

基于小波分析和鲁棒最小二乘的信息融合估计算法

朱培逸^{1,2}, 孙顺远¹, 刘 柱¹, 徐保国¹

(1 江南大学 物联网工程学院, 江苏 无锡 214122; 2 常熟理工学院 电气与自动化工程学院, 江苏 常熟 215500)

摘 要: 对于农田土壤的多传感器检测系统来说,其测量的信息具有非线性和时空变异等特性,因此对信息融合方法的基本要求是具有鲁棒性和并行处理能力.应用小波分析理论,对原始测量数据进行了降噪处理,使降噪后的数据更能反映土壤的本质及变化规律;应用鲁棒最小二乘估计技术可以对不同传感器数据进行综合处理,去除冗余,克服歧义,得到比任何单个传感器更全面、更准确、更可靠的信息.针对农业中的自然环境具有很强的不确定性和经验性,运用基于小波分析的降噪和现代信息融合思想,提出了一种基于小波降噪和鲁棒最小二乘的信息融合估计方法.通过实验分析,其结果表明该方法是可行和值得研究的.

关键词: 小波分析;鲁棒最小二乘;土壤信息;信息融合

中图分类号: TP27

文献标识码: A

文章编号: 1000-7180(2012)03-0018-04

Information Fusion Estimation Algorithm by Robust Least Squares and Wavelet Analysis

ZHU Pei-yi¹, SUN Shun-yuan¹, LIU Zhu¹, XU Bao-guo¹

(1 School of IOT, Jiangnan University, Wuxi 214122, China; 2 School of Electrical and Automation Engineering, Changshu Institute of Technology, Changshu 215500, China)

Abstract: To the multi-sensor detection system, its measurement information is non-linear and spatiotemporal variation, etc. so the basic requirement of information fusion method is robust and parallel processing capabilities. In this paper, the wavelet analysis theory is applied to denoise the original data, which makes the data after denoising is better to reflect the nature of soil and its Variation Law. Robust least squares estimation technique can be integrated with different sensors data, to remove redundancy, to overcome the ambiguity, so it can obtain more comprehensive, more accurate and reliable information than any single sensor. Considering highly uncertain and empirical of the natural environment in farmland soil, the information fusion estimation method based on wavelet denoising and robust least squares is proposed. The analysis of experimental results shows that the method is feasible and worth studying.

Key words: wavelet analysis; robust least squares; soil information; information fusion

1 引言

目前对于农田土壤属性信息的监测多利用各种先进的传感器技术,多种传感器的测量为这些信息的获取提供了不同的方法,但是也为后续信息处理提出了更高的要求,传统的实验室数据分析方法已经很难适应数字农业的要求.因此,如果能将测量

土壤属性信息的多传感器信息,经过有效的信息融合算法进行处理,得到一个全面而准确的信息,有利于农田的经营管理和决策,对农作物产量的预测也起到重要作用^[1-2].

对于农田的多传感器检测系统来说,其测量的信息具有非线性和时空变异等特性,因此对信息融合方法的基本要求是具有鲁棒性和并行处理能力;

收稿日期: 2011-07-11; **修回日期:** 2011-08-15

基金项目: 江苏省高校自然科学基金项目(10KJD510001)

此外还有方法的运算速度和精度,与前续预处理系统和后续信息识别系统的接口性能,与不同技术和方法的协调能力,对信息样本的要求等等.因此面向这种复杂应用背景(非线性、非高斯环境)的融合方法,则很少有文献报道,常规的卡尔曼滤波法、神经网络、小波分析等融合算法具有一定的优点,但这些算法存在线性化误差,估计精度不高,而且不适用于非高斯条件系统^[3-4].本文提出一种基于小波去噪和鲁棒最小二乘的融合估计算法,能保证其快速性和对多数据源数据处理能力等,从而将农田中的多传感器信息进行融合处理,直接对农作物的生产进行有效的指导.

2 小波降噪的原理

小波在对信号的处理中具有良好的时频局部化特性、特别的去噪能力以及便于提取弱信号的特点,在非平稳信号的处理以及对信号进行时频分析中得到了广泛的应用,在研究复杂非线性问题时,取得了很好的效果^[5].在各种小波降噪方法中非线性小波变换阈值法由于具有噪声几乎完全得到抑制、计算速度快、广泛的适应性的优点受到了科研工作者的青睐,本文基于小波分析理论,结合采集的农田土壤信息特点,采用非线性小波变换阈值法对测量的土壤信息进行观测.

基于多尺度分析的离散小波变换的基本思想是:将待处理的信号用正交变换在不同尺度进行分解,分解到粗尺度上的信号称之为平滑信号;在细尺度上存在,而在粗尺度上消失的信号称之为细节信号^[6].

在尺度 i 上,对给定的信号序列 $X(i, n) \in V_i$, $V_i \subset l^2(Z)$, $n \in Z$,通过一个脉冲响应是 $h(n)$ 的低通滤波器可以获得粗尺度上的平滑信号(尺度 $i-1$):

$$x_v(i-1, n) = \sum_i h(2n-k)x(i, k) \quad (1)$$

通过一个脉冲响应是 $g(n)$ 的高通滤波器可以获得细尺度上的信号 $x_D(i-1, n) \in W_{i-1}$

$$x_D(i-1, n) = \sum_i g(2n-k)x(i, k) \quad (2)$$

原始信号 $x(i, n)$ 可由 $x_v(i-1, n)$ 和 $x_D(i-1, n)$ 完全重构得到,即:

$$\begin{aligned} x(i, n) &= \sum_i h(2n-k)x_v(i-1, k) \\ &+ \sum_i g(2n-k)x_D(i-1, k) \end{aligned} \quad (3)$$

通过小波的分解和重构,即可达到对土壤信息的降噪的处理,该方法能得到原始信号的近似最优

估计,并且具有非常广泛的适应性.

3 基于鲁棒最小二乘的估计

给定一组相互独立的 n 个传感器对某性能参数的观测方程:

$$y_i = H_i x + e \quad (4)$$

其中: x 为估计参数, y 为 n 维测量向量, e 为 n 维噪声向量包含内部噪声和环境噪声, H 为已知的 n 常向量.最小二乘估计的准则是指在参数空间中寻找一个最优值,使得利用该估计值对测量的各分量的拟合误差平方和最小即:

$$\begin{aligned} J(\hat{x}) &= (Y - H\hat{x})^T (Y - H\hat{x}) \\ &= \sum_{i=1}^k (y_i - H_i \hat{x})^2 \end{aligned} \quad (5)$$

式(5)达到极小,令式(5)关于 $\hat{x}$ 的梯度为零可得到最小二乘估计.

最小二乘估计算法虽然具有一定的优势,但是当几个传感器精度相差很大时,最小二乘估计值的精度还不如只采用一个精度高的传感器测量的精度.为了在最小二乘的基础上得到更高精度的估计,对式(5)中引入鲁棒因子 γ ,可得到如下的优化命题:

$$J(\hat{x}) = \sum_{i=1}^n \gamma_i [(y_i - H_i \hat{x})]^2 \quad (6)$$

其中 γ_i 反应的是最小二乘估计的精度,称为鲁棒因子,其取值随最小二乘估计的精度变化而自适应变化,形式为:

$$\gamma_i = e^{-\tau(y_i - y_i)^2} \quad (7)$$

其中 τ 为鲁棒因子的鲁棒系数,取值范围为 $\tau \in [0, 1]$.当最小二乘估计有偏差,或者学习尚未收敛时,鲁棒因子就会根据 y 估计值和实际值误差的平方调整大小,引入较强的补偿机制,以减小传感器外界噪声干扰对估计精度的影响.因此,增强了估计的鲁棒性.

4 基于小波分析和鲁棒最小二乘的估计融合方法

基于小波分析和鲁棒最小二乘的融合估计的思想是:对测量的信号 $Y(N) = \{y(N, k)_{k \in z}\}$ 进行离散小波变换,根据系统的先验信息和细节信号的统计特性进行分析,对细节信号进行阈值处理,再由离散小波逆变换重构测量信号,作为最小二乘估计的测量值 $\hat{Y}(N) = \{\hat{y}(N, k)_{k \in z}\}$.利用离散小波变换,从原始尺度 N 开始分解 m ($1 \leq m \leq N-1$) 次,

得到尺度 $N-1, N-2, \dots, N-m$ 上的平滑信号 $Y_v(N-1), Y_v(N-2), \dots, Y_v(N-m)$ 和细节信号 $Y_D(N-1), Y_D(N-2), \dots, Y_D(N-m)$:

$$y_v(i-1, n) = \sum_i h(2n-k)y_v(i, k) \\ = \sum_i h(k)y_v(i, 2n-k) \quad (8)$$

$$y_D(i-1, n) = \sum_i g(2n-k)y_D(i, k) \\ = \sum_i g(k)y_D(i, 2n-k) \quad (9)$$

其中 $Y_D(i) = \{y_D(i, n)\}_{n \in z}$, $Y_v(i) = \{y_v(i, n)\}_{n \in z}$, $i = N, N-1, \dots, N-m+1$; $m = 1, 2, \dots, N-1$. 对 $Y_D(i)$ 进行预处理后记为 $\bar{Y}_D(i)$, 利用小波的重构公式, 尺度 i 上新的测量 $\bar{Y}_v(i)$ 可由尺度 $i-1$ 上的信号 $Y_v(i-1)$ 处理后的细节 $\bar{Y}_D(i-1)$ 重构得到:

$$\bar{y}_v(i, n) = \sum_i h(2n-k)y_v(i-1, k) \\ + \sum_i g(2n-k)\bar{y}_D(i-1, k) \quad (10)$$

同理可得到尺度 $N-1, N-2, \dots, N-m$ 上的细节信号 $Y_D(N-1), Y_D(N-2), \dots, Y_D(N-m)$ 的预处理后的新的细节信号 $\bar{Y}_D(N-1), \bar{Y}_D(N-2), \dots$, 来代替原来的细节信号, 再利用尺度 $N-m$ 上的平滑信号 $Y_v(N-m)$ 可重构尺度 N 上的测量值. 因此对测量信号原始尺度 N 上的 m 个预处理后的测量 $Y_v^{N-1}(N), Y_v^{N-2}(N), \dots, Y_v^{N-m}(N)$. 以 $Y_v^{N-1}(N), Y_v^{N-2}(N), \dots, Y_v^{N-m}(N)$ 中任一个或者其加权的平均为新的测量值, 记作: $\hat{Y}(N) = \{\hat{y}(N, k)\}_{k \in z}$, 用它来代替原始的测量信号 $Y(N) = \{y(N, k)\}_{k \in z}$ 进行鲁棒最小二乘估计:

$$\hat{\beta} = (X^T X)^{-1} X^T \hat{Y} \quad (11)$$

利用小波变换对信号进行去噪处理, 关键是选择一个好的小波以及选取适当的阈值. 选择小波时, 一般要求其具有良好的时频局部性. 阈值的选择, 一般分为硬阈值和软阈值. 一般来说两种阈值滤波方法中硬阈值比软阈值处理后的信号更粗糙一些, 所以本文采用软阈值处理方法.

5 实验及分析

本文采用 2 种比较典型的估计方法作为融合的基础, 包括平均估计法和加权最小二乘法(简称算法 1 和算法 2). 为了提高数据的可靠性和准确性, 本文通过数据预处理对错误和丢失数据进行识别和修复, 并利用指数平滑法去除尖锐的噪声. 本实验

采用山东东营垦利县土壤信息数据, 利用 10 个传感器对该位置 0~75 cm 深度的大地电导率进行 100 次测量, 各个传感器的测量噪声方差如表 1 所示.

表 1 传感器的测量噪声方差

No.	1	2	3	4	5	6	7	8	9	10
σ	0.05	0.07	0.10	0.20	0.30	0.25	0.10	0.10	0.20	0.30

采用一个数据集对上述常规估计方法进行测试与比较, 在相同条件下评价各方法的估计性能. 本文所用的估计误差包括均方根误差、平均绝对相对误差、最大绝对相对误差, 实验结果如表 2 所示.

表 2 估计误差的比较

方法	均方根误差	平均绝对相对误差	最大绝对相对误差
算法 1	0.398	0.031	0.058
算法 2	0.149	0.022	0.041
本文算法	0.095	0.013	0.017

从以上估计结果可以看出, 以上 3 种土壤电导率融合估计方法都可以进行相对准确的估计, 只是估计精度有所差异. 从估计的精度和稳定性看, 本文提出的融合估计方法的各项误差指标均小于其它两种估计方法, 结果证明该方法可以有效地提高传感器测量的估计精度.

图 1 分别绘出了三种方法的土壤电导率的估计曲线. 分别比较图中各曲线可见: 三种融合估计算法中本文算法较前两种算法而言具有更强的稳定性和鲁棒性, 而估计误差更小. 仿真验证了该算法的有效性.

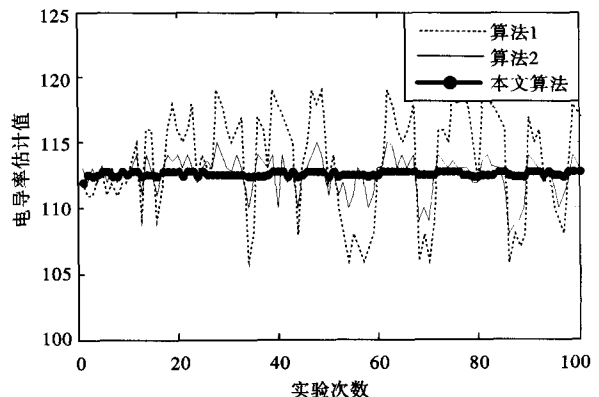


图 1 三种方法的估计曲线

6 结束语

理论推导与仿真研究表明,利用小波分析和与鲁棒最小二乘法相结合的多传感器信息融合估计算法对参数进行检测,能够充分利用量测数据,将噪声最大程度的抑制.在估计的无偏性和均方误差最小的前提下,该方法表现出较好的稳健性和鲁棒性,是一种有效的土壤信息融合估计方法.

参考文献:

- [1] 庞凤,李廷轩,王永东,等.土壤速效氮、磷、钾含量空间变异特征及其影响因子[J].植物营养与肥料学报,2009,15(1):114-120.
- [2] Sumfleth K, Duttman R. Prediction of soil property distribution in paddy soil landscape using terrain data and satellite information as indicators [J]. Ecological Indicators, 2008(8):485-501.
- [3] 王凤花,张淑娟.精细农业田间信息采集关键技术的研究进展[J].农业工程学报,2008,39(5):112-121.
- [4] Boonsompopphan Buree, Vearasilp Taweesak, Yos Russell S, et al. Field identification of soil series: in-

dexing and retrieving soil information while sharing experience and knowledge [J]. Field identification of soil series, 2008,173(10):736-744.

- [5] Donoho D L. Denoising by soft-threshold [J]. IEEE Transaction on Information, 1995(3):613-627.
- [6] 文成林.多尺度动态建模理论及其应用[M].北京:科学出版社,2008.

作者简介:



朱培逸 男,博士研究生.研究方向为智能信息处理、智能计算.

孙顺远 男,博士研究生.研究方向为无线传感器网络.

刘柱 男,硕士研究生.研究方向为过程控制.

徐保国 男,博士生导师,教授.研究方向为过程控制及优化.

(上接第17页)

从表2评估结果可知,本文提出的连锁经营企业网络安全评估模型可以网络规模和拓扑发生变化情况下,自动调整相似域的规模来消除网络变化对评估的影响.本文评估模型是一种有效的现代连锁经营企业网络安全评估方法.如果最后考虑连锁经营网络普遍存在的威胁因素(专家调查确定),那么网络安全综合风险值的评估精度会有所提高.

5 结束语

本文针对评估网络环境中实体风险相似性的原理,考虑到实体之间安全关联对整个网络风险的影响,通过将目标评估网络划分为若干个安全相似域进行风险评估,通过应用实例验证本文提出的评估思路是正确的,减少了网络安全计算量,很好的把握住系统安全风险的特点.但是该模型仍有很多问题需要进一步研究,如相似域的合理划分、描述语言的完备性等问题.

参考文献:

- [1] 孙强.基于定量安全风险评估模型的网络安全管理平台[J].微电子学与计算机,2010,27(5):70-73.
- [2] 徐礴晟,张傑穗,李生红.网络安全评估方法研究进展研究[J].信息安全与通信保密,2009(10):50-51.
- [3] 张伟,张宗华,赵霖.一种新型的网络安全策略模型[J].微电子学与计算机,2003,20(4):73-76.
- [4] 汤伟伟,姬敏.现代连锁经营与管理[M].北京:清华大学出版社,2001.
- [5] 邢栩嘉,林闯,蒋屹新.计算机系统脆弱性评估研究[J].计算机学报,2004,27(1):1-11.
- [6] 鲁智勇,冯超,余辉,等.网络安全性定量评估模型研究[J].计算机工程与科学,2009,31(10):100-104.
- [7] 徐源,戴青,郭靓基.基于安全相似域的风险评估模型[J].电子科学技术评论,2005,65(2):200-205.
- [8] 张海霞,连一峰,苏璞睿,等.基于安全状态域的网络评估模型[J].软件学报,2009,28(20):100-120.

作者简介:

蔡京玖 女,(1964—),硕士,讲师.研究方向为计算机网络技术与应用.