

一种基于构造性核覆盖的聚类算法

李丽芳,周鸣争

(安徽工程科技学院 计算机科学与工程系,安徽 芜湖 241000)

摘 要:基于构造性核覆盖学习方法的思路,提出了一种构造性核覆盖聚类算法。首先将原空间的待分类样本映射到一个高维的特征空间中,使得样本变得线性可分,然后在核空间采用构造性覆盖方法进行覆盖领域的构造,这组领域能将相似度小的样本分割开来,将相似度大的样本聚合在一起,通过定义一定的相似度度量标准和目标函数,达到聚类的效果。仿真实验也验证了该方法的有效性和可行性。

关键词:聚类分析;核覆盖;核函数;特征空间

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2009)01-0088-04

A Clustering Algorithm Based on Constructive Kernel Covering Algorithm

LI Li-fang, ZHOU Ming-zheng

(Dept. of Computer Sci. & Eng., Anhui University of Science and Technology, Wuhu 241000, China)

Abstract: The idea based on the constructive kernel covering study means, puts forward a constructive kernel covering clustering algorithm. Firstly, it maps the awaiting sort samples of original space to a high dimensional feature space, makes the samples linear separable. Then, in kernel space, the algorithm uses constructive covering method to construct covering domain. Every domain divides the minor semblance samples, makes the major semblance samples converge. By defining definite semblance standard and target function, it can get the clustering effect. Emulation experiment proved the validity and the feasibility of this algorithm.

Key Words: clustering analysis; kernel covering; kernel function; feature space

0 引言

聚类就是根据某种相似性准则将样本空间分成多个子空间,使每个子空间内部样本点尽可能相似,不同子空间内样本点之间差异尽可能大,其实质是寻找隐藏在数据中不同的数据模型,是一个无监督学习过程,能够实现样本空间的盲分类。其特点是输入空间的样本没有期望输出,聚类的过程完全依赖于样本之间的特征差别。聚类已广泛应用于统计、机器学习、模式识别、数据分析等领域,并越来越受重视。目前比较经典的聚类方法有传统的C-均值方法和模糊C-均值聚类方法^[1,2],这些方法都没有对样本的特征进行优化,而是直接利用样本的特征进行聚类。这样上述这些方法的有效性很大程度上取决于样本的分布情况。如果

一类样本散布较大,而另一类散布较小的话,这些方法的效果就比较差。如果样本分布更加混乱,则聚类的结果就会面目全非。

Girolami^[3]和张莉、焦李成等^[4]在将核方法应用于聚类算法方面做了开创性工作,他们通过把模式空间的数据非线性映射到高维特征空间,增加了模式的线性可分概率,即扩大模式类之间的差异,在高维特征空间达到线性可聚的目的。但这些方法存在有核函数的选取和核函数参数的优化以及聚类类别数必须事先确定的问题。

构造性核覆盖算法^[5]是一种将神经网络中的构造性学习方法与核函数法相结合,可处理高维海量数据的学习方法。其主要特点是在对具体数据的处理过程中,能同时得到系统结构的核参数。即系统的结构是在处理数据过程中逐步构造的,而不是在学习之前事先给定的。为了克服已有各种C-均值聚类方法所存在的不足,提高聚类的效果,我们基于构造性核覆盖学习方法的思路,提出了一种构造性核覆盖C-均值聚类方法,首先将原空间的待分类样本映射到一个高维

收稿日期:2008-04-18

基金项目:国家自然科学基金资助项目(60574028);安徽省自然科学基金资助项目(070412058);安徽高校省级自然科学研究重点项目(KJ2007A046)

作者简介:李丽芳(1983-),女,硕士研究生,研究方向为数据挖掘;周鸣争,教授,研究方向为人工智能与网络安全。

的特征空间(核空间)中,使得样本变得线性可分(或近似线性可分),然后在核空间采用构造性覆盖方法进行覆盖领域的构造,自动产生聚类类别数,通过定义一定的相似度量标准和目标函数,得到各类别的中心模值,由于利用构造性核覆盖算法使得聚类方法简单,可解释性强,算法的收敛速度也较快。仿真实验也验证了该方法的有效性和可行性。

1 构造性核覆盖聚类

1.1 构造性核覆盖算法

1.1.1 覆盖算法与核学习方法

1998年张铃等人提出了一种新的M-P神经元的几何意义解释^[6]——“球”领域模型,并以此给出了一种构造性的覆盖算法。

M-P神经元是一个 n 个输入、单输出的元件,其输入与输出的关系为:

$$Y = \text{Sgn}(\sum w_i x_i - \alpha) \quad (1)$$

若令 $\sum w_i x_i - \alpha = 0$,则此式表示为一个超平面方程。于是从几何意义上可将神经元看成是一个空间分类器,即落在正半空间的点对应的输出为1,落在负半空间上的点对应的输出为-1。若设输入样本的长度相等,即输入样本分布在 $n+1$ 维空间的某个球面 S^n 上(其中心点在原点,半径为 R)。那么这时 $(W * x - \theta) > 0$ (其中 W 是权向量, θ 是阈值),就表示球面上落在由超平面 P (其方程为 $(W * x - \theta) = 0$)所分割的正半空间的部分,这个部分恰好是球面上的某个“球形领域”。若取 W 与 x 等长,则这个“球形领域”的中心恰好是 W ,其半径为:

$$r(\theta) = \arccos(\theta/R^2) \quad (2)$$

若令 $\sigma'(x) = \begin{cases} 1, & \text{当 } x > 0 \\ 0, & \text{其它} \end{cases}$

且取神经元的激励函数为 $\sigma' = (W * x - \theta)$,则一个神经元的激励函数正好是它所代表的球面上“球形领域”的特征函数,这样,就将神经元与球面上的球形领域对应起来,利用神经元的这种几何意义,就能非常直观地进行神经网络的各种研究。

由上面给出的神经元的几何意义得知,构造一个网络,对给定的样本集能进行符合要求的聚类,等价于求出一组领域,对给定样本集 K 中的点,能按聚类的要求用领域覆盖将它们分割开来。这样,就将原先基于搜索机制的学习方法转变成构造性的学习方法。从而为处理海量数据提供了一种切实可行的方法。

当给定的输入向量的长度不相等时,可用下面给出的方法,将它变换成长度相等的情况。设输入的定义域为 n 维空间中的有界集合 D ,令 S^n 是 $n+1$ 维空间中

的 n 维的超球面,作变换 $T: D \rightarrow S^n, x \in D$

$$T(x) = (x, \sqrt{(d^2 - |x|^2)}) \quad (3)$$

其中 $d \geq \max\{|x|, x \in D\}$ 。

覆盖算法通过反复迭代求取最优的“球领域”覆盖中心,以使得球领域覆盖数最少。该算法的主要特点是,在对具体数据处理过程中,把求解样本集 S 的 k 类聚类问题转化成在样本空间构造覆盖簇 $\{C_i\}$,使每个覆盖 C_i 只盖住同一类点。它具有一次处理多分类问题的优点。但该算法存在以下两点不足:

(1)最终覆盖的质量受先验知识和初始种子样本的选择以及样本分布的影响较大;

(2)该算法相对复杂,实现比较困难。

核学习算法^[7],对线性可分情况给出一个用规划方法解得的最大间隔解。对线性不可分情况提出用核函数,将原问题映射到高维空间,在这个新空间中求取最优线性分类面,其所求得的分类函数形式上类似于一个神经网络,其输出是若干中间层节点的线性组合,而每一个中间层节点对应于输入样本与一个支持向量的内积。它不象传统方法,先试图将原输入空间降维(即进行特征选择和特征变换),而是设法将输入空间升维,以求在高维空间中,使问题变得线性可分。

从几何意义上看,核方法中求最优分类面问题就是在对应的核函数类中,求得一个其零值等高线是两类的分界线以及样本集到边界的距离最大的函数。也就是求划分边界线的问题。如果能将边界线“附近”的点分开,那么其它的点就自然而然地被分开。如果沿零等高线的两边,以最大间隔为宽度,划一条与之“平行”的线,即得到一条“边界河”,那么支持向量就必落在边界河的河沿上。由于不同的核函数所构成的“边界河”是不同的,因此不同类型的SVM利用的支持向量也是不同的。

支持向量机方法是到目前为止,统计学习理论最成功的实现,它是在统计学习理论的VC维理论和结构风险最小原理基础上,尽量提高学习机的泛化能力,能够较好地解决小样本、非线性和高维数的非线性分类问题。但是对SVM方法中存在的核函数选取问题(特别是核函数的参数选取问题)以及因要求二次规划而引起计算量大的问题,目前均未获得突破性的解决。

通过上述的分析,可以看出覆盖算法与核函数方法有如下关系:

(1)对于二分类问题,设覆盖法得到的输出为 $F(y)$,令 $K(x, y) = \langle T(x), T(y) \rangle$,则 $K(x, y)$ 是与覆盖算法对应的核函数。这个映射有如下特点:一个覆盖邻域中的点被映射到空间 Z 中的同一方向,而不同的覆盖邻域则被映射成相互正交的方向。这个性

质使求解问题大为简化,而核函数法得到的支持向量集,与每个支持向量相对应的方向则不一定正交。

(2) 核函数方法也可看成是一种特殊的覆盖算法,只要对每个支持向量 x_i 取覆盖的功能函数为 $\sigma' = (K(x, x_i))$, 其中:

$$\sigma'(x) = \begin{cases} 1, & \text{当 } x > 0 \\ 0, & \text{其它} \end{cases}$$

(3) 利用核函数方法求得的是最大间隔解,而覆盖算法在求解过程中只有局部求优的过程。

1.1.2 构造性核覆盖算法

基于上述核函数与覆盖算法的等价性特点,在覆盖算法中引入核函数^[5]。首先,任取一核函数 $K(x, y) = \langle T(x), T(y) \rangle$ 做以下变换 $T: D \rightarrow Z, x \in D$; 其中 D 为输入的定义域为 n 维空间的有界集合,共有 p 个样本。这种变换就是将 D 上的点映射到 P 维核空间上,记核空间的输入集为: $P_t, t = 1, 2, \dots, p$ 。在核空间中,不妨设输出集 Y 的前 k 个值均不相同。令所有输出为 $y_j (j \leq k)$ 的样本标号的集合为 I_j (即 $I_j = \{I | y_i = y_j\}$), 其对应的输入集合记为 $P_j, j = 0, 1, 2, \dots, k-1$ 。经过上面的一系列初始化后,即可开始求取一批核空间中的覆盖 $\{C_j^{(i)}, j = 1, 2, \dots, s-1; i = 1, 2, \dots, p\}$ 。令 $C_j = \bigcup C_j^{(i)}, i = 1, 2, \dots, p$, 则每个 C_j 表示一个类别的所有覆盖。因此构造性核覆盖算法的步骤为:

(1) 在样本集中任取一个尚未被覆盖的点 x_j , 使得 $x_j \in P_t$, 按式:

$$d_j^{(1)} = \min_{m \in I_j} \{K(x_j, x_m)\} \quad (4)$$

$$d_j^{(2)} = \max_{m \in I_j} \{K(x_j, x_m) | K(x_j, x_m) > d_j^{(1)}\} \quad (5)$$

$$d_j = [d_j^{(1)} + d_j^{(2)}] / 2 \quad (6)$$

$$\theta_j = [d_j^{(1)} - d_j^{(2)}] / 2 \quad (7)$$

计算,根据 x_j 和 d_j 构造一个覆盖 $C_j^{(i)}$, 该覆盖的中心为 x_j , 覆盖半径 $R = d_j$, 分类间隔为 θ_j 。

(2) $C_j^{(i)}$ 求出后,将 P_t 中所有的已被 $C_j^{(i)}$ 覆盖的点从 P_t 中删除,再在 P_t 中选择一个 $x_j (j \in I_j)$, 重复第一步操作,直到所有的 $x_j \in I_j$ 均已被删除为止。这样,便构造出一个类的所有覆盖领域。

(3) 对所求出的中心、半径分别为 x_1, x_2, \dots, x_m 和 d_1, d_2, \dots, d_m 的覆盖领域。令

$$K(x, y) = \exp(-\beta \|x - y\|^2 / d_i^2) \quad (8)$$

其中, d_i 表示以 x 为中心的领域的半径,求二次规划问题:

$$\max w(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j (K(d_i, d_j) + K(d_j, d_i)) / 2 \quad (9)$$

$$\text{s. t. } \sum_{i=1}^m a_i y_i = 0, a_i \geq 0, i = 1, 2, \dots, m \quad (10)$$

得到最优解:

$$a^* = \{a_1, a_2, \dots, a_m\}$$

(4) 用 a^* 构造超平面:

$$f(x) = \sum_{i=1}^m a_i y_i K(d_i, x) \quad (11)$$

其判别函数为:

$$F(x) = \text{Sign}(f(x) + b_0) \quad (12)$$

其中 b_0 为决策阈值。

(5) 对样本进行分类。对每一个样本,计算 $f(x)$ 的值,若 $f(x) > 0$, 则 x 属于正类,若 $f(x) < 0$, 则 x 属于负类,若 $f(x) = 0$, 称 x 被拒识。在数据有噪声或对精度要求很严格时,可以设定一个阈值 ϵ , 当 $|f(x)| < \epsilon$ 时认为 x 被拒识,这样可以减少误差所造成的损失。

与 SVM 算法相比,该算法具有如下的一些特点:

1) 对任意给定的样本集,该算法能构造出一次就可准确划分样本集的核函数。

2) 在对 $C_j = \bigcup C_j^{(i)}$ 求优时,求和项只对覆盖集取和,而不是对所有样本点取和。SVM 中最后其求和式只对支持向量集取和,但那是在求优之后才得出的,其在求解时是对所有样本点求和。一般覆盖个数要比样本点的个数少得多,这就是 SVM 计算量过大的原因。故该算法计算量比 SVM 少。

3) 通过覆盖后,在已求得一个允许解的情况下,可再在这个基础上求解最优,这比从任给的初始点开始求最优解要快得多。

1.2 构造性核覆盖聚类算法

用构造性核覆盖算法求聚类,等价于求出一组领域,这组领域能将相似度小的样本分割开来,将相似度大的样本聚合在一起,达到聚类的效果。即给定有 N 个未知标号的样本 (x_1, x_2, \dots, x_N) , 问题是如何根据样本的相似度,将样本分为 k 类: (C_1, C_2, \dots, C_k) 。每个类就是一个领域覆盖。对于领域覆盖比较少的样本点采用分层聚类算法中的最短距离法(采用欧式距离)对它们进行聚类,形成椭圆形覆盖领域,即选择圆心距离最近的一对覆盖合并成一个新的覆盖,根据实际需求和专业知识的球形覆盖,最后得到合理的覆盖划分,所有相似的点分布在一个领域(球形或椭球形)中。

构造性核覆盖聚类可以采用硬划分方法,也可采用模糊聚类的方法,这里仅给出硬划分的聚类方法。假设输入空间样本已被映射到特征空间 $T(x_1), T(x_2), \dots, T(x_n)$, 特征空间的欧式距离可表示为:

$$\begin{aligned} d_i(x, y) &= \sqrt{\|T(x) - T(y)\|^2} \\ &= \sqrt{T(x) \cdot T(x) - 2T(x) \cdot T(y) + T(y) \cdot T(y)} \\ &= \sqrt{T(x, x) - 2T(x, y) + T(y, y)} \end{aligned} \quad (13)$$

因此,将式(13)作为聚类的相似度量函数,而聚类的准则是使下面的目标函数最小。

$$J = \sum_{i=1}^C \sum_{j \in I_i} \{ T(x_j, x_j) - \frac{2}{N_i} \sum_{k=1}^{N_i} T(x_j, x_k) + \frac{1}{N_i^2} \sum_{k,p=1}^{N_i} T(x_k, x_p) \}$$
 (14)

其中 C 是聚类类别的个数, N_i 是第 C_i 类样本的个数, C_i 类中心的模为:

$$\| W_i \|^2 = \frac{1}{N_i^2} \sum_{k,p=1}^{N_i} T(x_k, x_p)$$
 (15)

根据上面的聚类准则(式(14))和距离相似度量(式(13)),可以得到一种基于构造性核覆盖的聚类算法。其思想是:任取一核函数 $K(x, y) = \langle T(x), T(y) \rangle$ 做以下变换 $T: D \rightarrow Z, x \in D$; 其中 D 为输入的定义域为 n 维空间的有界集合, 共有 p 个样本。这种变换就是将 D 上的点映射到 P 维核空间上, 记核空间的输入集为: $P_i, i = 1, 2, \dots, p$ 。目的是突出不同类别样本之间的特征差异, 使得样本在核空间中变的线性可分, 然后在这个高维的核空间中进行构造性覆盖得到聚类结果。其主要步骤为:

- (1) 用式(15)求所有未被覆盖的样本的中心 W_i 。
- (2) 以离中心 W_i 最近的样本 x_i 作为覆盖的圆心, 并按式(4)、(5)、(6)、(7) 计算, 根据 x_j 和 d_j 构造一个覆盖 $C_j^{(i)}$, 该覆盖的中心为 x_j , 覆盖半径 $R = d_j$, 分类间隔为 θ_j 。
- (3) $C_j^{(i)}$ 求出后, 将 P_i 中所有的已被 $C_j^{(i)}$ 覆盖的点从 P_i 中删除, 再重复第一步操作, 直到所有的 $x_j \in I_j$ 均已被删除为止。这样, 便构造出一个类的所有覆盖领域。其覆盖领域数便为聚类数 C 。
- (4) 对所求出的各覆盖领域(即聚类类别) C_i , 用式(15)求其类中心的模。
- (5) 根据实际允许误差, 将离的最近的两个覆盖采用最短距离法合并为一个新的覆盖。确定最后的聚类数。

2 仿真实验

为了验证构造性核覆盖聚类算法的有效性和可行性, 用实际数据进行了两组测试实验, 并与传统的 C-均值方法和核聚类方法进行了比较。

实验 1 用著名的 IRIS 实际数据^[8]作为测试样本集。IRIS 数据由四维空间中的 180 个样本组成, 分别隶属于 3 个不同类别, 每类 60 个样本。数据集中一类 IRIS 数据与其他两类间较好分离, IRIS 数据经常被作为标准的测试数据。该实验核函数选用径向基核函数 $T(X, Y) = \exp(-(X - Y)^2 / (2\sigma^2))$, 参数为 $m = 2$,

$\sigma^2 = 0.4$ 。共进行 10 次实验。文中算法平均仅聚错 6 个样本, 而 C-均值算法平均错聚个数达 14 个, 核聚类算法平均错聚个数达 8 个。构造性核覆盖聚类方法的收敛速度快于其它两种算法。实验结果见表 1。

表 1 三种聚类算法在 IRIS 数据样本集合上的结果比较

算法	平均迭代次数	平均错聚个数		
		类 1	类 2	类 3
C-均值聚类	12.6	0	2	12
核聚类	9.6	0	6	2
构造性核覆盖聚类	8.2	0	5	1

实验 2 用我校现有的评教数据作为测试样本集。该评教数据由三维空间中的教师评教数据平均值样本组成, 共 120 个样本分别隶属于 3 个(好中差)不同类别, 首先从以往的评教结果中抽出 3 个典型样本进行学习, 以 3 个典型样本作为 3 个不同类别的覆盖圆心实施核覆盖, 第 1 类 24 个样本(序号为 1~24), 第 2 类 72 个样本(序号为 25~96), 第 3 类 24 个样本(序号为 97~120)。该实验核函数选用 RBF 核函数, 将该组数据进行预处理后, 选参数为 $m = 2, \sigma^2 = 0.8$, 共进行 10 次实验。文中算法平均仅错聚 2 个样本(23 号和 24 号样本错聚为第 2 类), 而 C-均值算法平均错聚个数达 26 个, 核聚类算法平均错聚 4 个样本。结果见表 2。

表 2 三种聚类算法在评教数据样本集合上的结果比较

算法	平均迭代次数	平均错聚个数		
		类 1	类 2	类 3
C-均值聚类	15.6	24	2	0
核聚类	11.7	3	1	0
构造性核覆盖聚类	9.8	2	0	0

3 结束语

从上面的两个实验来看, 构造性核覆盖聚类方法在性能上, 比传统的 C-均值聚类算法有较大的改进, 比核聚类有更快的收敛速度以及更为准确的聚类效果。表明构造性核覆盖聚类方法是可行和有效的。

另外, C-均值聚类算法和核聚类算法一般需要事先给定聚类类别数 k , 但在多数情况下聚类类别数 k 事先无法确定, 这在一定程度上影响和限制了其应用合理性。在实际应用中, k 值是难以准确界定的, 用户无法知道采用什么样的 k 值聚类对自己更有利。但利用构造性核覆盖算法的特点, 可以有效地解决其 k 值

- [3] 赵 力, 钱向民, 邹彩荣, 等. 语音信号中的情感识别研究[J]. 软件学报, 2001, 12(7): 1050 - 1055.
- [4] Ververidis D, Kotropoulos C, Pitas I. Automatic Emotional Speech Classification[J]. IEEE ICASSP, 2004, 1(5): 1593 - 1596.
- [5] Niimi Y. Emotional Robot World[M]. Tokyo: Talk and Speak Press, 1995: 67 - 96.
- [6] Cowie R. Emotion recognition in human - computer interaction[J]. Signal Processing Magazine, 2001, 18(1): 32 - 80.
- [7] Xiao L, Yanqiu C, Soonleng L, et al. Recognition of Emotional State from Spoken Sentences[C]// presented at IEEE 3rd Workshop on Multimedia Signal Processing. [s. l.]: IEEE, 1999: 469 - 473.
- [8] Lee Chul Min, Narayanan S S. Toward Detecting Emotions in Spoken Dialogs[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(2): 293 - 303.
- [9] Ververidis D, Kotropoulos C. Emotional speech recognition: Resources, features, and methods[J]. Speech Communication, 2006(48): 1162 - 1181.
- [10] Lee H M, Chen C M, Chen J M. An efficient fuzzy classifier with features selection based on fuzzy entropy[J]. IEEE, 2001, 31(3): 426 - 432.
- [11] Mac Parthalan N, Fensen R, Qiang S. Fuzzy Entropy Assisted Fuzzy - Rough Feature Selection[C]// Proceedings of the 15th International Conference on Fuzzy Systems (FUZZ - IEEE'06). [s. l.]: IEEE, 2006: 423 - 430.
- [12] 周 洁. 语音信号中情感信息的分析与处理研究[D]. 南京: 东南大学, 2005.
- [13] Ujiie H, Omacchi S, Aso H. A Discriminant Function Based on Feature Transformation Considering Normality Improvement of Distribution[J]. IEICE Transactions on Information and Systems. Pt. 2 (Japanese Edition), 2003, 86(2): 441 - 449.
- [14] Dellart F, Polzin T, Waibel A. Recognizing Emotion in Speech[C]// International Conference on Spoken Language Processing (ICSLP'96). Philadelphia, PA, USA: IEEE, 1996: 1970 - 1973.
- [15] Takahashi K. Remarks on SVM - Based Emotion Recognition from Multi - Modal Bio - Potential Signal[C]// Proceeding of the 2004 IEEE International Workshop on Robot and Human Interactive Communication. [s. l.]: IEEE, 2004: 186 - 191.
- [16] 张石清. 基于模糊支持向量机的语音情感识别[J]. 台州学院学报, 2006, 12: 52 - 55.
- [17] 韩笑蕾, 王成儒, 贾晓光. 汉语语音情感识别的研究[C]// 中国通信理论与信号处理学术年会. 北京: 电子工业出版社, 2007: 570 - 575.
- [18] 余伶俐, 蔡自兴, 陈明义. 语音信号的情感特征分析与识别研究综述[J]. 电路与系统学报, 2007, 12(4): 76 - 83.
- [19] Lin Yi, Wei Gang. Speech emotion recognition based on HMM and SVM[C]// Proceedings of 2005 International Conference on Machine Learning and Cybernetics. Guangzhou: [s. n.], 2005: 4889 - 4901.
- [20] 国辛纯, 郭继昌, 窦修全. 基于 HMM 的语音信号的情感识别研究[J]. 电子测量技术, 2006, 10: 69 - 71.
- [21] Nwe T L, Foo S W, Silva L. Speech emotion recognition using hidden Markov models[J]. Speech Communication, 2003 (41): 603 - 623.
- [22] 赵 力. 语音信号处理[M]. 北京: 机械工业出版社, 2003.
- [23] Nicholson J, Takahaahi K, Nakatsu R. Emotion Recognition in Speech Using Neural Networks[J]. Neural Computing & Application, 2000, 12: 290 - 296.
- [24] 王 青. 基于神经网络的汉语语音情感识别的研究[D]. 杭州: 浙江大学, 2004.
- [25] Lee C M, Narayanan K. Emotion Recognition Using Data - Driven Fuzzy Inference System[M]// Eurospeech. [s. l.]: [s. n.], 2003: 157 - 160.
- [26] Wang Zhiping, Zhao Li, Zhou Caring. Support Vector Machines for emotion recognition in Chinese Speech[J]. Journal of Southeast University (English Edit), 2003, 19(4): 307 - 310.

(上接第 91 页)

的动态获取问题。但如何对所获取的 k 值进行优化, 是今后需进一步研究的问题。

参考文献:

- [1] 黄凤岗, 宋克欧. 模式识别[M]. 哈尔滨: 哈尔滨工业大学出版社, 1998.
- [2] 沈 清, 汤 霖. 模式识别导论[M]. 长沙: 国防科技大学出版社, 1991.
- [3] Girolami M. Mereer Kernel Based Clustering in Feature Space[J]. IEEE Trans on Neural Networks, 2002, 13(3): 780 - 784.
- [4] 张 莉, 周传达, 焦李成. 核聚类算法[J]. 计算机学报, 2002, 25(6): 587 - 590.
- [5] 吴 涛, 张 铃, 张燕平. 机器学习中的核覆盖算法[J]. 计算机学报, 2005, 28(8): 1295 - 1301.
- [6] 张 铃, 张 钺. M - P 神经元模型的几何意义及其应用[J]. 软件学报, 1998, 9(5): 334 - 338.
- [7] 瓦普尼克. 统计学习理论的本质[M]. 张学工译. 北京: 清华大学出版社, 1995.
- [8] Bezdek J C, Keller J M, Krishnapuram R, et al. Will the Real IRIS Data Please Stand Up[J]. IEEE Trans on Fuzzy System, 1999, 7(3): 368 - 369.