文章编号:1671-4598(2016)02-0053-04

DOI:10. 16526/j. cnki. 11-4762/tp. 2016. 02. 014

中图分类号:TM73

文献标识码:A

基于概率盒的电网故障诊断方法研究

刘彦君, 王清心, 丁家满, 司怀伟, 张晋芳

(昆明理工大学, 昆明 650500)

摘要:针对电网故障信息存在丢失、误动、拒动等不确定性问题,文章采用概率盒理论和支持向量机相结合的方法对电网故障进行诊断,充分利用概率盒在处理不确定问题上的优势;首先利用概率盒对故障录波、电气量等数据建模,然后利用融合规则将得到的多个概率盒进行融合,并提取特征向量;最后,利用支持向量机进行分类,并得出诊断结果;为了验证方法的有效性,采用仿真线路进行概率盒的故障诊断,实验验证该方法合理可行,且有较高的诊断率。

关键词: 电网; 故障诊断; 概率盒; 支持向量机; 信息融合

Fault Diagnosis of Power System based on Probability-box Theory

Liu Yanjun, Wang Qingxin, Ding Jiaman, Si Huaiwei, Zhang Jinfang (Kunming University of Science and Technology, Kunming 650500, China)

Abstract: The fault information of power system often has the problems of missing information, malfunction and failure to operation, in order to solve the problems and improve the diagnosis ability, a new method based on probability-box theory and support vector machine was proposed. Firstly, the data of fault record and electrical quantities were used as the information source to construct the p-box. Then to take advantage of the complementation of the information source, the different p-box were fused and extract the feature from the p-boxes. Finally, using SVM classification and obtain the result. In order to verify the validity of the method, the simulation circuit is adopted to improve the probability box of fault diagnosis, the experimental shows that the p-box theory has high degree of diagnosis accuracy and is characterized by fast diagnosis and good real-time.

Keywords: power system; fault diagnosis; probability box; support vector machine; information fusion

0 引言

电网发生故障后,生成大量的故障信息,快速定位故障设备,缩短故障处理时间对于电网恢复是至关重要的[1]。随着电网系统日益复杂,采用单一的智能诊断方法已经很难得到精确地结果,研究将多种数据源融合的故障诊断方法具有十分重要的现实意义[2-3]。在故障定位[4]和数据信息融合等故障诊断领域[5]已经获得一些实质性进展。

故障特征提取以及故障分类是故障诊断的重要环节,决定诊断的精度和效率。故障信息存在丢失、误动、拒动等不确定性问题,影响故障诊断的精确度。常用的故障诊断方法有专家系统、人工神经网络、优化算法等[6]。在当前电网运行中,保护动作的正确率高达 98%,对于故障诊断系统,剩余 2%的小样本才是关键^[7-6]。故障信息不完善的情况在很大程度上影响了对故障诊断的准确判断,从而影响维修人员对故障的处理^[8]。概率盒(Probability-box,P-box)是用来表征不确定数组成的偶然不确定性和认知不确定性,经常用于风险分析或者

收稿日期:2015-09-01; 修回日期:2015-09-30。

基金项目:国家自然科学基金资助项目(51467007);云南省应用基础研究所计划资助项目(2013FZ020)。

作者简介:刘彦君(1990-),女,河南沁阳人,硕士研究生,主要从事概率盒处理不确定性信息、故障诊断方向的研究。

王清心(1958-),男,辽宁丹东人,教授,主要从事决策理论与方法, 模式识别等应用研究。

丁家满(1974-),男,江苏于都人,副教授,主要从事决策理论与方法,不确定信息等应用研究。

是定量的不确定性建模[10]。概率盒可以用来解决小样本、概率分布类型未知、信息源依赖关系未知、测量不确定性、遗漏测量等不确定性问题。支持向量机(Support Vector Machine,SVM)是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的,根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷,以期获得最好的泛化能力。SVM 在解决小样本、非线性及高维模式识别中表现出许多特有的优势。

本文提出了基于概率盒理论和支持向量机的电网故障诊断方法。该方法利用概率盒进行多源信息融合并提取故障特征,解决小样本、故障信息不确定的问题,提高诊断率;再利用SVM方法对故障分类,得出诊断结果。实验结果证明该方法有较高的精确度和诊断效率。

1 基于概率盒的故障特征属性的概率界限分析

1.1 概率盒

由于故障诊断中各种特征存在很大的不确定性,影响故障 诊断的效率,概率盒可以根据记录对数据进行概率界限分析, 使得指标数据在一个界限范围内,并根据该界限提取数据的一 些特征向量,并将特征向量应用于故障诊断中。

概率盒的定义为: 假设 F和 F是实数域 [0,1] 上的非递减函数,且 $F(x) \leq F(x)$, $x \in \mathbf{R}$ (R为实数域),定义为包含所有非递减函数 $F \in [0,1]$ 的集合,且 $F(x) \leq F(x) \leq F(x)$ 。当 F 和 F 能表达一个不确定的概率分布时,定义 [F(x),

F(x)] 为概率盒^[6]。对于一个随机变量 X,其概率分布 F 未知,只知道 X 必然包含在一个概率盒中,那么 F(x) 就是 F(x) 的下边界,它是 $X \le x$ 发生的概率,同理 F(x) 就是 F(x) 的上边界。由 X 低概率测量方法 P 可以计算出 F(x) 和 F(x) ,公式如下所示。

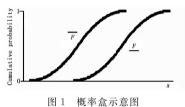
概率盒上界:

$$\bar{F}_X(x) = 1 - P(X > x) \tag{1}$$

概率盒下界:

$$F_X(x) = P(X \leqslant x) \tag{2}$$

概率盒的示意图中,左边界 F(x) 是概率分布的上边界,对应分量 X 的下边界;右边界 F(x) 是概率分布的下边界,对应分量 X 的上边界 (见图 1)。



1.2 DS 结构体

DS 结构体 (Dempster Shafer Structure, DSS) 由一个区间和一个信度组成,具体定义如下。

假设 Θ 是假设空间,在识别框架上的基本概率分配 (BPA) 是一个 $2^{\Theta} \rightarrow [0,1]$ 的函数 m ,称为 mass 函数。并且满足:

$$m(\emptyset) = 0 \coprod_{A \subseteq \Theta} m(A) = 1 \tag{3}$$

在假设空间 Θ上基于 BPA m 的信度函数 (Bel) 定义为:

$$Bel(A) = \sum_{B \in \mathcal{B}} m(B) \tag{4}$$

似然函数定义为:

$$Pls(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{5}$$

对于假设空间 Θ 中的某个假设 A,根据基本概率分配分别计算出有关该假设的信度函数和似然函数组成信任区间 $\lceil Bel(A), Pls(A) \rceil$,该区间用以表示信任程度。

基于 Dempster 规则可获取独立证据 m_1 , m_2 的组合或融合结果^[9],即

$$m(A) = \begin{cases} 0, & A = \emptyset \\ \sum_{\substack{A_i \cap B_j = A \\ 1 - K}} m_1(A_i) m_2(B_j), & A \neq \emptyset \end{cases}$$
 (6)

其中 $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)$ 为冲突项,多个证据组合时,Dempster 规则满足结合律和交换律,这有利于信息融合

时,Dempster 规则满足结合律和交换律,这有利于信息融合系统的分布式实现。

1.3 概率盒与 DSS 的关系

概率盒与 DSS 的关系可以理解为局部与整体的关系[10]。 把 DSS 的每个小区间左界累积在一起就是概率盒的左界,右 界 累 积 在 一 起 就 是 概 率 盒 的 右 界。如 $\{([x_1,y_1],m_1),([x_2,y_2],m_2)\cdots([x_n,y_n],m_n)\}$ 表示 DSS,右边界为信度函数 (Bel),左边界为似然函数 (Pls),变换可得,

$$F(z) = Bel(z) = \sum_{y_i < z} m_i$$
 (7)

$$\overline{F}(z) = Pls(z) = \sum_{x_i \le z} m_i \tag{8}$$

同一概率盒可以由不同的 DSS 组成,将概率盒分解为等信度的 DSS 的过程称为概率盒的离散化。

1.4 支持向量机故障诊断原理

一个完整的、可以处理线性和非线性,并能容忍噪音和离群点的 SVM 描述如下:

$$\min \frac{1}{2} \mid \mid \omega^{2} \mid \mid + C \sum_{i=1}^{n} \xi_{i}$$
 (9)

约束条件

$$y_i(\boldsymbol{\omega}^T x_i + b) \geqslant 1 - \boldsymbol{\xi}_i, i = 1, 2 \cdots, n$$

$$\boldsymbol{\xi}_i \geqslant 0, i = 1, 2 \cdots, n$$
 (10)

其中: ω 表示分类超平面的权系数向量; C 为惩罚因子,表示对错分样本比例和算法复杂度折衷; ξ , 为松弛变量,即数据点 x_i 允许偏离的函数间隔的量; b 为分类阈值。利用拉格朗日数乘法求解,得到原目标的对偶问题为:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_{i} - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_{i} \alpha_{j} y_{i} y_{j} \langle x_{i}, x_{j} \rangle K(x_{i}, x_{j})$$
 (11)

约束条件:

$$0 \leqslant \alpha_i \leqslant C, i = 1, 2, \dots, n$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$
(12)

其中K为核函数,主要用于将特征进行从低维到高维的转换; α ;为拉格朗日乘子。

2 基于概率盒的 SVM 故障诊断模型

2.1 构建概率盒

得到概率盒的方法有:专家估计,建模,鲁棒贝叶斯,约束缩小区间法以及测量观测法^[11]。测量观测法是根据实验数据的估计值进行累积的一种概率盒的建模方法。通过测量观测法进行概率盒指标的建模过程描述如下。

在测量观测法中通常用一个区间,如 [2.31,3.44] 表示观测的实验数据。这种区间通常会有一些说明,例如 [2.31,3.44,10%] 表示这段观测值的置信度。另外证据结构体的表示形式为 ([2.13,3.6],3.1),这 3 个数字可以表示为一个三角形,其中 [2.13,3.6] 代表间隔范围,3.1 表示峰值的最佳估计,这个三角形如图 2 中第一幅图所示。这 3 个数字不代表任何一个三角分布,只代表一个最佳点和一段区间。如图 2 第一幅图中 X 轴的三角形所示数值,三角形的宽度是不确定性数据的一种表示方法,表示数据集的可变性。图 2 第二幅图表示这些实验观测数据所得到的概率盒,横轴同样表示数据集,而纵坐标表示累积概率分布。概率盒是根据三角形的左边界和右边界两两分别做累积分布(CDF)得到的。对经验函数(EDF)进行累积,使得实验数据集中的所有数据都可以包含在概率盒中。

用简单的数据具体说明整个建模过程,假设我们有如下数据 3 ± 3 , 8 ± 5 , 11 ± 5 , 14 ± 4 , 15 ± 4 , 15 ± 4 , 16 ± 4 , 17 ± 3 , 19 ± 3 , 20 ± 3 , 21 ± 2 , 25 ± 2 , 28 ± 1 , 32 ± 1 , 38 ± 1 。在下图中用三角形的峰值表示最好的估计值,并且用峰值作为正负间隔的标准,例如: 3 ± 3 ,其中 3 作为最好估计值为三角形的峰值,它的正负区间分别为 $0\sim 3$ 、 $3\sim 6$ 。图 2 第二幅图表示用这些数据得到的概率盒。不确定变量的特征我们考虑是由正负区间导致的。上边界是累积区间的左端点,而下边界则是累积区间的右端点。综上所述,这条边界类似于经验分布函数。相关证据结构体为: $\{([0,6],1/15),([3,13],1/15),([6,16],1/15),([10,18],1/15),([11,19],2/15),([12,20],1/15),([14,20],1/15),([16,22],1/15),([17,23],1/15),([19,23],1/15),([31,33],1/15),([37,39],1/15))。$

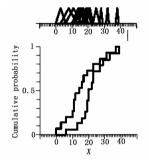


图 2 测量观测法所得概率盒示意图

当数据量足够大时边界线将是一条光滑的曲线。采用测量 观测法构建概率盒的步骤如下:

步骤 1: 观测采样。

步骤 2: 估计采样数据的峰值及其正负区间。

步骤 3:利用统计学方法获得经验分布函数以及证据结构 体以及置信区间。

步骤 4:根据累计所得的分布函数 (CDF) 或 DS 结构体构建概率盒。

如图 3 所示为非故障情况下得到的电压概率盒。

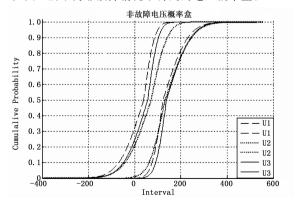


图 3 非故障情况下电压概率盒示意图

2.2 信息融合与特征提取

通过不同的方法或者在不同的条件下可以得到多个不同的 概率盒,将不同概率盒进行综合就可以得到相对完整、一致的 信息。概率盒常用的融合方法有:交集融合、均值融合、包 络、贝叶斯融合、DS证据结构体融合以及其改进算法等。本 文根据包络融合规则进行数据融合,得到非故障情况下的电 压、电流如图 4 所示。

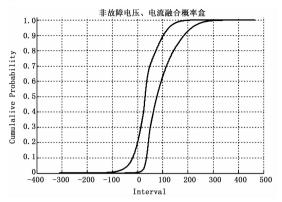


图 4 非故障电压、电流概率盒融合示意图

包络法融合概率盒原理简介如下。

假设有 n 个概率盒:

$$\begin{bmatrix} F_1(x), F_1(x) \end{bmatrix}, \begin{bmatrix} F_2(x), F_2(x) \end{bmatrix}, \cdots, \begin{bmatrix} F_n(x), F_n(x) \end{bmatrix}$$
它们的包络定义为:

$$\lceil F_n^*(x), \overline{F}_n^*(x) \rceil$$

则有

$$F_{\underline{}}^{*}(x) = \max(F_{\underline{1}}(x), F_{\underline{2}}(x), \dots, F_{\underline{n}}(x))$$

$$F^*(x) = \max(F_1(x), F_2(x), \dots, F_n(x))$$

这种方法可以扩展到 DS 结构体上,对每一个概率盒进行置信函数的累积和似然函数的累积。聚合这些概率盒然后进行离散化,即可转换到 DS 结构体。

将不同信息进行融合,提取融合后的特征量可以增加诊断精确度。累积不确定测量方法可以得到一些单一标量或者概率 盒上下界区间,用于 SVM 模式识别。结合文献 [10]、[11]、[12],对得到的概率盒进行 6 个特征向量提取,所得向量分别为:

向量1:累积宽度。

向量 2: 对数累积宽度。

向量 3: 累积区间边界值。

向量 4: 边界值。

向量 5: 矛盾区间统计。

向量 6: 置信区间。

2.3 SVM 故障诊断

随着采样频率的增加,概率盒的重叠部分随着不确定区间的减少而减小,但是仍然存在重叠部分,不可避免。采样频繁又会增加计算成本。为了解决该问题,采用支持向量机方法进行故障分类,解决概率盒分类的问题。SVM 故障分类过程如下。

步骤 1:建立训练数据和测试数据。训练数据和测试数据 应当同时、随机构造。

步骤 2: 归一化数据。将特征量归一化处理,作为 SVM 的输入。

步骤 3: 选取核函数和 SVM 所需参数。

步骤 4: 输入训练数据得到训练模型。

步骤 5: 输入测试数据,通过训练模型得到分类结果。

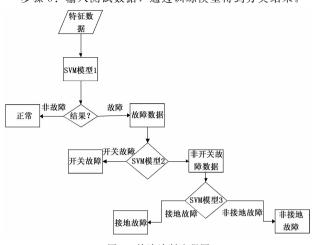


图 5 故障诊断流程图

基于 SVM 的故障诊断流程如图 5 所示,根据输入的特征向量依次诊断故障类型,每一次诊断之后,再进行一次二分类,直至识别出故障。

3 算例分析

选取电压等级为 220 kV, 长度 300 kM, 采样频率为 2 000 Hz的线路进行仿真实验。线路在 0.08 s 时发生故障, 0.24 s 故障切除。取故障前后 0.02 s 数据作为诊断数据。

针对电流、电压采用建模法构建概率盒,首先采用观测法进行采样。记录 30 s 内电压、电流的最佳估计值以及正负区间,并分为一组。出现已采样范围内的数据时采样结束,基于上述采样,得到经验分布函数和证据结构体。根据数据采样利用 2.1 节概率盒构建方法进行建模。进行分布函数累积或则取证据结构体左、右边界,即可得到概率盒。然后使用包络法进行融合。

故障情况下电压、电流概率盒融合如图 6 所示。

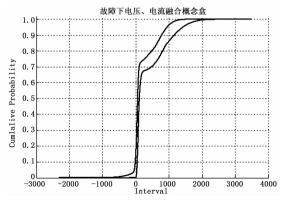


图 6 故障电压、电流概率盒融合示意图

图 6 中所示的每个概率盒都有不同形状,盒子的宽度、边界值等各不相同,这些即为概率盒的特征。选取累积宽度、对数累积宽度、累积区间边界值和边界值作为故障诊断的特征向量。

将得到的特征向量随机分为训练数据和测试数据作为

SVM 的输入。先对数据进行归一化处理,利用 SVM 分类方法进行训练,得到训练模型。再将测试数据送入训练模型,可以得到分类结果以及精确度。

SVM 常用的核函数有线性核函数、径向基核函数、多项 式核函数等,使用不同核函数得到的精确度不同,实验结果如 表1所示。

表 1 不同核函数精确度对比

核函数	精确度/%	
	故障	非故障
线性核函数	84.66	85. 12
径向基核函数	93.64	96.43
多项式核函数	85. 32	86.45

表 1 表明,使用径向基核函数精确度最高,为 93. 64%, 因此选用径向基核函数 (RBF)。在分类过程中,故障模式由 数据识别符标示。RBF 应用广泛,具有更多优势,例如线性 核函数是 RBF 的一个特例,RBF 可以将一个样本映射到更高 维的空间等。RBF 公式如下所示:

$$K(x,y) = \exp(-\gamma ||x-y||^2)$$
 (12)

其中:γ为核函数的宽度。

取电压、电流单一指标进行故障诊断,与融合后的结果进行对比,得到的诊断结果如表 2 所示。

表 2 不同指标的诊断结果对比

指标	故障	非故障
	精确度/%	精确度/%
电压	80.13	82. 27
电流	84.30	85. 22
电压、电流融合	93.64	96. 43

从表 2 可以看出,理想情况下采用概率盒融合后的实验结果更加精确,体现了使用概率盒方法进行特征提取的有效性,以及融合的必要性。概率盒在处理小样本、不确定性问题时有一定的优势,可以提高准确率。

4 结论

本文采用概率盒方法对多源故障信息进行融合、特征提取,充分利用概率盒处理不确定问题的优势,解决故障信息存在的拒动、误动等不确定性问题;并利用 SVM 方法对特征向量进行分类,得出诊断结果。实验证明该方法有较高的准确率。

结合更多故障信息,进行故障测距等诊断是本文的下个目标。

参考文献:

- [1] 高振兴,郭创新,彭建伟,等. 基于多源信息融合的电网故障诊断方法研究[J]. 电力系统保护与控制,2013,39(6):17-23.
- [2] 苏广宁, 张沛超, 胡 炎. 等. 基于多源信息的电网故障诊断新方法 [J]. 电力系统自动化, 2012, 36 (1): 61-64.
- [3] 夏可青, 陈根军, 李 力, 等. 基于多源信息融合的实时电网故障分析及实现「J]. 电力系统自动化, 2013, 37 (24); 81-88.

(下转第60页)

缩放系数为 0.69, 故障分离度达到 0.98。

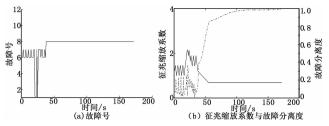


图 5 600 MW 故障 F8 验证结果

可见, 在 600 MW 额定工况协调控制方式下, 加入故障 程度与典型样本不同的低加故障,系统均能准确识别故障号 码,稳定后故障分离度能达到0.8以上。

为了验证该方法对不同负荷工况的适用性,进一步在不同 变负荷工况进行故障仿真诊断实验。

4.2 变工况故障仿真实验

4.2.1 540 MW 稳态工况故障仿真诊断

仿真机置于 540 MW 稳态工况协调控制方式。加入 5 号低 泄漏 25%的故障 (F1),实验结果如图 6 所示。结果表明,系 统在 200 s 左右准确识别出故障号码,稳定后故障系数缩放系 数为 0,22, 最终故障分离度为 0,79。

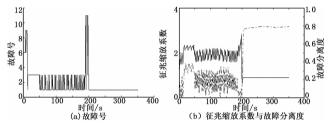


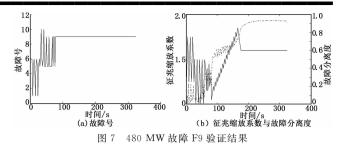
图 6 540 MW 故障 F1 验证结果

4.2.2 480 MW 稳态工况试验

仿真机置于 480 MW 稳态工况协调控制方式。加入 7A 号 低泄漏 10%的故障 (F9), 实验结果如图 7 所示。结果表明, 系统在 42 s 左右准确识别出故障号码,稳定后故障系数缩放 系数为 0.61, 最终故障分离度为 0.93。

结论

借助 600 MW 火电仿真机组系统,针对低加系统内部泄 漏和短路故障,结合征兆模糊计算,采用神经网络建立了低加



内部泄漏故障的诊断模型。将征兆缩放优化技术与神经网络诊 断模型相结合,实现了不同工况、不同程度低加泄漏故障的准 确实时诊断。仿真结果验证了本文方法的有效性。

参考文献:

- [1] 沙建峰, 刘继平, 严俊杰, 等. 低压加热器运行缺陷对经济性影响 的计算及其分析「J]. 河南电力, 2000, 27 (4): 9-10, 20.
- [2] 贾公礼, 王红梅, 胡乃文, 等. 低压加热器泄漏原因分 析及处理 [J]. 电站系统工程, 2008, 24 (3): 45-46.
- [3] 马良玉, 王兵树, 高建强, 等. 大旁路布置高压给水加热器系统故 障模糊知识库及其神经网络的诊断研究「门、动力工程,2002,22 (1): 1615 - 1621, 1588.
- [4] 曹 兴. 热力系统动态过程故障神经网络诊断方法研究 [D]. 保 定: 华北电力大学, 2012.
- [5] Frank P M. New Developments Using AI in Fault Diagnosis [J]. Engineering Application of Artificial Intelligence, 1997, 10 (1): 3 -
- [6] 马良玉, 马永光, 王兵树. 不同负荷下高加给水系统程度迥异故障 诊断的新方法「J]. 中国电机工程学报,2010,30(2):115-121.
- [7] 马良玉, 高建强, 王兵树, 等. 300 MW 机组给水高压加热 器系 统故障的模糊模式识别「J7. 中国电力,2002,35 (6):36-39.
- [8] 马良玉. 结合仿真技术的电站热力系统故障智能诊断研究 [D]. 保 定: 华北电力大学, 2004.
- [9] Xiaoyong C, Dejian Z, Zhaohua W, et al. Study on intelligent detecting technology for solder joint quality of SMT based on fuzzy diagnosis technique [A]. 11th International Conference on Electronic Packaging Technology & High Density Packaging (ICEPT-HDP) [C]. IEEE, 2010: 989 - 992.
- [10] Hatem A. Darwish. Development and implementation of an ANNbased fault diagnosis scheme for generator winding protection [J]. IEEE Transactions on Power Delivery, 2001, 16 (2): 208 - 214.

(上接第56页)

- [4] 杨 毅,曾祥君,徐舜等. 基于分布式行波检测的配电网单相接地 故定位方法 [J]. 电测与仪表, 2015, 52 (5): 59-64.
- [5] 郭创新, 彭明伟, 刘 毅. 多数据源信息融合的电网故障诊断新方 法「J]. 中国电机工程学报, 2009, 29 (31): 1-7.
- [6] 张 旭,魏 娟,赵冬梅,等. 电网故障诊断的研究里程及展望 「J7. 电网技术, 2013, 37 (10): 2746-2753.
- [7] 张 燕, 佘 维, 李 平, 等. 基于 Petri 网和故障树的电力系统故 障诊断模型 [J]. 计算机测量与控制,2015,23(8):3-3.
- [8] 李凤莲, 李园园, 金 铖, 等. 改进的 QPSO 算法在电网故障诊断

- 系统中的应用「J7. 计算机测量与控制, 2014, 22 (12): 3860 -3863.
- [9] 韩德强,杨 艺,韩崇朝. DS证据理论研究进展及相关问题探讨 [J]. 控制与决策, 2014, 29 (1): 1-11.
- [10] Scott Ferson, Vladik Kreinovich. Constructing Probability Boxes and Dempster-Shafer Structures [R]. Sand Report, 2003.
- [11] Fulvio Tonon. Using random set theory to propagate epistemic uncertainty through a mechanical system [J]. Reliability Engineering and System Safety, 2004, 85: 169-181.