

面向知识图谱的知识推理研究进展阅读

面向知识图谱的知识推理研究进展阅读

摘要

背景

知识推理简介

基于传统方法的推理

 基于传统规则的推理

 基于本体推理

单步推理

 基于分布式表示的推理

 基于神经网络的推理

 混合推理

多步推理

 基于规则的推理

 基于分布式表示的推理

 基于神经网络的推理

 混合推理

知识推理的应用

摘要

面向知识图谱的知识推理旨在根据已有的知识推理出新的知识或识别错误的知识.不同于传统的知识推理

分类:

- 根据推理类型
 - 单步推理
 - 多步推理
- 根据方法
 - 基于规则的推理
 - 基于分布式表示的推理
 - 基于神经网络的推理
 - 混合推理

背景

在当今互联网、物联网、云计算等技术不断发展的环境下,各类应用层出不穷,因此产生了海量的数据资源,其中包含大量有价值的知识.这吸引了许多研究人员对其进行深入挖掘和分析.如何组织表达这些知识,以便作进一步的计算和分析备受关注.知识图谱应运而生,将实体(包括概念、属性值)表示成图上的节点,节点之间的连边对应实体之间的关联关系,以一个网络化的结构表征所获得的知识,清晰、直观。

应用： KnowItAll[5]、YAGO[6][8]、DBpedia[9]、Freebase[10]、NELL[11]、Probase[12]等.这些知识图谱从大量数据资源中抽取、组织和管理知识,希望为用户提供能够读懂用户 需求的智能服务,例如理解搜索的语义,提供更精准的搜索答案。

知识推理简介

- 包括两种知识

- 已知的知识

(三段论: 大前提 - 小前提 - 结论)

例. 如果已知知识 p_1 的父亲是 p_2 , p_2 的母亲是 p_3 , 那么可以推出新知识 p_1 的祖母是 p_3 .

- 由已有的知识推出的新知识

- 传统的推理

- 演绎推理

- 归纳推理

- 默认推理

- 知识图谱补全:

- 实体预测: 给定三元组中任意两个元素, 试图推理出缺失的另外一个元素

- 关系预测: 给定头实体和尾实体, 找出与之形成有效三元组的关系

基于传统方法的推理

基于传统规则的推理

在知识图谱上运用简单规则或统计特征进行推理

知识图谱内部的推理组件采用一阶关系学习算法进行推理. 推理组件学习概率规则, 经过人工筛选过滤后, 带入具体的实体将规则实例化, 从已经学习到的其他关系实例推理新的关系实例

基于本体推理

利用更为抽象化的本体层面的频繁模式、约束或路径进行推理.

基于在知识图谱中的正确性统计, 计算每个候选的得分, 用大于阈值的候选作为规则补全知识图谱

- 优: 当规则、统计特征、本体频繁模式、本体约束/路径有效时准确率高

- 缺:

- 无论是规则还是抽象层面的本体约束, 都需要实例化, 可计算性比较差, 对于实例数量 很大的知识图谱而言, 代价很高.

- 有效并且覆盖率广的规则和本体约束难以获得, 导致推理结果的召回率通常比较低

- 统计特征过分依赖已有数据, 不易迁移, 难以处理样本稀疏的情况

- 当数据存在噪声时, 抽取的特征甚至误导推理. 因此, 面向知识图谱的知识推理逐渐发展出独有的具体推理方法.

单步推理

用直接关系即知识图谱中的事实元组进行学习和推理

基于分布式表示的推理

首先通过表示模型学习知识图谱中的事实元组,得到知识图谱的低维 向量表示;然后,将推理预测转化为基于表示模型的简单向量操作.基于分布式表示的单步推理包括基于转移、 基于张量/矩阵分解和基于空间分布等多类方法.

- 基于转移的表示推理
- 基于张量/矩阵分解的表示推理
- 基于空间分布的表示推理

基于神经网络的推理

利用神经网络直接建模知识图谱事实元组,得到事实元组元素的向量表 示,用于进一步的推理。该类方法依然是一种基于得分函数的方法,区别于其他方法,整个网络构成一个得分函 数,神经网络的输出即为得分值

用双线性张量 层代替传统的神经网络层,在不同的维度下,将头实体和尾实体联系起来,刻画实体间复杂的语义联系.其中,实 体的向量表示通过词向量的平均得到,充分利用词向量构建实体表示.具体地,每个三元组用关系特定的神经网 络学习,头尾实体作为输入,与关系张量构成双线性张量积,进行三阶交互,同时建模头尾实体和关系的二阶交 互.最后,模型返回三元组的置信度

如果头尾实体之间存在该特定关系,返回高的得分;否则,返回低的得分. 特别地,关系特定的三阶张量的每个切片对应一种不同的语义类型.一种关系多个切片可以更好地建模该关系 下不同实体间的不同语义关系

- 优: 基于神经网络的单步推理试图利用神经网络强大的学习能力建模知识图谱事实元组,获得很好的推理能力和泛化能力
- 缺: 神经网络固有的可解释性问题也依然存在于知识图谱的应用中,如何恰当地解释 神经网络的推理能力是一大难点

混合推理

充分利用不同方法的优势,例如基于规则推理的高准确 率、基于分布式表示推理的强计算能力、基于神经网络推理的强学习能力和泛化能力.

- 混合规则与分布式表示的推理
- 混合神经网络与分布式表示的推理

多步推理

多步关系是一种传递性约束

a 和 b 存在关系 r_1 , b 和 c 存在关系 r_2 ,该两步路径对应的直接关系是 a 和 c 存在 关系 r_3 .

多步关系的引入,建模了更多信息,往往比单步推理效果更好

基于规则的推理

- 基于全局结构的规则推理
- 引入局部结构的规则推理

基于分布式表示的推理

基于神经网络的推理

- 神经网络建模多步路径的推理
- 神经网络模拟计算机或人脑的推理

混合推理

- 混合PRA与分布式表示的推理
- 混合规则与分布式表示的推理
- 混合规则与神经网络的推理

知识推理的应用

- 现有的知识图谱由于数据来源的不全面以及知识获取的遗漏,不可能构建完备的知识图谱.通常,代表性的知识图谱中有 69%~99%的实体缺少至少一个属性信息三元组

例如,在 Freebase 中,93.8%的人没有出生地信息,78.5%的人没有国籍信息

解决办法之一就是通过知识推理方法,利用知识图谱中已有的知识去推理出新的事实(即隐含的知识),从而尽可能地对知识图

- 知识图谱自身而言,由于数据来源的噪声以及抽取过程的不准确,内部也存在噪声知识和知识矛盾现象

在NELL中,采集到的知识正确率随着时间推移不断下降,第一个月后,正确率达到0.9,第二个月后,正确率将为0.71

主要原因是知识导出的抽取模板不可靠,导致抽取错误的知识,错误的知识用于产生更多不可靠的模板。如此循环.NELL 也周期地使用人工监督移除不正确的事实元组,然而人工标注代价很高.这就需要通过知识推理方法,自动且高效地完成这一过程.

- 垂直搜索
- 智能问答
- 机器翻译
- 疾病诊断
- 金融反诈骗
- 数据异常分析

在垂直搜索领域,在国外,Google 公司提出了 Knowledge Graph 与 Knowledge Vault,Facebook 推出了 Graph Search,微软推出了 Bing Satori;在国内,搜狗提出了知立方,百度推出了中文知识图谱搜索.知识推理能够更好地理解用户的搜索意图,提供接近“专、精、深”的垂直搜索,回答复杂的推理问题.例如 Google 的搜索,在 Google 搜索引擎输入查询,搜索引擎在利用知识图谱直接给出推理得到的精确回答的同时,在搜索结果的右侧显示该 词条的深层信息.百度的搜索在知识图谱的支持下,也能更好地理解用户的搜索意图,类似地返回推理的精确答案,附带信息来源.

在智能问答领域,IBM的Watson、Google的Google Now、苹果公司的Siri、亚马逊的Alexa、微软的小娜和小冰以及百度的度秘等是近期代表性的智能问答系统.这些系统基于知识图谱的知识推理,提供精确、简洁的答案.例如,IBM 研发的超级计算机 Watson,2011 年,Watson 在美国知识竞赛节目“危险边缘 Jeopardy!”中上演人机问答大战,战胜了人类冠军选手 Ken 和 Brad.节目中的问题覆盖面广,参赛者需具备各个领域的知识,能够解析和推理隐晦含义、反讽与谜语等.这其中,面向知识图谱的知识推理发挥了重要作用.