

# 鲁棒性的模糊聚类神经网络\*

邓赵红<sup>1,2,3</sup>, 王士同<sup>1,2,3+</sup>

<sup>1</sup>(江南大学 信息工程学院,江苏 无锡 214036)

<sup>2</sup>(计算机软件新技术国家重点实验室(南京大学),江苏 南京 210093)

<sup>3</sup>(中国科学院 计算机科学重点实验室,北京 100080)

## Robust Fuzzy Clustering Neural Networks

DENG Zhao-Hong<sup>1,2,3</sup>, WANG Shi-Tong<sup>1,2,3+</sup>

<sup>1</sup>(School of Information Engineering, Southern Yangtze University, Wuxi 214036, China)

<sup>2</sup>(State Key Laboratory for Novel Software Technology at Nanjing University, Nanjing 210093, China)

<sup>3</sup>(Laboratory of Computer Science, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-510-8373140, E-mail: wxwangst@yahoo.com.cn, http://www.sytu.edu.cn

Received 2004-03-24; Accepted 2004-07-06

Deng ZH, Wang ST. Robust fuzzy clustering neural networks. *Journal of Software*, 2005,16(8):1415–1422.

DOI: 10.1360/jos161415

**Abstract:** In this paper a new robust fuzzy clustering neural networks (RFCNN) is presented to resolve the sensitivity of the fuzzy clustering neural network (FCNN) to outliers in real datasets. The new objective function of RFCNN is obtained by introducing Vapnik's  $\epsilon$ -insensitive loss function, and RFCNN's update rules are derived by using Lagrange optimization theory. Compared with the FCNN algorithm, RFCNN is much more robust to outliers in the datasets. Experimental results demonstrate the effectiveness of RFCNN.

**Key words:** fuzzy clustering; neural network;  $\epsilon$ -insensitive loss function; outliers; robustness

**摘要:** 针对模糊聚类神经网络 FCNN(fuzzy clustering neural network)对例外点(outliers)敏感的缺陷,通过引入 Vapnik's  $\epsilon$ -不敏感损失函数,重新构造网络的目标函数,并根据拉格朗日优化理论推导出新的鲁棒模糊聚类神经网络及其算法(robust fuzzy clustering neural networks,简称 RFCNN).RFCNN 有效地克服了 FCNN 对例外点敏感之缺点并且能得到合理的聚类中心.仿真实验结果表明,RFCNN 较之于 FCNN 有更好的鲁棒性.

**关键词:** 模糊聚类;神经网络; $\epsilon$ -不敏感损失函数;例外点;鲁棒性

中图法分类号: TP183 文献标识码: A

---

\* Supported by the National Natural Science Foundation of China under Grant No.60225015 (国家自然科学基金); the Natural Science Foundation of Jiangsu Province of China under Grant No.BK2003017 (江苏省自然科学基金); the 2004-New-Century Outstanding Young Teacher Grant of Ministry of Education of China under Grant No.NCET-040496 (2004 年度国家教育部新世纪优秀人才计划); the 2005 Key Research Project of Ministry of Education of China (2005 年度国家教育部科学研究重点项目)

作者简介: 邓赵红(1981 - ),男,安徽蒙城人,博士生,主要研究领域为人工智能,模式识别,模糊系统;王士同(1964 - ),男,博士,教授,博士生导师,主要研究领域为人工智能,模式识别,模糊系统,生物信息学.

聚类在模式识别、系统建模、图像处理及数据挖掘等工程领域扮演着重要的角色<sup>[1-3]</sup>。目前针对不同的应用,根据不同的理论已提出了许多各具特色的聚类算法<sup>[4-7]</sup>。业已证明多层前馈网络经过示例学习可以识别任意样式的聚类模式,但在某些应用中,样本点的类别是未知的、不确定的,这就需要一种可以根据样本特征自适应地对样本集进行聚类的算法。基于无监督学习的模糊聚类算法被广泛应用于对未知样本进行自适应地聚类。模糊聚类算法包含了一大类基于模糊技术的聚类算法,其中最著名的是 FCM 聚类算法。Kohonen<sup>[8]</sup>提出了一种处理模糊聚类问题的神经网络模型 LVQ。Zhang 和何丕廉等人对 Kohonen 模型进行了改进,提出了相应的学习算法 FCNN(fuzzy clustering neural networks)和不对称学习算法<sup>[9]</sup>。FCNN 算法和不对称学习算法能在某些条件下对未知样本集进行有效的聚类,但我们的实验也发现它们存在一些不足之处,其中一个最显著的缺陷是对样本集中的例外点(outliers)很敏感。而真实的数据集经常含有一定数目的例外点,FCNN 算法对例外点所具有的高敏感性极大地影响了它的适用性。

对于聚类的鲁棒性研究已有不少方法<sup>[10,11]</sup>,与现有的方法相比,本文方法的目的在于得到将 Vapnik's  $\varepsilon$ -不敏感损失函数及其优化方法引入到 FCNN 模糊聚类神经网络的研究。本文针对 FCNN 算法对例外点敏感这一缺陷,提出了一种鲁棒的模糊聚类神经网络 RFCNN(robust fuzzy clustering neural networks)。通过引入 Vapnik's  $\varepsilon$ -不敏感损失函数<sup>[12,13]</sup>重新构造网络的目标函数,并根据拉格朗日优化理论<sup>[14]</sup>推导出新的网络 RFCNN 及其学习算法。它能有效地克服 FCNN 的例外点敏感性。仿真结果表明,RFCNN 算法相对于 FCNN 算法有着更好的鲁棒性。

## 1 FCNN 算法及其例外点敏感性

Zhang<sup>[9]</sup>对 Kohonen 模型进行了改进,提出了相应的学习算法 FCNN,它可以等价地表述如下  
对构造的网络目标函数式(1),通过式(2)的向量修正公式进行网络学习。

$$J_m = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \frac{1}{\sigma^2} C_{ij} U_{ij} \| \mathbf{x}_j - \mathbf{v}_i \|^2 \quad (1)$$

$$\mathbf{v}_i(t+1) = \mathbf{v}_i(t) + \eta \sum_{j=1}^N \frac{1}{\sigma^2} C_{ij}(t) U_{ij}(t) (\mathbf{x}_j - \mathbf{v}_i(t)) \quad (2)$$

其中  $i=1, \dots, M, j=1, \dots, N$ 。

$$C_{ij}(t) = \exp\left(-\frac{d_{ij}^2(t)}{2\sigma^2}\right) \quad (3)$$

$$d_{ij}(t) = \| \mathbf{x}_j - \mathbf{v}_i \| \quad (4)$$

$$U_{ij}(t) = \frac{C_{ij}(t)}{\sum_{h=1}^M C_{hj}(t)} \quad (5)$$

$\mathbf{x}_j$  为第  $j$  个样本向量,  $\mathbf{v}_i$  为第  $i$  类的聚类中心,  $\eta, \sigma$  为常数,  $\|\cdot\|$  为欧氏距离,  $M$  为要聚类的类数,  $N$  为样本点的个数。

FCNN 算法由于网络所固有的缺陷使得该算法对例外点有很强的敏感性。对于含有例外点的数据集,FCNN 算法的聚类结果往往不能稳定于真正的聚类中心。如图 1 所示,其中图 1(a)为所示的二维数据集不含例外点时,FCNN 算法的聚类结果,图 1(b)为含有一定数目的例外点时,FCNN 算法的聚类结果,可以看出,当样本集含有一定数目的例外点时,FCNN 算法的聚类结果可能严重偏离真正的聚类中心。

## 2 具有更好鲁棒性的模糊聚类神经网络 RFCNN

### 2.1 基于 Vapnik's $\varepsilon$ -不敏感损失函数的网络学习目标函数的构造

FCNN 算法中用二次损失函数作为样本和聚类中心区分的度量尺度,选用二次损失函数作为尺度的目的是为了算法的简便性和较低的计算负担。但同时也造成了 FCNN 算法对于例外点有很强的敏感性。文献[15]中提出了许多鲁棒的损失函数,其中 Vapnik's  $\varepsilon$ -不敏感损失函数<sup>[13]</sup>由于其简便性最受关注。本文提出的模糊聚类

神经网络算法 RFCNN 就是通过引入 Vapnik's  $\varepsilon$ -不敏感损失函数来构造网络目标函数而得到的.



(a) The clustering result of FCNN without outliers  
(b) The clustering result of FCNN with outliers  
(a) 不含例外点时 FCNN 算法聚类效果  
(b) 含有一定数目的例外点时 FCNN 算法聚类效果

Fig.1

图 1

Vapnik's  $\varepsilon$ -不敏感损失函数可表示为

$$|f|_{\varepsilon} = \begin{cases} 0, & |f| \leq \varepsilon \\ |f| - \varepsilon, & |f| > \varepsilon \end{cases} \quad (6)$$

这里,  $\varepsilon \geq 0$  表示不敏感性参数,当  $\varepsilon = 0$  时,Vapnik's  $\varepsilon$ -不敏感损失函数变为绝对误差损失函数.

利用 Vapnik's  $\varepsilon$ -不敏感损失函数代替二次损失函数,同时引入不敏感参数惩罚项重新构造 FCNN 目标函数,此时目标函数表示为

$$J_{m\varepsilon} = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \frac{1}{\sigma^2} C_{ij} U_{ij} \| \mathbf{x}_j - \mathbf{v}_i \|_{\varepsilon_i} + \gamma \sum_{i=1}^M \| \varepsilon_i \| \quad (7)$$

这里,

$$\| \mathbf{x}_j - \mathbf{v}_i \|_{\varepsilon_i} = \sum_{k=1}^d |x_{jk} - v_{ik}|_{\varepsilon_{ik}} \quad \| \varepsilon_i \| = \sum_{k=1}^d | \varepsilon_{ik} |, \gamma \geq 0 \quad (8)$$

$\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{id}) (\varepsilon_{ih} \geq 0, h = 1, \dots, d)$  为第  $i$  类不敏感参数向量.式(8)定义的好处在于有很好的抗例外点干扰能力,相关文献<sup>[15,16]</sup>已有所论证.值得指出的是,我们发现新构造的目标函数中引入不敏感参数惩罚项具有如下优点:

(1) 新的目标函数可以在保证  $J_{m\varepsilon}$  的第 1 项最小的同时使不敏感参数惩罚项尽可能地小.通过调整  $\gamma$  可以灵活地根据需要确定不敏感参数惩罚项的影响.

(2) 引入不敏感参数惩罚项能够使推导出的新算法自动调节不敏感性参数,不再需要人为地确定不敏感性参数,这就避免了不敏感性参数选取的随意性.这一点由附录部分的定理证明可以看出.

根据所构造的上述目标函数,通过拉格朗日优化理论,我们就可以推导出 RFCNN 学习算法,它能有效地克服 FCNN 算法对例外点敏感这一缺点,即具有很好的鲁棒性.

## 2.2 RFCNN算法

定理. 给定  $M, N$ , 且  $U = [U_{ij}], i = 1, \dots, M, j = 1, \dots, N$ , 由式(3)~式(5)给定, 则  $J_{m\varepsilon}$  取最小时当且仅当

$$\begin{cases} v_{ik} = (x_{ik} + x_{mk})/2 \\ \varepsilon_{ik} = (x_{mk} - x_{ik})/2 \end{cases}, \forall \{l, m \mid \lambda_{ikl}^+ \in \Gamma_{ik}^+, \lambda_{ikl}^- \in \Gamma_{ik}^-\}, i = 1, \dots, M, k = 1, \dots, d, l, m = 1, \dots, N \quad (9)$$

这里,

$$\begin{cases} \Gamma_{ik}^+ = \{ \lambda_{ikj}^+ \mid \lambda_{ikj}^+ \in (0, C_{ij} U_{ij} / 2\sigma^2) \} \\ \Gamma_{ik}^- = \{ \lambda_{ikj}^- \mid \lambda_{ikj}^- \in (0, C_{ij} U_{ij} / 2\sigma^2) \} \end{cases}, j = 1, \dots, N \quad (10)$$

$$\left. \begin{aligned} & \min \sum_{j=1}^N (\lambda_{ikj}^+ - \lambda_{ikj}^-) x_{jk} \\ & \left\{ \begin{aligned} & \sum_{j=1}^N \lambda_{ikj}^+ = \sum_{j=1}^N \lambda_{ikj}^- \\ & \sum_{j=1}^N (\lambda_{ikj}^+ + \lambda_{ikj}^-) \leq \gamma \\ & \lambda_{ikj}^+, \lambda_{ikj}^- \in [0, C_{ij} U_{ij} / 2\sigma^2] \end{aligned} \right. \end{aligned} \right\} \quad (11)$$

基于上面的定理,可以得到具有更好鲁棒性的模糊聚类神经网络算法 RFCNN.

- (1) 设置参数  $M(1 < M < N), \varepsilon > 0, \sigma > 0, \gamma > 0$  以及精度级  $\zeta_a > 0$ , 初始化聚类中心  $V(0) = [v_{ik}(0)], i=1, \dots, M, k=1, \dots, d$  和不敏感参数  $\varepsilon(0) = [\varepsilon_{ik}], \varepsilon_{ik} > 0, i=1, \dots, M, k=1, \dots, d, s=0$ . 根据式(3)~式(5)计算  $U(s) = [U_{ij}(0)], i=1, \dots, M, j=1, \dots, N$ .
- (2) 由式(9)更新聚类中心和不敏感参数分别为  $V(s+1)$  和  $\varepsilon(s+1)$ , 同时更新  $U(s)$  为  $U(s+1)$ .
- (3) 若  $\|U(s+1) - U(s)\| \leq \zeta_a$ , 则停止; 否则, 置  $s = s + 1$  并返回(2).

对于 RFCNN 算法和 FCNN 算法中相同的参数  $\sigma$ , 可以理解为模糊度参数.  $\sigma$  取值越大, 样本点相对于各类的隶属关系越模糊, 对于  $M$  类样本, 当  $\sigma$  很大时, 样本点相对于各类的隶属度趋近于  $1/M$ , 当  $\sigma$  太小时, 两种算法对初始类中心都特别敏感, 因此  $\sigma$  取值不能太大也不能太小. 如何选取最佳的参数值  $\sigma$ , 目前还没有很好的理论依据.

式(11)中  $\gamma \geq 0$ , 但当  $\gamma = 0$  时, 由式(11)可知,  $\lambda_{ikj}^+, \lambda_{ikj}^- = 0$ , 此时由式(9)、式(10)可知  $v_{ik}, \varepsilon_{ik}$  得不到更新, 故  $\gamma > 0$  时算法才有实际意义.

### 3 仿真实验及实验分析

为了比较 RFCNN 算法和 FCNN 算法的性能, 基于人造数据集和基准数据集分别做了仿真实验.

实验 1. 图 1(a)所示的数据集包含了 3 类清晰可分的样本数据, 每类包含 20 个样本点, 各类真正的中心分别为 (2.9346, -1.7983), (2.9595, 3.2018), (4.9384, 3.0445). 为了实验的可比性, 实验中采用相同的初始类中心: (1.75, -0.25), (1.45, 1.55) 和 (4.5, 0.5), 模糊度参数  $\sigma$  取值为 1. 规定 RFCNN 算法执行的最大迭代次数为 20 次, FCNN 算法执行的最大迭代次数为 300 次. 通过在坐标(6, -1)处增加不同数目的例外点来测试两种算法的抗例外点干扰能力.

如图 2(a)所示为 FCNN 算法对数据集加不同数目例外点的聚类结果. 如图 2(b)所示为  $\gamma=5$  时, RFCNN 算法对数据集加不同数目例外点的聚类结果. 图中虚线表示迭代过程中各类中心的变化轨迹.



(a) The clustering results of FNCC with a varying number(0,3,6,9,12) (b) The clustering results of RFNCC ( $\gamma=5$ ) of outliers with a varying number(0,3,6,9,12) of outliers

(a) 例外点个数分别为 0,3,6,9,12 时,FCNN 算法的聚类效果 (b) 例外点个数分别为 0,3,6,9,12, $\gamma=5$  时,RFCNN 算法的聚类效果

Fig.2

图 2

如果真正的聚类中心用矩阵  $V_1$  表示, 存在例外点时算法执行得到的聚类中心用矩阵  $V_2$  表示, 则可用  $\Delta V = \|V_1 - V_2\|_2$  表示受例外点影响类中心的偏移量.  $\Delta V$  越小, 表示算法的抗例外点干扰能力越强. 表 1 给出了 FCNN 算法和 RFCNN 算法取不同参数值  $\gamma$  时类中心的偏移量.

从表 1 可以看出, FCNN 算法聚类结果受例外点影响很大, 随着例外点数目的增加, 偏移量  $\Delta V$  越来越

大.RFCNN 算法受例外点影响则较小,当  $\gamma \geq 3$  时,RFCNN 算法的聚类结果偏移量  $\Delta V$  很小,而且随着例外点数目在一定范围内的增加,偏移量  $\Delta V$  基本稳定在一定误差之内.同时从表中还可以看出, $\gamma=4$ 时在所有的  $\gamma$ 取值中聚类效果最好,当  $\gamma \geq 8$ 时偏移量  $\Delta V$  基本稳定不变且能得到合理的聚类中心,当  $\gamma \leq 2$ 时  $\Delta V$  随着例外点数目的增加,偏移量  $\Delta V$  变化虽不大,但 RFCNN 算法聚类中心稳定点已不合理.实验结果表明,当  $\gamma$ 取较大值时,RFCNN 算法能得到较理想的聚类结果.需要指出的是,对于该实验的数据集,当例外点数目很少时,FCNN 算法聚类的偏移量  $\Delta V$  可能较 RFCNN 算法聚类的偏移量  $\Delta V$  要小,但随着例外点数目的增加,选取合适参数  $\gamma$ ,RFCNN 算法聚类效果明显优于 FCNN 算法聚类效果.

**Table 1** Performance comparison between FCNN and RFCNN on the artificial dataset

表 1 FCNN 算法和 RFCNN 算法对人造数据集聚类效果比较

Outlier number $\gamma$	RFCNN $\Delta V$							FCNN $\Delta V$
	0.001	0.5	2	4	6	8	100	
0	2.741 3	1.236 0	0.875 9	0.100 8	0.136 4	0.288 6	0.288 6	0.134 9
3	2.904 1	1.236 0	0.878 4	0.100 8	0.136 4	0.288 6	0.288 6	0.134 9
6	2.904 1	1.236 0	0.878 4	0.100 8	0.136 4	0.288 6	0.288 6	0.141 5
9	2.904 1	1.236 0	0.878 4	0.072 0	0.136 4	0.288 6	0.288 6	5.982 1
12	2.904 1	1.678 6	0.878 4	0.125 1	0.136 4	0.288 6	0.288 6	5.982 6

实验 2.本实验采用著名的 IRIS 数据集.IRIS 数据集包含 3 类四维样本:setosa,versicolor, virginica,其中每类样本数目为 50.实验中保留了其中的第 3 维和第 4 维重新构造二维数据集,如图 3 所示.实验时,把数据重新分为两类,setosa 为一类,versicolor 和 virginica 为一类.采用 RFCNN 算法和 FCNN 算法对数据集聚类时,规定 RFCNN 算法执行的最大迭代次数为 20 次,FCNN 算法执行的最大迭代次数为 300 次.初始类中心分别初始化为 (3,0.4) 和 (5,-0.3),模糊度参数  $\sigma$ 取值为 1.5.实验中通过在坐标(6,-1)处人为地增加不同数目的例外点来考查两种算法的抗例外点干扰能力.表 2 给出了 FCNN 算法和 RFCNN 算法取不同参数值  $\gamma$ 时类中心的偏移量.

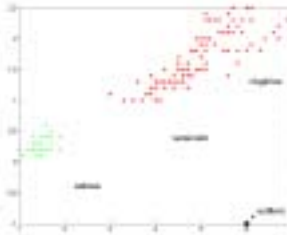


Fig.3 Distribution of data along the 3rd and the 4th dimensions in dataset IRIS

图 3 保留第 3 和第 4 维特征的 IRIS 数据集分布

**Table 2** The performance comparison between FCNN and RFCNN on dataset IRIS

表 2 FCNN 算法和 RFCNN 算法对 IRIS 数据集聚类效果比较

Outlier number $\gamma$	RFCNN $\Delta V$									FCNN $\Delta V$
	0.000 1	1	4	6	8	10	15	20	100	
0	2.889 7	2.828 7	2.183 0	0.087 4	0.052 3	0.076 4	0.083 9	0.095 1	0.095 1	0.259 8
2	2.864 9	2.828 7	2.228 6	0.087 4	0.052 3	0.076 4	0.083 9	0.142 2	0.142 2	0.259 8
4	2.864 9	2.961 9	2.274 4	0.087 4	0.052 3	0.112 8	0.083 9	0.095 1	0.095 1	0.245 3
6	2.864 9	2.961 9	2.301 3	0.120 6	0.068 1	0.112 8	0.083 9	0.095 1	0.095 1	0.237 7
8	2.864 9	2.822 9	2.347 8	0.120 6	0.068 1	0.112 8	0.083 9	0.095 1	0.095 1	0.238 0
10	2.864 9	2.822 9	2.348 2	0.120 6	0.068 1	0.112 8	0.083 9	0.095 1	0.095 1	0.254 0
12	2.864 9	2.822 9	2.348 2	0.1626 2	0.068 1	0.112 8	0.083 9	0.095 1	0.095 1	0.270 5
14	2.864 9	2.822 9	2.348 2	0.1626 2	0.068 1	0.112 8	0.083 9	0.095 1	0.095 1	0.287 4
16	2.864 9	2.822 9	2.207 9	0.1626 2	0.068 1	0.112 8	0.062 8	0.095 1	0.095 1	0.304 8
18	2.864 9	2.822 9	2.605 4	0.1467 8	0.107 4	0.112 8	0.062 8	0.095 1	0.095 1	0.322 7
20	2.864 9	2.822 9	2.605 4	0.1467 8	0.107 4	0.112 8	0.062 8	0.095 1	0.095 1	0.341 1

从表 2 可以看出,FCNN 算法聚类结果受例外点影响很大,随着例外点数目的增加,偏移量  $\Delta V$  越来越大.选择合适的参数  $\gamma$ ,RFCNN 算法受例外点影响则很较小.当  $\gamma \geq 6$ 时,RFCNN 算法均取得较好的聚类效果且偏移量  $\Delta V$  基本稳定在一个较小的范围内.

实验结果表明,当 $\gamma$ 取较大值时,较之文献[9]中的FCNN算法,RFCNN算法能够得到更理想的聚类结果.需要指出的是,当例外点数目很少时,FCNN算法聚类的偏移量 $\Delta V$ 可能比RFCNN算法聚类的偏移量 $\Delta V$ 要小(如实验1),但随着例外点数目的增加,选取合适参数 $\gamma$ ,RFCNN算法聚类效果明显优于FCNN算法聚类效果.

#### 4 结束语

针对模糊聚类神经网络FCNN对例外点敏感性的缺陷,通过引入Vapnik's  $\epsilon$ -不敏感损失函数,重新构造网络的目标函数,通过优化理论推导出新的具有更好鲁棒性的模糊聚类神经网络RFCNN,它有效地克服了FCNN对例外点敏感性的缺陷,并且能够得到合理的聚类中心.应当指出,尽管实验数据包含人造数据集和真实数据集,但例外点是人为加入的,我们目前正在将此方法应用于含脉冲噪声的图像处理中,限于篇幅,有关结果将另文表述.在未来的工作中,我们将对RFCNN算法做更深入的研究,并把本文的思想应用于建模等领域,以期有效地解决建模等领域的某些例外点敏感性问题.

致谢 在此,我们向对本文工作提出宝贵意见的评审专家表示衷心的感谢.

#### References:

- [1] Bezdek JC. Fuzzy models for pattern recognition: Methods that search for structures in data. New York: Institute of Electrical and Electronics Engineers, 1992.
- [2] Gabrys B, Bargiela A. General fuzzy min-max neural network for clustering and classification. IEEE Trans. on Neural Networks, 2000,11(3):769-783.
- [3] Krishnapuram R, Keller JM. A possibilistic approach to clustering. IEEE Trans. on Fuzzy Systems, 1993,1(2):98-110.
- [4] Xu JH, Zhang XG, Li YD. A nonlinear perceptron algorithm based on kernel functions. Chinese Journal of Computers, 2002, 25(7):689-695 (in Chinese with English abstract).
- [5] Chen JH, Chen CS. Fuzzy kernel perceptron. IEEE Trans. on Neural Networks, 2002,13(6):1364-1373.
- [6] Zhang L, Zhou WD, Jiao LC. Learning machine based on differential capacity control. Acta Electronica Sinica, 2003, 31(10): 1526-1531 (in Chinese with English abstract).
- [7] Wang ST. Fuzzy Systems, Fuzzy Neural Networks and Their Programming. Shanghai: Press of Shanghai Science and Technology, 1998 (in Chinese).
- [8] Kohonen T. Self organization and Associative Memory. Berlin: Springer-Verlag, 1989.
- [9] He PL, Hou YX. An asymmetric robust learning algorithm of fuzzy clustering neural networks. Journal of Computer Research and Development, 2001,38(3):296-301 (in Chinese with English abstract).
- [10] Shen HB, Wang ST, Wu XJ. Fuzzy kernel clustering with outliers. Journal of Software, 2004,15(7):1021-1029 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/15/1021.htm>
- [11] Keller A. Fuzzy clustering with outliers. In: Proc. of the NAFIPS 2000. 2000. 143-147.
- [12] Leski J. Towards a robust fuzzy clustering. Fuzzy Sets and Systems, 2003,137(2):215-233.
- [13] Vapnik V. Statistical Learning Theory. New York: John Wiley & Sons, 1998.
- [14] Krabs W. Optimization and approximation. New York: John Wiley & Sons, 1981.
- [15] Huber PJ. Robust Statistics. New York: John Wiley & Sons, 1981.
- [16] Steve RG. Support vector machines classification and regression. Technical Report, Department of Electronics and Computer Science, University of Southampton, 1998.

#### 附中文参考文献:

- [4] 许建华,张学工,李衍达.一种基于核函数的非线性感知器算法.计算机学报,2002,25(7):689-695.
- [6] 张莉,周伟达,焦李成.基于微分容量控制的学习机.电子学报,2003,31(10):1526-1531.
- [7] 王士同.模糊系统,模糊神经网络及应用程序设计.上海:上海科学技术文献出版社,1998.
- [9] 何丕廉,侯越先.模糊聚类神经网络的非对称性学习算法.计算机研究与发展,2001,38(3):296-301.
- [10] 沈红斌,王士同,吴小俊.离群模糊核聚类算法.软件学报,2004,15(7):1021-1029. <http://www.jos.org.cn/1000-9825/15/1021.htm>

附录. 定理证明

证明:如果  $M, N, U$  一定,结合式(7)、式(8),目标函数可表示为

$$J_{m\varepsilon} = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \frac{1}{\sigma^2} C_{ij} U_{ij} \sum_{k=1}^d |x_{jk} - v_{ik}|_{\varepsilon_{ik}} + \sum_{i=1}^M \sum_{k=1}^d \gamma \varepsilon_{ik} \tag{12}$$

$$= \sum_{i=1}^M \sum_{k=1}^d \left( \sum_{j=1}^N \frac{1}{2\sigma^2} C_{ij} U_{ij} |x_{jk} - v_{ik}|_{\varepsilon_{ik}} + \gamma \varepsilon_{ik} \right)$$

一般来说,不是所有的样本数据分量  $x_{jk}, j=1, \dots, N, k=1, \dots, d$  都满足不等式:

$$v_{ik} - x_{jk} \leq \varepsilon_{ik}, x_{jk} - v_{ik} \leq \varepsilon_{ik} \tag{13}$$

如果引入松弛变量  $\xi_{ikj}^+, \xi_{ikj}^- \geq 0$ ,对于所有的样本数据分量  $x_{jk}$  可满足下面的不等式

$$v_{ik} - x_{jk} \leq \varepsilon_{ik} + \xi_{ikj}^+, x_{jk} - v_{ik} \leq \varepsilon_{ik} + \xi_{ikj}^- \tag{14}$$

于是求式(12)最小可转化为求式(15)最小

$$J_{m\varepsilon} = \sum_{i=1}^M \sum_{k=1}^d \left( \sum_{j=1}^N \frac{1}{2\sigma^2} C_{ij} U_{ij} (\xi_{ikj}^+ + \xi_{ikj}^-) + \gamma \varepsilon_{ik} \right) \tag{15}$$

其中,约束条件为式(14)和  $\xi_{ikj}^+ \geq 0, \xi_{ikj}^- \geq 0$ .对含有上述约束条件的式(15)的拉格朗日函数为

$$\Phi_{m\varepsilon} = \sum_{i=1}^M \sum_{k=1}^d \left\{ \left[ \sum_{j=1}^N \frac{1}{2\sigma^2} C_{ij} U_{ij} (\xi_{ikj}^+ + \xi_{ikj}^-) + \gamma \varepsilon_{ik} \right] - \sum_{j=1}^N \lambda_{ikj}^+ (\varepsilon_{ik} + \xi_{ikj}^+ - v_{ik} + x_{jk}) - \sum_{j=1}^N \lambda_{ikj}^- (\varepsilon_{ik} + \xi_{ikj}^- + v_{ik} - x_{jk}) - \sum_{j=1}^N (\mu_{ikj}^+ \xi_{ikj}^+) - \sum_{j=1}^N (\mu_{ikj}^- \xi_{ikj}^-) - \beta_{ik} \varepsilon_{ik} \right\} \tag{16}$$

这里,  $\lambda_{ikj}^+, \lambda_{ikj}^-, \mu_{ikj}^+, \mu_{ikj}^-, \beta_{ik} \geq 0$  是拉格朗日乘子.问题求解的目标是拉格朗日函数式(16)相对于  $v_{ik}, \xi_{ikj}^+, \xi_{ikj}^-, \varepsilon_{ik}$  取最小,同时相对于拉格朗日乘子变量取最大.拉格朗日函数式(16)相对于  $v_{ik}, \xi_{ikj}^+, \xi_{ikj}^-, \varepsilon_{ik}$  最小时满足条件

$$\frac{\partial \Phi_{m\varepsilon}}{\partial v_{ik}} = \sum_{j=1}^N (\lambda_{ikj}^+ - \lambda_{ikj}^-) = 0 \tag{17}$$

$$\frac{\partial \Phi_{m\varepsilon}}{\partial \xi_{ikj}^+} = C_{ij} U_{ij} / 2\sigma^2 - \lambda_{ikj}^+ - \mu_{ikj}^+ = 0 \tag{18}$$

$$\frac{\partial \Phi_{m\varepsilon}}{\partial \xi_{ikj}^-} = C_{ij} U_{ij} / 2\sigma^2 - \lambda_{ikj}^- - \mu_{ikj}^- = 0 \tag{19}$$

$$\frac{\partial \Phi_{m\varepsilon}}{\partial \varepsilon_{ik}} = \gamma - \sum_{j=1}^N (\lambda_{ikj}^+ + \lambda_{ikj}^-) - \beta_{ik} = 0 \tag{20}$$

$\lambda_{ikj}^+, \lambda_{ikj}^-, \mu_{ikj}^+, \mu_{ikj}^- \geq 0$  和式(18)、式(19)表明,  $\lambda_{ikj}^+, \lambda_{ikj}^- \in [0, C_{ij} U_{ij} / 2\sigma^2]$ .由式(16)~式(20)可以得到

$$\min \Phi_{m\varepsilon} (v_{ik}, \xi_{ikj}^+, \xi_{ikj}^-, \varepsilon_{ik}) = \sum_{i=1}^M \sum_{k=1}^d \left[ - \sum_{j=1}^N (\lambda_{ikj}^+ - \lambda_{ikj}^-) x_{jk} \right] \tag{21}$$

由经典的拉格朗日对偶理论可知,

$$\max \sum_{i=1}^M \sum_{k=1}^d \left( - \sum_{j=1}^N (\lambda_{ikj}^+ - \lambda_{ikj}^-) x_{jk} \right) \tag{22}$$

$$\text{s.t.} \left\{ \begin{array}{l} \sum_{j=1}^N \lambda_{ikj}^+ - \sum_{j=1}^N \lambda_{ikj}^- = 0 \\ \sum_{j=1}^N (\lambda_{ikj}^+ + \lambda_{ikj}^-) \leq \gamma \\ \lambda_{ikj}^+, \lambda_{ikj}^- \in [0, C_{ij} U_{ij} / 2\sigma^2] \end{array} \right.$$

即为求解问题的所谓 Wolfe 对偶表达.令

$$g_{ik} = -\sum_{j=1}^N (\lambda_{ikj}^+ - \lambda_{ikj}^-) x_{jk} \tag{23}$$

式(22)可分解为式(24)的优化问题

$$\left. \begin{aligned} \max g_{ik} &= \max -\sum_{j=1}^N (\lambda_{ikj}^+ - \lambda_{ikj}^-) x_{jk} \\ \text{s.t.} \left\{ \begin{aligned} \sum_{j=1}^N \lambda_{ikj}^+ - \sum_{j=1}^N \lambda_{ikj}^- &= 0 \\ \sum_{j=1}^N (\lambda_{ikj}^+ + \lambda_{ikj}^-) &\leq \gamma \\ \lambda_{ikj}^+, \lambda_{ikj}^- &\in [0, C_{ij} U_{ij} / 2\sigma^2] \end{aligned} \right. \end{aligned} \right\} i = 1, \dots, M, k = 1, \dots, d \tag{24}$$

亦即

$$\left. \begin{aligned} \min \sum_{j=1}^N (\lambda_{ikj}^+ - \lambda_{ikj}^-) x_{jk} \\ \text{s.t.} \left\{ \begin{aligned} \sum_{j=1}^N \lambda_{ikj}^+ - \sum_{j=1}^N \lambda_{ikj}^- &= 0 \\ \sum_{j=1}^N (\lambda_{ikj}^+ + \lambda_{ikj}^-) &\leq \gamma \\ \lambda_{ikj}^+, \lambda_{ikj}^- &\in [0, C_{ij} U_{ij} / 2\sigma^2] \end{aligned} \right. \end{aligned} \right\} i = 1, \dots, M, k = 1, \dots, d \tag{25}$$

按照优化理论的 Kuhn-Tucker 定理,在鞍点,对偶变量与约束的乘积为 0,即

$$\left. \begin{aligned} \lambda_{ikj}^+ (\varepsilon_{ik} + \xi_{ikj}^+ - v_{ik} + x_{jk}) &= 0 \\ \lambda_{ikj}^- (\varepsilon_{ik} + \xi_{ikj}^- + v_{ik} - x_{jk}) &= 0 \\ \mu_{ikj}^+ \xi_{ikj}^+ &= 0 \\ \mu_{ikj}^- \xi_{ikj}^- &= 0 \\ \beta_{ik} \varepsilon_{ik} &= 0 \end{aligned} \right\} \tag{26}$$

又由式(18)、式(19)可得到

$$\left. \begin{aligned} \mu_{ikj}^+ &= C_{ij} U_{ij} / 2\sigma^2 - \lambda_{ikj}^+ \\ \mu_{ikj}^- &= C_{ij} U_{ij} / 2\sigma^2 - \lambda_{ikj}^- \end{aligned} \right\} \tag{27}$$

结合式(26)和式(27),进行如下推理:

当  $\lambda_{ikj}^+, \lambda_{ikj}^- \in (0, C_{ij} U_{ij} / 2\sigma^2)$  时,由式(27)可知,  $\mu_{ikj}^+, \mu_{ikj}^- \in (0, C_{ij} U_{ij} / 2\sigma^2)$ ,由式(26)中第 3,4 两式可知,此时  $\xi_{ikj}^+ = 0, \xi_{ikj}^- = 0$ ,再由式(26)第 1,2 两式可得:

$$\varepsilon_{ik} - v_{ik} + x_{jk} = 0, \varepsilon_{ik} + v_{ik} - x_{jk} = 0 \tag{28}$$

这样,可以根据任意样本分量  $x_{lk}, x_{mk}$ ,当  $\lambda_{ikl}^+, \lambda_{ikm}^- \in (0, C_{ij} U_{ij} / 2\sigma^2)$  时,由式(28)可得

$$\varepsilon_{ik} - v_{ik} + x_{lk} = 0, \varepsilon_{ik} + v_{ik} - x_{mk} = 0 \tag{29}$$

亦即

$$v_{ik} = (x_{lk} + x_{mk}) / 2, \varepsilon_{ik} = (x_{mk} - x_{lk}) / 2 \tag{30}$$

于是,聚类中心分量  $v_{ik}$  和不敏感性参数  $\varepsilon_{ik}$  的值可由式(30)确定.