

基于直觉模糊解析面积的聚类算法^①

程魏魏^{1,2}, 姚伟烈², 彭艳兵²

¹(武汉邮电科学研究院, 武汉 430070)

²(南京烽火软件科技有限公司, 南京 210019)

通讯作者: 程魏魏, E-mail: 1254937362@qq.com

摘要: 针对现有的直觉模糊聚类算法对权重的忽视或误用, 提出一种基于直觉模糊解析面积的聚类算法. 同时给出了直觉模糊集的解析面积和属性权重的计算方法, 然后构造了聚类算法的目标函数, 并给出聚类算法的详细步骤. 算例验证了所提出的算法的合理性与可行性.

关键词: 直觉模糊集; 解析面积; 属性权重; 直觉模糊聚类

引用格式: 程魏魏, 姚伟烈, 彭艳兵. 基于直觉模糊解析面积的聚类算法. 计算机系统应用, 2018, 27(4): 215-218. <http://www.c-s-a.org.cn/1003-3254/6278.html>

Cluster Method Based on Intuitionistic Fuzzy Analytic Area

CHENG Wei-Wei^{1,2}, YAO Wei-Lie², PENG Yan-Bing²

¹(Wuhan Research Institute of Posts and Telecommunications, Wuhan 430070, China)

²(Nanjing Fiberhome Software Science and Technology Co. Ltd., Nanjing 210019, China)

Abstract: Because of the neglect or misuse of the attribute weight of the existing intuitionistic fuzzy clustering algorithm, an intuitionistic fuzzy clustering algorithm based on the intuitionistic fuzzy analytic area is proposed, which introduces the analytic area of intuitionistic fuzzy sets and a method for calculating the attribute weights, then constructs the objective function of clustering and gives the clustering procedure. Finally, an example shows the validity and feasibility of this algorithm.

Key words: intuitionistic fuzzy sets; analytic area; attribute weights; intuitionistic fuzzy clustering

1 引言

聚类分析是将未知类别的事物根据其不同特征以及相似性等关系来进行分类. 在传统的聚类方法中, 样本所属的分类都是确定的, 但在实际当中大多数事物间的界限并不明显, 即不能清楚的表达非此即彼的性质. Zadeh^[1]于 1965 年提出了模糊集理论用来描述事物外延不清楚的亦此亦彼的模糊概念. 模糊集理论以模糊数学为基础, 研究有关非精确的现象, 最基本的特征是承认差异的中间过渡, 也即承认渐变的隶属关系. 模糊集理论在实际应用中几乎涉及到国民生产生活的各个方面, 例如, 对我国农村贫困线的测度^[2]等.

Ruspini^[3]于 1969 年首先提出了模糊规划概念, 将模糊理论引入到聚类分析中. 随后, 研究者提出了多种模糊聚类算法. 其中, Dunn^[4]和 Bezdek^[5]提出的模糊 C 均值 (FCM, fuzzy C-means) 算法是一种基于目标函数的聚类算法, 通过优化目标函数从而得到每个待分类对象对所有聚类中心的隶属度, 继而决定分类对象的类目进而达到自动分类的目的. 该算法被广泛应用于图像处理^[6]、模式识别^[7]和网络安全^[8]等领域.

Atanassov^[9]于 1986 年提出了直觉模糊集 (Intuitionistic Fuzzy Sets, IFS) 的概念, 它是在 Zadeh 提出的模糊集的基础上增加了非隶属度参数, 可更加灵

① 收稿时间: 2017-07-05; 修改时间: 2017-07-20; 采用时间: 2017-07-28; csa 在线出版时间: 2018-03-31

活细腻地刻画客观事物非此非彼的模糊概念,与模糊集比较,它能更好、更准确地表达模糊信息.同时,基于直觉模糊集的理论研究也如雨后春笋般显现^[10].因此,许多学者将FCM算法拓展到直觉模糊集领域进行研究^[11,12].文献[13]将FCM算法拓展到聚类对象和聚类中心都用直觉模糊集表示的聚类中,提出了新的直觉模糊C均值(Intuitionistic Fuzzy C-means, IFCM)聚类方法.虽然考虑到了指标权重的问题,但其聚类结果却是以实数形式表征,忽视了直觉模糊集的不确定性.文献[14]提出了一种基于集值统计的直觉模糊聚类方法,虽然其运用直觉模糊数来表征相似度,但没有考虑聚类对象的属性权重问题,而且计算量较大不能很好的应用于实际中.文献[15]在计算量方面进行了优化,但也没有考虑属性权重的问题,并且在得到直觉模糊相似矩阵后仅仅采用隶属度作为参考进行聚类,忽略了直觉模糊集的非隶属度和犹豫度,这势必会造成信息的丢失,影响聚类结果.文献[16]提出了一种基于新直觉模糊相似度的聚类算法,该算法引入了风险参数,需要决策者依赖自己的主观决定,降低了数据的客观性.虽然该聚类算法考虑到了属性权重,但是基于直觉模糊熵计算的属性权重并不能提高聚类效果.文献[17]提出了一种粒子群优化的直觉模糊核聚类算法,该算法存在很多不完善的地方,例如怎样确定算法的平滑参数以及算法的泛化性能差.文献[18]提出了一种基于相对熵的直觉模糊聚类方法,该算法对传统的相对熵进行了改进,但是属性权重却是以样本的平均值表示,忽视了属性权重对聚类的影响.文献[19]提出了一种基于新直觉模糊相似度量度的直觉模糊谱聚类算法,该算法定义了新的直觉模糊相似度量方法,未考虑属性权重对聚类的影响,且其计算结果是实数,忽视了直觉模糊集的不定性.

针对现有方法存在的问题,本文提出了一种基于直觉模糊解析面积的聚类算法,并且结合一种新的提高聚类效果的属性权重.该方法解决了现有直觉模糊聚类算法在聚类过程中没有考虑属性权重或属性权重计算不科学的问题.

2 预备知识

2.1 直觉模糊集

设 X 是一个给定的非空论域,集合 X 上的一个直觉模糊集^[9]表示为:

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \} \quad (1)$$

其中, $\mu_A(x)$ 和 $\nu_A(x)$ 分别表示元素 x 属于 A 的隶属度与非隶属度: $\mu_A: X \rightarrow [0, 1]$, $\nu_A: X \rightarrow [0, 1]$ 且满足: $0 \leq \mu_A(x) + \nu_A(x) \leq 1, \forall x \in X$ 则称 $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ 为元素 x 属于 A 的犹豫度;称 $\alpha = \langle \mu_A(x), \nu_A(x) \rangle$ 为直觉模糊数.

2.2 直觉模糊熵

设 $X = \{x_1, x_2, \dots, x_n\}$ 是一个给定的非空论域, $A = \{ \langle x_i, \mu_A(x_i), \nu_A(x_i) \rangle \mid x_i \in X \}$ 为论域上的直觉模糊集,则 A 的直觉模糊熵^[20]为:

$$E(A) = \frac{1}{n} \sum_{i=1}^n \frac{1 - |\mu_A(x_i) - \nu_A(x_i)| + \pi_A(x_i)}{1 + |\mu_A(x_i) - \nu_A(x_i)| + \pi_A(x_i)} \quad (2)$$

特别地,对于直觉模糊数 $\alpha = \langle \mu_A(x), \nu_A(x) \rangle$,其直觉模糊熵为:

$$E(\alpha) = \frac{1 - |\mu_A(x) - \nu_A(x)| + \pi_A(x)}{1 + |\mu_A(x) - \nu_A(x)| + \pi_A(x)} \quad (3)$$

对于任意的直觉模糊数 α ,其直觉模糊熵越大,说明该直觉模糊数的不确定性越大.依据直觉模糊熵确定的属性权重会影响聚类效果,并且客观性较差.

2.3 直觉模糊解析面积

设 A 和 B 是非空论域 X 上的两个直觉模糊集, $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \}$, $B = \{ \langle x, \mu_B(x), \nu_B(x) \rangle \mid x \in X \}$,则

$$S(A, B) = \sqrt{D_\mu^2 + D_\nu^2 + D_\pi^2} \quad (4)$$

为直觉模糊集 A 和 B 的解析面积^[21],其中 $D_\mu = |\nu_A(x)\pi_B(x) - \nu_B(x)\pi_A(x)|$, $D_\nu = |\mu_A(x)\pi_B(x) - \mu_B(x)\pi_A(x)|$, $D_\pi = |\mu_A(x)\nu_B(x) - \mu_B(x)\nu_A(x)|$.

并且,两个直觉模糊集的解析面积是从空间角度出发来刻画两个直觉模糊集的相似度.

2.4 直觉模糊解的交与并运算

设 $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \}$ 和 $B = \{ \langle x, \mu_B(x), \nu_B(x) \rangle \mid x \in X \}$ 为非空论域 X 上的两个直觉模糊集,则直觉模糊解的交与并运算^[22]如下所示:

$$A \cap B = \quad (5)$$

$$\{ \langle x, \min(\mu_A(x), \mu_B(x)), \max(\nu_A(x), \nu_B(x)) \rangle \mid x \in X \}$$

$$A \cup B = \quad (6)$$

$$\{ \langle x, \max(\mu_A(x), \mu_B(x)), \min(\nu_A(x), \nu_B(x)) \rangle \mid x \in X \}$$

$$R(A, B) = S(A \cap B, A \cup B) \quad (7)$$

其中,公式(7)是基于直觉模糊解析面积的极差.

3 基于直觉模糊解析面积的聚类算法

假设直觉模糊多属性聚类问题有 n 个分类对象 E_1, E_2, \dots, E_n , m 个评价属性 I_1, I_2, \dots, I_m , 分类对象 E_i 在评价属性 I_j 下的属性为直觉模糊数 $d_{ij} = \langle \mu_{ij}, \nu_{ij} \rangle$, 所以直觉模糊聚类矩阵为 $D = (d_{ij})_{n \times m}$.

评价属性权重的计算是按照属性的离散程度计算的,要使得分类明显并且客观,权重的计算必须使得数据容易分类.本文采用直觉模糊集的极差来计算权重 $W = \omega_1, \omega_2, \dots, \omega_m$, 如公式(8)所示.极差描述了一组数据的离散程度,极差越大,离散程度越大,反之,离散程度越小.而这里采用极差来计算属性权重,取 n 个分类对象的直觉模糊解析面积极差与在 m 个评价属性基础上的 n 个分类对象的直觉模糊解析面积极差的比值,这样能使得模糊集数据容易分类.

$$\omega_k = \frac{R\left(\bigcap_{i=1}^n d_{ik}, \bigcup_{i=1}^n d_{ik}\right)}{\sum_{j=1}^m R\left(\bigcap_{i=1}^n d_{ij}, \bigcup_{i=1}^n d_{ij}\right)}, k=1, 2, \dots, m \quad (8)$$

这里,为避免某个聚类对象属于离群点影响聚类的结果,故聚类中心的初始化则随机选择 $c(c < n)$ 个聚类对象,将其属性的直觉模糊数随机组合成新对象来作为聚类中心 $V_1 = (v_{ij})_{c \times m}$.各个分类对象 E_i 与某聚类中心 $V_1 = (v_{ij})_{c \times m}$ 的加权解析面积为:

$$WD_{il} = \sum_{j=1}^m \omega_j S(d_{ij}, v_{lj}) \quad (9)$$

聚类对象属于加权度量最小的一类,若某类没有任何对象则重新初始化聚类中心.一遍聚类结束后,则重新计算聚类中心,新聚类中心为该所有元素的各个隶属度与非隶属度的平均值.

聚类准则为求出合适的模糊聚类中心矩阵 V , 使目标函数 $J(D, V)$ 达到极小值.

$$J(D, V) = \sum_{i=1}^n \sum_{l=1}^c WD_{il} \quad (10)$$

综上所述可以得到如下直觉模糊聚类算法:

1) 由属性数据的离散程度,根据公式(8)计算各个属性的客观权重 $W = (\omega_1, \omega_2, \dots, \omega_m)$;

2) 确定聚类数目 c , 随机生成初始聚类中心 $V^{(0)} = (V_1^{(0)}, V_2^{(0)}, \dots, V_c^{(0)})^T$;

3) 计算各个分类目标与聚类中心的加权直觉模糊解析面积,判定聚类对象属于一类;

4) 重新计算聚类中心;

5) 比较 $V^{(l)}$ 与 $V^{(l+1)}$, 若取定精度 $\varepsilon > 0$, 如果 $\max\{|v_{ij}^{(l+1)} - v_{ij}^{(l)}|\} \leq \varepsilon$, 则停止迭代; 否则, 返回步骤3).

本文提出的直觉模糊聚类算法与现有方法进行对比,其特点是:

1) 本文提出的属性权重的计算是根据直觉模糊解析面积的极差来计算的,使得模糊集数据容易分类;

2) 本文提出的聚类方法将直觉模糊解析面积与聚类算法进行有效的结合,克服了现有算法针对属性权重计算的不客观性.

4 算例分析

为了便于对比,本文采用文献[13]的算例进行相应的分析.

某空中监视系统共发现 8 批目标,即 $A = \{A_1, A_2, \dots, A_8\}$, 每批目标用 4 个特征直觉模糊数表示,所有目标的目标特征数据如表 1 所示.

表 1 目标特征数据

目标	X_1	X_1	X_1	X_1
A_1	<0.56,0.34>	<0.40,0.50>	<0.30,0.40>	<0.71,0.10>
A_2	<0.41,0.40>	<0.08,0.80>	<0.05,0.75>	<0.20,0.50>
A_3	<0.38,0.52>	<0.90,0.10>	<0.80,0.10>	<0.01,0.80>
A_4	<0.31,0.60>	<0.40,0.50>	<0.30,0.50>	<0.63,0.15>
A_5	<0.31,0.61>	<0.74,0.22>	<0.70,0.25>	<0.00,0.90>
A_6	<0.44,0.45>	<0.11,0.80>	<0.06,0.80>	<0.31,0.25>
A_7	<0.58,0.30>	<0.37,0.52>	<0.30,0.50>	<0.45,0.35>
A_8	<0.43,0.45>	<0.14,0.72>	<0.07,0.70>	<0.25,0.55>

为了方便比较分析,聚类数目与初始聚类中心与文献[13]设置相同,将分类对象分成 3 类,即 $c=3$, 取初始聚类中心如表 2 所示,各个属性的交与并的结果如表 3 所示.由公式(8)可计算出各个属性的权重,如表 4 所示.

表 2 初始聚类中心设置

聚类中心	X_1	X_1	X_1	X_1
$V_1^{(0)}$	<0.7,0.1>	<0.2,0.7>	<0.1,0.2>	<0.7,0.2>
$V_2^{(0)}$	<0.2,0.7>	<0.1,0.2>	<0.1,0.7>	<0.1,0.7>
$V_3^{(0)}$	<0.1,0.2>	<0.1,0.7>	<0.7,0.2>	<0.2,0.1>

表 3 各个属性的交与并结果

交集结果	<0.58,0.30>	<0.90,0.10>	<0.80,0.10>	<0.71,0.10>
并集结果	<0.31,0.61>	<0.08,0.80>	<0.05,0.80>	<0.00,0.90>

表4 各个属性权重

属性	X_1	X_2	X_3	X_4
权重	0.2656	0.7202	0.6486	0.6628

取 $\varepsilon = 0.01$, 经过多次运算得到 $\max \left\{ \left| v_{ij}^{(3)} - v_{ij}^{(2)} \right| \right\} \leq \varepsilon$, 迭代终止并且得到最终的聚类中心如表5所示, 分类结果如表6所示。

表5 最终聚类中心

聚类中心	X_1	X_2	X_3	X_4
V_1	<0.48,0.42>	<0.39,0.51>	<0.30,0.47>	<0.60,0.20>
V_2	<0.34,0.56>	<0.82,0.16>	<0.75,0.18>	<0.01,0.85>
V_3	<0.43,0.43>	<0.11,0.77>	<0.06,0.75>	<0.25,0.43>

表6 聚类结果

聚类中心	元素
V_1	A_1, A_4, A_7
V_2	A_3, A_5
V_3	A_2, A_6, A_8

从表6可以看出第一类为 $\{A_1, A_4, A_7\}$, 第二类为 $\{A_3, A_5\}$, 第三类为 $\{A_2, A_6, A_8\}$, 可以看出, 所得到的分类结果与实际情况相同, 说明本方法是有效的。

该算法与文献[13]提出的新IFCM算法均考虑到了属性权重, 最终聚类的结果也是一致的。但文献[13]存在些缺陷, 其属性权重只是笼统的表示取值范围是[0-1]; 算法的迭代次数多。

本文提出的算法的迭代次数少于文献[13]提出的新IFCM算法, 说明本文提出的基于直觉模糊解析面积的聚类算法以及属性权重的选择比较恰当, 算法的复杂度低。

由于很多决策问题不能用纯数字描述, 只能用直觉模糊集描述, 这里该算法适用的数据集是一些需使用直觉模糊集描述的多属性决策问题或者是模糊语言描述的多属性决策问题。

5 结论与展望

本文将直觉模糊解析面积与聚类算法进行有效的结合, 提出了基于直觉模糊解析面积的聚类算法, 并且引入了提高聚类效果的属性权重。本文的权重计算完全依靠数据的离散程度而确定, 且克服了现有算法属性权重计算的不客观性, 最后通过实例验证了该算法的正确性和有效性。与此同时, 该算法还可以应用到多属性决策和机器学习的相关领域当中。

参考文献

1 Zadeh LA. Fuzzy sets. *Information and Control*, 1965, 8(3):

- 338-353. [doi: 10.1016/S0019-9958(65)90241-X]
- 2 张志国, 武力兵, 聂荣. 基于模糊集理论的中国农村贫困测度. *数学的实践与认识*, 2016, 46(23): 9-16.
- 3 Ruspini EH. A new approach to clustering. *Information and Control*, 1969, 15(1): 22-32. [doi: 10.1016/S0019-9958(69)90591-9]
- 4 Dunn JC. A graph theoretic analysis of pattern classification via Tamura's fuzzy relation. *IEEE Transactions on Systems, Man, and Cybernetics*, 1974, SMC-4(3): 310-313. [doi: 10.1109/TSMC.1974.5409141]
- 5 Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- 6 张翥. 基于模糊C均值聚类的医学MR图像分割研究[硕士学位论文]. 西安: 陕西师范大学, 2014.
- 7 虞文俊, 顾国华. 一种基于模糊C均值聚类 and 边缘提取算法的红外偏振图像的模式识别方法. *光子学报*, 2013, 42(10): 1244-1247.
- 8 严远驰. 基于FCM-C4.5组合过滤的入侵检测模型研究[硕士学位论文]. 广州: 广东工业大学, 2015.
- 9 Atanassov KT. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 1986, 20(1): 87-96. [doi: 10.1016/S0165-0114(86)80034-3]
- 10 江红莉, 何建敏, 庄亚明, 等. 基于直觉模糊集和证据理论的群决策方法. *控制与决策*, 2012, 27(5): 752-756.
- 11 雷英杰. *直觉模糊集理论及应用* (上册). 北京: 科学出版社, 2014.
- 12 雷英杰. *直觉模糊集理论及应用* (下册). 北京: 科学出版社, 2014.
- 13 贺正洪, 雷英杰. 直觉模糊C-均值聚类算法研究. *控制与决策*, 2011, 26(6): 847-850.
- 14 陈晓明, 姚泽清. 基于集值统计的直觉模糊聚类. *模糊系统与数学*, 2011, 25(3): 100-108.
- 15 Wang Z, Xu ZS, Liu SS, et al. A netting clustering analysis method under intuitionistic fuzzy environment. *Applied Soft Computing*, 2011, 11(8): 5558-5564. [doi: 10.1016/j.asoc.2011.05.004]
- 16 李鹏, 刘思峰, 朱建军. 基于新直觉模糊相似度的聚类方法. *控制与决策*, 2013, 28(5): 758-762.
- 17 余晓东, 雷英杰, 岳韶华, 等. 基于粒子群优化的直觉模糊核聚类算法研究. *通信学报*, 2015, 36(5): 74-80.
- 18 周广田, 曲科军, 徐为. 基于相对熵的直觉模糊聚类方法. *信息技术*, 2014, (1): 56-58, 61.
- 19 廖建平, 王卫民. 基于新直觉模糊相似度的直觉模糊谱聚类算法. *科技通报*, 2015, (4): 222-226.
- 20 Szmidi E, Kacprzyk J. Entropy for intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 2001, 110(3): 467-477.
- 21 姚伟烈, 林琳. 基于直觉模糊熵-解析面积的动态多属性决策方法. *模糊系统与数学*, 2015, 29(4): 103-108.
- 22 Zeng WY, Feng QL. On properties of intuitionistic fuzzy sets operations. *Fuzzy Systems and Mathematics*, 2008, 22(3): 97-104.