

# 用于鲁棒性建模的概率权重极限学习机

周 闯 范 彬 朱 蕾 陆新江

(中南大学机电工程学院 长沙 410083)

**摘 要** 极限学习机(ELM)在机器学习领域获得了很多的关注,并在应用方面取得了极大的成功。然而,极限学习机对训练数据中的异常值点和非高斯噪声非常敏感,从而大大阻碍了ELM的应用。概率权重ELM方法主要对含有异常值和非高斯噪声数据集进行建模,首先建立概率局部ELM模型,并在此基础上利用Parzen窗方法建立局部模型的概率分布,然后将概率分布作为权重来融合所有的局部模型以建立全局鲁棒性模型。该方法成功地应用了数学例子和UCI实例,并与传统ELM、正则化ELM和鲁棒ELM进行了比较分析,结果表明概率权重ELM表现出了较好的建模性能。

**关键词** 极限学习机,鲁棒性,噪声,概率分布,异常值

**中图分类号** TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.08.041

## Probabilistic Weighted Extreme Learning Machine for Robust Modeling

ZHOU Chuang FAN Bin ZHU Lei LU Xin-jiang

(College of Mechanical and Electrical Engineering, Central South University, Changsha 410083, China)

**Abstract** Extreme learning machine (ELM) has attracted a lot of attention in the machine learning field and gained great success in application. However, it is sensitive to outliers and non-Gaussian noise in the training dataset, which greatly hinder the application of ELM. Probabilistic weighted ELM was proposed to model the dataset with the outliers and non-Gaussian noise. First, a distributed local ELM modeling is developed, upon which the probability distribution function (PDF) of multiple local models is estimated by the Parzen window method. Then, the distribution function is further used as weights to integrate all local models to construct a global robust ELM model. The successful application of this robust probabilistic weighted ELM method to both artificial case and real life case as well as its comparison to traditional ELM, regularization ELM and robust ELM demonstrate its superiority in the modeling.

**Keywords** Extreme learning machine, Robustness, Noise, Probability distribution, Outlier

## 1 引言

毫无疑问,大数据是当前和未来的研究前沿之一,从大数据中可以提取出很多有用的信息<sup>[1-3]</sup>。但是实际获得的数据不一定是准确的,因为其中包含了大量噪声,比如测量误差、模型误差、操作误差等<sup>[4-6]</sup>,因此如何在噪声数据条件下进行鲁棒建模是研究大数据的基础和条件<sup>[7]</sup>。

极限学习机模型(ELM)<sup>[8]</sup>自从被Huang等人提出以来受到了越来越多人的关注与欢迎。这主要是由于极限学习机模型的第一层输入权值和阈值是随机产生的,而第二层输出权值通过公式定量求出,因此其计算复杂性低,具有较短的训练时间<sup>[9]</sup>。除此之外,相比传统的神经网络<sup>[10]</sup>和支持向量机<sup>[11-12]</sup>,ELM还拥有不易陷入局部最优、训练时间短且结构模型简单等优点,这些优点使得ELM在很多应用领域取得了极大的成功<sup>[13]</sup>。但是ELM对数据中的异常值点和非高斯

噪声点异常敏感<sup>[14]</sup>,这使得ELM无法对噪声数据进行鲁棒建模。

减小噪声影响的常见方法主要包括:1)在进行建模之前,通过统计或者经验的方法过滤掉“噪声点”,然后用传统的ELM对剩余的数据进行建模<sup>[15-16]</sup>;2)通过迭代的方法对不同的数据点施以不同的权重,即对噪声点施以较小的权重来减小噪声的影响<sup>[17]</sup>;3)通过改进惩罚函数增强模型的鲁棒性<sup>[18]</sup>。但是这些方法都脱离模型来进行噪声去除,耗费大量时间,在实际应用中的效果并不理想。

本文提出概率权重极限学习机,该方法通过计算子模型的概率误差分布来提高模型的鲁棒性。首先,通过重采样方法对数据集进行采样分组,并建立每个子数据集的模型;然后,用子模型对所有数据进行预测,并得到平均误差分布;最后,结合子模型的误差分布建立全局鲁棒性模型。将该方法与传统的极限学习机、正则化极限学习机和鲁棒极限学习机

到稿日期:2016-11-05 返修日期:2017-02-24 本文受中南大学中央高校基本科研业务费专项资金(2016zzts296),中南大学“创新驱动计划”项目(2015CX002),国家自然科学基金资助项目(51675539)资助。

周 闯(1991-),男,硕士生,主要研究方向为机器学习和数据挖掘,E-mail:zcjiyi@126.com;范 彬(1980-),男,博士,主要研究方向为数据建模和信号处理,E-mail:597161096@qq.com;朱 蕾(1991-),女,硕士生,主要研究方向为数据挖掘和经贸管理,E-mail:1016009967@qq.com;陆新江(1979-),男,博士,教授,博士生导师,主要研究方向为机器学习与智能控制,E-mail:211063@csu.edu.cn。

进行比较,结果表明该方法的性能优于对比方法。

## 2 极限学习机简介

掺杂有噪声的未知非线性系统可描述为:

$$y=f(x,\epsilon) \tag{1}$$

其中,  $x$  和  $y$  分别是系统的输入和输出,  $f$  是未知的非线性系统,  $\epsilon$  是多种多样的随机噪声。

假设 ELM 有  $L$  个隐层,则极限学习机模型可表述为:

$$y^* = h(x, a, b)\beta \tag{2}$$

其中,  $h=[h_1(x), h_2(x), \dots, h_L(x)]$ ;  $h(x)$  是激活函数;  $a$  是输入权重矩阵,用来连接输入层和隐层;  $b$  是截距。  $a$  和  $b$  均是随机选取的。

对于训练集  $\{x_i, y_i\}_{i=1}^n, x_i \in R^d, y_i \in R^m$ , ELM 的最终优化问题如式(3)所示:

$$J(\beta) = \min_{\beta} \| H\beta - T \|^2 \tag{3}$$

$$\text{其中, } H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_n) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & h_2(x_1) & \dots & h_L(x_1) & 1 \\ h_1(x_2) & h_2(x_2) & \dots & h_L(x_2) & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_1(x_n) & h_2(x_n) & \dots & h_L(x_n) & 1 \end{bmatrix},$$

$$T = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_n^T \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{bmatrix}.$$

根据式(3)求得:

$$\beta = H^\dagger \cdot T \tag{4}$$

其中,  $H^\dagger$  是  $H$  的 Moore-Penrose 广义逆矩阵<sup>[8]</sup>。

如果对 ELM 加上正则化项,容易推出<sup>[19]</sup>:

$$\beta = \left( \frac{I}{\gamma} + H^T \cdot H \right)^{-1} \cdot H^T \cdot Y \tag{5}$$

通过上式可知,传统的 ELM 对于异常值和非零均值噪声是非常敏感的。

## 3 概率权重极限学习机

一般地,异常值点或者噪声点应远离正常的的数据,或者相对正常的的数据较少<sup>[20]</sup>,因此噪声应该拥有较小的频率。采用有放回重采样的方法获得不同的子样本,子样本中噪声点的数量也应服从上述规律,即若噪声点数量多,则对子样本的影响大,但在所有子样本集合中的占比小;若噪声点数量少,则对子样本的影响小,在所有子样本集合中的占比大。基于以上特性,可以通过子样本集中的噪声分布来评估子样本的情况,从而建立鲁棒的概率权重 ELM,具体步骤如图 1 所示。

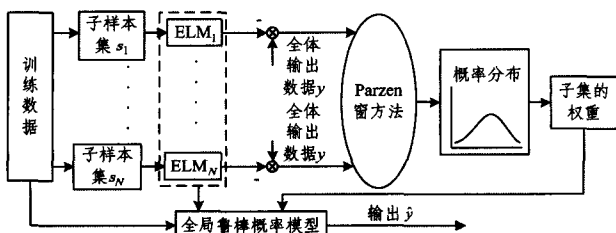


图 1 概率权重极限学习机

### 3.1 建立局部 ELM 模型

根据重采样定理,将训练数据随机有放回地分成  $N$  组,每一组则是一个子样本集,然后对子样本集建立 ELM 模型,第  $j$  个子模型可表示为:

$$\hat{y}_j(x_j) = h(x_j, a_j, b_j)\beta_j \tag{6}$$

其中,  $x_j$  和  $y_j$  是局部模型的输入和输出,  $a_j$  是局部输入权重矩阵,  $b_j$  是局部截距。

### 3.2 建立局部 ELM 的概率分布

采用总体数据评估的方法获得局部 ELM 的概率分布,主要步骤如下。

1) 用局部模型求出整体样本的均方根误差,其公式为:

$$E_j = \frac{1}{n} \sqrt{\sum_{i=1}^n \| y(x_i) - \hat{y}_j(x_i) \|^2} \tag{7}$$

其中,  $n$  和  $y$  分别是总体样本个数和所有训练数据的实际输出,  $\hat{y}_j$  是局部模型输出。

2) 用 Parzen 窗建立 RMSE 的概率分布,其中,窗函数定义为<sup>[21]</sup>:

$$\phi\left(\frac{E-E_j}{h}\right) = \begin{cases} 1, & \frac{|E-E_j|}{h} \leq \frac{1}{2} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

窗函数的中心和宽度分别为  $E$  和  $h$ ,并且窗的区域为  $-\frac{h}{2} \leq E-E_j \leq \frac{h}{2}$ ,所有落在区域内的 RMSE 用式(9)统计:

$$k = \sum_{j=1}^N \phi\left(\frac{E-E_j}{h}\right) \tag{9}$$

每一个 RMSE 的概率值可表示为:

$$P(E) = \frac{k}{N} = \frac{1}{N} \sum_{j=1}^N \phi\left(\frac{E-E_j}{h}\right) \tag{10}$$

因为 RMSE 是对局部 ELM 鲁棒性的一个度量, RMSE 的概率分布可看作是局部 ELM 鲁棒性的概率分布,而噪声点是局部 ELM 概率分布的直接影响因素,因此 RMSE 的概率分布反映了噪声点的影响程度。

### 3.3 目标函数的构建

正如上文所述,局部 ELM 模型的概率分布反映了噪声点的影响程度,因此为了得到鲁棒的 ELM,重新构造了一个充分考虑样本概率分布的新目标函数。具体如下:

$$J(\beta) = \min_{\beta} \frac{\lambda}{2} \| \beta \|^2 + \frac{1}{2} \sum_{j=1}^N \mu_j \| \beta - \beta_j \|^2 + \frac{\gamma}{2} \sum_{j=1}^N \mu_j \| e_j \|^2 \tag{11}$$

s. t.  $H_j \beta - y_j = e_j, j=1, 2, \dots, N$

其中,  $e_j$  是第  $j$  个子模型的模型误差向量;  $\gamma$  和  $\lambda$  是正则化因子;  $\beta_j$  是局部 ELM 模型的权重参数;  $\mu_j$  代表局部模型权重,等于  $P(E)$ 。这个目标函数具有以下特征:

1) 第一项  $\frac{\lambda}{2} \| \beta \|^2$  是结构风险项,它保证模型的全局泛化性并防止模型的过拟合;

2) 第二项  $\frac{1}{2} \sum_{j=1}^N \mu_j \| \beta - \beta_j \|^2$  是全局正则化项,它建立了所有局部 ELM 模型之间的相互关系,保证全局模型能够兼顾局部模型的特性;

3)第三项  $\frac{\gamma}{2} \sum_{j=1}^N \mu_j \|e_j\|^2$  是经验风险最小项,它最小化所有的训练误差,以提高模型的准确性。

4)权重  $\mu_j = P(E_j)$  用来减小噪声的影响,提高模型的鲁棒性。因为受噪声影响大的局部 ELM 会被赋予小的权重,受噪声影响小的局部 ELM 会被赋予大的权重,即通过这个权重可以尽可能地减小噪声点的影响。

### 3.4 求解过程

为了求解式(11)的优化问题,可将式(11)写成:

$$J(\beta) = \min_{\beta} \frac{\lambda}{2} \|\beta\|^2 + \frac{1}{2} \sum_{j=1}^N \mu_j \|\beta - \beta_j\|^2 + \frac{\gamma}{2} \sum_{j=1}^N \mu_j \|H_j \beta - y_j\|^2 \quad (12)$$

对式(12)进行求导可得:

$$\frac{\partial J}{\partial \beta} = 0 \Rightarrow \lambda \beta + \sum_{j=1}^N \mu_j (\beta - \beta_j) + \gamma \sum_{j=1}^N \mu_j H_j^T (H_j \beta - y_j) = 0 \quad (13)$$

$\beta_j$  已经通过式(5)求解。求解式(13)得:

$$\beta = (\lambda + 1 + \gamma \sum_{j=1}^N \mu_j H_j^T H_j)^{-1} (\sum_{j=1}^N \mu_j \beta_j + \gamma \sum_{j=1}^N \mu_j H_j^T y_j) \quad (14)$$

上述即为鲁棒概率权重 ELM 的建立及其求解过程,其步骤可归纳如下:

- 1)收集训练数据并将训练数据分成若干子样本集;
- 2)在每个子样本集中,建立局部 ELM;
- 3)计算每个局部 ELM 模型的 RMSE;
- 4)采用 Parzen 窗的方法建立 RMSEs 的概率分布,并计算每个子样本集的权重;
- 5)通过概率的方法综合局部 ELM 模型,建立全局 ELM 模型。

## 4 仿真分析

为了验证该方法的有效性,将该方法应用于一个数学例子和 UCI<sup>[22]</sup> 的实际例子,并且与传统的 ELM、正则化 ELM 以及鲁棒 ELM<sup>[23]</sup> 进行比较。

### 4.1 数学例子(非高斯噪声)

数据通过式(15)产生。

$$y = 150x^4(x^2 - 1)^2 e^{-x} + 5 + \epsilon \quad (15)$$

其中,  $x$  和  $y$  分别是输入和输出,  $\epsilon$  是服从瑞利分布(参数为  $\sigma$ )的噪声。随机产生 400 个输入  $x$ , 其中  $x \in [0, 1]$ ,  $\sigma = 0.2$ , 并计算相对应的输出  $y$ 。随机取 200 个数据作为训练数据,另外 200 个数据作为测试数据。将训练数据重采样组成 30 个子样本集,每个子样本集包含 50 个数据,用这些数据来建立局部 ELM。图 2 示出了局部 ELM 的 RMSE 的分布情况,将这些分布作为权重来建立全局鲁棒概率权重 ELM 模型,模型预测情况如图 3 所示。通过图 3 可以看到,提出的方法能够有效地建立非高斯噪声数据的模型。

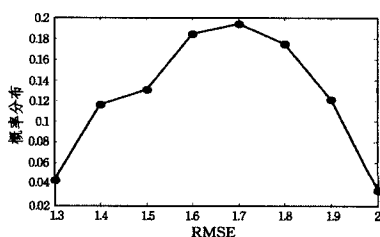


图 2 局部 ELM 的 RMSE 的分布

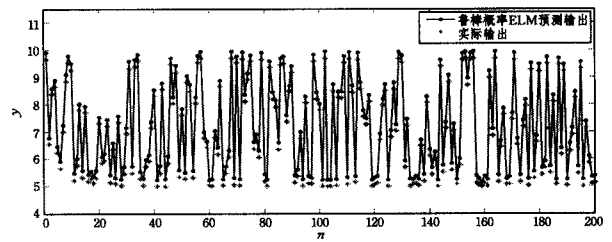


图 3 鲁棒概率 ELM 预测输出情况

除此之外,还将提出的方法与传统的 ELM、正则化 ELM 和鲁棒 ELM 进行对比,对比结果如图 4 所示,平均 RMSEs 如表 1 所列。通过图 4 和表 1 可以看出提出的新方法的性能优于其他方法。

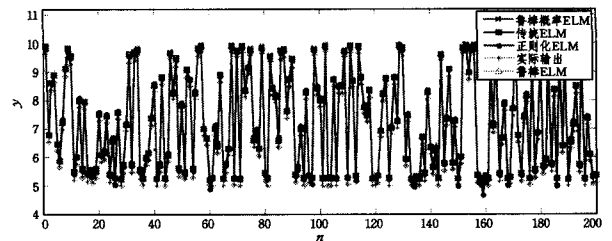


图 4 多种方法的预测情况比较

表 1 多种方法的 RMSE 比较

方法	概率权重 ELM	传统 ELM	正则化 ELM	鲁棒 ELM
RMSE	0.1527	0.1761	0.1615	0.1607

### 4.2 UCI 实际例子

这个例子搜集于 UCI 机器学习数据库,数据取自于某联合发电厂 2006—2011 年间的发电情况。每个数据包括 4 个输入和 1 个输出,具体参数特征如表 2 所列。

表 2 数据的输入输出特征

输入特征	输出特征
平均每小时环境温度/T	每小时的电能净输出/EP
环境压力/AP	
湿度/RH	
真空度/V	

从数据集中随机抽取 400 个数据,其中 200 个作为训练数据,另外 200 个作为测试数据。在训练过程中,训练数据被重采样分为 30 个子样本集,每个子样本集包含 50 个数据。图 5 示出了鲁棒概率 ELM 模型的预测情况,通过图 5 可以看出,提出的方法能够有效地建立实际数据的模型。

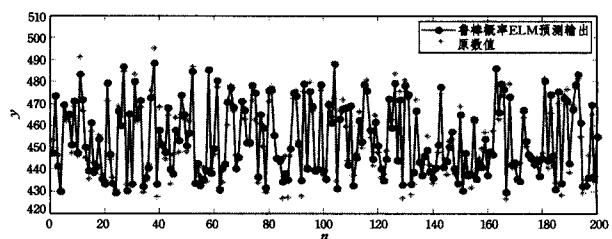


图 5 鲁棒概率 ELM 预测输出情况

鲁棒概率 ELM 与传统的 ELM、正则化 ELM 和鲁棒 ELM 的对比结果如图 6 和表 3 所示。通过图 6 和表 3 可以看出提出的新方法的性能优于其他方法。

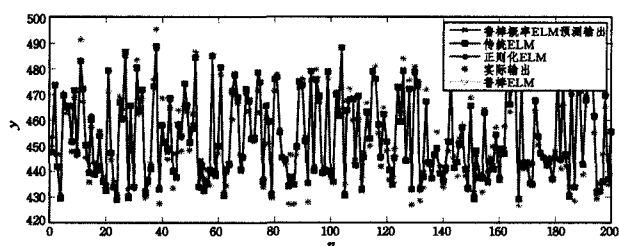


图 6 多种方法的预测情况比较

表 3 多种方法的 RMSE 比较

方法	鲁棒概率 ELM	传统 ELM	正则化 ELM	鲁棒 ELM
RMSE	3.3222	4.5781	3.9237	3.7496

**结束语** 本文提出了一种全新的鲁棒概率 ELM 建模方法,该方法对异常值点和非高斯噪声数据均有很好的模拟效果。该方法具有较好的鲁棒性的主要原因在于应用 Parzen 窗的方法建立了局部模型的概率分布,通过将概率分布作为权重融合了全部的局部模型以建立全局鲁棒 ELM,因此该方法能够非常有效地对多种噪声进行建模。将该方法应用于数学例子和 UCI 数据实例进行验证,可以发现鲁棒概率 ELM 相对于传统的 ELM、正则化 ELM 和鲁棒 ELM 具有更好的鲁棒性和准确度。

### 参 考 文 献

- [1] CHUANG C C, LEE Z J. Hybrid robust support vector machines for regression with outliers[J]. *Applied Soft Computing*, 2011, 11(1): 64-72.
- [2] WEN W, HAO Z, YANG X. Robust least squares support vector machine based on recursive outlier elimination[J]. *Soft Computing*, 2010, 14(11): 1241-1251.
- [3] YOU C C, ZHANG G G. An effective search method based on large data [J]. *Computer Science*, 2013, 40(6): 183-186. (in Chinese)  
尤川川, 张桂刚. 一种基于大数据的有效搜索方法[J]. *计算机科学*, 2013, 40(6): 183-186.
- [4] ZHANG G, LI H X, GAN M. Design a Wind Speed Prediction Model Using Probabilistic Fuzzy System[J]. *IEEE Transactions on Industrial Informatics*, 2012, 8(4): 819-827.
- [5] LI H X, LIU Z. A Probabilistic Neural-Fuzzy Learning System for Stochastic Modeling[J]. *IEEE Transactions on Fuzzy Systems*, 2008, 16(4): 898-908.
- [6] LIU Z, LI H X. A probabilistic fuzzy logic system for modeling and control[J]. *IEEE Transactions on Fuzzy Systems*, 2005, 13(6): 848-859.
- [7] BAI J, YANG L H, ZHANG X Y. An anti noise SVM parameter optimization method for speech recognition[J]. *Journal of Central South University (Natural Science Edition)*, 2013, 44(2): 604-611. (in Chinese)  
白静, 杨利红, 张雪英. 一种面向语音识别的抗噪 SVM 参数优化方法[J]. *中南大学学报(自然科学版)*, 2013, 44(2): 604-611.
- [8] HUANG G B, ZHU Q Y, SUEW C K. Extreme learning machine: Theory and applications [J]. *Neurocomputing*, 2006, 70(1-3): 489-501.
- [9] FOSSACECA J M, MAZZUCHI T A, SARKANI S. MARK-ELM: Application of a novel Multiple Kernel Learning framework for improving the robustness of Network Intrusion Detection [J]. *Expert Systems with Applications*, 2015, 42(8): 4062-4080.
- [10] FENG N Q, QIU Y H, WANG F. A new method for improving neural network generalization ability [J]. *Computer Science*, 2006, 33(2): 201-204. (in Chinese)  
冯乃勤, 邱玉辉, 王芳. 一种提高神经网络泛化能力的新方法 [J]. *计算机科学*, 2006, 33(2): 201-204.
- [11] ZHOU L, LAI K K, YU L. Least squares support vector machines ensemble models for credit scoring [J]. *Expert Systems with Applications*, 2010, 37(37): 127-133.
- [12] ABE D S. Support Vector Machines for Pattern Classification [J]. *Advances in Pattern Recognition*, 2010, 36(5): 7535-7543.
- [13] ZHU Q Y, QIN A K, SUGANTHAN P N, et al. Evolutionary extreme learning machine [J]. *Pattern Recognition*, 2005, 38(10): 1759-1763.
- [14] LU X J, ZHOU C, HUANG M H, et al. Regularized online sequential extreme learning machine with adaptive regulation factor for time-varying nonlinear system [J]. *Neurocomputing*, 2016, 174: 617-626.
- [15] CHEN C, YAN C, LI Y. A robust weighted least squares support vector regression based on least trimmed squares [J]. *Neurocomputing*, 2015, 168: 941-946.
- [16] TIAN S F, HUANG H K. A Simplification Algorithm to Support Vector Machines for Regression [J]. *Journal of Software*, 2002, 13(6): 1169-1172.
- [17] SUYKENS J A K, BRABANTER J D, LUKAS L, et al. Weighted least squares support vector machines: robustness and sparse approximation [J]. *Neurocomputing*, 2002, 48(1-4): 85-105.
- [18] CHEN X, YANG J, LIANG J, et al. Recursive robust least squares support vector regression based on maximum correntropy criterion [J]. *Neurocomputing*, 2012, 97(1): 63-73.
- [19] HUANG G B, WANG D H, LAN Y. Extreme learning machines: a survey [J]. *International Journal of Machine Learning & Cybernetics*, 2011, 2(2): 107-122.
- [20] BI H, LIANG H L, WANG Y. Resampling methods and machine learning [J]. *Journal of Computer*, 2009, 32(5): 862-877. (in Chinese)  
毕华, 梁洪力, 王珏. 重采样方法与机器学习 [J]. *计算机学报*, 2009, 32(5): 862-877.
- [21] ZHANG H, MENG X W, LIU L, et al. Improved SAR image target detection based on Parzen window algorithm [J]. *Computer Science*, 2015, 42(S2): 151-154. (in Chinese)  
张颖, 孟祥伟, 刘磊, 等. 改进的基于 Parzen 窗算法的 SAR 图像目标检测 [J]. *计算机科学*, 2015, 42(S2): 151-154.
- [22] LICHMAN M. UCI Machine Learning Repository [EB/OL]. (2014-02-26) [2016-07-15]. <http://archive.ics.uci.edu/ml/datasets.html>.
- [23] HORATA P, CHIEWCHANWATTANA S, SUNAT K. Robust extreme learning machine [J]. *Neurocomputing*, 2013, 102(2): 31-44.