S 计算机科学 COMPUTER SCIENCE

邻域双向聚合与全局感知的TKG链接预测模型

唐绍赛, 申德荣, 寇月, 聂铁铮

引用本文

唐绍赛, 申德荣, 寇月, 聂铁铮.邻域双向聚合与全局感知的TKG链接预测模型[J]. 计算机科学, 2023, 50(8): 177-183.

TANG Shaosai, SHEN Derong, KOU Yue, NIE Tiezheng. Link Prediction Model on Temporal Knowledge Graph Based on Bidirectionally Aggregating Neighborhoods and Global Aware [J]. Computer Science, 2023, 50(8): 177-183.

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)

基于时序知识图谱嵌入的短期地铁客流量预测

Short-term Subway Passenger Flow Forecasting Based on Graphical Embedding of Temporal Knowledge 计算机科学, 2023, 50(7): 213-220. https://doi.org/10.11896/jsjkx.220600120

基于对比预测的自监督动态图表示学习方法

Self-supervised Dynamic Graph Representation Learning Approach Based on Contrastive Prediction 计算机科学, 2023, 50(7): 207-212. https://doi.org/10.11896/jsjkx.220500093

混合曲率空间用于多关系异构知识图谱链接补全

Mixed-curve for Link Completion of Multi-relational Heterogeneous Knowledge Graphs 计算机科学, 2023, 50(4): 172-180. https://doi.org/10.11896/jsjkx.220500135

演化循环神经网络研究综述

Survey on Evolutionary Recurrent Neural Networks 计算机科学, 2023, 50(3): 254-265. https://doi.org/10.11896/jsjkx.220600007

基于机器学习的剩余使用寿命预测实证研究

Empirical Research on Remaining Useful Life Prediction Based on Machine Learning 计算机科学, 2022, 49(11A): 211100285-9. https://doi.org/10.11896/jsjkx.211100285



邻域双向聚合与全局感知的 TKG 链接预测模型

唐绍赛 申德荣 寂 月 聂铁铮

东北大学计算机科学与工程学院 沈阳 110167 (1533026814@gg.com)

时序知识图谱(Temporal Knowledge Graph, TKG)在推荐系统、搜索引擎和自然语言处理等领域有着广泛的应用前 摘要 景,然而其不完备性限制了它的应用,因此研究面向 TKG 的链接预测模型具有重要作用。针对已有的工作大多面向 TKG 补 全,无法预测未来的事实,提出了一种邻域双向聚合与全局感知的 TKG 链接预测模型。一方面,分别聚合实体的主动和被动 行为并通过循环神经网络建模其历时演变来捕捉实体的短期行为;另一方面,基于全局感知模块来捕捉实体的长期行为。在 4 个基准数据集上进行了测试,结果表明所提模型能够提升模型预测未来事实的性能。 关键词:时序知识图谱;链接预测;循环神经网络 **中图法分类号** TP391

Link Prediction Model on Temporal Knowledge Graph Based on Bidirectionally Aggregating **Neighborhoods and Global Aware**

TANG Shaosai, SHEN Derong, KOU Yue and NIE Tiezheng School of Computer Science and Engineering, Northeastern University, Shenyang 110167, China

Abstract Temporal knowledge graphs(TKG) have great potential of application in many fields, such as recommender systems, search engine and natural language processing, but the incompleteness of TKG limites its application, so it is important to study link prediction model on TKG. Most existing methods focus on TKG completion and can't predict future facts. This paper proposes a link prediction model on TKG, which is based on bidirectionally aggregating neighborhoods and global aware. On the one hand, the proposed model independently aggregates entity's recently active and positive behavior and models their temporal evolution by recurrent neural network (RNN). On the other hand, it captures the chronic behavior patterns of entities by global aware module. Experimental results on four benchmark datasets show that our proposed method can improve the performance of forecasting future facts.

Keywords Temporal knowledge graph, Link prediction, Recurrent neural network

引言 1

现今许多大型应用,如搜索引擎[1]、推荐系统[2]和自然语 言处理^[3]等,都以知识图谱(Knowledge Graph,KG)为底层知 识库,而链接预测是完善知识图谱的关键技术。已有的链接 预测方法大多假定事实是不变的[4],无法处理时序知识图谱 (TKG)。然而面向 TKG 的链接预测在很多应用中具有重要 作用,如疫情防控。

以图 1 为例说明面向 TKG 的链接预测的重要性。图 1 给出了某 TKG 的一张子图,t3表示未来时刻,其中事件,如四 元组(covid, infect, A, t_3),可以从四元组(covid, infect, B, t_1) $和(A, meet, B, t_2)$ 推出,即通过链接预测,我们得知了 A 在 t_3 时刻有可能感染新冠。对于疫情防控来说,这能让我们提前







目前,面向 TKG 的链接预测主要有两种设置,即内推和 外推^[5]。给定从 t₀到 t_T的时序知识图谱,内推旨在预测已知 时间段内的事件;而外推旨在预测 $t > t_T$ 的未来事实。此时图 结构未知,显然,外推问题更具有挑战性。

通信作者:申德荣(shendr@mail.neu.edu.cn)

到稿日期:2022-09-07 返修日期:2022-12-26

基金项目:国家自然科学基金(62172082,62072084,62072086);中央高校基本科研业务费(N2116008)

This work was supported by the National Natural Science Foundation of China (62172082, 62072084, 62072086) and Fundamental Research Funds for the Central Universities(N2116008).

研究早期,一些面向 TKG 的链接预测模型侧重于分别 计算每个快照的潜在表示^[6],而忽略了实体间的长期依赖和 丰富的语义交互信息,如 TTranse^[7]和 Hyte^[8]。近期,研究 人员提出了一些针对外推问题的解决方法。例如,RE-Net^[9] 使用基于 RGCN^[10]的聚合器来编码历史快照,但并没有区分 出入边;CyGNet^[6]首次引入了自然语言处理中的复制机制来 建模历史上规律性重复发生的事件,但没有考虑实体间丰富 的语义交互信息,制约了预测的准确性。

针对已有工作的不足,本文为处理外推问题提出了一种 邻域双向聚合与全局感知的链接预测模型。一方面,它通过 一个行为感知的聚合器来编码实体近期的主动和被动行为 (出入边),在时间线上对这两种行为分别建模来捕获其行为 模式的演变;另一方面,通过全局感知模块来捕捉事件在长期 时间段内规律性的重复行为。最后综合考虑二者以预测 结果。

本文的主要贡献如下:

(1)综合考虑了近期和长期的历史信息,提出了一种邻域 双向聚合与全局感知的 TKG 链接预测模型。

(2)提出了一种行为感知的聚合器,有效建模了主被动 行为间的不同作用机制。

(3)考虑了主被动行为在时间上的不同演化模式。

(4)在4个公共数据集上进行了实验测试,验证了所提模型能够有效提高预测精度。

2 相关工作

面向知识图谱的链接预测方法^[11]可被归纳为基于规则 的方法、基于表示学习的方法、基于神经网络的方法和混合方 法。我们的研究主要侧重于基于表示学习的链接预测方法, 因此本节主要介绍有关表示学习的相关研究。

2.1 静态知识图谱表示学习

静态知识图谱表示学习分为平移模型、语义匹配模型和 神经网络模型^[12]。

最早提出的平移模型是 TransE^[13]模型,其将实体和关 系视为同一连续空间中的向量,通过训练使尾实体的向量靠 近头实体和关系的向量之和。TransH^[14]模型在 TransE 模 型基础上将实体嵌入投影到特定于关系的超平面上;TransR^[15]则将嵌入在实体空间和关系空间中进行转换。TransE 模型和 TransH 模型实施简单、易扩展,但表达能力不足; TransR 表达能力强,但随着关系的增加会出现参数过多的 问题。

语义匹配模型的核心思想是将实体的隐藏语义和表示空间中已有的关系匹配,例如 DistMult^[16]和 ComplEx^[17],其中前者将关系矩阵限制为对角矩阵,后者在复数空间中训练模型。

神经网络模型 NTN^[18]通过三阶张量来表示关系,其表达能力强但相比向量计算复杂度更高;ER-MLP^[19]在其基础上进行了简化,同时保持了模型的性能。

以上静态方法不能处理 TKG,因为其要求实体和关系是 不变的。

2.2 时序知识图谱表示学习

目前,时序知识图谱表示学习并没有通用的处理方式。 研究早期,一些方法侧重于分别计算每个快照的潜在表示,忽 略了捕捉连续时间上实体间的长期依赖和丰富的语义交互信 息。例如,TTransE^[7]在评分函数中嵌入时间信息,扩展了 TransE;Hyte^[8]将实体和关系嵌入投影到特定于时间的超平 面上;DE-SimplE^[20]建立了一个历时的实体嵌入函数来表示 不同时间戳上的实体;TA-DistMult^[21]利用循环神经网络学 习关系的时间感知表示,结合因子分解方法预测时间事实。

一些方法使用时间点过程来建模 TKG 上事件的演变,如 Know-Evolve^[22],其特点是能够预测事件发生的具体时间。

最近提出的一些方法试图建模变化的图结构,例如 EvolveGCN^[23]利用循环神经网络获得未来时间的GCN参数 来捕获图序列的动态性;Re-Net^[9]通过RGCN^[10]聚合器更新 实体的嵌入表示,并建模其时间演变模式,但预测结果只依赖 于前 m 张子图,不考虑长期信息。还有一些方法从TKG的 特点出发提出相应的模型,例如CyGNet^[6]提供了一种复制 生成机制,利用历史事件进行预测,但在预测时没有考虑实体 间的语义交互信息。

本文提出的邻域双向聚合与全局感知的 TKG 链接预测 模型与上述方法的不同之处在于:1)定义了一种新的邻域聚 合器,考虑了出入边的不同作用;2)考虑了实体出入边的聚合 嵌入在时间上的不同演变模式;3)综合考虑了实体的近期行 为和长期行为,获取的信息更全面。

3 问题定义

本节主要介绍本文的一些定义和变量。

定义 1(时序知识图谱) 令 \forall , \Re 和 \mathcal{T} 分别表示实体、关 系和时间戳的有限集,事件使用四元组(*s*,*r*,*o*,*t*)表示,其中 *s*,*o* \in \forall , *r* \in \Re , *t* \in \mathcal{T} 。时序知识图谱则表示为一系列静态快照 的集合,*G* = {*G*₁, *G*₂, ..., *G*_t, ...},其中*G*_t = { \forall , \Re , ε _t }, ε _t 是时刻 *t* 的事件集合。

定义 2(时序知识图谱链接预测) 令 \mathcal{F} 表示基准事件 集,(s_q , r_q ,o, t_q) $\in \mathcal{F}$ 表示目标事件,给定缺失四元组(s_q , r_q ,?, t_q) 或(?, r_q , o_q , t_q),时序知识图谱链接预测定义为:在已观测 到事件集 \mathcal{O} ={(s_i , r_k , o_j , t_l) $\in \mathcal{F}$ | $t_l < t_q$ }的条件下,预测缺失 位置的尾(头)实体。具体地,我们将集合 V 中的所有实体视 为候选实体,将其与缺失四元组组合,计算出它们组成真实四 元组的可能性,并由此进行排序。

4 邻域双向聚合与全局感知的 TKG 链接预测模型

本节将介绍本文提出的模型——邻域双向聚合与全局感知的 TKG 链接预测模型。它由两部分组成,分别是邻域双向聚合模块和全局感知模块。前者用于捕捉实体间在近期时间线上复杂的交互语义,具体表现为,给定缺失四元组(*sq*, *rq*,?,*tq*),在距离*tq*最近的前*m* 个子图上利用出入边更新实体 *sq*的嵌入,再将出入边的聚合嵌入分别通过 RNN 来获得实体 当前时刻的嵌入表示;后者借鉴 CyGNet^[6]通过实体历史行 为得到历史向量,最后将这两个嵌入通过多层感知机获取候 选实体的概率分布。模型的整体架构如图 2 所示。



图 2 本文模型整体架构图

Fig. 2 Overall structure of the proposed model

4.1 邻域双向聚合模块

本节介绍了我们定义的邻域聚合器 DGAT(Directed Graph Attention Network),并阐明了它的作用机制,比较了 它和其他邻域聚合方法的差异,然后介绍了循环事件编码模 块。前者用于建模同一子图上实体间复杂的交互模式,后者 利用前者的输出,在时间轴上建模事件的演变。

4.1.1 DGAT 聚合器

聚合方法大多将有向图简单扩展为无向图来聚合出入 边,即实体的主动和被动行为,本文方法将出入边视为不同的 模式单独聚合,给定(*h*,*r*,?,*t*),通过计算事件的权重来更新 实体*h*的嵌入,具体如式(1)一式(4)所示:

$$\boldsymbol{m}_{ij} = \boldsymbol{W}_{\text{dire}} \left(\boldsymbol{h}_j + \boldsymbol{r}_{ij} \right) \tag{1}$$

 $e_{ij} = (\boldsymbol{W}_{Q}\boldsymbol{h}_{i})^{\mathrm{T}} * (\boldsymbol{W}_{K}\boldsymbol{m}_{ij})$ ⁽²⁾

$$\alpha_{ij} = \operatorname{softmax}_{j} (e_{ij}) \tag{3}$$

$$\boldsymbol{h}_{i}' = \sigma(\sum_{i \in \mathcal{N}} \alpha_{ij} \boldsymbol{m}_{ij} + \boldsymbol{W}_{\text{loop}} \boldsymbol{h}_{i})$$

$$\tag{4}$$

其中, $W_{dire} \in R^{d \times d}$ 为方向矩阵, W_Q , $W_K \in R^{d \times d}$ 为注意力参数 矩阵, $W_{loop} \in R^{d \times d}$ 为自环矩阵, σ 表示激活函数,本文选择 LeakyReLU。本文对出入边分开处理,以入边聚集为例说明 具体的计算方式,此时 N_i 表示入边集合,将 h_i 的入边邻居 h_j 和 关系 r_{ij} 相加,乘上方向矩阵 $W_{dire} = W_{in}$ 得到信息量 m_{ij} ,通过点 积计算边的注意力权重,最后加上节点自环的表示即为最终 的节点嵌入,记为 h_i^n ;对于出边,方向矩阵取 $W_{dire} = W_{out}$,以相 同的方式获得出边的节点表示 h_i^{out} 。通过以上步骤,可以得到 实体 h 在 t 时刻对出边和入边的聚合表示,记为 h_i^{out} 和 h_i^n ,这 两者将作为下文编码器的输入。

4.1.2 循环事件编码器

本节将介绍本文的编码器结构。给定缺失四元组(h, r,?,t_T),由于t_T时刻的图结构未知,因此实体当前可能采取 的行为主要取决于其近期时间段内的交互。我们使用上文 提到的 DGAT 聚合器,在 $t < t_T$ 的最近的 m 个子图上对 h 进行聚合操作,得到出入边相应的聚合嵌入,如下所示:

$$egin{array}{l} \{oldsymbol{h}_{T-m}^{ ext{in}}\,,oldsymbol{h}_{T-m+1}^{ ext{in}}\,,\cdots,oldsymbol{h}_{T-1}^{ ext{in}}\,ight\}$$

 $\{\boldsymbol{h}_{T-m}^{\text{out}}, \boldsymbol{h}_{T-m+1}^{\text{out}}, \cdots, \boldsymbol{h}_{t_{T-1}}^{\text{out}}\}$

鉴于主动行为和被动行为的作用机制不同,我们使用两 个独立的 RNN 建模出入边嵌入的时间演化关系,最后将它 们相加,获得给定实体当前时刻的隐藏状态。

$$\boldsymbol{H}_{t}^{\text{out}} = RNN_{1} \left(\boldsymbol{H}_{t-1}^{\text{out}}, \boldsymbol{h}_{t-1}^{\text{out}} \right)$$

$$(5)$$

$$\boldsymbol{H}_{t}^{\text{in}} = RNN_{2} \left(\boldsymbol{H}_{t-1}^{\text{in}}, \boldsymbol{h}_{t-1}^{\text{in}} \right)$$

$$\tag{6}$$

$$\mathbf{H}_{t} = \frac{1}{2} (\mathbf{H}_{t}^{\text{in}} + \mathbf{H}_{t}^{\text{out}}) \tag{7}$$

通过以上步骤获得给定实体当前时刻的最终状态后,将 其通过 MLP 获得候选实体的概率分布,记为行为得分*p*_B,如 式(8)和式(9)所示:

$$\boldsymbol{p}_{B} = \operatorname{softmax}(\boldsymbol{W}_{B}(\boldsymbol{H}_{t}:\boldsymbol{r}:\boldsymbol{t}_{T}))$$
(8)

$$\boldsymbol{t}_T = \boldsymbol{t}_{T-1} + \boldsymbol{t}_u \tag{9}$$

其中, $W_B \in R^{3d \times N}$ 为可训练的参数; t_u 表示基本的时间单元,且 $t_0 = t_u$ 。

4.2 全局感知模块

基于许多事实在历史上重复出现的观察^[6],我们通过实体的历史向量来捕捉实体长期的规律性行为模式,即实体的 全局模式,历史得分p₄的计算式如下:

$$\boldsymbol{p}_h = \operatorname{softmax}(\operatorname{tanh}(\boldsymbol{W}_h(\boldsymbol{h};\boldsymbol{r};\boldsymbol{t}_T)) + \boldsymbol{S}_{\boldsymbol{t}_T})$$
(10)

$$\mathbf{S}_{t_{T}} = f(\mathbf{s}_{1}^{(h,r)} + \mathbf{s}_{2}^{(h,r)} + \dots + \mathbf{s}_{t_{T-1}}^{(h,r)})$$
(11)

其中, S_{t_r} 表示实体 h 的历史向量, $s_t^{(h,r)}$ 表示长度为 [\mathscr{V}] 的 multi hot 向量,每一个索引位对应一个实体。给定查询(h,r,?,t), 如果 t 时刻快照中含有四元组(h,r,o_i,t),则将 $s_t^{(h,r)}$ 中索引位 i 对应的数置为 1;否则为 0。函数 f 的作用为将所有非 0 值 置 1,否则赋一个很小的负数。 具体计算方式如下:首先将给定的实体、关系和时间戳的 嵌入连接,将其和参数矩阵 W_h 相乘再通过 tanh 激活函数得到 索引向量 v,然后加上历史向量 S_{t_7} 得到v',此时在向量v'中,在 时刻 t_7 前没有和头实体 h 发生过关系 r 的尾实体 o_i ,其对应索 引位 i 的值就被置为一个很小的负数,最后对v'进行 softmax 操作得到概率分布。此时那些没有和头实体 h 发生过关系 r 的尾实体的概率是一个极小值,即全局感知模块只关注曾经 发生过的事件再次发生的概率。

4.3 链接预测

预测四元组($h, r, ?, t_T$)的尾实体时,邻域双向聚合模块 根据实体 h 近期的交互预测可能的尾实体,而全局感知模块 推断重复出现的事件的概率分布,这两者之间的权重通过超 参数 α 来调整,如式(12)所示:

 $p(o|s,r,t_T) = ap_B + (1-a)p$ (12) 其中,p(o|s,r,t_T)表示缺失四元组(h,r,?,t_T)尾实体的概率 分布。

4.4 损失函数

本文将链接预测视为多分类任务,每个类对应一个实体, 通过最小化训练期间所有事件的交叉熵损失来训练模型,损 失函数如式(13)所示:

$$\mathscr{L} = \sum_{i \in T} \sum_{i \in s} o_{ii} \ln p(y_i \mid s, r, t)$$
(13)

其中, o_{it} 表示快照 G_i 下的第i个事件中的尾实体的 one hot 向 量, $p(y_i | s, r, t)$ 表示模型对第i个事件的预测结果。算法 1 具体说明了本文方法是如何处理外推问题的。

算法1 链接预测算法

输入: t_0 至 t_T 的 TKG 快照序列 $\{G_0, G_1, \dots, G_T\}$,查询集合 $\{(h_q, r_q, ?, t_{T+1})\}$,聚合步长 m,权重系数 α

输出:查询集合 $\{(h_q, r_q, ?, t_{T+1})\}$ 的候选实体概率分布

- 1. for query in $\{(h_q,r_q,?,t_{T+1})\}$
- 2. for t in { t_0 , ..., t_T }
- 3. $S_t = S_{t+1} + s_t^{(h_q, r_q)} / / 计算h_q$ 的历史向量,见式(11)
- 4. end for
- 5. for t in $\{t_{T-m}, \dots, t_T\}$
- 6. \mathbf{h}_{t}^{out} , \mathbf{h}_{t}^{out} = DGAT($\mathbf{h}_{q}|\mathbf{G}_{t}$)//更新 \mathbf{h}_{q} 的出人边嵌入,见式(4)
- 7. end for

8.
$$\mathbf{H}_{T+1} = \frac{1}{2} (\text{RNN}_1 (\mathbf{h}_{T-m}^{\text{out}}, \cdots, \mathbf{h}_T^{\text{out}}) + \text{RNN}_2 (\mathbf{h}_{T-m}^{\text{in}}, \cdots, \mathbf{h}_T^{\text{in}})) / /$$

出入边独立建模

- 0_i←p(o|h_q,r_q,t_{T+1})//针对当前 query 预测,见式(8)、式(10)和 式(12)
- 10. end for
- 11. return 结果集

给定查询集 $\{(h_q, r_q, ?, t_{T+1})\}$,本文采用的链接预测算法 如算法1所示。首先输入已知的TKG快照序列、步长和权重 系数;第2-4行计算 t_{T+1} 时刻头实体 h_q 的历史向量;第5-7 行使用DGAT聚合器在 $[t_{T-m}, t_T]$ 的子图上对头实体 h_q 的出 入边分别聚集;第8行获取实体的当前状态;第9行结合邻域 双向聚合模块和全局感知模块计算候选实体的概率分布;最 后返回预测的结果集。

通过以上算法,我们可以从已知的 $t_0 \sim t_T$ 的 TKG 预测得 到时间为 t_{T+1} 的 TKG 的事件,而且本文方法也能以相同的方 式进行跨时间步的预测,只需已知查询时间最近的 m 张 TKG子图即可。例如,给定查询(h_q , r_q ,?, t_{T+3}),利用邻域双 向聚合模块得到实体当前的状态 H_i ,此时时间嵌入为 $t_{T+3} = t_T + 3 \times t_u$,将 H_i , r_q 和 t_{T+3} 连接,输入多层感知机中即可进行 预测。在测试时,我们将已知的图序列设置为训练集。

5 实验结果与分析

本节在 4 个数据集上进行了链接预测实验。首先介绍实 验设置,包括数据集细节、基线选择、评价指标和超参设置;然 后对比实验结果;最后进行了消融实验。

5.1 数据集

我们在4个广泛使用的公开数据集 ICEWS18^[24], ICEWS05-15^[24],WIKI^[24]和 YAGO^[24]上评估了本文方法在 链接预测上的表现。遵循已有的工作,我们将每个数据集按 时间顺序分为训练集、验证集和测试集,其比例分别为 80%, 10%,10%。其中,ICEWS18和 ICEWS05-15 是事件知识库 综合危机预警系统(ICEWS)的子集,用于训练的事件数量分 别为 373018和 369104,事件间的时间间隔为 24 h;WIKI 和 YAGO 两个数据集的原始数据是以(*s*,*r*,*o*,[*t*_{begin},*t*_{end}])的形 式保存的,我们遵照 CyGNet^[6]的处理方法,将其离散为时 间为 1 年的快照,用于训练的事件数量分别为 539 286 和 161 540。数据集的统计数据如表 1 所列。

表 1	数据集统计
Table 1	Dataset statistics

数据集	实体数	关系数	训练集	验证集	测试集	时间间隔	时间戳数量
ICEWS18	23033	256	373018	45995	49545	24 h	304
ICEWS05-15	10488	251	369104	46188	46037	24 h	4017
WIKI	12554	24	539286	67538	63110	1 year	232
YAGO	10623	10	161540	19523	20026	1 year	189

5.2 实验设置

(1)基线选择。我们分别对比了一些静态和动态知识图 谱链接预测的方法。

静态方法:TransE^[13],DistMult^[16]和 ComplEX^[17]。静态 方法和本文的研究并不相关,此处不再赘述。

动态方法:TTransE^[7]在评分函数中简单潜入了时间信息;DE-SimplE^[20]通过实体的时间函数来表达变化的实体;

TA-DistMult^[21]利用循环神经网络学习关系的时间表示;RE-Net^[9]和 CyGNet^[6]是最近提出的方法,其中 RE-Net 通过计 算事件的联合条件概率预测时间事实,CyGNet 提出了一种 复制生成机制来建模历史事件,两者均使用无时间感知的过 滤方法,CyGNet 表现优于 RE-Net,Ding 等^[24]解释了这种过 滤方法的不足。本文使用 Ding 等使用的时间感知的过滤 方法。 (2)评价指标。本文报告了我们的方法在这 4 个数据集上的平均倒数排名(Mean Reciprocal Rank, MRR)和命中率 HIT@1/3/10。对缺失四元组进行预测并将结果降序排列, MRR表示正确实体排名倒数的平均值, HIT@1/3/10表示 排名在前 1/3/10内的正确实体的比例,具体计算方法如下 所示:

$$HIT@k = \frac{\sum\limits_{i \in T_{test}} (rank_i < X? \ 1:0)}{T_{test}}$$
(14)

$$MRR = \frac{1}{T_{\text{test}}} \sum_{i \in T_{\text{test}}} \frac{1}{rank_i}$$
(15)

其中,T_{test}表示测试集的四元组个数,rank_i表示打分后的排序 名次。

(3)超参配置。本文根据每个验证集上的 MRR 值来调整α的值,具体为:系数α以 0.1 为步长从 0.1 调整到 0.9。

对 ICEWS18, YAGO 和 ICEWS05-15, a = 0.3, 对 WIKI, a = 0.5; 邻域双向聚合模块的快照序列长度 m = 10; 批次大小设置为 1024, 训练周期设置为 20 轮, 将嵌入维度设置为 200, 当 MRR 连续两轮没有增长时结束训练。

5.3 实验结果

表 2 和表 3 列出了本文方法(our)和基线方法在 4 个 TKG数据集上的链接预测结果,各项指标均乘以 100,其中 TransE,DistMult 和 ComplEx 来自 Han 等^[25]的实验结果,表 中没有提供 WIKI数据集的结果。粗体表示最佳结果,下划 线表示次优结果。其中 RE-Net,CyGNet 和我们的工作最相 关。可以看到,在 4 个数据集上,本文方法大多都达到了最优 效果,其中静态方法的表现大多远远落后于 RE-Net,CyGNet 和本文方法,因为静态方法无法捕获 TKG 在时间上的动态 变化。

表 2	在 ICEWS18 和 ICEWS05-15 上的链接预测实验结果
Table 2	Results of link prediction on ICEWS18 and ICEWS05-15

Madal	ICEWS18-filtered				ICEWS05-15-filtered			
Model —	MRR	HIT@1	HIT@3	HIT@10	MRR	HIT@1	HIT@3	HIT@10
TransE	12.24	5.84	12.81	25.10	22.55	13.05	25.61	42.05
DistMult	10.98	4.52	10.33	21.25	28.73	19.33	27.67	42.42
ComplEx	21.01	11.87	23.47	39.87	31.69	21.44	35.74	52.04
TTransE	8.22	1.64	8.13	21.09	21.24	4.74	31.29	49.55
TA-DistMult	11.20	5.34	11.86	22.85	24.16	14.57	27.42	44.07
DE-SimplE	19.02	11.23	21.45	34.52	35.46	26.23	39.23	53.87
RE-Net	27.62	18.20	30.79	45.96	39.62	29.64	44.36	59.29
CyGNet	24.63	15.70	28.04	42.31	35.84	26.23	40.34	54.75
our	27.45	18.45	31.14	45.81	40.21	31.15	44.52	58.89

表 3 在 WIKI 和 YAGO 上的链接预测实验结果

Table 3 Results of link prediction on WIKI and YAGO

M- 1-1	WIK1-filtered				YAGO-filtered			
Model	MRR	HIT@1	HIT@3	HIT@10	MRR	HIT@1	HIT@3	HIT@10
TransE	_	-	-	_	11.69	10.37	11.96	14.83
DistMult	—	_	_	—	11.98	10.20	12.31	14.93
ComplEx	—	—	_	—	12.01	10.42	12.36	14.82
TTransE	29.47	21.77	34.63	43.58	32.19	18.33	41.31	51.61
TA-DistMult	44.63	40.12	48.64	51.85	54.82	48.01	59.46	65.94
DE-SimplE	44.73	42.50	47.49	49.38	54.71	51.44	58.04	59.87
RE-Net	49.07	46.35	50.84	52.86	57.93	52.86	60.94	66.16
CyGNet	34.26	29.65	36.46	42.64	53.24	45.94	56.52	64.20
our	51.86	49.91	53.18	54.64	60.73	55.08	63.94	68.52

同时,对比 our 和次优方法 RE-Net 可以发现:在 4 个数 据集中,our 在大多数指标上都达到了最佳效果。其中,在 ICEWS18 和 ICEWS05-15 这两个数据集中,our 的 HIT@1 和 HIT@3 略有提升,但在 WIKI和 YAGO中,our 在所有指 标上都达到了最佳,其中 MRR 分别提升了 6%和 5%,这是 因为 WIKI和 YAGO中的事件更具有时间上的规律性,而 our 在捕获实体间的语义交互时区分了出入边,因此能够获 取更精确的信息,所以本文模型在这两个数据集上的提升 更大。

与 CyGNet 相比,our 在所有指标上均达到最佳效果,其 中在 HIT@1上的提升最大,在 ICEWS18,ICEWS05-15,WI-KI和 YAGO4 个数据集上分别提升了 17%,19%,68%和 20%。因为 HIT@1 表示模型给出的最可能的答案就是真实 答案的概率,而 our 能够更精细地建模实体间的主动和被动 行为,同时实体间的交互影响着实体未来的行为,从而能更精确地预测未来事实,因此 our 大大提高了预测的准确性。

5.4 消融实验

5.4.1 模块消融实验

为了解本文方法的不同模块对性能的影响,我们进行了 模块消融实验。表3列出了模块消融实验结果,其中 only-gl 表示只使用全局感知模块,only-bi表示只使用邻域双向聚合 模块,our表示综合考虑二者。可以看到,两个模块对实验结 果都有影响,其中邻域双向聚合模块的影响较大。当仅使用 邻域双向聚合模块而放弃全局感知模块时,MRR下降了 4%,其他指标也有着小幅度的下降,这表明实体确实有着长 期的规律性行为,建模这种全局的行为模式可以提升模型的 性能;放弃邻域双向聚合模块时,可以看到模型的表现急剧下 降,MRR下降了 21%,这表明近期事件对实体的影响比长期 历史事件更为深刻,关注实体之间最近的行为可以帮助模型 更好地建模实体的状态,从而提高预测的准确性。

表 4	模块消融实验结果

Table 4 Results of module's ablation study

	ICEWS18-filtered						
	MRR	HIT@1	HIT@3	HIT@10			
only-gl	21.64	13.60	24.92	36.39			
only-bi	26.33	17.39	29.70	43.78			
our	27.45	18.45	31.14	45.81			

5.4.2 参数敏感度分析

为了了解模型对超参的敏感性,我们对 DGAT 的层数 (DGAT Layers)、邻域双向聚合模块中快照序列长度(Sequence Length)和嵌入维数(Embedding Size)进行了敏感度 分析,实验结果为不同参数下相应指标和最佳指标的比值,如 图 3 所示。可以发现,在嵌入维数为 200 时,本文方法达到了 最优效果,减小嵌入维数会使得表达能力不足,而增大嵌入维 数则会导致过拟合的问题;在快照序列长度上,本文模型在序 列长度为 10 时表现最佳,在序列长度大于 10 时,MRR 先减 小后增大,但仍小于序列长度为 10 的结果,这说明序列长度 过长会引入过多的噪声,从而导致模型性能变差;在 DGAT 层数方面,本文模型在聚合层数为 1 时取得了最佳效果,增加 层数则会导致过拟合,从而使模型性能迅速下降。



Fig. 3 Experimental results of parameter sensitivity

结束语本文通过对实体的出入边独立建模捕获了实体 在被动和主动行为中的不同作用机制,并进一步结合实体的 全局行为模式来预测未来可能发生的事件。实验表明,本文 方法在预测未来事实上有明显的提升。在未来的工作中,如 何更好地表征主动和被动行为以及如何更好地将二者相结合 以更新实体的嵌入是需要探索的方向。

参考文献

- [1] LIU Z, XIONG C, SUN M, et al. Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval[J]. arXiv:1805.07591,2018.
- [2] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems [J]. Computer, 2009, 42(8): 30-37.
- [3] CHEN J C,CHEN J S,YU Z. Incorporating structured commonsense knowledge in story completion [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019,: 6244-6251.
- [4] WANG M, WANG H F, LI B H, et al. Survey of Key Technologies of New Generation Knowledge Graph [J/OL]. Journal of Computer Research and Development: 1-18. [2022-05-21].

http://kns.cnki.net/kcms/detail/11.1777.TP20220301.1217. 002.html.

- [5] PARK N,LIU F,MEHTA P,et al. EvoKG: Jointly Modeling Event Time and Network Structure for Reasoning over Temporal Knowledge Graphs[C] // Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022: 794-803.
- [6] ZHU C, CHEN M, FAN C, et al. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2021:6244-6251.
- [7] LEBLAY J.CHEKOL M W. Deriving validity time in knowledge graph[C] // Companion Proceedings of the Web Conference 2018. ACM, 2018; 1771-1776.
- [8] 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.
- [9] 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, 2020;6669-6683.
- [10] SCHLICHTKRULL M,KIPF T N,BLOEM P, et al. Modeling relational data with graph convolutional networks [C] // European Semantic Web Conference. 2018:593-607.
- [11] GUAN S P.JIN X L.JIA Y T.et al. Knowledge reasoning over knowledge graph: A survey [J]. Journal of Software, 2018, 29(10):2966-2994.
- [12] YANG D W,ZHOU G,LU J C,et al. Review of knowledge graph completion based on knowledge representation learning [J]. Journal of Information Engineering University, 2021, 22(5):558-565.
- [13] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C] // Neural Information Processing Systems(NIPS). New York: Curran Associates, 2013:2787-2795.
- [14] 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.
- [15] 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; 21881-2187.
- [16] YANG B, YIH W, HE X, et al. Embedding Entities and Relations for Learning and Inference in Knowledge Bases[C] // International Conference on Learning Representations. 2015:202-206.
- [17] TROUILLON T, WELBL J, RIEDEL S, et al. Complex Embeddings for Simple Link Prediction [C] // Proceedings of the 33nd International Conference on Machine Learning. 2016: 2071-2080.
- [18] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion [C] // 27th Annual Conference on Neural Information Processing Sys-

tems 2013.2013:926-934.

- [19] DONG X,GABRILOVICH E,HEITZ G,et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion [C] // The 20th ACM SIGKDD International Conference on knowledge Discovery and Data Miming, 2014;601-610.
- [20] 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.
- [21] GARCÍA-DURÁN A, DUMANČIĆ S, NIEPERT M. Learning sequence encoders for temporal knowledge graph completion [J]. arXiv:1809.03202,2018.
- [22] TRIVEDI R,FARAJTABAR M,WANG Y,et al. Know-Evolve: Deep Reasoning in Temporal Knowledge Graphs[C] // Proceedings of the 33nd International Conference on Machine Learning. 2017:3462-3471.
- [23] PAREJA A, DOMENICONI G, CHEN J, et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(4): 5363-5370.
- [24] DING Z, HAN Z, MA Y, et al. Learning neural ordinary equa-

tions for forecasting future links on temporal knowledge graphs [C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021:8352-8364.

[25] HAN Z,CHEN P,MA Y, et al. Explainable subgraph reasoning for forecasting on temporal knowledge graphs [C] // International Conference on Learning Representations. 2021.



TANG Shaosai, born in 1998, postgraduate. His main research interest is link prediction on temporal knowledge graph.



SHEN Derong, born in 1964, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include distributed database and web data management.

(责任编辑:杨雪敏)