

# 介绍 AIDoctor 整体架构

徐志城

2021 年 4 月 11 日

## 1 前后端交互

后端提供了一个核心接口 `get-response-dict`，这个接口的输入为用户输入的字符串，返回为一个字典，这个字典的 `key` 暂时只有“`msg`”，对应 `value` 为服务端回复给客户端的字符串。未来可以回复更多的有效信息，如推测用户心情、用户性格等。

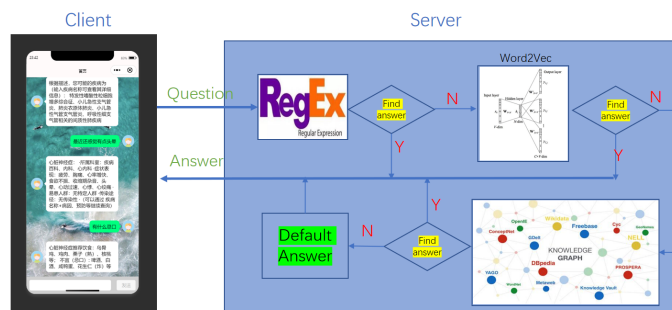


图 1: 整体架构图

## 2 后端 AI 模型介绍

从设计思路来说，客服端采用了类似责任链模式的方式来整合模型。链以及链的节点都实现了 search-answer 这个接口，接口输入为客户端输入，输出为将来会在服务端输出的字符串。

在运行时，只要链的一个节点找到 answer，就立即返回，就不会将客户端输入传入下一个节点。靠前的节点都是组合了响应速度快，原理简单的模型，目的是为了使得回复尽可能符合人的交流习惯。接下来就是按节点顺序介绍所用到的三个模型。

- 第一个是基于正则表达式的模板匹配模型。如果用户输入符合模板中的正则表达式，则随机选中指定的回答作为回复。
- 第二个是基于 word2vec 模型的句子相似度匹配模型。这个模型相当于第一个模型的优化。首先准备好语料库，一列为 Q(Question) 对应的另外一列为 A(Answer)。然后由于 word2vec 模型只能将词转化为对应向量，所以先对句子分词然后采用最经典的求和平均的方法得到句子向量。然后出于性能考虑，我事先准备了一个字典，key 为词，value 为包含该词的句子的索引 (句子属于 Q 列)。将客户端输入分词后，首先找到包含客户端词最多的前三名句子 (借助于那个字典)，然后获取这三个句子向量，然后求该句子向量与客户端输入的句子向量的余弦相似度，如果超过阈值，就用 A 列对应语句作为回复。
- 第三个为以知识图谱为核心的模型。这个模型通过正则表达式实现意图识别 (询问疾病? 描述症状? 询问忌口? ...), 使用 spacy 预训练模型实现命名实体识别 (各种医疗名称), 通过提取到的实体进行知识图谱召回, 通过存储上下文提取到的实体实现多轮对话, 在特定语境下 (如询问疾病相关信息是) 提供 pending action。知识图谱主要组成成分为 Node 和 Edge, Node 为实体 (如疾病, 症状, 推荐药物, 检查 ...), Edge 为实体间的关系, 通过字典 (key 为 Node1, value 为 a list of Node2) 存储, 本次知识图谱的空间复杂度为  $O(n^2)$  这是为了需要优化的地方。本应用最常遇到的情况为: 用户描述症状, 服务端召回对应的疾病及疾病相关描述。内部原理为, 通过症状与疾病所对应的 Edge 文件 (本质为字典), 找到多个症状所对应的疾病的交集, 即为召回的数据。

### 3 最终成果展示

