DREAM：用于时序知识图谱推理的基于注意力机制的自适应强化学习

Shangfei Zheng 苏州大学计算机科学与技术学院，中国苏州
sfzhengsuda@stu.suda.edu.cn

Hongzhi Yin*昆士兰大学信息技术与电气工程学院，澳大利亚布里斯班
db.hongzhi@gmail.com

Tong Chen 昆士兰大学信息技术与电气工程学院，澳大利亚布里斯班
tong.chen@uq.edu.au

Quoc Viet Hung Nguyen 格里菲斯大学信息与通信技术学院，澳大利亚黄金海岸
h
e
nry.nguyen@griffith.edu.au

Wei Chen
计算机科学与技术学院，苏州大学 苏州, 中国
robertchen@suda.edu.cn

Lei Zhao*
中国苏州苏州大学计算机科学与技术学院
zhaol@suda.edu.cn

arXiv:2304.03984v1 [cs.AI] 8 Apr 2023

摘要

时序知识图谱（TKG）对事件的时间演化进行建模，近年来受到了越来越多的关注。由于时序知识图谱本质上是不完整的，因此有必要推理出缺失元素。尽管现有的时序知识图谱推理方法能够预测缺失的未来事件，但它们无法生成明确的推理路径，并且缺乏可解释性。随着近年来传统知识图谱上的多跳推理强化学习（RL）开始展现出更优越的可解释性和性能，这为探索强化学习技术在时序知识图谱推理中的应用提供了机会。然而，强化学习时序知识图谱推理方法的性能受到以下因素的限制：（1）缺乏联合捕获时间演化和语义依赖的能力；（2）过度依赖人工设计的奖励。为克服这些挑战，我们提出了一种基于注意力机制的自适应强化学习模型（DREAM），用于预测未来的缺失元素。-具体而言，该模型包含两个组成部分：（1）一种多维注意力表示学习方法，用于联合捕获语义依赖和时间演化；（2）一个自适应强化学习框架，通过自适应学习奖励函数来进行多跳推理。实验结果表明，DREAM 在公开数据集上的表现优于最先进的模型。

CCS 概念

• **计算方法** → **信念与知识推理**；**强化学习**。

关键词

时序知识图谱、多跳知识推理、链接预测

ACM参考文献格式：

Shangfei Zheng, Hongzhi Yin, Tong Chen, Quoc Viet Hung Nguyen, Wei Chen,and LeiZhao.2023.DREAM：基于注意力机制的时序知识图谱推理自适应强化学习.在*SIGIR'23*论文集，*2023*年*7*月*23*日至*27*日，台北。ACM，美国纽约州纽约市，11页。https://doi.org/XXXXXXX.XXXXXXX

*通讯作者

允许为个人或课堂使用而免费制作本作品的全部或部分数字版或纸质副本，前提是不得为营利或商业利益制作或分发副本，且副本首页须附有本声明及完整引用信息。本作品中由非 ACM所有的组件，其版权必须得到尊重。经注明出处后，允许进行摘要。若要以其他方式复制、再版、发布到服务器或重新分发到列表，则需事先获得特定许可和/或支付费用。请通过permissions@acm.org申请许可。*SIGIR'23*，*2023*年*7*月*23*日至*27*日，台北
© 2023 计算机协会。ACMISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 引言

知识图谱（KG）存储了大量关于现实世界的信息，并在许多下游应用中取得了巨大成功，例如问答 [57]和推荐系统 [63, 66, 67]。传统知识图谱以静态关系三元组（e_s，r，e_d）的形式表示，其中 e_s和 e_d分别是主语实体和宾语实体，r 表示一种关系，例如（*COVID-19*，发生，市政厅）。然而，在现实世界中，事件不断呈现出复杂的时间动态 [56, 65]。为了反映事件的时间性质，传统知识图谱扩展为时序知识图谱，其进一步将关系三元组与时间戳（e_s，r，e_d，t）相关联，例如（*COVID-19*，发生，市政厅，2022年12月6日）。

鉴于时序知识图谱本身存在不完备性，近期的时序知识图谱推理研究致力于根据已知事件推断缺失事件 [8]。例如，可以通过时序知识图谱推理预测四元组查询(市政厅中的 *COVID-19* 发生 ?, 2022年–10–6)中的缺失元素。给定时间跨度介于 t_0 和 t_T之间的时序知识图谱，时序知识图谱推理方法可分为两类，即插值和外推 [24, 36]。前者旨在推理出观测时间戳t 对应的缺失事实，其中该时间戳不晚于 t_T（即 $t_0 \leq t \leq t_T$）。相比之下，后者侧重于预测未来时间戳$t > t_T$上的新事实。外推在实际应用中具有重要价值，有助于金融风险控制 [3],、疫情预测 [35],和灾害预警 [14]等诸多领域。此外，未来事件预测已成为人工智能研究中的一个重要课题 [43, 69], ，使得

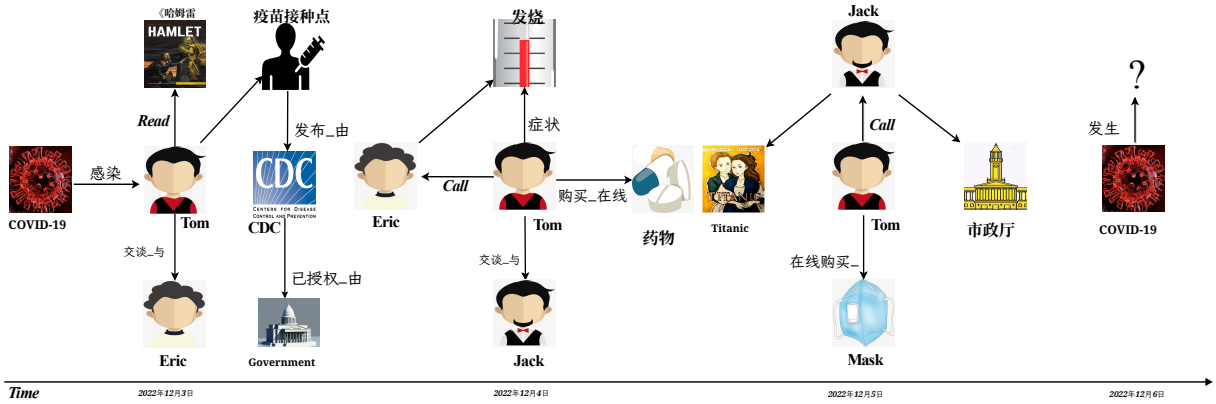


图 1：时序知识图谱中的一个小片段。包含实体和关系的快照序列按时间戳升序排列。通过利用历史和语义信息，可以推断出（*COVID-19*，发生，市政厅，2022年12月6日）。

基于外推的方法成为当前时序知识图谱推理研究的主流 [5]，也是我们工作的重点。

在文献中，关于外推式时序知识图谱推理已有大量研究，例如基于张量分解的方法 [25, 40],基于平移的方法 [13, 27], 基于神经网络的方法 [24, 36, 75], 等。尽管这些研究利用时序信息预测时序知识图谱中的缺失元素，但它们充当缺乏可解释性的黑盒系统 [53, 73]。传统知识图谱推理的最新进展表明，基于强化学习（强化学习）的多跳推理能够取得有前景且可解释的结果 [7, 76]。这一成功促使研究者将强化学习应用于外推式时序知识图谱推理。典型地，TPath [2],TAgent[48],TITer[47], 和 CluSTeR [35] 通过利用关系的符号组合与传递，学习多跳可解释路径。例如，通过连接（*COVID-19*，感染，*Tom*，2022年12月3日）、（*Tom*、交谈_与、*Jack*，2022年12月4日）和（*Jack*，拜访，市政厅，2022年12月5日），基于强化学习的模型可以推理出四元组（*COVID-19*，发生，市政厅，2022年12月6日）。尽管上述方法能够在时序知识图谱上以可解释的方式进行多跳推理，但其推理性能受到以下两个挑战的限制。

挑战 I。如何同时捕捉时间演化和语义依赖尚未得到探索。一些基于强化学习的时序知识图谱推理方法 [2, 47, 48] 无法捕捉**时间演化**（即历史事件的动态影响）。因此，CluSTeR [35] 在时间推理阶段考虑了历史事件，但通过采用循环神经网络（循环神经网络），仅利用了来自较小回溯窗口的有限历史信息（最近三天内发生的事件）。这是因为随着时间戳数量的增加，循环方法的扩展性较差 [52]。此外，基于循环神经网络的方法通常需要大量训练数据，这限制了CluSTeR [68]的性能。更进一步，现有的基于强化学习的时序知识图谱推理方法在很大程度上忽视了**语义依赖**。事实上，一个实体在不同关系中扮演着不同的角色 [54]。以图 1 为例，*Tom* 可能与 *Read* 等兴趣关系相连接，也可能与感染等临床关系相连接，而这些关系在响应查询（*COVID-19*，发生，？，2022年–10–6）时提供的信息并不相同。更加关注与查询关系高度相关的关系，有助于减少推理过程中的语义噪声 [46]。此外，一个实体之间的语义相关性实体及其邻居与它们的图距离 [70] 往往呈负相关。例如，Eric 的 4 跳邻居 *Government* 提供推理线索的可能性低于其 2 跳邻居 *COVID-19*。因此，同时利用时间演化和语义依赖来提升推理性能显然是必要的。

挑战 II。第二个挑战在于，前述基于强化学习的方法高度依赖人工设计的奖励，这很容易导致奖励稀疏困境、繁琐的设计过程 和性能波动 [10]。这三个问题中的任何一个都会直接阻碍基于强化学习的时序知识图谱推理模型的性能和泛化能力。(1) 奖励稀疏困境。仅考虑终止奖励所产生的稀疏奖励会导致学习缓慢甚至失败 [11, 37]。例如，TAgent由于其奖励是二元终止奖励而无法收敛 [48]。(2) 繁琐的设计过程。经验丰富的专家需要持续设计一些可能提升推理性能的候选辅助奖励，然后仔细选择最优的奖励函数（例如，CluSTeR加入了辅助束级奖励[35]）。然而，这种依赖领域专业知识的精细人工设计在不同数据集之间的泛化能力较低 [30, 55]。(3) 性能波动。人工设计的函数难以平衡探索与利用，从而不可避免地导致推理性能波动以及强化学习的决策偏差 [45, 49]。参考模型 TPath 的实验 [2]，在不同的时序知识图谱数据集上，所有评估指标均出现显著波动，其中一个主要原因是人工设计的路径多样性奖励所触发的决策偏差。

鉴于时序知识图谱推理面临的上述挑战，我们提出了一种名为 **DREAM**（基于注意力机制的自适应强化学习模型）的新模型。我们的创新受到以下两点启发：(1) 与循环方法相比，注意力机制更善于将多样的历史信息融入语义表示 [52]；(2) 生成对抗模仿学习能够从专家演示中自适应地学习策略 [22]。值得注意的是，DREAM 与现有多跳推理模型的主要区别在于，我们的模型不仅能够优雅地捕捉时间演化和语义依赖，还通过一种具有自适应奖励函数的新型自适应强化学习框架进行多跳推理。

具体而言，该模型包含两个组成部分。(1) 为解决 **挑战 I**，我们提出了一种多维注意力表示（MFAR）方法，以保留时序知识图谱的语义和时间属性。其关系感知衰减图注意力模块从多跳实体和关系中探索时序知识图谱的语义依赖。时序自注意力模块通过灵活地对历史上下文赋予不同权重，捕捉多个时间步上的演化。(2) 为解决 **挑战 II**，我们设计了一种基于生成对抗模仿学习的自适应强化学习框架（ARLF）。ARLF 旨在通过在语义和时序规则层级学习自适应奖励，预测缺失元素。总而言之，本文的贡献如下。

- 据我们所知，我们首次指出，在基于强化学习的时序知识图谱推理方法中预测缺失元素时，有必要联合利用时间演化和语义依赖，并设计自适应奖励函数。
- 我们提出了DREAM，这是一种有效且可解释的模型，通过利用关系的符号组合性来进行多跳推理。在DREAM中，我们提出 MFAR 来学习时序知识图谱的潜在特征表示。这些特征被输入 ARLF——一种新颖的基于强化学习的框架，该框架几乎不依赖手工设计的奖励函数——以实现可解释的推理结果。
- 我们在四个基准数据集上开展了广泛的实验 s. 实验结果展示了 DREAM 相较于最先进的基线方法的优越性。

2 相关工作

2.1 传统知识图谱推理

知识图谱是一种多关系图，反映现实中事件的内在联系 [76]。传统知识图谱由实体和关系构成的三元组表示。无论是人工构建还是自动抽取，知识图谱都存在不完整性[5, 37]。为解决不完整性问题，已有多种知识图谱推理方法 [8], , 例如基于平移的方法 [4],、基于神经网络的方法 [16]以及基于张量分解的方法 [71], 等。然而，上述方法仅适用于通过建模一步关系进行单跳推理，且缺乏可解释性 [53]。已知目前最先进的知识图谱推理方法主要是基于强化学习的多跳推理 [9]。这些方法探索多跳关系的语义组合，由此自然地产生可解释的推理结果 [8]。典型的基于强化学习的多跳推理方法包括 THML [73], 、DeepPath [61], 、MultiHop [37], 等。此外，基于规则的推理方法 [44, 64] 也通过学习逻辑规则实现可解释推理。尽管这些研究做出了巨大努力，但它们并非为时序知识图谱推理而设计 [24, 47]。

2.2 时序知识图谱推理

时序知识图谱推理方法通常分为插值和外推 [5]。前者推断历史时间戳处的缺失元素 [13, 15, 25, 27]。后者设置与我们的工作正交，我们尝试推断未来的缺失事实。通常，在预测未来事件时，插值式时序知识图谱推理模型的性能低于外推式模型 [13, 25, 27]。对于外推，一些研究 [19, 20, 50, 51]采用时间点过程或连续时间动态嵌入来估计连续时间下的条件概率。然而，这些方-法无法充分利用结构信息 [36]。通过结合卷积网络和课程学习，CEN应对了结构变化演化模式的挑战[33]。此外，RE-NET [24] 和 RE-GCN [36] 利用循环神经网络学习不同时间戳下的历史属性。同样，TiRGN [32] 和 HiSMatch [34] 分别采用循环神经网络和结构编码器来学习历史模式。然而，这些时序知识图谱推理方法不仅无法联合捕捉时间演化和语义依赖，而且由于循环神经网络的局限性，扩展性也较差 [28]。

另一类与我们的工作最为相关的基于强化学习的时序知识图谱推理方法，能够在缺乏多方面表示的情况下输出可解释的预测结果。形式上，基于强化学习的时序知识图谱推理方法将时序知识图谱推理视为马尔可夫决策过程，旨在学习用于推断时序知识图谱中缺失元素的最优策略。具体而言，TAgent [48] 仅采用二元终止奖励进行时序知识图谱推理，因此无法获得足够的奖励 [49]。为进一步提升奖励函数的质量，TPath [2] 和TITer [47] 分别引入了路径多样性奖励和时间塑形奖励。CluSTeR采用RNN学习结构信息，并将该信息整合到束级奖励函数中[35]。然而，这些模型的性能依赖于人工设计的奖励，因此受奖励稀疏困境、繁琐的设计过程以及性能波动的限制。尽管基于图学习的xERTE [18]和基于规则的TLogic [39] 也能进行可解释推理，但其推理性能通常低于基于强化学习的最先进推理模型。

2.3 强化学习

大多数基于强化学习的算法都利用智能体从环境中获取奖励，并通过马尔可夫决策过程（MDP）优化策略 [53, 74]。将强化学习应用于现实世界问题时，构建奖励函数是一项关键挑战 [10]。事实上，在基于强化学习的模型中，人工设计的奖励往往会面临奖励稀疏困境，导致学习缓慢甚至失败 [11, 37]。为缓解上述问题，奖励塑形受到了广泛关注 [55]。然而，这些费力方法的设计原则仍过度依赖人类经验，导致其在不同数据集上的泛化能力较低，且性能波动较大 [30, 49]。为解决上述问题，逆向强化学习（IRL）尝试从专家示例中自适应地学习奖励，但其学习循环在大型环境中会造成高昂的运行成本 [1]。因此，生成对抗模仿学习被应用于游戏模拟领域，通过生成对抗网络（GAN）消除所有中间的 IRL 步骤，并模仿专家示范。

3 预备知识

一个时序知识图谱 G = (E, R,T, Q) 是一个带时间戳的边连接实体的有向多关系图，其中 E、R和 T分别是实体、关系和时间戳的集合。Q = $\{(e_s, r_q, e_d, t_q) \mid e_s, e_d \in \mathrm{E}, r_q \in \mathrm{R}, t_q \in \mathrm{T}\}$是 G 中四元组的集合，其中 e_s、e_d、r_q 和 t_q 分别是源实体、目标实体、

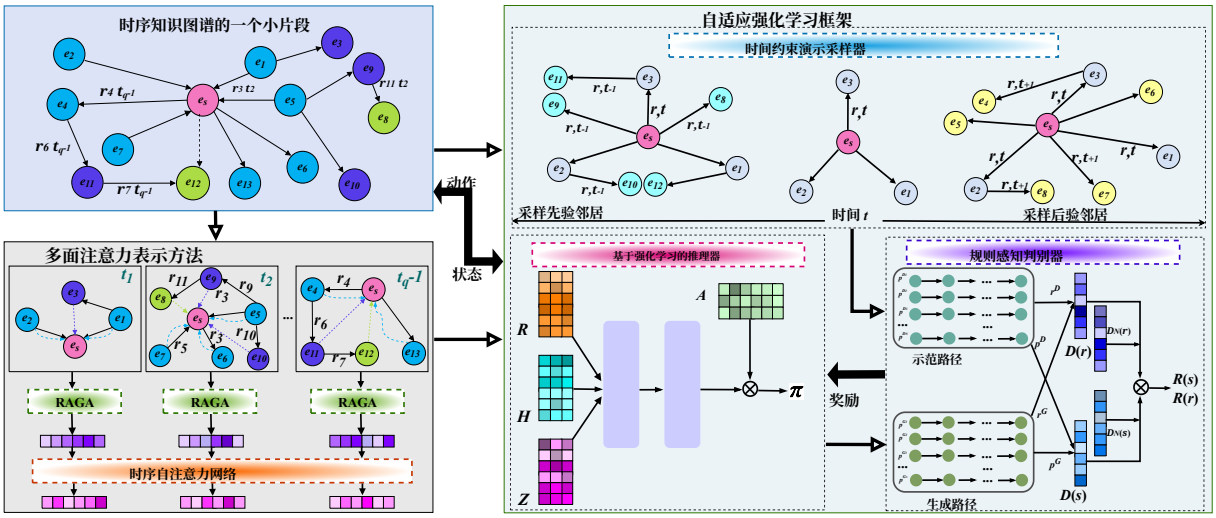


图2：DREAM 的详细示意图。将表示输入基于强化学习的推理器后，推理器输出推理路径。接下来，采样器通过双向采样获取示范路径。然后，判别器为推理器生成自适应奖励。最后，该推理器更新推理策略，并与时序知识图谱交互以完成预测。

分别表示它们之间的关系和一个时间戳。TKG 被视为按照时间戳升序排列的快照序列，即$\mathcal{G} = \{G1, G2, ..., G_T\}$。随后，[35, 47],逆关系也会附加到四元组中，即每个四元组（e_s, r, e_d, t）等价于四元组（e_d, r^{-1}, e_s, t）。在不失一般性的情况下，我们通过将 $(?, r_q, e_d, t_q)$转换为（$e_d, r_q^{-1}, ?, t_q$）来推断缺失的主体实体。本文研究的问题是外推式时序知识图谱推理，其被形式化为链接预测，旨在推断未来四元组。形式上，对于四元组查询（$e_s, r_q, ?, t_q$），外推式时序知识图谱推理的目标是给定 t_q 之前的历史事实集合（记为$\{G1, G2, ..., G^{t_q-1}\}$），通过生成一条可解释的 k 跳推理路径来预测缺失实体 e_d。例如，四元组查询（COVID- 19,发生, ?, 2022年12月6日）中的客体实体缺失。通过连接3跳路径COVID-19 $\xrightarrow[2022-10-3]{Infect}$ Tom $\xrightarrow[2022-10-4]{Talk_to}$ Jack $\xrightarrow[2022-10-5]{Visit}$ 市政厅，基于强化学习的模型获得四元组（COVID-19,发生,市政厅, 2022年12月6日）。我们将 TKG 问题表述如下：

- 输入是一个四元组查询（e_s、r_q、?、t_q）或（?、r_q、e_d、t_q），其中r_q 是查询关系，“?”是缺失的目标或源实体，t_q 是未来时间戳（即它仅出现在测试集）。
- 输出是通过不超过 k 跳数的推理路径获得的推断实体，$k > 1$。

4 方法论

4.1 DREAM 概述

如图 2 所示，DREAM包含两个组件：(1) 多维注意力表示（MFAR），该组件同时捕获时间演化和语义依赖，并生成多方面表示；(2) 带有自适应奖励机制的自适应强化学习框架（ARLF），基于从 MFAR 获得的表示进行可解释推理。下文将详细介绍这两个组件的设计。

4.2 多方面注意力表示

现有方法无法同时捕捉时间演化和语义依赖，这限制了时序数据 [6, 23] 的利用。为了解决这一问题，我们引入了一种多维注意力表示方法。MFAR 联合学习多跳邻居中的实体特征和关系特征，并通过灵活地对历史事件进行加权来捕捉潜在的时间演化。值得注意的是，与 TKGR 领域现有表示学习方法相比，我们的方法在技术上的差异主要体现在以下两点。(1) 我们通过计算多跳邻域内的实体和关系来扩展图注意力网络（GAT），而GAT仅考虑一跳邻域中的实体特征。此外，与传统知识图谱 [59, 72]，中基于关系感知图网络的嵌入方法相比，我们的方法不仅引入了衰减机制来模拟多跳实体 [73],贡献的逐渐减弱，而且计算成本更低。这是因为 MFAR通过对关系路径的嵌入求和，直接关联多跳邻居，而不是增加网络层数或引入代数空间 [26]。(2)我们通过学习注意力权重，而不是使用不可扩展的循环方法 [52]，从更长范围的先前时间戳中捕捉推理线索。

具体而言，MFAR 将图注意力和时间注意力解耦为独立模块，即关系感知衰减图注意力 （RAGA）和时间自注意力网络（ TSAN ）。在获取所有实体和关系的静态表示后，RAGA 会在每个时间戳通过衰减机制提取多跳邻域信息。随后，TSAN 捕捉长程时间演化，并依赖每个实体的历史信息。这种解耦式设计模式有助于在序列任务中实现高效、灵活的并行处理 [60]。

4.2.1 关系感知的衰减图注意力。通常，现有的时序知识图谱推理方法 [32–36, 47] 仅通过

以彼此割裂的方式关注实体和关系的表示来学习结构信息，无法利用事件三元组 [6] 的多样语义。更糟糕的是，由于忽视了语义衰减 [37]，现有方法会在推理过程中引入噪声。受 [41]，启发，我们提出 RAGA，以在本小节中解决上述语义依赖问题。RAGA 的输入是 G_t $\in \mathbb{R}^D$ $\forall e_i \in$ $_i \in \mathbb{R}^{F'}$ $\forall r_i \in$ ( ),

其中包含一组实体表示 $\{\boldsymbol{e}, \mathrm{E}\}$ 和关系表示 $\{\boldsymbol{r}, \mathrm{R}\}$。输出是一组新的实体表示 $\{\boldsymbol{e}_i^s \in \mathbb{R}^{D_e'}, \forall e_i \in \mathrm{E}\}$，其具有 D_e' 个维度，能够充分捕获 G_t 中的语义依赖。

具体而言，为了获得给定实体 e_i 到其 k 跳邻居实体的直接嵌入，我们在 e_i 与其 k 跳邻居实体 [61] 之间随机采样一条最短路径，然后对该路径中所有关系的嵌入求和，构造辅助关系嵌入 $\boldsymbol{r}_m$，即 $\boldsymbol{r}_m = \boldsymbol{r}_1 + ... + \boldsymbol{r}_k$。给定关系 r_k 表示时间戳 G_t 中从 e_i 到某个邻居实体的关系，其表示如下：

$$\boldsymbol{r}_k = \begin{cases} \boldsymbol{r} & k = 1 \\ \boldsymbol{r}_m & k > 1 \end{cases} \tag{1}$$

其中，k 表示关系路径中的跳数，$\boldsymbol{r}$ 表示 e_i 与直接邻居实体之间的关系嵌入。

我们首先通过拼接获得与实体 e_i 相关的每个三元组的向量表示，这样可以对给定快照 G_t 中实体在不同关系中扮演角色的差异进行建模。计算过程如下：

$$\boldsymbol{t}_{ikj} = \mathbf{W}_1[\boldsymbol{e}_i \oplus \boldsymbol{r}_k \oplus \boldsymbol{e}_j] \tag{2}$$

其中，$\oplus$ 表示拼接，$\boldsymbol{t}^{ikj}$ 表示给定快照 G_t 中的三元组向量表示，$\boldsymbol{e}^j$ 表示实体 e_i 在 G_t 中的直接流入邻居或多跳邻居的嵌入。

接下来，为了建模给定实体 e_i 在 G_t 中的 k 跳邻居实体的距离敏感贡献，鉴于高斯函数的稳定性和简洁性，我们定义了一个基于高斯函数的衰减系数，

$$w_{ij} = \exp(-\frac{k^2}{2b^2}) \tag{3}$$

其中，b 为高斯带宽。衰减系数 w_{ij} 确保距离 e_i 更近的实体 e_j 被赋予更高的权重。

然后，我们采用激活函数 LeakyReLU 来计算 G_t 中每个三元组 $\boldsymbol{t}^{ikj}$ 的重要性，记为 β_{ikj}，

$$\beta_{ikj} = \mathrm{LeakyReLU}(\mathbf{W}_2 \boldsymbol{t}_{ikj}) \tag{4}$$

其中，$\mathbf{W}_2$ 为线性变换矩阵。

随后，我们进一步计算 G_t 中每个三元组的衰减注意力权重 α_{ikj}，

$$\alpha_{ikj} = \frac{\exp(w_{ij}\beta_{ikj})}{\sum_{n \in \mathcal{N}_i} \sum_{r \in \mathcal{R}_{in}} \exp(w_{in}\beta_{irn})} \tag{5}$$

其中，N_i 表示实体 e_i 的直接邻居和多跳邻居，R 表示该实体 e_i 的直接入向关系和多跳关系。

基于能够提升学习过程稳定性的多头注意力机制，我们最终更新了新的实体 e_i 的嵌入，即 α_{ikj} 与 $\boldsymbol{t}^{ikj}$ 之间乘积的加权和，

$$\boldsymbol{e}_i^s = \Big\|_{m=1}^{M} \sigma\Big(\sum_{j \in \mathcal{N}_i} \sum_{k \in \mathcal{R}_{ij}} \alpha_{ikj}^m \boldsymbol{t}_{ikj}^m\Big). \tag{6}$$

其中，M 表示注意力头的最大数量，$\|$ 表示拼接。$\boldsymbol{e}_i^s$ 是实体嵌入，用于捕获实体 e_i 在时间戳 G_t 处的语义依赖。该嵌入会输入下文所述的时间注意力网络。

4.2.2 **时序自注意力网络。**现有的时序知识图谱表示方法大多采用循环网络聚合历史特征，但这类方法存在信息损失和计算效率低的问题 [28]。受注意力机制在序列任务中高效且表现显著优越的启发 [52]，，我们在 DREAM 中设计了时序自注意力网络（TSAN）。对于时间戳 t（$t<t_q$）处的特定实体 e_i，我们将 TSAN 的输入定义为表示序列 $\{\boldsymbol{e}_i^{s1}, \boldsymbol{e}_i^{s2}, ..., \boldsymbol{e}^s{}_{(t-1)\,i}, \boldsymbol{e}_i^{st}\}$，其中 e^s 的维度为 D'，并已在 RAGA 中充分捕获语义依赖。此外，TSAN 的输出是实体 e_i 在不同时间戳处的新表示序列 $\{\boldsymbol{z}_1^{e_i}, \boldsymbol{z}_2^{e_i}, ..., \boldsymbol{z}_{t-1}^{e_i}, \boldsymbol{z}_t^{e_i}\}$，$\boldsymbol{z} \in \mathbb{R}^{F'}$ 的维度为 $F_\prime$。形式上，将各时间戳上的表示打包在一起 t 后，输入和输出表示分别记为 $\mathbf{X}^s \in \mathbb{R}^{t \times D'}$ 和 $\mathbf{Z}^{e_i}{}_t \in \mathbb{R}^{t \times F'}$。

从技术角度来说，TSAN 的自注意力通过计算历史事件之间的相似度来捕捉时间演化。具体而言，查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 被视为输入实体表示。$\boldsymbol{Q}$、$\boldsymbol{K}$ 和值 $\boldsymbol{V} \in \mathbb{R}^{D' \times F'}$ 通过线性投影矩阵 $\mathbf{W}_q$、$\mathbf{W}_k$、$\mathbf{W}_v$ 分别变换查询、键和值，使它们具有相同的形状。受先前研究的启发 [24, 36]，，TSAN 允许每个时间戳 t 关注截至 t 的所有时间戳，从而保留自回归性质。

我们首先通过计算点积来建模时间演化的相关性，然后应用 Softmax 函数得到时间注意力值。接下来，时间戳 t 之前实体的新嵌入，是各表示按其时间注意力值加权后的总和。上述过程定义如下：

$$\mathbf{Z} = (\mathbf{X}^s\mathbf{W}_v)\mathrm{Softmax}((\mathbf{X}^s\mathbf{W}_q)^T(\mathbf{X}^s\mathbf{W}_k)) \tag{7}$$

最后，我们采用 $M_\prime$ 个注意力头，并组合得到时间戳 t 时实体 e_i 的多方面表示，

$$\mathbf{Z}_t^{e_i} = [\mathbf{Z}_1 \oplus \mathbf{Z}_2 \oplus ... \oplus \mathbf{Z}_{M'}] \tag{8}$$

其中，$M_\prime$ 是注意力头的最大数量。TSAN 的输出通过多头注意力机制进行更新。需要注意的是，RAGA 生成的实体语义表示被输入 TSAN 模块，以学习时间信息。在此基础上，MFAR 同时捕获时间演化和语义依赖。

4.3 自适应强化学习

基本上，现有的基于强化学习的时序知识图谱推理方案采用人工设计的奖励函数，导致推理过程既耗费大量劳力又缺乏成效 [10, 11, 37]。为解决这一问题，我们在本小节提出了一种新颖的自适应强化学习框架（ARLF）。ARLF 的关键创新在于以下方面

两点。(1) 对于不同的时序知识图谱数据集，ARLF 利用生成对抗网络通过模仿示范路径自适应地学习奖励，从而稳定推理性能并减轻人工干预。(2) ARLF 不仅将语义和时间知识引入生成对抗网络，还在时序知识图谱推理中明确建模网络的训练过程。这为基于强化学习的时序知识图谱推理方法提供了新的研究视角。

ARLF由三个模块组成，即基于强化学习的推理器、时间约束演示采样器 和 规则感知判别器。具体而言，推理器首先利用一个智能体输出关于缺失元素的多样化生成路径。接下来，时间约束演示采样器同时选择先验邻居和后验邻居，以生成高质量的示范路径。然后，规则感知判别器从语义和时序规则层级区分这些路径。最后，为了从规则感知判别器中获得更具自适应性的奖励，推理器尝试通过模仿演示样本生成路径，以欺骗规则感知判别器。基于此，基于强化学习的推理器不断更新推理策略，直到预测出缺失元素。

4.3.1 **基于强化学习的推理器。** 基于强化学习的推理器将推理过程视为马尔可夫决策过程，其目标是选择一系列最优动作并更新其推理策略（即对缺失元素的预测）。这一交互步骤直观地类似于"散步"，从而直观地形成时序知识图谱推理的可解释结果 [9]。因此，推理器通过马尔可夫决策过程的4元组（状态、动作、状态转移、奖励）来训练智能体与时序知识图谱进行交互。

状态。 状态由一些时序知识图谱元素组成，用于描述局部环境。形式上，在推理步 l 处的每个状态 s_l 表示为 $s_l = (e_l, t_l, e_s, t_q, r_q) \in \mathcal{S}$，其中$(e_l, t_l)$表示当前推理步 l的实体和时间戳，而(e_s, t_q, r_q)对应于原始查询中的实体、时间戳和关系，这些内容在所有步骤中保持不变。状态空间 S 可视为可用状态的集合。初始状态为$(e_s, t_q, e_s, t_q, r_q)$，因为基于强化学习的智能体从四元组查询的 e_s 开始。考虑到实体随时间持续演变，状态表示采用了多维注意力表示。因此，$\mathbf{Z}_{t_l}^{e_l}$表示时间戳 t_l处实体e_l的表示。此外，遵循时序知识图谱推理中的常见做法 [6, 24], ，我们使用$\mathbf{Z}_{t_q-1}^{e_s}$ 近似表示实体 e_s 在未来时间戳 t_q 处的多方面表示，因为推理过程中实际上无法观测到 t_q。

动作。 给定的 s_l的动作空间 A 包含推理步 l处的有效动作集合 A_l。 A_l定义为 $A_l = \{(r_\prime, e_\prime, t_\prime) \mid (e_l, r_\prime, e_\prime, t_\prime) \in \mathcal{A}, t_\prime \leq t_l, t_\prime \leq t_q\}$。此外，我们为每个 A 添加一条自环边。当推理展开至最大推理步 L时，自环的作用类似于 *STOP* 动作，从而避免推理过程无限展开。

状态转移。 新状态 s_{l+1}由先前状态 s_l和动作 A_l更新得到。形式上，$\mathrm{P}_r: \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ 定义为 $\mathrm{P}_r(s_l, A_l) = s_{l+1} = (e_{l+1}, t_{l+1}, e_q, t_q, r_q)$，其中 P 是状态转移函数。

奖励。 我们定义一种粗粒度终止奖励，以引导智能体选择基本策略，使其在初始训练期间能够接近目标实体 e_d。这种奖励是所有强化学习方法 [10]的基本奖励范式。如果目标实体是真实实体，则将终止奖励 R_t 设为 1。e_d。否则，终止奖励的值为 0。请注意，最终奖励 $R(s_l)$ 定义于公式（18）中。

策略网络。 推理器的目标是学习一个策略网络，以驱动上述由马尔可夫决策过程和时序知识图谱构成的4元组之间的交互。该策略网络以 MFAR 学得的嵌入作为输入，并输出下一个动作。与单跳方法不同，多跳推理方法会保留推理历史。形式上，推理历史 h_l由已拜访元素的序列构成，即 $(e_s, t_q), r_1, (e_1, t_1), ..., r_l, (e_l, t_l)$。我们定义了一种改进的路径表示$\mathbf{P}$，通过重塑$\mathbf{r}_l \in \mathbb{R}^{F'}$ 为$\mathbf{R}_l \in \mathbb{R}^{1 \times F'}$，

将关系嵌入与实体嵌入拼接起来$\mathbf{P}_l = [\mathbf{R}_l \oplus \mathbf{Z}_{t_l}^{e_l}]$。 h_l的历史嵌入可使用长短期记忆网络（LSTM）进行编码，

$$\mathbf{H}_l = \mathrm{LSTM}([\mathbf{H}_{l-1} \oplus \mathbf{P}_{l-1}]). \tag{9}$$

基于此，计算动作空间概率分布的策略网络 π 可定义如下，

$$\pi_\theta(a_l|s_l) = \sigma(\mathbf{A}_l(\mathbf{W}'' \mathrm{ReLu}(\mathbf{W}'[\mathbf{Z}_{t_l}^{e_l} \oplus \mathbf{H}_l \oplus \mathbf{R}_q]))) \tag{10}$$

其中，A_l通过堆叠所有动作的嵌入编码为$\mathbf{A}_l$，σ 是 Softmax 运算符，用于最大化动作空间 A_l中下一个动作的概率。

4.3.2 **时间约束演示采样器。** 在本小节中，我们提出一种时间约束演示采样器，以考虑时序知识图谱的时间属性，提取高质量的演示样本（专家路径）。具体而言，我们使用双向广度优先搜索（Bi-BFS），为每个事实探索 e_s 与 e_d 之间的最短路径，并将其作为演示样本。这是因为最短关系路径的组合能够准确表示两个实体之间的语义联系 [8, 37, 61]。然而，传统的双向广度优先搜索会均匀采样先验邻居和后验邻居，这可能会意外探索超出推理时间范围的邻居。为了解决这一问题，我们在双向广度优先搜索中设计了一种指数加权机制。例如，在双向广度优先搜索中考虑采样先验邻居时，指数加权机制的采样概率定义如下：

$$\mathcal{P}(e_i, r_i, e_j, t') = \frac{\exp(t' - t)}{\sum_{(e_i, r_l, e_k, t'') \in \mathcal{N}_q} \exp(t'' - t)} \tag{11}$$

其中，N_q 表示 e_i的先验邻居。$t_\prime$ 和$t_\prime$ 早于$t_\circ$ 本质上，通过指数加权，在采样过程中，来自时间戳更接近 t 的先验实体将被赋予更高的概率。

4.3.3 **规则感知判别器。** 受 [31], 启发，设计了规则感知判别器，分别从语义和时序逻辑层面评估推理器生成的路径以及演示样本。规则感知判别器包括语义判别器和时序规则判别器。

语义判别器。 为了在语义层面区分示范路径和生成路径，判别器首先通过采用策略网络中提到的路径拼接来学习路径的表示$\mathbf{P}$。受 ConvKB 启发，ConvKB 展示了卷积神经网络在提取知识图谱的语义特征方面的有效性 [42],

我们通过滑动卷积核$\boldsymbol{w}$来使用卷积层提取嵌入$\mathbf{P}$的语义特征，

$$D(s) = \sigma(\mathbf{W}_s(\mathrm{ReLu}(\mathrm{Conv}(\mathbf{P}, \boldsymbol{w}) + \boldsymbol{b}_s)) \tag{12}$$

其中，$\boldsymbol{b}_s$表示偏置项，Sigmoid函数 σ在区间 (0,1) 上输出路径语义特征 $D(s)$。

时序逻辑判别器。为提高路径的透明度和可信度，时序规则判别器进一步从时序知识图谱中学习时间逻辑规则，使推理器能够直接模仿时间约束示例中隐含的时序逻辑。时序逻辑规则 R_t 定义如下：

$$\wedge_{i=1}^{l}(E_i, r_i, E_{i+1}, T_i) \Rightarrow (E_1, r_h, E_{l+1}, T_{l+1}) \tag{13}$$

其中，$\wedge$ 表示逻辑合取，$T_1 < T_2 < \ldots < T_l < T_{l+1}$。 R_t的右侧称为规则头，r_h 表示头部关系。此外， R_t的左侧是规则体，由原子四元组 (e_s, r_1, e_2, t_1) $\wedge \ldots \wedge (e_i, r_i, e_{i+1},$

$t_i \wedge \quad \wedge e_l\ r_{l-1}\ e_{l-1}\ t_{l-1} \wedge e_{l-1}\ r_l\ e_l\ t_l \qquad _1, \qquad _l$) ... (, , ,) (, , ,) 的合取表示。我们用 Q ... 、 Q 表示上述四元组，并且 Q_h 表示规则头。

为了融入路径中隐含的时序规则的合理性，我们采用t-范数模糊逻辑 [17], , 将规则的真值度量定义为规则体中各真值度量的组合。为了获得四元组的真值度量 $I(\mathrm{Q}_i)$，我们首先获取关系和实体的拼接嵌入，然后使用多层神经网络作为特征提取器，表示如下：

$$I(Q_i) = \sigma(\mathbf{W}_r tanh([\mathbf{Z}_{t_i}^{e_i} \oplus \boldsymbol{R}_i \oplus \mathbf{Z}_{t_i}^{e_{i+1}})) \tag{14}$$

基于此，我们利用t-范数模糊逻辑来表示路径的时序逻辑特征，如下所示：

$$D(r) = I(Q_1) \cdot \ldots \cdot I(Q_l) \cdot I(Q_h) - I(Q_1) \cdot \ldots \cdot I(Q_l) + 1 \tag{15}$$

4.3.4 训练。本节遵循现有工作 [58]，定义最终奖励函数并形式化训练过程。具体而言，我们首先引入服从均匀分布的路径噪声 $D^N(s)$ 和逻辑噪声 $D^N(r)$，以剔除从基于强化学习的推理器获得的低质量路径，并从语义和时序逻辑层面获得自适应奖励。

$$R_s = \max(D^G(s) - D^N(s), 0) \tag{16}$$

$$R_r = \max(D^G(r) - D^N(r), 0) \tag{17}$$

其中， $D^G(s)$ 和 $D^G(r)$ 分别是基于强化学习的推理器获得的推理路径的语义特征和时序逻辑特征。 R_s和 R_r分别表示语义奖励和时序逻辑奖励。

接下来，我们将状态 s_l 下的自适应奖励 $R(s_l)$ 定义如下：

$$R(s_l) = R_t + \alpha R_s + (1 - \alpha) R_r \tag{18}$$

其中， α 是用于平衡 R_s 和 R_r的平衡因子。需要注意的是，智能体可以通过模仿语义层面和规则层面的演示样本，在每个推理步获得自适应奖励，从而消除稀疏奖励并确保性能稳定性 [1]。此外，ARLF通过从多样化路径中自动挖掘共通的元知识，提升了泛化能力，并避免了在不同数据集上的决策偏差 [45]。

表1：实验数据集的统计信息。

数据集	#训练集	#验证集	#测试集	#Ent	#Rel	*Time*
ICEWS14	74,845	8,514	7,371	6,869	230	1 天
ICEWS18	373,018	45,995	49,545	23,033	256	1 天
ICEWS05-15	368,868	46,302	46,159	10,094	251	1天
GDELT	1,734,399	238,765	305,241	7,691	240	15分钟

随后，我们通过降低训练损失来优化语义判别器和时序逻辑判别器，从而提升模仿学习的有效性。

$$\mathcal{L}_p = D^G(s) - D^D(s) + \lambda(\| \nabla_{\hat{p}} D(\hat{p}) \|_2 - 1)^2 \tag{19}$$

其中 λ是惩罚项，且 $\hat{p}$沿生成路径与专家示范之间的直线均匀采样。$D^G(s)$和 $D^D(s)$ 分别是从生成路径和专家示范中提取的语义特征。

对于时序逻辑判别器，我们将损失 $\mathrm{L}r$定义如下：

$$\mathcal{L}_r = -(\log D^D(r) + \log(1 - D^G(r)) \tag{20}$$

其中， $D^G(r)$和 $D^D(r)$ 分别是从生成路径和专家示范中提取的时序逻辑特征。

最后，为了最大化自适应强化学习的累积奖励并获得最优策略，目标函数如下：

$$\mathcal{J}(\theta) = \mathbb{E}_{(e_s, r_q, e_d, t_q)} \mathbb{E}_{a_1, \ldots, a_L \sim \pi_\theta}[R(s_l \mid e_s, r_q, t_q)] \tag{21}$$

5 实验

5.1 实验设置

5.1.1 数据集。 我们报告了 DREAM 在四个公开数据集 ICEWS14、ICEWS18、ICEWS05-15 和 GDELT上的推理性能。前三个数据集分别是综合危机预警系统（ICEWS）中发生于 *2014*年、*2018*年 和*2005*年至*2015*年的事件子集。最后一个数据集提取自全球事件与语言数据库 [29]。我们遵循 [47] 根据时间戳将数据集划分为训练集、验证集和测试集。数据集的数值信息如表1所示， *Time* 表示时间粒度。

5.1.2 评估指标。 我们采用两种广泛使用的评估指标来评估推理性能：平均倒数排名和Hits@N。请注意，为报告实验结果，我们采用更合理的时间感知过滤设置 [32]，而非过滤设置 [21, 62] 或原始设置 [12, 38]。这是因为时间感知过滤设置 仅过滤掉在查询时间 [34]发生的四元组。此外，时间感知过滤设置 能够模拟现实世界中的外推预测任务 [47]。

5.1.3 基线。 现有的外推式时序知识图谱推理研究 [18, 36, 47]已经充分证明，传统知识图谱推理方法（例如 TransE [4]）的性能低于时序知识图谱推理方法。因此，我们在实验中同时考虑了插值式时序知识图谱推理模型和外推式时序知识图谱推理模型。插值式时序知识图谱推理模型包括 TA-DistMult [13] 和 TNTComplEx [25]。此外，我们将模型与外推式时序知识图谱推理研究中的最新模型进行比较，包括非基于强化学习的模型 TANGO [19], 、XERTE [18], 、CyGNet [75], 、RE-GCN [36], 、RE-NET [24], 、TLogic[39], 和 HiSMatch [34]。最后，将基于强化学习的外推式时序知识图谱推理方法作为第三组

表2：公开时序知识图谱上的链接预测结果。

模型	ICEWS14				ICEWS05-15				ICEWS18				GDELT			
	平均倒数排名	H@1	H@3	H@10	平均倒数排名	H@1	H@3	H@10	平均倒数排名	H@1	H@3	H@10	平均倒数排名	H@1	H@3	H@10
TA-DistMult	25.8	16.9	29.7	43.0	24.3	14.6	27.9	44.2	16.7	8.6	18.4	33.6	12.0	5.8	12.9	23.5
TNTComplEx	34.1	25.1	38.5	50.9	27.5	9.5	30.8	42.9	21.2	13.3	24.0	36.9	19.5	12.4	20.8	33.4
CyGNet	35.1	25.7	39.0	53.6	36.8	26.6	41.6	56.2	24.9	15.9	28.3	42.6	18.5	11.5	19.6	32.0
RE-NET	36.9	26.8	39.5	54.8	43.3	33.4	47.8	63.1	28.8	19.1	32.4	47.5	19.6	12.4	21.0	34.0
RE-GCN	40.4	30.7	45.0	59.2	48.0	37.3	53.9	68.3	32.6	22.4	36.8	52.7	19.8	12.5	21.0	34.0
TANGO	36.8	27.3	40.8	55.1	42.9	32.7	48.1	62.3	28.9	19.4	32.2	47.0	19.2	12.2	20.4	32.8
XERTE	40.0	32.1	44.6	56.2	46.6	37.8	52.3	63.9	30.0	22.1	33.5	44.8	18.9	12.3	20.1	30.3
TLogic	42.5	33.2	47.6	60.3	47.0	36.2	53.1	67.4	29.6	20.4	33.6	48.1	19.8	12.2	21.7	35.6
HiSMatch	46.4	35.9	51.6	66.8	52.9	42.0	59.1	73.3	34.0	23.9	37.9	53.9	22.0	14.5	23.8	36.6
TPath	38.9	29.4	42.5	54.6	45.7	36.8	51.6	63.8	27.4	21.1	31.3	42.8	16.8	11.0	18.8	29.4
TAgent	39.1	30.7	43.2	55.2	45.1	36.3	51.0	63.4	27.2	20.1	30.7	42.3	17.7	11.8	19.6	30.6
TITer	40.9	32.1	45.5	57.6	47.7	38.0	52.9	65.8	30.0	22.1	33.5	44.8	20.2	14.1	22.2	31.2
CluSTeR	47.1	35.0	–	72.0	45.4	34.3	–	67.7	34.5	22.9	–	57.7	18.5	12.1	–	32.1
DREAM	**51.7**	**42.0**	**56.4**	**72.4**	**56.8**	**47.3**	**65.1**	**78.6**	**39.1**	**28.0**	**45.2**	**62.7**	**28.1**	**19.3**	**31.1**	**44.7**

包括 Tpath [2], TAgent [48], TITer [47], 和 CluSTer[35] 在内的基线。

5.1.4 实现细节。 我们提出的模型中包含一些超参数。维度 D 和 F，均设置为 200。λ 在所有数据集中均设置为 5。不同数据集上的注意力头数量 M 和M，均为 8。在模型训练中，最大训练轮数和卷积核大小分别为 400 和 3×5。奖励平衡因子 α、最大推理步数 L以及高斯带宽 b 的影响将在小节 **敏感性分析**中进行详细讨论。

5.2 TKGReasoning 上的结果

推理性能如表2所示（这些分数均以百分比表示），其中，具有竞争力的基线结果以下划线标示，最高性能结果以粗体突出显示。我们观察到，DREAM在所有评估指标和数据集上始终优于所有基线，这证明了其卓越的推理性能。具体而言，我们有以下详细观察结果。(1) 外推式时序知识图谱推理模型的表现通常优于第一组插值式时序知识图谱推理基线。我们认为，这一结果的原因在于，插值式时序知识图谱推理模型忽略了时间历史对于预测未来事件的重要性 [47]。(2) 我们提出的模型在所有情况下都优于第二组外推式基线。这是因为DREAM通过联合捕获时间演化和语义依赖，获得了多方面表示。此外，基于强化学习的DREAM能够利用最优推理策略，在不同推理线索之间进行多步推理，从而得到正确的目标。(3) 与当前最先进的基于强化学习的时序知识图谱推理方法相比，我们的模型在ICEWS05-15上的性能提升比在其他数据集上更加明显。(4) 由于实体是抽象概念，所有模型在GDELT上的推理性能通常都较低 [35]，但我们的模型仍然保持了具有竞争力的性能。这是因为DREAM能够学习多方面表示，以适应GDELT中频繁变化的历史事实，而GDELT具有最细粒度的时间粒度 [34]。

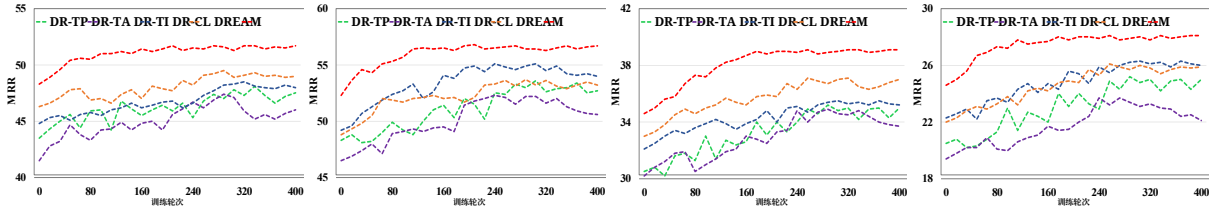
表3：四个数据集上的消融研究。

模型	ICEWS14	ICEWS05-15	ICEWS18	GDELT
不使用 MFAR	39.8(-11.9)	46.2(-10.6)	28.0(-11.1)	17.5(-10.6)
不使用 RAGA	49.5(-2.2)	53.4(-3.4)	36.8(-2.3)	26.0（-2.1）
不含 TSAN	46.2(-5.5)	51.0(-5.8)	33.5(-5.6)	21.8（-6.3）
不使用 ARLF	47.4（-4.3）	52.3(-4.5)	34.9(-4.2)	23.7(-4.4)
DREAM	**51.7**	**56.8**	**39.1**	**28.1**

5.3 消融研究

为研究 DREAM 的各个组件对性能的贡献，我们进行了消融分析。表3报告了 DREAM 及其变体的平均倒数排名结果。第一个变体版本 w/o MFAR 仅利用静态表示。w/o RAGA 仅使用时间注意力，而 w/o TSAN 仅考虑多方面表示方法中的语义依赖。需要注意的是，为评估 ARLF 的有效性，*w/o ARLF* 移除了 ARLF，仅保留带有基本终端奖励的基于强化学习的推理器。

我们有以下观察结果。(1)*w/oMFAR* 的平均倒数排名值远低于DREAM，这验证了同时捕获时间演化和语义依赖的必要性。(2) *w/o RAGA* 无法利用语义依赖，这导致其在稀疏数据集ICEWS05-15上的推理性能出现相对显著的下降。(3)*w/o TSAN* 在所有数据集上的性能下降均超过5%，这证明了TSAN在提升时序知识图谱上的推理性能方面的有效性。上述有效性有两个原因。首先，事件之间时间演化的客观存在是提升时序知识图谱推理性能的理论基础。其次，在ICEWS子集和GDELT中，时间间隔不超过60的查询相关实体分别占92%和83%。这一序列长度比例确保了自注意力机制在现有基准数据集上的技术有效性。(4) 尽管变体 *w/o ARLF* 采用多方面表示来进行时序知识图谱推理，但其性能仍然下降于



(a) ICEWS14 (b) ICEWS05-15 (c) ICEWS18 (d) GDELT

图3：T DREAM及其采用不同奖励函数的变体模型在数据集上的性能 ts.

所有数据集。这一实验结果可以解释为什么现有基于强化学习的时序知识图谱推理模型追求合理的奖励机制。下文将详细探讨奖励机制。

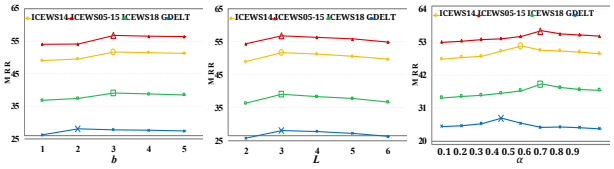
5.4 收敛性分析

为分析我们提出的自适应奖励与现有基于强化学习的时序知识图谱推理模型中的人工奖励之间的收敛速度和推理性能，我们设计了四种变体（即 DR-TP、DR-TA、DR-TI 和 DR-CL），通过在 DREAM 中移除自适应奖励并添加人工奖励构建而成。与 TAgent [48]和 TPath [2],类似，DR-TA 的奖励仅包含二元终止奖励，而 DR-TP 则额外加入路径多样性奖励。此外，DR-CL 和 DR-TI 分别加入了 CluSTeR [35] 的束级奖励和 TITer[47], 的时间塑形奖励。收敛速度如图3所示。采用自适应奖励的 DREAM 收敛最快，并且在不同的时序知识图谱数据集上表现最稳定。这是因为从语义和时间层面学习得到的自适应奖励函数能够消除稀疏奖励的困境，并避免在不同数据集上产生决策偏差 [30]。DR-TA 受到稀疏奖励的影响，因此收敛效果最差。此外，由于人工设计的多样性奖励会导致盲目探索，DR-TP 在不同数据集上的表现波动明显。尽管 DR-CL 和 DR-TI 最终能够收敛，但收敛速度较慢，且在所有数据集上的性能不稳定、泛化能力较弱。总之，DREAM 的自适应奖励机制不仅能够高效收敛，而且在所有数据集上都具有良好的泛化能力。

5.5 敏感性分析

在本小节中，我们研究了不同超参数下 DREAM 的性能波动。具体而言，如实现细节小节实现细节中所述，我们在图4中研究了模型对高斯带宽 b、最大推理步 L以及奖励平衡因子 α 值的敏感性。此外，根据公式（3）的定义，带宽值与衰减权重呈负相关。我们观察到，GDELT 上的最佳带宽值小于

在所有 ICEWS 子集上都是如此。一个可能的原因是，GDELT 中的数据较为密集，因此需要较大的衰减权重来过滤邻居中的噪声。如果 b 的值超过最优值，那么所有数据集上的平均倒数排名值都会相对稳定。这是因为高斯核的局部影响范围与带宽值呈正相关。超出该范围后，核函数的值变化很小，因此



(a) 带宽 (b) 推理步 (c) 平衡因子

图4：DREAM 在不同数据集上随关键超参数数量变化的性能。

b 对衰减权重的影响会趋于稳定。如图4(b)所示，当跳数超过 3 时，DREAM 的性能会相对下降，这与现有基于强化学习的时序知识图谱推理研究结果一致 [35, 47]。一种合理的解释是，3 跳之外的信息包含更多噪声，会对推理性能产生负面影响 [47]。图4(c)显示，四个数据集上的最优平衡因子各不相同。ICEWS14、ICEWS05-15、ICEWS18 和 GDELT 数据集上的最优平衡因子分别为 0.5、0.6、0.6 和 0.4。ICEWS 子集侧重于语义层面学习得到的奖励，这表明，在时序知识图谱数据集的时间粒度相同时，数据集规模与语义层面的自适应奖励呈正相关 [35]。

6 结论

本文提出了一种名为DREAM的新型基于强化学习的时序知识图谱推理模型，该模型克服了两个挑战：缺乏多方面表示以及过度依赖人工设计的奖励函数。针对前者，DREAM采用 MFAR，同时学习时间演化和语义依赖。为减轻对人工奖励的依赖，我们设计了一种新型ARLF，该框架通过模仿专家示范来进行多跳推理。大量实验表明，DREAM 在公开时序知识图谱数据集上具有出色的推理优势和有效性。

致谢

本研究部分得到澳大利亚研究理事会以下项目的资助：未来研究员项目（编号：FT210100624）、探索项目（编号：DP190101985）、探索早期职业研究奖（编号：DE200101465 和 DE230101033）以及工业转型培训中心（编号：IC200100022）。本研究还部分得到国家自然科学基金委员会（编号：62272332）、江苏省高等学校自然科学研究重大项目（编号：22KJA520006）的资助，

以及国家留学基金管理委员会高水平大学公派研究生项目（编号：202206920032）的资助。

参考文献

[1] Saurabh Arora 和 Prashant Doshi。2021年。逆向强化学习综述：挑战、方法与进展。《人工智能》297 (2021年)，103500–103512。[2] Luyi Bai、Wenting Yu、Mingzhuo Chen 和 Xiangnan Ma。2021年。使用强化学习在时序知识图谱的路径上进行多跳推理。《应用软计算》103 (2021年)，107144–107153。[3] Johan Bollen、Huina Mao 和 Xiao-Jun Zeng。2011年。Twitter 情绪可预测股票市场。《计算机科学杂志》2, 1 (2011年)，1–8。[4] Antoine Bordes、Nicolas Usunier、Alberto García-Durán、Jason Weston 和 Oksana Yakhnenko。2013年。用于多关系数据建模的嵌入转换。载于 NeurIPS。2787–2795。[5] Borui Cai、Yong Xiang、Longxiang Gao、He Zhang、Yunfeng Li 和 Jianxin Li。2022年。时序知识图谱补全：综述。CoRR abs/2201.08236 (2022年)，1–9。[6] Borui Cai、Yong Xiang、Longxiang Gao、He Zhang、Yunfeng Li 和 Jianxin Li。2022年。时序知识图谱补全：综述。CoRR abs/2201.08236 (2022年)。[7] Tong Chen、Hongzhi Yin、Guanhua Ye、Zi Huang、Yang Wang 和 Meng Wang。2020年。试试这个：个性化且可解释的替代推荐。载于 SIGIR。891–900。[8] Xiaojun Chen、Shengbin Jia 和 Yang Xiang。2020年。综述：知识图谱上的知识推理。《专家系统与应用》141 (2020年)，112948–112962。[9] Rajarshi Das、Shehzaad Dhuliawala、Manzil Zaheer、Luke Vilnis、Ishan Durugkar、Akshay Krishnamurthy、Alex Smola 和 Andrew McCallum。2018年。走一走，找到答案：使用强化学习在知识库中进行路径推理。载于 ICLR。1–18。[10] Rati Devidze、Goran Radanovic、Parameswaran Kamalaruban 和 Adish Singla。2021年。为强化学习智能体设计可解释的奖励。载于 NeurIPS。20118–20131。[11] Daniel Dewey。2014年。强化学习与奖励工程原则。载于 AAAI。1–7。[12] Yifu Gao、Linhui Feng、Linbo Qiao 和 Dongsheng Li。2022年。通过自编码器结构建模前兆以进行时序知识图谱推理。载于 IJCAI。2044–2051。[13] Alberto García-Durán、Sebastijan Dumancic 和 Mathias Niepert。2018年。学习用于时序知识图谱补全的序列编码器。载于 EMNLP。4816–4821。[14] Xingtong Ge、Yi Yang、Jiahui Chen、Weichao Li、Zhisheng Huang、Wenyue Zhang 和 Ling Peng。2022年。基于多源时空信息的灾害预测知识图谱。《遥感》14, 5 (2022年)，1214。[15] Rishab Goel、Seyed Mehran Kazemi、Marcus A. Brubaker 和 Pascal Poupart。2020年。用于时序知识图谱补全的历时嵌入。载于 AAAI。3988–3995。[16] Kelvin Guu、John Miller 和 Percy Liang。2015年。在向量空间中遍历知识图谱。载于 EMNLP。318–327。[17] Petr Hájek。1998年。模糊逻辑元数学。第4卷。共299页。[18] Zhen Han、Peng Chen、Yunpu Ma 和 Volker Tresp。2021年。用于时序知识图谱预测的可解释子图推理。载于 ICLR。1–24。[19] Zhen Han、Zifeng Ding、Yunpu Ma、Yujia Gu 和 Volker Tresp。2021年。学习神经常微分方程以预测时序知识图谱中的未来链接。载于 EMNLP。8352–8364。[20] Zhen Han、Yunpu Ma、Yuyi Wang、Stephan Günnemann 和 Volker Tresp。2020年。用于时序知识图谱预测的图 Hawkes 神经网络。载于 AKBC。1–14。[21] Yongquan He 、 Peng Zhang 、 Luchen Liu 、 Qi Liang 和 Chuang Zhang。2021年。HIP 网络：用于时序知识图谱外推推理的历史信息传递网络。载于 IJCAI。1915–1921。[22] Jonathan Ho 和 Stefano Ermon。2016年。生成对抗模仿学习。载于 NeurIPS。4565–4573。[23] Shaoxiong Ji、Shirui Pan、Erik Cambria、Pekka Marttinen 和 Philip S. Yu。2022年。知识图谱综述：表示、获取与应用。《IEEE 神经网络与学习系统汇刊》33, 2 (2022年)，494–514。[24] Woojeong Jin、Meng Qu、Xisen Jin 和 Xiang Ren。2020年。循环事件网络：时序知识图谱上的自回归结构推断。载于 EMNLP。6669–6683。[25] Timothée Lacroix、Guillaume Obozinski 和 Nicolas Usunier。2020年。用于时序知识库补全的张量分解。载于 ICLR。1–12。[26] Antonino Laudani、Gabriele Maria Lozito、Francesco Riganti Fulginei 和 Alessandro Salvini。2015年。前馈神经网络的训练效率与计算成本。《计算智能与神经科学》243: (2015年)，1–813。[27] Julien Leblay 和 Melisachew Wudage Chekol。2018年。在知识图谱中推导有效时间。载于 WWW。1771–1776。[28] John Boaz Lee、Ryan A. Rossi、Sungchul Kim、Nesreen K. Ahmed 和 Eunyee Koh。2019年。图中的注意力模型：综述。ACM《知识发现与数据挖掘汇刊》13, 6 (2019年)，1–25。[29] Kalev Leetaru 和 Philip A Schrodt。2013年。GDELT：1979–2012年事件、地点与情感倾向的全球数据。载于 ISA年会，第2卷。1–49。[30] Alexander C. Li、Lerrel Pinto 和 Pieter Abbeel。2020年。面向强化学习的广义后见经验。载于 NeurIPS。7954–7967。[31] Ruiping Li 和 Xiang Cheng。2019年。DIVINE：用于知识图谱推理的生成对抗模仿学习框架。载于 EMNLP-IJCNLP。2642–2651。[32] Yujia Li、Shiliang Sun 和 Jing Zhao。2022年。TiRGN：用于时序知识图谱推理的时间引导循环图网络及局部—全局历史模式。载于 IJCAI。2152–2158。[33] Zixuan Li、Saiping Guan、Xiaolong Jin、Wei Li、Jiafeng Guo 和 Xueqi Cheng。2022年。用于时序知识图谱推理的复杂演化模式学习。载于 ACL。290–296。[34] Zixuan Li、Zhongni Hou、Saiping Guan、Xiaolong Jin 和 Xueqi Cheng。2022年。HiSMatch：基于历史结构匹配的时序知识图谱推理。CoRR abs/2210.09708 (2022年)。[35] Zixuan Li、Xiaolong Jin、Saiping Guan、Wei Li、Jiafeng Guo、Yuanzhuo Wang 和 Xueqi Cheng。2021年。从历史中搜索，为未来推理：时序知识图谱上的两阶段推理。载于 ACL/IJCNLP。4732–4743。[36] 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。载于 SIGIR。408–417。[37] Xi Victoria Lin、Richard Socher 和 Caiming Xiong。2018年。通过奖励塑形进行多跳知识图谱推理。载于 EMNLP。3243–3253。[38] Kangzheng Liu、Feng Zhao、Hongxu Chen、Yicong Li、Guandong Xu 和 Hai Jin。[无日期。]DA-Net：用于时序知识图谱推理的分布式注意力网络。载于 CIKM。1289–1298。[39] Yushan Liu、Yunpu Ma、Marcel Hildebrandt、Mitchell Joblin 和 Volker Tresp。2022年。TLogic：用于时序知识图谱上可解释链接预测的时间逻辑规则。载于 AAAI。4120–4127。[40] Yunpu Ma、Volker Tresp 和 Erik A. Daxberger。2019年。情节式知识图谱的嵌入模型。《网络语义学杂志》59 (2019年)，100490–100499。[41] Deepak Nathani、Jatin Chauhan、Charu Sharma 和 Manohar Kaul。2019年。学习基于注意力的嵌入以进行知识图谱中的关系预测。载于 ACL。4710–4723。[42] Dai Quoc Nguyen、Tu Dinh Nguyen、Dat Quoc Nguyen 和 Dinh Q. Phung。2018年。一种基于卷积神经网络的全新知识库补全嵌入模型。载于 NAACL-HLT。327–333。[43] Yue Pan 和 Limao Zhang。2021年。人工智能在建筑工程与管理中的作用：关键综述与未来趋势。《建筑自动化》122 (2021年)，103517–103538。[44] Meng Qu、Junkun Chen、Louis-Pascal A. C. Xhonneux、Yoshua Bengio 和 Jian Tang。2021年。RNNLogic：学习用于知识图谱推理的逻辑规则。载于 ICLR。1–20。[45] Jinsheng Ren、Shangqi Guo 和 Feng Chen。2021年。强化学习中保持方向的奖励平衡。《IEEE神经网络与学习系统汇刊》(2021年)，1–15。[46] George Stoica、Otilia Stretcu、Emmanouil Antonios Platanios、Tom M. Mitchell 和 Barnabás Póczos。2020年。用于知识图谱链接预测的上下文参数生成。载于 AAAI。3000–3008。[47] Haohai Sun、Jialun Zhong、Yunpu Ma、Zhen Han 和 Kun He。2021年。TimeTraveler：用于时序知识图谱预测的强化学习。载于 EMNLP。8306–8319。[48] Ye Tao、Ying Li 和 Zhonghai Wu。2021年。通过强化学习进行时间链接预测。载于 ICASSP。3470–3474。[49] Amanda S Therrien、Daniel M Wolpert 和 Amy J Bastian。2016年。小脑损伤后，有效的强化学习需要在探索与运动噪声之间取得平衡。《脑》139, 1 (2016年)，101–114。[50] Rakshit Trivedi、Hanjun Dai、Yichen Wang 和 Le Song。2017年。Know-Evolve：动态知识图谱的深度时间推理。载于 ICML，第70卷。3462–3471。[51] Rakshit Trivedi、Mehrdad Farajtabar、Prasenjeet Biswal 和 Hongyuan Zha。2019年。DyRep：学习动态图库上的表示。载于 ICLR。1–25。[52] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N. Gomez、Lukasz Kaiser 和 Illia Polosukhin。2017年。注意力机制即所需的一切。载于 NeurIPS。5998–6008。[53] Guojia Wan、Shirui Pan、Chen Gong、Chuan Zhou 和 Gholamreza Haffari。2020年。像人类一样推理：用于知识图谱推理的分层强化学习。载于 IJCAI。1926–1932。[54] Heng Wang、Shuangyin Li、Rong Pan 和 Mingzhi Mao。2019年。将图注意力机制融入基于 Deep R 的知识图谱推理

einf orcemen t Learning. In *EMNLP-IJCNLP*. 2623 –2631。

[55] Kaixin Wang、Bingyi Kang、Jie Shao 和 Jiashi Feng。2020年。通过混合正则化提升强化学习的泛化能力。载于*NeurIPS*。7968–7978。[56] Quan Wang、Zhendong Mao、Bin Wang 和 Li Guo。2017年。知识图谱嵌入：方法与应用综述。*IEEE Transactions on Knowledge and Data Engineering* 29，12（2017年），2724–2743。[57] Qinyong Wang、Hongzhi Yin、Weiqing Wang、Zi Huang、Guibing Guo 和 Quoc Viet Hung Nguyen。2019年。利用神经记忆网络对知识图谱进行多跳路径查询。载于*DASFAA*，第11446卷。777–794。[58] Sheng Wang、Xiaoyin Chen 和 Shengwu Xiong。2021年。面向知识图谱推理的基于规则注入的生成对抗模仿学习。载于*PAKDD*，第12714卷。338–350。[59] Shen Wang、Xiaokai Wei、Cícero Nogueira dos Santos、Zhiguo Wang、Philip S. Yu 和 Isabel F. Cruz。2021年。用于知识图谱补全的混合曲率多关系图神经网络。载于*WWW*。1761–1771。[60] Tianwei Wang、Yuanzhi Zhu、Qianying Wang 和 Mingxiang Cai。2020年。用于文本识别的解耦注意力网络。载于*AAAI*。12216–12224。[61] Wenhan Xiong、Thien Hoang 和 William Yang Wang。2017年。DeepPath：一种用于知识图谱推理的强化学习方法。载于*EMNLP*。564–573。[62] Yi Xu、Junjie Ou、Hui Xu 和 Luoyi Fu。2022年。基于历史对比学习的时序知识图谱推理。*CoRR* abs/2211.10904（2022年）。[63] Hongrui Xuan、Yi Liu、Bohan Li 和 Hongzhi Yin。2023年。对比多行为推荐的知识增强。载于*WSDM*。195–203。[64] Fan Yang、Zhilin Yang 和 William W. Cohen。2017年。用于知识库推理的逻辑规则的可微学习。载于*NeurIPS*。2319–2328。[65] Hongzhi Yin 和 Bin Cui。2016年。社交媒体中的时空推荐。Springer。共114页。[66] Hongzhi Yin、Bin Cui、Zi Huang、Weiqing Wang、Xian Wu 和 Xiaofang Zhou。2015年。联合建模用户兴趣和移动模式以进行兴趣点推荐。在 *ACM MM*。819–822。[67] Junliang Yu、Hongzhi Yin、Xin Xia、Tong Chen、Lizhen Cui 和 Quoc Viet Hung Nguyen。2022。图增强是否必要？：用于推荐的简单图对比学习。在 *SIGIR*。1294–1303。[68] Yong Yu、Xiaosheng Si、Changhua Hu 和 Jianxun Zhang。2019。循环神经网络综述：长短期记忆单元与网络架构。*NeuralComput.* 31, 7（2019），1235–1270。[69] Caiming Zhang 和 Yang Lu。2021。人工智能研究：现状与未来展望。*J. Ind. Inf. Integr.* 23（2021），100224–100233。[70] Yao Zhang、Hongru Liang、Adam Jatowt、Wenqiang Lei、Xin Wei、Ning Jiang 和 Zhenglu Yang。2021。GMH：用于知识图谱补全的通用多跳推理模型。在 *EMNLP*。3437–3446。[71] Zhanqiu Zhang、Jianyu Cai 和 Jie Wang。2020。用于基于张量分解的知识图谱补全的对偶诱导正则化项。在 *NeurIPS*。1–12。[72] Zhao Zhang、Fuzhen Zhuang、Hengshu Zhu、Zhi-Ping Shi、Hui Xiong 和 Qing He。2020。采用层次注意力的关系图神经网络，用于知识图谱补全。在 *AAAI*。9612–9619。[73] Shangfei Zheng、Wei Chen、Pengpeng Zhao、An Liu、Junhua Fang 和 Lei Zhao。2021。难度何时产生影响：针对少样本关系的多跳知识图谱推理。在 *CIKM*。2688–2697。[74] Shangfei Zheng、Weiqing Wang、Jianfeng Qu、Hongzhi Yin、Wei Chen 和 Lei Zhao。2022。MMKGR：多跳多模态知识图谱推理。*CoRR* abs/2209.01416（2022）。[75] Cunchao Zhu、Muhao Chen、Changjun Fan、Guangquan Cheng 和 Yan Zhang。2021。从历史中学习：利用序列复制生成网络建模时序知识图谱。在 *AAAI*。4732–4740。[76] Xiaohan Zou。2020。知识图谱应用综述。在 *Journal of Physics*，第1487卷。12016–12028。