软件学报 ISSN 1000-9825, CODEN RUXUEW
2025,36(12):5572−5598 [doi: 10.13328/j.cnki.jos.007399] [CSTR: 32375.14.jos.007399]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

华谱通: 基于知识推理的家谱问答大语言模型*

吴信东[1,2], 卓兴锐[1,2], 常永泮[1,2], 吴共庆[1,2], 张 赞[1,2], 朱 毅[2,3]

[1](大数据知识工程教育部重点实验室 (合肥工业大学), 安徽 合肥 230009)

[2](合肥工业大学 计算机与信息学院, 安徽 合肥 230601)

[3](扬州大学 信息工程学院, 江苏 扬州 225009)

通信作者: 吴信东, E-mail: xwu@hfut.edu.cn

摘 要: 利用计算机技术实现家谱数据的智能化管理, 对传承和普及中华传统文化有着重要的意义. 近年来, 随着基于检索增强的大语言模型在知识问答领域被广泛应用, 通过大语言模型以对话的方式向用户展示多样的家谱文化已经成为一个备受关注的研究方向. 然而, 家谱数据的异构性、自治性、复杂性和演化性导致现有的知识检索框架难以在复杂的家谱信息中实现完备的知识推理. 针对上述问题, 提出一种基于知识图谱推理的大语言模型家谱问答系统——华谱通, 从推理逻辑完备性和信息筛选精准性两个方面, 构建适合大语言模型家谱问答的知识图谱推理框架. 在推理逻辑完备性方面, 以知识图谱作为家谱知识的载体, 并基于 Jena 框架提出一套完备的家谱知识推理规则, 以提升模型对家谱信息的检索召回率. 在信息筛选方面, 以家谱中的同名人物和多重亲属关系为场景, 提出基于问题-条件三元组的多条件匹配机制和基于大根堆的 Dijkstra 路径排序算法, 通过过滤冗余的检索信息, 达到对大语言模型精准提示的目的. 目前, 华谱通已经部署到公开的智能家谱网站——华谱网, 并通过真实的家谱数据验证了问答系统的有效性.

关键词: 家谱知识问答; 知识图谱推理; 大语言模型; 多条件匹配; 路径排序

中图法分类号: TP182

中文引用格式: 吴信东, 卓兴锐, 常永泮, 吴共庆, 张赞, 朱毅. 华谱通: 基于知识推理的家谱问答大语言模型. 软件学报, 2025, 36(12): 5572–5598. http://www.jos.org.cn/1000-9825/7399.htm

英文引用格式: Wu XD, Zhuo XR, Chang YP, Wu GQ, Zhang Z, Zhu Y. Huaputong: Large Language Model for Genealogical Question-answering with Knowledge Reasoning. Ruan Jian Xue Bao/Journal of Software, 2025, 36(12): 5572–5598 (in Chinese). http://www.jos.org.cn/1000-9825/7399.htm

Huaputong: Large Language Model for Genealogical Question-answering with Knowledge Reasoning

WU Xin-Dong[1,2], ZHUO Xing-Rui[1,2], CHANG Yong-Pan[1,2], WU Gong-Qing[1,2], ZHANG Zan[1,2], ZHU Yi[2,3]

[1](Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, Hefei 230009, China)

[2](School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China)

[3](School of Information Engineering, Yangzhou University, Yangzhou 225009, China)

Abstract: The use of computer technology for intelligent management of genealogy data plays a significant role in inheriting and popularizing Chinese traditional culture. In recent years, with the widespread application of retrieval-augmented large language model (LLM) in the knowledge question-answering (Q&A) field, presenting diverse genealogy scenarios to users through dialogues with LLMs has become a highly anticipated research direction. However, the heterogeneity, autonomy, complexity, and evolution (HACE) characteristics of genealogy data pose challenges for existing knowledge retrieval frameworks to perform comprehensive knowledge

* 基金项目: 国家自然科学基金 (62120106008); 教育部创新团队发展计划 (IRT17R32); 中央高校基本科研业务费专项资金 (JZ2023HGTB0270)
收稿时间: 2024-06-27; 修改时间: 2024-08-26; 采用时间: 2025-01-23; jos 在线出版时间: 2025-06-25
CNKI 网络首发时间: 2025-06-26

reasoning within complex genealogy information. To address this issue, Huaputong, a genealogy Q&A system based on LLMs with knowledge graph reasoning, is proposed. A knowledge graph reasoning framework, suitable for LLM-based genealogy Q&A, is constructed from two aspects: logic reasoning completeness and information filtering accuracy. In terms of the completeness of logic reasoning, knowledge graphs are used as the medium for genealogy knowledge, and a comprehensive set of genealogy reasoning rules based on the Jena framework is proposed to improve the retrieval recall of genealogy knowledge reasoning. For information filtering, scenarios involving name ambiguity and multiple kinship relations in genealogy are considered. A multi-condition matching mechanism based on problem-condition triples and a Dijkstra path ranking algorithm using a max heap are designed to filter redundant retrieval information, thus ensuring accurate prompting for LLMs. Huaputong has been deployed on the Huapu platform, a publicly available intelligent genealogical website, where its effectiveness has been validated using real-world genealogical data.

Key words: genealogy Q&A; knowledge graph reasoning; large language model (LLM); multi-condition matching; path ranking

一套家谱记载了一个姓氏族群的基本情况、家族起源变迁和家族文化等信息. 家谱与正史、地方志并列为我国历史研究的 3 大基石[1], 是中华文化多样性和延续性的重要依据. 随着互联网的发展, 电子家谱资源不断增多, 为大众了解古往今来的中华家族文化提供了便捷的窗口. 然而, 家谱信息的碎片化往往使得人们难以从中获取所需的信息. 为了应对上述问题, 现有的研究[2,3]以知识图谱 (knowledge graph, KG) 作为家谱信息的管理工具, 并基于大数据知识工程 BigKE[4]框架构建家谱服务系统 (https://www.zhonghuapu.com/), 以便实现智能化的家谱知识处理与展示.

近年来, 随着计算机算力的爆发式增长, 以 OpenAI、Facebook 和 Google 为首的人工智能科技公司先后发布了 GPT[5]、LLaMA[6]和 Gemini[7]系列大语言模型 (large language model, 下文简称大模型或 LLM), 将生成式人工智能的热潮推上了新的高度. 作为在大规模语料库上预训练的大型神经网络, LLM 能够通过 Transformer 架构[8]较为充分地捕获文本内容中的上下文信息, 同时拥有强大的人机对话能力. 因此, 相较于基于 KG 检索的家谱信息查询系统, 面向生成式大模型的家谱知识问答系统在人机交互方面往往更受用户青睐, 具有较高的研究意义和潜在社会价值.

然而, 由于预训练数据集的内容时效性和语料局限性, LLM 在处理特定领域的知识问答时仍显得力不从心. 以图 1 为例, 未在特定的家谱数据上做微调处理的 LLM 通常难以正确回答用户的问题, 一个常见的情况是生成难以理解的错误答案, 或称为“大模型幻觉”[9]. 该缺陷表明在没有额外保障的情况下, 将 LLM 部署在垂直领域是不切实际的[10]. 为此, 现有的研究提出了检索增强生成式 (retrieval-augmented generation, RAG) 框架[11], 通过从外部知识库中检索的相关信息, 实现对 LLM 答案生成的提示与约束. 然而, 在大数据环境下, 家谱数据存在 HACE 特性[2,12], 即异构性 (heterogeneity)、自治性 (autonomy)、复杂性 (complexity) 和演化性 (evolution), 该特性导致现有的 RAG 框架难以实现对家谱知识的全面检索和精准筛选, 这限制了 LLM 在家谱知识问答场景中的表现. 具体而言, 当前的 RAG 框架在家谱知识检索中面临着推理逻辑完备性和信息筛选精准性两方面的挑战.

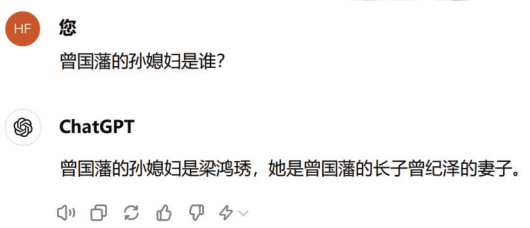


图 1 大模型幻觉案例 (孙媳妇不是儿子的妻子)

● 推理逻辑完备性挑战: 由于语义相似度匹配缺乏逻辑性, 基于文档检索的 RAG 框架[13]会受到各种家谱格式 (水平、垂直, 甚至树形) 和语义噪声的干扰. 此外, 受限于家族图谱中缺失的关系信息 (家族图谱通常只包含父母子女关系, 而不存储叔伯、祖先、后代等关系), 基于图推理机制的 RAG 框架很难通过 Text2Query[14]或节点上下

文获取[15]实现准确的知识推理. 上述问题导致目前的 RAG 框架难以利用完备的推理逻辑对家谱知识进行全面检索, 进而难以充分收集目标信息用于提示 LLM 回答问题.

● 信息筛选精准性挑战: 在家谱知识问答中, 同名人物 (找人/属性) 和多关系路径 (找人物关联) 问题尤为常见, 这使得 RAG 框架必须从众多的候选检索结果中筛选出最优信息. 现有的 RAG 框架大多使用重排序技术[16]筛选候选信息, 但当用户意图不明确或候选信息量过大时, 该方法的排序效率和质量都会被限制, 进而无法精准地筛选出合适的提示信息来启发 LLM 生成答案.

为了应对上述挑战, 本文提出一种基于知识图谱推理的大模型家谱问答系统——华谱通. 具体而言, 为了实现家谱知识推理逻辑的完备性, 本文将知识图谱作为家谱数据的存储载体, 并构建以 Jena 框架为核心的知识图谱推理机. 在此基础上, 通过专业化定制的 Jena 推理规则, 保证系统在家谱知识推理方面的完备性. 相较于在数据库端开发复杂的查询脚本, 在 Jena 推理机中定义的推理规则具有语义清晰、逻辑完备和扩展便利的优势. 此外, 为应对同名人物和多关系路径场景下出现家谱知识推理结果冗余的问题, 本文提出一种基于问题-条件三元组抽取的多条件匹配机制和基于大根堆的 Dijkstra 路径排序法, 以便系统从大量的候选结果中精准筛选有效的提示信息, 协助 LLM 准确地回答问题.

本文的贡献点如下.

1) 提出一种基于 Jena 框架的家谱知识图谱推理机制, 通过专业化定制的推理规则, 完善系统在复杂家谱问答场景中的知识推理逻辑完备性, 从而保障推理结果的完整性.

2) 针对冗余的推理结果, 提出一种基于多条件匹配和路径排序的信息筛选机制, 这有助于系统根据用户意图为 LLM 提供准确的家谱信息, 以此协助 LLM 问答.

3) 开发了“华谱通”系统 (https://www.zhonghuapu.com/index.php/datamgr/huaputong), 并将其成功部署到公开的中华谱网站, 并用现实世界的家谱数据验证了华谱通在家谱问答领域的有效性.

1 相关工作

本节将首先从大语言模型和 RAG 框架的发展现状入手, 介绍当前基于知识检索增强的 LLM 问答技术路线. 之后, 通过分析当前的家谱知识研究现状, 指出现有 RAG 框架在家谱知识检索方面的局限性, 进而说明华谱通的技术先进性.

1.1 大型语言模型发展现状

近年来, 自 Transformer 框架[17]被提出之后, 大量高效的自然语言处理 (natural language processing, NLP) 模型被提出, 其中尤以大型语言模型 (LLM)[18]最令人瞩目. 总体而言, LLM 通常指具有十亿量级参数的大型 NLP 模型, 它们在海量数据上进行预训练, 并被大量研究与实践证明了理解和生成文本的能力, 进而推动了各种自然语言处理研究的进展, 如文本分类[19]、人机协同[20]和信息检索[21]. 一般而言, 现有的 LLM 大多可以根据模型架构被划分为 3 大类: 编码器模型、解码器模型和编码-解码器模型.

以 BERT (bidirectional encoder representations from Transformers)[22]系列为代表的编码器语言模型通过将输入的文本编码到高维空间来处理文本信息. 这类大模型在预训练时通常采用基于掩码学习的自编码操作, 对关键特性进行双向编码处理, 这意味着它们在编码时可以同时考虑每个词缀的上下文信息, 进而使模型更好地理解每个词语在上下文中的含义, 这对于情感分析[23]、实体匹配[24]和完形填空[25]等自然语言理解任务至关重要.

与编码器语言模型相比, 基于解码器的语言模型仅依赖上文的信息实现对下文内容的生成与补全, 即以自回归的方式从左到右地生成文本. 因此, 基于解码器的 LLM 适合于有条件的生成式自然语言任务, 例如聊天对话[26]、自动化文章撰写[27]、图文交互[28]等. 由于具有强大的文本生成能力, 诸如 ChatGPT[29]、InstructGPT[30]和 PaLM[31]等基于解码器的大语言模型如今已经被广泛地应用于各大智能问答系统中, 以便实现更加灵活的人机互动.

基于编码-解码器的 LLM 吸收了前两类方法的优点, 在文本理解和生成上都有较好的表现, 例如 T5 (text-to-text transfer Transformer)[32]系列的编码-解码 LLM 能够将各种 NLP 任务转换为文本生成问题. 具体而言, T5 中的

编码器处理输入文本序列以捕获其意义, 而解码器则基于所编码的信息生成输出序列. T5 架构非常适合语言翻译[33]这类将一个序列转换为另一个序列的任务.

随着近几年国内各大人工智能企业对大模型自主研发热情的提升, 大量国产的优质 LLM 被相继提出. 例如, 由清华大学所提出的 GLM (general language model)[34]大模型利用自回归技术实现大模型的掩码预训练, 使其拥有强大的上下文理解能力和文本生成能力, 其商用版本 ChatGLM 已经成为一款备受关注的智能问答系统. 此外, 诸如阿里巴巴发布的 Qwen[35]、华为提出的盘古大模型[36]和科大讯飞提出的星火大模型 (https://xinghuo.xfyun.cn/)等, 都为大模型前沿技术在众多领域的落地提供了实际案例.

然而, 尽管当前国内外的 LLM 在日常人机问答中表现出了优异的性能, 但考虑到预训练语料库的局限性, 当前的 LLM 依旧难以应对特殊领域的问答场景. 因此, 将 LLM 实施任务领域的迁移成为一个极具挑战性的研究方向.

1.2 基于 RAG 的大模型领域迁移研究

为实现大模型在特定领域的落地, 早期的研究基于预训练-微调技术[37], 将特定领域的语料以文本或结构化知识[38]的形式注入 LLM 的预训练参数中. 尽管这种方法能够让 LLM 在特定的下游任务中保持良好的问题求解能力, 但高昂的算力和数据标注成本却限制了 LLM 在特定领域的部署效率. 为了实现高效的 LLM 领域迁移, 研究者们引入 RAG 框架, 通过问答的方式将外部领域知识库中的信息实时传输给 LLM, 以协助其完成开放域的下游任务.

检索过程是 RAG 框架的重中之重, 其决定着导入 LLM 的提示信息是否精准有效. 因此, 大量研究对 RAG 的检索过程提出了有建设性的改进方案. 例如, 在检索类型上, 一些稀疏检索方式[39]被直接用于 RAG 框架[40], 以提升系统的检索时效性. 但稀疏检索难以匹配问题和外部知识库中语义相关的内容, 因此, 现有的 RAG 框架大多采用基于语义相似度匹配的密集检索策略, 即利用预训练语言模型首先为问题和知识库做向量化处理, 再通过向量相似度选择与问题相关度高的知识片段. 然而, 相似度检索却容易因为缺乏逻辑性而造成检索不精准的问题.

为实现精准的知识检索, 基于知识图谱 (KG) 的 RAG 框架[41]被引入 LLM 问答领域, 这些方法大多通过 LLM 来生成与问题匹配的知识检索路径[42]或数据库查询语句[43], 进而实现对知识库的结构化检索. 然而, 基于 KG 的 RAG 框架需要基于完备的知识库才能实现精准的检索, 但大多数 KG 存在信息缺失或信息冗余的问题, 且维护成本较大, 这导致基于 KG 的 RAG 框架在复杂知识场景 (如家谱知识问答) 中难以通过检索完整的信息来提升 LLM 所生成答案的准确性.

为了应对因 KG 自身缺陷而导致的 RAG 检索质量下降问题, 本文通过 Jena 推理机的自定义推理规则来增强 RAG 框架在复杂知识中的推理逻辑完备性, 从而实现对不完整 KG 的全面检索, 并在家谱知识问答的场景中验证了所提出方法的有效性.

1.3 基于知识图谱的家谱知识研究进展

家谱反映了家族人物详细的个人信息和盘根错节的亲属关系, Wu 等人[2]认为, 家谱数据是互联网时代数据 HACE 原理[12]的典型案例, 即家谱数据存在异构性、自治性、复杂性和演化性这 4 大特点. 因此, 由家谱数据所构建的知识库是检验数据挖掘算法和知识工程系统有效性的一个经典场景.

现有专门针对家谱知识管理和分析的研究还普遍停留在“从数据到知识”的层面, 即通过数据获取与融合的方式, 将异构自治的家谱信息转化成统一的结构化知识形式, 以便于家谱信息的存储与管理[44]. 然而, 这种信息管理方式并不利于将家谱知识以自然语言的形式传递给一般用户. 此外, 在复杂数据分析方面, 受限于家谱中存在的信息缺失或内容冗余问题, 当前的知识挖掘算法[45,46]往往难以根据已有的家谱知识表示构建完备的推理逻辑, 进而限制了相关方法对家谱知识的精准挖掘.

由于上述问题的存在, 将繁杂的家谱知识以通俗易懂的自然语言形式展现在用户面前是极具挑战的. 因此, 本文拟结合完备的 Jena 推理逻辑框架与 LLM 强大的对话能力, 在实现精准的家谱信息检索的前提下, 构建一个具有人机交互能力的家谱知识问答系统.

2 华谱通系统框架

本节将详细阐述华谱通的系统架构, 包括家谱图谱构建、家谱知识图谱推理、冗余信息筛选以及基于提示学习的多轮问答框架. 此外, 为响应华谱系统“人人互联”的本质特征, 本节根据华谱通现有的技术框架, 对跨家谱知识问答做了技术路线的分析. 华谱通的问答框架如图 2 所示.

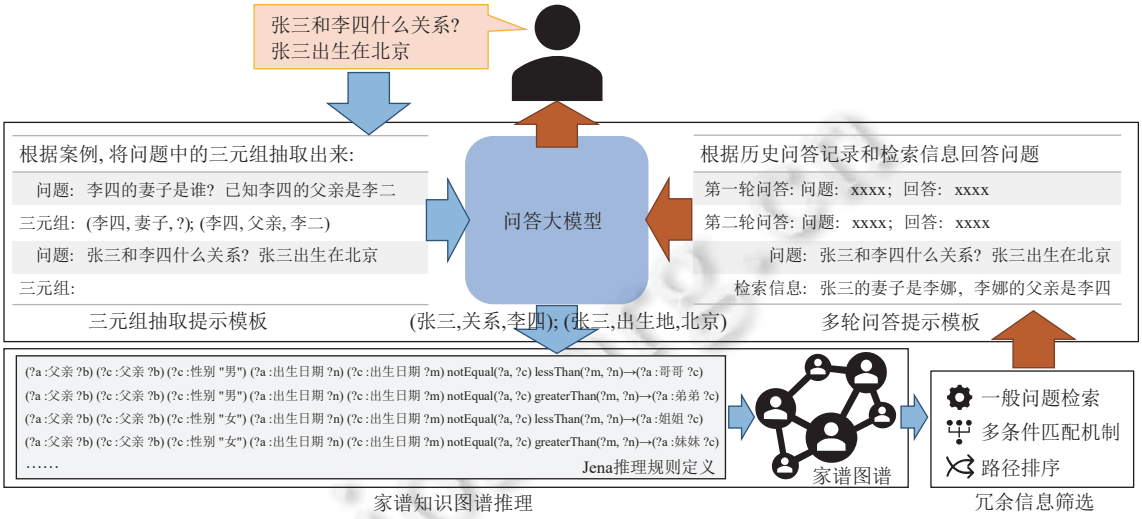


图 2 华谱通问答框架

2.1 家谱图谱构建

华谱通依托华谱平台的知识库来构建问答系统, 该知识库目前共收录 1 900 多份家谱, 录入的人物信息超过 1 800 万条, 并采用 Neo4j 图数据库存储人物的属性和关系信息. 然而, 由于华谱知识库中的数据来源于众多独立自治家谱文档, 其对人物之间的关系种类表述可能存在差异. 为了应对这一问题, 本文首先根据血缘和婚姻类型, 将常见的家族关系划分为如下所示的 3 类.

- 配偶关系: 指婚姻关系中的夫妻, 是其他亲属关系的基础.
- 血缘关系: 指有血缘关系的亲属, 包括直系血亲和旁系血亲. 直系血亲是与血缘直接相关的亲属. 一般来说, 直系血亲包括两类: 一方面, 它包括父母、祖父母、外祖父母和高世代的曾祖父母等. 另一方面, 它涉及子女、孙辈和较低级的曾孙辈. 旁系血亲则是指有血缘关系的非直系血亲, 如兄弟姐妹和叔伯姑侄等.
- 婚姻关系: 指通过婚姻维系的关系, 它通常可以被分为 3 类: (1) 血亲的配偶, 如嫂嫂; (2) 配偶的血亲, 如岳父; (3) 配偶血亲的配偶, 例如妻子兄弟的妻子等. 值得注意的是, 婚姻关系会随着家族成员的嫁娶关系而逐渐复杂.

如后文图 3 (a) 所示, 根据中华族谱的普遍规律, 本文规定“丈夫”“妻子”“父亲”“儿子”和“女儿”关系为华谱通家谱图谱的基础关系, 用于串联整个家谱的人物. 此外, 为支持华谱通的人物生平信息问答功能, 华谱系统对家谱人物分配了多个属性值, 包括生卒年月、字辈号、家庭排行、籍贯和住址等. 依照上述规则, 本文构造的家族图谱局部样例如后文图 3 (b) 所示.

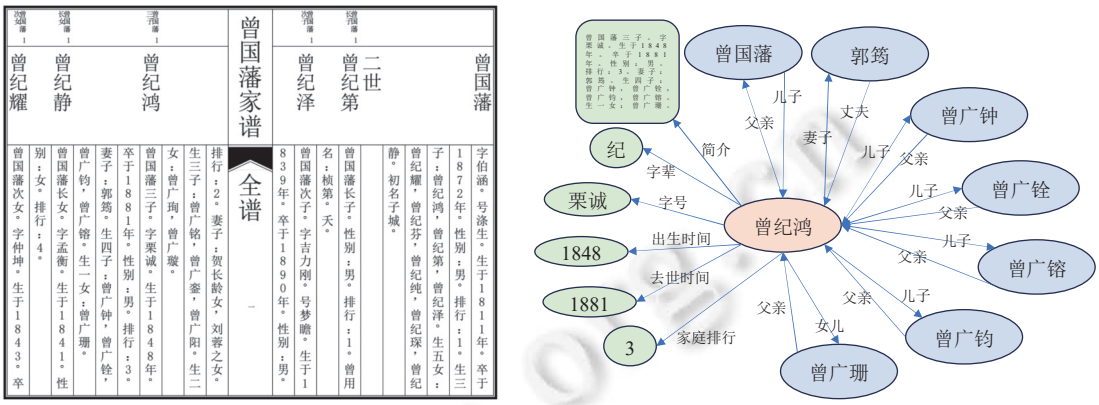
2.2 面向家谱的知识图谱推理

本节主要介绍华谱通中基于知识图谱推理的核心技术框架, 包括基础问题的 SPARQL 查询和基于 Jena 的完备推理逻辑定义. 该推理框架能够保证华谱通从家谱知识图谱中推理出的结果与用户的问题高度匹配, 以便更有效地提示大模型解答用户问题.

2.2.1 SPARQL 查询机制

为了更好地适应基于大模型的自然语言问答过程, 华谱通采用基于语义网络的知识图谱逻辑推理框架. 具体

而言, 华谱通以 SPARQL 语句作为知识图谱的查询载体, 这要求家谱数据的存储必须遵循资源描述框架 (resource description framework, RDF) 协议. 在 RDF 协议条件下, 家谱中的人物、属性和关系用唯一的统一资源标识符 (uniform resource identifier, URI) 确定, 进而通过关系和属性三元组构建符合 RDF 协议的家谱数据. 在如图 4 所示的曾国藩属性和关系案例中, 一个关系三元组和属性三元组被分别定义为<起始人物 URI, 关系 URI, 目标人物 URI>和<人物 URI, 属性 URI, 属性值字段>.



(a) 曾国藩家谱示例 (b) 华谱通人物知识结构

图 3 曾国藩家谱示例和华谱通人物知识结构

```
<zhonghuapu#person/14914777> <zhonghuapu#姓名> "曾国藩" .
<zhonghuapu#person/14914777> <zhonghuapu#出生日期> "1811年" .
<zhonghuapu#person/14914777> <zhonghuapu#去世日期> "1872年" .
<zhonghuapu#person/14914777> <zhonghuapu#性别> "男" .
<zhonghuapu#person/14914777> <zhonghuapu#民族> "汉族" .
<zhonghuapu#person/14914777> <zhonghuapu#字辈> "国" .
<zhonghuapu#person/14914777> <zhonghuapu#简介> "字伯涵。号涤生。生于1811年。卒于
<zhonghuapu#person/14914777> <zhonghuapu#排行> 1.
<zhonghuapu#person/14914777> <zhonghuapu#字> "伯涵" .
<zhonghuapu#person/14914777> <zhonghuapu#号> "涤生" .
<zhonghuapu#person/14914777> <zhonghuapu#出生地> "湖南长沙府湘乡县" .
<zhonghuapu#person/14914777> <zhonghuapu#住址> "湖南省娄底市双峰县荷叶镇富托村" .
<zhonghuapu#person/14914777> <zhonghuapu#祖籍> "江西永丰" .
<zhonghuapu#person/14914777> <zhonghuapu#家谱> "3030993" .
<zhonghuapu#person/14914777> <zhonghuapu#儿子> <zhonghuapu#person/14914900> .
<zhonghuapu#person/14914777> <zhonghuapu#儿子> <zhonghuapu#person/14914901> .
<zhonghuapu#person/14914777> <zhonghuapu#儿子> <zhonghuapu#person/14914902> .
<zhonghuapu#person/14914777> <zhonghuapu#女儿> <zhonghuapu#person/14914988> .
<zhonghuapu#person/14914777> <zhonghuapu#女儿> <zhonghuapu#person/14914989> .
<zhonghuapu#person/14914777> <zhonghuapu#女儿> <zhonghuapu#person/14914990> .
<zhonghuapu#person/14914777> <zhonghuapu#女儿> <zhonghuapu#person/14914991> .
<zhonghuapu#person/14914777> <zhonghuapu#女儿> <zhonghuapu#person/14914992> .
```

图 4 RDF 格式下的人物属性与关系案例

在此基础上, 本文定义了对人物基础关系和属性的 SPARQL 查询指令, 表 1 为一些查询指令的案例. 由于 SPARQL 查询语句以语义三元组“?头实体 : 关系 ?尾实体”或“?头实体 : 属性 ?尾实体”为查询条件, 且支持中文的关系和属性描述, 因此, 华谱通可以将用户的查询问题通过三元组抽取的方式快速地转换为条件完整的查询语句 (具体过程详见第 2.3.1 节), 这相对于一般基于大模型的查询指令生成方法在时效性和完整性上有着显著优势.

表 1 华谱通基础 SPARQL 语句查询示例

查询问题	SPARQL核心语句
曾国藩的父亲是谁?	SELECT ?name where{ ?n : 父亲 ?m. ?n : 姓名 曾国藩. ?m : 姓名 ?name.}
曾国藩的儿子是谁?	SELECT ?father where{?n : 儿子 ?m. ?n : 姓名 曾国藩. ?m : 姓名 ?name.}
曾国藩的祖籍在哪?	SELECT ?m where{?n : 祖籍 ?m. ?n : 姓名 曾国藩.}

2.2.2 基于 Jena 框架的完备推理逻辑定义

正如第 2.1 节所述, 一般的家谱图谱在人物关系上只会定义 “父亲”“儿子”“女儿”“丈夫”和“妻子”这 5 种关系, 但这些关系对知识图谱逻辑推理而言并不具备完备性. 因此, 在按照第 2.2.1 节构建的符合 RDF 协议的家谱图谱上, SPARQL 语句难以处理用户提出的复杂关系查询问题. 例如, 当用户询问“曾国藩的后代是谁?”时, 华谱通会抽取关系三元组“?n : 后代 ?m.”, 然而, 由于家谱数据中并不包含“后代”关系, 导致华谱通无法推理出正确的答案. 此外, 诸如“后代”和“祖先”这种需要递归处理的关系, 在 SPARQL 查询端是难以实现的, 这大大增加了华谱通中关系完备性推理功能的开发难度.

为了应对上述挑战, 华谱通采用 Jena 框架完善 RDF 家谱数据中的关系缺失问题. Jena 框架以 Fuseki 服务器管理 RDF 数据, 并支持开发者通过自定义推理规则来完善系统的推理逻辑. 相较于在 SPARQL 端开发复杂的查询语句, Jena 端构建的推理规则在语义清晰度、逻辑完备性和系统扩展性上都有显著的优势. 因此, 根据家谱数据中定义的 5 种关系, 华谱通在 Jena 框架上定义了 26 种关系的推理规则, 以支持完备的谱内亲属关系推理. 在系统实际运行过程中, Jena 框架能够编译开发者在脚本中定义的推理规则 (规则格式如表 2 所示), 编译结果会映射到家谱数据的 RDF 格式文件内, 并存储至计算机磁盘中, 用于后续的推理操作. 为便于读者直观地理解 Jena 推理规则的部署过程, 本文提供了华谱通部署 Jena 规则框架的核心代码与操作, 详见 https://github.com/lazyloafer/Huaputong/tree/main/Jena.

表 2 华谱通中定义的 26 种关系推理规则

关系	Jena推理规则	关系	Jena推理规则
父亲	(?a : 父亲 ?b)→(?a : 父亲 ?b)	姐妹	(?a : 姐姐 ?b)→(?a : 姐妹 ?b) (?a : 妹妹 ?b)→(?a : 姐妹 ?b)
儿子	(?a : 儿子 ?b)→(?a : 儿子 ?b)	兄弟姐妹	(?a : 兄弟 ?b)→(?a : 兄弟姐妹 ?b) (?a : 姐妹 ?b)→(?a : 兄弟姐妹 ?b)
女儿	(?a : 女儿 ?b)→(?a : 女儿 ?b)	姑姑	(?a : 父亲 ?b) (?b : 姐妹 ?c)→(?a : 姑姑 ?c)
丈夫	(?a : 丈夫 ?b)→(?a : 丈夫 ?b)	叔叔	(?a : 父亲 ?b) (?b : 弟弟 ?c)→(?a : 叔叔 ?c)
妻子	(?a : 妻子 ?b)→(?a : 妻子 ?b)	伯伯	(?a : 父亲 ?b) (?b : 哥哥 ?c)→(?a : 伯伯 ?c)
母亲	(?a : 父亲 ?b) (?b : 妻子 ?c)→(?a : 母亲 ?c)	侄子	(?a : 兄弟 ?b) (?b : 儿子 ?c)→(?a : 侄子 ?c)
父母	(?a : 父亲 ?b)→(?a : 父母 ?b) (?a : 母亲 ?b)→(?a : 父母 ?b)	侄女	(?a : 兄弟 ?b) (?b : 女儿 ?c)→(?a : 侄女 ?c)
子女	(?a : 儿子 ?b)→(?a : 子女 ?b) (?a : 女儿 ?b)→(?a : 子女 ?b)	爷爷	(?a : 父亲 ?b) (?b : 父亲 ?c)→(?a : 爷爷 ?c)
哥哥	(?a : 父亲 ?b) (?c : 父亲 ?b) (?c : 性别 "男") (?a : 出生日期 ?n) (?c : 出生日期 ?m) notEqual(?a, ?c) lessThan(?m, ?n)→(?a : 哥哥 ?c)	奶奶	(?a : 父亲 ?b) (?b : 母亲 ?c)→(?a : 爷爷 ?c)
弟弟	(?a : 父亲 ?b) (?c : 父亲 ?b) (?c : 性别 "男") (?a : 出生日期 ?n) (?c : 出生日期 ?m) notEqual(?a, ?c) greaterThan(?m, ?n)→(?a : 弟弟 ?c)	孙子	(?a : 儿子 ?b) (?b : 儿子 ?c)→(?a : 孙子 ?c)
兄弟	(?a : 哥哥 ?b)→(?a : 兄弟 ?b) (?a : 弟弟 ?b)→(?a : 兄弟 ?b)	孙女	(?a : 儿子 ?b) (?b : 女儿 ?c)→(?a : 孙女 ?c)
姐姐	(?a : 父亲 ?b) (?c : 父亲 ?b) (?c : 性别 "女") (?a : 出生日期 ?n) (?c : 出生日期 ?m) notEqual(?a, ?c) lessThan(?m, ?n)→(?a : 姐姐 ?c)	祖先	(?a : 父母 ?b)→(?a : 祖先 ?b) (?a : 祖先 ?b) (?b : 父母 ?c)→(?a : 祖先 ?c)
妹妹	(?a : 父亲 ?b) (?c : 父亲 ?b) (?c : 性别 "女") (?a : 出生日期 ?n) (?c : 出生日期 ?m) notEqual(?a, ?c) greaterThan(?m, ?n)→(?a : 妹妹 ?c)	后代	(?a : 子女 ?b)→(?a : 后代 ?b) (?a : 后代 ?b) (?b : 子女 ?c)→(?a : 后代 ?c)

表 2 定义的 26 种关系可以分为递归 (如祖先和后代) 与 Ad-hoc (如父亲~孙女) 两类关系. 根据 Jena 自定义规则的右赋值语法, 被定义的规则需要位于“→”符号的右侧, 其左侧为可推导出被定义规则的条件, 这些条件必须是数据集中已包含的关系/属性或 Jena 中已定义的规则. 例如, “父亲”“儿子”“女儿”“丈夫”和“妻子”这 5 种关系的推理规则可以直接映射为数据集中已包含的关系, 而其余的关系则需以这 5 种推理规则为基础, 结合人物自身的属性构造相应推理规则. 根据表 2 中定义的 26 种关系推理规则, 华谱通可以借助 SPARQL 查询实现完备的家谱内

人物关系推理.

2.3 冗余信息筛选机制

针对家谱数据中可能存在的同名人物和多关系路径场景, 本节讨论对第 2.2 节知识图谱推理框架的进一步优化方案, 包括多条件匹配机制和路径排序算法.

2.3.1 多条件匹配机制

多条件匹配机制主要针对家谱中的同名人物检索问题. 假设一份家谱里存在多个名叫“张三”的人物, 且用户仅简单询问其父亲是谁时, 华谱通在不依靠多条件匹配机制的情况下只能抽取出三元组 (张三, 父亲, ?), 并将其转换为 SPARQL 语句进行逻辑查询. 此时, 满足条件的查询结果会因为同名人物现象而增多, 且每个查询结果之间存在冲突 (比如一个人通常情况下不会出现两个父亲), 这种冲突的检索结果会误导大模型生成错误的结论.

为应对这一问题, 华谱通允许用户根据自己对同名人物的理解, 在问题中附加同名人物的具体信息, 以便实施更精确的家谱信息检索. 具体而言, 本文为华谱通设计了一种基于大模型提示学习的问题-条件三元组抽取方法, 以便系统在根据问题三元组抽取出所有同名人物的信息后, 能利用条件三元组匹配出最符合用户意图的一个同名人物, 用于辅助大模型解答用户的问题. 以下是所提出问题-条件三元组抽取方法的详细介绍.

首先, 华谱通利用大模型强大的自然语言理解和生成能力, 构造如图 5 所示的提示模板来指导大模型对用户的提问内容进行三元组抽取. 在提示模板中, 被抽取的三元组被划分为问题三元组和条件三元组, 以下是它们的定义.

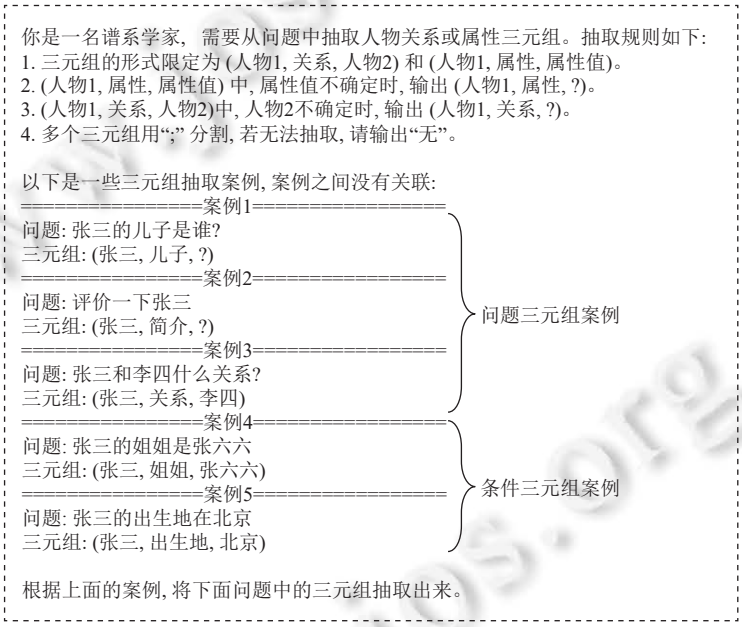


图 5 基于提示学习的问题-条件三元组抽取

● 问题三元组: 由用户的问题映射而来, 其中三元组的头实体必须是被询问的目标人物姓名, 关系和属性名称根据问题具体确定; 当用户的目的是找人或找人物属性时, 问题三元组的尾实体用“?”代替, 当用户的目的是找任意两人关系时, 尾实体则是另一个与头实体不同的人物姓名.

● 条件三元组: 由用户提供的陈述断言转化而来, 可以表示确定的人物关系和人物属性, 用于辅助系统辨析同名人物.

在实现对用户提问内容的问题-条件三元组抽取后, 华谱通会根据抽取的三元组进行同名人物筛选, 具体流程如算法 1 所示.

算法 1. 华谱通同名人物筛选算法.

输入: 问题三元组 (s*, r*, t*), 其中 s*是同名人物姓名; 条件三元组集合 P={(s_1, r_1, t_1), (s_2, r_2, t_2),…,(s_n, r_n, t_n)}; 关系类型集合 R; 属性类型集合 A;
输出: 最符合条件的同名人物 *max_URI*.

1. new List *same_people_uri_list*=*SPARQL*((*s**, 简介, ?)); /*保存所有候选同名人物 *URI**/
2. new Dict *same_people_dict*; /*用于保存同名人物的信息, 包括属性和人物关系*/
3. for *URI* in *same_people_uri_list*
4. for *r* in $R \cup A$
5. *same_people_dict*[*URI*]=*SPARQL*((*URI*, *r*, ?)); /*获取每个同名人物的所有信息*/
6. end for /*对应步骤 4*/
7. end for /*对应步骤 3*/
8. new List *same_people_match_score*; /*用于保存同名人物信息与条件信息的匹配度*/
9. for *URI* in *same_people_uri_list*
10. new List *candidate_information_list*;
11. for (s_i, r_i, t_i) in *P*
12. if *equal*(*URI*, s_1.*uri*)
13. *candidate_information_list.append*(t_i); /*获取每个同名人物的查询条件*/
14. end if /*对应步骤 12*/
15. end for /*对应步骤 11*/
16. *score*=*TextMatch*(*same_people_dict*[*URI*], *candidate_information_list*); /*匹配用户提供的条件和同名人物的知识库信息*/
17. *same_people_match_score.append*((*URI*, *score*)); /*存储所有同名人物与查询条件的匹配度*/
18. end for /*对应步骤 9*/
19. *max_URI* =*max*(*same_people_match_score*); /*选择最大 *score* 对应的 *URI* */
20. return *max_URI*

在算法 1 中, 步骤 1 的 *SPARQL*() 函数表示根据输入的三元组构造 SPARQL 查询语句并返回查询结果, 查询语句的构造过程参考表 1. 此外, 步骤 16 的 *TextMatch*(*A*, *B*) 函数被定义为公式 (1):

$$TextMatch(A,B)=\frac{|jieba(A)\cap jieba(B)|}{|jieba(B)|} \tag{1}$$

其中, A 是华谱通查询到的同名人物信息, B 是用户提供的同名人物查询条件. 为了便于进行词频匹配, 本文使用 jieba 分词软件对 A 和 B 中的文本进行分词, 并将 A 和 B 中共同存在的语段占 B 中语段的比重作为用户需要查询某个同名人物的概率, 最后选择概率最高的同名人物 *URI* 作为查询目标. 若无法匹配出最符合条件的同名人物, 华谱通会从候选的同名人物中随机选择一个进行查询和大模型问答; 之后, 在大模型生成答案的末端增加对其余同名人物的简介 (参考第 3.3.1 节中内容), 以方便用户做进一步选择. 为帮助读者更直观地了解算法中功能函数 (*SPARQL*() 和 *TextMatch*(*A*, *B*)) 的具体设计细节, 本文将华谱通同名人物多条件匹配机制的核心脚本上传至 GitHub 代码库, 详见 https://github.com/lazyloafer/Huaputong/tree/main/qa_demo.py.

2.3.2 关系路径排序

当进行人物关系查询时, SPARQL 可能会检索到多条关系路径来提示大模型生成答案. 然而, 大多数关系路径所表达的人物关系都不够直观, 且大模型对复杂关系的总结能力偏弱. 因此, 大量冗余的路径可能会干扰大模型对人物关系的总结, 图 6 展示了 ChatGPT 在多关系路径下错误总结人物关系的案例.

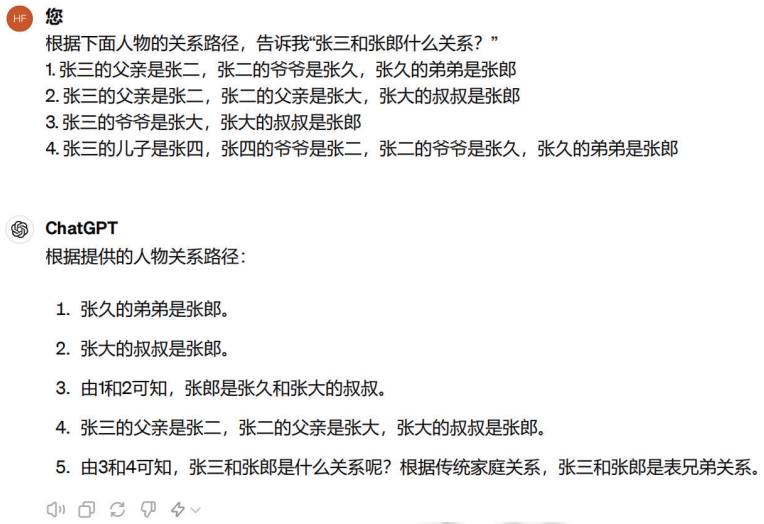


图 6　ChatGPT 对多条路径的错误归纳案例

为了应对大模型难以有效理解关系路径的问题, 本文对人物关系路径的检索过程做了进一步优化, 包括基于路径衰减的关系打分机制和基于大根堆的 Dijkstra 路径排序算法, 旨在检索到最能代表人物关系的路径来启发大模型回答用户问题.

关系分数是路径打分和排序的基础. 华谱通根据家谱中以“父子妻女”关系为核心的特性, 对第 2.2 节中定义的 26 种关系设计了层次打分规则. 如图 7 所示, 华谱通以“父亲”“儿子”“母亲”“女儿”“丈夫”和“妻子”为一个家庭的中心关系, 并向外扩展到兄弟姐妹、叔伯姑侄以及爷孙三组关系 (图 7 中蓝色箭头关联的表格); 此外, 华谱通还针对一些模糊的关系进行了打分 (图 7 中红色箭头关联的表格). 因为模糊关系检索的结果不及其他关系精准, 华谱通定义的模糊关系分数远低于其他关系.

图 7　层次关系打分

除了关系分数之外, 路径的长度是路径打分的另一个重要参考. 一般来说, 两个人之间的关系路径越长, 代表两人的关系越薄弱. 因此本文规定一个关系在路径的分数随着跳数的增加而减少. 具体的分数衰减原则如下.

1) 对任意关系, 其分数随其所在路径位置跳数的增加而递减. 以图 7 中的“父亲”关系为例, 当处于路径的一跳位置时, 其分数为 10 分. 二跳位置的“父亲”关系分数要小于 10 分, 三跳位置的“父亲”关系分数则要小于二跳位置的“父亲”关系分数, 依次类推.

2) 随着跳数的增大, 关系分数的衰减率趋近于 0, 即当路径无限长时, 关系对路径分数的影响趋近于 0.

3) 关系的衰减过程要相对平滑.

根据上述的原则, 本文用 $y(i)=e^{-i}$ 表示路径的衰减率, 其中 i 是路径的跳数. 在此基础上, 对一条长度为 N 的路径 *Path*, 本文用公式 (2) 计算其分数:

$$score(Path)=\frac{1}{N}\sum_{n=1}^{N}y(n-1)s_n \tag{2}$$

其中, s_n 表示路径中第 n 个关系在图 7 中对应的分数. 公式 (2) 综合考虑了路径长度和关系分数的层次性, 能够较为全面地评估一条路径的重要性.

以图 6 中的 4 条路径为例, 按照公式 (2) 的打分函数和图 7 中各个关系的基础分数, 4 条路径的分数可以按照如下步骤计算.

路径 1 (父亲→爷爷→弟弟): $\frac{1}{3}(9\times e^{-(1-1)}+8\times e^{-(2-1)}+7\times e^{-(3-1)})\approx 4.297$ 分.

路径 2 (父亲→父亲→叔叔): $\frac{1}{3}(9\times e^{-(1-1)}+9\times e^{-(2-1)}+5\times e^{-(3-1)})\approx 4.329$ 分.

路径 3 (爷爷→叔叔): $\frac{1}{2}(8\times e^{-(1-1)}+5\times e^{-(2-1)})\approx 4.920$ 分.

路径 4 (儿子→爷爷→爷爷→弟弟): $\frac{1}{4}(9\times e^{-(1-1)}+8\times e^{-(2-1)}+8\times e^{-(3-1)}+7\times e^{-(4-1)})\approx 3.344$ 分.

4 条路径分数由高到低分别是 4.920 分 (路径 3)>4.329 分 (路径 2)>4.297 分 (路径 1)>3.344 分 (路径 4), 即系统认为“张三的爷爷是张大, 张大的叔叔是张郎”能够最精炼地向大模型表述张三和张郎之间的关系, 这也符合大众对关系亲疏性的选择.

在所提出的路径打分方法的基础上, 本文设计了一种基于大根堆的 Dijkstra 路径排序算法, 以实现在动态变化的关系路径分数列表中快速查找到能连接目标人物的最优关系路径. 该排序算法的核心步骤为: 若当前长度为 N 的最高分路径无法连通两个人物, 就将该路径与表 2 中的所有关系组合后扩展为 26 条长度为 N+1 的路径, 与剩余长度为 N 的路径一起参与下一轮排序. 为了在一个动态更新的关系路径分数列表中高效地获取最高分路径, 本文采用大根堆对该分数表进行维护. 第 3.4.1 节具体分析了该算法的时间复杂度和有效性实验. 算法 2 为所提出路径排序算法的流程.

算法 2. 华谱通路径排序算法.

输入: 待查询的头尾人物姓名 *s**和 *t**, 所有关系集合 *R*, 对应的关系分数集合 *H*, 最大路径长度 *max_len*;
输出: 最优路径 *top_path*.

```
new Int hop=1;
new List path_list=R; /*初始化路径集合*/
new List score_list=H; /*初始化路径分数集合*/
new Heap path_heap = Heap(path_list, score_list); /*用一跳路径初始化大根堆*/
new List top_path=[None];
while True
    if equal(hop I, 1) and not path_heap.empty() /*先遍历完所有一跳路径*/
        pass;
    else
        if equal(hop I, 1) /*构建所有可能的二跳路径并入堆*/
            path_list = Group_relation(R, R); /*构建所有可能的二跳路径*/
            score_list = Calculate_score(H, H); /*计算 path_list 中所有二跳路径的分数*/
            path_heap.push(path_list, score_list); /*二跳路径入堆*/
```

14. *hop* ++;
15. end if /*对应步骤 10*/
16. if *top_path.len*()<*max_len*
17. *path_list* = *Group_relation*(*top_path*, *R*); /*扩展堆顶路径*/
18. *score_list* = *Calculate_score*(*top_ score*, *H*);
19. *path_heap.push*(*path_list*, *score_list*); /*将扩展的路径入堆*/
20. end if /*对应步骤 16*/
21. end if /*对应步骤 7*/
22. *top_path*, *top_score*=*path_heap.pop*(); /*堆顶路径出堆*/
23. *result* = *SPARQL*((*s**, *top_path*, *t**)); /*查询 *s**和 *t**之间是否存在路径 *top_path* */
24. if (*result* != None) or (*path_heap.empty*() and *top_path.len*() == *max_len*) /*找到最优路径或在最大路径长度内找不到最优路径*/
25. break; /*对应步骤 6*/
26. end if /*对应步骤 24*/
27. return *top_path*

因为一跳路径表示的关系最为重要, 因此在实际运行时, 华谱通会首先遍历完所有一跳路径, 再从二跳路径开始实施路径排序算法. 在算法 2 中, *Heap*()、*empty*()、*push*() 和 *pop*() 分别表示构建大根堆、判断堆是否为空、新路径入堆和堆顶路径出堆; *Group_relation*(*A*, *B*) 表示组合 *A* 和 *B* 中的关系路径, *Calculate_score*(*A*, *B*) 表示用公式 (2) 计算组合后路径的分数; *SPARQL*() 函数给功能与算法 1 中的类似, 用于构建对应的 SPARQL 语句并进行逻辑查询. 为便于读者获取华谱通路径排序算法中对不同跳数路径的处理细节, 本文将算法的核心脚本上传至 GitHub 代码库, 详见 https://github.com/lazyloafer/Huaputong/tree/ main/qa_demo.py.

2.4 基于提示学习的多轮问答框架

一般情况下, 用户在与大模型问答界面进行交互时, 很容易使用较为口语化的提问方式. 例如, 在询问完“曾国藩的儿子是谁?”后, 用户很可能会想知道曾国藩的女儿, 但不一定会提问“曾国藩的女儿是谁?”, 而是接着上一轮的问答历史直接询问“他女儿呢?”. 尽管第 2 个问题中的“他”, 在用户看来一定指代的是“曾国藩”, 但对问答系统而言, “他”在文本上与“曾国藩”并不匹配. 因此, 需要利用合适的提示信息, 让大模型能够利用其强大的上下文理解能力, 从历史问答记录中定位这个“他”所指的具体人物.

根据上述需求, 华谱通在知识推理逻辑和信息筛选的基础上, 构建了一套基于提示学习的多轮问答机制, 利用大模型强大的上下文理解能力, 从精准的知识库信息中挖掘有利于求解用户问题的依据.

如图 8 所示, 与第 2.3.1 节中用于问题信息抽取的提示模板不同, 本节所设计的提示模板包含大模型角色赋予、历史问答记录、用户当前问题和知识推理依据这 4 个部分. 在实际问答过程中, 该提示模板首先会填入具体的家谱简介信息、用户当前问题和图谱检索信息, 同时动态更新历史问答记录; 之后, 华谱通将更新好的提示模板作为大模型的输入信息, 大模型会根据提示模板的内容来回答用户的问题. 公式 (3) 展示了华谱通多轮问答的具体过程.

$$\begin{cases} ans_i = LLM(prompt(role, ques_i, retrieval_i, hist_{i-1})) \\ hist_i = \{(ques_k, ans_k)\}_{k=1}^{i} \end{cases} \tag{3}$$

在公式 (3) 中, $prompt(role, ques_i, retrieval_i, hist_{i-1})$ 表示构建如图 8 所示的提示模板, *role* 表示提示模板中的角色赋予信息; $ques_i$、ans_i 和 $retrieval_i$ 分别表示第 i 轮问答时用户提出的问题、大模型给出的答案和根据该问题检索到的家谱信息; $hist_{i-1}$ 存储前 i–1 轮的问答对, 作为历史信息导入大模型, 以增强问答的上下文情景, 这有助于启发大模型回答用户问题.

```
华谱系统致力于为华夏写史，助百姓寻根。期待和大家共同打造全球最大最全的跨姓华人家谱系统——中华之谱!
华谱通是华谱系统的特色应用之一，能够结合家谱中的信息来解答用户关于找人、找关系的问题。此外，华谱通也支持日常的聊天功能。
当前家谱信息如下:
【导入具体家谱简介信息】
你现在是一个华谱通小助手，请根据提供的附加信息，回答用户的问题。
==========第1轮问答==========
问题: xxxxxx
回答: xxxxxx
==========第2轮问答==========
问题: xxxxxx
回答: xxxxxx
…
==========当前用户问题==========
问题: 【导入用户当前问题】
附加信息: 【导入家谱图谱检索信息】
回答:
```

图 8 多轮问答提示模板

2.5 跨家谱问答技术路线分析

在华谱系统中, 除了家谱内的亲属关系, 还存在跨谱关系. 这些关系可以实现家谱之间的连接, 从而构建更加复杂的跨家谱人物关系网. 本节主要针对婚姻关系和社会关系, 从逻辑推理和关系生成两个角度对华谱通跨家谱问答的技术路线进行总体分析.

● 逻辑推理式跨家谱问答: 逻辑推理式跨家谱问答适用于跨谱的婚姻关系. 如第 2.1 节所述, 两个家族的成员会通过夫妻双方实现关系互联. 具体而言, 婚姻关系可以根据第 2.2 节提供的 26 种关系推理规则进行扩展. 例如, “岳父”关系和“亲家”关系的 Jena 推理规则可以被分别定义为“(?a : 妻子 ?b) (?b : 父亲 ?c)→(?a : 岳父 ?c)”和“(?a : 子女 ?b) (?b : 夫妻 ?c) (?c : 父母 ?d)→(?a : 亲家 ?d)”. 在此基础上, 基于婚姻关系的跨家谱知识推理结果可以被直接用于后续的信息筛选 (第 2.3 节) 和多轮问答 (第 2.4 节). 因此, 当前的华谱通技术框架可以快速地实现基于婚姻关系的跨家谱问答功能 (具体案例参考第 3.3.3 节).

● 关系生成式跨家谱问答: 这种问答方式主要针对人物之间的社会关系. 社会关系需要考虑家谱人物的生平经历和所处的时代背景. 因此, 相较于亲属关系有明确的规则定义, 人物之间的社会关系则显得更加复杂多样. 例如, 在镇压太平天国运动的背景下, 曾国藩和李鸿章存在“战友”“师生”等社会关系; 而在晚清朝堂上, 二人的社会关系则变为“同僚”或“上下级”. 此外, 当被查询的二人不属于同一个时代时, 对应的社会关系则表现得更加微弱. 例如, 在查询李四 (1998 年生于西安市) 与李鸿章的关系时, 可以认为二人是“同姓”关系, 也可以尝试检索出与李鸿章生活在同一时代的李四的某个祖先. 因此, 参考婚姻关系的 Jena 规则定义法是难以涵盖所有社会关系的. 在这种情况下, 目前一个可行的思路是对待查询二人的个人和亲属信息进行粗粒度地检索, 并利用大模型的自然语言理解能力和知识涌现能力, 从候选的检索结果中选择性地推理出与查询语句相关的社会关系 (具体案例参考第 3.3.3 节).

在实际问答过程中, 由于无法预判被查询人物之间的关系类型, 华谱通采取多线程的方式并行执行上述两个跨家谱问答模块. 逻辑推理式跨家谱问答能够提供精准的人物关系, 但需要对多种关系的多跳组合进行耗时的排序和遍历. 相比之下, 尽管关系生成式跨家谱问答模块中的大模型所推理出的人物关系存在模糊性, 但响应速度快. 因此, 在规定时间内请求不到逻辑推理式跨家谱问答模块的答案时 (可以认为被查询的二人之间关系薄弱或没有实际关联), 华谱通会调用关系生成式跨家谱问答模块中的答案, 以实现及时的人机交互.

2.6 系统异步请求响应机制

在实际人物关联问答时, 用户可能会提供关系薄弱 (如第 2.5 节中所介绍的关系生成式跨家谱问答场景) 或不存在于选定家谱中的待查询人物, 这会大大增加系统在路径选择上的时间开销, 进而导致系统响应超时的问题. 因

此, 在不改动第 2.3.2 节提出的路径排序算法的前提下, 华谱通部署了异步请求响应机制, 并协同百度百科检索功能, 以实现在规定时间内响应用户问题的需求. 如图 9 所示, 本节按照家谱推理和百度百科检索两个模块来详细介绍所提出的异步请求响应机制.

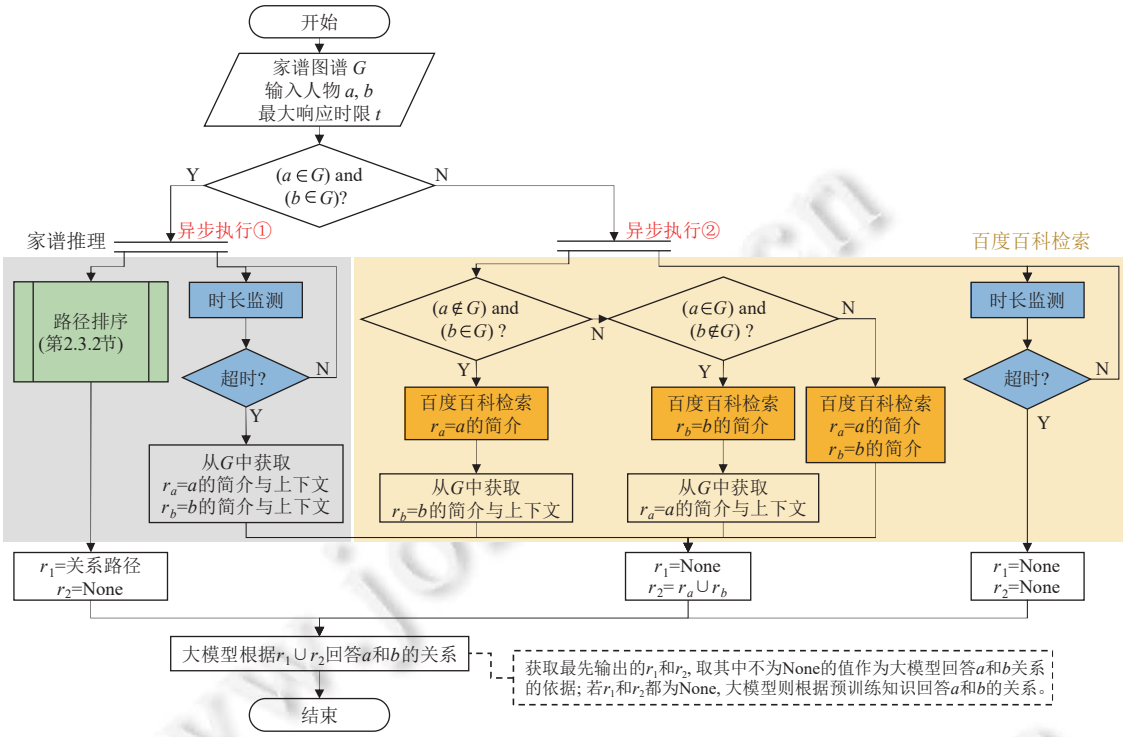


图 9 华谱通异步请求响应机制

● 家谱推理模块: 触发该模块的前提条件是待查寻关系的两个人物都存在于选定的家谱中. 此时系统后端会首先创建一个线程 (记为线程 1) 执行第 2.3.2 节中提出的路径排序算法, 用于为大模型提供人物之间精确的关系路径. 与此同时, 系统前端采用轮询机制, 每隔一段时间 (实际部署时间隔时间设为 3 s) 向后端发送请求信号, 以获取路径排序的最优结果. 当请求次数超过阈值时, 系统认为路径排序响应超时, 即认为待查寻的两个人物之间关系薄弱或没有实际关系. 此时, 系统后端会创建一个新的线程 (记为线程 2), 从指定家谱中直接获取人物的简介与上下文信息 (人物的 26 种亲属), 作为大模型生成人物关系的依据 (具体过程参考第 2.5 节 “关系生成式跨家谱问答” 部分). 需要注意的是, 上述的线程 1 和线程 2 之间的关系为异步执行 (图 9 中的“异步执行①”), 即系统只会将执行速度更快的线程所输出的内容作为大模型的输入信息, 同时忽略另一条线程滞后的输出结果.

● 百度百科检索模块: 当检测到不存在于选定家谱中的人物时, 系统会自动跳转到百度百科检索模块. 该模块借助百度百科查询端口获取非家谱人物的简介信息, 并将其与家谱人物信息 (如果存在) 融合, 以作为大模型的提示内容. 此外, 考虑到百度百科查询同样存在同名人物识别的问题 (例如, 目前百度百科中存在两份“十二金钗”的词条), 华谱通在首次查询到某个百度百科同名人物时, 会将所有同名人物的 URL 和简介存储到数据库表中, 以便复用第 2.3.1 节所提出的多条件匹配机制进行同名人物消歧. 然而, 在实际情况中, 一个同名的百度百科人物可能存在极多词条, 需要较长时间才能存储完成. 因此, 在百度百科检索模块, 华谱通也部署了一个类似上述家谱推理模块中的多线程异步执行功能 (图 9 中的“异步执行②”). 当系统检测到百度百科模块查询超时后, 系统会直接返回空值给后续的大模型问答模块, 即允许大模型根据自身的预训练知识来生成人物之间的关系. 此外, 由于前期负责百度百科检索与数据库表存储的线程并未被销毁, 系统后台会继续将剩余的同名人物信息全部存储到数据库表

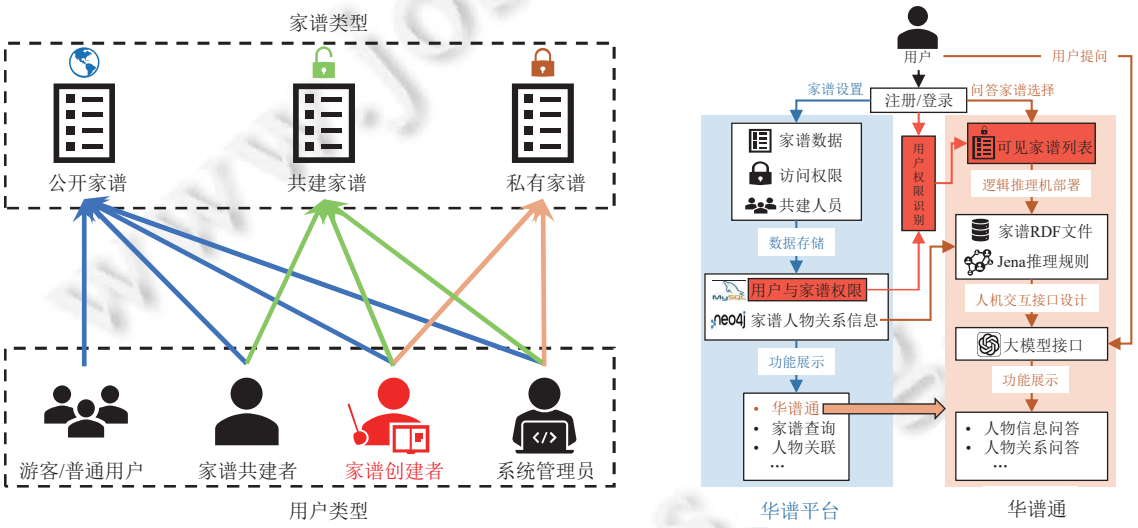
中, 以便后续的问答使用.

根据上述的两个模块, 华谱通可以充分利用异步请求响应机制, 在规定时间内为用户返回一个答案, 避免用户因系统响应过慢而导致不理想的使用体验.

2.7 华谱通数据隐私保护

华谱通问答系统以华谱平台为基座进行功能开发, 因此, 华谱通遵循华谱平台对家谱数据的访问权限管理框架[3], 以实现问答过程中的家谱数据隐私保护. 此外, 由于华谱通在问答过程中只会获取华谱平台上用户的注册 ID 编号, 且相关的大模型提示模板中没有使用用户的详细信息 (例如姓名、性别和联系方式等), 这可以保证在问答过程中不会泄露系统中的用户画像. 因此, 下文主要针对家谱数据的访问权限, 分析华谱通数据隐私保护机制的设计细节.

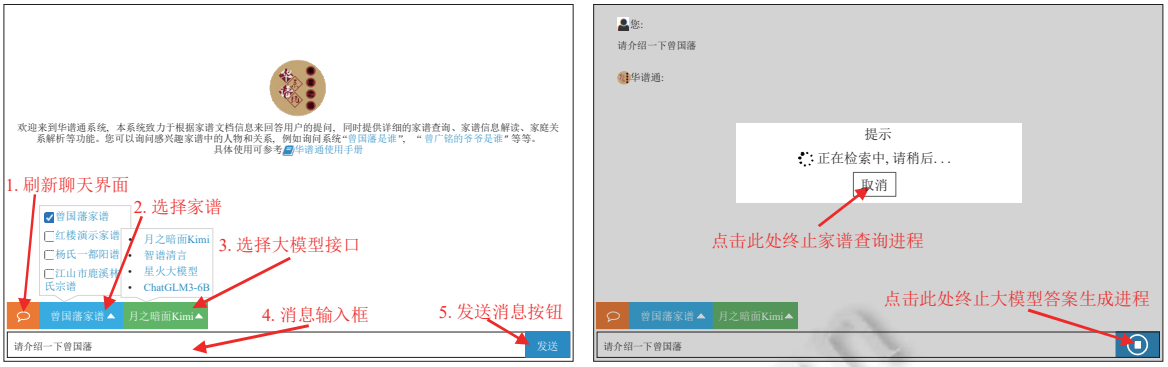
图 10(a) 展示了华谱平台对家谱数据的访问权限控制方案. 总体而言, 华谱平台将家谱的可访问等级划分为“公开”“共建”“私有”这 3 级, 并将用户划分为“游客/普通用户”“家谱共建者”“家谱创建者”“系统管理员”这 4 类进行对应. 在不考虑修改家谱信息的前提下, 我们以家谱创建者为例阐述该权限控制方案: 一个注册并登录的华谱用户, 可以进入“家谱建设”模块录入家谱信息并设置家谱可访问等级. 当可访问等级设置为“公开家谱”时, 该家谱的信息被所有人可见; 当为“共建家谱”时, 该家谱仅能被创建者和指定的共建者访问; 同理, 私有家谱仅由创建者可见; 而后台系统管理员对所有家谱都有访问权限. 需要注意的是, “游客/普通用户”“家谱共建者”“家谱创建者”这 3 类用户的角色会根据具体家谱的可访问等级而发生改变, 这能保证每份家谱只对符合访问权限的用户可见, 进而预防家谱信息泄露的隐患.



(a) 华谱平台用户与家谱访问权限　　(b) 华谱通调用华谱数据权限管理功能

图 10　华谱平台用户与数据访问权限设计与华谱通调用情况

在此基础上, 华谱通中用户可访问的家谱数据也遵循上述权限控制方案. 如图 10(b) 所示, 在用户登录华谱平台时, 后台系统会自动检测该用户可访问的家谱信息, 并整合为一份“可见家谱列表”. 当用户进入华谱通问答模块后, 该“可见家谱列表”会映射到前端的“选择家谱”区域 (参考图 11), 因此, 用户在华谱通界面中只能选择“可见家谱列表”中的家谱进行问答, 而不会访问到权限以外的家谱. 此外, 为防止大模型在多线程问答过程中出现家谱数据泄露的问题, 华谱通通过用户 ID 来唯一标识当前用户对应的线程. 在处理当前用户的问题时, 华谱通只允许大模型访问该用户 ID 对应线程内的历史问答记录, 并全称跟踪该线程直至该用户退出或刷新问答界面. 通过上述华谱平台数据访问权限设置和问答过程中的线程标记法, 华谱通可以从用户操作和大模型问答两个角度实现数据保护, 避免了家谱中隐私信息的泄露问题.



(a) 页面基础功能 (b) 问答中断机制

图 11 华谱通系统页面基础功能和问答中断机制

2.8 系统框架适用性分析

华谱通作为一个基于知识推理的大语言模型问答系统, 在家谱知识问答领域的表现已经达到了先进水平. 以华谱通为参考, 本节将大语言模型在特定领域知识上实现高效问答的技术路线做出以下总结.

① 领域知识图谱是大语言模型在特定领域实现精准问答的根本要求. 领域知识图谱的主旨是精准的知识归纳, 即不需要海量的知识储备, 但力求每条知识都是该领域的核心内容, 且形成的知识网能够概括该领域的主要内容. 例如, 华谱平台仅通过“父亲”“儿子”“女儿”“丈夫”和“妻子”这 5 种基础关系, 就能实现千万级别的人物关系存储, 这为后续的知识推理提供充足的知识储备.

② 完备的知识推理规则是保证领域知识被正确利用的必要条件. 与一般的知识检索增强框架仅依靠语义相似性获取大模型的提示信息不同, 基于知识图谱的推理框架可以通过核心知识点定义完备的推理逻辑, 以可解释的推理路径获取所有与用户问题高度相关的候选答案. 例如, 根据家谱中 5 种基础关系定义的 26 条 Jena 推理规则, 华谱通可以实现常见亲属关系的精准推理, 这极大提升了大模型提示信息的完整性.

③ 冗余信息筛选机制是满足知识推理结果确定性的重要条件. 由于用户提供问题的模糊性和领域知识图谱结构的复杂性, 知识推理规则在考虑推理完备性的同时, 势必会引入部分二义性信息干扰大模型解答, 如同名实体或冗余推理路径. 因此, 冗余信息筛选机制主要通过同名实体消歧与最优推理路径选择技术, 过滤对大语言模型问答没有实质帮助的知识. 例如, 华谱通中部署的多条件匹配机制和路径排序算法分别从同名人物和多关系路径的角度, 进一步提升用户问题条件下的知识推理结果确定性, 以便更有效地提示大模型回答问题.

④ 基于大语言模型的人机交互接口是大模型知识问答系统正常运转的基本保障. 用户提供的自然语言问题与知识图谱的结构化存储形式存在差异, 因此, 在实际问答过程中, 需要先将用户问题转化为知识三元组格式, 才能正常开展后续的知识推理与冗余信息筛选等环节. 此外, 为保证用户在问答过程中的体验感, 还需要将推理出的结构化知识以自然语言的形式反馈给用户. 为此, 华谱通利用大语言模型强大的自然语言理解能力与对话功能, 实现不同数据结构的格式转化, 以连通“用户问题→知识推理→系统反馈”的问答流程, 从而保障系统的正常运行.

此外, 如前文提到的系统响应异常处理机制 (第 2.6 节) 和数据隐私保护功能 (第 2.7 节), 都是保证华谱通系统鲁棒性的重要功能.

理论上, 在上述技术路线的协同运作下, 华谱通可以被作为一个基础的问答框架, 迁移到任意领域的知识问答场景中 (如医疗、物理、生物制药等), 这展示了华谱通系统框架广泛适用性. 然而, 现实生活中存在动态更新的领域知识, 而目前华谱通中的知识推理框架和路径排序算法需要通过领域专家定义推理规则和关系分数, 该过程限制了系统框架在动态变化知识上的适用性. 为实现更加灵活的大模型知识推理问答框架, 本文在总结部分论述了华谱通在可扩展性方向上的研究思路.

3 系统功能展示与问答场景分析

华谱通支持多种场景的家谱知识问答, 包括人物亲属查询、人物关联查询、同名人物筛选、指代模糊推理

以及部分跨家谱问答场景. 本节从系统功能展示和问答场景分析两个角度来体现华谱通在家谱问答任务中的有效性.

3.1 华谱通前端功能展示

如图 11 所示, 华谱通依托华谱平台而构建, 目前收录 4 份公开家谱与若干私有家谱 (因考虑用户隐私, 此处只展示公开家谱), 并支持 4 个开源大模型问答接口. 与 ChatGPT 等大模型的网页端功能类似, 华谱通也支持自然语言问答功能. 用户可以在选择合适的家谱和大模型接口之后, 向系统下方的消息输入框中输入想要询问的内容, 再点击右下角的“发送”按钮, 即可等待华谱通对问题进行解答.

此外, 考虑到问答过程中可能出现的家谱知识查询时间过长和大模型生成内容过多的问题, 华谱通允许用户随时中断当前问答进程. 如图 11 所示, 用户在前端点击相应的终止按钮, 华谱通会根据前端请求终止本次问答进程, 并创建一个新的进程作为下一轮问答的消息传输载体.

3.2 人物亲属查询和人物关联查询

人物亲属和关系查询是智能家谱问答最基础的功能, 问答的准确性直接反映了本文所提出的知识图谱推理框架是否能对大模型的答案生成起到正向的提示作用. 因此, 从问答精度测评方面, 本文根据华谱平台公开的《曾国藩家谱》《红楼演示家谱》《杨氏—都阳谱》和《江山鹿溪林氏宗谱》设计了一份包含 300 条问答对的测试数据集, 涉及第 2.2 节提出的 26 种亲属关系, 各种问题类型的数量分布如图 12 所示.

图 12 测试数据集各类问题数量分布

此外, 为例验证华谱通中知识图谱推理框架在家谱知识问答上的优越性, 本文选取了两个国内外先进的 LLM 文档问答框架 (智谱清言长文档解读功能 (https://chatglm.cn/main/doc) 和 LangChain RAG (https://github.com/langchain-ai/langchain)), 在《曾国藩家谱》《红楼演示家谱》《杨氏—都阳谱》和《江山鹿溪林氏宗谱》上与华谱通进行问答质量对比.

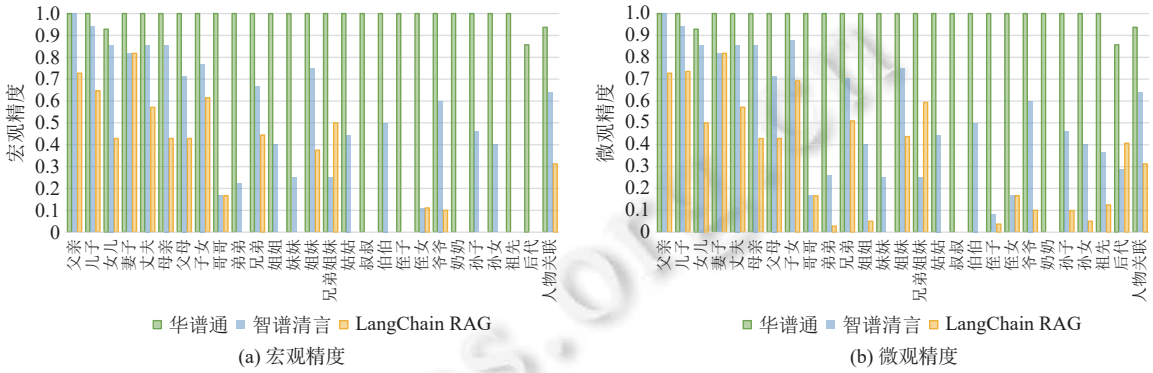
在问答测评时, 考虑到部分人物亲属问题的答案中可能包含多个人物信息, 本文从宏观问答精度 (完全答对算正确) 和微观问答精度 (按照答对内容占真实答案的百分比计算精度; 若回答内容存在不属于真实答案的部分, 则定义为答错) 两个层面评价华谱通的问答质量. 具体测试分析如下.

表 3 展示了华谱通和对比方法在曾国藩家谱上的总体问答精度, 图 13 展示了华谱通和对比方法在曾国藩家谱各类问题上的问答精度对比. 总体而言, 华谱通能够答对数据集中所有的人物亲属问题, 这得益于系统中完备的 Jena 推理规则, 它保证了华谱通能在家谱中检索出所有与指定关系相关的目标人物, 从而促使 LLM 生成准确且全面的答案 (在表 3 和图 13 中体现为华谱通拥有相同的宏/微观精度). 而智谱清言和 LangChain RAG 在不同人物亲属问题上的表现差异较大. 如图 13 所示, 它们在“兄弟姐妹”“叔伯姑侄”和“爷孙”这 3 类问题上的问答精度明

显低于华谱通, 我们将这种结果归因于智谱清言和 LangChain RAG 缺乏完备的知识推理逻辑, 进而难以检索到完整的候选信息用于提示大模型回答问题.

表 3 华谱通和对比方法在曾国藩家谱上的总体问答精度 (%)

模型	宏观精度	微观精度
华谱通	97.67	97.67
智谱清言	56.34	58.60
LangChain RAG	30.33	34.26



(a) 宏观精度 (b) 微观精度

图 13 华谱通和对比方法在曾国藩家谱上的问答精度

例如, 对“兄弟姐妹”和“叔伯姑侄”这类可能包含多个目标人物的问题, 单纯依靠文档检索是难以全面捕获所有答案的. 此外, 大模型本身可能难以理解“伯伯/叔叔”是“父亲的哥哥/弟弟”这一关系规则, 因此, 智谱清言和 LangChain RAG 在缺少完整的文档检索内容和关系规则提示的前提下, 难以在一些复杂关系的问答场景中取得较好的效果.

相比之下, 得益于完备的自定义 Jena 推理规则, 华谱通在家谱检索时能够准确区分“伯伯”和“叔叔”等复杂关系, 从而精准地定位到家谱中的正确信息. 在这种情况下, 华谱通传递给大模型的家谱知识在准确性和全面性上相较于其他两个对比方法而言具有显著优势. 从另一个角度来看, 由于能够检索到足够精确的家谱知识, 华谱通无需让大模型自身去解析复杂的关系, 这也是华谱通能在相对复杂的关系问答中给出正确答案的关键因素.

在人物关联问题上, 华谱通的问答质量有所下降, 但仍远高于其他两个方法. 我们认为华谱通在这类问题中出现错误的原因与路径排序有关. 不同于人物亲属问题中关系的唯一性, 人物关联问题可能涉及多条多跳关系路径, 尽管华谱通依靠 Jena 逻辑推理机制能够推理出正确的路径, 但受限于关系初始分数的设计, 所得到的路径可能难以直观地体现人物之间的关联. 例如, 根据第 2.3.2 节提出的算法 2, 原本应该是叔侄关系 (4 分) 的最优路径, 华谱通会检索出路径“父亲→兄弟” (5.36 分), 这导致华谱通难以完全准确地回答出两个人之间直接关系, 造成问答结果错误. 这也是华谱通在未来需要重点改进的功能.

3.3 复杂问答场景分析

本节将针对同名人物问答、指代模糊推理和跨家谱问答 3 个场景来分析华谱通在面对复杂问答场景时的有效性. 本节以智谱清言在上述场景的问答结果作为对照组, 用于和华谱通的问答结果进行比较. 为便于读者比较华谱通和智谱清言的问答质量, 本节在后文图 14 中提供了与问答内容相关的家谱人物信息以供参考.

3.3.1 同名人物问答

本节分析了华谱通和智谱清言在面对图 14 中同名人物“陈氏”时的问答场景.

华谱通的问答过程如图 15(a) 场景所示. 显然, 华谱通所回答的关于“陈氏”的信息与图 14 中两位“陈氏”的信息都能匹配, 尤其是针对“已知陈氏的父亲是陈彰绶, 请介绍一下她的丈夫”和“已知陈氏的父亲是陈池生, 请告诉我陈氏和曾国藩什么关系?”两个问题, 华谱通可以根据用户提供的条件, 精准地定位到两个不同的“陈氏”, 这得益于多条件匹配机制对同名人物的准确筛选.

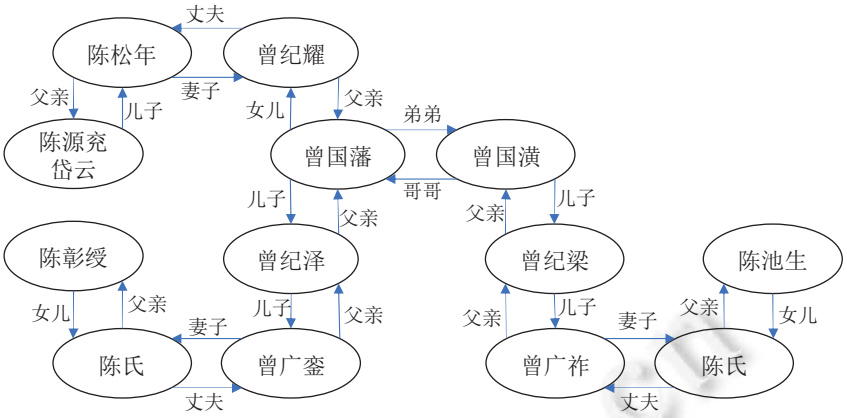
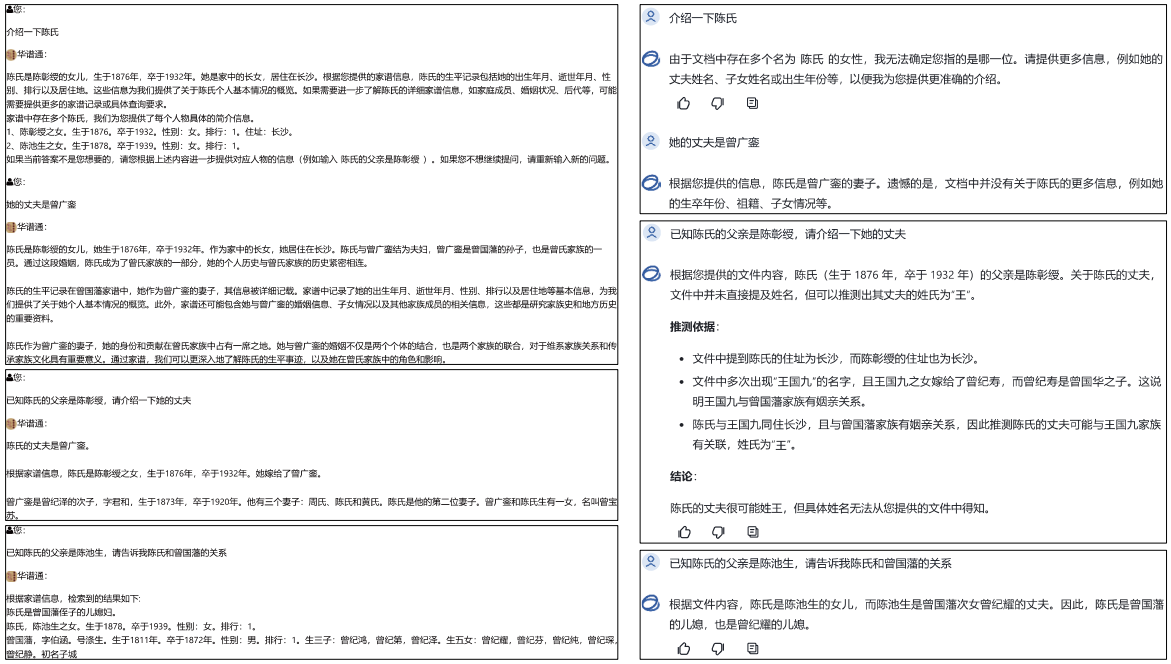


图 14 复杂场景问答涉及的人物关系



(a) 华谱通 (b) 智谱清言

图 15 华谱通和智谱清言同名人物问答场景

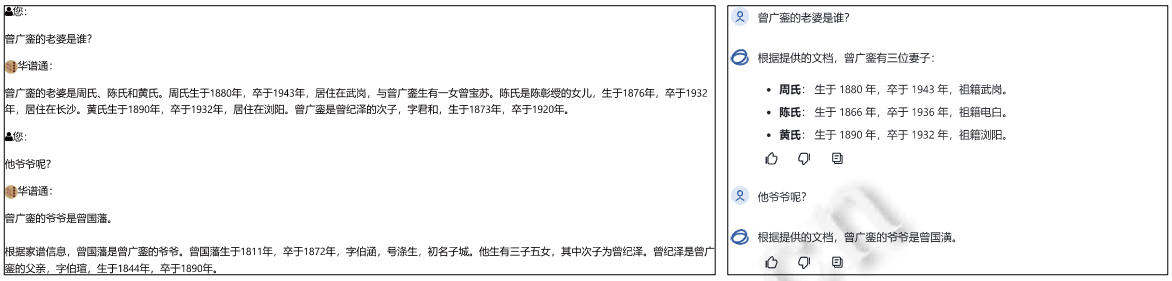
相比之下, 如图 15(b) 问答场景所示, 由于智谱清言所使用的文档检索机制在推理逻辑完备性上弱于华谱通, 导致其混淆了“陈彰绶之女”“陈池生之女”和“陈源兖岱云之子”, 进而将不完整或错误的检索结果提供给大模型进行问答, 造成最终问答结果与图 14 所示的人物信息存在不相符的问题.

3.3.2 指代模糊推理

如第 2.4 节所述, 用户在多轮问答中可能会用“他/她”等代词指代前期问答过程中的某个人. 为了让大模型能理解这种人物代指模糊的问题, 华谱通在知识图谱推理框架的基础上, 增加了基于历史问答记录提示的大模型多轮问答机制, 使华谱通能通过问答记录的上下文信息回答用户问题. 以下是针对华谱通和智谱清言在指代模糊的问答上的场景分析, 问答过程中涉及的人物信息同样以图 14 为准.

如图 16 所示, 华谱通能够处理指代模糊问题, 这得益于大模型对问答历史上下文信息的获取能力. 此外, 相较

于智谱清言在第 2 轮问答中误将曾国潢作为曾广銮的爷爷, 华谱通在明确问题中的“他”代指曾广銮之后, 仍然会执行一次家谱图谱查询的操作. 因此, 华谱通能够在多轮问答中同时利用历史问答记录和家谱逻辑查询结果来增加大模型生成正确答案的概率.



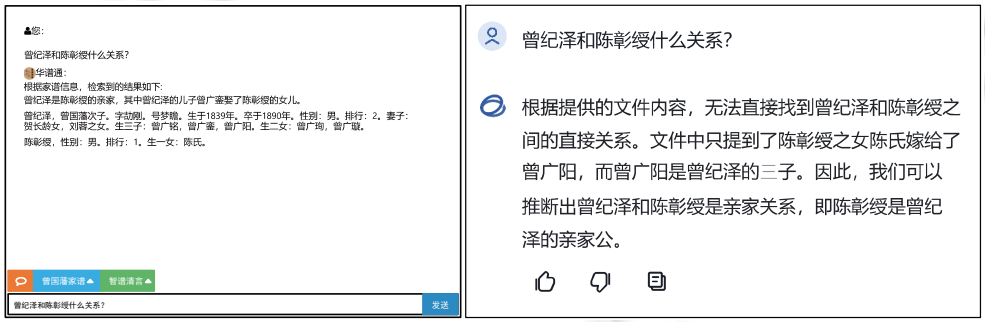
(a) 华谱通 (b) 智谱清言

图 16 华谱通和智谱清言指代模糊问答场景

3.3.3 跨家谱问答

为验证第 2.5 节所述的跨家谱问答技术路线的有效性, 本节分别从婚姻关系和社会关系两个场景来展示华谱通中逻辑推理式和关系生成式跨家谱问答模块的问答结果.

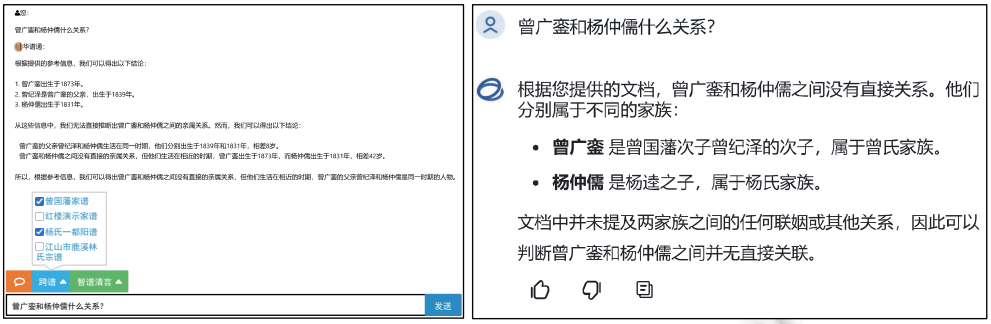
图 17 展示了华谱通和智谱清言在跨谱婚姻关系上的问答质量. 由于目前华谱系统中并未收集“陈彰绶”所在的陈氏家谱, 本文以《曾国藩家谱》中的“陈彰绶” (如图 14 所示) 作为跨谱问答的目标人物. 如问答结果所示, 尽管华谱通和智谱清言都能够回答出“曾纪泽和陈彰绶是亲家”这一关系, 但是智谱清言在利用大模型进行文档检索与语义理解时误将“曾广阳”当作“曾广銮”, 导致其提供的推理依据出现错误. 相比之下, 华谱通可以利用 Jena 规则推理出的可解释人物关系路径“曾纪泽→曾广銮 (儿子)→陈氏 (妻子)→陈彰绶 (父亲)”更直观地引导大模型进行关系总结, 从而提供准确的人物关系和推理依据.



(a) 华谱通 (b) 智谱清言

图 17 华谱通和智谱清言跨家谱问答场景 (婚姻关系)

图 18 展示了华谱通和智谱清言在社会关系方面的问答质量. 为了验证华谱通对无关联家谱人物之间的关系生成能力, 本文选择了两份公开的无关联家谱 (《杨氏—都阳谱》与《曾国藩家谱》) 进行跨谱问答. 在具体问答场景中, 用户提问“曾广銮和杨仲儒什么关系?”, 智谱清言首先检索并获取《曾国藩家谱》和《杨氏—都阳谱》中与二人有关的信息, 再通过大模型对检索结果进行总结. 由于两份家谱之间不存在有确定关联的人物, 智谱清言所回答的结果实际上是正确的. 相比之下, 为了体现华谱“人人互联”的特点, 华谱通设计了一种基于祖先年代匹配的模糊推理机制, 通过检索二人之间生活年代最接近的祖先或后代, 引导大模型生成模糊的社会关系. 根据图 18 的结果可以看出, 尽管曾广銮和杨仲儒之间没有确定关系, 但是华谱通生成了一条“曾纪泽 (曾广銮之父) 和杨仲儒生活于同一时期”的关系路径, 从而实现了曾广銮和杨仲儒之间关联, 这体现了华谱通在跨谱人物关联方面问答结果的多样性.

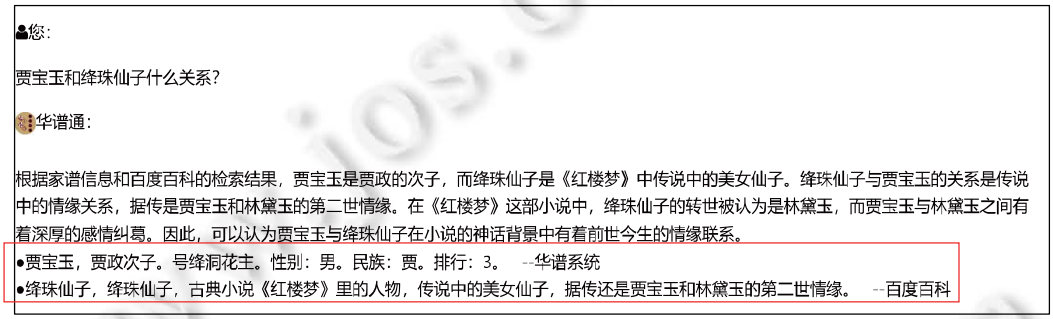


(a) 华谱通　　(b) 智谱清言

图 18　华谱通和智谱清言跨家谱问答场景 (社会关系)

3.3.4　百度百科问答场景展示与错误分析

为验证第 2.6 节所提出异步请求响应机制中百度百科检索模块对华谱通问答的有效性, 本节以《红楼演示家谱》中缺失的专有词语“绛珠仙子”和“十二金钗”为例进行测试, 图 19 展示了华谱通使用 Kimi 大模型回答测试案例的问答场景.



(a) 引入专有词语解释与来源, 提升问答的可解释性

(b) 家谱推理+百度百科检索　　(c) 家谱推理正确结果

图 19　百度百科检索模块问答案例展示

如图 19(a) 所示, 华谱通可以协同利用百度百科模块的检索结果和家谱中的相关信息, 充分调动大模型利用预训练知识来回答包含家谱缺失人物的问题. 同时, 当华谱通调用百度百科查询接口后, 前端会展示专有词语的来源, 以此提升问答的可解释性, 从人机交互的角度增强用户对大模型回答正确性的判断.

然而, 目前华谱通调用百度百科查询接口的功能还有待完善. 如图 19(b) 所示, 尽管华谱通根据百度百科检索模块可以获知“十二金钗”指林黛玉、薛宝钗、王熙凤、贾元春等人物, 但无法将这些的人物分别抽取出来, 与家谱中的“贾宝玉”进行基于路径排序的亲属关系推理. 因此, 图 19(b) 的回答是大模型参考《红楼演示家谱》中贾宝玉的简介和百度百科对“十二金钗”的介绍, 利用预训练的知识所生成的答案, 而并未从家谱中获取“十二金钗”每个人物与“贾宝玉”的准确关系, 这导致华谱通对大模型问答幻觉的约束力度降低, 进而导致大模型生成一些错误的答案, 例如“李纨是贾宝玉的亲嫂子”而非“堂嫂”. 相比之下, 如图 19(c) 所示, 将 3 个错误回答对应的问题单独输入系统后, 华谱通都能正确回答. 这是因为在单独询问这 3 个问题时, 待查询的人物都包含在《红楼演示家谱》内, 因此华谱通能够通过前文提出的逻辑推理框架和路径排序算法精准定位到人物之间的关系路径, 从而更加有效地提示大模型回答问题. 综上所述, 如何在使用百度百科检索模块的基础上进一步发挥华谱通中亲属关系推理规则的优势是华谱通在未来需要重点研究的方向.

3.4 系统时间性能分析

本节针对华谱通系统中部署的路径排序算法 (第 2.3.2 节) 和异步请求响应机制 (第 2.6 节), 展开详细的时间性能分析, 以证明华谱通在实际的问答场景中能够快速响应用户输入的问题.

3.4.1 关系路径排序算法时间复杂度有效性实验

为验证第 2.3.2 节中基于大根堆的 Dijkstra 算法 (Heap Dijkstra) 在排序时间上的优势, 本文基于《红楼演示家谱》构建了 4 条测试样例, 以测试算法在不同亲疏程度的关系路径上的排序时间. 同时, 本文设置了 Naïve Dijkstra 算法作为对照实验. Naïve Dijkstra 算法不使用任何措施维护路径分数表, 而是通过遍历更新后的无序分数表来获得最高分的路径. 具体实验结果如表 4 所示. 根据表 4 中不同样例所需的响应时间可得, 当待查询两个人物之间的关系强度越弱时 (关系路径越长), Heap Dijkstra 算法的优势越明显.

表 4 路径排序算法响应时间统计 (s)

测试样例	关系路径	Naïve Dijkstra	Heap Dijkstra
样例1: 贾宝玉和贾元春什么关系?	贾宝玉→**(姐妹)**→贾元春	**0.0878**	**0.0928**
样例2: 贾宝玉和贾演什么关系?	贾宝玉→**(父亲)**→贾政→**(爷爷)**→贾源→**(兄弟)**→贾演	6.4979	5.1663
样例3: 贾宝玉和林黛玉什么关系?	贾宝玉→**(父亲)**→贾政→**(姐妹)**→贾敏→**(女儿)**→林黛玉	11.3144	7.9809
样例4: 贾宝玉和贾珍什么关系?	贾宝玉→**(父亲)**→贾政→**(爷爷)**→贾源→**(侄子)**→贾代化→**(孙子)**→贾珍	321.2256	103.4212

注: 响应快的算法在每个样例上的时间用粗体表示

根据第 2.3.2 节的描述, 华谱通的路径排序算法需要在动态更新的无序路径分数列表中找到最高分的路径, 因此每一轮的排序都包含插入和查找两个步骤. 假设上一轮排序后, 列表中共存储 K 条关系路径, 在不考虑用大根堆维护路径分数列表时, Naïve Dijkstra 算法将新增的 26 条关系路径插入列表的时间为 O(1), 但需要遍历所有路径分数才能找到分数最大的路径, 即需要 O(K+26) 的查找时间. 相比之下, Heap Dijkstra 算法将插入 26 条新关系路径后的分数表重新调整为大根堆的时间为 O(log(K+26)), 且仅需将堆顶元素作为分数最大的路径, 因此查找时间为 O(1). 综上所述, Heap Dijkstra 算法 (O(log(K+26))+ O(1)) 的总时间复杂度要优于 Naïve Dijkstra 算法 (O(K+26)+O(1)).

3.4.2 异步请求响应机制有效性实验

为测试第 2.6 节所提出方法对华谱通问答系统响应时长的优化程度, 本节使用《红楼演示家谱》中的“贾宝玉”“贾代善”“林黛玉”“十二金钗”“绛珠仙子”和“神瑛侍者”构造测试样例进行有效性检验. 由于华谱平台暂未对人物的别称做专门的数据管理, 因此目前无法实现人物别称到具体人物实体的映射. 在这种情况下, “十二金钗”“绛珠仙子”和“神瑛侍者”可以被认为是《红楼演示家谱》中的缺失人物. 考虑到异步请求响应机制中家谱推理模块

和百度百科检索模块部署的多线程异步执行功能原理相同, 本节以家谱推理模块为例进行具体实验. 如表 5 所示, 本实验采用的 4 条样例拥有不同的路径排序速度和缺失人物数量, 有下划线和无下划线的“回答情况”分别代表大模型通过百度百科检索和家谱推理两个模块提供的信息回答对应的样例.

表 5　异步请求响应机制中家谱推理模块最大时限参数实验

测试样例	关系路径排序速度	家谱中缺失人物数	场景1: 无异步请求响应机制 (路径排序最大时限60 s)		场景2: 有异步请求响应机制 (路径排序最大时限10 s)		场景3: 有异步请求响应机制 (路径排序最大时限60 s)	
			页面响应时间 (s)	回答情况	页面响应时间 (s)	回答情况	页面响应时间 (s)	回答情况
样例1: 贾宝玉和贾代善什么关系?	快	0	22.66	回答正确	22.90	回答正确	21.66	回答正确
样例2: 贾宝玉和林黛玉什么关系?	适中	0	31.28	回答正确	16.15	回答正确	29.23	回答正确
样例3: 贾宝玉和十二金钗什么关系?	慢	1	超时	无法回答	46.45	回答正确	49.02	回答正确
样例4: 绛珠仙子和神瑛侍者什么关系?	慢	2	超时	无法回答	55.60	回答正确	59.11	回答正确

在无异步请求响应机制 (场景 1) 的情况下, 华谱通只能先依靠家谱推理模块中的路径排序来推理人物关系, 再将推理结果反馈给大模型进行问答. 此时, 家谱推理模块能在最大时限内推理出样例 1 和样例 2 中人物的关系路径, 这使得华谱通能准确回答问题. 然而, 当样例中包含指定家谱中缺失的人物时, 家谱推理模块无法推理出人物之间的关系, 且该结论需要在遍历完所有候选路径后才能得出, 这极大增加了系统的响应时间. 因此, 当路径排序最大时限为 60 s 时, 华谱通无法在场景 1 中对样例 3、4 进行回答.

相比之下, 在部署了异步请求响应机制后, 华谱通能在路径排序最大时限为 10 s 的情况下答对所有样例 (场景 2), 且总体响应时间远小于场景 1 的问答过程. 例如, 对样例 2 而言, 在路径排序超时的情况下, 华谱通直接利用“贾宝玉”和“林黛玉”在家谱中的个人信息来提示大模型回答问题. 由于“贾宝玉”和“林黛玉”之间的关系极有可能包含在公开数据集中而被大模型在预训练阶段获取, 华谱通提供的个人信息能够充分激发大模型利用预训练知识准确地回答问题, 且响应时长要优于场景 1 中华谱通对样例 2 回答情况. 此外, 针对场景 1 中无法回答的两个样例, 华谱通借助异步请求响应机制中的百度百科检索也能够准确且迅速地回复.

考虑到用户可能会查询一些大模型预训练知识以外的人物关系问题, 在系统的实际部署中, 家谱推理模块对路径排序的最大时限被设置为 60 s, 允许后台尽可能多地提供家谱中的准确信息, 以平衡华谱通问答的准确性和响应时间.

4 总　结

本文提出了一种基于知识图谱推理的大语言模型家谱知识问答系统——华谱通, 从推理逻辑完备性和信息筛选精准性两个角度完善了当前面向大语言模型的知识检索框架在家谱知识问答领域的缺陷. 与当前基于文档语义相似度匹配的知识检索框架不同, 华谱通通过自定义的 Jena 推理规则, 在完成可解释知识推理的同时, 实现了推理逻辑的完备性, 保证推理结果中能包含完整的候选知识. 此外, 考虑到现有知识检索框架缺乏对家谱中同名人物和多关系路径的过滤机制, 本文提出基于多条件匹配和路径排序算法的冗余信息筛选机制, 以过滤对大模型问答无用的家谱信息, 进一步提升华谱通的问答质量. 与智谱清言和 LangChain RAG 在《曾国藩家谱》《红楼演示家谱》《杨氏—都阳谱》和《江山鹿溪林氏宗谱》测试数据上的对比实验与案例分析体现了华谱通在家谱问答任务上的优越性.

在未来的工作中, 华谱通将主要从以下几个方面展开.

首先, 针对目前华谱通中因需要人工定义推理规则和关系分数而导致的适用性不足问题, 拟通过微调和提示学习结合的技术手段[42], 充分利用 LLM 的知识涌现能力来实现新规则的生成和关系路径重要性的排序, 使华谱通能适应动态变化的家谱信息, 进而将系统框架扩展到实时更新的知识领域实现精准问答.

其次, 构建一种 LLM 和知识图谱双向增强的问答系统. 目前的华谱通还处于知识图谱增强的 LLM 问答研究领域. 在知识生成方面, 华谱通目前还无法根据用户实时提供的信息抽取新的家谱知识. 后期拟研究基于 LLM 的知识图谱管理框架, 利用 LLM 自身的预训练参数化知识, 从知识图谱构建、知识编辑和知识验证等方面打通知识图谱与 LLM 的数据交互, 以构建更灵活的人机交互问答系统.

此外, 在跨家谱问答方面, 华谱通目前仅做了部分前瞻性的技术调研与实验分析, 以此验证跨家谱问答技术路线的可行性与有效性. 在后续的工作中, 拟参考亲属关系的 Jena 推理规则定义方法, 重点针对基础社会关系 (如同事、师生、朋友等) 设计一个跨谱关系推理机制, 并结合人物属性和亲属信息的粗粒度匹配结果, 引导大模型动态生成更多样的社会关系, 以保证华谱通在复杂人物关联问题上的求解灵活性.

最后, 在上述 3 个研究路线的基础上, 探索华谱通的泛化能力. 目前华谱通仅在家谱问答方面验证了完备性推理逻辑对 LLM 在垂直领域知识库上问答的有效性. 在未来, 拟考虑将华谱通的知识图谱推理框架应用到不同领域的问答场景中, 如医疗、金融和生物制药等.

References:

[1] Wu XD, Li J, Zhou P, Bu CY. Fusion technique for fragmented genealogy data. Ruan Jian Xue Bao/Journal of Software, 2021, 32(9): 2816–2836 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6010.htm [doi: 10.13328/j.cnki.jos.006010]

[2] Wu XD, Jiang TT, Zhu Y, Bu CY. Knowledge graph for China's genealogy. IEEE Trans. on Knowledge and Data Engineering, 2023, 35(1): 634–646. [doi: 10.1109/TKDE.2021.3073745]

[3] Wu XD, Sheng SJ, Jiang TT, Bu CY, Wu MH. Huapu-CP: From knowledge graphs to a data central-platform. Acta Automatica Sinica, 2020, 46(10): 2045–2059 (in Chinese with English abstract). [doi: 10.16383/j.aas.c200502]

[4] Wu XD, Chen HH, Wu GQ, Liu J, Zheng QH, He XF, Zhou AY, Zhao ZQ, Wei BF, Li Y, Zhang QP, Zhang SC. Knowledge engineering with big data. IEEE Intelligent Systems, 2015, 30(5): 46–55. [doi: 10.1109/MIS.2015.56]

[5] Brown TB, Mann B, Ryder N, *et al*. Language models are few-shot learners. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 1877–1901.

[6] Yin CL, Du KP, Nong Q, Zhang HC, Yang L, Yan B, Huang X, Wang XB, Zhang X. PowerPulse: Power energy chat model with LLaMA model fine-tuned on Chinese and power sector domain knowledge. Expert Systems, 2024, 41(3): e13513. [doi: 10.1111/exsy.13513]

[7] Teubner T, Flath CM, Weinhardt C, van der Aalst W, Hinz O. Welcome to the era of ChatGPT *et al*. Business & Information Systems Engineering, 2023, 65(2): 95–101. [doi: 10.1007/s12599-023-00795-x]

[8] Tay Y, Dehghani M, Bahri D, Metzler D. Efficient transformers: A survey. ACM Computing Surveys, 2023, 55(6): 109. [doi: 10.1145/3530811]

[9] Ji ZW, Lee N, Frieske R, Yu TZ, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P. Survey of hallucination in natural language generation. ACM Computing Surveys, 2023, 55(12): 248. [doi: 10.1145/3571730]

[10] Pan SR, Luo LH, Wang YF, Chen C, Wang JP, Wu XD. Unifying large language models and knowledge graphs: A roadmap. IEEE Trans. on Knowledge and Data Engineering, 2024, 36(7): 3580–3599. [doi: 10.1109/TKDE.2024.3352100]

[11] Chen JW, Lin HY, Han XP, Sun L. Benchmarking large language models in retrieval-augmented generation. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 17754–17762. [doi: 10.1609/aaai.v38i16.29728]

[12] Wu XD, Zhu XQ, Wu GQ, Ding W. Data mining with big data. IEEE Trans. on Knowledge and Data Engineering, 2014, 26(1): 97–107. [doi: 10.1109/TKDE.2013.109]

[13] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih WT, Rocktäschel T, Riedel S, Kiela D. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 9459–9474.

[14] Futia G, Vetro A, Melandri A, De Martin JC. Training neural language models with SPARQL queries for semi-automatic semantic mapping. Procedia Computer Science, 2018, 137: 187–198. [doi: 10.1016/j.procs.2018.09.018]

[15] Sun JS, Xu CJ, Tang LMY, Wang SZ, Lin C, Gong YY, Ni LM, Shum HY, Guo J. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. In: Proc. of the 12th Int'l Conf. on Learning Representations. Vienna: OpenReview.net, 2024. 1–31.

[16] Ma YB, Cao YX, Hong Y, Sun AX. Large language model is not a good few-shot information extractor, but a good reranker for hard samples. In: Proc. of the 2023 Findings of the Association for Computational Linguistics: EMNLP. Singapore: Association for

Computational Linguistics, 2023. 10572–10601. [doi: 10.18653/v1/2023.findings-emnlp.710]

[17] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

[18] Wang NY, Ye YX, Liu L, Feng LZ, Bao T, Peng T. Language models based on deep learning: A review. Ruan Jian Xue Bao/Journal of Software, 2021, 32(4): 1082–1115 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6169.htm [doi: 10.13328/j.cnki.jos.006169]

[19] Li ZY, Zhu HX, Lu ZR, Yin M. Synthetic data generation with large language models for text classification: Potential and limitations. In: Proc. of the 2023 Conf. on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, 2023. 10443–10461. [doi: 10.18653/v1/2023.emnlp-main.647]

[20] Li G, Peng X, Wang QX, Xie T, Jin Z, Wang J, Ma XX, Li XD. Challenges from LLMs as a natural language based human-machine collaborative tool for software development and evolution. Ruan Jian Xue Bao/Journal of Software, 2023, 34(10): 4601–4606 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7008.htm [doi: 10.13328/j.cnki.jos.007008]

[21] Zhao ZH, Fan WQ, Li JT, Liu YQ, Mei XW, Wang YQ, Wen Z, Wang F, Zhao XY, Tang JL, Li Q. Recommender systems in the era of large language models (LLMs). IEEE Trans. on Knowledge and Data Engineering, 2024, 36(11): 6889–6907. [doi: 10.1109/TKDE.2024.3392335]

[22] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019. 4171–4186. [doi: 10.18653/v1/N19-1423]

[23] Li SC, Wang ZQ, Zhou GD. LLM enhanced cross domain aspect-based sentiment analysis. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 644–659 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7156.htm [doi: 10.13328/j.cnki.jos.007156]

[24] Liang Z, Wang HZ, Dai JJ, Shao XY, Ding XO, Mu TY. Interpretability of entity matching based on pre-trained language model. Ruan Jian Xue Bao/Journal of Software, 2023, 34(3): 1087–1108 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6794.htm [doi: 10.13328/j.cnki.jos.006794]

[25] Ju SG, Huang FY, Sun JP. Idiom cloze algorithm integrating with pre-trained language model. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3793–3805 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6307.htm [doi: 10.13328/j.cnki.jos.006307]

[26] Shi D, You CB, Huang JT, Li TH, Xiong DY. CORECODE: A common sense annotated dialogue dataset with benchmark tasks for Chinese large language models. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 18952–18960. [doi: 10.1609/aaai.v38i17.29861]

[27] Xu P, Patwary M, Shoeybi M, Puri R, Fung P, Anandkumar A, Catanzaro B. MEGATRON-CNTRL: Controllable story generation with external knowledge using large-scale language models. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing. Pennsylvania: Association for Computational Linguistics, 2020. 2831–2845. [doi: 10.18653/v1/2020.emnlp-main.226]

[28] Verma G, Rossi R, Tensmeyer C, Gu JX, Nenkova A. Learning the visualness of text using large vision-language models. In: Proc. of the 2023 Conf. on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, 2023. 2394–2408. [doi: 10.18653/v1/2023.emnlp-main.147]

[29] Wu TY, He SZ, Liu JP, Sun SQ, Liu K, Han QL, Tang Y. A brief overview of ChatGPT: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica, 2023, 10(5): 1122–1136. [doi: 10.1109/JAS.2023.123618]

[30] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A, Schulman J, Hilton J, Kelton F, Miller L, Simens M, Askell A, Welinder P, Christiano P, Leike J, Lowe R. Training language models to follow instructions with human feedback. In: Proc. of the 2022 Annual Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 27730–27744.

[31] Chowdhery A, Narang S, Devlin J, *et al*. PaLM: Scaling language modeling with pathways. Journal of Machine Learning Research, 2023, 24(240): 1–113.

[32] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou YQ, Li W, Liu PJ. Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, 2020, 21(140): 1–67.

[33] Li S, Chen JJ, Yuan SY, Wu XY, Yang H, Tao SM, Xiao YH. Translate meanings, not just words: IdiomKB's role in optimizing idiomatic translation with language models. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 18554–18563. [doi: 10.1609/aaai.v38i17.29817]

[34] Du ZX, Qian YJ, Liu X, Ding M, Qiu JZ, Yang ZL, Tang J. GLM: General language model pretraining with autoregressive blank infilling. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: Association for Computational

Linguistics, 2022. 320–335. [doi: 10.18653/v1/2022.acl-long.26]

[35] Wang L, Ma YY, Bi WS, Lv HL, Li YX. An entity extraction pipeline for medical text records using large language models: Analytical study. Journal of Medical Internet Research, 2024, 26: e54580. [doi: 10.2196/54580]

[36] Bi KF, Xie LX, Zhang HH, Chen X, Gu XT, Tian Q. Accurate medium-range global weather forecasting with 3D neural networks. Nature, 2023, 619(7970): 533–538. [doi: 10.1038/s41586-023-06185-3]

[37] Li XL, Liang P. Prefix-tuning: Optimizing continuous prompts for generation. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Pennsylvania: Association for Computational Linguistics, 2021. 4582–4597. [doi: 10.18653/v1/2021.acl-long.353]

[38] Zhang ZY, Han X, Liu ZY, Jiang X, Sun MS, Liu Q. ERNIE: Enhanced language representation with informative entities. In: Proc. of the 57th Conf. of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 1441–1451. [doi: 10.18653/v1/P19-1139]

[39] Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond. Foundations and Trends® in Information Retrieval, 2009, 3(4): 333–389. [doi: 10.1561/1500000019]

[40] Agrawal S, Zhou CT, Lewis M, Zettlemoyer L, Ghazvininejad M. In-context examples selection for machine translation. In: Proc. of the 2023 Findings of the Association for Computational Linguistics. Toronto: Association for Computational Linguistics, 2023. 8857–8873. [doi: 10.18653/v1/2023.findings-acl.564]

[41] Qiao SJ, Yang GP, Yu Y, Han N, Tan X, Qu LL, Ran LQ, Li H. QA-KGNet: Language model-driven knowledge graph question-answering model. Ruan Jian Xue Bao/Journal of Software, 2023, 34(10): 4584–4600 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6882.htm [doi: 10.13328/j.cnki.jos.006882]

[42] Luo LH, Li YF, Haffari G, Pan SR. Reasoning on graphs: Faithful and interpretable large language model reasoning. In: Proc. of the 12th Int'l Conf. on Learning Representations. Vienna: OpenReview.net, 2024. 1–24.

[43] Avila CVS, Vidal VMP, Franco W, Casanova MA. Experiments with text-to-SPARQL based on ChatGPT. In: Proc. of the 18th IEEE Int'l Conf. on Semantic Computing. Laguna Hills: IEEE, 2024. 277–284. [doi: 10.1109/ICSC59802.2024.00050]

[44] Shao JX, Liu GL, Ji SW. An abnormal data analysis and processing method for genealogy graph databases. In: Proc. of the 2020 IEEE Int'l Conf. on Knowledge Graph. Nanjing: IEEE, 2020. 131–136. [doi: 10.1109/ICBK50248.2020.00028]

[45] Peng YW, Jiang H, Li RR, Peng ZY. PZXG: A genealogy data service platform for kinship management and application. In: Proc. of the 2020 IEEE Int'l Conf. on Knowledge Graph. Nanjing: IEEE, 2020. 505–512. [doi: 10.1109/ICBK50248.2020.00077]

[46] Dong BB, Zhang Z, Li J, Zhu Y, Bu CY, Wu XD. Hypernode: Entity fusion for data traceability and link prediction. In: Proc. of the 2022 IEEE Int'l Conf. on Data Mining. Orlando: IEEE, 2022. 111–120. [doi: 10.1109/ICDM54844.2022.00021]

附中文参考文献:

[1] 吴信东, 李娇, 周鹏, 卜晨阳. 碎片化家谱数据的融合技术. 软件学报, 2021, 32(9): 2816–2836. http://www.jos.org.cn/1000-9825/6010.htm [doi: 10.13328/j.cnki.jos.006010]

[3] 吴信东, 盛绍静, 蒋婷婷, 卜晨阳, 吴明辉. 从知识图谱到数据中台: 华谱系统. 自动化学报, 2020, 46(10): 2045–2059. [doi: 10.16383/j.aas.c200502]

[18] 王乃钰, 叶育鑫, 刘露, 凤丽洲, 包铁, 彭涛. 基于深度学习的语言模型研究进展. 软件学报, 2021, 32(4): 1082–1115. http://www.jos.org.cn/1000-9825/6169.htm [doi: 10.13328/j.cnki.jos.006169]

[20] 李戈, 彭鑫, 王千祥, 谢涛, 金芝, 王戟, 马晓星, 李宣东. 大模型: 基于自然交互的人机协同软件开发与演化工具带来的挑战. 软件学报, 2023, 34(10): 4601–4606. http://www.jos.org.cn/1000-9825/7008.htm [doi: 10.13328/j.cnki.jos.007008]

[23] 李诗晨, 王中卿, 周国栋. 大语言模型驱动的跨领域属性级情感分析. 软件学报, 2025, 36(2): 644–659. http://www.jos.org.cn/1000-9825/7156.htm [doi: 10.13328/j.cnki.jos.007156]

[24] 梁峥, 王宏志, 戴加佳, 邵心玥, 丁小欧, 穆添愉. 预训练语言模型实体匹配的可解释性. 软件学报, 2023, 34(3): 1087–1108. http://www.jos.org.cn/1000-9825/6794.htm [doi: 10.13328/j.cnki.jos.006794]

[25] 琚生根, 黄方怡, 孙界平. 融合预训练语言模型的成语完形填空算法. 软件学报, 2022, 33(10): 3793–3805. http://www.jos.org.cn/1000-9825/6307.htm [doi: 10.13328/j.cnki.jos.006307]

[41] 乔少杰, 杨国平, 于泳, 韩楠, 覃晓, 屈露露, 冉黎琼, 李贺. QA-KGNet: 一种语言模型驱动的知识图谱问答模型. 软件学报, 2023, 34(10): 4584–4600. http://www.jos.org.cn/1000-9825/6882.htm [doi: 10.13328/j.cnki.jos.006882]

吴信东(1963－), 男, 博士, 教授, 博士生导师, 主要研究领域为数据挖掘, 大数据分析, 知识工程.

吴共庆(1975－), 男, 博士, 教授, 主要研究领域为数据挖掘, 网络智能.

卓兴锐(1998－), 男, 博士生, 主要研究领域为知识图谱, 大语言模型.

张赞(1987－), 男, 博士, 副教授, 主要研究领域为因果推理, 弱监督学习.

常永洋(2000－), 男, 硕士生, 主要研究领域为数据挖掘, 多标签学习.

朱毅(1984－), 男, 博士, 副教授, 主要研究领域为自然语言处理, 知识图谱, 个性化推荐.