

居锦武. 基于 LS-SVM 的养殖水体氨氮含量分析预测模型[J]. 江苏农业科学, 2017, 45(2): 210-212.

doi:10.15889/j.issn.1002-1302.2017.02.061

基于 LS-SVM 的养殖水体氨氮含量分析预测模型

居锦武

(四川理工学院计算机学院, 四川自贡 643000)

摘要: 养殖水体的氨氮含量是水产养殖中的关键水质指标, 是评价水体污染情况的基本测量项目。设计一种自动氨氮监测系统, 自动抽取水样并定时测量, 使用最小二乘支持向量机算法 (LS-SVM) 建立了分析预测模型, LS-SVM 最小二乘支持向量机算法增强了泛化能力, 减少了异常样本的预测偏差, 提高了预测学习的精度。对基于神经网络模型、基于 LS-SVM 模型进行了仿真比较, 结果显示 LS-SVM 模型具有更优良的分析预测效果。

关键词: 支持向量机; 水产养殖; 氨氮; 分析预测模型; 比色法; 光强

中图分类号: TP391 **文献标志码:** A **文章编号:** 1002-1302(2017)02-0210-03

随着水产养殖业的快速发展, 养殖水体的污染情况越来越严重。各种饵料、化学物质、药物、排泄物残留在水中, 经腐败和分解, 引起水体水质的严重劣化, 使水体中氨氮等有害物质含量上升, 造成养殖对象发病或死亡, 严重影响产业的发展。我国《渔业水质标准 (GB 11607—1989)》中规定, 氨氮质量浓度小于 0.02 mg/L 时, 对鱼类的生命活动没有影响。养殖水体中铵离子 (NH_4^+) 的质量浓度不允许超过 5 mg/L。

笔者设计的自动氨氮含量监测系统^[1] 基于 HJ535—2009 标准。其基本原理是将被测水样与纳氏试剂混合, 纳氏试剂与水样中的氨反应, 采用紫外光度计对生成的化合物进行光度测量, 吸光度与水样中的氨氮含量成正比, 从而计算出水样的氨氮含量。此方法的优点为所需试剂种类少、生成化合物速度快、吸光度分辨率高, 是一种简单高效的测量方法, 特别适合设计成自动氨氮含量监测系统。

收稿日期: 2015-11-18

基金项目: 四川省院士(专家)工作站基金(编号: 2014YSGZZ01); 企业信息化与物联网测控技术四川省高校重点实验室基金(编号: 2014WYY01)。

作者简介: 居锦武(1976—), 男, 上海人, 硕士, 副教授, 主要从事嵌入式计算机系统设计研究。E-mail: jjwmail@163.com。

纳氏试剂比色法的缺点为试剂成本较高、测量数据量少、测量过程易受环境变化等不可预见因素的影响^[2], 重复性误差 $< \pm 10\%$, 数据可靠性还需提高。建立数据分析预测模型是提高数据可靠性的常用方法, 常用的分析预测模型包括神经网络^[3-4]、支持向量机^[5]等, 神经网络一般要求使用大量数据作为训练样本, 采用比色法测量氨氮含量所得数据样本较少, 符合支持向量机的特点。基本支持向量机 (SVM) 的泛化

3 结论

通过 Matlab 的快速傅立叶变化分析, 得出海棠果树的前四阶共振频率在 3.0、8~9、16~17、23~25 Hz 处。

能量在果树主干部分传递时能量衰减率为 $-1.5\% \sim 3\%$ 。在传递过程中, 越远离树根部分, 能量损失就越大。在果树第 1 层会有部分能量分流到第 2 层, 且主枝直径越大, 能量分流也就越大, 主枝与主干中心轴夹角角度越小, 分流越大。

比较树干 x 、 y 、 z 等 3 个方向能量传感数据发现, 在低位置小冲击时, z 向峰值明显大于 x 、 y 方向的峰值。当处于高位置大冲击时, z 向峰值与 x 、 y 方向峰值的差距在减小。当冲击位置距离地面为 40~60 cm, 冲击力为 900 N 时, 树干发生了共振, 此时的落果率最高。

参考文献:

- [1] 罗万杰. 新疆特色林果业发展现状、存在问题及对策[J]. 新疆农垦科技, 2012(8): 16-18.
- [2] 张敏敏. 果树粘弹性振动结构模型的研究[D]. 乌鲁木齐: 新疆大学, 2015: 9-30.
- [3] Paolo A, Catalano P, Ferruccio G, et al. Vibration analysis of an olives

- mechanical harvesting system[J]. Agricultural Engineering International, 2007(9): 1-9.
- [4] He L, Zhou J F, Du X Q, et al. Energy efficacy analysis of a mechanical shaker in sweet cherry harvesting[J]. Biosystems Engineering, 2013, 116(4): 309-315.
- [5] Awady M E, Genaidy M I, Rashowan M, et al. Modeling and simulating of olive-tree harvesting mechanism[J]. Farm Machinery and Power, 2008, 25(3): 712-722.
- [6] 贺磊盈. 面向振动采收的果树枝干三维重建方法及动力学特性研究[D]. 杭州: 浙江理工大学, 2014: 30-65.
- [7] 翁发云. 林果采收振动系统动态特性传递研究[D]. 南京: 南京林业大学, 2013: 12-19.
- [8] 钟声. 动态冲击实验中信号测试与结果分析[D]. 武汉: 武汉理工大学, 2013: 16-24.
- [9] Polat R, Gezer I, Guner M, et al. Mechanical harvesting of pistachio nuts[J]. Journal of Food Engineering, 2007, 79(4): 1131-1135.
- [10] Ferguson L, Rosa U A, Castro-Garcia S, et al. Mechanical harvesting of California table and oil olives[J]. Adv Hort Sci, 2010, 24(1): 53-63.
- [11] de Silva C W. 振动阻尼、控制和设计[M]. 北京: 机械工业出版社, 2013: 56-91.

能力不足,对于异常样本数据,预测数据偏差较大。最小二乘支持向量机(LS-SVM)^[6]对其进行了改进,优化了模型函数,具有更好的预测速度和效果。

笔者在设计自动氨氮监测系统时,针对测量数据的准确性波动问题,将LS-SVM算法应用到测量结果的分析中,建立了养殖水体氨氮含量测量结果的分析预测模型,获得了较好的效果。

1 支持向量机

Vapnik 等于 20 世纪 90 年代提出支持向量机(SVM)方法,该方法是一种机器学习方法,通过遵循结构风险最小原则,提高学习机的泛化能力,即使在统计样本数量较少的情况下也能得到良好的统计结果。LS-SVM 由 Suykens 等提出,改变了 SVM 使用的约束条件,同时使用新的目标优化函数,改变风险控制函数,进一步改善了泛化能力^[7-8]。

设有训练集 D , 内部包含 N 个样本, 将 D 表示为: $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 。其中 x, y 分别为输入、输出。通过 1 个非线性映射 $\phi(\cdot)$, 将样本数据从原空间映射到另一个高维特征空间 $\phi(x)$, 新空间表示为 $\phi(x) = [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]$ 。在空间 $\phi(x)$ 中设置最优决策函数 $y(x) = \omega\phi(x) + b + \xi_k$, 函数中的 ω, b 分别为权值向量、阈值。LS-SVM 的目标优化函数^[9]为:

$$\min J(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \nu i \xi_k^2; \quad (1)$$

$$\text{s. t. } y_k = \phi(x_k) \omega^T + b + \xi_k.$$

式中: C 是用于平衡误差与复杂度的正则化参数, 误差分量 $\xi_k \geq 0, k = 1, 2, \dots, N$ 。

对式(1)函数采用 Lagrange 法求解:

$$L(\omega, b, \xi, \alpha) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_k^2 - \sum_{i=1}^N \alpha_k [(\omega^T \phi(x_k) + b + \xi_k) - y_k]. \quad (2)$$

式中: $\alpha_k (k = 1, 2, \dots, N)$ 为 Lagrange 乘子, 分别对 ω, b, ξ, α 求偏导 = 0, 即

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & k(x_1, x_1) + 1/C & \dots & k(x_1, x_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & k(x_N, x_1) & \dots & k(x_N, x_1) 1/C \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ Y_1 \\ \vdots \\ Y_N \end{bmatrix}. \quad (3)$$

设 Ω 为 $N \times N$ 对称方阵, 即

$$\Omega = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix}. \quad (4)$$

定义核函数 $k(x_i, x_j)$, 满足以下 Mercer 条件:

$$\Omega_{ij} = k(x_i, x_j) = \phi(x_i) \phi(x_j) i, j = 1, \dots, N. \quad (5)$$

LS-SVM 的回归模型为:

$$f(x) = \sum_{k=1}^N \alpha_k k(x, x_k) + b. \quad (6)$$

选择径向基核函数(RBF)作为核函数, 模型具有学习速度快的特点, 公式为:

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right). \quad (7)$$

整个算法中有 2 个可调参数, 即正则化参数 C, σ 核宽度参数。在确定 2 个参数的值时, 通常先采用搜索法选择参数, 再采用交叉验证法进行最小误差寻优, 这样能够确定最佳的参数对。这 2 个参数将影响算法的泛化性能和学习能力。

2 养殖水体自动氨氮监测系统

养殖水体自动氨氮监测系统基于纳氏试剂比色法测量水体的氨氮含量, 其系统设计结构见图 1, 监测系统的实物照片见图 2。整个监测系统由 IPC、PLC、LCD、比色计、反应瓶、液量器、注射泵、多通阀及多种试剂等组成。IPC 通过 LCD 显示人机界面, 通过 PLC 控制液量器和多通阀的工作, 同时通过比色计读取反应瓶的色度数据, 再计算氨氮质量浓度。其基本操作流程为: (1) 控制电磁阀, 接入水样, 驱动注射泵, 水样进入液位计量器; (2) 液位达到规定值后切换电磁阀, 水样送入反应器; (3) 切换电磁阀, 掩蔽剂进入计量器, 达到规定液位后再进入反应器; (4) 切换电磁阀, 纳氏试剂抽入计量器, 达到规定液位后再进入反应器; (5) 等待 10 min, 测量比色计的输出数据, 即为测量光强; (6) 切换电磁阀, 反应后的溶液排到废液瓶中; (7) 切换电磁阀, 反应器抽入蒸馏水, 测量比色计的输, 即为参比光强; (8) 将废水排到废液瓶中。

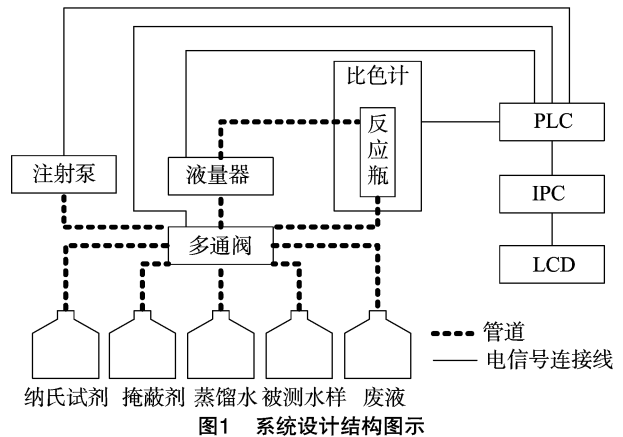


图1 系统设计结构图

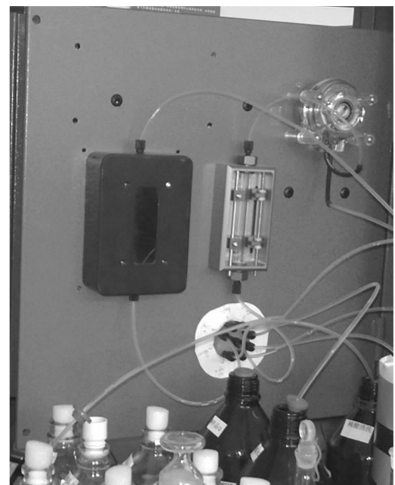


图2 监测系统实物

3 测量结果与分析

基于比色法设计的自动氨氮监测系统, 氨氮含量计算的

数据来源包括参比光强、测量光强。选取 40 组标准氨氮溶液,其含量从 0 mg/L 递增至 10 mg/L,间隔 0.25 mg/L。系统测量的总样本数据为 40 个,奇序列为训练样本,偶序列为测试样本,部分测量数据见表 1。使用神经网络模型和 LS-SVM 模型分别建立分析预测模型,仿真试验数据见表 2。

使用神经网络模型和 LS-SVM 模型分别对水体氨氮含

表 1 部分样本数据

样品序号	神经网络 预测值 (mg/L)	神经网络 误差 (%)	LS-SVM 预测值 (mg/L)	LS-SVM 误差 (%)	氨氮标 准含量 (mg/L)
2	0.55	10.42	0.51	1.64	0.50
4	0.97	-2.99	0.98	-2.12	1.00
6	1.31	-12.49	1.47	-2.30	1.50
8	2.10	5.15	2.00	0.24	2.00
10	2.41	-3.64	2.45	-2.00	2.50
12	3.10	3.32	3.01	0.45	3.00
14	3.86	10.34	3.48	-0.70	3.50
16	4.46	11.50	4.08	2.08	4.00
18	4.35	-3.27	4.51	0.21	4.50
20	4.94	-1.10	4.92	-1.68	5.00
22	5.89	7.04	5.55	0.90	5.50
24	5.77	-3.77	5.95	-0.91	6.00
26	5.67	-12.69	6.35	-2.37	6.50
28	7.63	8.95	6.92	-1.11	7.00
30	8.27	10.24	7.65	1.93	7.50
32	8.11	1.32	7.96	-0.54	8.00
34	9.45	11.20	8.56	0.69	8.50
36	9.48	5.35	8.74	-2.91	9.00
38	9.51	0.12	9.58	0.80	9.50
40	10.36	3.63	10.08	0.80	10.00

表 2 仿真试验结果

样品序号	氨氮标准含量 (mg/L)	参比光强 (lx)	测量光强 (lx)
1	0.25	1 493.26	2 091.15
3	0.75	1 492.65	1 849.05
5	1.25	1 493.88	1 637.00
7	1.75	1 483.90	1 438.39
9	2.25	1 477.09	1 266.54
11	2.75	1 468.72	1 114.02
13	3.25	1 490.52	1 000.07
15	3.75	1 484.62	881.15
17	4.25	1 466.60	769.99
19	4.75	1 486.51	690.37
21	5.25	1 473.80	605.47
23	5.75	1 481.49	538.39
25	6.25	1 468.36	472.03
27	6.75	1 489.58	423.59
29	7.25	1 493.79	375.76
31	7.75	1 478.69	329.03
33	8.25	1 469.56	289.26
35	8.75	1 480.75	257.82
37	9.25	1 486.84	229.00
39	9.75	1 474.90	200.95

量进行预测,氨氮实际含量与 2 组预测结果的对比如图 3。

由表 2、图 3 可知,神经网络模型预测结果的相对误差极值为 12.69%,平均值为 6.43%;LS-SVM 模型预测结果的相对误差极值为 2.91%,平均值为 1.32%。可见,使用 LS-SVM 预测模型获得了更高的精度。

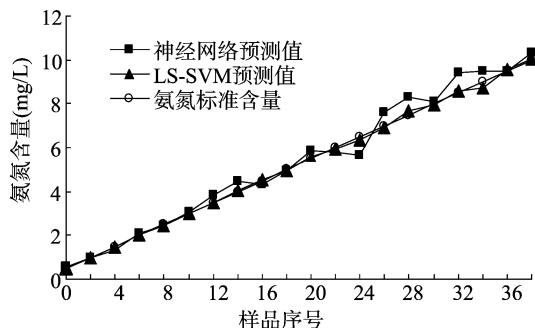


图 3 预测结果与实际含量比较

4 小结

设计了一种养殖水体自动氨氮监测系统,并将 LS-SVM 技术应用于测量结果的分析预测中。数据预测结果与实际含量的比较表明,该模型是有效的预测模型,用于实际的数据分析预测工作可明显提高数据的准确率。LS-SVM 预测模型的可变参数较少,只需确定正则化参数 C 和核宽度参数 σ ,可使用交叉验证法确定 2 个参数的值。LS-SVM 预测模型的精度高,可广泛应用于数据分析预测模型的建立。

参考文献:

- [1] 马淑勃,梁滨. 利用自动分析仪测定水中氨氮的方法研究[J]. 山东化工,2015(44):62-64.
- [2] 俞继梅. 纳氏试剂测氨氮影响因素的分析和研究[J]. 江西化工,2013(4):11-14.
- [3] 董敏明,张愉,齐美星. 基于 BP 神经网络的可燃混合气体分析方法的研究[J]. 计量学报,2006,27(2):169-171.
- [4] 侯迪波,陈玥,赵海峰,等. 基于 RBF 神经网络和小波分析的水质异常检测方法[J]. 传感器与微系统,2013,32(2):138-141.
- [5] 黄首,程小平. 基于支持向量机回归粒子滤波器的运动员实时跟踪方法[J]. 西南大学学报(自然科学版),2008,30(5):165-169.
- [6] 曾燕,成新文,王红旗,等. 基于加权最小二乘支持向量机的下水道可燃气体分析[J]. 西南师范大学学报(自然科学版),2014,39(9):94-99.
- [7] Suykens J, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters,1999,9(3):293-300.
- [8] Wen W, Hao Z, Yang X. A heuristic Weight-Setting strategy and iteratively updating algorithm for weighted LeastSquares support vector regression[J]. Neurocomputing,2008,71(16/18):3096-3103.
- [9] Suykens J, Vandewalle J. Recurrent least squares support vector machines[J]. IEEE Transactions on Circuits and System,2000,47(7):1109-1114.