

解析中心机 ACM 收敛性的研究*

朱嘉钢^{1,2} 王士同²

(南京理工大学计算机科学与工程系 南京 210094)¹ (江南大学信息工程学院 无锡 214122)²

摘要 从理论上讲,支持向量机(SVM)的解对应于样本空间最大超球体的中心,而最新的由 T. B. Trafalis 等人提出的解析中心机(ACM)的解对应于样本空间最大超球体的解析中心。理论和实验表明,解析中心机的泛化性能要优于支持向量机,但其收敛性问题尚待解决。本文研究了解析中心机的收敛性问题,证明了只要满足一定的条件,解析中心机是收敛的。

关键词 解析中心机,收敛性,支持向量机

Research on Convergence of Analytic Center Machine

ZHU Jia-Gang^{1,2} WANG Shi-Tong²

(Nanjing University of Science and Technology, Nanjing 210094)¹ (Southern Yangtze University, Wuxi 214122)²

Abstract From theoretical viewpoint, Support Vector Machine (SVM) solution corresponds to the center of the largest hypersphere inscribed in the version space, and Analytic Center Machine (ACM) solution corresponds to the analytic center of the largest hypersphere inscribed in the version space. Theoretical and experimental results demonstrate that analytic center machine has better generalization performance than support vector machine. However, its convergence has not been investigated yet. In this paper, the convergence is studied and an important conclusion is derived that under certain conditions, analytic center machine can converge.

Keywords Analytic center machine, Convergence, Support vector machines

1 问题的提出

支持向量机(SVM)以其优越的泛化性能(generalization capability)在机器学习和优化领域日益受到重视并得到广泛应用,其解对应于样本空间(version space)最大超球体的中心。最新的研究表明,如果采用样本空间的其他可能的中心,譬如引力中心,则可以进一步提高泛化性能。但是,计算多面体引力中心的有效算法还很缺乏。一种可以用牛顿方法有效计算的样本空间中心是凸多面体的解析中心。于是, T. B. Trafalis 等人^[1]于2002年提出了解析中心机(Analytic Center Machine),并给出了相应的求解算法(以下简称 ACM 算法)和实验结果。结果表明,ACM 机的性能要优于支持向量机^[1]。最近,又有学者对解析中心机进行了进一步的研究,提出了一种精确的基于解析中心机的分类器^[2]和解析中心回归机。然而, T. B. Trafalis 等人并没有给出 ACM 算法收敛性的证明。另一方面, Platt 研究了支持向量机的快速训练问题,提出了 SMO(Sequential Minimization Algorithm)算法。S. S. Keerthi 等人则进一步提出了改进的 SMO,并证明了一个泛化的 SMO 算法的收敛性,其改进的 SMO 算法可作为其泛化算法的一个特例^[3~5]。但是,ACM 模型不是凸二次规划问题,因此 S. S. Keerthi 等人的算法及其收敛性结论不能用于 ACM 模型。

ACM 的收敛性问题如同 FCM(模糊聚类)和 SVM 的收敛性问题一样是非常重要的。本文对 ACM 算法的收敛性进行了研究,分析了 ACM 算法中添加模式后可行解域和最优

解的变化规律,得到了与 ACM 算法有关的几个结论,证明了在一定的条件下 ACM 算法是收敛的,并通过实例分析,给出了解析中心机收敛条件的直观解释。本文的内容组织如下:第2节简要介绍 T. B. Trafalis 等人的 ACM 机及其求解算法,第3节研究 ACM 算法中的模式加入与松弛变量的使用,第4节讨论 ACM 算法的收敛性,最后给出一个实例。

2 ACM 机及其求解算法

2.1 ACM 模型

T. B. Trafalis 等人提出的 ACM 模型^[1]如下:

$$\begin{aligned} \min \Phi(\alpha) &= -\sum_{j=1}^l \ln(k_j^T \alpha) \\ \text{s. t. } h(\alpha) &= \frac{1}{2} \alpha^T \alpha - 1 = 0 \end{aligned}$$

其中

$$K = \begin{bmatrix} y_1 k_{11} & y_2 k_{12} & \cdots & y_l k_{1l} \\ y_1 k_{21} & y_2 k_{22} & \cdots & y_l k_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ y_1 k_{l1} & y_2 k_{l2} & \cdots & y_l k_{ll} \\ y_1 & y_2 & \cdots & y_l \end{bmatrix}$$

$$k_j = \begin{bmatrix} y_j k_{1j} \\ y_j k_{2j} \\ \vdots \\ y_j k_{lj} \\ y_j \end{bmatrix}, \forall j=1, \dots, l$$

$k_{ij} = h(x_i, x_j)$, 是核函数,

* 国家自然科学基金资助(60225015);江苏省自然科学基金(BK2003017)、南京大学软件新技术国家重点实验室开放课题、江苏计算机技术重点实验室开放课题和教育部跨世纪优秀人才的资助。朱嘉钢 博士生,副教授,主要研究领域为人工神经网络、模式识别。王士同 教授,博士生导师,主要研究领域为人工神经网络、模式识别、计算机视觉、模糊系统等。

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_l \\ b \end{bmatrix}$$

ACM模型的几何意义是,将 $k_j^T \alpha = 0$ 视作一个过原点的超平面($k_j^T \alpha$ 又称为一个模式), l 个超平面的正向一侧构成了一个锥体(cone)。超球面 $\frac{1}{2} \alpha^T \alpha - 1 = 0$ 落入锥体的部分组成了 ACM 模型的可行解域。所求的解析中心 α^* 满足:在可行解域,所有超平面对这一中心的引力之和最小。

在解出 α^* 后,对一个新样本 x 的决策函数为:

$$f(x) = \text{sign}(\sum_{i=1}^l \alpha_i k_i(x, x) + b)$$

可以验证,ACM 模型是凸规划问题。

2.2 ACM 求解算法

2.2.1 牛顿迭代公式

为了求解 ACM 模型,利用 KKT 优化条件,有

$$\nabla_{\alpha} L(\alpha) = \nabla_{\alpha} \Phi(\alpha) + u \cdot \nabla_{\alpha} h(\alpha)$$

$$h(\alpha) = \frac{1}{2} \alpha^T \alpha - 1 = 0$$

其中 u 是拉格朗日系数。

定义一个新变量 $Z(\alpha, u)$,并将上面优化条件写成矩阵形式,有

$$Z(\alpha, u) = \begin{bmatrix} \nabla_{\alpha} L(\alpha) \\ h(\alpha) \end{bmatrix} = \begin{bmatrix} \nabla_{\alpha} \Phi(\alpha) + u \cdot \nabla_{\alpha} h(\alpha) \\ \frac{1}{2} \alpha^T \alpha - 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

并得牛顿迭代公式:

$$Z(\alpha_k, u_k) + \nabla Z(\alpha_k, u_k) \begin{bmatrix} \alpha - \alpha_k \\ u - u_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (1)$$

其中,

$$Z(\alpha_k, u_k) = \begin{bmatrix} \nabla_{\alpha} \Phi(\alpha_k) + u_k \cdot \nabla_{\alpha} h(\alpha_k) \\ \frac{1}{2} \alpha_k^T \alpha_k - 1 \end{bmatrix}$$

$$\nabla Z(\alpha_k, u_k) = \begin{bmatrix} \nabla^2 L(\alpha_k, u_k) & \nabla h(\alpha_k, u_k) \\ \nabla h(\alpha_k, u_k)^T & 0 \end{bmatrix}$$

若引入对角阵 $S \in R^{l \times l}$, $s_j = k_j^T \alpha_k$, 和 $e = (1, 1, \dots, 1) \in R^l$, 则上述迭代公式又可以进一步写成易于计算的形式:

$$Z(\alpha_k, u_k) = \begin{bmatrix} -KS^{-1}e + u_k \cdot \alpha_k \\ \frac{1}{2} \alpha_k^T \alpha_k - 1 \end{bmatrix} \quad (2)$$

$$\nabla Z(\alpha_k, u_k) = \begin{bmatrix} KS^{-2}K^T + u_k \cdot I_{\alpha_k} \\ \alpha_k^T & 0 \end{bmatrix} \quad (3)$$

2.2.2 ACM 算法

如前所述, T. B. Trafalis 等人提出了以下的 ACM 算法(行号是本文作者加的):

- 1, 第 1 步: 初始化
- 2: l —— 模式的个数
- 3: j —— 当前模式, $j=1, \dots, l$
- 4: i —— 用于标识从 $1, \dots, j$ 的所有模式的下标
- 5: α_0 —— 初始起始点, 由元素为“1”组成的向量
- 6: u —— 用于投影到超球的初始拉格朗日系数
- 7, 第 2a 步, 求中心并添加模式
- 8: 置 $j=1$
- 9: 由 $k_j^T = (y_j k_{1j}, y_j k_{2j}, \dots, y_j k_{lj}, y_j)$ 计算最新的 k_j
- 10: 松弛变量 $s_j = k_j^T \alpha$, $\forall i=1, \dots, j$, 并置 $S = \text{diag}(s_i)$, $\forall i=1, \dots, j$
- 11: 置 $K = (k_1, \dots, k_j)$
- 12: if $s_j \leq 0$
- 13: 转第 2b 步: 纠正
- 14: else

- 15: 转第 2c 步: 更新
- 16: 第 2b 步: 纠正
- 17: 置 $total = -1.1 \cdot s_j$ 且 $s_{relax} = -0.1 \cdot s_j$ (问题的松弛量)
- 18: 置 $S_{relax} = \text{diag}(s_1, s_2, \dots, s_{j-1}, s_{relax})$
- 19: while ($s_j = k_j^T \alpha \leq 0$)
- 20: for newtoncounter = 1 to 5
- 21: 用迭代公式(2)和(3)计算 Z 和 ∇Z
- 22: 用牛顿近似公式(1)计算新的 (α, u)
- 23: if ($s = K^T \cdot \alpha > 0$)
- 24: 转第 2c 步: 更新
- 25: else
- 26: 更新第 j 个松弛变量, $s_j = total + s_j$ (s_j 仍是负的!)
- 27: 检查所有其它松弛变量 (s_1, \dots, s_{j-1}) 的可行性
- 28: while $\min(s_1, \dots, s_{j-1}) \leq 0$
- 29: 将原牛顿迭代步长减半
- 30: 计算 $s = K^T \cdot \alpha$ 并替换 $s_j, s_j = total + s_j$
- 31: endwhile
- 32: endfor
- 33: 置 $total = -1.1 \cdot s_j$ (构造新的松弛量)
- 34: 替换 $s_j, s_j = 0.1 \cdot s_j$
- 35: endwhile
- 36: 第 2c 步: 更新
- 37: 用迭代公式(2)和(3)计算 Z 和 ∇Z
- 38: 用牛顿近似公式(1)计算新的 (α, u)
- 39: $j = j + 1$, 转第 1 步

以上只是 ACM 算法的框架,完整的描述见文[1]。

如同研究 SVM 优化问题一样,一个重要问题是上述算法的收敛性问题。下面 3 节将研究 ACM 算法的收敛性问题。

3 模式加入与松弛变量的使用

T. B. Trafalis 等人将 $k_j^T, j=1, \dots, l$ 称为一个模式, 而将 $s_j = k_j^T \alpha, j=1, \dots, l$ 称为一个松弛变量。算法的思想是: 先求只有一个模式时 ACM 的最优解, 再添加一个模式, 求得含有两个模式时 ACM 的最优解。如此继续下去, 直至求得含有 l 个模式的 ACM 的最优解。

由于函数 $\ln(k_j^T \alpha)$ 要求每个模式 $k_j^T \alpha > 0, j=1, \dots, l$, 因此在每添加一个模式时, 如果原有的解不能使相应的松弛变量大于 0, 则需要进行“纠正”。“纠正”的思想是: 设一个常量 $\tau > 0$, 使 $k_j^T \alpha + \tau > 0$ (这里 k_j^T 是新添加的模式, $k_j^T \alpha$ 是对应的松弛变量值)。利用迭代不断求得新的解, 使 $k_j^T \alpha$ 的值不断增大(迭代时当然还要保证原有的松弛变量大于 0)。调整常量 τ , 使其不断减小, 直至 $k_j^T \alpha > 0$ 且 $\tau = 0$, 至此“纠正”过程结束。

假定所取的样本可以使 ACM 模型有解并使牛顿迭代公式(1)收敛, 我们有以下结论。

引理 1 考虑 2.1 的 ACM 模型

$$\min \Phi(\alpha) = -\sum_{j=1}^l \ln(k_j^T \alpha)$$

$$\text{s. t. } h(\alpha) = \frac{1}{2} \alpha^T \alpha - 1 = 0$$

若记 ACM_i 代表模型

$$\min \Phi(\alpha) = -\sum_{j=1}^i \ln(k_j^T \alpha), i=1, \dots, l-1$$

$$\text{s. t. } h(\alpha) = \frac{1}{2} \alpha^T \alpha - 1 = 0$$

记 α_i^* 是 ACM_i 的最优解, 则有

$$\sum_{j=1}^i \ln(k_j^T \alpha_i^*) \geq \sum_{j=1}^{i+1} \ln(k_j^T \alpha_{i+1}^*)$$

证明: 当 $l=i$ 时, α 的可行解域 $F_i = \{\alpha | (\frac{1}{2} \alpha^T \alpha - 1 = 0) \wedge (k_j^T \alpha > 0, j=1, \dots, i)\}$; 而当 $l=i+1$ 时, α 的可行解域 $F_{i+1} = \{\alpha | (\frac{1}{2} \alpha^T \alpha - 1 = 0) \wedge (k_j^T \alpha > 0, j=1, \dots, i+1)\}$ 。对任意的 $\alpha \in F_{i+1}$, 必有 $\alpha \in F_i$, 故 $F_i \supset F_{i+1}$ 。于是, 若 $\sum_{j=1}^i \ln(k_j^T \alpha_i^*) < \sum_{j=1}^{i+1} \ln(k_j^T \alpha_{i+1}^*)$

$(k_j^T \alpha_{i+1}^*)$, 这与 α_i^* 是 ACM_i 的最优解矛盾。故必有 $\sum_{j=1}^i \ln(k_j^T \alpha_i^*) \geq \sum_{j=1}^i \ln(k_j^T \alpha_{i+1}^*)$ 。

引理 1 表明, 在求得含有 i 个模式的 ACM 的最优解后, 无论新添加第 $i+1$ 个模式后 ACM 的最优解如何变化, 原有的 i 个松弛变量值的对数之和不会增大。

引理 2 设 α_i^* 是 ACM_i 的最优解, 但 $k_{i+1}^T \alpha_i^* < 0$ 。设常数 τ , 使 $k_{i+1}^T \alpha_i^* + \tau > 0$ 。将 $k_{i+1}^T \alpha_i^* + \tau$ 代替 $k_{i+1}^T \alpha_i^*$ 加入到模型 ACM_i 中, 用式(1)、(2)和(3)进行迭代, 记 $\alpha_{i+1, new}$ 为由 α_i^* 迭代后的值。若 $k_{i+1}^T \alpha_i^* > k_{i+1}^T \alpha_{i+1, new}$, 则有 $k_{i+1}^T \alpha_{i+1, new} > k_{i+1}^T \alpha_i^*$ 。

证明: 令 $\Phi_{i+1}(\alpha_{i+1}) = -\sum_{j=1}^i \ln(k_j^T \alpha_{i+1}) - \ln(k_{i+1}^T \alpha_{i+1} + \tau)$ 经迭代后有

$$\Phi_{i+1}(\alpha_{i+1, new}) \leq \Phi_{i+1}(\alpha_{i+1})$$

注意到这里最初的 α_{i+1} 就是 α_i^* , 且又已知 $k_{i+1}^T \alpha_i^* > k_{i+1}^T \alpha_{i+1, new}$, 则 $\alpha_i^* \neq \alpha_{i+1}$, 所以 $\Phi_{i+1}(\alpha_{i+1, new}) < \Phi_{i+1}(\alpha_{i+1})$, 即

$$-\sum_{j=1}^i \ln(k_j^T \alpha_{i+1, new}) - \ln(k_{i+1}^T \alpha_{i+1, new} + \tau) < -\sum_{j=1}^i \ln(k_j^T \alpha_{i+1}) - \ln(k_{i+1}^T \alpha_{i+1} + \tau)$$

由引理 1 即得

$$k_{i+1}^T \alpha_{i+1, new} > k_{i+1}^T \alpha_i^*$$

引理 2 表明, 只要满足条件 $k_{i+1}^T \alpha_{i+1}^* > k_{i+1}^T \alpha_i^*$, 即只要新添加模式的松弛变量对应于新最优解(含有 $i+1$ 个模式的最优解)的值大于原最优解(含有 i 个模式的最优解)的值, 上述“纠正”过程总能使新松弛变量的值增加。

引理 3 对 ACM 模型, 若 $k_j^T \alpha' > 0, j=1, \dots, l$, 则存在常数 $1 \geq \lambda > 0$, 使

$$\begin{bmatrix} \alpha'_{new} \\ u'_{new} \end{bmatrix} = \begin{bmatrix} \alpha' \\ u' \end{bmatrix} - \lambda \frac{Z(\alpha', u')}{\nabla Z(\alpha', u')}$$

满足

$$k_j^T \alpha'_{new} > 0, j=1, \dots, l$$

证明: 设 F 为可行解域。先取 $\lambda=1$ 进行迭代。若 $\alpha'_{new} \in F$, 则有 $k_j^T \alpha'_{new} > 0, j=1, \dots, l$ 。若 $\alpha'_{new} \notin F$, 但由 $k_j^T \alpha' > 0, j=1, \dots, l$ 可得 $\alpha' \in F$ 。设 f 是 F 边界上与 α' 之差的范数最小的点, 则存在 $\delta = \{\alpha \mid 0 < \|\alpha - \alpha'\| < \|f - \alpha'\|\}$ 。取 $\alpha'' \in \delta$, 则有 $k_j^T \alpha'' > 0, j=1, \dots, l$ 。若令 $\lambda = \frac{\|\alpha'' - \alpha'\|}{\|\alpha'_{new} - \alpha'\|}$, 则此 λ 就是满足上述条件的常数之一。

引理 3 表明, 如果迭代后出现某些松弛变量小于 0 的情形, 只要适当缩小迭代步长, 就可使迭代后所有的松弛变量大于 0。

上述 3 引理构成了 ACM 算法收敛性证明的基础。

4 ACM 算法的收敛性

定理 ACM 算法将在有限步内正确终止, 且收敛到 ACM 模型的最优解, 如果满足:

- (1) 不出现 $s_j = 0$ 的情况;
- (2) 确保 $k_j^T \alpha_i^* > k_j^T \alpha_{i-1}^*$ 。

证明:

先证 ACM 算法在满足上述 3 条件前提下将在有限步内结束。

显然, 如果算法的第 2b 步能在有限步内结束, 则第 2a 步、第 2b 步和第 2c 步将循环 l 次结束 (j 是循环变量, $j=1, \dots, l$)。

现在来证明第 2b 步能在有限步内结束。第 2b 步有三重

循环: 外层循环即第 19 行的 while 循环, 中间层循环即第 20 行的 for 循环, 内层循环即第 28 行的 while 循环。

内循环的结束条件是 $\min(s_1, \dots, s_{j-1}) > 0$ 。考察第 26 行和第 30 行的语句 $s_j = total + s_j$, 其作用是将 $total + s_j$ 作为松弛变量的新值。内层 while 循环的作用是, 如果 $\min(s_1, \dots, s_{j-1}) \leq 0$, 则反复缩小迭代步长。由于在执行第 21 行“用迭代公式(2)和(3)计算 Z 和 ∇Z ”和第 22 行“用牛顿近似公式(1)计算新的 (α, u) ”之前, α 的值满足 $k_j^T \alpha > 0, j=1, \dots, i, i$, 和 $s_{relax} = k_{i+1}^T \alpha + total > 0$ 。由引理 3 可知, 在不断缩小迭代步长的过程中, 总是存在一个 α' , 满足 $k_j^T \alpha' > 0, j=1, \dots, i$, 和 $s_{relax} = k_{i+1}^T \alpha' + total > 0$ 。注意到条件 $k_j^T \alpha > 0, j=1, \dots, i$, 与 $\min(s_1, \dots, s_{j-1}) > 0$ 等价, 故在每次中间层 for 循环中, 内层 while 循环可在有限步内结束。于是, 在每次外层 while 循环中, 中间层 for 循环至多循环 5 次后也将结束。

外层 while 循环的结束条件是 $s_j = k_j^T \alpha > 0$ 。由定理的后 2 个条件和引理 2 可知, 每次循环后 $s_j = k_j^T \alpha$ 总会增加, 而 $total$ 的值(相当于引理 2 中的) τ 不断接近 0。假如已加入了 i 个模式, 即 $j=i$, 则 $s_j = k_j^T \alpha$ 将收敛到 $s_i = k_i^T \alpha_i^* > 0$ 。所以, 外层 while 循环也将在有限步内结束, 即第 2b 步将在有限步内结束。

综上所述, ACM 算法将在有限步内结束。

再证明 ACM 算法将收敛到 ACM 模型的最优解。

在 ACM 算法中, 第 1 步保证被一一添加处理的模式个数为 1 个, 第 2a 步对当前模式执行添加操作, 第 2b 步保证当前的解在可行解域内, 第 2c 步对当前的解用牛顿法进行一次迭代。经过迭代的解可以看作是当前的最优解的逼近, 且收敛于最优解。对于每次 j 的增量, ACM 算法都求得 ACM_j 的最优解 α_j^* 的逼近, 且收敛于 ACM_j 的最优解。由于 $j=1, \dots, l$, 故 ACM 算法最终求得 ACM_l 的最优解的逼近, 且收敛于 ACM_l 的最优解。又由于 ACM_l 模型就是 ACM 模型, 故 ACM 算法收敛于 ACM 模型最优解。

现在来讨论当定理的条件不满足时对算法的影响。

第一, 当第 12 行 if 语句和第 19 行 while 语句的条件中出现了 $s_j = 0$ 的情况。由于接着执行“置 $total = -1.1 \cdot s_j$ 且 $s_{relax} = -0.1 \cdot s_j$ ”的操作, 此时有 $s_{relax} = -0.1 \cdot s_j = 0, \frac{1}{s_{relax}}$ 无定义。在算法的实际运行中, 若 s_{relax} 的值下溢而被当作机器零处理, 将产生中断, 算法不能正常结束。

第二, 当经第 17 行和第 18 行处理后, 出现了 $k_j^T \alpha_i^* \leq k_j^T \alpha_{i-1}^*$ 的情况。由于 $k_j^T \alpha_{i-1}^* < 0$, 则无论如何迭代, 总有 $k_j^T \alpha_i^* < 0$, 即第 2b 步中的外层 while 循环的两个结束条件 $s_j = k_j^T \alpha \leq 0$ 和 $s = K^T \cdot \alpha > 0$ 总是得不到满足, 将进入死循环。当第 33 行和第 34 行处理后, 出现了 $k_j^T \alpha_i^* \leq k_j^T \alpha_{i-1}^*$ 的情况时, 也有相同的结果。

第三, 在执行了“置 $total = -1.1 \cdot s_j$ 且 $s_{relax} = -0.1 \cdot s_j$ ”的操作后, 是否一定能保证 $k_j^T \alpha_i^* > k_j^T \alpha_{i-1}^*$? Theodore B. Trafalis 等人给出了几何意义。从直观上看, 当 $s_{relax} > 0$ 越小, 条件 $k_j^T \alpha_i^* > k_j^T \alpha_{i-1}^*$ 越能得到满足。但这仍然有待于证明。此外, 是否还有其他的方法也能满足 $k_j^T \alpha_i^* > k_j^T \alpha_{i-1}^*$, 也可以研究。

5 实例分析

为了说明上述结论, 下面给出一个实验结果, 进一步对 ACM 的收敛性进行分析。

用 IRIS 数据集作为实验数据。IRIS 数据集是已建立的用于演示分类算法性能的数据集。IRIS 数据集含有 150 个样本,包括三类对象的 4 个特征。这三类对象分别是 Setosa, Versicolor 和 Viginica,其样本点在图 1 中分别用“+”,“o”和“*”表示;4 个特征分别是 Sepal length, Sepal width, Petal length 和 Petal width(单位是 cm)。可以根据这 4 个特征将 IRIS 数据集分类。出于可视化的考虑,取 IRIS 数据集的两个特征 Petal length 和 Petal width 进行分类。由于 ACM 是二类分类器,为便于观察分类效果,分类时将 Versicolor 作为一类,将 Setosa 和 Viginica 作为另一类。略去了 IRIS 数据集的两个特征后,IRIS 数据集的 150 样本点出现了部分重复,分类前去掉这些重复的样本。经实验确定,核函数取指数核函数 $K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$,其中 $\sigma=1.3$ 。图 1 给出了 IRIS 数据集的数据分布和用 ACM 分类的结果,其中虚线是分类决策线。

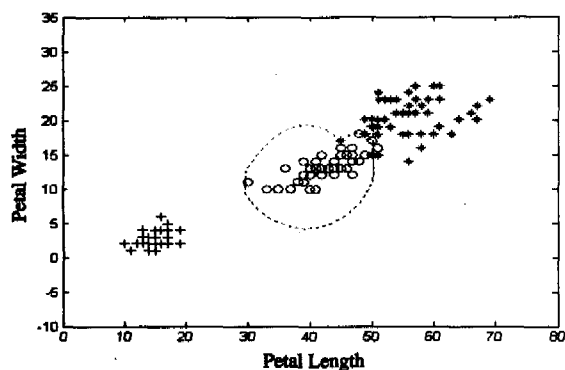


图 1 ACM 对 IRIS 数据的分类结果

我们还分别用二阶多项式核函数、高斯核函数和双曲正弦核函数进行分类,结果表明部分核函数的确可以使 ACM 发散。在使 ACM 收敛的核函数中,分类效果也不一样。限于篇幅,未将这些结果一一列出。

实验中显示,ACM 发散有两种情形:一是使 ACM 算法第 21 行中的矩阵 Z 成为奇异阵,从而使牛顿迭代不收敛,而

我们在证明 ACM 收敛时假定牛顿迭代是收敛的;二是经过 ACM 算法中从第 19 行到第 35 行的 while 循环, $k^T a_i$ 的值没有变化,这不满足上述定理的第二个条件。实验中还显示,ACM 收敛性与样本点集的分布、核函数及其参数的选择有关,对于给定的样本点集,如何选择合适的核函数及其参数,使得 ACM 收敛并有理想的分类效果,还有待于进一步研究。

结束语 T. B. Trafalis 等人提出了解析中心机 ACM 的方法和相应的求解算法。实验结果表明,ACM 机的性能要优于支持向量机。而 S. S. Keerthi 等人的泛化的 SMO 算法及其收敛性结论不能用于 ACM 模型。本文对 ACM 算法的收敛性进行了研究,证明了在一定的条件下 ACM 算法是收敛的,实例分析也支持了这一结论。此外,除了解析中心机外,是否还存在其它形式的中心机模型及其有效算法?另一个值得研究的问题是,当样本非常大时如何发展 ACM 优化问题的有意义的近似算法。作者目前正在研究这方面的问题。

参考文献

- 1 Trafalis T B, Malyscheff A M. An Analytic Center Machine. Machine Learning [J], 2002, 46: 203~223
- 2 曾凡仔,岳建海,裘正定. DRC-ACM:一种精确的基于解析中心的分类器. 计算机研究与发展[J], 2004, 41(5): 802~806
- 3 Keerthi S S. Convergence of a Generalized SMO Algorithm for SVM Classifier Design. Machine Learning [J], 2002, 46: 351~360
- 4 James Tin-Yau Kwok. The Evidence Framework Applied to Support Vector Machines. IEEE Transactions on Neural Networks [J], 2000, 11(5)
- 5 James Tin-Yau Kwok. Moderating the Outputs of Support Vector Machine Classifiers. IEEE Transactions on Neural Networks [J], 1999, 10(5)
- 6 边肇祺,张学工,等. 模式识别(第二版)[M]. 北京:清华大学出版社, 2002
- 7 关治,陆金甫. 数值分析基础[M]. 北京:高等教育出版社, 2002
- 8 钱颂迪,等. 运筹学(修订版)[M]. 北京:清华大学出版社, 1998

(上接第 181 页)

较和分析,使得格、超树等的有关性质在该模型中得到完美的结合,并且由超树 Stree 的几何特性和逻辑语言的代数性, $\langle Stree, U, L_{+base} \rangle$ 能把几何的直观性和代数的推理性较好结合在一起,为 KDD 的研究提供了一个统一而通用的理论框架。

结束语 不管是对 KDD 算法还是理论研究,探讨 KDD 的建模问题无疑都是很重要的。目前,从已有的文献看,在这方面的研究所取得的成果还是非常有限的。本文以粒度计算理论为工具,研究了这个问题,给出了一种“全粒度空间+正基语言”的粒度计算模型,并找出该模型中的一些关系定理以及导出了它的若干性质,最后把 KDD 任务归结为基于该模型的粒度计算问题。今后,我们将进一步对其完善和补充,并用以指导算法设计,真正推向实用化。

参考文献

- 1 杨炳儒,江亚东,申江涛. 基于双库协同机制的 KDD* 及其软件

实现. 系统工程与电子技术, 2000, 22(6): 69~72

- 2 Yao Y Y. On Modeling Data Mining with Granular Computing. In: 25th Annual International Computer Software and Applications Conf. (COMPSAC'01), October 08 - 12, Chicago, Illinois, 2001. 638~643
- 3 Pawlak Z. Rough Sets—Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, 1991
- 4 蒙祖强. 基于分类模型的知识发现机理和方法研究; [博士学位论文]. 长沙:中南大学信息科学与工程学院, 2004
- 5 张铃,张钊. 模糊商空间理论(模糊粒度计算方法). 软件学报, 2003, 14(4): 770~776
- 6 蒙祖强,蔡自兴. 一种面向个性化知识发现的属性约简算法. 小型微型计算机系统, 2005, 26(2): 209~213
- 7 Yao Y Y, Liao C J. A generalized decision logic language for granular computing. In: FUZZ-IEEE'02 in The 2002 IEEE World Congress on Computational Intelligence, 2002. 1092~1097