

类型增强的时态知识图谱表示学习模型

何 鹏^{1,2} 周 刚¹ 陈 静¹ 章梦礼¹ 宁原隆¹

¹(战略支援部队信息工程大学 郑州 450001)

²(郑州工程技术学院 郑州 450044)

(helen830209@163.com)

Type-Enhanced Temporal Knowledge Graph Representation Learning Model

He Peng^{1,2}, Zhou Gang¹, Chen Jing¹, Zhang Mengli¹, and Ning Yuanlong¹

¹(Information Engineering University, Zhengzhou 450001)

²(Zhengzhou University of Technology, Zhengzhou 450044)

Abstract Knowledge graph representation learning, which aims to represent elements (entities and relations) of knowledge graphs in low-dimensional continuous vector spaces, is able to effectively perform knowledge graph completion and improve computational efficiency. And it has been a significant issue throughout the whole process from knowledge graph construction to application. However, most existing approaches of knowledge graph representation learning have been developed only on the basis of static structured triples and ignore the time attribute of knowledge and entity type features, which limits their performance on knowledge graph completion and semantic computation. To conquer this problem, based on a famous tensor factorization technique, we propose a type-enhanced temporal knowledge graph representation learning model called T-Temp. The T-Temp model explicitly incorporates different forms of time information into the process of knowledge graph representation learning, and takes advantage of type compatibility between entities and their associated relations to fully explore the implicit type features of entities, so as to further improve the performance of knowledge graph representation learning. In addition, we prove that T-Temp is fully expressive and compared with competitive models, our model has lower time and space complexity. Extensive experiments on several real-world temporal knowledge graphs demonstrate the merits and advancement of our proposed model.

Key words knowledge graph; temporal knowledge graph; knowledge graph completion; temporal knowledge representation learning; knowledge graph embedding

摘 要 知识图谱表示学习旨在将知识图谱中的元素（实体和关系）表示在低维的连续向量空间中，可以有效地实现知识图谱补全并提高计算效率，是贯穿知识图谱构建和应用全过程的重要问题。现有的知识图谱表示学习模型大多基于静态的结构化三元组，忽略了知识的时间动态性和实体的类型特征，限制了它们在知识图谱补全和语义计算中的表现。针对这一问题，提出一种类型增强的时态知识图谱表示学习模型（type-enhanced temporal knowledge graph representation learning model, T-Temp），基于经典的张量分解技术，将不同形式的时间信息显式融合到知识图谱表示学习过程中。同时，利用实体与关系间的类型兼容性，充分挖掘隐含在实体中的类型特征，进一步提升表示学习的准确性。此外，证明 T-Temp 模型具有完全表达性，且与同类模型相比，具有较低的时空复杂度。在多个真实的时态知识图谱上进行的详细实验说明了 T-Temp 模型的有效性和先进性。

收稿日期：2021-12-15；修回日期：2022-06-19

基金项目：河南省科技攻关项目（222102210014）；河南省自然科学基金项目（222300420590）

This work was supported by the Scientific and Technological Key Project of Henan Province (222102210014) and the Natural Science Foundation of Henan Province (222300420590).

关键词 知识图谱; 时态知识图谱; 知识图谱补全; 时态知识图谱表示学习; 知识图谱嵌入

中图法分类号 TP391

知识图谱(knowledge graph)以多关系有向图的形式组织和存储现实世界的知识. 其中, 节点表示实体(人名、地名、机构名、概念等), 边表示实体间的语义关系. 因此, 知识图谱又可以看成是结构化三元组的集合, 对应有向图中的边(关系)及其相连的2个节点(实体). 自从2012年, 谷歌首次提出知识图谱的概念并将其成功应用于信息检索领域以来, 知识图谱在学术界和工业界得到了广泛的关注和研究. 现有的知识图谱包括 Freebase^[1]、Dbpedia^[2]、WordNet^[3]等, 已经在许多人工智能相关的应用中起到了巨大的推动和支撑作用, 如个性化推荐^[4]、智能问答^[5]、信息检索^[6]以及自然语言处理^[7]等.

现有的知识图谱规模庞大, 往往包含上亿条事实三元组, 不可避免的存在由数据缺失造成的不完整性问题. 为了解决此问题, 人们提出了各种知识图谱补全方法, 通过基于已有的事实来推理和预测知识图谱中缺失的链接. 知识图谱表示学习(knowledge graph representation learning)是一种有效的知识图谱补全方法, 它可以自动预测缺失的知识, 同时解决知识图谱中的数据稀疏性和计算效率问题, 为基于知识图谱的深度学习工作提供了极大的便利.

知识图谱表示学习又称为知识图谱嵌入(knowledge graph embedding), 旨在将知识图谱中的元素(实体和关系)映射到低维的连续向量空间中, 学习实体和关系的嵌入表示, 同时保持知识图谱的内在结构和语义信息, 即将知识图谱的符号化表示形式转换成数值化表示形式, 从而实现高效的语义计算. 现有的知识图谱表示学习模型通常是静态的, 忽略了事实的时间动态性和时序依赖性. 而在现实世界中, 时间是实体和关系所具有的重要属性, 事实往往随时间的变化而发生动态的演化. 常见的知识图谱 Wikidata^[8]和 YAGO^[9]都包含了知识的时间信息. ICEWS^[10]和 GDELT^[11]是2个带有时间信息的事件知识图谱. 将知识图谱中可提供的时间信息引入三元组中, 构成的带有时间戳的四元组集合称为时态知识图谱(temporal knowledge graph). 如图1所示, 连接实体的关系带有时间戳, 表明该事实发生的具体时间, 比如(美国, 总统, 奥巴马, [2009—2017])只在2009—2017年间有效. 当采用静态的知识图谱表示学习模型对时态知识图谱进行补全时, 很容易混淆相似实体的语义信息. 比如, 当对缺失实体的事实

(美国, 总统, ?, [1993—2001])进行链接预测时, 如果忽略了给定的时间戳[1993—2001], 则可能会混淆“克林顿”和其他历届美国总统, 给出错误的答案. 时态知识图谱表示学习模型通过将可提供的时间信息显式或隐式的融合到知识图谱表示学习过程中, 使得链接预测的结果随时间不同而产生不同的排序, 从而有效地区分相似语义的实体, 提高时态知识图谱补全的准确性.

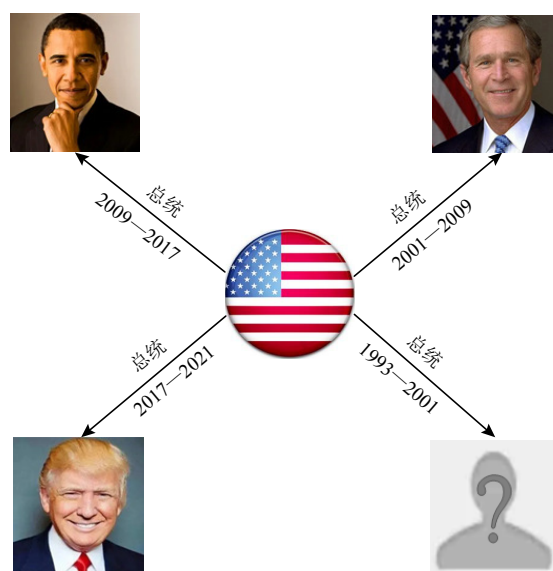


Fig. 1 An example of temporal knowledge subgraph extracted from the Wikidata

图1 从 Wikidata 抽取的时态知识图谱子图示例

由于上述优点, 最近几年时态知识图谱表示学习迅速成为知识图谱领域的研究热点. 已经有一些时态知识图谱表示学习模型通过利用时间信息来改善表示学习的效果, 但它们普遍存在一个或多个问题: 1) 不具备完全表达性, 即不能准确的区分事实(比如, “2009—2017年间的美国总统是奥巴马”)与非事实(如“2009—2017年间的美国总统是希拉里”)^[12], 从而限制了模型的表示能力. 2) 只考虑时间点形式的时间戳(如[2014-07-16]), 而没有考虑时间段形式的时间戳(如[2009—2017]). 比如, 在 Wikidata 和 YAGO 这2个著名的时态知识图谱中, 事实通常带有时间段形式的时间戳, 由于时间的连续性, 建模时间段形式的时间戳是一项具有挑战性的工作. 3) 存在冗余计算和时空复杂度高的问题. 4) 没有充分利用类型兼容性. 类型兼容性是指知识图谱中的实体除了表

示个体语义信息外,还隐含了一般的类型语义信息,且一个特定关系总是连接具有相同类型的实体.比如(美国,总统,?)中,“总统”这一关系总是连接“国家”类型和“人”类型的实体.在推理缺失信息时,利用类型兼容性这一先验知识可以判断缺失实体的类型应该是“人”,它在向量空间中应该与其他“人”类型的实体位置接近,从而可以更好地限制和优化实体嵌入.

为了解决这些问题,本文提出一种类型增强的时态知识图谱表示学习模型(type-enhanced temporal knowledge graph representation learning model, T-Temp),用于解决时态知识图谱中的知识补全和语义计算问题.模型基于张量分解技术,将时态知识图谱看成3阶张量,语义关系和时间信息联合索引其中的一个模式向量.同时,设计一种类型兼容性函数,自动捕获实体的类型特征并优化实体表示.此外,模型具有完全表达性,且可以建模时态知识图谱中常见的时间点形式和时间段形式的时间信息,具有普遍适用性.

本文的主要贡献包括3个方面:

1)提出了一种基于张量分解的时态知识图谱表示学习模型 T-Temp,将时间信息显式地编码到表示学习过程中,并利用实体和关系的类型兼容性,学习实体、关系、时间和类型的嵌入表示,提升表示学习的有效性.

2)理论上证明 T-Temp 模型具有完全表达性,并与现有的同类模型做对比分析,说明其具有较低的时间和空间复杂度.

3)在真实的时态知识图谱 ICEWS、Wikidata、YAGO 中抽取出来的4个公开数据集上开展广泛的实验.在链接预测任务上的结果表明 T-Temp 模型的性能较其他先进模型有显著提升,类型嵌入的可视化聚簇结果也表明 T-Temp 模型能够有效地捕获实体的类型特征.

1 相关工作

本节主要介绍与本文工作相关的知识图谱表示学习模型,包括静态知识图谱表示学习模型、时态知识图谱表示学习模型和类型增强的知识图谱表示学习模型,具体可参考综述文献[13–16].

1.1 静态知识图谱表示学习模型

现有的大多数知识图谱表示学习模型基于静态事实进行建模,大体上可分为2类:基于平移距离的模型和基于张量分解的模型.

基于平移距离的模型通常将关系建模成向量空间中的平移或旋转操作,用关系操作后的实体间距离度量事实的真实性.著名的基于平移距离的模型 TransE^[17] 及其变体模型 TransH^[18]、TransR^[19]、TransD^[20] 等均将关系建模为实数向量空间中的平移操作.而最近的 RotatE^[21] 和 HAKE^[22] 模型则将关系建模为复数向量空间中的旋转操作,从而推理知识图谱中的各种关系模式,提高模型的关系建模能力.然而,大部分平移距离模型不具备完全表达性^[12],限制了其表示能力,基于张量分解的 Simple^[12] 模型能够克服这一缺点. Simple 模型受到经典张量分解技术——典范多元(CANDECOMP/PARAFAC, CP)分解^[23]——的启发,将每个事实三元组对应3阶张量中的一个元素,其中,实体和关系分别索引该张量的一个模式向量,同时构建反向关系来统一不同位置的实体嵌入. Lacroix 等人^[24] 也提出了类似的基于 CP 分解的模型. ComplEx^[25] 模型是另一个基于张量分解的模型,它通过将实体映射到复数向量空间而非实数向量来建模实体间的关系. ComplEx 和 Simple 都具有完全表达性^[12,24],但与 Simple 相比, ComplEx 存在冗余计算问题.此外, Yang 等人^[26] 提出一个简化版的张量分解模型 DistMult, 王培妍等人^[27] 提出一种基于张量分解的知识超图模型 Typer.

尽管静态知识图谱表示学习模型取得了较好的表现,但它们没有考虑知识的时态演化性,容易造成相似语义实体的混淆.

1.2 时态知识图谱表示学习模型

最近,一些研究者通过对静态模型进行时态扩展,提出了时态知识图谱表示学习模型.比如, TTransE^[28]、TA-TransE^[29]、HyTE^[30]、Duration-HyTE^[31] 模型是对经典静态模型 TransE 的时态扩展,和 TransE 一样,这些动态模型不具备完全表达性.受到静态模型 RotatE 的启发, TeRo^[32] 模型将时间信息建模成复数空间中的旋转操作,实体通过沿不同角度的时间旋转来体现不同时期的时间特征. ChronoR^[33] 模型同样受到静态模型 RotatE 的启发,但实体的旋转变换由时间和关系共同决定,且采用向量间的角度而非距离来度量事实的真实性. DE-Simple^[34] 模型在静态模型 Simple 的基础上,通过引入 DE(diachronic embedding)函数^[35]来学习实体的时间演化特性. ConT^[36] 模型可以看成是静态模型 Tucker^[37] 的扩展,它用特定时间张量代替 Tucker 分解^[38] 中的核张量.由于其时间嵌入需要大量的参数,在训练过程中效率较低且容易产生过拟合. ChronoR、DE-Simple、ConT 模型都只能处理离

散的时间点信息,而没有考虑连续的时间段信息. TComplEx 和 TNTComplEx^[39] 模型将时态知识图谱表示成一个 4 阶张量,并添加一个时态模式向量来扩展静态模型 ComplEx. 同样, TComplEx 和 TNTComplEx 模型也存在冗余计算问题.

1.3 类型增强的知识图谱表示学习模型

实体的类型特征体现了实体的一般语义和类别,相关工作利用这一特征进一步优化知识表示的学习效果. TKRL^[40] 模型首次引入显式的实体类型来增强 TransE. JOIE^[41] 模型将知识图谱表示为本体视图(即类型信息)和实例视图(即实体信息),并联合编码这 2 个视图. TaRP^[42] 模型根据实体类型定义关系类型,并采用贝叶斯规则拟合关系类型和实体类型间的语义相似性. 上述模型均需要提供额外的类型信息. 最近, Jain 等人^[43] 认为实体类型普遍隐含在知识图谱中的实体和实体间的复杂语义关系中,并提出 TypeDM 和 TypeComplEx 模型,通过建模实体和关系间的类型兼容性扩展 DistMult 和 ComplEx 模型,自动学习实体的类型嵌入,不需要额外的类型信息. 然而,所有这些类型增强模型都没有考虑知识的时间动态性.

与上述工作相比,本文工作主要致力于解决时态知识图谱中的知识补全问题. 本文提出的类型增强的时态知识图谱表示学习模型 T-Temp 属于张量分解模型. 在表示学习过程中, T-Temp 模型可以以较低的时空消耗,充分利用各种形式的时间信息,并自动学习和表示实体的类型特征,不需要提供额外的类型信息. 此外,据我们所知, T-Temp 模型是为数不多的具有完全表达性的时态模型.

2 类型增强的时态模型 T-Temp

本节首先对时态知识图谱表示学习中的相关问题进行形式化定义,并对一些基本的概念和符号进行解释;然后详细介绍所提出的 T-Temp 模型.

2.1 问题定义

定义 1. 时态知识图谱. 时态知识图谱表示成一个带有时间信息的多关系有向图 $G = (E, R, T)$, 其中 E 是节点(实体)集, R 是边(关系)集, T 是时间戳集. 因此, 时态知识图谱又可以看成是四元组 $(h, r, t, \tau) \in G$ 或 $(h, r, t, [\tau_s, \tau_e]) \in G$ 的集合, 其中 $h, t \in E$ 分别称为头实体和尾实体, $r \in R$ 是它们之间的关系, $\tau \in T$ 或 $[\tau_s, \tau_e] \in T$ 是与事实相关联的时间戳. 具体来说, τ 表示事实发生在一个特定的时间点, $[\tau_s, \tau_e]$ 表示事实在一个开始时间为 τ_s 、结束时间为 τ_e 的持续时间段内均

有效.

定义 2. 时态知识图谱存在不完整性问题. 本文用 $W \subset E \times R \times E \times T$ 表示现实世界中的全部事实, 时态知识图谱 G 是 W 的子集(即 $G \subset W$), 时态知识图谱补全是一个根据 G 推理 W 的问题.

定义 3. 知识图谱表示学习模型通常定义 3 件事:

- 1) 嵌入函数——将知识图谱中的元素映射为向量、矩阵或张量等嵌入表示;
- 2) 得分函数——将上述嵌入表示作为输入, 通过数值运算获得输出, 作为评估事实真实性的得分;
- 3) 损失函数——通过最大化所有已知事实的得分来学习和优化各元素的嵌入表示.

定义 4. 典范多元(CP)分解是一个经典的张量分解技术, 类似于矩阵的低秩分解. 在 CP 分解过程中, 一个 N 阶张量 $X \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ 可以表示成 R 个秩-1 张量之和, 即 N 个因子向量 $u_r^{(i)}$ 的外积之和

$$X = \sum_{r=1}^R u_r^{(1)} \otimes u_r^{(2)} \otimes \dots \otimes u_r^{(N)}, \quad (1)$$

或元素形式:

$$X_{i_1 \dots i_N} = \sum_{r=1}^R u_{i_1, r}^{(1)} u_{i_2, r}^{(2)} \dots u_{i_N, r}^{(N)} = \langle u_{i_1}^{(1)}, u_{i_2}^{(2)}, \dots, u_{i_N}^{(N)} \rangle, \quad (2)$$

其中 $r \in \{1, 2, \dots, R\}$, $u_r^{(i)} \in \mathbb{R}^{I_i}$, \otimes 表示外积, $\langle \cdot, \cdot \rangle$ 表示哈达玛积(对应元素乘积)之和.

2.2 T-Temp 模型整体架构

本文基于张量分解的模型, 学习实体、关系和时间戳的向量化嵌入. 同时, 利用实体和关系的类型兼容性, 自动学习实体的类型嵌入, 进一步优化实体表示. 图 2 展示了模型的整体架构, 按照知识图谱表示学习的过程, 首先, 我们提出将已知四元组嵌入到实数向量空间中的嵌入函数; 接着, 定义基于 CP 分解的得分函数以及基于语义相似性的类型兼容性函数, 并将两者结合, 形成最终的得分函数; 最后, 设计一个带有正则化的交叉熵损失函数作为优化目标, 学习各元素的嵌入表示.

2.3 嵌入函数

嵌入函数又称为编码器. 本文将时态知识图谱中的实体、关系和时间戳元素显式的编码到维度为 d 的实数向量空间. 如图 2 所示, 对于给定四元组 (h, r, t, τ) , 向量 $h, t \in \mathbb{R}^d$ 分别是头实体 h 和尾实体 t 经过嵌入函数映射后得到的嵌入表示, 代表头/尾实体的个体语义特征, 向量 $r \in \mathbb{R}^{r_d}$ 和 $\tau \in \mathbb{R}^{\tau_d}$ 分别是关系 r 和时间戳 τ 经过嵌入函数映射后得到的嵌入表示, 代表关系和时间戳的语义特征, 其中 $r_d + \tau_d = d$.

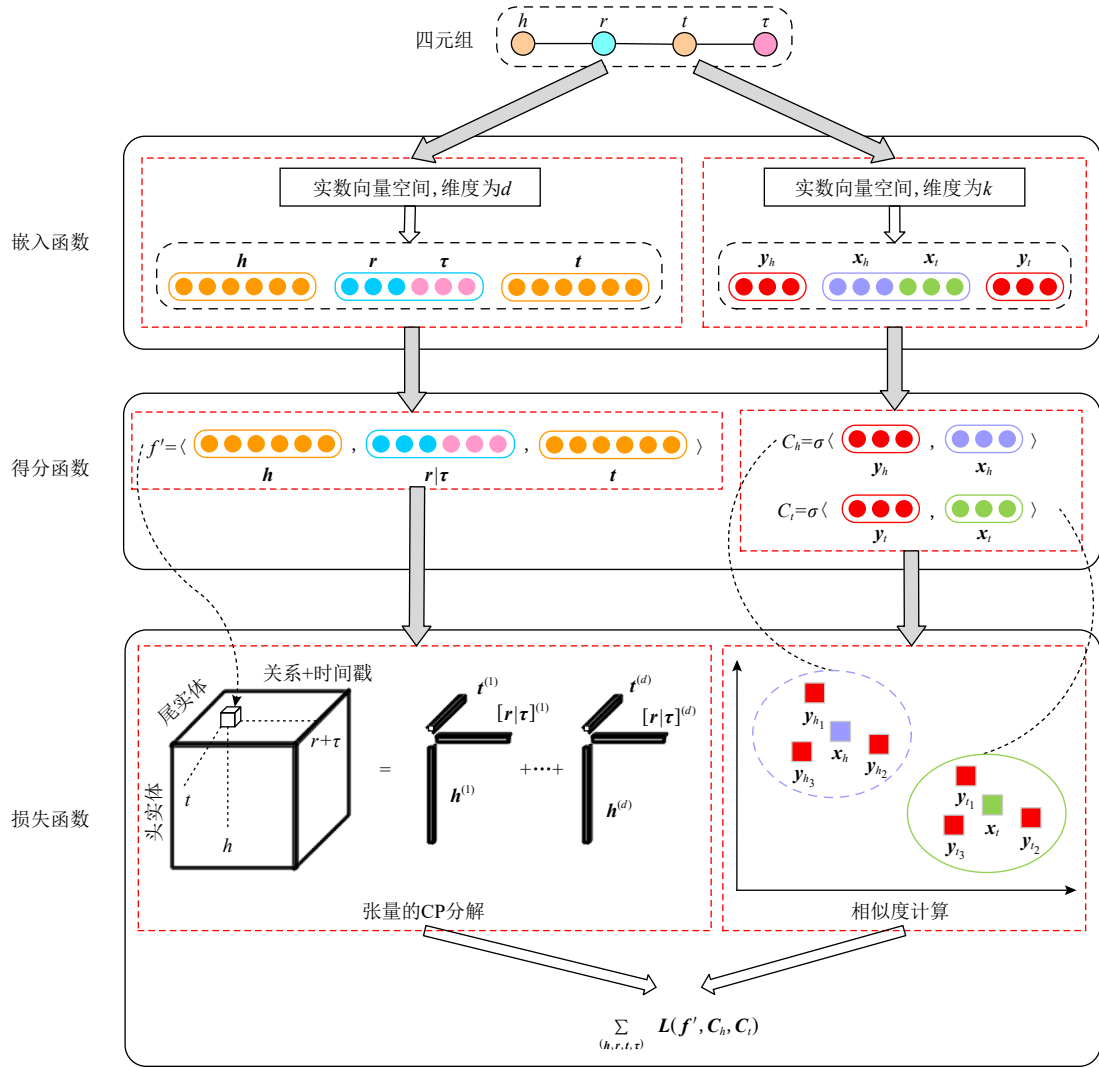


Fig. 2 T-Temp architecture

图2 T-Temp 架构

为了学习实体所隐含的类型特征并建模实体与关系间的类型兼容性,嵌入函数进一步将实体所属的类型信息编码到维度为 k 的实数向量空间中.具体来说,向量 $y_h, y_t \in \mathbb{R}^k$ 分别表示头实体 h 和尾实体 t 的类型嵌入,代表头/尾实体的类型特征.此外,构建关系的类型属性,根据一个特定关系总是连接具有相同类型的头实体和尾实体,令关系 r 期望连接的头实体类型为关系的头类型,关系 r 期望连接的尾实体类型为关系的尾类型.并定义向量 $x_h \in \mathbb{R}^k$ 为关系 r 的头类型嵌入,表示关系的头类型特征;向量 $x_t \in \mathbb{R}^k$ 为关系 r 的尾类型嵌入,表示关系的尾类型特征.实体类型体现了多个实体的一般语义信息,往往没有实体的语义丰富,因此通常情况下 $k \ll d$.

2.4 得分函数

本文定义基于CP分解的得分函数,将时态知识图谱 G 看成一个3阶张量 $\mathbf{X} \in \mathbb{R}^{|E| \times |R| \times |T| \times |E|}$,其中 $|E|$ 是实

体的个数, $|R||T|$ 是关系个数 $|R|$ 和时间戳个数 $|T|$ 的乘积.头/尾实体分别索引模式-1和模式-3向量,关系和时间戳联合索引模式-2向量.根据定义4所描述的CP分解方法,头实体向量、尾实体向量、关系向量和时间戳向量组成的多线性乘积可以用来估计张量 \mathbf{X} 中的各个元素,即四元组 (h, r, t, τ) 的得分函数为

$$f'(h, r, t, \tau) = \langle \mathbf{h}, [\mathbf{r}|\boldsymbol{\tau}], \mathbf{t} \rangle, \quad (3)$$

其中 \mathbf{h} 表示头实体向量, \mathbf{t} 表示尾实体向量, $[\mathbf{r}|\boldsymbol{\tau}]$ 表示关系向量 \mathbf{r} 和时间戳向量 $\boldsymbol{\tau}$ 的级联.该得分函数的值越大,说明四元组越真实.需要指出的是,这种原始的基于CP分解的方法存在同一实体位于头/尾不同位置时的嵌入向量不一致问题,为了解决这个问题,本文在训练时采用与静态模型 Simple 类似的,构建反向关系的方法统一实体的嵌入.

另一方面,考虑到本文所提出的模型不需要提供额外的类型信息,实体和关系的语义关联中其实

隐含了它们的类型特征. 为了可以自动学习和挖掘这些类型特征, 根据实体和关系的类型兼容性这一先验知识, 我们定义基于语义相似度的类型兼容性函数来建模实体和关系间的类型兼容性, 采用与余弦相似度的计算成比例的向量内积形式. 对于关系与其相连的头实体, 类型兼容性函数为

$$C_h(h, r) = \sigma(\mathbf{y}_h, \mathbf{x}_h), \quad (4)$$

其中 \mathbf{y}_h 表示头实体的类型向量, \mathbf{x}_h 表示关系 r 期望连接的头类型向量, σ 是 sigmoid 函数. 类似地, 对于关系与其相连的尾实体, 类型兼容性函数为

$$C_t(t, r) = \sigma(\mathbf{y}_t, \mathbf{x}_t), \quad (5)$$

其中 \mathbf{y}_t 表示尾实体的类型向量, \mathbf{x}_t 表示关系 r 期望连接的尾类型向量.

将式(3)与式(4)(5)相结合, 得到四元组 (h, r, t, τ) 的最终得分函数:

$$f(h, r, t, \tau) = f'(h, r, t, \tau) C_h(h, r) C_t(t, r) = \langle \mathbf{h}, [\mathbf{r}|\boldsymbol{\tau}], \mathbf{t} \rangle \sigma(\mathbf{x}_h, \mathbf{y}_h) \sigma(\mathbf{x}_t, \mathbf{y}_t). \quad (6)$$

式(6)中 2 个类型兼容性函数可以看成是 CP 分解得分函数的系数, 即根据四元组的类型兼容性调节该得分函数的结果.

除此之外, 常见的时态知识图谱通常是异构的, 也就是说, 除了时态感知关系, 还包含大量的非时态感知关系. 比如 Wikidata 中的三元组(奥巴马, 出生地, 夏威夷州), 其关系“出生地”就是一个非时态感知关系, 即奥巴马的出生地永远是夏威夷州, 不会随时间发生变化. 为了能更好地处理这种既包含时态关系, 又包含非时态关系的异构型时态知识图谱, 本文在得分函数中增加了一个非时态组件, 相应的得分函数变为

$$f(h, r, t, \tau) = (\langle \mathbf{h}, [\mathbf{r}|\boldsymbol{\tau}], \mathbf{t} \rangle + \langle \mathbf{h}, \mathbf{r}_s, \mathbf{t} \rangle) \sigma(\mathbf{x}_h, \mathbf{y}_h) \sigma(\mathbf{x}_t, \mathbf{y}_t) = \langle \mathbf{h}, [\mathbf{r}|\boldsymbol{\tau}] + \mathbf{r}_s, \mathbf{t} \rangle \sigma(\mathbf{x}_h, \mathbf{y}_h) \sigma(\mathbf{x}_t, \mathbf{y}_t), \quad (7)$$

其中 $\mathbf{r}_s \in \mathbb{R}^d$ 表示关系 r 的非时态向量, 而 \mathbf{r} 又称为关系的时态向量表示. 对于时态感知关系来说, 由于其通常出现在含有时间信息的四元组中, 因此, 该关系对应的时态向量 \mathbf{r} 应该使正四元组得分较高, 负四元组得分较低; 而其对应的非时态向量 \mathbf{r}_s 应该使大部分不含有时间信息的三元组得分较低. 相反, 对于非时态感知关系来说, 其对应的非时态向量 \mathbf{r}_s 应该使正三元组得分较高, 负三元组得分较低; 而其对应的时态向量 \mathbf{r} 应该使大部分含有时间信息的四元组得分较低.

2.5 损失函数与时间戳预处理

在 2.4 节定义的得分函数基础上, 对于缺失尾实体的四元组 $(h, r, ?, \tau)$, 我们可以估计任意候选实体 $t_i \in E$ 的真实性:

$$P(t = t_i | h, r, \tau) = \frac{\exp(f(h, r, t_i, \tau))}{\sum_{j=1}^{|E|} \exp(f(h, r, t_j, \tau))}. \quad (8)$$

对于缺失头实体的四元组 $(?, r, t, \tau)$ 也类似. 进而, 本文采用交叉熵损失函数来学习和优化各元素的向量嵌入:

$$L(G; \theta) = \sum_{(h, r, t, \tau) \in G} \left(\sum_{t_i, h_i \in E} -\ln(P_{t_i}) - \ln(P_{h_i}) \right), \quad (9)$$

其中 P_{t_i} 表示对缺失尾实体的四元组进行估计, P_{h_i} 表示对缺失头实体的四元组进行估计, θ 表示模型的所有参数, 包括实体的嵌入向量 \mathbf{h}, \mathbf{t} , 关系的时态和非时态向量 \mathbf{r}, \mathbf{r}_s , 实体所属的类型嵌入向量 $\mathbf{y}_h, \mathbf{y}_t$ 以及关系期望连接的头/尾类型嵌入向量 $\mathbf{x}_h, \mathbf{x}_t$. 与大多数现有模型^[12,17,30,43]不同, 该损失函数不需要负样本, 因此避免了由负采样过程引起的时空消耗, 极大地提高了模型的训练效率.

此外, 考虑到参数正则化可以提升模型的泛化能力, 避免对训练数据的过拟合, 本文采用类似于文献[37]中的张量核范数 $\Omega_p(\theta)$ 和基于先验知识的时间戳平滑 Δ_p 作为模型的正则化项:

$$\Omega_p(\theta) = \sum_{(h, r, t, \tau) \in G} (\|\mathbf{h}\|_p^p + \|\mathbf{r}|\boldsymbol{\tau}\|_p^p + \|\mathbf{r}_s\|_p^p + \|\mathbf{t}\|_p^p), \quad (10)$$

$$\Delta_p = \frac{1}{|T|-1} \sum_{i=1}^{|T|-1} \|\boldsymbol{\tau}_{i+1} - \boldsymbol{\tau}_i\|_p^p, \quad (11)$$

其中 $\|\cdot\|_p$ 表示向量的 p -范数, $\boldsymbol{\tau}_i$ 和 $\boldsymbol{\tau}_{i+1}$ 表示任意 2 个相邻的时间戳嵌入表示. 模型的最终优化目标是最小化带有正则化项的损失函数:

$$L_1(G; \theta) = L(G; \theta) + \lambda_1 \Omega_p(\theta) + \lambda_2 \Delta_p, \quad (12)$$

其中 λ_1 和 λ_2 是加权超参.

在模型进行训练之前, 需要考虑如何处理时态知识图谱中常见的 2 种时间信息形式, 即时间点时间戳和时间段时间戳. 在事件知识图谱 ICEWS 和 GDELT 中, 事实(事件)带有时间点形式的时间戳 τ , 用来表示该事件发生的具体时间. 由于时间点的离散性特点, 这种形式的时间信息可以直接适用于 T-Temp 模型. 而在时态知识图谱 YAGO 和 Wikidata 中, 与事实相关联的往往是时间段形式的时间戳 $[\tau_s, \tau_e]$, 如何建模这种连续性时间信息是一项具有挑战性的任务. 与 TComplEx 和 TNTComplEx 模型中采用的在时间段范围内均匀采样的方法不同, 我们直接用时间段的开始时间和结束时间作为新的时间戳来代替原有的时间戳, 从而将持续性时间信息做离散化处理. 本质上是利用扩充数据量的方法尽可能多地采

集连续性时间信息, 形成模型可处理的四元组形式. 算法 1 给出了 T-Temp 模型的伪代码.

算法 1. T-Temp 模型.

输入: 训练集 S_{train} , 实体集 E , 关系集 R , 时间戳集 T , 训练总轮数 N , 批次大小 β , 嵌入维度 k , d 和 r_d , τ_d , 加权超参 λ_1 和 λ_2 ;

输出: 所有实体的嵌入向量 $\mathbf{e} \in \{\mathbf{h}, \mathbf{t}\}$ 和类型嵌入向量 \mathbf{y}_e , 所有关系的嵌入向量 \mathbf{r} , \mathbf{r}_s 和头/尾类型嵌入向量 \mathbf{x}_e , 所有时间戳的嵌入向量 $\boldsymbol{\tau}$.

- ① 对带有连续性时间信息的四元组做离散化预处理;
- ② 使用 Xavier 对 \mathbf{e} , \mathbf{r} , \mathbf{r}_s , $\boldsymbol{\tau}$, \mathbf{x}_e , \mathbf{y}_e 进行初始化;
- ③ for $i = 1$ to N do
- ④ for each batch S_{train}^β of S_{train} do
- ⑤ for each $(h, r, t, \tau) \in S_{\text{train}}^\beta$ do
- ⑥ 计算得分函数 $f(h, r, t, \tau)$; /*式(7)*/
- ⑦ end for
- ⑧ end for
- ⑨ 计算损失函数 L ; /*式(12)*/
- ⑩ 计算关于 \mathbf{e} , \mathbf{r} , \mathbf{r}_s , $\boldsymbol{\tau}$, \mathbf{x}_e , \mathbf{y}_e 损失的梯度;
- ⑪ 通过反向传播更新 \mathbf{e} , \mathbf{r} , \mathbf{r}_s , $\boldsymbol{\tau}$, \mathbf{x}_e , \mathbf{y}_e ;
- ⑫ end for

2.6 完全表达性证明

完全表达性是知识图谱表示学习模型的一个重要属性, 现有工作^[12,34,38]已经证明, 基于张量分解的静态模型 ComplEx、Simple、Tucker 具有完全表达性, 最近提出的 DE-Simple 模型是第 1 个具有完全表达性的时态模型. 本文从理论上分析 T-Temp 模型的完全表达性.

定义 5. 知识图谱表示学习模型具有完全表达性, 当且仅当给定知识图谱中的已知事实(真事实)集合, 存在一种嵌入表示, 能够正确区分真事实与假事实.

Kruskal^[44]将 N 阶张量的秩定义为能够进行 CP 分解的最小 R 值. 虽然确定给定张量的秩是个 NP 难问题^[45], 但 Kruskal^[46]已经证明 3 阶张量的秩存在一个弱上界. 受到此证明启发, 本文证明 T-Temp 模型具有完全表达性, 并给出嵌入表示的边界.

定理 1. 给定实体集 E 、关系集 R 和时间戳集 T 上的已知事实集合 G , 存在维度为 $|E| \times |R| \times |T|$ 的嵌入表示, 使得 T-Temp 模型具有完全表达性.

证明. T-Temp 模型具备完全表达性的充分条件是, 真事实和假事实的得分永不相交, 即得分函数能够正确划分所有真事实与假事实. 由于得分函数中的 C_h 和 C_t 取值位于 0~1 之间, 不影响总体函数值的正

负, 为了简化证明, 我们只关注得分函数的第 1 部分 f' .

大小为 $|E| \times |R| \times |T|$ 的嵌入向量可以看成是 $|E|$ 个大小为 $|R| \times |T|$ 的块. 对于头实体 e_i , 令其嵌入向量 $\mathbf{e}_i \in \mathbb{R}^{|E| \times |R| \times |T|}$ 第 i 块中的所有元素值为 1, 其余块中的元素值均为 0. 于是, 只有第 i 块中的元素值对四元组 (e_i, r_k, e_j, t_l) 的得分有影响. 接下来, 进一步分析嵌入向量的第 i 块.

在大小为 $|R| \times |T|$ 的第 i 块中, 令关系 r_k 和时间戳 t_l 的嵌入向量级联 $[\mathbf{r}_k | \mathbf{t}_l] \in \mathbb{R}^{|R| \times |R| \times |T|}$ 的第 i 块第 $(k \times |T| + l)$ 个元素值为 1, 其余为 0. 因此, 嵌入向量 \mathbf{e}_i 与 $[\mathbf{r}_k | \mathbf{t}_l]$ 的哈达玛积中, 只有第 i 块第 $(k \times |T| + l)$ 个元素为 1, 其余为 0. 基于以上嵌入向量的元素值设置, 如果四元组 (e_i, r_k, e_j, t_l) 为真事实, 只需要令尾实体嵌入向量 $\mathbf{e}_j \in \mathbb{R}^{|E| \times |R| \times |T|}$ 的第 i 块第 $(k \times |T| + l)$ 个元素值为 1, 否则值为 -1, 即可得到真事实的得分 $\langle \mathbf{e}_i, [\mathbf{r}_k | \mathbf{t}_l], \mathbf{e}_j \rangle$ 为 1, 假事实为 -1, 互不相交. 证毕.

2.7 时间和空间复杂度分析

时间和空间复杂度是知识图谱表示学习模型的另一个重要属性, 会直接影响模型的训练效率和可扩展性. 如表 1 所示, 本文根据嵌入函数和得分函数分析比较 T-Temp 模型和几个现有的时态模型所需的参数量和计算时间消耗, 其中 γ 为调整时态特征权重的超参^[34]. 在时间复杂度方面, 除 ConT 模型外, 所有模型都消耗与嵌入维度呈线性的时间复杂度 $O(d)$. 由于 ConT 模型涉及 3 阶张量运算, 其时间复杂度为 $O(d^3)$. 在参数个数方面, 除了 DE-Simple 模型, 其他模型的参数个数均与时间戳的个数相关. 由于本文所提出的 T-Temp 模型可以自动学习类型特征, 模型需要与类型表示相关的参数量 $k(2|E| + 4|R|)$. 而通常情况下 $k \ll d$ (比如第 3 节实验中, $d = 2000, k = 20$), 因此该部分参数量可忽略不计. 又因为 $r_d + \tau_d = d$, 所以 T-Temp 模型所需的参数个数总体上与最先进的 ChronoR, TeRo, TNTComplEx 等模型相当甚至更少.

3 实验

链接预测是标准的知识图谱补全任务, 本节通过时态知识图谱上的链接预测任务对 T-Temp 模型进行有效性验证. 首先对实验中所采用的数据集、评价指标和基线模型等进行说明. 然后将实验分为 5 组以达到不同的实验目的:

1) 在 4 个通用的数据集上对 T-Temp 模型进行链接预测实验, 并将实验结果与之前的先进模型进行对比分析, 以评估 T-Temp 模型的有效性;

Table 1 Comparison of Our proposed models and State-of-the-Art Temporal KGE Models on Time Complexity and Space Complexity

表 1 本文模型与现有时态知识图谱表示学习模型的时间复杂度和空间复杂度对比

模型	时间复杂度	空间复杂度
TTransE	$O(d)$	$d(E + R + T)$
TA-TransE	$O(d)$	$d(E + R + T)$
HyTE	$O(d)$	$d(E + R + T)$
ConT	$O(d^3)$	$d(E + R) + d^3 T $
DE-Simple	$O(d)$	$2d((3\gamma + (1 - \gamma)) E + R)$
TComplEx	$O(d)$	$2d(E + 2 R + T)$
TNTComplEx	$O(d)$	$2d(E + 4 R + T)$
TeRo	$O(d)$	$2d(E + R + T)$
ChronoR	$O(d)$	$3d(E + R) + 3(r_d R + \tau_d T)$
T-Temp	$O(d)$	$2d(E + R) + (2r_d R + \tau_d T) + k(2 E + 4 R)$

2)通过消融实验分析类型兼容性和时间段时间戳的处理方法对 T-Temp 模型性能的影响;

3)对学习到的实体和类型嵌入进行聚簇实验,并将聚簇结果进行可视化展示,以验证 T-Temp 模型能够自动捕获实体的类型特征。

4)进行超参的敏感性分析实验,以验证模型的

性能对于超参设置的敏感度。

5)链接预测任务上的案例研究实验,更细粒度地展示 T-Temp 模型如何提升链接预测结果的准确性。

3.1 数据集

本文在 4 个抽取自真实时态知识图谱的公开数据集上对 T-Temp 模型进行评估,包括: ICEWS14^[29]、ICEWS05-15^[29]、YAGO11k^[30]、Wikidata12k^[30]。其中, ICEWS14 和 ICEWS05-15 是 Garcia-Duran 等人^[29]从事事件知识图谱 ICEWS 中抽取的 2 个子集。ICEWS 包含从 1995—2015 年发生的政治事件,通过头/尾实体(比如“国家”“总统”)和相连关系(比如“进行访问”“表达会面或谈判的意图”)以及时间点形式的时间戳(如 [2014-05-23])来表示。ICEWS14 和 ICEWS05-15 分别对应 2005—2015 年间发生的政治事件。YAGO11k 和 Wikidata12k 是时态知识图谱 YAGO 和 Wikidata 的子集,与事实相关联的是时间段形式的时间戳(如 [2006-11-18], [2012-08-20])。通过 YAGO11k 和 Wikidata12k 数据集,可以证明 T-Temp 模型能够有效地处理连续性时间信息。表 2 列出了 4 个数据集的详细统计信息,需要说明的是,表中 YAGO11k 和 Wikidata12k 数据集的时间戳个数是经过 2.5 节中介绍的离散化预处理后的值。

Table 2 Statistics of Datasets

表 2 数据集的统计信息

数据集	实体个数	关系个数	时间戳个数	时间范围/年	训练集个数	验证集个数	测试集个数
ICEWS14	7 129	230	365	2014	72 826	8 941	8 963
ICEWS05-15	10 488	251	4 017	2005—2015	368 962	46 275	46 092
YAGO11k	10 623	10	398	-453—2 844	16 408	2050	2000
Wikidata12k	12 554	24	557	1 479—2018	32 497	4 062	4 062

3.2 评价指标

为了准确评估模型在链接预测任务上的性能,本文采用 2 个广泛使用的评价指标:平均倒数排名 MRR (mean reciprocal rank)和命中率 $Hits@N$ ^[21-22,30-34, 37]。首先,用所有已知实体 $e \in E$ 分别替换测试集中每个四元组 (h, r, t, τ) 的头实体 h 和尾实体 t ,从而为每个四元组创建 2 个候选元组集合 (h', r, t, τ) 和 (h, r, t', τ) 。然后,用学到的嵌入表示和得分函数为所有候选元组计算得分,并按照得分进行降序排名。与文献 [17] 中的设置一样,只对训练集和验证集中均未出现过的候选元组进行排名。根据此排名, MRR 为测试元组在候选元组集合中排名倒数的平均值:

$$MRR = \frac{1}{2|test|} \sum_{(h, r, t, \tau) \in test} \left(\frac{1}{rank_h} + \frac{1}{rank_t} \right), \quad (13)$$

其中 $rank_h$ 和 $rank_t$ 分别表示测试元组在替换头/尾实体组成的候选元组集合中的排名。 $Hits@N$ 为排在前 N 名的测试集元组的平均个数:

$$Hits@N = \frac{1}{2|test|} \sum_{(h, r, t, \tau) \in test} [C(rank_h \leq N) + C(rank_t \leq N)], \quad (14)$$

其中, $C(\cdot)$ 是条件函数,当条件成立时值为 1,否则为 0。 MRR 和 $Hits@N$ 的值越大,说明模型在链接预测任务上的性能越好。

3.3 基线模型和实验设置

本文选取了当前被应用较多的静态和时态知识图谱表示学习模型作为基线模型。静态模型包括:基于平移距离的 TransE 和 RotatE 模型,基于 CP 分解的 DistMult、ComplEx、Simple 模型;时态模型包括:

TransE 的时态扩展模型 TTransE、TA-TransE、HyTE.

基于 Tucker 分解的时态模型 ConT 以及基于 CP 分解的时态模型 DE-Simple、TComplEx、TNTComplEx、TeRo、ChronoR. 这些模型均已在第 1 节中进行了详细介绍.

为公平起见, 本文在单个 NVIDIA Geforce RTX 2080Ti GPU 上运行 T-Temp 和部分基线模型, 使用 Ubuntu 16.04 LTS 操作系统, 配置 Intel Core i7-7700 3.60GHz CPU, 128GB 内存. 本文利用 PyTorch^[47] 实现 T-Temp 模型, Adam^[48] 作为优化器. 训练批次大小为 1 000, 训练总轮数为 50, 且每 5 轮验证 1 次模型, 选择验证集上 *MRR* 值最高的模型参数进行测试. 对于 ICEWS14 和 ICEWS05-15 数据集, 最佳参数设置为 $d = 2000, k = 20, \lambda_1 = \lambda_2 = 0.01, r_d/\tau_d = 0.25$, 学习率为 0.2; 对于 YAGO11k 数据集, 最佳参数设置为 $d = 1800, k = 30, \lambda_1 = 0.1, \lambda_2 = 1, r_d/\tau_d = 0.6$, 学习率为 0.1; 对于 Wikidata12k 数据集, 最佳参数设置为 $d = 2000, k = 30, \lambda_1 = \lambda_2 = 0.01, r_d/\tau_d = 0.6$, 学习率为 0.1.

鉴于部分基线模型与本文采用的数据集和评价指标相同, 直接引用其公开报告的实验结果^[32-33]. 对于其他基线模型^[36-37], 按照其公开论文中的最佳实验设置进行模型复现.

3.4 链接预测实验

表 3 和表 4 给出了 T-Temp 模型和其他先进模型在 4 个数据集上的链接预测结果对比. 总体来说, 时

态模型在评价指标上优于静态模型, 说明在知识图谱表示学习过程中引入时间信息能够显著提升模型性能. 本文提出的 T-Temp 模型在所有数据集上的结果均超过最先进的时态模型, 表明 T-Temp 模型能够有效完成时态知识图谱补全任务.

具体来说, 在 ICEWS14 和 ICEWS05-15 这 2 个带有时间点形式时间戳的数据集上, T-Temp 模型的 *MRR* 指标结果比最先进的时态模型 ChronoR 分别提升了 3.9% 和 1.8%, 比同样基于 CP 分解的时态模型 TNTComplEx 分别提升了 8.5% 和 16%. 这是因为 T-Temp 模型不仅利用事实的时间信息, 还充分挖掘实体的类型特征, 从而进一步优化实体嵌入. 而 ChronoR 和 TNTComplEx 模型仅仅融合了时间信息, 忽略了隐含在实体关系中的类型信息, 因此无法取得更优的结果. 此外, 由于 ConT 模型中的每个时间戳需要大量参数来表示, 容易过拟合, 所以 ConT 性能欠佳, 而且大量参数也使其训练速度尤其缓慢.

YAGO11k 和 Wikidata12k 是 2 个带有时间段形式时间戳的数据集, ChronoR、DE-Simple 和 ConT 模型不能处理这种连续性时间信息, 因此它们在 YAGO11k 和 Wikidata12k 这 2 个数据集上没有结果. 尽管 TeRo 模型采用与本文模型相同的方法来处理持续性时间信息, 但 T-Temp 模型的结果仍然比 TeRo 更优. 一方面是因为 TeRo 只考虑通过时间的旋转操作来建模实体的时态演化性, 而忽略了关系和整个事实的时

Table 3 Link Prediction Results on ICEWS14 and ICEWS05-15

表 3 在 ICEWS14 和 ICEWS05-15 上的链接预测结果

模型	ICEWS14				ICEWS05-15			
	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@3</i>	<i>Hits@10</i>	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@3</i>	<i>Hits@10</i>
TransE	0.280	0.094		0.637	0.294	0.090		0.663
DistMult	0.439	0.323		0.672	0.456	0.337		0.691
ComplEx	0.467	0.347	0.527	0.716	0.481	0.362	0.535	0.729
Simple	0.458	0.341	0.516	0.687	0.478	0.359	0.539	0.708
RotatE	0.418	0.291	0.478	0.690	0.304	0.164	0.355	0.595
TTransE	0.255	0.074		0.601	0.271	0.084		0.616
TA-TransE	0.275	0.095		0.625	0.299	0.096		0.668
HyTE	0.297	0.108	0.416	0.655	0.316	0.116	0.445	0.681
ConT	0.185	0.117	0.205	0.315	0.163	0.105	0.189	0.272
DE-Simple	0.526	0.418	0.592	0.725	0.513	0.392	0.578	0.748
TComplEx	0.560	0.470	0.610	0.730	0.580	0.490	0.640	0.760
TNTComplEx	0.568	0.460	0.610	0.740	0.600	0.500	0.650	0.780
TeRo	0.562	0.468	0.621	0.732	0.586	0.469	0.668	0.795
ChronoR	0.593	0.496	0.654	0.773	0.684	0.610	0.730	0.821
T-Temp	0.616	0.521	0.678	0.790	0.696	0.619	0.749	0.847

注: 最优结果用黑体标识.

Table 4 Link Prediction Results on YAGO11k and Wikidata12k**表 4 在 YAGO11k 和 Wikidata12k 上的链接预测结果**

模型	YAGO11k				Wikidata12k			
	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@3</i>	<i>Hits@10</i>	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@3</i>	<i>Hits@10</i>
TransE	0.100	0.015	0.138	0.244	0.178	0.100	0.192	0.339
DistMult	0.158	0.107	0.161	0.268	0.222	0.119	0.238	0.460
ComplEx	0.167	0.106	0.154	0.282	0.233	0.123	0.253	0.436
Simple	0.175	0.137	0.179	0.247	0.321	0.212	0.354	0.557
RotatE	0.167	0.103	0.167	0.305	0.221	0.116	0.236	0.461
TTransE	0.108	0.020	0.150	0.251	0.172	0.096	0.184	0.329
TA-TransE	0.127	0.027	0.160	0.326	0.178	0.030	0.267	0.429
HyTE	0.105	0.015	0.143	0.272	0.180	0.098	0.197	0.333
TComplEx	0.151	0.094	0.158	0.280	0.240	0.144	0.283	0.438
TNTComplEx	0.189	0.118	0.194	0.347	0.266	0.159	0.301	0.506
TeRo	0.187	0.121	0.197	0.319	0.297	0.197	0.325	0.510
T-Temp	0.261	0.176	0.281	0.410	0.329	0.208	0.364	0.578

注：最优结果用黑体标识。

间变化属性；另一方面也进一步验证了本文的模型自动挖掘并利用实体类型特征的优点。

3.5 消融实验

为了分析类型兼容性和连续性时间信息的处理方法对模型性能的影响，本文基于 T-Temp 模型设计了 2 个变体模型，T-Temp-Type 和 T-Temp-Type(Unif)，分别表示从 T-Temp 的得分函数中移除类型兼容性函数，以及在该变体模型基础上进一步将连续性时间信息的处理方法变为 TComplEx 和 TNTComplEx 模型中采用的均匀采样方法。T-Temp 模型及其变体模型在 YAGO11k 数据集上的链接预测结果如表 5 所示。

Table 5 Results of Ablation Experiments on YAGO11k**表 5 YAGO11k 上的消融实验结果**

模型	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@3</i>	<i>Hits@10</i>
T-Temp-Type (Unif)	0.201	0.124	0.205	0.379
T-Temp-Type	0.255	0.176	0.279	0.403
T-Temp	0.261	0.176	0.281	0.410

从表 5 中可以看出，T-Temp 模型比变体模型 T-Temp-Type 在 *MRR* 评价指标上高 2.4%，说明建模类型兼容性并自动学习实体的类型特征确实能提升模型效果。而变体模型 T-Temp-Type(Unif)的结果则进一步退化成与 TNTComplEx 相当。验证了我们设计的连续性时间信息的处理方法可以更多地采集时间的语义信息，有效改进模型性能。

3.6 聚簇和可视化实验

为了评估 T-Temp 模型能否有效捕获实体的类

型特征，本部分实验对 T-Temp 模型在 ICEWS14 和 YAGO11k 数据集上学习到的实体和类型嵌入进行聚簇，并将聚簇结果可视化。具体来说，我们采用 *k*-means^[49] 算法进行聚簇，采用 t-SNE^[50] 方法对实体和类型嵌入进行降维（原来的维度分别是 *d* 和 *k*），方便可视化。图 3 分别展示了在不同数据集上的聚簇结果，不同的簇用不同颜色标识。

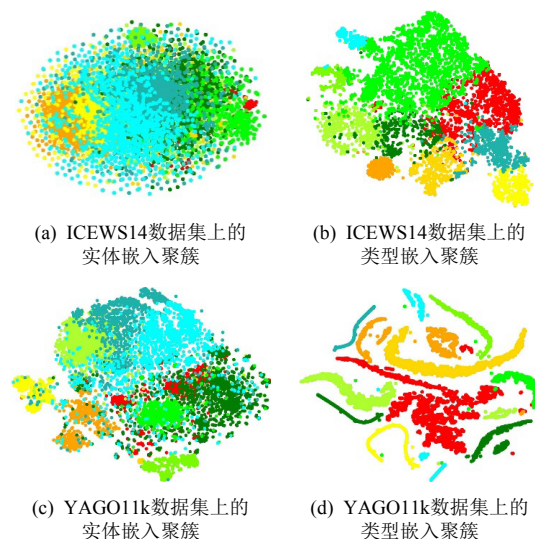


Fig. 3 The visualization of entity and type embedding clustering

图 3 实体和类型嵌入聚簇的可视化

从图 3 中可以看出，类型嵌入的簇比实体嵌入的簇更紧凑，且簇与簇之间的区分更明显，而实体嵌入却没有这种明显的聚簇现象。说明类型嵌入确实能够有效捕获实体的一般语义特征，学习实体中隐含

的类型信息.

3.7 参数敏感性分析

为了研究 T-Temp 模型中的参数对其性能的影响, 本文在 ICEWS14 数据集上对一些重要的超参进行了细粒度的分析和比较, 包括 2 个嵌入向量空间的维度 d 和 k 以及训练总轮数 N . 令嵌入向量空间的维度分别为 $d \in \{1\ 600, 1\ 800, 2\ 000, 2\ 200, 2\ 400, 2\ 600, 2\ 800\}$ 和 $k \in \{10, 15, 20, 25, 30, 35, 40\}$, 训练总轮数为 $N \in \{30, 40, 50, 60, 70, 80, 90, 100\}$. 为了实验公平, 除了当前研究的超参外, 其余超参的设置与 3.3 节中的相同, 实验结果如图 4 所示.

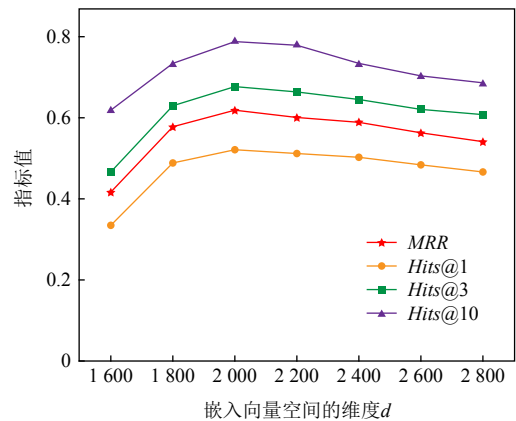
图 4(a) 显示了当维度为 d 的嵌入向量空间取不同维度值时 T-Temp 模型的各项评价指标变化趋势. 从图 4(a) 中可以看出, 当维度小于 2 000 时, 各项指标随维度的增加呈上升趋势, 并在维度取值为 2 000 时达到最大; 当维度大于 2 000 后, 各项指标随维度的增加呈缓慢下降趋势. 维度为 d 的嵌入向量空间用来表示实体的个体语义信息, 实验结果说明该向量空间的维度是个敏感参数, 当维度取值过小时, 可能造成欠拟合问题, 即向量空间无法充分表达所有实体的丰富语义信息; 而当维度取值过大时, 则可能出现过拟合现象, 从而导致性能变差.

图 4(b) 显示了当维度为 k 的嵌入向量空间取不同维度值时, T-Temp 模型的各项评价指标变化趋势. 从图中 4(b) 可以看出, 模型的各项指标同样随维度的增加先呈上升趋势, 然后在维度为 20 时达到最大值, 随后开始缓慢下降. 维度为 k 的嵌入向量空间用来表示实体所属的类型信息, 实验结果同样说明该向量空间的维度是个敏感参数, 其取值过小或过大可能造成模型的欠拟合或过拟合问题. 此外, 实体的类型特征所包含的一般语义信息通常没有实体的个体语义信息丰富, 因此实验中 2 个嵌入向量空间的最佳取值维度有 $k \ll d$.

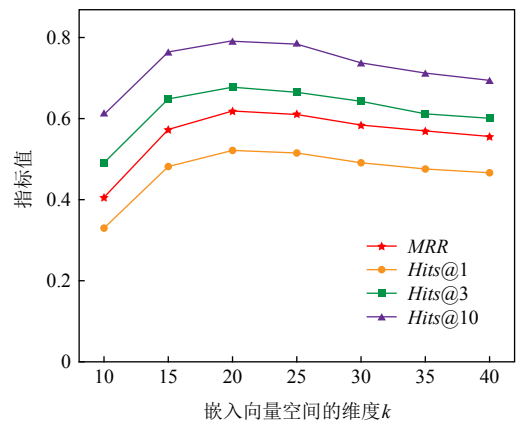
图 4(c) 显示了当训练总轮数 N 取不同值时, T-Temp 模型的各项评价指标变化趋势. 从图 4(c) 中可以看出, 当训练轮数增加时, 模型的各评价指标均有所提升. 而当训练轮数进一步增加时, 各项指标逐渐趋于平稳, 模型达到一个相对稳定的状态.

3.8 案例研究

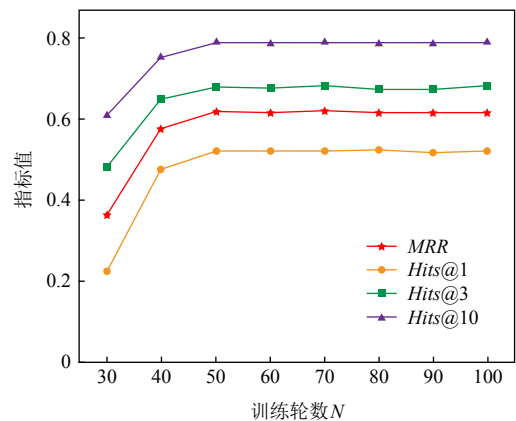
为了更细致地展示 T-Temp 模型确实可以利用时间信息来提高链接预测任务的准确性, 本文从 YAGO11k 数据集中抽取一些带有时间信息的四元组作为典型案例进行研究, 包括: (Ashley Cole, plays for, Arsenal, [1999—2006]) (Ashley Cole, plays for, Chelsea,



(a) 维度 d 对指标的影响



(b) 维度 k 对指标的影响



(c) 训练轮数 N 对指标的影响

Fig. 4 The sensitivity analysis of parameters

图 4 参数敏感性分析

[2006—2014]) (Ashley Cole, plays for, A.S. Roma, [2014—2016]) (Ashley Cole, plays for, LA Galaxy, [2016—2019]). 这些四元组的头实体和关系均是 “Ashley Cole” 和 “Playsfor”, 而尾实体则随着时间的不同而有所不同, 说明该球员在不同时期曾服务于不同的足球俱乐部.

实验分为 2 种情况进行对比: 1) 掩去四元组中的

尾实体,在已知头实体、关系和时间戳的情况下通过模型来预测尾实体,即回答问题:(Ashley Cole, plays for, ?, [1999—2006])(Ashley Cole, plays for, ?, [2006—2014])(Ashley Cole, plays for, ?, [2014—2016])(Ashley Cole, plays for, ?, [2016-2019]); 2)是将四元组中的尾实体和时间戳均掩去,通过训练好的模型来预测尾实体,即回答问题(Ashley Cole, plays for, ?).模型根据得分函数对所有候选实体组成的元组计算得分,并按照得分高低依次进行排序,实验结果如表6所示.

Table 6 Comparison of Link Prediction
表 6 链接预测对比

问题	预测结果排序
(Ashley Cole, plays for, ?, [1999—2006])	Arsenal , A.S. Roma
(Ashley Cole, plays for, ?, [2006—2014])	Chelsea , England national team
(Ashley Cole, plays for, ?, [2014—2016])	A.S. Roma , Chelsea
(Ashley Cole, plays for, ?, [2016—2019])	LA Galaxy , Crystal Palace
(Ashley Cole, plays for, ?)	A.S. Roma, England national team, Chelsea

注：正确的尾实体用黑色标识.

从表6中可以看出,在回答有具体时间范围限制的问题时,模型预测结果排名第一的候选实体均为正确答案,说明 T-Temp 模型确实可以利用可提供的时间信息实现更精准的链接预测.而回答不带有时间戳的三元组问题时,模型则给出了若干个候选实体作为答案.这些候选实体虽然无法给出用户精准的答案,但得益于 T-Temp 模型的类型限制作用,他们都属于“足球俱乐部”这一类型,一定程度上满足了用户的需求.

4 总结与展望

时态知识图谱表示学习是近几年的研究热点之一.本文基于张量的典范多元分解技术,提出了一种类型增强的时态知识图谱表示学习模型 T-Temp,用于解决时态知识图谱补全和语义计算问题,并证明了该模型在理论上具有完全表达性和较低的时空消耗. T-Temp 模型在表示学习的过程中不仅可以利用可提供的各种形式的时间信息,还可以利用实体和关系间的类型兼容性,自动学习实体、关系、时间和类型的嵌入表示,不需要提供额外的类型信息.在4个公开的时态知识图谱数据集上进行的大量实验结

果证明了 T-Temp 模型的有效性.

鉴于现有的时态知识图谱中,与事实相关联的时间戳存在大量的缺失现象,下一步,计划探索 T-Temp 模型对缺失的时间戳进行预测和补全的问题.另外,如何将 T-Temp 模型扩展到开放世界假设^[51],实现对未来事件的推理和预测,也是值得深入研究的方向.

作者贡献声明:何鹏提出了算法思路,完成实验并撰写论文;周刚、陈静、章梦礼、宁原隆提出指导意见并修改论文.

参 考 文 献

- [1] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge [C] //Proc of the 2008 ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2008: 1247–1250
- [2] Auer S, Bizer C, Kobilarov G, et al. Dbpedia: A nucleus for a web of open data [C] //Proc of the 6th Int Semantic Web. Berlin: Springer, 2007: 722–735
- [3] Miller G A. WordNet: A lexical database for English[J]. *Communications of the ACM*, 1995, 38(11): 39–41
- [4] Wang Xiang, He Xiangnan, Cao Yixin, et al. Kgat: Knowledge graph attention network for recommendation [C] //Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 950–958
- [5] Huang Xiao, Zhang Jingyuan, Li Dingcheng, et al. Knowledge graph embedding based question answering [C] //Proc of the 12th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2019: 105–113
- [6] Xiong Chenyan, Power R, Callan J. Explicit semantic ranking for academic search via knowledge graph embedding [C] //Proc of the 26th Int Conf on World Wide Web. New York: ACM, 2017: 1271–1279
- [7] Zhang Zhengyan, Han Xu, Liu Zhiyuan, et al. ERNIE: Enhanced language representation with informative entities [C] //Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 1441–1451
- [8] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase[J]. *Communications of the ACM*, 2014, 57(10): 78–85
- [9] Suchanek F M, Kasneci G, Weikum G. YAGO: A core of semantic knowledge [C] //Proc of the 16th Int Conf on World Wide Web. New York: ACM, 2007: 697–706
- [10] Ward MD, Beger A, Cutler J, et al. Comparing GDELT and ICEWS event data[J]. *Analysis*, 2013, 21(1): 267–297
- [11] Leetaru K, Schrodt P A. GDELT: Global data on events, location, and tone, 1979–2012 [C] //Proc of the 54th Int Studies Association

- Annual Convention. University Park, PA: Pennsylvania State University, 2013: 1–49
- [12] Kazemi S M, Poole D. Simple embedding for link prediction in knowledge graphs [C] //Proc of the 32nd Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2018: 4289–4300
- [13] Shen Yuming, Du Jianfeng. Temporal knowledge graph completion: Methods and progress[J]. Big Data Research, 2021, 7(3): 30–41 (in Chinese)
(申宇铭, 杜剑峰. 时态知识图谱补全的方法及其进展[J]. 大数据, 2021, 7(3): 30–41)
- [14] Zhang Tiancheng, Tian Xue, Sun Xianghui, et al. Overview of research on knowledge graph embedding technology [J/OL]. Journal of Software, 2022[2021-12-09]. <http://www.jos.org.cn/jos/article/abstract/6429> (in Chinese)
(张天成, 田雪, 孙相会, 等. 知识图谱嵌入技术研究综述[J/OL]. 软件学报, 2022[2021-12-09]. <http://www.jos.org.cn/jos/article/abstract/6429>(未发表))
- [15] Wang Quan, Mao Zhendong, Wang Bin, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724–2743
- [16] Ji Shanxiong, Pan Shirui, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494–514
- [17] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data [C] //Proc of the 26th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013: 2787–2795
- [18] Wang Zhen, Zhang Jianwen, Feng Jianlin, et al. Knowledge graph embedding by translating on hyperplanes [C] //Proc of the 28th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2014: 1112–1119
- [19] Lin Yankai, Liu Zhiyuan, Sun Maosong, et al. Learning entity and relation embeddings for knowledge graph completion [C] //Proc of the 29th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2015: 2181–2187
- [20] Ji Guoliang, He Shizhu, Xu Liheng, et al. Knowledge graph embedding via dynamic mapping matrix [C] //Proc of the 53rd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2015: 687–696
- [21] Sun Zhiqing, Deng Zhihong, Nie Jianyun, et al. Rotate: Knowledge graph embedding by relational rotation in complex space [C/OL] //Proc of the 7th Int Conf on Learning Representations. La Jolla, CA: ICLR, 2018[2021-12-09]. <https://openreview.net/forum?id=HkgEQnRqYQ>
- [22] Zhang Zhanqiu, Cai Jianyu, Zhang Yongdong, et al. Learning hierarchy-aware knowledge graph embeddings for link prediction [C] //Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 3065–3072
- [23] Hitchcock F L. The expression of a tensor or a polyadic as a sum of products[J]. Journal of Mathematics and Physics, 1927, 6(1/2/3/4): 164–189
- [24] Lacroix T, Usunier N, Obozinski G. Canonical tensor decomposition for knowledge base completion [C] //Proc of the 35th Int Conf on Machine Learning. Cambridge, MA: MIT Press, 2018: 2863–2872
- [25] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction [C] //Proc of the 33rd Int Conf on Machine Learning. Cambridge, MA: MIT Press, 2016: 2071–2080
- [26] Yang Bishan, Yih W, He Xiaodong, et al. Embedding entities and relations for learning and inference in knowledge bases [C/OL] //Proc of the 3rd Int Conf on Learning Representations. La Jolla, CA: ICLR, 2015[2021-12-09]. <https://arxiv.org/abs/1412.6575>
- [27] Wang Peiyan, Duan Lei, Guo Zhengshan, et al. Knowledge hypergraph link prediction model based on tensor decomposition[J]. Journal of Computer Research and Development, 2021, 58(8): 1599–1611 (in Chinese)
(王培妍, 段磊, 郭正山, 等. 基于张量分解的知识超图链接预测模型[J]. 计算机研究与发展, 2021, 58(8): 1599–1611)
- [28] Leblay J, Chekol M W. Deriving validity time in knowledge graph [C] //Proc of the 27th Web Conf 2018. New York: ACM, 2018: 1771–1776
- [29] Garcia-Duran A, Dumančić S, Niepert M. Learning sequence encoders for temporal knowledge graph completion [C] //Proc of the 2018 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2018: 4816–4821
- [30] Dasgupta S S, Ray S N, Talukdar P. Hyte: Hyperplane-based temporally aware knowledge graph embedding [C] //Proc of the 2018 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2018: 2001–2011
- [31] Cui Yuanning, Li Jing, Shen Li, et al. Duration-HyTE: A time-aware knowledge representation learning method based on duration modeling[J]. Journal of Computer Research and Development, 2020, 57(6): 1239–1251 (in Chinese)
(崔员宁, 李静, 沈力, 等. Duration-HyTE: 基于持续时间建模的时间感知知识表示学习方法[J]. 计算机研究与发展, 2020, 57(6): 1239–1251)
- [32] Xu C, Nanyeri M, Alkhoury F, et al. TeRo: A time-aware knowledge graph embedding via temporal rotation [C] //Proc of the 28th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2020: 1583–1593
- [33] Sadeghian A, Armandpour M, Colas A, et al. ChronoR: Rotation based temporal knowledge graph embedding [C] //Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 6471–6479
- [34] Goel R, Kazemi S M, Brubaker M, et al. Diachronic embedding for temporal knowledge graph completion [C] //Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 3988–3995
- [35] Bamler R, Mandt S. Dynamic word embeddings [C] //Proc of the 34th Int Conf on Machine Learning. Cambridge, MA: MIT Press, 2017: 380–389
- [36] Ma Yunpu, Tresp V, Daxberger E A. Embedding models for episodic knowledge graphs[J]. Journal of Web Semantics, 2019, 59: 90–99
- [37] Balažević I, Allen C, Hospedales T. Tucker: Tensor factorization for

- knowledge graph completion [C] //Proc of the 2018 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2019: 5185–5194
- [38] Tucker L R. Some mathematical notes on three-mode factor analysis[J]. *Psychometrika*, 1966, 31(3): 279–311
- [39] Lacroix T, Obozinski G, Usunier N. Tensor decompositions for temporal knowledge base completion [C/OL] //Proc of the 2020 Int Conf on Learning Representations. La Jolla, CA: ICLR, 2020[2021-12-09]. <https://openreview.net/forum?id=rke2P1BFwS>
- [40] Xie Ruobing, Liu Zhiyuan, Sun Maosong. Representation learning of knowledge graphs with hierarchical types [C] //Proc of the 25th Int Joint Conf on Artificial Intelligence. Amsterdam: Elsevier, 2016: 2965–2971
- [41] Hao Junheng, Chen Muhao, Yu Wenchao, et al. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts [C] //Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 1709–1719
- [42] Cui Zijun, Kapanipathi P, Talamadupula K, et al. Type-augmented relation prediction in knowledge graphs [C] //Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 7151–7159
- [43] Jain P, Kumar P, Chakrabarti S. Type-sensitive knowledge base inference without explicit type supervision [C] //Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 75–80
- [44] Kruskal J B. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics[J]. *Linear Algebra and Its Applications*, 1977, 18(2): 95–138
- [45] Håstad J. Tensor rank is NP-complete [C] // Proc of the 16th Int Colloquium on Automata, Languages and Programming. Berlin: Springer, 1989: 451–460
- [46] Kruskal J B. Rank, Decomposition, and Uniqueness for 3-way and N -Way Arrays [M] //Multiway Data Analysis. Amsterdam: Elsevier, 1989: 7–18
- [47] Paszke A, Gross S, Chintala S, et al. PyTorch: An imperative style, high-performance deep learning library [C] //Proc of the 33rd Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2019: 8024–8035
- [48] Kingma D P, Ba J. Adam: A method for stochastic optimization [C/OL] // Proc of the 3rd Int Conf on Learning Representations. Amherst, MA: UMASS, 2015 [2021-12-01]. <https://arxiv.org/abs/1412.6980>
- [49] MacQueen J. Some methods for classification and analysis of multivariate observations [C] //Proc of the 5th Berkeley Symp on Mathematical Statics and Probability. Oakland, CA: University of California Press, 1967: 281–297
- [50] Maaten L, Hinton G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9(11): 2579–2605
- [51] Shi B, Weninger T. Open-world knowledge graph completion [C] // Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2018: 1957–1964



He Peng, born in 1983. PhD, lecturer. Her main research interests include knowledge graph and big data analysis.

何 鹏, 1983 年生. 博士, 讲师. 主要研究方向为知识图谱和大数据分析.



Zhou Gang, born in 1974. PhD, professor. His main research interests include big data analysis, knowledge graph, and mass-data processing.

周 刚, 1974 年生. 博士, 教授. 主要研究方向为大数据分析、知识图谱和海量数据处理.



Chen Jing, born in 1990. PhD candidate, lecturer. Her main research interests include NLP, social network analysis, and knowledge engineering.

陈 静, 1990 年生. 博士研究生, 讲师. 主要研究方向为自然语言处理、社交网络分析和知识工程.



Zhang Mengli, born in 1993. PhD. His main research interests include NLP and text summarization.

章梦礼, 1993 年生. 博士. 主要研究方向为自然语言处理和文本摘要.



Ning Yuanlong, born in 1997. Master. His main research interests include knowledge graph and data mining.

宁原隆, 1997 年生. 硕士. 主要研究方向为知识图谱和数据挖掘.