

# Madaline 网络学习算法改进

盛新新, 颜伟琼, 王炳辉

(河海大学 计算机及信息工程学院, 江苏 南京 210098)

**摘要:**根据 Madaline 网络工作原理, 针对其网络特点和现有算法中存在的缺点, 提出将 MR II 算法与感知机学习规则相结合的改进的学习算法, 使每个样本学习后权值的改动尽量小, 从而减少它对其它样本学习产生的影响, 达到提高网络学习能力的目的。

**关键词:** Madaline 网络; MR II 学习算法; 感知机学习规则; 学习速率

**中图分类号:** TP183

**文献标识码:** A

**文章编号:** 1673-629X(2006)06-0224-03

## An Improved Learning Algorithm for Madaline Network

SHENG Xin-xin, YAN Wei-qiong, WANG Bing-hui

(Computer and Information Engineering College of Hohai University, Nanjing 210098, China)

**Abstract:** In this paper, according to the network characteristic and the shortcoming of the existing algorithm, by analyzing the Madaline network work principle and its learning algorithm MR II, propose an improvement algorithm that combines the MR II algorithm and the perception learning rule that will improve the performance of the network learning.

**Key words:** Madaline network; MR II learning algorithm; perception learning rule; learning rate

### 0 引言

神经网络发展至今天,吸引了许多科学家在这个领域进行研究,设计一个神经网络的重点在于网络结构的构成和学习算法的选择。一般来说,结构是根据所研究领域及要解决的问题确定的,因此对学习算法的研究就显得更有价值。所谓学习算法就是在网络学习过程中,不断地调整网络参数,直到输出结果满足要求。一个适合网络模型的学习算法将大大降低它的计算复杂度,提高网络的运行效率,这无疑具有重要的理论和现实意义。

最初的 Madaline 网络只有两层,后来, Madaline 由许多层神经元组成。每层的权值是可调的,且每层的输入输出均是二值。这样就使均方误差函数中存在许多的不连续点,它们是不可导的,因此不能再用 LMS 算法来解决,也不能用最速下降法或最大梯度来进行权系数调整。20 世纪 60 年代 Widrow 和 Winter<sup>[1]</sup>提出的针对 Madaline 网络特点的算法 MR II (Madaline rule II),其重点在于最小干扰原则的提出;国内的文献很少提到 MR II 算法,即使在引用 MR II 算法时,也未作详细阐述。文献[2,3]中提到的 MR II 算法是用 LMS 调权,但是由于 LMS 算法只适用于连续值,而 MR II 在比较误差时用的是输出层的出错个数,而不是出错误差的大小,这就使得 LMS 算法的使用受

到很大的局限;并且 LMS 算法的步幅  $a$  的选择有一定的随机性,若选择不当,则收敛速度很慢,甚至出现振荡现象。基于以上的原因,发现如果用感知机学习规则来实现翻转由 MR II 算法确定的隐层中需要翻转的 Adalines (adaptive linear element) 的输出,保持网络各层的输入输出均用二值的数字量进行计算和传递,将大大降低计算的复杂度;为了使前个样本学习所调整的权值对后一样本学习的影响尽量小,在感知机学习规则中提出学习速率这一概念,使每次权值的改变小于感知机学习规则所要求的权值的改变,降低了样本学习之间的互相影响,有效提高了网络学习效率。

### 1 Madaline 网络结构及工作原理

Adaline 是具有线性功能函数的神经元,是一种自适应线性元神经网络模型,其输入信号写成向量的形式  $X_K = [x_{0K}, x_{1K}, x_{2K}, \dots, x_{NK}]^T$ , 这个向量称为 Adaline 的输入信号向量或输入模式向量。一组输入信号相应的有一组权值  $W_K = [w_{0K}, w_{1K}, w_{2K}, \dots, w_{NK}]^T$ , 这个向量称为权向量。Adaline 的输出分为模拟输出和数字输出两种。其模拟输出量定义为  $y_k = w_k \cdot x_k$ , 数字输出量为  $q_k = \text{sgn}(y_k)$  (当  $y_k \geq 0$  时,  $q_k = 1$ ; 当  $y_k < 0$  时,  $q_k = -1$ )。Adaline 的训练过程是:将理想响应信号  $d_k$  送入 Adaline 中,与数字输出量  $q_k$  进行比较,并将差值送到学习机制中,以调整权向量  $W_K$ ,使得  $q_k$  和所要求的输出  $d_k$  相一致。若一致,就说明 Adaline 已训练完毕。

收稿日期:2005-09-13

作者简介:盛新新(1971-),女,江苏南京人,硕士研究生,研究方向神经网络;导师:曾晓勤,教授,研究方向神经网络和模式识别。

但是由于 Adaline 和感知机受同样的局限性影响: 只能解决线性可分问题, 而使单个的 Adaline 无法完成非线性划分问题, 为了克服硬限幅函数只能对两个线性可分集合或相交集合的分类, 需要将多个 Adaline 联结起来, 即用多层自适应线性网 Madaline (Many Adaline) 来实现非线性可分的逻辑函数。Madaline 是一种离散型的多层前向网络, 它由许多层神经元组成, 每层的权值系数是可调的, 且每层的输入和输出均为二值, 前一层的输出将作为后一层的输入, 是神经网络中应用最为典型和广泛的网络之一。图 1 为一个四层的 Madaline 网络。

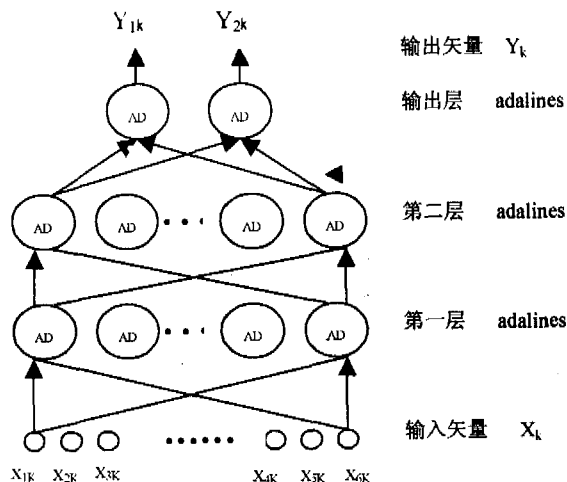


图 1 一个四层的 Madaline 网络

## 2 学习算法改进

### 2.1 MRII 算法和感知机学习规则

MRII 算法的算法思想是<sup>[1]</sup>: 给网络提供输入模式向量以及相应的理想输出, 训练的目标是尽可能地找出所有隐层中每个 Adaline 的理想输出, 使它转化为对每个 Adaline 的学习训练。当第一个训练模式提供给神经网络时, 网络得到一个输出, 比较该输出与理想输出的出错个数, 如果相同, 说明训练成功, 进入下一个样本学习; 如果不同, 按最小干扰原则, 在第一层上找出模拟量绝对值最小的那个神经元, 翻转它的数字量 (使其输出从 0 变为 1 或从 1 变为 0), 将翻转后的二值输出作为该 Adaline 的理想输出, 这样第二层的输入就发生了变化, 最终网络的输出也就改变了。逐层计算之后检查这是否减少了输出出错个数, 如果出错个数没有减少, 则恢复原来的数据; 如减小了, 这个尝试就被接收, 说明该神经元的翻转可以使输出更理想。在这种情况下, 假设翻转后的神经元输出就是该神经元的理想输出, 修改它的权值, 用新权值代替旧权值。如果这个神经元的翻转只是减少了出错个数并不能得到全部的理想输出, 按最小干扰原则找到加权 and 次接近于零的神经元, 然后对它重复上面的过程。如果单个这样的尝试完毕后仍未得到全部理想的要求, 再将第一层的神经元按最小干扰原则两个一组重复以上过程, 即将两个加权和最小的神经元同时翻转, 如果没有改变恢复加权和最小

的, 翻转加权和第三小的神经元, 计算输出情况, 依次三个一组、四个一组, 直到  $K$  个一组 ( $K$  为第一层的神经元个数), 所有这些进行完毕之后, 如误差仍不满足要求, 就在第二层上重复第一层的过程一直到输出层。

感知机学习规则是 Hebb 学习规则的变形。Hebb 学习规则的数学描述:  $W_{ij}$  表示神经元输入  $P_i$  到输出  $O_j$  的突触权值, Hebb 假设意味着, 如果一个正的输入要产生一个正的输出, 就应该增加  $W_{ij}$ , 这就是该假设的数学解释, 即:  $W_{ij}^{new} = W_{ij}^{old} + T_{iq}P_{jq}$ ,  $T_{iq}$  是第  $q$  个目标向量的第  $i$  个元素。感知机学习规则也遵循了这个基本数学思想。该算法是神经网络学习的一个重要算法, 并已被广泛应用<sup>[4]</sup>。现介绍如下:

感知机学习规则:

- (1) 选择一组初始权值  $w_i(0)$ ;
- (2) 计算某一输入模式对应的实际输出与期望输出的误差  $e$ ;
- (3) 如果  $e = 0$ , 输入下一个训练样本返回 (2), 否则继续;
- (4) 更新权值 (阈值可视为输入恒为 1 的一个权值);
- (5) 返回 (2); 重复, 直到对所有训练样本模式、网络输出均能满足要求。

算法中  $e$  是感知机的误差变量, 它是理想输出  $t$  与实际输出  $a$  之间的差值, 因此  $e$  的值会有以下 3 种情况 (假设它的数字量分别是 1 和 0):

\* 如果理想输出是 1 而实际输出是 0, 即  $e = 1 - 0 = 1$

\* 如果理想输出是 0 而实际输出是 1, 即  $e = 0 - 1 = -1$

\* 如果理想输出等于实际输出, 即  $e = 0$

这样, 把它的学习表达式分别写成以下 3 条规则:

当  $e = 1$  时,  $W_{ij}^{new} = W_{ij}^{old} + P_{jq}$

当  $e = -1$  时,  $W_{ij}^{new} = W_{ij}^{old} - P_{jq}$

当  $e = 0$  时,  $W_{ij}^{new} = W_{ij}^{old}$

由此, 它统一的权值更新计算可用以下表达式表示<sup>[3]</sup>:  $W_{ij}^{new} = W_{ij}^{old} + eP_{jq} = W_{ij}^{old} + (t - a)P_{jq}$ 。虽然感知机的学习规则非常简单, 但它的功能却十分强大, 它的有效性早已得到广泛验证。

### 2.2 学习算法改进

根据以上分析, 在 MRII 算法思想中, 最小干扰原则占有重要地位, 其目的是为了权值的改动尽量的小, 从而这个样本的学习不会对其它样本的学习造成大的影响。基于这个思想, 提出了学习速率这一概念, 目的是让一步就使神经元翻转的权值变化量分若干步进行迭代变化, 一旦新的权值达到了翻转的目的就退出迭代, 保证了在某一学习速率下权值的改动最小。实验证明用这个方法修改的权值远远小于其它的一些调权方法, 学习的收敛速度也大大提高。在 Widrow 和 Winter 的 MRII 算法中, 调权用了以下公式<sup>[5]</sup>:

$$W(k+1) = W(k), \text{ for } d(k)X^T(k)W(k) \geq \delta$$

$$W(k+1) = W(k) + \frac{\gamma d(k)}{n+1} X(k)[L - d(k)X^T(k)W(k)], \text{ for } d(k)X^T(k)W(k) < \delta$$

对式中各个参数的选取文中给出了具体的经验值。这个方法和感知机学习规则都遵循了 Hebb 学习规则的数学描述,新的权值是旧的权值加上或减去输入:

$$W_{ij}^{\text{new}} = W_{ij}^{\text{old}} + eP_{ij} \quad (e \text{ 可取正负值}) \quad (1)$$

并证明经过一步计算新的权值就可以翻转该神经元的数字输出量。为了使新的权值改动尽量的小,以保证它的改动对前面训练好的样本产生的影响尽量小,将(1)式改写成  $W_{ij}^{\text{new}} = W_{ij}^{\text{old}} + \gamma eP_{ij}$ ,把  $\gamma$  称为学习速率,  $\gamma$  在(0,1)之间取值,  $\gamma$  越小新的权值对旧的权值的改动也越小,但这样可能使一次迭代无法满足翻转神经元的数字输出量的要求,因此必须验证迭代后的新权值输出的正确性,一旦新权值满足了翻转要求,就把它作为下次学习的新的权值,这样既保证了翻转又保证了在特定学习速率下权值的改动最小。

由此可以得到带学习速率的感知机学习规则在 MR II 算法中进行调权的改进的 Madaline 学习算法,其步骤如下:

(1)初始化网络结构包括网络层数和各层结点数,用一个随机数作为网络中的各个权值的初始值。保存这些权值数据。

(2)随机输入一个样本矢量  $X_K$  和它的理想输出  $T_K$ ,按层一步一步地计算出每个 Adaline 的实际输出,分别保留它们的模拟量和数字量。

(3)比较理想的输出  $T_K$  和实际输出的不相同个数,如果出错个数为 0,则到(7);如果出错个数不为 0,则从第一层开始,根据最小干扰原则,即找出第一层中神经元模拟量最接近于 0(即其加权值的绝对值最小)的那个神经元。

(4)翻转该神经元(使其输出从 0 变为 1 或从 1 变为 0),即让它的输出改变符号。逐层计算网络最后的输出,比较它和理想输出之间的误差,如果输出误差个数减少,则接受这个翻转的尝试,记下此时该神经元的输出作为它

的理想输出,用带学习速率的感知机学习规则调整它的权值,每次迭代后新的权值要验证是否已经使该神经元翻转,一旦翻转就保留这个权值到(3);如果输出误差个数没有减少,就不接受它的翻转仍保留原来数据,即上一步改变的输出数字量的符号恢复,到(5)。

(5)转入下一个神经元,即其模拟输出次接近 0 的神经元,仍按(4)规则训练。

(6)当这一层的单个神经元训练结束后,再按两个一组地训练,然后再按三个一组,……,直到  $k$  个一组( $k$  为第一层神经元个数)。如仍然不能符合要求,再用同样的方法训练第二层,第三层……直到输出与要求响应之间的理想输出相吻合。保存符合该样本理想输出的数据。

(7)输入另一个新样本,用同样的方法训练。

### 3 总 结

计算和算法是人类自古以来十分重视的研究领域。在神经网络的发展中,找到适合特定网络模型的最佳算法,有效提高网络的运行能力,是很具有现实意义的工作。Madaline 网络由于具有明显的离散特点,且其学习算法 MR II 也为其改进提供了很好的平台。故恰当地运用这个特点,结合改进的感知机学习规则来有效减少权值的改变量,可达到提高学习能力的目的。同时,MR II 中还有很多细节值得更进一步的研究探讨。

### 参考文献:

- [1] Widrow B, Winter B, Rodney Gerard Madaline RULE II [M]. CA, USA: Stanford University, 1989.
- [2] 胡守仁. 神经网络导论 [M]. 长沙: 国防科技大学出版社, 1993.
- [3] 张立明. 人工神经网络模型及其应用 [M]. 上海: 复旦大学出版社, 1993.
- [4] Hagan M T, Demuth H B, Beale M H. 神经网络设计 [M]. 戴葵等译. 北京: 机械工业出版社, 2002.
- [5] Winter R G. Madaline Rule II: A new method for training networks of Adalines [D]. CA, USA: Stanford University, 1989.

(上接第 223 页)

是对常用搜索引擎的返回结果进行二次处理,消除重复网页,从而大大提高用户的搜索效率,具有较大的实用价值。

### 参考文献:

- [1] 李增智,李平均,王广荣. 计算机网络管理系统的若干重要问题[J]. 微机发展, 2000, 10(2): 7-10.
- [2] 葛新红. 数据挖掘软件应用分析[J]. 微计算机应用, 2005, 26(3): 374-377.
- [3] Shivakumar N, Garcia - Molina H. Finding near - replicas of

documents on the Web [A]. In proceedings of the Workshop on Web Databases [C]. [s.l.]: [s.n.], 1998. 204-212.

- [4] Shivakumar N. SCAM: A copy detection mechanism for digital documents [A]. In proceedings of 2nd International Conference in Theory and Practice of Digital Libraries [C]. Austin, Texas: [s.n.], 1995.
- [5] 王贵竹,李津生,洪佩琳. MD5 报文摘要算法与 IPv6 认证 [J]. 小型微型计算机系统, 2001, 22(1): 126-128.