第23卷第6期
2004年12月

红外与毫米波学报
J. Infrared Millim. Waves

Vol. 23, No. 6
December, 2004

文章编号:1001-9014(2004)05-0459-06

一种基于粗集理论的图像分割方法

刘宏建, 刘允才
(上海交通大学 图像处理与模式识别研究所,上海 200030)

摘要:提出了一种基于粗集理论的图像分割方法.在图像聚类过程中的对象往往是具有相似关系而不是等价关系的对象.在本文中将相似关系应用到粗集理论中来解决图像中的聚类问题.由于噪声的干扰,往往会影响到图像分割的效果.本方法提出了边界点的最大隶属原则并进而提出了边界点的粗糙度以及边界点的最大隶属原则,从而大大减小了噪声的干扰.在此基础上给出了聚类质量的评价函数.该方法为进行图像分割提供了一个崭新的视角.

关 键 词:粗集;图像分割;不可分辨测度;最大隶属原则;粗糙度
中图分类号:TN911.73 **文献标识码**:A

METHOD FOR IMAGE SEGMETNATION BASED ON ROUGH SETS

LIU Hong-Jian, LIU Yun-Cai
(Institute of Image Processing & Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200030, China)

Abstract: A method for image segmentation based on rough sets theory was presented. Objects in the clustering process often have similarity relation instead of equivalence relation. Rough sets theory was applied in similarity relation to solve clustering issue. In general, clustering results are easily corrupted by noises. In order to decrease noise disturbance, maximum membership principle of the boundary points and roughness are defined. Clustering evaluation function was presented on this base. The method provides us a new viewpoint on processing image.

Key words: rough sets; image segmentation; indiscernibility degree; maximum membership principle; roughness

引言

粗集理论是一种新型的处理数据的模糊性和不确定性的数学工具[1]能有效地分析和处理不粗确、不完整等各种不完备信息,并从中发现隐含的知识,揭示潜在的规律.不可分辨关系是粗集理论的最基本概念,在粗集理论中用上近似和下近似等概念来刻画不精确性与模糊性.

目前已有大量的聚类方法,比如K-平均算法,ISODATA算法,Hierarchical Clustering Method, Naive-Bayes, Fuzzy C-means[2,3]等.基于知识的粗集聚类方法仍然不多见,目前只在少数文献中可以看到[4].

在聚类问题中,由于数据之间大量存在的是相似,而不是等价关系.数据之间满足自反性与对称性,而不满足传递性,正是由于这种原因,所以类与类之间很容易受噪声的干扰而变得模糊不清.而粗集理论主要是解决满足等价关系的问题,对于相似关系则不太适合.为了解决这个问题,在本文提出了边界点的隶属度的概念,根据边界点对于相交各类的隶属度来最终确定边界点的归属,有效的解决了这个问题.同时提出了不可分辨测度的概念,可以动态的进行聚类.并在此基础上给出了聚类质量的评价函数.在最后一部分中将所得到的聚类算法应用于图像分割,分析了不同情况下产生的结果,证明了方法的有效性.

1 聚类过程

1.1 基于相似关系的粗集聚类

首先给出聚类过程中相似关系的定义:

收稿日期:2003-07-14,**修回日期**:2004-05-24 **Received date**: 2003-07-14, **revised date**: 2004-05-24
作者简介:刘宏建(1977-),男,山东泰安人.现于上海交通大学图像处理与模式识别研究所攻读博士学位.主要研究方向:机器视觉,模式识别与图像处理.

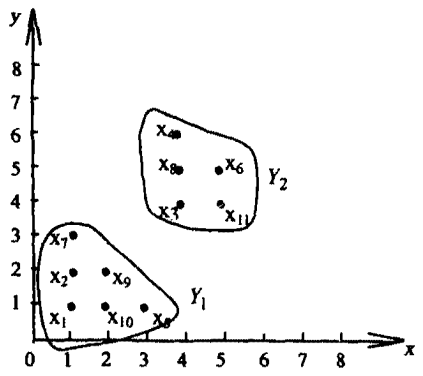

图1 例1中的分类结果
Fig. 1 Clustering results in example 1

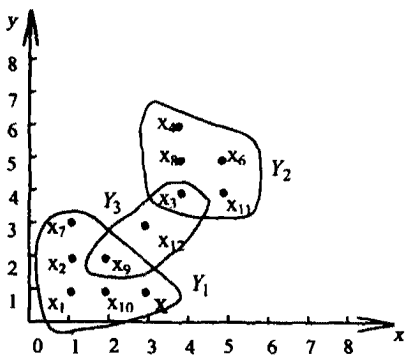

图2 例2中的分类结果
Fig. 2 Clustering results in example 2

定义1:设有论域 $U=\{x_1,x_2,\cdots,x_n\}$,则定义 U 上的关系 R 为:$R=\{X_1,X_2,\cdots X_n\}$ 其中:

$$X_i=\{x_j \mid s_c(x_i,x_j)\leqslant Th_c\},\forall x_j\in U,i\neq j. \tag{1}$$

显然关系 R 是一种相似关系. 因为 $s_c(x_i,x_i)\leqslant Th_c$,所以 R 自反的,同时由于 $s_c(x_i,x_j)\leqslant Th_c\Leftrightarrow s_c(x_j,x_i)\leqslant Th_c$,所以关系 R 是对称的,然而,$s_c(x_i,x_j)\leqslant Th_c, s_c(x_j,x_k)\leqslant Th_c$ 并不能推出 $s_c(x_i,x_k)\leqslant Th_c$ 即 R 不满足传递性. 因此关系 R 是一种相似关系而不是等价关系. 同时注意到在传统粗集的等价关系的条件下有 $X_i\cap X_j=\Phi$,但是在定义1中此条件并不满足.

对于这种相似关系,文献[5]中用扩展粗糙集来避免在数据库中的不完整数据的影响. 但是发现即使在精确完整的数据库中在很多情况下也遇到了这个问题,特别是像定义1那样的聚类过程. 这显然与传统粗集理论的等价关系是不一致的,但是在集合中这种相似关系是大量存在的.

基于上面的定义给出:

定义2:在论域 U 上,定义基于相似关系的粗集分类为:$R^-(X)=\{Y_1,Y_2,\cdots Y_f\}$ 其中:

$$Y_k=\bigcup_t\{X_i \mid X_i\cap X_j\neq\Phi,i,j\in t\}. \tag{2}$$

[例1]令 $U=\{\vec{x}_1,\vec{x}_2,\cdots,\vec{x}_9\}$ 是二维坐标上点集. 点的坐标如图1所示:

根据定义1,通过计算得到:

$X_1=\{x_1,x_2,x_9,x_{10}\}$, $X_7=\{x_2,x_7,x_9\}$,
$X_2=\{x_1,x_2,x_7,x_9,x_{10}\}$, $X_8=\{x_3,x_4,x_6,,x_8,x_{11}\}$,
$X_3=\{x_3,x_6,x_8,x_{11}\}$, $X_9=\{x_1,x_2,x_5,x_7,,x_9,x_{10}\}$,
$X_4=\{x_4,x_6,x_8\}$, $X_{10}=\{x_1,x_2,x_5,x_9,x_{10}\}$,
$X_5=\{x_5,x_9,x_{10}\}$, $X_{11}=\{x_3,x_6,x_8,x_{11}\}$.
$X_6=\{x_3,x_4,x_6,x_8,x_{11}\}$,

这里取欧氏距离 $Th_c=\sqrt{2}$.

根据定义2得到粗集分类为

$$\begin{aligned}Y_1&=X_1\cup X_2\cup X_5\cup X_7\cup X_9\cup X_{10}\\&=\{x_1,x_2,x_5,x_7,x_9,x_{10}\},\\Y_2&=X_3\cup X_4\cup X_6\cup X_8\cup X_{11}\\&=\{x_3,x_4,x_6,x_8,x_{11}\}.\end{aligned}$$

通过以上计算 $U=\{\vec{x}_1,\vec{x}_2,\cdots,\vec{x}_9\}$ 被聚类为两类 Y_1,Y_2.

可以看到用粗集的近似关系可以将距离接近的对象聚为一类. 但是发现,在聚类过程中存在着大量的并运算,以及比较运算. 为了使运算得以简化,提出以下定理.

定理1:如果根据定义2得到 $Y=\bigcup_t\{X_i \mid X_i\cap X_j\neq\Phi,i,j\in t\}$ 则可以直接得到 $Y=\{x_j \mid j\in t\}$.

证明:由题意知 $x_i\in X_i,x_j\in X_j$ 且 $X_i\cap X_j\neq\Phi\Rightarrow x_i,x_j\in X_i\cap X_j\Rightarrow x_i,x_j\in X_i$ 且 $x_i,x_j\in X_j\Rightarrow\{x_j \mid j\in t\}\subseteq Y=\bigcup_t\{X_i \mid X_i\cap X_j\neq\Phi,i,j\in t\}$

以上过程证明了 $\{x_j \mid j\in t\}$ 包含于 Y 下面证明 Y 中仅包含 $\{x_j \mid j\in t\}$:

假设有一变量 $x_d\in Y$ 但 $d\notin t$ 则由以上证明过程可知 x_d 必不属于任意两个 X 的交集,否则 $d\in t$. 所以 x_d 必定仅属于其中一类,不妨设 $x_d\in X_k$,但是根据题意必有 $x_d\in X_d, d\neq k$ 所以 $x_d\in X_d\cap X_k$ 推出 $d\in t$ 与假设矛盾. 所以 Y 中仅包含 $\{x_j \mid j\in t\}$.

由以上过程得:$Y=\{x_j \mid j\in t\}$.

如:在例1中根据定理1,可以直接得到:

$$\begin{aligned}Y_1&=X_1\cup X_2\cup X_5\cup X_7\cup X_9\cup X_{10}\\&=\{x_1,x_2,x_5,x_7,x_9,x_{10}\},\\Y_2&=X_3\cup X_4\cup X_6\cup X_8\cup X_{11}\\&=\{x_3,x_4,x_6,x_8,x_{11}\}.\end{aligned}$$

1.2 边界点的粗糙度

[例 2]再进一步讨论另外一种情况

如图 2 所示,在图 1 的基础上再增加一个点 x_{12} 利用定义 1,定义 2,得:

$$X_{12} = \{x_3, x_9, x_{12}\},$$

进而可以聚类得到:

$$Y = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}.$$

可以看到由于 x_{12} 而使原来的 Y_1, Y_2 两类变为一类. 显然这与我们的理解是不同的. 这种情况很容易使聚类结果受到噪声点的干扰. 为了减少噪声点的干扰,给出不可分辨测度的分类定义,如下:

定义 3:在论域 U 上,根据定义 1 定义基于相似关系的粗集分类为:$R^-(X) = \{Y_1, Y_2, \cdots Y_f\}$

其中

$$Y_t = \bigcup_t \{X_t \mid card(X_t \cap X_j) \geqslant indis,\ i,j \in t\}. \tag{3}$$

indis 为分类的不可分辨测度.

当 indis = 1 可以看到定义 3 就退化成了定义 2,因此定义 3 可以看作是定义 2 的一个推广形式. 根据定义 3,重新计算例 2,取 indis = 3

$$X_1 = \{x_1, x_2, x_9, x_{10}\}, \quad X_7 = \{x_2, x_7, x_9\},$$
$$X_2 = \{x_1, x_2, x_7, x_9, x_{10}\}, \quad X_8 = \{x_3, x_4, x_6, x_8, x_{11}\},$$
$$X_3 = \{x_3, x_6, x_8, x_{11}, x_{12}\}, \quad X_9 = \{x_1, x_2, x_5, x_7, x_9, x_{10}, x_{12}\},$$
$$X_4 = \{x_4, x_6, x_8\}, \quad X_{10} = \{x_1, x_2, x_5, x_9, x_{10}\},$$
$$X_5 = \{x_5, x_9, x_{10}\}, \quad X_{11} = \{x_3, x_6, x_8, x_{11}\},$$
$$X_6 = \{x_3, x_4, x_6, x_8, x_{11}\}, \quad X_{12} = \{x_3, x_9, x_{12}\}.$$

$$Y_1 = X_1 \cup X_2 \cup X_5 \cup X_7 \cup X_9 \cup X_{10} = \{x_1, x_2, x_5, x_7, x_9, x_{10}\},$$

$$Y_2 = X_3 \cup X_4 \cup X_6 \cup X_8 \cup X_{11} = \{x_3, x_4, x_6, x_8, x_{11}\},$$

$$Y_3 = Y_{12} = \{x_3, x_9, x_{12}\}.$$

从以上结果可以看出根据定义 3 得到 3 个粗分类 Y_1, Y_2, Y_3,其中 $\{x_9\} = Y_1 \cap Y_3$, $\{x_3\} = Y_2 \cup Y_3$ 称之为边界点集合. 也就是说对于边界点 $\{x_9\}$, $\{x_3\}$ 具体属于哪一类是不明确的.

为了确定边界点的归属,提出了以下定义:

定义 4:(边界点最大隶属原则)设通过定义 2 得到边界点集合为 $\{A_1, A_2, \cdots A_m\}$,其中 A_i 是由 $\{Y_1, Y_2, \cdots Y_t\}$ 共同形成的边界点集合,认为 A_i 相对隶属于 Y_a. 如果满足:

$$\mathrm{card}(R^-(A_i) \cap Y_a) = \max_j(\mathrm{card}(R^-(A_i) \cap Y_j)),\quad j \in [1,t]. \tag{4}$$

根据定义 4,得到粗分类的粗糙度:

定义 5:设通过定义 2 得到边界点集合为 $\{A_1, A_2, \cdots A_m\}$ 则由定义 4 得到相对隶属于 Y_a 的 A_i 的粗糙度为:

$$P(A_i) = \mathrm{card}(R^-(A_i) \cap Y_a) / \sum_j (\mathrm{card}(R^-(A_i) \cap Y_j)). \tag{5}$$

定义 6:在定义 5 的基础上,整个分类的粗糙度为:

$$P(A) = 1 - \frac{1}{\sum_i \mathrm{card}(A_i)} \cdot \left(\sum_i \mathrm{card}(A_i) P(A_i)\right). \tag{6}$$

例如:对于[例 2]的边界点集合 $A_1 = \{x_9\}$,得到

$$R^-(A_1) = X_1 \cup X_2 \cup X_5 \cup X_7 \cup X_9 \cup X_{10} X_{12} = \{x_1, x_2, x_3, x_5, x_7, x_9, x_{10}, x_{12}\},$$

$$\mathrm{card}(R^-(A_1) \cap Y_1) = \mathrm{card}\{x_1, x_2, x_5, x_7, x_9, x_{10}\} = 6,$$

$$\mathrm{card}(R^-(A_1) \cap Y_3) = \mathrm{card}(x_3, x_9, x_{12}) = 3.$$

由于 $card(R^-(A_1) \cap Y_1) > card(R^-(A_1) \cap Y_3)$

表 2　数据相似度

Table 2 Data similarity degree

$s(x_i, x_j)$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
x_1	0	0.1672	0.2355	0.0811	0.2014	0.2895	0.0549	0.2434	0.2198	0.1099	0.2139	0.2613
x_2	0.1672	0	0.1718	0.0974	0.0549	0.2197	0.1461	0.1099	0.1945	0.1442	0.0487	0.0974
x_3	0.2355	0.1718	0	0.1648	0.1296	0.0549	0.1820	0.1006	0.0487	0.1297	0.1945	0.2255
x_4	0.0811	0.0974	0.1648	0	0.1220	0.2198	0.0487	0.1623	0.1612	0.0648	0.1461	0.1948
x_5	0.2014	0.0549	0.1296	0.1220	0	0.1718	0.1672	0.0549	0.1621	0.1461	0.0648	0.1006
x_6	0.2895	0.2197	0.0549	0.2198	0.1718	0	0.2355	0.1296	0.0811	0.1820	0.2353	0.2593
x_7	0.0549	0.1461	0.1820	0.0487	0.1672	0.2355	0	0.2014	0.1648	0.0549	0.1948	0.2436
x_8	0.2434	0.1099	0.1006	0.1623	0.0549	0.1296	0.2014	0	0.1442	0.1672	0.1099	0.1296
x_9	0.2198	0.1945	0.0487	0.1612	0.1621	0.0811	0.1648	0.1442	0	0.1099	0.2255	0.2620
x_{10}	0.1099	0.1442	0.1297	0.0648	0.1461	0.1820	0.0549	0.1672	0.109	0	0.1903	0.2375
x_{11}	0.2139	0.0487	0.1945	0.1461	0.0648	0.2353	0.1948	0.1099	0.2255	0.1903	0	0.0487
x_{12}	0.2613	0.0974	0.2255	0.1948	0.1006	0.2593	0.2436	0.1296	0.2620	0.2375	0.0487	0

所以认为 $x_9 \in Y_1$ 同理可以得到 $x_3 \in Y_3$

所以最终得到分类为

$Z_1 = \{x_1, x_2, x_5, x_7, x_9, x_{10}\}$,

$Z_2 = \{x_3, x_4, x_6, x_8, x_{11}\}$,

$Z_3 = \{x_{12}\}$.

由定义5和定义6可以得到粗糙度为

$P(A_1) = 6/(6+3) = 2/3$,

$P(A_2) = 5/(5+3) = 5/8$,

$$P(A) = 1 - \frac{1}{2}(2/3 + 5/8) = 0.354.$$

1.3 实例分析

[例3]如表1所列,由于马氏距离对一切非奇异性变换都是不变的,这说明它不受特征量的纲的选取的影响,并且最平移不变的. 所以在这里使用 $s(x_i, x_j) = \frac{d_M(x_i, x_j)}{\max d_M(x_u, x_v)}$ 如果除了数值属性外还有其它非数值属性,可以定义其它的相似度.

表2中的数据是 x_i, x_j 之间的相似关系 $s(x_i, x_j)$. 对于阈值得的选取有很多方法[6]. 在这里仅简单的使用极小点阈值法. 首先将数据进行增序排列. 然后求出所有的极小点,并以第一个极小值点作为需要的阈值. 对于例3,经计算

$s(x_1, x_j) = 0.1099, s(x_2, x_j) = 0.0549, s(x_3, x_j) = 0.0549,$

$s(x_4, x_j) = 0.0648, s(x_5, x_j) = 0.0648, s(x_6, x_j) = 0.0811,$

$s(x_7, x_j) = 0.0487, s(x_8, x_j) = 0.0549, s(x_9, x_j) = 0.0487,$

$s(x_{10}, x_j) = 0.0648, s(s_{11}, x_j) = 0.0648, s(x_{12}, x_j) = 0.0487,$

$X_1 = \{x_1, x_4, x_7, x_{10}\}, X_2 = \{x_2, x_5, x_{11}\}, X_3 = \{x_3, x_6, x_9\},$

$X_4 = \{x_4, x_7, x_{10}\}, X_5 = \{x_2, x_5, x_8, x_{11}\}, X_6 = \{x_3, x_6, x_9\},$

$X_7 = \{x_4, x_7\}, X_8 = \{x_5, x_8\}, X_9 = \{x_3, x_9\},$

$X_{10} = \{x_4, x_7, x_{10}\}, X_{11} = \{x_2, x_5, x_{11}, x_{12}\}, X_{12} = \{x_{11}, x_{12}\}.$

如果设不可测度 indis = 1 即按定义2并由定理1进行聚类则可以得到聚类为

表1 数据的属性值

Table 1 Data attributes

U	att. 1	att. 2	att. 3	att. 4
x_1	0.2	0.1	0.1	0.2
x_2	0.3	0.4	0.2	0.5
x_3	0.6	0.2	0.5	0.3
x_4	0.3	0.2	0.2	0.3
x_5	0.4	0.4	0.3	0.5
x_6	0.7	0.2	0.6	0.3
x_7	0.3	0.1	0.2	0.2
x_8	0.5	0.4	0.4	0.5
x_9	0.6	0.1	0.5	0.2
x_{10}	0.4	0.1	0.3	0.2
x_{11}	0.3	0.5	0.2	0.6
x_{12}	0.3	0.6	0.2	0.7

$Y_1 = X_1 \cup X_4 \cup X_7 \cup X_{10} = \{x_1, x_4, x_7, x_{10}\}$,

$Y_2 = X_2 \cup X_5 \cup X_8 \cup X_{11} \cup X_{12} = \{x_2, x_5, x_8, x_{11}, x_{12}\}$,

$Y_3 = X_3 \cup X_6 \cup X_9 = \{x_3, x_6, x_9\}$.

如果设不可测度为 indis = 2 得到的聚类与 indis = 1 的相同.

如果设不可测度为 indis = 3 得到以下聚类:

$Y_1 = X_1 \cup X_4 \cup X_{10} = \{x_1, x_4, x_7, x_{10}\}$,

$Y_2 = X_2 \cup X_5 \cup X_{11} = \{x_2, x_5, x_8, x_{11}, x_{12}\}$,

$Y_3 = X_3 \cup X_6 = \{x_3, x_6, x_9\}$,

$Y_4 = X_7 = \{x_4, x_7\}, Y_5 = X_8 = \{x_5, x_8\}$,

$Y_6 = X_{12} = \{x_{11}, x_{12}\}, Y_7 = X_9 = \{x_3, x_9\}$.

根据定义4,得到最终的聚类结果为

$Z_1 = \{x_1, x_4, x_7, x_{10}\}$,

$Z_2 = \{x_2, x_5, x_8, x_{11}, x_{12}\}$,

$Z_3 = \{x_3, x_6, x_9\}$.

根据定义4,5,6得到聚类的粗糙度为

$P(A_1) = 4/6 = 2/3$, $P(A_2) = 5/7$,

$P(A_3) = 5/7$, $P(A_4) = 5/3$,

$P(A) = 0.06$.

此处:

$A_1 = \{x_4, x_7\}, A_2 = \{x_5, x_8\}, A_3 = \{x_{11}, x_{12}\}, A_4 = \{x_3, x_9\}$ 是边界点集合. 看到最终的聚类结果仍然与 indis = 1 时的结果相同.

1.4 聚类的评价函数

在本文中,可以看到最初聚类的数目最少,边界点的粗糙度最小,但是聚类往往不准确,随着 indis 增大,数目越多,但粗糙度越大,为此根据聚类的数目同时考虑边界点的粗糙度. 提出了一个更简单的聚类质量的评价函数:

定义7:设 R 是在 indis = 1 时的相似关系. R' 是在 indis = $t(t>1)$ 时的相似关系将 U 聚为 $R' = \{Y_1, Y_2, \cdots Y_m\}$ 则定义聚类质量为

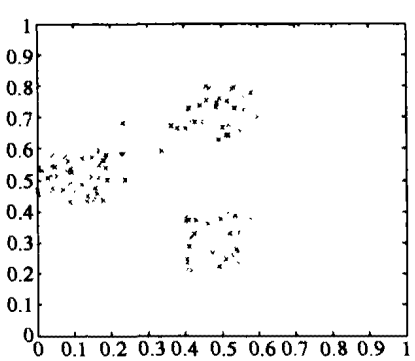

图3 初始数据

Fig. 3 Initial data

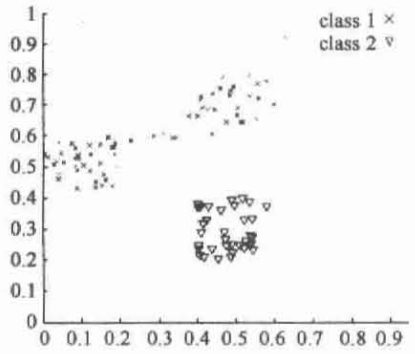

图 4　indis = 1 的聚类结果
Fig. 4　Clusters when indis = 1

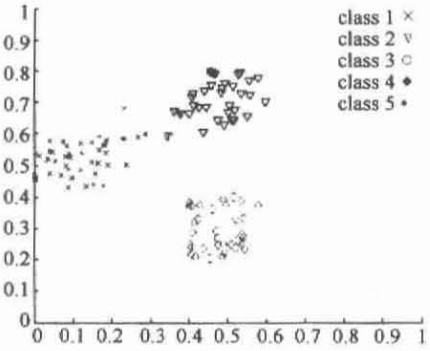

图 5　indis = 2 时边界点集合
Fig. 5　Boundary points set when indis = 2

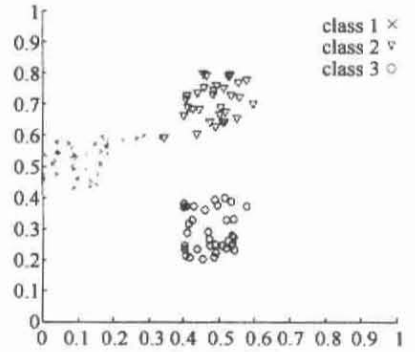

图 6　indis = 2 时边界点所属类别
Fig. 6　Classes of boundary points when indis = 2

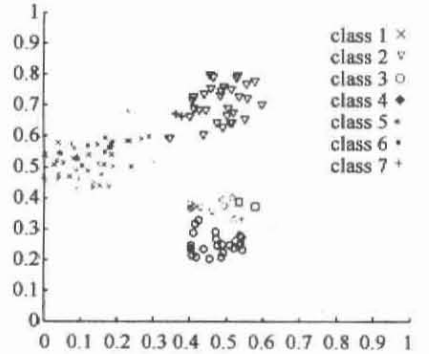

图 7　indis = 5 时的聚类结果
Fig. 7　Clusters when indis = 5

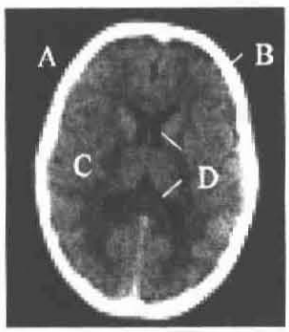

图 8　头部 CT 切片
Fig. 8　A CT slice of head

$$Q(R') = m(1 - P(A)). \tag{7}$$

其中:m 是在关系 R' 下聚类所得的类的数目,A 是在 R' 下聚类中的边界点集合,$P(A)$ 是在 indis = t(t > 1) 的边界点集合的粗糙度.

根据定义 7 可得到例 3 在 indis = 3 下的聚类质量为

$$Q(R') = 2.88.$$

同理例 2 在 indis = 3 下的聚类质量为

$$Q(R') = 1.938.$$

2　仿真结果

首先将方法应用到包含 120 个两维数据点的对象中. 图 3 ~ 图 7 给出了这 120 个点的聚类过程的仿真结果. 根据仿真结果可以看到数据点随不可分辨测度 indis 变化而变化的聚类过程. 图 3 是原始数据集合. 图 4 是在 indis = 1 即根据定义 2 得到的聚类结果. 在图 4 中可以看到由于类间噪声点的干扰,将去本来应该分开的类别误聚为一类. 而没有受噪声点干扰的对象则不受影响.

图 5 是 indis = 2 时的聚类结果. 可以看到在 indis = 2 时数据点被聚为五类,其中有三类属于边界点界合. 根据定义 4 的最大隶属原则,可以确定边界点的所属类. 进而可以得到 indis = 2 时最终的聚类结果(图 6). 但是当 indis = 5(图 7)时可以看到聚类结果已经开始恶化. 这说明不可能通过 indis 的提高来无限制的改善聚类的结果. 以上过程可以知道最初聚类的数目最少,边界点的粗糙度最小,但是聚类往往不准确,随着 indis 增大,数目越多,但粗糙度越大,因此在聚类时必须根据实际的情况来选择合适的 indis. indis 给人们提供了一个衡量聚类质量的参考尺度.

3　实验结果

下面以头部的一个 CT 切片(图 8)为例说明本

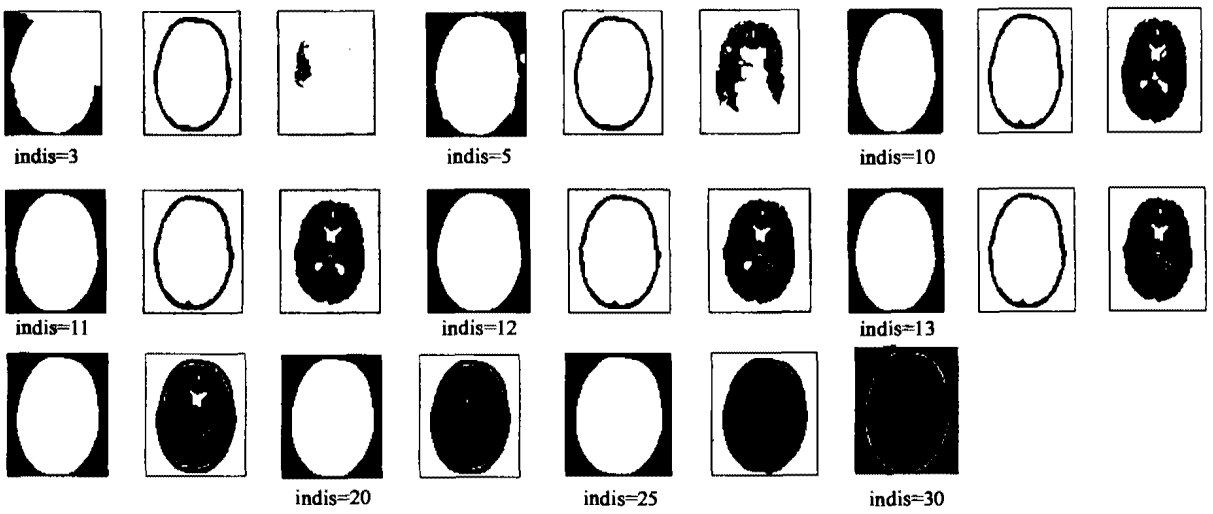

图9 不同 indis 对应的分割结果
Fig. 9 Clustering results with different indis

方法的应用过程. 其中,A 为背景,B 为骨质,C 为脑组织,D 为脊髓液.

图9表示了在不同的 indis 下的聚类结果. 实际的聚类数目为 3 ~ 10,但是仅表示出有代表性的几个样本. 用黑色表示分割出来的部分. 当 indis = 3 时可以看到,聚类质量是非常差的,甚至背景也无法从图像中分割出来,仅仅是骨组织可以分割出来,而脑组织仅可以得到一小部分. 当 indis = 5,分割结果得到改善,因为背景除了受到一小部分干扰外,大部分可以分割出来,此时脑组织可以分割出来的部分逐渐增多,但是脊髓液依然无法看到. 继续增加 indis 的值,当 indis = 10 与 indis = 11 时,可以清楚的看到骨组织,脑组织以及脊髓液已经可以完全的分割开来. 进一步增加 indis,当 indis = 12,13,15 时,可以看到脊髓液部分逐渐减小,到 indis = 20,25 时脑组织已经完全无法与脊髓液相互区分开来. 并且当 indis > 25 时,骨组织已经与脑组织融合,无法区分. 当 indis = 30 时,背景已完全无法从图像中分割出来,此时继续分割已丧失了任何意义.

4 结语

本文提出了基于相似关系的粗集聚类方法. 文章将粗集理论应用到相似关系中来解决聚类的问题,与传统方法相比,该方法在处理相似关系中的不精确,不确定问题上更具优势. 方法提出了边界点的最大隶属原则并进而提出了边界点粗糙度的概念. 根据边界点的最大隶属原则划归边界点的归属从而减少了噪声的干扰. 并在此基础上给出了聚类质量的评价函数. 仿真结果表明了聚类结果随不可分辨测度 indis 的变化过程,并分析了如何利用边界点的划分来防止噪声的干扰. 最后将本方法应用到医学图像的分割中,得到了很好的结果.

在本文中 indis 的选取是一个重要的内容,indis 选取的好坏将直接影响到聚类的结果,在本文中讨论了聚类结果随 indis 变化的过程. 在实际应用中 indis 需要根据实际情况进行选取.

REFERENCES

[1] PAWLAK Z. Vagueness and uncertainty: A Rough Set Prospective[J]. *Inter. J. of Computer Interlligence*, 1995, **11**(2):227—232.

[2] David Eppstein. Fast hierarchical clustering and other applications of dynamic closest pairs[C]. *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 619—628, California:San Francisco, 1998, 25—27.

[3] Pal N R, Bezdek J C. On cluster validity for the fuzzy c-means model[J]. *IEEE Transactions on Fuzzy Systems* 3, 1995, **3**:370—379.

[4] Choubey S K, Deogun J S, Raghavan V V, *et al*. A comparison of feature selection algorithms in the context of rough classifiers[C]. Proceedings of the Fifth IEEE International Conference on, 1996, **2**: 1122—1128.

[5] Fu shixing, Lu Zengxiang, Lu Haiming, *et al*. Rough set theory and its practice in knowledge discovery[C]. Proceedings of the 3rd World Congress on Intelligent Control and Automation. 2000, 2535—2539.

[6] Gonzalez R C, Woods R E. *Digital Image Processing*[M]. 3 edition, Addison Wesley Publishing Company, 1992.