DOI: 10.13656/j.cnki.gxkx.2004.01.005
广西科学 Guangxi Sciences 2004, 11 (1): 16~ 19

变参数激活函数可调神经元数学模型*

Variable Tunable Activation Function Neuron Mathematical Models

戴祯杰[1] 周永权[2]
Dai Zhenjie[1] Zhou Yongquan[2]

(1. 广西教育学院数学与计算机科学系 南宁市建政路 530023;
2. 广西民族学院计算机与信息科学学院 南宁市大学路 80号 530006)
(1. Dept. of Math.& Comp., Guangxi College of Education, Jianzhenglu, Nanning, Guangxi, 530023, China; 2. Dept. of Comp.& Info. Science Coll., Guangxi Univ. for Nationalities, 80 Daxuelu, Nanning, Guangxi, 530006, China)

摘要 提出一种能同时模拟包括 M-P与 TAF神经元在内的各种神经元通用的新数学模型,给出这种模型的一般形式. 该模型不但连接权值可调节,而且激活函数可根据实际问题动态地选取. 激活函数加入参数后,即成为变参数激活函数,大大地增强神经网络的灵活性.

关键词 神经元数学模型 激活函数 变参数 XOR问题
中图法分类号 TP18

Abstract A new kind of variable tunable activation function neuron mathematical model is designed, and the general form of variable tunable activation function is given. This form is different from *M-P* model, and the link weights is tunable. Since the variable tunable activation functions can be choosed randomly, so the non-linear capacity of neural networks could be improved. Finally, several given XOR problem examples show that the proposed new neuron mathematical model is effective and practical.

Key words neuron model, activation function, variable, XOR problem

1943年 McCulloch和 Pitts提出神经元的 M-P模型,这种模型用连接权值和非线性激活函数分别模仿生物神经元的突触和细胞体作用. 在学习过程中,神经网络的权值可调节,但激活函数不能调节,是事先固定的. 它通常选取一些简单的函数,如: 阶梯函数,分段线性函数、Sigmoid函数、双曲正切函数等. 但这种模型过于简单,它与生物神经元系统中神经元相差很大,其性能很难达到生物神经元的功能,因而阻碍着神经网络研究的进展.吴佑寿等人指出[1,2],这可能和神经元的模型选取有关, M-P模型的激活函数是固定不变,不适用求解所有问题. 并且认为,神经元模型的激活函数可以有各种形式,而不是固定不变的,在解决某些问题时,可根据问题的先验知识,选取与待解问题相适应的激活函数.但是,我们认为,先验知识往往不易得到,要经过多次的试验,才能获得先验知识,即使获得先验知识,并且不能适应所有问题; 其次,先验知识还受到研究者专家领域知识的限制,有可能由于获得先验知识不正确,导致产生错误的结果. 于是,文献 [3]提出变参数激活函数,通过改进激活函数,并在学习过程中调节连接权值的同时,调整激活函数和输入阈值,只是在某种程度上,仅对激活函数做了一点改进,使得神经网络具有较强的非线性映射能力.

本文在分析可调激活函数的神经元模型和变参数激活函数神经元功能基础上,将两者有机地结合起来,提出了一种能同时模拟包括 M-P与 TAF[1]神经元在内的各种神经元通用的新数学计算模型,并给出这种神经元数学模型的一般形式,我们称为变参数激活函数可调神经元数学模型. 一方面,在解决某些问题时,根据解决实际问题的先验知识,选取激活函数使之它与解问题相适应; 另一方面,在选定激活函数

2003-08-20收稿.
* 广西自然科学基金资助项目 (桂科基 0141034)和广西高校百名中青年学科带头人资助项目.

的同时，给激活函数加入适当参数，通过调整激活函数和输入阈值，使得所需网络会更为简单，网络性能会更优，泛化能力也会更强.

1 变参数激活函数可调神经元数学模型

图 1是变参数激活函数可调神经元数学模型（Variable Tunable Activation Function，以下简称 VTAF）. 仍是一多输入单输出的非线性元件，其输出:

$$O = f(X, W, T, U), \tag{1}$$

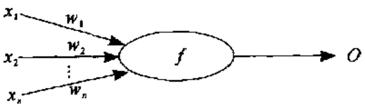

图 1 变参数激活函数可调神经元数学模型

Fig. 1 VTAF neuron mathematics model

式中，$W = (w_1, w_2, \cdots, w_n)$ 是 n维权矢量，$X = (x_1, x_2, \cdots, x_n)$ 是 n维输入信号矢量，X叫做外激活信号；T是一个可变矢量，我们称为外参变量，其用来确定激活函数类型的选取；U也是可变参矢量，我们称为内参变矢量，其用来改善函数的响应特性，以期得到更好的非线性表达能力.此时的 $f(X, W, T, U)$ 称为变参数可调激活函数神经元数学模型，在这个数学模型中，T, U都是可调节的参数，但它们的物理意义不同，在对网络训练时，可以调节这两个参数以改变激活函数 $f(X, W, T, U)$，使它与待解问题相适应，它是变参数可调激活函数神经元和一般 M-P模型的主要区别.

2 关于 VTAF神经元数学模型的讨论和依据

在（1）式，我们给出了变参数可调激活函数神经元数学模型的一般通式，以下讨论该模型包括 M-P与 TAF神经元在内的各种神经元通用的新数学计算模型的几种情况，在（1）式中:

（i）如果参数 T为常量，那么，变参数可调激活函数神经元退化为变参数激活函数神经元模型 -VAF神经元；

（ii）如果 U为常量，那么，变参数可调激活函数神经元退化为可调激活函数神经元模型[1,2]-TAF神经元；

（iii）如果 T, U同时不变，那么，变参数可调激活函数神经元模型退化为著名的 M-P模型.

由此，我们可看出: 新的 VTAF神经元数学模型是更一般的神经元数学模型，它对 M-P模型、TAF模型[1]、VAF模型[3]的功能增加了很大的自由度，因而扩大了神经元模型的适用范围.

事实上，在生物神经网络中，不可能所有神经细胞对输入有相同的反应. 生物的学习过程，除了各细胞间连接强度调节外，其本身响应特性也在变化. 例如，在日常生活中，有些人学习问题很快，有的人则不一样，这说明同一类型的神经元的细胞体对外来信号处理功能，因人而异，这一事实说明激活函数选取未必一致；另一方面，作为激活函数本身响应程度有所差别，说明往激活函数加入参数改善响应特性是合理的.

3 变参数可调激活函数的选取方法

一般认为，激活函数应该满足连续、可微、单调增的条件[4]. Hornik从理论上证明了任何连续、有界、非常值的函数都可作为激活函数[5]. 吴佑寿等人指出，在设计激活函数还应考虑实际问题，如函数尽量简单，便于计算，物理上易于实现. 归纳出作为神经元激活函数 f应满足三点:

（i）f应简单、易于运算、满足可微、有界、非常值的条件，在有界输入时能给出有界输出；

（ii）偏导数也应简单，最好是网络训练中计算结果的副产品；

（iii）尽量据待解问题的先验知识，使之与待求问题相匹配，以保证网络易于训练，泛化能力强，网络规模小，并且给出 TAF一种选择方法:

$$f(X, T) = \sum_{m=1}^{M} T_m h_m(X), \tag{2}$$

式中，$T_m\,(m = 1, 2, \cdots, M)$ 是实数，$h_m(X)$ 是 TAF的第 m个基函数，通常选用 Sigmoid函数、正态函数、指数函数等等.笔者认为，在生物神经网络中，不可能所有神经细胞对输入产生相同的影响，因此，在（2）式中基函数 h_m未必是同一类型的函数；其次，生物神经元在学习过程中，除了各神经细胞间连接强度的调整外，其本身特性也在变化.因此，在学习过程中，调整权值同时调整激活函数，即在（2）式中对基函数 $h_m(X)$ 加入适当的参数改善函数的响应特性，因此，笔者认为用作神经元的激活函数在文献 [1,2]提出以上三点基础上，应补充以下三点:

（iv）在（2）式中，基函数 $h_m(X)$ 可以是激活函数序列，它可由不同类型的函数构成，作为混合基函数序列；

（v）在（2）式中，每个基函数 $h_m(X)$ 可加入变参数矢量 U，使得基函数 $h_m(X)$ 成为 $h_m(X, U)$；

（vi）多维函数也可作为神经元的激活函数.

这样，（2）式可写成:

$$f(X, T, U) = \sum_{m=1}^{M} T_m h_m(X, U), \tag{3}$$

式中，T_m ($m = 1, 2, \cdots M$)，U_i ($i = 1, 2, \cdots, n$) 都是实数，$T = (T_1, T_2, \cdots, T_M)$，$U = (U_1 U_2, \cdots, U_n)$ 都是参变矢量，我们就得到 VATF神经元的一般表达式.

以下以 XOR问题为例，说明 VATF各种情形参数选取的方法.

例 **1** 图 2是 1个二层神经网络，其中神经元 f_1, f_2, f_3 的激活函数分别为 h_1, h_2, h_3，神经元 h 的激活函数为 h，则该网络的输出:

$$O = h(w_1 h_1(x_1) + w_2 h_2(x_2) + w_3 h_3(x_1, x_2) - \theta), \quad (4)$$

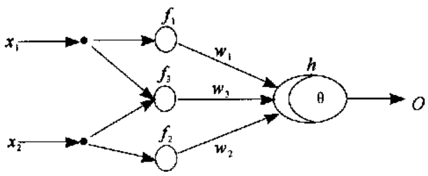

图 2 解决 XOR问题的二层神经网络

Fig. 2 Two layer neural network for solve XOR problem

式中，若取混合基函数序列: $h_1(u) = u$, $h_2(v) = v$, $h_3(u, v) = uv$，作为激活函数序列，则有:

$$O = h(w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 - \theta), \quad (5)$$

它可以模拟 XOR函数，若取权值矢量及阈值为: $W = (1, 1, -2)$，$\theta = 1$; (5) 式可转化为:

$$O = h(x_1 + x_2 - 2x_1 x_2 - 1),$$

对于二次曲线 $x_1 + x_2 - 2x_1 x_2 - 1 = 0$化为标准形:

$$-\frac{(x_1' - \sqrt{2}/2)^2}{(\sqrt{2}/2)^2} + \frac{(x_2')^2}{(\sqrt{2}/2)} = 1,$$ 它表示了一双曲线，可将 2组输入: {(0, 0), (1, 1)}, {(1, 0), (0, 1)} 分成二类，如图 3所示.

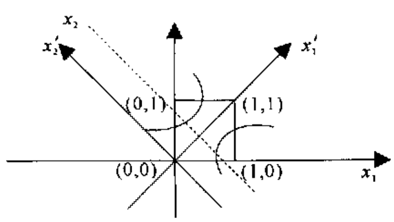

图 3 双曲函数把 XOR问题模式分类

Fig. 3 Hyperbolic function for XOR problem pattern classify

例 **2** 考察只有 1个可变激活函数神经元的网络如图 4所示，把 XOR问题模式分类.

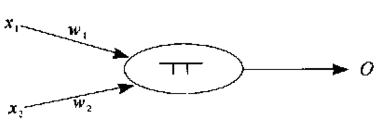

图 4 解决 XOR问题的 1个隐元网络

Fig. 4 A hidden neural network for solve XOR problem

设神经元 f的激活函数为 $h(u, v) = uv$为双曲函数，令 $f = h \prod$，其中 $\prod$ 视为一个映射，则该网络输出: $O = f(h(x_1, x_2) - \theta) = x_1 x_2 - \theta$，如果我们取权矢量 $W = (1, 1)$，$\theta = 1/2$，那么，我们可得到二次曲线 $x_1 x_2 = 1/2$，可化为标准形: $x_1^2 - x_2^2 = 1$，同样实现对 XOR问题的模式分类.

由例 1和例 2可看出，神经元激活函数选取，对同一个实际问题，可由不同类型的函数序列构成，也可以是多维函数 (如: 双曲函数等)，通过实例，回答了笔者提出关于神经元激活函数选取方法的 (4)和 (6).

以上例子说明如何动态地选取激活函数，一旦激活函数选定，能够适应大多数问题地需要，往激活函数中加人参数，使得成为 VTAF.

例 **3** 考察变参数 Sigmaid函数为: $f(x, \_, T) = \dfrac{1}{1 + \exp(-\frac{x - \_}{T})}$，其中，_ 表征自变量的偏移量，$T$ 是影响函数的“宽度”，T增大，函数的图形被横向“拉伸”，T缩小，函数的图形被横向“压缩”.这样，网络的信息不完全存贮于连接权值中，在学习中，调整 T，增强了神经元的自由度.

众所周知，只有 1个隐元的传统的 S-型网络，不能解决 XOR问题，但是，如果隐神经元是变参数 S型神经元激活函数，那么，只有 1个隐元的前向网络结构就能解决这个问题. 图 5就是这种网络结构.

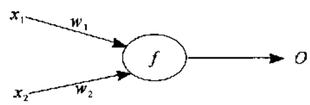

图 5 只有一个变参数神经元的网络

Fig. 5 A variable neural network model

该网络的输出: $O = f(w_1 x_1 + w_2 x_2 + \theta)$; 令 $S = w_1 x_1 + w_2 x_2 + \theta$，$(x_1, x_2)$ 是 1个 XOR码组，此时该网络的输出变为:

$$O = \frac{1}{1 + \exp(-\frac{S - \_}{T})}, \quad (6)$$

式中 _，T为可变参数，此时令: $S' = \frac{S - \_}{T}$，(6) 式可化为:

$$O = \frac{1}{1 + \exp(-S')}, \quad (7)$$

注意到 Sigmoid函数为单调递增连续函数，因此，存在反函数:

$$S' = \ln \frac{O}{1 - O}, \quad (8)$$

$$\text{式中，} S' = \frac{w_1 x_1 + w_2 x_2 + \theta - \_}{T}. \quad (9)$$

对于给定一组学习样本数据:

$$\{((x_{1,k}, x_{2,k}), O_k \mid k = 1, 2, \cdots, p\}, \quad (10)$$

将其代入 (10) 式，得到以 w_i ($i = 1, 2$) 为未知元的参

数为 _ , T 的线性方程组,其学习算法类似文献 [6],最终通过调整参数 _ , T, 可得到 XOR码组 (x_1, x_2).

以上重点以 XOR问题为例, 讨论了变参数动态激活函数的选取方法,有关这方面的其它应用见文献 [7~ 9].

4 结束语

本文提出一种新神经元数学模型 –VTAF神经元模型, 是一种能同时模拟包括 M–P与 TAF神经元在内的各种神经元通用的新数学模型.文中还给出这种神经元数学模型的一般形式, 在这种模型中, 激活函数可动态地选取, 一旦选定, 加入参数改善函数的响应特性.由于这种神经元构成的网络比目前常用的网络拥有更大的自由度,对推动神经网络新模型及理论的研究有很大的促进作用.

参考文献

1 吴佑寿,赵明生,丁晓青.一种激励函数可调的新人工神经网络及应用.中国科学 (E辑), 1997, 27(1): 55~ 60.

2 吴佑寿,赵明生.激活函数可调的神经元模型及其监督学习与应用.中国科学 (E辑), 2001, 31(3): 263~ 272.

3 龙欣海,肖田元,陈晓峰.采用变参数激励函数的人工神经网络.计算机工程, 2002, 27(12): 71~ 73.

4 Stork D G, A J D, et al. How to solve the N–bit parity problem with two hidden units. Neural Networks, 1992, 5 923~ 926.

5 Hornik K. Approxomation capabilities of multiplayer feed–forward networks. Neural Networks, 1991, 4 251~ 257.

6 李洪兴.数学神经网络 (II).北京师范大学学报, 1997, 33 (1): 35~ 42.

7 周永权.基于代数神经网络的多元多项式近似因式分解模型及学习算法.计算机研究与发展, 1999, 36(6): 668~ 674.

8 周永权.前向代数神经网络的函数逼近理论及学习算法.计算机研究与发展, 2000, 37(3): 264~ 271.

9 周永权,赵 斌.基于 FLANN自组织多项式网络学习算法.计算机研究与发展, 2001, 38(5): 587~ 590.

(责任编辑: 黎贞崇)

(上接第 15页 Continue from page 15)

的充要条件是存在一个正常数 k,使得

$$\int_{t_0}^{\infty} s^{n-1} | f(s,k) | \mathrm{d}s < \infty, \quad (3.6)$$

式中, $j \in \{1, 2, \cdots, \frac{n-1}{2}\}$.

注 3.1 本文所有的结论可以很容易地推广到如下形式的方程中去:

$$(x(t) + x(t - \tau))^{(n)} + f(t, x(t - \sigma_1), \cdots, x(t - \sigma_n)) = 0, t \geqslant t_0.$$

参考文献

1 Erbe L H, Kong Qingkai, Zhang B G. Oscillation theory for functional differential equations. New York: Marcel Dekker, 1995. 324~ 337.

2 Agarwal R P, Tang X H, Wang Z C. The existence of positive solutions to neutral differential equations. J Math Anal Appl, 1999, (240): 446~ 467.

3 Li W T, Fei X L. Classifications and existence of positive solutions of higher–order nonlinear delay differential equations. Nonlinear Analysis, 2000, (41): 433~ 445.

4 El–Metwally H, Kulenovic M R S, Had Siomerspahic. Nonoscillation solutions for system of neutral equation. Nonlinear Analysis, 2003, 54 63~ 81.

5 Ladde G S, Lakshmikantham V, Zhang B G. Oscillation theory of differential equations with deviating arguments. New York Marcel Dekkev, 1987.

(责任编辑: 黎贞崇)