

一种改进的线性判别分析算法 MLDA

刘忠宝^{1,2} 王士同¹

(江南大学信息学院 无锡 214122)¹ (山西大学商务学院信息工程系 太原 030031)²

摘 要 线性判别分析(LDA)是模式识别方法之一,已广泛应用于模式识别、数据分析等诸多领域。线性判别分析法寻找的是有效分类的方向。而当样本维数远大于样本个数(即小样本问题)时,LDA 便束手无策。为有效解决线性判别分析法的小样本问题,提出了一种改进的 LDA 算法——MLDA。该算法将类内离散度矩阵进行标量化处理,有效地避免了对类内离散度矩阵求逆。通过实验证明 MLDA 在一定程度上解决了经典 LDA 的小样本问题。

关键词 特征提取,线性判别分析(LDA),小样本问题,类间离散度矩阵,类内离散度矩阵,标量化

Modified Linear Discriminant Analysis Method MLDA

LIU Zhong-bao^{1,2} WANG Shi-tong¹

(School of Information, Jiangnan University, Wuxi 214122, China)¹

(Department of Information Engineering, Business College of Shanxi University, Taiyuan 030031, China)²

Abstract Linear Discriminant Analysis (LDA) is one of methods in pattern recognition, and is widely used in many fields such as pattern recognition and data analysis. LDA is to find an effective classification direction. While the sample dimension is much larger than its quantity, it is hard for LDA to deal with this problem. In order to effectively solve small sample size problem in LDA, this paper presented a modified LDA algorithm MLDA. This new algorithm turns within-class scatter matrix into scalarization in order to avoid computing the inverse of within-class scatter matrix. A series of experiments verify MLDA solves the small sample size problem to some extent.

Keywords Feature extraction, Linear Discriminant Analysis (LDA), Small sample size problem, Between-class scatter matrix, Within-class scatter matrix, Scalarization

1 引言

在模式识别中,由被识别的对象产生一组基本特征,这些特征是计算出来的(当识别对象是波形或数字图像时),也可以用仪表或传感器测量出来(当识别对象是实物或某种过程时),这样产生出来的特征叫做原始特征。原始特征的数量可能很大,或者说样本处于一个高维空间中,通过映射(或变换)的方法可以用低维空间来表示样本,这个过程叫做特征提取^[1-4]。线性判别分析法(LDA)^[5-8]是一个用于特征提取的统计方法。LDA 以样本的可分性为目标寻找一组线性变换,使得同一类的样本聚集在一起,而不同类的样本相对比较分散^[9]。因此,在理论上 LDA 适合于模式识别问题。然而,对于高维多类模式的识别问题,传统的 LDA 方法最大的缺点是小样本问题,即难以找到足够多的训练样本以保证类内离散度矩阵可逆。为了解决小样本问题,近几年提出很多方法,如 Hong 的扰动法^[10]、Liu 的正交补空间法、Chen 的零空间法和 Yang 的 D2LDA 方法^[11]等。这些算法存在维数很高的问题,在原始维数空间中求解最优鉴别矢量集计算量太大,导致求解困难^[12];另一种传统的方案是主元分析法(PCA)+LDA

的两步策略^[13],利用主成分分析将高维空间的样本投影到低维空间以保证样本类内离散度矩阵是非奇异的。然而该方法的问题在于 PCA 在降维的同时也失去很多帮助判别的有效信息。

基于以上分析,本文提出一种改进的线性判别分析算法 MLDA,该算法在原始 Fisher 准则的基础上对样本类内离散度矩阵进行标量化处理,避免对类内离散度矩阵求逆,可以有效地解决上述小样本问题^[14]。

2 线性判别分析法

2.1 经典的 LDA

LDA 的目的是从高维特征空间里提取出最具有判别能力的低维特征。“这些特征能帮助将同一个类别的所有样本聚集在一起,不同类别的样本尽量分开”,即选择使得样本类间离散度和类内离散度的比值最大的特征^[15]。经典线性判别分析使用的是 Fisher 准则函数,所以线性判别分析又被称为 Fisher 线性判别分析^[16](Fisher LDA)。Fisher 准则函数的定义是:

$$J(w) = \arg \max \frac{w^T S_b w}{w^T S_w w} = [w_1, w_2, \dots, w_d] \quad (1)$$

到稿日期:2009-12-09 返修日期:2010-02-10 本文受国家 863 项目(2007AA1Z158, 2006AA10Z313),国家自然科学基金项目(60773206/F020106, 60704047/F030304),2006 年江苏省 6 大人才高峰计划项目,2008 江苏省研究生创新计划课题资助。

刘忠宝(1981—),男,博士生,主要研究方向为人工智能与模式识别, E-mail: liu_zhongbao@hotmail.com; 王士同(1964—),男,教授,博士生导师,主要研究方向为人工智能、机器学习。

式中, S_b, S_w 分别为样本的类间离散度矩阵和类内离散度矩阵, 它们的定义如下:

样本类间离散度矩阵:

$$S_b = \sum_{i=1}^c \frac{N_i}{N} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (2)$$

样本类内离散度矩阵:

$$S_w = \sum_{i=1}^c \sum_{j=1}^{N_i} \frac{1}{N} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T \quad (3)$$

式中, N : 所有样本数; c : 样本类别数; $N_i (i=1, 2, \dots, c)$: 第 i 类样本数; $x_{ij} (i=1, 2, \dots, c; j=1, 2, \dots, N_i)$: 第 i 类的第 j 个样本; $\bar{x}$: 所有样本均值; $\bar{x}_i (i=1, 2, \dots, c)$: 第 i 类样本均值。

式(1)中 $J(w)$ 是广义的 Rayleigh 熵, 可用 Lagrange 乘子法求解。令分母等于非零常数, 即令 $w^T S_w w = C \neq 0$, 定义 Lagrange 函数为:

$$L(w, \lambda) = w^T S_b w - \lambda(w^T S_w w - C) \quad (4)$$

式中, λ 为 Lagrange 乘子。将式(4)对 w 求偏导得:

$$\frac{\partial L(w, \lambda)}{\partial w} = S_b w - \lambda S_w w \quad (5)$$

令式(5)偏导数为零, 得 $S_b w^* - \lambda S_w w^* = 0$, 即:

$$S_b w^* = \lambda S_w w^* \quad (6)$$

式中, w^* 为 $J(w)$ 取得最大值时的 w 。当 S_w 非奇异时^[17], 式(6)两边乘以 S_w^{-1} 得:

$$S_w^{-1} S_b w^* = \lambda w^* \quad (7)$$

求解式(7)等价于求解一般矩阵 $S_w^{-1} S_b$ 的特征值问题。

综上所述, 当 S_w 非奇异时, 在数学上求解函数(1)就等同于求解 $S_w^{-1} S_b$ 的特征值问题, 使 $J(w)$ 最大的变换矩阵 w 由 $S_w^{-1} S_b$ 的前 m 个最大特征值所对应的特征向量组成。

但对于人脸识别应用来说, 由于通常没有足够多的训练样本保证类内离散度矩阵 S_w 满秩, 此时产生小样本问题(即样本的维数很高, 训练样本很少)。小样本问题致使准则函数无法直接求解。

2.2 改进的线性判别分析法 MLDA

为解决经典 LDA 遇到的小样本问题, 本文提出了 LDA 的改进算法——MLDA。该算法在原始 Fisher 准则的基础上对类内离散度矩阵进行标量化处理, 避免对类内离散度矩阵求逆, 保证 S_w 在任何情况下都存在; 此外, 该算法将标量

$$w_k = \left\{ \frac{\sum_{i=1}^c N_i (\bar{x}_i^k - \bar{x}^k)^2}{\left[\sum_{k=1}^d \left(\frac{\sum_{i=1}^c N_i (\bar{x}_i^k - \bar{x}^k)^2}{\sum_{i=1}^c \sum_{j=1}^{N_i} (x_{ij}^k - \bar{x}_i^k)^2} \right)^{\frac{\alpha-2}{\alpha}} \right]^{\frac{\alpha-2}{\alpha}} \left[\sum_{i=1}^c \sum_{j=1}^{N_i} (x_{ij}^k - \bar{x}_i^k)^2 \right]^{\frac{1}{\alpha}}} \right\}^{\frac{1}{\alpha-2}} \quad (17)$$

上式等号右侧各变量均已知, 这样就得到样本空间各维的权重, 即各维在样本空间的重要程度。这里需要说明的是经典 LDA 及其一些改进算法往往以矢量为基本单位处理各对象, 通过求解相关矩阵的特征值和特征向量得到结果。而 MLDA 的出发点是训练样本各维重要程度各不相同, 因此赋予各维相应的权值, 将前若干个最大权值对应向量组成的矩阵作为降维后的结果。MLDA 彻底改变了以矢量为基本单位的处理方式, 把标量作为基本单位对问题进行求解, 不仅降低了运算难度, 提高了运算效率, 更为 LDA 的进一步研究开拓了新的思路。

3 实验结果及分析

为了验证 MLDA 的有效性, 本节从大样本情况和小样本情况两方面对算法进行了测试。在大样本情况下, 检验 ML-

作为基本单位处理各对象, 有效地降低了运算量, 提高了运算效率。

样本的类间离散度矩阵和类内离散度矩阵分别标量化处理后, 可得如下表达式:

$$S_b = \sum_{i=1}^c \sum_{k=1}^d \frac{N_i}{N} (\bar{x}_i^k - \bar{x}^k)^2 \quad (8)$$

$$S_w = \sum_{i=1}^c \sum_{j=1}^{N_i} \sum_{k=1}^d \frac{1}{N} (x_{ij}^k - \bar{x}_i^k)^\alpha (\alpha > 0 \text{ 且 } \alpha \neq 2) \quad (9)$$

式中, k 表示样本空间的维度。

样本空间各维加权后, 式(8)、式(9)分别变为如下表达式:

$$w S_b = \sum_{i=1}^c \sum_{k=1}^d w_k^2 (\bar{x}_i^k - \bar{x}^k)^2 \quad (10)$$

$$w S_w = \sum_{i=1}^c \sum_{j=1}^{N_i} \sum_{k=1}^d \frac{1}{N} w_k^\alpha (x_{ij}^k - \bar{x}_i^k)^\alpha (\alpha > 0 \text{ 且 } \alpha \neq 2) \quad (11)$$

式中, $w_k (k=1, 2, \dots, d)$ 为样本空间各维的权重。MLDA 的主要工作就是求 w_k 。

借鉴 Fisher 准则思想, 本文要找的 w_k 满足式(10)与式(11)之比最大。即:

$$y = \operatorname{argmax}_w \frac{w S_b}{w S_w} \quad (12)$$

式(12)等价于 $y = \max(w S_b)$ 且 $w S_w = 1$, 由 Lagrange 乘子法得:

$$y = w S_b + \lambda(1 - w S_w) \quad (13)$$

将式(13)对 w 的每个维度求偏导, 并令:

$$\frac{\partial y}{\partial w_k} = 0 \quad (14)$$

可求得:

$$w_k = \left[\frac{2 \sum_{i=1}^c N_i (\bar{x}_i^k - \bar{x}^k)^2}{\lambda \alpha \sum_{i=1}^c \sum_{j=1}^{N_i} (x_{ij}^k - \bar{x}_i^k)^\alpha} \right]^{\frac{1}{\alpha-2}} \quad (15)$$

由于 $w S_w = 1$, 由式(11)得:

$$\lambda = \left(\frac{1}{N} \right)^{\frac{\alpha-2}{\alpha}} \left[\sum_{k=1}^d \left(\frac{2 \sum_{i=1}^c N_i (\bar{x}_i^k - \bar{x}^k)^2}{\alpha \sum_{i=1}^c \sum_{j=1}^{N_i} (x_{ij}^k - \bar{x}_i^k)^2} \right)^{\frac{\alpha-2}{\alpha}} \right]^{\frac{\alpha-2}{\alpha}} \quad (16)$$

将式(16)带入式(15), 并去掉常数项得:

DA 在特征识别方面的有效性; 在小样本情况情况下, 在 ORL 人脸数据库上分别考察人脸识别率与训练样本数、特征维数、参数 α 的关系, 并通过实验得到识别率达到理想值的条件。

MLDA 基本流程是: 首先将人脸数据库根据一定规则分为训练样本和测试样本; 然后用 MLDA 对训练样本进行投影后, 得到最佳投影方向; 接着将测试样本投影到该最佳投影方向上, 得到投影后的测试样本; 最后将投影后的测试样本通过最近邻分类器与训练样本进行特征识别, 得到识别结果。算法流程如图 1 所示。

实验一 在大样本情况下, 验证 MLDA 在特征识别方面的有效性。

为了直观地考察 MLDA 在大样本情况下的有效性, 人为构造了一些数据, 然后通过 MLDA 对这些数据进行处理, 处理后的结果为算法发现的特征方向。通过数据本身的特点和

算法的结果相对比,发现 MLDA 可以很好地发现数据的本质特征。图 2 较长的线段表示 MLDA 发现的主特征,它忽略了右上角的数据点。

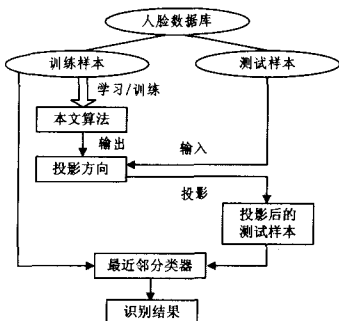


图1 MLDA流程图

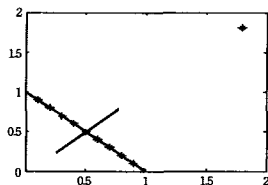


图2 MLDA特征发现示意图

通过该实验可以看出,MLDA 能发现数据的局部结构,因而找到了更适合的特征方向。

在大多数情况下,MLDA 在识别数据的本质特征方面也有良好的效果。如图 3 中数据的本质特征是两个几何圆形,用 MLDA 发现的主特征在竖直方向(投影点用“+”表示),这个方向的数据投影是完全分开的,可以被很好地识别。实验结果表明,MLDA 能识别数据的本质特征,即映射到低维空间后可以把原来分离的两个圆分开。

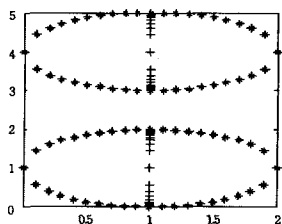


图3 MLDA特征识别示意图

下面的实验是在三维空间中人为构造两类数据集,每个数据集中数据分布均呈圆形,如图 4 所示。将数据降维后映射到二维空间,这种实验也叫 2D 可视化实验。

用 MLDA 对上述数据集进行降维后,得到图 5。不难发现两类数据被明显分开,显然 MLDA 能将不同的数据特征很好地区分开来。

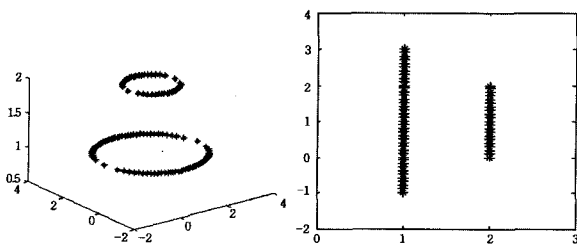


图4 人为构造的两类三维空间数据集

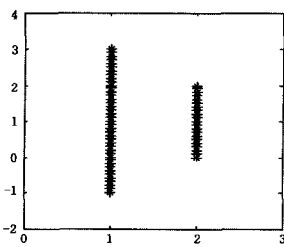


图5 三维空间数据集降维效果示意图

实验二 在小样本情况下,验证 MLDA 的有效性。

本文在 Olivetti-Oracle Research Lab (ORL)^[18] 人脸数据库上进行实验。ORL 人脸数据库是由 40 个人每人 10 幅共 400 幅图像组成,图像是在黑色背景下拍摄,光照条件基本不变,但头像的大小($\pm 10\%$)、视角(转动、倾斜最大 $\pm 20^\circ$)、表情(眼睛睁/闭、笑/非笑)和饰物(戴/不戴眼镜)有变化,原始图像大小为 112×92 像素,256 灰度级。

本文关注的是特征提取方法,因此在特征提取之后采用

最近邻分类器^[19]。最近邻分类器根据某种距离准则来比较未知测试样本与已知样本之间的距离,决策测试样本与离它最近的样本同类。本实验采用的是欧式距离分类器。

首先测试不同数量的训练样本对 MLDA 识别率的影响。本实验将人脸库中每个人的前 k 幅图像作为训练样本,剩下的图像作为测试样本。对于 ORL 人脸库,MLDA 保留特征向量个数为 39。本实验 k 分别取 3, 4, 5, 6, 7。图 6 表明了识别率与训练样本数量的关系。

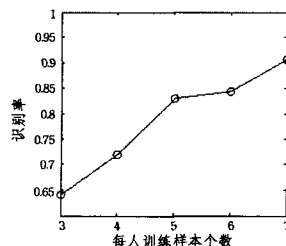


图6 识别率与训练样本数量的关系

由图 6 可以看出:随着训练样本数量的增加,识别率不断提高。在大多情况下,识别率在可接受的范围内。

下面的实验测试特征维数对识别率的影响。选取每人前 5 幅图像作为训练样本,剩下的 5 幅图像作为测试样本。图 7 表明了识别率与特征维数的关系。

由图 7 可以看出:特征维数达到 15 以后,MLDA 的识别率一直保持在较高的范围内;当特征维数为 35 时,达到最高识别率 83%。

MLDA 引入了参数 α , 参数 α 的变化引起识别率的变化。通过大量实验,得到的实验结果如图 8 所示。

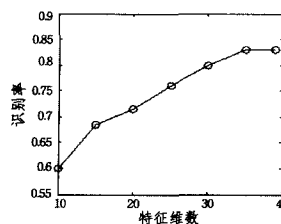


图7 识别率与特征维数的关系

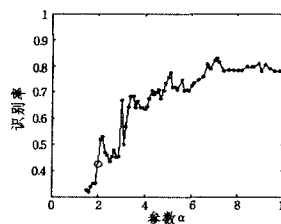


图8 识别率与参数 α 的关系

由图 8 可以看出:参数 α 对算法性能的影响较大。当 $\alpha = 2$ 时,识别率无法求解;当 $\alpha < 2$ 时,识别率不高;当 $\alpha > 2$ 时,识别率随着 α 的增大而呈现上升趋势;当 $\alpha = 7.01$ 时,MLDA 达到最大的识别率 83%。

特别地, $\alpha = 7.1$ 、特征维数在 32~39 之间变化时,识别率的变化情况如表 1 所列。

表 1 $\alpha = 7.1$ 、特征维数在 32~39 之间变化时识别率变化情况

特征维数	32	33	34	35	36	37	38	39
识别率	0.8	0.8	0.805	0.83	0.825	0.82	0.825	0.83

结束语 本文在原始 Fisher 准则的基础上,提出了一种改进的 LDA 算法——MLDA。该算法使最佳映射方向的确定不受类内离散度矩阵奇异的制约,从而有效地解决了在高维多类模式下特征提取中的小样本问题。MLDA 避免了对类内离散度矩阵求逆,同时也降低了运算复杂度,实现了有效、稳定的特征提取。实验结果证明了 MLDA 的有效性。然而 MLDA 也存在一些不足,如在识别率方面与 PCA+LDA 方法相比要低一些,算法性能受参数 α 的影响较大等。但 MLDA 对 LDA 的进一步研究开拓了新的思路。

参考文献

- [1] 严超, 苏光大. 人脸特征的定位与提取[J]. 中国图象图形学报, 1998, 3(5): 375-379
- [2] 张宏林, 蔡锐. Visual C++ 数字图像模式识别技术及工程实践[M]. 北京: 人民邮电出版社, 2003
- [3] 彭辉, 张长水. 基于 K-L 变换的人脸自动识别方法[J]. 清华大学学报: 自然科学版, 1997, 37(3): 67-700
- [4] 杨键, 杨静宇, 叶晖. Fisher 线性鉴别分析的理论研究及其应用[J]. 自动化学报, 2003, 29(4): 482-493
- [5] 孙伯成, 张文超. 基于部件的级联线性判别分析人脸识别[J]. 计算机工程与应用, 2006, 42(16): 67-69
- [6] 石跃祥, 蔡自兴, 王学武. 基于改进的 PCA 算法和 Fisher 线性判别的人脸识别技术[J]. 小型微型计算机系统, 2006, 27(9): 1731-1736
- [7] 郑宇杰, 於东军, 杨静宇. 一种基于 ICA 和 LDA 组合的人脸识别新方法[J]. 计算机科学, 2006, 33(4): 194-197
- [8] Zhang Xiaoxun, Jia Yunde. A linear discriminant analysis framework based on random subspace for face recognition[J]. Pattern Recognition, 2007, 40(9): 2585-2591
- [9] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720
- [10] Hong Z Q, Yang J Y, et al. Optimal discriminant plane for a

- small number of samoles and design method of classifier on the plane[J]. Pattern Recognition, 1991, 24(4): 317-324
- [11] Yu H, Yang J. A direct LDS algorithm for high dimensional data with application to face recognition[J]. Pattern Recognition, 2001, 34: 2067-2070
- [12] Etemad K, Chellappa R. Discriminant analysis for recognition of human face images[J]. Journal of the Optical Society of America A: Optics Image Science and Vision, 1997, 14(8): 1724-173
- [13] Belhumeur P, Hespanha J, Kriegman D. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[J]. IEEE Trans Patt Anal Mach Intel, 1997, 19(7): 711-720
- [14] Chen Li-fen, Mark Liao Hong-yuan, Ko Ming-kat, et al. A new LDA-based face recognition system which can solve the small samples size problem[J]. Journal of Pattern Recognition, 2000, 33(10): 1713-1726
- [15] 韩争胜, 李映, 张艳宁. 基于 LDA 算法的人脸识别方法的比较研究[J]. 微电子学与计算机, 2005, 22(7): 131-138
- [16] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000
- [17] 郭跃飞, 黄修武, 杨静宇. 一种求解 Fisher 最佳鉴别矢量的新方法及其人脸识别[J]. 中国图象图形学报, 1999, 4(2): 95-98
- [18] AT&T Laboratories Cambridge. The ORL Database of Faces [OL]. <http://www.cam-orl.co.uk/facedatabase.html>
- [19] 李刚, 高政. 人脸识别理论研究进展[J]. 计算机与现代化, 2003, (5): 1-6

(上接第 233 页)

信息粒减小, 使得粗糙熵随之单调减少。

推论 1 设 (U, A, F) 为集值信息系统, $B, C \subseteq A$. 若 $B =_< C$, 则 $E_r(B) = E_r(C)$ 。

但定理 5 和推论 1 的逆命题不一定成立。

由定理 2 及定理 5 易得:

推论 2 设 (U, A, F) 为集值信息系统. 则对任意的 $B \subseteq A$, 有 $E_r(A) \leq E_r(B)$ 。

定理 6 设 (U, A, F) 为集值信息系统. 则对任意的 $B, C \subseteq A$, 若 $E_r(B) = E_r(C)$ 且 $B \leq C$, 则 $B =_< C$ 。

证明: 由 $B \leq C$ 知对任意的 $x_i \in U$, $R_B^-(x_i) \subseteq R_C^-(x_i)$ 。假设 $B =_< C$ 不成立, 则 $R_B^- \neq R_C^-$ 。于是存在 $x_0 \in U$ 使得 $R_B^-(x_0) \subset R_C^-(x_0)$, 所以

$$\begin{aligned} E_r(B) &= - \sum_{i=1}^{|U|} \frac{|R_B^-(x_i)|}{|U|} \log_2 \frac{1}{|R_B^-(x_i)|} \\ &= \sum_{i=1}^{|U|} \frac{|R_B^-(x_i)|}{|U|} \log_2 |R_B^-(x_i)| \\ &= \sum_{x_i \neq x_{i_0}} \frac{|R_B^-(x_i)|}{|U|} \log_2 |R_B^-(x_i)| + \frac{|R_B^-(x_{i_0})|}{|U|} \log_2 |R_B^-(x_{i_0})| \\ &\leq \sum_{x_i \neq x_{i_0}} \frac{|R_C^-(x_i)|}{|U|} \log_2 |R_C^-(x_i)| + \frac{|R_B^-(x_{i_0})|}{|U|} \log_2 |R_B^-(x_{i_0})| \\ &< \sum_{x_i \neq x_{i_0}} \frac{|R_C^-(x_i)|}{|U|} \log_2 |R_C^-(x_i)| + \frac{|R_C^-(x_{i_0})|}{|U|} \log_2 |R_C^-(x_{i_0})| \\ &= \sum_{i=1}^{|U|} \frac{|R_C^-(x_i)|}{|U|} \log_2 |R_C^-(x_i)| = E_r(C) \end{aligned}$$

这与 $E_r(B) = E_r(C)$ 矛盾, 所以 $B =_< C$ 。

由定理 6 的证明及定理 5 可得:

推论 3 设 (U, A, F) 为集值信息系统, $B, C \subseteq A$, 若 $B < C$, 则有 $E_r(B) < E_r(C)$ 。

结束语 熵理论是信息系统中不确定性研究的有效工具之一。本文在集值信息系统中定义拟序关系, 在此基础上给出了覆盖集的性质, 引入了粗糙熵的概念, 研究了粗糙熵的极值性, 证明了粗糙熵随着信息粒的减小而单调减小。

参考文献

- [1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11: 341-356
- [2] Pawlak Z, Grzymala-Busse J W, Slowinski R, et al. Rough sets[J]. Communication of the ACM, 1995, 38(11): 89-95
- [3] Pawlak Z. Rough set theory and its application to data analysis[J]. Cybernetics and Systems, 1998, 9: 661-668
- [4] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001
- [5] 张文修, 梁怡, 吴伟志. 信息系统与知识发现[M]. 北京: 科学出版社, 2003
- [6] Shannon C E. The mathematical theory of communication [J]. The Bell System Technical Journal, 1948, 27(3/4): 373-423
- [7] 苗夺谦, 王珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报, 1999, 10(2): 113-116
- [8] 苗夺谦, 王珏. 粗糙集理论中知识熵与信息熵关系得讨论[J]. 模式识别与人工智能, 1998, 11(3): 34-40
- [9] Wang G Y. Algebra view and information view of rough sets theory[C]// Conference on Data Mining and knowledge Discovery: Theory, Tools, and Technology. Orlando, 2001, 4384: 200-207
- [10] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766
- [11] Liang J Y, Shi Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2004, 12(1): 37-46
- [12] 梁吉业, 李德玉. 信息系统中的不确定性与知识获取[M]. 北京: 科学出版社, 2004
- [13] 梁吉业, 钱宇华. 信息系统中的信息粒与熵理论[J]. 中国科学 E 辑: 信息科学, 2008, 38(12): 2048-2065