面向类别不平衡数据的主动在线加权极限学习机算法

王长宝 李青雯 于化龙
(江苏科技大学计算机科学与工程学院 镇江 212003)

摘 要 针对在样本类别分布不平衡场景下,现有的主动学习算法普遍失效及训练时间过长等问题,提出采用建模速度更快的极限学习机,即ELM(Extreme Learning Machine)作为主动学习的基分类器,并以加权ELM算法用于主动学习过程的平衡控制,进而在理论上推导了其在线学习的过程,大幅降低了主动学习的时间开销,并将最终的混合算法命名为AOW-ELM算法。通过12个基准的二类不平衡数据集验证了该算法的有效性与可行性。

关键词 主动学习,类别不平衡学习,极限学习机,加权极限学习机,在线学习

中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.12.040

Active, Online and Weighted Extreme Learning Machine Algorithm for Class Imbalance Data

WANG Chang-bao LI Qing-wen YU Hua-long
(School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

Abstract It is well known that most existing active learning algorithms often fail to provide excellent performance and cost much training time when they are used in the scenario of class imbalance. To deal with this problem, a hybrid active learning algorithm named AOW-ELM algorithm was proposed. The algorithm uses ELM (extreme learning machine) which has rapid modeling speed as base classifier in active learning. In addition, weighted ELM algorithm is adopted to guarantee the impartiality in the procedure of active learning. Next, to further accelerate the process of active learning, i.e., decreasing the time consumption of active learning, online learning procedure of weighted ELM algorithm was deduced in theory. Experimental results on 12 baseline binary-class imbalanced data sets indicate the effectiveness and feasibility of the proposed algorithm.

Keywords Active learning, Class imbalance learning, Extreme learning machine, Weighted extreme learning machine, Online learning

1 引言

主动学习(Active Learning, AL)是一种标准的机器学习范式,其主要被用于可轻易收集大量样本,但对这些样本进行类别标注却要消耗大量人力、物力以及财力的场景。举例来说,若要训练并开发一个基于网页文本内容的主题分类系统,可通过网络爬虫从Web上随意下载海量的网页,但若希望标注这些网页的主题类别以用于训练,则需要人类专家通过阅读网页内容的方式来完成,显然这是相当损时耗力的。为解决上述问题,不妨转换一下思路:在海量样本中,通常只有很少一部分样本与分类面的位置密切相关,若能将这些样本预先挑选出来并进行标注,则有望在不牺牲分类模型质量的前提下,大幅降低人力与时间成本的开销[1-2]。

主动学习为上述问题提供了解决方案,它采用迭代的方式来选取样本,从而可保证每一轮所选出的都是在当前信息量或不确定程度最大的样本,通过手动的方式为这些样本添加类别标注,并用其来更新分类模型,更新后的分类模型质量将获得很大的提升概率。重复上述过程,只需数轮便可得到一个高质量的分类模型。因此主动学习将有助于大幅降低训练样本的复杂度以及人力和时间成本的开销,同时得到高质量的分类模型。近年来,研究人员已从不同角度提出了多种主动学习算法[1],这些算法也已在广泛的应用领域中得到实际应用,如文本分类[3]、网络入侵检测[4]、语音识别[5]、图像分类[6]、视频分析[7]以及生物信息学等[8]。

毫无疑问,当所收集的样本在类别标记上是近似平衡时,主动学习能最大限度地提升分类模型的质量。换言之,若未标注样本的类别分布是不平衡的,主动学习算法则很可能会失效,进而生成一个质量较差的分类模型[9]。究其原因,不难发现:在主动学习算法的运行过程中,它所主动查询的样本几乎均位于分类边界附近,即不同类样本的交叠区域。对于类别不平衡数据而言,在上述区域,多数类样本通常会远多于少数类样本,这将导致更多的查询样本来自于多数类,进而将不

到稿日期:2016-12-01 返修日期:2017-03-16 本文受国家自然科学基金(61305058),江苏省自然科学基金(BK20130471),中国博士后特别资助计划项目(2015T80481),中国博士后科学基金(2013M540404),江苏省博士后基金(1401037B)资助。

王长宝(1963-),男,实验师,主要研究方向为嵌入式系统开发、模式识别、机器学习;李青雯(1992-),女,硕士生,主要研究方向为机器学习;于化龙(1982-),男,博士,副教授,主要研究方向为机器学习、数据挖掘,E-mail:yuhualong@just.edu.cn(通信作者)。

可避免地侵占少数类的利益，从而因类别不平衡问题导致最终所训练的分类模型失效。

针对上述问题，Zhu 和 Hovy[9] 进行了深入研究，并提出应在主动学习过程中引入平衡控制策略，以保证在主动学习迭代过程中每一轮所生成的分类模型都是公平公正的。他们考虑采用类别不平衡学习领域中最为常用的样本采样技术来实现平衡控制，即在每一轮分类模型更新前，均通过增加少数类样本或减少多数类样本的方式来保持已标注训练样本集的绝对平衡。结果表明，结合了此类技术的主动学习算法要优于未采用平衡控制策略的传统算法，但在不平衡比率较高的数据集上，降采样算法的性能通常有很大波动，缺乏稳定性，而过采样算法又会极大地降低分类器的建模速度，增加不必要的时间开销，特别是在大规模数据集上，这种现象体现得尤为明显。

为解决上述问题，本文采用极限学习机(Extreme Learning Machine，ELM)作为基分类器来执行主动学习的过程。众所周知，相比于诸多传统的分类算法，ELM 具有训练速度快、泛化能力强等两大优点[10-11]，能在至少不损失分类性能的前提下，极大地减少主动学习的时间开销[12]，因此采用加权 ELM 算法作为主动学习过程中的平衡控制策略。实际上，加权极限学习机(Weighted Extreme Learning Machine，WELM)是极限学习机在代价敏感场景下的一种改进算法，可以有效解决类别不平衡问题[13]。此外，为进一步减小主动学习的时间开销，在理论上推导了上述算法的在线学习模式并给出了相应的更新算法，从而保证在主动学习每轮迭代时，无需重新训练分类模型，而只需利用新标注样本对其进行微调即可。将上述的混合算法命名为 AOW-ELM 算法，即主动(Active)、在线(Online)、加权(Weighted)极限学习机算法。最后，在 12 个基准的二类不平衡数据集上对该算法与 5 种具有代表性的主动学习算法进行了实验比较，结果表明了该算法的有效性、可行性与优越性。

2 方法

2.1 主动学习

如前文所述，主动学习即通过主动获取样本的方式来进行学习，其主要目的是在不损害分类模型质量的前提下，尽可能地降低训练样本的复杂度。根据应用场景的不同，主动学习大致可以分为以下两类：基于流的主动学习模式[14]和基于池的主动学习模式[15]。本文侧重于基于池的主动学习模式，图 1 给出了在这一模式下的主动学习过程。

图 1 基于池的主动学习模式过程示意图

由图 1 不难看出在此类模式下，主动学习系统主要由以下 5 个基本构件组成：1)一个已标记样例集 L；2)一个未标记样例集 U；3)一个分类模型 S；4)一个查询算法 q；5)一个人工标注者 T。主动学习的运行过程如下。首先，人工标注者 T 仅随机地标注少量的未标注样本，并将其置于已标记样例集 L 中，进而训练一个初始的分类模型 S；然后，进入循环迭代过程：利用分类模型 S 评价未标记样例池 U 中的每一个样例，并利用查询算法 q 提取各样例的信息量，进而根据信息量的大小对其进行排序，选取一个或一批信息量最大的样例提交给人工标注者 T 进行标注。最后，再将标注的样例添加到已标记样例集 L 中，对分类模型 S 进行更新。上述过程循环往复，直至达到某个预设的停止条件[16]为止。

2.2 极限学习机

极限学习机(Extreme Learning Machine，ELM)由南洋理工大学的 Huang 等人[10]于 2006 年正式提出，经过近十年的发展，已经成为机器学习领域的研究热点之一。不同于传统的误差反传(Back-Propagation，BP)算法，极限学习机通过随机指定隐层参数，并利用最小二乘法求解输出层权重的方式来训练单隐层前馈神经网络(Single hidden-Layer Feedback Network，SLFN)，因此其具有泛化能力强、训练速度快等优点[10-11]。下面对 ELM 的工作机理做简要介绍。

不妨设训练集包括 N 个训练样本，表示为 $(x_i,t_i)\in R^n\times R^m$，其中，$x_i$ 表示 $n\times 1$ 维的输入向量，t_i 表示第 i 个训练样本的期望输出向量，对于分类问题而言，n 代表训练样本的属性数，m 代表样本的类别数。若一个具有 L 个隐层节点的单隐层前馈神经网络能以零误差拟合上述 N 个训练样本，则意味着存在 β_i，a_i 及 b_i，使得下式成立：

$$f_L(x_j)=\sum_{i=1}^{L}\beta_i G(a_i,b_i,x_j)=t_j,\ j=1,\cdots,N \tag{1}$$

其中，a_i 和 b_i 分别表示第 i 个隐层节点的权重与偏置，β_i 表示第 i 个隐层节点的输出权重，即第 i 个隐层节点到各输出节点的连接权重，而 G 则表示激活函数，式(1)可进一步简化为下式：

$$H\beta=T \tag{2}$$

其中，

$$H(a_1,\cdots,a_L,b_1,\cdots,b_L,x_1,\cdots,x_N)=\begin{bmatrix}G(a_1,b_1,x_1) & \cdots & G(a_L,b_L,x_1)\\ \vdots & \cdots & \vdots\\ G(a_1,b_1,x_N) & \cdots & G(a_L,b_L,x_N)\end{bmatrix} \tag{3}$$

$$\beta=\begin{bmatrix}\beta_1^{\mathrm{T}}\\ \vdots\\ \beta_L^{\mathrm{T}}\end{bmatrix}_{L\times m},\quad T=\begin{bmatrix}t_1^{\mathrm{T}}\\ \vdots\\ t_N^{\mathrm{T}}\end{bmatrix}_{N\times m} \tag{4}$$

其中，$G(a_i,b_i,x_j)$ 表示第 j 个训练样本在第 i 个隐层节点上的激活函数值；T 表示所有训练样本对应的期望输出矩阵，通常将每个样本所对应类别的输出节点的期望输出值设为 1，其他节点的输出值则设为 -1；H 被称为隐层输出矩阵，其第 i 列为第 i 个隐层节点在所有训练样本上的输出向量，第 j 行为第 j 个训练样本在整个隐藏层中对应的输出向量。

在极限学习机中，由于所有的 a_i 和 b_i 均是在 $[-1,1]$ 区间内随机生成的，因此输入样本、隐层权重与偏置、期望输出

(类别标记)均已知，则输出权重矩阵 β 的近似解$\hat{\beta}$可由下式直接计算得到：

$$\hat{\beta}=H^{\dagger}T \tag{5}$$

其中，$H^{\dagger}$ 为隐层输出矩阵的 Moore-Penrose 广义逆。根据其定义可推知，$\hat{\beta}$为该网络的最小范数最小二乘解。因此，极限学习机可通过一步计算得到，无需迭代训练，保证了神经网络的训练时间可被大幅缩减；同时，由于在求解过程中约束了输出权重矩阵 β 的 l_2 范数，使其最小化，因此可保证网络具有较强的泛化性能。

2012 年，极限学习机的优化版本[11]被提出，类似于支持向量机，其优化式可表示如下：

$$\begin{aligned}&\text{Min:}Lp_{ELM}=\frac{1}{2}\|\beta\|^2+C\frac{1}{2}\sum_{i=1}^{N}\varepsilon_i^2\\&\text{Subject to:}h(x_i)\beta=t_i-\varepsilon_i\end{aligned} \tag{6}$$

其中，ε_i 表示第 i 个训练样本的实际输出与期望输出之差；$h(x_i)$为第 i 个样例 x_i 在隐层上的输出向量；C 为惩罚因子，用于调控网络的泛化性与精确性之间的平衡关系。上述优化式可通过求解得到，给定一个具体的样例 x，其对应的实际输出向量可由下式求得：

$$f(x)=\begin{cases}h(x)H^{\mathrm{T}}(\dfrac{I}{C}+HH^{\mathrm{T}})^{-1}T, & \text{when } N<L\\ h(x)(\dfrac{I}{C}+H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}T, & \text{when } N\geqslant L\end{cases} \tag{7}$$

其中，I 表示单位矩阵；$f(x)=[f_1(x),\cdots,f_m(x)]$表示样例 x 的实际输出向量，可进一步通过下式确定该样例的预测类标：

$$label(x)=\arg\max_i f_i(x), i\in\{1,\cdots,m\} \tag{8}$$

2.3　主动极限学习机

Yu 等人[16]于 2015 年在考查了样本在 ELM 中的实际输出值与其至分类面之间距离的对应关系的基础上，提出了一种基于 ELM 分类器的主动学习算法，并将其命名为 AL-ELM 算法。该研究发现：实际上，ELM 的实际输出值与朴素贝叶斯的后验概率之间存在近似等价关系。图 2 给出了 ELM 分类器在一个人工二类数据集上的输出值等高线分布。从图 2 可以看出，样本通过 ELM 分类器得到的输出值越小，表明其距离当前的决策面越近，反之亦然。因此，在主动学习过程中，可采用当前 ELM 对所有未标记样本的决策输出值进行计算，并进行升序排列，进而选取排序靠前的样本，即当前最不确定的样本进行标注。因此，不难利用 ELM 作为基分类器来实现主动学习。

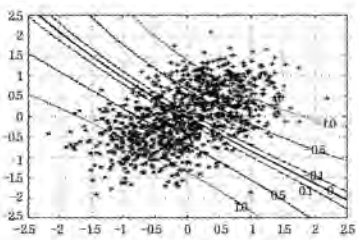

图 2　ELM 在人工数据集上的输出值等高线分布

2.4　加权极限学习机

Zong 等人[13]于 2013 年借鉴了代价敏感学习的思想，提出了一种加权极限学习机算法，即 WELM(Weighted Extreme Learning Machine)算法，用于解决类别不平衡问题。WELM 算法也是通过为不同类的样例赋予不同权重，从而改变惩罚因子的方式来有效降低少数类样例被错分的概率。在 WELM 算法中，式(6)被改写为如下形式：

$$\begin{aligned}&\min:LD_{ELM}=\frac{1}{2}\|\beta\|^2+C\frac{1}{2}\sum_{i=1}^{N}(W_{ii}\times\|\varepsilon_i\|^2)\\&\text{subject to:}h(x_i)\beta=t_i-\varepsilon_i\end{aligned} \tag{9}$$

其中，W 为一个 $N\times N$ 的对角阵，W_{ii} 代表第 i 个训练样例所对应的权重，若对少数类样例施以比多数类样例更大的权重，则会增大对其训练误差的惩罚力度，从而相应降低其被误分的概率。文献[13]提供了如下两种权重分配方式：

$$\text{WELM1:}W_{ii}=1/\#(t_i) \tag{10}$$

及

$$\text{WELM2:}W_{ii}=\begin{cases}0.618/\#(t_i), & \text{if } \#(t_i)>AVG\\ 1/\#(t_i), & \text{if } \#(t_i)\leqslant AVG\end{cases} \tag{11}$$

其中，$\#(t_i)$和 AVG 分别代表 t_i 类的训练样例数及所有类的平均训练样例数。显然，无论对于上述哪种权重分配方式，少数类样例都会被赋予更大的权重，且类不平衡比率越高，不同类样例间的权重就越高。求解式(9)，可得：

$$\beta=\begin{cases}H^{\mathrm{T}}(\dfrac{I}{C}+WHH^{\mathrm{T}})^{-1}WT, & \text{when } N<L\\ (\dfrac{I}{C}+H^{\mathrm{T}}WH)^{-1}WH^{\mathrm{T}}T, & \text{when } N\geqslant L\end{cases} \tag{12}$$

2.5　在线加权极限学习机

在主动学习过程中，新标注样本显然是以增量的方式进行添加的，因此可采用以下两种方式之一对分类器进行迭代更新：1)首先将新标记样本添加到已标记样例集中，然后采用全部已标记样本训练一个新分类器；2)采用增量学习算法，即仅利用新的已标记样本对分类器进行微调和更新。显然，在上述两种方式中，后者比前者更省时，尤其对于大规模数据集而言更是如此。文献[17]曾给出 ELM 的在线学习版本，但其主要是针对文献[10]中提出的早期 ELM 算法，对于 WELM 算法并不适用。因此，本文对 WELM 算法进行了理论分析，并提出了与之对应的在线学习算法。

根据 WELM 算法，显然有：

$$\beta=(H^{\mathrm{T}}WH+\frac{I}{C})^{-1}H^{\mathrm{T}}WT \tag{13}$$

对其进行变换，可得到等价式：

$$\beta=((\sqrt{W}H)^{\mathrm{T}}\sqrt{W}H+\frac{I}{C})^{-1}(\sqrt{W}H)^{\mathrm{T}}(\sqrt{W}T) \tag{14}$$

若令 $A=\sqrt{W}H$，$B=\sqrt{W}T$，则上式可表示为：

$$\beta=(A^{\mathrm{T}}A+\frac{I}{C})^{-1}A^{\mathrm{T}}B \tag{15}$$

对于增量式，可表示为：

$$\beta_{K+\Delta K}=(A_{K+\Delta K}^{\mathrm{T}}A_{K+\Delta K}+\frac{I}{C})^{-1}A_{K+\Delta K}^{\mathrm{T}}B_{K+\Delta K} \tag{16}$$

其中，$A_{K+\Delta K}^{\mathrm{T}}A_{K+\Delta K}$及 $A_{K+\Delta K}^{\mathrm{T}}B_{K+\Delta K}$ 分别可以拆解为如下形式：

$$\begin{aligned}&A_{K+\Delta K}^{\mathrm{T}}A_{K+\Delta K}=A_K^{\mathrm{T}}A_K+A_{\Delta K}^{\mathrm{T}}A_{\Delta K}\\&A_{K+\Delta K}^{\mathrm{T}}B_{K+\Delta K}=A_K^{\mathrm{T}}B_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}\end{aligned} \tag{17}$$

进一步，增量式可表示为：

$$\beta_{K+\Delta K}=(A_K^{\mathrm{T}}A_K+A_{\Delta K}^{\mathrm{T}}A_{\Delta K}+\frac{I}{C})^{-1}(A_K^{\mathrm{T}}B_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}) \tag{18}$$

当 $K=0$ 时，即尚未进行增量学习时，

$$\beta_0=(A_0^{\mathrm{T}}A_0+\frac{I}{C})^{-1}A_0^{\mathrm{T}}B_0 \tag{19}$$

令 $M_0=A_0^{\mathrm{T}}A_0+\frac{I}{C}$，则：

$$\beta_{0+\Delta 1}=(A_{\Delta 1}^{\mathrm{T}}A_{\Delta 1}+M_0)^{-1}(A_0^{\mathrm{T}}B_0+A_{\Delta 1}^{\mathrm{T}}B_{\Delta 1}) \tag{20}$$

即 $M_1=A_{\Delta 1}^{\mathrm{T}}A_{\Delta 1}+M_0$，以此类推，可得：

$$M_{K+\Delta K}=A_{\Delta K}^{\mathrm{T}}A_{\Delta K}+M_K=A_{\Delta K}^{\mathrm{T}}A_{\Delta K}+A_K^{\mathrm{T}}A_K+\frac{I}{C} \tag{21}$$

进而，有：

$$\beta_{K+\Delta K}=(M_K+A_{\Delta K}^{\mathrm{T}}A_{\Delta K})^{-1}(A_K^{\mathrm{T}}B_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}) \tag{22}$$

接下来，考虑第二项：

$$\begin{aligned}A_K^{\mathrm{T}}B_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}&=M_KM_K^{-1}A_K^{\mathrm{T}}B_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}\\&=M_K(A_K^{\mathrm{T}}A_K+\frac{I}{C})^{-1}A_K^{\mathrm{T}}B_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}\\&=M_K\beta_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}\end{aligned} \tag{23}$$

将其代入(22)，可得：

$$\beta_{K+\Delta K}=(M_K+A_{\Delta K}^{\mathrm{T}}A_{\Delta K})^{-1}(M_K\beta_K+A_{\Delta K}^{\mathrm{T}}B_{\Delta K}) \tag{24}$$

由于 M_K，$A_{\Delta K}$，$B_{\Delta K}$，β_K 均已知，因此可通过式(24)直接计算出 $\beta_{K+\Delta K}$ 的值，也即隐层输出矩阵的更新值，可利用其对 WELM 分类器进行在线更新。

在此，还要考虑一个重要的问题，即在在线学习过程中，新加入样本的权重设置问题。本文并未继承文献[13]中的权重设置方法，而是通过下式分别为少数类与多数类样本设置权重：

$$W_{ii}=\begin{cases}1-\dfrac{|N^+|}{|N^+|+|N^-|}, & \text{if } x_i \text{ 为少数类}\\ \dfrac{|N^+|}{|N^+|+|N^-|}, & \text{if } x_i \text{ 为多数类}\end{cases} \tag{25}$$

其中，N^+ 与 N^- 分别表示已标注的少数类与多数类的样本数。由式(25)不难看出，尽管本文与文献[13]的权重设置方式不同，但不同类样本的权重之比却是一致的，均等同于类别不平衡比率。需要特别指出的是：在主动学习过程中，会不断添加新的标注样本，因此 N^+ 与 N^- 的值也是在增量更新的。

2.6 算法流程

基于上述理论基础，可给出 AOW-ELM 算法的基本流程。

输入：初始有标记样本集 L、无标记样本集 U、主动学习的批处理样本规模 P

输出：加权极限学习机 S

1. 统计并保留 L 中少数类与多数类的样本数，分别记为$|N^+|$和$|N^-|$；
2. 利用式(25)计算初始的加权矩阵 W；
3. 训练初始的加权极限学习机分类器 S，并得到初始的隐层输出矩阵 β_0；
4. 删除有标记样本集 L；
5. 调用 AL-ELM 算法的查询样本规则在 U 中找到一个具有最大不确定度的无标记样本子集 V，其中 V 包括 P 个样本，并将该子集提交给人类专家进行手动标注；
6. U=U−V；
7. 统计 V 中少数类与多数类的样本数，并分别对已保留的$|N^+|$和$|N^-|$进行更新，同时利用式(25)计算出 V 中每个新添加样本的权重，构建新的权重矩阵 W；
8. 利用式(24)对隐层输出矩阵 β 进行更新并得到更新的极限学习机分类器 S；
9. 删除新标注样本集 V；
10. 判断是否达到预设的停止条件，若是，则退出，并输出最后的分类器 S；若否，则返回第 5 步，继续执行。

3 实验结果与讨论

3.1 数据集与实验设置

为验证本文所提出的 AOW-ELM 算法的有效性，采用 12 个基准的二类不平衡数据集对其性能进行测试。12 个数据集均采集于 Keel 数据库[18]，这些数据集的具体描述信息如表 1 所列。

表 1 本文所用数据集

数据集	特征数	样本数	不平衡比率
wisconsin	9	683	1.86
yeast3	8	1484	8.10
vehicle1	18	846	2.90
pima	8	768	1.87
abalone9_18	8	731	16.40
vowel0	13	988	9.98
yeast5	8	1484	32.73
abalone19	8	4174	129.44
led7digit_0_2_4_5_6_7_8_9_vs_1	7	443	10.97
ecoli_0_6_7_vs_5	6	220	10.00
winequality_red_4	11	1599	29.17
flare_F	11	1066	23.79

为模拟真实的主动学习环境，对于上述每个数据集，均随机抽取 15%的样本作为初始的已标注样本集 L，55%的样本作为初始的未标注样本集 U，采用剩余 30%的样本作为测试集。此外，考虑到样本抽取的随机性，在每个数据集上的实验均随机运行 20 次，结果以均值的形式给出。此外，考虑到在不平衡分类问题中，整体分类精度不能反映分类器的真实性能，因此采用 G-mean 测度作为分类性能的量度，其计算公式如下：

$$G\text{-}mean=\sqrt{Acc_+\times Acc_-} \tag{26}$$

其中，Acc_+ 与 Acc_- 分别代表少数类与多数类样本的分类精度，因此 G-mean 测度测试的是两类样本分类精度的平衡度。

为了验证 AOW-ELM 算法的优越性，将其与以下 5 种有代表性的主动学习算法进行了实验比较。

1)AL-ELM 算法[16]：即未考虑在主动学习过程中加入平衡控制策略，但学习的过程是增量的，通过文献[17]的 OS-ELM 算法实现。

2)ROW-ELM 算法：即同时考虑了在主动学习过程中采用平衡控制策略与增量学习，但“查询样本”的选取是随机的，没有考虑样本的不确定度。

3)AL-ELM-RUS 算法：即考虑在 AL-ELM 算法中加入随机降采样(Random Undersampling，RUS)算法作为平衡控制策略，但由于 RUS 算法需要在全部已标注样本上执行，因此该算法不是增量的。

4）AL-ELM-ROS 算法：即考虑在 AL-ELM 算法中加入随机过采样（Random Oversampling，ROS）算法作为平衡控制策略，但由于 ROS 算法需要在全部已标注样本上执行，因此该算法不是增量的。

5）AL-ELM-SMOTE 算法：即考虑在 AL-ELM 算法中加入 SMOTE（Synthetic Minority Oversampling Technique）算法[20]作为平衡控制策略，但由于 SMOTE 算法需要在全部已标注样本上执行，因此该算法不是增量的。此外，借鉴文献[19]，SMOTE 算法中的参数 K 取经验值 5。

上述算法各具代表性：AL-ELM 算法未考虑到在学习过程中引入平衡控制；ROW-ELM 算法未利用分类器的反馈信息来选取大信息量样本；AL-ELM-RUS，AL-ELM-ROS 及 AL-ELM-SMOTE 算法则是目前最为常用也是仅有的几种主动不平衡学习算法[9]。因此，通过与上述各类算法进行实验比较，可以有效测试本文算法在各方面的性能及有效性。

此外，为探究各算法在主动学习过程中的全貌，并未在实验中设置学习停止准则，而是令未标注样本耗尽为止。主动学习采用批处理标注模式，每轮标注 U 中 5%的样本，共标注 20 轮。

最后，为简化实验的过程并保证比较结果的公正性，对 ELM 分类器中的两个重要参数（隐层节点数及惩罚因子）以经验预设值的形式给出，分别预设为 100 和 10000。各比较算法中 ELM 分类器的参数设置均保持一致。

3.2　结果与讨论

图 3 给出了 6 种主动学习算法在 12 个数据集上的学习曲线。表 2 列出了这些学习曲线所对应的 ALC（Area under Learning Curve）[20]测度值。显然，该测度值越大表明对应的学习算法具有更好的性能。

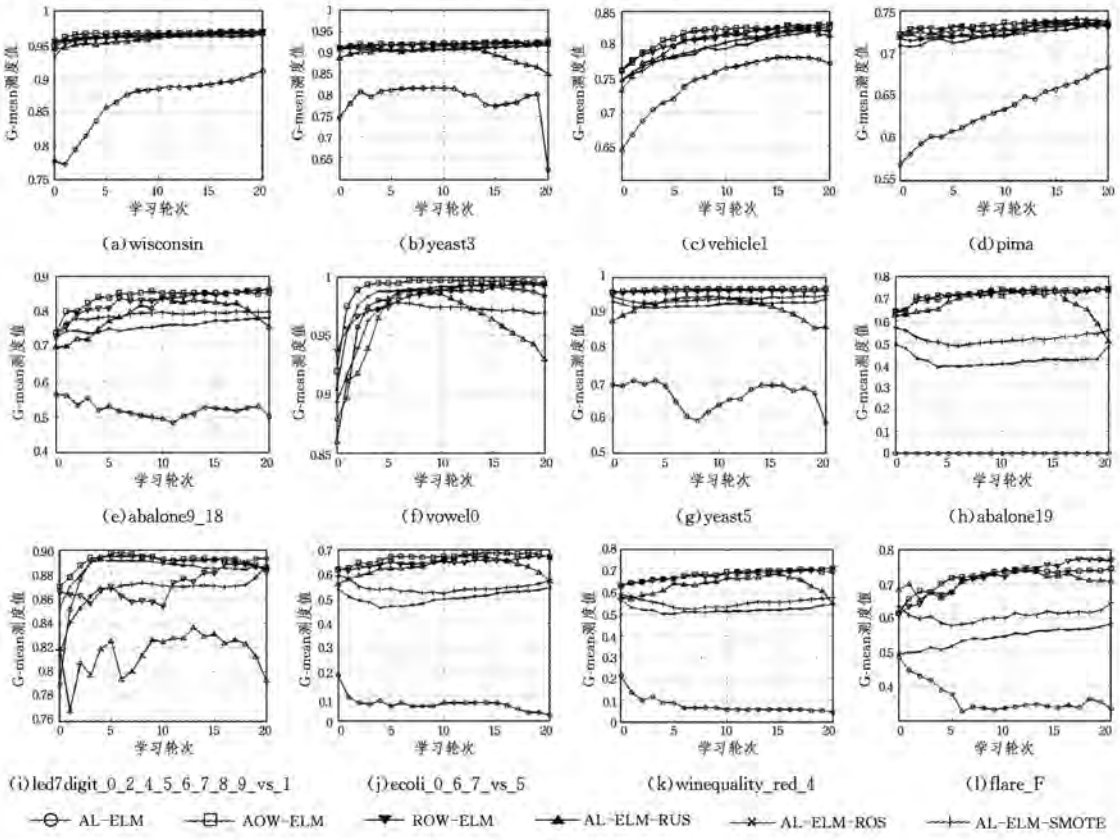

图 3　6 种比较算法在 12 个数据集上的学习曲线

表 2　6 种比较算法在 12 个数据集上的 ALC 测度值（最优结果以黑体标注）

数据集	学习算法					
	AL-ELM	ROW-ELM	AL-ELM-RUS	AL-ELM-ROS	AL-ELM-SMOTE	AOW-ELM
wisconsin	0.8666	0.9639	0.9590	0.9598	0.9630	**0.9654**
yeast3	0.7939	**0.9198**	0.8970	0.9080	0.9107	0.9191
vehicle1	0.7478	0.8104	0.8034	0.7946	0.7973	**0.8146**
pima	0.6325	0.7299	0.7283	0.7213	0.7244	**0.7314**
abalone9_18	0.5195	0.8292	0.7903	0.7579	0.7889	**0.8415**
vowel0	0.9782	0.9832	0.9604	0.9862	0.9664	**0.9929**
yeast5	0.6647	0.9597	0.9162	0.9215	0.9365	**0.9641**
abalone19	0.0000	0.7150	0.6839	0.4173	0.5159	**0.7203**
led7digit_0_2_4_5_6_7_8_9_vs_1	0.8870	0.8709	0.8157	0.8856	0.8669	**0.8909**
ecoli_0_6_7_vs_5	0.0682	0.6487	0.6343	0.5034	0.5441	**0.6702**
winequality_red_4	0.0715	**0.6836**	0.6375	0.5184	0.5517	0.6821
flare_F	0.3638	**0.7208**	0.7102	0.5432	0.6061	0.7199

由图 3 与表 2 的实验结果可以看出：

1)无论采用代价敏感还是样本采样技术作为主动学习的平衡控制策略，均可以大幅提升主动学习的性能，这一结果再次表明了在主动学习过程中引入平衡控制策略的必要性。显然，引入平衡控制策略可以保证在主动学习过程中所生成的每个分类面均是公正的，进而保证所采集的“查询样本”的有效性。

2)未引入平衡控制策略的 AL-ELM 算法在某些数据集上的表现极差，如 abalone19 数据集，但在如 vowel0 数据集上，其表现得则与其他算法大致相当。相信这一现象的出现可能与数据集的样本分布存在着密切的联系，当数据集中不同类的样本存在明显的可分性时，其并不会对 AL-ELM 算法的性能产生负面影响。

3)在 3 种采样算法中，SMOTE 算法显然要略优于 ROS 算法，而在大多数情况下，RUS 算法的性能要明显优于其他两种算法。相信这与 3 类算法的特点有关，ROS 算法仅通过复制少数类样本的方式来实现过采样，易于陷入过适应，而 SMOTE 算法则不会如此。至于 RUS 算法，若不考虑主动学习，则其在不平衡比率较高的数据集上的性能往往极不稳定，而鉴于主动学习引入了大量处于分类边界附近的样本，因此其性能会大幅提升。

4)相比于 3 种基于采样技术的主动学习算法，两种以代价敏感学习技术作为平衡控制策略的算法显然可以获得更优的性能，表明它们在保证分类面公正性方面要比前者做得更好。而对比采用了随机采样策略的 ROW-ELM 算法，以不确定度作为度量标准，进而提取“查询样本”的 AOW-ELM 算法明显更优，它在 9 个数据集上获得了最大的 ALC 测度值，而 ROW-ELM 算法仅在 3 个数据集上取得了最大的 ALC 测度值，且在这 3 个数据集上，其性能仅略好于前者。上述结果证明了以不确定度作为“查询样本”评价策略的正确性。

此外，我们也对各类算法在整个主动学习过程中所耗费的时间开销进行了统计，具体结果如表 3 所列。从表 3 的结果中不难看出，采用了在线学习模式的 3 种算法显然有更小的时间复杂度，在大多数数据集上，其运行时间甚至小于 AL-ELM-RUS 算法。具体而言，对比 AL-ELM 算法与 ROW-ELM 算法，AOW-ELM 算法的运行时间略长，这主要是由于其增加了不确定性评估与排序步骤所致。另外，我们发现：尽管 AL-ELM-RUS 算法没有采用在线学习的模式，但其时间复杂度相对较低，主要原因在于 RUS 算法会大幅降低已标记训练样本集的规模，从而节省分类器在每一轮建模时所需的时间开销。最后，从该表中还可以看出：采用过采样技术作为平衡控制策略的两种学习算法明显具有更大的时间开销，且随着数据集规模的增大以及不平衡比率的增加而呈现递增的趋势，这一现象也较易理解：在主动学习的迭代过程中，过采样技术会显著增加训练集的样本复杂度，进而增加分类器建模的时间开销。

表 3 6 种比较算法在 12 个数据集上的运行时间/s

数据集	学习算法					
	AL-ELM	ROW-ELM	AL-ELM-RUS	AL-ELM-ROS	AL-ELM-SMOTE	AOW-ELM
wisconsin	0.0806	0.0972	0.1789	0.2761	0.4716	0.0993
yeast3	0.1882	0.1950	0.2590	0.8471	2.0041	0.2080
vehicle1	0.1232	0.1243	0.2735	0.4852	0.7238	0.1394
pima	0.0972	0.1071	0.2272	0.3068	0.4202	0.1154
abalone9_18	0.1128	0.1217	0.1342	0.4482	0.8408	0.1321
vowel0	0.1524	0.1581	0.2153	0.7145	1.3364	0.1700
yeast5	0.2012	0.2090	0.2319	1.0509	2.2875	0.2179
abalone19	0.5273	0.5502	0.5424	3.8002	10.3642	0.5746
led7digit_0_2_4_5_6_7_8_9_vs_1	0.0952	0.1102	0.1113	0.3427	0.5304	0.1243
ecoli_0_6_7_vs_5	0.2298	0.2272	0.2808	1.3369	3.0363	0.2428
winequality_red_4	0.2101	0.2148	0.2486	1.2584	2.7550	0.2371
flare_F	0.1539	0.1784	0.1841	0.7810	1.6619	0.1836

对比 6 种比较算法，本文所提 AOW-ELM 算法显然具有最好的性能，它可以有效缓解样本不平衡分布对主动学习过程的负面影响，进而保证学习的公正性。此外，由于引入了在线学习模式，亦可保证上述算法具有较低的时间复杂度。实验结果证明了本文所提 AOW-ELM 算法的有效性与优越性。

当然，该算法也存在自身的缺点：1)学习的不可逆性，即由于算法采用了在线更新模式，一旦在学习过程中引入了较多的噪声信息，则该信息便会永远影响分类模型的质量，且会引发误差叠加效应；2)权重设置的经验性，即更新权重仅与样本的动态不平衡比率有关，而没有考虑样本分布的影响，这也是其相较采样技术的劣势所在，对于分布较为特殊的样本集，利用本文算法进行建模可能会存在一定风险。

结束语 针对类别不平衡分布数据会对主动学习的性能造成负面影响这一问题，提出了一种高效的解决方案，并将其命名为 AOW-ELM 算法。该算法结合了主动学习、代价敏感学习与在线学习这 3 类学习模式的优点，在保证缓解样本不平衡分布的负面影响的前提下，亦能大幅提升主动学习的时空效率。此外，该算法采用了 ELM 作为主动学习的基分类器，也可保证在不损害分类模型泛化性能的基础上进一步减少其建模的时间开销。通过多个基准的二类不平衡数据集验证了上述算法的有效性与可行性。

由于本文算法并未在多类不平衡数据集上进行验证，实验中也未采用具有行业背景的真实数据，因此希望未来的扩展工作能在上述两类问题上展开。

参 考 文 献

[1] SETTLES B. Active learning literature survey[R]. University of Wisconsin, Madison, 2010.

(下转第 254 页)

Research of Computers, 2009, 26(5): 1977-2000. (in Chinese)
陈自郁,何中市,张程. 基于冯诺依曼邻居的粒子群多阈值分割算法[J]. 计算机应用研究,2009,26(5):1977-2000.

[13] KENNEDY J. Small World and mega-minds: effects of neighborhood topology on particle swarm performance[C]// IEEE Congress on Evolutionary Computation. Piscataway, NJ, 1999: 1931-1938.

[14] RIGET J, VESTERSTROEM J S. A diversity guided particle swarm optimizer- the ARPSO[R]. Denmark: University of Aaehus, 2002.

[15] KOREL B. Dynamic method for software test data generation [J]. Software Testing, Verification and Reliability, 1992, 2(4): 203-213.

[16] TRACEY N, CLARK J, MANDER K, et al. An automated framework for structural test-data generation[C]// Proceeding of the 13th International Conference on Automated Software Engineering. Honolulu, HI, USA, 1998: 285-288.

[17] MCMINN P. Search-based software test data generation: A survey[J]. Software Testing, Verification and Reliability, 2004, 14(2): 105-156.

[18] HARMAN M, MCMINN P. A theoretical and empirical study of search-based testing: local, global, and hybrid search[J]. IEEE Transactions on Software Engineering, 2010, 36(2): 226-247.

[19] MA S L, YE D Y, YANG L L. A new design criteria of particle swarm topology[J]. Computer Engineering, 2015, 41(1): 200-206. (in Chinese)
马胜蓝,叶东毅,杨玲玲. 一种新的粒子群拓扑结构设计准则[J]. 计算机工程,2015,41(1):200-206.

(上接第 226 页)

[2] WANG M, HUA X S. Active learning in multimedia annotation and retrieval: a survey[J]. ACM Transactions on Intelligent System and Technology, 2011, 2(2): 210-231.

[3] SETTLES B, CRAVEN M. An analysis of active learning strategies for sequence labeling tasks[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2008: 1069-1078.

[4] LI Y, FANG B X, GUO L, et al. Supervised Intrusion Detection Based on Active Learning and TCM-KNN Algorithm [J]. Chinese Journal of Computers, 2007, 30(8): 1464-1473. (in Chinese)
李洋,方滨兴,郭莉,等. 基于主动学习和 TCM-KNN 方法的有指导入侵检测技术[J]. 计算机学报,2007,30(8):1464-1473.

[5] YU D, VARADARAJAN B, DENG L. Active learning and semi-supervised learning for speech recognition: A unified framework using the global entropy reduction maximization criterion[J]. Computer Speech & Language, 2010, 24(3): 433-444.

[6] NIU B, CHENG J, BAI X. A symmetric propagation based batch mode active learning for image retrieval[J]. Signal Processing, 2013, 93(6): 1639-1650.

[7] XU M X, SUN F M, LI H J. Online multi-label image classification with active learning[J]. Journal of Image and Graphics, 2015, 20(2): 237-244. (in Chinese)
徐美香,孙福明,李豪杰. 主动学习的多标签图像在线分类[J]. 中国图象图形学报,2015,20(2):237-244.

[8] MOHAMED T P, CARBONELL J G, GANAPATHIRAJU M K. Active learning for human protein-protein interaction prediction[J]. BMC bioinformatics, 2010, 11(Suppl 1): S57.

[9] ZHU J, HOVY E. Active Learning for Word Sense Disambiguation with Methods for Addressing the Class Imbalance Problem[C]// Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Prague, 2007: 783-790.

[10] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1-3): 489-501.

[11] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on System, Man and Cybernetics, Part B: Cybernetics, 2012, 42(2): 513-529.

[12] HUANG G B, WANG D H, LAN Y. Extreme learning machine: a survey[J]. International Journal of Machine Learning and Cybernetics, 2011, 2(2): 107-122.

[13] ZONG W, HUANG G B, CHEN Y. Weighted extreme learning machine for imbalance learning [J]. Neurocomputing, 2013, 101(3): 229-242.

[14] MCCALLU A, NIGRAM K. Employing EM in pool-based active learning for text classification[C]// Proceedings of the International Conference on Machine Learning. Morgan Kaufmann, 1998: 350-358.

[15] ZHU X, ZHANG P, LIN X, et al. Active learning from stream data using optimal weight classifier ensemble[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(6): 1607-1621.

[16] YU H L, SUN C Y, YANG W K, et al. AL-ELM: One uncertainty-based active learning algorithm using extreme learning machine[J]. Neurocomputing, 2015, 166: 140-150.

[17] LIANG N Y, HUANG G B, SARATCHANDRAN P. A fast and accurate online sequential learning algorithm for feedforward networks[J]. IEEE Transactions on Neural Networks, 2006, 17(6): 1411-1423.

[18] ALCALA FDES J, FEMANDEZ A, LUENGO J, et al. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework[J]. Journal of multiple-valued logic and soft computing, 2011, 17(2/3): 255-287.

[19] CHAWLA N V, BOWYER K W, HALL L O. SMOTE: Synthetic Minority Over-Sampling Technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1): 321-357.

[20] GUYON I, CAWLEY G C, DROR G, et al. Results of the Active Learning Challenge[C]// JMLR: Workshop and Conference Proceedings. 2011: 19-45.