

文章编号:1001-9081(2005)02-0348-04

## 模糊 K-Prototypes 算法中的加权指数研究

汪加才<sup>1</sup>, 朱艺华<sup>2</sup>

(1. 南京审计学院 计算机科学与技术系, 江苏 南京 210029;  
2. 浙江工业大学 信息智能与决策优化研究所, 浙江 杭州 310014)  
(jcwang@nau.edu.cn)

**摘 要:**模糊 K-Prototypes(FKP)算法融合了 K-Means 和 K-Modes 对数值型和符号型数据的处理方法, 适合于混合类型数据的聚类分析。同时, 模糊技术使得 FKP 适合于处理含有噪声和缺少数据的数据库。但是, 在使用 FCM(Fuzzy C-Means algorithm)或 FKP 算法时, 如何选取加权指数  $\alpha$  仍是悬而未决的问题。许多研究者基于他们的实验结果给出 FCM 中的最佳加权指数可能位于区间  $[1.5, 2.5]$ , 本文则提出了一个 FKP 中加权指数的探寻算法。在多个实际数据集上的实验结果表明, 为进行有效的聚类, FKP 中加权指数应该小于 1.5。

**关键词:**加权指数; FKP 算法; 聚类有效性

**中图分类号:** TP311.131 **文献标识码:** A

## Research on the weighting exponent in fuzzy K-Prototypes algorithm

WANG Jia-cai<sup>1</sup>, ZHU Yi-hua<sup>2</sup>

(1. Department of Computer Science and Technology, Nanjing Audit University, Nanjing Jiangsu 210029, China;  
2. Institute of Information Intelligence and Decision Optimization, Zhejiang University of Technology, Hangzhou Zhejiang 310014, China)

**Abstract:** Fuzzy K-Prototypes(FKP) algorithm integrating K-Means and K-Modes algorithm is suited for clustering mixed numeric and categorical valued data. The use of fuzzy techniques makes it robust against noise and missing values in the databases. But, it is an open problem how to select an appropriate weighting exponent  $\alpha$  when run FCM(Fuzzy C-Means algorithm) or FKP. Some researchers have suggested that the best choice for  $\alpha$  in FCM be probably in the interval  $[1.5, 2.5]$  based on their experimental results. In this paper, the algorithm for searching suitable  $\alpha$  in FKP was presented. The experimental results on several real datasets show that the valid clustering can be achieved when  $\alpha$  is under 1.5.

**Key words:** weighting exponent; FKP algorithm; clustering validity

## 0 引言

给定包含  $m$  个数据对象的数据库和所要形成的聚类个数  $K$ , 划分算法依据一个确定的划分标准将对象集合划分为  $K$  份, 其中, 每个划分均代表一个聚类。该标准使得一个聚类中的对象是相似的, 而不同聚类中的对象是不相似的。基于划分的聚类算法的典型代表是 K-means<sup>[1]</sup>、模糊 K-means<sup>[2]</sup>、ISODATA<sup>[3]</sup> 等。K-means 系列算法以其实现简单、效率高、适合于大数据集而广受欢迎。

为了克服 K-means 系列算法仅适合于数值型数据的不足, Huang 提出了 K-modes 算法<sup>[4]</sup>和模糊 K-modes 算法<sup>[5]</sup>。K-modes 算法用模(mode)来替换聚类中心, 采用符号匹配的差异性计算方法来处理符号量, 以及利用基于频率方法对各聚类模进行更新。模糊 K-modes 算法结合了 K-modes 和模糊 K-means 的长处, 通过对不同类内的对象赋予置信值(confidence)来确定类的核心对象和边缘对象, 从而为处理类

中的边缘对象提供了更多的有用信息。K-prototypes 算法<sup>[5,6]</sup>则是 K-means 算法和 K-modes 算法的结合, 可以对采用数值量和符号量描述的对象进行聚类分析。

在模糊 K-means 算法(通常也称为 FCM 算法)的目标函数中所含有的加权指数  $\alpha$  对算法的聚类效果有重要影响<sup>[7]</sup>。但在理论上, 如何选取合适的加权指数仍是一个悬而未决的问题。通过对聚类有效性函数的评估实验, 文献[8,9]宣称最佳的  $\alpha$  可能位于  $[1.5, 2.5]$  区间; 而文献[7]则给出了在所构造的 Normal-12 数据集上该区间内的加权指数将导致聚类算法失效的结论, 当  $\alpha$  为 1.2 时, 聚类效果较好。

对于模糊 K-modes 算法及模糊 K-Prototypes 算法(简称 FKP 算法), 合理的加权指数区间在哪里? 文献[5]针对 Soybean 疾病数据集(也称为 Soybean-small), 指出当  $\alpha = 1.1$  时可以得到较小的目标函数值。但文献[6]给出的定理 6 以及实验结果的事实表明, 随着  $\alpha$  的增加, 目标函数值是单调下降的。显然, 目标函数值不能作为选取  $\alpha$  的依据。文献[6]则

收稿日期: 2004-07-29; 修订日期: 2004-10-08

基金项目: 国家自然科学基金资助项目(60473097); 江苏省高校自然科学基金计划项目(03KJB520054)

作者简介: 汪加才(1962-), 男, 江苏人, 副教授, 博士, 主要研究方向: 数据挖掘、商业智能; 朱艺华(1961-), 男, 浙江人, 教授, 博士, 主要研究方向: 移动计算、决策支持系统研究。

利用聚类模型的正确分类率考察了  $\alpha$  的取值范围。对于 Soybean-small, 在  $\alpha$  超过 50 以后, 模型的平均正确分类率急剧下降; 最高正确分类率发生在  $\alpha = 4$  处。

正如文献[7]所指出的那样, 如何评价模糊聚类算法在不同加权指数时的聚类效果属于聚类有效性问题。实际上, 高正确分类率的聚类模型不一定是合适的模型。本文的目的是通过对一批实际数据集的 FKP 聚类, 利用聚类模型的正确分类率、划分系数<sup>[3]</sup>、划分熵<sup>[10]</sup>等指标, 探寻 FKP 算法的合理加权指数区间。

## 1 FKP 算法描述

**定义 1** 数据库表  $T$  是由  $n$  个属性  $A_1, \dots, A_n$  描述的一组待聚类对象集  $X, X = \{x_1, x_2, \dots, x_m\}$ , 对象  $x_i$  表示为  $(x_{i,1}, \dots, x_{i,n})$ 。将  $n$  个属性根据其取值的不同分为数值型和符号型两类, 设前者有  $p$  个, 后者有  $n - p$  个。属性  $A_j$  的值域记为  $DOM(A_j)$ , 对于  $1 \leq j \leq p, DOM(A_j) = [0, 1]$ ; 对于  $p + 1 \leq j \leq n, DOM(A_j) = \{1, 2, \dots, n_j\}$ , 其中  $n_j = |DOM(A_j)|$ , 为属性  $A_j$  符号值的个数。

**定义 2** 设  $X$  为定义 1 中的待聚类对象集, K-prototypes 的硬聚类算法 (HKP) 和模糊聚类算法 (FKP) 是将  $X$  划分为  $K$  个分类, 并使 (1) 式的目标成本函数最小。

$$F(W, Z) = \sum_{k=1}^K \sum_{i=1}^m w_{k,i}^\alpha d(Z_k, x_i) \quad (1)$$

满足:

$$0 \leq w_{k,i} \leq 1, 1 \leq k \leq K, 1 \leq i \leq m \quad (2)$$

$$\sum_{k=1}^K w_{k,i} = 1, 1 \leq i \leq m \quad (3)$$

$$0 < \sum_{i=1}^m w_{k,i} < m, 1 \leq k \leq K \quad (4)$$

式中,  $K (\leq m)$  是已知的目标聚类数;  $\alpha \in [1, \infty]$  为加权指数;  $Z$  为聚类中心矩阵,  $Z = [Z_1, Z_2, \dots, Z_K] \in R^{n \times K}, \forall k: 1 \leq k \leq K, j: 1 \leq j \leq n, Z_{k,j} \in DOM(A_j)$ ;  $W = [w_{k,i}]$  是一  $K \times m$  的实数关联矩阵, 表示对象  $x_i$  对聚类中心  $Z_k$  的隶属度;  $d(Z_k, x_i)$  是  $Z_k$  与  $x_i$  间的距离函数。

**定义 3** 设  $u, v$  为  $X$  或  $Z$  中的两个对象, 它们之间的距离定义为:

$$d(u, v) = d_e(u, v) + \gamma d_o(u, v) = \sqrt{\sum_{j=1}^p (u_j - v_j)^2} + \gamma \sum_{j=p+1}^n \delta_o(u_j, v_j)$$

式中,  $d_e(u, v)$  为对象  $u, v$  间在数值属性上的欧氏距离;  $d_o(u, v)$  为对象  $u, v$  间在符号属性上的重叠 (overlap) 距离; 对于符号属性  $A_j$ , 当  $u_j = v_j$  时  $\delta_o(u_j, v_j) = 1$ , 否则为 0;  $\gamma$  为平衡不同类型属性的权值因子。

**定理 1**<sup>[5,6]</sup> 设成本函数  $F(W, Z)$  中的  $Z$  固定为  $Z^*$ , 则满足 (2), (3), (4) 的  $W: W^* = \min_W F(W, Z^*)$ 。  $W^*$  的计算方法是:

1) 对于  $\alpha = 1$  (HKP):

$$w_{k,i}^* = \begin{cases} 1, & \text{若 } d(Z_k^*, x_i) \leq d(Z_l^*, x_i), 1 \leq l \leq K \\ 0, & \text{否则} \end{cases}$$

2) 对于  $\alpha > 1$  (FKP):

$$w_{k,i}^* = \begin{cases} 1, & \text{若 } Z_k^* = x_i \\ 0, & \text{若 } Z_l^* = x_i, l \neq k \\ \frac{1}{\sum_{l=1}^K \left[ \frac{d(Z_k^*, x_i)}{d(Z_l^*, x_i)} \right]^{1/(\alpha-1)}}, & \text{若 } Z_l^* \neq x_i, 1 \leq l \leq K \end{cases}$$

**定理 2**<sup>[5,6]</sup> 设成本函数  $F(W, Z)$  中的  $W$  固定为  $W^*$ , 则满足 (2), (3), (4) 的  $Z: Z^* = \min_Z F(W^*, Z)$ 。  $Z^*$  的计算方法是:

1) 对于数值型属性  $A_j (1 \leq j \leq p)$ :

$$Z_{k,j}^* = \frac{\sum_{i=1}^m w_{k,i}^* x_{i,j}}{\sum_{i=1}^m w_{k,i}^*}, 1 \leq k \leq K$$

2) 对于符号型属性  $A_j (p + 1 \leq j \leq n), Z_{k,j}^* = r \in DOM(A_j), 1 \leq k \leq K, r$  满足:

$$\sum_{i=1}^m (w_{k,i}^* | x_{i,j} = r) \geq \sum_{i=1}^m (w_{k,i}^* | x_{i,j} = t), 1 \leq t \leq n_j$$

**算法 1** FKP 算法

输入:  $X, K, \alpha$

输出:  $W, Z$

1) 选取初始聚类中心  $Z^{(0)}$ , 并按定理 1 确定  $W^{(0)}$ , 使  $F(W, Z^{(0)})$  最小,  $\tau = 0$ ;

2) 按定理 2 确定  $Z^{(\tau+1)}$ , 使  $F(W^{(\tau)}, Z)$  最小。若  $|F(W^{(\tau)}, Z^{(\tau+1)}) - F(W^{(\tau)}, Z^{(\tau)})| < \varepsilon$ , 算法停止, 返回  $(W^{(\tau)}, Z^{(\tau+1)})$ ;

3) 按定理 1 确定  $W^{(\tau+1)}$ , 使  $F(W, Z^{(\tau+1)})$  最小。若  $|F(W^{(\tau+1)}, Z^{(\tau+1)}) - F(W^{(\tau)}, Z^{(\tau+1)})| < \varepsilon$ , 算法停止, 返回  $(W^{(\tau+1)}, Z^{(\tau+1)})$ ;

4)  $\tau = \tau + 1$ , 转 2)。

## 2 聚类的有效性度量

在实际应用中, 需要对聚类效果进行度量, 这就是聚类的有效性度量问题。聚类的质量决定于初始值的设定、聚类数目及其他聚类参数的选择等诸多因素。下面给出三个在下节实验中将要用到的聚类有效性指标的定义。

**定义 4** 设  $y_i$  为对象  $x_i \in X$  的输入标注,  $C$  为最大标注数目;  $Class(Z_k)$  为聚类中心的归纳标注:  $Class(Z_k) = \arg \max_{c=1 \dots C} \left| \sum_{i: k = \arg \max_{l=1 \dots K} w_{l,i}} \delta_o(c, y_i) \right|$ ;  $f(x_i)$  为基于聚类的输出

标注:  $f(x_i) = Class(Z_k), k = \arg \max_{l=1 \dots K} w_{l,i}$ , 则聚类模型的正确率 (简称正确率) 定义为  $Ac(X, W, Z, K) = \frac{1}{m} \sum_{i=1}^m \delta_o(y_i, f(x_i))$ 。

**定义 5** 聚类的划分系数 (Partition Coefficient) 为

$PC(W, K) = \frac{1}{m} \sum_{k=1}^K \sum_{i=1}^m w_{k,i}^2$ ; 划分熵 (Partition Entropy) 为

$$PE(W, K) = -\frac{1}{m} \sum_{k=1}^K \sum_{i=1}^m w_{k,i} \log_K w_{k,i}.$$

划分系数反映了所有输入对象相对于聚类中心的接近程度。如果每个对象仅属于一类,且此时的  $w_{k,i}$  较大,则数据的不确定性就较小。划分熵反映了聚类结果的好坏:若所有的  $w_{k,i}$  接近 0 或 1,则所给出的聚类结果确定性高,熵就小;若所有的  $w_{k,i}$  接近  $1/K$ ,则聚类块的模糊程度高,从而熵就大,相应的聚类结果就差。

性质 1 FKP 算法的聚类有效性度量指标满足:

- 1)  $0 \leq PC(W, K), Ac(X, W, Z, K), PE(W, K) \leq 1$ ;
- 2) 对于 HKP:  $PC(W, K) = 1, PE(W, K) = 0$ ;
- 3)  $\lim_{\alpha \rightarrow +\infty} PC(W, K) = 1/K$ ;
- 4)  $\lim_{\alpha \rightarrow +\infty} PE(W, K) = 1$ 。

性质 1 中的 1)、2) 容易证明, 3)、4) 仅根据大量的实验结果归纳而出。对于 FCM 算法, 有定理<sup>[7]</sup>  $\lim_{\alpha \rightarrow +\infty} Z_k = \frac{1}{m} \sum_{i=1}^m x_i$ ,  $\lim_{\alpha \rightarrow +\infty} w_{k,i} = 1/K$  存在, 即当  $\alpha$  变得很大时, FCM 算法的聚类结果几乎总是数据集的中心, 各对象对于各分类 (聚类) 的隶属程度几乎相同, 而不管数据本身具有多么清晰的适合 FCM 算法的分类结构。因此, FCM 聚类的有效性可以从划分系数是否接近  $1/K$ , 划分熵是否接近 1 来判断。对于 FKP, 虽然无法从理论上证明性质 1 中的 3)、4), 但大量的实验结果表明 FKP 同样也存在这个规律。

### 3 实验结果

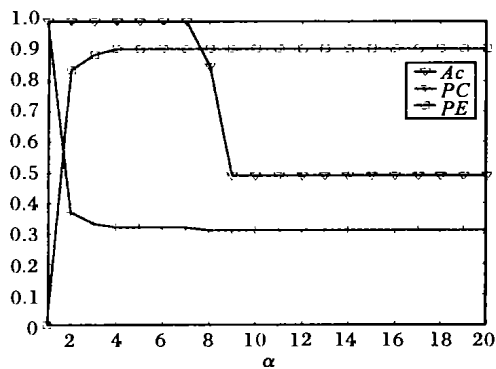


图 1  $\alpha$  与  $Ac, PC, PE$  的关系 ( $K = 4$ )

实验 1 第 1 个实验数据集是 Soybean-small<sup>[11]</sup>, 它包含 47 条记录, 由 35 个符号属性描述。数据集被分为 4 个疾病类 (在定义 4 中被称为输入标注), 分别含有 10, 10, 10 及 17 条记录。图 1 是  $K = 4$ ,  $\alpha$  由 1 变化到 20 时的聚类指标值。对于不同的  $\alpha$  值, 运行算法 30 次 (每次随机产生初始聚类中心), 得到平均的正确率、划分熵和划分系数。从图 1 可以观察到, 在  $Ac$  显著下降 ( $\alpha = 8$ ) 之前,  $PC$  早在  $\alpha = 2$  处下降到 0.37,  $PE$  则在  $\alpha = 4$  上升到 0.91。取  $\alpha = 20$  时第 1 类 10 个记录的隶属矩阵 (见表 1, 其中有几列不满足 3 式条件, 这是四舍五入造成的), 虽然它们可被一致地划分到 3 号聚类中, 但与其他聚类的隶属度差别已不分明。因此, 在给定  $K = 4$  的情况

下, 有理由相信, 合理的  $\alpha$  应该小于 2。图 2 是在  $K = 4$  时  $\alpha$  由 1 以 0.05 的间隔增到 2 时的运行结果。

表 1 Soybean-small 的部分隶属关系矩阵 ( $K = 4, \alpha = 20$ )

$K$	$W_{k,1}$	$W_{k,2}$	$W_{k,3}$	$W_{k,4}$	$W_{k,5}$	$W_{k,6}$	$W_{k,7}$	$W_{k,8}$	$W_{k,9}$	$W_{k,10}$
1	0.24	0.25	0.24	0.25	0.0	0.25	0.25	0.24	0.24	0.24
2	0.25	0.25	0.25	0.25	0.0	0.25	0.25	0.25	0.25	0.25
3	0.26	0.26	0.26	0.26	1.0	0.26	0.26	0.26	0.26	0.26
4	0.25	0.25	0.25	0.25	0.0	0.25	0.25	0.25	0.25	0.25

图 3 和图 4 给出了另外两组实验数据。图 3 和图 4 分别是固定  $\alpha = 1.1$  和  $\alpha = 2.0$ ,  $K$  由 4 变化到 20 时的运行结果。由图 3 可见, 对于较小的  $\alpha$  值,  $PC, PE$  基本不受  $K$  的影响, 显示了良好的聚类效果。

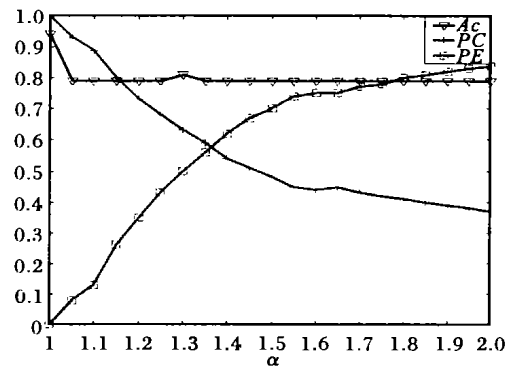


图 2  $\alpha$  与  $Ac, PC, PE$  的关系 ( $K = 4$ )

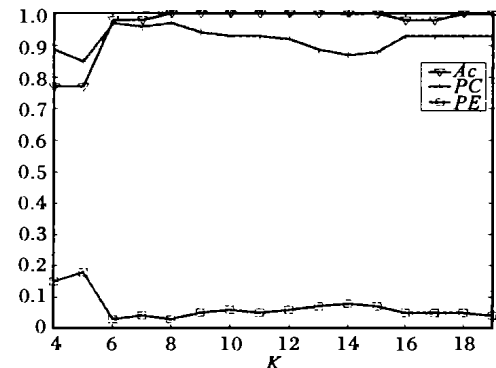


图 3  $K$  与  $Ac, PC, PE$  的关系 ( $\alpha = 1.1$ )

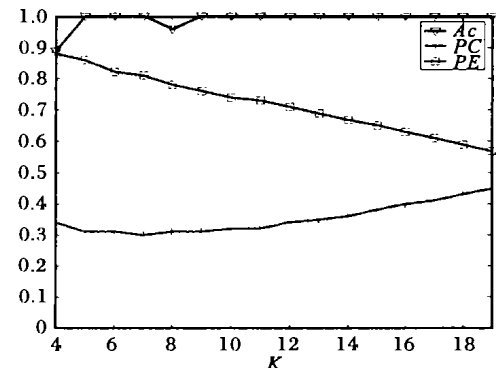


图 4  $K$  与  $Ac, PC, PE$  的关系 ( $\alpha = 2.0$ )

由图 4 可见, 随着  $K$  的递增,  $PE$  可以由不理想状态的 0.88 下降到可以接受的 0.57, 说明目标聚类数  $K$  对  $PC, PE$  的值具有一定的影响。如果我们主观地规定  $PE$  不能超过 0.8, 则在  $K < 8$  和  $\alpha = 2$  时就无法得到有效的聚类。为此, 要么增加  $K$ , 要么降低  $\alpha$  的值。

为了能够根据聚类有效性指标自动地探寻合理加权指数范围,对于数据集  $X$ , 给定  $K$  ( $C \leq K \leq K_{up}$ ,  $K_{up}$  为最大可接受目标聚类数), 判断是否存在  $\alpha$  ( $\alpha \in [1, 1, \alpha_0]$ ,  $\alpha_0$  为最大可能的加权指数) 使聚类有效。有效聚类的标准可以定义为:  $Ac > Ac_{low} \wedge PC > PC_{low} \wedge PE < PE_{up}$ , 其中  $Ac_{low}$ 、 $PC_{low}$ 、 $PE_{up}$  分别为分析者所认同的最低正确率、最低划分系数和最高划分熵。在判断标准中加入正确率限制分量是为了对所选择的  $K$  确保聚类模型符合数据集的聚类结构。算法 2 是本文给出的合理加权指数范围探寻算法 Searcha。

#### 算法 2 合理加权指数范围探寻算法 Searcha

输入:  $X, PE_{up}, PC_{low}, Ac_{low}, K_{up}, T, \alpha_0$

输出:  $\alpha_{up}, K, PC, PE, Ac$

- 1)  $t = 1$ ;
- 2)  $\alpha_t = \alpha_0 (1.1/\alpha_0)^{(t-1)/T}$ ;
- 3) for  $K = K_{up}$  downto  $C$ 
  - a)  $(W, Z) = FKP(X, K, \alpha_t)$ ;
  - b)  $PE = PE(W, K), PC = PC(W, K), Ac = Ac(X, W, Z, K)$ ;
  - c) if  $PE < PE_{up} \wedge PC > PC_{low} \wedge Ac > Ac_{low}$  then go 6);
  - 4)  $t = t + 1$ ;
  - 5) if  $t \leq T + 1$  go 2;
  - 6)  $\alpha_{up} = \alpha_{t-1}$

算法中,  $T$  为循环周期,  $\alpha$  由  $\alpha_0$  按指数下降, 直至找到满足有效聚类标准的加权指数  $\alpha_{up}$ 。算法应用了如下假设: 对于给定的  $K$ , 若加权指数  $\alpha_1 > \alpha_2$  且在  $\alpha_1$  时满足有效聚类标准, 则在  $\alpha_2$  时也一定满足。

表 2 数据集描述

数据集名称	对象数	属性数		类别数
		数值	符号	
Soybean-small	47	0	35	4
Soybean	683	0	35	19
Mushroom	8124	0	22	2
Tic-tac-toe	958	0	9	2
Vote	300	0	16	2
Zoo	101	1	15	7
Annealing	798	9	29	5
Crx	490	6	9	2
Labor-neg	40	8	8	2

表 3 合理加权指数范围

数据集名称	$\alpha_{up}$	$K$	$Ac$	$PC$	$PE$
Soybean-small	1.47	8	0.979	0.720	0.297
Soybean	1.27	38	0.753	0.771	0.137
Mushroom	1.27	4	0.818	0.710	0.374
Tic-tac-toe	1.27	4	0.654	0.732	0.340
Vote	1.47	4	0.890	0.613	0.471
Zoo	1.47	14	0.832	0.704	0.263
Annealing	1.27	10	0.812	0.775	0.197
Crx	1.27	4	0.782	0.781	0.283
Labor-neg	1.27	4	0.800	0.853	0.193

实验 2 从 UCI<sup>[11]</sup> 中选取 9 个含符号属性的实际数据集 (描述见表 2) 作为实验对象, 运行 Searcha 算法, 统一设定的参数是:  $PE_{up} = 0.8, PC_{low} = 0.6, Ac_{low} = 0.6, T = 20, \alpha_0 =$

20.0,  $K_{up} = 2C$ , 结果列于表 3。从运行的结果看, 3 个数据集的  $\alpha$  应该小于 1.47 ( $t = 19$ ), 而其他的 6 个则应该小于 1.27 ( $t = 21$ )。

实验结果显示, FKP 算法的合理加权指数范围并不象文献中对 FCM 给出 [1.5, 2.5] 的范围那么宽。由此, 也更加理解了 Z. Huang 在文献 [4] 的实验中采用 1.1 作为加权指数的合理性。

## 4 结语

在众多聚类算法中, 基于划分的模糊聚类算法 FCM 是模式识别中最常用的算法类型之一, 至今文献中仍不断有关的研究成果出现。为了能够处理如性别、颜色、形状、疾病类型等无顺序的符号类型数据, 可以选择 FKP 算法。

无论是 FCM, 还是 FKP, 加权指数  $\alpha$  对聚类效果均有重要影响。对于 FCM 中的  $\alpha$ , 已有相当多的研究, 如文献 [8, 9] 宣称最佳的  $\alpha$  可能位于 [1.5, 2.5] 区间; 而对于 FKP, 还未见文献明确讨论。

本文在前面给出了有效聚类的概念和判断标准, 并结合实验数据的分析提出了一个假设, 即对于给定的目标聚类数, 若加权指数  $\alpha_1 > \alpha_2$  且在  $\alpha_1$  时满足有效聚类标准, 则在  $\alpha_2$  时也一定满足。总体说, 有效聚类的判断标准是一个主观标准, 不同的分析者以及针对不同的数据对象可能会有不同的标准要求。

利用上述标准和假设, 本文提出了 FKP 中合适加权指数探寻算法。在多个实际数据集上的实验结果表明, 为进行有效的聚类, FKP 中加权指数应该小于 1.5 甚至更低, 这显然与 FCM 提出的最佳  $\alpha$  值不同。

## 参考文献:

- [1] JAIN A, DUBES R. Algorithm for clustering data[M]. Prentice - Hall, 1988.
- [2] RUSPINI E. A new approach to clustering[J]. Information & Control, 1969, 19: 22 - 32.
- [3] 史忠植. 知识发现[M]. 北京: 清华大学出版社, 2002.
- [4] HUANG Z. Extensions to the k-means algorithm for clustering large data sets with categorical values[J]. Data Mining Knowledge Discovery, 1998, 2(3): 283 - 304.
- [5] HUANG Z, NG M. A fuzzy k-modes algorithm for clustering categorical data[J]. IEEE Transaction on fuzzy systems, 1999, 7(4): 446 - 452.
- [6] CHEN N, CHEN A, ZHOU L. Fuzzy K-prototypes algorithm for clustering mixed numeric and categorical valued data (in English)[J]. 软件学报, 2001, 12(8): 1107 - 1119.
- [7] 于剑, 程乾生. 关于 FCM 算法中的权重指数  $m$  的一点笔记[J]. 电子学报, 2003, 31(3): 478 - 480.
- [8] PAL N, BEZDEK J. On cluster validity for the fuzzy c-means model [J]. IEEE Trans Fuzzy Systems, 1995, 3(3): 370 - 379.
- [9] 高新波, 裴纪红, 谢维信. 模糊  $c$ -均值聚类算法中加权指数  $m$  的研究[J]. 电子学报, 2000, 28(4): 80 - 83.
- [10] RUSPINI E. A new approach to clustering[J]. Information and Control, 1969, 15: 22 - 32.
- [11] BLAKE C, KEOGH E, MERZ C. UCI repository of machine learning database. 1998[EB/OL]. <http://www.ics.uci.edu/~mlearn/ML-Repository.html>, 2004.