

文章编号: 1000-8055(2008)08-1432-06

基于乏信息系统的本征融合技术

王雅红¹, 夏新涛², 王中宇³

(1. 洛阳理工学院 机电工程系, 洛阳 471023;

2. 河南科技大学 机电工程学院, 洛阳 471003;

3. 北京航空航天大学 仪器科学与光电工程学院, 北京 100191)

摘 要: 提出基于乏信息系统的本征融合技术以模拟小样本条件下的未知分布. 采用隶属函数法、最大隶属度法、自助法、灰自助法、算术平均值法和最大熵法等多种数学思想, 从不同侧面分析各个个体的特征信息并进行融合, 推断出总体的概率分布. 滚动轴承噪声的试验研究表明, 本征融合技术可以有效地模拟乏信息系统的概率分布.

关 键 词: 滚动轴承; 噪声; 概率分布; 小样本

中图分类号: TH 133 33; N 495

文献标识码: A

Eigen fusion technique based on information poor system

WANG Ya hong¹, XIA Xin tao², WANG Zhong yu³

(1. Department of Mechatronical Engineering,

Luoyang Institute of Science and Technology, Luoyang, 471023 China;

2. College of Mechatronical Engineering,

Henan University of Science and Technology, Luoyang, 471003 China;

3. School of Instrumental Science and Opto-Electronics Engineering,
Beijing University of Aeronautics and Astronautics, Beijing, 100191 China)

Abstract: The eigen fusion technique based on the information poor system was proposed to simulate an unknown distribution under the condition of small sample. Analyzing and fusing the characteristic information of every individual from different view sides, the probability distribution of a population was inferred, by using multifold mathematical ideas such as membership function, maximum membership grade, bootstrap, grey bootstrap, arithmetic mean, and maximum entropy. The experiment on rolling bearing noise shows that the eigen fusion technique can effectively simulate the probability distribution of the information poor system.

Key words: rolling bearing; noise; probability distribution; small sample

乏信息又叫贫信息, 是指研究对象呈现的特征信息不完备与不充分. 例如, 概率分布已知, 但仅有几个数据可供参考, 就属于乏信息问题; 无概率分布信息, 只有少量数据, 属于严重乏信息问题; 趋势项规律未知也属于乏信息范畴.

在某些滚动轴承性能试验分析中, 常常难以得到大量的数据. 比如一些新型航空、航天轴承的研制与生产, 由于品种多批量小, 每个品种每次只需求十几套甚至几套, 因此能用于性能试验的更少; 又因为缺乏概率分布和趋势项的先验信息, 所

收稿日期: 2007-11-09; 修订日期: 2008-05-16

基金项目: 国家自然科学基金项目(50375011; 50675011)

作者简介: 王雅红(1963-), 女, 河南洛阳人, 副教授. 主要从事滚动轴承与数据处理等方面的研究.

以不能用流行方法解决这类轴承的试验分析与评估问题。

目前的许多信息融合技术^[1-2]主要是针对真值估计而提出的, 且有很多准则。准则不同, 融合结果亦不同, 仅是对真值的某个侧面描述, 尚难以描述总体的概率分布。自助法^[3]虽然可以模拟未知分布, 但不能获取比原始数据更多的信息, 当数据个数很少时, 所产生的误差可以是无限的。最大熵法^[4]可以得到最无偏的概率分布估计, 但需要大量数据。

为了解决上述问题, 本文以滚动轴承噪声为研究对象, 提出一种新的信息融合技术——基于乏信息的本征融合技术。基于乏信息系统, 该技术的基本思想有以下两点:

1) 对个体的识别。对个体的识别是从个体表现出的少量数据中挖掘有用信息。由于信息不完备, 必须从多个侧面和不同角度去挖掘, 采用多种数学思想去分析每个个体的基本特征, 得到个体特征的集合。

2) 对总体的推断。对总体的推断实际上是对个体特征信息的融合而做出的决策。测量理论认为, 被测对象的真值总是唯一地客观存在的, 但很难获得。对个体的识别, 得到了每个个体特征信息, 但并不能代表总体的特征信息, 因而不能直接用于对总体的预报。也就是说, 各个个体的原始数据不同, 会导致它们具有不同的特征信息。但是, 从贝叶斯统计学的观点看, 它们的特征信息参数都可以看作随机变量, 这些随机变量虽然不等于总体特征参数的真值, 而只是总体真值的一个随机实现, 但隐含着总体真值的一些信息。因此可以对这些信息进行融合, 最终推断出总体特征信息。

1 数学模型

1.1 基本定义

本征融合是指, 在给定的论域 U 中, 设知识集为

$$F = (f^1, f^2, \dots, f^i, \dots, f^m) \quad (1)$$

知识集的本征信息子集为

$$f^i = (f_{i1}, f_{i2}, \dots, f_{ij}, \dots, f_{in}) \quad (2)$$

非空的乏信息集(序列)为

$$X = (x_1, x_2, \dots, x_l, \dots, x_w) \neq \Phi \quad (3)$$

本征融合序列为

$$X_{00} = (X_{01}, X_{02}, \dots, X_{0j}, \dots, X_{0n}) \quad (4)$$

问题集为

$$Q = (q^1, q^2, \dots, q^k, \dots, q^h) \quad (5)$$

记“属性一致性于”为符号“ $\subseteq$ ”, 在知识集 F 中, 存在且至少存在两个与问题集 Q 的一个或多个元素有某种映射关系的集合, 是满足非空信息集 X 的最终解 X_0 , 表示为

$$X_0 \mid X \text{ AND } Q \mid \text{Com_Fusion } F \text{ AND } X_{00} \mid f_i \subseteq F_0 \quad (6)$$

式中 F_0 : 系统属性的真值集合; $\mid X$: 在非空的乏信息集 X 下; $\text{Com_Fusion } F \text{ AND } X_{00}$: 与 F 和 X_{00} 有某种属性联合关系的融合; $\text{AND } Q$: 且针对与包含问题集 Q ; f_{ij} : 知识元素; $\mid f_i$: 依托于 f_i 。

一个乏信息系统的发展经历了多个重要阶段, 两个相临阶段的过渡状态称为通道。通道是系统发展的关键环节。由于信息的缺乏或严重缺乏, 通道被堵塞, 几乎没有信息流动。仅用一种数学工具难以打通所有通道, 必须根据不同的通道状态和特征, 采用不同的数学工具予以打通。实际上, 乏信息的本征融合是指, 将两种及两种以上数学思想有机地结合起来, 取长补短, 形成一种新的方法, 对系统进行分析, 直接得到总体特征。

1.2 融合方法

设有轴承噪声的信息集 X (单位: dB)

$$X = (x_1, x_2, \dots, x_l, \dots, x_w) \quad (7)$$

式中 x_l : 第 l 个数据, $l = 1, 2, \dots, w$ 。

用 m 种知识 f_i 估计出轴承噪声真值的 m 个数据序列 X_i

$$X_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in}); i = 1, 2, \dots, m \quad (8)$$

式中 X_i : 用第 i 种方法估计出的数据序列; x_{ij} : X_i 的第 j 个数据; n : X_i 的数据个数; m : m 种方法即共有 m 个数据序列。

因为是对同一对象进行的研究, 所以可将 m 个数据序列作为一个整体, 从而构成一个矩阵

$$\begin{bmatrix} X_1 & X_2 & \dots & X_i & \dots & X_{m-1} & X_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1,n-1} & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2,n-1} & x_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{i,n-1} & x_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{m-1,1} & x_{m-1,2} & \dots & x_{m-1,j} & \dots & x_{m-1,n-1} & x_{m-1,n} \\ x_{m1} & x_{m2} & \dots & x_{mj} & \dots & x_{m,n-1} & x_{mn} \end{bmatrix}_{m \times n} \quad (9)$$

从式(9)的矩阵中, 提取第 j 列向量, 转置后设为

$$\mathbf{Y}_j = \mathbf{X}_j^T = (x_{j1}, x_{j2}, \dots, x_{ji}, \dots, x_{jm})$$
$$i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (10)$$

用灰自助法^[5]对式(10)进行处理,对第*i*种方法,可以模拟出*B*个数据

$$\mathbf{X}_{\text{Bootstrap}} = (x_{1i}, x_{2i}, \dots, x_{bi}, \dots, x_{Bi}); b = 1, 2, \dots, B$$
$$(11)$$

由于*B*可以是一个很大的数,因此可以将式(11)数据从小到大排序,按一定的间隔分成*G*组,根据统计方法画直方图,就得到第*i*种方法估计出的频率值*F_i*

$$\omega_i = \omega_i(x_g); i = 1, 2, \dots, m; g = 1, 2, \dots, G$$
$$(12)$$

式中*x_g*:第*g*组中值.

上述研究中,*Y_j*是用各种方法对总体噪声真值估计的各个侧面信息,*ω*是基于各种方法的噪声总体分布的各个侧面信息.这些结果仅描述了总体的各个局部特征.

对*Y_j*进行一次融合处理,就得到*Y_j*的融合值*X_{00,j}*,于是有

$$\mathbf{X}_{00} = (X_{00,1}, X_{00,2}, \dots, X_{00,j}, \dots, X_{00,n}) \quad (13)$$

式中*X₀₀*:对总体噪声真值的估计序列.

再用灰自助法^[5]对式(13)进行处理,可以模拟出*B*个数据

$$\mathbf{X}_{\text{Bootstrap}} = (x_1, x_2, \dots, x_b, \dots, x_B); b = 1, 2, \dots, B$$
$$(14)$$

将式(14)数据从小到大排序,按一定的间隔分成*G*组,根据统计方法画直方图,就得到频率值

$$\omega = \omega(x_g); g = 1, 2, \dots, G \quad (15)$$

最后用最大熵法和基于区间映射的牛顿迭代法^[6],可以推断出噪声总体的概率密度函数

$$f(x) = \exp\left[\sum_{k=0}^M \lambda_k (ax + b)^k\right] \quad (16)$$

式中*M*:原点矩阶数;*λ_k*:第*k*个拉格朗日乘子;*a*,*b*:区间映射系数^[6].

噪声总体的概率分布函数为

$$F(x) = \int_{-\infty}^x f(x) \mathrm{d}x \quad (17)$$

式(13)~式(16)是对噪声总体的推断.

2 试验研究

有一小批稳定生产的滚动轴承,从中随机抽取6套进行噪声分析.

轴承噪声测量是在消声室进行的,采用了日本JISB1548标准.背景噪声为15~20 dB,比测量对象噪声低10~15 dB.驱动装置置于消声室外,轴承内圈转速为1 800 r/min,轴向载荷为20 N,传声器距离为70 mm,方向角为45°.测量设备与仪器见表1,噪声的测量方法见图1^[6].

表 1 噪声测量装置

Table 1 Measuring device of noise		
名称	型号	生产厂家
声压计	2 209	B & K
传声器	4 165	B & K
频谱分析仪	3 582A	HP
测速器	HT-446	上海转速表厂
静止变频器	ZYS NBJ10/350/2000	洛阳轴承研究所
驱动装置	—	洛阳轴承研究所

在测量轴承噪声时,为了准确地评价一套轴承的噪声声压级,采用正反两面测量.每面测量3次,这样,每套轴承有6个噪声数据,如表2所示.试验目的是根据这6套轴承的噪声试验值,获取该类轴承噪声的概率分布特征信息.

图 1 噪声测量

Fig. 1 Measurement of noise

表 2 滚动轴承噪声试验数据(单位: dB)^[6]

Table 2 Test data of rolling bearing noise (Unit: dB) ^[6]						
轴承	测量序号					
序号	1	2	3	4	5	6
1	27.0	26.0	26.5	27.0	27.0	27.0
2	28.0	28.0	27.5	27.5	27.5	27.5
3	26.5	26.0	26.5	26.5	26.0	26.0
4	27.0	26.5	27.5	26.0	26.0	26.5
5	25.5	26.0	26.0	27.0	27.0	27.0
6	27.5	26.5	26.5	27.5	27.5	28.0

用隶属函数法^[7]、最大隶属度法^[6]、自助法^[3]、灰自助法^[5]和算术平均法等 5 种方法分别处理表 2 数据, 得到表 3。由表 3 知, 对每一套轴承而言, 方法不同, 噪声真值估计的结果亦不同。

表 3 滚动轴承噪声真值估计(单位: dB)

Table 3 Estimating for truth of rolling bearing noise (Unit: dB)

轴承序号	各个侧面 信息序列	隶属函数法 计算结果	最大隶属度法 计算结果	自助法 计算结果	灰自助法 计算结果	算术平均法 计算结果
		X_1	X_2	X_3	X_4	X_5
1	Y_1	27.000	27.000	26.698	26.684	26.750
2	Y_2	27.625	27.625	27.649	27.676	27.667
3	Y_3	26.250	26.250	26.249	26.261	26.250
4	Y_4	26.250	26.250	26.598	26.562	26.583
5	Y_5	26.536	26.667	26.497	26.446	26.417
6	Y_6	27.250	27.167	27.177	27.149	27.250

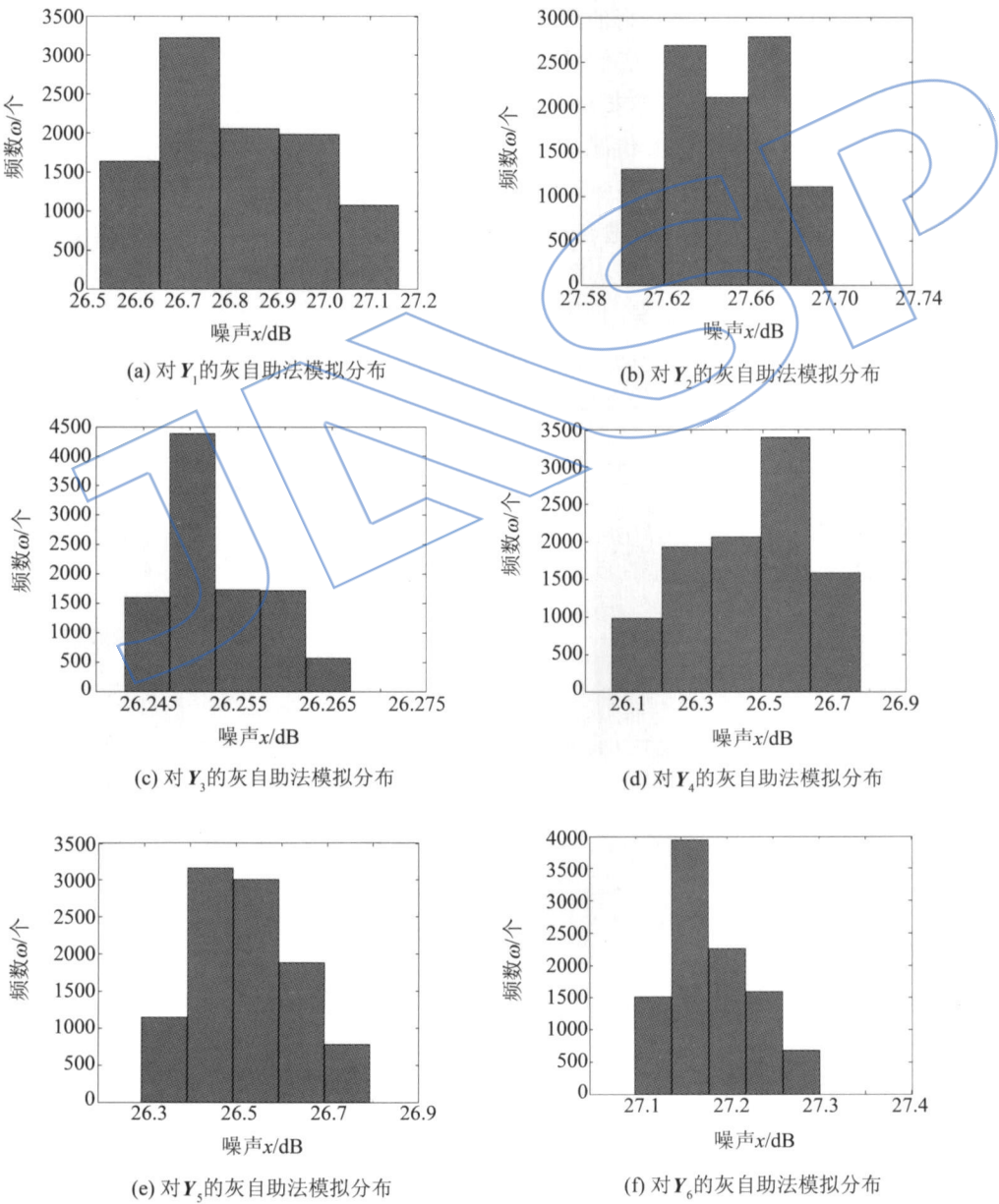


图 2 对个体的灰自助法模拟分布

Fig. 2 Simulated distribution for individual using grey bootstrap method

原因是: 在估计真值时, 隶属函数法关注数据之间的分布密度; 最大隶属度法关注数据之间分布最密集的数据; 自助法关注各个数据再抽样的等概率性; 灰自助法关注再抽样数据的发展势态与概率加权性; 最常用的算术平均法则关注各个数据的等权性.

该种轴承噪声的真值是唯一且客观存在的, 但其值是未知的, 特别是现有的研究仍然不能确认滚动轴承噪声的概率分布^[56]. 因此, 在这 5 种方法中, 不能说哪一种方法估计的结果最真实和最可信; 而只能说这 5 种方法描述了该种轴承噪声特性的 5 个不同的侧面. 若使用更多的数学方法描述, 则可以获取轴承噪声特性的更多侧面信息. 根据信息融合理论, 将关于总体特征的诸多侧面信息进行融合, 就可以获取总体特征信息的一个估计; 融合的信息越多(灰自助法要求 $m > 3$), 估计的结果就越接近总体特征的真实情况.

已知 $m=5, n=6$. 取 $B=10\ 000, G=5$, 用灰自助法对这 5 种方法的评估结果进行融合, 模拟出的各个个体的分布状态如图 2 所示, 从图 2 可以看出, 模拟出来的概率分布状态各异. 具体表现为: 图 2(a)、图 2(c)、图 2(e) 和图 2(f) 的 4 个分布状态有一定的相似性, 基本上属于左偏态分布(对

Y_1, Y_3, Y_5 和 Y_6 的描述); 图 2(b) 属于双峰分布(对 Y_2 的描述); 图 2(d) 属于右偏态分布(对 Y_4 的描述). 这些分布仅是对 $Y_1 \sim Y_6$ 中每一套轴承噪声分布的描述, 属于个体特征信息的描述, 是随机过程总体的 6 次随机实现. 因此, 得到这些不同的结果是正常现象. 而对该类轴承噪声分布的总体描述, 还需要做进一步的融合与分析.

用灰自助法对以上 5 种方法的结果进行融合时, 还可以得到估计真值的本征融合序列, 结果如下(单位: dB):

$$X_{01}=26.806; X_{02}=27.649; X_{03}=26.252X_{04}=26.472; X_{05}=26.523; X_{06}=27.186$$

即对真值的本征融合序列 X_{00} 为

$$X_{00}=(26.806, 27.649, 26.252, 26.472, 26.523, 27.186)$$

对 X_{00} 再进行一次灰自助法处理, 可推断出该种轴承噪声总体的真值估计与总体的分布特征. 此时, $m=6$. 取 $B=10\ 000, G=5$, 可以得到该种轴承噪声总体的真值估计为 $X_0=26.78\text{ dB}$.

该种轴承噪声总体的模拟分布特征如图 3 所示. 从图 3(a) 可以看出, 灰自助样本的 10 000 个模拟数据具有随机性. 从图 3(b) 和图 3(c) 可以看出, 概率密度函数不是正态分布. 这是正确的模

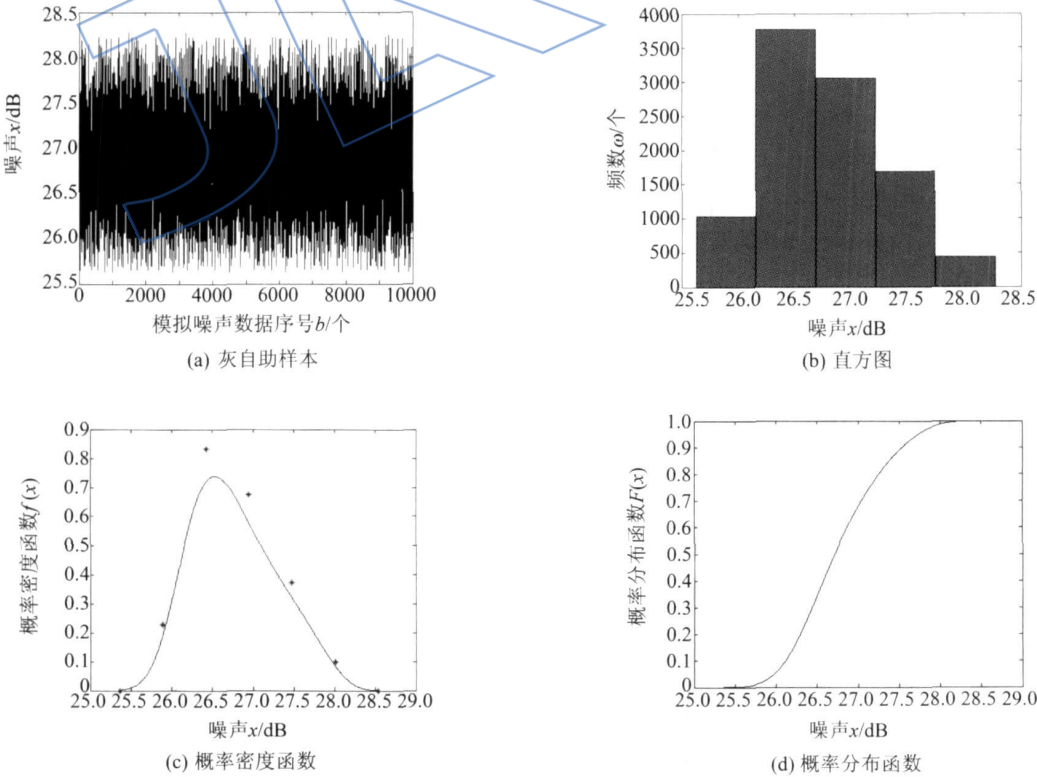


图 3 对滚动轴承噪声总体的灰自助法模拟分布

Fig. 3 Simulated distribution for population of rolling bearing noise using grey bootstrap method

拟, 因为现有的研究已经表明, 滚动轴承噪声声压级的概率分布与正态分布有一定的差距, 属于偏态分布^[6], 图 3(b) 和图 3(c) 的实际形状属于左偏态分布, 是非对称的, 也证实这种观点是正确的.

3 结束语

本征融合技术允许研究对象的概率分布未知, 也允许数据个数很少. 本征融合技术采用多种数学思想从不同的侧面分析个体的特征, 进而对个体特征信息进行融合而做出的决策, 最终推断出总体特征信息. 滚动轴承噪声的试验研究表明, 本征融合技术可以有效地模拟乏信息研究对象的分布特征.

参考文献:

[1] Hunter A, LIU Weiru. Fusion rules for merging uncertain information[J]. Information Fusion, 2006, 7: 97-134.
[2] WU Shengli, McClean S. Performance prediction of data

fusion for information retrieval[J]. Information Processing and Management, 2006, 42: 899-915.
[3] Efron B. Bootstrap methods[J]. The Annals of Statistics, 1979, 7: 1-36.
[4] Trebicki J, Sobczyk K. Maximum entropy principle and nonstationary distribution of stochastic systems [J]. Probabilistic Engineering Mechanics, 1996, 11: 169-178.
[5] 夏新涛, 陈晓阳, 张永振, 等. 滚动轴承振动的灰自启动态评估与诊断[J]. 航空动力学报, 2007, 22(1): 156-162.
XIA Xintao, CHEN Xiaoyang, ZHANG Yongzhen, et al. Dynamic assessment and diagnosis of vibration of rolling bearings using grey bootstrap[J]. Journal of Aerospace Power, 2007, 22(1): 156-162. (in Chinese).
[6] 夏新涛, 陈晓阳, 张永振, 等. 滚动轴承乏信息试验分析与评估[M]. 北京: 科学出版社, 2007.
[7] XIA Xintao, WANG Zhongyu, GAO Yongsheng. Estimation of nonstatistical uncertainty using fuzzy set theory [J]. Measurement Science and Technology, 2000, 11(4): 430-435.