可保证分类性能的最小二乘支持向量机

徐金宝[1],廖　雷[1],业巧林[2]
XU Jin-bao[1],LIAO Lei[1],YE Qiao-lin[2]

1.南京工程学院 计算机工程学院,南京 211167
2.南京林业大学 信息技术学院,南京 210037
1.School of Computer Science,Nanjing Institute of Technology,Nanjing 211167,China
2.School of Information Technology,Nanjing Forestry University,Nanjing 210037,China

XU Jin-bao,LIAO Lei,YE Qiao-lin.Least squares support vector machine classifiers with guaranteed classification performance.Computer Engineering and Applications,2009,45(21):48-50.

Abstract: Support Vector Machine (SVM) is one of focuses of research and application in classification.A new least-squares-based algorithm that introduces a within-class scatter with guaranteed classification performance(VSLSVM) in the design of least squares support vector machines(LS-SVM) is presented.This algorithm can obtain better correctness that reformulates primal LS-SVM problems with optimality criterion Min $\boldsymbol{w'Mw}$ where $\boldsymbol{w}$ is the weight vector corresponding the primal LS-SVM problems,$\boldsymbol{M}$ is the within-class scatter matrix.This method only requires to solve a linear system instead of a quadratic programming problem. Experiments are included to compare SVM and Suykens' approach.

Key words: least squares support vector machines(LS-SVM);with-class scatter;better correctness;linear system

摘　要:当前支持向量机是分类研究与应用的一个热点。提出了一个新的最小二乘支持向量机算法,该算法向最小二乘支持向量机(LS-SVM)优化模型中融入了类内散度(VSLSVM)思想,即用优化准则 Min $\boldsymbol{w'Mw}$ 对原 LS-SVM 进行重组合,$\boldsymbol{w}$ 为对应 LS-SVM 中的权向量,$\boldsymbol{M}$ 是类内散度矩阵。提出的方法仅仅需要求解一个线性系统而不是凸规划问题,实验主要对 SVM 和 Suykens 等人的方法进行了比较,并验证了提出的算法的有效性。

关键词:最小二乘支持向量机;类内散度;更好精度;线性系统
DOI:10.3778/j.issn.1002-8331.2009.21.012　**文章编号:**1002-8331(2009)21-0048-03　**文献标识码:**A　**中图分类号:**TP181

1　引言

支持向量机(Support Vector Machine-SVM)方法是建立在统计学习理论中的结构风险最小化和 VC 维基础上[1-2]。在实际问题中,往往原空间中的样本是线性不可分的,解决这种线性不可分的问题,SVM 是通过核映射方法,将原空间中样本映射到高维空间,使得样本在此高维空间中线性分开[3]。

SVM 以拥有较好的泛化能力而著称,其主要思想是寻找一个能使分类样本尽可能很好地分开的超平面。但是,标准 SVM 主要是通过一个二次规划问题求解的[4],由于需要大量的迭代次数,从而带来了大量的时间代价,为了改进标准 SVM 的计算时间代价的问题,Mangasarian 等人已经提出了许多快速算法,如 PSVM[5],LSVM[6]和 SSVM[7]。从分类精度角度出发,提出了一个可保证分类性能的 SVM 算法。提出的方法仅仅需要求解一个线性系统而不是凸规划问题,将类内散度的思想融入到最小二乘支持向量机(LS-SVM)中,即用优化准则 Min $\boldsymbol{w'Mw}$ 对原 LS-SVM 进行重组合,$\boldsymbol{w}$ 为对应 LS-SVM 中的权向量,$\boldsymbol{M}$ 是类内散度矩阵。在 UCI 数据库上的计算结果显示,与 SVM 和 Suykens 等人提出的方法相比,提出的方法有更好的分类性能。

2　支持向量机

SVM 的主要思想可以用下面的 QP 问题来表示:

$$\begin{aligned}&\text{Min}\ \frac{1}{2}\boldsymbol{w'w}\\&\text{subject to}\ \boldsymbol{D}(\boldsymbol{Aw}-\boldsymbol{e}b)\geqslant\boldsymbol{e}\end{aligned}\tag{1}$$

式(1)中的 $\boldsymbol{D}=\begin{bmatrix}y_1&0&0&\cdots&0\\0&y_2&0&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&y_n\end{bmatrix}_{m\times m}$,$y_i$ 是第 i 个样本的类标,$\boldsymbol{A}\in\boldsymbol{R}^{m\times n}$。但是在大多情况下式(1)由于存在一些不能线性分开的样本点,所以仅仅用式(1)求解是不可行的,这样,可以向式(1)引入一个松弛因子 $\boldsymbol{\xi}$,得:

$$\begin{aligned}&\text{subject to}\ \boldsymbol{D}(\boldsymbol{Aw}-\boldsymbol{e}b)+\boldsymbol{\xi}\geqslant\boldsymbol{e}\\&\qquad\boldsymbol{\xi}\geqslant 0\end{aligned}\tag{2}$$

根据式(1)和式(2),可以得到一个最优化问题:

作者简介:徐金宝(1970-),男,讲师,研究方向:Java 新技术与 MIS 研制、数据挖掘;廖雷(1964-),男,副教授,主要研究方向:软件工具、软件自动化;业巧林(1983-),男,硕士,主要研究方向:机器学习、数据挖掘。

收稿日期:2008-12-16　**修回日期:**2009-05-21

$$\begin{aligned}&\text{Min}\ \frac{1}{2}\boldsymbol{w}'\boldsymbol{w}+C\boldsymbol{e}'\boldsymbol{\xi}\\&\text{subject to}\ \boldsymbol{D}(\boldsymbol{A}\boldsymbol{w}-\boldsymbol{e}b)+\boldsymbol{\xi}\geqslant\boldsymbol{e}\\&\qquad\qquad\boldsymbol{\xi}\geqslant 0\end{aligned}\tag{3}$$

式(3)中的正常数 C 是惩罚因子。

3 VSLSVM 算法

为了有效地说明提出的 VSLSVM 算法,首先讨论一下最小二乘 SVM(LS-SVM)[8]:

$$\begin{aligned}&\min_{(w,b,\xi)}\frac{1}{2}(\boldsymbol{w}'\boldsymbol{w})+\frac{C}{2}\sum_{i=1}^{n}\xi_i^2\\&\text{s.t}\ y_i[<\boldsymbol{w}'\boldsymbol{\varphi}(\boldsymbol{x}_i)>+b]=1-\xi_i,i=1,\cdots,n\end{aligned}\tag{4}$$

式(4)中 $\varphi(\cdot)$ 是非线性函数。 现在,重点讨论一下最优化问题(4)。为了达到较好的分类性能的目的,对优化问题(式(4))作如下修正:

$$\begin{aligned}&\min_{(w,b,\xi)}\frac{1}{2}(\boldsymbol{w}'\boldsymbol{w})+\frac{C}{2}\sum_{i=1}^{n}\xi_i^2\\&\sum_{i=1}^{n_1}\boldsymbol{w}'(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_1^{\varphi})(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_1^{\varphi})'\boldsymbol{w}+\\&\sum_{i=1}^{n_2}\boldsymbol{w}'(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_2^{\varphi})(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_2^{\varphi})'\boldsymbol{w}\\&\text{s.t}\ y_i[<\boldsymbol{w}'\boldsymbol{\varphi}(\boldsymbol{x}_i)>+b]=1-\xi_i,i=1,\cdots,n\end{aligned}\tag{5}$$

式(5)中 n_1、n_2 分别为+1 类样本和-1 类样本数目,$\boldsymbol{m}_1^{\varphi}=\frac{1}{n_1}\sum_{i=1}^{n_1}\boldsymbol{\varphi}(\boldsymbol{x}_i)=\frac{1}{n_1}\boldsymbol{\varphi}(\boldsymbol{x}_1)\boldsymbol{1}_{n_1}$,$\boldsymbol{m}_2^{\varphi}=\frac{1}{n_2}\sum_{i=1}^{n_2}\boldsymbol{\varphi}(\boldsymbol{x}_i)=\frac{1}{n_2}\boldsymbol{\varphi}(\boldsymbol{x}_2)\boldsymbol{1}_{n_2}$($\boldsymbol{1}_{n_i}$ 由 n_i 个 1 组成的列向量,这里 $i=1,2$),$n=n_1+n_2$; 优化问题(式(5))可以保证尽可能得到较大间隔的同时,保证了最小的类内散度[9]。问题(式(5))不易求解,因此做如下组合:

$$\begin{aligned}&\min_{(w,b,\xi)}\frac{1}{2}(\boldsymbol{w}'\boldsymbol{w})+\frac{C}{2}\sum_{i=1}^{n}\xi_i^2+\frac{1}{2}\boldsymbol{w}'\boldsymbol{M}^{\varphi}\boldsymbol{w}\\&\text{s.t}\ y_i[<\boldsymbol{w}'\boldsymbol{\varphi}(\boldsymbol{x}_i)>+b]=1-\xi_i,i=1,\cdots,n\end{aligned}\tag{6}$$

式(6)中 $\boldsymbol{M}=(\boldsymbol{M}_1+\boldsymbol{M}_2)=\sum_{i=1}^{n_1}(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_1^{\varphi})(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_1^{\varphi})'+\sum_{i=1}^{n_2}(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_2^{\varphi})(\boldsymbol{\varphi}(\boldsymbol{x}_i)-\boldsymbol{m}_2^{\varphi})'$,在文献[9]中,Mika 等人已经证明了特征空间中的权向量可有训练样本线映射表示:

$$w=\sum_{i=1}^{n}a_i\boldsymbol{\varphi}(\boldsymbol{x}_i)=\boldsymbol{\varphi}(\boldsymbol{X})\boldsymbol{a}\tag{7}$$

式(7)中 $\boldsymbol{a}=[a_1,a_2,\cdots,a_n]'$,如同文献[3][10],通过替换权向量 $\boldsymbol{w}$,对式(6)进行同样修改,如下:

$$\begin{aligned}&\min_{(w,b,\xi)}\frac{1}{2}(\boldsymbol{a}'\boldsymbol{a})+\frac{C}{2}\boldsymbol{\xi}^2+\frac{1}{2}\boldsymbol{a}'\boldsymbol{S}^{\varphi}\boldsymbol{a}\\&\text{s.t}\ \boldsymbol{D}(\boldsymbol{K}\boldsymbol{a}+\boldsymbol{e}b)=\boldsymbol{e}-\xi\end{aligned}\tag{8}$$

式中 $\boldsymbol{D}=\begin{bmatrix}y_1&0&0&\cdots&0\\0&y_2&0&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&y_n\end{bmatrix}_{m\times m}$,$y_i$ 对应与第 i 个样本的类标,e 是由 n 个 1 组成的列向量,$\boldsymbol{S}^{\varphi}=\boldsymbol{K}_{n\times n_1}[\boldsymbol{I}_1-\frac{1}{n_1}\boldsymbol{1}_1\boldsymbol{1}'_1]\boldsymbol{K}'_1+\boldsymbol{K}_2[\boldsymbol{I}_2-\frac{1}{n_2}\boldsymbol{1}_2\boldsymbol{1}'_2]\boldsymbol{K}'_2$,$\boldsymbol{S}^{\varphi}$ 由下面式子得到:

$$\begin{aligned}&\boldsymbol{w}'\boldsymbol{M}^{\varphi}\boldsymbol{w}=\sum_{i=1}^{2}\boldsymbol{a}'\boldsymbol{\varphi}(\boldsymbol{X})'(\sum_{j=1}^{n_i}(\boldsymbol{\varphi}(\boldsymbol{x}_j^i)-\boldsymbol{e}_i^{\varphi})(\boldsymbol{\varphi}(\boldsymbol{x}_j^i)-\\&\boldsymbol{e}_i^{\varphi})')\boldsymbol{\varphi}(\boldsymbol{X})\boldsymbol{a}=\sum_{i=1}^{2}\boldsymbol{a}'\boldsymbol{\varphi}(\boldsymbol{X})'(\boldsymbol{\varphi}(\boldsymbol{X}_i)\boldsymbol{\varphi}(\boldsymbol{X}_i)'-\\&\frac{1}{n_i}\boldsymbol{\varphi}(\boldsymbol{X}_i)\boldsymbol{1}_{n_i}\boldsymbol{1}'_{n_i}\boldsymbol{\varphi}(\boldsymbol{X}_i)')\boldsymbol{\varphi}(\boldsymbol{X})\boldsymbol{a}=\\&\sum_{i=1}^{2}\boldsymbol{a}'\boldsymbol{\varphi}(\boldsymbol{X})'(\boldsymbol{\varphi}(\boldsymbol{X}_i)[\boldsymbol{I}_{n_i}-\frac{1}{n_i}\boldsymbol{1}_{n_i}\boldsymbol{1}'_{n_i}]\boldsymbol{\varphi}(\boldsymbol{X}_i)')\boldsymbol{\varphi}(\boldsymbol{X})\boldsymbol{a}=\\&\boldsymbol{a}'(\boldsymbol{K}_{n\times n_1}[\boldsymbol{I}_1-\frac{1}{n_1}\boldsymbol{1}_1\boldsymbol{1}'_1]\boldsymbol{K}'_1+\boldsymbol{K}_2[\boldsymbol{I}_2-\frac{1}{n_2}\boldsymbol{1}_{n_2}\boldsymbol{1}'_{n_2}]\boldsymbol{K}'_{n\times n_2})\boldsymbol{a}\end{aligned}\tag{9}$$

形成问题(式(8))对应的 Lagrangian 对偶问题:

$$L(\boldsymbol{a},b,\boldsymbol{\xi},\boldsymbol{\lambda})=\frac{1}{2}\boldsymbol{a}'\boldsymbol{a}+\frac{C}{2}\boldsymbol{\xi}+\frac{1}{2}\boldsymbol{a}'\boldsymbol{S}^{\varphi}\boldsymbol{a}-\boldsymbol{\lambda}(\boldsymbol{D}(\boldsymbol{K}\boldsymbol{a}+\boldsymbol{e}b)-\boldsymbol{e}+\boldsymbol{\xi})\tag{10}$$

式(10)中 λ 为 Lagrangian 乘子,分别对上式中的变量 $\boldsymbol{a},b,\boldsymbol{\xi}$ 求偏导,并设为 0,于是可以得到如下的 KKT 条件:

$$\begin{aligned}&\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{a}}=0\Rightarrow(\boldsymbol{I}+\boldsymbol{S}^{\varphi})\boldsymbol{a}=\boldsymbol{D}\boldsymbol{K}\boldsymbol{\lambda}\\&\frac{\partial\boldsymbol{L}}{\partial b}=0\Rightarrow-\boldsymbol{e}'\boldsymbol{D}\boldsymbol{\lambda}=0\\&\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{\xi}}=0\Rightarrow\boldsymbol{\lambda}=C\boldsymbol{\xi}\\&\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{\lambda}}=0\Rightarrow\boldsymbol{D}(\boldsymbol{K}\boldsymbol{a}+\boldsymbol{e}b)=\boldsymbol{e}-\boldsymbol{\xi}\end{aligned}\tag{11}$$

式(11)中 $\boldsymbol{I}$ 是 $n\times n$ 维的单位矩阵,矩阵$(\boldsymbol{I}+\boldsymbol{S}^{\varphi})$是正定的,因为 $\boldsymbol{S}^{\varphi}$ 是半正定的,并且矩阵 $\boldsymbol{I}$ 是正定的。对(11)中最后一个等式中的 $\boldsymbol{a},\boldsymbol{\xi}$ 进行替换,于是可以得到:

$$\begin{bmatrix}0&\boldsymbol{D}\boldsymbol{e}'\\\boldsymbol{e}\boldsymbol{D}&\boldsymbol{H}\end{bmatrix}\cdot\begin{bmatrix}b\\\boldsymbol{\lambda}\end{bmatrix}=\begin{bmatrix}0\\\boldsymbol{e}'\end{bmatrix}\tag{12}$$

式(12)中的 $\boldsymbol{H}=\boldsymbol{D}\boldsymbol{K}(\boldsymbol{I}+\boldsymbol{S}^{\varphi})^{-1}\boldsymbol{D}\boldsymbol{K}+C^{-1}$。这个矩阵是对称正定的,由于 $\boldsymbol{D}\boldsymbol{K}(\boldsymbol{I}+\boldsymbol{S}^{\varphi})^{-1}\boldsymbol{D}\boldsymbol{K}$ 是对称半正定的,而 C^{-1} 是正定的,式(12)中的矩阵不是正定的,因此,问题(式(12))不易求解。为了较方便地求解式(12),可以修正式(12),得:

$$\begin{bmatrix}\boldsymbol{D}\boldsymbol{e}'\boldsymbol{H}^{-1}\boldsymbol{e}\boldsymbol{D}&0\\0&\boldsymbol{H}\end{bmatrix}\cdot\begin{bmatrix}b\\\boldsymbol{\lambda}+\boldsymbol{H}^{-1}\boldsymbol{e}\boldsymbol{D}b\end{bmatrix}=\begin{bmatrix}\boldsymbol{D}\boldsymbol{e}'\boldsymbol{H}^{-1}\boldsymbol{e}'\\\boldsymbol{e}'\end{bmatrix}\tag{13}$$

对于问题(式(13))来说,可以用文献[11-12]中提到的 CG 方法求解。但 CG 存在许多缺陷,如分类性能的不稳定性,由于不适合的终止因子而导致的过长的计算时间[13-14],为了避免 CG 的缺陷,采用下面方法对 VSLSVM 求解:

VSLSVM 算法:

(1)求解 H:首先,求解$(\boldsymbol{I}+\boldsymbol{S}^{\varphi})\boldsymbol{x}=\boldsymbol{D}\boldsymbol{K}$ 中的中间变量 $\boldsymbol{x}$,可以得到 $\boldsymbol{x}=(\boldsymbol{I}+\boldsymbol{S}^{\varphi})\backslash\boldsymbol{D}\boldsymbol{K}$[5],所以,$\boldsymbol{H}=\boldsymbol{D}\boldsymbol{K}\boldsymbol{x}+C^{-1}$。

(2)从方程 $\boldsymbol{H}\boldsymbol{\delta}=\boldsymbol{e}\boldsymbol{D}$ 和 $\boldsymbol{H}\boldsymbol{\rho}=\boldsymbol{e}'$ 中分别求解 $\boldsymbol{\delta},\boldsymbol{\rho}$,即:$\boldsymbol{\delta}=\boldsymbol{H}\backslash(\boldsymbol{e}\boldsymbol{D})$,$\boldsymbol{\rho}=\boldsymbol{H}\backslash\boldsymbol{e}'$。

(3)计算 $\boldsymbol{D}\boldsymbol{e}'\boldsymbol{H}^{-1}\boldsymbol{e}\boldsymbol{D}=\boldsymbol{D}\boldsymbol{e}'\boldsymbol{\delta}$,$b=\boldsymbol{\delta}'\boldsymbol{e}'/\boldsymbol{D}\boldsymbol{e}'\boldsymbol{H}^{-1}\boldsymbol{e}\boldsymbol{D}$,$\boldsymbol{\lambda}=\boldsymbol{\rho}-\boldsymbol{\delta}b$。

(4)计算 $\boldsymbol{a}=(\boldsymbol{I}+\beta\boldsymbol{S}^{\varphi})\backslash\boldsymbol{D}\boldsymbol{K}\boldsymbol{\lambda}$。

4 实验结果与分析

为了说明 VSLSVM 算法的有效性,在 UCI 数据库[15]上进行了实验。Suykens 的方法中的终止因子设置为 0.001,计算机是运行在装有 MATLAB 2008 的 Windows XP 系统上,计算机内

存为512 MB。

表 1 中显示了高斯核(Gauss)SVM、VSLSVM 和 Suykens 方法的比较结果。在实验中,参数 C 与核参数 σ 分别是在$\{10^i|i=-3,-2,\cdots,4\}$与$\{10^i|i=-4,-3,-2,\cdots,4\}$范围内调节获得。表 1 中结果表明,提出的方法有较好的分类性能,且优于 SVM 和 Suykens 的方法。例如,在数据集 PIDD 上 VSLSVM 得到了 76.1 的测试精度,而在 SVM 与 Suykens 提出的方法上的测试精度分别为 68.0 和 67.8,可以看出 VSLSVM 分类性能的有效性。

表 1 在 UCI 数据库中的 6 个数据集上非线性 SVM、Suykens 方法和 VSLSVM 的分类精度

数据集	分类精度/(%)		
	SVM	Suykens'method[11]	VSLSVM
Spect(267×45)	55.6	**92.0**	**92.0**
WBCD(699×10)	**100**	**100**	**100**
WDBC(569×32)	94.1	75.5	**95.5**
Yeast(1 484×8)	**70.6**	69.0	70.1
PIDD(768×8)	68.0	67.8	**76.1**
BSWD(625×4)	**96.3**	63.4	91.6
ISD(2 130×19)	98.9	**99.6**	**99.6**
Wine(1 473×9)	83.8	70.0	**92.3**
Monk2(432×6)	**82.0**	57.6	**82.0**
Monk1(432×6)	**86.8**	52.0	**86.8**
Pidd(768×8)	68.0	67.9	**70.8**

注:黑体表示最好性能

5 结论

提出一个简单的分类算法 VSLSVM,与标准 SVM 相比,该方法能得到较好的分类性能,它主要通过求解线性方程组求解,而不是求解凸规划问题。将类内散度的思想融入最小二乘支持向量机(LS-SVM)中,使得在获得最大间隔的同时,保证类内散度最小化。实验结果也验证了 VSLSVM 的有效性能。

参考文献:

[1] Vapnik V N.The nature of statistical learning theory[M].New York:Springer-Verlag,1995.

[2] Cherkassky V,Mulier F.Learning from data:Concept,theory and method[M].NY:JohnViley&sons,1997.

[3] Lee Yuh-Jye,Mangasarian O L.SSVM:A smooth support vector machine,Technical Report 99-03[R].Data Mining Institute,Computer Sciences Department,University of Wisconsin,Madison,Wisconsin,September 1999.

[4] 李国正,王猛,曾华军.支持向量机导论[M].北京:电子工业出版社,2004:96-100.

[5] Fung G,Mangasarian O L.Proximal support vector machine classifiers[C]//Provost F,Srikant R.Proceedings KDD-2001:Knowledge discovery and data mining.San Francisco,CA,New York:Asscociation for Computing Machinery,2001:77-86.

[6] Mangasarian O L,Musicant D R.Lagrangian support vector machines,Technical Report 00-06[R].Data Mining Institute,Computer Sciences Department,University of Wisconsin,Madison,Wisconsin,2001:161-177.

[7] Shewchuk J R.An introduction to the conjugate gradient method without the agonizing pain,technical report CS-94-125[R].Carnegie Mellon University,Pittsburgh,1994.

[8] Suykens J A K,Van Gestel,Brabanter T.De,et al.Least Squares Support Vector Machines[M].Singapore:World Scientific Publishing Co,2002.

[9] Mika S,Ratsch G,Weston J,et al.Fisher discriminant analysis with kernels[C]//Hu Y H,Larsen J,Wilson E.Neural Networks for Signal Procesing IX,1999:41-48.

[10] Mangasarian O L.Generalized support vector machines[C]//Smola A,Bartlett P,Scholkopf B.Advances in Large MarginClassi_ers.Cambridge,MA:MIT Press,2000.

[11] Suykens J A K,Lukas L,Van Dooren P,et al.Least squares support vector machine classifiers:A large scale algorithm[J].Neural Processing Letters,1999,9(3):293-300.

[12] Wei C,Chong J O,Keerthi S.An improved conjugate gradient scheme to the solution of least squares SVM[J].IEEE Transactions on Neutal Networks,2004,1(11).

[13] Lee Yuh-Jye,Mangasarian O L.SSVM:A smooth support vector machine,Technical Report 99-03[R].Data Mining Institute,Computer Sciences Department,University of Wisconsin,Madison,Wisconsin,1999-09.

[14] MATLAB.User's Guide[M/OL].The MathWorks,Inc,1994-2001.http://www.mathworks.com.

[15] Newman D J,Hettich S,Blake C L.UCI repository of machine learning databases[DB/OL].(1992).http://www.ics.uci.edu/~mlearn/MLRepository.html.

(上接 43 页)

[8] Papadimitirou S,Kitagawa H,Gibbons P B.LOCI:Fast outlier detection using the local correlation integral[C]//Proceedings of the 19th International Conference on Data Engineering,Bangalore.Los Alamitos:IEEE Computer Society,2003:315-326.

[9] Sanjay C,Sun Pei.SLOM:A new measure for local spatial outliers[J].Knowledge and Information Systems,2006,9(4):412-429.

[10] He Z,Xu X,Deng S.Discovering Cluster-based Local outliers[J].Pattern Recognition Letters,2003,24(9-10):1642-1650.

[11] He Z,Xu X,Huuug J Z,et al.Mining Class outliers:Concepts,algorithms and applications in CRM[J].Experts System with Applications,2004,27(4):294-299.