

基于 Rough 集的信息融合目标识别方法*

曹治国,吴一飞,张天序

(华中科技大学 图像识别与人工智能研究所)

图像信息处理与智能控制教育部重点实验室,湖北 武汉 430074)

摘要:研究了 Rough 集理论在基于不同谱段光学信息融合的目标识别系统中的应用技术,在介绍了目标知识表示方法的基础上,提出了用 C 均值算法实现连续信息离散化的方法,并通过决策表属性的约简来生成决策表,以实现目标的识别。针对可见光、中/长波红外等三种不同谱段下的三种不同目标图像进行了实验,实验结果表明,该方法具有对先验知识获取的易实现性和对目标识别的有效性。

关键词: Rough 集; 信息融合; 目标识别

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1007-2276-(2005)01-0102-04

Approach of target recognition by data fusion based on Rough sets*

CAO Zhi-guo, WU Yi-fei, ZHANG Tian-xu

(Institute of Image Processing and Artificial Intelligence, State Key Laboratory of Education Ministry for Image

Processing and Intelligent Control, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: An application of Rough sets theory in target recognition system is researched based on data fusion of the multi-spectrum images. The method of expressing the information of the target is introduced, and the idea of dispersing the continued data by using the C-means method is presented. By reducing the decision-making table, targets recognition are realized. It is used to recognize the target from the light, middle and long wavelength infrared images in this paper. Experimental results prove that this approach is good for the target recognition and can easily get the pre-knowledge.

Key words: Rough sets; Data fusion; Target recognition

0 引言

常用的信息融合技术包括贝叶斯估计法、D-S 证据理论、模糊集理论、神经网络技术等^[1-3]。贝叶斯估计法的难点在于先验概率的确定,同时,它无法区分“不相信”与“缺乏可信度”等不确定性知识。D-S 证据

理论不仅能表示信息的已知性和确定性,还能够区分信息的未知性和不确定性,但当处理的信息种类较多时,存在着数据的组合爆炸问题。模糊集理论成功应用成功的关键在于找到合适的隶属度函数。

由于 Rough 集理论^[4]为不确定数据分析、推理,发现数据间的关系、提取有用特征,简化信息处理,

收稿日期:2004-01-17; 修订日期:2004-03-17

* 基金项目:国家 863 资助项目(2002 AA731212)

作者简介:曹治国(1964-),男,湖北武汉人,教授,博士,主要研究方向为图像处理、模式识别等。

研究不精确、不确定知识的表达、学习、归纳方法提供了一个有力的工具,因此本文提出了一种将 Rough 集理论应用于不同谱段光学信息融合的目标识别系统中的方法。针对可见光、中波红外、长波红外等三种不同谱段下的三种不同目标图像的实验结果表明,该方法在多源信息融合的目标识别领域有很大的应用前景。

1 Rough 集的理论基础

1.1 Rough 集的概念

设 U 是研究对象 X 的全集, $X \subseteq U$, R 是定义在 U 上的等价关系, $A=(U, R)$, 是一个近似空间, 在 A 上, 如果 X 是一些 R -基本类的并集, 则称 X 是 R -可定义的; 否则称 X 是 R -不可定义的。 R -可定义集是全集 U 上的一些子集, 这些子集在个体全集 U 上恰好是可被定义的, 而 R -不可定义集是子集 X 上不可能被恰好定义的。 R -可定义集被称作 R -一致集, 而 R -不可定义集被称作 R -不一致集, 即 R -Rough 集, 简称 Rough 集^[5]。如果存在一个等价关系 $R \in INDU$, 其中 $INDU$ 是 U 上给定的所有等价关系的集合族, 使得 $X \subseteq U$ 对任意 $R \in INDU$ 都是 R -Rough 的, 则 X 是 U 上的 Rough 集。

1.2 集合的下近似、上近似概念

设 $X, Y \subseteq U, R$ 是定义在 U 上的等价关系, 集合 X 关于 R 的下近似为:

$$R_*(X) = \cup \{Y \in U/R : Y \subseteq X\} \quad (1)$$

式中 $R_*(X)$ 是根据现有知识判断肯定属于 X 的等价类的最大并集, 称为 X 的 R -正区域, 被记为 $POS_R(X) = R_*(X)$ 。而根据现有知识判断肯定不属于 X 的等价类的最大并集称为负区域, 被记为 $NEG(X)$ 。

集合 X 关于 R 的上近似为:

$$R^*(X) = \cup \{Y \in U/R : Y \cap X \neq \phi\} \quad (2)$$

式中 ϕ 是空集; $R^*(X)$ 是可能属于 X 的等价类的最小集合。 X 集合的边界线集定义为:

$$BN_R(X) = R^*(X) - R_*(X) \quad (3)$$

它是那些通过等价关系 R 即不能被完全定义在 X 上的子集, 也不能被定义在 X 上的补集 \bar{X} 上的集合。显然, 如果 $BN_R(X)$ 是空集, 那么 X 关于 R 是一致的; 对于 Rough 集, $BN_R(X) \neq \phi$ 。为了精确表示 X 的这种

Rough 近似精度, 定义集合 X 的粗糙度为:

$$\alpha_R(X) = \text{card}(R_*(X)) / \text{card}(R^*(X)) \quad (4)$$

式中 $\text{card}(S)$ 表示 S 的基数, 即求集合成员个数的函数。

1.3 Rough 集的简约与核

任何目标的识别都是由它的众多属性决定的, 这些属性描述的信息在进行目标识别时, 需要得到一个对信息的决策规则。然而, 在描述同一现象的各种信息中, 包含着许多重复或者是矛盾的信息, 这可以利用 Rough 集理论对其进行简约, 从中找出描述现象本质的数据, 并将简约后的信息重新组合以产生新的决策规则。抽象地看, 各种属性描述可看成是一种记录关系。

定义 1: 设 R 是一个等价关系族, $r \in R$, 如果 $IND(R) = IND(R - \{r\})$, 则称 r 在 R 中是可被约去的知识; 如果 $P = R - \{r\}$ 是独立的, 则 P 是 R 中的一个简约。通常 R 的简约不止一个, 其所有简约族记为 $RED(R)$ 。

定义 2: 对任一属性 $r \in R$, 如果 $IND(R) \neq IND(R - \{r\})$, 则称 r 在 R 中是不可被约去的, R 中所有不可约去的关系称为核, 由它构成的集合称为 R 的核集, 记为 $CORE(R)$, 并且有 $CORE(R) = \cap RED(R)$ 。

2 基于 Rough 集理论的数据融合在图像目标识别中的应用

一个信息的知识表达可写为 $S=(U, R, C, D)$, 其中 U 是对象的集合, 即论域, R 是属性的集合, $C, D \subset R$, 是属性集 R 的两个子集, 分别称 C 和 D 为 R 的条件属性和决策属性。在图像的目标识别系统中, R 表示图像目标提取的信息表, C 表示目标的属性描述, D 表示目标的类型表述。

实际的目标图像识别问题中, 首先提取它的相关特征信息形成目标的知识表达。如某个图像样本的 5 个相关特征属性值: {均值, 方差, 能量, 熵, 对比度}, 分别用 C_1, C_2, C_3, C_4, C_5 表征, 谱段信息用 C_6 表征, 从而得到目标属性描述子集 C 的一个样本向量 $C = \{C_1, C_2, C_3, C_4, C_5, C_6\}$, 目标的类型为决策 D 子集, $D = \{1, 2, 3\}$ 。以 20 幅三种不同类型不同谱段目标图像为例, 形成表 1 的知识表达表。表 1 中 $C_6 = \{1, 2, 3\}$,

分别对应图像的{可见光,中波红外,长波红外}谱段信息。

由于 Rough 不能处理连续信息,需要对如上信息表进行离散量化。最简单的离散方法是对同一特征子集,如 C_1 子集,统计其连续属性值的分布,给定 N 个门限值划分连续属性值得到 $N+1$ 个离散结果。

表 1 实例数据表

Tab.1 Data of an example

U	C_1	C_2	C_3	C_4	C_5	C_6	D
1	170	1115	0.009	6.8	1.39	3	3
2	172	1086	0.009	6.8	1.37	3	3
3	170	1157	0.009	6.8	1.39	3	3
4	158	726	0.011	6.5	1.31	1	2
5	203	731	0.011	6.5	1.26	2	2
6	204	736	0.011	6.5	1.26	2	2
7	188	6918	0.011	6.5	1.26	2	2
8	194	851	0.013	6.4	1.31	2	2
9	184	1014	0.009	6.8	1.34	3	3
10	200	821	0.010	6.6	1.28	3	2
11	168	804	0.012	6.5	1.34	1	2
12	156	791	0.013	6.4	1.37	1	2
13	213	365	0.016	6.0	1.16	2	1
14	210	439	0.016	6.1	1.19	3	1
15	207	585	0.012	6.3	1.22	3	1
16	197	695	0.013	6.4	1.26	1	1
17	198	692	0.012	6.4	1.25	1	1
18	183	1359	0.011	6.7	1.43	2	3
19	193	10771	0.011	6.5	1.35	1	3
20	209	824	0.011	6.4	1.27	3	2

为了达到较好的离散效果以及对决策的指导作用,参考模式识别的动态 C -均值聚类思想^[6],提出动态均值量化算法来实现对连续数据的离散化。

若将子集 C_1 的所有连续属性值样本通过动态均值量化算法量化为 M 个离散值,即 M 级量化,若 N_i 是第 i 个离散值 Γ_i 的样本数目, m_i 是这些离散值为 Γ_i 的样本的原始连续属性的均值,定义 M 级离散化的误差平方和为:

$$J_M = \sum_{i=1}^M \sum_{y \in \Gamma_i} \|y - m_i\|^2 \quad (5)$$

式中 J_M 是样本集和量化集的函数,度量了用 M 个离散值量化连续数据时的总误差,使得 J_M 最小的量化

方法是量化级数为 M 时的最优方法。通常,对不同的量化级数,当量化级数 M 越大, J_M 越小,但当连续数据集是一个比较好的 x 级集中量化结果时, J_M 会随着量化级数从 1 增加到 x 再快速单调减小,而当量化级数继续增多时, J_M 的减小就变得十分缓慢,相当于将比较集中的量化细分开。所以,找到 x 这个比较适当的量化级数就能较好地达到对连续数据的离散化。

由于表 1 的数据量较小,对包含连续属性特征的 $C_1 \sim C_5$ 子集可采用两级分别量化, $C_1 \sim C_5$ 中的量化均值见表 2,这样表 1 的数据变为 0、1 两级离散值。

表 2 量化均值表

Tab.2 Means of classes

Classes	C_1	C_2	C_3	C_4	C_5
0	170.5	696.8	0.0108	6.42	1.245
1	201.7	1135.1	0.0142	6.79	1.364

表 3 简化后决策表与核值

Tab.3 Decision after reducing and core data

U	C_2	C_3	C_5	C_6	D
1	0	0	$\Delta 0$	$\Delta 1$	1
2	0	$\Delta 0$	0	$\Delta 2$	2
3	0	$\Delta 0$	0	$\Delta 3$	2
4	$\Delta 0$	0	$\Delta 1$	1	2
5	0	1	$\Delta 0$	1	1
6	0	$\Delta 1$	$\Delta 0$	2	1
7	0	$\Delta 1$	0	3	1
8	0	1	$\Delta 1$	1	2
9	0	1	$\Delta 1$	2	2
10	$\Delta 1$	0	1	1	3
11	1	0	1	2	3
12	1	0	1	3	3

由定义 1 和定义 2,对量化后的数据表进行相同决策和并与决策属性简约,并提取决策关系核,得到决策简约表,如表 3 所示,表中数据前标记“ Δ ”的为核值。可以看到,在这个决策系统中,属性关系 C_1 和 C_4 对于决策系统是可以被约去的。类似的,由已知的核值,通过简约与合并,约去简化规则中过剩的条件属性值,然后合并相同的最简决策规则,得到一个最小简约决策关系表,用于对目标的识别,如表 4 所示,“*”表示该条件属性的属性值不影响决策属性。

表 4 最简决策表

Tab.4 Finally decision

U	C_2	C_3	C_5	C_6	D
1	*	*	0	1	1
2	0	0	*	2	2
3	*	0	0	3	2
4	0	*	1	*	2
5	*	1	0	*	1
6	1	*	*	*	3

3 实验结果与分析

用 324 幅图像作为决策规则的训练集,训练集包括可见光、中波红外、长波红外等不同谱段下的三种不同目标图像。提取目标的均值、方差、能量、熵、对比度、一阶不变矩、二阶不变矩、三阶不变矩、四阶不变矩等 9 种不同特征,生成决策规则进行目标识别。在离散化连续特征时,有两种方法:(1)对每个连续特征子集用动态均值量化方法离散特征;(2)对每个有连续属性值的特征子集分别统计其样本值的分布,给出两个门限将连续值都离散化为 3 个不同的值。用另外 324 幅图像作为识别集,分别用两种不同方法得到的决策规则进行识别,识别结果如表 5 所示。

表 5 识别结果

Tab.5 Recognition results

Method	Recogni- tion times	Right times	Wrong times	Refused times	Recognition rate /%
C-Mean	324	266	5	53	82.1
$N=3$	324	223	24	77	68.8

动态均值量化方法对不同特性的属性子集的量化结果不同,在本例中,对得到的 9 个不同特征属性子集的量化级数分别为:5、4、7、5、5、5、4、7、5;而属性 C_6 本来就是离散的,不用量化。

另外,基于向量空间模型和最小距离方法的最近邻决策方法^[6]是一种常用的目标识别方法。使用和上面相同的两组训练样本图像和识别样本图像提取特征信息,用最近邻决策方法进行训练和识别,识别结果如表 6 所示。最近邻方法对识别目标必须分类,所以没有拒识类别。

表 6 识别结果比较

Tab.6 Comparison of the recognition results

Method	Recogni- tion times	Right times	Wrong times	Refused times	Recognition rate /%
Most-near neighbor method	324	172	152		82.1
Rough sets	324	266	5	53	68.8

从以上结果可以看出,利用基于 Rough 理论的数据融合方法进行目标识别,可以得到较好的识别效果,且运算量较小。两种不同连续属性值离散方法的识别结果表明,动态均值量化方法明显优化了识别过程,通过离散化,提高融合识别的效果。

4 结论

实验结果表明 Rough 集理论方法能在没有先验知识的情况下,从实际数据信息中提取规则,有很强的解决不精确知识的能力,且能获得较好的识别性能。但是,怎样减少拒识率,进一步提高 Rough 集理论用于数据融合中的识别率,以及怎样进行新知识的学习,是我们下一步将研究的内容。

参考文献:

- [1] Liu Rujie, Yuan Baozong. A D-S based multi-channel information fusion method using classifier's uncertainty measurement[A]. Proceedings of the 5th International Conference on Signal Processing, WCCC-ICSP 2000[C]. Beijing: Publishing House of Electronics Industry, 2000. 1297-1300.
- [2] Solaiman B, Debon R, Pipelier F, et al. Information fusion: application to data and model fusion for ultrasound image segmentation[J]. IEEE Transactions on Biomedical Engineering, 1999, 46(10): 1171-1175.
- [3] Barbera H M, Skarmeta A G, Tzquierdo M Z, et al. Neural networks for sonar and infrared sensors fusion[A]. Proceedings of the Third International Conference on Information Fusion, Fusion 2000[C]. Paris: ONERA Press, 2000. 18-25.
- [4] Pawlak Z. Rough Sets-Theoretical Aspect of Reasoning about Data [M]. Dordrecht: Kluwer Academic Publishers, 1991.
- [5] 刘清. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001.
- [6] 边肇祺, 张长水, 张学工. 模式识别[M]. 北京: 清华大学出版社 1999. 136-140, 235-237.