

结合全局和局部正则化的半监督二分类算法

吕佳^{1,2,3*}

(1. 内蒙古大学 数学科学学院, 呼和浩特 010021;

2. 重庆师范大学 计算机与信息科学学院, 重庆 400047; 3. 中国农业大学 理学院, 北京 100083)

(* 通信作者电子邮箱 lvjia@cqu.edu.cn)

摘要: 针对在半监督分类问题中单独使用全局学习容易出现的在整个输入空间中较难获得一个优良的决策函数的问题, 以及单独使用局部学习可在特定的局部区域内习得较好的决策函数的特点, 提出了一种结合全局和局部正则化的半监督二分类算法。该算法综合全局正则项和局部正则项的优点, 基于先验知识构建的全局正则项能平滑样本的类标号以避免局部正则项学习不充分的问题, 通过基于局部邻域内样本信息构建的局部正则项使得每个样本的类标号具有理想的特性, 从而构造出半监督二分类问题的目标函数。通过在标准二类数据集上的实验, 结果表明所提出的算法其平均分类正确率和标准误差均优于基于拉普拉斯正则项方法、基于正则化拉普拉斯正则项方法和基于局部学习正则项方法。

关键词: 半监督学习; 二分类问题; 全局正则化; 局部正则化; 平滑

中图分类号: TP18; TP391.4; TP301.6 **文献标志码:** A

Semi-supervised binary classification algorithm based on global and local regularization

LÜ Jia^{1,2,3*}

(1. School of Mathematical Sciences, Inner Mongolia University, Hohhot Nei, Mongolia 010021, China;

2. College of Computer and Information Science, Chongqing Normal University, Chongqing 400047, China;

3. College of Science, China Agricultural University, Beijing 100083, China)

Abstract: As for semi-supervised classification problem, it is difficult to obtain a good classification function for the entire input space if global learning is used alone, while if local learning is utilized alone, a good classification function on some specified regions of the input space can be got. Accordingly, a new semi-supervised binary classification algorithm based on a mixed local and global regularization was presented in this paper. The algorithm integrated the benefits of global regularizer and local regularizer. Global regularizer was built to smooth the class labels of the data so as to lessen insufficient training of local regularizer, and based upon the neighboring region, local regularizer was constructed to make class label of each data have the desired property, thus the objective function of semi-supervised binary classification problem was constructed. Comparative semi-supervised binary classification experiments on some benchmark datasets validate that the average classification accuracy and the standard error of the proposed algorithm are obviously superior to other algorithms.

Key words: semi-supervised learning; binary classification problem; global regularization; local regularization; smooth

0 引言

在解决机器学习和模式识别中的分类器学习问题时, 通常根据训练样本的类标号是否参与训练而将学习分成三类: 训练样本类标号要参与训练的, 称之为有监督学习^[1-2]; 训练样本类标号不参与训练的, 称之为无监督学习^[3]; 部分样本的类标号参与训练的, 称之为半监督学习^[4-5]。在实际应用中, 获取大量有标记数据通常费时且代价较高, 例如 Web 文本分类中^[6], 很容易收集到 Web 网页, 但标记不同的网页为对应的主题却费时费力, 因此旨在从大量无标记样本和少量有标记样本中学习的半监督学习在理论及实际应用中得到了极大关注。

到目前为止, 已有许多研究学者提出了各种半监督学习算法, 其中大部分算法都属于基于图的方法^[7-12], 通过估计一个以所有样本为节点、样本相似性为边的权重构成的图上的函数 f 来实现分类, f 要求同时满足以下条件: 一是应该与

其相邻有标记样本的标记接近; 二是应该在图上光滑。文献[7]提出一种基于 Gaussian 随机域和谐波函数的算法 (即基于拉普拉斯正则项方法 (Lap_Reg)) 来平滑无标记样本的标记信息; 文献[8]首次提出一种基于全局流形的正则化方法来学习无标记样本的类标号; 文献[9]提出一种基于局部和全局一致性的算法, 该算法利用基于正则化拉普拉斯正则项方法 (NLap_Reg) 在数据流形上得到无标记样本的类别; 文献[10]提出利用样本的邻居样本的信息来学习得到该样本类标号的基于局部学习正则项方法 (LL_Reg); 文献[11]提出利用局部样条回归方法来构建半监督分类算法; 文献[12]提出了一种通用的局部和全局正则化框架用来解决无标记样本的类标号标注问题, 以上算法实际上都是基于正则化框架下的半监督学习算法, 目标函数由损失函数和正则项构成, 它们的区别在于选择不同的损失函数和正则项。

本文在总结和分析上述各种方法的基础上, 在半监督二分类问题的目标函数中同时加入全局正则项和局部正则项,

提出了一种结合全局和局部正则化的半监督二分类算法。实验结果证明了本文算法的正确性和可行性。

1 半监督二分类问题

半监督二分类问题描述如下:

给定式(1)所示的训练集,据此寻找与 $\mathbf{x}_{l+1}, \dots, \mathbf{x}_n$ 对应的在 $\{+1, -1\}$ 中取值的输出 y_{l+1}, \dots, y_n 的值。

$$T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\} \cup \{\mathbf{x}_{l+1}, \dots, \mathbf{x}_n\} \quad (1)$$

其中: $\mathbf{x}_i \in \mathbf{R}^d (i = 1, 2, \dots, n), y_i \in \{+1, -1\} (i = 1, 2, \dots, l)$ 。

求解半监督二分类问题通常转化为求解如式(2)所示的无约束最优化问题^[7-12]。

$$\min_F ((F - Y)^T C (F - Y) + F^T H F) \quad (2)$$

其中:目标函数的第一项 $(F - Y)^T C (F - Y)$ 称为经验风险损失项, $C \in \mathbf{R}^{n \times n}$ 是一个对角矩阵,对角元素 $C_{ii} = C_i > 0 (i = 1, 2, \dots, l), C_{ii} = C_u \geq 0 (i = l + 1, \dots, n), Y = (y_1, \dots, y_n)^T, y_i \in \{+1, -1\} (i = 1, 2, \dots, l), y_i = 0 (i = l + 1, \dots, n)$ 为 $\mathbf{x}_{l+1}, \dots, \mathbf{x}_n$ 的初始类标号;第二项 $F^T H F$ 称为正则项, H 是 $n \times n$ 的正则项因子矩阵,如式(2)所示的无约束最优化问题的最优解 $F = (f_1, \dots, f_n)^T \in \mathbf{R}^n, F$ 是对应于 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的 n 维实值向量。最小化经验风险项的作用是使得最优解尽可能与其实际类别一致,而最小化正则项则是使得 $f_i (i = 1, 2, \dots, n)$ 具有理想的性质。

易知如式(2)所示的无约束最优化问题的解为:

$$F = (H + C)^{-1} C Y \quad (3)$$

$\mathbf{x}_{l+1}, \dots, \mathbf{x}_n$ 对应的 y_{l+1}, \dots, y_n 按式(4)推断:

$$y_i = \text{sign}(f_i); i = l + 1, \dots, n \quad (4)$$

符号函数 $\text{sign}(\cdot)$ 定义如下:

$$\text{sign}(t) = \begin{cases} +1, & t \geq 0 \\ -1, & t < 0 \end{cases} \quad (5)$$

2 全局正则化

半监督学习中有标记样本仅占少数,这样通过最小化如式(2)所示的无约束最优化问题无法保证充分学习,因此需要一些先验知识指导 F 的学习,全局正则项 $\|f\|_l^2$ 反映了 $p(\mathbf{x})$ 的内在几何分布信息^[8]。 $p(\mathbf{x})$ 为样本的分布概率, $p(y|\mathbf{x})$ 为在已知样本 \mathbf{x} 的条件下类标号为 y 的条件概率,分布较为集中的样本最可能具有相似的类标号,即 \mathbf{x}_1 和 \mathbf{x}_2 相邻,则 $p(y|\mathbf{x}_1) \approx p(y|\mathbf{x}_2)$, \mathbf{x}_1 和 \mathbf{x}_2 的类标号相似。也就是说, $p(y|\mathbf{x})$ 应当在 $p(\mathbf{x})$ 内在几何性质下非常光滑。 $\|f\|_l^2$ 为黎曼积分,形式如下:

$$\|f\|_l^2 = \int_{\mathbf{x} \in M} \|\nabla_m f\|^2 d\mathbf{p}(\mathbf{x}) \quad (6)$$

其中: M 代表低维数据流形, $\nabla_m f$ 是关于 M 的梯度 f , $\|f\|_l^2$ 反映了 f 的光滑性。 $\|f\|_l^2$ 可进一步近似表示为:

$$\|f\|_l^2 = \frac{\gamma_l}{n^2} \sum_{i,j} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 w_{ij} = \frac{\gamma_l}{n^2} F^T L F \quad (7)$$

其中: n 是样本总数, γ_l 是调节参数; $W = [w_{ij}] \in \mathbf{R}^{n \times n}$ 是一个无向加权图,其顶点对应样本,其边 w_{ij} 对应样本 \mathbf{x}_i 和 \mathbf{x}_j 的相似度。 L 是拉普拉斯项,计算方法如下:

$$L = D - W \in \mathbf{R}^{n \times n} \quad (8)$$

w_{ij} 计算公式为:

$$w_{ij} = \begin{cases} \exp\left(-\frac{1}{\delta} \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), & \mathbf{x}_j \in N_i \\ 0, & \text{其他} \end{cases} \quad (9)$$

其中: $\delta > 0, D$ 是一个对角矩阵,其对角元素 $D_{ii} = \sum_{j=1}^n w_{ij} (i = 1, 2, \dots, n)$ 。

3 局部正则化

文献[13]提出局部学习算法通常在性能上要优于全局学习算法,这是因为要寻找一个在整个样本空间内都行之有效的分类器是较为困难的,因此在许多实际应用中局部学习算法不失为一种更好的选择^[14-15]。局部学习的基本思想是一个样本的类别应能由其邻域内所有样本的学习来推断,即对应每一个样本 \mathbf{x}_i 的 f_i 值应该与建立在其邻域样本集 $\{(\mathbf{x}_j, f_j) | \mathbf{x}_j \in N_i\}$ 上学习得到的决策函数 $g_i(\mathbf{x})$ 的输出值 $g_i(\mathbf{x}_i) (i = 1, 2, \dots, n)$ 相同或相近,即希望 $\sum_{i=1}^n (f_i - g_i(\mathbf{x}_i))^2$ 尽可能小,该项对应着如式(2)所示的无约束最优化问题中的正则项^[10]。其矩阵形式为:

$$\|F - G\|^2; G = (g_1(\mathbf{x}_1), \dots, g_n(\mathbf{x}_n))^T \quad (10)$$

$g_i(\mathbf{x})$ 采用线性函数,形式如下:

$$g_i(\mathbf{x}) = \mathbf{w}_i^T (\mathbf{x} - \mathbf{x}_i) + b_i; \forall \mathbf{x} \in N_i \subset \mathbf{R}^d \quad (11)$$

其中: $\mathbf{w}_i \in \mathbf{R}^d, b_i \in \mathbf{R}, N_i$ 为 $\mathbf{x}_i (i = 1, 2, \dots, n)$ 的邻域。 $g_i(\mathbf{x})$ 可以通过求解如式(12)所示的无约束最优化问题得到。

$$\min_{\substack{\mathbf{w}_i \in \mathbf{R}^d \\ b_i \in \mathbf{R}}} \lambda \|\mathbf{w}_i\|^2 + \sum_{\mathbf{x}_j \in N_i} (\mathbf{w}_i^T (\mathbf{x}_j - \mathbf{x}_i) + b_i - f_j)^2 \quad (12)$$

其中: $\lambda > 0$ 。

下面的定理表明 $g_i(\mathbf{x}_i)$ 具有更加明确的解析表达式。

定理1 对 $\mathbf{x}_i (i = 1, 2, \dots, n), N_i$ 是 \mathbf{x}_i 的邻域, n_i 是 N_i 中样本个数,记 $X_i = (\mathbf{x}_j - \mathbf{x}_i)_{d \times n_i} \in \mathbf{R}^{d \times n_i}, \mathbf{x}_j \in N_i, \mathbf{w}_i \in \mathbf{R}^d, b_i \in \mathbf{R}, \mathbf{e} = (1, \dots, 1)^T \in \mathbf{R}^{n_i}, I$ 为 $n_i \times n_i$ 的单位矩阵,则 $g_i(\mathbf{x}_i) = \alpha_i^T F_i, F_i = (f_j)_{\mathbf{x}_j \in N_i}^T \in \mathbf{R}^{n_i}$ 且

$$\alpha_i^T = \frac{\mathbf{e}^T (\lambda I + X_i^T X_i)^{-1}}{\mathbf{e}^T (\lambda I + X_i^T X_i)^{-1} \mathbf{e}} \quad (13)$$

由式(13)知, α_i 仅与 X_i 有关,而与 f_j 无关,将 α_i 扩展成矩阵 $A = (a_{ij}) \in \mathbf{R}^{n \times n}$,当 $\mathbf{x}_j \in N_i$ 时, $a_{ij} = \alpha_{ij}$,当 $\mathbf{x}_j \notin N_i$ 时, $a_{ij} = 0$,则 $G = AF$ 。将其代入 $\|F - G\|^2$ 得到:

$$F^T (I - A)^T (I - A) F$$

其中 I 为 $n \times n$ 的单位矩阵。令:

$$H = (I - A)^T (I - A) \quad (14)$$

H 正是如式(12)所示的无约束最优化问题中的正则项因子。

4 结合全局和局部正则化的半监督二分类算法

局部学习算法在机器学习问题中已显示了良好的学习特性,但由于邻域样本数目通常较少有可能导致局部分类器学习不充分^[12],故本文将全局正则项和局部正则项整合在一起,形成了结合全局和局部正则化的半监督二分类问题,如下所示:

$$\min_F ((F - Y)^T D (F - Y) + \frac{\gamma_l}{n^2} F^T L F + F^T H F) \quad (15)$$

其中:第一项 $(F - Y)^T D (F - Y)$ 与如式(2)所示的无约束最优化问题中第一项相同;第二项 $\frac{\gamma_l}{n^2} F^T L F$ 是全局学习正则项,最小化该项来光滑样本标号;第三项 $F^T H F$ 是局部学习正则

项,最小化该项以使 F 具有希望的理想性质。

易知如式(15)所示的无约束最优化问题的最优解为:

$$F = ((I - A)^T(I - A) + \frac{\gamma_l}{n^2}(D - W) + C)^{-1}CY \quad (16)$$

结合全局和局部正则化的半监督二分类算法(GL_Reg)的算法框架如下:

- 1) 给定如式(1)所示的训练集。
- 2) 选择适当的参数 $D_l > 0, D_u \geq 0, \lambda > 0, \gamma_l > 0, \delta > 0$, 邻域中样本个数 $K > 0$ 。
- 3) 计算与 $x_i \in \mathbf{R}^d (i = 1, 2, \dots, n)$ 最近的 K 个样本 $x_{j_1}, x_{j_2}, \dots, x_{j_K}, j_k \in \{1, 2, \dots, n\}, k \in \{1, 2, \dots, K\}$, 记为 $X_i = (x_{j_1} - x_i, x_{j_2} - x_i, \dots, x_{j_K} - x_i) \in \mathbf{R}^{d \times K}$ 。
- 4) 根据式(13)计算 $\alpha_i = (\alpha_{ij_1}, \dots, \alpha_{ij_K})^T$, 将 α_i 扩展成矩阵 $A = (a_{ij}) \in \mathbf{R}^{n \times n}$, 若 x_j 是 x_i 邻域内样本, 令 $a_{ij} = \alpha_{ij}$; 否则 $a_{ij} = 0$ 。
- 5) 根据式(8)计算 L 。
- 6) 根据式(16)计算 $F = (f_1, \dots, f_n)^T$ 。
- 7) 根据式(4)确定 x_{l+1}, \dots, x_n 的类标号。

5 实验结果

5.1 数据集、方法与参数设置

实验中采用9个标准数据集,其详细描述见表1^[4,10-11]。

通常一个分类算法的优劣是以分类精度来度量的,在训练阶段,采用留一法(Leave-One-Out, LOO)选择参数。LOO实现方法是针对每一组参数,每次将第 $i (i = 1, 2, \dots, l)$ 个有标记样本改为无标记样本后再对训练集进行训练,判断第 i 个样本是否错分,算法迭代 l 次后统计样本错分个数,最终选择样本错分个数最少的那一组参数作为最优参数。

本文算法参数设置如下:

每个训练集随机地选择10%作为有标记样本:

$$D_l \in \{0.1, 1, 10, 100\}$$

$$D_u = 10^{-6}$$

$$\lambda \in \{0.1, 1, 10, 100\}$$

$$K \in \{5, 10, 20, 50, 100\}$$

$$\gamma_l \in \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1, 2^1, 2^2, 2^3, 2^4, 2^5\}$$

$$\delta \in \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1, 2^1, 2^2, 2^3, 2^4, 2^5\}$$

为了说明本文算法的有效性,将GL_Reg与文献[7, 9-10]中用到的Lap_Reg、NLap_Reg和LL_Reg比较分析。

表1 二类数据集描述

数据集	数据集规模	数据集特征数
BCI	400	117
Diabetes	768	8
Breast	683	10
Ionosphere	351	34
Digit1	1 500	241
USPS	1 500	241
G241c	1 500	241
G241d	1 500	241
G241n	1 500	241

5.2 实验结果

实验重复20次,得到4个算法在9个标准二类数据集上的平均分类正确率和标准误差,见表2。从表2可看出:在Digit1和G241c上GL_Reg分类正确率略低于LL_Reg的分类正确率,但是GL_Reg标准误差在G241c上要远小于LL_Reg的分类误差。而在其他数据集上,GL_Reg无论是分类正确率还是标准误差均要优于对比算法。

表2 4个算法的分类结果比较

数据集	Lap_Reg		NLap_Reg		LL_Reg		GL_Reg	
	正确率/%	标准误差	正确率/%	标准误差	正确率/%	标准误差	正确率/%	标准误差
BCI	52.40	2.29	52.78	2.31	68.85	2.30	72.85	2.11
Diabetes	67.20	2.54	68.07	2.71	72.37	2.40	74.61	1.33
Breast	96.75	1.87	96.60	1.65	97.00	1.13	98.00	0.87
Ionosphere	92.91	1.66	95.40	1.45	96.33	0.92	98.71	1.04
Digit1	96.98	0.84	97.09	0.59	97.37	0.66	97.15	1.10
USPS	92.91	3.37	95.40	2.04	96.33	1.24	98.71	1.04
G241c	61.00	2.23	55.00	3.92	78.64	3.67	78.35	1.20
G241d	63.88	1.50	56.69	3.30	77.49	1.79	78.79	1.23
G241n	58.81	1.14	55.35	0.48	76.62	1.27	78.19	0.57

6 结语

本文提出的结合全局和局部正则化的半监督二分类算法利用局部正则项使得最优解具有理想的性质,利用全局正则项来弥补局部正则项因邻域内样本较少可能导致的学习不充分的问题,从而平滑类标号。数值实验结果表明,与基于拉普拉斯正则项方法、基于正则化拉普拉斯正则项方法和基于局部学习正则项方法相比,本文算法分类精度更高。下一步工作将把全局正则项和局部正则项应用到半监督多类分类问题以及半监督多标记分类问题中。

参考文献:

- [1] 邓乃扬,田英杰. 数据挖掘中的新方法——支持向量机[M]. 北京: 科学出版社, 2004.
- [2] 邓乃扬,田英杰. 支持向量机: 理论、算法与拓展[M]. 北京: 科

学出版社, 2009.

- [3] 吕佳. 基于动态隧道系统的K-means聚类算法研究[J]. 重庆师范大学学报: 自然科学版, 2009, 26(1): 73-77.
- [4] CHAPPELLE O, SCHOLKOPF B, ZIEN A. Semi-supervised learning[M]. Cambridge: MIT Press, 2006.
- [5] ZHU X J. Semi-supervised learning literature survey [EB/OL]. [2010-05-10]. http://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf.
- [6] 吕佳. 基于改进分类模型的文本分类系统实现[J]. 重庆师范大学学报: 自然科学版, 2009, 26(2): 68-73.
- [7] ZHU X J, GHAHARMANI Z, LAFFERTY J. Semi-supervised learning using Gaussian fields and harmonic functions[C]// FAWCETT T, MISHRA N. Proceedings of 20th International Conference on Machine Learning, Menlo Park: AAAI Press, 2003: 912-919.

(下转第648页)

3.2 UCI 数据集实验

为了进一步验证 FCM-DC 算法的有效性,采用 UCI 数据集中的 IRIS 和 WINE 数据进行实验,因为这两组数据是国际公认的比较无监督聚类方法效果好坏的典型数据。其中 IRIS 数据包含 150 个 4 维的样本点,类别数为 3,每类 50 个样本点。第一类数据与其他两类数据离得较远,第二类数据与第三类数据离得较近,且部分重叠;WINE 数据包含 178 个 13 维数据,类别数也为 3,三类样本数目各为 59,71 和 48。

表 2~3 是 FCM 算法与 FCM-DC 算法对 UCI 数据集的聚类效果对比。在 IRIS 数据实验中,对于相隔较远的第一类数据,两种算法都能够完全做到没有错误,但对于有交叉重合的第二类和第三类数据,FCM-DC 算法表现出了更高的准确率,总体准确率比 FCM 提升了 12.6%;在 WINE 数据实验中,由于数据本身的高维稀疏特性,两种算法的聚类分错率都比较高,但 FCM-DC 算法仍然比 FCM 提升了 9.1% 的聚类准确率。实验说明,调节因子 w_i 起到了距离度量的修正优化效果,提升了聚类的准确率。同时,由于这两组数据集是典型的非球形结构数据,因此实验也说明 FCM-DC 算法适用面更广。

表 2 IRIS 数据的聚类效果对比

聚类算法	数据类别	分错个数	聚类分错率/%
FCM 聚类	第一类	0	10.67
	第二类	13	
	第三类	3	
FCM-DC 聚类	第一类	0	9.33
	第二类	9	
	第三类	5	

表 3 WINE 数据的聚类效果对比

聚类算法	数据类别	分错个数	聚类分错率/%
FCM 聚类	第一类	40	49.44
	第二类	25	
	第三类	23	
FCM-DC 聚类	第一类	36	44.94
	第二类	22	
	第三类	25	

4 结语

针对经典模糊 C 均值算法存在的等划分趋势的缺陷,本文提出了一种距离修正的改进算法——FCM-DC。基于样本点密度信息,引入了距离修正的调节因子,对 FCM 算法样本差异性度量进行了修正。通过人造数据集与 UCI 数据集两组实验,对比分析 FCM-DC 与 FCM 算法的性能,结果表明,

FCM-DC 算法对于非球形结构的数据同样适用,且具有更高的聚类准确率。

参考文献:

- [1] PEDRYCZ W. Conditional fuzzy C-means [J]. Pattern Recognition Letters, 1996, 17(6): 625-631.
- [2] GRAVES D, PEDRYCZ W. Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study [J]. Fuzzy Sets and Systems, 2010, 161(4): 522-543.
- [3] SONG Q, YANG X L, SOH Y C, et al. An information-theoretic fuzzy C-spherical shells clustering algorithm [J]. Fuzzy Sets and Systems, 2010, 161(13): 1755-1773.
- [4] LEE M, PEDRYCZ W. The fuzzy C-means algorithm with fuzzy P-mode prototypes for clustering objects having mixed features [J]. Fuzzy Sets and Systems, 2009, 160(24): 3590-3600.
- [5] 刘小芳,曾黄麟,吕炳朝. 点密度函数加权模糊 C-均值算法的聚类分析 [J]. 计算机工程与应用, 2004, 40(24): 64-65.
- [6] TANG C L, WANG S G, XU W. New fuzzy C-means clustering model based on the data weighted approach [J]. Data & Knowledge Engineering, 2010, 69(9): 881-900.
- [7] 王惠,申石磊. 一种改进的特征加权 K-means 聚类算法 [J]. 微电子学与计算机, 2010, 27(7): 161-163.
- [8] 李丹,顾宏,张立勇. 基于属性权重区间监督的模糊 C 均值聚类算法 [J]. 控制与决策, 2010, 25(3): 457-460.
- [9] BAI L, LIANG J Y, DANG C Y, et al. A novel attribute weighting algorithm for clustering high-dimensional categorical data [J]. Pattern Recognition, 2011, 44(12): 2843-2861.
- [10] 蔡静新, 谢福鼎, 张永. 基于自适应马氏距离的模糊 C 均值算法 [J]. 计算机工程与应用, 2010, 46(34): 174-176.
- [11] XIANG S M, NIE F P, ZHANG C S. Learning a Mahalanobis distance metric for data clustering and classification [J]. Pattern Recognition, 2008, 41(12): 3600-3612.
- [12] 王骏,王士同. 基于混合距离学习的双指数模糊 C 均值算法 [J]. 软件学报, 2010, 21(8): 1878-1888.
- [13] TSAI D M, LIN C C. Fuzzy C-means based clustering for linearly and nonlinearly separable data [J]. Pattern Recognition, 2011, 44(8): 1750-1760.
- [14] 于迪,李义杰. 基于减法聚类改进的模糊 C 均值算法的模糊聚类研究 [J]. 微型机与应用, 2010, 29(16): 14-16.
- [15] 李雷,罗红旗,丁亚丽. 自适应约束模糊 C 均值聚类算法 [J]. 模糊系统与数学, 2010, 24(5): 126-130.
- [16] YU J, CHENG Q S, HUANG H K. Analysis of the weighting exponent in the FCM [J]. IEEE Transactions on System, Man and Cybernetics: Part B: Cybernetics, 2004, 34(1): 634-639.
- [17] 肖满生,阳梯兰,张居武,等. 基于模糊相关度的模糊 C 均值聚类加权指数研究 [J]. 计算机应用, 2010, 30(12): 3388-3390.

(上接第 645 页)

- [8] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples [J]. Journal of Machine Learning Research, 2006, 7(11): 2399-2434.
- [9] ZHOU D Y, BOUSQUET O, LAL T N, et al. Learning with local and global consistency [C]// THURM S, SAUL L, SCHÖLKOPF B. Proceedings of the 18th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2003: 321-328.
- [10] WU M R, SCHÖLKOPF B. Transductive classification via local learning regularization [C]// MEILA M, SHEN X. Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. Cambridge: MIT Press, 2007: 624-631.
- [11] XIANG S M, NIE F P, ZHANG C S. Semi-supervised classification via local spline regression [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 32(11): 2039-2053.
- [12] WANG F. A general learning framework using local and global regularization [J]. Pattern Recognition, 2010, 43(9): 3120-3129.
- [13] VAPNIK V. The nature of statistical learning theory [M]. Berlin: Springer-Verlag, 1995.
- [14] BOTTOU L, VAPNIK V. Local learning algorithms [J]. Neural Computation, 1992, 4(6): 888-900.
- [15] VAPNIK V, BOTTOU L. Local algorithms for pattern recognition and dependencies estimation [J]. Neural Computation, 1993, 5(6): 893-909.