

支持向量机函数拟合在地下水动态预测中的应用

张霞¹, 王亚萍¹, 李占斌^{2,3}, 于国强⁴

(1. 陕西省环境科学研究院, 陕西 西安 710061; 2. 西安理工大学 西北水资源与环境生态教育部重点实验室, 陕西 西安 710048; 3. 中国科学院 水利部 水土保持研究所 黄土高原土壤侵蚀与旱地农业国家重点实验室, 陕西 杨凌 712100; 4. 中国地质调查局 西安地质矿产研究所, 陕西 西安 710054)

摘要: 为了深入探求陕西省洛惠渠灌区地下水动态变化情况, 在综合分析了各种地下水动态研究方法的基础上, 提出了基于支持向量机的灌区地下水动态预测方法。用 MATLAB 编制了相应的支持向量机程序, 建立了地下水动态预测模型。以灌区多年实测数据为学习样本和测试样本, 讨论了基于支持向量机的地下水动态分析方法及其可行性。结果表明, 支持向量机预测模型具有较高的精度, 可以很好地描述地下水动态与各因子之间响应关系, 各因子之间相互作用, 以及其复杂的耦合关系。研究表明, 这种方法运用到灌区地下水动态预测中是切实可行的。

关键词: 地下水动态; 洛惠渠灌区; 支持向量机; 函数拟合

文献标识码: B

文章编号: 1000-288X(2012)05-0277-04

中图分类号: S153.1

Application of Support Vector Machines Function Fitting in Groundwater Dynamic Prediction

ZHANG Xia¹, WANG Ya-ping¹, LI Zhan-bin^{2,3}, YU Guo-qiang⁴

(1. Shaanxi Provincial Academy of Environmental Science, Xi'an, Shaanxi 710061, China;

2. Key Laboratory of Northwest Water Resources and Environment Ecology of the MOE, Xi'an University of Technology, Xi'an, Shaanxi 710048, China; 3. State Key Laboratory of Soil Erosion and Dryland Farming on the

Loess Plateau, Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, China; 4. Xi'an Institute of Geology and Mineral Resources, CGS, Xi'an, Shaanxi 710054, China)

Abstract: To investigate groundwater dynamic in the Luohuiqu irrigation district of Shaanxi Province, different research methods for groundwater dynamic are comprehensively analyzed. The method of support vector machines(SVM) is proposed for predicting groundwater dynamic and building predicting model and two computer programs are compiled using MATLAB. Practical effectiveness of the theory of the SVM for groundwater dynamic prediction is discussed based on the irrigation data measured for many years. Results show that the SVM regression model can well express the complicated coupling relationship of groundwater dynamics. Therefore, the application of this method to the prediction in the irrigation district is feasible and practical.

Keywords: groundwater dynamic; Luohuiqu irrigation district; support vector machines; function fitting

我国西北内陆灌区由于水资源开发和管理中还存在不完善的方面, 出现了不良的生态环境问题, 如土壤次生盐渍化、土地沙漠化、地下水水质恶化、地面沉降等问题, 水资源贫乏和生态环境脆弱已成为制约经济发展的主要因素之一, 而地下水动态模拟是干旱内陆灌区水资源优化管理与调控的基础^[1-4]。对大型灌

区而言, 传统的地下水动态研究方法(解析法、数值法、物理模拟法)因水文地质因参数率定、边界条件等问题而难以实现。当以灌域为基本单位, 自然—人工—生物条件下的地下水系统时空变异性极为显著, 其影响因素繁多, 因果关系存在很强的非线性, 因而难以把握^[5-8]。

收稿日期: 2011-09-11

修回日期: 2011-12-20

资助项目: 中国博士后基金“白龙江流域暴雨泥石流运动机理及数值模型研究”(2011M501445); 陕西省自然科学基金基础研究计划项目“甘肃舟曲区域暴雨泥石流运动机理及数值模型研究”(2012JQ5001); 水利部公益性行业科研专项“小流域淤地坝坝系防洪风险评价技术项目”(201201084); 国家重点基础研究发展计划(973 计划)“黄土坡面土壤养分流失特征与控制机理”(2011CB411903)

作者简介: 张霞(1982—), 女(汉族), 内蒙古自治区包头市人, 硕士, 工程师, 从事生态环境研究工作。E-mail: zhangxia20002000@163.com。

通信作者: 李占斌(1962—), 男(汉族), 河南省南阳市人, 工学博士, 教授, 主要从事土壤侵蚀与水土保持方面的研究。E-mail: zhanbinli@126.com。

支持向量机作为一种新型机器学习方法,主要用于模式识别和非线性函数拟合。支持向量机的早期应用是模式识别问题,已有较成功的应用基础,例如手写汉字识别、面部识别等;与模式识别相比,支持向量机函数拟合的应用较少。支持向量机和神经网络相比,支持向量机是基于统计学习理论的小样本学习方法,采用结构风险最小化原则,具有很好的泛化性能,而神经网络是基于大样本的学习方法,采用经验风险最小化原则^[9-15]。本文采用支持向量机函数拟合方法来预测地下水动态,其二次优化问题采用 MATLAB 中的优化工具箱,并在 MATLAB 中编制了相应的程序。利用编制的支持向量机程序,通过对 24 个训练样本的学习和 12 个测试样本的预测,分析了自然—人工—生物复合因素对灌区地下水动态的影响,讨论了运用支持向量机进行地下水动态预测的可行性和有效性,为改善灌区地表生态环境提供了更精确的科学依据。

1 材料与方法

1.1 研究区概况

洛惠渠灌区位于陕西关中平原东部的渭洛河阶地,属大荔、蒲城、澄城 3 县管辖,灌溉面积 320 km²,海拔高度 329~533.5 m。灌区因地层北升南降,河流迁移,曲化、淘深,以及降水冲蚀,形成了塬梁的沟壑和河谷阶地堆积地貌。灌区属温带大陆性半干旱气候区,年平均气温 13.5 °C,年平均降雨量 480 mm,年平均蒸发量约 1 690 mm。灌区地下水补给来源主要是降雨和引水灌溉,其次是塬区潜水和微承压水,地面 40~50 m 以下的古湖沉积物中富集了大量的可溶盐类,在水循环作用下,古湖沉积物中富集的可溶盐上升,致使潜水矿化度偏高,不能引灌^[16-18]。

1.2 支持向量机的基本理论及算法

支持向量机的基本思想是通过用内积函数定义的非线性变换将输入空间变换到一个高维空间,在高维空间中寻找输入变量和输出变量之间的一种非线性关系。支持向量机采用结构风险最小化原则,具有很好的推广能力,其算法是一个凸二次优化问题,保证找到的解是全局最优解;能较好地解决小样本、非线性、高维数等实际问题。因此,支持向量机是当今研究的热点问题^[10-11]。

支持向量机最初是为分类问题而设计的,而支持向量机用在回归分析时,也同样具有很好的性能。支持向量机回归有线性回归和非线性回归,对于线性回归,考虑用线性回归函数 $f(x) = wx + b$ 拟合数据 $\{x_i, y_i\}, i=1, 2, \dots, n, x_i \in R^n, y_i \in R$, 时,为保证回

归函数平坦,须找到尽可能小的 w ,为此,采用最小化欧几里德空间泛数,假设所有的训练数据在精度 ϵ 下无误差地用线性函数拟合,可以解下面的优化问题,即:

$$\min \varphi(w) = \frac{1}{2} \|w\|^2 \quad (1)$$

满足约束条件:

$$\begin{cases} wx_i + b - y_i \leq \epsilon \\ y_i - wx_i - b \leq \epsilon \end{cases} \quad (i=1, 2, \dots, k) \quad (2)$$

在不能完全满足上述约束条件时,可以引入松弛变量 ξ_i 和 ξ_i^* ,优化问题可以转化为:

$$\min \varphi(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k (\xi_i + \xi_i^*) \quad (3)$$

满足约束条件:

$$\begin{cases} y_i - wx_i - b \leq \epsilon + \xi_i \\ wx_i + b - y_i \leq \epsilon + \xi_i^* \end{cases} \quad (i=1, 2, \dots, k) \quad (4)$$

优化函数 $\varphi(w)$ 为二次型,约束条件是线性的,因此是典型二次规划,可用拉格朗日乘子法求解。对于待分析的数据集, $(x_i, y_i), i=1, 2, \dots, n, x_i \in R^n$ 为输入, $y_i \in R$ 为输出,则支持向量机对偶形式为:

$$W(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^k (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) + \sum_{i=1}^k y_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^k (\alpha_i + \alpha_i^*) \quad (5)$$

优化目标是最小化 $\frac{1}{2} \|w\|^2 + C \sum_{i=1}^k (\xi_i + \xi_i^*)$, 其中,常数 $C > 0$, 表示对超出误差 ξ 的样本的惩罚程度。采用优化方法可以得到其对偶问题。约束条件为:

$$\begin{cases} \sum_{i=1}^k (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases} \quad (i=1, 2, \dots, k) \quad (6)$$

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

目前常用的核函数主要有:

(1) 多项式核函数:

$$K(x, y) = (xy + 1)^d \quad (d=1, 2, \dots, n) \quad (8)$$

(2) 径向基函数核函数:

$$K(x, y) = \exp(-\gamma |x - y|^2) \quad (9)$$

(3) Sigmoid 核函数:

$$K(x, y) = th[\varphi(xy) + \theta] \quad (10)$$

关于式该方法所对应的二次优化问题,目前已有多种方法求解,在支持向量机问题中常用的方法有内点算法^[13]、SMO 方法^[14]、分解方法等。

2 结果与讨论

2.1 地下水动态影响因素的确定

灌区地下水动态受多种因素的影响,针对灌区实

际情况,综合考虑各方面因素,选取降雨量、蒸发量、灌溉面积、灌溉引水量、地下水消耗量^[19]和作物水分盈亏量作为影响因素,综合体现了自然因素、人类活动因素以及生物因素对地下水动态的影响^[17-19]。

2.2 地下水动态模型的建立

假设有地下水动态实例 $(x_i, y_i) (i=1, 2, \dots, k)$, $x_i \in R^n$ 为影响灌区地下水动态的因素, y_i 为影响地下水动态的指标, $y_i \in R$ 表示地下水埋深。支持向量机地下水动态模型的建立,就是寻找 x_i, y_i 之间的关系:

$$f: R^n \rightarrow R, \quad y_i = f(x_i) \quad (i=1, 2, \dots, k)$$

式中: R^n ——影响地下水动态的 6 个因素,即降雨量、蒸发量、灌溉面积、灌溉引水量、地下水消耗量和作物水分盈亏量; R ——地下水动态的指标,即地下水埋深。

根据支持向量机理论,灌区地下水动态预测模型的建立,也即寻求如下的表达式成立:

$$f(x) = \sum_{i=1}^k (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (11)$$

式中: x ——要预测的地下水动态的影响因素; x_i —— k 个样本中第 i 个样本; $K(x, x_i)$ ——核函数。本文利用 MATLAB 中的优化工具箱解式(5),(6)组成的优化问题,并在 MATLAB 中编制了相应的支持向量机程序。

2.3 学习样本和测试样本的选择

本文所选样本数据来源于覆盖全灌区 78 个观测井 1971—2006 年观测资料,所有相关样本均参与网络的训练、检验。随机选取其中 24 组样本作为年地下水动态的预测模型学习样本,12 组样本用于测试建立的模型。

2.4 模型参数的选择及预测结果

学习样本集确定后,地下水动态预测模型的建立,主要是选择相应的支持向量机参数:核函数和惩罚系数 C ,它们对预测结果的影响很大,它们的合理确定直接影响模型的精度和推广能力。本文通过对各种核函数的测试,最终确定:地下水动态测模型中,预测模型的核函数为径向基函数且 $\gamma=2$;通过对大量 C 值的测试,确定 $C=1$;然后解得 α_i, α_i^*, b ,即可得到支持向量机地下水动态预测模型。利用上面建立的模型对 12 个样本进行预测,预测结果见表 1 和图 1,由预测结果可以看出,最大绝对误差小于 0.88 m,最大相对误差小于 10%。表明上述建立的预测模型是合理和可靠的。

2.5 模型参数对预测结果的影响及选取

在模型的建立过程中,考虑到支持向量机参数对预测结果的影响,对大量的参数进行了测试,最终确定了本文中的参数,图 2—3 表示在地下水动态预测

模型中 γ 和 C 对最大相对误差的影响。由于人工搜索支持向量机参数存在很大的盲目性,不能保证找到的参数是最优参数,因此,本研究将遗传算法和支持向量机结合,用遗传算法来搜索支持向量机参数。

表 1 支持向量机预测结果与实际值对比

样本序号	真实值/ m	预测值/ m	绝对误差/ m	相对误差/ %
1	5.84	6.28	-0.440 0	-7.53
2	6.05	6.36	-0.310 7	-5.13
3	6.29	6.20	0.091 2	1.45
4	6.30	6.21	0.091 2	1.45
5	7.03	7.39	-0.360 7	-5.13
6	7.86	8.48	-0.620 0	-7.89
7	8.53	8.12	0.410 0	4.81
8	10.13	9.25	0.880 0	8.69
9	12.21	11.50	0.710 0	5.81
10	8.48	8.80	-0.320 0	-3.77
11	9.49	9.42	0.065 4	0.69
12	5.67	6.21	-0.540 0	-9.52

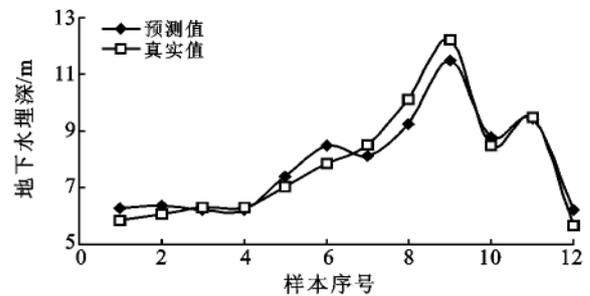


图 1 地下水埋深支持向量机预测结果

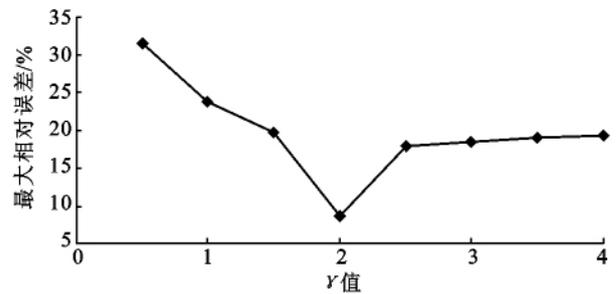


图 2 γ 值对预测结果的影响 ($C=1$)

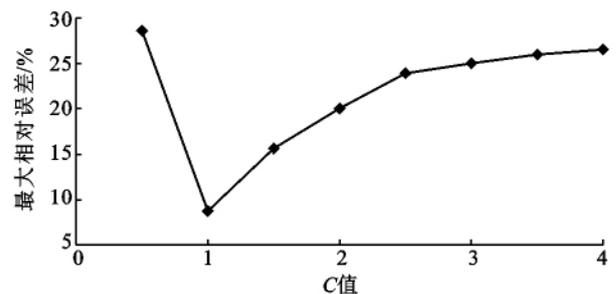


图 3 C 值对预测结果的影响 ($\gamma=2$)

3 结论

灌区地下水动态受很多因素的影响,很难在地下水埋深和这些因素之间建立一种确定的数学模型。一方面,这种关系是一种非常复杂的非线性关系,另一方面,地下水动态还与具体的工程措施有很大的关系。因此,充分利用以前的灌区特有实例,通过对实际工况分析,寻找地下水动态与各种因素之间的关系是非常重要的。支持向量机方法作为一种优秀的学习算法,具有很好的推广能力和较强的非线性动态数据处理能力。

本研究以大量灌区实例为学习预测的基础,建立支持向量机地下水动态预测模型,该模型具有较高的精度,能有效表征复合条件下地下水动态的变化情况以及地下水动态与各影响因子之间的响应关系;对灌区地下水动态分析具有很好的适用性,为解决灌区地下水分析开辟了一条新的途径,为改善灌区地表生态环境提供了更精确的科学依据。

[参 考 文 献]

- [1] 张祥伟,竹内邦良. 大区域地下水模拟的理论与方法[J]. 水利学报,2004,35(6):7-13.
- [2] 霍再林. 基于 ANN 与 FEFLOW 的民勤绿洲地下水动态模拟[D]. 北京:中国农业大学,2007.
- [3] 赵成义,王玉朝,李保国. 内陆河流域植被变化与地下水运动的耦合关系[J]. 水利学报,2003(12):59-65.
- [4] 刘国东,丁晶. BP 网络用于水文预测的几个问题探讨[J]. 水利学报,1999,30(1):65-70.
- [5] Joannis N D, Pauline C, Joannis K T. Groundwater level forecasting using artificial neural networks[J]. Journal of Hydrology, 2005,309(1/4):229-240.
- [6] Emery C J, Mary M P, Emmanuel C. Application of artificial neural networks to complex groundwater management problems[J]. Natural Resources Research, 2003,12(4):303-320.
- [7] Emery C J, Anthony J R, Mary M P. A neural network model for predicting aquifer water level elevations[J]. Ground Water,2005,43(2):231-241.
- [8] Neuman S P, Jacobson E A. Analysis of nonintrinsic spatial variability by residual kriging with application to regional ground water levels[J]. Math. Geol., 1984, 16(5):499-521.
- [9] 赵洪波,冯夏庭. 支持向量机函数拟合在边坡稳定性估计中的应用[J]. 岩石力学与工程学报,2003,22(2):241-245.
- [10] Platt J C. Sequential minimal optimization: A fast algorithm for training support vector machines[R]. Technical Report MSR-TR-98-14,1998.
- [11] Vladimir N V. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社,2000.
- [12] Toy F E H, Cao L J. Modified support vector machines in financial time series forecasting[J]. Neuro computing, 2002,48(1/4):847-861.
- [13] Flake G W, Lawrence S. Efficient SVM regression training with SMO[J]. Machine Learning, 2002, 41(1): 271-290.
- [14] Shevade S K, Keerthi S S, Bhattacharyya C. Improvements to SMO algorithm for SVM regression[J]. IEEE Transactions on Neural Networks, 2000,11(5):1188-1194.
- [15] Burge C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998,2(2):121-167.
- [16] 张霞,李占斌,李鹏. 洛惠渠灌区地下水位动态变化规律研究[J]. 西北农林科技大学学报:自然科学版,2007,35(8):223-226.
- [17] 于国强,李占斌,张霞,等. 土壤水盐动态的 BP 神经网络模型及灰色关联分析[J]. 农业工程学报,2009,25(11):74-79.
- [18] 于国强,李占斌,张霞,等. 洛惠渠灌区地下水位动态模拟研究[J]. 西北农林科技大学学报:自然科学版,2009,37(12):223-228.
- [19] 陈克强. 北方地区大面积作物生产对地下水的影响:以洛惠渠灌区为例[D]. 西安:西安理工大学,2005.