

基于模糊理论的高血压药物疗效预测模型研究

曹小凤 谢红薇* 安建成 郝晓燕 曹杰

(太原理工大学计算机科学与技术学院 山西 太原 030024)

摘要 针对评估期内药物占有率 MPR (Medication Possession Ratio) 和血压进行分析研究, 建立用于发现二者关系的贝塔分布模型。利用模糊理论和遗传算法通过交叉验证对模型进行优化并与线性分布模型进行对比。实验结果表明, 利用贝塔分布模型确定的 MPR 与血压值的关系能很好地对患者的用药疗效进行预测。对于大多数高血压患者, 只有接受长期的药物治疗, 才能使血压得到有效控制。

关键词 高血压 贝塔分布模型 遗传算法 交叉验证 模糊理论

中图分类号 TP18 文献标识码 A DOI:10.3969/j.issn.1000-386x.2016.02.013

ON PREDICTION MODEL OF HYPERTENSION DRUGS EFFICACY BASED ON FUZZY THEORY

Cao Xiaofeng Xie Hongwei* An Jiancheng Hao Xiaoyan Cao Jie

(School of Computer College of Science and Technology, Taiyuan University of Technology, Taiyuan 030024, Shanxi, China)

Abstract Aiming at the medication possession ratio (MPR) and blood pressure within evaluation period we carried out analysis and study, and built a beta distribution model for finding the relationship between them. Then we employed fuzzy theory, genetic algorithms (GA) and cross validation to optimise the model, and compared it with linear distribution model. Experimental results showed that the MPR and blood pressure value determined by using the beta distribution model can well forecast the medication efficacy on patients. For most hypertensives, only by long-term treatment with medicine can the blood pressure be effectively controlled.

Keywords Hypertension Beta distribution model Genetic algorithm Cross-validation Fuzzy theory

0 引言

2013 年世界卫生日主题“控制你的血压, 减少心脏病突发和卒中风险”, 可见高血压已成为全球瞩目的公共健康问题。高血压是一种常见的威胁人类健康的慢性疾病, 位居引起死亡的十大危险因素之首。因此, 有效控制血压对于提高人类的健康水平有重大现实意义。

《中国高血压防治指南》^[1]指出, 高血压是可以控制的疾病, 有效地控制血压可以减少患者心脑血管及其他并发症的发生, 进而提高患者的生存质量。Laura 等人^[2]通过研究发现, 在老年人中使用阿替洛尔这类降压药会增加中风的风险; 赵艳平^[3]通过对治疗原发性高血压的不同药物疗效进行分析, 发现洛沙坦和非洛地平的联合用药降低血压变异性更为明显, 患者预后更佳; 林彩美^[4]利用统计分析的方法, 研究抗高血压药物对治疗老年人高血压的效果, 实验结果表明, 卡托普利联合硝苯地平药物治疗高血压效果显著、不良反应少、易耐受, 是治疗老年人高血压的安全有效用药方法。进行药物治疗是控制高血压的有效方式, 通过对患者坚持服药程度和血压值进行分析。根据 Thusitha 等人提出的计算框架^[5]从电子处方中计算出患者的 MPR, 进而建立二者之间的关系, 利用遗传算法对所建立的模型进行评估优化, 最终得出患者服药时间与血压的模型图。此模型的建立, 不仅可以通过患者坚持服药程度对其血压趋势进行

预测, 还可以给患者提供坚持用药建议, 对于高血压慢性病的控制与治疗有一定的指导意义。

1 高血压定义和状态分类

在没有使用降压药物的情况下, 非同日测量 3 次血压, 收缩压 ≥ 140 mmHg 和 (或) 舒张压 ≥ 90 mmHg, 被诊断为高血压; 患者有既往高血压史, 且目前正在服用降压药物, 虽然血压低于 140/90 mmHg, 也被诊断为高血压^[1]。通过对文献^[6]进行研究及总结, 现将使用的收缩压 SBP (Systolic Blood Pressure) 及其状态定义如表 1 所示。

表 1 收缩压值与状态对应关系 单位: mm Hg

状态	收缩压
可接受血压 (Acceptable Blood Pressure, bpa)	90 ~ 129
中度高血压 (Middle Blood Pressure, bpm)	130 ~ 159
高血压 (High Blood Pressure, bph)	≥ 160

收稿日期: 2014-08-25。山西省国际合作计划项目 (2014081018-2); 山西省科技基础条件平台建设基金项目 (2013091003-0103); 山西省基础研究基金项目 (2012011011-2)。曹小凤, 硕士生, 主研领域: 人工智能, 医学信息学。谢红薇, 教授。安建成, 副教授。郝晓燕, 副教授。曹杰, 硕士生。

2 模型选择

根据山西省某医院提供的高血压数据,通过研究发现,患者的收缩压值和持续吃药时间近似符合贝塔分布曲线。由于正态分布的曲线分布特点是以均数为中心,左右两边呈现对称性并由均数处开始,分别向左右两侧呈逐渐均匀下降趋势。而相比于正态分布,贝塔分布具有多种不同的分布形状,其中包括对称的和非对称的分布,根据参数的不同呈现出完全不同的形状,体现出良好的适应性和普适性。在此,选择贝塔分布对数据进行研究分析。

2.1 贝塔分布特点

贝塔分布的密度函数定义如下:

$$f(x; a, b) = x^{a-1} (1-x)^{b-1} / B(a, b) \quad (1)$$

其中, a, b 表示待确定的参数, $B(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du$ 。由式(1)可知,贝塔分布函数的形状由参数 a 和 b 确定,属于广义上的正态分布,分布的形状是通过调节 a 和 b 的参数值而得到的。因此,主要通过对两个参数 a 和 b 的优化来确定最终的曲线。

2.2 MPR 值的计算

此处所研究的是患者持续吃药时间与收缩压的关系,由式(1)可知,该分布函数的自变量范围为(0,1),因此,文中将持续吃药时间转换为 MPR 进行研究。MPR 是计算有效药物供给时间在一个评估期^[5] EP 内所占有的比例。EP 指从病人的电子处方记录中选取感兴趣的一段时期(此处选择的是一年),根据 Thusitha 等人在文献^[5]中提出的一种方法^[4]来计算药物占有率,其计算如下:

$$MPR = \frac{EP \text{ 的天数} - \text{总的间隔天数}}{EP \text{ 的天数}} \times 100\% \quad (2)$$

3 基于模糊规则的模型

Takagi 与 Sugeno(1985)和 Sugeno 与 Kang(1988)^[7]给出了一个产生式规则系统的模糊推理工具。他们提出了一个多维模糊推理,该模糊模型是基于规则的^[8,9],其输出不是一个语言变量,而是输入变量的函数。规则库由 n 条规则组成,规则的形式如下:

$$R^i: \text{If } x_1 \text{ is } F_1^i \text{ and } x_2 \text{ is } F_2^i \text{ and } \dots \text{ and } x_n \text{ is } F_n^i \text{ Then } y \text{ is } G^i \quad (3)$$

其中, $x_i (i = 1, 2, \dots, n)$ 和 y 分别是过程的状态变量和控制变量。 F_j^i 和 G^i 分别是模糊隶属度函数 $u_{F_j^i}(x_i)$ 和 $u_{G^i}(y)$ 的语言术语。每一个 R^i 可以被看做一个模糊蕴含式,模糊规则集形式如下:

$$F_1^i \times F_2^i \times \dots \times F_n^i \rightarrow G^i \quad (4)$$

$$u_{R^i} = u_{(F_1^i, F_2^i, \dots, F_n^i \rightarrow G^i)}(x_1, x_2, \dots, x_n, y) \\ = [u_{F_1^i}(x_1) \text{ and } u_{F_2^i}(x_2) \text{ and } \dots \text{ and } u_{F_n^i}(x_n)] \text{ and } u_{G^i}(y) \quad (5)$$

这里使用的“and”是在模糊系统中经常使用的运算。通常使用由 Sugeno 提供的模糊规则^[10]得到复杂过程的输出,其形式描述如下:

$$R^i: \text{If } x_1 \text{ is } A_1^i \text{ and } x_2 \text{ is } A_2^i \dots \text{ and } x_n \text{ is } A_n^i \text{ Then } y^i = p_0^i + p_1^i x_1 + \dots + p_n^i x_n \quad i = 1, 2, \dots, m \quad (6)$$

这样,最终的输出 y 可以按下式计算:

$$y = \frac{\sum_{i=0}^n \min(u_{A_1^i}(x_1), \dots, u_{A_m^i}(x_m)) \cdot (p_0^i + p_1^i x_1 + \dots + p_m^i x_m)}{\sum_{i=0}^n \min(u_{A_1^i}(x_1), \dots, u_{A_m^i}(x_m))} \quad (7)$$

根据式(7),最终的输出是由参数决定的复杂的非线性函数,参数的确定不能由传统的方法得到。为了使用遗传算法^[11]确定式(7)的最优参数,必须建立适应度函数,表示如下:

$$J = 1 / \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (8)$$

其中, y_i 和 \bar{y}_i 分别是实际输出和期望输出, n 表示输出个数, $M = \sum_{i=1}^n (y_i - \bar{y}_i)^2$ 为目标函数。

4 实验过程及运行结果

4.1 实验过程

由于研究的是 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,需要对数据进行尺度变换,将模型得出的数据值模糊化为 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式子如下:

$$bph = u1/Maxu1 \times 70 + 160 \quad (9)$$

$$bpm = u2/Maxu2 \times 30 + 130 \quad (10)$$

$$bpa = u3/Maxu3 \times 40 + 90 \quad (11)$$

其中, u_i 是 n 维向量,其值是由式(2)计算出的 MPR 值根据初始的贝塔分布函数所确定的(初始贝塔分布的参数值说明如表 2 所示), $Maxu_i$ 是 u_i 中的最大值, $u_i/Maxu_i$ 表示相应规则(此处的规则是表 1 中所描述的 3 条规则)的隶属度。通过式(10) - 式(12)将相应的精确值分别模糊化为对应的规则区间。

根据山西某医院提供的 30 461 条患者高血压数据,经过数据清洗得到本实验所使用的 200 条相关数据,将数据分成 10 组,每组 20 条数据,选取其中的 8 组作为训练集,另外 2 组作为测试集。根据训练集 180 条数据得到参数初值如表 2 所示。

表 2 参数初值及种群初始范围

	a1	b1	a2	b2	a3	b3
初值	1	7	8	6	1	3
范围	[0.8, 1.2]	[5.6, 8.4]	[6.4, 9.6]	[4.8, 7.2]	[0.8, 1.2]	[2.4, 3.6]

实验在 MATLAB 7.0 平台下进行,根据表 2 的初值、范围和适应度函数式(8),交叉概率 $P_c = 0.8$,变异概率 $P_m = 0.05$,利用遗传算法对训练集进行多次训练得到参数的优化结果。表 3 给出了不同的训练集得到的优化参数结果(由于篇幅有限,只列举部分数据进行说明)。

表 3 仿真算法得到的优化参数值

	第 1 组	第 2 组	第 3 组	第 4 组
a1	1.14177	0.80573	1.01627	1.14177
b1	8.34785	8.09964	6.28199	7.94828
a2	7.27945	7.51947	6.95418	5.45959
b2	5.05606	4.92787	5.30812	6.74141
a3	3.40641	1.0013	1.01103	1.19255
b3	1.09092	3.58007	2.51293	3.27275

4.2 参数对比

研究的最终目的是确定 SBP 与 MPR 的关系,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 值所对应的隶属度对数据进行去模糊化处理,得到最终的输出结果。对最终结果好坏的评价,通常用平均误差来衡量,计算如下所示:

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (12)$$

在对大量数据进行研究和实验的基础上,根据式(12)分别计算不同组的平均误差如表 4 所示。

表 4 4 组模型对应的平均误差

	第 1 组	第 2 组	第 3 组	第 4 组
平均误差	5.11381	5.27472	5.51115	4.91254

通过比较发现第 4 组所确定的模型所得到的平均误差值最小,准确度最高;进行模糊化分段线性分布计算得到最终的 RMSE 为 5.74473,相应的参数 a4、b4、a5、b5、a6、b6 值如表 5 所示。因此将第 4 组对应的模型作为我们的最佳预测模型,用于对持续吃药时间和血压值的预测。

表 5 分段线性拟合的参数值

	a4	b4	a5	b5	a6	b6
参数值	-42.217	70.028	-59.496	142.114	-82.990	169.733

通过大量实验得到表 3 - 表 5 中的参数值,图 1 - 图 3 分别是对得到的贝塔分布函数,分段线性分布函数及贝塔分布函数和分段线性分布函数的对比进行描述。

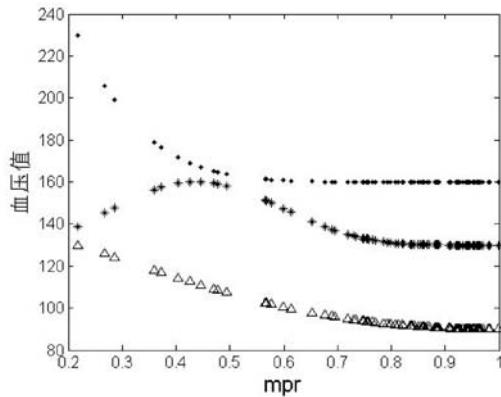


图 1 贝塔分布函数结果

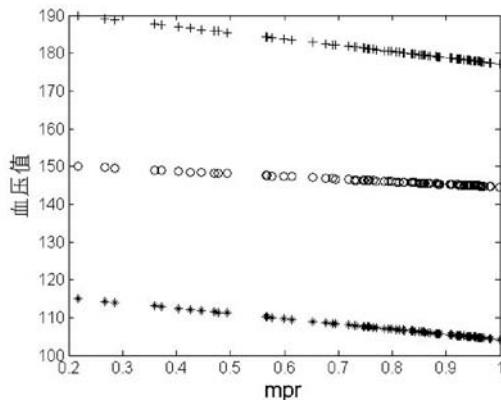


图 2 分段线性分布函数结果

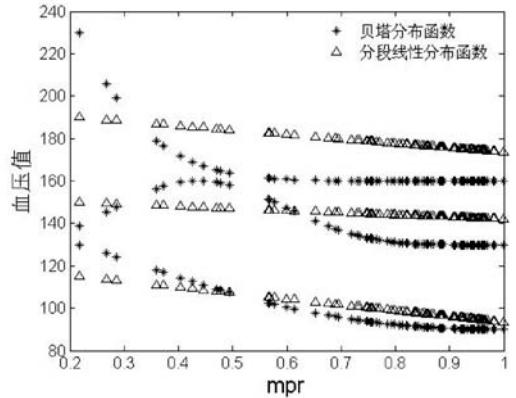


图 3 分段线性分布函数和贝塔分布函数比较结果

经过实验对比,确定图 1 为最终的持续吃药时间与血压关系模型图。由图 1 的描述可以得到:当患者血压水平处于 90 ~ 130 mmHg 时,患者持续服药 8 个月,可以使血压基本维持在较低水平;当患者血压水平处于 130 ~ 160 mmHg 时,由于患者体质不同,服用的降压药的差别可能在短期内会使患者血压有所上升,但服用半年后,这部分高血压患者的血压也可以维持在相对较低的水平;对于血压水平处于 160 ~ 230 mmHg 的高血压患者,在初期服用降压药的效果比较明显,但是需要进行长期的药物治疗,才能使血压维持在较低的水平。

5 结 语

通过对持续吃药时间和血压值的分析,建立用于预测二者关系的贝塔分布模型。与线性分布模型相比,该模型能够很好地进行血压趋势分析,增强患者可持续服药程度,更好地对血压进行控制。在此研究的基础上建立药物、时间与血压的模型,进而发现三者之间的关系对于指导患者用药,对于提高患者生存质量有重大现实意义,建立相关的模型是下一步所要进行的工作。

参 考 文 献

- [1] 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.
- [2] Laura M Kuyper, MD Nadia A Khan, MD MSc. Atenolol vs Nonatenolol β -Blockers for the Treatment of Hypertension; A Meta-analysis[J]. Canadian Journal of Cardiology, 2014, 30(5): S47-S53.
- [3] 赵艳平. 不同药物治疗原发性高血压的疗效分析与预后评价[J]. 医药论坛杂志, 2012, 33(7): 96-97.
- [4] 林彩美. 卡托普利联合硝苯地平药物治疗高血压的临床疗效分析[J]. 中外医疗, 2009, 28(8): 71.
- [5] Thusitha Mabotuwana, Jim Warren, John Kennelly. A computational framework to identify patients with poor adherence to blood pressure lowering medication[J]. International Journal of Medical Informatics, 2009, 78(11): 745-756.
- [6] 陈建华. 国内外的高血压诊断标准[J]. 中华实用中西医杂志, 2010, 23(4): 33-34.
- [7] Jose Luis Aznarte, Jesus Alcala-Fdez, Antonio Arauzo-Azofra, et al. Financial time series forecasting with a bio-inspired fuzzy model[J]. Expert Systems with Applications, 2012, 39(16): 12302-12309.
- [8] 张德丰. MATLAB 模糊系统设计[M]. 北京:国防工业出版社, 2009.

评估指标进行了分析,总结了这些指标的特点和适用范围以及优缺点,提出了针对不同应用场景,合适的能耗评估指标,方便研究者根据需要选择合适的指标来评估 WSNs 的能量有效性,并有助于协议的改进和 WSNs 的设计。我们下一步的工作是在此基础上,进一步改进评估指标,得到更加便于实现的轻量级能耗评估指标。

参 考 文 献

- [1] Akylidiz I F, Su W, Sankarasubramaniam Y, et al. A survey on sensor networks [J]. *Communications Magazine, IEEE*, 2002, 40 (8) : 102 - 114.
- [2] 司海飞, 杨忠, 王珺. 无线传感器网络研究现状与应用 [J]. *机电工程*, 2011, 28 (1) : 16 - 20.
- [3] Larios D F, Barbancho J, Rodriguez G, et al. Energy efficient wireless sensor network communications based on computational intelligent data fusion for environmental monitoring [J]. *Communications, IET*, 2012, 6 (14) : 2189 - 2197.
- [4] Cheng Z, Perillo M, Heinzelman W B. General network lifetime and cost models for evaluating sensor network deployment strategies [J]. *Mobile Computing, IEEE Transactions on*, 2008, 7 (4) : 484 - 497.
- [5] Gao D Y, Zhang L J, Wang H C. Energy saving with node sleep and power control mechanisms for wireless sensor networks [J]. *The Journal of China Universities of Posts and Telecommunications*, 2011, 18 (1) : 49 - 59.
- [6] Pantazis N A, Nikolidakis S A, Vergados D D. Energy-efficient routing protocols in wireless sensor networks: A survey [J]. *Communications Surveys & Tutorials, IEEE*, 2013, 15 (2) : 551 - 591.
- [7] Raghunathan V, Schurgers C, Park S, et al. Energy-aware wireless microsensor networks [J]. *Signal Processing Magazine, IEEE*, 2002, 19 (2) : 40 - 50.
- [8] Estrin D. Wireless sensor networks tutorial part IV: sensor network protocols [C] // *Proceedings of the 8th ACM Mobicom, Atlanta, Georgia, USA, Westin Peachtree Plaza*, 2002: 23 - 28.
- [9] Stojmenovic I, Lin X. Power-aware localized routing in wireless networks [J]. *Parallel and Distributed Systems, IEEE Transactions on*, 2001, 12 (11) : 1122 - 1133.
- [10] Hui W, Baoshu X. An Energy Efficient Routing Metric Algorithm Base on PCE for WSN [C] // *Intelligent Computing and Cognitive Informatics (ICICCI)*, 2010 International Conference on. *IEEE*, 2010: 90 - 93.
- [11] Cao Q, He T, Fang L, et al. Efficiency Centric Communication Model for Wireless Sensor Networks [C] // *Jordi Domingo-Pascual. 25th IEEE International Conference on Computer Communications Proceedings, Barcelona, Spain*, 2006: 1 - 12.
- [12] Zhao J, Govindan R. Understanding packet delivery performance in dense wireless sensor networks [C] // *Proceedings of the 1st international conference on Embedded networked sensor systems. ACM*, 2003: 1 - 13.
- [13] Akhtar A M, Nakhai M R, Aghvami A H. Power aware cooperative routing in wireless mesh networks [J]. *Communications Letters, IEEE*, 2012, 16 (5) : 670 - 673.
- [14] Rodoplu V, Meng T H. Minimum energy mobile wireless networks [J]. *Selected Areas in Communications, IEEE Journal on*, 1999, 17 (8) : 1333 - 1344.
- [15] Benini L, DeMicheli G. Dynamic power management: design techniques and CAD tools [M]. *Springer*, 1998.
- [16] Seo J, Kim T, Lee J. Optimal intratask dynamic voltage-scaling technique and its practical extensions [J]. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 2006, 25 (1) : 47 - 57.
- [17] Kim E J, Shon T, Park J J H, et al. Latency bounded and energy efficient MAC for wireless sensor networks [J]. *Communications, IET*, 2012, 6 (14) : 2120 - 2127.
- [18] Chiasserini C, Garetto M. An analytical model for wireless sensor networks with sleeping nodes [J]. *Mobile Computing, IEEE Transactions on*, 2006, 5 (12) : 1706 - 1718.
- [19] Schurgers C, Aberthorne O, Srivastava M. Modulation scaling for energy aware communication systems [C] // *Proceedings of the 2001 international symposium on Low power electronics and design. ACM*, 2001: 96 - 99.
- [20] 李焯, 张旗, 黄晓霞. 无线传感器网络综述 [J]. *先进技术研究通报*, 2009, 3 (11) : 2 - 6.
- [21] Li B, Wang W, Yin Q, et al. A new cooperative transmission metric in wireless sensor networks to minimize energy consumption per unit transmit distance [J]. *Communications Letters, IEEE*, 2012, 16 (5) : 626 - 629.
- [22] Shrotriya A, Nitawwre D. Optimizing Energy Consumption of Wireless Sensor Network Based on Hop Distance [J]. *IJCSC*, 2012, 3 (1) : 33 - 37.
- [23] Fonseca B J B, Gubner J A. Operation Cost as a Performance Metric of Wireless Sensor Networks [C] // *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on. IEEE*, 2007: 680 - 684.
- [24] Lindsey S, Raghavendra C, Sivalingam K M. Data gathering algorithms in sensor networks using energy metrics [J]. *Parallel and Distributed Systems, IEEE Transactions on*, 2002, 13 (9) : 924 - 935.
- [25] Dietrich I, Dressler F. On the lifetime of wireless sensor networks [J]. *ACM Transactions on Sensor Networks (TOSN)*, 2009, 5 (1) : 1 - 38.
- [26] Senouci M R, Mellouk A, Senouci H, et al. Performance evaluation of network lifetime spatial-temporal distribution for WSN routing protocols [J]. *Journal of Network and Computer Applications*, 2012, 35 (4) : 1317 - 1328.
- [27] Ozgovde A, Ersoy C. WCOT: A utility based lifetime metric for wireless sensor networks [J]. *Computer Communications*, 2009, 32 (2) : 409 - 418.
- [28] Ammer J, Rabaey J. The energy-per-useful-bit metric for evaluating and optimizing sensor network physical layers [C] // *Sensor and Ad Hoc Communications and Networks, 2006. SECON '06. 2006 3rd Annual IEEE Communications Society on. IEEE*, 2006, 2: 695 - 700.
- [29] Holland M, Wang T, Tavli B, et al. Optimizing physical-layer parameters for wireless sensor networks [J]. *ACM Transactions on Sensor Networks (TOSN)*, 2011, 7 (4) : 28.
- [30] 夏先进, 李士宁, 李志刚, 等. 基于做功系统的 WSN 能量有效性评价模型 [J]. *科学技术与工程*, 2013 (13) : 3647 - 3652.

(上接第 57 页)

- [9] Chen Chunhao, Hong Tzungpei, Tseng Vincent S. Fuzzy data mining for time-series data [J]. *Applied Soft Computing*, 2012, 12 (1) : 536 - 542.
- [10] Chamani M R, Pourshahabi S, Sheikholeslam F. Fuzzy genetic algorithm approach for optimization of surge tanks [J]. *Scientia Iranica A*, 2013, 20 (2) : 278 - 285.
- [11] 雷英杰. *MATLAB 遗传算法工具箱及应用* [M]. 西安: 西安电子科技大学出版社, 2005.