

灰关联分析与聚类在大豆食心虫 预报因子筛选中的应用*

席景会 张志生

(东北农业大学) (松花江地区植保站)

摘 要

本文应用灰色关联与聚类分析法,以黑龙江省双城市1987—1992年大豆食心虫的为害情况作为历史资料,对与大豆虫食率有关的十个因子进行了优选。根据符合程度将其分成五类,筛选出最有应用价值的二个因子,为大豆食心虫的科学预测提供了理论依据,并在预报因子选择方法上做出了新的尝试。

关键词 大豆食心虫;关联度;灰色聚类

大豆食心虫 [*Leguminivora glycinivorella* (Matsamura)] 是黑龙江省大豆的主要害虫之一,每年都有不同程度的发生,是影响大豆产量与品质的重要因素。为了经济、有效、科学地进行防治,准确地预测大豆食心虫的发生程度是首要前提,这就涉及到预报因子的选择,而传统的预报因子筛选大多采用统计学方法,计算相关系数,要求数据的分布是典型的,各因素之间呈线性关系。可是在客观世界中,许多因素之间的关系是灰色的,分不清哪些密切,哪些不密切。关联分析则根据灰色系统理论,进行系统发展势态的量化比较与分析来确定预报因子的优劣。基于上述原则,我们以大豆虫食率为因变量,与十个因子之间构成一个灰色系统,通过关联度计算与聚类分析,对大豆食心虫预报因子的优选方法进行了新的探讨。以期找到科学合理的测报方法,为农业生产服务。

一、方法与原理

设有参考数列 $X_0(R)$:

$$X_0(R) = (X_0(1), X_0(2), X_0(3), \dots, X_0(n))$$

* 本文得到东北农业大学植保系赵奎军副教授的精心指点,谨此致谢。

本文于1994年3月10日收到。

This paper was received on March 10, 1994.

有 i 个比较数列, $i=1, 2, 3, \dots, m, X_i(R)$

$X_1(R) = (X_1(1), X_1(2), \dots, X_1(n))$

$X_2(R) = (X_2(1), X_2(2), \dots, X_2(n))$

.....

$X_m(R) = (X_m(1), X_m(2), X_m(3), \dots, X_m(n))$

将参考序列和比较序列作数据预处理。

求参考数列 X_0 与比较数列 X_i 之间的差列 Δ 。

取分辨系数 $\delta = [0, 1]$ 计算相关系数

$$\xi_i(R) = \frac{\min_i \min_R |X_0(R) - X_i(R)| + \delta \max_i \max_R |X_0(R) - X_i(R)|}{|X_0(R) - X_i(R)| + \delta \max_i \max_R |X_0(R) - X_i(R)|}$$

求关联度

$$r_i = \frac{1}{N} \sum_{k=1}^N \xi_i(k) = \frac{1}{N} (\xi_i(1) + \xi_i(2) + \xi_i(3) + \dots + \xi_i(n))$$

全体 r_i 值构成关联度集 $R = (r_1, r_2, \dots, r_m)$, 关联度集为聚类分析的基础, 由关联度 r_i

定义分析论域中 X 中各元素之间的关联度差异矩阵 $Df: d_{ij} = \frac{|r_i - r_j|}{r_j}$

其中 $d_{ij} (i, j=1, 2, 3, \dots, m)$ 为差异系数

$$Df = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1m} \\ d_{21} & d_{22} & \dots & d_{2m} \\ \vdots & \vdots & & \vdots \\ d_{m1} & d_{m2} & \dots & d_{mm} \end{bmatrix}$$

由 Df 矩阵得差异距离矩阵 Ds , 其中 Ds 矩阵中的元素由下式计算得出:

$$S_{ij} = d_{ij} + d_j$$

$$Ds = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1m} \\ S_{21} & S_{22} & \dots & S_{2m} \\ \vdots & \vdots & & \vdots \\ S_{m1} & S_{m2} & \dots & S_{mm} \end{bmatrix}$$

由 Ds 得灰色相似关系矩阵 Rg

$$Rg = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix}$$

其中 $r_{ij} = 1 - s_{ij} / \max(Ds)$, 式中 $\max(Ds)$ 表示为矩阵 Ds 中的最大元素, 矩阵 Rg 满足 $r_{ii} = 1, r_{ij} = r_{ji}$ 。

根据灰色相似矩阵, 采用最大树法进行聚类分析。

二、关联因子的选择与聚类

根据 1987—1992 年在双城市调查得米的数据, 筛选出以下十个因子。

X_1 ——上一年脱荚孔荚率

X_2 ——上一年平方米脱荚孔数

X_3 ——上一年九月中、下旬降雨量

X_4 ——上一年九月中、下旬平均温度

X_5 ——当年七月份温雨系数

X_6 ——当年七月份平均湿度

X_7 ——当年七月上、中旬温度

X_8 ——当年八月上旬温度

X_9 ——当年七月下旬至八月上旬平均温度

X_{10} ——当年七月下旬至八月上旬平均湿度

视大豆虫食率的大小与各因子之间的关系为一灰色系统, 每个预报因子为系统中一个元素。以大豆虫食率所构成的序列为系统的参考序列, 其它因子构成的序列为比较序列, 采用均值法, 运用 Basic 语言程序, 在 PC—计算机上运行, 结果如下: (见下页表 1)

依各因子 X_i 对 X_0 的关联度大小, 得因子优劣顺序, 依次为: $X_3, X_5, X_6, X_9, X_4, X_7, X_8, X_{10}, X_2, X_1$ 。即上一年九月中、下旬的降雨量因子最好, 当年七月份温雨系数次之。

根据关联度大小建立灰色相似矩阵进行聚类。

依据关联度集, $R = (0.669, 0.798, 0.918, 0.871, 0.902, 0.877, 0.867, 0.863, 0.876, 0.855)$, 按 $d_{ij} = |r_i - r_j| / r_i$ 计算, 得差异矩阵 Df 。

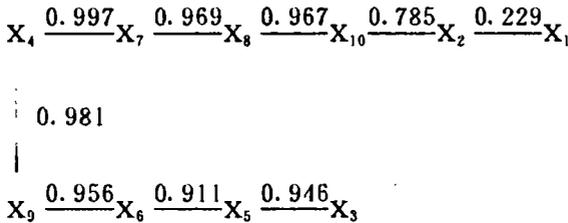
$$Df = \begin{bmatrix} 0 & 0.162 & 0.271 & \cdots & 0.236 & 0.218 \\ 0.193 & 0 & 0.131 & \cdots & 0.089 & 0.067 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0.309 & 0.098 & 0.046 & \cdots & 0 & 0.025 \\ 0.278 & 0.071 & 0.069 & \cdots & 0.024 & 0 \end{bmatrix}$$

$$Rg = \begin{bmatrix} 1 & & & & & & & & & \\ 0.448 & 1 & & & & & & & & \\ 0 & 0.563 & 1 & & & & & & & \\ 0.170 & 0.728 & 0.837 & 1 & & & & & & \\ 0.058 & 0.619 & 0.946 & 0.891 & 1 & & & & & \\ 0.148 & 0.706 & 0.857 & 0.978 & 0.911 & 1 & & & & \\ 0.185 & 0.742 & 0.821 & 0.984 & 0.877 & 0.964 & 1 & & & \\ 0.196 & 0.753 & 0.810 & 0.975 & 0.866 & 0.953 & 0.991 & 1 & & \\ 0.152 & 0.709 & 0.854 & 0.981 & 0.908 & 0.997 & 0.969 & 0.956 & 1 & \\ 0.229 & 0.785 & 0.778 & 0.942 & 0.833 & 0.922 & 0.956 & 0.967 & 0.924 & 1 \end{bmatrix}$$

采用最大树聚类法进行聚类,选 X_4 为顶点,构造聚类图

图1 聚类图

Classification figure



灰色相似关系确定后,采用最大树聚类法对 X_i 进行聚类,根据 (X_i, R_g) 二元素组构成最大树,取 $\lambda = 0.92$,砍断大于 λ 的各连通分枝,构成了在 λ 水平上的聚类,这样将十个预报因子分成五类。

最好: X_3

很好: X_2

一般: $X_6, X_9, X_4, X_7, X_8, X_{10}$

差: X_2

最差: X_1

三、分析与讨论

本文应用灰色关联分析法对大豆食心虫的预报因子进行了优选,根据关联度的大小与聚类分析可知上一年九月中、下旬降雨量与大豆食心虫的发生关系最为密切,为最优因子,当年七月份的温雨系数这个因子也比较好。九月中、下旬是大豆食心虫脱荚盛期,此时若降雨偏多,土壤湿润,有利于幼虫入土及越冬;七月份是越冬幼虫由地下向地表转移及作新蛹阶段,这时温度适中,雨量大,有利于幼虫的转移、化蛹,幼虫死亡率低。因此,关联分析与聚类的结果与大豆食心虫的生物学特性非常吻合。

灰色聚类分析表明, X_3, X_5 应用价值最大,可以作为以后预报大豆食心虫发生的主要

因子。

本文因子的选择比较少,如没有考虑天敌的影响,数据的年限也较短, λ 值的选择也有些人为的因素,可能会影响结果的可靠性,这些需在以后的工作实践中进一步改进。

参考文献

- [1] 贺仲雄,1982,《模糊数学及其应用》,159—162,天津科学技术出版社
- [2] 邓聚龙,1987,《农业系统灰色理论与方法》,42—60,山东科学技术出版社
- [3] 韩 宁,1987,《微机与农业系统工程应用软件》,244—253,山东科学技术出版社
- [4] 邓聚龙,1985,《灰色系统》,49—60,国防工业出版社
- [5] 顾成玉等,1992,大豆食心虫产量损失与防治指标的研究,病虫测报,(4)33—35

APPLICATION OF GREY RELATED ANALYSIS AND CLASSIFICATION TO THE SELECTION OF FORECASTING FACTORS OF *LEGUMINIVORA GLYCINIVORELLA*

Xi Jinghui

(Northeast Agricultural University)

Zhang Zhisheng

(Station of Plant Protection of Songhuajiang Region)

Abstract

The analysis of accumulated data on *leguminivora glycinivorella* damage was conducted by using application of grey theory to select forecasting factors at Shuangcheng city of Heilongjiang province during 1987—1992. The result showed that out of the ten factors rainfall on middle to late September and thermorainfall coefficient of July were the two factors most important for selecting forecast factors. This would provide a scientific basis and a new way for forecast of pests.

Key words *Leguminivora glycinivorella*; Related degree; Grey classification