一种基于最近邻决策的点集分类方法的确定与实现*)

马希荣[1,2] 王 嵘[1]

（天津师范大学计算机与信息工程学院 天津 300074）[1] （北京科技大学信息工程学院 北京 100083）[2]

摘 要 最近邻分类方法是识别中的重要方法。本文在最近邻理论的基础上，通过应用 Voronoi 图表[1]和 Delaunay 三角剖分[1]的特性，较好地将点集进行分类，运用一种新的方法进行实现。

关键词 最近邻，Voronoi 图，Delaunay 三角剖分

A New Method to Class a Point Set Based on Nearest-Neighbour Decision Boundaries

MA Xi-Rong[1,2] WANG Rong[1]

(College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300074)[1]

(Information Engineering School, UST Beijing, Beijing 100083)[2]

Abstract The measure of Nearest-Neighbour Decision Boundaries is very important in the identification. In this paper, based on the theory of Nearest-Neighbour, we could classify the point set using a new method by the peculiarity of Voronoi diagram and Delaunay triangulations.

Keywords Nearest-Neighbour, Voronoi diagram, Delaunay triangulations

1 引言

点集的分类是识别中的基础，贝叶斯分类、非线性判别函数分类等都属于这一类。其中最近邻法是较早提出的一种方法，在识别分类法中目前仍占有举足轻重的地位。

最近邻法将平面点集 S 分为蓝集 B 和红集 R。先前的方法就是简单地计算 S 的 Voronoi 图表[1]，这个图将平面划分成凸起的多边形 Voronoi 单元。在这个图中，点 $p(p\in S)$ 的 Voronoi 单元是所有距 p 点比距其他 S 中点近的点的集合。如果一个红点 r 的 Voronoi 单元完全被其他红点的 Voronoi 单元所包围，那么可以将这个红点从 S 中删除，这样做并不改变任何点在平面中的分类(图 1)。也就是说，这些点并不影响决策边界，而剩余的点影响决策边界。

本文提出的方法是在二维情况下，通过直接构造 Delaunay 三角剖分[1]，计算决策边界上的点，省去了构造 Voronoi 图[1]的过程。即先找到决策边界中的两个异色点，然后再逐点考查，每加一点就重新构造一下新的决策边界点集的 Delaunay 三角剖分[1]，通过判定定理验证此点是否属于决策边界的点集。同时，若新加点构成的 Delaunay 三角形的顶点都为同色，则这三个点所连接的 Delaunay 三角形的边不影响决策边界。

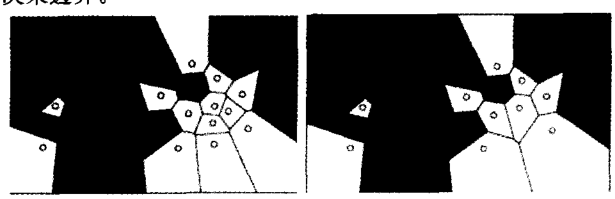


图 1 (a)点集的分布；(b) 在 S 中影响分类的点的分布

2 算法理论

在二维的最近邻决策边界问题中，集合 S 的 Voronoi 单元是凸起的多边形。运用 Delaunay 三角剖分[1]可以更方便地找出所有相邻异色点单元的 Voronoi 边界。但要假设集合 S 中的点都在一般的位置，即集合 S 中的任意四个点都不在同一个圆上。

在集合 S 中的一个 Delaunay 三角形是以 v_1, v_2, v_3 为顶点的三角形，且 v_1, v_2, v_3 所确定圆的内部不包含任何集合 S 中的点。一个集合 S 的 Delaunay 三角剖分[2]就是将集合 S 的凸壳划分成 Delaunay 三角形。同样，一个 Delaunay 边界就是一个线性的分割，这个边界的顶点(v_1, v_2)在集合 S 中。同时存在内部不包含任意集合 S 中点的圆，且 v_1, v_2 在此圆上。当集合 S 在一般的位置时，集合 S 的三角剖分是唯一的，它包含了所有的边界为 Delaunay 边界的三角形。在 Delaunay 三角剖分中，S 中的两个点由一条边界相连接，当且仅当它们 Voronoi 单元是相邻的。因此，Delaunay 三角剖分和 Voronoi 图表都是对应的。

通过以上可知，计算最近邻决策边界就等同于找出所有包含了至少一个红点和至少一个蓝点 Delaunay 边界。

因此可以通过如下定理，求得决策边界上的点集。

定理[2] 设 Q 是在决策边界上的点集，$P\subseteq Q$，且 $P\neq\emptyset$。下面的表达是等价的：

(1)在 P 集的 Delaunay 三角剖分中，每一个三角形，如果 t 的顶点中有一个蓝点(或红点)，那么 $C(t)$ 中就不能有 S 中的任何一个红点(或蓝点)(其中 $C(t)$ 为三角形 t 的外界圆)。

(2)$P=Q$

这就可以把求决策边界上的点集 Q 转化为求满足条件(1)的集合 P。

3 分类过程

在求两个集合的边界时，关键是找出 Delaunay 三角剖分的一个决策边界。选择任意一个红点 r 和任意一个蓝点 b，以 r 为原点过 b 做一条轴(一个点集中的轴当作一条射线，且最大圆的中心也在这条射线上，射线原点在这个圆的边界上，在圆内部没有集合 S 中的点(图 2))。通过 r 和 b'(可能 $b=b'$)做一个圆 C，没有任何蓝点在这个圆的内部(图 3 a)。以 b' 为

*)天津市科技攻关重点课题(编号：04310731R)、天津市自然科学基金项目(编号：60573059)。

原点且通过圆C的圆心做一条轴，通过b'和r'（可能$r=r'$）做一个圆C_1，没有任何S中的点在这个圆的内部（图3 b）。r'，b'两点连线即为Delaunay三角剖分的一条边界，也为决策边界上的一条Delaunay三角形边。这样，通过做两个轴即可确定Q集中的两个点。

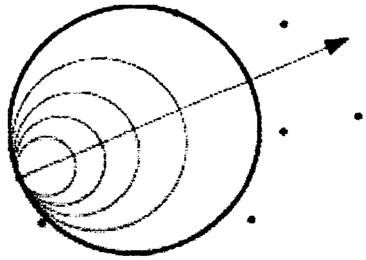
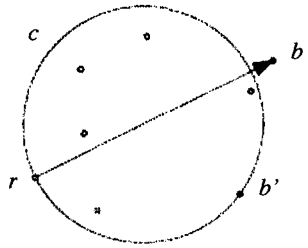


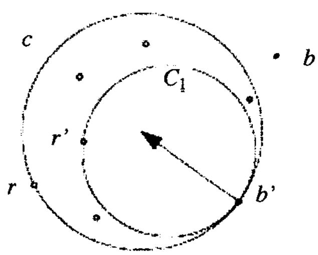
图2　点集中轴的确定

同样可以找到更多的边界，即通过找轴来确定点是否属于Q。因为找轴的过程就是定理条件(1)所表述的过程，所以可以将问题转化为通过验证点是否满足定理条件(1)来验证此点是否属于决策边界上的点集。每验证一点p，都要构造一下集合$p\cup Q$的*Delaunay*三角剖分。

通过以上过程就可以一边构造集合Q的Delaunay三角剖分，一边向集合Q中加入新找到的决策边界点。



a 确定内部无集合中蓝点的最大圆



b 确定内部无集合中任意点的最大圆

图3

4　图表表示

第3节的分类过程可用如图4表示。遍历集合中的每一点，将符合条件的点写入集合Q中。这样经过若干次循环后，Q集便是影响集合分类边界的点集。

5　实验过程及结果

对于提出的分类方法，作者运用MATLAB仿真进行实验。随机产生10个红点和10个蓝点，如图5。带圈的红点和蓝点为点r和点b，以黄点为圆心的圆是为确定第一条轴所做的圆，由此找到蓝色的星点b'，以绿点为圆心的圆是为确定第二条轴所做的圆，由此找到红色的星点r'，至此第一条边得出。同理可得第二条边，如此迭代，直至找到所有的边。这样便可把这些点中在决策边界点集中的异色点连接，这组直线扫过的区域便是边界区域，即可将两个点集中的边界找出。

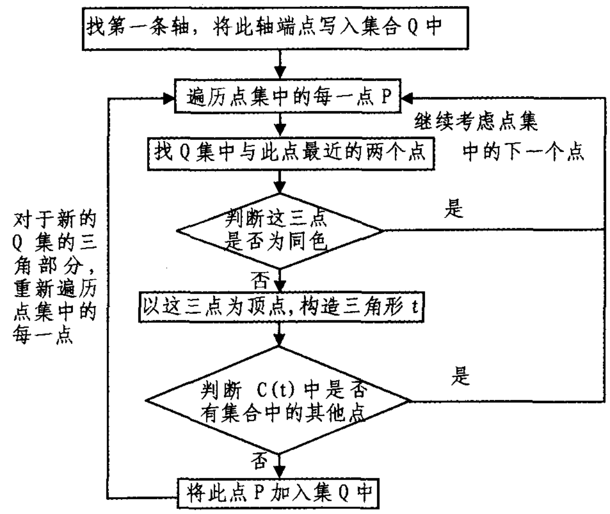


图4　确定边界点集算法流程图

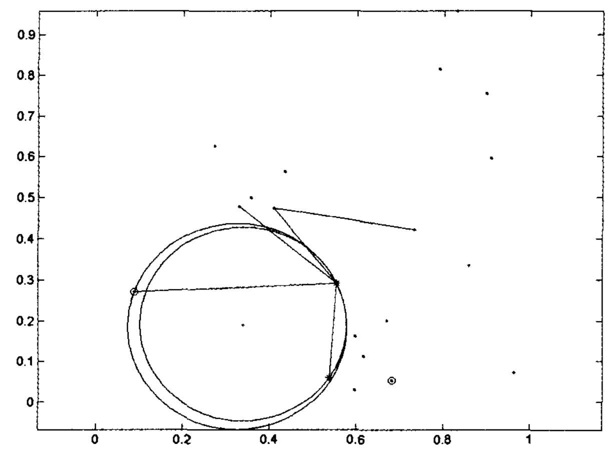


图5　随机产生20点的分类初期示意图

总结　本文在最近邻理论的基础上，借助Delaunay三角剖分和Voronoi图表的对偶性，较好地实现了对点集的划分，找到了其决策边界点集。本文提出的方法直接构造Delaunay三角剖分而不用通过Voronoi图表进行Delaunay三角剖分的构造，简单易行。应该指出的是，本方法适用于点集较小的分类。当点集规模较大时，需进一步改善、优化该算法，以提高其分类效率。

参 考 文 献

1　Bremner D, Demaine E, Erickson J, et al. Output-Sensitive Algorithms for Computing Nearest-Neighbour Decision Boundaries

2　Barber C B, Dobkin D P, Huhdanpaa H T. The Quickhull Algorithm for Convex Hulls. ACM Trans Mathematical Software, 1996, 22:469～483

3　Lee D T, Schachter B J. Two Algorithms for Constructing a Delaunay Triangulation. Int J Computer Information Sci., 1980, 9:219～242

4　周培德. 计算几何—算法分析与设计[M]. 清华大学出版社，广西科学技术出版社，88～132

5　李伟青，彭群生. 一个通用的快速三角化算法[J]. 计算机辅助设计与图形学学报，2001，13(9)：769～773

6　杨义军，孟祥旭，杨承磊，等. 复杂带状图像的快速三角剖分与骨架化算法[J]. 计算机辅助设计与图形学学报，2003，15(10)：1270～1274