<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Optimising ensemble combination based on maximisation of diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Mao, Shasha; Lin, Weisi; Chen, Jiawei; Xiong, Lin</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2017</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/45740">http://hdl.handle.net/10220/45740</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2017 Institution of Engineering and Technology (IET). This paper was published in Electronics Letters and is made available as an electronic reprint (preprint) with permission of Institution of Engineering and Technology (IET). The published version is available at: [<a href="http://dx.doi.org/10.1049/el.2017.0795">http://dx.doi.org/10.1049/el.2017.0795</a>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.</td>
</tr>
</tbody>
</table>
Optimising ensemble combination based on maximisation of diversity

Shasha Mao\textsuperscript{st}, Weisi Lin, Jiawei Chen and Lin Xiong

Balancing diversity and accuracy of individuals is crucial for improving the performance of an ensemble system, since they are two important but incompatible factors for ensemble learning. When multiple individuals are combined with the corresponding weights, the diversity should be dominated by individuals and their weights, whereas the weights are normally ignored in the analysis of diversity in most research. Inspired by this, the authors propose a novel ensemble method which seeks an optimal combination to maximise diversity and accuracy of weighted individuals with the constraint on the minimal ensemble error. Furthermore, a new expression is given based on the generated individuals and their weights to exploit the diversity of an ensemble. Experimental results illustrate that the proposed method outperforms relevant existing methods.

**Introduction:** The diversity among individuals and the accuracies of individuals are two crucial factors which determine the performance of an ensemble system \[1, 2\]. Generally, an ideal ensemble should be composed of multiple most diverse and accurate individuals. Unfortunately, pursuing higher diversity is incompatible with generating more accurate individuals \[3, 4\], since all individuals need to learn for the same target. Thereby, how to effectively balance the diversity and the accuracy of individuals become one crucial problem. At present, many methods \[4–7\] have been proposed to seek the optimal combination of individuals based on the analysis of diversity and accuracy. Partial, Miao et al. \[6\] proposed a greedy iterative algorithm via a constructed measurement for the diversity. Miao et al. \[7\] applied a greedy iterative algorithm to prune more similar individuals, meanwhile retaining more accurate individuals.

According to current methods \[8, 9\], it is known that a weight is normally assigned to one individual when they are combined, whereas the weight is normally assigned to one individual when they are combined. Generally, an ideal ensemble should be composed of multiple most diverse and accurate individuals. Unfortunately, pursuing higher diversity is incompatible with generating more accurate individuals \[3, 4\], since all individuals need to learn for the same target. Thereby, how to effectively balance the diversity and the accuracy of individuals become one crucial problem. At present, many methods \[4–7\] have been proposed to seek the optimal combination of individuals based on the analysis of diversity and accuracy. Partial, Miao et al. \[6\] proposed a greedy iterative algorithm via a constructed measurement for the diversity. Miao et al. \[7\] applied a greedy iterative algorithm to prune more similar individuals, meanwhile retaining more accurate individuals.

According to current methods \[8, 9\], it is known that a weight is normally assigned to one individual when they are combined, whereas the weight is normally assigned to one individual when they are combined. Generally, an ideal ensemble should be composed of multiple most diverse and accurate individuals. Unfortunately, pursuing higher diversity is incompatible with generating more accurate individuals \[3, 4\], since all individuals need to learn for the same target. Thereby, how to effectively balance the diversity and the accuracy of individuals become one crucial problem. At present, many methods \[4–7\] have been proposed to seek the optimal combination of individuals based on the analysis of diversity and accuracy. Partial, Miao et al. \[6\] proposed a greedy iterative algorithm via a constructed measurement for the diversity. Miao et al. \[7\] applied a greedy iterative algorithm to prune more similar individuals, meanwhile retaining more accurate individuals.

**Proposed Method:** Given a set of samples $X = \{x_1, \ldots, x_N\}$ and the corresponding label set $y = \{y_1, \ldots, y_N\}$, $L$ individuals are first generated by an ensemble strategy and basic classifier models, denoted by $L = \{I_1, \ldots, I_L\}$, where $I_i$ expresses the $i$th generated individual ($i \in \{1, \ldots, L\}$). Let $w \in \mathbb{R}^N$ denote a vector of weights corresponding to $L$ individuals, the classification error $E_{\text{sea}}$ of an ensemble system is normally calculated by

\[
E_{\text{sea}} = \frac{1}{N} \sum_{n=1}^{N} \left( \arg \max_{v \in \Omega} \left( \sum_{i=1}^{L} w_i I_i(x_n) \right) = y_n \right) \tag{1}
\]

where $I_i(x_n)$ expresses the output (or probability) that $x_n$ is classified by $I_i$, $w_i$ is the weight corresponding to $I_i$, $\Omega$ denotes an indicator function, $\Omega$ is the set of class labels, and $c \in \Omega$. From (1), it is observed that the ensemble of $L$ individuals can be also considered as an ensemble of $\{w_i I_{i_1}, \ldots, w_i I_{i_L}\}$ when they are combined with $w$, and then $w_i I_{i_L}$ is regarded as a new weighted individual. It indicates that the diversity of an ensemble should be dominated by both $w$ and $\Omega$. In Fig. 1, a simple example is given to discuss diversity and accuracy of generated individuals with weighted individuals, which is performed on an ensemble of 25 individuals for Liver dataset, where the horizontal axis is the kappa value between a pair of individuals and the vertical axis is the average error of them. Obviously, the diversity among weighted individuals is different from generated individuals, especially when $w < 0$, which implies the diversity should be measured based on $\{w_i I_i\}$ in the process of combining individuals rather than $\{I_i\}$.

Based on this, a new expression $S$ is introduced to exploit the diversity and the accuracies of weighted individuals in this Letter, shown as

\[
S = \sum_{x} \left( w_i I_i(x) \right)^2 + \left( 1 - \frac{1}{N} \sum_{n=1}^{N} \left( \arg \max_{c \in \Omega} \left( \sum_{i=1}^{L} w_i I_i(x_n) \right) = y_n \right) \right) \tag{2}
\]

where $h_i = \mathbb{I}(y = I_i(X))$, $h_l = \mathbb{I}(y = I_l(X))$, and $h_i, h_l \in \mathbb{Z}^{N}$. From (2), it is observed that $S_d$ denotes the maximal probability that samples can be correctly classified by two individuals $w_i I_i$ and $w_l I_l$. Actually, $S_d$ is equal to the sum of diversity and accuracy of individuals, rewritten by

\[
s_d = (w_i - w_l)^2 (p_i + p_l) + w_i w_l p_i p_l \tag{3}
\]

where $p_i = h_i / N$, $p_l = (1 - h_l) / N$, and $p_l = (1 - h_l) / N$. It expresses the accuracy of samples which are correctly classified by $I_i$ and $I_l$, $S_d$ is the accuracy that samples are correctly classified by $I_i$, and $p_l$ is contrary to $p_l$. Essentially, the first term of (3) defines the diversity between a pair of weighted individuals, and the second term implies the joint accuracy of them. It is obvious that $w$ determines the diversity and the accuracy of an ensemble.

Hence, combining (1) and (3), we construct an target function to seek an optimal combination of individuals by maximising the diversity among weighted individuals and the joint accuracy of them, with the constraint on the minimal ensemble error, formulated by

\[
\begin{align*}
\max_{w} & \quad \frac{1}{N} \sum_{n=1}^{N} s_{ij} \\
\text{s.t.} & \quad E_{\text{sea}} \leq \varepsilon
\end{align*} \tag{4}
\]

where $\varepsilon$ is a positive constant to prevent the over-fitting. However, (4) is NP hard. Thus, for solving the above model, (4) is converted as

\[
\begin{align*}
\max_{w} & \quad w^T Q w \\
\text{s.t.} & \quad H w = p \\
& \quad w^T w = 1
\end{align*} \tag{5}
\]

where $Q$ is a symmetric matrix, $p$ is a given vector and $H = [h_1, \ldots, h_L]$. Note that $w^T Q w = \sum s_{ij} / L^2$, where $q_i = (2L - 1) h_i / N L^2$ and $q_i = (1 - h_i) / K L^2$. In (5), an approximative expression of ensemble error \[11\] is utilised as the constraint instead of $E_{\text{sea}}$ in (4), where $H w$ expresses the probability that samples are correctly classified by an ensemble of $\{w_i I_i\}$ and $p$ denotes the ideal probability of correct classification [the most idea value is equal to 1]. In order to solve (5), the Lagrange multiplier method is applied, where $w$ is updated in by

\[
w^{t+1} = (\delta H + 2Q + (\delta p) - 2v_1) (v_1 H + \delta H p) \tag{6}
\]

where $\delta$ is a penalty factor. $B = (w^T w - 1)$, and the multipliers $v_1$ and $v_2$ is updated by $v_1^{t+1} = v_1^t - \delta (H w - p)$ and $v_2^{t+1} = v_2^t - \delta (w^T w - 1)$, respectively.

**Experiments and analyses:** In this section, the proposed method is performed on 12 UCI datasets [http://www.ics.uci.edu/ml/MLRepository.html] to validate its performance. For each dataset, all samples are randomly divided into training and testing sets. In this experiment, bootstrap sample strategy [12] is applied to generate multiple individuals, and classification and regression tree [13] is employed.
as the basic classifier model. Moreover, we compare the propose method with six classical ensemble methods: simple voting (SV) [8], weighted majority vote (WMV) [8], combining classifiers by using correspondence analysis (SCANN) [14], evolutionary ensemble classifiers (EVEN) [10], focused ensemble selection (FES) [6] and ensemble based on 0–1 matrix decomposition (SWENC) [11]. For MDOEC, we set $d = 2$, $v_i^0 = 1$, $v_i^0$ is an unit vector, and $p_i$ is randomly generated values from 0.5 to 1. The parameters of compared methods are set as the values mentioned in corresponding references.

Table 1 shows the average of classification errors obtained by repeating an ensemble of 100 individuals for 50 times, denoted by (mean ± std). From Table 1, it is seen that MDOEC achieves the lowest errors on 8 datasets (labelled with bold), but only SCANN and FES perform best for 3 and 1 datasets, respectively, which illustrates that MDOEC is more effective for seeking an optimal combination than compared methods. In Table 1, the last row shows the win–tie–loss between MDOEC and six compared methods, computed by $t$-test with confidence level 0.05. The win–tie–loss values also indicate MDOEC is superior to other compared methods for most datasets. Moreover, in order to visualise the performance, we draw box-plots of classification errors for Breast and Sonar datasets given by seven ensemble methods, shown in Fig. 2. It is obviously seen that the minimum and median values of box-plots performed by MDOEC are lower than compared methods. In short, all experimental results illustrate that the proposed method effectively improves the performance of classification via maximising diversity among weighted individuals.

In addition, we make an analysis on the relation between increasing the object value ($w^T Q w$) and ensemble error in the optimisation, shown in Fig. 3. It is shown that the experimental results of 80 iterations for Heartc Dataset, where the upper sub-diagram shows the object value of each iteration, and the lower two sub-diagrams show the accuracy of training and testing sets obtained by the weights ($w_i$) corresponding to the object values in iterations, respectively. From Fig. 3, it is obviously seen that the accuracy of an ensemble is increased with the growth of object function, which demonstrates that the proposed method can effectively improve the ensemble performance via maximising the diversity and accuracies of weighted individuals.

Fig. 2 Box-plots of classification errors obtained via 50 times ensemble

Table 1: Classification errors (%) obtained by ensemble methods

<table>
<thead>
<tr>
<th>Data</th>
<th>MDOEC</th>
<th>SV</th>
<th>WMV</th>
<th>SCANN</th>
<th>EVEN</th>
<th>FES</th>
<th>SWENC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>25.7 ± 1.8</td>
<td>26.2 ± 1.8</td>
<td>26.4 ± 1.7</td>
<td>25.9 ± 1.9</td>
<td>26.4 ± 1.9</td>
<td>27.7 ± 3.1</td>
<td>26.3 ± 1.7</td>
</tr>
<tr>
<td>Dna</td>
<td>6.2 ± 0.3</td>
<td>6.3 ± 0.3</td>
<td>6.3 ± 0.3</td>
<td>6.3 ± 0.3</td>
<td>6.3 ± 0.3</td>
<td>6.1 ± 0.3</td>
<td>6.3 ± 0.3</td>
</tr>
<tr>
<td>Heartc</td>
<td>19.9 ± 1.2</td>
<td>20.5 ± 1.4</td>
<td>20.5 ± 1.4</td>
<td>20.2 ± 1.2</td>
<td>20.5 ± 1.4</td>
<td>20.2 ± 1.6</td>
<td>20.5 ± 1.3</td>
</tr>
<tr>
<td>Liver</td>
<td>35.8 ± 2.2</td>
<td>36.0 ± 1.5</td>
<td>36.0 ± 1.5</td>
<td>36.2 ± 1.5</td>
<td>36.1 ± 1.5</td>
<td>36.8 ± 1.5</td>
<td>35.9 ± 1.5</td>
</tr>
<tr>
<td>Saltsage</td>
<td>16.1 ± 0.3</td>
<td>15.6 ± 0.2</td>
<td>15.6 ± 0.2</td>
<td>15.3 ± 0.3</td>
<td>15.7 ± 0.3</td>
<td>15.7 ± 0.3</td>
<td>15.6 ± 0.3</td>
</tr>
<tr>
<td>Segment</td>
<td>9.4 ± 0.4</td>
<td>11.5 ± 0.3</td>
<td>11.4 ± 0.3</td>
<td>11.3 ± 0.3</td>
<td>11.5 ± 0.3</td>
<td>10.2 ± 0.3</td>
<td>11.5 ± 0.3</td>
</tr>
<tr>
<td>Sonar</td>
<td>18.9 ± 2.3</td>
<td>20.4 ± 2.3</td>
<td>20.4 ± 2.3</td>
<td>20.1 ± 2.2</td>
<td>20.1 ± 2.3</td>
<td>20.1 ± 2.1</td>
<td>20.3 ± 2.3</td>
</tr>
<tr>
<td>Soybean</td>
<td>10.8 ± 1.1</td>
<td>11.3 ± 0.7</td>
<td>11.5 ± 0.7</td>
<td>9.7 ± 0.5</td>
<td>11.4 ± 0.7</td>
<td>11.4 ± 0.9</td>
<td>11.4 ± 0.7</td>
</tr>
<tr>
<td>Spambase</td>
<td>10.4 ± 0.3</td>
<td>10.6 ± 0.2</td>
<td>10.5 ± 0.2</td>
<td>10.5 ± 0.2</td>
<td>10.6 ± 0.2</td>
<td>10.6 ± 0.2</td>
<td>10.5 ± 0.2</td>
</tr>
<tr>
<td>Waveform</td>
<td>16.6 ± 0.3</td>
<td>16.5 ± 0.3</td>
<td>16.5 ± 0.3</td>
<td>16.3 ± 0.3</td>
<td>16.7 ± 0.3</td>
<td>16.6 ± 0.3</td>
<td>16.5 ± 0.3</td>
</tr>
<tr>
<td>Wdbc</td>
<td>5.0 ± 0.6</td>
<td>5.3 ± 0.6</td>
<td>5.4 ± 0.3</td>
<td>5.3 ± 0.6</td>
<td>5.4 ± 0.3</td>
<td>5.3 ± 0.6</td>
<td>5.4 ± 0.6</td>
</tr>
<tr>
<td>Wine</td>
<td>3.8 ± 1.0</td>
<td>4.1 ± 1.1</td>
<td>4.0 ± 1.1</td>
<td>5.4 ± 1.1</td>
<td>4.3 ± 1.1</td>
<td>4.0 ± 1.1</td>
<td>4.1 ± 1.1</td>
</tr>
<tr>
<td>MDENC-wt-tie-loss</td>
<td>7.4±1</td>
<td>7.4±1</td>
<td>6.4±2</td>
<td>8.3±1</td>
<td>7.3±2</td>
<td>7.4±1</td>
<td></td>
</tr>
</tbody>
</table>

References


© The Institution of Engineering and Technology 2017
Submitted: 2 March 2017 E-first: 26 June 2017
doi: 10.1049/el.2017.0795
One or more of the Figures in this Letter are available in colour online.
Shasha Mao and Weisi Lin (Nanyang Technological University, Singapore)
E-mail: ssmao@ntu.edu.sg
Jiawei Chen (Xi’an University, People’s Republic of China)
Lin Xiong (Panasonic R&D Center Singapore, Singapore)