

第 5 章

知识图谱概述

5.1 知识图谱的定义与分类

5.1.1 知识图谱的定义

2012 年,为了加强智能搜索,Google 提出了知识图谱(knowledge graph,KG)的概念。知识图谱是一种关系图谱,能够将不同种类的知识关联到一张图谱上。知识图谱的本质是语义网络的一种,其中,节点代表实体(entity)或概念,而边则代表它们之间的语义关系。知识图谱涵盖了语义网的一些特征,但具有更加广泛的数据来源,涵盖了知识表示、知识融合等多种技术。与知识库相比,尽管知识图谱具有与之相似的理论与方法,但涵盖了更为广泛的知识集合。

5.1.2 知识图谱的分类

知识图谱的分类方式有很多,按照知识图谱的覆盖范围,可以将其划分为通用知识图谱和领域知识图谱。一般来说,通用知识图谱的构建面向通用的领域,涵盖结构化的知识库和常识性知识,覆盖范围广泛;而领域知识图谱,也称为垂直知识图谱或行业知识图谱,一般面向特定领域,应用各种语义技术,为某一具体行业建立知识库,对领域知识的深度、标准化有较高的要求。如图 5.1 所示为当前常见的知识图谱。

而按照知识图谱研究内容,可以将知识图谱划分为文本知识图谱、视觉知识图谱和多模态知识图谱等,如图 5.2 所示。

1. 文本知识图谱

文本知识图谱主要将文本作为研究对象,将知识表示、知识推理等技术应用于文本知识。文本知识图谱应用于多个领域,如语义检索、深入搜索、情报分析等。

2. 视觉知识图谱

视觉知识图谱主要对图像进行知识表示、知识加工、推理更新等,在语义图像检索、判断文本真假关系等方面应用广泛。

3. 多模态知识图谱

多模态知识图谱的研究对象主要是多模态样本,对多模态样本进行知识表示、知识推理更新等,主要应用于将文本与视觉结合起来的知识问答领域,应用更加广泛。



图 5.1 当前典型知识图谱



图 5.2 知识图谱按研究内容划分

5.2 知识图谱的作用与意义

1994 年图灵奖获得者费根鲍姆提出知识工程的概念,“将知识集成到计算机系统从而完成只有特定领域专家才能完成的复杂任务”。历经 40 年的发展,知识工程的演进过程经历了前知识工程时期、专家系统时期、万维网 1.0 时期、群体智能时期以及知识图谱时期等几个时期,如图 5.3 所示。知识工程经历了极具挑战的发展,也为人工智能的发展做出了可观的贡献,演变而来的知识图谱已成为人工智能的一个重要分支。



图 5.3 知识工程发展历程

知识图谱结构化地对客观世界中的概念、实体及其关系加以描述,针对互联网上海量

的信息,知识图谱能够更好地对其组织、管理和理解,并转换为人类更好理解的形式。知识图谱技术是指建立和应用知识图谱的技术,是与机器学习、知识表示与推理、融合认知计算、自然语言处理、信息检索与抽取以及语义 Web 等研究相交叉的技术。知识图谱具有高效资源发现、逻辑推理、自然关联、可解释性、透明共享及可视化等优势,有着广阔的应用前景。

知识图谱不仅能够提高互联网语义搜索的效率,在智能问答系统中也有着不可替代的作用,同时极大地推动着互联网环境下知识驱动的智能应用的发展。随着互联网、人工智能等技术的飞速发展,大数据、深度学习以及知识图谱发挥着愈加重要的作用。其中,知识图谱的应用价值愈加突出,当前知识图谱的应用主要有 4 方面,如图 5.4 所示。



图 5.4 知识图谱应用价值

1. 语义集成

语义集成,即将知识图谱融合为统一的形式,能够为基于知识图谱的不同程序间的交互提供语义上的互操作性。语义集成包含的技术主要有本体匹配、实例匹配以及知识融合。知识图谱可以通过语义标注和链接,整合分布异构的大数据,提供以知识为中心的资源语义集成服务。语义集成对于提升知识图谱的智能化水平,并推动人工智能的发展具有重要意义。

2. 语义搜索和推荐

语义搜索,即在当前关键词搜索的基础上添加实体和关系的检索,可以更准确地捕捉用户的目的,通过语义搜索,可以更精准地检索出用户需要的答案。通过语义推荐,知识图谱可以根据用户输入的搜索词为用户搜索出满足其需求的结构化信息,而不是网页信息。

3. 问答系统

问答系统是一种可以自动回答用户问题的系统,其基于用户提出的问题,以对话的形式返回精准的答案。基于知识图谱的问答系统能够自动地将用户的问题转化为要查询的内容,通过到知识图谱中查询的方式,得到用户想要的答案。问答系统被看作是未来技术服务的颠覆性技术之一。

4. 大数据分析决策

大数据技术结合知识图谱进行大数据分析,能够分析出数据中的语义信息,使得大数据分析更加智能化。通过语义链接,知识图谱可以使大数据具有洞察力,为数据注入灵魂,进而为其提供决策支撑。

5.3 知识图谱的研究进展

5.3.1 知识图谱的研究现状

当前国内外关于知识图谱的研究主要集中于知识图谱关键技术的研究,知识图谱的关键技术主要包含知识抽取技术、知识表示技术、知识融合技术与知识推理技术,如图 5.5 所示。



图 5.5 知识图谱关键技术

1. 知识抽取技术

构建知识图谱的数据来源于文本、图像、视频等,而抽取方式主要有两种,一是从网页上获取,二是从数据库等数据集中抽取。知识图谱构建的关键技术主要涵盖怎样从各种数据源中抽取所需元素,如实体、关系、属性等,在构建知识图谱时,抽取的知识越完整,所构建的图谱就越全面,从而应用价值也越高。其中,常见的实体抽取方法主要包括基于规则、基于

统计机器学习和基于开放域的抽取方法。

2. 知识表示技术

知识表示是指使用符合逻辑表达的数据表达方式,对现实世界中的实体及其之间的相关关系进行建模,以提高人与计算机之间的沟通效率。目前常用的知识表示方法主要有基于语义网络,基于产生式规则和基于框架、基于逻辑、基于语义本体等表示方法,而对基于逻辑的表示方法研究较多。

3. 知识融合技术

从各个数据源获取知识后,本体被用来提供统一术语,以将不同来源的知识进行融合,从而生成一个大型知识库,并可以让用户灵活地建立或修改数据模型。在不同本体之间,往往会存在同一类数据用不同术语表示的情况,此时就需要本体融合技术来对这些本体进行融合,以解决此类问题。本体融合的基本流程如图 5.6 所示。通过本体解析、特征选取、相似度计算、结果表达、本体评价等流程,融合后形成的知识库往往需要一个存储和管理的解决方案。根据用户查询场景不同,所采用的存储管理方案往往不同,一般的存储架构包



图 5.6 本体融合流程

括 NoSQL 和关系数据库两种。大规模知识库往往具有大数据特征,因此可能需要传统大数据平台如 Spark 或者 Hadoop 来提高计算能力和性能以及支持快速运算。

4. 知识推理技术

知识推理往往应用于知识图谱数据来源不全面或抽取过程不准确的情况,知识推理可以从已有的知识图谱中进一步挖掘缺失的或者更深层次的实体与关系间的联系。其方法主要包括:

(1) 基于传统方法的推理,主要包括基于传统规则推理的方法和基于本体推理的方法。

(2) 单步推理,包括:①基于分布式表示的推理;②基于神经网络的推理;③混合推理。

(3) 多步推理,包括:①基于规则的推理;②基于分布式表示的推理;③基于神经网络的推理;④混合推理。

5.3.2 知识图谱的发展趋势

知识图谱具有广泛的应用价值,如图 5.7 所示,在人工智能、大数据的带动下,知识图谱在未来会有更大的发展;知识图谱与各领域、各行业相结合,也可以带动行业的发展,为各行业注入智能,未来知识图谱的更多创新应用有待开发。



图 5.7 知识图谱的发展趋势

1. 人工智能与知识图谱

人工智能追求的目标是利用机器快速地获取高质量信息,进而辅助智能化的应用,而人工智能的关键技术之一就是知识图谱,并且知识图谱是实现人工智能多元化应用的核心力量。目前,以深度增强学习为主导的通用智能发展越来越需要知识工程的辅助,知识图谱主导的可行的认知智能将成为未来战略技术的核心。

2. 大数据与知识图谱

大数据和知识图谱相辅相成,两者存在多个异同点:一是两者均是关于“结构化”和“关联”,不过大数据是关于数据结构化,而知识图谱则是关于知识结构化,并且大数据的关联在数据级别,而知识图谱的关联在知识级别;二是在应用场景上,两者一般都被应用于解决有关分析洞察的问题中,不过知识图谱更擅长处理“关系”问题,可以在解决此类问题时达到更加直观高效的效果。并且在分析和洞察问题中,知识图谱已经成为一种新的分析手段,在一些扩展能力上如设计多层次、多关系事物模型时,可突破传统数据分析的技术瓶颈。

3. 各行业与知识图谱

通过应用知识图谱技术智能化地处理海量信息,计算机能够更好地理解网络、用户、资

源等,从而最终提供新型智能化服务。在互联网行业以及传统行业中,构建并应用知识图谱具有重要意义,知识图谱的应用可以提升服务质量和效率。以搜索引擎为例,基于知识图谱的智能搜索引擎能够以理解用户意图的方式来呈现搜索结果,使搜索引擎更加“人性化”,并且使人和搜索引擎的交互更加自然。知识图谱还能升级传统行业,构建行业知识图谱,可以为传统行业注入新动能,实现行业升级。