第 39 卷第 12 期
2015 年 12 月

水 产 学 报
JOURNAL OF FISHERIES OF CHINA

Vol. 39, No. 12
Dec., 2015

文章编号:1000 - 0615(2015)12 - 1799 - 08　　DOI:10.11964/jfc.20150309788

基于径向基函数(RBF)神经网络的红鳍东方鲀体质量预测

王新安[1,2], 马爱军[1,2*], 赵艳飞[1], 岳 亮[1], 孙建华[1,2],
孟雪松[3], 刘圣聪[3]

(1. 中国水产科学研究院黄海水产研究所,农业部海洋渔业可持续发展重点实验室,
青岛市海水鱼类种子工程与生物技术重点实验室,山东 青岛 266071;
2. 青岛海洋科学与海洋技术国家实验室生物与技术功能实验室,青岛 266071;
3. 大连天正实业有限公司,辽宁 大连 116000)

摘要: 为解决基于表型性状预测红鳍东方鲀体质量时,由于不同表型性状间的自相关、部分性状和体质量之间的非线性关系以及线性回归方法自变量间的共线性,导致根据表型性状预测体质量误差过大的问题,本研究根据人工神经网络(artificial neural networks,ANN)建模原理,采用径向基函数(radial basis function,RBF)神经网络模型,利用 72 个红鳍东方鲀样本的表型数据,通过最近邻聚类算法,构建了基于 RBF 神经网络的红鳍东方鲀体质量预测模型,并采用线性回归检验法对所构建模型的可信度进行检验。结果显示,基于 RBF 神经网络预测模型的确定系数 R^2 为 0.992,接近于 1,而线性回归模型的确定系数为 0.949,比线性回归预测模型的确定系数提高 4.53%;此外,线性回归共线性检验显示,模型自变量间存在一定的共线性,性状体周长 1 和体高的自相关性较大。研究结果表明,通过 RBF 神经网络方法构建的预测模型,消除了线性回归分析中自变量的共线性问题,预测精度显著高于线性回归方法构建的预测模型,对红鳍东方鲀体质量的预测效果优于线性回归模型。基于 RBF 神经网络体质量预测模型的构建,为利用表型性状精确评估红鳍东方鲀的体质量提供了一种新的方法。

关键词: 红鳍东方鲀;体质量预测;径向基函数;神经网络

中图分类号: S 917.4　　**文献标志码:** A

在动物育种中,体质量是需要遗传改良的最重要的经济性状,也是评估选育进展的关键指标。对于陆生动物,可以采用体质量箱或电子秤等常规测量工具对其体质量直接测量获得相对精确的测量值,而对于水生动物,由于生存环境特殊,将其从水中取出进行测量时,鱼体表面带有较多的水分,导致测量结果的误差较大,另外,对于部分野生鱼类,一些操作可能需要在船上进行,由于船的波动对于体质量测量结果有一定影响;如果将其在水外放置较长时间以去掉水分,则会对大多数水生动物产生较强的胁迫性,在一定时期内影响摄食、生长,甚至导致死亡。数量遗传学研究表明,生物的生长性状受遗传基因调控,不同基因之间也存在着遗传连锁关系[1],因而可利用动物表型性状和体质量之间的相关性,通过建立数学模型,实现有效预估水生动物体质量的目的。

有关水生动物表型性状和体质量之间关系的数学模型构建,常规的方法是进行线性或者非线性回归分析,建立一元或多元回归方程,通过表型性状预估动物体质量[2-6]。然而,由于估测体质量的表型性状之间存在一定的自相关性、部分变量之间存在共线性[7],以及部分性状与体质量间存在非线性,通过线性回归预测体质量,会存在较大的误差,结果并不理想[8-11]。而具有对非线性

收稿日期:2015-03-23　　**修回日期:**2015-09-05

资助项目:国家现代农业技术体系资助项目(CARS - 50 - G01);中央级公益性科研院所基本科研业务费专项(20603022012005);大连金州新区科技计划(2012 - B1 - 012)

通信作者:马爱军,E-mail:maaj@ ysfri. ac. cn

函数最佳逼近和全局最优性能的神经网络模型则非常适用于非线性模型构建,很好地解决了这一问题。神经网络模型应用在畜牧体质量估测方面已有研究[12],但在国内外水生动物领域尚未见有报道。在神经网络模型中,径向基函数(radial basis function,RBF)神经网络具有计算量小、学习速度快、不易陷入局部极小等诸多优点[13-17],为非线性系统辨识与模型提供了一种有效的手段。本研究采用 RBF 神经网络方法,估测红鳍东方鲀(*Takifugu rubripes*)体质量,为利用表型性状预估红鳍东方鲀体质量提供参考资料。

1 材料与方法

1.1 实验材料

本实验于大连天正实业有限公司进行,实验用红鳍东方鲀为来源于日本进口鱼卵所培育的同一批次苗种。从众多个体中随机采集红鳍东方鲀样品 72 尾。逐尾用游标卡尺测量全长(TL)、体长(BL)、体高(BD)、头长(HL)、眼后头长(EH)、吻长(SL)、口宽(MW)、眼径(ED)、眼间距(IS)、尾柄长(CPL)、尾柄高(CPD)、尾柄宽(CPB)、体宽(BW)、躯干长(TR)、尾长(TA)、体周长 1(BG1)(沿背鳍前环躯干长度)、体周长 2(BG2)(沿胸鳍后环躯干长度)共 17 个形态性状,精确到 0.01 cm。体质量(BWH)用电子天平称量,精确到 0.01 g。

1.2 分析方法

线性回归分析 对上述测定资料以体质量(BWH)为依变量,以其他测定的形态性状为自变量,运用逐步引入-剔除法(stepwise)进行多元回归分析,并基于容许值(*Tol*)和方差膨胀因子(*VIF*)以及最大条件指数(maximum condition index)和最大方差比(maximum variance proportions)进行共线性检验。线性回归分析及共线性检验采用 SPSS 15.0 软件处理。引入自变量系数和常数项显著性检验结果,当 $P<0.05$ 时为显著差异,当 $P<0.01$ 时为极显著差异。

RBF 神经网络 RBF 神经网络是一种以函数逼近理论为基础的高效前馈式神经网络,具有其他前向网络所不具有的对非线性函数的最佳逼近性能和全局最优特性[18]。RBF 神经网络模型的原理为 RBF 神经网络由输入层、径向基层(也称隐含层)和输出层 3 层组成,以 RBF 作为隐单元的"基"构成隐含层空间;隐含层对输入矢量进行变换将低维的模式输入数据变换到高维空间内,使得在低维空间内的线性不可分问题在高维空间内线性可分。通常采用高斯函数作为径向基神经元的传递函数(又称基函数)。RBF 神经网络的输入层为 k 维向量 $X=\{X_1,X_2,\cdots,X_k\}$,隐含层为 l 维向量 $D=\{D_1,D_2,\cdots,D_l\}$,通过连续实验确定隐含层节点个数当误差达到理想状态时,实验结束。RBF 神经网络输出向量 $f(X)$ 的表达式为:

$$f(X)=\sum_{i=1}^{l}W_{iu}\exp\left(-\frac{\|X-c_i\|^2}{2\sigma_i^2}\right)$$

式中,$\|X-c_i\|$、c_i、σ_i、u 和 W_{iu} 分别表示欧氏范数、网络隐含层节点中心、用来调节神经网络灵敏度的径向基函数方差、隐含层中的径向基函数、输出神经元个数和第 i 个隐含层神经元到输出神经元的权值。

径向基神经元的传递函数选择高斯函数,其表达式:

$$D_i(X)=\exp\left(-\frac{\|X-c_i\|^2}{2\sigma_i^2}\right)$$

在 RBF 中,基于最近邻聚类算法求径向基神经元传递函数的中心 c_i。随机选取 i 个训练样本作为聚类中心 c_i;根据最近邻规则将输入的训练样本集合分组,根据 X 与中心 c_i 之间的欧几里得距离,将 X 分配到输入样本的各个聚类集合 A_b 中,并计算各 A_b 中样本均值,重新调整聚类中心,即得到新的类中心 c_i,当新的类中心保持恒定时,此时的 c_i 即为最终基函数中心 $c_{\max}$。

径向基函数方差 σ_i 的表达式为 $\sigma_i=c_{\max}/\sqrt{2l}$;隐含层各神经元到输出层权值的表达式:

$$W_{iu}=\exp\left(\frac{l}{c_{\max}^2}\|X-c_i\|^2\right)$$

为了检验所构建模型的可信度,需要对预测结果进行验证。通常采用误差百分比法和线性回归法对预测值与实测值的吻合程度进行检验。本研究采用线性回归方法,将实测值和预测值分别作为被解释变量和解释变量,构建一元回归方程:$y=ax+b$,并对此方程进行确定系数 R^2 检验。

2 结果与分析

2.1 回归分析的共线性检验

根据测定资料运用逐步引入-剔除法进行多

元回归分析。按照表型性状对体质量作用的显著程度,依次引入体周长 1(BG1)、全长(TL)、体高(BD)、尾柄高(CPD)、眼间距(IS)和眼后头长(EH),此时,确定系数 $R^2=0.949$。回归结果发现,在回归方程系数显著性检验中,对所建立的线性模型,除眼后头长的系数达到显著水平($P<0.05$),其他各项系数和常数项均达到极显著水平($P<0.01$)(表 1)。基于容许值(Tol)和方差膨胀因子(VIF)进行共线性检验,各形态性状都没有超过 $VIF<10$ 和 $Tol>0.10$(表 1),表明所建模型的共线性没有达到极其严重的程度[19];采用最大条件指数和最大方差比进行共线性检验,最大条件指数为89.909>10,且 TL 和 BD 的最大方差比分别为 0.65 和 0.52,均大于 0.5(表 2),表明模型自变量间存在一定的共线性;体周长 1 和体高的方差膨胀因子(VIF)分别为 2.646 和 2.107,均大于 2,而这两个性状相应的容许值(Tol)分别为 0.378 和 0.475,均小于 0.5。这表明,所建模型自变量间存在一定的共线性,其中体周长 1 和体高自相关较大。

表 1　回归方程系数显著性及基于容许值和方差膨胀因子的共线性检验

Tab. 1　Significance test of coefficients and collinearity statistics of independent variable based on tolerance (Tol) and variance inflation factor (VIF)

变量 variable	回归方程系数 regression coefficient	标准误差 standard error	t – test	显著性(Sig.) significance	共线性检验 collinearity diagnostics	
					容许值 Tol	方差膨胀因子 VIF
常数 constant	–2198.136	106.779	–20.586	0.000	–	–
体周长 1/cm BG1	42.754	3.423	12.489	0.000	0.378	2.646
全长/cm TL	33.140	3.109	10.660	0.000	0.642	1.558
体高/cm BD	63.291	12.337	5.130	0.000	0.475	2.107
尾柄高/cm CPD	41.928	12.581	3.333	0.001	0.548	1.824
眼间距/cm IS	47.067	15.922	2.956	0.004	0.586	1.706
眼后头长/cm EH	–27.175	13.351	–2.035	0.046	0.755	1.324

表 2　基于最大条件指数和最大方差比的共线性检验

Tab. 2　Collinearity statistics of independent variable based on maximum condition index and maximum variance proportions

维度 dimension	特征值 eigenvalue	条件指数 condition index	方差比 variance proportions						
			常量 constant	体周长 1/cm BG1	全长/cm TL	体高/cm BD	尾柄高/cm CPD	眼间距/cm IS	眼后头长/cm EH
1	6.985	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.006	35.481	0.00	0.05	0.01	0.10	0.00	0.00	0.46
3	0.003	45.174	0.00	0.00	0.01	0.19	0.36	0.01	0.25
4	0.003	50.444	0.16	0.10	0.07	0.00	0.18	0.05	0.23
5	0.002	68.099	0.20	0.09	0.05	0.03	0.00	0.73	0.00
6	0.001	82.302	0.16	0.56	0.21	0.35	0.40	0.09	0.04
7	0.001	89.909	0.47	0.19	0.65	0.52	0.05	0.13	0.02

2.2　BRF 神经网络建模与分析

网络设计　　根据性状间的相关性分析,选取与红鳍东方鲀体质量相关性达到极显著水平的全长(TL)、体长(BL)、体高(BD)、眼后头长(EH)、口宽(MW)、眼径(ED)、眼间距(IS)、尾柄高(CPD)、尾柄宽(CPB)、体宽(BW)、躯干长(TR)、尾长(TA)、体周长 1(BG1)和体周长 2(BG2) 14 个表型性状,作为网络输入样本,即 $k=14$,用矢量 $X=\{X_1,X_2,\cdots,X_{14}\}$ 表示(表 3);将红鳍东方鲀体质量实测值 $f(X)$ 作为网络唯一输出值,即 $u=1$;基于自动计算范围,在某个范围内查找最佳单位数的方法,反复进行网络训练,根据网络均方误差,确立隐含层数,建立输入层 14 个神经元、输出层 1 个神经元的 RBF 神经网络。

表 3 表型性状与体质量的相关系数

Tab. 3 The phenotype correlation coefficient between phenotypic traits and body weight

性状 traits	体质量/g BWH	性状 traits	体质量/g BWH	性状 traits	体质量/g BWH
全长/cm TL	0.622**	口宽/cm MW	0.549**	体宽/cm BW	0.809**
体长/cm BL	0.366**	眼径/cm ED	0.401**	躯干长/cm TR	0.580**
体高/cm BD	0.688**	眼间距/cm IS	0.614**	尾长/cm TA	0.275**
头长/cm HL	0.218*	尾柄长/cm CPL	0.247*	体周长 1/cm BG1	0.878**
眼后头长/cm EH	0.338**	尾柄高/cm CPD	0.653**	体周长 2/cm BG2	0.742**
吻长/cm SL	0.197*	尾柄宽/cm CPB	0.650**	体质量/g BWH	1

注：* 表示差异显著（$P<0.05$），** 表示差异极显著（$P<0.01$）

Notes: * means significant difference ($P<0.05$), ** means much significant difference ($P<0.01$)

网络训练　利用所采集数据中的 52 组数据作为训练样本。由于各输入自变量的量纲不一致，因此，在建模之前先将训练样本数据进行归一化处理，然后设定 RBF 神经网络的均方误差目标值，确定隐含层神经元的个数范围。基于归一化的径向基函数，对 RBF 神经网络的隐含层神经元个数进行调整，通过在测试数据集上的误差平方和的对比，确定最佳的隐含层神经元个数。经过实验，隐含层神经元个数为 10 时，RBF 神经网络的训练样本误差平方和为 1.591×10^{-31}，相对误差为 3.535×10^{-32}；测试样本误差平方和为 1.653，相对误差为 0.114，误差满足精度要求，具有最佳逼近效果（图 1）。隐藏层激活函数：Softmax；输出层激活函数：恒等。隐含层 10 个神经元到输出层神经元各个连接权值分别为 1.166、1.861、−1.096、−0.362、−0.266、0.821、0.249、0.114、−1.111 和 −1.430（表 4）。

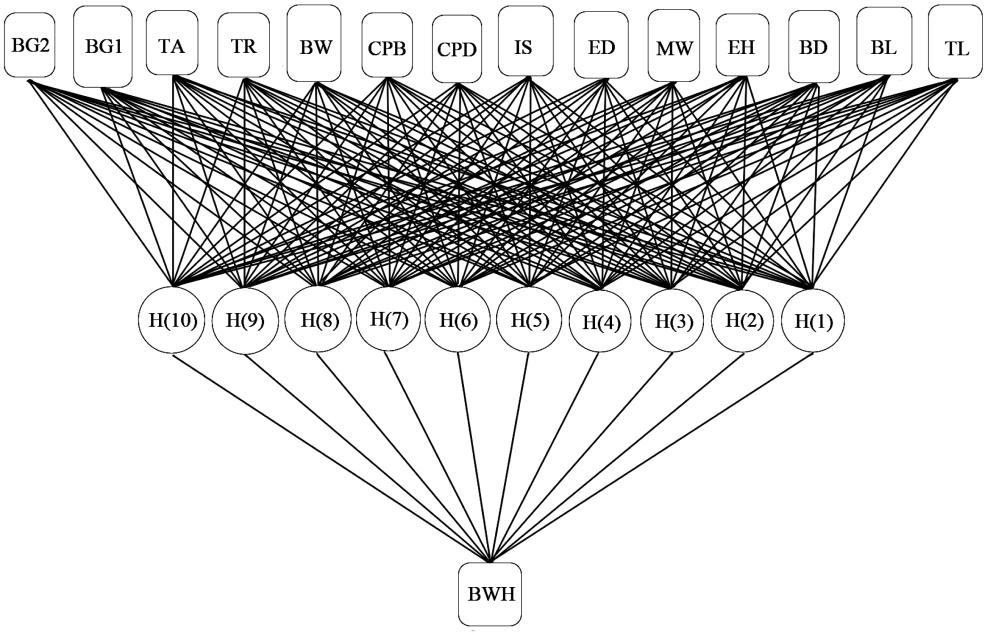

图 1 径向基函数网络图

注：TL，全长；BL，体长；BD，体高；HL，头长；EH，眼后头长；SL，吻长；MW，口宽；ED，眼径；IS，眼间距；CPL，尾柄长；CPD，尾柄高；CPB，尾柄宽；BW，体宽；TR，躯干长；TA，尾长；BG1，体周长 1；BG2，体周长 2；BWH，体质量

Fig. 1 Radical Basis Function network diagram

Notes: TL, total length; BL, body length; BD, body depth; HL, head length; EH, length between eye and head; SL, snout length; MW, mouth width; ED, eye diameter; IS, space between eye and eye; CPL, caudal peduncle length; CPD, caudal peduncle depth; CPB, caudal peduncle breadth; BW, body width; TR, trunk length; TA, trail length; BG1, body girth 1; BG2, body girth 2; BWH, body weight

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 hidden layer	全长/cm TL	体长/cm BL	体高/cm BD	眼后头长/cm EH	口宽/cm MW	眼径/cm ED	眼间距/cm IS	尾柄高/cm CPD	尾柄宽/cm CPB	体宽 BW/cm	躯干长/cm TR	尾长/cm TA	体周长 1/cm BG1	体周长 2/cm BG2	权值 W_{iu}
1	1.644	1.385	1.420	1.720	-0.187	0.621	0.896	0.851	0.051	1.262	1.350	0.912	0.755	1.284	1.166
2	1.178	1.274	1.036	-0.897	1.061	-1.449	1.710	1.541	0.307	0.977	1.612	0.747	1.466	0.950	1.861
3	-0.685	-0.554	-1.458	-0.150	-2.060	-1.449	-1.819	-0.299	0.307	-0.730	0.444	-1.078	-0.449	-0.092	-1.096
4	-1.709	-1.662	0.269	-1.084	0.437	0.621	0.353	0.851	2.097	-0.256	-1.096	-2.157	0.098	1.242	-0.362
5	0.247	0.388	-0.115	-0.150	-0.499	0.621	-0.733	-1.219	-0.460	-0.446	-0.824	1.327	-0.394	-0.258	-0.266
6	-0.033	-0.166	0.844	0.037	0.749	0.621	-0.462	0.391	0.307	0.028	0.444	0.083	0.645	-0.008	0.821
7	-0.079	-0.554	0.652	0.224	0.437	0.621	0.624	0.161	-0.972	1.072	-0.734	-0.083	0.810	0.492	0.249
8	0.759	1.052	-1.075	1.533	1.061	0.621	0.353	0.161	-0.205	0.693	0.263	0.083	-0.175	-0.800	0.114
9	-0.685	-0.831	-0.499	-1.271	-1.124	0.621	-0.733	-0.759	0.307	-0.730	-0.190	0.083	-0.667	-1.384	-1.111
10	-0.638	-0.332	-1.075	0.037	0.125	-1.449	-0.190	-1.679	-1.739	-1.869	-1.277	0.083	-2.090	-1.426	-1.430

模型验证　用预留的 20 组数据测试训练完成后的 RBF 神经网络模型,检验所构建 RBF 神经网络模型的准确性。将样本数据输入到模型中进行预测,并与实测值进行比较,以检验模型的准确性和稳定性。本研究采用线性回归法检验结果发现,决定系数 R^2 为 0.991 6(图 2),明显高于多元线性回归分析的 R^2 值(0.948 9),较多元线性回归模型具有较高的确定系数。根据预测值与实测值较高的确定系数,RBF 神经网络模型能够很好地解释红鳍东方鲀体质量实测值。

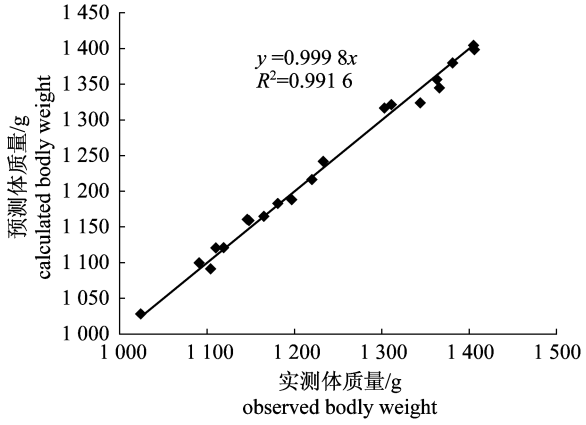

图 2　RBF 神经网络模型预测值与实测值回归分析

Fig. 2　Regression results of RBF neural network

3　讨论

研究表明,水生动物的部分表型性状和体质量之间存在较强的相关性[2-6]。基于表型性状和体质量的这种关系,可通过建立以表型性状为自变量,体质量为因变量的数学模型,实现预估体质量的目的。目前,在水产领域主要采用线性回归的方法来构建这类数学模型[2-6]。在多元回归分析中,不仅因变量与自变量之间存在着线性关系,而且自变量之间通常也存在着高度的线性相关,统计理论将这种现象称为多重共线性(multicollinearity)(简称共线性)。多重共线性是多元回归分析中一类比较复杂的问题,由于共线性的存在,会消弱回归系数估计值的准确性和稳定性,同时测量样本也可能存在着一些离群值,尤其是部分性状与体质量间存在非线性关系,从而给数据分析和预测带来很大的误差[20]。共线性诊断有多种方法,如:检查 R^2 值和 t 检验;自变量间的简单相关;特征根与条件数检验;方差膨胀因子(*VIF*)分析等。近年来,人工神经网络(artificial neural networks, ANN)以其大规模并行处理能力、分布式存储能力、自学习能力等优点而被广泛应用于各领域中的复杂非线性系统问题,其中,反向传播(back propagation, BP)神经网络和径向基函数神经网络更是渗透到各个领域,并取得了较好的成效[21],相比于 BP 神经网络,RBF 神经网络因其能够以任意精度逼近任意连续函数、计算量小、学习速度快、不易陷入局部极小等优点[13-17]为非线性建模提供了一种有效方法,得到了更加广泛的应用[22-25]。

本研究采用容许值和方差膨胀因子以及最大条件指数和最大方差比的方法对利用线性回归方法构建的模型进行共线性检验。基于容许值(*Tol*)和方差膨胀因子(*VIF*)对回归模型进行共线性检验发现,各形态性状都没有超过 $VIF < 10$

和 $Tol > 0.10$,这表明所建模型的共线性没有达到极其严重的程度[19]。由于容许值无明确的诊断界限,一般仅作为共线性诊断的参考指标;而方差膨胀因子则是诊断多元共线性严重程度的指标。所拟合的线性统计模型及相应参数的估计是否精确,还需要在检验容许值和方差膨胀因子的基础上,采用最大条件指数和最大方差比的方法进行共线性检验。通常情况下,若最大条件指数 $k \geqslant 10$ 且相应的最大方差比大于 0.5,可认为自变量间存在共线性[19]。结果显示,最大条件指数为 89.909,且 TL 和 BD 的最大方差比分别为 0.65 和 0.52,均大于 0.50,这表明,模型自变量间存在一定的共线性,通过多元线性回归预测体质量,存在一定误差。如果 $Tol < 0.5$,而 $VIF > 2$ 时,自变量间就可能存在很大的相关性。在此标准下观察发现,体周长 1 和体高自相关较大。显然,通过多元线性回归预测红鳍东方鲀的体质量,会导致一定的误差。本研究在对红鳍东方鲀线性回归模型统计分析的基础上,通过最近邻聚类算法,构建了基于 RBF 神经网络的红鳍东方鲀体质量预测模型,预估其体质量。用预留测试样本对所构建的 RBF 神经网络模型进行验证,发现基于 RBF 神经网络的红鳍东方鲀体质量预测模型的确定系数 R^2 为 0.991 6,而多元线性回归分析的 R^2 值为 0.948 9,这表明基于 RBF 神经网络方法构建的预测模型消除了线性分析中自变量的共线性问题,比线性回归预测模型的确定系数提高 4.53%,预测精度高于线性回归方法。基于 RBF 神经网络体质量预测模型的构建,为利用表型性状精确评估红鳍东方鲀的体质量提供了一种新的方法。

参考文献:

[1] Sheng Z L, Chen Y S. Quantitative genetics [M]. Beiing: Science Press, 2001: 132 – 137. [盛志廉,陈瑶生. 数量遗传学. 北京:科学出版社, 2001: 132 – 137.]

[2] Wang X A, Ma A J, Xu K, *et al*. Relationship between morphometric attributes and body weight of juvenile turbots *Scophthalmus maximus* [J]. Acta Zoologica Sinica, 2008, 54(3): 540 – 545. [王新安,马爱军,许可,等. 大菱鲆幼鱼表型形态性状与体重之间的关系. 动物学报, 2008, 54(3): 540 – 545.]

[3] Wang X A, Ma A J, Zhuang Z M, *et al*. Effects of morphometric attributes on body weight of *Fugu rubripe* (Temminck et Schlegel) [J]. Oceanologia et Limnologia Sinica, 2013, 44(1): 135 – 140. [王新安,马爱军,庄志猛,等. 红鳍东方鲀形态性状对体重的影响效果. 海洋与湖沼, 2013, 44(1): 135 – 140.]

[4] Lin Q, Wang Y J, Wang D W, *et al*. Comparison analysis of quantitative traits and path between two cultured populations of *Crassostrea gigas* and *C. angulata* [J]. Marine Science Bulletin, 2014, 33(1): 106 – 111. [林清,王亚骏,王迪文,等. 太平洋牡蛎和葡萄牙牡蛎养殖群体数量性状比较分析. 海洋通报, 2014, 33(1): 106 – 111.]

[5] Zhang C S, Li F H, Xiang J H. Path analysis of effects of morphometric attributes on body weight of *Exopalaemon carinicauda* [J]. Journal of Fisheries of China, 2013, 37(6): 809 – 815. [张成松,李富花,相建海. 脊尾白虾形态性状对体质量影响的通径分析. 水产学报, 2013, 37(6): 809 – 815.]

[6] Sun Z X, Chang L R, Xu J P. Effects of phenotypic morphometric trait on body weight andsoft-tissue weight of *Neverita Didyma* [J]. Oceanologia et Limnologia Sinica, 2010, 41(4): 513 – 518. [孙振兴,常林瑞,徐建鹏. 扁玉螺表型性状对体重和软体部重的影响效应分析. 海洋与湖沼, 2010, 41(4): 513 – 518.]

[7] Man J L, Yang W. Based on multiple collinearity processing method [J]. Mathematical Theory and Applications, 2010, 30(2): 105 – 109. [满敬銮,杨薇. 基于多重共线性的处理方法. 数学理论与应用, 2010, 30(2): 105 – 109.]

[8] Zhou Z Y, Zhang Y, Lin H K, *et al*. Weight model based on regression of principle component analysis of Xinjiang brown cattle [J]. China Cattle Science, 2012, 38(1): 1 – 4. [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. 中国牛业科学, 2012, 38(1): 1 – 4.]

[9] Cai Z H. Application of the least square method in estimating cattle body weight [J]. China Cattle Science, 2000, 26(5): 1 – 4. [蔡治华. 最小二乘法在黄牛体重估测中的应用. 中国牛业科学, 2000, 26(5): 1 – 4.]

[10] Zhang Y Y, He X. Multilevel nonlinear mixed effect model for evaluation of Muscovy ducks body weight for growth [J]. Scientia Agricultura Sinica, 2012, 45(3): 540 – 547. [张元跃,贺喜. 多层非线性混合效应模型评价番鸭体重生长. 中国农业科学, 2012, 45(3): 540 – 547.]

[11] Fu W S, Teng G H, Yang Y. Research on three-

dimensional model of pig's weight estimating[J]. Transactions of the CSAE,2006,22(suppl. 2):84 – 87.[付为森,滕光辉,杨艳. 种猪体重三维预估模型的研究. 农业工程学报,2006,22(增刊 2):84 – 87.]

[12] Liu T H,Li Y Z,Teng G H,*et al*. Prediction of pig weight based on radical basis function neural network [J]. Transactions of the Chinese Society for Agricultural Machinery,2013,44(8):245 – 249.[刘同海,李摇卓,滕光辉,等. 基于 RBF 神经网络的种猪体重预测. 农业机械学报,2013,44(8):245 – 249.]

[13] Du K L, Swamy M N S. Neural networks and statistical learning – radial basis function networks [M]. London: Springer-Verlag Press, 2014: 299 – 335.

[14] Muhmann M D. Radial basis functions: Theory and implementations [M]. Cambridge: Cambridge University Press,2003:2 – 5.

[15] Chen S,Fillings S A,Grant P M. Recursive hybrid algorithm for non-linear system identification using radial basis function networks[J].. International Journal of Control,1992,55(5):1051 – 1070.

[16] Muhmann M D. Radial basis functions[J]. Acta Numerica,2000,9:1 – 38.

[17] Jamshidi A A, Gear C W, Kevrekidis I G. Time varying radial basis functions [J]. Journal of Computational and Applied Mathematics, 2014, 266 (15):61 – 72.

[18] Park J,Sandberg J W. Universal approximation using radial basis functions network [J]. Neural Computation,1991,3(2):246 – 257.

[19] Ding Y L, Kong D L, Mao Z F. The common diagnosis of collinearity of multiple linear regression [J]. Journal of Mathematical Medicine, 2004, 17 (4):299 – 300.[丁元林,孔丹莉,毛宗福. 多重线性回归分析中的常用共线性诊断方法. 数理医药学杂志,2004,17(4):299 – 300.]

[20] Yang M,Xiao J,Cai H. The multicollinearity and its processing technique of multivariate analysis[J]. Chinese Journal of Health Statistics,2012,29(4): 620 – 624.[杨梅,肖静,蔡辉. 多元分析中的多重共线性及其处理方法. 中国卫生统计,2012,29 (4):620 – 624.]

[21] Huang H. Simulation test of principal component analysis – RBF neural network model based on SPSS [J]. Journal of Xinjiang Agricultural University, 2010,33(6):535 – 538.[黄华. 基于 SPSS 的 PCA-RBF 神经网络模型的仿真试验. 新疆农业大学学报,2010,33(6):535 – 538.]

[22] David W. Neural network credit scoring models[J]. Computers & Operations Research, 2000, 27 (11 – 12):1131 – 1152.

[23] Meng J E, Shiqian W, Juwei L, *et al*. Face recognition with radial basis function (RBF) neural networks [J]. Ieee Transact Ions on Neural Networks,2002,13(3):697 – 710.

[24] Tianshu B, Zheng Y, Fushuan W. On-line fault section estimation in power systems with radial basis function neural network[J]. Electrical Power and Energy System,2002,24(4):321 – 328.

[25] Sridhar S,Hassan K K. Output feedback control of nonlinear systems using RBF neural networks[J]. Ieee Transactions on Neural Networks,2000,11(1): 69 – 79.

Prediction of *Takifugu rubripes* weight based on radial basis function neural network

WANG Xin'an[1,2], MA Aijun[1,2*], ZHAO Yanfei[1], YUE Liang[1], SUN Jianhua[1,2], MENG Xuesong[3], LIU Shengcong[3]

(1. *Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences; Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture; Qingdao Key Laboratory for Marine Fish Breeding and Biotechnology, Qingdao* 266071, *China*; 2. *Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao* 266071, *China*; 3. *Dalian Tianzheng Industrial Co. Ltd., Dalian* 116000, *China*)

Abstract: *Takifugu rubripes* belongs to the bony fishes, Tetraodontiformers, Tetraodontoidei, Tetraodontidae, *Takifugu*. It is distributed mainly in Japan of the western north Pacific, the Korean peninsula and China's coastal areas. Due to its appealing taste, rich nutrition, low fat content and numerous trace elements, *Takifugu rubripes* represents one of the fish species with high economic value. In recent years, *Takifugu rubripes* are farmed in large numbers in Dalian, Qinhuangdao, Tangshan, and Tianjin regions, and has become the main cultured species of puffer fishery in china. There existed large errors due to self-correlation between different phenotypic traits, non-linear relationship between some traits and body weight and the collinearity among independent variables, when the linear regression model was used to predict *Takifugu rubripes* weight. As a solution, a *Takifugu rubripes* weight prediction RBF neural network model, according to Artificial Neural Networks theory and Radial Basis Function model, was constructed with the phenotypic traits (including total length, body length, body depth, head length, length between eye and head, snout length, mouth width, eye diameter, space between eye and eye, caudal peduncle length, caudal peduncle depth, caudal peduncle breadth, body width, trunk length, trail length, body girth 1, body girth 2 and body weight) of 72 *Takifugu rubripes* based on the nearest neighbor clustering algorithm, and the credibility of the neural network model constructed was tested by linear regression techniques. The results showed that the coefficient of determination R^2 of RBF neural network prediction model and the linear regression model for *Takifugu rubripes* weight were 0.992 (approximately 1) and 0.949, respectively. Obviously, the coefficient of determination R^2 of RBF neural network prediction model was improved by 4.53% compared with the linear regression model. In addition, the collinearity diagnostics of linear regression, based on tolerance and variance inflation factor as well as maximum condition index and maximum variance proportions, indicated that there existed certain collinearity among independent variables and self-correlation between body girth 1, and body depth. The results suggested that the RBF neural network technique was an effective method to construct the prediction model of *Takifugu rubripes*, and the collinearity of the independent variables, in RBF neural network analysis, was eliminated and it has higher accuracy than linear regression prediction model. Weight prediction model based on radial basis function neural network provides a new method for accurate prediction of *Takifugu rubripes* weight.

Key words: *Takifugu rubripes*; weight prediction; radial basis function; neural network

Corresponding author: MA Aijun. E-mail: maaj@ysfri.ac.cn

Funding projects: Earmarked Fund for Modern Agro-Industry Technology Research System (CARS-50-G01); Special Fund for Basic Scientific Research Business of Central Public Research Institutes (20603022012005); Research Program of Jinzhou Dalian, China (2012-B1-012)