

堆叠隐空间模糊 C 均值聚类算法

王 骏, 刘 欢, 蒋亦樟, 邓赵红, 王士同

(江南大学 数字媒体学院, 江苏 无锡 214122)

摘 要: 基于极限学习机理论, 将主成分分析技术与 ELM 特征映射相结合, 提出一种基于主成分分析的压缩隐空间构建新方法. 结合多层神经网络学习方法对隐空间进行多层融合, 进一步提出了堆叠隐空间模糊 C 均值聚类算法, 从而提高对非线性数据的学习能力. 实验结果表明, 所提出算法在处理复杂非线性数据时更加高效、稳定, 同时克服了模糊聚类算法对模糊指数的敏感性问题.

关键词: 隐空间映射; 极限学习机; 主成分分析; 模糊 C 均值聚类; 多层神经网络

中图分类号: TP181

文献标志码: A

Cascaded hidden space fuzzy C means clustering algorithm

WANG Jun, LIU Huan, JIANG Yi-zhang, DENG Zhao-hong, WANG Shi-tong

(School of Digital Media, Jiangnan University, Wuxi 214122, China. Correspondent: LIU Huan, E-mail: 1023206142@qq.com)

Abstract: In view of the good properties of the extreme learning machine(ELM) feature mapping, a novel technique of constructing to condensed hidden feature space is proposed by combining principal component analysis(PCA) with ELM feature mapping. The cascaded hidden space fuzzy C means clustering algorithm is proposed to improve the learning ability of the non-linear data. Experimental results show that the proposed algorithm is not only efficient and robust for high-dimension data, but also insensitive to the fuzzy index of fuzzy clustering algorithms.

Keywords: hidden-mapping space; extreme learning machine; principal component analysis; fuzzy C means clustering; multi-layer neural network

0 引 言

模糊聚类分析是一种重要的聚类学习方法. 作为聚类分析的有效工具, 模糊聚类技术已得到研究人员的广泛研究^[1-2]. 如今, 人们通过引入新的学习模型、改造目标学习准则等手段提出了各种新型模糊聚类方法^[3-7]. 这些方法虽然从不同的角度研究模糊聚类技术, 但对于实际应用中的各种难题, 仍然存在着局限性.

非线性数据在自然界中广泛存在. 在机器学习研究中, 核方法^[8]是处理非线性数据的一个有效手段. 通过向传统的线性学习算法中引入核函数来替换原算法中的内积运算, 可以使线性学习算法变成非线性学习算法, 这种思想已被成功地运用到诸如分类、回归、聚类、特征提取等学习任务中. 在模糊聚类分析的研究工作中, 核模糊 C 均值 (FCM)^[9]的研究为基于

划分的模糊聚类算法处理非线性数据提供了有益的思路. 但是, 在核 FCM 中, 核函数必须满足 Mercer 条件^[10]; 在缺乏先验知识的情况下如何合理选择核函数以及相关参数是一个经典难题, 这影响了核方法在模糊聚类技术中的应用. 根据 Cover 的模式可分性定理, 一个复杂的模式分类问题映射到高维空间后, 会比在低维空间里更容易变得线性可分^[11]. 除了核方法外, 是否还存在着其他的映射方法来生成高维特征空间用以解决模糊聚类时非线性数据的无监督学习问题, 是本文的研究目标.

近年来, 以极速学习机 (ELM) 为代表的单隐层前馈神经网络快速学习理论得到了人们的深入研究^[12-15]. 在 ELM 学习理论中, 隐节点张成的特征空间构成了 ELM 隐空间^[16]. 研究表明, 它具有如下优势: 1) 隐节点参数可以随机生成, 这有利于 ELM 隐空间

收稿日期: 2015-06-16; 修回日期: 2016-01-18.

基金项目: 国家自然科学基金项目(61300151); 江苏省自然科学基金项目(BK20130155); 江苏省高校自然科学基金项目(13KJB520001).

作者简介: 王骏(1978—), 男, 副教授, 博士, 从事人工智能、数据挖掘等研究; 刘欢(1993—), 男, 硕士生, 从事人工智能、数据挖掘的研究.

的快速构建; 2) 构建隐空间所使用的激励函数只要满足连续、有界且不是常量函数这几个简单条件即可, 这有利于用户灵活方便地选取激励函数。

不失一般性, 本文以 FCM 为代表展开研究工作, 同时基于 ELM 隐空间学习理论, 提出堆叠隐空间模糊聚类方法. 首先研究基于主成分分析 (PCA)^[17] 的压缩隐空间构建方法. 在此基础上, 结合多隐层神经网络学习方法^[18], 将单隐层压缩隐空间学习结构改造成堆叠隐空间学习结构, 并提出相应的堆叠隐空间模糊 C 聚类算法 (CHS-FCM). 该方法通过使用层次化的学习结构对数据对象在不同层次上的表达形式进行抽象, 并通过综合使用压缩和扩充手段对低层概念进行重组以得到新的信息表达形式, 从而有效提高模糊聚类算法处理复杂数据的能力. 经实验验证, 该算法能有效地弥补经典 FCM 的若干不足, 同时在保证学习精度和学习效率的前提下, 有效克服对模糊指数的敏感性问题。

1 相关工作

1.1 ELM 隐空间

单隐层前馈神经网络 (SLFN) 在许多诸如模式识别、信号处理以及短期预测的领域中得到了广泛应用, 人们相继提出了许多 SLFN 的训练算法. 这些学习算法中所有的网络隐层节点参数都是通过迭代过程进行多次优化最终确定的, 但是这些迭代步骤往往会大大增加参数训练过程的时间复杂度. 在 ELM 中, 隐节点所张成的特征空间构成了隐空间, 其映射过程^[19]如下:

1) 随机生成权重矩阵 $\mathbf{W} \in R^{L \times d}$ 和偏移量矩阵 $\mathbf{B} = [b_1, b_2, \dots, b_L]^T$, 其中 L 是随机 ELM 隐节点数, d 是原始数据的维数.

2) 将原始数据映射到 L 维的隐空间中. 每一个输入数据都是一个 d 维的向量, $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$. 该特征映射可以表示为

$$h(x) = [h_1(x), \dots, h_L(x)]^T = [G(w_1, b_1, \mathbf{x}), \dots, G(w_L, b_L, \mathbf{x})]^T, \quad (1)$$

其中 $G(x)$ 是激励函数, 其映射过程如图 1 所示.

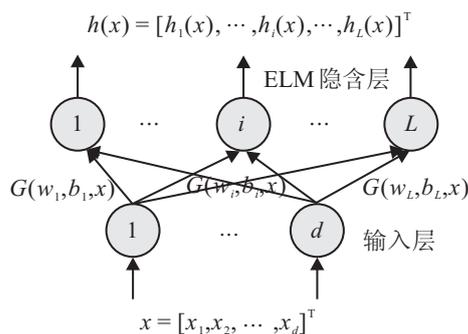


图 1 隐空间特征映射过程

常用的激励函数有以下几种:

Sigmoid 函数

$$G(x) = \frac{1}{1 + \exp(-x)}; \quad (2)$$

Sine 函数

$$G(x) = \sin x; \quad (3)$$

Gaussian 函数

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (4)$$

1.2 经典 FCM 及其问题分析

模糊 C 均值聚类算法是经典的模糊聚类算法, 1973 年由 Bezdek 等^[20]提出. 其思想直观, 便于实现, 是目前最常用的聚类方法之一. 但是, FCM 算法本身也存在问题: 1) 来自现实生活中的数据通常具有典型的非线性特性, 即至少包含了一个具有非凸形状边界的类簇, 这使得经典 FCM 算法在原数据空间中难以得到理想的聚类效果; 2) 模糊指数^[21]是 FCM 的一个重要参数, 其取值会严重影响算法的性能, 在没有先验知识的情况下, 用户通常很难选取合适的模糊指数来获得满意的聚类效果. 由于 FCM 算法存在以上问题, 近年来, 人们对 FCM 算法进行了大量改进, 其中, 基于核技术的模糊聚类方法通过引入核技巧将原始特征空间中的数据映射到高维空间中, 使非线性数据在高维空间中线性可分, 从而提高了算法线性可分的能力^[9, 22-23]. 但是, 在使用核方法的过程中, 由于缺乏先验知识, 用户通常难以选取合适的核函数以及核参数, 制约了核方法在模糊聚类算法中的使用及推广. 本文将 ELM 隐空间学习理论与特征降维技术相结合, 基于多隐层神经网络学习方法, 将传统的基于浅层学习结构的模糊聚类方法改造为多隐层学习结构, 从而提高了对非线性数据的无监督学习能力. 通过提出堆叠隐空间学习结构, 该方法能够有效精简冗余信息并及时补充必要的信息, 使得模糊聚类算法能够发挥出更好的性能。

2 堆叠隐空间模糊 C 均值聚类算法

2.1 基于主成分分析的压缩 ELM 隐空间

在 ELM 学习技术中, 可以通过随机赋值的方法来快速生成 ELM 隐空间, 随着隐节点数目的增加, 学习精度也不断提高^[13]. 但是, 随之而来的一个重要问题是计算效率会逐步降低. 此外, 由于隐节点生成过程中相关参数是随机生成的, 不可避免地会引入大量噪声. 针对这一问题, 本文中 ELM 隐空间技术与主成分分析相结合, 提出基于主成分分析的压缩隐空间构建方法, 其过程如下.

1) 根据图 1 所示过程将原始数据映射到高维空间 R^L 中, 得到矩阵 $\mathbf{H}^{(1)} \in R^{n \times L}$, n 为样本点个数.

2) 对 $H^{(1)}$ 的每一列 $H_j^{(1)}$ ($j = 1, 2, \dots, L$) 进行归一化, 得到矩阵 $A = [A_1, A_2, \dots, A_L]$, 其中 A_j 计算方法如下:

$$A_j = H_j^{(1)} - \frac{1}{n} \sum_{j=1}^n H_j^{(1)}, \quad j = 1, 2, \dots, L. \quad (5)$$

3) 计算矩阵 A 的协方差矩阵 $C = \text{cov}(A)$, 并计算矩阵 C 的特征值 λ_i 和特征向量 $V_i, i = 1, 2, \dots, L$.

4) 将特征值 λ_i 降序排序, 取前 p 个特征值对应的特征向量, 按列组成矩阵 \tilde{V} , PCA 压缩后的矩阵为

$$H^{(2)} = A\tilde{V}(1:p), \quad (6)$$

其结构可映射为如图 2 所示的前馈神经网络模型.

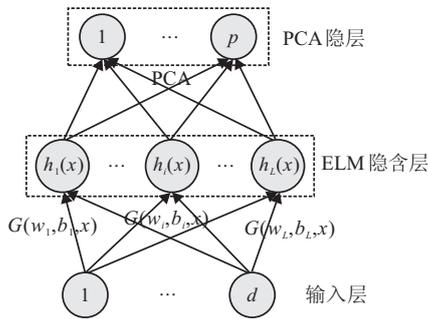


图 2 基于 PCA 的压缩 ELM 隐空间

2.2 堆叠隐空间模糊 C 均值聚类算法 (CHS-FCM)

上述压缩 ELM 隐空间构造过程可以转化为如图 2 所示的浅层结构的前馈神经网络模型. 在隐空间构建过程中, 隐节点数目的合理选取是一个重要问题. 为了使学习器得到更好的非线性表达能力和稳定的学习结果, 单隐层前馈神经网络学习模型通常会选择较多的隐节点数目, 但是, 这将增加额外的计算负担. 研究表明, 在保证学习器泛化能力的前提下, 将单隐层结构改造为多隐层结构是降低隐节点数目的有效方法. 为此, 本文对上述基于 PCA 的压缩 ELM 隐空间映射过程进行改造, 通过把 ELM 隐层中的隐节点分散到多个隐含层中, 并与 PCA 层的隐节点相结合, 形成新的混合隐含层. 再将若干个混合隐含层进行叠加, 得到堆叠隐空间学习结构, 如图 3 所示.

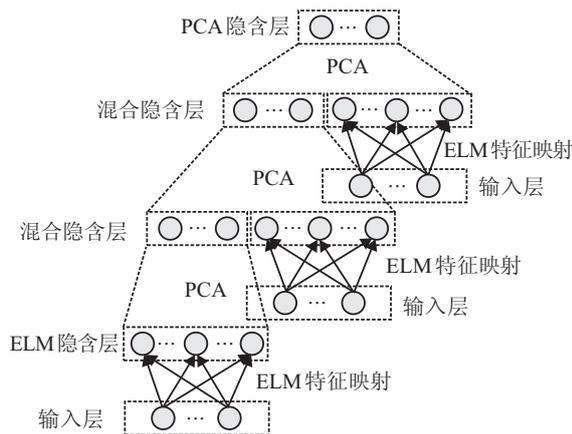


图 3 堆叠隐空间

CHS-FCM 算法流程如图 4 所示.

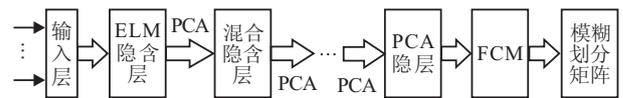


图 4 CHS-FCM 算法流程

输入: L 为随机 ELM 隐节点总数, d 为数据集 D 的维数, p 为 PCA 提取的特征数, f 为堆叠隐空间的层数, $T = L/f$ 为每层隐空间中随机 ELM 隐节点数, n 为样本点个数.

处理步骤如下:

- 1) 随机生成值在 $[0, 1]$ 之间的权重矩阵 $W \in R^{T \times d}$ 以及值在 $[0.5, 1]$ 之间的偏移量矩阵 $B \in R^{T \times 1}$;
- 2) 根据式 (1) 以及图 1 将数据集 D 进行特征映射, 得到高维数据矩阵 $H^{(1)} \in R^{n \times T}$;
- 3) 利用基于 PCA 的压缩隐空间方法对 $H^{(1)}$ 进行压缩, 得到维数为 p 的数据矩阵 $H^{(2)}$;
- 4) for $j = 1: f - 1$;
- 5) 重复步骤 1) 和步骤 2);
- 6) 将 $H^{(2)}$ 与 $H^{(1)}$ 合并存入矩阵 $H^{(3)}$ 中, $H^{(3)} = [H^{(2)}, H^{(1)}]$;
- 7) 利用 PCA 对 $H^{(3)}$ 进行压缩, 得到数据矩阵, 同样记为 $H^{(2)}$;
- 8) end;
- 9) 将最终获取的数据 $H^{(2)}$ 利用 FCM 算法进行聚类.

输出: 模糊划分矩阵 U .

CHS-FCM 算法将单隐层结构的隐空间改造为多隐层堆叠隐空间学习结构. 假设从第 2 层开始, 每个隐含层中的随机 ELM 隐节点数都是 T , 由前一层 PCA 压缩后得到的隐节点数为 p , 则易得隐节点总数为

$$L_0 = (T + p) \times (f - 1) + T, \quad (7)$$

其中随机 ELM 隐节点总数为

$$L = T \times f. \quad (8)$$

经过多次隐空间压缩, 每一层都会造成相应的信息损失, 但是经过每个混合层中随机 ELM 隐节点的信息弥补, 丢失的信息得以补偿, 因此能得到更好的学习效果.

2.3 CHS-FCM 算法复杂度分析

设聚类个数为 c , CHS-FCM 算法的时间复杂度由两部分构成: 1) 堆叠隐空间特征映射, 它包含 ELM 特征映射和 PCA 两个过程, 其时间复杂度为 $O((Tdn + p^2n + p^3)f)$, 其中 $O(Tdn)$ 为 ELM 特征映射的时间复杂度, $O(p^2n + p^3)$ 为 PCA 特征映射的时间复杂度; 2) FCM 聚类, 对于每次迭代, 其时间复杂度为

$O(np)$. 由此可知: CHS-FCM 算法的时间复杂度为 $O(nTdf + p^2nf + p^3f + npc)$; HS-FCM 算法的时间复杂度则由 ELM 特征映射和 FCM 聚类两部分构成, 即 $O(nLd + nLc)$. 在实际使用中, p 和 f 通常取较小值, p 又远小于 L , 所以 CHS-FCM 在后续的 FCM 聚类过程中所需的时间远少于 HS-FCM 的聚类时间. 在空间复杂度方面, HS-FCM 的空间复杂度可以表示为 $O(nL + cL)$, 而 CHS-FCM 算法的空间复杂度为 $O(n(T + p) + cp)$. 由于 p, T 远小于 L , 经过压缩隐空间特征映射后, FCM 学习过程更能节省空间.

3 实验研究与分析

3.1 实验数据集及其平台

本文选择 4 个具有代表性的数据集: 1) IRIS 和 X8D5K, 其共同特点是维数较低; 2) Libras movement 和 Synthetic control, 其共同特点是维数高, 分别达 90 维和 60 维. 为了研究不同类别数对聚类效果的影响,

本文将 Synthetic control 数据集中类标为 1、3 的数据点取出来, 构成包含 200 个样本点的子集 SC1; 将类标为 1、3、5 的数据点取出构成包含 300 个样本点的子集 SC2. 数据集的相关信息以及本文的实验平台如表 1、表 2 所示.

表 1 实验数据集

| 数据集 | 样本数 n | 特征数 d | 类别数 c | P | |
|-----------------|---------|---------|---------|---------|---------|
| | | | | PCA-FCM | CHS-FCM |
| IRIS | 150 | 4 | 3 | 2 | 3 |
| X8D5K | 1000 | 8 | 5 | 3 | 4 |
| Libras movement | 360 | 90 | 15 | 5 | 5 |
| SC1 | 200 | 60 | 2 | 4 | 5 |
| SC2 | 300 | 60 | 3 | 4 | 4 |

表 2 实验平台

| 结构 | 配置 |
|------|-----------------------------------|
| CPU | Intel(R)Core(TM) i5-4590 3.30 GHz |
| 内存 | 8.00 GB |
| 操作系统 | 64bit Windows7 |
| 软件平台 | Matlab R2012a |

表 3 各算法的说明以及相关参数设置

| 算法 | 算法说明及相关参数 |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| FCM ^[20] | 最大迭代次数 200 次, 收敛阈值 $1e-5$, 模糊指数 $m = 2$ |
| HS-FCM | 通过 ELM 隐空间映射后执行 FCM 算法. 随机 ELM 隐节点总数设为 1000, 激励函数使用 Sigmoid 函数; FCM 最大迭代次数 200 次, 收敛阈值 $1e-5$, 模糊指数 $m = 2$ |
| PCA-FCM | 通过 PCA 特征提取后执行 FCM 算法, PCA 提取特征数见表 1; FCM 最大迭代次数 200 次, 收敛阈值 $1e-5$, 模糊指数 $m = 2$ |
| KFCM-F ^[9] | 最大迭代次数 200, 收敛阈值 $1e-5$, 模糊指数 $m = 2$, 核函数如式 (4) 所示, 核参数由 $\sigma^2 = \frac{\sum_{i=1}^n \sum_{j=2}^n \ x_i - x_j\ ^2}{n^2}$ 计算得出 |
| CHS-FCM | 随机 ELM 隐节点总数设为 1000, 堆叠隐空间层数为 5, 激励函数 Sigmoid, PCA 提取特征数见表 1; FCM 最大迭代次数 200 次, 收敛阈值 $1e-5$, 模糊指数 $m = 2$ |

表 4 不同数据集集中的算法性能指标

| 数据集 | 性能指标 | FCM | HS-FCM | PCA-FCM | KFCM-F | CHS-FCM |
|-----------------|----------|----------------|----------------|----------------|----------------|----------------|
| IRIS | NMI_mean | 0.732 1 | 0.726 5 | 0.721 2 | 0.740 6 | 0.768 4 |
| | NMI_std | 0.043 2 | 0.013 1 | 0.020 8 | 0.042 7 | 0.001 1 |
| | RI_mean | 0.851 5 | 0.832 7 | 0.879 7 | 0.858 7 | 0.873 7 |
| | RI_std | 0.043 6 | 0.010 3 | 0.003 2 | 0.021 7 | 0.001 6 |
| X8D5K | NMI_mean | 0.895 7 | 0.876 9 | 0.877 5 | 0.829 1 | 0.894 2 |
| | NMI_std | 0.010 3 | 0.014 1 | 0.012 1 | 0.012 7 | 0.016 8 |
| | RI_mean | 0.887 9 | 0.914 5 | 0.894 5 | 0.881 4 | 0.927 2 |
| | RI_std | 0.005 5 | 0.001 8 | 0.010 4 | 0.010 5 | 0.005 5 |
| Libras movement | NMI_mean | 0.312 2 | 0.405 2 | 0.400 8 | 0.430 5 | 0.567 6 |
| | NMI_std | 0.037 6 | 0.025 2 | 0.041 3 | 0.031 3 | 0.013 2 |
| | RI_mean | 0.732 8 | 0.776 2 | 0.709 9 | 0.732 7 | 0.741 1 |
| | RI_std | 0.024 1 | 0.021 2 | 0.040 8 | 0.023 7 | 0.002 5 |
| SC1 | NMI_mean | 0.906 5 | 0.901 3 | 0.902 9 | 0.916 5 | 0.918 8 |
| | NMI_std | 0.013 5 | 0.043 1 | 0.032 4 | 0.014 2 | 0.024 7 |
| | RI_mean | 0.973 3 | 0.965 5 | 0.934 8 | 0.971 2 | 0.982 2 |
| | RI_std | 0.022 7 | 0.021 4 | 0.031 2 | 0.032 6 | 0.005 7 |
| SC2 | NMI_mean | 0.728 9 | 0.735 5 | 0.726 7 | 0.744 1 | 0.765 2 |
| | NMI_std | 0.021 6 | 0.034 3 | 0.023 2 | 0.013 8 | 0.011 4 |
| | RI_mean | 0.748 7 | 0.742 1 | 0.724 5 | 0.761 2 | 0.770 1 |
| | RI_std | 0.024 1 | 0.020 3 | 0.037 5 | 0.015 4 | 0.022 3 |

为了表明隐空间、PCA 技术、核方法等技术对 FCM 性能的影响, 本文将 CHS-FCM 算法与 FCM、HS-FCM、PCA-FCM、KFCM-F 进行比较, 以此验证本文算法的性能. 各算法相关说明及其参数设置如表 3 所示. 在 CHS-FCM 和 PCA-FCM 中, PCA 提取不同特征数目对聚类结果有较大影响, 实验中仅记录算法得到的最佳性能及相应的 PCA 所提取的特征数 p , 分别如表 4 和表 1 所示.

3.2 算法性能的评价指标

本文采用 NMI 和 RI 两个评价指标来分析各算法的性能^[24-25].

1) 归一化互信息 (NMI)

$$NMI = \frac{\sum_{i=1}^C \sum_{j=1}^C N_{i,j} \log N \times N_{i,j} / N_i \times N_j}{\sqrt{\sum_{i=1}^C N_i \times \log N_i / N \times \sum_{j=1}^C N_j \times \log N_j / N}} \quad (9)$$

其中: N_{ij} 为第 i 个聚类与第 j 类的契合程度, N_i 为第 i 个聚类所包含数据样本量, N_j 为类 j 所包含的数据样本量, N 为整个数据样本大小.

2) 芮氏指标 (RI)

$$RI = \frac{f_{00} + f_{11}}{N(N-1)/2} \quad (10)$$

其中: f_{00} 为数据点具有不同的类标签并且属于不同类的配对点数目, f_{11} 为数据点具有相同的类标签并且属于同一类的配对点数目, N 为整个数据样本的总

量大小.

以上两种评价指标的取值范围均为 $[0, 1]$, 数值越大, 算法的性能越好.

3.3 聚类效果对比实验

下面将 CHS-FCM 与 FCM、PCA-FCM、HS-FCM 以及 KFCM-F 这几个基于浅层学习结构的模糊聚类算法在如表 1 所示各数据集上进行实验. 实验平台如表 2 所示, 各算法的参数设置如表 3 所示, 每个算法重复运行 20 次, 各算法的 NMI、RI 指标的均值和标准差如表 4 所示 (其中粗体表示较好的实验结果).

从实验结果可以发现, 虽然各数据集内部结构复杂, Libras movement、SC1、SC2 这 3 个数据集的特征数分别达到 90、60 和 60, 但是本文所提出的 CHS-FCM 算法均可以取得良好且稳定的聚类效果. 其他基于浅层结构的 FCM 聚类算法 (FCM、HS-FCM、PCA-FCM 和 KFCM-F) 在个别数据集上也可以得到较为满意的聚类结果, 但总体性能不如本文所提出的 CHS-FCM 算法. 这也充分展示了堆叠隐空间学习结构处理复杂结构数据时的强大能力.

3.4 随机 ELM 隐节点总数相同时聚类效率对比

本实验将研究在随机 ELM 隐节点总数相同的前提下, 浅层学习结构和多层学习结构对算法性能的影响. 实验中分别执行 CHS-FCM 和 HS-FC 算法, 并为两个算法设置相同的随机 ELM 隐节点总数 (1 000、2 000). 实验中算法的其他参数如表 3 所示, 运行结果如表 5 所示.

表 5 CHS-FCM 与 HS-FCM 在聚类过程中的效率对比

| 数据集 | FCM 聚类过程所需的时间/s | | 隐节点总数 | | NMI_mean | | RI_mean | |
|-----------------|-----------------|----------------|--------|---------|----------------|----------------|----------------|----------------|
| | HS-FCM | CHS-FCM | HS-FCM | CHS-FCM | HS-FCM | CHS-FCM | HS-FCM | CHS-FCM |
| IRIS | 0.155 6 | 0.006 1 | 1 000 | 1 000 | 0.726 5 | 0.768 4 | 0.832 7 | 0.873 7 |
| | 0.369 5 | 0.019 2 | 2 000 | 2 000 | 0.728 2 | 0.735 2 | 0.821 5 | 0.856 2 |
| X8D5K | 2.126 3 | 0.019 9 | 1 000 | 1 000 | 0.876 9 | 0.894 2 | 0.914 5 | 0.927 2 |
| | 4.997 8 | 0.049 8 | 2 000 | 2 000 | 0.901 2 | 0.894 9 | 0.950 7 | 0.940 6 |
| Libras movement | 0.121 2 | 0.001 3 | 1 000 | 1 000 | 0.405 2 | 0.567 6 | 0.776 2 | 0.741 1 |
| | 0.278 3 | 0.002 4 | 2 000 | 2 000 | 0.432 4 | 0.557 9 | 0.755 4 | 0.756 5 |
| SC1 | 0.010 2 | 0.000 3 | 1 000 | 1 000 | 0.901 3 | 0.918 8 | 0.965 5 | 0.982 2 |
| | 0.020 9 | 0.000 7 | 2 000 | 2 000 | 0.902 7 | 0.901 4 | 0.946 7 | 0.935 6 |
| SC2 | 0.020 0 | 0.000 8 | 1 000 | 1 000 | 0.735 5 | 0.765 2 | 0.742 1 | 0.770 1 |
| | 0.047 1 | 0.001 8 | 2 000 | 2 000 | 0.726 8 | 0.725 2 | 0.735 4 | 0.723 1 |

从对本文算法的复杂度分析以及本次实验的结果中容易看出, HS-FCM 通过构建 ELM 隐空间将原始数据映射到高维隐空间, 在此过程中, 由于隐节点总数增加导致了特征数目的增加, 利用 FCM 聚类其效率会明显降低. 虽然 CHS-FCM 涉及了多次隐空间映射和压缩, 但是在该过程中隐空间中的冗余信息得到精简, 噪声被有效过滤, 同时每一层中进行了适度的信息弥补, 这使得该算法在后续的 FCM 聚类过

程中能够取得既快又好的学习效果. 由此可见, 随机 ELM 隐节点总数相同的情况下, 本文所提出的多层学习结构可以取得更好的学习效果.

3.5 参数敏感性实验

在模糊聚类算法中, 模糊指数 m 是一个重要指标, 也严重影响着算法的执行效果. 基于表 1 中的数据集以及表 2 中的实验平台, 本实验将比较 CHS-FCM、FCM、PCA-FCM、KFCM-F 以及 HS-FCM 等算法随

着模糊指数 m 的变化对算法性能的影响. 文中仅给出 NMI 指标的结果, RI 有类似结果, 受篇幅所限, 不再列出. 实验中, m 值依次取 $\{1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2\}$, 各算法的其他参数如表 3 所示, 实验结果如图 5 所示.

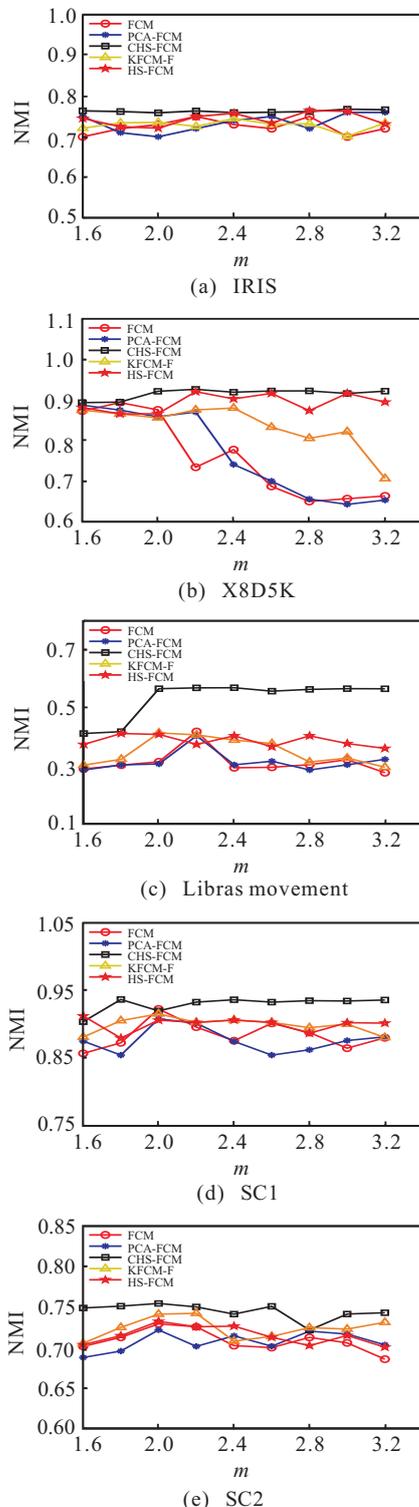


图5 不同数据集中各算法的NMI随 m 的变化情况

从图5结果中不难发现, 虽然模糊指数 m 在较为宽广的范围内变化, 但CHS-FCM在各数据集上均可取得令人满意且较为稳定的结果. 其他各算法的性能却随着 m 的变化有着较大的波动, 可见经过堆叠隐空

间的压缩与重组, CHS-FCM算法对模糊指数 m 的变化具有更好的鲁棒性, 这将有利于用户在实际应用中更方便地选取模糊指数.

CHS-FCM和KFCM-F均有着强大的非线性数据处理能力, 下面研究这两个算法的关键参数的动态变化对聚类结果产生的影响.

在CHS-FCM中, 隐节点数目是一个重要参数; 在KFCM-F中, 核参数 σ 是一个重要参数. 实验中, 随机ELM隐节点总数在 $\{200, 400, \dots, 1000, 1200, 1400, 1600, 1800\}$ 范围内变化, σ 值在 $\{0.1, 1, 5, 10, 50, 100, 150, 200, 1000\}$ 范围内变化, 算法的其他参数取值如表3所示, 实验结果如图6所示.

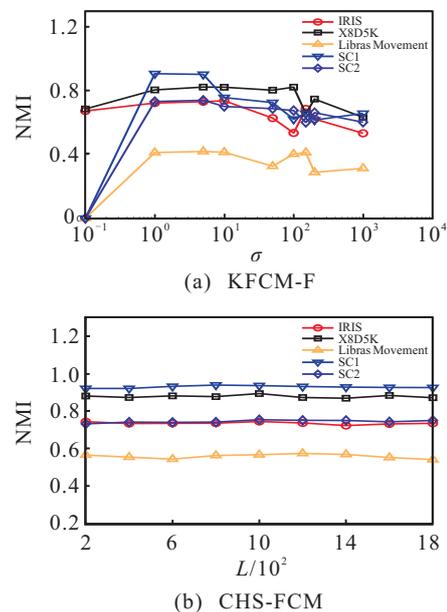


图6 KFCM-F和CHS-FCM算法随各自关键参数的变化情况

在表1所示的5个不同数据集上运行KFCM-F算法, 由图6的实验结果不难发现, 当 σ 变化时, KFCM-F的聚类效果出现了明显波动. 虽然随机ELM隐节点总数在较为宽广的区间内变化, 但CHS-FCM算法对该参数的变化并不敏感, 同时也能取得良好且稳定的聚类效果. 由此可见, CHS-FCM算法可以较为方便地选取相关参数, 且聚类效果令人满意. 这也是本文所提出压缩隐空间较传统核空间的一大优势.

4 结论

本文提出了堆叠隐空间模糊 C 均值聚类算法, 实验结果表明了该方法具有强大而高效的非线性数据的处理能力. 同时, 本文所提出的CHS-FCM算法对模糊指数和隐节点数目等重要参数不敏感, 这有利于用户的实际选用. 当前, 随着模糊聚类算法研究的不断深入, 人们基于不同理论提出了更先进的模糊聚类算法. 如何将本文所提出的堆叠隐空间技术应用于其他模糊聚类算法中, 将是今后研究的重点.

参考文献(References)

- [1] 王骏, 王士同, 邓赵红. 聚类分析研究中的若干问题[J]. 控制与决策, 2012, 27(3): 321-328.
(Wang J, Wang S T, Deng Z H. Survey on challenges in clustering analysis research[J]. Control and Decision, 2012, 27(3): 321-328.)
- [2] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48-61.
(Sun J G, Liu J, Zhao L Y. Clustering algorithm research[J]. J of Software, 2008, 19(1): 48-61.)
- [3] 蒋亦樟, 邓赵红, 王骏, 等. 熵加权多视角协同划分模糊聚类算法[J]. 软件学报, 2014, 25(10): 2293-2311.
(Jiang Y Z, Deng Z H, Wang J, et al. Collaborative partition multi-view fuzzy clustering algorithm using entropy weighting[J]. J of Software, 2014, 25(10): 2293-2311.)
- [4] Glenn T C, Zare A, Gader P D. Bayesian fuzzy clustering[J]. IEEE Trans on Fuzzy Systems, 2015, 23(5): 1545-1561.
- [5] 王骏, 王士同. 基于混合距离学习的双指数模糊 C 均值算法[J]. 软件学报, 2010, 21(8): 1878-1888.
(Wang J, Wang S T. Double indices-induced FCM algorithm based on hybrid distance metric learning[J]. J of Software, 2010, 21(8): 1878-1888.)
- [6] Wang J, Chung F L, Wang S T, et al. Double indices-induced FCM clustering and its integration with fuzzy subspace clustering[J]. Pattern Analysis and Applications, 2014, 17(3): 549-566.
- [7] Maraziotis I A A. Semi-supervised fuzzy clustering algorithm applied to gene expression data[J]. Pattern Recognition, 2012, 45(1): 637-648.
- [8] Zhang L, Cao Q. A novel ant-based clustering algorithm using the kernel method[J]. Information Sciences, 2011, 181(20): 4658-4672.
- [9] Graves D, Pedrycz W. Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study[J]. Fuzzy Sets and Systems, 2010, 161(4): 522-543.
- [10] Girolami M. Mercer kernel-based clustering in feature space[J]. IEEE Trans on Neural Networks, 2002, 13(3): 780-784.
- [11] Cover T M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition[J]. IEEE Trans on Electronic Computers, 1965, 14(3): 326-334.
- [12] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1): 489-501.
- [13] Huang G B, Zhou H, Ding X, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 2012, 42(2): 513-529.
- [14] Huang G B, Wang D H, Lan Y. Extreme learning machines: A survey[J]. Int J of Machine Learning and Cybernetics, 2011, 2(2): 107-122.
- [15] Huang G B, Chen L. Convex incremental extreme learning machine[J]. Neurocomputing, 2007, 70(16): 3056-3062.
- [16] Huang G B, Ding X, Zhou H. Optimization method based extreme learning machine for classification[J]. Neurocomputing, 2010, 74(1): 155-163.
- [17] Price A L, Patterson N J, Plenge R M, et al. Principal components analysis corrects for stratification in genome-wide association studies[J]. Nature genetics, 2006, 38(8): 904-909.
- [18] Zhou H, Huang G B, Lin Z, et al. Stacked extreme learning machines[J]. IEEE Trans on Cybernetics, 2014, 45(9): 2013-2025.
- [19] He Q, Jin X, Du C Y, et al. Clustering in extreme learning machine feature space[J]. Neurocomputing, 2014, 128(5): 88-95.
- [20] Bezdek J C. Pattern recognition with fuzzy objective function algorithms[M]. New York: Plenum Press, 1981: 203-239.
- [21] 于剑. 论模糊 C 均值算法的模糊指标[J]. 计算机学报, 2003, 26(8): 968-973.
(Yu J. On the fuzziness index of the FCM algorithms[J]. Chinese J of Computers, 2003, 26(8): 968-973.)
- [22] 沈红斌, 王士同, 吴小俊. 离群模糊核聚类算法[J]. 软件学报, 2004, 15(7): 1021-1029.
(Shen H B, Wang S T, Wu X J. Fuzzy kernel clustering with outliers[J]. J of Software, 2004, 15(7): 1021-1029.)
- [23] Zhang D Q, Chen S C. A novel kernelized fuzzy c -means algorithm with application in medical image segmentation[J]. Artificial Intelligence in Medicine, 2004, 32(1): 37-50.
- [24] Jing L, Ng M K, Huang J Z. An entropy weighting k -means algorithm for subspace clustering of high-dimensional sparse data[J]. IEEE Trans on Knowledge and Data Engineering, 2007, 19(8): 1026-1041.
- [25] Deng Z H, Choi K S, Chung F L, et al. Enhanced soft subspace clustering integrating within-cluster and between-cluster information[J]. Pattern Recognition, 2010, 43(3): 767-781.