基于双错测度的极限学习机选择性集成方法

夏平凡 倪志伟 朱旭辉* 倪丽萍

(合肥工业大学管理学院 合肥 230009)

(过程优化与智能决策教育部重点实验室 合肥 230009)

摘 要：极限学习机(ELM)具有学习速度快、易实现和泛化能力强等优点，但单个ELM的分类性能不稳定。集成学习可以有效地提高单个ELM的分类性能，但随着数据规模和基ELM数目的增加，计算复杂度会大幅度增加，消耗大量的计算资源。针对上述问题，该文提出一种基于双错测度的极限学习机选择性集成方法(DFSEE)，同时从理论和实验的角度进行了详细分析。首先，运用bootstrap 方法重复抽取训练集，获得多个训练子集，在ELM上进行独立训练，得到多个具有较大差异性的基ELM，构成基ELM池；其次，计算出每个基ELM的双错测度，将基ELM按照双错测度的大小进行升序排序；最后，采用多数投票算法，根据顺序将基ELM逐个累加集成，直至集成精度最优，即获得基ELM最优子集成，并分析了其理论基础。在10个UCI数据集上的实验结果表明，较其他方法使用了更小规模的基ELM，获得了更高的集成精度，同时表明了其有效性和显著性。

关键词：选择性集成；双错测度；极限学习机

中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2020)11-2756-09

DOI: 10.11999/JEIT190617

Selective Ensemble Method of Extreme Learning Machine Based on Double-fault Measure

XIA Pingfan NI Zhiwei ZHU Xuhui NI Liping

(*School of Management*, *Hefei University of Technology*, *Hefei* 230009, *China*)

(*Key Laboratory of Process Optimization and Intelligent Decision-making*, *Ministry of Education*, *Hefei* 230009, *China*)

Abstract: Extreme Learning Machine (ELM) has unique advantages such as fast learning speed, simplicity of implementation, and excellent generalization performance. However, the performance of a single ELM is unstable in classification. Ensemble learning can effectively improve the classification ability of single ELMs, but it may incur the rapid increase in memory space and computational overheads as the increase of the data size and the number of ELMs. To address this issue, a Selective Ensemble approach of ELM based on Double-Fault measure (DFSEE) is proposed, and it is evaluated by theoretical and experimental analysis simultaneously. Firstly, multiple training subsets extracted from a training dataset are obtained employing the bootstrap sampling method, and an initial pool of base ELMs is constructed by independently training multiple ELMs on different training subsets; Secondly, the ELMs in pool are sorted in ascending order according to their double-fault measures of those ELMs. Finally, it starts with one ELM and grows the ensemble by adding new base ELMs according to the order, the final ensemble of ELMs can be achieved with the best classification ability, and the theoretical basis of DFSEE is analyzed. Experimental results on 10 benchmark classification tasks show that DFSEE can achieve better results with less number of ELMs by comparing with other approaches, and its validity and significance.

Key words: Selective ensemble; Double-fault measure; Extreme Learning Machine (ELM)

收稿日期：2019-08-12；改回日期：2020-06-21；网络出版：2020-07-17

*通信作者： 朱旭辉 zhuxuhui@hfut.edu.cn

基金项目：国家自然科学基金(91546108, 71521001)，安徽省自然科学基金 (1908085QG298, 1908085MG232)，过程优化与智能决策教育部重点实验室开放课题，中央高校基本科研业务费专项资金(JZ2019HGTA0053, JZ2019HGBZ0128)

Foundation Items: The National Natural Science Foundation of China (91546108, 71521001), The Anhui Provincial Natural Science Foundation (1908085QG298, 1908085MG232), The Open Research Fund Program of Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, The Fundamental Research Funds for the Central Universities (JZ2019HGTA0053, JZ2019HGBZ0128)

1　引言

极限学习机(Extreme Learning Machine, ELM)是由Huang等人[1]首次提出的一种单隐层前馈神经网络学习算法，其无需对神经网络参数进行迭代调整，通过对神经网络中的输入权值和隐层节点的偏置进行随机赋值，并通过简单计算解析出网络的输出权值，大幅提高了神经网络的学习速度[2]，故ELM具有学习速度快、易实现和泛化能力强等特点。考虑到ELM的优势，它被广泛应用于诸多领域，并取得了良好的效果[3,4]。事实上，ELM可以被当作是一种随机神经网络，鉴于其隐层节点参数生成的随机性，故ELM存在分类性能不稳定的缺点[2]。

基于此，部分学者利用集成学习技术提高ELM分类性能。文献[5]提出了在线连续ELM集成方法(Ensemble of Online Sequential Extreme Learning Machine, EOS-ELM)，其性能明显优于OS-ELM；Ksieniewicz等人[6]将多个ELM进行集成，并应用于遥感和高光谱数据的分类中，提升了ELM的分类性能。诸多学者采用集成学习技术将多个ELM进行集成，ELM的分类性能虽然获得一定的提升，但是随着数据规模和ELM数目的增加，需要消耗大量的计算资源和开销[4,7]。为了解决这个难题，本文尝试运用选择性集成方法，其主要目的是选择集成系统中的部分基ELM进行综合集成，从而显著提升集成分类能力，并明显降低计算资源消耗。对于包含M个基分类器的集成，其有2^M-1个子集成，选择性集成是从中选择出性能最优的子集成，其实质是一个组合优化问题，也是一个NP难问题[8,9]。为了找到性能最优的子集成，学者们提出了很多选择性集成方法。

较经典的选择性集成方法有：Martínez-Muñoz等人[10,11]提出了基于差异性测度的排序集成方法，采用差异性测度度量每一个基分类器，并将测度高于某一预先设定阈值的基分类器进行综合集成，实现集成性能的提升；陆慧娟等人[4]将集成系统中不一致测度最大的基分类器进行剔除，并对余下的基分类器进行综合集成，以提高分类能力；Guo等人[12]提出了基于边界测度的选择性集成方法，保留部分边界测度较大的基分类器，进行集成；Dai 等人[13]提出了反向降错选择性集成方法(Reverse Reduce-Error, RRE)，将综合性能较优和最差的基分类器进行融合，获得最终的集成结果；Zhou等人[14]提出了基于遗传算法的选择性集成方法，先将随机权重赋予集成系统中每一个基分类器，并使用遗传算法(Genetic Algorithm, GA)进行反复迭代优化，选择权重超过预先设置阈值的基分类器进行集成；Cavalcanti等人[15]提出一种融合成对多样性测度的选择性集成算法(DivP)，计算基分类器的5种成对差异性测度，优化其权重构建新的组合测度，并引入图着色方法剔除冗余的基分类器，最后集成剩下的基分类器；Ykhlef等人[8]提出一种基于简单联盟博弈的选择性集成算法(Pruning methodology using non-monotone Simple Coalitional Games, SCGP)，使用基分类器的班扎夫权力系数大小衡量其差异性贡献，获得最小赢得联盟，并对联盟成员基分类器进行综合集成。

上述选择性集成方法虽然使得集成性能获得了一定的提升，但仅剔除了部分的冗余基分类器，不能较大限度地剔除综合性能较差的基分类器，且缺乏理论依据，未能同时从实验和理论的角度进行分析。针对现有方法的不足，本文采用双错测度对基分类器进行逐个排序集成，从大量的实验中发现了一条新的规律，即基于双错测度的排序集成，随着基分类器数目的增加，其集成精度先上升至某一数值后开始下降，并分析了其理论基础，同时从实验和理论的角度分析了方法的可行性。基于此，基于双错测度的ELM选择性集成方法，可以大幅度剔除冗余基分类器，获得集成性能较优的基分类器子集成，显著提高ELM的集成性能。

本文的主要工作如下：(1)提出了基于双错测度的ELM选择性集成方法(Selective Ensemble of ELM based on Double-Fault measure; DFSEE)，同时从实验和理论的角度分析了方法的有效性；(2)通过标准UCI数据集上的实验，发现一条规律，基于双错测度的排序集成，随着基ELM的规模的增加，其集成精度先上升后下降；(3)基于双错测度的选择性集成，可以有效地选择出性能和差异性均较好的基ELM，其先计算每个基ELM 的双错测度，再按照从小到大顺序进行排序集成，直至集成精度不再增加，即获得最终子集成。在10个UCI数据集上的实验结果表明，DFSEE所选择的基分类器规模较小，最终的集成分类能力更优，有效提升了ELM的分类性能，同时也验证了本文方法的有效性和显著性。

2　基于双错测度的ELM选择性集成

一个集成具有较优分类性能的关键在于分类器间的差异性。然而，应该采用何种标准衡量分类器间的差异性？如何权衡分类器间的差异性和基分类器的性能？为了增加基分类器间的差异性，而降低了基分类器的分类准确率，有必要吗？上述问题都是有待于解决的难题[4,10]。在实际应用中，我们应

该找到基分类器间差异性和其分类精度之间的某种稳态关系[16]。针对分类器间的差异性评估问题，学者们提出了很多差异性测度，但至今仍没有找到一种被普遍认可的，且具有良好效果的差异性测度[17]。基于此，当采用不同种类的差异性测度评估基分类器时，有必要同时从理论和实验的角度出发，进行全面、详细地分析。

假设有N个样本$X=\{x_1,x_2,\cdots,x_N\}$，M个基分类器$F=\{f_1,f_2,\cdots,f_M\}$，$Y=\{y_1,y_2,\cdots,y_N\}$为N个样本的类别，$f_i(x_k)(1\le i\le M,1\le k\le N)$为样本$x_k$被基分类器$f_i$分类后的输出结果。则第$i$个分类器$f_i$和第$j$个分类器$f_j$的联合分布[17]，见表1。

表 1 两个分类器的联合分布

	$f_i(x_k)=y_k$	$f_i(x_k)\ne y_k$
$f_j(x_k)=y_k$	a	b
$f_j(x_k)\ne y_k$	c	d

表1中，在f_i和f_j在样本X上的输出结果中，a是f_i和f_j均分类正确的样本规模；b是f_i分类正确，同时f_j分类错误的样本规模；c是f_i分类错误，同时f_j分类正确的样本规模；d是f_i和f_j均分类错误的样本规模，且满足$a+b+c+d=N$。目前较为常用的5种成对差异性测度分别为相关系数、Q统计量、Kappa测度、不一致测度、双错测度[18]，如式(1)—式(5)所示[15]。

$$\rho_{ij}=\frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \tag{1}$$

$$Q_{ij}=\frac{ad-bc}{ad+bc} \tag{2}$$

$$\mathrm{Kp}_{ij}=\frac{2(ad-bc)}{(a+b)(b+d)+(a+c)(c+d)} \tag{3}$$

$$\mathrm{Dis}_{ij}=\frac{b+c}{N} \tag{4}$$

$$\mathrm{DF}_{ij}=\frac{d}{N} \tag{5}$$

2.1 基于双错测度的ELM选择性集成方法(DFSEE)

基于双错测度的ELM选择性集成方法，先计算在相同样本上不同基ELM均分类错误的样本规模；再计算基ELM之间的双错测度，从而形成双错测度矩阵；接着根据双错测度矩阵，计算出各基ELM与全体基ELM之间的双错测度；最后根据各基ELM的双错测度，按照从小到大的顺序进行排序集成，选取集成效果最好的基ELM集合。M个基ELM之间的双错测度矩阵计算公式为

$$\mathbf{DF}=\begin{bmatrix}\mathrm{DF}_{11} & \mathrm{DF}_{12} & \cdots & \mathrm{DF}_{1M}\\ \mathrm{DF}_{21} & \mathrm{DF}_{22} & \cdots & \mathrm{DF}_{2M}\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{DF}_{M1} & \mathrm{DF}_{M2} & \cdots & \mathrm{DF}_{MM}\end{bmatrix} \tag{6}$$

由于基ELM f_i和f_j之间的双错测度$\mathrm{DF}_{ij}=\mathrm{DF}_{ji}$，故双错测度矩阵**DF**为对称矩阵。基ELM$f_i$与全体基ELM之间的双错测度为

$$\mathrm{DF}_i=\sum_{j=1}^{M}\mathrm{DF}_{ij} \tag{7}$$

令$\xi_i=\mathrm{DF}_i(i=1,2,\cdots,M)$，则可以得到各个基ELM的双错测度向量$\boldsymbol{\xi}=(\xi_1,\xi_2,\cdots,\xi_i,\cdots,\xi_M)$。根据第3节中的定理可知，剔除部分双错测度较大的基ELM，可以提高整体集成精度。但是，若剔除基ELM数目较少，则集成系统中必然存在其他冗余的基ELM，影响集成精度；若剔除数目太多，则最后所选择的基ELM个体规模太小，进而使得所选择的基ELM差异性较小，会降低集成分类器的分类性能。那么剔除多少数目的基ELM，才可以使得集成效果最好呢?

为解决上述问题，将各个基ELM的双错测度向量$\boldsymbol{\xi}=(\xi_1,\xi_2,\cdots,\xi_i,\cdots,\xi_M)$，按照从小到大规则进行排序，得到新的序列$\boldsymbol{\xi}'=(\xi'_1,\xi'_2,\cdots,\xi'_i,\cdots,\xi'_M)$，其中$\xi'_1\le\xi'_2\le\cdots\le\xi'_i\le\cdots\le\xi'_M$，对应的基ELM序列为$F'=\{f'_1,f'_2,\cdots,f'_M\}$，对应的分类准确率为$p'_1,p'_2,\cdots,p'_M$，保留前$M^*$个双错测度较小的基ELM。$M^*$的确定方式为

$$M^*=\left\{i\left|\left(1-p'_1-\sum_{j=2}^{i}\mathrm{df}_j\right)>0\right.\right.\\ \left.\cap\left(1-p'_1-\sum_{j=2}^{i+1}\mathrm{df}_j\right)<0\right\} \tag{8}$$

其中，df_j表示第j个基ELMf'_j与未加入f'_j之前$F'=\{f'_1,f'_2,\cdots,f'_{j-1}\}$的双错测度，$1\le j\le i\le M$，然后运用多数投票方法，对基ELM集合$\{f'_1,f'_2,\cdots,f'_{M^*}\}$进行综合集成。

2.2 DFSEE具体流程

DFSEE的基本思想是：先采用bootstrap抽样方法，独立训练多个基ELM；再计算每个基ELM的双错测度，并按照双错测度从小到大的顺序进行排序；最后选择出前M^*个双错测度较小的基ELM进行集成，即为集成性能最优的基ELM子集合。

算法 DFSEE基本过程如下：

输入：多数投票集成系统，训练样本，测试样本。

输出：集成分类器的输出结果p_{M^*}，所选择的基ELM子集成$F'=\{f'_1,f'_2,\cdots,f'_{M^*}\}$

步骤 1　运用bootstrap方法重复抽取训练样本M次，获得M个训练子集，并使用ELM进行独立训练，从而得到M个基ELM $F=\{f_1,f_2,\cdots,f_M\}$；

步骤 2　将M个基ELM在测试样本上进行测试，获得输出结果$\{\mathrm{pre}_1,\mathrm{pre}_2,\cdots,\mathrm{pre}_M\}$及分类准确率$\{p_1,p_2,\cdots,p_M\}$；

步骤 3　统计基ELMf_i和f_j之间的联合分布表，计算基ELMf_i和f_j之间的双错测度DF_{ij}；

步骤 4　计算基ELMf_i与全体基ELM的双错测度DF_i；

步骤 5　循环步骤3和步骤4，获得M个基ELM的双错测度向量$\boldsymbol{\xi}=(\xi_1,\xi_2,\cdots,\xi_i,\cdots,\xi_M)$；

步骤 6　将双错测度向量$\boldsymbol{\xi}=(\xi_1,\xi_2,\cdots,\xi_i,\cdots,\xi_M)$按照从小到大顺序进行排序，可得新的序$\xi'=(\xi'_1,\xi'_2,\cdots,\xi'_i,\cdots,\xi'_M)$，以及新的分类准确率序列$\{p'_1,p'_2,\cdots,p'_M\}$；

步骤 7　令$t=1$，将ξ'_t对应的基ELMf'_t加入到所选择的基ELM集合 $F'=\{f'_1,f'_2,\cdots,f'_t\}$中；

步骤 8　将所选择的基ELM集合F'中的基ELM，采用多数投票法进行集成，输出集成结果记为pre_t；

步骤 9　计算f'_{t+1}与F'的双错测度，记为df_{t+1}；

步骤 10　令$t=t+1$，若$\left(1-p'_1-\sum_{j=2}^{t}\mathrm{df}_j\right)>0\cap\left(1-p'_1-\sum_{j=2}^{t+1}\mathrm{df}_j\right)<0$，则循环步骤7—步骤9；否则终止循环；

步骤 11　令$M^*=t$，则 $F'=\{f'_1,f'_2,\cdots,f'_{M^*}\}$；

步骤 12　输出集成结果p_{M^*}和所选择的基ELM集合 $F'=\{f'_1,f'_2,\cdots,f'_{M^*}\}$。

3　理论分析

在基于差异性测度的选择性集成领域，缺乏同时从理论和实验的角度分析其有效性的研究。因此，应该从理论和实验的角度分析，集成系统中的双错测度与集成精度之间的关系[16]。实验分析将在第4节进行，理论分析将在本节展开。

定理 基于ELM的集成系统中，集成前M^*个双错测度较小的基ELM，则集成分类器的分类精度最高。

证明 在证明过程中所使用的基本符号定义和说明如下：

N为样本的规模；M为基ELM的规模；p_i为基ELM f_i在测试样本上的分类准确率，$0<p_i<1$；p_j为基ELMf_j在测试样本上的分类准确率，$0<p_j<1$；η_{ijk}为基ELM f_i和f_j在第k个测试样本上的双错测度；N_t为测试样本数目；c为样本的类别，$c=1,2,\cdots,C$；$\bar{p}$为基ELM的平均分类准确率，$\bar{p}=\dfrac{1}{M}\sum_{j=1}^{M}p_j$。

采用Bootstrap抽样方法从N个样本中随机选取$N'(N'<N)$个样本M次，分别采用ELM训练，则可以获得M个独立的基ELM。令$\eta_{ij}=N\cdot\xi_{ij}=N\cdot\mathrm{DF}_{ij}=d$，即$f_i$和$f_j$在$N_t$个测试样本上同时分类错误的数目，则$\eta_i=N\cdot\xi_i=N\cdot\mathrm{DF}_i$。因此，$\eta_i$的数学期望 $\mathrm{E}\eta_i=N\cdot\mathrm{E}\xi_i=N\cdot\mathrm{E}(\mathrm{DF}_i)$，即$\mathrm{E}\eta_i$与$f_i$的双错测度成正相关。

当第k个样本属于$c(c\le C)$类时，若$\eta_{ijk}=1$，即f_i和f_j同时分类错误，则此时的事件概率为$p_{+c}\{\eta_{ijk}=1\}=(1-p_i)\cdot(1-p_j)$；若$\eta_{ijk}=0$，即$f_i$和$f_j$同时分类正确或其中一个分类错误，则此时的事件概率为$p_{+c}\{\eta_{ijk}=0\}=p_i\cdot p_j+p_i\cdot(1-p_j)+(1-p_i)\cdot p_j$。考虑到$N_t$个测试样本为相互独立的，故$\eta_{ij}$在每一个测试样本上的期望都是相等的。假设$N_c(c\le C)$为属于第$c$类样本的数目，且 $N_1+N_2+\cdots+N_l+\cdots+N_c=N_t$，则

$$
\begin{aligned}
\mathrm{E}\eta_i&=\mathrm{E}\sum_{j=1}^{M}\eta_{ij}=\sum_{j=1}^{M}\mathrm{E}\eta_{ij}=\sum_{j=1}^{M}\sum_{k=1}^{N_t}\mathrm{E}\eta_{ijk}\\
&=\sum_{j=1}^{M}\left\{\sum_{k=1}^{N_1}[p_{+1}\{\eta_{ijk}=0\}\cdot0+p_{+1}\{\eta_{ijk}=1\}\cdot1]\right\}+\cdots\\
&\quad+\sum_{j=1}^{M}\left\{\sum_{k=1}^{N_c}[p_{+c}\{\eta_{ijk}=0\}\cdot0+p_{+c}\{\eta_{ijk}=1\}\cdot1]\right\}+\cdots\\
&\quad+\sum_{j=1}^{M}\left\{\sum_{k=1}^{N_C}[p_{+C}\{\eta_{ijk}=0\}\cdot0+p_{+C}\{\eta_{ijk}=1\}\cdot1]\right\}\\
&=\sum_{j=1}^{M}\sum_{c=1}^{C}\sum_{k=1}^{N_c}p_{+c}\{\eta_{ijk}=1\}\cdot1\\
&=\sum_{j=1}^{M}\sum_{c=1}^{C}\sum_{k=1}^{N_c}p_{+c}\{\eta_{ijk}=1\}\cdot N_c\\
&=\sum_{j=1}^{M}\sum_{c=1}^{C}\sum_{k=1}^{N_c}(1-p_i)\cdot(1-p_j)\cdot N_c\\
&=\sum_{j=1}^{M}(1-p_i)\cdot(1-p_j)\cdot\sum_{c=1}^{C}N_c\\
&=\sum_{j=1}^{M}(1-p_i)\cdot(1-p_j)\cdot N_t\\
&=N_t\cdot\sum_{j=1}^{M}(1+p_i\cdot p_j-p_i-p_j)\\
&=N_t\cdot M\cdot(\bar{p}-1)p_i+N_t\cdot M\cdot(1-\bar{p})
\end{aligned}
\tag{9}
$$

由于$0<\bar{p}<1$，即$\bar{p}-1<0$，则$\mathrm{E}\eta_i$为随p_i单调

递减。其仅仅表示f_i的双错测度的期望值与其分类准确率成负相关，而并非与集成系统的集成精度成负相关。从而说明，本文预剔除的双错测度较大的基ELM，其性能较差，同时也说明剔除双错测度较大的基ELM的合理性。

下面针对M个基ELM在第k个样本的分类结果进行分析，假设第k个样本属于第c类。

假设M个基ELM在第k个样本，分类结果为第$1,2,\cdots,c,\cdots,C$类的基ELM数目，分别$m_1,m_2,\cdots,m_c,\cdots,m_C$，则按照多数投票算法的集成结果为$c'=\underset{c'(1\le c'\le C)}{\arg\max}\,(m_{c'})$。若$c'=c$，则集成结果正确；若$c'\ne c$，则集成结果错误。

当剔除在第k个样本上双错测度较大的基ELM时，$m_1,m_2,\cdots,m_c,\cdots,m_C$中除了$m_c$之外，$m_1,m_2,\cdots,m_{c-1},m_{c+1},\cdots,m_C$数目均有可能减少。因此，条件$c=c'=\underset{c'(1\le c'\le C)}{\arg\max}\,(m_{c'})$成立的概率将会增加，即在第$k$个样本集成结果正确的准确率将会提高。进而推广到$N$个测试样本，即为剔除双错测度较大的基ELM，整体集成精度将会提高。

将数目M个基ELM按照双错测度从小到大排序序列为$f'_1,f'_2,\cdots,f'_M$，其双错测度序列为$\xi'_1,\xi'_2,\cdots,\xi'_i,\cdots,\xi'_M$，每个基ELM对应的分类准确率分别为$p'_1,p'_2,\cdots,p'_M$。按照2.2小节DFSEE的基本流程，将基ELM按照双错测度从小到大的顺序，逐个加入到基ELM集合F'中，不断地修正F'中分类错误的样本，从而达到提升集成分类性能的目的。

当将第1个基ELM f'_1加入到F'中，$F'=\{f'_1\}$，则集成后在N个样本上分类错误的样本数目为$(1-p'_1)\cdot N_t$；当将f'_2加入到F'时，f'_2在除了f'_1之外，其双错测度最小。即表示f'_2与F'在相同样本上同时分类错误的样本数目最小，其与F'的互补性最大，可以最大限度地修正F'中分类错误的样本。故f'_2可以修正的分类错误样本数目为$(1-p'_1-\mathrm{df}_2)\cdot N_t$，其中$\mathrm{df}_2$表示$f'_2$与$F'$的双错测度。$f'_2$加入到$F'$后，$F'=\{f'_1,f'_2\}$；当将$f'_3$加入到$F'$时，$f'_3$在除了$f'_1,f'_2$之外，其双错测度最小，则$f'_3$可以修正的分类错误样本数目为$(1-p'_1-\mathrm{df}_2-\mathrm{df}_3)\cdot N_t$，其中$\mathrm{df}_3$表示$f'_3$与$F'$的双错测度。$f'_3$加入到$F'$后，$F'=\{f'_1,f'_2,f'_3\}$；以此类推，当将$f'_i$加入到$F'$时，其可以修正的分类错误样本数目为$\left(1-p'_1-\sum_{j=2}^{i}\mathrm{df}_j\right)\cdot N_t$。

当$\left(1-p'_1-\sum_{j=2}^{i}\mathrm{df}_j\right)\cdot N_t>0$时，表示其可以修正$F'$中分类错误的样本；当$\left(1-p'_1-\sum_{j=2}^{i}\mathrm{df}_j\right)\cdot N_t<0$时，表示其不可以继续修正$F'$中分类错误的样本。因此，当$\left(1-p'_1-\sum_{j=2}^{i}\mathrm{df}_j\right)\cdot N_t>0\cap\left(1-p'_1-\sum_{j=2}^{i+1}\mathrm{df}_j\right)\cdot N_t<0$时，表示在加入$f'_i$后，再继续加入$f'_{i+1}$，不能进一步修正分类错误的样本。故在加入$f'_i$后，则停止加入。故关于保留双错测度较小基ELM数目M^*的理论计算方式为

$$M^*=\left\{i\left|\left(1-p'_1-\sum_{j=2}^{i}\mathrm{df}_j\right)>0\cap\left(1-p'_1-\sum_{j=2}^{i+1}\mathrm{df}_j\right)<0\right.\right\}\tag{10}$$

综上所述，所选择的基ELM集合$F'=\{f'_1,f'_2,\cdots,f'_{M^*}\}$，集成前$M^*$个双错测度较小的基ELM，其集成精度最优。 证毕

4 实验分析

为了检验DFSEE的一般性和显著性，在10个标准UCI数据集上进行测试。具体UCI数据集详见表2。使用5-折交叉验证技术[19]将UCI数据集随机划分为5份，选取4份作为训练集，剩下1份作为测试集。

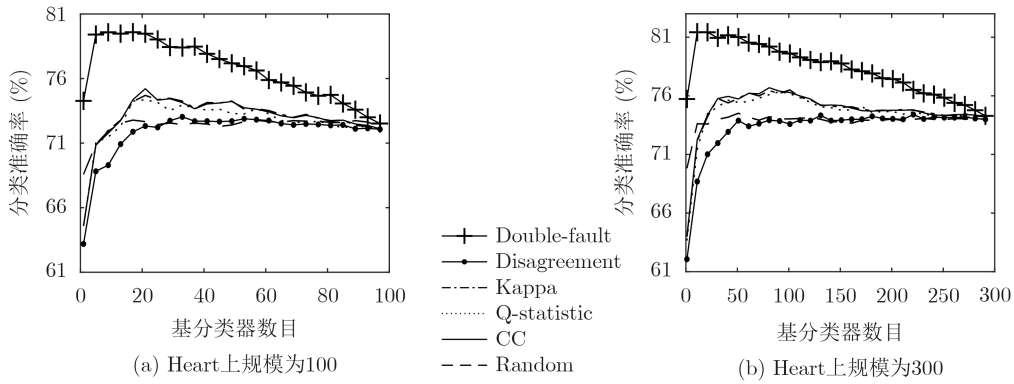
4.1 结果分析

本文实验部分所涉及的编码工作均采用Matlab R2017a软件编写，PC机的参数为: 64位Windows 10操作系统，Intel(R) Core(TM) i7-9700K 3.60 GHz CPU, 16.00 GB内存。为了降低随机性对试验结果的影响，在同一个UCI数据集上的集成结果均为独立重复30次试验结果的平均值。

图1展现了在Heart数据集上，不同规模基ELM (规模分别为100, 300)下，根据5种成对差异性测度大小进行排序集成，其集成分类准确率的变化趋势。将图1中的双错测度、不一致测度、Kappa

表 2 UCI数据集

数据集	实例个数	属性个数	类别
Heart	270	13	2
Cleveland	303	13	5
Bupa	345	6	2
Wholesale	440	7	2
Diabetes	768	8	2
German	1000	20	2
QSAR	1055	41	2
CMC	1473	9	3
Spambase	4601	57	2
Wineq-w	4898	11	7



(a) Heart上规模为100　　(b) Heart上规模为300

图 1 在不同基ELM集成规模下基于成对差异性测度排序集成分类准确率的趋势

测度、Q-统计量、相关系数和随机Bagging分别采用“Double-fault”，“Disagreement”，“Kappa”，“Q-statistic”，“CC”，“Random”表示。由图1可知，基于5种差异性测度的排序集成分类准确率的变化趋势，先逐渐上升到达某一值后，再逐渐下降。集成分类准确率在集成初期逐渐上升的原因是，由于在排序集成刚开始时，子集成中包含的基ELM规模较小，基ELM间差异性较小，直接导致集成精度不高。随着不断加入新的综合性能较优的基ELM，增加了基ELM的多样性，故其集成精度逐渐上升。在集成精度到达某一值后，集成精度逐渐下降，是由于此时继续加入基ELM后，会导致集成中出现综合性能较差的基ELM，使得集成精度逐渐下降。由图1可知，基于双错测度的排序集成分类准确率明显高于其余4种成对差异性测度和随机Bagging。基于上述分析，本文采用双错测度作为评估基ELM差异性的标准。

表3表明了DFSEE在不同规模基ELM (100, 200, 300)下的集成结果。其中“最高”，“平均”，“最低”分别表示基ELM的最高、平均和最低的分类准确率。从表3看出，DFSEE在10个UCI数据集上，集成分类准确率明显高于基ELM中的最高和平均分类准确率。表4展示了在不同UCI数据上不同基ELM规模下，DFSEE与Bagging在集成性能和集成规模上的对比分析，其中Bagging的集成规模分别为100, 200, 300; “n”表示DFSEE所选择的基ELM规模。由表4可以看出，DFSEE所选择的基ELM数目较少，但其集成分类准确率明显优于Bagging集成，同时表明了集成部分基ELM比集成全部的性能更优。

4.2 与其他选择性集成方法对比

为了检验DFSEE的集成分类性能，下面与AGOB[10], POBE[11], MOAG[12], EP-FP[20], SCG-P[8]进行对比分析。通过表4可以看出，DFSEE在基ELM规模达到200之后，总体上集成精度增加的幅度较小，换言之，继续增加基ELM规模，并不能显著提高集成分类器的性能，且会大幅增加计算复杂度。基于此，基ELM规模宜设置为200。由表5可知，当规模为200时，DFSEE在10个UCI数据集上，集成分类准确率均高于其他方法。综上所述，本文提出的DFSEE选择了较小规模的基ELM，取得了更高的集成分类准确率。

表 3　在不同规模基ELM (100, 200, 300)下的集成分类准确率(%)

数据集	100				200				300			
	DFSEE	最高	平均	最低	DFSEE	最高	平均	最低	DFSEE	最高	平均	最低
Heart	80.48	75.00	63.60	51.14	82.48	76.33	63.63	49.24	82.24	76.57	63.68	48.57
Cleveland	57.01	55.87	48.66	38.86	57.61	56.62	48.68	38.11	57.91	56.82	48.68	37.71
Bupa	75.37	70.67	60.21	48.18	76.95	71.51	60.15	47.09	77.40	72.18	60.23	46.49
Wholesale	94.04	89.56	82.78	74.19	94.78	90.26	82.78	73.48	95.15	90.59	82.73	73.04
Diabetes	71.88	70.62	61.83	52.64	73.10	71.49	61.73	51.03	73.77	71.81	61.74	50.65
German	77.40	75.17	69.61	63.63	78.08	76.12	69.63	62.83	78.58	76.40	69.64	62.62
QSAR	86.26	82.67	74.45	65.63	87.78	83.63	74.52	64.92	88.28	83.89	74.49	63.98
CMC	62.99	60.45	54.21	46.57	63.41	61.03	54.25	45.92	63.88	61.33	54.23	45.46
Spambase	80.78	77.57	70.13	63.32	81.55	78.17	70.12	62.79	81.70	78.42	70.13	62.53
Wineq-w	51.38	50.80	46.97	44.52	51.73	51.03	46.94	44.21	51.90	51.20	46.94	44.07

在基ELM规模为200的条件下，本文将DFSEE与其他方法在运行时间方面进行对比分析，如表6所示。由表6可知，本文提出的DFSEE在10个UCI数据集上的运行时间，明显低于其他方法，除了略高于POBE。DFSEE的运行时间略高于POBE的原因是：DFSEE在选择性集成时，需要不断地计算基ELM的双错测度，以进行相应的判断。而POBE直接采用策略选择部分基ELM，不需要计算基ELM的差异性测度并加以判断。但是DFSEE的分类性能明显优于POBE; DFSEE的运行时间明显低于其他方法的主要原因是：其他方法在进行选择性集成时，需要进行反复迭代搜索，以找到最优子集成，此过程需要消耗了大量的计算时间。通过上述分析表明，本文提出的DFSEE的运行效

表 4 在不同规模基ELM (100, 200, 300)下DFSEE与Bagging分类准确率对比分析(%)

数据集	100			200			300		
	Bagging	本文DFSEE	n	Bagging	本文DFSEE	n	Bagging	本文DFSEE	n
Heart	72.10	80.48	13	71.67	82.48	11	71.71	82.24	11
Cleveland	49.25	57.01	4	49.25	57.61	6	49.25	57.91	6
Bupa	65.61	75.37	12	64.14	76.95	12	64.98	77.40	15
Wholesale	86.44	94.04	8	86.00	94.78	10	86.11	95.15	11
Diabetes	63.79	71.88	7	63.51	73.10	7	63.63	73.77	8
German	74.13	77.40	14	74.40	78.08	9	74.38	78.58	9
QSAR	80.22	86.26	9	80.41	87.78	8	80.47	88.28	9
CMC	58.22	62.99	9	58.44	63.41	12	58.45	63.88	13
Spambase	73.34	80.78	12	73.37	81.55	13	73.46	81.70	11
Wineq-w	46.58	51.38	8	46.57	51.73	11	46.56	51.90	14

表 5 与其他方法在集成精度(%)和集成规模方面对比分析(基ELM规模200)

数据集	本文DFSEE	n	AGOB	n	POBE	n	MOAG	n	EP-FP	n	SCG-P	n
Heart	82.48	11	74.14	49	77.52	96	74.86	43	74.38	95	75.24	38
Cleveland	57.61	6	54.43	22	51.09	132	50.85	25	49.25	95	56.25	1
Bupa	76.95	12	69.93	37	72.95	99	69.47	59	65.89	66	76.89	48
Wholesale	94.78	10	89.67	36	92.74	99	88.59	27	86.11	96	87.85	9
Diabetes	73.10	7	66.03	26	68.99	102	66.27	52	63.73	89	65.30	58
German	78.08	9	75.15	36	76.60	96	75.30	38	74.47	86	75.18	54
QSAR	87.78	8	83.48	24	84.37	100	83.94	32	80.43	88	82.02	37
CMC	63.41	12	59.63	47	60.92	103	59.67	51	58.46	97	59.51	67
Spambase	81.55	13	76.18	32	79.12	97	76.47	67	76.64	93	76.66	58
Wineq-w	51.73	11	50.37	23	49.48	93	48.10	34	48.60	96	50.98	46

表 6 与其他方法在运行时间方面的对比分析(s)

数据集	本文DFSEE	AGOB	POBE	MOAG	EP-FP	SCG-P
Heart	0.80	10.96	0.79	0.87	18.24	0.86
Cleveland	0.77	22.36	0.73	1.10	2.77	1.10
Bupa	0.86	13.97	0.85	0.95	41.26	0.95
Wholesale	1.16	17.26	1.15	1.27	21.97	1.26
Diabetes	1.29	17.95	1.28	1.40	30.40	1.39
German	1.79	12.58	1.78	1.86	11.58	1.86
QSAR	2.29	14.01	2.29	2.37	23.55	2.37
CMC	2.25	24.44	2.21	2.62	30.85	2.61
Spambase	8.54	43.86	8.52	8.80	110.46	8.78
Wineq-w	7.71	79.18	7.58	8.85	48.56	8.82

率除了略低于POBE外，明显优于其他方法，并且DFSEE集成分类性能更优。因此，DFSEE在分类性能和运行效率方面均具有较突出的优势。

5　结束语

针对ELM分类性能不稳定，而集成学习计算开销较大的问题，本文提出了DFSEE，采用双错测度进行选择性集成，并给出了剔除集成系统中冗余基ELM的标准，大幅度约简了集成系统中综合性能较差的基ELM，显著提高了集成分类性能，并从理论上详细阐述了双错测度的有效性。在10个UCI数据集上的实验结果表明，与其他方法相比较，DFSEE所取得子集成中的基ELM数目更少，最终的集成分类准确率更优。进一步表明了集成部分基ELM远比集成全部基ELM的效果更好，为提高ELM的分类性能提供了新的研究思路。下一个阶段的研究工作，将尝试运用基于多种差异性测度的融合型测度进行选择性集成，进一步提升ELM的分类性能。

参 考 文 献

[1] HUANG Guangbin, ZHU Qinyu, and SIEW C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1/3): 489–501. doi: 10.1016/j.neucom.2005.12.126.

[2] YANG Yifan, ZHANG Hong, YUAN D, *et al*. Hierarchical extreme learning machine based image denoising network for visual Internet of Things[J]. *Applied Soft Computing*, 2019, 74: 747–759. doi: 10.1016/j.asoc.2018.08.046.

[3] 吴超, 李雅倩, 张亚茹, 等. 用于表示级特征融合与分类的相关熵融合极限学习机[J]. 电子与信息学报, 2020, 42(2): 386–393. doi: 10.11999/JEIT190186.

WU Chao, LI Yaqian, ZHANG Yaru, *et al*. Correntropy-based fusion extreme learning machine for representation level feature fusion and classification[J]. *Journal of Electronics & Information Technology*, 2020, 42(2): 386–393. doi: 10.11999/JEIT190186.

[4] 陆慧娟, 安春霖, 马小平, 等. 基于输出不一致测度的极限学习机集成的基因表达数据分类[J]. 计算机学报, 2013, 36(2): 341–348. doi: 10.3724/SP.J.1016.2013.00341.

LU Huijuan, AN Chunlin, MA Xiaoping, *et al*. Disagreement measure based ensemble of extreme learning machine for gene expression data classification[J]. *Chinese Journal of Computers*, 2013, 36(2): 341–348. doi: 10.3724/SP.J.1016.2013.00341.

[5] LAN Y, SOH Y C, and HUANG Guangbin. Ensemble of online sequential extreme learning machine[J]. *Neurocomputing*, 2009, 72(13/15): 3391–3395. doi: 10.1016/j.neucom.2009.02.013.

[6] KSIENIEWICZ P, KRAWCZYK B, and WOŹNIAK M M. Ensemble of Extreme Learning Machines with trained classifier combination and statistical features for hyperspectral data[J]. *Neurocomputing*, 2018, 271: 28–37. doi: 10.1016/j.neucom.2016.04.076.

[7] 李炜, 李全龙, 刘政怡. 基于加权的K近邻线性混合显著性目标检测[J]. 电子与信息学报, 2019, 41(10): 2442–2449. doi: 10.11999/JEIT190093.

LI Wei, LI Quanlong, and LIU Zhengyi. Salient object detection using weighted K-nearest neighbor linear blending[J]. *Journal of Electronics & Information Technology*, 2019, 41(10): 2442–2449. doi: 10.11999/JEIT190093.

[8] YKHLEF H and BOUCHAFFRA D. An efficient ensemble pruning approach based on simple coalitional games[J]. *Information Fusion*, 2017, 34: 28–42. doi: 10.1016/j.inffus.2016.06.003.

[9] CAO Jingjing, LI Wenfeng, MA Congcong, *et al*. Optimizing multi-sensor deployment via ensemble pruning for wearable activity recognition[J]. *Information Fusion*, 2018, 41: 68–79. doi: 10.1016/j.inffus.2017.08.002.

[10] MARTÍNEZ-MUÑOZ G, HERNÁNDEZ-LOBATO D, and SUÁREZ A. An analysis of ensemble pruning techniques based on ordered aggregation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 245–259. doi: 10.1109/TPAMI.2008.78.

[11] MARTÍNEZ-MUÑOZ G and SUÁREZ A. Pruning in ordered bagging ensembles[C]. The 23rd International Conference on Machine learning, New York, USA, 2006: 609-616. doi: 10.1145/1143844.1143921.

[12] GUO Li and BOUKIR S. Margin-based ordered aggregation for ensemble pruning[J]. *Pattern Recognition Letters*, 2013, 34(6): 603–609. doi: 10.1016/j.patrec.2013.01.003.

[13] DAI Qun, ZHANG Ting, and LIU Ningzhong. A new reverse reduce-error ensemble pruning algorithm[J]. *Applied Soft Computing*, 2015, 28: 237–249. doi: 10.1016/j.asoc.2014.10.045.

[14] ZHOU Zhihua, WU Jianxin, and TANG Wei. Ensembling neural networks: Many could be better than all[J]. *Artificial Intelligence*, 2002, 137(1/2): 239–263. doi: 10.1016/S0004-3702(02)00190-X.

[15] CAVALCANTI G D C, OLIVEIRA L S, MOURA T J M, *et al*. Combining diversity measures for ensemble pruning[J]. *Pattern Recognition Letters*, 2016, 74: 38–45. doi: 10.1016/j.patrec.2016.01.029.

[16] MAO Shasha, CHEN Jiawei, JIAO Licheng, *et al*. Maximizing diversity by transformed ensemble learning[J]. *Applied Soft Computing*, 2019, 82: 105580. doi: 10.1016/

j.asoc.2019.105580.

[17] TANG E K, SUGANTHAN P N, and YAO Xin. An analysis of diversity measures[J]. *Machine Learning*, 2006, 65(1): 247–271. doi: 10.1007/s10994-006-9449-2.

[18] GIACINTO G and ROLI F. Design of effective neural network ensembles for image classification purposes[J]. *Image and Vision Computing*, 2001, 19(9/10): 699–707. doi: 10.1016/S0262-8856(01)00045-2.

[19] FUSHIKI T. Estimation of prediction error by using *K*-fold cross-validation[J]. *Statistics and Computing*, 2011, 21(2): 137–146. doi: 10.1007/s11222-009-9153-8.

[20] ZHOU Hongfa, ZHAO Xuehan, and WANG Xiao. An effective ensemble pruning algorithm based on frequent patterns[J]. *Knowledge-Based Systems*, 2014, 56: 79–85. doi: 10.1016/j.knosys.2013.10.024.

夏平凡：女，1994年生，博士生，研究方向为机器学习、人工智能和集成学习等.

倪志伟：男，1963年生，教授，博士生导师，研究方向为人工智能、机器学习和云计算.

朱旭辉：男，1991年生，讲师，硕士生导师，研究方向为进化计算和机器学习.

倪丽萍：女，1981年生，副教授，硕士生导师，研究方向为分形数据挖掘、人工智能和机器学习.

责任编辑：马秀强