

doi:10.6041/j.issn.1000-1298.2013.08.042

基于 RBF 神经网络的种猪体重预测*

刘同海^{1,2} 李卓¹ 滕光辉¹ 罗城¹

(1. 中国农业大学水利与土木工程学院, 北京 100083; 2. 天津农学院计算机科学与信息工程系, 天津 300384)

摘要: 针对猪体生长参数之间具有一定的自相关性、部分参数与体重间呈非线性关系、通过线性回归模型预测猪体体重存在着自变量间共线性及拟合优度较低等问题,以 52 头长白母猪的生长参数为基础,通过最近邻聚类算法,构建了基于 RBF 神经网络的种猪体重预测模型。通过线性回归检验法对种猪体重预测值与实测值进行分析,发现基于 RBF 神经网络的长白种猪体重预测模型的拟合优度 R^2 为 0.998,而线性回归模型的 R^2 为 0.891。结果表明:通过 RBF 神经网络方法建模,消除了线性回归分析中自变量的共线性问题,预测效果优于线性回归模型。

关键词: 种猪体重 径向基函数 神经网络 回归分析 预测

中图分类号: TP183; S818 **文献标识码:** A **文章编号:** 1000-1298(2013)08-0245-05

Prediction of Pig Weight Based on Radical Basis Function Neural Network

Liu Tonghai^{1,2} Li Zhuo¹ Teng Guanghui¹ Luo Cheng¹

(1. College of Water Resources and Civil Engineering, China Agricultural University, Beijing 100083, China

2. Department of Computer Science and Information Engineering, Tianjin Agricultural University, Tianjin 300384, China)

Abstract: There is a certain correlation among the growth parameters of pigs, and some parameters have non-linear relationship with pig weights. When using the simple linear regression model to predict the pig weight, the collinearity among independent variables and low fit goodness were found. For these problems, based on the nearest neighbor clustering algorithm for RBF neural network, a pig weight prediction RBF neural network model was constructed with the growth parameters of 52 Landrace sows. The predicted value and measured value of the pig weight were compared by linear regression test. The regression analysis showed that the goodness of fit (R^2) of RBF neural network prediction model for Landrace pig weight was 0.998, while R^2 of the linear regression model was only 0.891. The results indicated that the RBF neural network-based modeling method was an effective way to build the prediction model of pig weight. It eliminated the collinearity of the independent variables in linear regression analysis, and forecasted better than linear regression model.

Key words: Pig weight Radical basis function Neural network Regression analysis Prediction

引言

猪体体重是种猪培育及生猪养殖所关注的主要生长指标之一,也是评价种猪生长性状的关键指标^[1]。传统的猪体体重测量方法一般是采用体重箱、磅秤或电子秤等工具直接进行称量,工作量大,

费时费力,且对猪体应激大。已有研究表明^[2~11],动物体尺、体重等生长参数之间存在相关性,通过建立数学模型,可有效预估动物体重。其研究方法主要是进行线性或者非线性回归分析,建立一元或多元回归模型,通过体尺等生长参数预估动物体重。但用于估测体重的生长参数间存在着一定的自相关

收稿日期: 2012-08-09 修回日期: 2012-10-23

* 国家自然科学基金资助项目(31072066)、国家公益性行业(农业)科研专项资助项目(201003011)和中央高校基本科研业务费专项资金资助项目(KYCX2011082)

作者简介: 刘同海, 博士生, 天津农学院副教授, 主要从事智能畜牧业及其空间数据挖掘研究, E-mail: tonghai_1227@163.com

通讯作者: 滕光辉, 教授, 博士生导师, 主要从事农业生物环境智能控制和数字畜牧研究, E-mail: futong@cau.edu.cn

性,部分变量之间存在共线性问题^[12],且测量样本可能存在一些离群值,尤其是部分参数与体重间存在非线性关系,通过简单的线性回归分析,拟合度较低^[13~16]。而神经网络模型适用于非线性模型构建,具有对非线性函数最佳逼近和全局最优的性能。神经网络模型应用在动物体重估测方面,国内外鲜见报道。

径向基函数(Radical basis function, RBF)神经网络以其计算量小、学习速度快、不易陷入局部极小等优点为非线性建模提供了一种有效方法。本文在天津惠康种猪场选择120日龄左右的52头种猪,测量其体长、体高、体宽、臀宽、臀高、背膘厚、眼肌厚及体重。在统计分析的基础上,建立RBF神经网络模型,估测种猪体重,为利用双目视觉技术进行无接触种猪体重预估奠定模型基础。

1 材料与方法

1.1 数据获取

本试验于2012年3~7月在天津市惠康种猪育种有限公司的规模种猪场进行,试验对象为长白母猪,日龄120d左右。采用CHISON 500J型猪场专用背膘仪、皮尺、磅秤、手杖、卷尺等测量工具,获取种猪的背膘厚、眼肌面积、体重、体高、体宽、胸围、管围、臀宽、臀高等共计52组数据,并记录对应的猪耳号和猪体图像。

1.2 RBF神经网络模型原理

RBF神经网络是一种单向传播的、具有3层及以上神经网络,具有对非线性函数最佳逼近和全局最优的性能^[17]。RBF神经网络模型包括输入层、隐含层和输出层,隐含层可有一个或多个,每层由多个神经元组成。其隐含层神经元功能函数是高斯函数,又叫基函数,输出层神经元功能函数是高斯函数的线性组合。

RBF神经网络的输入层为 k 维向量 $\mathbf{X} = \{X_1, X_2, \dots, X_k\}$ 。隐含层为 l 维向量 $\mathbf{D} = \{D_1, D_2, \dots, D_l\}$,隐含层节点个数一般通过不断试验来确定,直到误差达到满意为止。网络的输出向量为 h 维向量 $f(\mathbf{X})$,其输出为

$$f(\mathbf{X}) = \sum_{i=1}^l W_{iu} D_i(\mathbf{X}) \quad (i=1, 2, \dots, l; u=1, 2, \dots, h) \quad (1)$$

式中 $D_i(\mathbf{X})$ ——隐含层中的径向基函数

u ——输出神经元个数

W_{iu} ——第 i 个隐含层神经元到输出神经元的权值

径向基函数一般选择高斯函数,即

$$D_i(\mathbf{X}) = \exp\left(-\frac{\|\mathbf{X} - \mathbf{c}_i\|^2}{2\sigma_i^2}\right) \quad (2)$$

式中 $\|\mathbf{X} - \mathbf{c}_i\|$ ——欧氏范数

\mathbf{c}_i ——网络隐含层节点中心

σ_i ——径向基函数方差或宽度,用来调节网络的灵敏度

在径向基函数中,采用基于 K 均值聚类方法求取基函数的中心 \mathbf{c}_i 。随机选取 l 个训练样本作为聚类中心 \mathbf{c}_i ;将输入的训练样本集合按照最近邻规则分组,根据 \mathbf{X} 与中心为 \mathbf{c}_i 之间的欧式距离,将 \mathbf{X} 分配到输入样本的各个聚类集合 $A_b (b=1, 2, \dots, p)$ 中;计算各个聚类集合 A_b 中训练样本的平均值,重新调整聚类中心,即形成新的类中心 \mathbf{c}_i ,如果新的类中心不再发生变化,则所得的 \mathbf{c}_i 即为RBF神经网络最终的基函数中心 \mathbf{c}_{\max} 。

σ_i 的求解公式为

$$\sigma_i = \frac{c_{\max}}{\sqrt{2l}} \quad (3)$$

利用最小二乘法计算得到隐含层各神经元到输出层的权值为

$$W_{iu} = \exp\left(\frac{l}{c_{\max}^2} \|\mathbf{X} - \mathbf{c}_i\|^2\right) \quad (4)$$

本文的RBF神经网络模型的构建和训练是通过SPSS统计分析软件实现的。

1.3 预测结果验证方法

预测值与实测值吻合程度检验的常用方法有误差百分比法和线性回归法。本文采用线性回归法,线性回归法是将实测值作为被解释变量,将预测值作为解释变量,建立一元线性回归方程: $y = ax + b$ 。对回归方程进行拟合优度检验。若 R^2 接近1,或线性回归方程斜率 a 越接近1、 b 值越接近零,表示两者之间吻合度越高。

2 结果与分析

2.1 数据预处理

利用SPSS软件,对猪体实测数据进行相关分析,并使用皮尔逊双侧检验法检验其显著性,统计结果如表1所示。

从表1中可知,胸围、眼肌厚、臀高、臀宽、体长、体高、管围、背膘厚等8个参数值与猪体重在0.01水平下显著相关,其中体重与胸围、眼肌厚、臀高和臀宽等参数相关性较强,相关系数均在0.7以上,可以通过胸围、臀高等变量,建立线性模型,估算种猪体重。自变量间的相关性水平较高,眼肌厚、体长和胸围呈极显著相关;体重、臀宽、臀高、体高、体宽和眼肌厚与胸围在0.01水平下显著相关;而胸围与

表1 皮尔逊相关系数统计

Tab.1 Pearson correlations

	体重	体长	体高	体宽	胸围	管围	臀宽	臀高	眼肌厚	背膘厚
体重	1	0.586**	0.515**	0.412*	0.749**	0.490**	0.709**	0.718**	0.746**	0.479**
体长		1	0.046	0.091	0.327*	0.193	0.278	0.581**	0.437**	-0.116
体高			1	0.378*	0.607**	0.341*	0.326	0.653**	0.265	0.311
体宽				1	0.578**	0.402*	0.629**	0.612**	0.137	0.140
胸围					1	0.396*	0.705**	0.670**	0.463**	0.403*
管围						1	0.418	0.425	0.302	-0.048
臀宽							1	0.347	0.204	-0.100
臀高								1	0.239	0.025
眼肌厚									1	0.138
背膘厚										1

注: **为0.01水平下显著相关; *为0.05水平下显著相关。

背膘厚在0.05水平下显著相关。可见,估算体重时,自变量间存在一定的共线性问题。

2.2 线性回归分析

通过上述相关分析,本文选取与猪体体重相关性达极显著水平的胸围、眼肌厚、臀高、臀宽、体长、体高、管围、背膘厚等8个生长参数,利用最小二乘法进行多元线性回归分析,分析结果显示拟合优度 R^2 为0.935。但 R^2 的数值与自变量的数目有关,本文使用了8个自变量,为避免由于自变量数目对 R^2 评价能力的降低,本文的评价模型对体重的解释程度是利用剔除自变量数目影响后的修正 R^2 ,其值为0.891,拟合优度较高。以臀高、背膘厚、眼肌厚、臀宽、管围、体长、体高、胸围为自变量,体重为因变量,对其进行方差分析和回归方程系数显著性检验,结果如表2、3所示。

表2 方差分析

Tab.2 Analysis of various

方差来源	平方和	自由度	均方	F	显著性
回归	1 745.646	8	218.206	21.497	0
残差	121.806	43	10.151		
总和	1 867.452	51			

从方差分析(表2)可知,通过F检验,回归方程整体显著。但回归方程系数显著性检验中(表3),除了常数显著外,其他系数均不显著。可见,通过多元线性回归预测体重,存在一定误差。如果容忍度小于0.5,而膨胀系数大于2时,自变量间的相关可能性就很大。通过观察容忍度和变异量膨胀系数(VIF)发现,体长、体高、胸围、臀高和臀宽等自相关较大。

2.3 BRF神经网络建模与分析

2.3.1 网络设计

在相关分析的基础上,选取与猪体体重相关性达极显著水平以上的胸围、眼肌厚、臀高、臀宽、体

表3 回归方程系数显著性及共线性检验

Tab.3 Significance test of coefficients and collinearity statistics of independent variable

参数	回归方程系数	标准误差	t 检验	显著性	容忍度	VIF
常量	-213.728	26.112	-8.185	0		
背膘厚	1.025	0.488	2.100	0.058	0.679	1.473
眼肌厚	0.423	0.176	2.396	0.034	0.771	1.296
体长	0.361	0.147	2.458	0.030	0.519	1.928
体高	0.531	0.473	1.123	0.283	0.463	2.162
胸围	0.858	0.424	2.023	0.066	0.195	5.131
管围	1.663	1.144	1.454	0.172	0.630	1.587
臀宽	2.656	1.289	2.060	0.062	0.343	2.917
臀高	0.280	0.639	0.438	0.669	0.241	4.154

长、体高、管围、背膘厚8个猪体生长参数,作为网络输入样本,即 $k=8$,用矢量 $\mathbf{X}=\{X_1, X_2, \dots, X_8\}$ 表示;将猪体体重实测值 $f(\mathbf{X})$ 作为网络唯一输出值,即 $u=1$;通过自动计算范围,在某个范围内查找最佳单位数的方法,并进行多次网络训练,基于网络均方误差,确立隐含层数,建立输入层8个神经元、输出层1个神经元的RBF神经网络。

2.3.2 网络训练

将实测的52组数据中的32组数据作为训练样本,由于各个自变量之间单位的不一致,先将训练样本数据利用均值和标准差进行标准化处理,然后设定网络均方误差目标值,确定隐含层神经元的最大和最小个数。利用标准化径向基函数,调整RBF神经网络的隐含层神经元个数,直到均方误差在期望误差范围内为止。经试验,隐含层神经元个数为17时,RBF网络的训练样本误差平方和为 5.410×10^{-30} ,相对误差为 6.763×10^{-31} ;测试样本误差平方和为0.018,相对误差为0.045,误差满足精度要求,逼近效果最好。在训练模型的输入层中,胸围、眼肌厚、臀高、臀宽、体长、体高、管围、背膘厚8个神经元到隐含层的17个神经元权值及隐含层各神经元到输

出神经元猪体体重线性函数各个权值如表4所示。

表4 输入层、隐含层及输出层各神经元的权值计算结果

Tab.4 Weight of neurons in input layer, hidden layer and output layer

隐含层	背膘厚	眼肌厚	体长	体高	胸围	管围	臀宽	臀高	W_{iu}
1	0.473	0.288	0.512	2.325	1.440	0.183	1.126	1.387	1.815
2	-0.063	0.288	1.186	0.558	1.174	0.803	2.133	0.987	2.897
3	-0.063	0.502	0.377	1.442	0.642	2.045	-0.889	1.786	0.992
4	-1.134	-0.351	2.265	0.117	1.706	-0.438	1.126	1.387	0.595
5	-0.063	0.502	-0.567	0.117	0.908	0.803	1.126	0.188	0.731
6	-0.063	1.781	-0.298	1.000	0.642	0.803	1.126	0.588	-0.172
7	-1.670	-1.417	-0.163	-0.325	-1.753	-0.438	-0.889	0.188	-1.774
8	-1.670	1.781	-0.095	-0.766	-0.954	-1.680	-0.889	-0.611	-1.202
9	-1.134	-0.564	-2.590	-0.325	-1.221	-0.438	0.119	-1.810	-1.456
10	1.544	0.502	-0.972	0.117	0.110	0.803	0.119	-0.212	0.303
11	1.008	-0.991	0.377	-1.208	-0.156	-0.438	0.119	-1.410	1.153
12	-0.599	-0.564	0.512	0.117	-0.156	0.803	0.119	-1.011	-0.853
13	-0.063	0.715	-0.028	-1.208	-0.156	-0.438	-0.889	-0.611	1.574
14	0.473	-0.991	-0.028	-0.766	-1.221	-0.438	-0.889	-0.611	-3.780
15	1.544	0.075	-0.702	-1.649	0.110	-1.059	-0.889	-0.212	-1.307
16	1.008	-1.844	-0.028	0.338	-0.954	-1.680	-0.889	-0.212	0.204
17	0.473	0.288	0.242	0.117	-0.156	0.803	-0.889	0.188	0.553

2.3.3 模型验证

为了验证模型的准确性,用预留的20组数据测试训练完成后的RBF神经网络模型。将20组测试样本的数据输入到模型中,利用建好的神经网络模型进行预测分析,并与实测值进行比较分析,以检验RBF神经网络模型对猪体重预测的准确性和稳定性。本文通过线性回归法对预测结果和实测值的检验结果 R^2 为0.998,接近1。可见预测值与实测值拟合优度高,该模型很好地解释了猪体体重实测值变量。该值明显高于多元线性回归分析的 R^2 值0.891,较多元线性回归模型具有较高的拟合优度。RBF神经网络模型检验结果如图1所示。

3 结论

(1)在对52头长白母猪生长参数统计分析的基础上,通过最近邻聚类算法,构建了基于RBF神经网络的长白种猪体重预测模型,预估其体重。

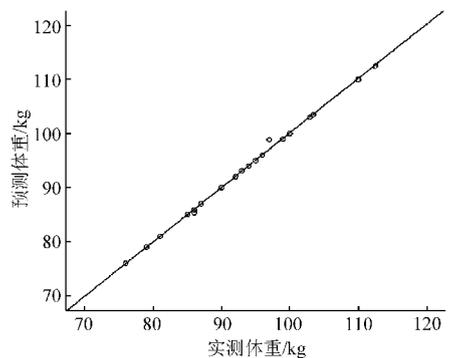


图1 RBF神经网络模型预测值与实测值回归分析

Fig.1 Regression results of RBF neural network

(2)用预留测试样本对模型进行验证,并与实测值进行回归分析,并对回归方程进行拟合优度检验,发现基于RBF神经网络的长白种猪体重预测模型的拟合优度 R^2 为0.998,预测效果优于线性回归模型。基于RBF神经网络方法建模,消除了线性回归分析中的原始自变量的共线性问题,是构建猪体体重预测模型的一种有效方法。

参 考 文 献

- 曹果清,景利芳,石建忠,等.猪体重和体尺性状与繁殖性能间的相关[J].山西农业大学学报,2003,23(2):136~138.
Cao Guoqing, Jing Lifang, Shi Jianzhong, et al. Study on correlation analysis between body weight and size and reproductive performance in swine[J]. Journal of Shanxi Agricultural University, 2003,23(2):136~138. (in Chinese)
- 徐宁迎,王先明,曹小英,等.长白猪主要选育性状间的典型相关分析[J].浙江大学学报:农业与生命科学版,2000,26(5):563~566.
Xu Ningying, Wang Xianming, Cao Xiaoying, et al. Canonical correlation analysis of main breeding traits for Landrace[J]. Journal of Zhejiang University: Agric. & Life Sci., 2000,26(5):563~566. (in Chinese)
- 侯建君,施正香,李保明.不同生长阶段白猪体型特征的主成分分析[J].中国农业大学学报,2006,11(3):56~60.
Hou Jianjun, Shi Zhengxiang, Li Baoming. Principal component analysis of Landrace's physical characteristics[J]. Journal of

- China Agricultural University, 2006,11(3): 56 ~ 60. (in Chinese)
- 4 李明丽,连林生,鲁绍雄,等. 撒坝猪部分生长发育性状的典型相关分析[J]. 家畜生态学报,2007,28(3):20 ~ 22,52.
Li Mingli, Lian Linsheng, Lu Shaoxiong, et al. Canonical correlation analysis on some growth traits of Saba pig[J]. Acta Ecologiae Animalis Domastici, 2007,28(3):20 ~ 22,52. (in Chinese)
 - 5 Keskin I, Dag B. Comparison of the linear and quadratic models for describing the growth of live weight and body measurements in Anatolian Merino male lambs during the fattening period[J]. Journal of Animal and Veterinary Advances,2006,5(1): 81 ~ 84.
 - 6 Zaky H I, Amin E M. Estimates of genetic parameters for body weight and body measurements in Bronze turkeys (Baladi) by using animal model [J]. Egyptian Poultry Science Journal,2007,27(1):151 ~ 164.
 - 7 Alade N K, Raji A O, Atiku M A. Determination of appropriate model for the estimation of body weight in goats [J]. Journal of Agricultural and Biological Science,2008,3(4):52 ~ 57.
 - 8 Ogah D M, Alaga A A, Momoh M O. Use of factor analysis scores in multiple regression model for estimation of body weight from some body measurement in Muscovy duck[J]. International Journal of Poultry Science, 2009,8(11): 1 107 ~ 1 111.
 - 9 Raji A O, Igwebuikwe J U, Kwari I D. Regression models for estimating breast, thigh and fat weight and yield of broilers from non invasive body measurements[J]. Agriculture and Biology Journal of North America,2010,1(4):469 ~ 475.
 - 10 Erat Serkan. Application of linear, quadratic and cubic regression models to predict body weight from different body measurements in domestic cats[J]. International Journal of Agriculture and Biology,2011,13(3):419 ~ 422.
 - 11 Chitra R, Rajendran S, Prasanna D, et al. Prediction of body weight using appropriate regression model in adult female Malabari goat[J]. Veterinary World,2012,5(7): 409 ~ 411.
 - 12 满敬鑫,杨薇. 基于多重共线性的处理方法[J]. 数学理论与应用,2010,30(2): 105 ~ 109.
Man Jingxuan, Yang Wei. Based on multiple collinearity processing method[J]. Mathematical Theory and Applications, 2010, 30(2):105 ~ 109. (in Chinese)
 - 13 周振勇,张杨,蔺宏凯,等. 基于主成分逐步回归法的新疆褐牛体重预测模型研究[J]. 中国牛业科学,2012,38(1):1 ~ 4.
Zhou Zhenyong, Zhang Yang, Lin Hongkai, et al. Weight model based on regression of principle component analysis of Xinjiang brown cattle[J]. China Cattle Science,2012,38(1):1 ~ 4. (in Chinese)
 - 14 蔡治华. 最小二乘法在黄牛体重估测中的应用[J]. 中国牛业科学,2000,26(5):1 ~ 4.
Cai Zhihua. Application of the least square method in estimating cattle body weight[J]. China Cattle Science, 2000,26(5):1 ~ 4. (in Chinese)
 - 15 张元跃,贺喜. 多层非线性混合效应模型评价番鸭体重生长[J]. 中国农业科学,2012,45(3):540 ~ 547.
Zhang Yuanyue, He Xi. Multilevel nonlinear mixed effect model for evaluation of muscovy ducks body weight for growth[J]. Scientia Agricultura Sinica, 2012,45(3):540 ~ 547. (in Chinese)
 - 16 付为森,滕光辉,杨艳. 种猪体重三维预估模型的研究[J]. 农业工程学报,2006,22(增刊2):84 ~ 87.
Fu Weisen, Teng Guanghui, Yang Yan. Research on three-dimensional model of pig's weight estimating[J]. Transactions of the CSAE, 2006,22(Supp. 2):84 ~ 87. (in Chinese)
 - 17 何晓群. 多元统计分析[M]. 北京:中国人民大学出版社,2004.