

自动交互检测法在营口水稻年景预报中的应用

葛日东 (营口市气象局 115001)

1 引言

在水稻年景预报实践中得知,水稻年景与气象因子及因子之间的相互组合之间是呈复杂的非线性关系的。因此,运用比较简便的非线性方法去建立效果较好的水稻年景预报模式,是农业气象预报服务工作所迫切需要的。

自动交互检测法(简称A·I·D方法)是解决非线性问题的比较简单的方法。该方法可与逐步回归、逐步判别方法一样,一步步地挑选因子,但不同的是每个因子只对若干个样本作出判别分类,因而用它建立的预报模式更能真实地反映出因子和预报对象之间的关系。

利用A·I·D方法,做营口市水稻年景预报,经验证和试报,效果较好。

2 A·I·D方法介绍

A·I·D方法是将因子按其取值大小顺序排列,交替地对于预报对象进行最优二分分割,使完成每一次分割对象的变差和最小。

设有一资料矩阵

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ Y \end{pmatrix} \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \\ Y_1 & Y_2 & \cdots & Y_m \end{pmatrix}$$

第1步:先将任意一因子 $X_{ij}(i=1,2,\dots,n$ 为因子数; $j=1,2,\dots,m$ 为样本数)按由大到小的顺序排列,即 $X_{i1} > X_{i2} > \dots > X_{im}$,并使预报对象 Y 作出相应排列后,可得到预报对象的一个新序列,将该序列在 C 点分割成 $A(X_{ij} \geq C)$ 、 $B(X_{ij} < C)$ 两组,分别计算每一组的平均值。

$$\bar{Y}_{(A)} = \frac{1}{mA} \sum_{i=1}^m Y_i, \quad \bar{Y}_{(B)} = \frac{1}{mB} \sum_{i=1}^m Y_i$$

mA 、 mB 分别是 A 、 B 两组中的样本数,有 $m = mA + mB$ 。

每一组的变差用下式求算:

$$S_{(A)} = \sum_{X_{ij} \geq C} (Y_i - \bar{Y}_{(A)})^2$$

$$S_{(B)} = \sum_{X_{ij} < C} (Y_i - \bar{Y}_{(B)})^2$$

然后得出变差和 $S_{(C)}, S_{(C)} = S_{(A)} + S_{(B)}$ 。对预报对象 Y 进行不同点的分割,得出 $m-1$ 个变差和,从中选取一个最小的 $S_{(C)}$,这就是因子 X_{ij} 对预报对象 Y 进行的最优二分分割。可以记为: $S_{(C)} = \min S_{(C)} \quad 1 \leq C \leq m-1$ 。其中, C^* 是以 C 点为最优分割点的分界点。这时,预报对象 Y 被分为两类,即 $(Y_1, Y_2, \dots, Y_{C^*})$ 、 $(Y_{C^*+1}, Y_{C^*+2}, \dots, Y_m)$ 。

第2步:同样方法,对其他因子重复上述步骤就可得到各因子对预报对象的最优二分分割的一组变差和 $S_{(C^*)}$ 序列 $S_{(1,C^*)}, S_{(2,C^*)}, \dots, S_{(n,C^*)}$ 。注意其中的每个 C^* 是不一样的($1, 2, \dots, n$ 是因子数),在得到的一组变差序列 $S_{(i,C^*)}(i=1, 2, \dots, n$ 因子数)中,挑选出最小的一个变差和 $S_{(i,C^*)}$,该因子入选。

第3步:对上述两类进行方差分析,如果通过 F 检验,表明以 C^* 点为分界点的分组具有显著性差异,分割有效,否则说明不必进行这种分割。 C^* 点就是最优分割的分组界值,至此完成一次分割。

第4步:在对预报对象 Y 分割成 A 、 B 两组后,再用其他因子分别对 A 、 B 两组重复上述步骤,随着因子不断入选(不排除重复选入同一个因子),不断分组,直至分割到不再显著时为止。经过 K 次分割,先后取得 K 个分组界值, C_1, C_2, \dots, C_k ,预报对象被分割成 $K-1$ 组,每一组都是因子和预报对象互相对应,各组间的这种对应都有显著差异。

3 用A·I·D方法做营口市水稻年景预报

3.1 年景等级与因子选择

水稻亩产为营口市统计局的社会产量。资料长度为21年(1970~1990年)。年景等级共分5级,分级方法是参照最近5年的平均产量进行评定。取评定年的产量(Y_i)与前5年平均产量($\bar{Y} = \sum_{j=1}^5 Y_j$)之比值 $Q: Q = Y_i/\bar{Y} \times 100\%$ 。分级标准如表1。

表1 营口市水稻年景评定指标Q

年景	丰	平	丰	平	平	歉	歉
Q(%)	≥ 120	119~110	109~100	99~90	90		
年景编号(级)	一	二	三	四	五		

根据相关普查^[2],选取如下3个因子: X_1 , 预报年的前一年1月份60°N、90°E500hPa距平值。 X_2 , 预报年的前一年1月份60°N、30°E500hPa距平值。 X_3 , 预报年的前一年9月份20cm平均地温值。

3.2 运算分割

首先将 X_1 按由大到小的顺序排列,将水稻产量 Y 依据 X_1 之对应顺序排序,并分割之。同时,将 X_2 、 X_3 分别对 Y 进行分割,因用 X_1 分割 Y 的总变差和最小,故 X_1 入选。其最优分割点是1。即将 Y 分割为两组: $X_1 > 9$ 时(Y_{16}), $X_1 \leq 9$ 时($Y_1 Y_2 \dots Y_{15} Y_{17} \dots Y_{21}$)。其 F 值为 $F = 10.2$,查 F 分布表知 $F_{0.01} = 8.18$,因此 $F > F_{0.01}$,说明该分割点在0.01水平下显著,分割有效。同理,将用 X_1 分割后形成的两组,再用 X_2 、 X_3 分别予以分割,并选取其中总变差和最小的那个因子,作为第二次分割的入选因子,其分割点即为第二次分割点,并进行方差检验。余此类推,直至分割到方差检验不显著时为止。最终得到营口市水稻年景预报结果。

3.2.1 一级年景(丰收年景)。当预报因子值出现下列3种方式组合时,预报水稻年景为一级年景。①当 $-6 \leq X_1 \leq 9$ 、 $-1 \leq X_2 \leq 2$ 、 $X_3 < 20.5$ 时;②当 $X_1 < -6$ 、 $X_2 \geq -1$ 、 $X_3 < 20.5$ 时;③当 $X_1 \leq 9$ 、 $X_2 < -1$ 、 $X_3 \geq 21.7$ 时。

3.2.2 二级年景(平丰年景)。当预报因子值出现下列两种方式组合,预报水稻年景为二级

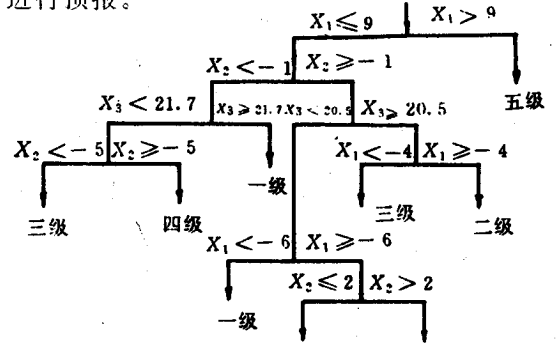
年景。①当 $-4 \leq X_1 \leq 9$ 、 $X_2 \geq -1$ 、 $X_3 \geq 20.5$ 时;②当 $-6 \leq X_1 \leq 9$ 、 $X_2 > 2$ 、 $X_3 < 20.5$ 时。

3.2.3 三级年景(平年年景)。当预报因子值出现下列两种方式组合时,预报水稻年景为三级年景。①当 $X_1 < -4$ 、 $X_2 \geq -1$ 、 $X_3 \geq 20.5$ 时;②当 $X_1 \leq 9$ 、 $X_2 < -5$ 、 $X_3 < 21.7$ 时。

3.2.4 四级年景(平歉年景)。当预报因子值以下述方式组合时,预报水稻年景为四级年景。即 $X_1 \leq 9$ 、 $-5 \leq X_2 < -1$ 、 $X_3 \geq 21.7$ 。

3.2.5 五级年景(歉年年景)。当预报因子 X_1 值为大于9时(即 $X_1 > 9$),则预报水稻年景为五级年景。

3.2.6 在业务工作中,可用下面的框图(附图)进行预报。



附图 自动交互检测法预报框图

4 结论

4.1 用自动交互检测法预报水稻年景与一般的线性回归等方法是有较大差异的,它可以用定量描述水稻年景与因子间的非线性关系,而且能够综合分析水稻年景变化中两个及其以上因子的交叉贡献,具有每个因子只对若干样本作出判别的特点,因此比较符合实际情况。本文所做的1970~1992年预报历史拟合率为95.2%(一级不差),另有4.8%只差一级(1987年)。用于实际预报结果为:1991年和1992年预报水稻年景均为二级年景(平丰年景),实际是1991年和1992年均均为二级年景,预报准确,可用于日常业务。预报及历史拟合表略。

4.2 由于本文所选因子是前一年的因子,因此在年底前就可做出下一年的水稻年景预报,这为有关部门安排部署下一年的水稻生产提供了决策依据。