

LM 神经网络在鱼类胚胎保存抗冻剂毒性试验设计中的应用

邵磊^{1,2}, 肖志忠³

(1. 中国海洋大学 信息科学与工程学院, 山东 青岛 266003; 2. 青岛大学 师范学院, 山东 青岛 266071; 3. 中国科学院 海洋研究所, 山东 青岛 266071)

摘要:采用 LM 算法神经网络对鱼类胚胎冷冻保存抗冻剂毒性试验数据进行了分析, 研究了孵化率和抗冻剂种类、浓度以及平衡时间等因素之间的对应关系, 为鱼类胚胎冷冻保存的研究提供了新的分析手段, 研究结果表明这一方法行之有效。

关键词:神经网络; LM 算法; 试验设计; 抗冻剂毒性; 低温保存

中图分类号: TP183; Q16; X176 **文献标识码:** A **文章编号:** 1000-3096(2007)07-0001-03

鱼类种质资源是优良品种培育、养殖生产及水产养殖业可持续发展的重要物质基础。通过低温保存技术将鱼类胚胎长期有效保存起来建立种质库, 有利于遗传基因多样性及生物种质资源的有效保护。目前, 对鱼类胚胎低温保存研究主要集中在抗冻剂种类、毒性浓度、胚胎内部结构的渗透能力和低温抵抗能力、降温方式及复温方式等方面。其中, 选择合理的抗冻剂种类、浓度和平衡时间, 是鱼类胚胎冷冻保存成功的关键因素之一。一般情况下, 几种不同抗冻剂以不同浓度组成的混合溶液能够降低单一成分抗冻剂的毒性^[1]。多种抗冻剂以不同浓度混合然后平衡不同时间, 可能的试验方案数目是巨大的, 采用穷举法或传统的正交试验方法进行生物试验是不可行的, 因而有必要研究新的方法对这一问题进行解决。不同抗冻剂的组合和平衡时间, 与冷冻保存结果之间存在某种潜在的规律性。如果能够从部分已有的试验数据中发现这种规律性, 那么可以通过仿真模拟的方法对所有潜在方案的效果进行预测, 然后将预测效果好的方案进行生物试验分析, 这无疑可以大大提高试验的效率。作者尝试采用神经网络方法对目前已有的鱼类胚胎保存抗冻剂毒性试验数据进行分析, 以孵化率作为抗冻剂毒性结果的评价指标, 以抗冻剂种类、浓度和平衡时间 3 个指标作为主要影响因素, 研究它们之间的对应关系, 并对研究的仿真分析结果进行展示。

1 神经网络

作者采用的神经网络是一种单向传播的多层前馈网络。作为一种高度参数化的非线性模型, 神经网络

可以对变量之间的影响关系进行良好的近似, 目前已经在多个领域的数据分析中获得了成功的应用。神经网络的结构通常包含三部分: 输入层, 中间层(隐藏层), 输出层, 结构如图 1 所示^[2,3]。

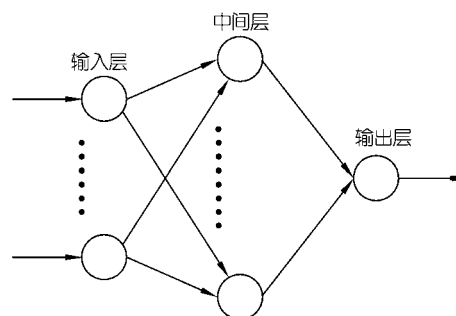


图 1 神经网络模型

Fig.1 Model of neural networks

输入层单元个数等于影响因素的个数, 在本研究中为抗冻剂、浓度、平衡时间, 输出层单元只有孵化

收稿日期: 2007-01-26; 修回日期: 2007-05-14

基金项目: 国家 863 计划资助项目(2003AA603510); 国家自然科学基金资助项目(30571427)

作者简介: 邵磊(1972-), 女, 山东青岛人, 讲师, 硕士研究生, 主要研究方向为数据挖掘技术在生物信息处理中的应用。
E-mail: lilishao@qingdaonews.com



率一个指标。对于输入层的每个单元,其输入值等于输出值,输入值由样本数据提供。对于中间层和输出层的单元,其输入等于前一层输出的线性加权和,公式如下:

$$I_j = \sum_i w_{ij} O_i + \theta_j$$

式中, I_j 为第 j 个神经元的输入值, O_j 为前一层中第 i 个单元的输出值, w_{ij} 为单元 i 到单元 j 的连接权重向量, θ_j 为第 j 个神经元的阈值向量,用于调节神经元的活性。中间层和输出层神经元的输出值与输入值之间存在如下的函数关系:

$$O_j = (1 + e^{-I_j})^{-1}$$

这一函数被称为 Logistic 激活函数,也可以根据研究问题的需要选择其它的激活函数形式,例如 Eolligt 函数等。研究的目标就是如何确定合适的连接权重向量 W 和阈值向量 θ ,从而也就确定了影响因素和目标变量之间的量化关系。经典的算法有 Back-propagation (BP) 算法、高斯-牛顿法等,文中采用 Levenberg-Marquardt (LM) 算法计算神经网络参数,并与 BP 算法的结果进行比较,结果表明这一方法的效果要明显优于 BP 算法。LM 算法是一种基于数值优化的学习算法,它是梯度法和高斯-牛顿法的结合,兼有两种算法的优点。其基本原理描述如下^[4,5],设 $x^{(k)}$ 表示第 k 次迭代的权值和阈值组成的向量,新的权值和阈值组成的向量 $x^{(k+1)}$ 可以根据下式求得:

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)}$$

选用平方误差函数作为目标函数,

$E(x) = -\frac{1}{2} \sum_{i=1}^N e_i^2(x)$, $e_i(x)$ 为第 i 个样本的预测值和实际观测值的偏差,此时,

$$\Delta x = -[J^T(x)J(x) + \mu I]^{-1} J(x)e(x)$$

式中, μ 为大于 0 的常数, I 为单位矩阵, $J(x)$ 为 Jacobian 矩阵,即

$$J(x) = \begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \dots & \frac{\partial e_1(x)}{\partial x_N} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \dots & \frac{\partial e_2(x)}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N(x)}{\partial x_1} & \frac{\partial e_N(x)}{\partial x_2} & \dots & \frac{\partial e_N(x)}{\partial x_N} \end{bmatrix}$$

任意给定算法的初始点 $x^{(0)}$,通过迭代更新 $x^{(k)}$ 的值直到满足算法的终止条件,从而获得最佳的神经

网络参数。

2 仿真分析

作者选取 110 条已有的鱼类胚胎冷冻保存试验数据验证神经网络方法的可行性。数据中包含 4 个变量列,其中抗冻剂、浓度、平衡时间为影响因素变量,孵化率为分析的目标变量。将数据随机分成训练样本和验证样本两部分,分别包含 99 条和 11 条数据。通过对训练样本的分析,获得神经网络模型所描述的变量关系,然后在验证样本上对这一关系进行验证,以评价方法的效果。经过多次分析比较,最终选定神经网络的中间层包含 4 个单元,激活函数形式为 Logistic 函数。对于同一模型,分别采用 LM 算法和 BP 算法进行训练,并将结果进行对比,如表 1 所示。

表 1 LM 与 BP 两种算法的结果对比

Tab.1 Comparison of results between LM and BP

算法	样本	分析指标		
		均为误差	AIC	SBC
LM	训练样本	0.021 5	-250.361 7	-81.678 9
	验证样本	0.049 9		
BP	训练样本	0.132 8	-69.877 7	98.805 1
	验证样本	0.223 6		

通过表 1 的模型训练结果可以看出,无论是均方误差,还是 AIC (Akaike's Information Criterion, 赤池信息准则) 和 SBC (Schwarz's Bayesian Criterion, 施瓦茨贝叶斯准则) 等指标 LM 算法的结果都明显地优于 BP 算法的结果,这里最重要的指标是验证样本上的均方误差,验证样本的误差较低说明了模型对于新数据有良好的预测能力,即模型的推广能力强。表 2 中列举了采用 LM 算法获得的模型在验证样本上孵化率的预测值和实际观测值的对比。

图 2 展示了 LM 算法的训练过程,通过对训练过程的跟踪,可以看出 LM 算法只需要很少的迭代步骤就可以得到一个较低的训练误差值,这说明在算法效率上也要远远地优于其它的算法。而且通过图 2 可以清楚地看到验证样本上的误差是先减小后增大的,也就是神经网络训练中出现了过拟合问题,文中选择验证样本上误差最低时的权值作为最终的训练结果,以保证模型获得最佳的预测能力。



表 2 验证样本孵化率预测值和实测值的对比

样本	预测值 (%)	实测值 (%)
1	32.11	52.00
2	58.27	3.27
3	99.23	94.00
4	91.30	92.19
5	7.50	0
6	4.53	0
7	89.50	99.33
8	73.50	1
9	67.00	97.51
10	96.34	96.00
11	75.59	60.50

3 结论

经研究,作者发现可以通过数据分析获得鱼类胚胎冷冻保存中抗冻剂毒性试验的有效设计方案。使得只需要进行少量的试验收集数据,进而通过对数据的分析获得仿真模型,然后通过模型的模拟预测,计算出孵化率最高的试验方案,并将这一预测结果用于指导新的生物试验,从而有效地避免了试验的盲目性,大大地提高了试验的效率。更重要的是,随着数据量的不断积累,不断地重复运用这一方法,可以有效地提高模型的预测精度,实现分析过程的动态改善。这一方法可以推广到冷冻保存试验设计的其它方面,具有广泛的借鉴和指导意义。

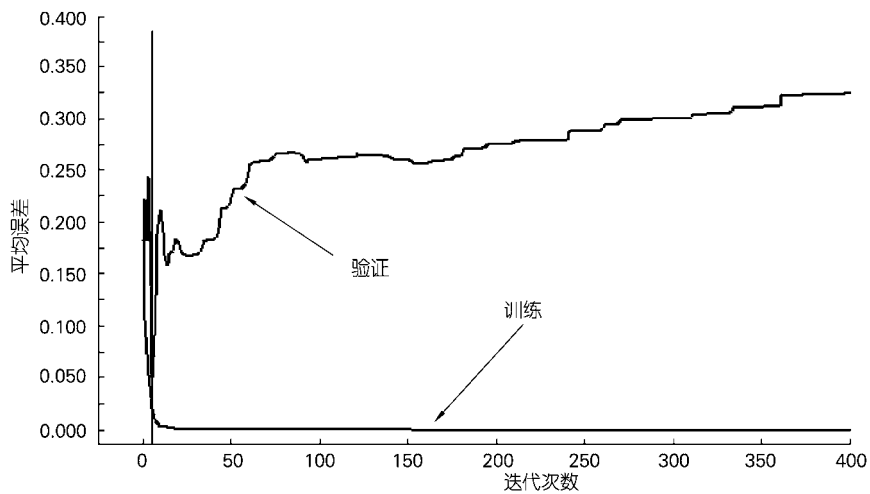


图 2 模型训练过程

Fig.2 Training process of the model

参考文献:

[1] 陈松林. 鱼类配子和胚胎冷冻保存研究进展及前景展望[J]. 水产学报, 2002, 26(2): 161-168.

[2] White H. Learning in artificial neural networks: a statistical perspective [J]. *Neural Computation*, 1989,1 : 425-464.

[3] Han Jiawei,Kamber M. Data Mining Concepts and Techniques [M].北京: 高等教育出版社, 2001.

[4] Marquardt D. An algorithm for least squares estimation of nonlinear parameters [J]. *Journal of the Society for Industrial and Applied Mathematics*, 1963, 11:431-441.

(下转第 24 页)

(上接第 3 页)

[5] 葛玲, 贾志成. 基于 LM 算法的神经网络语音识别[J]. 计算机工程与设计, 2006, 7 : 534-537.

[6] 徐冠, 夏克文, 徐乃勋. 基于 LM 算法的神经网络在冠心病诊断中的应用[J]. 微电子学与计算机, 2006, 23 (2): 189-192.

Study on the experimental design of cryoprotectant toxicity to fish embryo cryopreservation based on LM neural networks

SHAO Lei^{1,2}, XIAO Zhi - zhong³

(1. College of Information Science & Engineering , Ocean University of China, Qingdao 266003, China; 2. Teachers' College of Qingdao University, Qingdao 266071, China; 3. Institute of Oceanology, the Chinese Academy of Sciences, Qingdao 266071, China)

Received: Jan., 26, 2007

Key words: neural networks ; LM algorithm ; experimental design ; cryoprotectant toxicity ; cryopreservation

Abstract: In this paper, we analyze the cryoprotectant toxicity to the fish embryo cryopreservation based on LM neural networks, and study the influences of the types and thickness of the cryoprotectants and the balancing time on the hatching rate. Results show that this method is effective and the study gives a new analyzing method to study fish embryo cryopreservation.

(本文编辑 : 刘珊珊)