

可拓工程方法在储粮害虫分类识别中的应用研究

袁金丽¹, 吉海彦¹, 郭志涛²

(1. 中国农业大学信息与电气工程学院, 北京 100094; 2. 河北工业大学信息工程学院, 天津 300130)

摘要: 利用图像识别储粮害虫过程中, 需解决多种害虫的多特征参数和混和度大的分类识别问题。该文提出应用可拓工程方法, 构造储粮害虫特征数据的标准物元矩阵与节域物元矩阵, 计算待测粮虫与各类害虫的关联度, 以关联度的大小对储粮害虫进行分类判别, 程序计算结果表明, 该方法可行有效。可拓工程方法应用在储粮害虫分类识别上有应用价值。

关键词: 储粮害虫检测; 可拓工程; 可拓集合; 分类识别

中图分类号: O 29; S379.5

文献标识码: A

文章编号: 1002-6819(2004)05-0170-03

0 引言

及时、准确地检测识别储粮害虫, 才可对其进行有效防治, 邱道尹等人在储粮害虫检测和识别研究中开发了基于图像识别的粮仓害虫检测系统^[1-3]。利用图像处理模式和识别技术实现了害虫的特征提取, 并利用特征数据进行分类。在分类过程中他们先后应用了模糊分类算法^[1]、最近邻分类器^[2]和神经网络算法^[3]。对储粮害虫识别这样一个多参数、混合度大的分类问题, 应用前两种分类方法, 很难获得较高的识别率; 运用BP神经网络算法, 在害虫种类较少的情况下, 可以取得较为满意的效果, 而当害虫种类、特征参数增多时, 其误识率会升高, 且神经网络学习训练的时间会明显加长, 不利于在线检测的实现。可拓工程方法是将可拓方法应用于知识工程、模式识别、诊断、决策等领域的统称。该方法以物元与可拓集合理论为基础^[4-9], 适用于多参数、混合度大的综合分类问题^[10], 且在参数增多时, 计算量也不大。为此本文提出将可拓工程方法应用于储粮害虫的分类, 并在测取的数据基础上, 进行了应用研究与计算, 结果表明用这种方法识别储粮害虫的正确率较高, 且快速简便, 有相当的实用性。

1 可拓工程方法用于分类识别的理论基础

可拓识别研究的是利用物元理论对大系统的关键因素进行分析, 利用关联函数理论建立可拓相关矩阵, 并进行可拓聚类分析, 在此基础上建立可拓识别的物元模型和方法。多特征参数事物的分类识别是一个较为复杂的问题, 利用可拓识别的思想, 建立事物多特征参数的识别模型并以定量的数值表示评定结果, 能较完整地反映事物的实际类别, 且易于计算机进行规范化识别。

由物元与可拓集合理论, 若集合 P_0 是 P_e 的一个子集, 对于集合 P_e 中的任意元素 P_x , 判断 P_x 是否属于 P_0 , 并计算 P_x 属于 P_0 的程度, 其方法如下^[11]。

1.1 构造经典物元矩阵和节域物元矩阵

1) 由集合 P_0 的特征及其标准量值范围组成的物元矩阵称为经典物元矩阵, 记为

$$R_0 = \begin{bmatrix} P_0 & A_1 & [a_{01}, b_{01}] \\ & A_2 & [a_{02}, b_{02}] \\ & \cdots & \cdots \\ & A_n & [a_{0n}, b_{0n}] \end{bmatrix} \quad (1)$$

式中 A_1, A_2, \dots, A_n —— P_0 的 n 个不同特征; a_{0i}, b_{0i} —— 分别表示经典物元特征 A_i 量值的上、下限值。

2) 由经典物元加上可以转化为经典物元的事物及其特征, 再加上此特征相应拓广了的量值范围组成的物元矩阵称为节域物元矩阵, 记为

$$R_e = \begin{bmatrix} P_e & A_1 & [a_{e1}, b_{e1}] \\ & A_2 & [a_{e2}, b_{e2}] \\ & \cdots & \cdots \\ & A_n & [a_{en}, b_{en}] \end{bmatrix} \quad (2)$$

式中 a_{ei}, b_{ei} —— 分别表示节域物元特征 A_i 量值的上、下限值。显然, $P_0 \subset P_e$ 。

3) 待识别对象的物元矩阵, 记为

$$R_x = \begin{bmatrix} P_x & A_1 & X_1 \\ & A_2 & X_2 \\ & \cdots & \cdots \\ & A_n & X_n \end{bmatrix} \quad (3)$$

式中 X_i —— 待识别物元特征 A_i 的量值。

1.2 识别对象的综合关联度判别准则

物元各特征的量值为实数时, 判断某事物 P_x 符合集合 P_0 的程度, 由综合关联度 $K(P_x)$ 表示, $K(P_x)$ 由式(4)、(5)计算得出。

$$K_i(x_i) = \begin{cases} \frac{\rho(x_i, x_{0i})}{|x_{0i}|} & \text{当 } x = x_{0i} \\ \frac{\rho(x_i, x_{ei}) - \rho(x_1, x_{0i})}{\rho(x_i, x_{ei}) - \rho(x_1, x_{0i})} & \text{当 } x_i \neq x_{0i} \end{cases} \quad (4)$$

式中 $|x_{0i}| = |b_{0i} - a_{0i}|$ 为有界区间 $x_{0i} = (a_{0i}, b_{0i})$ 的模; $\rho(x_i, x) = |x - (a + b)/2| - (b - a)/2$ 为点 x_i 与区间 $x = (a, b)$ 上的距。

$$K(P_r) = \frac{1}{n} \sum_{i=1}^n \lambda K_i(x_i) \quad (5)$$

式中 λ —— 各特征的权系数, 表示各特征参数在识别中所起作用的重要程度。其数值一般由多次实验确定。

收稿日期: 2003-09-20 修订日期: 2004-08-10

作者简介: 袁金丽(1978-), 女, 河北故城县人, 主要从事计算机视觉技术在农产品无损检测中应用的研究。北京市海淀区圆明园西路2号 中国农业大学信息与电气工程学院, 100094。

Email: forestbraver@sohu.com

通信作者: 吉海彦, 男, 教授, 中国农业大学信息与电气工程学院, 100094。Email: instru@cau.edu.cn

待识别对象与某集合的综合关联度越大, 说明它符合该类集合要求的程度越高, 即它与该类集合中的事物愈接近, 故应判属该类。具体判别准则^[10] 如下:

当 $K(P_x) = 0$ 时, 判定 $P_x = P_0$; 且其数值愈大, 表明 P_x 符合 P_0 程度愈高, 即 P_x 属于 P_0 中事物的条件愈具备。

当 $-1 < K(P_x) < 0$ 时, 判定 $P_x = P_e$, 但 $P_x \neq P_0$; 且其绝对值愈小, 表明 P_x 可拓为 P_0 中事物的条件愈具备。

当 $K(P_x) = -1$ 时, 判定 $P_x \neq P_e$ 和 $P_x \neq P_0$; 且其绝对值愈大, 表明 P_x 符合 P_0 的条件愈差。

2 储粮害虫检测中可拓工程方法的应用

2.1 储粮害虫特征参数的定义

小麦粮仓中的主要害虫有米象类、赤拟谷盗、杂拟谷盗、扁谷盗、长头谷盗、锯谷盗、谷蠹。邱道尹等^[1] 对粮虫样本的二值图像进行特征提取, 本文依据他提取的面积、周长、复杂度三个特征的数据对害虫进行分类。这三个特征的定义如下。

1) 粮虫的面积: 即粮虫样本图像中粮虫所包含的像素点的个数, 设粮虫目标像素的灰度值为“0”, 则对于 $M \times N$ 大小的数字图像, 粮虫面积定义为

$$S = \sum_{i=1}^M \sum_{j=1}^N (1 - F(i, j)) \tag{6}$$

式中 M, N —— 分别表示以像素为单位的图像的长和宽; $F(i, j)$ —— 在点 (i, j) 处的二值化图像的灰度级 (只能取两个值: “0”或“1”);

2) 粮虫的周长: 即图像中粮虫区域边界像素的个数;

3) 粮虫的复杂度: 即图像中粮虫目标周长 L 的平方除以面积, 即复杂度。

$$F = L^2 / (K \times S) \tag{7}$$

式中 K —— 常数。

2.2 储粮害虫特征数据

邱道尹等人经实验提取的储粮害虫特征数据, 如表 1、表 2 所示。

表 1 七种粮虫图像特征的均值和方差

Table 1 Mean values and variance analysis of image characteristics of seven grain pests

虫名	均 值			方 差		
	面积/像素	周长/像素	复杂度	面积/像素	周长/像素	复杂度
赤拟谷盗	3268.89	410.70	4.15	237.54	23.59	0.23
杂拟谷盗	3350.40	397.60	3.79	140.64	12.36	0.23
米象类	2428.40	451.20	6.70	148.89	23.48	0.38
长头谷盗	2040.11	323.44	4.14	76.64	13.84	0.30
锯谷盗	1520.80	264.40	3.68	73.34	17.40	0.34
扁谷盗	902.00	193.38	3.37	25.41	11.95	0.36
谷 蠹	1693.70	205.60	2.03	36.35	17.40	0.18

注: 数据来源于参考文献[1]。

表 2 待识别粮虫图像样本的特征参数值

Table 2 Characteristic parameters of image samples of seven grain pests to be recognized

样本序号	面积/像素	周长/像素	复杂度
1	2715.00	363.00	3.86
2	2319.00	435.00	6.49
3	2638.00	372.00	4.17
4	3424.00	401.00	3.74
5	863.00	158.00	2.30
6	1538.00	195.00	1.97
7	3835.00	383.00	3.04

注: 数据来源于参考文献[1]。

2.3 构造 7 种储粮害虫特征的标准物元矩阵和节域物元矩阵

依据表 1 计算可得 7 种害虫特征值的变化区间 (表 3)。害虫特征的标准量值范围是特征值变化区间的最小值到特征平均值之间的范围, 害虫特征拓广后的量值范围是从零到特征值变化区间最大值之间的范围。

表 3 储粮害虫各特征值的变化区间

Table 3 Variation ranges of the grain pests' characteristic values

害虫种类	特征值变化范围		
	面积/像素	周长/像素	复杂度
赤拟谷盗	(3031.35, 3506.43)	(387.11, 434.29)	(3.92, 4.38)
杂拟谷盗	(3209.76, 3491.04)	(385.24, 409.96)	(3.56, 4.02)
米象类	(2279.51, 2577.29)	(427.72, 474.68)	(6.32, 7.08)
长头谷盗	(1963.65, 2116.75)	(309.6, 337.28)	(3.84, 4.44)
锯谷盗	(1447.46, 1594.14)	(247.28, 281.8)	(3.34, 4.02)
扁谷盗	(876.59, 927.41)	(181.43, 204.33)	(3.01, 3.73)
谷 蠹	(1657.4, 1730.1)	(195.99, 215.21)	(1.85, 2.21)

注: 数据由表 1 计算得来。

由各特征的标准量值范围和拓广后的量值范围构造害虫的标准物元矩阵和节域物元矩阵如下 (仅以赤拟谷盗为例, 其它 6 种依此类推)

标准物元矩阵 R_0

$$R_0 = \begin{pmatrix} \text{赤拟谷盗} & \text{面积} & (3031.35, 3268.89) \\ & \text{周长} & (387.11, 410.70) \\ & \text{复杂度} & (3.92, 4.15) \end{pmatrix}$$

节域物元矩阵 R_e

$$R_e = \begin{pmatrix} \text{赤拟谷盗} & \text{面积} & (0, 3506.43) \\ & \text{周长} & (0, 434.29) \\ & \text{复杂度} & (0, 4.38) \end{pmatrix}$$

2.4 用可拓方法将待测粮虫分类的计算及结果分析

2.4.1 计算结果

以待测粮虫样本为例 (表 2), 其物元模型为

$$R_{x1} = \begin{pmatrix} P_{x1} & \text{面积} & 2715.00 \\ & \text{周长} & 363.00 \\ & \text{复杂度} & 3.86 \end{pmatrix}$$

。计算各待测物元与以

上 7 种害虫的综合关联度。

各特征的权系数为: 面积 0.4; 周长 0.3; 复杂度 0.3 (对权系数取不同数值进行大量计算, 结果表明各特征的权系数按上述比例选取比较合适)。由公式 (4) 和 (5), 用 matlab 编程计算, 其结果见表 4。

表 4 待测粮虫与各害虫的综合关联度计算结果

Table 4 Calculated results of the correlative degree between the pest to be recognized and every kind of pests

害虫序号	赤拟谷盗	杂拟谷盗	米象类	长头谷盗	锯谷盗	扁谷盗	谷 蠹
1	- 0.4914	- 0.5528	- 1.1941	- 4.3999	- 8.5146	- 33.0950	- 16.4090
2	- 0.7537	- 1.3699	- 0.3333	- 4.1575	- 7.5743	- 28.6039	- 14.2609
3	- 0.4838	- 0.5375	- 0.9812	- 4.4753	- 8.2472	- 32.1012	- 15.8386
4	- 0.6862	- 0.3571	- 3.0692	- 9.2096	- 13.0376	- 45.2129	- 25.3885
5	- 0.7691	- 0.7771	- 0.7479	- 0.6897	- 0.6185	- 0.5103	- 0.6099
6	- 0.6511	- 0.6653	- 0.6385	- 0.6205	- 0.5240	- 10.4068	- 0.2673
7	- 1.2790	- 1.7056	- 4.1907	- 10.9706	- 14.9749	- 50.9561	- 29.3548

2.4.2 结果分析

依综合关联度判别准则,若 $-1 < K(P_x) < 0$, 则与 7 种害虫综合关联度绝对值最小的那种,即为待测害虫的种类;若 $K(P_x) = -1$, 则待测害虫不属于该类害虫。对表 4 数据分析可得: 1 号和 3 号害虫应为赤拟谷盗; 2 号害虫属于米象类; 5 号属扁谷盗; 6 号为谷蠹。1~6 号待测害虫的分析计算结果与实际完全相符,这说明可拓工程方法应用于害虫分类是很有效的。至于 7 号害虫,它的 7 个 $K(P_x)$ 都小于 -1 , 按理论分析不属于 7 种害虫。出现此类情况是由于各特征权系数的选取需要通过大量的实验来确定,权系数的选择会影响计算结果,从而形成上述的少数偏差。总体来看,识别快速、正确率较高是这一方法的突出优点。本文只对前人实验得出的数据作了处理,至于依据其它数据或当害虫特征增加时,该方法的正确率是否受到影响,应作进一步深入研究。

3 结 论

- 1) 粮虫图像具有多参数、多类别和混合度大的特征,实现正确分类识别难度很大。可拓工程方法为粮虫的识别提供了合理而又可行的新途径。
- 2) 用关联函数可取负值的特点进行储粮害虫图像的分类识别能全面分析对象属于集合的程度。不过关联函数的选取没有确定的形式,可以根据实际问题而采用不同形式,因而具有一定灵活性,但通常采用文中公式(5)进行计算。

3) 对文中害虫样本数据进行处理可知: 物元与可拓集合理论应用于粮虫分类识别,概念清晰、计算快速、分辨率高。

[参 考 文 献]

[1] 邱道尹, 张洪涛, 等. 模糊识别技术在粮储害虫检测中的应用[J]. 农业系统科学与综合研究, 2002, 18(2): 122- 125

[2] 徐 昉, 邱道尹, 沈宪章. 粮仓害虫的特征提取与分类的研究[J]. 郑州工业大学学报, 2000, 21(4): 62- 65

[3] 邱道尹, 张成花, 张红涛, 等. 神经网络在储粮害虫识别中的应用[J]. 农业工程学报, 2003, 19(1): 142- 144

[4] 蔡 文, 杨春燕, 林伟初. 可拓工程方法[M]. 北京: 科学出版社, 1997.

[5] 王洪伟. 可拓学的进展及应用研究[J]. 广东工业大学学报, 1998, 15(增刊): 111- 116

[6] 肖筱南. 基于物元变换的可拓决策分析方法[J]. 当代经济科学, 1991, (1): 81- 84

[7] 蔡 文. 可拓学概述[J]. 系统工程理论与实践, 1998, (1): 76- 84

[8] 倪大平, 刘金禄, 史开泉. 可拓集与关联函数空间[J]. 信息工程学院学报, 1995, 14(2): 39- 43

[9] 贾文新. 可拓集合与关联函数[J]. 华北水利水电学院学报, 1997, 18(4): 63- 65

[10] 李祚泳. 由红外遥感数据反演地物的物元可拓识别方法[J]. 激光与红外, 2000, 30(2): 98- 101

[11] 蒋 淳, 田 山, 陈化然, 等. 地震综合预测物元模型及其应用[J]. 地震学报, 2000, 22(4): 418- 423

Classification and recognition of stored-grain pests
by using extension engineering method

Yuan Jinli¹, Ji Haiyan¹, Guo Zhitao²

(1. College of Infomation and Electrical Engineering, China Agricultural University, Beijing 100094, China;

2. School of Infomation Engineering, Hebei University of Technology, Tianjin 300130, China)

Abstract Using image technology to recognize the stored-grain insects, the classification and recognition of multi-characteristic parameters and multi-compound degree of various pests must be settled. This paper presents a new approach—extension engineering method—to solve this problem. This approach is based on constructing standard and extension matter-element matrixes for various pests classification using matter-element and extension set, calculating the correlative degrees between the stored-grain pests to be recognized and various pests, and recognizing the pests according to magnitude of correlative degree. According to the calculated result, it is proved to be an effective and practical approach to classification and recognition of stored-grain pests.

Key words: stored-grain pests detection; extension engineering; extension set; classification and recognition