

# 时态知识图谱补全的方法及其进展

申宇铭, 杜剑峰

广东外语外贸大学信息科学与技术学院, 广东 广州 510420

## 摘要

时态知识图谱是将时间信息添加到传统的知识图谱而得到的。近年来,时态知识图谱补全受到了学术界的高度关注,并成为研究热点之一。总结了目前时态知识图谱补全的两大类方法,即基于符号逻辑的方法和基于知识表示学习的方法,比较分析了两类方法的优缺点,展望了未来时态补全方法的发展方向,还总结了7个用于时态知识图谱补全的基准数据集和若干代表性模型在基准数据集上的评测结果。

## 关键词

时态知识图谱;本体;表示学习

中图分类号:TP18

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2021024

## *Temporal knowledge graph completion: methods and progress*

SHEN Yuming, DU Jianfeng

School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou 510420, China

## *Abstract*

Temporal knowledge graph (TKG) are obtained by adding the time information of real-world knowledge to classical knowledge graphs. Recently, TKG completion has drawn much attention and become a hot topic in research. Two main methodologies for TKG completion were summarized, one based on symbolic logic whereas and the other based on knowledge representation learning. The pros and cons of these two different methodologies were discussed, highlighting some directions for enhancing TKG completion in future research. Also, seven benchmark datasets for TKG completion and evaluation results of several typical models on the benchmark datasets were introduced.

## *Key words*

temporal knowledge graph, ontology, representation learning

## 1 引言

在人工智能飞速发展的背景下,知识图谱(knowledge graph)被普遍地认为是人工智能技术和系统中的重要组成部分,在智能搜索、网络安全、金融风险控制及电子商务等诸多领域得到了广泛应用。传统的知识图谱以(实体,关系,实体)或(实体,属性,属性值)三元组集合的方式来表达现实世界的概念、实体、事件及三者之间的关系。比如,三元组(姚明,身高,2.26米)和(姚明,国籍,中国)。2012年5月,谷歌公司发布了谷歌知识图谱(Google knowledge graph),宣布以此为基础构建下一代智能化搜索引擎。这是知识图谱名称的由来,也标志着大规模知识图谱在语义搜索中的成功应用。事实上知识图谱技术渊源已久——从20世纪70年代的专家系统(expert system),到万维网之父Tim Berners-Lee提出的语义网(semantic web),再到他后来提出的链接数据(linked data),都是知识图谱的前身。可以说,知识图谱的升温是人工智能对数据处理与理解需求逐日增加而导致的必然结果。

时间是自然界中所有实体都具有的重要属性,不少知识图谱(如Freebase、Wikidata、DBpedia、YAGO等)具有时间标记的知识。将时间信息引入三元组中所构成的四元组集合称作时态知识图谱(temporal knowledge graph)。全球事件知识图谱<sup>[1]</sup>和冲突事件知识图谱<sup>[2]</sup>都是典型的基于事件的时态知识图谱。此类知识图谱不仅包含了事件之间的共指、因果和时序等关系,还描述了事件之间的规律和演化模式,对传统的知识图谱补充了时间信息,因而具有更大的应用价值。时态知识图谱示例如图1所示。

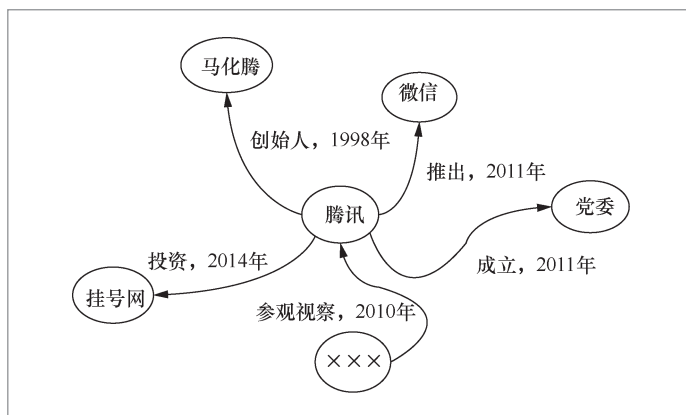


图1 时态知识图谱示例

与传统的静态知识图谱相似,时态知识图谱中的知识也是不完备的,为了实现最大价值,其需要不断地消化吸收新数据,以完善知识体系。近年来,时态知识图谱补全(temporal knowledge graph completion)方法应运而生,受到了学术界的高度关注,并成为研究热点之一<sup>[3]</sup>。这类方法基于时态知识图谱的现有四元组数据,建模预测真实存在的新四元组。这些预测得到的四元组经过验证后,可被添加到时态知识图谱中,作为其演化的结果。时态知识图谱补全的方法在构建大规模知识图谱、相似度计算、关系抽取,以及基于知识图谱问答等任务方面展现出巨大的应用潜力。

为了能够及时追踪到补全方法在时态知识图谱上的发展和应用,本文首先给出时态知识图谱补全的问题定义,回顾代表性的时态知识图谱补全方法,然后对不同类别的方法进行对比分析,给出可能的结合途径,最后总结当前时态知识图谱评测的7个基准数据集,并且给出几个代表性的补全模型在其中3个基准数据集上的评测结果。

## 2 问题定义

本节给出时态知识图谱及其补全过程

的定义,并介绍相关的评测任务。

**定义2.1** 时态知识图谱。一个时态知识图谱是一个四元组的集合。形式地说,任意给定实体集合 $E$ 、关系集合 $R$ 、有穷时间戳集合 $T$ ,时态知识图谱 $G$ 是笛卡儿积 $E \times R \times E \times T$ 的一个子集。

比如,事件“2016—2020年期间,特朗普是美国总统”可以表达为四元组(美国,总统,特朗普,[2016,2020])。类似地,事件“拜登于2021年当选美国总统”可以表达为四元组(美国,总统,拜登,2021)。

**定义2.2** 时态知识图谱补全。令 $E \times R \times E \times T$ 的一个子集 $W$ 表达现实世界中所有成立的事实, $G$ 为 $W$ 的一个真子集。时态知识图谱的补全是指:由 $G$ 出发,推理出不属于 $G$ 但是属于 $W$ 的事实。

**例1** 假设 $W = \{(a, \text{签订合同}, b, t_1), (a, \text{履行合同}, b, t_1+1), (a, \text{签订合同}, c, t_2), (a, \text{履行合同}, c, t_2+1)\}$ ,且 $G = \{(a, \text{签订合同}, b, t_1), (a, \text{履行合同}, c, t_2+1)\}$ ,则需要从 $G$ 出发推理出事实 $(a, \text{履行合同}, b, t_1+1)$ 和 $(a, \text{签订合同}, c, t_2)$ 。

时态知识图谱的补全评测任务主要有两个:

- 给定头实体 $a$ 、关系 $r$ 和时间戳 $t$ ,补全尾实体,即 $(a, r, ?, t)$ ;
- 给定尾实体 $b$ 、关系 $r$ 和时间戳 $t$ ,补全头实体,即 $(?, r, b, t)$ 。

与传统的知识图谱补全问题不同,时态知识图谱的补全更加强调事实成立的时效性,比如,四元组(特朗普,当选,美国总统,2016)和(特朗普,卸任,美国总统,2015),第一个四元组是真实事实,而第二个四元组是虚假事实,去掉时间戳后,所得的两个三元组(特朗普,当选,美国总统)和(特朗普,卸任,美国总统)都是真实事实。因此,如何将事实成立的时效性信息融入传统的补全模型是亟待解决的问题。

### 3 时态知识图谱补全方法

目前,依据对符号处理的不同方式,时态知识图谱的补全方法主要分为两大类:一类是基于符号逻辑的方法,此类方法通过构建领域本体,运用饱和度技术,推理出隐含存在的真实四元组;另一类是基于知识表示学习(knowledge representation learning)的方法,此类方法将符号映射到实值空间,采用数值运算评估四元组的真实程度。

#### 3.1 基于符号逻辑的方法

基于符号逻辑的时态数据查询回答(logic-based temporal query answering)方法又分为两类:一类以领域为中心,另一类以时间为中心。

以领域为中心的方法假定本体的构建语言为描述逻辑,查询语句为包含时态算子(next-time,previous-time,since-time,until-time,future-time)和否定联结词的一阶时态逻辑公式。以此为基础,Baader F等人<sup>[4]</sup>分析了基于描述逻辑SHQ本体的时态查询计算复杂性;Borgwardt S等人<sup>[5-6]</sup>提出了基于描述逻辑DL-Lite族和EL本体的时态查询推理算法,并分析了算法的计算复杂性。

以时间为中心的方法假定本体的构建语言为时态描述逻辑,查询语句为包含时态算子(next-time,previous-time,since-time,until-time,future-time)的一阶时态逻辑公式。以此为基础,Artale A等人<sup>[7]</sup>设计了一种时态描述逻辑语言TQL,其包含past-time和future-time两个一元时态算子,并在限定时态概念只出现在术语公理左侧的情形下,证明了该逻辑语言支

持时态查询的一阶重写。Gutiérrez-Basulto V等人<sup>[8]</sup>在限定时态查询语言为时态原子语句的情形下,证明了时态本体EL-LTL的查询问题是不可判定的,继而通过引入若干语法上的限制条件识别出EL-LTL的若干可判定子片段,并证明了在某些子片段上的时态查询回答是多项式时间可以完成的。Artale A等人<sup>[9]</sup>全面研究了时态DL-Lite-LTL族下时态查询的一阶重写问题,较完整地分析了时态查询问题的计算复杂性。

**例2** 假设本体只包括一条规则“对所有 $x, y, t$ :  $R(x, y, t) \rightarrow Q(x, y, t+1)$ ”,其中谓词 $R$ 、 $Q$ 分别表示签订合同和履行合同,考虑到例1中集合 $G = \{(a, \text{签订合同}, b, t_1), (a, \text{履行合同}, c, t_2+1)\}$ ,由上述规则可以推理出隐含事实 $(a, \text{履行合同}, b, t_1+1)$ 。对于时态查询 $q = (a, \text{履行合同}, ?, t_1+1)$ ,隐含的事实 $(a, \text{履行合同}, b, t_1+1)$ 满足该时态查询 $q$ 。

对比两类方法,以领域为中心的方法在查询语言的表达能力上要强于以时间为中心的方法,而以时间为中心的方法在本体的表达能力上要强于以领域为中心的方法。同时,基于符号逻辑的方法在实际应用中难以覆盖大量真实的四元组,召回率较低,而且构建本体时也要付出较高的人工成本。

### 3.2 基于知识表示学习的方法

本节先简要回顾针对传统知识图谱的表示学习代表性方法,再以此为基础,综述针对时态知识图谱的表示学习方法。知识表示学习方法的基本原理是将给定的三元组数据映射到低维、高密度的数值空间,通过数值运算评估未知三元组的真实程度。依据三元组评分函数的不同类型,传统知识图谱的表示学习方法大致分为3类:第一类是基于平移距

离模型的方法,第二类是基于矩阵分解模型的方法,第三类是基于神经网络模型的方法。

基于平移距离模型的方法根据三元组中头尾实体表示向量的距离来估计三元组的真实性。Bordes A等人<sup>[10]</sup>提出了第一个平移距离模型TransE。该模型采用损失函数 $\|h+r-t\|_{L1/L2}$ 来估计三元组的真实程度,其中 $h$ 为头实体向量, $r$ 为关系实体向量, $t$ 为尾实体向量,L1和L2分别表示1-范数和2-范数,而真实程度通常可以被定义为损失函数值的相反数。根据最优化目标,真实三元组的损失函数值应该趋向于零,因此TransE不太适用于一对多、多对一或者多对多的关系建模。针对TransE模型的局限性,此后陆续涌现了TransH<sup>[11]</sup>、TransR<sup>[12]</sup>、TransD<sup>[13]</sup>、TransG<sup>[14]</sup>、RotatE<sup>[15]</sup>等模型。

基于矩阵分解模型的方法采用 $h^T M_r t$ 形式的评分函数来评估三元组的真实程度,其中 $M_r$ 是关系依赖的矩阵, $h$ 和 $t$ 分别是头、尾实体向量。Nickel M等人<sup>[16]</sup>提出了第一个矩阵分解模型,即RESCAL模型,用于三元组预测。此后,他们又提出了参数更少的全息嵌入(holographic embedding, HolE)模型<sup>[17]</sup>。Yang B S等人<sup>[18]</sup>将关系依赖的矩阵看作以实数构成的对角矩阵,提出了DistMult模型。Trouillon T等人<sup>[19]</sup>使用复数而不是实数构造头尾实体向量,并将关系依赖的矩阵看作以复数构成的对角矩阵,提出了ComplEx模型。Liu H X等人<sup>[20]</sup>为关系依赖矩阵引入正态性和可交换性约束来表达类比性质(比如北京与中国的关系类似于巴黎与法国的关系),提出了ANALOGY模型,并证明了该模型是HolE和ComplEx等模型的一般化形式。此后,研究人员还陆续提出了Simple<sup>[21]</sup>、Tucker<sup>[22]</sup>等模型。

基于神经网络模型的知识表示学习方法采用神经网络形式的评分函数来评估三元组的真实程度。Bordes A等人<sup>[23]</sup>提出了语义匹配能量 (semantic matching energy, SME) 模型, 该模型将头实体与关系的交互模型和尾实体与关系的交互模型作为第一层网络, 再将两个交互模型的输出组合起来构成第二层网络。Socher R等人<sup>[24]</sup>提出了神经张量网络 (neural tensor network, NTN) 模型, 该模型采用头尾实体各自的线性变换模型以及它们之间交互的线性变换模型来构造神经网络, 其中所有线性变换模型都使用不同的关系依赖矩阵。Dong X等人<sup>[25]</sup>提出了多层感知器 (multi-layer perceptron, MLP) 模型, 该模型采用关系和头尾实体各自的线性变换模型共3个线性变换模型来构造神经网络, 3个线性变换模型中使用的变换矩阵都不依赖于关系。

除了这3类方法, 还有文献讨论了利用知识图谱外部信息的知识表示学习方法, 包括结合实体描述信息的方法、结合实体类型信息的方法、结合关系路径信息的方法、结合逻辑规则的方法等。更多的传统知识图谱的表示学习方法参见参考文献[3]。

传统知识图谱中的知识在大多数情况下只在特定的时间内有效, 而一些事实 (如演化的事件) 往往出现在一个时间序列中。为了对时间序列中的事实进行表示学习, 近年来涌现了不少针对时态知识图谱的补全方法。依据对时间戳的处理方式, 这些方法可以大致分为两类: 第一类是时间戳单独编码方法, 第二类是基于序列学习的方法。

时间戳单独编码方法显式地将时间戳建模为向量、矩阵或平面, 再将时间戳的信息直接用于知识图谱的补全。Jiang T S等人<sup>[26]</sup>率先提出结合时态信息知识图谱补全模型, 该模型由两部分构成, 其中一个

部分是由TransE获得关系的表示向量, 另一个部分由3种时态一致性约束 (先后顺序关系、时态不相交性、时态区间有效性) 构成。模型通过一个时态演化矩阵来刻画不同关系之间的时态依赖性, 具体地说, 任意给定两个时序依赖关系 $r_k$ 和 $r_l$ , 它们的时序评分函数定义为:  $f(r_k, r_l) = ||r_k T - r_l||_{L1/L2}$ , 其中矩阵 $T$ 是一个能够编码时序关系对的非对称矩阵, 此评分函数基本思想如图2所示。

在图2中,  $r_1$ 是先于 $r_2$ 的时态关系, 根据评分函数有 $r_1 T \approx r_2$ , 但是 $r_2 T \neq r_1$ 。Dasgupta S S等人<sup>[27]</sup>结合了模型TransE和TransH的特点, 提出了HyTE时态表示学习模型。该模型首先将时间戳建模为关系依赖的超平面, 然后利用TransH模型将头、尾实体投影到该平面, 最后利用TransE模型完成知识图谱的补全工作。Ma Y P等人<sup>[28]</sup>和Lacroix T等人<sup>[29]</sup>都将时间戳看作第4个维度, 分别扩展了Tucker和ComplEx张量分解模型, 再将时间戳的表示向量直接用于四元组真实程度的估计。Jain P等人<sup>[30]</sup>在Lacroix工作的基础上, 将先后顺序关系和循环关系 (比如奥运会每隔3年举办) 的信息增加到评分函数中, 用于图谱的补全。Xu C J等人<sup>[31]</sup>基于RotatE模型提出了时态旋转模型, 该模型将时间戳建模为旋转复向量, 将实体和关系表示为复向量, 通过旋转复向量与实体表示复向量的内积运算, 将时态信息融合到实体的表示向量中, 并利用基于距离TransE模型完成知识图谱的补全。时间戳单独编码方法将时间信息看成连通实体与实体、关系与实体及关系与关系的桥梁。

基于序列学习的方法先设计一个序列学习模型, 将时态信息融合到实体或关系的表示向量中, 再用已有的表示学习模型估计带有时态信息三元组的真实程



度,从而完成时态知识图谱的补全任务。Garcia-Duran A等人<sup>[32]</sup>将关系和时间戳的特征(年、月、日)构成一个关系序列,通过一个线性层函数,将关系和时间戳特征映射为同维数的向量,然后把该序列向量输入一个长短期记忆(long short-term memory, LSTM)网络进行编码,学习到融合时间信息的关系表示向量,该具体过程如图3所示。

在图3中,关系“bornIn”与日期“1986”经过LSTM模型后形成了融合时间信息的关系表示向量 $e_{pseq}$ ;最后,依据DistMult模型的评分函数 $f(s, p_{seq}, o) = (e_s \circ e_o) e_{pseq}^T$ 或TransE模型的评分函数 $f(s, p_{seq}, o) = ||e_s + e_{pseq} - e_o||_2$ 对三元组 $(s, p_{seq}, o)$ 完成补全,这里 $e_s$ 、 $e_o$ 和 $e_{pseq}$ 分别表示头实体 $s$ 、尾实体 $o$ 和关系 $p_{seq}$ 的表示向量。Goel R等人<sup>[33]</sup>将实体的表示向量分为静态和动态两个部分,并利用SimplE模型完成知识图谱的补全,其中实体表示向量的静态部分表达实体在演化过程中固定不变的特征,动态部分则结合正弦激活函数来调控不同时间点状态的闭合,进而表达演化过程中变化的特征。Wu J P等人<sup>[34]</sup>利用鲁棒性图卷积神经网络(robust graph convolutional network, RGCN)模型将不同时间同一实体的邻居结构化信息进行融合,获得了该实体的一个序列表示向量,然后将该表示向量序列输入时态递归神经网络中,获取该实体融合时态信息的表示向量,最后利用静态的补全模型完成补全的工作。Jung J等人<sup>[35]</sup>提出了一种时态图神经网络(temporal graph neural network, TGNN)模型。该模型对时态知识图谱及查询分别进行预训练,完成时态信息与实体表示向量的融合,并计算其邻居的注意力分布,然后利用子图采样的方法获得每个实体及与查询相关的邻居的子图结构,过滤与查询不相关的实体,再利用图神经网络模型,更

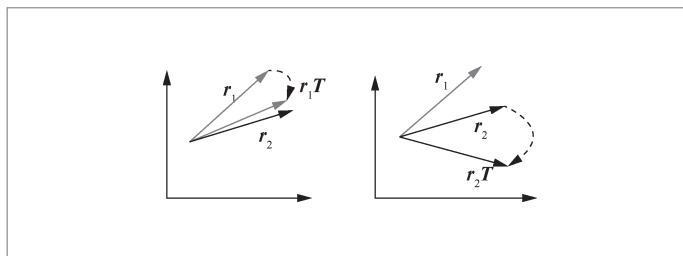


图2 时态演化矩阵

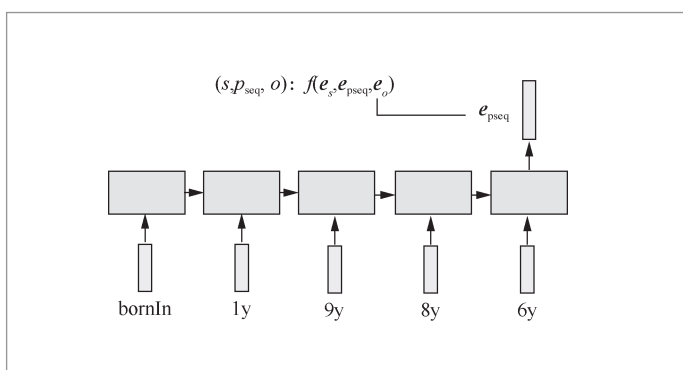


图3 融合时间信息的关系表示向量

新子图上实体的表示向量,并结合基于路径遍历的方法更新实体邻居的注意力分布,最后依据最高概率推理出实体间隐藏的关系。与更新实体或关系的表示向量不同, Xu Y R等人<sup>[36]</sup>考虑了时态知识图谱中增加新实体的情况,设计了一种策略递归地更新模型参数。Xu C等人<sup>[37]</sup>考虑了知识图谱时态演化过程中的不确定性因素,在每个时间点采用高斯分布函数来表达实体和关系的不确定性,再结合时间序列的线性模型来刻画实体和关系表示向量随时间演化的趋势,最后通过计算实体和关系概率分布的距离来完成补全。Han Z等人<sup>[38]</sup>则将补全的工作从传统的欧氏空间拓展到黎曼流形(Riemannian manifold)上完成。相对于第一类方法,基于序列学习的方法更加强调不同实体和关系间的历史关系,即实体或关系之间带有时间戳的序列之间的交互。

## 4 两类方法的比较

基于符号逻辑的方法可以从已有的知识图谱出发,结合本体中的规则,推理出新的实体间关系;同时,还可以对演化后的知识图谱进行逻辑一致性检查,使得推理结果具备透明、可靠及可解释性强等特点。为了表达时态的知识,这一类方法通常需要引入时态算子来提升本体的表达能力,而表达能力的提升通常会导致如下两种局限性。

- 不可判定性:即不存在有限时间可终止的算法,使得该算法能够判定相关的推理问题是否可证。比如,在描述逻辑EL中,引入时态算子到本体中会导致其时态查询回答是不可判定的<sup>[8]</sup>。
- 高计算复杂性:比如,在描述逻辑EL中,交查询的回答是多项式时间的,但引入时态算子到查询语言后,时态交查询回答却是NP难的<sup>[5]</sup>。

由此可见,基于符号逻辑的方法在推理效率方面难以满足日益增长的数据需求。

基于知识表示学习的方法将研究对象的语义信息表示为低维稠密的实值向量。在低维向量空间中能够高效地计算实体和关系的语义关系,显著地提高推理性能。但是,此类方法的推理过程不透明,推理结果的可解释性低。此外,大多数表示学习模型的表达能力有限。比如,参考文献[39]指出数值嵌入模型不能表达本体中的存在规则(existential rule),而这类规则恰好对应于轻量级描述逻辑EL或DL-Lite族的术语或角色公理;进一步地,参考文献[40]指出,就算表示学习模型能够区分所有真实的三元组和错误的三元组,也不能确保正确区分出本体中的上下位关系。由此可见,基于知识表示学习的推理不能完全替代基于符号逻辑的推理。

从推理的方式来看,基于符号逻辑的推理属于演绎推理,而基于表示学习的推理属于不完全归纳推理,两种推理方式各有各的优缺点,但最终的目的都是将不完备的知识库(incomplete knowledge base)演化为完备的知识库(complete knowledge base)。为了发挥两种推理方式的优势,未来的知识图谱补全研究方向可以聚焦于解决表示学习模型无法习得存在规则逻辑的结论的问题。在这一方向上,Du J F等人<sup>[41]</sup>提出了逻辑背景预完备技术来融入关系特征定义,并提出了区分头尾实体的投影函数来解决关系表示向量不可区分的问题;进一步地,参考文献[42]引入了逻辑一致性规则预完备技术,解决了部分排位靠前的三元组与逻辑一致性规则相违背的问题。

## 5 基准测试数据集

当前,时态知识图谱补全研究领域有7个基准测试数据集,它们是在Wikidata、YAGO、GDELT和综合早期危机预警系统(integrated crisis early warning system, ICEWS)4个数据库上构建的。这7个数据集分别是GDELT-500、ICEWS14、ICEWS05-15、YAGO15k、Wikidata11k、YAGO11k和Wikidata12k,其中YAGO和Wikidata中的事实是基于时间区间的,而GDELT和ICEWS中的事实是基于时间点的。

- GDELT: GDELT数据库记录了从1969年至今,每个国家大约100多种语言的新闻媒体中印刷、广播和Web形式的新闻,并且每隔15 min更新一次数据。GDELT主要包含两大数据库,即事件数据库(event database)和全球知识图谱(global knowledge graph)。目前,用于时态知识图谱补全研究的数据集是GDELT-500。

● ICEWS: ICEWS数据库涵盖了100多个数据源以及250个国家和地区的政治事件,并且每天更新一次数据。用于时态知识图谱补全研究的数据集是ICEWS14和ICEWS05-15。

● Wikidata: Wikidata是维基媒体基金会主持的一个自由的协作式多语言辅助知识库,旨在为维基百科、维基共享资源以及其他的维基媒体项目提供支持。目前,用于时态知识图谱补全研究的数据集是Wikidata11k和Wikidata12k。

● YAGO: YAGO是由德国马克斯·普朗克研究所研制的链接数据库。该数据库主要集成了Wikipedia、WordNet和GeoNames 3个来源的数据。YAGO将WordNet的词汇定义与Wikipedia的分类体系进行了融合集成,使得YAGO具有

更加丰富的实体分类体系。YAGO还考虑了时间和空间知识,为很多知识条目增加了时间和空间维度的属性描述。目前,用于时态知识图谱补全研究的数据集是YAGO11k和YAGO15k。上述7个数据集的统计结果见表1。

表2给出了代表性的补全模型在3个基准数据集ICEWS14、ICEWS05-15和Wikidata11k上的评测结果。表2中上标为\*的评测结果来自参考文献[32],上标为+的评测结果来自参考文献[33],上标为#的评测结果来自参考文献[29]。其中-表示所在行的模型在所在列的基准数据集上没有公布评测结果。评测指标MRR表示所有正确答案预测排名的倒数的均值;Hit@*k*表示正确答案在前*k*位预测三元组中的百分比。

表1 基准数据集的统计结果

数据集	实体数/个	关系数/条	时间跨度	更新时间间隔
GDELT-500	500	20	2015—2016年	15 min
ICEWS14	7 128/6 869	230	2014年	每天
ICEWS05-15	10 094	251	2005—2015年	每天
Wikidata11k	11 134	95	25—2020年	每年
Wikidata12k	12 554	24	1479—2018年	每年
YAGO11k	10 623	10	公元前453—2844年	每年
YAGO15k	15 403	34	1513—2017年	每年

表2 若干代表性补全模型的评测结果

模型	ICEWS14			ICEWS05-15			Wikidata11k		
	MRR	Hit@1	Hit@10	MRR	Hit@1	Hit@10	MRR	Hit@1	Hit@10
TransE*	0.280	9.4	63.7	0.294	9.0	66.3	0.316	18.1	65.9
DistMult*	0.439	32.3	67.2	0.456	33.7	69.1	0.316	18.1	66.1
ConT+	0.185	11.7	31.5	0.163	10.5	27.2	—	—	—
TTransE*	0.255	7.4	60.1	0.271	8.4	61.6	0.488	33.9	80.6
HyTE+	0.297	10.8	65.5	0.316	11.6	68.1	—	—	—
TA-TransE*	0.275	9.5	62.5	0.299	9.6	66.8	0.484	32.9	80.7
TA-DistMult*	0.477	36.3	68.6	0.474	34.6	72.8	0.700	65.2	78.5
DE-Simple+	0.526	41.8	72.5	0.513	39.2	74.8	0.310	18.4	62.5
TComplex#	0.560	47.0	73.0	0.580	49.0	76.0	0.731	67.3	84.5
TNTComplex#	0.560	46.0	74.0	0.600	50.0	78.0	0.718	65.4	85.6



## 6 结束语

给定某个时间区间 $[t_0, t_1]$ 及其对应的时态知识图谱 $G$ , 补全任务是针对某个时刻 $t$  ( $t_0 \leq t \leq t_1$ ) 的推理任务。比如, 在例2中, 可以由 $(a, \text{履行合同}, c, t_2+1)$ 补全出 $(a, \text{签订合同}, c, t_2)$ 。与补全任务相对的另一个任务是时态知识图谱的预测任务。即给定某个时间区间 $[t_0, t_1]$ 及其对应的时态知识图谱 $G$ , 预测出 $t > t_1$ 时刻图谱 $G$ 的演化结果。相比而言, 时态知识图谱的预测任务比补全任务更具有挑战性。限于文章的篇幅, 请读者阅读参考文献[3]了解预测任务的解决方案和相关技术。

综合考虑时态知识图谱的补全方法不难发现, 基于知识表示学习的方法是目前的主流方法。这类方法具有计算效率高和召回率高的特点, 但是在表达能力上还存在弱点。因此, 未来的研究工作可以关注如下两个可能的完善方向: ①融合本体推理, 运用基于符号逻辑的推理弥补知识表示学习在表达能力上的不足, 完成知识表示学习模型的精准训练; ②设计表达能力更强的神经网络模型, 用于表达本体中所有可能的规则。

## 参考文献:

- [1] LEETARU K, SCHRODT P. GDELT: global data on events, location and tone, 1979–2012[C]// Proceedings of the ISA 2013 Annual Convention. [S.l.:s.n.], 2013: 1–49.
- [2] BOSCHÉ E, LAUTENSCHLAGER J, BRIEN S, et al. ICEWS coded event data[Z]. 2015.
- [3] KAZEMI S M, GOEL R, JAIN K, et al. Representation learning for dynamic graphs: a survey[J]. Journal of Machine Learning Research, 2020, 21(70): 1–73.
- [4] BAADER F, BORGWARDT S, LIPPMANN M.

Temporal query entailment in the description logic SHQ[J]. Journal of Web Semantics, 2015, 33(8): 71–93.

- [5] BORGWARDT S, THOST V. Temporal conjunctive query answering in the extended DL-Lite family[J]. arXiv preprint, 2020, arXiv:2003.09508.
- [6] BORGWARDT S, THOST V. Temporal query answering in the description logic EL[C]// Proceedings of the 24th Joint Conference on Artificial Intelligence (IJCAI). [S.l.:s.n.], 2015: 2819–2825.
- [7] ARTALE A, KONTCHAKOV R, WOLTER F, et al. Temporal description logic for ontology-based data access[C]// Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI). [S.l.:s.n.], 2013: 711–717.
- [8] GUTIÉRREZ-BASULTO V, JUNG J C, SCHNEIDER T. Lightweight temporal description logics with rigid roles and restricted TBoxes[C]// Proceedings of the 24th Joint Conference on Artificial Intelligence (IJCAI). [S.l.:s.n.], 2015: 3015–3021.
- [9] ARTALE A, KONTCHAKOV R, KOVTUNOVA A, et al. First-order rewritability of ontology-mediated queries in linear temporal logic[J]. arXiv preprint, 2020, arXiv:2004.07221.
- [10] BORDES A, USUNIER N, GARCÍA-DURÁN A, et al. Translating embeddings for modeling multi-relational data[C]// Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS). [S.l.]: Curran Associates Inc., 2013: 2787–2795.
- [11] WANG Z, ZHANG J W, FENG J L, et al. Knowledge graph embedding by translating on hyperplanes[C]// Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2014: 1112–1119.
- [12] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion[C]//

- Proceedings of the 29th AAAI Conference on Artificial Intelligence(AAAI). Palo Alto: AAAI Press, 2015: 2181–2187.
- [13] JI G L, HE S Z, XU L H, et al. Knowledge graph embedding via dynamic mapping matrix[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL). [S.l]: Association for Computational Linguistics, 2015: 687–696.
- [14] XIAO H, HUANG M L, ZHU X. TransG: a generative model for knowledge graph embedding[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL). [S.l]: Association for Computational Linguistics, 2016: 992–998.
- [15] SUN Z Q, DENG Z H, NIE J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space[C]// Proceedings of the 7th International Conference on Learning Representations (ICLR). [S.l:s.n.], 2019.
- [16] NICKEL M, TRESP V, KRIEDEL H. A three-way model for collective learning on multi-relational data[C]// Proceedings of the 28th International Conference on Machine Learning (ICML). Madison: Omnipress, 2011: 809–816.
- [17] NICKEL M, ROSASCO L, POGGIO T. Holographic embeddings of knowledge graphs[C]// Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2016: 1955–1961.
- [18] YANG B S, YIH W, HE X D, et al. Embedding entities and relations for learning and inference in knowledge bases[C]// Proceedings of the 3rd International Conference on Learning Representations (ICLR). [S.l:s.n.], 2015.
- [19] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]// Proceedings of the 33rd International Conference on Machine Learning (ICML). Madison: Omnipress, 2016: 2071–2080.
- [20] LIU H X, WU Y X, YANG Y M. Analogical inference for multi-relational embeddings[C]// Proceedings of the 34th International Conference on Machine Learning (ICML). Madison: Omnipress, 2017: 2168–2178.
- [21] KAZEMI S M, POOLE D. Simple embedding for link prediction in knowledge graphs[C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS). [S.l.]: Curran Associates Inc., 2018: 4289–4300.
- [22] BALAŽEVIĆ I, ALLEN C, HOSPEDALES T. Tucker: tensor factorization for knowledge graph completion[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2019: 5185–5194.
- [23] BORDES A, GLOROT X, WESTON J, et al. A semantic matching energy function for learning with multi-relational data – application to word-sense disambiguation[J]. Machine Language, 2014, 94(2): 233–259.
- [24] SOCHER R, CHEN D Q, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]// Proceedings of 27th Annual Conference on Neural Information Processing Systems (NIPS). [S.l.]: Curran Associates Inc., 2013: 926–934.
- [25] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion[C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2014: 601–610.
- [26] JIANG T S, LIU T Y, GE T, et al. Towards time-aware knowledge graph completion[C]// Proceedings of the 26th International Conference on Computational Linguistics (COLING).

- [S.l.:s.n.], 2016: 1715–1724.
- [27] DASGUPTA S S, RAY S N, TALUKDAR P. HyTE: hyperplane-based temporally aware knowledge graph embedding[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2018: 2001–2011.
- [28] MA Y P, TRESP V, DAXBERGER E A. Embedding models for episodic knowledge graphs[J]. *Journal of Web Semantics*, 2018, 59.
- [29] LACROIX T, OBOZINSKI G, USUNIER N. Tensor decompositions for temporal knowledge base completion[C]// Proceedings of the 8th International Conference on Learning Representations (ICLR). [S.l.:s.n.], 2020.
- [30] JAIN P, RATHI S, MAUSAM M. Temporal knowledge base completion: new algorithms and evaluation protocols[C]// Proceedings of the 2020 International Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2020: 3733–3747.
- [31] XU C J, NAYYERI M, ALKHOURY F, et al. TeRo: a time-aware knowledge graph embedding via temporal rotation[C]// Proceedings of the 28th International Conference on Computational Linguistics (COLING). [S.l.:s.n.], 2020: 1583–1593.
- [32] GARCIA-DURAN A, DUMANCIC S, NIEPERT M. Learning sequence encoders for temporal knowledge graph completion[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2018: 4816–4821.
- [33] GOEL R, KAZEMI S M, BRUBAKER M, et al. Diachronic embedding for temporal knowledge graph completion[C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2020: 3988–3995.
- [34] WU J P, CAO M, CHEUNG J C K, et al. TeMP: temporal message passing for temporal knowledge graph completion[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2020: 5730–5746.
- [35] JUNG J, JUNG J, KANG U. T-GAP: learning to walk across time for temporal knowledge graph completion[J]. *arXiv preprint*, 2020, arXiv:2012.10595.
- [36] XU Y R, E H H, SONG M N, et al. RTFE: a recursive temporal fact embedding framework for temporal knowledge graph completion[J]. *arXiv preprint*, 2020, arXiv:2009.14653.
- [37] XU C, NAYYERI M, ALKHOURY F, et al. Temporal knowledge graph completion based on time series gaussian embedding[C]// Proceedings of the 19th International Semantic Web Conference (ISWC). [S.l.:s.n.], 2020.
- [38] HAN Z, CHEN, P, MA Y P, et al. DyERNIE: dynamic evolution of Riemannian manifold embeddings for temporal knowledge graph completion[C]// Proceedings of 2020 the Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2020: 7301–7316.
- [39] G U T I É R R E Z – B A S U L T O V , SCHOCKAERT S. From knowledge graph embedding to ontology embedding? An analysis of the compatibility between vector space representations and rules[C]// Proceedings of the 16th International Conference on the Principles of Knowledge Representation and Reasoning (KR). [S.l.:s.n.], 2018: 379–388.
- [40] FATEMI B, RAVANBAKHS S, POOLE D. Improved knowledge graph embedding using background taxonomic information[C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2019: 3526–3533.

[41] DU J F, QI K X, WAN H, et al. Enhancing knowledge graph embedding from a logical perspective[C]// Proceedings of the 7th Joint International Semantic Technology Conference (JIST). [S.l.:s.n.], 2017: 232–247.

[42] DU J F, QI K X, SHEN Y M. Knowledge graph embedding with logical consistency[C]// Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data (CCL). [S.l.:s.n.], 2018: 123–135.

#### 作者简介



**申宇铭** (1976– ), 男, 博士, 广东外语外贸大学教授, 主要研究方向为知识表示与推理、知识图谱。主持或参与多项国家自然科学基金和省部级项目。近年来在《计算机学报》《软件学报》等国内重要期刊, 以及国际重要期刊和国际会议上发表论文20余篇。担任CCKS、AAAI、EMNLP等国内外重要学术会议的程序委员会委员。



**杜剑峰** (1976– ), 男, 博士, 广东外语外贸大学教授, 中国中文信息学会语言与知识计算专业委员会委员, 主要研究方向为知识表示与推理、数据挖掘和自然语言处理。在AAAI、WWW、ISWC、CIKM和KAIS等学术会议上发表数十篇文章, 获得多项国家自然科学基金项目资助。担任*Journal of Web Semantics*编委, 长期担任CCKS、CSWS、IJCAI、AAAI、ISWC、JIST等学术会议的程序委员会成员, 曾担任CSWS 2014程序委员会主席。

**收稿日期:** 2021-03-02

**通信作者:** 申宇铭, ymshen2002@163.com

**基金项目:** 国家自然科学基金资助项目 (No.61876204); 广东省自然科学基金资助项目 (No.2018A030313777)

**Foundation Items:** The National Natural Science Foundation of China (No.61876204), Guangdong Natural Science Foundation (No.2018A030313777)