

集成特征选择的广义粗集方法与多分类器融合

孙亮¹ 韩崇昭² 沈建京¹ 戴宁³

摘要 为改善多分类器系统的分类性能,提出了基于广义粗集的集成特征选择方法. 为在集成特征选择的同时获取各特征空间中的多类模式可分性信息,研究并提出了关于多决策表的相对优势决策约简,给出了关于集成特征选择的集成属性约简(Ensemble attribute reduction, EAR)方法,结合基于知识发现的 KD-DWV 算法进行了高光谱遥感图像植被分类比较实验. 结果表明,EAR 方法与合适的多分类器融合算法结合可有效提高多分类器融合的推广性.

关键词 集成特征选择, 多分类器融合, 广义粗集, 高光谱
中图分类号 TP391.4, TP75

Generalized Rough Set Method for Ensemble Feature Selection and Multiple Classifier Fusion

SUN Liang¹ HAN Chong-Zhao² SHEN Jian-Jing¹ DAI Ning³

Abstract For improving the performance of multiple classifier system, a novel method of ensemble feature selection is proposed based on generalized rough set. In the paper, the relative dominance decision reduct (RDDR) with respect to multiple decision tables is presented to obtain the best feature subsets and interclass separability from different feature spaces. Then, the ensemble attribute reduction (EAR) method is proposed for ensemble feature selection. Using the KD-DWV algorithm based on knowledge discovery, the effectiveness of EAR was examined with the vegetation classification on a hyperspectral image. The result of the comparison experiment shows that EAR can be used to improve the generalization of multiple classifier system by combining appropriate multiple classifier fusion algorithm.

Key words Ensemble feature selection, multiple classifier fusion, generalized rough set, hyperspectral

研究与实验表明,不同分类器间的相异性(Diversity)对提高多分类器系统的分类精度以及推广能力至关重要,而提高相异性的手段之一就是采用具有互补分类信息的多个特征集^[1-2],这些特征集可以是同一特征集的不同子集,也可以是异类或不同特征空间中的特征子集.所谓的集成特征选择,就是按一定策略和准则为多分类器系统中各基分类器进行特征子集选择,以期有效提高多分类器系统的总体分类能力^[2].

利用粗集属性约简方法进行特征选择有利于多类分类知识的发现、更新和管理,这对基于集成特征选择的多分类器融合具有重要价值.其应用中的问题主要有:1)经典的属性约简大多在协调的决策

表中进行,而大量的决策表是不协调的^[3];2)论域上的关系多利用各种距离测度进行构造,一般是非等价的,通常只能得到一个论域上的覆盖而非划分;3)对不同特征集或特征空间,往往需分别建立决策表,而关于多决策表的集成特征选择尚待研究.

特征选择方法主要有两种框架,即 Filter 法(或开环法)和 Wrapper 法(或闭环法).前者通常为基于类间可分性判据的方法,后者则是一种基于预测器的反馈方法.为解决上述问题,本文采用广义粗集模型讨论了不协调决策表中的广义属性约简,研究了基于广义属性约简的多决策表集成特征选择 Filter 方法,最后将集成特征选择中获得的多类模式可分性引入到多分类器决策融合中,以期进一步提高多分类器融合的质量.

1 特征选择的粗集方法

1.1 特征选择的经典粗集方法

应用粗集方法进行特征选择的关键是合理定义所需的判别准则.常见的准则定义主要是基于特征关联(Relevance of features)、互信息(Mutual information)和类别可分性(Interclass separability)^[4]的.其中,基于类别可分性的属性约简方法对了解特征空间中多类模式的分布结构十分有利,其基本出发点是 Fisher 线性判决思想,即:一个好的特

收稿日期 2007-05-30 收修改稿日期 2007-10-05
Received May 30, 2007; Year; in revised form October 5, 2007
国家自然科学基金(60574033),国家重点基础研究发展计划(973 计划)(2006CB504605, 2007CB311006)资助
Supported by National Natural Science Foundation of China (60574033), National Basic Research Program of China (973 Program) (2006CB504605, 2007CB311006)
1. 信息工程大学理学院 郑州 450001 2. 西安交通大学电子与信息工程学院综合自动化研究所 西安 710049 3. 郑州大学数学系 郑州 450052
1. Institute of Science, Information Engineering University, Zhengzhou 450001 2. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049 3. Department of Mathematics, Zhengzhou University, Zhengzhou 450052
DOI: 10.3724/SP.J.1004.2008.00298

征应该使类内离散度尽可能小, 而类间离散度尽可能大. 基于这一思想, Skowron 在文献 [5] 中提出了所谓的 D -约简, 该方法旨在保持总体类间可分性的同时完成特征选择.

1.2 特征选择的广义粗集方法

粗集属性约简往往依赖于离散化的属性值, 考虑到数据的不确定性和随机性, 一般离散化造成的分类信息丢失对多类和复杂分类问题往往难以容忍, 而基于类内、类间距离的特征选择又涉及非等价关系下的属性约简, 相应的决策表常为不协调的. 对此, 文献 [6] 研究了非离散属性值下的属性约简; 文献 [3, 7] 基于论域上的等价关系研究了不协调决策表上的属性约简; 文献 [8] 则基于广义粗集模型研究了非等价关系下特征选择的广义属性约简方法, 解决了 D -约简中无法根据距离度量生成可分辨矩阵的问题. 这些研究为进一步探讨关于集成特征选择的广义属性约简方法打下了基础.

2 关于集成特征选择的集成属性约简 (Ensemble attribute reduction, EAR)

利用粗集属性约简进行集成特征选择, 需建立关于多决策表的广义属性约简模型.

2.1 广义近似空间上的粗集近似

研究非等价关系下的广义属性约简问题, 可借助广义近似空间模型^[9].

设 (U, I, ν) 为广义近似空间, 其中 U 为有限论域, I 为集值函数, $\nu: \mathcal{P}(U) \times \mathcal{P}(U) \rightarrow [0, 1]$ 为 $\mathcal{P}(U)$ 上的包含度 ($\mathcal{P}(U)$ 为 U 的幂集), 即 $\forall Y, Z \in \mathcal{P}(U)$, 有

$$\nu(Y, Z) = \begin{cases} \frac{|Y \cap Z|}{|Y|}, & Y \neq \emptyset \\ 1, & \text{其他} \end{cases}$$

则对任意 $X \subseteq U$, 其关于 (U, I, ν) 的广义上、下近似可分别定义为

$$\begin{aligned} \bar{I}(X) &= \{x \in U : \nu(I(x), X) > 0\} \\ \underline{I}(X) &= \{x \in U : \nu(I(x), X) = 1\} \end{aligned}$$

由此, 可给出决策表上基于距离测度的广义粗集近似.

设 $(U, A \cup \{d\})$ 为决策表 ((U, A) 称为信息表), 其中 $A = \{a_1, \dots, a_m\}$ 为有限条件属性集, d 为决策属性, $\forall x \in U$, 其关于属性的数据模式可表达为 $\mathbf{a}(x) = (a_1(x), \dots, a_m(x))$, 这里, $a_l: U \rightarrow V_l$ ($l \leq m$) 为对象属性值映射, V_l 为 a_l 的值域. 决策属性值为 $d(x)$, 对 c 类分类问题, 其值域记为 $V_d = \{1, \dots, c\}$, $c \leq n$. 对 U 上的等价关系

$$R_d = \{(x, y) \in U \times U : d(x) = d(y)\}$$

和容差关系

$$\begin{aligned} \tau_B^h &= \{(x, y) \in U \times U : u_l(a_l(x), a_l(y)) \leq h \\ &\quad \forall a_l \in B\}, B \subseteq A \end{aligned} \quad (1)$$

($u_l: V_l \times V_l \rightarrow \mathbf{R}_+$ 是某种距离度量, $h > 0$ 为给定的阈值), 若定义集值函数

$$I_B^h(x) = \{y \in U : (y, x) \in \tau_B^h\} \quad (2)$$

则 R_d 、 τ_B^h 分别生成 U 上的划分 $\mathcal{A}(d) = U/R_d = \{X_1, \dots, X_c\}$ (称为专家分类) 和覆盖 $\mathcal{I}^h(B) = \{I_B^h(x) : x \in U\}$. 由此, $\forall X_j \in \mathcal{A}(d)$, 其关于 B 的广义上、下近似分别为

$$\begin{aligned} \bar{I}_B^h(X_j) &= \{x \in U : \nu(I_B^h(x), X_j) > 0\} \\ \underline{I}_B^h(X_j) &= \{x \in U : \nu(I_B^h(x), X_j) = 1\} \end{aligned}$$

2.2 广义优势决策约简 (Generalized dominance decision reduct, GDDR)

$\forall x \in U$, 记

$$\theta_B^h(x) = \{i : \nu(I_B^h(x), X_i) \neq 0, X_i \in \mathcal{A}(d)\} \quad (3)$$

为 x 关于属性集 $B \subseteq A$ 的 h 广义决策, 由此可给出多种广义属性约简定义.

定义 1. 设 $(U, A \cup \{d\})$ 为决策表, 对给定的 $h > 0$, 称 $B \subseteq A$ 为 h 决策约简, 若 $\forall x \in U$, $\theta_B^h(x) = \theta_A^h(x)$, 且 $\forall a \in B$, $\exists x \in U$ 使 $\theta_{B-\{a\}}^h(x) \neq \theta_A^h(x)$.

显然, h 决策约简保持了对每个对象的 h 广义决策不变 (即每个对象 x 的可能归类不变).

设 $(U, A \cup \{d\})$ 为决策表, $\forall x_p, x_q \in U$ 和任意 $B \subseteq A$, $B \neq \emptyset$, 记

$$\Lambda_{pq}^B = \{l : a_l \in B, u_l(a_l(x_p), a_l(x_q)) > h\} = \{l_1, \dots, l_k\}$$

其中 $0 < k = |\Lambda_{pq}^B| \leq |A|$, 并定义

$$\begin{aligned} dis_B(x_p, x_q) &= \\ &f(u_{l_1}(a_{l_1}(x_p), a_{l_1}(x_q)), \dots, u_{l_k}(a_{l_k}(x_p), a_{l_k}(x_q))) = \\ &f(u_l(a_l(x_p), a_l(x_q)) : l \in \Lambda_{pq}^B) \end{aligned} \quad (4)$$

为对象距离函数. 这里, $f: \mathbf{R}_+^k \rightarrow \mathbf{R}_+$ 为一适当函数, 使当 $\Lambda_{pq}^B = \emptyset$ 时 $f(0) = 0$, $\Lambda_{pq}^B \neq \emptyset$ 时 $f(\cdot) > h$ (如 $f(t_1 \cdots t_k) = (t_1^2 + \dots + t_k^2)^{1/2}$). $\forall x_p, x_q \in U$, $B \subseteq A$ 和 $E \subseteq B$, 若对象距离函数具有性质 $dis_E(x_p, x_q) \leq dis_B(x_p, x_q)$, 则称其关于属性单调.

定义 2. 设 (U, A) 为信息表, 对给定的阈值 $h > 0$, 称矩阵 $C = (C_{pq})_{p,q=1,2,\dots,n}$ 为 (U, A) 的距离可分辨矩阵, 若

$$C_{pq} = \begin{cases} \{a_i \in A : u_i(a_i(x_p), a_i(x_q)) > h\}, & p \neq q \\ \emptyset, & p = q \end{cases} \quad (5)$$

定义 3. 设 $(U, A \cup \{d\})$ 为决策表, 对给定的阈值 $h > 0$, 称非负矩阵 $D = (D_{pq})_{p,q=1,2,\dots,n}$ 为 $(U, A \cup \{d\})$ 的距离矩阵, 若 $dis_A(x_p, x_q) \neq 0$ 时 $h < D_{pq} \leq dis_A(x_p, x_q)$; $dis_A(x_p, x_q) = 0$ 时 $D_{pq} = 0$.

由定义 3 可见, D 设定了 U 中任意两对象间的距离下界. $\forall X_i \in \mathcal{A}(d)$, 进一步记

$$r_i(A) = \frac{|I_A^h(X_i)|}{|X_i|} \leq 1, \quad i = 1, \dots, c \quad (6)$$

并称 $K = \{k_0 : r_{k_0}(A) = \max_{k \leq c} r_k(A)\}$ 为优势决策集, $U_K = \bigcup_{i \in K} X_i$ 为优势对象集, 则基于定义 1 可进一步给出关于 D 的 GDDR 定义.

定义 4. 设 $(U, A \cup \{d\})$ 为决策表, 对给定的距离阈值 $h > 0$, K 为 $(U, A \cup \{d\})$ 的非空优势决策集, U_K 为相应的优势对象集, $D = (D_{pq})_{p,q=1,2,\dots,n}$ 为其距离矩阵, 称 $B \subseteq A$ (不唯一) 为关于 D 的 GDDR, 若 B 是满足下列条件的 A 的最小子集:

- 1) $\forall x \in U_K, \theta_B^h(x) = \theta_A^h(x)$,
- 2) $dis_B(x_p, x_q) \geq D_{pq}, x_p \in X_i, x_q \notin X_i, i \in K$.

定义 4 实际给出了一种特殊的特征选择判据, 在保持 h 广义决策不变的约束下, 以此选出的属性子集与同等基数的其他属性子集相比, 可使 $X_i (i \in K)$ 中的对象与 U 中其他对象间的距离相对最大. 显然, 这是一种保持局部模式最佳可分性的监督学习策略.

定理 1. 设 $(U, A \cup \{d\})$ 为决策表, K 为其非空优势决策集, $D = (D_{pq})_{p,q=1,2,\dots,n}$ 为给定的距离矩阵, 对象距离函数关于属性单调, 则关于 D 的 GDDR 总是存在的.

证明. 因 D 为给定的距离矩阵, $K \neq \emptyset$, 故 $\forall x_p \in X_i, x_q \notin X_i (i \in K), dis_A(x_p, x_q) \geq D_{pq}$ 成立. 又对象距离函数关于属性单调, 故 $\forall a \in A$, 若 $\exists x_p \in X_i, x_q \notin X_i (i \in K)$ 使 $\theta_{A-\{a\}}^h(x_p) \neq \theta_A^h(x_p)$ 或 $dis_{A-\{a\}}(x_p, x_q) < D_{pq}$, 则 A 本身就是 GDDR; 若对 $A - \{a\}$, 定义 2 的条件得以满足, 则研究 $B_1 = A - \{a\}$, 若 $\forall b_1 \in B_1, \exists x_p \in X_i, x_q \notin X_i (i \in K)$ 使 $\theta_{B_1-\{b_1\}}^h(x_p) \neq \theta_{B_1}^h(x_p)$ 或 $dis_{B_1-\{b_1\}}(x_p, x_q) < D_{pq}$, 则 B_1 就是 GDDR, 否则再研究 $B_2 = B_1 - \{b_1\}$, 重复以上过程. 由于 A 是有限集, 故总可找到 $B \subseteq A$ 符合定义 4, 则 B 即为关于 D 的 GDDR. \square

GDDR 的存在使从多个决策表中选出保持各

特征空间局部模式最佳可分性且具有互补分类能力的多个特征子集成为可能. 由此, 引入 EAR 概念.

2.3 相对优势决策约简 (Relative dominance decision reduct, RDDR) 与 EAR

设 $(U, A_1 \cup \{d\}), \dots, (U, A_N \cup \{d\})$ 分别为对应 N 个特征集的决策表, 记为 DT_1, \dots, DT_N , 它们具有相同的论域和专家分类, 相应的 $V_d = \{1, \dots, c\}$, 距离阈值分别为 h_1, \dots, h_N . 依式 (2) 定义集值函数, 分别记为 $I_{A_1}^{h_1}, \dots, I_{A_N}^{h_N}$; 依式 (4) 定义相应的对象距离函数, 分别记为 $dis_{A_1}, \dots, dis_{A_N}$; 由定义 2 和 3 生成相应的距离可分辨矩阵 $C^{(j)}$ 和距离矩阵 $D^{(j)} (j = 1, \dots, N)$. 设 U 上的专家分类 $\mathcal{A}(d) = \{X_1, \dots, X_c\}, \forall A_j (j \leq N)$ 和 $\forall X_i \in \mathcal{A}(d)$ 记

$$r_i(A_j) = \frac{|I_{A_j}^{h_j}(X_i)|}{|X_i|} \quad (7)$$

$$J_i = \{j_0 : r_i(A_{j_0}) = \max_{j \leq N} r_i(A_j)\} \quad (8)$$

$$J = \bigcup_{i=1}^c J_i \quad (9)$$

显然, $|J| \leq N$. 对决策表 $DT_j (j \in J)$, 基于 Fisher 线性判决思想构造准则函数

$$F_j(x_p) = \frac{S_B^{(j)}(x_p)}{S_W^{(j)}(x_p)} \quad (10)$$

其中, $\forall x_p \in X_i (i \in V_d)$,

$$S_B^{(j)}(x_p) = \min_{x_q \notin X_i} dis_{A_j}(x_p, x_q),$$

$$S_W^{(j)}(x_p) =$$

$$\begin{cases} \max_{x_q \in X_i} dis_{A_j}(x_p, x_q), \exists x_q \in X_i, dis_{A_j}(x_p, x_q) \neq 0 \\ h_j, \quad \forall x_q \in X_i, dis_{A_j}(x_p, x_q) = 0 \end{cases}$$

$\forall j \in J$, 进一步记

$$Z_j = \{i : x_p \in X_i \text{ 且 } F_j(x_p) = \max_{j' \in J} F_{j'}(x_p)\} \quad (11)$$

$$U_j = \{x_p : x_p \in X_i \text{ 且 } F_j(x_p) = \max_{j' \in J} F_{j'}(x_p)\} \quad (12)$$

并称 Z_j 为 DT_j 的相对优势决策集. 若 $Z_j \neq \emptyset$, 则称 DT_j 为相对优势决策表. 进一步记 $\mathcal{Z} = \{Z_j : Z_j \neq \emptyset\}$, 则当 $\bigcup_{Z_j \in \mathcal{Z}} Z_j = V_d$ 时, 称所有 $DT_j (Z_j \in \mathcal{Z})$ 完全互补, 否则为不完全互补. 对获得的全部 U_j 进行检验, 若它们相交不空, 则作修剪处理, 以使 $\bigcup_{Z_j \neq \emptyset} U_j$ 中的任意对象在所有 U_j 中仅出现一次, 并称检验处理后的 U_j 为 DT_j 的相对优势对象集.

从特征选择的角度看, J_i 指出对第 i 类分类利用哪些决策表进行属性约简最好, 故由式 (8) 和 (9) 可对全部决策表进行预选以减少后续计算量, J 为

预选出的决策表序号集; Z_j 标定利用 DT_j 进行属性约简对哪些类别分类最佳; U_j 显示依据 DT_j 中的哪些对象进行属性约简对提高第 i ($i \in Z_j$) 类的分类精度最有利. 如 $Z_2 = \{1, 3\}$ 、 $U_2 = \{x_1, x_2, x_6\}$ ($x_1, x_2 \in X_1, x_6 \in X_3$) 说明, 在所有决策表中, 利用 DT_2 里的数据模式 $\mathbf{a}(x_1)$ 、 $\mathbf{a}(x_2)$ 进行属性约简对提高第 1 类分类精度最好, 而利用 $\mathbf{a}(x_6)$ 对第 3 类分类最有利.

定义 5. 设 $(U, A_j \cup \{d\})$ 为相对优势决策表, 对给定的距离阈值 $h_j > 0$, Z_j 为其相对优势决策集, U_j 为其相对优势对象集, $D^{(j)} = (D_{pq}^{(j)})_{p,q=1,2,\dots,n}$ 为其距离矩阵. 称 $B \subseteq A_j$ (不唯一) 为关于 $D^{(j)}$ 的相对优势决策约简 (RDDR), 若 B 是满足下列条件的 A_j 的最小子集:

$$1) \forall x \in U_j, \theta_B^{h_j}(x) = \theta_{A_j}^{h_j}(x),$$

$$2) dis_B(x_p, x_q) \geq D_{pq}^{(j)}, x_p \in U_j \cap X_i, x_q \notin X_i, i \in Z_j.$$

根据属性约简的布尔推理方法, 对相对优势决策表 DT_j 和给定的距离矩阵 $D^{(j)}$, 其 RDDR 可由辨识公式

$$L_R^{(j)} = \bigwedge \{ \bigvee \{ \wedge E : E \in \mathcal{M}_{pq}^{(j)} : x_p \in U_j \cap X_i, x_q \notin X_i, i \in Z_j \} \} \quad (13)$$

($\mathcal{M}_{pq}^{(j)}$ 是 $C_{pq}^{(j)}$ 中满足 $dis_E(x_p, x_q) \geq D_{pq}^{(j)}$ 的全部最小非空子集 E 构成的集类) 确定, 记为 $B_R^{(j)} \in RED(D^{(j)})$.

记 $K_j \neq \emptyset$ 为相对优势决策表 DT_j 的优势决策集, $U_{K_j} \neq \emptyset$ 为其优势对象集, 则由定义 4、5 和定理 1 易证, 因 $\emptyset \neq Z_j \subseteq K_j$ 及 $\emptyset \neq U_j \subseteq U_{K_j}$, 故关于 $D^{(j)}$ 的 RDDR 也总是存在的. 显然, RDDR 的获取与所有相对优势决策表有关, 全部 RDDR 的获取过程称为 EAR.

若对 DT_1, \dots, DT_N 定义同样的距离函数, 并使其关于属性单调, 同时给定相同的距离阈值 h , 则根据选出的所有相对优势决策表可生成所谓的集成距离可分辨矩阵

$$C^e = [C_1^{(j_1)} \quad C_2^{(j_2)} \quad \dots \quad C_n^{(j_n)}]^T$$

$$C_q^{(j_q)} = [C_{1q}^{(j_q)} \quad C_{2q}^{(j_q)} \quad \dots \quad C_{nq}^{(j_q)}]^T, q \leq n \quad (14)$$

其中, $C_q^{(j_q)}$ 为 $C^{(j_q)}$ 的第 q 列, 决策表序号 j_q 应满足 $x_q \in U_{j_q}$ (U_{j_q} 为 DT_{j_q} 的相对优势对象集), 否则 $C_q^{(j_q)} = [\emptyset \dots \emptyset]_{n \times 1}^T$. 显然, $C_q^{(j_q)}$ 生成 C^e 的第 q 行, C^e 为 $n \times n$ 矩阵, 相应的集成距离矩阵为

$$D^e = [D_1^{(j_1)} \quad D_2^{(j_2)} \quad \dots \quad D_n^{(j_n)}]^T$$

$$D_q^{(j_q)} = [D_{1q}^{(j_q)} \quad D_{2q}^{(j_q)} \quad \dots \quad D_{nq}^{(j_q)}]^T, q \leq n$$

$$D_{pq}^{(j_p)} = dis_G(x_p, x_q), G = C_{pq}^{(j_p)} \quad (15)$$

建立 C^e 、 D^e 的目的在于更方便、清晰地进行 EAR 计算 (也可用 βD^e ($0 < \beta \leq 1$) 来整体控制 EAR 的质量, 其中 β 的取值应保证 D^e 中的最小非零元乘上 β 后仍大于 h). 运用 EAR 方法进行集成特征选择首先要根据训练样本建立有效的决策表. 对此, 可采用两种方法:

1) 统计法: 以训练样本子集 (可依某种距离度量聚类生成, 也可由专家选定) 标号构成论域; 各特征集组成相应的条件属性集, 属性/特征值应为某种统计参数值; 决策属性值为类别标号.

2) 直接法: 论域为训练样本标号集; 各特征集组成相应的条件属性集, 特征值就是属性值; 决策属性值为类别标号.

统计法适用于样本数较多且特征统计参数容易获取的情况; 直接法则适用于小样本或特征统计参数难以获得的场合. 无论哪种方法, 其成功应用的关键有两点: 1) 属性值映射的合理定义; 2) 对象距离函数 dis_B 的构造. 关于直接法, 其应用针对性较强, 特别是 dis_B 的构造尚需结合具体应用进一步研究. 本文根据课题需求和遥感图像数据的统计特性, 仅就统计法进行了实验, 相应的 EAR 步骤为:

Step 1. 对 c 类分类问题, 按统计法分别就 N 个数据源和特征集建立相应的决策表, 记为 DT_1, \dots, DT_N ;

Step 2. 对 DT_1, \dots, DT_N , 按式 (2) 和 (4) 定义相应的集值函数 $I_{A_1}^{h_1}, \dots, I_{A_N}^{h_N}$ 和对象距离函数 $dis_{A_1}, \dots, dis_{A_N}$, 并使其关于属性单调;

Step 3. 按式 (7) 计算 $r_i(A_j)$, $i = 1, \dots, c$, $j = 1, \dots, N$, 并按式 (8) 和 (9) 预选决策表 (决策表较少的场合可省略该步);

Step 4. 对所有预选出的决策表, 按式 (10) 对任意的 $j \in J$, $x_p \in U$ 计算 $F_j(x_p)$;

Step 5. 依式 (11) 和 (12) 计算 Z_j, U_j ($j \in J$) 以确定所有的相对优势决策表及其相对优势对象集;

Step 6. 按定义 2、3 计算 $C^{(j)}$ 、 $D^{(j)}$ ($Z_j \neq \emptyset$), 其中, $\forall x_p, x_q \in U, h < D_{pq}^{(j)} \leq dis_{A_j}(x_p, x_q) = dis_{C_{pq}^{(j)}}(x_p, x_q)$;

Step 7. 根据定义 5, 按式 (13) 计算各相对优势决策表的 RDDR, 并将其作为对应基分类器的输入. 若对各相对优势决策表定义同样的距离函数, 并给定相同的 h , 则可由式 (14) 和 (15) 得到相应的 C^e 、 D^e , 进而在 βD^e 的约束下统一计算各相对优势决策表的 RDDR.

3 基于 EAR 的多分类器融合实验

基于 EAR 的多分类器融合包括两部分: 1)

EAR 及相应基分类器选择/设计; 2) 关于 EAR 的多类模式可分性度量及其在决策级分类融合中的应用. 对于 1), 各基分类器选择/设计应由定义的距离函数决定, 记为 $CL_j (j \leq N)$, 则多分类器系统由所有 CL_j 组成; 对于 2), 可在多分类器融合中采用基于知识发现的 KD-DWV 算法^[10] 予以实现.

为检验 EAR 方法的有效性, 实验数据选用了一幅美国 Jasper Ridge 地区的高光谱图像 (400×400 , 60 个波段, 波长范围为 $441 \sim 1321 \text{ nm}$, AVIRIS 拍摄, 见图 2(a)). 根据地面资料^[11], 就三类植被及裸地选择了 10 个感兴趣区 (Region of interest, ROI) 进行监督分类, 其中 4 个留作测试区, 相应的像元数见表 1.

表 1 训练、测试数据
Table 1 Training and testing data

标号	训练像元集标签	训练像元数	测试像元数
常青林	1 x_1, x_2	$76(x_1) + 75(x_2)$	679
落叶林	2 x_3, x_4	$70(x_3) + 126(x_4)$	711
灌木丛	3 x_5	71	574
裸地	4 x_6	72	1308

首先, 数据预处理采用 ASD 方法^[12] 进行波段子空间分解, 得到三个波段子集, 相应的波段序号集分别为 $\{1, \dots, 19\}$, $\{20, \dots, 46\}$, $\{47, \dots, 60\}$. 进一步用 PCA 方法在三个波段集中进行特征提取, 分别得到三个主成分波段集 $A_1 = \{a_1, \dots, a_{10}\}$, $A_2 = \{b_1, \dots, b_{10}\}$, $A_3 = \{c_1, \dots, c_{10}\}$. 因高光谱图像中的植被等光谱可看成正态型数据, 故由 6 个训练像元集和 $A_1 \sim A_3$ 按前述 Step 1 建立决策表 DT_1 , DT_2 和 DT_3 . 其中 DT_3 如表 2 所示, 表中属性值 $c_l(x_p) = (\mu_p(l), \sigma_p(l))$, $\mu_p(l)$, $\sigma_p(l)$ 分别为 x_p 在第 l 主成分 c_l 上的均值和标准差. 由表 2 易得 $A(d) = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5\}, \{x_6\}\}$.

按 Step 2, 对任意 $B \subseteq A$, 采用 Bhattacharyya 距离的特定形式^[13] 定义对象距离函数

$$dis_B(x_p, x_q) = \sum_{\substack{a_l \in B \\ u_l(a_l(x_p), a_l(x_q)) > h}} u_l(a_l(x_p), a_l(x_q))$$

$$u_l(a_l(x_p), a_l(x_q)) = \frac{1}{4}(\mu_p(l) - \mu_q(l))^2(\sigma_p^2(l) + \sigma_q^2(l))^{-1} + \frac{1}{2} \ln\left(\frac{\sigma_p^2(l)}{2} + \frac{\sigma_q^2(l)}{2}\right)(\sigma_p^2(l)\sigma_q^2(l))^{-\frac{1}{2}}$$

显然, 对象距离函数关于属性单调. 取 $h = 1$, 计算 $I_{A_1}^1$, $I_{A_2}^1$ 和 $I_{A_3}^1$ 如表 3.

表 2 决策表 $DT_3 \sim (U, A_3 \cup \{d\})$

Table 2 Decision table $DT_3 \sim (U, A_3 \cup \{d\})$

U	c_1	c_2	\dots	c_{10}	d
x_1	(-227.5, 1096.6)	(204.3, 118.3)	\dots	(-2.7, 3.9)	1
x_2	(-782.1, 1143.4)	(178.3, 105.8)	\dots	(-1.8, 4.3)	1
x_3	(796.4, 652.6)	(586.0, 88.5)	\dots	(-6.6, 4.0)	2
x_4	(242.1, 1823.1)	(399.4, 165.5)	\dots	(-3.2, 4.2)	2
x_5	(-1091.9, 510.3)	(-161.2, 33.6)	\dots	(4.6, 3.9)	3
x_6	(2988.8, 533.85)	(-313.0, 53.9)	\dots	(-1.9, 4.4)	4

表 3 $I_{A_1}^1, I_{A_2}^1$ 和 $I_{A_3}^1$ 的映射集合值

Table 3 Set values of $I_{A_1}^1, I_{A_2}^1$ and $I_{A_3}^1$

U	$I_{A_1}^1$	$I_{A_2}^1$	$I_{A_3}^1$
x_1	$\{x_1\}$	$\{x_1\}$	$\{x_1, x_5, x_6\}$
x_2	$\{x_2\}$	$\{x_2, x_4\}$	$\{x_2, x_3\}$
x_3	$\{x_3, x_4, x_5, x_6\}$	$\{x_3\}$	$\{x_2, x_3, x_5\}$
x_4	$\{x_3, x_4, x_6\}$	$\{x_2, x_4\}$	$\{x_4\}$
x_5	$\{x_3, x_5, x_6\}$	$\{x_5, x_6\}$	$\{x_1, x_3, x_5, x_6\}$
x_6	$\{x_3, x_4, x_5, x_6\}$	$\{x_5, x_6\}$	$\{x_1, x_5, x_6\}$

按 Step 3 预选决策表. 由式 (7) 计算 $r_i(A_j)$ 如表 4, 由式 (8) 和式 (9) 计算得 $J_1 = \{3\}$, $J_2 = \{3\}$, $J_3 = \{1, 2, 3\}$, $J_4 = \{1, 2, 3\}$, 故 $J = \{1, 2, 3\}$, 三个决策表均可用.

按 Step 4 计算 $F_j(x_p)$ 如表 5.

表 4 $r_i(A_j)$

Table 4 $r_i(A_j)$

	A_1	A_2	A_3
$r_1(\cdot)$	0	0	1
$r_2(\cdot)$	0	0.5	1
$r_3(\cdot)$	1	1	1
$r_4(\cdot)$	1	1	1

表 5 $F_j(x_p)$

Table 5 $F_j(x_p)$

	x_1	x_2	x_3	x_4	x_5	x_6
$F_1(\cdot)$	0	0	0	0	1.36	12.49
$F_2(\cdot)$	0	0	2.42	0	4.96	8.6
$F_3(\cdot)$	1.15	1.43	2.79	1.15	2.53	4.45

按 Step 5 计算 Z_j, U_j 得: $Z_1 = \{4\}$, $Z_2 = \{3\}$, $Z_3 = \{1, 2\}$; $U_1 = \{x_6\}$, $U_2 = \{x_5\}$, $U_3 = \{x_1, x_2, x_3, x_4\}$, 则 DT_1 , DT_2 , DT_3 均为相对优势决策表, 且它们完全互补.

按 Step 6 计算 $C^{(j)}$ 分别为

$$C^{(1)} =$$

$$\begin{bmatrix} \emptyset & \emptyset & \emptyset & \emptyset & \{a_2\} & \{a_1, a_2, a_3\} \\ & \emptyset & \{a_6, a_7\} & \emptyset & \{a_1, a_2\} & \{a_1, a_2, a_3\} \\ & & \emptyset & \emptyset & \{a_1, a_2, a_6, a_7, a_8\} & \{a_1, a_2, a_3, a_7\} \\ & & & \emptyset & \{a_1, a_2\} & \{a_1, a_2, a_3\} \\ & & & & \emptyset & \{a_1, a_3, a_6\} \\ & & & & & \emptyset \end{bmatrix}$$

$$C^{(2)} = \begin{bmatrix} \emptyset & \emptyset & \{b_2\} & \emptyset & \{b_1, b_2, b_7\} & \{b_2\} \\ & \emptyset & \{b_1, b_2\} & \emptyset & \{b_2, b_7\} & \{b_2\} \\ & & \emptyset & \emptyset & \{b_1, b_2\} & \{b_1, b_2\} \\ & & & \emptyset & \{b_1, b_2\} & \{b_2\} \\ & & & & \emptyset & \{b_1, b_2\} \\ & & & & & \emptyset \end{bmatrix}$$

$$C^{(3)} = \begin{bmatrix} \emptyset & \emptyset & \{c_2, c_6\} & \{c_6\} & \{c_2\} & \{c_1, c_2\} \\ & \emptyset & \{c_2, c_6\} & \{c_6\} & \{c_2\} & \{c_1, c_2\} \\ & & \emptyset & \emptyset & \{c_1, c_2, c_3, c_6\} & \{c_1, c_2\} \\ & & & \emptyset & \{c_2, c_6\} & \{c_2\} \\ & & & & \emptyset & \{c_1, c_2\} \\ & & & & & \emptyset \end{bmatrix}$$

按 Step 7 计算相应的 C^e 、 D^e , 得

$$C^e = [C_1^{(3)} \ C_2^{(3)} \ C_3^{(3)} \ C_4^{(3)} \ C_5^{(2)} \ C_6^{(1)}]^T$$

$$C_1^{(3)} = [\emptyset \ \emptyset \ \{c_2, c_6\} \ \{c_6\} \ \{c_2\} \ \{c_1, c_2\}]^T$$

$$C_2^{(3)} = [\emptyset \ \emptyset \ \{c_2, c_6\} \ \{c_6\} \ \{c_2\} \ \{c_1, c_2\}]^T$$

$$C_3^{(3)} = [\{c_2, c_6\} \ \{c_2, c_6\} \ \emptyset \ \emptyset \ \{c_1, c_2, c_3, c_6\} \ \{c_1, c_2\}]^T$$

$$C_4^{(3)} = [\{c_6\} \ \{c_6\} \ \emptyset \ \emptyset \ \{c_2, c_6\} \ \{c_2\}]^T$$

$$C_5^{(2)} = [\{b_1, b_2, b_7\} \ \{b_2, b_7\} \ \{b_1, b_2\} \ \{b_1, b_2\} \ \emptyset \ \{b_1, b_2\}]^T$$

$$C_6^{(1)} = [\{a_1, a_2, a_3\} \ \{a_1, a_2, a_3\} \ \{a_1, a_2, a_3, a_7\} \ \{a_1, a_2, a_3\} \ \{a_1, a_3, a_6\} \ \emptyset]^T$$

$$D^e =$$

$$\begin{pmatrix} 0 & 0 & 2.79 & 1.15 & 2.53 & 5.96 \\ 0 & 0 & 3.61 & 1.43 & 2.62 & 6.75 \\ 2.79 & 3.61 & 0 & 0 & 19.29 & 20.59 \\ 1.15 & 1.43 & 0 & 0 & 4.33 & 4.45 \\ 5.19 & 5.84 & 21.84 & 4.96 & 0 & 8.60 \\ 12.49 & 17.94 & 26.02 & 14.16 & 12.99 & 0 \end{pmatrix}$$

以 C^e 、 D^e 为约束, 取 $\beta = 1$, 由式 (13) 得各决策表的 RDDR: $B_R^{(1)} = \{a_1, a_2, a_3, a_6, a_7\}$ 、 $B_R^{(2)} = \{b_1, b_2, b_7\}$ 、 $B_R^{(3)} = \{c_1, c_2, c_6\}$ (注: 此例中若再减小 β , 所得 RDDR 将与 PCA 方法结果相同, 即 $B_R^{(1)} = \{a_1, a_2, a_3\}$ 、 $B_R^{(2)} = \{b_1, b_2\}$ 、 $B_R^{(3)} = \{c_1, c_2\}$).

分类比较实验针对以下三种方法进行:

1) PCA 方法: 在 A_1 、 A_2 、 A_3 上分别按 PCA 方法提取与 $B_R^{(1)}$ 、 $B_R^{(2)}$ 、 $B_R^{(3)}$ 基数相同的主成分, 并组成特征集 $A = \{a_1, a_2, a_3, a_4, a_5, b_1, b_2, b_3, c_1, c_2, c_3\}$, 然后用 A 对图像做最大似然分类;

2) PCA_PV 方法: 分别用主成分集 $\{a_1, a_2, a_3, a_4, a_5\}$ 、 $\{b_1, b_2, b_3\}$ 和 $\{c_1, c_2, c_3\}$ 作为三个最大似然分类器的输入, 然后用经典的多数表决法 (Plurality voting, PV) 进行分类融合;

3) EAR_KD-DWV 方法: 根据定义的距离函数, 仍选用三个最大似然分类器, 分别用 $B_R^{(1)}$ 、 $B_R^{(2)}$ 和 $B_R^{(3)}$ 做相应输入. 对应三个特征子集的多类模式可分性度量分别由启发性融合函数 (其中系数 α 的取值参见文献 [10], 本例取 $\alpha = 0.8$) 转换为对三个分类器不同输出的加权值, 并用 KD-DWV 算法完成动态加权表决融合.

训练区和测试区的分类精度比较见图 1. 结果显示, 在训练区, PCA 方法和 PCA_PV 方法与基于 EAR 的 KD-DWV 方法的总体分类精度相当; 但在测试区, 后者的总体分类精度比前两者有显著提高, 相应的分类图如图 2(b)~2(d) (见下页) 所示.

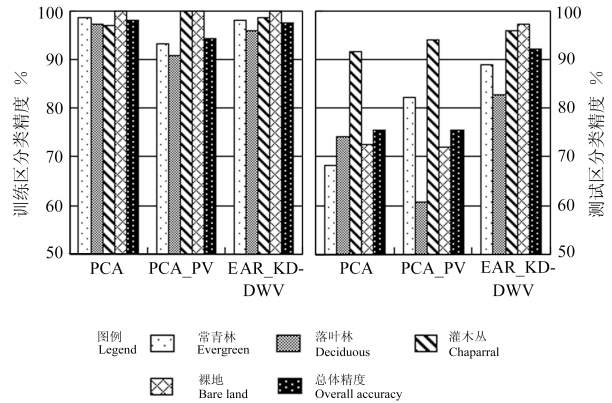


图 1 三种方法分类精度比较

Fig. 1 Classification accuracies for the three methods

4 结论

基于粗集属性约简的集成特征选择可方便分类知识的发现和利用, 因而对复杂和多类模式分类问题研究具有重要价值, 但其涉及不协调决策表和非等价关系下的属性约简. 本文在文献 [3, 7-8] 的基础上, 利用广义粗集模型研究了不协调决策表中关于容差关系的广义属性约简问题, 提出了 RDDR 概念和集成特征选择的 EAR 方法, 结合 KD-DWV 算法进行了高光谱遥感图像分类比较实验. 结果显示, 基于 EAR 的多分类器融合可有效提高分类的推广性.

由于问题的复杂性, 本文仅初步研究了 EAR 的 Filter 方法. 对复杂或小样本模式分类问题, 可进一步研究基于决策表构建“直接法”的 EAR 方法. 此

外, 基于 EAR 的分类知识发现、更新和管理也有很大的研究价值。

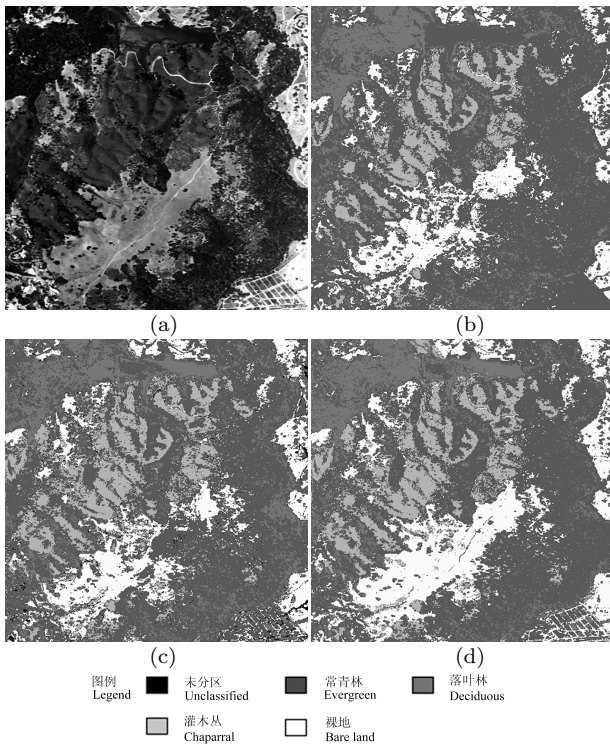


图 2 三种方法分类结果比较

((a) 原始图像; (b) PCA 方法分类图; (c) PCA_PV 方法分类图; (d) EAR_KD-DWV 方法分类图)

Fig. 2 Classification result comparison of the three methods

((a) Original image; (b) PCA method; (c) PCA_PV method; (d) EAR_KD-DWV method)

References

- Kittler J, Hatef M, Duin R P W, Matas J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(3): 226–239
- Tsybmal A, Pechenizkiy M, Cunningham P. Diversity in search strategies for ensemble feature selection. *Information Fusion*, 2005, **6**(1): 83–98
- Kryszkiewicz M. Comparative study of alternative types of knowledge reduction in inconsistent systems. *International Journal of Intelligent Systems*, 2001, **16**(1): 105–120
- Swiniarski R W, Skowron A. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*, 2003, **24**(6): 833–849
- Skowron A. Extracting laws from decision tables: a rough set approach. *Computational Intelligence*, 1995, **11**(2): 371–388
- Xiao Di, Hu Shou-Song. Real rough set theory and attribute reduction. *Acta Automatica Sinica*, 2007, **33**(3): 253–258 (肖迪, 胡寿松. 实域粗糙集理论与属性约简. *自动化学报*, 2007, **33**(3): 253–258)
- Zhang Wen-Xiu, Mi Ju-Sheng, Wu Wei-Zhi. Knowledge reductions in inconsistent information systems. *Chinese Journal of Computers*, 2003, **26**(1): 12–18 (张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简. *计算机学报*, 2003, **26**(1): 12–18)
- Sun Liang. Studies on Knowledge Discovery Methods for Fusion and Classification of High-dimensional Remote Sensing Data [Ph.D. dissertation], Xi'an Jiaotong University, 2006 (孙亮. 高维遥感数据融合与分类的知识发现方法研究 [博士学位论文], 西安交通大学, 2006)
- Polkowski L. *Rough Set: Mathematical Foundations*. Heidelberg: Physica-Verlag, 2002. 23–24
- Sun Liang, Han Chong-Zhao. Dynamic weighted voting for multiple classifier fusion: a generalized rough set method. *Journal of System Engineering and Electronics Technology (English Edition)*, 2006, **17**(3): 487–494
- ENVI tutorial: vegetation hyperspectral analysis [Online], available: http://www.rsinc.com/Envi/docs/tutorials/Vegetation_Analysis.pdf, May 30, 2007
- Zhang Y, Mita D, Zhang J P, Jin M. Adaptive subspace decomposition for hyperspectral data dimensionality reduction. In: *Proceedings of IEEE International Conference on Image Processing*. Kobe, Japan: IEEE, 1999. 326–329
- Richards J A, Jia X P. *Remote Sensing Digital Image Analysis: An Introduction (Third Edition)*. Berlin, Germany: Springer-Verlag, 1999. 331



孙亮 副教授. 主要研究方向为多源信息融合、智能模式识别与遥感图像处理、智能检测技术. 本文通信作者.

E-mail: sun_liang@people.com.cn

(SUN Liang Associate professor.

His research interest covers multisource information fusion, pattern recognition and remote sensing image processing,

and intelligent detection technology. Corresponding author of this paper.)



韩崇昭 教授. 主要研究方向为多源信息融合、随机控制与自适应控制、非线性频谱分析.

E-mail: nianming@gmail.com

(HAN Chong-Zhao Professor. His

research interest covers multisource information fusion, stochastic control and adaptive control, and nonlinear spectral

analysis.)



沈建京 教授. 主要研究方向为人工智能、计算智能、分布式计算和新一代 Web 技术. E-mail: shenjj@vip.sina.com

(SHEN Jian-Jing Professor. His

research interest covers artificial intelligence, computational intelligence, distributed computing, and new generation Web technology.)



戴宁 副教授. 主要研究方向为随机集和粗糙集理论、随机过程.

E-mail: dning@zzu.edu.cn

(DAI Ning Associate professor. Her

research interest covers random set and rough set theory, and stochastic processes.)