

一种向量索引支持的时态知识图谱高效搜索方法

朱红¹, 胡新雨², 高莉莎², 张强³, 徐晓轶², 朱盟盟⁴

(1. 国网南京供电公司, 江苏南京 210009;

2. 国网南通供电公司, 江苏南通 226000;

3. 国网智能电网研究院有限公司, 北京 102200;

4. 苏州华天国科电力科技有限公司, 江苏苏州 215000)

摘要:知识图谱嵌入(Knowledge Graph Embedding, KGE)将实体和关系表示为低维、连续的向量,使机器学习模型能够轻松适应知识图谱(Knowledge Graph, KG)的搜索任务。然而,在大规模知识库(Knowledge Base, KB)的搜索密集型应用中,现有的模型大多侧重于提高在静态KG上搜索的准确性,忽略了在动态时态知识图谱(Temporal Knowledge Graph, TKG)上搜索的时间效率。为此,提出了一种向量索引支持的TKG高效搜索方法,以提高在TKG上的搜索效率。具体来说,首先,将实体、关系和时间信息映射到向量空间,并利用长短期记忆神经网络(Long Short-Term Memory, LSTM)学习关系类型的时间感知,从而建立了具有时间信息感知与关系联合编码的TKG向量库。然后,利用向量数据库建立大规模TKG的向量索引库(IndexIVFFlat)。注意,该索引通过聚类操作来划分搜索空间,以提高知识的搜索效率。最后,在拥有高效索引的TKG上通过相似度计算执行近似性搜索与实验评估。结果显示,该方法在时间效率上优于未建立索引的方法,且在搜索准确度上优于一些强相关的方法。表明,该向量索引库的建立保证了搜索准确性的前提下提高了在TKG上的搜索效率。

关键词:知识图谱嵌入;时态知识图谱;索引;搜索;向量数据库;机器学习

中图分类号: TP305

文献标识码: A

文章编号: 1673-629X(2025)02-0138-08

doi: 10.20165/j.cnki.ISSN1673-629X.2024.0305

An Efficient Search Method on Temporal Knowledge Graph Supported by Vector Indexing

ZHU Hong¹, HU Xin-yu², GAO Li-sha², ZHANG Qiang³, XU Xiao-yi², ZHU Meng-meng⁴

(1. State Grid Nanjing Power Supply Company, Nanjing 210009, China;

2. State Grid Nantong Electric Power Co., Ltd., Nantong 226000, China;

3. State Grid Smart Grid Research Institute Co., Ltd., Beijing 102200, China;

4. Suzhou Huatian Guoke Electric Power Technology Co., Ltd., Suzhou 215000, China)

Abstract: Knowledge graph embedding (KGE) represents entities and relations as low-dimensional, continuous vectors, thus enabling machine learning models to be easily adapted to knowledge graph (KG) search task. However, in search-intensive applications of large-scale knowledge base (KB), most of the existing models focus on improving the accuracy of searching on static KGs, while neglecting the time efficiency of searching on dynamic temporal knowledge graph (TKG). To this end, an efficient search method for TKG supported by vector indexing is proposed to improve the search efficiency on TKG. Specifically, firstly, the entity, relation and time information are mapped to the vector space, and the time-awareness of relation types is learned using the Long Short-Term Memory (LSTM) neural network, which leads to the establishment of the TKG vector database with the joint encoding of time information awareness and relation. Then, the vector database is utilized to build a vector indexing database (IndexIVFFlat) for large-scale TKGs. Note that the index divides the search space by clustering operations to improve the search efficiency of knowledge. Finally, approximation search and experimental evaluation are performed by similarity computation on the TKG with efficient indexing. Results show that the proposed method outperforms the unindexed method in time efficiency and outperforms some strongly correlated methods in search accuracy. It is

收稿日期: 2024-07-03

修回日期: 2024-11-06

基金项目: 国家电网公司总部科技项目(5700-202318598A-3-2-ZN)

作者简介: 朱红(1971-),女,研究员级高级工程师,硕士,国网公司首批人工智能首席专家,研究方向为知识图谱与电力应用;通信作者:朱盟盟(1989-),男,工程师,研究方向为人工智能。

demonstrated that the establishment of this vector indexing database improves the search efficiency on the TKG under the guarantee of the search accuracy.

Key words: knowledge graph embedding; temporal knowledge graph; indexing; search; vector database; machine learning

0 引言

随着信息时代的来临,知识图谱(Knowledge Graph, KG)已经成为连接信息的重要工具。KG是一种结构化的数据表示方法,它将实体和关系之间的语义信息表示为图模型,其基本结构可以形式化表征为三元组 $\langle \text{subject}, \text{predicate}/\text{relation}, \text{object} \rangle$ 或 $\langle s, p/r, o \rangle$,其中 $s, p/r, o$ 分别表示主语、谓语/关系和宾语。KG被广泛应用于搜索推荐^[1]、问答系统^[2]、自然语言处理^[3]、地铁客流量预测^[4]等领域。

尽管网络上不断涌现的知识获取系统为用户提供了大量结构化的数据源,使得用户能够通过发布图形查询来寻找他们感兴趣的项目,但与此同时,一个日益凸显的问题是空答案现象,其根本原因在于这些系统普遍存在着信息不完备性,使得部分查询无法得到满意的答案^[5]。因此,知识图谱嵌入(Knowledge Graph Embedding, KGE)^[6]应运而生,从经典的翻译模型TransE^[7]及它的变体^[8-11]模型开始,KGE逐渐成为一个非常活跃的研究领域。KGE旨在将KG中的实体和关系转化为低维、连续的向量表示,从而简化KG上的操作,并保留其固有的结构信息和语义关系。它为KG赋予了更具智能的表达能力,从而使机器学习模型能够轻松适应KG的补全和查询任务。

近年来,随着数据量的不断增加和应用场景的扩展,越来越多的研究工作开始关注动态时态知识图谱(Temporal Knowledge Graph Embedding, TKG)的建模和推理问题^[12-16]。与传统的静态KG不同,TKG中的实体、关系和事实会随时间变化而演化。这种变化可能是周期性的,也可能是突发性的,具有复杂性和多样性。例如,一个人的工作、学习、家庭等状态都会随时间变化而改变,而一个公司的股东、业务范围、市值等属性也会随时间发生变化。这种时态性质使得TKG在很多实际应用中具有重要的意义。

利用嵌入技术来回答不完整TKG上的图查询是一种很有前途的方法,但其也存在两个挑战:(1)在大规模知识库(Knowledge Based, KB)的搜索密集型应用中,当嵌入维数过高时,TKGE学习的密集向量对于相似性计算是低效的;(2)现有的KGE模型大多侧重于提高在KG上搜索的准确性,而忽略了在动态KG上搜索的时间效率^[17-18],尤其在面对具有时间敏感性的TKG上。这些缺陷阻碍了TKGE模型在现实世界KG搜索中的广泛应用。

为此,如何在保证搜索准确性的前提下,提高

TKG的搜索效率,成为了当前研究的一个重要问题。针对这一问题,提出了一种向量索引支持的TKG高效搜索方法,从而在保证查询准确性的前提下提高了搜索效率。具体来说,该方法将实体、关系和时间信息联合编码到向量空间,并利用长短期记忆神经网络(Long Short-Term Memory, LSTM)学习关系类型的时间感知,从而建立了TKG向量库;然后,利用该向量数据库建立大规模TKG的向量索引库(IndexIVFFlat)。注意,该索引库通过聚类操作来划分搜索空间,从而大大降低了搜索复杂度。最后,在拥有高效索引的TKG上通过相似度计算执行了近似性搜索与实验评估。实验结果表明,该向量索引库的建立大大提高了在TKG上的搜索效率,并且能够在保证搜索准确性的情况下,加速查询过程。这种方法可以为动态TKG的搜索和推理任务提供有力支持。主要贡献如下:

(1)提出了一种向量索引支持的TKG高效搜索方法,该方法能够在保证搜索准确性的前提下,大大提高查询效率。

(2)利用LSTM学习关系类型的时间感知,将实体、关系和时间信息联合编码到向量空间中,提高了知识表示的精度和时态性。

(3)建立了大规模TKG的向量索引库(IndexIVFFlat),通过聚类操作将搜索空间划分为多个小空间,从而达到快速搜索的目的。此外,在多个实验数据集上进行了近似搜索和实验评估,表明该方法在TKG搜索效率上具有显著优势。

1 相关工作

1.1 时态知识图谱嵌入

事实的动态演化激发了动态KGE的研究。大量的工作研究了具有时间敏感性的TKGE模型。这些模型基于不同的方法(例如,投影^[12],旋转^[13,19],几何代数^[14]和神经网络^[15-16])将动态时间扩展到静态三元组中。

上述研究可以分为三类:第一类是基于几何变换的TKGE, Dasgupta等^[12]提出了HyTE模型。时间向量被作为法向量,实体和关系通过时间向量所投影到超平面。然而,该模型不能很好地建模关系模式。为此,TeRo模型^[13]通过旋转操作将时态信息融入到实体和关系向量中。TKG被嵌入到复向量空间,以有效地捕获实体和关系的多种时态关系模式,即:对称性、

自反性等。基于旋转操作,一种上下文时序关联的 TKG 嵌入模型被提出^[19],该研究将细粒度时序信息细分为知识开始时间的相关性和知识时间间隔的一致性。第二类是基于代数的 TKGE,例如,Xu 等^[14]采用几何代数思想,提出了适用于静态和时态 KG 的嵌入模型,能够处理复杂关系,并映射到嵌入空间中。第三类是基于神经网络的 TKGE,例如,一种基于序列编码器的 TKGE 方法被提出^[15]。该方法利用序列编码器来捕获知识图谱中实体和关系的时态演化,并通过学习实体和关系的序列表示来完成 TKG 的补全任务。基于上述思想,Hou 等^[16]开发了 T-GAE 模型,利用图注意力机制捕获实体和关系之间的复杂依赖关系,并考虑了时间跨度对知识图谱的影响。

尽管这些研究为 TKGE 领域提供了强有力的技术框架,但是这些方法在大规模 TKG 上的一些下游任务上(如搜索、推理)存在效率低下的缺点。

1.2 向量索引支持的知识图谱搜索

向量索引支持的知识图谱搜索是指利用向量索引技术来提高知识图谱搜索的效率和准确性。在这种方法中,知识图谱中的实体和关系被表示为高维向量,并通过给所表征的向量建立索引结构以达到加速搜索的过程。

一些工作探究了基于哈希函数的 KG 搜索方法^[17,20-21]。这些工作都利用哈希函数从高维的知识图谱向量数据中学习紧凑的二进制代码,并通过测量汉明距离或点积来提高近似搜索效率。此外,文献[17,20]也将本体信息融入到 KG 中,以捕捉实体之间的相似性。

与之不同的是,文献[22]提出了一种基于嵌入的知识图谱搜索方法 EMBLOOKUP。EMBLOOKUP 是基于深度度量学习和三元组损失,以支持准确高效的知识图谱实体查找,从而提升了知识图谱检索的效率。随后,为了减少在知识图谱上逻辑搜索对时间的消耗,

文献[18]提出查询计算树优化(OTO)。QTO 通过在树状计算图上的前向-后向传播找到最优解,并利用编码在查询计算树中的独立性来减少搜索空间,其中,在优化期间仅涉及局部计算。

此外,文献[23]比较了六种不同的枢纽减少技术和八种不同的(近似)最近邻搜索算法以实现知识图谱的实体对齐。

然而,尽管这些研究探究了基于向量索引支持的静态知识图谱搜索方法,但是他们并没有将向量索引应用到 TKG 的近似搜索上。

综上所述,一些工作研究了 TKGE,但是其存在效率低下的问题;一些工作研究了基于索引的静态 KG 搜索方法,但是其不能对具有动态时间敏感性的 KG 执行高效的查询,而该文提出了一种索引支持的 TKG 搜索方法,可以分别弥补上述两类研究的缺陷。

2 方法

2.1 模型总述和问题定义

TKG 是传统静态知识图谱的扩展,用于表示和组织与时间相关的事实和关系。该文旨在解决基于索引支持的 TKG 的高效搜索问题。与以往的 TKG 查询不同,基于索引支持的 TKG 查询可以通过给 TKG 建立索引来提高对实体的检索效率。因此,下面给出了该方法的总述和问题定义。

模型总述:图 1 阐述了向量索引支持的 TKG 搜索模型的整体架构。具体来说:(1)该模型将实体、关系和时间信息联合编码到向量空间,并利用 LSTM 学习关系类型的时间感知;(2)利用向量数据库建立大规模 TKG 的向量索引库(IndexIVFFlat)。该索引库通过聚类操作来划分搜索空间,从而极大降低了搜索复杂度;(3)在拥有高效索引的 TKG 上通过相似度计算设计近似性搜索方法。

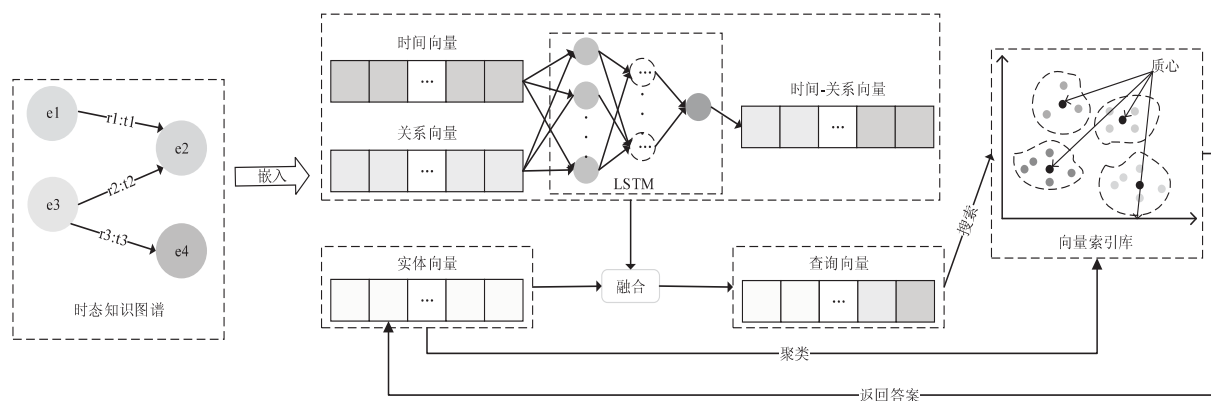


图 1 向量索引支持的 TKG 搜索模型的整体架构

问题定义:一个 TKG 被定义为 $G = \langle s, r, o, t \rangle$, 其中, $s, o \in E$ 分别表示某一个事实的主语和宾语; r

$\in R$ 表示该事实的关系,用于描述 s 和 o 的语义; $t \in T$ 表示该事实成立的有效时间; E 表示实体集合, r 表示

关系集合, T 表示时间戳集合。给定一个缺失的查询四元组 $\langle ?, r, o, t \rangle / \langle s, r, ?, t \rangle$, 其目标是通过所建立的索引以高效地找到正确的实体 $? \in E$ 。

2.2 时间感知表征

给定一个时态 KG 三元组, 即, (s, r, o, t) , 一些三元组用时间信息 t 扩充, 可以将给定的时间分解成由以下时间标记组成的序列。

$$\begin{cases} \text{year} \leftarrow 0 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \\ \text{month} \leftarrow 01 \cdot 02 \cdot 03 \cdot 04 \cdot 05 \cdot 06 \cdot 07 \cdot 08 \cdot 09 \cdot \\ \quad 10 \cdot 11 \cdot 12 \\ \text{day} \leftarrow 0 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \end{cases} \quad (1)$$

其中, year, month 和 day 分别表示给定时间的年, 月和日。时态标记的词汇量大小总共为 32。此外, 对于每个三元组, 可以提取一个关系标记序列, 该序列总是由关系类型标记和时态修饰符标记组成。此外, 图 2 描述了时态信息被融入到三元组的组成结构。

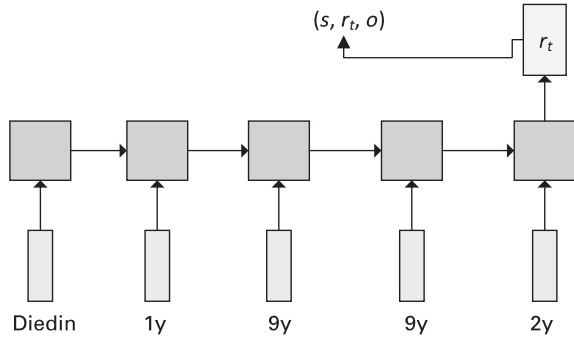


图 2 TKG 的组成结构

2.3 基于 LSTM 的时间序列编码

给定一个时态 KG 三元组, 即 (s, r, o, t) , 将关系序列和时间序列的联合称为时间-关系序列 p_t 。由此, 时态 KG 三元组可以表示为形式 (s, r_t, o) 的三元组的集合, 其中关系序列可以包括时间信息。基于上述定义的时间序列标记, 即 year, month 和 day, 其被用作 LSTM 的输入。

LSTM 是一种神经网络架构, 特别适合于对序列数据进行建模。定义 LSTM 的公式为:

$$\begin{cases} i = \sigma_g(h_{n-1}U_i + x_nW_i) \\ f = \sigma_g(h_{n-1}U_f + x_nW_f) \\ o = \sigma_g(h_{n-1}U_o + x_nW_o) \\ g = \sigma_c(h_{n-1}U_g + x_nW_g) \\ c_n = f \circ c_{n-1} + i \circ g \\ h_n = o \circ \sigma_h(c_n) \end{cases} \quad (2)$$

其中, i, f, o 和 g 分别是输入、遗忘、输出和输入调制门。 c 和 h 分别是单元和隐藏状态。所有向量均在 \mathbb{R}^h 。 $x_n \in \mathbb{R}^d$ 是序列的第 n 个元素的表示。注意, 设置 $h = d$ (输入向量的维度), σ_g, σ_c 和 σ_h 是激活函数。

输入序列 r_t 的每个标记首先通过线性层映射到其对应的 d 维嵌入, 并将所得嵌入序列用作 LSTM 的输入。每个长度为 N 的关系序列由 LSTM 的最后一个隐藏状态表示, 即 $e_n = h_N$ 。定义时间-关系向量嵌入为:

$$r_t(r, t) = \text{LSTM}(r, t) \quad (3)$$

其中, r_t 表示时间-关系向量序列。

2.4 TKGE 的打分函数和损失函数

打分函数: 携带时间信息的关系序列表示可以与经典打分函数中的主语和宾语嵌入结合使用, 打分函数被定义为:

$$\begin{cases} \text{Score}(q, e) = \|q - e\|_{L_1/L_2} \\ q = e + r_t \end{cases} \quad (4)$$

其中, 使用随机梯度下降与 LSTM 的参数共同学习该打分函数的所有参数。 L_1/L_2 表示 1/2 范数。

损失函数: 其对应的损失函数被定义为:

$$L = \sum_{\xi} \sum_{\xi'} [\gamma + \text{Score}(q, e) - \text{Score}(q, e')] \quad (5)$$

其中, $(s, r_t, o) \in \xi$ 和 (s', r_t, o') , $(s, r_t, o') \in \xi'$ 分别为时态三元组的正样本和负样本。通过打碎正样本中的主语或者宾语来生成负样本。 $\gamma > 0$ 表示间隔超参数。

2.5 索引构建与近似搜索

为了加快搜索速度, 可以将训练完备的实体向量库进行聚类, 以划分搜索空间。在 d 维向量空间中定义 Voronoi 单元格, 并且每个实体向量都落入其中一个单元格中。在搜索时, 只有查询向量所在单元中包含的实体向量与少数几个相邻查询向量进行比较, 从而建立了高效的实体向量索引库 IndexIVFFlat。

IndexIVFFlat 索引需要一个训练的过程, 即, 使用函数 train(), 可以在与实体向量具有相同分布的任何向量集合上执行。同时, 该索引还需要另一个索引 IndexFlatL2, 即量化器 (quantizer), 它将实体向量分配给 Voronoi 单元。每个单元由一个质心定义, 找到一个实体向量所在的 Voronoi 单元包括在质心集中找到该实体向量的最近邻居。

索引构建: 基于上述时态 KGE 模型输出的结果, 将实体向量集合输入到索引库中, 其实体索引被定义为:

$$\begin{cases} \text{quantizer} = \text{IndexFlat2}(d) \\ \text{index} = \text{IndexIVFFlat}(\text{quantizer}, d, \text{nlist}, L2) \\ \text{index.train}(e) \\ \text{index.add}(e) \end{cases} \quad (6)$$

其中, quantizer 表示量化器。 d 表示嵌入的实体向量的

维数。add() 函数表示将实体向量集合的 id 添加到所构建的索引库中,以使实体向量和对应的索引相对应。nlist 表示划分单元的数量(聚类中心数量)。

近似搜索:基于 TKG 训练完备并组合的包含时间信息的查询向量集合,利用 IndexIVFFlat 索引库进行近似性搜索,给定一个时态 KG 查询向量 q ,并设置要返回最相似的前 k 个实体向量,其近似搜索被定义为:

$$\begin{cases} \text{index.nprobe} = \text{nprobe} \\ \text{Distance, Index} = \text{index.search}(q, \text{top_k}) \end{cases} \quad (7)$$

其中, nprobe 表示执行搜索访问的单元格数。Distance 和 Index 分别表示返回的前 k (top_k) 个最相似向量的距离和索引。

在使用 IndexIVFFlat 索引结构时, nlist 和 nprobe 是两个关键参数,需要根据具体的应用需求来选择合适的取值。nprobe 是调整搜索效率和准确度的一种方式。当 nprobe = nlist 时,该搜索方式与暴力搜索的结果相同。注意:(1)通常可以设置 nlist 的值为数据集大小的平方根左右。对于较大的数据集,可能需要增加 nlist 的值,以提高搜索结果的准确性。针对不同数据集的大小和特征的分布情况以调优来找到最佳的取值。(2)一般来说, nprobe 被设置为 nlist 的一个较小的比例,例如, nlist/10 或者 nlist/20。对于较大的数据集或者对搜索准确性要求较高的情况,可以适当增加 nprobe 的值。注意,增加 nprobe 的值会增加搜索的计算和内存访问开销,可能导致搜索时间变长,则需要根据具体的性能要求和资源限制进行权衡。

2.6 算法总述

算法 1 描述了向量索引支持的 TKG 搜索过程。具体来说,输入为多个 TKG 的训练子图 G_1, G_2, \dots, G_n , 以及实体集 E , 关系集 R , 时间戳集 T , 超参数 H ; 输出为搜索的结果,即前 k 个最相似实体向量的距离和索引。02 行表示对输入子图的向量化。03 行表示使用 LSTM 将关系向量和时间戳向量合并。04 和 05 行分别表示计算查询向量与实体(答案)的分数和损失。06 行表示索引的建立,07 行表示基于建立索引的近似搜索。

算法 1: IndexSearch(g)

输入: $G_1 = \langle s_1, r_1, o_1, t_1 \rangle \cup G_2 = \langle s_2, r_2, o_2, t_2 \rangle \cup \dots \cup G_n = \langle s_n, r_n, o_n, t_n \rangle$; E, R, T, H

输出: top_k 个最相似实体向量的距离和索引

01 for all G_1, G_2, \dots, G_n do

02 $G_1, G_2, \dots, G_n \leftarrow \text{Embedding}(G_1, G_2, \dots, G_n)$

03 $r_i \leftarrow \text{LSTM}(r_i, t_i)$

04 $\begin{cases} \text{Score}(q, e) = \|q - e\|_{L_2/L_2} \\ q = e + r_i \end{cases}$

05 $L = \sum_{\xi} \sum_{\xi} [\gamma + \text{Score}(q, e) - \text{Score}(q, e')]$

```

06 {
    quantizer = IndexFlat2(d)
    index = IndexIVFFlat(quantizer, d, nlist, L2)
    index.train(e)
    index.add(e)
07 {
    index.nprobe = nprobe
    Distance, Index = index.search(q, top_k)
08 end for

```

3 实验与分析

为了评估该方法的性能,设计了一些实验对其进行了验证与评估。

3.1 实验设置与数据集

该实验在 Window10_64 系统上运行,该系统配备了 Intel(R) Core(TM) i7-9700 CPU @ 3.00 GHz 3.00 GHz 处理器和 32.0 GB 内存。该模型使用 PyTorch 实现。

执行细节:对于所有数据集,该模型对超参数的调优方式如下:嵌入维度 $d \in \{100, 200, 300, 400, 500\}$; 间隔 $\gamma \in \{1, 2, 3, 5, 10\}$; 学习率 $\text{lr} \in \{0.1, 0.01, 0.001\}$; 数据集的批量大小 $b \in \{1024, 512\}$; 负样本 $n \in \{100, 200, \dots, 500\}$; 迭代次数 $\text{epoch} \in \{100, \dots, 500\}$; 划分单元的数量(聚类中心的数量) $\text{nlist} \in \{84, 101, 106\}$; 执行搜索访问的单元格数量 $\text{nprobe} \in \{8, 10, 11\}$ 。这里,数据集 ICEWS14/ICEWS15/WIKIDATA 所对应的默认参数分别为: $\text{epoch} = 500, b = 1024, d = 400, \text{lr} = 0.001, \gamma = 1, n = 500, \text{nlist} = 84/101/106, \text{nprobe} = 8/10/11$ 。模型的优化器选择 Adagrad。注意,每 20 个 epoch 验证一次,每当验证集上的 MRR 值下降时,模型停止学习。

数据集:使用三种数据集对模型进行验证,包括: ICEWS14^[15], ICEWS15^[15], WIKIDATA^[24]。具体来说:

ICEWS14: ICEWS14 是 Integrated Crisis Early Warning System(ICEWS)的一个版本,它是一个用于全球政治事件监测和预测的系统,更多信息可以参考 <http://www.icews.com/>。ICEWS14 数据集中包含了国际事件的信息,这些信息涉及到各种类型的事件,如政治抗议、冲突、外交关系等,其包含了 2014 年所有事件的短程版本。该数据集提供了事件发生的时间、地点、参与者等关键信息,通常被用于训练和评估 TKG 补全/预测任务。

ICEWS15: ICEWS15 是 ICEWS 的另一个版本,与 ICEWS14 相比,它可能包含更新的数据和更准确的事件信息。ICEWS15 数据集通常也被用于 TKG 补全/预测任务,并且可能具有更好的数据质量和覆盖范围,其包含了 2005-2015 年之间发生的所有事件的远程版本。

WIKIDATA: WIKIDATA 是一个包含了丰富的实体

和关系信息的开放式知识图谱。该数据集涵盖了各种领域的知识,包括人物、地点、事件、组织等。在TKG补全/预测任务中,WIKIDATA数据集可以作为一个额外的资源,用于丰富实体和关系之间的信息,从而提高

预测的准确性。

表1统计了所有数据的规模。其中,每一种数据集按照8:1:1比例划分为训练集、验证集和测试集。

表1 TKG数据集的规模统计

数据集	实体	关系	时间跨度	训练集	验证集	测试集
ICEWS14	7 128	284	2014	70 390	8 799	8 799
ICEWS15	10 302	283	2005-2015	368 902	46 113	46 113
WIKIDATA	11 145	108	25-2020	120 055	15 007	15 007

3.2 评估指标与基线方法

评估指标:为了评估该模型的效果,使用时态KGE常用的评测指标。本质上,该检索任务是给定查询下的排序问题,且给定查询所返回答案的排名分数越高,查询效果就越好。给定一个测试查询集 q ,对于它的每个答案集 a ,利用查询函数来检测 $v \in a$ 的排名。用 $\text{Rank}(v)$ 来计算 v 的排名,然后计算回答查询 q 的度量。使用不同的指标作为回答查询 q 的度量,即 K 的命中(Hits@K)和平均倒数排名(MRR),如下:

$$\text{Hits@K}(q) = \frac{1}{|a|} \sum_{v \in a} \mathbb{I}(\text{Rank}(v) \leq K) \quad (8)$$

$$\text{MRR} = \frac{1}{|a|} \sum_{v \in a} \text{Rank}(v)^{-1} \quad (9)$$

其中, $\mathbb{I}[x \leq K]$ 是indicator函数,如果 $x \leq K$ 为真,则它等于1,否则等于0,Hits@K,MRR $\in [0, 1]$ 的值越大,性能越好。

基线方法:该研究考虑了两类KGE基线,包括经典KGE搜索模型和时态KGE搜索模型。该文的目的是验证在给基于嵌入的KG搜索模型添加索引以提升检索效率,因此,选择了最为相关的上述两类对比模型。同时,由于经典/时态KGE搜索模型较为成熟,因此,对这两类基线分别考虑了广泛使用的模型,即经典KGE搜索模型:TransE^[7]、RotatE^[10]和QuatCNNE^[11];时态KGE搜索模型:HyTE^[12]、TA-TransE^[15]和CTATKGE^[19]。这些模型与文中模型在一些方面具有强相关性,比如:基于距离的模型架构等。

3.3 结果与分析

3.3.1 效率分析

为了验证添加向量索引可以提高TKG搜索模型的检索效率,本节分别计算了本模型在三个数据集(ICEWS14,ICEWS15,WIKIDATA)上的搜索效率,并与其它模型进行了对比。其中,搜索效率如图3所示。Ours表示该文的搜索模型,Ours(-index)移除索引的搜索模型。结果显示,Ours模型的搜索效率是最优的,这说明建立向量索引IndexIVFFlat到时态知识向量库中可以显著加快对实体在TKG上的搜索效率。

具体来说,可以从图3中看出,Ours在三个数据集

ICEWS14、ICEWS15和WIKIDATA上的搜索时间分别是22.94 s、199.17 s和69.18 s;同时,Ours(-index)在三个数据集ICEWS14、ICEWS15和WIKIDATA上的搜索时间分别是45.67 s、319.23 s和107.79 s。可以发现,对比Ours(-index)在三个数据ICEWS14、ICEWS15和WIKIDATA上的搜索时间,Ours的搜索时间效率分别提升了49.77%、37.16%和35.82%。原因是:IndexIVFFlat索引是将训练完备的实体向量进行聚类以划分为不同的搜索空间,在搜索时,查询向量只需要与所在聚类单元中的这些实体向量进行比较。这极大地减少了查询向量与所有候选实体向量进行比较的时间。因此,IndexIVFFlat索引对缩短对TKG搜索的时间是有效的。

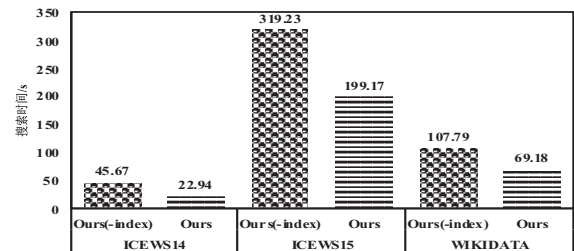


图3 模型的搜索效率

3.3.2 有效性评估

为了验证模型的有效性,本节使用评估指标(MRR(%)和Hits@10(%))在三个数据集(ICEWS14,ICEWS15,WIKIDATA)上对本模型进行了搜索评估,并与对比模型进行了比较。其中,评估结果如表2所示。黑体表示最优的结果,下划线表示次优结果。

表2展示了Ours与对比模型(TransE和RotatE;HyTE和TA-TransE)在三个数据集(ICEWS14,ICEWS15,WIKIDATA)上的搜索指标。结果显示,Ours模型的搜索效果是最优/次优的,证明了该方法的有效性。

具体来说:(1)与经典KGE搜索模型(TransE,RotatE和QuatCNNE)相比,Ours的指标分别得到了不同程度的上升。例如,在ICEWS14上,从RotatE到Ours的MRR指标上升了15.39%。其主要原因是

Ours 考虑了 KG 的时间敏感性。(2)与时态 KGE 搜索模型(HyTE,TA-TransE 和 CTATKGE)相比,Ours 的指标也分别得到了不同程度的上升。例如,在 WIKIDATA 上,从 TA-TransE 到 Ours 的 Hits@10 指标上升了 33.88%。其主要原因是模型可以更有效地学

习到不同聚类空间中的向量表示。

综上所述,与强相关的模型进行对比,文中模型在三个数据集上拥有显著的搜索性能,这证明了模型的有效性。

表 2 模型在 ICEWS14、ICEWS15 和 WIKIDATA 上的搜索效果 %

数据集 指标	ICEWS14		ICEWS15		WIKIDATA	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
TransE ^[7]	14.32	38.57	28.39	68.59	26.17	50.29
RotatE ^[10]	21.37	41.78	29.35	61.56	32.49	68.39
QuatCNNE ^x ^[11]	21.64	40.32	29.72	59.40	32.90	65.99
HyTE ^[12]	15.18	39.66	30.51	70.45	26.46	49.40
TA-TransE ^[15]	14.06	37.84	28.87	69.11	26.17	63.64
CTATKGE ^[19]	25.22	61.22	32.02	70.82	43.95	76.25
Ours	24.66	61.98	31.31	71.70	70.88	85.20

3.3.3 参数敏感性

为了验证参数对模型的影响,本节通过改变嵌入维度 d 的大小以计算模型在数据集 ICEWS14 上的搜

索效率和搜索效果(MRR(%) 和 Hits@10(%))。其中,不同嵌入维度的大小对模型的影响结果如图 4 所示。

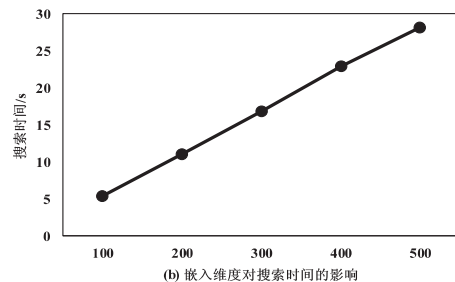
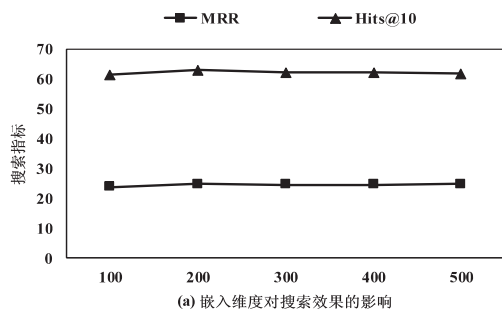


图 4 嵌入维度对模型的搜索效果和搜索效率的影响

图 4(a) 展示了文中模型(Ours)在数据集 ICEWS14 上的嵌入维度对模型的搜索效果的影响。可以看出,随着嵌入维度的增加,搜索指标 MRR 和 Hits@10 轻微波动,证明了嵌入维度对模型轻微敏感。图 4(b) 展示了 Ours 在数据集 ICEWS14 上的嵌入维度对模型的搜索效率的影响。可以发现,随着嵌入维度的增加,搜索效率显著增加。原因是较大的嵌入维度导致更大的内存消耗,使得模型训练的时长增加。因此,选择合适的参数可以降低模型的搜索开销。

4 结束语

探究基于向量索引的 TKG 的高效搜索方法,为大规模 KB 的搜索密集型需求提供了一种有效的解决方案。该文提出了一种针对大规模 TKG 的高效搜索方法。具体来说,通过将实体、关系和时间信息融合到向量空间中,并利用 LSTM 学习关系类型的时间感知,从而得到训练良好的时态知识向量库。同时,利用向量数据库构建了一个索引库(IndexIVFFlat),通过聚类操作划分搜索空间,从而建立了 TKG 的索引库。最后,

使用实验(包括:效率分析、有效性验证和参数敏感性分析)证明了该方法的有效性。实验结果显示,该方法在 TKG 上执行近似性搜索时,具有较高的准确性和效率。表明,该研究为解决大规模 KB 搜索密集型应用中的时间效率问题提供了一种新的思路和方法。

未来的工作可以进一步探索如何进一步优化搜索算法,提高搜索效率,并将该方法应用于更广泛的领域和应用场景中。

参考文献:

- [1] SUN J, CHEN J, DING G, et al. Spectrum recommendation in cognitive internet of things: a knowledge-graph-based framework[J]. IEEE Transactions on Cognitive Communications and Networking, 2024, 10(1): 21-34.
- [2] TAN S, GE M, GUO D, et al. Knowledge-based embodied question answering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 11948-11960.
- [3] SUN Y, WANG J, LIN H, et al. Knowledge guided attention and graph convolutional networks for chemical-disease relation extraction[J]. IEEE/ACM Transactions on Computational

- Biology and Bioinformatics, 2021, 20(1):489–499.
- [4] 毛慧慧,赵小乐,杜圣东,等.基于时序知识图谱嵌入的短期地铁客流量预测[J].计算机学报,2023,50(7):213–220.
- [5] WANG Q, MAO Z, WANG B, et al. Knowledge graph embedding: a survey of approaches and applications [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12):2724–2743.
- [6] JI S, PAN S, CAMBRIA E, et al. A survey on knowledge graphs: representation, acquisition, and applications [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2):494–514.
- [7] BORDES A, USUNIER N, GARCIA-DURÁN A, et al. Translating embeddings for modeling multi-relational data [C]//Proceedings of the 26th international conference on neural information processing systems. Lake Tahoe: Curran Associates Inc, 2013:2787–2795.
- [8] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes [C]//Proceedings of the twenty-eighth AAAI conference on artificial intelligence. Québec City: AAAI, 2014:1112–1119.
- [9] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion [C]//Proceedings of the twenty-ninth AAAI conference on artificial intelligence. Austin: AAAI, 2015:2181–2187.
- [10] SUN Z, DENG Z H, NIE J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space [C]//International conference on learning representations. New Orleans: OpenReview, 2018:1–18.
- [11] 熊伟,陈浩,苏鸿宇. QuatCNNE: 一种面向多关系模式的知识图谱嵌入模型[J/OL]. 计算机学报:1–11 [2024-08-17]. <http://kns.cnki.net/kcms/detail/11.1826.tp.20240701.1016.002.html>.
- [12] 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: Association for Computational Linguistics, 2018:2001–2011.
- [13] XU C, NAYYERI M, ALKHOORY F, et al. TeRo: a time-aware knowledge graph embedding via temporal rotation [C]//Proceedings of the 28th international conference on computational linguistics. Barcelona: International Committee on Computational Linguistics, 2020:1583–1593.
- [14] XU C, NAYYERI M, CHEN Y Y, et al. Geometric algebra based embeddings for static and temporal knowledge graph completion [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(5):4838–4851.
- [15] GARCIA-DURAN A, DUMANČIĆ S, NIEPERT M. Learning sequence encoders for temporal knowledge graph completion [C]//Proceedings of the 2018 conference on empirical methods in natural language processing. Brussels: Association for Computational Linguistics, 2018:4816–4821.
- [16] HOU X, MA R, YAN L, et al. T-gae: a timespan-aware graph attention-based embedding model for temporal knowledge graph completion [J]. Information Sciences, 2023, 642:119225.
- [17] WANG M, CHEN W, WANG S, et al. Efficient search over incomplete knowledge graphs in binarized embedding space [J]. Future Generation Computer Systems, 2021, 123:24–34.
- [18] BAI Y, LV X, LI J, et al. Answering complex logical queries on knowledge graphs via query computation tree optimization [C]//International conference on machine learning. Honolulu: PMLR, 2023:1472–1491.
- [19] 陈小英,熊盛武,王盛,等.基于上下文时序关联的时序知识图谱嵌入方法[J].武汉大学学报:理学版,2023,69(2):249–257.
- [20] WANG M, SHEN H, WANG S, et al. Learning to hash for efficient search over incomplete knowledge graphs [C]//2019 IEEE international conference on data mining (ICDM). Beijing: IEEE, 2019:1360–1365.
- [21] WANG M, WU T, QI G. A hash learning framework for search-oriented knowledge graph embedding [C]//European conference on artificial intelligence. Santiago de Compostela: IOS Press, 2020:921–928.
- [22] ABUODA G, THIRUMURUGANATHAN S, ABOULNAGA A. Accelerating entity lookups in knowledge graphs through embeddings [C]//2022 IEEE 38th international conference on data engineering (ICDE). Kuala Lumpur: IEEE, 2022:1111–1123.
- [23] OBRACZKA D, RAHM E. Fast hubness-reduced nearest neighbor search for entity alignment in knowledge graphs [J]. SN Computer Science, 2022, 3(6):501.
- [24] LEBLAY J, CHEKOL M W. Deriving validity time in knowledge graph [C]//Companion proceedings of the web conference. Lyon: ACM, 2018:1771–1776.