

doi: 10.3969/j.issn.1006-1576.2012.11.007

主成分聚类分析在航材分类中的应用

张作刚¹, 胡新涛¹, 刘望²

(1. 海军航空工程学院青岛分院, 山东 青岛 266041; 2. 海军 92367 部队, 山东 青岛 266041)

摘要: 针对航材种类多、规格复杂而导致航材管理复杂的问题, 提出一种基于主成分聚类分析的航材分类方法。在综合考虑传统航材分类指标和航材自身特点的基础上, 构建了一种新的航材分类指标体系。根据指标集运用主成分分析法进行降维处理, 运用 SPSS 软件的系统聚类法对其聚类分析, 得出航材综合分类。仿真实例结果表明: 该方法减少了数据的冗余, 得到的分类结果客观、合理。

关键词: 主成分分析; 聚类分析; 航材分类

中图分类号: TJ04 **文献标志码:** A

Application of Principal Component Clustering Analysis in Aviation Spares Classification

Zhang Zuogang¹, Hu Xintao¹, Liu Wang²

(1. Qingdao Branch, Naval Aeronautical & Astronautical University, Qingdao 266041, China;
2. No. 92367 Unit of Navy, Qingdao 266041, China)

Abstract: Aiming at the complex management of large amount of aviation spares with complex specifications, introduced a spares classification method based on principal component clustering analysis. Based on traditional index of classification and characteristic of spare, establish a new aviation spares classification index system. According to index set, use principal component clustering analysis method to reduce dimensions. Adopt SPSS software system clustering method to carry out clustering analysis and get aviation spares integrated clustering. The simulation results show that the method can reduce data redundancy and the clustering results is objective and reasonable.

Key words: principle components analysis; clustering analysis; spare classification

0 引言

目前,“多机种,少批量”的趋势,使航材的品种、数量急剧增加,出现了“三多一少”的特点,即:“机种多、特种设备多、专用航材种类多,而数量少”,因而使库存结构更趋复杂,航材备件储存、管理的难度也相应增大^[1]。据统计,一个场站航材股储备的航材一般都在数万项,且这些航材性质不一,用途各异^[2]。为方便航材的管理,提高工作效率,必须把它们分类。目前较常用的分类方法有 ABC 分类法、FAHP 与模糊综合评判相结合的方法等^[3]。

传统的分类方法要么过于简单,要么过于复

杂,或者引入过多定性的主观判断,要么需要大量的历史数据^[4]。笔者根据航材的保障特性和航材自身特点,构建航材分类指标,运用主成分聚类分析对其进行分类,并通过实例验证该指标体系和分类方法的可行性。

1 分类指标体系构建

笔者综合考虑航材的特点,选取了以下 8 个反映航材特点的指标变量: 1) 器材单价 x_1 ; 2) 采购提前期 x_2 ; 3) 平均故障间隔时间 x_3 ; 4) 年均消耗量 x_4 ; 5) 危害度 x_5 ; 6) 历史消耗量 x_6 ; 7) 供应商等级 x_7 ; 8) 航材保质期 x_8 。如表 1 所示。

表 1 航材分类指标体系

属性	属性值	说明
单价 x_1	元	航材的采购单价,以最近一次的采购单价计算
采购提前期 x_2	d	航材的采购提前期,以最近一次的采购发单到航材验收入库的时间来计算
平均故障间隔时间 x_3	10^3 h	可修复航材在相邻两次故障之间的工作时间是其工作寿命,该寿命的数学期望即为平均故障间隔时间
年均消耗量 x_4	个	航材近五年领用量的平均值
危害度 x_5	0.1~1	由航材部门按照航材对飞行安全影响的程度和缺件时造成损失的程度进行综合评定
保质期 x_6	d	航材的保质期
历史消耗量 x_7	个	航材历年的总领用数量
供应商等级 x_8	0.1~1	航材部门按航材供应商供货质量、服务水平等进行综合评定

收稿日期: 2012-05-28; 修回日期: 2012-06-25

作者简介: 张作刚(1961—),男,山东人,硕士研究生导师,从事航空装备、器材勤务技术研究。

2 主成分聚类分析模型

2.1 主成分分析模型

主成分分析是利用降维的思想，将多个变量转化为少数几个综合变量(即主成分)，其中每个主成分都是原始变量的线性组合，各主成分之间互不相关，从而这些主成分能够反映原始变量的绝大部分信息，且所含的信息互不重叠^[5]。计算步骤如下：

1) 计算相关系数矩阵。

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (1)$$

$r_{ij} (i, j = 1, 2, \dots, p)$ 为原变量 x_i 与 x_j 的相关系数。 R 为实对称矩阵($r_{ij} = r_{ji}$)。

2) 计算特征值与特征向量。

解特征方程 $|\lambda I - R| = 0$ ，通常用雅比克法求出特征值 $\lambda_i (i = 1, 2, \dots, p)$ ，并使其按大小顺序排列，即 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ；求出对应于特征值 λ_i 的特征向量 $e_i (i = 1, 2, \dots, p)$ ，并要求 $\|e_i\| = 1$ ，即 $\sum_{j=1}^p e_{ij}^2 = 1$ ，

其中 e_{ij} 表示向量 e_i 第 j 个分量。

3) 计算主成分贡献率及累计贡献率。

$$\text{贡献率: } \lambda_i / \sum_{k=1}^p \lambda_k \quad (2)$$

$$\text{累计贡献率: } \sum_{k=1}^i \lambda_k / \sum_{k=1}^p \lambda_k \quad (3)$$

根据选取主成分个数的原则，特征值要求大于 1 且累计贡献率达 80% ~ 95% 的特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 所对应的 $1, 2, \dots, m (m \leq p)$ ，其中整数 m 即为主成分的个数。

4) 计算主成分载荷。

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} \quad (i, j = 1, 2, L, P) \quad (4)$$

5) 各主成分得分。

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix} \quad (5)$$

2.2 聚类分析模型

在不知道类型的个数或对于各种类型的结构未作任何假设的情况下，聚类分析能够将一批样本变量自动进行分类。并且每一类代表具有相似性的集合，不同类之间具有明显的非相似性。分类构成中，不必事先给出分类的标准，聚类分析能够从样本变量出发，客观地决定分类标准^[6]。计算步骤如下：

- 1) 构造 n 个类，每个类只包含一个样本点。
- 2) 计算 n 个样品两两之间的距离 $\{d_{ij}\}$ ，记作

$$D^{(0)} = (d_{ij}^{(0)})_{n \times n}。$$

3) 合并距离最近的两类为一新类，称做第 $n+1$ 类，并取消刚合并的那两类，这样得到 $n-1$ 类。

4) 计算新类与剩余各类的距离，其他各类间距离不变，得到降一阶的新距离矩阵 $D^{(1)} = (d_{ij}^{(1)})_{(n-1) \times (n-1)}$ 。若样本类的个数等于 1，则转到步骤 5)；否则，回到步骤 3)。

- 5) 画聚类图。
- 6) 决定类的个数和类。

2.3 主成分聚类分析的步骤

- 1) 分类指标的选取及数据收集；
- 2) 将选取的指标进行归一化处理；
- 3) 对归一化后的指标体系进行主成分分析，剔除存在相关性、信息重叠的指标，以减少数据的冗余；
- 4) 将主成分分析后的指标，进行聚类分析。

3 实例仿真分析

为验证航材分类指标的有效性以及航材分类方法的可行性，笔者收集了某航材仓库部分航材的指标信息，如表 2 所示。

表 2 航材指标信息

器材	属性							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
A	59	12	0.485	2	0.3	26	0.1	20 000
B	144	13	0.445	6	0.2	124	0.6	20 000
C	845	15	0.593	4	0.5	25	0.5	7 300
D	211	9	1.154	1	0.1	7	0.4	7 300
E	93	15	0.661	3	0.6	34	0.5	2 520
F	87	15	0.351	2	0.5	21	0.4	2 520
G	69	11	0.159	4	0.6	43	0.6	2 520
H	1	13	0.691	4	0.6	45	0.6	20 000
I	199	7	0.659	4	0.4	56	0.4	20 000
J	1 366	37	1.121	1	0.9	12	0.9	20 000
K	46 654	92	0.675	3	0.7	21	0.8	20 000
L	31 591	91	0.751	4	0.9	34	0.6	20 000
M	55	33	0.697	3	0.5	36	0.4	3 650
N	815	39	0.785	9	0.5	61	0.9	20 000
O	87	24	0.984	1	0.6	16	0.3	20 000
P	416	32	0.895	4	0.5	20	0.9	20 000

3.1 主成分分析

笔者借助于 SPSS19 统计软件完成分类指标数据的处理及分析, 将 16 个航材的指标数据进行主成分分析, 分析指标间的关系, 剔除一些没有明显分异的指标或相互间存在明显线性相关关系的指标, 以确定新的指标。

为了减少分类指标的数量又不损失太多的信息, 这里取累计贡献率为 80% 以上^[7], 提出了 4 个主成分。运行结果如表 3、表 4 所示。

表 3 航材分类指标的解释总方差

成分	初始特征值			提取平方和载入		
	合计	方差的百分比/%	累积百分比/%	合计	方差的百分比/%	累积百分比/%
1	2.935	36.691	36.691	2.935	36.691	36.691
2	2.099	26.241	62.933	2.099	26.241	62.933
3	1.206	15.079	78.012	1.206	15.079	78.012
4	0.843	10.544	88.556	0.843	10.544	88.556
5	0.471	5.882	94.438			
6	0.209	2.611	97.050			
7	0.202	2.521	99.571			
8	0.034	0.429	100.000			

表 4 航材分类指标的主成分系数

分类指标	主成分系数			
	1	2	3	4
x_1	0.472 7	-0.001 7	-0.342 3	0.401 3
x_2	0.542 5	0.013 4	-0.206 2	0.187 3
x_3	0.214 7	-0.297 2	0.695 4	-0.059 9
x_4	0.045 2	0.634 3	0.022 7	-0.185 3
x_5	0.434 9	-0.115 1	-0.228 7	-0.375 2
x_6	-0.136 1	0.616 0	0.010 9	0.182 1
x_7	0.375 3	0.268 1	0.195 0	-0.597 4
x_8	0.292 8	0.211 5	0.515 9	0.485 6

从主成分系数表中可以看成, x_1 、 x_2 、 x_5 在第一主成分上载荷较大; x_4 、 x_6 在第二主成分上载荷较大; x_3 、 x_8 在第三主成分上载荷较大; x_7 在第四主成分上载荷较大。故将 8 个指标降为 4 个, 即重要度主成分、消耗量主成分、可靠性主成分和供应商等级主成分。

主成分表达式为:

$$F_1 = 0.472 7ZX_1 + 0.542 5ZX_2 + \dots + 0.292 8ZX_8 \quad (6)$$

$$F_2 = -0.001 7ZX_1 + 0.013 4ZX_2 + \dots + 0.211 5ZX_8 \quad (7)$$

$$F_3 = 0.401 3ZX_1 + 0.187 3ZX_2 + \dots + 0.485 6ZX_8 \quad (8)$$

式中的 $\{ZX_i\}$ 为 $\{x_i\}$ 的标准化数据。

笔者采用加权综合的方式进行评价, 如式 (9) 所示, 权重为各信息量的比重。主成分得分如表 5。

$$F = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} F_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} F_2 + \dots + \frac{\lambda_4}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} F_4 \quad (9)$$

即 $F = 0.414 4F_1 + 0.296 3F_2 + \dots + 0.119F_4$ 。

表 5 主成分的得分

器材	F_1	F_2	F_3	F_4	F
A	-1.650 0	-0.246 7	-0.075 4	1.767 8	-0.559 3
B	-1.438 6	-0.484 6	0.421 7	0.951 7	-0.554 7
C	-0.835 4	0.300 0	-0.510 8	-0.553 3	-0.410 1
D	-1.477 6	-0.772 4	1.326 0	0.371 6	-0.571 2
E	-0.845 7	-0.467 9	-0.737 1	-0.907 3	-0.722 6
F	-1.421 2	-0.207 3	-1.544 9	-0.399 1	-0.961 0
G	-1.194 1	0.322 7	-1.924 3	-1.112 0	-0.859 2
H	-0.091 8	0.146 7	0.591 4	-0.141 8	0.089 3
I	-1.017 6	-0.188 6	0.597 5	0.765 0	-0.284 8
J	1.984 1	-0.343 5	1.386 9	-1.267 7	0.805 8
K	3.757 2	0.595 5	-1.200 3	1.045 7	1.653 4
L	3.311 7	0.196 8	-0.978 2	0.733 1	1.351 4
M	-0.784 4	-0.575 9	-0.685 0	-0.275 9	-0.645 2
N	0.860 9	1.663 4	1.038 3	-0.903 5	0.918 9
O	-0.039 9	-0.820 4	0.974 9	0.727 2	-0.007 0
P	0.882 4	0.882 0	1.319 5	-0.801 4	0.756 3

3.2 新指标聚类分析

对所筛选数据进行运算, 运用 SPSS 软件, 基于新航材分类指标, 采用系统聚类法对其进行分类, 聚类结果如图 1、表 6 所示。

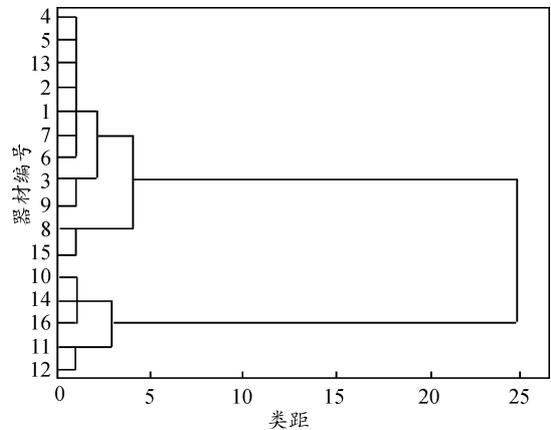


图 1 航材分类的聚类

表 6 航材分类

器材名称	类别
J、K、L、N、P	A 类
A、B、C、D、E、F、G、I、M	B 类
H、O	C 类

对于航材 P, 按照传统方法划为 C 类航材, 而分类模型将其划为 A 类航材。由表 2 观察可知航材 P 的价格虽低, 但是采购提前期、危害度和消耗量都较高, 所以归为 A 类比较合适。

同理, 按照传统方法航材 B、D、I 划为 C 类,

而分类模型将其划为 B 类航材。根据分析该方法是客观、合理的。

4 结论

仿真结果表明：利用主成分分析的结果作为聚类分析的样本数据，减少了数据的冗余，计算简单、原理清晰，所得的结论客观、合理。同时可以消除航材传统分类方法的弊端，能够大大提高航材管理的效率，实现航材的科学分类，为航材的筹措、供应、统计和管理提供依据。

参考文献：

[1] 张作刚. 海军航材库存管理[M]. 青岛：海潮出版社，

(上接第 24 页)

通过扩展有限元程序模拟计算，可以得到裂纹扩展的整个动态图。以侵彻时间 $t=118 \mu s$ 为例，可得到侵彻图 5。从图 5 可以观察到裂纹已经发生了严重的变形。原先的直线裂纹已经扩展成不规则形状。随着杆弹的继续侵彻，裂纹会发生进一步的形变和开裂。

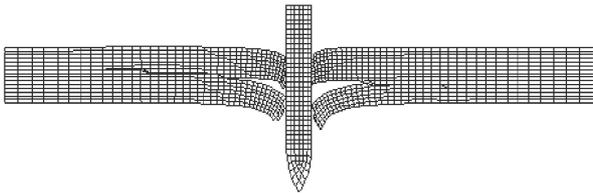


图 5 卵形杆弹侵彻靶板过程中网格分布

为了分析不同时刻杆弹和靶板的变形程度，仍以 $t=118 \mu s$ 为例，得到此刻的等效塑性应变图 6。由图 6 可知，靶板的变形集中于自身中线两侧，并随离中线距离的增加变形逐渐减小；由于钢质杆弹强度较大，变形集中于弹头部分，弹体基本不发生形变。

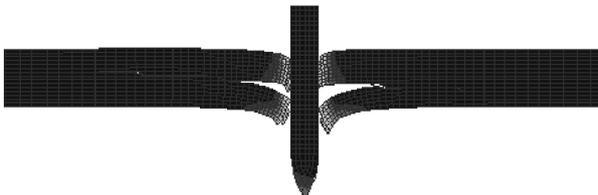


图 6 等效塑性应变

4 结论

通过模拟计算可以获得侵彻过程中裂纹的整个

2008: 2-3.

[2] 虞文胜. 聚类分析在航材分类上的应用[J]. 价值工程, 2011, 20(30): 309.

[3] 姜晨, 徐廷学, 余仁波. 基于 DAGSVM 的装备不常用备件分类方法[J]. 舰船科学技术, 2011, 33(7): 135-136.

[4] 韩增奇, 于俊杰, 王朝阳. 航空兵异地驻训航材需求预测模型[J]. 四川兵工学报, 2010, 31(2): 35.

[5] 朱星宇, 陈勇强. SPSS 多元统计分析方法及应用[M]. 北京: 清华大学出版社, 2011: 241.

[6] 鲍燕, 胡振琪, 柏玉, 等. 主成分聚类分析在土地利用生态评价中的应用[J]. 农业工程学报, 2006, 22(8): 87-88.

[7] 张永利, 傅俊伟. 基于主成分分析方法的聚类分析方法在灾情综合分类中的应用[J]. 佳木斯大学学报, 2011, 29(2): 298.

扩展情况，说明扩展有限元方法和水平集理论相结合能精确地重现裂纹的扩展。程序计算结果可以显示侵彻过程中任一时刻的裂纹扩展图，这些扩展图精确和直观地给出此刻裂纹扩展的形状和裂纹处在靶板中的具体位置。据此即可分析出相比其他方法更为精确的裂纹扩展规律，进而可以对杆弹侵彻靶板机理做更精确和细致的研究。

参考文献：

[1] Daux C, Moes N, Dolbow J, et al. Arbitrary branched and intersecting cracks with the extended finite element method[J]. International Journal for Numerical Methods in Engineering, 2000, 48: 1741-1760.

[2] Melenk J M, Babuska I. The Partition of Unity Finite Element Method: Basic Theory and Applications[C]. Seminar fur Angewandte Mathematic, Eidgenossische Technische Hochschule, Research Report No.96-01, January, CH-8092 Zurich, Switzerland.

[3] Ventura G, Xu J X, Belytschko T. A vector level set method and new discontinuity approximations for crack growth by EFG[J]. International Journal for Numerical Method in Engineering, 2002, 54: 923-944.

[4] Stolarska. Modeling Crack Growth by level sets[D]. Northwestern University, 2002(12).

[5] 朱传锐. 基于扩展有限元和水平集理论的裂纹扩展问题研究[D]. 河南理工大学硕士学位论文, 2010.

[6] Piekutowski A J, Forrestal M J, Poormon K L, et al. Perforation of aluminum plates with ogive-nose steel rods at normal and oblique impacts[J]. Int J Impact Eng, 1996, 18(7-8): 877-887.

[7] 王峰. 有限元方法及其在高速碰撞中的应用[D]. 合肥: 中国科学技术大学博士学位论文, 2007.