Pattern Recognition 48 (2015) 1688–1706

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Weighted classifier ensemble based on quadratic form

Shasha Mao [a,b,*,1], Licheng Jiao [a], Lin Xiong [a,1], Shuiping Gou [a], Bo Chen [a], Sai-Kit Yeung [b]

[a] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, PR China
[b] The Vision, Graphics and Computational Design Group at the Information Systems Technology and Design, Singapore University of Technology and Design, Singapore

ARTICLE INFO

Article history:
Received 23 July 2013
Received in revised form
20 August 2014
Accepted 12 October 2014
Available online 20 November 2014

Keywords:
Ensemble learning
Weighted classifier ensemble
Quadratic form

ABSTRACT

Diversity and accuracy are the two key factors that decide the ensemble generalization error. Constructing a good ensemble method by balancing these two factors is difficult, because increasing diversity is at the cost of reducing accuracy normally. In order to improve the performance of an ensemble while avoiding the difficulty derived of balancing diversity and accuracy, we propose a novel method that weights each classifier in the ensemble by maximizing three different quadratic forms. In this paper, the optimal weight of individual classifiers is obtained by minimizing the ensemble error, rather than analyzing diversity and accuracy. Since it is difficult to minimize the general form of the ensemble error directly, we approximate the error in an objective function subject to two constraints ($\sum w_i = 1$ and $-1 < w_i < 1$). Particularly, we introduce an error term with a weight vector $\mathbf{w}_0$, and subtract this error with the quadratic form to obtain our approximated error. This subtraction makes minimizing the approximation form equivalent to maximizing the original quadratic form. Theoretical analysis finds that when the value of the quadratic form is maximized, the error of an ensemble system with the corresponding optimal weight $\mathbf{w}^*$ will be smallest, especially compared with the ensemble with $\mathbf{w}_0$. Finally, we demonstrate improved classification performance from the experimental results of an artificial dataset, UCI datasets and PolSAR image data.

1. Introduction

Ensemble learning as an active research indeed improves the performance of a single learner by combining multiple learners [1,2]. In recent years, it has been widely used in fields of not only supervised learning but also unsupervised learning [3–5]. Classifier ensemble [6–8] is considered as a classical application that ensemble learning is employed to combine multiple classifiers in supervised learning in order to improve the accuracy and stability of a single classifier. Moreover, it is also named as multiple classifier system [8]. In a classifier ensemble system, a classifier as a learner is called an individual classifier or an individual. At present, many ensemble methods [9–13,37,38] have been proposed and they roughly fall into two basic categories. One lays emphasis on how to construct individual classifiers, and the other lays emphasis on how to combine individual classifiers.

For the former, it concentrates on making different training subsets for individual classifiers, and many classical ensemble strategies have been proposed, such as bagging [2], AdaBoost [14], random forest [7], rotation forest [9] and so on. For the latter, it engages in how to combine the outputs of individual classifiers and is considered as a research hotspot of classifier ensemble recently. From the point of the value of classifiers' coefficients, the existing methods about combining classifiers are roughly divided into three categories: (a) *simple vote strategy* [8,11]: It combines all individual classifiers' outputs with same probability. In other words, all individual classifier are given a same weight coefficient in simple vote strategy. Especially, it is equivalent to majority vote [8,15] as a most popularly used rule; (b) *weighted classifier ensemble* (*WCE*) [8,16,17,39–44]: It combines individual classifiers with different weight coefficients, and the value of each weight coefficient is not equal to zero. In WCE, it indicates that each individual classifier is supposed to have a different contribution for improving the performance; (c) selective or pruning classifier ensemble [18–20,45–48]. It combines individual classifiers with a weight vector including a zero coefficient at least, which indicates that some individual classifiers have negative or insignificant effects on boosting the performance. According to Zhou et al. [21], it demonstrates that an ensemble of partial individual classifiers is better than all. In particular, our work focuses on designing a weighted classifier ensemble method in this paper.

In general, the ensemble generalization error is decided by diversity among individual classifiers and accuracy of them in an ensemble system [22,14,7]. On this basis, many ensemble

* Corresponding author at: The Vision, Graphics and Computational Design Group at the Information Systems Technology and Design, Singapore University of Technology and Design, Singapore. Tel.: +86 15399028270.
E-mail address: skymss0828@gmail.com (S. Mao).
[1] Contributed equally to this work.

http://dx.doi.org/10.1016/j.patcog.2014.10.017
0031-3203/

algorithms [9,19,38] have been proposed based on accuracy of individuals themselves or diversity among individuals. It illustrates that a good ensemble method depends on not only high accuracy of individual classifiers but also high diversity between a pair of individual classifiers. However, according to Krogh and Vedelsby [22], Zhou et al. [23] and Zhang et al. [18], it is known that the more individual classifiers are of high accuracy rates, the less the diversity among them becomes, because the class label of the target sample is uniform. In other words, enhancing diversity among individual classifiers is at the expense of decreasing their accuracy. Thus constructing a good ensemble is difficult based on diversity and accuracy.

On the other hand, some ensemble algorithms increase diversity among individual classifiers by producing different training subsets. Yet perhaps it may be resulted in an unexpected problem that individual classifiers corresponding to different training subsets gain the same outputs. It means that diversity is not always enhanced while creating classifiers by different training subsets. Moreover, the paper [24] illustrates that accuracy of individual classifiers is the leading factor in improving the ensemble performance compared against diversity among individuals. It means that some classifiers with the better performance are more helpful than ones with the higher diversity but poor performance in an ensemble. In brief, it is tough and inconclusive to design an ensemble method via balancing and analyzing diversity and accuracy. In fact, the initial intention of an ensemble of classifiers is the improvement of the classification performance, and the analysis of diversity and accuracy is also to boost the performance of an ensemble system. Consequently, it is fascinating to see whether or not we construct a method by the explicit analysis on the performance of classifier ensemble rather than facing a dilemma of balancing diversity and accuracy.

In this paper, we propose a novel weighted classifier ensemble method based on quadratic forms, which is also named by QFWEC method. In the proposed method, the ensemble error is directly utilized to seek the optimal weight vector of classifiers instead of analyzing diversity and accuracy, whereas it is difficult to obtain the optimal solution via the minimization of the ensemble error, especially for a binary classification problem. Thereby the QFWEC method converts the minimization of the ensemble error into a new optimization problem that contains an approximation form and two constraints. The approximation form is considered as the target function of seeking the optimal weight vector of the classifiers. Furthermore, the approximation form is decomposed into two parts by introducing a given weight vector. The first part is the ensemble error gained by the introduced weight vector, and the second part is a quadratic form. The value of the approximation form is equal to the one by subtracting the second part from the first part. Specifically, the first part is independent of the solving weight vector. Consequently, the process of minimizing the approximation form is transformed into maximizing the quadratic form. Finally, an optimal weight vector is sought by maximizing the quadratic form in the QFWEC method. In addition, it is found that when the value of the quadratic form is larger, the ensemble error gained by the sought weight vector is lower than the one gained by the introduced weight vector. In addition, the experimental results demonstrate that the proposed method obtains a better performance against other ensemble methods.

The organization of this manuscript is as follows. Section 2 introduces several proposed weighted classifier ensemble algorithms. Section 3 introduces the proposed method in detail, including how to change the minimization of the ensemble error into maximizing a real quadratic form and how to seek the optimal weight vector based on three different optimizations. In Section 4, the experimental results of an artificial dataset, UCI datasets and PolSAR image data are shown to illustrate that the proposed method improves the classification performance. Lastly, Section 5 concludes our work and proposes some future works.

2. Related works

As the latter part of classifier ensemble, combining individual classifiers is actually equal to assemble the predictions obtained by individual classifiers, as important as constructing individual classifiers. Suppose a given training sample set $\mathbf{X}$ with $N \times d$, $\mathbf{X} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, where y_n is the true label of $\mathbf{x}_n$ ($\mathbf{x}_n \in \mathbb{R}^d$), $y_n \in \{\omega_1, \ldots, \omega_C\}$, ω_j expresses the jth class, and C is the number of classes. In an ensemble system, Ψ denotes a set of individual classifiers, $\Psi = \{\mathscr{L}_1, \ldots, \mathscr{L}_L\}$, where $\mathscr{L}_i$ $(i = 1, \ldots, L)$ expresses an individual classifier and L is the number of individual classifiers. Then a general form of combining individual classifiers is given as follows:

$$H(\mathbf{x}_T) = \underset{\omega_j \in \{\omega_1, \ldots, \omega_C\}}{\arg\max} \left(\sum_{i=1}^{L} p_{ij}(\mathbf{x}_T) * w_i \right) \tag{1}$$

where $H(\mathbf{x}_T)$ expresses the predictive label of an unlabeled sample $\mathbf{x}_T$ ($\mathbf{x}_T \in \mathbb{R}^d$) given by an ensemble, $p_{ij}(\mathbf{x}_T)$ is the probability of $\mathbf{x}_T$ classified to ω_j by an individual classifier $\mathscr{L}_i$, and w_i denotes the weight coefficient of $\mathscr{L}_i$. When the processed problem is a two-class classification problem $(y_n \in \{-1, +1\})$, Eq. (1) is also presented in the following formula:

$$H(\mathbf{x}_T) = \mathrm{sgn}\left(\sum_{i=1}^{L} f_i(\mathbf{x}_T) * w_i \right) \tag{2}$$

where $f_i(\mathbf{x}_T)$ expresses the predictive label of $\mathbf{x}_T$ given by $\mathscr{L}_i$.

2.1. Simple vote rule

Simple vote rule [8,11] has been widely used to combine the outputs of individual classifiers in many ensemble strategies which focus on constructing different individual classifiers, such as bagging [2], random subspace method [28], rotation forest [9], and so on. In general, each individual classifier is considered to be of an effect as same as others in simple voting rule. In fact, it is equivalent to giving a same coefficient to all individuals in an ensemble system, such as $w_i = 1$ $(i = 1, \ldots, L)$ in Eq.(1). In other words, each individual is important for improving ensemble performance as same as others while simple voting rule is employed to combine classifiers. Particularly, the ensemble predictive label of an unlabeled sample $\mathbf{x}_T$ is given by the formula $H(\mathbf{x}_T) = \mathrm{sgn}\left(\sum_{i=1}^{L} f_i(\mathbf{x}_T) \right)$ for a two-class classification problem.

2.2. Weighted majority vote

Weighted majority vote [8] achieves the final decision of classifier ensemble by giving more power to more individual classifiers. It indicates that an individual has a different effect from others in an ensemble when it is not of the identical classification performance. According to Kuncheva [8] and Rodriguez [15], the probability that a sample is classified into each class is computed by weighted majority vote, shown as follows:

$$p_j^{wmv}(\mathbf{x}_T) = \sum_{i=1}^{L} w_i d_{ij}, \quad j \in \{1, \ldots, C\} \tag{3}$$

where $p_j^{wmv}(\mathbf{x}_T)$ expresses the probability that $\mathbf{x}_T$ is classified into ω_j by weighted majority vote, w_i is the weight coefficient of a classifier $\mathscr{L}_i$ and satisfies for the condition $\sum_{i=1}^{L} w_i = 1$, and if $\mathbf{x}_T$ is classified into ω_j by $\mathscr{L}_i$, $d_{ij} = 1$, otherwise, $d_{ij} = 0$. Finally, the label