

SEU 知识图谱概论

2019年10月28日 14:08

- **1.知识图谱概念**

- **知识的图表示形式**

- 包括：实体、属性、关系
- 节点是实体，节点有属性标签（可以包含类型）
- 两个节点之间的边是实体键的关系
- 强调实体，但是也可以描述概念

- **深度学习 VS 知识图谱**

	 深度学习	 知识图谱
智能原理	<ul style="list-style-type: none">• 对人类智能(大脑)的隐性模拟	<ul style="list-style-type: none">• 对人类智能(思考)的显性模拟
场景	<ul style="list-style-type: none">• 擅长处理感知智能领域特定任务• 围棋、语音、图像、视频	<ul style="list-style-type: none">• 广泛用于各种任务• 搜索、人机交互
特点	<ul style="list-style-type: none">• 海量训练数据• 强大算力• 难解释	<ul style="list-style-type: none">• 海量知识• 可解释• 可理解
进展	<ul style="list-style-type: none">• 在一些任务上接近或超过人类	<ul style="list-style-type: none">• 在知识量上超过人类• 在知识推理上不如人类
趋势	<ul style="list-style-type: none">• 未来两种智能技术深度融合	

- **深度学习面临的问题**

- **算力红利**

GPU/TPU等硬件的计算速度的摩尔定律开始失效

- **大数据红利**

大数据的获取、清洗、标注已经出现瓶颈

- **算法红利**

新模型的提出很难；现有的模型的优化和潜力挖掘存在瓶颈

- **平台红利**

DL开发平台日趋成熟；难以有更大的提升空间

- **知识表示**

- **语义网络**

- **框架**

- **概念图**

- 有数学和逻辑支撑的知识表示
- 一阶逻辑的图接口，基于图的知识表示和推理

- **形式化概念分析 (formal concept analysis)**

- 格 (Lattice) 偏序集合 // 广泛应用在数据挖掘、机器学习、语义web

- **语义网**

- **Semantic web = data + link**

- 语义网：蛋糕模型

- 本体：领域共享知识的描述方式，是语义web、语义搜索、知识工程和很多人工智能的应用基础

- **2.经典知识图谱**

- **Cyc**

cyc知识图谱主要由术语Terms和断言Assertions组成

Terms包含概念、关系、实体的定义。

Assertions用来建立Terms之间的关系//既包含事实fact 也包含规则rule的描述

- **WordNet**

词典知识库，主要用于词义消歧

WordNet主要定义了名词、动词、形容词和副词之间的语义关系。

- **ConceptNet**

ConceptNet是常识知识库，以三元组形式的关系型知识构成。

与cyc相比，ConceptNet采用了非形式化、更接近自然语言的描述。而不是像Cyc那样采用形式化的谓词逻辑。

- **Freebase**

目标包含19亿三元组，提供多语言查询接口

- **Wikidata**

个人可编辑的知识库

核心概念是：**entity**

item都有**label**、**description**、**aliases**

item中的的具体数据为**statement = property + value + qualifier + reference** 等

- **DBPedia**

DBPedia是早期的语义网项目

DBPedia意指数据库版本的Wikipedia，是从wikipedia中抽取出来的链接数据集

DBPedia采用RDF语义数据模型

- **YAGO**

集成了wikipedia wordnet geonames 三个来源的数据

考虑了时间和空间的共识

- **NELL**

主要采用互联网挖掘的方法从Web自动抽取三元组知识

机器学习的方式不断从web中学习和抽取新的知识

- **OpenIE**

- **ZhiShi.me**

百度百科 互动百科 维基百科中文部分为基础构建的中文知识图谱

- **CN-DBPedia**

类似于zhishi.me

- **BabelNet**

类似于wordnet的多语言词典知识库

- **3.知识图谱应用**

- 知识图谱怎么用

结构化数据	查询、搜索、统计、分析、挖掘
图数据	查询、搜索、可视化、图分析、挖掘、推荐
知识	重用、认知（语义理解、解释、推理、理性）

- 知识图谱应用场景

- 辅助搜索--精准回答
 - 提高搜索精度
 - 语义搜索
 - 搜索意图理解
 - 多模态搜索

知识图谱就是高质量数据，最好的应用方式就是直接用它！

- 辅助问答--人机互动

- 提高问答精度
- 改善问答体验
- 引导对话走向
- 多轮对话

问答就是人工智能的重要应用场景

- 辅助数据集成--智能数据整合

- 针对大规模多源易购数据集成机制
- 建立和文本挖掘数据内在的关联
- 通用性高
- 可扩展性强
- 形式灵活
- 只是重用
- 不干涉数据源

知识图谱实现了数据-信息-知识的智能数据整合机制

- 辅助决策--智能决策

- 收集和整理数据
- 建立数据关联
- 知识挖掘
- 知识发现
- 知识推理
- 知识图谱辅助自然语言理解
- 辅助各种AI场景
- 广泛用于国防 金融 制造 商务等

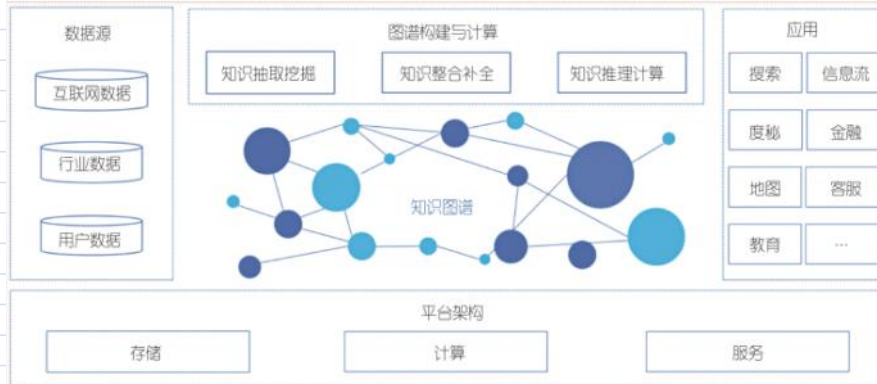
知识图谱和各种AI技术综合使用能更好的发挥AI作用

- 知识图谱技术的核心价值

- 集成异构数据源
 - 知识图谱构建基于分布和异构数据源
 - 知识图谱的图模型是对异构数据的一种灵活集成
- 描述数据之间的关联

- 实体关系
- 事件
- 实现实体链接
 - 建立知识和实体间的桥梁
 - 更理智的智能
- 大规模的知识推理
 - 发现隐含知识
 - 可解释的智能（可推理数据缺失/稀疏）

• E.G. 百度知识图谱



- 实体图谱entity graph：用于精准回答
- 关注点图谱attention graph：用于文本理解
- 意图图谱 intent graph：用于对话

• E.G. IBM Watson问答系统

- 开放域
- 复杂语言
- 高精度
- 高置信度
- 高速问答

• 句子分析 “who is the 44th president of united states?”

- 问题分析
 - 答案类型：人
 - 焦点：who
 - 关键词：44th 总统 美国
- 假设生成
 - 给定问题 生成候选答案
 - 基于证据收集和假设评分模块给候选答案打分
- 候选生成
 - 搜索结构化/非结构化资源 寻找证据
 - 计算候选答案的正确概率
- 打分
 - 每个证据贡献支持或者反对
 - 综合得分

- 其他知识图谱的应用举例

- 高考机器人

- 金融知识图谱

- 金融知识的搜索和智能问答
 - 金融事件关联分析
 - 辅助信贷审查 整合多元信息
 - 金融反欺诈 发现诈骗团伙间隐藏关系
 - 商业情报知识图谱

- 医学知识图谱

- 辅助医疗诊断
 - 医学知识服务
 - 辅助抑郁症治疗
 - 中医药知识图谱

- 学术知识图谱

- 学术数据源
 - 学术知识图谱本体层
 - schema局部，其他还包括学术活动、论文信息等
 - schema可根据引用扩展或裁剪
 - 学术知识图谱实例层
 - 学者和论文的智能搜索
 - 问题专家搜索

- 唐诗知识图谱

- 唐诗的语义搜索
 - 诗人图谱
 - 作品热点图谱
 - 诗人轨迹图
 - 诗人社交网络图

- 构建知识图谱是一项复杂的系统工程

应用层	语义搜索	智能问答	辅助决策		
服务层	知识存储	知识检索	知识推理	知识演化	
知识层	知识图谱	知识抽取	知识融合	知识众包	知识表示
数据层	非结构化数据	结构化数据库	半结构化数据	多媒体数据	多模态数据

- 知识层的支撑技术：自然语言处理、机器学习、深度学习、图数据库、分布式计算等

- 知识抽取与挖掘：任务和问题

挖掘层	表示学习	实体链接	实体消解	链接预测	
知识层	实体识别	类型推断	本体构建		
信息层	实体识别	关系抽取	事件抽取	数据爬取	信息抽取
数据层	非结构化数据	结构化数据库	半结构化数据	多媒体数据	多模态数据

- **Summary**

- 知识图谱是一项实用的人工智能前沿技术,是认知智能的核心技术之一
- **知识图谱的作用:**
 - 灵活集成异构数据
 - 建立数据间关联
 - 实现实体链接
 - 进行大规模知识推理
- **知识图谱的应用场景**
 - 精准回答
 - 人机互动
 - 数据整合
 - 智能决策
- **知识图谱的门槛**
 - 复杂的技术门槛
 - 知识获取和处理门槛
 - 应用门槛