

用一组单输出的子网络代替多输出的 BP 网络

A New Architecture of Neural Networks Array Replacing of a Multiple Outputs BP Model Networks

王秀坤 张晓峰

(大连理工大学计算机科学与工程系 大连116024)

Abstract In this paper, a new architecture of single output multilayered feedforward neural networks array and a learning algorithm based this architecture are proposed. This architecture can be used to replace of a multiple outputs multilayered feedforward neural networks. Theoretical analyses show this method outperforms the traditional multiple outputs multilayered feedforward neural networks.

Keywords Neural networks, Architecture of networks, Learning algorithm

BP 网络是使用最广泛的神经网络模型^[1]。在 BP 网络模型中,输入信息 $(x_1, x_2, \dots, x_m)$ 首先通过各输入结点前向传播到下一层的各结点,这下一层就是所谓的隐含层,隐含层可以是一层也可以是多层,最后通过输出层得到输出信息 $(y_1, y_2, \dots, y_n)$ 。在 BP 网络中,输入结点和输出结点的个数由具体问题决定,隐层单元的个数是可变的。如果具体问题的目标输出是多维的, BP 网络就具有多个输出结点。本文提出了一种子网组的结构,用来代替多输出的单个 BP 网络。子网组即是对问题目标输出的每一维建立一个单输出的 BP 网络,整个子网组结构的输出由这 n 个子网络组合而成。本文还给出了一种子网组结构的学习算法,并证明了该学习算法能够达到更好的学习效果。

1 BP 网络的结构

BP 网络的结构如图1所示。

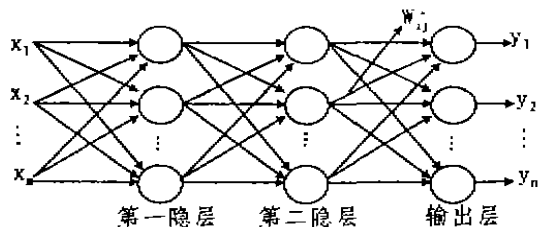


图1 BP 网络的结构

BP 网络每层单元只接受前一层的输出信息并输出给下一层的各结点,各结点的特性为 Sigmoid 型(它是连续可微的)。对于第 l 层的第 j 个结点,当输入第 k

个样本时,结点 j 的输入为:

$$\text{net}_k^j = \sum_i w_{ij} O_k^{i-1} \quad (1)$$

其中 O_k^{i-1} 表示当输入第 k 个样本时, $l-1$ 层第 i 个结点的输出, w_{ij} 表示 $l-1$ 层第 i 个结点和第 l 层第 j 个结点之间连接权的权值,结点 j 的输出为:

$$O_k^j = f(\text{net}_k^j) \quad (2)$$

网络产生的输出误差为:

$$E_k = 1/2 \sum_j (y_k^j - y_k^j)^2 \quad (3)$$

y_k^j 是单元 j 的理想输出, y_k^j 是单元 j 的实际输出,总误差为:

$$E_g = 1/n \sum_k E_k \quad (4)$$

BP 算法或其它改进的前馈神经网络学习算法的目标就是要使总误差 E 达到最小。

2 子网组结构

子网组结构即是对问题目标输出的每一维建立一个单输出的 BP 网络,这一组共 n 个单输出的 BP 网络构成子网组结构。对于任意一个多输出 BP 网络 N 都能够建立相应的子网组结构,其中子网组结构所包含的每一个子网络与 BP 网络 N 都有相同的输入结点个数、相同的隐层数、相同的隐层结点数,但是每一个子网络的输出结点个数都为1。每个输入样本同时作为每一个子网的输入样本,整个子网组结构的输出由这 n 个子网络的输出组合而成 $(y_1, y_2, \dots, y_n)$,其中 y_i 为第 i 个子网的输出。我们把 BP 网络 N 称为子网组结构对应的 BP 网络原型,子网组结构如图2所示。

王秀坤 博士导师,主要研究方向为分布式数据库、决策支持系统,张晓峰 硕士,主要研究方向为分布式数据库、决策支持系统。

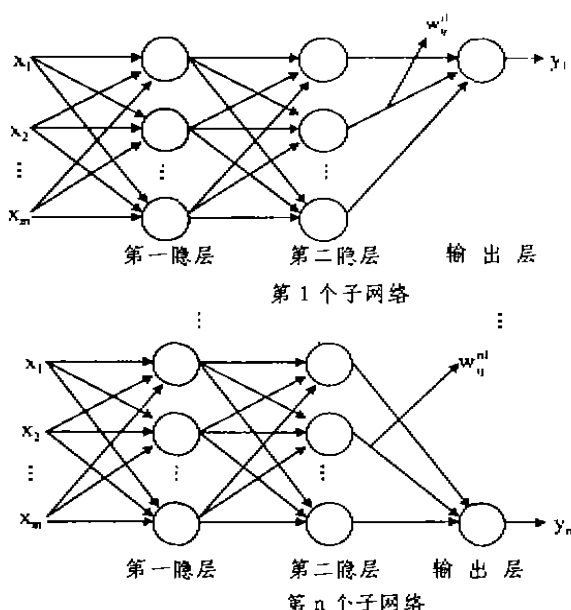


图2 子网组结构

当输入第 k 个样本时,网络产生的输出误差为:

$$E_k = 1/2 \sum_j (y_{jk}^* - y_{jk})^2 \quad (5)$$

y_{jk}^* 是子网 j 的理想输出, y_{jk} 是子网 j 的实际输出。总误差为:

$$E_g = 1/n \sum_k E_k \quad (6)$$

(6)式与该子网组结构对应的 BP 网络原型具有完全相同的形式。

子网组结构与 BP 网络所不同的是 BP 网络中输入层与隐层结点之间、隐层结点之间的连接权同时对所有输出结点的输出起作用,而子网组结构每一维的输出只与各自独立的连接权有关系。

3 子网组结构的一种学习算法及理论分析

3.1 一种学习算法

这种子网组的学习算法主要可以分为三个步骤:

步骤1 使用 BP 算法或其它改进的前馈神经网络学习算法对子网组结构对应的 BP 网络原型进行训练,在训练完成后转入步骤2。

步骤2 首先使用步骤1中 BP 网络原型训练完成后的最终权值初始化子网组结构中各个子网络连接权的权值,具体方法如下

a) 对于输入层与隐层结点之间、隐层结点之间的连接权。由于每一个子网络与 BP 网络原型具有相同的输入结点个数、相同的隐层数、相同的隐层结点数,所以每个子网输入层与隐层结点之间、隐层结点之间的连接权与 BP 网络原型的连接权对应相同。对这部

分连接权,我们只要取 BP 网络原型中对应的连接权的最终权值为其初始权值即可。赋值表达式如下

$$w_{ij}^l = w_{ij}^l \quad (7)$$

其中 p 表示第 p 个子网, w_{ij}^l 表示子网 p 中第 $l-1$ 层第 i 个结点与第 l 层第 j 个结点之间连接权的权值, w_{ij}^l 表示 BP 网络原型中第 $l-1$ 层第 i 个结点与第 l 层第 j 个结点之间连接权的权值。

b) 对于隐层结点和输出结点之间的连接权。由于每个子网络只有一个输出结点,所以这部分连接权的权值只需要取为 BP 网络原型中对应输出结点和隐层结点之间连接权的权值即可。赋值表达式如下

$$w_{ij}^l = w_{ij}^l \quad (8)$$

步骤3 在步骤2完成以后对每个子网分别使用 BP 算法或其它改进的前馈神经网络学习算法对每个子网络进行训练。每一个子网络训练完成以后,整个子网组结构的训练也就完成了。

3.2 理论分析

根据子网组结构的构造方法,对于任意一个 BP 网络我们都可以构造出相应的子网组结构。所以我们只要证明任意一个子网组结构与其相应的 BP 网络原型相比可以达到更好的学习效果即可,也就是证明步骤3完成以后得到的子网络组结构比步骤1中得到的 BP 网络有更好的学习效果。

首先根据步骤2中子网组结构的连接权权值的获得方法,易知步骤2完成以后得到的子网组结构和步骤1中得到的 BP 网络对任意的输入都会得到完全相同的输出结果。所以关键是要证明步骤3能进一步改进子网组的学习效果,我们从两个方面加以考虑。

情况1 由于 BP 算法是基于求解优化问题的梯度下降法的基础之上,所以 BP 算法往往只能求出局部极小点,在局部最小点时,对网络中所有权值 w_{ij}^l 有:

$$\partial E_g / \partial w_{ij}^l = 0 \quad (9)$$

此时梯度下降法已经无法得到更好的权值。BP 网络的网络总误差 E_g 还可以写成如下形式

$$E_g = 1/n (E_{1g} + E_{2g} + \dots + E_{ng}) \quad (10)$$

其中 E_{mg} 代表对于所有样本网络输出第 m 维产生的总误差

$$E_{mg} = 1/2 \sum_k (y_{mk}^* - y_{mk})^2 \quad (11)$$

所以(9)式又可以写成如下形式

$$\partial E_g / \partial w_{ij}^l = 1/n (\partial E_{1g} / \partial w_{ij}^l + \partial E_{2g} / \partial w_{ij}^l + \dots + \partial E_{ng} / \partial w_{ij}^l) = 0 \quad (12)$$

在 BP 网络中输入层与隐层结点之间、隐层结点之间连接权对每个输出结点的输出值都起作用,上式中每一项 $\partial E_{mg} / \partial w_{ij}^l$ ($1, 2, \dots, n$) 都有可能不等于0,所以虽然(12)式中所有项的和等于0,但是其中的任意一项却

不一定等于0。下面我们再看看子网组的情况,由于在子网组结构中每一维输出都是由一个独立的单输出BP网络产生的。子网组结构中任意连接权只对整个子网组结构总的输出中的一维起作用,于是可得

$$\partial E_{\text{总}} / \partial w_{ij}^p = \partial E_{p\text{总}} / \partial w_{ij}^p \quad (13)$$

其中 w_{ij}^p 是子网组结构中的任意连接权的权值, $E_{p\text{总}}$ 代表第 p 个子网产生的总误差。

由于子网组结构的构造方法可知,子网组结构每一维输出与其相应的BP网络原型相比都是与相同数量的连接权有关系,与这些连接权之间关系表达的形式是完全相同,每一维输出产生的总误差与这些连接权之间关系表达的形式也就完全相同,因此总误差与这些连接权的偏导数也有相同的关系表达式。这些连接权之间是一一对应的。又根据步骤2中子网组结构初始权值的获得方法可知,于是我们可以得到如下等式

$$\partial E_{p\text{总}} / \partial w_{ij}^p = \partial E_{m\text{总}} / \partial w_{ij}^p \quad (14)$$

其中 $p=m$, w_{ij}^p 代表在步骤2完成以后第 p 个子网中第 $l-1$ 层第 i 个结点与第 l 层第 j 个结点之间连接权的权值, $E_{p\text{总}}$ 代表第 p 个子网输出产生的总误差, w_{ij}^m 代表步骤1中得到的BP网络中第 $l-1$ 层第 i 个结点与第 l 层第 j 个结点之间连接权的权值, $E_{m\text{总}}$ 代表步骤1中得到的BP网络第 m 个输出结点产生的总误差。

上面已经说明如果 w_{ij}^p 为输入层与隐层结点之间或隐层结点之间的连接权,则 $\partial E_{m\text{总}} / \partial w_{ij}^p$ 有可能不等于0,根据(13)式和(14)式可知对于子网组结构 $\partial E_{\text{总}} / \partial w_{ij}^p$ 有可能不为0,即当步骤3开始时每一个子网络都有可能尚未达到局部极小,所以基于梯度下降的BP算法仍能够进一步得到更好的解,而步骤3完成以后得到的子网组结构比步骤1中得到的BP网络就有了更好的学习效果。

情况2 假设在某些神经网络学习算法能够达到全局最小点,但是由于子网组结构的误差函数有可能有函数值更小的全局最小点,所以我们仍可以希望子网组结构能达到更好的学习效果。现证明如下:

我们可以把子网络组结构的学习问题看成求 $E_{\text{总}}$ 最小值的优化问题

$$\min E_{\text{总}} = f(w) \quad (15)$$

其中 W 代表子网络组中所有连接权组成的向量。子网络组结构相应的BP网络原型的学习问题,可以看成如上优化问题加上一组等式约束条件所构成的另一个优化问题,这一组约束条件具有如下形式

$$w_{ij}^1 = w_{ij}^2 = \dots = w_{ij}^m \quad (16)$$

其中 w_{ij}^1 代表第1个子网中第 $l-1$ 层第 i 个结点与第 l 层第 j 个结点之间连接权的权值,第 $l-1$ 层是输入层或隐层,第 l 层是隐层, w_{ij}^2 代表第2个子网中相同位置连接权的权值,其余依此类推。这一组约束条件的意义

就是不同子网的输入层与隐层结点之间、隐层结点之间连接权权值对应相等。

由步骤2中方法可知,BP网络原型可以实现的映射关系,子网组结构都可以实现,即BP网络原型的误差函数的值域是子网组结构的误差函数值域的一子集。但是如果子网组结构的误差函数其全局最小值所在的点不满足上述一组约束条件,则BP网络原型的误差函数就无法取得此全局最小值,即子网组结构的误差函数的全局最小值小于其相应的BP网络原型的误差函数的全局最小值。

Hornik等人^[2]已经证明,只需一个具有单隐层的前馈网络就可以逼近任意复杂度的函数,但在实际应用中,隐结点的数量不能太多,网络误差函数的最优解并不能无限趋近于0,所以与本文上述结论并不矛盾。

综上所述,我们证明了子网组结构比多输出的BP网络能达到更好的学习效果,3.1节所提出的学习算法的一个主要目的是证明子网组结构的有效性,在实践中也可以让每个子网络直接使用随机的初始权值进行训练。

4 实验结果

本文对如下编码问题进行训练。该问题有6位输入,每一位为0或1,其中前3位为干扰位,后3位为信息位,输出为2位,每一位也为0或1。输入和输出的映射关系如表1所示。

表1 编码问题输入输出映射表

输入	输出	输入	输出
* * * 000	00	* * * 100	11
* * * 001	01	* * * 101	10
* * * 010	10	* * * 110	01
* * * 011	11	* * * 111	00

该问题共有64个样本,取40个作为训练样本,24个作为检验样本。本文分别使用传统的BP网络方法和子网络组结构对该问题进行训练,当使用传统的6输入2输出的BP网络方法时,如果2位输出位的误差都小于0.3,则认为结果正确。BP网络方法对训练样本的正确率最好为87.5%,对检验样本的正确率最好为79.17%。当使用子网络组结构时,对两个子网直接使用随机的初始权值和BP算法分别进行训练,如果2位输出位的误差都小于0.05,则认为结果正确。子网络组结构法对训练样本的正确率最好为100%,对检验样本的正确率最好也为100%。

上述结果表明,子网络组结构与传统的多输出的

(下转第18页)

再有任何的数据拷贝,就这一点来说,VIA 实现了真正的零数据拷贝。

VIA 的软件模拟与硬件实现

目前的商业网络接口卡还没有实现对 VIA 的硬件支持,M-VIA 是在软件层支持 VIA 规范的,它利用 ERCL 实现了这一层的功能。硬件集成 VIA 的功能需要网络接口卡能够识别 VI 描述符格式,能够支持硬件门铃机制,能够一定程度地实现卡上的 VI 调度和管理。VIA 标准的制定之初也是为了整个体系足够简单以使其能够在硬件中实现其功能。构建于 VI-NIC 之上的 VIA 会更大程度地提高机群通讯的性能。

国内 VIA 的研究成果

清华大学计算机系已经开始 VIA 规范和 VI-NIC 的研究,并且已经取得了初步的成果。THVIA^[9]是基于 VIA 规范,借鉴 M-VIA,由清华大学计算机系,根据实际情况实现了部分功能的 VIA 系统,该系统在硬件上一定程度实现了 VI-NIC 的支持,包括根据描述符进行轮询发送,VI 门铃机制等等;在软件上实现了半用户层通讯,即发送通过用户层,而接收还经由操作系统。该系统正处于初步实现的阶段,在不久之后就会推出能够运行的版本。

总结 虚拟接口体系 VIA 实现了用户层通讯的工业规范,它向用户提供了简单的数据操作集、方便的应用编程接口;它消除了消息传输关键路径上的多次数据拷贝,实现了真正意义上的零拷贝;除了核心代理的必要的功能之外,它的所有功能都在用户层实现,避免了项操作系统核心的陷入、避免了进程上下文切换带来的 CPU 周期的开销和可能由此引起的 cache 抖动;它使进程通过虚拟接口能直接访问网卡的硬件资

源,实现了并发进程对网络接口的复用;它采用了对短消息的特殊处理,这一点有更大的运用潜能;它消除了 I/O 操作触发中断的必要,有效提高了 CPU 周期的利用率和减小了通讯延迟;所以这些实现,使得 VIA 成为轻量级用户层通讯的规范 and 标准。

参考文献

- 1 Virtual Interface Architecture Specification. Version 1.0. Compaq, Intel and Microsoft Corporations, Dec. 1997. <http://www.viarch.org>
- 2 Martin R P, et al. Effects of Communication Latency, Overhead and Bandwidth in a Cluster Architecture. Computer Architecture News, May 1999. 85~97
- 3 Gusella R. A Measurement Study of Diskless Workstation Traffic on an Ethernet. IEEE Trans. Communications, 1990, 38(9): 1557~1568
- 4 Virtual Interface Architecture for Clustered Systems Dell Computer Corporation white paper, September 1998. <http://www.dell.com/us/en/hied/topics/vectors-1998-wpvi-a.htm>
- 5 Virtual Interface (VI) Architecture-The New Open Standard for Distributed Messaging Within a Cluster. Compaq white papers, September 1998. <http://www5.compaq.com/support/techpubs/whitepapers/ecg0980998.html>
- 6 Berkeley VIA Project. <http://www.lcs.berkeley.edu/~philip/via/>
- 7 Buonadonna P, Geweke A. An Implementation and Analysis of the Virtual Interface Architecture. University of California at Berkeley, Computer Science Department, Berkeley, California, May 1998
- 8 戈弋. 网络并行互连和通讯机制及其研究. [清华大学博士学位论文]. 2000

(上接第63页)

单个 BP 网络相比能够达到更好的学习效果。

结论 本文提出一种子网组结构用来代替多输出的 BP 网络,并提出了一种子网组结构的学习算法。理论分析表明,子网组使用这种学习算法与其相应的 BP 网络原型相比能够达到更好的学习效果。本文提出的思想和方法实际上表明了,在很多情况下,多输出的 BP 网络可以用一组单输出的 BP 网络代替,而且能够

达到更好的学习效果。

参考文献

- 1 刘署光,等. 前馈神经网络中的反向传播算法及改进:进展与展望. 计算机科学, 1996, 23(1): 70~79
- 2 Hornik K M, Stinchcombe M, White H. Multilayer feed-forward networks are universal approximators. Neural Networks, 1989, 2(2): 359~366