第 28 卷 第 2 期
2002 年 3 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 28, No. 2
Mar., 2002

# 熵约束广义学习矢量量化神经网络和软竞争学习算法[1)]

张志华 郑南宁 张淮峰 于海霞

(西安交通大学人工智能与机器人研究所 西安 710049)

(E-mail: zhhaa@aiar.xjtu.edu.cn)

**摘 要** 结合广义学习矢量量化神经网络的思想和信息论中的极大熵原理，提出了一种熵约束广义学习矢量量化神经网络，利用梯度下降法导出其学习算法，该算法是软竞争格式的一种推广. 由于亏损因子和尺度函数被定义为同一个模糊隶属度函数，它可以有效地克服广义学习矢量量化网络的模糊算法存在的问题. 文中还给出熵约束广义学习矢量量化网络及其软竞争学习算法的许多重要性质，以此为依据，讨论拉格朗日乘子的选取规则.

**关键词** 学习矢量量化，极大熵原理，软竞争学习，拉格朗日乘子

**中图分类号** TP18

# ENTROPY-CONSTRAINED GENERALIZED LEARNING VECTOR QUANTIZATION NEURAL NETWORK AND SOFT COMPETITIVE LEARNING ALGORITHM

ZHANG Zhi-Hua ZHENG Nan-Ning ZHANG Huai-Feng YU Hai-Xia

(*Institute of AI & Robotics, Xi'an Jiaotong University, Xi'an* 710049)

(E-mail: zhhaa@aiar.xjtu.edu.cn)

**Abstract** According to the generalized learning vector quantization (GLVQ) network and the maximum-entropy principle, an entropy-constrained generalized learning vector quantization (ECGLVQ) neural network is proposed in this paper. A learning algorithm of the network, a generalization of the soft-competition scheme (SCS), is derived via the gradient descent method. Because the loss factor and the corresponding scaling function are defined as the same fuzzy membership function, it can overcome the problems for fuzzy algorithms of GLVQ network possess. Many important properties of the ECGLVQ network and its soft competitive learning algorithm are given. Thereby, the rule for choosing the Lagrangian multiplier is designed.

**Key words** Learning vector quantization, maximum-entropy principle, soft competitive learning, Lagrangian multiplier

1) 国家自然科学重点基金(69735010)和博士点专项基金(98069825)资助

收稿日期 1999-11-22 收修改稿日期 2000-07-06

# 1　引言

近十年来，由于 Kohonen 在学习矢量量化(Learning Vector Quantization，LVQ)算法和自组织特征映射(Self-Organization Feature Mapping，SOFM)网络两方面[1]的许多开创性的工作，把 LVQ 和神经网络相融合研究 LVQ 网络的工作变得十分流行. 但是由于传统的 LVQ 网络的训练规则基于胜者为王机制，其存在神经元未被充分利用以及输入样本和竞争单元之间的信息被浪费等两个主要问题[2]. Huntsberger 和 Ajjimarangsee[3]首次把 SOFM 同 LVQ 相结合，为试图克服上述问题提出了一种新的思路. 但同样也存在两大缺点[4]：一是哪些节点应被考虑；二是每个非获胜节点应怎样发挥影响. 其主要原因是由于这些算法的学习规则是通过启发式过程导出的.

自然，利用最优化过程来导出学习规则是克服由启发式过程带来的问题的一个有效途径. Pal 等首先据此思想提出了广义的学习矢量量化[5](Generalized Learning Vector Quantization，GLVQ)网络，Karayiannis 等修正了 GLVQ 并提出了它的模糊算法[6,7]——GLVQ-F，FALVQ1，FALVQ2 和 FALVQ3. 虽然其目标函数中的亏损因子被定义为模糊隶属度函数，但由此导出的学习规则中的尺度函数的性能却异常复杂，致使算法性能十分不稳定[8].

本文提出了一种新的研究途径. 主要思想是：首先利用 GLVQ 思想，通过极大熵原理构造一个聚类准则函数，利用最优化方法导出学习规则，将亏损因子和尺度函数均定义为隶属度函数. 因而可以克服 GLVQ-F 和 FALVQ 等存在的问题.

# 2　构造 GLVQ 网络的极大熵方法

记 $X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p\}\subset R^n$ 为输入样本集合，设 $X$ 中的类数(表现型的个数)为 $c$. GLVQ 网络的学习规则是从最优化一个目标函数导出的. 该目标函数定义为输入样本 $\boldsymbol{x}$ 的一个亏损函数 $L_x$

$$L_x = L(\boldsymbol{x};\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_c) = \sum_{r=1}^{c} u_r \| \boldsymbol{x} - \boldsymbol{v}_r \|^2 \tag{1}$$

其中，表现型是 $V=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_c\}\subset R^n$ 我们要寻找的矢量量化器，$u_r$ 是 $\boldsymbol{x}$ 相对表现型 $\boldsymbol{v}_r$ 的亏损因子，它不同的定义就会导出不同的学习算法. Karayiannis 等[6,7]把 $u_r=u_r(\boldsymbol{x})(r=1,2,\cdots,c)$ 定义为一个隶属度函数. 然而由于亏损因子和尺度函数之间的关系很复杂，不能保证尺度函数同时也是隶属度函数. 在实际计算中恰恰需要尺度函数为隶属度函数，这点在文献[8]关于 GLVQ-F 算法的性能分析中得到了验证.

将最优化问题(1)考虑为去寻找亏损因子 $u_r$ 的一个分布，在满足一定程度随机性的条件下来最小化目标函数 $L(\boldsymbol{x};\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_c)$. 其中令亏损因子 $u_r$ 属于单纯形集合

$$\Lambda = \left\{u \geqslant 0 : \sum_{r=1}^{c} u_r = 1\right\} \tag{2}$$

自然随机程度可以用信息论中的 Shannon 熵来度量，其值由下式给出

$$H(u) = -\sum_{j=1}^{c} u_j \ln u_j \tag{3}$$

于是利用拉格朗日方法，可以构造一个复合极小问题

$$\min_{u \in \Lambda}\{L_T(\boldsymbol{v},u) = L(\boldsymbol{x};\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_c) - TH(u)\} \tag{4}$$

其中，拉格朗日乘子 $T \in (0,\infty)$ 相当于两个目标函数的加权系数，当它趋于零时，式(4)将退化为式(1). 在实际应用中，总是希望函数 $L_T$ 不小于零，但由于 $-H(u)$ 是负的，这个条件不能够得到保证. 然而我们知道 $H(u) = -\sum_{j=1}^{c} u_j \ln u_j \leqslant \ln c$，且 $\min(-H(u))$ 同 $\min(\ln c - H(u))$ 是等价的. 于是可以把最优化问题(4)等价地变为

$$\min_{u \in \Lambda}\{L_T^*(\boldsymbol{v},u)\} = L(\boldsymbol{x};\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_c) + T(\ln c - H(u))\} \tag{5}$$

通过简单计算，即可得到该问题的一个解析解

$$u_j = u_j(\boldsymbol{x}) = \frac{\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{v}_j\|^2}{T}\right)}{\sum_{l=1}^{c}\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{v}_l\|^2}{T}\right)},\ \forall\ j \tag{6}$$

将其回代到问题(5)中 $L_T^*(\boldsymbol{v},u)$ 的，消去 $u$，得到下面函数

$$J_T(\boldsymbol{v}) = \min_u L_T^*(\boldsymbol{v},u) = T\left(\ln c - \ln\sum_{j=1}^{c}\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{v}_j\|^2}{T}\right)\right) \tag{7}$$

把式(6)代入式(1)，我们定义此时 $L(\boldsymbol{v},u)$ 为 $L(\boldsymbol{v})$.

这样就把问题(1)变为了一个目标函数为 $J_T(\boldsymbol{v})$ 的无约束优化问题，由此构造了一种新的 GLVQ 网络，由于它是基于极大熵原理得到的，我们称之为熵约束(Entropy-Constrained)广义学习矢量量化(ECGLVQ)网络. ECGLVQ 的主要思想是：首先把无约束最优化问题(1)通过极大熵原理转变为约束条件为(2)的约束最优化问题(5)，利用拉格朗日方法求出(5)关于亏损因子的解析解，将其回代到(5)得到一个新的关于表现型的无约束最优化问题(7). 同文献[6,7]中的 GLVQ 的模糊算法相比主要区别是：前者的亏损因子是一个约束最优化问题的解析解，后者的是事先根据某种规则定义的.

现在来看 $J_T(\boldsymbol{v})$ 和亏损函数 $L(\boldsymbol{v})$ 之间的关系，首先给出亏损因子 $u_j$ 的一个性质.

**定理 1.** 由式(6)定义的亏损因子 $u_j(j=1,2,\cdots,c)$ 具有下面性质

1) $\lim\limits_{T\to 0} u_j = \begin{cases} 1, & j = \arg\min\limits_i \|\boldsymbol{x}-\boldsymbol{v}_i\|^2 \\ 0, & \text{otherwise} \end{cases}$；

2) $\lim\limits_{T\to\infty} u_j = \frac{1}{c}(j=1,2,\cdots,c)$.

**定理 2.** 1) $J_T(\boldsymbol{v})$ 随 $T$ 的减少单调减少；

2) $\min\limits_j \|\boldsymbol{x}-\boldsymbol{v}_j\|^2 \leqslant L(\boldsymbol{v}) \leqslant J_T(\boldsymbol{v}) \leqslant \frac{1}{c}\sum_{j=1}^{c}\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$；

3) $J_T(\boldsymbol{v}) \leqslant \min\limits_j \|\boldsymbol{x}-\boldsymbol{v}_j\|^2 + T\ln c \leqslant L(\boldsymbol{v}) + T\ln c$；

4) $\lim\limits_{T\to 0} L(\boldsymbol{v}) = \lim\limits_{T\to 0} J_T(\boldsymbol{v}) = \min\limits_j \|\boldsymbol{x}-\boldsymbol{v}_j\|^2$；

5) $\lim\limits_{T\to\infty} L(\boldsymbol{v}) = \lim\limits_{T\to\infty} J_T(\boldsymbol{v}) = \frac{1}{c}\sum_{j=1}^{c}\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$.

定理 2 的证明略. 定理 1 和 2 是重要的结果. 它们表明当 $T\to\infty$ 时 $J_T(\boldsymbol{v})$ 的极限是 $\frac{1}{c}\sum_{j=1}^{c}\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$,这正是当均匀分配输入样本 $\boldsymbol{x}$ 给所有的表现型,即 $u_j=\frac{1}{c}(j=1,2,\cdots,c)$ 时函数 $L(\boldsymbol{v},u)$ 的值,这意味着聚类的随机程度最大;而当 $T\to 0$ 时 $\min\limits_j\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$ 是 $J_T(\boldsymbol{v})$ 和 $L(\boldsymbol{v})$ 的极限值,即完全分配 $\boldsymbol{x}$ 给与之最近的表现型,此时表现出聚类的非随机性,或所谓的"硬". 因而在聚类过程中,应先选取较大的 $T$ 值,然后逐渐降低它的值,使聚类的随机程度逐渐变小,以至于趋近非随机状态. 从最优化理论的角度来看,是用一列可微的函数 $J_T(\boldsymbol{v})$ 去逼近不可微的函数 $\min\limits_{\boldsymbol{v}}\ \min\limits_j\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$,进而用求 $J_T(\boldsymbol{v})$ 最小值得到的 $\boldsymbol{v}$ 来逼近求 $\min\limits_{\boldsymbol{v}}\ \min\limits_j\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$ 最小值得到的 $\boldsymbol{v}$. 下面定理是这个思想的保证.

**定理 3.** 设 $\boldsymbol{v}^{(T)}$ 满足 $J_T(\boldsymbol{v}^{(T)})=\min\limits_{\boldsymbol{v}}J_T(\boldsymbol{v})$,$\bar{\boldsymbol{v}}^{(T)}$ 满足 $L(\bar{\boldsymbol{v}}^{(T)})=\min\limits_{\boldsymbol{v}}L(\boldsymbol{v})$,则

$$\lim_{T\to 0}J_T(\boldsymbol{v}^{(T)})=\lim_{T\to 0}L(\bar{\boldsymbol{v}}^{(T)}).$$

**证明.** 根据定理 2 和定理 3 的已知条件,一方面有 $J_T(\boldsymbol{v}^{(T)})\leqslant J_T(\bar{\boldsymbol{v}}^{(T)})\leqslant L(\bar{\boldsymbol{v}}^{(T)})+T\ln c$;另一方面有 $L(\bar{\boldsymbol{v}}^{(T)},u)\leqslant L(\bar{\boldsymbol{v}}^{(T)})\leqslant J_T(\boldsymbol{v}^{(T)})$. 所以 $L(\bar{\boldsymbol{v}}^{(T)})\leqslant J_T(\boldsymbol{v}^{(T)})\leqslant L(\bar{\boldsymbol{v}}^{(T)})+T\ln c$,于是 $\lim\limits_{T\to 0}J_T(\boldsymbol{v}^{(T)})=\lim\limits_{T\to 0}L(\bar{\boldsymbol{v}}^{(T)})$.　　证毕.

## 3　熵约束 GLVQ 网络的软竞争学习算法

本节研究 ECGLVQ 网络的学习算法,也即考虑求 $J_T(\boldsymbol{v})$ 最小值的最优化方法. 我们采用最速下降法,由此网络的第 $t$ 次学习规则为

$$\boldsymbol{v}_j(t)=\boldsymbol{v}_j(t-1)-\eta(t-1)\frac{\partial J_T(\boldsymbol{v})}{\partial\boldsymbol{v}_j(t-1)}=\boldsymbol{v}_j(t-1)+2\eta(t-1)u_{jk}(\boldsymbol{x}_k-\boldsymbol{v}_j(t-1))\qquad(8)$$

其中,尺度函数 $u_{jk}=u_j(\boldsymbol{x}_k)$ 由式(6)定义. 下面给出算法的详细步骤.

步骤 1. 已知训练样本集 $X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_K\}\subset R^N$. 给定误差下限 $\varepsilon$, $\varepsilon_1$ 和最大迭代次数 Max$t$, Max$t1$; 初始化 $\boldsymbol{v}_j(0)=(v_{j1}(0),\cdots,v_{jn}(0))(j=1,2,\cdots,c)$ 和拉格朗日乘子 $T(0)$.

步骤 2. a) 修正拉格朗日乘子 $T(t)$,

b) 令 $t1=0$,

c) 对于固定的 $T(t)$,最小化 $J_{T(t)}(\boldsymbol{v})$,

1) For $k=1,2,\cdots,K$,

For $j=1,2,\cdots,c$,

按式(8)调整 $\boldsymbol{v}_j$, 此时式(8)中的 $t$ 改为 $t1$,

2) 若 $J_{T(t)}(\boldsymbol{v})>\varepsilon_1$ 且 $t1<\text{Max}t1$,令 $t1=t1+1$,转入步骤 c);否则,转入步骤 3.

步骤 3. 若 $J_{T(t)}(\boldsymbol{v})>\varepsilon$ 且 $T<\text{Max}t$,令 $t=t+1$,转入步骤 2;否则停机.

其中,$\eta(t)$ 是步长. 上述算法与 Yair 等[9]提出的软竞争格式(SCS)相似,其中的尺度函数 $u_{jk}$(文献[9]中记为 $p_{jk}$)都满足 Gibbs 分布. 而拉格朗日乘子 $T(t)$ 对应于 SCS 中的参数 $\beta(t)$ 的倒数. 如果我们令 Max$t1=1$,即在每一次迭代 $t$ 中,迭代次数 $t1$ 都只进行一次,此时学习算法就与 SCS 完全一致. 所以,这里称上面算法为 ECGLVQ 网络的软竞争学习算法(Soft

Competitive Learning Algorithm，SCLA）.

从上面分析可知，SCLA 中的亏损因子和相应的尺度函数都取相同的值，即都是由式(6)定义的 $u_j(j=1,2,\cdots,c)$. 而式(6)表明可以把 $u_j(j=1,2,\cdots,c)$看作一种模糊隶属度函数. 所以在 SCLA 中亏损因子是隶属度函数，同时尺度函数也是隶属度函数. 这正是上节分析 GLVQ 的模糊算法之后所期望的. 因而利用极大熵原理，构造了一个结合 GLVQ 网络和软竞争学习算法各自优点的途径. 下面进一步从理论上来讨论软竞争学习算法的性质. 首先根据定理 1，直接可得到下面的基本定理.

**定理 4.** 若中心向量 $\boldsymbol{v}_l(l=1,2,\cdots,c)$的学习规则由式(8)定义，则

1) $\lim\limits_{T(t)\to 0}\boldsymbol{v}_j(t)=\begin{cases}\boldsymbol{v}_j(t-1)+2\eta(t)(\boldsymbol{x}_k-\boldsymbol{v}_j(t-1)), & j=\arg\min\limits_r\|\boldsymbol{x}_k-\boldsymbol{v}_r(t-1)\|,\\ \boldsymbol{v}_j(t-1), & \text{其它},\end{cases}$

2) $\lim\limits_{T(t)\to +\infty}\boldsymbol{v}_j(t)=\boldsymbol{v}_j(t-1)+2\,\dfrac{\eta(t)}{c}(\boldsymbol{x}_k-\boldsymbol{v}_j(t-1)),\ j=1,2,\cdots,c.$

固定拉格朗日乘子 $T$，来考虑尺度函数之间的关系，下面定理是显然的.

**定理 5.** $\|\boldsymbol{x}-\boldsymbol{v}_j\|^2\leqslant\|\boldsymbol{x}-\boldsymbol{v}_l\|^2$，$\forall j,l\in\{1,2,\cdots,c\}$，则 $1\geqslant u_j\geqslant u_l\geqslant 0$.

再从另一个角度分析尺度函数的性能，$\forall j\in\{1,2,\cdots,c\}$，固定 $\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$，把由式(6)定义的函数 $u_j$ 当作拉格朗日乘子 $T$ 的函数进行分析. 先计算 $u_j$ 关于 $T$ 的导数. 对式(6)两边求对数，有

$$\ln u_j=-\frac{\|\boldsymbol{x}-\boldsymbol{v}_j\|^2}{T}-\ln\sum_{l=1}^{c}\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{v}_l\|^2}{T}\right),$$

对于上式两边关于 $T$ 求导数，可以获得$\dfrac{\mathrm{d}u_j}{\mathrm{d}T}\cdot\dfrac{1}{u_j}=\dfrac{\|\boldsymbol{x}-\boldsymbol{v}_j\|^2}{T^2}-\sum\limits_{l=1}^{c}u_l\dfrac{\|\boldsymbol{x}-\boldsymbol{v}_l\|^2}{T^2}$，于是

$$\frac{\mathrm{d}u_j}{\mathrm{d}T}=u_j\frac{\|\boldsymbol{x}-\boldsymbol{v}_j\|^2}{T^2}-u_j\sum_{l=1}^{c}u_l\frac{\|\boldsymbol{x}-\boldsymbol{v}_l\|^2}{T^2}=u_j\left(\sum_{l=1}^{c}u_l\left(\frac{\|\boldsymbol{x}-\boldsymbol{v}_j\|^2}{T^2}-\frac{\|\boldsymbol{x}-\boldsymbol{v}_l\|^2}{T^2}\right)\right),$$
$$j=1,2,\cdots,c\ .$$

根据上式，容易得到下面结论.

**定理 6.** 令 $l=\arg\min\limits_j\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$，$i=\arg\max\limits_j\|\boldsymbol{x}-\boldsymbol{v}_j\|^2$，则 $u_l$ 为 $T$ 在区间$(0,\infty)$的递减函数，而 $u_i$ 为 $T$ 在区间$(0,\infty)$的递增函数.

# 4 数值实验及结果分析

这里用著名的 IRIS 数据[10]来验证本文提出的 ECGLVQ 网络的软竞争学习算法的性能. IRIS 数据经常被用来检验聚类(无监督)或分类(有监督)模式识别方法的性能，其共有 150 个数据，每个数据又由 4 个分量组成，IRIS 数据分为 3 类，每类有 50 个数据. 对有监督设计标准的错误个数为 0～5，而无监督设计标准的个数在 15 左右.

表 1 给出了下面实验中采用的表现型的初始值. 我们分别使用了 1-$NP$(最近邻)算法，传统的硬 LVQ 网络，SCS 和本文提出 SCLA 算法进行了实验比较，实验结果分别见表 1 和表 2. 在实验中，我们定义拉格朗日乘子 $T(t)=\dfrac{1}{\beta(t)}$，其中 $\beta(t)$的选取规则为

$$\beta(t)=\beta_0+\frac{\beta_{\max t}-\beta_0}{\max t}t,$$

其中，$\max t=1\ 000$ 为最大迭代次数，$t$ 代表当前迭代次数；$\beta_{\max t}=10.0$ 和 $\beta_0=0.5$ 分别是当 $t$ 取最大迭代次数 $\max t$ 和 0 时 $\beta$ 的值.

表 1　数值实验中表现型的初始值及最近邻算法实验结果

| 表现型的初始值 | 表现型的最终值 | 混淆矩阵 | | | 错误率(%) |
|---|---|---|---|---|---|
| 5.006, 3.428, 1.462, 0.246 | 5.006, 3.428, 1.462, 0.246 | 50 | 0 | 0 | 7.3 |
| 5.936, 2.770, 4.260, 1.326 | 5.936, 2.770, 4.260, 1.326 | 0 | 46 | 4 | |
| 6.588, 2.974, 5.552, 2.026 | 6.588, 2.974, 5.552, 2.026 | 0 | 7 | 43 | |

表 2　硬 LVQ，SCS 和本文的 SCLA 三种算法的实验结果

| 算法模型 | 表现型的最终值 | 混淆矩阵 | | | 错误率(%) |
|---|---|---|---|---|---|
| 硬 LVQ 算法 | 5.005 773, 3.427 690, 1.462 131, 0.246 190 | 50 | 0 | 0 | 10.7 |
| | 5.900 532, 2.748 040, 4.397 150, 1.438 331 | 0 | 48 | 2 | |
| | 6.905 461, 3.074 274, 5.741 103, 2.072 295 | 0 | 14 | 36 | |
| Yair 等的 SCS | 5.005 773, 3.427 689, 1.462 133, 0.246 191 | 50 | 0 | 0 | 11.3 |
| | 5.892 710, 2.744 258, 4.398 185, 1.437 611 | 0 | 48 | 3 | |
| | 6.908 681, 3.077 306, 5.727 011, 2.067 521 | 0 | 14 | 36 | |
| 本文的 SCLA | 5.006 088, 3.427 901, 1.462 023, 0.246 169 | 50 | 0 | 0 | 4.0 |
| | 5.925 553, 2.836 748, 4.503 774, 1.463 558 | 0 | 46 | 4 | |
| | 5.900 000, 3.000 000, 5.100 000, 1.800 000 | 0 | 2 | 48 | |

从表 2 可以看到，硬 LVQ 算法和 Yair 等的软竞争格式的聚类效果都不是很好，文献[10]分析了软竞争格式对 IRIS 数据的聚类性能不佳的原因. 然而由于 SCLA 在拉格朗日乘子取每一个具体的值时，都要关于此时对应的目标函数求最小值而得到相应的表现型的值，然后，降低拉格朗日乘子的值，再以前面得到的表现型的值作为初始的表现型迭代最小化这时的目标函数来求表现型的一组新值，如此重复. 这可以有效地改善软竞争格式的全局收敛性，从而增强算法的聚类能力.

## 5　结束语

ECGLVQ 网络是在 GLVQ 网络的思想上，利用极大熵原理构造的，而它的基于梯度下降法导出的学习算法却是 Yair 等提出的 SCS 的一种推广，GLVQ 和 SCS 两者都是为克服传统的 LVQ 算法存在的问题提出的，所以 ECGLVQ 为克服 LVQ 的问题设计了一个新的更为有效的思路.

ECGLVQ 是一个目标函数由式(5)定义和满足约束条件式(2)的约束最优化问题，而式(5)是由式(1)加($\ln c-H(u)$)组成的. 不难发现式(1)定义的是类内距离和，它有一个单调趋势，当 $c=N$ 时(也就是一个样本为一类)它有一个最小值零. 因而，若是在聚类中，事先并不知道类数，要自动确定类数 $c$，GLVQ 显然是无效的. 另一方面，我们知道($\ln c-H(u)$)在所有的表现型取相同的值时达到最小值零，这实际上蕴涵着 $c=1$，即所有样本为一类. 因而式(1)的最小化和($\ln c-H(u)$)的最小化相对于 $c$ 是矛盾，若我们选取一个合适的 $T$ 值，可以克服 GLVQ 中存在的所谓数据“过拟合”(Overfitting)问题. ECGLVQ 的这个重要特性是今后需要进一步深入研究的.

## 参 考 文 献


1 Kohonen T. Self-Organization Maps. Berlin: Springer-Verlag, 1995

2 Chung F L, Lee T. Fuzzy competitive learning. *Neural Networks*, 1994, **7**(3):539～551

3 Huntsberger T, Ajjimarangsee P. Parrallel self-organizing feature maps for unsupervised pattern recognition. *Int. J. General Systems*, 1989, **16**(2):357～372

4 Tsao E C K, Bezdek J C, Pal N R. Fuzzy Kohonen clustering networks. *Pattern Recognition*, 1994, **27**(5):757～764

5 Pal N R, Bezdek J C, Tsao E C K. Generalized clustering networks and Kohonen's self-organizing scheme. *IEEE Trans. Neural Networks*, 1993, **4**(4):549～558

6 Karayiannis N B, Bezdek J C, Pal N R, Hatharay R J. Repair to GLVQ: A new family of competitive learning schemes. *IEEE Trans. Neural Networks*, 1996, **7**(5):1062～1071

7 Karayiannis N B, Pai P I. Fuzzy algorithms for learning vector quantization. *IEEE Trans. Neural Networks*, 1996, **7**(5):1196～1211

8 张志华，郑南宁，王天树. 广义LVQ神经网络的性能分析及其改进. 自动化学报，1999，**25**(5):582～588

9 Yair E, Zeger K, Gersho A. Competitive learning and soft competition for vector quantizer design. *IEEE Trans. Signal Processing*, 1992, **40**(2):294～308

10 Bezdek J C, Pal N R. Two soft relatives of learning vector quantization. *Neural Networks*, 1995, **8**(5):729～743



**张志华** 讲师，现在西安交通大学攻读博士学位. 主要研究领域为模式识别、图像处理、及神经计算.

**郑南宁** 博士，中国工程院院士. 主要研究领域为模式识别、图像处理及计算机视觉的理论与应用.