:1008 1542(2018)02 0159 07　　doi:10 7535/hbkd 2018yx02010

,

(兰州交通大学数理学院,甘肃兰州　)

:为了解决以欧氏距离作为相似性准则的传统模糊聚类算法对多维数据处理不利的问题,采用马氏距离代替欧氏距离,对基于马氏距离的模糊聚类算法进行优化研究,以增强基于马氏距离的模糊聚类算法的聚类效果和能力。通过构造启发式搜索与 k 算法结合的初始优化方法,利用可以自动调节最佳聚类数的有效性函数,提出了一种优化算法 ,并将此新算法与 , , 聚类算法在 个标准数据集上进行了实验。结果表明, 算法有效,聚类精度比 , , 高,对高维数据聚类识别能力强,具有全局优化作用,并且聚类个数无需提前设定。新算法可为基于马氏距离的模糊聚类算法的优化提供参考。

:算法理论;模糊聚类;马氏距离;初始优化;聚类个数

: :

Abstract

Keywords

c

1

1 1 (FCM)

c- c-

$$J = \sum_{i=}^{c} \sum_{j=}^{n} u_{ij} \quad x_j - v_i \quad u_{ij}$$

$$J = \sum_{i=}^{c} \sum_{j=}^{n} u_{ij} \quad x_j - v_i \quad u_{ij}$$ m J

c-

$\boldsymbol{X}$ $\boldsymbol{x}$ $\boldsymbol{x}$ $\boldsymbol{x}_i$ $\boldsymbol{x}_n$ n $\boldsymbol{X}$

c $\boldsymbol{S}$ $\boldsymbol{S}$ $\boldsymbol{S}_i$ $\boldsymbol{S}_c$ $\boldsymbol{V} = \boldsymbol{v}$ $\boldsymbol{v}$ $\boldsymbol{v}_i$ $\boldsymbol{v}_c$ c u_{ij}

$\boldsymbol{x}_j$ $\boldsymbol{S}_i$

$$J^m \quad \boldsymbol{U} \ \boldsymbol{V} \ \boldsymbol{X} \quad = \sum_{i=}^{c} \sum_{j=}^{n} u_{ij}^m d_{ij} = \sum_{i=}^{c} \sum_{j=}^{n} u_{ij}^m \quad \boldsymbol{x}_j - \boldsymbol{v}_i$$

u_{ij}

$$\begin{cases} \sum_{i=}^{c} u_{ij} = & j \quad n \\ u_{ij} & i \quad c \quad i \quad j \quad n \end{cases}$$

$\boldsymbol{U} = u_{ij}$ $c \times n$ $\boldsymbol{V} = \boldsymbol{v}$ $\boldsymbol{v}$ $\boldsymbol{v}_i$ $\boldsymbol{v}_c$ $s \times c$ d_{ij} $\boldsymbol{x}_j$ $\boldsymbol{v}_j$

$m =$

1 c c n m m $+$ $\boldsymbol{V}$

ε $L =$ L

2 $$u_{ij} = \frac{}{\sum_{k=}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{\overline{m-}}}$$ $\boldsymbol{U}^{L+}$

3 $$\boldsymbol{V}_i = \frac{\sum_{j=}^{n} u_{ij} \ ^m \boldsymbol{X}_j}{\sum_{j=}^{n} u_{ij} \ ^m}$$ $\boldsymbol{V}^{L+}$ $L = L +$

ஞ

4 $\mathbf{V}^{L} - \mathbf{V}^{L-}$

ε L

c

1 2

k- k

p

k- k

k

k

c $\sqrt{n}$ n

O x x

x x A O

A a a O O a x

x A O A $\frac{n}{\sqrt{n}}$

O x' x' A O

m A A A_m m $\sqrt{n}$ n

$m = \sqrt{n}$

A A A_n a a a_n a a a_n k-

$\sqrt{n}$

c

2 2 KM-FCM

$\boldsymbol{\Sigma}_i^{-}$ |

$$J^m\ \boldsymbol{U}\ \boldsymbol{C}\ \boldsymbol{X}\quad \sum_{i=}^{c}\sum_{j=}^{n} u_{ij}^m\ \ \boldsymbol{x}_j - \boldsymbol{c}_i\ '\boldsymbol{\Sigma}_i^{-}\ \ \boldsymbol{x}_j - \boldsymbol{c}_i\ \ -\ \ |\boldsymbol{\Sigma}_i^{-}\ |$$

$$u_{ij}\quad \sum_{i=}^{c} u_{ij} = \quad \sum_{j=}^{n} u_{ij}\quad n\quad i = \quad c\quad j = \quad n$$

$$J\quad \sum_{i=}^{c}\sum_{j=}^{n} u_{ij}^m\ \ \boldsymbol{x}_j - \boldsymbol{c}_i\ '\boldsymbol{\Sigma}_i^{-}\ \ \boldsymbol{x}_j - \boldsymbol{c}_i\ \ -\ \ |\boldsymbol{\Sigma}_i^{-}\ | + \sum_{j=}^{n} \lambda_j\ \ - \sum_{i=}^{c} u_{ij}$$

J $\boldsymbol{c}_i$ u_{ij} $\boldsymbol{\Sigma}_i$

$$u_{ij} = \sum_{s=}^{c} \frac{\boldsymbol{x}_j - \boldsymbol{c}_i\ '\boldsymbol{\Sigma}_i^{-}\ \ \boldsymbol{x}_j - \boldsymbol{c}_i}{\boldsymbol{x}_j - \boldsymbol{c}_s\ '\boldsymbol{\Sigma}_i^{-}\ \ \boldsymbol{x}_j - \boldsymbol{c}_s}\ \frac{}{m-}\ -\quad i = \quad c\quad j = \quad n$$

$$\boldsymbol{c}_i = \sum_{j=}^{n} u_{ij}^m\ ^{-}\ \sum_{j=}^{n} u_{ij}^m \boldsymbol{x}_j\quad i = \quad c$$

$$\boldsymbol{\Sigma}_i = \frac{\sum_{j=}^{n} u_{ij}^m\ \ \boldsymbol{x}_j - \boldsymbol{c}_i\ \ \boldsymbol{x}_j - \boldsymbol{c}_i\ '}{\sum_{j=}^{n} u_{ij}^m}\quad i = \quad c$$

GD

c $\sqrt{}$ $m =$

L $\alpha = .$ L $L =$

c $\boldsymbol{C}$

$\boldsymbol{U}$

$\boldsymbol{C}$

$\boldsymbol{U}$ $\boldsymbol{C}$

$L = L +$

$$GD\ \boldsymbol{C} = \alpha CD\ \boldsymbol{C} + \ -\alpha\ \frac{}{SD\ \boldsymbol{C}}$$

$$CD\ \boldsymbol{C} = \frac{}{n} \sum_{i=}^{c} \sum_{j=}^{n} u_{ij}^{m} d_{ij} \quad i = \quad c \quad j = \quad n$$

$$SD\ \boldsymbol{C} = \frac{\sum_{i\ s=\ \ i\ \ s}^{c} d_{is}}{c\ \ c -} \quad i\ \ s = \quad c$$

GD

c

c c c L

GD

3

m L L

表 对于 数据集的 次实验 种聚类算法比较

表 对于 数据集的 次实验 种聚类算法比较

表 对于 数据集的 次实验 种聚类算法比较

4

k

/References:

C

C

C

K

K

C

C

C

C

k

C C