

基于知识图谱的知识推理 研究综述汇报

——张志强 2022.05.27

论文依据

- Xiaojun Chen, Shengbin Jia, Yang Xiang “A review: Knowledge reasoning over knowledge graph ” Expert Systems With Applications (SCI 1区) 2018
- Guan SP, Jin XL, Jia YT, Wang YZ, Cheng XQ. Knowledge reasoning over knowledge graph: A survey. Ruan Jian Xue Bao/Journal of Software, 2018,29(10):(in Chinese)

研究动机

- 面向知识图谱的知识推理是知识图谱的研究热点之一,已在**垂直搜索、智能问答**等应用领域发挥了重要作用.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,由于知识图谱中知识表达形式的简洁直观、灵活丰富,面向知识图谱的知识推理方法也更加多样。尽管已有诸多知识图谱综述文献,但仍然**缺乏对面向知识图谱的知识推理研究进行系统、深入地梳理与总结的工作**.

Venue Google Scholar Top 300	All 300	I 164	Selected 66
ACL	39	23	7
EMNLP	54	49	17
NAACL	46	27	5
ISWC	60	42	3
CIKM	45	29	4
AAAI	36	36	17
IJCAI	41	41	7
ICML	15	14	4
NIPS	34	34	6
WWW	30	26	6
ICLR	5	5	3
COLING	9	9	2
Conference	414	335	81
All	714	499	147

研究框架

➤ 对面向知识图谱知识推理的最新研究进展进行归纳总结,并展望未来发展方向和前景, 具体研究框架如下

- 知识推理简介
- 基于知识图谱的传统知识推理
- 基于知识图谱的单步推理
 - 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
- 基于知识图谱的多步推理
 - 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
- 知识图谱推理的应用
- 总结与未来研究展望

知识推理简介

- 知识推理的基本概念

推理是人们对各种事物进行分析、综合和决策,从已知的事实出发,通过运用已掌握的知识,找出其中蕴含的事实,或归纳出新的事实的过程. 严格地说,就是按照某种策略由已知判断推出新的判断的思维过程。总的来说,知识推理就是利用已知的知识推出新知识的过程.

- 传统的知识推理

传统的知识推理方法主要是基于逻辑、规则的推理,逐渐发展为最基本的通用推理方法,除了一般的短语和句子级的推理之外,另一大类受到广泛关注的推理是本体推理.本体是共享概念的模型 明确的形式化规范说明

知识推理简介

- 面向知识图谱的知识推理

- 面向知识图谱的知识推理旨在基于已有的知识图谱事实,推理新的事实或识别错误知识. (补全知识图谱和知识图谱去噪)
- KG 上的知识推理定义: 给定一个知识图 $KG = \langle E, R, T \rangle$ 和关系路径 P , 其中 E , T 表示实体集合, R 表示关系集合, R 中的边连接两个节点形成三元组 $(h, r, t) \in T$, 生成一个在 KG 中不存在的三元组 $G' = \{(h, r, t) | h \in E, r \in R, t \in T, (h, r, t) \notin G\}$.
- 其目标是使用机器学习方法推断实体对之间的潜在关系, 并根据现有数据自动识别错误知识, 以补充 KG .

基于知识图谱的传统知识推理

- 基于传统规则推理的方法

基于传统规则推理的方法主要借鉴传统知识推理中的规则推理方法,在知识图谱上运用简单规则或统计特征进行推理

- 基于本体推理的方法

基于本体推理的方法主要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。

总结：面向知识图谱的知识推理可以借鉴传统的知识推理方法,特别是本体推理方法.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。但是具有很多缺点：**可计算性比较差**,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.

基于知识图谱的单步推理

- 单步推理是指用**直接关系**即知识图谱中的**事实元组**进行学习和推理,根据所用方法的不同,具体可分为**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**以及**混合推理**.
- 其中,基于规则的单步推理主要借鉴传统的规则推理方法.鉴于这些方法的**可计算性差**、**代价高**等问题,基于分布式表示的推理、基于神经网络的推理以及混合推理得到了更广泛的关注和发展.

基于分布式表示的推理

单步推理中,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

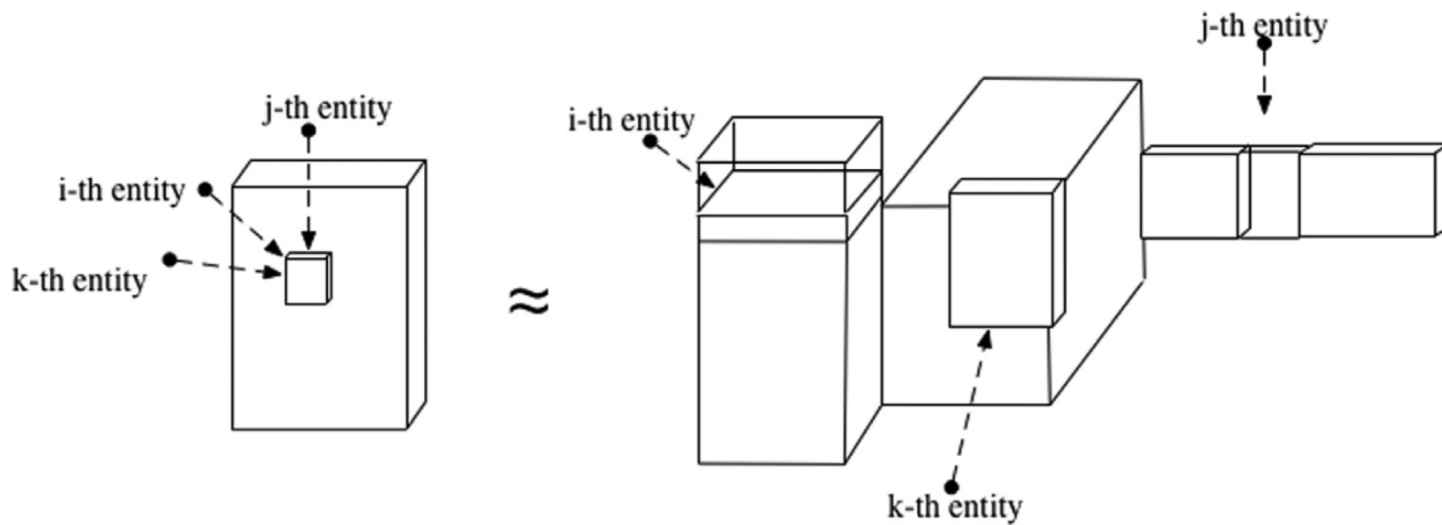
基于转移的表示推理根据转移假设设计得分函数,衡量多元组有效的可能性,得分越高,多元组越可能有效.也即正例元组的得分高,负例元组的得分低.

其中最经典的方法是基于转移的表示模型 TransE, 也掀起了一系列研究热潮, 主要思想是:如果三元组(头实体,关系,尾实体)成立,头实体向量 h 与关系向量 r 的和与尾实体向量 t 相近,否则远离.

基于分布式表示的推理

- 基于张量/矩阵分解的表示推理

基于张量/矩阵分解的表示推理将(头实体,关系,尾实体)三元组看成张量/矩阵中的元素构建张量/矩阵,通过张量/矩阵分解方法进行表示学习.分解得到的向量表示相乘重构成张量/矩阵,元素值即为对应三元组有效 与否的得分,可以认为得分大于特定阈值的三元组有效,或候选预测按照得分排序,选择得分高的候选作为推理结果.



基于分布式表示的推理

- 基于空间分布的表示推理

基于空间分布的表示推理建立模型拟合知识图谱中实体和关系的空间分布特征,使得在向量表示空间中,实体和关系的空间分布尽可能地与原知识图谱一致.该方法通过设计对应的能够反映空间分布特征的**得分函数**,采用与基于转移方法类似的学习和推理过程。

总结：基于分布式表示的单步推理研究工作比较全面,然而建模时通常只考虑满足知识图谱事实元组的约束,未考虑更深入的组合语义信息,推理能力受限.

- 基于转移的表示推理方法简单、有效,计算效率高；
- 基于张量/ 矩阵分解的表示推理,虽然可解释性强,性能相对较好,但时间复杂度较高,难以推广应用；
- 基于空间分布的表示推理很好地拟合了知识图谱中三元组的空间分布,推理能力强

基于神经网络的推理

- 单步推理中,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.

总结：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.然而,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

单步推理中,混合推理通过混合多种单步推理方法,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
 - 在分布式表示辅助规则发现方面,一些传统的推理规则发现方法通过计算关系之间的分布式相似度实现,其中,关系表示为对应实例的特征向量,如上下文敏感的推理规则发现方法
 - 在规则辅助基于分布式表示的推理方面,如在表示模型中无缝嵌入逻辑规则和物理规则,形式化推理为整型线性规划问题(ILP).

混合推理

- 混合神经网络与分布式表示的推理

- 一种是用神经网络方法建模外部信息,例如相关的外部文本、实体描述等,表示模型建模知识图谱中的三元组;
- 另一种是用神经网络方法建模知识图谱,其输出进一步用于表示模型

总结：混合单步推理通过混合不同单步推理方法实现**优势互补**。然而,目前的混合单步推理还停留在两种方法的**浅层混合**,即:以一种方法为主,另一种为辅的推理,尚缺乏更深层次的混合模式以充分利用各方法的优势。同时,混合推理方法不同于其他具体的某类方法,它更像是一种**策略**,主要基于对各类方法的透彻分析找到优势互补的方法,进行混合,可创新的点在选择哪些方法进行混合以及混合模式上

基于知识图谱的多步推理

- 多步推理是在单步推理建模直接关系的基础上进一步建模间接关系,即多步关系.多步关系是一种传递性约束,

实例: 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.

多步推理按不同的推理方法划分,同样分为基于规则的推理、基于分布式表示的推理、基于神经网络的推理以及混合推理.

基于规则的推理

多步推理中,基于规则的推理不同于基于规则的单步推理,后者用到的是类似关系 r_1 推出关系 r_2 的简单经验规则或一些基于统计的频繁模式.而多步推理用到的规则更为复杂,如传递性规则.鉴于人工获取有效且覆盖率广的传递性规则代价比较高,这些规则一般通过挖掘的实体间路径来近似.

- 基于全局结构的规则推理

基于全局结构的规则推理在整个知识图谱上进行**路径挖掘**,将一些路径近似地看成规则.实体间的路径进一步作为判断实体间是否存在**指定关系的特征**训练学习模型.这类工作的一大部分内容是**基于随机行走的规则挖掘**,其他的还可以用知识图谱上的其他规则挖掘模型获取规则,包括**整型线性规划**、**关联规则挖掘**等.

基于规则的推理

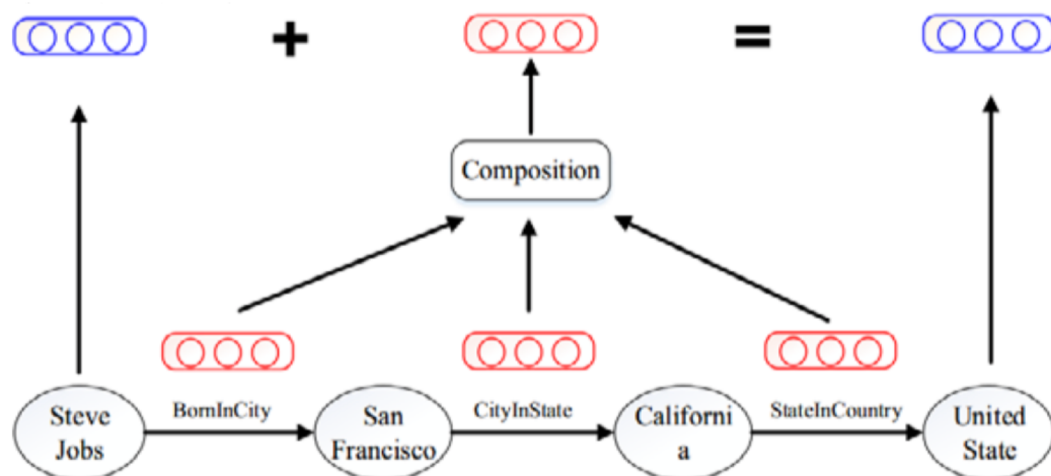
- 引入局部结构的规则推理

在整个知识图谱上进行推理,一方面代价往往比较高,而知识的推理一般主要与局部区域的知识更相关,因此可以在局部结构上进行推理;另一方面,全局结构信息粒度比较粗,结合细粒度的局部结构信息将有助于推理.

总结：基于规则的多步推理不同于基于规则的单步推理采用简单经验知识或统计特征进行推理,所用规则为以传递性约束为主的更复杂的规则或特征,一般通过知识图谱上的规则挖掘算法得到,避开了人工构建规则的高代价.但是,挖掘的规则即使通过一定的过滤机制进行筛选,仍然不可避免地引入了一些信息量很低、作用很小、甚至是噪声的规则,误导了推理.并且,目前使用的规则主要是传递性约束,有待进一步挖掘更多样、更有效的复杂规则

基于分布式表示的推理

- 多步推理中,基于分布式表示的推理与基于分布式表示的单步推理类似,都是通过向**量化知识图谱进行推理**.不同的是,多步推理在学习向量表示的过程中引入了**多步关系约束**,使得学到的向量表示更有助于实体和关的推理预测



总结：基于分布式表示的多步推理主要是在基于分布式表示的单步推理基础上加入多步关系的建模,采用补充建模或共同建模的方式.补充建模的方式以单步关系(直接关系)为主,多步关系辅助学习,用于调整空间中向量的位置.而共同建模的方式同等地对待直接关系和间接关系,获得更好的向量表示,但可能引入级联误差,即,路径上中间步骤关系建模引起的误差将一直传递累计到最终结果

基于神经网络的推理

多步推理中,基于神经网络的推理旨在用神经网络建模学习多步推理过程,包括建模多步路径以及模拟计算机或人脑的推理.

- 神经网络建模多步路径的推理

该类方法用神经网络建模路径,充分学习多步路径的向量表示,得分函数关联于路径的表示与直接关系表示的相似度,希望正例对应的相似度大,即乘积大,负例小。可以使用RNN或者强化学习来解决

- 神经网络模拟计算机或人脑的推理

利用神经网络强大的学习能力,模拟计算机或人脑的知识存储和处理方式.一般用一个存储结构模拟人脑的存储记忆,用一个控制器模拟人脑的控制处理中心.通过对知识图谱中已知三元组的学习记忆,希望神经网络能够具有人脑的推理能力,推理出新的三元组

基于神经网络的推理

总结：基于神经网络的多步推理直接建模多步路径或推理过程,相比于基于神经网络单步推理,研究工作更丰富,可解释性更强,效果更好.然而,可解释性需要进一步增强.

- 直接建模多步路径的方式,由于多步路径可以看成是关系或关系和实体的一个序列,主要通过 RNN 建模
- 直接建模推理过程的方式可以模拟人的学习和推理过程,而人具备强大的学习推理能力,因此这将是一个很热门的研究方向.并且,人的学习和推理过程相对复杂,使其也成为具有很大的挑战性的问题.

混合推理

多步推理中,混合推理通过混合不同多步方法进行推理,实现优势互补.分布式表示方法由于其计算的便捷性,通常被用于与其他方法混合.混合多步推理具体包括混合 PRA 与分布式表示、混合规则与分布式表示以及混合规则与神经网络的推理.

- 混合 PRA 与分布式表示的推理

混合 PRA 与分布式表示的推理主要用分布式表示方法从相关文本语料中学习连接实体的关系隐性表示, 增广知识图谱,降低知识图谱的不连通性和稀疏性,帮助路径发现过程,从而提升推理效果

- 混合规则与分布式表示的推理

在基于规则分布式表示学习的推理方面, 提出通过矩阵分解直接学习一阶逻辑规则的表示; 在图分布式表示作为先验的推理方面, 提出通过矩阵分解直接学习一阶逻辑规则的表示; 在共同建模规则和知识图谱的推理方面, 提出了共同表示知识图谱和逻辑规则的模型

混合推理

- 混合规则与神经网络的推理

混合规则与神经网络的推理主要将规则转化为向量操作,应用于强学习能力的神经网络方法中,实现一个可微的模型。

总结：混合多步推理相比于混合单步推理内容更为丰富,其中,混合规则和图分布式表示的推理得到了比较全面的研究,产生了更有效的混合模式:共同建模,取得了更好的效果.但目前的混合推理依然局限于两种方法的混合.各类混合方法中,混合规则和神经网络的推理具有很大的发展空间,规则方法的高准确率和可解释性以及神经网络方法的强学习和高泛化能力,使得二者的结合可以得到高准确率的微模型,避开了传统规则方法的计算难题,一定程度上也增加了神经网络方法的可解释性

知识图谱推理的应用

知识图谱推理方法从现有的三元组中推断未知关系，不仅为大规模异构知识图谱中的资源提供了高效的相关性发现能力，而且还完善了知识图谱。一致性推理等技术确保了知识图谱的一致性和完整性。推理技术可以通过对领域知识和规则建模来进行领域知识推理，从而支持自动决策、数据挖掘和链接预测。由于强大的智能推理能力，知识图谱可以广泛用于许多下游任务。

- 知识图谱补全

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^[149].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^[150].解决办法之一就是知识推理方法,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谱进行补全.

知识图谱推理的应用

- 知识图谱去噪

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^[151].例如在 NELL 中,采集到的知识正确率随着时间推移不断下降,第 1 个月后,正确率达到 0.9;第 2 个月后,正确率降为 0.71^[17].主要原因是知识导出的抽取模版不可靠,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版,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.

- 医疗领域

目前,医学领域已经成为知识图谱被积极使用的领域,也是人工智能领域的研究热点。当应用于医学知识图谱时,知识推理方法可以帮助医生收集健康数据、诊断疾病和控制错误

知识图谱推理的应用

- 互联网金融

金融也是使用知识图谱的一个活跃领域。知识图谱中的投资关系和雇佣关系可以通过聚类算法来识别利益相关者群体。当部分节点发生变化或发生大事件时，可以通过路径排序和子图发现方法推断变化实体之间的关联。在金融行业，反欺诈是一项重要工作。通过知识推理，人们可以验证信息的一致性，从而提前识别欺诈行为

- 智能问答系统

基于知识库的问答（KBQA）分析查询问题，然后从知识库中找到答案。但是，KBQA 也需要推理技术的支持，因为知识图谱是不完整的。例如，Watson 在 Jeopardy 中击败了人类，其中知识推理发挥了重要作用。Jeopardy 的问题涵盖各个领域，要求考生分析和推理蕴涵、讽刺和谜语。智能问答系统，如苹果的 Siri、微软的 Cortana、亚马逊的 Alexa，都需要知识图谱推理的支持。

知识图谱推理的应用

- 推荐系统

基于知识图谱的推荐系统将用户和物品连接起来，可以整合多个数据源来丰富语义信息。可以通过推理技术获得隐含信息，以提高推荐的准确性。基于知识图谱推理方法的推荐有几个典型案例，如购物推荐、电影推荐和音乐推荐。

- 其他应用

知识推理技术在其他一些智能场景中也发挥着重要作用。例如，知识推理技术可用于了解用户在搜索引擎中的查询意图。此外，它还可用于其他计算语言学任务，例如抄袭检测、情感分析、文档分类、口语对话系统。

总结与未来研究展望

- 知识推理总结

类别	子类	方法描述	方法优点	方法缺点
单步推理	基于规则的推理	利用简单规则、统计特征等进行推理	可解释性强; 规则正确时,准确率高	规则不易获得,不全面,导致推理结果召回率低;统计特征不易迁移,当数据存在噪声时,可能误导推理;需要一个一个地推理候选实例,可计算性比较差
	基于分布式表示的推理	利用知识图谱表示模型,得到知识图谱的低维向量表示,通过向量操作进行推理预测	计算方便快捷	通常只考虑满足知识图谱事实元组的约束,未考虑更深入的组合语义信息,推理能力受限
	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	推理能力较强	相比基于分布式表示的推理,复杂度更高,可解释性更弱
	混合推理	混合多种单步推理方法	不同推理方法优势互补	目前还停留在浅层的混合,即:以一种方法为主,另一种为辅的推理,缺乏更深层次的混合模式
多步推理	基于规则的推理	通过挖掘全局或局部复杂规则,主要是传递性规则进行推理	自动挖掘复杂规则,而不是通过专家构建	挖掘的规则不可避免地存在噪声,误导推理
	基于分布式表示的推理	向量化事实元组中实体和关系的同时,建模多步关系约束	获得知识图谱更好的向量表示,推理预测能力提高	可能存在级联误差
	基于神经网络的推理	利用神经网络学习多步推理,直接建模多步路径或推理过程	表达能力丰富,推理能力强	需要进一步增强可解释性
	混合推理	混合多种多步推理方法	充分利用不同多步推理方法的优势,出现了更有效的混合模式:共同建模,推理能力强	局限于两种方法的混合

总结与未来研究展望

- 面向多元关系的知识推理

相比于一元与二元关系,多元关系结构多样、语义复杂并且难以处理.因此,现有的知识处理方面的工作主要集中在二元关系未来有必要研究如何进行多元关系的形式化表达(representation)与表示学习(embedding)以及如何兼顾表达能力与推理能力.表达能力越强,表达结构越复杂,灵活性越差,推理能力相应地越弱.

- 融合多源信息与多种方法的知识推理

融合多源信息的知识推理,通过结合文本语料或其他知识图谱,利用更多的额外信息降低知识图谱的不连通性和稀疏性,进行有效推理.融合多种方法的知识推理,通过在更深层次混合不同方法,优势互补,提升推理性能。如何同时融合多源信息与多种方法,进一步提升推理性能,也将成为未来的一大研究方向.其中,融合模式,即以什么方式融合,是一大难点.

总结与未来研究展望

- 基于小样本学习的知识推理

现有的知识推理模型往往需要大量高质量的样本进行训练学习,这需要耗费很大的代价去获取样本.在实际应用中,甚至难以获得大量的训练样本,极大地限制了现有知识推理模型的应用范围.另一方面,人往往凭借相关先验知识只需少量样本就能快速学习推理.基于小样本学习的知识推理将成为未来研究的重要方向,即:如何以已建立的知识图谱(如 YAGO[6-8]、Freebase[10]和 NELL[11])等为先验,通过少量高质量样本进行新的学习和推理.

- 动态知识推理

目前的知识推理方法主要针对静态知识图谱进行.然而,知识图谱随着时间的推移往往动态变化,包括增、删和改操作,例如,“谁是现任中华人民共和国主席?”在 2008 年 3 月~2013 年 3 月期间,该问题的答案是胡锦涛,2013 年 3 月至今,该问题的答案是习近平.因此,如何进行动态知识推理成为未来的一个研究热点。