

doi: 10.7690/bgzd.2015.05.001

## 基于可拓数据挖掘技术的故障诊断和预防方法

张金春<sup>1</sup>, 张家宾<sup>2</sup>, 李超亚<sup>2</sup>, 李日华<sup>1</sup>

(1. 海军航空工程学院基础部, 山东 烟台 264001; 2. 海军航空工程学院研究生大队, 山东 烟台 264001)

**摘要:** 针对发动机转子故障诊断和故障预测的问题, 提出一种可拓数据挖掘技术。将该技术应用于发动机转子系统的故障诊断和预防领域, 利用可拓遗传算法从大量原始数据中挖掘出转子系统的故障规则, 根据可拓学中挖掘蕴含知识的原理, 发现变化条件下故障规则的转变趋势。结果表明: 该方法有效、可行, 能实现对故障的预防, 为探究故障诊断中故障类型的变化问题提出了一种新的思路。

**关键词:** 可拓学; 数据挖掘; 故障诊断和预防; 转子振动故障

**中图分类号:** TJ04 **文献标志码:** A

## Methods of Fault Diagnosis and Prevention of Equipment Based on Extenics and Data Mining Technology

Zhang Jinchun<sup>1</sup>, Zhang Jiabin<sup>2</sup>, Li Chaoya<sup>2</sup>, Li Rihua<sup>1</sup>

(1. Department of Basic, Naval Aeronautical & Astronautical University, Yantai 264001, China;

2. Brigade of Postgraduate, Naval Aeronautical & Astronautical University, Yantai 264001, China)

**Abstract:** Aiming at the fault diagnosis and forecast of engine rotor, put forward the extenics data mining technology. Use the technology in fault diagnosis and forecast of engine rotor system, adopt extenics algorithm to excavate rotor system fault rule form lots of raw data. According to the principle of extenics mining contain knowledge, find the transformation trend of the fault rule. The result shows that the method is effective and feasible. It can be used for fault prevention and give a new method for fault type transformation in fault diagnosis.

**Keywords:** extenics; data mining; fault diagnosis and prevention; rotor vibration fault

### 0 引言

军事装备的故障率高, 不仅使装备的作战效能降低, 同时维修所消耗的人力、财力、物力等也造成了资源的极大浪费。低水平的装备完好率, 不仅影响平时的训练, 在战时可能贻误战机, 对战争胜负产生很大的影响<sup>[1]</sup>。若能依据设备的工作机理以及结构特征, 以传感器所探测的数据为基础, 分析研究故障发生的起因, 并能有针对性地采取相应的预防措施, 将故障防患于未然, 则能在节约资源的前提下, 提高部队装备的作战效能。利用可拓数据挖掘方法可以对故障模式进行判断并有针对性地采取相应的维修手段。但如果在故障发生之前, 能及时发现故障隐患, 对设备进行有效维护和调整, 则可能把故障遏制在即将发生的状态。可拓学中的挖掘蕴含规则原理能发现故障系统潜在的故障变化。

由故障诊断知识可以得到一系列的故障诊断规则, 即在模型的特征值处于不同的特征值区间时, 对应的故障种类是固定的。实际应用过程中, 往往会出现因为某些特征的改变而使整个系统的故障类

型发生变化。因此监测这种变化并及时发现故障类型的转变非常重要。另一方面, 若能根据故障诊断系统中特征值的某些变化, 判断出故障类型的转变趋势, 那么就可以利用这种故障转变趋势, 对可能发生的故障进行预防研究<sup>[2]</sup>。

这就要求在进行装备的后勤保障以及后期的维修管理时, 要密切关注设备故障类型的变化。故障种类的变化可以反映到故障模型对应的特征值的变化, 因此可以通过研究故障模式对应的特征值的变化来监测对应故障类型的变化。可拓数据挖掘技术可以研究在变化的条件下故障类型的变化。在存在已知若干条规则的条件下, 若存在对条件的变换, 则存在对结论的传导变换<sup>[3]</sup>。可拓学中对这种蕴含型知识的挖掘, 为探究故障诊断中故障类型的变化问题提供了一种新的思路。

### 1 发动机转子的故障诊断模型

笔者以发动机转子系统的故障诊断为例, 来说明可拓数据挖掘技术在故障预防领域的应用。

收稿日期: 2014-12-10; 修回日期: 2015-01-28

作者简介: 张金春(1960—), 男, 辽宁人, 硕士, 教授, 从事复杂系统的建模分析与仿真研究。

航空发动机是飞行器的核心部件，由于其工作在高压、高温以及高负载的条件下，发动机出现故障的概率较高。转子是发动机的核心部件，由振动导致的发动机转子的故障占发动机故障总数的 80%。随着发动机技术的飞速发展，建立发动机转子的故障诊断和预防系统显得尤为重要<sup>[4]</sup>。

建立转子系统的物元模型为

$$M = (O, c, v) = (\text{故障, 相应特征集, 经典域})$$

对发动机转子系统进行可拓分析，确定其特征集为

$$C = \begin{cases} c_1 \text{ 奇异谱熵} \\ c_2 \text{ 双通道奇异谱熵} \\ c_3 \text{ 功率谱熵} \\ c_4 \text{ 小波能谱熵} \\ c_5 \text{ 双通道复数小波能谱熵} \\ c_6 \text{ 小波空间特征谱熵} \\ c_7 \text{ 双通道复数小波空间特征谱熵} \end{cases}$$

几种常见的故障类型为

$$O_m = \begin{cases} O_1 \text{ 转子裂纹,} \\ O_2 \text{ 转子碰摩,} \\ O_3 \text{ 裂纹碰摩耦合故障,} \\ O_4 \text{ 转子不对中,} \\ O_5 \text{ 转子不平衡,} \\ O_6 \text{ 转子支座松动} \end{cases}$$

利用六通道转子实验台采集故障数据如表 1 所示，表格节选了数据的一部分。

数据的第 27 行，第 3 个传感器采集的幅值数据为 80.583，超出了信号的幅值范围，为错误数据，将其剔除并选择序列的前后数据的平均值补充代替。第 21 行数据缺失（6 个传感器都未采集到信号，可能是实验台的软件系统在这个采样点出现问题），可以直接忽略（因为采样点比较密集，少量数据的缺失不影响结果）；而第 2 列缺失数据较多，则通过判定数据序列的走向，并对数据进行平滑处理，填补缺失值。第 16 行的第 4 个传感器，所采集到的数据为 -46.433，和邻近的数据出入很大，为离群点数据，应予以剔除<sup>[5-8]</sup>。

对所采集的数据进行清洗后，数据序列如表 2。

表 1 碰磨故障状态下振动信号的幅值序列

| 序列 | 通道     |        |        |         |        |        |
|----|--------|--------|--------|---------|--------|--------|
|    | 1      | 2      | 3      | 4       | 5      | 6      |
| 1  | 30.973 | 28.884 | 29.408 | 30.156  | 30.465 | 28.437 |
| 2  | 33.861 | 31.781 | 32.314 | 33.010  | 33.483 | 31.463 |
| 3  | 36.861 | 34.669 | 35.323 | 35.881  | 36.629 | 34.514 |
| 4  | 39.647 | 37.652 | 38.237 | 38.950  | 39.612 | 37.506 |
| 5  | 42.372 | 40.498 | 41.065 | 41.701  | 42.397 | 40.205 |
| 6  | 45.114 | 43.463 | 44.014 | 44.392  | 44.985 | 43.154 |
| 7  | 47.530 | 46.120 | 46.378 | 46.885  | 47.590 | 45.707 |
| 8  | 49.610 | 48.578 | 48.707 | 49.378  | 49.971 | 48.200 |
| 9  | 51.321 | 50.616 | 50.616 | 51.441  | 51.974 | 50.298 |
| 10 | 52.980 | 52.163 | 52.447 | 53.263  | 53.461 | 52.421 |
| 11 | 54.364 | 53.478 | 53.994 | 54.527  | 55.051 | 54.029 |
| 12 | 55.765 | 55.172 | 55.662 | 55.937  | 56.599 | 55.567 |
| 13 | 56.736 | 56.676 | 56.762 | 57.218  | 57.338 | 56.590 |
| 14 | 57.347 | 57.493 | 57.587 | 58.043  | 57.828 | 57.579 |
| 15 | 57.218 | 57.837 | 57.914 | 57.897  | 57.974 | 58.112 |
| 16 | 57.011 | 57.863 | 57.768 | -46.433 | 57.536 | 57.983 |
| 17 | 56.040 | 0      | 56.900 | 56.994  | 56.762 | 57.708 |
| 18 | 55.112 | 0      | 55.971 | 56.117  | 55.903 | 56.805 |
| 19 | 53.917 | 0      | 54.837 | 55.017  | 54.338 | 55.404 |
| 20 | 52.361 | 0      | 53.710 | 53.461  | 53.091 | 54.149 |
| 21 | 0      | 0      | 0      | 0       | 0      | 0      |
| 22 | 48.527 | 0      | 49.971 | 49.653  | 49.249 | 50.977 |
| 23 | 46.429 | 48.415 | 47.753 | 47.530  | 46.627 | 48.578 |
| 24 | 44.014 | 45.827 | 45.595 | 44.968  | 44.426 | 46.283 |
| 25 | 41.555 | 43.506 | 42.741 | 42.595  | 41.469 | 43.644 |
| 26 | 38.916 | 40.609 | 40.171 | 39.818  | 39.200 | 41.306 |
| 27 | 36.449 | 38.237 | 80.583 | 37.446  | 36.311 | 38.555 |
| 28 | 33.655 | 35.262 | 35.004 | 34.428  | 33.904 | 35.976 |
| 29 | 31.342 | 33.010 | 32.288 | 31.910  | 30.861 | 33.156 |
| 30 | 28.609 | 30.044 | 29.786 | 29.159  | 28.609 | 30.629 |
| 31 | 26.408 | 27.818 | 27.053 | 27.010  | 25.875 | 27.878 |
| 32 | 23.184 | 24.869 | 24.577 | 24.207  | 23.520 | 25.376 |
| ⋮  | ⋮      | ⋮      | ⋮      | ⋮       | ⋮      | ⋮      |

表 2 数据清洗后的振动信号的幅值序列

| 序列 | 通道     |        |        |        |        |        |
|----|--------|--------|--------|--------|--------|--------|
|    | 1      | 2      | 3      | 4      | 5      | 6      |
| 1  | 30.973 | 28.884 | 29.408 | 30.156 | 30.465 | 28.437 |
| 2  | 33.861 | 31.781 | 32.314 | 33.010 | 33.483 | 31.463 |
| 3  | 36.861 | 34.669 | 35.323 | 35.881 | 36.629 | 34.514 |
| 4  | 39.647 | 37.652 | 38.237 | 38.950 | 39.612 | 37.506 |
| 5  | 42.372 | 40.498 | 41.065 | 41.701 | 42.397 | 40.205 |
| 6  | 45.114 | 43.463 | 44.014 | 44.392 | 44.985 | 43.154 |
| 7  | 47.530 | 46.120 | 46.378 | 46.885 | 47.590 | 45.707 |
| 8  | 49.610 | 48.578 | 48.707 | 49.378 | 49.971 | 48.200 |
| 9  | 51.321 | 50.616 | 50.616 | 51.441 | 51.974 | 50.298 |
| 10 | 52.980 | 52.163 | 52.447 | 53.263 | 53.461 | 52.421 |
| 11 | 54.364 | 53.478 | 53.994 | 54.527 | 55.051 | 54.029 |
| 12 | 55.765 | 55.172 | 55.662 | 55.937 | 56.599 | 55.567 |
| 13 | 56.736 | 56.676 | 56.762 | 57.218 | 57.338 | 56.590 |
| 14 | 57.347 | 57.493 | 57.587 | 58.043 | 57.828 | 57.579 |
| 15 | 57.218 | 57.837 | 57.914 | 57.897 | 57.974 | 58.112 |
| 16 | 57.011 | 57.863 | 57.768 | 57.847 | 57.536 | 57.983 |
| 17 | 56.040 | 57.003 | 56.900 | 56.994 | 56.762 | 57.708 |
| 18 | 55.112 | 56.152 | 55.971 | 56.117 | 55.903 | 56.805 |
| 19 | 53.917 | 55.189 | 54.837 | 55.017 | 54.338 | 55.404 |
| 20 | 52.361 | 53.710 | 53.710 | 53.461 | 53.091 | 54.149 |
| 21 | 50.444 | 51.407 | 52.390 | 51.743 | 52.592 | 52.864 |
| 22 | 48.527 | 50.066 | 49.971 | 49.653 | 49.249 | 50.977 |
| 23 | 46.429 | 48.415 | 47.753 | 47.530 | 46.627 | 48.578 |
| 24 | 44.014 | 45.827 | 45.595 | 44.968 | 44.426 | 46.283 |
| 25 | 41.555 | 43.506 | 42.741 | 42.595 | 41.469 | 43.644 |
| 26 | 38.916 | 40.609 | 40.171 | 39.818 | 39.200 | 41.306 |
| 27 | 36.449 | 38.237 | 37.734 | 37.446 | 36.311 | 38.555 |
| 28 | 33.655 | 35.262 | 35.004 | 34.428 | 33.904 | 35.976 |
| 29 | 31.342 | 33.010 | 32.288 | 31.910 | 30.861 | 33.156 |
| 30 | 28.609 | 30.044 | 29.786 | 29.159 | 28.609 | 30.629 |
| 31 | 26.408 | 27.818 | 27.053 | 27.010 | 25.875 | 27.878 |
| 32 | 23.184 | 24.869 | 24.577 | 24.207 | 23.520 | 25.376 |
| ⋮  | ⋮      | ⋮      | ⋮      | ⋮      | ⋮      | ⋮      |

采集不同故障模式下的数据幅值，并进行数据预处理。利用信息熵的计算方法得到各个信息熵值的序列。对所得到的信息熵值序列进行数据挖掘，取得各个信息熵序列的最小最大值如表 3 所示。

表 3 信息熵序列的区间

| 信息熵   | 区间       | 信息熵   | 区间       |
|-------|----------|-------|----------|
| $c_1$ | (35, 85) | $c_5$ | (45, 80) |
| $c_2$ | (35, 70) | $c_6$ | (45, 75) |
| $c_3$ | (25, 80) | $c_7$ | (5, 40)  |
| $c_4$ | (20, 85) |       |          |

## 2 基于可拓规则挖掘的故障诊断方法

基于规则挖掘的故障诊断方法，是在物元模型建立的前提下，采用数据挖掘的算法挖掘出特征值和故障类型之间的规则。通过样本和规则的比较来确定待判断样本的类型。

同样以转子的可拓故障诊断模型为基础，采用基于遗传算法的数据挖掘技术对转子故障诊断系统进行规则的挖掘。基于规则挖掘的故障诊断方法要求特征值为离散的向量，因此先要对特征值数据进行离散化处理。以奇异谱熵特征值的离散化为例，由表 3 可知其最大最小值区间为(35, 85)对数据进行 10 等分(之所以 10 等分，是在综合考虑计算的复杂度和计算的准确度后决定的)，并分别进行 1~10 的赋值。按同样方法对其他特征值进行离散化处理。

遗传算法一般采用二进制编码来表示样本，并通过样本之间的交叉、变异和重组等步骤来挖掘相应的规则<sup>[9-10]</sup>。在本问题中，每个信号的信息熵特征对应 10 个离散数据值，因此每种信息熵都可看作存在 10 种不同的特征状态，对这 10 种不同的特征状态值进行二进制编码，以信号的奇异谱熵特征为例，其对应的 10 种状态的二进制编码如表 4 所示。

表 4 奇异谱熵值的二进制编码

| 原区间      | 离散值 | 编码值  | 原区间      | 离散值 | 编码值  |
|----------|-----|------|----------|-----|------|
| (35, 40) | 1   | 0000 | (60, 65) | 6   | 0101 |
| (40, 45) | 2   | 0001 | (65, 70) | 7   | 0110 |
| (45, 50) | 3   | 0010 | (70, 75) | 8   | 0111 |
| (50, 55) | 4   | 0011 | (75, 80) | 9   | 1000 |
| (55, 60) | 5   | 0100 | (80, 85) | 10  | 1001 |

继续划分其余几类信息谱熵值区间，并分别进行二进制编码。对 6 种故障类型进行二进制编码

$$D = \begin{bmatrix} \text{转子裂纹} \\ \text{转子碰摩} \\ \text{裂纹碰摩耦合} \\ \text{转子不对中} \\ \text{转子不平衡} \\ \text{转子支座松动} \end{bmatrix} = \begin{bmatrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \end{bmatrix}$$

因此每条遗传算法的信息由 28 位条件信息和 3 位结论信息组成。例如任取一条样本信息，其遗传信息表示为

$$M = [(10001001100010000111011001001000), (000)]$$

设定适应度函数为可拓关联函数，训练的收敛条件为连续 3 代所得的最优解都未发生改变或整体的适应值提高不足 1.5%，交叉概率  $P_c=0.9$ ，变异概率  $P_m=0.02$ ，种群规模  $NP=16$ 。经遗传算法所得的规则如表 5 所示( $C_1 \sim C_7$  代表 7 种信息熵特征， $D$  代表相对应的故障种类)。

表 5 转子故障诊断系统规则对应表

| $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ | $C_7$ | $D$ |
|-------|-------|-------|-------|-------|-------|-------|-----|
| 8     | 9     | 8     | 8     | 6     | 4     | 8     | 0   |
| 6     | 7     | 5     | 9     | 3     | 9     | 4     | 1   |
| 9     | 9     | 8     | 8     | 6     | 4     | 8     | 2   |
| 5     | 7     | 6     | 9     | 3     | 9     | 4     | 3   |
| 3     | 4     | 2     | 1     | 7     | 3     | 6     | 4   |
| 3     | 2     | 4     | 1     | 9     | 4     | 7     | 5   |

取待测样本经数据预处理后，计算其 7 种信息熵值，采用相同的数据离散方法算出样本各信息熵特征的离散值。将样本的 7 种信息熵离散值和表 5 中故障对应的规则相比较，选择距离最近的故障模式，即为待测样本的故障类型。

## 3 对故障类型转变的故障预防方法研究

### 3.1 挖掘蕴含型规则的可拓学原理

从已知的蕴含型知识中挖掘潜在的新的规则的过程，被称为挖掘基于蕴含型知识的可拓挖掘过程。

给定 2 条规则

$$A \Rightarrow P$$

$$B \Rightarrow N$$

其中： $A$ 、 $B$  代表条件； $P$ 、 $N$  代表结论。

若存在 1 个对条件产生的可拓变换  $\varphi$ ，有

$$\varphi A \Rightarrow B$$

那么会产生 1 个对结论的传导变换  ${}_A T_P$ ，使得下式成立

$${}_A T_P P \Rightarrow N$$

整体的规则为

$$(\varphi A \Rightarrow B) \Rightarrow ({}_A T_P P \Rightarrow N)$$

可以表示为

$$\text{IF } \varphi A \Rightarrow B \text{ THEN } {}_A T_P P \Rightarrow N$$

### 3.2 转子系统的故障预防方法

根据上文，由基于可拓规则挖掘的故障诊断方法得到的规则如表 5 所示。

以  $L_{ij}$  代表规则的条件,  $D_i$  代表规则的结论, 则有

$$L_{i1} \wedge L_{i2} \wedge L_{i3} \wedge L_{i4} \wedge L_{i5} \wedge L_{i6} \wedge L_{i7} \Rightarrow D_i$$

即在 7 种条件的共同作用下, 可以得出相应的结论。

从表中可以得出转子裂纹故障和转子裂纹碰磨耦合故障对应的规则分别为:

$$8 \wedge 9 \wedge 8 \wedge 8 \wedge 6 \wedge 4 \wedge 8 \Rightarrow D_1$$

$$9 \wedge 9 \wedge 8 \wedge 8 \wedge 6 \wedge 4 \wedge 8 \Rightarrow D_3$$

若存在对条件的变换  $\varphi$ , 使得如下的条件变换成立。

$$\varphi L_{11} = L_{31}$$

则根据上述挖掘蕴含规则的可拓学原理, 必有对结论的传导变换, 使得对结论的传导变换  $l_1 T_{D_1}$  成立。即为

$$l_1 T_{D_1} D_1 \Rightarrow D_3$$

这个变换具体解释为在转子系统发生裂纹故障时, 存在向裂纹碰磨耦合故障转化的潜在趋势, 而转子的奇异谱熵值的变化可以将这种趋势表现出来。若裂纹故障状态下的奇异谱熵值在某种变换条件下, 从(70, 75)区间变化为(75, 80)区间, 则相应的也会发生对故障类型的传导变换, 故障类型由裂纹故障向裂纹碰磨耦合故障转变。这要求在维修裂纹故障时, 应时刻注意转子系统的奇异谱熵值的变换, 看其是否也发生了碰磨故障, 以便采取正确的维修方法。另一方面在发现转子处于裂纹故障状态时, 也应预防其再发生碰磨故障, 从而实现对其他可能发生的故障的预防<sup>[11]</sup>。

同理, 从表 5 中可知转子不对中故障和转子碰磨故障的规则分别为:

$$5 \wedge 7 \wedge 6 \wedge 9 \wedge 3 \wedge 9 \wedge 4 \Rightarrow D_4$$

$$6 \wedge 7 \wedge 5 \wedge 9 \wedge 3 \wedge 9 \wedge 4 \Rightarrow D_2$$

若存在对规则的变换  $\varphi'$ , 使得以下 2 条规则同时成立

$$\varphi' L_{43} = L_{23}$$

$$\varphi' L_{41} = L_{21}$$

根据上述原理, 必有对结论的传导变换使得下面对结论的变换成立

$$L_{43} \wedge L_{41} T_{D_4} D_4 \Rightarrow D_2$$

该变换蕴含的潜在规则为当转子的双通道奇异谱熵值由区间(55, 60)变换到(60, 65), 并且小波能谱熵值由区间(59, 62.5)变化到(62.5, 66), 则转子系统的不对中故障会引发转子的碰磨故障。这也警示设备维修人员, 转子系统故障发生后, 若不及时采取相应的维修措施, 很有可能导致其他故障的出现。

可拓学可以挖掘变化条件下的蕴含规则, 根据蕴含规则发现故障潜在的趋势变化, 从而进行故障的预防。

### 4 结束语

笔者采用基于可拓学的遗传算法挖掘出转子故障系统的诊断规则, 并在此基础上讨论了挖掘蕴含型知识的可拓学原理, 通过挖掘故障系统的蕴含规则, 发现潜在的故障趋势从而实现对故障的预防。

基于可拓学的故障诊断方法可以将故障诊断问题形式化, 因此其应用范围广泛, 受所研究对象的限制较小, 同时该方法还可应用于其他模式识别问题的研究。由于目前对已知的故障知识所知较少, 限制了该方法的应用, 这也是进一步研究的内容。

### 参考文献:

- [1] 陈予恕, 张华彪. 航空发动机整机动力学研究进展与展望[J]. 航空学报, 2011, 32(08): 1371-1391.
- [2] 周德新, 谢晓敏. 基于可拓规则的故障诊断专家系统推理机的研究[J]. 计算机测量与控制, 2011, 19(2): 266-268.
- [3] 马丽丽. 基于可拓理论的 Apriori 数据挖掘算法分析[J]. 计算机与数字工程, 2010, 38(3): 9-11.
- [4] 陈果, 李成刚, 王德友. 航空发动机转子-滚动轴承-支承-机匣耦合系统的碰磨故障分析与验证[J]. 航空动力学报, 2008, 23(7): 1304-1311.
- [5] 苗润华. 基于聚类和孤立点检测的数据预处理方法的研究[D]. 北京: 北京交通大学, 2012: 12-18.
- [6] 方洪鹰. 数据挖掘中数据预处理的方法研究[D]. 重庆: 西南大学, 2009: 23-26.
- [7] 李萌. 基于粗集理论的数据挖掘的数据预处理研究[D]. 武汉: 华中师范大学, 2004: 18-19.
- [8] 朱可恒. 滚动轴承振动信号特征提取及诊断方法研究[D]. 大连: 大连理工大学, 2013: 41-44.
- [9] 向长城, 黄席樾, 殷礼胜, 等. 基于遗传算法与可拓神经网络的故障诊断[J]. 计算机仿真, 2008, 25(4): 249-252.
- [10] 郑伟. 航空发动机转子系统振动故障诊断技术研究[D]. 德阳: 中国民用航空飞行学院, 2013: 14-15.
- [11] 付琪. 基于信息融合的转子振动故障诊断方法研究[D]. 沈阳: 沈阳航空航天大学, 2013: 37-39.