

融合知识图谱和大语言模型的罗氏沼虾养殖问答系统的设计与实现

郑豪¹, 杨国伟¹, 李飞², 郭建林²

(1. 杭州电子科技大学通信工程学院, 浙江 杭州 310018;

2. 浙江省淡水水产研究所, 浙江 湖州 313001)

摘要: 互联网存储着大量罗氏沼虾的相关信息, 借助网络搜索、大语言模型 (large language model, LLM) 很难高效获取养殖问题的准确解答。基于实体粒度的属性拼接向量检索策略, 设计并实现了一个面向罗氏沼虾养殖行业的智能问答系统。该系统主要通过融合知识图谱和大语言模型来实现罗氏沼虾养殖领域的友好问答。首先, 借助 LLM 来完成知识抽取功能, 将收集到的原始数据转化为结构化数据, 构建一个罗氏沼虾养殖相关的知识图谱, 将图谱中实体名称与其属性拼接形成语义丰富的描述文本, 进行向量化表示并导入 Neo4j。然后, 通过 LLM 搭建一个问题过滤模块, 对用户提出的问题先进行语义优化和分类后进行向量化处理, 从而提升噪声环境下的问答鲁棒性。最后, 通过语义向量检索方式, 将检索到的知识传递给 LLM 进行专业回答, 从而实现融合知识图谱和 LLM 的智能问答系统。实验测试展示了所提方法可以有效提高回答的准确性和鲁棒性。该系统有望助力于提供高效、准确、友好的罗氏沼虾专业信息检索, 以期提升养殖决策效率和增加养殖生产效益。

关键词: 知识图谱; 大语言模型; 罗氏沼虾养殖; 智能问答

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202532

Design and implementation of an intelligent question answering system for *Macrobrachium rosenbergii* aquaculture based on knowledge graph and large language model

ZHENG Hao¹, YANG Guowei¹, LI Fei², GUO Jianlin²

1. School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

2. Zhejiang Institute of Freshwater Fisheries, Huzhou 313001, China

Abstract: As a vast amount of information about *Macrobrachium rosenbergii* is available on the Internet, it makes the use of web searches or general-purpose large language model (LLM) inefficient to obtain accurate answers to aquaculture-specific questions. An intelligent question answering (QA) system for the *Macrobrachium rosenbergii* aquaculture field based on an entity-level attribute-concatenated vector retrieval strategy was proposed and implemented. A knowledge graph with LLM was integrated to support professional, user-friendly QA. Firstly, raw, unstructured aquaculture data was transformed into structured triples by an LLM-based knowledge extraction module to construct a knowledge graph. The entity names and their attributes were concatenated into semantically rich textual descriptions, which were vectorized and stored in a Neo4j database for similarity-based retrieval. Then, a question filtering module based on an LLM was introduced to optimize and classify user queries before vectorization, thereby enhancing answer robustness in noisy or ambiguous input environments. Finally, semantic vector retrieval was performed over the knowledge graph, and the retrieved knowledge was passed to an LLM for generating professional, context-aware responses. Experimental results demonstrate

收稿日期: 2025-04-07; 修回日期: 2025-06-26

通信作者: 杨国伟, guowei.yang@hdu.edu.cn

基金项目: 浙江省“三农九方”科技协作计划项目 (No.2023SNJF073)

Foundation Item: “Three Rural and Nine Party” Science and Technology Cooperation Plan of Zhejiang Province (No.2023SNJF073)

that the proposed approach effectively improves response accuracy and robustness. The system shows promise in enabling efficient, accurate, and user-friendly knowledge access in aquaculture, potentially improving decision-making and enhancing production outcomes.

Key words: knowledge graph, large language model, *Macrobrachium rosenbergii* aquaculture, intelligent question answering

0 引言

罗氏沼虾凭借其肉质优良、生长速度快和市场价格较高,逐渐成为全球范围内广泛养殖的重要淡水虾品种之一^[1]。罗氏沼虾的需求在全球市场中持续增长,成为许多地区的重要经济来源^[2]。然而,罗氏沼虾的养殖涉及多项复杂技术,养殖者在进行信息获取时常面临两大问题:一方面,信息来源分散、缺乏系统性,导致答案多样且表述不一;另一方面,养殖者在表达问题时常语义模糊,使得基于规则的检索系统难以准确响应。这不仅影响养殖效率,也提高了决策失误的风险,从而影响经济效益。

针对上述问题,知识图谱与大语言模型(large language model, LLM)的结合为养殖知识获取提供了新的技术路径。一方面,知识图谱作为一种系统性、结构化的知识表达方式,可以有效整合和关联分散的数据,为特定领域构建知识网络^[3-4]。另一方面,作为自然语言处理领域的前沿技术,LLM不仅能够提取文本的语义特征,还可以根据输入的问题生成专业的回答,为用户提供实时的、针对性的指导,为智能化问答系统提供了全新的可能性^[5]。

为此,本文提出并实现一个融合知识图谱与LLM的罗氏沼虾养殖智能问答系统,旨在为垂直领域提供专业、交互式知识服务。本文通过设计一种基于实体粒度拼接属性的向量化检索方法,从非结构化文本中抽取实体及属性信息,将其拼接为语义完整的文本单元;在文本单元与用户问题向量化后,通过语义相似度匹配对应实体节点,将匹配结果作为Prompt输入LLM生成自然语言回答。该方法结合了知识图谱的结构优势与LLM的语义理解能力,既摆脱了结构化查询对用户表达方式的限制,又优化了传统向量化检索语义一致性不足的局限。本文主要开展以下3个方面的工作。

(1) 构建罗氏沼虾养殖知识图谱,覆盖罗氏沼虾相关的核心内容,形成结构化、可拓展的知识网络。

(2) 设计实体粒度拼接属性的向量化检索机制,提升系统对自然语言问题的语义匹配能力,实

现知识图谱内容与用户问题之间的高效语义匹配。

(3) 实现融合知识图谱与LLM的智能问答系统,对比常见垂直领域的大模型问答系统,验证了本文方法在准确性、鲁棒性方面的优势。

1 相关工作

面向垂直领域的问答系统是人工智能(artificial intelligence, AI)的重要研究方向,在早期,知识图谱问答系统通常通过抽取实体和关系构建图谱,用户输入的问题经语义解析生成结构化查询语句,再从图数据库中检索答案。由于其结构清晰、回答准确,知识图谱问答系统在医疗、金融、农业等领域得到广泛应用^[6]。然而,这类系统高度依赖模板或规则,难以处理语言多样性强、语义含糊或模糊的问题,且自然语言交互能力有限,难以满足复杂交互式问答需求^[7]。

2022年,OpenAI推出的预训练生成式大语言模型ChatGPT引起了广泛关注,紧随其后许多LLM开始涌现。LLM通过在大规模通用语料上的预训练,具备了强大的语言理解和生成能力,能够在多轮对话、复杂推理、开放问答等任务中表现出色,这使LLM成为各类智能辅助决策、问答系统的核心技术^[8-9]。尽管LLM在大部分通用场合取得了不错的进展,但其应用于垂直领域仍面临诸多挑战^[10]。一方面,通用LLM训练语料多来自开放文本,对专业术语与特定领域知识的覆盖有限;另一方面,垂直领域知识具有较强的专业性和复杂性,通用模型往往难以准确理解和应用这些知识。

为缓解上述问题,Lewis等^[11]提出了检索增强生成(retrieval-augmented generation, RAG)技术。该方法将用户输入向量化,并在外部文档集合中检索语义相关文本,作为上下文输入大语言模型以生成回答,从而提升了回答准确性。RAG架构融合了信息检索与语言生成的优势。然而,该技术难以充分利用领域知识的结构化特点,文本切片可能造成语义割裂,且难以合理控制切片粒度,容易导致上下文缺失或冗余,影响答案的连贯性和准

确性^[12]。

近年来,研究者开始尝试将结构化的知识图谱与大语言模型(LLM)结合,以弥补传统RAG方法在结构化知识应用方面的不足。为此,KG-RAG(knowledge graph retrieval-augmented generation)方法应运而生。例如,Soman等^[13]利用生物医学知识图谱SPOKE,提取与问题相关的图谱信息,并自动生成优化后的提示词,帮助模型更好地理解生物学术语和复杂语义。该方法不仅减少了无关内容的干扰,还提升了回答的准确性和效率。Sanmartín^[14]也提出了类似方法,将知识图谱中的结构信息与生成模型结合,使模型在复杂问答任务中更有效地使用领域知识,回答更加连贯可靠。尽管KG-RAG在垂直领域问答表现方面取得了一定进展,但仍存在一些局限性。首先,知识图谱中的实体和关系与自然语言输入之间的对齐存在较大难度,尤其在该领域口语或术语较多的情况下,导致匹配不到或匹配错误。其次,不支持表达复杂意图,且系统整体流程构建复杂度高。

本文提出了一种融合知识图谱和大语言模型的罗氏沼虾养殖问答系统,基于实体粒度拼接属性的向量化知识检索策略,既避免了文本切片导致的语义切割,又利用了领域知识的结构化特点。首先,从垂直领域文档中抽取实体及其属性信息,构建结构化的知识图谱;随后,以实体为单位,将其名称及所有属性和属性值拼接成文本,并作为检索单元

进行向量化存储。相较于以整段或整句为单位的传统方式,这种“实体粒度”策略更能精准命中与用户问题相关的信息,减少冗余干扰,提升语义匹配精度。

2 系统框架概述

本文设计与开发的罗氏沼虾养殖问答系统,旨在满足该领域的专业知识管理与问答需求。本文系统采用智谱AI和清华大学发布的ChatGLM3-6B对话预训练模型作为系统进行回答时的基座模型^[15],实现从数据处理、知识建模、语义理解到自然语言交互的完整流程。罗氏沼虾养殖问答系统流程如图1所示,整体工作流程分为5个阶段。

(1) 原始数据处理与知识图谱构建:将收集的罗氏沼虾养殖相关文档转为标准文本格式,利用OneKE知识抽取框架提取实体与属性,生成结构化数据,构建涵盖生物学特性、养殖方法、病害等主题的知识图谱。采用Neo4j图数据库进行存储与管理,支持可视化展示与查询。

(2) 实体粒度拼接与向量化表示:以图数据库中每个实体为单位,将其名称、属性及属性值拼接为语义完整的描述文本,使用jina-embeddings-v3模型进行向量化,构建语义增强的向量知识库。

(3) 问题处理与过滤:用户通过前端输入问题,系统对其进行语义优化,并判断其是否与罗氏沼虾养殖相关。若相关,则进行向量化处理;否则

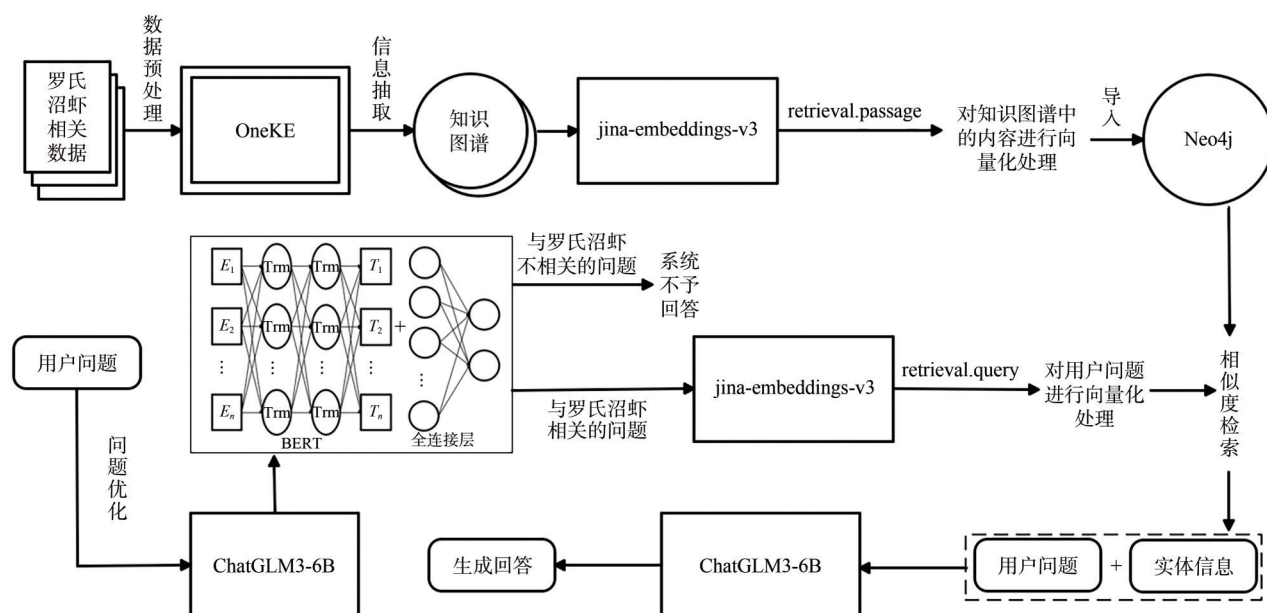


图1 罗氏沼虾养殖问答系统流程

提示用户重新输入。

(4) 相似度检索：计算问题向量与实体向量的语义相似度，提取 Top- K 相关实体作为检索结果。

(5) Prompt 构造与回答：将用户问题与检索到的信息拼接构建 Prompt，输入 ChatGLM3-6B 模型生成回答。

3 系统构建方法

3.1 罗氏沼虾知识图谱构建

罗氏沼虾知识图谱构建流程如图2所示。

3.1.1 数据获取与知识建模

数据获取与预处理是构建罗氏沼虾养殖知识图谱的关键环节。本文从多个来源获取了结构化和非结构化的数据，包括水产养殖专业书籍，如《罗氏沼虾健康养殖技术》《罗氏沼虾养殖技术》等；学术文章，如《罗氏沼虾生物学研究》《罗氏沼虾的营养需要研究》《罗氏沼虾主要病害研究概况》等；以及互联网上的公开信息，如百度百科、维基百科。收集的数据可能包含重复、不完整或错误的信息，因此需要对数据进行清洗，包括去除重复项、填充缺失值、校正错误数据等。对于非结构化的文本数据，统一转化成文本形式，以便为后续的信息抽取做准备。

在知识建模过程中，本文以“生物学特性”“养殖方式”和“疾病”三大主题为主线，提取出7个核心实体类别，分别为物种、生长阶段、养殖模式、管理措施、饲料、疾病、药物，并针对不同实体类别，定义了多个不同属性，实体类别及属性见表1。

表1 实体类别及属性

实体	实体描述	属性
物种	罗氏沼虾作为知识中心节点	学名、生长周期、繁殖方式等
生长阶段	生长过程的不同阶段分类	持续天数、生长速度、饲养重点等
养殖模式	不同的养殖技术方案	适应区域、经济效益、水体深度等
管理措施	管理和操作行为	操作内容、执行频率、建议条件等
饲料	常见饲料种类	蛋白含量、适用阶段、喂养方式等
疾病	常见疾病种类	病因、传播途径、典型症状等
药物	用于疾病治疗防控	治疗疾病、推荐剂量、副作用等

在实体之间，本文设计了6种关系类别以保持结构一致性和语义连贯性，分别为：经历阶段、可食用、推荐投喂、包含、推荐治疗药物、常见疾病。关系类别见表2。

表2 关系类别

关系	头实体	尾实体
经历阶段	物种	生长阶段
可食用	物种	饲料
推荐投喂	生长阶段	饲料
包含	养殖模式	管理措施
推荐治疗药物	疾病	药物
常见疾病	物种	疾病

3.1.2 信息抽取

信息抽取是自然语言处理领域的一个重要分支，是实现从文本到结构化数据转化的核心步骤，其目的是从非结构化或半结构化的文本中自动提取结构化信息，使其以实体、关系或事件的形式呈现，如实体名称、实体之间的关系等^[16]。

随着深度学习技术的发展，信息抽取性能得到了显著提升。本文基于 OneKE 进行罗氏沼虾知识

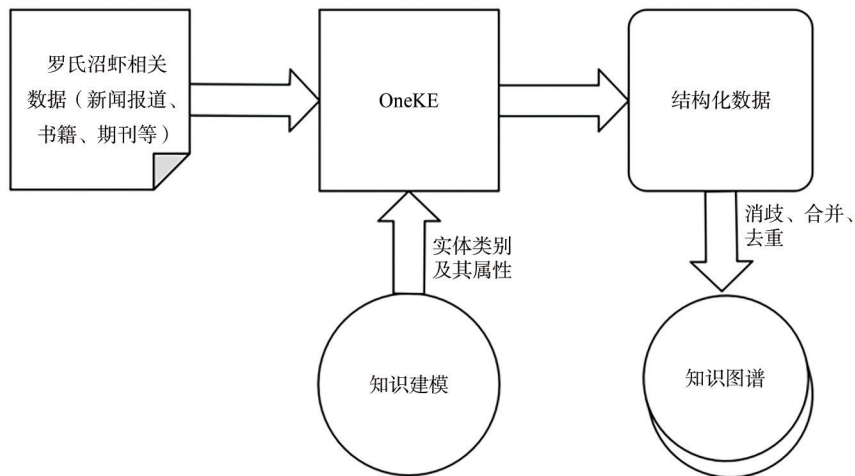


图2 罗氏沼虾知识图谱的构建流程

图谱的信息抽取。OneKE是由蚂蚁集团和浙江大学联合研发的大模型知识抽取框架，具备中英文双语、多领域多任务的泛化知识抽取能力，并提供了完善的工具链支持^[17]。OneKE通过不同定制指令来进行不同抽取任务，指令格式采用了类JSON字符串的结构，它由以下3个字段构成：（1）"instruction"，即任务描述，以自然语言指定模型扮演的角色以及需要完成的任务；（2）"schema"，是一份需提取的标签列表，明确指出了待抽取信息的关键字段，反映用户需求，是动态可变的；（3）"input"，指的是用于信息抽取的源文本。

首先，定义好"instruction"，指定对实体和属性的抽取，并要求以JSON格式的数据返回抽取完成的结果；然后，根据不同实体和其对对应属性来定义"schema"，一共制定了7条指令；最后，对每段文本执行这些指令即可得到所需JSON格式的结构化数据。OneKE信息抽取见表3。

3.1.3 数据处理和导入

在完成实体与属性的结构化抽取后，为高效、规范地构建知识图谱，需要进一步对抽取结果进行系统化的数据处理与整理，并按照Neo4j图数据库的要求批量导入。抽取得到的结果为多条结构化的JSON数据，每条记录包含一个实体的类型、名称及其对应的属性信息。由于同一实体可能在不同文本中以不同形式出现，且属性信息存在冗余或不一致的情况，需对抽取结果进行进一步的规范化处理，以确保知识图谱数据的一致性和完整性^[18]。本部分主要包括实体标准化、属性合并、关系三元组构建及图数据库加载4个步骤。

首先进行的是实体标准化处理。为了便于后续的数据去重与统一管理，本文将抽取出的实体按照类型分类保存，分别建立7类核心实体的文件，包括

“物种.txt”“生长阶段.txt”“养殖模式.txt”“管理措施.txt”“饲料.txt”“疾病.txt”和“药物.txt”。每类实体文件采用3列结构：“实体类别 实体名称 实体ID”。对所有实体进行统一编号，并通过字符串匹配与人工审核相结合的方式去重和同义归一处理。例如，“白斑病”与“白斑综合征”在确认语义一致的前提下归为同一实体，分配相同ID。

其次是属性合并与整理。抽取阶段的每一条记录通常同时包含实体与属性，但若将属性直接存入实体文件中，可能导致大量重复实体及属性混乱。因此，本文将属性信息独立保存，建立7类属性文件，每个文件存储某一类实体的所有属性数据。属性文件采用“实体ID 属性名 属性值”的3列结构，便于对相同ID的多个属性进行合并和去重。同时，对于多值属性以列表形式存储。

接下来是关系三元组的构建。基于预定义的6种实体关系类型来进行构建，关系文件格式采用标准三元组结构，形式为：“实体1 ID 关系类型 实体2 ID”，后续通过实体ID对照实体文件进行匹配，确保关系中的实体已在图谱中定义。

最后将所有清洗后的数据导入Neo4j图数据库。为此，本文使用Python进行数据处理和导入，基于py2neo库操作Neo4j的节点和关系。导入过程分两个阶段：首先批量创建实体节点，并附加其对应的属性信息；然后读取关系文件，创建对应的关系边。导入完成后，需对知识图谱进行初步查询和验证，确保数据正确连接和展示。

3.2 问答系统的构建

3.2.1 向量化模块

向量化是提升问答系统检索效率的关键步骤^[19]。传统的知识图谱问答依赖Cypher等图查询语言，存在对关键词匹配依赖强、缺乏语义理解、

表3 OneKE信息抽取

阶段 信息抽取	
输入	"白斑病主要由嗜水气单胞菌引起,常见于罗氏沼虾的幼虾阶段,表现为虾体白斑、活动迟缓,易通过水体传播。通常推荐使用氟苯尼考和水杨酸钠进行治疗,春季是该病的高发季节。"
指令	"instruction": "你是一个图谱实体知识结构化专家。根据输入实体类型(entity type)的 schema 描述,从文本中抽取相应的实体实例和其属性信息,不存在的属性不输出,属性存在多值就返回列表,并输出为可解析的 json 格式。" "schema": [{ "entity_type": "疾病", "attributes": ["病原体","症状","传播途径","易感阶段","推荐药物","发病季节"]}]
输出	{ "疾病": { "白斑病": { "病原体": "嗜水气单胞菌", "症状": ["虾体白斑","活动迟缓"], "传播途径": "水体传播", "易感阶段": "幼虾阶段", "推荐药物": ["氟苯尼考","水杨酸钠"], "发病季节": "春季" } }

扩展性差等问题。例如，用户查询“罗氏沼虾的饵料成分”时，若图谱中使用的是“饲料成分”，尽管语义相近，仍可能因关键词不一致而导致查询失败。构建同义词库可在一定程度上缓解此类问题，但该方法依赖大量人工收集和分类词汇，难以覆盖所有表达方式，尤其在专业术语丰富或涉及新词、地方用词时，语义匹配问题仍难以解决。

向量化技术有效解决了传统方法的局限。近年来，基于神经网络的词嵌入模型，如 Word2Vec 和 GloVe^[20-21]，将词汇映射到固定维度的向量空间，语义相近的词距离更近，能够捕捉词义并处理文本多样性。当前主流方法采用 LLM 进行向量化，通过对上下文建模生成句子或段落级别的高维向量，显著提升语义表示的准确性^[22]。然而，在实际应用中，向量化效果受文本切割粒度影响。传统密集向量化检索常以文档、段落或句子为单位，不同粒度对检索性能和问答效果影响显著。粗粒度切割包含更多上下文，但易引入噪声，降低语义匹配准确性；细粒度切割信息简洁但稀疏，难以表达复杂语义，导致信息不完整。主流向量知识库框架如 LangChain+FAISS 通常采用固定窗口切片策略，可能破坏上下文连续性，导致召回文段冗余或偏离问题意图，影响回答效果。

本文提出基于实体粒度的属性拼接向量化检索策略。该策略以实体为单位，将所有属性及属性值拼接为完整描述文本，整体向量化表示。相比传统段落切片，能完整保留完整语义，避免信息割裂。本文系统采用基于 Jina-XLM-RoBERTa 架构的 jina-embeddings-v3 多语言文本嵌入模型^[23]，对知识图谱内容采用 retrieval.passage 方式向量化，对用户问题采用 retrieval.query 方式。相似度计算采用余弦相似度，召回结果设定为 Top-K=3，以平衡精度与效率，供后续问答生成模块处理。向量化模

块执行流程如图3所示。

3.2.2 问题过滤模块

在罗氏沼虾养殖问答系统，问题过滤模块可以保证系统仅处理与罗氏沼虾相关的问题，从而限制 LLM 的回答范围，减少不相关输入对系统的干扰和“幻觉事实”的产生^[24]。系统通过将输入问题的全集 Q 限制到罗氏沼虾养殖问答系统的领域问题集合 R，确保只有与罗氏沼虾相关的问题进入问答流程，提升整体的准确性与鲁棒性。

本文系统采用的 BERT 模型为经过中文语料训练的预训练模型，它能够有效地处理中文语句的上下文信息，并能为后续分类任务提供特征表示^[25]。具体来说，BERT 模型用于提取问题的语义信息，生成 [CLS] token 的输出向量，该向量凝聚了整个句子的语义表示。在 BERT 模型的输出层后，添加一层全连接层来执行二分类任务，将 BERT 模型的输出映射为“相关”和“无关”两个类别，问题过滤模块网络结构如图4所示。全连接层的输入为 BERT 模型输出的 [CLS] token 的向量化表示，输出为两个类别的概率分布。最终，模型根据输出的概率值，选择最可能的标签作为问题的分类结果。训练目标是最大化正确标签的概率，并通过交叉熵损失函数进行优化：

$$P(L|H) = \text{softmax}(WH) \quad (1)$$

其中， L 表示分类任务的标签，即问题类别； H 为 BERT 模型输出的 [CLS] token 的向量化表示； W 为全连接层的权重矩阵。为了训练问题过滤模块，笔者收集了大量与罗氏沼虾养殖相关和无关的中文问句数据，数据来源包括相关书籍、学术文章等。这些数据经过标注，分为“相关”和“无关”两类，构成了模型的训练和测试数据集。

为了解决养殖用户在提出问题时常存在“表达模糊”“语义不清”等问题，本文在传统 BERT 分

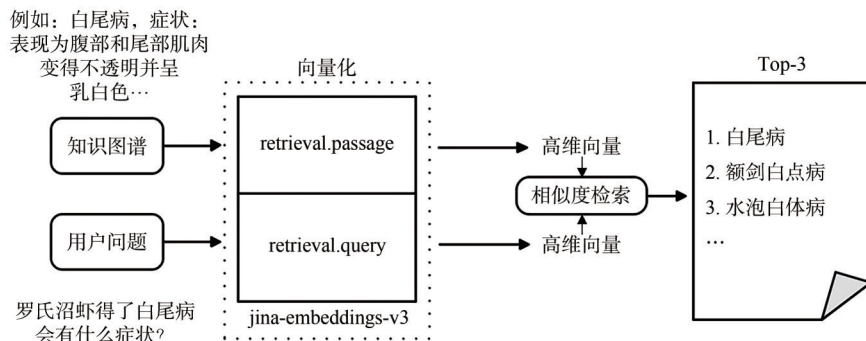


图3 向量化模块执行流程

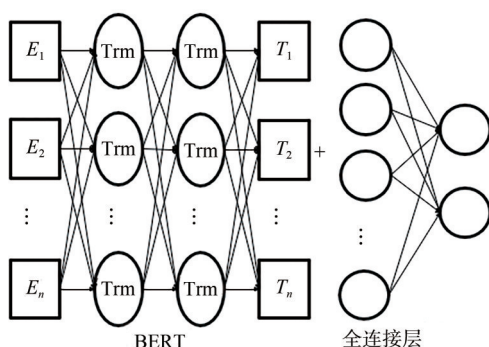


图4 问题过滤模块网络结构

类器的基础上引入 LLM，先对问题优化预处理。具体来说，系统在接收到用户输入的原始问题后，首先将其输入 ChatGLM3-6B 基座模型中，由模型生成语义更加明确、结构更合理的优化版本。该优化过程不仅保留原始意图，还能消除常见的语言噪声，如冗余词汇、省略表达、口语化描述等。将 LLM 优化后的问题再传入 BERT 分类器进行分类，可以带来两方面好处：一方面，有效提升了 BERT 分类器对边界样本和模糊问题的判别能力，使系统对隐含但相关的问题表达更加敏感，从而减少错误分类的概率；另一方面，由于优化后的问题结构清晰、语义聚焦，也为后续的向量化表示和相似度检索提供了更可靠的输入语料，提升检索精度。

问题过滤模块执行流程如图5所示，具体步骤如下阐述。

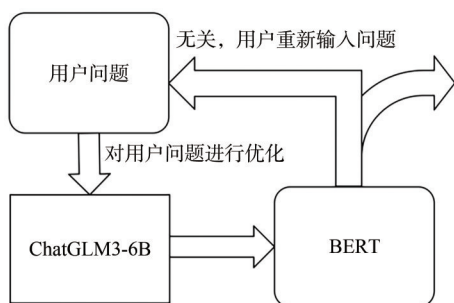


图5 问题过滤模块执行流程

(1) 问题优化：用户输入的问题先经过 LLM 进行优化。

(2) 问题处理：将优化后的问题通过 BERT 进行分类和编码。

(3) 特征提取：BERT 模型将这些编码后的数据输入，并生成 [CLS] token 的特征向量。

(4) 分类判断：该向量被传入全连接层，计算出问题属于“相关”或“无关”的概率。

(5) 输出结果：如果问题被判定为“相关”，则将该问题传递给后续的知识图谱查询和专业回答生成模块；若问题被判定为“无关”，则系统返回提示，要求用户重新提问或选择相关问题。

4 实验分析

问答系统演示如图6所示。



图6 问答系统演示

4.1 专业问答评估

本节实验主要评估本文系统在垂直领域内的专业问答能力，收集并整理了有关罗氏沼虾的 50 道选择题和 50 道简答题，题目由水产所专家参考实际养殖需求设计，包括基础生物学知识、养殖管理、疾病等多个维度。对比模型包括罗氏沼虾养殖问答系统、ChatGLM3-6B 以及“ChatGLM3-6B+向量知识库”（使用 Milvus 检索并拼接 Top-3 片段作为提示输入）。

在选择题的验证试验中，通过计算系统回答正确的概率，评估系统的客观性能。准确率以正确问题个数除以总问题个数进行计算。在简答题的验证试验中，采用 ROM-Base 和 coROM 两种语义相关性模型^[26]。ROM-Base 模型是一个通用领域的中文语义相关性模型，它能够量化两段文本之间的相似度，其值介于 0~1，数值越接近 1 表示文本间的语义相似度越高。coROM 模型则基于 Dual Encoder 框架，通过计算两段文本之间的余弦距离来评估它们的相似性，值越大表明相似性越高。问答系统的回答输入这两种模型，对比不同系统在各自指标的平均数表现，并分析 ROM-Base 输出的相似度指标超过 50% 的比例 R_{50} ，问答系统演示结果见表 4。

表4 专业问答评估

系统	准确率	余弦距离 平均数	相似度 平均数	R_{50}
罗氏沼虾养殖问答系统	0.86	0.897	0.707	86%
ChatGLM3-6B+向量知识库	0.82	0.883	0.689	82%
ChatGLM3-6B	0.7	0.842	0.534	62%

4.2 抗噪声干扰评估

为了评估系统在处理含噪声干扰（如错别字、口语化表述、语序颠倒等）时的鲁棒性，利用ChatGPT对4.1专业问答评估中的原题进行修改，构造了与原题目语义等价但形式干扰较强的“扰动问题”，输入系统进行问答测试。本文主要通过准确率和 R_{50} 来评估性能。对比模型包括罗氏沼虾养殖问答系统、“ChatGLM3-6B+向量知识库”系统，结果见表5。

表5 抗噪声干扰评估

系统	准确率	R_{50}
罗氏沼虾养殖问答系统	0.8	84%
ChatGLM3-6B+向量知识库	0.64	76%
ChatGLM3-6B	0.7	60%

4.3 总结

专业问答评估结果表明，罗氏沼虾养殖问答系统在准确率、余弦距离平均数、相似度平均数及 R_{50} 这4个关键指标上，整体表现略优于“ChatGLM3-6B+向量知识库”系统，且显著优于原始ChatGLM3-6B模型。相较于通用LLM或结合向量知识库的方案，本文系统在针对罗氏沼虾垂直领域的专业问答任务中，具备更强的语义理解能力与信息召回能力。

抗噪声干扰评估结果表明，针对准确率、 R_{50} 这两个指标关键上，引入噪声干扰后，罗氏沼虾养殖问答系统与ChatGLM3-6B的变化不大，“ChatGLM3-6B+向量知识库”系统的波动比较明显。一方面说明通用LLM本身的抗噪声能力良好；另一方面说明在问题过滤模块中结合LLM对用户问题进行语义优化后，能在一定程度上提高检索准确性，从而避免LLM在面对无关、冗余的上下文输入时，回答质量降低的情况。

虽然本文系统在专业领域和鲁棒性方面取得较好的表现，但在响应时间上，由于系统包含问题过滤、向量化检索与大模型多次调用等处理环节，整

体响应时间相对较长。相比于ChatGLM-6B模型的即时响应能力，本文系统存在一定延迟。未来将通过引入向量缓存机制与异步处理框架来优化系统性能。

5 结束语

本文设计并实现了融合知识图谱和LLM的罗氏沼虾养殖智能问答系统，借助知识图谱的专业性与大语言模型的灵活性，显著提升了在垂直领域问答中的准确性和实用性。借助问题过滤模块改善“幻觉事实”问题，确保了回答内容的相关性与可信度。罗氏沼虾疾病、养殖等方面的选择题与简答题的实验测试结果表明，融合知识图谱和LLM来进行问答，能够在一定程度上提高LLM对垂直领域的理解和应用能力，验证了在罗氏沼虾养殖方面智能问答的性能与优势，能够满足行业用户的实际需求。未来工作中，可以进一步扩展知识图谱的覆盖范围、提高系统对问题推理的能力与响应速度，以提升智能问答系统在农业与水产养殖领域的应用价值。

参考文献：

- [1] 黄玉玲. 我国虾类淡水养殖概述[J]. 水利渔业, 2003, 24(1): 1-3.
HUANG Y L. Summary of freshwater shrimp culture in China[J]. Reservoir Fisheries, 2003, 24(1): 1-3.
- [2] 杨国梁, 陈雪峰, 王军毅, 等. 罗氏沼虾产业在中国持续增长的经济与社会原因分析[J]. 浙江海洋学院学报(自然科学版), 2011, 30(5): 450-457.
YANG G L, CHEN X F, WANG J Y, et al. Social and economical factors of sustained growth of macrobrachium rosenbergii industry in China's mainland[J]. Journal of Zhejiang Ocean University (Natural Science), 2011, 30(5): 450-457.
- [3] 刘翥, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.
LIU Q, LI Y, DUAN H, et al. Knowledge graph construction techniques[J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.
- [4] 张吉祥, 张祥森, 武长旭, 等. 知识图谱构建技术综述[J]. 计算机工程, 2022, 48(3): 23-37.
ZHANG J X, ZHANG X S, WU C X, et al. Survey of knowledge graph construction techniques[J]. Computer Engineering, 2022, 48(3): 23-37.
- [5] ZHAO W X, ZHOU K, LI J, et al. A survey of large language models[EB]. 2023.
- [6] 高锐涛, 林达伟, 郭亮, 等. 基于知识图谱的水稻种植智能问答系统设计与实现[J]. 计算机工程, 2024, 50(12): 133-141.
GAO R T, LIN D W, GUO L, et al. Design and implementation of rice planting intelligent question-answering system based on knowledge graph[J]. Computer Engineering, 2024, 50(12): 133-141.
- [7] YANI M, KRISNADHI A A. Challenges, techniques, and trends of simple knowledge graph question answering: a survey[J]. Information,

- 2021, 12(7): 271.
- [8] 李炎英, 王新宇, 王晓, 等. 基于大语言模型的交通异常事件检测与辅助决策[J]. 智能科学与技术学报, 2024, 6(3): 347-355.
LI Y Y, WANG X Y, WANG X, et al. Traffic anomaly event detection and auxiliary decision-making based on large language models[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(3): 347-355.
- [9] 张志成, 王静, 张阳, 等. OrthoGPT: 面向精准诊疗的多模态骨科大模型[J]. 智能科学与技术学报, 2024, 6(3): 338-346.
ZHANG Z C, WANG J, ZHANG Y, et al. OrthoGPT: multimodal generative pre-trained transformer models for precise diagnosis and treatment of orthopedics[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(3): 338-346.
- [10] LING C, ZHAO X J, LU J Y, et al. Domain specialization as the key to make large language models disruptive: a comprehensive survey[EB]. 2023.
- [11] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[J]. Advances in Neural Information Processing Systems, 2020, 33: 9459-9474.
- [12] CHEN T, WANG H W, CHEN S H, et al. Dense X retrieval: what retrieval granularity should we use?[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2024:15159-15177.
- [13] SOMAN K, ROSE P W, MORRIS J H, et al. Biomedical knowledge graph-optimized prompt generation for large language models[J]. Bioinformatics, 2024, 40(9): btac560.
- [14] SANMARTIN D. KG-RAG: bridging the gap between knowledge and creativity[EB]. 2024.
- [15] ZENG A H, XU B, WANG B W, et al. ChatGLM: a family of large language models from GLM-130B to GLM-4 all tools[EB]. 2024.
- [16] 李保利, 陈玉忠, 俞士汶. 信息抽取研究综述[J]. 计算机工程与应用, 2003, 39(10): 1-5.
LI B L, CHEN Y Z, YU S W. Research on information extraction: a survey[J]. Computer Engineering and Applications, 2003, 39(10): 1-5.
- [17] ZHANG N Y, XU X, TAO L K, et al. DeepKE: a deep learning based knowledge extraction toolkit for knowledge base population[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, December 2022, Abu Dhabi, UAE. Stroudsburg: Association for Computational Linguistics, 2022: 98-108.
- [18] HOGAN A, BLOMQUIST E, COCHEZ M, et al. Knowledge graphs [J]. ACM Computing Surveys (CSUR), 2021, 54(4): 1-37.
- [19] WANG Q, MAO Z D, WANG B, et al. Knowledge graph embedding: a survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [20] CHURCH K W. Word2Vec[J]. Natural Language Engineering, 2017, 23(1): 155-162.
- [21] PENNINGTON J, SOCHER R, MANNING C. Glove: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL Press, 2014: 1532-1543.
- [22] REIMERS N, GUREVYCH I. Sentence-BERT: sentence embeddings using Siamese BERT-networks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, Hong Kong, China. Stroudsburg: Association for Computational Linguistics, 2019: 3982-3992.
- [23] STURUA S, MOHR I, AKRAM M K, et al. Jina-embeddings-v3: multilingual embeddings with task LoRA[EB]. 2024.
- [24] 张鹤译, 王鑫, 韩立帆, 等. 大语言模型融合知识图谱的问答系统研究[J]. 计算机科学与探索, 2023, 17(10): 2377-2388.
ZHANG H Y, WANG X, HAN L F, et al. Research on question answering system on joint of knowledge graph and large language models[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(10): 2377-2388.
- [25] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Stroudsburg: ACL Press, 2019: 4171-4186.
- [26] QIU Y F, LI H Y, QU Y Q, et al. DuReader-retrieval: a large-scale chinese benchmark for passage retrieval from web search engine[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, Abu Dhabi, United Arab Emirates. Stroudsburg: Association for Computational Linguistics, 2022: 5326-5338.

[作者简介]



郑豪 (2001-), 男, 杭州电子科技大学通信工程学院硕士生, 主要研究方向为知识图谱、大语言模型等。



杨国伟 (1984-), 男, 博士, 杭州电子科技大学通信工程学院副教授, 主要研究方向为神经网络、物联网技术等。



李飞 (1986-), 男, 浙江省淡水水产研究所副研究员、渔业经济与大数据中心主任, 主要研究方向为水产动物繁殖育种与养殖、渔业数据分析和规划编制等。



郭建林 (1981-), 男, 浙江省淡水水产研究所高级工程师、副所长, 主要研究方向为智慧渔业、水产营养等。