

# 智能辅助诊疗平台设计与探索\*

李强 吴裕雄 古国照 陈锡林 陈晔

(广东东软学院计算机学院 佛山 528225)

〔摘要〕 设计开发智能医生助理平台,应用贝叶斯分类、人工神经网络和关联规则算法,分别实现智能分诊、智能诊断以及智能优选处方等医疗辅助功能,从而改善就医环境,提升医疗服务水平。

〔关键词〕 智慧医疗;贝叶斯算法;神经网络算法;关联规则算法

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.11.007

**Design and Exploration of Intelligent Auxiliary Diagnosis and Treatment Platform** LI Qiang, WU Yuxiong, GU Guozhao, CHEN Xilin, CHEN Ye, Department of Computer Science, Guangdong Neusoft University, Foshan 528225, China

〔Abstract〕 The intelligent doctor assistant platform has been designed. It realizes intelligent triage, intelligent diagnosis and intelligent preferred prescription and other medical auxiliary functions with Bayesian classification, artificial neural network and association rule algorithms in order to improve hospital environment and medical service level.

〔Keywords〕 intelligent medical; bayesian algorithm; neural network algorithm; association rule algorithm

## 1 引言

随着我国互联网迅速发展,“互联网+”、大数据和人工智能(Artificial Intelligence, AI)等技术越来越多地介入到医疗领域,智慧医疗是最近兴起的专有医疗名词,通过打造健康档案区域医疗信息平台,利用最先进的物联网技术,实现患者与医务人员、医疗机构、医疗设备之间的互动,逐步达到信息化。统计数据表明人工智能针对一些疾病的诊断

效果已经达到甚至超越传统的人工治疗方案,对于提升多种疾病的筛查和诊断效率作用明显,在一定程度上缓解医生少、患者多的看病难问题。在现有系统中,有致力于特殊大病诊断治疗服务的产品,如推想科技的推想人工智能致力于进行肺癌辅助筛查,腾讯医疗人工智能实验室推出帕金森病AI辅助诊断新技术等;同时也有一些软件提供特定精准诊断服务,如深圳医软科技舌诊功能,利用人工智能技术,通过医生线上观察患者舌头和得到舌的纹理信息运用算法进行分析,从而得出更科学、准确的结论;此外有大量的在线医生平台提供在线诊断服务,如丁香网、好大夫和好医生等知名网站<sup>[1]</sup>。前者技术相对专业,准确度很高,优势有时同时也是主要缺点,专业针对性强,受众面较窄,费用高昂;对于后者来说,基本上是利用了“互联网+”平台,从诊断到后续服务都需要实体医生和助理进行跟踪处理,就诊费用较高,很难从根本上解决看

〔收稿日期〕 2019-05-06

〔作者简介〕 李强,硕士,副教授,发表论文15篇。

〔基金项目〕 2017年度广东省科技计划协同创新与平台环境建设项目“基于大数据环境下的协同医疗数据平台研发构建”(项目编号:2017A040406001);2016年度广东省佛山市科技创新项目(项目编号:2016AG100792)。

病难问题。综合以上因素，本研究课题组基于大数据和人工智能技术设计开发智能医生助理平台，旨在为公众提供辅助智能医疗平台，从一定程度上改善就医环境、提升治疗服务水平<sup>[2]</sup>。

## 2 设计思路

### 2.1 概述

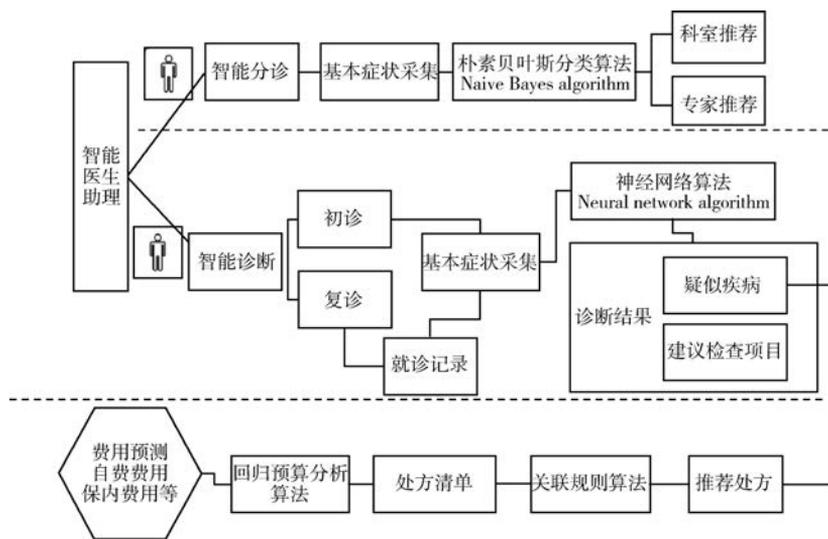


图 1 智能医生助理系统框架

### 2.2 主要功能

2.2.1 智能分诊 基于大数据，使用朴素贝叶斯分类算法建立模型，结合初诊患者提供的常见症状，由于不同患者描述症状的用词不同，需使用中文分词功能实现关键字的提取，最终提供建议就医的门诊科室，同时为就医者就近推荐本专科擅长的医院和优选医生供参考选择<sup>[3-4]</sup>。

2.2.2 智能诊断处方 分为初诊和复诊。对于不方便到医院就医或者暂时挂不到号的患者，应用人工神经网络算法，通过大数据分析模型结合不同诊室问诊症状情况进行模拟临床诊断推荐，判断就医对象的健康情况，可以得出 3 种不同诊断结果：没有明显患病的亚健康状态，系统提供健身养生建议方案；对于不确定的疑似患病和基本确定的患病给出病症名称；根据确定的疾病，通过关联规则算法，实现每种疾病的有效处方推荐，根据推荐指数或费用预测进行排序，同时根据需要推荐医院进行

智能医生助理平台目前主要提供两大核心功能，即诊前的智能分诊和智能诊断处方，此外还包括历史医疗数据统计分析报告，限于篇幅原因不再赘述。本系统平台所需大数据来源为在线医疗网站，如好大夫、丁香网、好医生等，通过爬取网站公开的大量疾病病症特征以及诊疗方案作为数据基础建立模型，其测试准确率达到 66%。系统框架，见图 1。

具体项目检查。对于含有处方药物的处方，基于患者安全考虑，建议尽快到医院就诊<sup>[5-6]</sup>。

## 3 核心算法

### 3.1 智能分诊

主要根据就诊用户输入的一些病理特征，系统智能推荐一个就诊科室，同时就近推荐相关知名医院和优选医生。算法主要由采用拆词、创建单词向量表、单词匹配、统计相关单词词频等模块构成。算法首先将系统内所有门诊类别中具体科室的相关疾病的每一种特征进行提取。根据提取到的疾病特征进行拆词，拆成以每一个独立字作为键值，对应的值是键值的词频。根据词频键值对进行进一步处理得到系统所有疾病的单词向量表。将对应的门诊和其中具体某个科室通过字典的形式添加到其所对应的疾病单词向量表中。将用户输入的病理特征进行拆词处理，随后与前一步骤的字典单词向量表中

的每个门诊中某个科室的具体疾病匹配。根据匹配度的高低决定该用户应前往哪个门诊中哪个科室的具体某个疾病科就诊<sup>[7]</sup>。

### 3.2 智能诊断

定义神经网络的输入、参数和输出,定义前向传播过程。Deep\_1=5:设置第1层神经网络的深度为5;Biases\_1=5:设置第1层神经网络的偏置维度为5\*1,以正态分布概率生成计算第1层神经网络的输出;Deep\_2=1:设置第2层神经网络的深度为1(如果转one\_hot编码则为3,为方便取出具体的预测值来进行以下损失函数的定义并没有转,如果转码最后还要使用np.argmax函数取值,所以这一层神经网络权值维度是5\*1,与5个病理特征相对应,也方便最后使用这5个权值与对应输入的特征值加权累加得出是否患病的预测);Biases\_1=1:设置第1层神经网络的偏置维度为1\*1,以正态分布概率生成计算第2层神经网络的输出。定义损失函数及反向传播方法。Loss\_mse:算法使用的损失函数是均方误差。Train\_step:算法使用的优化方法是动量优化算法MomentumOptimizer,学习率是0.0001,衰减率decay是0.9,这个值越大模型越趋于稳定,所以设置值接近于1。生成会话,训练STEPS轮。STEPS=2000:代表算法迭代2000次,每次训练8条数据。等待神经网络训练完后可得到一个训练好的特征权值,也就是最后一层神经网络层训练后的权值,其维度与每个特征对应是5\*1<sup>[10]</sup>。神经网络算法将进一步处理好的目标权值通过函数返回。系统计算出诊断结果。目前系统可提供多达22种疾病的辅助诊断功能(多次测试结果算法预测准确率约70%左右)。核心代码如下:

```
tx = tf.placeholder (tf.float32, [None, 5])
ty = tf.placeholder (tf.float32, [None])
with tf.variable_scope (weights1):
    tw = tf.Variable (tf.truncated_normal (shape = [5],
mean = 1.0, stddev = 0.1), name = "weights1")
    tf.summary.histogram (weights1', tw)
    with tf.variable_scope (weights2):
        sw = tf.nn.softmax (tw, name = "weights2")
        tf.summary.histogram (weights2', sw)
    with tf.variable_scope (res):
```

```
res = tf.reduce_sum ( (tx * sw), name = "res")
tf.summary.histogram (res', res)
with tf.variable_scope (loss):
    loss = tf.abs (res - ty, name = "loss")
    tf.summary.histogram (loss', loss)
optimizer = tf.train.RMSPropOptimizer (0.1)
train = optimizer.minimize (loss)
merged = tf.summary.merge_all ()
batch_size = 20
init = tf.global_variables_initializer ()
with tf.Session () as sess:
    sess.run (init)
    writer = tf.summary.FileWriter (C:/p1', sess.graph)
    for i in range (1):
        for j in range (int (len (X1) /batch_size) - 1):
            s = (j * batch_size) % (len (X1) - batch_size)
            e = s + batch_size
            back1 = X1 [s: e]
            back2 = Y1 [s: e]
            result = sess.run ([train, tx, ty, tw, sw, res, loss],
feed_dict = {tx: back1, ty: back2})
            resu, _ = sess.run (merged, feed_dict = {tx: back1,
ty: back2})
            writer.add_summary (resu, i)
        writer.close ()
```

### 3.3 智能优选处方

Apriori算法是最有影响的挖掘布尔关联规则频繁项集,基于两阶段频繁思想的递推算法。该关联规则在分类上属于单维、单层、布尔关联规则。所有支持度大于最小支持度的项集称为频繁项集,简称频繁集<sup>[8]</sup>。关联规则的产生主要分为两个过程:一是基于数据产生频繁项集,对于每一个频繁项集来说,该项集在数据集中所出现的频率必须满足一定要求。项集在数据集中出现频率称之为支持度,所需满足的要求称之为最小支持度。支持度可以表示关联规则是否普遍适用,通过设置最小支持度使最终的关联规则具有更广的适用面,也使最终获得的结果更具有价值<sup>[9]</sup>。二是产生关联规则,通过所发现的频繁项集产生规则,计算每个规则的置信度,规则的置信度若满足最小置信度,那么这条规则就称之为关联规则。置信度的大小表示使用该关联规则对数据进行推理的准确性。运用关联规则算法于医疗大数据,发现医

疗数据中疾病的有效处方关联,同时根据置信度可推出多个有效处方并设置其推荐指数。

## 4 实现

### 4.1 智能分诊

对用户输入的症状数量不作限制,越详细越准确,见图 2。



图 2 智能分诊示例

### 4.2 智能诊断

用户选择科室和门诊后,针对具体门诊类别,回答 1-5 个问题以描述主要的症状情况,系统将初步判断结果,可能为亚健康状态或疑似病例。

### 4.3 处方优选

根据初步诊断的疾病,系统显示出多个可供选择的处方并根据预测费用或推荐指数进行排序。

## 5 结语

智能医生助理平台能为患者带来便捷和可靠的

服务,有助于推动人工智能化医疗的建设,也将助力于我国建设覆盖城乡居民的基本医疗卫生制度,为公众提供安全、有效、方便、价廉的医疗卫生服务。智能化医疗信息系统以前所未有的速度发展,将对医疗卫生行业乃至全人类的健康产生重大影响<sup>[10]</sup>。此外基于大数据技术构建的机器学习模型能够提供分诊及初步诊断功能,但由于数据来源于互联网上的可爬取内容,目前无法真正获取权威和官方的基础数据,对于分诊和智能诊断的准确率有一定影响,因此平台只能起到辅助作用,最终结果需听取医生的专业建议。该平台已基本从需求框架到技术模型实现智能就医的功能和条件,希望研究成果能得到业内认可和改进并得到相关数据支撑,为实现真正的“智能医生”提供基础条件。

## 参考文献

- 1 袁静,张华,刘长兴. 基于智能导医的数字化门诊建设[J]. 中国医疗设备, 2014 (9): 81-83.
- 2 许筱毓,郝婧宇. 基于大数据的 His 系统——智能自助分诊及预知病情的设计研究及开发[J]. 数码世界, 2018 (9): 79.
- 3 梁聪. 基于改进贝叶斯网络的健康大数据分类模型[J]. 计算机与现代化, 2017 (12): 61-64.
- 4 王双成,高瑞,杜瑞杰. 基于高斯核函数的朴素贝叶斯分类器依赖扩展[J]. 控制与决策, 2015 (12): 2280-2284.
- 5 程广,王晓峰. 基于 MapReduce 的并行关联规则增量更新算法[J]. 计算机工程, 2016 (2): 21-25, 32.
- 6 崔妍,包志. 关联规则挖掘综述[J]. 计算机应用研究, 2016, 33 (2): 330-334.
- 7 曹通. 一种基于语义分析的热点新闻发现方法[J]. 计算机与现代化, 2017 (6): 30-33.
- 8 李庆鹏,张龙军,耿新元. I-Apriori: 一种基于 Spark 平台的改进 Apriori 算法[J]. 科学技术与工程, 2017, 27 (12): 243-248.
- 9 谢志明,王鹏. 基于 MapReduce 架构的并行矩阵 Apriori 算法[J]. 计算机应用研究, 2017 (2): 401-404.
- 10 胡红濮. 我国全民健康信息化发展历程及展望[J]. 医学信息学杂志, 2019, 40 (7): 2-6.