

文章编号: 1006- 5342( 2010) 12- 0059- 02

# 基于模糊模式识别的蠓虫分类数学模型<sup>\*</sup>

王 琪

( 电子科技大学 中山学院计算机工程系, 广东 中山 528436)

摘 要: 首先在模糊聚类的前提下根据所给的 5 组蠓虫实际数据将其分为三个类别作为母本库, 然后再利用模糊模式识别的贴近度法将待识别的三组数据进行贴近度计算, 最后按照最大隶属度原则将它们归类到母体库中, 具有很好的应用价值和广阔的使用前景.

关键词: 模糊聚类; 贴近度; 隶属度

中图分类号: O 141

文献标识码: A

0 引言

模式识别是属于计算机应用科学范畴的一门交叉学科, 具有丰富的研究内容和广泛的应用背景. 所谓模式识别问题是指已知事物的各种类别 ( 标准模式 ), 判断对给定的或新得到的对象应该属于哪一类, 或是否为一个新的类别的问题. 如医生对病人的病情诊断; 科学家对生物种群的识别; 接收电话时对方的语音识别等, 都可以归结为模式识别问题.

在日常生活和实际问题中, 有些模式界线是明确的, 如识别男女性别时, 其模式是非常清楚的; 而有些模式界线是不很明确的, 如识别一个人的“胖”与“瘦”其模式“胖”和“瘦”界线是不明确的; 我们把这种界线不明确的模式称为模糊模式, 相对应的识别问题称为模糊模式识别问题, 而利用模糊集理论来处理模糊模式识别问题的方法称为模糊模式识别方法.

1 模糊模式识别方法原理及其步骤

1. 1 模糊模式识别的分类

模糊模式识别问题大致分为两种: 一种是所有已知模式的全体 ( 母体库 ) 是模糊的, 而待识别对象是分明的模式识别问题; 另一种母体库和待识别对象都是模糊的模式识别问题. 解决前一种模糊模式识别问题的方法称为模糊模式识别的直接方法, 而解决后一种的方法称为模糊模式识别的间接方法.

设  $U = \{ u_1, u_2, u_3, \dots, u_n \}$  为待识别的对象所构成的集合,  $U$  中每一个待识别对象  $u_i$  有  $m$  个指标  $u_{i1}, u_{i2}, \dots, u_{im}$  记  $u_i = ( u_{i1}, u_{i2}, \dots, u_{im} ) \quad ( i = 1, 2, \dots, n )$

如果待识别对象集合  $U$  可以分成  $p$  个类别, 且每一类别均为  $U$  上的一个模糊集, 记为  $A_1, A_2, A_3, \dots, A_p$  我们将其称为模糊模式.

1. 2 构造模糊模式

若现在有两种蠓虫  $A_f$  和  $A_{pf}$  已由生物学家根据它们的触角长和翼长加以区分, 观测的数据如表 1 其中 1~ 3 号的为  $A_f$  蠓, 4~ 5 号的为  $A_{pf}$  蠓. 现在根据上述数据来建立模糊模式.

1. 2. 1 数据规格化

由于  $m$  个指标的量纲和数量级可能不太一致, 所以一般在用数据之前先将数据规格化, 使每一指标值都统一到某种共同的数值特性范围. 本文使用最大值规格化.

( 1 ) 先把所给数据整理成一个特征指标矩阵, 如下所

示:

表 1 蠓虫触角长和翼长数据

序号	1	2	3	4	5
触角长	1. 24	1. 36	1. 56	1. 26	1. 30
翼长	1. 72	1. 74	2. 08	2. 00	1. 96

$$U^* = \begin{pmatrix} 1. 24 & 1. 72 \\ 1. 36 & 1. 74 \\ 1. 56 & 2. 08 \\ 1. 26 & 2. 00 \\ 1. 30 & 1. 96 \end{pmatrix}$$

( 2 ) 数据规格化

在特征指标矩阵中作变换  $x'_{ij} = \frac{x_{ij}}{M_j} \quad i = 1, 2, \dots, 5 \quad j = 1, 2$  其中  $M_1 = 1. 56, M_2 = 2. 08$  ( 表示每一列的最大数 ), 那么上述矩阵就可以变成:

$$U^* = \begin{pmatrix} 0. 79 & 0. 83 \\ 0. 87 & 0. 84 \\ 1 & 1 \\ 0. 81 & 0. 96 \\ 0. 83 & 0. 94 \end{pmatrix}$$

1. 2. 2 构造模糊相似矩阵

将已经规格化的数据  $x'_{ij} = ( i = 1, 2, \dots, 5; j = 1, 2 )$  利用最大最小法可构造模糊相似矩阵  $R = ( r_{ij} )_{5 \times 5}$  这里:

$$r_{ij} = \frac{\sum_{k=1}^2 ( x'_{ik} \wedge x'_{jk} )}{\sum_{k=1}^2 ( x'_{ik} \vee x'_{jk} )}$$

经计算可得到:

$$R = \begin{pmatrix} 1 & 0. 95 & 0. 81 & 0. 92 & 0. 92 \\ 0. 95 & 1 & 0. 86 & 0. 90 & 0. 92 \\ 0. 81 & 0. 86 & 1 & 0. 89 & 0. 89 \\ 0. 92 & 0. 90 & 0. 89 & 1 & 0. 98 \\ 0. 92 & 0. 92 & 0. 89 & 0. 98 & 1 \end{pmatrix}$$

1. 2. 3 利用平方自合成运算求  $t(R)$

由于由上述方法构造出的对象与对象之间的模糊关系

矩阵  $R$  一般来说只是一个模糊相似矩阵, 而不一定具有传递性, 从而未必是模糊等价矩阵. 而模糊相似矩阵  $R$  的传递闭包  $t(R)$  就是一个模糊等价矩阵, 因此可以以  $t(R)$  为基础进行分类.

$$R^2 = \begin{pmatrix} 1 & 0.95 & 0.81 & 0.92 & 0.92 \\ 0.95 & 1 & 0.86 & 0.90 & 0.92 \\ 0.81 & 0.86 & 1 & 0.89 & 0.89 \\ 0.92 & 0.90 & 0.89 & 1 & 0.98 \\ 0.92 & 0.92 & 0.89 & 0.98 & 1 \end{pmatrix}$$
$$\bigcirc \begin{pmatrix} 1 & 0.95 & 0.81 & 0.92 & 0.92 \\ 0.95 & 1 & 0.86 & 0.90 & 0.92 \\ 0.81 & 0.86 & 1 & 0.89 & 0.89 \\ 0.92 & 0.90 & 0.89 & 1 & 0.98 \\ 0.92 & 0.92 & 0.89 & 0.98 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0.95 & 0.89 & 0.92 & 0.92 \\ 0.95 & 1 & 0.89 & 0.92 & 0.92 \\ 0.89 & 0.89 & 1 & 0.89 & 0.89 \\ 0.92 & 0.92 & 0.89 & 1 & 0.98 \\ 0.92 & 0.92 & 0.89 & 0.98 & 1 \end{pmatrix} \in R$$
$$R^4 = \begin{pmatrix} 1 & 0.95 & 0.89 & 0.92 & 0.92 \\ 0.95 & 1 & 0.86 & 0.92 & 0.92 \\ 0.89 & 0.89 & 1 & 0.89 & 0.89 \\ 0.92 & 0.92 & 0.89 & 1 & 0.98 \\ 0.92 & 0.92 & 0.89 & 0.98 & 1 \end{pmatrix}$$
$$\bigcirc \begin{pmatrix} 1 & 0.95 & 0.89 & 0.92 & 0.92 \\ 0.95 & 1 & 0.86 & 0.92 & 0.92 \\ 0.89 & 0.89 & 1 & 0.89 & 0.89 \\ 0.92 & 0.92 & 0.89 & 1 & 0.98 \\ 0.92 & 0.92 & 0.89 & 0.98 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0.95 & 0.89 & 0.92 & 0.92 \\ 0.95 & 1 & 0.86 & 0.92 & 0.92 \\ 0.89 & 0.89 & 1 & 0.89 & 0.89 \\ 0.92 & 0.92 & 0.89 & 1 & 0.98 \\ 0.92 & 0.92 & 0.89 & 0.98 & 1 \end{pmatrix} = R^2$$

所以  $t(R) = R^2$ .

1.2.4 动态聚类, 建立母本库

适当选取置信水平值  $\lambda \in [0, 1]$ , 求出  $t(R)$  的  $\lambda$  截矩阵  $t(R)_\lambda$ , 然后按  $t(R)_\lambda$  进行分类, 具体的聚类原则如下.

设  $t(R) = (r_{ij})_{5 \times 5}$ ,  $t(R)_\lambda = (r_{ij}(\lambda))_{5 \times 5}$

则  $r_{ij}(\lambda) = \begin{cases} 1 & r_{ij} \geq \lambda \\ 0 & r_{ij} < \lambda \end{cases}$ .

(1) 首先, 把  $t(R)$  中的元素从大到小的顺序编排如下:  
 $1 > 0.98 > 0.95 > 0.92 > 0.89$

当  $\lambda = 1$  时, 5 个样本分为 5 类, 显然不合理; 若  $\lambda = 0.98$  则 5 个样本分为 3 类

$A_1$  类:  $x_1, x_2$   $A_f$  蠼螋;  $A_2$  类:  $x_3$   $A_f$  蠼螋;  $A_3$  类:  $x_4, x_5$   $A_{pf}$  蠼螋

如果取  $\lambda < 0.98$  的任意值, 则会出现  $A_f$  蠼螋和  $A_{pf}$  蠼螋就会混为一类, 使分类结果和实际情况不相符合. 因此, 取  $\lambda = 0.98$ , 既能使分类数最少, 又能使  $A_f$  和  $A_{pf}$  蠼螋分开, 所以为最佳分类.

(2) 建立母本库

将上面的分类结果直接作为母本库中的模糊集, 就有下面的模糊模式.

2 模糊模式识别过程及方法

我们知道两个模糊集合之间的接近程度可以用贴近度来度量, 两个模糊集合之间的贴近度越大, 他们就越接近.

因此可以利用贴近度来解决待识别对象与已知模式都是论域  $U$  上的模糊集合的模糊模式识别问题.

现有三只蠼螋的触角长和翼长的数据  $B_1(1.24, 1.80)$ ,  $B_2(1.28, 1.84)$  和  $B_3(1.40, 2.04)$ . 先根据所给的模糊模式来确定其类别.

表 2 母本库指标平均值

类别	蠼螋	母体样本	指标平均值	
			触角长	翼长
1	$A_f$	$x_1, x_2$	1.30	1.73
2	$A_f$	$x_3$	1.56	2.08
3	$A_{pf}$	$x_4, x_5$	1.28	1.98

2.1 择近原则

若  $A_1, A_2, A_3$  为论域  $U$  上的 3 个模糊模式 (母本库),  $B \in F(U)$  为一个待识别的对象, 若  $\sigma(B, A_i) = \max\{\sigma(B, A_1), \sigma(B, A_2), \sigma(B, A_3)\}$ , 则认为  $B$  应该属于模式  $A_i$ , 其中  $\sigma$  为  $F(U)$  的某种贴近度. 下面的计算中取:

$$\sigma(B, A_i) = \frac{\sum_{i=1}^2 (A(u_i) \wedge B(u_i))}{\sum_{i=1}^2 (A(u_i) \vee B(u_i))}$$

有:

$$\sigma(B_1, A_1) = 0.958 \quad \sigma(B_1, A_2) = 0.835$$
$$\sigma(B_1, A_3) = 0.933$$
$$\sigma(B_2, A_1) = 0.959 \quad \sigma(B_2, A_2) = 0.857$$
$$\sigma(B_2, A_3) = 0.957$$
$$\sigma(B_3, A_1) = 0.881 \quad \sigma(B_3, A_2) = 0.945$$
$$\sigma(B_3, A_3) = 0.944$$

由上面所给的最大隶属于可得:

$\sigma(B_1, A_1) = \max\{\sigma(B_1, A_1), \sigma(B_1, A_2), \sigma(B_1, A_3)\}$ , 故  $B_1(1.24, 1.73)$  属于  $A_1$  类, 且为  $A_f$  蠼螋.

$\sigma(B_2, A_1) = \max\{\sigma(B_2, A_1), \sigma(B_2, A_2), \sigma(B_2, A_3)\}$ , 故  $B_2(1.28, 1.84)$  属于  $A_1$  类, 且为  $A_f$  蠼螋.

$\sigma(B_3, A_2) = \max\{\sigma(B_3, A_1), \sigma(B_3, A_2), \sigma(B_3, A_3)\}$ , 故  $B_3(1.40, 2.04)$  属于  $A_2$  类, 且为  $A_f$  蠼螋.

3 总结

文章通过模糊聚类得到了模糊模式, 然后通过贴近度模式识别从而判断待识别对象的归类情况. 具有很好的示范性和可操作性, 能够在实际生活中起到一定的作用, 但是在实际操作过程计算量非常的大, 若能够通过计算机来求解的话能够减少很多工作量, 节约很多时间.

参考文献:

[1] 张兴华. 模糊聚类分析的新算法 [J]. 数学的实践与认识, 2005, 35(3): 138~141

[2] 吴善杰. 关于模糊聚类分析方法的进一步思考 [J]. 华北科技学院学报, 2008, 5(1): 108~111

[3] 王霞. 模糊聚类分析的一个改进算法及其应用 [J]. 天津科技大学学报, 2009, 12(6): 71~73

[4] 陈水利, 李敬公, 王向公. 模糊集理论及其应用 [M]. 北京: 科学出版社, 2005.

[5] 刘承平. 数学建模方法 [M]. 北京: 高等教育出版社, 2002.