面向儿科医学试题的答案解析自动推荐方法研究*

王 娟[1] 侯 丽[1] 孙月萍[1] 李佳明[1] 杨 丽[1] 董良广[2] 李云汉[3]

([1] 中国医学科学院/北京协和医学院医学信息研究所 北京 100020 [2] 人民卫生出版社有限公司 北京 100021
[3] 北京市测绘设计研究院 北京 100038)

〔摘要〕 **目的/意义** 探索并实现儿科医学试题答案的自动化解析，提高试题答案解析编撰效率与质量。**方法/过程** 提出一种隐性语义索引、MC－BERT 和 CoSENT 模型相结合的方法。首先使用基于隐性语义索引的方法和 MC－BERT 模型从参考文档中抽取多个候选答案解析，然后利用 CoSENT 模型计算候选解析、试题题干和答案选项之间的相似度，选取相似度最高的候选解析作为最终答案解析。**结果/结论** 该方法答案解析精确率达到 72.6%，相较单一方法或模型明显提高查全率和精确率，有效提高了编撰试题答案解析的效率，减轻教育工作者负担，并可为教育研究提供重要的数据支持。

〔关键词〕 医学试题；答案解析；隐性语义索引；MC－BERT；CoSENT；自然语言处理
〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2024.10.003

Study on the Automatic Recommendation Method for Answer Analysis of Pediatric Medical Examination Questions

WANG Juan[1], *HOU Li*[1], *SUN Yueping*[1], *LI Jiaming*[1], *YANG Li*[1], *DONG Liangguang*[2], *LI Yunhan*[3]
[1] *Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China*; [2] *People's Medical Publishing House Co. Ltd., Beijing 100021, China*; [3] *Beijing Institute of Surveying and Mapping, Beijing 100038, China*

〔**Abstract**〕 **Purpose/Significance** To explore and implement the automated interpretation of pediatric medical exam answers, so as to enhance the efficiency and quality of answer explanation compilation. **Method/Process** The paper proposes a method that combines latent semantic indexing (LSI), MC－BERT, and the CoSENT model. Initially, multiple candidate answer explanations are extracted from reference documents using the LSI method and the MC－BERT model. Subsequently, the CoSENT model is employed to calculate the similarity between the candidate explanations and the question stems as well as the answer options. The candidate explanation with the highest similarity is then selected as the final answer explanation. **Result/Conclusion** The experimental results show that the method presented in this paper achieves a precision rate of 72.6%. Compared to single methods or models, it significantly improves the recall and precision of answer parsing, effectively enhances the efficiency of compiling question answer explanations, reduces the burden on educators, and provides significant data support for educational research.

〔**Keywords**〕 medical test questions; answer analysis; latent semantic indexing (LSI); MC－BERT; CoSENT; natural language processing (NLP)

〔修回日期〕 2024－09－09
〔作者简介〕 王娟，研究实习员，发表论文5篇；通信作者：侯丽，研究员。
〔基金项目〕 中国医学科学院医学与健康科技创新工程项目（项目编号：2021－I2M－1－001）；医学融合出版知识技术重点实验室。

1 引言

医学试题用于评估学生对医学知识和技能的掌握情况，答案解析能够引导学生理解和分析问题，促进深层次学习。自动解析能提供详细的答案解释和背景知识，帮助学生迅速理解试题要点和医学概念，提高学习效率，同时减轻教育工作者的负担。为此，本文以儿科医学试题为例，提出一种结合隐性语义索引（latent semantic indexing，LSI）、中文生物医学文本挖掘的概念化表示学习（conceptualize representation learningapproach，MC－BERT）与 CoSENT（cosine sentence BERT）模型的自动推荐方法，以提高试题编撰效率并提供可选择的答案解析。

2 相关研究

2.1 基于预训练模型的方法

预训练模型可以利用大量数据学习深层次语言特征，提高解析获取能力。Dimitriadis D 等[1]利用预训练的提取式问答模型自动获取“是/否”型问答数据集中的答案解析；Yu S 等[2]提出一种基于解析过滤的方法，使用双向编码器表征（bidirectional encoder representations from transformers，BERT）模型处理上下文、问题等，隐式地从上下文提取解析，并进行对比和筛选；Wang Q 等[3]提出一种包含实体掩码机制的两阶段解析检索方法，从全局语义和实体方面获取解析；Alkhaldi T 等[4]提出一种联合训练方法，训练问答模型和答案解析检索模型，加强文档间的桥接；Huang Y 等[5]引入一种新型递归消除记忆网络，利用预训练的生成模型为问题生成定制的候选解析，通过一个模块来细化解析，递归地删除不相关的低质量解析。

2.2 基于图神经网络的方法

图神经网络（graph neural networks，GNN）能够有效处理结构化数据，捕捉实体之间的复杂关系和依赖性，适合于解析推荐任务。Christmann P 等[6]构建大型异构图，结合问题级别的注意力机制使用 GNN 对图进行迭代处理，使图逐渐被减小，直至提炼出最佳答案解析；Zhu P 等[7]收集问题相关的解析来构建图，使用 ELECTRA 编码解析，利用 GATv2 算法在图上进行推理，以生成最终的答案解析；Jia M 等[8]构建动态推理图，结合句子间的显式关键词和隐式语义信息，通过 GNN 预测句子级别的答案解析；Yang N 等[9]提出一种多跳注意力 GNN 模型，结合答案与解析的对比学习损失，通过答案与解析的相似性区分解析句子与其他混淆句子；Si J 等[10]利用图卷积神经网络（graph convolutional networks，GCN）合并多条解析之间的拓扑交互信息，以学习解析表示。

2.3 基于信息检索技术的方法

使用先进信息检索技术和自然语言处理技术的集成方法从大规模数据中提取、处理并理解复杂的文本信息，从而生成更准确、更相关的答案解析。这种集成方法强化了系统的整体性能，能够更有效地处理复杂的查询和信息检索任务。Chen Z 等[11]提出 EvidenceNet 模型，从不同级别的参考文档信息中捕获全局上下文信息，利用门控机制过滤冗余信息，并提出对称的交互注意力机制，以识别问题与解析之间的相互作用；Zhao Y 等[12]综合集成多粒度答案解析，并将这些答案解析与原始段落相结合；Fan Y 等[13]首先对答案解析进行直观判断，过滤不相关的解析以减少搜索空间，然后设计一个演绎－溯因双向推理模块，并构建交叉验证和多视角对比学习；You H 等[14]提出无监督多粒度解析检索方法，考虑以多种交互获得答案解析框架，以分数融合机制建模句子级和段落级检索之间的单向引导关系，以门控选择机制检索解析段落；Yadav V 等[15]首先从 BM25 信息检索模型选出前 n 个句子，然后由这些句子组成多个候选答案解析句集合，根据公式对集合排序，得分最高的句子集合则成为最终的答案解析；Yadav V 等[16]还采用 GloVe 嵌入的无监督对齐方法，利用对齐模型找到与当前查询（初始化的问题和候选答案文本）在嵌入空间中最接近的解析句。

预训练模型能快速利用大量数据的深层语言特征，提升解析的准确性和效率，在多项自然语言处理

任务中表现优异，但其训练和微调需较多计算资源。GNN 擅长处理结构化数据，能捕捉复杂关系，适用于密集连接的数据，但对非结构化数据效率较低，且构建精确图结构需大量前期工作。信息检索技术侧重从大规模文本中高效提取信息，但对语义复杂或多义性强的查询可能难以确保准确性，且对算法优化要求较高。本文根据以上研究启发，调研使用较多的预训练模型 BERT 及相关变种，结合本研究任务特点，选择更合适的 MC－BERT 模型和 CoSENT 模型，同时结合信息检索技术中发展成熟的 LSI，将以上模型、方法进行集成，提高试题解析效率。

3 研究方法

3.1 任务分析

3.1.1 任务基本情况 本研究任务是基于试题和答案选项进行查询，从参考文档中推荐答案解析。部分试题已按照《儿科学》教材章节进行标注，见表 1，示例试题所属章节为“内分泌疾病－性早熟－诊断与鉴别诊断”，从而缩小检索范围。对于缺少章节标注的试题，则需采用其他方法进行答案解析推荐。

3.1.2 数据基本情况 该任务包含 3 种数据成分：试题题干、答案选项和参考文档。参考文档来自合作单位提供的组织成段落的半结构化文本《儿科学》XML 电子教材。将内容划分为 135 小节，每节包含序号、标题、上级章节、子章节和具体内容，便于系统理解和检索。数据包括 10 239 道试题，每道试题包含试题 ID、所属章节、题干、选项及正确答案。错误选项在参考文档中可能没有相关解析，因此不计入评价指标。

表 1 儿科医学试题答案解析推荐示例

试题所属章节	题干	正确答案选项	答案解析
内分泌疾病、性早熟、诊断和鉴别诊断	由于下丘脑对性激素的负反馈作用下降，使促性腺激素释放激素过早分泌所致的性早熟为（ ）。	A. 特发性性早熟	第十五章 内分泌疾病 第四节 性早熟（1）特发性性早熟：又称体质性性早熟，是下丘脑对性激素负反馈的敏感性下降、促性腺素释放激素过早增加分泌所致。

3.1.3 数据分析 分析数据集发现，试题涉及广泛的医学专业知识。儿科学作为综合性学科，领域复杂，答案解析的检索需理解上下文。部分试题数据包含所属章节。因此提出一种自动答案解析推荐方法，同时利用 LSI 和 MC－BERT 模型两种方法捕捉文本语义获取候选答案解析，其中 LSI 可以通过章节信息限制检索范围，最后利用 CoSENT 模型对解析进行排序，提升性能。

3.2 方法设计

提出一种集成推荐方法，结合 LSI、MC－BERT 和 CoSENT 模型的优势，以提高答案解析的速度和准确性。方法流程，见图 1。数据预处理后，首先进行初步答案解析提取，基于 LSI 与章节信息的答案解析推荐方法（answer analysis recommendation method based on LSI and chapter information，ARM－LCI）进行快速文本筛选，从章节内容中提取相关信息。基于 MC－BERT 的答案解析推荐方法（answer analysis recommendation method based on MC－BERT，ARM－MC－BERT）进行深度语义分析，以捕捉复杂的语义联系和深层含义。并行处理上述两种方法，允许系统快速且全面地收集多个候选答案解析。然后对答案解析进行评估与选择，基于 CoSENT 的答案解析排序方法（answer analysis ranking method based on CoSENT，ARM－CoSENT）计算候选答案解析与试题题干及选项的相似度。最后，选择最匹配的答案解析作为最终推荐，确保解析的准确性和相关性。

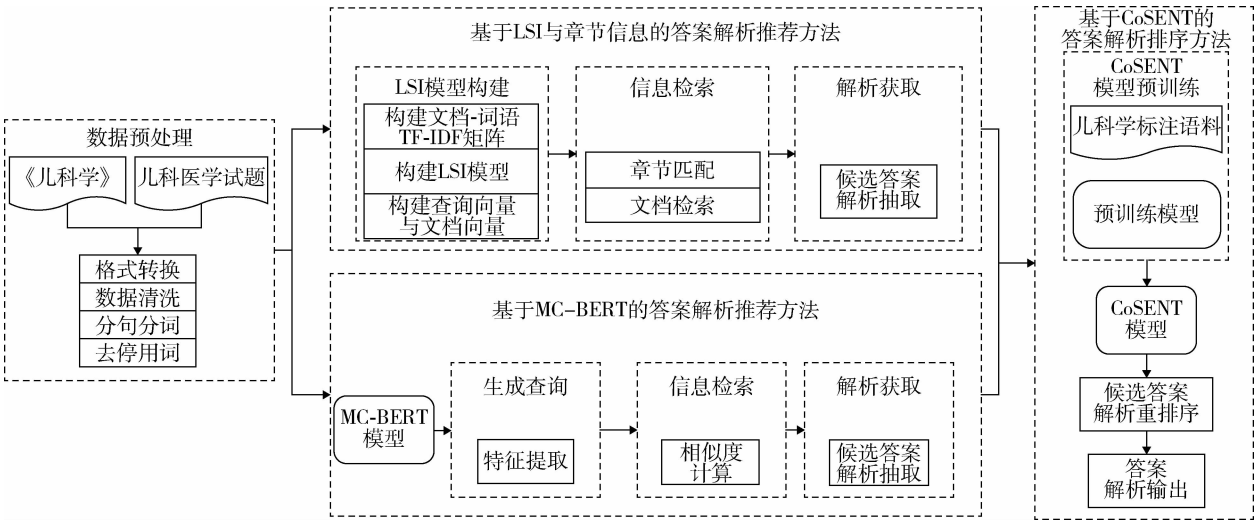

图1 儿科医学试题答案解析自动推荐方法流程

3.2.1 基于LSI与章节信息的答案解析推荐方法 LSI是一种文本挖掘和信息检索技术，旨在发现文本数据中的潜在语义结构。通过奇异值分解，识别并提取文档中的潜在语义结构，从而改善文本检索和分析性能。采用LSI根据章节信息快速过滤和定位相关答案解析，为每道试题生成初步候选答案解析，具体步骤如下。一是文本预处理，生成文档和试题的文本表示。二是构建文档-词语的词频-逆向文件频率（term frequency - inverse document frequency，TF-IDF）矩阵，计算词频和逆文档频率。三是使用奇异值分解对矩阵降维，得到文档-概念矩阵（U矩阵）和词语-概念矩阵（V矩阵），以及查询向量（每个试题的题干和答案选项）与文档概念表示。四是根据章节信息筛选文档，若题目包含章节信息，则根据章节筛选相关文档，缩小检索范围；若不存在章节信息，则跳过该步骤。五是通过计算查询向量与文档概念表示的余弦相似度，找到最相关的句子，作为支持答案选项的候选解析。通过ARM-LCI方法得到的候选答案解析将作为中间结果，利用ARM-CoSENT方法进行排序。

3.2.2 基于MC-BERT的答案解析推荐方法 虽然LSI能够处理文本中的语义相关性，降维以减少冗余信息，但其对大规模数据集的处理效率较低，无法捕捉更复杂的语义关系，且部分试题章节信息缺失，无法缩小检索范围，因此应用MC-BERT模型进一步提高答案解析的准确性。MC-BERT是针对中文医学文本优化的预训练语言模型，基于BERT架构，通过大量医学文献和临床报告训练而成，能更好地理解医学领域的语言特点和术语，尤其在疾病诊断分析和医学信息提取等任务中表现优于通用模型。具体步骤如下。一是特征提取，将参考文档中的每个句子作为输入，使用MC-BERT提取其特征表示。对于每篇参考文档中的每个句子 $X = [x_1, x_2, \ldots, x_m]$，其表示为：

$$H_X = \text{MC-BERT}(X) \tag{1}$$

同样地，对于每道试题其题干和对应选项拼接后形成查询。假设查询为 $T = [t_1, t_2, \ldots, t_n]$，其表示为：

$$H_T = \text{MC-BERT}(T) \tag{2}$$

二是相似度计算，计算每道试题与每个参考文档中句子的余弦相似度。三是提取解析，对于每道试题的查询，选取与之相似度最高的前3个句子作为候选答案解析，以提高方法的查全率。基于MC-BERT的答案解析推荐方法可以避免依赖试题的章节信息，因此无论试题是否标注章节信息，均可提供相应的答案解析。

3.2.3 基于CoSENT的答案解析排序方法 以上两种方法均可能得到多个答案解析内容，为提高准确性，采用基于CoSENT的答案解析排序方

法对试题与答案解析进行重排序。CoSENT[17]是一种有监督的句向量模型，用于计算句子级别的语义相似度，能解决 BERT 在语义相似度检索中的时间开销问题，同时弥补句子表征在无监督任务中的不足。CoSENT 与 Sentence - BERT[18]结构相似，训练目标与推理任务目标一致，通过统计正负样本的余弦值差异来优化模型。一是特征提取，使用预训练后的 CoSENT 模型生成嵌入向量：试题题干、对应选项拼接形成的查询 $T = [t_1, t_2, \ldots, t_n]$ 向量和候选答案解折 $E = [e_1, e_2, \ldots, e_m]$ 向量：

$$C_T = \text{CoSENT}(T) \tag{3}$$

$$C_E = \text{CoSENT}(E) \tag{4}$$

二是相似度计算，计算每个答案解析与对应查询的相似度。三是答案解析重排序，对于每个选项 T_i 根据计算出的相似度对候选答案解析 E 进行排序：

$$S(T_i) = \text{sort}(E_1, E_2, \ldots, E_k) \tag{5}$$

通过以上过程，利用 CoSENT 的语义嵌入能力，将每个选项的相关答案解析按照与选项的语义相似度重排序，从而提高选项答案解析的质量和相关性。

4 实验设计与结果分析

4.1 评价指标

为判断答案解析是否有效，使用精确率（precision，P）、召回率（recall，R）、$F1$ 值对实验结果进行评价。

$$\text{P} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{R} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$F1 = 2 \cdot \frac{\text{P} \cdot \text{R}}{\text{P} + \text{R}} \tag{8}$$

其中 TP 表示真正例，即正确检索出的与答案相关的解析句子；FP 表示假正例，即错误地检索出的与答案无关的解析句子；FN 表示假负例，即未能检索出的与答案相关的解析句子。

引入 Spearman 相关系数评价答案解析的重排序。Spearman 相关系数是一个非参数统计度量，用于衡量两个变量排序之间的单调关系（不一定是线性）。其基于变量的排名进行计算，适用于非线性关系和非正态分布的数据。Spearman 系数的取值范围从 -1 到 1，其中 1 表示完全正相关，-1 表示完全负相关，0 表示无相关性。

$$\text{Spearman}(T) = 1 - \frac{6\sum r_i^2}{k(k^2 - 1)} \tag{9}$$

其中，r_i 表示第 i 个答案解析在人工排序和模型排序中的排序差，k 是答案解析的数量，$\sum r_i^2$ 是所有排序差的平方和。

4.2 模型参数

基于 CoSENT 模型的答案解析推荐方法预训练模型为 Chinese - MacBERT - base[19]，MacBERT 是 BERT 的改进版本，引入了纠错型掩码语言模型（MLM as correction，Mac）预训练任务，缓解了“预训练 - 下游任务”不一致的问题。使用预标注的《儿科学》部分试题答案和答案解析语句构建相应的模型训练数据集，最终形成训练集 8 200 条、验证集 1 900 条和测试集 1 900 条，相关实验参数，见表 2。

表 2 基于 CoSENT 的答案解析推荐方法实验参数

参数	参数描述	参数值
learn_rate	学习率	2e - 5
batch_size	数据批处理量	100
train_epochs	训练迭代数	20

4.3 结果分析

选取 2 名具有医学背景的学生独立评估答案解析，判断其是否准确反映答案信息并与试题相关。第 3 名标注者负责一致性检验，比较 2 名标注者的结果。如一致，标注结果即被认为可靠；如不一致，由第 3 名标注者最终评审。儿科医学试题答案解析推荐实验结果，见表 3。同时为了验证本文提出方法的有效性，选择较多文献中使用的 BERT 模型和 RoBERTa 模型进行对比实验。

表 3 儿科医学试题答案解析推荐实验结果（%）

模型	P	R	*F*1
BERT	52.6	91.7	66.9
RoBERTa	57.1	92.3	70.6
ARM – LCI	59.5	98.1	74.1
ARM – MC – BERT	61.9	93.5	74.5
ARM – LCI + ARM – MC – BERT	72.6	98.1	83.4

为确保答案解析的相关性和精确性，本任务中精确率至关重要。LSI 方法因其结构简单、计算复杂度低，适合快速检索和识别未登录词，能够结合章节信息，为接近半数问题提供正确解析。虽然 LSI 在语义理解方面不及 MC – BERT，但其广泛的文档覆盖提高了召回率。MC – BERT 则能更好地理解复杂的语义和句子结构，提供超过一半问题的正确解析。MC – BERT 通过深度学习减少假阳性，提高了精确率，但降低了召回率。将 LSI 与 MC – BERT 结合，精确率可达 72.6%。这种结合利用了 LSI 在章节信息处理方面的优势和 MC – BERT 在语义理解方面的能力，提高了系统的整体鲁棒性和解析准确性。

BERT 模型和 RoBERTa 模型的 *F*1 分数分别达到 66.9% 和 70.6%，展示了这两种模型在本文任务上的基本效力。RoBERTa 模型的精确率、查全率及 *F*1 值均略优于 BERT 模型，反映了 RoBERTa 模型相对于 BERT 模型在结构和训练策略方面的优化改进。实验结果表明，本文提出的多种模型相结合的方法在精确率和查全率方面均优于以上模型；通过适当地结合不同类型的模型，可以显著提升系统的整体性能。

答案解析排序实验结果，见表4，使用 Sentence – BERT 作为对比模型。CoSENT 模型改进了训练策略，有助于更好地理解和表示句子含义。实验结果表明 CoSENT 模型的效果好于 Sentence – BERT 模型，在答案解析排序任务上表现更佳。

表 4 答案解析排序实验结果（%）

模型	Spearman
Sentence – BERT	58.3
CoSENT	66.7

儿科医学试题答案解析自动推荐结果示例，见图 2。其中“B:”为试题 ID 和试题考察知识所属《儿科学》章节；“Q:”为试题题干；“R:”为正确答案选项；“W:”为错误答案选项。

B:8a36170724114c29a4c6b0c02c57d611 生长发育 神经心理发育 神经系统的发育
Q:4岁左右儿童可出现
R:神经髓鞘发育完成
978-7-117-26664-2(2018.7);TX;2;_2-5_42-45.txt 神经髓鞘的形成和发育约在4岁完成,在此之前,尤其在婴儿期,各种刺激引起的神经冲动传导速度缓慢,且易于泛化
W:Babinski征阳性可为生理现象
978-7-117-26664-2(2018.7);TX;2;_2-5_42-45.txt 3~4个月前的婴儿肌张力较高,凯尔尼格征可为阳性,2岁以下儿童巴宾斯基征阳性亦可为生理现象
W:握持反射消失
978-7-117-26664-2(2018.7);TX;2;_2-5_42-45.txt 握持反射应于3个月时消失
W:腹壁反射能稳定引出
978-7-117-26664-2(2018.7);TX;2;_2-5_42-45.txt 婴儿肌腱反射较弱,腹壁反射和提睾反射也不易引出,到1岁时才稳定
W:提睾反射不易引出
978-7-117-26664-2(2018.7);TX;2;_2-5_42-45.txt 婴儿肌腱反射较弱,腹壁反射和提睾反射也不易引出,到1岁时才稳定。

图 2 儿科医学试题答案解析自动推荐结果示例

尽管本文提出的方法在实际应用中有效，但对错误示例的分析揭示了一些问题。首先，由于试题多为多选题，方法在某些情况下未能提供完整的正确解析，影响了精确率。其次，部分试题内容超出参考文档范围，无法提供相关解析。最后，《儿科学》文档中包含大量图片和表格，当前方法未针对这些数据进行深入分析，可能影响评估结果。未来的优化应重点关注多选题的全面解析、扩展参考文档范围以及提取图片和表格中的信息。

5 结语

为自动提供儿科医学试题的答案解析推荐，本文提出了一种融合 LSI、MC – BERT 和 CoSENT 模型的方法。该方法首先使用 LSI 和 MC – BERT 模型提取候选答案解析，然后通过 CoSENT 模型选择最相似的解析。实验结果表明，该方法的精确率达 72.6%，展现了减轻教育工作人员负担的潜力，并为医学生学习和教师教学模式创新提供技术支持。

未来研究将重点解决多选题解析、扩展参考文档数据源以及处理图片和表格信息的问题，如增加数据来源的多样性和采用多模态学习方法，以进一步完善自动答案解析推荐系统。

作者贡献：王娟负责技术实施、论文撰写与修订；侯丽负责技术指导、论文修订；孙月萍负责技术指导；李佳明、杨丽、董良广负责数据处理；李云汉负责技术指导。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 DIMITRIADIS D, TSOUMAKAS G. Enhancing yes/no question answering with weak supervision via extractive question answering [J]. Applied intelligence, 2023, 53 (22): 27560 – 27570.

2 YU S, ZHANG H, JING W, et al. Context modeling with evidence filter for multiple choice question answering [C]. Singapore: 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.

3 WANG Q, FENG J, XU G, et al. Two – phase semantic retrieval for explainable multi – hop question answering [C]. Changsha: International Conference on Neural Information Processing, 2023.

4 ALKHALDI T, CHU C, KUROHASHI S. Flexibly focusing on supporting facts, using bridge links, and jointly training specialized modules for multi – hop question answering [EB/OL]. [2024 – 03 – 14]. https://ieeexplore.ieee.org/document/9580589.

5 HUANG Y, FANG M, ZHAN X, et al. Rem – net: recursive erasure memory network for commonsense evidence refinement [C]. online: The AAAI Conference on Artificial Intelligence, 2021.

6 CHRISTMANN P, SAHA ROY R, WEIKUM G. Explainable conversational question answering over heterogeneous sources via iterative graph neural networks [C]. Taipei: The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023.

7 ZHU P, YUAN Y, CHEN L. ELECTRA – based graph network model for multi – hop question answering [J]. Journal of intelligent information systems, 2023, 61 (3): 819 – 834.

8 JIA M, LIAO L, WANG W, et al. Keywords – aware dynamic graph neural network for multi – hop reading comprehension [J]. Neurocomputing, 2022, 501 (8): 25 – 40.

9 YANG N, YANG M. Multi – hop attention GNN with answer – evidence contrastive loss for multi – hop QA [C]. Queensland: 2023 International Joint Conference on Neural Networks (IJCNN), 2023.

10 SI J, ZHU Y, ZHOU D. Exploring faithful rationale for multi – hop fact verification via salience – aware graph learning [C]. Washington DC: The AAAI Conference on Artificial Intelligence, 2023.

11 CHEN Z, HUI S C, ZHUANG F, et al. Evidencenet: evidence fusion network for fact verification [C]. New York: The ACM Web Conference 2022, 2022.

12 ZHAO Y, ZHAO H, DUAN S. Multi – grained evidence inference for multi – choice reading comprehension [EB/OL]. [2024 – 03 – 14]. https://ieeexplore.ieee.org/document/10254212.

13 FAN Y, ZHANG H, LI R, et al. FRVA: fact – retrieval and verification augmented entailment tree generation for explainable question answering [C]. St. Julian's, Malta: Findings of the Association for Computational Linguistics ACL 2024, 2024.

14 YOU H. Multi – grained unsupervised evidence retrieval for question answering [J]. Neural computing and applications, 2023, 35 (28): 21247 – 21257.

15 YADAV V, BETHARD S, SURDEANU M. Quick and (not so) dirty: unsupervised selection of justification sentences for multi – hop question answering [C]. Kerrville: The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP – IJCNLP), 2019.

16 YADAV V, BETHARD S, SURDEANU M. Unsupervised alignment – based iterative evidence retrieval for multi – hop question answering [C]. oneline: The 58th Annual Meeting of the Association for Computational Linguistics, 2020.

17 HUANG X, PENG H, ZOU D, et al. CoSENT: consistent sentence embedding via similarity ranking [EB/OL]. [2024 – 03 – 14]. https://ieeexplore.ieee.org/document/10531646.

18 REIMERS N. Sentence – BERT: sentence embeddings using siamese BERT – Networks [EB/OL]. [2024 – 07 – 12]. https://arxiv.org/abs/1908.10084.

19 CUI Y, CHE W, LIU T, et al. Revisiting pre – trained models for Chinese natural language processing [EB/OL]. [2024 – 03 – 14]. https://arxiv.org/abs/2004.13922.