

从人工智能到知识图谱

从人工智能到知识图谱

人工智能

聪明的AI

有学识的AI

AI的发展和未来

知识

知识表示

KR的用途

知识抽取

主要方法

知识推理

知识融合

人工智能

聪明的AI

能力体现在学习能力中，算法上的提升

- 感知
- 识别：图像识别、语音识别
- 判断：控制、规划

有学识的AI

能力主要体现在推理能力

- 如何利用小数据进行深度的分析
- 缺失数据时也能更好的处理问题（这也是人有别于其他机器的地方）

AI的发展和未来

- 联结 + 符号
- 学习 + 推理
- 感知 + 认知

知识

人类的自然语言、绘画创作、数学语言、物理模型、化学公式等都是人类知识的表示形式和传承方式

AI的核心是研究怎样用计算机易于处理的方式表示各种各样的知识

知识表示

(KR) 用易于计算机处理的方式描述人脑的知识

- **理解：**
 - 不是数据格式（不是json、xml）
 - 不等同于数据结构（不是链表、树）
 - 也不是编程语言
- **数据与知识的区别：** KR支持推理
 - 有明晰的语法和语义
 - 以及基于语法上各种操作实现语义层面的推理
- **语义网：** 链接文本 -> 链接数据
 - 文本有歧义性，对计算机太复杂

KR的用途

1. **标示 | Entity ID：** 对事物进行标示，并使得这个表示有唯一性、无歧义性
2. **概念建模：** 描述事物的属性、特征
3. **支持推理**的表示基础
4. 高效计算的**数据结构**
5. **人可以理解**的机器语言

知识抽取

1. 各种非结构化文本数据
2. 文本预处理
3. 分词：词性标注、句法分析、依存分析
4. NER命名实体识别、Entity Linking实体链接
5. 关系抽取、事件抽取
6. KR：三元组、多元关系、模态识别.....

主要方法

- 知识工程
 - 正则表达式
 - 模板匹配
 - BootStrap：从种子数据中识别Patterns，用于抽取更多数据，提取更多Patterns
 - 规则约束
 - POS、NER约束、距离约束

- 基于本体的抽取
 - 知识挖掘（推理）
- 基于模型的抽取
 - 模型
 - SVM
 - 条件虽机场
 - 循环神经网络
 - 训练
 - 有监督学习
 - 无监督局类
 - 远程监督：从已有知识库获取少量训练数据

知识推理

基于已知事实推出未知事实的计算过程

- 解决方法
 - 基于描述逻辑
 - 基于规则挖掘
 - 基于概率逻辑
 - 基于表示学习与神经网络
- 推理类型
 - 缺省推理
 - 连续变化推理
 - 空间推理
 - 因果关系推理

知识融合

在不同数据集中找出同一实体的描述记录，形成更加全面的实体信息

- 将数据稍作修改，但是仍然认为二者是相同的数据