互联网教育智能技术及应用
国家工程实验室

互联网教育智能技术及应用国家工程实验室

北京師範大學 BEIJING NORMAL UNIVERSITY 清華大學 Tsinghua University 中国移动 China Mobile 网龙华渔教育 科大讯飞 iFLYTEK

Empower MOOCs with AI

Jie Tang

Tsinghua University

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

Big Data in MOOC

- **149** partners
- 2400+ courses
- **33,000,000** users

- **1,000+** courses
- **10,000,000** users
- Chinese EDU association

- **110** partners
- 1,800 courses
- **14,000,000** users
- 10+ MicroMaster

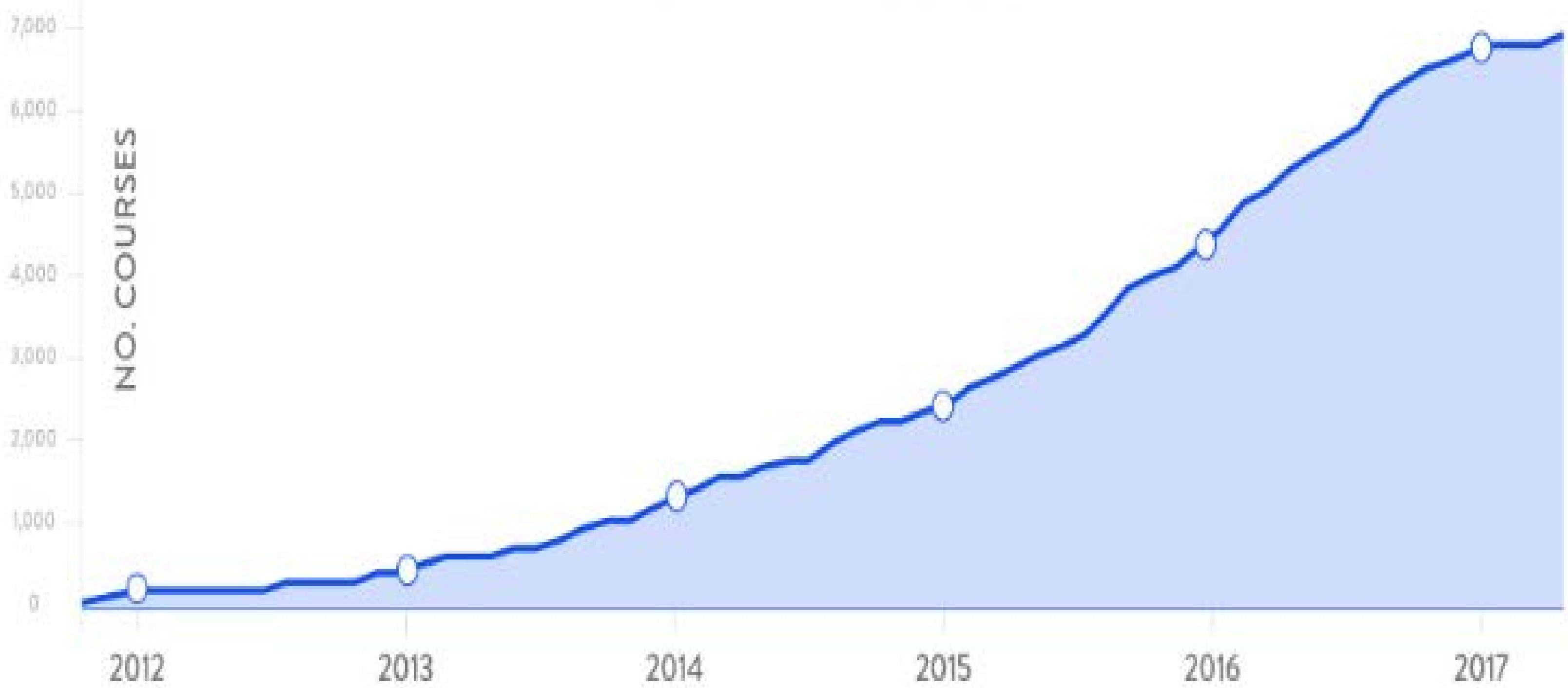
MOOC

- host >1,000 courses
- millions of users

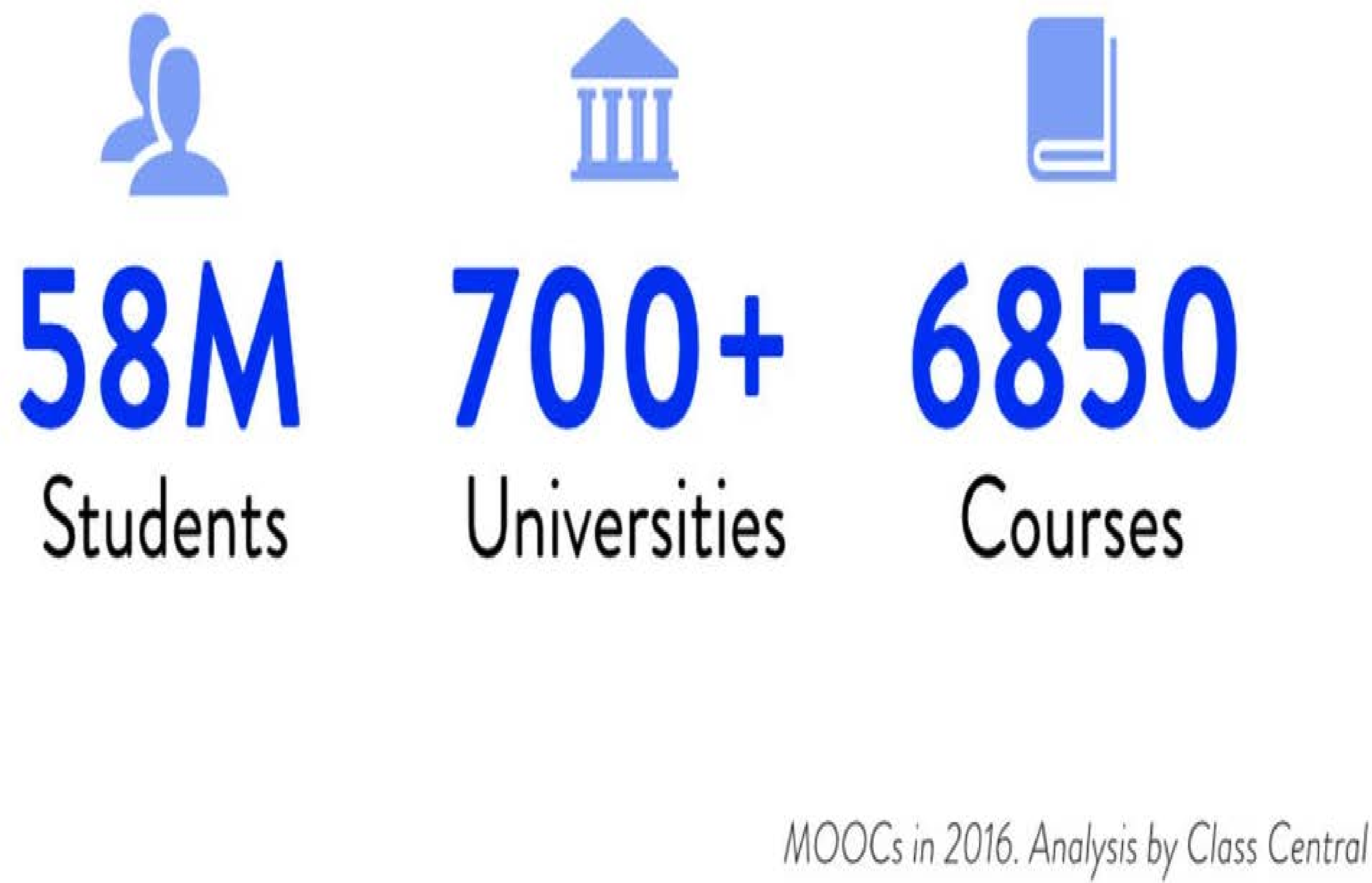
- **~10** partners
- 40+ courses
- **1.6 million** users
- **"nanodegree"**

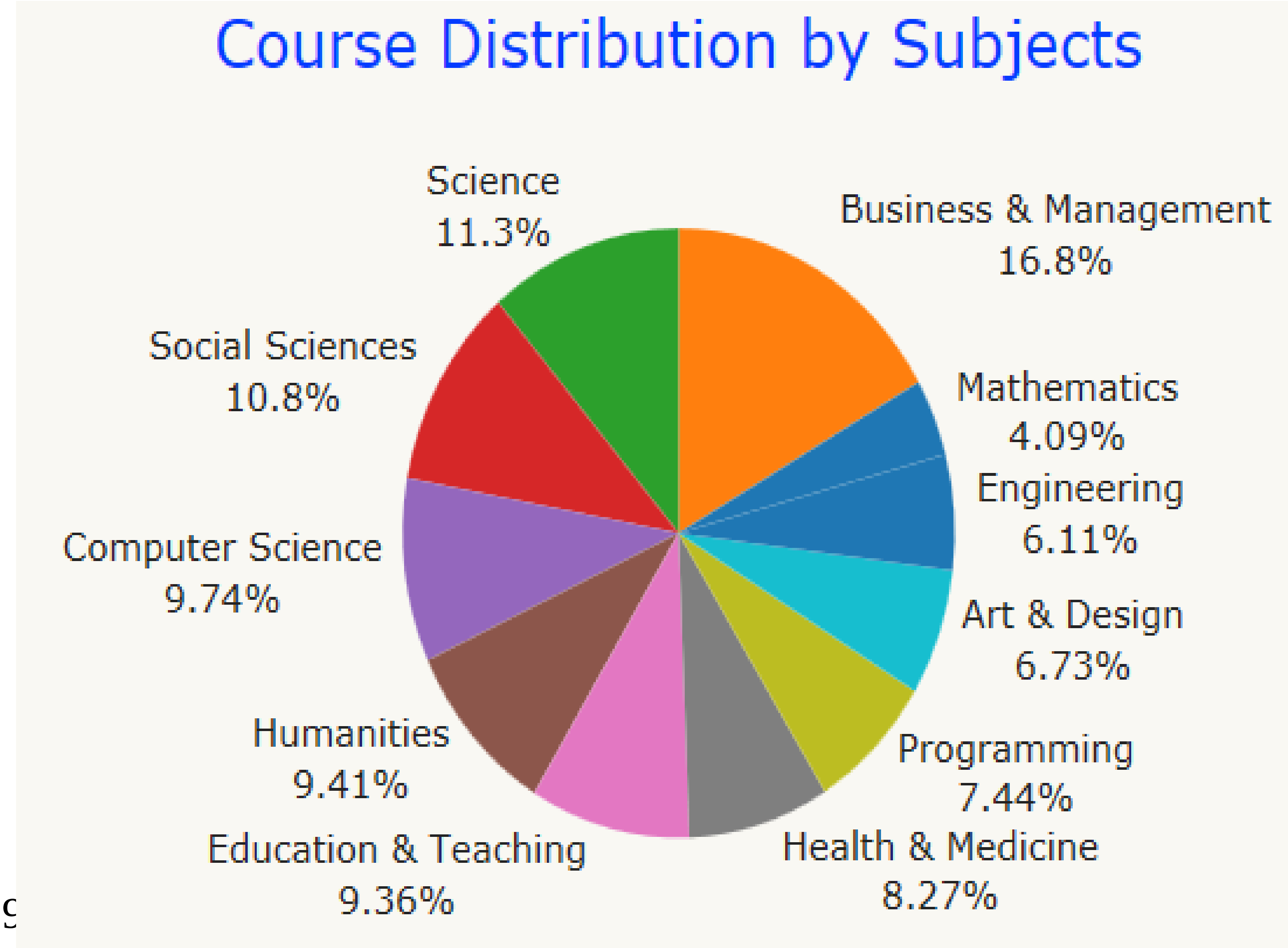
……

Growth of MOOCs



Source CLASS CENTRAL


58M
Students
700+
Universities
6850
Courses
MOOCs in 2016. Analysis by Class Central


Course Distribution by Subjects
Science
11.3%
Business & Management
16.8%
Social Sciences
10.8%
Mathematics
4.09%
Engineering
6.11%
Computer Science
9.74%
Art & Design
6.73%
Humanities
9.41%
Programming
7.44%
Education & Teaching
9.36%
Health & Medicine
8.27%

2019

Coursera


coursera
目录
搜索目录
合作伙伴
登录
注册
在网上学习全世界最好的课程。
免费加入
Penn
JOHNS HOPKINS UNIVERSITY
UNIVERSITY OF MICHIGAN
Stanford
UC San Diego

Neural Networks for Machine Learning

总览

授课大纲

制作方

评分和审阅

Neural Networks for Machine Learning

Starts 11月 28

助学金仅对无法承担费用的学生提供。
了解更多并申请。

关于此课程： Learn about artificial neural networks and how they're being used for machine learning, as applied to speech and object recognition, image segmentation, modeling language and human motion, etc. We'll emphasize both the basic algorithms and the practical tricks needed to get them to work well.

更多

制作方： 多伦多大学

UNIVERSITY OF TORONTO

教学方： Geoffrey Hinton, Professor
Department of Computer Science

Deep Learning in Python
via *DataCamp*

Stanford University
Machine Learning
via *Coursera* 5-7 hours a week , 11 weeks long

Goldsmiths, University of London
Machine Learning for Musicians and Artists
via *Kadenze* 7 weeks long

Google
Deep Learning
via *Udacity* 6 hours a week , 12 weeks long

University of Washington
Machine Learning Foundations: A Case Study Approach
via *Coursera* 6 weeks long

University of California, Berkeley
CS188.1x: Artificial Intelligence
via *edX* 12 weeks long

Johns Hopkins University
Practical Machine Learning
via *Coursera* 4-9 hours a week , 4 weeks long
Earn A Credential Part of the Data Science Specia…

Stanford University
Introduction to Artificial Intelligence
via *Udacity* 6 hours a week , 16 weeks long

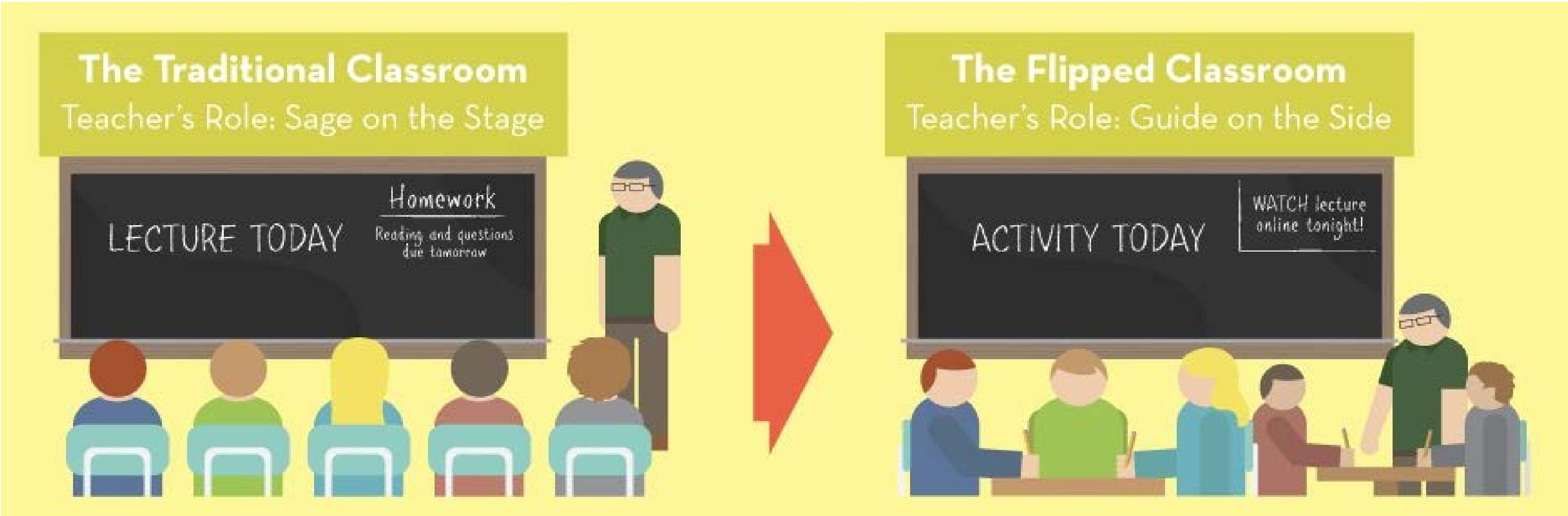
University of Toronto
Neural Networks for Machine Learning
via *Coursera* 7-9 hours a week , 16 weeks long

XuetangX

Launched in 2013

Some exciting data…

- Every day, there are 10,000+ 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…

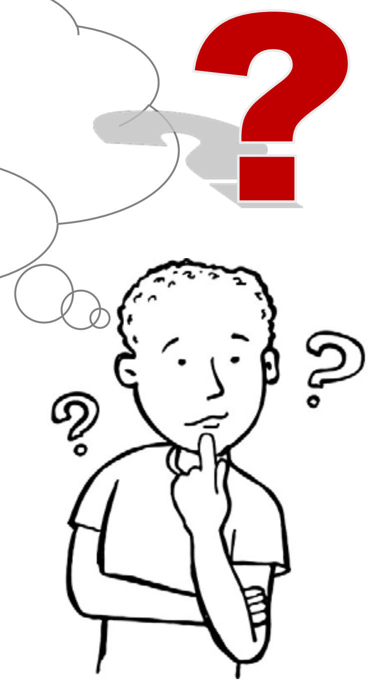
- Every day, there are 5,000+ new students
- An MOOC course can reach 100,000+ students
- >35% of the XuetangX users are using mobile
- traditional->flipped classroom->online degree
- "**Network+** EDU" (O2O)
  - edX launched 10+ MicroMaster degrees
  - Udacity launched NanoDegree program
  - GIT+Udacity 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?

科学问题1：MOOC user = Student?

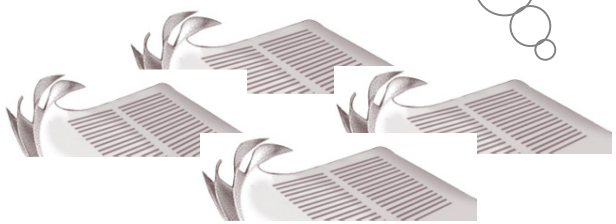
How to learn more effectively and more efficiently?



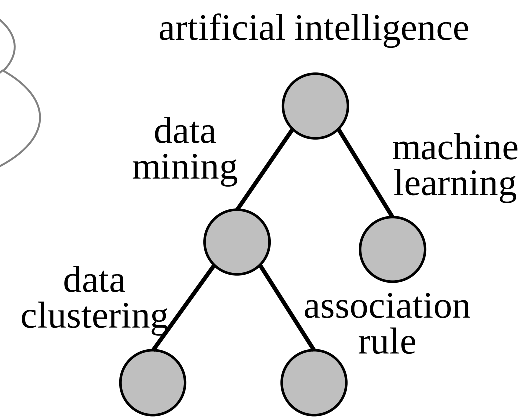
- **Who is who? background, where from?**
- **Why MOOC? motivation? degree?**
- **What is personalization? preference?**

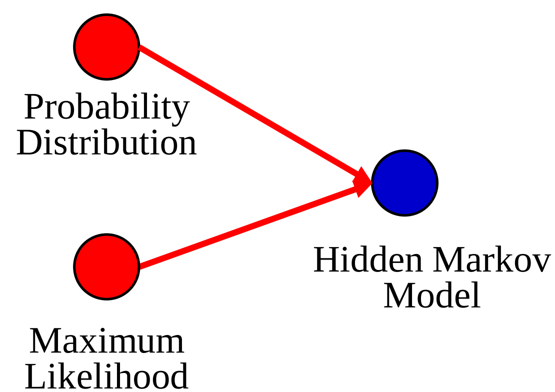
科学问题2：MOOC course = University course?

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



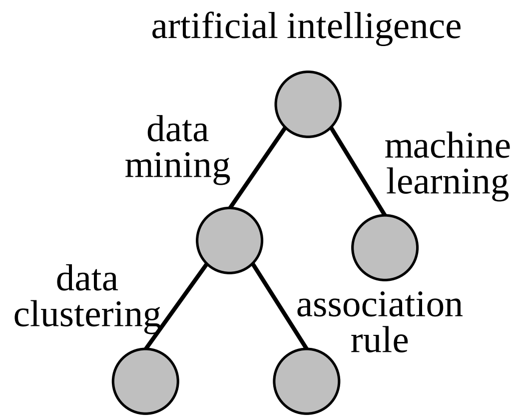
Thousands of Courses





How to leverage the external knowledge?

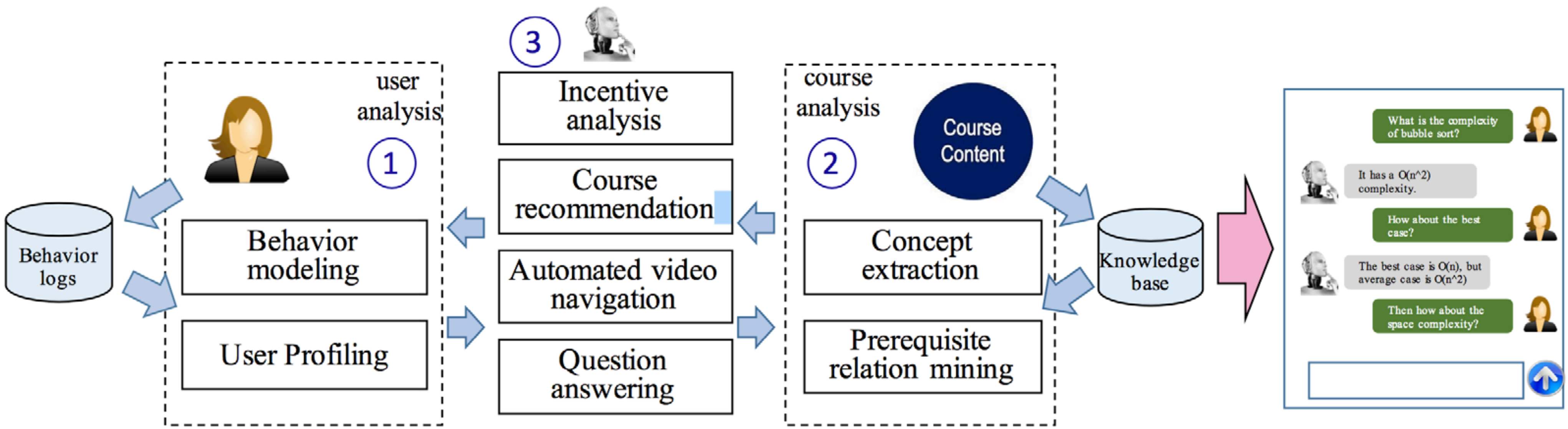
科学问题3：从交互到干预



User

Knowledge

AI驱动的在线教育


3
Incentive analysis
Course recommendation
Automated video navigation
Question answering
user analysis
1
Behavior modeling
User Profiling
Behavior logs
course analysis
2
Course Content
Concept extraction
Prerequisite relation mining
Knowledge base
What is the complexity of bubble sort?
It has a O(n^2) complexity.
How about the best case?
The best case is O(n), but average case is O(n^2)
Then how about the space complexity?

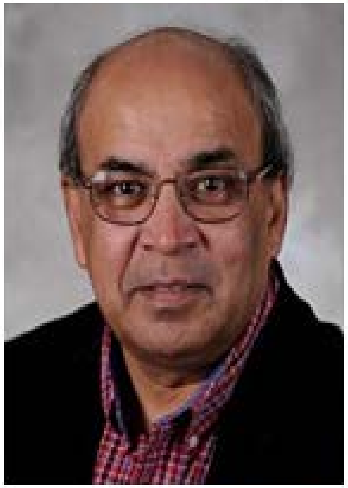
XiaoMU (“小木”)
—AI driven Learning Assistant

What is AI? Watson?

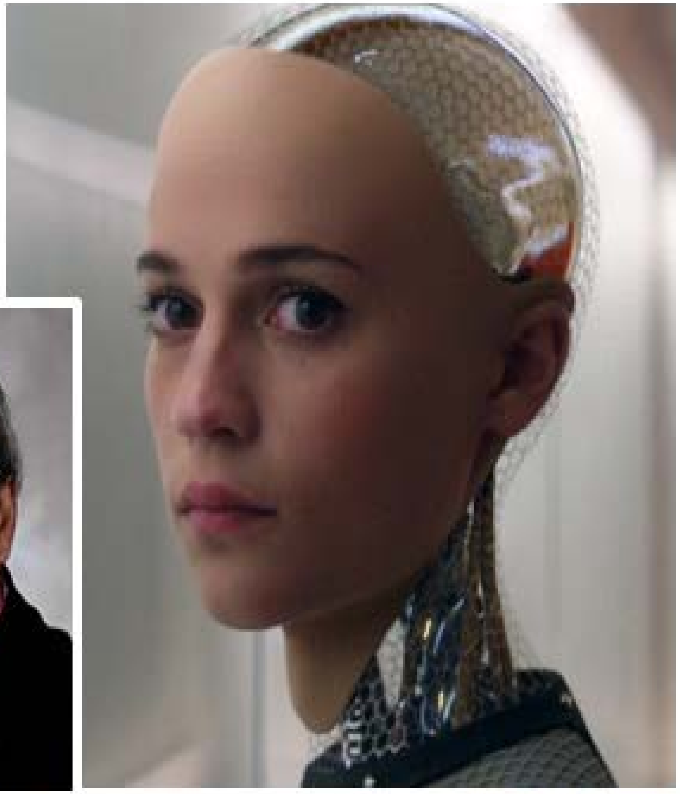
Georgia Tech | College of Computing

Jill Watson: Our Newest TA

- Creation of Prof. Ashok Goel
- TA for CS 7637: Knowledge-Based Artificial Intelligence
- Based on IBM Watson platform
- Anticipate that Jill will be able to answer 40% of ~10,000 questions posted to online forum



Ashok Goel

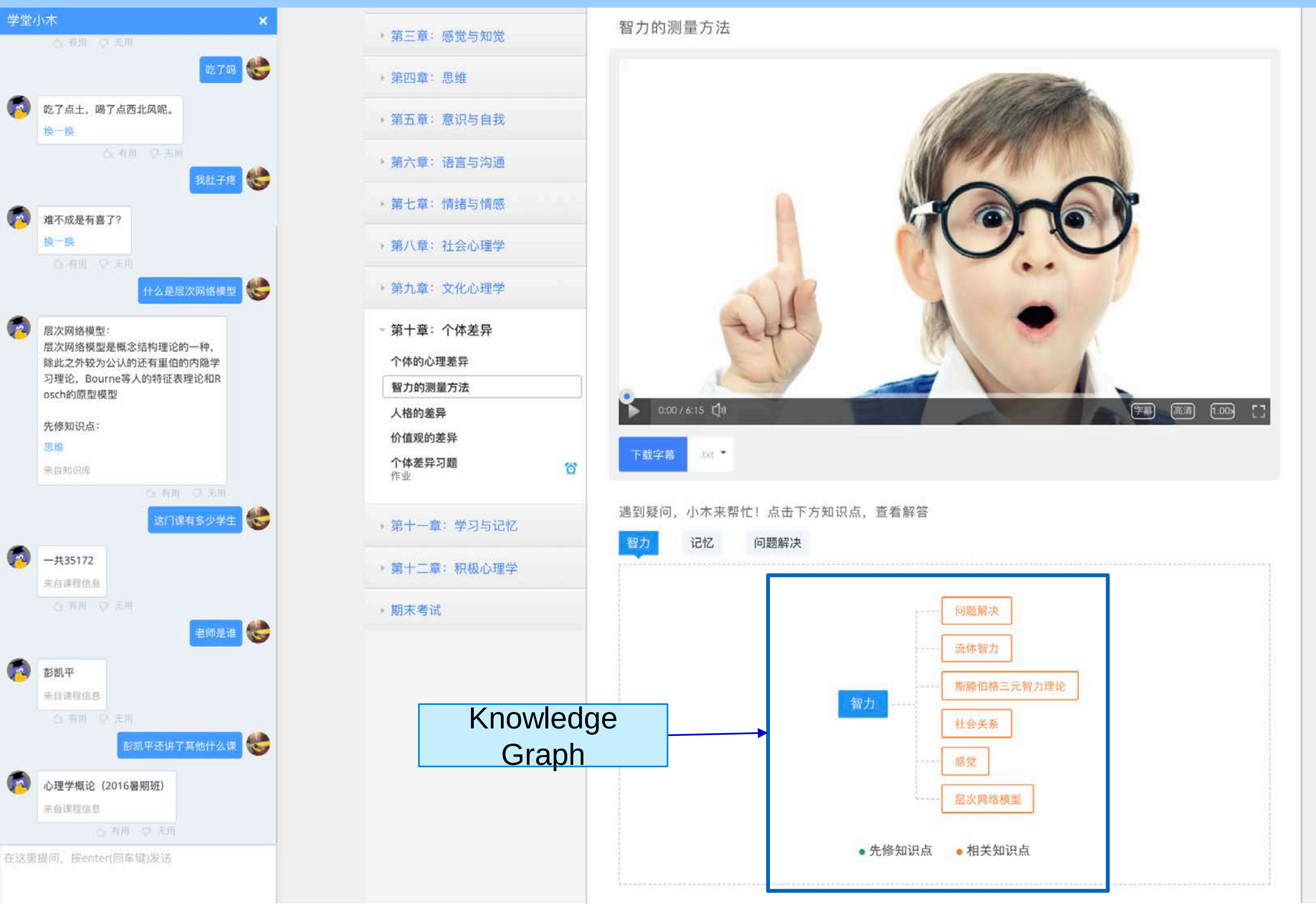


From the 2015 film, *Ex Machina*

What is XiaoMU ("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

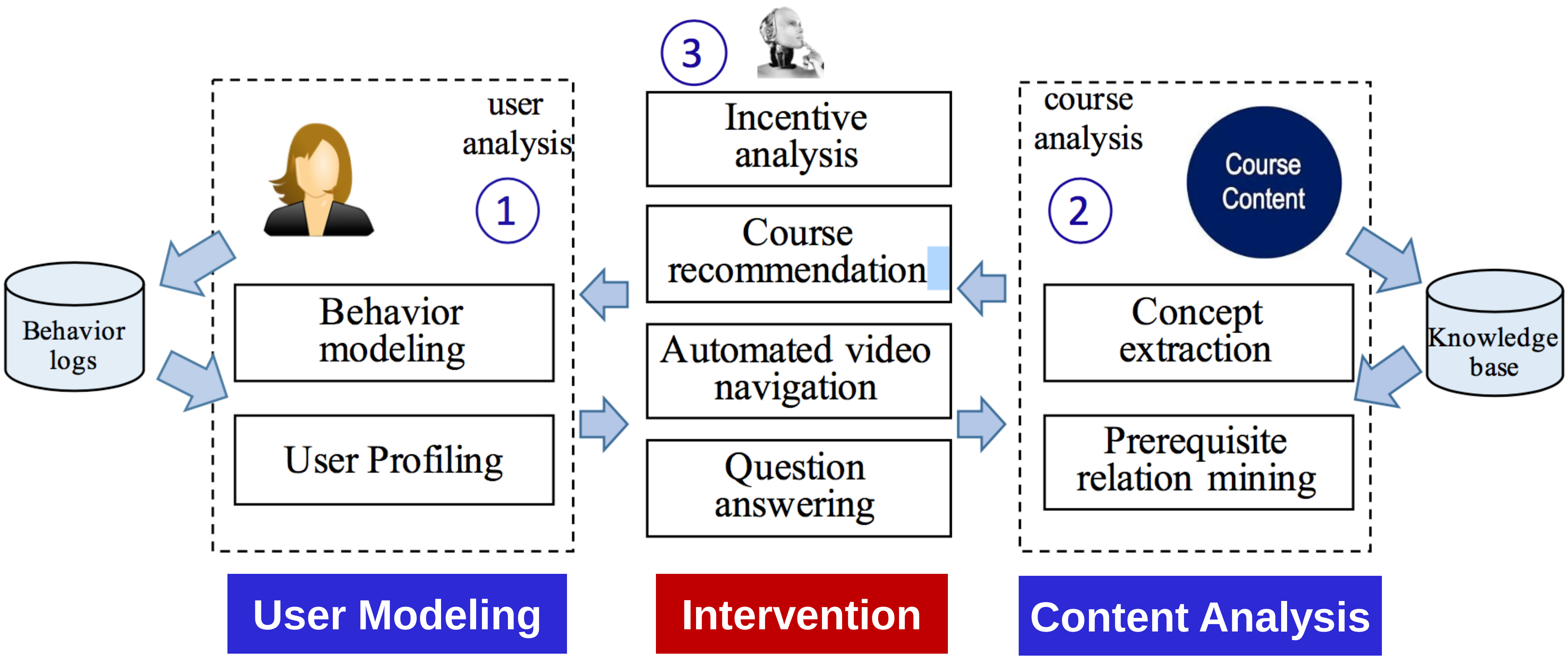
- Instead,"小木" is more like a learning peer
  - Tell you some basic knowledge in her mind
  - Tell you what the other users are thinking/learning
  - Try to understand your intention
  - Teach "小木" what you know


学堂小木
有用 无用
吃了吗
吃了点土，喝了点西北风呢。
换一换
我肚子疼
难不成是有喜了？
换一换
什么是层次网络模型
层次网络模型：
层次网络模型是概念结构理论的一种，除此之外较为公认的还有里伯的内隐学习理论，Bourne等人的特征表理论和Rosch的原型模型
先修知识点：
思维
来自知识库
这门课有多少学生
一共35172
来自课程信息
老师是谁
彭凯平
来自课程信息
彭凯平还讲了其他什么课
心理学概论（2016暑期班）
来自课程信息
在这里提问，按enter(回车键)发送
第三章：感觉与知觉
第四章：思维
第五章：意识与自我
第六章：语言与沟通
第七章：情绪与情感
第八章：社会心理学
第九章：文化心理学
第十章：个体差异
个体的心理差异
智力的测量方法
人格的差异
价值观的差异
个体差异习题
作业
第十一章：学习与记忆
第十二章：积极心理学
期末考试
智力的测量方法
0:00 / 6:15
字幕
高清
1.00x
下载字幕
.txt
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
Knowledge Graph
智力
问题解决
流体智力
斯腾伯格三元智力理论
社会关系
感觉
层次网络模型
先修知识点
相关知识点

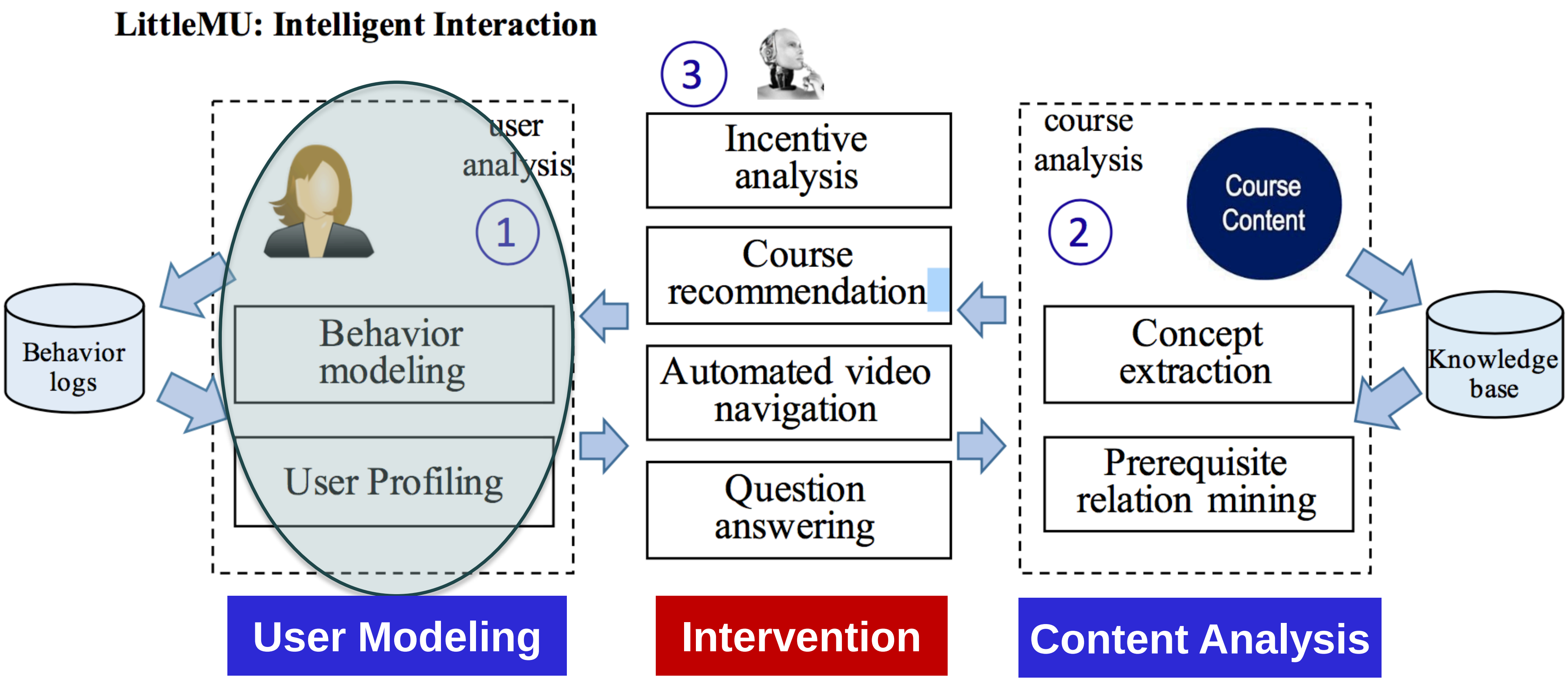
Acrostic Poem: 小木作诗—by 九歌

XiaoMU (小木)

But most existing systems focus on passively interactions…



XiaoMU (小木)

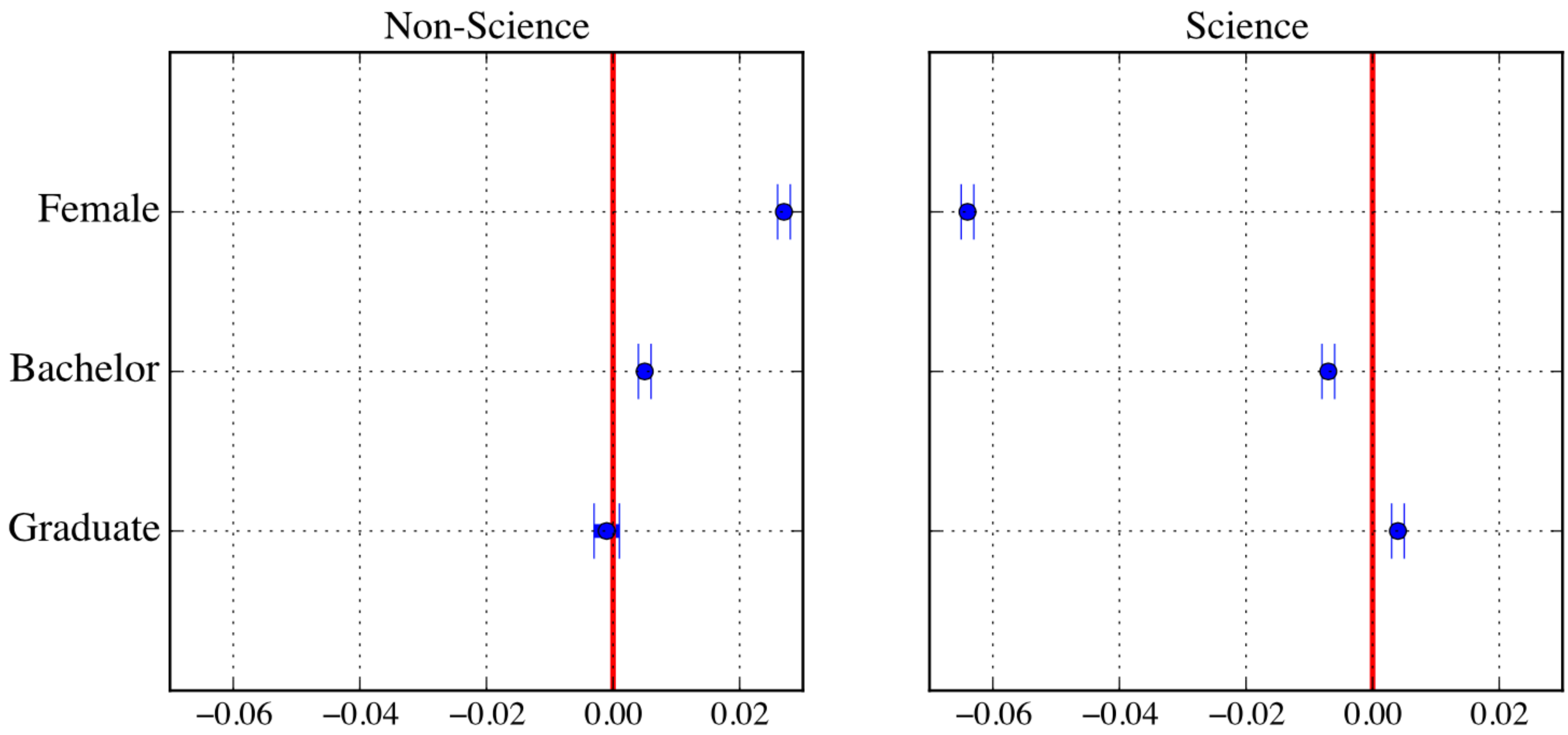


LittleMU: Intelligent Interaction
3
user analysis
1
Behavior logs
Behavior modeling
User Profiling
Incentive analysis
Course recommendation
Automated video navigation
Question answering
course analysis
2
Course Content
Concept extraction
Prerequisite relation mining
Knowledge base
User Modeling
Intervention
Content Analysis

MOOC user

- **Who is who? background, where from?**
- **Why MOOC? motivation? degree?**
- **What is personalization? preference?**

Basic Analysis



Observation 1 – Gender Difference

Table 4: Regression Analysis for Certificate Rate: All Users

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

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

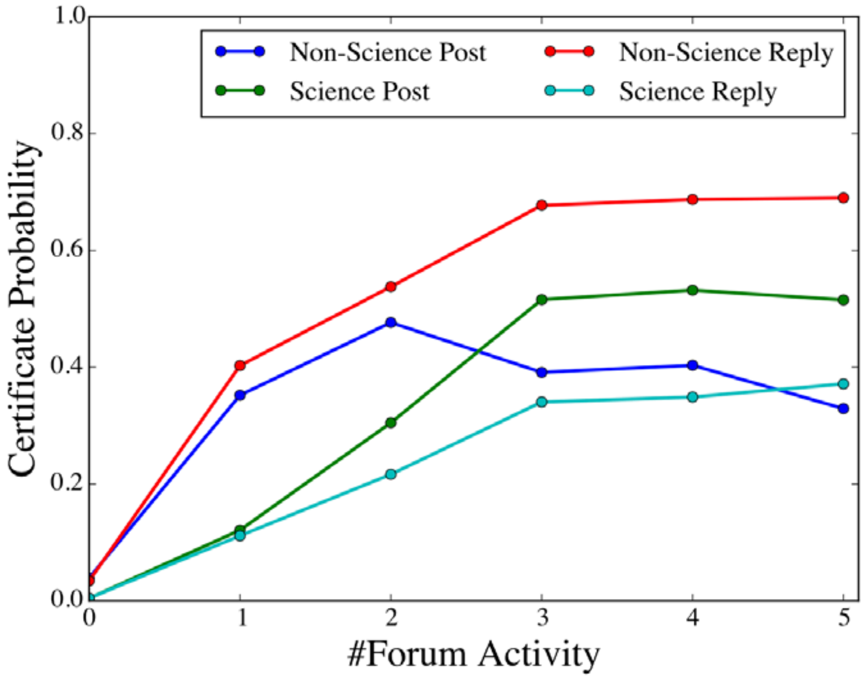
	Model 1		Model 2	
	Non-Science (1)	Science (2)	Non-Science (3)	Science (4)
Female	0.014*** (0.002)	-0.003 (0.002)	0.002* (0.001)	0.001 (0.002)
New Post	—	—	0.004*** (0.001)	0.038*** (0.008)
Reply	—	—	0.004** (0.002)	0.001* (0.001)
Video	—	—	0.000*** (0.000)	-0.000 (0.000)
Assignment	—	—	0.003*** (0.000)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

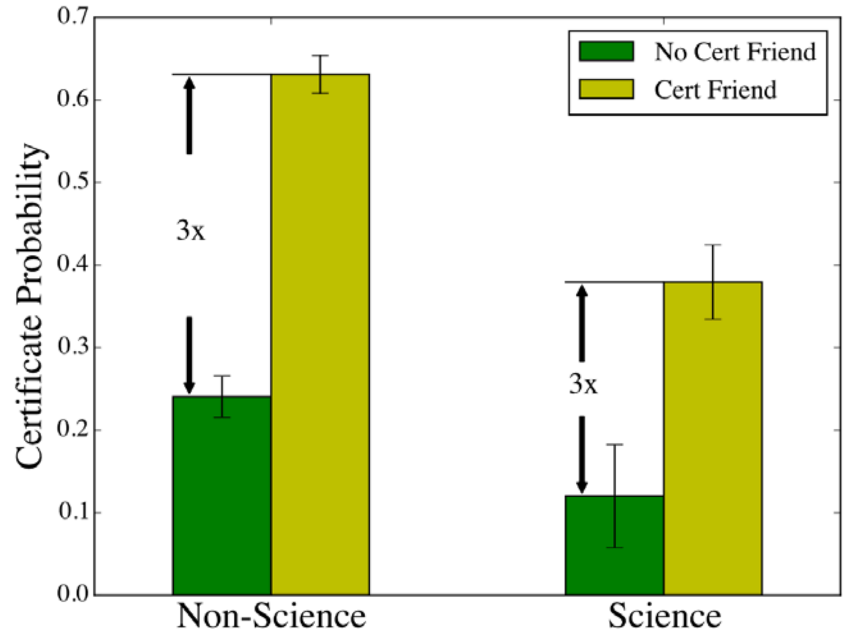
Model 1: Demographics vs Certificate

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





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)

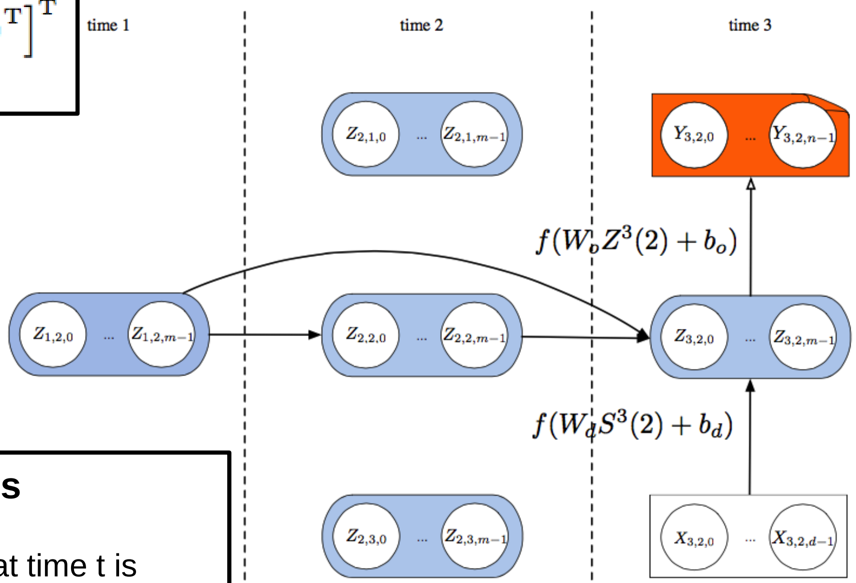
Dynamic Factor Graph Model

Model: incorporating "embedding" and factor graphs

$$Y^t(i)^* = f(W_o Z^t(i) + b_o)$$
$$Z^t(i)^* = f(W_d S^t(i) + b_d)$$
$$S^t(i) = \left[\mathbf{Z}_{t-p}^{t-1}(i)^{\mathrm{T}}, X^t(i)^{\mathrm{T}}\right]^{\mathrm{T}}$$

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

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



Latent learning states

Every student's status in at time t is associated with a vector representation

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

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

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

[1] Jiezhong Qiu, Jie Tang, Tracy Xiao Liu, Jie Gong, Chenhui Zhang, Qian Zhang, and Yufei Xue. Modeling and Predicting Learning Behavior in MOOCs. **WSDM'16**, pages 93-102.

Certificate Prediction

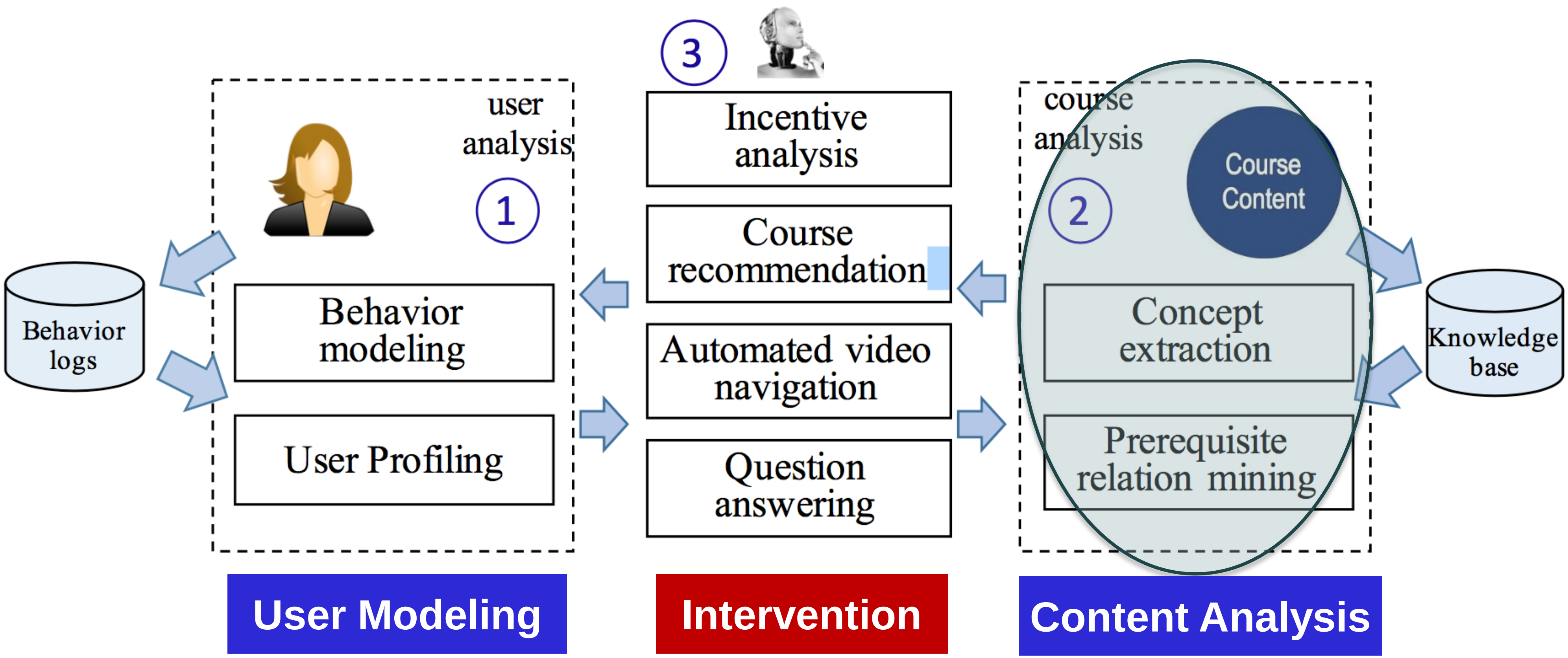
Category	Method	AUC	Precision	Recall	F1-score
Science	LRC	92.13	**83.33**	46.51	59.70
	SVM	92.67	52.17	83.72	64.29
	FM	94.48	61.54	74.42	67.37
	LadFG	**95.73**	73.91	**79.07**	**76.40**
Non-Science	LRC	94.16	76.93	89.20	82.57
	SVM	93.94	76.96	88.60	82.37
	FM	94.87	**80.22**	86.23	83.07
	LadFG	**95.54**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

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

Predicting more

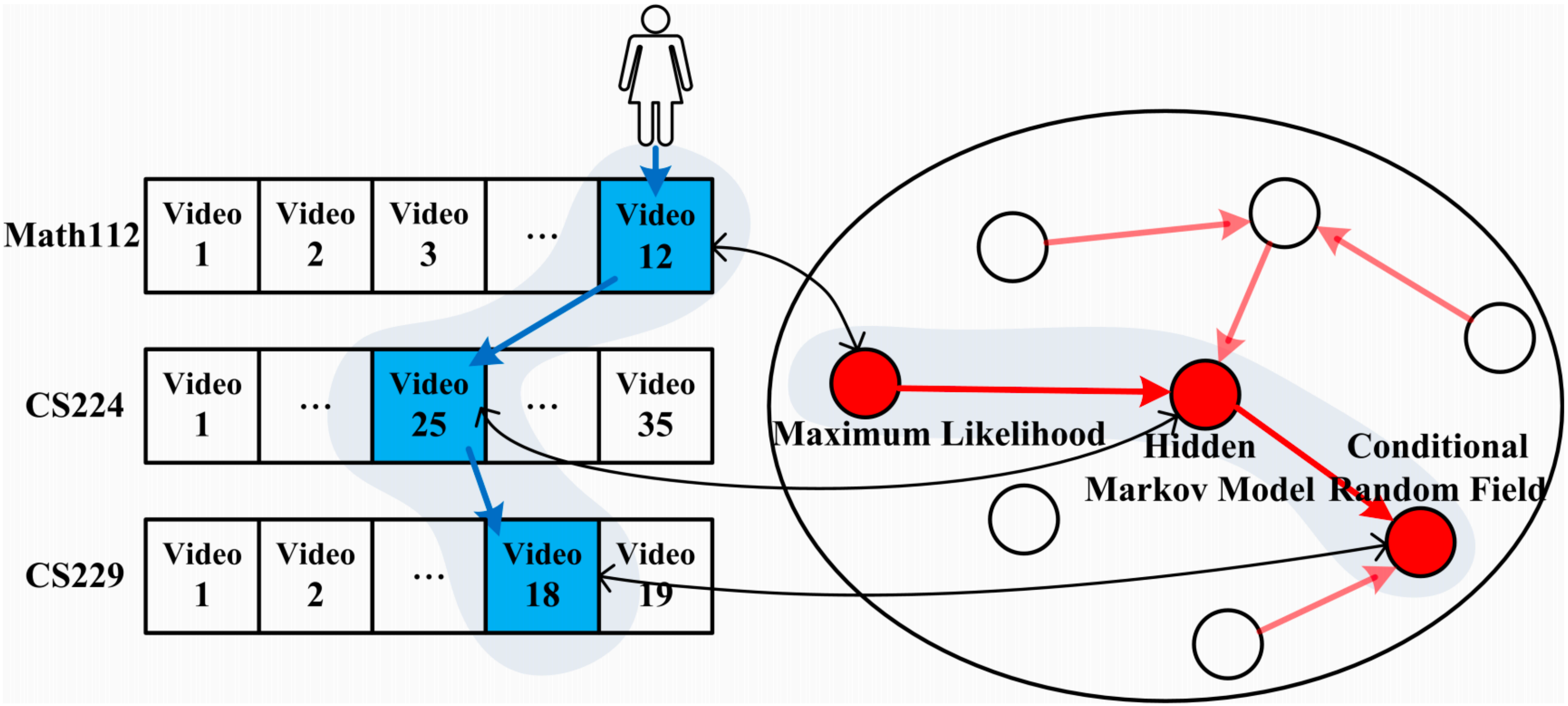
- Dropout
  - KDDCUP 2015, 1,000+ teams worldwide
- Demographics
  - Gender, education, etc.
- User interests
  - computer science, mathematics, psychology, etc.
- …

XiaoMU (小木)



3
user analysis
1
Behavior logs
Behavior modeling
User Profiling
Incentive analysis
Course recommendation
Automated video navigation
Question answering
course analysis
2
Course Content
Concept extraction
Prerequisite relation mining
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**.

Concept Extraction

Candidate Concept Extraction → **Semantic Representation Learning** → **Graph-based Ranking**

In this course, we will teach some basic knowledge about data mining and its application in business intelligence.

Video script

data mining

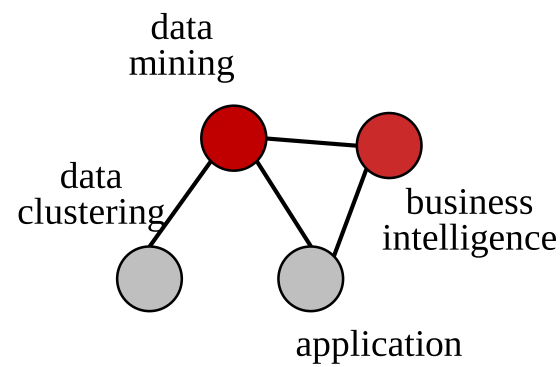
0.8	0.2	0.3	…	0.0	0.0

business intelligence

0.1	0.1	0.2	…	0.8	0.7

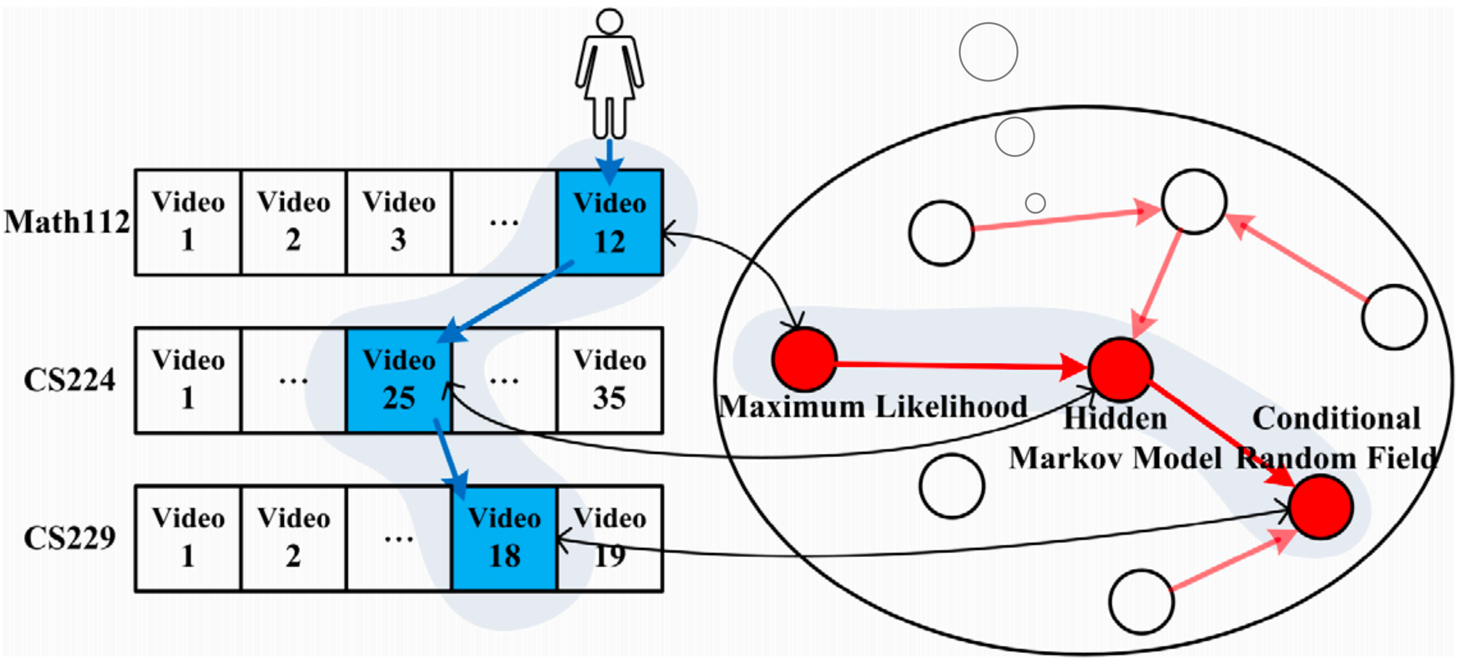
Vector representation

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

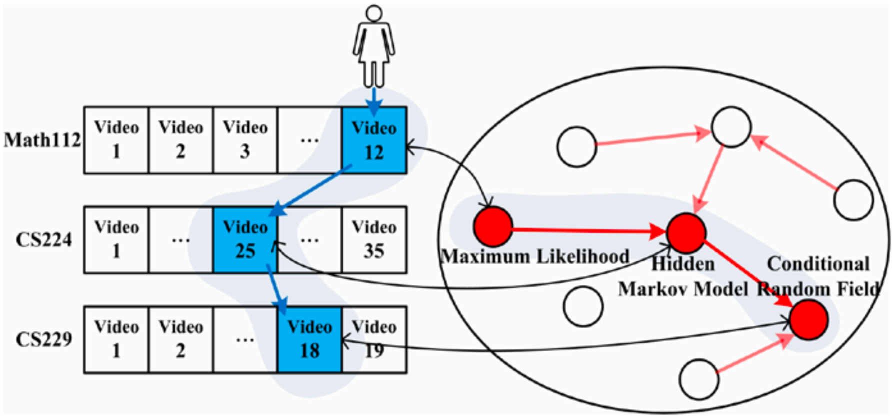
business intelligence

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

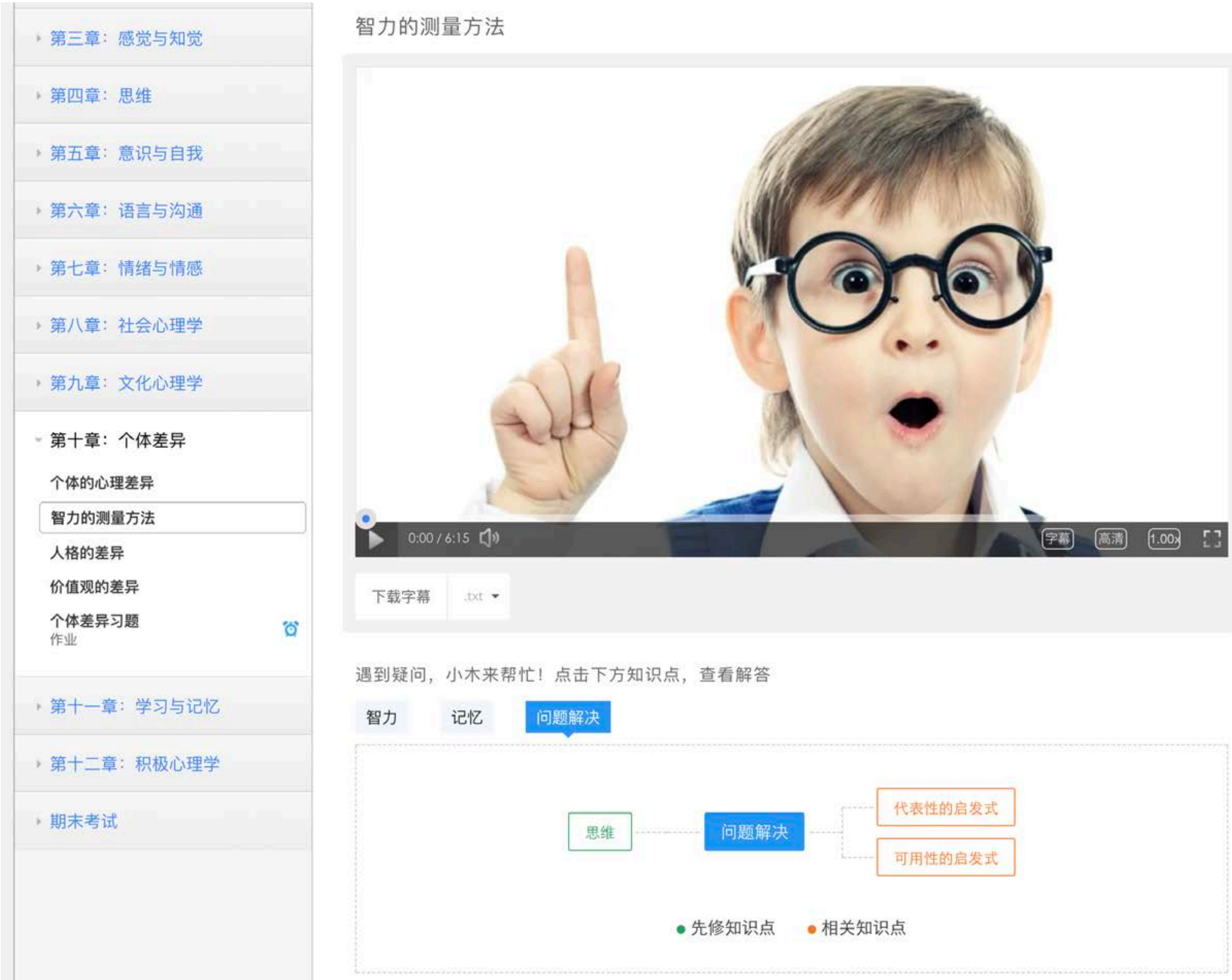
Classifier		ML		DSA		CAL	
	M	1	10	1	10	1	10
SVM	P	63.2	60.1	60.7	62.3	61.1	61.9
	R	68.5	72.4	**69.3**	67.5	**67.9**	68.3
	F_1	65.8	65.7	64.7	64.8	64.3	64.9
NB	P	58.0	58.2	62.9	62.6	60.1	60.6
	R	58.1	60.5	62.3	61.8	61.2	62.1
	F_1	58.1	59.4	62.6	62.2	60.6	61.3
LR	P	66.8	67.6	63.1	62.0	62.7	63.3
	R	60.8	61.0	64.8	66.8	63.6	64.1
	F_1	63.7	64.2	63.9	64.3	61.6	62.9
RF	P	**68.1**	**71.4**	**69.1**	**72.7**	**67.3**	**70.3**
	R	**70.0**	**73.8**	68.4	**72.3**	67.8	**71.9**
	F_1	**69.1**	**72.6**	**68.7**	**72.5**	**67.5**	**71.1**

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

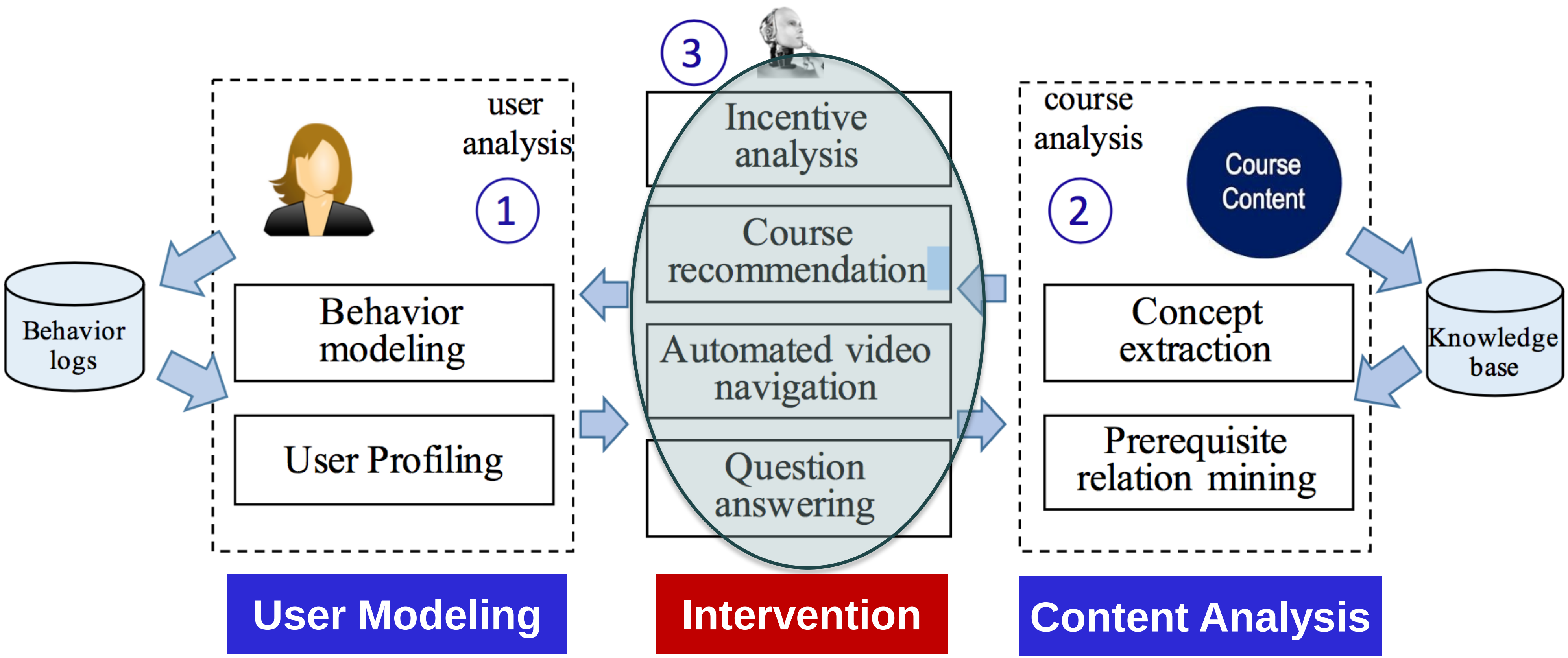
- SVM, NB, LR, and RF are different classification models
- It seems that with the defined distance functions, RF achieves the best

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

System Deployed

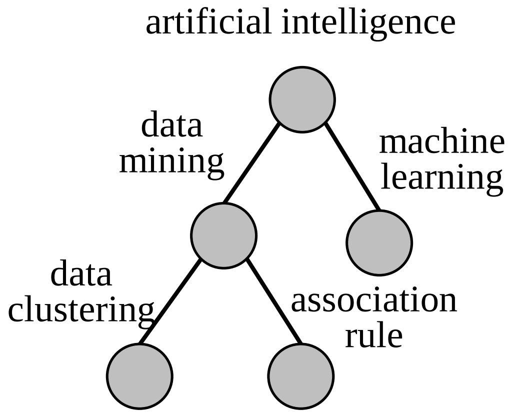


XiaoMU (小木)



user analysis
1
Behavior logs
Behavior modeling
User Profiling
3
Incentive analysis
Course recommendation
Automated video navigation
Question answering
course analysis
2
Course Content
Concept extraction
Prerequisite relation mining
Knowledge base
User Modeling
Intervention
Content Analysis

What we can do?

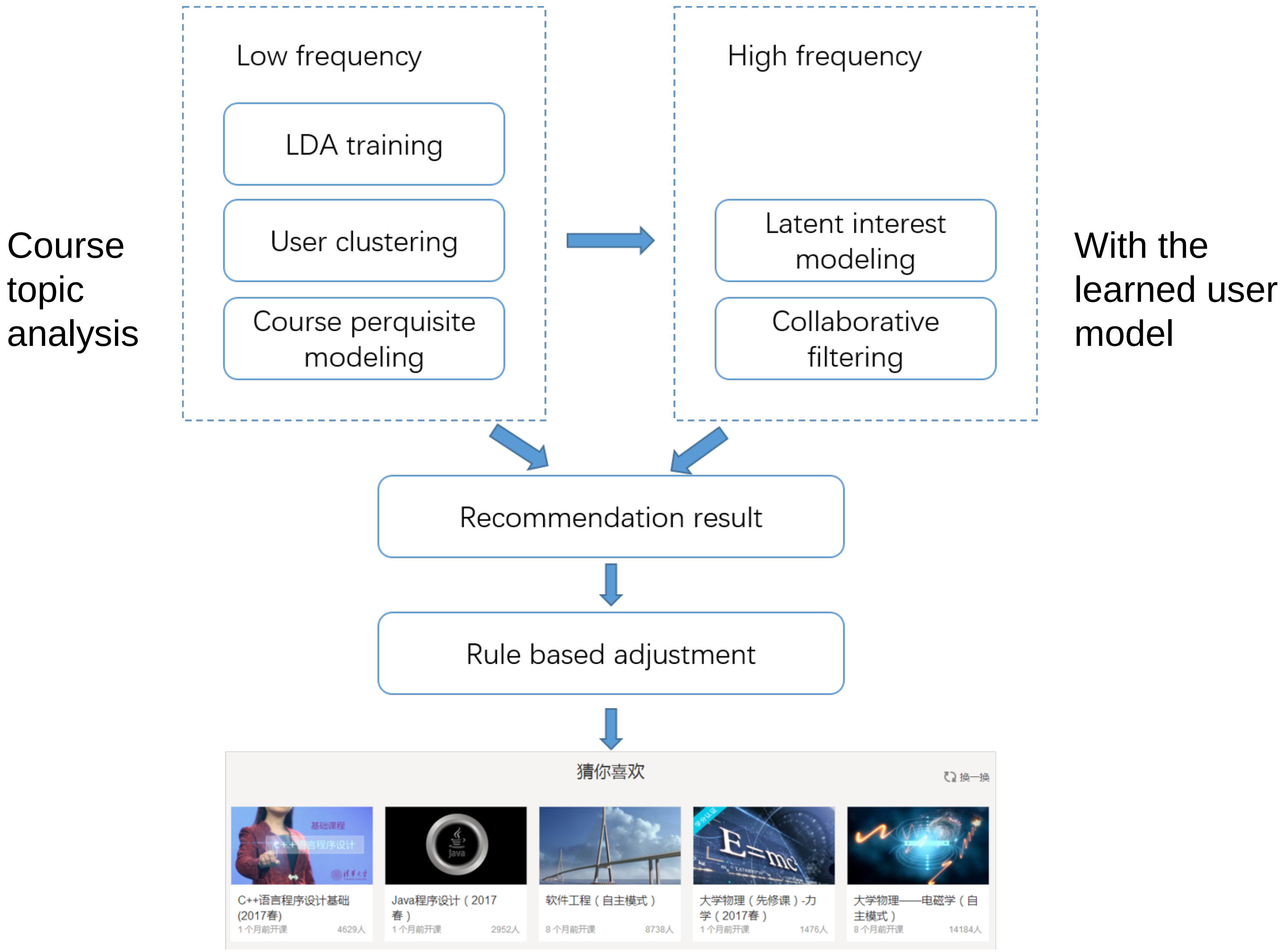


User modeling

Knowledge

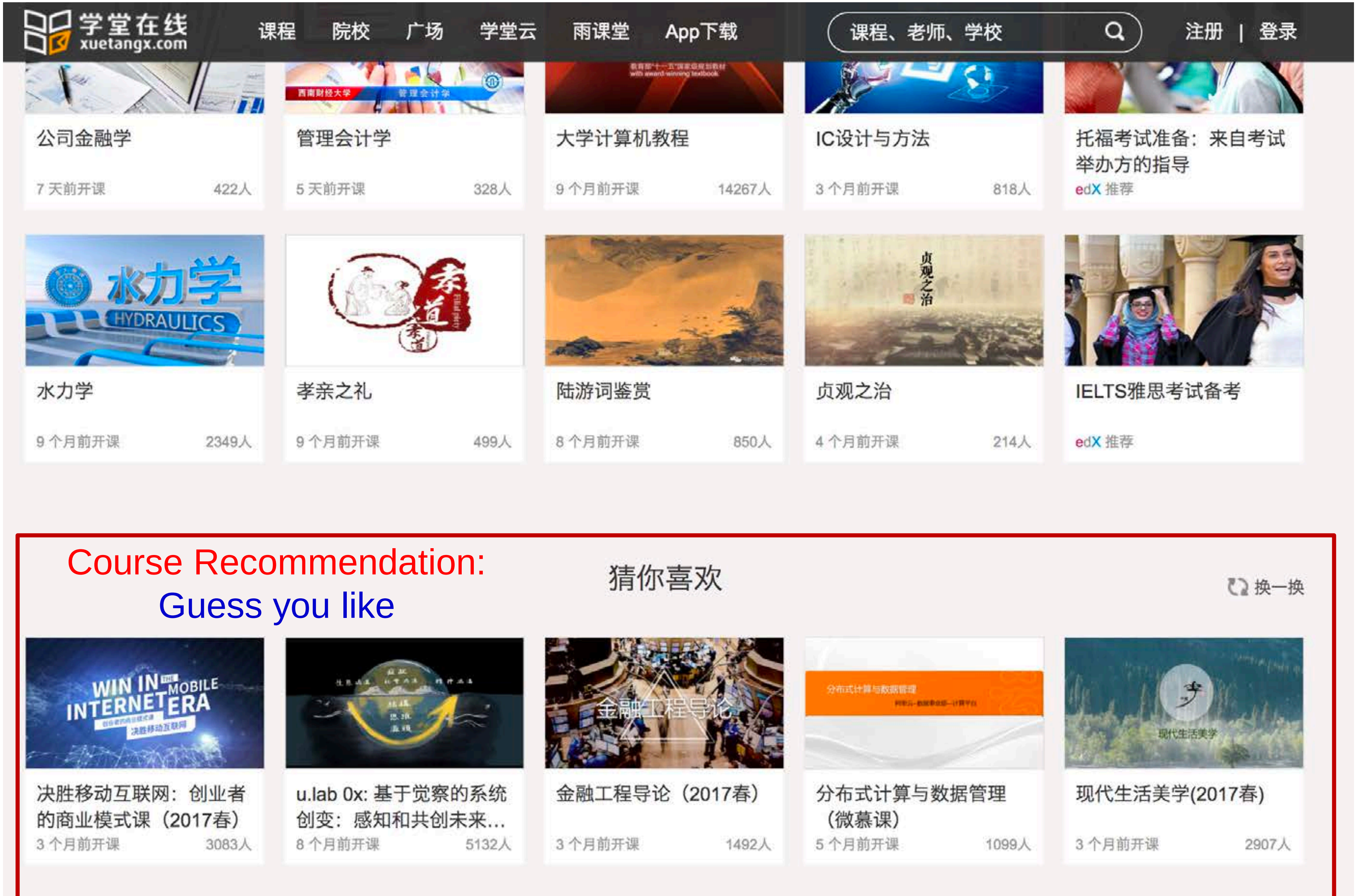
- Let start with a simple case
  - Course recommendation based on user interest

Course Recommendation



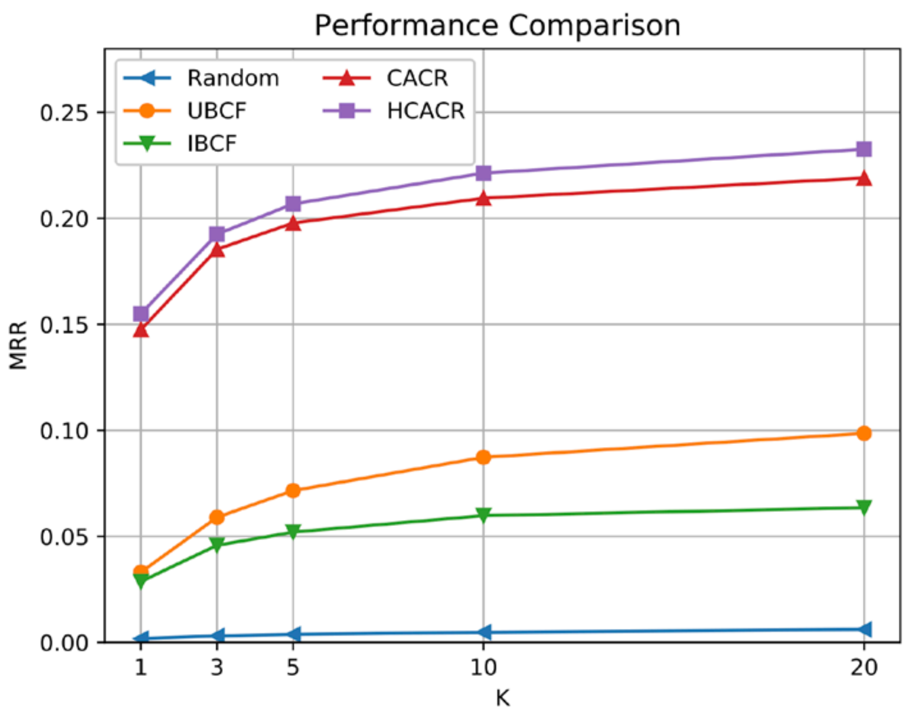
[1] Xia Jing, Jie Tang, Wenguang Chen, Maosong Sun, and Zhengyang Song. Guess You Like: Course Recommendation in MOOCs. **WI'17**.

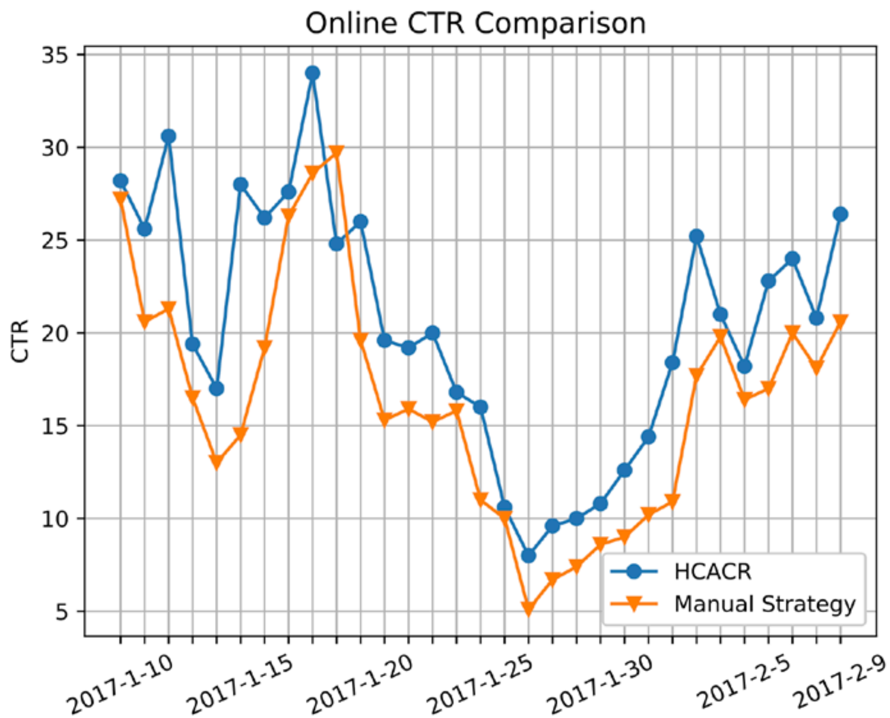
Course Recommendation



Course Recommendation:
Guess you like

Online A/B Test





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

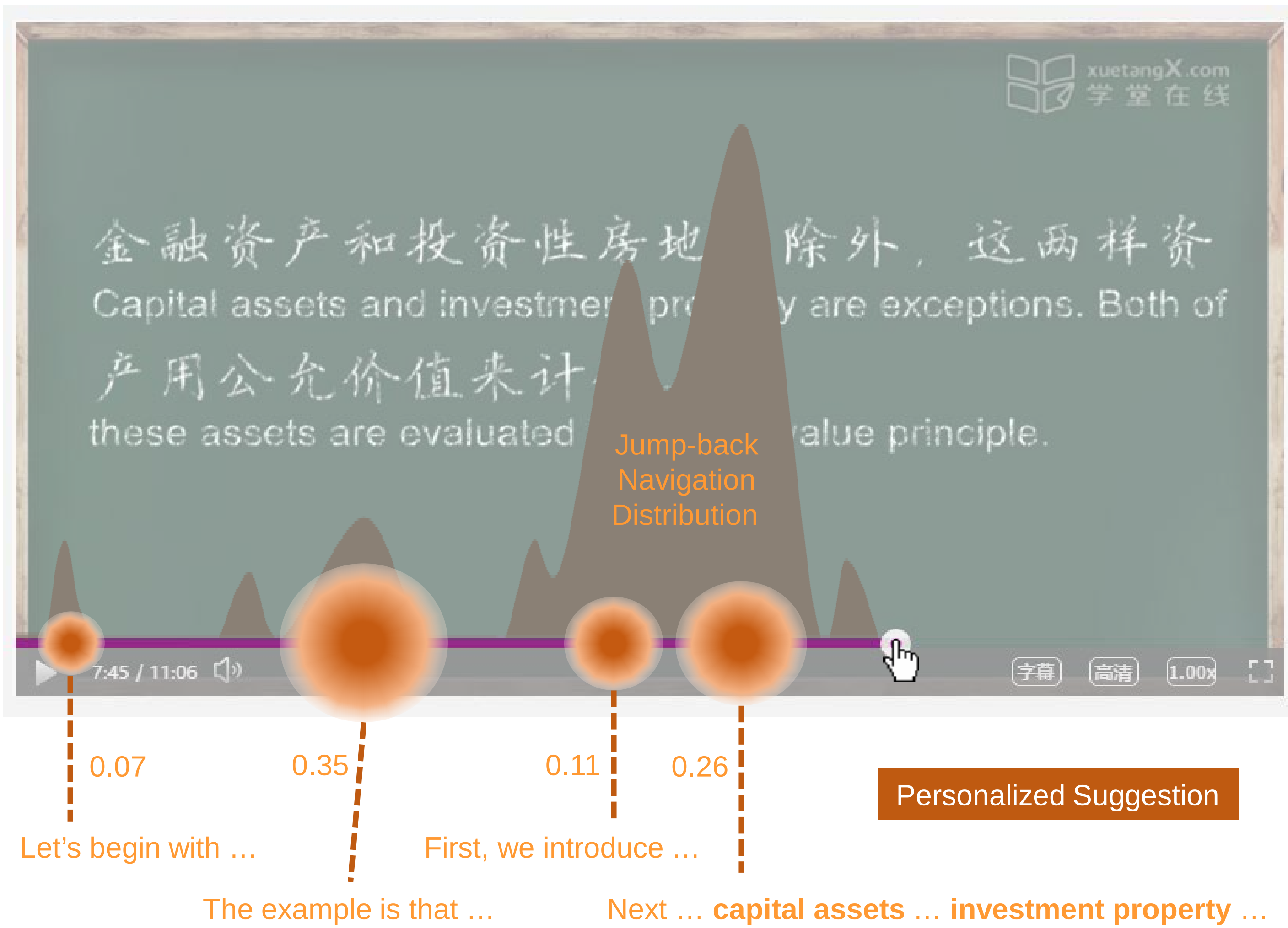
Comparison methods:

HCACR – Our method

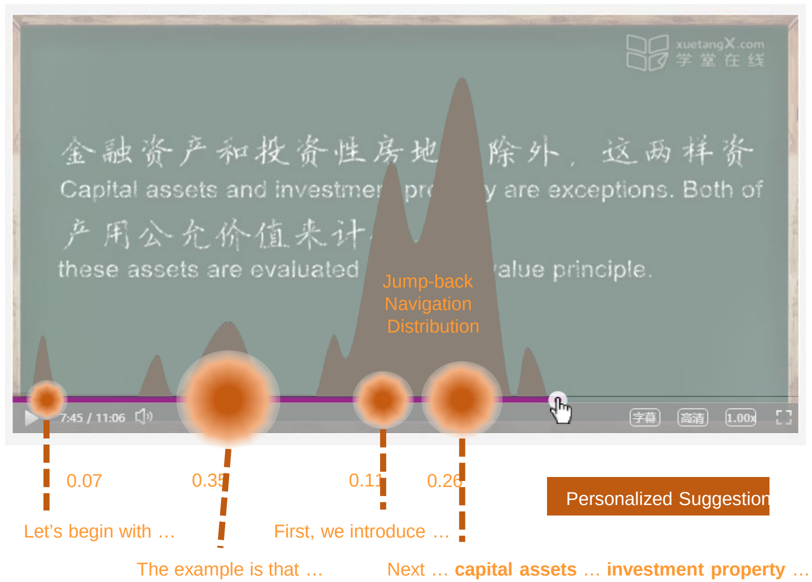
Manual strategy

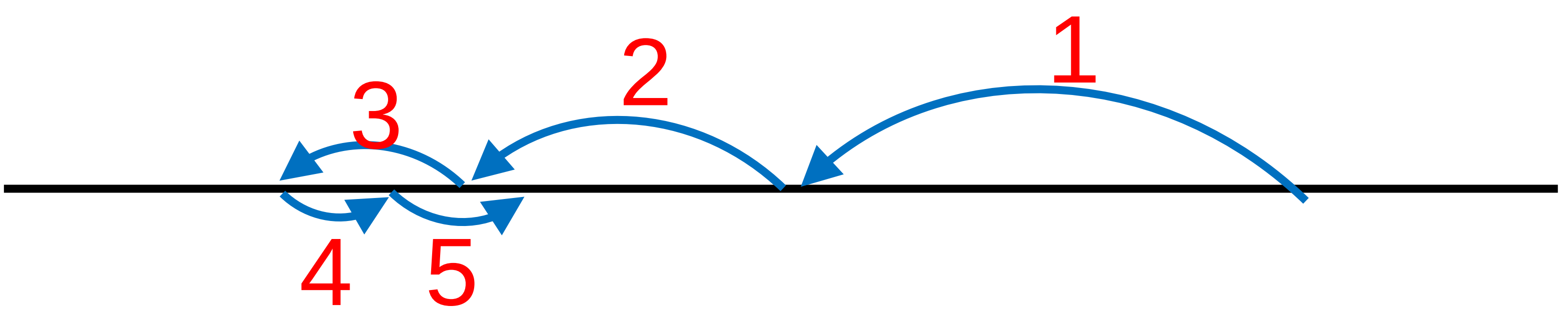
- Let start the simplest case
  - Course recommendation based on user interest
- What can we else?
  - Interaction when watching video?

Smart Jump
—Automated suggestion for video navigation



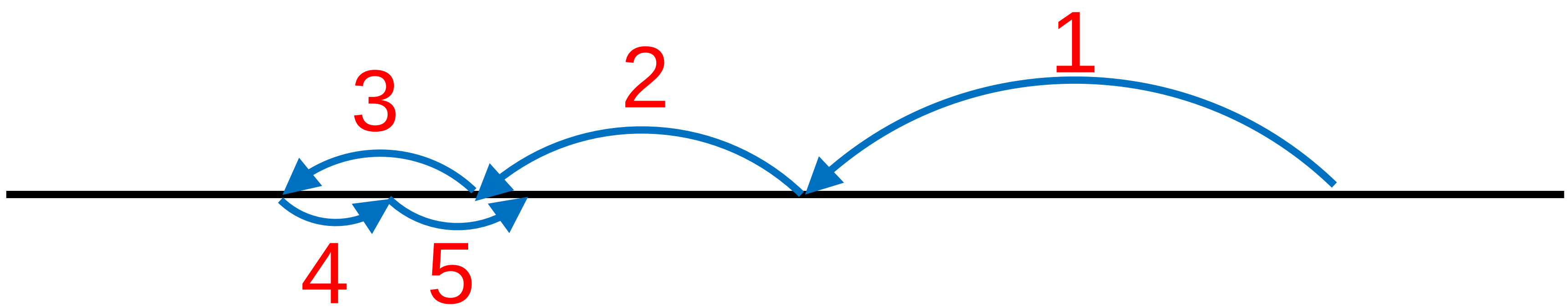
Average Jump



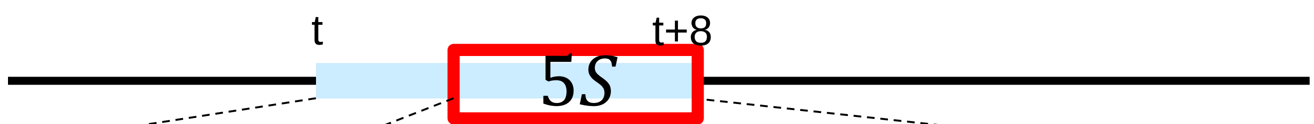


On Average: 2.6 Clicks = 5 seconds

Two Numbers



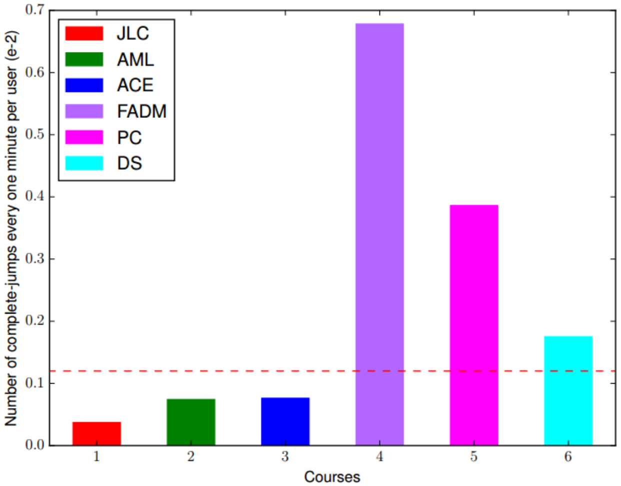
On Average: 2.6 Clicks = 5 seconds

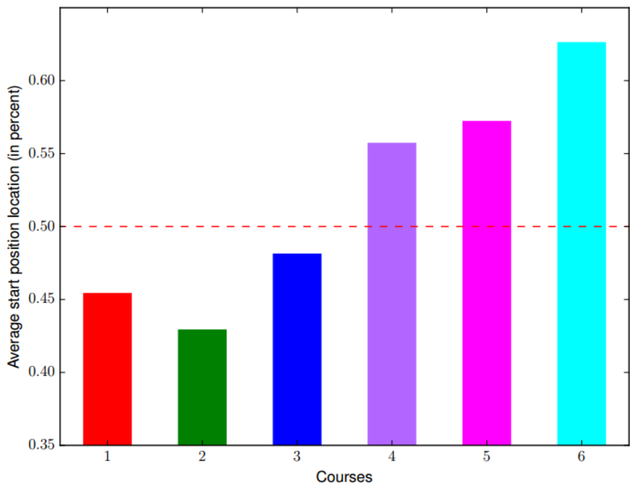


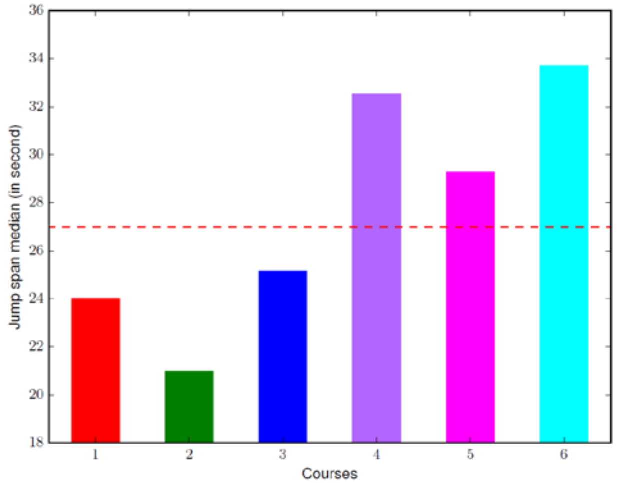
According to what we have discussed we find that the fifth activity belongs to cash outflow of a business activity.

$$5S \times 8{,}000{,}000\ users = 1.3\ years$$

Observations – Course Related





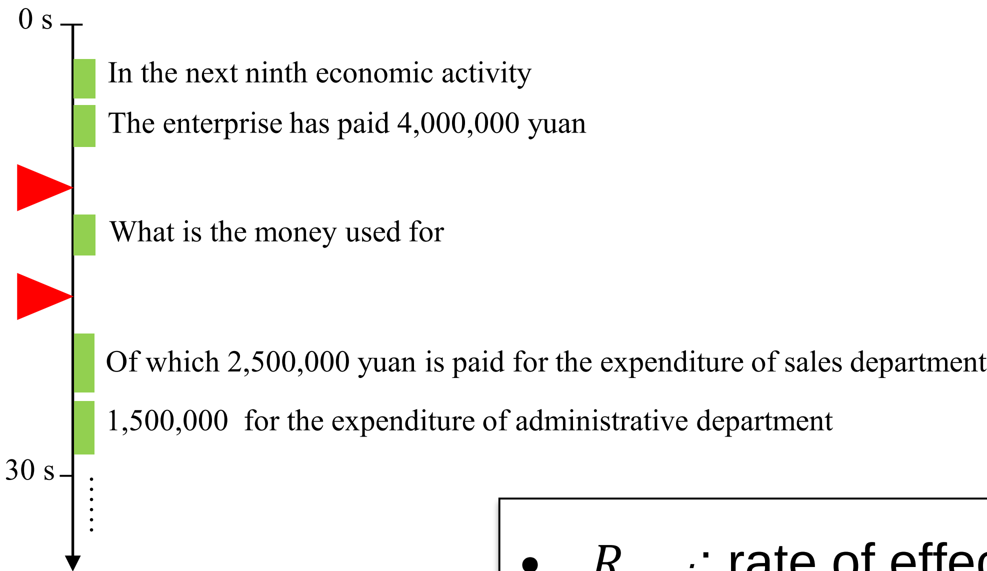


Science courses contain much more frequent jump-backs than non-science courses.

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.

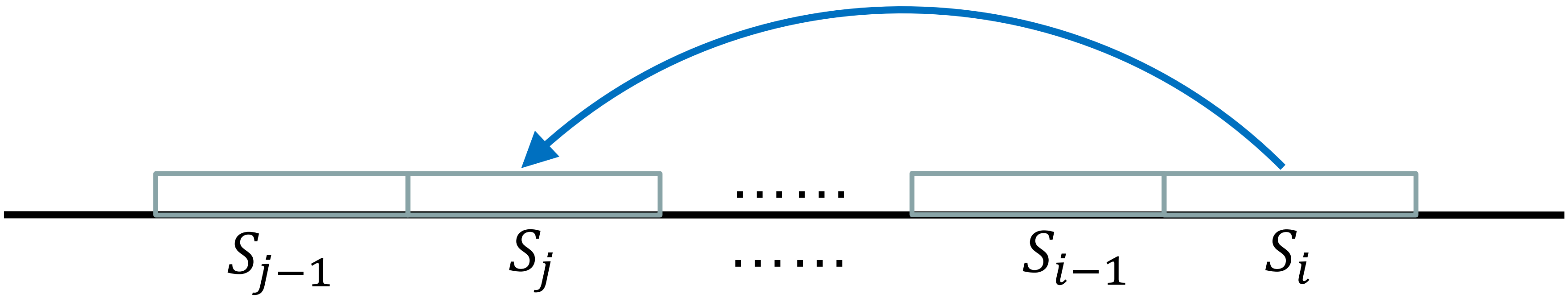
Video Segmentation



$$\underset{\Delta t}{\operatorname{argmax}}\ 2\frac{\boxed{R_{e\_cj}} \cdot \boxed{R_{n\_s}}}{R_{e\_cj} + R_{n\_s}}$$

- 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



$$\operatorname*{argmax}_{\Theta} P(s_j | u, v, s_i; \Theta)$$

[1] Han Zhang, Maosong Sun, Xiaochen Wang, Zhengyang Song, Jie Tang, and Jimeng Sun. Smart Jump: Automated Navigation Suggestion for Videos in MOOCs. **WWW'17**, pages 331-339.

Prediction Results

Course	Model	AUC	P@1	P@3	P@5
Science	**LRC**	72.46	35.95	65.54	80.13
	SVM	71.92	35.45	66.15	81.99
	FM	74.02	37.61	**76.04**	**89.59**
Non-science	**LRC**	72.59	69.23	73.23	89.32
	SVM	73.52	68.39	76.64	91.30
	FM	73.57	67.56	**88.43**	**96.05**

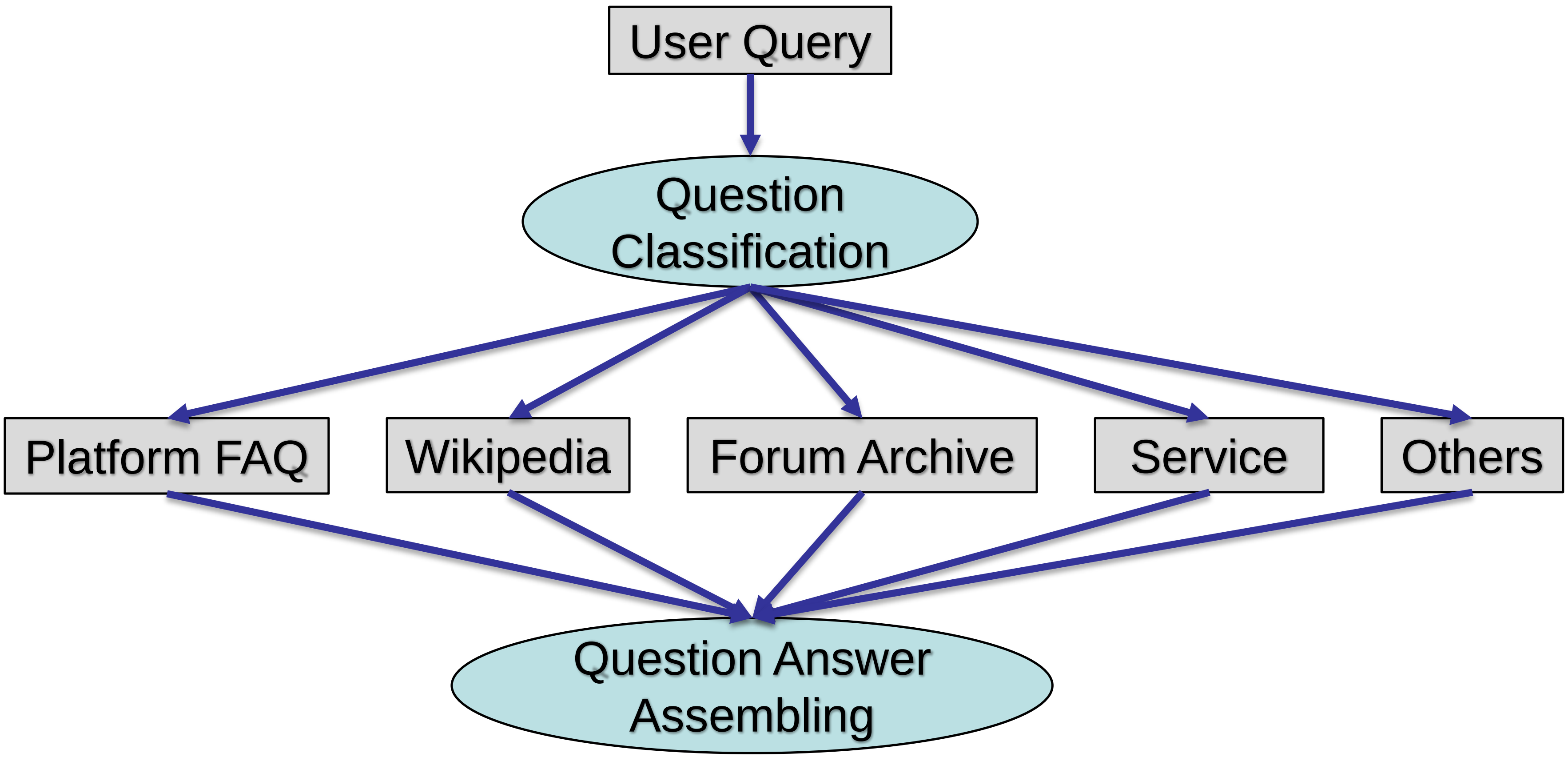
- LRC, SVM, and FM are different models
- FM is defined as follows

$$\hat{y}(\mathbf{x}_i) = w_0 + \sum_{j=1}^{d} w_j x_{i,j} + \boxed{\sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} x_{i,j} x_{i,j'} \langle \mathbf{p}_j, \mathbf{p}_{j'} \rangle}$$

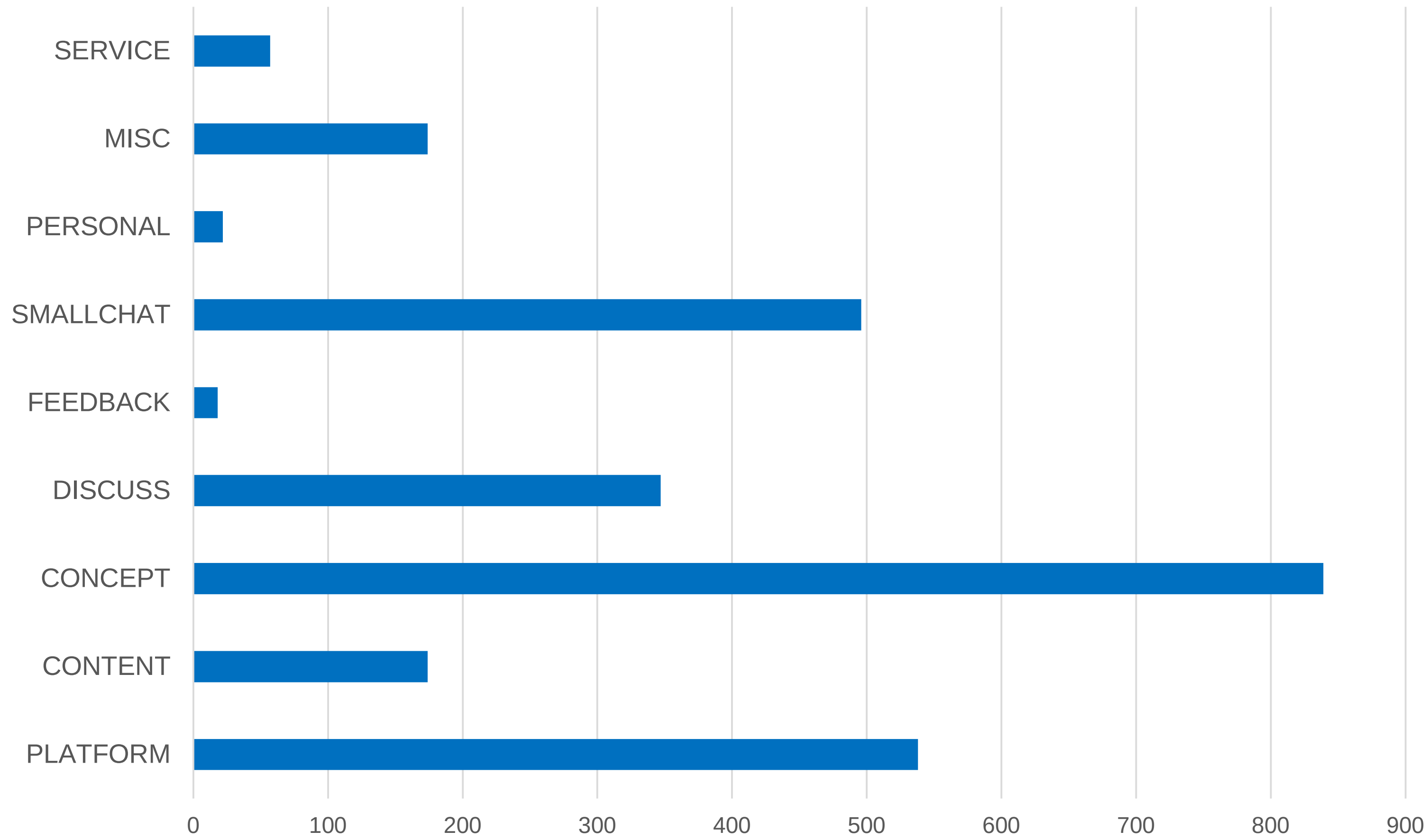
More

- 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?

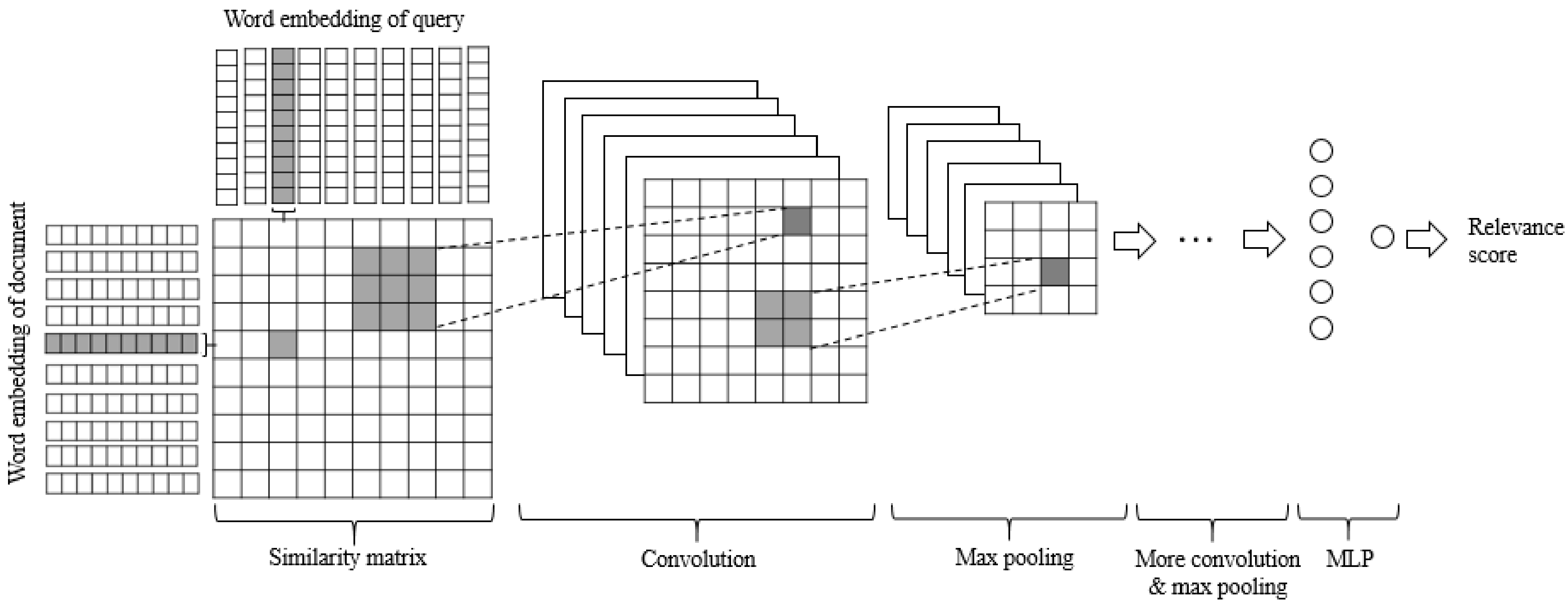
Question Answering



Category Distribution



Question Classification



- #Training (March 2017 – August 2017): 2162
- #Test (September 2017): 499

  Precision: 0.77, Recall: 0.78

Answer Selection

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

	MRR	Hit @ 1	Hit @ 3	Hit @5
ES (TF-IDF)	0.617	0.558	0.698	0.748
Word2vec + WMD	0.695	0.602	0.745	0.817
Word2vec + Cosine	0.653	0.577	0.685	0.726
1.0*WMD+1.5*ES	0.728	0.640	0.781	0.845

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?
  - Interaction->intervention

XiaoMU would like to ask you

Question: What are the shortcomings of Raven Progressive Test? (3 users thumbs up)

Fundamental Challenges (3W):

- When
- to Whom
- ask What (question)

Bandit Learning with Implicit Feedback

[1] Yi Qi, Qingyun Wu, Hongning Wang, Jie Tang, and Maosong Sun. Bandit Learning with Implicit Feedback. **NIPS'18**.

Bandit (Online) Learning

An online learning framework: contextual bandit

- Click/Buy etc. as reward, features of content/user/etc. as context.
- Adaptively and sequentially learning
- Successfully deployed for recommender system and ad displaying.

The problem is about the definition of rewards

- Is non-click indicates negative impression?
- Examination hypothesis:
  - Click occurs if and only if examination happens
  - Implication: no-click DOES NOT necessarily mean negative feedback

Model

Classical bandit model with linear reward:

$$\mathbf{E}[r_{t,a}|\mathbf{x}_{t,a}] \quad = \quad \mathbf{x}_{t,a}^{\top}\boldsymbol{\theta}_a^*.$$

- Reward is 1 if clicked, 0 if non-clicked.
  - Inevitably linear regret.

Our model: E-C bandit (Examination-click bandit)

$$\mathbb{P}(C_t = 1|E_t = 0, \mathbf{x}_{C,t}) = 0$$

$$\mathbb{P}(C_t = 1|E_t = 1, \mathbf{x}_{C,t}) = \rho(\mathbf{x}_{C,t}^{\mathsf{T}}\boldsymbol{\theta}_C^*)$$

$$\mathbb{P}(E_t = 1|\mathbf{x}_{E,t}) = \rho(\mathbf{x}_{E,t}^{\mathsf{T}}\boldsymbol{\theta}_E^*)$$

Thus:

$$\mathbb{E}[C_t|\mathbf{x}_t] = \rho(\mathbf{x}_{C,t}^{\mathsf{T}}\boldsymbol{\theta}_C^*)\rho(\mathbf{x}_{E,t}^{\mathsf{T}}\boldsymbol{\theta}_E^*).$$

The common goal: regret minimization

$$\text{BayesRegret}(T, \pi) = \sum_{t=1}^{\mathsf{T}} \mathbb{E}\big[\max_{a \in \mathcal{A}_t} f_{\boldsymbol{\theta}^*}(\mathbf{x}^a) - f_{\boldsymbol{\theta}^*}(\mathbf{x}^{a_t})\big]$$

Evaluation – Empirical data

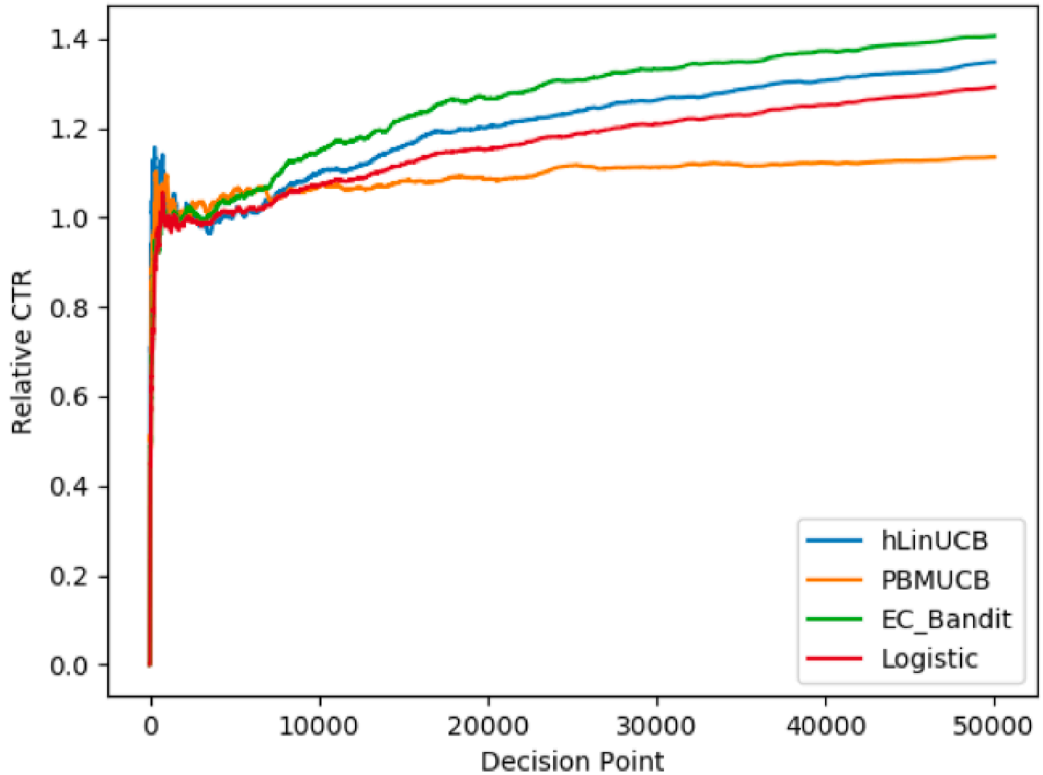
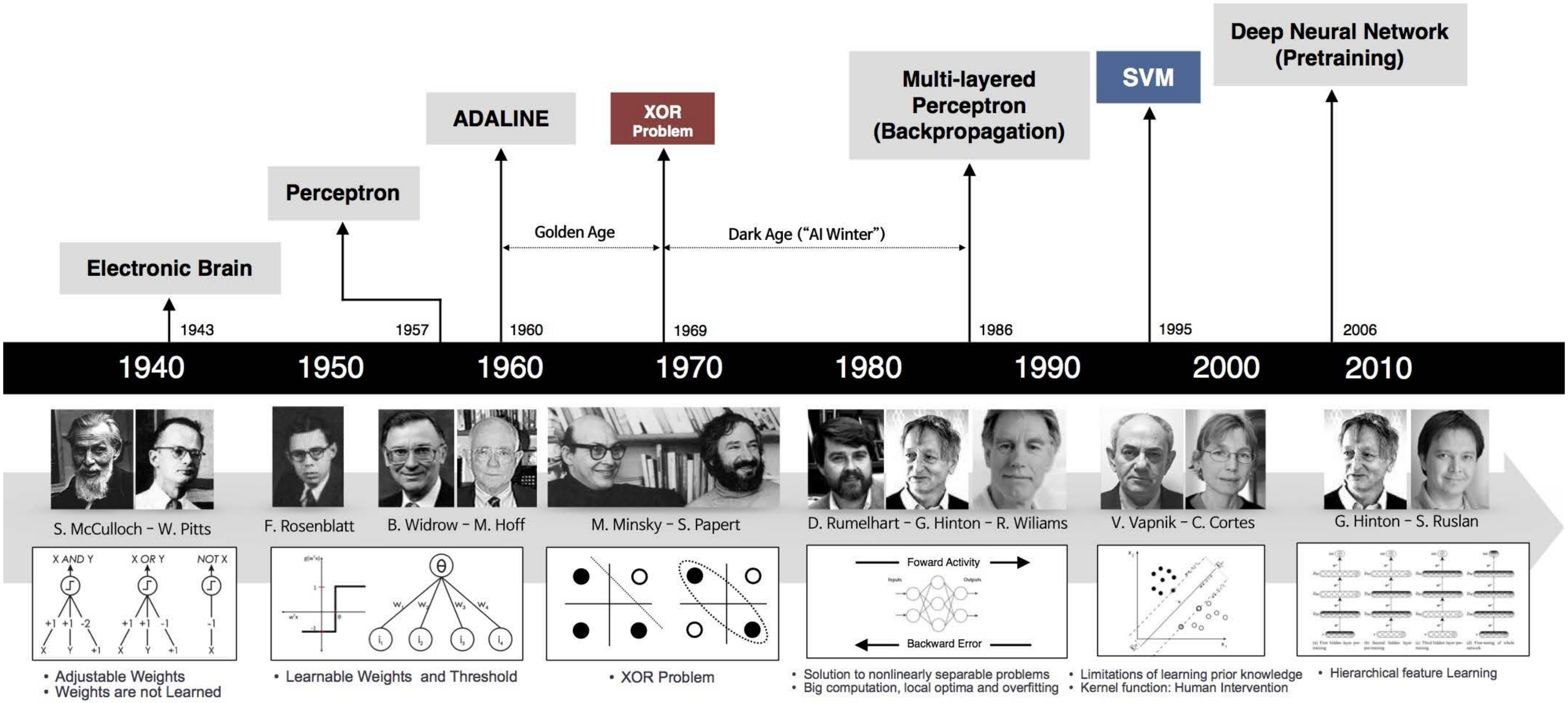
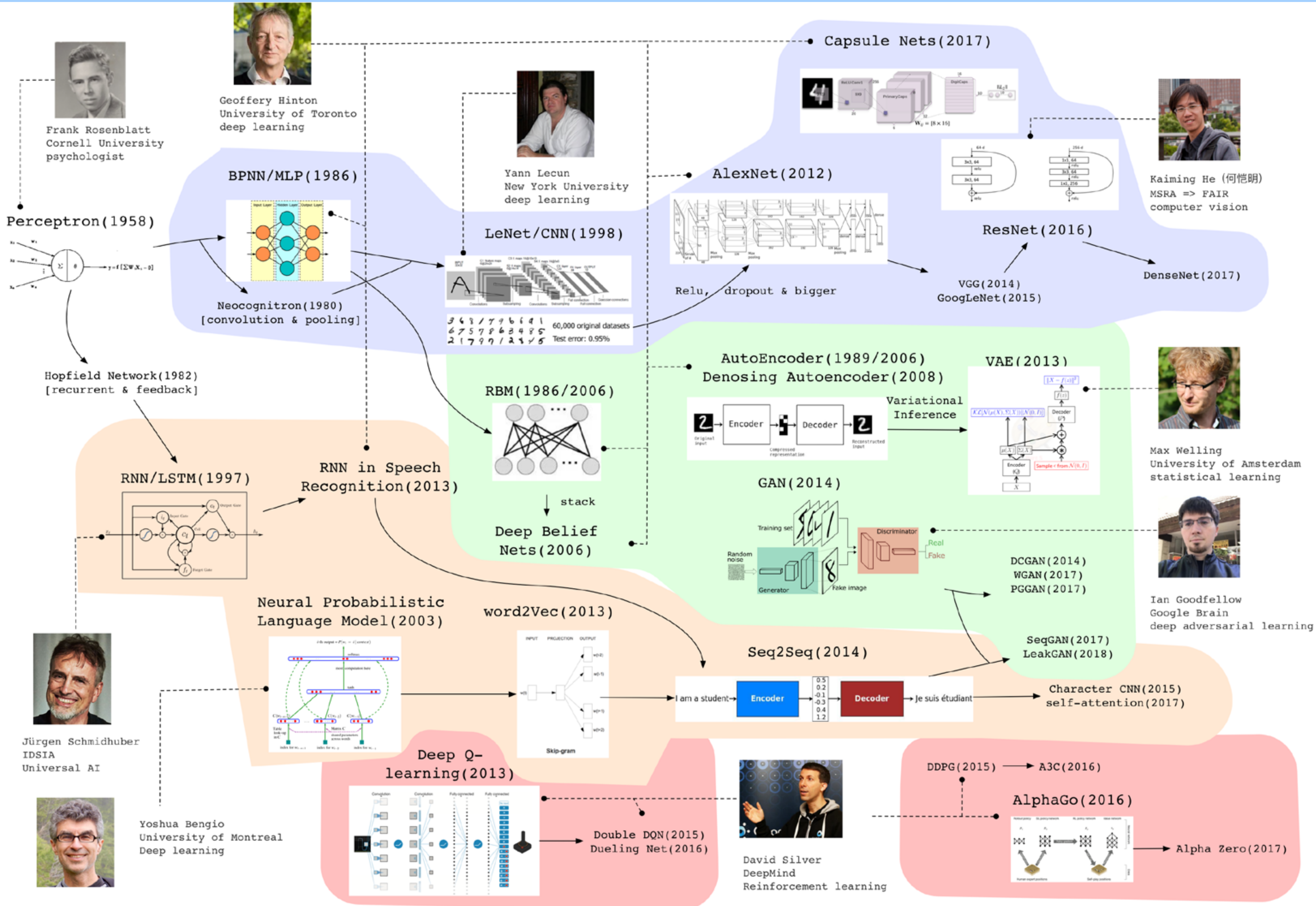


Figure 4: Performance comparison on MOOC videos' data

History of Neural Networks





Capsule Nets(2017)
Geoffrey Hinton
University of Toronto
deep learning
Frank Rosenblatt
Cornell University
psychologist
Yann Lecun
New York University
deep learning
Kaiming He (何恺明)
MSRA => FAIR
computer vision
BPNN/MLP(1986)
AlexNet(2012)
Perceptron(1958)
LeNet/CNN(1998)
ResNet(2016)
VGG(2014)
GoogLeNet(2015)
DenseNet(2017)
Relu, dropout & bigger
Neocognitron(1980)
[convolution & pooling]
60,000 original datasets
Test error: 0.95%
Hopfield Network(1982)
[recurrent & feedback]
AutoEncoder(1989/2006)
Denosing Autoencoder(2008)
VAE(2013)
RBM(1986/2006)
Variational
Inference
Encoder
Decoder
Max Welling
University of Amsterdam
statistical learning
RNN/LSTM(1997)
RNN in Speech
Recognition(2013)
stack
GAN(2014)
Deep Belief
Nets(2006)
Training set
Discriminator
Real
Fake
Random noise
Generator
Fake image
DCGAN(2014)
WGAN(2017)
PGGAN(2017)
Ian Goodfellow
Google Brain
deep adversarial learning
Neural Probabilistic
Language Model(2003)
word2Vec(2013)
Skip-gram
Seq2Seq(2014)
SeqGAN(2017)
LeakGAN(2018)
I am a student
Je suis étudiant
Character CNN(2015)
self-attention(2017)
Jürgen Schmidhuber
IDSIA
Universal AI
Deep Q-
learning(2013)
DDPG(2015)
A3C(2016)
AlphaGo(2016)
Double DQN(2015)
Dueling Net(2016)
Alpha Zero(2017)
Yoshua Bengio
University of Montreal
Deep learning
David Silver
DeepMind
Reinforcement learning

认知思考：AI Trend

- From perceptron to cognition

Storage & Computing

Recognize text, images, objects, voices

Organize and generate knowledge, reasoning

Computing → Perception → Cognition

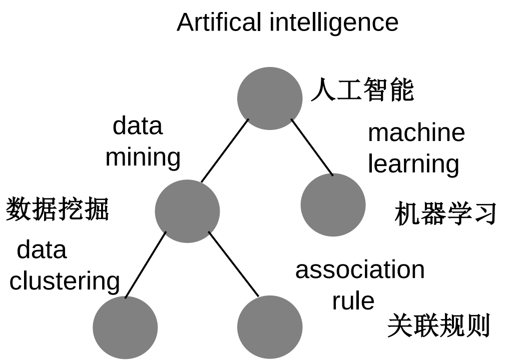
Stochastic vs Deterministic

Uncertainty!

Data → knowledge → Cognition

Knowledge graph:
A graph of concepts

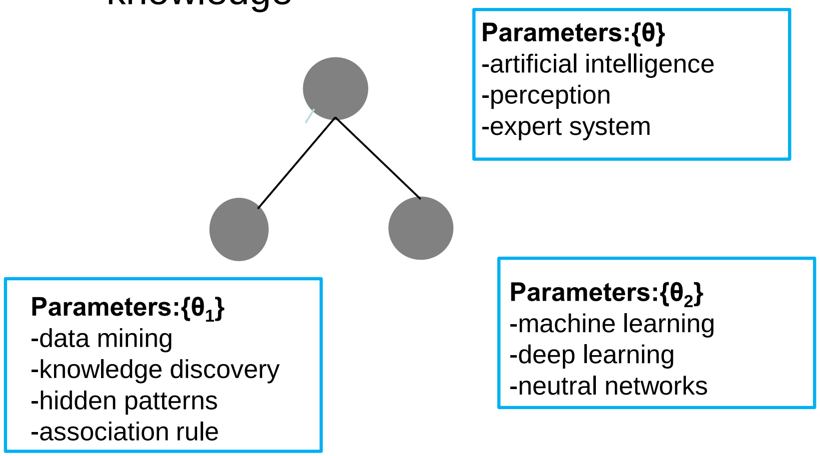
To organize
knowledge



Allow computers to
process data automatically

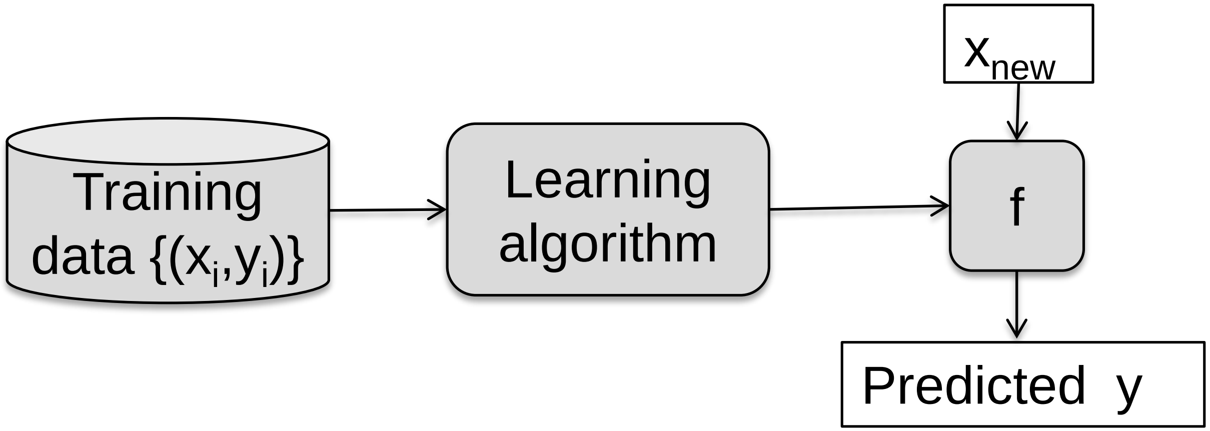
Cognitive graph:
A graph of generators

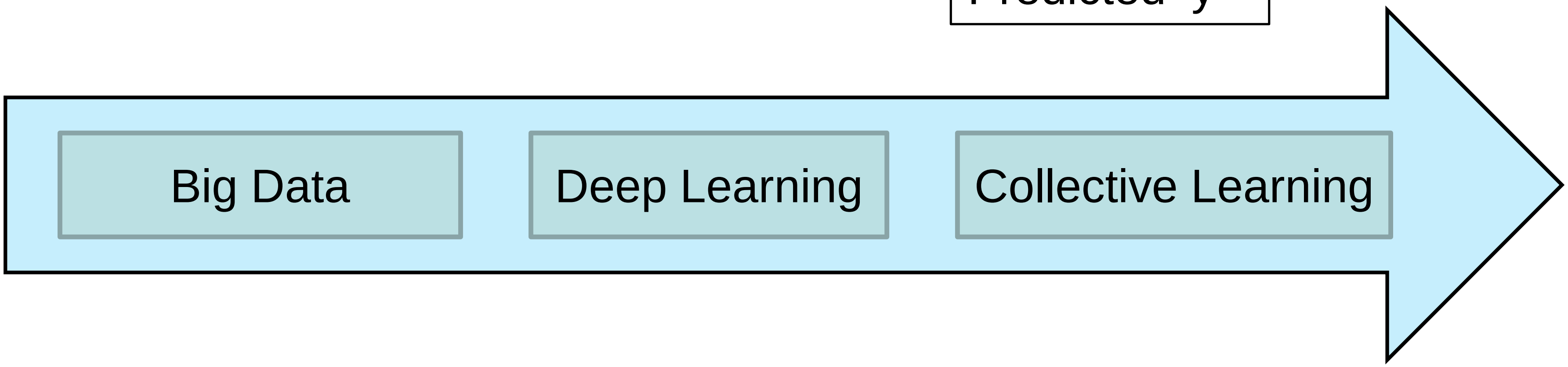
To generate
knowledge

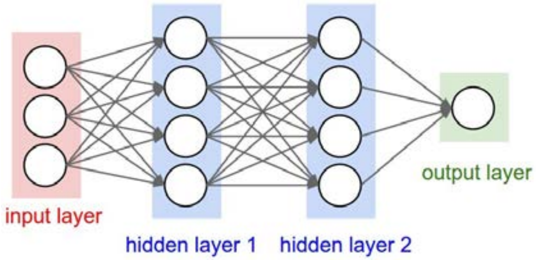


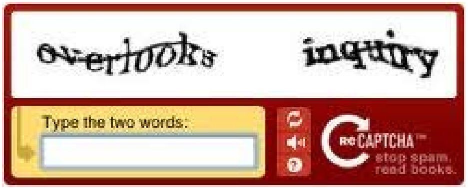
Allow computers to
process data autonomously

技术思考：Machine Learning









数据思考：MoocData.org

· 821 Teams

KDD Cup 2015

Final Submissions

2015-05-01 2015-07-12

Home > Competitions >

- Information
- Introduction
- Rules
- Data
- Evaluation
- Timeline
- Prizes
- Organizers
- Rank
- Discussion Board
- Submission
- Make a submission
- Make a final submission

If you have any ques

support@kddcup201

Update, August 4 :

Here is the Schedule

KDD Cup 2015 Workshop Sc

9:00 – 9:30 **Opening:** Informa

9:30 – 10:30 **Invited Talk:** Ja

10:30 – 11:00 Coffee Break

11:00 – 11:25 **10th Prize:** Ikki

MOOCs Dropout Prediction.

11:25 – 11:50 **9th Prize:** Chih

and Sin-Ya Peng A Linear En

Prediction.

MoocData

data Publications Challenges About

Name	Size	Keywords	Download link	Reference
Xiaomu Questioning	237.7 MB	Bandit Learning, Implicit Feedback	data	NeurIPS'18

Xiaomu Questioning

Data Description:

Xiaomu Questioning is an innovative feature which actively and adaptively displays banner containing questions in lecture videos, aiming at improving learners' engagement.

Learner is allowed to click on the displayed banner so as to see answer to the question.

Random selection strategy on positions and questions are used and the data of displayed question/clicks/feature/candidate items are collected.

The data is in .pkl form and must be deserialized by the package "pickle" of python3.

Example code:

> data = pickle.load(open("rawData.pkl", 'rb'))

The data is collected from Xiaomu Questioning, consisting of 7 columns.

The description of each column:

1st column: vidx - integer, the ID of video;

2nd column: aid - integer, the ID of recommended question;

3rd column: reward - integer, the observed feedback(1 for click, 0 for non-click)

4th column: candidate list of arms - list of integer, the available candidates at this decision point;

5th column: feature of relevance - two dimensional numpy array, each row representing the corresponding arm in the candidate list;

6th column: feature of examination - two dimensional numpy array, each row representing the corresponding arm in the candidate list;

7th column: position of the recommended arm - integer;

学科交叉思考

- 2014年，教育部在清华成立在线教育研究中心
  - 计算机系
  - 交叉信息研究院
  - 教育研究院
  - 心理系、社会学系
  - 学堂在线
- 主任：袁驷(原副校长)
- 专注计算与教育的交叉融合研究

教育部副部长林蕙青、中国高教学会会长瞿振元、清华大学党委书记陈旭、清华大学校务委员会副主任袁驷、清华大学副校长薛其坤，清华大学副秘书长聂风华等出席成立大会

Representative Publications

- Yi Qi, Qingyun Wu, Hongning Wang, Jie Tang, and Maosong Sun. Bandit Learning with Implicit Feedback. NIPS'18.
- Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. DeepInf: Modeling Influence Locality in Large Social Networks. In KDD'18.
- Yutao Zhang, Fanjin Zhang, Peiran Yao, and Jie Tang. Name Disambiguation in AMiner: Clustering, Maintenance, and Human in the Loop. In KDD'18.
- Golnoosh Farnadi, Jie Tang, Martine De Cock, and Marie-Francine Moens. User Profiling through Deep Multimodal Fusion. In WSDM'18.
- Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Kuansan Wang, and Jie Tang. Network Embedding as Matrix Factorization: Unifying DeepWalk, LINE, PTE, and node2vec. In WSDM'18.
- Yutao Zhang, Robert Chen, Jie Tang, Jimeng Sun, and Walter Stewart. LEAP: Learning to Prescribe Effective and Safe Treatment Combinations for Multimorbidity. In KDD'17, pages 1315-1324.
- Jie Tang, A.C.M. Fong, Bo Wang, and Jing Zhang. A Unified Probabilistic Framework for Name Disambiguation in Digital Library. IEEE Transaction on Knowledge and Data Engineering (TKDE), 2012, Volume 24, Issue 6, Pages 975-987.
- Jie Tang, Tiancheng Lou, Jon Kleinberg, and Sen Wu. Transfer Learning to Infer Social Ties across Heterogeneous Networks. ACM Transactions on Information Systems (TOIS), 2016, Volume 34, Issue 2, Article No. 7.
- Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In KDD‘08, pages 990-998.
- Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In KDD‘09, pages 807-816.
- Chi Wang, Jiawei Han, Yuntao Jia, Jie Tang, Duo Zhang, Yintao Yu, and Jingyi Guo. Mining Advisor-Advisee Relationships from Research Publication Networks. In KDD‘10, pages 203-212.
- Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In KDD‘11, pages 1397-1405.
- Jie Tang, Sen Wu, Jimeng Sun, and Hang Su. Cross-domain Collaboration Recommendation. In KDD‘12, pages 1285-1293. (Full Presentation & Best Poster Award)
- Jie Tang, Sen Wu, and Jimeng Sun. Confluence: Conformity Influence in Large Social Networks. In KDD‘13, pages 347-355.
- Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V. Chawla. Inferring User Demographics and Social Strategies in Mobile Social Networks. In KDD‘14, pages 15-24.

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

CIT 互联网教育智能技术及应用国家工程实验室

远程教学交互
研发中心
中国移动 China Mobile

知识建模与
分析研发中心
清华大学 Tsinghua University

学习环境设计
与评测研发中心
网龙华渔教育

科大讯飞 iFLYTEK 学习者建模与
学习分析研发中心

http://cit.bnu.edu.cn cit@bnu.edu.cn 010-58807205
北京市海淀区学院南路12号 北京师范大学南院 京师科技大厦A座3层和12层

扫描二维码 关注公众号

THANKS