

文章编号:1671-5896(2026)04-0889-07

半监督学习下医疗术语信息关键词模糊检索算法

朱越石¹, 郭凌辉²

(1. 江苏省人民医院 信息处, 南京 210029; 2. 河南大学 人工智能学院, 郑州 450046)

摘要: 医疗领域术语具有强异构性, 且标注数据稀缺, 导致语义特征表达较弱, 进而使检索精度较低, 为此, 提出了半监督学习下医疗术语信息关键词模糊检索算法。利用词向量对原始医疗术语信息进行标准化处理与映射, 基于核心元数据信息节点和用户查询信息节点, 确定检索样本之间的置信度距离, 结合向量空间模型构建医疗术语信息元数据特征空间。在元数据特征空间内, 依据元数据的时间属性, 对各类数据进行合并处理, 以此同步数据索引, 并结合多级索引集群与上下文编码增强语义特征。引入基于一致性正则化与熵最小化原则的 MixMatch 半监督学习算法, 利用少量标注数据对无标签数据进行数据增强, 从而生成无标签数据的低熵伪标签, 增强语义特征表达, 并计算检索词与候选信息的相似度, 实现跨类别或近义术语的高效匹配。实验结果表明, 应用设计的方法进行医疗术语信息关键词检索, 输出的检索结果与输入关键词适配度高于 95%, 可以提高信息检索的精度。

关键词: 半监督学习; 医疗术语; 路径; 信息关键词; 模糊检索

中图分类号: TP391

文献标志码: A

Semi-Supervised Learning-Based Fuzzy Retrieval Algorithm for Medical Term Information Keywords

ZHU Yueshi¹, GUO Linghui²

(1. Information Department, Jiangsu Province Hospital, Nanjing 210029, China;

2. School of Artificial Intelligence, Henan University, Zhengzhou 450046, China)

Abstract: Medical terminology is characterized by strong heterogeneity and scarce labeled data, leading to weak semantic feature representation and low retrieval accuracy. To address these issues, a semi-supervised learning-based fuzzy retrieval algorithm for medical term information keywords is proposed. Word embeddings are utilized to standardize and map original medical term information. Based on core metadata information nodes and user query information nodes, the confidence distance between retrieval samples is determined, and a vector space model is employed to construct a metadata feature space for medical term information. Within this metadata feature space, data are merged according to their temporal attributes, thereby synchronizing data indexes. Semantic features are enhanced through multi-level index clustering and contextual encoding. The MixMatch semi-supervised learning algorithm, based on consistency regularization and entropy minimization principles, is introduced. A small amount of labeled data is used for data augmentation of unlabeled data, generating low-entropy pseudo-labels for unlabeled data to enhance semantic feature representation. The similarity between retrieval terms and candidate information is then calculated, enabling efficient matching across categories or synonymous terms. Experimental results demonstrate that applying the proposed method for medical term

收稿日期: 2025-12-27

基金项目: 江苏省社科应用研究精品工程重点基金资助项目(25SYA-025)

作者简介: 朱越石(1990—), 男, 南京人, 江苏省人民医院(中级)工程师, 硕士, 主要从事医学信息学、电子信息工程研究, (Tel) 86-13951085612 (E-mail) zhuyueshinj@163.com。

information keyword retrieval yields a retrieval result relevance exceeding 95% , effectively improving information retrieval accuracy.

Key words: semi-supervised learning; medical terminology; path; information keywords; fuzzy retrieval

0 引 言

在医疗信息化飞速发展的当下,海量医疗数据不断涌现,其中医疗术语包含了疾病名称、症状描述、药物名称等关键内容,通过对医疗信息的分类存储有利于促进医疗水平的提高和学术发展,但也给信息关键词检索造成了搜索效率低下、检索效果差等困难^[1]。传统的信息关键词模糊剪碎技术通常采用线性映射与特征匹配机制。其一,将多模态信息或数据统一映射至公共表示区间,通过最大化各模态数据间的相关性,提取与关键词对应的信息;第 2 种方法是通过计算检索词与预先制定特征库中的数据项进行匹配度计算,获取检索结果。但这些方法仅关注文本信息与关键词字面意义上的机械匹配,对具有同义词、近义词以及表述差异等术语,无法实现精准匹配,导致检索结果不全面,影响检索准确性。

因此,部分学者借助智能模型以及自然语言处理等先进技术,实现了对信息关键词的深度理解与挖掘,李岩等^[2]建立了一个端到端网络模型,将不规则文本提取和跨模态相似度学习统一到一个框架内,利用学习到的相似度对检测的文本实例排序,从而实现对不规则文本的检索。该技术借助端到端匹配与优化,可降低文本检索过程中检测误差的累积,提升模型的推理速率,但当非文本区域与待检索区域的相似度偏高时,该方法会获取大量错误的检索候选项,使检索准确度下降。师广田等^[3]借助图卷积网络将文本标签信息映射到特征层,同时补充样本缺失的语义信息,结合量化差异性保持损失函数获得判别性较强的检索结果。该方法能有效利用语义信息,擅长处理呈长尾分布的罕见类别。此外,其采用多标签相似性约束,可过滤大部分非语义噪声。然而,该方法忽略了文本标签稀疏性问题,无法充分识别标签间的潜在关联,导致语义嵌入完整度不足,进而使跨模态检索时模态间的语义差距增大,检索精度下降。贾义等^[4]基于去偏对比学习和知识蒸馏相整合的方法,完成了对大规模数据集的检索。该方式采用全监督模型训练弱监督模型的学习策略,可规避假负例的干扰。但若将模型从通用领域迁移至垂直领域,由于源域数据与目标域数据分布差异较大,会导致检索精准度降低。李骛等^[5]提出了基于多标签语义相似性的检索方法,利用模态特征学习与标签生成的相似性约束,输出检索信息列表,该方法能同时处理单标签和多标签数据,提升了检索效率。然而,当样本的多标签间存在语义冲突时,保护间隔与相似性约束难以协调冲突标签,导致生成的哈希码混乱,影响检索准确性。

针对以上问题,笔者以医疗信息资源的智能检索方法为研究对象,基于半监督学习技术,结合实际应用场景开展实验与对比分析。

1 医疗术语信息关键词模糊检索算法设计

1.1 医疗术语信息资源元数据特征空间构建

医疗数据分散在电子病历、影像报告、科研文献等不同系统中,由于术语使用缺乏标准化,导致其存在多义性、同义词、缩写及专业变体等情况,为确保检索结果能覆盖所有相关术语,可基于术语的语义关系、同义词集、上下位词等关联,构建元数据的特征空间,将用户输入的关键词映射到语义相近的术语集合,实现模糊匹配,从而提升检索结果的查全率^[6]。

从电子病历、医学文献、术语库、临床指南等资源中筛选医疗术语,去除拼写错误、非术语内容后,保留符合医学命名规范的术语,将其映射到医学系统命名法-临床术语(SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms)标准术语集中,基于词向量和编辑距离完成术语数据的标准化与映射处理,之后依据元数据信息节点和用户查询的信息节点构建术语元数据特征结构空间^[7],如图 1 所示。

假设医疗术语类型包括 s 类,各类型数据的属性数量最大值为 n ,根据这些属性信息,在元数据特征结构中计算不同样本间的置信距离^[8]。公式为

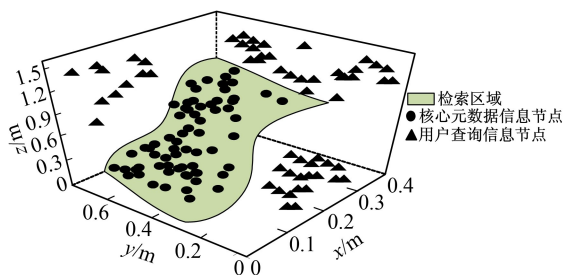


图1 医疗术语元数据特征结构空间

Fig. 1 Metadata feature structure space of medical terminology

$$d(a_i, a_j) = \sqrt{\sum_{k=1}^n \sum_{e=1}^s w_k r_e (N_1 - N_2)^2}, \quad (1)$$

其中 w_k 表示待检索样本的第 k 个属性值, r_e 表示第 e 类术语样本个数, N_1 、 N_2 分别表示医疗术语资源的核心元数据信息节点数量和用户查询的信息节点数量, $d(a_i, a_j)$ 表示样本 a_i 与 a_j 的置信距离。

接下来,将量化后的样本置信距离组织为结构化空间,并采用向量空间模型构建元数据特征空间^[9]。具体公式如下:

$$A = \left[\sum_{t=1}^m E_t(\boldsymbol{\beta} + f)^2 + d(a_i, a_j) \mid \varphi_0^2 \right], \quad (2)$$

其中 m 表示医疗术语元数据特征项个数, E_t 表示第 t 个特征项的检索周期, $\boldsymbol{\beta}$ 表示术语的词向量, f 表示术语信息的储存配置参数, φ_0 表示术语信息切片量, A 表示医疗术语元数据特征空间。

通过系统化地收集、清洗、标准化及量化术语,形成了结构化、可计算的语义表示特征空间,这一空间为模糊检索和跨系统互操作奠定了坚实基础。进一步地,通过挖掘术语间的隐含语义关联,可增强术语的语义信息。

1.2 医疗术语信息语义特征增强

医疗术语具有高度的专业性和复杂性。同一术语可能存在多种表达形式,而不同术语之间也可能存在语义上的关联或重叠。为实现对这些术语进行更深入的语义解析和特征提取,从而更准确地理解其含义和上下文关系,可以在术语元数据特征空间中增强术语信息的语义特征。这样做有助于检索算法识别术语之间的等价关系,避免出现误检情况^[10]。

在元数据特征空间内,首先构建多模态检索索引集群,清晰界定术语的注册信息、数据源、数据类型以及域类型等核心要素,同时整合字段中的时间属性,达成多模态术语信息的索引同步^[11]。该过程可由

$$T_y = \frac{F(x)\boldsymbol{\lambda}}{\sum_{y=1}^Y A + z_y \boldsymbol{\psi}} \quad (3)$$

表示。其中 $F(x)$ 表示特征初始化函数, $\boldsymbol{\lambda}$ 表示文档-主题相关矩阵, Y 表示系统中包含的文档集数量, A 表示术语有数据特征空间, z_y 表示文档 y 中包含的词数量, $\boldsymbol{\psi}$ 表示文档主题向量, T_y 表示文档 y 的信息索引。

依据数据集、数据坐标点以及时间范围,按照数据索引结构构建多级索引,同时将结构化元数据转化为数值特征向量^[12]。即:

$$\begin{cases} R = \{p \in P \mid \exists p_\mu \in \chi\}, \\ O = \{q \in Q \mid \exists q_\varepsilon \in \gamma\}, \end{cases} \quad (4)$$

$$L = T_y \sum_{e=1}^s l_e(X_1 X_2, M, R, O), \quad (5)$$

其中 p 表示待检索数据的域类型, P 表示域类型总数, p_μ 表示第 μ 个域类型的数据源量, χ 表示数据源规模, q 表示域节点, Q 表示总域节点数, q_ε 表示第 ε 个域节点的坐标, γ 表示域节点空间范围, R 、 O 分别表示一级索引和二级索引; l_e 表示第 e 类术语权重, X_1 、 X_2 分别表示文本特征和结构特征; M 表示结构化元数据特征; L 表示数值特征。

采用上下文编码方法对术语信息特征进行语义增强^[13]。表达式为

$$L' = \sum U(L, K) \max(\theta \mid \omega_r - \omega_u \mid), \quad (6)$$

其中 U 表示词嵌入维度, K 表示被掩码的词位置集合, θ 表示输出的上下文表示, ω_r 、 ω_u 分别表示距离函数与边际阈值, L' 表示增强后的术语语义特征。

基于此,运用半监督学习技术达成术语信息关键词的模糊检索。

图2给出了医疗术语语义特征增强过程中的可视化结果。图2中通过向量空间或特征投影的方式,直观呈现了增强前后术语特征在语义空间中的分布变化。可以看出,原始术语特征因异构性和模糊性呈现出较为分散的分布状态;而经上下文编码、多级索引与混合增强处理后,语义相似或相关的术语在特征空间中趋于聚集,类别边界更加清晰,这有效验证了笔者所提特征增强方法在提升术语语义表示区分度与一致性方面的作用,为后续基于半监督学习的模糊匹配奠定了高质量的语义基础。

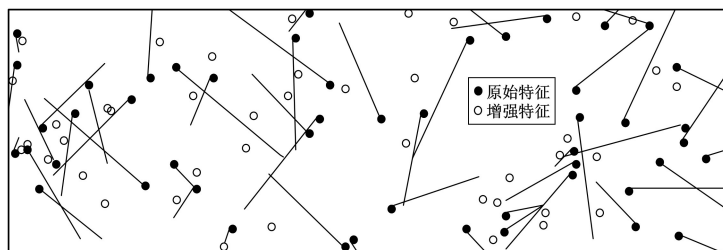


图2 术语语义特征增强可视化

Fig.2 Visualization of semantic feature enhancement for terms

1.3 基于半监督学习的医疗术语关键词检索

笔者运用基于一致性正则化与熵最小化原则的 MixMatch 半监督学习算法开展术语信息模糊检索。MixMatch 算法通过添加正态高斯噪声的方式对每个无标签数据进行 b 次弱增强^[14]。表达式为

$$D = W_1 + L' \frac{1}{(2\pi)^b} \exp\left(-\frac{W_2}{2z}\right), \quad (7)$$

其中 W_1 表示标签术语数据集, L' 表示增强后的术语语义特征, W_2 表示无标签术语数据集, z 表示噪声缩放系数, D 表示增强数据集。

对每个无标签样本 ι_g , 计算其 b 次增强的平均预测概率作为“猜测标签”, 获取低熵伪标签。公式如下:

$$\varphi(\iota_g) = \frac{1}{b} \sum_{\kappa=1}^b p'(\omega | \alpha) + D, \quad (8)$$

其中 p' 表示模型预测的概率分布; ω 表示温度参数 ($\omega \rightarrow 0$ 时输出趋近于 one-hot), α 表示锐化函数, $\varphi(\iota_g)$ 表示无标签样本 ι_g 的低熵伪标签。

将带标签数据 W_1 和无标签数据 W_2 合并后随机打乱, 生成混合数据集, 并结合带标签数据的交叉熵损失 L_1 和无标签数据的损失 L_2 生成总损失, 进而计算查询关键词与候选信息的模糊语义相似度^[15-17]。公式为

$$Q(\sigma, \tau) = \frac{1}{\varphi(\iota_g) |L_1 L_2|} \|\nu - \zeta\|_2^2, \quad (9)$$

其中 ν, ζ 分别表示有标签和无标签数据的加权因子, $Q(\sigma, \tau)$ 表示检索关键词 σ 与候选信息 τ 的相似度。

当检索词与信息的相似度越高时, 表明该信息与检索关键词的匹配度越高, 将相似度最高的信息呈现给用户。

2 实验分析

2.1 数据集描述

为确保实验对比的公平性, 实验选取医疗术语检索领域常用的 PubMed 医学文献摘要数据集作为研究数据。该数据集包含 3 000 万余条医疗文献摘要信息, 涵盖生物医学、临床医学、公共卫生等领域。遵循文献数据选取方式, 经脱敏抽取后, 共获得约 120 万条医疗术语记录, 总字数约 2.4 亿字。随机抽取 5 万条附带术语边界标注及同义词关系的术语信息作为有标签数据, 剩余 115 万条数据作为半监督学习中的无标签数据。

数据集集中的医疗术语呈现长尾效应,其中“发热”“咳嗽”等高频术语占比超60% ,“嗜酸性粒细胞增多症”等低频术语占比不足5%。术语类型涵盖疾病名称、症状表现、检查项目和药物名称等。为便于实验与分析,将英文术语转换为中文,并删除停用词,通过专家标注构建同义词环,即为每个术语生成5个同义词表述。

2.2 实验准备

在计算机上部署自然语言处理工具包(NLTK: Natural Language Toolkit)的 word_tokenize 和 punkt 分词器,对原始数据进行清洗、分词与标注处理,划分训练集、验证集和测试集,生成新的数据集,具体信息如表1所示。

表1 数据集信息
Tab.1 Dataset information

信息	训练集	验证集	测试集	信息	训练集	验证集	测试集
标注数据/万条	16	2	2	文献数/篇	1 500	200	500
未标注数据/万条	144	18	18	最大特征维度	5	5	5
稀疏度	0.5	0.48	0.33	特征数	20	50	50
术语类型数/种	50	40	40				

基于医疗术语元数据的基本信息,构建多级索引集群,具体结构如表2所示。

表2 医疗术语元数据索引结构
Tab.2 Medical terminology metadata index structure

域	域类型	是否存储	域说明
term_id	string	true	医疗术语的唯一标识符
term_name	string	true	医疗术语的具体名称
term_type	string	true	医疗术语所属的类型
synonyms	text_ik	true	医疗术语的同义词集合
definition	text	true	医疗术语的详细定义
related_terms	text_ik	true	与当前医疗术语相关的其他术语集合
clinical_department	string	true	该医疗术语主要涉及的临床科室
source_literature	text_ik	true	包含该医疗术语的文献信息集合
update_time	date	true	医疗术语信息的最后更新时间
popularity_score	int	true	医疗术语的热度评分

采用基于一致性正则化和熵最小化原则的 MixMatch 半监督学习算法对建立索引集群的未标注数据添加高斯噪声随机扰动,参数设置如下:高斯混合系数 $\varepsilon'=0.75$, 温度参数 $\omega=0.5$, 无标签数据使用比例 $W_1:W_2=1:3$, 动量 $\iota'=0.9$, 学习率 $\theta_o=0.001$, 权重衰减 $\varpi=0.01$, 未标注损失权重 $\gamma_w=100$ 。每轮对标注数据做 $b=1$ 次增强,对未标注数据做 $b=2$ 次增强,取模型预测的平均值,再通过锐化生成伪标签。伪标签生成的可视化结果如图3所示。

利用有标签数据对未标注数据开展半监督训练,在测试集上生成了10类伪标签数据,每个颜色对应一个类别,由此可见,半监督算法具备较强的类型区分能力,能获得更高质量的标签表示,为后续术语模糊检索奠定基础。生成伪标签后,与标注数据混合进行 MixUp 生成新样本,计算模型总损失后求取检索词与候选项的相似度。

2.3 实验结果

为测试笔者提出的半监督学习算法在实际检索应用中的有效性,将其作为实验组,将文献[2]基于跨模态相似度学习的端到端不规则文本检索方法和文献[3]标签驱动语义感知学习的跨模态哈希检索

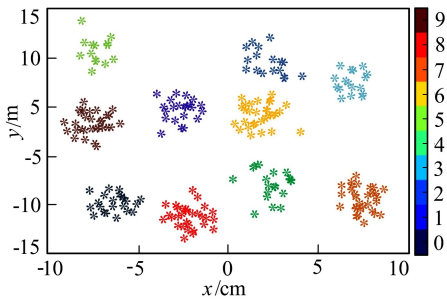


图3 未标注数据伪标签生成
Fig.3 Pseudo-label generation for unlabeled data

方法作为对照组,主要对比各种方法的检索精度。以 29 个医疗术语关键词作为输入,采用 3 种方法进行检索后,统计检索信息与关键词的适配度,结果如图 4 所示。

通过图 4 可知,文献[2]方法由于过度依赖不同模态数据的质量和一致性,导致对不规则文本信息的语义特征提取不够全面,检索结果与关键词的匹配精度偏低;在文献[3]方法的检索过程中,将高维数据映射为低维哈希编码时会造成信息损失,进而导致检索结果与输入关键词的适配度下降;笔者方法可充分利用大量有标签数据,挖掘无标签数据的内在结构与分布特征,进而提升模型对医疗术语的理解与检索能力,使检索结果与关键词的适配度达到 95% 以上,保持了优异的检索性能。

在此基础上,采用归一化折损累积增益作为评估指标,其数值越接近 1.0,表示检索结果的质量越好,检索信息越准确。实验结果如图 5 所示。

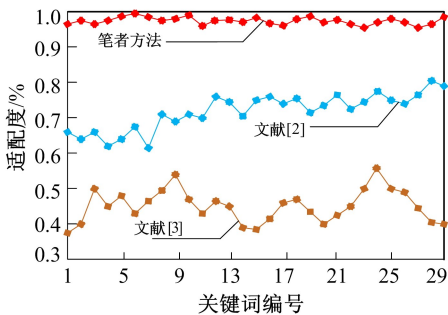


图4 各方法关键词检索的适配度对比

Fig.4 Comparison of adaptability of keyword retrieval across various methods

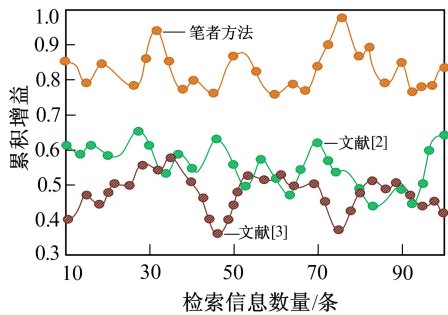


图5 归一化折损累积增益对比结果

Fig.5 Normalized cumulative gain comparison results

由图 5 可知,在针对医疗术语信息关键词模糊检索场景中,文献[2-3]方法的累积增益分别稳定在 0.42 ~0.68 和 0.34 ~0.59 区间内,而笔者方法的累积增益大小在 0.76 ~0.98,明显高于对照组方法,这一结果有效地证明了笔者所应用的半监督学习算法在医疗术语信息检索方面展现出了更强的搜索能力,检索精度得到提升。

3 结 语

为帮助用户准确、高效地定位所需的医疗信息,笔者研究一种基于半监督学习的医疗信息关键词模糊检索算法。通过构建医疗术语元数据特征空间与增强信息语义特征,通过半监督学习确定检索结果与关键词的相似度,并由此实现对医疗信息关键词的有效检索。通过与当前主流方法进行对比验证分析,所提方法在检索精度上展现出显著优势,进一步深化了半监督学习技术在医疗领域的应用,尤其适用于海量医疗信息资源中对关键词检索结果一致性要求较高的场景,为医疗领域重用历史数据、辅助医疗数据分析提供了有力参考。

参考文献:

[1]唐秀,伍赛,侯捷,等. 面向多模态模型训练的高效样本检索技术[J]. 软件学报, 2024, 35(3): 1125-1139.
TANG X, WU S, HOU J, et al. Efficient Sample Retrieval Techniques for Multimodal Model Training [J]. Journal of Software, 2024, 35(3): 1125-1139.

[2]李岩,张敏艺,宿汉辰,等. 基于跨模态相似度学习的端到端不规则文本检索方法[J]. 无线电工程, 2023, 53(3): 501-507.
LI Y, ZHANG M Y, SU H C, et al. End-to-End Irregular Text Retrieval via Cross-Model Similarity Learning [J]. Radio Engineering, 2023, 53(3): 501-507.

[3]师广田,张峰,张辉,等. 标签驱动语义感知学习的跨模态哈希检索方法[J]. 重庆邮电大学学报(自然科学版), 2025, 37(6): 870-883.
SHI G T, ZHANG F, ZHANG H, et al. Label-Driven Semantic-Aware Learning for Cross-Modal Hashing Retrieval [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2025, 37(6): 870-883.

[4]贾义,毛佳昕. 基于弱监督学习的稠密向量检索模型[J]. 中文信息学报, 2025, 39(10): 156-166.

- JIA Y, MAO J X. Dense Retrieval Models Based on Weakly Supervised Learning [J]. Journal of Chinese Information Processing, 2025, 39(10): 156-166.
- [5] 李骛, 谢委衡, 高天宇, 等. 基于多标签语义相似性引导的跨模态哈希检索方法 [J]. 光电子·激光, 2025, 36(10): 1063-1077.
- LI A, XIE W H, GAO T Y, et al. Cross-Modal Hashing Retrieval Method Based on Multi-Label Semantic Similarity Guiding [J]. Journal of Optoelectronics · Laser, 2025, 36(10): 1063-1077.
- [6] 肖泉彬, 陈源, 吴毅坚, 等. 基于代码克隆差异分析的函数模板挖掘和检索方法 [J]. 软件学报, 2025, 36(6): 2774-2793.
- XIAO Q B, CHEN Y, WU Y J, et al. Function Template Mining and Retrieval Based on Code Clone Difference Analysis [J]. Journal of Software, 2025, 36(6): 2774-2793.
- [7] 彭姣, 贺月, 商笑然, 等. 基于渐进原型匹配的文本-动态图片跨模态检索算法 [J]. 计算机科学, 2025, 52(9): 276-281.
- PENG J, HE Y, SHANG X R, et al. Text-Dynamic Image Cross-Modal Retrieval Algorithm Based on Progressive Prototype Matching [J]. Computer Science, 2025, 52(9): 276-281.
- [8] 王丹, 张峰, 张辉, 等. 基于信息互补与交叉注意力的跨模态检索方法 [J]. 计算机应用研究, 2025, 42(7): 2032-2038.
- WANG D, ZHANG F, ZHANG H, et al. Information Complementarity and Cross-Attention for Cross-Modal Retrieval [J]. Application Research of Computers, 2025, 42(7): 2032-2038.
- [9] 陈曦, 彭姣, 张鹏飞, 等. 基于预训练模型和编码器的图文跨模态检索算法 [J]. 北京邮电大学学报, 2023, 46(5): 112-117.
- CHEN X, PENG J, ZHANG P F, et al. Cross-Modal Retrieval Algorithm for Image and Text Based on Pre-Trained Models and Encoders [J]. Journal of Beijing University of Posts and Telecommunications, 2023, 46(5): 112-117.
- [10] 夏鑫雨, 朱磊, 聂秀山, 等. 典型概念驱动的模式缺失深度跨模态检索 [J]. 计算机辅助设计与图形学学报, 2025, 37(3): 519-532.
- XIA X Y, ZHU L, NIE X S, et al. Typical Concept-Driven Modality-Missing Deep Cross-Modal Retrieval [J]. Journal of Computer-Aided Design & Computer Graphics, 2025, 37(3): 519-532.
- [11] 叶亚芬, 原德巍. 基于数据化和文本检索技术的档案资源智能聚类研究 [J]. 浙江档案, 2023(8): 17-19.
- YE Y F, YUAN D W. Research on Intelligent Clustering of Archival Resources Based on Data Digitization and Text Retrieval Technology [J]. Zhejiang Archives, 2023(8): 17-19.
- [12] 杨梦雅, 赵琰, 薛亮. 基于改进的 Vision Transformer 深度哈希图像检索 [J]. 陕西科技大学学报, 2025, 43(4): 183-191.
- YANG M Y, ZHAO Y, XUE L. Deep Hashing Method Based on Improved Vision Transformer [J]. Journal of Shaanxi University of Science & Technology, 2025, 43(4): 183-191.
- [13] 齐丹丹, 王长征, 郭少茹, 等. 基于主题多视图表示的零样本实体检索方法 [J]. 广西师范大学学报(自然科学版), 2025, 43(3): 23-34.
- QI D D, WANG C Z, GUO S R, et al. Topic-Based Multi-View Entity Representation for Zero-Shot Entity Retrieval [J]. Journal of Guangxi Normal University (Natural Science Edition), 2025, 43(3): 23-34.
- [14] 肖玥. 跨学科知识生成导向的交互式信息检索模型构建 [J]. 图书馆建设, 2025(2): 143-153.
- XIAO Y. Construction of Interactive Information Retrieval Model Guided by Interdisciplinary Knowledge Generation [J]. Library Development, 2025(2): 143-153.
- [15] 黄溪, 王先兵, 林海, 等. 基于自集成视觉 Transformer 的图像检索 [J]. 武汉大学学报(工学版), 2024, 57(12): 1795-1802.
- HUANG X, WANG X B, LIN H, et al. Image Retrieval Based on Self-Ensemble Vision Transformer [J]. Engineering Journal of Wuhan University, 2024, 57(12): 1795-1802.
- [16] 唐莹莹, 陈玉玲, 罗运, 等. 基于全同态加密的可验证多关键词密文检索方案 [J]. 计算机工程, 2025, 51(4): 188-197.
- TANG Y Y, CHEN Y L, LUO Y, et al. Verifiable Multi-Keyword Ciphertext Retrieval Scheme Based on Fully Homomorphic Encryption [J]. Computer Engineering, 2025, 51(4): 188-197.
- [17] 赵周颖, 余正涛, 黄于欣, 等. 基于要素关联图的汉越跨语言事件检索方法 [J]. 现代电子技术, 2024, 47(7): 127-132.
- ZHAO Z Y, YU Z T, HUANG Y X, et al. Chinese-Vietnamese Cross-Lingual Event Retrieval Method Based on Arguments Relational Graph [J]. Modern Electronics Technique, 2024, 47(7): 127-132.