

文章编号: 2095-2163(2020)07-0005-06

中图分类号: TP274

文献标志码: A

非线性量测数据融合与自校准方法

傅惠民, 崔轶

(北京航空航天大学 小样本技术研究中心, 北京 100191)

摘要: 本文提出一种非线性量测数据融合方法, 给出多传感器非线性量测数据融合公式与计算步骤, 解决了在非非线性量测数据融合时传统的极大似然方法和最小二乘方法需要求解非线性方程(组)的难题; 同时提出一种非线性量测数据自校准融合方法, 给出量测数据未知系统误差(未知输入)的自识别自校准公式与计算步骤, 能够对非线性量测数据中事先无法校准的系统误差进行自动识别、估计、补偿和修正。文中详细讨论了秩采样量测数据融合方法和 Sigma 点采样量测数据融合方法, 秩采样量测数据自校准融合方法和 Sigma 点采样量测数据自校准融合方法。大量实例计算和仿真模拟验证表明, 本文方法不但比传统方法计算简单, 便于工程应用, 而且能够有效减小误差, 显著提高精度。

关键词: 非线性; 量测; 数据融合; 自校准; 多传感器

Nonlinear Measurement Data Fusion and Self-Calibration Methods

FU Huimin, CUI Yi

(Research Center of Small Sample Technology, Beihang University, Beijing 100191, China)

[Abstract] A nonlinear measurement data fusion method is proposed. The multisensor nonlinear measurement data fusion formulas and calculation steps are given, which solve the problem that the traditional maximum-likelihood and least-squares methods need to solve nonlinear equations in nonlinear measurement data fusion. At the same time, a nonlinear measurement data self-calibration fusion method is proposed. The self-calibration formulas and calculation steps of the unknown system error (unknown input) in measurement data are given, which can automatically recognize, estimate, compensate and correct system errors that cannot be calibrated beforehand in nonlinear measurement data. In this paper, the rank sampling and Sigma-point sampling measurement data fusion methods as well as the rank sampling and Sigma-point sampling measurement data self-calibration fusion methods are discussed in detail. A large number of example calculations and simulation validations show that the presented methods in this paper are not only simpler in calculation than traditional methods, but also more convenient for engineering applications, and can effectively reduce the errors and greatly improve the accuracy.

[Key words] Nonlinear; Measurement; Data fusion; Self-calibration; Multisensor

0 引言

随着在线监测、人工智能、物联网、导航控制、遥感探测等技术的快速发展, 多传感器数据融合已经越来越重要^[1]。但是, 目前比较成熟的是线性量测数据(量测方程为线性)之间的数据融合, 而非线性量测数据(量测方程为非线性)之间的数据融合, 虽然国内外研究人员也提出了许多方法, 但在实际工程应用中仍面临诸多问题。例如: 基于极大似然方法和最小二乘方法的非线性数据融合, 常常需要求解非线性方程(组); 加权融合方法则存在科学确定权重的问题; 贝叶斯数据融合方法又需要先验信息^[2]; 神经网络数据融合则需要大量数据训练等^[3]。为此, 本文进行了深入系统的研究, 基于秩

采样^[4]和 Sigma 点采样^[5], 提出一种新的非线性量测数据融合方法, 包括秩采样量测数据融合方法和 Sigma 点采样量测数据融合方法, 该方法在非非线性量测数据融合时, 不但计算更加简单, 而且精度也更高。此外, 由于受环境因素的影响、测量设备的不稳定(如陀螺仪和加速度计的漂移)等原因, 导致量测数据中往往含有事先无法校准的系统误差(偏差), 这严重影响量测精度的提高。量测数据自校准融合方法^[6], 自识别自校准滤波方法^[7], 含缓变未知输入的数据自校准滤波方法^[8], 这些方法解决了线性 and 可线性化的量测数据自校准融合问题。本文进一步提出一种非线性量测数据自校准融合方法, 包括秩采样量测数据自校准融合方法和 Sigma 点采样量

基金项目: 国家重点基础研究发展计划(2012CB720000); 工信部 2018 年智能制造综合标准化项目《基于数字仿真的机械产品可靠性测试方法标准研究与试验验证》。

作者简介: 傅惠民(1956-), 男, 博士, 教授, 博士生导师, “长江学者”奖励计划特聘教授, 主要研究方向: 小样本可靠性技术、自校准滤波方法、在线监测等。

通讯作者: 傅惠民 Email: fhm@buaa.edu.cn

收稿日期: 2020-05-18

测数据自校准融合方法,能够对非线性量测数据中事先无法校准的系统误差自动识别、估计、补偿和修正,从而提高量测估计精度。

1 线性和非线性量测数据融合方法

1.1 线性和非线性量测方程

设线性和非线性量测方程分别为式(1)和式(2):

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_{Y,k}, \quad (1)$$

$$\mathbf{Z}_k = \mathbf{h}_k(\mathbf{X}_k) + \mathbf{V}_{Z,k}. \quad (2)$$

其中, \mathbf{X}_k 为 m 维状态向量; \mathbf{Y}_k 为 n_Y 维量测向量; \mathbf{Z}_k 为 n_Z 维量测向量; \mathbf{H}_k 为量测矩阵, $\mathbf{h}_k(\cdot)$ 为非线性量测函数; $\mathbf{V}_{Y,k}$ 和 $\mathbf{V}_{Z,k}$ 是均值为零,协方差矩阵分别为 $\mathbf{R}_{Y,k}$ 和 $\mathbf{R}_{Z,k}$ 的量测噪声向量,并且满足式(3):

$$\begin{cases} E(\mathbf{V}_{Y,k}) = 0, & E(\mathbf{V}_{Y,k} \mathbf{V}_{Y,j}^T) = \mathbf{R}_{Y,k} \delta_{k,j}; \\ E(\mathbf{V}_{Z,k}) = 0, & E(\mathbf{V}_{Z,k} \mathbf{V}_{Z,j}^T) = \mathbf{R}_{Z,k} \delta_{k,j}; \\ E(\mathbf{V}_{Y,k} \mathbf{V}_{Z,j}^T) = 0. \end{cases} \quad (3)$$

其中, $E(\cdot)$ 为数学期望符号, $\delta_{k,j}$ 为 δ 函数,当 $k=j$ 时, $\delta_{k,j} = 1$, 当 $k \neq j$ 时, $\delta_{k,j} = 0$ 。

1.2 秩采样量测数据融合方法

首先,由线性量测方程式(1),基于加权最小二乘原理求得 \mathbf{X}_k 的初步估计 $\hat{\mathbf{X}}_k^{(0)}$ 为式(4):

$$\hat{\mathbf{X}}_k^{(0)} = (\mathbf{H}_k^T \mathbf{R}_{Y,k}^{-1} \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{R}_{Y,k}^{-1} \mathbf{Y}_k. \quad (4)$$

初步估计误差协方差矩阵为式(5):

$$\mathbf{P}_{X,k}^{(0)} = (\mathbf{H}_k^T \mathbf{R}_{Y,k}^{-1} \mathbf{H}_k)^{-1}. \quad (5)$$

采用秩采样方法,得到秩采样点集 $\{\mathbf{x}_{k,i}\}$ 为式(6):

$$\mathbf{x}_{k,i} = \begin{cases} \hat{\mathbf{X}}_k^{(0)} + u_{p_1} (\sqrt{\mathbf{P}_{X,k}^{(0)}})_i, & (i = 1, 2, \dots, m); \\ \hat{\mathbf{X}}_k^{(0)} - u_{p_1} (\sqrt{\mathbf{P}_{X,k}^{(0)}})_{i-m}, & (i = m+1, \dots, 2m); \\ \hat{\mathbf{X}}_k^{(0)} + u_{p_2} (\sqrt{\mathbf{P}_{X,k}^{(0)}})_{i-2m}, & (i = 2m+1, \dots, 3m); \\ \hat{\mathbf{X}}_k^{(0)} - u_{p_2} (\sqrt{\mathbf{P}_{X,k}^{(0)}})_{i-3m}, & (i = 3m+1, \dots, 4m). \end{cases} \quad (6)$$

其中, $(\sqrt{\mathbf{P}_{X,k}^{(0)}})_i$ 为 $\mathbf{P}_{X,k}^{(0)}$ 的平方根的第 i 列向量, u_{p_1} 和 u_{p_2} 为标准正态偏量, $u_{p_1} = 0.4823$, $u_{p_2} = 1.1281$ 。

最后,求得 \mathbf{X}_k 的状态估计 $\hat{\mathbf{X}}_k$ 为式(7):

$$\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_k^{(0)} + \mathbf{K}_k (\mathbf{Z}_k - \hat{\mathbf{Z}}_k) \quad (7)$$

其中: $\hat{\mathbf{Z}}_k$ 由式(8)计算:

$$\hat{\mathbf{Z}}_k = \frac{1}{4m} \sum_{i=1}^{4m} \mathbf{Z}_{k,i}, \quad (8)$$

$$\mathbf{Z}_{k,i} = \mathbf{h}_k(\mathbf{x}_{k,i}), \quad i = 1, 2, \dots, 4m. \quad (9)$$

增益矩阵 \mathbf{K}_k 由式(10)计算:

$$\mathbf{K}_k = \mathbf{P}_{XZ,k} \mathbf{P}_{Z,k}^{-1}. \quad (10)$$

其中, $\mathbf{P}_{XZ,k}$ 和 $\mathbf{P}_{Z,k}$ 由式(11)和式(12)给出:

$$\mathbf{P}_{XZ,k} = \frac{1}{\omega} \sum_{i=1}^{4m} [(\mathbf{x}_{k,i} - \hat{\mathbf{X}}_k^{(0)}) (\mathbf{Z}_{k,i} - \hat{\mathbf{Z}}_k)^T] \quad (11)$$

$$\mathbf{P}_{Z,k} = \frac{1}{\omega} \sum_{i=1}^{4m} [(\mathbf{Z}_{k,i} - \hat{\mathbf{Z}}_k) (\mathbf{Z}_{k,i} - \hat{\mathbf{Z}}_k)^T] + \mathbf{R}_{Z,k}, \quad (12)$$

$$\omega = 2(u_{p_1}^2 + u_{p_2}^2) = 3.0105. \quad (13)$$

状态估计误差协方差矩阵为式(14):

$$\mathbf{P}_{X,k} = \mathbf{P}_{X,k}^{(0)} - \mathbf{K}_k \mathbf{P}_{Z,k} \mathbf{K}_k^T. \quad (14)$$

1.3 Sigma 点采样量测数据融合方法

首先,初步估计 $\hat{\mathbf{X}}_k^{(0)}$ 及其误差协方差矩阵 $\mathbf{P}_{X,k}^{(0)}$ 仍分别由式(4)和式(5)给出。计算得到 Sigma 点集 $\{\mathbf{x}_{k,i}\}$ 为式(15):

$$\mathbf{x}_{k,i} = \begin{cases} \hat{\mathbf{X}}_k^{(0)} - (\sqrt{(m+\lambda) \mathbf{P}_{X,k}^{(0)}})_i, & (i = 1, 2, \dots, m); \\ \hat{\mathbf{X}}_k^{(0)}, & (i = m+1); \\ \hat{\mathbf{X}}_k^{(0)} + (\sqrt{(m+\lambda) \mathbf{P}_{X,k}^{(0)}})_{i-m-1}, & (i = m+2, \dots, 2m+1). \end{cases} \quad (15)$$

其中, λ 为比例参数,满足 $\lambda = \alpha^2(m + \kappa) - m$; κ 为调节参数,一般取 0 或 $3 - m$; α 为调节参数。

最后,根据式(7)进行状态估计,其中: $\hat{\mathbf{Z}}_k$ 由式(16)计算:

$$\hat{\mathbf{Z}}_k = \sum_{i=1}^{2m+1} w_M^i \mathbf{Z}_{k,i}, \quad (16)$$

$$\mathbf{Z}_{k,i} = \mathbf{h}_k(\mathbf{x}_{k,i}) \quad i = 1, 2, \dots, 2m+1, \quad (17)$$

$$w_M^i = \begin{cases} \frac{1}{2(m+\lambda)}, & (i = 1, 2, \dots, m, m+2, \dots, 2m+1), \\ \frac{\lambda}{m+\lambda}, & (i = m+1). \end{cases} \quad (18)$$

增益矩阵 \mathbf{K}_k 仍由式(10)计算,其中: $\mathbf{P}_{XZ,k}$ 和 $\mathbf{P}_{Z,k}$ 由式(19)和式(20)给出:

$$\mathbf{P}_{XZ,k} = \sum_{i=1}^{2m+1} [w_C^i (\mathbf{x}_{k,i} - \hat{\mathbf{X}}_k^{(0)}) (\mathbf{Z}_{k,i} - \hat{\mathbf{Z}}_k)^T], \quad (19)$$

$$\mathbf{P}_{Z,k} = \sum_{i=1}^{2m+1} [w_C^i (\mathbf{Z}_{k,i} - \hat{\mathbf{Z}}_k) (\mathbf{Z}_{k,i} - \hat{\mathbf{Z}}_k)^T] + \mathbf{R}_{Z,k},$$

$$w_C^i = \begin{cases} \frac{1}{2(m+\lambda)}, & (i = 1, 2, \dots, m, m+2, \dots, 2m+1), \\ \frac{\lambda}{m+\lambda} + (1 - \alpha^2 + \beta), & (i = m+1). \end{cases} \quad (20)$$

其中, β 为调节参数。状态估计误差协方差矩阵由式(14)给出。

2 非线性和非线性量测数据融合方法

设两组非线性量测方程分别为式(22)和式(23):

$$\mathbf{Y}_k = \mathbf{h}_{Y,k}(\mathbf{X}_k) + \mathbf{V}_{Y,k}, \quad (22)$$

$$\mathbf{Z}_k = \mathbf{h}_{Z,k}(\mathbf{X}_k) + \mathbf{V}_{Z,k}. \quad (23)$$

其中, $\mathbf{h}_{Y,k}(\cdot)$ 和 $\mathbf{h}_{Z,k}(\cdot)$ 为非线性量测函数, $\mathbf{X}_k, \mathbf{Y}_k, \mathbf{Z}_k, \mathbf{V}_{Y,k}$ 和 $\mathbf{V}_{Z,k}$ 的意义与式(1)和式(2)相同。

由于估计误差的协方差信息无法直接通过非线性方程求得,所以需对其中某一量测方程进行线性化,设非线性量测方程式(22)更便于线性化(如相邻两次量测时间间隔更短或方程非线性程度较低等)。对于线性化带来较大误差的情况,可引入未知输入 \mathbf{d}_k ,再采用下文中的线性和非线性量测自校准融合方法处理。

当 $k \geq 2$ 时,对式(22)在 $\hat{\mathbf{X}}_{k-1}$ 处进行泰勒展开,并保留1阶项,得式(24):

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{U}_k + \mathbf{V}_{Y,k}. \quad (24)$$

其中, \mathbf{H}_k 和 \mathbf{U}_k 满足式(25)和式(26):

$$\mathbf{H}_k = \left. \frac{\partial \mathbf{h}_{Y,k}}{\partial \mathbf{X}} \right|_{\mathbf{X}=\hat{\mathbf{X}}_{k-1}}, \quad (25)$$

$$\mathbf{U}_k = \mathbf{h}_{Y,k}(\hat{\mathbf{X}}_{k-1}) - \mathbf{H}_k \hat{\mathbf{X}}_{k-1}. \quad (26)$$

并令 $\mathbf{Y}_k^* = \mathbf{Y}_k - \mathbf{U}_k$, 则有式(27):

$$\mathbf{Y}_k^* = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_{Y,k}. \quad (27)$$

由式(23)和式(27)可知,已将非线性和非线性量测数据融合问题转化为线性和非线性量测数据融合问题,可以采用线性和非线性量测数据融合方法处理。

当 $k = 1$ 时,初步估计由式(28)给出:

$$\hat{\mathbf{X}}_1^{(0)} = \mathbf{h}_{Y,1}^{-1}(\mathbf{Y}_1). \quad (28)$$

其中, $\mathbf{h}_{Y,1}^{-1}(\cdot)$ 为非线性量测函数 $\mathbf{h}_{Y,1}(\cdot)$ 的逆函数。对于逆函数求解过程复杂的情况,可通过数值计算得到 $\hat{\mathbf{X}}_1^{(0)}$ 。并令 $\mathbf{H}_1 = \left. \frac{\partial \mathbf{h}_{Y,1}}{\partial \mathbf{X}} \right|_{\mathbf{X}=\hat{\mathbf{X}}_1^{(0)}}$, 再由式(5)计算得到 $\mathbf{P}_{X,1}^{(0)}$ 。此外,当逆函数 $\mathbf{h}_{Y,1}^{-1}(\cdot)$ 不

存在时,可采用极大似然方法求得 $\hat{\mathbf{X}}_1^{(0)}$ 。由于 $\hat{\mathbf{X}}_1^{(0)}$ 的精度对后面的计算结果影响不大,所以也可以采用其他方法获取 $\hat{\mathbf{X}}_1^{(0)}$ 。

3 线性和非线性量测自校准融合方法

设线性量测方程(含未知系统误差)和非线性量测方程分别为式(29)和式(30):

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{d}_k + \mathbf{V}_{Y,k}, \quad (29)$$

$$\mathbf{Z}_k = \mathbf{h}_k(\mathbf{X}_k) + \mathbf{V}_{Z,k}. \quad (30)$$

其中, $\mathbf{X}_k, \mathbf{Y}_k, \mathbf{Z}_k, \mathbf{H}_k, \mathbf{h}_k(\cdot), \mathbf{V}_{Y,k}$ 和 $\mathbf{V}_{Z,k}$ 的意义与式(1)和式(2)相同, \mathbf{d}_k 为 n_Y 维未知量测系统误差向量。

式(29)是含未知系统误差的量测方程,这种系统误差(如未知输入、环境因素影响等)事先无法校准;式(30)为不含系统误差的量测方程。由于 \mathbf{Y}_k 中含有系统误差,如果直接将其与 \mathbf{Z}_k 融合,所得状态估计将不是无偏估计,因此需要对未知的量测系统误差 \mathbf{d}_k 进行估计和修正。

工程实际中,当相邻两次量测的时间间隔足够小时,量测系统误差 \mathbf{d}_k 往往不会发生较大的变化。因此,当 $k \geq 2$ 时,系统误差 \mathbf{d}_k 可近似表示为式(31):

$$\mathbf{d}_k \approx \mathbf{d}_{k-1}. \quad (31)$$

首先,由式(29)和式(31)可得系统误差向量 \mathbf{d}_k 的初步估计 $\hat{\mathbf{d}}_k^{(0)}$ 为式(32):

$$\hat{\mathbf{d}}_k^{(0)} = \mathbf{Y}_{k-1} - \mathbf{H}_{k-1} \hat{\mathbf{X}}_{k-1}. \quad (32)$$

然后,考虑到偶然误差的影响,系统误差向量 \mathbf{d}_k 的估计 $\hat{\mathbf{d}}_k$ 由式(33)给出:

$$\hat{\mathbf{d}}_{k,j} = \begin{cases} 0, & \text{当 } |\hat{\mathbf{d}}_{k,j}^{(0)}| < c\sigma_{k,j} \text{ 时} \\ \hat{\mathbf{d}}_{k,j}^{(0)}, & \text{当 } |\hat{\mathbf{d}}_{k,j}^{(0)}| \geq c\sigma_{k,j} \text{ 时} \end{cases}. \quad (33)$$

其中, $\hat{\mathbf{d}}_{k,j}$ 和 $\hat{\mathbf{d}}_{k,j}^{(0)}$ 分别为 $\hat{\mathbf{d}}_k$ 和 $\hat{\mathbf{d}}_k^{(0)}$ 的第 j 个分量 ($j = 1, 2, \dots, n_Y$); $\sigma_{k,j} = \sqrt{\mathbf{R}_{Y,k}(j,j)}$, $\mathbf{R}_{Y,k}(j,j)$ 为协方差矩阵 $\mathbf{R}_{Y,k}$ 第 j 行、第 j 列的元素; c 为根据工程需要确定的偶然误差影响阈值调节参数。通常 c 取3(即 3σ 准则);当未知系统误差与量测噪声大小相当时,可以酌情取0~3之间的某一数值;当确有未知系统误差存在时,则取0。

在量测系统误差 \mathbf{d}_k 自识别自校准估计完成后,可以对线性和非线性量测数据进行融合。对 \mathbf{Y}_k 中的系统误差进行补偿,得到修正后的量测数据为式(34):

$$\mathbf{Y}_k^* = \mathbf{Y}_k - \hat{\mathbf{d}}_k. \quad (34)$$

则有式(35):

$$\mathbf{Y}_k^* = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_{Y,k}. \quad (35)$$

由式(30)和式(35)可知,已将线性和非线性量测自校准融合问题转化为线性和非线性量测数据融合问题,可以采用线性和非线性量测数据融合方法处理。

当 $k=1$ 时,只通过 \mathbf{Z}_1 进行状态估计得式(36):

$$\hat{\mathbf{X}}_1 = \mathbf{h}_1^{-1}(\mathbf{Z}_1). \quad (36)$$

其中, $\mathbf{h}_1^{-1}(\cdot)$ 表示非线性量测函数 $\mathbf{h}_1(\cdot)$ 的逆函数。对于逆函数求解过程复杂的情况,可通过数值计算得到 $\hat{\mathbf{X}}_1$ 。同样,当逆函数 $\mathbf{h}_1^{-1}(\cdot)$ 不存在时,可采用极大似然方法求得 $\hat{\mathbf{X}}_1$ 。由于 $\hat{\mathbf{X}}_1$ 的精度对后面的计算结果影响不大,所以也可以直接令 $\mathbf{d}_1 = 0$, 由式(29)求得 $\hat{\mathbf{X}}_1$ 。

4 非线性和非线性量测自校准融合方法

设两组非线性量测方程(其中一组含未知系统误差)分别为式(37)和式(38):

$$\mathbf{Y}_k = \mathbf{h}_{Y,k}(\mathbf{X}_k) + \mathbf{d}_k + \mathbf{V}_{Y,k}, \quad (37)$$

$$\mathbf{Z}_k = \mathbf{h}_{Z,k}(\mathbf{X}_k) + \mathbf{V}_{Z,k}. \quad (38)$$

其中, $\mathbf{X}_k, \mathbf{Y}_k, \mathbf{Z}_k, \mathbf{V}_{Y,k}$ 和 $\mathbf{V}_{Z,k}$ 的意义与式(1)和式(2)相同; $\mathbf{h}_{Y,k}(\cdot)$ 和 $\mathbf{h}_{Z,k}(\cdot)$ 的意义与式(22)和式(23)相同; \mathbf{d}_k 的意义与式(29)相同。

当 $k \geq 2$ 时,与上文同理,系统误差向量 \mathbf{d}_k 的初步估计 $\hat{\mathbf{d}}_k^{(0)}$ 为式(39):

$$\hat{\mathbf{d}}_k^{(0)} = \mathbf{Y}_{k-1} - \mathbf{h}_{Y,k-1}(\hat{\mathbf{X}}_{k-1}). \quad (39)$$

考虑到偶然误差的影响,系统误差向量 \mathbf{d}_k 的估计 $\hat{\mathbf{d}}_k$ 仍由式(33)给出。

对非线性量测方程式(37)在 $\hat{\mathbf{X}}_{k-1}$ 处进行线性化,得式(40):

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{U}_k + \mathbf{d}_k + \mathbf{V}_{Y,k}. \quad (40)$$

其中, \mathbf{H}_k 和 \mathbf{U}_k 由式(41)和式(42)给出:

$$\mathbf{H}_k = \left. \frac{\partial \mathbf{h}_{Y,k}}{\partial \mathbf{X}} \right|_{\mathbf{X}=\hat{\mathbf{X}}_{k-1}}. \quad (41)$$

$$\mathbf{U}_k = \mathbf{h}_{Y,k}(\hat{\mathbf{X}}_{k-1}) - \mathbf{H}_k \hat{\mathbf{X}}_{k-1}. \quad (42)$$

并令 $\mathbf{Y}_k^* = \mathbf{Y}_k - \mathbf{U}_k - \hat{\mathbf{d}}_k$, 则有式(43):

$$\mathbf{Y}_k^* = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_{Y,k}. \quad (43)$$

由式(38)和式(43)可知,已将非线性和非线性量测自校准融合问题转化为线性和非线性量测数据融合问题,可以采用线性和非线性量测数据融合方法处理。

当 $k=1$ 时,由式(36)进行状态估计。同样,当

逆函数 $\mathbf{h}_1^{-1}(\cdot)$ 不存在时,可采用极大似然方法求得 $\hat{\mathbf{X}}_1$ 。由于 $\hat{\mathbf{X}}_1$ 的精度对后面的计算结果影响不大,所以也可以采用其他方法获取 $\hat{\mathbf{X}}_1$ 。

5 仿真算例

5.1 非线性量测数据融合方法算例

考虑式(44)和式(45)的非线性量测系统:

$$Y_k = 0.5X_k + V_{Y,k}. \quad (44)$$

$$\mathbf{Z}_k = \begin{pmatrix} \hat{e} \\ \hat{e} \\ \hat{e} \\ \hat{e} \end{pmatrix} = \begin{pmatrix} 5\sin X_k - \frac{1}{X_k} \\ 0.9X_k + \frac{25X_k}{1+X_k^2} + \cos(0.004k) \\ \dot{u} \\ \dot{u} \end{pmatrix} + \mathbf{V}_{Z,k}. \quad (45)$$

其中, X_k 表示系统状态值, Y_k 和 \mathbf{Z}_k 表示量测值; $V_{Y,k}$ 和 $\mathbf{V}_{Z,k}$ 是服从高斯分布的零均值随机量测误差,其方差分别为 $R_{Y,k} = 16$ 和 $\mathbf{R}_{Z,k} = \begin{pmatrix} 64 & 0 \\ 0 & 80 \\ 0 & 0 \\ 0 & 48 \end{pmatrix}$ 。图1给出了系统状态 X_k 真值变化规律。

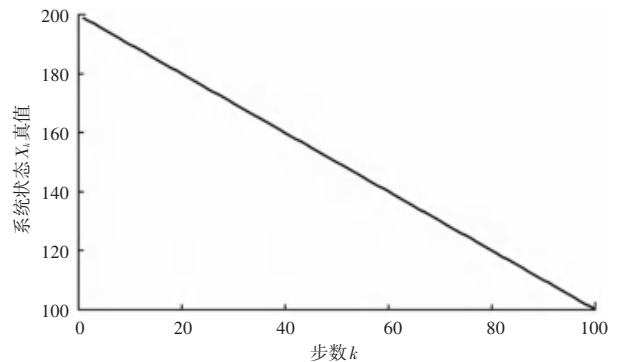


图1 系统状态 X_k 真值变化规律

Fig. 1 Change rule of true values of system state X_k

现分别采用传统极大似然估计方法和本文方法对系统状态进行估计,其中本文方法1为秩采样量测数据融合方法,本文方法2为Sigma点采样量测数据融合方法。由于工程实际中状态真值是未知的,所以极大似然估计方法的迭代初始值由式(44)求得。对系统进行5000次蒙特卡洛仿真模拟,分别求得上述方法的均方根误差,绘于图2,均方根误差的均值见表1。

表1 数据融合均方根误差均值比较

Tab. 1 Comparison of means of root mean square errors in data fusion

方法	均方根误差均值
传统方法	7.7660
本文方法1	2.1439
本文方法2	3.0512

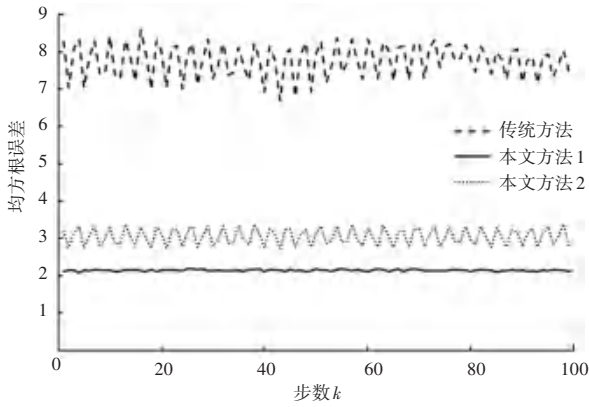


图 2 数据融合均方根误差比较

Fig. 2 Comparison of root mean square errors in data fusion

从图 2 和表 1 可以看到, 本文方法 1 的精度比传统方法的精度提高了 2.6 倍, 本文方法 2 的精度比传统方法的精度提高了 1.5 倍。其中, 秩采样量测数据融合方法的精度又略高于 Sigma 点采样量测数据融合方法, 这是因为秩采样跟踪非线性量测方程变化的能力要高于 Sigma 点采样。也正因为这个原因, 图 2 中本文方法 1 的均方根误差曲线波动更小。

此外, 从本例平均单次计算时间来看, 传统方法的时间为 0.020 7 s, 本文方法 1 的时间为 0.001 3 s, 本文方法 2 的时间为 0.001 1 s。与传统方法相比, 本文方法在提高精度的同时大幅减少了计算时间。

5.2 非线性量测数据自校准融合方法算例

考虑式(46)和式(47)的非线性量测系统:

$$Y_k = 0.5X_k + d_k + V_{Y,k}, \quad (46)$$

$$Z_k = \begin{pmatrix} \hat{e}_k \\ \hat{e}_k \\ \hat{e}_k \\ \hat{e}_k \end{pmatrix} = \begin{pmatrix} 5\sin X_k - \frac{1}{X_k} \\ 0.9X_k + \frac{25X_k}{1 + X_k^2} + \cos(0.004k) \end{pmatrix} \begin{pmatrix} \hat{u}_k \\ \hat{u}_k \end{pmatrix} + V_{Z,k}. \quad (47)$$

其中, d_k 为未知的系统误差, $V_{Y,k}$ 和 $V_{Z,k}$ 所对应的方差分别为 $R_{Y,k} = 16$ 和 $R_{Z,k} = \begin{pmatrix} 64 & 0 \\ 0 & 4 \end{pmatrix}$ 。系统状态 X_k 真值变化规律仍由图 1 给出, 并设置量测系统误差 d_k 为式(48):

$$d_k = \begin{cases} 0, & (1 \leq k \leq 25); \\ 18, & (26 \leq k \leq 50); \\ 0, & (51 \leq k \leq 75); \\ 36, & (76 \leq k \leq 100). \end{cases} \quad (48)$$

如式(46)所示, 本算例中的线性量测方程含有未知量测系统误差, 传统方法已无法对这种未知的系统误差进行估计和修正。现分别采用传统极大似然估计方法和本文方法对系统状态进行估计, 其中

本文方法 1 为秩采样量测数据自校准融合方法, 本文方法 2 为 Sigma 点采样量测数据自校准融合方法。对系统进行 5 000 次蒙特卡洛仿真模拟, 分别求得上述方法的均方根误差, 绘于图 3, 均方根误差的均值见表 2。

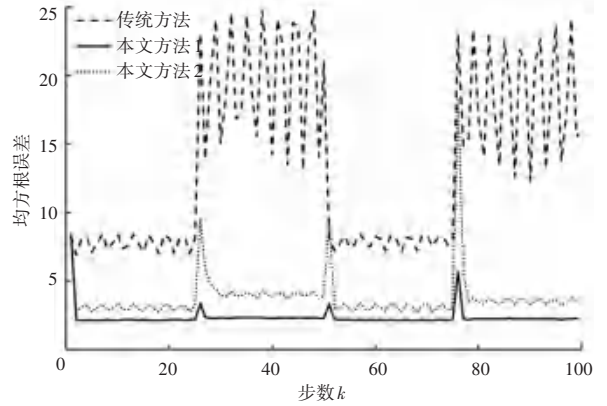


图 3 自校准融合均方根误差比较

Fig. 3 Comparison of root mean square errors in self-calibration fusion

表 2 自校准融合均方根误差均值比较

Tab. 2 Comparison of means of root mean square errors in self-calibration fusion

方 法	均方根误差均值
传统方法	13.425 4
本文方法 1	2.320 3
本文方法 2	3.765 7

从图 3 和表 2 可以看到, 本文方法 1 的精度比传统方法的精度提高了 4.7 倍, 本文方法 2 的精度比传统方法的精度提高了 2.5 倍。同样, 秩采样量测数据自校准融合方法的精度略高于 Sigma 点采样量测数据自校准融合方法。

此外, 从本例平均单次计算时间来看, 传统方法的时间为 0.020 7 s, 本文方法 1 的时间为 0.001 5 s, 本文方法 2 的时间为 0.001 2 s。与传统方法相比, 本文方法既可提高精度又能减少计算时间。

6 结束语

(1) 本文的非线性量测数据融合方法, 与传统方法相比, 不但计算更加简单而且具有更高的精度。其中, 秩采样量测数据融合方法的精度又略高于 Sigma 点采样量测数据融合方法, 这是因为秩采样跟踪非线性量测方程变化的能力要高于 Sigma 点采样。此外, 这两种方法都适用于高斯分布情况, 式(6)虽然采用了标准正态偏量, 但大量仿真验证表明, 秩采样量测数据融合方法对于非高斯分布情况也具有很高的估计精度;

(2)本文的非线性量测数据自校准融合方法,包括秩采样量测数据自校准融合方法和 Sigma 点采样量测数据自校准融合方法,能够对非线性量测数据中事先无法校准的系统误差自动进行识别、估计、补偿和修正,从而提高量测估计精度。同样,秩采样量测数据自校准融合方法的精度也略高于 Sigma 点采样量测数据自校准融合方法。

参考文献

- [1] MALAKAR B, ROY B K. Train localization using an adaptive multisensor data fusion technique [J]. *Transport*, 2019, 34(4): 508-516.
- [2] 华海亮,关维国,刘志建,等. 基于优化贝叶斯的室内 WiFi 与蓝牙融合定位算法 [J]. *计算机工程*, 2016, 42(11): 114-119.

(上接第 4 页)

3.4 实验结果

为了评价本文 CEVC 方法的性能,将其与现有分析方法 CETR 进行对比试验。实验通过对中文与英文数据集中的网页文件进行内容提取并计算性能指标,从而得出两种方法的提取性能,实验结果如表 1 所示。实验结果表明,CEVC 在中文数据集和英文数据集上的提取性能均优于 CETR。中文数据集的实验结果比英文数据集更好,研究发现其原因在于,英文数据集的 Web 页面内容结构较为扁平,标题、作者等内容信息可能以主要内容的兄弟节点形式呈现,给自动的提取过程带来了性能损失。随着 Web 技术的快速演进,网页文件倾向于使用更多复杂的嵌套结构来包装主要内容。中文数据集中的内容节点的平均深度大致为 8,而在英文数据集中则为 6。实验结果表明,CEVC 更善于处理复杂的 Web 网页结构。

4 结束语

基于包含停用词的文本更有可能是 Web 真实内容的实践总结,本文提出了一种区别于模板的方法来提取新闻和博客内容。将 Web 网页文件解析为 DOM 树后,递归计算每个节点的有效字符,重复地进入验证字符密集的子节点,最终在主内容节点上根据有效字符的比例设置阈值。本文方法 CEVC 在内容提取过程中,不需要训练数据和额外的页面,不对网页布局或结构做任何假设,通过结合超链接与停止词的统计信息,有效地将内容与噪声区别开来。实验结果表明,它可以有效地处理来自不同领域和语言的各种新闻和博客网页。与现有方法相比,本文方法在性能指标方面表现更优。

参考文献

- [1] GIBSON D, PUNERA K, TOMKINS A. The volume and evolution

- [3] 黄衍标,罗广岳,何铭金. BP 神经网络在巡逻机器人多传感器数据融合中的应用 [J]. *传感技术学报*, 2016, 29(12): 1936-1940.
- [4] 傅惠民,肖强,吴云章,等. 秩滤波方法 [J]. *机械强度*, 2014, 36(4): 521-526.
- [5] JULIER S J, UHLMANN J K, DURRANT-WHYTE H F. A new approach for filtering nonlinear systems [C]//*Proceedings of the American Control Conference*, Seattle, WA, USA, Jun. 1995: 1628-1632.
- [6] 傅惠民,杨海峰,文歆磊. 量测数据自校准融合方法 [J]. *航空动力学报*, 2019, 34(8): 1759-1763.
- [7] 傅惠民,杨海峰,付越帅. 自识别自校准滤波方法 [J]. *机电产品开发与创新*, 2019, 32(6): 8-12.
- [8] 傅惠民,杨海峰,付越帅,等. 含缓变未知输入的数据自校准滤波方法 [J]. *智能计算机与应用*, 2020, 10(1): 7-11.
- [9] of web page templates [C]//*Special Interest Tracks & Posters of the International Conference on World Wide Web*. 2005: 830-839.
- [2] KILGARRIFF A, GREFFENSTETTE G. Introduction to the special issue on the web as corpus [J]. *Computational Linguistics*, MIT Press, 2003, 29(3): 333-347.
- [3] BALUJA S. Browsing on small screens; recasting web - page segmentation into an efficient machine learning framework [C]//*Proceedings of the 15th International Conference on World Wide Web*. 2006: 33-42.
- [4] SAHUGUET A, AZAVANT F. Building intelligent web applications using lightweight wrappers [J]. *Data & Knowledge Engineering*, Elsevier, 2001, 36(3): 283-316.
- [5] BAR-YOSSEF Z, RAJAGOPALAN S. Template detection via data mining and its applications [C]//*Proceedings of the 11th International Conference on World Wide Web*. 2002: 580-591.
- [6] DEBNATH S, MITRA P, GILES C L. Automatic extraction of informative blocks from webpages [C]//*Proceedings of the 2005 ACM Symposium on Applied Computing*. 2005: 1722-1726.
- [7] MANTRATZIS C, ORGUN M, CASSIDY S. Separating XHTML content from navigation clutter using DOM-structure block analysis [C]//*Proceedings of the 16th ACM Conference on Hypertext and Hypermedia*. 2005: 145-147.
- [8] GOTTRON T. Combining content extraction heuristics: the CombinE system [C]//*Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services*. 2008: 591-595.
- [9] DAVISON B D. Recognizing nepotistic links on the web [J]. *Artificial Intelligence for Web Search*, 2000: 23-28.
- [10] FINN A, KUSHMERICK N, SMYTH B. Fact or Fiction: Content Classification for Digital Libraries. [C]//*DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*. 2001.
- [11] GOTTRON T. Content code blurring: A new approach to content extraction [C]//*The 19th International Workshop on Database and Expert Systems Applications*. IEEE, 2008: 29-33.
- [12] WENINGER T, HSU W H, HAN J. CETR: content extraction via tag ratios [C]//*Proceedings of the 19th International Conference on World Wide Web*. 2010: 971-980.
- [13] SUN F, SONG D, LIAO L. Dom based content extraction via text density [C]//*Proceedings of the 34th International ACM Conference on Research and Development in Information Retrieval*. 2011: 245-254.
- [14] WU G, LI L, HU X, et al. Web news extraction via path ratios [C]//*Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. 2013: 2059-2068.