最小二乘支持向量机在洪水预报中的应用

丁海蛟,车文刚　(昆明理工大学信息工程与自动化学院,云南昆明 650500)

摘要　由于洪水这种自然灾害严重影响着正常的农业生产,该研究提出基于最小二乘支持向量机(LS - SVM)的洪水预报方法。利用LS - SVM 泛化能力强、收敛速度快、预测精度高的特性,结合四川省自贡市某水文站实测水位和流量的数据,建立了单输入单输出(水位—水位)和双输入单输出(水位和流量—水位)2 种洪水预报模型。结果表明:这2 种模型都可以满足洪水预报精度的要求,并且双输入单输出(水位和流量—水位)模型的预报精度比单输入单输出(水位—水位)模型的预报精度高。

关键词　水位;流量 ;LS-SVM;洪水预报
中图分类号　S422　　**文献标识码**　A　　**文章编号**　0517 - 6611(2015)23 - 350 - 04

The Application of LS-SVM in Flood Prediction

DING Hai-jiao, CHE Wen-gang　(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500)

Abstract　Since the flood natural disasters seriously affect the normal agricultural production, this paper gives a LS-SVM method for flood prediction. Using the features of strong generalization ability, fast convergence rate, high prediction accuracy of LS-SVM, combining with a hydrological station in Sichuan Zigong measured data of water level and flow set up a single input single output (water level-water level) and double input single output (water level and flow-water level) two kinds of flood prediction model. Experimental results show that both models can meet the requirements of flood prediction accuracy, and the accuracy of the double input single output (water level and flow-water level) prediction model is higher than the single input single output (water level-water level) prediction model.

Key words　Water level; Flow; LS-SVM; Flood prediction

目前在全世界范围内,洪水是最具毁灭性的自然灾害之一,严重影响了正常的农业生产。为了保障正常的农业生产,防洪是非常必要的。所谓防洪,就是指根据具体的灾情特点以及水文事件演进的变化规律,研究并且采用相关的方法和对策,通过预防来减轻洪水灾害所造成的损失。防洪的主要措施有工程措施和非工程措施。工程措施造价高,而且实施的周期长,所以洪水预报作为一项重要的防洪非工程措施,它在预防洪水和防洪预案的制定中有着非常重要的作用。作为洪水预报研究对象之一的河道洪水预报可以分为水位预报和流量预报2 种。笔者以河道的水位预报为研究对象,提出了基于最小二乘支持向量机(LS - SVM)的洪水预报方法。

1　洪水预报研究概况

目前有很多洪水预报的方法,例如,法国、英国和丹麦的科学家们共同合作研制了十分具有代表性的分布式水文物理模型,也就是 SHE 模型[1];雷晓云等研究了 BP 神经网络在河道流量预报中的应用[2];许剑华将最小二乘滤波算法构造时变参数的动态处理模式用在汇流预报模型中的回归型差分方程模型中,并将这种方法在石桥站中进行试用,达到了预期的提高模型预报精度目的[3]。近年来神经网络在预测领域被普遍的应用,但人工神经网络的局部极小点、类型和结构的选择都十分的依赖于经验,这就在很大程度上限制了其更加广泛的应用。

与人工神经网络基于经验风险最小化原则不同,基于统计学习理论的支持向量机方法根据结构风险最小化原则,在最大程度上提高了其泛化能力,其算法的局部最优解就是全局最优解[4]。最小二乘支持向量机(LS - SVM)是支持向量机的改进形式,采用平方项作为优化指标,并且用等式约束代替标准支持向量机的不等式约束,也就是将二次规划问题(QP)转化为线性方程组的求解问题,降低了计算的复杂性、提高了求解的速度。该研究应用 LS - SVM 建立了单输入单输出(水位—水位)和双输入单输出(水位和流量—水位)2 种河道水位预测模型,并且对这2 种模型进行了对比。

2　LS - SVM 算法

样本数据的个数决定了标准支持向量机算法的复杂程度,因此样本数据越大,相应的二次规划问题(QP)将变得更加的复杂,计算的速度会更慢[5]。而最小二乘支持向量机,采用了不同的优化目标函数,并且用等式约束代替了原来的不等式约束,这不仅降低了计算的复杂性,还提高了求解的速度。

设训练样本集为:$\{x_i,y_i\}_{i=1}^n$,其中,$x_i\in R^d$,$y_i\in\{+1,-1\}$,ϕ 为核函数的匿名映射,则 LS - SVM 可以表示成如下的约束优化问题:

$$\min_{\omega,b,e}\phi(\omega,b,e)=\frac{1}{2}\|\omega\|^2+\gamma\frac{1}{2}\sum_{i=1}^{n}e_i$$
$$s.t.\ y_i[w^T\varphi(x_i)+b]=1-e_i,i=1,\cdots,n \quad (1)$$

用 Lagrange 法求解上述优化问题,则可以转化为求解如下的线性方程组:

$$\begin{bmatrix}I&0&0&-H^T\\0&0&0&-Y^T\\0&0&rI&-I\\H&Y&I&0\end{bmatrix}\begin{bmatrix}\omega\\b\\e\\\alpha\end{bmatrix}=\begin{bmatrix}0\\0\\0\\\overrightarrow{1}\end{bmatrix} \quad (2)$$

式中,$H=[\varphi(x_1)^Ty_1\cdots\varphi(x_n)^Ty_n]$,$Y=[y_1,\cdots,y_n]$,$\overrightarrow{1}=[1,\cdots,1]$,$e=[e_1,\cdots,e_n]$,$\alpha=[\alpha_1,\cdots,\alpha_n]$,消去 ω、e,得到下式:

$$\begin{bmatrix}0&-Y^T\\Y&\Omega+\gamma^{-1}I\end{bmatrix}\begin{bmatrix}b\\\alpha\end{bmatrix}=\begin{bmatrix}0\\\overrightarrow{1}\end{bmatrix} \quad (3)$$

作者简介　丁海蛟(1989 -),男,辽宁朝阳人,硕士研究生,研究方向:智能控制。
收稿日期　2015-06-15

式(3)中,$\Omega = HH^T$。设 $K(x_k, x_l) = \varphi(x_k)^T\varphi(x_l)$,满足 Mercer 条件,就是核函数,则:

$$\Omega_{kl} = y_k y_l \varphi(x_k)^T \varphi(x_l) = y_k y_l K(x_k, x_l) \quad (4)$$

所求的 LS－SVM 的输出是:

$$y(x) = \mathrm{sign}\left[\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b\right], \alpha_k \in R^N \quad (5)$$

核函数的类型较多,比较常用的有 4 种[6],该研究选取最常用的核函数,即 RBF 核函数:$K(x_i, x) = \exp\{-|x - x_i|^2 / 2\sigma^2\}$。其中,正则化参数 γ 和核宽度 σ 是 LS－SVM 必须要进行调整的 2 个参数,因为正则化参数 γ 和核宽度 σ 是作为一个整体使用,所以 γ 和 σ 的取值将直接决定了 LS－SVM 的训练和泛化的能力。

a 水位—水位预测模型

b 水位和流量—水位预测模型

图 1　2 个 LS－SVM 预测模型

3　LS－SVM 预测实验

该研究建立了 2 个预测模型——单输入单输出(水位—水位)预测模型和双输入单输出(水位和流量—水位)预测模型,如图 1 所示。

为了验证LS－SVM预测方法的可行性以及有效性,以四川省自贡市某水文站 2008 年 7 月～2014 年 7 月(即分别取每年 7 月份汛期)每天平均的水位数据和流量数据作为 LS－SVM 模型的样本数据。其中 2008 年 7 月～2013 年 7 月的数据作为模型学习和训练的数据,2014 年 7 月的数据作为验证数据。

3.1　样本数据预处理　将样本数据集的所有数据进行归一化处理,这样做的目的是:①避免一些特征值的范围过小,而另一些特征值的范围过大;②避免在进行训练的时候,为了计算核函数而计算内积的时候而引起数值计算困难,可能会影响建模的效果和计算的速度。因此就要将样本数据重新缩放到一个适当的范围内,该研究将样本数据缩放到[0,1]的范围内。归一化公式如下:

$$x_i = \frac{|x_i - x_{\min}|}{|(x_{\max} - x_{\min})|} \quad (6)$$

在训练结束时,再对数据进行反归一化,这样做的目的是:可以直观地看到预测数据的实际值而并非在[0,1]这个范围内的数值。

3.2　预测结果的评定标准　为了验证预测模型的可行性和有效性,在这里先给出预测结果的评定标准。该研究所采用的精度评定标准来自于中华人民共和国水利部《水文情报预报规范》(SL250－2000)[7]。这里给出一些比较重要的参数取值范围:

(1)结合现场的实际情况和《水文情报预报规范》的有关规定,设置水位许可误差为 0.2 m。

(2)合格率。计算公式为:

$$QR = \frac{n}{m} \times 100\% \quad (7)$$

式中,n 是合格预报次数;m 是预报总次数。其中:$QR \geqslant 85.0\%$时,为甲级精度;$70.0\% \leqslant QR < 85.0\%$ 时,为乙级精度;$60.0\% \leqslant QR < 70.0\%$时,为丙级精度。

(3)确定性系数

$$DC = 1 - \frac{\sum_{i=1}^{n}(y_c(i) - y_0(i))^2}{\sum_{i=1}^{n}(y_0(i) - \bar{y}_0)^2} \quad (8)$$

式中,$y_0(i)$为实测值;$y_c(i)$为预报值;y_0 为实测值均值;n 为资料序列长度。其中:$DC \geqslant 0.90$ 时,为甲级精度;$0.70 \leqslant DC < 0.90$ 时,为乙级精度;$0.50 \leqslant DC < 0.70$ 时,为丙级精度。

(4)绝对误差

$$AE = y_c(i) - y_0(i) \quad (9)$$

相对误差

$$RE = \frac{y_c(i) - y_0(i)}{y_0(i)} \times 100\% \quad (10)$$

式中,$y_0(i)$为实测值;$y_c(i)$为预报值。

(5)作业预报精度 GI 为预报误差和许可误差之比。$GI \leqslant 25.0\%$,为优秀;$25.0\% < GI \leqslant 50.0\%$,为良好;$50.0\% < GI \leqslant 100.0\%$,为合格;$GI > 100.0\%$,为不合格。

(6)经过精度评定,洪水预报方案的精度达到甲、乙 2 个等级者,可以用于发布正式的预报;方案精度达到丙级者,可以用于参考性预报;丙级以下者,只能用于参考性估报。

3.3　单输入单输出水位预测模型(水位—水位)　首先选取核函数。对于最小二乘支持向量机核函数的选择问题,目前在国际上还没有形成一个统一的选择模式,一般凭借经验来进行选择,而且一般优先考虑径向基函数(RBF 核函数),所以该研究中也选择 RBF 核函数。其次,确定 γ 和 σ^2 2 个参数。根据参考文献[8],采用网格搜索法寻找最优参数,最终确定 $\gamma = 295.6$,$\sigma^2 = 301.2$。利用 2008 年 7 月～2013 年 7 月的数据对模型进行训练后,对 2014 年 7 月测试集的数据进行预测,预测结果见图 2。

对模型的输出进行分析,结果见表 1。

为了验证输入的水位影响因子的多少对水位预测结果产生的影响,接下来增加一个影响因子——流量,即把模型变为双输入单输出(水位和流量—水位)。

3.4　双输入单输出水位预测模型(水位和流量—水位)　首先进行核函数的选取,该研究选择 RBF 核函数;然后确定 γ 和 σ^2 2 个参数。根据参考文献[8],采用网格搜索法寻找最优参数,最终确定 $\gamma = 256.7$,$\sigma^2 = 300.0$。利用 2008 年 7 月～2013 年 7 月的数据对模型进行训练后,对 2014 年 7 月测试集的数据进行预测,结果见图 3。

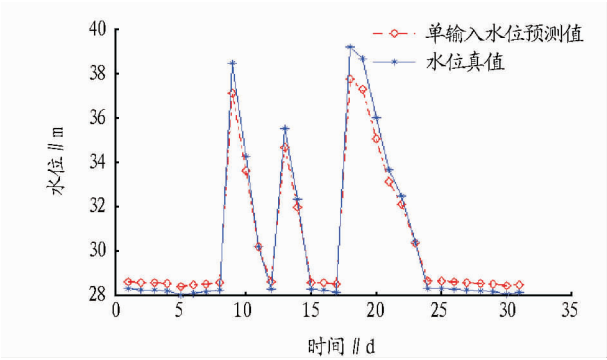

图 2 单输入单输出 LS – SVM 水位预测结果

表 1 单输入单输出水位分析

验证数据 m	LS – SVM 预测数据 // m	LS – SVM 绝对误差 // m	LS – SVM 相对误差 // %	预报是否合格
28.29	28.41	0.12	0.42	是
28.22	28.35	0.13	0.46	是
28.23	28.36	0.13	0.46	是
28.20	28.33	0.13	0.46	是
28.02	28.16	0.14	0.49	是
28.09	28.22	0.13	0.46	是
28.15	28.28	0.13	0.46	是
28.22	28.35	0.13	0.46	是
38.46	37.97	0.49	1.27	否
34.27	34.03	0.24	−0.70	否
30.18	30.19	0.01	0.03	是
28.27	28.39	0.12	0.42	是
35.53	35.22	0.31	−0.87	否
32.32	32.20	0.12	−0.37	是
28.25	28.37	0.12	0.42	是
28.23	28.36	0.04	0.14	是
28.14	28.27	0.13	0.46	是
39.21	38.68	0.53	−1.35	否
38.70	38.20	0.50	−1.29	否
36.01	35.67	0.34	−0.94	否
33.67	33.47	0.20	−0.59	是
32.49	32.36	0.13	−0.40	是
30.39	30.40	0.01	0.03	是
28.30	28.42	0.12	0.42	是
28.30	28.42	0.12	0.42	是
28.27	28.39	0.12	0.42	是
28.24	28.37	0.13	0.56	是
28.19	28.32	0.13	0.46	是
28.17	28.30	0.13	0.46	是
28.05	28.19	0.14	0.59	是
28.13	28.26	0.13	0.46	是

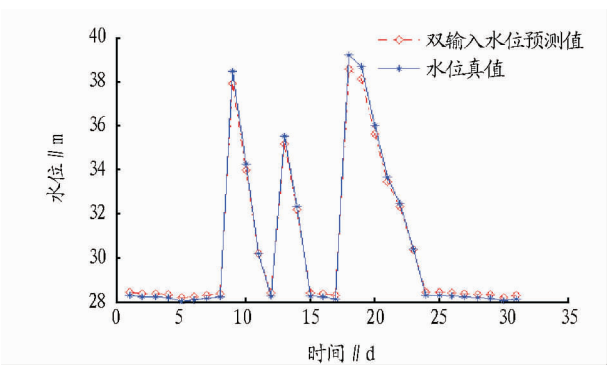

图 3 双输入单输出 LS – SVM 水位预测结果

对模型的输出进行分析,结果见表 2。

表 2 双输入单输出模型水位分析

验证数据 m	LS – SVM 预测数据 // m	LS – SVM 绝对误差 // m	LS – SVM 相对误差 // %	预报是否合格
28.29	28.43	0.14	0.49	是
28.22	28.37	0.15	0.53	是
28.23	28.38	0.15	0.53	是
28.20	28.35	0.15	0.53	是
28.02	28.18	0.16	0.57	是
28.09	28.25	0.16	0.57	是
28.15	28.30	0.15	0.53	是
28.22	28.37	0.15	0.53	是
38.46	37.90	0.56	−1.46	否
34.27	34.10	0.17	−0.49	是
30.18	30.19	0.01	0.03	是
28.27	28.41	0.14	0.50	是
35.53	35.37	0.16	−0.45	是
32.32	32.18	0.14	−0.43	是
28.25	28.39	0.14	0.50	是
28.23	28.38	0.15	0.53	是
28.14	28.29	0.15	0.53	是
39.21	38.59	0.62	1.58	否
38.70	38.12	0.58	−1.50	否
36.01	35.62	0.39	1.08	否
33.67	33.54	0.13	−0.39	是
32.49	32.34	0.15	−0.46	是
30.39	30.38	0.01	−0.03	是
28.30	28.44	0.14	0.49	是
28.30	28.44	0.14	0.49	是
28.27	28.41	0.14	0.50	是
28.24	28.39	0.15	0.53	是
28.19	28.34	0.15	0.53	是
28.17	28.32	0.15	0.53	是
28.05	28.21	0.16	0.57	是
28.13	28.28	0.15	0.53	是

3.5 结果分析 根据表 1、2 数据,计算得到 2 种模型的预测精度:单输入水位预测模型的合格率 $QR = 80.6\%$,确定系数 $DC = 0.996\,4$,$GI = 94.80\%$;双输入水位预测模型的合格率 $QR = 87.10\%$,确定系数 $DC = 0.998\,5$,$GI = 91.45\%$。计算表 1、2 中 2 种模型的绝对误差与相对误差的平均值,得出如表 3 所示的 2 种模型精度对比。

表 3 2 种模型精度对比

模型	*QR* %	*DC*	*GI* %	绝对误差均值 // m	相对误差均值 // %
LS – SVM 单输入单输出水位预测模型	80.6 (乙)	0.996 4 (甲)	94.80 (合格)	0.189 6	0.580 3
LS – SVM 双输入单输出水位预测模型	87.10 (甲)	0.998 5 (甲)	91.45 (合格)	0.182 9	0.571 0

根据表 3 结果,依据《水文情报预报规范》(SL250 – 2000)[7] 的相关规定可知:通过 LS – SVM 得到的 2 种预测模型预测的效果都比较好,均可以用于发布正式的洪水预报。进一步比较得知:基于 LS – SVM 构建的双输入单输出(水位和流量—水位)预测模型要比单输入单输出(水位—水位)预测模型的预测精度高。也就是说,输入的水位影响因子越多,得到的水位预测模型就会越精确。由于该研究只得到了水位和流量的数据,所以只考虑了水位和流量这 2 种影响因子,若日后得到其他的水位影响因子的数据,比如河道的断面、降雨量、泥沙量等,就可以得到更加精确的预测模型。

4 结论

该研究应用最小二乘支持向量机(LS – SVM)建立了河

道水位预测模型,通过实例分析表明:基于 LS－SVM 建立的河道水位预测模型是合理可行的,模型的精度等级都达到了乙级以上,可以用于发布正式的洪水预报,这就为农业生产中防洪方案的制定提供了非常有效的支持。由于最小二乘支持向量机需要确定的模型参数仅有 γ 和 σ^2 2 个,收敛速度快,预测精度高,因此,LS－SVM 的应用前景十分广阔。

参考文献

[1] ABBOTT M B,BATHURST J C,CUNGE J A,et al. An introduction to the European hydrological system-systeme hydrologique Europeen,"SHE",2: Structure of a physically-based,distributed modelling system[J]. Journal of Hydrology,1986,87(1):61－77.

[2] 雷晓云,张丽霞,梁新平. 基于 MATLAB 工具箱的 BP 神经网络年径流量预测模型研究——以塔城地区乌拉斯台河为例[J]. 水文,2008,28(1):43－46.

[3] 许剑华. 一般线性汇流模型实时预报方法的初步探讨[J]. 水利学报,1986(9):72－73.

[4] 孙德山,吴今培,侯振挺,等. 加权支持向量回归算法[J]. 计算机科学,2003,30(11):38－39.

[5] 段华. 支持向量机的增量学习算法研究[D]. 上海:上海交通大学,2008.

[6] 李方方,赵英凯,颜昕. 基于 Matlab 的最小二乘支持向量机的工具箱及其应用[J]. 计算机应用,2006,26(S2):358－360.

[7] 中华人民共和国水利部. SL250－2000 水文情报预报规范[S]. 北京:中国水利水电出版社,2000.

[8] 宋小杉,蒋晓瑜,罗建华,等. 基于类间距的径向基函数－支持向量机核参数评价方法分析[J]. 兵工学报,2012,33(2):203－208.

(上接第 349 页)

仅可以体现学校管理实验室的特色、能力和水平,也可以指导和影响创新型先进管理理念的形成和运行。学校本着培养一个完整的社会人的教育理念,强化以德育为基础的技术技能培养、知识分析和再学习能力、专业岗位培养和就业零距离培养。每个学校都有自己的办学理念,把这种办学的理念融入到管理制度中去,潜移默化地影响着实验室的建设和发展,规范师生的行为,形成一种连续性的学校特色。

2.2　规范制度　优秀的管理制度不仅是实验室良性运行的体现,也能从侧面体现实验室领导管理水平与特色。实验室的规范与制度分为两个范畴,一类是普适性的制度,如学生实验实训守则、实验室安全管理办法、实验实训室运行规定、校企合作守则等;另一类是具有专业和院校特色的制度,如行业工作守则、行业操守、专业流程、系部实验室共享办法、大型仪器设备校内外开放办法等。这些规范和制度的建立和设置是为了标准化实验室的管理运行,为社会培养更加规范、合格的全面发展的人才。制定规章制度要客观、可行、合理和科学,要以人为本,以学校人才培养理念为纲,且诸制度间不能产生自相矛盾的冲突。实施规章制度要加强宣传与普及,要创造多种实施条件,并且必须加大执行力度,严格执行奖惩措施,使规范制度的功效最大化。

3　精神文化

实验室的精神文化是最能体现实验室管理与运行质量的衡量标准之一,主要体现在实训室的管理风格、教学风格、学习风格、文化传统、精神追求等方面。

3.1　实训室管理风格　实训室管理风格的好坏与实训室运行效率及特色的发挥成正相关。管理领导的行事风格及形成的办事风格会有效影响下属的办事风格。清晰有效的指令能促进下属办事效率的提升,和睦融洽的上下属氛围是最佳的管理风格,能在诸多方面有效地解决问题,能缓解突发事件的不良影响,能提升办事效率和办事过程中的沟通效率。

3.2　实训教学风格　实训教学风格是实训教师在长期教学、科研实践活动中形成的鲜明特点、风格,是教师道德品质、知识水平、技能、业务修养等素质的综合反映。具有特色的教学风格能达到知识的有效传递,对于高职实训课程的开展有良好的促进作用,同时可以增加师生之间的情感与信任,可以形成良好的课堂氛围。实训教学风气与理论教学同出一脉,要关心青年教师的授业技能成长和加强青年教师的道德素质与修养的培养。在注重教师培养的同时,给予教师适当的个性发展空间,调动青年教师潜在的积极性和主动性。

3.3　实训学习风格　实训教师的学习风格是保持与时俱进的原动力,是知识更新过程中表现出来的态度和方法,反映了教师的行为习惯。实训科科长及系部分管主任需要有良好的感染力来提升学习风气,其行为可以感染员工,做到个人行为文化与群体精神文化的传递,故应以严谨的治学实验实训态度与以身作则的担当把自己分管的本职工作做到最好。良好的学风也是通过长期的学习、教育、培养、教学、科研等活动过程积累形成的良好行为习惯。

3.4　文化传统与精神追求　高职随着办学特色的逐渐凸显,日益强调专业的特色化。相关的实训文化最能体现出办学理念的特色性。对于涉农专业传承下来的专业特色传统,需要取其精华,发扬光大。尤其是具有物候农时特色的实训运行风格,能更加有效和适应专业实训的开展。实训体系的发展也应有远大的精神追求,不应安于现状,固步自封。高职院校的办学历史相对较短,在实训运行的相关方面仍需要向本科院校借鉴和学习,即追求的目标要高,要向本科级别的院校看齐。

4　结语

实训文化营造质量的高低直接影响着学校实训教学质量的水准。不断提高实训文化,不断打造具有涉农特色的实训文化,对于打造特色性发展高职院校具有促进作用。在营建的过程中,不仅为实训室的内涵充实了内容,也会成为高职校园文化的主导和中坚力量。

参考文献

[1] 韩振生. 高职院校实训室文化建设的思考[J]. 黑龙江农业工程职业学院学报,2009(1):35－36.

[2] 朱发仁. 高职院校校园文化的特色研究[D]. 长沙:中南大学,2007:62－63.

[3] 高峰. 高校实验室文化建设之我见[J]. 鸡西大学学报,2006(6):56－57.

[4] 金涛. 校内实训室文化建设的探析[J]. 中小企业管理与科技,2009(10):47－48.