

基于 LTP 的糖尿病智能问答系统的 研究与实现

易法令^{1,2}, 孙晓翠^{1,2}, 陈珊珊¹

(1.广东药科大学医药信息工程学院, 广东 广州 510006)

(2.广东普通高校工程技术研究中心-医药信息真实世界工程技术研究中心, 广东 广州 510006)

[摘要] 广泛收集健康网站用户实际提出的糖尿病相关问题,以 LTP 自然语言处理平台为基础对问句进行综合分析,建立糖尿病基本诊疗、指标、用药、饮食、保健品、预防及疑问(判断)词汇等方面的分类词典。应用词语匹配、依存句法分析、相关词语组合分析等方法进行语义关键词提取。在此基础上进行问题分类理解,同时结合问题分类及相关的决策树构建糖尿病的相关知识图谱及问题的判断流程,实现了问题与回答的有效匹配。系统实际问答测试结果是识别率为 91.3%、准确率为 83.6%。

[关键词] 糖尿病,知识图谱,问答系统,语义关键词,LTP,自然语言处理

[中图分类号] TP391 **[文献标志码]** A **[文章编号]** 1672-1292(2023)03-0060-07

Research and Implementation of Diabetes Intelligent Question Answering System Based on LTP

Yi Faling^{1,2}, Sun Xiaocui^{1,2}, Chen Shanshan¹

(1.College of Medical Information Engineering, Guangdong Pharmaceutical University, Guangzhou 510006, China)

(2.Engineering and Technology Research Center of Guangdong Universities-Real World Engineering
and Technology Research Center of Medical Information, Guangzhou 510006, China)

Abstract: Through extensive collection of diabetes related questions actually raised by users on health websites, and comprehensive analysis of questions based on LTP natural language processing platform, build a classification dictionary of basic diagnosis and treatment, indicators, medication, diet, health care products, prevention and question (judgment) vocabulary of diabetes. The semantic keywords are extracted by word matching, dependency syntax analysis, and related word combination analysis. On this basis, the problem classification is carried out. At the same time, combined with the problem classification and the related decision tree, the knowledge graph of diabetes and the judgment process of the problem are constructed, which realizes the effective matching of questions and answers. The actual question and answer test results of the system show that the recognition rate is 91.3% and the accuracy rate is 83.6%.

Key words: diabetes, knowledge graph, Q&A system, semantic keywords, LTP, natural language processing

随着人口老龄化和生活方式的改变,糖尿病在我国患病率逐年升高。按照中国 2 型糖尿病防治指南(2020 年版)中的流行病学调查数据,我国成年人群糖尿病患病率为 11.2%^[1],糖尿病防治成为“健康中国”的主要任务之一,防治工作需要大众进行健康教育和健康管理。健康知识的来源除了基层医疗卫生机构提供的健康教育服务外,公众还会通过互联网技术搜索满足自身需求的健康信息,因此有必要针对糖尿病高发人群或糖尿病患者建立一个自主获取糖尿病相关知识的服务平台,而智能问答系统在知识服务方面的便利、高效及针对性使之成为平台建设的主要选择^[2-3]。

收稿日期:2023-04-24.

基金项目:广州市基础研究计划基础与应用基础研究项目(202102080300)。

通讯作者:易法令,博士,教授,研究方向:人工智能、健康信息管理等。E-mail:flyi@gdpu.edu.cn

1 研究现状

智能问答是为用户提出的自然语言问题自动提供答案.近年来,随着大数据及深度学习技术的不断发展,自然语言处理技术也有了长足进步,各种智能语音助手、智能问答系统层出不穷^[4].在医疗领域,智能问答系统的研究也在不断发展,大部分系统都是在建立相关医疗领域的知识图谱基础上,应用深度学习的相关模型进行问题语义识别与答案生成^[5].谭威等^[6]将基于深度学习的语义解析模型和规则模型融合起来,在医疗基本知识的知识图谱基础上构建了一个医疗问答系统.贾丽娜等^[7]提出一种将 CNN 网络和 BiGRU 网络相结合并引入注意力机制的混合模型—ABiGRU-CNN,该模型有效提高了问答对的匹配效率.吴丹等^[8]在心血管知识图谱基础上,采用 BERT+BiLSTM+CRF 模型进行医学实体识别,构建了一个心血管疾病的智能问答系统.洪海蓝等^[9]以海洋中药知识图谱为基础,综合利用 AhoCorasick, Word2Vec 等技术获取海洋中药实体,构建了海洋中药智能问答系统.目前,医疗问答系统研究主要集中在利用深度学习技术提取问句的主要特征,且多在通用数据集或小规模的领域相关数据上进行测试^[10].对普通用户关注的与典型慢病相关的实际问题缺乏系统研究.为此,本系统以研究分析普通用户针对糖尿病提出的实际问题为基础,构建了一个有针对性的糖尿病自我管理和服务的问答系统.

在技术实现方面,与当前常见的智能问答系统有所不同.当前智能问答系统的知识图谱一般是以相应领域自身的知识体系为基础,分层分类构建.本系统是以当前糖尿病患者和糖尿病易感人群所关注的主要问题为基础,通过对大量实际问句进行统计分析,发现了关键词汇的聚集性,以及关键词汇组合与问句语义之间的高度关联性.以此为基础,提出了“语义关键词”概念,并通过语义关键词的组合进行问题分类,同时构建相应知识体系.

2 问答系统的技术流程

问答系统的数据来源于主流健康网站上大众提出的糖尿病相关问题及对应的医生回答,通过网络爬虫技术,以“糖尿病”为关键词,在 39 健康网、好医生在线等网站上共爬取与糖尿病相关问题及对应的回复 6 537 条,经过排除无意义、不相关问题、去重等处理后,纳入研究的问题共 4 788 条.通过对问题中间句的方式和用词特点及回答的知识范围两方面进行分析,形成了有效的问题理解方法及问题回答模式.

由于深度学习侧重于问题的分类,且要求训练的样本数据量尽可能多.而问答系统更多强调对问题的精准识别,其中有效的关键词提取尤为重要.因此,问答系统没有采用当下比较流行深度学习的分析方法,而是基于比较成熟的语言技术平台——LTP^[11].哈尔滨工业大学研发的语言技术平台(LTP)提供一套包括中文分词、词性标注、命名实体识别、依存句法分析等丰富、高效、精准的自然语言处理模块,在有效关键词识别方面具有较强优势.问答系统应用 Python 作为编程语言,通过 LTP 系统进行问题的自然语言的分析,利用 Neo4j 数据库进行糖尿病知识图谱的构建.最终通过 Neo4j 数据库提供的查询语言 Cypher 来对糖尿病知识库进行检索,并返回问题答案.

系统的技术流程如图 1 所示,总体包括两大部分:问题理解与问题解答.其中问题理解包括 4 个模块,即:LTP 分析、语义关键词提取、问题分类及判断查询条件生成、问题判断与实体链接.问题解答包括 3 个模块,即:查询生成、图谱查询及答案生成.

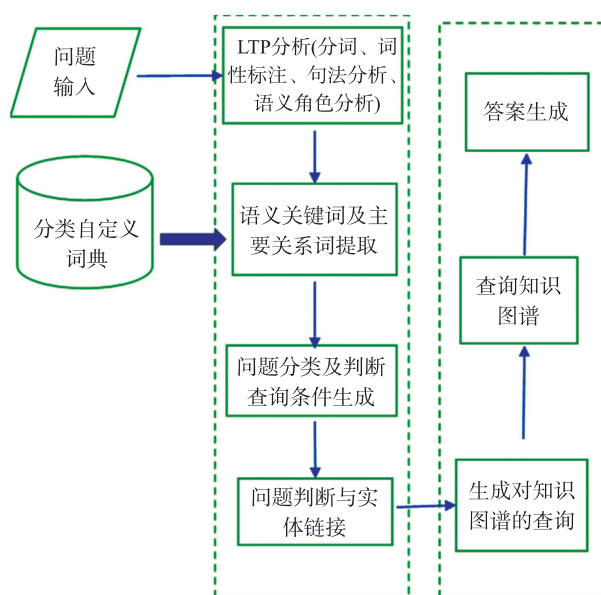


图 1 问答系统的技术流程

Fig. 1 Technical process of Q&A system

3 问题分析

本系统属于针对专用领域的专用人群的系统,其受众主要为糖尿病患者和糖尿病易感人群,问题大多集中在基本诊疗、用药、饮食、保健品、预防及一些基础知识方面,通过分析用户的问题可以发现绝大多数问题有两个特点:(1)问题一般包括两个部分,即:前置条件与问句.有时前置条件是隐含的或包含在问句中,但必须是问句.(2)问题中的用词有比较明显的聚集特征.因此,这里提出一个语义关键词的概念,语义关键词是指去除句子中不影响基本语义表达的修饰或辅助词后,剩下的能够完整的表达整句话语义的关键词.这类关键词以实体关键词为主,还包含有辅助的疑问(判断)词、动词等.比如以下语句:

我是糖尿病患者,应该吃什么水果好?
语义关键词:糖尿病 患者 吃 什么 水果
前置条件:糖尿病 患者
问句:吃 什么 水果

3.1 语义关键词分析

为具体分析语义关键词,从 4 788 个问题中选取各类有代表性的问题 300 个,应用 LTP 进行分词,然后再通过人工提取每个问句的语义关键词.经过统计分析:300 个问句一共提取了 1 423 个关键词(包含指标数值),表 1 是部分主要关键词的统计分析.

表 1 主要语义关键词(部分)出现的频度表
Table 1 Frequency table of main semantic keywords(part)

分类	关键词	出现次数
病名类	糖尿病、2 型糖尿病、1 型糖尿病、妊娠期糖尿病、妊娠糖尿病	180
判断词类	吗、是不是、能不能、会不会、要不要、能否	146
疑问词类	什么、啥、怎么办、怎样、如何、咋办、为什么、哪些、咋回事	152
指标类	血糖、空腹血糖、血糖餐后、血糖空腹、饭后 2 h 血糖、餐后血糖、餐后 2 h 血糖、餐前血糖、血糖餐前…	97
用药相关词类	药、吃药、用药、药物、降糖药、中药、服药、中成药、药品、降压药、药片、开药、换药、停药、打针	66
药品名类	胰岛素、二甲双胍(瓜)、吡格列酮、拜糖平、沃丽汀、葡萄糖、优必淋、糖适平、消渴片、格列本脲、诺和灵…	74
饮食类(含保健品)	饮食、水果、保健品、黄豆、食物、茶、金施尔康、栗子、西餐、牛排、沙拉、枸杞、榴莲、苹果、几丁聚糖、豆浆…	58
并发症及症状类	并发症、糖尿病足、低血糖、(手脚)麻木、(眼睛)模糊、神经病变、(大腿)刺痛、眼底出血、酮症、脑梗塞…	48
常用相关行为动词类	治疗、控制、注意、预防、引起、治愈、降、调整、治、诊断、传染、检查、影响	107
与用药及饮食相关的动词类	吃、打、注射、喝、用、服用、使用	147

从前面的问句及表 1 的分类中可以看出,语义关键词有以下特点:(1)由于问题大多具有口语化的特点,所以存在较多不同关键词表达同样意思的现象,也就是同义词,其中“指标类”词语中表现最明显,比如,糖尿病的两个重要判断指标:空腹血糖和餐后血糖,在表中列出了 9 种描述(血糖、空腹血糖、血糖餐后、血糖空腹、饭后 2 h 血糖、餐后血糖、餐后 2 h 血糖、餐前血糖、血糖餐前).通过对问题及有关回答进行分析,“血糖”“空腹血糖”“血糖空腹”“餐前血糖”“血糖餐前”等都是表示“空腹血糖”,其它则是表示“餐后血糖”.因此,对类似的实体关键词需要统一表示.(2)不同类型的语义关键词的组合与问题的分类之间的关系比较清晰,在大多数的问句中,只要抽取了几个关键词就能理解整个问题的语义.(3)有些类型语义关键词,与问句中的其它词语的组合比较固定,比如:“指标类”的关键词后面一般都是数值或者是描述指标高低程度的词;含有“药品名类”关键词的问句中,一般会有相关的效果评价词或“指标类”词等.

3.2 问题分类

问题分类以前置条件和问句的组合为基础,通过语义关键词进行区分.其中,前置条件的语义关键词包括指标类、症状类、用药类、饮食类、疾病名类、基本信息类以及相关的一些动词.

- (1)指标类:主要指空腹血糖和餐后血糖指标及其对应的数值,有些对指标没有具体数值说明,而是口语化的指标程度描述.比如,“高”“不好”“超标”等.
- (2)症状类:主要指疾病(含并发症)症状的描述,症状描述有些是单一词汇.比如,“眼底出血”.有些则是包括主体及症状词两部分.比如,“足大拇指”“溃烂”.
- (3)用药类:包括各类的口服降糖药、胰岛素的名称,以及一些口语化的描述.比如,“打针”一般指

“注射胰岛素”。

(4) 饮食类:包括各类的主食、辅食、水果、以及保健品的名称。

(5) 疾病名类:包括各种类型的糖尿病及其并发症的名称。比如,“妊娠糖尿病”“糖尿病足”等。

问句一般由疑问(判断)词及问句核心词汇构成,其中,问句核心词汇一般为名词或动词,在表 1 中列举的疑问词及判断词基本上包含了大部分的相关词汇。

分类提取前置条件关键词及问题核心词汇的组合构成了问题的分类。根据当前收集的所有问题以及专业医生的建议,将整个问题分为 5 个大类,17 个小类,小类还可以进一步细分。具体分类体系如表 2 所示。表 3 则列出了实际问题的分类及语义关键词示例。

表 2 问题分类

Table 2 Problem classification

问题大类	包含的小类				
诊治类	单指标	双指标	症状		
饮食推荐类	主食	肉食	水果	蔬菜	保健品
药品推荐类	一般推荐类		条件推荐类		
	一般换药类		条件换药类		
药品使用类	口服降糖药		胰岛素		
基础知识类	疾病一般知识				
	预防、护理、日常生活等知识				
	其它相关知识				

表 3 实际问题分类示例

Table 3 Examples of actual problem classification

问题	分类	条件关键词	问句关键词
空腹血糖 7.6 是糖尿病吗?	单指标诊断	空腹血糖 7.6	糖尿病 吗
我刚确诊了糖尿病,吃什么药好?	药物一般推荐类	糖尿病	什么 药
血糖 18 点,打胰岛素 5 个单位,现经常便秘,脚有点肿有点麻木。请问有什么药治疗	药物条件换药类(指标、药物)	血糖 18 打胰岛素 5 便秘 脚 肿 麻木	什么 药 治疗
用了胰岛素出现不良反应怎么办	药品使用类的胰岛素类	胰岛素 不良反应	怎么办

4 问题理解

4.1 语义关键词提取方式

如何从问题中提取语义关键词是问题理解的关键,根据前面的分析,语义关键词主要分为 3 种类型,分别采用不同提取方式。

(1) 糖尿病相关的专有名词,药品、饮食类的相关名词,相关症状类的常用词汇,以及疑问(判断)词类。由于这类词汇的聚集性及领域的专业性,所以这种类型的关键词可以采用穷举法建立相应的词典。与前置语义关键词及疑问(判断)词相对应系统一共建立了 7 个词典,包括:指标类、症状类、用药类、饮食类、疾病名称类、基本信息类及疑问(判断)词类词典。同时,考虑到有些专有名词在 LTP 分词中会被切分,需要使用 LTP 的自定义词典功能自定义专有词汇。比如,指标“空腹血糖”,LTP 分词就是“空腹”和“血糖”。因此,需要自定义包括药品名、指标等一些专有名词,同时一些同义词也要进行标注。

(2) 与专有名词相关的动词,比如与疾病名相关的“预防治疗”,与药品名相关的“使用”“服用”,与食品名词相关的“吃”“饮用”等。由于这类词比较广泛,难以建立相关词典,因此应用 LTP 句法分析提取相关词汇,并通过 Word2Vec 进行词语相似度计算确定对应的关键词。

(3) 与特定关键词相关的词汇,主要是指与指标关键词或者药品类关键词相关的数值,通过提取句子中的数词和相关组合关系的方式提取。

以上 3 种方式中,方式(1)主要是通过自定义分类词典的关键词匹配的方式来实现,方式(3)则根据不同类型关键词的搭配方式采取不同的策略,下面具体分析方式(2)。

4.2 LTP 依存句法分析与相关关键词筛选

依存句法的关键概念在于依存,句子各个成分之间存在着支配和从属的关系,其中处于支配地位的就是句子的核心词(用 Head 表示),以核心词为基础提取其主要关系词汇,这里的关系主要指并列关系(COO)、主谓关系(SBV)、动宾关系(VOB)以及左、右附加关系(L,RAD)。从依存句法理论的角度来分析,核心词的并列关系词也应该是语句中的另外一个核心,核心词及其并列关系词的主谓和动宾关系词应该是整个句子的主干成分,在实际应用中提取上述关系词汇能够覆盖方式(2)中所涉及的关键词汇^[12]。

算法 1 基于依存句法分析的关键词提取算法:

输入:问句

输出:keyL 列表(包含语义关键词及候选问句关键词)

①LTP 分词、句法分析,获取其分词列表 seg、句法分析列表 dep,初始化关键词列表 keyL.

②查找 seg 列表中的 Head 关系词,加入到 keyL 列表.

③查找列表中 Head 关系词的 COO 关系词. 如果存在,则加入到 keyL 列表.

④查找 keyL 列表中所有词的 SBV 及 VOB 关系词. 如果存在,则加入到 keyL 列表.

⑤查找 keyL 列表中所有词的 COO 关系词. 如果存在,则加入到 keyL 列表.

⑥查找 seg 列表中的疑问词. 如果存在,则加入到 keyL 列表. 同时查找疑问词的 SBV、ATT 及 ADV 关系. 如果存在,则加入到 keyL 列表,同时标记这些词为候选问句关键词. 如果疑问词不存在,则查找 seg 列表中的判断词. 如果判断词存在,则转⑦.

⑦如果判断词是助词列表(“吗”“么”)中的词,则查找其 RAD 关系词. 如果存在,则加入到 keyL 列表,同时标记这些词为候选问句关键词. 如果判断词不是助词列表中的词,则查找其 ADV 关系及 VOB 关系词. 如果存在,则加入到 keyL 列表,同时标记这些词为候选问句关键词.

算法 1 中所有的关键词加入都保持原来语句中的顺序,且不重复加入. 在实际程序运行过程中通过上述 3 种方式提取的关键词都按照在语句中的顺序加入到 keyL 列表,但不能重复加入. 在 keyL 列表中有些词汇可能是非关键词,但是所有的语义关键词都在 keyL 列表中. 因此,为了准确理解问题语义还需要从 keyL 列表中筛选出最终的语义关键词.

语义关键词的筛选以上述算法中生成的 keyL 列表为基础,进行 2 个部分的数据筛选处理. (1) 针对 keyL 列表中的候选问句关键词提取问句核心词. (2) 针对上述算法中前 5 个步骤提取的动词筛选出与专有名词相关的动词. 筛选算法主要根据问句中匹配的主要关键词的匹配规则通过 Word2Vec 的相关算法进行词语相似度计算确定对应的关键词.

4.3 问题类型与判断、查询条件

问题理解的最终目标是确定问题类型及相应的判断与查询条件. 问题的类型可以根据前置条件关键词类型和问句核心关键词类型来确定. 在表 3 中已列出了相关的示例,从中可以看出不同的关键词类型的组合决定了不同的问题类型.

问题类型决定了问题回答的处理流程,其中包括问题可能的判断决策及数据库的查询,具体而言就是有些类型的问题可能需要通过决策树进行分析决策,有些问题可能需要查询数据库,因此需要确定问题的可能判断条件及与数据库查询相对应的实体和属性. 图 2 是根据前置关键词确定药物推荐类问题的基本流程图.

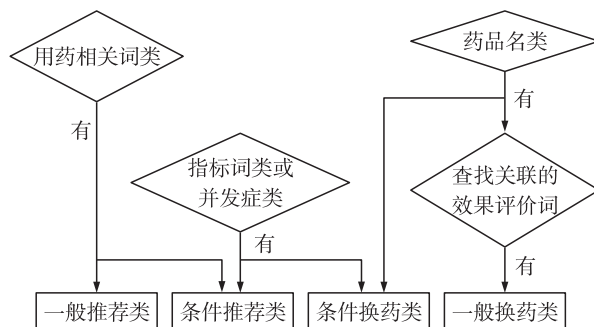


图 2 药品推荐类型问题的判断流程图

Fig. 2 Flow chart for determining the drug recommendation type

5 问题解答

问题回答分为 2 个部分,第 1 部分是要根据具体条件进行判断、选择. 比如,根据具体的血糖指标进行判断,或者根据具体情况推荐降糖药物;这部分需要通过决策树或者推理算法来实现;第 2 部分是与糖尿病相关实体(包括药物、饮食、并发症等)的具体属性描述,属于与糖尿病相关的知识点. 问题回答的相关知识点主要以中国 2 型糖尿病防治指南(2020 年版)为基础^[14],综合相关网站问题的回答及专业医生的意见形成的. 一般一个问题与一个或多个知识点相对应. 这些知识点按照问题的分类构成相应的知识体系.

5.1 问题答案知识体系构建

在构建问题回答的知识体系时,第 2 部分可以通过构建知识图谱来实现. 本系统应用 Neo4j 数据库构建知识图谱,图 3 是与药品推荐及使用相关的部分知识图谱,限于篇幅这里不具体描述构建过程. 如何将

第1部分的判断推理与第2部分的知识图谱相结合,构建问题答案的知识体系,同时建立规范的问题答案,是问题解答的关键.



图3 药品推荐及使用相关知识图谱(部分)

Fig. 3 Knowledge graph of drug recommendation and use(part)

以单指标类问题为例具体说明如何进行两个部分的融合. 单指标问题就是指问题中只包含单个指标的数值, 一般是指“空腹血糖(FBS)”, 根据这个值来诊断是否糖尿病、如何用药、如何预防等方面的问题. 为了与整个知识图谱相衔接, 建立了一个单指标结点, 该结点的属性值有 2 个特点: (1) 根据糖尿病诊断与治疗的问题不同的回答方式将糖尿病程度判断与糖尿病的治疗建议分开. (2) 考虑到不同条件的药品推荐可能不同, 把推荐的具体药品名词与治疗建议的描述分开. 下面是在 Neo4j 数据库中建立结点描述.

CREATE(singleIndex:单指标类{

indexType:“空腹血糖”.

normal:“空腹血糖指标正常,应该不是糖尿病”.

impairedFastingGlucose:“空腹血糖受损,是糖尿病前期,应该通过饮食控制和运动来控制血糖指标,并进行一段时间的血糖监测,如果较长时间血糖无法达到正常状态,则考虑药物治疗”。

lightDiag:"空腹血糖指标轻微超标,应该是糖尿病,需要进一步检查确诊".

medHighDiag:"空腹血糖指标超标较高,是糖尿病,需要从生活方式和药物两方面进行共同干预".

highDiag:"空腹血糖指标超标严重,是糖尿病,建议加大生活方式的干预,和加大药物剂量,需尽快控制血糖".

lightDiabete:"空腹血糖指标轻微超标,需要严格控制饮食,并通过口服降糖药来降低血糖值,可以服用".

medHighDiabetes:"空腹血糖指标超标,需要从生活方式和药物两方面进行共同干预,可以服用".

highDiabetes:"血糖指标严重超标,需要加大生活方式的干预,药物治疗方面,可采取多种降糖药物联合治疗或者用胰岛素进行治疗.推荐口服降糖药".

 $\})$

在“单指标类”结点中,normal(FBS≤6.1)、impairedFastingGlucose(FBS>6.1 &&<7.0)、lightDiag(FBS>=7.0 &&<=8)、medHighDiag(FBS>8 &&<11.1)、highDiag(FBS>=11.1)分别对应不同指标时的糖尿病判断,lightDiabete、medHighDiabetes、highDiabetes 则对应是糖尿病时的用药推荐,注意其中属性值的最后还需要药品名,具体推荐的药品名,则根据问题中的前置条件与药品推荐的算法具体实现。

5.2 问题解答实现

问题解答部分是根据问题理解部分确定的语义关键词,通过前置条件关键词与问句核心关键词确定

问题类型,然后按照不同的问题类型的问题回答模板,确定判断条件及相应的实体和属性,建立知识图谱查询语句,访问 Neo4j 数据库获取相应的内容,并按照一定格式组合后形成答案并输出,其中 Neo4j 数据库中已经构建了针对问题回答的知识图谱。

5.2.1 查询语句的建立

Neo4j 数据库的查询语言是 Cypher, Cypher 是一个描述性的图形查询语言,通过模式匹配来匹配图数据库中的结点和关系。一般的 Cypher 查询语句通常包括 match、where 及 return 子句。其中,match 子句主要定义匹配模式,对数据库中的数据进行匹配,以获取满足查询条件的数据。where 子句定义过滤条件,对结点或关系的属性进行判断,过滤 match 的查询的结果。return 子句指定需要返回的内容。

确定了问题所对应的实体(Entity)和属性(Attribute),则一般可生成如下对应 Cypher 查询语句。比如,描述对应药品功能的查询语句,其中的 iKey 就是实际抽取的药品名词:

```
" match ( p: 药物类 ) where p. class = '{ 0 } '
return p.description".format( iKey ).
```

5.2.2 知识图谱查询及答案生成

通过查询语句链接存放在 Neo4j 数据库中的知识图谱进行数据查询,然后将返回结果与相关判断结合进行必要的格式整理后生成最终答案。图 4 是问答系统的测试截图,其中 4 个问题对应不同的类型,问题 1 是单指标判断类型。问题 2 是条件(含单指标、并发症)药推荐类型。问题 3 是治疗的一般知识类型。问题 4 含药品名的条件换药类型。

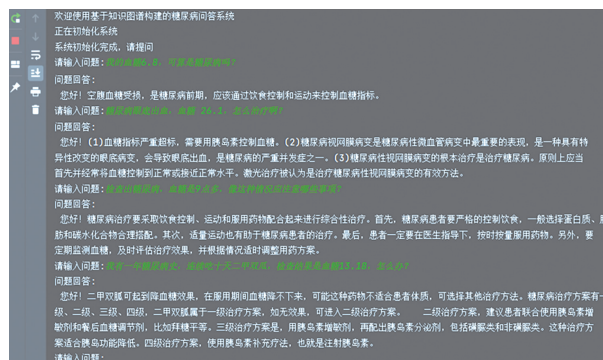


图 4 问答系统运行测试截图

Fig. 4 Screenshot of Q&A system operation test

6 结论

问答系统测试了整理后的 4 788 个问句,其中小部分问句由于抽取的关键词组合无法确定对应的问题分类,因此问题无法被识别,问题总体识别率达到 91.3%。针对已识别的问题,通过查询语句获取问题答案,与问题对应的标准知识点相比,其回答的准确率为 83.6%。其中回答出现误差的原因主要有两个方面:(1)问题分类的差错。(2)词语相似度匹配方面,有些词语因为不同的语境有不同的含义导致词语理解错误。

[参考文献](References)

- [1] 中华医学会糖尿病学分会. 中国 2 型糖尿病防治指南(2020 年版)[J]. 中华内分泌代谢杂志, 2021, 37(4): 315-409.
- [2] 何延, 张宁. 智能医疗问答系统的设计与实现[J]. 中国医疗设备, 2021, 36(9): 100-103, 108.
- [3] 庄莉, 苏江文, 卢伟龙, 等. 专业领域智能问答系统设计与应用[J]. 电子技术与软件工程, 2022(4): 210-213.
- [4] 刘佳, 王路路. 标准化服务智能问答系统研究[J]. 信息技术与标准化, 2022(10): 88-92.
- [5] 侯梦薇, 卫荣, 陆亮, 等. 知识图谱研究综述及其在医疗领域的应用[J]. 计算机研究与发展, 2018, 55(12): 2587-2599.
- [6] 谭威, 刘成良. 基于知识图谱和模型融合的医疗问答系统的构建[J]. 中华医学图书情报杂志, 2021, 30(11): 1-9.
- [7] 贾丽娜, 陈恒, 李冠宇. 基于注意力混合模型的中文医疗问答匹配[J]. 计算机应用与软件, 2021, 38(11): 148-154.
- [8] 吴丹, 周作建. 基于知识图谱的心血管疾病智能问答系统[J]. 软件导刊, 2022, 21(3): 160-164.
- [9] 洪海蓝, 李文林, 杨涛, 等. 基于知识图谱的海洋中药智能问答系统的设计与实现[J/OL]. 世界科学技术-中医药现代化[2023-07-15]. <https://kns.cnki.net/kcms/detail/11.5699.r.20230112.1159.003.html>.
- [10] 吴宗友, 白昆龙, 杨林蕊, 等. 电子病历文本挖掘研究综述[J]. 计算机研究与发展, 2021, 58(3): 513-527.
- [11] 郑捷. NLP 汉语自然语言处理原理与实践[M]. 北京: 电子工业出版社, 2017.
- [12] KONG Z, YUE C X, SHI Y, et al. Entity extraction of electrical equipment malfunction text by hybrid NLP algorithm[J]. IEEE Access, 2021(9): 40216-40226.

[责任编辑:陈 庆]