

叠加马尔可夫链模型在雨季强度指数预测中的应用

沈永梅 夏乐天

(河海大学理学院 ,江苏 南京 210098)

摘要 :根据雨季降水过程中的不确定性和非精确性特点 ,用有序聚类法(即最优分割法)建立华北地区雨季强度指数的状态分类标准 ,并以 1951 ~ 1994 年华北地区雨季强度指数所处的状态为依据 ,用叠加马尔可夫链预测方法来预测 1995 年和 1996 年华北地区雨季强度指数所处的状态 .结果表明 ,叠加马尔可夫链预测方法计算方便 ,结果准确 ,可为雨季强度指数状态预测分析提供一种新的途径 .

关键词 :华北地区 ;雨季强度指数 ;马尔可夫链模型 ;有序聚类法

中图分类号 :O21 文献标识码 :A 文章编号 :1000-1980(2006)04-0477-04

我国北方大部分地区的降水主要集中在夏季 ,特别是北方东部地区 ,此现象更为明显 .本文首先利用有序聚类法^[1]建立雨季强度指数状态分类标准 ,然后以华北地区雨季强度指数^[2]前 44 个时段所处的状态为依据 ,预测了第 45 和第 46 个时段强度指数所处的状态^[3] .

1 雨季强度指数状态分类

本文利用有序聚类法(即最优分割法)^[1]对华北地区雨季强度指数状态进行分类 .

假设以 $x_1, x_2, \dots, x_n$ 表示 n 个有序样品 ,则每一类的形式为 $\{x_i, x_{i+1}, \dots, x_j\} (i < j)$.若将 n 个样品分成 k 类 ,设某种分法是

$$P(n, k) : \{x_{i_1}, x_{i_1+1}, \dots, x_{i_2-1}\}, \{x_{i_2}, x_{i_2+1}, \dots, x_{i_3-1}\}, \dots, \{x_{i_k}, x_{i_k+1}, \dots, x_{i_{k+1}-1}\}$$

可简写成

$$P(n, k) : \{i_1, i_1 + 1, \dots, i_2 - 1\}, \{i_2, i_2 + 1, \dots, i_3 - 1\}, \dots, \{i_k, i_k + 1, \dots, i_{k+1} - 1\}$$

其分点为

$$1 = i_1 < i_2 < \dots < i_k < i_{k+1} - 1 = n$$

则这种分类的目标函数可定义为

$$d[P(n, k)] = \sum_{j=1}^k D(i_j, i_{j+1} - 1) \tag{1}$$

其中

$$D(i, j) = \sum_{k=i}^j (x_k - \bar{x}_{ij})^2 \quad \bar{x}_{ij} = \frac{1}{j-i+1} \sum_{\lambda=i}^j x_\lambda$$

$d[P(n, k)]$ 为各类的离差平方和 .当 n 与 k 固定时 , $d[P(n, k)]$ 越小则各类的离差平方和越小 ,此时的分类是合理的 .样品到底分成几类 ,取决于目标函数关于分类数 k 的图像 .当无法确定 k 值时 ,可按矩阵 $\{d[P(i, j)]\}_{n \times n}$ 中的最后一行数值画出 $d[P(n, k)]$ 关于 k 的图像 ,转折点处的 k 则是所要找的最优分割类数 .据此 ,按表 1 雨季强度指数由小到大排列后计算所得的数据画出目标函数关于分类数 k 的图像 ,如图 1 所示 .由图 1 可知 ,最优分类数为 5 .

根据有序聚类的基本思想 ,其分类状况为

$$\{1, \dots, 9\}, \{10, \dots, 19\}, \{20, \dots, 36\}, \{37, \dots, 41\}, \{42, 43, 44\}$$

由于所取的状态分点应保证上述分类成立,取第1个分点为 $[\alpha(9)+\alpha(10)]/2$.以此类推,可得到将雨季强度指数分为5个状态的4个分点.具体状态分类标准见表2.

表1 华北地区雨季强度指数所处状态^{2]}

Table 1 State of rainfall intensity index for rainy season of North China

年份	开始日期 (月-日)	结束日期 (月-日)	长度/旬	指数	状态	年份	开始期 (月-日)	结束期 (月-日)	长度/旬	指数	状态
1951	07-03	08-03	3	1.99	2	1974	07-03	08.02	3	2.69	3
1952	07-03	08-01	2	1.81	1	1975	07-01	08-02	4	3.24	3
1953	07-02	08-03	4	3.39	3	1976	07-02	09-01	4	3.93	3
1954	06-01	08-03	8	6.18	5	1977	06-03	08-02	5	5.00	4
1955	08-01	09-03	5	4.61	4	1978	07-03	09-01	4	3.31	3
1956	06-01	08-02	7	5.19	5	1979	07-02	08-03	3	2.81	3
1957	07-01	07-03	2	1.78	1	1980	06-01	06-02	1	0.95	1
1958	07-01	08-02	4	3.84	3	1981	07-01	07-03	2	1.76	1
1959	07-01	09-01	6	4.84	4	1982	07-03	08-02	3	2.18	2
1960	07-02	08-01	2	1.89	2	1983	07-03	08-02	2	1.63	1
1961	07-02	08-03	4	3.48	3	1984	08-01	08-03	1	2.84	3
1962	07-02	08-03	4	3.37	3	1985	08-02	09-01	2	2.00	2
1963	07-01	08-02	4	6.31	5	1986	06-03	07-01	4	1.43	1
1964	07-02	09-03	7	5.13	4	1987	08-02	09-01	2	1.89	2
1965	07-01	07-02	1	1.00	1	1988	07-01	08-03	3	3.86	3
1966	07-02	09-01	5	3.61	3	1989	07-02	08-01	2	1.90	2
1967	07-02	08-03	3	2.91	3	1990	07-01	08-01	2	2.49	2
1968	07-02	07-03	1	1.03	1	1991	07-02	08-01	3	2.16	2
1969	07-01	09-01	4	4.12	3	1992	07-03	08-02	4	1.64	1
1970	07-02	08-02	3	2.60	2	1993	07-01	08-02	4	2.84	3
1971	06-03	08-01	3	3.12	3	1994	07-01	08-02	2	3.96	3
1972	07-02	08-01	2	1.82	2	1995	07-02	08-03	1	3.46	3
1973	07-01	09-01	5	4.59	4	1996	07-01	08-02	4	4.29	3

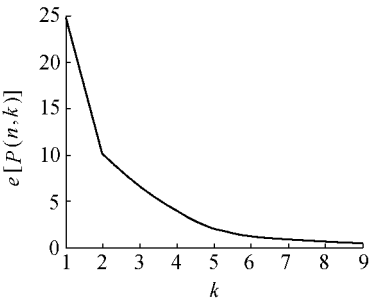


图1 $e[P(n,k)]$ 关于 k 的图像

Fig.1 $e[P(n,k)]\sim k$ curve

表2 华北地区雨季强度指数分类

Table 2 Classification of rainfall intensity index for North China

状态	级别	区间
5	强	$x > 5.160$
4	偏强	$4.355 \leq x < 5.160$
3	一般	$2.645 \leq x < 4.355$
2	偏弱	$1.815 \leq x < 2.645$
1	弱	$x < 1.815$

2 雨季强度指数预测的基本步骤

- 本文利用叠加马尔可夫链预测雨季强度指数所处状态,其基本步骤^[3-7]如下:
- a. 将华北地区雨季强度指数序列由小到大排列后用有序聚类法生成强度指数的最优分级标准.
 - b. 按所生成的分级标准确定各年雨季强度指数所处的状态,对应的状态空间为 I ,并求出雨季强度指数时间序列所对应的状态序列.
 - c. 对华北地区雨季强度指数序列进行马氏性检验.设 $(f_{ij})_{m \times m}$ 为1步转移频率矩阵,则

$$p_{\cdot j} = \frac{\sum_{i=1}^m f_{ij}}{\sum_{i=1}^m \sum_{j=1}^m f_{ij}} \text{ 且 } p_{ij} = \frac{f_{ij}}{\sum_{j=1}^m f_{ij}}$$

(2)

当 n 较大时,统计量

$$\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \ln \frac{p_{ij}}{p_{\cdot j}} \right| \sim \chi^2 [(m-1)^2] \quad (3)$$

对于给定的显著性水平 α ,查表可得分位点 $\chi^2_{\alpha} [(m-1)^2]$ 的值.若计算所得的统计量 χ^2 的观测值大于分位点的值,则认为该序列符合马氏性,从而可以用马氏链预测方法进行预测.

d. 按步骤 b 所得到的状态序列计算出不同步长的马尔可夫链的 1 步转移概率矩阵,再用 SPMCP 预测出该年雨季强度指数所处的状态.

e. 各种步长的马尔可夫链的特征分析.本文以相依性最强的步长为 1 的马尔可夫链的特征分析为例,得出了极限分布及各状态的复现周期.

3 实例分析

以承德、北京、天津、保定、德州、安阳和新乡 7 个代表站的降水量强度指数资料^[2]的分析预测为例来说明叠加马尔可夫链预测方法的预测效果.所取数据状态见表 1,华北地区雨季强度指数的分级标准和状态所代表的区间见表 2.

取显著性水平为 0.05,查表得 $\chi^2_{0.05}(16) = 26.296$.统计量的观测值为 31.545,大于 26.296,符合马氏性,故可以利用马氏链预测方法进行预测分析.

本文分别根据 1990~1994 年和 1991~1995 年华北地区雨季强度指数状态及其相应的状态转移概率矩阵对 1995 年和 1996 年华北地区雨季强度指数状态进行了预测,结果见表 3.

表 3 1995 年和 1996 年华北地区雨季强度指数状态预测结果

Table 3 Predicted results of state of rainfall intensity index for North China in 1995 and 1996

预测年份	初始年	滞时(年)	状态	状 态					概率来源
				1	2	3	4	5	
1995	1994	1	3	0.125 0	0.250 0	0.375 0	0.125 0	0.125 0	$P^{(1)}$
	1993	2	3	0.266 7	0.133 3	0.266 7	0.266 7	0.066 7	$P^{(2)}$
	1992	3	1	0.375 0	0.250 0	0.250 0	0.125 0	0	$P^{(3)}$
	1991	4	2	0.111 1	0.222 2	0.444 4	0.222 2	0	$P^{(4)}$
	1990	5	2	0.250 0	0.250 0	0.250 0	0.125 0	0.125 0	$P^{(5)}$
1996	1995	1	3	0.117 6	0.235 3	0.411 8	0.117 6	0.117 6	$P^{(1)}$
	1994	2	3	0.250 0	0.125 0	0.312 5	0.250 0	0.062 5	$P^{(2)}$
	1993	3	3	0.133 3	0.333 3	0.333 3	0.133 3	0.066 7	$P^{(3)}$
	1992	4	1	0	0.500 0	0.375 0	0	0.125 0	$P^{(4)}$
	1991	5	2	0.222 2	0.222 2	0.333 3	0.111 1	0.111 1	$P^{(5)}$

由表 3 可知,系统在 1995 年和 1996 年转移到各状态的概率和向量分别为

$$W = (0.127\ 8\ 0.105\ 6\ 0.586\ 1\ 0.863\ 9\ 0.316\ 7)$$
$$W = (0.723\ 2\ 0.415\ 8\ 0.765\ 9\ 0.612\ 1\ 0.482\ 9)$$

由此可知,系统 1995 年和 1996 年都是处于状态 3 的概率最大.因此,1995 年和 1996 年华北地区雨季强度指数的预测状态均为 3.这与实际相符(1995 年的实际雨季强度指数为 3.46,1996 年的实际强度指数为 4.29,均处于状态 3).

综上所述,将叠加马尔可夫链预测方法用于华北地区雨季强度指数状态预测是可行和有效的.

为进一步了解序列特性,现进行马尔可夫链的特征分析.由于本例各步标准化自相关系数向量 $W = (0.407\ 49\ 0.096\ 611\ 0.165\ 03\ 0.068\ 267\ 0.262\ 6)$,本文以相依性最强的步长为 1 的马尔可夫链的特征分析为例.由 1 步转移概率矩阵可知,该链是遍历的(非周期、不可约、正常返).因此,该链存在唯一的平稳分布,且此时的平稳分布即为其极限分布.

设该链的平稳分布为 $\{\pi_j\ j \in I\}$ 则

$$\sum_{j \in I} \pi_j = 1 \qquad \pi_j = \sum_{i \in I} \pi_i p_{ij} \quad (j \in I) \qquad (4)$$

参照步长为 1 的转移概率矩阵可得各状态的重现期,如表 4 所示.

表 4 平稳分布和各状态的重现期
Table 4 Stationary distribution and recurrence period of various states

状态(j)	π_j	$\mu_j = 1/\pi_j$
1	0.204 0	4.902 9
2	0.209 6	4.770 4
3	0.394 3	2.535 9
4	0.119 0	8.404 5
5	0.073 1	13.681 9

表4结果是符合客观规律的,即中间状态比较容易出现,而比较极端的情况出现的时间间隔相对较长.例如,状态5在整个序列中出现的次数较少.

4 结 束 语

SPMCP是一种行之有效的预测方法,只要一个时间序列满足“马氏性”,就可以用叠加马氏链的预测方法对其进行处理,而且能得出令人满意的预测结果.但在某些场合下,人们需要对所研究的对象做出点值预测.对于这种情况,时间序列分析理论提供了多种线性模型.可以利用某种线性模型进行点值预测,也可以利用某种线性模型与叠加马氏链预测方法相结合的综合方法进行预测.

参考文献:

[1]王玲玲,周纪芾.常用统计方法[M].上海:华东师范大学出版社,1994:250-253.
[2]赵振国.中国夏季旱涝及环境场[M].北京:气象出版社,1999:13-16.
[3]夏乐天.马尔可夫链预测方法及其在水文序列中的应用研究[D].南京:河海大学,2005.
[4]孙才志,张戈,林学钰.加权马尔可夫链在降水丰枯状况预测中的应用[J].系统工程理论与实践,2003(4):100-105.
[5]张宗国.马尔可夫链在太湖流域梅雨期预报中的应用[J].河海大学学报:自然科学版,2006,34(1):116-118.
[6]顾岚.时间序列分析在经济中的应用[M].北京:中国统计出版社,1998:78-184.
[7]王艳明,许启发.时间序列分析在经济预测中的应用[J].统计与预测,2001(5):32-34.

SPMCP model and its application to prediction of rainfall intensity index for rainy season

SHEN Yong-mei , XIA Le-tian

(College of Sciences , Hohai University , Nanjing 210098 , China)

Abstract :In consideration of the uncertainty and inaccuracy of precipitation process in the rainy season , a classification standard for rainfall intensity index of the rainy season was established by use of the sequential cluster method . Based on state of rainfall intensity index for the rainy season of North China from 1951 to 1994 , the state of the index of North China for 1995 and 1996 was predicted by use of SPMCP model . The results show that SPMCP model is simple and accurate in calculation , and it provides a new way to predict the state of rainfall intensity index for the rainy season .

Key words :North China ; rainfall intensity index of rainy season ; Markov chain model ; sequential cluster method