·计算机工程与应用·

关于Real AdaBoost算法的分析与改进

付忠良

(中国科学院成都计算机应用研究所 成都 610041)

【摘要】采用一种新的技术，对Real AdaBoost算法的有效性、误差估计、算法流程和弱分类器训练进行了分析和证明。证明了可用加权组合弱分类器对Real AdaBoost算法进行改进，并得到了近似最佳组合系数；指出Real AdaBoost算法的样本权值调整和弱分类器训练方法的真实目的是确保弱分类器的独立性；基于Bayes统计推断对Real AdaBoost算法进行了多分类推广，得到了算法公式和误差估计，给出了便于使用的弱分类器训练简化方法。得到了Gentle AdaBoost算法的误差估计公式。UCI数据实验验证了所提算法和改进算法的效果。

关 键 词 分类器组合；集成学习；Gentle AdaBoost；Real AdaBoost

中图分类号 TP391 **文献标识码** A doi:10.3969/j.issn.1001-0548.2012.04.013

Analysis and Improvement on Real AdaBoost Algorithm

FU Zhong-liang

(Chengdu Institute of Computer Application, Chinese Academy of Sciences Chengdu 610041)

Abstract The effectiveness, error formula, algorithm flow, and weak classifiers training of Real AdaBoost algorithm are analyzed and proved by a new technique. Real AdaBoost algorithm is improved by weighted combination of weak classifiers and the approximately best combination coefficients are obtained. It is proved that the function of sample weight adjusting method and weak classifiers training method is to guarantee the independence of weak classifiers. Multi-class Real AdaBoost algorithm is proposed based on Bayes statistics deduction. The formula of algorithm and the estimation of classification error are discussed. The training method of weak classifiers is simplified. The estimation of classification error of Gentle AdaBoost is obtained. The effectiveness of the proposed algorithms is verified by the experiment on UCI dataset.

Key words classification combination; ensemble learning; Gentle AdaBoost; Real AdaBoost

AdaBoost(adaptive boosting)算法[1]是于1995年提出的一种Boost算法，可自动选取弱分类器组合成强分类器以提升分类精度。文献[2]对AdaBoost算法进行了改进并提出了Real AdaBoost算法，其将AdaBoost算法从处理二值判定推广到具有连续置信度输出。连续置信度能更精确地刻画分类边界[3]，理论上，Real AdaBoost算法效果应更好。文献[4-6]均是基于AdaBoost算法应用成功后提出用Real AdaBoost算法进行改进。

AdaBoost算法在人脸检测系统得到的成功应用[7-9]促进了对AdaBoost的研究，但目前对Real AdaBoost算法的研究及应用比AdaBoost算法少，缺乏对Real AdaBoost算法的实质分析以及使用算法时必要的适应性调整指导，可能是其主要原因。文献[10-11]就集成学习算法确保算法有效对弱分类器选取和样本抽样的条件要求进行了比较详细的论述，类似观点如果能对Real AdaBoost算法进行解释，将有助于其应用发展。文献[11]讨论了分类器最佳组合问题并得到了近似最佳组合系数，Real AdaBoost算法属于一种分类器组合，可否采用加权组合对算法进行改进值得研究。文献[2]推广得到的多分类Real AdaBoost算法与Bayes统计推断不一致，但AdaBoost算法却与Bayes统计推断一致，于是依据Bayes统计推断推广Real AdaBoost算法是否可行也值得研究。

本文用一种新的方法分析并证明了Real AdaBoost算法的有效性条件和置信度公式，并从一

收稿日期：2010－09－09；修回日期：2011－09－19
基金项目：四川省科技支撑计划项目(2008SZ0100, 2009SZ0214)
作者简介：付忠良(1967－)，男，研究员，博士生导师，主要从事机器学习和图形图像处理方面的研究.

种新的角度分析了Real AdaBoost算法实质；基于分类器最优组合定理[11]提出了对Real AdaBoost算法的改进，得到了改进后的误差估计，实验数据也验证了改进效果。分析指出可以基于Bayes统计推断得到多分类Real AdaBoost算法。同时讨论了Gentle AdaBoost算法的有效性条件、误差估计和改进方法，其中误差估计公式属于首次得到。

1 Real AdaBoost算法简介

训练样本集 $S=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$，考虑二分类问题，$y_i\in\{-1,+1\}$，弱分类器空间记为$H$。Real AdaBoost算法流程为：

1) 初始化权值：$\omega_i^1=1/m$，$i=1,2,\cdots,m$。

2) DO FOR $t=1,2,\cdots,T$

① 基于带 ω_i^t 的训练集S训练弱分类器：

a. 对S进行划分，$S=S_1\cup S_2\cup\cdots\cup S_n$，$i\neq j$ 时，$S_i\cap S_j=\varnothing$；

b. 统计 S_j 中 +1 和 –1 类累积样本权，$W_+^{jt}=\sum_{i:(x_i\in S_j)\wedge(y_i=1)}\omega_i^t$，$W_-^{jt}=\sum_{i:(x_i\in S_j)\wedge(y_i=-1)}\omega_i^t$；

c. 定义 $h_t(x)$，$\forall x\in S_j$，令 $h_t(x)=\frac{1}{2}\ln(W_+^{jt}+\delta)/(W_-^{jt}+\delta)$，$j=1,2,\cdots,n$，$\delta$ 为平滑因子(一小正数)；

d. 选取 $h_t(x)$，最小化 $Z_t=2\sum_{j=1}^{n}\sqrt{W_+^{jt}W_-^{jt}}$，选取 $h_t(x)=\arg\min_{h\in H}Z_t$。

②调整样本权，$\omega_i^{t+1}=\frac{\omega_i^t}{Z_t}\exp(-y_ih_t(x_i))$。

3) 强分类器：$H(x)=\mathrm{sign}(f(x))$，其中 $f(x)=\sum_{t=1}^{T}h_t(x)$。

Real AdaBoost算法训练错误率估计[2]为：

$$\frac{1}{m}\left|\{i:H(x_i)\neq y_i\}\right|\leqslant\prod_{t=1}^{T}\left(2\sum_{j=1}^{n}\sqrt{W_+^{jt}W_-^{jt}}\right)\tag{1}$$

算法采取给错分样本更大权值。当采用二段划分时($n=2$)，$W_+^{1t}+W_-^{2t}$ 为 $h_t(x)$ 在分布为 ω_i^t 下的训练正确率，$W_+^{2t}+W_-^{1t}$ 为训练错误率。如果+1类被错分成–1类与–1类被错分成+1类的错误率相等，即 $W_+^{2t}=W_-^{1t}=\varepsilon_t/2$，$W_+^{1t}=W_-^{2t}=(1-\varepsilon_t)/2$，其中 ε_t 为 $h_t(x)$ 在分布为 ω_i^t 下的训练错误率，则：

$$h_t(x)=\frac{1}{2}\begin{cases}\ln((1-\varepsilon_t)/\varepsilon_t) & \text{if } x\in S_1\\ -\ln((1-\varepsilon_t)/\varepsilon_t) & \text{if } x\in S_2\end{cases}\tag{2}$$

$H(x)$等同 $\mathrm{sign}\left(\sum_{t=1}^{T}\alpha_th_t'(x)\right)$，$h_t(x)=\begin{cases}+1 \text{ if } x\in S_1\\ -1 \text{ if } x\in S_2\end{cases}$，$\alpha_t=\frac{1}{2}\ln(1-\varepsilon_t/\varepsilon_t)$，这正是AdaBoost。因此AdaBoost`是Real AdaBoost的特例。

$$\sqrt{W_+^{1t}W_-^{1t}}+\sqrt{W_+^{2t}W_-^{2t}}\leqslant\sqrt{(W_+^{1t}+W_-^{2t})(W_+^{2t}+W_-^{1t})}=\sqrt{(1-\varepsilon_t)\varepsilon_t}$$

所以，Real AdaBoost比AdaBoost有更小的训练错误率，因为有：

$$\frac{1}{m}\left|\{i:H(x_i)\neq y_i\}\right|\leqslant\prod_{t=1}^{T}\left(2\sum_{j=1}^{2}\sqrt{W_+^{jt}W_-^{jt}}\right)\leqslant 2^T\prod_{t=1}^{T}\sqrt{(1-\varepsilon_t)\varepsilon_t}\tag{3}$$

2 Real AdaBoost算法分析与改进

2.1 Real AdaBoost算法分析

文献[2]提出了Real AdaBoost算法，下面用一种不同的技术来分析和证明Real AdaBoost算法。

$h_t(x)$ 在划分 $S=S_1\cup S_2\cup\cdots\cup S_n$ 上，$\forall x\in S_j$，$h_t(x)=\alpha_j$。W_+^{jt} 和 W_-^{jt} 定义同前所述，$H(x)=\mathrm{sign}(f(x))$，$f(x)=\sum_{t=1}^{T}h_t(x)$。定义随机变量：

$$R_t=\begin{cases}\alpha_j & \text{if } y=+1,x\in S_j\\ -\alpha_j & \text{if } y=-1,x\in S_j\end{cases}\quad j=1,2,\cdots,n\tag{4}$$

令 $R=\sum_{t=1}^{T}R_t$，则有如下定理：

定理 1 $H(x)$的训练错误率 ε 等于随机变量R小于等于零的概率，即 $\varepsilon=P_r[R\leqslant 0]$。

证明：$H(x)$ 错分 x 当且仅当 $yf(x)\leqslant 0$，因 $yf(x)=\sum_{t=1}^{T}yh_t(x)$，$\sum_{t=1}^{T}yh_t(x)\leqslant 0$ 与 $\sum_{t=1}^{T}R_t\leqslant 0$ 等价，于是 $\varepsilon=P_r[R\leqslant 0]$。

证毕。

由式(4)有 $P_r[R_t=\alpha_j]=P_r[y=1,x\in S_j]=W_+^{jt}$，$P_r[R_t=-\alpha_j]=P_r[y=-1,x\in S_j]=W_-^{jt}$。$\forall x\in S$，若各 $h_t(x)$ 的取值相互独立(简称 $h_t(x)$ 相互独立)，此时 R_t 也相互独立，在该假设下有定理：

定理 2 当 $h_t(x)$ 相互独立，则 $\alpha_j=0.5\ln(W_+^{jt}/W_-^{jt})$ 时，$H(x)$ 的训练错误率 $\varepsilon\leqslant\prod_{t=1}^{T}\left(2\sum_{j=1}^{n}\sqrt{W_+^{jt}W_-^{jt}}\right)$。

证明：记 $g(r)$ 为R的概率密度函数，由定理1

$$\varepsilon = P_r[R \leqslant 0] = \int_{-\infty}^{0} g(r)\mathrm{d}r \leqslant \int_{-\infty}^{0} \mathrm{e}^{-r} g(r)\mathrm{d}r \leqslant \int_{-\infty}^{+\infty} \mathrm{e}^{-r} g(r)\mathrm{d}r = E[\mathrm{e}^{-R}] = \prod_{t=1}^{T} E[\mathrm{e}^{-R_t}] = \prod_{t=1}^{T}\left(\sum_{j=1}^{n}(W_+^{jt}\mathrm{e}^{-\alpha_j} + W_-^{jt}\mathrm{e}^{\alpha_j})\right) \tag{5}$$

$W_+^{jt}\mathrm{e}^{-\alpha_j} + W_-^{jt}\mathrm{e}^{\alpha_j} \geqslant 2\sqrt{W_+^{jt}W_-^{jt}}$，当 $W_+^{jt}\mathrm{e}^{-\alpha_j} = W_-^{jt}\mathrm{e}^{\alpha_j}$，式(5)取极小值，此时 $\alpha_j = 0.5\ln(W_+^{jt}/W_-^{jt})$，代入，即得定理。

证毕。

定理2结论同文献[2]，此处用一种新技术得到了Real AdaBoost算法，包括置信度和误差估计。定理结论强烈地依赖弱分类器的独立性条件，看似很难满足，但正如文献[11]指出的，集成学习算法的样本权值调整和弱分类器选取策略正是为了尽量确保训练的弱分类器满足独立性条件而制定的。为此下面将分析Real AdaBoost的样本权值调整和弱分类器选取。

仿照文献[11]的“组合累积权”定义，称 $y_i f(x_i)$ 为样本 x_i 的“带标签累积置信度”。新增 $h_t(x)$ 后，x_i 的带标签累积置信度将增加 $y_i h_t(x_i)$，在集成学习算法中，新增 $h_t(x)$ 总是为了提升 $H(x)$ 的训练正确率，即 $y_i f(x_i) > 0$的样本数增加。对于带标签累积置信度为负的样本，总是希望通过增加新的弱分类器使其改变为正，而带标签累积置信度越小的样本，在下一轮弱分类器选取时应给予越多的关注，因此用 $y_i f(x_i)$ 的单调递减函数来调整权是合适的。当用负指数函数可得权值调整公式 $\omega_i^{t+1} = \omega_i^t / Z_t \exp(-y_i h_t(x_i))$，其正是Real AdaBoost的样本权调整公式。可见，Real AdaBoost算法的样本权值调整目的正是为了各个弱分类器可正确分类样本分布尽量均匀，即让弱分类器尽量满足独立性条件。而最小化 Z_t 来选取弱分类器，也体现了新的弱分类器更加关注有较大权值的样本，而错分样本有较大权值，可见弱分类器选取策略也是为了同样的目的。

基于上述分析还可得到如下的简化策略：

$$h_t(x) = \arg\min_{h \in H} \varepsilon_t \tag{6}$$

式中，$\varepsilon_t = \sum_{i=1}^{m} \omega_i^t [\mathrm{sign}(h_t(x_i)) \neq y_i]$ 为训练错误率。

AdaBoost与Bayes统计推断具有等价性[1]，Real AdaBoost也与Bayes统计推断等价，具体分析如下：

$\forall x_i \in S$，当其位于 $h_t(x)$ 对应划分第 j 段，有 $P_r[y_i = 1] = W_+^{jt}/(W_+^{jt} + W_-^{jt})$ 和 $P_r[y_i = -1] = W_-^{jt}/(W_+^{jt} + W_-^{jt})$。若 $h_t(x)$ 相互独立，x_i 位于 $h_t(x)$ 对应划分的第 j_t 段，有 $P_r[y_i = 1] = \prod_{t=1}^{T} W_+^{j_t t} / (W_+^{j_t t} + W_-^{j_t t})$ 和 $P_r[y_i = -1] = \prod_{t=1}^{T} W_-^{j_t t} / (W_+^{j_t t} + W_-^{j_t t})$。根据Bayes统计推断，其将输出概率最大对应的标签，于是对各项取对数并略掉公共项后可得到 $\mathrm{sign}(P_r[y_i=1] - P_r[y_i=-1]) = \mathrm{sign}(\ln(P_r[y_i=1]) - \ln(P_r[y_i=-1])) = \mathrm{sign}\left(2\sum_{t=1}^{T} h_t(x)\right)$，这正是算法的强分类器。

2.2　Real AdaBoost算法的改进

Real AdaBoost的强分类器为各个弱分类器之和，现采用加权和对其进行改进。强分类器定义为：

$$H(x) = \mathrm{sign}\left(\sum_{t=1}^{T} \beta_t h_t(x)\right) \tag{7}$$

R_t 由式(4)定义，其均值和方差分别为：

$$\mu_t = \frac{1}{2}\sum_{j=1}^{n}((W_+^{jt} - W_-^{jt})\ln(W_+^{jt}/W_-^{jt})) \tag{8}$$

$$\sigma_t^2 = \sum_{j=1}^{n}(W_+^{jt}(0.5\ln(W_+^{jt}/W_-^{jt}) - \mu_t)^2 + W_-^{jt}(0.5\ln(W_+^{jt}/W_-^{jt}) + \mu_t)^2) \tag{9}$$

令 $R = \sum_{t=1}^{T} \beta_t R_t$，其均值和方差记为 μ 和 σ^2，则有定理：

定理 3　当 $h_t(x)$ 相互独立且 $\mu > 0$，组合系数 $\beta_t = \mu_t/\sigma_t^2$ 时，$H(x)$ 的训练错误率 $\varepsilon \leqslant 1\Big/\sum_{t=1}^{T}(\mu_t^2/\sigma_t^2)$；当 T 很大时，如果 μ_t^2/σ_t^2 有界，但 $\sum_{t=1}^{T}(\mu_t^2/\sigma_t^2)$ 很大，该组合系数近似为最优组合系数。

证明：类似定理1的分析，仍然有 $\varepsilon = P_r[R \leqslant 0]$。记 $g(r)$ 为 R 的概率密度函数，则：

$$\varepsilon = P_r[R \leqslant 0] = \int_{-\infty}^{0} g(r)\mathrm{d}r \leqslant \int_{-\infty}^{0} (r-\mu)^2/\mu^2\, g(r)\mathrm{d}r \leqslant \int_{-\infty}^{+\infty} (r-\mu)^2/\mu^2\, g(r)\mathrm{d}r = \sigma^2/\mu^2 = \left(\sum_{t=1}^{T}\beta_t^2\sigma_t^2\right)\Big/\left(\sum_{t=1}^{T}\beta_t\mu_t\right)^2 \tag{10}$$

$\left(\sum_{t=1}^{T}\beta_t\mu_t\right)^2 = \left(\sum_{t=1}^{T}\beta_t\sigma_t\frac{\mu_t}{\sigma_t}\right)^2 \leqslant \sum_{t=1}^{T}(\beta_t\sigma_t)^2\sum_{t=1}^{T}\frac{\mu_t^2}{\sigma_t^2}$，即

$\beta_t=\mu_t/\sigma_t^2$ 时，σ^2/μ^2 取到极小值，此时式(10)的分子分母都有的公共项 $\sum_{t=1}^{T}\beta_t^2\sigma_t^2$ 被约掉了，于是定理前部分成立。

再证明 $T\to\infty$ 时，σ^2/μ^2 的极小值点就是 $\varepsilon=P_r[R\leqslant 0]$ 的极小值点。$\beta_t=\mu_t/\sigma_t^2$ 时，$\beta_t R_t$ 的均值和方差都为 μ_t^2/σ_t^2，记 $\theta=\sum_{t=1}^{T}\mu_t^2/\sigma_t^2$，则 $\sigma^2/\mu^2=1/\theta$。μ_t^2/σ_t^2 有界，由极限定理，$T\to\infty$ 时，$\frac{1}{T}R=\frac{1}{T}\sum_{t=1}^{T}\beta_t R_t$ 是均值为 θ/T、方差为 θ/T^2 的正态分布，则 $(-R/T+\theta/T)/(\sqrt{\theta}/T)=(-R+\theta)/\sqrt{\theta}$ 为标准正态分布。对标准正态分布的随机变量 Y，当 $v\to\infty$ 时，$P_r[Y\geqslant v]\approx\left(v\sqrt{2\pi}\right)^{-1}\exp(-v^2/2)$ 是 v 的单调函数。$\varepsilon=P_r[R\leqslant 0]=P_r[(-R+\theta)/\sqrt{\theta}\geqslant\sqrt{\theta}]$，当 $T\to\infty$，且 $\theta\to\infty$ 时，$\varepsilon\approx(\sqrt{2\pi\theta})^{-1}\exp(-\theta/2)$。因此，$\sigma^2/\mu^2=1/\theta$ 取到极小值时，ε 取到极小值，定理后部分成立。

证毕。

定理3指出了Real AdaBoost算法的一种改进，即用 $y_i h_t(x_i)$ ($i=1,2,\cdots,m$)数据的均值与方差比为加权系数，强分类器为弱分类器的之加权和。

$\beta_t=\mu_t/\sigma_t^2$ 时，定理要求的条件 $\mu>0$ 满足。μ_t^2/σ_t^2 有界而 $\lim_{T\to\infty}\theta=\infty$ 条件也容易满足，对样本集 S 进行分段时，每一段内都有两种样本即可保证满足条件，而每个 $h_t(x)$ 对 S 的划分段数是可以不一样的。

参与组合的弱分类器很多时，定理3给出的组合近似为最优组合，这一结论表明：引入加权系数可以改进Real AdaBoost算法。

均值方差比可以定义为随机变量的一种归一化，因为 $y_i\beta_t h_t(x_i)$ 的均值方差比为1，定理3表明在归一化条件下构造弱分类器的置信度更好。引入加权系数 β_t 后样本权值调整公式需同步调整为 $\omega_i^{t+1}=\omega_i^t/Z_t\exp(-y_i\beta_t h_t(x_i))$。

2.3 Gentle AdaBoost算法分析与改进

类似定理3证明方法可得到Gentle AdaBoost算法的置信度、误差估计及改进方法。分析如下：

R_t 仍由式(4)定义，$R=\sum_{t=1}^{T}R_t$，只要 R 的均值大于零，$h_t(x)$ 相互独立时，$H(x)$ 的训练错误率满足 $\varepsilon\leqslant\sum_{t=1}^{T}\sigma_t^2\Big/\left(\sum_{t=1}^{T}\mu_t\right)^2$。$\alpha_j$ 取值使其分子越小越好，注意到 $\sigma_t^2=\sum_{j=1}^{n}(W_+^{jt}(\alpha_j-\mu_t)^2+W_-^{jt}(\alpha_j+\mu_t)^2)$，把 μ_t 当常数对待时，σ_t^2 的极小值点可作为近似极小值点，于是可得置信度和误差估计分别为：

$$\alpha_j=(W_+^{jt}-W_-^{jt})/(W_+^{jt}+W_-^{jt}) \tag{11}$$

$$\varepsilon\leqslant\sum_{t=1}^{T}\sigma_t^2\Big/\left(\sum_{t=1}^{T}\mu_t\right)^2=\sum_{t=1}^{T}(\mu_t-\mu_t^2)\Big/\left(\sum_{t=1}^{T}\mu_t\right)^2=1\Big/\sum_{t=1}^{T}\mu_t-\sum_{t=1}^{T}\mu_t^2\Big/\left(\sum_{t=1}^{T}\mu_t\right)^2\leqslant 1\Big/\sum_{t=1}^{T}\mu_t \tag{12}$$

式中，$\mu_t=\sum_{j=1}^{n}((W_+^{jt}-W_-^{jt})^2/(W_+^{jt}+W_-^{jt}))$，满足 R 均值大于零条件；上面用到 $\sigma_t^2=(\mu_t-\mu_t^2)$，其证明只需展开 σ_t^2，并注意到 $\sum_{j=1}^{n}(W_+^{jt}+W_-^{jt})\equiv 1$。

$$\sigma_t^2=\sum_{j=1}^{n}(W_+^{jt}(\alpha_j-\mu_t)^2+W_-^{jt}(\alpha_j+\mu_t)^2)=\sum_{j=1}^{n}((W_+^{jt}+W_-^{jt})\alpha_j^2-2(W_+^{jt}-W_-^{jt})\alpha_j\mu_t+(W_+^{jt}+W_-^{jt}))\mu_t^2=\mu_t-2\mu_t\mu_t+\mu_t^2 \tag{13}$$

式(11)正是Gentle AdaBoost算法的置信度公式[12]。此处不仅推导出了置信度公式，且得到了Gentle AdaBoost算法训练错误率估计。Gentle AdaBoost算法的弱分类器选取为：

$$h_t(x)=\arg\max_{h\in H}\mu_t \tag{14}$$

对Gentle AdaBoost算法，同样可以引入加权系数 $\beta_t=\mu_t/\sigma_t^2=1/(1-\mu_t)$ 对算法进行改进，可以验证改进后的Gentle AdaBoost算法训练错误率估计为 $\varepsilon\leqslant 1\Big/\sum_{t=1}^{T}\left(\mu_t/(1-\mu_t)\right)$。

3 多分类Real AdaBoost算法分析

3.1 多分类Real AdaBoost算法

设训练样本集 $S=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$，$y_i\in\{1,2,\cdots,K\}$。$h_t(x,l)$ 表示 $h_t(x)$ 输出标签 l 的置信度，$l=1,2,\cdots,K$。

$\forall x_i\in S$，当位于 $h_t(x)$ 对 S 的划分的第 j 段，则 $P_r[y_i=l]=W_l^{jt}\Big/\sum_{k=1}^{K}W_k^{jt}$，其中 $W_k^{jt}=\sum_{i:(y_i=k)\wedge(x_i\in S_j)}\omega_i^t$。

若 $h_t(x)$ 相互独立，同标签的概率直接相乘，Bayes统计推断输出概率最大对应的标签，取对数并略掉每个标签都有的公共项，于是可得到如下的 K 分类

Real AdaBoost算法：

1) 初始化权值：$\omega_i^1 = 1/m$，$i = 1,2,\cdots,m$。

2) DO FOR $t = 1,2,\cdots,T$

① 基于带ω_i^t的训练集S训练弱分类器：

a. 对S进行划分，$S = S_1 \cup S_2 \cup \cdots \cup S_n$，$i \neq j$时，$S_i \cap S_j = \varnothing$。

b. 统计S_j中标签为k的累积样本权，$W_k^{jt} = \sum_{i:(x_i \in S_j) \wedge (y_i = k)} \omega_i^t$，$k$=1,2,⋯；$K$=1,2,⋯,$m$；$j$=1,2,⋯,$n$。

c. 定义$h_t(x,l)$：$\forall x \in S_j$，令$h_t(x,l) = \ln(W_l^{jt})$，$l = 1,2,\cdots,K$；$j = 1,2,\cdots,n$。

d. 选取$h_t(x)$，最小化$Z_t = K\sum_{j=1}^{n}\sqrt[K]{\prod_k W_k^{jt}}$，选取$h_t(x)$，即$h_t(x) = \arg\min_{h \in H} Z_t$。

② 调整样本权，$\omega_i^{t+1} = \frac{\omega_i^t}{Z_t}\exp\times\left(-\ln(W_{y_i}^{jt}) + \frac{1}{K}\sum_{k=1}^{K}\ln(W_k^{jt})\right)$。

3) 强分类器：$H(x) = \arg\max_l f(x,l)$，其中，$f(x,l) = \sum_{t=1}^{T} h_t(x,l)$。

具体使用时同样需引入平滑因子δ(一小正数)，即$h_t(x,l) = \ln(W_l^{jt} + \delta)$。下面对算法合理性进行分析。

如果弱分类器相互独立，前面的分析已表明算法等价于Bayes统计推断，而样本权值调整和弱分类器选取能否使选取的弱分类器相互独立是方法有效的关键。分析算法权值调整公式当$W_{y_i}^{jt} = \max_k W_k^{jt}$，$\arg\max_l h_t(x_i,l) = y_i$，因此，该调整系数同样体现了给错分样本更大权值，而弱分类器选取策略也体现了聚焦错分样本的特性，两者结合便可最大限度地使新选取的弱分类器相互独立。

定义一般形式的$h_t(x,l)$：$x \in S_j$时，$h_t(x,l) = \alpha_l^{jt}$。假设$h_t(x)$输出正确标签概率大于$1/K$(比随机猜测略好)，输出其他标签概率小于等于$1/K$，则有定理：

定理 4　当$h_t(x)$相互独立，则$\alpha_l^{jt} = \ln(W_l^{jt})$，$T \to \infty$时，$H(x)$的训练错误率$\varepsilon \leqslant \prod_{t=1}^{T}\left(K\sum_{j=1}^{n}\sqrt[K]{\prod_l W_l^{jt}}\right)$。

证明：

定义随机变量：

$$R_t = \left(\alpha_l^{jt} - \frac{1}{K}\sum_{k=1}^{K}\alpha_k^{jt}\right) \quad if\,(y = l, x \in S_j) \tag{15}$$

则$P_r\left[R_t = \left(\alpha_l^{jt} - \frac{1}{K}\sum_{k=1}^{K}\alpha_k^{jt}\right)\right] = W_l^{jt}$。$\frac{1}{K}\sum_{k=1}^{K}\alpha_k^{jt}$是$x \in S_j$时$h_t(x)$输出置信度的平均值，对$f(x,l) = \sum_{t=1}^{T} h_t(x,l)$而言，当$T \to \infty$时，由大数定理，正确标签对应的频率(加权频率)将大于$1/K$，其他标签的频率将小于$1/K$，当只有一个弱分类器时其意指$\alpha_{y_i}^{jt} \Big/ \sum_{k=1}^{K}\alpha_k^{jt} \geqslant \frac{1}{K}$的概率很大。因此，$x \in S_j$时，$H(x) = \arg\max_l f(x,l)$的错误率应比$R = \sum_{t=1}^{T} R_t \leqslant 0$的概率小，即$\varepsilon \leqslant P_r[R < 0]$。类似定理2的证明，有：

$$\varepsilon \leqslant P_r[R < 0] \leqslant \prod_{t=1}^{T} E[e^{-R_t}] = \prod_{t=1}^{T}\left(\sum_{j=1}^{n}\sum_{l=1}^{K} W_l^{jt}\exp\left(-\alpha_l^{jt} + \frac{1}{K}\sum_{k=1}^{K}\alpha_k^{jt}\right)\right) \tag{16}$$

由算术平均大于几何平均，等号成立条件为各数相等，$\alpha_l^{jt} = \ln(W_l^{jt})$时取到极小值，把该值代入式(16)即得定理。

证毕。

增加$h_t(x)$，每个标签的累积置信度都会增加，标签为l的增加$\ln(W_l^{jt})$，当增加值比所有标签增加值之平均值大，就可增加输出正确标签的可能性。类似前面的分析，标签y_i增加的置信度与所有标签增加的置信度平均值之差，能够较好地度量x_i对应标签y_i的置信度增加量，该值的负指数函数作为样本权值调整是合理的，于是有：

$$\omega_i^{t+1} = \frac{\omega_i^t}{Z_t}\exp\left(-\ln(W_{y_i}^{jt}) + \frac{1}{K}\sum_{k=1}^{K}\ln(W_k^{jt})\right) \tag{17}$$

定理4表明K分类Real AdaBoost算法是科学的。$K = 2$时算法与二分类Real AdaBoost算法完全一样，包括误差估计和置信度公式。因此，上述K分类Real AdaBoost算法应该是最自然的一种推广。

类似分析可得多分类Gentle AdaBoost的$h_t(x,l)$为：$\forall x \in S_j$，$h_t(x,l) = W_l^{jt} \Big/ \sum_{k=1}^{K} W_k^{jt}$，$l = 1,2,\cdots,K$，$j = 1,2,\cdots,n$，权值调整公式为：

$$\omega_i^{t+1} = \frac{\omega_i^t}{Z_t}\exp\left(-W_{y_i}^{jt} \Big/ \sum_{k=1}^{K} W_k^{jt}\right) \tag{18}$$

3.2 多分类Real AdaBoost算法使用简化

Real AdaBoost算法的弱分类器选取策略为 $h_t(x)=\arg\min_{h\in H} Z_t$，在 Z_t 中如果存在标签 b 使 $W_b^{jt}=0$，即 S_j 中无 b 标签样本，则 $\prod_k W_k^{jt}=0$，此时 Z_t 并不能真实反映 $h_t(x)$ 对样本集的划分情况，仍然依据 $h_t(x)=\arg\min_{h\in H} Z_t$ 选取弱分类器将可能难于满足弱分类器的独立性条件，即直接使用Real AdaBoost算法会出现问题，这可能是导致Real AdaBoost算法目前没得到广泛应用的原因之一。

为克服该问题，经过理论分析和实验验证发现，一种可替代策略是最小化式(19)来选取弱分类器。其理由是分析Real AdaBoost算法中弱分类器选取策略 $h_t(x)=\arg\min_{h\in H} Z_t$，由 Z_t 表达式可知，该策略实际上是尽量不要选取各个 W_k^{jt} 相等的情况(此时取不到极小值)，而式(19)具有同样功能，但能适应 $W_b^{jt}=0$。

$$h_t(x)=\arg\min_{h\in H}\left(K\sum_{j=1}^{n}\sqrt[K]{\prod_k(1+W_k^{jt})}\right) \tag{19}$$

与二分类Real AdaBoost算法类似，权值调整配合弱分类器选取都是为了使选取的弱分类器相互独立，因此，算法中弱分类器选取策略可统一简化为：

$$h_t(x)=\arg\min_{h\in H}\varepsilon_t \tag{20}$$

其中，$1-\varepsilon_t=\sum_{i=1}^{m}\omega_i^t\left[\!\left[\arg\max_l h_t(x_i,l)=y_i\right]\!\right]$。

权值调整也可直接仿照二分类AdaBoost方法：

$$\omega_i^{t+1}=\omega_i^t\times\begin{cases}\mathrm{e}^{-\alpha_t} & \arg\max_l h_t(x_i,l)=y_i\\ \mathrm{e}^{\alpha_t} & \arg\max_l h_t(x_i,l)\neq y_i\end{cases} \tag{21}$$

其中，$\alpha_t=\ln\left((1-\varepsilon_t)/\varepsilon_t\right)/K$。对多分类问题，有：

$$\omega_i^{t+1}=\omega_i^t\times\begin{cases}\mathrm{e}^{-\alpha_t(K-1)/K} & \arg\max_l h_t(x_i,l)=y_i\\ \mathrm{e}^{\alpha_t/K} & \arg\max_l h_t(x_i,l)\neq y_i\end{cases} \tag{22}$$

其中 $\alpha_t=\ln((1-\varepsilon_t)/(\varepsilon_t/(K-1)))$，其为多分类AdaBoost算法调整公式[11]，可用于多分类Real AdaBoost算法。Real AdaBoost和AdaBoost可以混合使用，利用Real AdaBoost的连续置信度可精细刻画分类边界的优点，利用AdaBoost的样本权值调整和弱分类器选取策略的易使用优点，二者结合能适应 $W_b^{jt}=0$。

4 实验与分析

4.1 实验内容和条件

理论上对Real AdaBoost算法进行了分析、改进与简化，再通过实验进一步验证。以下为算法名称说明。

Improved Real AdaBoost：引入组合系数改进的Real AdaBoost。Simple Real AdaBoost：弱分类器选取用式(6)(二类)和式(19)(三类)的Real AdaBoost。Practical Real AdaBoost：弱分类器选取用式(20)的Real AdaBoost。STW AdaBoost：权值调整改用式(21)的Real AdaBoost。Improved Gentle AdaBoost：引入组合系数改进的Gentle AdaBoost，其中Gentle AdaBoost的权值调整用式(14)。

实验数据选取了UCI数据集上的Ionosphere、Sonar和Wine数据集。训练集和测试集按数据集中不同类同比例随机抽取，重复多次后计算测试错误率的均值和方差，均值能反映算法效果，方差能反映算法稳定性。实验时训练集和测试集比率为6:4。所有算法都训练30个分类器，随机重复40次。

样本集的划分考虑基于单属性的简单划分，具体为Ionosphere数据和Sonar数据采取4段划分，正类样本均值与负类样本均值的平均值为阈值。划分样本集为上下两部分：最大值与该阈值的平均值为阈值划分上半部分为两段；最小值与该阈值的平均值为阈值分下半部分为两段。Wine数据采取3段划分，即计算3类样本各自均值，相邻两均值的平均为分段阈值完成数据3段划分。实验结果见表1～表3。

表1 Ionoshpere数据集上的实验结果

	测试错误率	测试错误率方差
AdaBoost	0.189 5	0.023 6
Real AdaBoost	0.106 8	0.023 7
Improved Real AdaBoost	0.093 9	0.017 4
Simple Real AdaBoost	0.103 4	0.020 5
Gentle AdaBoost	0.105 0	0.021 0
Improved Gentle AdaBoost	0.094 5	0.020 3

表2 Sonar数据集上的实验结果

	测试错误率	测试错误率方差
AdaBoost	0.253 3	0.044 5
Real AdaBoost	0.234 6	0.039 2
Improved Real AdaBoost	0.230 0	0.041 2
Simple Real AdaBoost	0.230 7	0.042 2
Gentle AdaBoost	0.233 7	0.037 2
Improved Gentle AdaBoost	0.230 5	0.045 3

表3 Wine数据集上的实验结果

	测试错误率	测试错误率方差
AdaBoost	0.072 2	0.028 0
STW AdaBoost	0.088 3	0.024 0
Real AdaBoost	0.207 0	0.036 8
Practical Real AdaBoost	0.054 6	0.024 3
Simple Real AdaBoost	0.051 4	0.029 7
Gentle AdaBoost	0.073 3	0.032 1

4.2 实验结果分析

对二分类问题(如表1和表2)，Real AdaBoost明

显好于AdaBoost，特别是在Ionosphere数据集上，测试错误率降低超过70%。Improved Real AdaBoost对Real AdaBoost有一定的改进，而Simple Real AdaBoost与Real AdaBoost一样好，因此在实际使用时，可直接使用Simple Real AdaBoost算法。Gentle AdaBoost与Real AdaBoost效果几乎一样，而Improved Gentle AdaBoost比Gentle AdaBoost约好。测试错误率方差都很小，说明算法是稳定的。

对多分类问题(如表3)，权值调整取式(22)(AdaBoost)好于取式(21)(STW AdaBoost)。Real AdaBoost似乎失效，正如前面的分析，其原因是出现了$W_b^{jt}=0$情形，当用Practical Real AdaBoost时，结果得到明显改进，测试错误率降低超过30%。Simple Real AdaBoost与Practical Real AdaBoost的效果几乎一样，实际使用时可直接使用Simple Real AdaBoost算法，不用顾虑特殊边界情况($W_b^{jt}=0$)。而Gentle AdaBoost与 Real AdaBoost效果几乎一样。测试错误率方差都很小，说明算法是稳定的。

5 总　结

本文把分类器组合问题转换为随机变量组合问题，以此证明了Real AdaBoost算法中置信度选取的科学性，同时还证明了可以采用加权组合来改进Real AdaBoost并得到近似最佳组合系数。还得到了Gentle AdaBoost算法的误差估计，基于Bayes统计推断推广得到了多分类Real AdaBoost算法。对多分类问题中弱分类器可能出现的边界问题提出了解决办法。结论表明，只要正确使用，Real AdaBoost好于AdaBoost。

参　考　文　献

[1] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.

[2] SCHAPIRE R E, SINGER Y. Improved boosting algorithms using confidence-rated predictions[J]. Machine Learning, 1999, 37(3): 297-336.

[3] SCHAPIRE R E, FREUND Y, BARTLETT P, et al. Boosting the margin: a new explanation for the effectiveness of voting methods[J]. The Annals of Statistics, 1998, 26(5): 1651-1686.

[4] 钟向阳, 凌捷. 基于多阀值弱学习的AdaBboost检测器[J]. 计算机工程与应用, 2009, 45(19): 160-162.
ZHONG Xiang-yang, LING Jie. AdaBoost detector based on multiple thresholds for weak classifier[J]. Computer Engineering and Applications, 2009, 45(19): 160-162.

[5] 武勃, 黄畅, 艾海舟, 等. 基于连续AdaBoost算法的多视角人脸检[J]. 计算机研究与发展, 2005, 42(9): 1612-1621.
WU Bo, HUANG Chang, AI Hai-zhou, et al. A multi-view face detection based on real AdaBoost algorithm[J]. Journal of Computer Research and Development, 2005, 42(9): 1612-1621.

[6] 孙士明, 潘青, 纪友芳. 多阈值划分的连续AdaBoost人脸检测[J]. 计算机应用, 2009, 29(8): 2098-2100.
SUN Shi-ming, PAN Qing, JI You-fang. Real AdaBoost face detection method based on multi-threshold[J]. Journal of Computer Applications, 2009, 29(8): 2098-2100.

[7] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of IEEE Conference Computer Vision and Pattern Recongnition. Marriott, Hawaii: IEEE, 2001.

[8] VIOLA P, JONES M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.

[9] 梁路宏, 艾海舟, 徐光祐, 等. 人脸检测研究综述[J]. 计算机学报, 2002, 25(5): 449-458.
LIANG Lu-hong, AI Hai-zhou, XU Guang-you, et al. A survey of human face detection[J]. Chinese Journal of Computer, 2002, 25(5): 449-458.

[10] 付忠良. 关于AdaBoost有效性的分析[J]. 计算机研究与发展, 2008, 45(10): 1747-1755.
FU Zhong-liang. The effectiveness analysis of AdaBoost[J]. Journal of Computer Research and Development, 2008, 45(10): 1747-1755.

[11] 付忠良. 分类器线性组合的有效性和最佳组合问题的研究[J]. 计算机研究与发展, 2009, 46(7): 1206-1216.
FU Zhong-liang. Effictive property and best combination of classifiers linear combination[J]. Journal of Computer Research and Development, 2009, 46(7): 1206-1216.

[12] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Additive logistic regression: a statistical view of boosting[J]. Annals of Statistics, 2000, 28(2): 337-374.

编　辑　税　红