

文章编号:1671-6833(2012)02-0028-03

确定非线性隶属度函数的单次聚类方法

王杰¹, 刘琪², 党勤华¹

(1. 郑州大学 电气工程学院, 河南 郑州 450001; 2. 周口师范学院 物理与电子工程系, 河南 周口 466001)

摘要: 针对隶属度函数及其确定方法研究中存在的问题, 通过采用高斯函数和 sigmoid 函数表示隶属度函数, 提出了基于密度指标的单次聚类方法, 它的计算量仅与输入数据数目成线性关系, 且能快速得到聚类中心和数据的隶属度; 在实例仿真中, 确定了鄂北石油钻井现场数据 EB_Da51 的隶属度函数, 解决了隶属度函数难以确定的难题。

关键词: 隶属度函数; 单次聚类; 曲线拟合

中图分类号: TP181

文献标志码: A

doi:10.3969/j.issn.1671-6833.2012.02.007

0 引言

具有不确定性的现象与事物, 普遍存在于自然界和人类社会中. 如何表示和处理不确定性现象与事物, 一直是自然科学研究的热点和重点, 也是一个瓶颈性问题^[1]. Zadeh 博士^[2]1965 年发表了模糊集合的开创性成果, 提出了利用隶属度来表示事物的亦此亦彼性, 隶属度函数的形式主要包括分段线性函数和非线性函数^[3]两种.

隶属度函数的确定方法^[4]一直没有得到根本解决. 实际应用中, 隶属度的确定一般采用主观取值, 是不可靠的, 若采用统计的方法, 往往成本很高, 有时根本不可能做到. 笔者提出了基于密度指标的单次聚类方法, 通过曲线拟合, 确定了非线性隶属度函数, 方法简单, 易于推广.

1 非线性隶属度函数

非线性隶属度函数主要有高斯型、sigmoid 型、S 型、抛物线型等. 其中高斯型隶属度函数和 sigmoid 型隶属度函数具有很好的光滑性, 图形没有零点而且具有比较清晰的物理意义, 是描述不确定性最为常用的隶属度函数.

高斯型隶属度函数^[3]表达式如式(1), 其主要参数是函数中心 a 和函数曲线的宽度 σ , 如图 1 所示.

$$y = \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]. \quad (1)$$

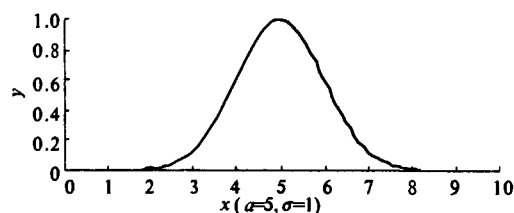


图1 高斯型隶属度函数

Fig.1 Gaussian membership function

sigmoid 型隶属度函数^[3]表达式如式(2), 适用于“很”、“很不”修饰的语言值的隶属度函数, 由 b, c 两个参数确定, 当 b 为正, sigmoid 型隶属度函数曲线向右开口; 当 b 为负, sigmoid 型隶属度函数曲线向左开口, 如图 2 所示.

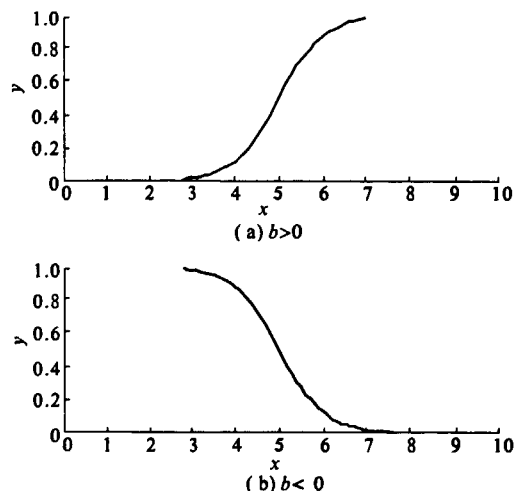


图2 sigmoid 型隶属度函数

Fig.2 Sigmoid membership function

收稿日期:2011-10-01; 修订日期:2011-12-12

基金项目:国家自然科学基金资助项目(60974005); 河南省自然科学基金计划项目(2010A120008).

作者简介:王杰(1959-), 男, 河南周口人, 郑州大学教授, 博士, 博士生导师, 主要研究方向为智能控制与智能计算, 信息与计算机网络安全, E-mail: wj@zzu.edu.cn.

$$\gamma = \frac{1}{1 + \exp[-b(x-c)]}. \quad (2)$$

2 单次聚类方法

模糊聚类分析是许多分类和系统建模的基础. 聚类的目的是从大量的数据中抽取固有的特征, 以获得系统行为的简洁表示. Bezdek 提出用隶属度确定数据点属于某个聚类程度的模糊 C 均值聚类(FCM)方法^[5]是目前比较成熟的模糊聚类方法^[6], 但算法结果严重依赖随机生成的初始聚类中心. 笔者提出一种基于密度指标的单次算法, 它的计算量与输入数据数目成线性关系.

对于 $X = \{x_1, x_2, \dots, x_n\} \subset R^S$ 的样本点集合, 单次聚类方法如下:

Step1: 设定邻域半径 r_a 、模糊指数 b 和聚类中心个数 m ;

Step2: 计算每一个点 x_i 的密度指标:

$$D_i = \sum_{j=1}^n \exp\left[-\frac{\|x_i - x_j\|^2}{(0.5r_a)^2}\right], \quad (3)$$

其中,

$$r_a = \frac{1}{2} \min_k \{\max_i \{\|x_i - x_k\|\}\} \quad (4)$$

是该点的领域半径, 选取最高的密度指标点 x_{c_1} 作为首个聚类中心;

Step3: 对每个数据点的密度指标进行修正:

$$D_i = D_i - D_{c_1} \exp\left[-\frac{\|x_i - x_{c_1}\|^2}{(0.5r_b)^2}\right]. \quad (5)$$

其中, r_b 是密度指标函数显著减小的邻域半径, 一般为 $r_b = 1.2r_a$, 取最高的密度指标点 $x_{c_{k+1}}$ 作为新的聚类中心;

Step4: 再修订每个数据点的密度指标, 直到得到 m 个聚类中心, 将聚类中心定义为 $w_k, k = 1, 2, \dots, m$;

Step5: 计算 x_i 对 w_k 的隶属度 u_{ik} :

$$u_{ik} = \frac{(1/\|x_i - w_k\|^2)^{1/b-1}}{\sum_{i=1}^n (1/\|x_i - w_k\|^2)^{1/b-1}}. \quad (6)$$

在上述方法中, 领域半径 r_a 是某点的密度贡献显著减小的度量, 一般采用式(4)确定, 实际分析中, 只要能得到 m 个聚类中心, 其值大小对聚类结果没有影响, 其值越大聚类中心出现越早, 但如果过大, 得到的聚类中心数目将会小于 m . 模糊指数 b 的确定比较复杂, 实际分析表明, 聚类结果对模糊指数 b 的变化并不敏感, 最佳选取区间为 $[1.5, 2.5]$, 一般情况下可取 $b = 2$.

3 实例仿真

单次聚类方法可以得到聚类中心和每个数据点对聚类中心的隶属度, 这些参数可以确定隶属度函数, 仿真实例选用鄂北石油钻井^[7]现场总体积数据 EB_Da51, 确定总体积变化量的隶属度函数, 总体积变化量是石油钻井井涌、井漏等事故的主要判别参数之一. EB_Da51 中的 200 个数据分布如图 3.

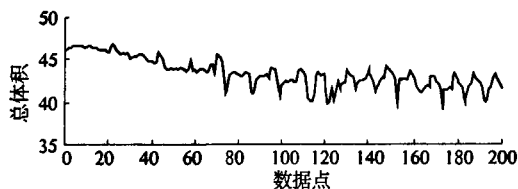


图3 石油钻井中 EB_Da51 数据

Fig.3 Data EB_Da51 of the petroleum drilling

单次聚类方法预设定的聚类中心个数 m 、聚类得到聚类中心和数据点对聚类中心的隶属度等参数确定隶属度函数的具体方法如下:

Step1: 数据统一论域, 并根据实际情况确定语言值数目, 笔者寻求石油钻井中总体积变化量的隶属度函数, 将 EB_Da51 数据映射到统一的论域空间, 取无、小、中、大 4 个语言值, 即 $m = 4$.

Step2: 单次聚类, 求解聚类中心 w_k 和数据对聚类中心的隶属度 u_{ik} , 如图 4 所示. 可以看出, 中间语言值基本服从正态分布, 可用高斯型隶属度函数表征, 两极语言值出现了边缘失真, 用 sigmoid 型隶属度函数曲线表征可以避免这种失真.

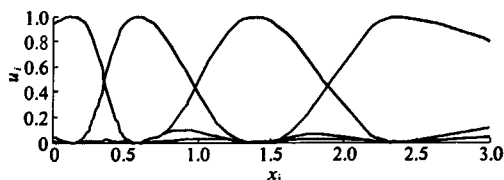


图4 数据变化量的隶属度

Fig.4 Membership degree of data variable quantity

Step3: 将中间语言值(本例中为“小”和“中”)的论域用高斯型隶属度函数曲线拟合, 得到其参数. 笔者采用 Matlab7 中曲线拟合工具箱(Curve Fitting)的 Trust-Region 方法进行拟合得到参数, 如表 1 所示.

Step4: 将两极语言值(本例中为“无”和“大”)的论域用 sigmoid 型隶属度函数曲线拟合, 得到其参数. 笔者采用 Matlab7 中曲线拟合工具箱(Curve Fitting)的 Trust-Region 方法进行拟合

得到参数如表 1 所示.

表 1 隶属度函数的参数

Tab. 1 Parameters of membership function

语言值	函数类型	函数模型	参数
无	sigmoid 型	$y = \frac{1}{1 + \exp[-b(x - c)]}$	$b = -19.930\ 0$ $c = 0.357\ 4$
小	高斯型	$y = \exp\left[-\frac{(x - a)^2}{2\sigma^2}\right]$	$a = 0.641\ 5$ $\sigma = 0.300\ 7$
中	高斯型	$y = \exp\left[-\frac{(x - a)^2}{2\sigma^2}\right]$	$a = 1.425\ 0$ $\sigma = 0.473\ 2$
大	sigmoid 型	$y = \frac{1}{1 + \exp[-b(x - c)]}$	$b = 9.211\ 0$ $c = 1.899\ 0$

Step5:根据曲线拟合得到的参数,可以确定隶属度函数如图 5 所示.

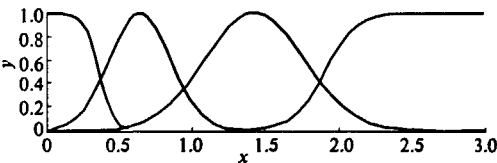


图 5 EB_Da51 的隶属度函数

Fig. 5 Membership function of data EB_Da51

4 结论

经实例研究表明,笔者提出的单次聚类方法不需迭代,避免了模糊均值聚类方法容易陷入局

部极小点的可能性,快速得到了聚类中心和数据的隶属度,它的计算量仅与输入数据数目成线性关系.采用单次聚类方法,通过曲线拟合确定了高斯型和 sigmoid 型隶属度函数,解决了非线性隶属度函数难以确定的难题.

参考文献:

[1] 李德毅. 知识表示中的不确定性[J]. 中国工程科学, 2002(10): 73 - 79.

[2] ZADEH L A. Fuzzy sets [J]. Information and Control, 1965(8): 338 - 353.

[3] 黄士涛,胡全义,马进元,等. 自适应神经模糊推理系统在起重机稳钩控制中的应用[J]. 郑州大学学报:工学版, 2007,28(1):21 - 25.

[4] 陆建江,张亚非,宋自林. 模糊关联规则的研究与应用[M]. 北京:科学出版社,2008:43 - 51.

[5] 刘琪,张鸿辉,黄贞贞. 基于模糊 C 均值聚类方法的石油钻井工作状态分析[J]. 周口师范学院学报, 2009(5):47 - 49.

[6] 黎灿兵,曲芳,王晓宁,等. 基于模糊聚类的电力系统负荷特性分析[J]. 郑州大学学报:工学版, 2010, 29(1):107 - 110.

[7] 王杰,陈东海,朱晓东,等. 基于分层模糊系统的石油钻井事故预警系统:中国,200710055178. 6[P]. 2007 - 06 - 01.

Nonlinear Membership Function Established by Single-Shot Clustering Method

WANG Jie¹, LIU Qi², DANG Qin-hua¹

(1. School of Electric Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. Department of Physics and Electronic Engineering, Zhoukou Normal University, Zhoukou 466001, China)

Abstract: The research of membership function and its determination method is of great significance. For determining the membership function expressed by Gaussian or sigmoid function, this paper develops a density index based on single-shot clustering method. The calculation amount depends on the number of input data linearly and the center and the membership degree of engineering data can be easily obtained. Through simulation, membership function of the data EB_Da51 from the petroleum drilling in Ebei is established, and the problem that membership function is hard to be described has been solved in paper.

Key words: membership function; single-shot clustering; curve fitting