

## 基于人工智能的新型冠状病毒肺炎流行病学信息采集系统\*

王维笑<sup>①</sup> 贾露荣<sup>②</sup> 陶焜<sup>③</sup> 李艳<sup>④</sup> 闫海荣<sup>①\*</sup>

\*基金项目：国家自然科学基金（U1736210）“复杂声学环境下声学事件检测与音频场景识别方法研究”

①清华大学自动化系 北京 100084

②中日友好医院信息中心 北京 100029

③北京爱医声科技有限公司 北京 100085

④福州数据技术研究院有限公司 福建 福州 350200

**[摘要] 目的：**结合新型冠状病毒肺炎(COVID-19)疫情下现场流行病学调查工作的实际需求，开发一种移动端流行病学智能采集系统，为一线流行病学调查工作者提供高效、便捷的信息采集服务，最大限度地利用智能化、信息化手段提高疫情防控效率，降低交叉感染风险。**方法：**系统通过人工智能(AI)人机对话引擎进行对话管理、知识库管理以及 AI 技术管理，收集患者发病前后的暴露情况、接触情况、活动轨迹以及就医情况等信息，为及时采取相应防控措施提供数据基础。**结果：**系统可模仿流行病学调查工作者以自然对话的形式进行多轮引导式问答，实现流行病学信息采集、化验单拍照识别、调查表单自动生成以及疫情资料智能归档等功能。**结论：**基于 AI 技术的移动端流行病学智能采集系统对传统的流行病学信息采集模式进行创新，使采集工作更加及时化、便捷化、规范化及精准化，在疫情防控的关键时期，极大地提高了一线流行病学调查人员的工作效率与调查质量。

**[关键词]** 新型冠状病毒肺炎(COVID-19)；现场流行病学调查；信息采集；人机交互；疫情防控

*An epidemiological information collection system of COVID-19 based on artificial intelligence/WANG Wei-xiao, JIA Lu-rong, TAO Kun, et al//China Medical Equipment,2020*

**[Abstract] Objective:** Based on the actual needs of epidemiologic field investigations during COVID-19(novel coronavirus pneumonia 2019), a mobile intelligent epidemiological information collection system was developed to provide efficient and convenient information collection services for frontline investigators, which can maximize the use of intelligent and information technology to improve the efficiency of epidemic control and prevention, and reduce the risk of cross infection. **Methods:** The system manages dialogues, knowledge bases, and artificial intelligence technologies through an intelligent human-computer dialogue engine, and collects information about exposures, contacts, activity trajectories, and medical treatments before and

after the onset of the disease, so as to provide a data basis for timely taking corresponding prevention and control measures. **Results:** The system can conduct multiple rounds of guided questions and answers in the form of natural conversation just like real investigators. Functions such as epidemiological information collection, identification of test sheets, automatic generation of survey forms, and intelligent archiving of epidemic information, are all available. **Conclusion:** The system is innovative in the traditional epidemiological information collection mode, making the collection work more timely, convenient, standardized and accurate. During the critical period of epidemic control and prevention, the system will greatly improve the work efficiency and quality of epidemiologic field investigation.

**[Key words]** COVID-19; Epidemiologic field investigation; Information collection; Human-computer interaction; Epidemic control and prevention

**[First-author's address]** Department of Automation, Tsinghua University, Beijing 100084, China.

流行病学调查作为新型冠状病毒肺炎(COVID-19)疫情防控的重要手段,对于锁定感染者、搜索密切接触者以及梳理传播链条发挥着不可替代的作用<sup>[1-5]</sup>。在全国多个地区先后启动重大突发公共卫生事件一级响应机制时,各地疾病预防控制机构派出流行病学专业人员,深入现场展开调查和疫情分析,与患者面对面交流,询问患者发病前后的暴露情况、接触情况、活动轨迹以及就医情况,寻找与传染源以及传播途径有关的蛛丝马迹,为判定密切接触者并采取隔离措施、划定消毒范围提供依据<sup>[6-7]</sup>。然而,随着确诊病例不断增多,疫情进入暴发期后,持续、多点位的流行病学调查工作使得流行病学调查人员的工作量极大增加<sup>[8]</sup>。传统的人工问卷调查方式暴露出效率低、准确性差、数据传输不及时和利用率低等诸多问题。此外,COVID-19以呼吸道飞沫传播和密切接触传播<sup>[9]</sup>为主要传播途径,传统的信息采集方式无法避免与被调查者的大量交流与接触,增加了现场流行病学调查人员交叉感染的风险。因此,开发一种针对COVID-19的流行病学智能采集系统,将现场流行病学调查工作中的信息收集流程智能化、标准化,帮助流行病学调查人员及时收集完备信息的同时最大限度地减少人员之间的交流与接触,将极大地提升流行病学调查工作效率,避免交叉感染,遏制疫情蔓延。

近年来,随着人工智能(artificial intelligence, AI)技术的飞速发展,越来越多的医疗场景中出现了智能化应用。人机对话系统作为AI领域的核心技术之一,也被广泛地应用在居家自诊、分诊导诊及病史采集等医疗场景中。人机对话系统通过自然语言处理相关技术来模仿学习人与人之间对话的方式和内容,让机器通过自然语言与人进行信息交互<sup>[10]</sup>。语音识别、自然语言理解、对话状态跟踪及策略优化、自然语言生成以及语音合成是人机自然语言交互的关键组成部分<sup>[11]</sup>。本研究针对此次疫情传播速度快、传播途径多及传播范围广等特点,利用多模态的移动端智能人机交互技术,构建智能信息采集系统,辅助现场流行病学调查人员进行便捷高效的信息采集,将有效降低交叉感染风险,缓解调查人员工作压力,提升数据质量与安全性,并为后续的数据挖掘与共享提供基础。

## 1 流行病学智能采集系统技术架构

COVID-19 流行病学智能采集系统利用智能人机交互技术,根据《新型冠状病毒感染的肺炎防控方案》《中华人民共和国传染病防治法》以及《突发公共卫生应急条例》等文件要

求，并结合现场实践建立标准化的流行病学信息收集模板，构建文字与图片智能识别平台，并实现结构化信息与预设模板的自动映射与分析，为流行病学现场调查提供更加智能高效的信息采集模式。

智能人机交互系统的核心是人机对话服务引擎，其主要由对话管理系统、知识库管理系统及 AI 技术管理系统三部分组成。前端可根据现场需要采用智能手机 APP、智能平板 APP 等为载体。系统技术框架见图 1；手机 APP 应用界面见图 2。

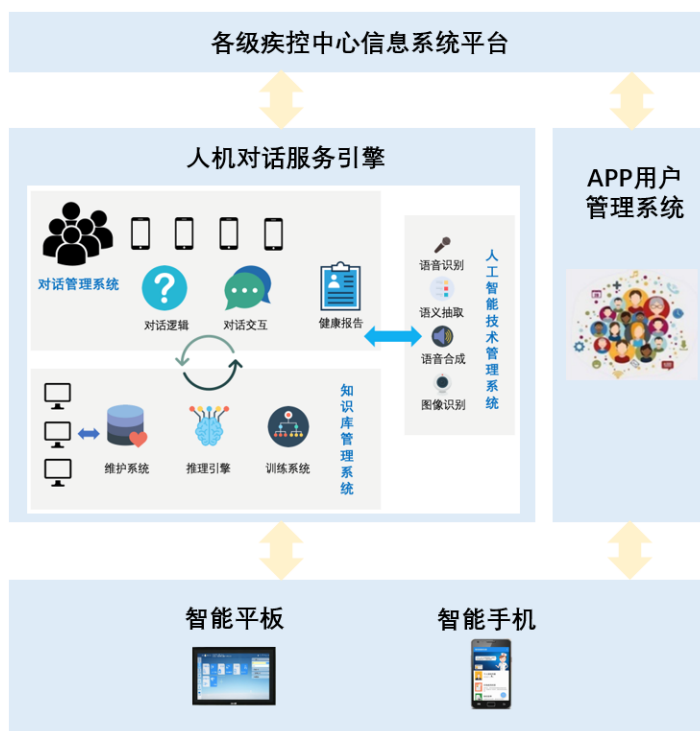


图 1 COVID-19 流行病学智能采集系统技术框架



注：图中 a 为系统交互界面，体现语音交互的形式；b 为系统交互界面，体现点选交互的形式；c 为报告查看界面

图 2 COVID-19 流行病学智能采集系统手机 APP 应用界面

### 1.1 对话管理系统

对话逻辑模块负责生成问题，通过与后台知识库管理系统交互获得新的问题。对话交互模块负责展示问题并回收答案，同时负责与 AI 技术管理系统进行交互，调用对应的接口获得识别结果。健康报告模块负责将所有对话内容转换为符合相关文件规范的调查报告或表单。

### 1.2 知识库管理系统

知识库管理系统以服务器的形式供前端访问。知识库维护系统负责流行病学知识库的创建、读取、更新和删除操作。知识库推理引擎负责返回问题结构体，供移动端对话逻辑模块调用。知识库模型训练系统利用机器学习算法，基于知识库结构，训练高效的推理算法，以最简洁的问题得到最完善的调查信息。该过程在服务器端自行完成训练。

### 1.3 AI 技术管理系统

图像识别、语音识别与合成调用相应的语音/图像服务应用程序编程接口(application programming interface, API)完成。语义抽取接口负责识别患者表述中的症状、体征、暴露史等相关内容，转换为语义结构体。

## 2 流行病学智能采集系统功能模块

COVID-19 流行病学智能采集系统集成流行病学信息采集、化验单拍照识别、调查报告和(或)表单自动生成以及疫情资料智能归档等功能于一体，支持语音输入、点选、普通输入法输入及手写识别等多模态交互方式。

### 2.1 流行病学信息采集

系统模仿流行病学调查人员以自然对话的形式进行多轮友好的引导式问答，根据每个问题的性质(单选题、多选题及开放式等)提供相对应的最便捷的交互方式(点选、语音输入等)，快速和全面收集患者发病与就诊情况、暴露史、危险因素以及密切接触者等重要信息，并进行结构化存储，为寻找传染源、梳理传播链条、锁定密切接触者、控制疫情传播范围等工作提供有力的数据支持。智能人机交互模式不仅可以最大限度地减少流行病学调查人员与被调查者不必要的接触，降低交叉感染风险，还可以提高数据准确性与规范性，提高个案调查质量。

### 2.2 化验单拍照识别

对于已完成的实验室检测，被调查者可以通过拍照上传检测结果。系统利用图像处理和文字识别技术，对用户上传的化验单进行自动识别和结构化转换，通过图像增强和自动校准技术，对用户拍摄的照片清晰度和几何畸变进行智能优化。通过光学字符识别(optical character recognition, OCR)技术和版式智能分析技术自动分析化验单表格相关字段，并将识别结果映射到结构化表格中进行保存。

### 2.3 调查报告和(或)表单自动生成

根据所采集的患者基本信息、发病与就诊以及危险因素与暴露史等内容，自动进行结构化映射，生成规范的 COVID-19 病例个案调查表或流行病学调查报告。所采集的所有数据可与调查信息系统对接，并同步至电脑端，流行病学调查人员可以在电脑端进行调阅、审核与修订，无需重复录入，优化采集流程，提高工作效率。

### 2.4 疫情资料智能归档

按相关要求对所采集数据进行自动化归档，实现对海量数据的存储管理。疫情资料具备重要的医学研究价值，通过对疫情资料的分析、查证，有助于探究突发公共卫生事件的发生

预兆和成因,规范突发公共卫生事件的处置方法,为科学防范和控制突发公共卫生事件提供科学依据<sup>[12]</sup>。

### 3 结论

COVID-19 流行病学智能采集系统打破传统的单一信息采集模式,将多模态的移动端智能人机交互方式应用到现场流行病学调查的实际场景中,解决当前流行病学调查工作中信息采集效率低、准确性差、数据利用率与安全性不足以及交叉感染风险高等问题,让战“疫”时期的信息采集变得更加智能、高效、安全且便捷。系统所开发的人机对话与信息结构化算法引擎,可以准确理解患者描述的病情信息,引导患者以自然对话的形式完成信息采集,并进行结构化映射,不仅可以在疫情期间的流行病学调查工作中发挥重要作用,在日常的导诊、预问诊、随访及满意度调查等诸多医疗场景中,也可以发挥重要的应用价值。

日益发展的 AI 技术正成为抗击 COVID-19 疫情战线上的“英勇战士”,在这场没有硝烟的战斗中“冲锋陷阵”。疫情终将被战胜,被应用到抗“疫”战斗中的新技术也将日趋成熟,让更多人获益。

### 参考文献

- [1]国家卫生健康委员会.国家卫生健康委关于新型冠状病毒肺炎暂命名事宜的通知[EB/OL].国家卫生健康委员会(2020-02-07)[2020-02-27].[http://www.gov.cn/zhengce/zhengceku/2020-02/08/content\\_5476248.htm](http://www.gov.cn/zhengce/zhengceku/2020-02/08/content_5476248.htm).
- [2]国家卫生健康委员会.国家卫生健康委关于修订新型冠状病毒肺炎英文名称事宜的通知[EB/OL].国家卫生健康委员会(2020-02-21)[2020-02-27].[http://www.gov.cn/zhengce/zhengceku/2020-02/22/content\\_5482019.htm](http://www.gov.cn/zhengce/zhengceku/2020-02/22/content_5482019.htm).
- [3]The Lancet. Emerging understandings of 2019-nCoV[J]. The Lancet, 2020, 395(10221).
- [4]国家卫生健康委员会.中华人民共和国国家卫生健康委员会公告:2020年第1号[EB/OL].国家卫生健康委员会(2020-01-21)[2020-02-27].[http://www.gov.cn/zhengce/zhengceku/2020-01/21/content\\_5471164.htm](http://www.gov.cn/zhengce/zhengceku/2020-01/21/content_5471164.htm).
- [5]李亚品,宋文静,高东旗,等.新型冠状病毒感染肺炎疫情现场流行病学调查技术及调查实施方案探讨[J/OL].解放军医学院学报:1-10(2020-02-24)[2020-03-16].<http://kns.cnki.net/KCMS/detail/10.1117.R.20200224.1324.006.html>.
- [6]人民日报评论员.坚定信心坚决打赢疫情防控阻击战[N].人民日报,2020-01-26(001).
- [7]杨芊,徐小林,赵鸿辉,等.公共卫生学科作用在新冠肺炎疫情防控中的凸显[J].治理研究,2020(2):75-80.
- [8]彭宗超,黄昊,吴洪涛,等.新冠肺炎疫情前期应急防控的“五情”大数据分析[J].治理研究,2020(2):6-20.
- [9]国家卫生健康委办公厅,国家中医药管理局办公室.关于印发新型冠状病毒肺炎诊疗方案(试行第六版):国卫办医函〔2020〕145号[EB/OL].(2020-02-18)[2020-03-12].<http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2.shtml>.
- [10]Young S, Gasic M, Thomson B, et al. POMDP-Based Statistical Spoken Dialog Systems: A Review[J]. Proceedings of the IEEE, 2013, 101(5):1160-1179.

[11]赵宇婧. 移动机器人基于自然语言的人机交互技术研究[D]. 哈尔滨:哈尔滨工程大学, 2017.

[12]计晓丽. 医院突发公共卫生事件档案管理的优化策略[J]. 兰台内外, 2019(21):68-69.

**\*通信作者:** lvhairong@tsinghua.edu.cn

**作者简介:** 王维笑, 女, (1992- ), 硕士, 从事多模态人机交互、自然语言理解及知识图谱等技术的应用研究。