

文章编号: 2095-2163(2023)07-0168-05

中图分类号: TP391.1

文献标志码: A

面向多类型知识库的多引擎智能问答系统设计

李春豹

(中国电子科技集团公司第十研究所, 成都 610041)

摘要: 针对现有问答系统存在的问题理解不清、问答引擎单一、答案溯源困难等不足, 该文设计了一种面向多类型知识库的多引擎智能问答系统。根据用户数据特点将其组织存储为问答对库、文档库、图谱库和结构化数据库, 并分别设计对应的智能问答引擎, 结合复杂问题理解与意图识别分类结果和问答引擎选择策略, 完成答案及其来源获取。此外, 系统还支持基于用户评价反馈的自优化, 在使用中不断提升问答效果。在构建的新冠疫情数据集上进行了实验验证, 结果表明, 该系统能够准确理解用户输入问题并快速精准给出答案, 提升了用户信息获取效率。

关键词: 多类型知识库; 智能问答系统; 意图识别分类; 答案溯源

Design of multi-engine intelligent question-answering (QA) system for multi-type knowledge bases

LI Chunbao

(The 10th Research Institute of China Electronics Technology Group Corporation, Chengdu 610041, China)

【Abstract】 In order to solve the problems of the existing QA system, such as unclear understanding of question, single QA engine and difficult tracing of answers, this paper designs a multi-engine intelligent QA system for multi-type knowledge bases. According to the characteristics of the data, it is organized and stored as frequently asked question library, document library, knowledge graph library and databases, and the corresponding intelligent QA engine is designed respectively. By combining the results of complex question understanding, intention classification and the engine selection strategy, the final answer and its source are obtained. In addition, the system also supports automatic optimization based on the user's evaluation and feedback, and constantly improves the performance of QA in use. Experimental verification is carried out on the constructed COVID-19 dataset, and in the results, it is demonstrated that the system can accurately understand complex questions and give answers quickly, improving the efficiency of user information acquisition.

【Key words】 multi-type knowledge base; QA system; intent recognition classification; answer tracing

0 引言

随着信息技术的飞速发展, 海量的互联网信息不断涌现, 无论是数据样式、还是数据种类都变得繁多, 给互联网用户带来更加丰富多样信息的同时加大了用户信息获取的难度。为了应对传统检索方式存在检索结果仍需人工筛选、研读、答案确认等耗时耗力的问题, 与此同时进一步提高用户准确获取信息的效率, 智能问答系统应运而生。

根据应用场景的不同, 现有智能问答系统类型主要包括电商客服类、闲聊陪护类^[1]、虚拟助理类、检索式问答^[2]和综合问答类。针对军事领域特定场景数据, 诸如 IBM watson、Wolfram Alpha 等的综合问答类系统已经取得了一些成果, 但是距离实际

使用仍然有一定差距^[3-4], 现有研究的不足主要体现在以下 3 个方面:

(1) 复杂问题理解不清晰: 用户输入问题时存在描述不完整、描述冗余、字词错误等情况, 而现有系统大多未进行针对性处理导致问题理解偏差或错误, 进而返回错误答案。

(2) 现有系统大多是针对某一类特定类型数据进行设计: 例如, 面向用户常见问题对形式的智能客服, 还有一系列针对知识图谱、数据库表或自由文档而提出的问答系统, 然而, 实际应用场景中的数据往往是以多种类型存储。

(3) 答案来源不可追溯: 现有智能问答系统仅返回用户输入问题对应答案, 用户验证给定答案的正确性比较困难。

作者简介: 李春豹(1989-), 男, 博士, 工程师, 主要研究方向: 自然语言处理、视觉目标跟踪。

收稿日期: 2023-01-10

为了应对上述问题,本文设计了一种面向多类型知识库的多引擎智能问答系统。在充分理解用户输入复杂问题的基础上,针对特定领域的多来源、多类型数据分别设计智能问答引擎以分别获取用户输入问题的候选答案,完成候选答案筛选排序,以文本、图像和可视化图谱等方式展示最终答案。同时,将答案获取过程和答案内容在原文中位置进行可视化,让用户更直观、更便捷地分析返回答案的准确性和全面性。

1 面向多类型知识库的多引擎智能问答系统设计

面向多类型知识库的多引擎智能问答系统的整体流程如图1所示。由图1可知,针对用户输入的自然语言描述问题,首先进行复杂问题理解,结合文本描述、外部业务知识、外部业务常识知识、上下文对话信息以及问题所属主题等对问题进行深度剖析,合理补充用户输入问题并抽取用户问题中的核心要素,进而对该问题的真实意图进行准确识别、理解与分类。然后,基于理解补充后的问题、用户问题意图图分类结果和业务问答知识库构建多引擎智能问答模型,完成面向不同类型知识库的候选答案快速获取,结合特定规则完成最终答案及其来源的获取,同时为了方便用户进行答案溯源,从用户问题输入、问题理解、问答引擎答案获取到答案组织的全过程进行展示。此外,还支持用户对给定答案进行评分、反馈与入库,并基于入库语料对问答模型进行自优化,持续提升问答系统的性能。

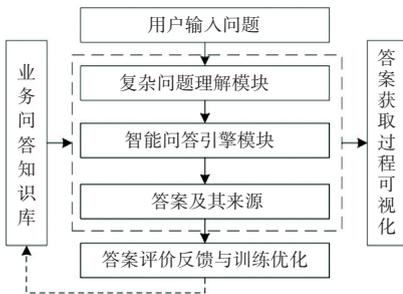


图1 面向多类型知识库的多引擎智能问答系统流程图

Fig. 1 Flowchart of the multi-engine intelligent QA system for multi-type knowledge bases

1.1 复杂问题理解设计

复杂问题理解的目标是在补充问题的同时结合问题中的核心要素识别用户问题真实意图,辅助多引擎智能问答模型的问答策略选择,其详细流程如图2所示。

首先,对用户输入的文本问题进行字词纠错、事

实性纠错与语法纠错。其中,字词纠错采用面向大规模字词纠错语料集的 Seq2Seq 深度卷积神经网络模型^[5]完成问题中多字、漏字、错字等错误的纠正。事实性纠错则是采用基于外部常识知识库的事实检测方法对问题中偏离常识知识的错误进行纠正。语法纠错则是采用基于语义搭配的错误检测方法完成问题中语义搭配错误检测和纠正。

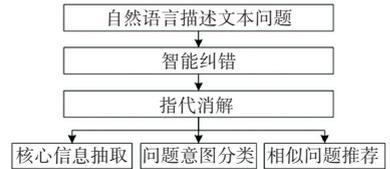


图2 复杂问题理解流程图

Fig. 2 Flowchart of the complex question understanding

当用户输入问题中出现“他”、“她”、“它”、“呢”等代词和语气助词或者省略实体等情况时,还需要结合上文对话信息采用基于规则的方法进行指代消解与问题补充,得到的新问题作为复杂问题理解的输出,同时采用基于 Bert 预训练模型对问题中的时间、国家、人物、军种、专用名词等核心实体进行识别与抽取。针对军事领域数据,将用户问题意图划分为事实描述类、属性查询类、统计计算类和分析总结类,并将其设置为意图分类结果标签,标注训练数据,对基于 Bert 的 TextCNN 分类算法^[6]进行训练后获取用户问题的意图分类结果,用于指导多引擎智能问答模型的问答策略选择。此外,复杂问题理解还提供了相似问题和热点问题推荐功能,其中相似问题推荐结果是基于 BM25 和语义相似度算法计算用户输入问题与历史问题库的相似度排序得到,热点推荐问题结果则是根据近一周、一月和近三月的问答匹配次数排序得到。

1.2 多引擎智能问答模型设计

面向多来源、多类型问答知识库构建多引擎智能问答模型,多个引擎面向不同类型知识库分别设计问答算法,包括问答对库问答算法、阅读理解问答算法、知识图谱问答算法和数据库表问答算法,以帮助用户全面、高效、精准获取问题答案、答案置信度及其答案来源。

针对军事领域的用户业务数据,问答知识库按照知识不同类型分别进行存储,以供用户进行提问。知识库类型主要分为以下4类:业务活动中积累的领域常见问题与典型问题及其答案构成问答对库,存储在 Elastic Search 库中;针对领域中的结构化数据,尤其是基础参数、性能、属性等一段时间内

固定不变的静态数据,将其组织为知识图谱,采用 Neo4j 进行存储与展示;一些日常业务活动相关的动态数据,以数据记录的形式存储在 MySQL 或 Oracle 结构化数据库中;而一些非结构化的文档、资

料等构成文档库,存储在 Elastic Search 中。针对上述 4 种不同类型的知识库分别提供最适合该类知识的问答引擎,4 种问答引擎共同构建了多引擎智能问答模型,如图 3 所示。

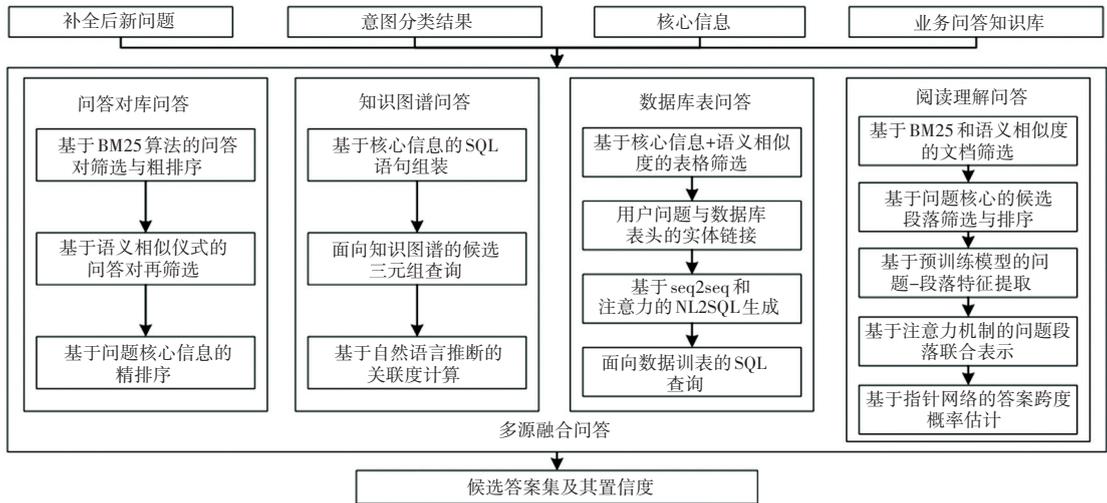


图3 多引擎智能问答模型结构图

Fig. 3 Structure of the multi-engine intelligent QA model

针对问答对库数据,设计了问答对库问答引擎。首先,面向 Elastic Search 库采用 BM25 算法完成问答对的初步筛选,并根据评分完成排序,取前 Top-k 问答对;然后,基于 Simhash 计算文本语义相似度,并根据评分完成排序,取前 Top-n 问答对,完成再筛选;最后,结合问题中抽取的时间、国家、人物、军种或专用名词等核心实体对前 Top-n 问答对问题进行精确匹配,剔除不包含核心实体的问答对以得到最终候选答案集。

针对知识图谱数据,构建了知识图谱问答引擎。首先,根据核心信息中的实体,采用 BM25 算法从 Elastic Search 库中存储的知识图谱实体集中检索出达到指定阈值的相似度最高的实体,若未找到符合要求的实体,则直接向用户推荐相似实体;然后,采用 cypher 语句查询 Neo4j 中对应实体的所有关系;最后,基于自然语言推断计算输入问题与各关系的关联度以得到候选答案三元组结果。其中,自然语言推断采用 LSTM 网络^[7]进行计算,问题和答案整体的显式匹配度由 logistic 回归层计算:

$$r_e = \sigma(W_e \cdot h_q + b_e) \quad (1)$$

其中, W_e 和 b_e 为逻辑斯蒂回归的参数, h_q 为输入问题的浅层语义向量表示。问题和答案的潜层关联度用张量神经网络计算获得:

$$r_i = \sigma(W_i \cdot T(h_q, h_a) + b_i) \quad (2)$$

其中, T 表示张量变换; W_i 和 b_i 为逻辑斯蒂回

归的参数; h_a 为三元组答案的浅层向量表示。

使用动态加权求和的方式从问题和答案的显式匹配度和浅层语义关联度计算其整体的关联度得分:

$$r = g \cdot r_e + (1 - g) \cdot r_i \quad (3)$$

$$g = \sigma(W_g \cdot [h_q, h_a] + b_g) \quad (4)$$

其中, g 表示门限权重,即通过问题和答案的浅层语义向量动态地估计显式匹配度和浅层语义关联度的有效程度。

针对存储在结构化数据库中的数据记录,采用了数据库表问答引擎^[8]。首先,基于语义相似度模型对补全后问题与数据库表名计算最长公共子串,同时与表名-列名组成的语义向量做相似度匹配,获取 Top-k 数据表;然后,将修正后问题与数据表名-列名连接后输入构建的 NL2SQL 模型^[9]中以生成 SQL 查询语句,其中 NL2SQL 模型包括 2 个独立的模型:模型一将问题与表头名字采用标志位分割并连接后输入 Bert-wwm 中文预训练模型进行编码,随后针对聚合操作、组合关系和比较关系三个不同的任务分别连接稠密网络和交叉熵损失函数层,最后通过最小化 3 个任务的损失之和完成模型一的训练;模型二将问题与模型一选择的列输入 Bert-wwm 中文预训练模型进行组合获取候选<列,操作,值>,然后通过一层稠密层对所有候选组合进行分类,选出最终组合后得到 SQL 语句;最后,采用 SQL

语句对数据库表进行查询的结果即为最终答案。

针对自由文档数据,构建了面向多文档的阅读理解问答引擎。首先,面向存储海量非结构化自由文档的 Elastic Search 库,针对补全后新问题完成基于 IK-BM25 算法的检索,获取 top - k 文档数据;然后,结合核心信息中的国家、时间、地点、人物、装备、任务和其他专有实体等进行再筛选,得到 top - n 文档数据;第三,将 top - n 文档按照 512 长度进行分段并与问题连接输入 Bert 中文预训练模型中完成对段落的筛选;第四,将问题与筛选后的段落连接后输入 Bert 中文预训练模型和注意力网络^[10],提取问题-段落的联合特征表示;最后,输入指针网络层估计各段落中答案的开始位置及跨度,并按照评分答案置信度进行排序。

面向多种类型的问答知识库,将补全后问题输入多引擎智能问答模型中,结合意图分类结果自动选择问答引擎,实现候选答案、答案置信度及其答案来源的获取。其中,意图分类结果与问答引擎对应关系:事实描述类问题采用问答对库、知识图谱和阅读理解问答引擎;属性查询类问题采用知识图谱问答和阅读理解问答引擎;数据计算类问题采用数据库表问答和知识图谱问答引擎;统计分析类问题则采用数据库表问答引擎。当采用多问答引擎时,返回更快、更准的问题答案及其置信度,即采用多线程方式同时请求上述不同意图问题对应问答引擎,最先返回满足该引擎方式设定的答案阈值的答案结果即为最终答案。

同时,该系统从用户问题输入、问题理解和问题推荐到问答引擎答案获取和答案组织进行全流程展示。其中,复杂问题理解展示用户输入问题、问题推荐结果、指代消解结果、意图问题结果、核心词及关键短语;问答对库问答展示 Top5 粗检索结果与精排序问题答案对;知识图谱问答展示问题对应图谱中的实体、cypher 查询语句、候选三元组和图谱答案;数据库表问答展示选表结果、SQL 查询语句、答案及答案对应表格记录;阅读理解问答展示文档检索与筛选后 Top5 结果、答案及其对应原文。

1.3 答案评价反馈与训练优化设计

答案评价反馈与训练优化是为了让用户在使用问答系统的同时提升其性能,其流程如图 4 所示。用户可以对系统给定答案进行评分,若评分高于给定阈值(例如,给定阈值三颗星,满分五颗星)或者评分低于给定阈值同时用户对给定问题的答案进行反馈,则将答案及其来源存入问答对库以使用户下

次提问时能给出正确答案,同时存入训练语料库。当用户反馈的语料达到一定数量或者达到指定训练间隔时间时,提醒用户进行问答模型手动训练评估或者根据用户设定完成自动训练评估,且当准确率或 F1 值等评估指标优于线上问答模型时,支持一键上线或自动上线新版本以提供更优的问答服务。

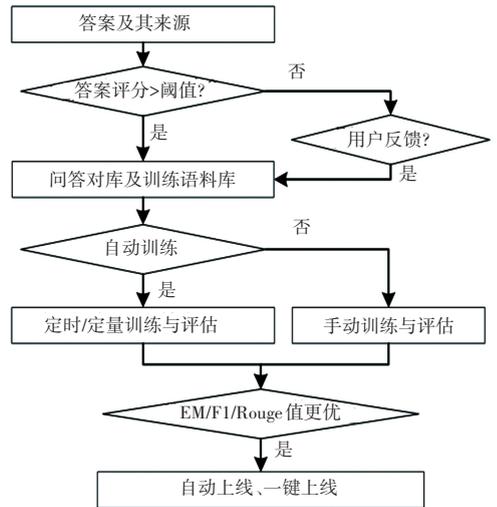


图 4 用户反馈与模型训练流程图

Fig. 4 Flowchart of user feedback and model training

2 实验

为验证本文构建的多引擎智能问答系统的有效性,收集构建了新冠疫情数据集,包含 5 000 余条数据,其内容涵盖了国内、日本和美国半年内每日疫情确诊数据、疫情防控政策数据、政要疫情言论数据、政要基础信息数据和疫情常识数据,并将其组织存储为问答对、知识图谱、自由文档和数据库中记录的形式。由于构建的多引擎智能问答系统支持面向上述 4 类数据的问答,且未发现类似的开源综合问答系统,因而在该疫情数据集下实验验证多引擎智能问答系统下各引擎的问答效果,其准确率 EM 和 F1 指标结果见表 1。

从表 1 中可以看出,本系统在构建的新冠疫情数据集上取得了较好的结果,尤其是问答对库问答表现最好,这主要得益于基于核心实体的粗排序-精排序思想,既可以得到语义最相近的答案,又可以筛选问题句式相似、但核心实体错误的候选答案。而知识图谱问答、阅读理解问答和数据库表问答的测试结果也取得了与主流算法模型相当的得分。此外,本文设计的多源融合问答方式(Top1 答案)准确率得分 84.40%略低于针对不同类型问题分别采用对应问答引擎进行测试得分的平均值 87.80%,这主

要是因为该引擎采用了返回更快、更准答案的方式, 损失了极少的答案精度, 但可以有效提升系统的响

应速度。同时, 提供了融合问答 (Top3 答案) 方式以满足特定场景下对精准问答的需求。

表 1 本系统模型的不同引擎问答效果对比

问答引擎	问答对库 问答	知识图谱 问答	阅读理解 问答	数据库表 问答	多源融合问答	
					top1 答案	top3 答案
EM 值	94.67	92.57	84.57	77.34	84.40	89.60
F_1 值	95.24	94.19	87.85	81.46	86.64	91.73

3 结束语

面向多类型知识库的多引擎智能问答系统集成数据管理、智能问答、用户反馈及模型训练优化于一体, 针对不同格式数据分别存储并构建对应问答引擎, 同时提供方便、快捷、友好的用户评价反馈方式, 让用户在使用中不断帮助该系统提升问答效果, 从实验结果可以看出, 本系统提供了一种效果更优、速度更快的多类型知识智能问答解决途径, 更切合实际使用场景。未来可不断从数据自动组织、常识知识辅助问答等方面继续完善系统, 为多引擎智能问答系统的工程落地应用提供支撑。

参考文献

- [1] WESTON J, BORDES A, CHOPRA S, et al. Towards AI-complete QA: A set of prerequisite toy tasks[J]. arXiv preprint arXiv:1502.05698, 2015.
- [2] 魏泽林, 张帅, 王建超. 基于知识图谱问答系统的技术实现[J]. 软件工程, 2021, 24(02): 38-44.

- [3] 李代祜, 盛杰, 刘运星, 等. 基于知识图谱的军事武器问答系统[J]. 指挥信息系统与技术, 2020, 11(05): 58-65.
- [4] LIU Shanshan, ZHANG Xin, ZHANG Sheng, et al. Neural machine reading comprehension: Methods and trends [J]. arXiv preprint arXiv:1907.01118, 2019.
- [5] ZHENG Liying, DENG Yue, SONG Weishun, et al. An alignment-agnostic model for Chinese text error correction [J]. arXiv preprint arXiv:2104.07190, 2021.
- [6] 韩东方, 吐尔地·托合提, 艾斯卡尔·艾木都拉. 问答系统中问句分类方法研究综述[J]. 计算机工程与应用, 2021, 57(06): 10-21.
- [7] 刘依红, 杨波, 孙宇宇, 等. 基于 BiLSTM 的婚姻法自然语言问答[J]. 计算机工程与设计, 2019, 40(04): 1190-1195.
- [8] 李智, 王震, 杨赋庚, 等. 基于表格的自动问答研究与展望[J]. 计算机工程与应用, 2021, 57(13): 67-76.
- [9] DOU Longxu, GAO Yan, PAN Mingyang, et al. UniSar: A unified structure-aware autoregressive language model for text-to-SQL[J]. arXiv preprint arXiv:2203.07781, 2022.
- [10] ZHOU Xiangming. A study of machine reading comprehension based on attention mechanism [C]//2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China: IEEE, 2021:1058-1061.

(上接第 167 页)

参考文献

- [1] 朱家. 复杂环境下军事目标检测与威胁度评估[D]. 太原: 中北大学, 2020.
- [2] 刘英. 基于卷积神经网络的陆战场目标分类算法研究[D]. 成都: 电子科技大学, 2020.
- [3] 童雪东. 面向无人作战车辆的图像敏感目标检测技术的研究与实现[D]. 南京: 南京理工大学, 2019.
- [4] 朱朔. 基于机器视觉的无人平台目标识别与跟踪系统设计[D]. 南京: 南京理工大学, 2018.
- [5] 王韧. 基于机器视觉的仪表识别算法研究[D]. 西安: 西安理工大学, 2021.
- [6] 周锡祥. 基于机器视觉的无人平台目标识别方法与视频信息处理的研究[D]. 重庆: 重庆理工大学, 2021.
- [7] 朱凯. 基于深度学习的装甲车辆识别技术研究[D]. 南京: 南京理工大学, 2018.
- [8] 王仁晓, 罗建华, 周宇航, 等. 浅析基于图像信息的打击效果评

估技术[J]. 科技与创新, 2021(22): 142-143, 146.

- [9] 张海峰, 于忠琳, 谢宇鹏. 基于贝叶斯——熵权法的岸舰导弹作战效能研究[J]. 海军航空工程学院学报, 2020, 35(04): 310-315.
- [10] 彭庭睿, 刘海鹏, 刘彦, 等. 基于模糊层次分析法和图像对比毁伤评估法的目标权重计算方法[J]. 兵工学报, 2021, 42(S1): 173-180.
- [11] 魏燕明, 张雪艳, 岳奕丹, 等. 基于模糊物元分析法的坦克作战效能评估[J]. 火力与指挥控制, 2017, 42(11): 151-155.
- [12] 邓雪亚. 基于深度学习的战场目标检测和跟踪研究[D]. 西安: 西安工业大学, 2021.
- [13] 梁凌. 多视角聚类新算法研究[D]. 无锡: 江南大学, 2021.
- [14] 詹炜, RAMATOV I, 崔万新, 等. 基于候选区域的深度学习目标检测算法综述[J]. 长江大学学报(自然科学版), 2019, 16(05): 108-115.
- [15] 杨荣, 王明伟, 刘思铭. 基于图像处理算法的目标识别、定位与跟踪系统设计与实现[J]. 物联网技术, 2020, 10(09): 75-79.