



# 不同土壤营养条件下棉花估产 RNN 模型研究

## Study on the Regression Neural Network Model for Cotton Yield Prediction under Different Soil Nutrition

苏里坦<sup>1,2</sup>, 宋郁东<sup>1</sup>, 张展羽<sup>2</sup>

(1. 中科院新疆生态与地理研究所, 乌鲁木齐 830011; 2. 河海大学水利水电工程学院, 南京 210098)

神经网络模型是一种基于生理学的智力仿生模型, 由大量神经元互联组成的非线性、大规模、自适应动力学系统。它具有自组织、自适应和自学习能力, 以及非线性、非局域性、非定常性和非凸性等特点。本文在 BP 网络的基础上, 加入偏差单元, 建立了回归神经网络(Regression Neural Network, RNN), 利用网络的内部状态反馈来描述系统的非线性动力学行为, 大大提高了学习速度。

### 1 材料和方法

#### 1.1 试验方案

采用田间小区试验, 重复 3 次, 随机排列, 小区面积 33 m<sup>2</sup>。试验方案按目前认为的最佳施肥量设计最佳处理, 然后设置减量和加量处理。N 肥各处理留 104.5 kg · hm<sup>-2</sup>, 2003 年 6 月 28 日灌 1 水前开沟追施。其余 N 肥与 P、K 肥在播种前一次施入作基肥。试验方案如表 1。

表 1 影响棉花产量的主要土壤营养因素

Table 1 Major nutrition factors for affecting cotton yield

kg · hm<sup>-2</sup>

处理	1	2	3	4	5	6	7	8	9	10	11	12
N	138	241.5	345	241.5	241.5	241.5	241.5	241.5	0	270	241.5	345
P <sub>2</sub> O <sub>5</sub>	138	138	138	138	138	60	207	138	0	140	0	60
K <sub>2</sub> O	60	60	60	0	120	60	60	120	0	0	60	60
对应产量	4099	4581	4393	3745	4423	3848	4157	4490	3277	4419	4245	3758

注: 表中棉花产量为子棉产量, 下同

#### 1.2 种植方式与田间管理

试验地点为新疆阿克苏地区新和县塔什力克乡阿其东村一组吾帕棉田, 土壤为灌淤土。供试棉花品种为 89-7, 窄膜种植, 株行距为 (60+30) cm × 12.5 cm。2003 年 4 月 8 日播种, 全生育期中耕 3 次, 化控 4 次, 灌水 3 次。

### 2 棉花估产回归神经网络模型

#### 2.1 模型的响应函数

在考虑带有偏差单元回归神经网络的权系数调整规则时, 可以借用 BP 算法。设 NI 和 NH 分别为输入节点数和隐含节点数(除偏差节点处) I<sub>j</sub>(k), 是带有偏差单元回归神经网络在时间 k 的第 j 个输入, x<sub>j</sub>(k) 是第 j 个隐层节点的输出,

y(k) 是带有偏差单元的回归神经网络的输出向量, 则带有偏差单元的回归神经网络可由下列数学公式描述:

$$y(k) = \sum_{j=1}^{NH} WO_j x_j(k) + WO_{bias} \quad (1)$$

$$x_j(k) = \sigma[S_j(k)] \quad (2)$$

$$S_j(k) = \sum_{i=1}^{NI} WR_{ij} x_i(k-1) + \sum_{i=1}^{NI} WI_{ij} I_i(k) + WI_{jbias} \quad (3)$$

式中  $\sigma[S_j(k)]$  是隐层节点的非线性激活函数, WI, WR, WO 分别为从输入层到隐层、回归信号、从隐层到输出层的权系数, WI<sub>bias</sub>、WO<sub>bias</sub> 分别为在隐层和输出层上的偏差单元的权系数。由方程(1)~(3)可以看出, 隐层节点的输出可视为动态系统的状态。带有偏差单元的回归神经网络的

隐层节点能够存储过去的输入输出信息。

这里采用  $S$ (Sigmoid) 函数作为网络响应函数,其数学表达式为:

$$f(x) = [1 + \exp(x)]^{-1} \quad (4)$$

## 2.2 模型中权重的初始化

在灌水量一定的条件下,直接影响棉花产量的土壤营养因素有: N、 $P_2O_5$ 、 $K_2O$ 。本文设定 3 层回归神经网络,输入层有 3 个节点,分别代表了 3 个影响因子;中间层有 1 个隐含层,其隐含节点数为 3 个;输出层有 1 个节点,代表公顷产子棉。因此,网络结构为 3-3-1 型。

## 3 结果与分析

本文主要探讨新和县棉花生产的合理 N、P、K 施用量和养分配比以及按照试验结果建立一个棉花产量预测的回归神经网络模型,该模型通过样本集训练好后,可以模拟试验结果并能预测不同肥料投入下的棉花产量,其模拟和预测精度相当高(表 2)。建立此模型旨在找出棉花生长中土壤养分氮、磷、钾的最佳配合比并确定其最高产量。

表 2 回归神经网络训练和检验结果对比  
Table 2 Comparison of training and testing results

kg · hm<sup>-2</sup>

处理号	实际产量	网络预测产量	误差值	相对误差/%
1	4099	4097.0371	+1.9629	0.048
2	4581	4582.3697	-1.3697	0.030
3	4393	4389.1016	+3.8984	0.089
4	3745	3747.3065	-2.3065	0.062
5	4423	4425.9270	-2.9270	0.066
6	3848	3846.1744	+1.8256	0.047
7	4157	4159.2311	-2.2311	0.054
8	4490	4485.6571	+4.3429	0.097
9	3277	3325.0852	+1.9148	0.058
10	4419	4686.1170	-267.1170	6.0
11	4245	4091.2048	+153.7952	3.6
12	3758	4050.9875	-292.9875	7.8

从表 2 可见,样本训练后的最大学习误差为 0.097%,将训练好的网络参数冻结后,即可应用于土壤营养因素的最佳配合量和棉花产量的估算中。该网络拟定的 N、P、K 最佳配合比为 1 : 0.60 : 0.29,即其肥料投入比例分别为 240 : 143 : 69;而试验观测所定的 N、P、K 最佳配合比为 1 : 0.57 : 0.25,其肥料投入比例分别

为 241.5 : 138 : 60。可见,试验观测和网络模拟的肥料配合比无太大差别,比较接近。本例将处理 10、11、12 中的土壤营养因素(氮、磷、钾)试验资料代入训练好的模型,得出了相应的棉花产量估算值如表 2(检验),与实测棉花产量相比,其误差不超过 10%,表明通过样本训练后的网络,可以应用于实际的棉花产量估算中。 ●