

大规模在线教育中的知识智能

清华大学知识工程实验室 于济凡

| 讲者介绍



- 清华大学知识工程实验室
- 研究生二年级
- 导师为李涓子教授、唐杰教授
- 研究兴趣包括概念获取与扩展, 智能教育



| 大规模在线教育中的知识智能

Knowledge-driven Online Education



- 大规模在线课程MOOC
- 以知识为核心的在线教育资源组织
- 在线教育智能交互的知识服务

| 大规模在线课程MOOC

Massive Open Online Course

- MOOC (Massive Open Online Courses)
- 针对于大众人群，强调知识共享
- 自2008年诞生以来，服务超过1亿人次



大规模在线课程MOOC

Massive Open Online Course



- MOOC (Massive Open Online Courses)



第一章：绪论

第二章：心理学历史

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

智力的测量方法

智力到底如何进行测量呢？

心理学概论

Introduction to Psychology

第十讲 个体差异

0:01 / 6:15

字幕 高清 1.25x

下载字幕 .txt

| 大规模在线课程MOOC

Massive Open Online Course



MOOC (Massive Open Online Course)



- **Massive**: 数据规模**大**
 - **Open**: 面向人群的**背景各异**
 - **Online**: 用户**交互**数据充分
 - **Course**: 核心在于**知识**的传递与分享
- 挑战
- 重点



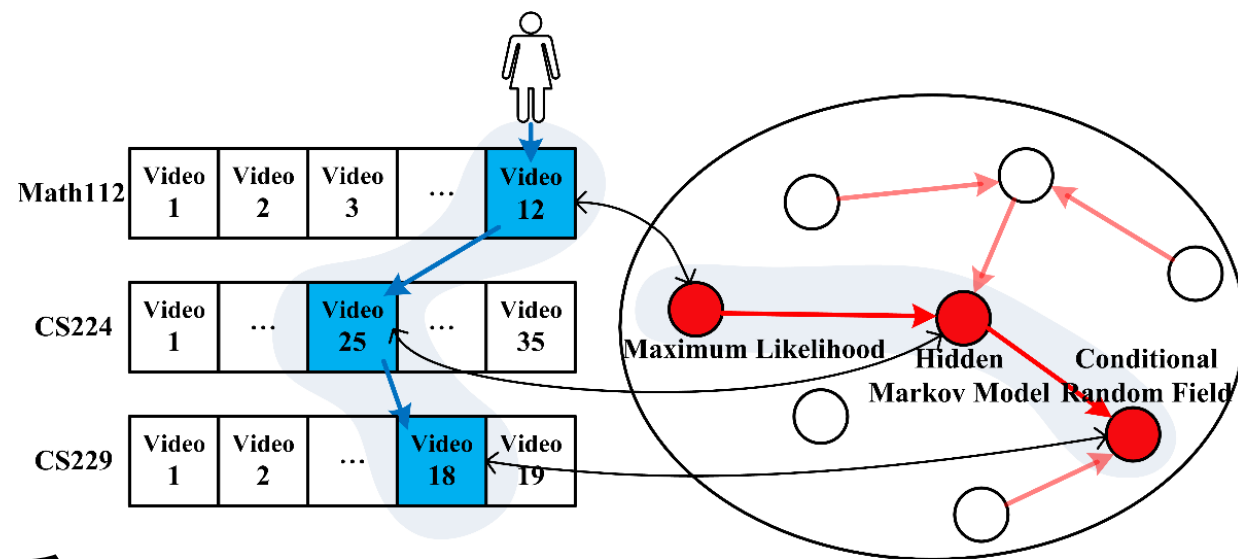
大规模在线课程MOOC

Massive Open Online Course



示例: 智能个人学习路径推荐

- 应用目标:
 - 1) 帮助学生更好地掌握更多知识
 - 2) 帮助提升教学资源的质量
- 建立以知识为核心的在线课程资源体系
- 充分考虑 学生的需求, 知识状态, 构建智能可交互的应用



| 大规模在线教育中的知识智能

Knowledge-driven Online Education



- 大规模在线课程MOOC
- 以知识为核心的在线教育资源组织
- 在线教育智能交互的知识服务

| 以知识为核心的在线教育资源组织

Knowledge centered online education resource organization



- 课程中蕴含大量的知识概念

- 示例:“数据结构”课程

You might learn how to write a **bubble sort** and learn why a **bubble sort** is not as good as a **heapsort**. Next, we are going to talk about the **quick sort** algorithm. **Quicksort** is an algorithm invented in the 1960s by doctor Tony Hoare. It is also called the **partition exchange sort**, and is a typical algorithm based on **divide-and-conquer**.

.....

Now we have the first version of **Q sort**. After we make an analysis on its performance, performance, we will find that **quicksort** is an **unstable sorting algorithm**. Fortunately, the **quick sort** has an average **time complexity** of $n \log n$, and in most cases, it can achieve its optimal performance. We first estimate its performance under **independent uniform distribution**.

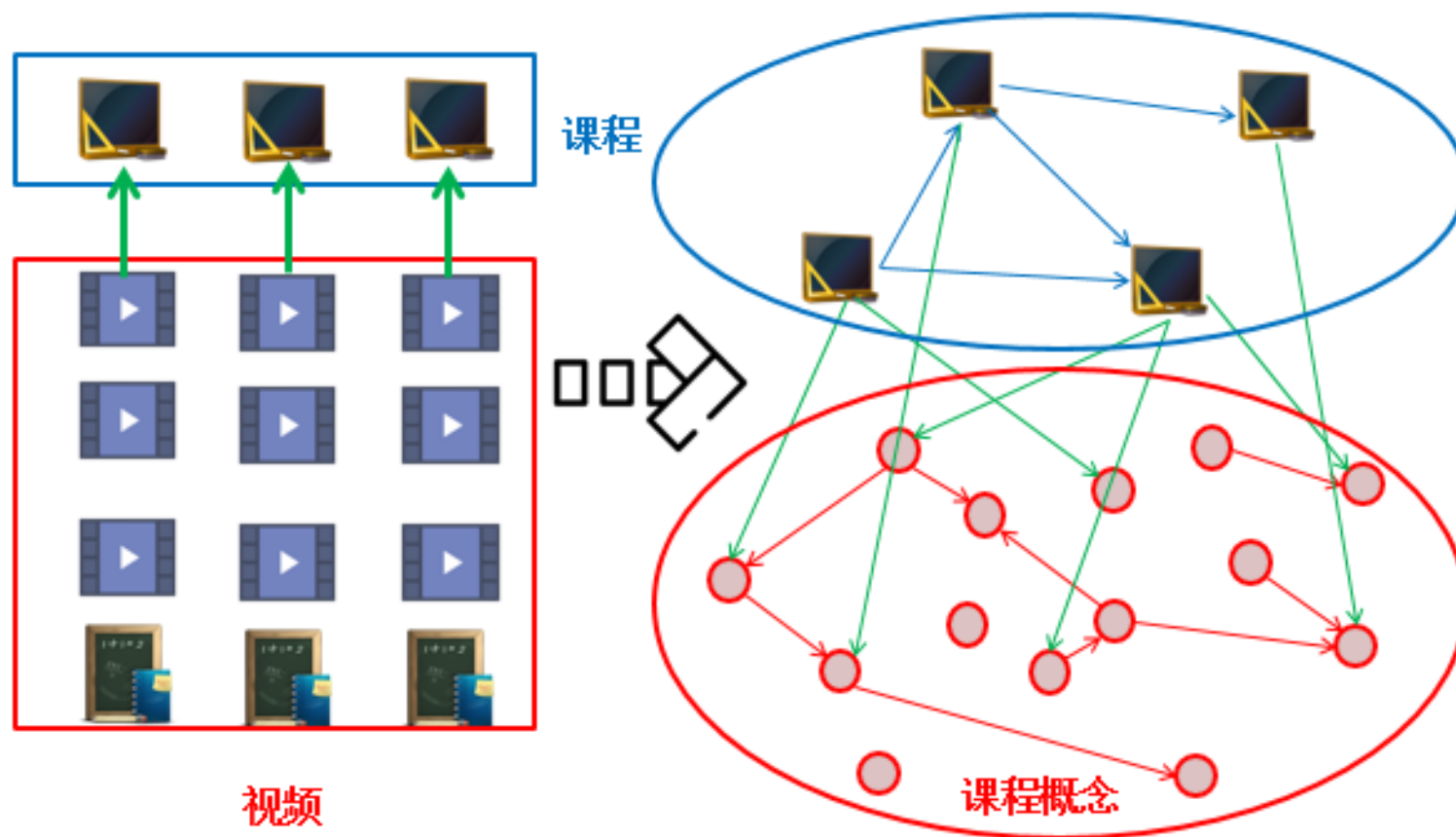
以知识为核心的在线教育资源组织

Knowledge centered online education resource organization



- 从“课程”为核心 到 “知识”为核心

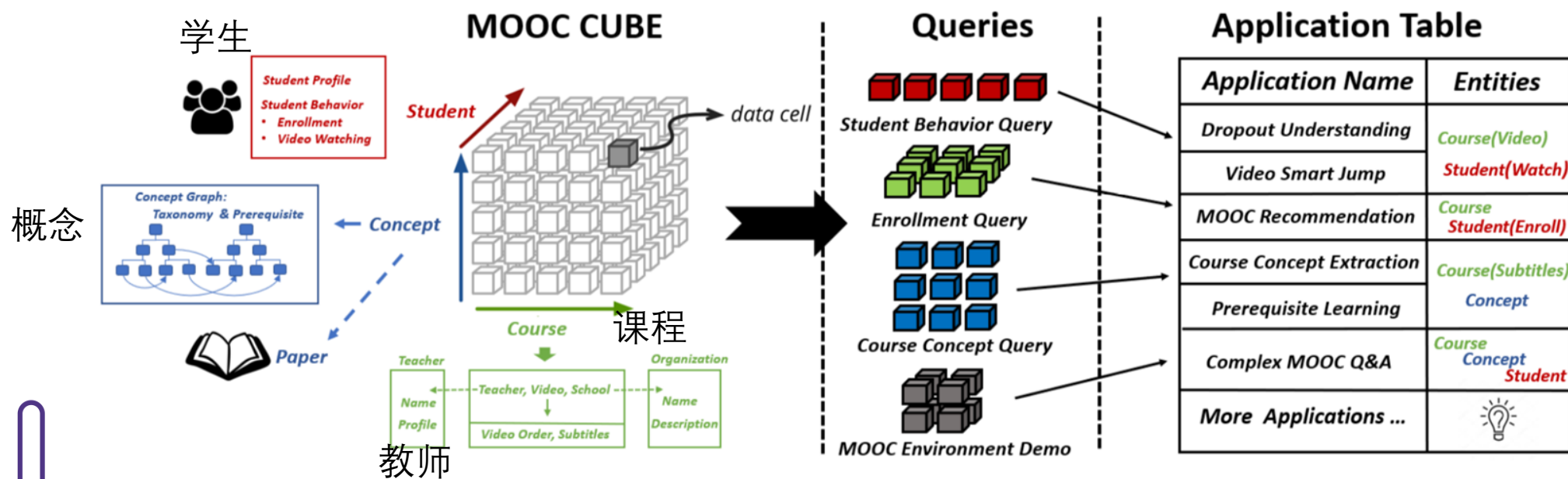
学生行为?



以知识为核心的在线教育资源组织

Knowledge centered online education resource organization

- MOOCCube: 服务于MOOC相关研究的开源大规模数据仓库
- 包括：学生行为记录，课程视频资料，知识概念体系



| 以知识为核心的在线教育资源组织

Knowledge centered online education resource organization



- MOOCCube: 细粒度学生行为记录



Student Profile
Student Behavior
• Enrollment
• Video Watching

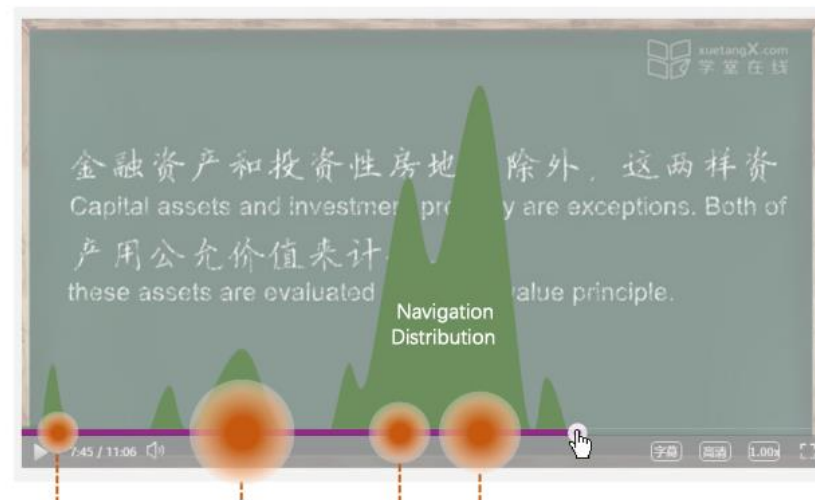
- 用户个人画像:

- 性别, 地区...

- 用户视频学习记录(精确到毫秒):

- 学习时长, 学习次数, 学习视频的区间(t_1, t_2)

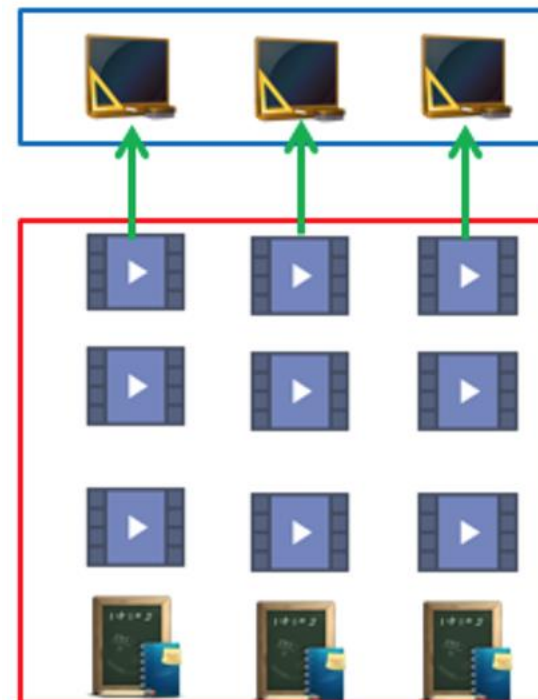
- 199,199名学生, 4,874,298人次观看视频记录



以知识为核心的在线教育资源组织

Knowledge centered online education resource organization

- MOOCCube: 课程视频资源
- 课程：
 - 课程介绍，授课教师，开课机构
- 视频：
 - 视频顺序，字幕文本
- 706门课程，38,181个视频



课程

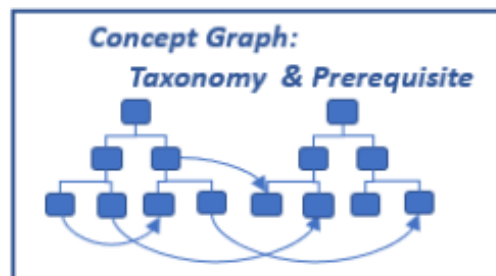
视频

| 以知识为核心的在线教育资源组织

Knowledge centered online education resource organization



- MOOCCube: 知识概念体系
- 课程概念集
- 概念先后修关系
- 概念的相关资源(论文集)
- 106,056个知识概念，并被进一步补充



在线教育知识体系构建

Construction of online education knowledge system



- 课程概念 + 先后序关系  知识体系

定义

- 课程视频中所教授的知识概念

例子

You might learn how to write a **bubble sort** and learn why a **bubble sort** is not as good as a **heapsort**. Next, we are going to talk about the **quick sort** algorithm. **Quicksort** is an algorithm invented in the 1960s by doctor Tony Hoare. It is also called the **partition exchange sort**, and is a typical algorithm based on **divide-and-conquer**.

.....

Now we have the first version of **Q sort**. After we make an analysis on its performance, performance, we will find that **quicksort** is an **unstable sorting algorithm**. Fortunately, the **quick sort** has an average **time complexity** of $n \log n$, and in most cases, it can achieve its optimal performance. We first estimate its performance under **independent uniform distribution**.

在线教育知识体系构建

Construction of online education knowledge system

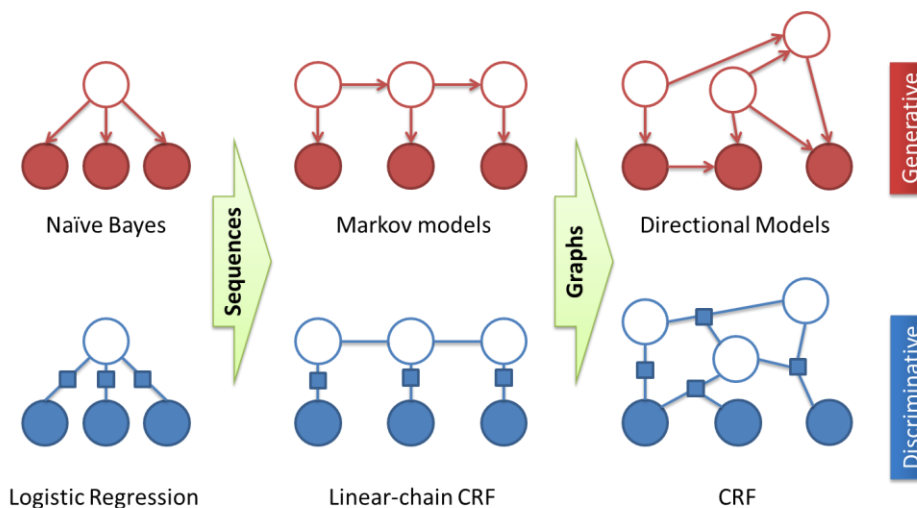


课程概念 + 先后序关系 知识体系

定义

➤ 知识概念间的学习依赖关系

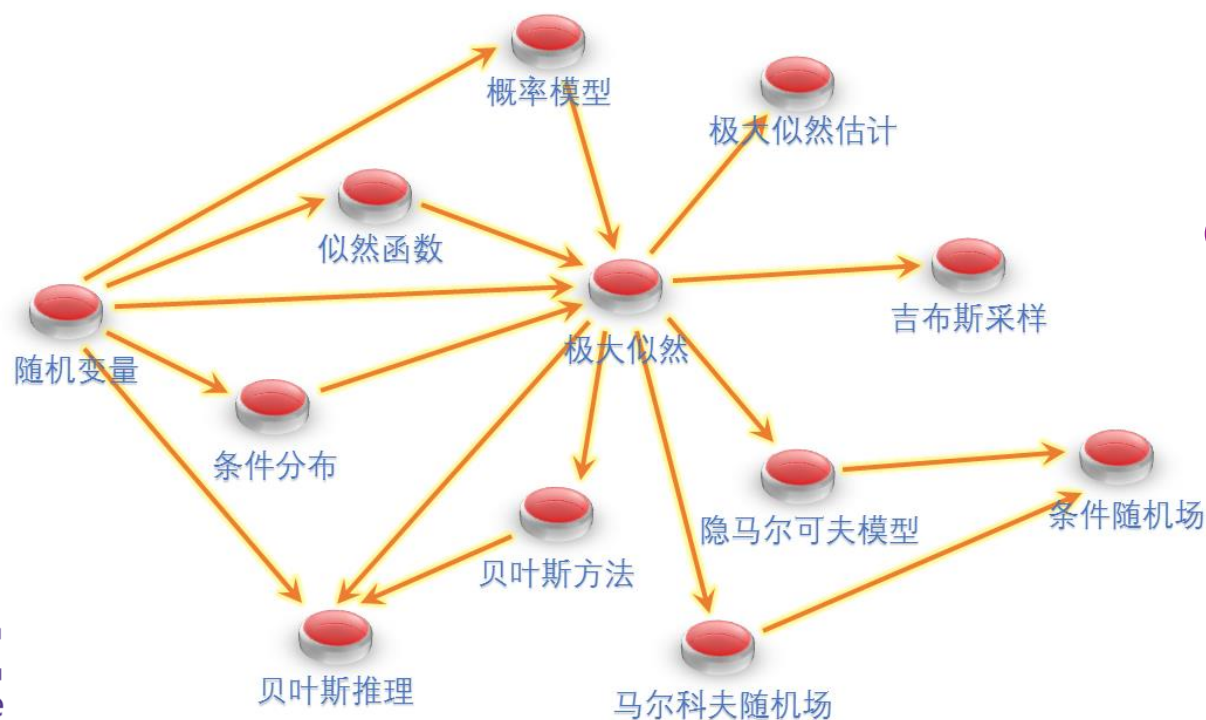
例子



在线教育知识体系构建

Construction of online education knowledge system

- 课程概念 + 先后序关系  知识体系



课程知识体系

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取
 - 从哪里获取?
 - 如何保证质量?
- 先后修关系挖掘

课程概念的三个特性

短语性

- 课程概念应该是一个**语法正确的词组或短语**

信息性

- 课程概念在语义上表示一个**科学或技术术语**

相关性

- 课程概念应该与MOOC课程语料中的**课程相关**

在线教育知识体系构建

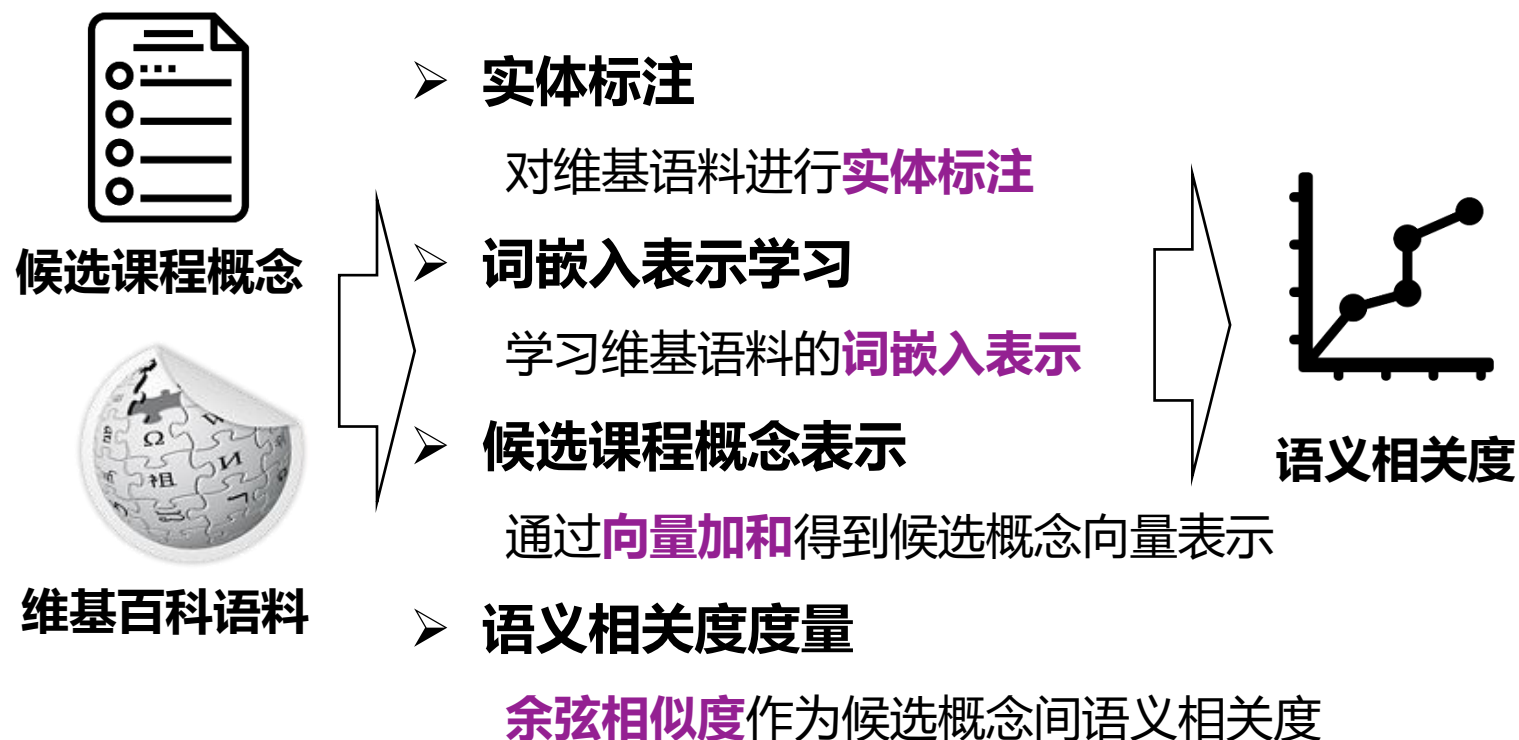
Construction of online education knowledge system



- 课程概念获取(来自课程字幕)

- 分词+组合
- 词向量训练

- 先后修关系挖掘



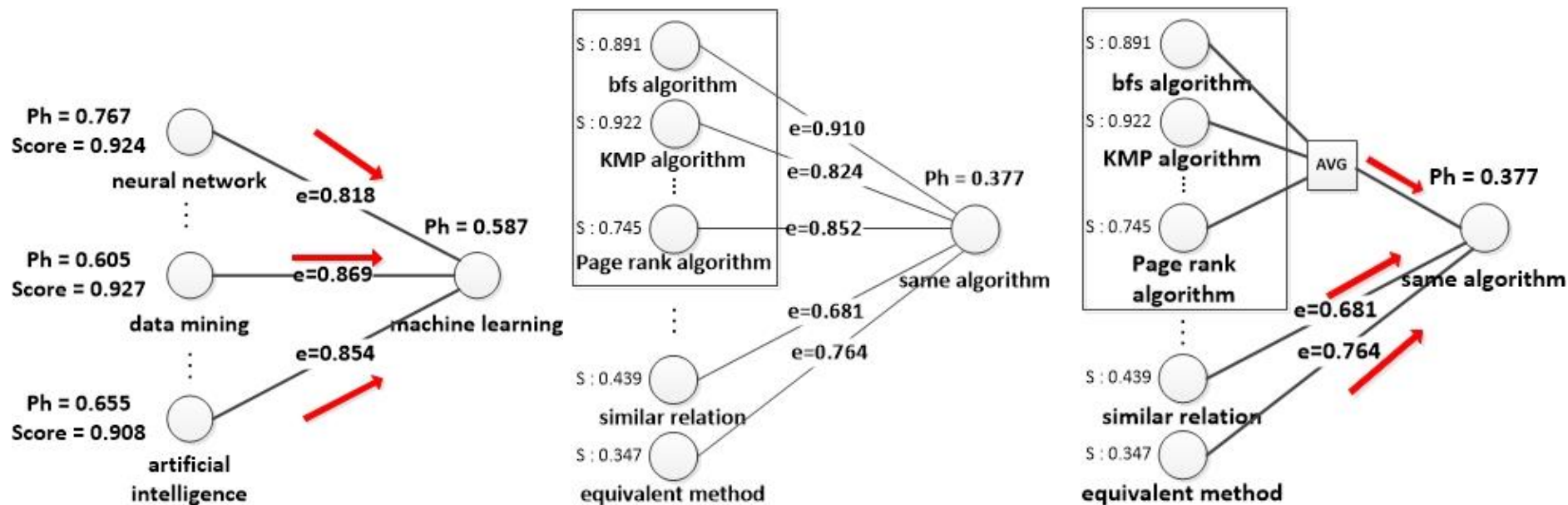
在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取
- 先后修关系挖掘

课程概念图中，课程概念与其他课程概念以高语义相关度相连接
可以通过少数种子概念通过置信度传播找出更多课程概念



在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取
- 先后修关系挖掘

模型表现:

方法	评价指标	数据集			
		CSEN	EcoEN	CSZH	EcoZH
TF-IDF	R_p	0.125	0.303	0.118	0.198
	MAP	0.105	0.232	0.109	0.145
PMI	R_p	0.239	0.222	0.246	0.179
	MAP	0.141	0.197	0.187	0.121
TextRank	R_p	0.151	0.290	0.142	0.161
	MAP	0.137	0.263	0.131	0.115
TPR	R_p	0.284	0.414	0.305	0.303
	MAP	0.255	0.387	0.267	0.288
CCP	R_p	0.443	0.427	0.434	0.435
	MAP	0.432	0.365	0.416	0.423

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取(来自外部知识)
 - 基于已有课程概念
 - 概念扩展
- 先后修关系挖掘

Course subtitles

... there should be data structures that would give it to you; probably a **binary search tree** is the first thing you'd want to consider...

Hidden related concepts

Heap **Tango Tree**
Sorting Algorithms
Priority Queue

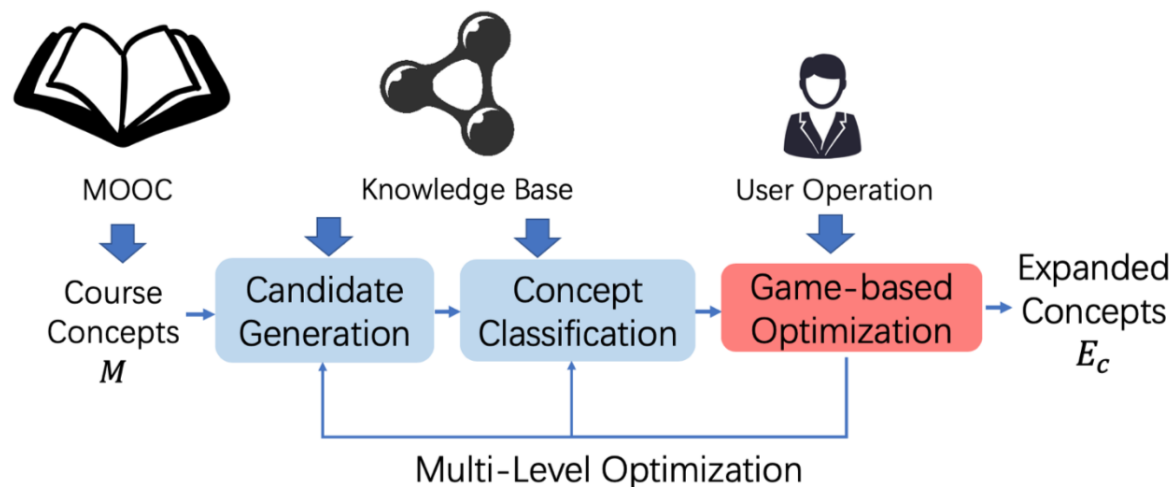
大量课程相关概念并不出现在课程文本中

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取(来自外部知识)
 - 基于已有课程概念
 - 概念扩展
- 先后修关系挖掘



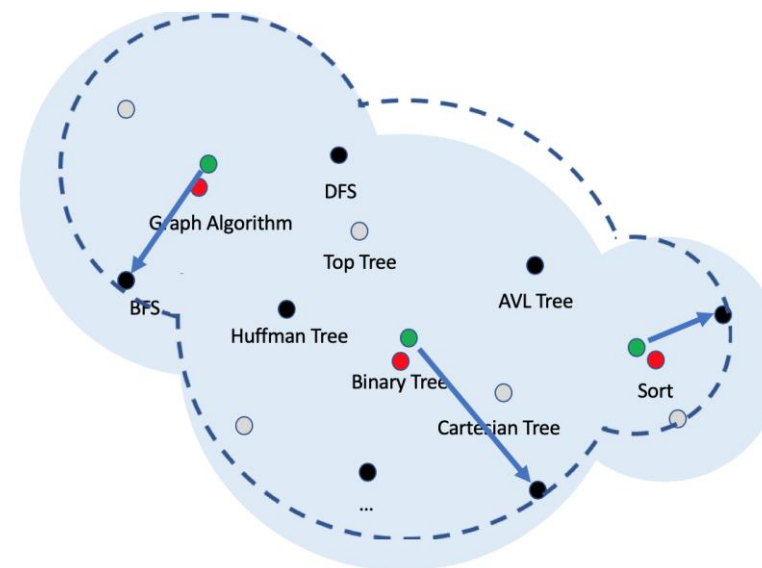
利用外部知识库等资源，进行概念扩展补充

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取(来自外部知识)
 - 语义漂移?
 - 与课程无关的术语?
- 先后修关系挖掘



课程概念空间

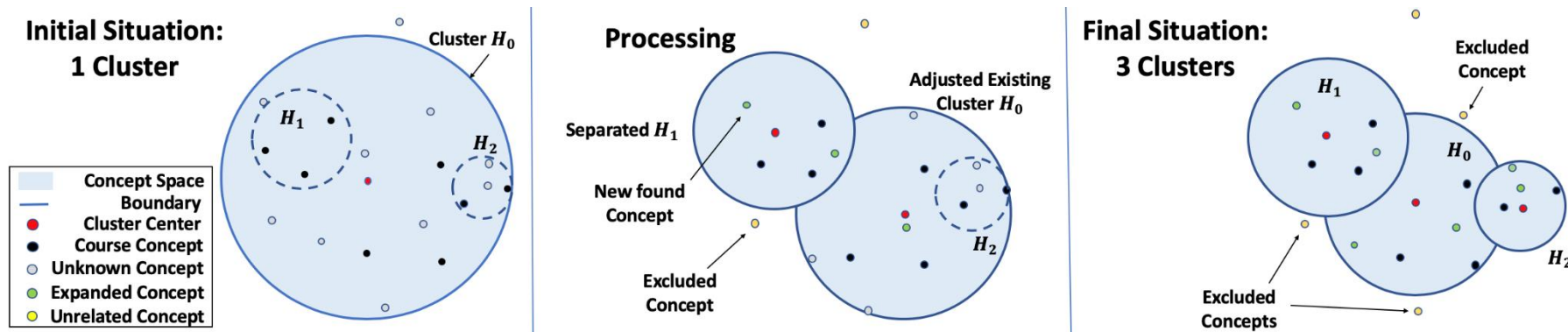
E.g. “数据结构”课主要内容：
排序算法，二叉树，图

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取(来自外部知识)
 - 语义漂移
 - 与课程无关的术语
- 先后修关系挖掘



基于动态聚类，构建课程概念空间

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取(来自外部知识)
 - 引入用户交互
- 先后修关系挖掘

The screenshot displays an online education platform interface. On the left is a sidebar menu with chapters: 第十章: 个体差异 (Chapter 10: Individual Differences), 第十一章: 学习与记忆 (Chapter 11: Learning and Memory), 第十二章: 积极心理学 (Chapter 12: Positive Psychology), and 期末考试 (Final Exam). The main content area features a video player at the top. Below the video is a section titled '遇到疑问, 小木来帮忙! 点击下方知识点, 查看解答' (Have a question, Little Mu will help! Click the knowledge points below to view the answer). This section includes a horizontal navigation bar with tabs: 心理 (Psychology), 智力 (Intelligence), 言语 (Verbal), 智力测验 (Intelligence Test), 智商 (IQ), 心理年龄 (Psychological Age), 心理能力 (Psychological Ability), and 记忆 (Memory). Below this is another row of tabs: 智力超群 (Intelligence Outstanding), 社会阶层 (Social Class), 社会 (Society), 文化 (Culture), 本质 (Essence), and 教育 (Education). The '学霸模式' (Academic Mode) is turned ON. A game card for '倔强青铜' (Stubborn Bronze) shows a score of 6 and a ranking of 8/8. To the right is a knowledge map with '心理' (Psychology) at the center, connected to various sub-topics: 心理动力学 (Psychodynamic), 心理测验 (Psychological Test), 心理健康 (Mental Health), 心理病理 (Psychopathology), 心理年龄 (Psychological Age), 心理防卫机制 (Psychological Defense Mechanism), 潜意识心理学 (Psychoanalytic Psychology), and 认知神经心理学 (Cognitive Neuropsychology). A legend at the bottom right indicates that orange dots represent '相关知识点' (Related Knowledge Points).

“学霸模式”小游戏，放置于视频下，
吸引用户对模型进行优化

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取(来自外部知识)
- 先后修关系挖掘

模型表现:

	ZH-PSY	ZH-DSA	ZH-Avg	EN-PSY	EN-DSA	EN-Avg
PR	0.849	0.519	0.684	0.833	0.822	0.827
SEISA	0.877	0.448	0.663	0.814	0.890	0.852
EBM	0.875	0.421	0.648	0.777	0.858	0.817
PUL	0.855	0.680	0.768	0.734	0.795	0.765
MOOC	0.894	0.785	0.839	0.781	0.933	0.857
MOOC-C	0.939	0.804	0.872	0.922	0.976	0.949
MOOC-CG	0.954	0.819	0.886	-	-	-

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取
- 先后修关系挖掘
 - 非事实型关系
 - 如何建模?



提及

梯度下降



~~提及~~

后向传播算法

先后修关系的核心：不对称性



在线教育知识体系构建

Construction of online education knowledge system

- 课程概念获取
- 先后修关系挖掘
 - 多种特征
 - 基于图? 更多资源?

语义特征



上下文特征



结构特征



利用MOOC字幕, 视频顺序, 课程依赖等多种信息

在线教育知识体系构建

Construction of online education knowledge system



- 课程概念获取
- 先后修关系挖掘
 - 多种特征
 - 基于图? 更多资源?

模型表现:

方法	评价指标	ML	DSA	CAL	W-ML	W-DSA	W-CAL
HPM	P	67.3	71.4	69.5	79.9	72.3	73.5
	R	18.4	14.8	16.5	25.5	27.3	23.3
	F_1	29.0	24.5	26.7	38.6	39.6	35.4
RD	P	–	–	–	73.4	77.8	74.4
	R	–	–	–	42.8	44.8	43.1
	F_1	–	–	–	54.1	56.8	54.6
T-SRI	P	61.4	62.3	62.5	58.1	60.1	62.7
	R	62.9	64.6	65.5	67.6	65.3	67.9
	F_1	62.1	63.4	64.0	62.5	62.6	65.2
F-SRI	P	–	–	–	64.3	64.3	64.8
	R	–	–	–	62.1	65.6	65.2
	F_1	–	–	–	63.2	64.9	65.0
MOOC	P	71.4	72.7	70.3	72.8	68.4	71.4
	R	73.8	72.3	71.9	71.3	72.0	70.8
	F_1	72.6	72.5	71.1	72.0	70.2	71.1

| 以知识为核心的在线教育资源组织

Knowledge centered online education resource organization



- 优势：
 - 更加利于教育资源的整合、建模，从而支持智能教育应用
- 难点：
 - 概念的特殊性：可能是实体，也可能是已有知识图谱中未出现过的实体短语
 - 课程概念的模糊性：“指针数组”是否是《数据结构》的概念？“指针”呢？
- 方向：充分利用外部资源和已有课程资源，结合教学，学习行为，优化模型

| 大规模在线教育中的知识智能

Knowledge-driven Online Education



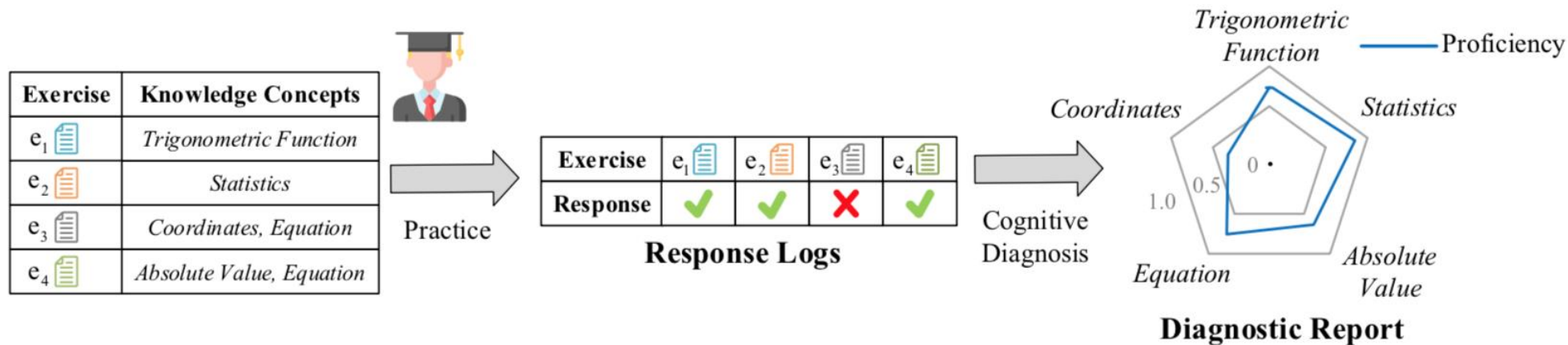
- 大规模在线课程MOOC
- 以知识为核心的在线教育资源组织
- 在线教育智能交互的知识服务



大规模在线教育中的知识智能

Knowledge-driven Online Education

- 学生 $\Rightarrow$ 学生-知识
- 学生的知识追踪/认知诊断
 - 任务描述：根据学生的行为，做题结果，建模学生的知识掌握情况

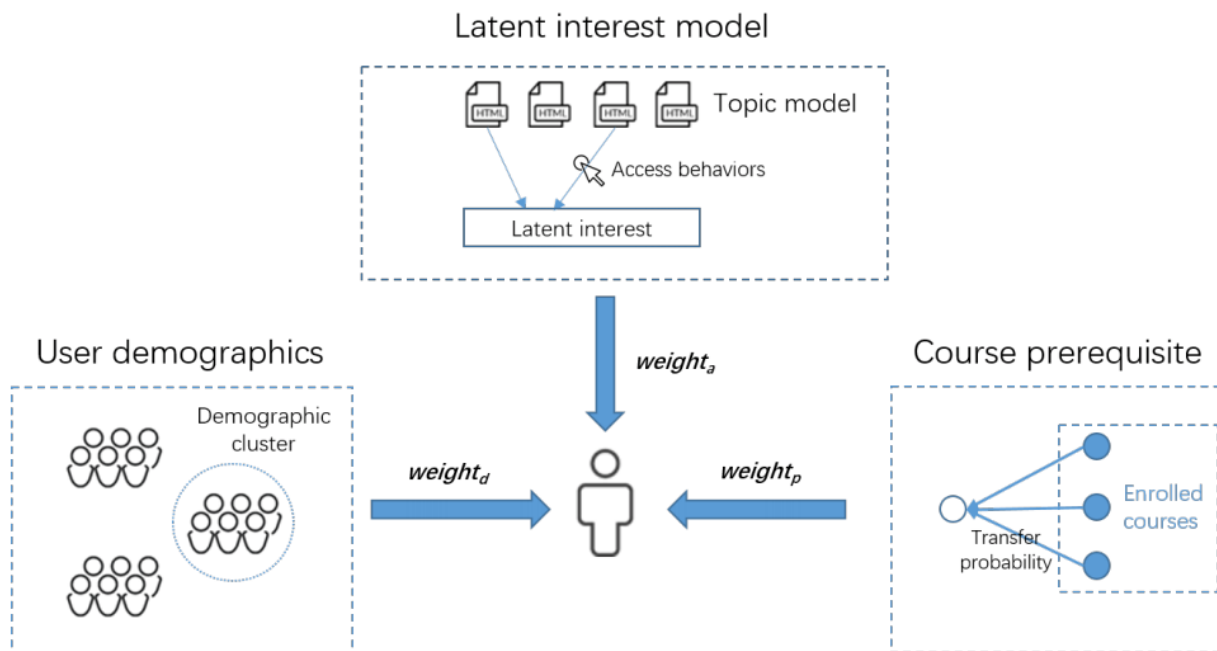


大规模在线教育中的知识智能

Knowledge-driven Online Education



- 课程-学生 $\Rightarrow$ 课程-知识-学生
- 知识图谱增强的课程/视频/教育资源推荐
 - 任务描述：根据学生的学习目标与学习掌握情况，学习兴趣，推荐相关的教育资源

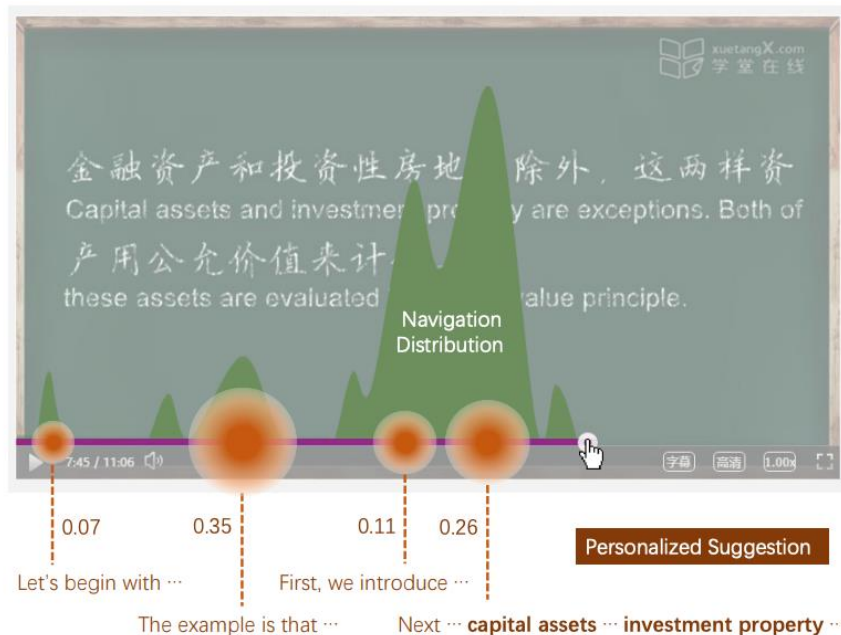


大规模在线教育中的知识智能

Knowledge-driven Online Education



- 课程视频-学生 $\implies$ 视频-知识-学生
- 智能视频导航
 - 任务描述：利用学生行为，以及课程知识分析，自动建议重要学习位置

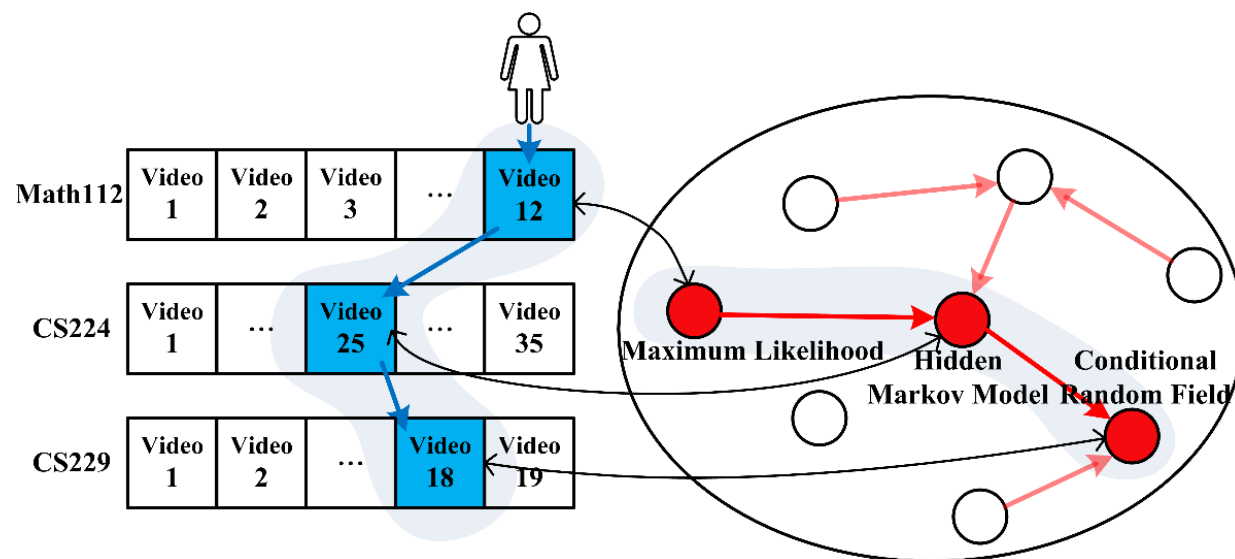


大规模在线教育中的知识智能

Knowledge-driven Online Education



- 学生认知诊断 + 教育资源推荐 + 智能视频导航 + 更多知识
- 个性化智能学习助手/学习路径规划
 - 任务描述：综合考虑学生的需求、兴趣，课程的内容以及知识的依赖，完成智能教学辅助



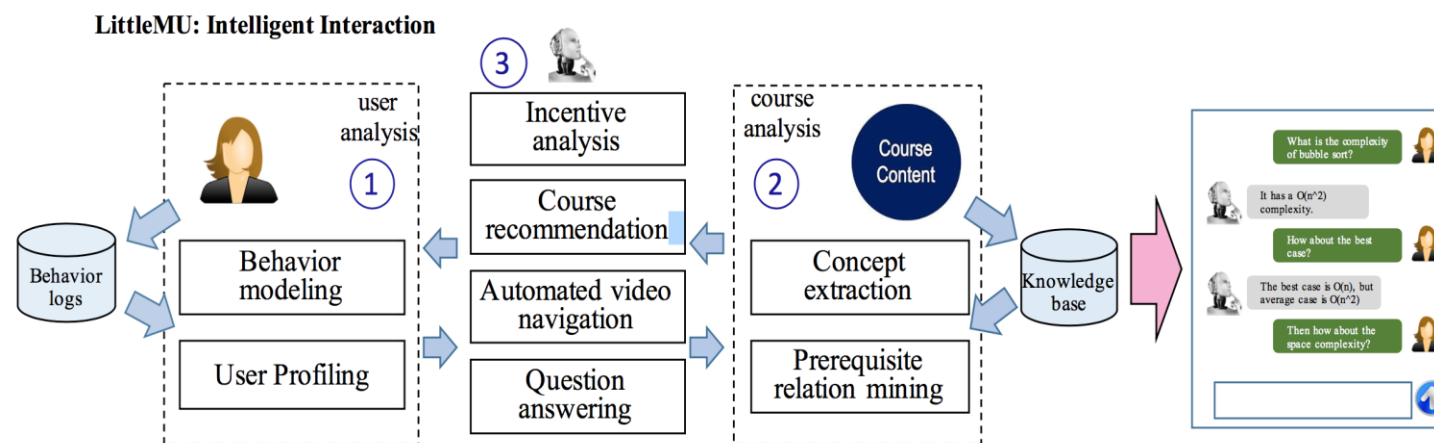
大规模在线教育中的知识智能

Knowledge-driven Online Education



• 智能助教机器人 小木(Little Mu)

- 自动答疑
- 内容推荐
- 主动干预
- 甚至可以“开玩笑”



大规模在线教育中的知识智能

Knowledge-driven Online Education



- 智能助教机器人 小木(Little Mu)

- 自动答疑



二叉树；英文名为：binary tree；树的一种情况。可以递归地定义为：该树或者为空，或者是由一个根节点和最多两棵互不相交的称为左子树或右子树的二叉树组成。属于[计算机科学技术](#)领域，其先修概念有[B树](#)等。概念延伸：[研究二叉树的学者](#)。

什么是二叉树

- 主动干预



| 大规模在线教育中的知识智能

Knowledge-driven Online Education



- 帮助学生找到“志趣”相投的同学？
- 根据知识的分布以及学生反馈，为老师提供建议？
- 自动生成思考题，激发学生兴趣？
- ...
- 更多的研究任务值得挖掘。 欢迎联系我们，

- 我们的团队主页，包含相关论文与数据集：<http://moocdata.cn/>



Q & A