

## FCM在飞机动态使用评价中的应用

李克武, 郭建胜, 田 舫

(空军工程大学工程学院, 陕西西安 710038)

摘要: 提出了一种在数据预处理后利用模糊C-均值法进行飞机动态使用评价的聚类分析方法, 论述了该方法的聚类过程, 给出了实例并对其结果进行了分析, 结果表明该方法科学合理, 切合实际。

关键词: 航空维修信息; 模糊聚类分析; 模糊C-均值

中图分类号: V37 文献标识码: A 文章编号: 1009-3516(2005)03-0016-03

航空维修信息既多又杂, 其中某架飞机的使用状况是机务人员所关心的, 平时认为某架飞机好或差, 只是感觉上的, 没有对其进行量化分析。而且, 现在使用的飞机完好率指标是根据飞机完好架日与飞机停飞架日的关系计算得出, 也只是反映某单位所有飞机总体完好情况。

模糊聚类分析方法可以根据对象的属性进行聚类, 而在基于目标函数的聚类方法中, C-均值法<sup>[1]</sup>最为常用, 可用于飞机的使用性能评价。它包括硬C-均值法(HCM)和模糊C-均值法(FCM)。前者将每个数据划分到一个, 也只能到一个类别中; 后者则可表示出每个数据属于各个类别的程度, 精确地反映出实际的数据分布情况, 这对于增加飞机使用信息分析的敏感度是一个重要的特征。

## 1 模糊聚类分析——FCM 算法

FCM 算法将给定的  $p$  维空间  $R^p$  中的  $n$  个向量  $X = \{x_i | x_i \in R^p, i = 1, 2, \dots, n, p$  是样本的属性个数} 分为  $c$  ( $1 < c < n$ ) 类, 其分类矩阵  $U = \{\mu_{ij} | i = 1, 2, \dots, n, j = 1, 2, \dots, c\}$ , 元素  $\mu_{ij}$  表示第  $j$  个数据点属于第  $i$  类的隶属度, 并求每个类的聚类中心  $c_i$  ( $i = 1, 2, \dots, c$ ) 使得非相似性(或距离)指标的目标函数达到最小, 满足如下条件:

随机初始化分类矩阵  $U$ 。  $\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, n$  目标函数式  $\min J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}$ 。式中  $u_{ij} \in [0, 1]$ ;  $c_i$  是聚类中心  $i$ ;  $d_{ij}$  是第  $j$  个数据点与第  $i$  个聚类中心的距离, 一般采用欧式距离:  $d_{ij} = \|X_j - C_i\|^2$ , 欧式距离适合于类内数据点为超球形分布的情况, 采用不同的距离定义, 可用于不同分布类型数据的聚类分析;  $m \in (1, \infty)$  是权重指标。

为了达到目标函数最小, 要满足式(1)和式(2)两个条件。

$$c_i = \left( \sum_{j=1}^n u_{ij}^m x_j \right) / \sum_{j=1}^n u_{ij}^m \quad (1) \quad u_{ij} = 1 / \sum_{k=1}^c (d_{ij} / d_{kj})^{2/(m-1)} \quad (2)$$

算法可用迭代过程描述为<sup>[2]</sup>: ①初始化  $\mu_{ij}$ ; ②用式(1)计算新的聚类中心; ③根据欧几里德距离公式<sup>[3]</sup>计算每一个向量  $x_i$  到所属类的聚类中心  $c_j$  的距离; ④用式(2)计算每一个向量的隶属度; ⑤根据  $|\mu_{ij}^{(l+1)} - \mu_{ij}^{(l)}| \leq \varepsilon$  判断隶属度是否收敛, 其中  $l$  表示第  $l$  次迭代,  $\varepsilon$  为收敛水平。如果上式成立, 分类迭代结束, 选取

收稿日期: 2004-12-31

基金项目: 军队科研基金资助项目

作者简介: 李克武(1968-), 男, 湖南长沙人, 讲师, 硕士, 主要从事系统管理与决策研究;

郭建胜(1965-), 男, 河南灵宝人, 教授, 主要从事智能信息处理研究。

合适的聚类水平  $\lambda$  值,输出分类结果;否则转到第②步。

## 2 属性数据的模糊处理

飞机现处状况信息存储在关系数据库中,很多属性属于非数值型字段,不能直接参与计算,但又对数据分析很重要。飞机现处状况分类,不仅跟飞机的飞行时间、起落次数等数值型字段相关,还与飞机的出厂日期、可否作战等日期型、逻辑型字段类型相关,因此,在飞机总体性能分类时,有必要对原始数据作模糊处理。

### 2.1 非数值型数据

对于非数值型数据可以利用某个元素属于某类属性的隶属度来表示<sup>[4-5]</sup>。具体方法是:设  $n$  是元素总数, $m$  是属性的分类数, $C_i$  是第  $i$  个类, $N(C_i)$  是类  $i$  包含的属性值个数, $C_{ij}$  是第  $i$  类中第  $j$  个属性值,则属性值  $\mu_{ij}$  的隶属度为

$$\mu_{ij} = N(C_i)/n \quad (3)$$

对于布尔型数据,分类数  $m = 2$ ;其它的属性值应根据距离分布进行分类处理。分类数  $m$  可以采取划分区间的办法,把相同属性值划分为同一类,如果其离散性很小,把同一区间内的所有属性值划分为一类,经过类划分后,就可以计算各个类对于属性的隶属情况。

### 2.2 数值型数据

对于数值型数据可以利用式(3),也可以采用百分制,将数据化为从域  $X = \{x_1, x_2, \dots, x_n\}$  到  $[0, 1]$  闭区间的映射。

## 3 实例分析

某单位有关飞机出厂年月、飞行时间、飞行场次、故障次数的统计资料如表1所示(以8架飞机、4个指标为例),现将飞机现处状况按照上述4个指标分为完好、一般、欠佳三类。

### 3.1 数据预处理

根据式(3)计算飞机出厂年月数据。这里  $n = 8$ ,区间划分为: $d_1 = [1970.5, 1971.4]$ ,  $d_2 = [1974.1, 1974.8]$ ,  $d_3 = [1980.5, 1980.5]$ 。则  $\mu(A) = \mu(C) = \mu(F) = \mu(H) = 4/8 = 0.5$ ,  $\mu(B) = \mu(E) = 2/8 = 0.25$ ,  $\mu(D) = \mu(G) = 2/8 = 0.25$

利用百分制计算数值型字段,结果见表2。

表1 飞机状态统计表

| 飞机 | 出厂年月   | 飞行时间  | 场次 | 故障次数 |
|----|--------|-------|----|------|
| A  | 1974.3 | 72.50 | 25 | 2    |
| B  | 1980.5 | 80.34 | 25 | 0    |
| C  | 1974.8 | 73.00 | 24 | 1    |
| D  | 1970.5 | 65.22 | 23 | 2    |
| E  | 1980.5 | 79.20 | 24 | 0    |
| F  | 1974.2 | 72.38 | 23 | 1    |
| G  | 1971.4 | 63.11 | 24 | 2    |
| H  | 1974.1 | 74.25 | 24 | 1    |

表2 飞机状态数据处理后表

| 飞机 | A1   | A2     | A3   | A4  |
|----|------|--------|------|-----|
| A  | 0.5  | 0.7250 | 0.25 | 0.2 |
| B  | 0.25 | 0.8034 | 0.25 | 0   |
| C  | 0.5  | 0.7300 | 0.24 | 0.1 |
| D  | 0.25 | 0.6522 | 0.23 | 0.2 |
| E  | 0.25 | 0.7920 | 0.24 | 0   |
| F  | 0.5  | 0.7238 | 0.23 | 0.1 |
| G  | 0.25 | 0.6311 | 0.24 | 0.2 |
| H  | 0.5  | 0.7425 | 0.24 | 0.1 |

### 3.2 模糊聚类

设初始分类共分3类,随机选取飞机A、B、C归于一类,飞机D、E、F归于一类,飞机G、H归于一类,初始隶属度为  $\mu^{(0)}$ 。

$$\mu^{(0)} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\mu^{(7)} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

令  $m = 2$ ,利用模糊C-均值法的聚类过程如表3所示。

表3 分类迭代隶属度表

| 飞机 | 第1次迭代 |       |       | 第2次迭代 |       |       | ... | 第6次迭代 |       |       | 第7次迭代 |       |       |
|----|-------|-------|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|
| A  | 0.073 | 0.875 | 0.052 | 0.032 | 0.948 | 0.020 | ... | 0.048 | 0.911 | 0.041 | 0.048 | 0.912 | 0.040 |
| B  | 0.486 | 0.305 | 0.209 | 0.545 | 0.291 | 0.164 | ... | 0.863 | 0.093 | 0.044 | 0.865 | 0.092 | 0.043 |
| C  | 0.216 | 0.672 | 0.112 | 0.099 | 0.851 | 0.050 | ... | 0.019 | 0.967 | 0.014 | 0.020 | 0.965 | 0.015 |
| D  | 0.184 | 0.263 | 0.553 | 0.096 | 0.143 | 0.761 | ... | 0.053 | 0.099 | 0.848 | 0.055 | 0.102 | 0.843 |
| E  | 0.517 | 0.292 | 0.191 | 0.591 | 0.268 | 0.141 | ... | 0.930 | 0.049 | 0.021 | 0.928 | 0.050 | 0.022 |
| F  | 0.037 | 0.936 | 0.027 | 0.058 | 0.904 | 0.038 | ... | 0.061 | 0.885 | 0.054 | 0.060 | 0.867 | 0.053 |
| G  | 0.220 | 0.292 | 0.488 | 0.163 | 0.225 | 0.612 | ... | 0.062 | 0.104 | 0.834 | 0.060 | 0.102 | 0.838 |
| H  | 0.522 | 0.361 | 0.117 | 0.356 | 0.540 | 0.104 | ... | 0.161 | 0.748 | 0.091 | 0.161 | 0.748 | 0.091 |

从表3的迭代隶属度可以看出,当迭代到第7次时,隶属度已经收敛( $\varepsilon = 0.05$ )。根据表3的第7次迭代结果,给定截集水平  $\lambda = 0.748$ ,得到  $\mu^{(7)}$ : 第一类{B, E}, 第二类{A, C, F, H}, 第三类{D, G}。结果表明飞机 B、E 属于完好,飞机 A、C、F、H 属于一般,飞机 D、G 欠佳。

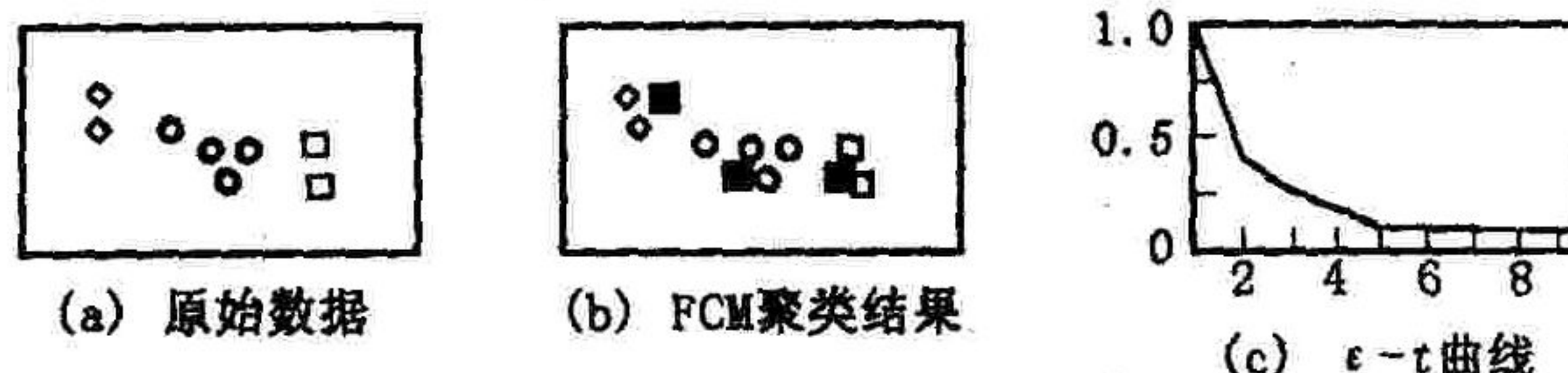


图1 飞机现处状况数据聚类结果

图1(a)表示原始数据,图1(b)中的“■”表示类别中心,是迭代7次的结果,图1(c)是收敛值  $\varepsilon$  与迭代次数  $t$  的关系曲线。根据图1(a)和图1(b)可看出结果是合理的,该单位的维修经验也表明了结论是符合实际的。

## 4 结论

对属性数据的模糊化处理和归一化处理后,应用模糊 C-均值算法能够较好地实现飞机使用性能评价。另外,随着聚类水平  $\lambda$  的取值由 1 减少到 0,分类由细到粗,表明对象以不同的隶属度可以同时属于不同的几个类,对于飞机使用信息分析这是一个有用的特征。可见,模糊聚类分析法可作为航空维修信息系统的重要支持工具。

## 参考文献:

- [1] Westphal C, Blaxton T. Data Mining Solutions: Methods and Tools for Solving Real - World Problems[M]. New York: Wiley Computer Publishing, 1998.
- [2] Jang J S R, Sun C T, Mizutani E. Neuro - Fuzzy and Soft Computing [M]. New York: Prentice Hall, 1997.
- [3] 杨伦标,高英仪. 模糊数学——原理及应用[M]. 广州:华南理工大学出版社,1995.
- [4] 邹喜正,朱明华. 对模糊聚类分析中原始数据预处理研究[J]. 煤矿开采,1994,(2):43-45.
- [5] 张 丽,高虹霓,梁颖亮. 模糊综合评价管理信息系统[J]. 空军工程大学学报(自然科学版),2001,2(5):91-94.

(编辑:姚树峰)

Application of FCM in Performance Evaluation for Aircraft Dynamic Use

L[ Ke - wu, GUO Jian - sheng, TIAN Shah

(The Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710038, China)

Abstract: In this paper, a method applied to performance evaluation for aircraft dynamic use based on fuzzy c - means clustering after data pretreatment is presented. The process of clustering - analysis and the result of clustering are described in detail by an example. This application result shows that this method is scientific rational and corresponds to actual situation.

Key words: aviation maintenance information; fuzzy clustering analysis; fuzzy c - means clustering