Journal of Computer Applications
计算机应用,2014,34(8):2166-2169
ISSN 1001-9081
CODEN JYIIDU
2014-08-10
http://www.joca.cn

文章编号:1001-9081(2014)08-2166-04
doi:10.11772/j.issn.1001-9081.2014.08.2166

新的模糊聚类有效性指标

郑宏亮*,徐本强,赵晓慧,邹丽
(辽宁师范大学 计算机与信息技术学院,辽宁 大连 116081)
(*通信作者电子邮箱 zheng-hl@263.net)

摘　要:在经典的模糊C均值(FCM)算法中,聚类数需要预先给出,否则算法无法工作,这在一定程度上限制了FCM算法的应用范围。针对FCM算法中聚类数需要预先设定问题,提出了一种新的模糊聚类有效性指标。首先,通过运行FCM算法得到隶属度矩阵;然后,通过隶属度矩阵计算类内紧密性和类间重叠性;最后,利用类内的紧密性和类间的重叠性定义了一个新的聚类有效性指标。该指标克服了FCM算法中类数需要预先设定的缺点,利用该指标可以发现最符合数据自然分布的类的数目。通过对人工数据集和实际数据集的测试表明,对于模糊因子取1.8,2.0和2.2三个不同的常用值,均能发现最优聚类数。

关键词:模糊聚类;模糊C均值算法;有效性指标;模糊因子;最佳聚类数

中图分类号: TP18; TP391　　**文献标志码**:A

Novel validity index for fuzzy clustering

ZHENG Hongliang*, XU Benqiang, ZHAO Xiaohui, ZOU Li
(*School of Computer and Information Technology, Liaoning Normal University, Dalian Liaoning* 116081, *China*)

Abstract: It is necessary to pre-define a cluster number in classical Fuzzy C-means (FCM) algorithm. Otherwise, FCM algorithm can not work normally, which limits the applications of this algorithm. Aiming at the problem of pre-assigning cluster number for FCM algorithm, a new fuzzy cluster validity index was presented. Firstly, the membership matrix was got by running the FCM algorithm. Secondly, the intra class compactness and the inter class overlap were computed by the membership matrix. Finally, a new cluster validity index was defined by using the intra class compactness and the inter class overlap. The proposal overcomes the shortcomings of FCM that the cluster number must be pre-assigned. The optimal cluster number can be effectively found by the proposed index. The experimental results on artificial and real data sets show the validity of the proposed index. It also can be seen that the optimal cluster number are obtained for three different fuzzy factor values of 1.8, 2.0 and 2.2 which are general used in FCM algorithm.

Key words: fuzzy clustering; Fuzzy C-Means (FCM) algorithm; validity index; fuzzy factor; optimal cluster number

0　引言

聚类分析是数据挖掘领域中用于数据处理的重要方法之一。作为一种无监督的分类过程,没有类的先验知识可用。因此,如何发现最符合数据自然分布的类数,是研究聚类问题的一个最基本的问题。一般来说,聚类结果往往依赖于算法中的参数[1-3]。如何定义一个指标来发现最优的聚类数目,揭示数据集的内在结构,在过去的几十年里,许多学者借助于模糊C均值(Fuzzy C-Means,FCM)算法,对该问题进行了许多的研究工作[4-11]。

1974年,Bezdek利用FCM算法的隶属度矩阵,借助Dunn[12]的分离性指标第一次提出了有效性指标的划分系数(Partition Coefficient, PC)[13-14]和划分熵(Partition Entropy, PE)[15]的概念;Dave[16]为了减弱PC与PE指标随聚类个数的单调变化趋势,对其公式作了修改;Windham[17]利用隶属度最大值提出了比例指数(Windham Proportion Exponent, WPE)的有效性定义。另外,利用数据集类内紧密性和类间分离性的不同计算方法,产生了不同的有效性函数,如Fukuyama等[18]提出的V_{FS}有效性指标,Gunderson[19]的分离系数指标等。1979年,Davies等[20]利用类间的Fisher距离定义了分离性测度DB指标。Gath等[21]对于非欧氏距离的情况,基于模糊C均值引入了模糊超体积和模糊密度的概念,提出了(Fuzzy Hypervolume, FHV)指标。1991年,Xie等[22]利用FCM的优化目标函数和类间距离,定义了Xie-Beni聚类有效性指标。

对于具有良好分离性的数据集,已提出的各类有效性指标均能发现优化的聚类数目,这是因为这些指标充分考虑了类内的紧致性和类间的分离性。但是,对于多数情形下,数据的分布是非均匀的或类间有重叠的情况,已提出的各种有效性指标不能有效地发现理想的聚类数,因为这些指标没有考虑类间重叠的情况。而类间的数据重叠正是引起误分的原因之一,且在已存在的各类有效性指标公式中,均没有涉及到类间重叠的计算问题。针对这一问题,本文提出了一个新的有效性指标,充分考虑了类内的紧致性和类间的重叠度。在类间有重叠的情形下,利用所提出的指标,可以发现最优的聚类数目。部分克服了在类间有重叠的情形下,难以得到优化聚

收稿日期:2014-04-20;**修回日期**:2014-05-20。　**基金项目**:国家自然科学基金资助项目(61105059)。
作者简介:郑宏亮(1970-),男,辽宁铁岭人,讲师,硕士,主要研究方向:人工智能、数据挖掘;　徐本强(1978-),男,黑龙江双城人,讲师,硕士,主要研究方向:人工智能;　赵晓慧(1987-),女,辽宁大连人,硕士研究生,主要研究方向:数据挖掘;　邹丽(1971-),女,辽宁大连人,副教授,博士,CCF会员,主要研究方向:智能信息处理。

类数的弱点。实验结果表明，利用该指标可以有效地发现数据的最优聚类数。

1 新的模糊聚类有效性指标

本文提出的模糊聚类有效性指标，同时考虑了类内紧密性和类间重叠性的特点。聚类结果的类内紧密性越好，紧密性函数值越大；类间重叠性越低，重叠性函数值越小。显然，可以构造一个指标 F，对于理想的聚类结果使其达到最大值。

1.1 FCM 算法

聚类算法大体可以分成两类：硬聚类和软聚类。硬聚类是将数据集中的每个数据对象严格地划分到某一类中，划分界限十分鲜明；软聚类也被称作模糊聚类，每个数据对象以不同的隶属度被指派到每个类中。其中，应用范围最广、效果最为突出的是模糊 C 均值算法。

其算法描述如下。

给定数据集合 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \subset \mathbf{R}^s$，其中：$\boldsymbol{x}_i = [x_{1i}, x_{2i}, \cdots, x_{si}]^{\mathrm{T}} \in \mathbf{R}^s$，$n$ 为数据对象个数。则公式为：

$$\min J_m(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{X}) = \sum_{j=1}^{n} \sum_{i=1}^{C} u_{ij}^m \| \boldsymbol{x}_j - \boldsymbol{v}_i \|_A^2 ; 1 < m < \infty ,$$
$$\sum_{i=1}^{C} u_{ij} = 1, 1 \leqslant j \leqslant n, 1 \leqslant i \leqslant C, u_{ij} \in [0,1] \tag{1}$$

其中：$\boldsymbol{U} = [u_{ij}]_{C\times n}$，是数据对象 $\boldsymbol{x}_j$ 属于第 i 类的隶属度矩阵；C 为聚类个数$(1 < C < n)$；$\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_C]_{s\times C}$ 是类中心矩阵；m 是模糊因子，控制隶属度的模糊性，通常取 $m = 2$；矩阵范数 A 定义为数据对象 $\boldsymbol{x}_j$ 与第 i 类类中心的相似性度量规则，一般使用欧氏距离。μ_{ij} 和 $\boldsymbol{v}_i$ 的计算公式分别如下：

$$u_{ij} = 1 \Big/ \sum_{k=1}^{C} \left(\frac{\| \boldsymbol{x}_j - \boldsymbol{v}_i \|}{\| \boldsymbol{x}_j - \boldsymbol{v}_k \|} \right)^{\frac{2}{m-1}} ; 1 \leqslant i \leqslant C, 1 \leqslant j \leqslant n \tag{2}$$

$$\boldsymbol{v}_i = \left(\sum_{j=1}^{n} u_{ij}^m \boldsymbol{x}_j \right) \Big/ \left(\sum_{j=1}^{n} u_{ij}^m \right); \quad 1 \leqslant i \leqslant C \tag{3}$$

则 FCM 算法的计算步骤如下：

步骤 1　给定模糊因子 m，初始化类中心集合 $\boldsymbol{V}$ 及隶属度矩阵 $\boldsymbol{U} = [u_{ij}]_{C\times n}$。

步骤 2　根据式(3) 更新类中心矩阵 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_C]_{s\times C}$。

步骤 3　根据式(2) 更新隶属度矩阵 $\boldsymbol{U} = [u_{ij}]_{C\times n}$。

步骤 4　计算 J_m，ε 为终止阈值。若 $|J_m - J_m^{\text{previous}}| \leqslant \varepsilon$，则停止；否则转到步骤 2。

1.2 紧密性

类内数据的紧密程度是衡量模糊聚类结果有效性的重要标准和基本条件之一。基于 FCM 算法，本文提出了模糊聚类有效性指标中的紧密性定义，其公式如下：

$$Comp(k, \boldsymbol{U}) = \frac{1}{S} \times \sum_{i=1}^{n} \delta\left(\max_{1\leqslant c\leqslant k} u_{ic} \right);$$
$$\delta\left(\max_{1\leqslant c\leqslant k} u_{ic} \right) = \begin{cases} 1, & \max\limits_{1\leqslant c\leqslant k} u_{ic} \geqslant \alpha \\ \max\limits_{1\leqslant c\leqslant k} u_{ic}, & \beta \leqslant \max\limits_{1\leqslant c\leqslant k} u_{ic} < \alpha \\ 0, & \max\limits_{1\leqslant c\leqslant k} u_{ic} < \beta \end{cases} \tag{4}$$

其中：S 为最大隶属度 $\max\limits_{1\leqslant c\leqslant k} u_{ic} \geqslant \beta$ 的数据对象个数；n 为数据集中所有数据对象的个数；$\boldsymbol{U}$ 为隶属度矩阵；k 为聚类个数，其最大值一般取 $\sqrt{n}$；α 和 β 是两个参数。当数据对象的最大隶属度大于阈值 α 时，令 $\delta\left(\max\limits_{1\leqslant c\leqslant k} u_{ic} \right)$ 的值为 1，表明该数据对象属于对应的类；当最大隶属度介于 α 和 β 之间时，令 $\delta\left(\max\limits_{1\leqslant c\leqslant k} u_{ic} \right)$ 的值等于最大隶属度，记录了该数据对象最有可能属于某个类的程度；当最大隶属度小于 β 时，表明该数据对象隶属于某个类的程度较低，可能处于类间重叠区域。通过对 $\delta\left(\max\limits_{1\leqslant c\leqslant k} u_{ic} \right)$ 值的计算可以获得类内紧密程度。$Comp(k, \boldsymbol{U})$ 值越大，表明模糊聚类的类内紧密程度越高。

1.3 重叠性

对于 $\delta\left(\max\limits_{1\leqslant c\leqslant k} u_{ic} \right) = 0$，即 $\max\limits_{1\leqslant c\leqslant k} u_{ic} < \beta$ 的数据对象 $\boldsymbol{x}_i$，其最大隶属度没有达到阈值 β，所以它可能处于多个类边界的重叠区域。为了找到这样的数据点，设阈值 γ，对于 $\forall 1 \leqslant p \leqslant k$，$1 \leqslant q \leqslant k$，若存在 $|u_{ip} - u_{iq}| \leqslant \gamma$，则认为数据对象 $\boldsymbol{x}_i$ 处于类 p 和类 q 的重叠区域。重叠性定义为式(5)：

$$Overlap(k, \boldsymbol{U}) = \frac{1}{n} \times \frac{2}{k(k-1)} \times \sum_{i=1}^{n} \sum_{\forall 1\leqslant p,q\leqslant k}^{R} \varphi(|u_{ip} - u_{iq}|) \tag{5}$$

其中：$\varphi(|u_{ip} - u_{iq}|)_{1\leqslant p,q\leqslant k} = \begin{cases} 1, & \max\limits_{1\leqslant c\leqslant k} u_{ic} < \beta \text{ 且 } |u_{ip} - u_{iq}| \leqslant \gamma \\ 0, & \text{其他} \end{cases}$；

R 为满足 $\max\limits_{1\leqslant c\leqslant k} u_{ic} < \beta$ 且 $|u_{ip} - u_{iq}| \leqslant \gamma$ 条件的矩阵元素的个数；n 为数据集中全部数据对象的个数；$\boldsymbol{U}$ 为隶属度矩阵；k 为聚类个数，其最大值一般取 $\sqrt{n}$。设定阈值 γ，当最大隶属度小于 β 时，即该数据对象处于多个类边界的重叠区域，若同时满足 $|u_{ip} - u_{iq}| \leqslant \gamma$，表明该数据对象隶属于这两个类的程度相等，令此时的 $\varphi(|u_{ip} - u_{iq}|)_{1\leqslant p,q\leqslant k}$ 值等于 1，将所有符合以上条件的 $\varphi(|u_{ip} - u_{iq}|)_{1\leqslant p,q\leqslant k} = 1$ 相加求平均，则获得重叠性定义 $Overlap(k, \boldsymbol{U})$。$Overlap(k, \boldsymbol{U})$ 值越小，聚类重叠性程度越低。

1.4 提出的有效性指标

基于紧密性和重叠性定义，本文提出了一种新的模糊聚类有效性指标。对于紧密性和重叠性的计算，取 $c = 2, 3, \cdots, c_{\max}$，得到式(6) ~ (7)：

$$Comp(k, \boldsymbol{U}) = \{Comp(2, \boldsymbol{U}), Comp(3, \boldsymbol{U}), \cdots, Comp(c_{\max}, \boldsymbol{U})\} \tag{6}$$

$$Overlap(k, \boldsymbol{U}) = \{Overlap(2, \boldsymbol{U}), Overlap(3, \boldsymbol{U}), \cdots, Overlap(c_{\max}, \boldsymbol{U})\} \tag{7}$$

分别得到最大值如下：

$$Comp_{\max} = \underbrace{\max}_{k} Comp(k, \boldsymbol{U}) \tag{8}$$

$$Overlap_{\max} = \underbrace{\max}_{k} Overlap(k, \boldsymbol{U}) \tag{9}$$

利用最大值，对两者进行归一化处理，得到式(10) ~ (11)：

$$Comp^*(k, \boldsymbol{U}) = Comp(k, \boldsymbol{U}) / Comp_{\max} \tag{10}$$

$$Overlap^*(k, \boldsymbol{U}) = Overlap(k, \boldsymbol{U}) / Overlap_{\max} \tag{11}$$

其中：$Comp^*(k, \boldsymbol{U}) \in [0,1]$，$Overlap^*(k, \boldsymbol{U}) \in [0,1]$。结合式(10) 和式(11)，得到模糊聚类有效性指标，如式(12) 所示：

$$F = Comp^*(k, \boldsymbol{U}) - Overlap^*(k, \boldsymbol{U}) \tag{12}$$

该指标表明，模糊聚类的类内紧密程度越大，$Comp^*(k, \boldsymbol{U})$ 值越大；类间重叠程度越小，$Overlap^*(k, \boldsymbol{U})$ 值越小。由上述特点可得，聚类结果越好，F 值越大。因此，可以通过得到的最大 F 值，发现理想的聚类结果。本文的参数取值分别为：

$\alpha = 0.7$, $\beta = 0.6$, $\gamma = 0.11$。

2 实验结果与分析

为了证明该指标的可行性和有效性，本文进行了仿真数据和真实数据的测试，实验平台为主频 2.2 GHz，内存 1.00 GB，Windows XP 操作系统的电脑，测试软件使用 Matlab 2007。通过 FCM 算法对数据集进行聚类，模糊因子取 $m = 1.8, 2.0, 2.2$。由于篇幅有限，这里仅列出了部分结果。

仿真数据是聚类个数分别为 4 类与 5 类的高斯分布数据集，每类均含 100 个数据点，如图 1～2 所示。

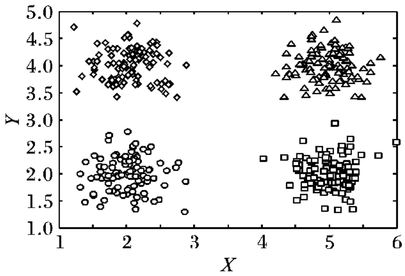

图 1 分 4 类的高斯分布数据集

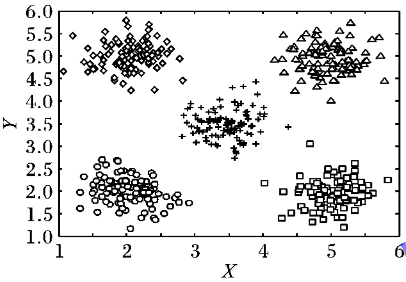

图 2 分 5 类的高斯分布数据集

使用本文提出的模糊聚类有效性指标式(12)，对这两个数据集进行聚类分析，得到的结果如表 1～2 所示。

表 1 4 类的高斯分布数据集的 F 值

m	C									
	2	3	4	5	6	7	8	9	10	11
1.8	**1.000**	0.995	**0.996**	0.955	0.887	0.790	0.641	0.648	0.515	0.513
2.0	**1.000**	0.986	**0.994**	0.921	0.853	0.733	0.661	0.575	0.432	0.395
2.2	**1.000**	0.949	**0.990**	0.917	0.799	0.776	0.488	0.439	0.306	0.325

表 2 5 类的高斯分布数据集的 F 值

m	C										
	2	3	4	5	6	7	8	9	10	11	12
1.8	0.719	0.965	0.726	**0.995**	0.910	0.840	0.800	0.727	0.531	0.443	0.332
2.0	0.838	0.917	0.836	**0.985**	0.904	0.802	0.778	0.701	0.656	0.492	0.363
2.2	0.842	0.920	0.884	**0.968**	0.866	0.804	0.677	0.647	0.463	0.374	0.262

由表 1 可知，当聚类个数为两类时，F 值最大，其次是聚类个数为 4 类的 F 值，说明最佳聚类个数为两类，其次数据集分布也可能具有 4 类的特点。结合图 1，数据分布最佳情况为左右两类，也可以分成 4 类，与实际情况吻合。当 $C = 2$ 时，F 值为其取值范围内的最大值 1.000，聚类个数为两类的情况最为理想。取 FCM 算法的不同模糊因子 m 值，对此模糊聚类有效性的计算并不产生影响，说明此指标具有较好的稳定性。

在表 2 中可以看到，当聚类个数为 5 类时，对应的 F 值最大，说明此时的聚类效果最好，$C = 5$ 为最佳聚类数，该结果与图 2 数据的实际分布吻合，并且模糊聚类有效性指标的计算不依赖于 m 值的变化。

下面对 6 个实际的数据集进行了测试。测试数据集为 UCI 中的 Iris 数据集、Wine 数据集、WBCD 数据集和 WDBC 数据集以及 Eamonn Keogh 提供的 SonyAIBORobot Surface 和 CBF。Iris 数据集中的每个数据对象均为四维，数据集共分 3 类，每一类含有 50 个样本，第一类与后两类线性分离，后两类之间存在重叠；WDBC 数据集有 569 个数据对象，分两类，每个数据对象有 30 个特征；Wine 数据集有 178 个数据对象，分 3 类，每个数据对象有 13 个特征；SonyAIBORobot Surface 分成两类，每条时间序列长度为 70，共有 20 条时间序列；CBF 分为 3 类，数据长度为 128，共有 30 个序列。对 6 个数据集进行模糊聚类，结果如表 3～5 所示，并将其与 7 个常用的模糊聚类的有效性指标：FS(Fukuyama and Sugeno)[14]、XB(Xie and Beni)[22]、SC(Separation Coefficient)[24]、FHV(Fuzzy Hyper Volume)[21]、PCAES(Partition Coefficient and Exponential Separation)[25]、PBMF(Pakhira Bandyopadhyay Maulik Fuzzy)[26] 和 CWB(Compose Within and Between scattering)[1] 进行比较，结果如表 6 所示(C^* 为原始的类数)。

表 3 Iris 数据集的 F 值

m	C										
	2	3	4	5	6	7	8	9	10	11	12
1.8	0.908	**0.906**	0.904	0.793	0.765	0.654	0.507	0.243	0.262	0.040	−0.033
2.0	0.935	**0.988**	0.801	0.787	0.621	0.472	0.258	0.373	0.116	−0.057	−0.114
2.2	0.940	**0.970**	0.784	0.715	0.592	0.511	0.359	0.136	0.003	0.123	0.115

表 4 Wine 数据集的 F 值

m	C					
	2	3	4	5	6	7
1.8	0.315	**0.781**	0.574	0.551	0.405	0.454
2.0	0.782	**1.000**	0.801	0.796	0.663	0.461
2.2	0.732	**0.956**	0.751	0.662	0.630	0.375

表 5 WDBC 数据集的 F 值

m	C										
	2	3	4	5	6	7	8	9	10	11	12
1.8	**1.000**	0.992	0.931	0.816	0.771	0.670	0.668	0.625	0.578	0.458	0.424
2.0	**1.000**	0.979	0.930	0.861	0.755	0.691	0.626	0.594	0.569	0.441	0.398
2.2	**0.933**	0.929	0.894	0.829	0.821	0.727	0.655	0.695	0.644	0.615	0.627

表 6 聚类个数结果对比

数据集	C^*	FS	XB	SC	FHV	PCAES	PBMF	CWB	F 值
Iris	3	5	2	3	3	2	3	2	3
Wine	3	13	3	2	3	2	3	2	3
WDBC	2	12	2	2	22	3	2	5	2
WBCD	2	12	2	2	2	2	2	2	2
CBF	3	3	3	3	3	3	3	3	3
SonyAIBORobot Surface	2	2	2	2	2	2	2	2	2

本文所提出的聚类有效性指标是基于 FCM 算法得到的隶属度计算出来的，但由于 FCM 算法不能有效地聚类高维数据[23]，所以本文并没有对于高维数据集进行测试。实验结果表明，对于实际的数据集，可以根据所提出的指标发现最佳的聚类数目，且与实际结果相吻合。对于 3 个常用的模糊因子 m，都可以发现正确的聚类数，表明本文提出的有效性指标对模糊因子 m 具有良好的稳定性。

3 结语

借助于 FCM 算法，本文提出了一个聚类的有效性指标。

利用所提出的指标,可以自动获得最佳聚类数,实现聚类的无监督学习过程。所提出的指标的优点是:1)克服了 FCM 算法中需要预先指定类数的缺点;2)克服了在使用距离度量的情况下,由于数据分布的不均匀性导致计算模糊程度不准确的缺点。实验结果表明了所提出指标的有效性,且所提出的指标对模糊子具有较好的鲁棒性。由于数据分布特征的多样性,如何定义一个更好的有效性指标,快速地发现最符合数据自然分布的聚类结果,仍是一个值得深入研究的问题。

参考文献:

[1] REZAEE B. A cluster validity index for fuzzy clustering [J]. Fuzzy Sets and Systems, 2010, 161(23): 3014 -3025.

[2] ZHANG Y J, WANG W N, ZHANG X, *et al.* A cluster validity index for fuzzy clustering [J]. Information Sciences, 2008, 178(4): 1205 -1218.

[3] ZALIK K R. Cluster validity index for estimation of fuzzy clusters of different sizes and densities [J]. Pattern Recognition, 2010, 43 (10): 3374 -3390.

[4] KIMA D W, LEE K H, LEE D. On cluster validity index for estimation of the optimal number of fuzzy clusters [J]. Pattern Recognition, 2004, 37(10): 2009 -2025.

[5] WANG W, ZHANG Y. On fuzzy cluster validity indices [J]. Fuzzy Sets and Systems, 2007, 158(19): 2095 -2117.

[6] YUE S, WANG J, WU T, *et al.* A new separation measure for improving the effectiveness of validity indices [J]. Information Sciences, 2010, 180(5): 748 -764.

[7] HUANG K. Applications of an enhanced cluster validity index method based on the fuzzy C-means and rough set theories to partition and classification [J]. Expert Systems with Applications, 2010, 37 (12): 8757 -8769.

[8] MASSON M, WU T. ECM: an evidential version of the fuzzy C-means algorithm [J]. Pattern Recognition, 2008, 41(4): 1384 -1397.

[9] WU K, YANG M. Robust cluster validity indexes [J]. Pattern Recognition, 2009, 42(11): 2541 -2550.

[10] SUN X, ZHAO Y, WANG H-L, *et al.* Sensitivity of digital soil maps based on FCM to the fuzzy exponent and the number of clusters [J]. Geoderma, 2012, 171/172: 24 -34.

[11] MITRA S, PEDRYCZ W, BARMAN B. Shadowed C-means: integrating fuzzy and rough clustering [J]. Pattern Recognition, 2010, 43(4): 1282 -1291.

[12] DUNN J C. Well-separated clusters and the optimal fuzzy partitions [J]. Journal of Cybernetics, 1974, 4(1): 95 -104.

[13] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms [M]. Norwell: Kluwer Academic Publishers, 1981.

[14] BEZDEK J C. Cluster validity with fuzzy sets [J]. Journal of Cybernetics, 1974, 3(3): 58 -73.

[15] BEZDEK J C. Numerical taxonomy with fuzzy sets [J]. Journal of Mathematical Biology, 1974, 1(1):57 -71.

[16] DAVE R N. Validating fuzzy partitions obtained through C-shells clustering [J]. Pattern Recognition Letters, 1996, 17(6): 613 - 623.

[17] WINDHAM M P. Cluster validity for the fuzzy C-means clustering algorithm [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1982, 4(4): 357 -363.

[18] FUKUYAMA Y, SUGENO M. A new method of choosing the number of clusters for the fuzzy C-means method [C]// Proceedings of the 5th Fuzzy Systems Symposium. Kobe: [s. n.], 1989: 247 - 250.

[19] GUNDERSON R. Applications of fuzzy ISODATA algorithms to star-tracker printing systems [C]// Proceedings of the 7th Triennial World IFAC Congress. Helsinki: [s. n.], 1978: 1319 -1323.

[20] DAVIES D L, BOULDIN D W. A cluster separation measure [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1979, 1(2): 224 -227.

[21] GATH I, GEVA A B. Fuzzy clustering for the estimation of the parameters of the components of mixtures of normal distributions [J]. Pattern Recognition Letters, 1989, 9(2): 77 -86.

[22] XIE X, BENI G. A validity measure for fuzzy clustering [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(4): 841 -847.

[23] WINKLER R, KLAWONN F, KRUSE R. Fuzzy C-means in high dimensional spaces [J]. International Journal of Fuzzy System Applications, 2011, 1(1): 1 -16.

[24] ZAHID N, LIMOURI M, ESSAID A. A new cluster validity for fuzzy clustering [J]. Pattern Recognition, 1999, 32(7): 1089 - 1097.

[25] WU K, YANG M. A cluster validity index for fuzzy clustering [J]. Pattern Recognition Letters, 2005, 26(9): 1275 -1291.

[26] PAKHIRA M K, BANDYOPADHYAY S, MAULIK U. Validity index for crisp and fuzzy clusters [J]. Pattern Recognition, 2004, 37(3): 487 -501.

(上接第 2160 页)

[11] LIU J, MARTINEZ L, CALZADA A, *et al.* A novel belief rule base representation, generation and its inference methodology [J]. Knowledge-Based Systems, 2013, 53: 129 -141.

[12] HU Y. The application of the 80-20 rule in the evaluation of key performance indicators [J]. Construction Machinery and Maintenance, 2009(5): 116 -117. (胡玉美. 二八法则在关键绩效指标考评中的应用[J]. 工程机械与维修, 2009(5): 116 -117.)

[13] CHEN Y-W, YANG J-B, XU D-L, *et al.* Inference analysis and adaptive training for belief rule based systems [J]. Expert Systems with Applications, 2011, 38(10): 12845 -12860.

[14] YANG J-B. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties [J]. European Journal of Operational Research, 2001, 131(1): 31 -61.

[15] CHEN Y-W, YANG J-B, XU D-L, *et al.* On the inference and approximation properties of belief rule based systems [J]. Information Sciences, 2013, 234: 121 -135.

[16] LIU J, MARTINEZ L, RUAN D, *et al.* Optimization algorithm for learning consistent belief rule-base from examples [J]. Journal of Global Optimization, 2011, 51(2): 255 -270.

[17] JIN Y, von SEELEN W, SENDHOFF B. On generating FC^3 fuzzy rule systems from data using evolution strategies [J]. IEEE Transactions on System, Man and Cybernetics, Part B: Cybernetics, 1999, 29(6): 829 -845.