

玉米性状的计算机神经网络模型研究

王笑娟¹, 孟泉山², 肖杰¹, 王占森¹, 杜金昆¹, 张旭¹, 刘彩凤^{3*}

(1. 中国农业大学, 北京 100193; 2. 河北省蔚县农牧局, 河北张家口 075700; 3. 烟台职业学院, 山东烟台 264670)

摘要 利用神经网络模型研究了品种产量与其他性状的关系, 发现神经网络模型在不同次训练结果中的平均绝对误差和平均绝对相对误差的变异系数较小, 但在不同次训练结果中的各个自变量相对重要性的变异系数较大。因此, 神经网络较适合于产量的预测, 在用于研究各个性状对产量的相对重要性时, 应采取多次重复求平均值的方法, 以减少分析结果的误差。在所有的性状中, 穗粗、穗长对产量的相对重要性最大, 随后是百粒重、株高和生育期等。同时, 还分析了使产量最大时各个性状的最佳组合。该研究结果对黄淮海地区玉米育种目标的制定具有一定的参考价值。

关键词 玉米; 性状; 神经网络模型

中图分类号 S126 文献标识码 A

文章编号 0517-6611(2019)08-0228-03

doi: 10.3969/j.issn.0517-6611.2019.08.060



开放科学(资源服务)标识码(OSID):

Research on Computer Neural Network Model of Maize Traits

WANG Xiao-juan¹, MENG Quan-shan², XIAO Jie¹ et al (1. China Agricultural University, Beijing 100193; 2. Agriculture and Animal Husbandry Bureau of Wei County, Zhangjiakou, Hebei 075700)

Abstract In order to solve the problems of low level of information and lack of depth mining of experimental data in maize breeding, the relation of yield and other characters was conducted in this study by using neural network model. It was showed that the CV (the coefficient of variation) of the mean absolute error and the mean absolute relative error was small, and the CV of relative importance of each independent variable was larger. Therefore, the neural network model was suitable for production forecast. When applying in the researches on relative importance of each trait to yield, we should calculate the average by repetition, in order to to reduce the error of results. Among all the traits, ear diameter and ear length showed the greatest relative importance to yield, followed with 100-grain weight, plant height and growth period. Meanwhile, the optimal combination of the traits for the maximum yield was studied. The research results had certain reference value for the formulation of maize breeding targets in Huanghuaihai Area.

Key words Maize; Trait; Neural network model

神经网络模型是一个简单通用的可模拟复杂非线性关系的计算机模型, 只需要训练样本, 不需要建模或任何假定, 因而弥补了传统的回归模型缺乏对非线性问题研究能力的问题, 并在很多领域得到广泛的应用。提高产量是作物育种的首要目标。在玉米育种中, 了解产量及其相关性状之间的关系是十分必要的。但各种相关性状对产量的贡献大小不一, 因子间还存在着复杂的相互联系和影响, 这就给育种工作带来了许多困难。研究人员从利用相关、回归、通径、灰色关联等不同方法对影响玉米产量的因素和性状进行了探讨^[1-3], 但由于所在地区或所用玉米品种的不同, 以及研究方法的不同, 研究结果很不一致。同时, 目前在其他领域逐步得到广泛应用的计算机神经网络等分析方法, 在玉米育种试验数据的分析中应用不够。为了解决目前玉米育种试验数据得不到有效分析的实际问题, 该研究将应用各种计算机模型对玉米品种产量与其他性状之间的关系进行模拟, 探讨不同性状对产量的相对重要性, 以及在最高产量水平下各种性状的最佳组合, 以期对玉米育种目标和育种策略的制定提供指导和帮助。

1 材料与方

1.1 试验数据

试验数据为 2015 年黄淮海地区 26 个试验

点的玉米区试的各种性状数据资料, 包括来自不同育种单位的 80 个玉米品种。品种的选取兼顾了不同的育种单位和不同种质基础, 试验点则涵盖了黄淮海地区的各种生态类型, 因此试验材料和种植地点都具有较好的代表性。数据共计 1 420 组, 每组包括生育期性状, 株高、穗位等植株性状, 倒折株数、空秆株数、病指数等抗逆性性状, 以及穗粗、穗长、秃尖、穗行数、百粒重和产量等相关性状。

1.2 分析方法 以产量为因变量, 其他性状为协变量, 利用 SPSS 22.0 进行多层感知器神经网络模型的分析。由于各个性状的原始数据的量纲和级别不同, 首先要进行标准化处理。多层感知器神经网络模型的实际应用效果受训练样本数量、模型结构等因素影响较大。因此, 首先还要进行隐藏层数和训练样本/检验样本比例的优选。隐藏层数一般设置为 1 或 2 层, 训练样本/检验样本的比例设置 9:1、8:2、7:3、6:4、5:5、4:6、3:7、2:8、1:9 共 9 个处理。由于神经网络存在多解问题, 因此有必要对不同次训练结果之间预测精度的变异性以及各个自变量相对重要性的变异性进行研究。最后分析不同性状对产量的相对重要性, 以及在最高产量水平下各种性状的最佳组合。

2 结果与分析

2.1 神经网络模型结构的优选 隐藏层数和训练与检验样本分区比率对预测误差的影响见图 1。从图 1 可以看出, 当训练与检验样本分区比率 $\geq 7:3$ 时, 预测误差较小, 而且此时隐藏层数对预测误差影响不大。若训练与检验样本分区比率 $< 7:3$, 随着比率的减小, 误差逐渐增大, 而且此时 2 层隐藏

基金项目 北京市粮经作物产业创新团队资助项目(BAIC09-2018)。
作者简介 王笑娟(1971—), 女, 河北康保人, 助理工程师, 硕士, 从事信息技术与应用研究。* 通信作者, 讲师, 硕士, 从事计算机网络技术研究。

收稿日期 2018-11-04

层比 1 层隐藏层的预测误差要大。可见,训练分区样本的数量较少会影响训练的效果。因此,该试验选用训练、检验和预测样本分区比率为 5:2:3 和 1 层隐藏层的网络结构进行模拟。建立的多重传感器神经网络模型含有 1 个隐藏层,输入

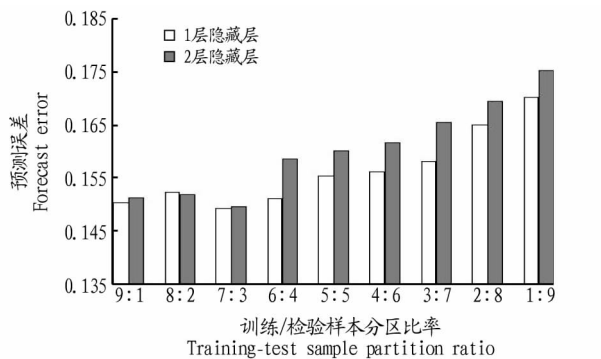


图 1 模型结构对预测误差的影响

Fig.1 Effects of model structure on forecast error

层有 10 个神经元节点,即自变量个数,输出层有 1 个节点,隐藏层自动设置 7 个节点,构成一个 10-7-1 的神经网络模型。模型的基本情况见表 1,系统自动生成的各个节点的权重系数见表 2。

表 1 多重传感器神经网络模型的基本情况

模型结构 Model structure	内容 Content	说明 Instruction
输入层 Input layer	变量数	10
隐藏层 Hidden layer	自变量的重新计算方法	标准化
	隐藏层的数目	1
	隐藏层的节点数	7
输出层 Output layer	启动函数	双曲正切函数
	变量数	1
	尺度相依的重新计算方法	标准化
	启动函数	恒等函数
	错误函数	平方和

表 2 各个节点的权重系数矩阵

Table 2 Weight coefficient matrix of each node

结构 Structure	项目 Item	隐藏层 Hidden layer							输出层 Output layer
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	
输入层 Input layer	偏差	0.087	0.740	0.124	0.231	0.211	0.361	-0.160	
	生育期	-0.133	-0.015	0.156	0.233	0.373	-0.486	-0.521	
	株高	0.250	-0.045	0.373	0.376	0.561	-0.536	-0.533	
	穗位	-0.235	-0.231	-0.363	-0.051	0.152	-0.115	-0.173	
	倒折株数	-0.280	-0.325	0.127	-0.303	-0.234	-0.208	0.230	
	空秆株数	-0.376	-0.155	-0.013	-0.598	-0.569	0.816	-0.385	
	穗粗	0.354	0.391	-0.011	-0.439	0.444	-0.110	-0.049	
	穗长	0.263	0.659	0.022	0.484	-0.227	0.038	-0.302	
	秃尖	0.135	0.120	0.208	-0.288	-0.447	0.305	-0.207	
	百粒重	0.753	0.001	-0.225	-0.150	0.666	-0.394	-0.244	
病指数	-0.575	0.204	-0.488	0.284	0.018	0.568	0.070		
隐藏层 Hidden layer	偏差								-0.289
	H(1:1)								-0.011
	H(1:2)								0.623
	H(1:3)								0.091
	H(1:4)								0.079
	H(1:5)								0.326
	H(1:6)								-0.675
	H(1:7)								0.348

2.2 不同次训练结果的变异度分析 模拟结果中预测误差和自变量重要性的变异度比较见表 3。从表 3 可以看出,不同次训练结果的平均绝对误差和平均绝对相对误差的变异系数较小,但不同次训练结果中各个自变量相对重要性的变异系数较大。因此,神经网络较适用于进行产量的预测,在用于研究各个性状对产量的相对重要性时,应采取多次重复求平均值的方法,以减少分析结果的误差。

2.3 各个性状的相对重要性 神经网络模型预测的各种性状的相对重要性见图 2。从图 2 可以看出,穗粗和穗长的相对重要性最高,其次是百粒重、株高和生育期,穗位和倒折株数的相对重要性中等,而秃尖、倾斜株数、病指数和穗行数的

相对重要性较低。

2.4 各个性状的最佳组合 3 种模型的预测结果中,产量最大时各个性状的最佳组合的结果基本一致,除了秃尖、倒折株数和空秆株数等性状的变异系数较大之外,其他性状的变异系数均较小,而这 3 个性状的取值都很小。产量最大时的各个性状的取值分别为生育期 103 d 左右、株高 300 cm 左右、穗位 120 cm 左右、倒折株数 2~3 株、空秆株数接近 0 株、穗粗 5.6 cm、穗长 20 cm、秃尖 0.5 cm、穗行数 16 行、百粒重 39 cm、病指数 1.0。在该合下,预测的最高产量为 9 750 kg/hm²左右。

表3 模拟结果中预测误差和自变量重要性的变异度比较

Table 3 Comparison of the variation degree of forecast error and independent variable importance in simulation results

测试编号 Test code	平均绝对误差 Mean absolute error	平均绝对相对误差 Mean absolute relative error	生育期 Growth period	株高 Plant height	穗位 Ear height	倒折株数 Lodging plants	空秆株数 Barren plants	穗粗 Ear width	穗长 Ear length	秃尖 Baretip	百粒重 100-grain weight	病指数 Disease index
1	73.012	0.149	0.061	0.089	0.059	0.060	0.196	0.159	0.138	0.085	0.098	0.054
2	70.743	0.151	0.060	0.113	0.039	0.080	0.183	0.195	0.101	0.063	0.112	0.053
3	73.043	0.150	0.101	0.079	0.048	0.053	0.240	0.139	0.096	0.095	0.105	0.044
4	72.617	0.148	0.086	0.120	0.065	0.058	0.213	0.145	0.058	0.080	0.105	0.069
5	71.952	0.150	0.084	0.076	0.045	0.073	0.209	0.145	0.105	0.088	0.105	0.070
6	72.076	0.147	0.087	0.120	0.071	0.051	0.190	0.146	0.095	0.081	0.093	0.066
7	72.586	0.150	0.058	0.146	0.075	0.060	0.218	0.109	0.138	0.070	0.077	0.048
8	72.073	0.148	0.093	0.115	0.090	0.050	0.199	0.125	0.100	0.077	0.092	0.058
9	69.759	0.144	0.090	0.103	0.042	0.051	0.210	0.158	0.108	0.087	0.097	0.052
10	72.245	0.150	0.077	0.134	0.049	0.069	0.184	0.119	0.099	0.085	0.111	0.072
平均值 Mean	72.011	0.149	0.080	0.110	0.058	0.061	0.204	0.144	0.104	0.081	0.100	0.059
均方差 Mean square error	1.028	0.002	0.015	0.023	0.017	0.010	0.018	0.024	0.023	0.009	0.011	0.010
变异系数 Variable coefficient // %	1.537	1.428	19.001	20.893	28.577	17.093	8.575	16.795	21.874	11.441	10.554	17.012

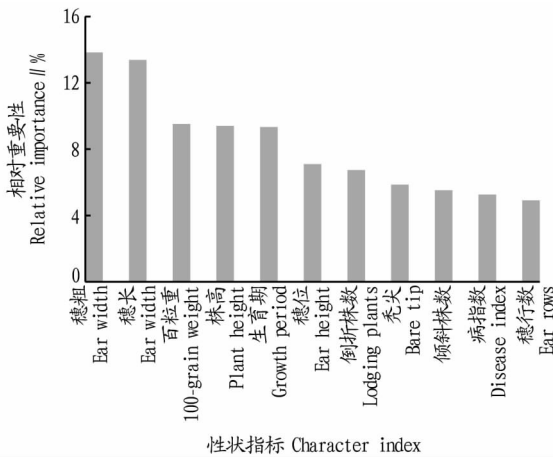


图2 各个性状的相对重要性

Fig.2 Relative importance of each character

3 结论与讨论

神经网络模型是近年来发展起来的一种变量关系模拟和预测方法,由于它具有良好的非线性映射能力和很强的自学习适应能力,因而为解决未知不确定性非线性关系问题提供了有效的方法^[4]。这些模型对育种家确定不同环境下的育种目标、比较不同育种方法的优劣、设计育种的选择方案和优化策略开展分子辅助选择育种研究提供了强有力的辅助工具。然而,作物育种学是一门实践科学,育种方法的研究来源于育种实践的需要。在以后的研究中,计算机模拟应与育种实践紧密结合,不断发现并解决育种工作中存在的问题,才能真正为育种家服务。

玉米产量除受自然环境影响外,遗传因素也具有很重要地位,在玉米育种工作中,通过研究主要农艺性状及其与产量的相互关系,找出影响产量的主要因素,可以为优良杂交种的选育提供选择依据^[5]。玉米产量是各农艺性状综合作用的结果,影响产量的因素很多,有些与产量关系密切,而有些作用较小^[6-7]。对玉米杂交种主要农艺性状的遗传相关及其与产量的关系已有很多报道,前人大多利用各自不同的育种材料,基于一种方法进行研究,因此结果各不相同^[8-9]。该

研究发现,在所有的性状中,穗粗、穗长对产量的相对重要性最大,随后是百粒重、株高和生育期等,这与张泽民等^[10]、丁山等^[11]和卓德众等^[12]的研究结果基本一致。其次,研究还发现,产量最大时各个性状的最佳组合为生育期 103 d 左右、株高 300 cm 左右、穗位 120 cm 左右、倒折株数 2~3 株、空秆株数接近 0 株、穗粗 5.6 cm、穗长 20 cm、秃尖 0.5 cm、穗行数 16 行、百粒重 39 cm、病指数 1.0 等,这较符合玉米育种目标的常规认识和思路。刘帆等^[13]、陈发波等^[14]、闫海霞等^[15]的结果与该研究也基本一致。针对特定的原始材料和黄淮海地区特殊的生态条件,对该玉米育种项目和黄淮海地区的玉米育种目标和育种策略的制定具有一定的参考价值。

参考文献

- [1] 鲁珊,肖荷霞,毛彩云,等.玉米杂交种主要农艺性状的相关和通径分析[J].安徽农业科学,2017,45(21):26-27,58.
- [2] 史新海,李可敬,孙为森,等.山东省不同年代玉米杂交种主要农艺性状演变规律的研究[J].玉米科学,2000,8(2):33-35.
- [3] 李晓花,林永明,谢淑芳,等.玉米产量与主要农艺性状的灰色关联度分析[J].云南农业科技,2016(1):13-15.
- [4] 焦李成,杨淑媛,刘芳,等.神经网络七十年:回顾与展望[J].计算机学报,2016,39(8):1697-1716.
- [5] 佟屏亚.中国玉米生产的发展方向——质疑“超级玉米”“超高产”[J].农业科技通讯,2005(8):11-13.
- [6] 崔俊明.新编玉米育种学[M].北京:中国农业科学技术出版社,2007:12-57.
- [7] 周得宝,王娟,王五洲,等.夏玉米品系(种)的产量比较及主要农艺性状的关联度分析[J].安徽农业科学,2017,45(26):48-51,78.
- [8] 何文昭.玉米产量和主要农艺性状杂种优势遗传分析[D].北京:中国农业科学院,2017.
- [9] 史新海,赵格.山东省紧凑型玉米杂交种主要农艺性状对产量的影响及其演变规律的研究[J].玉米科学,2003,11(2):59-61,85.
- [10] 张泽民,刘丰明,李雪英.河南省 1963~1993 年玉米杂交种籽粒产量及其组成性状的遗传增益[J].作物学报,1998,24(2):182-186.
- [11] 丁山,郭去,宋军,等.玉米主要性状与产量的回归模型及相关分析[J].西南农业学报,2008,21(5):1226-1230.
- [12] 卓德众,刘启华,郭红甫.灰色关联度分析法在玉米育种中的应用[J].玉米科学,1996,4(3):31-34.
- [13] 刘帆,石海春,余学杰.玉米果穗主要性状与产量间的相关与通径分析[J].玉米科学,2005,13(3):17-20.
- [14] 陈发波,杨克诚,荣廷昭,等.西南及四川区试玉米组合主要性状分析及育种对策探讨[J].玉米科学,2007,15(4):41-45.
- [15] 闫海霞,柳家友,吴伟华,等.含非洲血缘玉米杂交种主要穗部性状与单株产量的相关及通径分析[J].杂粮作物,2007,27(6):379-381.