrative department

$\text{argmax} \Delta t \frac{R_{e_{cj}}}{R_{e_{cj}}} \cdot \frac{R_{n_{s}}}{R_{n_{s}}}$

- $R_{e_{cj}}$: rate of effective complete-jumps (start position and end position located in different segments).
- $R_{n_{s}}$: rate of non-empty segments (contains at least one start position or end position of some complete-jumps).
Problem Formulation

\[ \text{argmax } P(s_j \mid u, v, s_i; \Theta) \]

### Prediction Results

<table>
<thead>
<tr>
<th>Course</th>
<th>Model</th>
<th>AUC</th>
<th>P@1</th>
<th>P@3</th>
<th>P@5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Science</strong></td>
<td>LRC</td>
<td>72.46</td>
<td>35.95</td>
<td>65.54</td>
<td>80.13</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>71.92</td>
<td>35.45</td>
<td>66.15</td>
<td>81.99</td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>74.02</td>
<td>37.61</td>
<td><strong>76.04</strong></td>
<td><strong>89.59</strong></td>
</tr>
<tr>
<td><strong>Non-science</strong></td>
<td>LRC</td>
<td>72.59</td>
<td>69.23</td>
<td>73.23</td>
<td>89.32</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>73.52</td>
<td>68.39</td>
<td>76.64</td>
<td>91.30</td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>73.57</td>
<td>67.56</td>
<td><strong>88.43</strong></td>
<td><strong>96.05</strong></td>
</tr>
</tbody>
</table>

- LRC, SVM, and FM are different models
- FM is defined as follows

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

<table>
<thead>
<tr>
<th>类别</th>
<th>统计量</th>
<th>7.15-8.15</th>
<th>8.16-10.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>用户数量</td>
<td>总共用户数量</td>
<td>14875</td>
<td>20043</td>
</tr>
<tr>
<td>受触发回看事件的用户数量</td>
<td>781</td>
<td></td>
<td>1025</td>
</tr>
<tr>
<td>视频数量</td>
<td>总共视频数量</td>
<td>235</td>
<td>235</td>
</tr>
<tr>
<td>受触发回看事件的视频数量</td>
<td>234</td>
<td></td>
<td>235</td>
</tr>
<tr>
<td>总的回看次数</td>
<td>7772</td>
<td></td>
<td>10369</td>
</tr>
<tr>
<td>回看路径不包含推荐点的回看</td>
<td>回看次数</td>
<td>3809</td>
<td>5325</td>
</tr>
<tr>
<td>平均回跳次数</td>
<td>1.657653</td>
<td></td>
<td>1.722441</td>
</tr>
<tr>
<td>回看路径包含但未点击推荐点的回看</td>
<td>回看次数</td>
<td>3408</td>
<td>4333</td>
</tr>
<tr>
<td>平均回跳次数</td>
<td>1.784918</td>
<td></td>
<td>1.803831</td>
</tr>
<tr>
<td>点击推荐点开始看视频的回看</td>
<td>回看次数</td>
<td>196</td>
<td>297</td>
</tr>
<tr>
<td>平均回跳次数</td>
<td>1.882653</td>
<td></td>
<td>1.845118</td>
</tr>
<tr>
<td>点击推荐点后继续跳转的回看</td>
<td>回看次数</td>
<td>359</td>
<td>414</td>
</tr>
<tr>
<td>平均回跳次数</td>
<td>2.788301</td>
<td></td>
<td>3.135266</td>
</tr>
</tbody>
</table>
Data statistics

效果好的统计量：
点击推荐点后开始看视频的回看比例有所上升：35.3% -> 41.7%
点击推荐点后开始看视频的回看的平均回跳次数：1.882653 -> 1.845118

效果不好的统计量：
回看路径不包含推荐点的回看
回看路径包含但未点击推荐点的回看
点击推荐点后继续跳转的回看
More

- Let start the simplest case
  - Course recommendation based on user interest
- What can we else?
  - Interaction when watching video?
  - What kind of questions did the users ask?
Question Answering

User Query

Question Classification

Platform FAQ  Wikipedia  Forum Archive  Service  Others

Question Answer Assembling
Query Categories

- **PLATFORM**: XuetangX platform
- **CONTENT**: enrollments, courses, teachers
- **CONCEPT**: simple knowledge point
- **DISCUSS**: general discussion, comparison
- **FEEDBACK**: suggestions, complains
- **SMALLCHAT**: small chat
- **CUSTOMER**: personal questions (e.g., account)
- **MISC**: meaningless questions (e.g., asjedkjqw)
- **SERVICE**: poem, recommendation
Category Distribution

- SERVICE
- MISC
- PERSONAL
- SMALLCHAT
- FEEDBACK
- DISCUSS
- CONCEPT
- CONTENT
- PLATFORM
Candidate Dataset

- Wikipedia: 892,185
- Forum Archive: 65,001
- Platform FAQ: 137
- Zhihu: 1,000+
- CSDN: 670
- Course Structure: 8 types
**Question Classification**

- **#Training (March 2017 – August 2017):** 2162
- **#Test (September 2017):** 499
  
  Precision: 0.77, Recall: 0.78
### Online Result

<table>
<thead>
<tr>
<th></th>
<th>#Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total_request</td>
<td>20604</td>
</tr>
<tr>
<td>feedback</td>
<td>470</td>
</tr>
<tr>
<td>Feedback_ratio</td>
<td>0.023</td>
</tr>
<tr>
<td>User-thumb_up</td>
<td>245</td>
</tr>
<tr>
<td>User-thumb_down</td>
<td>225</td>
</tr>
<tr>
<td>Thumb_ratio</td>
<td>0.52</td>
</tr>
</tbody>
</table>
Question Retrieval

• Queries in PLATFORM category: 538
• Q-A pairs in Candidate Set: 77

<table>
<thead>
<tr>
<th>Method</th>
<th>MRR</th>
<th>Hit @ 1</th>
<th>Hit @ 3</th>
<th>Hit @ 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES (TF-IDF)</td>
<td>0.617</td>
<td>0.558</td>
<td>0.698</td>
<td>0.748</td>
</tr>
<tr>
<td>Word2vec + WMD</td>
<td>0.695</td>
<td>0.602</td>
<td>0.745</td>
<td>0.817</td>
</tr>
<tr>
<td>Word2vec + Cosine</td>
<td>0.653</td>
<td>0.577</td>
<td>0.685</td>
<td>0.726</td>
</tr>
<tr>
<td>1.0<em>WMD+1.5</em>ES</td>
<td>0.728</td>
<td>0.640</td>
<td>0.781</td>
<td>0.845</td>
</tr>
</tbody>
</table>
• Let start the simplest case
  – Course recommendation based on user interest
• What can we else?
  – Interaction when watching video?
  – What kind of questions did the users ask?
  – Interaction->intervention
Active Question

Question: What is time complexity?

- 常数 (constant function)

\[ 2 = 2013 = 2013 \times 2013 = O(1), \text{ 甚至 } 2013^{2013} = O(1) \]
Active Question

Question: What is Random Vector?
Bot->Mindsets

• are those interventions really useful?
  – not enough…
Active Question with Social Pressure

Example: Thumb_up Class (with #thumbup)
Active Question

On-line experiment Setting:

<table>
<thead>
<tr>
<th>Time</th>
<th>Classified Type</th>
<th>Total user count</th>
<th>User Count per Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/14 – 9/17</td>
<td>On/Off</td>
<td>266</td>
<td>On</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>137</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>129</td>
</tr>
<tr>
<td>9/23 – 9/30</td>
<td>Social/Thumb_up/None</td>
<td>1150</td>
<td>Social</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Thumb_up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>365</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>414</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>371</td>
</tr>
</tbody>
</table>

1. Each question lasts for 10 seconds;
2. Displaying time is notated manually to ensure strong connection with the on-going content;
## Active Question

### Positive Direct Feedback:

<table>
<thead>
<tr>
<th>Time</th>
<th>Classified Type</th>
<th>Feedback ratio(at least once)</th>
<th>Thumb_up Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0914 -- 0917</td>
<td>On/Off</td>
<td>12.4%(17/134)</td>
<td>31.2%(10/32)</td>
</tr>
<tr>
<td>0923 -- 0930</td>
<td>Social/Thumb_up/None</td>
<td>17.5%(151/864)</td>
<td>47.1%(113/240)</td>
</tr>
</tbody>
</table>

1. Each question lasts for 10 seconds;
2. Appearing time is notated manually to ensure strong connection with the on-going content;
Active Question

New Peaks in in-video interaction:

Vertical line:
- Red: start of question
- Green: end of question

Curve:
- Yellow: without question displaying
- Blue: with question displaying

(Since the course is on-going, a full comparison is not available for now)
Active Question

A specific case of jumping back to the question time

X-axis: video time axis
Y-axis: event time axis

Bottom blue line:
• Red: start of question
• Green: end of question

Other lines:
• User’s jump span

Dots:
• Other event, e.g., playing, pausing.
Active Question

Longer Video Watching Time in total:

<table>
<thead>
<tr>
<th>Class</th>
<th>Median Watch Time(second)</th>
<th>Average Watch Time(second)</th>
<th>User Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>On</td>
<td>1329.5</td>
<td>3497.4</td>
<td>137</td>
</tr>
<tr>
<td>Off</td>
<td>1864.0</td>
<td>2946.3</td>
<td>129</td>
</tr>
</tbody>
</table>

(t-test, p=0.303)
Active Question

The fixed strategy has some major shortcomings:
1. It does not scale up well;
2. User difference is not considered;
3. The displaying time and duration is chosen intuitively and far from optimal.

Reinforcement learning may help.
1. Using users’ history for personalization;
2. Iteratively update the strategy by users’ feedback;
   • Careful design needed to integrate both explicit feedback (thumb_up or exit button) and implicit feedback (watching time, etc.);
3. Experiment is still on the way.
LittleMU (小木)

LittleMU: Intelligent Interaction

1. User analysis
   - Behavior modeling
   - User Profiling

2. Course analysis
   - Incentive analysis
   - Course recommendation
   - Automated video navigation
   - Question answering

3. Concept extraction
   - Course Content
   - Prerequisite relation mining

User Modeling | Intervention | Content Analysis
Recent Publications

- Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang. Prerequisite Relation Learning for Concepts in MOOCs. In ACL'17.
- Xia Jing, Jie Tang, Wenguang Chen, Maosong Sun, and Zhengyang Song. Guess You Like: Course Recommendation in MOOCs. WI'17.
- Jie Tang, Tracy Xiao Liu, Zhenyang Song, Xiaochen Wang, Xia Jing, Jiezhong Qiu, Zhenhuan Chen, Chaoyang Li, Han Zhang, Liangmin Pan, Yi Qi, Xiuli Li, Jian Guan, Juanzi Li, and Maosong Sun. LittleMU: Enhancing Learning Engagement Using Intelligent Interaction on MOOCs. submitted to KDD.
- 薛宇飞, 敬峡, 裘捷中, 唐杰, 孙茂松. 一种在线课程中的作业互评方法：中国，201510531490.2. （中国专利申请号）
- 唐杰, 张茜, 刘德兵. 用户退课行为预测方法及装置. 201610292389.0 （中国专利申请号）
Thank you!

Collaborators: Jian Guan, Xiuli Li, Fenghua Nie (XuetangX)
Jie Gong (NUS), Jimeng Sun (GIT)
Wendy Hall (Southampton)
Maosong Sun, Tracy Liu, Juanzi Li (THU)
Xia Jing, Zhenhuan Chen, Liangmin Pan, Jiezhong Qiu, Han Zhang,
Zhengyang Song, Xiaochen Wang, Chaoyang Li, Yi Qi (THU)

Jie Tang, KEG, Tsinghua U,
Download all data & Codes,
http://keg.cs.tsinghua.edu.cn/jietang
http://arnetminer.org/data
http://arnetminer.org/data-sna