

论文写作指导课程

结课论文

成绩: _____

二〇二〇年五月

基于卡尔曼滤波的发酵温度数据优化处理

刘云腾

(江南大学物联网工程学院, 江苏无锡 214000)

摘要: 固态酿酒发酵过程中的窖池温度直接影响酒质和出酒率, 由于发酵周期较长, 传感器中温度数据的采集过程中会出现孤立型野值点和斑点型野值点。因此提出了一种基于卡尔曼滤波联合全选主元高斯消去法的野值剔除方法。对于孤立型野值点问题, 先进行全选主元高斯消去法拟合, 然后进行卡尔曼滤波; 对于斑点型野值点问题, 先进行分段线性替代野值点, 然后进行全选主元高斯消去法拟合, 最后进行卡尔曼滤波。实验结果表明, 该方法有效地剔除了野值, 保证了数据的准确性, 满足了工业生产对数据精度的要求。

关键词: 全选主元高斯消去法; 孤立型野值和斑点型野值; 卡尔曼滤波; 曲线拟合;

Optimization of fermentation temperature data based on kalman filter

Tengyun Liu

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China;)

Abstract: Solid wine during fermentation pits temperature directly affects the quality of the wine and liquor yield, since the longer the fermentation period, the temperature data acquisition process appears in solitary speckled pattern outliers and outliers. For isolated point type outliers, to all the main yuan gauss elimination fitting, and then to Kalman filter; To the problem of point type spot outliers to piecewise linear replacement outliers point, then all principal gaussian fitting across France, finally to kalman filter. The experimental results show that the method Can effectively eliminate the outliers, ensures the accuracy of the data, satisfies the requirements of industrial production on the accuracy of the data.

Key words:all principal gauss elimination; solitary outliers and speckled outliers; Kalman filtering; curve fitting

1 引言

无线传感网络(WSN)获取的发酵温度数据一般会出现野值。在实际工况下, 野值产生的原因有很多, 包括嘈杂的生产环境, 节点间的通信异常以及操作人员的过失等^[1]。如果不对数据中的野值进行有效处理, 无客观反映发酵温度数据与时间的关系, 使得后期分析得到的固态发酵模型是错误的^[2], 因此必须要把野值全部剔除。

固态发酵温度测量异常数据根据其表现形式可以分为孤立型野值和斑点型野值两类^[3]。孤立型野值表现为一个个不连续的跳点, 斑点型野值即野值成片的出现, 并且野值的幅值相近。为了解决这个问题传统的做法如文献^[4]所示, 通过调节卡尔曼滤波的增益进行去噪, 可以减小误差, 但是总的误差依然很大。文献^[5]提出了基于新息的卡尔曼滤波方法, 通过比较预测值与观测值的残差是否超

过阈值来判断观测值是不是野值, 该方法在野值幅值较大的情况下, 可以有效地去除孤立型野值点, 但该方法用前一点的预测值代替野值, 当含有斑点型野值时, 其处理的效果不好。文献^[6]提出了卡尔曼滤波与最小二乘相结合的方法, 该方法在野值较少且幅值较小的情况下, 能够取得较好效果, 对于斑点型野值, 最小二乘算法会严重偏离真实值, 导致滤波效果不好。文献^[7]通过M-型估计-拉格朗日极值法来对野值进行剔除, 取得的效果较好, 但是对于野值点的起始点和结束点的求解还是不够完善。

文献^[8]提出基于卡尔曼滤波算法的不可信点判断方法来剔除野值, 其实现效果与文献^[5]类似。基于此, 文献^[9]分别从孤立型野值点和斑点型野值点的角度出发, 对现有的野值去除方法进行了比较, 对于孤立型的野值点, 差分法和多项式外推拟合法^[10]则能够取得较好的效果。但是对于斑点型野

值点的辨识和去除的效果不好，会存在去除不彻底或者剔除了有效点等问题。因此，对于野值点问题，单一的处理方法不能取得很好的处理效果。

对此，提出了以全选主元高斯消去法与卡尔曼滤波相结合的方法。全选主元高斯消去法具有抗干扰能力强、稳定性高等特性，所以可以通过它来去除幅值较大的野值，然后引入卡尔曼滤波算法，来消除噪声和低幅值的野值。本文通过仿真全选主元高斯消去法和卡尔曼滤波处理数据的过程，比较误差结果，分析产生误差的原因。

2 全选主元高斯消去法

曲线拟合即通过现有的观测数据选择出较好的连续曲线，使用解析式逼近已知离散点的方法，并且用获得的曲线方程分析变量之间的关系，常用的算法是最小二乘法。由于最小二乘法在求取的过程中。会把所有样本点的值代入到表达式中求解。当样本点的值不含有野值的时候，会取得较好的结果。但当遇到含有较多野值数据的样本时，结果偏差较大，因此选用全选主元高斯消去法，它具有抗干扰能力强、算法稳定性较好等优点，其求解过程描述如下。一般情况下，要求取的曲线方程可转变为求取线性方程组，如表达式(1)所示：

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \quad (1)$$

用向量表示上述表达式，即： $Ax = b$ ，其中

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \text{ 满足约束}$$

$$A \in R^{n \times n}, x, b \in R^n$$

全选主元高斯消去法的目标就是将上式转化为如表达式(2)所示的形式：

$$\begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{(n)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(2)} \\ \vdots \\ b_n^{(n)} \end{pmatrix} \quad (2)$$

因此得到方程组的增广矩阵如式(3)所示：

$$B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j1} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2j1} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj1} & \cdots & a_{nn} & b_n \end{pmatrix} \quad (3)$$

并满足约束 $B \in R^{n \times (n+1)}$

其求解过程描述为：首先在原始矩阵 A 中选取绝对值最大的元素作为主元素，根据选取的主元素就可以对其 进行归一化处理，接着交换增广矩阵 B 的

第一行与第 i_1 行，经过第一次消元计算可以得：

$B \rightarrow B^{(2)}$ ，依次类推，直到第 $k-1$ 次选取主元完成，则此时的增广矩阵 B 转化为如式(4)所示：

$$(A^{(k)} \quad b^{(k)}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} & b_1 \\ & a_{22} & \cdots & \cdots & a_{2n} & b_2 \\ & & \ddots & & \vdots & \vdots \\ & & & a_{kk} & \cdots & a_{kn} & b_k \\ & & & \vdots & & \vdots & \vdots \\ & & & & a_{nk} & \cdots & a_{nn} & b_n \end{pmatrix} \quad (4)$$

最后将原方程转化为如式(5)所示：

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad (5)$$

其中 y_1, y_2, \dots, y_n 的次序为未知 x_1, x_2, \dots, x_n 调换后的次序，回带求解得最终结果，如式(6)所示：

$$\begin{cases} y_n = b_n / a_{nn} \\ y_i = (b_i - \sum_{j=i+1}^n a_{ij}y_j) / a_{ii} (i = n-1, \dots, 1) \end{cases} \quad (6)$$

3 卡尔曼滤波

卡尔曼滤波是线性无偏最小方差准则下的最优估计，由于其求解的过程是不断迭代的，因此非常适

合于对含有噪声的数据进行处理。它的基本出发点就是要获取系统的状态方程和观测方程。由于应用的场景较为特殊，固态发酵模型受多因素影响。为了获取发酵温度与时间的关系，其状态方程由全选主元高斯消去法拟合获得。卡尔曼滤波的实现过程即是依据存在的动态噪声和测量噪声的大小(动态噪声和测量噪声默认都是高斯白噪声)，每一时刻都要根据上一时刻的系统均方误差最小的准则动态地调整滤波增益系数，最后对上述各变量进行更新。其(离散的)状态空间模型描述的动态方程如：

$$X(k+1) = \phi X(k) + \tau W(k) \quad (7)$$

$$Y(k) = HX(k) + V(k) \quad (8)$$

式中，k 为离散时间，系统在时刻 k 的状态为

$X(k)$ ， $Y(k)$ 为对应状态的观测信号， $W(k)$ 为输入的高斯白噪声， $V(k)$ 为观测噪声。称式(1)为状态方程，式(2)为观测方程。 ϕ 称为状态转移矩阵， Γ 为噪声驱动矩阵， H 为观测矩阵。

说明的是，上述方程中涉及到的参数 X 为卡尔曼估计值， Y 为实际观测值。 ϕ 、 Γ 和 H 均为和系统相关的模型参数，所以当系统模型足够准确，卡尔曼滤波才能正常工作；卡尔曼滤波认为观测值和真实值有误差，这个误差的来源有两部分，即过程噪声

W 和观测噪声 V。什么是过程噪声呢？过程噪声就是外界对于系统的干扰，例如：为温度测量系统中，过程噪声是由于人体干扰、阳光照射、风等因素造成的；测量噪声与传感器测量精度息息相关，即由传感器的测量误差引起。测量噪声和过程噪声的特性均可由统计得到。基本假设：

假设 1: W 和 V 是均值为零，方差各为 Q 和 R 的不相关白噪声；

假设 2: 初始状态 $X(0)$ 不相关于 W 和 V

总结：卡尔曼滤波问题可归结为：基于观测信号 $\{Y(1), Y(2), Y(3), \dots, Y(k)\}$ ，求状态的线性方差估计值其算法设计如下：

状态的一步预测： $\hat{X}(k+1|k) = \phi \hat{X}(k|k)$ (9)

状态更新：

$$\begin{aligned} \hat{X}(k+1|k+1) &= \hat{X}(k+1|k) + \hat{X}(k+1)\varepsilon(k+1) \\ \varepsilon(k+1) &= Y(k+1) - H\hat{X}(k+1|k) \end{aligned} \quad (10)$$

滤波增益矩阵：

$$K(k+1) = P(k+1|k)H^T[HP(k+1|k)H^T + R^T]^{-1} \quad (11)$$

一步预测协方差阵：

$$P(k+1|k) = \phi P(k|k)\phi^T + \Gamma Q \Gamma^T \quad (12)$$

协方差阵更新：

$$P(k+1|k+1) = [I_n - K(k+1)H]P(k+1|k) \quad (13)$$

其中， $X(0|0) = \mu_0$, $P(0|0) = P_0$ ，R 是观察噪声方差，Q 是过程噪声方差。

4 实验结果分析

图 1 是两个发酵周期的原始温度数据，

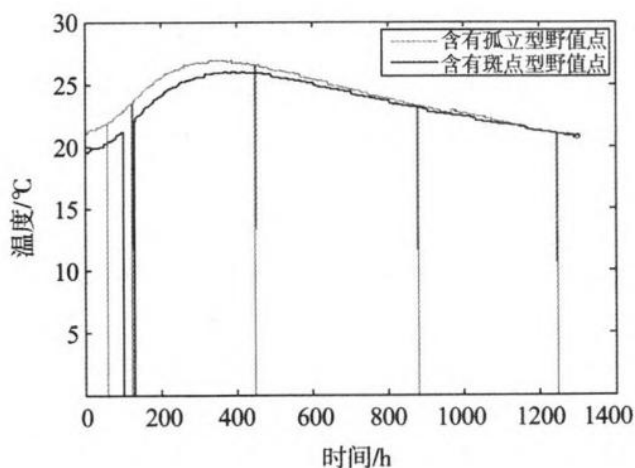


图 1 原始发酵温度数据曲线图

由图可知，一条曲线含有一定数量的孤立型野值点，另一条曲线含有斑点型野值点，这是在接收到的发酵数据中经常会出现的两种情况。图 3 是主元高斯消去法使用的次幂与其拟合值与实际值的均方误差的对应关系，通过图可以看出，当多项

式的次幂大于 5 以后，均方误差几乎都小于 1，依据计算方便快捷原则，所以接下来的实验中，全选主元高斯消去法中多项式的最高次幂都选取 6。接下来实验分为两个部分进行：①去除孤立型野值点；②去除斑点型野值点。图 4 显示了含有孤立型野值的原始数据曲线、卡尔曼滤波曲线、基于新息的卡尔曼滤波曲线和本文算法实现的滤波曲线，由图可以看出，单独使用卡尔曼滤波算法进行去噪误差较大，由图 5 可以看出，基于新息的卡尔曼滤波算法和本文算法误差较小。通过表 1 所示，本文算法比卡尔曼滤波平均误差减少 0.07996，比基于新息的平均误差减少 0.00022。图 6 分别显示了含有斑点型野值的原始曲线、卡尔曼滤波曲线、基于新息的卡尔曼滤波曲线和本文先进行局部线性曲线代替野值，然后进行拟合和滤波的曲线图。由图可知，基于新息和本文算法能够较好地去除斑点型野值。通过图 7 可以看出，在野值处，卡尔曼滤波产生较大偏差，基于新息和本文算法产生的偏差较小。表 2 表明在含有斑点型野值时，本文算法比卡尔曼滤波的平均误差减少 0.41392，比基于新息的平均误差减少 0.00253。

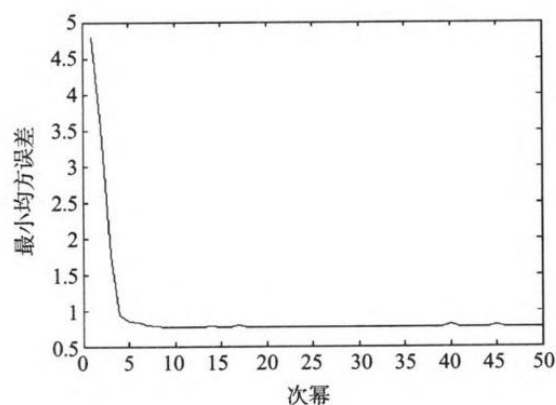


图 2 最高次幂与最小均方误差的对应关系

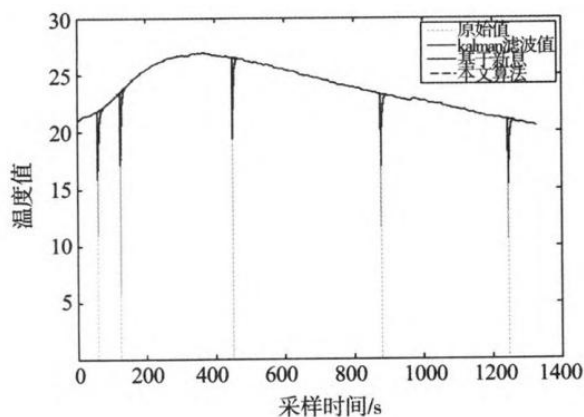


图 3 含有孤立型野值点各算法滤波对比

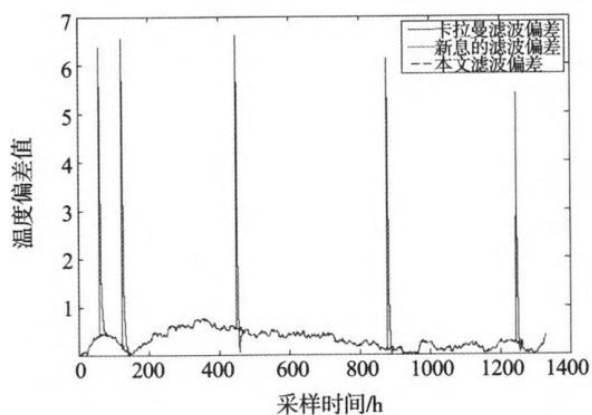


图4 含有孤立型野值点各算法去噪误差对比

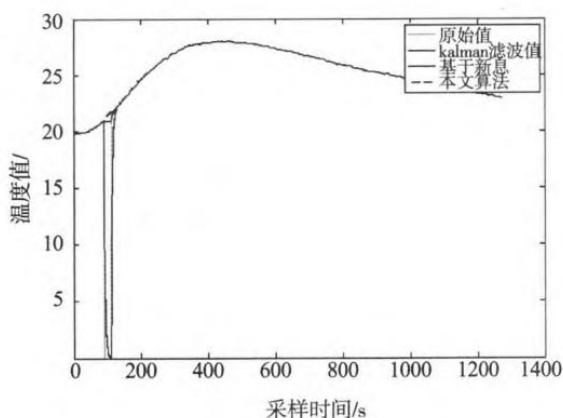


图5 含有斑点型野值点各算法滤波对比

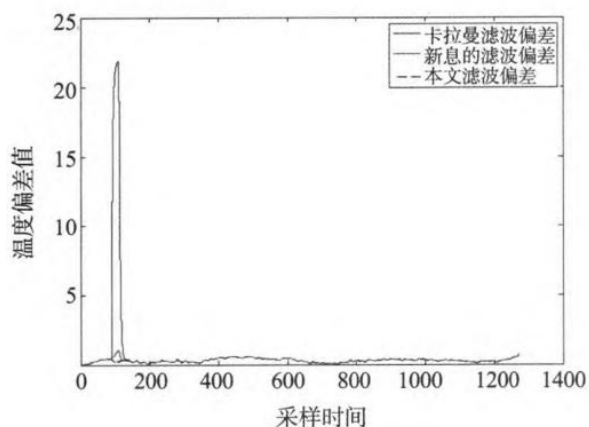


图6 含有斑点型野值点各算法误差对比

5 结束语

在实际的白酒发酵过程中,WSN中的温度数据中出现野值的现象经常发生。本文提出的全选主元高斯消去法联合卡尔曼滤波算法能够有效地去除野值,并且获得更贴近真实的有效值,保证发酵温度数据与时间之间变化的准确性,在实际的工程实践中有很大的应用价值。

参考文献:

- [1] 余龙江. 发酵工程原理与技术[M].北京:高等教育出版社,2016. (Yu Long-Jiang, Principle and Technology of Fermentation Engineering[M]. Beijing: Higher Education Press, 2016,3-4.)
- [2] Aleixandretudo J L, Nieuwoudt H, Aleixandre J L, et al. Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using different infrared spectroscopy techniques.[J]. Talanta, 2018, 176:526-536.
- [3] Vagge A, Capris P, Traverso C E. Choroidal Abnormalities Detected by Near-Infrared Imaging (NIR) in Pediatric Patients With Neurofibromatosis Type 1 (NF1) [J]. Investigative Ophthalmology & Visual Science, 2016, 57(3): 774.
- [4] Xue Z, Zeng S, Hao J. Non-invasive through-skull brain vascular imaging and small tumor diagnosis based on NIR-II emissive lanthanide nanoprobe beyond 1500 nm [J]. Biomaterials, 2018, 171: 153-163.
- [5] 于霜,刘国海,夏荣盛,江辉. 基于 Adaboost 及谱回归判别分析的近红外光谱固态发酵过程状态识别[J]. 光谱学与光谱分析,2016,36(01):51-54. (Yu Shuang, Liu Guo-Hai, Xia Rong-Sheng, Jiang Hui. State Recognition of Near Infrared Spectrum Solid State Fermentation Process Based on Adaboost and Spectral Regression Discriminant Analysis[J]. Spectroscopy and Spectral Analysis, 2016, 36(01): 51-54.)
- [6] Wang L, Liu T, Chen Y, et al. [On-line monitoring of biomass in 1,3-propanediol fermentation by Fourier-transformed near-infrared spectra analysis][J]. Chinese Journal of Biotechnology, 2017, 33(1):68.
- [7] Oren A, Chefetz B. Sorptive and Desorptive Fractionation of Dissolved Organic Matter by Mineral Soil Matrices[J]. Journal of Environmental Quality, 2012, 41(2):526-533.
- [8] 白雁, 宋瑞丽, 陈志红等. NIR 结合 OPUS 软件建立山药中尿囊素定量模型[J]. 中国医院药学杂志, 2008, 28(22):1945-1948. (Bai Yan, Song Rui-li, Chen Zhi-hong, et al. Establishment of quantitative model of allantoin in yam by NIR and OPUS software[J]. Chinese Journal of Hospital Pharmacy, 2008, 28(22):1945-1948.)
- [9] Li T, He X. Quantitative Analysis of Salidroside and p-Tyrosol in the Traditional Tibetan Medicine Rhodiola crenulata by Fourier Transform Near-Infrared Spectroscopy.[J]. Chemical & Pharmaceutical Bulletin, 2016, 64(4):289-296.
- [10] Li H D , Xu Q S , Liang Y Z . Random frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification[J]. Analytica Chimica Acta, 2012, 740:20-26.
- [11] Li H D , Xu Q S , Liang Y Z . libPLS: An integrated library for partial least squares regression and linear discriminant analysis[J]. Chemometrics & Intelligent Laboratory Systems, 2018, 176:34-43.
- [12] 任东, 瞿芳芳, 陆安祥等. 近红外光谱分析技术与应用 [M]. 北京: 科学出版社, 2016. 4-6. (Ren dong, Qu Fang-fang, Lu An-xiang, et al. Near Infrared Spectroscopy Analysis Technology and Its Application[M]. Beijing: Science Press, 2016. 4-6.)
- [13] 褚小立, 陆婉珍. 近五年我国近红外光谱分析技术研究与应用进展 [J]. 光谱学与光谱分析, 2014, 34(10): 2595-2605. (Chu Xiao-li, Lu Wan-zhen. Research and Application Progress of Near Infrared Spectroscopy in China in Recent Five Years[J]. Spectroscopy and Spectral Analysis, 2014, 34(10): 2595-2605.)

- [14] 刘红卫.微信小程序应用探析[J].无线互联科技, 2016(23): 11-12+40. (Liu Hong-wei. Analysis on the Application of Wechat Small Program[J]. Wireless Internet Technology, 2016(23): 11-12+40.)
- [15] 薛陇彬,刘钊远.基于 WebSocket 的网络实时通信[J]. 计算机与数字工程, 2014, 42(03): 478-481. (Xue Long-bin , Liu Zhao-yuan. Real-time Network Communication Based on WebSocket[J]. Computer&Digital Engineering, 2014, 42(03): 478-481.)