文章编号：1003-1421(2026)11-0047-08
引用格式：杨扬，朱韦桥. 基于多元嵌入组合表示的轨道交通四网融合数据实体对齐模型[J]. 铁道运输与经济，2026，48(11)：47-54.
YANG Yang，ZHU Weiqiao. Entity Alignment Model for Rail-Transit "Four-Network Integration" Data Based on Multi-View Embedding Combination[J]. Railway Transport and Economy，2026，48(11)：47-54.

基于多元嵌入组合表示的轨道交通四网融合数据实体对齐模型

杨　扬，朱韦桥

(中国铁道科学研究院集团有限公司 电子计算技术研究所，北京　100081)

摘　要：针对轨道交通"四网融合"背景下综合交通枢纽数据实体的对齐难题，提出一种基于多元嵌入组合表示的实体对齐模型EAMEC。模型通过谓词对齐及知识图谱合并，统一关系谓词空间；在此基础上，集成语义摘要、TransER-GCN关系结构及句法增强属性嵌入3个模块，深度表征交通枢纽中数据实体的中心语义、路网拓扑及专业术语特征。此外，引入动态加权的多视图融合机制，根据数据特征自适应平衡语义、结构和属性信息的贡献，有效缓解了传统实体对齐模型对单一特征的过度依赖以及在属性稀疏场景下的对齐不稳定问题。在交通数据集CH-DT、JA-DT和AM-DT上的实验结果表明，EAMEC模型在*Hits*@1和*Hits*@10等关键指标上的表现均优于主流模型。本研究为构建统一交通知识图谱，实现跨系统数据治理提供了鲁棒的解决方案。

关键词：交通数据；知识图谱；实体对齐；BERT；属性嵌入
中图分类号：U293.3；TP391.1　　**文献标识码**：A　　**DOI**：10.16668/j.cnki.issn.1003-1421.20250928002

Entity Alignment Model for Rail-Transit "Four-Network Integration" Data Based on Multi-View Embedding Combination

YANG Yang，ZHU Weiqiao

(Institute of Computing Technologies, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China)

Abstract: To address the challenge of aligning data entities in comprehensive transportation hubs under the "four-network integration" context of rail transit, this paper proposed an entity alignment model using multi-view embedding combination (EAMEC). The model first aligned predicates and merged knowledge graphs to unify the relation-predicate space. Building on this foundation, it integrated three modules—semantic abstraction, TransER-GCN relational structure, and syntax-enhanced attribute embeddings—to deeply represent the core semantics, network topology, and domain-specific terminology of transportation-hub data entities. Moreover, a dynamic-weight multi-view fusion mechanism was introduced, which adaptively balanced the contributions of semantic, structural, and attribute information according to data features. This effectively mitigated the over-reliance on single features and the alignment instability observed in traditional entity-alignment models under attributesparse conditions. Experiments on the transportation datasets CH-DT, JA-

收稿日期：2025-09-28
通信作者：朱韦桥(1982—)，男，安徽太湖人，中国铁道科学研究院集团有限公司电子计算技术研究所研究员。
基金项目：中国国家铁路集团有限公司科技研究开发计划课题(2025F005)

DT, and AM-DT demonstrate that EAMEC outperforms mainstream models on key metrics such as *Hits*@1 and *Hits*@10. This research provides a robust solution for constructing a unified transportation knowledge graph and achieving cross-system data governance.
Keywords: Traffic Data; Knowledge Graph; Entity Alignment; BERT; Attribute Embedding

0　引言

轨道交通“四网融合”需要对干线铁路、城际铁路、市域(郊)铁路和城市轨道交通4种制式的数据进行整合。由于不同制式交通系统在建设年代、管理主体及数据模型上的差异，导致同一交通枢纽内的设施、服务等实体在各系统中呈现碎片化状态。例如，国家铁路系统里“上海虹桥站”在地铁系统中对应“虹桥火车站”；检票闸机、出站口等实体在不同系统中命名规则不一(如“检票口1”和“1号闸机群”)。这种语义鸿沟限制了综合交通枢纽精准换乘导航的实现。研究大型综合交通枢纽内异构设施的实体对齐[1]，对于实现四网融合具有实践价值与意义。

实体对齐是指找出指向相同现实世界对象的不同异构数据实体之间的关系，为构建统一的交通知识图谱[2-3]、实现数据互联互通提供技术支撑。现有实体对齐方法大多依赖于知识图谱的结构和属性信息。MTransE[4]、IPTransE[5]等翻译模型仅利用关系三元组，在属性信息丰富的场景中表现不佳；AttributeE[6]等模型过度依赖属性的字面匹配，难以应对跨语言场景。虽然MultiKE[7]等多视图融合模型整合了多种特征，但静态权重策略无法应用于交通领域数据属性稀疏[8]、结构复杂多变的动态场景，且未针对交通领域的长尾实体(如专业设备名称)设计表示方法，领域适配性不足。

基于此，以通用领域(如跨语言检索、医疗知识分析)多视图嵌入与知识图谱对齐为理论基础，提出一种基于多元嵌入组合表示的原创数据实体对齐模型(Entity Alignment Model based on Multivariate Embedding Combination，EAMEC)。相较于现有方法，EAMEC首次将句法增强与动态加权融合应用于轨道交通领域，用以解决专业术语歧义及属性稀疏难题。通过语义摘要、基于翻译的带关系投影的实体关系模型(Translation-based Entity-Relation Model with Relational projection，TransER)与图卷积神经网络(Graph Convolutional Network，GCN)联合框架及句法增强属性嵌入并行学习，显著提升了模型对轨道交通领域专业知识的理解深度与泛化能力。

1　综合交通枢纽实体对齐需求与数据特征

在四网融合背景下，综合交通枢纽的实体对齐需满足高精度安全关联、名称匹配及运营逻辑对齐的要求，并支持异构语义下的精准映射，以支撑枢纽内联程联运导航与应急疏散决策。综合交通枢纽的数据特征呈现出显著的复杂性。

(1) 命名规则及术语差异。不同制式交通系统对综合交通枢纽实体有不同的命名方式。铁路系统中多以“城市名+方位词+站”的形式命名，城市轨道交通站点多采用地理标识或地标命名。如铁路系统的“上海虹桥站”，在地铁系统中则命名为“虹桥火车站”。

(2) 空间属性稀疏。各系统开放程度不一，部分枢纽设施可能仅具备基础名称及编码，缺乏精确的坐标描述，对模型的推理能力提出了极高要求。

(3) 层级结构不一致。不同制式交通网络具有不同的拓扑层级，如国铁网络层级(枢纽-站场-站台)与城市轨道交通网络层级(站点-出口-闸机)无法一一对应匹配。这种异构性要求实体对齐模型具备理解不同粒度的业务分类体系的能力。

(4) 长尾实体的特殊性。轨道交通领域包含一些出现频率较低的实体，如“CTCS-3级列控车载设备”及特定的“站内继电器型号”等。这类实体具有全局分布稀疏但内部语义高密度性的特征。由于在全局语料库中出现频次低，基于统计学习的向量表示难以捕捉其特征；但其名称内部往往蕴含着复杂的层级结构、技术制式及核心功能等信息。这对实体对齐模型的深度语义解析能力提出了挑战。

2　轨道交通四网融合数据实体对齐模型EAMEC

EAMEC模型通过融合多元嵌入组合及动态权重机制实现数据实体对齐。EAMEC模型架构如图1所示，模型包含3部分：谓词对齐与知识图谱合

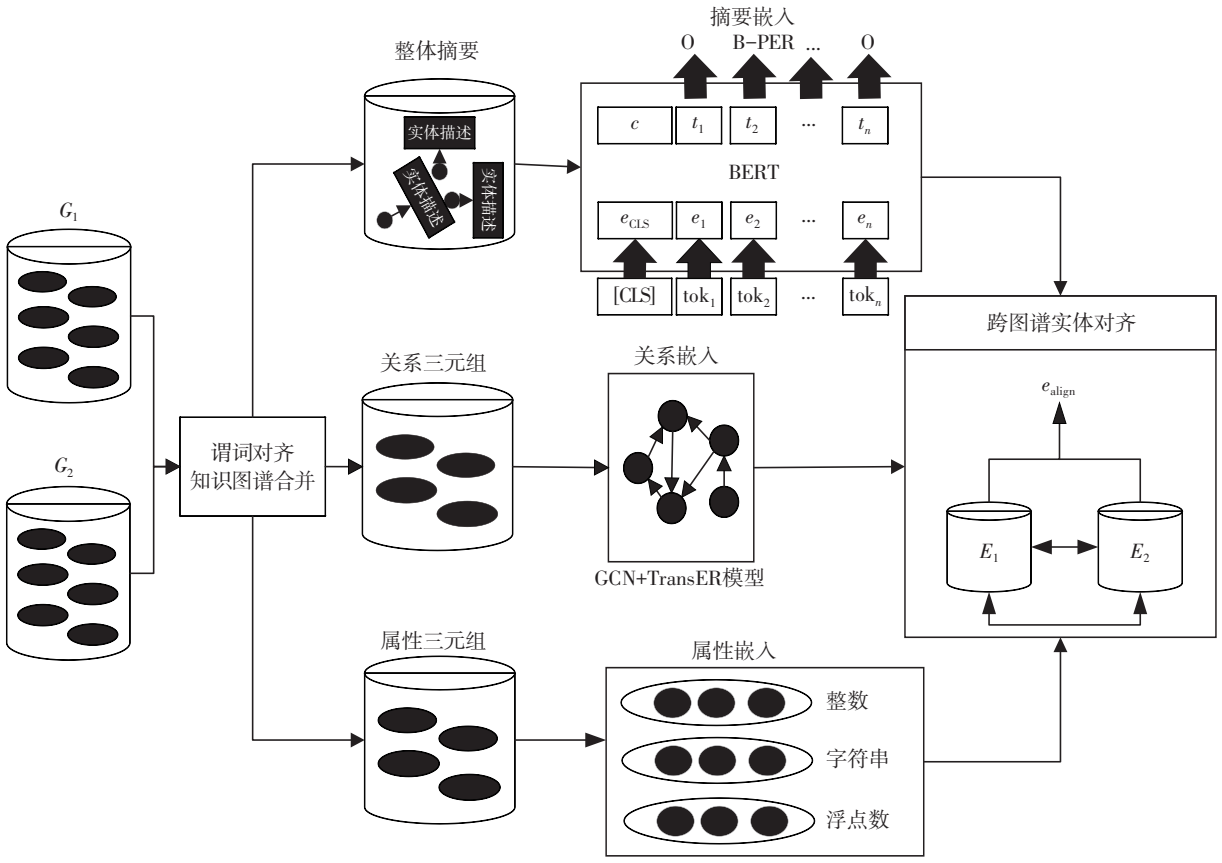

图1　EAMEC模型架构

Fig.1　EAMEC model architecture

并、多视图嵌入学习和自适应跨图谱实体对齐。以2个异构的知识图谱作为输入，首先经过“谓词对齐与知识图谱合并”模块，生成具有统一关系谓词空间的知识图谱。随后，该图谱并行输入到3个嵌入学习模块：“语义摘要嵌入”模块利用BERT处理实体文本描述；“关系结构嵌入”模块通过TransER-GCN联合框架学习；“属性语义嵌入”模块进行句法增强的编码。最后，3个模块输出的嵌入表示经“动态加权融合”机制进行组合，执行“跨图谱对齐匹配”，输出最终的对齐结果。

2.1　谓词对齐与知识图谱合并

进行实体对齐之前，首先需统一源图谱 $G_1=(E_1, R_1, T_1)$ 与目标图谱 $G_2=(E_2, R_2, T_2)$ 的谓词空间，以避免语义偏移。通过构建统一的谓词映射规则[9]，为后续并行嵌入模块的学习奠定一致性的结构基础。

2.1.1　语义与领域知识双重映射

为兼顾通用语义的泛化性和领域术语的专业性，通过对名称嵌入相似度与交通词典匹配度加权求和，计算谓词间的语义相似度。首先，通过预训练的FastText模型捕捉谓词间在自然语言层面的关联(如“表”与“数据项”可能具有相同含义)；其次，结合交通领域词典，对具有特定含义的同义词进行强制对齐。这种双重映射能够避免基于通用语料库建立的词向量在处理专业术语时的误判。

2.1.2　层次化对齐策略

针对轨道交通数据中相对复杂的谓词层级(如“一级业务”与“业务活动”)，采用父类定位、子类细化的分层对齐策略。首先，基于预定义的领域核心词列表提取谓词中的父类标识符(如从“一级业务”中提取“业务”)，保证全局语义方向上一致(如优先保证两者均为“业务”)；然后，在细粒度修饰词层面计算残差相似度。层次化对齐策略模拟了专家处理跨制式数据时的逻辑过程，优先保证宏观层级对齐，减少因业务命名层级不同导致的错误映射，提高复杂关系对齐的准确率。

2.1.3　全局标识符生成与图谱合并

完成上述过程后，通过设定的动态阈值筛选出

高置信度的谓词对，为每个谓词对分配全局唯一标识符，确保跨系统谓词在统一空间中的无歧义映射。然后，将2个异构图谱的三元组合并到一个统一的知识图谱中。知识图谱合并降低了后续并行嵌入模块的搜索空间复杂度，同时通过结构重组实现了信息增广，从而在原始数据属性缺失的情况下，也能根据合并后的关系进行补全推理。

2.2　多元组合嵌入学习

在四网融合场景下，实体属性在各空间中分布不均衡，因此无法通过单一维度的特征提取实现鲁棒对齐。模型设计了3个并行嵌入学习模块，能够在独立表示各维度信息的同时，对部分缺失信息起到一定的弥补作用。当某个数据源出现属性严重缺失或拓扑结构极度稀疏的情况时，其他模块仍能提供有效的判别信息，保证模型在复杂场景下的鲁棒性。

2.2.1　基于语义摘要的实体嵌入

采用BERT预训练语言模型对实体文本描述进行语义提取，得到语义摘要向量，表征实体的核心语义信息。不同于简单的名称匹配，语义摘要能体现实体之间的内在联系(如识别出“检票闸机”与“自动检票系统”具有同一功能)。通过设计联合评分函数，将语义摘要嵌入与图谱结构化嵌入进行联合优化，显式关联文本语义和业务逻辑信息。同时，引入“摘要-结构”混合翻译约束，将文本语义空间投影到统一的图谱拓扑空间，以语义摘要作为跨图谱对齐的“全局锚点”，保证模型在面对未见过的“新实体”时，即使缺乏拓扑关联，也能依据文本语义进行“零样本”的初步定位。

2.2.2　基于TransER-GCN的关系结构嵌入

为兼顾轨道交通路网拓扑特征和复杂的业务层级关系，设计了“翻译模型+图卷积网络”的双重框架。提出的TransER模型在传统翻译实体-关系模型基础上，通过为头尾实体h和t设置类型感知的投影矩阵，能够精确刻画交通枢纽中“铁路局集团公司-站段-车站”等跨层级复杂业务关系，其投影逻辑如公式(1)所示。

$$\begin{cases} h' = h\boldsymbol{M}_{hr} = h(\boldsymbol{I} + h_p r_p^{\mathrm{T}}) \\ t' = t\boldsymbol{M}_{tr} = t(\boldsymbol{I} + t_p r_p^{\mathrm{T}}) \end{cases} \tag{1}$$

式中：h'与t'为投影后的头实体和尾实体向量；$\boldsymbol{M}_{hr}$与$\boldsymbol{M}_{tr}$为将头实体和尾实体映射到特定关系空间的动态投影矩阵；$\boldsymbol{I}$为单位矩阵；h_p，$t_p \in \mathbb{R}^d$为头尾实体的类型特征向量，决定实体的个性化形变权重；$r_p \in \mathbb{R}^d$为关系r的投影向量，决定当前关系在投影中限定的空间维度。

并行的双层GCN架构则通过聚合邻域节点特征，捕捉交通枢纽的空间分布模式与通用结构模式。以“北京南站”为例，GCN不仅聚合站点自身属性(如等级、客流)，还聚合相邻站点(如“北京站”“北京丰台站”)的特征，学习枢纽站之间相连的空间分布规律。这种“局部逻辑”与“全局拓扑”的并行组合，弥补了单一评分函数无法对对称关系建模的局限，如“换乘关联”是双向对等的，传统翻译模型侧重单向层级关系，难以刻画这种对称性。通过联合优化，模型能够跨越不同系统的业务分类体系差异，实现对异构层级结构的深度挖掘。

2.2.3　基于句法增强的多层次属性嵌入

针对交通领域长尾实体的特征，本研究提出一种基于句法增强的属性嵌入方法，不再将专业术语视为不可分割的整体，而是通过构建依存关系矩阵，识别术语中的核心词与修饰词，并利用图注意力网络(GAT)对词汇进行特征增强。

首先对属性值文本进行依存句法分析，得到依存关系矩阵$\boldsymbol{D} \in \{0, 1\}^{n \times n}$，其中$D_{ij} = 1$表示词$i$与$j$存在依存关系。对于每个词$i$，定义其依存邻域$N(i) = \{ j | D_{ij} = 1 \}$。随后，采用图注意力网络(GAT)聚合邻域信息，得到增强后的表示e_i'。

$$e_i' = \alpha_{\mathrm{pos}(i)} \cdot \boldsymbol{W}_1 e_i + \sum_{j \in N(i)} \gamma_{ij} \cdot \boldsymbol{W}_2 e_j \tag{2}$$

式中：$\alpha_{\mathrm{pos}(i)}$为根据其词性(如名词、动词)动态分配的权重，可通过词性嵌入层学习得到；$\boldsymbol{W}_1$，$\boldsymbol{W}_2 \in \mathbb{R}^{d \times d}$为可学习的权重矩阵；$e_i$和$e_j$分别为词$i$和$j$的初始向量；$N(i)$为词$i$的邻居节点集合；$\gamma_{ij}$为核心注意力权重，衡量在重构词$i$的特征时，邻居词$j$的语义贡献度。

核心注意力权重γ_{ij}计算如公式(3)所示。

$$\gamma_{ij} = \mathrm{softmax}_j(\mathrm{LeakyReLU}(a^{\mathrm{T}}[W_1 e_i \| W_2 e_j])) \tag{3}$$

式中：softmax为归一化函数，确保词i分配给所有邻居j的注意力权重总和为1；LeakyReLU为激活函数，负责引入非线性计算；$a \in \mathbb{R}^{2d}$为注意力机制的

可学习参数向量，用于将高维拼接特征映射为计算注意力大小的标量得分；$\|$为特征拼接操作符，用于评估中心词特征$W_1 e_i$与邻居词特征$W_2 e_j$之间的匹配度。

通过上述机制，模型能够自适应地增加核心功能词在嵌入表征中的权重。随后，将增强后的词向量序列输入BERT进一步编码。本模块解决了通用模型对轨道交通专业词汇理解不准确的问题，通过深度解析复杂名词短语的内部结构与功能信息，提升模型对低频、高频专业术语的语义理解准确度及对轨道交通领域的适应度。

2.3　自适应跨图谱实体对齐

多视图特征的简单叠加会导致噪声累积。EAMEC模型创新性地提出了一种基于一致性的动态加权融合机制，协同优化3个并行嵌入视图，实现对异构交通数据的高效对齐。

动态权重通过计算各并行视图嵌入向量的“共识度”来实时调节。首先，将实体e在M种视图(语义摘要、关系结构、属性语义)中的嵌入表示进行线性加权，生成组合嵌入向量$e_{\text{cmb}}^{(e)}$。

$$\boldsymbol{e}_{\text{cmb}}^{(e)} = \sum_{i=1}^{M} \omega_i \cdot \boldsymbol{e}^{(i)} \tag{4}$$

式中：$M=3$为视图总数；$\omega_i \in [0,\ 1]$为第i种视图的动态融合权重且满足$\sum_{i=1}^{M} \omega_i = 1$；$\boldsymbol{e}^{(i)} \in \{e_{\text{summary}},\ e_{\text{relational}},\ e_{\text{attribute}}\}$为第$i$种视图的嵌入向量。

其次，为评估各视图嵌入的一致性并自适应分配权重，计算多视图嵌入的均值向量$\bar{e}$作为一致性评估的参考中心量。

$$\bar{e} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{e}^{(i)} \tag{5}$$

随后，通过计算各视图嵌入与中心向量$\bar{e}$的余弦相似度，通过softmax函数进行归一化，生成动态权重ω_i。

$$\omega_i = \frac{\exp\left(\dfrac{\text{sim}(\boldsymbol{e}^{(i)},\ \bar{e})}{\tau}\right)}{\sum_{j=1}^{M}\exp\left(\dfrac{\text{sim}(\boldsymbol{e}^{(j)},\ \bar{e})}{\tau}\right)} \tag{6}$$

式中：$\text{sim}(a,\ b) = \dfrac{a \cdot b}{\|a\|_2 \|b\|_2 + \varepsilon}$为余弦相似度计算函数；$\varepsilon = 1\text{e}^{-8}$用于防止除零错误；$\tau > 0$为温度系数，用于控制权重分布尖锐度。

该步骤确保了语义一致性更高的视图获得更大权重，从而有效抑制属性稀疏场景下的噪声。

在获得实体的组合嵌入表征后，模型遵循最大相似度原则执行匹配逻辑。通过计算源图谱G_1中的实体e_1与目标图谱G_2中所有候选实体$\{e_2^{(k)}\}$之间的余弦相似度，选择相似度最高的目标实体作为对齐结果e_{align}。这种基于全局一致性优化的匹配方案，确保了跨系统、跨制式数据在逻辑层面上的深度互联，显著增强了模型在极端环境下的鲁棒性与匹配精度。

$$e_{\text{align}} = \arg\max_{e_2^{(k)} \in KG_2} \frac{\boldsymbol{e}_{\text{cmb}}^{(e_1)} \cdot \boldsymbol{e}_{\text{cmb}}^{(e_2^{(k)})}}{\left\|\boldsymbol{e}_{\text{cmb}}^{(e_1)}\right\|_2 \left\|\boldsymbol{e}_{\text{cmb}}^{(e_2^{(k)})}\right\|_2} \tag{7}$$

3　实验结果与分析

3.1　数据集与实验环境

为了充分模拟轨道交通“四网融合”背景下大型综合交通枢纽内异构设施实体的对齐任务，对中国、日本、美国3个国家交通场景特征进行分析，分别对应构建了CH-DT，JA-DT和AM-DT 3个数据集，数据集统计特征如表1所示。

表1　数据集统计特征

Tab.1　Statistics of datasets

数据集	实体数	关系三元组数量	属性三元组数量	对齐位词数量
CH-DT	100 000	448 774	389 815	63
JA-DT	100 000	428 952	751 646	42
AM-DT	100 000	502 563	418 376	51

数据集包含了不同技术标准、管理主体及语言背景的交通枢纽信息，采用公开交通数据平台与专家标注相结合的方式构建，按照6∶2∶2比例划分训练集、验证集和测试集，种子对齐比例为30%，目的在于模拟交通枢纽对齐所面临的3类问题。①管理层级冲突：CH-DT数据集偏重于模拟中国大型交通枢纽中“国家铁路-地铁”双重管理下的数据异构。②冗余信息辨析：针对日本新宿、东京等极高密度的商业化综合枢纽，JA-DT设计了很高的属性密集度(属性三元组约为关系三元组的1.75倍)，在该场景下，枢纽内部包含大量设施及重叠的拓扑关系，模型需通过多视图嵌入深度挖掘实

体的微小差异，实现“同名异构”设施的精准对齐。③数据缺失与标准差异：AM-DT模拟了美国交通环境中不同制式间由于标准差异导致的数据缺失，该数据集的属性/关系比为0.83，存在严重的字段缺失，且涉及“英里”与“公里”等度量单位的不统一，该数据集能够验证模型在缺乏详细描述信息时，利用“语义摘要”与“拓扑结构”进行互补推理的鲁棒性。

3.2　实验设置与评价指标

模型基于PyTorch 2.1深度学习框架构建。硬件使用AMD EPYC 7742处理器(64核心/128线程)及双路NVIDIA GeForce RTX 3090 GPU，可提供48 GB高速显存池。软件栈采用CUDA 12.1驱动加速及分布式数据并行(DDP)架构。参数统一设置为：嵌入维度 $d=100$，学习率 $\alpha=1$，损失函数边际 $\gamma=5$，训练轮次为200，批处理大小为1 024。

试验分为3个部分。①CH-DT和JA-DT上的多视图嵌入组合分析实验。②CH-DT和JA-DT上的主流模型对比实验。③AM-DT上的属性稀疏场景验证。采用*Hits*@*k*(正确对齐实体位于前*k*位的比例)与*MR*(平均秩)作为核心指标，计算公式如下。

$$Hits@k=\frac{1}{|Q|}\sum_{e\in Q}\mathbb{I}(rank(e)\leqslant k) \tag{8}$$

$$MR=\frac{1}{|Q|}\sum_{e\in Q}rank(e) \tag{9}$$

式中：Q为查询实体集；$|Q|$为查询实体的总数；$\mathbb{I}(\cdot)$为示性函数；$rank(e)$表示实体e的真实匹配对象在模型预测候选列表中的具体排名；k为截断排名阈值，本实验中取1和10。

3.3　多视图嵌入组合实验分析

为验证多视图信息的互补性，在CH-DT与JA-DT上对比了6种嵌入组合，不同嵌入组合的实体对齐结果如表2所示。

当运用摘要、属性和关系嵌入时，EAMEC模型在2个数据集上均发挥了最优性能(CH-DT的*Hits*@1=92.04%；JA-DT的*Hits*@1=90.86%)，相较于单视图，多视图融合策略使性能得到极大提升。原因在于，一方面，多视图融合弥补了单一视图的局限性，如JA-DT在仅使用关系嵌入时效果不佳(*Hits*@1=33.49%)，表明在部分交通网络场景中，仅依赖结构化关系难以区分业务功能相似的实体，EAMEC通过多视图融合，模拟了人类专家综合多方面信息进行判断的思路；另一方面，动态权重机制平衡了不同视图的贡献，在JA-DT中，尽管单独使用属性嵌入时，*Hits*@1(62.52%)远高于关系嵌入(33.49%)，但在“属性+关系”组合中，模型通过动态权重抑制了冗余属性的干扰，并利用关系嵌入捕捉了主-外键关联上的优势，帮助模型分辨出更多的同名异构实体。

3.4　基线模型对比实验

为全面评估模型性能，选取MTransE，IPTransE，JAPE[10]，BootEA[11]，KDCoE[12]，AttributeE和MultiKE共7种主流模型作为基线进行对比实验。实体对齐模型的性能比较如表3所示，可以看出，EAMEC在CH-DT和JA-DT数据集上均取得了最佳性能(*Hits*@1分别为92.04%和90.86%)。

基于翻译的模型(MTransE，IPTransE)仅利用关系三元组构建语义信息，在属性密集的JA-DT上性能不佳(*Hits*@1<30%)。JAPE和BootEA等对种子对齐敏感的方法在使用30%种子对时，性能下降显著(超过14%)。跨语言模型KDCoE由于设计目的不同，*Hits*@1仅为42.71%。AttributeE在属性稀疏的

表2　不同嵌入组合的实体对齐结果

Tab.2　Entity alignment results with different embedding combinations

嵌入组合	CH-DT			JA-DT		
	MR	*Hits*@1/%	*Hits*@10/%	*MR*	*Hits*@1/%	*Hits*@10/%
摘要嵌入	927	70.40	80.34	1 108	64.89	78.49
关系嵌入	152	54.24	65.08	58	33.49	59.89
属性嵌入	6 774	61.13	72.08	1 001	62.52	73.89
属性+关系嵌入	145	67.83	77.98	138	56.34	69.63
摘要+关系嵌入	207	74.40	85.34	110	74.39	83.18
摘要+属性+关系嵌入	43	92.04	97.83	29	90.86	96.64

表3　实体对齐模型的性能比较

Tab.3　Performance comparison of entity alignment models

模型	CH-DT			JA-DT		
	MR	*Hits*@1/%	*Hits*@10/%	*MR*	*Hits*@1/%	*Hits*@10/%
MTransE	656	28.12	51.95	512	25.15	49.29
IPTransE	265	34.85	63.84	158	29.74	55.76
JAPE	266	31.84	58.88	189	23.57	48.41
BootEA	109	74.79	89.84	34	76.10	89.44
KDCoE	182	57.19	69.53	137	42.71	48.30
AttributeE	142	68.77	80.78	108	57.05	70.64
MultiKE	114	91.45	95.19	35	88.03	96.63
EAMEC	43	92.04	97.83	29	90.86	96.64

CH-DT(属性/关系比≈0.87)上表现一般(*Hits*@1=68.77%)，MultiKE可能存在跨语言命名差异引起的噪声，*Hits*@1比EAMEC低1.59%。相比之下，EAMEC的动态融合机制展示出更强的鲁棒性。

3.5　属性稀疏场景有效性验证

为验证EAMEC在属性稀疏场景下的表现，在AM-DT数据集(属性/关系比为0.83)上进行了测试，基线模型包括TransE，MTransE，JAPE和AttributeE。使用AM-DT数据集的结果分析如表4所示，可以看出，EAMEC的*Hits*@1达到93.35%，明显优于AttributeE和其他基线模型(均低于35%)。当出现术语差异(如“数据项名称”与“字段名”)时，摘要嵌入通过语义泛化达到了83.6%的跨术语匹配。当面临跨语言差异(例如“公里”与“mile”混用)时，动态权重机制有效地抑制了稀疏场景下的噪声干扰，将无效属性的平均贡献度抑制在18.3%左右，保障了对齐结果的准确性。

表4　使用AM-DT数据集的结果分析

Tab.4　Results analysis using AM-DT dataset

模型	*MR*	*Hits*@1/%	*Hits*@10/%
TransE	24 809	1.22	3.54
MTransE	7 105	33.46	34.32
JAPE	5 296	33.35	33.37
AttributeE	26	91.02	92.17
EAMEC	24	93.35	97.89

4　结论

针对轨道交通“四网融合”背景下综合交通枢纽的实体对齐难题，通过集成多视图特征并引入动态加权机制，构建了多元嵌入组合实体对齐模型EAMEC。主要研究结论如下。

（1）多元特征协同表示在复杂交通实体对齐中具有重要作用。单一维度的特征难以完整刻画交通枢纽中异构实体的复杂性。通过集成语义摘要、TransER-GCN结构及句法增强属性3个模块，能够从全局语境、网络拓扑及专业术语3个维度实现实体特征的深度表征，为跨系统异构数据对齐提供了基础。

（2）动态加权机制对属性稀疏与噪声干扰具有抑制作用。多视图动态加权机制能够根据数据局部特征自适应评估并分配各视图的贡献权重，克服了传统模型在属性缺失或稀疏环境下的性能衰减问题，显著提升了实体对齐的鲁棒性。

参考文献：

[1] 张子豪. 基于表示学习的知识图谱实体对齐方法研究[D]. 济南：齐鲁工业大学，2024.

[2] 黄　勃，吴申奥，王文广，等. 图模互补：知识图谱与大模型融合综述[J]. 武汉大学学报(理学版)，2024，70(4)：397-412.
HUANG Bo，WU Shen'ao，WANG Wenguang，et al. KG-LLM-MCom：A Survey on Integration of Knowledge Graph and Large Language Model[J]. Journal of Wuhan University (Natural Science Edition)，2024，70(4)：397-412.

[3] 迟　棠，车　超. 融合迭代式关系图匹配和属性语义嵌入的实体对齐方法[J]. 计算机科学，2023，50(S2)：69-74.
CHI Tang，CHE Chao. Entity Alignment Method Combining Iterative Relationship Graph Matching and Attribute Semantic Embedding[J]. Computer Science，2023，50(S2)：69-74.

[4] ZHU Y，LIU H Z，WU Z H，et al. Relation-Aware

Neighborhood Matching Model for Entity Alignment[J]. Proceedings of the AAAI Conference on Artificial Intelligence，2021，35(5)：4749-4756.

[5] KOUAGOU N J，HEINDORF S，DEMIR C，et al. ROCES：Robust Class Expression Synthesis in Description Logics *via* Iterative Sampling[J]. Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence，2024：4335-4343.

[6] PENG Y H，ZHANG J，ZHOU C Q，et al. Knowledge Graph Entity Alignment Using Relation Structural Similarity[J]. Journal of Database Management，2022，33(1)：1-19.

[7] ZENG W X，ZHAO X，TAN Z，et al. Matching Knowledge Graphs in Entity Embedding Spaces：An Experimental Study[J]. IEEE Transactions on Knowledge and Data Engineering，2023，35(12)：12770-12784.

[8] ZENG K S，DONG Z H，HOU L，et al. Interactive Contrastive Learning for Self-Supervised Entity Alignment[C]// Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Atlanta，2022：2465-2475.

[9] TRISEDYA B D，QI J Z，ZHANG R. Entity Alignment between Knowledge Graphs Using Attribute Embeddings[C]// Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. ACM，2019：297-304.

[10] MUNNE R F，ICHISE R. Entity Alignment *via* Summary and Attribute Embeddings[J]. Logic Journal of the IGPL，2023，31(2)：314-324.

[11] MUNNE R F. Entity Alignment and Attribute Enhancement between Knowledge Graphs[D]. Kanagawa：The Graduate University for Advanced Studies，2023.

[12] CHEN L H，KAGEURA K. Multilingualization of Medical Terminology：Semantic and Structural Embedding Approaches[C]// Proceedings of the Twelfth Language Resources and Evaluation Conference. Marseille，2020：4157-4166.

(责任编辑　刘宁馨)