一种基于神经网络的广义熵模糊聚类算法

李 凯,曹 喆
(河北大学计算机科学与技术学院,河北保定,071000)

摘 要: 以模糊聚类为基础,将广义熵引入到模糊聚类的目标函数中,提出一种基于模糊熵的模糊聚类的统一形式,即广义熵模糊聚类模型;利用增广拉格朗日求解方法,以及 Hopfield 神经网络和复突触神经网络解决了基于广义熵的目标函数的优化问题,提出了基于神经网络的广义熵模糊聚类算法,表明了使用神经网络求解的收敛性;同时,给出一种用于确定增广拉格朗日乘子的迭代方法. 实验中选取人工生成数据集和 UCI 标准数据集对提出的算法进行了实验研究,并与常用的聚类算法进行了性能比较.

关键词: 模糊聚类; 广义熵; 增广拉格朗日方法; 神经网络
中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2016)08-1881-06
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2016.08.016

A Fuzzy Clustering Algorithm with Generalized Entropy Based on Neural Network

LI Kai, CAO Zhe
(*School of Computer Science and Technology, Hebei University, Baoding, Hebei* 071000, *China*)

Abstract: Based on fuzzy clustering, an unified form is presented for fuzzy clustering algorithm based on fuzzy entropy by introducing the generalized fuzzy entropy into objective function of fuzzy clustering, namely generalized entropy's fuzzy clustering model. Optimization problem for generalized entropy's objective function is solved using Hopfield neural network and multiple synapses based on augmented Lagrange method. After that, the generalized entropy's fuzzy clustering algorithm based on neural network is presented. And convergence of neural network is shown. At the same time, iterative method is given to determine Lagrange multipliers. In experiments, a synthetic data set and some standard UCI data sets are chosen to conduct some experimental studies. And clustering performance is compared with commonly used clustering algorithms.

Key words: fuzzy clustering; generalized entropy; augmented Lagrange method; neural network

1 引言

聚类作为一种重要的数据分析工具,已被广泛应用于实际问题中,例如模式识别,图像分割,市场研究,数据挖掘,遥感图像检测等[1,2]. 到目前为止,研究人员提出了很多不同的聚类算法,而模糊 c 均值聚类算法(Fuzzy C-Means, FCM)是较常用的一种方法,它的目标函数为 $J_m(\boldsymbol{U},\boldsymbol{V}) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \| \boldsymbol{x}_k - \boldsymbol{v}_i \|^2$,其中 u_{ik} 为第 k 个样本 $\boldsymbol{x}_k$ 隶属于第 i 个簇的程度,$\boldsymbol{U}$ 为隶属度构成的矩阵,$\boldsymbol{v}_i$ 为簇中心,V 为簇中心构成的矩阵,m 为加权指数,n 为数据样本的个数. 理论研究与实验表明,模糊聚类算法 FCM 存在许多缺陷,为此,人们通过修改目标函数或约束条件,提出了一些改进算法[3~6]. 为了便于表述,在下面的内容中,使用了 $J_1(\boldsymbol{U},\boldsymbol{V})$ 和 $J_2(\boldsymbol{U},\boldsymbol{V})$,它们分别代表 $J_m(\boldsymbol{U},\boldsymbol{V})$ 中 $m=1$ 和 2 时的目标函数,也就是 $J_m(\boldsymbol{U},\boldsymbol{V})$ 的特例.

为了克服 FCM 对噪声的影响,一些学者将模糊熵引入到目标函数 $J_1(\boldsymbol{U},\boldsymbol{V})$ 中,提出了基于熵的聚类算法[7~11];另外,一些学者基于目标函数 $J_2(\boldsymbol{U},\boldsymbol{V})$,通过加入调整项,提出了竞争凝聚算法 CA,以此解决聚类数的自动确定问题[12],之后,学者们在 CA 聚类算法的基础上,将约束条件或样本间的度量引入到 $J_2(\boldsymbol{U},\boldsymbol{V})$ 中,

收稿日期:2015-12-08;修回日期:2016-03-18;责任编辑:蓝红杰
基金项目:国家自然科学基金(No. 61375075)

提出了半监督聚类算法[13-14];最近,Maraziotis[15]通过对约束进行量化,并将其引入到 $J_2(\boldsymbol{U},\boldsymbol{V})$,提出了一种半监督模糊聚类算法. 可以看到,以上介绍的聚类算法,学者们主要采用 $J_m(\boldsymbol{U},\boldsymbol{V})$ 的特殊形式 $J_1(\boldsymbol{U},\boldsymbol{V})$ 或 $J_2(\boldsymbol{U},\boldsymbol{V})$,通过引入调整项,并借助拉格朗日方法,获得相应的模糊聚类算法. 为了研究一般情形下的聚类算法,Pedrycz 等[16]将监督信息引入到 $J_m(\boldsymbol{U},\boldsymbol{V})$ 中,给出了一般形式的目标函数,然而,他们并未对此种优化问题给出求解方法,只是使用了特殊的目标函数 $J_2(\boldsymbol{U},\boldsymbol{V})$;另外,Wei 等[17]试图使用神经网络解决一般目标函数的模糊聚类,遗憾的是该方法却存在一些问题[18]. 针对此种情况,本文研究了一般情形下的目标函数构成的优化问题的模糊聚类.

2 广义熵模糊聚类的目标函数

给定数据集 $X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$,其中 $\boldsymbol{x}_i\in R^N$,定义 c 是一个大于 1 的整数,u_{ik} 为第 k 个样本隶属于第 i 个簇的程度,$\boldsymbol{U}$ 为隶属度矩阵,则广义熵被定义为 $H(\boldsymbol{U},\alpha)=\sum_{k=1}^{n}(2^{1-\alpha}-1)^{-1}(\sum_{i=1}^{c}u_{ik}^{\alpha}-1)$,其中 α 称为广义熵指数,满足 $\alpha>0,\alpha\neq1$ 条件. 由定义并根据洛比大法则可知,当 $\alpha\to1$ 时,$H(\boldsymbol{U},\alpha)$ 变为模糊熵,即 $H(\boldsymbol{U},1)=-\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\log_2 u_{ik}$. 在下面的内容中,将 $H(\boldsymbol{U},\alpha)$ 引入到 $J_m(\boldsymbol{U},\boldsymbol{V})$ 中,从而获得了广义熵模糊聚类的目标函数 $J_G(\boldsymbol{U},\boldsymbol{V})$,即

$$\begin{aligned}J_G(\boldsymbol{U},\boldsymbol{V})&=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}\|\boldsymbol{x}_k-\boldsymbol{v}_i\|^2\\&+\delta\sum_{k=1}^{n}(2^{1-\alpha}-1)^{-1}(\sum_{i=1}^{c}u_{ik}^{\alpha}-1)\end{aligned}\tag{1}$$

其中 $\alpha>0$ 且 $\alpha\neq1,\delta$ 为参数. 因此,广义熵模糊聚类的优化问题为

$$\min_{U,V}J_G(\boldsymbol{U},\boldsymbol{V}),\text{s.t}\sum_{i=1}^{c}u_{ik}=1\tag{2}$$

由式(1)知,当 $m=1$ 且 $\alpha\to1$ 时,式(1)成为 Karayiannis[9]等提出的聚类算法. 当 $\alpha\to1$ 时,式(1)为 Wei 等[17]提出的聚类算法,因此,优化问题式(2)可以视为它们的统一形式.

另外,由式(2)中的目标函数可以知道,当 $m=\alpha$ 时,该优化问题可以通过拉格朗日方法求解,对于此种情况,我们已经进行了研究[19];然而,当 $m\neq\alpha$ 时,使用传统拉格朗日求极值方法对式(2)求解是不可行的. 实际上,对于优化问题式(2),可以使用启发式方法求解,例如,禁忌搜索、变邻域搜索等. 本文主要使用神经网络方法对其进行求解.

3 广义熵模糊聚类的神经网络方法

为了求解优化问题式(2),本文使用增广拉格朗日方法,Hopfield 神经网络与复突触神经网络求解.

3.1 使用 Hopfield 网络求解聚类中心

对于优化问题式(2),其增广拉格朗日函数为

$$\begin{aligned}L(\boldsymbol{U},\boldsymbol{V};\boldsymbol{\lambda})&=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}\|\boldsymbol{x}_k-\boldsymbol{v}_i\|^2\\&+\delta\sum_{k=1}^{n}(2^{1-\alpha}-1)^{-1}(\sum_{i=1}^{c}u_{ik}^{\alpha}-1)\\&+\sum_{k=1}^{n}\lambda_k(\sum_{i=1}^{c}u_{ik}-1)+\sum_{k=1}^{n}\gamma(\sum_{i=1}^{c}u_{ik}-1)^2\end{aligned}\tag{3}$$

其中 $\lambda_k(k=1,2,\cdots,n)$ 为拉格朗日乘子,$\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_n)$,$\gamma$ 是一个充分大的数. 由式(3)知道,它是一个关于聚类中心 $\boldsymbol{v}_i=(\boldsymbol{v}_{i,1},\boldsymbol{v}_{i,2},\cdots,\boldsymbol{v}_{i,N})(i=1,2,\cdots,c)$ 的二次函数,通过将二维下标转化为一维下标,且使用具有 $q=c\times N$ 个神经元的 Hopfield 神经网络求解,下面给出了求解聚类中心的 Hopfield 神经网络的权值,外部输入和激励函数.

$$\boldsymbol{NET}=\boldsymbol{W}\cdot\boldsymbol{V}+\boldsymbol{I}$$

其中 $\boldsymbol{NET}=(net_1,net_2,\cdots,net_q)^{\mathrm{T}}$,

$\boldsymbol{W}=(w_{ij})_{q\times q},\boldsymbol{V}=(v_1,v_2,\cdots,v_q)^{\mathrm{T}},\boldsymbol{I}=(i_1,i_2,\cdots,i_q)^{\mathrm{T}}$,

$$net_j=\sum_{i=1}^{q}w_{ji}v_i+i_j,\quad j=1,2,\cdots,q\tag{4}$$

$$i_{(i-1)\times N+l}=2\sum_{k=1}^{n}u_{ik}^{m}x_{kl},\quad i=1,2,\cdots,c;l=1,2,\cdots,N\tag{5}$$

$$w_{ji}=\begin{cases}-2\sum_{k=1}^{n}u_{\lceil\frac{i}{N}\rceil k}^{m},&\text{if } i=j\\0,&\text{otherwise}\end{cases}\tag{6}$$

$j=1,2,\cdots,q;i=1,2,\cdots,q$,

$$\begin{aligned}v_j^{(g+1)}&=f(net_j^{(g)})\\&=\begin{cases}v_j^{(g)}+\delta_j,&\text{if}\quad net_j^{(g)}\geqslant0,\\v_j^{(g)}-\delta_j,&\text{if}\quad net_j^{(g)}<0,\end{cases}\quad j=1,2,\cdots,q\end{aligned}\tag{7}$$

其中上角标的 g 为迭代次数,δ_j 是一个较小的可调正数,详细推导可参见文献[17].

3.2 使用复突触神经网络求解模糊隶属度

令 $d_{ik}=\|\boldsymbol{x}_k-\boldsymbol{v}_i\|^2$,则式(3)变为

$$\begin{aligned}L(\boldsymbol{U},\boldsymbol{V};\boldsymbol{\lambda})&=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}d_{ik}+\delta\sum_{k=1}^{n}(2^{1-\alpha}-1)^{-1}(\sum_{i=1}^{c}u_{ik}^{\alpha}-1)\\&+\sum_{k=1}^{n}\lambda_k(\sum_{i=1}^{c}u_{ik}-1)+\sum_{k=1}^{n}\gamma(\sum_{i=1}^{c}u_{ik}-1)^2\end{aligned}\tag{8}$$

将式(8)中的二维下标(i,k)转换为一维下标$(k-1)\times$

$c+i$ 并展开后得到

$$\begin{aligned}\sum_{k=1}^{n}[&(u_{(k-1)\times c+1})^m d_{(k-1)\times c+1}+(u_{(k-1)\times c+2})^m d_{(k-1)\times c+2}+\cdots\\&+(u_{(k-1)\times c+c})^m d_{(k-1)\times c+c}\\&+\gamma(u_{(k-1)\times c+1}+u_{(k-1)\times c+2}+\cdots+u_{(k-1)\times c+c})^2\\&+\delta(2^{1-\alpha}-1)^{-1}((u_{(k-1)\times c+1})^\alpha+(u_{(k-1)\times c+2})^\alpha+\cdots\\&+(u_{(k-1)\times c+c})^\alpha)\\&+(\lambda_k-2\gamma)(u_{(k-1)\times c+1}+u_{(k-1)\times c+2}+\cdots+u_{(k-1)\times c+c})\\&+\gamma-\lambda_k-\delta(2^{1-\alpha}-1)^{-1}]\end{aligned}\tag{9}$$

由 m 和 α 的取值可知,函数式(9)可能为一个高次的函数,鉴于此种情况,本文使用复突触神经网络对其进行优化. 根据聚类中心 $\boldsymbol{v}_i(i=1,2,\cdots,c)$ 在第二层循环中其值不变以及最后一行为常数,因此,在下面的优化中对这些项进行删除.

复突触神经网络[17]是一种在两个神经元之间具有多个突触且神经元具有多个激活输出的一种网络,其网络结构如图 1 所示. 可以看到,两个神经元间具有三个不同的权值 w_{ji}、z_{ji} 和 y_{ji},它们都是从第 i 个神经元的输出连接到第 j 个神经元的输入,并且三个权值是相互独立的,该神经网络的工作方式可以表示如下:

$$u_j^{(g+1)}=f(net_j^{(g)})=f(\sum_{i=1}^{s}(w_{ji}u_i^{(g)}+z_{ji}u_i^{(g)}+y_{ji}u_i^{(g)})+i_j),\quad j=1,2,\cdots,s\tag{10}$$

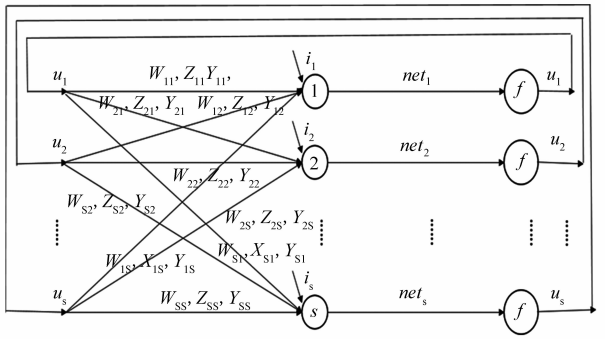


图1　复突触神经网络结构

网络输入的矩阵形式可表示为:

$$\boldsymbol{NET}=\boldsymbol{W}\cdot\boldsymbol{U}+\boldsymbol{Z}\cdot\boldsymbol{U}+\boldsymbol{Y}\cdot\boldsymbol{U}+\boldsymbol{I}\tag{11}$$

其中 $\boldsymbol{Z}=(z_{ij})_{s\times s}$,$\boldsymbol{Y}=(y_{ij})_{s\times s}$,$\boldsymbol{U}=(u_1,u_2,\cdots,u_s)^{\mathrm{T}}$,$\boldsymbol{I}=(i_1,i_2,\cdots,i_s)^{\mathrm{T}}$,$s=n\times c$.

定义矩阵 $\boldsymbol{U}_{<m-1>}$ 和 $\boldsymbol{Y}_{<\alpha-1>}$ 如下:

$\boldsymbol{U}_{<m-1>}=(u_1^{m-1},u_2^{m-1},\cdots,u_s^{m-1})^{\mathrm{T}}$,$m>1$ 且 $\boldsymbol{U}_{<1>}=\boldsymbol{U}$,$\boldsymbol{Y}_{<\alpha-1>}=(y_1^{\alpha-1},y_2^{\alpha-1},\cdots,y_s^{\alpha-1})^{\mathrm{T}}$,$\alpha>1$ 且 $\boldsymbol{Y}_{<1>}=\boldsymbol{Y}$.

为了获得求解隶属度的神经网络,利用式(11)得到如下的能量函数:

$$\begin{aligned}E=&-(\frac{1}{m})\boldsymbol{U}_{<m-1>}^{\mathrm{T}}\cdot\boldsymbol{W}\cdot\boldsymbol{U}-(\frac{1}{2})\boldsymbol{U}^{\mathrm{T}}\cdot\boldsymbol{Z}\cdot\boldsymbol{U}\\&-(\frac{1}{\alpha})\boldsymbol{Y}_{<\alpha-1>}^{\mathrm{T}}\cdot\boldsymbol{W}\cdot\boldsymbol{Y}-\boldsymbol{U}^{\mathrm{T}}\cdot\boldsymbol{I}\end{aligned}\tag{12}$$

比较式(9)和式(12),获得了如下的对应关系:

$$w_{ji}=\begin{cases}-md_i, & i=j\\0, & i\neq j\end{cases},i,j=1,2,\cdots,s\tag{13}$$

$$z_{ji}=\begin{cases}-2\gamma, & (\lceil\frac{i}{c}\rceil-1)\cdot c<j\leqslant\lceil\frac{i}{c}\rceil\cdot c,i,j=1,2,\cdots,s\\0, & \text{otherwise}\end{cases}\tag{14}$$

$$y_{ji}=\begin{cases}-\alpha\cdot\delta\cdot(2^{1-\alpha}-1)^{-1}, & i=j\\0, & i\neq j\end{cases}\quad i,j=1,2,\cdots,s\tag{15}$$

$$i_j=2\gamma-\lambda_k\ j=1,2,\cdots,s;k=\lceil j/c\rceil\tag{16}$$

其中$\lceil x\rceil$为不小于 x 的最小整数. 利用对称矩阵和向量具有 $L^{\mathrm{T}}\boldsymbol{AR}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{AL}$ 关系(其中 $\boldsymbol{A}$ 为一个对称矩阵,$\boldsymbol{L}$ 和 $\boldsymbol{R}$ 分别为列向量),则式(12)变为

$$\begin{aligned}E=&-(\frac{1}{m})\boldsymbol{U}^{\mathrm{T}}\cdot\boldsymbol{W}\cdot\boldsymbol{U}_{<m-1>}-(\frac{1}{2})\boldsymbol{U}^{\mathrm{T}}\cdot\boldsymbol{Z}\cdot\boldsymbol{U}\\&-(\frac{1}{\alpha})\boldsymbol{U}^{\mathrm{T}}\cdot\boldsymbol{Y}\cdot\boldsymbol{U}_{<\alpha-1>}-\boldsymbol{U}^{\mathrm{T}}\cdot\boldsymbol{I},\end{aligned}\tag{17}$$

对于此式,可以理解为将 $-(1/m)\boldsymbol{U}^{\mathrm{T}}$、$-(1/2)\boldsymbol{U}^{\mathrm{T}}$、$-(1/\alpha)\boldsymbol{U}^{\mathrm{T}}$ 和 $-\boldsymbol{U}^{\mathrm{T}}$ 分别乘以式(18)中右边第一项、第二项、第三项和第四项,这就意味着将 $\boldsymbol{U}_{<m-1>}$ 反馈给权值 W,将新的 U 反馈给权值 Z,将 $\boldsymbol{U}_{<\alpha-1>}$ 反馈给权值 Y,从而得到一种具体的复突触神经网络,其中网络输入的权值矩阵为:

$$\boldsymbol{NET}=\boldsymbol{W}\cdot\boldsymbol{U}_{<m-1>}+\boldsymbol{Z}\cdot\boldsymbol{U}+\boldsymbol{Y}\cdot\boldsymbol{U}_{<\alpha-1>}+\boldsymbol{I}\tag{18}$$

激励函数 f 按照如下方法确定,其中 ω_j 是一个较小的可调整正数,

$$u_j^{(g+1)}=f(net_j^{(g)})=\begin{cases}u_j^{(g)}+\omega_j, & \text{if } net_j^{(g)}\geqslant 0\\u_j^{(g)}-\omega_j, & \text{if } net_j^{(g)}<0\end{cases},\quad j=1,2,\cdots,s.\tag{19}$$

另外,利用式(18),(19)以及 $\frac{\partial E}{\partial u_j}=-net_j$,且根据 $\Delta E=(\nabla\boldsymbol{E})^{\mathrm{T}}\Delta\boldsymbol{u}=-(\boldsymbol{NET})^{\mathrm{T}}\Delta\boldsymbol{u}$,可以得到复突触神经网络的收敛性,即不论 $net_j^{(g)}\geqslant 0$ 还是 $net_j^{(g)}<0$,总有 $\Delta\boldsymbol{E}<0$,其中 $\Delta\boldsymbol{E}$ 为能量变化量,$\Delta\boldsymbol{u}$ 为隶属度的变化向量,$\nabla\boldsymbol{E}$ 为梯度向量.

3.3　增广拉格朗日乘子的确定

为了确定式(16)中的拉格朗日乘子,考虑如下的优化问题:

$$\begin{aligned}\min_{U,V}J_G(\boldsymbol{U},\boldsymbol{V})=&\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^m d_{ik}\\&+\delta\sum_{k=1}^{n}(2^{1-\alpha}-1)^{-1}(\sum_{i=1}^{c}u_{ik}^\alpha-1),\end{aligned}\tag{20}$$

$$\text{s.t}\ \varphi_k(U)=\sum_{i=1}^{c}u_{ik}-1=0,k=1,2,\cdots,n$$

其增广拉格朗日函数为:

$$L(\boldsymbol{U},\boldsymbol{V};\boldsymbol{\lambda}) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{m} d_{ik} + \delta\sum_{k=1}^{n}(2^{1-\alpha}-1)^{-1}\Big(\sum_{i=1}^{c} u_{ik}^{\alpha} - 1\Big) + \sum_{k=1}^{n}\lambda_k\varphi_k(\boldsymbol{U}) + \sum_{k=1}^{n}\gamma\varphi_k^2(\boldsymbol{U}) \tag{21}$$

关于式(20)和(21)有如下结论[20]:如果 $\boldsymbol{U}^*$,$\boldsymbol{V}^*$ 是问题式(20)的最优解,λ^* 为由拉格朗日乘子构成的向量,则当 γ 充分大时,$\boldsymbol{U}^*$,$\boldsymbol{V}^*$ 是式(21)的极小值点,且式(20)和(21)等价. 为此,针对式(21),先给定 λ^* 的一个初始值 $\lambda^{(p)}$,并用该值求解优化问题式(21),设其解为 $\boldsymbol{U}^{(p+1)}$,这样可以得到如下等式:

$$\begin{aligned}
&\nabla_U L(\boldsymbol{U}^{(p+1)},\boldsymbol{V};\lambda^{(p)})\\
&= \nabla J_G(\boldsymbol{U}^{(p+1)},\boldsymbol{V}) + \sum_{k=1}^{n}\lambda_k^{(p)}\nabla\varphi_k(\boldsymbol{U}^{(p+1)})\\
&\quad + 2\gamma\sum_{k=1}^{n}\varphi_k(\boldsymbol{U}^{(p+1)})\nabla\varphi_k(\boldsymbol{U}^{(p+1)})\\
&= \nabla J_G(\boldsymbol{U}^{(p+1)},V) + \sum_{k=1}^{n}(\lambda_k^{(p)}\\
&\quad + 2\gamma\varphi_k(\boldsymbol{U}^{(p+1)}))\nabla\varphi_k(\boldsymbol{U}^{(p+1)})\\
&= 0
\end{aligned}$$

可以看到,如果 $\boldsymbol{U}^*$ 是式(20)的最优解,则有

$$\nabla J_G(\boldsymbol{U}^*,\boldsymbol{V}) + \sum_{k=1}^{n}\lambda_k^*\nabla\varphi_k(\boldsymbol{U}^*) = 0,$$

从而有

$$\lambda_k^{(p+1)} = \lambda_k^{(p)} + 2\gamma\varphi_k(\boldsymbol{U}^{(p+1)}). \tag{22}$$

3.4 广义熵模糊聚类算法 GEFCM

通过上面的推导,下面给出具体的算法,并将该算法记为 GEFCM(Generalized Entropy FCM):

算法 1 GEFCM(Generalized Entropy FCM)算法

step1 初始化参数,分别给 c、m、γ、δ_{vj}、δ_{uj}、Δv、Δu、ε 一个定值.

step2 随机初始化聚类中心 $\boldsymbol{v}_j$ 和隶属度 u_j,其中 $\boldsymbol{v}_j$ 是从数据点中随机选择,u_j 是从 0 到 1 的范围内随机选择;随机初始化 $\boldsymbol{\lambda}^{(0)}$.

step3 设定初始值 $net_j^{(0)}=0$,以及 Hopfield 网络中的迭代变量 $g=1$.

step4 利用式(4)、(5)和(6)分别计算 ***NET***、矩阵 ***I*** 和权值矩阵 ***W***.

step5 for $j=1$ to q

{

if $net_j^{(g)} \cdot net_j^{(g-1)} <= 0$ then $\delta_{vj}=\delta_{vj}/2$.

if $net_j^{(g)} >= 0$ then $\boldsymbol{v}_j=\boldsymbol{v}_j+\delta_{vj}$ else $\boldsymbol{v}_j=\boldsymbol{v}_j-\delta_{vj}$.

}

if $((\delta_{v1}<=\Delta v)\&(\delta_{v2}<=\Delta v)\&\cdots\&(\delta_{\Delta q}<=\Delta v))$ then go step6 else {$g=g+1$, go step4.}

step6 初始化复突触神经网络的迭代变量 $t=1$.

step7 利用式(13),(14),(15),(16)和(18)分别计算权值矩阵 ***W***、***Z***、***Y***、***I*** 和 ***NET***.

step8 for $j=1$ to s

{

if $net_j^{(t)} \cdot net_j^{(t-1)} <= 0$ then $\delta_{uj}=\delta_{uj}/2$.

if $net_j^{(t)} >= 0$ then $u_j=u_j+\delta_{uj}$ else $u_j=u_j-\delta_{uj}$.

}

if $((\delta_{u1}<=\Delta u)\&(\delta_{u2}<=\Delta u)\&\cdots\&(\delta_{us}<=\Delta u))$ then go step9 else {$g=g+1$,利用式(22)更新 λ, go step7.}

step9 if $\|\boldsymbol{U}^{(t)}-\boldsymbol{U}^{(t-1)}\|<=\varepsilon$ then stop else go step3.

4 实验结果与分析

为了表明提出的算法 GEFCM 的有效性,实验中选取了 7 个数据集进行了研究,其中 Arti3 是人工生成的三维线性可分数据集,具有 3 个类别和 150 个数据样本,且每类中的数据点个数均为 50. 另外 6 个数据集来自于 UCI,分别为 Iris,Breast-w,Wine,Heart,Ionosphere 和 Australian. 实验中选择的参数值分别为 $\gamma=1000$,$\delta_v=0.5$,$\delta_u=0.2$,$\varepsilon=0.001$,$\Delta v=0.0001$ 和 $\Delta u=0.0001$. 为了评价聚类的性能,实验中选择了聚类正确率作为评价指标,即正确聚类的个数除以样本总数. 实验结果如表 1 至表 7 所示.

表 1 Arti3 数据集的实验结果

加权指数 m	广义熵指数 α	广义熵系数 δ	正确率(%)
1.1	1.1	-1000	100
1.1	100	-10000	100
2	100	-10000	100
3	15	-150	100
8	8	-10	99.33
12	100	-100	72.67
15	15	-1	65.33

表 2 Iris 数据集的实验结果

加权指数 m	广义熵指数 α	广义熵系数 δ	正确率(%)
1.1	1.1	10	92.7
2	100	-10000	96
5	25	-1000	96
8	8	-10	96
8.2	15	-200	96.67
10	5	-100	66
12	10	-20	72
15	2	-2	40

表 3 Breast-w 数据集实验结果

加权指数 m	广义熵系数 α	广义熵系数 δ	正确率(%)
1.1	1.1	10000	97.4
2	2	1	96.78
3	10	-20	95.17
5	10	-1000	95.46
8	15	-200	96.63
10	100	-100	96.93
15	20	-1000	96.93
20	10	2	43.78

表 4　Wine 数据集实验结果

加权指数 m	广义熵指数 α	广义熵系数 δ	正确率(%)
1.1	1.1	5	70.2
2	2	-10	85.96
3	7	-25	71.35
5	10	-20	74.16
10	100	-1000	73.03
12	150	-10000	74.16
20	20	2	41.57

表 5　Heart 数据集实验结果

加权指数 m	广义熵指数 α	广义熵系数 δ	正确率(%)
1.1	1.1	10000	60.7
2	2	-100	58.89
5	10	-1000	71.11
8	20	-5000	75.93
10	50	-2000	75.19
12	40	-2000	72.96
15	100	-10000	74.07

表 6　Ionosphere 数据集实验结果

加权指数 m	广义熵指数 α	广义熵系数 δ	正确率(%)
1.1	200	-20000	71.23
2	2	-20	56.13
5	15	-200	68.66
8	100	-20000	62.39
10	50	-2000	69.8
12	50	-10000	57.27

表 7　Australian 数据集实验结果

加权指数 m	广义熵指数 α	广义熵系数 δ	正确率(%)
1.1	1.1	2	56.1
2	3	-50	78.26
5	10	-100	79.42
8	20	-1000	75.65
10	20	-2000	79.86
12	60	-12000	63.41

由实验结果可以看到,对于可分性数据集 Arti3,当加权指数和广义熵指数相等且接近于 1 时,获得了较好的聚类结果;然而,对于选择的 UCI 标准数据集,当加权指数和广义熵指数相等且接近于 1 或加权指数等于 2 时,其聚类结果并不一定是最好的,例如,对于 Wine 数据集,当 $m=2$ 且 $\alpha=2$ 时,获得了最好的聚类正确率 85.96,然而,当 $\alpha=2$ 时,此时的广义熵并不是模糊熵,表明了基于模糊熵的聚类未必获得更好的聚类结果,从而充分说明引入广义熵模糊聚类的重要性.

另外,为了验证提出的算法 GEFCM 的性能,实验中也选取了 FCM、PPSO-FCM[21] 以及 SSWFCM 算法[22] 在标准数据集进行了比较,实验结果如图 2 所示.

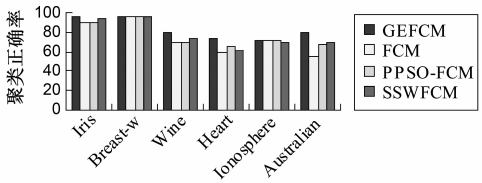


图2　不同算法在标准数据集上的实验结果比较

可以看到,基于广义熵的模糊聚类算法在选取的数据集上的聚类正确率都高于 FCM、PPSO-FCM 和 SSWFCM,且此时加权指数 m 的取值并不仅限制在 1 或 2 两个值,从而表明提出的广义熵目标函数聚类的合理性.

5　结论

本文通过引入广义熵,提出了广义熵模糊聚类模型,统一了基于熵的模糊聚类算法,使用 Hopfield 神经网络和复突触神经网络并结合增广拉格朗日乘子法解决了基于广义熵的模糊聚类问题,提出了基于神经网络的广义熵模糊聚类算法,给出了拉格朗日乘子的确定方法. 在实验中,选取人工数据集和 UCI 数据集对提出的算法性能进行了测试,并与其它算法进行了性能比较,表明了提出的算法对数据聚类的有效性.

参考文献

[1] Timothy C H, James C B, Christopher L, et al. Fuzzy c-Means algorithms for very large data[J]. IEEE Transactions on Fuzzy Systems, 2012, 20(6): 1130 - 1146.

[2] 慕彩红,霍利利,刘逸,刘若辰,焦李成. 基于小波融合和 PCA-核模糊聚类的遥感图像变化检测[J]. 电子学报, 2015, 43(7): 1375 - 1381.
Mu Caihong, Huo Lili, Liu Yi, Jiao Licheng. Change detection for remote sensing images based on wavelet fusion and PCA-kernel fuzzy clustering[J]. Acta Electronica Sinica, 2015, 43(7): 1375 - 1381. (in Chinese)

[3] Yang M S, Tsai H S. A gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction[J]. Pattern Recognition Letters, 2008, 29(12): 1713 - 1725.

[4] 刘兵,夏士雄,周勇,韩旭东. 基于样本加权的可能性模糊聚类算法[J]. 电子学报, 2012, 40(2): 371 - 375.
Liu Bing, Xia Shixiong, Zhou Yong, Han Xudong. A sample-weighted possibilistic fuzzy clustering algorithm[J]. Acta Electronica Sinica, 2012, 40(2): 371 - 375. (in Chinese)

[5] Jacek M L. Fuzzy c-ordered-means clustering[J]. Fuzzy Sets and Systems,2016,286,114 – 133.

[6] Ding Y,Fu X. Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm[J]. Neurocomputing,doi:10.1016/j.neucom.2015.01.106.

[7] Krishnapuram R,Keller,J M. A possibilistic approach to clustering[J]. IEEE Transactions on Fuzzy Systems,1993,1(2):98 – 110.

[8] Krishnapuram R,Keller J M. The possiblistic means algorithms:insights and recommendation[J]. IEEE Transactions on Fuzzy Systems,1996,4(3):385 – 393.

[9] Karayiannis N B. MECA:maximum entropy clustering algorithm[A]. IEEE world congress on computational intelligence,Proceedings of the third IEEE conference on fuzzy systems[C]. Orlando:IEEE,1994. 630 – 635.

[10] Ichihashi H,Miyagishi K,Honda K. Fuzzy c-means clustering with regularization by K-L information[A]. Proceedings of the 10th IEEE international conference on fuzzy systems[C]. Melbourne:IEEE,2001. 924 – 927.

[11] Yasuda M,Furuhashi T,Matsuzaki M,et al. A study on statistical mechanical characteristics of fuzzy clustering [A]. Proceedings of IEEE International Conference on Systems,Man,and Cybernetics[C]. Tucson:IEEE,2001. 2415 – 2420.

[12] Frigui H,Krishnapuram R. Clustering by competitive agglomeration[J]. Pattern Recognition,1997,30 (7):1109 – 1119.

[13] Grira N,Crucianu M,Boujemaa N. Active semi-supervised fuzzy clustering[J]. Pattern Recognition,2008,41 (5):1834 – 1844.

[14] Gao C F,Wu X J. A new semi-supervised clustering algorithm with pairwise constraints by competitive agglomeration[J]. Applied Soft Computing,2011,11 (8):5281 – 5291.

[15] Maraziotis I A. A semi-supervised fuzzy clustering algorithm applied to gene expression data[J]. Pattern Recognition,2012,45:637 – 648.

[16] Pedrycz W,Amato A,Lecce V D,et al. Fuzzy clustering with partial supervision in organization and classification of digital images[J]. IEEE Transactions on Fuzzy Systems,2008,16(4):1008 – 1026.

[17] Wei C,Fahn C. The multisynapse neural network and its application to fuzzy clustering[J]. IEEE Transactions on Neural Networks,2002,13(3):600 – 618.

[18] Yu J,Hao P W. Comments on "The multisynapse neural network and its application to fuzzy clustering"[J]. IEEE Transaction on Neural Networks,2005,16(3):777 – 778.

[19] Li K,Ma H Y,Wang Y. Unified model of fuzzy clustering algorithm based on entropy and its application to image segmentation[J]. Journal of Computational Information Systems,2011,7(15):5476 – 5483.

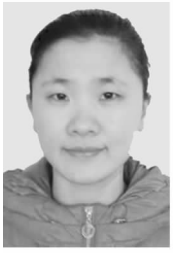
[20] 刘健,王晓明著. 鞍点规划与形位误差评定[M]. 大连:大连理工大学出版社,1996.
Liu Jian,Wang Xiaoming. Saddle Point Programming and Geometric Error Evaluation[M]. Dalian:Dalian University of Technology Press,1996.

[21] Liu H,Yih J M,Wu D B,Liu S W. Fuzzy c-mean clustering algorithms based on picard iteration and particle swarm optimization [A]. Proceedings of International Workshop on Education Technology and Training [C]. Washington:IEEE,2008. 838 – 842.

[22] Zhang X B,Huang H,Zhang S. A FCM clustering algorithm based on semi-supervised and point density weighted [A]. Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems[C]. Xiamen:IEEE,2010. 710 – 713.

作者简介

李 凯 男,1963 年出生,河北保定人. 2005 年毕业于北京交通大学计算机与信息技术学院,并获得工学博士学位. 主要从事机器学习、模式识别、数据挖掘等方面的研究.
E-mail:likai@hbu.edu.cn



曹 喆 女,1991 年出生,河北邢台人,硕士研究生. 主要从事机器学习和数据挖掘等方面的研究.