

基于回归分析的航天器异常检测方法

史晓云^{✉1}, 陈 军¹, 郭小红¹, 符叶丹², 王新广¹

(1 航天器在轨故障诊断与维修重点实验室 西安 710043

2 宇航动力学国家重点实验室 西安 710043)

摘要: 针对某在轨航天器电源温控系统典型故障案例, 提出基于 Pearson 相关系数回归分析的航天器多元遥测数据异常检测方法。首先, 对相关遥测数据进行特征提取; 其次, 对正常状态下的遥测数据进行拟合建模, 并用该模型检测后续遥测数据; 最后, 采用基于 Pearson 相关系数的回归分析方法对实测数据进行仿真分析。结果表明, 针对航天器首次发生的异常, 该方法能从多维遥测数据中快速提取与异常相关的特征遥测参数, 可有效建立遥测参数预警模型及检测遥测数据的异变趋势, 并能够推广应用于其他航天器。

关键词: 多元数据; 回归分析; 遥测数据; 异常检测

中图分类号: V557+3 文献标识码: A 文章编号: CN11-1780(2022)02-0057-06

DOI: 10.12347/j.ycyk.20211107001

引用格式: 史晓云, 陈军, 郭小红, 等. 基于回归分析的航天器异常检测方法[J]. 遥测遥控, 2022, 43(2): 57-62.

Spacecraft anomaly detection method based on regression analysis

SHI Xiaoyun¹, CHEN Jun¹, GUO Xiaohong¹, FU Yedan², WANG Xinguang¹

(1. Key Laboratory of Spacecraft In-Orbit Fault Diagnosis and Maintenance, Xi'an 710043, China;

2. State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China)

Abstract: Aiming at a typical fault case of a spacecraft power supply temperature control system, an anomaly detection method based on Pearson correlation coefficient regression analysis is proposed. Firstly, the feature extraction of relevant telemetry data is carried out. Secondly, the normal telemetry data are fitted and modeled, and the subsequent telemetry data are detected by the model. Finally, the regression analysis method based on Pearson correlation coefficient is used to simulate the measured data. The results show that the method can quickly extract the abnormal phase from the multi-dimensional telemetry data for unknown spacecraft anomalies, it can be applied to other spacecraft.

Key words: Multivariate data; Regression analysis; Telemetry data; Anomaly detection

DOI: 10.12347/j.ycyk.20211107001

Citation: SHI Xiaoyun, CHEN Jun, GUO Xiaohong, et al. Spacecraft anomaly detection method based on regression analysis[J]. Journal of Telemetry, Tracking and Command, 2022, 43(2): 57-62.

引 言

航天器遥测数据能够实时动态反映航天器的运行规律, 但是其数据关系非常复杂且专业性强^[1], 故总体上呈现海量、多源、异构、时空多序列等特点, 隐含着大量有价值的信息没有得到充分的挖掘和应用。因此, 如何精准、快速定位航天器故障, 预测航天器状态和改进航天器服务, 已成为目前航天领域遥测分析中亟待解决的问题。

在轨航天器遥测数据一般按照缓慢变化、周期变化和突变进行分类, 总的来说, 它的变化呈现出趋势性、季节性、周期性以及随机性^[2]。在轨管理的实际过程中, 主要通过阈值法对航天器进行异常检测, 其优点是检测过程简单, 但是存在虚警、漏警现象, 时效性也差。现阶段, 在针对大量遥测数据的研究中, 主要基于时间序列模型, 将实际遥测值与模型输出值比对, 实现对遥测数据的异常检测。文献[3]

✉通信作者: 史晓云 (814706781@qq.com)

收稿日期: 2021-11-07 收修改稿日期: 2022-01-30

对航天器遥测数据的变化趋势进行分类, 研究了多种主流的数据预测理论并分析了其适应性, 总结提出一种基于历史遥测数据的在轨卫星故障预警系统构架; 文献[4]利用 M-VMD 方法对遥测数据进行分解, 提取各模态分量的熵特征, 通过计算遥测数据与样本数据对应的各模态分量的 JRD 距离, 实现在轨航天器异常状态识别。随着信息技术的发展, 基于专家知识和物理模型的故障诊断方法已经难以满足高实时性和高准确性的故障处置需求, 尤其是随着大数据时代的来临, 航天器系统的各种特征属性、运行机理以及一些无法解释的复杂机理, 都可以通过实时数据间的某种关联关系进行分析研究。

本文通过回归分析方法对某航天器电源温控系统相关实测数据进行尝试性研究, 通过特征提取和数据预处理, 对正常状态下的时序参数值进行建模, 并与异常时段的时序参数值进行对比。实验结果表明, 该算法以多元数据为背景可以有效检测遥测数据趋势, 可及时发现遥测时序数据潜藏的异常变化, 从中挖掘出各部件及器件缓慢变化的的异变信息^[1], 能有效提升在轨航天器的管理能力, 从而确保航天器在轨安全可靠运行。

1 回归分析方法

航天器参数之间互相关联、互相耦合, 它的数值和系统运行模式变化受多种遥测参数的共同影响。在实际问题中, 常常会遇到相互联系、相互制约的变量, 包括确定性和非确定性两种关系, 回归分析就是研究有一定的关系但又不能用确定的函数进行表达的统计方法, 它既可以探索和检验自变量和因变量的因果关系, 也能用自变量的取值变化来预测因变量的取值, 是一种定量处理多个变量相关关系的方法^[1]。在现实工作中, 回归分析的优点在于它可以通过统计手段对于干扰因素进行控制, 发现自变量和因变量的关系, 从而及时掌握航天器的当前状态。

1.1 特征选取及预处理

为了简化分析, 一般只保留权重影响因子较大的特征遥测参数。一方面可以简化预测过程, 另外对今后设计改进有着重要的参考价值。由于遥测数据受多种因素影响, 采用 3σ 标准差原则进行野值剔除。

1.2 回归模型

航天器遥测参数种类繁多, 找出参数间隐含的相关关系有利于在轨卫星管理人员根据航天器实时状态实施安全有效的管控^[5]。一般用皮尔逊系数对线性相关进行判别, 用互信息法对非线性相关进行判别, 对于强相关性的参数通过构建映射函数或模型进行预测^[6]。回归模型是一种统计分析方法, 因变量根据自变量进行“预测”, 从而使“预测值”以最大可能性接近因变量的“观测值”^[5]。很显然, 由于随机误差等因素, 回归方程中的预测值不可能和观测值完全相同, 回归分析的特点在于它把观测值分为结构和随机两部分^[5], 结构部分是因变量与自变量间的结构关系; 随机部分是未被结构项解释的剩余部分。

一元线性回归模型: 只有一个特征是输入变量, 被称作单变量线性回归问题^[5]。回归模型表示如下:

$$Y = \beta_0 + \beta_1 \times x + \xi \quad (1)$$

多元线性回归模型: 在航天器的实际工作中, 遥测因变量的变化随着重要因素的变化而变化, 此时就需要用两个或是更多的影响因素作为自变量来解释因变量的变化, 多元回归就是研究多个随机变量与一个可控变量间的相关关系, n 代表特征参数的数量, x_n 代表第 n 个特征的遥测参数的取值, 支持多变量的预测函数可表示为^[7]:

$$Y = \beta_0 + \beta_1 \times x_1 + \dots + \beta_i \times x_i + \dots + \beta_n \times x_n + \xi \quad (2)$$

式中, Y 是因变量, 代表随机观测值, 未知参数 $\beta_0, \beta_1, \dots, \beta_i, \dots, \beta_n$ 为偏回归系数, $\beta_i (i=1, 2, \dots, n)$ 则是在其他自变量固定不变的情况下, $x_i (i=1, 2, \dots, n)$ 每改变一个单位, 其单独引起因变量 Y 变化的平均改变量, ξ 为去除 k 个自变量对 Y 影响后的随机误差^[5]。

在多元线性回归方程中, 模型必须满足以下条件:

- ① 自变量对因变量有显著的影响, 并表示出非确定关系 (线性关系);
- ② 样本的容量 > 回归系数个数;
- ③ 残差服从正态分布。

多元线性回归模型建成之后,必须进行评价和检验,才能判定模型是否可以应用。模型的检验能够通过考察模型的绝对平均差、线性相关等指标及模型预测误差图,来综合评价模型的可应用性并比较模型的效果。

1.3 评价回归方程的参数

由上述方法确定了模型的相关参数后,需对其进行一定的评价,来确定模型的预测情况,一般常用的评价指标有:协方差矩阵、离差平方和、残差平方和、回归平方和、拟合优度检验^[5]。

① 有多个变量时,可建立归一化的协方差矩阵,用以表示多变量间的相关关系,并用于航天器参数异常检测^[5]。协方差:假设有随机变量 X 、 Y ,其协方差为:

$$\text{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\} \quad (3)$$

② 协方差矩阵:设 n 维随机变量 (x_1, x_2, \dots, x_n) 的二阶混合中心矩^[5]:

$$c_{ij} = \text{Cov}(x_i, x_j) = E\{[x_i - E[x_i]][x_j - E[x_j]]\} \quad (4)$$

$i, j = 1, 2, \dots, n$ 都存在,则称矩阵:

$$c = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix} \quad (5)$$

为 n 维随机变量 (x_1, x_2, \dots, x_n) 的协方差矩阵,显然有 $c_{ij} = c_{ji}$ 。

③ 总偏差平方和表征的是每个因变量的实际值与因变量平均值的差的平方和,即反映了因变量取值的总体波动情况,即被回归方程解释的因变量的总变异^[8]:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (6)$$

④ 回归平方和表征的是因变量的预测值对其平均值的差的平方和,它是由于自变量 x 的变化引起的 y 的变化,反映了 y 的总偏差中由于 x 与 y 之间的线性变化引起的 y 的变化部分,可以由回归直线解释^[8]:

$$SSR = \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 \quad (7)$$

⑤ 残差平方和又称误差平方和,表征的是因变量的各实际观测值对其回归值的差的平方和,它是除了 x 对 y 的线性影响之外的其他因素对 y 变化的作用,不能由回归直线来解释^[8]:

$$SSE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (8)$$

⑥ 拟合优度反映回归方程所解释的变异在因变量总变异中所占的比例,判断回归方程的拟合程度,表示拟合直线能多大程度反映因变量的波动^[8],变化范围为 $0 \sim 1$,决定系数越大,说明方程的预测准确性越高,原因变量对结果变量的解释程度越高,越接近 1 说明拟合效果越好。

$$R^2 = \frac{SSR}{SSE} \quad (9)$$

拟合完方程的参数后,就要对回归模型进行检验,在简单线性回归中用测定系数来验证方程的拟合程度,而在多元线性回归中,如果在模型中增加一个自变量,模型 R^2 往往会相应地增加^[9]。

2 仿真实验及结果分析

为验证本文提出的回归分析方法的正确性和可行性,选取某型号在轨运行航天器驱动机构滑环典型故障时的电源温控系统相关遥测参数进行实验验证,分别对正常状态下的时序曲线、异常发生后的时序曲线进行建模分析。

某在轨航天器运行过程中蓄电池充电电流减少了 $2/3$,卫星进入最小能源安全模式,有效载荷全部自主关机,航天器姿态稳定,其他分系统工作正常。本文针对某在轨航天器电源温控系统典型故障案例,绘出了在轨故障过程中的主要遥测参数曲线。

如图 1 所示,当航天器出现异常时,与异常相关的遥测参数值时序图形态发生畸变,目标遥测参数

在异常出现前后有明显变化。由于本文所选遥测参数值均为连续变量, 回归分析更简单有效, 故本文利用回归分析方法, 通过遥测参数间的相关性对航天器趋势进行预测, 有助于及时发现航天器异常情况并予以相应的控制。

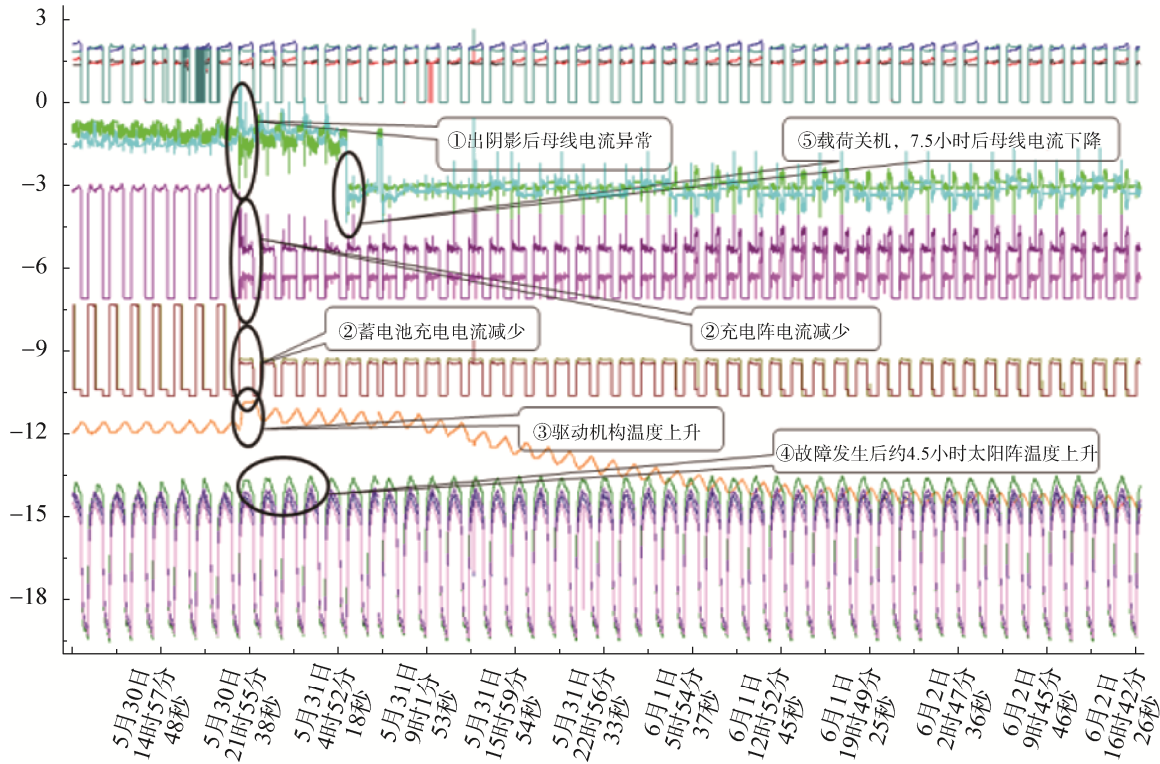


图 1 在轨故障过程主要遥测曲线

Fig. 1 Main telemetry curve during the in-track fault process

2.1 特征选取

为了减少遥测数据的特征数量, 本文使用皮尔逊相关系数对某航天器在轨运行过程中电源温控系统相关 40 个遥测参数进行选取。经过实验仿真, 选择了与航天器驱动机构温度高度相关的特征参数: 蓄电池 A、B 组的充电电流、充电阵 A、B 阵电流、太阳电池板温度及驱动机构温度遥测参数作为特征变量, 变量均为连续型。通过对正常状态和异常状态时的相关系数进行对比发现, 当航天器出现部件异常时, 遥测参数间的相关关系已经发生变化。

如表 1 所示, 可以看到正常状态时 7 个自变量与因变量的驱动机构温度相关系数均超过 0.8, 当航天器各分系统正常运行时, 相应的遥测参数具有确定的关系, 它们之间的关系为高度正相关; 异常状态时, 如表 2 所示, 相关系数出现了明显变化。因此, 本实验可以通过回归分析对驱动机构温度变化进行预测。

表 1 正常状态 Pearson 相关性

Table 1 Normal-state Pearson correlations

coef	A_1	A_2	A_3	A_4	A_5	A_6	B_1	B_2
A_1	1	0.96	-0.96	-0.96	-0.97	-0.92	-0.98	-0.98
A_2	0.96	1	-0.85	-1	-0.99	-0.98	-0.99	-0.97
A_3	-0.965	-0.85	1	0.86	0.88	0.8	0.89	0.92
A_4	-0.962	-0.99	0.86	1	0.99	0.97	0.99	0.98
A_5	-0.973	-0.99	0.88	0.973	1	0.97	0.996	0.99
A_6	-0.92	-0.98	0.8	0.975	0.97	1	0.98	0.95
B_1	-0.977	-0.99	0.89	0.99	0.99	0.97	1	0.99
B_2	-0.982	-0.97	0.92	0.979	0.99	0.95	0.99	1

表 2 异常状态 Pearson 相关性

Table 2 Abnormal-state Pearson correlations

coef	A_1	A_2	A_3	A_4	A_5	A_6	B_1	B_2
A_1	1	0.99	0.99	0.99	0.41	0.83	-0.41	1
A_2	0.99	1	0.99	0.99	0.39	0.83	-0.4	0.99
A_3	0.99	0.99	1	0.99	0.36	0.82	-0.39	0.99
A_4	0.99	0.99	0.99	1	0.38	0.82	-0.39	0.99
A_5	0.41	0.39	0.36	0.39	1	0.45	-0.64	0.41
A_6	0.83	0.83	0.82	0.82	0.44	1	-0.11	0.83
B_1	-0.41	-0.4	-0.39	-0.41	-0.64	-0.11	1	-0.41
B_2	-0.28	-0.26	-0.24	-0.26	-0.86	-0.24	0.66	1

2.2 数据预处理

航天器遥测数据受多种因素干扰，会出现数据不连续，野值或者明显不合理的值，这些情况对建模有一定的干扰，因此需要对遥测数据进行预处理，本文通过用 3 倍标准差剔除不合理的点；根据时间顺序，依次对正常状态时的 500 个样本点和异常发生后 264 个样本点进行实验。

数据预处理：根据参数范围，首先根据 3δ 标准差原则剔除数据中因解码等原因造成的野值，并将上述 8 个重要参数分别按 30 s 一条进行挑点存储。

本实验以驱动机构温度为因变量，基于回归的预测方法以其余电源温度相关遥测参数为自变量进行回归分析，建立驱动机构因变量相对应的回归方程，进而求解预测值。

2.3 回归建模

根据最小二乘法做线性回归拟和，如表 3 所示，可得回归方程为

表 3 回归系数

Table 3 The regression coefficients

	β_0	A_1	A_2	A_3	A_4	A_5	A_6	B_1
正常	9.133 9	-0.074 8	0.023 3	-0.004 7	0.039 3	0.019 9	0.183	-0.387
异常	16.946	0.055 4	-0.532 3	-0.420 2	-0.002 9	0.001 56	0.386 4	0.488 9

$$Y = 9.134 - 0.075 \times X_1 + 0.023 \times X_2 - 0.005 \times X_3 + 0.023 \times X_4 + 0.039 \times X_5 + 0.02 \times X_6 + 0.183 \times X_7 \quad (10)$$

$$Y = 16.946 - 0.055 \times X_1 - 0.532 \times X_2 - 0.42 \times X_3 - 0.0029 \times X_4 + 0.0016 \times X_5 + 0.386 \times X_6 + 0.489 \times X_7 \quad (11)$$

由式 (10)、式 (11) 可以得出，该典型异常发生前回归模型与正常状态下的回归模型不匹配，影响因素的权重发生了明显变化，从而可以说明航天器遥测参数趋势已经发生改变。

2.4 显著性检验

根据表 4 实验结果所示，由 OLS 最小二乘法可以得到，正常状态时模型的拟合优度为 99.3%，说明该回归模型拟合精度高， $p > |t|$ 对应的是 0.007，说明回归显著，回归结果可靠；异常状态时模型的拟合优度为 0.779，说明该回归模型拟合精度较好， $p > |t|$ 对应的是 0，说明回归显著，回归结果可靠。

表 4 OLS 回归结果

Table 4 OLS regression results

	$Coef_1$	Std_1	$p_1 > t $	$Coef_2$	Std_2	$p_2 > t $
const	9.133 9	3.351	0.007	16.945 7	0.794	0
X_1	-0.074 8	0.009	0	0.055 5	0.032	0.080
X_2	0.023 3	0.003	0	-0.532 4	0.597	0.373
X_3	-0.004 7	0.002	0.014	-0.420 2	1.007	0.677
X_4	0.039 4	0.002	0	-0.002 9	0	0
X_5	0.02	0.003	0	0.001 6	0.001	0.014
X_6	1.183 3	0.119	0	0.386 4	0.547	0.481
X_7	-0.387 4	0.134	0.004	0.488 9	0.507	0.336
R-squared	0.993			0.779		

3 结束语

本文根据某航天器典型异常案例, 对遥测数据进行了回归分析的异常检测, 主要做了数据预处理, 回归系数、数据拟合建模及显著性检验, 通过对不同时间段的航天器遥测数据的回归方程进行检验, 从而得出航天器的状况是否前后一致。它主要解决以下问题: ① 从数据角度出发, 确认遥测参数间是否具有相关性, 如果有, 确定它们之间究竟满足何种数学关系, 即经验公式或是回归方程, 并对它的可信度做统计检验; ② 从对同一个遥测参变量有影响的多个特征变量中, 选取具有显著系数的参数; ③ 利用航天器遥测数据去推算回归函数, 尤其是对随机变量 Y 的观测值做出点预测和区间预测。

本文提出的方法可用于故障检测, 通过遥测数据的不断积累, 对回归方程不断进行更新迭代, 不断提高检测的准确性, 从而可推广应用于其他航天器。

参考文献

- [1] 杜莹, 王飞, 孙超, 等. 基于多分布特征的航天遥测动态加权异常检测算法[J]. 兵器装备工程学报, 2018, 39(7): 98-105.
DU Ying, WANG Fei, SUN Chao, et al. Dynamic weighted anomaly detection algorithm for spacecraft telemetry data based on multi distribution features[J]. Journal of Sichuan Ordnance, 2018, 39(7): 98-105.
- [2] 王晓燕. 几种常用的异常数据挖掘方法[J]. 甘肃联合大学学报(自然科学版), 2010, 24(4): 68-71.
WANG Xiaoyan. Analysis of several commonly used methods on outlier data mining[J]. Journal of Gansu Lianhe University(Natural Sciences), 2010, 24(4): 68-71.
- [3] 秦巍, 郭永富. 一种基于历史遥测数据的在轨卫星故障预警系统[J]. 航天器工程, 2010, 19(6): 40-45.
QIN Wei, GUO Yongfu. A failure warning system based on historical telemetry data for satellites on orbit[J]. Spacecraft Engineering, 2010, 19(6): 40-45.
- [4] 姜海旭, 张科, 王靖宇, 等. 基于形态变分模态分解和JRD的航天器异常状态识别[J]. 西北工业大学学报, 2018, 36(1): 20-27.
JIANG Haixu, ZHANG Ke, WANG Jingyu, et al. Spacecraft anomaly recognition based on morphological variational mode decomposition and JRD[J]. Journal of Northwestern Polytechnical University, 2018, 36(1): 20-27.
- [5] 邵峰晶. 数据挖掘原理与算法[M]. 北京: 中国水利水电出版社, 2003.
- [6] 熊宇虹, 贺彦博, 郑庆晖, 等. 基于数据挖掘的航天遥测数据分析系统的设计[J]. 空间电子技术, 2019, 16(6): 24-27.
XIONG Yuhong, HE Yanbo, ZHENG Qinghui, et al. Design of spacecraft telemetry data analysis system based on data mining[J]. Space Electronic Technology, 2019, 16(6): 24-27.
- [7] 刘斌, 王鹏飞. 基于多元回归分析的煤的吸附瓦斯含量影响因素研究[J]. 华北科技学院学报, 2019, 16(1): 12-15.
LIU Bin, WANG Pengfei. Study on influencing factors of adsorption gas content of coal based on multivariate regression analysis[J]. Journal of North China Institute of Science and Technology, 2019, 16(1): 12-15.
- [8] 用 EXCEL 做回归分析的详细步骤[EB/OL]. (2019-02-28). <http://www.itongji.cn/detail?type=1395>.
- [9] 孟静. 异常数据挖掘算法研究与应用[D]. 无锡: 江南大学, 2013.

[作者简介]

- 史晓云 1986 年生, 硕士, 工程师, 主要研究方向为航天器在轨故障分析与评估。
 陈 军 1977 年生, 硕士, 正高级工程师, 主要研究方向为航天器在轨故障分析与评估。
 郭小红 1978 年生, 博士, 高级工程师, 主要研究方向为航天器故障诊断方法研究。
 符叶丹 1989 年生, 硕士, 工程师, 主要研究方向为航天器软件开发。
 王新广 1992 年生, 硕士, 工程师, 主要研究方向为航天器在轨故障分析与评估。

(本文编辑: 傅 杰)