

# 贝叶斯分类器在液体火箭发动机 故障诊断中的应用

李京浩, 胡小平, 韩泉东

(国防科技大学 航天与材料工程学院, 湖南, 长沙 410073)

**摘 要:** 研究数据挖掘技术在火箭发动机故障诊断中的应用, 利用两种典型的贝叶斯分类器——朴素贝叶斯分类器和 TAN 分类器对液体火箭发动机故障进行分类, 对某型号液体火箭发动机的试车数据和仿真数据进行了故障诊断, 结果和实际试车情况相符, 从而验证了贝叶斯分类器可以应用于液体火箭发动机故障诊断。

**关键词:** 数据挖掘; 贝叶斯分类器; 液体火箭发动机; 故障诊断

**中图分类号:** V434

**文献标识码:** A

**文章编号:** (2007) 06-0012-06

## Application of Bayesian classifier to fault diagnosis of liquid propellant rocket engine

Li Jinghao, Hu Xiaoping, Han Quandong

(Inst. of Aerospace and Material Engineering, National Univ. of Defense Technology, Changsha 410073, China)

**Abstract:** The application of data mining in fault diagnosis of rocket engine is analyzed in this paper. Two representative Bayesian classifier——Naive Bayesian classifier and TAN classifier are used to calssify the faults of liquid propellant rocket engine, and the trial run data of a rocket engine and simulation data are diagnosed. The results is correspond to real hot run. Therefore, the verified Bayesian classifier can be used in fault diagnosis of liquid rocket engine.

**Key words:** date mining; Bayesian classifier; fault diagnosis; liquid propellant rocket engine

收稿日期: 2007-07-25; 修回日期: 2007-10-26。基金项目: 国家自然科学基金资助项目(50376073)。

作者简介: 李京浩 (1980—), 男, 硕士, 研究领域为数据挖掘、火箭发动机故障诊断。

## 1 引言

数据挖掘是从大量的数据中提取隐含在其中人们事先未知的、但潜在有用的信息和知识的过程<sup>[1]</sup>。分类在数据挖掘中是一项非常重要的任务, 分类的目的是学会一个分类函数或分类模型 (也称作分类器)。

介绍了贝叶斯分类器中具有代表性的两个分类器——朴素贝叶斯分类器和 TAN (Tree-Augmented Naive Bayesian) 分类器, 并通过对火箭发动机数据的挖掘, 测试验证了两种分类方法的可行性和分类精度。

## 2 贝叶斯分类器

分类通常被认为是把一组事物分成子集合, 而子集合的成员相互之间比其他成员之间具有更大的“相似性”。

贝叶斯分类器<sup>[2]</sup>是一种典型的基于统计方法的分类方法, 它可以预测类成员关系的可能性, 如给定样本属于一个特定类的概率。分类包括规则分类 (查询) 和非规则分类 (有指导学习)。贝叶斯分类是非规则分类, 是通过训练集 (已分类的实例集) 训练 (学习) 而归纳出分类器, 利用分类器对没有分类的数据进行分类。

贝叶斯分类具有如下三个特点:

(1) 贝叶斯分类并不把一个对象绝对地指派给某一类, 而是通过计算得出属于某一类的概率, 具有最大概率的类便是该对象所属的类;

(2) 一般情况下在贝叶斯分类中所有的属性都潜在地起作用, 即并不是一个或几个属性决定分类, 而是所有的属性都参与分类;

(3) 贝叶斯分类的对象的属性可以是离散的、连续的、也可以是混合的。

贝叶斯分类器的这些特点很适合液体火箭发动机的故障诊断。

### 2.1 朴素贝叶斯分类器

Duda 和 Hart 于 1973 提出了基于贝叶斯公式的朴素贝叶斯分类器 NBC (Naive Bayesian Classi-

fier)<sup>[3]</sup>。NBC 是一个简单有效而且在实际使用中比较成功的分类器。其结构如图 1 所示。

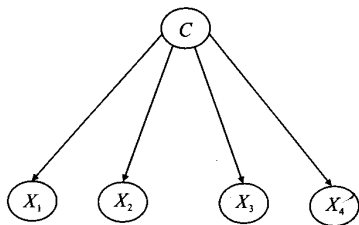


图 1 朴素贝叶斯模型

Fig.1 The model of Naive Bayes

朴素贝叶斯分类假定一个属性值对给定类的影响独立于其他属性的值。这一假设称作类条件独立。做此假定是为了简化所需的计算, 并在此意义下称为“朴素的”。

设有变量集  $U=\{A_1, \dots, A_n, C\}$ , 其中  $A_1, \dots, A_n$  为实例的属性变量;  $C$  为取  $m$  个值的类变量。假设所有的属性都条件独立于类变量  $C$ 。

分类问题描述为: 给定一组训练  $D_1, \dots, D_p$ , 其中  $D_k$  ( $k=1, 2, \dots, p$ ) 由  $U$  中的变量组  $\{A_1, \dots, A_n, C\}$  的一组值描述。给定一个新的具有属性  $a_1, \dots, a_n$  的实例, 判别这个新实例的类别。

使用朴素贝叶斯分类器进行分类的做法是: 通过概率计算, 从待分类实例的属性值  $a_1, \dots, a_n$  求出最可能的分类目标值。即计算各类  $c_j \in C$  对于这组属性的条件概率  $P(c_j | a_1, \dots, a_n)$ , 其中  $j=1, 2, \dots, m$ , 并输出条件概率最大的类标签作为目标值。应用贝叶斯定理和条件独立假设:

$$\begin{aligned} P(c_j | a_1, \dots, a_n) &= \frac{P(c_j | a_1, \dots, a_n) P(c_j)}{P(a_1, \dots, a_n)} \\ &= a \cdot P(c_j) \cdot \prod_{i=1}^n P(a_i | c_j) \end{aligned} \quad (1)$$

式中,  $a$  为正规化常数。以后验概率作为分类指标, 即输出具有最大后验概率的类标签  $C_{NB}$

$$C_{NB} = \arg \max P(c_j) \cdot \prod_{i=1}^n P(a_i | c_j) \quad (2)$$

式中,  $C_{NB}$  为朴素贝叶斯分类器输出的目标值;

常数  $a$  可以省略。上式通常作为朴素贝叶斯分类器的定义，实际计算时，式中项  $P(c_j)$  可以通过计算训练例中  $c_j$  出现的频率来估计， $P(a_i|c_j)$  的数目等于属性的数目乘以类的数目，也可以通过计算训练例中出现的频率来估计。

2.2 TAN 分类器

TAN 分类器<sup>[4]</sup>是由 Nir Friedman 等首次提出的，是对朴素贝叶斯分类器的扩展。TAN 分类器是树扩展朴素贝叶斯网络 (Tree-Augmented Naive Bayesian Network) 的简称，是以类变量为根结点，每个属性变量以类变量和最多一个属性变量为父结点的贝叶斯网络。其基本思想是在朴素贝叶斯分类器的基础上，在属性之间增添连接弧，以消除朴素贝叶斯分类器关于类条件独立的假设，其中  $\{X_1, \dots, X_n\}$  构成一棵树，如图 2 所示。

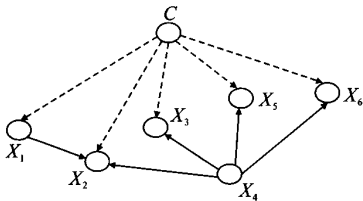


图 2 TAN 分类模型示意图  
Fig.2 The model of TAN

TAN 贝叶斯网络被这棵树所惟一确定，而树可被函数  $\pi: \{1, \dots, n\} \rightarrow \{0, 1, \dots, n\}$  (使  $\pi(i)=0$  的结点为父结点，不存在序列  $i_1, \dots, i_k$  使  $\pi(i_j)=i_{j+1}$ ，其中  $i \leq j < k$ ，而且  $\pi(i_k)=i_1$ ) 所确定，当  $\pi(i) > 1$  时， $\Pi_{\pi_i} = \{X_{\pi(i)}\}$ ，当  $\pi(i)=0$  时  $\Pi_{\pi_i} = \Phi$ ，因此，函数  $\pi$  就定义了 TAN 贝叶斯网络，给定属性结点之间的条件互信息函数：

$$I_p(X; Y | Z) = \sum_{x,y,z} P(x,y,z) \cdot \log \frac{P(x,y|z)}{P(x|z)P(y|z)} \tag{3}$$

TAN 贝叶斯网络构造过程如下：

- (1) 通过训练集计算属性对之间的条件互信息  $I_{p_i}(X_i; X_j | C)$ ；
- (2) 建立一个以  $I_{p_i}(X_i; X_j | C)$  为弧的权重的加权完全无向图；
- (3) 建立一个最大权重跨度树；

(4) 选择一个根结点，设置所有边的方向是由根结点向外，把无向树转换为有向树；

(5) 增加一个类变量结点及类变量结点与属性结点之间的弧。

建立最大权重跨度树的方法是：先把边按权重由大到小排序，后遵照选择的边不能构成回路的原则，按照边的权重由大到小的顺序选择边，这样由所选择的边构成的树便是最大权重跨度树。

3 仿真试验及结果

3.1 数据预处理

数据预处理是数据挖掘的一个极其重要的环节。液体火箭发动机试车测得的参数种类很多，数据量非常大，数据采样的步长也可能不一致，这些给数据的选取带来了极大困难。

这里通过选取多次试车数据中共有的且参数值相对稳定的数据，从 80 多个测量参数中初步提取出了 30 个参数的数据。然后采用属性选择的方法（主要选取与易发生故障的组件所对应的参数），将上述 30 个参数进一步压缩到 19 个，用英文字母表示。其具体情况见参考文献 [5]。

试车实测数据可能变化较大，为了减小阈值的搜索范围以提高数据挖掘的效率，采用了

$(X_{ij} - \bar{X}_j) / \bar{X}_j$  形式的函数把每个数据转换成与正常数据的统计均值的偏差值，式中， $X_{ij}$  表示第  $i$  个样本中第  $j$  个参数的测量值，表示该型号发动机多次正常试车稳态段第  $j$  个参数的统计均值。

3.2 利用贝叶斯分类器进行故障分类

用来挖掘的数据是某大型氢氧发动机的 23 次热试车稳态段的数据（即 3 次故障试车数据和 20 次正常试车数据，共 76005 个样本）。三组稳态故障数据的故障类型分别为氧副文氏管出现多余物、氢涡轮破坏以及氢泵次同步振动。这里从三次故障数据中选择故障发生段的数据作为故障样本，把所有正常试车数据作为正常数据样本。该样本集中，测量参数是实例的属性变量，故障类别是实例类变量。因此，样本集拥有上述 19 个属性变量和 4 个类变量——氧副文氏管出现多余物；氢涡轮破坏；氢泵同步振动；正常。

表 1 对试车数据训练集进行分类结果  
Tab.1 The result of classifying training data

数据	朴素贝叶斯分类器	TAN 分类器
训练生成模型 所需的时间	0.33 秒	2.16 秒
错分矩阵	a b c d <--被分到	a b c d <--被分到
	5 0 0 0   a=氧副文氏管出现多余物	5 0 0 0   a=氧副文氏管出现多余物
	0 4 0 0   b=氢涡轮破坏	0 4 0 0   b=氢涡轮破坏
	0 0 3 0   c=氢泵同步振动	0 0 3 0   c=氢泵同步振动
	0 0 0 6617   d=正常	0 0 0 6617   d=正常

本次试验所用的工具是依据国家自然科学基金资助项目“数据挖掘和知识发现在液体火箭发动机故障诊断中的应用研究”开发的数据挖掘平台“液体火箭发动机故障检测和诊断数据挖掘系统”。

分别利用朴素贝叶斯分类器和 TAN 分类器对训练集进行分类,其结果如上表 1 所示。

生成的朴素贝叶斯模型和 TAN 分类模型如图 3 和图 4 所示。

可以看出,朴素贝叶斯分类器和 TAN 分类器都能够准确地对故障进行分类,TAN 分类器需要的时间更长一些。可见,火箭发动机的属性参数只有 19 个的时候各属性之间满足朴素贝叶斯分类器的条件独立性假设,这说明属性变量之间的依赖相对于属性变量与类变量之间的依赖是可以忽略的,这就是为什么朴素贝叶斯分类器应用的最优范围比想象的要大得多的一个主要原因。

利用训练集所形成的分类器对 23 组试车数据进行挖掘,能够准确分辨出故障数据和正常数据。如表 2 所示。

3.3 利用仿真数据挖掘验证

为了更好地验证贝叶斯分类器的性能,还采用一组液体火箭发动机仿真数据进行试验。仿真数据通过建立火箭发动机的数学模型,仿真了发动机可能发生的各种故障类型,每个样本包含火箭发动机的 69 个参数。

这里选取 8 组故障数据:氧稳压阀出口泄漏、发生器氧副控阀泄漏、发生器氢副控阀泄漏、氧涡轮入口燃气泄漏、氢涡轮入口燃气泄漏、氢涡轮出口燃气泄漏、氧泵后泄漏及氧主文氏管后泄漏等作为训练集(为方便起见分别取代号 Fault\_0 到 Fault\_7)。每种故障选取 40 个样本,则共有 320 个样本。

分类结果如表 3 所示。

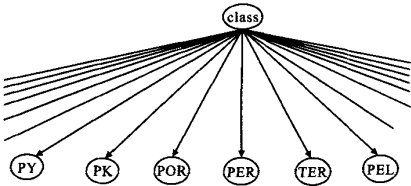


图 3 生成的朴素贝叶斯模型  
Fig.3 Naive Bayes model of data

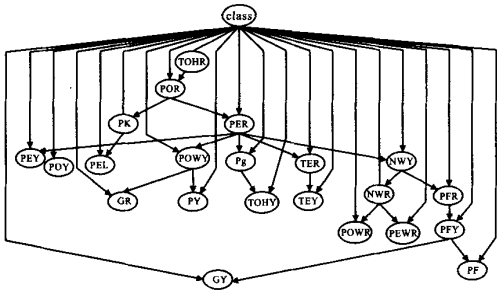


图 4 生成的 TAN 网络模型  
Fig.4 TAN model of data

表 2 故障诊断结果  
Tab.2 The result of fault diagnosis

氧副文氏管出现多余物					氢涡轮破坏				
a	b	c	d	<--被分到	a	b	c	d	<--被分到
0	0	0	0	a=氧副文氏管出现多余物	0	0	0	0	a=氧副文氏管出现多余物
0	0	0	0	b=氢涡轮破坏	0	0	0	0	b=氢涡轮破坏
0	0	0	0	c=氢泵同步振动	0	0	0	0	c=氢泵同步振动
7	0	0	27	d=正常	0	5	0	2719	d=正常
氢泵同步振动					正常				
a	b	c	d	<--被分到	a	b	c	d	<--被分到
0	0	0	0	a=氧副文氏管出现多余物	0	0	0	0	a=氧副文氏管出现多余物
0	0	0	0	b=氢涡轮破坏	0	0	0	0	b=氢涡轮破坏
0	0	0	0	c=氢泵同步振动	0	0	0	0	c=氢泵同步振动
0	0	3	270	d=正常	0	0	0	4948	d=正常

表 3 对仿真数据训练集进行分类结果  
Tab.3 The result of classifying simulate traing data

数据	朴素贝叶斯分类器									TAN 分类器								
训练生成模型所需的时间	0.36 秒									7.33 秒								
错分矩阵	a	b	c	d	e	f	g	h	<--被分到	a	b	c	d	e	f	g	h	<--被分到
	38	0	0	0	0	0	0	2	a=Fault_0	38	0	0	0	0	0	0	2	a=Fault_0
	0	26	14	0	0	0	1	0	b=Fault_1	0	39	0	0	0	0	1	0	b=Fault_1
	0	0	40	0	0	0	0	0	c=Fault_2	0	0	40	0	0	0	0	0	c=Fault_2
	0	0	0	33	0	7	0	0	d=Fault_3	0	0	0	40	0	0	0	0	d=Fault_3
	0	3	14	0	23	0	0	0	e=Fault_4	0	0	0	0	40	0	0	0	e=Fault_4
	0	0	0	31	0	9	0	0	f=Fault_5	0	0	0	6	0	34	0	0	f=Fault_5
	7	0	0	0	0	0	33	0	g=Fault_6	4	0	0	0	0	0	36	0	g=Fault_6
	0	0	0	0	0	0	0	40	h=Fault_7	0	0	0	0	0	0	0	40	h=Fault_7

3.4 结果分析

可以看出当火箭发动机的属性参数增加到 69 个时，随着属性数目的增加，各属性间的条件独立性假设不能满足。因此，利用朴素贝叶斯分类器进行分类，分类的误差比较大，不能满足分类的需求。TAN 分类器虽然也有误差，但是其分类精度明显比朴素贝叶斯分类器要好。

TAN 分类器是朴素贝叶斯分类器进行有效改进的分类器，它既有朴素贝叶斯分类器的简单

性，又有比朴素贝叶斯分类器更好的分类性能。但是也因为如此，TAN 分类器进行所需要的时间也比朴素贝叶斯分类器多很多。

以上试验可以看出贝叶斯分类器可挖掘液体火箭发动机数据中的故障数据并进行分类，与参考文献 [5] 采用的数值型关联规则算法及参考文献 [6] 采用的支持向量机方法进行对比，所得结果也保持一致，故该方法可信。

(下转第 42 页)

从冷却通道的应力场可以看出——冷却通道内的应力分布有以下特点:

(1) 在沿扩压器横截面上, 各冷却通道的应力分布具有对称性, 这与扩压器几何形状的对称性相吻合。

(2) 在沿扩压器轴向上, 应力在扩压器出水端附近最大, 这一特点符合扩压器的受力情况, 与经典材料力学分析得到的结论一致。

(3) 单个冷却通道内, 应力在槽形撑与内筒交界处达到最大, 应力值最大为 19.2MPa, 此处也是扩压器整体上的应力最大值点。

## 5 结束语

在进行发动机的高模试验时, 扩压器所承受的力学环境十分恶劣, 而扩压器能否在这种环境下安全可靠地工作, 是关系到发动机高模试验成败的重要问题。扩压器作为大型结构件, 结构强

度是保证其结构可靠性最重要的指标。

本文利用大型有限元分析软件 ANSYS9.0, 对扩压器筒体结构, 建立了较为合理可行的有限元计算模型, 选取壳单元 SHELL63, 采用映射网格划分方法, 通过静态强度分析, 准确形象地描述出扩压器筒体的应力分布情况。

扩压器筒体的应力场分布符合其受力情况, 与经典材料力学分析得到的结论一致。对于筒体结构中应力集中的位置, 采取了增加加强筋的办法来提高结构强度, 经过实际使用效果很好。在发动机高模试验时, 高温高速燃气对扩压器结构有很大影响, 因此扩压器结构在高温高速燃气下的结构强度分析是今后研究的重点之一。

## 参考文献:

- [1] 任重. ANSYS 实用分析教程[M]. 北京: 北京大学出版社, 2003.

(编辑: 王建喜)

(上接第 16 页)

## 4 结束语

利用数据挖掘可以挖掘液体火箭发动机的故障, 贝叶斯分类是数据挖掘的重要分支。介绍了贝叶斯分类中重要的两种分类器: 朴素贝叶斯分类器和 TAN 分类器, 并利用这两种方法对某型液体火箭发动机的试车数据和基于数学模型建立的仿真数据进行故障分类, 取得了很好的效果。很好地对比和验证了两种分类方法的分类精度。TAN 分类器是一种很好的分类器, 但是, 当数据库中的数据量十分庞大时, TAN 分类器的使用就受到了限制。在实际应用大规模数据库的时候, 贝叶斯分类器还有很多待究的问题。给出有效实用的贝叶斯网络分类器结构学习方法与算法是进一步研究的目标。

## 参考文献:

- [1] 韩家伟. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2001.
- [2] 陈琴. 贝叶斯数据挖掘技术及其在反垃圾邮件中的应用[D]. 南宁: 广西大学, 2004.
- [3] 张剑飞. 贝叶斯网络学习方法和算法研究[D]. 长春: 东北师范大学, 2005.
- [4] 周颜军, 王双成, 王辉. 基于贝叶斯网络的分类器研究[J]. 东北师大学报(自然科学版), 2003, 35(2): 21-27.
- [5] 李京浩, 胡小平, 韩泉东. 液体火箭发动机故障的数值型关联规则挖掘[J]. 火箭推进, 2007, 33(2): 7-11.
- [6] 韩泉东, 胡小平, 李舟军. 决策树和支持向量机方法在液体火箭发动机故障诊断中的应用[J]. 计算机科学, 2006, 33(8): 275-277.

(编辑: 马 杰)