

# 《知识图谱：概念与技术》

## 基于知识图谱的搜索与推荐

---

# 课程通知

---

- 最终有41位同学选了本课程。上台报告形式：10分钟PPT展示（不得超时，严格计时）。
- 12–16周课堂时间进行上台最新报告展示，报告范围是最近三年的ISWC和ESWC的论文。
- 请报名的于5月5日（含）前将报告题目发送至我研究生的邮箱553101282@qq.com。邮件标题格式为"[报告选题]-[姓名]-[学号]"
- 汇报顺序：<https://bnu05pp.github.io/KGGraduateCourse/presentation.html>

# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 基于知识图谱的搜索

- 互联网搜索的进化



传统搜索引擎的搜索结果是什么？

搜索用户到底想搜什么内容？

除了返回姚明的信息还能提供其他信息么？

# 基于知识图谱的搜索

## • 搜索的演化



The image shows a side-by-side comparison of Google and Baidu search results for the query "姚明的身高" (Yao Ming's height).

**Google Results:**

- Search bar: 姚明的身高
- Results count: 找到约 5,220,000 条结果 (用时 0.49 秒)
- Summary card: 姚明 / 身高 (姚明, 7' 6")
- Text: 7' 6"
- Text: 直接显示答案
- Related searches: 叶莉 6' 6", 沙奎尔·奥尼尔 7' 1", 林书豪 6' 3"
- Footnote: 姚明- 维基百科, 自由的百科全书
- Footnote: <https://zh.wikipedia.org/zh-hans/姚明>
- Footnote: 跳到身高争议 - 以前在中国CBA的秩序册上, 姚明的官方身高是2米26 (7英尺5英寸), 而到了美国后, 姚明的身高变成了2米29 (7英尺6英寸), 这是因为NBA球员 ...

**Baidu Results:**

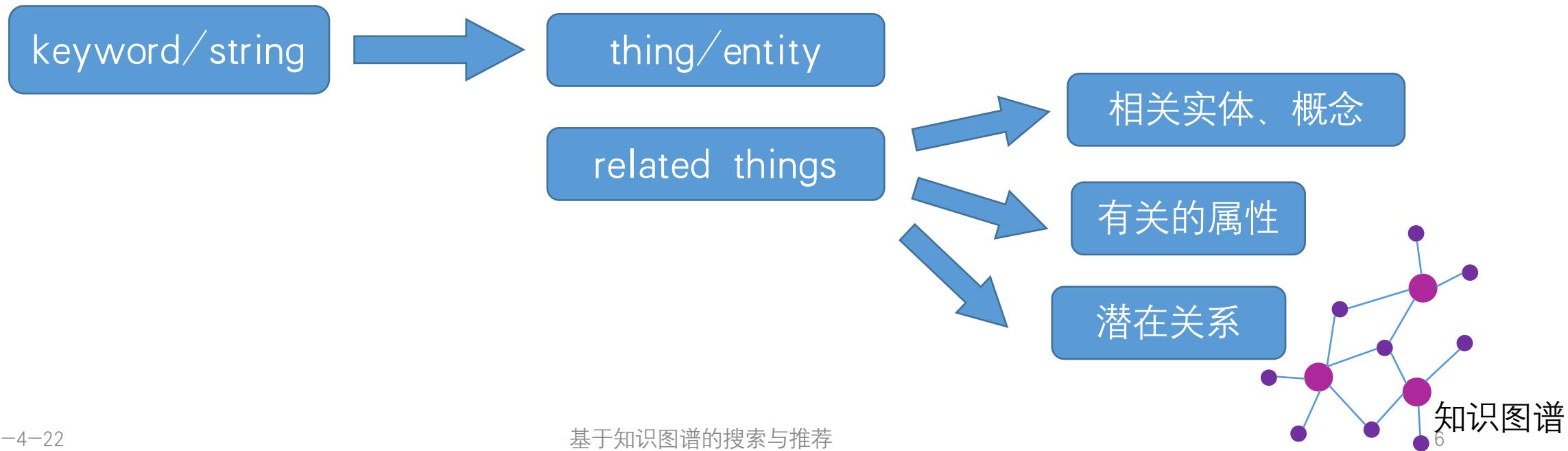
- Search bar: 姚明的身高
- Results count: 百度为您找到相关结果约3,080,000个
- Summary card: 姚明 (姚明, 篮球运动员, 226cm)
- Text: 姚明身高: 226cm
- Text: 姚明 (Yao Ming), 1980年9月12日出生于上海市徐汇区, 祖籍江苏省苏州市吴江区震泽镇, 前中国职业篮球运动员, 司职中锋, 现任中职联公司董事长兼总经理。1998年4月, 姚明... 详情>>
- Text: 来自百度百科
- Section: 知名体育人物 (巨人卡里, 梁天云, 郭艾伦, 韩德君, 科比)
- Section: 篮球名将 (叶莉, 科比, 韩德君)

# 基于知识图谱的搜索

- 搜索的演化

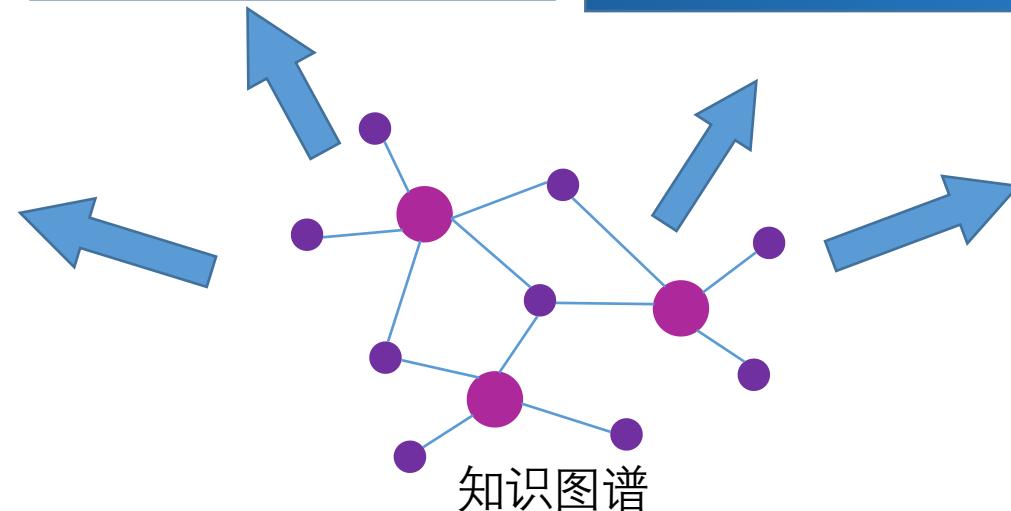


# “things, not strings”



# 基于知识图谱的搜索

## • 基于实体的搜索流程



# 基于知识图谱的搜索

- 基于实体的搜索流程



- 准确捕获用户的搜索意图（目标）是前提
- 如果目标实体不唯一，还需将目标实体排序后再返回给用户
- 如果无法明确目标实体，只能寻找一些相关实体，同样需要实体排序
- 从改善用户体验考虑，搜索结果除了展现目标实体外，最好还能以合适的形式展现相关实体、概念，以及它们之间的关系

# 内容提纲

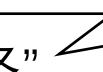
---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - **搜索意图理解**
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 搜索意图理解

- 何谓搜索意图?
  - 用户查询中包含的意图、查询目标和动机等，也有学者将其界定为用户与搜索引擎进行查询交互的过程中所表达出来的情感、认知与情景的目标，目前还没有标准的定义
- 搜索意图理解的挑战
  - 歧义搜索意图的理解，例如：查询“苹果 价格”

到底是哪个“苹果”实体?


- 基于搜索关键词的意图理解可视为根据查询文本形成用户意图“**理解**”的过程
- 根据“理解”的形式可分为：
  - 主题分类
  - 语法解析
  - 语义解析

# 搜索意图理解

---

- 主题分类
  - 对候选的实体类别进行排序来求解实体搜索的主题分类
- 语法解析
  - 输出查询中关键词的词性标注或者语法修饰关系
    - 通过分词、词性标注、词语序列标注、命名实体识别、概念标记等NLP技术抽取出查询文本中的term、mention、实体/概念名称、词性等
    - 基于知识图谱来设计模板，以确定各关键词之间的语法修饰关系
    - 例如： popular<sub>[modifier]</sub> smart cover<sub>[head]</sub> iPhone X<sub>[constraint]</sub>
- 语义解析
  - 输出查询中关键词对应的语义角色，比语法解析更加关注查询关键词之间的语义关联
  - 其中一个重要任务就是查询关键词中的实体理解，包括实体识别与实体链接

# 搜索意图理解

- 实体链接
  - 为什么是一个挑战?
  - 同一个实体在广泛的文本中可能有多个指代
    - Barack Obama
    - Barack H. Obama
    - President Obama
    - Senator Obama
    - President of the United States
  - 同一个指代可能指代多个不同实体
    - Michael Jordan 到底是指篮球巨星还是机器学习大牛?
    - “Apple”是指能吃的水果还是时尚的公司/手机? “Big Apple”是水果么?



都是指美国前任总统奥巴马

# 搜索意图理解

- 短文本实体链接的算法目标<sup>[1]</sup>

- 利用实体指代词 $m$ 与候选实体 $e$ 的上下文等相关特征计算两者的匹配度分数 $\varphi(e, m)$ ，按分数进行排序，并选择分数最大的实体 $e_{best}$ 作为 $m$ 的链接结果，即

$$e_{best} = \operatorname{argm}_{e} \operatorname{ax} \varphi(e, m)$$

- 局部模型

- 为短文本中的每个指代词及其链接的实体单独计算 $\varphi(e, m)$ ，每个链接实体都是独立产生

- 全局模型

- 考虑文本中多个指代词所链接的实体间联系，对上下文内所有歧义的实体指代一同消歧

- 令 $\Gamma = \{(m_1, e_1), (m_2, e_2) \dots\}$ 为一个全局实体链接方案，则目标函数为：

$$\Gamma_{best} = \operatorname{argm}_{\Gamma} \operatorname{ax}_{\Gamma} O(\Gamma) = \operatorname{argm}_{\Gamma} \operatorname{ax}_{\Gamma} \sum_{i=1}^N (\varphi(m_i, e_i) + \sum_{t_j \in \Gamma} \psi(e_i, e_j))$$

$\psi(e_i, e_j)$ 是实体 $e_i, e_j$ 之间的相关度分数

# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - **目标查找与结果呈现**
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 目标查找

---

- 前一步骤搜索意图理解的结果作为目标实体的查询条件
- 根据查询条件生成数据库的查询语句
  - SQL或者SPARQL语句
- 或设计特定的算法从KG中找出目标实体
  - 大多是基于图的搜索算法
  - 例如：查询“上海 985高校”

# 结果呈现

---

- 两个子任务
  - 结果排序（目标实体不唯一时）
  - 内容的分类与组织
- 实体排序的依据
  - 在KG网络结构中的重要性，可利用PageRank等算法
  - 实体流行度
  - 与查询的相关性
    - 可利用前一章介绍的计算  $p(e|c) = \frac{n(e,c)}{\sum_{e_i} n(e_i,c)}$ ，即针对给定的查询概念c，计算候选实体e的典型性（代表性）
    - 候选实体的排序分值可计算为： $p(e)p(e|c)$ ，为实体e的流行度

# 结果呈现

---

- 搜索结果内容的分类与组织
  - 一个目标实体可展现的内容信息可以很多，如何合理的分类呈现是提升搜索用户体验的关键
    - 参考百科词条页面的内容组织
    - 对实体属性的重要性进行排序是关键
      - 优先展现重要的属性的内容
        - 例如：为考生提供报考高校的搜索服务时，平台展现各大学实体的各类信息时如何分先后？
        - 针对高频的实体类别（概念）可由人工设定关键属性
        - 对于低频的实体类别，则可以基于用户点击日志等挖掘该类别的关键属性

# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 实体探索

---

- 目的在于拓展目标实体之外的相关内容并向用户有效地呈现（往往是展现相关实体）
- 是提升**搜索多样性**的关键
- 包括：
  - 相关实体发现
  - 实体摘要 (entity summarization)
  - 相关实体解释
  - .....

# 相关实体发现

- 不仅限于单个实体的搜索



Google 姚明的身高

全部 图片 新闻 视频 更多

找到约 5,220,000 条结果 (用时 0.49 秒)

姚明 / 身高

7' 6"

姚明 / 身高

相关实体

用户还搜索了

叶莉 6' 3" 沙奎尔·奥尼尔 7' 1" 林书豪 6' 3"

姚明 - 维基百科, 自由的百科全书  
<https://zh.wikipedia.org/zh-hans/姚明>

跳到身高争议 - 以前在中国CBA的秩序册上, 姚明的官方身高是2米26 (7英尺5英寸), 而到了美国后, 姚明的身高变成了2米29 (7英尺6英寸), 这是因为NBA球员 ...

Baidu 姚明的身高

网页 图片 知道 视频 新闻 贴吧 文库 音乐 地图 更多»

百度为您找到相关结果约3,080,000个

姚明

篮球运

姚明，市吴江中国国职业联赛篮球城”。现月，姚职业篮

生于：人民共

身高：

臂展：

配偶：

NBA選

子女：

你負責資源嗎

姚明身高：  
226cm

姚明 (Yao Ming)，1980年9月12日出生于上海市徐汇区，祖籍江苏省苏州市吴江区震泽镇，前中国职业篮球运动员，司职中锋，现任中联公司董事长兼总经理。1998年4月，姚明... [详情»](#)

来自百度百科

姚明的身高\_百度图片

姚明与叶莉、科比等人的合影

相关实体

知名体育人物 展开

巨人卡里 梁天云 郭艾伦

摔跤手演员 和健美选手 前国家少年 篮球比赛中锋 篮球运动员 翘楚

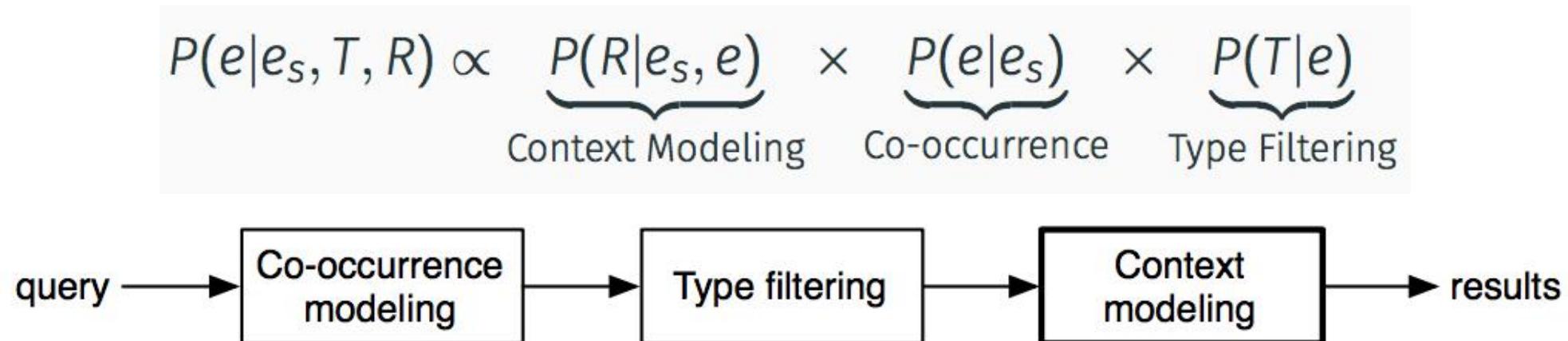
篮球名将 展开

叶莉 科比 韩德君

前中国著名 篮球运动员 NBA最有价 值球员奖 被称为CUBA 第一中锋

# 相关实体发现

- 问题定义<sup>[2]</sup>
  - 对于给定的一个实体 $e_s$ （实际上是前述步骤的目标实体），针对目标实体 $e$ （是本步骤要找出的相关实体）所属的类型 $T$  及其与给定实体的关系描述 $R$ ，为目标实体计算如下的概率，最后按照此概率对所有相关的目标实体进行排序并输出



# 相关实体发现

- 实体共现的程度

$$P(e|e_s) = \frac{f(e, e_s)}{\sum_{e' \in E} f(e', e_s)}$$

e与 $e_s$  共现的频率

- 基于类型的相关性  $P(T|e)$ 
  - 利用百科实体页面中的分类信息
  - 利用命名实体识别工具
- 基于关系的相关性

$$P(R|e, e_s) = \prod_{t \in R} P(t|\Theta_{ee_s})$$

$t$  是关系描述  $R$  中的一个词， $\Theta_{ee_s}$  是实体  $e$  与  $e_s$  的共现语言模型， $t$  越多地出现在  $e$  与  $e_s$  的共现文档集合中，则  $P(t|\Theta_{ee_s})$  越大

# 实体摘要

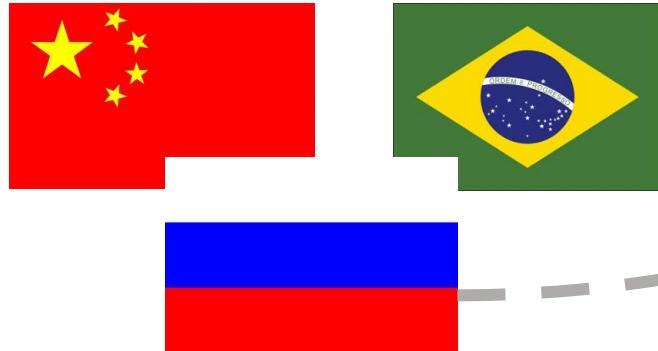
- 多种呈现形式
  - 文本式摘要
    - 例如维基百科和百度百科的Infobox
  - 图形化摘要
    - 针对相关实体进行层次化聚类<sup>[4]</sup>
    - 针对相关实体与搜索的目标实体相关性进行排序



CN-DBpedia的Curiosity实体探索界面

# 相关实体解释

- 应用实例



应该推荐什么相关实体?



# 相关实体解释

- 问题定义<sup>[3]</sup>

- 给定由一组实体代表的查询 $q$ , 不仅能推荐出与 $q$  最相关的实体 $e$ , 而且还能产生一个(组)概念 $c_i$ 能完美解释 $e$ 与 $q$  中实体间的潜在联系

您已经搜索了



¥5299.00

京品手机 Apple iPhone XR (A2108)  
128GB 白色 移动联通电信4G手机 双卡双



¥6959.00

三星 Galaxy S10+ 8GB+128GB皓玉白  
(SM-G9750) 3D超声波屏下指纹超感官



向您推荐以下手机, 因为它们都属于高端商务机



¥5488.00

华为 HUAWEI P30 Pro 超感光徕卡四摄10  
倍混合变焦麒麟980芯片屏内指纹



¥5069.00

京品手机 华为 HUAWEI Mate 20 Pro  
(UD)屏内指纹版麒麟980芯片全面屏超大

# 相关实体解释

## • 算法描述

- 寻找的概念 $c_i$ 与 $e$ 应满足下述目标

1. Probabilistic Relevance Model

$$\operatorname{argmax}_{e \in E - q} r_e(q, e) = \sum_i P(e|c_i)P(c_i|q)\delta(c_i)$$

利用Probbase发现概念  
与实体间的关系

找到的概念既要有代  
表性又要能很好地在  
**最优的粒度层级**上解  
释所推荐的相关实体

2. Relative Entropy Model

$$\operatorname{argmin}_{e \in E - q} KL(P(C|q), P(C|q, e)) = \sum_{i=1}^n \delta(c_i)P(c_i|q)\log\left(\frac{P(c_i|q)}{P(c_i|q, e)}\right)$$

# 相关实体解释

- 计算  $P(c_i | q)$

1. Naïve Bayes Model

$$P(c_i | q) = \frac{P(q|c_i)P(c_i)}{P(q)} \propto \prod_{e_j \in q} P(e_j | c_i)P(c_i)$$

$$\propto P(c_i) \prod_{e_j \in q, n(e_j, c_i) > 0} \lambda P(e_j | c_i) \prod_{e_j \in q, n(e_j, c_i) = 0} (1 - \lambda) P(e_j)$$

2. Noisy-or Model

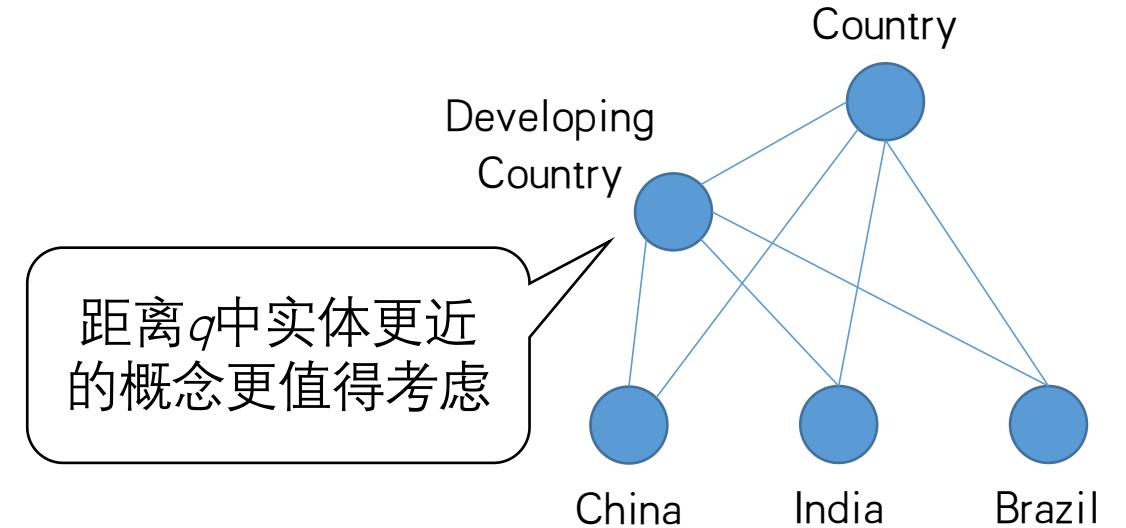
$$P(c_i | q) = 1 - \prod_{e_j \in q} (1 - P(c_i | e_j))$$

# 相关实体解释

- 计算 $\delta(c_i)$ 
  - 用于度量目标概念的粒度，好的概念既不能太一般化也不能太具体化

Concept	Number of Entities
Country	2648
Developing country	149
Growing market	18

Entity-based Approach



Hierarchy-based Approach

# 相关实体解释

---

- 计算 $\delta(c_i)$

1. Entity-based Approach

- Penalize popular concepts

- $\delta(c_i) = \frac{1}{P(c_i)}$

2. Hierarchy-based Approach (Average first passage time)

- $\operatorname{argmax}_{C_q^k} \sum_{c \in C_q^k} \sum_{q_i \in q} h(q_i | c)$

- $$\begin{cases} h(q_i | c) = 0, & \text{if } q_i = c \\ h(q_i | c) = 1 + \sum_{c' \in c(c')} P(c' | q_i) h(c' | c) & \text{if } q_i \neq c \end{cases}$$

# 参考文献

---

- [1] L. Chen, J. Liang, C. Xie and **Yanghua Xiao**. “Short Text Entity Linking with Fine-grained Topics”, in CIKM 2018.
- [2] Bron, Marc, K. Balog, and M. D. Rijke. “Ranking related entities: components and analyses”, in CIKM 2010.
- [3] Y. Zhang, **Yanghua Xiao** et al. “Entity Suggestion with Conceptual Expanation”, in IJCAI 2017.
- [4] Xie C, Chen L, Liang J, **Yanghua Xiao** et al. “Automatic Navbox Generation by Interpretable Clustering over Linked Entities”, in CIKM 2017.

# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 推荐系统的价值

## 推荐系统应用广泛

侏罗纪世界2 Jurassic World:



导演: 胡安·安东尼奥·巴亚纳  
编剧: 科林·特雷沃罗 / 德里克·康纳利  
主演: 克里斯·帕拉特 / 布莱丝·达拉斯·霍华德 / 伊莎贝拉·瑟蒙 / 拉菲·斯波  
类型: 动作 / 科幻 / 冒险  
官方网站: www.jurassicworld.com  
制片国家/地区: 美国 / 西班牙  
语言: 英语  
上映日期: 2018-06-15(中国大陆) / :  
喜欢这部电影的人也喜欢 ······



电影 (豆瓣)

推荐系统实践 平装 - 20

4.5星 94 条商品评价

立即购买

Kindle ¥12.00

使用我们的 免费Kindle阅读软件

促销信息: 买书满150元返

库存中仅剩 5 件 (更多商品正

浏览此商品的顾客也同时浏览



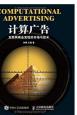
推荐系统技术、评估及高效  
算法  
弗朗西斯科·里奇 ...  
★★★★☆ 11

平装  
¥144.56



集体智慧编程  
托比·西格兰 (T...  
★★★★☆ 39

平装  
¥65.10 prime



计算广告: 互联网商业变现出的  
市场与技术  
刘鹏 王超  
★★★★☆ 19

平装  
¥57.80 prime

图书 (亚马逊)

美食 火锅 西餐 日本菜 霸王餐

全部 »



蟹的稻田屋

日本料理

人均: 574元/人



蟹的稻田屋

722条点评

人均: 574元/人



蟹的稻田屋

1440条点评

人均: 490元/人

全部 »



尚9一滴水

人均: 534元/人



尚9一滴水

4354条点评



尚9一滴水

人均: 170元/人

餐馆 (大众点评)

热门推荐 华语 | 流行 | 摆滚 | 民谣



胖友们, 又到了吃脂肪  
的季节啦!



锦鲤歌单|考研静心|美  
梦成真



没机会现场看麦豆的演  
唱会? 一张歌单过足瘾



电台节目|新世相 | 我, 斗  
战胜佛

音乐 (网易云音乐)

中国电信 中国移动 2G 3G 4G 100% 早上8:12

搜索: 日本战国名将 | 中国男足 | 意甲派系  
发布

关注 推荐 上海 视频 热点 娱乐

习近平: 全面加强新时代我军党的领导和党的建设工作



置顶 央视网新闻 53评论

“温比亚”强势北上 豫皖苏鲁多地暴雨破历史极值



热 中国新闻网 939评论 7分钟前

本月发三起猪瘟疫情 农业农村部: 正调查是否有联系



热 环球网 100评论

姚均晟神仙球韦世豪建功, 国足2-1逆转阿联酋, 多角度疯狂3球回放



首页 新闻 (今日头条) 我的

# 推荐算法概述

---

- 推荐算法的目标

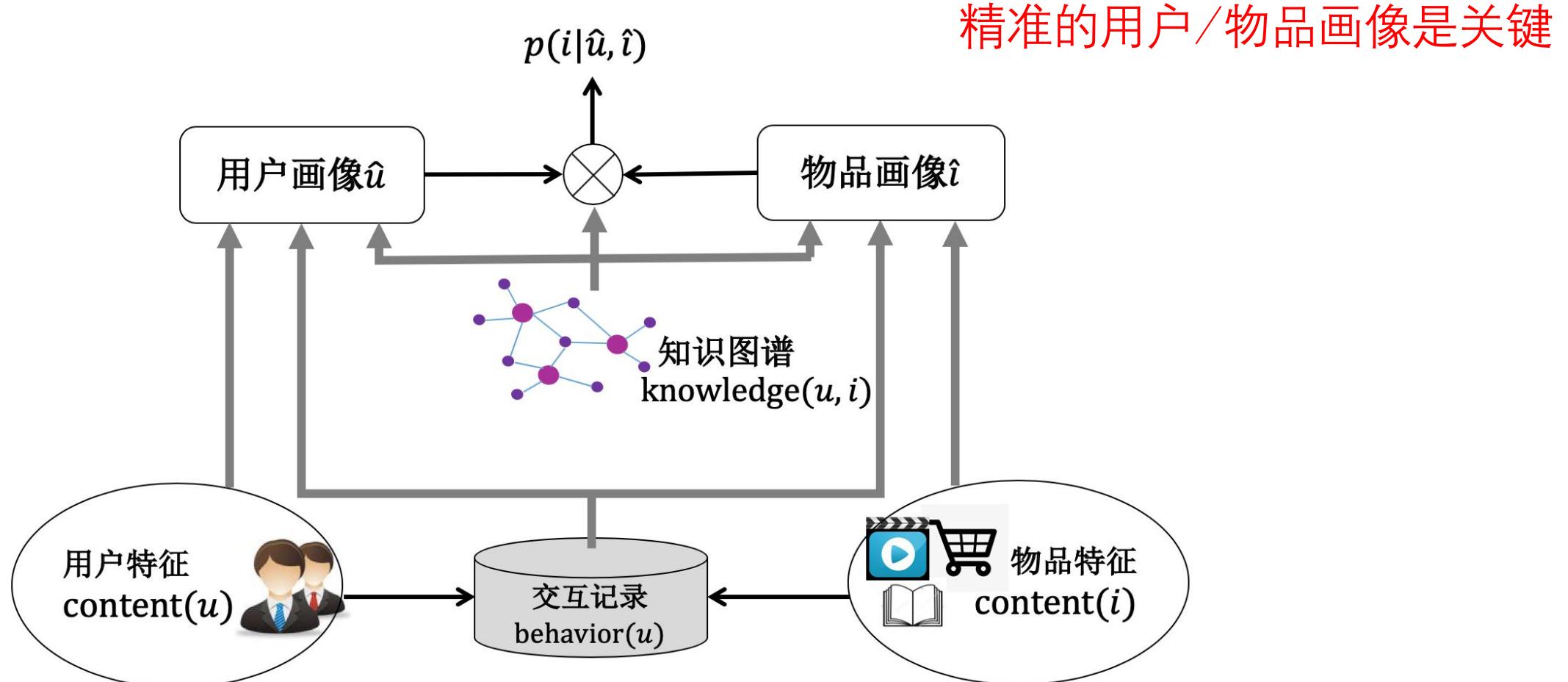
$$\underset{i \in I}{\operatorname{argmax}} p(i | u) \text{ 或者 } f: U \times I \rightarrow \mathbb{R}$$

- 推荐算法分类

- 基于协同过滤:  $p(i|u, behavior(u))$ 
  - 基于记忆
  - 基于模型
- 基于内容:  $p(i|u, content(u, i))$
- 混合方法
- 基于知识:  $p(i|u, knowledge)$

# 推荐算法概述

- 推荐算法的基本框架



# 基于知识的推荐

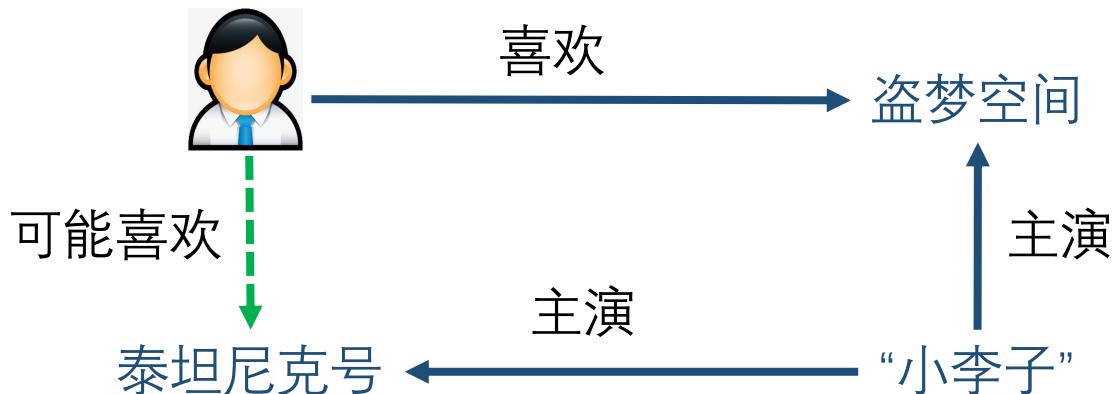
---

- 传统推荐算法的挑战
  - 基于协同过滤
    - 冷启动
    - 数据稀疏
    - 可扩展性
    - .....
  - 基于内容
    - 特征描述
    - 同义/多义词
      - “十面埋伏”是电影？小说？还是成语？
    - 结果的同质性（缺乏多样性）
    - .....

# 基于KG的推荐

推荐系统中引入知识图谱的优势：

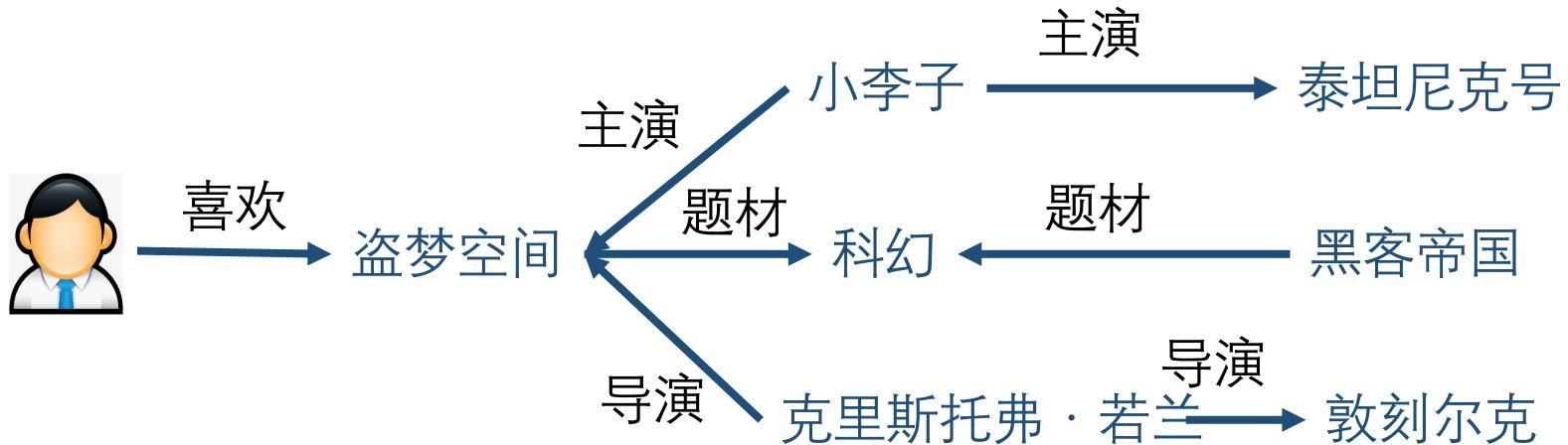
- 提高推荐的精准度
  - 知识图谱为物品引入了更多的语义关系
  - 知识图谱可以深层次地发现用户兴趣



# 基于KG的推荐

推荐系统中引入知识图谱的优势：

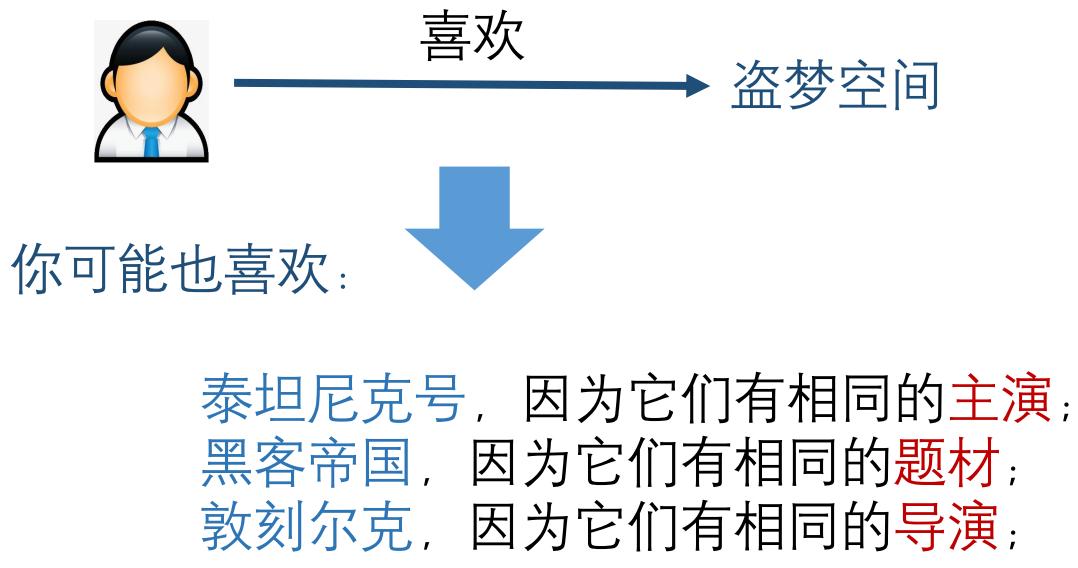
- 增加多样性
  - 知识图谱提供了物品间不同类型的关联
  - 有利于产生多样的推荐结果，避免推荐结果局限于单一类型（信息茧房）



# 基于KG的推荐

推荐系统中引入知识图谱的优势：

- 可解释性
  - 可以从知识图谱中找到用户的兴趣历史与推荐结果之间的关联
  - 提高用户对推荐结果的满意度和接受度，增强用户对推荐系统的信任



# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - **基于嵌入向量的方法**
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 基于KG的推荐方法分类<sup>[1]</sup>

---

- 基于嵌入向量的方法 (embedding-based)
- 基于路径的方法 (path-based)
- 统一/集成的方法 (unified)

# Embedding-based方法

- 从KG中抽取的知识信息来丰富用户、物品的表示（向量）
- 一般采用KG嵌入表示算法（KGE）来为KG中的每个节点生成嵌入表示向量
  - KGE包括Trans系列模型，如TransE、TransD、TransH、TransR等
  - 基于Skip-gram模型，如Node2Vec、Metapath2Vec、HIN2Vec等
- 用户的表示向量产生方式可以细分为
  - 用户节点连同用户-物品间的交互边，以及用户-用户间的社交/共同购买/共同点评等关系都作为实体间的关系一起加入KG，使得KGE能直接产生用户节点表示向量
  - 用户节点不加入KG，则用户的表示向量基于其交互过的物品表示向量生成
- 将目标用户表示向量  $\mathbf{u}_i$  和候选物品表示向量  $\mathbf{v}_j$  输入推荐框架中产生最终推荐结果

$$\hat{y}_{i,j} = f(\mathbf{u}_i, \mathbf{v}_j)$$

也可纳入关系的表示向量  $\mathbf{r}$

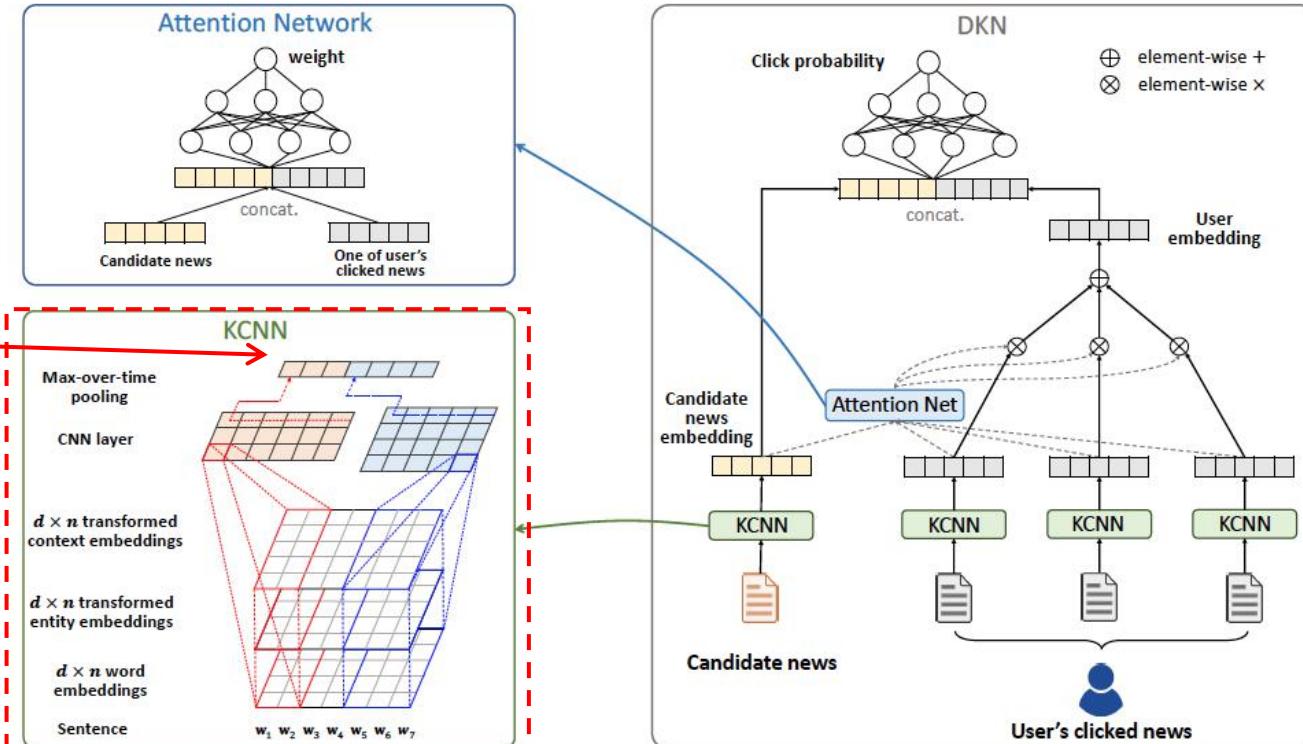
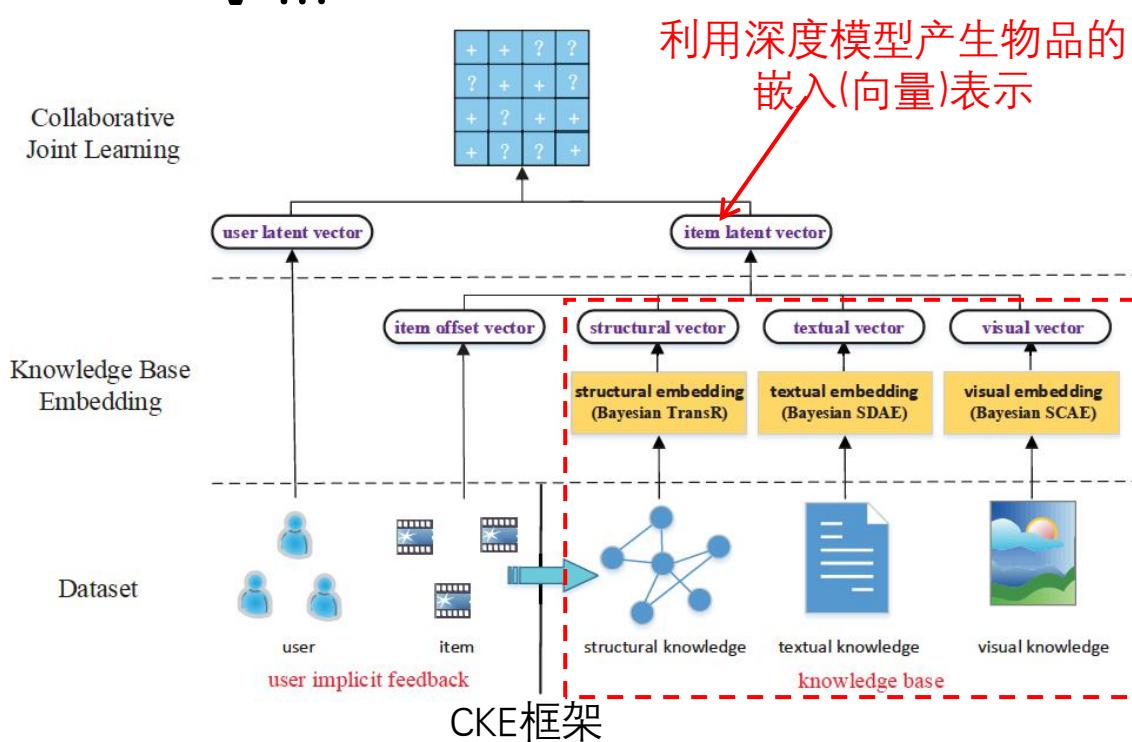
$f$  可以是向量操作函数。  
也可以是某种神经网络

$$\hat{y}_{i,j} = f(\mathbf{u}_i, \mathbf{v}_j, \mathbf{r})$$

# Embedding-based方法

- 代表模型

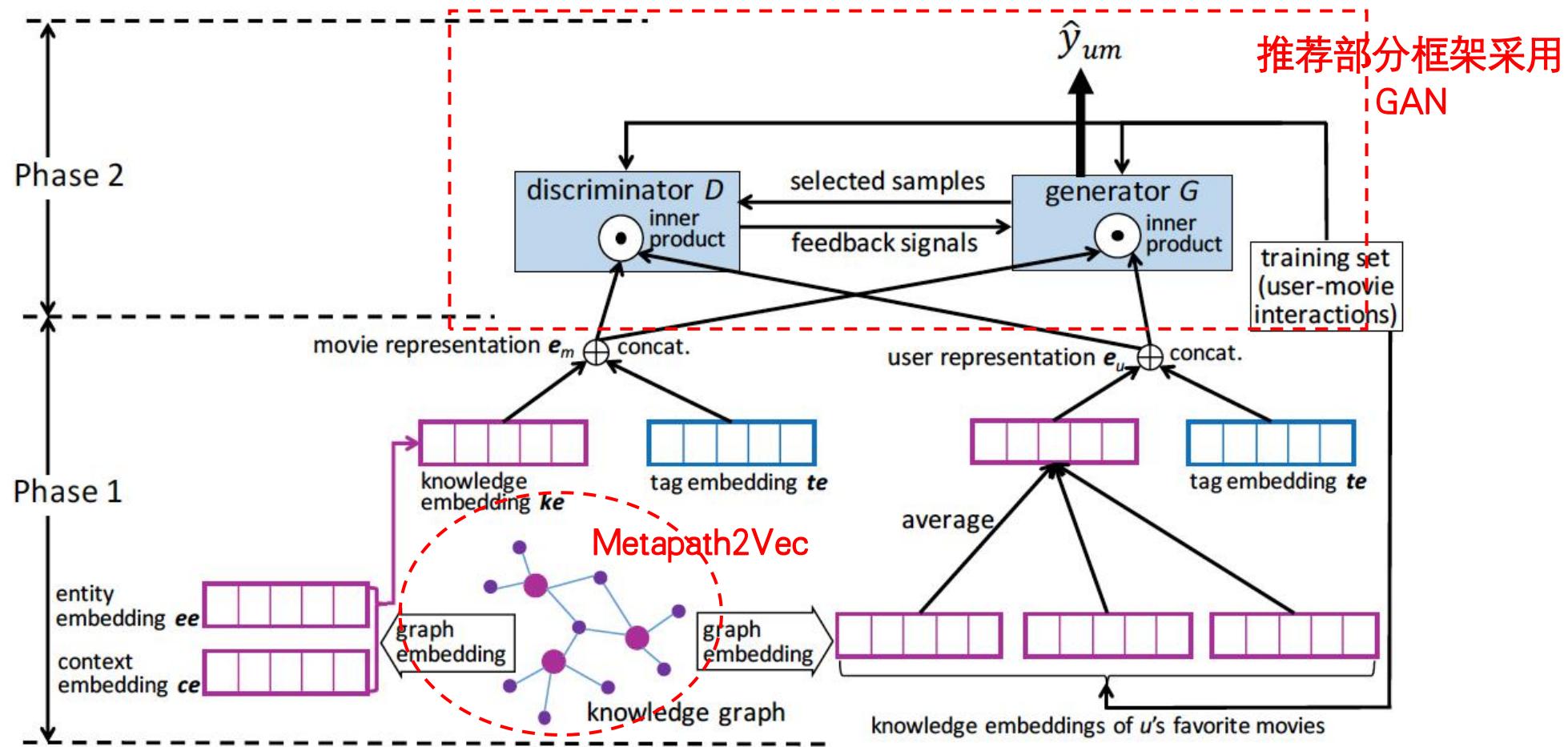
- CKE<sup>[2]</sup>
- DKN<sup>[3]</sup>
- KSR<sup>[4]</sup>
- ...



DKN框架

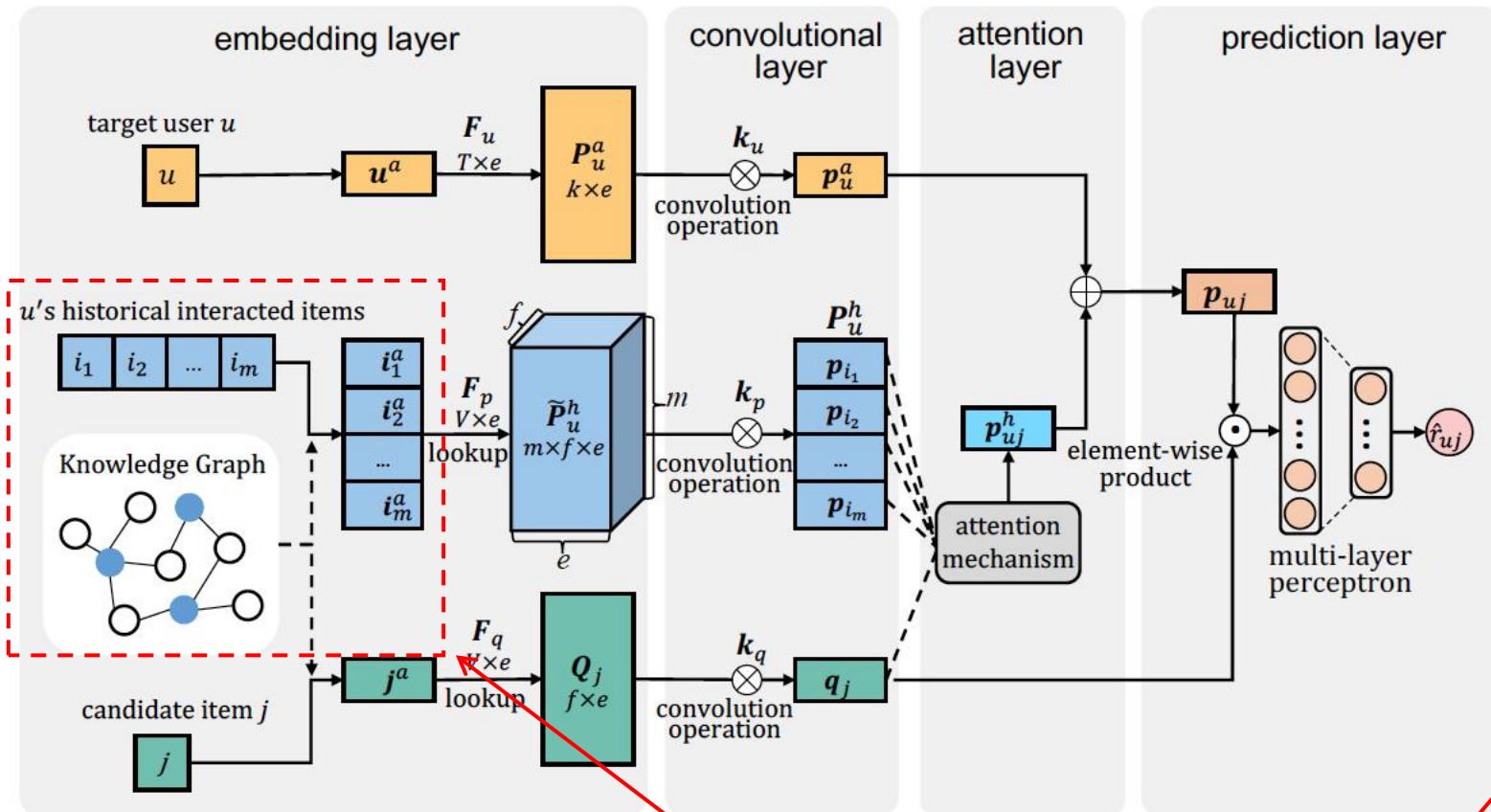
# Embedding-based方法

- 代表模型KTGAN<sup>[5]</sup>

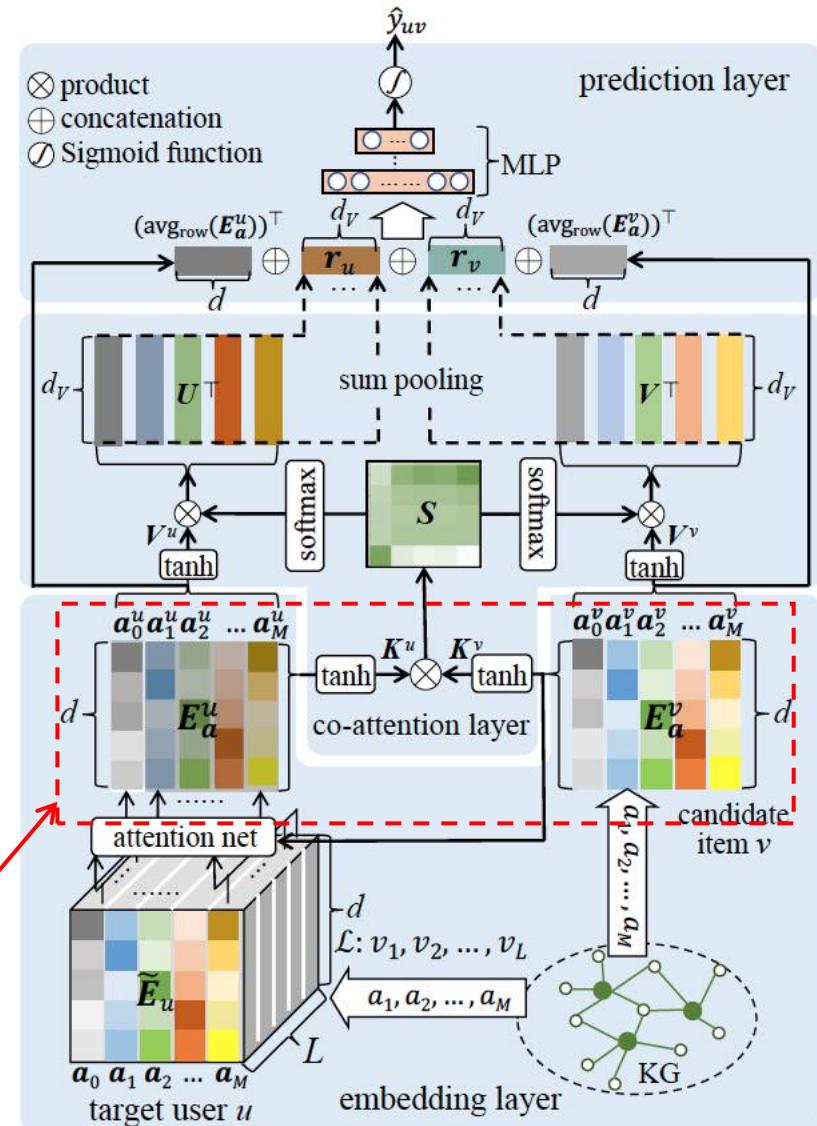


# Embedding-based方法

- 代表模型KAC<sup>[6]</sup>和ACAM<sup>[7]</sup>



从KG中获得的物品属性生成嵌入向量以丰富用户和物品的初始表示



# 内容提纲

---

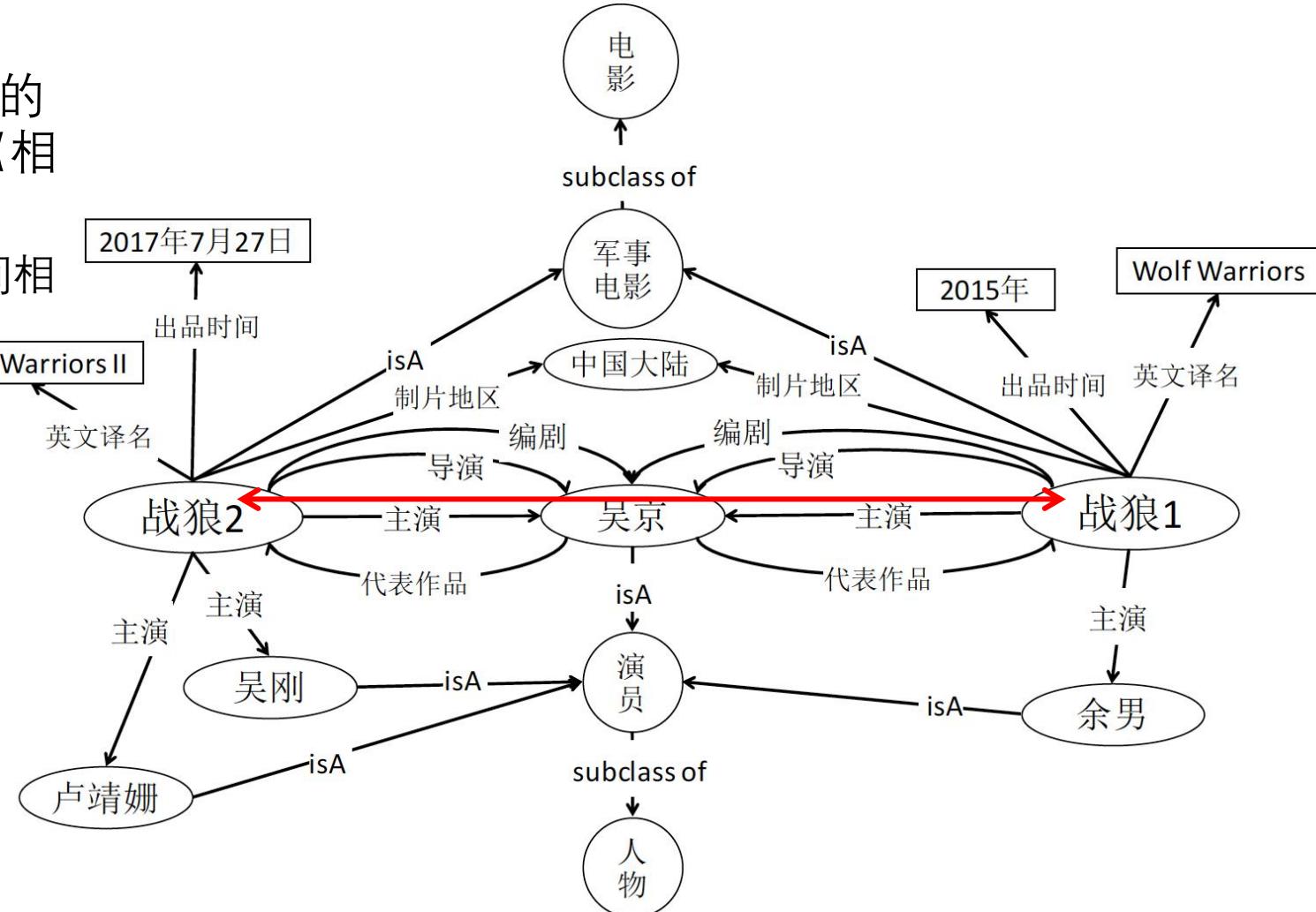
- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - **基于路径的方法**
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# Path-based方法

- 基于KG的信息构建包含用户和物品的异构信息网络 (HIN)
- 基于HIN中联接不同用户/物品节点的meta-path计算节点间语义关联度 (相似度)
  - 如文献[8]中用下式计算两个物品间相似度

$$sim_p(v, v') = \frac{2 \times |S_p(v) \cap S_p(v')|}{|S_p(v)| + |S_p(v')|}$$

电影 $v$ 在属性 $p$ 上所  
关联的实体集合



# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - **集成的方法**
  - 跨领域推荐
  - 可解释推荐

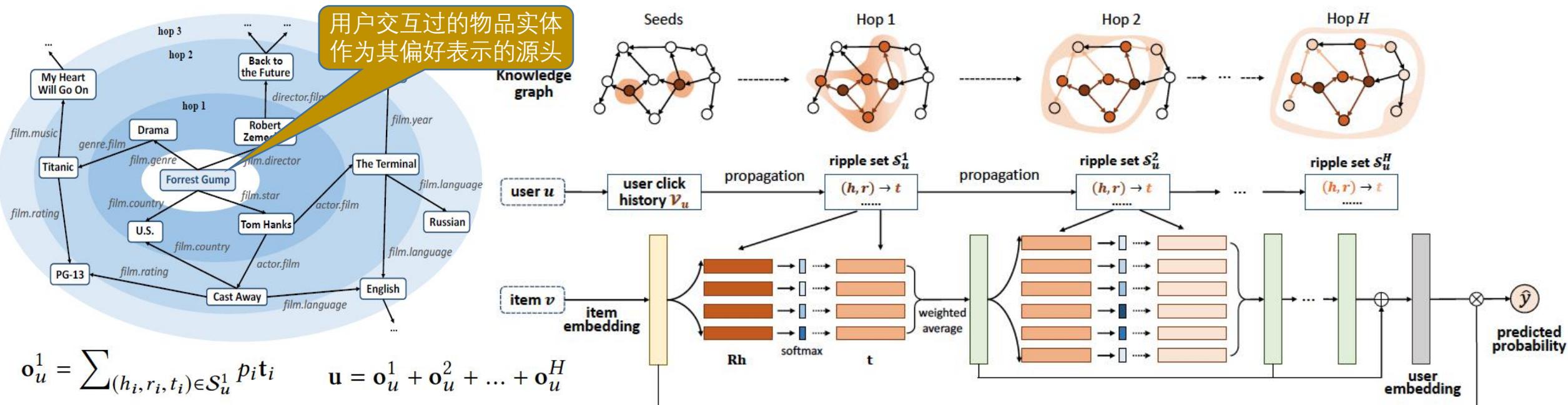
# Unified方法

---

- 融合前两类方法的特点
  - 利用KG中的实体表示+KG/HIN中实体(节点)的关联信息
  - 用KG中的联接结构特性来监督实体表示的优化
- 代表模型
  - RippleNet<sup>[9]</sup>
  - KGAT<sup>[10]</sup>
  - KGCN<sup>[11]</sup>
  - ...

# Unified方法

- RippleNet模型
  - 模拟用户兴趣在知识图谱上的传播过程



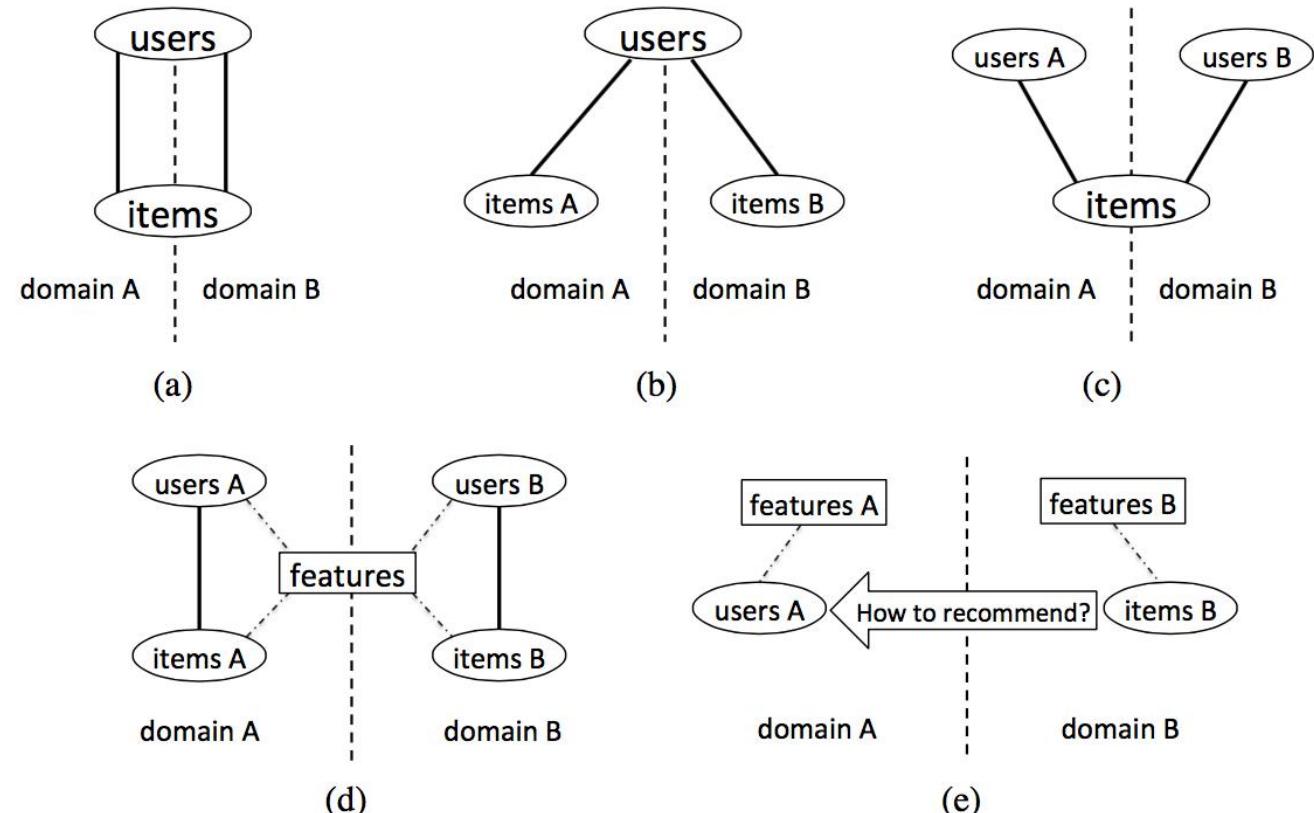
# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 基于知识的跨领域推荐

- 跨领域推荐的几种场景
  - 用户/物品完全重叠
    - 能比较容易地将两个领域合并成一个领域
  - 用户/物品只有部分重叠
    - 重叠的用户/物品成为新兴共享和迁移学习的桥梁
  - 用户/物品完全不重叠
    - 需要挖掘领域间隐含的共同用户/物品或其他关系进行迁移学习



几类跨领域推荐场景[10]

# 基于知识的跨领域推荐

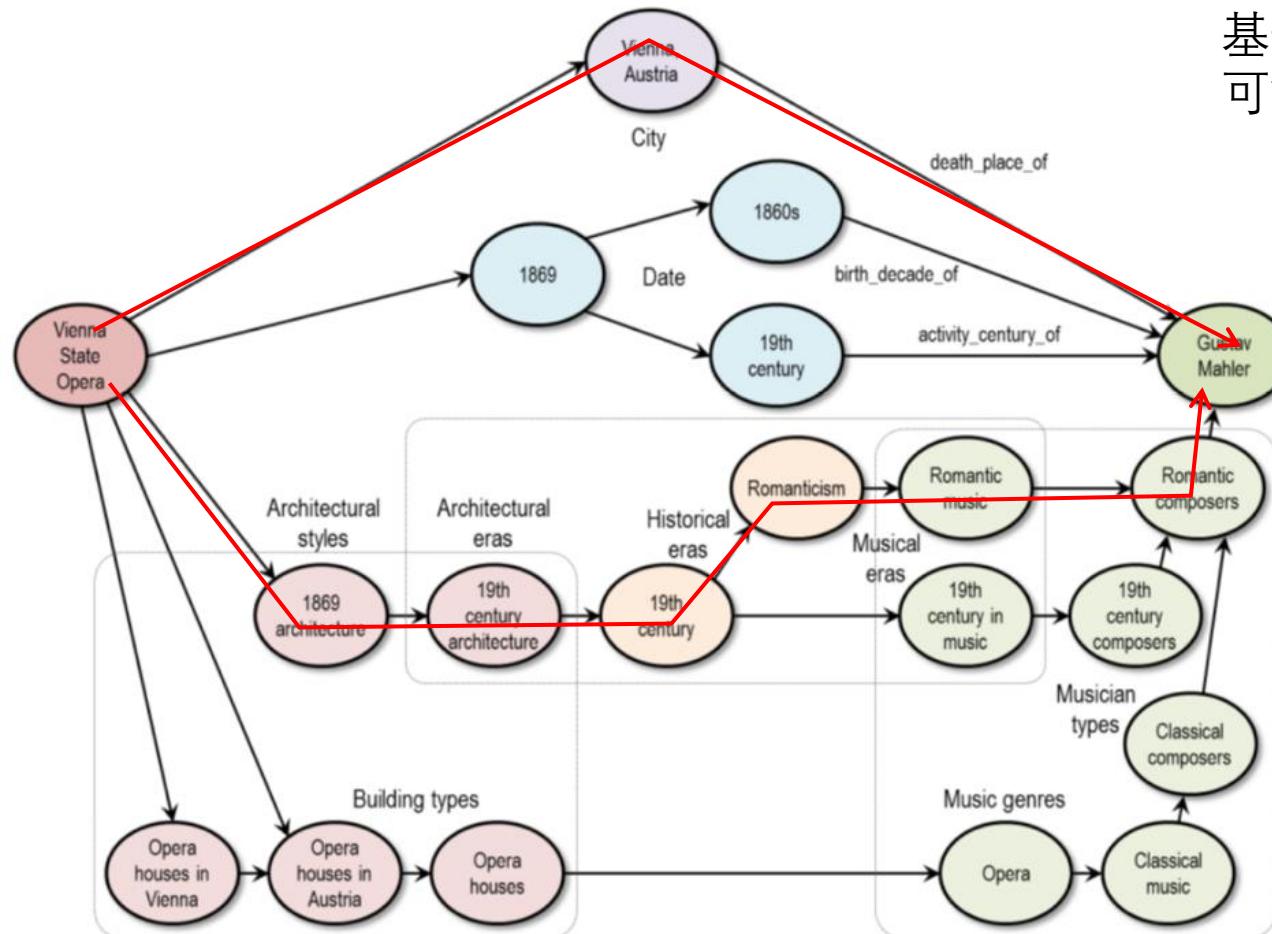
---

- 主流的跨领域推荐算法
  - 基于协同过滤
  - 基于语义关系
  - 基于深度学习
- 实现跨领域推荐的关键假设
  - 用户的兴趣偏好或项目特征在领域之间存在一致性或相关性

通过知识图谱发现

# 基于知识的跨领域推荐

- 跨领域的实体关联<sup>[12]</sup>
  - 利用知识图谱发现不同类实体间的关联路径，并计算相关度

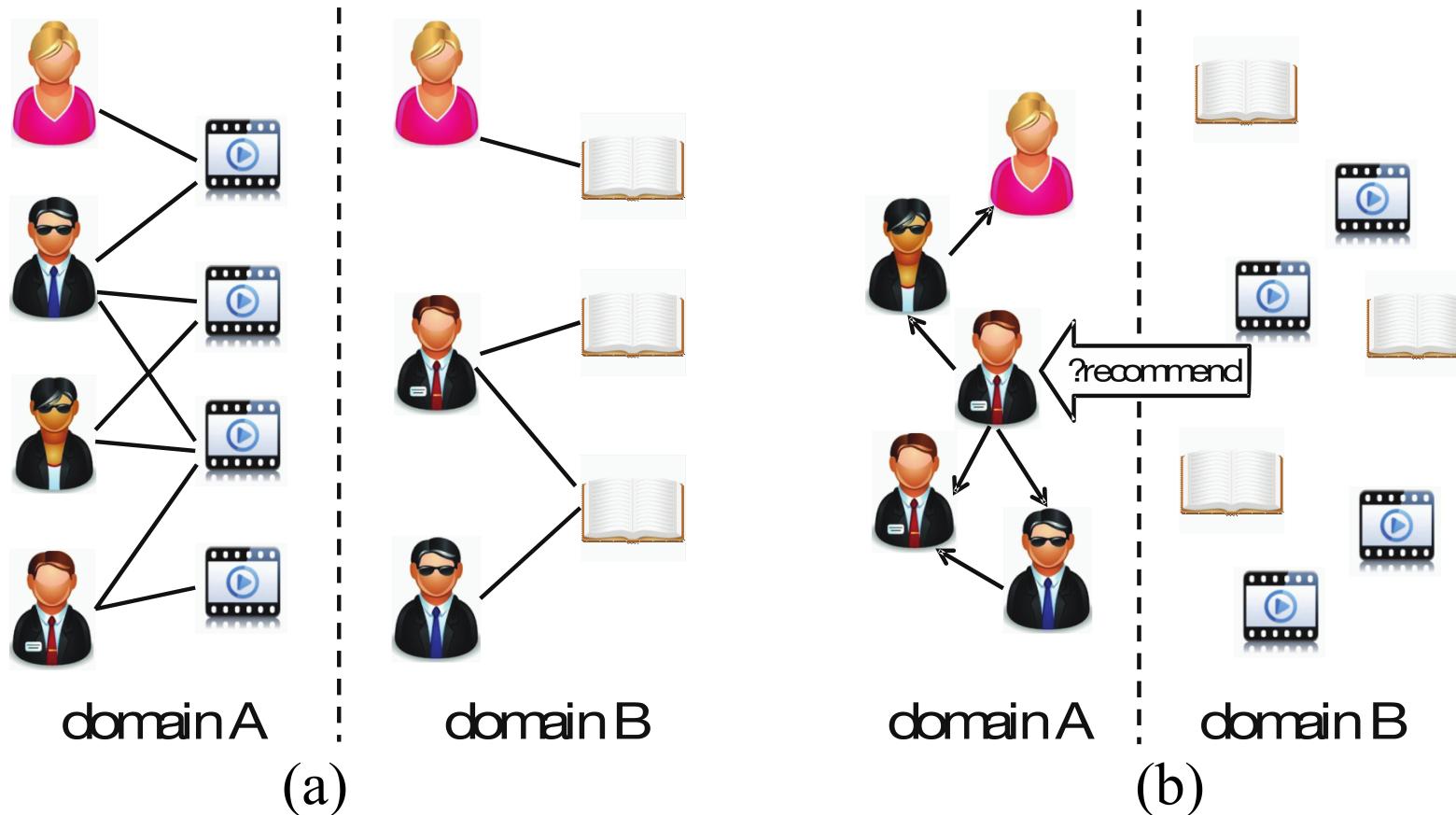


基于迭代的权重传播（类似PageRank）  
可算出头尾实体结点的相关程度

维也纳歌剧院到音乐家Mahler的关联图谱[9]

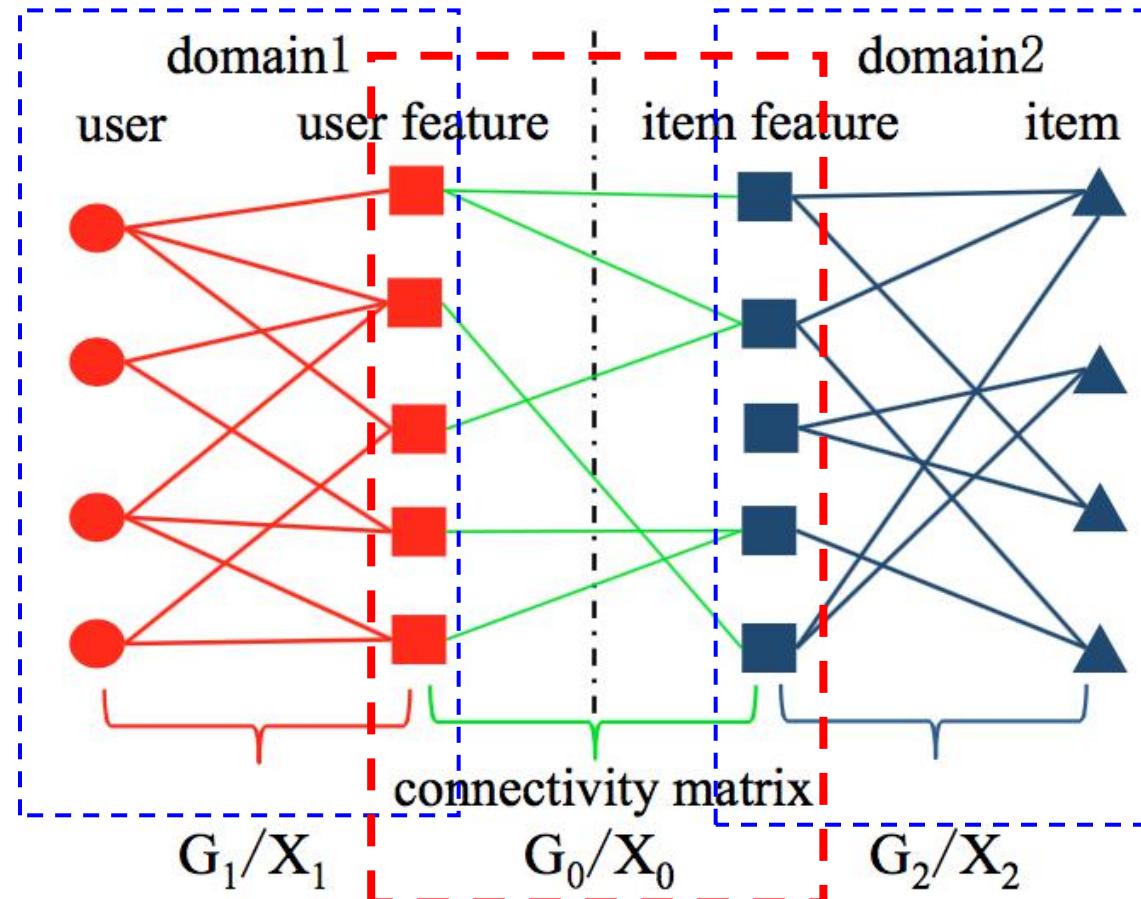
# 基于知识的跨领域推荐

- 跨领域的特征语义关联<sup>[13, 14]</sup>
  - 针对用户和物品完全不重叠的特殊跨领域场景



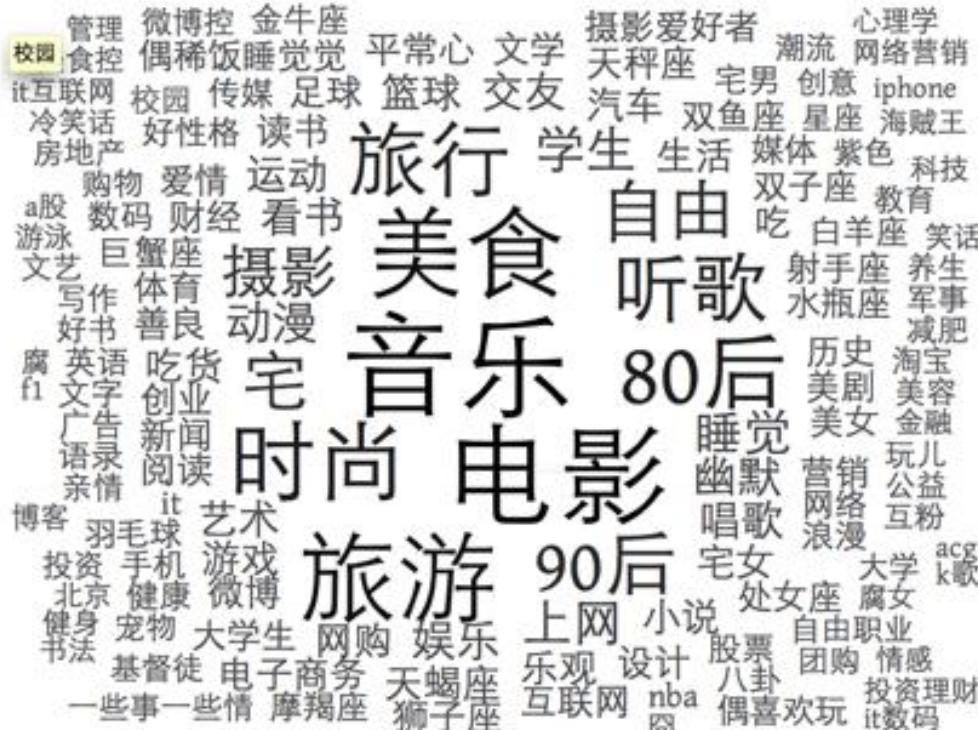
# 基于知识的跨领域推荐

- 跨领域的特征语义关联
  - 首先构建一个多部图 $G$ (multi-partite graph)<sup>[13]</sup>
  - 基于Manifold算法找出领域1的用户与领域2的物品之间的潜在关联



# 基于知识的跨领域推荐

- 跨领域的特征语义关联
    - 实现该算法的挑战在于如何找出两个领域特征（标签）之间的语义关联？即构建 $X_0$ 矩阵



## 微博用户标签



豆瓣电影标签

不同领域的特征（标签）用词有很大不同

# 基于知识的跨领域推荐

- 跨领域的特征语义关联
  - 利用CN-DBpedia语料结合**ESA模型**找出不同标签间的语义关联

- 原理：CN-DBpedia的实体（标签对应的）摘要文本中，出现的参考实体与本实体都有一定的语义关联
- 若两个实体在很多实体的摘要文本中被共同作为参考实体引用，则该两个实体在语义上非常相近

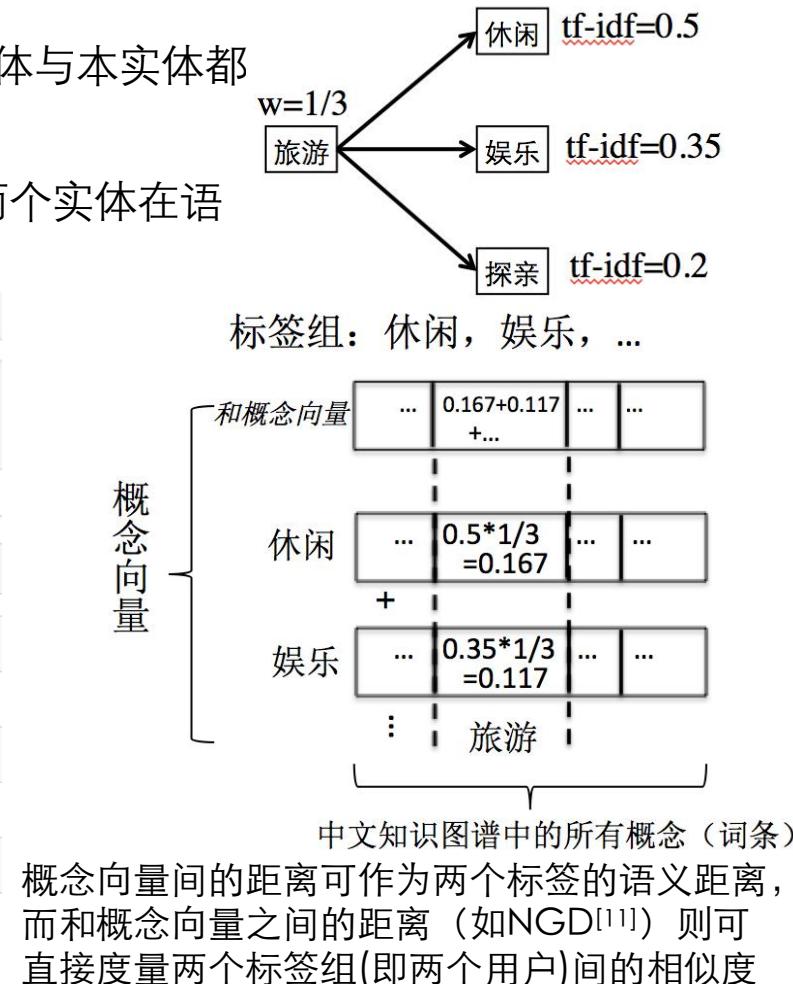
Information

“旅”是旅行，外出，即为了实现某一目的而在空间上从甲地到乙地的行进过程；“游”是外出游览、观光、娱乐，即为达到这些目的所作的旅行。二者合起来即旅游。所以，旅行偏重于行，旅游不但有“行”，且有观光、娱乐含义。

旅游的参考实体

Q InfoBox

世界旅游日	9月27日
中文名	旅游
外文名称	travel
拼音	lǚ yóu
旅游活动	团队旅游和散客旅游



# 内容提纲

---

- 基于知识图谱的搜索
  - 互联网搜索的演化
  - 搜索意图理解
  - 目标查找与结果呈现
  - 实体探索
- 基于知识图谱的推荐
  - 基于知识的推荐概述
  - 基于嵌入向量的方法
  - 基于路径的方法
  - 集成的方法
  - 跨领域推荐
  - 可解释推荐

# 基于知识的可解释推荐

---

- 推荐结果的可解释性是衡量推荐效果好坏的重要指标
- 可解释的推荐结果更容易被用户接受，提升推荐系统的商业价值
- 传统推荐算法往往缺乏可解释性

可解释推荐算法/系统是可解释人工智能（XAI）的重要体现

# 基于知识的可解释推荐

- 已有的可解释推荐可分为两类<sup>[15]</sup>:
  - A: 设计可解释的推荐模型，使得推荐结果可解释
  - B: 为已有的推荐结果寻找可解释依据（信息），并用合适的形式展现出来
- 可解释的依据来源
  - 相关的用户/物品
    - 展现自己评分高的物品与当前物品的相似程度
    - 展现好友（邻居）对当前物品的评分
  - 用户/物品的特征
    - 例如物品的性能参数
  - 文本信息
    - 包括解释性/描述性语句、标签、词云等
  - 图像信息
  - 社交信息
    - 例如展现喜欢当前物品的好友列表

不一定是可解释模型产生的

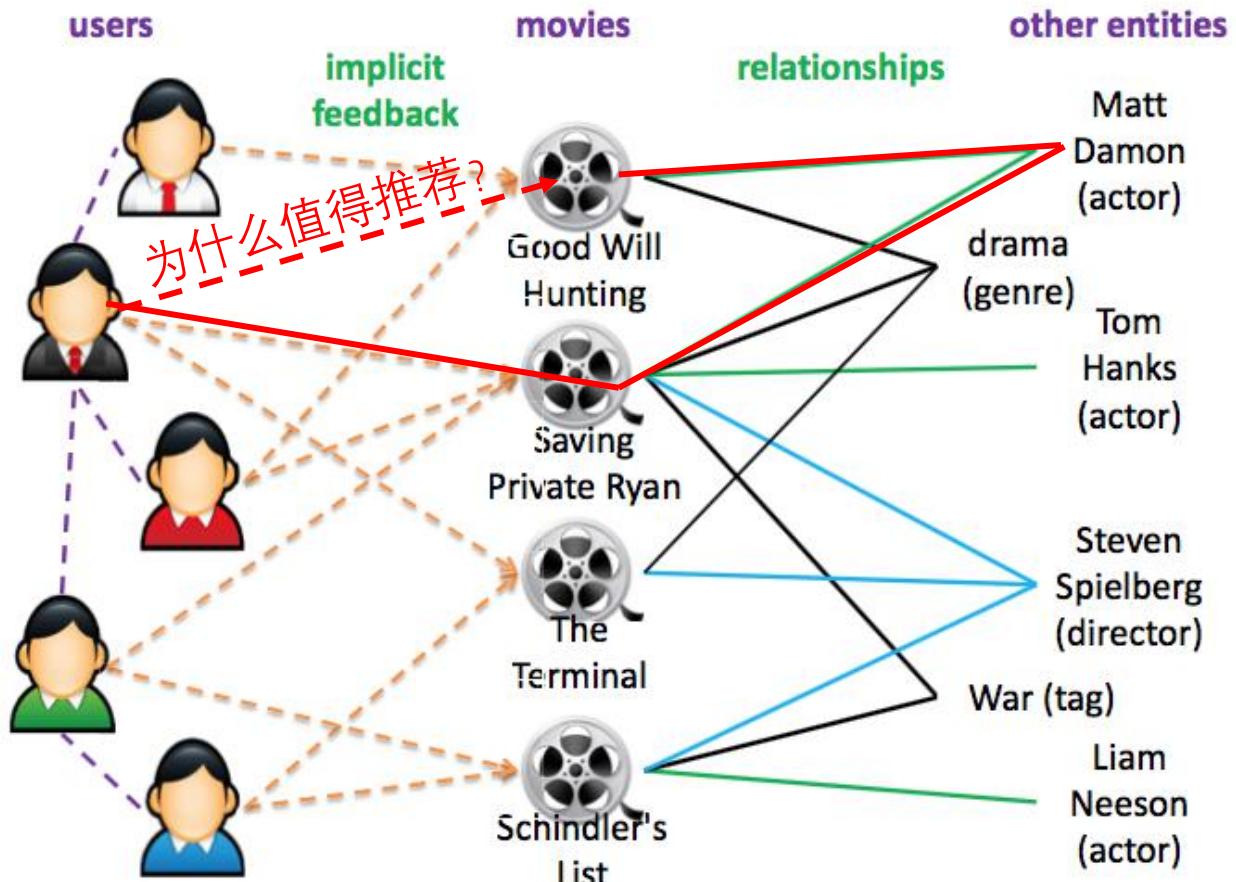
事后(post-hoc)  
可解释推荐

# 基于知识的可解释推荐

- 基于知识图谱的A类可解释推荐
  - 将用户和物品都看作是实体加入知识图谱，基于图谱中已有的实体关联，能够发现连接用户和物品的路径
  - 用户与物品间的关联路径从知识角度上解释了用户会喜欢该物品的原因，基于这些路径设计的模型就是可解释的推荐模型

$$p(u, i | P_{KG}(u, i))$$

图谱中连接用户 $u$ 和物品 $i$ 的路径

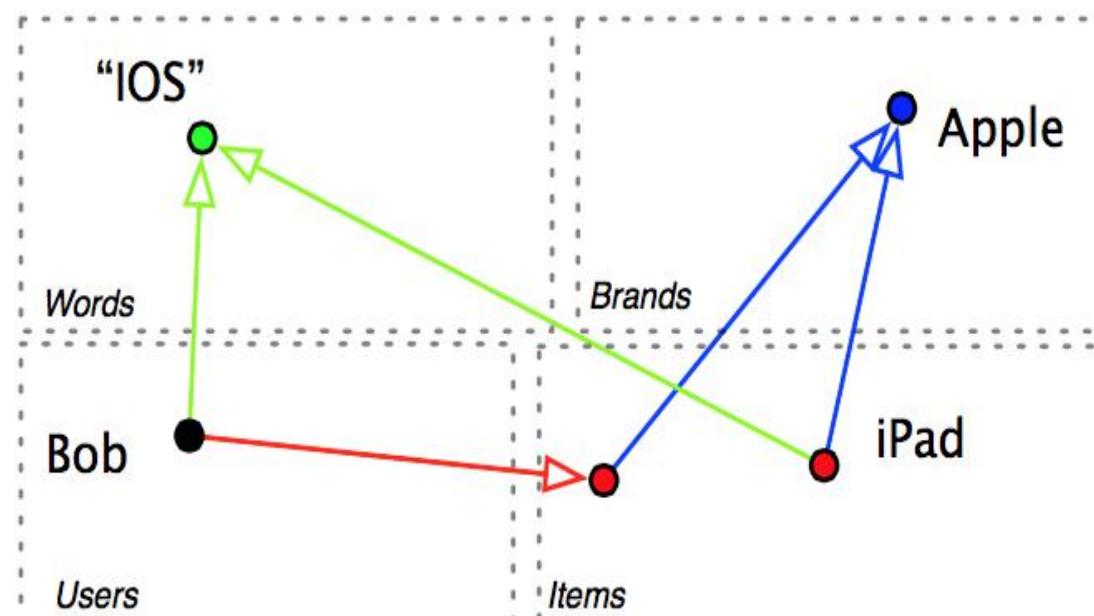


近两年出现的基于深度学习模型的可解释推荐<sup>[16]</sup>也是基于本框架

# 基于知识的可解释推荐

- 基于知识图谱的B类可解释推荐
  - 对于在知识图谱中发现的用户-物品关联路径（可解释路径），设计合理的排序算法，选出最具代表性的路径作为可解释依据

→ *Mention* → *Produced\_by* → *Purchase*



商品知识图谱中用户与商品间的可解释路径<sup>[17]</sup>

# 参考文献

- [1] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, Qing He : A Survey on Knowledge Graph—Based Recommender Systems. CoRR abs/2003.00911 (2020)
- [2] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.—Y. Ma, “Collaborative knowledge base embedding for recommender systems,” in Proceedings of KDD, 2016.
- [3] H. Wang, F. Zhang, X. Xie, and M. Guo, “DKN: Deep knowledge-aware network for news recommendation,” in Proceedings of World Wide Web Conference, 2018.
- [4] J. Huang, W. X. Zhao, H. Dou, J.—R. Wen, and E. Y. Chang, “Improving sequential recommendation with knowledge—enhanced memory networks,” in Proceedings of SIGIR, 2018.
- [5] Deqing Yang, Z. Guo, Z. Wang, J. Jiang, Y. Xiao, and W. Wang: A knowledge—enhanced deep recommendation framework incorporating GAN—based models , in ICDM 2018.
- [6] Deqing Yang, Chenlu Shen, Baichuan Liu, Lyuxin Xue, Yanghua Xiao: Generating Knowledge—based Attentive User Representations for Sparse Interaction Recommendation, in TKDE 2020.
- [7] Deqing Yang, Zengchun Song, Lvxin Xue, Baichuan Liu, Yanghua Xiao: A Knowledge—Enhanced Recommendation Model with Attribute—Level Co—Attention, in SIGIR 2020.
- [8] Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu: Personalized Entity Recommendation: A Heterogeneous Information Network Approach, in WSDM 2014.

# 参考文献

- [9] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo, "RippleNet: Propagating user preferences on the knowledge graph for recommender systems," in Proceedings of CIKM, 2018.
- [10] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "KGAT: Knowledge graph attention network for recommendation," in Proceedings of KDD, 2019.
- [11] H. Wang, M. Zhao, X. Xie, W. Li, and M. Guo, "Knowledge graph convolutional networks for recommender systems," in Proceedings of WWW, 2019.
- [12] Ignacio Fernández-Tobías, Kaminskas M, Ricci F.: A generic semantic-based framework for cross-domain recommendation, International Workshop on Information Heterogeneity and Fusion in Recommender Systems 2011.
- [13] Deqing Yang, J. He, H. Qin, Yanghua Xiao, et al.: A Graph-based Recommendation across Heterogeneous Domains, in CIKM 2015.
- [14] Deqing Yang, Yanghua Xiao, Y. Song, et al.: Semantic-Based Recommendation Across Heterogeneous Domains, in ICDM 2015.
- [15] Yongfeng Zhang, Xu Chen. Explainable Recommendation: A Survey and New Perspectives. arXiv:1804.11192v4 [cs.IR]. 2018.
- [16] Xiang Wang, Dingxian Wang, Canran Xu, et al.: Explainable Reasoning over Knowledge Graphs for Recommendation, in AAAI 2019.
- [17] Ai Q, Azizi V , Chen X , et al. Learning Heterogeneous Knowledge Base Embeddings for Explainable Recommendation[J]. 2018.

# 《知识图谱：概念与技术》

Q & A

