

基于支持向量机的 GPS 似大地水准面拟合

吴兆福^{1,2}, 官 鹏³, 高 飞², 王 依²

(1. 南京大学 城市与资源学系, 江苏 南京 210093; 2. 合肥工业大学 土木建筑工程学院, 安徽 合肥 230009;
3. 美国伯克利加州大学 环境科学、政策与管理系, 加州)

GPS Quasi Geoid Fitting Based on Support Vector Machine Technology

WU Zhao-fu^{1,2}, GONG Peng³, GAO Fei², WANG Nong²

(1. Department of Urban and Resources Sciences, Nanjing University, Nanjing 210093, China; 2. School of Civil Engineering, Hefei University of Technology, Hefei 230009, China; 3. Dept. of ESPM, University of California Berkeley, California, U. S. A.)

Abstract: Plan imetric control results of GPS surveying have been used widely in all kinds of engineering, while its height information is being researched at present. This paper introduces statistical learning theory and support vector machine, proposes a new method, support vector machine technology, to simulate quasi-geoid. Based on real GPS surveying datum, we did an experiment with support vector machine, neural network and polynomial technology to simulate quasi-geoid. The compared and analyzed test results show that the simulating accuracy of support vector machine achieved the same accuracy of neural network and polynomial. Moreover, support vector machine can settle many questions that neural network must face, such as real-time handling data, over-learning, convergence too slowly, falling into regional maximum easily etc.

Key words: support vector machine; statistical learning theory; neural network; GPS height; function approximation

摘 要: GPS 高精度平面控制成果在各类工程中已经得到了广泛应用, 然而其高程信息目前正在作进一步的研究。介绍统计学习理论和支持向量机, 提出利用支持向量机技术进行似大地水准面拟合。以实测 GPS 定位数据为试验资料, 对支持向量机和神经网络以及多项式拟合的结果进行比较分析: 支持向量机技术拟合数据的精度达到了神经网络和多项式拟合的精度, 并且解决了神经网络技术不能实时处理数据、过学习、收敛速度慢、易陷于局部极值等问题。

关键词: 支持向量机; 统计学习理论; 神经网络; GPS 高程; 函数逼近

1 引 言

全球定位技术是 20 世纪最伟大的技术之一, 其

平面相对定位精度已达到 $10^8 \sim 10^9$, 绝对定位精度达到毫米级, 可以取代传统的平面控制测量。但是 GPS 在高程测量方面还不能完全取代几何水准测

量,其高精度高程信息还没有得到很好的应用。

利用 GPS 确定测站点正常高的关键在于高程异常 ζ 的测定。目前计算高程异常有直接法和拟合法,直接法需要有一定精度、分布良好的重力数据和地形数据,这在实际工程中往往很难满足。拟合法则是通过 GPS 测量获得各控制点的大地高,并利用水准测量求得正常高,进而计算出这些点的高程异常,然后用内插的方法获得待定点的高程异常值。拟合法简单易行,这在当前的工程中得到了广泛的应用。由于模型的差异,产生了不同的拟合方法:加权平均法、多项式拟合、多面函数法、以及神经网络法^[1,2]等。其中神经网络法转换精度高于二次多项式拟合法,并且转换的精度比较稳定^[1]。但是神经网络法是一种事后处理方法,不能实时处理 GPS 数据,并且网络还存在过学习、收敛速度慢、易陷于局部极值等问题。本文将利用小样本统计学习理论,采用支持向量机进行似大地水准面的拟合,克服神经网络在似大地水准面拟合中的缺陷。

2 统计学习理论与支持向量机^[3,4]

在函数逼近机器学习问题中,通常采用经验风险最小化(ERM)准则。然而 ERM 准则并没有经过充分的理论论证,只是直观上想当然做法,在实际问题中学习机器的 VC 维越高则置信范围越大,导致真实风险与经验风险之间可能的差别越大。这就是为什么会出现过学习现象的原因。机器学习过程不但要使经验风险最小,还要使 VC 维尽量小以缩小置信范围,才能取得较小的实际风险,从而能够对未来样本有较好的推广性。传统方法中选择学习模型和算法的过程就是调整置信范围的过程,如果模型比较适合现有的训练样本,则可以取得比较好的效果。但因为缺乏理论指导,这种选择只能依赖先验知识和经验,造成了如神经网络等方法对使用者“技巧”的过分依赖。统计学习理论提出了一种新的策略:把函数集构造为一个函数子集序列,使各个子集按照 VC 维的大小排列;在每个子集中寻找最小经验风险,在子集间折衷考虑经验风险和置信范围,取得实际风险的最小。这种思想称作结构风险最小化(即 SRM 准则)。

支持向量机(SVM)是一种基于统计学习理论的机器学习算法,采用 SRM 准则,在最小化样本点误差的同时缩小模型泛化误差的上界,从而

提高模型的泛化能力。在函数拟合问题中,对于数据集 $\{x_i, y_i\} i = 1, 2, \dots, n, x_i \in R^d, y_i \in R$, 如果用线性函数 $f(x) = w \cdot x + b$ 来拟合,拟合精度为 ε , 引入松弛因子 $\zeta_i \geq 0$ 和 $\zeta_i^* \geq 0$, 根据 SRM 准则,拟合函数 $f(x)$ 应使 $\frac{1}{2} |w|^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*)$ 最小($C > 0$ 为惩罚因子),同时要满足条件 $y_i - w \cdot x_i - b \leq \varepsilon + \zeta_i, w \cdot x_i + b - y_i \leq \varepsilon + \zeta_i^*, \zeta_i \geq 0$ 和 $\zeta_i^* \geq 0 \quad i = 1, 2, \dots, n$ (1)

采用 Lagrange 优化方法可以得到该优化目标对偶问题,则在条件 $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, n$ (2)

下,对 Lagrange 因子 α_i, α_i^* 最大化目标函数 $W(\alpha, \alpha^*) = -\varepsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) + \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i \cdot x_j)$ (3)

得回归函数为 $f(x) = (w \cdot x) + b = \sum_{i=1}^n (\alpha_i^* - \alpha_i) (x_i \cdot x) + b^*$ (4)

这里 α_i, α_j^* 只有小部分不为 0,它们对应的样本就是支持向量,在样本集中剔除非支持向量重新拟合,拟合结果应不变。对于非线性函数拟合,采用适当的内积函数 $k(x_i, x_j)$ 就可以实现某一非线性变换 $\varphi(x)$ 后的线性逼近,此时拟合函数变为

$f(x) = (w \cdot x) + b = \sum_{i=1}^n (\alpha_i^* - \alpha_i) (\varphi(x_i) \cdot \varphi(x)) + b^* = \sum_{i=1}^n (\alpha_i^* - \alpha_i) k(x_i \cdot x) + b^*$ (5)

SVM 中不同的内积核函数将形成不同的算法,目前常用的核函数主要有多项式核函数、径向基函数(RBF)和 Sigmoid 核函数等^[5]。

3 基于支持向量机的似大地水准面拟合

3.1 试验研究

由上述统计学习理论可知,支持向量机是一种新的机器学习算法,具有较好的模型泛化能

力,可以很好地进行函数拟合。本文利用支持向量机对似大地水准面进行了拟合。试验中采用的数据来自安徽合肥某开发区 GPS 实地测量成果。测区控制点覆盖面积约为 19 km²,实测 GPS 控制点 61 个,全部进行了水准联测,各 GPS 控制点空间分布如图 1 所示。支持向量机、神经网络和多项式均用相同的 37 个学习集和 37 个检测集分别进行训练和检测,试验中采用的是 5 层 BP 神经网络,其中隐层有 100 个神经元,学习率初始值 0.5,动量常数为 0.9,各个学习集训练了 3 000 次。支持向量机采用的是二阶多项式核函数。

3.2 粗差检验

在拟合似大地水准面中,实测几何水准点高程的可靠性尤为重要,因此需要进行粗差检验。这里利用拟合残差分别采用 Grubbs 判别法和 Dixon 检验法^[6]进行检验。

以 $n = 30, \alpha = 0.05$ 为引数查表可知临界值 $G_0(30, 0.05) = 2.881$,利用拟合残差实际计算出 Grubbs 统计量 $g_n = 2.123, g'_n = -2.343$ 。其绝对值均小于临界值 $G_0(30, 0.05)$ 。

以 $n = 30, \alpha = 0.05$ 为引数查表可知临界值 $r_0(30, 0.05) = 0.376$,利用拟合残差实际计算出 Dixon 统计量 $r_{22} = 0.181, r'_{22} = 0.160$ 。均小于临界值 $r_0(30, 0.05)$ 。

故可以认为参与拟合的几何水准点实测高程是可靠的。

3.3 精度分析

分别比较不同检测集中拟合中误差、最大拟合残差和拟合计算时间,比较结果如表 1 所示,其中 V 表示拟合残差, V_{max} 为最大拟合残差。同时在不同检测集中比较支持向量机拟合误差小于神经网络和多项式拟合误差的样本数,比较结果如图 2 所示。由表 1 可知 3 种拟合方法拟合中误差大致相同,但是笔者在试验中采用神经网络方法要达到该精度消耗的时间约为 4~5 h,而采用支持向量机和多项式方法则只需 4~5 min,可见支持向量机方法大大地提高了数据拟合的速度,满足实际工程计算的需要。由图 2 中检测集样本数比较曲线可知,多数情况下支持向量机拟合误差小于神经网络和多项式拟合误差的样本数要超过检测集样本数的一半,尤其是同多项式拟合比较,可见支持向量机拟合完全能够达到神经网络和多项式拟合精度。图 3 为拟合中误差曲线,由

该图可知多项式拟合中误差略大于支持向量机和神经网络拟合中误差。把检测集样本点的拟合误差与四等和等外水准测量限差比较,结果如表 2 所示,可见 3 种方法拟合误差都达到了等外水准的精度要求。

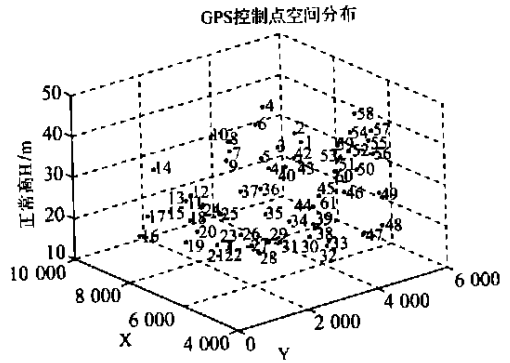


图 1 GPS 控制点空间分布图

Fig. 1 Spatial distribution map of GPS points

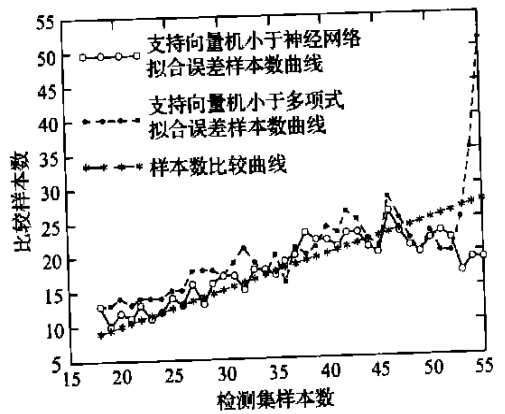


图 2 样本比较曲线

Fig. 2 Samples comparison curve

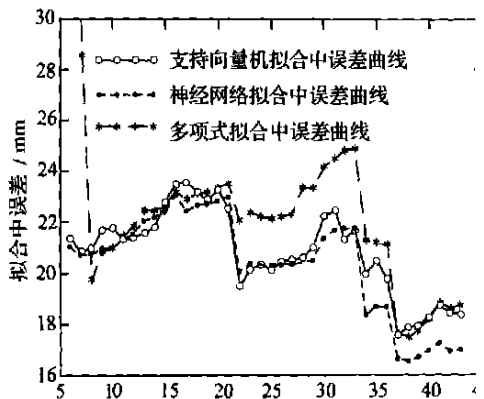


图 3 拟合中误差曲线

Fig. 3 Error curve of fitting error

表1 不同拟合模型计算结果的比较

Tab. 1 Comparison of calculating results between different models

检测集1	拟合中误差/mm	拟合点数	V_{max} / mm	$V > 20$ mm 个数	计算时间
支持向量机法	17.5	24	58	8	4~5 min
神经网络法	16.6	24	57	10	4~5 h
多项式拟合法	17.6	24	60	8	4~5 min
检测集2	拟合中误差/mm	拟合点数	V_{max} / mm	$V > 20$ mm 个数	计算时间
支持向量机法	17.8	23	59	8	4~5 min
神经网络法	16.5	23	57	10	4~5 h
多项式拟合法	17.4	23	60	8	4~5 min

表2 拟合精度比较表

Tab. 2 Comparison of simulating results

序号	至最近水准点距离/m	水准限差/mm	等外水准限差/mm	支持向量机拟合误差/mm	神经网络拟合误差/mm	多项式拟合误差/mm
1	382.204 0	12.4	24.7	14.2	24.3	12.4
2	586.523 4	15.3	30.6	6.3	5.1	7.3
3	836.733 0	18.3	36.6	22.2	17.4	15.3
4	331.368 6	11.5	23.0	12.4	21.4	12.1
5	459.756 6	13.6	27.1	13.8	19.5	14.8
6	547.713 5	14.8	29.6	0.4	1.0	2.4
7	402.954 6	12.7	25.4	5.4	2.8	1.4
8	298.600 8	10.9	21.9	6.7	1.2	5.2
9	369.385 0	12.2	24.3	3.7	7.8	4.7
10	532.415 8	14.6	29.2	17.3	15.3	18.9
11	369.385 0	12.2	24.3	4.8	7.3	8.2
12	405.036 0	12.7	25.5	5.6	6.7	11.3
13	385.183 6	12.4	24.8	24.2	19.5	12.3
14	389.322 2	12.5	25.0	8.7	12.6	21.1
15	489.740 2	14.0	28.0	20.5	18.6	14.5

4 结论

1. GPS拟合水准测量精度主要受到3部分影响:GPS大地高测量误差、已知水准点高程测量误差和高程异常拟合误差。其中前两项的精度

目前可以得到较好的保证;而拟合误差是影响GPS水准测量精度的一个重要因素,它与测区高程异常变化、已知水准点的密度及其分布、拟合模型精度等因素有关。图3中3种拟合法拟合误差曲线变化趋势大致相同,可见拟合误差与已知水准点的分布有关。

2. 对于似大地水准面比较复杂的测区来说,单纯采用某个多项式模型可能很难较好地拟合,而支持向量机和神经网络均为机器学习方法,它们具有良好的自适应性、高度非线性性和较好的泛化能力,支持向量机方法转换GPS高程的精度达到了神经网络方法转换GPS高程的精度。

3. 神经网络方法属于事后处理,支持向量机方法避免了神经网络方法中存在的网络过学习、收敛速度慢、易陷于局部极值等问题,因而可以将该拟合法集成到GPS数据处理软件中,具有实际使用和推广价值。

参考文献:

- [1] YANG Ming-qing, JIN Fan, ZHU Da-cheng, *et al.* Conversion of GPS Height by Artificial Neural Network Method[J]. *Acta Geodaetica et Cartographica Sinica*, 1999, 28(4): 301-307. (in Chinese)
- [2] HU Wei-sheng, HUA Xi-sheng, BAO Xing-nan. Neural Network Method for GPS Height Transformation[J]. *Journal of Hohai University*, 2001, 29(6): 87-89. (in Chinese)
- [3] ZHANG Xue-gong. Introduction to Statistical Learning Theory and Support Vector Machines[J]. *Acta Automatica Sinica*, 2000, 26(1): 32-42. (in Chinese)
- [4] ZHU Guo-qiang, LIU Shi-rong, YU Jin-shou. Support Vector Machine and Its Applications to Function Approximation[J]. *Journal of East China University of Science and Technology*, 2002, 28(5): 555-559. (in Chinese)
- [5] CHRISTIANINI N, SHAW T J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods[M]. London: Cambridge University Press, 2000.
- [6] ZHANG Fang-ren, ZHANG Jin-tong. Statistic Distribution and Test of Surveying Error[M]. Beijing: Publishing House of Surveying and Mapping, 2001. (in Chinese)