

基于扩散概率分布的时序知识图谱推理*

周光有, 李鹏飞, 谢鹏辉, 罗昌银

(华中师范大学 计算机学院, 湖北 武汉 430079)

通信作者: 周光有, E-mail: gyzhou@mail.ccnu.edu.cn



摘要: 时序知识图谱推理旨在补充知识图谱中缺失的链接(事实), 其中每个事实都与时间戳进行绑定. 基于变分自动编码器的动态变分框架在这项任务中显示出独特的优势. 通过将实体和关系基于高斯分布进行联合建模, 该方法不仅具备很强的可解释性, 而且解决了复杂的概率分布问题. 然而, 传统的变分自动编码器方法在训练过程中容易出现过拟合问题, 从而不能精确捕捉实体语义的演化过程. 为了解决这个问题, 提出基于扩散概率分布的时序知识图谱推理模型. 具体来讲, 建立一个双向的迭代过程, 将实体语义建模过程分为多个子模块. 其中, 每个子模块通过一个正向的加噪变换和反向的高斯采样组成, 负责建模实体语义的一个微小演变过程. 相对基于变分自动编码器的方法, 通过多个子模块联合建模显示地学习度量空间中实体语义随时间的动态表示, 能够得到更为精确的建模. 与基于变分自动编码器的方法相比, 对于评估指标 *MRR*, 模型在 Yago11k 数据集和 Wikidata12k 数据集分别提高 4.18% 和 1.87%, 在 ICEWS14 和 ICEWS05-15 数据集上分别提高 1.63% 和 2.48%.

关键词: 知识图谱; 时序知识图谱; 链接预测; 扩散模型

中图法分类号: TP18

中文引用格式: 周光有, 李鹏飞, 谢鹏辉, 罗昌银. 基于扩散概率分布的时序知识图谱推理. 软件学报, 2024, 35(11): 5083-5097. <http://www.jos.org.cn/1000-9825/7002.htm>

英文引用格式: Zhou GY, Li PF, Xie PH, Luo CY. Temporal Knowledge Graph Reasoning Based on Diffusion Probability Distribution. Ruan Jian Xue Bao/Journal of Software, 2024, 35(11): 5083-5097 (in Chinese). <http://www.jos.org.cn/1000-9825/7002.htm>

Temporal Knowledge Graph Reasoning Based on Diffusion Probability Distribution

ZHOU Guang-You, LI Peng-Fei, XIE Peng-Hui, LUO Chang-Yin

(School of Computer, Central China Normal University, Wuhan 430079, China)

Abstract: Temporal knowledge graph reasoning aims to fill in missing links or facts in knowledge graphs, where each fact is associated with a specific timestamp. The dynamic variational framework based on variational autoencoder is particularly effective for this task. By jointly modeling entities and relations using Gaussian distributions, this method not only offers high interpretability but also solves complex probability distribution problems. However, traditional variational autoencoder-based methods often suffer from overfitting during training, which limits their ability to accurately capture the semantic evolution of entities over time. To address this challenge, this study proposes a new temporal knowledge graph reasoning model based on a diffusion probability distribution approach. Specifically, the model uses a bi-directional iterative process to divide the entity semantic modeling process into multiple sub-modules. Each sub-module uses a forward noisy transformation and a backward Gaussian sampling to model a small-scale evolution process of entity semantics. Compared with the variational autoencoder-based method, this study can obtain more accurate modeling by learning the dynamic representation of entity semantics in the metric space over time through the joint modeling of multiple submodules. Compared with the variational autoencoder-based method, the model improves by 4.18% and 1.87% on the Yago11k dataset and Wikidata12k dataset for evaluating the

* 基金项目: 国家自然科学基金 (61972173); 中央高校基本科研业务费 (CCNU22QN015); 武汉市知识创新专项基础研究项目 (2022010801010278)

收稿时间: 2023-02-28; 修改时间: 2023-04-25; 采用时间: 2023-07-10; jos 在线出版时间: 2023-11-15

CNKI 网络首发时间: 2023-11-17

MRR of the indicator and by 1.63% and 2.48% on the ICEWS14 and ICEWS05-15 datasets, respectively.

Key words: knowledge graph (KG); temporal knowledge graph (TKG); link prediction; diffusion model

知识图谱 (knowledge graph, KG) 是一种结构化的语义网络, 它以三元组 (头实体, 关系, 尾实体) 的形式将自然界中复杂的知识体系进行清晰的表达和灵活的建模^[1], 已成为当下研究的重点.

作为知识图谱的一个重要组成部分, 知识图谱表示学习是知识获取与应用的前提^[2]. 关于知识图谱表示学习, 开始是通过资源描述框架 (resource description framework, RDF)、网络本体语言 (Web ontology language, OWL) 等语言来描述的. 随着深度学习的迅速崛起, 人们开始尝试更简单的方法, 于是向量被引入进来. 因此, 当下的知识图谱表示学习的研究目标便是指通过机器学习的方式将图片、视频、文本等语义信息通过低维稠密的实体向量来表示^[3]. 基于向量的知识图谱表示学习方法给研究带来诸多便利. 例如, 用向量来表示知识图谱中的内容时, 实体和关系实现了数值化, 使得复杂任务简单化, 显著提高了运算效率. 随着研究不断深入, 嵌入方式不仅更加准确高效, 而且灵活多变, 适应能力更强^[4].

然而, 先前的很多研究都是基于静态知识图谱所提出的, 但随着时间的推移, 实体与关系会发生变化^[5]. 一些新的实体可能会被添加进 KGs; 一些旧的实体可能会从 KGs 中消失; 新的实体之间会产生新的关系. 如图 1 所示, 姚明在 2008 年是一名篮球运动员, 他效力于休斯顿火箭队, 其中一名队友是麦迪; 2012 年的时候姚明已经退役并前往上海交通大学读书, 而此时他的前队友麦迪也转投亚特兰大老鹰队; 而在 2020 年的时候, 姚明任职中国篮协主席. 这说明随着时间的推移, 实体的语义以及不同实体间的关系都可能发生变化. 在这种情况下, 当在进行查询时, 涉及时间点或者时间间隔时候, 基于静态知识图谱表示学习的模型便会暴露很多缺陷.

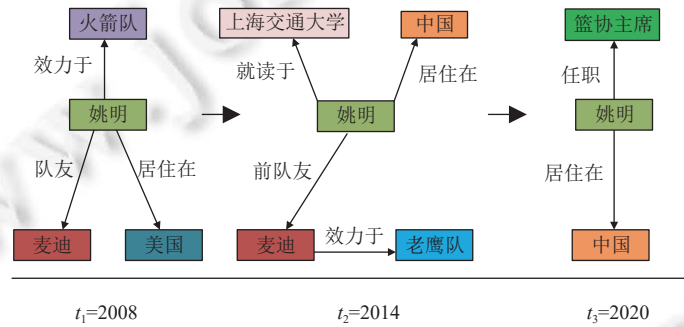


图 1 时序知识图谱的一个例子

按照功能划分, 时序知识图谱 (temporal knowledge graph, TKG) 可分为两大类, 分别为插值 (interpolation) 和外推 (extrapolation). 基于插值方法针对时间范围为 $[t_0, t_T]$ 之间缺失的事实进行补全; 基于外推方法针对时间范围为 $[t_T, \infty]$ 的未来事实进行预测. 以图 1 为例, 在对时间戳范围为 $[2008, 2020]$ 的四元组事实进行学习之后, 查询 (姚明, 居住于, ?, 2009) 所处时间段在已经学习的时间范围之内, 该任务是基于插值的方法; 查询 (姚明, 居住于, ?, 2021) 所处时间段在已经学习的时间范围之后, 该任务是基于外推的方法. 由此可以看到, 基于插值的方法侧重于从已知事实进行学习, 从而预测缺失的事实; 基于外推的方法侧重于学习过去时间戳发生事实所包含的结构信息, 以此作为依据对未来做出准确的预测. 相对而言, 基于外推的方法更具有挑战性^[6].

基于外推的方法面临的主要挑战是如何有效捕捉实体随时间的语义演化过程, 从而对未来事实做出精准的预测. 针对外推任务所面临的挑战, 先前学者做了很多研究来进行探索. Trivedi 等人^[4]提出了 Know-evolve 及其扩展 DyRep 模型, 通过使用循环神经网络对连续时间进行建模, 但是这种方法无法捕捉结构信息, 不能处理并发事件. Zhu 等人^[7]提出了 CyGNet 模型, 在训练过程中将历史重复事实全都添加到历史词汇表中, 从而利用已经发生的事实去预测未来, 但是这种对历史记录进行单独编码的方式不仅降低了效率, 而且无法做到对于没有先前事件的未来事实进行预测. Jin 等人^[8]提出了 RE-NET 模型, 使用循环神经网络来捕获并顺序编码实体预测中实体进行直接参与的历史事实, 但是它并没有学习实体语义的演化过程. Liao 等人^[9]提出了 DBKGE 模型, 它基于变分自动

编码器 (variational auto-encoder, VAE), 利用潜在变量, 通过设置高斯转移分布来动态地捕捉实体语义随时间的演化过程. 然而, 这种基于高斯分布进行建模的方式在训练过程中容易出现过拟合现象, 导致无法精准捕捉实体的语义演化过程.

为了解决 DBKGE 模型存在的问题, 本文提出了基于扩散概率分布的时序知识图谱推理模型 (temporal knowledge graph inference model based on diffusion probability distribution, DPMTKG). 如图 2 所示, X 表示未知概率分布的一个 TKG 事实 (即对于给定四元组的初始嵌入, 通过扩散模型去得到最终嵌入表示), Z_1, Z_2, \dots, Z_t 代表经过加噪或采样得到的潜在变量. 对于按照时间顺序排列的 TKG 事实 X , 本文通过迭代的方式向 X 添加高斯噪声, 然后得到潜在变量 Z_1, Z_2, \dots, Z_t . 这些潜在变量的数据分布逐渐向高斯分布靠拢 (例如, Z_t 服从高斯分布), 它们将实体语义建模的过程分成多个子模块, 共同建模 TKG 事实. 在反向过程中, 从 Z_t 开始, 每个子模块依次建立起一个高斯采样过程, 目的是捕捉实体语义随时间变化的过程. 通过正向和反向两个过程, 本文建立起一个马尔可夫链, 通过多个子模块联合建模显式地学习度量空间中实体语义随时间的动态表示. 不同于 DBKGE, 本文将建模过程分为多个子模块, 其中每个子模块均由一个加噪和与之对应的高斯采样组成. 通过这种方式, 每个子模块仅负责建模实体语义的一个微小演化过程. 在这种情况下, 使用高斯分布便可实现对每个子模块足够精确的建模, 从而解决了 DBKGE 存在的过拟合问题. 经过大量实验, 本文证实了该方法的有效性. 本文工作的主要贡献如下.

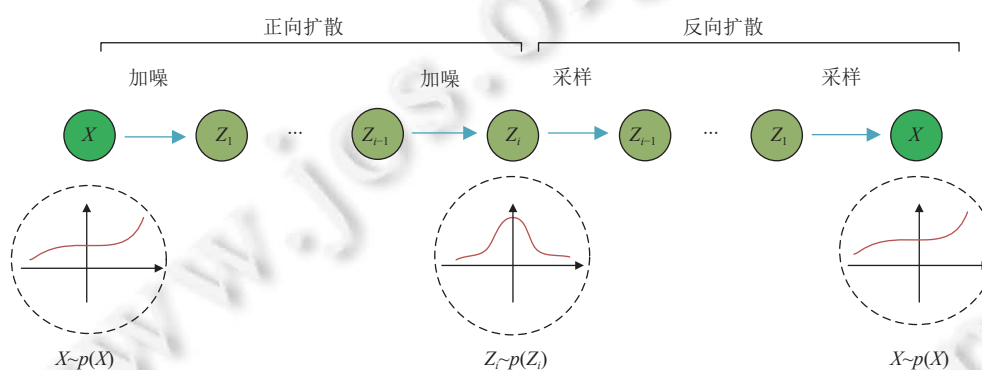


图 2 基于扩散模型的生成过程

(1) 对于按照时间顺序依次到达的序列, 本文将 t 时刻的四元组数据 (即 t 时刻的四元组的初始嵌入) 作为前向扩散输入部分. 在前向扩散阶段, 不断向四元组中添加噪声. 在这个迭代过程中, 后一步数据均由前一步添加噪声之后生成. 在经过多次迭代之后, 四元组的分布被逐渐引入的噪声污染, 直到成为完全随机噪声 (如图 2 所示).

(2) 在反向扩散过程中, 将前向扩散得到的数据作为输入, 并在每个时间步进行采样, 逐步去除噪声. 在这个迭代过程中, 通过训练学习得到每一步采样所需的均值和方差, 并逐步建立起噪声分布到数据分布之间的关系, 从而得到下一时刻的四元组嵌入. 此外, 为了提升采样速度, 对采样序列取任意子序列来进行采样.

(3) 本文在 4 个公开的数据集上进行了大量实验, 实验结果表明本文提出的方法优于基于变分自动编码器的方法. 具体来说, 对于评估指标 MRR , 模型在 Yago11k 和 Wikidata12k 分别提高了 4.18% 和 1.87%, 在 ICEWS14 和 ICEWS05-15 数据集上分别提高了 1.63% 和 2.48%.

1 相关工作

传统知识图谱表示学习不考虑时间信息, 它通过建立相关模型, 将知识图谱当中的关系与实体映射至连续的向量空间中, 捕捉并学习当中潜在的语义信息^[10]. 在现实世界中, 实体的语义信息往往随着时间发生变化, 而传统知识图谱表示学习并没有考虑实体的时间特征. 不同于传统知识图谱表示学习, 时序知识图谱表示学习将时间特征进行融合到表示学习中, 在链接预测时表现出比传统时序图谱更好的性能. 时序知识图谱表示学习主要分为两大类-插值 (interpolation) 和外推 (extrapolation).

插值方法针对时间范围为 $[t_0, t_T]$ 之间缺失的事实进行预测. 先前有许多学者将静态 KG 模型扩展到时序 KG, 并且重新设计了嵌入模块和与时间相关的评分函数. Leblay 等人^[11]将时间与关系融合后进行编码, 从而将时间转换为影响实体与关系嵌入的系数. 与 TTransE 类似, García-Durán 等人^[12]也将时间与关系进行融合, 他们提出了 TA-DistMult 模型, 通过循环神经网络去捕捉关系类型的时间感知表示. Dasgupta 等人^[13]提出了 HyTE 模型, 他们通过将时间进行分割, 将三元组投影到特定时间所对应的超平面上. 为捕获特定时间的实体特征, Goel 等人^[14]提出了 De-simple 模型, 他们将实体与时间同时进行输入并结合历史嵌入函数, 以得到每一个时间所对应的实体的一个隐藏的表示. Xu 等人^[15]提出了 ATISE 模型, 将时间信息分解后融入实体和关系的嵌入当中, 并且利用高斯分布的协方差去解决时间不确定性问题. Xu 等人^[16]提出了 TeRo 模型, 通过利用复杂空间中的旋转去映射实体随时间的演化过程. Ma 等人^[17]提出了 BiLSTM 模型, 首先将构造了一个时间字典, 然后根据构造的时间字典将分解获得的时间信息转为相应的序列, 然后将时间嵌入于关系嵌入组合以获得时间关系序列. Han 等人^[18]首先致力于 TKG 双曲嵌入的工作, 他们借助黎曼流形乘积中定义的速度矢量来捕捉实体随时间的演变过程, 从而来学习实体的嵌入表示. Xu 等人^[19]提出了 TeLM 模型, 该模型通过一个线性事件正则化器, 对 TKG 进行了四阶张量分解; Liu 等人^[20]将所有事实视为一个完整整体, 利用神经网络进行语义建模, 并利用目标事实去挑选最有用的上下文. Shao 等人^[21]提出了 TuckERT 模型, 提出时间可用三元组进行隐式表示, 并基于张量分解从三元组中的实体或关系中提取出时间信息.

外推方法针对时间范围为 $[t_T, \infty]$ 的未来事实进行预测. 先前的工作通过对历史事实进行建模来模拟实体的演化过程, 从而去预测未来的事实. Trivedi 等人^[4]提出了 Know-evolve 模型, 他们通过循环神经网络建模过去时间戳中发生的所有事实, 并尝试计算未来事实的条件概率. Zhu 等人^[7]提出了 CyGNet 模型, 他们在训练阶段使用复制网络对所有历史重复事实进行建模, 以使重复事实能够更好地预测未来时间戳中的事实. Li 等人^[22]通过建立基于图卷积网络的循环演化网络将序列的结构信息、顺序模式、静态属性全部进行整合, 以处理对将来的时间戳进行事实的预测问题. 以上这些方法缺乏对其预测的解释, 并且无法处理以前看不见的实体. Han 等人^[23]提出了 xERTE 模型, 通过子图采样技术建立了一个具有可解释性的推理框架. Sun 等人^[24]提出了 TITer 模型, 通过设计基于时间的强化学习模型来处理 and 预测未知的实体和事实. Deng 等人^[25]将所有事实视为一个完整整体, 利用神经网络进行语义建模, 并利用目标事实去挑选最有用的上下文. Li 等人^[26]提出了 CluSTeR 模型, 采用两阶段方法, 首先利用强化学习从历史事实搜索并归纳线索, 然后利用图卷积网络从线索中推理答案. Park 等人^[27]将事件时间与演化网络分别进行建模, 以联合概率乘积的形式将二者统一, 从而学习丰富的互补信息. Jung 等人^[28]提出了 T-GAP 模型, 融合了每个实体的邻域信息, 通过每个节点的注意力信息向可达到的节点进行路径搜索. Liao 等人^[9]提出了 DBKGE 模型, 他们首次提出借助基于动态贝叶斯的高斯分布去学习实体和关系的演化过程. 虽然 DBKGE 在时序知识图谱的链接预测方面取得了很好的结果, 但是它在训练过程中容易出现过拟合问题, 从而不能精确捕捉实体语义的演化过程. 针对这一问题, 本文建立了一个双向的迭代过程, 将实体语义建模分为多个子模块. 其中, 每个子模块通过一个正向的加噪变换和反向的高斯采样组成, 负责建模实体语义的一个微小演变过程. 不同于 DBKGE, 本文提出的 DPMTKG 通过多个子模块联合建模显式地学习度量空间中实体语义随时间的动态表示, 能够得到更为精确的建模.

2 DPMTKG 模型介绍

2.1 问题定义

定义 1 (时序知识图谱). 时序知识图谱是一种多关系有向图, 它是由不同时刻下的事实所组成的序列. 时序知识图谱中的每一个事实都可以用一个四元组 (s, r, o, t) 来进行表示. 其中 s 代表头实体, o 代表尾实体, r 代表头实体与尾实体之间的关系, t 代表该事实发生的时间. 此外, 头实体 s 和尾实体 o 属于集合 \mathcal{E} , 关系 r 属于集合 \mathcal{R} , 时间 t 属于集合 \mathcal{T} . 对于整个历史事实序列, 用 $X_t = (E_t, R_t)$ 表示 t 时刻的实体和关系的序列嵌入, 其中 E_t, R_t 分别代表 t ($1 \leq t \leq T$) 时刻的实体和关系的嵌入表示. 表 1 展示了主要符号.

定义 2 (时序知识图谱推理任务). 本文的主要任务从已知的历史序列 $X_{\leq T}$ 学习实体的语义变化情况, 并对 $T+1$ 时刻缺失的头实体 $(?, r, o, T+1)$ 或者尾实体 $(s, r, ?, T+1)$ 进行概率评估. 该过程表示如下:

$$p(X_{T+1}|X_{\leq T}) = F(X_{\leq T}),$$

其中, $F(\cdot)$ 代表由历史序列 $X_{\leq T}$ 到 $T+1$ 时刻序列嵌入的映射函数.

表 1 符号列表

符号	描述	符号	描述
\mathcal{E}	实体集合	t	时间
\mathcal{R}	关系集合	X_t	在时间 t 的序列嵌入
\mathcal{T}	时间集合	E_t	在时间 t 的实体嵌入
s	头实体	R_t	在时间 t 的关系嵌入
r	关系	Z_t^J	在第 J 步加噪得到的样本
o	尾实体		

2.2 模型概述

图 3 显示了本文所提出的 DPMTKG 模型的结构. DPMTKG 主要有两部分组成, 分别为基于加噪变换的前向扩散和基于高斯采样的反向扩散. 在前向扩散的过程中, 本文以迭代的方式不断向四元组中添加噪声. 在这个迭代过程中, 后一步数据均由前一步添加噪声之后生成. 在经过多次迭代之后, 四元组的分布被逐渐引入的噪声污染, 直到成为完全随机噪声. 在反向扩散的过程中, 本文使用基于高斯采样的方法来逆向地学习噪声分布与数据分布之间的关系, 从而通过学习得到下一时刻四元组的嵌入表示. 对于时间 t 的每个四元组, 评分函数如下:

$$f_t(s, r, o) = \|s_t + r_t - o_t\| \tag{1}$$

其中, s_t 、 o_t 和 r_t 分别表示在时间 t 时头实体、尾实体和关系的嵌入.

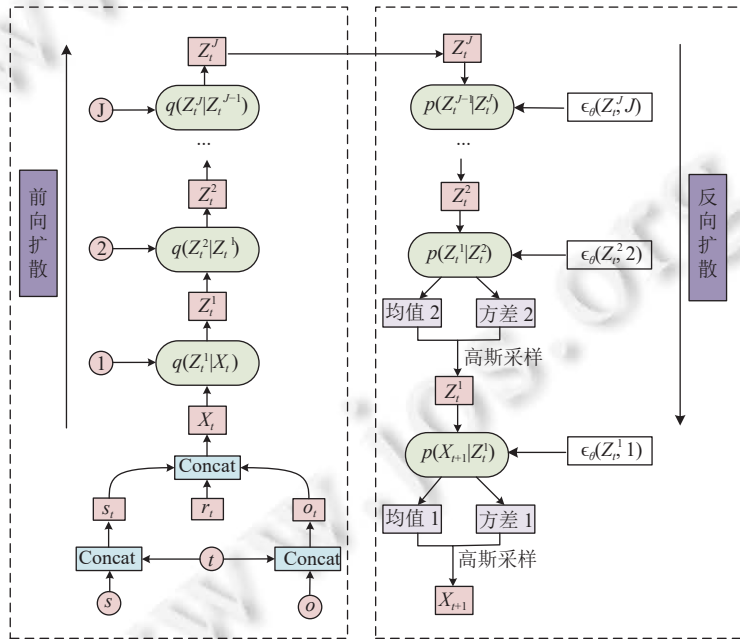


图 3 基于扩散概率分布的时序知识图谱推理模型

对于模型优化, 主要目标是使正样本损失不断接近于 0, 而负样本的损失尽可能大. 与之前方法类似^[16], 本文采用一个边界参数 γ 来控制损失范围. 在时间 t 发生的历史事实的概率分布如下:

$$p(X_t) = \sum_{(s,r,o,t) \in X_t} \sum_{(s',r,o') \in X_{t-}^c} [f_t(s,r,o) + \gamma - f_t(s',r,o')]_+ \quad (2)$$

在这里 X_{t-}^c 表示在时间 t 之前和时间 t 时出现的所有负样本的集合.

2.3 基于加噪变换的前向扩散

对于输入数据, 在得到实体和关系的嵌入后, 为了使实体具备时间特性, 将实体与时间的嵌入进行拼接, 计算方式如下所示:

$$s_t = W_1 [s; t] + b_1 \quad (3)$$

$$o_t = W_2 [o; t] + b_2 \quad (4)$$

其中, W_1, b_1, W_2, b_2 是可学习的参数, $[\cdot]$ 代表拼接.

为了方便接下来的介绍, 本文用 $X_t = [s_t, r_t, o_t]$ 来统一表示在 t 时刻四元组 (s, r, o, t) 的嵌入表示.

在对四元组进行建模时, 对于未知的数据分布, 首先通过前向扩散来打乱四元组的数据分布, 然后通过反向扩散进行结构恢复, 从而生成下一时刻四元组的嵌入表示. 在设置参数化的模型之后, 利用训练来不断缩小真实数据分布与生成数据分布的散度估计. 前向扩散的目标是不断向原始数据 X_t 中添加高斯噪声, 目的是使得原本有意义的分布逐渐接近高斯噪声分布^[29]. 在进行 J 步的噪声叠加操作之后, 依次得到的潜在变量为 $Z_t^1, Z_t^2, \dots, Z_t^J$. 具体而言, 前向扩散过程采用的是一个马尔可夫链 (Markov chain), 其中每一步的转移过程都是一个线性的高斯函数前向扩散过程定义如下所示:

$$q(Z_t^1, Z_t^2, \dots, Z_t^J | X_t) = q(Z_t^1 | X_t) \prod_{j=2}^J q(Z_t^j | Z_t^{j-1}) \quad (5)$$

其中, $q(Z_t^1 | X_t)$ 与 $q(Z_t^j | Z_t^{j-1})$ 是马尔可夫扩散核. 关于加噪变换的过程, Z_t^1 是由 X_t 添加噪声得到, 后面每一步均由其前一项添加噪声, 并且在下一时刻对数据分布进行建模. Z_t^1 以及 $Z_t^j (1 < j \leq J)$ 的条件概率分布定义如下所示:

$$q(Z_t^1 | X_t) = \sqrt{a_1} X_t + \sqrt{1-a_1} \epsilon_1 \quad (6)$$

$$q(Z_t^j | Z_t^{j-1}) = \sqrt{a_{j-1}} Z_t^{j-1} + \sqrt{1-a_{j-1}} \epsilon_j \quad (7)$$

其中, ϵ_1, ϵ_j 是服从高斯分布的高斯噪声, $\sqrt{a_1}, \sqrt{1-a_1}, \sqrt{a_{j-1}}, \sqrt{1-a_{j-1}}$ 为权重系数, 用于控制上一步数据与高斯噪声的占比. 当 j 不断增大时, X_t 逐渐接近纯噪声. 因此, 当 j 足够大的时候, Z_t^j 逐渐趋近收敛于标准正态分布 $\mathcal{N}(0, I)$. 因此, 随着前向扩散的推进, 为了模糊原始数据, 应该增加高斯噪声的占比. 因此, 随着 $j (1 \leq j \leq J)$ 的增加, $\sqrt{a_j}$ 应不断减小. 本文通过设置初始值与终止值并设计了与 $j (1 \leq j \leq J)$ 负相关的线性函数来解决这个问题. 对于 $a_j (1 \leq j \leq J)$, 计算如下:

$$a_j = c - \frac{j}{J}(c-d) \quad (8)$$

其中, $c (0 < c < 1)$ 为初始值, $d (0 < d < 1)$ 为终止值, 且满足 $c > d$. 由于每次加入的噪声都服从高斯分布并且它们之间是相互独立的, 因此在进行 j 步添加操作后, 相对于初始数据而言, 该步所添加的噪声依然服从高斯分布. 利用这一高斯分布的良好性质, 可以由初始输入数据以及 a_j 直接计算得到任意一步的潜在变量, 而不需要进行中间步骤的计算, 从而可以帮助加快前向扩散的速度. 由于每一步的数据均由上一步计算得到, 因此通过借助公式 (5)、公式 (6), 从输入数据开始进行迭代计算, 从而得到关于 $Z_t^j (1 < j \leq J)$ 与初始数据 X_t 的条件概率分布如下所示:

$$q(Z_t^j | X_t) = \tilde{a}_j X_t + \sqrt{1-\tilde{a}_j} \epsilon \quad (9)$$

$$\tilde{a}_j = \prod_{k=1}^j a_k \quad (10)$$

其中, ϵ 服从标准正态分布, \tilde{a}_j 代表由 X_t 直接得到 Z_t^j 的权重参数.

2.4 基于高斯采样的反向扩散

通过前向扩散, 对 t 时刻的数据不断进行加噪操作, 从而使得数据接近高斯分布. 而在反向扩散操作中, 目标

是进行去噪, 并生成 $t+1$ 时刻的数据分布 X_{t+1} . 因此, 本文通过反向扩散来近似前向扩散的逆过程, 从而实现去噪. 与前向扩散类似, 采用一个高斯核的反向马尔可夫链来生成下一时刻的历史数据分布; 不同之处在于在反向扩散过程中需要逐渐从之前得到的噪音中进行数据结构恢复, 这需要通过训练来完成, 因此高斯函数的均值和方差均被神经网络进行参数化. 具体而言, 本文从前向扩散最终得到的 Z_t^j 出发进行迭代采样. 反向扩散过程表述如下所示:

$$p_\theta(Z_t^1, Z_t^2, \dots, Z_t^J, X_{t+1}) = p(Z_t^J) p_\theta(X_{t+1}|Z_t^J) \prod_{j=2}^J p_\theta(Z_t^{j-1}|Z_t^j) \quad (11)$$

其中, $p(Z_t^j)$ 服从标准正态分布; θ 代表逆向扩散中相关的高斯分布参数, 它可以通过学习得到. 上述过程可以理解: 将 Z_t^j 作为输入, 然后去预测高斯分布的均值和方差, 然后基于预测的分布进行随机采样得到 Z_t^{j-1} . 在经过不断地进行预测和采样之后, 得到 $t+1$ 时刻的数据分布 X_{t+1} . $p_\theta(X_{t+1}|Z_t^1)$ 与 $p_\theta(Z_t^{j-1}|Z_t^j)$ ($2 \leq j \leq J$) 全都服从高斯分布:

$$p_\theta(X_{t+1}|Z_t^1) = \mathcal{N}(X_{t+1}; \mu_\theta(Z_t^1, 1), \sigma_\theta(Z_t^1, 1)I) \quad (12)$$

$$p_\theta(Z_t^{j-1}|Z_t^j) = \mathcal{N}(Z_t^{j-1}; \mu_\theta(Z_t^j, j), \sigma_\theta(Z_t^j, j)I) \quad (13)$$

其中, $\mu_\theta(Z_t^1, 1)$ 、 $\mu_\theta(Z_t^j, j)$ 与 $\sigma_\theta(Z_t^1, 1)$ 、 $\sigma_\theta(Z_t^j, j)$ 分别代表高斯分布的均值与方差, 均被神经网络进行参数化. 由于 $p_\theta(X_{t+1}|Z_t^1)$ 和 $p_\theta(Z_t^{j-1}|Z_t^j)$ 的均值与方差无法直接计算得出, 因此用 $q(Z_t^1|X_t)$ 与 $q(Z_t^{j-1}|Z_t^j, X_t)$ 进行替代计算. $q(Z_t^1|X_t)$ 可由公式 (5) 直接得到; 借助全概率公式, $q(Z_t^{j-1}|Z_t^j, X_t)$ 的计算如下所示:

$$q(Z_t^{j-1}|Z_t^j, X_t) = \frac{q(Z_t^j|Z_t^{j-1})q(Z_t^{j-1}|X_t)}{q(Z_t^j|X_t)} \quad (14)$$

其中, $q(Z_t^j|Z_t^{j-1})$ 可由公式 (6) 计算得到, $q(Z_t^{j-1}|X_t)$ 、 $q(Z_t^j|X_t)$ 可由公式 (9) 计算得到. 最终, $\mu_\theta(Z_t^1, 1)$ 、 $\mu_\theta(Z_t^j, j)$ 与 $\sigma_\theta(Z_t^1, 1)$ 、 $\sigma_\theta(Z_t^j, j)$ 的计算如下所示:

$$\mu_\theta(Z_t^1, 1) = \sqrt{a_1}X_t \quad (15)$$

$$\sigma_\theta(Z_t^1, 1) = \sqrt{1-a_1} \quad (16)$$

$$\mu_\theta(Z_t^j, j) = \frac{1}{\sqrt{a_j}} \left(Z_t^j - \frac{1-a_j}{\sqrt{1-\tilde{a}_j}} \epsilon_\theta(Z_t^j, j) \right) \quad (17)$$

$$\sigma_\theta(Z_t^j, j) = \frac{1-\tilde{a}_{j-1}}{1-\tilde{a}_j} (1-a_j) \quad (18)$$

其中, \tilde{a}_j 与 \tilde{a}_{j-1} 的计算可由公式 (10) 得到, $\epsilon_\theta(Z_t^j, j)$ 代表噪声.

随着 J 的不断增大, 反向扩散需要更多的迭代步骤. 对于拥有 J 个扩散训练步骤的模型, 原本需要进行 J 次采样迭代, 相应序列值为 $\{1, 2, \dots, J-1, J\}$. 为了提升模型性能, 本文遵循 Nichol 等人所提出的加速采样方法^[30]. 虽然目前已经有一些更快的采样方法被提出, 但是 Nichol 等人所提出的加速采样方法存在如下优点: 首先, Nichol 等人提出的加速采样方法简单易实现并且具有很好的可扩展性, 已经被应用于很多领域; 其次, Nichol 等人提出的加速采样方法在许多实际问题中已经被证明是高效稳定的. 因此, 本文采用 Nichol 等人提出的加速采样方法. 具体来说, 使用给定 n ($1 < n < J$) 值大小的任意子序列进行采样得到序列 $\{S_1, S_2, \dots, S_{n-1}, S_n\}$, 从而实现加速采样的目的.

2.5 损失函数

通过上述介绍, 本文展示了 DPMTKG 模型如何利用前向扩散、反向扩散生成下一时刻的实体和关系表示. 在进行优化的时候, 通常借助变分推断中的变分下界 (variational lower bound, VLB) 技巧进行推导, 相应的似然函数如下所示:

$$\begin{aligned} -\log(p_\theta(X_{t+1})) &\leq -\log(p_\theta(X_{t+1})) + KL(q(Z_t^1, Z_t^2, \dots, Z_t^J|X_t) || p_\theta(Z_t^1, Z_t^2, \dots, Z_t^J|X_{t+1})) \\ &= -\log(p_\theta(X_{t+1})) + E_{X_t} \left[\log \frac{q(Z_t^1, Z_t^2, \dots, Z_t^J|X_t)}{p_\theta(Z_t^1, Z_t^2, \dots, Z_t^J, X_{t+1})} + \log(p_\theta(X_{t+1})) \right] = E_{X_t} \left[\log \frac{q(Z_t^1, Z_t^2, \dots, Z_t^J|X_t)}{p_\theta(Z_t^1, Z_t^2, \dots, Z_t^J, X_{t+1})} \right] \quad (19) \end{aligned}$$

为了方便计算, 借助公式 (5)、公式 (11) 对上述公式进行继续转换:

$$\begin{aligned}
 E_{X_t} \left[\log \frac{q(Z_t^1, Z_t^2, \dots, Z_t^J | X_t)}{p_\theta(Z_t^1, Z_t^2, \dots, Z_t^J, X_{t+1})} \right] &= E_{X_t} \left[\log \frac{q(Z_t^1 | X_t) \prod_{j=2}^J q(Z_t^j | Z_t^{j-1})}{p(Z_t^1) p_\theta(X_{t+1} | Z_t^1) \prod_{j=2}^J p_\theta(Z_t^{j-1} | Z_t^j)} \right] \\
 &= E_{X_t} \left[-\log p(Z_t^1) + \log \frac{q(Z_t^1 | X_t)}{p_\theta(X_{t+1} | Z_t^1)} + \prod_{j=2}^J \log \frac{q(Z_t^j | Z_t^{j-1})}{p_\theta(Z_t^{j-1} | Z_t^j)} \right] \\
 &= E_{X_t} \left[-\log p(Z_t^1) + \log \frac{q(Z_t^1 | X_t)}{p_\theta(X_{t+1} | Z_t^1)} + \prod_{j=2}^J \log \left(\frac{q(Z_t^{j-1} | Z_t^j, X_t)}{p_\theta(Z_t^{j-1} | Z_t^j, X_t)} \cdot \frac{q(Z_t^j | X_t)}{q(Z_t^{j-1} | X_t)} \right) \right] \\
 &= E_{X_t} \left[-\log p(Z_t^1) + \log \frac{q(Z_t^1 | X_t)}{p_\theta(X_{t+1} | Z_t^1)} + \log \frac{q(Z_t^1 | X_t)}{q(Z_t^1 | X_t)} + \prod_{j=2}^J \log \left(\frac{q(Z_t^{j-1} | Z_t^j, X_t)}{p_\theta(Z_t^{j-1} | Z_t^j, X_t)} \right) \right] \\
 &= E_{X_t} \left[\log \frac{q(Z_t^1 | X_t)}{p_\theta(Z_t^1)} + \prod_{j=2}^J \log \left(\frac{q(Z_t^{j-1} | Z_t^j, X_t)}{p_\theta(Z_t^{j-1} | Z_t^j, X_t)} \right) - \log p_\theta(X_{t+1} | Z_t^1) \right] \quad (20)
 \end{aligned}$$

通过公式 (9)、公式 (11)、公式 (12), 可以将公式 (20) 中的每一项都计算出来. 最终损失函数为:

$$\mathcal{L}(X_{<T}) = E_{X_t} \left[\sum_{t=1}^T \log(p(X_t)) + \|\epsilon - \epsilon_\theta(\sqrt{a_t} X_t + \sqrt{1-a_t} \epsilon)\| \right] \quad (21)$$

其中, ϵ 代表前向扩散的噪声, ϵ_θ 代表反向扩散的噪声. 损失函数的第 1 项为基于翻译模型的打分函数, 第 2 项为基于语义扩散的打分函数. 其中, 损失函数的第 1 项通过最小化正样本与负样本的距离之和来优化参数, 第 2 项通过最大化边际似然的变分下界来优化参数.

本文采用 Adam 优化器^[31]训练模型, 其因为在训练过程中为了尽快通过学习获得高质量的四元组嵌入, 需要选择能够快速收敛的优化器. 而 Adam 优化器在进行更新的时候对于梯度的均值和方差进行了综合考虑, 不仅收敛速度快, 而且内存需求少、易调参^[25].

3 实验

3.1 数据集

本文在 4 个公开的 TKG 数据集进行了大量的链接预测实验. 本文所使用的 4 个数据集分别为 Yago11k 数据集、Wikidata12k 数据集、ICEWS14 数据集以及 ICEWS05-15 数据集. 其中, Yago11k 数据集是从 YAGO 数据集中提取最频繁的 10 个关系所得到的; Wikidata12k 数据集是从 Wikidata 数据集选取了前 24 个时间关系所提取的; ICEWS14 数据集是从 ICEWS 数据集中提取的 2014 年发生的所有事件; ICEWS05-15 数据集是从 ICEWS 数据集中提取的 2005–2015 年发生的所有事件.

本文采用与 Jin 等人^[8]相同的处理方式, 将这些数据集进行离散化. 此外, 本文按照时间顺序对数据集进行了排序并且分为了训练集、验证集和测试集, 占比情况分别为: 80%, 10%, 10%. 这样便有: $T_{\text{train}} < T_{\text{valid}} < T_{\text{test}}$. 表 2 列出了 4 个数据集的具体统计结果.

表 2 数据集统计

数据集	#实体	#实体间的关系	#训练集	#验证集	#测试集
Yago11k	10623	10	161540	19523	20026
Wikidata12k	12554	24	539286	67538	63110
ICEWS14	6869	230	74845	8514	7371
ICEWS05-15	10094	251	368868	46302	46159

3.2 评估指标

对于未来时间戳的链接预测任务旨在预测未来时间戳序列下四元组 (s, r, o, t) 缺失的头实体 s 或者尾实体 o .

在测试过程中, 对于一个需要测试的四元组来说, 以预测头实体为例, 将 (s, r, o, t) 中的 s 用知识图谱中的时间 t 之前未出现过的实体所替代, 然后利用得分函数计算分数, 并对所有分数进行排列. 这样便得到了对于一个测试集中的四元组的正确答案的排名, 显然, 排的越前说明越好. 本文使用常用的两种评价指标 MRR 与 $Hits@m$ 进行测试. MRR 是所有正确四元组排名倒数的平均值. $Hits@N$ 是所有正确四元组的前 N 个等级的比例. 在本实验中, 设置 $N = 1, 3, 10$.

在训练过程中, 本文利用 t 时刻及之前出现的实体随机替换四元组 (t 时刻) 的头实体或尾实体来生成负样本. 在测试过程中, 对于 t 时刻发生的事实四元组有可能作为损坏四元组参与排名, 这会影响到正确四元组的排名. 先前的一些模型, 例如 CyGNet^[7], 使用传统的静态过滤, 在不考虑时间因素的情况下删除了数据集中的所有四元组, 不适用于时序 KG. 因此, 继先前工作, 在测试过程中采用时间感知过滤设置^[28]. 时间感知过滤设置的作用是在所有的负样本中过滤掉在时间 t 发生的事实. 例如, 对于两个四元组 (姚明, 居住于, 美国, 2008 年) 以及 (姚明, 居住于, 中国, 2020 年), 四元组 (姚明, 居住于, 美国, 2008 年) 没有在 2020 年出现. 因此, 在 2020 年这个时间戳, 时间感知过滤会将 (姚明, 居住于, 美国, 2008 年) 过滤掉, 而传统的静态过滤会将二者都过滤掉.

3.3 参数设置

本文使用 NVIDIA GTX A5000 GPU 实现该模型. 在实验中, 使用 Adam 优化器^[31]来学习参数. 对于这两个数据集, 使用网格搜索策略调整参数以找到验证子集上的最佳值. 具体地, 调整嵌入维度 $dim = \{100, 150, 200, 250\}$, 学习率 $lr = \{0.00001, 0.00003, 0.00005, 0.00008, 0.0001\}$, 负样本数 $eta = \{100, 150, 200\}$, 训练过程中的批大小 $batch = \{256, 512, 1024\}$, 扩散迭代次数 $J = \{500, 1000, 1500, 2000, 2500\}$, 当 $J = 2000$, 采样次数 $n = \{200, 400, 600, 800, 1000\}$, 前向扩散的初始值 $c = \{0.8, 0.85, 0.9, 0.95, 1\}$ 和终止值 $d = \{0, 0.05, 0.1, 0.15, 0.2\}$. 表 3 展示了模型在 4 个数据集上的最佳设置.

表 3 参数最佳配置列表

符号	Yago11k	Wikidata12k	ICEWS14	ICEWS05-15
dim	200	250	200	200
lr	0.00008	0.00003	0.0001	0.00008
eta	150	150	200	200
$batch$	1024	1024	1024	1024
J	2000	2000	2000	2000
n	800	600	800	800
c	0.9	0.95	0.95	0.95
d	0.1	0.1	0.1	0.1

3.4 基线系统

本文将 DPMTKG 模型与两类基线进行比较: 传统 KGE 模型和时序 KGE 模型. 其中, 传统 TKG 模型中忽略了时间信息, 其代表模型包括 TransE^[32]、DistMult^[33]、R-GCN^[34]、ConvE^[35]、ConvTransE^[36]和 RotaE^[37]. 该部分实验结果来自 Li 等人^[22]. 时序 TKG 模型按照可分为插值和外插两部分. 基于插值的模型针对时间范围为 $[t_0, t_r]$ 之间缺失的事实进行预测, 其代表模型主要包括 TTransE^[11]、HyTE^[13]和 TA-DistMult^[12]. 该部分实验结果来自 Li 等人^[22]. 外推方法针对时间范围为 $[t_r, \infty)$ 的未来事实进行预测, 其代表模型主要包括 RE-NET^[8]、CyGNet^[7]和 TITer^[24]、DBKGE^[9]、EvoKG^[27]. 关于 RE-NET、CyGNet 和 TITer、EvoKG, 本文复现了代码并给出实验结果.

3.5 实验结果与分析

表 4 和表 5 展示了 Yago11k、Wikidata12k、ICEWS14 和 ICEWS05-15 数据集的实验结果. 基线被划分为 3 类, 其中第 1 类为静态方法; 第 2 类为基于插值的方法; 第 3 类为基于外推的方法. 从表 4 和表 5 中, 有以下 3 个重要发现.

表 4 在 Yago11k 和 Wikidata12k 数据集上时序知识图谱推理结果 (%)

Baseline	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	41.39	26.78	55.27	57.66	41.25	37.51	44.48	46.13
DistMult	44.05	—	49.70	59.94	27.96	—	32.45	39.51
R-GCN	20.25	—	24.01	37.30	13.96	—	15.75	20.25
ConvE	41.22	—	47.03	59.90	26.03	—	30.51	39.18
ConvTransE	46.67	—	52.22	62.52	30.89	—	34.30	41.45
RotaE	42.08	—	46.77	59.39	26.08	—	31.63	38.51
TTransE	26.10	—	36.28	47.73	20.66	—	23.88	26.10
TA-DistMult	44.98	—	50.64	61.11	26.44	—	31.36	38.97
HyTE	14.42	—	39.73	46.98	25.40	—	29.16	37.54
RE-NET	45.87	35.45	51.74	61.44	30.45	22.89	33.48	41.16
CyGNet	44.91	36.19	50.46	61.52	43.12	37.81	46.65	52.52
TITer	41.90	26.89	50.85	69.67	27.27	15.02	32.45	51.95
EvoKG	48.34	40.98	54.69	66.68	44.07	38.48	51.29	54.68
DBKGE	46.10	33.03	55.49	63.47	43.30	38.24	51.29	52.31
DPMTKG	50.28	42.19	56.34	66.26	45.17	39.67	51.80	55.24

表 5 在 ICEWS14 和 ICEWS05-15 数据集上时序知识图谱推理结果 (%)

Baseline	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	13.42	3.11	17.32	34.55	16.75	8.61	18.41	33.59
DistMult	20.32	6.13	27.59	46.61	19.91	5.63	27.22	47.33
R-GCN	28.03	19.42	31.95	44.83	27.13	18.83	30.41	43.16
ConvE	30.30	21.30	34.42	47.89	31.40	21.56	35.70	50.96
ConvTransE	31.50	22.46	34.98	50.03	30.28	20.79	33.80	44.95
RotaE	25.71	16.41	29.01	45.16	19.01	10.42	21.35	36.92
TTransE	12.86	3.14	15.72	33.65	16.53	5.51	20.77	39.26
TA-DistMult	26.22	16.83	29.72	45.23	27.51	17.57	31.46	47.32
HyTE	16.78	2.13	24.84	43.94	16.05	6.53	20.20	34.72
RE-NET	35.77	25.99	40.10	54.87	36.86	26.24	41.85	57.60
CyGNet	34.68	25.35	38.88	53.16	35.46	25.44	40.20	54.47
TITer	38.73	32.70	46.46	58.44	39.98	32.77	41.86	55.01
EvoKG	37.24	30.94	44.84	57.67	36.81	38.48	40.58	54.92
DBKGE	38.39	26.34	45.01	58.22	39.79	38.55	41.46	57.28
DPMTKG	40.02	28.32	47.15	59.47	42.27	39.04	43.22	58.49

• 与时间基线方法相比较,前 6 行静态方法的性能相对较差,是因为他们忽略了时间因素.

• 基于插值的模型(如 TTransE, TA-DistMult)在四元组的基础上对时间进行建模,因此具有一定的推理能力.基于外推的模型(如 RE-NET、CyGNet 和 TITer)对实体的时序特征信息捕获更加全面,因此在预测未来事实方面比基于插值的模型取得了更好的效果.

• 与其他基于外推的方法(如 RE-NET、CyGNet 和 TITer)相比,DPMTKG 模型在 4 个数据集上都取得了最好的效果.对于评估指标 *MRR*,与 DBKGE^[9]模型相比,DPMTKG 模型在 Yago11k 和 Wikidata12k 分别提高了 4.18% 和 1.87%,在 ICEWS14 和 ICEWS05-15 数据集上分别提高了 1.63% 和 2.48%.

表 6 展示了本文在 Yago11k 数据集和 Wikidata12k 数据集选取的两个查询(Florence Augusta Merriam Bailey, graduate from, ?, 1882)以及(Raphael Leo, playfor, ?, 2010)进行预测时,本文提出的 DPMTKG 模型和 DBKGE 模型^[9]候选实体中排名前 5 的实体的对比情况.在查询(Florence Augusta Merriam Bailey, graduatefrom, ?, 1882)时,

除了查询实体 Smith College 之外, 其他排名往前的几个实体与 Smith College 语义接近, 均为大学. 然而, DBKGE 得到的前 5 个候选实体中, Smith College 排名第 3, 并且排名第 5 的实体为 New York, 相较于其他 4 个实体而言语义相差较大. 当查询 (Raphael Leo, play for, ?, 2010) 的时候, 除了查询实体 Sporting Clube de Portugal B 之外, 其他几个实体均与 Raphael Leo 存在密切关联 (Raphael Leo 之前为这几个球队效力过). 与之相比, DBKGE 在查询得到的 5 个候选实体中, Sporting Clube de Portugal B 排名第 2, 并且 Raphael Leo 不曾效力过 Aue FC in the Erzgebirge 和 Spanish Club. 以上结果表明, 本文通过多个子模块联合建模显示地学习度量空间中实体语义随时间的动态表示, 可以精准捕捉实体语义随时间变化的过程.

表 6 关于 DPMTKG 和 DBKGE 在 Yago11k 和 Wikidata12k 数据集的一个查询实例分析

要查询的四元组	DPMTKG 的前 5 个候选实体	DBKGE 的前 5 个候选实体
(Florence Augusta Merriam Bailey, graduate from, ?, 1882)	Smith College , University of Maryland, Columbia University, Yale University, University of Arizona	University of Cambridge, Columbia University, Smith College , Yale University, New York
(Raphael Leo, play for, ?, 2010)	Sporting Clube de Portugal B , Arsenal_F.C Academy, Portugal national under-16 football team, Spanish Club, Spanish national football team	Portugal national under football team, Sporting Clube de Portugal B , Catalan national football team, Aue FC in the Erzgebirge, Spanish Club

注: 粗体代表查询答案

3.6 消融研究

为了研究模型各个部分对实验产生的影响, 本文进行了消融实验. 如表 7 和表 8 所示, DPMTKG w/o forward 表示未使用前向扩散, DPMTKG w/o backward 表示未使用反向扩散. 实验结果表明前向扩散与反向扩散在对模型的性能提升上都起了积极作用. DPMTKG w/o forward 通过反向扩散直接对原始数据进行迭代采样处理学习得到下一时刻的数据分布. 实验结果表明, 在对数据加噪进行模糊处理后能够帮助更好的捕捉数据的真实分布. DPMTKG w/o backward 只对数据进行前向扩散处理操作后便生成得到了下一时刻四元组的数据分布. 实验结果表明, 由于 DPMTKG w/o backward 在进行一系列的加噪过程后打乱了数据的真实分布, 因此 DPMTKG w/o backward 取得了较差的实验结果. 实验结果证实了模型的有效性.

表 7 在 Yago11k 和 Wikidata12k 数据集上消融研究的实验结果 (%)

模型	Yago11k				Wikidata12k			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
DPMTKG w/o forward	36.67	27.10	48.75	54.12	28.36	22.02	33.40	42.78
DPMTKG w/o backward	31.54	24.68	37.66	48.01	20.58	14.39	25.01	33.34
DPMTKG	50.28	42.19	56.34	66.26	45.17	39.67	51.80	55.24

表 8 在 ICEWS14 和 ICEWS05-15 数据集上消融研究的实验结果 (%)

模型	ICEWS14				ICEWS05-15			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
DPMTKG w/o forward	25.38	16.02	35.37	45.09	24.00	20.45	28.39	42.78
DPMTKG w/o backward	20.17	9.04	28.66	40.68	19.44	10.01	26.54	38.93
DPMTKG	40.02	28.32	47.15	59.47	42.27	39.04	43.22	58.49

3.7 超参数对实验的影响

• 迭代次数 J 的影响. 在前向扩散过程中, 原始数据的分布被逐渐添加的噪音所打乱, 直至接近高斯分布. 在探究迭代次数对实验的影响过程中, 设置 $J = \{500, 1000, 1500, 2000, 2500\}$, 并探究 $MRR/Hits@1/Hits@3/Hits@10$ 随 J 的变化情况. 如表 9 和表 10 所示, 在 4 个数据集上, 随着 J 的增加, MRR 都得到提升. 当 $J = 2000$ 的时候, 实验

都取得了较好的结果. 当 $J > 2000$ 的时候效果仍有小幅度提升, 但是由于迭代次数过多实验性能会受到影响. 实验结果表明, 通过增加迭代次数之后依次进行数据加噪、去噪生成下一时刻的数据分布会对实验结果产生积极影响.

表 9 在 Yago11k 和 Wikidata12k 数据集上不同迭代次数 J 的实验结果 (%)

J	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>
500	40.39	31.74	42.26	47.66	28.35	21.52	32.28	38.13
1000	43.04	35.38	45.60	55.64	36.96	29.38	44.45	44.51
1500	47.35	39.40	54.01	63.30	39.96	33.45	49.75	53.25
2000	50.28	42.19	56.34	66.26	45.17	39.67	51.80	55.24
2500	50.67	42.57	55.29	66.52	44.89	40.39	52.30	56.45

表 10 在 ICEWS14 和 ICEWS05-15 数据集上不同迭代次数 J 的实验结果 (%)

J	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>
500	29.45	22.67	32.01	39.54	30.20	25.33	33.87	42.29
1000	34.87	24.06	35.42	46.74	36.96	29.79	39.90	49.55
1500	38.12	27.94	42.47	52.69	39.96	32.97	41.93	54.01
2000	40.02	28.32	47.15	59.47	42.27	39.04	43.22	58.49
2500	40.11	28.98	46.29	58.35	42.98	38.91	43.01	59.23

• 采样次数 n 的影响. 在反向扩散中, 随着 n 的增大, 采样次数也要增加. 如果 n 比较大, 那么采样次数不断增加会给实验带来性能压力. 本实验探究了采样次数对 *MRR/Hits@1/Hits@3/Hits@10* 的影响. 在 $J = 2000$ 的情况下, 设置采样次数 $n = \{200, 400, 600, 800, 1000\}$. 如表 11 和表 12 所示, 随着采样次数增多, *Hits@10* 逐渐增大. 对于 Yago11k 数据集来说, 当 $n = 800$ 的时候, 性能趋于稳定; 对于 Wikidata12k 数据集来说, 当 $n = 600$ 的时候, 性能趋于稳定; 对于 ICEWS14 数据集来说, 当 $n = 800$ 的时候, 性能趋于稳定; 对于 ICEWS05-15 数据集来说, 当 $n = 600$ 的时候, 性能趋于稳定. 上述实验表明, 通过子序列采样方式在提高采样速度的同时也能够达到性能的稳定.

表 11 在 Yago11k 和 Wikidata12k 数据集上不同采样次数 n 的时序知识图谱推理结果 (%)

n	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>
200	21.25	18.65	24.44	30.53	20.02	17.52	24.89	32.48
400	35.67	30.44	38.62	43.39	38.44	32.09	44.60	49.34
600	44.39	36.32	49.02	60.58	45.17	39.67	51.80	55.24
800	50.28	42.19	56.34	66.26	44.88	39.90	51.01	54.58
1000	50.48	41.96	56.01	66.85	44.46	39.38	50.96	55.36

表 12 在 ICEWS14 和 ICEWS05-15 数据集上不同采样次数 n 的时序知识图谱推理结果 (%)

n	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>
200	10.19	7.96	13.12	20.34	15.78	12.49	21.48	24.21
400	24.87	19.45	29.02	37.50	30.01	23.50	36.46	42.19
600	33.08	25.74	41.95	52.38	39.89	35.42	40.17	53.69
800	40.02	28.32	47.15	59.47	42.27	39.04	43.22	58.49
1000	41.15	27.09	47.78	59.26	41.50	39.51	42.28	58.68

4 结 论

本文提出基于扩散概率分布的时序知识图谱推理模型. 对于按照时间顺序排列的历史序列信息, 本文建立起一个马尔可夫链, 通过多个子模块联合建模显示地学习度量空间中实体语义随时间的动态表示. 通过这种方式, 每

个子模块仅负责建模实体语义的一个微小演化过程. 在这种情况下, 使用高斯分布便可实现对每个子模块足够精确的建模. 本文在 4 个公开的数据集上进行了大量实验, 实验结果表明本文提出的方法优于基于变分自动编码器的方法. 具体来说, 对于评估指标 MRR, 模型在 Yago11k 和 Wikidata12k 分别提高了 4.18% 和 1.87%, 在 ICEWS14 和 ICEWS05-15 数据集上分别提高了 1.63% 和 2.48%.

未来主要有两个方面需要探索. 首先, TKG 通常使用现成的信息提取工具构建, 可能包含各种错误, 对 TKG 补全任务产生严重的负面影响. 未来的工作将重点考虑如何构建事实验证与时序知识图谱推理的联合模型, 为下游任务提供可靠的知识资源. 其次, 本文提出的方法仍然缺乏有效的反馈机制, 无法对学习的四元组进行及时和准确的评估. 未来的工作将重点考虑如何引入人类反馈强化学习^[38], 从而进一步提升模型对历史事实的归纳推理能力.

References:

- [1] Trivedi R, Dai HJ, Wang YC, Song L. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In: Proc. of the 34th Int'l Conf. on Machine Learning. Sydney: JMLR.org, 2017. 3462–3471.
- [2] Guan SP, Jin XL, Jia YT, Wang YZ, Cheng XQ. Knowledge reasoning over knowledge graph: A survey. Ruan Jian Xue Bao/Journal of Software, 2018, 29(10): 2966–2994 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5551.htm> [doi: 10.13328/j.cnki.jos.005551]
- [3] Wang X, Chen WX, Yang YJ, Zhang XW, Feng ZY. Research on knowledge graph partitioning algorithms: A survey. Chinese Journal of Computers, 2021, 44(1): 235–260 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2021.00235]
- [4] Trivedi R, Farajtabar M, Biswal P, Zha HY. DyRep: Learning representations over dynamic graphs. In: Proc. of the 2019 Int'l Conf. on Learning Representations. New Orleans: OpenReview.net, 2019.
- [5] Lin WQ, Wang M, Wang W, Wang CN, Jin SC. A survey to knowledge graph and its military application. Journal of Chinese Information Processing, 2020, 34(12): 9–16 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2020.12.003]
- [6] Zhang ZC. Incremental construction and visualization of time-series knowledge graph [MS. Thesis]. Kunming: Yunnan University, 2021 (in Chinese with English abstract). [doi: 10.27456/d.cnki.gyndu.2021.000446]
- [7] Zhu CC, Chen MH, Fan CJ, Cheng GQ, Zhan Y. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks. Proc. of the AAAI Conf. on Artificial Intelligence, 2020, 35(5): 4732–4740. [doi: 10.1609/aaai.v35i5.16604]
- [8] Jin W, Qu M, Jin XS, Ren X. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 2020. 6669–6683. [doi: 10.18653/v1/2020.emnlp-main.541]
- [9] Liao SY, Liang SS, Meng ZQ, Zhang Q. Learning dynamic embeddings for temporal knowledge graphs. In: Proc. of the 14th ACM Int'l Conf. on Web Search and Data Mining. ACM, 2021. 535–543. [doi: 10.1145/3437963.3441741]
- [10] Ma A, Yu YH, Yang SL, Shi C, Li J, Cai XX. Survey of knowledge graph based on reinforcement learning. Journal of Computer Research and Development, 2022, 59(8): 1694–1722 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.20211264]
- [11] Leblay J, Chekol MW. Deriving validity time in knowledge graph. In: Proc. of the 2018 Web Conf. Lyon: Int'l World Wide Web Conf. Steering Committee, 2018. 1771–1776. [doi: 10.1145/3184558.3191639]
- [12] García-Durán A, Dumančić S, Niepert M. Learning sequence encoders for temporal knowledge graph completion. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 4816–4821. [doi: 10.18653/v1/D18-1516]
- [13] Dasgupta SS, Ray SN, Talukdar P. HyTE: Hyperplane-based temporally aware knowledge graph embedding. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 2001–2011. [doi: 10.18653/v1/D18-1225]
- [14] Goel R, Kazemi SM, Brubaker M, Poupart P. Diachronic embedding for temporal knowledge graph completion. Proc. of the AAAI Conf. on Artificial Intelligence, 2020, 34(4): 3988–3995. [doi: 10.1609/aaai.v34i04.5815]
- [15] Xu CJ, Nayyeri M, Alkhoury F, Yazdi H, Lehmann J. Temporal knowledge graph completion based on time series Gaussian embedding. In: Proc. of the 19th Int'l Semantic Web Conf. Athens: Springer, 2020. 654–671. [doi: 10.1007/978-3-030-62419-4_37]
- [16] Xu CJ, Nayyeri M, Alkhoury F, Yazdi HS, Lehmann J. TeRo: A time-aware knowledge graph embedding via temporal rotation. In: Proc. of the 28th Int'l Conf. on Computational Linguistics. Barcelona: Int'l Committee on Computational Linguistics, 2020. 1583–1593. [doi: 10.18653/v1/2020.coling-main.139]

- [17] Ma SX, Li AP, Zhao XJ, Song YC. Learning BiLSTM-based embeddings for relation prediction in temporal knowledge graph. *Journal of Physics: Conf. Series*, 2021, 1871(1): 012050. [doi: [10.1088/1742-6596/1871/1/012050](https://doi.org/10.1088/1742-6596/1871/1/012050)]
- [18] Han Z, Chen P, Ma YP, Tresp V. DyERNIE: Dynamic evolution of riemannian manifold embeddings for temporal knowledge graph completion. In: *Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2020. 7301–7316. [doi: [10.18653/v1/2020.emnlp-main.593](https://doi.org/10.18653/v1/2020.emnlp-main.593)]
- [19] Xu CJ, Chen YY, Nayyeri M, Lehmann J. Temporal knowledge graph completion using a linear temporal regularizer and multivector embeddings. In: *Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2021. 2569–2578. [doi: [10.18653/v1/2021.naacl-main.202](https://doi.org/10.18653/v1/2021.naacl-main.202)]
- [20] Liu Y, Hua W, Qu JF, Xin KX, Zhou XF. Temporal knowledge completion with context-aware embeddings. *World Wide Web*, 2021, 24(2): 675–695. [doi: [10.1007/s11280-021-00867-6](https://doi.org/10.1007/s11280-021-00867-6)]
- [21] Shao PP, Zhang DW, Yang GH, Tao JH, Che FH, Liu T. Tucker decomposition-based temporal knowledge graph completion. *Knowledge-based Systems*, 2022, 238: 107841. [doi: [10.1016/j.knosys.2021.107841](https://doi.org/10.1016/j.knosys.2021.107841)]
- [22] Li ZX, Jin XL, Li W, Guan SP, Guo JF, Shen HW, Wang YZ, Cheng XQ. Temporal knowledge graph reasoning based on evolutionary representation learning. In: *Proc. of the 44th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*. Association for Computing Machinery, 2021. 408–417. [doi: [10.1145/3404835.3462963](https://doi.org/10.1145/3404835.3462963)]
- [23] Han Z, Chen P, Ma YP, Tresp V. Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In: *Proc. of the 2020 Int'l Conf. on Learning Representations*. OpenReview.net, 2020.
- [24] Sun HH, Zhong JL, Ma YP, Han Z, He K. TimeTraveler: Reinforcement learning for temporal knowledge graph forecasting. In: *Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing*. Punta Cana: Association for Computational Linguistics, 2021. 8306–8319. [doi: [10.18653/v1/2021.emnlp-main.655](https://doi.org/10.18653/v1/2021.emnlp-main.655)]
- [25] Deng S, Rangwala H, Ning Y. Dynamic knowledge graph based multi-event forecasting. In: *Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining*. ACM, 2020. 1585–1595. [doi: [10.1145/3394486.3403209](https://doi.org/10.1145/3394486.3403209)]
- [26] Li ZX, Jin XL, Guan SP, Li W, Guo JF, Wang YZ, Cheng XQ. Search from history and reason for future: Two-stage reasoning on temporal knowledge graphs. In: *Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers)*. Association for Computational Linguistics, 2021. 4732–4743. [doi: [10.18653/v1/2021.acl-long.365](https://doi.org/10.18653/v1/2021.acl-long.365)]
- [27] Park N, Liu FC, Mehta P, Cristofor D, Faloutsos C, Dong YX. EvoKG: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs. In: *Proc. of the 15th ACM Int'l Conf. on Web Search and Data Mining*. ACM, 2022. 794–803. [doi: [10.1145/3488560.3498451](https://doi.org/10.1145/3488560.3498451)]
- [28] Jung J, Jung J, Kang U. Learning to walk across time for interpretable temporal knowledge graph completion. In: *Proc. of the 27th ACM SIGKDD Conf. on Knowledge Discovery & Data Mining*. ACM, 2021. 786–795. [doi: [10.1145/3447548.3467292](https://doi.org/10.1145/3447548.3467292)]
- [29] Luo ST, Hu W. Diffusion probabilistic models for 3D point cloud generation. In: *Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 2836–2844. [doi: [10.1109/CVPR46437.2021.00286](https://doi.org/10.1109/CVPR46437.2021.00286)]
- [30] Nichol A, Dhariwal P. Improved denoising diffusion probabilistic models. arXiv:2102.09672, 2021.
- [31] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2017.
- [32] Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: *Proc. of the 26th Int'l Conf. on Neural Information Processing Systems*. Lake Tahoe: Curran Associates Inc., 2013. 2787–2795.
- [33] Yang BS, Yih WT, He XD, Gao JF, Deng L. Embedding entities and relations for learning and inference in knowledge bases. arXiv:1412.6575, 2015.
- [34] Schlichtkrull M, Kipf TN, Bloem P, van den Berg R, Titov I, Welling M. Modeling relational data with graph convolutional networks. In: *Proc. of the 15th Int'l Conf. on the Semantic Web*. Heraklion: Springer, 2018. 593–607. [doi: [10.1007/978-3-319-93417-4_38](https://doi.org/10.1007/978-3-319-93417-4_38)]
- [35] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. *Proc. of the AAAI Conf. on Artificial Intelligence*, 2018, 32(1): 1811–1818. [doi: [10.1609/aaai.v32i1.11573](https://doi.org/10.1609/aaai.v32i1.11573)]
- [36] Shang C, Tang Y, Huang J, Bi JB, He XD, Zhou BW. End-to-end structure-aware convolutional networks for knowledge base completion. *Proc. of the AAAI Conf. on Artificial Intelligence*, 2019, 33(1): 3060–3067. [doi: [10.1609/aaai.v33i01.33013060](https://doi.org/10.1609/aaai.v33i01.33013060)]
- [37] Sun ZQ, Deng HZ, Nie YJ, Tang J. RotatE: Knowledge graph embedding by relational rotation in complex space. arXiv:1902.10197, 2019.
- [38] Griffith S, Subramanian K, Scholz J, Isbell CL, Thomaz A. Policy shaping: Integrating human feedback with reinforcement learning. In: *Proc. of the 26th Int'l Conf. on Neural Information Processing Systems*. Lake Tahoe: Curran Associates Inc., 2013. 2625–2633.

附中文参考文献:

- [2] 官赛萍, 靳小龙, 贾岩涛, 王元卓, 程学旗. 面向知识图谱的知识推理研究进展. 软件学报, 2018, 29(10): 2966–2994. <http://www.jos.org.cn/1000-9825/5551.htm> [doi: 10.13328/j.cnki.jos.005551]
- [3] 王鑫, 陈蔚雪, 杨雅君, 张小旺, 冯志勇. 知识图谱划分算法研究综述. 计算机学报, 2021, 44(1): 235–260. [doi: 10.11897/SP.J.1016.2021.00235]
- [5] 林旺群, 汪淼, 王伟, 王重楠, 金松昌. 知识图谱研究现状及军事应用. 中文信息学报, 2020, 34(12): 9–16. [doi: 10.3969/j.issn.1003-0077.2020.12.003]
- [6] 张子辰. 时序知识图谱的增量构建与可视化 [硕士学位论文]. 云南: 云南大学, 2021. [doi: 10.27456/d.cnki.gyndu.2021.000446]
- [10] 马昂, 于艳华, 杨胜利, 石川, 李劫, 蔡修秀. 基于强化学习的知识图谱综述. 计算机研究与发展, 2022, 59(8): 1694–1722. [doi: 10.7544/issn1000-1239.20211264]



周光有(1983—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为自然语言处理, 信息检索.



谢鹏辉(1998—), 男, 硕士, 主要研究领域为自然语言处理, 计算机科学与技术.



李鹏飞(1997—), 男, 硕士, 主要研究领域为自然语言处理, 计算机科学与技术.



罗昌银(1983—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为大数据管理, 数据挖掘.