第 8 卷第 4 期
1997 年 4 月

软 件 学 报
JOURNAL OF SOFTWARE

Vol. 8, No. 4
Apr. 1997

多层反馈神经网络的 FP 学习和综合算法*

张 铃

（安徽大学人工智能研究所 合肥 230039）

（智能技术与系统国家重点实验室 北京 100084）

张 钹

（清华大学计算机系 北京 100084）

摘要 本文给出多层反馈神经网络的 FP 学习和综合算法，并讨论此类网络的性质，指出将它应用于聚类分析能给出不粒度的聚类，且具有收敛速度快（是样本个数的线性函数）、算法计算量少（是样本个数和输入、输出维数的双线性函数）、网络元件个数少、权系数简单（只取 3 个值）、网络容易硬件实现等优点. 作为聚类器的神经网络的学习和综合问题已得到较圆满地解决.

关 键 词 多层反馈神经网络，学习算法，聚类.

中图分类号 TP18

众所周知，从应用观点来看，神经网络的最主要功能为：①作为联想记忆器，②作为分类器，③用于求优化的解. 但至今尚无有效地、快速地使所得到的网络具有良好性能的神经网络的学习算法. 我们在文献[1]中讨论了多层前馈网络的 FP 学习和综合算法，指出用此法得到的网络是一个极好的通用联想记忆器. 但是当给定具体的样本集时，按 FP 算法得到的网络的元件个数往往有过多的隐层元，于是对一给定的样本集的情况下，如何求性能好、规模小的网络的问题仍需研究.

本文将转来讨论多层反馈神经网络的 FP 学习和综合算法，将证明所得到的网络作为分类器（聚类分析器）. 具有计算量少、收敛速度快、元件个数少、网络性能好等优点.

1 作为聚类分析器的神经网络

1.1 问题的提出

聚类分析问题：设在 N 维空间中，给定一有限点集 $K=\{x^0,x^1,x^2,...,x^p\}$，再给出一个量度相似性的函数 $D(x,y)$ 和一个相似度阈值 d. 要求将 K 进行分类并满足下面 2 个条件：

A. 将 K 分成子类 $\{K_1,K_2,...,K_m\}$，要求这些子类构成 K 的一个划分.

B. 若 $D(x,y)>d$，$x,y\in K$，则 x 和 y 属于同一类.

为了使神经网络能实现聚类的功能，将 K 中的各元素作为一组输入，即将 K 作为样本

* 本文研究部分得到国家攀登计划基金和国家自然科学基金资助. 作者张铃，1937 年生，教授，主要研究领域为人工智能理论，人工神经网络理论，应用数学. 张钹，1935 年生，教授，博士导师，中国科学院院士，主要研究领域为人工智能，计算机应用技术.

本文通讯联系人：张铃，合肥 230039，安徽大学人工智能研究所

本文 1996-04-16 收到修改稿

集，然后对网络进行学习，使经过学习后的网络，能完成分类的功能，即分在同一类中元素有相同的输出，并满足聚类的 2 个条件.

对于神经网络，通常我们取 $D(x,y)=\langle x,y\rangle$ 为相似性函数，($\langle x,y\rangle$ 表 x 与 y 的内积).

1.2　FP 算法简介

我们将利用文献[1]中提出的 FP 算法来解上述问题，下面简单介绍一下 FP 算法的一些本文将用到的性质.

作 p 个 $(p-1)$ 维向量如：$z^0=(-1,-1,\ldots,-1)^T$，$z^1=(+1,-1,\ldots,-1)^T$，$\ldots$，$z^{p-1}=(-1,-1,\ldots,+1)^T$. y，z 向量各分量取值 $\{-1,1\}$.

性质 1. 任给 p 个 m 维输出向量 $\{y^0,y^1,\ldots,y^{p-1}\}$，则存在 $m\times(p-1)$ 维权矩阵 U 和阈值向量 ξ：

即
$$u_j^i=\frac{y_i^j-y_i^0}{2}\qquad i=1,2,\ldots,m;\ j=1,2,\ldots,p-1. \tag{10}$$
$$\xi_i=-(y_i^0+u_1^i+\ldots+u_{p-1}^i)$$

使得
$$y^j=G(z^j),\qquad j=0,1,\ldots,p-1.$$

即
$$y^j=Sgn(U*z^j-\xi_j)$$

2　多层反馈神经网络的 FP 学习和综合算法

2.1　总的思路

(1)利用内积作为相似性函数；(2)利用 FP 学习算法对待分类的集合进行预分类；(3)利用反馈网络的反馈作用完成最后的分类.

2.2. 网络线路的设计

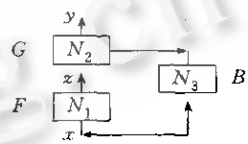

图1

整个网络线路如图 1. 其中第 1 元件层 N_1，有 p 个元件，其权矩阵和阈值向量分别为 $W(1)$、$\theta(1)$，其对应的关系为 F. 第 2 元件层 N_2，有 m 个元件，其权矩阵和阈值向量分别为 $W(2)$、$\theta(2)$，其对应的关系为 G. 反馈元件层 N_3，有 n 个元件，其权矩阵和阈值向量分别为 $W(3)$、$\theta(3)$，其对应的关系为 B.

在本文中设神经元 A 有 n 个输入和一个输出，其对应关系为

$$y=Sgn(x)\qquad\qquad Sgn(x)=\begin{cases}+1, & x>0,\\ -1, & x\leqslant 0\end{cases}$$

2.3　各层元件的学习和综合

我们按照 FP 算法的原则，从后向前逐层确定各层元件的个数和各元件的权和阈值.

A. 第 1 元件层的学习和综合

设样本集为 $K=\{x^0,x^1,\ldots,x^p\}$，x^i 是 n 维单位向量(一般可设 x^i 具有相同的模)

以 x 为输入向量，其对应的输出向量为 z(z 是 p 维向量，其分量取值于 $\{-1,1\}$).

$$z=F(x)$$

令
$$d_i=\max\{\langle x^i,x^j\rangle,i\neq j\},i=0,1,\ldots,p.$$

取 $D(x,y)=\langle x,y\rangle$，$\forall\ x,y\in K$，为相似性函数

给定相似性阈值为 d($d<1$，显然，当 $d\geqslant 1$ 时，所有样本都被分成不同的类).

作权矩阵 $W^i(1)=(x^i)^T, i=1,2,\ldots,p.$

$$\theta_i(1)=d, \qquad i=1,2,\ldots,p. \tag{1}$$

则有下面命题成立.

命题 1. 设 F 是第 1 元件层 N_1 的对应关系,按(1)式定义的权矩阵和阈值,则有

(i)若 $\langle x^i,x^j\rangle>d$,则 $z^j=F(x^j)$ 和 $z^i=F(x^i)$ 的第 i,第 j 分量均为 1.

(ii)$\forall\ i$,均有 $z_i^i=1$

(iii)若 $d_i<d$,则 $z^i=F(x^i)=(-1,-1,\ldots,-1,+1,-1,\ldots,-1)$,即 z 的第 i 个分量为 $+1$,其余分量为 -1.

证明:由 $\langle x^i,x^j\rangle>d$,及(1)得 $\langle x^i,x^j\rangle>d=\theta_i(1)$,

得 $$z_i^j=F_i(x^j)=Sgn(\langle w^i,x^j\rangle-\theta_i(1))=Sgn(\langle x^i,x^j\rangle-\theta_i(1))=1,$$

同理得 $$z_j^i=F_j(x^i)=Sgn(\langle w^j,x^i\rangle-\theta_i(1))=1.$$

故(i)式得证.

显然,对一切 i 有 $z_i^i=F_i(x^i)=Sgn(\langle w^i,x^i\rangle-\theta_i(1))=Sgn(1-\theta_i(1))=1.$

故(ii)式得证.

若 $d_i<d$,对任意 $j\neq i$,有 $\langle x^i,x^j\rangle\leqslant d_i<d=\theta_i(1)$

故有 $$z_j^i=F_j(x^i)=Sgn(\langle w^j,x^i\rangle-\theta_i)=-1.$$

(iii)式得证.

命题 1 说明:若 x^i 与样本中其它的点的内积均小于 d,则 $z^i=F(x^i)=(-1,-1,\ldots,+1,-1,\ldots,-1)$,即 z 的第 i 个分量为 $+1$,其余分量为 -1.

若 x^i 与样本中某些的点的内积大于 d,则 $z^i=F(x^i)$ 的分量至少有 2 个分量为 1,故第 1 元件层完成了区别各样本的相似性是否大于 d 的作用.

B. 第 2 元件层的学习和综合

设第 2 层有 p 个元件,其输入向量为 z,输出向量为 y,其对应关系为 G.

作权矩阵 $$w_j^i(2)=\begin{cases}-1, & 当\ j<i.\\ 1, & 当\ j=i.\\ 0, & 当\ j>i.\end{cases}$$

阈值向量 $$\theta_i(2)=(i-1) \tag{2}$$

则有如下命题

命题 2. 按(2)式定义第 2 层的权矩阵和阈值向量,则第 2 元件层的对应关系 G 满足如下性质:

令 $e^0=(-1,-1,\ldots,-1)^T, e^1=(+1,-1,\ldots,-1)^T,\ldots,e^p=(-1,-1,\ldots,+1)^T.$

(i)若 z^i 至少有 2 个分量为 1,设 z^i 第 1 个为 1 分量的下标为 $k(i)$,则有 $G(z^i)=e^{k(i)}$

(ii)若 z^i 只有第 i 个分量为 1,则有 $G(z^i)=e^i$

(iii)若 z^i 的所有分量均为 -1,则有 $G(z^i)=e^0$

证明:设 z^i 的第 $k(i)$ $(k(i)<i)$ 个分量为 1,由 $W(2)$ 的定义知:

当 $t=i$ 时,$\sum_j w_j^i(2)z_j^i-\theta_i(2)\leqslant(k(i)-1)-1+(i-k(i)-1)+1-(i-1)=\ 1,$

即 $y_i^i=-1.$

当 $t>i$ 时,$\sum w_j^t(2)z_j^i-\theta_t(2)\leqslant(k(i)-1)-1+(t-k(i))-(t-1)=-1,$

即 $y_t^i=-1$.

当 $t=k(i)$ 时，$\Sigma w_j^t(2)z_j^i-\theta_t(2)\geqslant(k(i)-1)+1-(k(i)-1)=1$,

即 $y_{k(i)}^i=1$.

当 $t<k(i)$ 时，$\Sigma w_j^t(2)z_j^i-\theta_t(2)=(t-1)-1-(t-1)=-1$,

即 $y_t^i=-1$.

当 $k(i)<t<i$ 时，$\Sigma w_j^t(2)z_j^i-\theta_t(2)\leqslant(k(i)-1)-1+(t-k(i))-(t-1)=-1$,

即 $y_t^i=-1$.

综合上述得 $G(z^i)=e^{k(i)}$

(ii)、(iii)类似可证.

由命题 2 可知，任给 2 个输入，若其相似度大于相似度阈值 d，则经过两层元件的映射均被映射到同一类的输出，故聚类分析的第 2 个条件已经基本满足.

即设固定 x^i，设与 x^i 内积大于 d 的 x^j 的最小上标为 $k(i)$，则有

$$G(F(x^i))=(-1,-1,\ldots,+1,\ldots,-1)=e^{k(i)}$$

即　　$G(F(x^i))$ 的第 $(k(i))$ 个分量为 $+1$，其余分量均为 -1.

另一方面，我们知道，给定集合的一个划分就给出一个等价关系，而等价关系必须满足传递律，故单单 N_1、N_2 两个元件层还不能完成聚类分析的任务. 为此我们再加上一个反馈层，利用反馈层来完成等价关系的传递律.

C. 反馈层的学习和综合

设反馈层有 n 个元件，其输入向量为 y，输出向量为 x，即反馈到 N_1 层的输入.

由命题 2 知，当输入 $\{x^0,x^1,\ldots,x^p\}$ 时，其对应的输出 $\{y^0,y^1,\ldots,y^p\}$. 则 $\{y^0,y^1,\ldots,y^p\}$ 是 $\{e^0,e^1,\ldots,e^p\}$ 的子集.

由性质 1，以 $\{y^0,y^1,\ldots,y^p\}$ 和 $\{x^0,x^1,\ldots,x^p\}$ 为 N_3 层的输入、输出. 并且有

$$x^i=B(e^i),i=0,1,\ldots,p.$$

的权、阈值的网络是存在的，下面证明这样网络即可完成等价关系的传递律.

命题 3. 以 y 和 x 为 N_3 层的输入和输出，作权矩阵和阈值向量如下

$$w_j^i(3)=\frac{x_i^j-x_i^0}{2}\qquad\theta_i(3)=-(x_i^0+\sum_j w_j^i(3)).\tag{3}$$

则有　　$x^i=B(e^i),\ i=1,2,\ldots,n;\ j=1,2,\ldots,p.$

证明：由性质 1 直接得.

2.4　网络性能的分析

本节将证明按照命题 1～3 构造的网络能完成聚类的任务.

当给定相似度阈值 d，称满足聚类分析条件(A)、(B)的聚类分析为相似度为 d 的聚类分析.

先介绍称之为最大树聚类分析方法(见文献[2]).

设 N 维空间有一点集 $K=\{x^0,x^1,\ldots,x^p\}$，设 x^i 与 x^j 的相似度为 $d_{ij}(d_{ij}=d_{ji})$，给定 d，则相似度为 d 的聚类可由下面方法完成.

设所有大于 d 的 d_{ij} 按大小，从大到小排列，记为 $\{d_{ij}(1),d_{ij}(2),\ldots,d_{ij}(k)\}$. 依次将以对应于 $d_{ij}(t)$ 的 x^i,x^j 用边相连，若到某步在其之前没有出现回路，而连上边 e 时就出现回

路，则跳过 e 边不相连，再继续上述步骤，直至所有 $d_{ij}(t)$ 都连完为止. 这样得到的图中的每个连通成分即是分类的一个子类.

下面证明我们的网络能完成上述的算法.

定理. 给定样本集 $K=\{x^0,x^1,...,x^p\}$ 及相似度阈值 d，若按公式(1)，(2)，(3)公式分别定义 N_1、N_2、N_3 各层的权和阈值，则图 1 网络构成相似度为 d 的聚类器.

证明：这只要证明同一子类的元素当网络运行达到稳定状态时，其对应的输出是相同的，反之，不同子类的元素，其对应的输出是不同的.

不妨设子类 $K_1=\{x^{i_1},x^{i_2},...,x^{i_t}\}$，(设 $i_1\leqslant i_2\leqslant...\leqslant i_t$). 下面证明至多经过 $(p-1)$ 步，K_1 中的元素对应的输出均相同(即对应于 x^{i_1} 的输出).

由命题 1、2 得：经过 N_1、N_2 元件层运行一步后，$x^{i_1},x^{i_2},...,x^{i_t}$ 的不同输出至多只有 $(t-1)$ 个(因为 x^{i_t} 的输出必与某个 x^{i_h}，$(t>h)$ 的输出相同). 再经过反馈层将子类对应的 y 又映到 X 对应的子类的某个子集上. 这样经过一个循环，子类元素的不同输出个数至少减少一个，故至多经过 $(t-1)$ 步，K_1 必被映到同一输出上. 之后，K_1 子类就稳定在这个状态上.

再由，最大树聚类法所分的各连通子类是互不相同的，故不同的子类被映到不同的输出上. 因为不然的话，设有两子类 K_1，K_2 被映到某同一输出 y 上，设到某步 K_1 中的 x^{1_i} 和 K_2 中的 x^{2_j} 被映到同一输出 y 上，这只有 2 种可能，一是在通过 N_1 和 N_2 层时，x^{1_i} 和 x^{2_j} 被映到同一输出上，这只有当 2 个输入的相似度大于 d 时才发生，但是因为 K_1 与 K_2 是不连通的，故得 x^{1_i} 与 x^{2_j} 的相似度必小于 d，矛盾. 那么一定是在通过反馈层时 2 个输入被映到同一输出上，即 2 个输入的为 1 的分量的最小下标是相同的，设为 k，则 x^{1_i} 与 x^k 是连通的，同理 x^{2_j} 与 x^k 是连通的，换句话说，x^{1_i} 与 x^{2_j} 是连通的，矛盾. 定理得证.

下面举一个简单的例子.

设给定样本集 $K=\{x^1,x^2,...,x^6\}$，

$x^1=(+1,+1,+1,+1,+1,+1)$　　$x^4=(-1,+1,-1,-1,-1,+1)$

$x^2=(+1,-1,-1,+1,+1,+1)$　　$x^5=(+1,-1,-1,+1,+1,-1)$

$x^3=(-1,-1,+1,+1,-1,+1)$　　$x^6=(-1,-1,-1,-1,-1,-1)$

容易知道：取相似度阈值 d 不同，可得到不同的分类.

$d=6,5,4$ 时有分类：$\{1,2,3,4,5,6,\}$　　$d=1,0$ 时有分类：$\{(1,2,5),(3),(4,6)\}$

$d=3,2$ 时有分类：$\{1,(2,5),3,4,6\}$　　$d=-1$ 时有分类：$\{(1,2,3,4,5,6)\}$

例取 $d=1$，

$F(x^1)=(+1,+1,-1,-1,-1,-1)\to G(F(x^1))=(+1,-1,-1,-1,-1,-1)=e^1$
$\to B(G(F(x^1)))=x^1\to F(x^1)=(+1,+1,-1,-1,+1,-1)\to G(F(x^1))=e^1$.

$F(x^2)=(+1,+1,-1,-1,+1,-1)\to G(F(x^2))=(+1,-1,-1,-1,-1,-1)=e^1$
$\to B(G(F(x^2)))=x^1\to F(x^1)=(+1,+1,-1,-1,-1,-1)\to G(F(x^1))=e^1$.

$F(x^3)=(-1,-1,+1,-1,-1,-1)\to G(F(x^3))=(-1,-1,+1,-1,-1,-1)=e^3$
$\to B(G(F(x^3)))=x^3\to F(x^3)=(-1,-1,+1,-1,-1,-1)\to G(F(x^3))=e^3$.

$F(x^4)=(-1,-1,-1,+1,-1,-1)\to G(F(x^4))=(-1,-1,-1,+1,-1,-1)=e^4$
$\to B(G(F(x^4)))=x^4\to F(x^4)=(-1,-1,-1,+1,-1,-1)\to G(F(x^4))=e^4$.

$F(x^5)=(-1,+1,-1,-1,+1,-1)\to G(F(x^5))=(-1,+1,-1,-1,-1,-1)=e^2$
$\to B(G(F(x^5)))=x^2\to F(x^2)=(+1,+1,-1,-1,-1,-1)\to G(F(x^2))=e^1$.
$F(x^6)=(-1,-1,-1,+1,-1,+1)\to G(F(x^6))=(-1,-1,-1,+1,-1,-1)=e^4$
$\to B(G(F(x^6)))=x^4\to F(x^4)=(-1,-1,-1,+1,-1,-1)\to G(F(x^4))=e^4$.

由上面得，网络经过一次循环后，将样本集予分为{(1,2),(3),(4,6),(5)}，再经过一次循环，将样本集分成为{(1,2,5),(3),(4,6)}. 这时网络就稳定了. 即网络完成了相似度为 $d=1$ 的聚类分析.

3　多层反馈网络的功能

上面已经讲过，多层反馈网络可以作为聚类器. 下面我们将指出这种网络不但有普通的聚类分析能力，而且有一定的容错能力.

为简单计，设 n 维输入向量的分量只取 $-1,+1$ 两个值.

再设样本集 $K=\{x^1,x^2,\ldots,x^p\}$，给定相似度阈值 d. 定义：

$$g_i=\max\{\langle x^i,x^j\rangle \mid \langle x^i,x^j\rangle\leqslant d, i\neq j\},\quad i=1,2,\ldots,p.$$
$$h_i=\max\{d,(n-(d-g_i))\},\quad i=1,2,\ldots,p.$$
$$D(x^i)=\{x \mid \langle x,x^i\rangle>h_i\},\quad i=1,2,\ldots,p.$$

命题 4. 图 1 网络按上述所给的 FP 学习和综合算法得到的神经网络，不但能进行所给相似度为 d 的聚类分析，而且有如下的容错能力：

当 $x\in D(x^i)$，则 x 与 x^i 必属于同一分类.

证明：先给出一个不等式：

对任意 n 维向量 x,y,z，其模平方为 n，则有下式成立

$$\langle x,z\rangle\leqslant n+\langle y,z\rangle-\langle x,y\rangle \tag{4}$$

设 $x\in D(x^i)$，由(4)式，当 $\langle x^j,x^i\rangle\leqslant d$ 时得

$$\langle x,x^j\rangle\leqslant n+\langle x^i,x^j\rangle-\langle x,x^i\rangle\leqslant n+g_i-\langle x,x^i\rangle<n+g_i-(n-(d-g_i))=d$$

即
$$F_j(x)=F_j(x^i)=-1,\ j\neq i$$

又当 $x\in D(x^i)$，有
$$\langle x,x^i\rangle>h_i=\max\{d,(n-(d-g_i))\}\geqslant d$$

得
$$F_i(x)=F_i(x^i)=+1,$$

对于 $\langle x^i,x^j\rangle>d, j\neq i$ 时，$F_j(x)=+1$ 或 -1，若有某 j 使 $F_j(x)$ 为 $+1$，设其坐标最小者为 $k(x)$，显然 $G(F(x))=e^{k(x)}$. 而 $x^{k(x)}$ 与 x^i 是属于同一类的，故得 x 与 x^i 是属于同一类的. 若没有 $j(j\neq i)$，使 $F_j(x)=+1$，显然 $F(x)=F(x^i)$. 便有 x 与 x^i 属于同一类的. 命题证毕.

由此可见，我们所得到的网络作为聚类器比一般的聚类器有很大的优点. 它具有一定的容错能力.

图 1 网络若将 N_3 断开，就得到多层前馈网络. 故图 1 网络可以看成是多层前馈网络的推广. 这样，对神经网络的 2 个主要功能，我们都给出了很有效的学习算法. 今后，主要是研究作为求优化的神经网络的学习和综合的有效算法问题.

4　结　论

本文是研究神经网络的学习算法总的课题的组成部分，即研究作为聚类器的神经网络

的学习和综合算法(或研究多层反馈神经网络的学习和综合算法问题).

我们给出的算法,具有以下的优点:(1)权系数、阈值简单;(2)进行聚类分析时,操作方便且可得到不同粒度的聚类;(3)用于聚类分析时,收敛速度快(最多经过($p-1$)步),且有一定的容错能力;(4)学习速度快(是样本个数的线性函数),网络的元件个数少.

我们的方法与已有的学习方法(如 BP 算法[3]、δ-算法等)相比,不但有以上所述的优点,更主要的是在方法论上的区别,一般的学习方法是在局部上进行调整搜索,已经有理论证明(见文献[4]):即使在树上进行启发性搜索,在一般情况下其计算量也是指数阶的,故按这种思路进行求解很难得到好的结果的.而我们学习方法是在全局上进行规划求解,事实证明:使用全局的观点能得到好的结果的.

综上所述,神经网络的 FP 算法为神经网络的学习和综合提供了强有力的方法,也为神经元硬件化提供了理论基础.

参考文献

1 张铃,吴福朝,张钹.多层前馈神经网络的学习和综合算法.软件学报,1995,6(7):440~448.

2 汪培庄.模糊集合论及其应用.上海:上海科学技术出版社,1983.

3 Rumelhart D E, McClelland J L. Parallel distributed processing. Vol. 1 & Vol. 2, MIT Press, Cambridge, MA, 1986.

4 Pearl J. Heuristics. New York: Addison-Wesley Company, 1984.

A FORWARD PROPAGATION LEARNING ALGORITHM OF MULTILAYERED NEURAL NETWORKS WITH FEEDBACK CONNECTIONS

*※ZHANG Ling　　**※ZHANG Bo

*(*Institute of Artificial Intelligence Anhui University Hefei* 230039)

**(*Department of Computer Science Tsinghua University Beijing* 100084)

※(*State Key Laboratory of Intelligent Technology and Systems Beijing* 100084)

Abstract A forward propagation learning algorithm (FP) of multilayered neural networks with feedback connections is presented in this paper. And the properties of cluster networks are discussed. A cluster with different grain-sizes can be obtained by applying FP to cluster. Its convergent speed is just a linear function of sample size. Its computational complexity is a bilinear function of simple size and the dimension of imput vectors. The network constructed by the algorithm uses a comparatively fewer number of elements and its weight simply has one of three values, i.e., −1, 0, 1. Thus, it can be easily implemented into electronic circuits. The authors also discuss the properties of the network and show it is an ideal pattern classifier.

Key words Multilayered neural network with feedback connections, learning algorithm, clustering.

Class number TP18