

大型目标红外遥测数据相关性分析与补全方法

蔡 伟, 蓝宋江, 聂 超

火箭军工程大学导弹工程学院, 陕西 西安

Email: 807442843@qq.com

收稿日期: 2021年3月21日; 录用日期: 2021年4月5日; 发布日期: 2021年4月20日

摘 要

针对大型目标红外特性测试中因采用超低功耗无线遥测方案和信道劣化而导致部分测试节点的温度数据缺失问题, 在考虑大型目标红外特征分布的基础上, 通过分析温度数据的时间相关性和空间相关性, 提出了基于单节点时间相关、多节点空间相关、环境相似条件趋势相关的数据补全算法, 实现对多类缺失数据的有效补全, 该算法同样适用于对测试数据中粗大误差的剔除与补全。为了验证算法有效性, 通过对完整实测数据进行随机挖孔生成测试样本, 将补全数据与实测数据进行比对。结果表明, 补全数据的各项统计指标及与实测数据的曲线拟合度均达到较优的性能, 为遥测数据的分析补全提供了一种行之有效的方案。

关键词

红外遥测, 相关性分析, 数据补全

Completion and Correlation Analysis of Infrared Telemetry Data for Large Targets

Wei Cai, Songjiang Lan, Chao Nie

College of Missile Engineering, Rocket Force University of Engineering, Xi'an Shaanxi

Email: 807442843@qq.com

Received: Mar. 21st, 2021; accepted: Apr. 5th, 2021; published: Apr. 20th, 2021

Abstract

In terms of the loss of temperature data in some test nodes caused by adopting ultra-low power wireless telemetry and the degradation of channel, this paper puts forward the data completion algorithm related to single node time, multiple nodes space and environmental similarity condi-

tions trend by analyzing the correlation between the time and space of temperature data on the basis of considering the infrared characteristic distribution of large targets. It effectively completes many types of data. This algorithm also applies to gross error in the data of the test strip and completion. In order to verify the effectiveness of the algorithm, the test samples were generated by random digging of the complete measured data, and the completed data were compared with the measured data. The results show that the statistical indexes of the data and the curve fitting degree of the measured data all achieve better performance, which provides an effective scheme for the analysis and completion of telemetry data.

Keywords

Infrared Telemetry, Correlation Analysis, Completion

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

大型目标红外特性测试是目标红外隐身性能分析与优化的关键环节，通常采用超低功耗无线遥测方案，数据传输为非应答模式，即安装在目标表面不同位置的温度测试节点将测试的实时数据通过无线方式发送后，无需等待接收端反馈即可关闭无线通信功能进入超低功耗模式，从而大幅降低功耗，在不更换电池条件下可显著延长节点寿命，但非应答模式也存在因信道质量导致部分测试数据缺失的问题，对后续的红外特征分析建模造成不利影响。

因此，缺失数据的处理是一个不可缺少的数据预处理过程，目前一般分为两类，一类是直接删除含有缺失值的数据点，这种方法简单易操作，但缺点是在缺失比例较高时，该方法会造成信息的大量流失从而降低有效性。另一类是缺失值填补方法，用估计值来代替缺失值[1]。一般地，缺失值填补分为基于统计技术和基于机器学习技术的，机器学习技术包括：决策树、贝叶斯法、期望值最大化、回归法等。广泛应用的统计技术包括：均值或众数填补和回归法。其中大多数方法是利用已有的正常数据作为样本，从而建立预测模型，以估计缺失数据[2]。然而，一旦预测值不精确，将导致数据重建的精度受到累积误差的限制。

针对大型目标红外遥测数据缺失问题，本文提出了基于单节点时间相关、多节点空间相关、环境相似条件趋势相关的数据补全算法，实现对多类缺失数据的有效补全，该算法同样适用于对测试数据中粗大误差的剔除与补全。在采样长度不变的基础上，对测试数据进行缺失项有效补全。

2. 研究对象与数据前期处理

在一个典型应用中，大型目标表面不同位置布设有 100 个温度测试节点，各节点以 10 分钟为采样间隔，连续测 10 天，共形成 100 组温度数据，以一定的时间差传给两个不同的电脑数据处理系统，第一组温度数据汇总样本用 TEMP1 命名，第二组温度数据汇总样本用 TEMP2 命名，最终数据汇总样本用 TEMP 命名。

鉴于测试时间段连续，对 TEMP1、TEMP2 数据进行前期准备处理，主要分为以下三个步骤：首先，利用 MATLAB 软件导入两组数据形成两组矩阵，删除其重复时间段所在行温度数据；其次，扩增矩阵 TEMP1 和矩阵 TEMP2，使其维度相同，确保两组矩阵第一列时间点相同；最后，进行缺失值互补，以

矩阵 TEMP1 中时间点为基准, 查找所在行缺失项元素, 调用矩阵 TEMP2 对应元素进行填充, 用 TEMP 命名形成的最终数据样本。

3. 温度数据的相关性分析

3.1. 单节点温度数据时间相关

由于单节点测量温度具有时间持续久、周期性且间隔较短, 使得温度变化的形式不是突然、迅速或一瞬间完成的剧烈变化形式, 而是表现为连续、缓慢、逐渐的, 是一种平稳的变化形式。根据时间相近, 温度变化幅度不大的特点便于缺失数据补全。

3.2. 多节点温度数据空间相关

多节点温度数据空间相关判定需建立温度分布理论模型, 也就是目标表面温度计算模型, 从外界环境辐射和目标自身辐射两个方面分别计算节点表面温度。对于无内能源的目标, 当测点表面材质、方位角、俯仰角都相同, 它们具有热交换相似性, 即温度相似。利用该性质推断出空间相关的多节点, 分析测得数据之间差值规律并进行补全。

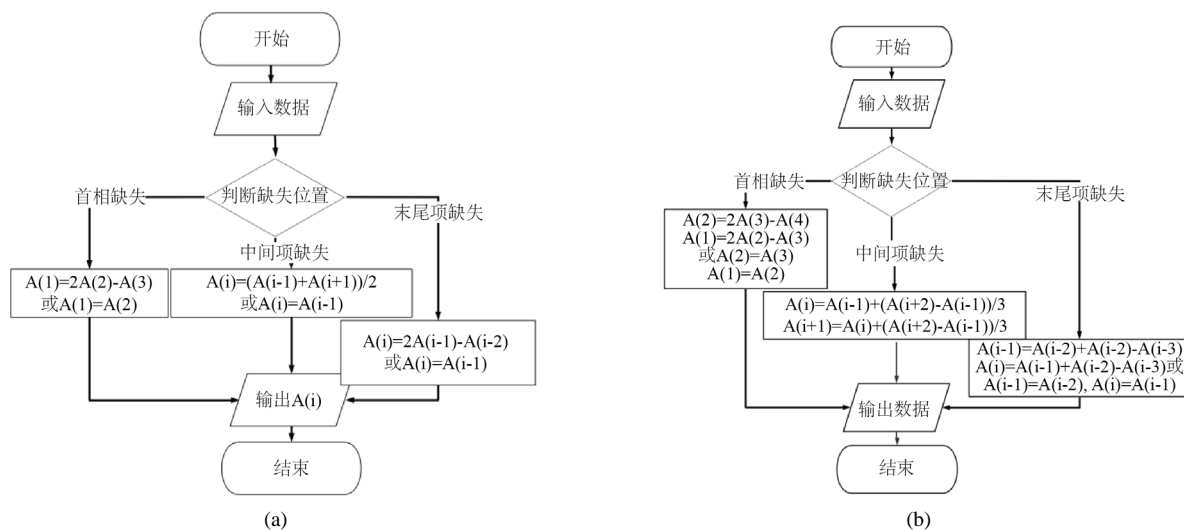
3.3. 环境相似条件温度变化趋势相关

同一节点在环境、日照等因素大致相同的情况下, 传感器在发生数据缺失的值和不同周期内的同一位置的数值很接近, 温度幅值不尽相同, 但温度变化趋势呈现相似性, 即是一种平稳的周期性变化。借鉴整体时间 - 温度曲线, 截取符合该规律的对应时间段完善缺失的温度数据。

4. 算法模型设计

通过对每组数据观察, 将缺失情况分为单值、双值、三值及三值以上缺失。每一种数据缺失状况可分为三类: 一、首相缺失; 二、中间项缺失; 三、末尾项缺失。运用热卡填充法(Hot deck imputation) [3]、内插填充法(interpolation method) [4]、多重插补法(multiple interpolation) [5]及相关性分析法(correlation analysis) [6]对缺失项补全。

给定数值缺失样本中共 N 列数组, 以第 k ($k \in [1, N]$) 列数组为例, A 为该数组集合, i 为缺值位置序号, A(i)为缺值。算法流程图如下(见图 1)。



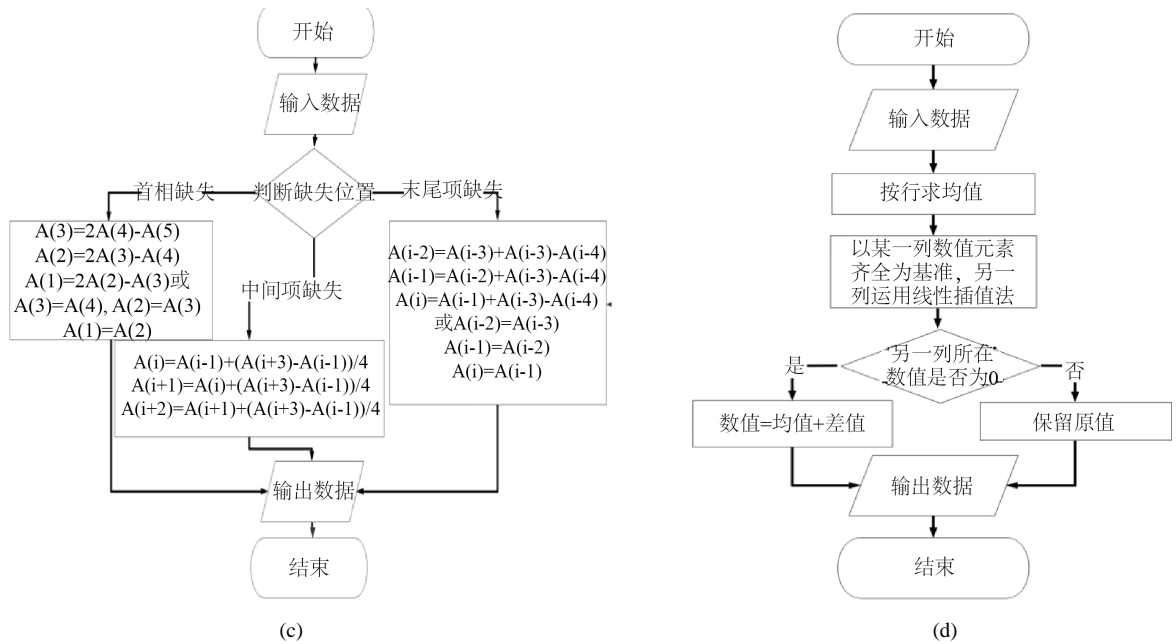


Figure 1. Data completion algorithms. (a) Completion of single value; (b) Completion of double values; (c) Completion of triple values; (d) Completion of multiple values

图 1. 缺失数据补全算法流程。(a) 单值补全；(b) 双值补全；(c) 三值补全；(d) 三值以上补全

5. 算法验证

5.1. 实验方法

从 TEMP 样本数据中选取一组含 500 个元素的完整数据，每个元素分别标记序号 1 至 500，通过 MATLAB 随机函数产生并截取 5 个单值缺失、5 个双值缺失、3 个三值缺失、3 个四值缺失数据段(见表 1)，导入 MATLAB 中运行算法，得出补全后新的数据表，分析对比处理前后数据误差，利用数据折线图对比，验证算法的准确性和可靠性。

Table 1. Types and sequence of lost values

表 1. 数据缺值类型及相应序号

缺值类型	缺值点序号
单值缺失	9、220、335、400、459
双值缺失	66 67、195 196、380 381、473 474、499 500
三值缺失	1 2 3、164 165 166、296 297 298
四值缺失	4 5 6 7、100 101 102 103、305 306 307 308、415 416 417 418

5.2. 数据误差分析

下面展示四值以下缺失补全情况，从数据误差分析角度，具体由方差、平均误差、最大误差、相对误差[7]体现(见表 2)，分析不同数据段不同位置缺失情况下，对比结果及相应温度折线对比图(见图 2)。

局部示意如图 3 所示。

从上述表中的数据误差分析对比及 MATLAB 折线对比图可以看出算法运行结果很好地补全了原样本缺失数据，与原样本拟合基本一致，误差较小，表明通过本文算法进行数据缺失补全是可行的。

Table 2. Completion and its error magnitude
表 2. 缺值补全及误差值示意

	方差	平均误差	最大误差	相对误差
单值补全	2.34	1.22	2.71	5.09%
双值补全	3.15	1.42	3.24	5.34%
三值补全	5.74	1.91	5.65	10.62%
四值补全	8.35	2.30	8.91	15.73%

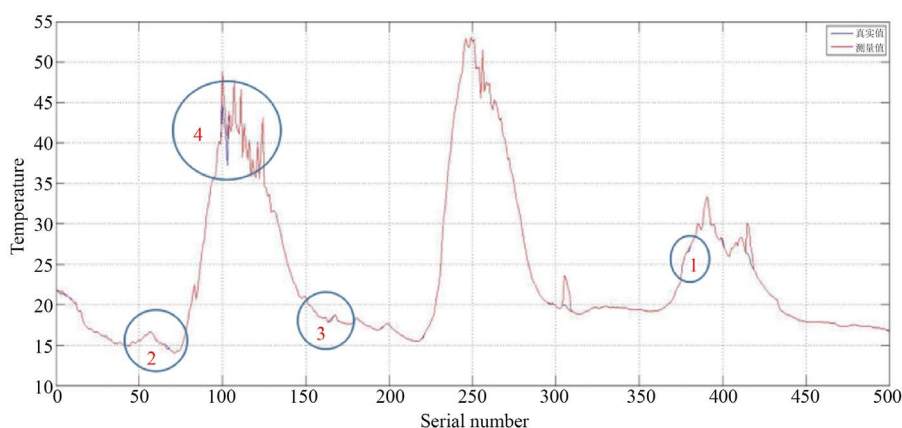


Figure 2. Comparison of broken lines before and after completion
图 2. 补全前后折线对比图

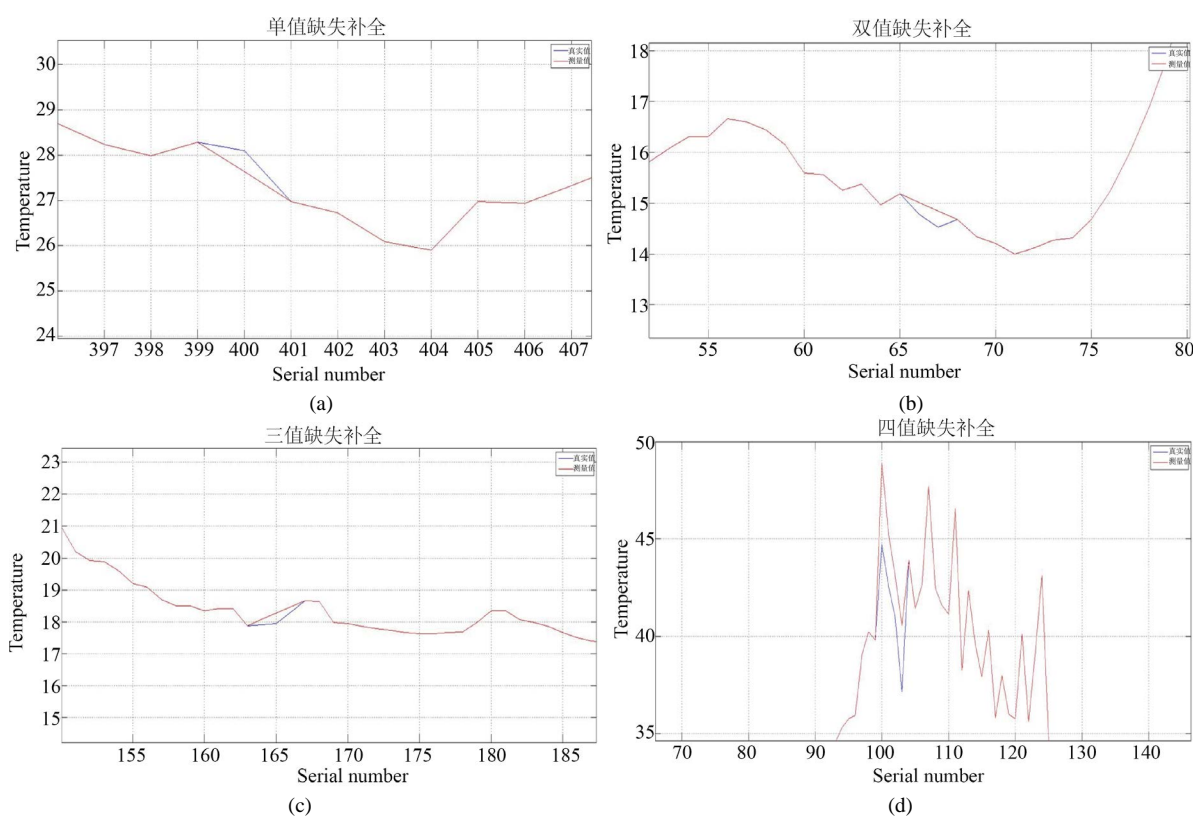


Figure 3. Partial enlarged detail
图 3. 局部放大图

6. 结论

数据集的完备化是数据挖掘过程中必须解决的首要问题。对缺失的数据进行补全,尽可能与原系统信息保持一致,才能确保数据的准确性和可靠性。本文所介绍的大型目标红外遥测数据分析与补全方法对于一般性数据补全也具有很好的借鉴意义,利用 MATLAB 仿真验证并进行对比分析,结果表明,补全数据的各项统计指标及与实测数据的曲线拟合度均达到较优的性能,表明该算法为大型目标红外测试数据分析补全的实现打下了坚实的基础。

参考文献

- [1] 刘莎, 杨有龙. 基于灰色关联分析的类中心缺失值填补方法[J]. 四川大学学报(自然科学版), 2020, 57(5): 53-60.
- [2] Xu, X., Zhang, Z., Chen, Y., *et al.* (2017) HMM-Based Predictive Model for Enhancing Data Quality in WSN. *International Journal of Computers and Applications*, **6**, 1-9.
- [3] 宋亮, 万建洲. 缺失数据插补方法的比较研究[J]. 统计与决策, 2020, 36(18): 10-14.
- [4] 帅亮乾. 直线内插法在工程热力学课程中的应用[J]. 大学教育, 2013(16): 84-85.
- [5] Libasin, Z. and UI-Saufie, A.Z., *et al.* (2020) Single and Multiple Imputation Method to Replace Missing Values in Air Pollution Datasets: A Review. *IOP Conference Series: Earth and Environmental Science*, **616**, 1-8.
<https://doi.org/10.1088/1755-1315/616/1/012002>
- [6] 李绍坚, 等. 基于多维度相关性分析的电压缺失数据辨识方法研究[J]. 电气自动化, 2021, 43(1): 63-66.
- [7] 吴石林, 等. 误差分析与数据处理[M]. 北京: 清华大学出版社, 2010.