

DOI:10.3969/j.issn.1673-4785.201410004.

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20150113.1130.006.html>

# 一种新的聚类有效性函数:模糊划分的模糊熵

卿铭<sup>1</sup>,孙晓梅<sup>2</sup>

(1.西南交通大学 数学学院,四川 成都 600031; 2.河南工程学院 数学与物理科学系,河南 郑州 451191)

**摘要:**模糊聚类分析结果是否合理的问题属于模糊聚类有效性判定课题,其核心是模糊聚类有效性函数的构造。文中基于序关系定义了模糊划分模糊熵来描述模糊划分的模糊程度。考虑到现有的一类有效的模糊聚类有效性函数就是基于数据集的模糊划分的,因此文中也用模糊划分的模糊熵作为聚类有效性函数。实验表明,模糊划分的模糊熵作为模糊聚类的有效性函数是合理的、可行的。

**关键词:**模糊  $C$  均值聚类;模糊划分的模糊熵;聚类有效性;聚类分析;模糊划分;模糊熵;熵函数;模糊集

**中图分类号:** TP O235 **文献标志码:** A **文章编号:** 1673-4785(2015)01-0075-06

中文引用格式:卿铭,孙晓梅.一种新的聚类有效性函数:模糊划分的模糊熵[J].智能系统学报,2015,10(1):75-80.

英文引用格式:QING Ming, SUN Xiaomei. A new clustering effectiveness function: fuzzy entropy of fuzzy partition[J]. CAAI Transactions on Intelligent Systems, 2015, 10(1): 75-80.

## A new clustering effectiveness function: fuzzy entropy of fuzzy partition

QING Ming<sup>1</sup>, SUN Xiaomei<sup>2</sup>

(1.School of Mathematics, Southwest Jiaotong University, Chengdu 610031, China; 2. Department of Mathematical and Physical Science, Henan Institution of Engineering, Zhengzhou 451191, China)

**Abstract:** In this paper, the determination that whether a fuzzy clustering analysis result is reasonable or not is decided by the effectiveness of fuzzy clustering and its core is the construction of fuzzy clustering effectiveness function. This paper proposed a new concept of fuzzy entropy for a fuzzy partition to describe fuzzy degree of a fuzzy partition based on an order relation. Fuzzy entropy for a fuzzy partition is also considered as a clustering effectiveness function because some existing fuzzy clustering effectiveness functions are based on fuzzy partition of data sets. The experiments demonstrated that it is reasonable and practicable to utilize fuzzy entropy for a fuzzy partition as the effectiveness function of a fuzzy clustering.

**Keywords:** fuzzy  $C$ -means clustering; fuzzy entropy of fuzzy partition; clustering effectiveness; clustering analysis; fuzzy partition; fuzzy entropy; entropy function; fuzzy set

聚类是一个在人类认识世界的过程中产生的古老问题。在面对纷繁的大千世界时,关注事物间的相似性并以此来区别不同的事物,对于人类来说是一种清晰而直观的思路。而每个概念的最初形成无不借助于事物间的聚类分析。因此聚类分析的研究不仅具有重要的理论意义,而且具有重要的工程应用价值和人文价值。对于给定的数据集,首先要判

断有无聚类结构,这属于聚类趋势研究的课题;如果已经确认有聚类结构,则需要用算法来确定这些结构,这属于聚类分析的研究课题;得到结构后,需要分析、判断聚类的结果是否合理,这属于聚类有效性研究的课题<sup>[1]</sup>。

传统的聚类分析是一种硬划分。它将每个待辨识的对象严格地划分到某个类中,其聚类结果具有非此即彼的性质。而实际上大多数对象并没有严格的属性,它们在形态和类属方面存在着中介性,往往

收稿日期:2014-10-08. 网络出版日期:2015-01-13.

基金项目:中央高校基础研究基金资助项目(2682014ZT28).

通信作者:卿铭. E-mail: qingming@swjtu.edu.cn.

具有亦此亦彼的性质,因此适合进行软划分。而模糊集理论的提出,对聚类分析软划分的发展起到了重要作用。用模糊的方法来处理聚类问题,称之为模糊聚类分析,是当前聚类分析研究的主流。

关于聚类有效性问题的研究通常是基于硬  $C$  均值聚类算法和模糊  $C$  均值聚类算法进行的。在模糊聚类有效性函数中,基于数据集的模糊划分的有效性函数是很重要的一类,其观点是:好的分类聚类应对应于数据集较分明的划分,划分的分明性越好,分类的不确定性就越小。尽管以 Dunn<sup>[4]</sup> 的划分系数为代表的聚类有效性函数取得一定的效果然而仍存在缺陷:划分系数有随类数增加而递减的趋势。为克服这一缺点,范九伦<sup>[2-3]</sup>、Dave<sup>[5]</sup> 和 Robubens<sup>[6]</sup> 作了很多的工作。但仍未较好解决这一难题——聚类有效性函数随类数增加而递减(增)的趋势。其他一些学者也对模糊聚类的有效性做了深入的讨论<sup>[7-20]</sup>,但都不完善,有继续改进的余地。

文中借鉴好的聚类应对应于数据集较分明的划分,划分的分明性越好,分类的不确定性就越小的学术思想,基于序关系定义模糊划分的模糊熵来反映模糊划分的模糊程度,并以其作为聚类有效性函数,实验证明模糊划分的模糊熵作为模糊聚类的有效性函数是合理、有效的。

### 1 FCM 算法

$X = \{x_1, x_2, \dots, x_n\} \subset R^s$  是数据集,  $n$  是数据的个数,  $c$  是事先确定的数据集的分类数目 ( $1 < c < n, c$  的最大取值限通常为  $c_{\max} \leq \sqrt{n}$ ),  $d_{ij} = ||x_i - v_j||$  是样本点  $x_i$  与聚类中心或聚类原型 (clustering prototype)  $v_j$  的某个距离度量 (反映了样本  $x_i$  与  $v_j$  的失真度),  $v_j \in R^s (1 \leq j \leq c)$ 。  $u_{ij}$  是第  $i$  个样本属于第  $j$  个中心的隶属度,则模糊聚类问题可表示成下面的数学规划问题<sup>[1]</sup>:

$$\begin{cases} \min J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m d_{ij}^2 \\ \text{s.t. } U \in M_{fc} \end{cases} \quad (1)$$

这里

$$M_{fc} = \{U \in R^{nc} \mid u_{ij} \in [0, 1], \forall i, j$$

$$\sum_{j=1}^c u_{ij} = 1, \forall i$$

$$0 < \sum_{i=1}^n u_{ij} < n, \forall j\}$$

是  $X$  的模糊  $c$  划分。因为  $X = \{x_1, x_2, \dots, x_n\}$  有限,

$\sum_{j=1}^c u_{ij} = 1, u_{ij} \in [0, 1], 1 \leq i \leq n$ , 所以  $U$  代表了  $X$  的一个模糊  $c$  划分,  $U$  的一列就是一个  $X$  上的模糊集。式(1)中参数  $m$  称为平滑因子,控制着模式在模糊类中的分享程度<sup>[8]</sup>,因此,模糊聚类要得到好的效果就必须选定一个合适的  $m$ 。然而,最佳  $m$  的选取目前尚缺乏理论指导。Bezdek 提出了交替优化算法(即人们所熟知的 FCM 算法)来解数学规划问题。Pal 和 Bezdek<sup>[9]</sup> 从聚类有效性的实验研究中得出  $m$  合理取值范围为  $[1.5, 2.5]$ 。

下面给出文中采用的从随机初始化模糊划分矩阵开始迭代 FCM 算法步骤<sup>[8-9]</sup>。

1) 初始化: 给定聚类类别数  $c (2 \leq c \leq n, n$  为样本点个数)、模糊加权指数  $m$ , 设定迭代停止阈值  $\varepsilon$ , 随机初始化划分矩阵  $U^{(0)}$ , 设置迭代计数器  $b = 0$ 。

2) 计算或更新聚类原型模式矩阵  $P^{(b)}$  :

$$p_i^{(b)} = \frac{\sum_{k=1}^n \mu_{ik}^{(b)} \cdot x_k}{\sum_{k=1}^n \mu_{ik}^{(b)}}, i = 1, 2, \dots, c$$

3) 更新模糊划分矩阵  $U^{(b+1)}$  ;

4) 若  $||U^{(b+1)} - U^{(b)}|| < \varepsilon$ , 输出划分矩阵  $U$  和聚类原型  $P$  等聚类信息, 算法结束; 否则, 令  $b = b + 1$ , 转向 2)。其中  $||\cdot||$  为合适的矩阵范数。

算法从随机初始化模糊划分矩阵开始迭代, 这样做的原因主要有 3 个: 1) 标准 FCM 算法本身并不能保证一定收敛到全局最优点, 算法可能收敛于某个极值点或局部最优点, 由于模糊划分矩阵的随机性选取, 通过对 FCM 算法的多次调用, 可能会发现全局最优点。 2) 若从初始化聚类原型模式矩阵开始迭代, 初始聚类原型的选取有很多种方法, 考虑到算法的通用性, 很难抉择一个优秀的选取方案, 并且一旦选取方案 (这里不包括随机初始化聚类原型) 确定了, 迭代过程就随之固定, 算法可能始终收敛于某个极值点或局部最优点。 3) 随机初始化聚类原型显然没有随机初始化划分矩阵方便, 而且也没有什么客观的理由。

### 2 模糊划分的模糊熵

文中基于序关系定义模糊划分的模糊熵反映了一个模糊划分的模糊程度。用  $F(X)$  表示  $X$  上所有模糊集, 对任意  $A, B \in F(X)$ , 称  $A \leq_m B$  当且仅当  $\bar{A} \leq \bar{B}$ , 其中  $\bar{A} = A \cap A^c, \bar{B} = B \cap B^c, A^c, B^c$  分别为  $A、$

$B$  的补集,  $\cap$  表示交集运算,  $\wedge$  是取小运算。

**定义 1** 若  $U_1 = [A_1, A_2, \dots, A_c], U_2 = [B_1, B_2, \dots, B_c]$  是  $X$  上的 2 个模糊  $c$  划分, 且满足  $A_j \leq_M^F B_j, j=1, 2, \dots, c$ , 则称  $U_1$  是  $U_2$  的分明修改, 记为  $U_1 <_M^F U_2$ 。

记  $X$  上的所有模糊划分为  $FP(X)$ 。  $[a] \in FP(X)$  表示每个元素都是常数  $a$  的划分矩阵。

**定理 1** ( $FP(X), <_M^F$ ) 满足以下性质:

- 1) ( $FP(X), <_M^F$ ) 是拟序集不是偏序集;
- 2) 硬划分是其极小元, 任何 2 个硬划分都可比, 且在同一个等价类中;
- 3)  $[1/c]$  是其最大元。

**证明** 1) 设  $U_1 = [u_{ij}], U_2 = [v_{ij}]$ 。

由定义 1,  $U_1 <_M^F U_2$  当且仅当任意  $j = 1, 2, \dots, c, A_j \leq_M^F B_j$ 。

即  $U_1 <_M^F U_2$  当且仅当任意  $i = 1, 2, \dots, n, j = 1, 2, \dots, c,$

$$\min\{u_{ij}, 1 - u_{ij}\} \leq \min\{v_{ij}, 1 - v_{ij}\} \leq 1/2。$$

对  $X$  上的任意模糊划分  $U, U <_M^F U$  显然成立。

若  $U_1 <_M^F U_2, U_2 <_M^F U_3$ , 则  $U_1 <_M^F U_3$ 。所以, ( $FP(X), <_M^F$ ) 是一个拟序集。显然 ( $FP(X), <_M^F$ ) 不是一个偏序集, 因为对任意 2 个不同的硬划分  $U_1, U_2$ , 都有  $U_1 <_M^F U_2$  且  $U_2 <_M^F U_1$ , 但  $U_2 \neq U_1$ 。

2) 证明略。

3) 由于  $U_1, U_2$  是模糊划分, 即  $\sum_{j=1}^c u_{ij} = 1,$

$\sum_{j=1}^c v_{ij} = 1$ , 所以, ( $FP(X), <_M^F$ ) 的最大元是  $[1/c]$  而不是  $[1/2]$ 。由于  $[1/c]$  是 ( $FP(X), <_M^F$ ) 最大元, 所以  $[1/c]$  是最模糊的划分。

**定义 2** 一个实函数  $E: FP(X) \rightarrow R^+$  称为有限论域  $X$  上模糊  $c$  划分的模糊熵 (简称模糊划分模糊熵), 如果  $E$  满足:

- 1)  $E(U, c) = 0$  当且仅当  $U$  是硬划分;
- 2) 若  $U_1$  和  $U_2$  都是模糊  $c$  划分, 且  $U_1 <_M^F U_2$ , 则  $E(U_1, c) \leq E(U_2, c)$ 。

若定义

$$E_p(U; c) = \frac{4}{c^p} \sum_{j=1}^c \sum_{i=1}^n u_{ij}(1 - u_{ij}) \quad (2)$$

式中: 参数  $p > 0$ 。

显然, 下述的定理 2、3 是成立的。

**定理 2**  $E_p(U; c)$  是有限论域  $X$  上模糊划分模糊熵。

**定理 3** 当  $1 < c < n$  时,  $E_p(U; c)$  有如下性质:

- 1)  $0 \leq E_p(U; c) \leq \frac{4n(c-1)}{c^{p+1}}$ ;
- 2)  $E_p(U; c) = 0 \Leftrightarrow U$  是硬划分;
- 3)  $U = [1/c] \Leftrightarrow E_p(U; c) = \frac{4n(c-1)}{c^{p+1}}$ 。

**证明** 1) 因为  $E_p(U; c) \geq 0$  明显成立, 只须证  $E_p(U; c) \leq \frac{4n(c-1)}{c^{p+1}}$ 。用拉格朗日乘数法:

$$L = \frac{4}{c^p} \sum_{j=1}^c \sum_{i=1}^n u_{ij}(1 - u_{ij}) + \lambda(1 - \sum_{j=1}^c u_{ij}),$$

则

$$\begin{cases} \frac{\partial L}{\partial u_{ij}} = \frac{4}{c^p}(1 - 2u_{ij}) - \lambda = 0 & (1 \leq i \leq n) \\ \sum_{j=1}^c u_{ij} = 1 & (1 \leq j \leq c) \end{cases},$$

解出  $u_{ij} = \frac{1}{c}, 1 \leq i \leq n, 1 \leq j \leq c$  为唯一驻点,

显然  $E_p(U; c) \leq E_p([1/c]; c) = \frac{4n(c-1)}{c^{p+1}}$ 。

性质 2)、3) 显然成立。

若记熵函数  $h(u) = 4u(1 - u)$ , 则

$$e(A_j) = \sum_{i=1}^n h(A(x_i))$$

为第  $j$  个类模糊集  $A_j$  的模糊熵,  $j = 1, 2, \dots, c$ , 则

$$E_p(U; c) = \frac{1}{c^p} \sum_{j=1}^c e(A_j)。$$

3) 如果采用其他的熵函数, 也可以得到类似的结果。

范九伦和吴成茂<sup>[3]</sup> 基于模糊熵来定义聚类有效性函数:

$$E_{FW}(U; c) = \sum_{j=1}^c e(A_j)。 \quad (4)$$

式中:  $e(A) = \frac{1}{n} \sum_{i=1}^n h(A(x_i))$ 。

显然式(3)与式(4)有类似之处。式(3)中  $p = 0$  即得式(4)。但强调 2 点: 1) 基于序关系得到公式(3); 2) 参数  $p$  在聚类有效性分析中也起作用, 给出参数  $p$  的初衷是克服划分系数及其他类似的聚类有效性函数的缺陷。

聚类有效性准则  $\Omega_c$  表示  $U \in M_{jc}$  的“最优”的有限集合, 若存在  $(U^*, c^*)$  满足  $E_p(U^*; c^*) = \min_c \{ \min_{\Omega_c} E_p(U; c) \}$  则  $(U^*, c^*)$  为最佳的有效性聚类,  $c^*$  为最好的分类数目。

### 3 模糊划分模糊熵作为聚类有效性函数的实验结果

文中采用 3 组不同的数据,对  $E_p(U;c)$  的聚类有效性进行实验。实验所用数据如表 1 所示。它们都是适合作为聚类有效性分析的典型数据集。

表 1 3 组典型实验数据

Table 1 Three types of classical experiment data

编号	样本点维数	最佳分类数	每类样本数	样本总数	说明
I	二维	6	100	600	由可分为 6 类的 42 个样本点扩展生成,每类数据呈均匀分布
II	三维	5	200	1 000	由三维空间中的 5 个点扩展生成,每类数据呈均匀分布
III	四维	4	100	400	Pal 和 Bezdek 使用过的四维正态分布数据

对数据集 I,调用 FCM 算法在不同分类数目  $c$  下的情况见图 1 至图 8,其中  $m=2, \varepsilon=10^{-4}$ 。

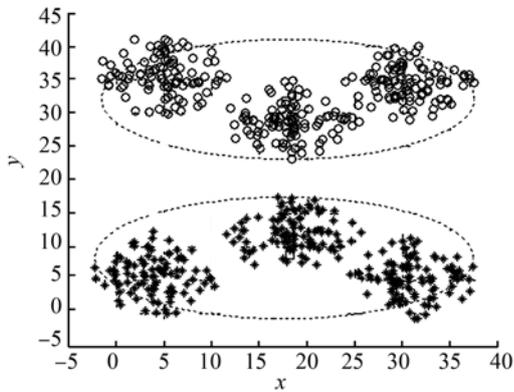


图 1  $c=2$  的聚类结果  
Fig.1 Clustering result for  $c=2$

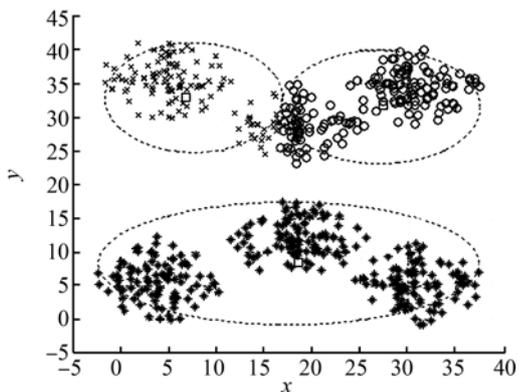


图 2  $c=3$  的聚类结果  
Fig.2 Clustering result for  $c=3$

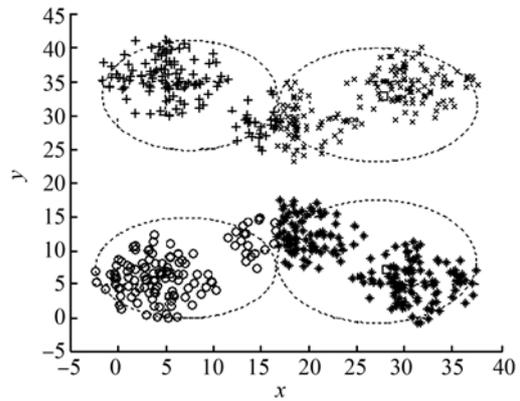


图 3  $c=4$  的聚类结果  
Fig.3 Clustering result for  $c=4$

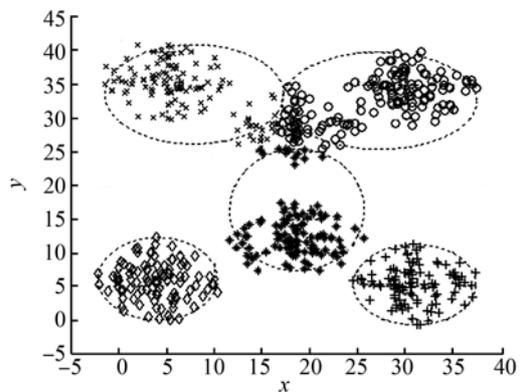


图 4  $c=5$  的聚类结果  
Fig.4 Clustering result for  $c=5$

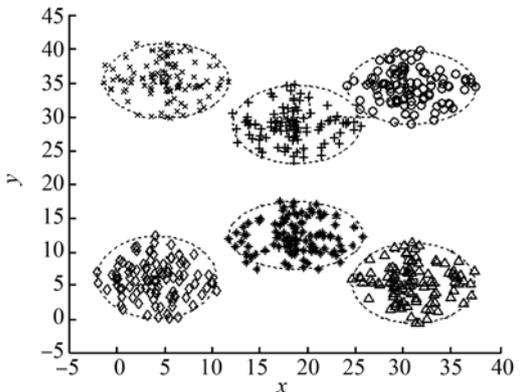


图 5  $c=6$  的聚类结果  
Fig.5 Clustering result for  $c=6$

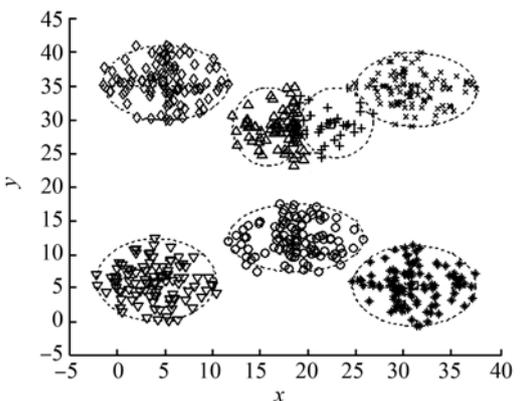


图 6  $c=7$  的聚类结果  
Fig.6 Clustering result for  $c=7$

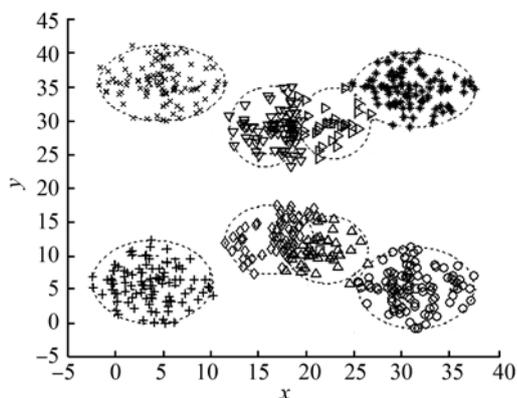


图 7  $c = 8$  的聚类结果

Fig.7 Clustering result for  $c = 8$

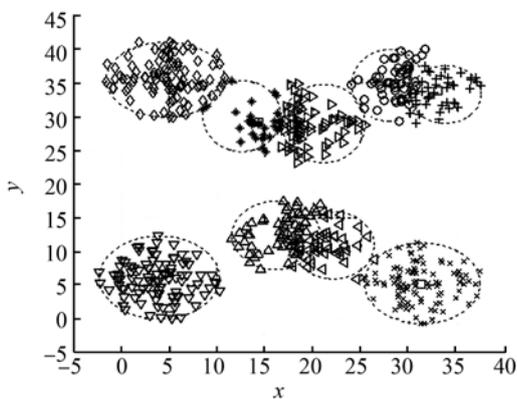


图 8  $c = 9$  的聚类结果

Fig.8 Clustering result for  $c = 9$

对数据集 II 与 III,取  $m = 2, \varepsilon = 10^{-4}$ 调用 FCM 算法在不同分类数目  $c$  下的聚类情况略去。

对数据集 I、II 与 III,取  $p = 1/n, \varepsilon = 10^{-4}$ 且  $m$  分别取 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5 时调用 FCM 算法在不同分类数目  $c$  下  $E_p(U;c)$  的取值情况做了计算。部分情况见表 2~4。表中黑体部分对应  $E_p(U;c)$  的最小值,相应的  $c$  值即聚类算法决定的“最优”分类数。

表 2 数据集 I  $E_p(U;c)$  的值

Table 2 Values of  $E_p(U;c)$  for data I

c	m		
	1.5	2.0	2.5
2	163	<b>563</b>	<b>793</b>
3	240	697	1011
4	256	664	1011
5	212	632	1052
6	<b>101</b>	574	1087
7	170	709	1225
8	246	795	1322
9	300	879	1395

可以知道数据集 I 真正的最优分类数是 6,由此可见此时  $E_p(U;c)$  在  $m = 1.5$  时有效,在  $m = 2.0$  和  $m = 2.5$  时未得到最优分类,得到的是次优分类。

表 3 数据集 II  $E_p(U;c)$  的值

Table 3 Values of  $E_p(U;c)$  for data II

c	m		
	1.7	2.0	2.3
2	507	<b>871</b>	<b>1152</b>
3	723	1175	1547
4	459	1036	1725
5	<b>448</b>	1019	1555
6	679	1319	1863
7	894	1541	2088
8	1105	1710	2264
9	1105	1710	2352

可以知道数据集 II 真正的最优分类数是 5,由此可见,此时  $E_p(U;c)$  在  $m = 1.7$  时有效,在  $m = 2.0$  和  $m = 2.3$  时未得到最优分类,得到的是次优分类。

表 4 数据集 III  $E_p(U;c)$  的值

Table 4 Values of  $E_p(U;c)$  for data III

c	m		
	1.6	2.0	2.4
2	511	<b>690</b>	<b>766</b>
3	514	790	934
4	<b>420</b>	780	980
5	571	929	1102
6	658	1 020	1 180
7	720	1 082	1 234
8	778	1 128	1 274
9	816	1 165	1 307

可以知道数据集 III 真正的最优分类数是 4,由此可见,此时  $E_p(U;c)$  在  $m = 1.6$  时有效,在  $m = 2.0$  和  $m = 2.4$  时未得到最优分类,得到的是次优分类。

#### 4 结束语

基于序关系定义了数据集模糊划分的模糊熵并将其作为模糊聚类有效性函数。从 3 个实验数据集的模糊聚类有效性验证来看,  $E_p(U;c)$  中的参数取  $p = 1/n$ , 也总存在合适的参数  $m$  (对数据集 I,  $m = 1.5$ ; 对数据集 II,  $m = 1.7$ ; 对数据集 III,  $m = 1.6$ ) 使得模糊划分的模糊熵  $E_p(U;c)$  作为聚类有效性函数是有效、可行的。下一步,将考虑参数  $p$  与  $m$  的最优组合问题。

#### 参考文献:

[1] 范九伦. 模糊聚类新算法与聚类有效性问题研究[D]. 西安:西安电子科技大学, 1998: 52-77.  
FAN Jiulun. The studies on some new algorithms of fuzzy

- clustering and clustering validity [D]. Xian: Xidian University, 1998: 1-165.
- [2] 范九伦, 吴成茂. 可能性划分系数和模糊变差相结合的聚类有效性函数[J]. 电子与信息学报, 2002, 24(8): 1017-1021.  
FAN Jiulun, WU Chengmao. Clustering validity function based on possibilistic partition coefficient combined with fuzzy variation [J]. Journal of Electronics and Information Technology, 2002, 24(8): 1017-1021.
- [3] 范九伦, 吴成茂. 基于模糊熵的聚类有效性函数[J]. 模式识别与人工智能, 2001, 14(4): 390-394.  
FAN Jiulun, WU Chengmao. Clustering validity function based on fuzzy entropy [J]. Pattern Recognition and Artificial Intelligence, 2001, 14(4): 390-394.
- [4] DUN J C. Some recent investigations of a new fuzzy partitions algorithm and its application to classification problems [J]. J Cybernet, 1974(4): 1-15.
- [5] DAVE R N. Validating fuzzy partition obtained through c-shells clustering [J]. Pattern Recognition Letters, 1996(17): 613-623.
- [6] ROBUBENS M. Pattern classification problems and fuzzy sets [J]. Fuzzy Sets Systems, 1978, 1: 239-253.
- [7] BEZDEK J C, HARRIS J. Fuzzy partitions and relations— an axiomatic basis for clustering [J]. Fuzzy Sets and Systems, 1978(1): 111-126.
- [8] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms [M]. New York: Plenum Press, 1981: 105-133.
- [9] PAL N R, BEZDEK J C. On cluster validity for fuzzy  $C$ -means model [J]. IEEE Trans Fuzzy Systems, 1995(3): 370-379.
- [10] ZHANG Huangxiang, LU Jin. Creating ensembles of classifiers via fuzzy clustering and deflection [J]. Fuzzy Sets and Systems, 2010(160): 1790-1802.
- [11] YU Jian, YANG Minshen, LEE E S. Sample-weighted clustering methods [J]. Computers and Mathematics with Applications, 2011(62): 2200-2208.
- [12] 岳士弘, 黄媿, 王鹏龙. 基于矩阵特征值分析的模糊聚类有效性指标[J]. 天津大学学报, 2014, 47(8): 689-696.  
YUE Shihong, HUANG Ti, WANG Penglong. Matrix eigenvalue analysis-based clustering validity index [J]. Journal of Tianjin University, 2014, 47(8): 689-696.
- [13] 张大庆, 徐再花. 一种新的模糊聚类有效性指标[J]. 沈阳农业大学学报, 2012, 14(5): 636-639.  
ZHANG Daqing, XU Zaihua. A new validity index for fuzzy clustering [J]. Journal of Shenyang Agricultural University, 2012, 14(5): 636-639.
- [14] 朱文婕, 吴楠, 胡学钢. 一个改进的模糊聚类有效性指标[J]. 计算机工程与应用, 2011, 45(5): 206-209.  
ZHU Wenjie, WU Nan, HU Xuegang. Improved cluster validity index for fuzzy clustering [J]. Computer Engineering and Applications, 2011, 45(5): 206-209.
- [15] 胡春春, 孟令奎, 谢文君, 等. 空间数据模糊聚类的有效性评价[J]. 武汉大学学报, 2007, 32(8): 740-743.  
HU Chunchun, MENG Lingkuai, XIE Wenjun, et al. Validity measure on fuzzy clustering for spatial data [J]. Journal of Wuhan University, 2007, 32(8): 740-743.
- [16] 郑弘亮, 徐本强, 赵晓慧, 等. 新的模糊聚类有效性指标[J]. 计算机应用, 2014, 34(8): 2166-2169.  
ZHENG Hongliang, XU Benqiang, ZHAO Xiaohui, et al. Novel validity index for fuzzy clustering [J]. Journal of Computer Applications, 2014, 34(8): 2166-2169.
- [17] 王丽娜, 马晓晓. 一种改进的模糊聚类有效性指标[J]. 微电子学与计算机, 2014, 31(4): 68-74.  
WANG Lina, MA Xiaoxiao. An improved validity index for clustering [J]. Microelectronics and Computer, 2014, 31(4): 68-74.
- [18] 张宇献, 刘通, 董晓, 等. 基于改进划分系数的模糊聚类有效性函数[J]. 沈阳工业大学学报, 2014, 36(4): 431-435.  
ZHANG Yuxian, LIU Tong, DONG Xiao, et al. Validity function for fuzzy clustering based on improved partition coefficient [J]. Journal of Shenyang University of Technology, 2014, 36(4): 431-435.
- [19] 普运伟, 金炜东, 朱明, 等. 核模糊  $C$  均值算法的聚类有效性研究[J]. 计算机科学, 2007, 34(2): 207-210.  
PU Yunwei, JIN Weidong, ZHU Ming, et al. On cluster validity for kernelized fuzzy  $C$ -means algorithm [J]. Computer Science, 2007, 34(2): 207-210.
- [20] 姚晓红, 任珂珂, 赵花妮, 等. 一种新的模糊聚类有效性指标的验证[J]. 洛阳理工学院学报, 2012, 22(3): 76-79.  
YAO Xiaohong, REN Keke, ZHAO Huani, et al. The verification of a new fuzzy clustering validity index [J]. Journal of Luoyang Institute of Science and Technology, 2012, 22(3): 76-79.

#### 作者简介:



卿强, 男, 1971 年生, 副教授, 主要研究方向为智能信息处理、系统可信性分析, 发表学术论文 20 余篇。



孙晓梅, 女, 1962 年生, 副教授, 主要研究方向为组合最优化。