

模糊时间序列模型在短期气候预测中的应用

王永弟¹

摘要

将模糊时间序列模型引入短期气候预测,利用重庆34个地面气象观测站的逐日观测资料(1971—2007年)和重庆市旱涝灾害监测预警决策服务系统计算的干旱指数和洪涝指数等资料,运用模糊时间序列模型分别对2001—2007年重庆市城口县1月降水、1月平均气温的预报结果(年度预测)和重庆市春旱指数的预报结果(年度预测)进行了模糊时间序列分析,预测了2004—2007年的发展趋势,用2004—2007年实测值与预测结果进行了比较,并与加权集成、人工神经网络集成、数据挖掘集成等模型进行了精度比较和分析.结果表明:模糊时间序列模型各项精度评定指标优良,并且计算简单,具有一定的实用价值.

关键词

模糊时间序列;模糊技术;短期气候预测;预报模型;模型精度评定

中图分类号 P456

文献标志码 A

收稿日期 2011-10-04

资助项目 江苏省高校自然科学基金(1K-JB420002);南京信息工程大学科研基金(S811-0063001);南京信息工程大学科研基金(Y649)

作者简介

王永弟,男,博士,讲师,主要从事气候学研究. ydwang2003@163.com

0 引言

模糊技术正越来越引起各国科技界、工程界和企业界的高度重视.模糊数学在气象科学中的应用探索始于1979年^[1].楼世博等^[2]应用模糊聚类分析方法做了上海地区春季连阴雨短期预报;宗荣祥^[3]利用模糊向量运算,对内蒙古白云地区夏季总降水量做了长期预报;王式中^[4]用模糊聚类方法对年-月天气周期进行了划分;曹鸿兴等^[5]依据不同环流型的主要特征,建立隶属函数,然后根据最大隶属原则对天气图进行了直接识别;章文茜^[6]运用模糊数的概念、贴近度原理和模糊推理提出了分别适用于单站天气预报和区域预报的评分方案.实践证明:模糊子集论在气象科学中的应用是适宜的、有前途的,它的诞生为探索气象科学问题提供了一个新工具.

关于模糊时间序列的研究,Song等^[7]利用Zadeh的模糊集理论发展了模糊时间序列预测模型,并将其运用在预测Alabama大学的入学人数上;Sun等^[8]的模型提出了计算以绝对误差的平均数来划分区间,但是模型中没有考虑到历史数据的分布对于模糊集定义的影响;Chen等^[9]的模型中对模糊集的定义考虑了历史数据的分布,但是该模型在划分区间时没有考虑区间间隔对于预测精度的影响;Singh^[10]的模型中采用了差分算子来进行预测,但是对于区间间隔仍然是简单的均分;吴铭峰等^[11]的模型吸取了Sun模型和Chen模型的优点,预测精度再次提高;张钰敏等^[12]改进了现有的模糊时间序列模型,在模糊集的定义上使用Chen的方法,在区间间隔的划分上采用Sun的方法,最后用Singh的预测规则进行预测,综合了3个模型的优点,模型更简便且易于计算,并再次提高了预测的精度.本文在此基础上,将模糊时间序列预测模型应用在短期气候预测中,然后通过实例数据进行计算验证,并与其他方法进行了预测精度的比较,最后做出模型精度的评定.

1 模糊时间序列模型简介

1.1 模糊时间序列的定义

Song等^[7]在1993年提出了基于模糊集的模糊时间序列定义,该定义如下:定义 U 为论域,给定 U 的一个次序分割集为 $U = \{u_1, u_2, \dots, u_n\}$,定义 A 为论域 U 上的语义变量集,并记为

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n, \quad (1)$$

¹ 南京信息工程大学 遥感学院,南京,210044

其中, f_A 是定义在 A 上的隶属函数, $f_A: U \rightarrow [0, 1]$, u_k 是模糊集的 A_i 元素, 表示 u_k 到 A_i 的隶属度, 并且 $f_A(u_i) \in [0, 1], 1 \leq i \leq n$.

设 $X(t)$ 是 \mathbf{R} 的子集, 并且 $X(t)$ 是由模糊集 $f_i(t)$ 所定义的论域, 如果 $F(t)$ 由 $f_1(t), f_2(t), \dots$ 构成, 则 $F(t)$ 是定义在 $X(t) (t = 1, 2, \dots)$ 上的模糊时间序列.

1.2 模糊时间序列预测的基本步骤

首先定义论域并进行区间的划分, 然后定义论域上的模糊集和相应的模糊语义变量, 从而使历史数据模糊化, 接着建立模糊关系, 最后通过一定的方法预测并去模糊化. 模糊时间序列预测模型的预测步骤如下^[11-12].

1) 定义论域并划分区间. 论域 $U = [D_{\min} - \sigma_1, D_{\max} + \sigma_2]$, 其中, D_{\max}, D_{\min} 分别为历史数据的最大值和最小值, σ_1, σ_2 为两个合适的正数 (即不足整数的部分), 然后计算所有相邻两个数的差的绝对值, 再求这些绝对值的平均数, 以这个平均数的一半作为长度, 并根据表 1 规定的范围进行取整运算, 确定区间长度.

表 1 取整映射表^[8]

Table 1 The rounded down mapping table^[8]

范围	0.1 ~ 1	1 ~ 10	10 ~ 100	100 ~ 1000
取整	0.1	1	10	100

2) 统计历史数据, 根据历史数据的分布再次划分区间. 如果一个区间中有 m 个历史数据, 则将这个区间 m 等分. 这样, 对于 n 个数据, 就得到了 n 个子区间 u_1, u_2, \dots, u_n , 且每一个历史数据都与一个区间对应, 容易看出各子区间长度不一, 这反映了历史数据在论域上的不均匀分布.

3) 定义模糊集并模糊化数据. 根据第 2 步所得到的区间定义模糊集 $A = \{A_1, A_2, \dots, A_n\}$, 其中 A_i 表示模糊集 A 的语义变量. 将历史数据按下式模糊化:

$$\begin{cases} A_1 = \frac{f_{11}}{u_1} + \frac{f_{12}}{u_2} + \dots + \frac{f_{1n}}{u_n}, \\ A_2 = \frac{f_{21}}{u_1} + \frac{f_{22}}{u_2} + \dots + \frac{f_{2n}}{u_n}, \\ \vdots \\ A_n = \frac{f_{n1}}{u_1} + \frac{f_{n2}}{u_2} + \dots + \frac{f_{nn}}{u_n}. \end{cases} \quad (2)$$

4) 建立模糊逻辑关系并预测. 经过模糊化后, 如果第 n 月的模糊隶属为 A_i , 第 $(n+1)$ 月的模糊隶

属为 A_j , 那么它们的模糊逻辑就为 $A_i \rightarrow A_j$.

预测规则中使用的符号如下:

$[A_j^*]$ —— 模糊集 A_j 中隶属度为 1 的区间 u_j ;

$U[A_j^*]$ —— 区间 u_j 的上边界;

$L[A_j^*]$ —— 区间 u_j 的下边界;

$l[A_j^*]$ —— 区间 u_j 的长度;

$M[A_j^*]$ —— 区间 u_j 的中点值.

对于模糊逻辑 $A_i \rightarrow A_j$, A_i 和 A_j 分别表示第 n 月和第 $(n+1)$ 月的模糊化数据, E_i 表示第 n 月的实际数据, F_j 表示第 $(n+1)$ 月的预测值.

文献[12]采用的是三阶模型, 即用第 $(n-2)$ 月、第 $(n-1)$ 月和第 n 月的数据来预测第 $(n+1)$ 月的数据. 预测规则如下. 首先计算如下数据:

$$D_i = \|E_i - E_{i-1}\| - \|E_{i-1} - E_{i-2}\|, \quad (3)$$

$$X_i = E_i + D_i/2, \quad (4)$$

$$XX_i = E_i - D_i/2, \quad (5)$$

$$Y_i = E_i + D_i, \quad (6)$$

$$YY_i = E_i - D_i. \quad (7)$$

然后进行去模糊化, 用于数据累加的临时变量 P, S 的取值按照表 2 第一列各项 (X_i, XX_i, Y_i, YY_i) 是否在 $[A_j^*]$ 中分别进行判断.

表 2 P, S 的取值判断

Table 2 Value determination of variable P and S

名称	在 $[A_j^*]$ 中		不在 $[A_j^*]$ 中	
	P	S	P	S
X_i	X_i	1	0	0
XX_i	$P + XX_i$	$S + 1$	$P + 0$	$S + 0$
Y_i	$P + Y_i$	$S + 1$	$P + 0$	$S + 0$
YY_i	$P + YY_i$	$S + 1$	$P + 0$	$S + 0$

最后, 将 $M[A_j^*]$ 和临时变量 P, S 的值代入式 (8) 得出第 $(n+1)$ 月的预测值:

$$F_j = (P + M[A_j^*]) / (S + 1). \quad (8)$$

1.3 模型精度评定

1) 方法 1. 残差大小检验, 即用均方误差 (MSE)、平均误差 (记为 Δ) 2 个指标来进行评价^[12].

记实测值和预测值分别为 y_i 和 \hat{y}_i , 残差 $e_i = |y_i - \hat{y}_i|$, 均方误差和平均误差分别为

$$E_{MS} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} = \frac{\sum e_i^2}{n}, \quad (9)$$

$$\Delta = \frac{\sum_{i=1}^n |[(y_i - \hat{y}_i)/y_i] \times 100\%|}{n} =$$

$$\frac{\sum (e_i/y_i) \times 100\%}{n} \quad (10)$$

2) 方法 2. 用后验差比值 C 和小误差概率 P 共同描述^[3].

记原始数列及残差的方差分别为 S_1^2, S_2^2 , 则后差比值为

$$C = S_2/S_1. \quad (11)$$

小误差概率为

$$P = \{ |e(k)| < 0.6745S_1 \}, \quad k = 1, 2, \dots, n. \quad (12)$$

模型精度等级判别式^[13]为

模型精度等级 = Max { P 所在的级别, C 所在的级别 }.

表 3 模型精度等级

Table 3 Model accuracy grades description

模型精度等级	P	C
1 级(好)	$0.95 \leq P$	$C \leq 0.35$
2 级(合格)	$0.80 \leq P < 0.95$	$0.35 < C \leq 0.5$
3 级(勉强)	$0.70 \leq P < 0.80$	$0.5 < C \leq 0.65$
4 级(不合格)	$P < 0.70$	$0.65 < C$

2 应用实例

本例取自文献[14]中的应用实例. 实验中各种算法所使用的输入数据是 74 项环流特征量资料(北半球副高面积指数、北非副高面积指数、北半球副高强度指数等), 是由国家气候中心气候系统诊断预测室再处理资料, 起始时间是 1951—2007 年, 资料数据全都为整型. 实验中各种算法的预测对象为: 降水、平均气温、5 种干旱指数(春旱、夏旱、伏旱、秋旱、冬旱)

和洪涝指数. 降水、平均气温等要素的数据来自于重庆 34 个地面气象观测站的逐日观测资料(1971—2007 年), 各种干旱指数和洪涝指数来自于重庆市旱涝灾害监测预警决策服务系统的计算结果.

运用本文所介绍的方法, 分别对 2001—2007 年重庆市城口县 1 月的降水预报结果(年度预测)、城口县 1 月的平均气温的预报结果(年度预测)、重庆市春旱指数的预报结果(年度预测)进行了模糊时间序列分析, 预测了 2004—2007 年的发展趋势, 并用 2004—2007 年实测值与预测结果进行比较. 计算结果见表 4—6.

从表 4—6 的计算结果来看, 在精度上基本与加权集成、神经网络集成、数据挖掘集成等集成模型精度相当.

模糊时间序列模型的优点如下.

1) 精度较高. 在各项精度指标中均表现良好, 预报精度与集成预报基本相当.

2) 计算量小. 适用于小样本数据, 时间序列中仅有 3 个以上数据便能进行预报, 从计算量的角度来看也是相当优越的.

本模型的评级结果显得有些不合理, 主要是因为样本数量比较小, 需要更多的数据来进行模型等级的评定; 另外, 该模型仅适用于气候的短期预测, 即只能进行一步预测, 对于更长时间的预测, 还需要结合其他方法进行综合预测.

总之, 本文模型不但能够得到较高的预报精度, 而且计算简便, 是一种很好的短期预报方法, 非常实用.

表 4 城口 1 月降水预报结果对比

Table 4 Precipitation forecast result comparison of Chengkou city in January

年份	实际值/ mm	加权集成							神经网络集成						
		预报值/ mm	绝对误差/mm	精度评级				预报值/ mm	绝对误差/mm	精度评级					
				MSE	Δ	C	P 等级			MSE	Δ	C	P 等级		
2004	0.1	0.450	0.350					0.447	0.347						
2005	0.3	0.394	0.094	0.113	1.214	0.547	0 3 级	0.430	0.130	0.106	1.150	0.515	0 3 级		
2006	0.3	0.540	0.240					0.415	0.115						
2007	0.6	0.455	0.145					0.411	0.189						
年份	实际值/ mm	数据挖掘集成							模糊时间序列模型						
		预报值/ mm	绝对误差/mm	精度评级				预报值/ mm	绝对误差/mm	精度评级					
				MSE	Δ	C	P 等级			MSE	Δ	C	P 等级		
2004	0.1	0.147	0.047					0.291 7	0.192						
2005	0.3	0.317	0.017	0.091	0.224	0.443	50% 2 级	0.295 0	0.195	0.091	0.529	0.441	50% 2 级		
2006	0.3	0.309	0.009					0.298 3	0.002						
2007	0.6	0.396	0.204					0.495 0	0.195						

表5 城口1月平均气温预报结果对比

Table 5 Average temperature forecast result comparison of Chengkou city in January

年份	实际值/ ℃	加权集成							神经网络集成						
		预报值/ ℃	绝对误差/ ℃	精度评级				预报值/ ℃	绝对误差/ ℃	精度评级					
				MSE	Δ	C	P 等级			MSE	Δ	C	P 等级		
2004	3.4	2.080	1.320					3.415	0.015						
2005	3.3	3.067	0.233	0.544	0.156	2.287	75% 3级	3.448	0.148	0.146	0.049	0.615	100% 1级		
2006	3.8	3.299	0.501					3.431	0.369						
2007	3.3	3.190	0.110					3.458	0.158						

年份	实际值/ ℃	数据挖掘集成							模糊时间序列模型						
		预报值/ ℃	绝对误差/ ℃	精度评级				预报值/ ℃	绝对误差/ ℃	精度评级					
				MSE	Δ	C	P 等级			MSE	Δ	C	P 等级		
2004	3.4	3.393	0.007					3.383	0.617						
2005	3.3	3.026	0.274	0.164	0.058	0.687	100% 1级	3.725	0.325	0.312	0.084	1.311	75% 3级		
2006	3.8	3.921	0.121					3.775	0.475						
2007	3.3	2.923	0.377					3.950	0.150						

表6 重庆市春旱指数预报结果对比

Table 6 Drought index forecast result comparison of Chongqing in January

年份	实际值	加权集成							神经网络集成						
		预报值	绝对误差	精度评级				预报值	绝对误差	精度评级					
				MSE	Δ	C	P 等级			MSE	Δ	C	P 等级		
2004	0.118	0.488	0.370					0.380	0.262						
2005	0	0.425	0.425	0.104	4.322	0.623	25% 3级	0.499	0.499	0.205	7.263	1.224	25% 4级		
2006	0.029	0.299	0.270					0.587	0.558						
2007	0.368	0.559	0.191					0.248	0.120						

年份	实际值	数据挖掘集成							模糊时间序列模型						
		预报值	绝对误差	精度评级				预报值	绝对误差	精度评级					
				MSE	Δ	C	P 等级			MSE	Δ	C	P 等级		
2004	0.118	0.318	0.200					0.025	0.093						
2005	0	0.222	0.222	0.074	1.383	0.442	25% 2级	0.115	0.115	0.110	4.306	0.659	25% 4级		
2006	0.029	0.086	0.057					0.366	0.337						
2007	0.368	0.188	0.180					0.555	0.187						

3 结束语

本文简要介绍了模糊技术在气象应用中的研究状况,并将模糊时间序列模型引入短期的气候预测中,然后通过实例计算,验证了方法的可行性,最后分别利用两种方法对模型的精度进行了评定.与其他集成模型相对照,结果表明:模糊时间序列模型用于短期气候预测是可行的;模糊时间序列模型的预报精度与加权集成、神经网络集成、数据挖掘集成等集成模型精度相当;模糊时间序列模型建模简单,所需数据样本很小,且计算简单,使用方便,非常适合短期预报.中长期预报研究是下一步需要做的工作.

参考文献

References

- [1] 陈国范. 模糊数学在气象中的应用简述[J]. 气象科技, 1982(6): 8-14
CHEN Guofan. The brief introduction of fuzzy mathematics application in meteorology[J]. Meteorological Science and Technology, 1982(6): 8-14
- [2] 楼世博, 陈化成. Fuzzy 聚类与天气预报[J]. 模糊数学, 1981(1): 79-90
LOU Shibo, CHEN Huacheng. Fuzzy clustering and weather forecasting [J]. Fuzzy Mathematics, 1981(1): 79-90
- [3] 宗荣祥. 模糊数学在长期天气预报中的应用[J]. 数学的实践与认识, 1982(1): 1-4

- ZONG Rongxiang. The application of fuzzy mathematics in the long-term weather forecast [J]. *Mathematics in Practice and Theory*, 1982(1):1-4
- [4] 王式中. 用模糊数学作天气周期分类和天气周期配置年型等预报 [C] // 模糊数学在气象中的应用会议, 1981:95-102
- WANG Shizhong. Using the fuzzy mathematics for the weather cycle classification and weather cycle in configuration of forecast [C] // Proceedings of the Conference on Fuzzy Mathematics Applications in Meteorology, 1981: 95-102
- [5] 曹鸿兴, 陈国范. 运用隶属原则和择近原则对天气类型进行模糊识别 [C] // 长江流域规划办公室. 中长期水文气象预报文集, 1981:28-34
- CAO Hongxing, CHEN Guofan. Using the subjection principle and choose nearly principle to recognize weather type by fuzzy [C] // The Yangtze River Watershed Planning Office. Mid-to-long-term Hydrological Meteorological Forecast Essays, 1981:28-34
- [6] 章文茜. 试用 Fuzzy 数和贴进度做天气预报评分 [J]. *模糊数学*, 1982(1):83-90
- ZHANG Wenqian. Trial Fuzzy number and close to the degrees do the weather forecast score [J]. *Fuzzy Mathematics*, 1982(1):83-90
- [7] Song Q, Chissom B S. Fuzzy time series and its models [J]. *Fuzzy Sets System*, 1993, 54(3):269-277
- [8] Sun X H, Li Y M. Average-based fuzzy time series models for forecasting Shanghai compound index [J]. *International Journal of Applied Science and Engineering*, 2005, 3:234-244
- [9] Chen S M, Hsu C C. A new method to forecast enrollments using fuzzy time series [J]. *International Journal of Applied Science and Engineering*, 2004, 2(3):234-244
- [10] Singh S R. A simple method of forecasting based on fuzzy time series [J]. *Applied Mathematics and Computation*, 2007, 186(1):330-339
- [11] 吴铭峰, 蒋勋. 基于模糊时间序列的预测模型: 以上证指数为例 [J]. *无锡南洋学院学报*, 2008, 7(3):57-63
- WU Mingfeng, JIANG Xun. Forecasting model based on fuzzy time series: With an example of Shanghai stock index [J]. *Journal of Wuxi South Ocean College*, 2008, 7(3):57-63
- [12] 张钰敏, 张羽, 沈晓羽. 国际原油期货价格的模糊时间序列预测 [J]. *石油天然气学报*, 2011, 33(5):313-317
- ZHANG Yumin, ZHANG Yu, SHEN Xiaoyu. Research of international petroleum futures price forecasting [J]. *Journal of Oil and Gas Technology*, 2011, 33(5):313-317
- [13] 傅立. 灰色系统理论及其应用 [M]. 北京: 科学技术文献出版社, 1992
- FU Li. Grey system theory and its application [M]. Beijing: Science and Technology Documentation Publishing House, 1992
- [14] 李学明, 郭尚坤, 王剑柯, 等. 新的集成预报及其在短期气候预测中的应用 [J]. *重庆大学学报*, 2010, 33(12):119-126
- LI Xueming, GUO Shangkun, WANG Jianke, et al. A new integrating forecast and its application in short-term climate prediction [J]. *Journal of Chongqing University*, 2010, 33(12):119-126

Application of fuzzy time series model in short-term climate prediction

WANG Yongdi¹

¹ School of Remote Sensing, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract The fuzzy time series model was introduced to short-term climate prediction. Daily ground observations (from 1971 to 2007) from 34 weather stations of Chongqing, and the drought index and the flood index computed by Chongqing drought and flood monitoring and warning decision making system are used. The precipitation, average temperature of Chengkou county of Chongqing in January, and the Spring drought index of Chongqing from 2001 to 2007 are analyzed by fuzzy time series model. The trend in precipitation, average temperature, and the Spring drought index of 2004—2007 are predicted by fuzzy time series method, and the results are compared with that of other models like weight integrating forecast model, artificial neural networks, data mining integrating forecast. The comparison shows that fuzzy time series model is better in accuracy and simpler for computation. The fuzzy time series method will be more valuable in future application.

Key words fuzzy time series; fuzzy technology; short-term climate prediction; forecasting model; model accuracy assessment