

改进的加权极速学习机

邢胜^{1,2} 王晓兰³ 赵士欣^{1,4} 赵彦霞^{1,5}

(河北大学管理学院 保定 071002)¹ (沧州师范学院计算机科学与工程学院 沧州 061001)²

(沧州职业技术学院信息工程系 沧州 061001)³ (石家庄铁道大学数理系 石家庄 050043)⁴

(河北经贸大学信息技术学院 石家庄 050061)⁵

摘要 针对加权极速学习机人为固定权重可能会错失更优权重的问题,提出了改进的加权极速学习机。该方法的多数类的初始权重设为1,使用多数类与少数类样例数的比值作为少数类的初始权重,然后通过多数类或者少数类中添加权重调节因子,从缩小和扩大两个方向去调节权重,最后通过实验结果选出最优的权重。实验分别使用原加权极速学习机、其他权重的极速学习机和新方法在改进的UCI数据集上进行比较。结果表明新方法无论是在F-measure还是G-mean上都要优于其他加权极速学习机。

关键词 不平衡学习,加权极速学习机,代价敏感学习,单隐层前馈网络

中图分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.04.057

Improved Weighted Extreme Learning Machine

XING Sheng^{1,2} WANG Xiao-lan³ ZHAO Shi-xin^{1,4} ZHAO Yan-xia⁵

(College of Management, Hebei University, Baoding 071002, China)¹

(College of Computer Science and Engineering, Cangzhou Normal University, Cangzhou 061001, China)²

(Department of Information Engineering, Cangzhou Technical College, Cangzhou 061001, China)³

(Department of Mathematics and Physics, Shijiazhuang Tiedao University, Shijiazhuang 050043, China)⁴

(College of Information and Technology, Hebei University of Economics and Business, Shijiazhuang 050061, China)⁵

Abstract If the weight of the original weighted extreme learning machine is fixed artificially, the more optimal weight may be missed. Aiming at this problem, an improved weighted extreme learning machine was proposed. The new method uses the ratio of the sample number of different classes as the initial weight. The weight is adjusted by the weight adjustment factor from two directions of reducing and enlarging the weight ratio. Finally, the optimal weights are selected by the experimental results. The experiments were carried out on the transformed UCI data set using the original weighted extreme learning machine, the weighted extreme learning machine with other weights and the new method respectively. The experimental results indicate that the improved weighted extreme learning machine has better classification performance.

Keywords Imbalanced learning, Weighted extreme learning machine, Cost sensitive learning, Single hidden layer feed-forward networks

1 引言

不平衡数据在现实生活中(如网络入侵检测^[1]、医疗诊断^[2]、反垃圾邮件^[3]等领域)大量存在,而传统的机器学习方法往往偏重多数类,忽略少数类,所以针对不平衡数据的学习成为研究热点。目前不平衡学习主要包括集成学习^[4]、重采样技术^[5-6]和代价敏感学习^[7]。其中代价敏感学习(Cost Sen-

sitive Learning, CSL)根据不同的错分类型给予不同的惩罚代价,以总体误分代价最小化为目的来设计分类器。近年来多种代价敏感学习方法被提出,如Etheridge等^[8]提出了概率神经网络的代价敏感学习算法, Fu等^[9]提出了多分类问题代价敏感AdaBoost算法, Masnadi等^[10]提出了代价敏感的SVM算法等, 2013年, Zong Wei-wei等^[11]提出了加权极速学习机, 它也是一种代价敏感学习方法, 虽然该方法取得了较好的实

到稿日期:2016-11-07;返修日期:2017-02-16 本文受国家自然科学基金资助项目(71371063, 61170040, 61672205), 河北省人力资源社会保障科研合作课题(JRSHZ-2016-07038)资助。

邢胜(1982—),男,博士生,讲师,CCF学生会员,主要研究方向为机器学习, E-mail: cssxing@126.com; 王晓兰(1983—),女,硕士,讲师,主要研究方向为机器学习、模式识别; 赵士欣(1978—),女,博士生,讲师,主要研究方向为机器学习; 赵彦霞(1970—),女,博士生,讲师,主要研究方向为机器学习。

验结果,但是固定权重的方法缺乏灵活性,也错过了更好的权重选择。针对该问题,本文提出了基于添加权重调节因子的两种方法,首先以多数类与少数类的样例数比值作为少数类的初始权重,1为多数类的初始权重。方法1在少数类的权重上添加了调节因子,多数类的权重保持不变;方法2固定少数类的权重,给多数类的权重添加调节因子。将两种方法和原始的加权极速学习机、其他权重极速学习机进行实验比较,实验结果表明改进的加权极速学习机的整体分类效果更好。

2 加权极速学习机

Zong等在极速学习机的基础上提出了加权的极速学习机。极速学习机^[12-13]是由黄广斌提出的一种随机赋权前馈神经网络,因为该方法不用迭代,对单层前馈神经网络的输入层权重和偏置随机赋值,通过求解广义逆的方法直接计算出隐含层到输出层的权重,因此极速学习机的学习速度非常快,被广泛研究和应用。原极速学习机最优化问题如式(1)所示。

$$\text{Minimize: } \frac{1}{2} \|\beta\|^2 + C \frac{1}{2} \sum_{i=1}^N \|\xi_i\|^2 \quad (1)$$

$$\text{s. t. } h(x_i)\beta = t_i^T - \xi_i^T, i=1, \dots, N$$

其中, β 是隐含层到输出层的权重, C 是正则化参数,用来权衡训练误差的最小化和边际距离的最大化, $\xi_i = [\xi_{i1}, \dots, \xi_{im}]^T$ 是 m 个输出节点关于训练样例 x_i 的训练误差向量, $h(x) = [h_1(x), \dots, h_L(x)]$ 是输入样例 x 的隐含层(有 L 个隐含节点)输出, t_i 是样例 x_i 的目标向量, N 是总的样例个数。

而加权极速学习机在式(1)的基础上增加了权重 W ,从而得到了最优化问题(2)。

$$\text{Minimize: } \frac{1}{2} \|\beta\|^2 + CW \frac{1}{2} \sum_{i=1}^N \|\xi_i\|^2 \quad (2)$$

$$\text{s. t. } h(x_i)\beta = t_i^T - \xi_i^T, i=1, \dots, N$$

文献[11]中权重矩阵 W 对角线上的元素 W_{ii} 由式(3)人为设定, W 的其他元素均设为0。

$$W_{ii} = 1/\text{num}(c_j), \text{ when } t_i = c_j, i=1, \dots, N; j=1, \dots, m \quad (3)$$

其中, c_j 是第 j 个类别, $\text{num}(c_j)$ 是类别 c_j 的样例个数, N 同上, m 是类别个数。

最近的研究表明该方法可以应用于污水处理在线故障诊断^[14]、变压器故障的诊断^[15]、移动医疗保健^[16]等不平衡数据领域,并取得了较好的效果。虽然该方法的应用较多,但是针对该方法的固定权重的改进较少,杨泽平^[17]对文献[11]中的权重进行了修改,当类别 $m=2$ 时,权值如下:

$$W_{ii} = 1 - \frac{\text{num}(c_j)}{\sum_{j=1}^m \text{num}(c_j)}, j=1, \dots, m \quad (4)$$

实验结果显示采用方法所得的分类性能有所提高,但是权重设定单一,如何设定合适的权值使得分类效果达到最好还没有相应的理论依据。本文拟添加调节因子,使权重在一定范围内合理变化,通过实验结果来选取最适合的权重。

3 改进的加权极速学习机

由于加权极速学习机的权重是人为设定的固定值,其结

果不一定能达到最好的效果。而改进的加权极速学习机(Improved Weighted Extreme Learning Machine, IWELM)通过添加权重调节因子矩阵 Φ ,使得权重的设定由实验结果确定,相较于单一固定值更为合理。相似于加权极速学习机求解最优化问题的情形,改进的方法在最优化问题(2)的基础上添加了权重调节因子矩阵 Φ :

$$\text{Minimize: } \frac{1}{2} \|\beta\|^2 + C(W+\Phi) \frac{1}{2} \sum_{i=1}^N \|\xi_i\|^2$$

$$\text{s. t. } h(x_i)\beta = t_i^T - \xi_i^T, i=1, \dots, N \quad (5)$$

为求解式(5),引入拉格朗日函数:

$$L(\beta, \alpha, \xi) = \frac{1}{2} \|\beta\|^2 + C(W+\Phi) \frac{1}{2} \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N \alpha_i (h(x_i)\beta - t_i + \xi_i) \quad (6)$$

其中, α_i 是拉格朗日乘子。然后对变量 (β, ξ, α_i) 分别求偏导并令其等于零。

$$\frac{\partial L_{D_{ELM}}}{\partial \beta} = 0 \rightarrow \beta = \sum_{i=1}^N \alpha_i h(x_i)^T = H^T \alpha \quad (7)$$

$$\frac{\partial L_{D_{ELM}}}{\partial \xi_i} = 0 \rightarrow \alpha_i = C(W+\Phi)\xi_i, i=1, \dots, N \quad (8)$$

$$\frac{\partial L_{D_{ELM}}}{\partial \alpha_i} = 0 \rightarrow h(x_i)\beta - t_i + \xi_i = 0, i=1, \dots, N \quad (9)$$

通过式(7)一式(9)可以求解出关于左伪逆或右伪逆的两个版本 β :

$$\beta = H^T \left(\frac{1}{C} + (W+\Phi)HH^T \right)^{-1} (W+\Phi)T, \text{ when } N < L \quad (10)$$

$$\beta = \left(\frac{1}{C} + H^T(W+\Phi)H \right)^{-1} H^T(W+\Phi)T, \text{ when } L < N \quad (11)$$

其中, H 是隐含层的输出矩阵, T 是目标矩阵。

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix}, T = \begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix} \quad (12)$$

在权重的初始设定上,使用不同类别样例个数的比值来替代原来的权重。原方法的多数类权重是 $W_{\text{Max}} = 1/\text{num}(c_{\text{Max}})$,少数类的权重是 $W_{\text{Min}} = 1/\text{num}(c_{\text{Min}})$ ($\text{num}(c_{\text{Max}})$ 表示多数类的样例个数, $\text{num}(c_{\text{Min}})$ 表示少数类的样例个数)。为了使权重调节得更合理,将多数类的权重 W'_{Max} 设定为1,少数类的权重 W'_{Min} 设定为多数类与少数类的比值,即:

$$\frac{W'_{\text{Max}}}{W'_{\text{Min}}} = \frac{1}{\text{num}(c_{\text{Max}})/\text{num}(c_{\text{Min}})} \quad (13)$$

设 $\text{ratio} = \text{num}(c_{\text{Max}})/\text{num}(c_{\text{Min}})$,式(13)可以转化为式(14):

$$\frac{W'_{\text{Max}}}{W'_{\text{Min}}} = \frac{1}{\text{ratio}} \quad (14)$$

于是新方法中的 W'_{ii} 设定如下:

$$W'_{ii} = \begin{cases} 1, & \text{when } t_i = c_{\text{Max}} \\ \text{ratio}, & \text{when } t_i = c_{\text{Min}} \end{cases} \quad (15)$$

对权重调节因子矩阵 Φ 的对角线上元素 Φ_{ii} ($i=1, \dots, N, N$ 是样例个数)采用两种方法进行调节, Φ 的其他元素设为0。

第一种方法(IWELM1):固定多数类的权重,只对少数类的权重做适当调节,因为少数类的权重较大,所以调节因子从-1变化到1,步长为0.1,顺序选取。

$$\Phi_{ii} = \begin{cases} 0, & \text{when } t_i = C_{Max} \\ \{\varphi | \varphi \in (-1, -0.9, \dots, 1)\}, & \text{when } t_i = C_{Min} \end{cases} \quad (16)$$

式(16)加上新设定的权重式(15)即得式(17):

$$W_{ii} + \Phi_{ii} = \begin{cases} 1, & \text{when } t_i = C_{Max} \\ ratio + \varphi, & \text{when } t_i = C_{Min} \end{cases} \quad (17)$$

第二种方法(IWELM2):少数类的权重保持不变,改变多数类的权重,因为多数类的原来的权重是1,所以调节因子不能设定得太大,调节因子从-0.5变化到0.5,步长为0.05,同样是顺序选取。相较于方法1,方法2对权重的调节幅度更大,因为当调节因子 $\varphi=0.5$ 或 $\varphi=-0.5$ 时,相当于类别权重之间的比值缩小了一半或者扩大了2倍,从而使权重的选择面更广。

$$\Phi_{ii} = \begin{cases} \{\varphi | \varphi \in (-0.5, -0.45, \dots, 0.5)\}, & \text{when } t_i = C_{Max} \\ 0, & \text{when } t_i = C_{Min} \end{cases} \quad (18)$$

式(18)加上式(15)得到式(19):

$$W_{ii} + \Phi_{ii} = \begin{cases} 1 + \varphi, & \text{when } t_i = C_{Max} \\ ratio, & \text{when } t_i = C_{Min} \end{cases} \quad (19)$$

将式(17)和式(19)分别代入式(10)或者式(11),就得到了两种方法中 β 的求解公式。

4 实验及结果分析

4.1 实验设计

实验在改造的4个UCI数据集和1个平衡数据集上进行。对于多类数据集,保留其中一类,将其余的类别归为另一类,从而形成不平衡的两类数据集,其中4个改造的数据集为ecoli1, haberman1, yeast1, glass1。除此之外,还包括类别平衡的banana数据集,以此来检验新方法对于平衡数据集是否同样有效。实验中所用数据集的具体信息如表1所列(类别不平衡比为多数类样例个数与少数类样例个数的比值)。

表1 实验的数据集信息

数据集	属性个数	训练集 样例个数	测试集 样例个数	类别 不平衡比
ecoli1	7	268	68	3.3934
haberman1	3	244	62	2.8125
yeast1	8	757	190	30.5417
glass1	9	171	43	1.8500
banana	2	800	200	1

实验对比了原始的加权极速学习机(以下简称WELM)、杨泽平修改的加权极速学习机(以下简称Y_WELM)以及本文提出的两种改进的极速学习机IWELM1和IWELM2。为了方便比较,实验中隐含层节点数统一设定为100,激活函数采用sigmoid函数,C在 $\{2^{-50}, 2^{-49}, \dots, 2^{50}\}$ 中依次取值,且每次运行原始加权极速学习机时保留输入层到隐含层的权重和偏置,并将其应用到Y_WELM和新方法的IWELM1和IWELM2中,由此得到的结果仅与权重的调节有关,最后根

据实验结果找出最合适的权重。

传统的评价方法采用准确率来度量分类器的分类效果,但是该评价标准往往侧重多数类的分类表现,忽略了少数类,所以不平衡学习的评价标准一般采用F-measure^[18]和G-mean^[19]来对多数类和少数类的分类效果进行整体评价。在介绍非平衡的评价标准时,需要先引入混淆矩阵,如表2所列。

表2 混淆矩阵

	预测为正类	预测为负类
实际为正类	TP	FN
实际为负类	FP	TN

F-measure 的计算公式:

$$F\text{-measure} = \frac{(1 + \beta^2) * recall * precision}{\beta^2 * recall + precision} \quad (20)$$

其中, β 是可调参数,一般设为1。

$$precision = \frac{TP}{TP + FP} \quad (21)$$

$$recall = \frac{TP}{TP + FN} \quad (22)$$

G-mean 的计算公式:

$$G\text{-mean} = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{TN + FP}} \quad (23)$$

因为F-measure侧重于正类(少数类)的分类效果,而G-mean一般用来评价分类器对多数类和少数类的整体分类效果,所以本文使用两种评价指标对多个分类器进行评价。

4.2 实验结果分析

从表3—表6的实验结果中可以看出通过IWELM1和IWELM2得到的F-measure值、G-mean值都要高于原来的加权极速学习机得到的。对比IWELM1和IWELM2两种方法的实验结果可知,在F-measure值方面,IWELM1在1个数据集(haberman1)上略0,在1个数据集(glass1)上与IWELM2持平,在3个数据集(ecoli1, yeast1, banana)上低于IWELM2;在G-mean值方面,IWELM1在2个数据集(haberman1和glass1)上略高,在1个数据集(ecoli1)上与IWELM2持平,在2个数据集(yeast1和banana)上低于IWELM2方法(表中粗体数据为两种方法中的较大值,有下划线的数据表示IWELM1和IWELM2中的较大值)。

表3 IWELM1和WELM的F-measure比较

数据库	WELM		IWELM1		
	C	F-measure	C	φ	F-measure
ecoli1	$2^{\wedge}6$	0.8107	$2^{\wedge}(-4)$	-0.9	0.8267
haberman1	$2^{\wedge}5$	0.5901	$2^{\wedge}(-1)$	-0.2	0.5957
yeast1	$2^{\wedge}16$	0.2120	$2^{\wedge}(-11)$	-1.0	0.2392
glass1	$2^{\wedge}9$	0.5946	$2^{\wedge}2$	0.1	0.6011
banana	$2^{\wedge}21$	0.9611	$2^{\wedge}39$	0.7	0.9639

表4 IWELM2和WELM的F-measure比较

数据库	WELM		IWELM2		
	C	F-measure	C	φ	F-measure
ecoli1	$2^{\wedge}6$	0.8107	$2^{\wedge}(-4)$	0.40	0.8280
haberman1	$2^{\wedge}5$	0.5901	$2^{\wedge}(-1)$	0.10	0.5944
yeast1	$2^{\wedge}16$	0.2120	$2^{\wedge}(-9)$	0.15	0.2571
glass1	$2^{\wedge}9$	0.5946	$2^{\wedge}2$	-0.05	0.6011
banana	$2^{\wedge}21$	0.9611	$2^{\wedge}18$	-0.50	0.9700

表5 IWELM1和WELM的G-mean比较

数据库	WELM		IWELM1		
	C	G-mean	C	φ	G-mean
ecoli1	2 [^] 6	0.8916	2 [^] (-3)	-0.4	<u>0.8968</u>
haberman1	2 [^] 5	0.7112	2 [^] (-4)	0.6	<u>0.7148</u>
yeast1	2 [^] 15	0.7516	2 [^] 6	-1.0	<u>0.7533</u>
glass1	2 [^] 9	0.6183	2 [^] 37	0.2	<u>0.6243</u>
banana	2 [^] 21	0.9610	2 [^] 39	0.7	<u>0.9634</u>

表6 IWELM2和WELM的G-mean比较

数据库	WELM		IWELM2		
	C	G-mean	C	φ	G-mean
ecoli1	2 [^] 6	0.8916	2 [^] (-4)	0.20	<u>0.8968</u>
haberman1	2 [^] 5	0.7112	2 [^] (-3)	-0.20	<u>0.7128</u>
yeast1	2 [^] 15	0.7516	2 [^] 4	0.20	<u>0.7571</u>
glass1	2 [^] 9	0.6183	2 [^] 2	-0.05	<u>0.6204</u>
banana	2 [^] 21	0.9610	2 [^] 18	-0.50	<u>0.9694</u>

表7—表10对比了Y_WELM, IWELM1和IWELM2的实验结果。从实验结果可以看出, Y_WELM和WELM的分类效果相差不大, 而本文提出的IWELM1和IWELM2方法均要优于Y_WELM方法。

表7 IWELM1和Y_WELM的F-measure比较

数据库	Y_WELM		IWELM1		
	C	F-measure	C	φ	F-measure
ecoli1	1	0.8096	2 [^] (-4)	-0.9	<u>0.8267</u>
haberman1	2 [^] 1	0.5880	2 [^] (-1)	-0.2	<u>0.5957</u>
yeast1	2 [^] 11	0.2123	2 [^] (-11)	-1.0	<u>0.2392</u>
glass1	2 [^] 4	0.5946	2 [^] 2	0.1	<u>0.6011</u>
banana	2 [^] 14	0.9602	2 [^] 39	0.7	<u>0.9639</u>

表8 IWELM2和Y_WELM的F-measure比较

数据库	Y_WELM		IWELM2		
	C	F-measure	C	φ	F-measure
ecoli1	1	0.8096	2 [^] (-4)	0.40	<u>0.8280</u>
haberman1	2 [^] 1	0.5880	2 [^] (-1)	0.10	<u>0.5944</u>
yeast1	2 [^] 11	0.2123	2 [^] (-9)	0.15	<u>0.2571</u>
glass1	2 [^] 4	0.5946	2 [^] 2	-0.05	<u>0.6011</u>
banana	2 [^] 14	0.9602	2 [^] 18	-0.50	<u>0.9700</u>

表9 IWELM1和Y_WELM的G-mean比较

数据库	Y_WELM		IWELM1		
	C	G-mean	C	φ	G-mean
ecoli1	1	0.8931	2 [^] (-3)	-0.4	<u>0.8968</u>
haberman1	2 [^] 1	0.7087	2 [^] (-4)	0.6	<u>0.7148</u>
yeast1	2 [^] 11	0.7525	2 [^] 6	-1.0	<u>0.7533</u>
glass1	2 [^] 4	0.6183	2 [^] 37	0.2	<u>0.6243</u>
banana	2 [^] 14	0.9600	2 [^] 39	0.7	<u>0.9634</u>

表10 IWELM2和Y_WELM的G-mean比较

数据库	Y_WELM		IWELM2		
	C	G-mean	C	φ	G-mean
ecoli1	1	0.8931	2 [^] (-4)	0.20	<u>0.8968</u>
haberman1	2 [^] 1	0.7087	2 [^] (-3)	-0.20	<u>0.7128</u>
yeast1	2 [^] 11	0.7525	2 [^] 4	0.20	<u>0.7571</u>
glass1	2 [^] 4	0.6183	2 [^] 2	-0.05	<u>0.6204</u>
banana	2 [^] 14	0.9600	2 [^] 18	-0.50	<u>0.9694</u>

整体来看, IWELM2的分类效果比IWELM1略好, 其原因可能是IWELM2在多数类上的权重为1, 少数类的权重为ratio的基础上只改变多数类的权重, 使多数类和少数类权重之间的比值变化幅度比IWELM1大, 从而有更好的权重可供选择。而个别数据集在IWELM1上的效果更好, 可能的原因是只改变少数类的权重, 权重比的变化幅度较小, 但是变化得更细, 从而在合理的区间能够挑选出较好的权重。此外, 在平衡数据集banana上应用新方法也得到了更好的实验结果, 更

验证了新方法不仅适用于不平衡数据, 对于平衡数据集同样有效。

从以上实验结果中可以看到, 增加权重调节因子后确实能够找到更好的分类结果。那么权重调节因子的取值范围和步长选择是如何影响最后结果的呢?

为了更直观地了解调节因子对实验结果的影响, 固定参数C=1, φ 分别在区间[-0.5, 0.5]中依次取值, 步长为0.05。同时比较WELM和Y_WELM和IWELM2 3种方法的实验结果。以glass1数据集为例: WELM方法中多数类的权重为1/111, 少数类的权重为1/60; Y_WELM方法中多数类的权重为60/171, 少数类的权重为111/171; 而IWELM2方法中的多数类的权重为1+ φ , 少数类的权重为1.85。因为WELM和Y_WELM方法中不存在调节因子, 为了更好地进行比较, 在 $\varphi=0$ 时分别标记glass1应用WELM和Y_WELM方法所得的G-mean值。从图1中可以看出IWELM2方法在 $\varphi=0.05$ 时取到了最优G-mean值, 优于前两种方法。另外, φ 的取值区间也不宜过大, 当 $\varphi=-0.5$ 或 $\varphi=0.5$ 时, 得到的G-mean值均显著下降。观察实验发现, 对于大部分数据集, IWELM2方法中的 φ 在[-0.5, 0.5]的取值区间里能够获得最优的G-mean值。

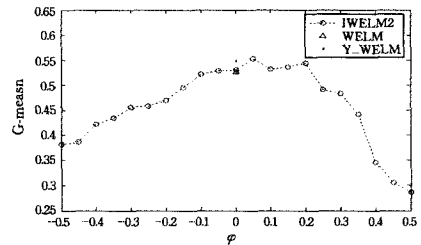


图1 对glass1数据集应用3种方法的实验结果

接下来讨论步长大小对结果的影响。同样以glass1数据集应用IWELM2方法为例。为了方便比较, 仍然固定系数C为1, 然后观察步长的大小对结果的影响。 φ 在区间[-0.5, 0.5]中取值, 其中的步长分别设置为0.01, 0.05和0.1。实验结果表明, 当步长设置为0.01时, 区间[-0.5, 0.5]有101个值, 虽然得到的结果更精确, 但是需要计算的数据过多; 当步长设置为0.1时, 在 $\varphi=0.2$ 时得到了比 $\varphi=0$ 时更高的G-mean值, 但却低于步长设置为0.05和0.01时得到的最优G-mean值。当步长为0.05时, 区间[-0.5, 0.5]有21个值, 也得到了与步长为0.01时相同的最优G-mean值, 并且计算的数据量在可接受的范围内, 如图2所示。通过实验观察发现, 步长设置太大容易错过最优值且与最优值相差较多; 如果步长设置太小, 又会增加更多的运行时间, 为了在可接受的运行时间里得到最优值或接近最优值的结果, 设置的步长使调节因子在区间中能够取到21个值。

此外, 之所以同时提出了IWELM1和IWELM2两种方法, 也是为了尽可能地接近最优结果。因为IWELM1中权重之间的比值变化较小, 可以在较小范围内寻找最优结果; 而IWELM2中权重之间的比值变化较大, 可以在较大范围内寻找最优结果, 因此两种方法可以互为补充。

IWELM2方法中 φ 设置不同步长的实验结果如图2所示。

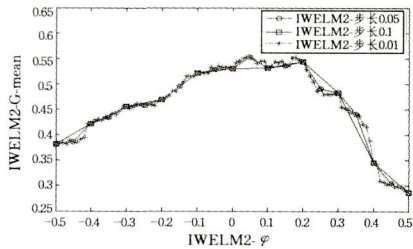
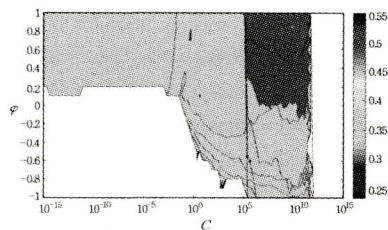
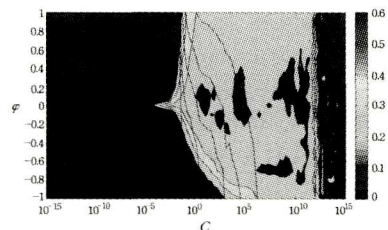


图 2 IWELM2 方法中 φ 设置不同步长的实验结果(glass1 数据集)

如果不固定参数 C , 那么调节因子 φ 和参数 C 如何影响 IWELM1 和 IWELM2 的分类性能? 同样以 glass1 数据集为例, 将 IWELM1 和 IWELM2 方法分别得到的 F-measure 和 G-mean 值(IWELM1 的调节因子的取值区间是 $[-1, 1]$, 步长是 0.1; IWELM2 的调节因子的取值区间是 $[-0.5, 0.5]$, 步长是 0.05; 参数 C 从 $\{2^{-50}, 2^{-49}, \dots, 2^{50}\}$ 中依次取值)投影到调节因子 φ 和参数 C 的平面上, 得到了 F-measure 和 G-mean 值的等高线。

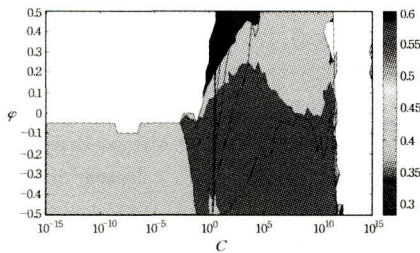


(a)

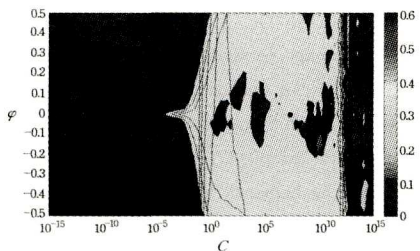


(b)

图 3 IWELM1 方法在 glass1 数据集上的测试结果(F-measure(a) 和 G-mean(b)在调节因子 φ 和参数 C 平面上的等高线)



(a)



(b)

图 4 IWELM2 方法在 glass1 数据集上的测试结果(F-measure(c) 和 G-mean(d)在调节因子 φ 和参数 C 平面上的等高线)

如图 3 和图 4 所示, 结合表 3—表 6 中两种方法的数据, 可以观察到实验结果中最优的 F-measure 和 G-mean 值所对应的调节因子 φ 和参数 C 均落在了等高线图中部黑色的极值区域内, 也印证了 IWELM1 和 IWELM2 方法的有效性; 而且从图 3 和图 4 中也可以看出, F-measure 和 G-mean 的极值点(即中部黑色区域)分布并不规律, 且分布不均匀, 所以通过实验的方法寻找最优值是目前可行的方案。

结束语 加权极速学习机虽然适合处理不平衡数据, 但是单一固定权重的方法存在遗漏更优权重的问题。改进的极速学习机提供了两种方法来解决此问题, 首先改用不同类别样例数的比值作为初始权重, 然后分别在多数类和少数类权重上添加调节因子, 最后从实验结果中选择出最优的调节因子形成最终权重。实验结果验证了新方法的有效性, 提高了原方法的 F-measure 值和 G-mean 值, 所以相较于加权极速学习机, 改进的加权极速学习机更适合处理不平衡数据, 同时对于平衡数据集同样有效。

参 考 文 献

- [1] THOMAS C. Improving intrusion detection for imbalanced network traffic[J]. Security & Communication Networks, 2013, 6(6):309-324.
- [2] DROSOU K, GEORGIU S, KOUKOUVINOS C, et al. Support vector machines classification on class imbalanced data: a case study with real medical data[J]. Journal of Data Science, 2014, 12(4):143-155.
- [3] FAWCETT T. "In vivo" spam filtering: a challenge problem for data mining[J]. SIGKDD Explorations, 2003, 5(2):140-148.
- [4] ZHANG C X, ZHANG J S. A Survey of Selective Ensemble Learning Algorithms[J]. Chinese Journal of Computers, 2011, 34(8):1399-1410. (in Chinese)
张春霞, 张讲社. 选择性集成学习算法综述[J]. 计算机学报, 2011, 34(8):1399-1410.
- [5] TOMEK I. Two modifications of CNN[J]. IEEE Transactions on Systems, Man and Communications, 1976, 6(11):769-772.
- [6] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1):321-357.
- [7] SUN Y, WONG A K C, KAMEL M S. Classification of imbalanced data: a review[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23(4):687-719.
- [8] ETHERIDGE H L, SRIRAM R S, HSU H Y K. A comparison of selected artificial neural networks that help auditors evaluate client financial viability[J]. Decision Sciences, 2007, 31(2):531-550.
- [9] FU Z L. Cost-sensitive AdaBoost Algorithm for Multi-class Classification Problems[J]. Acta Automatica Sinica, 2011, 37(8):973-983. (in Chinese)
付忠良. 多分类问题代价敏感 AdaBoost 算法[J]. 自动化学报, 2012, 37(8):973-983
- [10] MASNADI H, VASCONCELOS N, IRANMEHR A. Cost-sensitive support vector machines[J]. Journal of Machine Learning Research, 2015, 1(1):1-26.
- [11] ZONG W W, HUANG G B, CHEN Y. Weighted extreme learning machine for imbalance learning[J]. Neurocomputing, 2013, 101(3):229-242.

- [12] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. *Neurocomputing*, 2006, 70(1-3): 489-501.
- [13] HUANG G B, WANG D H, LAN Y. Extreme learning machines: a survey [J]. *International Journal of Machine Learning & Cybernetics*, 2011, 2(2): 107-122.
- [14] XU Y G, DENG W K, CHEN L D. Online fault diagnosis in wastewater treatment process by kernel-based weighted extreme learning machine [J]. *CIESC Journal*, 2016, 67(9): 3817-3825. (in Chinese)
许玉格, 邓文凯, 陈立定. 基于核函数的加权极限学习机污水处理在线故障诊断 [J]. *化工学报*, 2016, 67(9): 3817-3825.
- [15] YU B J, ZHU Y L. Transformer fault diagnosis using weighted extreme learning machine [J]. *Computer Engineering & Design*, 2013, 34(12): 4340-4344. (in Chinese)
遇炳杰, 朱永利. 加权极限学习机在变压器故障诊断中的应用 [J]. *计算机工程与设计*, 2013, 34(12): 4340-4344.
- [16] GAO X, CHEN Z, TANG S, et al. Adaptive weighted imbalance learning with application to abnormal activity recognition [J]. *Neurocomputing*, 2015, 173: 1927-1935.
- [17] YANG Z P. Neural Network Based Classification Methods for Imbalanced Datasets [D]. Shanghai: South China University of Technology, 2015. (in Chinese)
杨泽平. 基于神经网络的不平衡数据分类方法研究 [D]. 上海: 华东理工大学, 2015.
- [18] DASKALAKI S, KOPANAS I, AVOURIS N. Evaluation of classifiers for an uneven class distribution problem [J]. *Applied Artificial Intelligence*, 2006, 20(5): 381-417.
- [19] KUBAT M, MATWIN S. Addressing the curse of imbalanced training sets: one-sided selection [C] // *Proceedings of the Fourteenth International Conference on Machine Learning*. 2000: 179-186

(上接第 255 页)

歧模型上来验证本文所提方法的有效性, 并且测试其他模型的词义消歧性能, 持续改善有监督词义消歧模型的性能, 力争达到实用的目的。

参 考 文 献

- [1] CHAN Y S, NG H T. Scaling up word sense disambiguation via parallel texts [C] // *AAAI*. 2005: 1037-1042.
- [2] WANG R Q, KONG F S. Research on unsupervised word sense disambiguation [J]. *Journal of Software*, 2009, 20(8): 2138-2152. (in Chinese)
王瑞琴, 孔繁胜. 无监督词义消歧研究 [J]. *软件学报*, 2009, 20(8): 2138-2152.
- [3] NAVIGLI R. Word sense disambiguation: A survey [J]. *ACM Computing Surveys (CSUR)*, 2009, 41(2): 1-69.
- [4] YAROWSKY D. Unsupervised word sense disambiguation rivaling supervised methods [C] // *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1995: 189-196.
- [5] LEACOCK C, MILLER G A, CHODOROW M. Using corpus statistics and WordNet relations for sense identification [J]. *Computational Linguistics*, 1998, 24(1): 147-165.
- [6] MIHALCEA R, MOLDOVAN D I. A method for word sense disambiguation of unrestricted text [C] // *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. Association for Computational Linguistics, 1999: 152-158.
- [7] MIHALCEA R, MOLDOVAN D I. An automatic method for generating sense tagged corpora [C] // *AAAI/IAAI*. 1999: 461-466.
- [8] AGIRRE E, MARTINEZ D. Unsupervised WSD based on Automatically Retrieved Examples: The Importance of Bias [C] // *EMNLP*. 2004: 25-32.
- [9] BRODY S, LAPATA M. Good neighbors make good senses: Exploiting distributional similarity for unsupervised WSD [C] // *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008: 65-72.
- [10] BROWN P F, PIETRA S A D, Pietra V J D, et al. Word-sense disambiguation using statistical methods [C] // *Proceedings of the 29th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1991: 264-270.
- [11] HE Q Z, WANG H F. Chinese Word Sense Disambiguation Based on Maximum Entropy Model with Feature Selection [J]. *Journal of Software*, 2010, 21(6): 1287-1295. (in Chinese)
何径舟, 王厚峰. 基于特征选择和最大熵模型的汉语词义消歧 [J]. *软件学报*, 2010, 21(6): 1287-1295.
- [12] LU W P, HUANG H Y. Word sense disambiguation based on dependency fitness with automatic knowledge acquisition [J]. *Journal of Software*, 2013, 24(10): 2300-2311. (in Chinese)
鹿文鹏, 黄河燕. 基于依存适配度的知识自动获取词义消歧方法 [J]. *软件学报*, 2013, 24(10): 2300-2311.
- [13] LU W P, HUANG H Y, WU H. Word Sense Disambiguation with Graph Model Based on Domain Knowledge [J]. *Acta Automatica Sinica*, 2014, 40(12): 2836-2850. (in Chinese)
鹿文鹏, 黄河燕, 吴昊. 基于领域知识的图模型词义消歧方法 [J]. *自动化学报*, 2014, 40(12): 2836-2850.
- [14] YANG Z Z, HUANG H Y. Graph Based Word Sense Disambiguation Method Using Distance Between Words [J]. *Journal of Software*, 2012, 23(4): 776-785. (in Chinese)
杨陟卓, 黄河燕. 基于词语距离的网络图词义消歧 [J]. *软件学报*, 2012, 23(4): 776-785.
- [15] YANG Z Z, HUANG H Y. WSD Method Based on Heterogeneous Relation Graph Network [J]. *Journal of Computer Research and Development*, 2013, 50(2): 437-444. (in Chinese)
杨陟卓, 黄河燕. 基于异构关系网络图的词义消歧研究 [J]. *计算机研究与发展*, 2013, 50(2): 437-444.
- [16] ESCUDERO G, MÁRQUEZ L, RIGAU G. Naive Bayes and exemplar-based approaches to word sense disambiguation revisited [OL]. <http://www.cs.opc.edu/~escudero/wsd/00-ecai.pdf>.
- [17] JIN P, WU Y, YU S. Semeval-2007 task 05: Multilingual chinese-english lexical sample [C] // *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 2007: 19-23.
- [18] XING Y. SRCB-WSD: Supervised Chinese word sense disambiguation with key features [C] // *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 2007: 300-303.