

基于粗糙集全局离散和粒子群算法的启发式航材订货模型

周家萱，徐常凯

(空军勤务学院航材四站系，江苏徐州，221000)

摘要 为了解决目前航材订货模型单纯依托人为经验,对航材消耗规律变化反应滞后且可信度不高工作量过大的问题,提出在原始订货模型的框架下,使用 PAM 聚类、粗糙集全局离散和粒子群算法,计算模型中的航材分类、属性离散和区间权重,从而建立启发式航材订货模型,并在测试集上对比新老模型的均方离差。结果表明,启发式航材订货模型可以使人摆脱依靠经验确定模型的繁琐工作,并且能有效提高模型的准确性和及时性,从而提高航材订货工作效率。

关键词 订货;聚类;粗糙集;离散;粒子群

DOI 10.3969/j.issn.1009-3516.2019.04.005

中图分类号 V37;E939 **文献标志码** A **文章编号** 1009-3516(2019)04-0033-06

The Air Materials Heuristic Ordering Model Based on Rough Set Global Discretization and PSO

ZHOU Jiaxuan, XU Changkai

(Department of Air Materials and Four Stations,
Air Force Logistic College, Xuzhou 221000, Jiangsu, China)

Abstract: Aimed at the problems that at the present time the model of air materials ordering is relying solely on human experience, the stagnancy responded to variation of consume rules exists in and the credibility is low, and the amount of work is too heavy, a heuristic ordering model is established to calculate the classification, the properties' discretion and the interval weight by using PAM cluster, Rough Set global discretion and PSO under the framework of ordinary ording model. Then, the MSE between the two models is compared. The result indicates that the heuristic model could help people get rid of the tedious work of calculate model by experience, and improve the veracity and response timeliness of model.

Key words: ordering; clustering; rough set; discretization; PSO

航材订货工作是各航空兵部队航材保障的重要环节,合理、准确的航材订货数量可以有效满足日常训练备战的航材供应需求。目前航材订货主要依靠

根据航材决策工作人员长期累积的工作经验,对航材订货中某些关键属性值进行离散并赋权,通过经验公式计算订货数量。但是,航材消耗是一个快速

收稿日期: 2018-07-05

作者简介: 周家萱(1995—),男,江西吉安人,硕士生,主要从事航材保障决策与信息化研究。E-mail:zjxwss@163.com

引用格式: 周家萱,徐常凯. 基于粗糙集全局离散和粒子群算法的启发式航材订货模型[J]. 空军工程大学学报(自然科学版), 2019, 20(4): 33-38. ZHOU Jiaxuan, XU Changkai. The Air Materials Heuristic Ordering Model Based on Rough Set Global Discretization and PSO[J]. Journal of Air Force Engineering University (Natural Science Edition), 2019, 20(4): 33-38.

动态变化的过程,特别是当前航空兵部队飞行任务变化可能导致消耗规律突变,长期使用一个依靠人为经验的静态模型难以及时、准确地反映航材消耗规律,并且依靠人为经验确定模型费时费力,严重分散航材订货决策人员的工作精力。为了解决这些问题,本文提出一种基于航材历史消耗、维修和可靠性数据,根据航材保障的实施情况进行动态调整,不断优化重构航材运算模型的组成结构和参量运用方式的启发式航材订货模型。

1 航材订货原始模型

航材订货流程如下:①首先需要对所有航材做预测基数计算,根据 3 年平均消耗区分正常消耗航材和低消耗航材,使用不同预测算法计算基数;②然后对所需预测的航材进行分类,分为消耗件、可修件和有寿件。进行消耗件预测时,库存折算只需考虑储存期因素,预测影响因子中需要计算订货到货时间影响因素;进行可修件预测时,库存折算需要考虑储存期和修理次数因素,预测影响因子中需要计算订货到货时间、平均修理周期和航材修复率,如果该可修件是直属库类器材,则还应计算直属库库存占比因素;进行有寿件预测时,大体流程与可修件相同,区别在于在库存折算中还需考虑装机日历时间折算因素;③最后将所有参数带入启发式航材订货模型公式^1:

$$M = (X_{\text{消耗}1} + S_3 | X_{\text{平均}3} + S_3 | X_{\text{平均}3}\beta_1) \cdot \prod_{k=2}^5 \beta_k - (\sum_{i=1}^n (\prod_{j=1}^3 \alpha_j^i)) \quad (1)$$

式中: $X_{\text{消耗}1}$ 表示某型航材近 1 年消耗; $X_{\text{平均}3}$ 表示近 3 年消耗平均; S_3 表示近 3 年消耗的标准差; $\beta_k (k=1, 2, \dots, 5)$ 分别表示低消耗影响因子,订货到货时间影响因子,修理周期影响因子,修复率影响因子和直属库影响因子; $\alpha_j^i (j=1, 2, 3)$ 分别表示库存第 i 件航材的储存期影响因子、修理次数影响因子和寿控影响因子。在订货预测中,影响因子的取值采用将该属性值离散后,对每一区间依照订货经验赋予不同权重的方式获得,例如,订货到货时间影响因子的取值可由表 1 确定。

表 1 订货到货时间影响因子

条件	影响因子取值
300 < 订货到货时间 ≤ 365	1.5
200 < 订货到货时间 ≤ 300	1.3
150 < 订货到货时间 ≤ 200	1.0
100 < 订货到货时间 ≤ 150	0.9
70 < 订货到货时间 ≤ 100	0.8
订货到货时间 ≤ 70	0.7

可以看出,航材订货原始模型存在以下 3 点问

题:①模型的离散和赋值依靠人为经验,其科学性和准确性难以得到验证;②经验是长期累积的过程,难以对不断变化的航材消耗规律做出快速反应;③依靠经验确定具体参数需要人工反复验证,由于数据量过大,此项工作将会耗费大量时间。因此,需要针对这些问题对原始模型做出改进。

2 启发式航材订货模型建模流程

启发式航材订货建模流程如下:①首先根据航材混合属性数据,使用 PAM 算法进行聚类;②然后使用粗糙集理论中的数据离散方法,对需要离散操作的属性依据第 1 步的聚类结果进行有监督离散;③最后根据需求重新构建启发式订货模型,使用粒子群优化算法求解不同区间的最优权重。最终建立启发式航材订货模型。

2.1 混合属性下的 PAM 聚类算法

2.1.1 混合属性不相似度计算

航材订货特征属性中包含分类变量和连续型变量,而欧式距离、曼哈顿距离等常用的距离度量方法无法适用在混合变量的场合,因此在计算器材之间的不相似度时,本文选择使用 Gower 距离。

计算 Gower 距离首先需要将分类变量的属性值按照 $0, 1, \dots, n$ 的序列离散,然后按式(2)标准化^[2]:

$$W_{ij} = \begin{cases} \frac{p_{ij} - \bar{p}_{..j}}{S_{..j}}, & \text{连续变量} \\ \sqrt{2} \times \frac{p_{ij} - \bar{p}_{..j}}{S_{..j}}, & \text{分类变量} \end{cases} \quad (2)$$

然后代入式(3)得到每种器材在样本中的不相似度:

$$d(i) = \sum_{j=1}^m W_{i,j} \quad (3)$$

2.1.2 PAM 聚类算法

PAM 算法是 K-means 算法的改进版本,相比与传统的 K-means 算法最小化二次欧式距离,PAM 算法的目标函数是最小化不相似度,拥有更好的鲁棒性。PAM 算法首先从 m 个样本中随机选择 k 个对象作为聚类的初始中心点,剩余样本对象按与各中心的距离就近分配成一簇,然后反复用非中心点代替中心点以提高聚类的质量(聚类过程中聚类质量的评估依据为代价函数),直至每一聚类簇中的对象不再变化^[3]。

设 U_i, U_p 为中心点, U_j, U_k 为非中心点,定义函数 $R(A, B)$ 为重心点替换函数, C_{jk} 为聚类过程中执行替换操作的代价函数。由文献[4]可知,执行 $R(U_i, U_k)$ 操作的总代价为:

$$\text{Total}(C) = \sum_{j=1}^{m-k} C_{jk} \quad (5)$$

从式(5)可以看出,聚类簇数 K 是评价聚类的重要参数。在聚类算法的实际应用中,经常选择使用轮廓系数^[5]评价聚类簇数选取的质量,具体为^[6]:

$$S_i = \frac{1}{K} \sum_{i=1}^k \frac{b_i - a_i}{\max(a_i, b_i)} \quad (6)$$

式中: a_i 为在簇 i 中,某一对象与其他对象的平均距离; b_i 为该对象在所有对象中的最小距离^[5]。轮廓系数越大,聚类的质量越好。因此在聚类时,应当先计算轮廓系数求出最优 k 值,再进行聚类操作。

2.2 基于断点分辨力的启发式全局离散算法

2.2.1 断点分辨力定义

定义为集合 X 中由断点 (a, c) 得到的可分辨对象数, $l_j^X(a, c)$ 为属性值为 j 的实例中,属于集合 X 且属性 a 的值大于等于断点 c_m^a 的实例个数,类似的,定义 $r_j^X(a, c)$ 为属性值为 j 的实例中,属于集合 X 且属性 a 的值小于等于断点 c_m^a 的实例个数,对于 $\forall (a, c) \in C_a, X \in U, W^X(a, c)$ 的断点分辨力为^[11]:

$$W^X(a, c) = l^X(a, c)r^X(a, c) - \sum_{i=1}^r l_i^X(a, c)r_i^X(a, c) \quad (7)$$

2.2.2 启发式全局离散算法

全局离散是将决策表中所有连续值属性进行离散。假设 X_1, X_2, \dots, X_m 为表 $S|_D$ 不可分辨关系 $\text{IND}(C|_D)$ 下的等价类,定义 $\text{PART}(D) = \{X_1, X_2, \dots, X_m\}$ 为论域 U 在断点集合 D 下分为 m 个不相交子集,此时,在断点集合 D 下,所有对象中只由断点 (a, c) 分辨的数量为^[12]:

$$W_D(a, c) = W^{X_1}(a, c) + W^{X_2}(a, c) + \dots + W^{X_m}(a, c) \quad (9)$$

全局离散算法可以形式化描述如下^[13]:

Input: 决策表 A
Output: 决策表 A 上所有属性一个分类 P
Data Structure: D -决策表 A 上最优断点集合;
 $L = \text{PART}(D)$ -论域 U 上由断点集合 D 得到的所有等价类; C_A -决策表 A 上所有可能的断点集合。
Method:
1) $D = \emptyset; L = \{U\}; A_1 =$ 属性集合 A 上的初始断点集合。

2) For $c \in C_A$ do 计算 $W_D(c)$ 。

3) 选择 A_1 中使得 $W_D(c)$ 最大的断点值。

$$D = D \cup \{c_{\max}\}$$

$$c_{\max} \text{ and set } C_A = C_A \setminus \{c_{\max}\}$$

4) For $A \in L$ do。

if X 包含某一决策类中的对象 then 将 X 从 L 中移除;

if c_{\max} 将集合 X 分为 X_1, X_2 两部分 then 将 X 从 L 中移除 AND 将集合 X_1, X_2 加入到 L 。

5) if L 为空 then 退出 else 返回 2)。

可以看出,全局离散算法是将 $W_D(a, c)$,即断点分辨力作为算法的启发式方程,分辨力强的断点优先进入结果断点集 D ,遍历论域 U 上所有对象后算法结束。

2.3 粒子群算法

粒子群算法由 Kennedy 和 Eberhart 在 1995 年提出,属于群智能算法范畴。相比于其他群智能算法,它的优势在于简单,容易实现,无需梯度信息,参数少,特别是其天然的实数编码特点特别适合于处理实优化问题^[14]。PSO 的基本公式为^[15-16]:

$$X_{k+1, d} = X_{d, i}^k + V_{d, i}^{k+1} \quad (10)$$

$$V_{d, i}^{k+1} = V_{d, i}^k + c_1 r_1 (p_{d, i}^k - x_{d, i}^k) + c_2 r_2 (p_{g, i}^k - x_{d, i}^k)$$

式中: $V_{d, i}^k$ 为第 k 次迭代粒子 i 速度矢量第 d 维分量; $X_{d, i}^k$ 为粒子 i 位置矢量第 d 维分量; $p_{d, i}^k$ 为粒子 i 经历过的最佳位置; c_1, c_2 为学习因子或加速系数,一般为 2; r_1, r_2 取值范围 $[0, 1]$,可以自行确定。

3 启发式航材订货模型构建

3.1 数据准备

启发式航材订货模型的数据来源是航材管理信息系统的日常业务流水数据。本文从流水数据中选择了 1 000 项需要统一制定订货计划的典型器材,每项器材选取 21 个属性,其中包括 7 个分类变量和 14 个连续型数值变量,用于表征不同航材的多方面订货特性,见表 2。

表 2 启发式订货模型属性

属性名	属性类型	属性名	属性类型	属性名	属性类型
直属库占比	数值	订货到货时间	数值	是否有寿	分类
3 年消耗平均	数值	专业名称	分类	予收数	数值
3 年消耗标准差	数值	册号	分类	近 1 年消耗	数值
平均修理时间	数值	订货 AB 类	分类	近 2 年消耗	数值
修复入库次数	数值	是否成套	分类	近 3 年消耗	数值
未修退回次数	数值	是否进口	分类	一级品库存	数值
修复率	数值	是否可修	分类	二级品库存	数值

在统计数据中,存在若干空缺值,有以下 3 种情况:

1) 直属库占比空缺:即该器材不属于直属库储存的器材,空缺值用 0 填充。

2)平均修理时间、修复入库次数空缺:即该器材不属于可修件,空缺值用 0 填充。

3)订货到货时间空缺:该器材的统计数据不完整,空缺值按样本中无空缺的到货时间平均值填充。

3.2 订货航材 PAM 聚类

3.2.1 订货航材不相似度计算

按式(3)计算订货航材 Gower 距离 Gower_dist,得到 Gower_dist 的五值见图 2。

```
> summary(gower_dist)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.04348 0.16990 0.23980 0.23100 0.28400 0.69570
```

图 2 Gower_dist 五值

3.2.2 最优聚类簇数选择

在 PAM 聚类方法下,本文分别计算聚类簇数从 2 到 200 时的聚类轮廓系数,得到轮廓系数随聚类簇数的变化和极值点附近轮廓系数见图 3 和表 3。

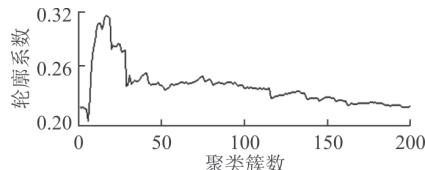


图 3 轮廓系数变化图

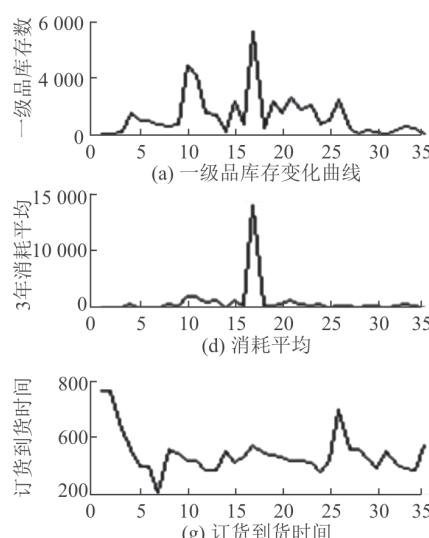


图 4 第 17 类器材各属性统计图

3.3 订货航材相关属性离散

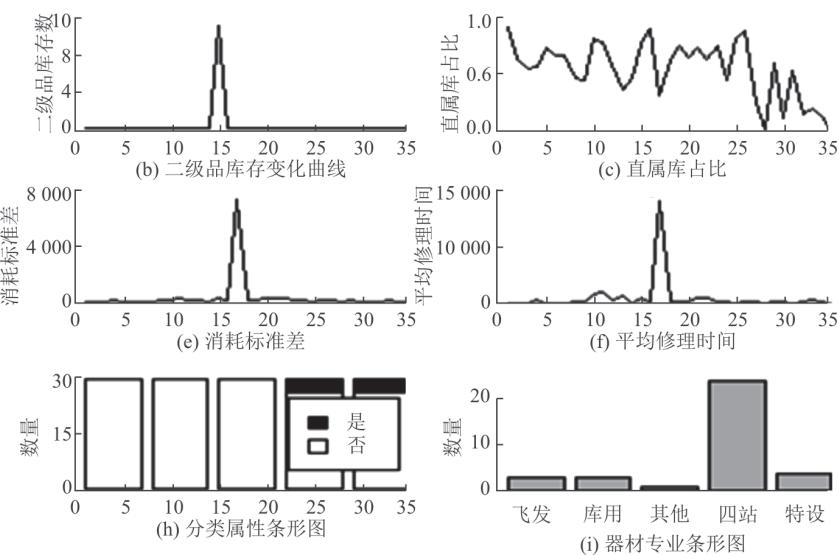
订货航材混合属性中,需要进行离散操作的属性有直属库占比、平均修理时间、修复率和订货到货时间 4 个属性,依据 3.2 节的聚类结果,建立粗糙集决策表。将决策表输入离散算法程序可得断点和各断点分辨能力见图 5。属性直属库占比上的断点集 $C_{a_1} = \{0.245, 0.405, 0.605\}$, 属性平均修理时间上的断点集 $C_{a_2} = \{1.5, 10.5, 41.5\}$, 属性平均修复率上的断点集 $C_{a_3} = \{35\%\}$, 属性订货到货时间上的

表 3 极值附近轮廓系数值

聚类簇数	轮廓系数	聚类簇数	轮廓系数
10	0.281 597 0	16	0.305 241 3
11	0.296 300 1	17	0.307 289 4
12	0.299 591 3	18	0.306 384 8
13	0.299 637 8	19	0.304 232 4
14	0.293 413 9	20	0.273 216 1
15	0.299 900 9	21	0.277 301 8

可知当聚类簇数为 17 时,轮廓系数最大,为 0.307 289 4,因此在进行聚类操作时,k 值选取为 17。由于此次聚类的对象数量多、维度高、聚类簇数大,难以对聚类结果赋予实际含义,也难以将聚类结果直观呈现,因此我们选择其中一类,统计该类中各属性数据进行分析。第 17 类航材各属性统计数据见图 4。

第 17 类中共有 35 项器材,其中大部分为四站装(设)备,除离群点外,此类器材的消耗量适中,且较为稳定,库存一级品和直属库占比很高,同时,此类器材的订货到货时间平均超过 300 天且大部分不可修。可以预见,此类器材在订货需要设立更高的安全库存量。



断点集 $C_{a_1} = \{330, 417.5, 452.5\}$ 。用数字序列将各条件属性离散,最终得到的粗糙集离散决策表见表 4。

3.4 模型参数计算

3.4.1 粒子群适应度函数

模型的适应度函数表示模型计算值与实际订货值的距离,比照原始订货模型计算公式,用粒子群算法求解启发式航材订货模型参数的适应度函数如下:

表4 离散决策表

对象 器材	条件属性					决策属性			
	直属库占比	类别	平均修理时间	类别	修复率/%	类别	订货到货时间	类别	分类
应答主机	0.08	0	63	3	99.43	1	245	0	7
收发信机	0.28	1	18	2	100	1	423	2	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
专用计算机	1	3	0	0	0	0	448	2	10
数据采集器	0.83	3	7	2	100	1	468	3	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

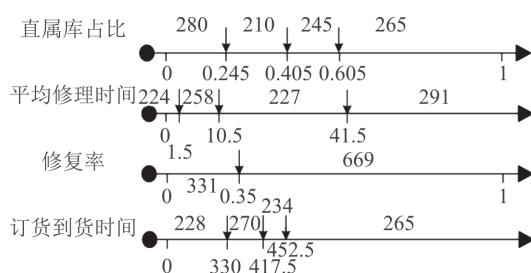


图5 断点示意图

$$\text{Fitness} = [D_r - (C \times \prod_{k=2}^5 \beta_k - \sum_{i=1}^n (\prod_{j=6}^8 \beta_j))]$$

式中: D_r 表示实际订货量(未在属性中列出); C 表示消耗基数,由原模型式(1)确定; $\beta_{2,3,4,5}$ 与式(1)的表示相同; β_6 表示一级品库存折算参数; β_7 表示二级品库存折算参数; β_8 表示修理次数折算参数。

3.4.2 参数计算

本文在求解某一变量单个区间的参数值时,采用先控制变量多次实验,再综合结果取平均值的方法计算参数。下面以计算订货到货时间参数 β_2 在第一区间[0, 330]上的参数值为例说明。

求解目标区间上的参数值时,需要组合其余3个属性的区间值,其中 β_3 有4个区间, β_4 有2个区间, β_5 有4个区间,即共需做32组实验。从原数据表中随机抽取800条数据构成训练集,将对应实验数据从训练集中筛选出,32组实验中样本容量不为0的16组实验最优解中 β_2 的值如表5。

表5 参数取值表

属性	区间			
	0	1	2	3
β_2	0.317 79	0.824 57	1.352 58	1.527 54
β_3	1.009 33	0.873 37	0.925 38	1.597 84
β_4	4.219 27	1.527 27		
β_5	1.826 30	1.458 58	1.229 84	1.011 24
β_6		0.840 26		
β_7		0.386 82		
β_8		0.308 44		

可得 β_2 在第一区间[0, 330]上的参数值为

0.317 79。类似的,可以求出所有参数在对应区间上的取值。

4 模型比较

将原始表中剩余的200条数据抽出构成测试集,分别将测试集输入新老模型计算预测订货数,与实际订货数比较,计算各条数据均方误差(MSE),两模型均方误差比较见图6。

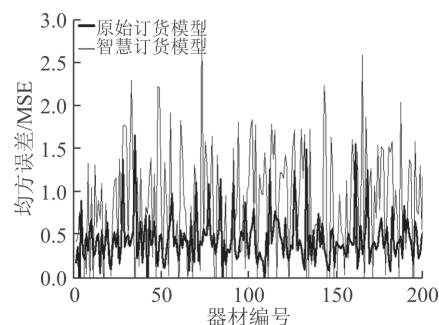


图6 均方误差对比图

经计算,对测试集2002项器材预测结果的均方误差(MSE),原始订货模型 MSE、启发式订货模型 MSE 的标准差和均值分别为 0.786 4、0.274 5 和 0.799 3、0.429 5。启发式订货模型 MSE 的标准差和均值分别下降 65.09% 和 46.24%。

可以明显看出,启发式订货模型均方误差明显降低,且其方差也显著降低,说明其预测精度更高且更加稳定。

相比于依靠经验的原始订货模型,启发式订货模型的优势在于可以随时根据具体的航材消耗情况自动调整订货航材分类,连续属性上的离散点和各离散区间上的权值,既能保证订货模型及时适应不断变化的消耗规律,也能有效减少利用经验在确定具体参数上的人为误差,同时,整个模型建立过程中,只有数据准备需要人工处理,大大减少订货决策人员的工作量。因此,启发式订货模型能够有效解决原始订货模型滞后性、不准确性和人工工作量大的问题,更加适应目前航材消耗规律快速变化的现状。

5 结语

航材订货是一项涉及多个学科和专业的复杂工作。本文提出使用启发式智能算法代替人为经验,依托航材历史累积数据,科学计算订货模型中航材分类、属性离散和离散区间权值。理论证明,启发式订货模型可以实时反映对航材消耗规律,减少由人为经验对订货模型带来的误差,并且使订货决策人员从繁琐的模型确定工作中解脱出来,更适应由飞行任务变化带来的消耗规律突变。同时,在实验中,其准确性和稳定性也较经验模型有更好的表现。综上,启发式航材订货模型可以有效提升订货预测的准确性和及时性,进而提高航材订货的工作效率。

参考文献(References) :

- [1] 何亚群,柏林,赵福军,等.航材供应学[M].徐州:空军勤务学院,2012.
- HE Y Q, BO L, ZHAO F J, et al. Supplement of Air Material [M]. Xuzhou: The Air Force Logistics College, 2012. (in Chinese)
- [2] GOWER J C. A General Coefficient of Similarity and Some of Its Properties[J]. Biometrics, 1971, 27(4): 857-871
- [3] 王燕飞.基于 PAM 聚类分析方法的路径优化算法研究与实现[J].山西科技,2013,28(1):85-88.
- WANG Y F. Study on and Implement of Path Optimization Algorithm Based on PAM Cluster Analysis[J]. Shanxi Science and Technology, 2013, 28(1):85-88. (in Chinese)
- [4] 杨延璞,顾蓉,古玉峰.基于 PAM 聚类的工业设计方案评价共识度达成方法[J].计算机集成制造系统,2017,23(7):1386-1393.
- YANG Y P, GU R, GU Y F. Method for Reaching a Consensus in Industrial Design Evaluation Based on PAM Clustering[J]. Computer Integrated Manufacturing Systems, 2017, 23(7):1386-1393. (in Chinese)
- [5] KAUFMAN L, ROUSSEEUW P J. Finding Groups in Data: An Introduction to Cluster Analysis[M]. New York: John Wiley & Sons, 1990.
- [6] AMORIM R C D, HENNIG C. Recovering the Number of Clusters in Data Sets with Noise Feature Rescaling Factors[J]. Information Sciences, 2015, 324 (c): 126-145.
- [7] PAWLAK Z. Information Systems-Theoretical Foundations[J]. Information Sciences, 1981, 6:205-218.
- [8] 张靖,段富.优化初始聚类中心的改进 k-means 算法[J].计算机工程与设计,2013,34(5):1691-1694.
- ZHANG J, DUAN Y. Improved k -Means Algorithm with Meliorated Initial Centers [J]. Computer Engineering and Design, 2013, 34 (5): 1691-1694. (in Chinese)
- [9] LALA S R, ANDRZAJ J, CHRISTOPH B, et al, Jose M B. Implementing Algorithm of Rough Set Theory and Fuzzy Rough Set Theory in the R Package‘Rough-Sets’[J]. Information Science, 2014, 287:68-89.
- [10] JAN G B, HUNG S N, SINH H N, et al. Rough Set Algorithm in Classification Problem [J]. Rough Set Methods and Applications, 2001, 1:49-88.
- [11] 刘静.基于粗糙集理论的连续值属性离散化方法研究[D].重庆:重庆邮电大学,2009.
- LIU J. Discretization of Continuous Attribute Values in Rough Set Theory[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2009. (in Chinese)
- [12] NGUYEN H S. On Efficient Handling of Continuous Attributes in Large Data Bases [J]. Fund Inform, 2001, 48:61-81.
- [13] NGUYEN H S. Discretization Methods in Data Mining[J]. Rough Sets in Knowledge Discovery, 1998, 1: 451-482.
- [14] 沈林成,霍霄华,牛铁峰.离散粒子群优化算法研究现状综述[J].系统工程与电子技术,2008, 30 (10): 1986-1994.
- SHEN L C, HUO X H, NIU Z F. Survey of Discrete Particle Swarm Optimization Algorithm[J]. Systems Engineering and Electronics, 2008, 30(10):1986-1994. (in Chinese)
- [15] CLERC M, KENNEDY J. The Particle Swarm-Exploration, Stability, and Convergence in a Multidimensional Complex Space[J]. IEEE Transaction on Evolutionary Computation, 2002, 6(1):58-73.
- [16] SHI Y, EBERHART R C. A Modified Particle Swarm Optimizer[C]// IEEE International Conference on Evolutionary Computation. Piscataway, NJ: IEEE Press, 1998:69-73.

(编辑:徐敏)