

## 时序知识图谱表示学习

徐涌鑫, 赵俊峰, 王亚沙, 谢冰, 杨恺

### 引用本文

徐涌鑫, 赵俊峰, 王亚沙, 谢冰, 杨恺. [时序知识图谱表示学习](#)[J]. 计算机科学, 2022, 49(9): 162-171.

XU Yong-xin, ZHAO Jun-feng, WANG Ya-sha, XIE Bing, YANG Kai. [Temporal Knowledge Graph Representation Learning](#)[J]. Computer Science, 2022, 49(9): 162-171.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于无监督集群级的科技论文异质图节点表示学习方法](#)

Scientific Paper Heterogeneous Graph Node Representation Learning Method Based on Unsupervised Clustering Level

计算机科学, 2022, 49(9): 64-69. <https://doi.org/10.11896/jsjcx.220500196>

### [基于 Key-Value 关联记忆网络的知识图谱问答方法](#)

Key-Value Relational Memory Networks for Question Answering over Knowledge Graph

计算机科学, 2022, 49(9): 202-207. <https://doi.org/10.11896/jsjcx.220300277>

### [基于自注意力模型的本体对齐方法](#)

Ontology Alignment Method Based on Self-attention

计算机科学, 2022, 49(9): 215-220. <https://doi.org/10.11896/jsjcx.210700190>

### [融合知识图谱的多层次传承影响力计算与泛化研究](#)

Multi-level Inheritance Influence Calculation and Generalization Based on Knowledge Graph

计算机科学, 2022, 49(9): 221-227. <https://doi.org/10.11896/jsjcx.210700144>

### [基于安全多方计算和差分隐私的联邦学习方案](#)

Federated Learning Scheme Based on Secure Multi-party Computation and Differential Privacy

计算机科学, 2022, 49(9): 297-305. <https://doi.org/10.11896/jsjcx.210800108>

# 时序知识图谱表示学习

徐涌鑫<sup>1,2</sup> 赵俊峰<sup>1,2,3</sup> 王亚沙<sup>1,2,3</sup> 谢冰<sup>1,2,3</sup> 杨恺<sup>1,2,3</sup>

1 北京大学计算机学院 北京 100871

2 高可信软件技术教育部重点实验室 北京 100871

3 北京大学(天津滨海)新一代信息技术研究院 天津 300450

(xuyx@stu.pku.edu.cn)

**摘要** 知识图谱作为一种结构化的人类知识形式,对海量多源异构异质的数据语义互通起到了很好的支撑作用,并有效地支持了数据分析等任务,成为了近年来学术界和工业界的研究热点。目前大多数知识图谱都是根据非实时的静态数据构建,没有考虑实体和关系的时间特性。然而社交网络通信、金融贸易、疫情传播网络等应用场景的数据具有实时动态的特点以及复杂的时间特性,如何利用时序数据构建知识图谱并且对该知识图谱进行有效建模是一个具有挑战性的问题。目前,有许多研究工作利用时序数据中的时间信息丰富知识图谱的特征,赋予知识图谱动态特征,将事实三元组拓展为(头实体,关系,尾实体,时间)的四元组表示,使用时间相关四元组进行知识表示的知识图谱被统称为时序知识图谱。文中对时序知识图谱相关文献进行整理和分析,并对时序知识图谱表示学习的工作进行了全面综述。具体地,首先简单介绍了时序知识图谱的背景与定义;其次总结了时序知识图谱表示学习方法相比传统知识图谱表示学习方法的优点;然后从事实的建模方法角度详细阐述了时序知识图谱表示学习的主要方法,并且介绍了上述方法使用到的数据集;最后对该技术的主要挑战进行了总结,并对其未来研究方向进行了展望。

**关键词:** 知识图谱;深度学习;表示学习;时间信息;动态过程

**中图法分类号** TP311

## Temporal Knowledge Graph Representation Learning

XU Yong-xin<sup>1,2</sup>, ZHAO Jun-feng<sup>1,2,3</sup>, WANG Ya-sha<sup>1,2,3</sup>, XIE Bing<sup>1,2,3</sup> and YANG Kai<sup>1,2,3</sup>

1 School of Computer Science, Peking University, Beijing 100871, China

2 Key Laboratory of High Confidence Software Technologies, Ministry of Education, Beijing 100871, China

3 Peking University Information Technology Institute(Tianjin Binhai), Tianjin 300450, China

**Abstract** As a structured form of human knowledge, knowledge graphs have played a great supportive role in supporting the semantic intercommunication of massive, multi-source, heterogeneous data, and effectively support tasks such as data analysis, attracting the attention of academia and industry. At present, most knowledge graphs are constructed based on non-real-time static data, without considering the temporal characteristics of entities and relationships. However, data in application scenarios such as social network communication, financial trade, and epidemic spreading network are highly dynamic and exhibit complex temporal properties. How to use time series data to build knowledge graphs and effectively model them is a challenging problem. Recently, numerous studies use temporal information in time series data to enrich the characteristics of knowledge graphs, endowing knowledge graphs with dynamic features, expanding fact triples into quadruple representation(head entity, relationship, tail entity, time). The knowledge graph which utilizes time-related quadruples to represent knowledge are collectively referred to as temporal knowledge graph. This paper summarizes the research work of temporal knowledge graph representation learning by sorting out and analyzing the corresponding literature. Specifically, it first briefly introduce the background and definition of temporal knowledge graph. Next, it summarizes the advantages of the temporal knowledge graph representation learning method compared with the traditional knowledge graph representation learning method. Then it elaborates on the recent method of temporal knowledge graph representation learning from the perspective of the method modeling facts, introduces the dataset used by the above method and summarizes the main challenges of this technology. Finally, the future research direction is prospected.

**Keywords** Knowledge graph, Deep learning, Representation learning, Temporal information, Dynamic process

到稿日期:2021-10-22 返修日期:2022-05-16

基金项目:国家自然科学基金(62172011)

This work was supported by the National Natural Science Foundation of China(62172011).

通信作者:赵俊峰(zhaojf@pku.edu.cn)

## 1 引言

现实世界中,当我们进行深入数据分析时,需要对多源异构数据的离散信息进行融合,并从海量数据中检索出有价值的信息。如何提高信息检索的准确性并对客观规律进行归纳总结,使数据能够更好地服务于其他应用成为了一个挑战。谷歌为了提升其搜索引擎返回的答案质量,于2012年提出了知识图谱的概念。知识图谱是一种结构化的人类知识表示形式,由实体、关系和语义描述构成。谷歌提出的“things not strings”的语义搜索范式<sup>[1-2]</sup>改进了只关注字符串匹配的检索方法,使得信息检索更关注字符串背后的语义信息,对信息检索模式产生了深远的影响。随着知识图谱的兴起,易贝<sup>[3]</sup>、优步<sup>[4]</sup>、亚马逊<sup>[5]</sup>等公司也纷纷推出企业知识图谱产品。知识图谱已成为了人工智能的前沿研究问题。

结合人类知识来提升算法性能是人工智能的研究方向之一。机器能够理解并利用好知识的关键在于将知识转化为符合其语义的、机器可读的表示。大部分知识图谱使用资源描述框架(Resource Description Framework, RDF)三元组建模实体以及实体之间的关系,三元组的基本结构为(头实体,关系,尾实体),如(颐和园,坐落于,北京市海淀区)。然而,使用这种知识表示形式存在图算法计算复杂度、可扩展性差的计算效率问题以及罕见实体难以进行语义计算的数据稀疏问题<sup>[6]</sup>。以深度学习为代表的表示学习技术将知识图谱中的实体和关系投影到低维连续向量空间,以捕获实体和关系的潜在语义,有效地解决了上述问题。在医疗健康、天气预报、生物学、股票价格预测等领域,由于其领域应用多与时间相关,因此产生了较多与时间密切相关的时序数据,具有高度动态的特点,表现出复杂的时间特性。随着深度学习技术的快速发展,越来越多的深度学习模型能够对时间序列数据进行编码,并且学习到知识,以更完整的方式学习到更复杂的数据模式。然而对于动态演化的大数据背景,使用传统知识图谱表示学习方法仍然面临着很多挑战,主要包括以下两个方面。

### (1) 语义相似关系混淆问题

传统的表示学习方法虽然能够学习到实体和关系的语义信息,但是无法区分具有相似语义的关系,预测时易出现混淆,如(人,出生于,地点)、(人,去世于,地点)这两种类型的三元组中的“出生于”和“去世于”这两种关系的头实体都是人名,尾实体都是地名,语义信息相似。传统的表示学习方法没有利用数据中的时间信息增强实体和关系的表示,经常会混淆这种语义相似关系,使得算法预测性能不理想。

### (2) 干扰项过多

传统的表示学习方法在进行三元组链接预测时,会对候选三元组集中的每个三元组计算一个得分,并选择得分最高的三元组作为预测结果。然而候选三元组中通常存在大量干扰项,传统的表示学习方法无法捕获时间信息对实体和关系演化的影响,使得模型预测时在时序数据上下文中无法捕获知识图谱中实体的历史行为对其当前状态的影响,且不能根据三元组中头实体和尾实体的状态直接对候选三元组的干扰项进行排除,一定程度上也降低了模型预测的性能。

近年来,有许多研究工作利用时间信息提升知识图谱

表示学习的质量,对知识图谱的动态过程进行建模,利用时间信息编码节点、边的表示,将事实三元组拓展为 $(h, r, t, \tau)$ 的四元组表示,表示三元组在时间间隔 $\tau$ 内有效或者在时间点 $\tau$ 时有效: $(h, r, t, \tau)$ 。使用时间相关四元组进行知识表示的知识图谱是时序知识图谱,通过引入时间信息对语义相似关系进行了一定程度的区分,并通过时间一致性信息直接对候选三元组中的干扰项进行排除,提升了实体、关系表示学习的质量以及知识推理的准确性。这一类研究工作被统称为时序知识图谱表示学习。经过调研发现,目前缺乏对时序知识图谱表示学习任务全面的综述。

本文对时序知识图谱表示学习的工作进行了系统性的总结与概括。本文第2节首先介绍时序知识图谱的定义与基本概念以及时序知识图谱表示学习的优势;第3节根据对事实的建模方法、模型对知识推理的设置、模型对数据的处理方法3个角度对该领域的研究工作进行总结整理,为了方便读者理解模型的主要思想和方法,本文对事实建模方法的角度详细地阐释相关研究工作的思想和方法;第4节介绍相关研究工作使用的数据集;第5节对时序知识图谱表示学习仍然存在的困难与挑战进行分析;最后总结全文并展望未来。

## 2 时序知识图谱表示学习的概念与特点

在介绍时序知识图谱表示学习的主要方法、数据集和主要挑战之前,本文先对时序知识图谱表示学习进行简要介绍,包括时序知识图谱的定义与基本概念以及时序知识图谱表示学习的优势。

### 2.1 时序知识图谱的定义与基本概念

传统的知识图谱研究主要集中在静态知识图谱上,静态知识图谱的事实不会随着时间的变化而变化,而对知识图谱中实体和关系的时间动态演化少有研究。然而大多数结构化知识仅在特定时期内有效,例如:针对诸如事件类这种与时间特性紧密相关的知识,需要用一种包含时间信息的动态的知识图谱进行描述。因此,时间信息在知识图谱表示学习中尤为重要。时序知识图谱是关系上带有时间戳信息的多关系有向知识图谱,Jiang等<sup>[7]</sup>于2016年首次将时间信息编码到知识图谱的表示学习中,并通常将时序知识图谱中带有时间戳信息的三元组直接建模为四元组,表示三元组在时间间隔 $\tau$ 内有效或者在时间点 $\tau$ 时有效: $(h, r, t, \tau)$ ,例如四元组(贝拉克·侯赛因·奥巴马,总统,美国,[2008年11月4日,2017年1月20日])表示实体“奥巴马”与“美国”之间在2008年11月4日至2017年1月20日这个时间间隔内具有“总统”关系,后续的时序知识图谱表示学习研究工作也使用Jiang等定义的知识表示形式。我们可以将整个时序知识图谱表示为 $\mathcal{K} = \{(h, r, t, \tau) | h, t \in \mathcal{E}, r \in \mathcal{R}, \tau \in \mathcal{T}\}$ ,其中 $\mathcal{E}$ 是实体集合,包含 $|\mathcal{E}|$ 种不同实体; $\mathcal{R}$ 是关系,包含 $|\mathcal{R}|$ 种不同关系; $\mathcal{T}$ 是时间戳信息集合,包含 $|\mathcal{T}|$ 种不同时间戳。

### 2.2 时序知识图谱表示学习的主要优点

相比传统的静态知识表示学习方法,时序知识图谱表示学习主要具有以下优点。

#### (1) 缓解语义相似关系混淆

针对知识图谱传统表示学习方法无法捕获知识图谱的

动态演化的问题,以及由于基于静态事实的假设无法学习事实的时间动态的问题,时序知识图谱表示学习引入时间信息编码实体和关系的表示演化,捕获知识图谱中存在时间信息和现实世界事实的动态性质。时序知识图谱表示学习可以捕获到实体和关系演化的顺序信息,如对于时序知识图谱的同一实体,其作为头实体的三元组序列通常具有一定的顺序,如(爱因斯坦,出生于,乌尔姆,1879)表示爱因斯坦出生于乌尔姆的事实发生在1879年,(爱因斯坦,获奖,诺贝尔奖,1922)表示爱因斯坦获得诺贝尔奖的事实发生在1922年,(爱因斯坦,去世于,普林斯顿,1955)表示爱因斯坦去世于普林斯顿的事实发生在1955年。将事实发生的时间戳信息引入到表示学习中使得模型能够学习关系表示的因果模式,如在“出生于—>毕业于—>工作于—>去世于”这种描述关系演变发展的时序顺序链中,引入时间信息能帮助模型对“出生于”和“去世于”这种语义相似的关系进行更细粒度的区分。

(2)减少模型预测的干扰项

时序知识图谱表示学习捕获了知识图谱中的时间一致性信息。知识图谱中的很多事实都是在一定时间内有效,例如某个人在某球队效力只是在一段时间内有效,一个国家不能同时拥有两个总统。传统的知识表示学习方法获得的候选预测中不可避免地包含许多不正确的预测。捕获的时间一致性信息也能够帮助我们在预测时排除掉一些错误选择,从而得到更准确的预测结果。通过将事实的时间戳信息融入到表示学习的过程,时序知识图谱表示学习方法在检测表示学习质量的多个链接预测任务中展示出了比传统知识图谱表示学习方法更优异的性能。

(3)辅助时间序列预测

利用好时序知识图谱中的时间信息可以很好地帮助时间序列进行预测。社交网络通信、金融贸易、疫情传播网络等应用场景数据具有高度动态的特点,表现出复杂的时间特性,

例如国家之间的贸易信息可以用时序知识图谱表示,国家之间贸易的信息会影响国家的货币汇率。我们可以利用时序知识图谱的信息对已发生的事实动态以及事实相关性进行建模,以更好地提升货币汇率这种时间序列预测问题的准确性。

3  时序知识图谱表示学习的主要方法

知识图谱能够对实体及其丰富关系的结构化信息进行编码。尽管典型的大型知识图谱可能包含数百万个实体和数十亿个关系事实,但是知识图谱在应用中仍然存在关系缺失或者属性缺失等不完备性问题,以及构建时采用统计方法带来的知识错误问题。为了解决上述问题,时序知识图谱相关研究工作利用知识图谱中事实发生的时间信息增强实体和关系的表示,并在此基础上对推导出的实体间关系进行打分,从而学习到新的实体关系,以增强知识图谱的完备性并且对知识图谱构建过程中出现的错误进行检测。近年来,随着知识图谱的应用越来越广泛,其也引起了越来越多的研究者的关注,使用深度学习方法分析时序数据的热潮为时序知识图谱表示学习的研究带来了新的机遇。

从对事实的建模方法来看,时序知识图谱表示学习的方法可以分为翻译模型、双线性模型、旋转模型、时序点过程、概率分布、图神经网络以及其他模型;从知识推理的设置来看,时序知识图谱表示学习的方法可以分为插值(Interpolation)和外推(Extrapolation),当给定时间戳为 $t_0$ 到 $t_T$ 的知识图谱序列,插值设置对时间戳范围为 $[t_0, t_T]$ 中的缺失事实进行预测,外推设置对时间戳范围为 $(t_T, \infty)$ 的未来事实进行预测;从对时序知识图谱的建模方式来看,时序知识图谱表示学习的方法可以分为使用静态子图的方法和不使用静态子图的方法。使用静态子图的方法即把时序知识图谱按照离散时间点分割为多个静态子图。表1列出了模型对事实的建模方法、发表的会议、模型对知识推理的设置、模型对数据的处理方法。

表 1   相关工作总结  
Table 1   Summary of related works

| 方法            | 会议     | 年份   | 对事实的建模方式   | 推理设置 | 对图谱的建模方式 |
|---------------|--------|------|------------|------|----------|
| t-TransE      | EMNLP  | 2016 | 翻译模型       | 插值   | 不使用静态子图  |
| Trans-TAE     | COLING | 2016 | 翻译模型       | 插值   | 不使用静态子图  |
| KnowEvolve    | ICLR   | 2017 | 时序点过程      | 外推   | 不使用静态子图  |
| TA-DisMult    | EMNLP  | 2018 | 双线性模型      | 插值   | 不使用静态子图  |
| HyTE          | EMNLP  | 2018 | 翻译模型       | 插值   | 使用静态子图   |
| DyRep         | ICLR   | 2019 | 时序点过程      | 外推   | 不使用静态子图  |
| DiachronicEmb | AAAI   | 2020 | 翻译模型/双线性模型 | 插值   | 不使用静态子图  |
| ReNet         | EMNLP  | 2020 | 概率分布       | 外推   | 使用静态子图   |
| TIMEPLEX      | EMNLP  | 2020 | 双线性模型      | 插值   | 不使用静态子图  |
| DArtNet       | IJCAI  | 2020 | 概率分布       | 外推   | 使用静态子图   |
| ChronoR       | AAAI   | 2021 | 旋转模型       | 外推   | 不使用静态子图  |
| CyGNet        | AAAI   | 2021 | 其他模型       | 外推   | 不使用静态子图  |
| RE-GCN        | SIGIR  | 2021 | 图神经网络      | 外推   | 使用静态子图   |

为了方便读者理解模型的主要思想和方法,本文从对事实的建模方法的角度来介绍时序知识图谱表示学习的主要方法。

3.1  翻译模型

受自然语言处理中 Word Embedding 词向量空间的平移不变性的启发<sup>[6]</sup>,Bordes 等<sup>[8]</sup>提出了 TransE 模型,将实体和关系映射到低维嵌入空间来刻画其潜在语义特征,并将关系

嵌入解释为对实体嵌入的翻译操作。他们认为头实体向量经过关系的位移后应尽可能接近尾实体向量,即: $\vec{h} + \vec{r} \approx \vec{t}$ (其中 $\vec{h}$ 为头实体表示, $\vec{r}$ 为关系表示, $\vec{t}$ 为尾实体表示)。而对于非事实三元组,头实体经过关系位移后的表示应远离尾实体表示。TransE 构造非事实三元组集合的方式如式(1)所示:

$$S'_{(h,\ell,t)} = \{(h',\ell,t) \mid h' \in E\} \cup \{(h,\ell,t') \mid t' \in E\}$$

(1)

其中, TransE 对事实三元组的头实体或尾实体进行随机的实体替换, 以此构成非事实三元组样本集合, 训练的损失函数如下:

$$\mathcal{L} = \sum_{(h, t, r) \in S} \sum_{(h', t', r') \in S'_{(h, t, r)}} [\gamma + d(h + \ell, t) - d(h' + \ell, t')]_+ \quad (2)$$

损失函数的设计思路是:  $d(\cdot)$  为在低维实体嵌入空间中向量的距离度量, 而事实三元组中头实体与关系嵌入之和与尾实体嵌入距离越小, 产生的损失就越小; 非事实三元组中头实体与关系嵌入之和与尾实体嵌入距离越大, 产生的损失就越小。通过该损失函数进行训练并不断更新实体和关系的嵌入, 最终研究者使用该算法在大型知识图谱 Freebase 和 WordNet 的链接预测问题上取得了不错的效果。

TransE 虽然因其简单、高效的特性而广受欢迎, 但是也存在一定的缺陷, 如在建模自反关系中, 一对一、一对多、多对一、多对多等复杂关系会由于算法表示空间的局限性而出现性能下降的问题。近年来, 已经有不少研究工作针对 TransE 的缺陷进行了改进, 比如 TransH<sup>[9]</sup>, TransR<sup>[10]</sup>, TransD<sup>[11]</sup>, TranSparse<sup>[12]</sup> 等, 在链接预测等问题上取得了非常不错的效果。然而实体和关系往往是随时间动态演变的, 上述方法无法对时间信息编码并建模知识图谱中实体和关系的动态演化。Jiang 等<sup>[7]</sup> 在三元组的基础上增加了时间维度, 形成四元组  $(h, r, t, \tau)$ , 表示这个事实在时间间隔  $\tau$  内或者时间点  $\tau$  时有效, 并且假设先前的关系向量可以通过时间转换演变为后续的关系向量, 提出了 TransE-TAE-ILP 模型。

如图 1 所示, TransE-TAE-ILP 模型使用转移矩阵 (演化矩阵)  $\mathbf{M}$  捕获关系之间的时间顺序信息。比如“出生于”可以演化为“毕业于”和“去世于”, 但是“去世于”不能演化为“出生于”。这样, 在学习过程中, 语义和时间信息都被嵌入到连续的向量空间中。对于共享头实体的一对四元组, 为了使正确时间顺序的关系比错误的关系的损失更低, TransE-TAE-ILP 定义了如下对于时间顺序的损失函数:

$$g(r_i, r_j) = \|r_i \mathbf{M} - r_j\|_1 \quad (3)$$

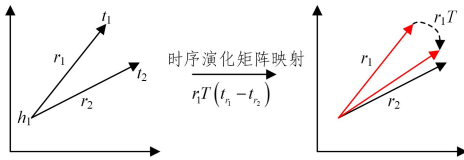


图 1 TransE-TAE-ILP 模型的关系演化矩阵

Fig. 1 Relationship evolution matrix of TransE-TAE-ILP model

TransE-TAE-ILP 将时间顺序的损失以及三元组嵌入学习的损失函数结合起来, 形成新的损失函数进行模型训练。

$$\mathcal{L} = \sum_{x^+ \in \Delta} \sum_{x^- \in \Delta'} [\gamma_1 + f(x^+) - f(x^-)]_+ + \sum_{y^+ \in \Omega_e} \sum_{y^- \in \Omega'_e} [\gamma_2 + g(y^+) - g(y^-)]_+ \quad (4)$$

其中,  $\Delta$  是合法三元组集合,  $\Delta'$  是替换相应实体的错误三元组集合, 对于四元组  $(e_i, r_i, e_j, t_x)$ ,  $y^+ = (r_i, r_j) \in \Omega_e$  被定义为合法的时序关系对。  $\Omega_e$  被定义为:

$$\Omega_e = \{(r_i, r_j) \mid (e_i, r_i, e_j, t_x) \in \Delta, (e_i, r_j, e_k, t_y) \in \Delta, t_x < t_y\} \quad (5)$$

其中,  $r_k$  和  $r_l$  共享相同头实体  $e_i$ , 并且  $y^- = (r_j, r_i) \in \Omega'_e$  是通过

对合法的时序关系对进行逆转操作构造的错误时序关系对。

在对时序关系对进行建模的基础上, TransE-TAE-ILP 考虑了 3 种时间一致性信息。

(1) 时序不相交性: 任何两个具有相同一对一关系以及相同头实体的事实的时间间隔是不能重叠的, 例如一个人在一段时间内只能是另一个人的配偶。

(2) 关系时间顺序: 对于某些具有时间顺序的关系, 一个事实三元组总会先于另一个事实三元组发生, 例如一个人出生于某地肯定是发生在他毕业于某学校之前。

(3) 时序范围性: 有些事实仅在特定时间段内是正确的, 超出该时间范围则无效。

TransE-TAE-ILP 将时间感知推理公式化为具有时间约束的 ILP 问题。TransE-TAE-ILP 使用  $w_{ij}^{(k)}$  表示嵌入模型计算的可靠性; 使用  $x_{ij}^{(k)}$  作为布尔决策变量表示四元组是否成立。TransE-TAE-ILP 的目标是找到对决策变量的最佳分配, 在遵守所有时间约束的同时, 使总体置信度最大化。具体定义如下:

$$\max \sum_{x_{ij}^{(k)}} w_{ij}^{(k)} x_{ij}^{(k)} \quad (6)$$

$$x_{ij}^{(k)} + x_{il}^{(k)} \leq 1, \forall k \in \mathcal{C}^d, t_{x_{ij}^{(k)}} \cap t_{x_{il}^{(k)}} \neq \emptyset \quad (7)$$

$$x_{ij}^{(k)} + x_{il}^{(k')} \leq 1, \forall (k, k') \in \mathcal{C}^p, t_{x_{ij}^{(k)}} \geq t_{x_{il}^{(k')}} \quad (8)$$

$$x_{ij}^{(k)} + x_{il}^{(k')} \leq 1, \forall (k, k') \in \mathcal{C}^o, t_{x_{ij}^{(k)}} \geq t_{x_{il}^{(k')}} \quad (9)$$

其中, 式(7)一式(9)分别为时序不相交性、关系时间顺序、时序范围性对布尔决策变量的约束。

TransE-TAE-ILP 模型的方法也可以移植到 TransH 和 TransR 等模型。由于对于关系的演化过程进行了建模以及利用时间一致性信息进行了约束, TransE-TAE-ILP 模型在实体预测、链接预测等问题上的性能均优于传统的 TransE 模型。自 TransE-TAE-ILP 模型被提出以来, 越来越多的工作开始聚焦于时序知识图谱的表示学习, 并对 TransE-TAE-ILP 模型进行改进与扩展。

基于翻译模型的方法可以灵活地对逆反关系 (如“上位词”和“下位词”)、复合关系 (如“父亲”是“母亲”和“丈夫”的复合关系) 进行建模, 但是无法对对称关系 (如“婚姻”) 灵活建模。

### 3.2 双线性模型

近年来, 部分知识图谱表示学习方法基于线性映射假设, 即将三元组中的头实体和尾实体建模为向量, 把关系建模为从头实体到尾实体的线性映射矩阵, 并且认为如果三元组是正确的, 那么头实体经过线性映射后的向量与尾实体距离相近。线性映射假设将头实体经过线性映射后的向量与尾实体内积构建得分函数:

$$f_r(h, t) = h^T \mathbf{M}_r t \quad (10)$$

其中,  $\mathbf{M}_r \in \mathbb{R}^{d \times d}$  是关系  $r$  的双线性映射矩阵, 基于线性映射假设的具有代表性的传统知识表示学习方法主要包括将关系建模为满秩矩阵的 RESCAL 模型<sup>[13]</sup>、将关系建模为对角矩阵的 DisMult 模型<sup>[14]</sup>、将实体和关系映射到复数空间的 ComplEx 模型<sup>[15]</sup>。

为了改善传统双线性模型忽略时间信息、仅从时间未知的事实三元组中学习、预测性能较差的问题, García-Durán

等<sup>[16]</sup>提出了 TA-DisMult 模型。TA-DisMult 模型将时间戳分解为由一些时间 token 组成的序列(例如 1961 年可以分解为[1y, 9y, 6y, 1y]),并且对于每个时序知识图谱四元组,都提取出一个谓词序列  $p_{\text{seq}}$ (由关系和“自从”“直到”等时间谓词以及时间戳分解的 token 序列组成)。TA-DisMult 模型将  $p_{\text{seq}}$  中的每一个 token 映射到低维向量空间,并将  $p_{\text{seq}}$  转化的 Embedding 序列作为长短期记忆网络(Long Short-Term Memory, LSTM)的输入,长度为  $N$  的谓词序列由 LSTM 的最后一个隐状态表示成如下定义:

$$\mathbf{e}_{p_{\text{seq}}} = \mathbf{h}_N \quad (11)$$

然后,TA-DisMult 将谓词序列  $p_{\text{seq}}$  与事实的头实体和尾实体重新建模为三元组  $(s, p_{\text{seq}}, o)$ ,将得到的谓词序列表示  $\mathbf{e}_{p_{\text{seq}}}$  与头实体和尾实体的嵌入共同输入双线性模型 DisMult 中进行训练:

$$f(s, p_{\text{seq}}, o) = (\mathbf{e}_s \circ \mathbf{e}_o) \mathbf{e}_{p_{\text{seq}}}^T \quad (12)$$

重新建模后的三元组既包含了头实体和尾实体之间发生的关系信息,也包含了时间信息,相比传统双线性模型大大提升了知识图谱补全的准确率。

基于双线性模型的方法可以灵活地对逆反关系和对称关系进行建模,但是不能很好地建模复合关系。

### 3.3 旋转模型

知识图谱中具有多种类型的关系有助于根据知识图谱的已知事实推理出知识图谱的连接模式,包括对称关系(如婚姻)、非对称关系(如孝顺)、语义相反的关系(如上位词和下位词)、复合关系(如我妈的妹妹是我姨妈)。传统的翻译模型可以对语义相反的关系和复合关系进行建模;传统的双线性模型可以很好地对称关系进行建模。然而,没有一种模型能够建模和推断所有类型的关系。Sun 等<sup>[17]</sup>受欧拉分解( $e^{i\theta} = \cos \theta + i \sin \theta$ )的启发,将实体和关系投影到复数向量空间,并且将关系建模为从头实体到尾实体的旋转,提出了 RotatE 模型。对于三元组  $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ ,RotatE 希望头实体经过关系的旋转更加接近于尾实体:

$$\mathbf{t} = \mathbf{h} \circ \mathbf{r}, \text{ where } |\mathbf{r}_i| = 1 \quad (13)$$

其中,  $\mathbf{h}$  为头实体嵌入,  $\mathbf{r}$  为关系嵌入,  $\mathbf{t}$  为尾实体嵌入,  $\circ$  为 Hadmard 积。RotatE 利用了翻译模型和双线性模型的优势,能够建模所有关系类型,在链接预测等问题中取得了新的突破。Sadeghian 等<sup>[18]</sup>受 RotatE 的启发,提出了 ChronoR 模型,将关系和时间信息建模为正交矩阵的一部分,对头实体进行基于关系和时间信息的旋转变换之后,它就会落在其对应的尾实体附近,具体表示如下:

$$\mathbf{Q}_{r,\tau}(\mathbf{h}) = \mathbf{t} \quad (14)$$

$$\mathbf{Q}_{r,\tau} = [\mathbf{r} | \boldsymbol{\tau}]^4$$

其中,  $\mathbf{Q}_{r,\tau}$  是被关系  $r$  和时间戳  $\tau$  参数化的正交矩阵。由于正交矩阵的性质,正交矩阵实质上就是直角坐标系的  $n$  维扩展,可以把正交矩阵看作一个  $k$  维的旋转因子。并且,不同于 RotatE,ChronoR 去掉了矩阵或者向量范数等于 1 的设置,这样既可以旋转又可以缩放。ChronoR 的得分函数定义如下:

$$\begin{aligned} \langle \mathbf{A}, \mathbf{B} \rangle &:= \text{tr}(\mathbf{A}\mathbf{B}^T) \\ \cos(\theta) &:= \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\sqrt{\langle \mathbf{A}, \mathbf{A} \rangle \langle \mathbf{B}, \mathbf{B} \rangle}} \\ g(\mathbf{h}, \mathbf{r}, \mathbf{t}, \tau) &:= \langle \mathbf{Q}_{r,\tau}(\mathbf{h}), \mathbf{t} \rangle \end{aligned} \quad (15)$$

因此,ChronoR 希望头实体进行基于关系和时间信息的旋转变换之后,与尾实体的夹角的余弦值更大,也就是夹角更小。ChronoR 模型不仅继承了旋转模型能够建模多种关系模式的优势,也将时间信息编码到实体的旋转过程中,捕获了实体和关系的动态演化。

基于旋转模型的方法可以建模逆反关系、复合关系和对称关系,但是对于同一实体,只能利用其一跳邻居进行表示学习和推理,缺少了对知识图谱中多跳路径信息的捕获。

### 3.4 时序点过程

世界上的很多事件都是相互关联的,过去的事件可能会导致或者阻止未来事件的发生,发现事件之间的关联对于通过时间序列的历史事件预测未来事件的发生至关重要。时序点过程是一类用于时间序列建模的常用方法,该方法能够捕获事件之间的关联,对历史事件进行建模,实现对未来事件的预测。时序点过程已被广泛应用于模拟地震和余震<sup>[19-22]</sup>、计量经济学的研究<sup>[23-24]</sup>、对犯罪分子网络进行建模<sup>[25]</sup>等应用场景。使用条件强度函数是表征时序点过程的一种重要方法,条件强度函数是给定所有先前事件的下一个事件时间的随机模型。 $\lambda^*(t)dt$  是给定历史事件  $\mathcal{H}_t$ ,在时间窗口  $[t, t+dt)$  中有新时间发生的概率,如下式所示:

$$\lambda^*(t)dt = \mathbb{P}\{\text{event in } [t, t+dt) \mid \mathcal{H}_t\} \quad (16)$$

鉴于时序点过程的以上优点,Trivedi 等<sup>[26]</sup>将时序点过程引入知识表示学习,提出了知识进化网络 Know-Evolve 模型。Know-Evolve 模型将时间建模为一个随机变量,并使用时序点过程对知识图谱中事实的发生进行建模,Know-Evolve 使用以下条件强度函数表征时序点过程:

$$\begin{aligned} \lambda_{r^{e^s, e^o}}(t \mid \bar{t}) &= f(g_{r^{e^s, e^o}}(\bar{t})) * (t - \bar{t}) \\ g_{r^{e^s, e^o}}(t) &= \mathbf{v}^{e^o}(t)^T \cdot \mathbf{R}_r \cdot \mathbf{v}^{e^s}(t^-) \end{aligned} \quad (17)$$

其中,  $t$  是现在事件的时间,  $\bar{t} = \max(t^{e^s} -, t^{e^o} -)$  是在时间  $t$  之前头实体和尾实体在事件中出现的最近的时间点;  $g_{r^{e^s, e^o}}(t)$  是关系  $r$  下所涉及实体的相关性得分;  $\mathbf{R}_r$  是关系权重矩阵,其能够捕获特定关系空间中两个实体的交互,并且能够在基于实体之间的事件历史中的累积知识。然后,Know-Evolve 模型使用新颖的知识进化网络建模动态演化的实体表示。

$$\mathbf{m} = (e^s, \mathbf{r}, e^o, t)_m \in \mathcal{D} \quad (18)$$

$$\mathbf{v}^e(t_p) = \sigma(\mathbf{W}_t^e(t_p - t_{p-1}) + \mathbf{W}^{th} \cdot \mathbf{h}^{e^s}(t_p -)) \quad (19)$$

$$\mathbf{h}^e(t_p -) = \sigma(\mathbf{W}^h \cdot [\mathbf{v}^e(t_{p-1}) \oplus \mathbf{v}^{e^s}(t_p -) \oplus \mathbf{r}_{p-1}^e]) \quad (20)$$

$$\mathbf{v}^{e^o}(t_q) = \sigma(\mathbf{W}_t^{e^o}(t_q - t_{q-1}) + \mathbf{W}^{th} \cdot \mathbf{h}^{e^o}(t_q -)) \quad (20)$$

$$\mathbf{h}^{e^o}(t_q -) = \sigma(\mathbf{W}^h \cdot [\mathbf{v}^{e^o}(t_{q-1}) \oplus \mathbf{v}^{e^s}(t_q -) \oplus \mathbf{r}_{q-1}^{e^o}])$$

对于一个四元组事件  $m$ ,  $m$  是头实体  $e^s$  的第  $p$  次事件,  $m$  是尾实体  $e^o$  的第  $q$  次事件,上述计算方式体现了在较短的时间内,实体倾向于与其行为相似的其他实体形成关系的假设。

对于给定的在时间窗口  $[0, T)$  中记录的事实集合,Know-Evolve 模型通过最小化条件强度函数的负对数似然来学习。

$$\begin{aligned} \mathcal{L} = & - \sum_{p=1}^N \log(\lambda_{r^{e^s, e^o}}(t_p \mid \bar{t}_p)) + \\ & \underbrace{\sum_{r=1}^r \sum_{e^s=1}^n \sum_{e^o=1}^n \int_0^T \lambda_{r^{e^s, e^o}}(\tau \mid \bar{\tau}) d\tau}_{\text{survival term}} \end{aligned} \quad (21)$$

其中,第一项为最大化两个实体之间特定类型事件的概率;第二项为惩罚给定观察窗口中所有可能的实体对之间不可能出现的事件类型。每个关系需要计算所有可能实体对的生存概率。通过时序点过程的方法进行时序知识图谱表示学习,捕获了事件之间的复杂时间依赖性,并能从历史实体交互和关系演化中捕获知识。

时序点过程模型虽然能够捕获知识图谱中事件之间的关联,对历史事件进行建模并对未来事件进行预测,但是不能对发生在同一时间窗口的并发事件进行建模。

### 3.5 概率分布

针对基于时序点过程方法的缺陷,Jin 等<sup>[27]</sup>将事件的发生建模为以知识图谱的历史时间序列为条件的概率分布,提出了 Re-Net 这个新颖的时序知识图谱补全框架。Re-Net 将时序知识图谱表示为序列,然后在序列上以自回归的方式定义了事件的联合概率分布:

$$\begin{aligned} p(G) &= \prod_t \prod_{(s_t, r_t, o_t) \in G_t} p(s_t, r_t, o_t \mid G_{t-m:t-1}) \\ &= \prod_t \prod_{(s_t, r_t, o_t) \in G_t} p(s_t \mid G_{t-m:t-1}) \cdot p(r_t \mid s_t, G_{t-m:t-1}) \cdot \\ &\quad p(o_t \mid s_t, r_t, G_{t-m:t-1}) \end{aligned} \quad (22)$$

RE-NET 把联合概率分布分解成了一系列的条件概率分布。假设  $t$  时刻事件发生的概率只取决于之前  $m$  步发生的事件,对于每个条件概率  $p(G_t \mid G_{t-m:t-1})$ ,进一步假设  $G_t$  中的事件在  $G_{t-m:t-1}$  发生的情况下是相互独立的。

Re-Net 使用循环神经网络(Recurrent Neural Networks, RNN)建模了知识图谱的全局表示  $\mathbf{H}_t$  和局部表示  $\mathbf{h}_t(s)$  和  $\mathbf{h}_t(s, r)$ 。全局表示汇总了整个知识图谱直到时间戳  $t$  的全局信息;局部表示更多地关注相应的头实体  $s$  及其连接关系  $r$ , 捕获与相应实体和关系相关的知识。具体表示如下:

$$\begin{aligned} \mathbf{H}_t &= \text{RNN}^1(g(G_t), \mathbf{H}_{t-1}) \\ \mathbf{h}_t(s, r) &= \text{RNN}^2(g(\mathbf{N}_t^{(s)}), \mathbf{H}_t, \mathbf{h}_{t-1}(s, r)) \\ \mathbf{h}_t(s) &= \text{RNN}^3(g(\mathbf{N}_t^{(s)}), \mathbf{H}_t, \mathbf{h}_{t-1}(s)) \end{aligned} \quad (23)$$

Re-Net 模型使用 R-GCN<sup>[28]</sup> 聚合器,可以聚合来自多种关系和多跳邻居的信息,从而更新每个节点的表示:

$$g(\mathbf{N}_t^{(s)}) = \mathbf{h}_s^{(t+1)} = \sigma(\sum_{r \in R} \sum_{o \in \mathbf{N}_t^{(s,r)}} \frac{1}{C_s} \mathbf{W}_r^{(d)} \mathbf{h}_o^{(t)} + \mathbf{W}_0^{(d)} \mathbf{h}_s^{(t)}) \quad (24)$$

对于知识图谱补全,给定  $(s, r)$ , 预测尾实体  $o$  可以被视为一次多分类任务,每一类都对应一个实体。在给定  $s$  和  $o$  的情况下预测关系  $r$ ,也可以被看作多分类任务。模型采用多分类的交叉熵损失函数进行训练:

$$\begin{aligned} \mathcal{L} = - \sum_{(s, r, o, t) \in G} \log p(o_t \mid s_t, r_t) + \lambda_1 \log p(r_t \mid s_t) + \\ \lambda_2 \log p(s_t) \end{aligned} \quad (25)$$

由于将事件的发生建模为以知识图谱的历史时间序列为条件的概率分布,基于概率分布的方法可以执行有效的多步推断,以预测遥远的未来看不见的实体关系,提高了时序知识图谱表示学习的质量和补全的准确性。但是基于概率分布的方法对全局表示和局部表示使用 RNN 进行建模,无法解决长期依赖问题,并且所有时间步的表示对于当前表示的贡献相同。

### 3.6 图神经网络

图卷积神经网络(Graph Neural Networks, GCN)<sup>[29]</sup>、

图注意力网络(Graph Attention Networks, GAT)<sup>[30]</sup> 等图神经网络模型及其变种是在图形结构中捕获内容和结构特征的代表性模型,此类模型能够利用图形中的高阶连接信息增强节点的表示学习。知识图谱也是一种图形结构,因此有很多研究工作将图神经网络模型应用到知识图谱中。Li 等<sup>[31]</sup>通过关系感知 GCN 来捕获每个时间戳的知识图谱内的结构依赖性,通过门控循环组件自回归建模历史知识图谱序列,以有效地捕获时间序列模式,并且结合静态信息的约束提升知识表示学习的质量,提出了一种基于图卷积网络的新型循环进化网络,称为 RE-GCN。

RE-GCN 由一个进化单元和多任务评分函数组成。其中,进化单元由一个关系感知 GCN、两个门控循环组件和一个静态图约束组件组成。关系感知 GCN 尝试在每个时间戳捕获知识图谱内的结构依赖性;两个门循环组件自动回归对历史知识图谱序列进行建模。进化单元的输入为最近  $m$  个时间戳的知识图谱序列  $(\{G_{t-m+1}, \dots, G_t\})$ , 并将其映射为实体嵌入矩阵序列  $(\{\mathbf{H}_{t-m+1}, \dots, \mathbf{H}_t\})$  和关系嵌入矩阵序列  $(\{\mathbf{R}_{t-m+1}, \dots, \mathbf{R}_t\})$ 。

RE-GCN 使用  $\omega$ -层关系感知 GCN 来对结构依赖性进行建模。对于在时间戳为  $t$  的知识图谱,尾实体在第  $l+1$  层的表示由其自身在第  $l$  层的表示、头实体节点表示以及关系表示通过消息传递框架得到:

$$\vec{h}_{o,t}^{l+1} = f\left(\frac{1}{C_o(s,r)} \sum_{(s,r,o) \in e_j} \mathbf{W}_1^l (\vec{h}_{s,t}^l + \vec{r}_t) + \mathbf{W}_2^l \vec{h}_{o,t}^l\right) \quad (26)$$

潜在的序列模式是通过堆叠  $\omega$ -层关系感知 GCN 来建模的。RE-GCN 使用一个时间门控循环组件来解决 GCN 大量堆叠导致的梯度消失问题,其中实体嵌入矩阵  $\mathbf{H}_t$  由两部分组成:关系感知 GCN 在时间戳  $t$  处最后一层的输出以及前一个时间戳的实体嵌入矩阵。

$$\begin{aligned} \mathbf{H}_t &= \mathbf{U}_t \otimes \mathbf{H}_t^r + (1 - \mathbf{U}_t) \otimes \mathbf{H}_{t-1} \\ \mathbf{U}_t &= \sigma(\mathbf{W}_4 \mathbf{H}_{t-1} + \mathbf{b}) \end{aligned} \quad (27)$$

RE-GCN 使用 GRU<sup>[32]</sup> 捕获关系的序列模式,通过对关系  $r$  在时间戳  $t-1$  处的相关实体进行平均池化操作,并将其在静态知识图谱中的嵌入进行拼接作为 GRU 的输入,最终得到时间戳  $t$  处的关系嵌入矩阵。

$$\begin{aligned} \vec{r}'_t &= [\text{pooling}(\mathbf{H}_{t-1, V_{r,t}}), \vec{r}] \\ \mathbf{R}_t &= \text{GRU}(\mathbf{R}_{t-1}, \mathbf{R}'_t) \end{aligned} \quad (28)$$

RE-GCN 将静态图合并到进化表示的建模中,对实体的名称字符串进行解析,构建出静态知识图谱,并使用 R-GCN 模型获得实体的静态嵌入:

$$\vec{h}_s^t = \gamma \left( \frac{1}{C_i(r', j)} \sum_{(i, r', j) \in e'} \mathbf{W}_t \vec{h}_{i'}^t(j) \right) \quad (29)$$

使用 ConvTransE 作为解码器计算实体预测和关系预测的概率,构成多任务学习的得分函数,并计算损失:

$$\begin{aligned} \tilde{p}(o \mid s, r, \mathbf{H}_t, \mathbf{R}_t) &= \sigma(\mathbf{H}_t \text{ConvTransE}(\vec{s}_t, \vec{r}_t)) \\ \tilde{p}(r \mid s, o, \mathbf{H}_t, \mathbf{R}_t) &= \sigma(\mathbf{R}_t \text{ConvTransE}(\vec{s}_t, \vec{o}_t)) \\ \mathcal{L}^e &= \sum_{t=0}^{T-1} \sum_{(s, r, o, t+1) \in e_{t+1}} \sum_{i=0}^{|V|-1} y_{t+1,i}^e \log p_i(o \mid s, r, \mathbf{H}_t, \mathbf{R}_t) \\ \mathcal{L}^r &= \sum_{t=0}^{T-1} \sum_{(s, r, o, t+1) \in e_{t+1}} \sum_{i=0}^{|R|-1} y_{t+1,i}^r \log p_i(r \mid s, o, \mathbf{H}_t, \mathbf{R}_t) \end{aligned} \quad (30)$$

以 RE-GCN 为代表的基于图神经网络的时序知识图谱表示学习方法,通过如图卷积网络的消息传递框架捕获了相同时间戳下知识图谱中实体之间的结构依赖性,并通过图神经网络的堆叠捕获序列信息模式,提高了计算效率。但是基于图神经网络模型的方法由于建模图结构的动态演变需要堆叠很多层,因此会出现过平滑(Over-smoothing)和过拟合(Overfitting)的问题。

3.7 其他模型

除上述相关研究工作以外,也有研究者将文本摘要生成中的复制机制引入时序知识图谱表示学习中。Zhu 等<sup>[33]</sup>认为更精确的时序事实推论应该充分利用历史中已知的事实,提出了一种新的基于时间感知复制机制的时序知识图谱的表示学习模型 CyGNet。该模型不仅能够预测未来的事实,还可以识别重复出现的事实。CyGNet 首先针对头实体、关系、时间戳定义了历史词汇  $H_{t_k}^{(s,p)}$ ,具体表示如下:

$$H_{t_k}^{(s,p)} = \{o \mid (s, p, o, t_k) \in \mathcal{G}_{(t_1, t_{k-1})}\} \tag{31}$$

其中,历史词汇  $H_{t_k}^{(s,p)}$  表示在时间戳  $t_k$  以前与头实体  $s$  建立关系  $p$  的所有尾实体集合,可以用一个 **multi-hot** 向量表示,如果在集合中,则相应实体位置的向量值为 1。

CyGNet 模型由两部分组成:复制模式和生成模式。复制模式识别重复出现的事实,并通过从历史已知事实中进行复制来相应地预测将来的事实;生成模式从整个实体用户中预测实体,CyGNet 将复制模式和生成模式中的概率预测组合在一起,以输出最终预测。

对于复制模式,如果正在进行的预测  $(s, p, ?, t_k)$  具有历史词汇量  $H_{t_k}^{(s,p)}$ ,CyGNet 将会增加在历史词汇表中被选择的对象实体的估计概率,复制模式首先使用多层感知机生成一个索引向量,表示如下:

$$v_q = \tanh(W_c[s, p, t_k] + b_c) \tag{32}$$

同时为了缩小候选空间,CyGNet 对历史词汇进行更改,将复制模式不感兴趣的实体的索引所对应的 **multi-hot** 值由 0 更改为一个较小的负数,并输出预测尾实体的概率分布。

生成模式则负责从整个实体词汇表  $\epsilon$  中选择对象实体来预测事实。通过生成模式将预测的事实视作全新的事实,无须参考历史,具体表示如下:

$$\begin{aligned} g_g &= W_g[s, p, t_k] + b_g \\ P(g) &= \text{softmax}(g_g) \end{aligned} \tag{33}$$

通过复制模式和生成模式的结合,CyGNet 模型既捕获了一些事实的循环复发性,从历史中学习到了实体和关系的演化知识,也能够从全局所有可能的结果中预测事实,并在知识图谱实体预测、链接预测等任务上较之前的模型有一定的提升。这种方法为时序知识图谱表示学习提供了全新的视角,值得关注。

4 时序知识图谱表示学习数据集

本节主要介绍时序知识图谱表示学习研究工作使用的数据集 ICEWS(Integrated Crisis Early Warning System),YA-

GO15K,GDELT(Global Database of Events, Language, and Tone)。

4.1 ICEWS

综合危机早期预警系统(ICEWS)是一个存储库,其中包含具有特定时间戳的政治事件。这些政治事件通过逻辑谓词(如“访问”“谈判”等)将实体(如“国家”“总统”等)与许多其他实体联系起来<sup>1)</sup>。该存储库以转储形式组织,其中包含从 1995 年到 2015 年发生的事件。García-Durán 等<sup>[16]</sup>根据 ICEWS 创建了两个时序知识图谱数据集,分别称为 ICEWS 2014 和 ICEWS 2005-15。其中,前者覆盖时间范围较短,只包含 2014 年发生的所有事件;后者覆盖时间范围较长,包含 2005—2015 年之间发生的所有事件。

4.2 YAGO15K

García-Durán 等<sup>[16]</sup>将 FB15K<sup>[8]</sup>中的实体与包含时间信息的 YAGO 中的实体对齐,最终数据集是所有成功对齐的事实的结果。

4.3 GDELT

全球事件、语言、语气数据库(GDELT)是世界上最大的政治事件开放数据库。GDELT 主要包含两大数据集:Event Database(事件数据库)和 Global Knowledge Graph(GKG,全球知识图谱)。它记录了从 1969 年至今的新闻,并且每 15 min 更新一次数据,代表政治领袖、组织和国家等实体一个月內不同类型的事件<sup>2)</sup>。

5 时序知识图谱表示的主要挑战

以 Trans-TAE-ILP 为代表的模型提升了时序知识图谱表示学习的质量,相比传统的知识表示学习方法提升了知识图谱补全等任务的准确性,现已成为知识图谱表示学习的研究热点。然而,随着研究的深入,时序知识图谱表示学习的研究面临着一些新的挑战。

5.1 时间信息编码

Trans-TAE-ILP 等模型使用转移矩阵捕获关系随时间的演化过程,只学习了关系的时序顺序链,并将其作为传统静态知识图谱嵌入学习方法的约束来提高对时序知识图谱补全的准确性,其本质上对时间信息的使用不直接,并不是时间感知的。此外,Trans-TAE-ILP 模型只对关系的时间先后顺序进行建模,并没有考虑关系之间的时间间隔,这也会对模型的预测性能产生影响。

如何对事实的准确时间戳进行编码并将其引入到实体和关系的表示中,使模型能够学习到更多的知识,对实体和关系的动态演化进行更准确的建模是一个挑战。近年来有一些研究工作尝试解决这个问题。

Dasgupta 等<sup>[34]</sup>受 TransH<sup>[9]</sup>思想的启发提出了一种新颖的时间感知的 KG 嵌入方法——HyTE 模型,该模型将一定时间范围内的动态知识图谱分为多个静态子图,每一个子图与一个时间戳相对应,并且将实体和关系的嵌入投影到时间戳特定的超平面上,直接在所学得的嵌入中编码时间信息。

1) <http://www.icews.com/>

2) <http://www.gdeltproject.org/>

Jain 等<sup>[35]</sup>提出了 TIMEPLEX 方法,该方法将实体、关系和时间嵌入到复数空间,将时间差异作为显式特征,使学习到的嵌入能够捕获关系之间的隐式时间关系。另外,其使用高斯分布建模复发关系的固定循环周期以及特定实体关系之间的时间差异。

上述两种方法均解决了 Trans-TAE-ILP 编码时间信息不够全面的问题,在相关数据集上的实验结果表明,上述两种方法的性能均取得了显著提升,其中 HyTE 模型在 MR (Mean Rank) 指标上相比 SOTA 模型提升了 6%,TIME- PLEX 模型在 MRR (Mean Reciprocal Ranking) 指标上相比 SOTA 模型提升了 10%。

5.2 实体关系动态演化建模

Trans-TAE-ILP 等模型对关系或者时间戳进行时序建模,但是对每个实体学习的表示仍然是使用静态知识表示学习的方法,没有对实体随着时间的变化进行建模。然而实体也会随着时间的推移而发生动态演变,例如,已知事实四元组 (Mary, Liked, God Father, 1995),但是 2021 年“Mary”和“God Father”的特征(如性格)会发生很大变化,会对链接预测产生一定的影响。因此,如何根据同一实体在不同时间戳的上下文让实体学习到既符合知识上下文又符合实体动态演变进程的表示,以及如何在对实体状态进行动态建模的过程中对关系的动态演化过程同步建模也是一个挑战。

Goel 等<sup>[36]</sup>受自然语言处理的 Diachronic Word Embedding 启发,认为实体有些特征是随时间变化的,有些特征是固定不变的。将实体嵌入构建为一个函数,输入为该实体与时间戳,输出为该实体表示,在静态知识图谱表示学习方法的基础上配备了 Diachronic Embedding 功能来构建时序知识图谱补全,提出了 DE-Simple 模型。DE-Simple 模型解决了 Trans-TAE-ILP 模型无法捕获实体特征演化的问题,显著提升了模型性能,在 MRR 指标上相比 SOTA 模型提升了 10%。

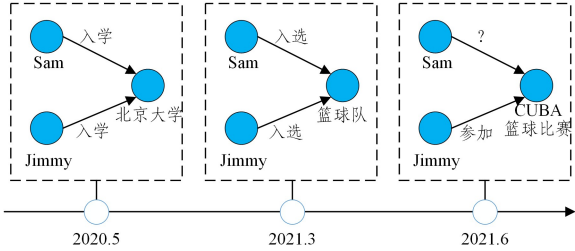
5.3 全局演化建模

TransE-TAE-ILP 等模型将整个时序过程的关系记录下来,通过时间戳来进行区分。这种时间建模的方式非常细粒度,只针对实体的上下文动态演化过程对其表示进行建模,即当前模型仅关注局部上下文的变化过程,但是很难捕获知识图谱全局的实体和关系演化动态,也不能对整体图结构的演化进行建模。然而整体图结构的变化表明了整个知识图谱随时间的进化趋势,对于图谱中局部节点和关系的表示学习质量的提升大有裨益。近年来,有很多方法使用静态子图分割的方式处理时序知识图谱,即将时序知识图谱按照离散时间点分割为多个静态子图,表示为:

$$G=\{G_1,\cdots,G_T\}$$
 (34)

其中, $G_T$ 表示在时间戳  $T$  下所有有效的事实集合中的实体和关系构成的子图。通过静态子图分割的方法可以捕获全局图结构的拓扑结构变化以及图中节点的相互作用变化,这种方法的本质是将时序知识图谱建模为一个对应于不同时间戳的知识图谱序列,每个知识图谱都包含在同一时间戳共同发生的事实。其中每个知识图谱中的所有并发事实在结构上具有依赖性,而在时间上,相邻的事实具有信息序列模式。

近年来,已有不少研究工作使用静态子图分割的方式处理时序知识图谱。Jin 等<sup>[27]</sup>把时序知识图谱根据时间戳分解为静态子图序列,并且根据历史子图序列预测概率分布  $p(G)$ 。根据静态子图序列进行知识推理的示例如图 2 所示。



注:在 2021 年 6 月 Sam 与 CUBA 篮球比赛会发生什么?

图 2 基于静态子图序列的推理示例

Fig. 2 Example of reasoning based on sequence of static subgraphs

Garg 等<sup>[37]</sup>将时序知识图谱根据时间戳分解为静态子图序列,并将单个静态子图分解为静态属性子图和静态结构子图,将动态知识图的信息纳入到时间序列预测中,学习了图中每个节点的静态嵌入以及依赖于时间属性值的动态嵌入,对属性预测和链接预测的嵌入进行了联合训练,提出了 DArt-Net 模型。

上述方法通过新颖的处理方式捕获了知识图谱的全局动态演化信息,为时序知识图谱的表示学习提供了新的思路。

6 时序知识图谱表示学习未来研究方向展望

时序知识图谱表示学习方法在知识图谱实体预测、链接预测等任务上表现出了卓越的性能,并且相对于传统知识表示学习方法而言,其适合更多的应用场景,在金融、医疗、生物信息学、交通等领域的应用有着非常好的前景。除了上一节中提到的面临的主要挑战外,还有一些问题有待解决。本节将对时序知识图谱表示学习的未来方向进行展望。

6.1 利用逻辑规则增强时序知识图谱表示学习的准确性和可解释性

基于连续向量的表示学习方法虽然能够发现人们不易总结出来的隐性知识和潜在假设,并且能够更容易与其他深度学习模型集成,以进行更多任务的实现,但是可解释性非常差。对于这个问题,图结构中的关联路径和规则为知识表示学习和推理提供了可解释性。近年来,也有越来越多的工作<sup>[12,38-39]</sup>将表示学习与规则学习结合起来,用两种方法的优势互相弥补对方的不足,以提高推理的鲁棒性和效率,并为表示学习提供可解释性。然而,在时序知识图谱背景下并没有研究工作结合逻辑规则对表示学习进行约束以提升准确性和可解释性。如何将逻辑规则以及时间信息同时编码到实体和关系的表示是一个值得研究的问题。

6.2 增强稀疏实体编码能力的时序知识图谱表示学习

时序知识图谱中存在许多稀疏实体,即在所有的已知事实集合中出现频率较低的实体,因此已知事实集合中关于该实体随时间的演化信息同样是稀疏的。时序知识图谱表示学习方法对稀疏实体的编码能力较差,从而对模型的整体性能产生了一定的影响。我们需要探索一种能够增强稀疏实体编码能力的时序知识图谱表示学习方案。

6.3  时序知识图谱快速在线表示学习

时序知识图谱的表示学习方法虽然能够通过分布式表示显著提升计算效率,但是在处理大规模数据时差强人意。在金融等许多应用场景中每天都有大量的增量数据,每当有新的实体添加到知识图谱中,基于表示学习的方法需要花费大量时间进行重新训练,以实现为新实体学得表示。因此,探索一种时序知识图谱快速在线表示学习方法对于知识图谱在金融等领域的应用具有重要意义。

6.4  时序知识图谱在推荐系统的应用

为了缓解推荐系统的数据稀疏和冷启动问题,有很多工作<sup>[40-51]</sup>利用知识图谱中的语义信息增强对用户和商品的学习,从而提升推荐系统的准确性与可解释性。然而,目前绝大多数工作仅局限于传统的静态知识图谱,推荐系统中用户与商品的交互具有很强的时效性(用户在不同时间对商品的收藏、购买等行为刻画了不同的用户偏好),如果能够利用好时间信息,将时序知识图谱表示学习方法与推荐系统相结合,对提升推荐效果具有重要意义。

**结束语**  通过对时序知识图谱表示学习的相关研究工作进行梳理,我们认为时序知识图谱表示学习具有重要意义。现实世界的绝大多数应用场景数据具有高度动态的特点,表现出复杂的时间特性。时序知识图谱表示学习将时间信息编码到实体和关系的表示中,能够更好地应用到现实应用场景中的知识图谱补全等任务中,已成为知识表示学习的热点和趋势。

本文对时序知识图谱表示学习的主要方法进行了梳理,对其最新进展进行了综述,对每类方法的核心思想以及代表性算法进行了对比分析,总结了时序知识图谱表示学习方法的研究规律,简要介绍了时序知识图谱表示学习面临的挑战以及可能的解决方案,并且对其未来研究方向进行了探讨。

参 考 文 献

[1] NOY N,GAO Y,JAIN A,et al. Industry-scale knowledge graphs: lessons and challenges [J]. Communications of the ACM,2019,62(8):36-43.

[2] SINGHAL A. Introducing the Knowledge Graph: things, not strings [EB/OL]. Google Blog. <https://www. blog. google/ products/ search/ introducing-knowledge-graph-things-not/>.

[3] PITTMAN R J,SRIVASTAVA A,HEWAVITHARANA S, et al. 2017. Cracking the Code on Conversational Commerce [EB/OL]. eBay Blog. <https://www. ebayinc. com/stories/ news/cracking-the-code-on-conversationalcommerce/>.

[4] HAMAD F,LIU I,ZHANG X X. Food Discovery with Uber Eats: Building a Query Understanding Engine [EB/OL]. Uber Engineering Blog. <https://eng. uber. com/uber-eats-query-understanding/>.

[5] KRISHNAN A. Making search easier: How Amazon’s Product Graph is helping customers find products more easily [EB/OL]. Amazon Blog. <https://blog. aboutamazon. com/innovation/ making-search-easier>.

[6] LIU Z Y,SUN M S,LIN Y K,et al. Knowledge Representation Learning: A Review[J]. Journal of Computer Research and Development,2016,53(2):247-261.

[7] JIANG T,LIU T,GE T,et al. Encoding temporal information

for time-aware link prediction [C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:2350-2354.

[8] BORDES A,USUNIER N,GARCIA-DURAN A,et al. Translating embeddings for modeling multi-relational data [C] // Neural Information Processing Systems(NIPS). 2013:1-9.

[9] WANG Z,ZHANG J,FENG J,et al. Knowledge graph embedding by translating on hyperplanes [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2014:1112-1119.

[10] LIN Y,LIU Z,SUN M,et al. Learning entity and relation embeddings for knowledge graph completion [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2015: 2181-2187.

[11] JI G,LIU K,HE S,et al. Knowledge graph completion with adaptive sparse transfer matrix [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2016:985-991.

[12] ZHANG W,PAUDEL B,WANG L,et al. Iteratively learning embeddings and rules for knowledge graph reasoning [C] // The World Wide Web Conference. 2019:2366-2377.

[13] NICKEL M,TRESP V,KRIEDEL H P. A three-way model for collective learning on multi-relational data [C] // ICML. 2011: 809-816.

[14] YANG B,YIH W,HE X,et al. Embedding entities and relations for learning and inference in knowledge bases [J]. arXiv:1412. 6575,2014.

[15] TROUILLON T,WELBL J,RIEDEL S,et al. Complex embeddings for simple link prediction [C] // International Conference on Machine Learning. PMLR,2016:2071-2080.

[16] GARCÍA-DURÁN A,DUMANČIĆ S,NIEPERT M. Learning sequence encoders for temporal knowledge graph completion [J]. arXiv:1809. 03202,2018.

[17] SUN Z,DENG Z H,NIE J Y,et al. Rotate: Knowledge graph embedding by relational rotation in complex space [J]. arXiv: 1902. 10197,2019.

[18] SADEGHIAN A,ARMANDPOUR M,COLAS A,et al. ChronoR: Rotation Based Temporal Knowledge Graph Embedding [J]. arXiv:2103. 10379,2021.

[19] HAWKES A G. Point spectra of some mutually exciting point processes [J]. Journal of the Royal Statistical Society: Series B (Methodological),1971,33(3):438-443.

[20] HAWKES A G. Spectra of some self-exciting and mutually exciting point processes [J]. Biometrika,1971,58(1):83-90.

[21] HAWKES A G,OAKES D. A cluster process representation of a self-exciting process [J]. Journal of Applied Probability,1974, 11(3):493-503.

[22] OGATA Y. Space-time point-process models for earthquake occurrences [J]. Annals of the Institute of Statistical Mathematics, 1998,50(2):379-402.

[23] BACRY E,IUGA A,LASNIER M,et al. Market impacts and the life cycle of investors orders [J]. Market Microstructure and Liquidity,2015,1(2):1550009.

[24] BACRY E,JAISSON T,MUZY J F. Estimation of slowly decreasing hawkes kernels: application to high-frequency order book dynamics [J]. Quantitative Finance,2016,16(8):1179-1201.

[25] CHANDRA R,ZHANG M. Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction [J]. Neurocomputing,2012,86:116-123.

- [26] TRIVEDI R, FARAJTABAR M, WANG Y, et al. Know-evolve: Deep reasoning in temporal knowledge graphs[J]. arXiv:1705.05742, 2017.
- [27] JIN W, QU M, JIN X, et al. Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020:6669-6683.
- [28] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//European Semantic Web Conference. Cham:Springer, 2018:593-607.
- [29] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [30] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv:1710.10903, 2017.
- [31] LI Z, JIN X, LI W, et al. Temporal Knowledge Graph Reasoning Based on Evolutional Representation Learning[J]. arXiv:2104.10353, 2021.
- [32] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv:1412.3555, 2014.
- [33] ZHU C, CHEN M, FAN C, et al. Learning from History: Modeling Temporal Knowledge Graphs with Sequential Copy-Generation Networks[J]. arXiv:2012.08492, 2020.
- [34] 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. 2018:2001-2011.
- [35] JAIN P, RATHI S, CHAKRABARTI S. Temporal Knowledge Base Completion: New Algorithms and Evaluation Protocols [J]. arXiv:2005.05035, 2020.
- [36] GOEL R, KAZEMI S M, BRUBAKER M, et al. Diachronic embedding for temporal knowledge graph completion [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020:3988-3995.
- [37] GARG S, SHARMA N, JIN W, et al. Temporal attribute prediction via joint modeling of multi-relational structure evolution [J]. arXiv:2003.03919, 2020.
- [38] LIN Y, LIU Z, LUAN H, et al. Modeling relation paths for representation learning of knowledge bases[J]. arXiv:1506.00379, 2015.
- [39] GUO S, WANG Q, WANG L, et al. Knowledge graph embedding with iterative guidance from soft rules[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018:4816-4823.
- [40] ZHANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016:353-362.
- [41] WANG H, ZHANG F, HOU M, et al. SHINE: signed heterogeneous information network embedding for sentiment link prediction [C]//Proceedings of the 11th ACM International Conference on Web Search and Data Mining. 2018:592-600.
- [42] HUANG J, ZHAO W X, DOU H, et al. Improving sequential recommendation with knowledge-enhanced memory networks [C]//Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018:505-514.
- [43] WANG H, ZHANG F, WANG J, et al. RippleNet: propagating user preferences on the knowledge graph for recommender systems [C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018:417-426.
- [44] YU X, REN X, SUN Y, et al. Personalized entity recommendation: a heterogeneous information network approach [C]//Proceedings of the 7th ACM International Conference on Web Search and Data Mining. 2014:283-292.
- [45] SUN Z, YANG J, ZHANG J, et al. Recurrent knowledge graph embedding for effective recommendation [C]//Proceedings of the 12th ACM Conference on Recommender Systems. 2018:297-305.
- [46] WANG X, WANG D, XU C, et al. Explainable reasoning over knowledge graphs for recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019:5329-5336.
- [47] SONG W, DUAN Z, YANG Z, et al. Explainable knowledge graph-based recommendation via deep reinforcement learning [J]. arXiv:1906.09506, 2019.
- [48] HUANG X, FANG Q, QIAN S, et al. Explainable interaction-driven user modeling over knowledge graph for sequential recommendation [C]//Proceedings of the 27th ACM International Conference on Multimedia. 2019:548-556.
- [49] WANG H, ZHANG F, WANG J, et al. Exploring high-order user preference on the knowledge graph for recommender systems [J]. ACM Transactions on Information Systems, 2019, 37: 1-26.
- [50] WANG M, LIU M, LIU J, et al. Safe medicine recommendation via medical knowledge graph embedding [J]. arXiv:1710.05980, 2017.
- [51] PALUMBO E, RIZZO G, TRONCY R. Entity2rec: learning user-item relatedness from knowledge graphs for top-n item recommendation [C]//Proceedings of the 11th ACM Conference on Recommender Systems. 2017:32-36.



**XU Yong-xin**, born in 1998, postgraduate. His main research interests include knowledge graph and so on.



**ZHAO Jun-feng**, born in 1974, Ph.D., research professor, is a member of China Computer Federation. Her main research interests include big data analysis, knowledge graph, urban computing and so on.