第24卷 第6期
1998年11月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 24, No. 6
Nov., 1998

短 文

一类二次 Sigmoid 神经元构建的分类器分类能力的理论分析[1)]

王保云　　何振亚
(南京邮电学院信息工程系　南京　210003)　(东南大学无线电系　南京　210096)

摘　要　针对含二次 Sigmoidal 神经元的前向神经网络，研究了它的模式分类能力. 通过分析它在解决 Dichotomy 问题所需的规模，从而得到了这种神经元模型对神经网络分类器的分类能力的改进程度.

关键词　神经元模型，二分类，阈值，分类能力.

1　引言

人工神经网络中的基本信息处理单元——神经元模型是决定网络的信息处理能力的重要因素之一. Chiang 等为构建性能较好的模式分类器，提出以二次 Sigmoid 函数作为神经元的作用函数，由此得到两种新型神经元模型——单阈值二次 Sigmoid 神经元及多阈值二次 Sigmoid 神经元. 这种神经元模型是从工程应用的角度提出的，它在训练效率及泛化能力方面表现出较好的性能.

已经知道，Committee Machine 对分(Dichotomize)定义在有 k 个训练模式的训练集上的任意二分类(Dichotomy)至多需要 $k-1$ 个隐结点. Chiang 等通过构造法证明了在隐层含至多 $k+1$ 个单阈值二次 Sigmoid 神经元及一个多阈值二次 Sigmoid 输出神经元的网络可对分定义在含至少 $4k+1$ 个模式的训练集上的二分集. 本文对文[6]的结果进行了进一步地研究，统一分析了含二次 Sigmoid 神经元的这类网络的分类能力，明确指出了引入二次神经元对网络分类能力的改进程度.

2　二次 Sigmoid 型神经元及 Dichotomy 问题的表述

定义1. 设输入为 $\boldsymbol{x}=(1,x_1,x_2.\cdots,x_n)$，相应权值、阈值分别为 $\boldsymbol{w}=(w_0,w_1,\cdots,w_n)$ 和 $\boldsymbol{\theta}=(\theta_0,\theta_1,\cdots,\theta_n)$，则作用函数为

$$f(net,\boldsymbol{\theta})=\frac{1}{1+\exp(net^2-g(\boldsymbol{\theta},\boldsymbol{x}))}$$

1)国家自然科学基金资助项目.

收稿日期　1996-09-02

的神经元称为多阈值二次 Sigmoid 神经元. 这里

$$net = \boldsymbol{w} \cdot \boldsymbol{x} = w_0 + \sum_{i=1}^{n} w_i x_i, \quad g(\boldsymbol{\theta}, \boldsymbol{x}) = \theta_0 + \sum_{i=1}^{n} \theta_i x_i.$$

定义2. 以函数

$$H(net, \boldsymbol{\theta}) = \begin{cases} 1, & net \geq g(\boldsymbol{\theta}, \boldsymbol{x}), \\ 0, & \text{otherwise} \end{cases}$$

为作用函数的神经元称为多阈值二次 Heaviside 神经元.

在定义1和2中，若 $g(\boldsymbol{\theta}, \boldsymbol{x})$是一常数，则相应的神经元称为单阈值二次型神经元. 二次 Sigmoid 神经元与 σ-π 模型、高斯模型有着密切的关系.

Dichotomy 问题是一个抽象的模式分类问题，可很好地检验一类网络的模式分类能力. n 维欧氏空间 R^n 的二分类问题可描述为[4] $R^n = D_+ \cup D_-$，$D_+ \cap D_- = \varnothing$；相应的目标函数为 $f(\boldsymbol{x})=1$，当 $\boldsymbol{x} \in D_+$；$f(\boldsymbol{x})=0$，当 $\boldsymbol{x} \in D_-$. 文[5]指出神经网络学习 n 维空间二分类问题可化为1维情况进行研究.

3　含二次 Sigmoid 神经元的前向网络的分类能力的理论分析

对一训练样本集 $S = \{\boldsymbol{x}^i \mid \boldsymbol{x}^i \in R, i = 0, 1, \cdots, 4k+2\}$，且有 $\boldsymbol{x}^i < \boldsymbol{x}^{i+1}$，$f(\boldsymbol{x}^i) = 1 - f(\boldsymbol{x}^{i+1})$，这里 $f(\cdot)$为相应的目标值.

构造如下区间序列

$\alpha_- < I_0 < \alpha_0 < I_1 < \beta_0 < I_2 < \gamma_0 < I_3 < \gamma_0' < I_4 < \beta_0' < I_5$，　$\alpha_1 < \cdots < I_{4k+2} < \alpha_+$

满足 $\boldsymbol{x}^i \in I_i$，网络的权值参数由 $\alpha_-, \alpha_0, \beta_0, \gamma_0, \cdots, \alpha_+$ 加以确定.

考虑三层前向网络，设网络的输入到隐层的连接权向量为 $\boldsymbol{w}_i = (w_{i0}, w_{i1})$，隐层神经元的阈值为 $\boldsymbol{\theta}_i^1 = (\theta_{i0}^1, \theta_{i1}^1)$，隐层到输出层的权为 $\boldsymbol{u} = (u_0, u_1, \cdots, u_{k+1})$，输出神经元的阈值为 $\boldsymbol{\theta}^2 = (\theta_0^2, \theta_1^2, \cdots, \theta_{k+1}^2)$.

对网络作如下两个假设：

1）第 i 个隐层神经元接受的加权输入为 $q_i(\boldsymbol{x}) = (\theta_0^1 + \theta_1^1 \boldsymbol{x}) - (w_{i0} + w_{i1}\boldsymbol{x})^2$，其两个零点不包含在集合 $\bigcup_j I_j$ 中；

2）$(u_0 + \sum_i u_i H[q_i(x)])^2 \neq (\theta_0^2 + \sum_t \theta_t^2 H[q_i(\boldsymbol{x})])$，对 $\boldsymbol{x} \in \bigcup_j I_j$.

定理1. 训练集

$$S = \{x_i \mid x_i \in R, i = 0, 1, \cdots, 4k+2\}, \quad \boldsymbol{x}^i < \boldsymbol{x}^{i+1}, \ f(\boldsymbol{x}^i) = 1 - f(\boldsymbol{x}^{i+1}), \tag{1}$$

对由二次 Sigmoid 神经元组成的三层网络，至少需 k 个隐结点才能正确学习 S. 对含二次 Heaviside 神经元及输入输出直接连接的三层网络有如下结果.

引理1. 对任意一含输入输出直接连接的三层神经网络，若其隐层神经元为 k 个单阈值或多阈值二次 Heaviside 神经元，输出层为一多阈值二次 Heaviside 神经元，则无论其权值及阈值怎样选取，它至多只能把 R 分成$4k+3$个子区间以使得

$$S^i \cap S^j = \varnothing, \quad f(\boldsymbol{x}) = 1 - f(y), \quad \forall\, \boldsymbol{x} \in S^i, y \in S^{i+1}, \quad i = 1, \cdots, 4k+2, \tag{2}$$

这里 $f(\boldsymbol{x})$为网络的输出值.

证明. 设 q_i^-, q_i^+ 为 $q_i(\boldsymbol{x}) = (\theta_0^1 + \theta_1^1 \boldsymbol{x}) - (w_{i0} + w_{i1}\boldsymbol{x})^2$的两个零点，$q_i(\boldsymbol{x})$仅在区间$[q_i^-$，

q_i^+]内非负. 首先考虑 $q_i^+ < q_{i+1}^-(i=1,2,\cdots,k-1.)$ 的情形,这时$[q_i^-, q_i^+](i=1,2,\cdots,k.)$互不相交,所以实数集 R 被分成3k 个子区间,记为 $I_1, I_2, \cdots, I_{3k}$,如图1(a)所示. 因此隐层神经元的输出应该为

$$x \in I_{3i-1}, \quad \begin{cases} h_i = 1, \\ h_j = 0, \end{cases} \quad j \neq i, \quad i = 1,2,\cdots,k, \tag{3a}$$

$$x \in \bigcup_{i=1}^{k} (I_{3i-2} \cup I_{3i}), \quad h_j = 0, \quad j = 1,2,\cdots,k. \tag{3b}$$

进一步,整个网络的最后输出是

$$O = H_q^e(u_0 + vx, \theta^2) = \begin{cases} 1, |u_0 + vx| \leq \sqrt{\theta_0^2}, \\ 0, |u_0 + vx| > \sqrt{\theta_0^2}, \end{cases} \quad \text{if} \quad x \in \bigcup_{i=1}^{k} (I_{3i-2} \cup I_{3i}), \tag{4}$$

$$O = H_q^e(u_0 + vx + u_i, \theta^2) = \begin{cases} 1, (u_0 + vx + u_i)^2 \leq \theta_0^2 + \theta_i^2, \\ 0, (u_0 + vx + u_i)^2 > \theta_0^2 + \theta_i^2, \end{cases} \text{if } x \in I_{3i-1}. \tag{5}$$

由式(5)可知,I_{3i-1}至多被输出神经元分成三个区间 $I_{3i-1}^1, I_{3k-1}^2, I_{3k-1}^3$,如图1(c)所示. 显然 $I_{3i} \cup I_{3(i+1)-2}$与 I_{3i-1}^3及 $I_{3(i+1)-1}^1$相邻,式(4)中的参数 u_0, v, θ_0与 i 无关,因此应有

$$O(x) = H_q^e(u_0 + vx, \theta^2) = 1, \text{ if } x \in \bigcup_{i=2}^{k-1} (I_{3i-2} \cup I_{3i}). \tag{6}$$

对区间 I_1及 I_{3k},适当选取参数 u_0, v, θ_0可使式(6)成立,同时可至多把 I_1及 I_{3k}分成 $I_1^+ \cup I_1^-$, $I_{3k}^+ \cup I_{3k}^-$,且有

$$O = H_q^e(u_0 + vx, \theta^2) = 1, \text{if} \quad x \in I_1^+ \cup I_{3k}^+; O = H_q^e(u_0 + vx, \theta^2) = 0, \backslash\text{if} \quad x \in I_1^- \cup I_{3k}^-.$$

综合上面的分析,可以看到所述网络至多可把 R 分成满足式(2)的4k+3个小区间

$$I_1^-, \quad I_1^+, \quad I_{3k}^-, \quad I_{3k}^+, \quad I_{3i-1}^1, \quad I_{3i-1}^2, \quad I_{3i-1}^3, \quad I_{3i} \cup I_{3i+1}.$$

假如存在 i_0, $q_{i_0}^+ > q_{i_0+1}^-$,如图1(b)所示,显然经输出神经元的划分后所得子区间的最大数目与图1(a)所示的情形相比不会增加. 引理得证.

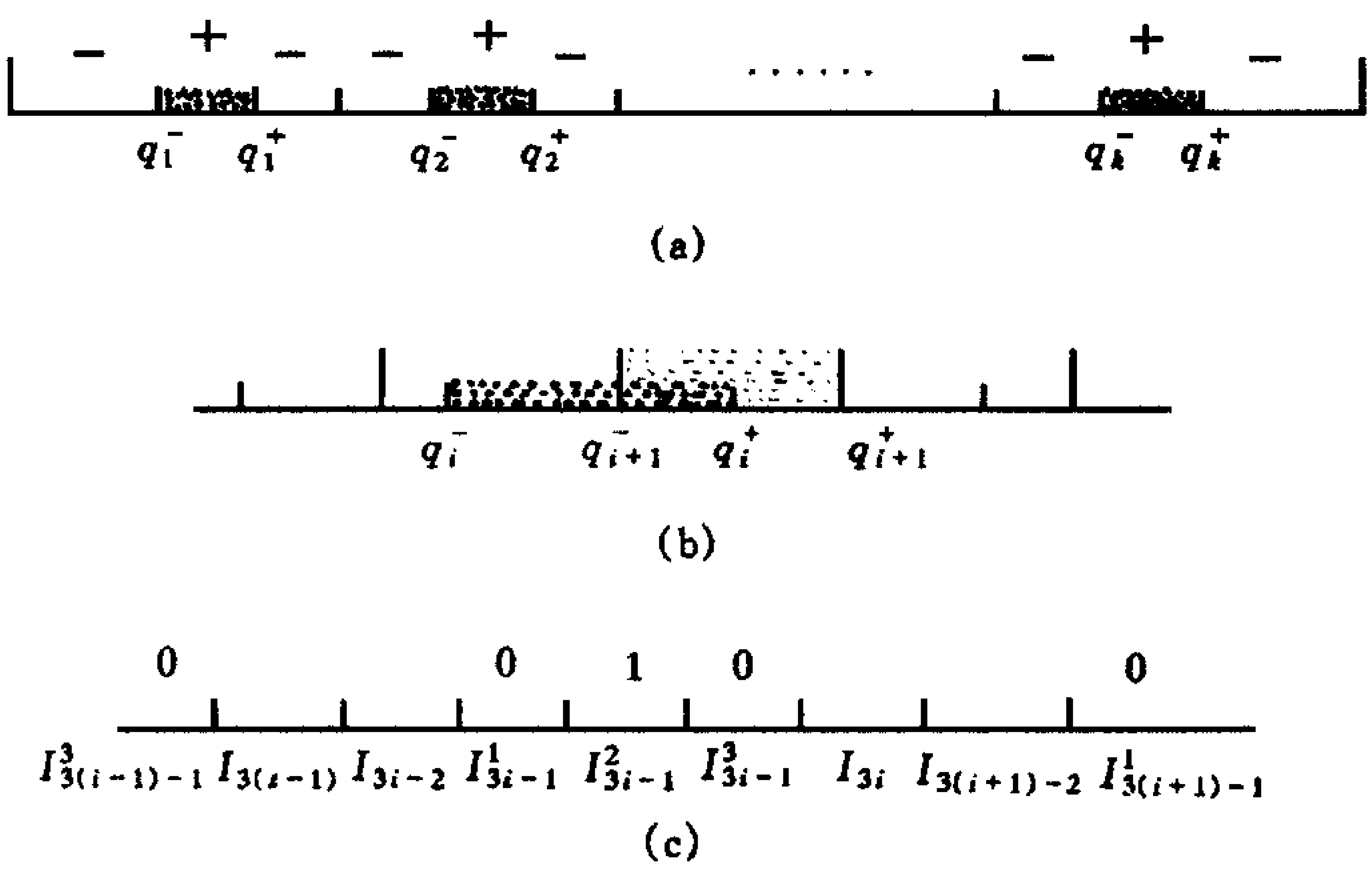

图1 (a)隐层神经元对 R 的划分,(b)隐层神经元划分成的小区间有重叠情形,(c)输出神经元对(a)中小区间的进一步划分

定理1的证明. 用反证法. 假设有一个三层网络在其隐层有少于 k+1个二次 Sigmoid 神经元可以学习所述的训练集,由上面的引理1及文[5,6]中的引理可知,单(多)阈值的二次 Sigmoid 神经元可由一单(多)阈值的二次 Heaviside 神经元来替换,其中隐层的一个单阈值的二次 Sigmoid 神经元可由输入输出直接连接来逼近. 因此应存在一个有输入输出直接连接的三层网络,在其隐层有少于 k 个的二次 Heaviside 神经元能将实数集 R 分成4k+3个小区间. 这与引理相矛盾. 证毕.

文[5]的结果表明二次 Sigmoid 神经元可以将网络的分类能力提高到原来的4倍,而

本文的定理1则指出这种神经元只能提高到4倍,从而明确了引入二次 Sigmoid 神经元对前向网络分类能力的改进程度,并给出了这种神经元在构建模式分类器中作用的一个评价.

参 考 文 献

1 Nilsson N J. Learning Machines: Foundation of Trainable Pattern Classifying Systems, New York: McGraw-Hill, 1965

2 Huang, S C, Huang. Y F. Bounds on the number of hidden neurons in multilayer perceptrons. *IEEE Trans. Neural Networks*, 1991, **2**:47－55

3 Chiang. C C, Fu H C. A variant of second-order multilayer perceptron and its application to function approximation. In:Proc IJCNN, Baltimore, 1992, Ⅲ:887－892

4 Amari S. A universal theorem on learning curves. *Neural Networks*, 1993, **6**:161－166

5 Chiang C C, Fu H C. Using multithreshold quadratic sigmoid neurons to improve classification capability of multilayer perceptrons. *IEEE Trans. Neural Networks*, 1994, **5**:516－519

6 Baoyun Wang, Zhenya He. Can the classification capability of network be further improved by using quadratic neurons? *Pattern Recognition*, 1998, **31**(9) to appear

THEORETICAL ANALYSIS ON CLASSIFICATION CAPABILITY OF CLASSIFIER CONTAINING QUADRATIC SIGMOID NEURONS

WANG BAOYUN

(*Dept. of Information Engineering, Nanjing University of Post and Telecommunication, Nanjing* 210003)

HE ZHENYA

(*Dept. of Radio, Southeast University, Nanjing* 210096)

Abstract This paper addresses the classification capability of network containing quadratic sigmoidal neurons. By analysing the size of the network required for solving Dichotomy problem, we can figure out the degree of the improvement on the classification capability of network by using quadratic sigmoidal neurons.

Key words Neuron model, dichotomy, threshold, classification capability.