

3.1 引言

机载参数飞行记录系统采集的飞行数据存在噪声干扰、数据缺失等问题,如果不经数据预处理,难以满足飞行数据智能分析的要求。飞行数据预处理是后续如飞行动作识别、飞行状态监控等数据分析工作的基础。由于飞行数据来源于发动机部件布置的多个传感器,存在传感器不稳定和网络传输延迟等问题,数据中一般会存在少量的噪声和空缺值,而且机载参数飞行记录系统采集的飞行数据具有多源性和冗余性的特点,所以,数据智能分析前,先要对数据进行预处理。

3.2 飞行数据滤波

目前的飞行数据滤波算法通常是针对单一飞行参数的滤波方法,如小波变换、人工神经网络等,这些方法缺乏普适性,如果参数发生改变,需要重新建模。对多个传感器采集到的同一参数的测量值,运用合理的融合滤波算法通常会得到更加可靠的数据。

飞机上采集记录飞行参数数据的设备较多,测量手段丰富。因此,飞行数据具有多源性和冗余性的特点,这就为应用多源数据融合技术处理飞行数据提供了信源保证。一方面,航空飞行器上常常有多套飞行数据采集装置,如惯导系统、航姿系统等,所以不同的数据记录

装置可以在同一时刻对于同一个参数提供多个数据记录值;另一方面,同一任务记录装置的不同信源针对同一参数的观测值不同,比如惯导系统提供的飞机俯仰角和航姿系统提供的飞机俯仰角,气压计测量的飞机高度和无线电高度等。

目前数据融合算法的分类方法很多,按照是否依赖先验知识可以分为两类:一类是依赖先验知识的数据融合算法,如贝叶斯方法、卡尔曼滤波方法等,这些方法的融合效果要依赖于提供的先验知识,但是在实际工程应用中,先验知识的获取往往是困难的;另一类是不依赖于先验知识的方法,如基于支持度一致性的融合算法、基于关系矩阵的融合算法、基于参数之间相互距离的融合算法等。这些算法可以在难以获取先验知识的场合下应用。但是随着数据量的增加,一方面算法的运算效率减慢,另一方面,由于旧数据的增多,递推算法失去修正能力。本节选用基于改进支持度的记忆融合滤波算法对单一飞行参数滤波。既保证了滤波的效果,也能提高算法的运算效率。

3.2.1 基于斜率距离的支持度

支持度是衡量多源系统中两个信号源在某一时刻之间的相似程度。在由 n 个信号源观测的某个参数 X 的多源系统中,其在 k 时刻第 i 个信号源的观测值为 $z_i(k)$,根据飞行数据统一误差模型 $z_i(k) = X(k) + V(k)$,其中, $V(k)$ 表示参数第 k 时刻的所有误差之和, $E[V(k)]$ 和 $D[V(k)]$ 等分布情况均未知。此时可以采用数据融合的方法估计真实数据,即运用支持度来表示数据源观测值的可信程度,给每个数据源分配恰当的权重,进而融合各个数据源的观测值,消除数据的误差,得到飞行参数 X 真实值的估计值。

数据源的相似程度可以用数据源之间的差异性来衡量,其差异程度采用时间序列的绝对距离计算。在 k 时刻,数据源 i 和数据源 j 的绝对距离为 $d_{ij}(k)$,表达式为

$$d_{ij}(k) = |z_i(k) - z_j(k)| \quad (3-1)$$

为了便于后续运算,统一量纲,对绝对距离 $d_{ij}(k)$ 归一化,记作 $d'_{ij}(k)$,表达式为

$$d'_{ij}(k) = \frac{|z_i(k) - z_j(k)|}{\max_{i,j}(|z_i(k) - z_j(k)|)} \quad (3-2)$$

运用常用的指数衰减函数来量化两个数据源之间的观测支持度存在两个问题:一是支持度函数含有人为设定的参数,包含一定的主观成分;二是未考虑同一数据源在观测区间采集数据的可信度。为了更可靠衡量两个数据源之间的支持度,提出一种改进支持度函数,引入灰色接近度理论,定义数据源 i 的自支持度为

$$s_i(k) = \frac{1}{1 + \sigma_i(k)} \quad (3-3)$$

式(3-3)中, $\sigma_i(k)$ 为第 i 个信号源的观测值 $z_i(k)$ 的方差,所以 k 时刻数据源 i 和数据源 j 的观测支持度为 $a_{ij}(k)$,即:

$$a_{ij}(k) = \exp[-ad'_{ij}(k)^2] \quad (3-4)$$

$$\alpha = \sqrt{s_i(k)s_j(k)} \quad (3-5)$$

由式(3-4)可知,如果 $z_i(k)$ 与 $z_j(k)$ 相差较大,则 $a_{ij}(k)$ 较小,数据源 i 和数据源 j 之间的相互支持度较低;如果 $z_i(k)$ 与 $z_j(k)$ 相差较小,则 $a_{ij}(k)$ 较大,数据源 i 和数据源 j 之间的相互支持度较高。式(3-5)中, α 为支持度系数,反映支持度和数据源相差大小的关系。

应用式(3-5)可以得到在一定时间段内,参数各个数据源 X 支持度矩阵为

$$SD(k) = \begin{bmatrix} 1 & a_{12}(k) & \cdots & a_{1n}(k) \\ a_{21}(k) & 1 & \cdots & a_{2n}(k) \\ \vdots & \vdots & & \vdots \\ a_{n1}(k) & a_{n2}(k) & \cdots & 1 \end{bmatrix} \quad (3-6)$$

支持度矩阵每 i 行元素之和 $\sum_{j=1}^n a_{ij}(k)$,反映第 i 个数据源和其他数据源的一致程度,若 $\sum_{j=1}^n a_{ij}(k)$ 大,表明第 i 个数据源和大多数数据源是一致的,反之,则表示与大多数数据源存在偏差。

定义 k 时刻数据源 i 观测值与其他数据源观测值的一致性度量 $r_i(k)$,表达式为

$$r_i(k) = \frac{\sum_{j=1, j \neq i}^n a_{ij}(k)}{n-1} \quad (3-7)$$

根据支持度的定义,通常选择在整个时段观测一致性度量 $r_i(k)$ 较大的数据源作为可靠的数据。但是数据源的观测一致性在不同时刻的变化较大,很有可能出现在某一时刻观测一致性较大,在其他时刻观测一致性较小的情况。所以,仅比较单一时刻的观测一致性无法确定哪个数据源更为准确和可靠,要衡量整个时刻数据源观测一致性的变化规律,运用统计理论,通过比较信号源 i 的观测一致性均值和观测一致性方差,来确定数据融合的权重。

k 时刻数据源 i 观测一致性均值为

$$\overline{r_i(k)} = \frac{1}{k} \sum_{t=1}^k r_i(t) \quad (3-8)$$

k 时刻数据源 i 观测一致性方差为

$$\sigma_i^2(k) = \frac{1}{k} \sum_{t=1}^k [\overline{r_i(k)} - r_i(k)]^2 \quad (3-9)$$

定义 k 时刻数据源 i 的加权系数为 $q_i(k)$,在应用过程中,观测一致性均值大,观测一致性方差小的数据源往往被认为更加可靠,所以观测一致性均值大,观测一致性方差小的数据源的权重应该较大,即加权系数 $q_i(k)$ 与数据源 i 观测一致性均值 $\overline{r_i(k)}$ 成正比,与观测一致性方差 $\sigma_i^2(k)$ 成反比。因此,加权系数 $q_i(k)$ 的表达式为

$$q_i(k) = [1 - \lambda \sigma_i^2(k)] \overline{r_i(k)} \quad (3-10)$$

式中, λ 为可调参数, 可以调节方差对于权重大小的影响, 所以基于支持度的融合估计表达式为

$$\hat{X}(k) = \frac{\sum_{i=1}^n q_i(k) z_i(k)}{\sum_{i=1}^n q_i(k)} \quad (3-11)$$

3.2.2 剔除异常支持度

即使考虑数据源的稳定程度, 若想获得更为精准的融合结果, 还需要在进行数据融合之前, 剔除异常支持度对融合结果的影响。选用格拉布斯(Grubbs)法剔除支持度中的异常值。

k 时刻数据源 i 的支持度的均值和方差分别为

$$\bar{a}_i(k) = \frac{1}{n} \sum_{i=1}^n a_{ij}(k) \quad (3-12)$$

$$\sigma'_i(k) = \sqrt{\frac{1}{n} \sum_{j=1}^n [\bar{a}_{ij}(k) - a_{ij}(t)]^2} \quad (3-13)$$

$a_{ij}(k)$ 对应的 Grubbs 统计量为

$$T_{ij} = \frac{\|a_{ij} - \bar{r}_i(k)\|}{\sigma_i(k)} \quad (3-14)$$

若 $T_{ij}(k) \geq T(N, a)$, 则令 k 时刻数据源 i 的支持度 $a_{ij}(k)$ 为 0。

为了防止数据饱和问题, 引入遗忘函数, 时间间隔越大的数据, 其影响越小; 反之, 时间间隔越小的数据, 其影响越大。由遗忘函数的定义可知, 任何单调非增函数都可以是遗忘函数。

3.2.3 遗忘函数

根据指数遗忘函数的特点构造遗忘函数为

$$A_i(k) = e^{-at} \quad (3-15)$$

式中, t 为当前时间差与过去某个时点的时间差, 构造的遗忘函数具有形式简单、工程上易于应用的特点。

引入遗忘函数之后, 此时的融合估计表达式为

$$\hat{X}(k) = \frac{\sum_{i=1}^n w_i(k) z_i(k)}{\sum_{i=1}^n w_i(k)} \quad (3-16)$$

式中, $w_i(k)$ 为记忆融合值的权值, 表达式为

$$w_i(k) = \frac{1}{k} \sum_{u=1}^k r_i(u) A_i(u) \quad (3-17)$$

3.2.4 实验与结果分析

选取某型飞机发动机转速为实验样本, 分别截取飞行参数系统的主记录器、快速记录器和事故记录器中关于发动机转速的数据, 共 400 组, 分别记作数据源 A、数据源 B 和数据源 C, 真实值为 6000r/min, 数据源的部分观测数据如表 3.1 所示。

表 3.1 数据源观测数据

观测次数	数据源 A	数据源 B	数据源 C	观测次数	数据源 A	数据源 B	数据源 C
1	218.7	1001.1	6003.6	11	6007.9	6002.1	6004.9
2	6009.1	5352.2	6005.3	12	5989.2	6003.1	6004.3
3	5990.5	5401.5	6005.0	13	6010.0	6003.4	6005.0
4	5989.7	6003.1	6000.8	14	6010.0	5999.0	6001.2
5	5992.1	5900.8	5996.3	15	6008.3	6001.5	5995.9
6	5992.5	5998.9	5997.0	16	5989.3	5998.6	5996.5
7	6008.5	5998.3	5996.8	17	6008.5	6002.1	5996.7
8	6010.0	6001.5	6005.2	18	5989.2	6001.7	6003.2
9	6005.2	6000.4	6003.2	19	6005.1	6006.5	6001.5
10	6003.2	6099.5	6000.4	20	6001.5	6004.3	6000.3

采用本算法, 与数据源比较如图 3.1 所示, 可以看出应用本算法可降低数据源本身存在的噪声, 提高了数据源的精度。下面将本算法与其他算法比较, 如图 3.2 所示。

从图 3.2 可以看出与平均值法和支持度融合算法相比, 本算法拥有较好的适应性, 滤波结果更接近实际值, 数据更加平稳, 总体误差小于其他两种方法, 极大提高了飞行数据的精度。

对于机载参数飞行记录系统采集的飞行数据原值存在噪声干扰问题, 根据飞行数据具有多源性和冗余性的特点, 提出在已有的支持度算法上的改进支持度记忆融合滤波算法, 通过对改进的支持度计算和添加遗忘因子, 使得算法的效率增加, 提高了数据的精度。

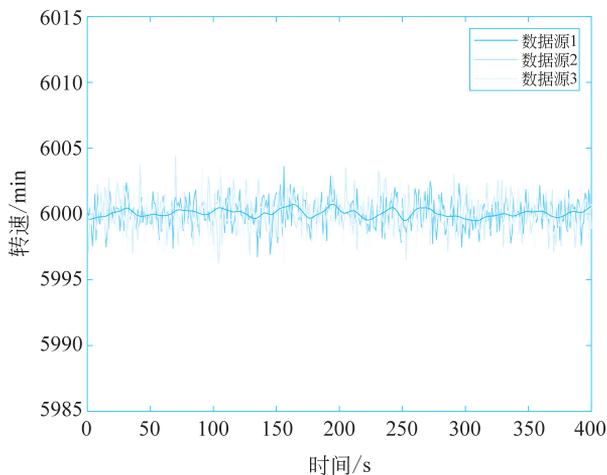


图 3.1 融合结果与数据源输入值比较

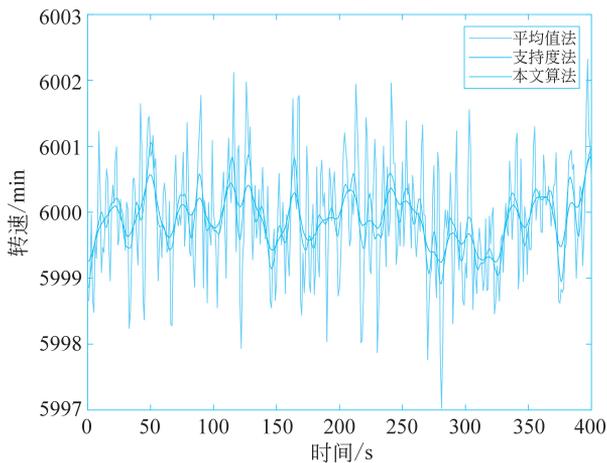


图 3.2 3种算法比较

3.3 飞行数据归一化处理

如航空飞行器飞行经度、高度、纬度等多个状态参数的数值范围不同,参数的数据量级不同可能导致训练的网络模型不准确,所以要对多个状态参数进行归一化。常见的3种数据归一化方法为 Z-score、Min-max 和 MaxAbs 标准化方法,3种方法的计算方法和适用范围如表 3.2 所示。

表 3.2 数据标准化方法对比

名称	计算方法	适用范围
Z-score	$x' = \frac{x - x_{\text{mean}}}{x_{\text{std}}}$ (状态参数的测量值减去其均值,再除以其标准差)	适用于状态参数测量值的最大值和最小值无法判断的情况,或者是测量值存在过多的异常数据的情况
Min-max	$x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$ (状态参数的测量值减去其最小值,得到的差再除以最大值与最小值之间的差)	适用于状态参数测量值的最大值最小值都已知的情况
MaxAbs	$x' = \frac{x}{x_{\text{max}}}$ (状态参数的原始数据除以其最大值)	适用于已经零中心化的状态数据或稀疏数据

3.4 本章小结

本章首先针对飞行数据采集过程中传感器测量噪声变化导致的数据精度下降问题,提出一种基于改进支持度的记忆融合滤波算法,通过改进支持度计算各个数据源的权重系数,引入遗忘函数减少历史数据的影响,通过与典型算法仿真对比,验证结果表明该算法有效地提高了数据的精度和稳定性;然后,针对数据量纲不一致的问题,阐述了飞行数据归一化处理的方法,为实现飞行数据的智能化分析处理提供可靠信源保障。