

孙家清,张志薇,艾文文. BP 神经网络在油菜花期预报中的应用[J]. 气象与环境科学,2019,42(4):22-26.

Sun Jiaqing, Zhang Zhiwei, Ai Wenwen. Application of BP Neural Network in the Prediction of Oilseed Rape Florescence[J]. Meteorological and Environmental Sciences, 2019, 42(4): 22-26.

doi:10.16765/j.cnki.1673-7148.2019.04.004

# BP 神经网络在油菜花期预报中的应用

孙家清, 张志薇, 艾文文

(江苏省气象服务中心, 南京 210009)

**摘要:** 建立花期预报模型, 发布观赏性植物的精准花期预报, 为旅游活动提供重要参考依据, 已经成为气象服务领域一个新的发展方向。为了解高淳旅游区油菜花期的变化规律, 探索其预报方法, 指导高淳油菜花节旅游活动, 根据 1985—2010 年高淳站日最高气温、日最低气温、日平均气温、日降水量、日日照时数、日平均 5 cm 地温、日平均相对湿度和日小型蒸发量等气象观测数据, 利用主成分分析法, 得到其与油菜花期相关系数较大的 3 个主成分, 即温度因子、天气因子和辐射因子, 以此为输入因子, 建立基于 BP 神经网络的油菜花期预报模型, 探讨 BP 神经网络在花期预报领域的应用。结果表明, 传统的有效积温方法预报结果与实际开花期平均相差 4.25 天, BP 神经网络方法预报结果与实际开花期平均相差 1.5 天, 与有效积温预报油菜花期的方法相比, BP 神经网络技术具有预测结果准确率高和操作简单等特点, 在花期预报领域具有广阔的应用前景。

**关键词:** BP 神经网络; 主成分分析; 油菜花期; 预报模型

**中图分类号:** S165+.29

**文献标识码:** A

**文章编号:** 1673-7148(2019)04-0022-05

## 引言

随着经济的发展和物质生活水平的提高, 旅游赏花已成为人们享受美好生活的方式之一。“樱花节”“油菜花节”“桃花节”等时令旅游活动的举办依赖于实际的开花时间<sup>[1]</sup>, 赏花活动的举办日期早于或者晚于实际的花期直接影响旅游点的收入和游客满意度。关于植物花期与气候、气候变化的关系和观赏花期的预测技术国内外有诸多研究<sup>[2-7]</sup>。王焕炯等<sup>[8]</sup>模拟了过去 50 年中国白蜡树展叶始期时空变化。Gonsamo 等<sup>[9]</sup>模拟了加拿大 19 种植物 1948 年来的始花期距平变化。张增信等<sup>[10]</sup>对南京市 8 种常见观赏树木的初花、盛花和末花期进行了特征分析。Morin 等<sup>[11]</sup>预测了未来 100 年北美 22 种木本植物展叶始期变化。叶海龙等<sup>[12]</sup>分析了龙游油菜花开花期影响较大的气象因子。仲舒颖等<sup>[13]</sup>对中国观赏性植物花期建立基于温度的时空预测模型。针对观赏性植物花期预报, 现有研究主

要以樱花为主, 而对其他观赏性植物的花期模拟却较少<sup>[14-15]</sup>。高淳区位于南京市南端, 被誉为南京的后花园和南大门。每年春季 1.3 余万公顷的油菜花已经成为高淳地区一张亮眼的名片, 至今已成功举办 10 届油菜花节。油菜花属于春季观赏性植物, 开花期集中在每年的 3—4 月。近 10 年的研究表明, 其开花期与前一年冬季(12 月—翌年 2 月)的气象因子关系比较紧密<sup>[16-18]</sup>。

目前, 油菜花期预报主要采用的是有效积温法则, 即植物开花所需要的满足植物发育阈值温度后一定时间累计的有效积温<sup>[16]</sup>。通过积温判断油菜花花期的方法, 操作简单, 有一定的生物依据, 但是由于该方法忽略了日照、降水等气象因子对油菜花生长的影响, 致使预测的花期预报准确率较低。随着计算科学的发展, 人工神经网络技术为花期的模拟预测提供了新的途径。本文利用主成分分析法分析气象因子对油菜花期的各大主成分, 综合考虑气象因子与油菜花期的关系, 建立基于神经网络的油

收稿日期: 2018-06-15; 修订日期: 2019-05-31

基金项目: 青年基金项目(Q201704)

作者简介: 孙家清(1990-), 男, 江苏南京人, 助理工程师, 硕士, 从事智能算法、计算机应用、气象信息安全研究。E-mail: jqsun115@163.com

菜花期预测模型。

### 1 资料和方法

#### 1.1 研究资料

本文选取南京市高淳气象站 1985—2010 年日最高气温、日最低气温、日平均气温、日降水量、日日照时数、日平均 5 cm 地温、日平均相对湿度和日小型蒸发量等气象观测资料,通过计算得到 1986—2010 年逐年冬季要素的平均值。同时从相关部门获得高淳相对应年份的油菜开花期日期。

#### 1.2 研究方法

##### 1.2.1 主成分分析

主成分分析(Principal Component Analysis, PCA)最早是由 Pearson 于 1901 年发明<sup>[19]</sup>,旨在通过对协方差矩阵进行特征分析,达到在减少数据维数的同时,保持数据集对方差贡献最大的目的。通过降维把原来指标转化为一个或几个综合指标,即主成分。各个主成分间互不相关。

采用主成分分析法对油菜花期相关气象因子进行分析的步骤如下。

步骤 1:选取油菜花期相关的气象因子,对其原始数据进行 z-score 标准化(zero-mean normalization)处理<sup>[20]</sup>,其公式如下:

$$x_{ij}^* = (x_{ij} - \bar{x}_j) / s_j \quad (1)$$

式中,  $x_{ij}^*$  为标准化后的各气象因子,  $x_{ij}$  是各气象因子的原始数据,  $\bar{x}_j$ 、 $s_j$  分别为各气象因子的均数和标准差。

步骤 2:计算出标准化后气象因子数据的相关系数矩阵  $R$ :

$$R_{ij} = \frac{\sum_{k=1}^n (|x_{ki} - \bar{x}_i| |x_{kj} - \bar{x}_j|)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (2)$$

式中,  $R_{ij}$  为标准化第  $i$  个指标与第  $j$  个指标的相关系数。

步骤 3:根据气象因子相关系数矩阵  $R$  求出特征值、主成分贡献率和累计方差贡献率,确定主成分的个数;解特征方程

$$|\lambda E - R| = 0$$

求出特征值  $\lambda_i (i=1, 2, \dots, p)$ ,  $p$  为主成分的个数;特征值是各主成分之间的方差,它反映各个主成分的影响力。主成分  $Z_i$  的共享率为

$$W_i = \lambda_i / \sum_{j=1}^p \lambda_j$$

累计贡献率为

$$\sum_{j=1}^m \lambda_j / \sum_{j=1}^p \lambda_j$$

根据选取主成分个数原则,选取特征值大于 1 且累计贡献率大于 85% 的特征值。

##### 1.2.2 BP 神经网络

BP (Black Propagation) 神经网络是由 Rumelhart 和 McClelland 为首的科学家提出的概念,是一种按照误差逆向传播算法训练的多层前馈网络,是至今为止应用最广泛的神经网络。它的基本思想是梯度下降法,利用梯度搜索技术,以期使网络的实际输出值和期望值的误差均方差为最小。如图 1 为 BP 神经网络拓扑结构,包含输入层、隐层与输出层,若输出层得不到期望输出则转入误差反向传播,根据预测误差调整神经网络的网络权值、阈值,以期逼近期望输出。

图 1 中,  $X_1, X_2, \dots, X_m$  是 BP 神经网络的输入参数,  $Y_1, Y_2, \dots, Y_n$  是输出参数,  $\omega_{ij}$  与  $\omega_{jk}$  为神经网络权值,  $b$  为神经网络阈值,其中  $a = f_1(\omega p + b_1)$ ,  $y = f_2(\omega a + b_2)$ 。

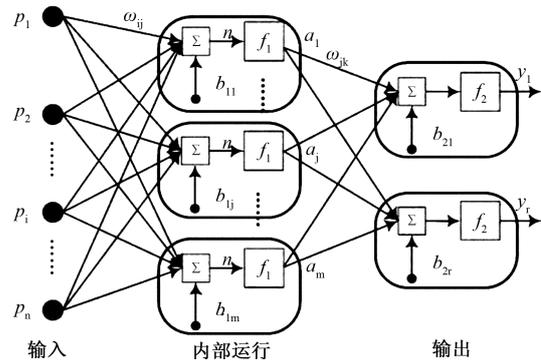


图 1 BP 神经网络拓扑结构

研究表明, BP 神经网络模型能较好地表征复杂的非线性问题<sup>[21-22]</sup>,而且一个 3 层的 BP 神经网络即可模拟任意复杂的非线性问题<sup>[23-24]</sup>。因此,本文也采用 3 层的 BP 神经网络模型,即 1 个输入层、1 个隐层和 1 个输出层,其中转移函数均为单极性 Sigmoid 函数。建立基于气象主成分因子的开花期预报模型,首先分析输入和输出条件。输出条件为油菜花的开花日期,输入条件是选取的主成分,同时采用 8 个气象因子原始数据作为输入,并对两种不同输入方式进行了比较。

隐层神经元节点数选择是 BP 神经网络设计中的关键步骤,节点数过少可能得不到期望结果,节点数过多又易导致网络训练复杂化,训练时间延长。网络预

测误差一般随着隐层节点数增加呈先下降后升高的趋势,数量确定一般通过试凑法和以下公式确定:

$$l=2m+1 \text{ 或 } l < \sqrt{(m+n)} + a \quad (3)$$

式中, $l$ 为隐层节点数, $m$ 为输入层节点数, $n$ 为输出层节点数, $a$ 为0—10的常数<sup>[25]</sup>。

## 2 结 果

### 2.1 主成分结果分析

采用1985—2010年高淳站的气象观测数据,对数据进行标准化处理,主成分分析代码运行于 matlab R2010b 中,求出其相关矩阵的特征值和各主成分的贡献率,结果见表 1。一般选取特征根大于 1 且累计贡献率大于 85%的  $m$  个主成分。从表 1 中

所知,前 3 个主成分的特征根大于 1,且累计贡献率达到 89.16%,基本上可以反映原气象指标的大部分信息,能够降低原始数据的复杂性,达到降维的目的。

表 1 1985—2010年南京市高淳站 24 个样本  
相关系数矩阵的特征值

成分	特征根	贡献率/%	累计贡献率/%
1	3.7560	46.95	46.95
2	2.2964	28.71	75.56
3	1.0804	13.50	89.16
4	0.4739	5.92	95.08
...	...	...	...

3 个主成分因子的特征向量列于表 2。

表 2 南京市高淳气象站1985—2010年气象资料主成分特征向量

气象因子	第 1 主成分	第 2 主成分	第 3 主成分
$x_1$ : 冬季平均日最低气温	0.4019	0.3697	-0.0123
$x_2$ : 冬季平均日最高气温	0.4973	-0.0603	0.1230
$x_3$ : 冬季平均日平均气温	0.4968	0.1609	0.0721
$x_4$ : 冬季平均日降水量	0.1534	-0.5523	0.0528
$x_5$ : 冬季平均日日照时数	-0.0992	0.5796	0.2066
$x_6$ : 冬季平均日平均 5 cm 地温	0.4879	0.1091	-0.1036
$x_7$ : 冬季平均日平均相对湿度	0.2153	-0.4008	0.4906
$x_8$ : 冬季平均日小型蒸发量	-0.1630	0.1425	0.8263

由表 2 可知,在第 1 主成分的特征向量中,特征值较大且为正的因子是冬季平均日最低气温、冬季平均日最高气温、冬季平均日平均气温和冬季平均日平均 5 cm 地温。由此可以看出,花期与温度紧密相关,因此第 1 主成分可以称为温度因子。在第 2 主成分的特征向量中,特征值较大且为正的因子是冬季平均日日照时数,较大且为负的因子是冬季平均日降水量和冬季平均日相对湿度,所以,花期是与日日照时数呈明显的正相关,与降水量和相对湿度呈负相关,第 2 主成分称为天气因子。第 3 主成分的特征向量中,特征值较大且为正的因子是冬季平均日小型蒸发量,可以称为辐射因子。根据主成分的特征向量(表 2),可获得 3 个主成分与气象因子间的线性方程:

$$F_1 = 0.4019x_1 + 0.4973x_2 + 0.4968x_3 + 0.1534x_4 - 0.0992x_5 + 0.4879x_6 + 0.2153x_7 - 0.163x_8 \quad (4)$$

$$F_2 = 0.3697x_1 - 0.0603x_2 + 0.1609x_3 - 0.5523x_4 + 0.5796x_5 + 0.1091x_6 - 0.4008x_7 + 0.1425x_8 \quad (5)$$

$$F_3 = -0.0123x_1 + 0.123x_2 + 0.0721x_3 + 0.0528x_4 + 0.2066x_5 - 0.1036x_6 + 0.4096x_7 + 0.8263x_8 \quad (6)$$

### 2.2 BP 神经网络建模结果与分析

1985—2010年高淳站共 24 个样本数据(1989 年缺测),其中 20 个作为训练样本,4 个作为检验样本。利用主成分分析法得到影响油菜花期的 3 个主成分因子作为 BP 神经网络的输入层,共 3 个节点;输出层为油菜花开花日期,即 1 个节点;隐层的节点数通过试凑法取 8 个节点。故 BP 神经网络的拓扑结构为 {3, 8, 1} (即网络输入层为 3,隐层为 8,输入层为 1)。在对所确定的网络结构进行训练的过程中,为防止出现“过渡训练”,网络采用如下的收敛规则:若满足绝对误差小于 2 的样本达到总样本的 85%,则停止训练,否则规定最大的训练次数(10000 次),当训练达到要求后则利用样本对网络进行验证。

当以 3 大主成分因子作为网络输入时,考虑的是气象因子的综合情况,放大主要因子,减弱次要因子。同时,本文也通过 8 大气象因子的原始数据作为网络的输入,重新确定 BP 神经网络结构,即 BP 神经网络的拓扑结构为 {8, 17, 1},收敛标准相同。

根据两种不同拓扑结构的 BP 神经网络进行油菜开花期预报建模,程序运行于 matlab R2010b 中,

图2和图3为两种建模对油菜花期的模拟结果。由图2和图3可以看出,采用3大主成分因子为输入的BP神经网络模拟的平均误差为2.2天,略优于误差为2.45天的8个气象因子为输入的BP神经网络模拟。利用2007—2010年的油菜花期的数据样本,对以上两种拓扑结构的BP神经网络建模方法和传统的有效积温法则方法进行验证。三种方法的验证结果见表3。

由表3可知,采用3大主成分因子{3,8,1}作为网络输入的花期预报的误差为1.5天,8大气象因子{8,17,1}作为BP神经网络的输入的误差为2.75天,3大主成分因子作为网络输入的预测的开花日期更接近实际开花日期。由此可以看出,这种通过主成分分析过滤后的气象因子作为模型输入的建模方法提高了花期预报精度。与传统有效积温法则预报油菜花期的误差为4.25天相比,两种BP神经网络建模在油菜花期预报方面精确性更好,且具有很强的可操作性。

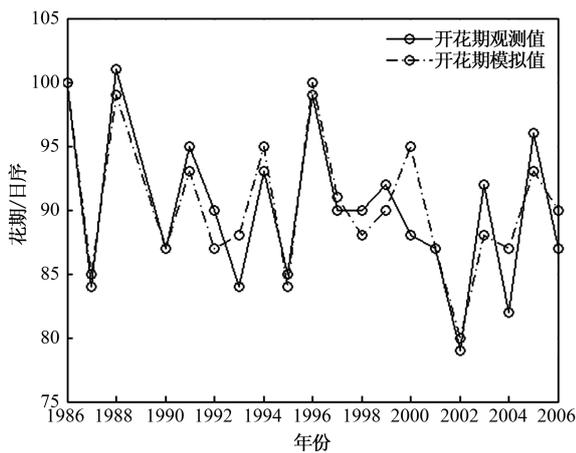


图2 3个主成分因子为输入的基于BP神经网络的油菜花期模拟图

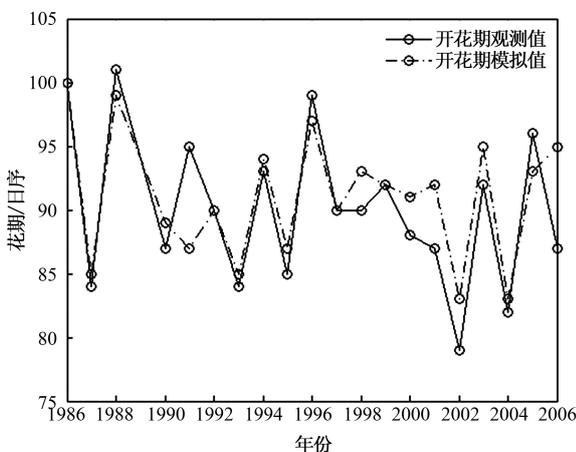


图3 8个气象因子为输入的基于BP神经网络的油菜花期模拟图

表3 传统有效积温预报模型和BP神经网络花期预报模型预测结果

年份	实际开花期	有效积温预测	BP结构{8,17,1}预测	BP结构{3,8,1}预测
2007	3月28日	3月19日	3月26日	3月31日
2008	3月24日	3月27日	3月28日	3月22日
2009	3月24日	3月22日	3月28日	3月25日
2010	3月24日	3月27日	3月25日	3月24日

### 3 结 语

通过主成分分析法分析影响油菜花期的8个气象因子,得到3个主成分因子,分别是温度影响较大的第1主成分因子(温度因子),降水、日照和湿度系数影响较大的第2主成分因子(天气因子)和蒸发量影响为主的第3主成分因子(辐射因子)。将BP神经网络引入到油菜花期预报应用中,油菜开花期为输出层,3大主成分因子为输入层,构建BP神经网络花期预报模型。传统的有效积温方法预报结果与实际开花期平均相差4.25天,3大主成分因子作为输入的BP神经网络预报结果与实际开花期平均相差1.5天。由此可见,相较于传统有效积温法则预报,BP神经网络构建的花期预报模型具有较高的可靠性和准确性。

人工神经网络具有强大的预测能力、大规模并行处理特性、高度的容错性和自适应性,其在花期预报方面的应用前景广阔。在今后花期预报研究中,进一步建立相关因子选择体系,除气象因子之外,添加人工因子、环境因子等,从而构建准确度更高的花期预报模型。

### 参考文献

- [1]刘俊,李云云,刘浩龙,等.气候变化对成都桃花观赏旅游的影响与人类适应行为[J].地理研究,2016,35(3):504-512.
- [2]王连喜,陈怀亮,李琪,等.植物物候与气候研究进展[J].生态学报,2010,20(2):447-454.
- [3]徐雨晴,陆佩玲,于强.气候变化对植物物候影响的研究进展[J].资源科学,2004,26(1):129-136.
- [4]张倩,李新建,吉春容,等.基于逐步回归的库尔勒香梨始花期预测模型[J].沙漠与绿洲气象,2013,7(2):43-46.
- [5]刘维,侯英雨,吴门新,等.WOFOST模型在东北春玉米产区的验证与适应性评价[J].气象与环境科学,2017,40(3):8-14.
- [6]李彤霄.河南省气候变化对大豆生育期的影响研究[J].气象与环境科学,2015,38(2):24-28.
- [7]罗天虎.基于GIS的赤水市金钗石斛农业气候区划[J].气象与环境科学,2014,37(2):70-73.
- [8]王焕炯,戴君虎,葛全胜.1952-2007年中国白蜡树春季物候时空变化分析[J].中国科学:地球科学,2012,42(5):701-710.

- [9] Gonsamo A, Chen J M, Wu C Y. Citizen science: Linking the recent rapid advances of plant flowering in Canada with climate variability [J]. *Scientific Reports*, 2013, 3(7): 2239.
- [10] 张增信, 王言鑫, 杨艳蓉, 等. 南京地区 8 种常见观赏树木花期可预报研究[J]. *生态科学*, 2014, 33(4): 642-648.
- [11] Morin X, Gspurger M J, Carol, et al. Leaf phenology in 22 North American tree species during the 21st century[J]. *Global Change Biology*, 2010, 15(4): 961-975.
- [12] 叶海龙, 吴海镇. 气象因子预测油菜盛花期的探讨[J]. *浙江农业科学*, 2013, 1(9): 1080-1081.
- [13] 仲舒颖, 葛全胜, 戴君虎, 等. 中国典型观赏植物花期模型建立及过去花期变化模拟[J]. *资源科学*, 2017, 39(11): 2116-2129.
- [14] Chung U, Mack L, Yun J I, et al. Predicting the timing of cherry blossoms in Washington, DC and Mid-Atlantic states in response to climate change[J]. *Plos One*, 2011, 6(11): 27439.
- [15] Aono Y. Cherry blossom phenological data since the seventeenth century for Edo (Tokyo), Japan, and their application to estimation of March temperatures[J]. *International Journal of Biometeorology*, 2015, 59(4): 427-434.
- [16] 汪如良, 邓德文, 郭瑞鸽, 等. 鄱阳湖旅游地区油菜花花期的分析研究[J]. *江西科学*, 2015, 33(5): 708-712.
- [17] 傅玮东, 李兰, 顾雅文. 库车日平均气温  $\geq 3$   $^{\circ}\text{C}$  初日与杏树开花期的关系[J]. *沙漠与绿洲气象*, 2017, 11(4): 86-89.
- [18] 王琳, 郑育桃, 伍艳芳, 等. 气象因子对芳香植物始花期的影响[J]. *经济林研究*, 2017, 35(2): 194-199.
- [19] Jolliffe I T. Principal component analysis[J]. *Journal of Marketing Research*, 2002, 87(100): 513.
- [20] 舒晓惠, 刘建平. 利用主成分回归法处理多重共线性的若干问题[J]. *统计与决策*, 2004(10): 25-26.
- [21] 刘洪兰, 张俊国, 阙龙凯, 等. 基于 BP 神经网络的张掖国家湿地公园水域结冰厚度预报模型[J]. *高原气象*, 2014, 33(3): 832-837.
- [22] 张长卫. 基于 BP 神经网络的单站总云量预报研究[J]. *气象与环境科学*, 2009, 32(1): 68-71.
- [23] Durrani T S, Owens D H, Johnson M A, et al. Adaptive systems in control and signal processing 1989[M]. Amsterdam: Elsevier Ltd, 1990.
- [24] Hecht-Nielsen R. Theory of the backpropagation neural network[J]. *Neural Networks*, 1988, 1(1): 445.
- [25] 李鑫斐, 赵林. 人工神经网络在溶解度预测方面的应用[J]. *化学通报*, 2015, 78(3): 208-214.

## Application of BP Neural Network in the Prediction of Oilseed Rape Florescence

Sun Jiaqing, Zhang Zhiwei, Ai Wenwen

(Meteorological Service Center of Jiangsu Province, Nanjing 210009, China)

**Abstract:** Developing flowering phenological models is conducive to the accurate simulation of flowering periods of ornamental plants and could provide basis for seasonal flowering tourism events. This has become a new direction for the business of meteorological services. This paper analyzes the variation of oilseed rape florescence in Gaochun tourist area and explores its florescence forecasting methods to guide the tourism activities during the Gaochun rape flower festival. Based on meteorological observations, such as daily maximum temperature, daily minimum temperature, daily mean temperature, daily mean rainfall, daily sunshine hours, daily mean 5 cm soil temperature, daily mean air humidity and daily mean evaporation from 1985 to 2010, three principal components of meteorological elements including temperature factor, weather factor and radiation factor are obtained by using the principal component analysis (PCA) method. Then, taking these as input factors, we establish a prediction model of rape flower florescence based on BP neural network and discuss the application of BP neural network in the prediction of flower florescence. The results demonstrate that, compared with 4.25 d flowering forecast error by the traditional effective accumulated temperature method, the error by the BP neural network method is 1.5 d on average, so the BP neural network method has higher accuracy in predicting florescence of flower, and is easy to utilize. It can be widely used in the operation of florescence forecasting.

**Key words:** BP neural network; principal component analysis; oilseed rape florescence; prediction model