Can AI/DM help MOOCs?

Jie Tang
Computer Science
Tsinghua University

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

- **Coursera**
  - 149 partners
  - 2,000+ courses
  - 24,000,000 users

- **edX**
  - 110 partners
  - 1,270 courses
  - 10,000,000 users
  - 10+ MicroMaster

- **XuetangX.com**
  - 1,000+ courses
  - 8,000,000 users
  - Chinese EDU association

- **Udacity**
  - ~10 partners
  - 40+ courses
  - 1.6 million users
  - “nanodegree”

- **网易公开课**
  - host >1,000 courses
  - millions of users
launched in
2013
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/DM 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?

artificial intelligence

data mining

machine learning

data clustering

association rule

Probability Distribution

Maximum Likelihood

Hidden Markov Model
However to improve the engagement?

User

Knowledge

- artificial intelligence
- machine learning
- association rule
- data clustering
- data mining
LittleMU (小木)
LittleMU (小木)
LittleMU (小木)

LittleMU: Intelligent Interaction

1. User analysis
   - Behavior modeling
   - User Profiling

2. Course analysis
   - Concept extraction
   - Prerequisite relation mining

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

Behavior logs → User Modeling → Intervention → Content Analysis → Knowledge base
LittleMU (小木)

LittleMU: Intelligent Interaction

1. User Profiling
   - Behavior modeling
   - User modeling

2. Course analysis
   - Course Content
   - Concept extraction
   - Prerequisite relation mining

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

Behavior logs

Knowledge base
MOOC user

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

Non-Science

Science

Female

Bachelor

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

\[
Y^t(i) = f(W_o Z^t(i) + b_o)
\]
\[
Z^t(i) = f(W_s S^t(i) + b_s)
\]
\[
S^t(i) = \left[ Z_{t-1}^T, X^t(i)^T \right]^T
\]

**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

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

**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.

• …
LittleMU (小木)

LittleMU: Intelligent Interaction

1. User analysis
   - Behavior modeling
   - User Profiling

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

3. Incentive analysis

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

data mining
data clustering
application
business intelligence
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

<table>
<thead>
<tr>
<th>Concept</th>
<th>Vector Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>data mining</td>
<td>0.8 0.2 0.3 ... 0.0 0.0</td>
</tr>
<tr>
<td>business intelligence</td>
<td>0.1 0.1 0.2 ... 0.8 0.7</td>
</tr>
</tbody>
</table>

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>$P$</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>$F_1$</td>
<td>65.8</td>
<td>65.7</td>
</tr>
<tr>
<td>NB</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>$F_1$</td>
<td>58.1</td>
<td>59.4</td>
</tr>
<tr>
<td>LR</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>$F_1$</td>
<td>63.7</td>
<td>64.2</td>
</tr>
<tr>
<td>RF</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>$F_1$</td>
<td>69.1</td>
<td>72.6</td>
</tr>
</tbody>
</table>

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

- SVM, NB, LR, and RF are different classification models
- It seems that with the defined distance functions, RF achieves the best

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

LittleMU: Intelligent Interaction

1. Behavior modeling
   - User Profiling
   - User analysis

2. Course recommendation
   - Automated video navigation
   - Question answering

3. Incentive analysis
   - Course analysis
     - Course Content
     - Concept extraction
     - Prerequisite relation mining

Behavior logs

User Modeling

Intervention

Content Analysis

Knowledge base
What we can do?

User modeling

Knowledge

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

Course topic analysis

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

High frequency
- Latent interest modeling
- Collaborative filtering

Recommendation result

Rule based adjustment

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

<table>
<thead>
<tr>
<th>Course Name</th>
<th>Duration</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>公司金融学</td>
<td>7 天前开课</td>
<td>422人</td>
</tr>
<tr>
<td>管理会计学</td>
<td>5 天前开课</td>
<td>328人</td>
</tr>
<tr>
<td>大学计算机教程</td>
<td>9 个月前开课</td>
<td>14267人</td>
</tr>
<tr>
<td>IC设计与方法</td>
<td>3 个月前开课</td>
<td>818人</td>
</tr>
<tr>
<td>托福考试准备：来自考试举办方的指导</td>
<td>3 个月前开课</td>
<td>818人</td>
</tr>
<tr>
<td>水力学</td>
<td>9 个月前开课</td>
<td>2349人</td>
</tr>
<tr>
<td>孝亲之礼</td>
<td>9 个月前开课</td>
<td>499人</td>
</tr>
<tr>
<td>陆游词鉴赏</td>
<td>8 个月前开课</td>
<td>850人</td>
</tr>
<tr>
<td>贞观之治</td>
<td>4 个月前开课</td>
<td>214人</td>
</tr>
<tr>
<td>IELTS雅思考试备考</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Course Recommendation:
Guess you like

<table>
<thead>
<tr>
<th>Course Name</th>
<th>Duration</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>决胜移动互联网：创业者商业模式课（2017春）</td>
<td>3 个月前开课</td>
<td>3083人</td>
</tr>
<tr>
<td>u.lab 0x：基于觉察的系统创变：感知和共创未来...</td>
<td>8 个月前开课</td>
<td>5132人</td>
</tr>
<tr>
<td>金融工程导论（2017春）</td>
<td>3 个月前开课</td>
<td>1492人</td>
</tr>
<tr>
<td>分布式计算与数据管理（慕课）</td>
<td>5 个月前开课</td>
<td>1099人</td>
</tr>
<tr>
<td>现代生活美学（2017春）</td>
<td>3 个月前开课</td>
<td>2907人</td>
</tr>
</tbody>
</table>
Online A/B Test

Performance Comparison

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 CTR Comparison

Online Click-through Rate
Comparison methods:
HCACR – Our method
Manual strategy
More Analysis

Distribution by age

Distribution by 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

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

On Average: 2.6 Clicks = 5 seconds
Two Numbers

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.

\[ 5S \times 8,000,000 \text{ users} = 1.3 \text{ years} \]
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 the sales department, and 1,500,000 yuan for the expenditure of the administrative department.

- $R_{e_cj}$: 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).

\[
\arg\max_{\Delta t} \frac{R_{e_cj}}{R_{e_cj} + R_{n_s}}
\]
Problem Formulation

\[
\arg\max_{\Theta} P(s_j|u, v, s_i; \Theta)
\]

### Prediction Results

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

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

• Let start the simplest case
  – Course recommendation based on user interest
• What can we else?
  – Interaction when watching video?
  – Interaction->intervention
Question: What is time complexity?
Active Question

Question: What is Random Vector?
## Preliminary Results

<table>
<thead>
<tr>
<th></th>
<th>#Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total_request</td>
<td>30991</td>
</tr>
<tr>
<td>feedback</td>
<td>569</td>
</tr>
<tr>
<td>Feedback_ratio</td>
<td>0.0184</td>
</tr>
<tr>
<td>User-thumb_up</td>
<td>132</td>
</tr>
<tr>
<td>User-cancel</td>
<td>503</td>
</tr>
<tr>
<td>Thumb_ratio</td>
<td>0.24</td>
</tr>
</tbody>
</table>
LittleMU (小木)
Acrostic Poem: 小木作诗

学堂小木

Hi, DashChen, 我是智能助教小木，根据您目前的章节进度，献上藏头诗一首，看看藏的是什么词？

数声茅屋两三家
据石桥边日又斜
结客不来春已醉
构堂风雨一窗纱

学堂小木

Hi, DashChen, 我是智能助教小木，根据您目前的章节进度，献上藏头诗一首，看看藏的是什么词？

风雨萧萧两鬓秋
流光冉冉五湖游
天花满地无人扫
子弟携琴独上楼

学堂小木

Hi, DashChen, 我是智能助教小木，根据您目前的章节进度，献上藏头诗一首，看看藏的是什么词？

冒雨浮生又一年
泊icias惨白云边
排空行尽青山外
序齿童心亦可怜

学堂小木

Hi, DashChen, 我是智能助教小木，根据您目前的章节进度，献上藏头诗一首，看看藏的是什么词？

网罗不惜黄金缕
络绎何嫌白玉京
技能于今无一事
术疏元自有前生
LittleMU (小木)
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)
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