

基于 FN—SPA 模型的降雨径流预测

黄建国¹, 王东明²

(1. 安徽水利水电职业技术学院 基础部, 安徽 合肥 231603; 2. 皖西学院 金融与数学学院, 安徽 六安 237012)

摘 要: 为了提高降雨径流的预测精度, 提出一种基于泛函网络与集对分析(FN—SPA)的预测模型。该模型采用集对分析对网络训练样本进行优化, 然后使用泛函网络进行预测。计算机仿真结果表明, 与神经网络模型相比较, 该模型的预测精度更高。

关键词: 泛函网络; 集对分析; 降雨径流

中图分类号: O177.92

文献标识码: A

文章编号: 1009-9735(2020)05-0038-04

降雨径流是指落到地面的降雨, 除去蒸发、地表和地下调蓄的水量, 通过地面径流、表层流和地下流汇集到出口断面的汇流过程。影响降雨径流的因素很多, 如降雨量、降雨时间、地形地貌、土壤湿度等等。因此, 降雨径流的预测是一个比较困难的过程。目前国内外降雨径流的预测方法可归纳分为数学物理模型、概念模型和黑箱模型, 如瞬时单位线法^[1]、综合单位线法、等流时线法、调蓄函数法、tank 模型、人工神经网络模型等等^[2-3], 但各种预测方法都有一定的局限性。泛函网络(Functional Networks, 简称 FN)是一种“灰箱”模型, 具有很好的非线性逼近能力。集对分析(SetPair Analysis, 简称 SPA)具有很好的信息预处理能力。为了提高降雨径流的预测精度, 把泛函网络与集对分析结合起来, 提出了一种 FN—SPA 模型。计算机仿真结果显示, 与神经网络模型相比较, 在预测降雨径流洪峰流量时, 该模型的预测精度更高。

1 FN—SPA 模型理论

1.1 泛函网络简介

1998 年 Castillo E. 提出了泛函网络^[4], 它是对神经网络的推广。一个典型的泛函网络模型主要组成如下: 1) 输入单元层。它是输入数据的一层单元, 如图 1 中的 $\{x_1, x_2, x_3\}$ 。2) 泛函神经元。每个泛函神经元是一个计算单元, 对一个或多个输入的信息进行处理产生输出, 如图 1 中的 $\{f_1, f_2, f_3\}$ 。3) 中间存储单元。它存储由神经元产生的信息, 如图 1 中的 $\{x_4, x_5\}$ 。4) 一个输出层。输出网络的结果, 如图 1 中的

$\{x_6\}$ 。5) 有向连接线。它们连接输入层、中间层神经元和输出层, 如图 1 中的箭头表示, 箭头的方向表示信息流的方向。

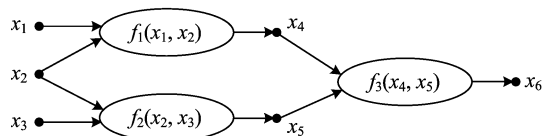


图 1 典型的泛函网络模型

1.2 泛函网络学习算法

可分离泛函网络是目前应用最广泛的泛函网络之一, 可分离是指它的表达式为输入变量分离作用效果的组合。用于 FN—SPA 模型的可分离泛函网络如图 2, 它的输出函数为

$$t = F(x, y, z) = f_1(x) \times f_2(y) + f_3(z) \quad (1)$$

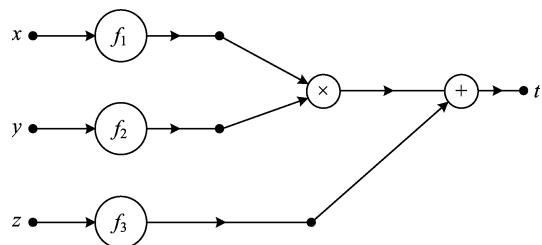


图 2 用于 FN—SPA 模型的可分离的泛函网络结构

* 收稿日期: 2020-06-22

基金项目: 安徽高校自然科学基金重点项目(KJ2015A377)的研究成果。

作者简介: 黄建国(1962—), 男, 安徽怀宁人, 副教授, 研究方向: 高等数学教学及应用。

定理^[5] 方程 $\sum_{i=1}^n f_i(x_i)g_i(y_i) = 0$ (写成向量

形式为 $f^T(x)g(y) = 0$) 的所有解可以表示成 $f(x) = A\varphi(x), g(y) = B\Psi(y)$ 的形式, 式中 A 和 B 是常数矩阵(维数分别为 $n \times r$ 和 $n \times (n-r)$) 且 $A^TB = 0, \varphi(x) = (\varphi_1(x), \varphi_2(x), \dots, \varphi_r(x))$ 和 $\Psi(y) = (\Psi_{r+1}(y), \dots, \Psi_n(y))$ 是两个由任意相互线性独立的函数组成的系统, 其中 r 是 0 和 n 之间的整数。

利用最小化误差平方和的方法来优化网络系数 a_{ik} ,

$$E_0 = \sum_{j=1}^m e_j^2 = \sum_{j=1}^m [t_j - \sum_{i=1}^n f_i(x_j)]^2$$

$$= \sum_{j=1}^m [t_j - \sum_{i=1}^n \sum_{k=1}^p a_{ik} \varphi_i(x_j) \varphi_k(x_j)]^2 \quad (2)$$

式中, t_j 为第 j 个训练模式输出。

为了得到唯一的网络表达式, 需要给出一些泛函网络的初始值。设泛函网络的初始值为 $f_i(x_0) = u_{i0}, i=1, \dots, n$ 。

$$E = E_0 + \sum_{i=1}^n \lambda_i [f_i(x_0) - u_{i0}] \quad (3)$$

对(3)式求偏导, 可得

$$\frac{\partial E}{\partial a_{ik}} = -2 \sum_{j=1}^m e_j \varphi_i(x_j) \varphi_k(x_j) + \sum_{i=1}^n \lambda_i [f_i(x_0) - u_{i0}]$$

$$\equiv 0, i=1, \dots, n, k=1, \dots, p \quad (4)$$

$$\frac{\partial E}{\partial \lambda_i} = f_i(x_0) - u_{i0} = 0, i=1, \dots, n, \quad (5)$$

训练泛函网络就是求解(4)、(5)组成的方程组。

1.3 集对分析

集对分析是赵克勤先生在 1989 年提出的一种处理不确定性问题的理论和方法^[6]。四元联系数是对组成集对的两个集合特性作同一性、偏同性、偏反性和对立性分析, 用联系数来刻画两个集合的联系程度, 其联系度的表达式为

$$u = \frac{A}{N} + \frac{B}{N}i + \frac{C}{N}j + \frac{D}{N}k \quad (6)$$

式中, N 为集合的特性总数; A 为同一性个数; B 为偏同性个数; C 为偏反性个数; D 为对立性个数; $i \in [0, 1]; j \in [-1, 0]; k = -1$ 。

1.4 FN—SPA 模型原理

FN—SPA 模型的输入变量有总降雨量 x_1 、降雨历时 x_2 、河道基流 x_3 、单位时间平均降雨量 x_4 , \dots, x_n 。把需要预测值所对应的输入作为 B 集合, 即 $B = \{x_1, x_2, \dots, x_n\}$ 。将历史资料时间序列样本

集的输入分别作为 A_t , 即 $A_t = \{x_{t1}, x_{t2}, \dots, x_{tm}\}, t=1, \dots, m$ 。按照联系数的大小, 在 A_1, A_2, \dots, A_t 中, 寻找与 B 集合最为相似的 k 个集合, 把这 k 个集合和其对应的输出作为训练集。由于训练集的输入与需要预测值的输入条件类似, 即减少了极端数据作为训练样本对预测结果带来的不利影响, 因此预测的结果相对会更精确一些。

2 FN—SPA 模型算法

step 1. 预留部分历史资料时间序列为检测集。

将最后 l 个历史资料时间序列作为检测集, 其余的历史资料时间序列都作为样本数据。

step 2. 构建集对 B_i 与 A_t 。将检测集中第 i 个预测值所对应的输入作为集合 B_i , 即 $B_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ 。把各样本数据的输入作为集合 $\{A_t\}, t=1, \dots, m$, 其中 $A_t = \{x_{t1}, x_{t2}, \dots, x_{tm}\}$ 。

step 3. 计算集合 B_i 与每一个集合 A_t 的联系数。如果 $|A_{ti} - B_i| \leq 15\% B_i$, 认为分量 A_{ti} 与分量 B_i 具有同一性; 如果 $15\% B_i < |A_{ti} - B_i| \leq 25\% B_i$, 认为分量 A_{ti} 与分量 B_i 具有偏同性; 如果 $25\% B_i < |A_{ti} - B_i| \leq 35\% B_i$, 认为分量 A_{ti} 与分量 B_i 具有偏反性; 如果 $35\% B_i < |A_{ti} - B_i|$, 认为分量 A_{ti} 与分量 B_i 具有对立性。逐一对集合 B_i 与每一个集合 A_t 的同一性、偏同性、偏异性和对立性进行分析, 计算出 B_i 与每一个集合 A_t 的联系数。按照联系数的大小, 在 A_1, A_2, \dots, A_t 中, 找出与 B_i 集合最为相似的 k 个集合, 将此集合与其对应的输出作为训练集。

step 4. 数据归一化处理。归一化方法:

$$T = \frac{X}{X_{\max}} \quad (7)$$

式中, X 为原始数据; X_{\max} 为原始数据的最大值; T 为变换后的数据, 即目标数据。

step 5. 网络训练。首先给出任意两两相互线性独立的函数组成的系统, 即网络基函数 $\varphi(x) = (\varphi_1(x), \varphi_2(x), \dots, \varphi_r(x))$ 。其次, 将训练集中归一化的数据输入泛函网络, 进行网络训练, 得到最佳的网络参数。

step 6. 网络逼近能力检测。利用最佳的网络参数, 把 $B_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ 输入网络, 得到所预测的第 i 个时间序列的洪峰流量。

3 计算机仿真

应用 FN—SPA 模型对降雨径流进行预报的实

例资料是来源日本一柳(Ichiyanagi)先生在日本中部史资料分成样本数据集和检测样本数据集,如表 1 和大井川河上游某地区降雨径流资料^{[7](P100-101)}。将历史资料分成样本数据集和检测样本数据集,如表 1 和表 2。

表 1 样本数据集

降雨次数	时间(年、月、日)	总降雨量/mm	降雨历时/h	河道基流/(m^3/s)	洪峰流量/(m^3/s)
1	1980.04.06	102	20	17.6	278
2	1980.04.12	148	46	25.2	356
3	1980.05.15	145	32	35.0	213
4	1980.05.25	122	30	35.0	272
5	1980.09.10	197	52	17.4	316
6	1980.10.19	120	54	24.3	122
7	1981.05.11	91	30	41.0	259
8	1981.05.17	44	21	38.0	92
9	1981.06.21	103	39	18.2	173
10	1981.06.25	117	71	40.2	124
11	1981.07.02	127	74	39.1	186
12	1981.09.03	53	35	22.5	118
13	1981.10.08	205	38	24.1	380
14	1981.10.23	97	20	20.0	136
15	1982.05.02	84	46	19.0	141
16	1982.05.14	71	18	31.6	148
17	1982.08.26	123	34	20.6	238
18	1982.09.24	52	29	40.8	107
19	1983.04.15	230	63	26.7	425
20	1983.05.06	109	24	37.2	311

表 2 检测样本数据集

降雨次数	时间(年、月、日)	总降雨量/mm	降雨历时/h	河道基流/(m^3/s)	洪峰流量/(m^3/s)
1	1983.05.15	207	38	43.0	504
2	1983.06.20	160	28	12.5	438
3	1983.07.20	159	72	70.9	215
4	1983.09.26	347	68	24.9	958
5	1983.10.08	154	70	34.0	187

计算两个集合联系度时,把总降雨量、降雨历时、河道基流和单位时间平均降雨量作为集对的元素。(6)式中, i 和 j 分别取 0.5 与 -0.5。选取降雨历时、河道基流和单位时间平均降雨量作为网络训练输入变量。采用 $\varphi_i(x) = (1, x)$ 作为泛函网络的基函数。为了便于求解最优网络参数,同时也发挥泛函网络是灰箱的优点,设网络函数为

$$F(x, y, z) = a_1 xy + a_2 x + a_3 y + a_4 z + a_5 \quad (8)$$

式中, x 为降雨历时; y 为单位时间平均降雨量; z 为河道基流; a_i 为网络参数, $i=1, \dots, 5$ 。网络的初始值设为: $F(0, 0, z) = z$ 。

仿真的结果见表 3。由表 3 可知, FN—SPA 模型对降雨径流洪峰流量预测的精度要高于人工神经网络模型,说明 FN—SPA 模型预测降雨径流洪峰流量是有效的,而且预测精度更高。

表 3 预报误差

降雨次数	时间 (年、月、日)	洪峰流量 /(m ³ /s)	FN—SPA 模型预测 /(m ³ /s)	ANN 模型预测 /(m ³ /s)
1	1983. 05. 15	504	450. 8	373. 0
2	1983. 06. 20	438	350. 2	332. 9
3	1983. 07. 20	215	238. 5	275. 2
4	1983. 09. 26	958	622. 3	603. 5
5	1983. 10. 08	187	231. 7	314. 2

注:ANN 模型预测数据是文献中峰值误差百分比数据转化来的^{[7](P103)}。

4 结语

本文的 FN—SPA 模型是对降雨径流洪峰流量预报方法的一种探索。FN—SPA 模型优点是既利用集对分析量化解决不确定性问题,又利用泛函网络良好非线性逼近能力。仿真的结果显示 FN—SPA 模型预测精度更高,表明该模型预报降雨径流是有效的。

FN—SPA 模型还存在一些不足:集对分析中, i 和 j 具体选取什么值最优,目前还没有理论指导。在集合 A_i 中,具体界定哪些集合与 B 相似,还处于经验阶段。泛函网络基函数的选取也缺乏理论指导。

参考文献:

[1]罗萍萍. 降雨径流预报瞬时单位线推求分析[J]. 黑龙江水利科技,2016,44(6):41-42.
[2]金保明,卢光毅,王伟,等. 基于弹性梯度下降算法的 BP 神

经网络降雨径流预报模型[J]. 山东大学学报(工学版), 2020,50(3):117-124.
[3]殷兆凯,廖卫红,王若佳,等. 基于长短时记忆神经网络(LSTM)的降雨径流模拟及预报[J]. 南水北调与水利科技,2019,17(6):1-9.
[4]Castillo E. Functional Networks[J]. Neural Processing Letters, 1998, 7(3): 151-159.
[5]Castillo E., Gutierrez J M. Nonlinear Time Series Modeling and Prediction Using Functional Networks[J]. Extracting Information Masked by Chaos, Phys., Lett. A, 1998(244): 71-84.
[6]赵克勤. 集对分析及其初步应用[J]. 大自然探索,1994,13(1):18-23.
[7]苑希民,李鸿雁,刘树坤,等. 神经网络和遗传算法在水科学领域的应用技术[M]. 北京:中国水利水电出版社,2002.

Rainfall Runoff Forecasting based on FN-SPA Model

HUANG Jianguo¹, WANG Dongming²

(1. Department of Basic Education, Anhui Water Conservancy Technical College, Hefei 231603, China;
2. Faculty of Finance and Mathematics, West Anhui University, Lu'an 237012, China)

Abstract: In order to improve the forecast accuracy of rainfall runoff, a prediction model is presented based on set pair analysis and functional network(FN -SPA). The model uses set pair analysis to optimize the network training sample and then uses functional network to predict rainfall runoff. The simulation results demonstrate that the predicted date by the FN -SPA model is more accurate than neural network model.

Key words: functional networks; set pair analysis; rainfall runoff