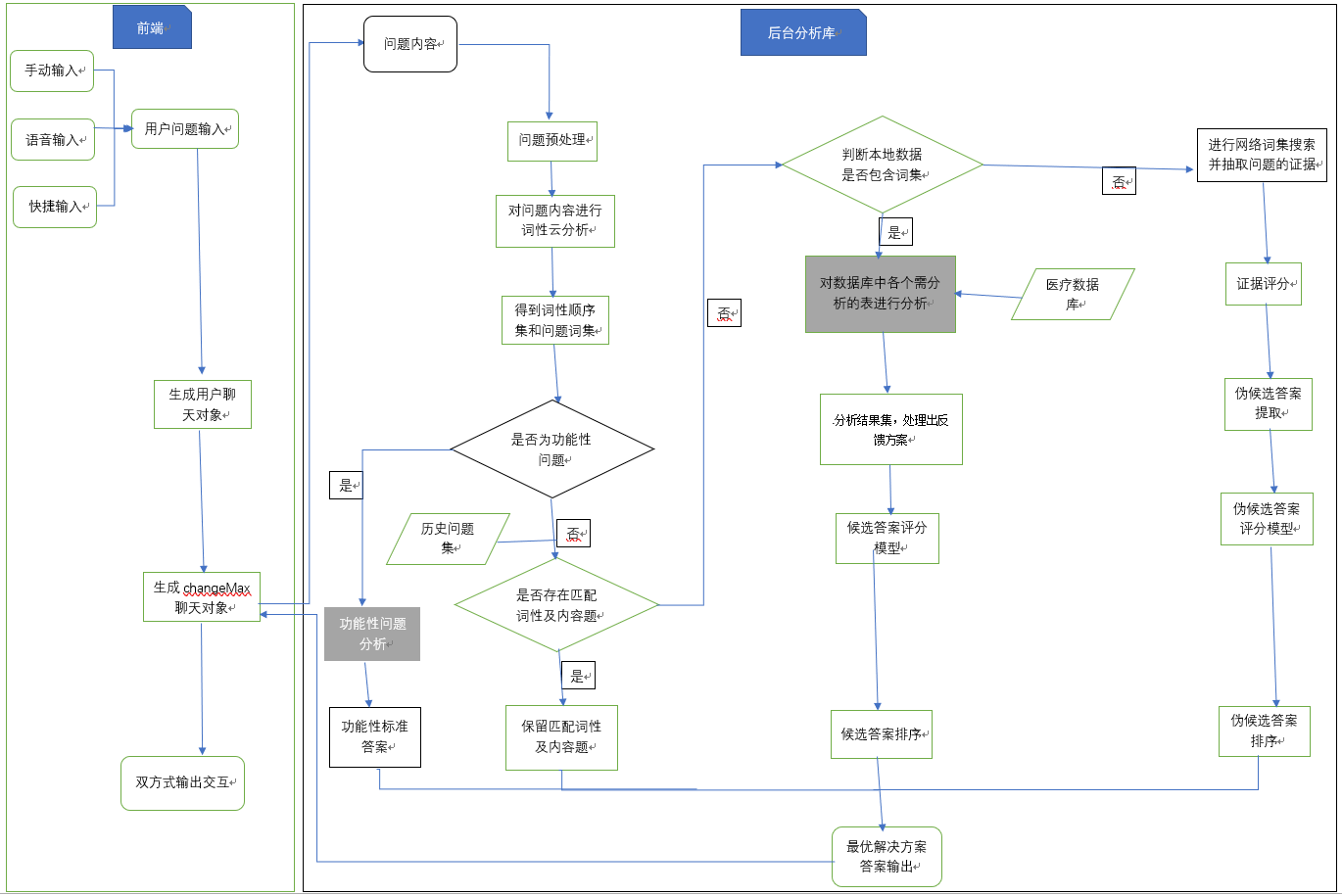
|  |
| --- |
| changeMax未来科技 |
| changMax医疗健康助手技术实现 |
| [项目工作原理] |

|  |
| --- |
| 佚名  2017-9-4 |

# changeMax医疗健康助手技术实现

changeMax医疗健康助手（以下简称为医疗助手）是一个基于android开发平台的对话式人工智能，能够通过中文自然语言对话式和文字输入式两种交互方式来实现人机交互。旨在对用户的生理、心理进行医学辅导，并给出有趣的、亲和的反馈内容。本项目说明方案，主要根据项目实现过程中实现的技术难点、分析思路，解决流程，项目成型，项目结构等多方面进行概述。

1. **项目工作原理：**



changeMax系统分析流程图

上图主要为医疗助手的实现分析流程图，主要可以划分成为前端的消息展示及交互方式、后台分析库的常规功能性问题匹配分析、历史问题匹配分析、问题内容预处理分析、大数据多维度分析、互联网数据分析五主要子模型，其中大数据多维度分析为本系统主要核心分析内容：

* **常规功能性问题匹配分析**：通过用户所提出的语言进行分析，判断是否触发常规性固定内容，医疗助手将做出常规性固定回答方案。分为反馈可点击跳转新视图方案和语言内容触发式响应方案。（如果在此分析流程中，未匹配成功数据，未生成功能性问题反馈，那么将反馈为空，将进入下一分析步骤中。如果匹配，将直接异步请求，生成反馈，终止分析。）
  + 反馈可点击跳转新视图方案

其中主要有：对用户进行定位，搜索周围医院进行导航功能（调用BaiDu开发性地图资源进行运用）；关键词性触发医疗百科模式，反馈一个强大的后台医疗数据库展现视图（医疗百科，在后面将做详细介绍）；第三方应用跳转链接，将反馈一个官方QQ客服跳转链接（点击将启动用户的腾讯qq应用，直接响应官方客服qq聊天界面）；

* + 语言内容触发式响应方案

当用户消息中出现触发关键词内容，将反馈对应随机内容。其中主要包括：用户对医疗助手说明需要看病，将启动系统性诊疗模式（系统性诊疗，在后面将做详细介绍）；用户询问“你会做什么”，医疗助手将使用随机反馈内容进行回复（小范围内随机反馈）；后台管理员登入模式；

* + 两种方案结合：

通过内置2000条专业医疗症状名称和2000条专业医疗疾病名称进行匹配分析，当用户语句中含有这4000个名称时，将对其进行可能需要反馈，当用户语句和4000个名称完全匹配时，说明用户就是需要对该医疗对象进行查询，将对用户反馈该对象的简介内容缩写，并对该反馈添加可点击性事件，用于用户了解详细信息。

* + 第三方问答系统api

通过调用第三方智能问答系统（通过发送http请求给第三方平台服务器，然后回调数据），实现多领域智能问答、中文聊天对话、场景对话、情感交互、基础知识库等多种功能。），对其进行日常正常交流反馈，配合该分析系统，实现基本医疗交流。

* **历史问题匹配分析**：在每一段人机交互过程中，系统将默认保留每段交互内容。那么在分析过程中将用户问题内容和历史问答内容进行匹配，如果在数据库历史问题集表中包含用户问题内容，那么将直接调用历史问题的反馈消息进行反馈。如果匹配，将直接异步请求，生成反馈，终止分析。（此分析阶段暂时关闭，将会导致一个bug：使得每次交互内容反馈将一步步走向唯一性。约束了反馈内容）。
* **问题内容预处理**：将用户的消息内容进行大数据分析前，先进行一段预处理过程（对用户的消息内容进行词性标注，然后进行统计，再进行医学关键词匹配，进行系统减负载分析（系统减负载分析，在后面将做出详细介绍），通过整理，进行序列化操作，将其预处理结果进行进行数据库存储，生成一段专属uid（钥匙），反馈给接下来的大数据分析模型）。
  + 语言云提供分析支持。词性标注(Part-of-speech Tagging, POS)是给句子中每个词一个词性类别的任务。 这里的词性类别可能是名词、动词、形容词或其他。 下面的句子是一个词性标注的例子。 其中，v代表动词、n代表名词、c代表连词、d代表副词、wp代表标点符号。

【国务院/ni 总理/n 李克强/nh 调研/v 上海/ns 外高桥/ns 时/n 提出/v ，/wp 支持/v 上海/ns 积极/a 探索/v 新/a 机制/n 。/wp】

将对用户的问题串内容进行语音云分析，生成“自然语言分析云分析词集”；

将“自然语言分析云分析词集”进行医学关键词筛选，对数据库中“医学词汇表”（由THUOCL：清华大学开放中文词库提供）进行匹配得到“医学关键词集”；

将“自然语言分析云分析词集”进行减负载关键词筛选，将其中含有标点符号，你我他等干扰分析词进行删减，得到“减负载关键词集”；

将以上分析集进行综合整理，生成Question对象传递。

以上为常规医疗医疗预处理，在本系统中还存在系统性诊疗预处理：

对患者综合信息进行分析处理：

1. 姓名词性标注
2. 性别词性标注
3. 年龄词性标注
4. 部位器官词性标注
5. 患者症状描述词性标注
6. 患者疾病描述词性标注

注：以上词性标注均通过语言云分析，通过开启线程池Executor框架进行词性标注。

1. 姓名减负载分析
2. 性别减负载分析
3. 年龄减负载分析
4. 部位器官减负载分析
5. 症状医学关键词集分析
6. 症状减负载关键词集分析
7. 疾病医学关键词集分析
8. 疾病减负载关键词集分析

注：减负载：对症状描述和疾病描述进行医疗关键词预分析，但是对于个别对诊断有帮助的非医疗词汇也会被舍弃：例如：早上，睡前，天冷。所以为避免对诊断减低帮助，所以对自然语言词性标注集也应保留，但是需要去除广义上无法对诊断进行帮助的词汇，反而会对数据匹配产生误解、负重的词汇。

* **大数据多维度分析：**
  + 在本数据分析阶段，存在两种大数据分析模型：常规性诊疗数据分析、系统性诊疗综合数据分析。
  + **常规性诊疗数据分析**

通过对各数据表进行对应属性模糊匹配，得到模糊匹配的id集以及对应属性，（模糊匹配：通过数据库模糊查询，匹配字符串内容）

然后进行交叉分析（交叉分析：通过分析出来的症状集合来进行进一步分析，分析疾病），增加结果可信度

基于评分模型对其进行筛选，该模型的目标是评价候选答案可信度

前者预分析出

1.所有词词集；2.减负载关键词词集；3.医疗关键词词集

获取各集对象

1.对用户的症状描述集进行精准模糊分析用户的症状(精简)

2.对用户的症状描述集进行模糊分析用户的症状(精简)

3.对用户的症状减负载集进行连续性双词模糊匹配分析用户的症状(精简)

4.对患者所描述的患处部位减负载集进行模糊分析用户的症状

5.取分值症状集合前5替换原分值症状集合，为最终结果

6.构建疾病id集

7.传递分析：症状得出疾病

8.对用户的症状描述集进行精准分析出用户的疾病集，进行加分(精简)

9.对用户的症状描述集进行模糊分析用户的疾病集，进行用户疾病加分

对用户的疾病集进行加分操作，不会产生新IDScore，不需要排序，筛选

10.对用户的疾病减负载集进行连续性双词模糊匹配分析用户的症状(精简)

（连续性双词：指连续两个关键词合成一个关键词）

11.对用户的患者部位器官描述分析用户的疾病集，进行加分

对用户的疾病集进行加分操作，不会产生新IDScore，不需要排序，筛选

12.取分值集合前10替换原分值集合

14.分析id集，处理出结果

15.反馈数据分析结果

(以上精简：在系统分析过程中，会出现大量id，且出现一定分值差距，需对其进行精简，减少后期对比，系统负担)

将id集进行数据结果分析，调用数据库中对应id疾病症状，对应疾病症状解决方案进行文本整理，反馈给用户。

* + **系统性诊疗综合数据分析**

通过对各数据表进行对应属性模糊匹配，得到模糊匹配的id集以及对应属性，然后进行交叉分析，增加结果可信度

基于评分模型对其进行筛选，该模型的目标是评价候选答案可信度

前者预分析出（通过反序列化操作，从数据库中获取到内存中）

同常规性诊疗分析，也存在一个数据精简问题，在大量数据分析打分排序中，在某子子模型分析过后，数据的分值差距明显区别，所以，对于系统的分析负载问题来说，应当对数据进行低分删减，只取部分。

1.称呼减负载关键词集，2.性别减负载关键词集，3.年龄减负载关键词集，4.部位器官减负载关键词集，5.症状医学关键词集，6.症状减负载关键词集，7.疾病医学关键词集，8.疾病减负载关键词集

注：系统分支分析结果均未排序，结果分析末端选择取值时才排序

1. 对用户的症状描述集进行精准分析用户的症状
2. 对用户的症状描述集进行模糊分析用户的症状
3. 对用户的疾病描述集进行精准分析用户的症状
4. 对用户的疾病描述集进行模糊分析用户的症状

（精准分析：指关键词和词库中疾病名称必须相同；模糊分析：值词库中疾病名称包含该关键词）

1. 对用户的症状减负载集进行连续性双词模糊匹配分析用户的症状
2. 对用户的疾病减负载集进行连续性双词模糊匹配分析用户的症状

（连续性双词：指连续两个关键词合成一个关键词）

1. 对患者所描述的患处部位减负载集进行模糊分析用户的症状
2. 取分值症状集合前300替换原分值症状集合

注意：以下分析前提均为有症状分析出来的可能病症集下分析，不能进行新增

1. 构建疾病id集
2. 传递分析：症状得出疾病
3. 对用户的症状描述集进行精准分析出用户的疾病集，进行加分
4. 对用户的症状描述集进行模糊分析出用户的疾病集，进行加分
5. 对用户的疾病描述集进行精准分析出用户的疾病集，进行加分
6. 对用户的疾病描述集进行模糊分析出用户的疾病集，进行加分
7. 对用户的患者性别描述分析用户的疾病集，进行加分
8. 对用户的患者年龄描述分析用户的疾病集，进行加分
9. 对用户的患者部位器官描述分析用户的疾病集，进行加分
10. 并发症分析，对用户的患有疾病分析用户的疾病集，进行加分
11. 取分值集合前100替换原分值集合
12. 分析id集，处理出结果

将id集进行数据结果分析，调用数据库中对应id疾病症状，对应疾病症状解决方案进行文本整理，反馈给用户。

* **互联网评分分析**：如果在以上分析中都不存在匹配数据，那么将向互联网中搜索引擎（谷歌，雅虎，百度）从互联网上抓取与该问题相关的片段，并抽取、整理出其中的正文作为该问题的支撑证据。（目前只提供搜索页面的超链接，因为该分析系统为顺序执行流程，通常来说，前面分析99%可以得出结果反馈。）

1. **主要数据结构：**

本医疗助手主要数据结构有：PersonChat（消息条例对象）、Question（用户所提出的问题）、Patient（患者信息）、Symptom（症状）、Disease（疾病）

|  |  |  |
| --- | --- | --- |
|  | 属性 | 类型 |
| 消息条例 | Id | Int |
| Uid | String |
| 消息内容 | String |
| 是否为用户发送 | Boolean |
| 消息发送时间 | String |
| 消息音频内容 | String |
| 问题 | Id | Int |
| Uid | String |
| 问题内容 | String |
| 答案内容 | String |
| 问题生成时间 | String |
| 所有词集 | List |
| 减负载关键词次词集 | List |
|  | 医学关键词集 | List |
| 患者信息 | Id | Int |
| Uid | String |
| 患者信息生成时间 | String |
| 患者名称 | String |
| 患者性别 | String |
| 患者年龄段 | String |
| 患者患疾部位 | String |
| 患者症状描述 | String |
| 患者疾病描述 | String |
| 症状 | 症状id | Int |
| 症状名称 | String |
| 症状简介 | String |
| 症状起因 | String |
| 可能伴随疾病集 | List |
| 诊断详述内容 | String |
| 建议就诊科室 | String |
| 疾病 | 疾病id | Int |
| 疾病名称 | String |
| 疾病别名 | String |
| 疾病简介 | String |
| 疾病发病部位集 | String |
| 疾病传染性 | String |
| 疾病多发人群 | String |
| 疾病早期症状 | String |
| 疾病晚期症状 | String |
| 疾病相关症状 | String |
| 疾病相关症状介绍 | String |
| 疾病并发症 | String |
| 疾病并发症介绍 | String |
| 疾病就诊科室 | String |
| 疾病治愈率 | String |

**三 项目关键技术**

在本项目中，运用三层架构，所以可分为前端表示层、业务逻辑层、数据层。

前端表示层：对话式人工智能交互。关键词分析处理显示

业务逻辑层：异步处理用户消息反馈系统、高效线程死循环监听异步结果、系统性综合数据收集系统、自定义随机反馈库、第三方高智能度问答系统、自然语言处理、医疗数据多维度计分分析、第三方地图接入分析、后台响应任务、序列化和反序列化技术；

数据层：医疗大数据库。

* 对话式人工智能交互：对话式交互，顾名思义，就是像两个人对话一样的交互方式。可以是文字的，也可以是语音的。我们每天都用的微信，你说一句，我说一句，就是典型的对话式交互。本医疗助手中将人工智能和用户运用该交互方式。从键盘鼠标，触屏、到语音，此人机交互的操作过程变得更加简单高效，使用人群扩展至儿童和老人，人机交互门槛降低，人机交互方式变得更加丰富。交互方式也越来越自然，尤其是语音时代的人机交互，体现出了更大价值：人们获取信息的方式更加便捷，获取信息的时间、空间得以无限延展，受交互成本、设备的影响减少。

实时语音转写（Real-time ASR）：基于深度全序列卷积神经网络框架，通过 WebSocket 协议，建立应用与语言转写核心引擎的长连接，将音频流数据实时转换成文字流数据结果

语音识别技术(Auto Speech Recognize，简称ASR)所要解决的问题是让机器能够“听懂”人类的语音，将语音中包含的文字信息“提取”出来，相当于给机器安装上“耳朵”，使其具备“能听”的功能。

* 关键词分析处理显示：后台分析出的反馈结果，为挺高用户的可读性，对关键词的标识符进行选择性删减；为提升用户的app体验性，对一些非礼貌用语进行选择性屏蔽。通过字符匹配问题
* 异步处理用户消息反馈系统：通过用户每次主动发起交互请求，生成对应的反馈需求，向分析系统发送一个异步消息请求，由系统进行分析用户消息，反馈结果，生成异步处理结果。
* 高效线程死循环监听异步结果：由系统生成异步消息请求的同时，开辟新线程，通过死循环监听异步结果，只要异步线程一完成，那么将退出该新线程，同时将异步线程结果处理反馈给前端交互。
* 系统性综合数据收集系统：通过智能交互方式，向用户一句一句的收集患者信息，对其进行判断，是否为有意义信息，然后进行序列化操作，存储到数据库中。供后期系统性诊疗综合综合数据分析。
* 自定义随机反馈库：实现了机器更拟人化，将功能性语句反馈实现随机性。
* 第三方高智能度问答系统：调用图灵机器人api接口，将其和本分析处理系统整合，实现日常人机交互体验。
* 自然语言处理：(N LP , Natural Language Processing)是使用自然语言同计算机进行通讯的技术, 因为处理自然语言的关键是要让计算机“理解”自然语言,所以自然语言处理又叫做自然语言理解(NLU ,Natural Language Understanding), 也称为计算语言学(Computational Ling uistics)。一方面它是语言信息处理的一个分支 , 另一方面它是人工智能(AI , Artificial Intelligence)的核心课题之一 。
* 医疗数据多维度计分分析：向用户收集患者信息，进行语言云词性预分析，然后进行医学关键词筛选，减负载筛选（筛选方式由患者的各种信息种类进行对应筛选），然后进行医疗症状疾病匹配，进行多维度匹配分析，按照计分制进行排序，分值最高者为准诊疗候选结果。
* 第三方地图接入分析：通过调用BaiDu开放性平台的地图数据，实现自定义地图定位导航，为用户提供就近医疗机构，提升就医效率。
* 后台响应任务：当用户使用医疗助手开始起系统后台进行计时任务，如果用户长时间未对本医疗助手进行操作，那么医疗系统将主动向用户发起交互请求，如果连续5次交互请求，用户都不响应，那么系统将自动回收，退出应用，对用户内存回收机制一种帮助，降低用户平台的功耗。
* 序列化和反序列化技术：在数据传输过程中，通过将数据保存到数据库中，生成一个32位uid，将内存中的用户信息保存到本地硬盘中（序列化），在使用的过程中，通过32位uid来提取用户信息（反序列化）。
* 医疗全面数据库：由症状疾病百科提供，具有2000个·疾病详细信息，2000个临床症状，上万种疾病，症状关系数据等，每种疾病具有对应的简介、起因，早晚期症状，应对策略等全面数据，在医疗百科中，疾病和症状环环相扣，环环跳转。实现医疗数据的整合性。

**四 代码实现**

|  |  |  |
| --- | --- | --- |
| **代码所在包** | **分类** | **内容** |
|  |  |  |
|  | **Admin** | **管理员界面** |
| **Animation** | **启动app动画** |
| **App Introduction** | **首次启动app应用介绍** |
| **Map** | **App导航界面** |
| **Mea** | **医疗百科界面** |
| **intro** | **App介绍界面** |
| **Web** | **App自定义浏览网页** |
| **MainActivity** | **App主界面** |

|  |  |  |
| --- | --- | --- |
|  | **Dao\_external** | **外部数据库连接对象** |
| **Dao\_internal** | **内部数据库连接对象** |

|  |  |  |
| --- | --- | --- |
|  | **DataAnalysis** | **数据分析层** |
| **FunctionAnalysis** | **常规功能分析层** |
| **HistoryQuestionMatch** | **历史问题分析层** |
| **Internet** | **互联网数据分析** |
| **LtpCloud** | **语言云分析层** |
| **Pretreatment** | **预分析层** |
| **MSEntrance** | **医疗分析入口** |

|  |  |  |
| --- | --- | --- |
|  | **Object\_external** | **外部数据库实体类** |
| **Object\_internal** | **内部数据库实体类** |

|  |  |  |
| --- | --- | --- |
|  | **Service\_external** | **外部数据库操作** |
| **Service\_internal** | **内部数据库操作** |

|  |  |  |
| --- | --- | --- |
|  | **Adapter\_util** | **消息适配器工具类** |
| **File\_util** | **文件操作工具类** |
| **http\_util** | **网页访问工具类** |
| **keyBoardUtils** | **软键盘操作工具类** |
| **RegisteredNoUtil** | **挂号工具类** |
| **ReturnRandomWord** | **系统随机语言工具类** |
| **Tool** | **32位uid工具类** |
| **Overlay\_util** | **地图覆盖工具类** |
| **Permissions** | **权限工具类** |
| **Plugin\_util** | **Loading工具类** |
| **Runnable\_util** | **异步医疗分析工具类** |
| **Speech\_util** | **对话式工具类** |
| **Sqlite\_util** | **数据库操作工具类** |
| **UI\_util** | **界面设计工具类** |

|  |  |  |
| --- | --- | --- |
|  | **spashActivity** | **启动图界面** |

**五. 项目总结及可行性分析**

首先，行业存在持续痛点，这就说明有商业机会和需求，如果能为之提供解决方案，自然有人愿意买单。在传统医疗行业中，所存在的弊端很多，那么在短时间内，这些弊端，通过常规慢慢填上是很难。所以，人工智能医疗将可以颠覆传统医疗，那么其各方面发展不可估量。可行性发展背景：

1. 国家政策全力支持发展
2. 传统医疗时效
3. 传统医疗资源配置不合理，不均
4. 国民医疗意识淡薄
5. 医疗资源短缺

……

对于人工智能的发展，必不可少：算力，数据和算法。那么在现阶段发展迅速的现在，就差医疗大数据的引爆。而数据的获取，就差一个让人们愿意去使用医疗人工智能产品，愿意将人工智能医疗替代传统医疗的思想灌输的一个突破口。而这个突破口谁创造了，那么谁将拥有主导力量，人们将信赖这个人工智能医疗机构，那么他将开启一个新时代，将取得这个机遇。那么突破口就是移动端应用。

智能设备的普及率现状很高，在2017年是数据显示：中国达到68%。所以说让用户能够很简单的体验到、很乐意去接触医疗人工智能的产物，就得通过移动设备。

再者，据数据分析，现30-50岁人群更加愿意主动接触健康生活的知识，但是在网络流传的信息不具有科学性，规范性，很容易出现谣言，恶传。所以，有一个权威的医疗常识普及平台也是很有必要。

通过changeMax健康医疗助手的医疗常识传播，普及，通过健康助手的数据分析，实现用户的简单诊疗，让基层用户很简单的就可以体验到人工智能的产物，让人工智能医疗进入每一个人的眼球，这样将大大推进人工智能医疗的发展。