



Exploiting diversity for optimizing margin distribution in ensemble learning



Qinghua Hu^{a,*}, Leijun Li^a, Xiangqian Wu^a, Gerald Schaefer^b, Daren Yu^c

^a Biometric Computing Research Centre, School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

^b Department of Computer Science, Loughborough University, UK

^c School of Energy Science and Engineering, Harbin Institute of Technology, Harbin 150001, China

ARTICLE INFO

Article history:

Received 9 December 2013

Received in revised form 29 March 2014

Accepted 9 June 2014

Available online 18 June 2014

Keywords:

Ensemble learning
Margin distribution
Diversity
Fusion strategy
Rotation

ABSTRACT

Margin distribution is acknowledged as an important factor for improving the generalization performance of classifiers. In this paper, we propose a novel ensemble learning algorithm named Double Rotation Margin Forest (DRMF), that aims to improve the margin distribution of the combined system over the training set. We utilise random rotation to produce diverse base classifiers, and optimize the margin distribution to exploit the diversity for producing an optimal ensemble. We demonstrate that diverse base classifiers are beneficial in deriving large-margin ensembles, and that therefore our proposed technique will lead to good generalization performance. We examine our method on an extensive set of benchmark classification tasks. The experimental results confirm that DRMF outperforms other classical ensemble algorithms such as Bagging, AdaBoostM1 and Rotation Forest. The success of DRMF is explained from the viewpoints of margin distribution and diversity.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Ensemble learning has been an active research area in pattern recognition and machine learning domains for more than twenty years [1,29,38,45,59]. Ensemble learning, also referred to as multiple classifier systems, committees of learners, decision forest or consensus theory, is based on the idea of training a set of base classifiers or regressors for a given learning task and combining their outputs through a fusion strategy.

A significant amount of works have been focused on designing effective ensemble classifiers [15,20,25,39]. However, an exact explanation of the success of ensemble strategies is still an open problem. Some researchers explored how an ensemble's effectiveness is related to the large margin principle, which is regarded as an important factor for improving classification [42,51]. In this paper, we propose a novel ensemble learning algorithm named Double Rotation Margin Forest (DRMF), which is designed to improve the margin distribution of ensembles by enhancing the diversity of base classifiers and exploiting this diversity using an optimization technique.

In general, there are two well-accepted viewpoints – diversity and margin – to explain the success of ensemble learning. Roughly

speaking, the margin of a sample is its distance from the decision boundary and thus reflects the confidence of the classification. The margin distribution is acknowledged as an important factor for improving the generalization performance of classifiers [2,6,11,43,49]. In [43], Shawe-Taylor et al. gave an upper bound of generalization error in terms of the margin, while in [6] a similar bound was derived for neural networks with small weights. The large margin principle has been employed to design classification algorithms in [8,14,21,26,50].

The performance of ensemble learning methods, especially boosting, has been attributed to the improvement of the margin distribution of the training set [42,51]. In AdaBoost, each new base classifier is trained by taking into account the performance of the previous base classifiers. Training samples that are misclassified by the current base classifiers play a more important role in training the subsequent one. The success of Adaboost can thus be explained from the margin distribution, where the optimization objective is to minimize a margin distribution based exponential loss function. In [42], an upper bound of the generalization error was derived in terms of the margins of the training samples, and it was shown that the generalization performance was determined by the margin distribution, the number of training samples and the number of base classifiers. The efficacy of AdaBoost thus lies in its ability of effectively improving the margin distribution. In [51], Wang et al. showed that a larger Equilibrium margin (Emargin) and a smaller Emargin error can reduce the generalization error,

* Corresponding author. Tel.: +86 22 27401839.

E-mail addresses: huqinghua@hit.edu.cn (Q. Hu), lileijun1985@163.com (L. Li), xqw@hit.edu.cn (X. Wu), Gerald.Schaefer@ieee.org (G. Schaefer), yudaren@hit.edu.cn (D. Yu).

and demonstrated that AdaBoost can produce a larger Emargin and a smaller Emargin error.

It is acknowledged that the diversity among the members of an ensemble is crucial for performance improvement. Intuitively, no improvement can be achieved when a set of identical classifiers are combined. Diversity thus allows different classifiers to offer complementary information for classification, which in turn can lead to better performance [28]. A number of techniques have been proposed to introduce diversity. In general, we can divide these into two categories: classifier perturbation and sample perturbation approaches. Classifier perturbation refers to the adoption of instability of learning algorithms [10,36] such as decision trees and neural networks. Since they are sensitive to initialization, trained predictors may converge to different local minima if started from different initializations, and diversity can thus be generated from trained classifiers. Sample perturbation techniques train classifiers on different sample subsets or feature subsets, and include bagging, boosting, random subspaces and similar approaches [4,7,17,41].

Since both diversity and margin are argued to explain the success of ensemble learning, it appears natural to question whether there is a connection between the two. Tang et al. [46] proved that maximizing the diversity among base classifiers is equivalent to optimizing the margin of an ensemble on the training samples if the average classification accuracy is constant and maximal diversity is achievable. Consequently, increasing the diversity among base classifiers is an effective method to improve the margin of ensembles. Our work is motivated by this conclusion, and our aim is to improve the margin distribution of ensembles.

In our proposed approach, we enhance the diversity of base classifiers by perturbing the samples using double random rotation. This idea is inspired by the PCA rotation proposed in the Rotation Forest algorithm [39]. In Rotation Forest, a candidate feature set is randomly split into K subsets and Principal Component Analysis (PCA) is conducted on each subset to create diverse training samples. Diversity is thus promoted through the random feature splits for different base classifiers. In our work, the feature sets are also randomly split into K subsets. In order to introduce further diversity between the base classifiers, we apply PCA and Locality Sensitive Discriminant Analysis (LSDA). In particular, we first perform unsupervised rotation with PCA, and then employ supervised large-margin rotation with LSDA. LSDA [12], as a supervised method, is able to derive a projection which maximizes the margin between data points from different classes. Our experimental results show that the applied Double Rotation can consistently enhance the diversity in a set of base classifiers.

We further exploit the diversity and improve the margin distribution with an optimal fusion strategy. In principle, there are two kinds of fusion strategies. One approach is to combine all available classifiers, e.g., in simple (plurality) voting (SV) [28] or through linear or non-linear combination rules [5,9,19,48]. The other method is to derive selective ensembles, or pruned ensembles such as LP-AdaBoost [23] or genetic algorithm (GA)-based approaches [53], which only select a fraction of the base classifiers for decision making and discard the others. Clearly, the key problem here is how to find an optimal subset of base classifiers [32]. In the GASEN approach [55], neural networks are selected based on evolved weights to constitute the ensemble. In [54], the subset selection problem is formulated as a quadratic integer programming problem, and semi-definite programming is adopted to select the base classifiers. Both GASEN and semi-definite programming are global optimization methods and thus their computational complexity is rather high. Suboptimal ensemble pruning methods were proposed to overcome this drawback, including reduce-error pruning [31], margin distance minimization (MDM) [33], orientation ordering [34], boosting-based ordering [35], and expectation propagation [13]. In practice, users would prefer sparse ensembles since computational

resources are often limited [57]. In this paper, we introduce a technique to improve the margin distribution by minimizing the margin induced classification loss. In our pruned ensembles, the weights of base classifiers are trained with L_1 regularized squared loss [56]. The base classifiers are then sorted according to their weights, and those with large weights are selected in the final ensemble.

Our presented work comprises three major contributions. First, since diversity is considered to be an important factor which affects the classification margin, Double Rotation is proposed to enhance the diversity among base classifiers. Second, we present a new pruned fusion method based on the Lasso technique for generating ensembles with optimal margin and sparse weight vectors, where the weights are learned through minimization of the regularized squared loss function. Third, we present an extensive set of experiments to evaluate the effectiveness and explain the rationality of the proposed algorithm. We convincingly show that it can improve the margin distribution to a great extent and lead to powerful ensembles.

The remainder of the paper is organized as follows. Related work is introduced in Section 2. Section 3 describes our proposed algorithm, while an analysis in terms of parameter sensitivity

Table 1
Statistics of classification tasks.

Dataset	Instances	Discrete features	Continuous features	Classes
Australian	690	8	6	2
Crx	690	9	6	2
Cmc	1473	7	2	3
Derm	366	0	34	6
German	1000	13	7	2
Glass	214	0	9	6
Heart	270	0	13	2
Horse	368	15	7	2
ICU	200	16	4	3
Iono	351	0	34	2
Iris	150	0	4	3
Movement	360	0	90	15
Pima	768	0	8	2
Rice	104	0	5	2
Spectf	269	0	44	2
Thyroid	215	0	5	3
Wiscon	699	0	9	2
Wdbc	569	0	30	2
Yeast	1484	0	7	2
Zoo	101	15	1	7

Table 2
Classification performance of DRMF with different numbers of splits.

Data set	$K=2$	$K=3$	$K=4$	$K=5$
Australian	88.11 ± 3.48	87.24 ± 4.10	87.10 ± 2.94	87.97 ± 3.63
Crx	86.37 ± 13.66	86.22 ± 15.01	86.53 ± 13.58	86.37 ± 14.28
Cmc	54.24 ± 3.25	54.45 ± 3.75	55.54 ± 3.43	52.89 ± 2.65
Derm	96.75 ± 3.84	95.95 ± 4.28	96.19 ± 3.22	95.71 ± 3.20
German	77.80 ± 2.94	76.70 ± 3.33	75.60 ± 3.63	76.40 ± 4.01
Glass	76.64 ± 10.61	73.37 ± 8.64	75.69 ± 11.01	62.06 ± 14.65
Heart	84.44 ± 4.88	83.70 ± 5.00	86.30 ± 3.51	80.00 ± 6.34
Horse	93.49 ± 3.83	91.85 ± 5.25	92.94 ± 4.06	93.21 ± 3.88
ICU	93.56 ± 4.80	89.45 ± 11.53	93.61 ± 3.99	90.98 ± 8.51
Iono	93.47 ± 4.76	95.24 ± 3.69	93.52 ± 5.05	95.21 ± 4.56
Iris	98.67 ± 2.81	94.00 ± 4.92	96.00 ± 3.44	96.00 ± 3.44
Movement	82.44 ± 16.29	81.56 ± 17.95	81.00 ± 16.63	80.22 ± 19.09
Pima	78.78 ± 3.76	77.35 ± 4.90	77.87 ± 4.61	73.57 ± 3.65
Rice	89.82 ± 13.17	79.05 ± 10.04	79.96 ± 10.70	79.05 ± 10.04
Spectf	82.48 ± 7.91	83.28 ± 3.08	82.58 ± 7.00	83.94 ± 7.63
Thyroid	96.26 ± 4.86	93.48 ± 7.92	93.96 ± 8.17	93.48 ± 5.93
Wiscon	97.86 ± 2.36	97.57 ± 3.16	96.57 ± 3.24	92.99 ± 3.71
Wdbc	97.72 ± 1.66	95.80 ± 3.42	97.21 ± 2.49	96.84 ± 2.58
Yeast	73.25 ± 3.47	72.57 ± 3.73	70.68 ± 5.80	70.68 ± 5.80
Zoo	94.39 ± 8.39	92.39 ± 11.40	93.14 ± 9.63	94.39 ± 8.39
Average	86.83	85.06	85.60	84.10

The best result for each data set is highlighted in bold face.

Download English Version:

<https://daneshyari.com/en/article/405080>

Download Persian Version:

<https://daneshyari.com/article/405080>

[Daneshyari.com](https://daneshyari.com)