

面向知识融合的本草典籍知识图谱实体对齐研究

李贺^{1*} 邵文诗¹ 刘嘉宇¹ 张津源¹ 沈旺¹ 王桂敏²

(1. 吉林大学商学与管理学院, 吉林 长春 130012; 2. 吉林省图书馆, 吉林 长春 130022)

摘要: [目的/意义] 针对本草典籍知识图谱实体对齐任务中图谱异构、术语易混淆及高质量标注稀缺等挑战, 提出融合生成对抗网络与模糊语义辨识的实体对齐模型 GAFL-Align, 旨在实现多源知识自动化融合。[方法/过程] 该模型通过 BERT 与图注意力网络融合实体语义与拓扑结构, 利用生成对抗网络进行领域自适应以消除异构引发的特征分布差异, 采用模糊边界负采样策略强化对易混淆术语的细粒度辨识, 并结合迭代自训练机制利用高置信度结果扩充样本, 有效降低对人工标注的依赖。[结果/结论] 实验表明, 该模型在自建数据集上的核心指标均优于基线方法。在此基础上构建的多源融合图谱实现了典籍间知识的互补与增值, 为本草典籍知识自动化融合提供了有力的技术支持。

关键词: 知识融合; 实体对齐; 本草典籍; 知识图谱; 深度学习

DOI: 10.3969/j.issn.1008-0821.2026.03.003

[中图分类号] G254.0; TP391.1; R281.2 [文献标识码] A [文章编号] 1008-0821(2026)03-0030-14

Research on Entity Alignment in Chinese Materia Medica Knowledge Graphs for Knowledge Fusion

Li He^{1*} Shao Wenshi¹ Liu Jiayu¹ Zhang Jinyuan¹ Shen Wang¹ Wang Guimin²

(1. School of Business and Management, Jilin University, Changchun 130012, China;
2. Jilin Provincial Library, Changchun 130022, China)

Abstract: [Purpose/Significance] The digital transformation of Chinese Materia Medica (CMM) classics is critical for bridging ancient pharmacological wisdom with modern drug discovery. However, existing Knowledge Graphs (KGs) for CMM are often constructed in isolation, resulting in fragmented information silos that hinder global data interoperability. While Entity Alignment (EA) has become a focal point in the international Semantic Web community, specific research targeting the alignment of ancient CMM literature remains a significant gap. Moreover, current state-of-the-art models—primarily designed for modern, high-resource languages—struggle to address the unique challenges of ancient Chinese texts. These challenges include severe structural heterogeneity caused by disparate historical writing styles, high terminological ambiguity where distinct medical concepts share similar characters, and a critical scarcity of high-quality annotated datasets. This study aims to fill this gap by proposing a domain-specific deep learning framework designed to automate the fusion of multi-source historical medical knowledge. [Method/Process] To overcome these barriers, this paper proposed the Generative Adversarial Fuzzy-boundary Learning (GAFL-Align) model. The study utilized two representative classics spanning different historical eras: Shennong Bencao Jing and Tangye Bencao. After data cleaning, the datasets comprised 3 771 and 3 910 normalized entities, respectively, focusing on core categories such as herbs, symptoms, and diseases.

收稿日期: 2025-12-11

基金项目: 国家社会科学基金冷门绝学研究专项学术团队研究项目“本草典籍整理、知识组织与智慧化建设研究”(项目编号: 23VJXT024)。

作者简介: 邵文诗 (1999-), 女, 硕士研究生, 研究方向: 中医古籍数字人文。刘嘉宇 (1995-), 男, 助理研究员, 博士, 研究方向: 数据挖掘。张津源 (2001-), 女, 博士研究生, 研究方向: 中医古籍数字人文。沈旺 (1983-), 女, 教授, 博士, 博士生导师, 研究方向: 用户信息行为。王桂敏 (1976-), 女, 研究馆员, 硕士, 研究方向: 科技信息服务。

通信作者: 李贺 (1964-), 女, 教授, 博士, 博士生导师, 研究方向: 信息行为分析、知识管理。

The technical architecture integrated BERT for deep semantic encoding with Graph Attention Networks (GAT) to capture topological structures. To handle distribution shifts across heterogeneous texts, the model employed a Generative Adversarial Network (GAN) for domain adaptation, mapping entities into a unified feature space. Furthermore, a novel fuzzy boundary negative sampling strategy was developed to distinguish “hard negatives”—terms with high lexical similarity but distinct medical meanings. To address data scarcity, an iterative self-training mechanism with confidence-aware filtering was implemented to augment the training set from a limited number of expert-annotated seed pairs. [Result/Conclusion] Experimental results indicated that GAFL-Align achieved a Hits@1 score of 83.59%, significantly outperforming nine baselines, including translation-based models, GNN variants, and Large Language Models (LLMs)-augmented approaches like ChatEA. The model successfully constructed a fused KG containing 6 826 entities, effectively merging heterogeneous data while preserving unique source-specific attributes. These findings demonstrate that combining adversarial domain adaptation with fine-grained semantic differentiation offers a superior solution for low-resource historical knowledge fusion compared to generic LLMs. Ultimately, this research provides a robust technical foundation for the intelligent organization of CMM heritage, offering significant implications for digital humanities and the global standardization of traditional medicine data.

Keywords: knowledge fusion; entity alignment; Chinese Materia Medica classics; knowledge graph; deep learning

中医本草典籍承载着数千年的传统医学智慧，其丰富的药物知识体系既为中医临床实践提供理论指导，也为现代药物研发与药理研究提供重要参考^[1]。为有效利用这些宝贵资源，将海量、异构的典籍文献转化为结构化的知识图谱已成为一项关键研究路径^[2]。然而，现有的本草典籍知识图谱往往局限于单一典籍或特定的研究视角来构建^[3]。这种各自独立的构建模式虽具有局部价值，却切断了典籍间的内在联系，导致中医知识体系碎片化，不仅阻碍跨典籍知识的融通，也使基于全局视角的智能推理与知识发现难以实现^[4]。因此，整合分散的知识资源，构建更为全面、一致的本草典籍知识图谱，已成为该领域的迫切需求^[5]。实体对齐(Entity Alignment)正是应对此挑战的关键技术，主要通过识别并链接不同图谱中指代同一现实世界对象的实体，为实现多源知识的有效融合与互补提供了核心方法^[6]。

尽管实体对齐技术在通用领域已取得显著进展，但将其直接应用于中医本草典籍时，仍面临诸多领域特有的挑战。首先，图谱结构异构问题限制了模型的鲁棒性。不同典籍在著述风格、术语体系乃至图谱拓扑结构上存在显著差异^[7]，而现有模型往往忽视这种跨图谱分布偏差^[8-9]，导致对齐精度受限。其次，术语易混淆是该领域的另一大难题^[10]。本草典籍中充斥着大量形近义远的困难样本，使得现有依赖随机负采样策略的模型容易产生误判^[11-13]，难以精准捕捉细微的语义差别。最后，高质量标注数据的稀缺成为制约技术应用

的主要瓶颈^[4]。由于本草领域的专业门槛导致标注成本较高，依赖大规模监督信号的传统深度学习方法在低资源环境下往往难以奏效。

针对上述挑战，本文提出一种融合生成对抗网络与模糊语义辨识的实体对齐模型(Generative Adversarial Fuzzy-boundary Learning, GAFL-Align)，旨在通过跨域特征自适应机制、模糊边界负采样策略(Fuzzy Negative Sampling)以及迭代自训练方法(Bootstrapping)，攻克本草典籍实体对齐中图谱异构、术语混淆及标注数据稀缺的难题。在此基础上，本研究构建了一个融合知识图谱，有效实现了不同典籍间异构知识的互补与增值，为本草典籍知识的系统性整合提供了可行的技术路径。

1 相关工作

1.1 本草典籍知识图谱相关研究

近年来，面向本草典籍的知识图谱研究主要围绕知识图谱的构建与应用两大方向展开^[14]。在构建方面，早期研究侧重于领域本体的构建^[15-16]以及依赖专家经验的手工知识抽取^[17-18]。随着对构建规模与效率要求的不断提升，研究重心逐渐转向自动化知识抽取技术。历经了从以条件随机场为代表的传统机器学习方法的探索^[19]，到以BERT-BiLSTM-CRF为代表的深度学习模型的引入^[20]，并且随着大语言模型(Large Language Models, LLMs)的兴起，以此为驱动的全自动化构建方式亦开始浮现^[21]。而在应用方面，基于已构建的知识图谱，研究者在智能问答^[22]、可视化分析^[23]和辅助决策^[24]等方面开展了诸多有益尝试，拓展了本草典

籍知识在实际场景中的使用价值。

然而,当前本草典籍知识图谱研究或局限于单一典籍^[19],知识覆盖严重不足;或在整合多部典籍时仅停留在表层数据的汇集^[21-23],同一实体常以不同名称在图谱中冗余且割裂地存在,破坏了知识图谱的语义一致性。知识融合特别是实体对齐在该领域的研究相对滞后,少数对齐尝试主要停留在人工对齐^[25]或基于表层规则与字符串匹配^[26]的阶段,难以应对本草典籍中复杂的语义现象,迫切需要引入能够理解深层语义的智能化自动实体对齐模型。

1.2 实体对齐相关研究

实体对齐旨在识别并链接不同知识图谱中的等价实体,是实现知识图谱融合的关键技术^[27]。早期的研究主要基于以TransE及其变体为代表的翻译模型^[28],如MTransE^[11]等。随着图神经网络的兴起,基于图结构的对齐方法成为主流^[29]。GCN-Align模型^[30]首次利用GCN通过聚合邻域信息来生成实体的结构化嵌入。随后,RREA模型^[8]引入关系感知机制,显著增强了实体在不同关系下的语义表征能力。为弥补单一结构信息的不足,语义增强成为另一重要趋势。例如,BERT-INT模型^[9]通过微调BERT捕获细粒度文本语义。近期,利用大语言模型强大的泛化与推理能力来提升对齐效果,已成为这一领域的重要探索方向。例如,LEA模型^[31]通过LLM增强实体的文本化语义描述,将对齐转化为检索任务。

在基础架构之外,针对实体对齐中特定难题的优化策略也相继被提出。首先,在嵌入特征优化方面,对抗学习机制被引入以试图缓解分布异构问题。其中,SEA模型^[32]主要致力于减少由实体度分布差异引发的嵌入偏差,而GAEA模型^[33]则通过模拟伪结构分布进行对抗训练,旨在改善无监督场景下的噪声干扰问题。其次,针对负采样质量不高的问题,研究者提出了从困难样本挖掘^[34]到自适应负样本混合^[35]等策略,旨在提升模型对困难样本的辨识能力。最后,为缓解数据稀疏问题,迭代自训练机制被广泛采用,BootEA模型^[36]率先确立了利用高置信度预测结果扩充训练集的自训练方法,而后续研究如STEA模型^[37]则改进了数据扩充策略,通过引入实体依赖关系来校验预测结果的结构兼容性,进而生成高质量的依赖

感知伪标签以持续丰富训练样本。

1.3 研究述评

针对本草典籍领域的深层语义对齐研究尚处于起步阶段,直接迁移现有模型仍面临诸多局限。具体而言,现有对抗模型多聚焦局部图结构的去噪,未能从典籍源异构的宏观视角,有效消除因著述风格与时代变迁引发的整体特征分布的差异,致使模型难以建立跨典籍实体间的准确映射关系。同时,针对形近义远术语的细粒度辨识机制依然缺失,通用的随机负采样策略往往难以挖掘处于语义模糊边界的困难样本,无法满足本草知识对齐的精确性要求。此外,本草典籍实体对齐受制于繁重的人工校验与复杂的规则构建,尚未引入迭代自训练等半监督机制以挖掘未标注数据的价值,限制了模型在低资源场景下的泛化能力。

2 本草典籍知识图谱实体对齐模型

本文设计并构建了一个融合生成对抗学习与模糊语义辨识的实体对齐模型GAFL-Align,如图1所示。该模型首先利用BERT(Bidirectional Encoder Representations From Transformers)预训练语言模型捕捉本草典籍术语的深层语义信息,生成富含先验知识的节点初始特征。随后将图注意力网络(Graph Attention Networks, GAT)作为结构编码器与GAN生成器,先深度聚合实体的语义内容与图谱的拓扑结构生成嵌入表示,继而通过与判别器的动态对抗博弈,消除由不同典籍著述风格带来的特征分布差异,将实体映射到统一的特征空间中。与此同时,针对本草典籍中大量形近义远的易混淆术语,本研究设计了模糊边界负采样策略,迫使模型关注语义空间中距离极近但非同义的困难样本,以提升对细粒度语义差异的辨识能力。最后,结合迭代自训练策略,利用模型的高置信度预测结果自动扩充训练样本,从而有效缓解对大规模人工标注数据的依赖。

2.1 问题定义

本文将本草典籍知识图谱定义为 $G=(E,R,T)$,其中, E 代表实体集, R 代表关系集, $T\subseteq E\times R\times E$ 表示三元组集合,三元组 $(h,r,t)\in T$ 表示头实体 h 通过关系 r 连接到尾实体 t 。给定两个来源不同且异构的本草知识图谱 $KG1=(E_1,R_1,T_1)$ 和 $KG2=(E_2,R_2,T_2)$,实体对齐的任务是发现两个图谱中指代同一现实对象的等价实体对集合 A 。该集合定义为式(1):

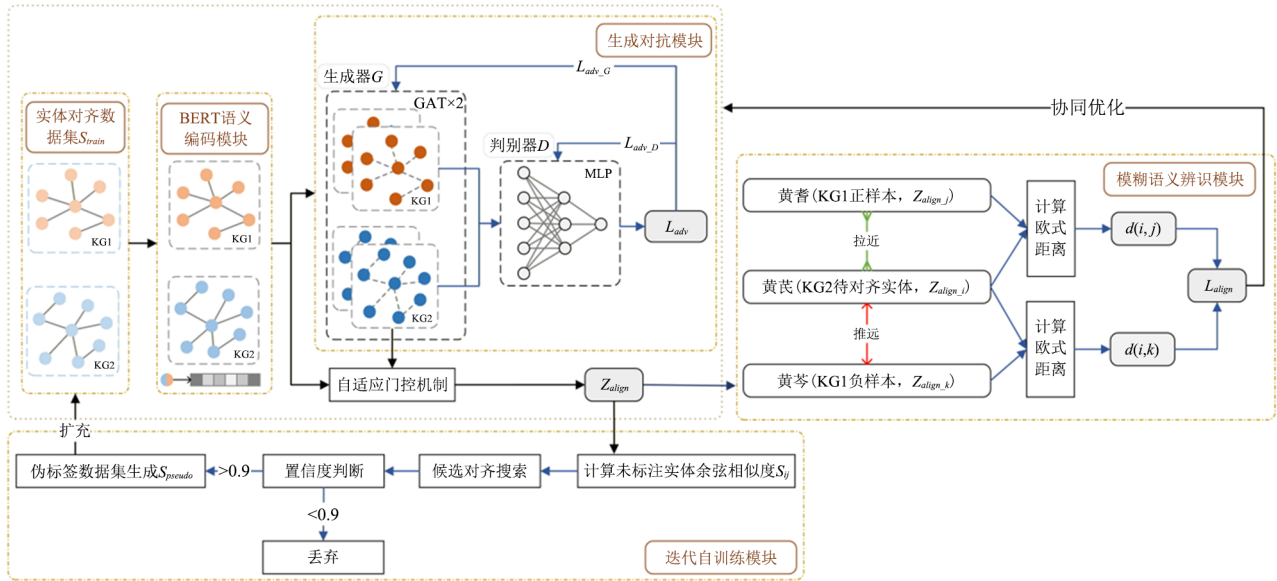


图1 融合生成对抗学习与模糊语义辨识的实体对齐模型

Fig. 1 Entity Alignment Model Integrating Generative Adversarial Learning and Fuzzy Semantic Identification

$$A = \{(e_1, e_2) | e_1 \in E_1, e_2 \in E_2, e_1 \equiv e_2\} \quad (1)$$

其中, \equiv 表示等价关系。在本研究的半监督设置下, 假设已有一组少量的预对齐实体对作为训练种子集 $S_{train} \subset A$, 模型的目标是利用 $KG1$ 、 $KG2$ 及 S_{train} 自动识别出剩余未知的对齐实体对。

2.2 基于BERT的本草典籍实体语义编码模块

为精确捕捉本草典籍术语名称中蕴含的深层语义信息, 本模型采用BERT预训练语言模型对所有图节点进行初始特征向量化。对于知识图谱中的任一节点 v , 其文本名称 $text_v$, 输入至预训练的 bert-base-chinese 模型, 并提取与分类任务相关的 [CLS] 标记所对应的输出向量, 作为该节点的初始特征向量 X_v , 该过程形式化表示为式 (2):

$$X_v = BERT(text_v) \quad (2)$$

其中, $X_v \in \mathbb{R}^d$, 在本研究中嵌入维度 d 设定为 768。通过此步骤, 模型将本草典籍实体映射为连续的高维稠密向量, 不仅解决了特定术语的语义表示问题, 更为后续图神经网络融合提供了高质量的语义特征。

2.3 生成对抗学习模块与自适应门控机制

2.3.1 聚合拓扑结构的图注意力生成器

仅依赖BERT提取的文本语义不足以完全表示实体全部信息, 因为在本草典籍知识图谱中, 结构上下文蕴含着关键的判别信息。考虑到不同邻居节点对实体对齐的贡献度存在差异, 本研究采用GAT作为结构编码器。GAT既起到深度融合语义与图谱结构的功能, 又作为生成对抗网络中的生成器。

作为编码器, 该模块旨在实现文本语义与拓扑结构的深度融合。其接收由BERT初始化的语义特征矩阵 X 与图谱邻接矩阵 A 作为输入。GAT利用自注意力机制, 根据节点间语义特征的关联强度动态分配结构聚合的权重。对于任意节点 i 及其邻居 j , 其注意力系数 α_{ij} 计算如式 (3) 所示:

$$\alpha_{ij} = \frac{\exp(LeakyReLU(a^T [Wh_i || Wh_j]))}{\sum_{k \in \mathcal{N}_i} \exp(LeakyReLU(a^T [Wh_i || Wh_k]))} \quad (3)$$

其中, W 为可学习的权重矩阵, a 为注意力向量, \parallel 表示向量拼接操作。

作为生成器, 模型利用上述注意力系数对邻域特征进行加权聚合, 直接构建用于对抗博弈的全局特征矩阵 Z 。对于矩阵中的任意行向量 $z_i \in Z$, 其计算公式定义为式 (4):

$$z_i = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} Wh_j \right), \forall z_i \in Z \quad (4)$$

2.3.2 消除分布差异的领域自适应判别器

尽管前述GAT模块能够有效融合局部结构信息, 但由于不同典籍成书年代悬殊, 其著述风格、用词习惯及潜在的图拓扑特征存在差异。这种异构性导致生成的实体嵌入 Z_{KG1} 和 Z_{KG2} 在特征空间中服从不同的概率分布, 若直接在此基础上计算相似度, 这种分布偏差将掩盖实体间真实的语义关联, 导致对齐精度受限。为此, 本节引入基于对抗性领域自适应 (Adversarial Domain Adaptation) 机制, 旨在迫使模型学习到一个与具体典籍来源无关的统一特征空间。为了消除 $KG1$ 与 $KG2$ 之间的整体分布差异, 本模型引入了一个二分类判别器 D 。该判

别器由多层感知机构成，其目标是尽可能精准地分辨输入的实体嵌入 Z_i 是来自源图谱还是目标图谱。训练过程遵循最小化—最大化博弈策略，判

$$\mathcal{L}_{adv_D} = -\mathbb{E}_{Z_i \sim \{Z_{KG1} \cup Z_{KG2}\}} [c_i \log D(Z_i) + (1 - c_i) \log(1 - D(Z_i))] \quad (5)$$

其中， c_i 为实体的真实来源标签。与之相对，生成器的目标则是生成具有混淆性的嵌入以误导判

$$\mathcal{L}_{adv_G} = -\mathbb{E}_{Z_i \sim \{Z_{KG1} \cup Z_{KG2}\}} [(1 - c_i) \log D(Z_i) + c_i \log(1 - D(Z_i))] \quad (6)$$

通过上述 \mathcal{L}_{adv_D} 与 \mathcal{L}_{adv_G} 的交替优化，模型最终达到纳什均衡状态。此时，生成的实体嵌入不仅保留了语义与结构特征，还消除了由典籍来源引起的分布差异，从而将异构的 $KG1$ 与 $KG2$ 映射到了一个统一的向量空间中。

2.3.3 语义与结构的自适应门控融合机制

经过前述对抗训练优化后，生成器输出的嵌入 Z 虽然消除了图谱间的分布差异，具备良好的跨域不变性，但这种为了迎合判别器而进行的对抗性调整，可能会在一定程度上导致原始实体特征丢失了部分对于精准对齐至关重要的细粒度语义信息。

为解决消除图谱异构与保留语义信息之间的潜在冲突，本模型引入了一个自适应门控融合机制，动态权衡原始高保真语义特征 X 与对抗适应后的结构特征 Z 。门控系数 g 通过式 (7) 学习：

$$g = \sigma(W_g [X \oplus Z] + b_g) \quad (7)$$

其中， \oplus 表示向量拼接操作， W_g 和 b_g 为可学习的权重与偏置， σ 为 Sigmoid 激活函数。随后，利用该门控系数对两种特征进行加权融合，生成最终用于对齐任务的实体嵌入 Z_{align} ，如式 (8) 所示：

$$Z_{align} = g \odot X + (1 - g) \odot Z \quad (8)$$

其中， \odot 表示元素级乘法。通过这一机制，模

$$S_{neg}^{(i)} = \{e_k \in E \mid sim(BERT(e_i), BERT(e_k)) > \tau \wedge e_k \neq e_i \wedge (e_i, e_k) \notin A\} \quad (9)$$

其中， $sim(\cdot)$ 表示余弦相似度， A 为已知的对齐集合。在此基础上，为了最大化梯度的判别效力，模型在训练过程中实施 Top-K 困难样本挖掘策略。不同于随机策略，该算法对 $S_{neg}^{(i)}$ 中的实体按相似度进行降序排列，并截取前 K 个实体作为困难负样本。这种机制确保了模型始终聚焦那些外观最相似但实质不同的干扰项，迫使模型在统一的特征空间中必须捕捉更细微的语义差异才能完成区

$$\mathcal{L}_{align} = \sum_{(i,j) \in S_{min}} \max(0, m + d(Z_{align_i}, Z_{align_j}) - d(Z_{align_i}, Z_{align_k})) \quad (10)$$

其中， (i, j) 为正样本对， k 为候选集 $S_{neg}^{(i)}$ 中选取的困难负样本， $d(\cdot)$ 为 L_2 欧氏距离度量， m 为预设

别器通过最小化二元交叉熵损失来进行优化，其损失函数 \mathcal{L}_{adv_D} 定义为式 (5)：

别器，其对抗损失 \mathcal{L}_{adv_G} 定义为式 (6)：

型能够智能地对特征进行选择，对于语义歧义较大但结构清晰的实体，模型会自动赋予 Z 更高的权重，而对于那些语义独特但结构噪声较大的实体，则更多地保留 X 中的原始信息，最终得到的 Z_{align} 兼具了语义丰富性与分布一致性。

2.4 面向易混淆术语的模糊语义辨识模块

2.4.1 模糊负样本筛选算法

尽管前述生成对抗网络有效解决了宏观图谱异构，但其忽略了微观细粒度语义差异，且传统随机负采样因引入大量简单负样本，难以区分特征空间中极近的混淆项。为此，本研究提出模糊边界负采样策略，旨在挖掘语义模糊地带的困难负样本，将优化重心转向局部精细化边界，迫使模型学习细粒度辨识能力。

具体而言，模型针对任意锚点实体 e_i 构建模糊负样本候选集。该过程利用 BERT 初始语义向量计算锚点与图谱中其他实体的余弦相似度，并引入语义阈值来划定模糊边界。在此筛选过程中，算法排除了真实的对齐目标，以防止将正样本误标记为负例而干扰模型优化。即只有相似度高于该阈值且并非正确对齐的实体，方可被纳入模糊负样本候选集 $S_{neg}^{(i)}$ ，其公式如式 (9) 所示：

分，从而缓解了复杂本草术语的细粒度辨识难题。

2.4.2 协同优化目标与梯度隔离策略

为了实现模型中不同组件的协同优化，本研究设计了基于分层梯度的交替迭代训练策略。基于筛选出的 Top-K 困难负样本，本研究构建了基于间隔的排序损失函数 (Margin-based Ranking Loss)，旨在拉近正样本对在特征空间中的距离，同时推远处于模糊边界的困难负样本。其形式化定义为式 (10)：

的对齐间隔参数，其核心作用是强制要求负样本对的距离至少要比正样本对的距离大 m ，从而确保

表2 模型参数
Tab. 2 Model Parameters

参数类别	参数	设定值	简述
模型结构	嵌入维度	768	节点由 Bert 初始化的特征维度
	GAT隐藏层维度	128	图注意力网络的隐藏层大小
	注意力头数	4	GAT中的多头注意力头数
损失函数权重	对齐损失间隔	1.0	FuzzyMarginLoss 中的间隔超参数
	模糊采样相似度阈值	0.8	筛选高相似度困难负样本的语义阈值
	对抗训练权重	0.2	整体损失中 $L_{adv,c}$ 的平衡权重
对齐与迭代参数	伪标签生成阈值	0.9	迭代过程中筛选高质量对齐样本的相似度门槛
	初始最大伪标签数	20	迭代初期每轮新增的伪标签数量上限
	伪标签衰减率	0.95	每轮迭代中最大伪标签数的动态衰减系数, 防止噪声干扰
学习率参数	BERT学习率	0.00005	BERT的学习速率
	GAT及门控学习率	0.001	GAT编码器与门控融合机制的学习速率
	判别器学习率	0.0005	对抗网络判别器的学习速率

3.2.2 评估指标

本研究采用实体对齐领域的通用指标 Hits@k ($k = 1, 3, 10$) 和平均倒数排名 (Mean Reciprocal Rank, MRR) 评估模型性能。其中, Hits@1 直观反映对齐的准确率, Hits@3 与 Hits@10 考察候选列表的覆盖能力, MRR 则衡量整体排序质量。

3.2.3 基线模型

为评估 GAFL-Align 的有效性, 本研究选取了 9 个涵盖实体对齐技术演进历程的基线模型, 旨在考察其在结构异构、语义复杂及低资源场景下的综合竞争力。其中, MTransE^[11] 作为平移距离模型的基准, GCN-align^[30]、RDGCN^[38] 与 MuGNN^[39] 代表了图神经网络的不同变体, 引入此类模型旨在验证深层聚合多视角拓扑结构信息对于克服典籍结构异构的重要性。在语义与半监督方面, 选取 BERT-INT^[9] 以评估脱离图结构约束下纯文本语义对齐的独立贡献, 选取 BootEA^[36] 作为迭代自训练的经典代表, 用于衡量在缺乏噪声控制机制时半监督策略的实际收益。此外, 本研究还引入了 LEA^[31]、AutoAlign^[40] 与 ChatEA^[41] 这三类大语言模型增强的前沿方法, 分别代表了基于语义检索、零样本全自动对齐以及基于思维链推理的方法。

3.3 实验结果与分析

3.3.1 对比实验

本研究将 GAFL-Align 模型与上述基线模型在

自建的本草典籍数据集上进行了对比实验, 实验结果如表 3 所示。

表3 GAFL-Align 与基线模型的对比实验结果
Tab. 3 Comparative Experimental Results of GAFL-Align and Baseline Models

模型	Hits@1/ %	Hits@3/ %	Hits@10/ %	MRR/ %
MTransE	74.10	81.54	87.18	77.84
GCN-align	75.38	82.31	88.21	79.03
RDGCN	78.97	85.38	90.51	82.48
MuGNN	80.50	85.13	90.00	83.71
BootEA	77.69	84.10	89.49	81.75
BERT-INT	78.21	84.62	89.74	82.12
LEA	82.05	85.13	91.03	84.82
AutoAlign	81.28	86.41	90.76	83.98
ChatEA	82.82	87.43	92.05	85.35
GAFL-Align (本模型)	83.59	88.97	92.56	86.23

从整体实验结果来看, GAFL-Align 模型在所有核心指标上均取得了最优性能, 显著超越了所有基线方法。通过横向对比基于结构的方法可以发现, 以 MTransE 为代表的翻译模型表现相对最弱, 这主要归因于简单的线性变换假设难以有效处

理本草典籍中复杂的非线性映射关系。相比之下，GCN-align、RDGCN和MuGNN等图神经网络变体的性能呈现出阶梯式上升的趋势。其中，MuGNN利用多通道机制融合了多视角信息，将Hits@1指标提升至80.50%。这一趋势证明了在本草典籍实体对齐任务中，深度挖掘和聚合邻域拓扑结构信息对于弥补单一典籍的描述缺失具有决定性作用。而BERT-INT模型虽然利用了强大的预训练语义特征，但其78.21%的Hits@1指标仍低于部分GNN模型。这说明脱离了图谱结构约束的纯语义匹配，极易受困于古籍中名异实同或名同实异的语言模糊性陷阱。同时，经典的迭代自训练模型BootEA表现一般，其Hits@1指标为77.69%，这表明在缺乏有效噪声控制机制的情况下，直接引入伪标签容易导致错误累积。这一对比反证了GAFL-Align不仅通过BERT与GAT实现了语义与结构的深层互补，还通过置信度感知的动态迭代策略，规避了传统半监督学习中固有的错误扩散问题。

在大语言模型相关方法的对比中，ChatEA虽凭借强大的语义推理能力，在基线模型中表现最佳，优于LEA和AutoAlign。但GAFL-Align依然保持了约0.77%的优势。这种专用小模型超越通用大模型的现象，主要归功于本模型针对领域特性的深度定制。LLM方法虽然通识语义能力极强，但往往缺乏对本草典籍垂直领域知识图谱异构结构的感知能力，容易忽略隐式的拓扑关联。而GAFL-Align通过生成对抗网络主动消除了跨典籍的特征分布差异，并利用模糊边界负采样策略强迫模型在细粒度上区分形近义远的困难样本。这种针对性的领域自适应设计，使其在处理本草典籍特有的异构与混淆挑战时，比通用大语言模型具有更高的精确度。

3.3.2 消融实验

为了验证GAFL-Align模型中各个核心组件的有效性，本文设计了消融实验，实验结果如表4所示。

表4 GAFL-Align的消融实验结果
Tab. 4 Ablation Study Results of GAFL-Align

模型	Hits@1/%	Hits@3/%	Hits@10/%	MRR/%
移除 Bert 语义模块	77.18	82.31	87.18	79.94
移除生成对抗网络模块	82.31	84.87	88.72	84.02
移除门控机制	81.79	83.07	88.46	83.51
移除模糊匹配负采样 FuzzyMarginLoss	83.08	87.69	90.77	85.46
移除迭代对齐训练策略	83.08	88.21	91.54	85.88
GAFL-Align(本模型)	83.59	88.97	92.56	86.23

从表4的实验数据可以观察到，GAFL-Align模型在四项评估指标上均优于移除了任一单一组件的变体模型。当移除BERT语义模块后，模型性能下降最为显著。这表明，由BERT提供的深层语义信息是模型理解本草典籍实体内涵的基石，缺乏高质量的语义输入，后续的结构和对抗学习将无从谈起。移除门控机制导致的性能降幅位居第二，证明了简单的特征拼接无法替代动态权衡策略，门控机制保障了模型在消除分布差异的同时保留关键语义细节。而移除生成对抗网络模块导致的模型性能下降，则印证了对抗学习在宏观层面上消

除典籍著述风格差异、缓解图谱结构异构方面的核心作用。此外，当用常规随机负采样替换模糊边界负采样损失后，模型性能也出现了下降。这证实了针对本草领域形近义远术语设计的模糊采样策略，对于提升模型在处理易混淆实体时的精细分辨能力至关重要。最后，移除迭代对齐训练策略同样导致了性能的降低，验证了半监督学习思想在低资源场景下的应用价值。

3.3.3 参数敏感性分析

本研究选取模糊负采样语义阈值、伪标签置信度阈值以及对抗生成网络权重3个核心参数进行

敏感性分析，以验证GAFL-Align模型在不同超参数配置下的鲁棒性，实验结果如图4所示。针对模糊负采样语义阈值，模型性能随该值提升呈现先升后降趋势，并在0.8处达到峰值。这表明过低的阈值难以筛选出具有梯度的困难样本，过高的阈值则导致有效负样本匮乏，限制了模糊边界的优化效果。在伪标签置信度阈值方面，模型在参数设为0.9时表现最优。较低的阈值会引入噪声伪标签

导致误差累积，过高的阈值则大幅缩减了训练样本的扩充规模，削弱了自训练收益。此外，对抗生成网络权重在0.2时取得了最佳平衡。权重过小无法有效消除典籍间的分布异构，权重过大会使原始语义丢失，破坏实体原始特征结构。综上所述，适度的参数设定能有效平衡样本质量与模型约束，验证了模型在特定参数区间内的稳定性。

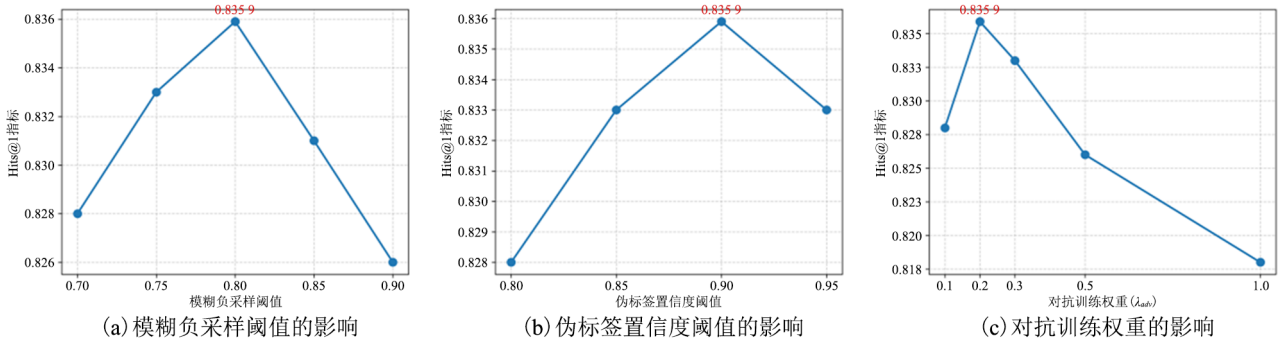


图4 核心超参数对模型性能的影响分析

Fig. 4 Analysis of Impact of Core Hyperparameters on Model Performance

3.3.4 错误案例分析

为深入探究GAFL-Align模型的误差边界与性能

瓶颈，本节对测试集中典型Top-1预测错误案例进行了分类统计与定性分析，如表5所示。

表5 错误案例统计与分析

Tab. 5 Statistics and Analysis of Error Cases

错误类型	局部字面重叠	生僻字干扰	知识缺失
数量	12	17	24
源实体(SKG1)	头眩痛	腹中症坚痞结	薯蓣
真实对齐(SKG2)	头风眩痛	腹癥痃坚积	山药
错误预测(SKG2)	头肿痛	腹中坚痛	地胆
错因分析	错误项“头肿痛”与源实体共享首尾高频字“头”和“痛”，导致字面相似度虚高。模型未能准确捕捉核心病机词“眩”与“肿”的语义差异。	目标实体包含生僻异体字“癥”，“薯蓣”与“山药”为古今异名，字面无一字相同。模型仅凭语义表征与结构特征均无法建立有效关联，导致随机性误判。	

尽管消融实验已证实模糊边界负采样策略降低了本草领域形近义远术语的整体误判率，但在处理字面重叠度极高的长尾样本时，模型性能仍会受限。首先，当候选实体与源实体出现多个相同字符时，BERT的注意力机制倾向于聚焦这些共现特征，从而在一定程度上掩盖了核心差异字的语义。但此类错误的出现，恰恰印证了模型已高效剔除了绝大多数常规的中低难度混淆干扰。其

次是生僻字干扰造成的识别盲区，由于通用BERT模型自身的词表存在一定的局限性，本草典籍中部分异体字常被编码为未知标记，导致关键语义信息缺失，迫使模型只能依赖剩余的常用字进行对齐判别，从而引发预测偏差。最后，由于知识缺失引起语义关联失效，针对少部分古今异名的实体，模型仅凭字面和结构信息难以建立联系，这表明纯数据驱动的方法仍存在性能瓶颈。

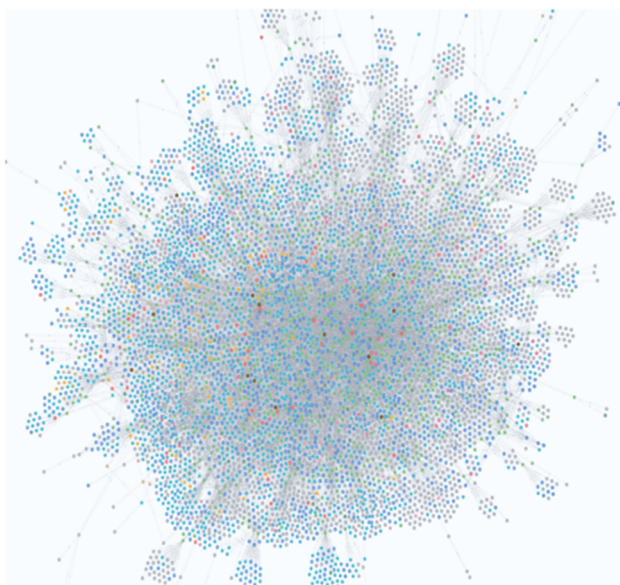
3.4 融合知识图谱构建与结构特征分析

本节应用GAFL-Align模型，实现《神农本草经》与《汤液本草》知识图谱的融合，并定量分析融合效果。构建过程采取分层混合对齐策略：针对“性味”“产地”“炮制方法”等语义明确且表述固定的实体，实施基于规则的精确匹配；对于“中药”“症状”“疾病”以及“功效”存在深层语义异构的复杂实体，部署GAFL-Align模型进行细粒度对齐。此外，为保障知识的完备性，两部典籍各自特有的知识均予以完整保留。最终，将所有被识别为等价

的实体对合并为单一节点并迁移关联关系，生成融合图谱SKG-Fused。

3.4.1 融合知识图谱的可视化展示与案例分析

如图5所示，SKG-Fused在整体规模与拓扑复杂性上远超原始图谱，图右的“白薇”局部网络进一步展示了融合的具体价值。融合后的“白薇”节点汇集了来自两部典籍的知识，其关系网络得到了极大丰富，同时涵盖了功效、别名、产地以及采收方式等多个维度。这证明融合图谱成功构建了一个远比单一典籍更全面、更立体的实体知识网络。



(a) SKG-Fused的整体结构



(b) SKG-Fused的局部示例

图5 融合知识图谱SKG-Fused的整体结构与局部示例

Fig. 5 Overall Structure and Local Example of Fused Knowledge Graph SKG-Fused

SKG-Fused中“白薇”实体的知识来源构成如图6所示。针对“黄芪”与“黄耆”这对经典异名以及“寒热酸痛”与“寒热酸疝”这类存在古今字形差异的样本，模型利用BERT的深层语义编码与GAT的结构聚合能力，成功跨越了字面符号的差异并将其精准识别为等价实体。对于“卒中”和“神昏”这两类已标准化的实体，虽然其在两部典籍中名称一致，但各自关联的药物与症状网络存在显著异构性，GAFL-Align通过对抗学习消除分布偏差，并从结构层面识别出来源不同的两个节点实为同一对象从而实现了合并。这些关键实体在融合后被标记为同时来自两部典籍，这种基于模型计算的融合不仅实现了知识的统一，更通过多源记载的交叉验证显著增强了核心本草知识的可信度。

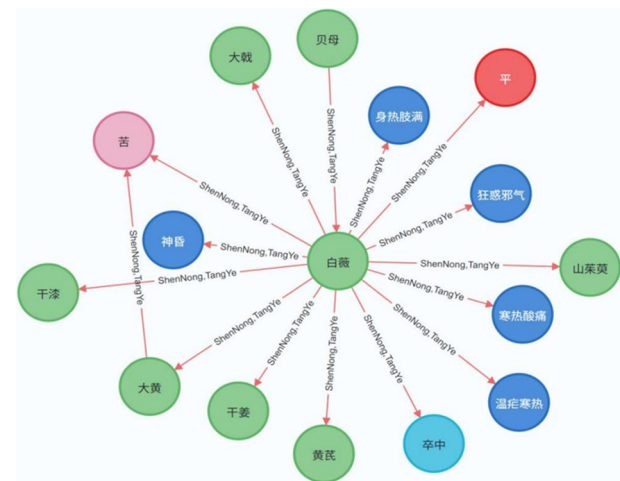


图6 基于典籍来源构成的白薇实体知识融合与对齐效果示例

Fig. 6 Example of Knowledge Fusion and Alignment Effects for Baiwei Entity Based on Classic Sources

3.4.2 融合知识图谱的结构特征定量分析

本节对融合图谱SKG-Fused与两个源图谱进行了定量指标对比,如表6所示。融合图谱在规模上接近源图谱总和,其中855对核心实体通过混合对齐策略被成功合并。分析网络指标可以发现,SKG-Fused的平均度达到了3.7209,高于两个源图谱的3.6424与3.3714。这一结果证明了实体对齐的增益效应,即通过合并等价实体,融合后的节点继承了更丰富的关系属性从而充实了其知识内涵。同时,

融合图谱的平均聚类系数为0.0631,体现了对两个源图谱的有效中和与继承,说明融合策略能够将结构稀疏与紧密的图谱进行有机整合。值得注意的是,融合图谱的图密度有所下降,但这属于符合图论规律的预期现象。此外,SKG-Fused的最大连通分量占比高达98.71%,表明图谱的核心主体保持了高度连通。知识融合并未导致图结构的碎裂,而是在保持主体结构完整性的基础上提升了实体的平均知识丰裕度。

表6 源图谱与融合图谱的网络结构指标统计对比

Tab. 6 Comparison of Network Structural Metrics Between Source and Fused Knowledge Graphs

图谱名称	实体数量	关系数量	平均度	平均聚类系数	图密度	LCC占比/%
《神农本草经》SKG1	3771	7011	3.6424	0.0068	0.000924	97.74
《汤液本草》SKG2	3910	6616	3.3714	0.0825	0.000853	98.68
融合图谱SKG-Fused	6826	13108	3.7209	0.0631	0.000593	98.71

4 讨论

4.1 研究贡献与价值

本研究在方法创新层面突破了本草典籍实体对齐主要依靠人工对齐与表层规则匹配的局限,构建了面向古籍复杂语义特性的自动化对齐框架。针对传统规则方法难以应对典籍异构与术语易混淆的特有挑战,本研究提出了GAFL-Align模型,通过生成对抗网络与模糊边界辨识机制有效攻克了深层语义识别难题。此外,针对古籍领域专家标注资源稀缺且昂贵的困境,本研究引入了置信度感知的迭代自训练策略,利用高置信度伪标签在保证精度的同时大幅降低了对大规模人工标注数据的依赖。本模型实现了低资源场景下的高精度、自动化对齐,为古籍知识的智能化组织提供了高效的方法论支撑。

在知识组织层面,本研究改善了既往本草知识图谱研究中构建与应用环节衔接不足的现状,确立了以融合为核心的知识增值路径。区别于现有研究多局限于单一典籍知识图谱的构建或局部应用开发,本研究将重心聚焦于连接异构图谱的语义桥梁的构建,通过实体对齐技术,实现了《神农本草经》与《汤液本草》典籍内容的对齐与融合。这一工作打通了从上游图谱构建到下游智能应用的数据路径,构建了更加完整有序的本草知识网

络,提升了本草知识的可信度,并为后续典籍可视化分析、智能问答及辅助诊疗等下游任务提供更为完备的数据支持。

4.2 与大语言模型方案的权衡分析

相比于大语言模型方案,GAFL-Align在处理效率上具备显著优势。作为专用小模型,其利用图注意力网络进行大规模并行计算,从而避免了通用大模型将对齐转化为序列化文本检索带来的高推理延迟,更适合海量典籍实体的快速处理。除了速度优势,本模型在可解释性上也提供了独特的算法视角。尽管大模型能生成直观的自然语言解释,但其内部逻辑仍具有一定的黑盒属性,而GAFL-Align能够通过自适应门控机制动态展示语义与结构特征的权重分配,体现决策时的侧重依据,提供了基于算法机制的透明度。更重要的是,本模型摆脱了对高性能计算集群的依赖,仅需单显卡即可高效运行。这种轻量化架构显著降低了硬件门槛与运行开销,为本草古籍数字化整理工作的低成本、规模化开展提供了更具可行性的技术路径。

4.3 局限性与未来工作

尽管GAFL-Align模型在本草典籍实体对齐任务中表现优异,但仍存在以下局限。首先,典籍间的知识组织逻辑存在较大差异时,基于对抗学习的特征映射效果便会受到限制。其次,通用预训练

语言模型难以全面覆盖专业领域生僻字, 导致关键术语的核心语义信息表征缺失, 模型被迫过度依赖邻域结构, 可能会产生预测偏差。最后, 对于字面完全不同的古今异名实体, 纯数据驱动的方法不能利用外部背景知识建立有效的隐式关联。

未来工作将致力于扩展实验规模, 引入更多典籍数据以验证模型在极端异构场景下的泛化能力; 构建本草专用嵌入模型并引入汉字字形或偏旁部首等视觉特征, 解决生僻字编码难题; 探索轻量化模型与大语言模型的协同机制, 为对齐结果提供直观的自然语言事后解释, 实现算法透明度与用户可理解性的双重提升。

5 结论

本研究聚焦于本草典籍知识图谱的实体对齐任务, 提出了一种融合生成对抗网络与模糊语义辨识的实体对齐模型 GAFL-Align。在方法论层面, 本模型通过引入对抗性领域自适应机制, 有效消除了不同典籍间的跨图谱分布偏差, 实现了深层的语义对齐。同时, 本研究设计了模糊边界负采样策略与置信度感知的迭代自训练机制, 不仅攻克了形近义远术语的细粒度辨识难题, 更验证了在低资源环境下降低对人工标注依赖、实现自动化知识聚合的可行性。实验结果表明, GAFL-Align 在 Hits@1 与 MRR 等关键指标上均优于当前主流的基线方法, 证明了其在处理垂直领域复杂异构数据方面的优势。在此基础上, 本研究整合了《神农本草经》与《汤液本草》中分散的知识资源, 构建了一个多源融合知识图谱。综上所述, 本研究不仅提升了本草知识体系的完整性, 更为本草典籍的学术溯源、系统性整理以及传统医学智慧的数字化活化提供了坚实的方法支撑与数据基础。

参 考 文 献

[1] 沈旺, 刘嘉宇, 李贺, 等. 学科领域生命周期视域下国内中医古籍数字化研究进展 [J]. 图书情报工作, 2022, 66 (22): 4-15.

[2] 赵辉, 张华敏, 张磊, 等. 中医古籍文献研究方法述要 [J]. 中国中医药图书情报杂志, 2024, 48 (1): 228-233.

[3] 谢云霏, 贾李蓉. 中医药知识图谱构建技术及应用的研究进展 [J]. 中国医药导报, 2024, 21 (20): 62-66.

[4] 王松, 李正钧, 杨涛, 等. 中医药知识图谱研究现状及发展

趋势 [J]. 南京中医药大学学报, 2022, 38 (3): 272-278.

[5] Weng H, Chen J L, Ou A H, et al. Leveraging Representation Learning for the Construction and Application of a Knowledge Graph for Traditional Chinese Medicine: Framework Development Study [J]. JMIR Medical Informatics, 2022, 10 (9): e38414.

[6] Sun Z Q, Zhang Q H, Hu W, et al. A Benchmarking Study of Embedding-Based Entity Alignment for Knowledge Graphs [J]. Proceedings of the VLDB Endowment, 2020, 13 (12): 2326-2340.

[7] 翟东升, 姜莹, 阙慧敏, 等. 基于多源异构数据的中医药知识图谱构建与应用研究 [J]. 数据分析与知识发现, 2023, 7 (9): 146-158.

[8] Mao X, Wang W T, Xu H M, et al. Relational Reflection Entity Alignment [C]//Proceedings of the 29th ACM International Conference on Information and Knowledge Management. New York: Association for Computing Machinery, 2020: 1095-1104.

[9] Tang X B, Zhang J, Chen B, et al. BERT-INT: A BERT-Based Interaction Model for Knowledge Graph Alignment [C]//Bessiere C. Proceedings of the 29th International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2020: 3174-3180.

[10] 阮彤, 孙程琳, 王昊奋, 等. 中医药知识图谱构建与应用 [J]. 医学信息学杂志, 2016, 37 (4): 8-13.

[11] Chen M H, Tian Y T, Yang M H, et al. Multilingual Knowledge Graph Embeddings for Cross-Lingual Knowledge Alignment [C]//Sierra C. Proceedings of the 26th International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2017: 1511-1517.

[12] Sun Z Q, Hu W, Li C K. Cross-Lingual Entity Alignment via Joint Attribute-Preserving Embedding [C]//Damato C, Fernandez M, Tamma V, et al. Semantic Web-ISWC 2017. Cham: Springer, 2017: 628-644.

[13] Zhang Q H, Sun Z Q, Hu W, et al. Multi-View Knowledge Graph Embedding for Entity Alignment [C]//Kraus S. Proceedings of the 28th International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 5429-5435.

[14] 杨一帆, 刘忠博, 周玉新. 中医知识图谱构建及应用综述 [J]. 软件导刊, 2025, 24 (8): 17-26.

[15] 于彤, 刘静, 贾李蓉, 等. 大型中医药知识图谱构建研究 [J]. 中国数字医学, 2015, 10 (3): 80-82.

[16] Zhou Y, Qi X L, Huang Y, et al. Research on Construction and Application of TCM Knowledge Graph Based on Ancient Chinese Texts [C]//Barnaghi P, Gottlob G, Katsaros D, et al. Proceedings of the 2019 IEEE International Conference on Bioin-

- formatics and Biomedicine. Piscataway: IEEE, 2019: 144-147.
- [17] 李芊芊, 付兴, 杨凤, 等. 基于“病脉证并治”诊疗思维的《伤寒论》知识图谱构建与应用 [J]. 世界科学技术—中医药现代化, 2022, 24 (9): 3613-3621.
- [18] 张卫东, 张晓晓. 中医古籍数字资源知识组织与可视化研究——以《金匱要略》为例 [J]. 情报科学, 2022, 40 (8): 107-117.
- [19] 李贺, 祝琳琳, 刘嘉宇, 等. 基于本体的简帛医药知识组织研究 [J]. 图书情报工作, 2022, 66 (22): 16-27.
- [20] 时倩如, 李贺, 于雯倩, 等. 中医文本命名实体识别研究综述 [J]. 现代情报, 2025, 45 (2): 4-16.
- [21] 李明, 罗晓兰, 朱邦贤. 中医古籍方剂数据挖掘与知识问答系统构建 [J]. 图书馆论坛, 2025, 45 (4): 49-59.
- [22] 许天翼, 石进. 基于循环-过滤检索增强生成技术的中医知识问答系统研究 [J/OL]. 数据分析与知识发现: 1-18 [2026-02-02]. <https://link.cnki.net/urlid/10.1478.G2.20250815.1141.006>.
- [23] 沈旺, 于琳, 冯欣, 等. 基于动态知识图谱的中医疫病古籍知识演化研究 [J]. 现代情报, 2025, 45 (2): 26-37.
- [24] 杨凤, 张艺, 陶晓华, 等. 基于知识元理论与临床需求深度融合的中医古籍数智化研究路径和范式 [J]. 中医杂志, 2024, 65 (12): 1201-1207.
- [25] 齐佳杰, 魏东升, 刘雨婷, 等. 基于“病机链”理念的《脾胃论》湿热成痿肺金受邪论Neo4j知识图谱构建 [J]. 中医药导报, 2024, 30 (11): 172-177.
- [26] 张君冬, 杨松桦, 严颖, 等. 跨医学体系下医疗知识图谱的构建与药物预测研究——以动脉粥样硬化为例 [J]. 情报理论与实践, 2024, 47 (2): 178-188.
- [27] 张富, 杨琳艳, 李健伟, 等. 实体对齐研究综述 [J]. 计算机学报, 2022, 45 (6): 1195-1225.
- [28] Jiang W, Liu Y N, Deng X Y. Fuzzy Entity Alignment via Knowledge Embedding With Awareness of Uncertainty Measure [J]. Neurocomputing, 2022, 468: 97-110.
- [29] Wang C D, Wang W H, Liang Q Y, et al. Local and Global Structure-Aware Contrastive Framework for Entity Alignment [J]. Neurocomputing, 2025, 624: 129445.
- [30] Wang Z C, Lv Q S, Lan X H, et al. Cross-Lingual Knowledge Graph Alignment via Graph Convolutional Networks [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 349-357.
- [31] Long M X, Wang J H, Ma J X, et al. LLM-Augmented Entity Alignment: An Unsupervised and Training-Free Framework [J]. Neural Networks, 2026, 194: 108139.
- [32] Pei S C, Yu L, Hoehndorf R, et al. Semi-Supervised Entity Alignment via Knowledge Graph Embedding With Awareness of Degree Difference [C]//Proceedings of the World Wide Web Conference. New York: Association for Computing Machinery, 2019: 3130-3136.
- [33] Li Y F, Chen L, Liu C F, et al. Generative Adversarial Network for Unsupervised Multi-Lingual Knowledge Graph Entity Alignment [J]. World Wide Web-Internet and Web Information Systems, 2023, 26 (5): 2265-2290.
- [34] Mao X, Wang W T, Wu Y B, et al. Boosting the Speed of Entity Alignment 10x: Dual Attention Matching Network With Normalized Hard Sample Mining [C]//Proceedings of the 30th World Wide Web Conference. New York: Association for Computing Machinery, 2021: 821-832.
- [35] Meng X Z, Jiao X X, Li J H, et al. CELLMEA: A Collaboratively Enhanced Large Language Model-Based Entity Alignment for Aircraft Fault Maintenance [J]. Expert Systems With Applications, 2025, 282: 127630.
- [36] Sun Z Q, Hu W, Zhang Q H, et al. Bootstrapping Entity Alignment With Knowledge Graph Embedding [C]//Lang J. Proceedings of the 27th International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2018: 4396-4402.
- [37] Liu B, Lan T C, Hua W, et al. Dependency-Aware Self-Training for Entity Alignment [C]//Proceedings of the 16th ACM International Conference on Web Search and Data Mining. New York: Association for Computing Machinery, 2023: 796-804.
- [38] Wu Y T, Liu X, Feng Y S, et al. Relation-Aware Entity Alignment for Heterogeneous Knowledge Graphs [C]//Kraus S. Proceedings of the 28th International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 5278-5284.
- [39] Cao Y X, Liu Z Y, Li C J, et al. Multi-Channel Graph Neural Network for Entity Alignment [C]//Korhonen A, Traum D, Marquez L. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 1452-1461.
- [40] Zhang R, Su Y X, Trisedya B D, et al. AutoAlign: Fully Automatic and Effective Knowledge Graph Alignment Enabled by Large Language Models [J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36 (6): 2357-2371.
- [41] Jiang X H, Shen Y H, Shi Z C, et al. Unlocking the Power of Large Language Models for Entity Alignment [C]//Ku L W, Martins A, Srikumar V. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2024: 7566-7583.

(责任编辑: 孔婧媛)