

DOI: 10.3969/j.issn.1004-6933.2014.03.006

河流健康评价的回归支持向量机模型及应用

刘艳

(云南省水文水资源局文山分局, 云南 文山 663000)

摘要: 建立河流健康评价指标体系、分级标准及回归支持向量机(SVR)河流健康评价模型,并以云南省文山州清水河健康评价为例进行研究。首先,利用层次分析法(AHP)从水文水资源、物理结构、水质、水生生物和社会服务功能5个方面遴选出13个评价指标,构建3个层次的河流健康评价指标体系和5个等级的分级标准;其次,基于SVR原理,利用随机生成和随机选取的方法,在等级标准阈值间构造5种不同容量大小的训练样本和检验样本,提出5种不同容量方案的SVR河流健康评价模型,设计合理的输出模式,并构建具有良好性能的RBF(radial basis function neural network)回归模型作为对比模型,利用模型随机5次运行的平均相对误差绝对值、最大相对误差绝对值和运行时间对各方案模型性能进行评价;最后,利用达到期望精度的SVR模型对实例进行评价分析。结果表明:①无论是训练样本还是检验样本,5种方案的SVR模型的预测精度和泛化能力均优于RBF模型。在相同参数设置条件下,SVR模型随着样本容量的增加其精度和泛化能力变化不大;而RBF模型随着样本容量的增加其精度和泛化能力均有提高。表明SVR模型具有较高的精度和泛化能力,可以用于河流健康评价,尤其在小样本情况下,SVR模型的精度和泛化能力是RBF模型不可比拟的。②5种方案的SVR模型对清水河2011—2012年3次调查的评价结果均为健康,但已接近于亚健康。

关键词: 河流健康; 指标体系; 分级标准; 回归支持向量机; 综合评价; 云南省

中图分类号: X824 **文献标志码:** A **文章编号:** 1004-6933(2014)03-0025-06

Regression support vector machine for river health assessment and its application

LIU Yan

(Wenshan Branch of Yunnan Provincial Hydrology and Water Resources Bureau, Wenshan 663000, China)

Abstract: A river health assessment index system, grading standards, and a support vector regression (SVR) river health assessment model are proposed for health assessment of the Qingshui River in Wenshan, in Yunnan Province. In this study, first, 13 evaluation indices were selected with the analytic hierarchy process (AHP) in terms of hydrology and water resources, physical structure, water quality, aquatic organisms, and social services, in order to construct a three-level river health assessment index system as well as five-level grading standards. Then, based on the SVR principle, the random generation and random selection methods were used to construct five training and testing samples with different capacities in grading thresholds. Five models with different capacity solutions were developed for the SVR river's health assessment. A reasonable output mode was designed, and the corresponding radial basis function neural network (RBF) regression model, which showed a good performance, was built as a comparison model. After the model ran five times stochastically, the absolute value of the average relative error, the absolute value of the maximum relative error, and the runtime were used to evaluate the performance of the model in each program. Finally, the SVR model that achieved the desired accuracy was evaluated and analyzed in a case study. The results are as follows: (1) For either the training sample or the testing sample, the SVR model in five programs had a higher prediction accuracy and better generalization ability than the RBF model. Under the same parameter setting conditions, as the sample size increased, the SVR model's accuracy and generalization ability changed insignificantly, while the RBF model's accuracy and generalization ability improved, indicating that the SVR model has higher accuracy and better generalization ability and can be used for river health assessment, especially in the cases of small samples. In this regard, the RBF model is totally uncompetitive. (2) The SVR model in the five programs was applied to the evaluation of the Qingshui River during the period from 2011 to 2012. The results of the survey carried out three times show that the river was healthy, but nearly sub-healthy.

Key words: river health; index system; grading standards; support vector regression machine; comprehensive assessment; Yunnan Province

1 研究背景

河流水系是地表水资源最重要的载体,是维系生态系统健康的主要因子,在维系地球水循环、能量平衡、气候变化和生态良性循环中具有重要作用。近年来,随着我国经济社会的快速发展,以及工业化、城市化进程的持续推进,水资源开发过度、水质下降、河湖萎缩和生态环境功能退化等问题日益凸显,河流生态安全受到严重威胁。开展河流健康评价对了解河流生态状况,掌握河流健康变化规律,保障河流水生态系统健康,实现水资源的可持续利用具有重要意义^[1]。河流健康状况是多种因素综合影响的结果,是一个涉及自然、社会、经济和生态等多方面的庞大系统工程,其评价表现出多指标、高维、非线性等特征。人工神经网络(artificial neural network,以下简称 ANN)是一种模仿动物神经网络行为特征,进行分布式并行信息处理的数学模型。ANN 依据系统的复杂程度,通过调整内部大量节点之间相互连接的关系来处理信息。在处理复杂、高维、非线性系统和不确定性问题方面,ANN 有着广泛的应用^[2]。然而,由于传统 ANN 算法是基于渐近理论,是当样本容量趋向于无穷大时的解决方案,因此,在实际应用中存在着当样本容量有限时可能出现较差的推广能力、结构及参数选择困难、收敛速度慢和易陷入局部极值等问题^[3-4]。支持向量机(support vector machine, SVM)是 20 世纪 90 年代中后期发展起来的基于统计学习理论构建的典型神经网络,是一种通用的前馈神经网络,用于解决模式识别和非线性映射问题。SVM 具有严谨的数学基础,通过统计学习中的 VC 维(vapnik-chervonenkis dimension)理论和寻求结构风险最小化原理来提高泛化能力,已成为继 ANN 之后机器学习领域新的研究热点,尤其在解决小样本容量问题时,很大程度上解决了传统 ANN 在模型选择、高维和局部极值等方面的问题,在模式识别和回归预测中有着广泛的应用^[5-6]。将 SVM 引入河流健康评价主要存在着 2 方面的问题:一是我国河流健康评价工作正处于试点阶段,与发达国家相比还存在较大差距,主要表现在未建立统一、完善的河流健康评价指标体系和分级标准,评价方法主要以主观赋分或赋权法为主,存在一定的缺陷;二是如何科学、合理评价河流健康评价模型的性能,以及如何解决数据样本、样本容量及输出模式等问题。

笔者参考文献[7-10],利用层次分析法(analytical hierarchy process, AHP)构建包含目标层、准则层和指标层 3 级的河流健康评价指标体系和

“理想状态”、“健康”、“亚健康”、“不健康”和“病态”5 个等级的分级标准;基于回归支持向量机(support vector regression, SVR)基本原理,提出基于不同样本容量的 SVR 河流健康评价模型,并构建具有良好性能的 RBF(radial basis function neural network, RBF)回归模型作为对比模型,以云南省文山州清水河为例进行验证,为即将全面发展的河流健康评价提供借鉴和参考。

2 河流健康内涵及评价指标体系

2.1 河流健康内涵

自从“河流健康”概念被提出以来,其内涵至今仍存在着争议,但由于“健康”比喻对于大众而言易于理解,引起了人们对河流生态状况的广泛关注。河流健康评价在西方发达国家和一些发展中国家得到了应用,以欧盟水框架指令、美国河湖健康评价、澳大利亚河流及湿地健康评价、南非河流健康计划等的影响最大^[1,8]。我国河流健康评价正处于试点阶段。从国外河流健康评价历程来看,河湖健康评价中提出的社会背景是人类经济社会的快速发展和水资源的过度开发严重威胁到河流生态系统生态服务功能的可持续提供,同时也严重威胁到河流生态系统所支撑维持的生物多样性。因此,河流健康可理解为:河流自然生态状况良好,同时具有可持续的社会服务功能。河流健康概念源于人类对河流环境退化的关注,其内涵不仅包含了河流为人类所提供的服务,还包含维持河流自身复杂特征,满足人类需求。河流健康的主体由河流生态系统以及所支撑的经济社会系统构成^[11-12]。

2.2 指标体系的建立

我国在构建评价指标体系过程中需考虑以下制约因素:①由于我国流域内人口密集,生态用水和经济社会用水竞争激烈,导致在健康河流评价指标上更强调维持河流自身生命及其功能的健康,更关注和突出人水关系;②在河流健康评价时,仍侧重于借助化学手段以及少量生物监测手段评估河流水质状况;③实例研究多以单条河流为主,缺乏对流域、水系、河流不同空间尺度的探讨;④迄今国内尚未形成统一或公认的河流健康评价指标体系。笔者充分考虑我国河流生态系统的特征及经济社会发展背景,遵循科学性、系统性、层次性、独立性和指标定量性与可操作性的指标选取原则,利用 AHP 方法从水文水资源(HD)、物理结构(PF)、水质(WQ)、水生生物(AL)和社会服务功能(SS)5 个方面遴选出 13 个评价指标构成河流健康评价指标体系,将河流健康评价分为目标层 A、准则层 B 和指标层 C 3 个层次。

河流健康评价指标体系见表 1。

2.3 指标体系等级的建立

按照上述所构建的河流健康评价指标体系,参考文献[7,9-10],构建河流健康评价指标分级标准,将河流健康评价分为理想状态、健康、亚健康、不健康和病态 5 个等级,分别用 5~1 级表示,见表 2。

3 SVR 评价模型

SVM 最初是为研究线性问题而提出的,其用于模式识别的基本思想是通过非线性变换将输入空间变换到一个高维空间,在此新空间通过求解凸二次规划问题,寻求最优线性分类超平面;而用于回归预测时,其基本思想不再是寻找最优分类面来将样本分开,而是寻找一个最优超平面,使得所有训练样本离该最优超平面距离最短,这个超平面可看作拟合

好的曲线。将 SVM 用于逼近函数的方法称为 SVR。SVR 实现回归预测步骤归纳如下^[5,13]。

步骤 1 设含有 l 个训练样本的集合 $\{(x_i, y_i), i=1, 2, \dots, l\}$, 其中 $x_i(x_i \in R^d)$ 是第 i 个训练样本的输入列向量, $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T$, $y_i \in R$ 为对应输出值。在高维特征中建立的线性回归函数为

$$f(x) = w\Phi(x) + b \quad (1)$$

式中: $f(x)$ 为回归函数返回的预测值; $\Phi(x)$ 为非线性映射函数; w 为超平面的权值向量, b 为偏置项。

步骤 2 定义 ε 线性不敏感损失函数为

$$L(f(x), y, \varepsilon) = \begin{cases} 0 & |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & |y - f(x)| > \varepsilon \end{cases} \quad (2)$$

式中: y 为对应真实值; ε 为不敏感损失函数所定义的误差。

表 1 河流健康评价指标体系

目标层 A	准则层 B	指标层 C	指标选取或量化方法
河流健康评价	水文水资源(HD)	流量过程变异程度 C_1	采用评价年逐月实测径流量与天然月径流量的平均偏离程度表达
		生态流量满足程度 $C_2(\%)$	采用实测日径流量占多年平均径流量的最低百分比来表征生态流量满足程度
	物理结构(PF)	河岸带状况 $C_3(\%)$	对岸坡稳定性(0.25)、河岸植被覆盖率(0.5)和河岸人工干扰程度(0.25)3个指标分别赋分,采用加权平均得到河岸带状百分比
		河流纵向连续性 $C_4(\%)$	将水利设施分为4类情况:完全阻隔、严重阻隔、阻隔、轻度阻隔。对评价河段每个水利设施按照阻隔分类分别赋分,取其均值作为河流纵向连续性百分比
	水质(WQ)	DO水质状况 $C_5(\text{mg/L})$	采用8~10个断面均值表示
		耗氧有机污染状况 $C_6(\%)$	对高锰酸盐指数、化学需氧量、氨氮分别赋分。取3个水质项目赋分的均值作为耗氧有机污染状况赋分值
		重金属污染状况 $C_7(\%)$	对汞、镉、铬、铅及砷分别赋分,取5个水质项目最低赋分值作为该项指标赋分值污染状况指标赋分
	水生生物(AL)	Pielou均匀度指数 C_8	选取对健康评价最不利指标 Pielou 均匀度指数;采用8个调查河段的 Pielou 均值表示
		附着硅藻指数 C_9	选取对健康评价最不利指标硅藻生物指数(IBD);采用7~8个调查河段的 IBD 均值表示
	社会服务功能(SS)	水功能区达标率 $C_{10}(\%)$	采用符合水功能区水质标准的水功能区数量与流域水功能区监测总数之比表示
水资源开发利用率 $C_{11}(\%)$		采用流域内供水量占流域水资源量的百分比表示	
防洪指标 $C_{12}(\%)$		采用达标河段长度占总河段长度的百分比表示	
		公众满意程度 $C_{13}(\%)$	收集分析公众调查表,统计有效调查表调查成果,根据公众类型和公众总体评价赋分得到公众满意度指标

表 2 河流健康评价指标分级标准

等级	状态	C_1	$C_2/\%$	$C_3/\%$	$C_4/\%$	$C_5/(\text{mg} \cdot \text{L}^{-1})$	$C_6/\%$	$C_7/\%$
5级	理想状态	<0.1	≥80	≥80	≥80	≥7.5	≥90	≥80
4级	健康	[0.1 0.3)	[60 80)	[60 80)	[60 80)	[6 7.5)	[70 90)	[60 80)
3级	亚健康	[0.3 1.5)	[40 60)	[40 60)	[40 60)	[5 6)	[50 70)	[40 60)
2级	不健康	[1.5 3.5)	[20 40)	[20 40)	[20 40)	[3 5)	[30 50)	[20 40)
1级	病态	≥3.5	<20	<20	<20	<3	<30	<20
等级	状态	C_8	C_9	$C_{10}/\%$	$C_{11}/\%$	$C_{12}/\%$	$C_{13}/\%$	
5级	理想状态	≥0.5	≥17	≥90	<20	≥95	≥90	
4级	健康	[0.4 0.5)	[13 17)	[80 90)	[20 40)	[80 95)	[70 90)	
3级	亚健康	[0.3 0.4)	[9 13)	[60 80)	[40 50)	[60 80)	[50 70)	
2级	不健康	[0.2 0.3)	[5 9)	[40 60)	[50 60)	[40 60)	[30 50)	
1级	病态	<0.2	<5	<40	≥60	<40	<30	

步骤3 类似于 SVM 分类情况,引入松弛变量 ξ_i, ξ_i^* , 并将上述寻找 w, b 的问题用数学语言描述出来,即,

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t.} \begin{cases} y_i - w\Phi(x_i) - b \leq \varepsilon + \xi_i \\ -y_i + w\Phi(x_i) + b \leq \varepsilon + \xi_i^* \quad (i = 1, 2, \dots, l) \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \end{cases} \quad (3)$$

式中: C 为惩罚因子, C 越大表示对训练误差大于 ε 的样本惩罚越大, ε 规定了回归函数的误差要求, ε 越小表示回归函数的误差越小。求解式(3)时,同时引入 Lagrange 函数,并转换成对偶形式:

$$\begin{cases} \max \left[-\frac{1}{2} \sum_{i=1}^l \sum_{i=j}^l (a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) - \sum_{i=1}^l (a_i + a_i^*)\varepsilon + \sum_{i=1}^l (a_i - a_i^*)y_i \right] \\ \text{s. t.} \begin{cases} \sum_{i=1}^l (a_i - a_i^*) = 0 \\ 0 \leq a_i \leq C \\ 0 \leq a_i^* \leq C \end{cases} \end{cases} \quad (4)$$

式中: $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$, 为核函数; a_i, a_i^* 为拉格朗日乘子。

步骤4 设求解式(4)得到的最优解为 $a = [a_1, a_2, \dots, a_l], a^* = [a_1^*, a_2^*, \dots, a_l^*]$, 则有:

$$w^* = \sum_{i=1}^l (a_i - a_i^*) \Phi(x_i) \quad (5)$$

$$b^* = \frac{1}{N_{\text{nsv}}} \left\{ \sum_{0 < a_i < C} [y_i - \sum_{x_j \in \text{SV}} (a_i - a_i^*)K(x_i, x_j) - \varepsilon] + \sum_{0 < a_i^* < C} [y_i - \sum_{x_j \in \text{SV}} (a_i - a_i^*)K(x_i, x_j) + \varepsilon] \right\} \quad (6)$$

式中, N_{nsv} 为支持向量个数。

步骤5 将 w^*, b^* 代入式(1)得到回归函数为

$$f(x) = \sum_{i=1}^l (a_i - a_i^*)K(x_i, x) + b^* \quad (7)$$

其中,只要部分参数 $(a_i - a_i^*)$ 不为 0, 其对应的样本 x_i 即为问题中的支持向量。

表3 5种方案的样本及期望输出设计

训练(检验)样本/组					评价等级	状态	输出结果	
方案1	方案2	方案3	方案4	方案5			SVR	RBF
1~5	1~10	1~20	1~50	1~100	5级	理想状态	0.1	0.1
6~10	11~20	21~40	51~100	101~200	4级	健康	0.2	0.2
11~15	21~30	41~60	101~150	201~300	3级	亚健康	0.3	0.3
16~20	31~40	61~80	151~200	301~400	2级	不健康	0.4	0.4
21~25	41~50	81~100	201~250	401~500	1级	病态	0.5	0.5

4 网络训练与性能评价

4.1 指标数据归一化处理

为了消除表2中各评价指标不同量纲及“方向”对评价结果的影响,需对评价指标数据进行归一化和一致性处理。对于指标值越大河流健康状态越理想的指标,按式(8)进行归一化处理;对于指标值越小河流健康状态越理想的指标,对其取倒数后乘以100,再按式(8)进行处理。

$$\hat{x} = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (8)$$

式中: \hat{x} 为经过归一化处理的数据; x 为原始数据;为了使各评价指标具有相同的权重,这里取 x_{\max} 为各评价指标上限阈值的2倍; x_{\min} 为各评价指标下限阈值的10%。经过标准化处理后,数据处于0~1范围之内,有利于网络训练。

4.2 样本及输出设计

在解决小样本回归预测问题上,基于结构风险最小化原则的 SVR 比基于经验风险最小化原则的传统 ANN 有着较大优势。为验证不同样本容量对 SVR 模型性能的影响,采用随机生成样本的方法在各评价等级阈值间生成5种不同容量大小的样本方案,随机选取样本总量的3/5作为训练样本,余下的样本作为检验样本。设计各模型的输出模式见表3。

4.3 网络训练及性能评价比较

本文基于 Matlab 环境和 libsvm 工具箱,创建 SVR 及 RBF 回归模型对表3设计的5种方案的样本进行训练和检验,经过反复调试,SVR 及 RBF 模型参数设置如下时,模型具有较好的预测精度(为使模型在不同样本容量情况下具有可比性,SVR 及 RBF 的5种模型方案除样本容量大小不一致外,其余参数设置均相同)。

SVR 模型:SVR 在选定核函数条件下,模型中的惩罚因子 C 和核函数参数 g 的选取对模型的识别精度有着关键性影响。参考文献[4-5,10],选择径向基核函数为 SVR 的核函数,设置惩罚因子 C 和核函数参数 g 的搜索空间均设置为 $2^{-2} \sim 2^6$,交叉验证折数 K 设置为5, g 和 C 的步进大小均取0.5,不敏感系数 ε 为0.001(其他参数采用默认值),利用

交叉验证法 (cross validation, CV) 确定模型中的惩罚因子 C 和核函数参数 g (由于采用随机生成及随机选取样本, 因此每次运行的参数可能不同)。

RBF 模型: 编写循环训练算法程序, 最终确定 RBF 神经网络在径向基函数分布密度 spread 和期望误差分别为 1 和 0.0001 时, 模型性能达到最优。

本文以平均相对误差绝对值、最大相对误差绝对值和模型运行时间作为评价 SVR 及 RBF 回归模型性能优劣的指标。由于训练样本和检验样本是通过随机生成和随机选取产生, 因此模型每次运行结果均不一样。某次随机连续 5 次运行的统计指标平均值见表 4。

表 4 SVR 及 RBF 回归模型误差及运行时间

方案	平均相对误差绝对值/%				最大相对误差绝对值/%				运行时间/s	
	SVR		RBF		SVR		RBF		SVR	RBF
	训练样本	检验样本	训练样本	检验样本	训练样本	检验样本	训练样本	检验样本		
1	1.65	3.06	1.86	6.91	7.76	11.91	6.79	23.38	0.112	1.34
2	1.67	2.26	2.71	5.49	10.01	9.60	10.56	20.56	0.381	1.36
3	2.13	3.1	2.51	4.04	11.05	12.95	11.56	15.28	1.370	1.46
4	2.49	2.82	2.51	3.47	16.46	13.96	17.69	17.21	9.180	3.28
5	1.86	2.58	2.64	3.06	13.38	14.52	17.22	17.06	40.800	4.36

由表 4 可得出以下结论: ①从 SVR 模型方案 1 至方案 5 的误差值来看, 其训练样本及检验样本的平均相对误差绝对值和最大相对误差绝对值分别在 1.65% ~ 2.49%、2.26% ~ 3.10% 和 7.76% ~ 16.46%、9.60% ~ 14.52% 之间, 均优于对应的 RBF 模型, 表明 SVR 回归模型具有较好的预测精度和泛化能力, 将其用于回归预测评价是合理可行的。②从同一模型不同方案的误差值来看, SVR 模型随着样本容量的增加, 其预测精度和泛化能力变化不大; RBF 模型随着样本容量的增加, 其预测精度和泛化能力均有提升, 表明 SVR 模型对样本容量大小不敏感, 尤其在小样本情况下, SVR 模型的预测精度和泛化能力是 RBF 所不可比拟的。③从模型的运行时间来看, SVR 模型随着样本容量的增加运行时间显著增加, RBF 模型随着样本容量的增加运行时间略有增加, 表明增加样本容量无助于提高 SVR 模型的预测精度和泛化能力, 反而使 SVR 模型性能下降。

本文主要基于模型的预测精度和泛化能力考

表 6 SVR 模型河流评价等级临界值输出结果 (随机 5 次平均)

河流等级	状态	临界值				
		方案 1	方案 2	方案 3	方案 4	方案 5
5 级	理想状态	>0.4285	>0.4371	>0.4345	>0.4355	>0.4135
4 级	健康	(0.3427 0.4285]	(0.3476 0.4371]	(0.3502 0.4345]	(0.3548 0.4355]	(0.3599 0.4135]
3 级	亚健康	(0.2415 0.3427]	(0.2465 0.3476]	(0.2495 0.3502]	(0.2538 0.3548]	(0.2596 0.3599]
2 级	不健康	(0.1400 0.2415]	(0.1457 0.2465]	(0.1485 0.2495]	(0.1492 0.2538]	(0.1428 0.2596]
1 级	病态	≤0.1400	≤0.1457	≤0.1485	≤0.1492	≤0.1428

虑, 选取不同样本容量的 SVR 模型作为文山州清水河健康评价模型。

5 实例应用

5.1 研究区概况

清水河又名北门河, 位于云南省丘北县境内, 发源于丘北县八道哨乡五家寨西缘, 自源地向东流经普者黑湖, 于天星乡小法白村汇入清水江, 属珠江流域西江水系, 为南盘江的二级支流。清水河河长 61.5 km, 落差 727 m, 平均坡降 0.77%, 有清平河、旧城河、高枳槽河、碧松就河等多条支流, 集水面积 1533.5 km², 多年平均流量 18.2 m³/s, 多年平均径流量 5.11 亿 m³。清水河上游建有红旗水库, 又称摆龙湖, 总库容 5400 万 m³; 中游是普者黑湖, 为河道型湖泊群, 水面面积约为 6 km², 水深 3 ~ 4 m。近年来, 随着流域内经济社会及旅游业的迅猛发展, 清水河受到了不同程度的污染, 河流生态系统结构、社会服务功能等随之发生了变化。因此, 开展清水河健康评价对了解河流生态健康状况、掌握健康变化规律、制定流域规划具有重要意义^[1]。清水河 2011—2012 年 3 次调查评价指标数据详见表 5。

表 5 清水河健康评价指标数据^[1]

调查时间	C_1	$C_2/\%$	$C_3/\%$	$C_4/\%$	$C_5/$ (mg · L ⁻¹)	$C_6/\%$	$C_7/\%$
2011-04	0.16	35.3	29.6	83.3	6.7	86.7	100
2012-04	0.16	36.2	29.6	83.3	7.9	85.0	100
2012-10	0.16	35.7	29.6	83.3	7.3	98.2	100
调查时间	C_8	C_9	$C_{10}/\%$	$C_{11}/\%$	$C_{12}/\%$	$C_{13}/\%$	
2011-04	0.697	17.6	66.7	6.99	11.5	48.4	
2012-04	0.641	11.9	100	6.99	11.5	48.4	
2012-10	0.783	11.6	100	6.99	11.5	48.4	

5.2 评价结果及分析

利用上述训练好的 5 种样本容量方案的 SVR 模型对清水河健康状况进行评价, 并将表 2 中河流健康评价分级阈值进行模拟计算, 将输出结果作为划分河流健康等级的依据。输出及评价结果见表 6 和表 7。

从表 6 和表 7 可以得出以下结论: ① 5 种样本容量方案的 SVR 模型对清水河的 3 次调查评价结果均为 4 级, 即处于健康状态, 但从输出结果来看, 已接近于亚健康状态。评价结果基本反映了目前清

表 7 SVR 模型清水河健康评价结果(随机 5 次平均)

调查时间	方案 1			方案 2			方案 3			方案 4			方案 5		
	输出值	评价等级	状态												
2011-04	0.3505	4 级	健康	0.3570	4 级	健康	0.3607	4 级	健康	0.3628	4 级	健康	0.3929	4 级	健康
2012-04	0.3483	4 级	健康	0.3598	4 级	健康	0.3638	4 级	健康	0.3655	4 级	健康	0.3955	4 级	健康
2012-10	0.3544	4 级	健康	0.3677	4 级	健康	0.3704	4 级	健康	0.3742	4 级	健康	0.3992	4 级	健康

水河健康状态。②从清水河健康调查评价指标值可以看出,随着流域内水资源的开发利用和人类活动的加剧,河流在生态流量满足程度、河岸稳定性、植被覆盖率,以及防洪设施、公众满意度等方面表现较差,导致清水河目前虽处于健康状态,但已接近亚健康状态。

6 结 语

河流健康是一个相对概念,具有一定的主观性。依据《湖泊健康评估指标、标准与方法(试点工作)》,从河流的功能属性出发,以河流健康概念为基础,笔者提出了包括水文完整性、物理结构完整性、化学完整性、生物完整性和服务功能完整性 5 个方面的河流健康评价指标体系和分级标准,构建了不同样本容量的数据样本,提出基于 SVR 回归原理的河流健康评价模型,对清水河健康状况进行了评价。但由于河流健康评价属于多学科交叉综合评价问题,其评价指标体系涉及多个领域,内容广泛,加之制约河流健康的因素众多,因此,本文所提出的河流健康指标体系及评价方法还有待进一步研究和完善。

参考文献:

[1] 云南省水文水资源局. 云南省文山州清水河健康评估试点报告(送审稿)[R]. 昆明:云南省水文水资源局, 2012.

[2] 田景文,高美娟. 人工神经网络算法研究及应用[M]. 北京:北京理工大学出版社,2006.

[3] 田雨波. 混合神经网络技术[M]. 北京:科学出版社, 2009.

[4] 王雷. 支持向量机在汽轮机状态监测中的应用[M]. 北京:北京师范大学出版社,2012.

[5] 崔东文. 支持向量机在湖库营养状态识别中的应用研究[J]. 水资源保护, 2013, 29(4): 26-30. (CUI Dongwen. Application of support vector machine to lake and reservoir trophic status recognition [J]. Water Resources Protection, 2013, 29(4): 26-30. (in Chinese))

[6] 张楠,夏自强,江红. 基于多因子量化指标的支持向量机径流预测[J]. 水利学报, 2010, 41(11): 1318-1323. (ZHANG Nan, XIA Ziqiang, JIANG Hong. Prediction of runoff based on the multiple quantity index of SVM[J]. Journal of Hydraulic Engineering, 2010, 41(11): 1318-1323. (in Chinese))

[7] 办资源[2010]484 号 河流健康评估指标、标准与方法(试点工作)[S].

[8] 孙雪岚,胡春宏. 河流健康评价指标体系初探[J]. 泥沙研究, 2007(4): 21-27. (SUN Xuelan, HU Chunhong. River health evaluation index system [J]. Journal of Sediment Research, 2007(4): 21-27. (in Chinese))

[9] 崔东文. RBF 与 GRNN 神经网络模型在河流健康评价中的应用;以文山州区域中小河流健康评价为例[J]. 中国农村水利水电, 2012(3): 56-61. (CUI Dongwen. RBF and GRNN neural network model in the evaluation of river health[J]. China Rural Water and Hydropower, 2012(3): 56-61. (in Chinese))

[10] 崔东文,郭荣. 基于 GRNN 模型的区域水资源可持续利用评价;以云南文山州为例[J]. 人民长江, 2012, 43(5): 26-31. (CUI Dongwen, GUO Rong. Evaluation of sustainable utilization of regional water resources based on GRNN neural network model: case of Wenshan Prefecture of Yunnan Province[J]. Yangtze River, 2012, 43(5): 26-31. (in Chinese))

[11] 杨文慧,严忠民,吴建华. 河流健康评价的研究进展[J]. 河海大学学报:自然科学版, 2005, 33(6): 607-611. (YANG Wenhui, YAN Zhongmin, WU Jianhua. Advances in river health assessment[J]. Journal of Hohai University: Natural Sciences, 2005, 33(6): 607-611. (in Chinese))

[12] 蔡守华,胡欣. 河流健康的概念及指标体系和评价方法[J]. 水利水电科技进展, 2008, 28(1): 23-27. (CAI Shouhua, HU Xin. Concept of river health and index system for its evaluation [J]. Advances in Science and Technology of Water Resources, 2008, 28(1): 23-27. (in Chinese))

[13] 史峰,王辉,郁磊,等. MATLAB 智能算法 30 个案例分析[M]. 北京:北京航空航天大学出版社,2011.

(收稿日期:2013-08-26 编辑:彭桃英)

