第 34 卷　第 6 期　　火 箭 推 进　　Vol.34, No.6
2008 年 12 月　　JOURNAL OF ROCKET PROPULSION　　Dec. 2008

液体火箭发动机故障诊断中的特征优选策略

袁晓峰[1]，许化龙[1]，陈淑红[2]
(1 第二炮兵工程学院研四队 西安 710025；2 第二炮兵装备研究院第三研究所 北京 100085)

摘　要：针对液体火箭发动机故障诊断中的特征选择问题，提出了一类最小代价模糊决策模型。在应用中，该模型能够融合各个故障特征参数的先验信息、故障程度，并允许故障间的关联存在，从而比现有特征选择方法具有广泛的适用性，特征选择结果更贴近工程实际。同时，给出了基于贪婪算法的模型求解方法，并对某发动机的故障特征集进行了特征优选实验。实验证明了该方法的有效性。

关键词：液体火箭发动机；粗糙集；模糊决策；贪婪算法

中图分类号：V434　　**文献标识码**：A　　**文章编号**：(2008) 06-0001-05

An optimized feature selection strategy in liquid rocket engine fault diagnosis

Yuan Xiaofeng[1], Xu Hualong[1], Chen Shuhong[2]
(1 Second Artillery Engineering Institute, Xi′an 710025, China;
(2 Second Artillery Armament Academy, Beijing 100085, China)

Abstract: To solve the problem of feature selection in liquid rocket engine fault diagnosis, a fuzzy decision model constrained by the minimum cost is proposed in this paper. In this model, the prior information of each feature and fault intensity is taken into consideration, and the association between different faults is allowed. Accordingly, it is more flexible than other models, and the feature selection results are more suitable for engineering applications. In addition, a greedy algorithm is provided to solve the model. Experiments on a feature set of a liquid rocket engine have validated the effectiveness of the model.

Key words: liquid rocket engine; rough set; fuzzy decision, greedy algorithm

---

收稿日期：2008-04-22；**修回日期**：2008-06-30。

作者简介：袁晓峰（1975—），男，博士生，工程师，研究领域为数据挖掘与故障诊断。

万方数据

0 引言

液体火箭发动机的工作状态参数有几十个甚至上百个，各个参数分布在发动机的不同部位，在监测中涉及诸多方面的问题。因此，在故障诊断中只选取部分特征参数。如何针对发动机故障诊断的工程实际，综合考虑各种因素进行特征选择是一个复杂而重要的问题。

从本质上讲，特征选择的目的是在保持故障分类信息不变的条件下，从原始故障特征集合中尽可能多地去除冗余特征，获得最优的特征子集。这里所指的最优是从多个方面来评价的，诸如灵敏度、稳定性、测试量和计算量以及识别率等[1]。目前，液体火箭发动机故障诊断特征选择方法主要包括基于 Shannon 互信息熵的方法[2]、基于样本集综合离散度的方法[3] 和粗糙集 (Rough Set) 方法[4]。其中，互信息熵的方法中采用的准则是，类特征互信息量大且子集中特征之间的互信息量小；样本集综合离散度法采用的准则是，特征子集对原始特征集合具有最大的类内与类间的综合离散度。而粗糙集方法是一种近年出现的新方法，它通过属性约简实现特征选择。简而言之，这些方法虽然基于不同理论，但它们都是从特征相互关系的角度，最大程度上去除冗余特征的，而无法实现对具体特征量在测试量、获得的难易程度、检测精度和检测成本等因素综合考虑。同时，故障间关联性和故障程度无法体现。例如，在某些特殊情况下一组发动机监测参数可能对应不同程度的几个故障。因此，通过这些方法进行特征选择，很难实现真正意义上的最优。有鉴于此，本文提出了一类最小代价模糊决策模型，以期提高特征选择的灵活性和适用性，并实现特征选择的优化。

1 最小代价模糊决策模型

1.1 模型定义

定义 1[5]：称 D 为包含度，若满足以下条件：

$$\begin{cases} 0 \leqslant D(A_2/A_1) \leqslant 1 \\ A_1 \subseteq A_2 \Rightarrow D(A_2/A_1)=1 \\ A_1 \subseteq A_2 \subseteq A_3 \Rightarrow D(A_1/A_3) \leqslant D(A_2/A_3) \end{cases} \tag{1}$$

包含度刻画了两个模糊集合之间的关系，而粗糙模糊集的上/下近似只是包含度的特殊情形。包含度在实际应用中，可以根据需要给出具体定义，在此定义如下，

$$D(A_2/A_1)=\max\{A_2(x) \wedge A_1(x) \mid x \in U\} \tag{2}$$

令论域 U 上模糊集合的全体为 $F(U)$，U 上经典集合的全体为 $P(U)$。

定义 2：设 $S=\langle U,R,V,W,f\rangle$ 为一知识表达系统，U 是论域，$R=A\cup D$，$A\cap D=\varnothing$，$A\in P(U)$为条件属性集，$D\in F(U)$为模糊决策属性集，V 为属性值的集合；W 为与条件属性的一一对应，且所有元素均为正实数的权值集合，它反映了条件属性的代价；$f:U\times R\rightarrow V$ 为信息函数，它指定 U 中每一个对象 x 的属性值。称 S 为加权模糊决策系统。

对于加权模糊决策系统，令

$$R_B=\{(x_i,x_j) \mid f_m(x_i)=f_m(x_j)\},a_m\in B\} \quad B\subseteq A \tag{3}$$

则 R_B 为 U 上的等价关系，包含 x_i 的等价类为：$[x_i]_B=\{xj|(xi,xj)\in RB\}$。设 D 是 $F(U)$的包含度，令[6]

$$\begin{cases} D_B(D_k)(x_i)=D(D_k/[x_i]_B) \quad (k\leqslant r) \\ H_B(x_i)=(D_B(D_1)(x_i),D_B(D_2)(x_i),\cdots,D_B(D_r)(x_i)) \\ M_B(x_i)=\{D_k \mid D_B(D_k)(x_i)=\max(D_B(D_k)(x_i),j\leqslant r)\} \end{cases} \tag{4}$$

其中，r 为决策属性的个数。有模糊决策规则：对于 $x_i\in[x]_B$，x_i 的决策为 $M_B(x_i)$。

从决策协调集[7]概念出发，对加权模糊决策系统的最优决策约简定义如下：

定义 3：$S=\langle U,R,V,W,f\rangle$为加权模糊决策系统，若对于任意 $x_i\in U$，有 $B\subseteq A$，$H_B(x_i)=H_A(x_i)$，称 B 为 S 的决策协调集。若 B 为 S 的决策协调集，且其任何真子集都不是 S 的决策协调集，则称 B 为 S 的决策约简。若 B 为 S 的一个决策约简，且其对应代价在所有约简中最小，称 B 为 S 的最优决策约简，对应代价为最优决策代价，由此，系统 S 亦称作最小代价模糊决策系统。

万方数据

定义 4：设最小代价模糊决策系统 S 的决策协调集构成的集合为 S_c，B 为 S 的决策协调集，有：

$$\begin{cases}\mathrm{Cost_{opt}}(S_c)=\min\left\{\mathrm{cost}(B)=\sum_{b_i\in B} b_i\cdot W(b_i)\,\middle|\,B\in S_c\right\}\\ B_{opt}=\{B\,|\,\mathrm{cost}(B)=\mathrm{Cost_{opt}}(S_c),B\in S_c\}\end{cases}\quad(5)$$

则称 B_{opt} 为 S 的最小代价属性协调集。

1.2 模型的性质

性质 1：最小代价模糊决策系统 S 的最小代价属性协调集 B，则 B 一定是 S 为决策约简，且为最优决策约简。

证明：反证法，假设最小代价决策协调集 B 不是 S 的决策约简，则必存在决策协调集 $B'=B-\{b_k\}$，其中 $b_k\in B$；显然，有 $\mathrm{cost}(B')<\mathrm{cost}(B)$，这与已知条件中 B 为 S 的最小代价属性协调集相矛盾，所以假设不成立。因此，B 为 S 的决策约简，且根据定义 (3) 和 (4) 可知为 B 为 S 的最优决策约简。

性质 2：最小代价模糊决策系统 $S=\langle U,R,V,W,f\rangle$,则 B 是 S 的决策协调集当且仅当[7]，

$$\begin{cases}B\cap Q_A(x_i,x_j)\neq\emptyset\quad(x_i,x_j\in U)\\ Q_A(x_i,x_j)=\begin{cases}\{a_m\in A\,|\,f(x_i)\neq f(x_j)\}& H_A(x_i)\neq H_A(x_j)\\ A & H_A(x_i)=H_A(x_j)\end{cases}\end{cases}\quad(6)$$

推论 1：分辨矩阵中所有元素 $Q_A(x_i,x_j)$构成的集合为 Q，令 Q 中包含属性 a_k 的所有矩阵元素所构成的集合为 $S(a_k)$，属性 a_k 对应的代价为 $W(a_k)$，则决策协调集 B 必满足 $\cup_{a_k\in B}S(a_k)=Q$。

证明：B 满足 $\cup_{a_k\in B}S(a_k)=Q$，则对于 Q 中的任意元素 Q_A (x_i,x_j)，至少存在一个属性 $a_n\in B$ 使 $Q_A(x_i,x_j)\in S(a_n)$，即 $B\cap Q_A(x_i,x_j)=a_n\neq\emptyset$；由性质 2 得 B 为 S 的属性协调集，证毕。

由推论 1 结合定义 (3) 和 (4)，有以下推论。

推论 2：集合 B 为最优决策约简的充要条件为：集合 B 满足 $\cup_{a_k\in B}S(a_k)=Q$ 并且使 $\sum_{a_k\in B}W(a_k)$最小。

令矩阵 Q 中元素的个数为 m，条件属性的个数为 n。上述问题可通过布尔矩阵进行描述：将集合 S (a_k)转化为由元素 0 和 1 构成的列向量 S_k。具体转化原则为：如果 $S(a_k)$包含 Q 中的某个元素 q，则将 S_k 中对应位置的元素置 1，否则置 0。例如 S $(a_1)=\{q_1,q_3,q_5\}$，则 $S_1=\{1,0,1,0,1\}$。S_k 中对应权重为 $W(a_k)$，为了表示方便这里用 W_k 表示。信息表中由列向量 S_k 构成的布尔矩阵（表 1)表示为 $P=(p_{ij})_{m\times n}$。本文中所指的向量均为列向量。

表 1 布尔矩阵

Tab.1 Illustration of Boolean matrix

Q	w_1	w_2	w_3	w_4	w_t
	S_1	S_2	S_3	S_4	S_5
$q_1=\{a_1\ a_2\ a_5\}$	1	1	0	0	1
$q_2=\{a_2\ a_3\}$	0	1	1	0	0
$q_3=\{a_1\ a_2\ a_3\}$	1	1	1	0	0
$q_4=\{a_4\ a_6\}$	0	0	0	1	0
$q_5=\{a_1\ a_2\ a_4\ a_5\}$	1	1	0	1	1

2 最优决策约简求解算法

最优决策约简为 NP 完全问题，从理论上讲它在多项式时间只能获得最优解或近似最优解，下面给出问题求解的贪婪算法。

贪婪算法是解决最优化问题的一种基本方法，它将全局最优问题进行了转化，通过逐步构造最优解的方法，在每个阶段都做出一个在贪婪准则下最优决策；而不考虑在最后看来这种选择是否合理。贪婪算法只能获得问题的最优解或近似最优解。

求解最优决策约简的贪婪算法：在循环过程的每一步中，算法从当前从未选中的所有条件属性中，计算每个属性的有效代价 α，并选择 α 最小的属性加入到约简 Reduct 中，直到满足 $\cup_{a_k\in \mathrm{Reduct}}S(a_k)=Q$。贪婪算法整个过程如下：

步骤 1 $C=\emptyset$, Reduct=$\emptyset$;

步骤 2 while $Q\neq\emptyset$ do

{找到属性 a_k 使 $\alpha=w(a_k)/|Q\cap S(a_k)|$ 最小;

$C=C\cup S(a_k)$; $Q=Q\setminus S(a_k)$;

Reduct=Reduct$\cup$ $\{a_k\}$;}

步骤 3 对 Reduct 去冗余

Resort (Reduct) ;//对 Reduct 中的元素按照其

万方数据

对应属性代价从大到小排序

For $a_k \in$ Reduct

{If $\cup_{z\in Reduct-\{ak\}} S(z)=Q$;

Then Reduct=Reduct-{a_k};}

采用贪婪算法求解最优决策约简的优点在于该算法计算量较小，因而能够在实际应用中快速求解，尤其当属性矩阵 Q 的维数较大时。

3 故障特征的优选

3.1 实现过程

下面以实例说明最小代价模糊决策模型在发动机故障特征优选中的应用。数据由液体发动机稳定工作状态的故障仿真模型生成。涉及 22 个参数保持稳定状态。这些参数如表 2、表 3 所示。

表 2 发动机参数及其代号

Tab.2 Parameters and their names of an engine

代号	参数名称	代号	参数名称
a_1	氧化剂流量	a_{12}	冷却套流量
a_2	燃料流量	a_{13}	涡轮泵转速
a_3	燃烧室氧化剂流量	a_{14}	燃烧室燃料流量
a_4	燃气发生器氧化剂流量	a_{15}	头部隔板流量
a_5	燃气发生器燃料流量	a_{16}	发动机总流量
a_6	燃料主导管分支处压力	a_{17}	涡轮燃气流量
a_7	蒸发器流量	a_{18}	氧化剂泵功率
a_8	燃烧室压力	a_{19}	燃料泵功率
a_9	氧化剂泵压力增量	a_{20}	涡轮功率
a_{10}	燃料泵压力增量	a_{21}	推 力
a_{11}	燃烧室氧喷前压力	a_{22}	声速喷嘴流量

表 3 故障代号与名称

Tab.3 Name and symbol of the faults

代 号	故障名称
d_1	氧化剂主活门故障
d_2	燃料主活门故障
d_3	氧化剂喷注器阻塞故障
d_4	燃气发生器氧化剂管路过滤网阻塞故障
d_5	氧化剂泵效率降低
d_6	燃料泵效率降低

步骤 1：参数离散化。参数的离散化处理非常重要，它直接关系到特征选择的计算量和结果的可用性。这里对发动机诊断决策表的离散化采用如下方法：对于参数 a 的正常工况值 $V_n(a)$和当前工况值 $V_c(a)$，令 $\Delta V(a)=V_c(a)-V_n(a)$为故障导致的偏离量，阀值为 Thr，则其离散化后对应的值为 $S(a)$，表示如下：

$$S(a)=\begin{cases}1 & if\quad \Delta V(a)/V_n(a)>Thr \\ 0 & if\quad \Delta V(a)/V_n(a)|\leqslant Thr \\ -1 & if\quad \Delta V(a)/V_n(a)|\leqslant -Thr\end{cases} \tag{7}$$

这里，将阀值 Thr 设置为 0.5%。阀值的设置对诊断效果有一定影响，在实际诊断应用中，若阀值设置过小将会对检测数据的精度提出更高的要求，阀值设置过大将会影响诊断的灵敏度。在此，对该问题不作展开分析。

步骤 2：建立诊断决策表。

表 4 中包含了发动机参数和对应的故障症状。每一行可看作是一条故障记录 (共计 40 条记录，这里作了省略)。诊断决策表可认为是一个最小代价模糊决策系统。其中发动机参数对应条件属性，故障名对应决策属性。表中的最下面一行所示，每个发动机参数对应一个权值，可依据该参数对检测精度等因素为其赋权值。决策属性对应 0 到 1 之间的模糊值，它反映了故障程度。从表 4 中可观察到，一条故障记录可以允许对应 1 个以上的故障症状，这样使诊断决策表更加合理，更能反映实际。毕竟由于故障间的种种联系，几种故障同时出现也是存在的。

表 4 故障诊断决策表

Tab.4 Decision table for fault diagnosis

a_1	a_2	…	a_{14}	…	a_{19}	a_{22}	d_1	d_2	d_3	d_4	d_5	d_6
-1	1	…	1	…	1	0	1	0	0	0	0	0
-1	1	…	1	…	0	-1	1	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…		…
-1	1	…	1	…	0	0	0.6	0	1	0	0	0
…	…	…	…	…	…	…	…	…	…	…		…
0.7	0.6	…	1.	…	1.6	1.4	左侧为属性权值 w_1-w_{22}					

万方数据

步骤 3：采用贪婪算法进行特征优选。

对于表 4 进行特征选择获得的最优解为 $\{a_{11}, a_{13}, a_{15}, a_{18}, a_{19}, a_{20}, a_{21}\}$，对应的最小代价为 8.3，占所有条件属性总代价的 31.6%。经验结果证符合实验要求。可由此得出诊断规则，例如对于氧化剂喷注器阻塞故障 d_3 对应的诊断规则为：

$$R_{13}=\{1,0,1,-1,0,-1,-1\}\rightarrow\{d_3\},\ M_B(R_{13})=d_3$$

$$R_{17}=\{1,0,1,-1,1,0,-1\}\rightarrow\{d_3\},\ M_B(R_{17})=d_3$$

$$R_{19}=\{1,-1,1,-1,1,0,-1\}\rightarrow\{d_3,\ d_1(0.6)\},\ M_B(R_{19})=d_3$$

从中可分析出，该故障的显著特征是燃烧室氧喷前压力增大，氧化剂泵功率降低，发动机推力降低。这些特征与文献[3]所描述的故障征兆吻合。第 3 条规则给出：当满足规则前件时故障 d_3 或 d_1 出现，故障 d_1 对应故障程度为 0.6。其中，M_B $(R_{19})=d_3$ 表示该条规则对应的最严重的故障为 d_3，应在诊断中给予优先关注。

有关其它故障诊断规则的提取和分析，均可参照上述故障 d_3 的实例进行，这里不予赘述。

3.2 几点说明

如果两条记录的条件属性相同但决策属性不同，则认为两条记录是不相容的。由于各种干扰等因素的影响，这种情况在实际中也时有发生。在基本的粗糙集模型中这种情况是不允许的，而最小代价模糊决策模型允许这种情况的存在，特征选择算法实际是遵循不漏报原则进行了处理，该过程体现在式(2)中。

特征选择前后，决策属性对应的各个故障之间的逻辑关系保持不变。即故障之间是“or”还是“and”逻辑关系不变。

在实际应用中，模型中的权值可以根据需要灵活定义。例如，可以定义为重要性，模型的求解依旧采用贪婪算法。模型中的决策属性可根据故障程度赋值，如轻微、一般、严重等，尽量不要划分过细。

4 结论

针对发动机故障诊断的实际，提出了一种最小代价模糊决策模型，它具有两个特点：一是能够对属性的代价进行描述；二是决策的重要性通过模糊隶属函数进行描述。因此，在特征优选中，能够使监测参数的代价 (或重要性)、故障程度、甚至故障间的关联性均在模型中有所体现。这些特点是粗糙集模型和其它特征选择方法不具备的。该模型在应用中具有很强的灵活性，基本的粗糙集模型可视为该模型的特殊形式。

参考文献：

[1] 张惠军. 液体火箭发动机故障检测与诊断技术综述[J]. 火箭推进, 2004, 30(5): 40–45.

[2] 刘冰, 张育林. 液体火箭发动机故障诊断中的特征选择[J]. 推进技术, 1998, 19(2): 5–8.

[3] 张育林, 吴建军, 朱恒伟, 等. 液体火箭发动机健康监控技术[M]. 长沙: 国防科技大学出版社, 1998.

[4] 胡小平, 张丽娟, 王艳梅, 等. 液体火箭发动机故障检测和诊断中数据挖掘策略的分析[J]. 国防科技大学学报, 2005, 27(3): 1–5.

[5] 乔斌. 粗糙集理论分层递阶约简算法的研究[D]. 杭州: 浙江大学, 2003.

[6] 张文修, 仇国芳. 基于粗糙集的不确定决策[M]. 北京: 清华大学出版社, 2005.

[7] Zhang Mei, Wu Weizhi. Knowledge Reduction in Information Systems with Fuzzy Decisions [J]. Journal of Engineering Mathematics. 2003, 20(2): 53–58.

(编辑：王建喜)

万方数据