

M-P 神经元模型的几何意义及其应用*

张铃^{1,3} 张钺^{2,3}

¹(安徽大学人工智能研究所 合肥 230039)

²(清华大学计算机科学与技术系 北京 100084)

³(清华大学智能技术与系统国家重点实验室 北京 100084)

摘要 给出M-P神经元模型的几何意义,这个几何的诠释,给神经元一个非常直观的理解,利用这个直观的理解,给出两个颇为有趣的应用:(1)用此法给出三层前向神经网络的学习能力的基本定理的新的证明;(2)给出前向网络的拓扑结构设计的新方法。

关键词 前向神经网络,领域覆盖,可测函数。

中图法分类号 TP18

1 人工神经元的几何意义

1943年,McCulloch和Pitts根据神经元传递中的“0,1律”和神经传递中信号不但有不同的强度,而且有兴奋和抑制两种情况,第1次提出了神经元的数学模型^[1],此模型一直沿用至今,其M-P神经元的模型如下:

将神经元看成是1个有 n 个输入和1个输出的元件,此元件的功能函数可表示为 $y = \sigma(W * x - \theta)$,其中 σ 是符号函数,即 $\sigma(x) = \begin{cases} 1, & x > 0 \\ -1, & x \leq 0 \end{cases}$ 。现在让我们来分析一下这个模型的几何意义。

1.1 超平面表示

由神经元的功能函数的定义可知,它是由两个函数复合而成的,其第1个函数为 $W * x - \theta$,若令其等于零,得: $W * x - \theta = 0$,这方程在 n 维空间中表示一个超平面 P ,当 $(W * x - \theta) > 0$ 时,表点 x 落在超平面的正半空间内,此时, $\sigma(W * x - \theta) = 1$;当 $(W * x - \theta) < 0$ 时,表示点 x 落在 P 的负半空间内,此时, $\sigma(W * x - \theta) = -1$ 。于是,一个M-P神经元的功能,可看成是一个由超平面划分的空间的位置的识别器。这给神经元一个相当直观的理解,人们也曾利用这个直观的理解,来进行神经网络的学习的研究。^[2,3]当两三个平面时,尚可直观理解,而当 n, m 较大时, n 维空间中 m 个超平面的相交情况非常复杂,很不直观。总之,我们很难用超平面的模型,帮助我们直观地理解神经网络的内在性质。故此,很少有人再用神经元的几何意义(指用超平面的表示)来帮助进行神经网络的学习研究了。

1.2 球面上“领域”表示

几何的直观往往是研究的很好的向导,上面用超平面的方法,虽遭到失败,但是只要我们对此稍加改变,将会给我们提供新的直观帮助。

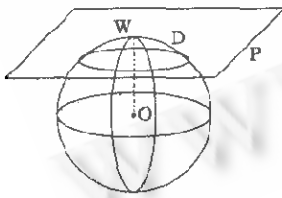


图1

若我们限定输入向量的长度相等,即输入向量是限定在 n 维空间的某个球面上,那么这时, $(W * x - \theta) > 0$,就表示球面上落在 P 的正半空间的部分,这个部分恰好是球面上的某个“球形领域”,若取 W 与 x 等长,则这个“球形领域”的中心恰好是 W ,其半径为: $r(\theta)$ 是 θ 的单调下降的函数(如图1所示)。若我们取 $\sigma(x) = \begin{cases} 1, & \text{当 } x > 0 \\ 0, & \text{其他} \end{cases}$,且取神经元的激励函数为: $\sigma(W * x - \theta)$,则一个神经元的功能函数,正好是它所代表的球面上“球形领域”的特征函数。这样,我们就能非常直观地进行神经网络的各种研究。

* 本文研究得到国家自然科学基金、国家863高科技项目基金和国家“攀登”计划基金资助。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。张钺,1935年生,教授,博士生导师,中国科学院院士,主要研究领域为人工智能理论,计算机应用。

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1997-11-06收到原稿,1998-02-25收到修改稿

2 神经网络的几何意义的应用

2.1 多层前馈网络(作为联想记忆器或分类器)的综合和学习

多层神经网络的综合学习问题是人工神经网络理论中的研究热点之一,目前用得最多的算法要算是 BP 算法^[4],此算法理论基础坚实、推导严密、形式优美、物理概念清晰、应用范围广,但是 BP 算法也存在着学习速度慢、所得到的网络性能差、算法不完备等缺点。

下面,我们从另一角度给出多层前向网络(作为联想记忆器或分类器)的综合和学习算法。

设样本集 K 为: $K = \{x^t = (x^t, y^t), t = 0, 1, 2, \dots, s\}$, 那么,所谓“记忆”,就是当输入为 x^t 时网络的输出为 y^t , 记住 x^t 与 y^t 之间的对应关系;至于“联想”,就是当输入为 $x^t + \Delta^t$ 时,其输出仍为 y^t (其中 Δ^t 是噪声或微扰)。有上面两种功能的网络,就可以认为是一个“联想记忆器”。

另一方面,若将“当输入为 $x^t + \Delta^t$ 时,其输出仍为 y^t ”理解为:输入落在 x^t 的附近时,其输出应为 y^t , 否则其输出就不为 y^t , 若将在“ x^t 附近”看作是 x^t 的一个球形领域,那么,由神经元的几何意义,知:若取以 x^t 为中心,以 Δ^t 为半径的“球形领域”,则其对应的神经元,就能完成上述的联想记忆的任务。

若将多层前向神经网络作为“分类器”来进行设计,那么其设计就相当于用若干“球形领域”将输入 x^t 按其所属的类把它们划分开来。一个最简单的方法,就是对每一类,用一组球形领域将属于该类的 x^t 覆盖住,又不覆盖不属于该类的 x^t , 于是不同类的输入被不同组的球形领域所覆盖。然后再将属于同组球形领域对应的神经元的输出用或门集中起来。这样就给出一个分类器的设计。

按照这样的理解,我们可给出这个设计方法的粗略的框架,暂称之为领域覆盖算法(至于此方法的精细、深入的研究,将在另文进行)。

2.1.1 领域覆盖算法(最简单的情况)

设样本集 K 为: $K = \{x^t = (x^t, y^t), t = 0, 1, 2, \dots, p-1\}$, K 中不同的 y^t 只有 k 个,不妨设 y^t 的头 k 个的值均不相同,令样本的输出为 y^t 的样本标号的集合为 $I(t)$ (即 $I(t) = \{i | y^t = y^i\}$) 其对应的输入集合记为 $P(t)$, $t = 0, 1, 2, \dots, k-1$ 。并将输入样本的上标按 $I(0), I(1), \dots, I(k-1)$ 顺序排序。

于是,若我们能够取一批“球形领域” $\{C_j^t, t = 0, 1, \dots, k-1; j = 1, 2, \dots, k_t\}$, 令 $C^t = \cup C_j^t$, 使得: C^t 只覆盖 j 属于 $I(t)$ 的 x^i , 而不覆盖 j 不属于 $I(t)$ 的 x^i , 且 C^t 互不相交。这样的球形领域就能完成分类的要求。

固定 t , 令 $P(t) = \{x^i | i \in I(t)\}$, 对样本输入 $x^i \in P(t)$, 令

$$\begin{aligned} d^1(i) &= \max_{j \in I(t)} \{ \langle x^i, x^j \rangle \} \\ d^2(i) &= \min_{j \notin I(t)} \{ \langle x^i, x^j \rangle > d^1(i) \} \\ d(i) &= \frac{d^2(i) + d^1(i)}{2} \\ \theta_i &= d(i) \quad W = (w^i), \theta = (\theta_i) \quad i = 1, 2, \dots, k_t \end{aligned} \quad (1)$$

其中 $\langle x, y \rangle$ 表示 x, y 的内积。

定义 1. 令 $D_i(t) = \{x | \langle x, x^i \rangle > \theta_i\}$, 称之为以 x^i 为中心的邻域, 以式(1)定义权、阈值的神经元记为 A_i^t , 称为对应于 x^i 的神经元, 其中 $i \in I(t)$ 。

定义 2. 令 $D(t) = \cup_{i \in I(t)} D_i(t)$, 称 $D(t)$ 为对应于 y^t 的领域集。

由 FP 算法^[5]知, 当样本输入 $x^i \in P(t)$ 时, 则 $x^i \in D(t)$; 当 $x \in P(t)$ 时, 则 $x \in D(t)$, $t = 1, 2, \dots, k$ 。

若我们可以从 $D(t)$ 中选出 $P(t)$ 的子覆盖 $D'(t)$, 令 $D'(t)$ 中邻域中心 x^i 的标号的集合为 $I'(t)$, 于是在网络第 1 层只留下 $A_i^t, i \in \cup I'(t) (t = 1, 2, \dots, k-1)$ 的神经元, 就可完成当 $x \in P(t)$ 时, $A_i^t(x) = 1, i \in \cup I'(t) (t = 1, 2, \dots, k-1)$ 的性质。

网络的第 2 层, 取 m 个元件 $B_i, i = 1, 2, \dots, m$. 其权、阈值取如下值即可。

$$\begin{aligned} u_i^j &= \frac{1}{2} (y_j^t - y_j^s), \xi_j = -[y_j^s + \frac{1}{2} \sum_{i=1}^m u_i^j] \\ U &= (u_i^j), \xi = (\xi_j) \end{aligned} \quad (2)$$

显然, 一个覆盖中的球形领域, 对应于第 1 隐层中的一个隐层元, 所以, 若我们每次从 $D(t)$ 中, 能选出最小覆盖 $D'(t)$, 则在一定意义下, 我们就给出了隐层元最少的网络结构。

可惜, 求最小覆盖问题是一个典型的 NPC 问题, 故目前尚无有效的求解方法。

注:上面是以不同的 y^j 值来分类,若不同的 y^j 值的个数 k 比较大(如 $k \gg m$)且 y 分量是二值的情况时,则也可对 y 的每个分量进行分类.另外,用来覆盖的领域也不一定要以 x^i 为中心,以及其他一些技巧均可使隐层的元件个数减少.

从以上分析得,任何一种求(次优)最小覆盖的方法与“球形领域”法相结合,都将给出一个神经网络的学习算法,称这种学习算法为领域覆盖算法.

现将学习步骤总结如下:

领域覆盖算法(已知样本集 $K = \{r^t = (x^t, y^t), t = 0, 1, 2, \dots, p-1\}$,且每个 x^t 的长度相等):

- (1) 按定义 2 求各 y^j 对应的领域集 $D(t), t = 0, 1, 2, \dots, k-1$ (k 为样本的不同输出的个数).
- (2) 按某种求最小(次小)覆盖的算法,从 $D(t)$ 中求出 $P(t)$ 的最小(或次小)覆盖 $D'(t), t = 0, 1, 2, \dots, k-1$.
- (3) 令 $I'(t)$ 是 $D'(t)$ 中领域对应的指标集,对 $D'(t)$ 中的领域 $D_i(t)$,做对应的神经元 $A_i^t, i \in \cup I'(t)$.
- (4) 构造网络:第 1 元件层,由元件 $A_i^t (i \in \cup I'(t), t = 0, 1, \dots, k-1, i = 1, 2, \dots, k)$ 构成.
- (5) 若输出是实值向量时,增加一个隐含层,其元件设为 O_s, O_s 是或门,其输入为 $A_i^t (i \in I'(t))$ 的输出.不然转入(6).
- (6) 输出层,取 m (m 是输出向量的维数)个神经元 $B_s, s = 1, 2, \dots, m$.

(6.1) 若输出是实值向量时,取 B_s 为线性元件,其对应函数为

$$f(x) = \sum_{s=0}^{k-1} y_s^t O_s(A_i^t(x)), s = 1, 2, \dots, n \tag{5}$$

注:单独符号 A_i, O_i 等表示神经元,而符号 $A_i(\cdot), O_i(\cdot)$ 表示对应神经元的功能函数.其中 $A^t(x) = (A_1^t(x), A_2^t(x), \dots, A_k^t(x))$ 表示一向量; $O_i(A^t(x))$ 表示 $O_i(\cdot)$ 的输入是 $A^t(x)$ 时输出时值, $i = 1, 2, \dots, k, B_s$ 与 O_s 的连接权为 y_s .

(6.2) 若输出是二值向量,则其功能函数按 FP 算法的第 2 层元件定义.

即
$$u_i^j = (\frac{1}{2}(y_i^j - y_i^0)), \xi_j = -[y_j^0 + \frac{1}{2} \sum_{i=1}^m u_i^j]$$

取权、阈值分别为 U, ξ

$$U = (u_i^j), \xi = (\xi_j) \tag{4}$$

显然,按领域覆盖算法构成的函数,有: $f(x^t) = y^t$, 当 $i \in I(t)$.

注:在领域覆盖法中,当输入是实值情况时,我们在网络中增加了第 2 元件层,这是为了构造方便起见.因为在 $D'(t)$ 中的各领域 $D_i(t)$ 可能有相互重叠的情况,用第 2 元件层(或门)就能克服这种现象,最后再用一个线性元件就能完成整个网络的合成.当然,若我们不考虑元件个数的多少,每个 $x_i(t)$ 均用一个互不相交的领域覆盖的话,则第 2 元件层(或门)可以不需要.

2.2 三层前向网络能力的基本定理的证明

1989年, Funabashi, Arzi 和 Hecht-Nielsen 在文献[6~8]中分别证明(其证明相当复杂)了三层前向网络能任意逼近紧集上的连续函数和平方可积函数(定理:三层前向神经网络,可以任意逼近 $L^2(D')$ 中的函数,其中 D' 为 n 维欧氏空间中的单位超立方体).

本节用上面神经元的几何理解(球形领域的特征函数),非常直观地证明上述定理.下面简单介绍一下用到的一些概念和引理(下面的测度,均指 n 维欧氏空间中的勒贝格测度).

定义 3. 设 D 是 S^n 上的有界可测集, $f: D \rightarrow R$, 且 $f(x) = \sum_{i=1}^m a_i d(D_i)$, 其中 $d(D_i)$ 是 D_i 的特征函数, D_i 是可测集, 则称 f 为阶梯函数(其中 S^n 是 $n+1$ 维空间中的 n 维球面,下同). 若每个 D_i 均是球形领域,则称 f 为球形领域阶梯函数.

从定义 1 知,若每个 D_i 都是球形领域,则由神经元的几何意义得, $d(D_i)$ 可用一神经元表示.于是,若第 1 层放置 D 对应的神经元 $A_i, i = 1, 2, \dots, m$. 第 2 层用一线性元件,这一神经元起着线性组合的作用.于是这样的三层前向网络就可实现阶梯函数 $f(x)$. 另一方面,由实变函数论知,任一平方可积函数可用阶梯函数任意逼近,故要证明基本定理,只要能证明任一可测集能用“球形领域”逼近即可(而这在实变函数论中是一个现成的结论),于是,只要将这些结论重新稍加改写,即可得出基本定理的证明.

定义 4. 设 D 是 R^n 上的有界可测集, $f: D \rightarrow R$ 的有界可测函数, $g(x): D \rightarrow R$, 若满足: $\mu(\{x | |f(x) - g(x)| > \epsilon\}) < \delta, \mu$ 是 R^n 上的测度, 则称 $g(x)$ 在 D 上 (δ, ϵ) -测度逼近于 $f(x)$, 或称 $g(x)$ 是 $f(x)$ 的 (δ, ϵ) -逼近.

引理 1. 设 D 是 R^n 上的有界可测集, 任给 $\epsilon > 0$, 则存在有限个开领域 C^i , 使得: 令 $C = \cup C^i, D \subset C$, 有 $\mu(C/D) < \epsilon$.

引理 2. 设 D 是 R^n 上的有界可测集, $f: D \rightarrow R$ 上的有界可测函数, 则任给 $\epsilon > 0$, 存在阶梯函数 g , 且 g 是 f 的 $(0, \epsilon)$ -测度逼近.

引理 3. 设 D 是 R^n 上的有界可测集, $f: D \rightarrow R$ 上的几乎有限可测函数, 任给 $\delta, \epsilon > 0$, 则存在阶梯函数 h , h 是 f 的 (δ, ϵ) -测度逼近.

引理 4. 设 D 是 S^n 上的有界可测集, 任给 $\epsilon > 0$, 存在有限个互不相交的“球形领域” $C_i, i=1, 2, \dots, t$, 令 $C = \cup C_i$, 则 C 与 D 的对称差的测度小于 ϵ , 即 $\mu(C/D) \cup (D/C) < \epsilon$. 其中 $C/D = \{x | x \in C \& x \notin D\}$.

引理 5. 设 D 是 R^n 上的有界可测集, $f: D \rightarrow R$ 上的几乎有限可测函数, 任给 $\epsilon > 0$, 则存在 $\delta, \epsilon_1 > 0$, 当 $g(x)$ 是 f 的 (δ, ϵ_1) -测度逼近时, $d_{L^2}(f(x), g(x)) < \epsilon$. 其中 $d_{L^2}(\dots)$ 是 L^2 空间的距离.

有了上述几个引理(除引理 4 外(引理 4 的特殊之处是要求各 C_i 不相交), 都是实变函数论中最一般的常识)之后, 我们就可以非常直观地给出基本定理的证明轮廓.

设 f 是平方可积函数 $\rightarrow f$ 是几乎有限可测函数 \rightarrow 存在其 (δ, ϵ) -测度逼近的阶梯函数 $g \rightarrow$ 将 g 的定义域中的可测集用互不相交的“球形领域”逼近 \rightarrow 得到 f 的新的 (δ, ϵ_1) -测度逼近函数 g_1 (g_1 是球形领域阶梯函数) $\rightarrow g_1$ 就是 f 在 L^2 空间中的一个 ϵ 逼近且 g_1 可由三层前馈网络来实现. 这样就得出基本定理的证明.

定理 1. 设 D 是 S^n 上的有界可测集, $f: D \rightarrow R$ 上的几乎有限可测函数, 任给 $\delta, \epsilon > 0$, 则存在一个三层前向神经网络, 其对应功能函数 $g(x)$ 是 f 的 (δ, ϵ) -测度逼近.

证明: 任给 $\delta, \epsilon > 0$, 由引理 3 得, 存在其 $(\delta/2, \epsilon)$ -逼近函数 h , 设 $h(x) = \sum_{i=1}^m a_i d(D_i)$, 再由引理 4 对每个 D_i , 存在有限个互不相交的“球形领域” C_j^i , 设其并为 C^i , 且有 $\mu(C^i/D_i \cup D_i/C^i) < \delta/2m$, 令 $E = \cup (C^i/D_i \cup D_i/C^i)$.

结构一个三层网络, 其第 1 元件层放置元件 $A_j^i, i=1, 2, \dots, m; j=1, 2, \dots, i_i$, 其中 A_j^i 是球形领域 C_j^i 对应的神经元. 其第 2 元件层仅为一个线性元件 B , 若元件 A_j^i 的功能函数是特征函数, 则其对应的权值均为 $a_i, j=1, 2, \dots, i_i$.

因为 C_j^i 互不相交, 故有 $d(C^i) = \sum_j d(C_j^i)$, 这样的三层网络, 其映射函数为

$$g(x) = \sum_{i=1}^m a_i \left[\sum_j d(C_j^i)(x) \right] = \sum_{i=1}^m a_i d(C^i) \tag{5}$$

若元件 A_j^i 的功能函数是符号函数, 利用符号函数与特征函数之间的关系, 也可得到其对应的权、阈值.

因为 $d(C_j^i)(x) = \frac{[A_j^i(x)+1]}{2}$, 其中 $A_j^i(x)$ 表示元件 A_j^i 对应的功能函数. 将此关系代入式(5), 得

$$\begin{aligned} g(x) &= \sum_{i=1}^m a_i \left\{ \sum_j \frac{[A_j^i(x)+1]}{2} \right\} \\ &= \sum_{i=1}^m \left\{ \sum_j \frac{a_i}{2} [A_j^i(x)+1] \right\} \\ &= \sum_{i=1}^m \left\{ \left[\sum_j \frac{a_i}{2} A_j^i(x) \right] + \frac{a_i}{2} k_i \right\} \\ &= \left\{ \sum_{i=1}^m \left[\sum_j \frac{a_i}{2} A_j^i(x) \right] \right\} + \sum_{i=1}^m \frac{a_i}{2} k_i \end{aligned}$$

其中 k_i 是固定 i, C_j^i 的个数.

于是, 对应于元件 B 与 A_j^i 的连线的权值为 $a_i/2$, 元件 B 的阈值为 $\sum_{i=1}^m \frac{a_i}{2} |I(i)|$.

显然, 当 $x \in D_i/E$ 时, 有 $g(x) = h(x), i=1, 2, \dots, m$. 且 $\mu(E) < \delta/2$, 再由 $h(x)$ 是 f 的 $(\delta/2, \epsilon)$ -测度逼近. 于是得, $g(x)$ 是 f 的 (δ, ϵ) -测度逼近. □

推理 1. 设 D 是 R^n 上的有界可测集, $f: D \rightarrow R$ 上的平方积函数, 任给 $\epsilon > 0$, 则存在三层前向神经网络, 令其功能函数为 $g(x)$, 且有 $d_{L^2}(f(x), g(x)) < \epsilon$.

注: 推理 1 就是有名的 Hecht-Nielsen 定理.

参考文献

- McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 1943, 5: 115~133
- Rujan P, Marchand M. A geometric approach to learning in neural networks. In: Proceedings of the International Joint Conference on Neural Networks '89, Vol II. Washington, DC: IEEE TAB Neural Network Committee, 1989. 105~110
- Ramacher U, Wesseling M. A geometrical approach to neural network design. In: Proceedings of the International Joint

- Conference on Neural Networks'89, Vol II. Washington, DC: IEEE TAB Neural Network Committee, 1989. 147~154
- 4 Rumelhart D E, McClelland J L. *Parallel distributed processing* (Vol. 1~2). Cambridge, MA: MIT Press, 1985
- 5 张铃, 张钹. 多层前馈神经网络的综合和学习算法. 软件学报, 1997, 8(4): 252~258
(Zhang Ling, Zhang Bo. A forward propagation learning algorithm of multilayered neural networks with feedback connections. *Journal of Software*, 1997, 8(4): 252~258)
- 6 Funabashi K. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 1989, 2: 183~192
- 7 Arai M. Mapping abilities of three-layer neural networks. In: *Proceedings of the International Joint Conference on Neural Networks'89, Vol I*. Washington, DC: IEEE TAB Neural Network Committee, 1989. 419~423
- 8 Hecht-Nielsen R. Theory of the backpropagation neural networks. In: *Proceedings of the International Joint Conference on Neural Networks'89, Vol I*. Washington, DC: IEEE TAB Neural Network Committee, 1989. 593~605

A Geometrical Representation of M-P Neural Model and Its Applications

ZHANG Ling^{1,3} ZHANG Bo^{2,3}

¹(Artificial Intelligence Institute Anhui University Hefei 230039)

²(Department of Computer Science and Technology Tsinghua University Beijing 100084)

³(State Key Laboratory of Intelligent Technology and Systems Tsinghua University Beijing 100084)

Abstract In this paper, a geometrical representation of M-P neural model is presented. From the representation, a clear visual picture and interpretation of the model can be seen. Two interesting applications based on the interpretation are discussed. They are (1) a new design principle of feedforward neural networks, and (2) a new proof of mapping abilities of three-layer feedforward neural networks.

Key words Feedforward neural networks, neighborhood covering, measurable functions.