arXiv:2305.04034v1 [cs.AI] 6 May 2023

Wasserstein-Fisher-Rao 嵌入：兼顾局部比较与全局传输的逻辑查询嵌入

Zihao Wang[*†]、**Weizhi Fei**[*♣]、**Hang Yin**[♣]、**Yangqiu Song**[†]、**Ginny Y. Wong**[★]、**Simon See**[★]

† 计算机科学与工程系，香港科技大学，香港特别行政区，中国 ♣ 数学科学系，清华大学，中国北京 ★NVIDIA 人工智能技术中心（NVATIC），NVIDIA，美国圣克拉拉 zwanggc@cse.ust.hk，{fwz22,hyin-20}@mails.tsinghua.edu.cn yqsong@cse.ust.hk，{gwong,ssee}@nvidia.com

摘要

在知识图谱上回答复杂查询十分重要，但由于数据不完整，这一任务尤其具有挑战性。查询嵌入方法通过基于学习的模型来解决这一问题，并利用集合算子模拟逻辑推理。以往研究主要关注特定形式的嵌入，但嵌入之间的评分函数尚未得到充分探索。不同于现有的、受局部比较或全局传输启发的评分函数，本文基于非平衡最优传输理论，研究局部与全局之间的权衡。具体而言，我们将集合嵌入为 $\mathbb{R}$ 中的有界测度，并采用受 Wasserstein-Fisher-Rao 度量启发的评分函数。这种设计还支持在嵌入空间中构造闭式集合算子。此外，我们提出了一种基于卷积的算法，以在线性时间内完成计算，并设计了分块对角核来施加这种权衡。结果表明，在标准数据集、包含组合复杂查询的评测集以及层次化知识图谱上，WFRE 均能优于现有的查询嵌入方法。消融研究表明，找到更好的局部与全局权衡对于提升性能至关重要。[1]

1 引言

知识图谱存储现实世界的事实知识，以实体节点和关系边的形式表示（Miller，1995年；Bollacker 等人，2008年；Vran-de˘ci´c 和 Krötzsch，2014年）。它们还为许多下游任务提供支持（Xiong 等人，2017a；Wang 等人，2019年；Lin 等人，2020年）。值得注意的是，回答复杂逻辑查询是利用知识图谱中存储知识的一种重要方式（Ren 等，2020年，2021年）。

严格来说，复杂逻辑查询可以通过一阶逻辑（Ren 等，2020年；Marker，2002）来表达。某些查询组的谓词和逻辑连接词可以转换为集合运算（Wang 等人，2021年），由于其语义清晰而特别受到关注。因此，回答复杂查询的逻辑推理过程被转换为在算子树中执行集合投影和运算（Ren 等，2020年；Wang 等人，2021年）。图1 展示了查询“赢得金球奖或奥斯卡金像奖的非美国籍导演是谁”的算子树。

<escape>这项任务的难点在于知识图谱的数据不完整性。现代大规模知识图谱天然存在不完整性，因为它们是通过众包（Bollacker 等人，2008年；Vrande˘ci´c 和 Krötzsch，2014年）或自动化信息抽取流程（Carlson 等人，2010年）构建的。这一问题被称为开放世界假设（Libkin 和 Sirangelo，2009年）（OWA）。由于数据不完整，针对完整数据库应用查询回答算法无法得到完整的答案。此外，这些算法也无法利用观测到的不完整知识图谱来剪枝搜索空间，从而导致较高的计算成本（Ren 等，2020年）。当在包含数十亿条边的大规模知识图谱上回答逻辑查询时，这一问题会变得更加棘手（Ren 等，2022年）。关于知识图谱上的逻辑查询，读者可参阅近期的综述文章以了解更多信息（Wang 等人，2022b；Ren 等，2023年）。</escape>

固定维度空间中的查询嵌入方法（Hamilton 等人，2018年；Ren 等，2020年）旨在克服上述困难。通过将可学习的集合嵌入和算子泛化到未见数据，解决了数据不完整问题（Ren 等，2020年；Ren 和 Leskovec，2020年）。计算成本不会随着……而增长。开发高效的集合嵌入和算子形式是近期的研究重点之一。

[*] 共同贡献[1]我们的实现可在以下网址找到：https://github.com/HKUST-KnowComp/WFRE。

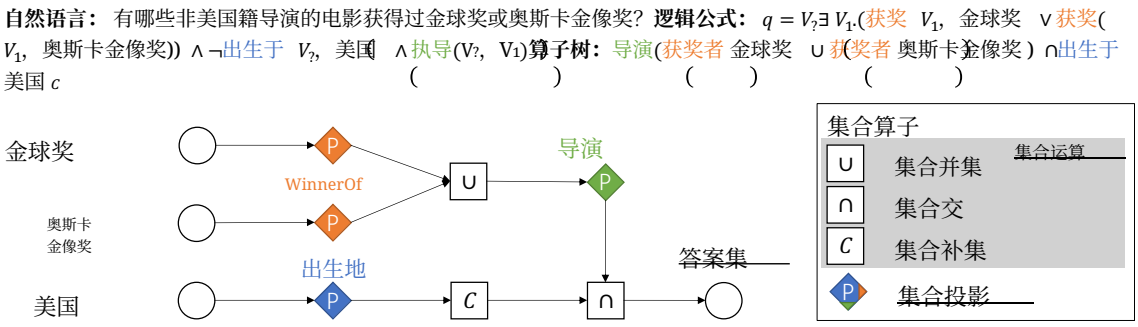

图1：在知识图谱上回答逻辑查询。自然语言句子可以解释为逻辑公式，然后转换为集合算子树（Wang 等人，2021年）。

cuses（Choudhary 等人，2021a；Zhang 等，2021；Alivanistos 等人，2022；Bai 等，2022；Chen 等人，2022；Yang 等人，2022）。然而，集合之间的评分函数虽然同样刻画集合嵌入，并在模型训练中发挥着重要作用，但在现有文献中尚未得到充分研究。现有评分函数可分为两类，分别强调局部比较（Ren 和 Leskovec，2020；Amayuelas 等人，2022）或几何区域之间的全局传输（Ren 等人，2020；Choudhary 等人，2021b；Zhang 等，2021）。以下示例促使我们为嵌入开发兼顾局部性与全局性权衡的评分函数。

示例 1.1。考虑四个维度为 $d = 100$的“独热”向量：

$$A = [1, 0, 0, ..., 0], \quad (1)$$
$$B = [0, 1, 0, ..., 0], \quad (2)$$
$$C = [0, 0, 1, 0, ..., 0], \quad (3)$$
$$D = [0, \ldots, 0, 1]. \quad (4)$$

我们观察到：

- 局部函数（例如，欧几里得距离）L 无法区分 A、 B、 C和 D之间的不同相似性。具体而言， $L(A, B) = L(A, C) = L(A, D) = L(B, C) = L(B, D) = L(C, D) = 1$。
- 全局函数（例如 Wasserstein度量） G能够区分。具体而言， $G(A, B) = 1 < G(A, C) = 2 < G(A, D) = 99$。然而，G 对优化而言存在风险。例如，如果 $G(A, D) + G(A, B)$ 出现在某个批次的目标函数中，那么 $G(A, D)$ 将主导 $G(A, B)$，因为它大 100 倍，导致优化无效。
- 局部与全局权衡函数（例如本文提出的WFR评分函数）

通过将传输限制在窗口大小范围内，降低了这种风险。我们的论文发现，合适的窗口大小为 5，从而截断了 A 和 D等远距离样本之间的传输距离。然后，$WFR(A, D) = 5$，优化过程也因此趋于稳定。

本文受 Wasserstein-Fisher-Rao（WFR）度量（Chizat 等人，2018a）的启发，提出了一种更有效的评分函数，该度量引入了局部与全局之间的权衡。我们提出将集合嵌入为 $\mathbb{R}$ 中的有界测度，其中每个集合嵌入都可以离散化为定义在大小为 d 的均匀网格上的有界直方图。这种集合嵌入可以从局部角度进行解释，因此可以通过逐元素模糊逻辑 t-范数（Hájek，1998）轻松定义集合交、并集和否定。我们提出了一种高效的基于卷积的算法，可在 $O(d)$ 时间内实现熵正则化 WFR 的计算，并采用块对角核来实现局部与全局之间的权衡。我们在大量数据集上进行了广泛实验：(1) 三个知识图谱上的标准复杂查询回答数据集（Ren 和 Leskovec，2020）；(2) 强调模型在组合式复杂查询方面组合泛化能力的大规模评估集（Wang 等人，2021）；以及 (3) 分层知识图谱上的复杂查询（Huang 等，2022）。消融研究表明，通过在局部比较与全局传输之间选择更好的权衡，可以显著提升复杂查询回答的性能。

2 相关工作

本节讨论固定维度下的其他查询嵌入方法以及最优传输。复杂查询回答的其他方法

详见附录 A，

2.1 查询嵌入

作为回答逻辑查询的主流方式，查询嵌入（Hamilton 等，2018）将答案集嵌入连续空间，并利用神经网络对集合运算进行建模。查询嵌入方法可处理的逻辑查询范围不断扩大：从合取查询（Hamilton 等，2018），扩展到存在正一阶（EPFO）查询（Ren 等人，2020；Choudhary 等人，2021a,b），再到一阶（FO）查询（Ren 和 Leskovec，2020；Zhang 等，2021；Amayuelas 等，2022；Bai 等，2022；Yang 等人，2022；Wang 等人，2023；Yin 等人，2023；Bai 等人，2023）。

各种形式的集合嵌入已得到广泛研究，例如向量（Amayuelas 等，2022；Chen 等人，2022；Huang 等，2022）、几何区域（Ren 等人，2020；Zhang 等，2021；Choudhary 等人，2021b）以及概率分布（Ren 和 Leskovec，2020；Choudhary 等人，2021a；Yang 等人，2022）。尽管嵌入的形式多种多样，但其评分函数所捕获的要么是局部比较，例如欧几里得距离（Amayuelas 等人，2022）、内积（Chen 等人，2022）和 KL 散度（Ren 和 Leskovec，2020；Yang 等人，2022）；要么是全局传输，例如几何区域之间的启发式距离（Ren 等人，2020；Choudhary 等人，2021b；Zhang 等，2021）、马氏距离（Choudhary 等人，2021a），或目标粒子与点云中最近粒子之间的相似度（Bai 等，2022）。

在这项工作中，我们基于非平衡最优传输理论提出了一种新颖的评分函数（Chizat 等人，2018年a）。作为最优传输的一种变体，它继承了最优传输的优点，并兼顾局部比较与全局传输。

2.2 用于嵌入的最优传输

最优传输（OT）（Peyré 等人，2019年）引入了概率分布之间的幂度量，并促进了其在语言和图数据领域的诸多应用（Alvarez-Melis 和 Jaakkola，2018年；Zhao 等人，2020a；Xu 等人，2021年；Li 等人，2021年；Tang 等人，2022年；Wang 等人，2022a；Li 等人，2022年；Tan 等人，2023年）。将图顶点和词语嵌入为 Wasserstein 空间中的概率分布时，这种方法尤其高效（Muzellec 和 Cuturi，2018年；Frogner 等人，2018年）。

Wasserstein-Fisher-Rao（WFR）度量（Chizat 等人，2018a）通过引入传输半径，在局部比较和全局传输之间取得平衡，将分布之间的最优传输推广到了测度之间 η。已有研究（Zhao 等人，2020b）证明，WFR 度量是衡量嵌入对齐的稳健且有效的方法。以往的工作在 WFR 空间中测量预训练嵌入（Wang 等人，2020年），而本文首次在 WFR 空间中学习嵌入。此外，我们还在查询嵌入的背景下验证了 WFR 空间的优势。

3 预备知识

3.1 知识图谱与复杂查询

知识图谱 $\mathcal{KG} = \{(h,\ r,\ t) \in \mathcal{V} \times \mathcal{R} \times \mathcal{V}\}$是由三元组构成的集合，其中 $h,\ t \in \mathcal{V}$ 是实体节点，$r \in \mathcal{R}$是关系。

知识图谱上的复杂查询可以用一阶公式定义。遵循先前的研究（Ren 和 Leskovec，2020年），我们考虑一个查询 Q，它包含一个自由变量节点 $V_?$ 和量化节点 V_i，$1 \leq i \leq n$。任意逻辑公式都可以转换为前束范式和析取范式，如下所示（Marker，2002）。

$$Q[V_?] = \square V_1 \cdots \square V_n . c_1 \vee \cdots \vee c_l, \qquad (5)$$

其中，每个量词 $\square$ 要么是 $\exists$，要么是 $\forall$；每个c_i，$1 \leq i \leq l$ 都是满足$c_i = y_{i1} \wedge \cdots \wedge y_{im_i}$的合取子句；每个$y_{ij}$，$1 \leq j \leq m_i$表示一个原子公式或其否定。也就是说，$y_{ij} = r(a,\ b)$或 $\neg r(a,\ b)$，其中 $r \in \mathcal{R}$、a 和 b可以是变量 $V.$，也可以是 $\mathcal{V}$中的实体。

3.2 使用集合算子树回答查询

可通过集合算子回答的查询尤其值得关注（Ren 和 Leskovec，2020年）。答案可通过按自底向上顺序执行集合算子得到。每棵算子树的叶节点都是已知实体，可将其视为只含一个元素的集合。每个集合算子的输入和输出都是集合。我们注意到，由于存在额外假设以保证查询能够转换为运算树，可通过集合算子求解的查询仅是一阶查询的一个片段（Wang 等人，2021年）。此外，用于表示整个类别的集合算子并非唯一。本文聚焦于以下算子：

集合投影 由关系推导而来。**集合运算**：

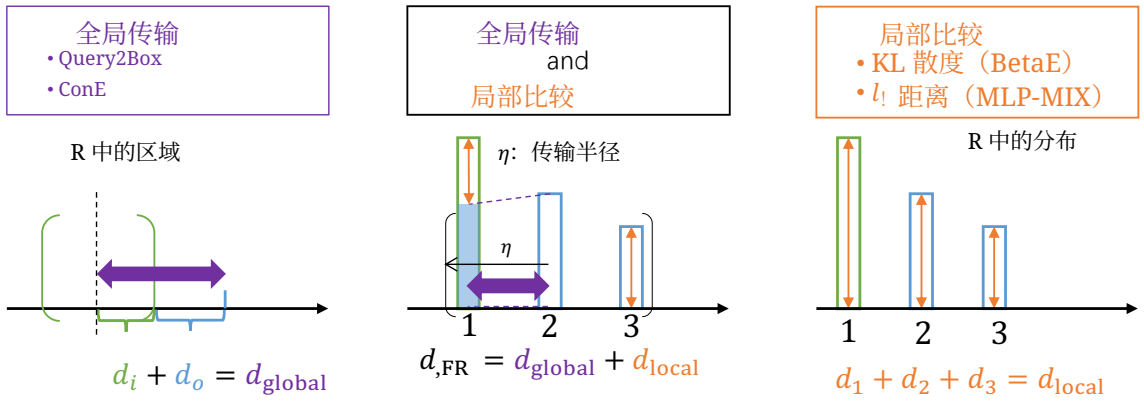

图 2：不同评分函数的示意图。**左侧**：全局传输，通过将质量从一处移至另一处的方式来衡量差异（紫色箭头）；**右侧**：局部比较，通过原位比较来衡量差异（黄色箭头）；**中间**：局部与全局之间的权衡，先在传输半径 η内移动质量，再比较未填充的质量。

集合交 由合取推导而来 .
集合并集 源自析取。**集合补集** 源自否定。

3.3 Wasserstein-Fisher-Rao（WFR）度量

Wasserstein-Fisher-Rao 度量定义了两个测度之间的距离（Chizat 等人，2018年a）。考虑 $\mathbb{R}^d$中的两个离散测度，即 $\mu = \sum_{i=1}^{M} u_i \delta_{x_i}$ 和 $\nu = \sum_{j=1}^{N} v_i \delta_{y_j}$，其中 δ 是狄拉克函数，u_i，$v_j \geq 0$，以及 x_i，$y_j \in \mathbb{R}^d$ 是 $1 \leq i \leq M$ 和 $1 \leq j \leq N$对应的坐标。简言之，$\mathbf{u} = [u_1, ..., u_M]^\top$和$\mathbf{v} = [v_1, ..., v_N]^\top$ 表示列质量向量。那么，WFR 度量可通过求解以下最小化问题来定义。

$$\mathcal{WFR}(\mu, \nu; \eta) = \min_{P \in \mathbb{R}_+^{M \times N}} J(P; \mu, \nu, \eta), \quad (6)$$

其中 $P \in \mathbb{R}^{M \times N}$ 是传输计划，且 P_{ij}表示从 x_i 到y_j的传输质量。我们将问题（6）的全局最小值称为 WFR 最优传输计划。目标函数如下：

$$J(P; \mu, \nu, \eta) = \sum_{i=1}^{M} \sum_{j=1}^{N} C_{ij} P_{ij} + \mathcal{D}(P \mathbb{1}_N \| \mathbf{u}) + \mathcal{D}(P^\top \mathbb{1}_M \| \mathbf{v}), \quad (7)$$

其中，$\mathbb{1}_N$ 是 $\mathbb{R}^N$ 中所有元素均为 1 的列向量，$\mathcal{D}(\cdot\|\cdot)$ 是 KL 散度。 $C \in \mathbb{R}_+^{M \times N}$是代价矩阵，$C_{ij}$ 表示从 x_i 到y_j的代价，

$$C_{ij} = -2 \log \left(\cos_+ \left(\frac{\pi}{2} \frac{\|x_i - y_j\|}{\eta} \right) \right). \quad (8)$$

其中，如果 $|x| < \pi/2$，则 $\cos_+(x) = \cos(x)$，否则 $\cos_+(x) = 0$。 η 是传输半径的超参数。

WFR 度量的一项关键性质可通过 WFR 空间中的测地线来理解，正如Chizat 等人(2018年a) 的定理 4.1 所述。具体而言，对于位于 x 和 y处的两个质量点，只有当 $\|x - y\| < \eta$时才会发生传输，例如图 2 中的位置 1 和 2；否则，仅计入局部比较。我们看到，η 控制着传输过程的范围。

3.4 熵正则化 WFR 解方案

公式（6）中的 WFR 度量可以通过 Sinkhorn 算法计算，并加入额外的熵正则化项（Chizat 等人，2018年b）。具体而言，可以通过以下熵正则化优化问题估计 WFR:

$$\min_{P \in \mathbb{R}_+^{M \times N}} J(P; \mu, \nu, \eta) + \overbrace{\epsilon \sum_{ij} P_{ij} \log P_{ij}}^{\text{Entropic Regularization}}. \quad (9)$$

广义 Sinkhorn 算法（Chizat 等人，2018年b）求解问题（9）的无约束对偶问题，该问题旨在最大化目标函数

$$D_\epsilon(\phi, \psi; \mathbf{u}, \mathbf{v}, K_\epsilon) = \langle 1 - \phi, \mathbf{u} \rangle + \langle 1 - \psi, \mathbf{v} \rangle + \epsilon \langle 1 - (\phi \otimes \psi)^{\frac{1}{\epsilon}}, K_\epsilon \rangle, \quad (10)$$

其中 $K_\epsilon = e^{-\frac{C}{\epsilon}}$是核矩阵，$\phi \in \mathbb{R}^M$ 和 $\psi \in \mathbb{R}^N$ 是对偶变量。 j-th Sinkhorn 的第 ($l + 1$) 步更新过程

迭代是

$$\phi^{(l+1)} \leftarrow \left[\mathbf{u} \oslash \left(K_\epsilon \psi^{(l)}\right)\right]^{\frac{1}{1+\epsilon}}, \tag{11}$$

$$\psi^{(l+1)} \leftarrow \left[\mathbf{v} \oslash \left(K_\epsilon^\top \phi^{(l+1)}\right)\right]^{\frac{1}{1+\epsilon}}. \tag{12}$$

设 ϕ^* 和 ψ^* 为通过收敛的 Sinkhorn 算法得到的最优对偶变量。最优传输计划可通过以下方式恢复：

$$P^* = \mathrm{diag}(\phi^*) K_\epsilon \mathrm{diag}(\psi^*). \tag{13}$$

我们可以看到，Sinkhorn 算法采用矩阵-向量乘法，其耗时为$O(MN)$。与 Wasserstein度量不同，后者可以通过一维切片 Wasserstein 近似（Carriere 等人，2017年；Kolouri 等人，2019年），耗时为 $O((M+N)\log(M+N))$。相比之下，目前尚无已知的次二次时间算法能够近似 WFR 度量，这限制了它在大规模场景中的应用。下一节中，我们将集合嵌入限制为 $\mathbb{R}$中的有界测度。我们还将利用核矩阵的稀疏结构，进一步开发一种 $O(d)$ 算法 K_ϵ。

4 Wasserstein-Fisher-Rao 嵌入

本节旨在介绍如何将集合嵌入作为 $\mathbb{R}$中的有界测度来处理复杂查询。设 S 为任意集合，其中包括仅含有一个实体 e的单元素集合 $\{e\}$，其嵌入为 $m[S]$。我们将所有有界测度的集合记为 $\mathcal{BM}(\mathbb{R})$。我们的讨论从将测度$m[S] \in \mathcal{BM}(\mathbb{R})$离散化为直方图 $m^S \in \mathcal{BM}_d$开始，其中 $\mathcal{BM}_d$ 表示含有d 个柱条的有界直方图集合。接着，我们讨论如何利用 $\mathcal{BM}_d$ 中的嵌入对集合算子进行参数化，并在 $\mathcal{BM}_d$中高效计算评分函数。最后，我们介绍如何学习集合嵌入和算子。

4.1 将 BM1D 离散化为直方图

我们将每个 $m[S] \in \mathcal{BM}(\mathbb{R})$ 离散化为定义在 $\mathbb{R}$ 上均匀网格中的直方图。不失一般性，直方图中柱条的最大长度为 1，网格间距为 Δ。通过这种方式，每个 $m[S] = \sum_{i=1}^d m_i^S \delta_{i\Delta}$，其中 $m_i^S \in [0,1]$ for $1 \le i \le d$。因此，只需存储离散化后的质量向量 $\mathbf{m}^S = [m_1^S, \ldots, m_d^S] \in \mathcal{BM}_d$，因为支撑集 $\{i\Delta\}_{i=1}^d$ 对所有 $m[S] \in \mathcal{BM}(\mathbb{R})$ 都是固定的。接下来，我们讨论 $\mathcal{BM}_d$ 上的集合算子

4.2 $\mathcal{BM}_d$上的集合算子

非参数集合运算 需要强调的是，质量向量 $\mathbf{m}^S \in \mathcal{BM}_d$ 可以进行局部解释，其中$\mathbf{m}^S$ 的每个元素都被视为模糊逻辑中的连续真值。因此，$\mathcal{BM}_d$ 上的集合运算**交集** $\cap$ **并集** $\cup$和**补集**，通过质量向量$\mathbf{m}^S$上的逐元素 t-范数进行建模。对于质量向量的第 i 个元素， $1 \le i \le d$,

$$\text{Intersection} \quad m_i^{S_1 \cap S_2} = m_i^{S_1} \top m_i^{S_2}, \tag{14}$$

$$\text{Union} \quad m_i^{S_1 \cup S_2} = m_i^{S_1} \perp m_i^{S_2}, \tag{15}$$

$$\text{Complement} \quad m_i^{S^C} = 1 - m_i^S, \tag{16}$$

其中， $\top$ 是一个 t范数， $\perp$ 是相应的t余范数。

神经集合投影 每个集合**投影** 都被建模为将一个质量向量映射到另一个质量向量的函数，映射由关系 r给定。我们采用基分解（Schlichtkrull 等人，2018年）来定义一个从 $[0,1]^d$ 到$[0,1]^d$的多层感知机（MLP）。对于每个输入为$\mathbf{m}^{S,(l)} \in [0,1]^{d_l}$的全连接层，经由关系 r 得到的输出 $\mathbf{m}^{S,(l+1)} \in [0,1]^{d_{l+1}}$计算如下：

$$\mathbf{m}^{S,(l+1)} = \sigma(W_r^{(l)} \mathbf{m}^{S,(l)} + b_r^{(l)}), \tag{17}$$

其中， σ 是激活函数， $W_r^{(l)}$和 $b_r^{(l)}$分别是第 l层关系 r对应的权重矩阵和偏置向量。具体而言，

$$W_r^{(l)} = \sum_{j=1}^K V_j^{(l)} r_j, \quad b_r^{(l)} = \sum_{j=1}^K a_j^{(l)} r_j. \tag{18}$$

K 是基的数量，$\mathbf{r} \in \mathbb{R}^K$ 是关系嵌入。 $V_j^{(l)} \in \mathbb{R}^{d_{l+1} \times d_l}$ 和 $a_j^{(l)} \in \mathbb{R}^{d_{l+1}}$ 分别是第 l层的基权重矩阵和基偏置向量。

集合补集上的随机失活 受能够提高泛化能力的神经网络随机失活技术启发，我们提出将随机失活应用于集合补集运算。其核心思想是在执行补集运算之前，随机改变质量向量中的元素，即随机将其值设为1_2。通过这种方式，取补后的元素也将为1_2。该技术能够提高集合补集算子的泛化能力。

4.3 $\mathcal{BM}_d$的评分函数

考虑 $\mathbf{m}^{S_1}$,$\mathbf{m}^{S_2} \in \mathcal{BM}_d$。通过 $\mathcal{WFR}(\mathbf{m}^{S_1}, \mathbf{m}^{S_2}; \eta)$ 对这一对进行评分非常直接。

然而，直接应用 Sinkhorn 算法需要 $O(d^2)$ 时间，这限制了 WFR 度量的大规模计算。在本节中，我们介绍：(1) 基于卷积的 Sinkhorn，将复杂度降至 $O(d)$ 时间；以及 (2) 块对角传输，作为传输半径之外的一种额外机制，用于权衡局部 与 全局 之间的关系 η。需要指出的是，我们的贡献不同于近期提出的线性时间“快速”Sinkhorn 算法（Liao 等人, 2022a,b），后者不适用于$\mathcal{BM}_d$中的非平衡最优传输。

基于卷积的 Sinkhorn 方程（11）和（12）所示的 Sinkhorn 更新，其计算瓶颈在于矩阵-向量乘法。当比较 中的离散化测度时， K_ϵ 呈现出对称和对角结构。

$$K_{\epsilon,ij} = \begin{cases} \cos\left(\frac{\pi}{2}\frac{|i-j|}{\eta/\Delta}\right)^{\frac{2}{\epsilon}} & |i-j| < \frac{\eta}{\Delta} \\ 0 & \text{o.w.} \end{cases} \quad (19)$$

设 $\omega = \lfloor \frac{\eta}{\Delta} \rfloor$ 为窗口大小，则矩阵-向量乘法 $K_\epsilon \mathbf{v} = K_\epsilon^\top \mathbf{v}$ 可简化为离散卷积 $H(\beta, \omega) * \mathbf{v}$，其中核函数 $[H(\beta, \omega)]_k = \cos\left(\frac{\pi\beta}{2\omega}k\right)$，$-\omega \le k \le \omega$ 且$\beta := \lfloor \overline{\frac{\eta}{\Delta}} \rfloor / \frac{\eta}{\Delta} \in (1 - \frac{1}{\omega+1}, 1]$是另一个超参数。具体而言， $H * \mathbf{v}$的第 i 个元素为

$$[H(\beta,\omega) * \mathbf{v}]_i = \sum_{k=-\omega}^{+\omega} H_k v_{i+k} \mathbf{1}_{1\le i+k\le d}, \quad (20)$$

其中， $\mathbf{1}_{1\le i+k\le d} = 1$ 当且仅当 $1 \le i+k \le d$。然后，Sinkhorn 算法可简化为

$$\phi^{(l+1)} \leftarrow \left[\mathbf{u} \oslash \left(H(\beta,\omega) * \psi^{(l)}\right)\right]^{\frac{1}{1+\epsilon}}, \quad (21)$$

$$\psi^{(l+1)} \leftarrow \left[\mathbf{v} \oslash \left(H(\beta,\omega) * \phi^{(l+1)}\right)\right]^{\frac{1}{1+\epsilon}}. \quad (22)$$

因此，Sinkhorn 算法的时间复杂度可降低至 $O(\omega d)$。在我们的设定中，ω 是窗口大小，用于在全局传输和局部比较之间进行插值；在所有设定中， β 均取为 1。

基于卷积的Sinkhorn算法收敛后，我们可以通过 D_ϵ 以及最优的 ϕ^* 和 ψ^*来近似 WFR 度量。对于复杂查询的答案，最终答案会根据其距离进行排序（距离越小越好）。这一过程可以通过基于 WFR 的原始-对偶剪枝加速，用于 k近邻（Wang 等人，2020年）或 Wasserstein 字典学习（Schmitz 等人，2018年）。

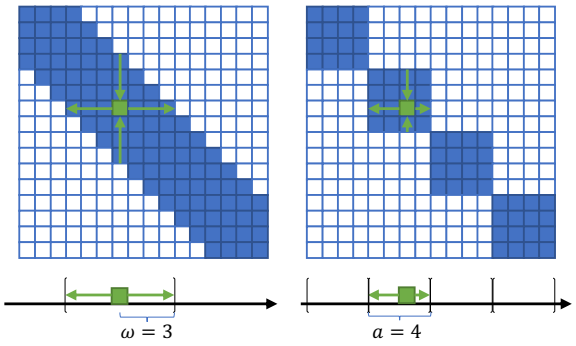

图3：两种机制生成的 16×16 传输计划矩阵示例。零元素以白色块表示，而（可能的）非零元素则以彩色表示。图中箭头展示了一个样本质量点（绿色块）的传输范围。左：WFR 传输以窗口大小 $\omega = 3$定义的相对范围；右：块大小为 $a = 4$的块对角核定义的绝对范围。

块对角传输 除了控制传输范围的窗口大小 ω（相对于每个质量点）之外，我们还提供了另一种机制，通过每个质量点的绝对位置来限制传输范围。具体而言，我们考虑由 K_ϵ^b 个 b块构成的块对角核矩阵，且 $a = d/b$ 表示每个对角块的大小。从公式（13）可以看出，块对角核会产生块对角传输计划。图3 通过可能的传输计划，展示了这两种限制全局传输机制之间的差异。

计算评分函数 我们提出如下评分函数定义：Dist 由采用块对角核的基于卷积的 Sinkhorn 算法计算。需要强调的是，大小为 9 的问题（ $d \times d$ ）可视为在块对角问题下求解 b 个相互独立、大小为 $a \times a$ 的问题。这一特性有助于进一步并行化 Sinkhorn 迭代（21）和（22）。我们假设 $a > \omega$，以确保每个块都至少包含一个窗口大小的 WFR 传输，从而使这两种机制能够协同工作。鉴于一维卷积具有并行特性，整个距离计算可通过图形处理器实现高度并行化。具体而言，评分函数 Dist见算法 1。

算法1 $\mathcal{BM}_d$ 上的评分函数（PyTorch-风格）

输入： 两个有界测度 $\mathbf{m}^{S_1}$，$\mathbf{m}^{S_2} \in \mathcal{BM}_d$，熵正则化 ϵ，窗口大小 ω, 块数 b，使得块大小为 $a = d/b \geq \omega$，迭代次数 L。

过程 Dist($\mathbf{m}^{S_1}, \mathbf{m}^{S_2}, \epsilon, \omega, a, b, L$)
    $M_1 \leftarrow \mathbf{m}^{S_1}$.reshape($1, b, a$)
    <escape>$M_2 \leftarrow$ </escape>$\mathbf{m}$<escape>^{S_2}</escape>.reshape(1,<escape> b</escape>,<escape> a </escape>).
    初始化 $H \leftarrow H(1, \omega)$。
    初始化 $\psi \leftarrow$ ones($1, b, a$).
    对于 $l = 1, ..., L$ **do**
        $\phi \leftarrow M_1$/一维卷积(ψ, H)。
        $\psi \leftarrow M_2$/一维卷积(ϕ, H)。
    结束循环
    $\phi^* \leftarrow \phi$.reshape(d))
    $\psi^* \leftarrow \psi$.reshape(d))
    <escape>return D_ϵ**</escape>**$(\phi^*, \psi^*; \mathbf{m}^{S_1}, \mathbf{m}^{S_2}, K_\epsilon^b)$。
结束过程

4.4 在 $\mathcal{BM}_d$中学习嵌入

设 $\mathbf{m}^Q \in \mathcal{BM}_d$ 为查询的查询嵌入$Q[V_?]$，而 $\mathbf{m}^e \in \mathcal{BM}_d$ 为幺正集合 $\{e\}$（其元素为 e）的集合嵌入。我们遵循 Ren 和 Leskovec (2020年) 中的做法，通过负采样训练参数化投影和嵌入。对于查询 Q，我们采样一个答案 a和 K_{neg} 个负样本 $\{v_k\}_{k=1}^{K_{\text{neg}}}$。目标函数为

$$L = -\log \sigma\left(\gamma - \rho \text{Dist}\left(\mathbf{m}^a, \mathbf{m}^Q\right)\right) - \sum_{k=1}^{K_{\text{neg}}} \frac{1}{K_{\text{neg}}} \log \sigma\left(\rho \text{Dist}\left(\mathbf{m}^{v_k}, \mathbf{m}^Q\right) - \gamma\right), \tag{23}$$

其中，γ 是间隔，ρ 是尺度，σ 是 Sigmoid 函数。

5 实验

本节从三个方面评估 WFRE 在复杂查询回答任务中的表现：(1) 在三个知识图谱上，使用常见数据集将 WFRE 与其他最先进水平的查询嵌入方法进行比较（Ren 和 Leskovec, 2020年）；(2) 在 301 种查询类型上评估 WFRE，以验证其组合泛化能力（Wang 等人, 2021年）；(3) 在 WordNet（Miller, 1995年）这一词汇知识图谱上训练并评估 WFRE，该图谱的关系通常具有层次结构（Huang 等人, 2022年）。方面 (2) 和 (3) 分别着重考察不同的查询类型和底层知识图谱。这些结果为 WFRE 能够应用于各种查询类型和知识图谱提供了实证依据。此外，我们还在消融研究中考察了 WFRE 在 局部与全局之间的权衡，包括 ω和 a。其他结果列于附录中。

5.1 实验设置

所有实验均遵循Ren 和 Leskovec（2020年）提出的训练与评估惯例。我们使用训练数据训练查询嵌入，在验证数据上选择超参数，并报告测试数据上的分数。训练与评估协议的详情见附录B。对于 WFRE，超参数列于附录C中并作了讨论。所有实验均使用 PyTorch（Paszke 等人，2019年）在一块配备 32G 内存的 V100 GPU 上进行。

5.2 基准数据集

Ren 和 Leskovec（2020年）提出的 FB15k-237（Bordes 等人，2013a）、FB15k（Toutanova 和 Chen，2015年）以及 NELL（Xiong 等人，2017b）数据集常用于评估查询嵌入方法的有效性。我们将 WFRE 与采用局部比较和全局传输的基线方法进行比较，包括 BetaE（Ren 和 Leskovec, 2020年）、ConE（Zhang 等，2021年）、MLP-MIX（Alivanistos 等人，2022年）、Q2P（Bai 等，2022）和 GammaE（Yang 等人，2022年）。为确保公平，我们将并集算子与 Ren 和 Leskovec（2020年）提出的 DNF处理进行比较，该方法将答案分数合并自包含这些答案的合取查询的分数。关于并集算子的其他处理方式见附录 E。关于数据集和基线方法的详细讨论见附录 D.1。表 1显示，与现有方法相比，WFRE 在包含逻辑否定和不包含逻辑否定的查询的平均分数方面均大幅领先。

5.3 查询的组合泛化能力

我们还探究了 WFRE 在一个用于评估查询嵌入方法组合泛化能力的基准测试中，对复杂查询组合空间的泛化能力（Wang 等人，2021年）。数据集的详细信息见附录 D.2

301 种不同查询类型的结果，按锚点节点数量和算子树最大深度取平均值，并以可视化形式呈现在

表1：在 FB15k、FB15k-237 和 NELL 上回答所有任务的平均倒数排名。基线方法的分数取自其原始论文。粗体表示最佳分数。 A_{P} 是不带否定的查询（EPFO 查询）的平均分数。 A_{N} 是带否定的查询的平均分数。

数据集	QE	1P	2P	3P	2I	3I	PI	IP	2U	UP	2IN	3IN	INP	PIN	PNI	A_{P}	A_{N}
FB15k	BetaE	65.1	25.7	24.7	55.8	66.5	43.9	28.1	40.1	25.2	14.3	14.7	11.5	6.5	12.4	41.6	11.8
	ConE	75.3	33.8	29.2	64.4	73.7	50.9	35.7	55.7	31.4	17.9	18.7	12.5	9.8	15.1	49.8	14.8
	MLPMIX	69.7	27.7	23.9	58.7	69.9	46.7	30.8	38.2	24.8	17.2	17.8	13.5	9.1	15.2	43.4	14.8
	Q2P	82.6	30.8	25.5	65.1	74.7	49.5	34.9	32.1	26.2	21.9	20.8	12.5	8.9	17.1	46.8	16.4
	GammaE	76.5	36.9	**31.4**	65.4	75.1	53.9	39.7	**53.5**	30.9	20.1	20.5	13.5	11.8	17.1	51.3	16.6
	WFRE	**81.1**	**37.7**	30.5	**68.5**	**78.0**	**56.3**	**41.8**	48.0	**33.1**	**26.1**	**26.5**	**15.6**	**13.7**	**19.4**	**52.8**	**20.2**
FB15k-237	BetaE	39.0	10.9	10.0	28.8	42.5	22.4	12.6	12.4	9.7	5.1	7.9	7.4	3.6	3.4	20.9	5.4
	ConE	41.8	12.8	11.0	32.6	47.3	25.5	14.0	14.5	10.8	5.4	8.6	7.8	4.0	3.6	23.4	5.9
	MLPMIX	42.4	11.5	9.9	33.5	46.8	25.4	14.0	14.0	9.2	6.6	10.7	8.1	4.7	4.4	22.9	6.9
	Q2P	39.1	11.4	10.1	32.3	47.7	24.0	14.3	8.7	9.1	4.4	9.7	7.5	4.6	3.8	21.9	6.0
	GammaE	43.2	13.2	11.0	33.5	47.9	27.2	15.9	13.9	10.3	6.7	9.4	**8.6**	4.8	**4.4**	24.0	6.8
	WFRE	**44.1**	**13.4**	**11.1**	**35.1**	**50.1**	**27.4**	**17.2**	**13.9**	**10.9**	**6.9**	**11.2**	**8.5**	**5.0**	4.3	**24.8**	**7.2**
NELL	BetaE	53.0	13.0	11.5	37.6	47.5	24.1	14.3	12.2	8.5	5.1	7.8	10.0	3.1	3.5	24.6	5.9
	ConE	53.1	16.1	13.9	40.0	50.8	26.3	17.5	15.3	11.3	5.7	8.1	10.8	3.5	3.9	27.2	6.4
	MLPMIX	55.4	16.5	13.9	39.5	51.0	25.7	18.3	14.7	11.2	5.1	8.0	10.0	3.1	3.5	27.4	5.9
	Q2P	56.5	15.2	12.5	35.8	48.7	22.6	16.1	11.1	10.4	5.1	7.4	10.2	3.3	3.4	25.5	6.0
	GammaE	55.1	17.3	14.2	**41.9**	51.1	26.9	18.3	15.1	11.2	6.3	8.7	11.4	4.0	**4.5**	27.9	7.0
	WFRE	**58.6**	**18.6**	**16.0**	41.2	**52.7**	**28.4**	**20.7**	**16.1**	**13.2**	**6.9**	**8.8**	**12.5**	**4.1**	4.4	**29.5**	**7.3**

表2：WN18RR 上不同查询嵌入方法的平均倒数排名分数。 A_{p} 是 1P、2P 和 3P 查询分数的平均值； A_{ℓ} 是不带否定的其他查询分数的平均值； A_{N} 是带否定的查询分数的平均值。分数取自Huang 等人（2022年）。

QE	1P	2P	3P	2I	3I	IP	PI	2IN	3IN	INP	PIN	PNI	2U	UP	A_{p}	A_{ℓ}	A_{N}
BetaE	44.13	9.85	3.86	57.19	76.26	17.97	32.59	12.77	59.98	5.07	4.04	7.48	7.57	5.39	19.28	32.83	17.87
LinE	45.12	12.35	6.70	47.11	67.13	14.73	24.87	12.50	60.81	7.34	5.20	7.74	8.49	6.93	21.39	28.21	18.72
WFRE	52.78	21.00	15.18	68.23	88.15	26.50	40.97	18.99	69.07	14.29	11.06	11.01	15.14	15.81	29.65	42.46	24.88

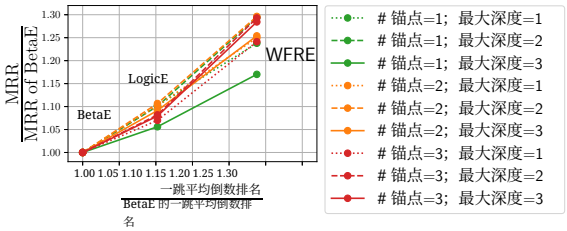

图4：不同查询嵌入方法在组合泛化能力基准测试中的可视化结果（Wang 等人，2021年）。BetaE 和 Log- icE 的结果取自Wang 等人 （2021年）。直线的斜率表示复杂查询的性能如何随单跳查询的性能提升而增长。

图4。为了展示模型对复杂查询的组合泛化能力，我们根据坐标轴标签所示，将各查询类型的分数归一化为相对于 BetaE 的分数（见图4）。随后，我们按照锚点节点数量和最大深度，将结果绘制成折线。同一模型的分数位于同一条垂直线上。我们发现，与 LogicE（Luus 等人，2021年）相比，WFRE 不仅显著提升了性能，而且在组合复杂查询上的泛化能力也更强，表现为更大的斜率。

5.4 层次化知识图谱上的复杂查询

上述评估仅限于三种常用知识图谱。接下来，我们转向另一类底层知识图谱，其特点是关系的层次结构。我们在Huang 等人 (2022年) 提出的 Word- Net（Miller，1995年）复杂查询数据集上训练并评估 WFRE。该数据集的详细信息见附录D.3。我们将 WFRE 与 LinE（Huang 等人，2022年）进行比较。LinE 是另一种基于直方图的查询嵌入方法，旨在解决层次型知识图谱上的查询，但不使用全局传输。表2展示了 WR18RR 上的结果。可以看出，WFRE 的表现显著优于 LinE 和 BetaE。尤其是在较长的多跳查询（即 1P、2P 和 3P 查询）上，WFRE 显著提升了 BetaE 和 LinE 的性能。值得强调的是，LinE 与 WFRE 一样使用了直方图，但其训练所用的评分函数仅以局部比较为依据。这一结果表明，WFRE 适合建模层次关系，因为评分函数中局部与全局之间的权衡使 WFRE 学到了更好的嵌入。这也证实了 Wasserstein 空间能够提高嵌入的效率（Frogner 等人，2018年）。

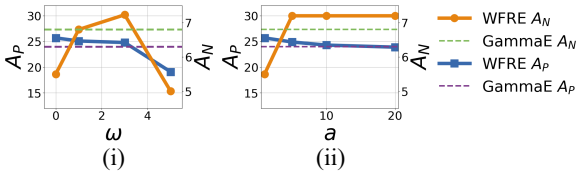

图5：超参数 ω 和 a的影响。默认选择为（ω、 a）＝（3、5）。

5.5 局部与全局权衡

我们进一步研究限制传输的两种机制，即传输窗口大小 ω和块大小 a 对性能的影响。我们在 Ren 和 Leskovec(2020年) 抽取的 FB15k-237 查询上进行了实验。我们改变其中一个值，并固定另一个值。默认设置为$(\omega, a) = (3, 5)$。图5 展示了这两个超参数的影响。

与最新的最先进水平查询嵌入方法 GammaE（Yang 等人，2022年）相比，结果证实了局部比较与全局传输之间权衡的重要性。当块大小 $a = 5$时，我们发现，较大的窗口大小 ω会损害否定查询的性能。与此同时，不带否定的查询（EPFO 查询）的性能在适当选择 $\omega = 3$时达到最大值。当窗口大小固定为$\omega = 3$ 且 a 较小时，我们看到 EPFO 查询和否定查询的性能遵循我们关于窗口大小的观察结果。进一步增大块大小 a 对 EPFO 查询的影响很小，但也会损害否定查询的性能。这表明，当 ω 固定时，适当的 a 对于性能而言是必要的。这一观察结果有助于提高基于卷积的 Sinkhorn 算法的并行化程度。

6 结论

本文提出了 WFRE，这是一种面向知识图谱复杂查询的新型查询嵌入方法。与以往方法相比，WFRE 的关键特点在于其评分函数能够平衡局部比较与全局传输。实证结果表明，WFRE 是复杂查询回答领域的最先进查询嵌入方法，并且对组合复杂查询和层次化知识图谱具有良好的泛化能力。消融研究验证了局部与全局权衡的重要性。

7 局限性

WFRE 存在现有查询嵌入方法的常见缺陷。这些方法能够解决的查询仅是一阶查询的有限子类。在归纳设置下，如何将 WFRE 应用于未见过的实体和关系也尚不明确。

8 伦理声明

作为一种查询嵌入方法，WFRE 对不同查询类型和知识图谱具有更强的泛化能力。本文中的实验和评估不涉及任何伦理问题，甚至与任何人类实体都无关。WFRE 可能被用于从工业级知识图谱中高效推断隐私信息。这是针对数据不完整性和链接预测的方法普遍存在的潜在风险。

致谢

本文作者获得了以下项目的资助：中国国家自然科学基金委员会的国家自然科学基金（U20B2053）、香港研究资助局的研究影响基金（R6020-19 和 R6021-20）及通用研究基金（16211520 和 16205322）、香港创新科技署的医疗卫生及健康科技基金（MHP/001/19），以及国家重点研发计划（2019YFE0198200）。特别感谢香港多媒体及人工智能协会和 CUSBLT。此外，我们还感谢 NVIDIA 人工智能技术中心（NVAITC）以及大学教育资助委员会研究配对补助金（RMGS20EG01-D、RMGS20CR11、RMGS20CR12、RMGS20EG19、RMGS20EG21、RMGS23CR05、RMGS23EG08）提供的支持。

参考文献

Dimitrios Alivanistos、Max Berrendorf、Michael Cochez 和 Mikhail Galkin。2022年。超关系知识图谱上的查询嵌入。载于国际学习表征会议。

David Alvarez-Melis 和 Tommi Jaakkola。2018年。词嵌入空间的 Gromov-Wasserstein 对齐。载于 *2018*年自然语言处理实证方法会议论文集，第 1881–1890 页，比利时布鲁塞尔。计算语言学协会。

Alfonso Amayuelas、Shuai Zhang、Susie Xi Rao 和 Ce Zhang。2022年。《知识图谱上的逻辑推理神经方法》。载于第十届国际学习表征会议，*ICLR 2022*，线上会议，*2022*年*4*月*25*日至*29*日。Open-Review.net。

Erik Arakelyan、Daniel Daza、Pasquale Minervini 和 Michael Cochez。2021年。《使用神经链接预测器回答复杂查询》。载于国际学习表征会议。

Jiaxin Bai、Zihao Wang、Hongming Zhang 和 Yangqiu Song。2022 年。Query2Particles：利用粒子嵌入进行知识图谱推理。载于计算语言学协会会议论文集：*NAACL 2022*，第2703–2714页，美国西雅图。计算语言学协会。

Jiaxin Bai、Tianshi Zheng 和 Yangqiu Song。2023年。用于知识图谱复杂查询回答的序列查询编码。*arXiv*预印本 *arXiv:2302.13114*。

Kurt D. Bollacker、Colin Evans、Praveen K. Paritosh、Tim Sturge 和 Jamie Taylor。2008年。Freebase：一个由协作创建、用于组织人类知识的图数据库。载于*ACM SIGMOD*数据管理国际会议论文集，*SIGMOD 2008*，加拿大不列颠哥伦比亚省温哥华，*2008*年*6*月*10*日至*12*日，第1247–1250页。ACM。

Antoine Bordes、Nicolas Usunier、Alberto Garcia-Duran、Jason Weston 和 Oksana Yakhnenko。2013a。用于多关系数据建模的翻译嵌入。载于神经信息处理系统进展，第 26 卷。Curran Associates, Inc.

Antoine Bordes、Nicolas Usunier、Alberto García- Durán、Jason Weston 和 Oksana Yakhnenko。2013b。用于建模多关系数据的嵌入翻译。载于神经信息处理系统进展 *26*：第 *27* 届神经信息处理系统年会（*2013*）。会议论文集，*2013* 年 *12* 月 *5* 日至 *8* 日于美国内华达州太浩湖举行，第 2787–2795 页。

Andrew Carlson、Justin Betteridge、Bryan Kisiel、Burr Settles、Estevam R. Hruschka Jr 和 Tom M. Mitchell。2010。《迈向永不止息的语言学习架构》。载于 第二十四届人工智能大会论文集，*AAAI 2010*，美国佐治亚州亚特兰大，*2010*年*7*月*11*日至*15*日。AAAI出版社。

Mathieu Carriere、Marco Cuturi 和 Steve Oudot。2017年。用于持久性图的切片Wasserstein核。在国际机器学习会议上，第664–673页。PMLR。

Xuelu Chen、Ziniu Hu 和 Yizhou Sun。2022年。基于模糊逻辑的知识图谱逻辑查询回答。载于：第三十六届*AAAI*人工智能会议，*AAAI 2022*，第三十四届人工智能创新应用会议，*IAAI 2022*，人工智能教育进展研讨会第十二届会议，*EAAI 2022*线上会议，*2022*年*2*月*22*日至*3*月*1*日，第3939–3948页。AAAI出版社。

LenaicChizat、GabrielPeyré、BernhardSchmitzer 和 François-Xavier Vialard。2018年。最优传输与 Fisher–Rao度量之间的一种插值距离。《计算数学基础》，18(1):1–44。

Lenaic Chizat、Gabriel Peyré、Bernhard Schmitzer 和 François-Xavier Vialard。2018b。非平衡最优传输问题的缩放算法。计算数学，87(314):2563–2609。

Nurendra Choudhary、Nikhil Rao、Sumeet Katariya、Karthik Subbian 和 Chandan Reddy。2021a。用于知识图谱推理的概率实体表示模型。在神经信息处理系统进展中，第34卷，第23440–23451页。Curran Associates, Inc.。

Nurendra Choudhary、Nikhil Rao、Sumeet Katariya、Karthik Subbian 和 Chandan K. Reddy。2021b。基于知识图谱逻辑查询的自监督双曲面表示。载于*2021*年万维网大会论文集，WWW '21，第1373–1384页，纽约州纽约市，美国。计算机协会。

Daniel Daza 和 Michael Cochez。2018年。消息传递查询嵌入。

Charlie Frogner、Farzaneh Mirzazadeh 和 Justin Solomon。2018年。学习嵌入到熵 Wasserstein空间。载于国际学习表征会议。

Albert Gu、Frederic Sala、Beliz Gunel 和 Christopher Ré。2018年。在乘积空间中学习混合曲率表示。载于国际学习表征会议。

Petr Hájek。1998年。《模糊逻辑元数学》，第4卷，《逻辑学趋势》。Springer Netherlands，多德雷赫特。

William L. Hamilton、Payal Bajaj、Marinka Zitnik、Dan Jurafsky 和 Jure Leskovec。2018年。在知识图谱上嵌入逻辑查询。载于神经信息处理系统进展*31*：神经信息处理系统年度会议 *2018*，神经信息处理系统大会 *2018*，*2018*年*12*月*3*日至*8*日，加拿大蒙特利尔，第2030–2041页。

ZijianHuang、Meng-FenChiang 和 Wang-Chien Lee。2022年。LinE：层次化知识图谱上的逻辑查询推理。收录于第 *28* 届 *ACM SIGKDD* 知识发现与数据挖掘会议论文集，KDD '22，第 615–625 页，纽约，纽约州，美国。计算机协会。

Soheil Kolouri、KimiaNadjahi、UmutSimsekli、Roland Badeau 和 Gustavo Rohde。2019年。广义切片Wasserstein距离。神经信息处理系统进展，32。

Bhushan Kotnis、Carolin Lawrence 和 Mathias Niepert。2021年。使用双向序列编码器回答知识图谱中的复杂查询。*AAAI*人工智能会议论文集，35(6):4968–4977。

Manling Li、Tengfei Ma、Mo Yu、Lingfei Wu、Tian Gao、Heng Ji 和 Kathleen McKeown。2021年。基于事件图压缩和时间感知最优传输的时间线摘要。载于：*2021*年自然语言处理实证方法会议论文集，第 6443–6456 页，线上及多米尼加共和国 Punta Cana。计算语言学协会。

Xinhang Li、Zhaopeng Qiu、Xiangyu Zhao、Zihao Wang、Yong Zhang、Chunxiao Xing 和 Xian Wu。2022年。Gromov-Wasserstein 引导的跨域推荐表示学习。载于：第 *31* 届 *ACM* 信息与知识管理国际会议论文集，第 1199–1208 页。

Qichen Liao、Jing Chen、Zihao Wang、Bo Bai、Shi Jin 和 Hao Wu。2022a。Fast Sinkhorn I：一种用于 Wasserstein-1度量的 $o(n)$ 算法。数学科学通讯，20(7)：2053–2067。

Qichen Liao、Zihao Wang、Jing Chen、Bo Bai、Shi Jin 和 Hao Wu。2022b。Fast Sinkhorn II：共线三角矩阵与最优传输的线性时间精确计算。*ar Xiv* 预印本 *arXiv:2206.09049*。

Leonid Libkin 和 Cristina Sirangelo。2009。数据交换中的开放世界与封闭世界假设。载于第 *22* 届国际描述逻辑研讨会论文集（*DL 2009*），英国牛津，*2009* 年 *7* 月 *27—30* 日，第 477 卷，收录于*CEUR* 研讨会论文集。CEUR-WS.org。

Xuan Lin、Zhe Quan、Zhi-Jie Wang、Tengfei Ma 和 Xiangxiang Zeng。2020年。KGNN：用于药物-药物相互作用预测的知识图谱神经网络。载于 *IJCAI*，第 380 卷，第 2739–2745 页。

Xiao Liu、Shiyu Zhao、Kai Su、Yukuo Cen、Jiezhong Qiu、Mengdi Zhang、Wei Wu、Yuxiao Dong 和 Jie Tang。2022年。Mask and Reason：面向复杂逻辑查询的知识图谱变换器预训练。载于第 *28* 届 *ACM SIGKDD*知识发现与数据挖掘会议论文集，KDD ’22，第 1120–1130 页，美国纽约。计算机协会。

Francois Luus、Prithviraj Sen、Pavan Kapanipathi、RyanRiegel、Ndivhuwo Makondo、Thabang Lebese 和 Alexander Gray。2021年。用于复杂查询回答的逻辑嵌入。*arXiv:2103.00418[cs]*。

D. Marker。2002年。模型论：入门。数学研究生教材第217卷。Springer，NewYork。

George A. Miller。1995年。WordNet：英语词汇数据库。*ACM*通信，38(11)：39–41。Boris Muzellec和Marco Cuturi。2018年。使用椭圆分布的 Wasserstein 空间推广点嵌入。神经信息处理系统进展，31。

Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Te-jani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。2019年。PyTorch：一种命令式风格的高性能深度学习库。载于神经信息处理系统进展 *32*，第 8024–8035 页。Curran Associates, Inc.。

Gabriel Peyré、Marco Cuturi 等。2019年。计算最优传输：及其在数据科学中的应用。《机器学习基础与趋势®》，11(5-6):355–607。

Hongyu Ren、Hanjun Dai、Bo Dai、Xinyun Chen、Michihiro Yasunaga、Haitian Sun、Dale Schuurmans、Jure Leskovec 和 Denny Zhou。2021年。LEGO：用于知识图谱多跳问答的潜在执行引导推理。发表于第*38*届国际机器学习会议论文集，第8959–8970页。PMLR。

Hongyu Ren、Hanjun Dai、Bo Dai、Xinyun Chen、Denny Zhou、Jure Leskovec 和 Dale Schuurmans。2022年。SMORE：大规模知识图谱中的知识图谱补全与多跳推理。发表于第*28*届 *ACM SIGKDD* 知识发现与数据挖掘会议论文集，KDD ’22，第1472–1482页，美国纽约。计算机协会。

Hongyu Ren、Mikhail Galkin、Michael Cochez、Zhaocheng Zhu 和 Jure Leskovec。2023年。神经图推理：复杂逻辑查询回答与图数据库相遇。*ar Xiv* 预印本 *arXiv:2303.14617*。

Hongyu Ren、Weihua Hu 和 Jure Leskovec。2020年。Query2Box：使用盒嵌入在向量空间中对知识图谱进行推理。载于第*8*届国际学习表征会议，*ICLR 2020*，亚的斯亚贝巴，埃塞俄比亚，*2020*年*4*月*26*日至*30*日。OpenReview.net。

Hongyu Ren 和 Jure Leskovec。2020年。用于知识图谱多跳逻辑推理的 Beta 嵌入。载于神经信息处理系统进展第*33*卷：*2020*年神经信息处理系统年会，*NeurIPS 2020*，*2020*年*12*月*6*日至*12*日，线上。

Michael Schlichtkrull、Thomas N. Kipf、Peter Bloem、Riannevan den Berg、Ivan Titov 和 Max Welling。

2018年。 使用图卷积网络对关系数据进行建模。载于 语义网，《计算机科学讲义》，第 593–607 页，尚贝里。施普林格国际出版。

Morgan A Schmitz、Matthieu Heitz、Nicolas Bonneel、Fred Ngole、David Coeurjolly、Marco Cuturi、Gabriel Peyré 和 Jean-Luc Starck。2018年。Wasserstein 字典学习：基于最优传输的无监督非线性字典学习。*SIAM*影像科学期刊，11(1):643–678。

Zhiquan Tan、Zihao Wang 和 Yifan Zhang。2023年。Seal：同时探索和学习标签层级。*arXiv*预印本 *arXiv:2304.13374*。

Peggy Tang、Kun Hu、Rui Yan、Lei Zhang、Junbin Gao 和 Zhiyong Wang。2022年。OTExtSum：基于最优传输的抽取式文本摘要。载于计算语言学协会论文集：*NAACL 2022*，第1128–1141页，美国西雅图。计算语言学协会。Kristina Toutanova 和 Danqi Chen。2015年。知识库与文本推理中的观测特征和潜在特征。载于第三届连续向量空间模型及其组合性研讨会论文集，第57–66页，中国北京。计算语言学协会。Denny Vrandeˇci´c 和 Markus Krötzsch。2014年。维基数据：一个自由协作的知识库。《*ACM*通讯》，57(10)：78–85。Xiang Wang、Xiangnan He、Yixin Cao、Meng Liu 和 Tat-Seng Chua。2019年。KGAT：用于推荐的知识图谱注意力网络。载于第*25*届*ACM SIGKDD*国际知识发现与数据挖掘会议论文集，第950–958页。Zihao Wang、Jiaheng Dou 和 Yong Zhang。2022a。基于组合短语语义的无监督句子文本相似度。载于第*29*届国际计算语言学会议论文集，第4976–4995页。Zihao Wang、Yangqiu Song、Ginny Y Wong 和 Simon See。2023年。基于原子公式单跳推理的逻辑消息传递网络。*arXiv*预印本 *arXiv:2301.08859*。Zihao Wang、Hang Yin 和 Yangqiu Song。2021年。知识图谱复杂查询回答的组合泛化能力基准测试。载于神经信息处理系统大会数据集与基准赛道论文集第*1*卷，神经信息处理系统大会数据集与基准 *2021*，*2021*年*12*月，线上。Zihao Wang、Hang Yin 和 Yangqiu Song。2022b。知识图谱上的逻辑查询：不完整关系数据的新兴接口。《数据工程》，第3页。

Zihao Wang、Datong Zhou、Ming Yang、Yong Zhang、Chenglong Rao 和 Hao Wu。2020年。基于 Wasserstein-Fisher-Rao 度量的稳健文档距离。载于亚洲机器学习会议，第721–736页。PMLR。

Chenyan Xiong、Russell Power 和 Jamie Callan。2017a。通过知识图谱嵌入实现学术搜索的显式语义排序。载于第*26*届国际万维网大会，第1271–1279页。

Wenhan Xiong、Thien Hoang 和 William Yang Wang。2017b。DeepPath：一种用于知识图谱推理的强化学习方法。载于*EMNLP*。

Jingjing Xu、Hao Zhou、Chun Gan、Zaixiang Zheng 和 Lei Li。2021年。通过最优传输进行神经机器翻译的词汇学习。载于第*59*届计算语言学协会年会暨第*11*届国际自然语言处理联合会议论文集（第*1*卷：长篇论文），第7361–7373页，线上。计算语言学协会。

Zezhong Xu、Wen Zhang、Peng Ye、Hui Chen 和 Hua-jun Chen。2022年。用于复杂查询回答的神经符号纠缠框架。载于神经信息处理系统进展。

Dong Yang、Peijun Qing、Yang Li、Haonan Lu 和 Xiaodong Lin。2022年。GammaE：用于知识图谱逻辑查询的 Gamma 嵌入。

Hang Yin、Zihao Wang 和 Yangqiu Song。2023年。关于知识图谱上存在一阶查询的推理。*arXiv* 预印本 *arXiv:2304.07063*。

Zhanqiu Zhang、Jie Wang、Jiajun Chen、Shuiwang Ji 和 Feng Wu。2021年。ConE：用于知识图谱多跳推理的锥嵌入。载于神经信息处理系统进展，第34卷，第19172–19183页。Curran Associates, Inc.

Xu Zhao、Zihao Wang、Hao Wu 和 Yong Zhang。2020a。基于双向交互的半监督双语词典归纳。载于*2020*年自然语言处理经验方法会议论文集（*EMNLP*），第2973–2984页。

Xu Zhao、Zihao Wang、Yong Zhang 和 Hao Wu。2020b。用于无监督 bli 的宽松匹配过程。载于计算语言学协会第*58*届年会论文集，第3036–3041页。

Zhaocheng Zhu、Mikhail Galkin、Zuobai Zhang 和 JianTang。2022年。面向知识图谱逻辑查询的神经符号模型。

A 其他复杂查询答案方法

尽管使用神经集合算子计算查询嵌入，但也有其他方法被提出用于推导答案。 Daza 和 Cochez (2020年); Liu 等 (2022年探索了图表示，利用图神经网络回答逻辑查询；而 Kotnis 等 (2021年则将逻辑查询表示为序列。Arakelyan 等 (2021 年通过求解神经链接预测器所引出的连续优化问题来求解逻辑查询。然而，这些讨论仅限于不含 逻辑否定的 EPFO 查询。目前尚不清楚这些方法如何处理一阶查询。

与此同时，神经符号方法会估计每个实体是答案集的概率（即使在每个中间步骤也是如此）（Zhu 等人，2022年；Xu 等人，2022 年）。因此，要为给定查询推导答案，需要$O(|\mathcal{V}| + |\mathcal{T}|)$ 的空间和时间，其中 $\mathcal{V}$ 和 $\mathcal{T}$分别是知识图谱的实体集和三元组集。与仅需要 $O(d)$的查询嵌入方法相比，其中 d 是嵌入空间的固定维度，将神经符号方法扩展到大规模知识图谱上的逻辑查询颇具挑战性（Ren 等，2022年）。

B 训练与评估协议

我们遵循 EFO-1 查询回答中常用的实验设置，旨在在不完整图中找到非平凡答案，并泛化到未见类型的查询。

给定一个底层知识图谱 $\mathcal{G} = (\mathcal{V}, \mathcal{R})$ 及其三元组集 $\mathcal{T}$，我们通过改变三元组 $\mathcal{T}_{\text{train}} \subset \mathcal{T}_{\text{valid}} \subset \mathcal{T}_{\text{test}} = \mathcal{T}$ 的范围来采样三个子图。遵循标准评估协议，我们旨在找出无法通过遍历图直接发现的非平凡答案。我们将查询 q 在训练图中的答案集记为 $[q]_{\text{train}}$，我们关注的答案集为 $[q]_{\text{test}} \backslash [q]_{\text{train}}$，这些都是只能通过推理或预测得出的简单答案。困难答案为 $[q]_{\text{test}} \setminus [q]_{\text{valid}}$。随后，我们将简单答案（困难答案）与所有非答案集 $\mathcal{V}/[q]_{\text{valid}}$（$\mathcal{V}/[q]_{\text{test}}$）进行排序。得到排名 r 后，我们计算平均倒数排名（MRR）：$\frac{1}{r}$ 和 Hits@K（Hits@K）：$\mathbb{1}_{r<K}$，作为衡量模型性能的指标。

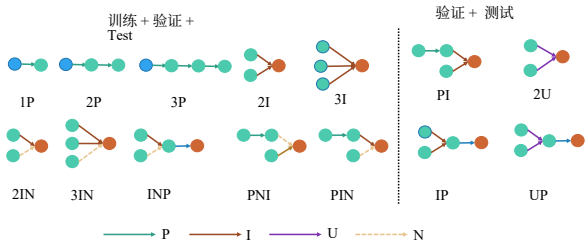

图6：逻辑查询结构的可视化。左侧的查询仅出现在训练阶段，所有查询均用于验证和测试阶段。

C WFRE 的设置

我们的框架使用 PyTorch 实现。我们的代码基于 EFO-1-QA 基准的处理流程（Wang 等人，2021年），并使用 AdamW 作为优化器。

代码中还有一些超参数。我们在投影网络上应用丢弃，并将丢弃概率记为 Drop$_p$。Sinkhorn算法的最大迭代次数记为 K_S。根据实验结果，我们将投影多层感知机的层数设为 1。WFRE 中的超参数及其相关信息列于表 3。我们针对四个数据集微调了超参数，结果见表4。希望这两张表能帮助你快速了解我们模型的超参数。

D 数据集和基线方法

本节介绍三个实验中的基线方法。表 5 展示了所有基准数据集上不同查询的基本统计信息。

D.1 基准数据集

对于常用数据集（Ren 和 Leskovec，2020年），训练数据集中包含 1P、2P、3P、2I、3I、2IN、3IN、INP、PNI、PIN 这十种查询类型，验证集和测试数据集中还包含 IP、PI、2U 和 UP 这四种未见过的查询结构。相关查询结构如图 6 所示。在验证集和测试查询中引入未见类型，旨在测试神经集合算子的组合泛化能力。

在本节中，我们选择以下支持任意 EFO1 查询的复杂查询嵌入方法：

BetaE (Ren 和 Leskovec,2020年) 基于 Beta 分布的嵌入，其评分函数为

表3：WFRE使用的超参数

超参数	评论	选项
学习率	模型收敛	$\{0.0001, 0.0005, 0.001\}$
训练步数	模型收敛	$\{240000, 300000, 360000\}$
负样本数量 K_{neg}	模型收敛	$\{32\}$
权重衰减	模型正则化	$\{0.001, 0.005, 0.01\}$
Drop_p	投影操作的正则化	$\{0.05\}$
Drop_n	否定操作的正则化	$\{0.05, 0.15, 0.25\}$
实体维度 d	实体表示	$\{400, 800, 1600\}$
关系基数量 K	关系表示	$\{70, 90, 120\}$
间隔 γ	损失函数	$\{37.5\}$
尺度 ρ	损失函数	$\{90, 120, 150\}$
对角块大小 a	实体表示	{5, 10, 20}
窗口大小 ω	WFR 距离的传输区域	$\{1, 3, 5\}$
Sinkhorn 正则化 ϵ	WFR 距离的熵正则化	$\{0.1\}$
Sinkhorn 最大迭代次数 K_S	Sinkhorn 算法的收敛性	$\{10, 15, 30\}$

表 4：每个数据集上的最佳超参数

	学习率	训练步数	K_{neg}	权重衰减	丢弃$_p$	丢弃$_n$	d	K	γ	ρ	a	ω	ϵ	K_S
FB15k	0.0005	360000	32	0.01	0.05	0.1	1600	90	37.5	150	5	3	0.1	10
FB15k-237	0.0005	240000	32	0.01	0.05	0.1	1600	120	37.5	120	5	3	0.1	10
NELL	0.0005	240000	32	0.01	0.05	0.1	1600	70	37.5	180	5	3	0.1	10
WN18RR	0.001	120000	32	0.01	0.05	0.1	800	70	37.5	120	5	3	0.1	15

基于 KL 散度进行局部比较

GammaE（Yang 等人，2022年）其评分函数基于使用 KL 散度进行局部比较的 GammaE 分布嵌入。**ConE（Zhang 等人，2021年）**二维锥嵌入，其评分函数基于通过锥之间的旋转距离进行全局传输。**MLP-MIX（Amayuelas 等人，2022年）**向量嵌入，其评分函数基于使用欧几里得距离进行局部比较。**Q2P（Bai 等人，2022年）**多粒子嵌入，其评分函数基于通过将目标粒子与点云中的最近粒子进行比较来实现全局传输。

这些分数直接取自相应的论文。

D.2 查询上的组合泛化能力

Wang 等人 (2021年) 提出了一个新数据集，其中包含 301 种不同的查询类型，用于评测复杂查询回答模型的组合泛化能力。基于由 OpsTree 表示的 EFO-1 查询，通过实体、投影、交集、并集和否定等操作生成 EFO-1 公式。为了使查询更加贴近现实，投影/否定链的最大长度以及锚点节点的数量都限制为不超过 3。基线方法为 BetaE 和 LogicE（Luus 等人，2021年）。分数直接取自 Wang 等人 (2021年)。

D.3 层次化知识图谱上的复杂查询

<escape>WN18RR 最初被引入为一个从 WN18 创建的链接预测数据集（Bordes 等人，2013年），而 WN18 是 WordNet 的一个子集。WN18RR 包含 93,003 个三元组、40,943 个实体和 11 种关系类型，其中大多数关系具有层次性。在表 6 中，我们可以看到，11 种关系中有 7 种具有较高的反对称性 $Khs_{\mathcal{G}_r}$ Huang 等人(2022年）和负传递分数 $\xi_{\mathcal{G}_r}$ (Gu 等人，2018年），因此被视为层次关系。Huang 等人 (2022年）将复杂逻辑查询扩展到 WN18RR，详细的查询统计数据见表 5。 Huang 等人(2022年）从层次化知识图谱 WN18RR 中生成了 14 种类型的查询，旨在研究查询嵌入在层次知识图谱中的推理能力。我们选择 BetaE 和 LinE（Huang 等人，2022年）作为基线方法，其得分也取自 Huang 等人 (2022年）。值得注意的是，LinE（Huang 等人，2022年）也是一种基于相同闭式集合运算的基于直方图的查询嵌入方法。LinE 与 WFRE 的主要区别在于，WFRE 鼓励局部与全局权衡。</escape>

表5：针对不同查询结构生成的训练、验证和测试查询数量 .

数据集	训练		验证		Test	
	1P/2P/3P/2I/3I	2IN/3IN/INP/PIN/PNI	1P	其他	1P	其他
FB15k	273,710	27,371	59,097	8,000	67,016	8,000
FB15k-237	149,689	14,968	20,101	5,000	22,812	5,000
NELL995	107,982	10,798	16,927	4,000	17,034	4,000
WN18RR	103,509	10,350	5,202	1,000	5,356	1,000

表 6：WN18RR 中的层次关系

关系	$Khs_{\mathcal{G}_r}$	$\xi_{\mathcal{G}_r}$	层次
成员整体关系	1.00	-2.90	✓
上位词	1.00	-2.46	✓
具有部分	1.00	-0.82	✓
实例上位词	1.00	-0.78	✓
属于领域区域	1.00	-0.78	✓
属于领域用法	1.00	-0.78	✓
同义词集所属领域主题	0.99	-0.69	✓
另见	0.36	-0.29	×
派生相关形式	0.07	-3.84	×
类似于	0.07	-1.00	×
动词组	0.07	-0.50	×

E 并集建模：DNF 和 DM

处理并集运算有两种方式。利用德摩根律（DM），可以自然地将并集运算建模为 $S_1 \cup S_2$ 与 $S_1 \cap S_2$。（Ren 等，2020年）将查询转换为析取范式（DNF），并仅在最后一步计算并集运算。因此，CQA 模型通常训练交集和补集逻辑运算。虽然 WFRE 支持封闭形式的并集运算，但采用 DNF 的 WFRE 性能更好，因为训练查询中不包含并集运算。

F 补充结果

此外，我们还与两种 QE 方法 FuzzQE（Chen 等人，2022年）和 Gam- maE（Yang 等人，2022年）进行了比较。Yang 等人（2022年）采用自注意力机制开发了一种新的并集运算方法，性能优于 DNF 和 DM。论文中缺少 FuzzQE 在 FB15k 上的结果，也没有给出其在 FB15k-237 上推荐的超参数设置。由于我们无法复现 FuzzQE 在 NELL 上的结果，因此列出了论文中的结果以及我们复现的结果。在表7中，除论文中报告的 FuzzQE 结果外，WFRE 的表现均优于这两种模型。

表8还给出了我们模型输出的均值和标准差。所有分数均根据使用不同随机种子的四次运行计算得出。可以看出，标准差比均值小四个数量级。这表明 WFRE 非常稳定，并且显著优于以往的基线方法。

表 7：在 FB15k、FB15k-237 和 NELL 上的额外基准比较（平均倒数排名）。

数据集	QE	1P	2P	3P	2I	3I	PI	IP	2U	UP	2IN	3IN	INP	PIN	PNI	A_{P}	A_{N}
FB15k	GammaE	76.5	36.9	31.4	65.4	75.1	53.9	39.7	57.1	34.5	20.1	20.5	13.5	11.8	17.1	52.3	16.6
	WFRE	81.1	37.7	30.5	68.5	78.0	56.3	41.8	48.0	33.1	26.1	26.5	15.6	13.7	19.4	52.8	20.2
FB15k-237	GammaE	43.2	13.2	11.0	33.5	47.9	27.2	15.9	15.4	11.3	6.7	9.4	8.6	4.8	4.4	24.3	6.8
	WFRE	44.1	13.4	11.1	35.1	50.1	27.4	17.2	13.9	10.9	6.9	11.2	8.5	5.0	4.3	24.8	7.2
NELL	GammaE	55.1	17.3	14.2	41.9	51.1	26.9	18.3	16.5	12.5	6.3	8.7	11.4	4.0	4.5	28.2	7.0
	FuzzQE（我们的方法）	55.5	16.8	14.4	37.3	46.9	24.0	19.1	15.0	11.7	7.3	9.1	11.1	4.1	4.9	26.7	7.3
	FuzzQE（报告值）	58.1	19.3	15.7	39.8	50.3	28.1	21.8	17.3	13.7	8.3	10.2	11.5	4.6	5.4	29.3	8.0
	WFRE	58.6	18.6	16.0	41.2	52.7	28.4	20.7	16.1	13.2	6.9	8.8	12.5	4.1	4.4	29.5	7.3

表 8：WFRE：指标的均值（$\times 10^{-2}$）和标准差（$\times 10^{-6}$，粗体）。

数据集	QE	1P	2P	3P	2I	3I	PI	IP	2U	UP	2IN	3IN	INP	PIN	PNI	A_{P}	A_{N}
FB15k	MRR	81.1	37.7	30.5	68.5	78.0	56.3	41.8	48.0	33.1	26.1	26.5	15.6	13.7	19.4	52.8	20.2
		0.073	**0.33**	**0.39**	**0.45**	**1.5**	**0.37**	**1.6**	**1.3**	**2.5**	**0.74**	**1.2**	**1.8**	**0.37**	**2.8**	**0.16**	**0.046**
	HITS@1	73.0	27.4	21.2	58.6	70.1	45.7	31.1	36.1	23.1	16.5	16.4	8.4	7.1	10.9	42.9	11.9
		0.096	**0.35**	**2.0**	**1.6**	**5.4**	**1.0**	**1.8**	**2.5**	**8.1**	**1.0**	**1.4**	**1.3**	**0.32**	**4.8**	**0.67**	**0.11**
	HITS@3	87.8	42.0	33.8	74.5	83.5	62.0	46.6	54.5	36.6	28.5	28.9	16.4	13.8	20.8	58.0	21.7
		0.18	**1.3**	**0.35**	**2.0**	**0.69**	**1.9**	**9.8**	**3.5**	**2.6**	**0.18**	**6.8**	**3.0**	**0.16**	**1.7**	**0.11**	**0.32**
	HITS@10	93.5	58.0	48.5	86.5	92.2	76.3	62.4	70.4	52.7	45.5	47.3	29.9	26.6	36.2	71.2	37.1
		0.38	**6.5**	**1.2**	**0.60**	**0.95**	**0.062**	**1.1**	**0.60**	**2.9**	**0.97**	**0.20**	**0.91**	**4.3**	**0.16**	**0.0086**	**0.29**
FB15k-237	MRR	44.1	13.4	11.1	35.1	50.1	27.4	17.2	13.9	10.9	6.9	11.2	8.5	5.0	4.3	24.8	7.2
		0.032	**2.1**	**2.4**	**1.0**	**3.1**	**1.6**	**0.36**	**0.89**	**0.12**	**0.60**	**1.1**	**0.20**	**0.42**	**0.56**	**0.016**	**0.17**
	HITS@1	33.7	7.3	5.5	23.9	39.6	18.5	10.8	7.5	5.2	2.8	5.2	3.7	1.6	1.4	16.9	2.9
		0.056	**5.7**	**5.7**	**1.6**	**5.5**	**2.1**	**0.70**	**2.3**	**0.37**	**0.071**	**1.2**	**0.76**	**2.0**	**0.56**	**0.15**	**0.044**
	HITS@3	48.9	13.8	11.3	39.7	55.2	29.9	18.0	14.0	11.0	6.3	10.8	8.2	4.4	3.6	27.0	6.7
		1.2	**0.36**	**0.92**	**4.1**	**4.8**	**0.69**	**1.8**	**0.11**	**0.49**	**3.7**	**0.42**	**5.1**	**0.20**	**1.1**	**0.15**	**0.73**
	HITS@10	64.5	25.6	21.9	57.8	71.0	45.6	29.8	23.1	17.5	14.4	23.1	17.5	10.9	9.0	40.5	15.0
		0.15	**3.9**	**1.0**	**13**	**1.6**	**18**	**0.53**	**0.69**	**0.48**	**2.5**	**4.6**	**0.95**	**0.37**	**0.35**	**0.17**	**0.24**
NELL	MRR	58.6	18.8	16.0	41.2	52.7	28.4	20.7	16.1	13.2	6.9	8.8	12.5	4.1	4.4	29.5	7.3
		0.74	**0.44**	**0.056**	**0.72**	**6.0**	**2.6**	**0.96**	**0.22**	**0.62**	**0.0026**	**0.24**	**1.6**	**0.046**	**0.047**	**0.52**	**0.095**
	HITS@1	49.1	12.6	10.6	29.5	41.4	20.6	14.1	9.4	7.8	2.3	3.2	6.1	1.1	1.4	21.7	2.8
		1.3	**0.27**	**0.40**	**12**	**0.77**	**5.3**	**0.34**	**1.1**	**0.29**	**0.12**	**0.16**	**1.9**	**0.24**	**0.19**	**0.44**	**0.25**
	HITS@3	64.2	19.9	16.8	46.5	58.3	30.8	22.2	17.2	13.8	6.0	7.6	12.9	3.2	3.7	32.2	6.7
		1.3	**8.9**	**1.6**	**2.4**	**2.9**	**3.5**	**4.1**	**5.9**	**2.1**	**0.25**	**0.70**	**3.2**	**0.10**	**0.73**	**1.71**	**0.18**
	HITS@10	76.1	31.0	26.3	64.6	75.0	43.9	33.7	29.4	24.0	15.6	19.7	24.8	8.6	9.1	44.5	15.5
		1.1	**6.2**	**6.5**	**5.1**	**1.6**	**1.3**	**6.3**	**3.0**	**0.27**	**0.49**	**0.062**	**1.9**	**1.2**	**1.5**	**0.33**	**0.24**