



# 计算机科学

COMPUTER SCIENCE

## 基于大语言模型的电力知识库智能问答系统构建与评价

张金营, 王天堃, 么长英, 谢华, 柴林政, 刘书恺, 李彤亮, 李舟军

引用本文

张金营, 王天堃, 么长英, 谢华, 柴林政, 刘书恺, 李彤亮, 李舟军. [基于大语言模型的电力知识库智能问答系统构建与评价](#)[J]. 计算机科学, 2024, 51(12): 286-292.

ZHANG Jinying, WANG Tiankun, YAO Changying, XIE Hua, CHAI Linzheng, LIU Shukai, LI Tongliang, LI Zhoujun. [Construction and Evaluation of Intelligent Question Answering System for Electric Power Knowledge Base Based on Large Language Model](#) [J]. Computer Science, 2024, 51(12): 286-292.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

### [基于大语言模型的移动应用可访问性增强方法](#)

Large Language Model-based Method for Mobile App Accessibility Enhancement

计算机科学, 2024, 51(12): 223-233. <https://doi.org/10.11896/jsjcx.240400077>

### [文本人格检测研究综述](#)

Study on Text-based Personality Detection – A Review

计算机科学, 2024, 51(12): 209-222. <https://doi.org/10.11896/jsjcx.240500071>

### [一种基于集成学习的开源许可证检测与兼容性判断的方法](#)

Ensemble Learning Based Open Source License Detection and Compatibility Assessment

计算机科学, 2024, 51(12): 79-86. <https://doi.org/10.11896/jsjcx.231200100>

### [汽车验证电控系统中的测试用例自动生成方法](#)

Automatic Test Case Generation Method for Automotive Electronic Control System Verification

计算机科学, 2024, 51(12): 63-70. <https://doi.org/10.11896/jsjcx.240900093>

### [面向大语言模型的推荐系统综述](#)

Survey of Recommender Systems for Large Language Models

计算机科学, 2024, 51(11A): 240800111-11. <https://doi.org/10.11896/jsjcx.240800111>

# 基于大语言模型的电力知识库智能问答系统构建与评价

张金莹<sup>1</sup> 王天堃<sup>1</sup> 么长英<sup>2</sup> 谢华<sup>2</sup> 柴林政<sup>3</sup> 刘书恺<sup>3</sup> 李彤亮<sup>4</sup> 李舟军<sup>3</sup>

1 国电电力发展股份有限公司 辽宁 大连 116000

2 国能信控互联技术有限公司 北京 100039

3 北京航空航天大学计算机学院 北京 100191

4 北京信息科技大学计算机学院 北京 102206

(jinying.zhang@chnenergy.com.cn)

**摘要** 大语言模型是近年来自然语言处理领域的一个重大突破,已成为该领域研究的一种新范式。在金融、法律等垂直领域,基于 FinGPT,ChatLaw 等垂直领域大模型的智能问答系统,促进了大模型技术在相关领域的学术研究与应用落地。然而,由于电力领域缺乏相关的高质量数据,相关的大模型问答系统的构建工作遇到了较大阻碍。为了构建电力领域的智能问答系统,提出了基于大语言模型的电力知识库智能问答系统 ChatPower。为了确保问答效果,ChatPower 充分利用了电力管理各环节的数据。通过语义化理解,梳理和整合了大量的电力专业知识,精心设计和构建了一个较大规模的电力系统知识库。该知识库覆盖电力相关规章制度、安全生产管理体系以及发电设备故障知识等方面的内容。此外,通过参考检索到的电力知识,ChatPower 显著缓解了问答中存在的模型幻觉问题,并在检索系统中引入了 BM25 检索、向量库检索与重排相结合的方法,有效降低了单纯依赖向量库检索的不准确性。同时,ChatPower 结合基于大模型的提示工程技术,提升了对于规章制度类型问题生成回复的条理性。为了对问答系统进行评价,构建了一个电力知识问答的测试数据集,并对其进行了测试验证,测试结果表明:基于大语言模型的电力知识库问答系统 ChatPower 能够有效提升电力相关知识的检索和问答的准确性。

**关键词:** 大语言模型;知识库问答系统;信息检索;自然语言生成

**中图分类号** TP391

## Construction and Evaluation of Intelligent Question Answering System for Electric Power Knowledge Base Based on Large Language Model

ZHANG Jinying<sup>1</sup>, WANG Tiankun<sup>1</sup>, YAO Changying<sup>2</sup>, XIE Hua<sup>2</sup>, CHAI Linzheng<sup>3</sup>, LIU Shukai<sup>3</sup>, LI Tongliang<sup>4</sup> and LI Zhoujun<sup>3</sup>

1 Guodian Power Development Co., Ltd., Dalian, Liaoning 116000, China

2 Guoneng Xinkong Internet Technology Co., Ltd., Beijing 100039, China

3 School of Computer Science and Engineering, Beihang University, Beijing 100191, China

4 Computer School, Beijing Information Science and Technology University, Beijing 102206, China

**Abstract** Large language model is a major breakthrough in the field of natural language processing in recent years and have become a new paradigm for research in this field. In vertical fields such as finance and law, intelligent question and answering systems based on large models in vertical fields such as FinGPT and ChatLaw have promoted the academic research and application of large model technology in related fields. However, due to the lack of relevant high-quality data in the electric power field, the construction of related large-model question answering systems has encountered great obstacles. In order to build an intelligent question and answering system in the electric power field, an intelligent question and answering system for electric power knowledge base ChatPower based on a large language model is proposed. In order to ensure the Q&A effect, ChatPower fully utilizes data from all aspects of power management, sorts out and integrates a large amount of power professional knowledge through semantic understanding, and carefully designs and constructs a large-scale power system knowledge base. The knowledge base covers power-related rules and regulations, production safety management systems, and power generation equipment failure knowledge. In addition, by referring to the retrieved electricity knowledge, ChatPower significantly reduces the problem of model illu-

到稿日期:2024-03-14 返修日期:2024-05-15

基金项目:国家自然科学基金(62276017);软件开发环境国家重点实验室自主研究项目(SKLSDE-2021ZX-18)

This work was supported by the Natural Science Foundation of China(62276017) and State Key Laboratory of Complex & Critical Software Environment(SKLSDE-2021ZX-18).

通信作者:李舟军(lizj@buaa.edu.cn)

sion in question and answering, and introduces a method that combines BM25 retrieval, dense retrieval and rerank in the retrieval system, effectively reducing the the inaccuracy of relying solely on vector library retrieval. At the same time, ChatPower combines prompt engineering technology based on large models to improve the orderliness of generating responses to rules and regulations type questions. In order to evaluate the Q&A system, a test data set for electric power knowledge question and answering is constructed, and ChatPower is tested and verified. The test results show that the electric power knowledge base question and answering system ChatPower based on a large language model can effectively improve the accuracy of retrieval of power-related knowledge and Q&A.

**Keywords** Large language model, Knowledge base question answering system, Information retrieval, Natural language generation

## 1 引言

随着信息技术的迅猛发展,大数据时代的到来使得人们面临着前所未有的信息规模。在专业领域中,各类知识不断涌现,而获取、管理和应用这些知识变得愈发复杂。在这一背景下,构建一个高效且智能的专业知识问答系统,成为了解决专业领域知识管理难题的迫切需求。

Molla 等<sup>[1]</sup>将问答系统(Question and Answering, QA)定义为一个能回答任意自然语言形式问题的自动机。其输入是自然语言形式的问题,输出是一个简洁的答案或者可能答案的列表。问答系统主要有两种形式:基于文本的问答和基于知识库的问答。其中,基于知识库的问答系统(Knowledge Base Question Answering, KBQA)可以极大地提高问答系统解决专业领域知识管理的能力。它的核心是知识库,这是一种结构化的信息集合,通常以三元组的形式表示,如三元组(中国,首都,北京)描述了“中国的首都是北京”这个知识。一些大规模的知识库,如 Freebase<sup>[2]</sup>、DBPedia<sup>[3]</sup>和 Wikidata<sup>[4]</sup>,已经服务许多下游任务。目前,在工业垂直领域,基于知识库的智能问答系统已经得到广泛的应用,例如 Facebook 社交知识图谱<sup>[5]</sup>、Amazon 商品知识图谱<sup>[6]</sup>、Bing 的网络搜索图谱<sup>[7]</sup>等。

传统的基于知识库的问答系统可以很好地完成对简单问题的回答,但对于复杂问题的处理能力有限。过去的研究总结了一系列方法来解析复杂问题,如基于模版匹配的 QUINT<sup>[8]</sup>和 BB-KBQA<sup>[9]</sup>模型、基于语义解析的 Sequence-to-Action<sup>[10]</sup>模型和基于深度学习的 embedding<sup>[11]</sup>模型。这些研究虽然都取得了一定成效,但是仍然存在准确率不高、通用性不强等问题。

而随着大语言模型的崛起,智能问答领域发生了巨大的变革。大语言模型在问答系统等任务上展现出强大的潜能,其可以使计算机更好地理解 and 生成自然语言,这有助于增强问答系统对用户提出的复杂问题的理解能力。目前基于大语言模型的智能问答系统已经取得巨大的进步。在金融、法律等垂直领域,基于 FinGPT<sup>[12]</sup>、Chatlaw<sup>[13]</sup>等垂直领域大模型的智能问答系统促进了大模型技术在相关领域的应用落地,为领域专业知识的智能化处理提供了强大的支持。

然而,在电力领域,我们面临着挑战,即缺乏高质量的相关专业数据,这给构建相应的大模型问答系统带来了较大的阻碍。

为了应对电力领域中知识处理的独特挑战,本文提出了一种创新性的解决方案——基于大语言模型的电力知识库智能问答系统,我们称之为 ChatPower。通过深度理解语义

信息,ChatPower 能够充分利用电力管理各环节的数据,以梳理和整合大量的电力专业知识,从而确保问答的准确性和实用性。我们精心设计和构建了一个庞大的电力系统知识库,该知识库覆盖了电力相关规章制度、安全生产管理体系以及发电设备故障排查等多方面的内容。

与此同时,为了解决问答系统中存在的模型幻觉问题,ChatPower 采用了一系列有效的技术手段。通过引入 BM25 检索、向量库检索与重排相结合的方法,有效降低了单纯依赖向量库检索的不准确性。此外,ChatPower 还结合基于大模型的提示工程技术,显著提升了对规章制度类型问题的回答条理性。

在实验阶段,本文构建了一个电力知识问答的测试数据集,并对 ChatPower 进行了测试验证。测试结果表明:基于大语言模型的电力知识库问答系统 ChatPower 能够有效提升电力相关知识的检索和问答的准确性,为电力制度管理流程的优化提供了有力支持。通过 ChatPower 的研发与应用,我们在电力领域智能化方面取得了较好成绩,为未来的智能化电力管理进行了有益的探索。

## 2 相关工作

本章将依次介绍大语言模型和知识库问答技术的现有方法,这些方法都已取得良好的效果,对相关领域的发展起到了重要推动作用,对本文的工作也有着很强的借鉴意义。

### 2.1 大语言模型

虽然传统的预训练语言模型(PLM),如 BERT<sup>[14]</sup>,通过微调的方法已经在许多的自然语言处理任务中取得了良好的效果。但为了更好地挖掘预训练语言模型的能力,研究人员尝试扩展预训练语言模型的规模,且发现随着模型规模的增大,模型在解决一系列下游的复杂任务时展现了更加出色的能力<sup>[15]</sup>,研究人员将这些大规模预训练语言模型命名为大语言模型(LLM)。大语言模型一般指包含百亿、千亿(或更多)级参数的 Transformer<sup>[16]</sup>语言模型,这些模型除了会在大规模文本数据上进行预训练,还会通过微调与人类价值进行对齐<sup>[17]</sup>。目前常见的大语言模型包含 GPT3.5、GPT-4<sup>[18]</sup>、LLaMA-2<sup>[19]</sup>和 ChatGLM-2<sup>[20]</sup>等。

大语言模型展现了强大的自然语言理解能力和解决复杂任务的能力,许多传统的自然语言处理任务在大语言模型的帮助下正在变得更加简化,且取得了更好的效果。以信息抽取任务为例,Wei 等<sup>[21]</sup>提出了基于 LLM 的信息抽取模型 ChatIE,其使用了上下文学习<sup>[22]</sup>和思维链(COT)<sup>[23]</sup>两种技术,在几个数据集上(例如 NYT11-HRL)的

性能超过了监督模型。

尽管大语言模型在传统自然语言处理任务上表现出色,但其也面临着如幻觉生成<sup>[24]</sup>等问题,即生成的信息与现有来源冲突(内在幻觉)或无法通过现有来源进行验证(外部幻觉)。为了缓解这个问题,研究人员提出了如 Chain-of-Thought (CoT), Tree of Thoughts (ToT)<sup>[25]</sup>, Graph-of-Thought(GoT)<sup>[26]</sup>和 Program of Thoughts(PoT)<sup>[27]</sup>等解决方案,在一定程度上减轻了这些幻觉,但仍会发生事实错误。因此,基于大模型的知识库智能问答是解决大语言模型幻觉问题的有效途径。

## 2.2 知识库问答

虽然传统的问答系统的技术发展已经取得较大的成就,但是由于知识覆盖的限制,传统的问答系统在回答一些领域特定问题时可能会不准确或不完整。近年来,知识库技术的发展引起了众多专家学者的兴趣。知识库(KB)是一个数据库,其中包含了关于特定主题或领域的知识和信息,通常包含结构化和非结构化两种数据存储形式。而知识库问答技术(KBQA)旨在利用预先构建的知识库作为知识来源来回答用户的提问。目前存在两种的主流的技术方案:基于语义解析(SP-based)的方法和基于信息检索(IR-based)的方法。前者通过符号逻辑的形式对问题进行表示,然后通过知识库执行该问题以获得答案。为了解析复杂问题的语法语义信息,Luo等<sup>[28]</sup>通过编码方向依存路径,将句法特征和局部语义特征串联在一起,形成全局的问题表征。Zhu等<sup>[29]</sup>则研究了

查询的结构特性,提出了基于查询-问题匹配的 KBQA 方法。他们使用结构感知编码器对查询中的实体或关系上下文进行建模,以促进查询和问题之间的匹配。基于信息检索的方式主要从知识库中提取问题对应的三元组或文本,并将其作为参考来生成确定答案。同样是为了解决复杂问题,Miller等<sup>[30]</sup>使用键-值记忆网络来实现动态的指令更新,以更好地表示复杂问题的组合语义。Shi等<sup>[31]</sup>则提出了一个用于关系图多跳问答的透明有效框架,在标签形式的 MetaQA 的 2-hop 和 3-hop 问题上实现了 100% 的准确率。

随着预训练语言模型和大语言模型的兴起,大语言模型可以作为知识库问答任务的底座模型。Luo等<sup>[32]</sup>提出基于大模型的生成-检索知识问答框架 ChatKBQA,利用生成再检索代替检索再生成的方式,解决了检索效率低、检索误导生成、KBQA 任务解决方案复杂等痛点,在 WebQSP 和 CWQ 两个知识库问答基准(Benchmark)上均取得了优异的表现。

## 3 电力知识库问答系统 ChatPower

### 3.1 问题描述

首先,给出了知识库问答任务的形式化的描述:给定用户问题  $x$ ,系统  $S$  将根据用户问题  $x$  和知识库  $D$ ,生成该问题对应的答案  $y$ 。

### 3.2 系统框架

如图 1 所示,本文提出的 ChatPower 系统主要由 3 个模块组成:数据处理模块、数据存储模块和人机交互模块。

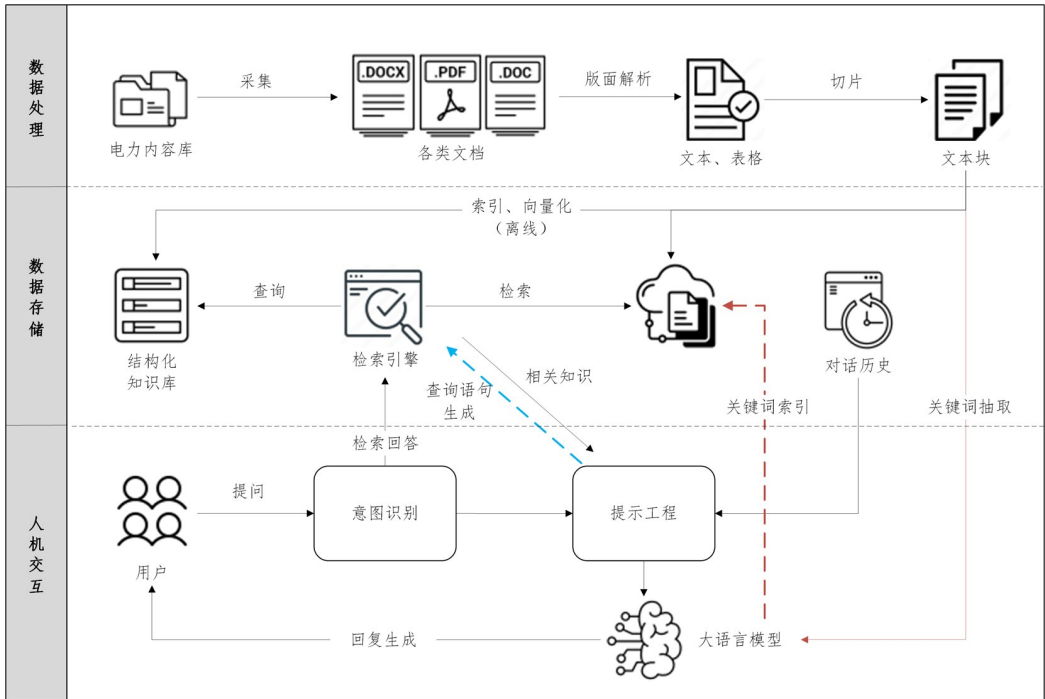


图 1 系统的总体框架

Fig. 1 Overall system framework

### 3.2.1 数据处理模块

数据处理模块的主要功能是对电力企业内的相关电力领域文档进行处理,以构建电力领域知识库。首先该模块对文档系统中的文档进行筛选,选出包含电力相关内容的各类文档(包含 txt, word, pdf 以及 excel 等格式);然后结合版面解析技术,

提取文档中的内容,提取的内容主要包含表格和文本两类。

由于文档内容长度差别较大,较长文档无法直接输入到大模型中,因此需要对文档内容进行切片处理。模块将根据文档的布局和段落信息以及预设的片段最大长度对文档内容进行切分。

### 3.2.2 数据存储模块

数据存储模块的主要功能是存储已整理的电力知识以及多轮的对话记录,并在进行问答时完成对相关知识的检索与整合。

电力知识的存储方式包含结构化和非结构化两种。对于表格类型数据,本系统将其存储在数据库中,通过查询语句完成对表格中相关数值的检索以及分析计算型的检索。对于文本型数据,系统将其存储在非结构化数据库中。

### 3.2.3 人机交互模块

人机交互模块即知识库问答的入口,用户在向系统进行提问后,意图识别模块对用户问题类别进行判别,包含4个类别:直接生成、非结构化查询、结构化查询以及制度条例查询。

1)直接生成:根据用户的问题,通过提示工程模块直接构造相应的提示,并输入到大模型,由大模型生成问题的回复。

2)非结构化查询:检索引擎根据用户的问题,检索与该问题相关的文本片段;然后提示工程模块根据用户问题以及检索到的文本片段,构造相应的提示;最后,将其输入到大模型中进行相关问题的回答。

3)结构化查询:该类别针对结构化知识库的问答。首先,提示工程模块根据用户的问题,构造生成 SQL 查询语句的指令;然后大模型生成相应 SQL 语句;接着检索引擎执行查询语句得到相应的数据;最后,提示工程模块根据用户问题和查询到的结果,构造相应的提示,输入到大模型以生成相应的回复。

4)制度条例查询:该类别根据用户问题对制度知识库进行检索,并设计相应的指令来引导大模型有条理地生成相应的制度解释和回答。

### 3.3 系统实现

#### 3.3.1 电力知识库构建

为确保 ChatPower 的问答效果,结合电力管理的各环节数据,通过语义化理解,梳理和整合了大量电力专业知识,构建了大规模的电力知识库。该知识库覆盖电力相关规章制度、安全生产管理体系以及发电设备故障知识等方面。

本文采集了电力相关文档 173 篇,并对文档进行了归类,主要包含电力系统相关规章制度、安全生产管理和发电设备使用说明 3 方面的文档。文档的具体数据如表 1 所列。

表 1 知识库文档统计信息

Table 1 Knowledge base document statistics

文档类别	文档数	总页数
规章制度	10	1023
安全生产管理	68	859
发电设备使用	95	738

通过版面解析算法,对这些文档进行结构化解析,将其中的文本进行切片得到文本块,用于构建非结构化知识库;对其中的表格进行抽取,用于构建结构化知识库。

#### 3.3.2 数据存储系统构建

数据存储系统的整体架构如图 2 所示,主要包含索引和检索两部分。本文基于 Elastic-Search 框架和 MySQL 框架分别构建了非结构化知识存储系统和结构化知识存储系统。如图 2 所示,数据存储部分主要分为离线索引部分和在线检索部分。在离线检索部分,系统分别将经过数据处理模块处理得到的非结构化的文档片段和表格数据录入到非结构化数据库和结构化数据库中。

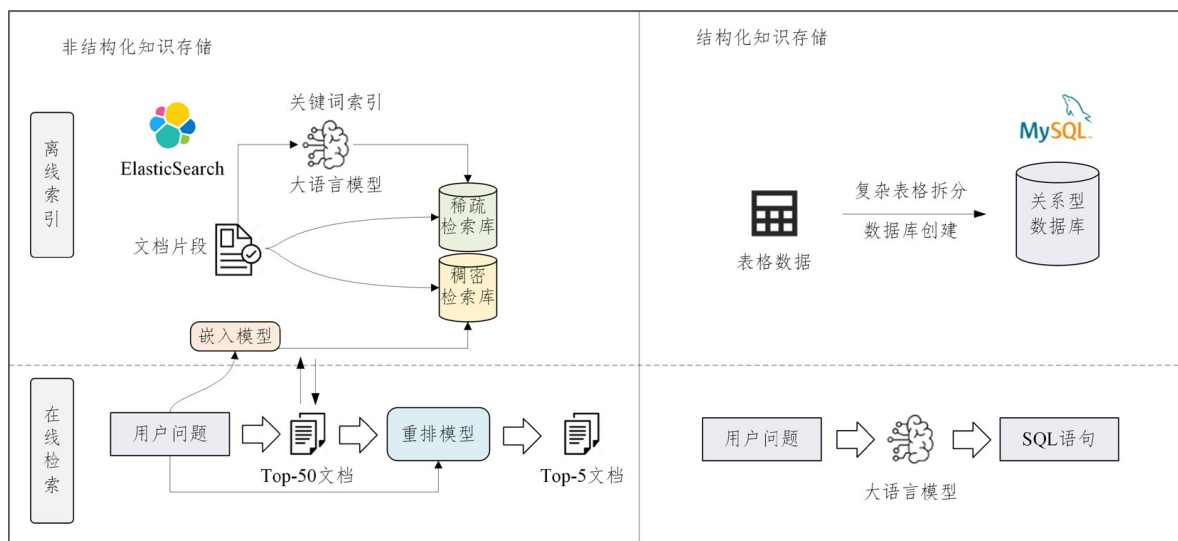


图 2 知识库构建与检索流程

Fig. 2 Process of knowledge base construction and retrieval

非结构化数据库的索引采用了稀疏检索和稠密向量检索相结合的方式。稀疏检索向量库的构建采用了 BM25 算法和关键词检索相结合的方式。

BM25 索引是一种词袋检索功能,它根据每个文本片段中出现的查询语句的频数对一组文档进行排名,而不管它们在文本片段中的接近程度如何。给定用户问题  $Q$ , 包含词  $q_1, \dots, q_n$ , 一个文本片段  $D$  的 BM25 分数为:

$$score(D, Q) = \sum_{i=1}^n IDF(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)} \quad (1)$$

其中,  $f(q_i, D)$  是  $q_i$  在文本片段  $D$  中出现的频数;  $|D|$  是文本片段  $D$  的长度;  $avgdl$  表示文本片段集合的平均长度;  $k_1$  和  $b$  是超参数;  $IDF(q_i)$  是每个词  $q_i$  的逆文档频率, 其

计算式如式(2)所示:

$$IDF(q_i) = \ln\left(\frac{N - n(q_i) + 0.5}{n(q_i) + 0.5} + 1\right) \quad (2)$$

其中,  $N$  表示文本片段集合的总数量;  $n(q_i)$  表示包含  $q_i$  的文档数。

关键词索引过程中, 系统借助大模型生成每一个文本片段  $D$  中的关键词  $Key_D$  来更加简要地表征文本片段的内容。

稠密索引过程中, 系统使用了 BGE Embedding 模型对每个文本片段  $D$  进行编码, 得到每个文本片段对应的向量  $v_D$ 。用户问题  $Q$  通过 BM25 检索得到每一个文本片段对应的相关分数  $score_{bm25}(D, Q)$ 。同时将用户问题通过句子向量模型进行编码得到向量  $v_Q$ , 然后  $v_Q$  与所有文本片段对应的向量  $v_D$  通过内积计算相关分数  $score_{dense}(D, Q)$ 。根据两种检索方法, 计算混合分数:

$$score_{hybrid} = \beta \cdot score_{bm25}(D, Q) + (1 - \beta) \cdot score_{dense}(D, Q) \quad (3)$$

根据混合分数, 选取分数最高的 50 个文本片段。

上述检索过程中, 算法更加侧重于共现词汇, 对于语义的相关性的检索能力较差, 因此本文使用 ROM 语义相关性模型对检索到的文本片段进行重排。将问题与 50 个文本片段分别输入 ROM 中, 计算相关性分数, 选取分数最高的 5 个文本片段。

对于结构化数据库的索引, 首先将表格进行预处理, 对于普通表格, 直接构造相应的数据库表格; 对于存在合并列的表格, 对表格进行拆分, 重组成多个表格后再录入到数据库中。

对于结构化数据库的检索, 首先使用大模型将用户问题转化为 SQL 语句, 然后通过查询语句查询相应的数据库得到相应的答案。

### 3.3.3 人机交互流程构建

本文假设用户的问题可以分为两大类: 第一类问题为非结构化问答, 问答可以直接回答或者相应答案包含在非结构化知识库中; 第二类问题对应的答案包含在结构化知识库中。因此, 系统首先需要对用户的意图进行识别, 从而决定回答流程所使用的流程。针对电力所处的场景, 本文设计了直接生成、非结构化查询、结构化查询以及制度条例查询 4 类意图以及相应的问答流程。用户进行提问后, 首先经过基于大模型的意图识别模块进行意图的识别, 然后根据识别的意图来选择不同的答案生成方式。

1) 直接生成。用户的提问不需要进行相关电力知识的检索, 可以直接经过提示工程构造相应的指令后回答。

2) 非结构化查询。用户问题首先通过混合检索的方式从非结构化知识库中检索和问题最相关的  $k$  个文本片段。然后问题与最相关的文本片段通过提示工程模块生成指令, 将其输入到大语言模型中生成回复返回给用户。

3) 结构化查询。用户问题首先通过混合检索的方式从结构化知识库中检索和问题最相关的  $k$  个表格。然后根据问题最相关的表格信息通过提示工程模块生成 SQL 语句生成指令, 将其输入到大语言模型中生成相应的 SQL 语句。接着, 再根据查询语句去关系数据库中查询相应的结果。如果能够

查询到结果, 则根据查询结果结合问题再次经过提示工程模块构造回复生成的指令, 将其输入到大模型中生成相应的回复; 如果不能查到结果或者查询语句查询失败, 则通过大模型生成无法回答的回复返回给用户。

4) 制度条例查询。由于电力系统中存在部分条例类型的问题, 使用普通的指令构造方法通常无法使大模型稳定生成按点列出的相关条例结果。因此, 本文对条例类的查询的指令构造进行了优化, 使得大模型可以稳定输出条例结果。

具体实现上, 系统实现环境为 Linux 下的 Centos7, CPU 为 IntelXeon(R) CPU E5-2698 v4 @ 2.20GHz, 显卡为 A100 40GB×2, 内存为 512GB。在大模型的选择上, 本文综合考量了模型在中文任务上的表现, 如模型推理速度以及安全性, 选择了通义千问的 Qwen-14b-int4 模型。此外, 本文结合开源框架 FastChat 对大模型进行部署, 并在无网环境中进行了测试, 确保数据的安全性。

## 4 实验验证

### 4.1 数据集

本文根据表 1 所列文档, 构建了一个电力知识问答的测试数据集, 其中包括非结构化问答和结构化问答两大类。非结构化问答包含直接生成、非结构化查询和制度条例查询 3 种具体的查询类型; 非结构化问答包含制度条例查询。具体的测试数据集信息如表 2 所列。

表 2 数据集问题统计

Table 2 Dataset question statistics

问答类型	具体类别	问题数
非结构化问答	直接生成	20
非结构化问答	非结构化查询	50
结构化问答	结构化查询	50
非结构化问答	制度条例查询	50

### 4.2 参数设置

在 BM25 算法中,  $k_1$  设置为 5,  $b$  设置为 1,  $\beta$  设置为 0.6。大模型的生成中, 采样算法为贪心采样。

### 4.3 评价指标

本文使用准确率来评价问答系统的性能。

$$Acc = \frac{N_{correct}}{N_{total}}$$

其中,  $N_{correct}$  表示系统能够正确回答的问题数量,  $N_{total}$  表示所有问题数量。

### 4.4 实验结果及分析

表 3 和表 4 分别列出了本系统在非结构化知识库问答以及结构化知识库问答上的表现。其中, BM25 表示 BM25 稀疏索引算法; Keyword 表示关键词检索算法; Dense 表示稠密检索算法; Rerank 表示精排算法。

表 3 非结构化问答实验结果

Table 3 Experimental results of unstructured question answering (%)

Methods	Acc
BM25	63.3
BM25+Keyword	68.3
BM25+Keyword+Dense	72.5
BM25+Keyword+Dense+Rerank	<b>79.2</b>

表4 结构化问答实验结果

Table 4 Experimental results of structured question answering

Methods	Acc (%)
BM25	52.0
BM25+Keyword	54.0
BM25+Keyword+Dense	60.0
BM25+Keyword+Dense+Rerank	<b>68.0</b>

实验结果表明,在BM25检索的基础上,关键词、向量检索以及重排能够有效提升检索的准确率,从而提升大模型最终生成的答案的准确性。其中,精排算法的引入能够有效提升知识召回的质量,从而有效提升问答的准确率。在非结构化问答以及结构化问答任务上,本系统分别达到了79.2%以及68.0%的准确率,能够满足电力相关的基本问答需求。

**结束语** 本文系统地介绍了基于大语言模型的电力知识库智能问答系统的构建方法,并给出了其中算法部分的实验验证。在面向特定的垂直专业领域时,首先应梳理和整合相应领域的专业数据,构建相应的专业知识库;然后根据领域的特点,设计知识检索和人机交互模块的问答流程;最后,可以参考本文提供的系统框架,实现一个面向垂直专业领域的知识库智能问答系统。

## 参 考 文 献

- [1] MOLLA D, VICEDO J L. Question Answering in Restricted Domains: An Overview[J]. Computational Linguistics, 2007, 33(1): 41-61.
- [2] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: A collaboratively created graph database for structuring human knowledge[C]// Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD). Canada: ACM, 2008: 1247-1250.
- [3] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia: A large-scale, multilingual knowledge base extracted from Wikipedia[J]. Semantic Web, 2015, 6(2): 167-195.
- [4] PELLISSIER T T, VELIČKOVIĆ D, SCHAFFERT S, et al. From Freebase to Wikidata: The Great Migration[C]// Proceedings of the International World Wide Web Conference (WWW). Canada: ACM, 2016: 1419-1428.
- [5] NOY N F, GAO Y, JAIN A, et al. Industry-scale Knowledge Graphs: Lessons and Challenges [J]. ACM Queue, 2019, 17(2): 20.
- [6] ARUN K. Making search easier: How Amazon's Product Graph is helping customers find products more easily. Amazon Blog. [EB/OL]. <https://blog.aboutamazon.com/innovation/making-search-easier>.
- [7] SAURABH S. Bring rich knowledge of people, places, things and local businesses to your apps [EB/OL]. <https://blogs.bing.com/search-quality-insights/2017-07/bring-rich-knowledge-of-people-places-things-and-local-businesses-to-your-apps>.
- [8] ABUJABAL A, YAHYA M, RIEDEWALD M, et al. Automa-

ted template generation for question answering over knowledge graphs[C]// Proceedings of the International World Wide Web Conference (WWW). Perth: ACM, 2017: 1191-1200.

- [9] LIU A, HUANG Z, LU H, et al. BB-KBQA: BERT-based knowledge base question answering [C]// Proceedings of the 18th China National Conference on Chinese Computational Linguistics. Kunming: Springer, 2019: 81-92.
- [10] CHEN B, SUN L, HAN X P. Sequence-to-Action: End-to-end semantic graph generation for semantic parsing [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: ACL, 2018: 766-777.
- [11] BORDES A, CHOPRA S, WESTON J. Question answering with subgraph embeddings [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014: 615-620.
- [12] YANG H, LIU X Y, DAN W C. FinGPT: Open-Source Financial Large Language Models [C]// FinLLM Symposium at IJCAI 2023. 2023.
- [13] CUI J, LI Z, YAN Y, et al. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases [J]. arXiv: 2306.16092, 2023.
- [14] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of Deep bidirectional Transformers for Language Understanding [J]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, 1(2): 4171-4186.
- [15] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners [EB/OL]. [https://cdn.openai.com/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf).
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). Montreal, Canada: NIPS, 2017: 5998-6008.
- [17] OpenAI. Our approach to alignment research [EB/OL]. <https://openai.com/blog/our-approach-to-alignment-research>.
- [18] OPENAI, JOSH A, STEVEN A, et al. GPT-4 Technical Report [J]. arXiv: 2303.08774, 2023.
- [19] TOUVRON H, LOUIS M, KEVIN S, et al. Llama 2: Open foundation and fine-tuned chat models [J]. arXiv: 2307.09288, 2023.
- [20] ZENG A, LIU X, DU Z X, et al. Glm-130b: An open bilingual pre-trained model [J]. arXiv: 2210.02414, 2022.
- [21] WEI X, CUI X Y, CHENG N, et al. Zero-shot information extraction via chatting with chatgpt [J]. arXiv: 2302.10205, 2023.
- [22] BROWNT B, MANNB, RYDERN, et al. Language models are few-shot learners [C]// Advances in Neural Information Processing Systems. Curran Associates, Inc, 2020: 1877-1901.
- [23] WEI J, WANG X Z, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models [C]// Advances in Neural Information Processing Systems. Curran Associates, Inc, 2022: 24824-24837.
- [24] WELLECK S, KULIKOV I, ROLLER S, et al. Neural text ge-

- neration with unlikelihood training[C]// International Conference on Learning Representations. Addis Ababa: OpenReview.net;2020.
- [25] YAO S Y, YU D, ZHAO J, et al. Tree of thoughts: Deliberate problem solving with large language models[J]. arXiv: 2305.10601, 2023.
- [26] YAO Y, LI Z C, ZHAO H. Beyond Chain-of-Thought: Effective Graph-of-Thought Reasoning in Large Language Models[J]. arXiv: 2305.16582, 2023.
- [27] CHEN W H, MA X G, WANG X Y, et al. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks[J]. arXiv: 2211.12588, 2022.
- [28] LUO K, LIN F, LUO X, et al. Knowledge base question answering via encoding of complex query graphs[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018; 2185-2194.
- [29] ZHU S, CHENG X, SU S. Knowledge-based question answering by tree-to-sequence learning[J]. Neurocomputing, 2020, 372: 64-72.
- [30] MILLER A H, FISCH A, DODGE J, et al. Key-value memory networks for directly reading documents[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: The Association for Computational Linguistics, 2016: 1400-1409.
- [31] SHI J X, CAO S L, HOU L, et al. TransferNet: An effective and transparent framework for multi-hop question answering over relation graph[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana: The Association for Computational Linguistics, 2021: 4143-4158.
- [32] LUO H R, EH H, TANG Z C, et al. ChatKBQA: A Generate-then-Retrieve Framework for Knowledge Base Question Answering with Fine-tuned Large Language Models[J]. arXiv: 2310.08975, 2023.



**ZHANG Jinying**, born in 1984, Ph.D. His main research interests include artificial intelligence, distributed control systems, and fieldbus control systems.



**LI Zhoujun**, born in 1963, Ph.D, professor, is a member of CCF(No. 06618S). His main research interests include artificial intelligence, natural language processing, network and information security.

(责任编辑:喻黎)