

doi: 10.3969/j.issn.2095-1914.2014.03.012

## 加权马尔科夫链在福建省森林火灾预测中的应用研究

曹彦 何东进 洪伟 纪志荣 朱学平 游巍斌 连素兰

(福建农林大学林学院 福建 福州 350002)

**摘要:**以福建省1987—2008年森林火灾发生次数的数据资料为基础,运用加权马尔科夫链预测方法建立预测模型,对2009年福建省森林火灾发生次数进行预测,并利用该方法对2007—2008年森林火灾发生次数进行检验。结果表明,预测结果与实际结果高度吻合,而且预测精度较高,都大于85%,效果较好,说明加权马尔科夫链方法可以用于短期森林火灾预测。

**关键词:**加权马尔科夫链;森林火灾;预测;应用;福建省

中图分类号: S762.2

文献标志码: A

文章编号: 2095-1914(2014)03-0062-05

## Application of Weighted Markov-Chain in Prediction of Forest Fire Disaster in Fujian Province

CAO Yan ,HE Dong-jin ,HONG Wei ,JI Zhi-rong ,ZHU Xue-ping ,YOU Wei-bin ,LIAN Su-lan

( College of Forestry ,Fujian A&F University ,Fuzhou Fujian 350002 ,China)

**Abstract:** Based on the data for 1987 - 2008 of the forest fire disaster in Fujian province ,the prediction model was established by using the weighted Markov-chain prediction method , and applied the prediction model to predicting the number of the forest fire disaster in 2009. Then use this method to validate the data for 2007 - 2008. The result showed that the predicted result was highly consistent with the actual result ,and the forecasting precision was more than 85% ,which indicated that the weighted Markov-chain prediction method could be used in short-term prediction of the forest fire disaster.

**Key words:** weighted Markov-chain; forest fire disaster; forecast; application; Fujian Province

福建省地处亚热带,自然条件优越,十分适合林木生长,森林覆盖率为60%左右,是我国主要的林区之一。但由于气候变化、人类破坏和人为干扰等因素的影响,福建省森林火灾发生次数频繁且危害损失严重,因此,对其森林火灾发生规律及预测预报进行研究有着重要的意义。目前,国内外学者已经对森林火灾开展了大量的研究工作,如张典铨<sup>[1]</sup>以福建省建阳市森林火灾资料为例建立了森林火灾的灰色拓扑预测模型;傅泽强等<sup>[2]</sup>对内蒙古大兴安岭林区森林火灾发生序列建立了灰色灾变

GM(1,1)预测模型;姜学鹏等<sup>[3]</sup>建立了我国火灾相对变动模糊马尔科夫预测模型;李炳华等<sup>[4]</sup>提出了基于小样本数据信息扩散的重大火灾频度估算方法;朱学平等<sup>[5]</sup>运用突变级数法对福建省森林火灾损失评价指标体系进行研究;谭三清等<sup>[6]</sup>基于灰色联度方法研究分析了森林火灾危害程度及其相关影响因子。我国学者对火灾预测的方法主要集中于定性预测法、灰色预测法、人工神经网络预测法和回归预测法等,在对森林火灾预测研究中,运用加权马尔科夫链方法开展预测预报的研究则未见

收稿日期: 2013-12-11

基金项目: 国家自然科学基金项目(31370624、41301203、30870435)资助;教育部博士学科点专项基金项目(20103515110005、20133515120007)资助;福建农林大学林学院青年科学基金资助项目(6112C035F)资助。

第1作者:曹彦(1990—),女,硕士生。研究方向:资源调查与统计。Email:1304086643@qq.com。

通信作者:何东进(1969—),男,教授,博士生导师。研究方向:统计学与生态学。Email:fjhdj1009@126.com。

报道。为此,本项研究以前人相关研究为基础,采用加权马尔科夫链的方法对福建省森林火灾发生次数进行预测,以期对森林火灾预测方法提供新的思路 and 理论依据。

## 1 研究数据来源

研究数据来源于《福建省统计年鉴》<sup>[7]</sup>,相关资料见表1。从表1数据可以看到,福建省1987—2009年发生森林火灾总次数为12 007次,其中1987年、1988年和2004年3年发生次数均超过1 000次,1988年达到最高为1 344次,而1998年火灾发生次数最少,仅有156次。

表1 福建省1987—2009年森林火灾发生次数统计

年份	火灾次数/次	年份	火灾次数/次
1987	1 042	1999	318
1988	1 344	2000	915
1989	845	2001	247
1990	325	2002	478
1991	344	2003	522
1992	613	2004	1 164
1993	488	2005	309
1994	329	2006	158
1995	434	2007	237
1996	473	2008	520
1997	167	2009	579
1998	156		

## 2 研究方法

马尔科夫链是一个时间连续状态离散的马尔可夫过程,它具有“马氏性”,其转移概率与初始时刻状态无关,则称该马尔科夫链具有齐次性。传统的马尔科夫链预测法<sup>[8-9]</sup>,一般默认所研究的马尔科夫链满足“齐次性”,但实际研究中所论及的随机变量序列虽然满足“马氏性”,但一般都不满足“齐次性”。另外该方法没有充分利用已知历史数据的信息,忽略了各阶马尔科夫链的绝对分布在预测中所起的作用。而加权马尔科夫链<sup>[10-11]</sup>方法弥补了传统马尔科夫链方法的不足,根据各阶自相关系数能够刻画各种滞时状态间相关关系强弱的原理,先分别依据前面若干时段的指标值状态对该时段指标值的状态进行预测,然后按照各阶自相关系数进行加权求和,

最后按照“概率最大”原则判断该时段所属状态。

### 2.1 系统状态的确定

确定系统状态即对指标值进行分级,目前运用最广泛的是均值-标准差法<sup>[12]</sup>。

设指标值序列为  $x_1, x_2, \dots, x_n$ , 样本均值为  $\bar{x}$ , 样本标准差为:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

一般将指标值的变化区间表示为  $(-\infty, \bar{x} - \sigma_1 s)$ ,  $(\bar{x} - \sigma_1 s, \bar{x} - \sigma_2 s)$ ,  $(\bar{x} - \sigma_2 s, \bar{x} + \sigma_2 s)$ ,  $(\bar{x} + \sigma_2 s, \bar{x} + \sigma_1 s)$ ,  $(\bar{x} + \sigma_1 s, +\infty)$ 。  $\sigma_1$  和  $\sigma_2$  的取值范围一般为  $[1.0, 1.5]$  和  $[0.3, 0.6]$ 。

### 2.2 马尔科夫转移概率矩阵的计算

根据分级标准确定各时段指标值所对应的状态,建立转移概率矩阵。设所研究的马尔科夫链的一个指标值序列包含了  $m$  个状态,即状态空间  $E = \{1, 2, \dots, m\}$ ; 用  $f_{ij}$  表示指标值序列中从状态  $i$  经过一步转移到状态  $j$  的频数,  $i, j \in E$ , 转移的步长为  $m$  ( $m=1, 2, \dots$ ) 个时间单位。由  $f_{ij}$  ( $i, j \in E$ ) 组成的矩阵  $(f_{ij})_{i, j \in E}$  称为转移频数矩阵。将转移频数矩阵的第  $i$  行第  $j$  列元素除以各行的总和所得的值称为转移概率,记为  $p_{ij}$  ( $i, j \in E$ ), 即

$$p_{ij} = \frac{f_{ij}}{\sum_{j=1}^m f_{ij}} \quad (1)$$

则转移概率矩阵可以表示为:

$$P = (p_{ij})_{i, j \in E} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

### 2.3 随机变量序列的马氏性检验

检验随机变量序列是否具有马氏性是应用马尔科夫链模型进行预测的重要前提,通常用  $\chi^2$  统计量进行检验。将转移频数矩阵的第  $j$  列之和除以各行各列的总和所得的值称为边际概率,记为  $p_{\cdot j}$ , 即:

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

当  $n$  充分大时,统计量  $\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \lg \frac{p_{ij}}{p_{\cdot j}} \right|$  服

从自由度为  $(m-1)^2$  的  $\chi^2$  分布。给定显著性水平  $\alpha$ , 查表可得分位点  $\chi_{\alpha}^2 ((m-1)^2)$  的值,若  $\chi^2 > \chi_{\alpha}^2$

$((m-1)^2)$ , 则可以认为序列符合马氏性, 否则不能对该序列用马尔科夫链来处理<sup>[13]</sup>。

## 2.4 计算各阶自相关系数

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

式中:  $r_k$  为滞后  $k$  ( $k \in E$ ) 阶时的自相关系数;  $x_i$  为指标值;  $\bar{x}$  为指标均值;  $n$  为指标序列的个数。

对各阶自相关系数进行归一化, 即:

$$W_k = |r_k| / \sum_{k=1}^m |r_k| \quad (3)$$

将  $W_k$  作为各种步长的马尔科夫链的权重 ( $m$  为按时间需要计算到的最大阶数)。

## 2.5 进行加权马尔科夫链预测

分别以前面若干时段的指标值为初始状态, 结

合其相应的各阶转移概率矩阵即可预测出该时段指标值的状态概率  $p_i^{(k)}$  ( $i \in E$ )  $k=1, 2, \dots, m$ 。将同一状态各预测概率加权和作为指标处于该状态的预测概率, 即:

$$p_i = \sum_{k=1}^m W_k p_i^{(k)} \quad (i \in E) \quad (4)$$

$\max(p_i, i \in E)$  所对应的  $i$  即为该时段指标值的预测状态。将预测结果加入初始状态向量序列得到新的序列, 可重复以上过程预测较长时间间隔的状态。

## 3 结果与分析

以福建省 1987—2008 年森林火灾数据为基础, 其变化趋势见图 1。

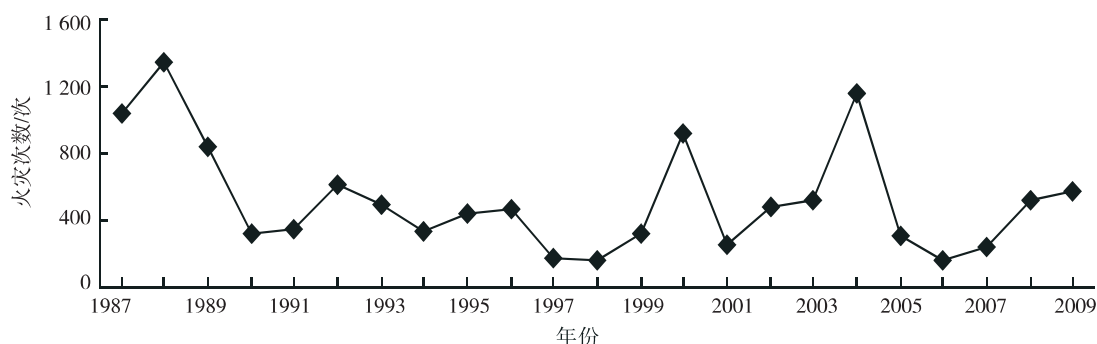


图1 福建省1987—2009年森林火灾发生次数趋势

从图 1 中可以看出, 福建省森林火灾发生次数变化很不稳定, 出现了 3 个峰值, 1988—1999 年间林火次数呈明显下降趋势。利用加权马尔科夫链预测模型对 2009 年森林火灾发生次数进行预测, 步骤如下。

1) 根据均值标准差法将森林火灾次数划分为 5 个状态, 即确定马尔科夫链的状态空间  $E = \{1, 2, 3, 4, 5\}$ 。分级情况见表 2, 各年所属状态见表 3。

2) 根据(1) 式计算得到步长从 1 到 5 的一步转移概率矩阵如下:

表2 福建省森林火灾数据分级

状态	级别	火灾次数区间
1	轻微灾年	$0 \leq x < 182.63$
2	一般灾年	$182.63 \leq x < 351.04$
3	较大灾年	$351.04 \leq x < 687.87$
4	重大灾年	$687.87 \leq x < 957.33$
5	特大灾年	$957.33 \leq x < 1400$

表3 福建省 1987—2008 年森林火灾发生次数序列及其状态

年份	林火次数/次	状态等级
1987	1 042	5
1988	1 344	5
1989	845	4
1990	325	2
1991	344	2
1992	613	3
1993	488	3
1994	329	2
1995	434	3
1996	473	3
1997	167	1
1998	156	1
1999	318	2
2000	915	4
2001	247	2
2002	478	3
2003	522	3
2004	1 164	5
2005	309	2
2006	158	1
2007	237	2
2008	520	3

$$P_1 = \begin{bmatrix} 0.333\ 3 & 0.666\ 7 & 0 & 0 & 0 \\ 0.142\ 9 & 0.142\ 9 & 0.571\ 4 & 0.142\ 9 & 0 \\ 0.166\ 7 & 0.166\ 7 & 0.5 & 0 & 0.166\ 7 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$
$$P_2 = \begin{bmatrix} 0 & 0.333\ 3 & 0.333\ 3 & 0.333\ 3 & 0 \\ 0 & 0.333\ 3 & 0.666\ 7 & 0 & 0 \\ 0.4 & 0.2 & 0.2 & 0 & 0.2 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0.333\ 3 & 0.333\ 3 & 0 & 0.333\ 3 & 0 \end{bmatrix}$$
$$P_3 = \begin{bmatrix} 0 & 0.5 & 0 & 0.5 & 0 \\ 0.166\ 7 & 0.166\ 7 & 0.666\ 7 & 0.5 & 0.166\ 7 \\ 0.333\ 3 & 0.333\ 3 & 0.333\ 3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$
$$P_4 = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0.333\ 3 & 0.333\ 3 & 0.166\ 7 & 0.166\ 7 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0.333\ 3 & 0.666\ 7 & 0 & 0 \end{bmatrix}$$

$$P_5 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0.2 & 0.2 & 0.4 & 0 & 0.2 \\ 0.333\ 3 & 0.333\ 3 & 0.166\ 7 & 0.166\ 7 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

结合步长为 1 的一步转移概率矩阵对火灾发生次数序列进行马氏性检验。由(2)式计算得火灾次数序列统计量 $\chi^2 = 34.52 > \chi^2_{\alpha}(16) = 32.00$ ,故该序列通过马氏性检验,可以对其进行马尔科夫链预测。

3) 利用(3)式和(4)式计算各阶自相关系数得:

$$r = (0.320\ 0, -0.107\ 1, -0.191\ 5, 0.226\ 5, -0.045\ 3)$$

标准化后可得各步长马尔科夫链的权重向量为:

$$W = (0.359\ 6, 0.120\ 3, 0.215\ 1, 0.254\ 2, 0.050\ 8)$$

4) 依据 2004、2005、2006、2007、2008 年的指标值所属状态及其相对应的状态转移概率矩阵,对福建省 2009 年森林火灾发生次数状态进行预测,结果见表 4。

表 4 2009 年森林火灾发生次数状态预测

初始年份	初始状态	滞时年	权重	状态 1	状态 2	状态 3	状态 4	状态 5
2004	5	5	0.050 8	0	0	1.000 0	0	0
2005	2	4	0.254 3	0.200 0	0.400 0	0.400 0	0	0
2006	1	3	0.215 1	0	0.500 0	0	0.500 0	0
2007	2	2	0.120 3	0	0.333 0	0.667 0	0	0
2008	3	1	0.359 4	0.166 7	0.166 7	0.500 0	0	0.166 7
$P_i$				0.110 8	0.309 3	0.412 5	0.107 5	0.059 9

由表 4 可以看出,将同一状态各预测概率加权求和后 $\max\{P_i, i \in E\} = 0.412\ 5$ ,即 2009 年福建省森林火灾发生次数预测状态是 3,说明 2009 年属于福建省发生森林火灾一般灾年,发生次数应该落在 $351.04 \leq x < 687.87$ 范围内,取 351.04 和 687.87 的平均值得到 519.45,即预计 2009 年福建省发生森

林火灾次数为 519.45 次。而 2009 年的实际火灾发生次数为 579 次,预测值与真实值的相对误差为 $(579 - 519.45) / 579 = 0.102\ 8$ ,模型的预测精度达到 90%。

利用该方法对福建省 2007 和 2008 年森林火灾发生次数状态进行回验,所得结果见表 5。

表 5 2007—2009 年福建省森林火灾发生次数状态预测结果

年份	真实状态	预测状态	属于状态 1 概率	属于状态 2 概率	属于状态 3 概率	属于状态 4 概率	属于状态 5 概率
2007	2	2	0.321 2	0.415 6	0.190 0	0.073 2	0.000 0
2008	3	3	0.080 3	0.321 3	0.414 4	0.130 2	0.053 8
2009	3	3	0.110 8	0.309 3	0.412 5	0.107 5	0.059 9

从表5可以看出,2007—2009年的预测结果与真实结果完全吻合。且经计算得到2007、2008年预测值与真实值的相对误差分别为 $|237 - 272.68| / 237 = 0.1505$ 和 $|520 - 519.43| / 520 = 0.0011$ ,预测精度都大于85%。这不仅说明加权马尔科夫链模型可以用于对森林火灾发生次数的预测,而且模型预测的准确度和精度都较高,预测效果较好。

#### 4 小结与讨论

根据森林火灾发生次数序列突变性强、无后效性的特点,对其进行马尔科夫链预测是一个大胆的尝试。研究表明,加权马尔科夫链方法在森林火灾预测领域是可行的,且预测效果较为理想。其中,应用均值-标准差法对原序列指标状态进行划分,充分利用了森林火灾次数序列的数据结构,能够有效刻画序列的内在分布规律,并且客观反映每年森林火灾的严重程度,使得划分的状态级别较为科学合理。预测结果从森林火灾发生次数的一个状态区间通过区间端点的均值转化为一个具体数值,在保证可靠性程度的情况下使模型的预测精度得到了提高。

加权马尔科夫链以各种步长的自相关系数为权重,使相关分析和马尔科夫链有机的结合起来进行预测,与普通马尔科夫链预测相比更充分合理地利用了相关信息,所以预测效果更好。该模型比较适用于短期预测,且历史资料数据越多,预测精度越高。若用于长期预测,则可将预测年的预测状态和预测值计入原数据序列重新构置马尔科夫转移概率矩阵进行逐年预测。

鉴于森林火灾发生次数受多方面因素影响,如天气、气候等自然因素,以及当地的林火管理水平和政府重视程度等人为因素,因此林火发生次数序列波动性较大且极不稳定,偶尔会出现极端值或者异常值。而马尔科夫链预测仅仅是根据历史火灾发生次数进行建模,没有将其他影响因素进行综合考虑,所以对于一些异常年份火灾的预测可能会出现偏差。这也是该方法所存在的局限性,在今后的研究中将继续对此方法进行探讨与改进。

#### [参 考 文 献]

- [1] 张典铨. 灰色拓扑预测方法在森林火灾预测中的应用[J]. 福建林学院学报, 2005, 25(1): 1-4.
- [2] 傅泽强, 孙启宏, 蔡运龙, 等. 基于灰色系统理论的森林火灾预测模型研究[J]. 林业科学, 2002, 38(5): 95-100.
- [3] 姜学鹏, 徐志胜, 冷彬. 火灾预测的模糊马尔柯夫模型[J]. 灾害学, 2006, 21(3): 27-32.
- [4] 李炳华, 朱霁平, 小出治. 一种基于小样本数据信息扩散的重大火灾频度估算方法[J]. 火灾科学, 2010, 19(2): 82-88.
- [5] 朱学平, 何东进, 丁福利, 等. 基于突变级数法的福建省森林火灾评价分析[J]. 福建林学院学报, 2011, 31(4): 295-299.
- [6] 谭三清, 张贵, 刘大鹏. 基于灰色联度分析的森林火灾危害程度及相关因子研究[J]. 中南林业科技大学学报, 2010, 41(7): 10-14, 51.
- [7] 福建省统计局. 福建省统计年鉴(2009) [M]. 北京: 中国统计出版社, 2009: 67-98.
- [8] Sen Z. Critical drought analysis by second-order Markov chain[J]. Journal of Hydrology (Amsterdam), 1990, 120(1): 183-202.
- [9] Peng Z, Bao C, Zhao Y, et al. Weighted Markov chains for forecasting and analysis in incidence of infectious diseases in Jiangsu Province, China [J]. Journal of Biomedical Research, 2010, 24(3): 207-214.
- [10] 钱会, 李培月, 王涛. 基于滑动平均-加权马尔科夫链的宁夏石嘴山市年降雨量预测[J]. 华北水利水电学院学报, 2010, 31(1): 6-9.
- [11] 仲远见, 李靖, 王龙. 改进马尔科夫链降雨量预测模型的应用[J]. 济南大学学报: 自然科学版, 2009, 23(4): 402-405.
- [12] 金峻炎, 陈进. 加权马氏链在房地产投资回收期预测中的应用[J]. 湖南科技大学学报: 自然科学版, 2010, 25(2): 63-66.
- [13] 罗积玉, 邢瑛. 经济统计分析及预测[M]. 北京: 清华大学出版社, 1987: 347-348.

(责任编辑 韩明跃)