面向时序知识图谱推理的大语言模型知识蒸馏

Wang Xing[1], Wei Song[2], Siyu Lin[3],Chen Wu[2],ZhesiLi[4],Man Wang[2][1]西安电子科技大学计算机科学与技术学院[2]西南交通大学计算与人工智能学院
[3]重庆交通大学信息科学与工程学院[4]长安大学信息工程学院

arXiv:2601.00202v1 [cs.CL] 1 Jan 2026

摘要

时序知识图谱（TKG）推理对于提升智能决策系统的效率和可靠性至关重要，已成为未来人工智能应用的关键技术基础。尽管近期取得了进展，现有时序知识图谱推理模型通常依赖较大的参数规模和密集计算，导致硬件成本和能耗较高。这些限制阻碍了模型部署到资源受限、低功耗且需要实时推理的分布式平台。此外，现有的大多数模型压缩和蒸馏技术都是针对静态知识图谱设计的，无法充分捕捉时序知识图谱中固有的时序依赖关系，往往导致推理性能下降。为应对这些挑战，我们提出了一种专为时序知识图谱推理设计的蒸馏框架。我们的方法利用大语言模型作为教师模型来引导蒸馏过程，从而有效地将结构推理和时序推理能力迁移到轻量级学生模型中。通过将大规模公共知识与任务特定的时间信息相结合，所提出的框架增强了学生模型对时序动态的建模能力，同时保持了紧凑高效的架构。在多个公开可用的基准数据集上开展的大量实验表明，我们的方法持续优于强基线，在推理准确率、计算效率和实际可部署性之间实现了良好的平衡。

1 引言

时序知识图谱（时序知识图谱）是一种带有显式时间属性的知识存储结构，已广泛应用于现实场景中（Hamilton等人，2018年；Kazemi等人，2020）。随着自然语言处理的快速发展，时序知识图谱已日趋成熟，并得到越来越广泛的应用于系统运行、语义匹配和运行场景规划（Cai等人，2022年）。具有代表性的时序知识图谱包括马克斯·普朗克研究所开发的YAGO知识库（Mahdisoltani等人，2013），以及维基媒体基金会提出的多语言知识库WIKI（Vrande˘ci´c和Krötzsch，2014年）。

静态知识图谱推理模型处理不包含时间属性的事件三元组，使模型能够理解知识图谱并进行推理，从而执行链接预测和路径预测等任务（Bordes 等人，2013）。TransE 是该领域的先驱方法之一，它将关系映射为实体嵌入之间的平移（Bordes 等人，2013）。RotatE 将这一思想扩展为在复向量空间中将关系建模为旋转，从而增强了表示能力（Sun 等人，2019）。ComplEx 进一步在复数空间中编码实体和关系（Trouillon 等人，2016年），而 SimplE 则采用向量空间中的嵌入，以丰富实体和关系表示（Kazemiand Poole，2018年）。这些方法推动了静态知识图谱推理的发展，也为时序知识图谱推理奠定了坚实基础。

时序知识图谱推理模型以平移向量的形式表示知识，将两个实体嵌入之间的关系表示为一个平移向量（Hamilton等人，2018年）。一些方法采用基于深度神经网络的方法，例如在嵌入空间中应用反馈机制或卷积层（Jin等人，2020）。然而，这些继承自静态知识图谱研究的推理方法并未明确考虑时间属性，因而限制了其对事件动态演化进行建模的能力。除了对静态知识图谱模型进行基于嵌入的扩展外，近期研究还探索了引入显式的时序逻辑结构，以

增强对时序知识图谱的推理。可微逻辑规则学习框架（Xiong等人，2024b）通过逻辑规则对时序依赖关系进行建模，并支持可解释的时序推理和时间预测。

时序知识图谱推理方法可分为两种不同的设定，即插值和外推（Li等人，2021年）。给定一个时序知识图谱，基于插值的模型旨在解决时间区间 $[t_0, t_N]$内的推理问题，这也称为时序知识图谱补全（Hamilton等人，2018年）。相比之下，基于外推的模型侧重于预测未来时间戳 t_{N+1}及之后可能发生的事件（Trivedi等人，2017年）。具有代表性的插值模型包括 HyTE 及相关方法（Dasgupta等人，2018年）。

Chronos 旨在推断缺失的实体关系，但无法有效捕捉时序变化，也难以泛化到未见过的时间段（Mavromatis等人，2022年）。Know-Evolve 将时序演化表示为一个连续过程（Trivedi等人，2017年），而 ExTTE 提供可解释的推理与预测，但其性能会受到数据稀疏性的影响（Li等人，2021年）。CyGNet 引入循环演化网络，以建模时序知识图谱中的周期模式（Zhu等人，2021年）。其他方法采用图神经网络或循环神经网络来捕捉时序动态，例如 RE-NET（Jin等人，2020）。

尽管时序知识图谱推理已得到广泛研究，但许多方法是为大规模知识图谱设计的，并依赖于训练参数量巨大的模型（Cai等人，2022年）。这导致计算和存储需求高，限制了其在资源受限设备上的部署。为解决这一问题，研究者探索了轻量化建模和压缩技术（Hinton 等人，2015）。

模型加速与压缩已成为活跃的研究课题。知识蒸馏已广泛应用于计算机视觉和推荐系统等领域（Hinton 等人，2015）。在知识图谱领域，蒸馏已被用于将知识从大型模型迁移到紧凑的学生模型。然而，现有的大多数蒸馏方法聚焦于静态知识图谱，且未明确纳入时间信息（Park等人，2019）。

大语言模型（LLMs），例如 GPT- 3（Brownet al.，2020）、PaLM（Chowdhery et al.，2023）和 LLaMA（Touvron et al.，2023），凭借大规模预训练，展现出强大的推理和泛化能力。近期的进展，包括指令微调和基于人类反馈的强化学习，进一步增强了它们的推理能力（Ouyang et al.，2022年）。近期研究还表明，大语言模型能够获得相当可观的推理能力，并泛化至复杂的现实世界推理场景（Xiongetal.，2024a；Yangetal.，2024a）。除了本身的推理能力外，研究还发现，大语言模型能够支持结构化推理，并借助内部世界模型进行审慎规划（Yanget al.，2024b；Xionget al.，2025年）。这些特性使大语言模型成为知识蒸馏中很有潜力的教师模型。

受上述观察启发，我们提出了一种专为时序知识图谱推理设计的蒸馏方法。我们的方法利用大型语言模型作为教师，将公共知识和任务特定的时序知识融入蒸馏过程。通过以丰富的时序推理信号引导轻量级学生模型，该方法能够在低功耗分布式设备上实现高效部署。实验结果表明，我们的方法始终优于传统时序知识图谱推理模型，同时显著减小模型规模并降低计算成本。

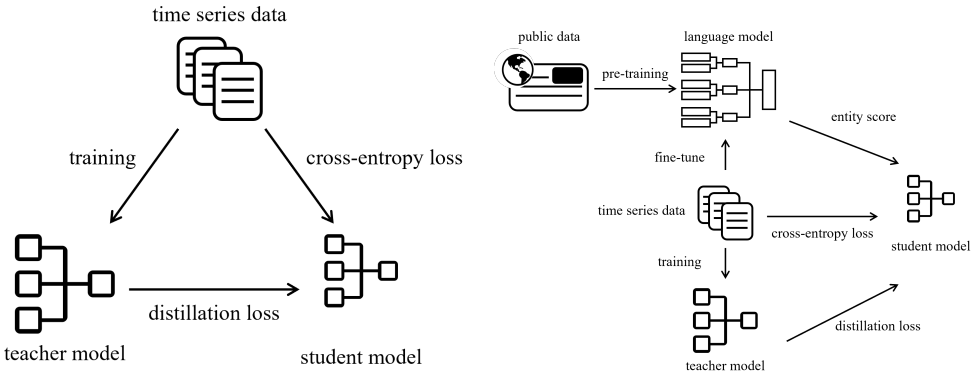
2 时序知识图谱推理模型的蒸馏方法

2.1 问题定义

时序知识图谱定义为一组时序事实四元组（s, p, o, t），其中 s 表示主语实体，p 表示关系，o 表示宾语实体，t 表示时间戳。图1展示了一个国家关系演变过程的示例，其中一系列事件发生在不同的时间戳上。一些实体事件会在不同的时间点多次发生，而另一些则只发生一次，从而形成动态时序结构。

给定一个时序知识图谱 $\mathcal{G}$，时序知识图谱推理的目标是在给定时间戳预测缺失事实，或根据历史观测预测未来事件。这项任务通常称为时序知识图谱推理，可表述为受时序约束的链接预测问题。

未来时序预测旨在回答



(a) 传统知识蒸馏方法　　(b) 基于大语言模型的蒸馏方法

图1：不同蒸馏方法的比较。

诸如未来某个时间点乌克兰与美国之间的关系将会如何这类查询 t_4。

与传统知识图谱不同，时序知识图谱显式地融入了时间信息。在时序知识图谱中，每条事实都表示为一个四元组（s，p，o，t），其中 $s \in E$ 表示主语实体，$o \in E$ 表示宾语实体，$p \in R$ 表示关系，$t \in T$ 表示时间戳。此处，E 和 R 分别表示实体集和关系集，T 表示时间戳集。令 $\mathcal{G}_t$ 表示时间 t处的时序知识图谱，并令 $g =$（s，p，o，t）表示 $\mathcal{G}_t$中的一条四元组事实。时序知识图谱是按时间排序的事实集序列，记为 $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T\}$，其中每个 $\mathcal{G}_t$ 包含时间 t发生的所有事实。

时序知识图谱推理侧重于预测图谱中的缺失事实。具体而言，给定一组观测事实，模型旨在推断缺失的宾语实体（s，p，？t）或缺失的主语实体（？p，o，t）。在本文的实验设置中，推理模型应用于时序事实预测任务，其中主语实体或宾语实体中的一个缺失。

在时间 t，时序知识图谱推理任务构建一组缺失事实四元组 $\mathcal{G}'_t$，其中每个四元组都是通过随机替换原始事实集 $\mathcal{G}_t$ 中的主语实体或宾语实体生成的。这两类不完整事实集共同构成完整时序知识图谱 $\mathcal{G}_c$，定义如下：

$$\mathcal{G}_c = \mathcal{G} \cup \mathcal{G}'. \tag{1}$$

形式上，不完整事实集合定义如下

$$\begin{aligned}\mathcal{G}' = &\big\{(s', p, o, t) \mid s' \in E, s' \neq s\big\} \\ &\cup \big\{(s, p, o', t) \mid o' \in E, o' \neq o\big\}.\end{aligned} \tag{2}$$

2.2 模型概述

受传统知识蒸馏方法启发，我们提出了基于大型语言模型的时序知识图谱推理蒸馏框架。在蒸馏过程中，选择一个高容量推理模型作为教师模型，同时训练一个轻量级模型作为学生模型。首先在时序数据上预训练教师模型，随后通过最小化由教师模型导出的蒸馏损失来优化学生模型。

除教师模型外，还引入了一个大型语言模型以增强蒸馏过程。该大型语言模型在大规模公共数据集上进行预训练，并进一步在时序知识图谱数据上进行微调，使其能够充当辅助教师模型。在训练过程中，学生模型同时向传统教师模型和大型语言模型学习。

在推理阶段，教师模型、大语言模型和学生模型可以独立评估。为更好地对齐学生模型与教师模型之间的预测分布，我们采用两阶段蒸馏策略。第一阶段侧重于使学生模型与传统教师模型对齐，第二阶段则利用大语言模型生成的预测分布，进一步优化学生模型。

2.3 损失函数

在训练过程中，引入了三种类型的损失函数。最终目标函数被定义为这些损失的加权和。令 L_1 表示学生模型与

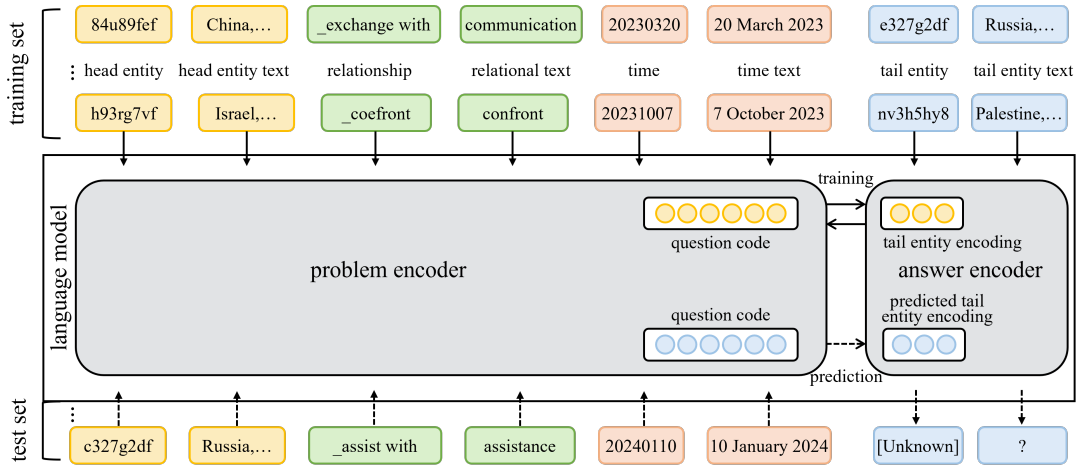


图 2：使用大型语言模型对四元组知识进行训练和预测。

传统教师模型之间的蒸馏损失，L_2 表示学生模型与大型语言模型之间的蒸馏损失，L_3 表示基于真实标签的监督损失。总损失定义为

$$L_{\text{total}} = L_1 + L_2 + \beta L_3. \tag{3}$$

第一个损失函数 L_1 旨在将时序推理知识从传统教师模型迁移到学生模型。采用编码器-解码器架构，以对齐两个模型生成的隐藏表示。给定一个事实四元组 $(s,\ p,\ o,\ t)$，蒸馏损失定义为

$$\begin{aligned} L_1 = \alpha\, \mathcal{C}\big(g \log f_T(s,p,o,t)\big) \\ + (1-\alpha)\, \mathcal{C}\big(g \log f_S(s,p,o,t)\big). \end{aligned} \tag{4}$$

其中，$f_T(\cdot)$ 和 $f_S(\cdot)$ 分别表示教师模型和学生模型的评分函数，α 是一个平衡系数，$\mathcal{C}(\cdot)$ 表示 SoftMax 函数。

为缩小学生模型与大型语言模型预测分布之间的差异，第二个损失函数 L_2 采用 Huber 损失，该损失在保持对小误差敏感的同时，对离群值具有鲁棒性。损失定义为

$$L_2 = \begin{cases} \frac{1}{2}(f_T - f_S)^2, & |f_T - f_S| \le \delta, \\ \delta |f_T - f_S| - \frac{1}{2}\delta^2, & |f_T - f_S| > \delta, \end{cases} \tag{5}$$

其中，f_T 和 f_S 分别表示大型语言模型和学生模型的预测分数，δ 是一个阈值超参数。

第三个损失函数 L_3 是基于真实标签的监督损失。大型语言模型用于编码实体关系语义，并生成稠密向量表示。预测过程定义如下：

$$\mathbf{E}(p) = \sigma\big(w_p\, \mathbf{LLM}(e) + b_p\big), \tag{6}$$

$$P(p) = \sigma\big(w \cdot \mathbf{E}(p) + b\big), \tag{7}$$

其中，LLM(·) 表示大型语言模型的输出嵌入，w_p、b_p、w 和 b 是可学习参数，σ 表示 SoftMax 函数。

最后，使用均方误差计算监督损失

$$L_3 = \|f_S(p) - P(p)\|_2^2. \tag{8}$$

大型语言模型的输出表示为一个向量。

大型语言模型的评分函数通过软蒸馏损失纳入蒸馏过程，该损失使用均方误差计算

$$L_3^{\text{LLM}} = \left\| f^{\text{LLM}}(E(s), P(p), E(o), t) - f^S \right\|_2^2. \tag{9}$$

这里，f^{LLM} 表示大型语言模型的评分函数。

用于训练学生模型的最终损失函数定义为上述三个损失项的加权组合

$$L_{\text{total}} = L_1 + \alpha L_2 + \beta L_3^{\text{LLM}}. \tag{10}$$

3 实验设置

本节介绍实验设置，包括数据集、基线模型、评估指标和训练配置。

表 1：两个数据集的统计数据

数据集	实体	关系	训练集	验证集	Test
YAGO	10,623	10	161,540	19,523	20,026
WIKI	12,544	24	539,286	67,538	63,110

表2：经典时序知识图谱模型介绍

模型	编码器	解码器
TTransE	–	$\|s+p-o+t\|_2$
TADistMult	$p_{\text{seq}} = \text{LSTM}(p:t)$	$(s \circ o)\, p_{\text{seq}}^{\top}$

3.1 数据集

我们选取了知识图谱领域中两个公开可用的基准数据集，即 YAGO11k 和 WIKIdata12k。它们广泛用于评估时序知识图谱推理模型（Hamilton等人，2018年；Cai等人，2022年）。这些数据集包含带有时间戳标注的时序事件，并支持插值和外推设置。

YAGO 数据集源自马克斯·普朗克研究所开发的知识库，其中包含从维基百科、WordNet 和 GeoNames 收集的事实数据（Mahdis- oltani 等人，2013）。该数据集包含丰富的时序和关系信息，适合用于评估时序推理性能。我们的实验使用的数据集是 YAGO11k。

WIKI 数据集基于维基数据和维基共享资源构建，其中维基数据是维基百科的结构化知识库（Vrande˘ci´c 和 Krötzsch，2014）。该数据集包含大量多语言事实和时序标注。我们的实验中用于时序知识图谱推理的数据集是 WIKI- data12k。

3.2 基线模型

BKD 是一种经典的知识蒸馏方法，通过最小化教师模型与学生模型输出分布之间的 Kullback-Leibler散度（Hinton 等人，2015）。它是基线蒸馏方法。

FitNet 在 BKD 的基础上引入中间层监督，使学生模型能够从教师网络的多个层学习表征，从而提供更丰富的训练信号（Srivastava 等人，2015）。

RKD 是一种基于结构的蒸馏方法，通过匹配教师模型与学生模型嵌入向量之间的距离和角度来传递关系知识，从而明确捕捉样本间的关系因素。

3.3 模型选择

所选的知识图谱嵌入模型为 TTransE 和 TADistMult。这些模型是具有代表性的时序知识图谱推理方法，且在时序链接预测任务中表现出色（Hamilton等人，2018年；Cai等人，2022年）。这些模型的编码器和解码器配置遵循其原始实现。

在这些模型中，s，$o \in \mathbb{R}^d$ 分别表示主体和客体实体的嵌入，而 $p \in \mathbb{R}^d$ 表示关系类型的嵌入。变量 t 表示时间嵌入，算子 $\circ$ 表示逐元素乘法。

3.4 评估指标

我们使用两个公开可用的数据集进行训练和评估。报告的指标包括平均排名（MR）、平均倒数排名（MRR）和前 k 名排名准确率，即 Hits@1、Hits@3 和 Hits@10。这些指标是知识图谱推理和链接预测任务的标准评估指标（Bordes 等人，2013；Trouillon 等人，2016年）。Hits@k 衡量正确实体在预测列表中排名位于前 k 位的测试查询所占比例。

对于所有指标，平均排名越低表示性能越好，而平均倒数排名和 Hits@k 的值越高，则表示推理能力越强（Sun 等人，2019）。

3.5 训练设置

所有实验均在一台配备 Intel Core i9 10900K CPU 和 NVIDIA GeForce RTX 3090 Ti GPU 的机器上进行。模型基于扩展版 OpenKE 框架（Han 等人，2018 年）实现，该框架构建于 PyTorch 之上。

为展示所提出的方法在教师模型与学生模型容量差异显著时的有效性，教师模型的嵌入维度设为 400，而学生模型的嵌入维度设为 25。这一配置遵循先前关于知识蒸馏和轻量级知识图谱嵌入模型的研究中的标准设置（Hinton 等人，2015）。

训练期间，批次大小设为 1024，且

表3：两个数据集上不同模型方法组合的蒸馏结果

模型	方法	YAGO MRR	YAGO MR	YAGO 命中率@1	YAGO 命中率@3	YAGO 命中率@10	WIKI MRR	WIKI MR	WIKI 命中率@1	WIKI 命中率@3	WIKI 命中率@10
TTransE	BKD	7.65	1410.12	3.50	7.83	15.61	**7.94**	2383.67	4.75	8.22	14.04
	FitNet	7.59	1201.69	3.06	7.18	16.48	7.86	2148.86	3.93	7.78	14.67
	RKD	7.01	**1186.27**	3.56	6.95	13.47	7.89	2052.37	4.72	7.49	12.85
	Ours	**7.69**	1193.15	**3.61**	**7.89**	**16.57**	7.92	**1985.63**	**4.86**	**8.36**	**14.94**
TADistMult	BKD	**61.90**	973.89	58.51	64.13	67.59	45.89	3150.11	42.46	48.87	51.18
	FitNet	58.44	986.92	54.71	60.29	65.34	43.92	3158.20	39.77	47.38	50.18
	RKD	58.15	1089.57	54.48	61.72	65.17	42.72	3287.49	36.32	43.92	47.28
	Ours	61.87	**965.35**	**58.73**	**64.15**	**67.68**	**46.03**	3142.85	42.50	**49.16**	51.14

为确保充分收敛，训练轮次的最大值设为10000。在蒸馏过程中，温度参数设为 7，遵循知识蒸馏中的常见做法（Hinton 等人，2015）。采用 Adagrad 优化器来处理稀疏数据并自适应调整学习率（Duchi 等人，2011 年）。

4 结果与分析

4.1 实验结果

在公开的 YAGO 和 WIKI 数据集上评估了所提出蒸馏方法的有效性。报告了两种数据集在不同模型和方法配置下的实验结果。除平均排名指标外，所有报告结果均以百分比表示。在相同的实验设置下，最佳结果以粗体标出，次佳结果以下划线标出。

实验结果表明，采用所提出蒸馏方法训练的学生模型始终优于采用传统蒸馏方法训练的模型。与基线 BKD 方法相比，所提出的方法在几乎所有评估指标上都取得了提升。对于 YAGO 数据集上的 TTransE 模型，五项评估指标分别提升了 0.5%、15.4%、3.1%、0.8% 和 6.1%。同样，在 WIKI 数据集上，平均排名、Hits@1、Hits@3 和 Hits@10 分别提升了 16.7%、2.3%、1.7% 和 6.4%。

TADistMult 模型也呈现出持续的性能提升。这些结果表明，所提出的蒸馏框架是有效的，并且与传统蒸馏方法相比，更能将知识迁移到轻量级学生模型中。

通过将所提出的方法与两种经典知识蒸馏方法 FitNet 和 RKD 进行比较，我们发现，所提出的方法在两个公共数据集上均能取得稳定且优异的性能。具体而言，对于 TADistMult 模型，所提出的方法在全部五项评估指标上均优于 FitNet。在 YAGO 数据集和 WIKI 数据集上，平均提升分别达到 2.77% 和 3.28%，进一步验证了所提出方法的有效性。

值得注意的是，所提出的方法在 Hits@1 和 Hits@3 上取得了最佳表现。这表明，采用所提出框架训练的学生模型更善于将正确实体排在更靠前的位置，这对于时序知识图谱推理至关重要。这些结果共同证明了所提出的方法在应对时序知识图谱推理模型蒸馏挑战方面的优势。

在少数实验设置中，所提出的方法在某些指标上并未取得最佳性能。例如，在 YAGO 数据集上评估 TTransE 模型时，其蒸馏效果在平均排名指标上略逊于 RKD 基线。同样，在 WIKI 数据集上，RKD 的表现也出现了明显波动。这种不稳定性可能源于大语言模型生成的预训练表示存在差异，导致在某些评估指标下难以有效迁移知识。此外，学生模型相对简单的结构也可能是造成这一现象的原因。实验结果表明，增加训练轮次有助于缓解这一问题。

4.2 消融分析

我们进一步基于实验结果开展消融分析，以评估基于大语言模型的知识蒸馏对学生模型训练的影响。通过将所提出的方法与 BKD 基线进行比较，我们验证了将大语言模型融入蒸馏框架的有效性。

在 YAGO 数据集上，以 TTransE 作为骨干模型时，所提出的方法在平均倒数排名、平均排名、Hits@1、Hits@3 和 Hits@10 上分别提升了 0.5%、15.4%、3.1%、0.8% 和 6.1%。对于 TADistMult 模型，在相应指标上观察到的提升分别为 7.0%、4.0%、1.0% 和 0.2%。

在 WIKI 数据集上，与 BKD 基线相比，使用所提出的蒸馏方法训练的学生模型在大多数评估指标上也取得了稳定提升。这些结果证实，从大语言模型中注入的知识能够有效提升轻量级学生模型的训练效果。

5 结论

本文提出了一种基于大语言模型的时序知识图谱推理蒸馏框架，旨在实现其在资源受限设备上的高效部署。通过利用大语言模型的推理和知识整合能力，所提出的方法能够有效地将时序推理知识迁移至轻量级学生模型。在多个数据集和骨干模型上的实验结果表明，该方法在显著降低计算和存储需求的同时，持续提升了推理准确率，优于现有的基于蒸馏的方法。

局限性

尽管所提出的框架效果显著，但仍存在若干局限。首先，蒸馏效果取决于大型语言模型的可用性和能力，这可能会增加训练期间的计算开销。其次，当前框架侧重于时序链接预测，并未明确处理更复杂的推理任务，例如多跳时序推理或逻辑规则归纳。

最后，性能提升可能因不同的骨干模型和数据集而异，尤其是在学生模型容量极其有限时。解决这些局限并将该框架扩展到更广泛的推理场景，仍是未来工作的重要方向。

参考文献

Antoine Bordes、Nicolas Usunier、Alberto Garcia-Duran、Jason Weston 和 Oksana Yakhnenko。2013年。用于多关系数据建模的平移嵌入。神经信息处理系统进展，26。

Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、PrafullaDhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell 及其他 1 人。2020年。语言模型是少样本学习者。神经信息处理系统进展，33:1877–1901。

Borui Cai、Yong Xiang、Longxiang Gao、He Zhang、Yunfeng Li 和 Jianxin Li。2022年。时序知识图谱补全：一项综述。*arXiv* 预印本 *arXiv:2201.08236*。

Aakanksha Chowdhery、Sharan Narang、Jacob Devlin、Maarten Bosma、Gaurav Mishra、Adam Roberts、Paul Barham、Hyung Won Chung、Charles Sutton、Sebastian Gehrmann 及其他 1 人。2023年。PaLM：利用路径扩展语言建模。机器学习研究期刊，24(240):1–113。

Shib Sankar Dasgupta、Swayambhu Nath Ray 和 Partha Talukdar。2018年。Hyte：基于超平面的时序感知知识图谱嵌入。收录于*2018*年自然语言处理经验方法会议论文集，第 2001–2011 页。

John Duchi、Elad Hazan 和 Yoram Singer。2011年。用于在线学习和随机优化的自适应次梯度方法。机器学习研究期刊。

Will Hamilton、Payal Bajaj、Marinka Zitnik、DanJuraf- sky 和 Jure Leskovec。2018年。在知识图谱上嵌入逻辑查询。神经信息处理系统进展，第 31 卷。

Xu Han、Shulin Cao、Xin Lv、Yankai Lin、Zhiyuan Liu、Maosong Sun 和 Juanzi Li。2018年。Openke：用于知识嵌入的开放工具包。收录于*2018*年自然语言处理经验方法会议论文集：系统演示，第 139–144 页。

Geoffrey Hinton、Oriol Vinyals 和 Jeff Dean。2015。蒸馏神经网络中的知识。*arXiv* 预印本 *arXiv:1503.02531*。

Woojeong Jin、Meng Qu、Xisen Jin 和 Xiang Ren。2020。循环事件网络：时序知识图谱上的自回归结构推断。发表于*2020* 年自然语言处理经验方法会议（*EMNLP*）论文集，第 6669–6683 页。

Seyed Mehran Kazemi、Rishab Goel、Kshitij Jain、Ivan Kobyzev、Akshay Sethi、Peter Forsyth 和 Pascal Poupart。2020年。动态图的表示学习：综述。机器学习研究期刊，21(70)：1–73。

Seyed Mehran Kazemi 和 David Poole。2018年。知识图谱中的链接预测简单嵌入。神经信息处理系统进展，31。

Zixuan Li、Xiaolong Jin、Wei Li、Saiping Guan、Jiafeng Guo、Huawei Shen、Yuanzhuo Wang 和 Xueqi Cheng。2021年。基于演化表示学习的时序知识图谱推理。在第*44*届国际 *ACM SIGIR* 信息检索研究与开发会议论文集中，第408–417页。

Farzaneh Mahdisoltani、Joanna Biega 和 Fabian M Suchanek。2013年。Yago3：来自多语言维基百科的知识库。在*CIDR*中。

Costas Mavromatis, Prasanna Lakkur Subramanyam, Vassilis N Ioannidis, Adesoji Adeshina, Phillip R Howard, Tetiana Grinberg, Nagib Hakim, and George Karypis. 2022年。Tempoqr：知识图谱上的时间问题推理。载于*AAAI*人工智能大会论文集，第36卷，第5825–5833页。

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022年。通过人类反馈训练能够遵循指令的语言模型。神经信息处理系统进展，35:27730–27744。

Wonpyo Park, Dongju Kim, Yan Lu, and Minsu Cho. 2019。关系知识蒸馏。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第3967–3976页。

Rupesh K Srivastava、Klaus Greff 和 Jürgen Schmidhuber。2015年。训练极深网络。神经信息处理系统进展，28。

Zhiqing Sun、Zhi-Hong Deng、Jian-Yun Nie 和 Jian Tang。2019年。RotatE：在复数空间中通过关系旋转进行知识图谱嵌入。*arXiv* 预印本 *arXiv:1902.10197*。

Hugo Touvron、Thibaut Lavril、Gautier Izacard、Xavier Martinet、Marie-Anne Lachaux、Timothée Lacroix、Baptiste Rozière、Naman Goyal、Eric Hambro、Faisal Azhar及另外1位作者。2023年。Llama：开放且高效的基础语言模型。*arXiv* 预印本 *arXiv:2302.13971*。

Rakshit Trivedi、Hanjun Dai、Yichen Wang和Le Song。2017年。Know-evolve：面向动态知识图谱的深度时序推理。载于国际机器学习大会，第3462–3471页。PMLR。Théo Trouillon、Johannes Welbl、Sebastian Riedel、Éric Gaussier和Guillaume Bouchard。2016年。用于简单链接预测的复数嵌入。载于国际机器学习大会，第2071–2080页。PMLR。Denny Vrandeˇci´c和Markus Krötzsch。2014年。Wikidata：一个免费的协作知识库。*ACM*通讯，57(10):78–85。

Siheng Xiong、Ali Payani、Ramana Kompella 和 Faramarz Fekri。2024a。大语言模型能够学习时序推理。载于计算语言学协会第 *62* 届年会论文集（第 *1* 卷：长篇论文），第 10452–10470 页。

Siheng Xiong、Ali Payani、Yuan Yang 和 Faramarz Fekri。2025年。语言模型中的审慎推理：基于准确世界模型的结构感知规划。载于计算语言学协会第 *63* 届年会论文集（第 *1* 卷：长篇论文），第 31900–31931 页。

Siheng Xiong、Yuan Yang、Faramarz Fekri 和 James Clayton Kerce。Tilp：在知识图谱上可微地学习时间逻辑规则。载于第十一届国际学习表征会议。

Siheng Xiong、Yuan Yang、Ali Payani、James C Kerce 和 Faramarz Fekri。2024b。Teilp：通过逻辑推理进行知识图谱时间预测。载于人工智能 *AAAI* 会议论文集，第 38 卷，第 16112–16119 页。

Yuan Yang、Siheng Xiong、Ali Payani、Ehsan Shareghi 和 Faramarz Fekri。2024a。大语言模型能否借助程序在真实场景中进行推理？载于计算语言学协会研究成果：*EMNLP 2024*，第 9806–9829 页。

Yuan Yang、Siheng Xiong、Ali Payani、Ehsan Shareghi 和 Faramarz Fekri。2024b。利用大语言模型的能力，将自然语言翻译为一阶逻辑。载于第*62*届计算语言学协会年会论文集（第*1*卷：长篇论文），第 6942–6959 页。

Cunchao Zhu、Muhao Chen、Changjun Fan、Guangquan Cheng 和 Yan Zhang。2021。《从历史中学习：使用序列复制生成网络对时序知识图谱进行建模》。载于*AAAI*人工智能大会论文集，第35卷，第4732–4740页。