一种基于模糊聚类和粗糙集的产品族设计知识约简方法 *

华尔天, 肖军军, 刘科红, 周 科
(浙江工商大学 计算机与信息工程学院, 杭州 310018)

摘 要: 针对产品族设计中产品族共性与产品个性化之间相互冲突的问题,提出了一种基于模糊聚类和粗糙集的产品族设计知识约简方法。通过模糊聚类对设计参数进行离散化和分类,以粗糙集进行知识约简区分出产品族设计参数和个性化产品设计参数,实现产品族综合性能损失最小。通过对通用电机产品族的实例分析,并经与 PPCEM 方法比较,验证了该方法的高效性和可行性。

关键词: 知识约简; 模糊聚类; 粗糙集; 设计参数; 产品族
中图分类号: TP391 **文献标志码**: A **文章编号**: 1001-3695(2011)11-4064-03
doi:10.3969/j.issn.1001-3695.2011.11.016

Method of knowledge reduction for product family design based on fuzzy clustering and rough set

HUA Er-tian, XIAO Jun-jun, LIU Ke-hong, ZHOU Ke
(*School of Computer Science & Information Engineering, Zhejiang Gongshang University, Hangzhou* 310018, *China*)

Abstract: Aiming at the conflict problem between product family commonality and product individuation in product family design, this paper presented a method of knowledge reduction for product family design based on fuzzy clustering and rough sets. Scattered and classified the design parameters through fuzzy clustering, used the rough sets to reduce knowledge, and then distinguished product family design parameters and individuation product design parameters, achieved minimal loss of overall performance of product family design. The feasibility and effectiveness of this method is verified by an example of universal electric motors product family design and the comparison against the PPCEM method.

Key words: knowledge reduction; fuzzy clustering; rough sets; design parameters; product family

产品族作为一个多元化,以客户需求为中心的设计战略得到了工业界和学术界的广泛重视。产品族设计的目的是使一组产品的共性达到最大化而又不牺牲个性化产品的性能,通过产品族共性的增加简化生产并降低生产成本[1]。由于产品族共性与产品个性化之间经常相互冲突,为了研究一种既能兼顾产品族共性,又能实现产品个性化,最大限度地满足用户需求的设计方法,许多学者对产品族设计进行了深入研究。文献[2]提出了产品平台概念探索方法,在已知产品族设计参数分类的情况下,用相容决策支持问题分别求出产品族参数值和个性化产品参数值。文献[3]提出了面向可适应性的参数化产品平台构建方法,通过相似度计算和聚类分析确定了产品族的多组平台。文献[4]提出了基于非支配排序遗传算法,使用该方法能够求得多目标优化问题的 Pareto 集,通过优化设计的两个阶段分别确认产品族参数、个性化产品参数及它们的值。文献[5]构建了稳健优化模型,通过参数灵敏度分析和模糊均值聚类确定产品族设计参数的多值共享,再以性能稳健指数作为约束条件,构建产品族稳健设计方案,实现产品族设计目标与设计稳健性的综合优化。

本文提出一种基于模糊聚类与粗糙集理论的产品族设计知识约简方法,通过模糊聚类进行产品族参数分类,采用阈值优化策略克服属性值划分时的人为主观性;通过以产品性能为决策变量的粗糙集知识约简降低计算复杂度,确定产品族设计参数和个性化产品设计参数,建立了具有可行性的产品族设计方案。

1 产品族设计及方法描述

1.1 产品族设计

产品族是以产品平台为基础,对满足顾客多样化、个性化需求的变异和不可变异模块分类而衍生出的一系列产品[6]。其中变异模块是产品族中实现多样化细分产品、满足顾客需求的组合模块,包括:

a)产品族设计参数。它是指不同产品在一定范围内取值相同的设计参数,用以体现产品族中影响产品性能的共性需求。设有 S 个产品族设计参数指标,组成产品族设计参数矢量 $x_c = \{x_{c1}, x_{c2}, \cdots, x_{cs}\}$。

收稿日期: 2011-04-12; **修回日期**: 2011-05-19 **基金项目**: 国家自然科学基金资助项目(61070059);国家"973"计划资助项目(2009CB326204);浙江省自然基金资助项目(Y1091096);浙江工商大学科研创新基金资助项目(1130XJ1510123)

作者简介:华尔天(1963-),男,浙江慈溪人,研究员,博士,主要研究方向为产品设计及理论、机电一体化技术(huaertian@mail.zjgsu.edu.cn);肖军军(1986-),男,硕士研究生,主要研究方向为数据挖掘、产品设计;刘科红(1986-),女,硕士研究生,主要研究方向为数据挖掘、产品设计;周科(1986-),男,硕士研究生,主要研究方向为数据挖掘、产品设计.

b）个性化产品参数。它是指不同产品在一定范围内取值不相同的设计参数，这些参数值可以由设计人员优化，也可根据客户需求进行定制，实现个性化的产品。设有 t 个个性化产品参数指标，组成个性化产品参数矢量 $x_v=\{x_{v1},x_{v2},\cdots,x_{vt}\}$。

1.2　粗糙集

粗糙集（rough set）理论[7,8]是由波兰数学家 Pawlak 于 1982 年提出来的，它是一种刻画不完全性和不确定性的数学工具。粗糙集理论与证据理论、概率理论等方法相比，优势在于：除所需处理的数据集合，粗糙集无须其他任何的先验信息，给产品族设计参数的知识约简带来了便利。

1.3　模糊聚类

模糊聚类分析[9]是根据客观事物间的不同特性、亲疏程度和相似性等关系，通过建立模糊相似关系对客观事物进行分类的数学方法。模糊聚类分析的方法主要有直接聚类法、传递闭包法、最大树法等[10]。

2　基于模糊聚类和粗糙集的知识约简方法

2.1　产品族设计知识约简模型

产品族设计知识约简模型如图 1 所示。一个产品族可以用一组能代表产品主要技术特征的设计参数及阈值来表示：$x=\{x_1,x_2,\cdots,x_i,\cdots,x_m\}$。其中，$x_i$ 表示第 i 个设计参数，m 是所有设计参数的总数。对产品族设计参数进行知识约简，得到 S 个产品族设计参数 x_c 和 t 个性化产品参数 x_v，其中 $s+t=m$。

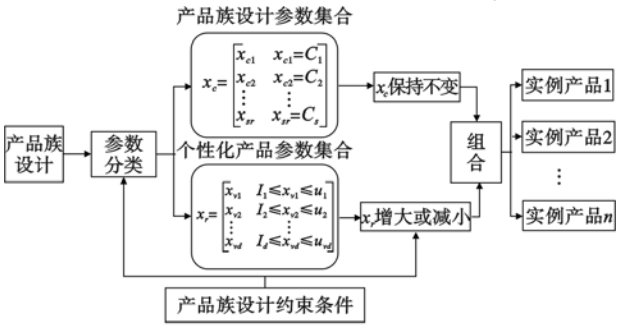

图1　产品族设计知识约简模型

产品族设计就是要对这些设计参数进行分类并确认这些参数的值，满足一定的约束条件，找到产品族设计参数和个性化产品参数之间可接受的平衡点。设计时需考虑的系统约束主要是参数范围要求 $\{l,u\}$ 以及客户对产品性能的需求 $\{g(x),h(x)\}$，这里 u 和 l 分别表示参数的上限和下限，$g(x)$ 表示不等式约束，$h(x)$ 表示等式约束。

综上所述，产品族设计知识约简模型可描述为（模型Ⅰ）

$$\text{find } x_c,x_v$$
$$\min \sum \Delta f(x_c,x_v)$$
$$\text{s.t. } g(x)\leqslant a,h(x)=b,l\leqslant x\leqslant u$$

其中，目标函数表示产品族设计综合性能损失之和最小。

2.2　知识约简方法描述

基于模糊聚类和粗糙集的产品族设计参数知识约简的基本思想是，采用模糊聚类对产品族中的各个设计参数进行分析，通过阈值的选择进行分类和离散，使用粗糙集约简算法选取对产品性能影响度大的参数，确定产品族设计参数和个性化产品参数。具体方法描述如下：

a）产品族单目标优化。根据产品族目标性能与设计参数之间的函数关系建立产品族单目标优化模型（模型Ⅱ）。

$$\text{find } x_j$$
$$\max f_\eta(x_j)$$
$$\text{s.t. } g(x_j)\leqslant a_j,h(x_j)=b_j,l_j\leqslant x_j\leqslant u_j$$
$$j=1,2,\cdots,m;\eta=1,\cdots,n$$

式中，η 表示性能目标的个数，求解后可得每个产品族单目标优化的最优解。

b）模糊聚类对设计参数分类。通过模糊聚类算法对设计参数值离散化处理，形成决策表，具体步骤如下：

（a）对每个参数进行极差变换，消除矩阵中量纲的影响，得到归一化矩阵 P。设

$$P=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{1}$$

式中，每个设计参数有 n 个参数值，即 $x_j=(x_{1j},x_{2j},\cdots,x_{nj})^{\mathrm{T}}$。

（b）在相似系数法、距离法、主观评分法等标定方法中，选择一种来确定相似系数，对每个设计参数分别建立模糊相似矩阵，即 x_{kj} 与 x_{lj} 的相似程度 $r_{ij}=R_j(x_{kj},x_{lj})$。

（c）选择传递闭包法将各设计参数的模糊相似矩阵 R_j 分别进行模糊等价矩阵聚类，从 R_j 出发，依次求平方：

$$R_j\to R_j^2\to R_j^4\to\cdots\to R_j^{2k}=R_j^* \tag{2}$$

得到各个参数的动态聚类图以及各个阈值 λ。

（d）根据动态聚类图，用 F 统计量法计算 λ 的变化率：

$$F=\frac{\sum_{i=1}^{r}n_i\,\|x^{(i)}-\bar{x}\|^2/(r-1)}{\sum_{i=1}^{r}\sum_{j=1}^{n_i}\|x_j^{(i)}-x^{(i)}\|^2/(n-r)} \tag{3}$$

其中，r 为对应 λ 值的类数，n_i 为第 i 类的样本数，$\bar{x}$ 为总体样本的中心，即 $\bar{x}=\frac{1}{n}\sum_{j=1}^{n}x_{jl}$，$x^{(i)}=\frac{1}{n_i}\sum_{k=1}^{n_i}x_{kl}^{(i)}$（$k=1,2,\cdots,n_i;l=1,2,\cdots,m$）。如果 $F>F_\alpha(r-1,n-r)$，说明类与类之间差异是显著的，即分类比较合理，得到 λ 的最佳值，确定最佳分类，形成决策表。

c）通过知识约简区分各设计参数。采用基于遗传算法的属性约简算法对决策表进行属性约简，得到关于产品性能重要度的产品设计参数知识约简集和知识核，将并集后的产品设计参数确定为个性化产品参数 x_v。

约简时若只得到单个设计参数的知识核，将该参数从决策表提出，再次进行属性约简，直到得到知识的约简集，将结果与之前提出的核并集，确定为个性化产品参数 x_v，剩余的设计参数为产品族设计参数 x_c。

d）选择设计参数值，建立产品族设计方案。通过上述分析，已经确认产品族的产品族设计参数 x_c 和个性化产品参数 x_v。产品族设计参数的值可根据其分类进行均值计算，确定产品族设计参数的值后，可根据模型Ⅱ针对个性化产品参数进行单目标优化，得到最终产品族设计方案。

对比产品族设计方案的综合性能损失率，若在允许范围，内则产品族设计参数知识约简完成，若超出范围则转到步骤 b）再次确认阈值。

3　实例分析

通用电机是一种典型的参数化产品族，本文参照文献[11]描述的通用电机模型，对其单目标优化设计方案后的设计参数进行基于模糊聚类和粗糙集的产品族知识约简。在这一实例中，一共有 10 台电机组成一个产品族，设计参数包括 $\{Nc, Ns, A_{wa}, A_{wf}, R, t, I, L\}$，分别是电枢上的线圈绕数 Nc；磁极的线圈绕数 Ns；电枢线圈的横截面积 A_{wa}；磁极线圈的横截面积 A_{wf}；定子的外径 R；定子的厚度 t；电流 I 和堆栈长度 L。具体参数值如表 1 所示。

表 1　电机初步设计参数

序号	设计参数								性能	
	Nc	Ns	A_{wa}	A_{wf}	R	t	I	L	η	T
1	1 019	61	0.161	0.236	18.03	5.99	3.41	11.73	74.5	0.050
2	930	58	0.190	0.190	18.64	4.66	3.83	18.96	70.0	0.100
3	992	62	0.212	0.203	19.83	5.25	3.83	19.56	70.0	0.125
4	1 004	66	0.229	0.223	21.01	5.62	3.83	20.95	70.0	0.150
5	1 016	71	0.259	0.255	22.72	6.07	3.83	23.94	70.0	0.200
6	1 107	76	0.288	0.288	24.38	6.97	3.85	24.43	69.6	0.250
7	1 160	76	0.295	0.295	25.39	7.38	4.00	25.15	67.0	0.300
8	1 199	75	0.303	0.301	26.01	7.50	4.14	26.14	64.6	0.350
9	1 236	75	0.310	0.309	26.93	8.01	4.25	26.39	63.0	0.400
10	1 295	75	0.324	0.324	28.27	8.57	4.49	27.47	59.6	0.500

3.1　各参数模糊聚类

针对表 1 的原始数据进行极差变换，消除量纲化影响，获得标准化的数据矩阵。本文采用直接欧几里德距离法：

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \tag{4}$$

计算后得到的模糊相似矩阵 R，通过传递闭包法得到模糊等价矩阵 R^*。将 λ 从大到小进行聚类，得到动态聚类图。选取第一个设计参数 Nc 为例（图 2），得出该参数的所有分类。

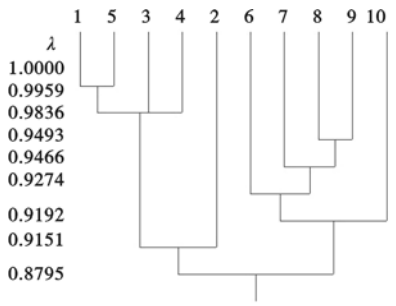

图2　Nc动态聚类图

3.2　最佳阈值 λ 的确认

根据动态聚类图，用 F 统计量法计算 λ 的变化率，得到 λ 的最佳值，确定每个设计参数的最佳分类。以第一个设计参数 Nc 为例，最优分类如表 2 所示。其最佳 $\lambda = 0.9192$，c 表示分类数。

表 2　最佳值及分类数

设计变量	λ	c	分　类
Nc	0.919 2	3	{1,5,3,4}、{2}、{6,7,8,9,10}

3.3　离散化获得决策表

通过动态聚类分类，知道各个设计参数分别被分为 3、2、4、2、4、2、4、3 类；而效率参数 η 被分为三类，可分别表示为“低，中，高”。通过离散化设计参数，可得到决策表，如表 3 所示。以 Nc 为例，其被分为三类，可用“少，中，多”表示。

表 3　设计参数决策表

序号	Nc	Ns	A_{wa}	A_{wf}	R	t	I	L
1	0	0	0	0	0	0	0	0
2	1	0	1	0	0	0	1	1
3	0	0	1	0	0	0	1	1
4	0	0	1	0	0	0	1	1
5	0	1	2	0	1	0	1	2
6	2	1	3	1	2	1	1	2
7	2	1	3	1	2	1	2	2
8	2	1	3	1	2	1	2	2
9	2	1	3	1	2	1	2	2
10	2	1	3	1	3	1	3	2

3.4　基于遗传算法的粗糙集属性约简

本文在传统的粗糙集方法上采用遗传算法对粗糙集进行属性约简。约简前的决策表如表 3 所示，决策变量 $\eta = \{0,1,1,1,1,1,2,2,2,3\}$，通过遗传算法计算，得到约简后的结果只有一个参数 $\{I\}$，其支持度为 100，确定 I 为知识的核。单个核不足以完全作为产品族设计参数的约简集，在此基础上，将 I 首先提取出决策表，对新的决策表再次进行属性约简，可得到两个约简集 $\{R, L\}$ 和 $\{A_{wa}, R\}$，确认该电机产品族的知识约简集为 $\{I\} \vee \{R, L\} \vee \{A_{wa}, R\}$，可求得电机产品族的个性化产品参数为 $\{I, R, A_{wa}, L\}$。

3.5　设计参数优化

通过动态聚类及粗糙集分析，确认通用电机产品族的个性化产品参数为 $\{I, R, A_{wa}, L\}$，产品族设计参数为 $\{Nc, Ns, A_{wf}, t\}$。表 4 是单目标优化后的产品族设计方案。

表 4　产品族设计方案

序号	设计参数								性能	
	Nc	Ns	A_{wa}	A_{wf}	R	t	I	L	η	T
1	1 007	62	0.215	0.221	15.18	5.52	3.03	11.32	84.5	0.050
2	930	62	0.190	0.221	18.36	5.52	3.68	17.96	73.3	0.100
3	1 007	62	0.206	0.221	20.69	5.52	3.75	18.86	69.5	0.125
4	1 007	62	0.219	0.221	21.48	5.52	3.84	20.69	68.3	0.150
5	1 007	75	0.232	0.221	22.74	5.52	3.92	22.58	66.0	0.200
6	1 200	75	0.240	0.303	23.58	7.69	4.02	24.43	62.6	0.250
7	1 200	75	0.259	0.303	24.76	7.69	4.16	25.35	61.8	0.300
8	1 200	75	0.274	0.303	25.83	7.69	4.14	26.14	62.0	0.350
9	1 200	75	0.293	0.303	26.85	7.69	4.38	26.49	60.7	0.400
10	1 200	75	0.305	0.303	27.78	7.69	4.66	26.79	58.4	0.500

3.6　优化结果对比

以性能中的效率为参数，同单目标优化得到的原始数据[11]及产品平台概念探索法（product platform concept exploration method，PPCEM）[2]进行对比，结果如表 5 所示。从表 5 的数据中可以看出，基于粗糙集和聚类的产品族设计知识约简得到的方案比单目标优化方案在综合平均性能上降低了 1.82%，属于可接受范围；同 PPCEM 相比，平均效率提高了 2.13%，其方案只有 R 和 t 两个平台参数，本文（下转第 4080 页）

了。

2.2　计算长度的选择

前面分析了 MATLAB 时频工具箱的时频分布与计算 FFT 的个数有关,即当计算 FFT 时的长度与实际信号的长度不同,就会出现在时间轴发生位移,甚至是错误的,如图 3 所示。而采用本文改进的算法,则不存在这样的问题。

图 3 和 4 仍然是以图 1 的信号为输入信号,采用不同的方法计算的时频分布图。从图 3 中可知,当 FFT 计算长度与实际信号长度不一样时,计算的时频分布图在时间轴上发生了位移,在频率轴上具有散射现象,使频率的分辨率模糊了。从图 4 中可知,采用改进后的算法,则时频分布的分辨率不受计算长度的影响,甚至比利用原始信号长度 1 801 个点计算的时频分辨率还高。因此改进后的算法更适合计算实际的地震信号。

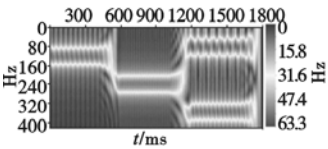

图3　基2FFT算法1 024点时频域图像

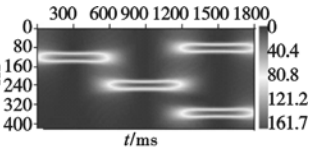

图4　改进后的算法1 024点时频域图像

3　结束语

本文结合实际地震勘探资料处理的需要,改进了 MATLAB 时频工具箱中时频分布函数的离散短时傅里叶变换的算法。通过实验表明,改进后的算法的计算效率与时窗长度有关,时窗越小,计算效率越高;同时,改写的 Fortran 程序受计算机的内存影响较小,因此,从理论上来讲,可以计算任意大容量的地震数据,而不像 MATLAB 中由于数据太大而出现内存不足,导致计算终止。

参考文献:

[1] PANERAS D E, MANI R, NAWAB S H. STFT computation using pruned FFT algorithms[J]. IEEE Signal Processing Letters, 1994, 1(4): 61-63.

[2] 黄德济,赵宪生. 时频域中主参数剖面的计算及应用[J]. 物探与化探计算技术, 1994, 16(3): 197-204.

[3] STOCKWELL R G, MANSINHA L, LOWE R P. Localization of the complex spectrum: the S transform[J]. IEEE Trans on Signal Processing, 1996, 44(4): 998-1001.

[4] 徐天吉,程冰洁. 瞬时谱分析技术与气水识别方法研究[J]. 石油天然气学报, 2008, 30(4): 73-78.

[5] 刘喜武,刘婉莹,刘洪,等. 地震信号广义时频分析及其数值实现[J]. 物探与化探计算技术, 2007, 29(5): 386-390.

[6] MARFURT K J, KIRLIN R L. Narrow-band spectral analysis and thin-bed tuning[J]. Geophysics, 2001, 66(4): 1274-1283.

[7] 田仁飞,曹俊兴,方磊. 时频分析技术及其在识别含气层中的应用[J]. 天然气地球科学, 2008, 19(2): 72-75.

[8] 向富强. 地震相干分析和时频分析方法及其在储层描述中的应用[D]. 成都: 成都理工大学, 2008.

[9] 何坤,周激流,李健,等. STFT 与图像压缩的探讨[J]. 计算机应用研究, 2005, 22(3): 158-163.

[10] 葛哲学,陈仲生. MATLAB 时频分析技术及其应用[M]. 北京: 人民邮电出版社, 2007.

[11] 王宏禹,陈喆. 非平稳随机信号分析与处理[M]. 2 版. 北京: 国防工业出版社, 2008.

[12] 丁玉美. 数字信号处理[M]. 西安: 西安电子科技大学出版社, 2005.

(上接第 4066 页)有四个,通用性从 25% 增加到了 50%。以上分析说明,本文的方法是可行的,能够有效且快速地实现产品族设计的知识约简。

表 5　电机产品族效率对比

序号	单目标优化	PPCEM	本文优化	差异 1	差异 2
	η_1/%	η_2/%	η/%	η 同 η_1/%	η 同 η_2/%
1	74.5	74.7	84.5	13.42	13.12
2	70.0	72.1	73.3	4.71	1.66
3	70.0	71.1	69.5	-0.71	-2.25
4	70.0	70.1	68.3	-2.43	-2.57
5	70.0	67.5	66.0	-5.71	-2.22
6	69.6	64.6	62.6	-10.06	-3.10
7	67.0	62.2	61.8	-7.76	-0.64
8	64.6	59.9	62.0	-4.02	3.51
9	63.0	57.7	60.7	-3.65	5.20
10	59.6	53.8	58.4	-2.01	8.55
平均变化率				-1.82	+2.13

4　结束语

产品族设计方法中,设计参数和个性化参数的确定及它们之间的平衡点是其中的关键问题。本文提出基于模糊聚类和粗糙集的产品族设计知识约简方法,通过模糊聚类对参数进行离散化处理及分类,用粗糙集确定各设计参数的重要程度,找到产品族共性和产品个性化之间可接受的平衡点,避免了优化模型所产生的复杂算法,实现产品族的通用性最大和产品综合性能损失最小。通过通用电机实例进行验证,证明了该方法的有效性和可行性。

参考文献:

[1] 李柏姝,雒兴刚,唐加福. 基于可调节参数的产品族设计多阶段方法[J]. 计算机集成制造系统, 2010, 16(6): 1138-1143.

[2] SIMPSON T W, MAIER J R A, MISTREE F. Product platform design: method and application[J]. Research in Engineering Design, 2001, 13(1): 2-22.

[3] 陈永亮,褚巍丽,徐燕申. 面向可适应性的参数化产品平台设计[J]. 计算机集成制造系统, 2007, 13(5): 877-884.

[4] 李忠凯,谭建荣,冯毅雄,等. 基于多目标遗传算法的可调节变量产品族优化[J]. 浙江大学学报: 工学版, 2008, 42(6): 1015-1020.

[5] 丁力平,冯毅雄,谭建荣,等. 基于性能稳健指数的产品族稳健优化设计[J]. 计算机集成制造系统, 2010, 16(6): 1121-1130.

[6] 朱斌,江平宇. 面向产品族的设计方法学[J]. 机械工程学报, 2006, 42(3): 1-7.

[7] 王哲,王仕成,张金生,等. 一种地磁匹配制导适配性特征参数选取方法[J]. 宇航学报, 2009, 30(3): 1057-1063.

[8] 苗夺谦,李道国. 粗糙集理论、算法与应用[M]. 北京: 清华大学出版社, 2008.

[9] 吴静,吴晓燕,高忠长. 基于模糊聚类和粗糙集的仿真可信性模糊综合评估[J]. 系统工程与电子技术, 2010, 32(4): 770-773.

[10] 谢季坚,刘承平. 模糊数学方法及应用[M]. 武汉: 华中科技大学出版社, 2006.

[11] DAI Zhi-huang, SCOTT M J. Product platform design through sensitivity analysis and cluster analysis[J]. Journal of Intelligent Manufacturing, 2007, 18(1): 97-113.