

DOI:10.3969/j.issn.1673-4785.201403063
网络出版地址: <http://www.cnki.net/kcms/doi/10.3969/j.issn.16734785.201403063.html>

邻域同调学习算法

赵梦梦, 李凡长

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘 要: 目前已有的边缘学习算法对边缘可变的数据划分问题存在一些不足, 这些算法在分类过程中不能有效地保证数据的结构特征不变。因而文章首先通过引进同调代数中的单形划分理论, 从机器学习的角度对分类问题中的边缘划分进行研究, 提出了一种邻域同调学习算法。算法给出了图形的邻域复形的构造方法和判断 2 个给定图形相似性的判定标准。最后通过在 USPS_ALL 手写数字集数据库和 MPEG7 CE 图像库上与 SVM、TVQ 算法的对比实验验证了本算法的有效性。

关键词: 同调学习; 同调代数; 机器学习; 边缘划分; 边缘同调学习; 邻域同调学习算法; 邻域复形图; 相似性

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2014)03-0336-07

中文引用格式: 赵梦梦, 李凡长. 邻域同调学习算法[J]. 智能系统学报, 2014, 9(3): 336-343.
英文引用格式: ZHAO Mengmeng, LI Fanzhang. Neighborhood homology learning algorithm[J]. CAAI Transactions on Intelligent Systems, 2014, 9(3): 336-343.

Neighborhood homology learning algorithm

ZHAO Mengmeng, LI Fanzhang

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: At present, the existing margin learning algorithms still have some affects when attempting to solve the data partitioning problem of variable margins. These algorithms can not effectively maintain the structure feature of datas in classification.. At present, the existing margin learning algorithms still have defects when attempting to solve the data partitioning problem of variable margins. As a consequence, this paper initially proposes a neighborhood homology learning algorithm through using the monomorphic division theory in homology algebra. The neighborhood homology learning algorithm reasearchs the margin partitioning problem from the perspective of machine learning. The neighborhood homology learning algorithm includes the method of structuring the neighborhood complex, and the criterion for judging the similarity between two given graphs. Finally, this algorithm is justified through the experimental results contrasted with SVM and TVQ on an image dataset named MPEG7 CE and a database of handwritten digits named USPS_ALL.

Keywords: homology learning; homology algebra; machine learning; margin partitioning; margin homology learning; neighborhood homology learning algorithm; neighborhood complex graphs; similarity

边缘学习^[1-5]是机器学习的核心问题之一。目前比较成功的边缘学习算法主要有 V.Vapnik 提出的 SVM 最大边缘算法^[6-8], Fabio Ailli、Alessandro

Sperduti 提出的 TVQ 算法^[9]和同调边缘胞腔学习算法等^[10-11]。

支持向量机(support vector machine, SVM)是一种努力最小化结构风险^[12-13]的算法, 主要有线性可分支持向量机^[12]、近似线性可分支持向量机^[12]和线性不可分支持向量机。线性可分支持向量机又叫

硬间隔支持向量机^[12],它是在正确划分训练集的超平面中,采用极大间隔分类超平面的方法^[12,14-15],根据最大间隔原则,找出最优的决策超平面。近似线性可分支持向量机即软间隔支持向量机(soft margin SVM)^[15],是在线性可分的情况下建立起来的,即在线性可分支持向量机的最优化问题上添加松弛因子 ξ_i 和惩罚因子 C ,允许有错分样本存在。线性不可分^[12,15]是通过引入一个核函数^[12]来实现的,该方法是将原始数据通过非线性映射投影到高维空间,使得数据在高维空间变得线性可分。

在 TVQ 算法中,提出了一个对提升边缘问题的很好的方案,该方案是建立在边缘理论原则上,是基于样本加权策略。其算法实例包含了基于 1-NN 分类器的切线距离^[16],其中 1-NN 分类器加入了一种量子化的切距离原型。该算法中,引入的切距离原型在关于标准切线模式的泛化误差中得到了明显的改进。该方法与 SVM 方法进行比较,可以观察到,当此方法和 SVM 方法具有从经验风险到理想风险中一致收敛的统计学习理论概念时,该方法将输入分布直接工作在线性模型上而不是采取预定义核的方式。该方法和 Boosting 算法^[17-18]很相似,但是在 Boosting 算法中,采取的是多种组合假设;而在 TVQ 算法中,更加关注的是独立假设。

SVM 和 TVQ 在一定程度上都提高了边缘划分效果和准确度。但是这些方法对边缘可变的划分问题就显得不足了,因此本文提出了一种邻域同调学习算法。邻域同调学习算法是通过引进同调论思想,从机器学习角度出发研究的一种边缘学习新方法。该方法针对图集的分类问题,在分类过程中能够有效地保证图形的结构特征不变。该方法主要是通过判断给定图形的邻域复形的各阶同调群是否同构,对图形进行邻域同调分类。

1 邻域同调的基本概念

同调^[19-20]是源于代数拓扑学^[21]的一个分支。20 世纪 40 年代,代数拓扑学的一些概念和方法被引进代数领域,形成了系统的理论,它的兴起对群、李代数^[22]与结合环的研究都起了非常重要的作用。同调的研究对象基本上是模范畴^[23-24]所派生一些 Abel 范畴,其核心是同调以及同调函子。代数拓扑学中的复形、同调群等概念也由此引入。其中起重要作用的一类模^[23-24]是投射模、内射模和平坦模,利用它们定义了扩张函子和 n 级张量函子以及同调维数^[24]等概念。在环论研究中,利用同调性质来刻画环自身结构取得了一系列重要成果。计

算群、环、代数的同调群是同调的重要研究课题。

为了方便讨论,先介绍一些基本概念。

定义 1 设点集 $\{a_0, a_1, \dots, a_q\} \subset R^n$, R^n 表示线性的 n 维欧氏空间,若矢量组 $a_1 - a_0, a_2 - a_0, \dots, a_q - a_0$ 线性无关,则称点集 $\{a_0, a_1, \dots, a_q\} \subset R^n$ 在 R^n 中是几何独立的。

设点集 $\{a_0, a_1, \dots, a_q\}$ 在 R^n 中几何独立,令

$$A^q = \left\{ x \in R^n; x = \sum_{i=0}^q \lambda_i a_i, \lambda_i \geq 0, \sum_{i=0}^q \lambda_i = 1 \right\}$$

称 A^q 为 q 维几何单形,简称 q 维单形,记作 $\langle a_0, a_1, \dots, a_q \rangle$, a_0, a_1, \dots, a_q 称为 q 维单形 A^q 的顶点,实数 λ_i 称为点 x 在 q 维单形 A^q 中的重心坐标。

定义 2 设 $K = A_i^q, q = 0, 1, \dots, n, i = 1, 2, \dots, \alpha_q$ 是 R^n 中有限个单形的集合,若满足下列 2 个条件,则称 K 为单纯复形。

- 1) K 中任意 2 个单形都是规则相处的;
- 2) K 中任意单形的任意一个面仍是 K 中的单形。

单纯复形 K 中的 0 维单形称为 K 中的顶点。 K 的维数是 K 中单形维数的最大值,即 $\dim K := \max\{q; A^q \in K\}$ 。对于一个 q 维单形的全体顶点 a_0, a_1, \dots, a_q ,它们的全部排列可以分成 2 组:同一组(如:同是顺时针方向)中的排列彼此差一个偶置换,不同组之间的排列,彼此差一个奇置换。这两组叫做 A^q 的 2 个定向。选定定向的单形叫做有向单形,那么可以将该有向单形中的一组记作 $+\sigma^q$ (简记作 σ^q),另一组相反指向的记作 $-\sigma^q$ 。

现有一 n 维单纯复形 K , $\{A_i^q; q = 0, 1, \dots, n, i = 1, 2, \dots, \alpha_q\}$ 是它的全部单形。对每个 A_i^q 取定一个指向,这个有向单形记作 σ_i^q (另一个指向决定的有向单形就是 $-\sigma_i^q$)。则称 $\{\sigma_i^q; q = 0, 1, \dots, n, i = 1, 2, \dots, \alpha_q\}$ 为 K 的有向单形的一个基本组。

$$x^q = \sum_{i=1}^{\alpha_q} \lambda_i \sigma_i^q, \lambda_i \in Z, q = 0, 1, \dots, n \quad (1)$$

叫做 K 的一个 q 维链。链(1)与同维链 $y^q = \sum_{i=1}^{\alpha_q} \mu_i \sigma_i^q$ 的和定义为链:

$$x^q + y^q = \sum_{i=1}^{\alpha_q} (\lambda_i + \mu_i) \sigma_i^q \quad (2)$$

显然,在式(2)的加法运算下, q 维链的全体构成一个自由交换群,记作 $C_q(K)$, 即

$$C_q(K) = \left\{ \sum_{i=1}^{\alpha_q} \lambda_i \sigma_i^q; \lambda_i \in Z, q = 0, 1, \dots, n \right\} \quad (3)$$

$C_q(K)$ 叫作 K 的 q 维(整数)链群。且对于任意 K 的 q 维链 $c^q = \sum_{i=1}^{\alpha_q} \lambda_i \sigma_i^q \in C_q(K)$, 令

$$\partial_q c^q = \partial_q \left(\sum_{i=1}^{\alpha_q} \lambda_i \sigma_i^q \right) = \sum_{i=1}^{\alpha_q} \lambda_i \partial \sigma_i^q \in C_{q-1} K \quad (4)$$

则 $\partial_q: C_q(K) \rightarrow C_{q-1}(K)$ 是群同态。这里的 ∂_q 即为 q 维边缘算子或边缘同态。其实,边缘同态是几何上概念的代数化,是用来反映图形的边界关系的。

在此基础上,定义出一个单纯链复形 $C_*(K) = \{C_q(K), \partial_q\}$ 。 $C_*(K)$ 即是一串 q 维(整数)链群 $C_q(K)$ 和一串边缘同态 $\partial_q: C_q(K) \rightarrow C_{q-1}(K)$,排成一个序列:

$$\begin{aligned} \cdots \rightarrow 0 \rightarrow C_q(K) \xrightarrow{\partial_q} C_{q-1}(K) \xrightarrow{\partial_{q-1}} \cdots \\ \rightarrow C_1(K) \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0 \end{aligned} \quad (5)$$

满足条件对每个维数 q 都有 $\partial_q \partial_{q+1} = 0$,也就是说“2次边缘为零”。

定义3 单纯复形 K 的 q 维闭链群和边缘链群依次为

$$\begin{aligned} Z_q(K) &= \text{Ker} \partial_q = \{z^q \in C_q(K); \partial_q z^q = 0\} \quad (6) \\ B_q(K) &= \text{Im} \partial_{q+1} = \\ &\{b^q \in C_q(K); \exists c^{q+1} \in C_{q+1}(K), \text{使 } \partial c^{q+1} = b^q\} \end{aligned} \quad (7)$$

$Z_q(K)$ 的元 z^q 叫做 q 维闭链, $B_q(K)$ 的元 b^q 叫做 q 维边缘链。

这里 $B_q(K) \subset Z_q(K)$, 而且作为自由群 $C_q(K)$ 的子群,它们也都是自由的。 $Z_q(K)$ 是描述有可能成为边的图形, $B_q(K)$ 是描述真的是边的图形。那么“有可能成为边,又不真是边”这一事实,用群论的语言来表达,就应该是:在有可能成为边的 $Z_q(K)$ 中,模真是边的部分 $B_q(K)$ 。这样就有商群:

$$H_q(K) = Z_q(K) / B_q(K) = \text{Ker} \partial_q / \text{Im} \partial_{q+1}$$

$H_q(K)$ 叫做单纯复形 K 的 q 维(单纯)同调群。

同调群 $H_q(K)$ 中的元素为 $Z_q(K)$ 中元素的模 $B_q(K)$ 的等价类,因此,由单纯复形 K 上 q 维闭链 z^q 所决定的模 $B_q(K)$ 的等价类是 $z^q + B_q(K)$, 记作 $[z^q]$, 称其为 z^q 的 q 维同调类,并且 $H_q(K) = \{[z^q]; z^q \in Z_q(K)\}$ 。同一同调类中的元 z_1^q 和 z_2^q 称为同调的,记作 $z_1^q \sim z_2^q$ 或 $z_1^q - z_2^q \sim 0$ 。

由定义2可以看出单纯复形是由简单的图形——单形构成的。那么有了一个单纯复形之后,由此出发也可构成相应的新复形。其实仔细观察单纯复形的定义就会发现,单纯复形实际上完全由其顶点和顶点构成的子集组(单形)所决定,当然这时顶点的子集组要适合某种条件。据此,定义4给出了抽象复形的定义,该定义则为本文算法中提到的构造邻域复形提供了理论基础。

定义4 以有限个元素 c_0, c_1, \dots, c_q 的子集做

为元素所得的集 \mathbf{K} ,叫做以 c_0, c_1, \dots, c_q 为顶点的抽象复形,如果满足以下2个条件:

1)由单个元素 c_i 所组成的子集属于 \mathbf{K} ,这里 $i = 0, 1, \dots, q$;

2)若 \mathbf{A} 是 \mathbf{K} 的元,则 \mathbf{A} 的子集也是 \mathbf{K} 的元。

抽象复形 \mathbf{K} 中的元 \mathbf{A} 叫做抽象单形,它的维数是其顶点数减1。 \mathbf{A} 的子集叫做 \mathbf{A} 的面。 \mathbf{K} 的维数是 \mathbf{K} 中诸单形的最大维数。显然,每个单纯复形 K 都自然地决定一个抽象复形 \mathbf{K} 。实际上, K 的每个单形 $(a_0 a_1 \cdots a_q)$ 规定 \mathbf{K} 的一个抽象单形 $(c_0 c_1 \cdots c_q)$ 。这样规定的 \mathbf{K} ,显然是一个抽象复形,且称 K 为 \mathbf{K} 的一个几何实现。

2 邻域同调学习算法

设 $G = (V(G), E(G))$ 是一个简单连通图。显然,一个简单连通图就是一个单纯复形,根据上述定义即可以看成是有限个单形的集合, $V(G) = \{a_0, a_1, \dots, a_{\alpha_0-1}\}$ 表示图 G 的 α_0 个顶点的集合也就是图 G 中所有0维单形的集合, $E(G)$ 表示顶点之间的所有连接边的集合。令 X 和 F 分别是 G 的一个顶点子集和边子集,就以 $G - X$ 表示 G 去掉 X 中的点及其关联边而得的图, $G - F$ 表示 G 去掉 F 中的边而得到的图。设 a 是 G 的任意一个顶点,以 $N(a)$ 表示 a 的邻域,那么 $N(X)$ 表示 X 中所有点邻域的并,即 $N(X) = \bigcup_{a \in X} N(a)$ 。

基于此,可以构造出一幅给定图形 G 的邻域复形 $N(G)$, $N(G)$ 是以 G 的顶点为顶点,以 G 中具有公共邻接点的子集为单形的一个抽象复形。根据抽象复形的概念,可以看出任意的抽象复形都可以在欧氏空间实现^[25]。因此邻域复形在欧氏空间中是有其相对应的单纯复形的,这里仍以 $N(G)$ 记邻域复形所对应的单纯复形。除非特殊说明,二者不加区别,统称为 G 的邻域复形。下文中的 $N(G)$ 即指邻域复形在欧氏空间中所对应的单纯复形。

由于构造出的任意图形的邻域复形 $N(G)$ 可以看成是单纯复形,那么可以将 $N(G)$ 剖分成有限个单形,根据定义2可表示为 $N(G) = \{A_i^q; q = 0, 1, \dots, \dim(N(G)), i = 1, 2, \dots, \alpha_q\}$; 对每个单形取定一个指向并记为 σ_i^q , 根据定义1,每个单形 σ_i^q 都可以根据顶点坐标线性表示出来;再由式(1)将 q 维单形线性组合就能得出 $N(G)$ 的 q 维链,这些 q 维链组成的自由交换群即为 $N(G)$ 的 q 维链群 $C_q(N(G))$ 。再通过计算同调群,就可以得出2个图形之间的一种同调关系。设有2个图 G 、 H ,如果 $N(G)$ 和 $N(H)$ 的各阶同调群分别同构,就称 G 和 H 是邻域同调的,记为 $G \stackrel{N}{\simeq} H$ 。

当然,不同类型的连通图,它们是邻域同调的充分必要条件也是不一样的。例如,2 个仙人掌图是邻域同调的,当且仅当这 2 个图都是二分图或都不是二分图且它们的非 4-圈^[26]的圈数相同^[27];同样对于两幅立方图(每个顶点的次数都等于 3 的有限简单连通图)来说,它们是邻域同调的充要条件是两幅图的分二性相同且 D 值相同^[28];而两个不含 4-圈的图是邻域同调的充要条件则是它们的二分性相同且圈秩相同^[29]。

基于上述理论,文中给出了一种邻域同调学习算法(neighborhood homology learning algorithm, NHLA)。该算法主要是针对简单联通图(不含孤立顶点的有限简单图)的分类,通过图形的邻域复形的同构性判断出图形是否是邻域同调,从而进行邻域同调分类。

假设有一个由 n 幅图组成的样本集 $G = \{G_1, G_2, \dots, G_n\}$ 。根据邻域复形的概念,为每一个 $G_i, i \in [1, n]$ 构造出其邻域复形 $N(G_i)$, 并计算出 $N(G_i)$ 的 q 阶同调群, 其中 $0 \leq q \leq \dim(N(G_i))$ 。然后判断 $N(G_i)$ 和 $N(G_j)$ 的各阶同调群是否同构,进而判断出 $G_i (i \in [1, n - 1])$ 与 $G_j (j \in [i + 1, n])$ 是否邻域同调的。最终根据图之间的邻域同调性,可将图分为 k 类,则输出集合可表示为 $H = \{H_1, H_2, \dots, H_k\}$ 。具体算法步骤如下:

算法 邻域同调学习算法

输入:待测样本集合 $G = \{G_1, G_2, \dots, G_n\}$ 。

输出:分类后的集合 $H = \{H_1, H_2, \dots, H_k\}$ 。

1) 置每个样本 G_i 为一个类,则 G_1, G_2, \dots, G_n , 共形成 n 个类;

2) 找出 G_i 的任意一个顶点的邻域,那么 G_i 中的所有顶点邻域的并集即为所要构造的邻域复形 $N(G_i)$;

3) 根据式(3)和式(4)计算出每个 G_i 邻域复形 $N(G_i)$ 的 q 维链群 $C_q(N(G_i))$, 此时可以得到形如式(5)的一个序列:

$$\begin{aligned} \dots \rightarrow 0 \rightarrow C_q(N(G_i)) &\xrightarrow{\partial_q} C_{q-1}(N(G_i)) \rightarrow \dots \\ \rightarrow C_1(N(G_i)) &\xrightarrow{\partial_1} C_0(N(G_i)) \xrightarrow{\partial_0} 0 \end{aligned}$$

4) 再由式(6)和式(7)分别可以求出邻域复形 $N(G_i)$ 的 q 维闭链群 $Z_q(N(G_i))$ 和 q 维边缘链群 $B_q(N(G_i))$;

5) 根据 $H_q(N(G_i)) = Z_q(N(G_i)) / B_q(N(G_i))$ 计算出 $N(G_i)$ 的 q 维同调群;

6) 如果 $G_i, i \in [1, n - 1]$ 与 $G_j, j \in [i + 1, n]$ 的各阶同调群分别同构,则 $G_i \overset{N}{\simeq} G_j$ 。那么就将 G_i 与 G_j 合并为一个新类,此时现有的类数为

$k = n - 1$ 类。循环判断至所有样本都分类完毕;

7) 通过上述步骤可将样本分成 k 类,输出 $H = \{H_i\}_{i=1}^k$ 。

由上面的算法可得定理 1。

定理 1 NHLA 算法的所需时间复杂度为 $O(n^2)$, 空间复杂度为 $O(n)$ 。

证明 假设有一任意图形 G , 顶点集为 $V(G) = \{a_0, a_1, \dots, a_{\alpha_0-1}\}$, 根据每个顶点的邻域可以构造出它在欧几里德空间里的邻域复形 $N(G) = \{A_i^q; q = 0, 1, \dots, \dim(N(G)), i = 1, 2, \dots, \alpha_q\}$ 。

根据步骤 2), 对于 n 个待测的图形样本, 需构造出 n 个邻域复形, 因此本文可以得到 2) 的时间复杂度为 $O(n)$, 空间复杂度为 $O(n)$ 。

根据步骤 3)~5), 求出每个邻域复形相对应的 q 维同调群, 此时的时间复杂度为 $O(n)$, 空间复杂度为 $O(n)$ 。

根据步骤 6) 程序主体, 通过 *for* 循环进行判断分类直到分类完毕并最终输出分类结果, 这里时间复杂度为 $O(n^2)$, 空间复杂度为 $O(n)$ 综上可得, NHLA 算法的时间复杂度为 $O(n^2)$, 空间复杂度为 $O(n)$ 。

3 实验分析

为了验证本文算法的有效性, 本文进行了 3 组实验。第 1 组实验是人工数据, 第 2 组是手写数字识别实验, 第 3 组是实物图像。在实验中, 本文方法同时也和 SVM、TVQ 方法进行了对比。本文实验均在 MATLAB 平台上实现。

3.1 人工数据

本文给出了一次随机生成的二维的连通图集数据, 该人工数据是 3 类服从高斯分布的随机数据, 每次生成数据时, 每类样本是 100 个。其中符号 ‘*’, ‘o’ 和 ‘∇’ 分别表示不同的类别, 将其可视化如图 1(a) 所示。图中的每个点表示每个二维图形所在的位置, 是用图形的重心坐标表示的。

分类之前首先对生成图形 G 进行处理, 即需要先提取图形边缘点。将提取的边缘点用 $V(G)$ 表示, 这些边缘点间的连接边用 $E(G)$ 表示。接着根据邻域复形构造原理构造出 $N(G)$, 然后再利用上述的邻域同调学习算法进行分类, 分类结果如图 1(b) 所示。图 1(c) 和图 1(d) 分别是用 SVM 和 TVQ 算法分类后的结果。图 1(a) 中随机生成的这几类数据有交叉重叠部分, 且从图 1(b)~(d) 可以看出, 经过 NHLA、SVM 和 TVQ 等 3 种算法分类后, 数据都能不同程度的分开。但比较分类后的 3 幅图可以看出, 本文提出的邻域同调学习算法分类后的效果是明显好于另 2 种算法。

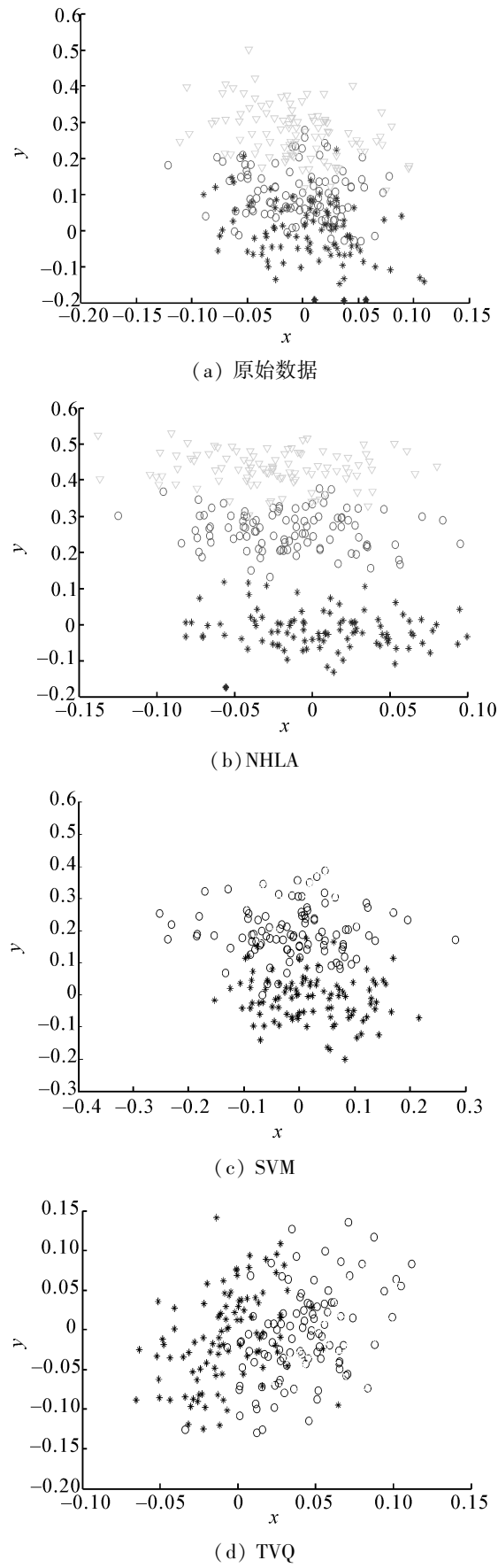


图 1 随机数据和可视化结果

Fig.1 Random data and the result of visualization

3.2 手写数字识别实验

本实验采用 USPS_ALL 数据库的手写数字集进行实验。此手写数字集包含 10 个数字(0,1,...,9)的图片,每个数字含有 1 100 张字迹不同的图片,每张图片的像素大小为 16×16。部分图像如图 2。



图 2 USPS_ALL 数据库部分图像实例
Fig.2 Part of the images in USPS_ALL

实验中,从每类数字中选取一定数量的图像构成样本集。样本集大小从 1 500 变化到 4 500。为消除选择样本的随机独立性,独立重复实验 100 次,最后取平均识别率作为最终识别率。图 3 给出了几种算法在数字手写数据集上不同样本点个数情况下的实验结果。分类前对图像的边缘处理和人工数据实验是相同的。图 3 中,随着选取样本个数的增加,各个方法的识别精度也是随之增加的,而且本文提出的邻域同调算法始终是好于 SVM 和 TVQ 的。

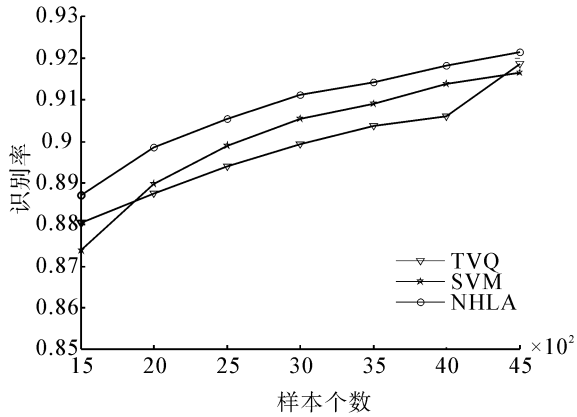


图 3 手写数字集上的识别率

Fig.3 Recognition accuracy on the USPS_ALL database

3.3 实物图像

采用 MPEG7 CE-Shape1 Part B 图像库对实物图像进行分类。该图像库是一个图像模版库,包含了各种形状特征的物体模版图,由 70 类物体、每类 20 幅共 1 400 幅实物图像组成。部分图像如图 4。

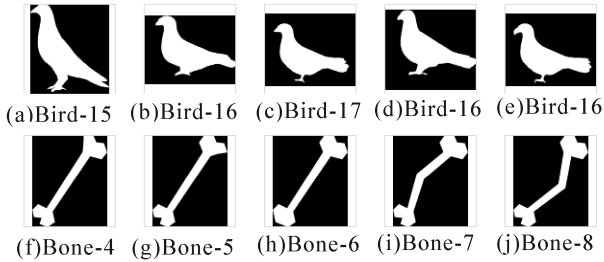


图 4 部分实物图像

Fig.4 Part of the images in MPEG7 CE

本文在 MPEG7 CE-Shape1 Part B 图像库中选

取其中 16 类,即共 320 幅图像进行实验。从每类中选取一定数量的图像构成样本集。其中样本集的大小从 80 变化到 320。为了消除选择样本的随机独立性,独立重复实验 100 次,最后取平均识别率作为最终识别率。图 5 给出了在 MPEG7 CE-Shape1 Part B 图像库上,几种算法在选取不同样本点个数情况下的实验结果。

从图 5 中可以看出,选取的样本个数越多,各个方法的识别精度整体趋势也是随之增加。此外在样本较少的情况下,由于部分图像受到噪声的干扰,本文提出的 NHLA 算法识别率受到一定的影响,其识别效果不稳定,有一点降低的趋势。但是与其他算法相比较,NHLA 依然具有较好的识别精度。而且随着样本的增多,识别率受噪声干扰的程度也减小,仍是处于增长趋势的。

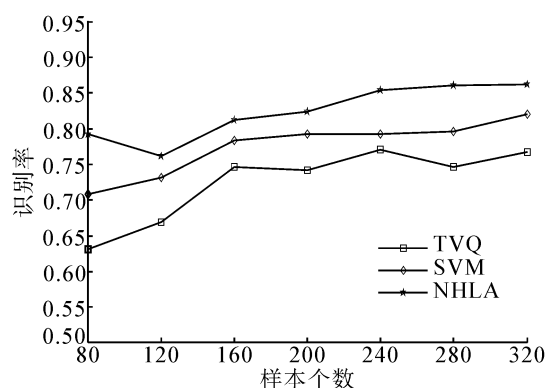


图 5 MPEG7 CE 上 3 种算法的识别率

Fig.5 Recognition accuracy on the MPEG7 CE database

4 结束语

针对边缘可变的划分问题,本文提出了邻域同调学习算法,该算法引进了同调理论,利用复形之间的性质,通过构造图形的邻域复形,并判断这些邻域复形是否同构来进行对图集的邻域同调分类。利用该算法进行分类能够有效地保证在分类过程保持图形原有的结构特征不变。

通过实验对比可以看出,本文提出的算法的识别率较高,具有一定的优势。但是,本文的算法目前只实验于二维图像上,对于更高维的图像分类还有待完善,另外本算法对噪声和扰动的鲁棒性也有待提高。总之,邻域同调学习算法具有比较坚实的理论基础,是一种可持续发展的学习方法。

参考文献:

[1] 周志华,王珏.机器学习及其应用研究[M].北京:清华大学出版社,2007:1-161.
 [2] 李凡长,张莉,杨季文.李群机器学习[M].合肥:中国科学技术大学出版社,2013:281-296.

[3] PERNKOPF F, WOHLMAYR M, TSCHIATSCHEK S. Maximum margin Bayesian network classifiers[J]. Pattern Analysis and Machine Intelligence, 2012, 34(3): 521-532.
 [4] PANWAR H, GUPTA S. Advances in Computer Science, Engineering and Applications[M]. Berlin: Springer, 2012: 385-392.
 [5] EIDELMAN V. Optimization strategies for online large-margin learning in machine translation[C]//Proceedings of the 7th Workshop on Statistical Machine Translation, Montreal, Canada, 2012: 480-489.
 [6] VAPNIK V. Statistical learning theory[M]. New York: Wiley, 1998: 1-768.
 [7] LEE Y J, HSIEH W F, HUANG C F. e-SSVR: A smooth support vector machine for e-insensitive regression [J]. IEEE Transactions on Knowledge and data Engineering, 2005, 17(5): 678-685.
 [8] LIN C F, WANG S D. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(3): 466-471.
 [9] FABIO A, ALESSANDRO S. A re-weighting strategy for improving margins [J]. Artificial Intelligence, 2002, 137: 197-216.
 [10] 鲜敏,李凡长.一种同调边缘学习算法[J].计算机工程与应用,2008,44(21): 192-194.
 XIAN Min, LI Fanzhang. Learning algorithm based on homology margin [J]. Computer Engineering and Applications, 2008, 44(21): 192-194.
 [11] 杨季文,李哲硕,何书萍,等.胞腔同调边缘学习算法研究[J].计算机研究与发展,2013,50(5): 1005-1011.
 YANG Jiwen, LI Zheshuo, HE Shuping et al. On cellular homology margin learning algorithms [J]. Computer Research and Development, 2013, 50(5): 1005-1011.
 [12] 邓乃杨,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,2005:1-408.
 [13] BURGESS C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
 [14] SHAW-TAYLOR J, CRISTININI N. Kernel methods for pattern analysis[M]. Cambridge University Press, 2004: 289-436.
 [15] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
 [16] SIMARD P Y, LECUN Y, DENKER J. Efficient pattern recognition using a new transformation distance[J]. Advanced in Neural Information Processing System, 1993, 5: 50-58.
 [17] SCHAPIRE R E, SINGER Y. Improved boosting algorithms using confidence-rated predictions [J]. Machine Learning, 1999, 37(3): 297-336.
 [18] DUFFY N, HELMBOLD D. Boosting methods for regres-

- sion[J]. Machine Learning, 2002, 47(2/3): 153-200.
- [19] 南基洙, 王颖. 同调代数导论[M]. 大连: 大连理工大学出版社, 2011: 112-222.
- [20] 姜伯驹. 同调论[M]. 北京: 北京大学出版社, 2006: 1-90.
- [21] 江泽涵. 拓扑学引论[M]. 上海: 上海科学技术出版社, 1978: 25-169.
- [22] 孟道骥, 白承铭, 李群[M]. 北京: 科学出版社, 2003: 28-92.
- [23] 黄保军. 同调与同伦原理[M]. 合肥: 中国科学技术出版社, 2005: 16-59.
- [24] 沈信耀. 同调论——代数拓扑学之一[M]. 北京: 科学出版社, 2002: 19-132.
- [25] 谢力同. 组合拓扑方法在图论和拟阵理论中的应用[J]. 数学进展, 1983, 12(3): 1-5.
XIE Litong. Applications of combinatorial topology method in graph theory and matroid theory[J]. Advances in Mathematics, 1983, 12(3): 1-5.
- [26] 方冬云. 不包含 $\{4, 8, 9\}$ -圈平面图结构的性质[J]. 长春工业大学学报, 2011, 32(5): 485-488.
FANG Dongyun. Nature of stucture planar graph without $\{4, 8, 9\}$ -cycles[J]. Journal of Changchun University of Technology, 2011, 32(5): 485-488.
- [27] 薛秀谦. 仙人掌图的邻域同调分类[J]. 山东大学学报, 1994, 29(1): 13-17.
XUE Xiuqian. The neighborhood homology classification of cactus[J]. Journal of Shandong University, 1994, 29(1): 13-17.

- [28] 薛秀谦. 立方图的邻域同调分类[J]. 中国矿业大学学报, 1995, 24(4): 110-112.
XUE Xiuqian. The neighborhood homology classification of cubic graphs[J]. Journal of China University of Mining and Technology, 1995, 24(4): 110-112.
- [29] 薛秀谦. 无 C_4 图的邻域同调分类[J]. 系统科学与数学, 1998, 18(1): 101-103.
XUE Xiuqian. The neighborhood homology classification of the C_4 free graphs[J]. Systems Science and Mathematics, 1998, 18(1): 101-103.

作者简介:



赵梦梦, 女, 1991年生, 硕士研究生, 主要研究方向为机器学习。



李凡长, 男, 1964年生, 教授, 博士生导师, 中国人工智能学会理事, 中国人工智能学会的机器学习专委会常务委员, 机器感知与虚拟现实专委会委员, 智能系统工程专委会委员, 粗糙集与软计算专委会常委, 中国计算机学会高级会员, 中国计算机学会的理论计算机科学专委会委员, 人工智能与模式识别专委会委员。主要研究方向是动态模糊逻辑和李群机器学习等, 先后承担国家自然科学基金重点、面上及省级项目8项, 获省级科技奖2项, 发表学术论文150余篇, 出版专著7部。

2014 IEEE 模糊系统国际会议

2014 IEEE International Conference on Fuzzy Systems

The IEEE World Congress on Computational Intelligence (IEEE WCCI) is the largest technical event in the field of computational intelligence. IEEE WCCI 2014 will host three conferences: The 2014 International Joint Conference on Neural Networks (IJCNN 2014), the 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2014), and the 2014 IEEE Congress on Evolutionary Computation (IEEE CEC 2014). IEEE WCCI 2014 will engage in cross-fertilization among the three big areas and provide a stimulating forum for scientists, engineers, educators, and students from all over the world to discuss and present their research findings on computational intelligence.

IEEE WCCI 2014 will be held in Beijing, the capital of the People's Republic of China. Beijing is the nation's political, economic, and cultural center as well as China's most important center for international trade and communications. With the biggest central square in the world - Tiananmen Square, the Forbidden City - the largest and best-preserved imperial palace complex, a superbly preserved section of the Great Wall, as well as the largest sacrificial complex in the world - the Temple of Heaven, Beijing attracts both domestic and foreign visitors who all come to wonder at its long history and unique cultural relics.

Contact Us

E-mail: wcci2014@gmail.com

Website: <http://www.ieee-wcci2014.org/>