

金翅夜蛾亚科的判别分析^{*}

周静芋

(西北农林科技大学 生命科学学院, 陕西 杨凌 712100)

[摘 要] 对金翅夜蛾亚科的2种分类情况分别以多种变量作为分类自变量, 用判别分析法建立了几个回判正确率高的判别函数, 并对其进行了讨论和比较。

[关键词] 金翅夜蛾亚科; 判别分析; 回判正确率; 主成分分析

[中图分类号] Q 969.436.4; O 212.4 [文献标识码] A [文章编号] 1671-9387(2005)02-0086-03

金翅夜蛾亚科(Plusiinae)是多食、杂食性的农业害虫, 危害多种农作物, 在农业上有一定重要性, 因此建立判别金翅夜蛾亚科分类的判别函数具有重要意义。著名昆虫学家周尧教授用歧序分类法把其中的18个属分为5类^[1]。周静芋等^[2]用多种聚类方法对其进行了数值分类研究, 结果表明周尧教授对金翅夜蛾亚科的分类是比较合理和科学的; 同时对葫芦夜蛾属和淡银夜蛾属的分类提出值得重视和探讨的问题, 并将18个属分为6类。周静芋等^[3]还用主成分分析法对其进行了分类研究, 再次证明文献^[2]中对葫芦夜蛾属和淡银夜蛾属的分类结果是有充分理由的。本研究用判别分析方法对2种分类结果分别建立判别函数, 对金翅夜蛾亚科的分类作进一步研究, 这个工作在昆虫分类中很少见到。

1 材料与方法

用文献^[2]中的18个属作为已知分类结果的样品, 分5类时各类分别为: 第1类 隐金夜蛾属(1号); 第2类 异纹夜蛾属(2), 弧铜夜蛾属(3), 印铜夜蛾属(4); 第3类 葫芦夜蛾属(5), 金弧夜蛾属(6), 金翅夜蛾属(7); 第4类 富丽夜蛾属(8), 银辉夜蛾属(9), 粉纹夜蛾属(10), 隐纹夜蛾属(11), 银纹夜蛾属(12), 淡银纹夜蛾属(13), 珠纹夜蛾属(14); 第5类 金斑夜蛾属(15), 丫纹夜蛾属(16), 黑银纹夜蛾属(17), 银锭夜蛾属(18)。分6类时各类分别为: 第1类 隐金夜蛾属(1号); 第2类 异纹夜蛾属(2), 弧铜夜蛾属(3), 印铜夜蛾属(4); 第3类 葫芦

夜蛾属(5); 第4类 金弧夜蛾属(6), 金翅夜蛾属(7); 第5类 富丽夜蛾属(8), 银辉夜蛾属(9), 粉纹夜蛾属(10), 隐纹夜蛾属(11), 银纹夜蛾属(12), 珠纹夜蛾属(14); 第6类 淡银纹夜蛾属(13), 金斑夜蛾属(15), 丫纹夜蛾属(16), 黑银纹夜蛾属(17), 银锭夜蛾属(18)。

分别用文献^[3]中得到的5个主成分(在用协方差矩阵做的主成分分析中, 前5个主成分的贡献率为82.40%), 用对分类起最重要作用的前4个特征变量(抱器、导精管、皮肤突起和金银斑)和文献^[2]中的20个特征变量作为分类的自变量。

用费歇尔和贝叶斯判别分析法建立判别函数。无特别说明时, 建立判别函数用全模型法。在费歇尔判别中, 选择使用合并组内协方差矩阵作为分类使用的协方差矩阵, 判别准则是用所求的几个判别函数计算样品与各类的马氏距离, 判样品归距离最大的一类; 在贝叶斯判别中, 设各类的先验概率相等, 判别准则是计算样品属于各类的验后概率, 判样品归验后概率最大的一类^[4]。2种方法的回判正确率一样, 故在本研究中对各种情况只给出判别所用的几个费歇尔判别函数; 用贝叶斯判别时, 各样品属于各类的验后概率。

2 结 果

2.1 用5个主成分建立判别函数

分5类时, 费歇尔判别函数为(全文计算用SPSS 10.0统计软件):

$$y_1 = 3.351f_1 - 0.142f_2 + 0.278f_3 - 0.514f_4 + 0.574f_5, \\ y_2 = 0.055f_1 - 1.241f_2 + 1.310f_3 + 1.126f_4 - 0.253f_5,$$

* [收稿日期] 2004-03-01

[作者简介] 周静芋(1944-), 女, 浙江鄞县人, 教授, 主要从事农业应用数学研究。

$$y_3 = 0.196f_1 + 1.127f_2 + 0.608f_3 + 0.331f_4 - 0.865f_5,$$

$$y_4 = -0.046f_1 + 0.518f_2 + 0.079f_3 + 0.704f_4 + 0.989f_5,$$

结果表明, 回判正确率为100%。具体判别情况是: 1, 2, 3, 8, 11 和12 号属于原来类的验后概率均为1.000 00; 4~ 7, 9, 10 和13~ 18 号属于原来类的验后概率依次为 0.999 99, 0.999 58, 0.993 30,

$$y_1 = 2.686f_1 - 2.529f_2 + 1.588f_3 + 0.256f_4 + 0.314f_5,$$

$$y_2 = 1.743f_1 + 1.782f_2 - 0.166f_3 - 0.459f_4 + 0.407f_5,$$

$$y_3 = 0.305f_1 + 0.007f_2 - 0.788f_3 + 1.539f_4 + 0.177f_5,$$

$$y_4 = -0.126f_1 + 0.512f_2 + 1.058f_3 + 0.642f_4 - 0.673f_5,$$

$$y_5 = -0.327f_1 + 0.139f_2 + 0.514f_3 + 0.194f_4 + 1.160f_5.$$

结果表明, 回判正确率为100%。具体判别情况是: 7 和15~ 17 号属于原来类的验后概率依次为0.999 93, 0.998 87, 0.999 76 和0.999 97; 其余各样品属于原来类的验后概率均为1.000 00。

根据各样品的前两个判别函数分数值可画出样

$$y_1 = 1.139x_2 + 0.678x_8 - 1.410x_{16} + 0.929x_{20} - 3.787,$$

$$y_2 = 0.885x_2 + 0.365x_8 + 1.029x_{16} + 0.051x_{20} - 3.884,$$

$$y_3 = -0.740x_2 + 1.267x_8 + 0.519x_{16} + 1.163x_{20} - 2.188,$$

$$y_4 = -0.023x_2 - 1.049x_8 + 0.624x_{16} + 0.610x_{20} - 0.663,$$

结果表明, 判错1 个, 回判正确率为94.4%。将属于第4 类的13 号以0.734 49 的验后概率判为第5 类; 其余判为原来类的验后概率依次为: 0.999 21, 0.996 64, 0.953 51, 0.996 34, 0.847 99, 0.652 02, 0.

$$y_1 = 0.395x_2 + 0.309x_8 + 4.221x_{16} - 0.264x_{20} - 5.535,$$

$$y_2 = 1.526x_2 + 0.570x_8 - 0.012x_{16} + 0.499x_{20} - 5.418,$$

$$y_3 = -0.207x_2 + 2.209x_8 - 0.792x_{16} - 0.561x_{20} + 0.121,$$

$$y_4 = -0.641x_2 + 0.836x_8 + 0.538x_{16} + 1.241x_{20} - 2.150.$$

结果表明, 回判正确率为100%。18 个属分别以验后概率 0.999 98, 0.999 97, 0.786 90, 0.999 32, 1.000 00, 0.999 96, 0.956 09, 0.999 99, 0.999 50, 0.997 02, 0.999 50, 0.999 99, 0.999 96, 1.000 00, 0.963 12, 0.969 46, 0.999 74 和0.998 30 判为原来所属类。

2.3 用20 个特征变量为分类的自变量

对分5 类, 经各变量的均值和方差检验后保留变量为1~ 6, 8~ 10, 13, 15, 16 和18 共13 个变量, 判别函数略。结果表明, 回判正确率为100%, 18 个属均以验后概率1.000 00 判为原来所属类。

对分6 类, 经各变量的均值和方差检验后保留变量为1~ 10, 12 和15 共12 个变量, 判别函数略。结果表明, 回判正确率为100%, 18 个属均以验后概率1.000 00 判为原来所属类。

0.984 84, 0.999 99, 0.995 60, 0.978 97, 0.999 99, 0.591 28, 0.994 38, 0.995 78 和0.999 91。

分6 类时, 费歇尔判别函数为:

品散点图, 分5 类时的散点图中, 第3 与第5 类挤在一起; 分6 类时的散点图中, 各类明显分开, 下面的情况类同。

2.2 用4 个特征变量建立判别函数

分5 类时, 费歇尔判别函数如下:

711 62, 0.998 82, 0.976 11, 0.757 15, 0.976 11, 0.998 82, 0.999 92, 0.602 72, 0.708 78, 0.934 26 和0.971 98。

分6 类时, 费歇尔判别函数如下:

如果不是用全模型法, 而是用Wilks' lambda 逐步判别分析法, 使用F 值作为判据统计量, 加入模型的变量需满足 $F \geq 3.84$, 移出模型的变量需满足 $F \leq 2.71$, 结果如下:

对分5 类, 进入模型的变量(括号里的数表示变量号)为: 前翅臀齿栉(9), A₃₋₄节SV 毛(14), 皮肤突起(16)和下唇须(18)。判别函数略。结果表明, 错判3 个: 将原属第2 类的2 号, 第4 类的13 号和第5 类的15 号依次以验后概率0.585 32, 0.991 24 和0.585 32 分别判为第3 类, 第5 类和第3 类; 7 号以验后概率0.585 32 判对, 其余均以验后概率大于0.99判对, 回判正确率为83.3%。

对分6 类, 进入模型的变量为: 中室下纹(7), 金银斑(8), 前翅臀齿栉(9), 皮肤突起(16)。判别函数略。结果表明, 2 号和15 号均以验后概率0.999 66 判

为原来所属类,其余的均以验后概率1.000 00判为原来所属类,回判正确率为100%。

3 结论与讨论

1)用5个主成分和20个特征变量作为分类自变量建立判别函数,对分5类和6类的回判正确率均为100%,这首先说明建立的判别函数是有效的,其次有力地说明文献[2]和[3]提出的5个主成分确实可以代表20个特征变量起到降维的作用,当然用20个特征变量时,验后概率高于用5个主成分,这是因为前者比后者利用了更多的原有信息。

2)用4个特征变量作为分类自变量建立判别函数的回判正确率,对分5类时为94.4%,对分6类时为100%,这也说明建立的判别函数是有效的,同时说明文献[3]提出的4个特征变量确实是重要的。

3)用5个主成分和4个特征变量分别作为分类自变量建立判别函数,从回判正确率尤其是验后概率都可以明显看出,分6类的判别效果较分5类好。

用5个主成分时回判正确率皆为100%。在分5类中,15号被判为与16、17和18号同属一类的验后概率仅为0.591 28(判为与6和7号同属一类的概率为0.408 71),其余均大于0.9;在分6类中,各判别的验后概率均大于0.99。用4个特征变量时,在分5类时,回判正确率为94.4%,其中错将13号以0.734 49的验后概率判为与16~18号为一类,其余判对的有5个的验后概率小于0.80;在分6类时,回判正确率为100%,其中只有1个验后概率为0.786 90,其余的均大于0.95。在文献[2]和[3]中特别指出,13号应与16~18号为一类。以上结果从判别分析的角度说明,提出分6类的具体分法是值得重视和探讨的。

4)值得继续研究的问题是,在用20个特征变量为分类的自变量和逐步判别分析法时,保留的变量并不都是文献[3]中提出的重要变量,特别是文献[3]中提出的重要变量导精管均未作为保留的变量,其内在原因值得进一步探讨。

[参考文献]

- [1] 周尧,卢笋.金翅夜蛾亚科分类系统的讨论[J].昆虫分类学报,1979,1(2):71-78
- [2] 周静芋,宋世德.金翅夜蛾亚科(鳞翅目:夜蛾科)数值分类研究[J].西北大学学报,1995,23(6):28-32
- [3] 周静芋,宋世德,袁志发.用主成分分析法分析金翅夜蛾亚科各性状对分类的重要性和状态变异方向[J].生物数学学报,1996,11(4):77-82
- [4] 袁志发,周静芋.多元统计分析[M].北京:科学出版社,2002.220-241.

Discriminant analysis of Plusiinae

ZHOU Jing-yu

(College of Life Sciences, Northwest A & F University, Yangling, Shaanxi 712100, China)

Abstract: This article constructed several discriminant functions by varying grouping independent variables for two classification results of Plusiinae. The discriminant functions are provided with original grouped cases correctly classified. Detailed features of those functions were discussed and compared in the article.

Key words: Plusiinae; discriminant analysis; original grouped cases correctly classified; principal component analysis