·学术科普·

漫威电影中的知识图谱

Knowledge Graph in Marvel Cinematic Universe

王元卓　中国科学院计算技术研究所
张　琨　中国科学院计算技术研究所
陆　源　竞技世界（北京）网络技术有限公司

在电影《复仇者联盟2：奥创纪元》中，人工智能奥创引发了重大的灾难。奥创作为电影中的超级反派，有着强大的战斗力，它可以精准地分析复仇者联盟中每位英雄的各项数据、属性和历史战役，找出其各自的弱点，借此打败了多个英雄，如图1所示。其中一个可能的原因就是奥创拥有强大的知识计算能力，而知识计算的重要基础就是知识图谱。

图1　电影《复仇者联盟 2：奥创纪元》中的片段

那么，什么是知识图谱呢？

知识图谱本质上是一种揭示事物之间关系的语义网络。这么说大家可能还不太理解。那首先，我们来了解一下什么是知识。说到知识，我们需要把它与信息区分开。信息是指对外部客观事实的描述，比如说这里有一瓶水，现在温度是7℃，这就是一个信息。但知识是对外部客观规律的归纳和总结，比如说，水在0℃时会结冰。因此，在信息的基础上，建立事物之间的联系，就能形成知识。知识图谱是由一条条知识组成的网状结构，每条知识可以用两个表示事物的实体和它们之间关系的三元组来表示。

知识图谱背后的思想，最早可以追溯到二十世纪五六十年代提出的一种知识表示形式——语义网络。语义网络由相互连接的节点和边组成，其中节点表示概念或者对象，边表示它们之间的关系。

2012年，谷歌正式提出了知识图谱的概念，其初衷是优化搜索引擎返回的结果，增强用户搜索的质量和体验。假设我们想知道“互联网的发明者是谁”，搜索引擎会准确返回Tim Berners-Lee的信息，说明搜索引擎理解了用户的意图，这就是知识图谱的一个重要应用。

了解了知识谱图的发展历程，接下来，我们看一看知识图谱是怎么构建的，如图2所示。

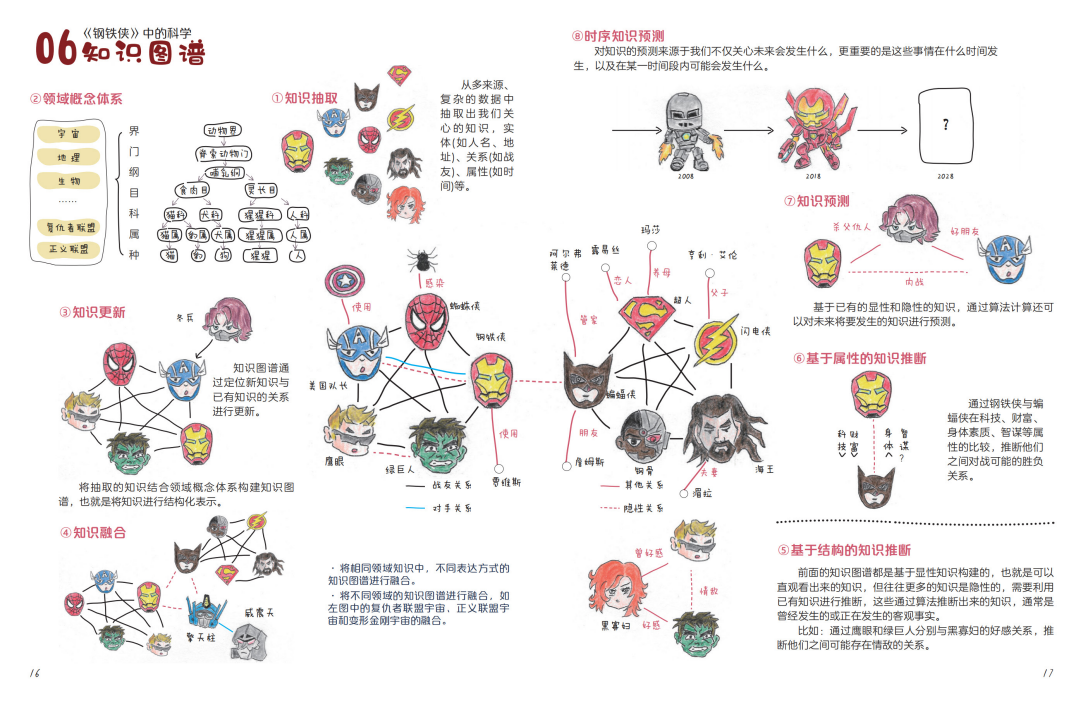

图 2　知识图谱讲解图

（选自《科幻电影中的科学：科学家奶爸的AI手绘》）

当我们构建某个领域的知识图谱时，首先，我们需要根据该领域的知识构建领域概念体系。所谓的领域概念体系实际上是知识图谱的一种逻辑结构，是描述该领域里概念和概念之间的关系的一种体系。以生物为例，大家熟知的界门纲目科属种就是一种领域概念体系。建立好领域概念体系后，我们就可以根据领域概念体系向知识图谱里补充实例化数据。比如，动物界下面有脊索动物门，脊索动物门下有哺乳纲，哺乳纲下有食肉目和灵长目，等等。然而，现实生活中，我们的数据往往来源于互联网中的海量数据，这些数据没有统一的格式，可能是半结构化或者非结构化的。这就需要我们从多来源、复杂的数据中抽取出所关心的实体、关系、属性等知识要素，这就是知识抽取技术。举例来说，在电影《复仇者联盟2：奥创纪元》中，我们从钢铁侠和美国队长的对话中可以抽取出，雷神与他们是战友关系。当发现新的知识后，我们需要把新的知识加入已有的知识图谱进行更新，这就是知识更新技术。实现这一操作需要定位新知识与知识图谱中已有知识的联系，以此进行更新。因为知识图谱中的知识来源广泛，存在知识质量良莠不齐、来自不同数据源的知识可能重复、知识间的关联不够明确等问题，所以必须要进行知识融合。知识融合大致可以分为两类：一类是在相同领域中，将不同表达方式的知识图谱进行融合；另一类是将不同领域的知识图谱进行融合，比如将复仇者联盟宇宙、变形金刚宇宙和正义联盟宇宙之间的知识进行融合。

完成上述操作后，知识图谱的构建步骤就基本完成了。那么在知识图谱的基础上我们可以实现哪些有意思的应用呢？

首先可以通过算法，让计算机对知识图谱中没有显性表达出来的知识（也就是隐含知识）进行推理，知识推理主要包括基于属性的推理和基于结构的推理两大类。那么如何通过对知识图谱中属性的计算来实现知识推理呢？举一个简单的例子，我们都知道“钢铁侠大战蝙蝠侠”是不可能实现的，但这两人的战斗力到底谁更强呢？通过知识推理的算法可以实现钢铁侠与蝙蝠侠，这两个不同宇宙中的英雄

的战斗力比较，通过算法实现科技、财富、身体素质、智谋等属性值的量化比较，计算机可以推理出他们之间对战可能的胜负关系。

此外，还可以通过基于结构的推理来计算曾经发生或正在发生的客观事实。比如，通过分析鹰眼和绿巨人分别与黑寡妇的好感关系，推理出他们之间可能存在情敌关系。虽然这些推理人脑分析起来并不费劲，但要想让计算机可以准确计算出来，还需要更多深入的研究工作。

除了对已有可能存在的知识的推理外，还可以通过算法实现对未来可能发生事件的预测。比如，在电影《美国队长2：冬日战士》中，知识图谱显性地表示出冬兵和钢铁侠是杀父仇人关系，而冬兵与美国队长是好朋友关系，那么算法就可以预测出在未来钢铁侠很可能会与美国队长发生冲突。这一点也在电影《美国队长3》中得到了验证。

进一步对知识图谱的时间属性等规律进行计算和比较，可以实现知识的时序分析，从而对未来的时序知识进行预测。比如，在相关电影中钢铁侠于2008年发布了第一款战甲马可1代。这款战甲外表比较简陋，武器装备也很简单，可抵御一般轻量级武器袭击。到2018年时，钢铁侠发布了马可85代战甲，该战甲就已经装备了高杀伤性武器，并且可抵御反派的超强进攻。那么根据战甲的知识图谱时序数据，计算机可以预测出到2028年时战甲可能具备的功能特性。

在现在的日常生活中，知识图谱已经被成功应用于很多人工智能产品中。比如，在电子商务领域，可以通过构建商品知识图谱来精准地匹配用户的购买意愿和商品候选集合；在公安刑侦领域，可以通过分析证据实体和实体之间的关系来准确推断出特定目标的隐含线索；知识图谱类技术还被广泛应用于智慧城市、智能穿戴、网络安全、智能问答等实际系统。

作者简介

王元卓（1978- ），男，博士，中国科学院计算技术研究所研究员、博士生导师，中科大数据研究院院长，中国科普作家协会副理事长，中国计算机学会科学普及工作委员会主任，主要研究方向为大数据与人工智能。

张琨（1997- ），男，中国科学院计算技术研究所博士生，主要研究方向为智能问答。

陆源（1990- ），男，现就职于竞技世界（北京）网络技术有限公司，从事数据产品工作，主要研究方向为大数据与社交网络。热心科普创作，科普畅销书《科幻电影中的科学：科学家奶爸的宇宙手绘》《科幻电影中的科学：科学家奶爸的AI手绘》作者之一。

中图分类号：TP399　　文献标志码：E　　doi: 10.11959/j.issn.2096-0271.2022083