文章编号：2096-1472(2017)-05-39-03

基于遗传算法的药物疗效评价模型研究

曹小凤
（太原工业学院计算机工程系，山西 太原 030008）

摘　要：高血压是引起冠心病、中风、心力衰竭等疾病的重要风险因素，并已成为影响人类健康的重要疾病，对患者进行药物治疗是控制血压的常见手段。本文通过对高血压患者的服药情况及血压值进行分析，建立两者之间的(β,beta)分布评价模型，利用遗传算法对该模型进行优化。通过与线性拟合相比较，表明该模型能很好地预见患者的血压情况。

关键词：高血压；β分布；遗传算法；交叉验证
中图分类号：TP18　　**文献标识码**：A

Research on the Evaluation Model of Drug Efficacy Based on the Genetic Algorithm

CAO Xiaofeng
(*Department of Computer Engineering,Taiyuan Institute of Technology,Taiyuan* 030008,*China*)

Abstract: Hypertension is an important risk factor for coronary heart disease,stroke,heart failure and other diseases,and has become an important disease affecting human health.Through the analysis of the medication and the blood pressure,the paper establishes the beta distribution model,and adopts the genetic algorithm to optimize the model.The comparison results with the linear fitting algorithm show that this model can well predict the blood pressure of patients.

Keywords:hypertension;beta distribution;genetic algorithm;cross-validation

1　引言(Introduction)

心脑血管疾病是全球威胁人类健康的重大疾病，其发病和死因一半以上与高血压有关，同时居民脑卒中和冠心病发病最重要的危险因素也是高血压，所以控制高血压是防治心脑血管病、脑卒中和冠心病的关键。《中国高血压防治指南》指出，患者有效地控制血压可以减少心脑血管及其他并发症的发生，从而提高生存质量。医学研究表明：对高血压进行早期预防和早期稳定的治疗及健康的生活方式，可使75%的高血压及并发症得到预防和控制。Wilbert S.Aronow[1]对治疗老年人高血压情况进行研究，发现抗高血压药物治疗能有效降低心血管、脑卒中死亡率；Huan M.Nguyen、Karlene Ma[4]等人在研究治疗成人严重高血压中发现，氯维地平能有效控制术期的血压，且副作用很小。目前高血压病人的知晓率、控制率和治疗率一直处于较低的水平，本文从提高高血压患者的控制率方面，构建高血压药物疗效评价模型，发现持续用药与血压的关系，为医生开药、病人持续用药提供指导，增强病人的可持续用药程度，从而提高患者的控制率和治疗率，提高患者的生活水平。

2　高血压水平定义(Definition of hypertension)

对高血压的定义，国内外的标准稍有不同，本文的研究采用文献[3]中对高血压水平的定义，即在不使用降压药物的情况下，非同日3次测量，收缩压SBP≥140mmHg和(或)舒张压DBP≥90mmHg；患者有既往高血压史，而目前正在使用降压药物的，即使当前的血压值低于140/90mmHg，也判定为患有高血压。如果收缩压SBP≥140mmHg，舒张压DBP<90mmHg可以判定为单纯的收缩期高血压。收缩压升高比舒张压升高给人带来的影响更大，因此本文主要是研究收缩压与用药时间的关系，其值与相应关系的描述如表1所示。

表1 收缩压值与状态对应关系

Tab.1 The relation between the value of SBP and state

状态	收缩压
轻度高血压(bpl)	90—129 mmHg
中度高血压(bpm)	130—159 mmHg
高血压(bph)	≥160 mmHg

3 模型构建(Model constructing)

本文的研究数据来源于山西省某医院，通过对数据清洗、筛选，研究发现患者的收缩压值与服药时间近似服从β分布。由于β分布根据参数的不同可呈现出多种不同的形状，有很好地适应性，因此选择此模型对收缩压和服药时间进行研究。

3.1 β分布函数

在概率论与数理统计中，β分布是指定义在(0,1)区间的连续分布，里面包含两个参数(a>0,b>0)。β分布的密度函数为：

$$
\begin{aligned}
f(x;a,b) &= \frac{x^{a-1}(1-x)^{b-1}}{\int_0^1 u^{a-1}(1-u)^{b-1}du} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \\
&= \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}
\end{aligned}
\tag{1}
$$

此处，$f(x)$代表收缩压值，代表持续用药时间，由于β分布的定义域限制，根据Jim Warren等人在文献[2]中的研究，现将本文中的用药时间转换为药物所占比例(Medication Possession Ratio,MPR)。

3.2 计算MPR值

MPR描述的是患者持续用药时间占所评估时间(Evaluation Period,EP)的比例，此处研究的评估时间是一年，它的值可以反映患者的药物坚持程度。根据Thusitha Mabotuwana、Jim Warren等人在文献中所提出的方法，其时间关系如图1所示。

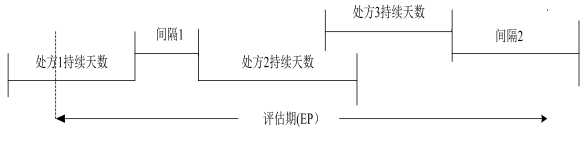

图1 MPR计算的时间关系图

Fig.1 Time relation diagram for MPR calculation

间隔1是指患者从服用处方1结束到开始服用处方2的一段间隔期。患者在间隔1内没有药物供给。MPR的具体计算如式(2)所示。

$$
MPR = \frac{EP\text{天数} - \text{总间隔天数}}{EP\text{天数}} \times 100\% \tag{2}
$$

4 数据处理(Data processing)

4.1 数据模糊化

由于医疗数据内在的数据复杂性，通常用户查询请求中具有模糊性或不确定性，许多研究工作处理语义数据中的不确定性和模糊检索，引入了模糊描述逻辑。模糊模型利用IF-THNE形式的规则来描述系统，使模糊系统能够解决传统的数学方法不适宜的一些问题，模糊系统的输入和输出都是模糊量，但是在实际的工程应用中，测量资料几乎都是精确量，如速度、位置等。将精确量输入模糊系统，需要首先将其模糊化，具体过程如下：

①测量：测量输入变量的值。

②变换：将测量值进行尺度变换，使其转换到各自的论域范围。

③模糊化：将已经变换到论域范围的输入量进行模糊处理，使原先精确的输入量变成模糊量，并用相应的模糊集合来表示。

本文研究的是SBP与MPR之间的关系，根据前面的模型得出的数据值应该符合SBP的取值范围，因此，本文需要对数据进行尺度变换，将模型得出的数据值通过模糊化为SBP值所符合的范围，即是转换为第一部分所列出的范围。数据模糊化过程中用到的三个规则如下：

$$bph=u1/Maxu1*70+160 \tag{3}$$

$$bpm=u2/Maxu2*30+130 \tag{4}$$

$$bpa=u3/Maxu3*40+90 \tag{5}$$

其中，ui/Max(ui)表示属于该规则的隶属度。通过式(3)—式(5)将相应的精确值模糊化为对应的规则区间。

4.2 数据去模糊化

本文的最终目的是确定三个范围的SBP与MPR的关系，根据我们得到的模型，每一个MPR值对应于三个属于不同范围的SBP值，也就是有三条不同的曲线。因此需要根据不同范围的SBP值所占的权重对数据进行去模糊化处理，本文中我们用到的数据去模糊化模型为TSK模型[5]。

在TSK模型中，每个规则的输出是输入变量加一个常数项的线性组合，它最终得到一个清晰量。最终的输出是每个

规则输出的加权平均，如式(6)所示：

$$y=\frac{\sum_{i=1}^{u} y^i \times w_i}{\sum_{i=1}^{u} w_i} \tag{6}$$

其中，u为模糊规则数，W_i为相应规则的隶属度。对数据进行去模糊化后，利用得到的期望值与实际值之间的误差对模型进行评估。下面我们将选择算法对模型的参数进行训练，确定最佳模型作为SBP的评价模型。

5 基于遗传算法的参数优化设计(Parameter optimization design based on genetic algorithm)

5.1 遗传算法

对于复杂问题或缺乏先验知识的问题，传统的基于专家经验的建模方法难以实现，而遗传算法具有很强的全局搜索能力，因此，本文提出了将遗传算法用于模糊建模。遗传算法有其强大的全局搜索能力，能够同时处理群体中的多个个体，也即对搜索空间中的多个解进行评估，减少了陷入局部最优解的风险，遗传算法利用进化过程获得的信息自行组织搜索时，适应度大的个体具有较高的生存概率，并获得更适应环境的基因结构。本文主要通过遗传算法对上述模型进行参数优化，找出问题的最优解，确定最佳模型。

5.2 参数评估

本文所提出的参数优化算法的思想是将模型参数作为优化对象，通过遗传算法迭代并使用适应值评价函数对求解性能进行评价，通过交叉、变异不断进行解的优化。

评价最终模型的标准，通常根据预测数据与原始数据之间的误差来判断。均方根误差又叫标准误差，它是观测值与真值偏差的平方和观测次数n比值的平方根。在实验中，观测次数n总是有限的，真值只能用最佳值来代替，标准误差对一组观测中的特大或特小误差反应非常敏感，所以，标准误差能很好地反映出测量的准确度。因此，本实验选择则标准误差最为评估标准，其计算公式如式(7)所示：

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n} w_i (y_i-\overline{y_i})^2} \tag{7}$$

6 实验结果与分析(Experimental results and analysis)

为验证本文提出的参数优化算法的有效性，在MATLAB7.0平台下进行算法实现，利用遗传算法求解该问题的最优参数，通过交叉验证比较在不同参数下遗传算法的性能，并与在线性拟合条件下对参数的求解性能进行比较。

6.1 参数优化算法仿真

在对本文提出的参数优化算法的仿真过程中，参数a1、b1、a2、b2、a3、b3为优化的对象，其参数初始种群范围如表2所示。算法的最大迭代次数，设为100。

表2 参数初始种群取值范围

Tab.2 The range of initial population of parameters

参数	取值范围
a1	[0.5,1.5]
b1	[3.5,10.5]
a2	[4,12]
b2	[3,9]
a3	[0.5,1.5]
b3	[1.5,4.5]

根据山西省某医院提供的高血压数据，经过数据清洗、整理得到部分的数据用于本实验。将数据分成10组，采用交叉验证的方法，经过反复的训练、测试。对得到不同参数下的模型，计算其均方误差，从而确定最终的评价模型。参数优化过程中，部分组所得到的平均误差及准确率如表3所示。

表3 4组模型对应的平均误差

Tab.3 The average error corresponding to the four sets of models

名称	第1组	第4组	第7组	第9组
平均误差	5.1138	5.2516	5.5095	4.9125
准确率	85.2%	83.8%	82.4%	86.7%

表3列出了不同的模型对应的平均误差值，通过比较我们发现第9组所确定的模型所得到的误差值最小，准确度最高；同样将数据进行线性拟合，其平均误差为5.7447，准确率为77.9%。因此我们将第9组对应的模型作为我们的最佳评价模型，相应的模型图2所示。

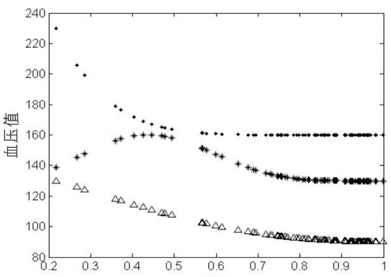

图2 第9组模型图

Fig.2 Model diagram of the ninth group

6.2 结果分析说明

由图2的描述可以看出，总体来讲，患者用药一段时间后能够显著降低血压水平，但对于中度高血压患者来说，可能由于初期药物的不合理性或者患者体质的影响等因素，会导致血压的暂时升高，这部分患者在药物选择过程中还有待进一步研究。对于大部分高血压患者来说，如果坚持服药，能使血压维持在较低的水平，从而为高血压患者坚持用药提供了理论依据。

7 结论(Conclusion)

通过对持续吃药时间和血压值的研究，建立二者之间的评价模型。本文的研究工作从一定程度上能够增强病人的可持续用药程度，从而提高高血压疾病的治疗效果，提升患者及其家人的幸福感。可以考虑将本文的研究方法用于其他疾病的研究。

参考文献(References)

[1] Wilbert S.Aronow.Treating hypertension and prehypertension in older people:When,whom and how[J]. Maturitas,2015,80(1):31-36.

[2] Thusitha Mabotuwana,Jim Warren.ChronoMedIt——A computational quality audit framework for better management of patients with chronic conditions[J]. Journal of Biomedical Informatics,2010,43(1):144-158.

[3] 刘力生,王文,姚崇华.中国高血压防治指南(2010年基层版)[J].中华高血压杂志,2011,18(1):11-18.

[4] Huan M.Nguyen,et al.Clevidipine for the Treatment of Severe Hyper-tension in Adults[J].Clinical Therape-utics,2010,32(1):11-23.

[5] Tak-chung Fu.A review on time series data mining[J]. Engineering Applications of Artificial Intelligence,2011, 24(1):164-181.

[6] Michel Burnier,et al.Measuring,Analyzing,and Managing Drug Adherence in Resistant Hypertension[J].Hypertension, 2013,62:218-225.

[7] M.Akhil jabbar,B.L Deekshatulua,Priti Chandra.Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm[J].Procedia Technology,2013,10:85-94.

作者简介：

曹小凤(1989-)，女，硕士，助教.研究领域：人工智能，数据挖掘.

(上接第46页)

[4] 王竹君.基于人脸检测的移动点名系统研究与实现[D].云南:云南大学,2014.

[5] 王姗.课堂智能点名系统[D].南阳:南阳理工学院,2011.

[6] 王攀,李少波.基于肤色和FBLBP算法的人脸检测[J].计算机与发展,2017,27(1):44-47.

[7] 王志一,杨大利.人脸识别中发型遮挡检测方法研究[J].微型机与应用,2016,35(2):32-34.

[8] 白蕾,郭清菊.HTML5与CSS3的设计模式[J].智能计算机与应用,2015,6(2):104-109.

[9] 张卫明.ASP_NET2_0缓存技术在动态网页设计中的应用[J].软件导刊,2010,9(12):138-139.

[10] 李凌春.浅谈Access中Select语句的应用[J].湖北经济学院学报(人文社会科学版),2012,9(7):133-134.

[11] 梅炳夫.基于ASP的通用网页查询函数的设计[J].现代计算机(专业版),2008:131-132.

作者简介：

荆 洲(1996-)，男，本科生.研究领域：电气工程及其自动化.

权 伟(1982-)，男，博士，讲师.研究领域：控制科学与工程.本文通讯作者.

唐 杰(1997-)，男，本科生.研究领域：电气工程及其自动化.

晏 寒(1995-)，男，本科生.研究领域：电气工程及其自动化.