

文章编号: 1001-0920(2012)03-0459-05

一种支持向量机的快速分类算法

张战成^{1,3}, 王士同^{1,2}, 邓赵红¹, 钟富礼²

(1. 江南大学 数字媒体学院, 江苏 无锡 214122; 2. 香港理工大学 电子计算学系, 香港; 3. 中国科学院 苏州纳米技术与纳米仿生研究所, 江苏 苏州 215123)

摘要: 鉴于传统支持向量机分类过程的计算量和支持向量的个数成正比, 为了提高分类决策的速度, 提出一种约简支持向量的快速分类算法, 该算法对原始的支持向量进行特定比例的模糊 C 均值聚类操作, 按照分类误差最小的原则构建最小线性二乘回归模型, 求解新的支持向量系数和决策函数的偏置. 人造数据集和标准数据集上的实验表明, 约简 50% 支持向量后, 可以在保持分类精度在无统计意义的明显损失的前提下, 使得分类速度提高 50%.

关键词: 分类; 支持向量机; 快速

中图分类号: TP391.4

文献标识码: A

A fast decision algorithm of support vector machine

ZHANG Zhan-cheng^{1,3}, WANG Shi-tong^{1,2}, DENG Zhao-hong¹, CHUNG Fu-li²

(1. School of Digital Media, Jiangnan University, Wuxi 214122, China; 2. Department of Computing, Hong Kong Polytechnic University, Hong Kong, China; 3. Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Science, Suzhou 215123, China. Correspondent: ZHANG Zhan-cheng, E-mail: cimszhang@163.com)

Abstract: The number of support vectors(SVs) of support vector machine(SVM) is usually large, which results in a substantially slower classification speed than many other approaches. The less SVs means the more sparseness and the higher classification speed. Therefore, an algorithm called FD-SVM is proposed, which employs FCM to cluster a dense SVs set to a sparse set. Then, aiming to minimize the classification gap between SVM and FD-SVM, a linear least square programming model is built for obtaining the optimal coefficients of the new sparse SVs. Experiments on toy and real-world data sets demonstrate that, after reducing 50% of SVs, an increase of 50% on the classification speed is achieved, while the performance of losing maintains a statistically insignificant level.

Key words: classification; support vector machine; fast

1 引言

基于结构风险最小化原则, Vapnik^[1]提出了支持向量机(SVM), 其目标是在 VC 维的意义下构造稀疏模型. 与其他学习机相比, SVM 具有良好的推广能力, 在处理非线性识别和小样本学习方面具有良好的特性. 标准 SVM 训练问题可以归结成一个带约束的凸二次规划(QP)问题, SVM 的分类问题仅与稀疏化后支持向量(SVs)及其对应的权重系数有关, 分类过程的计算量和支持向量的个数成正比. 通常支持向量的个数反映其稀疏性, 支持向量越少, 稀疏性越好^[2-3].

目前, 模式识别应用中计算能力的增长远远不及数据规模的增长. 对于一个具有 N 个样本的训练集, 求解 SVM 的 QP 问题的时间复杂度为 $O(N^3)$, 空

间复杂度为 $O(N^2)$, 所以 SVM 在处理大规模数据集时训练速度和分类速度较慢, 在一定程度上影响了 SVM 的推广应用. 针对训练速度慢的问题, Platt^[4]提出了时间复杂度为 $O(N) \sim O(N^{2.3})$ 的 SMO(sequential minimal optimization)算法, 该算法主要是针对二次规划中计算量和存储量过大的问题, 能够以渐近的局部最小值收敛到全局最小值. Suykens 等人^[5]提出了最小二乘支持向量机(LS-SVM), 采用最小二乘线性系统作为损失函数, 代替传统 SVM 的二次规划方法, 最终等价于求解 $N+1$ 维的线性方程组. 这些方法简化了计算的复杂性, 降低了 SVM 训练过程的复杂度, 但是在实际使用中, 尤其是一些如网络入侵检测、垃圾邮件检测等在线应用, 其训练过程可以作为一个采用

收稿日期: 2010-10-07; 修回日期: 2011-02-14.

基金项目: 国家自然科学基金项目(60903100, 61103128); 中央高校基本科研业务费专项资金项目(JUDCF09034).

作者简介: 张战成(1977-), 男, 博士生, 从事人工智能、模式识别的研究; 王士同(1964-), 男, 教授, 博士生导师, 从事人工智能、机器学习等研究.

增量学习算法的异步执行的后台任务. 这类应用对于分类决策的实时性要求高于训练的实时性要求, 而分类过程的速度取决于支持向量的个数.

沿着精简支持向量提高 SVM 分类效率这个方向, 研究人员提出了很多精简支持向量的算法. Schölkopf 等人^[6]提出了 ν -SVM, 通过调节参数 ν 来控制支持向量的个数. Burges 等人^[7-8]提出一种提高分类速度的方法, 该方法在分类函数中不使用支持向量, 而使用一个约简的向量集. 与标准 SVM 不同的是, 该向量集中的向量既不是训练样本也不是支持向量, 而是经过变换的特殊向量. 基于线性变换约简支持向量的思想, 刘向东等人^[9]提出的快速支持向量机分类算法通过适当变换特征空间中的内积矩阵, 进一步变换分类函数的形式, 以减少分类函数中的支持向量, 提高了分类速度, 同时通过变换矩阵保留其他支持向量的信息, 使全部支持向量的信息保留在分类函数中, 以保证在不损失分类精度的前提下通过减少支持向量提高分类速度. 但是如何找出合适的支持向量集划分需要不断的迭代, 每次迭代中求解变换矩阵的代价也是较大的. 方景龙等人^[10]通过减少错分间隔 ξ 的自由度、运用二次训练支持向量的方法来减少支持向量的数量, 其二次训练过程可以看成是改进了约束规则的 SVM 再训练的过程. 另一种思想是充分利用样本的几何结构信息约简支持向量, 文献[11]使用 KNN 和聚类方法预选取样本, 有效减少了训练样本数目, 显著加快了学习速度, 并在保证理想分类精度的同时提高了分类的速度.

本文主要研究标准的 SVM 训练得到分类模型后, 利用样本的几何结构信息约简支持向量, 在尽量减少分类精度损失的前提下, 高效地构造分类函数 (FD-SVM). 利用目前应用广泛的 FCM (fuzzy C-means) 聚类方法将原始 SVM 训练得到的 l_1 个支持向量聚集为 l_2 个 ($l_2 \ll l_1$), 以聚类的中心为新的“支持向量”, 然后在误差最小的原则下构造一个最小二乘线性回归问题, 求解出新的“支持向量”的权重系数, 使得 FD-SVM 的计算代价减小为原分类函数的 l_2/l_1 倍.

2 快速分类算法 FD-SVM

传统二分类支持向量机的分类通常由一个实值函数 $f: X \subseteq \mathbf{R}^n \mapsto \mathbf{R}$ 按照以下方式操作: 当 $f(x) \geq 0$ 时, 赋予输入 x 正类, 否则赋予负类. 核化情况下, 分类函数为

$$f(x) = \sum_{i=1}^{l_1} \alpha_i y_i K(s_i, x) + b.$$

其中: $s_i \in \mathbf{R}^n$ 为支持向量; $\alpha_i \in \mathbf{R}$ 为拉格朗日系数, 且 $\alpha_i > 0$; $y_i \in \{+1, -1\}$ 为对应的 s_i 的标签,

$\sum_{i=1}^{l_1} \alpha_i y_i = 0$; $b \in \mathbf{R}$ 为偏移量, $K(\cdot)$ 为核函数. 为了表示方便, 将 $\alpha_i y_i$ 直接改写为 α_i , 即

$$f(x) = \sum_{i=1}^{l_1} \alpha_i K(s_i, x) + b. \quad (1)$$

从式 (1) 可以看出, SVM 分类过程的计算量取决于 l_1 的大小. 传统的 SVM 分类函数中包含全部的支持向量, 而较多支持向量又导致分类时的计算量较大, 分类速度必然较慢, 所以提高分类速度的关键是减少支持向量的数量 l_1 . 定义新的快速分类函数为

$$g(x) = \sum_{j=1}^{l_2} \beta_j K(\acute{s}_j, x) + \acute{b}.$$

其中: $l_2 < l_1$, $\{\acute{s}_j\}_{j=1}^{l_2}$ 为新的“支持向量集”, $\acute{s}_j \in \mathbf{R}^n$, β_j 为对应的系数, \acute{b} 为新的偏移量.

标准 SVM 将训练样本集 $\{x_k\}_{k=1}^N$ 稀疏化为支持向量集 $\{s_i\}_{i=1}^{l_1}$, FD-SVM 则将支持向量集 $\{s_i\}_{i=1}^{l_1}$ 再稀疏化为新的支持向量集 $\{\acute{s}_j\}_{j=1}^{l_2}$. 其区别在于, SVM 的训练结果是从 N 个样本中选择 l_1 个合适的样本作为支持向量; 而 FD-SVM 再次训练的结果是进一步利用 $\{s_i\}_{i=1}^{l_1}$ 构造出新的更稀疏的支持向量集 $\{\acute{s}_j\}_{j=1}^{l_2}$. 在不影响分类精度的前提下, 如何构造 \acute{s}_j 并求解对应的 β_j 和 \acute{b} 是本文的关键. 考虑到 FCM 是一种直接方便的数据压缩方法, 选用 FCM 将支持向量聚为 l_2 类, 聚类中心为 \acute{s}_j , l_2/l_1 称为压缩率. 压缩率越小, 支持向量越少, 稀疏程度越高, 分类速度越快.

因为快速分类函数 $g(x)$ 要保证分类精度的损失最小, 所以对于训练集 $\{x_k\}_{k=1}^N$, 定义如下损失函数:

$$L = \min_{\beta, \acute{b}} \sum_{k=1}^N \|g(x_k) - f(x_k)\|^2 = \min_{\beta, \acute{b}} \sum_{k=1}^N \|L_k\|^2 = \min_{\lambda} \sum_{k=1}^N \|H_k \lambda - f_k\|^2 = \min_{\lambda} \|H \lambda - f\|^2. \quad (2)$$

其中

$$\lambda = [\beta_1, \beta_2, \dots, \beta_{l_2}, \acute{b}]^T,$$

$$f = [f(x_1), f(x_2), \dots, f(x_N)]^T.$$

λ 的系数矩阵为

$$H = [H_1, H_2, \dots, H_N]^T,$$

$$H_k = [K(\acute{s}_1, x_k), K(\acute{s}_2, x_k), \dots, K(\acute{s}_{l_2}, x_k), 1]^T,$$

$$k = 1, 2, \dots, N.$$

式 (2) 所示的线性最小二乘问题可以使用梯度下降法求解, 有

$$\frac{\partial L}{\partial \lambda_j} = 2 \sum_{k=1}^N L_k \frac{\partial L_k}{\partial \lambda_j} = 0,$$

$$\frac{\partial L_k}{\partial \lambda_j} = H_{kj}, \quad j = 1, 2, \dots, l_2 + 1.$$

即

$$\sum_{k=1}^N (H_k \lambda - f_k) H_{kj} = 0, j = 1, 2, \dots, l_2 + 1;$$

$$\sum_{k=1}^N H_{kj} \sum_{j=1}^{(l_2+1)} H_{kj} \lambda_j = \sum_{k=1}^N H_{kj} f_k.$$

整理为矩阵形式为 $H^T H \lambda = H^T f$. 最终求得

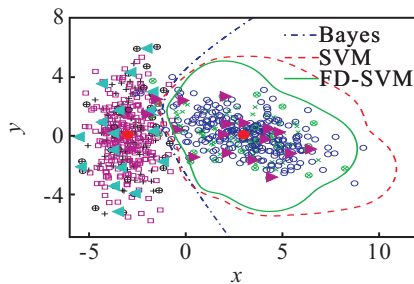
$$\lambda = (H^T H)^{-1} H^T f. \quad (3)$$

文献[12]详细讨论了式(3)的QR和SVD分解的求解方式,本文实验中使用matlab的lsqin函数QR方式求解.

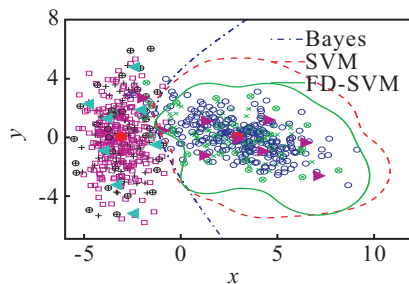
3 实验结果分析

3.1 人工数据集

为了直观地展示选用FCM聚类算法在几何上对于支持向量的压缩,构造如图1所示的数据集.



(a) 支持向量压缩率为50%



(b) 支持向量压缩率为25%

图1 模拟数据集分类结果

图1中,左边数据为正类,由中心为 $[-4, 0]$ 协方差为 $[1, 0; 0, 5]$ 的470个正态分布的数据点组成;右边

数据为负类,由中心为 $[4, 0]$ 协方差为 $[1, 0; 0, 5]$ 的470个正态分布的数据点逆时针旋转 $3\pi/8$ 组成,这样两类数据在几何上有部分交叉,可以用来检验分类算法的识别效果.每类数据随机划分为120个训练数据和250个测试数据. $+$ 和 \square 分别表示正类的训练数据和测试数据, \times 和 \circ 分别表示负类的训练数据和测试数据,在 $+$ 和 \times 上加标 \circ 的点分别表示正类和负类经SVM训练得到的支持向量, \blacktriangleleft 和 \blacktriangleright 分别表示正类和负类的支持向量聚类中心,即新的支持向量.

SVM高斯核为 $K(x_i, x_j) = \exp(-\sigma \|x_i - x_j\|^2)$,在训练集上使用10折交叉验证,利用 $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ 和 $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ 的间隔网格搜索^[13]得到核参数 $\sigma = 0.5$,惩罚系数 $C = 1$.

得到最优参数后,运行10次该实验,每次按照#train/#test为120/250的比例随机划分,最终分类精度取10次的平均值. Bayes分类器和标准SVM的平均分类精度分别为98.2%和98.0%,同时,FD-SVM按照(0.5, 0.25, 0.125)的压缩率取得的平均分类精度分别为(97.8%, 97.6%, 96.8%).可以看出,随着支持向量的减少,分类精度有所下降,但是当压缩率为0.5时FD-SVM和SVM的精度相当.由图1中的分类线也可以看出,FD-SVM取得了与SVM相当分类效果,FD-SVM分类精度的损失主要在类的边界线上.

3.2 真实数据集

为了进一步验证FD-SVM的分类性能,选取UCI标准数据集^[14]上的10套数据集,使用高斯核将FD-SVM与标准SVM作对比测试,各个数据集的规模见表1.实验中,压缩率 l_2/l_1 取50%,SVM的惩罚系数 C 和高斯核参数 γ 按照 $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ 和 $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ 的间隔网格搜索10折交叉验证取得最好平均精度时的数值.表1中为最优参数时10折交叉验证分类精度的均值和标准差以及10次分类累计消耗的CPU时间.为了检验FD-SVM和SVM的分类精度是否一致,在0.05的显著性水平下,

表1 UCI数据集相关参数和分类结果

data set	size	dim	γ	C	#SVs		test times/s		accuracy/%		t-test	
					l_1	l_2	SVM	FD-SVM	SVM	FD-SVM	t-value	H
breast	569	30	0.125	2	80.5±2.6	39.7±1.4	0.92	0.44	97.89±1.89	97.89±1.54	0.0119	0
glass	146	9	8	1	106.4±2.2	52.4±1.0	0.31	0.16	78.90±6.76	77.52±5.65	1.4989	0
heart-disease	225	13	0.5	0.5	98.3±4.2	48.6±2.0	0.47	0.17	88.06±5.08	88.06±5.08	-	0
iosphere	351	34	0.5	32768	173.9±2.0	86.5±1.0	1.22	0.59	94.01±4.18	89.74±8.27	2.5584	1
musk	416	167	0.125	32768	373.0±3.0	186.2±1.7	5.33	2.64	93.46±2.68	84.01±3.24	9.3989	1
parkinsons	195	22	2	32768	130.8±3.4	65.0±1.7	0.69	0.28	92.89±5.58	91.78±4.75	0.4706	0
pima	768	8	0.125	0.5	433.5±4.3	216.4±2.3	6.36	3.09	77.09±2.87	77.09±2.87	-	0
sonar	208	60	0.125	32768	136.7±2.7	67.8±1.3	0.66	0.31	89.45±8.69	84.68±6.86	2.5345	1
vote	435	16	0.125	8	86.0±2.2	42.5±1.1	0.70	0.34	94.94±3.06	94.02±3.41	1.1766	0
yeast	892	6	2	0.5	634.7±4.4	316.8±2.1	10.28	5.08	66.59±3.20	65.69±3.70	2.0622	0

对两种方法在每个数据集的 10 折分类精度作 t 检验, H 为 0 表示一致, H 为 1 表示不一致.

从统计结果上看, 10 个数据集中除了 *iosphere*, *musk* 和 *sonar* 以外, 其他 7 个数据集上的分类精度差异并不显著. 可以认为, 压缩率为 50% 时, FD-SVM 的分类精度保持与 SVM 相当, 并且 FD-SVM 分类消耗的时间仅是 SVM 的一半.

进一步分析压缩率对于分类精度的影响, 考虑到 *iosphere*, *musk* 和 *sonar* 数据集上的分类精度差异较大, 分别为 4.28%, 9.45% 和 4.77%, 在 3 个数据集上将压缩率按照 (0.5, 0.6, 0.7, 0.8, 0.9) 的序列递增, 增加新支持向量的数量, 考察分类精度是否会随着压缩率的增加而增加. 另外选取分类差异小的 *heart-disease* 数据集, 按照 (0.5, 0.4, 0.3, 0.2, 0.1, 0.09, 0.08, 0.07, 0.06) 的次序减低压缩率, 考察分类精度是否会随着压缩率的减低而减小.

图 2 为不同压缩率对于分类精度的影响. 从图 2

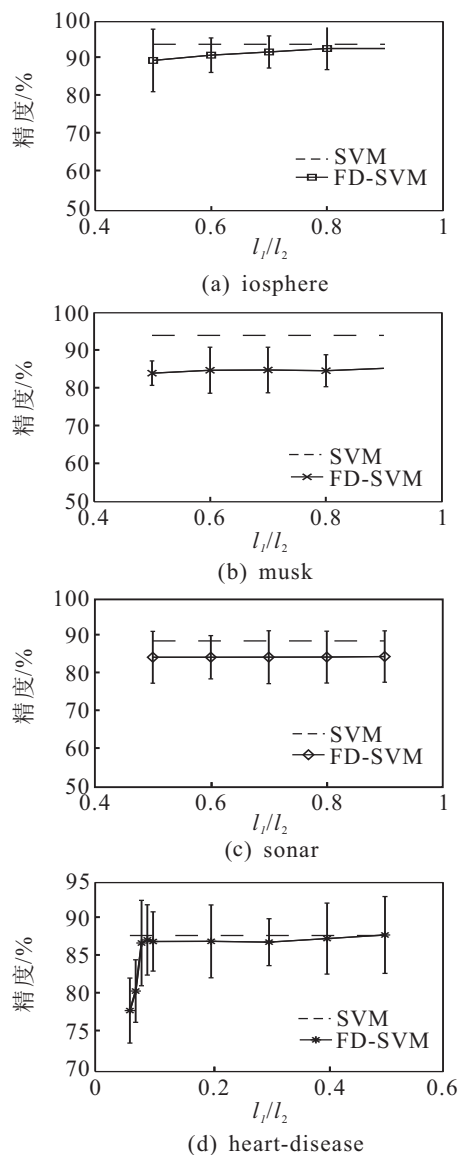


图 2 不同压缩率对于分类精度的影响

可以看出, 在 *musk* 和 *sonar* 数据集上, 随着压缩率的增加, FD-SVM 分类精度的增长并不明显. 可能是因为这些数据集上经 SVM 训练得到的支持向量本身已经很稀疏, 无法再约简, 或者其空间几何结构不适合使用 FCM 压缩, 所以 FD-SVM 在这两个数据集上的分类精度损失较大. 注意到图 2(a) 中, *iosphere* 数据集随着压缩率的增加分类精度有明显的提升, 说明对该数据集直接使用 FCM 压缩损失了较多信息. 另外, 图 2(d) 中 *heart-disease* 数据集在压缩率低于 0.08 后, FD-SVM 的分类精度急剧下降, 但是在 [0.09, 0.5] 的范围内, FD-SVM 的分类精度依然保持较小的分类差异 (平均 0.5%), 说明标准 SVM 训练得到的支持向量有很大的压缩空间, 尤其对于一些空间几何结构呈聚簇状分布的数据集. 虽然 FD-SVM 简单地使用了 FCM 压缩方法, 但在这类数据集上能够以较低的压缩率得到高效的分类函数, 并保持分类精度损失较小.

4 结 论

本文提出了一种新的支持向量机快速分类算法 FD-SVM. 该算法使用 FCM 聚类算法将原始支持向量机训练得到的 l_1 个支持向量按照一定的压缩率稀疏化后得到较少的 l_2 个新的支持向量; 然后按照误差最小原则构造了一个线性最小二乘问题求解新的支持向量的权重系数和决策函数的偏置. 在人工数据集和 UCI 真实数据集上进行实验的结果表明, FD-SVM 以 50% 的压缩率能够保持与标准 SVM 相当的分类精度, 同时将分类决策消耗的时间减少 50%, 尤其在一些空间几何结构呈聚簇状分布的数据集上, 当支持向量集约简到 8% 时, FD-SVM 依然能保持与 SVM 相当的分类精度.

根据不同数据集的分布特点, 在保证分类精度的前提下, 采用有效的压缩算法自动调整压缩率并将支持向量集压缩到最小, 同时加快 SVM 的训练速度和分类速度是制约 SVM 实际应用的瓶颈之一, 值得进一步深入研究.

参考文献(References)

- [1] Vladimir N Vapnik. Statistical learning theory[M]. USA: Wiley-Interscience, 1998: 20-29.
- [2] Huang Kai-zhu, Zheng Da-nian, Sun Jun, et al. Sparse learning for support vector classification[J]. Pattern Recognition Letters, 2010, 31(13): 1944-1951.
- [3] Zhang Kai, Kwok J T. Simplifying mixture models through function approximation[J]. IEEE Trans on Neural Networks, 2010, 21(4): 644-658.
- [4] Platt J C. Fast training of support vector machines using sequential minimal optimization[C]. Advances in Kernel

- Methods—Support Vector Learning. Cambridge: MIT Press, 1999: 185-208.
- [5] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [6] Schölkopf B, Smola A J, Williamson R C, et al. New support vector algorithms[J]. *Neural Computation*, 2000, 12(5): 1207-1245.
- [7] Burges C J C. Simplified support vector decision rules[C]. *Proc of 13th Int Conf on Machine Learning*. San Mateo: Morgan Kaufmann, 1996: 71-77.
- [8] Burges C J C, Schölkopf B. Improving the accuracy and speed of support vector machines[C]. *Advances in Neural Information Processing Systems*. Vancouver: MIT Press, 1997: 375-381.
- [9] 刘向东, 陈兆乾. 一种快速支持向量机分类算法的研究[J]. *计算机研究与发展*, 2004, 41(8): 1327-1332. (Liu X D, Chen Z Q. A fast classification algorithm of support vector machines[J]. *Chinese J of Computer Research and Development*, 2004, 41(8): 1327-1332.)
- [10] 方景龙, 陈铄, 潘志庚, 等. 复杂分类问题支持向量机的简化[J]. *电子学报*, 2007, 35(5): 858-861. (Fang J L, Chen S, Pan Z G, et al. A simplification to support vector machine for complicated recognition problem[J]. *Chinese J of Electronics*, 2007, 35(5): 858-861.)
- [11] 韩德强, 韩崇昭, 杨艺. 基于 k -最近邻的支持向量预选取方法[J]. *控制与决策*, 2009, 24(4): 494-498. (Han D Q, Han C Z, Yang Y. Approach for pre-extracting support vectors based on k -NN[J]. *Control and Decision*, 2009, 24(4): 494-498.)
- [12] Lawson C L, Hanson R J. *Solving least-squares problems*[M]. Prentice-Hallpage, 1974: 161-163.
- [13] Chang Chih-chung, Lin Chih-jen. LIBSVM: A library for support vector machines[ED/OL]. (2008-10-31)[2008-11-04]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [14] Asuncion A, Newman D J. UCI machine learning repository[DB/OL]. (2008-08-10)[2008-11-01]. <http://archive.ics.uci.edu/ml/>.

~~~~~

(上接第458页)

- [6] 王坚强. 模糊多准则决策方法研究综述[J]. *控制与决策*, 2008, 23(6): 601-607. (Wang J Q. Overview on fuzzy multi-criteria decision-making approach[J]. *Control and Decision*, 2008, 23(6): 601-607.)
- [7] 王坚强, 张忠. 基于直觉模糊数的信息不完全的多准则规划方法[J]. *控制与决策*, 2008, 23(10): 1145-1148. (Wang J Q, Zhang Z. Programming method of multi-criteria decision making based on intuitionistic fuzzy number with incomplete certain information [J]. *Control and Decision*, 2008, 23(10): 1145-1148.)
- [8] 王坚强, 张忠. 基于直觉梯形模糊数的信息不完全确定的多准则决策方法[J]. *控制与决策*, 2009, 24(2): 226-230. (Wang J Q, Zhang Z. Multi-criteria decision making method with incomplete certain information based on intuitionistic fuzzy number[J]. *Control and Decision*, 2009, 24(2): 226-230.)
- [9] Wang J Q, Zhang Z. Aggregation operators on intuitionistic trapezoidal fuzzy number and its application to multi-criteria decision making problems[J]. *J of Systems Engineering and Electronics*, 2009, 20(2): 321-326.
- [10] 万树平, 董九英. 多属性群决策的直觉梯形模糊数法[J]. *控制与决策*, 2010, 25(5): 773-776. (Wan S P, Dong J Y. Method of intuitionistic trapezoidal fuzzy number for multi-attribute group decision[J]. *Control and Decision*, 2010, 25(5): 773-776.)
- [11] Wei G W. Some arithmetic aggregation operators with intuitionistic trapezoidal fuzzy numbers and their application to group decision making[J]. *J of Computer*, 2010, 5(3): 345-351.
- [12] 徐泽水, 达庆利. 一种基于可能度的区间判断矩阵排序法[J]. *中国管理科学*, 2003, 11(1): 63-65. (Xu Z S, Da Q L, A possibility based method for priorities of interval judgment matrices[J]. *Chinese J of Management Science*, 2003, 11(1): 63-65.)
- [13] Li D F, Wang Y C. Mathematical programming approach to multiattribute decision making under intuitionistic fuzzy environments[J]. *Int J of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2008, 16(4): 557-577.