

基于 CNN 和多头注意力关系嵌入的时序知识图谱补全

陈 美¹ 许运生¹
CHEN Mei XU Yunsheng

摘 要

时序知识图谱对大多人工智能应用至关重要,但它们通常存在不完整性。为解决知识或事件的真实性会随时间发生动态变化的问题以及更好地捕捉实体和关系动态演化的特性,提出基于 CNN 和多头注意力关系嵌入的时序知识图谱补全方法(CNN and multi-head attention embedding, CMAE),先通过 CNN 提取关系的复杂时间特征,再通过多头注意力机制并行处理多个关系子空间的语义信息,并结合了 TransE、DisMult 和 SimpIE3 种现有的评分函数模型来完成对实体的预测。通过在两个公开的 ICEWS 14 和 ICEWS 05-15 数据集上进行实验,结果表明,提出的 CMAE 方法不仅能与现有的评分函数模型相结合,还能相应地提升它们的性能。同时,与 DE-SimpIE、HiSMatch 和 TTransE 等方法进行实验对比,提出的 CMAE 方法在 MRR 和 Hit@N 值上表现较优,能够实现更准确的时序知识图谱补全任务。

关键词

知识图谱补全;时序知识图谱;CNN;多头注意

doi: 10.3969/j.issn.1672-9528.2024.12.039

0 引言

目前,知识图谱补全(knowledge graph completion, KGC)方法大多只适用于静态知识图谱。然而,在现实世界中,知识或事件的真实性会随时间发生动态变化。因此,知识图谱通过引入时序信息可以更好地描述知识或事件本身的性质,将其称为时序知识图谱。在实际应用中,时序知识图谱所包含的信息通常是稀疏且复杂的,为了更好地契合真实场景,需要不断补充数据以保证其质量和规模,这就使得时序知识图谱补全(temporal knowledge graph completion, TKGC)迅速成为近几年的研究热点^[1]。

在 TKGC 中,尽管已有方法通过引入时间信息取得了一定的进展,但仍存在一些挑战。例如,现有模型可能未充分考虑关系特征的重要性。此外,尽管注意力机制在 KGC 中得到了广泛应用,但这些模型主要关注实体特征,而忽略了实体间关系随时间变化的时间信息。这些限制将会影响模型在动态场景下的性能和预测准确性。为了解决上述问题,提出一种基于 CNN 和多头注意力关系嵌入模型(CNN and multi-head attention embedding, CMAE)。通过这一方法,能够在预测给定时间内提供实体特征的情况下,同时关注关系的语义信息,最终提高实体预测的准确性。

1 相关工作

时序知识图谱是在知识图谱的基础上,引入了时间维度,

用 $G = (E, R, T)$ 来表示时序知识图谱,其中 E 表示实体的集合, R 表示关系的集合, T 表示所有时间戳的集合^[2]。每一个事件由四元组 (s, r, o, t) 来表示,其中 $s \in E$ 和 $o \in E$ 分别表示头实体和尾实体, $r \in R$ 表示它们之间的关系, $t \in T$ 代表时间信息。

根据 TKGC 是否预测未来事件可以分为基于插值和基于外推的两种方法。基于插值的 TKGC 方法是通过分析 TKG 中的已知知识来完成缺失项,而基于外推的 TKGC 方法专注于连续的 TKG,通过从历史快照中学习实体和关系的嵌入来预测未来事件^[3]。CMAE 主要解决的问题是基于插值的 TKGC,在插值方法中,存在两个重要问题:如何有效将时间信息融入到知识图谱的演化过程中,以及如何处理时间戳以充分利用知识图谱的语义信息^[4]。

近年来,TKGC 方法受到广泛关注,并且提出了多种方法来解决这些问题。例如, Ni 等人^[5]提出的时间嵌入方法,能够将时间信息融入到知识图谱的演化过程中,并且能够处理非连续有效期间,这充分利用了时间戳的语义信息。为了解决对历史上未发生过的事件预测不够准确的问题,许智宏等人^[6]提出基于历史对比学习的时序知识图谱补全模型,引入时序结构信息嵌入模块和分开编码两类信息对比方法,使得模型能够更有效地补全实体和关系,提高模型的预测性能。此外,深度学习算法也能够处理 TKG 中的时态信息。例如, HyTE^[7]将每个时间戳与相应的超平面相关联,预测关系事实缺失时间注释的时间范围。然而,HyTE 忽略了潜在的时间属性和关系的多样性, Leblay 等人^[8]提出一种能够处理任

1. 南宁师范大学计算机与信息工程学院 广西南宁 530100

意精度的时间信息方法，将时间戳与实体和关系相关联，以探索时间演化并通过 HyTE 完成缺失项。

2 CMAE 时态知识图谱补全模型

CMAE 模型如图 1 所示，由编码器和对事件进行评分的解码器组成，其主要任务是对实体进行预测。

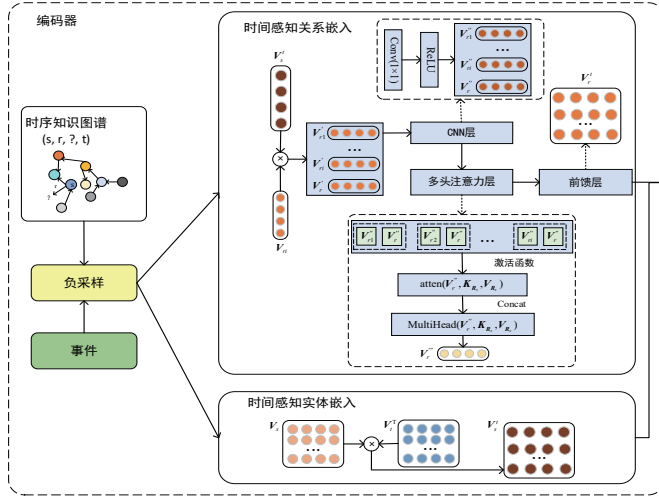


图 1 CMAE 模型示意图

以四元组 $(s, r, ?, t)$ 为例，在编码器中，将实体、关系 r 和时间戳 t 映射到嵌入空间中，并对事件进行负采样。时间感知关系嵌入由 CNN 和多头注意力嵌入方法来计算，时间感知实体嵌入由历时嵌入方法^[9]来计算。在解码器中，采用 3 种原始的评分函数来估计事件的似然性，并计算训练的损失。

2.1 编码器

(1) 将时序知识图谱中的实体、关系和时间戳映射成嵌入向量，并对事件进行负采样。在时间感知实体嵌入中，将负采样之后的实体向量 V_s 和时间戳向量 V_t 通过计算来构造时间感知实体嵌入向量 V_s' ，计算公式为：

$$V_s' = V_s V_t^T \quad (1)$$

式中： V_t^T 表示时间向量的转置。

(2) 在时间感知关系嵌入中，考虑到当前只知道头实体 s 的信息，所以将与 s 相关关系的聚类向量表示 R_s 。具体计算步骤如下：

① 将每个关系向量 $V_{r_i} \in \{r \cup R_s\}$ 映射到时间感知实体向量空间得到初始的关系向量 V_{r_i}' ，计算公式为：

$$V_{r_i}' = V_s^T V_{r_i} \quad (2)$$

② 通过将初始的关系向量 V_{r_i}' 输入到 CNN 层得到新的关系向量 V_r'' ，计算公式为：

$$V_r'' = \sigma(V_{r_i}' * W_{f_i} + b_{f_i}) \quad (3)$$

式中： $\sigma(\cdot)$ 表示 ReLU 激活函数； $*$ 表示卷积操作； W_{f_i} 表示

1×1 卷积核的权重矩阵； b_{f_i} 是卷积层的偏置项。

③ 使用多头注意力^[10]方法来计算多个关系间的注意力。首先，利用缩放点积注意力和 softmax 函数将关系注意力分数映射到 $[0, 1]$ ，计算公式为：

$$\text{atten}(V_r'', K_{R_s}, V_{R_s}) = \text{softmax}\left(\frac{V_r''^T V_{R_s}}{\sqrt{k}}\right) \quad (4)$$

式中： V_r'' 表示查询关系 r 向量； K_{R_s} 、 V_{R_s} 分别表示注意力机制的键向量和值向量； $\text{atten}(V_r'', K_{R_s}, V_{R_s})$ 表示多头注意力的关系输出向量； V_{r_i} 表示每个关系向量 $r \in R_s$ 的映射向量； $V_r''^T V_{R_s}$ 表示注意力权重的计算过程； $\sqrt{k}$ 表示查询和键的长度。

④ 通过公式 (5) 和 (6) 合并计算多头注意力的结果 V_r''' ：

$$\text{head}_i = \text{atten}(V_r'', K_{R_{s_i}}, V_{R_{s_i}}), i = 1, 2, 3, 4 \quad (5)$$

$$V_r''' = \text{MultiHead}(V_r'', K_{R_s}, V_{R_s}) \\ = \text{Concat}(\text{head}_1, \text{head}_2, \text{head}_3, \text{head}_4) \quad (6)$$

⑤ 将 V_r''' 输入到一个完全连接的前馈层 F_r 和层归一化函数 $\text{LN}(\cdot)$ 得到最终的关系编码向量 F_r' 。最后，采用加权和的方法来计算最终的关系嵌入 V_r^t ，计算公式分别为：

$$F_r = \sigma(V_r''' W_f + b_f) W_{ff} + b_{ff} \quad (7)$$

$$F_r' = \text{LN}(F_r + V_r''') \quad (8)$$

$$V_r^t = (1 - \beta) V_r + \beta F_r' \quad (9)$$

式中： $\sigma(\cdot)$ 表示激活函数； W_f 和 W_{ff} 表示两个线性变换矩阵； b_f 和 b_{ff} 表示两个线性变换的偏置向量； β 表示嵌入权重。

2.2 解码器

在训练中，将 CMAE 分别与 TransE、DisMult 和 SimpIE 模型结合，来完成实体预测任务。下面以 CMAE-TransE 模型为例，采用 DE-Trans 的方法来完成实体嵌入，计算公式为：

$$V_{e_r} = \text{DT}(V_e, t) = V_e + \sum_{i=1}^t \Delta V_{e_i} \quad (10)$$

式中： $\text{DT}(V_e, t)$ 是实体 e 的嵌入函数； V_e 是实体 e 的嵌入向量； t 是时间步的索引； ΔV_{e_i} 是实体 e 在时间步 i 的时间嵌入向量。

遵循 TransE 模型的评分函数来计算解码器中事件的可能性，如公式所示，同理 $\text{Sco}(V_s', V_o, V_r)$ 。

$$\text{Sco}(V_s, V_r, V_o') = \text{sum}((V_s^t + V_r^t - V_o') + (V_s^t + V_r^t - V_o')^{-1}) \quad (11)$$

对于 $(s, r, ?, t)$ ，生成一个候选答案集 neg ，它由训练三

元组组成，其中头部或尾部被随机实体替换，并不是同时替换。此外，采用交叉熵损失作为模型训练的损失函数，计算预测尾实体的损失函数 $L(s, r, t)$ ，如公式 (12) 所示。同理，对于预测头实体 $(?, r, o, t)$ 是同样的计算。最终的目的是尽量减少整体损失，计算公式分别为：

$$L(s, r, t) = - \sum_{G_{\text{train}}} \frac{\exp(\text{Sco}(V_s, V_o', V_r)_o)}{\sum_{i \in \text{neg}} \exp(\text{Sco}(V_s, V_o', V_r)_i)} \quad (12)$$

$$\text{Loss} = L(s, r, t) + L(o, r, t) \quad (13)$$

$$L_{\text{min}} = \min(L_{\text{min}} + \text{Loss}) \quad (14)$$

式中： G_{train} 是训练集；neg 是负采样的候选集。

3 实验结果及分析

3.1 实验环境及数据集

实验运行环境：CPU 为 Intel(R) Xeon(R) Gold 6133 CPU @ 2.5 GHz，显卡 GPU 为 NVIDIA GeForce RTX 3090，内存为 24 GB，采用编程语言为 Python。

实验采用带时间戳的政治事件数据集综合危机预警系统（integrated crisis early warning system, ICEWS）^[11] 中的 ICEWS14 和 ICEWS05-15，如表 1 所示。

3.2 评价指标

为了评估 CMAE 模型的有效性，本次实验选择平均倒数排名（mean reciprocal rank, MRR）和 Hit@N{N=1,3,10} 作为模型评价指标，MRR 值越高表明排名越符合事实情况，模型嵌入的效果越好。因此，较高的 MRR 和 Hit@N 值反映了模型对实体的预测性能越好。计算公式为：

$$\text{MRR} = \frac{1}{2 \times |\text{test}|} \sum_{\text{test}} \left(\frac{1}{f_s} + \frac{1}{f_o} \right) \quad (15)$$

$$\text{Hit@N} = \frac{1}{2 \times |\text{test}|} \sum_{\text{test}} [(\zeta f_s \leq N) + (\zeta f_o \leq N)] \quad (16)$$

式中： f_s 和 f_o 分别表示头实体 s 和尾实体 o 查询的对应排名， $\zeta(\cdot)$ 表示指标函数，当 $f_s \leq N$ 时， $\zeta(\cdot)=1$ 成立，反之则为 0。

表 1 数据集的详细统计数据

	ICEWS14	ICEWS05-15
实体数	7182	10 488
关系数	230	251
时间戳数	365	4017
训练集	72 826	386 962
验证集	8941	46 275
测试集	8963	46 092

3.3 实验设计与对比

实验参数的设置：训练次数 epoch 为 500，学习率 lr 为 0.001，负采样数量 neg_ratio 为 500，嵌入维度 emb_dim 为 100，时间维度 t_emb_dim 为 64，批大小 batch_size 为 512，关系嵌入权重 r_emb_weight 为 0.4，多头注意力的数量 n_head 为 4，dropout 概率为 0.4。

将 CMAE 与现有的 3 种评分函数结合进行实验，分别表示为：CMAE-TransE、CMAE-DistMult 和 CMAE-SimpIE，并选取与近几年的 14 个模型^[12-13] 展开比较，结果如表 2 所示。

表 2 在 ICEWS14 和 ICEWS05-15 数据集上的 TKGC 结果

模型	ICEWS14				ICEWS05-15			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
TransE*	0.280	9.4	—	63.7	0.294	9.0	—	66.3
DisMult*	0.439	32.3	—	67.2	0.456	33.7	—	69.1
TTransE	0.255	7.4	—	60.1	0.271	8.4	—	61.6
TA-DisMult	0.477	36.3	—	68.6	0.474	34.6	—	72.8
DE-SimpIE	0.526	41.8	59.2	72.5	0.513	39.2	57.8	74.8
HyTE	0.297	10.8	41.6	65.5	0.316	11.6	44.5	68.1
RE-NET	0.370	27.0	39.6	54.8	0.440	34.4	49.0	64.0
REGCN	0.415	30.8	46.6	62.4	0.464	35.1	52.7	67.6
HiSMatch	0.464	35.9	51.6	66.8	0.528	42.0	59.0	73.2
Re-Temp	0.480	37.32	53.60	68.90	0.563	45.4	62.8	77.1
CMAE-TransE	0.366	18.45	49.22	68.72	0.359	16.5	49.0	70.9
CMAE-DistMult	0.529	42.29	59.49	72.68	0.499	38.2	56.3	72.7
CMAE-SimpIE	0.549	44.4	61.6	74.2	0.558	44.7	63.9	77.4

注：“*”模型为 KGC，未做标记的模型为 TKGC

由表 2 可知，CMAE-SimpIE 在 ICEWS4 数据集上的 MRR 和 Hit@N 指标均优于基准模型，在 ICEWS05-15 数据集的 Hit@3 和 Hit@10 指标均优于基准模型，但 MRR 和 Hit@1 指标略低于 Re-Temp，其原因是 Re-Temp 在每个时间戳之后引入了跳跃信息流，这使得模型能够跳过对预测不必要的信息，从而提高推理效率。总而言之，CMAE 能够提高 TKGC 任务的性能，并证明了动态时间和复杂关系在 TKGC 任务中的重要性。下面将具体分析 CMAE -SimpIE 优于基准模型原因：

（1）从时间信息角度，TKGC 模型利用时间信息来约束事实的相似性，从而有效地分离具有不同时间戳的相似事实，因此大多数 TKGC 模型比静态 KGC 取得了更好的结果。

（2）与 TTransE、TA-DisMult 和 DE-SimpIE 相比，CMAE 与 TransE、DisMult 和 DE-SimpIE 结合的方法在两个数据集上都优于其他相应的基线，实验结果表明，该方法可以提高大多数平移模型的性能，对 TKGC 有很强的适用性。

（3）HiSMatch 只使用查询的完全相同的关系来构建查询子图，这忽略了关系之间的潜在相似性，而 CMAE 使用关

系的嵌入来计算注意力权重。

在 ICEWS14 数据集上进行了消融实验,设计了一个变体,即: CMAE_A, 其只使用多头注意力方法来学习复杂关系的特征, 实验结果如表 3 所示。

表 3 在 ICEWS14 数据集上的消融实验

ICEWS14——DES				
模型	MRR	Hit@1	Hit@3	Hit@10
CMAE	0.520	41.2	58.7	72.5
CMAE_A	0.549	44.4	61.6	74.2

实验表明, 只使用多头注意力方法的 CMAE_A 预测性能较差, 而使用 CNN 方法的 CMAE 能够提取关系的复杂时间特征, 增强模型对时序数据的关注度。

4 结语

本文提出基于 CNN 和多头注意力关系嵌入模型, 先通过 CNN 用于提取关系的重要特征, 再利用多头注意力机制学习多个关系子空间的语义信息, 证明了 CMAE 方法不仅能与现有的 TKGc 模型相结合, 还能相应地提升它们的性能。未来, 将深入探究时序知识图谱中实体与关系的时间演化特性, 构建更准确的模型来探索时态知识图推理的任务。

参考文献:

[1] WU J. Deep learning for temporal knowledge graph completion[M]. Canada: McGill University, 2021.

[2] BAI L Y, MA X N, MENG X X, et al. RoAN: a relation-oriented attention network for temporal knowledge graph completion[J]. Engineering applications of artificial intelligence, 2023, 123(8): 106308.

[3] CAI B R, XIANG Y, GAO L X, et al. Temporal knowledge graph completion: a survey[DB/OL]. (2022-01-16)[2024-04-11].<https://doi.org/10.48550/arXiv.2201.08236>.

[4] WANG J P, WANG B Y, QIU M K, et al. A survey on temporal knowledge graph completion: Taxonomy, progress, and prospects[DB/OL].(2023-08-04)[2024-07-03].<https://doi.org/10.48550/arXiv.2308.02457>.

[5] NI R Y, MA Z G, YU K H, et al. Specific time embedding for temporal knowledge graph completion[C]//2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC). Piscataway:IEEE,2020: 105-110.

[6] 许智宏, 邱鹏林, 王利琴, 等. 基于历史对比学习的时序

知识图谱补全[J]. 计算机工程与应用, 2024,60(22):154-161.

[7] 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. Brussels: ACL, 2018: 2001-2011.

[8] LEBLAY J, CHEKOL M W, LIU X. Towards temporal knowledge graph embeddings with arbitrary time precision[C]//Proceedings of the 29th ACM international Conference on Information & Knowledge Management. NewYork: ACM, 2020: 685-694.

[9] GOEL R, KAZEMI S M, BRUBAKER M, et al. Diachronic embedding for temporal knowledge graph completion[DB/OL]. (2019-07-06)[2023-0509].<https://doi.org/10.48550/arXiv.1907.03143>..

[10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[11] WARD M, BEGER A, CUTLER J, et al. Comparing GDELT and ICEWS event data[J]. Analysis, 2013, 21(1): 267-297.

[12] LI Y C, ZHANG X M, ZHANG B, et al. Each snapshot to each space: space adaptation for temporal knowledge graph completion[C]//21st International Semantic Web Conference. Berlin: Springer, 2022: 248-266.

[13] WANG K Z, HAN S C, POON J. Re-temp: relation-aware temporal representation learning for temporal knowledge graph completion[DB/OL]. (2023-10-24)[2024-07-19].<https://doi.org/10.48550/arXiv.2310.15722>.

【作者简介】

陈美 (2000—) 女, 广西陆川人, 硕士研究生, 研究方向: 语义网 Web、机器学习。

许运生 (1992—) 男, 广东惠州人, 硕士研究生, 研究方向: 计算机应用、语义网 Web。

(收稿日期: 2024-09-14)