Model Optimization of Air Quality with M-ELM

Jianxiong Ye\textsuperscript{1,a} Wenzhen Zhou\textsuperscript{2,b} Zhigang Li\textsuperscript{3,c} Jinlan Zhou\textsuperscript{1,d}

\textsuperscript{1}Jiangxi Province Key Laboratory of precision drive & control, Nanchang Institute of Technology, China 330099

\textsuperscript{2}School of Information Science and Engineering, Central South University, China 330013

\textsuperscript{3}School of Mechanic Engineering, East China Jiaotong University, China 330013

\textsuperscript{a}jxlpyjx@163.com, \textsuperscript{b}315578522@qq.com, \textsuperscript{c}214504207@qq.com, \textsuperscript{d}214503760@qq.com

Keywords: Extreme learning machine, M-estimator, Model performance, Air quality model

Abstract. The extreme learning machine (ELM), which was originally proposed for "generalized" single-hidden layer feedforward neural networks (SLFNs), provides efficient unified learning solutions for the applications of clustering, regression, and classification. But when the training data have been contaminated, ELM can't guarantee the model accuracy. A novel hybrid way called M-ELM is proposed to adjust the output matrix of ELM model, this way combined M-estimator with ELM to reduce the noise influence. Experimental results on UCI (University of California at Irvine) datasets and air quality detection indicate that M-ELM performs competitively good, it can be used on design of air cleaner.

I Introduction

In resent years, due to its extremely fast training, and universal approximation capability, ELM has been becoming an significant research topic for machine learning. Single-layer ELM algorithms \cite{1-4} and multi-layer ELM have been set up based on widespread SLFNs \cite{5-11}. But when the samples are heavily polluted by noise, ELM can't ensure a good fitting accuracy, so M-estimation is used to solve this problem.

II Basic-ELM Introduction

2.1 ELM structure and its mathematical express

ELM is a feedforward networks, it includes only one input layer and a output layer, one or more hidden layers, Studies have proved that multi-layer network can learn any \( N \) distinct samples with any arbitrarily small error by no more than \( 2\sqrt{(m+2)N} \) hidden neurons, where \( m \) is the number of output neurons\cite{4}, this result means that the required number of hidden neurons of ELM can be significantly decreased in application; and the most popular form is single-hidden layer feedforward networks (SLFNs), the structure of SLFN can be described as figure 1.

\begin{center}
\includegraphics[width=\textwidth]{fig1.png}
\end{center}

Fig. 1. Network structure constructed by ELM
As shown in Figure 1, L means the number of hidden neurons, g indicates activation function and m means the number of output neurons, 
\[ f(x) = \sum_{i=1}^{L} \beta_i g(a_i x + b_i) = t, j = 1, \ldots, N, \]
it can be rewritten as \( H\beta = T \), where
\[
H(a_1, \ldots, a_L, b_1, \ldots, b_L, x_1, \ldots, x_N) = \begin{bmatrix}
g(a_1 x_1 + b_1) & \cdots & g(a_L x_1 + b_L) \\
\vdots & \ddots & \vdots \\
g(a_1 x_N + b_1) & \cdots & g(a_L x_N + b_L)
\end{bmatrix}, \beta = \begin{bmatrix}
\beta_1^T \\
\vdots \\
\beta_L^T
\end{bmatrix}, T = \begin{bmatrix}
t_1 \\
\vdots \\
t_L
\end{bmatrix}
\]
\( a_i = [w_{i1}, w_{i2}, \ldots, w_{im}] \), so output matrix can be simply represented as follows:
\[
\hat{\beta} = H^T T \quad (1)
\]
Where \( H^T \) is the Moore-Penrose generalized inverse of \( H \)\(^{[2,15]}\).

### 2. Main Problems in Learning

According to equation (1), how to calculate \( H^T \) is important, and several ways have been proposed such as iterative method, orthogonalization method, orthogonal projection method and singular value decomposition (SVD)\(^{[12]}\), the problem lies in that if the training data is polluted, all these resolutions are meaningless. How to get a better output matrix needed further study.

### III Combination of M-estimation and ELM

#### 3.1 Working Process of M-ELM

The error is inevitable for sampling data, there are two kinds of them, one is small and random error, which is also called system error; another is gross error, it also call outliers, M-estimator is a kind of robust estimator and used to reduce the influence of this kind of error.

Consider regression model \( T = H\beta + e \), a new method called ME-ELM is designed:
\[
\hat{\beta} = (H^T PH)^{-1} H^T P T \quad (2)
\]
Where P is the coefficient matrix, which should be tuned according to the training errors. For noise-free samples, all the coefficients of P are set as 1, it implies that ME-ELM works in the same way as ELM; if part of samples are moderately polluted, corresponding weights are less than 1, so as to weaken the poor influence; for those with gross errors, the relating weights tend to become much more smaller, so the bad influence can be greatly reduced. Generally speaking, adjusting matrix P plays an vital role in improving model accuracy.

The weights of P are regulated by estimation function \( \psi \), which is derivation of static function \( \rho; \psi(x) = \frac{\partial \rho(x)}{\partial \beta} \), and optimization objective function \( Q(\beta) \) can be defined as:
\[
Q(\beta) = \sum_{i=1}^{N} \rho(e_i) = \sum_{i=1}^{N} \rho(T_i - H\beta_i) \quad (3)
\]
Solutions of (3) are called M-estimators:
\[
\hat{\beta} = \arg \min_{\beta} \left( \sum_{i=1}^{N} \rho(T_i - H\beta_i) \right) \quad (4)
\]
Many estimation functions \( \psi \) have been proposed, and most of them have similar results in improving model accuracy, Huber function is a popular one and is used in our research\(^{[13]}\).
\[
\psi(x) = \begin{cases}
x & |x| < k (k > 0) \\
k & |x| > k (k > 0)
\end{cases}, \quad \rho(x) = \begin{cases}
x^2 / 2 & |x| \leq k \\
k |x| - k^2 / 2 & |x| > k
\end{cases}, \quad k = 1.345 \quad (5)
\]

#### 3.2 Experimental Verification
The performance of M-ELM are tested with datasets shown in table 1, these datasets are all from UCI\textsuperscript{[14-17]}. Segmentation dataset is used to predict classification of images according to 19 attributes, the Iris is for species prediction of iris from four physical sizes, the Abalone is to estimate the age of abalone from 8 physical measurements; Bike Share dataset is used to determine the relation between bike rented hour and the data of weather, season and environment etc; CCPP contains 9568 data points taken from various sensors from 2006 to 2011 and is used to predict the energy output of electrical plant. All theses datasets are different in types, sizes and complexities, and for all the problems here, both the training data and testing data are chosen out randomly for each trial.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Type Classification/Regression</th>
<th>#Train</th>
<th>#Test</th>
<th>#Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation</td>
<td>C</td>
<td>110</td>
<td>100</td>
<td>19</td>
</tr>
<tr>
<td>Iris</td>
<td>C</td>
<td>80</td>
<td>70</td>
<td>5</td>
</tr>
<tr>
<td>Abalone</td>
<td>C/R</td>
<td>2000</td>
<td>2177</td>
<td>8</td>
</tr>
<tr>
<td>Bike Sharing (hours only)</td>
<td>R</td>
<td>10000</td>
<td>7397</td>
<td>13</td>
</tr>
<tr>
<td>CCPP</td>
<td>R</td>
<td>23920</td>
<td>23920</td>
<td>5</td>
</tr>
</tbody>
</table>

The relationship between the number of hidden neurons and network performance, as well as the working efficiency of M-ELM are studied. results are shown in table 3 when $N=30$, 60 and 100. It can be seen that testing accuracy improves with the increasing of hidden neurons number, and the training time increases synchronously.

<table>
<thead>
<tr>
<th>N=20</th>
<th>N=50</th>
<th>N=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris 40</td>
<td>0.1286</td>
<td>0.5858</td>
</tr>
<tr>
<td>Segmentation 40</td>
<td>0.1300</td>
<td>0.1872</td>
</tr>
<tr>
<td>Abalone 40</td>
<td>0.0743</td>
<td>0.4368</td>
</tr>
<tr>
<td>Bike Sharing 400</td>
<td>0.00423</td>
<td>152.9278</td>
</tr>
<tr>
<td>CCPP 400</td>
<td>0.0503</td>
<td>76.4873</td>
</tr>
</tbody>
</table>

IV Conclusions

Normal ELM works on the minimal least square error(LSE) principle, it can approximate N arbitrary samples with zero error, but if the data are contaminated, the network constructed by ELM and other algorithms can't work well, comparison tests manifest that M-ELM is useful in this case, it can weaken the outliers influence and achieve a better output matrix of SLFN. Now we are using it in air quality prediction, the samples from different sensors are often contaminated by random dust, so it is meaningful to put M-ELM in this situation.

V Acknowledgements

This work is partially founded by Science Project of Jiangxi Education Department (GJJ151129), Open Project Program of Jiangxi province key laboratory of precision drive & control, (PLPDC-KFKT-201625), Natural Science Foundation of China (CN) (51665016), Natural Science Foundation of Jiangxi Province (CN) (20151BAB207047).

References


