文章编号:0253-9993(2011)12-2097-05

基于遗传-支持向量回归的煤层底板突水量预测研究

曹庆奎[1],赵　斐[1,2]

(1. 河北工程大学 经济管理学院,河北 邯郸　056038;2. 北京科技大学 东凌经济管理学院,北京　100083)

摘　要:针对煤层底板突水问题的小样本、非线性特点,采用支持向量回归算法对突水量进行预测,避免了定性分析的局限性。利用遗传算法全局搜索能力的优势,提出了基于遗传算法的支持向量回归参数寻优方法,并建立煤层底板突水量预测的遗传-支持向量回归模型。该模型首先通过遗传算法对训练样本的学习,得到支持向量回归机的最优参数值,然后运用遗传-支持向量回归模型对测试样本进行突水量预测。测试结果表明:与神经网络,传统支持向量回归机的预测值相比,煤层底板突水量预测的遗传-支持向量回归模型精度高,具有较强的泛化能力。

关键词:煤层底板;突水量预测;遗传算法;支持向量机;支持向量回归机

中图分类号:TD742.1　　**文献标志码**:A

Forecast of water inrush quantity from coal floor based on genetic algorithm-support vector regression

CAO Qing-kui[1],ZHAO Fei[1,2]

(1. *School of Economics and Management, Hebei University of Engineering, Handan*　056038, *China*; 2. *Dongling School of Economics and Management, University of Science and Technology Beijing, Beijing*　100083, *China*)

Abstract: The problem of water inrush from coal floor was characterized by small samples, nonlinear, and using support vector regression algorithm avoided the limitations of qualitative analysis to predict the water inrush quantity. Support vector regression parameters optimization method was proposed based on genetic algorithm using the advantages of the global search capability of the genetic algorithm, and established genetic algorithm-support vector regression model of water inrush quantity prediction from coal floor. First, the model got the optimal support vector regression parameters by genetic algorithm to learn the training samples, and then used genetic algorithm-support vector regression model to predict the water inrush quantity of test samples. The test results show that, compared with the predictive values of neural network and the traditional support vector regression, the genetic algorithm-support vector regression model has higher prediction accuracy and good generalization ability.

Key words: coal floor; water inrush quantity prediction; genetic algorithm; support vector machine; support vector regression

近些年来随着煤矿开采深度的增加,我国许多煤田水文地质条件越来越复杂,煤层底板带压开采突水的危险性不断加大,煤矿水害事故时有发生。因此,对煤层底板突水量进行预测对加强煤矿的防治水工作、防止和减少突水事故、保障煤矿职工生命安全具有重要意义。现已形成了多种理论和方法:国家安全生产监督管理总局在《煤矿防治水规定》中所规定的突水系数法;陈红江等[1]和付玉华等[2]将距离判别分析理论用于煤层底板突水量预测中,并证明该方法的可行性;王连国等[3]和王凯等[4]将煤层底板突水看作是由底板岩层失稳而发生的不连续的突变现象,通过建立煤层底板突水的尖点突变模型探求煤层底

收稿日期:2011-01-24　　**责任编辑**:韩晋平
基金项目:国家自然科学基金资助项目(60874116);河北省自然科学基金资助项目(F2010001047)
作者简介:曹庆奎(1963—),男,河北乐亭人,教授。E-mail:Caoqk@126.com

板突水机制;邱秀梅等[5]和冯利军等[6]建立的人工神经网络底板突水预测模型,利用神经网络输入和输出间的非线性映射能力、自学习和自适应功能,实现了矿井的突水预测;闫志刚等[7]和姜谙男等[8]针对煤层底板突水的非线性和不确定性,采用支持向量机模型进行预测,得到样本点的突水等级;石秀伟等[9]、张和生等[10]和武强等[11]利用 GIS 强大的空间数据统计处理功能,采用多因素地学信息空间复合叠置方法确定煤层底板突水灾害模式,构建煤层底板突水危险性分区模式图。

目前对煤层底板突水预测分析大多局限于定性研究,最终得到煤层底板突水的等级划分,而对实际突水量预测的定量研究却很少。基于统计学习理论的支持向量机实现了经验风险和置信范围的最小化,从而在样本数量较少的情况下获得良好的统计规律和泛化能力;且已被广泛地用于数据分类、回归预测等方面,能较好地解决煤层底板突水中的小样本、非线性难题。可以避免神经网络的局部极小值问题、过学习以及过分依赖经验等缺陷。将遗传算法和支持向量机中的回归算法相结合,利用遗传算法对支持向量回归参数进行优化,建立煤层底板突水量预测模型,并以实例验证其有效性。

1 支持向量回归机

支持向量机(Support Vector Machine,SVM)最初是针对模式识别问题,由 Vapnik 等人于 20 世纪 90 年代提出的一种新型机器学习方法,主要用于对数据分类问题的处理[12-14]。SVM 基于 VC 维理论和结构风险最小原理,根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折中,以期获得最好的推广能力。SVM 理论分为支持向量分类(Support Vector Classification,SVC)和支持向量回归(Support Vector Regression,SVR),其中 SVR 算法用于时间序列的预测、非线性建模与预测、优化控制等方面。

训练样本集 $\{(x_i,y_i)\mid i=1,2,\cdots,l\}$,其中 $x_i\in X=R^n$ 为输入向量,$y_i\in Y=R$ 为输出向量。对于非线性支持向量回归,通过核函数 $k(x_i,x_j)=\varphi(x_i)\times\varphi(x_j)$ 将输入样本空间非线性映射到高维特征空间进行线性回归,非线性回归估计函数为

$$f(x)=\omega\varphi(x)+b \tag{1}$$

式中,ω 为权值向量;φ 为非线性映射函数;b 为阈值。

为了在支持向量回归机中保持较好的稀疏性,引入不敏感损失函数 ε 最小化经验风险,得到损失函数为

$$c(x,y,f)=|y-f(x)|_\varepsilon=\max\{0,|y-f(x)|-\varepsilon\} \tag{2}$$

即如果点 x 的观察值 y 和回归预测函数值 $f(x)$ 之间的差别小于 ε 时,则损失为 0。

支持向量回归机实际上是求解 ω 和 b,在满足不敏感损失函数 ε 前提下最小化 $\frac{1}{2}\|w\|^2$。引入松弛变量 ξ_i、ξ_i^*,SVR 转化为以下目标函数的最小化问题,即

$$\min\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)$$
$$\text{s.t.}\begin{cases}y_i-\omega\varphi(x_i)-b\leqslant\varepsilon+\xi_i\\ \omega\varphi(x_i)+b-y_i\leqslant\varepsilon+\xi_i^*\\ \xi_i,\xi_i^*\geqslant 0,i=1,2,\cdots,l\end{cases} \tag{3}$$

对式(3)进行拉格朗日变换得到其对偶问题,即

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

其中,a_i 和 a_i^* 为拉格朗日算子;C 为惩罚参数。$(a_i-a_i^*)\neq 0$ 的训练样本为支持向量,求解得到回归函数为

$$f(x)=\sum_{i=1}^{N}(a_i-a_i^*)K(x_i,x_j)+b \tag{5}$$

式中,N 为支持向量个数,核函数采用径向基函数 $K(x_i,x_j)=\exp\{-|x_i-x_j|^2/(2\sigma^2)\}$。

2 遗传-支持向量回归模型

SVR 中包含 3 个参数:惩罚参数 C、核参数 σ、不敏感损失函数 ε,目前它们往往靠经验或测试给定,还没有好的方法。遗传算法(Genetic Algorithm,GA)模拟生物进化过程中的自然选择和遗传变异,具有很强的全局搜索能力,利用这一优点,对支持向量回归机中的参数进行优化,构造遗传-支持向量回归模型(GA-SVR),为 SVR 参数选取问题的解决提供了一种有效方法[15]。GA-SVR 模型的步骤,如图 1 所示。

(1)参数编码。SVR 参数的寻优过程是一个复杂的连续参数优化问题,采用实数编码方式避免了二进制编码操作时解码、编码的操作;克服了二进制字符串有限长度问题,从而提高遗传算法的性能和求解

精度。

(2)适应度函数定义为训练数据上的交叉验证后的均方误差平均值,即

$$F = \frac{1}{l}\sum_{i=1}^{l}(y_i - f_i)^2 \tag{6}$$

式中,l 为训练集的样本数;y_i 为实际值;f_i 为预测值。

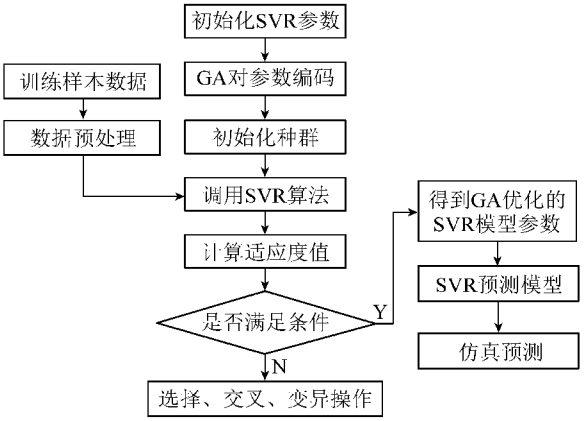

图 1　GA-SVR 流程

Fig. 1　GA-SVR process

(3)遗传操作。

① 选择操作。基于排序的适应度分配原则。将种群内的个体按照适应度值进行排序,按照式(7)确定各个个体被选择的概率 P_i。

$$P_i = r(1-r)^{i-1} \tag{7}$$

其中,i 为个体排序序号;r 为排序第 1 的个体的选择概率。选择概率仅仅取决于个体在种群中的序位,而不是实际的适应度值。

② 交叉操作。线性组合的交叉操作。以某一概率 a(0 和 1 之间的随机数)对某两个染色体 x_1 和 x_2 进行交叉操作,即

$$\begin{cases} x_1 = ax_1 + (1-a)x_2 \\ x_2 = (1-a)x_1 + ax_2 \end{cases} \tag{8}$$

③ 变异操作。均匀变异,在将变异的染色体中随机选择一个变异位 j 设置为一个归一化的随机数 $U(a_i, b_i)$,a_i 和 b_i 为对应该变异位的上下限,即

$$x_j = \begin{cases} U(a_i, b_i) & (i = j) \\ x_i & (i \neq j) \end{cases} \tag{9}$$

(4)终止规则:采用最大进化代数法,即迭代计算到指定代数时,算法终止计算,并返回当前最优解。

3　GA-SVR 模型的实例应用

3.1　样本数据的获取与处理

在文献[5]中选取 16 个典型的煤层底板突水资料,其中前 13 个作为训练样本,另外 3 个为测试样本。由于各个预测指标的量纲不同及各指标值的差异,在用 GA-SVR 进行煤层底板突水量预测时,需将原始数据进行标准化处理,从而消除量纲对样本数据可比性的影响。此处,采用比例转换法,对于正向指标,转换公式为 $x' = (x - x_{\min})/(x_{\max} - x_{\min})$;对于逆向指标,转换公式为 $x' = (x_{\max} - x)/(x_{\max} - x_{\min})$。标准化后的数据,见表 1。

表 1　原始数据标准化

Table 1　Standardization of the raw data

序号	工作面名称	水压	含水层	隔水层厚度	底板采动破坏深度	断层落差	突水量
1	淮南谢一矿 33 采区	0.305 01	0	0.612 36	0.411 76	0.100 00	0.198 62
2	焦作九里山矿 12031	0.261 44	0	0.731 90	0.367 65	0	0.335 73
3	肥城陶阳矿 9901	0	0	0.834 36	0.095 59	0.533 33	0.198 10
4	肥城大封矿 9204	0.104 58	0	0.842 90	0.676 47	0.213 33	0.337 78
5	肥城陶阳矿 9906	0.178 65	0	0.685 79	0.580 88	0	0.028 19
6	淄博夏庄矿 1007	1.000 00	1	0.170 08	0.713 24	0.466 67	0.947 21
7	焦作王封矿 1441	0.108 93	0	0.783 13	0.088 24	1.000 00	0.704 77
8	峰峰二矿 2682	0.501 09	0	0.441 60	1.000 00	0	0.142 23
9	新汶协庄矿 31104	0.152 51	0	0.612 36	0.808 82	0.326 67	0.422 86
10	淄博龙泉矿 149	0.753 81	1	0	0.639 71	0.666 67	0.308 05
11	肥城陶阳矿 9903	0.054 47	0	0.730 19	0.485 29	0.026 67	0
12	淮北杨庄矿Ⅱ617	0.546 84	0	0.368 17	0.522 06	0.233 33	0.728 60
13	峰峰一矿 1532	0.370 37	1	1.000 00	0	0	1.000 00
14	峰峰二矿 2671	0.479 30	0	0.441 60	0.566 18	0.400 00	0.256 28
15	肥城查庄矿 7505	0.089 32	0	0.817 28	0.323 53	0	0.070 94
16	焦作韩王矿 2131	0.108 93	0	0.851 43	0.051 47	0	0.151 20

3.2 GA-SVR 预测模型参数的确定

SVR 从训练样本中获取知识,利用遗传算法选择最优的惩罚参数 C、核参数 σ、不敏感损失函数 ε,建立煤层底板突水量预测的 GA-SVR 模型。在 MATLAB 中实现 GA-SVR 预测模型参数的优化,其中,支持向量回归机采用 3-e-SVR,核函数采用径向基核函数。遗传算法进化代数为 200,种群规模 20,交配概率和变异概率分别为 0.4 和 0.01,以式(6)作为适应度函数。惩罚参数 C 的取值范围为[0,100],径向基核函数参数 σ 的取值范围为[0,1 000],不敏感损失函数 ε 的取值范围为[0,1]。

将表 1 中训练样本的水压、含水层、隔水层厚度、底板采动破坏深度、断层落差指标标准化后的数据作为训练函数的输入向量,突水量为目标向量。模型训练结束后,得到 GA-SVR 预测模型的最优参数:惩罚参数 $C=30.936$、径向基核函数的参数 $\sigma=12.021$ 和不敏感损失函数 $\varepsilon=0.115$。利用该模型对训练样本进行预测,预测值和原始突水量的拟合曲线,如图 2 所示。

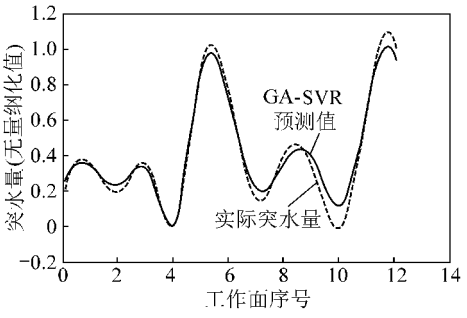

图 2 预测值对训练目标向量的拟合曲线

Fig. 2 The fitting curves of predictive values to training target vector

3.3 测试样本的突水量预测

利用训练所得到的最优参数建立煤层底板突水量预测的 GA-SVR 模型,通过 svmpredit 函数对测试样本进行预测,得出预测值,并对其反归一化。将 GA-SVR 的预测值分别与神经网络、传统支持向量回归机的预测值进行比较,其结果见表 2。

表 2 不同模型的煤层底板突水量预测值比较

Table 2 Comparison of predictive values of water inrush from coal floor based on different models

序号	实际突水量/($m^3 \cdot h^{-1}$)	神经网络模型		传统 SVR 模型		GA-SVR 模型	
		预测值/($m^3 \cdot h^{-1}$)	预测误差/%	预测值/($m^3 \cdot h^{-1}$)	预测误差/%	预测值/($m^3 \cdot h^{-1}$)	预测误差/%
14	1 310.0	1 259.1	3.90	1 265.3	3.41	1 285.2	1.89
15	586.8	513.4	12.50	530.1	9.66	571.6	2.59
16	900.0	644.5	28.40	798.7	11.26	873.4	2.96

4 结 论

(1)煤层底板突水量与其影响因素之间存在着非线性映射关系,支持向量回归机是建立在统计学习理论基础上的模式识别方法,对于这种非线性映射问题可以通过核函数将其映射到高维特征空间,并在高维特征空间对这种非线性关系进行线性回归,确定突水量。

(2)支持向量回归机参数的选择对突水量预测的准确性存在影响,利用遗传算法的全局寻优能力确定支持向量回归机的最优参数,利用建立的 GA-SVR 预测模型对测试样本的突水量预测,结果表明 GA-SVR 模型比传统的 SVR 模型预测精度高。

(3)由于神经网络是一种基于无限样本的渐近理论,只有当样本数量趋于无穷大时理论上才收敛于全局最优值;而支持向量回归机是建立在结构风险最小化原则之上的小样本学习方法,更满足对煤层底板突水量的预测。测试结果也验证了 SVR 模型和 GA-SVR 模型的预测误差更小,且 GA-SVR 模型的预测效果最好,这为煤层底板突水量预测提供了一个可行的方法。

参考文献:

[1] 陈红江,李夕兵,刘爱华,等. 煤层底板突水量的距离判别分析预测方法[J]. 煤炭学报,2009,34(4):487-491.
Chen Hongjiang,Li Xibing,Liu Aihua,et al. Forecast method of water inrush quantity from coal floor based on distance discriminant analysis[J]. Journal of China Coal Society,2009,34(4):487-491.

[2] 付玉华,董陇军. 基于 Fisher 判别的煤层底板突水量预测研究[J]. 矿业研究与开发,2009,29(3):70-72.
Fu Yuhua,Dong Longjun. Study on forecast of water bursting volume from coal seam floor based on Fisher discriminant[J]. Mining Research and Development,2009,29(3):70-72.

[3] 王连国,宋 扬,缪协兴. 基于尖点突变模型的煤层底板突水预测研究[J]. 岩石力学与工程学报,2003,22(4):573-577.
Wang Lianguo,Song Yang,Miao Xiexing. Study on prediction of water-inrush from coal floor based on cusp catastrophic model[J]. Chinese Journal of Rock Mechanics and Engineering,2003,22(4):573-577.

[4] 王 凯,位爱竹,陈彦飞,等. 煤层底板突水的突变理论预测方

法及其应用[J]. 中国安全科学学报,2004,14(1):11-14.
Wang Kai,Wei Aizhu,Chen Yanfei,et al. Predicting method and its application of water inrush from coal floor based on catastrophe theory[J]. China Safety Science Journal,2004,14(1):11-14.

[5] 邱秀梅,王连国. 煤层底板突水人工神经网络预测[J]. 山东农业大学学报,2002,33(1):62-65.
Qiu Xiumei,Wang Lianguo. ANN forecast for water-inrush from coal floor[J]. Journal of Shandong Agricultural University,2002,33(1):62-65.

[6] 冯利军,郭晓山. 人工神经网络在矿井突水预报中的应用[J]. 西安科技学院学报,2003,23(4):369-372.
Feng Lijun,Guo Xiaoshan. The application of artificial neural network theory to mine water inrush prediction[J]. Journal of Xi'an University of Science and Technology,2003,23(4):369-372.

[7] 闫志刚,白海波,张海荣. 一种新型的矿井突水分析与预测的支持向量机模型[J]. 中国安全科学学报,2008,18(7):166-170.
Yan Zhigang,Bai Haibo,Zhang Hairong. A novel SVM model for the analysis and prediction of water inrush from coal mine[J]. China Safety Science Journal,2008,18(7):166- 170.

[8] 姜谙男,梁　冰. 基于最小二乘支持向量机的煤层底板突水量预测[J]. 煤炭学报,2005,30(5):613- 617.
Jiang Annan,Liang Bing. Forecast of water inrush from coal floor based on least square support vector machine[J]. Journal of China Coal Society,2005,30(5):613-617.

[9] 石秀伟,胡耀青,张和生. 基于 GIS 的煤层底板突水预测理论模型[J]. 太原理工大学学报,2008,39(S2):244-247.
Shi Xiuwei, Hu Yaoqing, Zhang Hesheng. GIS-based forecasting model of floor water bursting in coal mines[J]. Journal of Taiyuan University of Technology,2008,39(S2):244-247.

[10] 张和生,薛光武,石秀伟,等. 基于地学信息复合叠置分析对煤层底板突水的预测[J]. 煤炭学报,2009,34(8):1 100-1 104.
Zhang Hesheng,Xue Guangwu,Shi Xiuwei,et al. Prediction of water inrush from coal seam floor confined based on geo-information composite overlay analysis[J]. Journal of China Coal Society,2009,34(8):1 100-1 104.

[11] 武　强,庞　炜,戴迎春,等. 煤层底板突水脆弱性评价的 GIS 与 ANN 耦合技术[J]. 煤炭学报,2006,31(3):314-319.
Wu Qiang,Pang Wei,Dai Yingchun,et al. Vulnerability forecasting model based on coupling technique of GIS and ANN in floor ground water bursting[J]. Journal of China Coal Society,2006,31(3):314-319.

[12] 邓乃扬,田英杰. 支持向量机:理论、算法与拓展[M]. 北京:科学出版社,2009:75-123.
Deng Naiyang,Tian Yingjie. Support vector machines:theory,algorithms and development[M]. Beijing: Science Press,2009:75-123.

[13] 曹庆奎,张方明. 基于支持向量机的供应链合作伙伴评价[J]. 河北工程大学学报,2009,26(1):93-97.
Cao Qingkui,Zhang Fangming. Evaluation of supply chain partners based on support vector machines[J]. Journal of Hebei University of Engineering,2009,26(1):93-97.

[14] Vapnik V,ladimir N. The nature of statistical learning theory[M]. NewYork:Springer2-Verlag,Inc,2000.

[15] 刘　胜,李妍妍. 自适应 GA-SVM 参数选择算法研究[J]. 哈尔滨工程大学学报,2007,28(4):398-402.
Liu Sheng, Li Yanyan. Parameter selection algorithm for support vector machines base on adaptive genetic algorithm[J]. Journal of Harbin Engineering University,2007,28(4):398-402.

2010 年《煤炭学报》评价指标

根据《2011 年版中国科技期刊引证报告(核心版)》最新数据统计,2010 年《煤炭学报》的评价指标如下:

总被引频次(学科排名)	影响因子(学科排名)	即年指标	他引率	学科影响指标	来源文献量	文献选出率	参考文献量	平均引文数	地区数	机构分布数	基金论文比	引用半衰期
2603(1)	1.109(1)	0.215	0.60	0.68	409	0.85	6 639	16.23	23	116	0.83	6.20