

知识图谱补全算法综述

丁建辉 贾维嘉
上海交通大学计算机科学与工程系 上海 200240

摘要 知识图谱补全算法能让知识图谱变得更加完整，目前是人工智能领域的一个研究热点。为了更好地给出补全算法综述，文章按照能否处理新实体或者新关系，将知识图谱补全算法分成两类：静态知识图谱补全算法以及动态知识图谱补全算法。前者仅能处理实体以及关系都是固定的场景，扩展性较差。后者可以处理含有新实体或者新关系的场景，能够构造动态的知识图谱，具有更好的现实意义。文章主要结合相关工作对动态知识图谱补全算法进行系统性的总结，发现动态知识图谱算法的研究热度逐渐增加，如何更好更快地构建动态知识图谱是一个较好的研究点。

关键词 人工智能；表示学习；知识图谱；分布式表示

引言

知识图谱这个概念最早由Google在2012年提出，Google认为“things, not strings”，即对于搜索引擎，世界中的各种物体不应该仅仅是strings，而是具有实际含义的things，例如“苹果”这个词，既可以代表美国的苹果公司，也可以代表一种水果。借助知识图谱，Google的搜索引擎实现了从strings到things的变化，使得机器能更好地理解用户搜索词所代表的具体含义。知识图谱通常以高度结构化的形式表示，描述了现实世界中各种实体之间的关系^[1]，图1展示了歌手周杰伦的知识图谱。目前，知识图谱已经广泛地应用于人工智能的多个领域，例如自动问答、搜索引擎、信息抽取等。典型的知识图谱由大量结构化的三元组构成，例如(奥巴马, 国籍, 美国)，该三元组描述了“奥巴马的国籍是美国”这件事实。

基金项目：国家重点自然科学基金资助项目(61532013)。

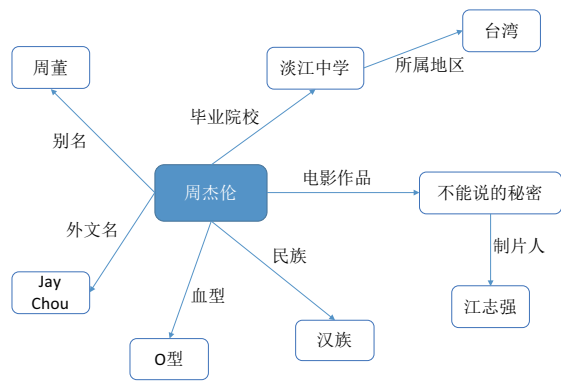


图1 周杰伦的知识图谱

虽然知识图谱能提供高质量的结构化数据，但是大部分开放知识图谱，例如Freebase^[2]、DBpedia^[3]都是由人工或者半自动的方式构建，这些图谱通常比较稀疏，大量实体之间隐含的关系没有被充分地挖掘出来。在Freebase中，有71%的人没有确切的出生日期，75%的人没有国籍信息^[4]。由于知识图谱具有高质量的结构化数据，是很多人工智能应用的基石，因此，近期很多工作都在研究如何利用机器学习算法更好地表示知

识图谱,并以此为基础进行知识图谱补全,从而扩大知识图谱的规模。本文借助知识图谱补全这个任务,来介绍知识图谱表示学习的研究进展。

1 知识图谱补全简介

知识图谱补全的目的是预测出三元组中缺失的部分,从而使知识图谱变得更加完整。对于知识图谱 G ,假设 G 中含有实体集 $E=\{e_1, e_2, \dots, e_M\}$ (M 为实体的数量)、关系集 $R=\{r_1, r_2, \dots, r_N\}$ (N 为关系的数量)以及三元组集 $T=\{(e_i, r_k, e_j) | e_i, e_j \in E, r_k \in R\}$ 。由于知识图谱 G 中实体和关系的数量通常是有限的,因此,可能存在一些实体和关系不在 G 中。记不在知识图谱 G 中的实体集为 $E^*=\{e_1^*, e_2^*, \dots, e_S^*\}$ (S 为实体的数量),关系集为 $R^*=\{r_1^*, r_2^*, \dots, r_T^*\}$ (T 为关系的数量)。

根据三元组中具体的预测对象,知识图谱补全可以分成3个子任务:头实体预测、尾实体预测以及关系预测。对于头(尾)实体预测,需给定三元组的尾(头)实体以及关系,然后预测可以组成正确三元组的实体,例如(姚明,国籍,?), (?, 首都,北京)。对于关系预测,则是给定头实体和尾实体,然后预测两个实体之间可能存在的关系,例如(姚明,?, 中国)。

根据三元组中实体和关系是否均属于知识图谱 G ,可以把知识图谱补全分成两类:1)静态知识图谱补全(Static KGC),涉及的实体 $entity \in E$ 以及关系 $relation \in R$,该场景的作用是补全已知实体之间的隐含关系;2)动态知识图谱补全(Dynamic KGC),涉及不在知识图谱 G 中的实体或关系($entity \in E^*$ 或者 $relation \in R^*$),该场景能够建立知识图谱与外界的关联,从而扩大知识图谱的实体集、关系集以及三元组集。

2 表示学习的相关理论

为了进行知识图谱补全,首先得给知识图谱中的实体和关系选择合适的表示,即构建出合适的特征对实体和关系进行编码。在机器学习中,特征构建通常有两种方法:一种是手工构建,这种方法需要较多的人工

干预,并且需要对所涉及的任务有深入的了解,才可能构建出较好的特征。对于较为简单的任务,该方法是可行的,但对于较为复杂的任务,构建出合适的特征可能需要耗费大量的人力物力。另一种方法是表示学习,该方法需要较少的人工干预,直接通过机器学习算法自动地从数据中学得新的表示,能够根据具体的任务学习到合适的特征。表示学习其实是一个比较广泛的概念,机器学习中不少算法都属于某种形式的表示学习,例如目前人工智能领域的研究热点——深度学习,就是一类常见的表示学习算法。随着硬件的升级以及大数据时代的到来,深度学习在很多领域(图像识别、语音识别、机器翻译等)都击败了传统的机器学习算法,例如经典的多层感知机算法、基于统计学的方法等。但是深度学习也不是万能的,基于深度学习的模型通常拥有大量的参数,加大模型容量的同时也引入了过拟合的风险;因此,为了增强模型的泛化能力,基于深度学习的模型通常需要在大规模数据集上进行训练。此外,深度学习比较适合用于原始特征是连续的且处于比较低层次的领域,例如语音识别、图像识别,深度学习能基于低层次的特征构造出适合任务的高层次语义特征,从而产生较大的突破。而对于自然语言处理领域,语言相关的特征通常已经处于高层次,例如语法结构、依存关系等特征。此外,语言的特征通常是离散的,并且存在多义性;因此,深度学习在该领域的突破相对要小一点。

手工构建和表示学习这两种方法各有利弊,前者虽然需要较多的人工干预,但是构建出的特征通常具有较好的可解释性,有利于研究人员对模型起作用的原因以及任务的本质有更深入的认识。例如计算机视觉领域著名的HOG特征、SIFT特征,其背后就有严谨的数学原理。表示学习虽然在较少的人工干预下能自动地根据任务构建特征,但构建出的特征的可解释性通常比较差,例如现在应用十分广泛的卷积神经网络(CNN),虽然在很多领域都取得了突破性的成果,但是学术界目前也还没从数学角度严谨地证明CNN能起作用的原因。最近的一种研究趋势是把这两种构建方式结合起来,将手

工构建的特征作为先验知识去指导或者优化表示学习算法进行特征的学习,这种做法在不少任务上取得了较好的效果。例如在知识图谱补全这个任务上,不少工作就利用了规则、实体类型、多跳路径等信息构造出高质量的先验知识,并将这些先验知识融合到表示学习上。

接下来,本文将分别介绍Static KGC以及Dynamic KGC这两类场景的相关工作。由于不少综述性文章^[1]都已经介绍过Static KGC场景的很多工作,因此本文重点介绍能解决Dynamic KGC的工作。

3 知识图谱表示学习

3.1 静态知识图谱补全——Static KGC

知识图谱可以看成是一个有向图,实体是结点,而有向边则代表了具体的关系。对于Static KGC场景,其实就是给知识图谱中不同的结点寻找潜在的有向边(关系)。早期的不少工作都属于基于图的表示学习方法,这些方法在小规模的知识图谱上表现良好。然而,随着知识图谱规模的扩大,数据稀疏问题会加重,算法的效率也会降低。为了能适应大规模Static KGC,人们陆续提出多种基于知识图谱结构特征(三元组)的表示学习算法,这些算法将知识图谱中的实体和关系嵌入到低维稠密空间^[1],然后在这个空间计算实体和关系的关联,从而

进行Static KGC。

最经典的工作是Bordes等人于2013年提出的翻译模型——TransE^[5],如图2所示,该模型认为正确的三元组(h, r, t)(h代表头实体的向量, r代表关系的向量, t代表尾实体的向量)需满足 $h + r \approx t$,即尾实体是头实体通过关系平移(翻译)得到的。TransE不仅简单高效,而且还具有较好的扩展性。然而,通过深入分析可以得知,TransE不适合对复杂关系进行建模。例如“性别”这类“N-1”(多对一)型关系,如图3所示,当训练数据中含有三元组(张三, 性别, 男)以及(李四, 性别, 男)时,经过TransE训练后,张三和李四这两个实体的向量可能会比较接近。然而张三和李四在其他方面可能存在差异,例如年龄、籍贯等属性,而TransE无法对这些信息进行有效地区分,导致TransE在复杂关系上的表现比较差。

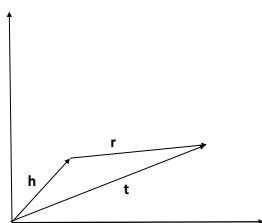


图2 TransE的基本思想

$$\begin{aligned} v(\text{张三}) + v(\text{性别}) &= v(\text{男}) \\ \dots\dots\dots \\ v(\text{李四}) + v(\text{性别}) &= v(\text{男}) \end{aligned}$$



$$v(\text{张三}) \approx v(\text{李四})$$

图3 TransE的复杂关系建模

为了能在复杂关系下有较好的

表现,不少工作考虑实体在不同关系下应该拥有不同的向量。其中,文献[6]设计了TransH模型,该模型将实体投影到由关系构成的超平面上。文献[7]提出TransR模型,该模型则认为实体和关系存在语义差异,它们应该在不同的语义空间。此外,不同的关系应该构成不同的语义空间,因此TransR通过关系投影矩阵,将实体空间转换到相应的关系空间。文献[8]沿用了TransR的思想,提出了TransD模型,该模型认为头尾实体的属性通常有比较大的差异,因此它们应该拥有不同的关系投影矩阵。此外,考虑矩阵运算比较耗时,TransD将矩阵乘法改成了向量乘法,从而提升了运算速度。文献[9]基于实体描述的主题分布来构造实体的语义向量,并且将实体的结构向量投影到对应的语义向量上,从而增强了模型的辨别能力。文献[10]考虑了实体多语义的性质,认为实体应该拥有多个语义向量,而语义向量则是根据实体所处的语境动态生成的。此外,文献[10]通过实体类型构造了关系的类型信息,并将实体与关系、实体与实体之间的相似度作为先验知识融合到表示学习算法中。此外,还有不少工作将规则、路径等信息融入到表示学习中。这些工作通过更加细致的建模以及引入先验知识,在静态知识图谱补全任务上取得了一定的提升。值得一提的是,文献[11]设计了一个基于共享

memory的网络架构IRNs(Implicitly ReasonNets),在向量空间中进行了多跳推理,该模型在复杂关系上取得了目前最好的结果。

3.2 动态知识图谱补全—Dynamic KGC

前面的表示学习算法均属于“离线”算法,它们有一个共同的局限性,只能在训练过程中得到实体以及关系的向量。当实体或关系不在训练集中时,就无法获得它们的向量。然而,为了维持知识图谱的可靠性以及扩大它的规模,我们通常需要对知识图谱中的数据进行“增”、“删”、“改”操作。对于“离线”算法,一旦知识图谱中的数据发生变化,就得重新训练所有实体以及关系的向量,扩展性较差并且耗时耗力。

对于大规模开放知识图谱,例如Freebase、DBPedia,随着时间的变化,它们所含有的事实类三元组可能会发生改变。例如2007年和2017年的美国总统不是同一个人,显而易见,(奥巴马,总统,美国)这个三元组在2017年就是错误的。为了维持知识图谱的可靠性,我们需要不定期地对知识图谱中与时间相关的事实类三元组进行更新。文献[12]考虑了知识图谱中三元组的时间有效性,并提出了一个时间敏感型(time-aware)知识图谱补全模型——TAE,该模型融合了事实的时序信息,将三元组扩展成四元组——(头实体,关系,尾实体,时间)。其中,时间信息能够有效约束向量空间的几何结构。

现有知识图谱中实体和关系的数量通常是有限的。然而,大部分表示学习模型只能在知识图谱中的实体和关系之间进行补全,因此这些模型无法自动地引入新实体或者新关系来扩大知识图谱的规模。新实体和知识图谱中的实体通常拥有丰富的额外信息,例如名称、描述、类型等,这些信息从不同角度对实体进行了刻画。为了能够实现自动向知识图谱中添加新实体的需求,不少方法结合了额外信息来获得新实体的向量,从而建立新实体与现有知识图谱的关联。对于新关系,若有高质量的额外信息,同样可以通过这些信息来建立相应的向量,从而实现关联。

向知识图谱中添加新实体或者新关系的场景其实可以抽象为迁移学习(Transfer Learning)中的零数据学习(Zero-Shot Learning)问题,知识图谱中的实体和关系为源域(Source Domain),新实体和新关系为目标域(Target Domain)。实现迁移学习的基本前提是源域与目标域之间要存在相关性,即要共享相同或者类似的信息(例如特征),否则迁移效果就会比较差,甚至出现负迁移的情况。例如,对于实体的描述信息,若两个域中实体的描述均是英文,由于两个域之间可能共享一些相同或者相似的英文单词,因此可以通过对描述信息建模从而实现迁移,迁移的效果取决于两个域之间英文单词的共享程度;若实体的描述信息不属于同一种语言,例如源域中实体的描述是英文,而目标域中实体的描述是中文,两个域之间共享的信息就非常少,直接进行迁移学习会非常困难;因此,为了能提高添加新实体或者新关系场景的准确率,需要寻找两个域之间所共享的额外信息,然后结合源域中的三元组数据对这些额外信息进行建模,再将学习到的模型迁移到目标域中,得到目标域中实体的向量,从而实现动态知识图谱补全。

在现实世界中,关系的数量一般远少于实体的数量;因此,现有知识图谱的不完整性主要来源于实体的缺失。为了提升知识图谱的完整性,大部分工作主要研究如何准确地向知识图谱中添加新实体,而添加新关系这个场景的相关工作目前还比较少。当向知识图谱中添加新实体时,可以根据知识图谱中的实体以及新实体所拥有的额外信息分成两类场景。(1)新实体拥有丰富的文本信息,例如实体名称、实体描述以及类型;(2)新实体与知识图谱中的实体以及关系有显性的三元组关联,这些三元组通常被称为辅助三元组。辅助三元组不会参与模型的训练过程,它们的作用在于借助训练好的模型推理出新实体的向量。

对于场景(1),相关工作主要通过建立实体与额外信息的映射关系来挖掘以及增强源域与目标域之间的关联。例如,对于源域中的实体A,若它的描述中出现“人口总量”、“国土面积”等词汇,说明实体A很有

可能代表一个国家。根据实体与词汇之间的映射关系，当实体B的描述中也出现这些词汇时，表明实体B很有可能也是一个国家，那么实体B应该具备实体A的一些属性。早期的模型主要通过将知识图谱中的结构信息(实体、关系)与额外信息统一到同一个空间来建立两者的关联。

文献[13]提出了首个联合对齐模型L，该模型分为3个子模型：知识图谱模型K、文本模型T以及对齐模型A。

$$L=K+T+A \quad (1)$$

知识图谱模型K主要通过条件概率 $\Pr(h|r,t)$ 、 $\Pr(r|h,t)$ 、 $\Pr(t|h,r)$ 来获得实体和关系的向量，其中 $z(h,r,t)=7-0.5 \times \|h+r-t\|_2^2$ ，E代表实体集，R代表关系集。

$$K=-\sum_{(h,r,t)}[\log \Pr(h|r,t) + \log \Pr(t|h,r) + \log \Pr(r|h,t)] = -\sum_{(h,r,t)}[\log \frac{\exp\{z(h,r,t)\}}{\sum_{h' \in E} \exp\{z(h',r,t)\}} + \log \frac{\exp\{z(h,r,t)\}}{\sum_{r' \in E} \exp\{z(h,r,t')\}} + \log \frac{\exp\{z(h,r,t)\}}{\sum_{t' \in E} \exp\{z(h,r,t')\}}] \quad (2)$$

文本模型T借鉴了word2vec^[14]的skip-gram算法，从而获得额外信息(实体名称以及维基百科anchors)的词向量，其中V代表词库， $z(w,v)=7-0.5 \times \|w-v\|_2^2$ ，

$$T=-\sum_{(w,v)} \log \Pr(w|v) = -\sum_{(w,v)} \log \frac{\exp\{z(w,v)\}}{\sum_{w' \in V} \exp\{z(w',v)\}} \quad (3)$$

对齐模型A的作用是建立实体与其额外信息中词语之间的映射，其中De代表实体e描述中的词， $z(e,w)=7-0.5 \times \|e-w\|_2^2$ ，

$$A=-\sum_{e \in E} \sum_{w \in D_e} [\log \Pr(w|e) + \log \Pr(e|w)] = -\sum_{e \in E} \sum_{w \in D_e} [\log \frac{\exp\{z(w,e)\}}{\sum_{w' \in V} \exp\{z(w',e)\}} + \log \frac{\exp\{z(w,e)\}}{\sum_{e' \in E} \exp\{z(w,e')\}}] \quad (4)$$

然而，这个对齐模型依赖于维基百科的anchors，因此应用范围受到了限制。文献[15]对它进行了改进，将额外信息替换成实体的描述信息。相比维基百科的anchors，实体的描述信息更加常见，因此能够增大模型的应用范围。

此外，不少工作利用神经网络来获得实体的向量。文献[16]提出一种张量神经网络，用实体名称中所有词

的词向量的平均作为该实体的向量，从而让拥有类似名称的实体能够共享文本信息。文献[17]使用了两种表示学习方法，连续词袋模型(CBOW)以及卷积神经网络模型(CNN)来建立基于实体描述的语义向量。文献[18]结合了知识图谱的结构信息以及实体的描述信息，并提出了基于门机制(Gate-based)的联合学习模型。此外，文献[18]认为实体具有多语义，在不同场景(关系)下可能偏向于某一种语义，而语义则通过实体描述中的词体现。因此，文献[18]设计了一种注意力机制来计算实体描述中的词在不同关系下的权重，使得实体在不同关系下拥有不同的语义向量。

对于场景(2)，文献[19]提出了一种基于图神经网络(Graph-NNs)的模型。该模型分为两部分：传播模型以及输出模型。其中，传播模型负责在图中的节点之间传播信息，而输出模型则是根据具体任务定义了一个目标函数。对于知识图谱补全任务，文献[19]将图谱中相邻(头/尾)实体的向量进行组合，从而形成最终的向量。对于输出模型，本文使用了经典的翻译模型—TransE。为了模拟场景(2)，文献[19]构造了3组测试集：仅三元组的头实体是新实体，仅尾实体是新实体以及头尾实体都是新实体。此外，给每个新实体设计了相应的辅助三元组(头尾实体中仅含有一个新实体)，用于获得新实体的向量。

为了更好地构造知识图谱，实现对数据进行“增”、“删”的需求，文献[20]提出一个新颖的在线(online)知识图谱表示学习模型—puTransE (Parallel Universe TransE)。相比经典的翻译模型，例如TransE、TransR，puTransE具有更好的鲁棒性以及扩展性，并且对超参数不太敏感，具有更好的实用价值。puTransE利用分而治之的思想，通过生成多个向量空间，将语义或结构相似的三元组放在同一个空间中进行训练。此外，每个空间中的超参数均是在给定范围内随机生成，因此，不需要进行大规模的超参数调优。在多个大规模数据集上的测试结果表明，puTransE在效率以及准确率上均优于翻译模型。

4 结语

本文借助知识图谱补全任务,将知识图谱表示学习算法进行了大致梳理。早期的工作主要集中在静态知识图谱补全,以TransE为代表的翻译模型在这个场景上获得了较好的效果。然而,这些模型对超参数比较敏感,并且扩展性也比较差。在真实世界中,可能会不间断地产生新实体以及新关系,翻译模型无法满足自动添加新实体以及新关系的需求,因此,大家逐渐把重心转移到动态知识图谱补全上,从而能自动地扩大知识图谱的规模。相比Static KGC, Dynamic KGC能建立现有知识图谱与外界的有效关联,并且能对知识图谱中的数据进行更新,具有更好的现实意义。因此,如何设计高效的在线学习算法来解决Dynamic KGC是目前一个较好的研究点。

参考文献

- [1] 刘知远,孙茂松,林衍凯,等.知识表示学习研究进展[J].计算机研究与发展, 2016, 53(2):247-261
- [2] BOLLACKER K D, EVANS C, PARITOSH P, et al.: Freebase: a collaboratively created graph database for structuring human knowledge[C]//the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008:1247-1250
- [3] AUER S, BIZER C, KOBILAROV G, et al. DBpedia: A Nucleus for a Web of Open Data[C]//The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + Aswc 2007, Busan, Korea, 2007:722-735
- [4] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion[C]//In: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, 2014:601-610
- [5] BORDES A, USUNIER N. Translating embeddings for modeling multi-relational data[C]//Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013:2787-2795
- [6] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[C]//the Twenty-Eighth AAAI Conference on Artificial Intelligence, Canada, 2014:1112-1119
- [7] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]//the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015:2181-2187
- [8] JI G, HE S, XU L, et al. Knowledge graph embedding via dynamic mapping matrix[C]//the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Beijing, China, 2015:687-696
- [9] XIAO H, HUANG M, MENG L, et al. SSP: Semantic space projection for knowledge graph embedding with text descriptions[C]//the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, U SA, 2017:3104-3110
- [10] MA S, DING J, JIA W, et al. Transt: Type-based multiple embedding representations for knowledge graph completion[C]//Machine Learning and Knowledge Discovery in Databases - European Conference, ECML-PKDD 2017:717-733
- [11] SHEN Y, HUANG P, CHANG M, et al. Modeling large-scale structured relationships with shared memory for knowledge base completion[C]//the 2nd Workshop on Representation Learning for NLP, Rep4NLP@ACL 2017:57-68
- [12] JIANG T, LIU T, GE T, et al. Towards time-aware knowledge graph completion[C]//COLING 2016, 26th International Conference on Computational Linguistics, Japan, 2016:1715-1724
- [13] WANG Z, ZHANG J, FENG J, et al. Knowledge graph and text jointly embedding[C]//the 2014 Conference on Empirical Methods in Natural Language Processing, 2014:1591-1601
- [14] Tomas Mikolov, Kai Chen, Greg Corrado, et al. Efficient estimation of word representations in vector space[EB/OL]. [2018-02-06]. <https://www.researchgate.net/>

- publication/234131319_Efficient_Estimation_of_Word_Representations_in_Vector_Space
- [15] ZHONG H,ZHANG J,WANG Z,et al.Aligning knowledge and text embeddings by entity descriptions[C]//the 2015 Conference on Empirical Methods in Natural Language Processing,2015:267-272
- [16] SOCHER R,CHEN D,MANNING C D,et al.Reasoning with neural tensor networks for knowledge base completion[C]//Advances in Neural Information Processing Systems 26:27th Annual Conference on Neural Information Processing Systems,2013:926-934
- [17] XIE R,LIU Z,JIA J,et al.Representation learning of knowledge graphs with entity descriptions[C]//the Thirtieth AAAI Conference on Artificial Intelligence, 2016:2659-2665
- [18] XU J,QIU X,CHEN K,et al.Knowledge graph representation with jointly structural and textual encoding[C]//the Twenty-Sixth International Joint Conference on Artificial Intelligence,2017:1318-1324
- [19] HAMAGUCHI T,OIWA H,SHIMBO M,et al.Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach[C]//the Twenty-Sixth International Joint Conference on Artificial Intelligence,2017:1802-1808
- [20] TAY Y,LUU A T,HUI S C.Non-parametric estimation of multiple embeddings for link prediction on dynamic knowledge graphs[C]//The Thirty-First AAAI Conference on Artificial Intelligence,2017:1243-1249

作者简介



丁建辉

硕士研究生在读，研究方向为知识图谱表示学习。



贾维嘉

博士，上海交通大学全职致远讲习教授、博士生导师，主要研究方向为网络空间实体对象智慧传感、搜索与大数据处理、多媒体通信等领域。

The Research Advances of Knowledge Graph Completion Algorithm

Ding Jianhui
Jia Weijia

Shanghai Jiao Tong University, Shanghai 200240, China

Abstract Knowledge graph completion algorithm can make knowledge graph more complete, which is a research focus in the field of artificial intelligence. To give better reviews of these algorithms, based on whether they can deal with the scenes that include new entities or new relations, we divide them into two categories: static and dynamic knowledge graph completion algorithm. The former can only handle the scenes that both entities and relations are fixed, thus the extensibility of it is weak. The latter can handle the scenes that involve new entities or new relations, which is of better practical significance. This paper mainly conducts a systematic summary of dynamic knowledge graph completion algorithm, we find that the research heat of dynamic knowledge graph completion algorithm increases gradually, and how to construct dynamic knowledge graphs faster and better is a good research point.

Keywords Artificial Intelligence; Representation Learning; Knowledge Graph; Distributed Representation
