Optimising ensemble combination based on maximisation of diversity

Mao, Shasha; Lin, Weisi; Chen, Jiawei; Xiong, Lin

2017

Mao, S., Lin, W., Chen, J., & Xiong, L. (2017). Optimising ensemble combination based on maximisation of diversity. *Electronics Letters*, 53(15), 1042-1044. doi: 10.1049/el.2017.0795

<https://hdl.handle.net/10356/88377>

<https://doi.org/10.1049/el.2017.0795>

© 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: [<http://dx.doi.org/10.1049/el.2017.0795>]. 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.

Downloaded on 26 Apr 2023 12:27:19 SGT

Optimising ensemble combination based on maximisation of diversity

Shasha Mao[✉], 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. Partalas *et al.* [6] proposed a greedy selection ensemble algorithm via a constructed measurement for the diversity. Mao *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, where the weight is obtained via the analysis for the diversity and the accuracy of generated individuals [4, 10, 11]. Actually, when generated individuals are integrated with corresponding weights, the individual with its weight can be regarded as a new individual. It indicates that the diversity among individuals should be dominated by both individuals and the corresponding weights. However, the weights of individuals are always neglected in the analysis of diversity in most methods. Inspired by this, we propose a novel ensemble method that optimising ensemble combination based on the maximisation of diversity (MDOEC), which employs both generated individual and the corresponding weights to explore the diversity. In MDOEC, a new metric is introduced to measure the diversity among weighted individuals and the joint accuracy of them. Then, an objective function is constructed based on the maximisation of diversity and accuracy to seek an optimal combination of individuals, especially with the constraint on the minimal ensemble error. The details of the proposed method will be given in the following sections. Note that an individual denotes an individual classifier.

Proposed Method: Given a set of samples $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and the corresponding label set $\mathbf{y} = \{y_1, \dots, y_N\}$, L individuals are first generated by an ensemble strategy and basic classifier models, denoted by $\mathcal{L} = \{\mathcal{I}_1, \dots, \mathcal{I}_L\}$, where $\mathcal{I}_i$ expresses the i th generated individual ($i \in \{1, \dots, L\}$). Let $\mathbf{w} \in \mathbb{R}^L$ denote a vector of weights corresponding to L individuals, the classification error E_{en} of an ensemble system is normally calculated by

$$E_{\text{en}} = \frac{1}{N} \sum_{n=1}^N \mathbb{I} \left(\arg \max_{c \in \Omega} \left(\sum_{i=1}^L w_i \mathcal{I}_i(\mathbf{x}_n) \right) = y_n \right) \quad (1)$$

where $\mathcal{I}_i(\mathbf{x}_n)$ expresses the output (or probability) that $\mathbf{x}_n$ is classified by $\mathcal{I}_i$, w_i is the weight corresponding to $\mathcal{I}_i$, $\mathbb{I}$ denotes an indicator function, Ω 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 $\{w_1 \mathcal{I}_1, \dots, w_L \mathcal{I}_L\}$ when they are combined with $\mathbf{w}$, and then $w_i \mathcal{I}_i$ is regarded as a new weighted individual. It indicates that the diversity of an ensemble should be dominated by both $\mathbf{w}$ and $\mathcal{L}$. In Fig. 1, a simple example is given to compare 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 $\{w_i \mathcal{I}_i\}$ in the process of combining individuals rather than $\{\mathcal{I}_i\}$.

Fig. 1 Diagrams of diversity and accuracy. Left-hand side diagram is drawn based on 25 generated individuals, and right-hand side one is based on 25 weighted individuals where the weights are obtained by MDOEC and located on $(-1, 1)$

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

$$\mathcal{S}: s_{ij} = \frac{1}{N} \left(w_i^2 \mathbf{h}_i^t \mathbf{h}_i + w_j^2 \mathbf{h}_j^t \mathbf{h}_j - w_i w_j \mathbf{h}_i^t \mathbf{h}_j \right) \quad (2)$$

where $\mathbf{h}_i = \mathbb{I}(\mathbf{y} = \mathcal{I}_i(\mathbf{X}))$, $\mathbf{h}_j = \mathbb{I}(\mathbf{y} = \mathcal{I}_j(\mathbf{X}))$, and $\mathbf{h}_i, \mathbf{h}_j \in \mathbb{Z}^N$. From (2), it is observed that s_{ij} denotes the maximal probability that samples can be correctly classified by two individuals $w_i \mathcal{I}_i$ and $w_j \mathcal{I}_j$. Actually, s_{ij} is equal to the sum of diversity and accuracy of individuals, rewritten by

$$s_{ij} = (w_i - w_j)^2 (p_{\bar{i}j} + p_{ij}) + w_i w_j p_{ij} \quad (3)$$

where $p_{ij} = \mathbf{h}_i^t \mathbf{h}_j / N$, $p_{\bar{i}j} = \mathbf{h}_i^t (\mathbf{1} - \mathbf{h}_j) / N$, and $p_{\bar{j}i} = (\mathbf{1} - \mathbf{h}_i)^t \mathbf{h}_j / N$. p_{ij} is the accuracy that samples are correctly classified by $\mathcal{I}_i$ and $\mathcal{I}_j$, $p_{\bar{i}j}$ is the accuracy that samples are correctly classified by $\mathcal{I}_i$ but incorrectly classified by $\mathcal{I}_j$, and $p_{\bar{j}i}$ is contrary to p_{ij} . 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 $\mathbf{w}$ determines the diversity and the accuracy of an ensemble.

Hence, combining with (1), 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

$$\begin{aligned} \max_{\mathbf{w}} \quad & \sum_{i=1}^L \sum_{j=1}^L s_{ij} \\ \text{s.t.} \quad & E_{\text{en}} \leq \varepsilon \end{aligned} \quad (4)$$

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

$$\begin{aligned} \max_{\mathbf{w}} \quad & \mathbf{w}^t \mathbf{Q} \mathbf{w} \\ \text{s.t.} \quad & \mathbf{H} \mathbf{w} = \mathbf{p}_o \\ & \mathbf{w}^t \mathbf{w} = 1 \end{aligned} \quad (5)$$

where $\mathbf{Q}$ is a symmetric matrix, $\mathbf{p}_o$ is a given vector and $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_L]$. Note that $\mathbf{w}^t \mathbf{Q} \mathbf{w} = \sum_i \sum_j s_{ij} / L^2$, where $q_{ii} = (2L - 1) \mathbf{h}_i^t \mathbf{h}_i / NL^2$ and $q_{ij} = -\mathbf{h}_i^t \mathbf{h}_j / NL^2$. In (5), an approximative expression of ensemble error [11] is utilised as the constraint instead of E_{en} in (4), where $\mathbf{H} \mathbf{w}$ expresses the probability that samples are correctly classified by an ensemble of $\{w_i \mathcal{I}_i\}$ and $\mathbf{p}_o$ 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 $\mathbf{w}$ is updated in by

$$\mathbf{w}^{k+1} = (\delta \mathbf{H}^t \mathbf{H} - 2\mathbf{Q} + (\delta \beta_k - 2\nu_2^k) \mathbf{I})^{-1} (\nu_1^k \mathbf{H} + \delta \mathbf{H}^t \mathbf{p}_o) \quad (6)$$

where δ is a penalty factor, $\beta_k = (\mathbf{w}^k)^t \mathbf{w}^k - 1$, and the multipliers ν_1^k and ν_2^k is updated by $\nu_1^{k+1} = \nu_1^k - \delta(\mathbf{H} \mathbf{w} - \mathbf{p}_o)$ and $\nu_2^{k+1} = \nu_2^k - \delta((\mathbf{w}^k)^t \mathbf{w}^k - 1)$, respectively.

Experiments and analyses: In this section, the proposed method is performed on 12 UCI datasets [http://www.ics.uci.edu/mllearn/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 $\delta = 2$, $v_2^0 = 1$, v_1^0 is a unit vector, and p_o 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 $\pm$ 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.

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

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

Data	MDOEC	SV	WMV	SCANN	EVEN	FES	SWENC
Breast	25.7 $\pm$ 1.8	26.2 $\pm$ 1.8	26.4 $\pm$ 1.7	25.9 $\pm$ 1.8	26.4 $\pm$ 1.9	27.7 $\pm$ 3.1	26.3 $\pm$ 1.7
Dna	6.2 $\pm$ 0.3	6.3 $\pm$ 0.3	6.3 $\pm$ 0.3	6.3 $\pm$ 0.3	6.3 $\pm$ 0.3	6.1 $\pm$ 0.3	6.3 $\pm$ 0.3
Heartc	19.9 $\pm$ 1.2	20.5 $\pm$ 1.4	20.5 $\pm$ 1.4	20.2 $\pm$ 1.2	20.5 $\pm$ 1.4	20.2 $\pm$ 1.6	20.5 $\pm$ 1.3
Liver	35.8 $\pm$ 2.2	36.0 $\pm$ 1.3	36.0 $\pm$ 1.5	36.2 $\pm$ 1.1	36.1 $\pm$ 1.5	36.8 $\pm$ 1.8	35.9 $\pm$ 1.5
Satimage	16.1 $\pm$ 0.3	15.6 $\pm$ 0.2	15.6 $\pm$ 0.2	15.3 $\pm$ 0.2	15.6 $\pm$ 0.3	15.7 $\pm$ 0.3	15.6 $\pm$ 0.2
Segment	9.4 $\pm$ 0.4	11.5 $\pm$ 0.3	11.4 $\pm$ 0.3	11.3 $\pm$ 0.3	11.5 $\pm$ 0.4	10.2 $\pm$ 0.6	11.5 $\pm$ 0.3
Sonar	18.9 $\pm$ 2.3	20.6 $\pm$ 2.0	20.4 $\pm$ 2.3	20.1 $\pm$ 2.2	20.1 $\pm$ 2.3	20.1 $\pm$ 2.1	20.3 $\pm$ 2.1
Soybean	10.8 $\pm$ 1.1	11.3 $\pm$ 0.7	11.5 $\pm$ 0.7	9.7 $\pm$ 0.5	11.4 $\pm$ 0.7	11.4 $\pm$ 0.9	11.4 $\pm$ 0.7
Spambase	10.4 $\pm$ 0.3	10.6 $\pm$ 0.2	10.5 $\pm$ 0.2	10.5 $\pm$ 0.2	10.5 $\pm$ 0.3	10.6 $\pm$ 0.4	10.5 $\pm$ 0.2
Waveform	16.6 $\pm$ 0.3	16.5 $\pm$ 0.3	16.5 $\pm$ 0.3	16.5 $\pm$ 0.3	16.7 $\pm$ 0.3	16.6 $\pm$ 0.4	16.5 $\pm$ 0.3
Wdbc	5.0 $\pm$ 0.6	5.3 $\pm$ 0.6	5.4 $\pm$ 0.6	5.3 $\pm$ 0.6	5.4 $\pm$ 0.7	5.3 $\pm$ 0.6	5.4 $\pm$ 0.6
Wine	3.8 $\pm$ 1.0	4.1 $\pm$ 1.1	4.02 $\pm$ 1.0	5.4 $\pm$ 1.1	4.3 $\pm$ 1.1	4.00 $\pm$ 1.0	4.1 $\pm$ 1.1
MDENC:win–tie–loss		7-4-1	7-4-1	6-4-2	8-3-1	7-3-2	7-4-1

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 $\{w_t\}$ 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. 3 Diagram of the relation between objective function and accuracy in optimisation

Conclusion: This Letter introduces a new ensemble method based optimisation which employs both individuals and their corresponding weights to measure the diversity and the accuracy. An objective function is constructed to maximise the diversity among and the accuracy of weighted classifiers with the constraint on the minimal ensemble error. Experimental results illustrate that the proposed method is superior to relevant existing methods, which demonstrates that employing both generated individuals and their weights is more effective for measuring the diversity than only individuals.

© 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 (Xidian University, People's Republic of China)

Lin Xiong (Panasonic R&D Center Singapore, Singapore)

References

- 1 Zhou, Z.H.: 'Ensemble learning', Encyclopedia of Biometrics, 2015, pp. 411–416
- 2 Ren, Y., Zhang, L., and Suganthan, P.N.: 'Ensemble classification and regression-recent developments, applications and future directions [Review Article]', *IEEE Comput. Intell. Mag.*, 2016, **11**, (1), pp. 41–53
- 3 Zhang, M.L., and Zhou, Z.H.: 'Exploiting unlabeled data to enhance ensemble diversity', *Data Min. Knowl. Discov.*, 2013, **26**, pp. 98–129
- 4 Zhang, Y., Burer, S., and Street, W.N.: 'Ensemble pruning via semi-definite programming', *J. Mach. Learn. Res.*, 2006, **7**, (7), pp. 1315–1338
- 5 Zhou, Z.H., Wu, J.X., and Tang, W.: 'Ensembling neural networks: many could be better than all', *Artif. Intell.*, 2002, **137**, (1–2), pp. 239–263
- 6 Partalas, I., Tsoumakas, G., and Vlahavas, I.P.: 'Focused ensemble selection: a diversity based method for greedy ensemble selection'. 19th European Conf. on Artificial Intelligence, Patras, Greece, August 2008, pp. 117–121
- 7 Mao, S.S., Jiao, L.C., Xiong, L., and Gou, S.P.: 'Greedy optimization classifiers ensemble based on diversity', *Pattern Recogn.*, 2011, **44**, (6), pp. 1245–1261
- 8 Kuncheva, L.I.: 'Combining pattern classifiers: methods and algorithms' (John Wiley & Sons, Inc., Hoboken, NJ, 2004)
- 9 Zhou, Z.-H.: 'Ensemble methods: foundations and algorithms' (CRC press, 2012)
- 10 Chawla, N.V., and Sylvester, J.: 'Exploiting diversity in ensembles: improving the performance on unbalanced datasets', *Multi. Classif. Syst. Lect. Notes Comput. Sci.*, 2007, **4472**, pp. 397–406
- 11 Mao, S.S., Xiong, L., Jiao, L.C., Zhang, S., and Chen, B.: 'Weighted ensemble based on 0-1 matrix decomposition', *Electron. Lett.*, 2013, **49**, (2), pp. 116–118
- 12 Breiman, L.: 'Bagging predictors', *Mach. Learn.*, 1996, **24**, (2), pp. 123–140
- 13 Loh, W.Y.: 'Classification and regression trees', *Interdiscip. Rev. Data Min. Knowl. Discov.*, 2011, **1**, (1), pp. 14–23
- 14 Merz, C.J.: 'Using correspondence analysis to combine classifiers', *Mach. Learn.*, 1999, **36**, (1–2), pp. 33–58