

# 基于神经网络的数据压缩研究

马 恋, 何 铭

(长沙理工大学 计算机与通信工程学院, 湖南 长沙 410076)

**摘 要** 数据压缩 = 建立模型 + 编码, 是传统数据压缩的最基本思想。传统的设计思路主要基于减少信息中的冗余着力于编码设计, 其压缩比一直在百分之几十左右徘徊。经过传统编码压缩的信息, 很难再经类似的方法进一步压缩。为了解决上述的问题, 可用占用信号空间较少的信息表示占用信号空间较多的信息, 以实现数据压缩。这就是数据压缩新的设计思路, 这样, 就突破了传统的只依靠编码减少数据冗余的局限, 实现了更高的压缩比。同时, 经过这样压缩后的信息可再重复压缩, 并有很好的性能。这就是文中所介绍的神经网络与数据压缩结合的基本思想。

**关键词** 数据压缩; 神经网络; BP 算法

中图分类号: TP311

文献标识码: A

文章编号: 1673-629X(2007)02-0012-04

## Research on NN - Based Data Compression

MA Lian, HE Pei

(Computer and Comm. Eng. Dept., Changsha Univ. of Sci. and Techn., Changsha 410076, China)

**Abstract** Data compression = model formation + coding, which is the basic thought of the traditional data compression. The traditional ideas of designing data compression which is mainly based upon reducing the redundancy of information exert to coding and designing. The compression rate is very low and the compressed information is difficult to be recompressed. The problems can not be resolved by traditional method. In order to solve those problems, need find the new way what the new data compression designing ideas. That breaks through the traditional localization, achieving higher compression rate. At the same time, this method also comes true recompression and has good capability. This is the basic ideas that NN combines with data compression.

**Key words** data compression; NN; BP algorithm

## 0 引言

数据压缩是信息处理的基础问题之一。所谓数据压缩, 就是以最少的数码表示信源所发的信号, 以减少容纳给定消息集合或数据采样集合的信号空间<sup>[1]</sup>。数据压缩的很多技术已经很成熟, 加上硬件技术的深入发展, 以至于有人认为已经没有继续研究的必要。但现代信息爆炸, 全球网络的兴起, 多媒体技术的全面发展, 各种高新技术不断涌现, 使人们对信息的数量和质量的要求越来越高。由于人类的愿望没有止境, 基于图形图像、声音的信息越来越多, 加上很多领域对实时性的要求, 人们对处理这些信息所占的时间和空间日益不能忍受。这就要求人们为更好、更快地处理多媒体信息开发出更好、更实用的数据压缩方法来。再有, 现代信息多是超文本的, 子信息种类极其繁多, 很可能需要在一个压缩任务中针对各个子信息模块采取

多种相应适合的压缩技术以达到最佳压缩效果。为实现上述目的就需要一个能充分调动各种手段以节省信号空间的智能数据压缩系统。如果能有好的数据压缩方法, 把大量的数据信息缩成尽可能小的部分, 则一来可以减少存储空间, 二来可以降低数据流量。但是, 由于数据信息种类各异, 对存储和处理的要求(精度、速度等)也不相同, 再加上这些大不相同的信息经常是混杂在一起要求一并处理, 而现在又没有一种通用的适合于所有数据种类的好的压缩算法, 所以有理由认为, 可以用人工智能的方法使适合不同压缩要求的压缩方法自动被选择应用到不同类型的数据集合中去。文中所要研究的正是神经网络与数据压缩的结合。

## 1 传统的数据压缩思路

一条消息中的冗余信息占用额外的位来编码, 并且如果要去除额外信息, 将减少消息的大小。为了更好地压缩数据, 需要预言符号高概率的模型。具有高概率的符号就有低的信息内容, 并将只需要较少的位来编码, 这就是传统数据压缩的基本思想, 即: 数据压

缩 = 建立模型 + 编码。传统的设计思路主要基于减少信息中的冗余,着力于编码设计,已经取得了相当多的成就,在此不再详述。但是,以传统的设计思路,压缩比一直在百分之几十左右徘徊,很难取得进一步的进展。还有,经过传统编码压缩的信息,很难再经类似的方法进一步压缩。这两个问题是传统的设计思路所无法解决的。

2 新的数据压缩思路

为了解决上述的问题,可换个角度来考虑问题。既然通过合适的编码可减少信息表示的冗余以达到数据压缩的目的,那么,是否也可找到合适的方法用另外一些占用信号空间较少的信息表示占用信号空间较多的信息以实现数据压缩呢?答案是可以。关键是要建立起占用信号空间较少的信息到占用信号空间较多的信息之间的映射关系。这就是数据压缩新的设计思路。这样,就突破了传统的只依靠编码减少数据冗余的局限,实现了更高的压缩比。同时,经过这样压缩后的信息可再重复压缩,并有很好的性能。从理论上讲,这样的压缩方法可以重复压缩直到得到满意的压缩比为止。

数据压缩的新思路突破了传统的障碍,使人们能够更进一步地压缩信息。但它的意义并不仅此。现在有很多很优秀的压缩算法,它们各有优缺点,很难找出各个方面都是最好的、最通用的算法,而且使用算法的环境、目的都是时刻变化的,对压缩的要求也是不同的。各个算法都有各自的使用场合,要有针对性地研究和改进某一特定领域的信息处理方式和方法,基于已有算法的压缩成果,按新思路用更少的信息表示已经把冗余减到最低限度的信息,以最大限度地压缩信号空间,这样既可利用各个领域的已有压缩成果,又可节省比原先多得多的信号空间。这就是新思路的真正意义所在。

3 神经网络与数据压缩的结合

通过对数据压缩技术的思想与神经网络的某些算法作了认真的分析和比较,发现有些神经网络算法和思想可满足上述新思路的要求<sup>[2]</sup>。神经网络中有一个很知名的算法:BP 算法<sup>[3,4]</sup>。BP 算法是一个很有效的算法,许多问题都可由它来解决。1985 年 Rumelhart 发展了 BP 算法。它的基本思想是:

- (1) 状态向前传播,输入模式由输入单元传到隐层单元。
- ① 隐层单元逐层处理后传到输出单元;
- ② 输出单元产生输出模式。

(2) 误差反向传播,由网络实际输出模式与期望输出模式产生误差信号。

- ① 误差信号沿连接通路反向传播;
- ② 逐层修改各层神经元间的连接权值。

(3) 反复训练网络,对于给定的训练模式,重复上述过程,直到误差满足要求为止。

通过一个多层前馈的网络,在教师信号的指导下对输入的信息进行学习。网络不仅有输入层节点、输出层节点,还有隐层节点(可以是一层或多层)。对于输入信号,要先向前传播到隐层节点,经过传递函数后,再把隐层节点的输出信息传播到输出节点,最后给出输出结果。隐层节点的传递函数通常选取 Sigmoid 函数,如  $f(x) = 1/(1 + e^{-x})$  等。BP 模型已成为神经网络的重要模型之一。其网络拓扑结构见图 1。

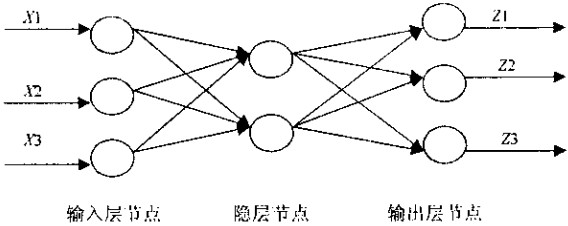


图 1 BP 网络拓扑结构

BP 网络的工作原理如下:

- 1) 定义记号:
- $k$  为样本序号;
- $Y_k$  为输出样本;
- $\hat{Y}_k$  为网络的实际输出;
- $O_{ik}$  为单元  $i$  的输出;
- $W_{ij}$  为由单元  $i$  到单元  $j$  的连接权值;
- $net_{jk}$  为单元  $j$  的总输入;
- $\xi$  为误差集;
- $\eta > 0$  为学习因子。

2) 权值修正公式:

$$\Delta_k W_{ij} = \eta \xi_{jk} O_{ik}$$

3) 输出层单元的误差调整公式:

$$\xi_{jk} = (Y_k - \hat{Y}_k) f'(net_{jk})$$

其中,  $f'$  为传递函数。

4) 隐层单元的误差调整公式:

$$\xi_{jk} = f'(net_{jk}) \sum \xi_{mk} \cdot W_{mj}$$

其中,  $m$  为反馈上层的神经元的编号。

既然神经网络可以通过学习得到输入与指定输出的映射,如果隐层单元个数相对于输入层个数少很多时,建立一个输入与输出完全一样的映射,则相对较少的隐层单元个数就可以代表较多的输出层单元<sup>[5]</sup>。这就是神经网络与数据压缩结合的基本思想,即选择有

尽可能少的隐层单元数的 BP 网络结构,将可能出现的二进制数集(如字符的 ASCII 码字)作为输入,也可直接对欲压缩文件进行学习,按扫描顺序选择一定长度字符串的 ASCII 码作为输入模式,指定其输出信号为其自身并进行学习。学习结束后,就可以进行压缩了。将欲压缩文件从头到尾扫描一遍,按照学习方式,以字符或字符串为单位依次输入,按照学习后的网络权值和输出函数得到隐层单元的中间结果。顺次保存该中间结果,就得到了原文件的一个映射文件。由于隐层单元远少于输入层的单元数,故该映射文件的大小就远小于原文件,达到了压缩效果。解压时,由于 BP 网络稳定性较好,将映射文件的数值作为 BP 网络隐层单元的中间结果,在原有 BP 网络的作用下导出输出层,得出与原文件相同的输出结果,达到了还原的效果。基于这思想,通过修改 BP 算法流程,以实现数据压缩的目的。

4 算法流程

(1) 学习流程。流程图见图 2。形式化学习流程,有如下过程:

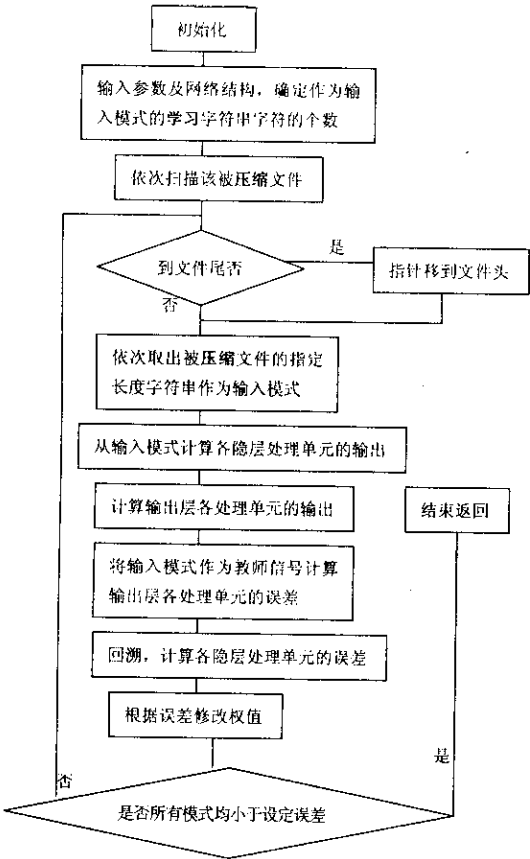


图 2 学习流程图

① 定义网络结构的模块

Netstruct Def( netlayernum ,Anetlayer ,modelnum ,

inputset ,outputset ,... );  
其中 netlayernum 为网络层数目 ;Anetlayer 为一数组 ,指明各层所含单元个数 ;modelnum 为所学习模式个数 ;inputset ,outputset 为二维数组 ,指明每一输入集及对应的输出集。

```
②Q[i]=GetSTR( sourcefile );  
WHILE NOT eo( sourcefile )  
BEGIN  
    computetoooutlayer ;  
    If successif( Q[i] ) then  
        BEGIN Q[i]=GetSTR( sourcefile );  
            continue ;  
        END ;  
    Else  
        BEGIN returntoinlayer ;  
            changeweight ;  
        END ;  
END ;
```

将被压缩文件内容依次取出一个模式长度的字符串( GetSTR )作为 BP 网的输入( Q[i] );再计算网络输出( computetoooutlayer )及输入与教师信号的误差,若小于给定误差限( successif = TRUE ),则本模式学习结束,可以进入下一个模式串的学习。否则( successif = FALSE ),误差反馈( returntoinlayer ),修改权值( changeweight ),重复上述学习过程。重复上述步骤,直至被压缩文件( sourcefile )所有模式均被学习完为止。其中 Q[i] ( 1<i<=netlayernum )为 BP 网络第 i 层各单元的当前值。

(2) 压缩流程。流程图见图 3。形式化压缩流程,有如下过程:

①SaveNetParameter( targetfile );  
把学习好的网络参数保存在目标文件( targetfile )的头部。

```
②WHILE NOT eo( sourcefile )  
BEGIN  
    Thide=Computetohidlayer( GetSTR( sourcefile ) );  
    Writetoooutputfil( Thide ,targetfile );  
END ;
```

用学习好的网络一段一段地压缩( Computetohidlayer )被压文件( sourcefile )。压缩后在目标文件( targetfile )存放的是隐层的参数。

(3) 解压流程。流程图见图 4。形式化解压流程,有如下过程:

①ResetBpNe( targetfile );  
从压缩后的目标文件( targetfile )中读出网络参数,重建 BP 网络。

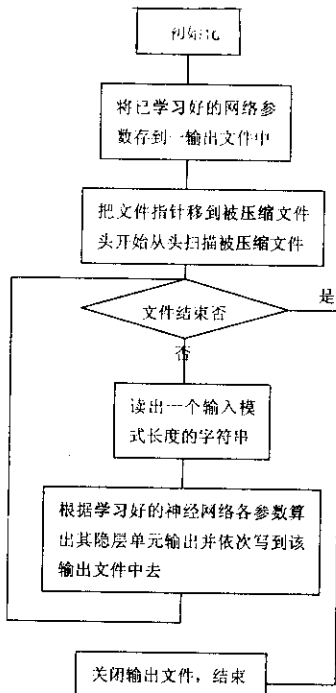


图 3 压缩流程图

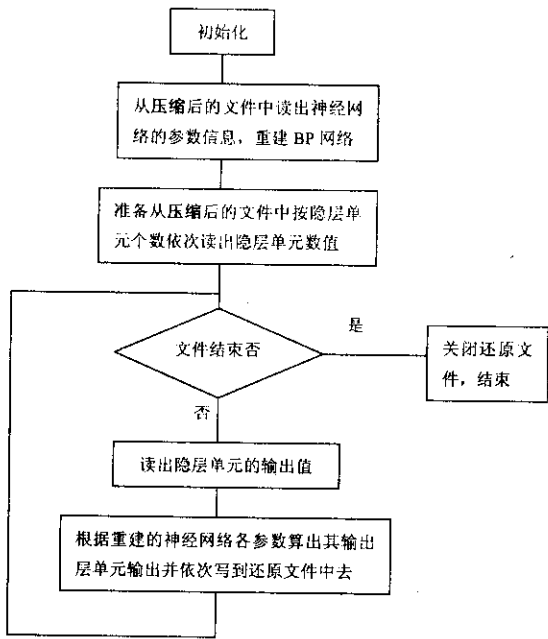


图 4 解压流程图

②WHILE NOT eo(targetfile)

```
BEGIN
Thide= readhideunitva( );
Tout= computetoooutput( Thide );
Writetoretfil( Tout ,retfile );
```

END ;

从被压缩文件( targetfile )中还原( readhideunitval )出隐层单元值 Thide ,再根据重建的 BP 网络得到与输入模式一致的输出 Tout( computetoooutput ) ,再将其写入到解压文件( retfile )中去( Writetoretfile )。此处 , Thide ,Tout 均为值集。

5 关于再压缩

从理论上说 ,按新思路的方法可以重复多次压缩信息 ,以达到理想的压缩比 ,大大减少信息传播所需的时间和空间。但因多次压缩需经多次解压才能还原 ,且每次压缩都可能产生信息缺损 ,故应注意压缩次数不要超过一定限度 ,以避免压缩时间、解压时间和信息缺损增加到不可忍受的地步。

6 结束语

用神经网络的方法实现数据压缩的新思路 ,这本身就是一个突破。它必将带动一个新领域研究的发展。这种方法也促使人们更多地考虑到对已有压缩成果的再压缩 ,对压缩极限有了新的认识 ,开创了压缩理论的新天地。数据压缩技术是现代信息社会所急需的热门技术之一 ,仍然还有很多的方面值得人们去研究。现在 ,不仅是上述的方法 ,很多新理论、新技术也层出不穷。小波变换、分形压缩等等已经成为数据压缩的新热点。上面提到的与神经网络结合的数据压缩技术也将占有一席之地 ,并将以此为基础 ,不断促进信息处理的智能化、自动化。

参考文献 :

[ 1 ] 吴乐南. 数据压缩的原理与应用[ M ]. 北京 :电子工业出版社 ,1996.

[ 2 ] 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.

[ 3 ] 王彦春 ,余钦范 ,段云卿. 改进的快速 BP 算法[ J ]. 物探与化探 ,1999 ,23( 2 ) :133 - 137.

[ 4 ] Werbos P J. Back Propagation Through Time :What It Does and How to Do It[ J ]. Proc. of the IEEE ,1990 ,78( 10 ) :1550 - 1560.

[ 5 ] 侯 阳. 数据压缩技术与 C 语言实例[ M ]. 北京 :学苑出版社 ,1994.

(上接第 11 页)

[ 2 ] 吴望名. 弗晰图与弗晰树[ J ]. 数学的实践与认识 ,1980 ( 4 ) :13 - 16.

[ 3 ] 王耀南 ,李树涛. 计算机图像处理与识别技术[ M ]. 北京 :高等教育出版社 ,2001.

[ 4 ] 杨炳儒. 图论概要[ M ]. 天津 :天津科学出版社 ,1990.

[ 5 ] 严蔚敏 ,吴伟民. 数据结构[ M ]. 北京 :清华大学出版社 ,1992.