关于单体模糊神经网络感知机收敛定理的讨论

梁久祯[1], 何新贵[2], 黄德双[2]
(1. 浙江师范大学计算机科学研究所, 浙江金华 321004; 2. 北京系统工程研究所, 北京 100101)

摘 要: 考虑单体模糊神经网络模型, 讨论了单体模糊神经网络的感知机收敛定理与传统神经网络感知机收敛定理的关系, 指出了在证明单体模糊神经网络的感知机收敛定理时存在的问题, 提出了两种修改方案, 即给出了两个改进的感知机收敛定理, 并给出相应的证明和说明.

关键词: 模糊; 神经网络; 感知机; 学习算法; 单体模糊神经网络
中图分类号: TP183 文献标识码: A 文章编号: 0372-2112 (2002) 03-0407-03

On the Perceptron Convergence Theorem of the Monolithic Fuzzy Neural Networks

LIANG Jiu-zhen[1], HE Xin-gui[2], HUANG De-shuang[2]
(1. *Computer Science Institute, Zhejiang Normal University, Jinhua, Zhejiang* 321004, *China;*
2. *Beijing Institute of System Engineering, Beijing* 100101, *China*)

Abstract: The monolithic fuzzy neural networks model is considered. The contrasts are showed between the monolithic fuzzy neural networks and the traditional neural networks on the perceptron convergence theorem. Some issues are discussed on the perceptron convergence theorem of monolithic fuzzy neural networks. Two improving methods are investigated, which are presented as two modified perceptron convergence theorems, and the corresponding arguments or explanations support these modifications.

Key words: fuzzy; neural networks; perceptron; learning algorithm; monolithic fuzzy neural networks

1 引言

单体模糊神经网络(MFNNs)是将传统的神经网络中$<\cdot,\Sigma>$算子改为$<\wedge,\vee>$算子而构成的一种模糊神经网络. 由文[1]知, Zadeh算子$<\wedge,\vee>$为模糊神经元算子, 满足对输入信息处理的交换律、结合律和0-元律. 不同于$<\cdot,\Sigma>$算子型神经网络, MFNNs不是将所有的输入进行加权累积, 而是由于取小($\wedge$)和取大($\vee$)运算, 从而忽略了输入和权值的部分信息, 因此, MFNNs与传统的神经网络的计算和映像机理有很大区别. 近年来有人致力于MFNNs的研究, 并将该网络应用于智能控制[2,3]. 但是, 在对MFNNs的研究中也存在许多未解决的问题, 如感知机收敛定理对MFNNs是否成立? 单体模糊神经网络能否直接求解"异或"问题? 单体模糊神经网络的非线性映像能力怎样? 传统神经网络的结果不能直接照搬到MFNNs上, 否则会得出一些比较荒谬的结论. 对这些问题认识不清楚或简单地挪用$<\cdot,\Sigma>$算子型神经网络的结论将对单体模糊神经网络的应用带来不良的后果, 因此很有必要开展对这些问题的进一步讨论.

本文在文[2]的基础上展开对上述问题的进一步讨论, 指出文[2]的不足和若干错误. 文中重点讨论了单体模糊神经网络的感知机收敛定理, 提出了两种修改方案, 并给出相应的证明和说明.

2 单体模糊神经网络模型

考虑多输入单输出的单体模糊神经元, 对第i个神经元, 有M个输入$x_j\in R(j=1,2,\cdots,M)$, 一个输出$y_i\in\{-1,+1\}$. 连接权值为$\{w_{ji},j=1,2,\cdots,M\}\in R$, 阈值$t_i\in R$. 若记$X=(x_1,x_2,\cdots,x_M)$, $W=(w_{1i},w_{2i},\cdots,w_{Mi})$, 则

$$y_i=f(\sigma_i-t_i) \tag{1}$$

其中σ_i为节点i的净输入, 即

$$\sigma_i=X\circ W=\bigvee_{j=1}^{M}(x_j\wedge w_{ji}) \tag{2}$$

"$\wedge$"代表求小运算, "$\vee$"代表求大运算, $f(\cdot)$为硬限幅函数或sigmoid型函数. 由上述单体模糊神经元按照一定的拓扑结构组成的神经网络称单体模糊神经网络.

3 单体模糊感知机的收敛定理

文[1]考虑了如下单体模糊神经元的模型: M个输入$X\in R^M$, 一个输出Y, 连接权值W, 节点函数采用硬限幅形式:

$$Y=\mathrm{sgn}(W\circ X)=\begin{cases}+1, & \bigvee_{m=1}^{M}(w_m\wedge x_m)\geq 0\\ -1, & \bigvee_{m=1}^{M}(w_m\wedge x_m)<0\end{cases} \tag{3}$$

收稿日期: 1999-11-29; 修回日期: 2001-06-25

设 D 为样本期望输出, 权值修改法则为

设初始权值 $W^0 \neq 0$;

若 $W^k \circ X < 0$ 且 $D > 0$, 则 $W^{k+1} = W^k + X$; (4)

若 $W^k \circ X > 0$ 且 $D < 0$, 则 $W^{k+1} = W^k - X$; (5)

其它情况, 则 $W^{k+1} = W^k$. (6)

定义 1 若存在 W 及 t 使得 $f = \mathrm{sgn}(W \circ X - t)$, 则称函数 $f: R^n$(或 R^n 的子集) $\to \{-1, +1\}$ 为线性可分.

文[1]将模糊感知机收敛定理表述为: 若被学习函数是线性可分的, 经过预处理使样本幅值满足 $\bigvee_{m=1}^{M} x_m = 1, x_m \in [0, 1]$, $(m = 1, 2, \ldots, M)$, 则模糊感知机通过学习过程(4) - (6)在有限次迭代后可收敛到正确的数值.

然后, 文[1]利用与文[4, 5]中感知机收敛定理的证明同样的方法叙述了该定理的证明过程. 定理证明中对样本及函数作如下简化:

(a) 令 $\| X \| = 1$(即训练样本都是单位向量)

(b) $t = 0$

(c) 若 $Y < 0$, 则用 $-X$ 代替 X, 使 $Y > 0$

笔者认为, 其中第(c)条假设是不成立的. 否则

$Y < 0 \Leftrightarrow \bigvee_{m=1}^{M} (w_m \wedge x_m) < 0 \Leftrightarrow w_m \wedge x_m < 0 \overset{x_m \in [0,1]}{\Leftrightarrow} w_m < 0, (m = 1, 2, \ldots, M)$

另一方面, 当用 $-X$ 代替 X 时, $-x_m \in [-1, 0]$, 故必有

$w_m \wedge (-x_m) < 0, (m = 1, 2, \ldots, M) \Rightarrow \bigvee_{m=1}^{M} (w_m \wedge (-x_m)) < 0 \Leftrightarrow Y < 0$

这与 $Y > 0$ 矛盾, 因此第 c 条假设不成立. 这说明文[1]给出的三条假设(a)、(b)、(c)与定理表述中的 $x_i \in [0, 1]$ 是不兼容的, 因此需对此进行改造. 下面提出两种改造方案, 表述为两个定理.

定理 1 若被学习函数是线性可分的, 经过预处理使样本幅值满足 $\bigvee_{m=1}^{M} x_m = 1, x_m \in [-1, 1]$, $(m = 1, 2, \ldots, M)$, 且存在 $1 \leqslant p \leqslant M$, 使得 $x_p < 0$, 则模糊感知机通过学习过程(4) - (6)在有限次迭代后可收敛到正确的数值.

证明: 将学习过程简化为:

(1) 置 $k = 1$, 初始化 $W^k \neq 0$;

(2) 任取 $i \in \{1, 2, \ldots, N\}$, $X^k = (x_{i1}, x_{i2}, \ldots, x_{iM})$;

(3) 若 $W^k \circ X^k \geqslant 0$, 则返回(2), 否则执行(4);

(4) $W^{k+1} = W^k + X^k$, $k = k + 1$; 返回(2).

在这里, 对样本及函数作的简化同上(a)、(b). 当 $Y > 0$ 时, 证明过程同文[1]; 当 $Y < 0$ 时, 以 $-X^k = (-x_{i1}, -x_{i2}, \ldots, -x_{iM})$ 代替 $X^k = (x_{i1}, x_{i2}, \ldots, x_{iM})$, 则

$$Y < 0 \Leftrightarrow \bigvee_{m=1}^{M} (w_m \wedge x_m) < 0 \Leftrightarrow w_m \wedge x_m < 0, (m = 1, 2, \ldots, M) \Rightarrow w_p \wedge x_p < 0$$

若 $w_p \geqslant 0$, 则

$$w_p \wedge (-x_p) \geq 0 \Rightarrow \bigvee_{m=1}^{M} w_m \wedge (-x_m) \geq 0 \Leftrightarrow Y > 0$$

若 $w_p < 0$, 由于 $-x_p > 0$, 由上述迭代过程(2) ~ (4)知, $w_p^{k+1} = w_p^k + (-x_p)$ 经有限次迭代后, 必有 $w_p \geqslant 0$. 因此, 这时上述假设(c)成立. 定理证毕.

注: 定理 1 对样本的输入突破了限制 $x_i \in [0, 1]$, 而且必须存在 $x_p < 0$, 这对要求模糊输入 $x_i \in [0, 1]$ 的情况不易实现, 故提出如下第二种改造方案.

将条件假设(b)改为 $t > 0$. 由于 $x_i \in [0, 1]$, 所以当 $Y < 0$ 时我们不能以 $-X$ 代替 X 就能得到 $Y > 0$, 故简单的假设(c)一般不成立.

将 $Y > 0$ 的输入点记为 $X^+ = \{X_1^+, X_2^+, \ldots, X_{n_1}^+\}$; 将 $Y < 0$ 的输入点记为 $X^- = \{X_1^-, X_2^-, \ldots, X_{n_2}^-\}$, 则 $n_1 + n_2 = N$ 为样本个数. $\forall X_i^+ \in X^+$, $X_i^+ = \{x_{1i}^+, x_{2i}^+, \ldots, x_{M,i}^+\}$ 由

$Y > 0 \Leftrightarrow \bigvee_{m=1}^{M} (w_m \wedge x_{mi}^+) \geq t \Leftrightarrow \exists p_i$ 使 $w_{p_i} \wedge x_{p_i,i}^+ = \bigvee_{m=1}^{M} (w_m \wedge x_{mi}^+) \geq t \Leftrightarrow w_{p_i} \geq t$ 且 $x_{p_i,i}^+ \geq t$, $i = 1, 2, \ldots, n_1$.

另一方面, $\forall X_j^- \in X^-$, $X_j^- = \{x_{1j}^-, x_{2j}^-, \ldots, x_{M,j}^-\}$, 由 $Y < 0 \Leftrightarrow \bigvee_{m=1}^{M} (w_m \wedge x_{mj}^-) < t \Leftrightarrow \forall m, w_m \wedge x_{mj}^- < t, j = 1, 2, \ldots, n_2$.

若 $\{p_1, p_2, \ldots, p_{n_1}\} = \{1, 2, \ldots, M\}$, 则 $w_1, w_2, \ldots, w_M \geqslant t$, 于是有

$$x_{mj}^- < t, \text{对 } m \in \{1, 2, \ldots, M\} \tag{7}$$

若 $\{p_1, p_2, \ldots, p_{n_1}\} \neq \{1, 2, \ldots, M\}$, 对指标集 $\{p_1, p_2, \ldots, p_{n_1}\}$ 进行整理, 使其中的元素两两不等, 即

$\{p_1, p_2, \ldots, p_{n_1}\} = \{s_1, s_2, \ldots, s_{n_3}\}$, 其中 $s_i \neq s_j$, 对 $i \neq j$. 不妨设

$\{1, 2, \ldots, M\} = \{s_1, s_2, \ldots, s_{n_3}\} \cup \{q_1, q_2, \ldots, q_{n_4}\}$, 且 $\{s_1, s_2, \ldots, s_{n_3}\} \cap \{q_1, q_2, \ldots, q_{n_4}\} = \phi$

其中 $q_i \neq q_j$, 对 $i \neq j$. 则对 $m = s_1, s_2, \ldots, s_{n_3}$ 有 $w_1, w_2, \ldots, w_{n_3} \geqslant t$ 且

$$x_{s_1,j}^-, x_{s_2,j}^-, \ldots, x_{s_{n_3},j}^- < t \tag{8}$$

对 $m = q_1, q_2, \ldots, q_{n_4}$, 有

$$w_m \wedge x_{mj}^- < t \Leftrightarrow w_m < t \text{ 或 } x_{mj}^- < t \tag{9}$$

由上述分析知, 若取 $t \in (0, 1)$, 需要对样本输入作如下预处理:

(1) 若使 $Y > 0$ 的输入点 $\{p_1, p_2, \ldots, p_{n_1}\} = \{1, 2, \ldots, M\}$, 则使 $Y < 0$ 的输入点必须满足式(7);

(2) 若 $\{p_1, p_2, \ldots, p_{n_1}\} \neq \{1, 2, \ldots, M\}$, 则对 $m = s_1, s_2, \ldots, s_{n_3}$, 使 $Y < 0$ 的输入点必须满足式(8);

(3) 若 $\{p_1, p_2, \ldots, p_{n_1}\} \neq \{1, 2, \ldots, M\}$, 则对 $m = q_1, q_2, \ldots, q_{n_4}$, 使 $Y < 0$ 的输入点必须满足式(9).

由于对式(7)、(8)两种情况输入输出关系与权值的调整无关, 所以只对式(9)的情况考虑学习过程. 即

(Ⅰ) 初始化 w_m 使其对 $m = s_1, s_2, \ldots, s_{n_3}$ 有 $w_1, w_2, \ldots, w_{n_3} \geqslant t$

(Ⅱ) 任取 $m \in \{q_1, q_2, \ldots, q_{n_4}\}$, $j = 1, 2, \ldots, n_2$;

(Ⅲ) 若 $w_m \wedge x_{mj}^- < t$ 则返回(Ⅱ), 否则执行(Ⅳ);

(Ⅳ) $w_m = w_m - x_{mj}^-$; 返回(Ⅱ).

迭代过程(Ⅱ)－(Ⅳ), $w_m = w_m - x^{-}_{mj}$ 经有限次迭代后, 必有 $w_m \wedge x^{-}_{mj} < t$, $m = q_1, q_2, \cdots, q_{n_4}$. 故模糊感知机收敛定理可修改为:

定理 2　若被学习函数是线性可分的, 经过预处理使样本幅值满足 $\bigvee_{m=1}^{M} x_m = 1$, $x_m \in [0, 1]$, $(m = 1, 2, \cdots, M)$, 及式(7)、(8), 则模糊感知机通过学习过程(Ⅰ)－(Ⅳ)在有限次迭代后可收敛到正确的数值.

参考文献:

[1] 李晓忠, 汪培庄, 罗承忠. 模糊神经网络 [M]. 贵州: 贵州科技出版社, 1994.

[2] 王振峰, 靳东明, 李志坚. 单体模糊神经网络自学习问题研究 [J]. 电子学报, 1997, 25(2): 33－38.

[3] 王振峰, 靳东明, 李志坚. 单体模糊神经网络智能控制系统及其 VLSI 研究 [D]. 博士学位论文, 北京: 清华大学, 1995.

[4] 阎平凡, 黄端旭. 人工神经网络——模型、分析与应用 [M]. 合肥: 安徽教育出版社, 1993.

[5] 焦李成. 神经网络系统理论 [M]. 西安: 西安电子科技大学出版社, 1996.

作者简介:

梁久祯　男. 1968 年生于山东东平. 1991 年 7 月毕业于大庆石油学院获理学学士学位, 1996 年 3 月毕业于哈尔滨工业大学获理学硕士学位, 2001 年 6 月毕业于北京航空航天大学, 获工学博士学位, 研究方向为: 智能计算及其应用.

何新贵　男. 1938 年生于浙江浦江. 博士生导师, 研究员, 研究方向为: 模糊理论与技术.

黄德双　男. 1964 年生于安徽肥东. 博士, 教授, 研究方向为: 神经网络与信号处理.