



# 那些你不说我也知道的事

## 浅谈知识图谱中的关系推理

主讲人：薇拉vera@勾陈安全实验室

# 先扯个闲话

## 数据挖掘 (Data Mining)

从数据中挖出有用的信息

## 机器学习 (Machine Learning)

一种工具，可以变身成各种样子

## 深度学习 (Deep Learning)

机器学习里神经网络算法的衍生，本质上还是属于机器学习的一类

## 大数据(Big Data)

无法用现有的软件工具提取、存储、搜索、共享、分析和处理的海量的、复杂的数据集合



但总而言之  
最大的BOSS是

# 人工智能

让  
计算机  
展现  
智慧  
的研究

知识图谱

语音识别

自然语言处理

翻译

图片搜索引擎

计算机视觉

AR

图像识别

机器学习

深度学习

数据挖掘

机器人





# 知识图谱

- 本质上是语义网络。
- 一种基于图的数据结构，由节点(Point) (“实体”) 和边(Edge) (“关系”) 组成。
  - 每个节点：现实世界中存在的“实体”
  - 每条边为实体与实体之间的“关系”。
- 知识图谱是关系的最有效的表示方式。
- 通俗地讲，知识图谱就是把所有不同种类的信息连接在一起而得到的一个关系网络。





网页

新闻

贴吧

知道

音乐

图片

视频

地图

文库

更多»

百度为您找到相关结果约2,380,000个

搜索工具



比尔·盖茨妻子：

梅琳达·盖茨

梅琳达·盖茨（Melinda Gates，1964年8月15日~），毕业于美国杜克大学计算机系，后获得MBA学位，然后如愿进入了自己曾经实习过的微软公司，很快崭露头角，取得了骄人的业... [详情>>](#)

来自百度百科 | 报错

【专访】比尔盖茨:与妻子结婚是自己做的最好决定 访谈 亿智蘑菇



2016年3月9日 - 盖茨在Reddit论坛上说,与妻子结婚是他有生以来做过最好的决定。... 据CNN网站报道,作为世界上最富有的人,比尔盖茨可以花钱买下任何东西,但是他却说,并... [www.yzmg.com/fangtan/1...](http://www.yzmg.com/fangtan/1...) [Vi](#) - 百度快照



全部

图片

新闻

视频

地图

更多

设置

工具

找到约 1,720,000 条结果（用时 0.49 秒）

比尔·盖茨 / 配偶

梅琳达·盖茨

结婚时间：1994 年



梅琳达·盖茨，婚前原名梅琳达·安·法兰奇。美琳达·盖茨出生于美国德克萨斯州达拉斯，并在达拉斯长大，其丈夫为世界首富比尔·盖茨，现居麦地那。 [维基百科](#)

更多有关“梅琳达·盖茨”的信息

反馈

时代周刊的风云人物

[比尔·盖茨](#)

[扎克伯格](#)

相关人物

[斯特凡妮·宙赫尔](#)

[普莉希拉·陈](#)



# 知识图谱的应用

更加丰富的相关信息推荐



Baidu 百度 达芬奇 百度一下 百度浏览器, 打开网页快2秒!

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约25,000,000个

为您推荐: [达芬奇的恶魔](#) [达芬奇词典](#) [达芬奇藏品](#) [达芬奇词典软件](#)

[达芬奇](#) 百度百科

姓名: 列昂纳多·达·芬奇  
生卒: 公历1452年4月23日, 儒略历1452年4月15日-1519年5月2日  
简介: 莱昂纳多·达·芬奇(儒略历1452年4月15日—1519年5月2日, 公历1452年4月23日—1519年5月), 又译列昂纳多·达·芬奇、李奥纳多·达·芬奇, 是意大利文艺复兴时期的一个杰出天才。  
[简介](#) [生平](#) [科学](#) [工程](#) [创作](#) [成就](#) [遗产](#) [大事记](#) [更多>>](#)  
[baike.baidu.com/ 2014-10-18](#)

[达芬奇](#) 百度图片 - 壁纸图片

[image.baidu.com](#) - 查看全部89,000张图片

[达芬奇的恶魔第二季](#) - 高清在线观看 - 腾讯视频

[达芬奇](#)在第二季中为探险《叶之书》(Book Of Leaves)的秘密将前往一个陌生之地, 利用天才的头脑展开一系列危机四伏的新冒险, 完成一

其他人还搜 [展开](#)

[蒙娜丽莎](#) [达芬奇调色](#) [不做菜的菜](#)

相关人物 [展开](#)

[伦勃朗](#) [卓别林](#) [苏拉·哈德森](#)

相关画作 [展开](#)

[渔光曲](#) [土耳其浴室](#) [最后的晚餐](#)



通过知识图谱建立事物之间的关联, 扩展用户搜索结果, 发现更多内容。

立即体验



# 知识图谱的应用

多维度的信息展示



通过知识图谱梳理事态的进展，为用户在广度和深度上提供全方位的内容。

立即体验

# 知识图谱的应用

## 场景化的搜索结果

 全聚德 百度一下

在北京市搜索全聚德\_百度地图



**A 全聚德(前门店)**  
¥37起 ★★★★★ 2844条评论  
地址: 前门大街32号前门步行街北口内...  
电话: 010-67011379 [订座](#)

**B 全聚德(奥运村店)**  全景  
¥162起 ★★★★★ 781条评论  
地址: 朝阳区大屯路慧忠北里309号楼...  
电话: (010)64801686

**C 全聚德(清华园店)**  全景  
¥139起 ★★★★★ 511条评论  
地址: 北京海淀区中关村东路1号院清...  
电话: (010)82150018  
[查看全部39条结果>>](#)  
[map.baidu.com](http://map.baidu.com)



借助于知识图谱，结合用户行为信息，为用户提供更符合当前场景的搜索结果。

立即体验

# 知识图谱的应用

直接给出答案



汪涵的妻子是谁

百度一下

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约2,240,000个



汪涵妻子:

**杨乐乐**

杨乐乐，1978年6月23日出生于重庆，中国内地节目主持人。1998年从四川师范大学表演及节目主持专业毕业，在四川电视台黄金十频道担任《好运周末》及生活时尚类节目《生活... [详情>>](#)

来自百度百科 | 报错



妻子

?

通过知识图谱建立实体之间的属性与关系，让搜索引擎更懂用户的意图，直接解答用户的疑惑。

立即体验



# 起源

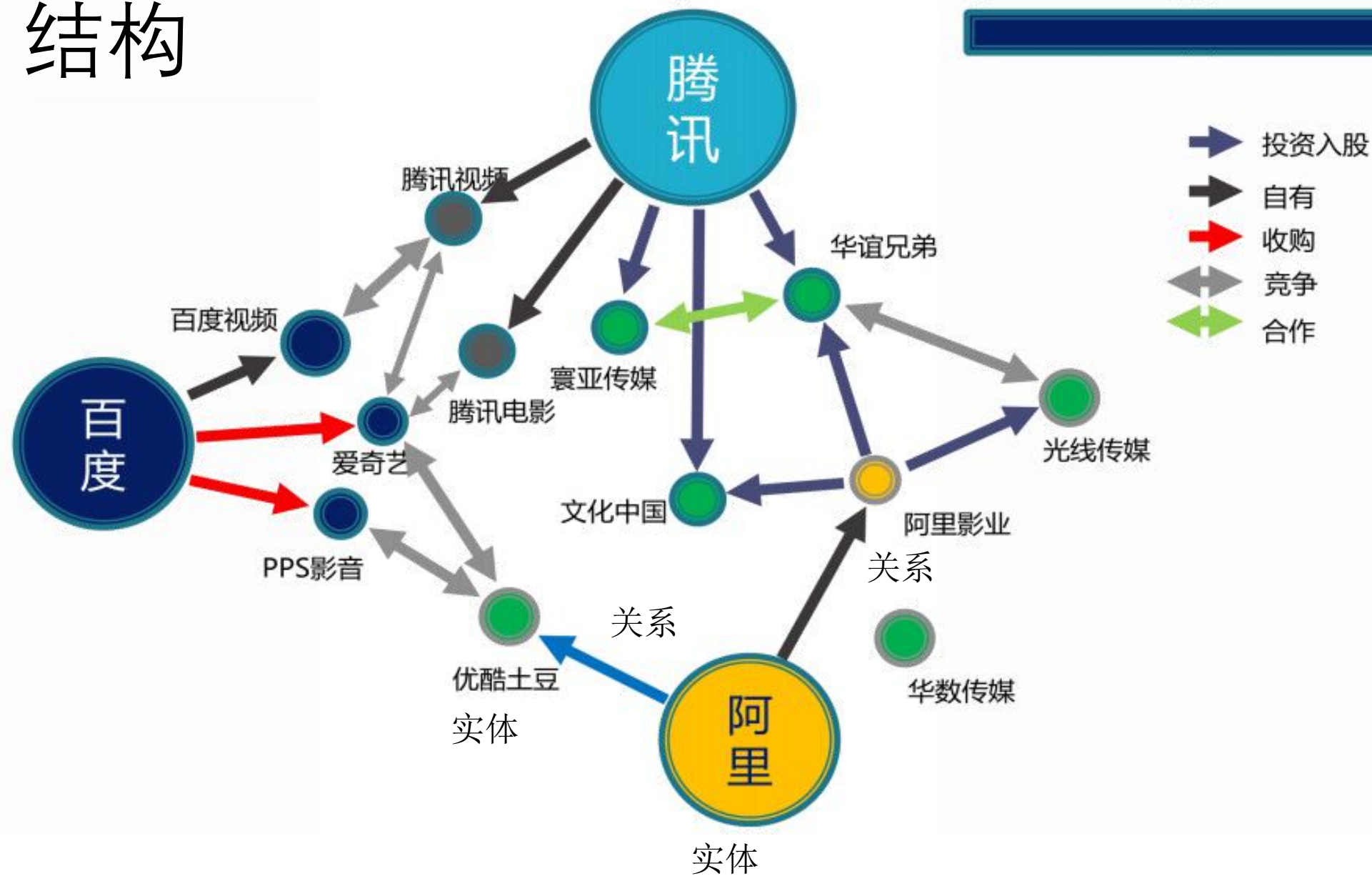
---

- 知识图谱最早的起源：**知识工程**（通过智能软件建立的专家系统）
- 知识图谱这个概念最早由Google提出，主要是用来优化现有的搜索引擎。
  - 不同于基于关键词搜索的传统搜索引擎，知识图谱可用来更好地查询复杂的关联信息，从**语义层面**理解用户意图，改进搜索质量。比如在Google的搜索框里输入Bill Gates的时候，搜索结果页面的右侧还会出现Bill Gates相关的信息比如出生年月，家庭情况等等。

# 数据来源

- 百科类数据
- 结构化数据
- 半结构化数据挖掘AVP
- 通过搜索日志进行实体和实体属性等挖掘

# 结构



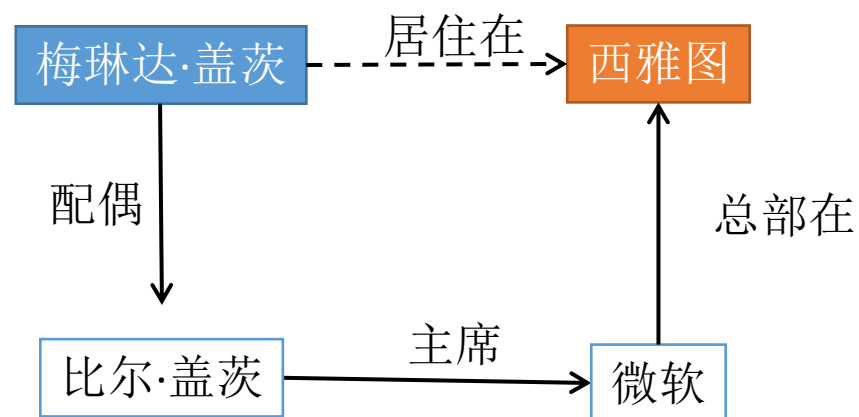


# 但是……

- 虽然目前的知识图谱上已经有了非常多的实体对和关系事实，但是由于数据的更新迭代以及不完整性，注定了这个知识图谱的不完整，同样，他里面也隐藏着我们难以轻易发现的信息。



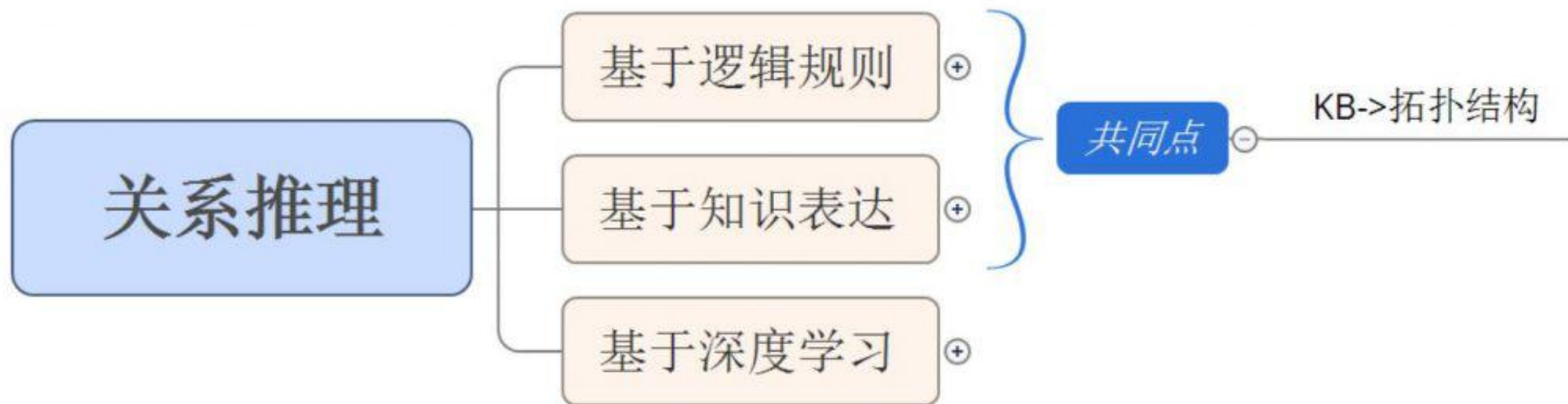
?



怎么实现？

# 关系推理





# 基于逻辑规则的关系推理

- 建模依据：采用抽象或具象的Horn子句
- 本质：基于逻辑规则进行推理
- 优势：能够模拟人类的逻辑推理能力，有可能引入人类的先验知识辅助推理
- 缺点：尚未有效解决优势所带来的的一系列问题，包括专家依赖、复杂度过高等问题
- 发展趋势
  - 逐渐摒弃对人工规则的依赖
  - 转而借助模式识别的方式进行规则（模式特征）发现
  - 采用机器学习方法进行特征建模

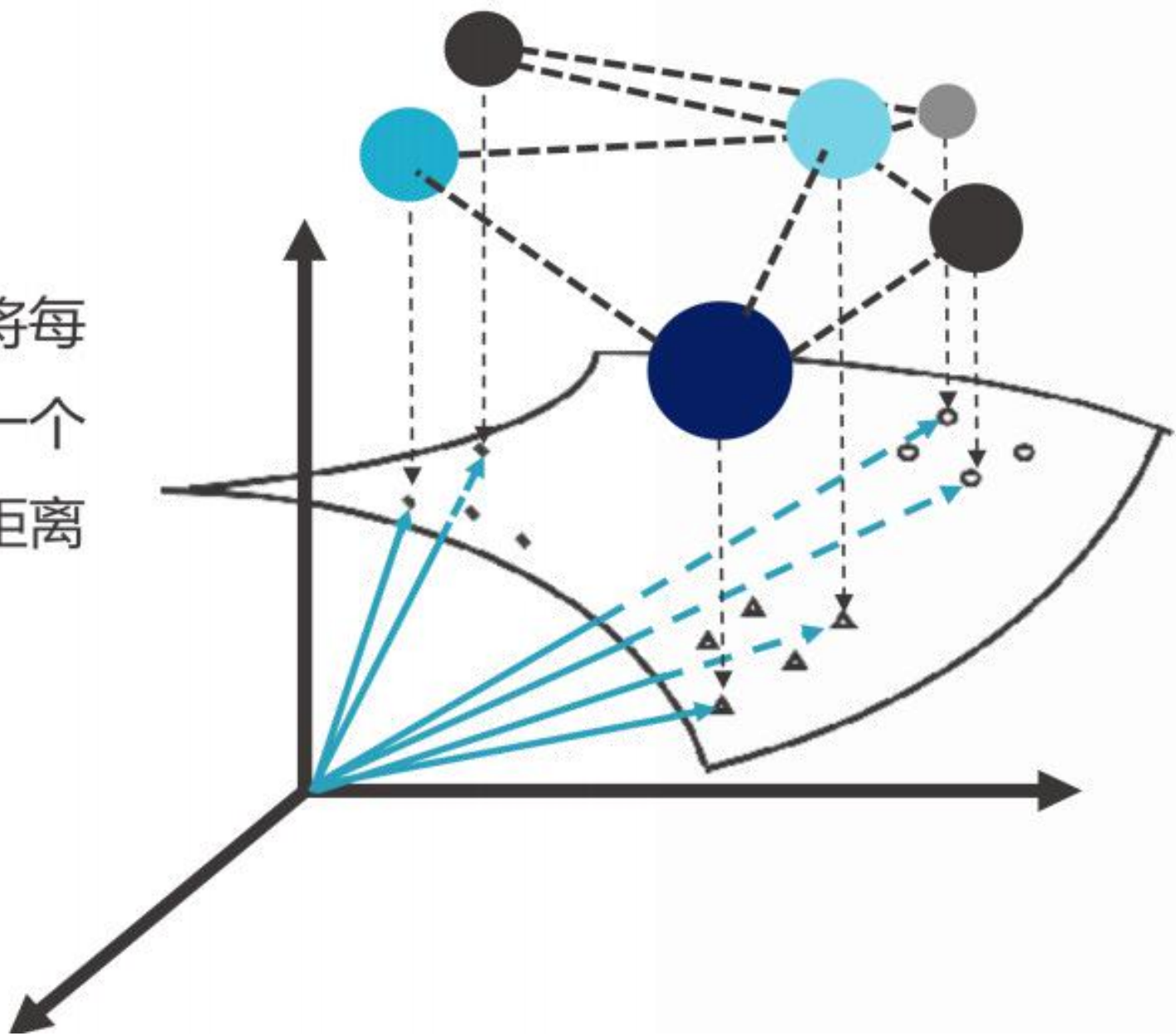
代表性工作：

- 马尔代夫逻辑网络模型
- 基于贝叶斯网络的概率关系模型
- 基于统计机器学习的FOIL算法
- PRA算法
- SFE算法
- HiRi算法

# 基于知识表达的关系推理

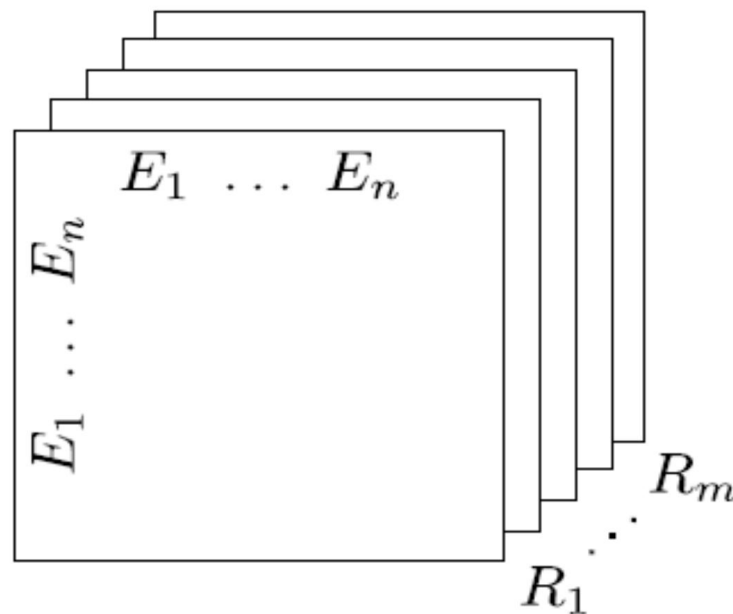
- 建模依据：将实体和关系映射到一个低维的embedding空间中，基于知识的语义表达进行推理建模
- 代表性工作：
  - RESCAL张量分解模型 (Tensor Factorization Model)
  - SE (Structured Embedding) 关系推理算法
  - TransE (Translating Embedding) 算法及其系列算法
- 优势：生成知识表达时能够充分利用知识图谱已有的结构化信息
- 缺点：建模方法着眼于实体间的直接关联关系，难以引入并利用人类的先验知识实现逻辑推理

**基本思想：**用低维的向量空间将每个实体都表示到空间里面的某一个位置，从而可以利用空间中的距离来衡量实体之间的语义关系。



# Representation Learning

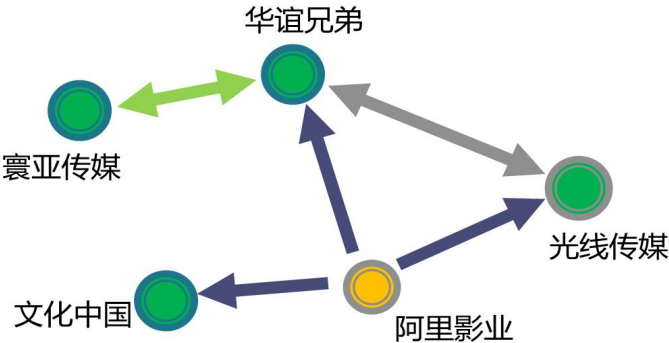
**基本思想**：  $E_1, E_2, \dots, E_n$  代表所有实体，  $R_1, R_2, \dots, R_m$  代表所有关系，故可以使用如下三维矩阵（张量）表示知识图谱：





# Representation Learning

示例：



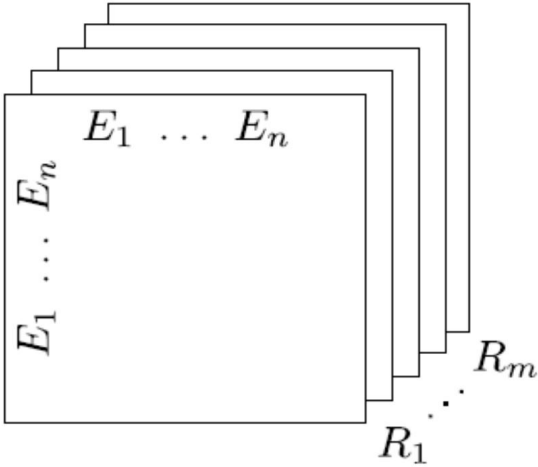
头实体	关系	尾实体
阿里影业	投资	光线传媒
阿里影业	投资	文化中国
阿里影业	投资	华谊兄弟
华谊兄弟	合作	寰亚传媒
华谊兄弟	竞争	光线传媒

	阿里影业	文化中国	寰亚传媒	华谊兄弟	光线传媒
阿里影业	0	1	0	1	1
文化中国	0	0	0	0	0
寰亚传媒	0	0	0	0	0
华谊兄弟	0	0	0	0	0
光线传媒	0	0	0	0	0

投资关系 R1

	阿里影业	文化中国	寰亚传媒	华谊兄弟	光线传媒
阿里影业	0	0	0	0	0
文化中国	0	0	0	0	0
寰亚传媒	0	0	0	0	0
华谊兄弟	0	0	0	0	1
光线传媒	0	0	0	1	0

竞争关系 R2



# Representation Learning

**已知** : 三元组  $\langle h, r, t \rangle$

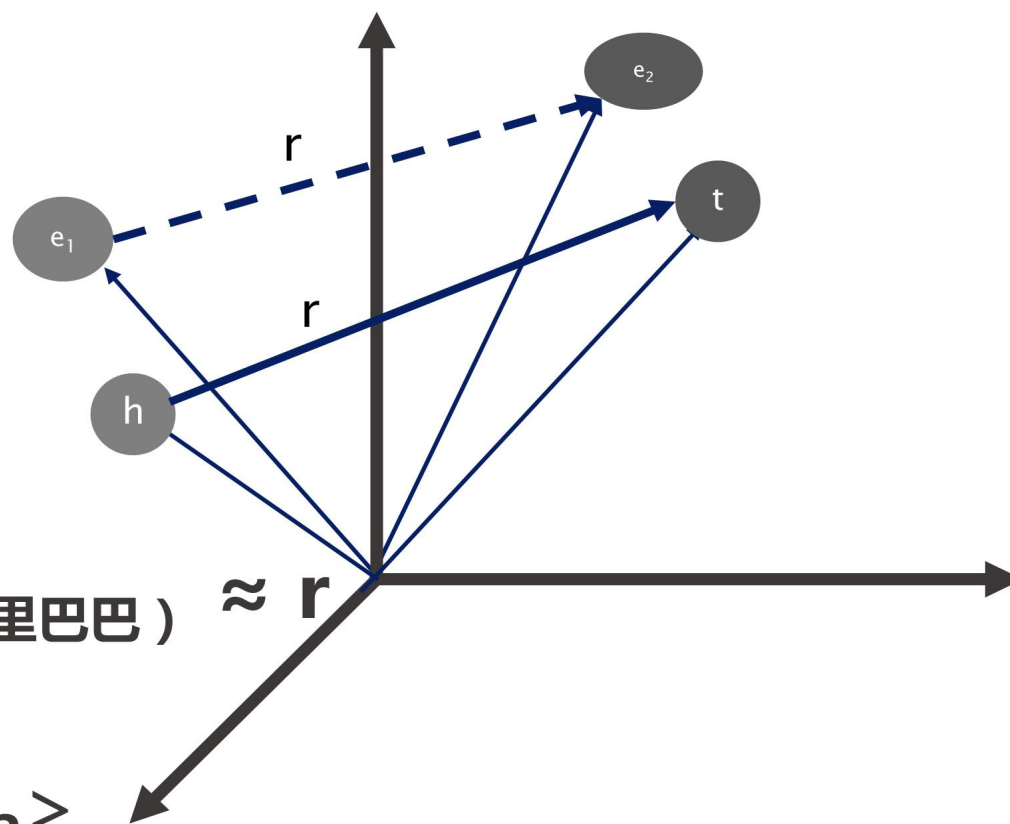
$h$  = 腾讯       $t$  = 寰亚传媒

$h + r = t$  ;  $r$  = 投资入

**假设** :  $e_2$  ( 华数传媒 ) -  $e_1$  ( 阿里巴巴 )  $\approx r$



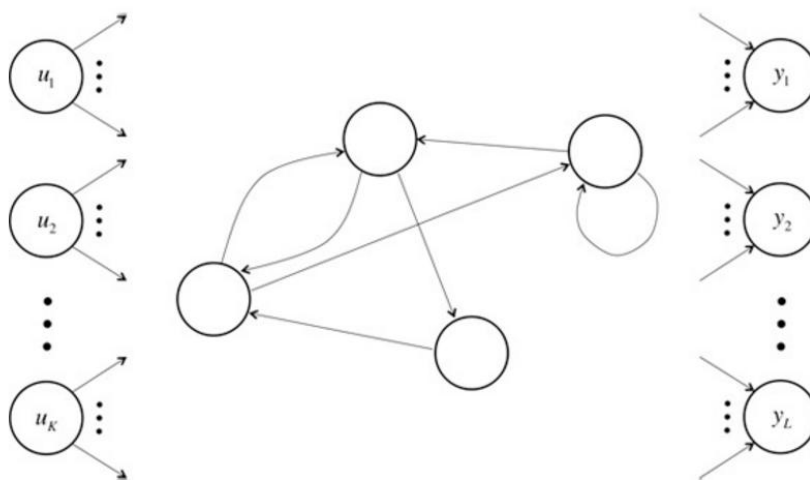
新三元组  $\langle e_1, r, e_2 \rangle$



# 基于深度学习的关系推理

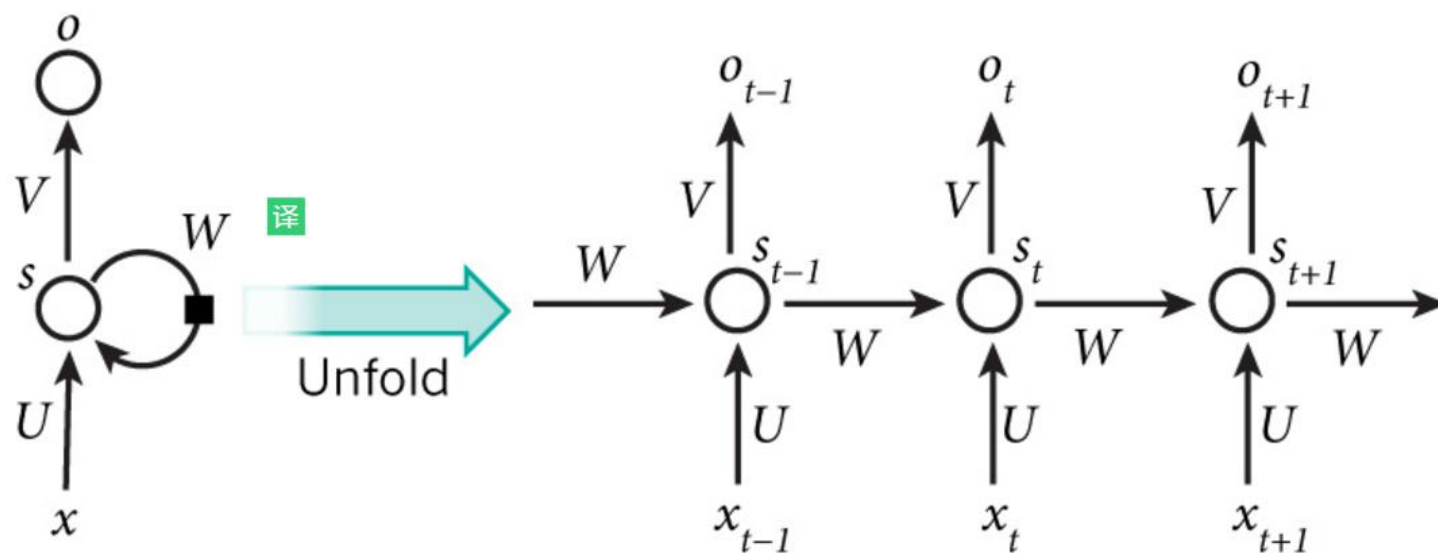
- 代表性工作：

- 单层感知机模型SLM (Single Layer Model)
- NTN神经张量模型 (Neural Tensor Networks)
- DKRL (Description-Embodied Knowledge Representation Learning) 模型
- Path-RNN模型



# 基于深度学习的关系推理

- 输入集标记为  $\{x_0, x_1, \dots, x_t, x_{t+1}, \dots\}$ ,
- 输出集标记为  $\{y_0, y_1, \dots, y_t, y_{t+1}, \dots\}$
- 隐藏单元的输出集标记为  $\{s_0, s_1, \dots, s_t, s_{t+1}, \dots\}$
- $x_t$  表示第  $t, t=1, 2, 3 \dots$  步(step) 的输入
- $s_t$  为隐藏层的第  $t$  步的状态, 它是网络的记忆单元。
- $O_t$  是第  $t$  步的输出
- 共享参数  $U, V, W$



# 基于深度学习的关系推理

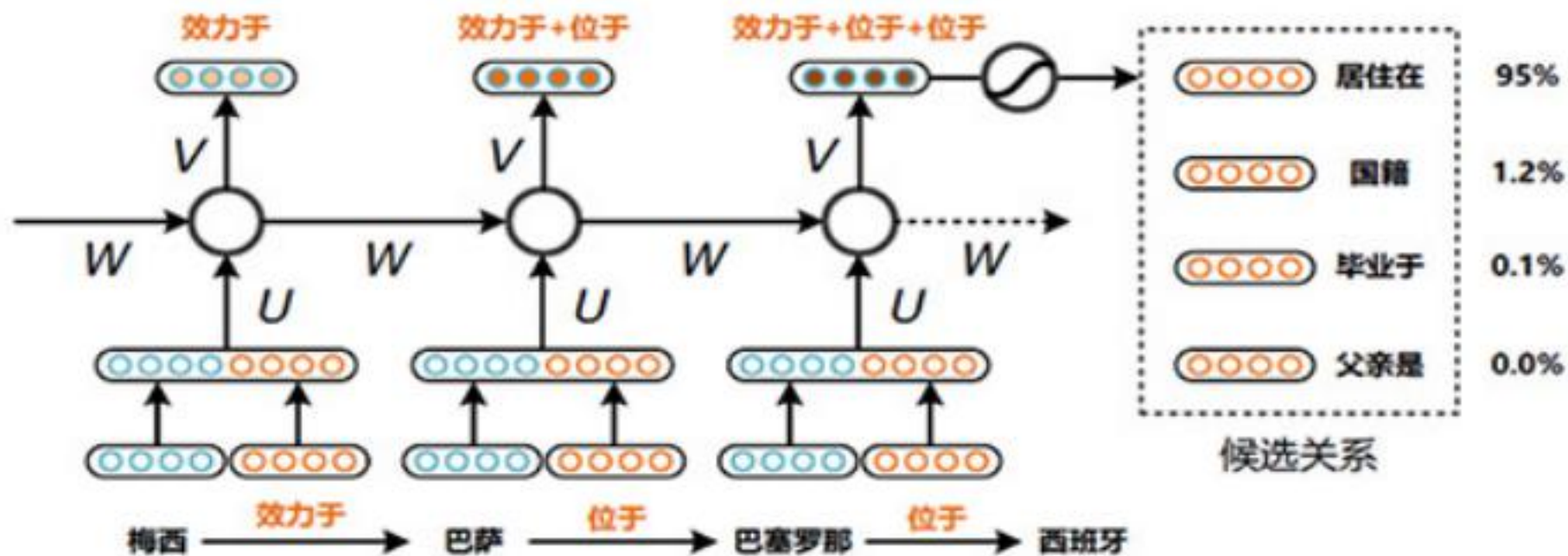


图 3-3 基于深度学习的关系推理技术路线示意图



# 相关资料：

- [数据挖掘、机器学习、深度学习，脸盲症入坑指南](#)
- [知识图谱中的关系推理，究竟是个什么玩意儿？](#)
- [知识图谱技术综述](#)
- [知识图谱技术原理介绍](#)
- [知识图谱的应用](#)

谢谢~

主讲人：薇拉vera@勾陈安全实验室