

智能导诊服务机器人的设计与实现

崔浩 刘丰源

(上海计算机软件技术开发中心 上海 201112)

摘要 针对就诊人流集中、医院导诊人力资源不足的问题,设计一款医院智能导诊服务机器人。机器人融合了智能语义理解分析、图形化可视界面设计、传感器应用、路径规划、脸部识别等技术,可实现患者就诊过程的导诊咨询、医院各科室的路径指引以及常规的体温心率数据的测量。从功能结构设计、关键技术设计与实现、医疗知识库构建、系统演示等方面,详细介绍其设计开发过程。实验表明,该智能导诊机器人图形化界面设计清晰友好,工作运行稳定,能对医院的患者进行有效合理的分流,大大提高医院患者就诊效率,解决医院患者高峰就诊慢的问题。

关键词 导诊机器人 智能问询 人机交互 医疗信息化

中图分类号 TP3 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2020.07.055

DESIGN AND IMPLEMENTATION OF INTELLIGENT GUIDANCE SERVICE ROBOT

Cui Hao Liu Fengyuan

(Shanghai Computer Software Technology Development Center, Shanghai 201112, China)

Abstract In view of the problem of centralized flow of patients and insufficient human resources for hospital guidance, we designs an intelligent guidance service robot for hospital. The robot integrated several technologies such as intelligent semantic understanding and analysis, graphical visual interface design, sensor application, path planning, and face recognition. It could realize the guidance and consultation of patients in the process of consultation, the path guidance of hospital departments and the measurement of routine body temperature and heart rate data. The design and development process of the robot was introduced in detail from the aspects of functional structure design, key technology design and implementation, medical knowledge base construction and system demonstration. Experiments show that the graphical interface design of the intelligent guidance robot is clear and friendly, and its operation is stable. It can effectively distribute the hospital patients reasonably, greatly improve the efficiency of hospital patients, and solve the problem of slow peak visits of hospital patients.

Keywords Guidance robot Intelligent inquiry Human-computer interaction Medical informatization

0 引言

随着互联网技术的迅速发展和大数据时代的到来,电子医疗信息剧增,然而在医疗导诊的工作中对电子医疗信息的利用率始终较低,传统的人工医疗导诊繁杂低效依然是大多数医院的痛点。仅仅依靠医护人员的导诊辅助缺乏一定的高效性,导致病人满意度下降。与此同时,近年来人们对健康关注度显著提高,医院人流高峰出现频次越来越高,医疗导诊人力资源明

显不足,利用机器人智能医疗导诊能有效地对医院患者进行合理分流,同时能让患者享受到智慧医疗的全新体验。

随着医疗信息化的发展,导诊工具也逐渐进步^[1],机器人已经渐渐走进人们的视野,然而目前还没有医疗行业大规模实际应用机器人智能导诊方案,究其原因主要有以下几点:

1) 传统医院导诊工具操作繁琐,可视化图形界面理解较为麻烦,不够友好。

2) 传统医院导诊工具功能单一,仅仅集成了简单

的查询功能,不能有效地提供患者所需信息。

3) 导诊知识库信息数据少,缺少人机交互的过程。

智能导诊服务机器人是依据医院导诊服务要求而设计开发的专业领域服务应用机器人^[2]。它由感知、分析、诊断和优化四部分组成,可通过与患者实时语交互,依据多维度疾病模型,引导患者了解病情,给出合理的导诊建议。该机器人基本实现自助分诊,提升医院服务质量,大大减轻医务员负担。

1 系统功能结构设计

1.1 系统架构

基于医疗知识库的系统架构设计如图1所示。



图1 导诊机器人系统架构

导诊机器人架构总体包含辅助预诊、科室导航、疾病症状自查、就诊流程导诊、娱乐休闲、智能迎宾、专家推荐查询和数据测量等主要模块,其功能分别如下:

1) 辅助预诊:实现机器人与患者的诊前辅助,由机器人代替医生进行患者诊前基本信息的确认,后续转接给医生确诊。

2) 科室导航:通过医护人员对机器人进行路径规划录入,机器人保存相关科室路径,后续患者问询,机器人可实现科室带路指引功能。

3) 疾病症状检索:系统拥有丰富的医疗知识数据库,包含疾病、症状、药品、检查项目等多种百科知识,从多维度对患者所询问题进行解答和图形化展示。

4) 就诊流程导诊:机器人系统整合了医院的预约、挂号及医疗卡使用等功能,可实现在机器人端进行挂号、预约等功能。

5) 娱乐休闲:机器人整合了多个第三方聊天接口,能在各个方面满足患者的问询,包含天气、交通等通用功能模块。

6) 智能迎宾:机器人基于人脸信息数据库来识别范围内捕捉到的人脸进行比较识别并自定义迎宾操作。

7) 专家推荐查询:根据患者的问询进行信息收集,结合患者的情况进行就诊专家推荐以及专家信息查询。

8) 医院常规数据测量:通过机器人连接外部电子测量仪器,进行智能化测试体温、心率等常规操作。

系统通过三种路径进行数据的获取,分别为语音识别、人脸识别以及传感器感知,获取的信息进入系统对话管理模块,通过系统内部的多模匹配算法以及相关自然语言处理技术结合已有的医疗知识库,识别患者意图或者给出咨询结果。系统依据医疗知识库中的症状数据和病例挂号数据,构建症状倒序索引表和科室推荐表^[3],患者与系统交互进行语音问诊后,采集语音输入信号进行语音识别,转换为文本并提取症状信息,依据症状信息与症状索引表比对,按照相关度对可能的疾病进行排序并得出诊断结果。允许用户对诊断结果打分和反馈意见录入,系统将反馈结果作为优化诊断结果的依据。结合疾病诊断结果和科室推荐表为患者推荐挂号科室,同时可以查看疾病更多相关信息。

1.2 访问流程

系统构建了以特征词库、对话知识库、特征向量库为基础的医疗知识库,结合专家矫正后的语料学习,智能化地解决患者问询。系统访问流程如图2所示。

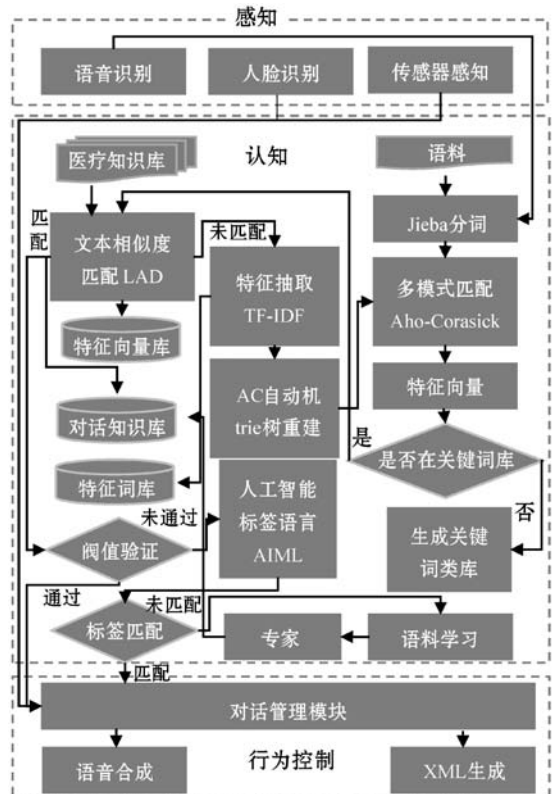


图2 系统访问流程

系统数据源通过语音识别、人脸识别以及传感器感知获取,转换成文本数据,通过自然语言处理技术处

理后与医疗知识信息进行多模式匹配^[4],得出最优答案,优化后反馈给患者用户。

2 关键技术

2.1 基于语料库的句法分析

句法分析的传统方法是通过添加分析规则来实现的。基于规则的分析方法在大数据量的句法分析过程中存在弊端^[5],因此本系统采用基于语料库的句法分析方法来改良。首先基于系统试运行阶段的初期语音采集信息作为基本语料库,其次依靠人工实现基于依存句法的基本语料库的词性标注工作,然后将已经标注好的依存关系构建句法树矩阵,并依据矩阵抽取每个词汇的语法结构信息和搭配信息构建成为开放可扩展的知识库。

在已有知识库的情况下对句法进一步分析。首先对于采集的句子中的每个词汇进行分词,依照知识库中每个词所对应的依存关系,构建所有可能的句法树,形成特殊矩阵。其次根据知识库和句法公理,筛选合适的中心词并剔除不合理的依存关系。然后依据句法公理,对剩余的句法树进行剪枝操作,经过剪枝的句法树会残缺,在最后搜索环节中就可以去除所有残缺的句法树,大大提高检索的效率。最后需要对句法分析的效果进行评价和优化,通过结构合理加权系数、词间适应性系数、比邻度系数和整体评价四个评价维度进行最终的评价^[6]。

2.2 基于追问与反馈的意图识别

用户在进行信息检索时的需求已经不仅仅局限在基于关键词模糊查询的传统搜索方法中,还需要系统能对用户的意图进行识别。意图识别首先需要确定意图分类及其特征词,其大致可以区分为信息检索型、事务型和导航型三大类,其中信息检索型意图又可以细分为其特征词采用分类算法进行学习归类^[7],最终获得意图分类及特征词划分表如表1所示。

表1 意图分类及特征词划分表

查询意图标识	查询意图类别	意图特征词
SI1	信息检索	查找、搜索、列出……
SI2	知识询问	有哪些、在哪里、多少、是什么、怎么定义……
SI3	咨询建议	如何、怎样、怎么办……
SI4	资源下载	下载……
SI5	导航	跳转、导航……

用户输入咨询内容后,系统基于采集到的信息,分词得到一系列的意图特征词,并根据意图特征词的分

类尝试明确用户搜索意图。若用户给出的关键词不在已有的意图特征词列表中,则系统会根据已有特征词的同义词表进行动态匹配,并提供用户相应词语替换建议,辅助用户进行点选操作,实现有效问题的构建。若在同义词库中也无法匹配到对应的关键词,则将该关键词作为新的意图特征词插入意图特征词库,并基于特征词分类算法实现新的特征词的类别划分。

系统可以设置多种针对患者的多轮问答场景由机器人主动进行智能问询及推荐,更加智能化地获取收集患者的就诊信息,为患者提供较好的就诊体验。

2.3 基于 Aho-Corasick 多模匹配的文本匹配

Aho-Corasick 算法是多模式匹配算法中的经典算法,目前在实际应用中较多,该算法对应的数据机构是 Aho-Corasick 自动机,简称 AC 自动机^[8]。

为了实现与已有医疗知识库的数据匹配,系统实现了一套以 Aho-Corasick 为基础的文本匹配算法。

根据获取的患者的问询信息,首先需要进行分词处理,提取出来的症状信息需要被结构化存储为索引表,基于索引可以大大加速症状检索,提高系统的可用性。症状索引表需要至少包含三个属性,分别是症状词典编码、国际疾病分类编码和对应疾病类型数。后两个属性在进行诊断时使用。

其次通过多模式匹配获取到对应的特征向量,基于索引表检索其是否存在于已有的关键词库,若存在,则将得到的所有与症状相关的疾病都取出,存入到可能结果集合中,作为初始结果集。

然后进行文本相似度的匹配,通过设定阈值验证,匹配出知识库中最相近的问答^[9]。相关度函数为:

$$S = \frac{1}{t} \sum (q \times d) \quad (1)$$

式中: S 为症状与疾病的相关度; t 为症状类型数量,由于不同疾病的描述症状数量有所不同,描述症状较多的疾病被检索的概率会高于描述症状少的疾病,因此为了消除此类差异,需要采用症状数量的平均数来平衡疾病被检索概率差异问题; q 为每个症状的特有权重,权重的确定以该症状被检索频率为依据,高频症状被赋予较高的权重,稀有症状被赋予较低权重,并以此权重判定疾病诊断的主要症状和次要症状; d 为每个症状在每种疾病的症状描述中出现频率,用于表征某症状相对于某疾病的重要程度。对相关度 S 进行最大-最小值归一化处理,并将相对相关度作为最终判断依据按由大到小进行正向排序。

若没有匹配成功,则对问题使用 TF-IDF 算法进行重新提取特征关键字^[10],再次进行 AC 多模式匹配,若

知识库中没有结果,则存储收集到的新的特征值,后续补充特征值对应的答案。

2.4 诊断结果优化

系统给出初步诊断结果以后,用户可根据自己的实际情况选择结果列表里最符合病情的几种疾病,系统需要回传用户的反馈给疾病自查系统,重新进行排序展示。

为实现诊断结果的优化,系统会按照如下两个维度进行自我调整。首先将预诊结果疾病症状与当前输入症状进行比对,如果该疾病存在主流症状但并没有被用户作为症状输入,则将该遗漏症状加入查询项。其次结合反馈的疾病列表,对当前输入的症状权重进行调整,次数多的症状相应提升权重^[11]。

反馈的信息与原来的症状信息权重比为 α :

$$\alpha = 0.5 \times ((x/\max) + 1) \quad (2)$$

式中: x 为某种疾病在反馈疾病列表里的位置, \max 为该疾病在诊断结果集中的位置。当二者比值越接近1(即 α 越接近1)时,说明原诊断结果正确性越高,反之则说明用户选择了预诊结果集中靠后的疾病,即原预诊结果症状信息权重不够合理,理应把用户选择的疾病的权重加大,以加强反馈的优化效果。

3 系统实现

3.1 系统主界面

智能导诊服务机器人系统主页面如图3所示。



图3 智能导诊服务机器人系统主页面

3.2 语音人机交互模块

这一模块主要实现系统与用户的互动,系统内置主动问答流程,可针对“有什么症状”、“症状持续时间”、“症状严重程度”等进行提问,同时可以设置多轮对话场景,由机器人更加智能化地问询来收集用户的就诊信息,将收集到的信息传送到远程端统一处理,结合后天分析,及时有效地反馈给用户准确的答案或者推荐。语音采集页面如图4所示。

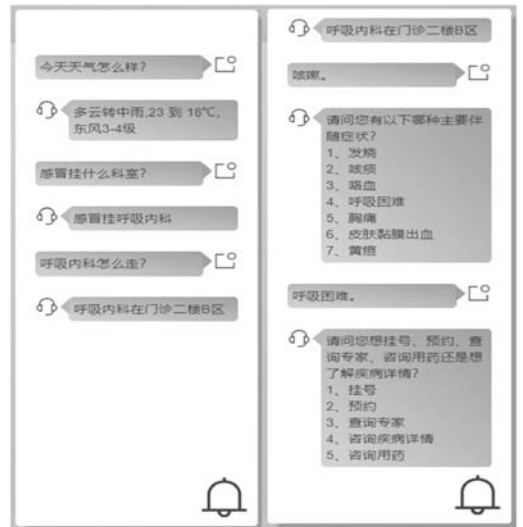


图4 语音采集页面

3.3 诊断结果

诊断结果展示页面分两部分,第一部分是上方的搜索框,该框可以接收患者输入的疾病,点击立即搜索后下面的数据随即发生变化。在患者无输入即默认的情况下,该输入框中提示:请输入疾病关键字。当有输入后,该框中显示的是患者所输入的疾病名称。疾病搜索页面如图5所示。



图5 疾病搜索页面

第二部分为下方的数据显示区域,包括疾病名称、疾病介绍、疾病对应症状和四个链接,其中四个链接可以跳转到疾病对应的疾病简介、典型症状、临床检查和治疗方法的位置。在患者无输入即默认的情况下,疾病通过疾病表中的权重字段的内容进行排序,其中权重字段中的数据是疾病的热度值。当患者输入疾病关键字后,下面列表的显示是从相似度最高到最低来排列。

当患者输入疾病名称时,按相似度显示疾病的信息。从疾病表中获取全部疾病的名称,存到列表中,再从疾病别名表中获取全部疾病的别名,存到列表中。再对列表进行去重处理。接下来,将从前端获取到的参数和列表中全部的疾病进行相似度匹配,规定的界线为0.5,将所有满足相似度大于0.5的疾病进行从

高到低的排序,再将所有的疾病名称存入列表中返回给前端。

当患者点击搜索出来的疾病后,例如感冒,就会跳转到疾病相关信息展示界面,包含疾病简介、典型症状、发病原因、预防、临床检查、鉴别、治疗方法、护理、饮食保健和并发症等相关信息。疾病相关信息展示页面如图6所示。



图6 疾病相关信息展示页面

3.4 智能导诊

依据预诊后给出的结果,与科室推荐表联合查询后给出推荐挂号就诊的科室列表,系统依据科室列表中概率推荐最高的科室,高亮动态展示医院地图、患者所在位置、最佳通行途径等导航信息。同时机器人可以根据已有规划路径进行路径指引,实现智能导诊功能。机器人路径规划页面和智能导诊页面分别如图7和图8所示。

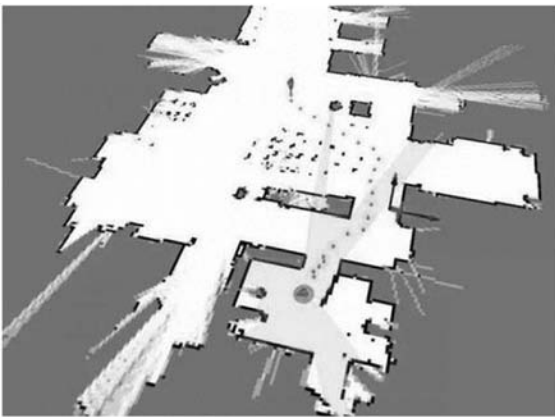


图7 机器人路径规划页面

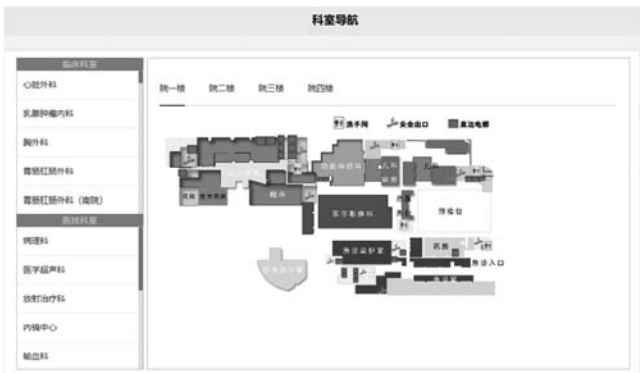


图8 智能导诊展示页面

4 结语

本文介绍了一款基于医疗知识库的智能化导诊机器人,并详细介绍了其系统架构、访问流程、关键技术原理、数据库设计,给出了关键功能的系统展示。智能机器人问诊提升了传统分诊系统的交互性差的问题,可以有效解决医院中人力成本浪费、高峰时段人力资源不足、患者就诊体验差等问题。

系统目前在患者咨询问题收集训练数据量上还有所不足,同时医疗知识库目前只对接了少量医院的在线电子病历数据和部分互联网数据资源,数据量仍有提高空间,因此未来工作可以在基于大量数据的深度学习和知识库数据量提升方面进行改进。

参考文献

- [1] 万薇,邓秋迎.我国急诊分诊工具的研究现状与对策[J].全科护理,2016,14(5):450-453.
- [2] 陈曦,杨雅慧,赵嘉麒,等.智能导诊服务机器人的设计与实现[J].物联网技术,2017,7(11):70-73.
- [3] 马钰,王宏志,高宏.基于医疗知识库的辅助诊疗系统[J].计算机应用,2016,36(S1):217-219.
- [4] 曹为政,葛蒙蒙.多模式匹配算法研究和优化[J].智能计算与应用,2018,8(2):129-133.
- [5] 史兆鹏,邹徐熹,向润昭,等.基于依存句法分析的多特征词义消歧[J].计算机工程,2017,43(9):210-213.
- [6] 马骥.面向依存句法分析优化技术的研究[D].沈阳:东北大学,2015.
- [7] 杨艺,周元.基于用户查询意图识别的Web搜索优化模型[J].计算机科学,2012,39(1):264-267.
- [8] 汪泓才,李训根.一种基于分类存储的空间高效Aho-Corasick算法[J].计算机应用与软件,2017,34(5):279-282,316.
- [9] 谷重阳,徐浩煜,周晗,等.基于词汇语义信息的文本相似度计算[J].计算机应用研究,2018,35(2):391-395.
- [10] 贺科达,朱铮涛,程昱.基于改进TF-IDF算法的文本分类方法研究[J].广东工业大学学报,2016,33(5):49-53.
- [11] 钟旭东.网页分类中的标签权重自动优化研究[D].合肥:中国科学技术大学,2015.

(上接第301页)

- [13] 曹震震,顾小卓,顾梦鹤.面向Ad Hoc网络的无证书认证组密钥协商协议[J].计算机应用,2019,39(2):476-482.
- [14] 张笛.基于ABE的群组密钥管理的研究[D].西安:西安电子科技大学,2018.
- [15] 张辉,石胜友,石倩.基于LKH树的组播密钥管理方案设计[J].计算机工程与设计,2019,40(2):312-316.