

新型前馈网络学习算法在语音识别中的应用

李 原¹ 邓辉文²

(西南大学计算机与信息科学学院 重庆 400715)¹ (西南大学逻辑与智能研究中心 重庆 400715)²

摘 要 简要介绍一种全新的前馈神经网络学习算法——代数算法,以该理论提出者针对这一算法做出的多方面的理论证明为基础,将此算法与传统 BP 算法相比较,利用其在网络结构和获取全局最优点方面具有传统 BP 算法不可比拟的优点,在 matlab7.0 实验平台上,分别把传统 BP 算法和此种新型学习算法应用于语音识别,通过仿真实验说明了新型算法的有效性,语音识别的精度在一定程度上得到了改善。

关键词 人工神经网络,代数算法,代价函数,精确映射,严格对角占优矩阵

Application of Speech Recognition System Based on Algebra Algorithm

LI Yuan¹ DENG Hui-wen²

(Faculty of Computer & Information Science, Southwest University, Chongqing 400715, China)¹

(Research Center of Logic and Intelligence, Southwest University, Chongqing 400715, China)²

Abstract It introduces a novel learning algorithm. Compared with backpropagation algorithm, the novel learning algorithm for neural networks in search of global optimal point is much better. On matlab7.0 experimental platform, backpropagation algorithm and this novel learning algorithm are applied into speech recognition. From validating result, the speech recognition's accuracy is improved with this novel algorithm. It is an efficient method to be used into speech recognition.

Keywords Artificial neural networks, Algebra algorithm, Cost function, Accurate mapping, Strictly diagonally dominant matrix

1 引言

随着现代信息科学的迅猛发展,智能化工具进入人们的生活,致使人们渴望人与机器的交互,甚至期盼人与机器的对话能达到人与人之间交流的那种自然的水平。在种种思潮的触动下,对人机交互的研究近几年来趋于活跃^[1]。实现对语音的识别是人机交互研究的核心之一,也是它的一个重要研究课题。

语音识别的研究,从最初的孤立数字识别到今天的大词连续语音识别^[1],已经开展了 40 多年。在其间各种各样的理论和算法应运而生,并相继被运用于语音识别当中,人工神经网络(Artificial Neural Networks, ANN)便是其中之一。ANN 以其特有的性能,活跃地应用于语音信号处理^[3]。在多种类型的 ANN 中,前馈网络的反向传播(Back Propagation, BP)算法得到了广泛应用并被深入研究,但因其收敛性、全局最优点获取及规模等方面存在着本质上无法克服的缺点^[4,5],在实际应用中受到了限制。考虑到代数算法这种新型前馈神经网络学习算法具有很高的精度、较快的运算速度,能在网络训练之前根据给定的问题确定网络结构等一些 BP 算法不可比拟的优点,我们将其运用到语音识别中。我们把用代数算法训练的神经网络简称为代数神经网络。

本文探讨了用代数神经网络实现对非特定人语音的识别。首先介绍了代数算法这种新型前馈神经网络学习算法的基本原理,并将其与传统 BP 算法相比较;其次是对引入了该新型算法的语音识别系统的介绍及具体的实验设计;然后介绍了实验的统计结果并做出分析;文章最后对这种新型算法在语音识别中的性能进行了总结。

2 代数算法的基本原理

代数算法^[2],这一全新的神经网络学习算法的得名,是因为其将复杂的非线性优化问题转化为一组线性代数方程组来求解。

2.1 神经网络的拓扑结构

本文探讨的代数神经网络,采用了目前广泛使用的三层前馈网络模型,它与传统神经网络不同之处在于,网络训练开始之前就能根据具体问题准确给出隐层神经元个数,这也是代数神经网络具有的独特优点之一。下面介绍网络结构如何确定。

本三层的代数神经网络的拓扑结构与传统三层网络相同,由输入层、隐层、输出层构成。用 X_0 代表输入样本矩阵。分别用 Y_2 和 Z_2 代表实际输出样本矩阵和目标样本矩阵。神经网络的输入输出关系为

$$\begin{cases} X_1 = \bar{W}_0^T X_0 \\ Y_1 = f(X_1) \\ Y_2 = W_1^T Y_1 \end{cases}$$

其中 Y_1 是隐层实际输出矩阵。

我们首先按一定方式选择自由权 $\bar{W}_0$ ^[2],通过神经网络的输入输出关系式求得 Y_1 ,然后令 $Y_2 = Z_2$ 。由此可得一组变量为 W_1 的线性方程组: $Z_2 = W_1^T Y_1$,若 Y_1 列满秩,则此方程组有解。用 l 代表隐层神经元的个数, N 代表待训练样本对个数,要保证 Y_1 列满秩,则有 $\text{rank}(Y_1) = N$,根据线性代数知,该式成立的一个必要条件是 Y_1 的行数不小于列数,即 $l \geq N$ 。为简化神经网络结构,保证隐层神经元个数尽可能小,我们取 $l = N$ 。

所以代数神经网络的结构参数在网络训练之前可以确定为 $m-l-n$, 输入和输出神经元个数设置与传统神经网络一样, 输入样本的维数 m 作为输入层神经元个数, 实际输出样本的维数 n 作为输出层神经元个数, l 为待训练样本对的个数, 作为隐层神经元的个数。

2.2 代数算法

代数算法的基本思想是按一定方式选择自由权, 而后求解隐层和输出层间的权值, 最终实现网络的精确映射, 即训练后的网络代价函数 $J(\mathbf{E}) := \|\mathbf{E}\|_F^2 = \|\mathbf{Y}_2 - \mathbf{Z}_2\|_F^2$ 为 0。将隐层神经元数目确定为待训练样本对的个数, 这是网络实现精确映射的一个充分条件^[7], 隐层实际输出矩阵即是 N 阶方阵, 故只要 $\mathbf{Y}_1$ 非奇异, 等式 $\mathbf{Z}_2^T = \mathbf{Y}_1^T \mathbf{W}_1$ 必然存在唯一解。至此, 神经网络训练的问题就简化成求解线性方程组及如何保证 $\mathbf{Y}_1$ 是非奇异矩阵的问题。

由于将类支集函数神经元引入隐层, 可以方便地调整自由权从而使含待求权的线性方程组是良态的, 便于求解。在代数神经网络中, 采用类支集函数作为网络的激励函数。

在选取类支集函数作为隐层激励函数的前提下, 选取自由权 $\tilde{\mathbf{W}}_0$ 之后, 就能保证隐层输出矩阵 $\mathbf{Y}_1$ 是严格对角占优矩阵^[2], 若 $\mathbf{Y}_1$ 严格对角占优, 则 $\mathbf{Y}_1$ 也是非奇异矩阵^[9]。即是说, 如果 $\mathbf{Y}_1$ 严格对角占优, 那么 $\mathbf{Z}_2^T = \mathbf{Y}_1^T \mathbf{W}_1$ 必有唯一解。同时考虑到严格对角占优的线性方程组的诸多优越性, 它可以用多种方法求解, 当用迭代法求解时, 它对任何初始值都是收敛的。因此, 我们将隐层输出矩阵考虑为严格对角占优矩阵。当隐层激励函数 $f(x)$ 是 $x = c$ 的类支集函数时, 要使 $\mathbf{Y}_1$ 严格对角占优, $\tilde{\mathbf{W}}_0$ 应满足如下条件^[2]:

$$\begin{cases} \tilde{x}_{0j} \cdot \tilde{w}_{0j} - c = 0, & j=1, 2, \dots, l(l=N) \\ |\tilde{x}_{0p} \cdot \tilde{w}_{0j} - c| \geq \alpha, & p \neq j, p=1, 2, \dots, l(l=N) \end{cases}$$

只要根据 $\tilde{\mathbf{W}}_0$ 应满足的条件, 并且在保证输出的 $\mathbf{Y}_1$ 严格对角占优的前提下, 确定了输入和隐层间的权矩阵, 等式 $\mathbf{Z}_2^T = \mathbf{Y}_1^T \mathbf{W}_1$ 的 $\mathbf{W}_1$ 必然有唯一解, 就可以得到严格对角占优的含待求权 $\mathbf{W}_1$ 的线性方程组。三层网络结构及求得的权值 $\tilde{\mathbf{W}}_0, \mathbf{W}_1$ 构成了所求网络的全部参数, 实现了精确映射。

2.3 代数算法与传统 BP 算法的性能比较

将代数算法运用于神经网络与传统 BP 算法相比, 具有以下优点:

神经网络在训练之前就能根据具体问题准确的确定隐层神经元的个数, 进而确定网络的结构。

采用类支集函数作为隐层的激励函数, 不要求函数可导, 使用更方便, 求解更简单, 增强了神经网络训练的稳定性。

由于隐层输出考虑了严格对角占优矩阵, 网络的抗噪能力得到提高。

可以很方便地求得全局最优点, 使训练过的样本, 实现精确映射, 即代价函数等于 0。

神经网络训练的主要工作是求解一个多项式阶的线性方程组, 时间效率更高, 有更快的运算速度。

3 基于代数算法的语音识别系统

本语音识别系统(图 1)是为识别非特定人的语音而设计。在本系统中引入了代数神经网络, 通过此网络来实现对语音特征参数的学习和计算, 最终给出识别结果。此系统主要由三个部分构成, 语音特征参数的提取, 语音样本的学习和对语音的识别。

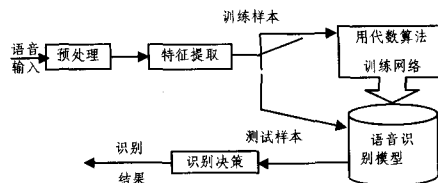


图 1 基于代数算法的语音识别系统

3.1 语音信号特征参数提取

我们通过采样、预处理和特征提取这三个阶段, 实现模拟的语音信号到时域离散的数字信号的转换以及提取出信号中的特征参数。

预处理阶段, 首先将语音信号通过一个一阶的高通滤波器, 用于抑制语音信号的低频分量, 以提高语音识别系统的性能; 然后对语音信号进行分帧处理, 保证语音参量有更好的连续性; 再采用端点检测中常用的双门限法^[8]来确定语音的起始位置, 并将提取出的有用的语音部分用汉明窗进行加窗处理。

特征参数的提取阶段, MFCC 参量比 LPCC 参量能够更好地提高系统的识别性能^[1], 因而本系统采用了美尔倒系数 MFCC。首先对预处理阶段获得的语音信号进行短时傅立叶变换, 得到其频谱; 然后求频谱幅度的平方, 并将获得的能量谱通过有 12 个滤波器的滤波器组; 再对滤波器组的输出取对数, 最后做 24 点逆傅立叶变换, 就得到了 MFCC 参数。标准的 MFCC 参数只能反应语音参量的静态特性, 所以, 再对每阶的 MFCC 参数计算出其一阶差分参数, 通过差分参数来描述语音参量的动态特性。然后把 MFCC 参数和各阶的差分参数合并作为语音特征参数。最后对语音特征参数进行时间规整, 最终将经过规整的语音特征参数作为语音模型。

3.2 语音样本的学习

本语音识别系统对样本学习的过程也即是对神经网络权值和阈值各个参数进行调整的过程。在本系统中, 使用代数算法对神经网络进行训练。根据前面对代数算法理论的介绍, 首先确定神经网络隐层神经元的数目, 即为语音训练样本对的个数; 然后设定隐层的激励函数, 我们选取的类支集函数是 Delta 函数:

$$\text{Delta}(x; \epsilon, c) = \frac{1}{\pi} \frac{\epsilon}{(x-c)^2 + \epsilon^2}$$

由于选取类支集函数能实现代价函数为 0 的精确映射, 因此我们将训练样本对的最理想输出矩阵设定为目标矩阵。最后将训练样本对经过特征提取后得到的语音模型作为输入, 经过精确映射, 得到网络的权值和阈值, 最终实现对样本的学习。

4 仿真实验与分析

本实验将代数算法引入到语音识别当中, 目的是为了证明语音识别中通过代数算法训练的神经网络, 能更加精确地记忆语音特征, 而提高语音的识别率。

实验搜集了 150 个训练样本和 150 个测试样本, 在 Matlab7.0 的实验平台上对样本提取特征参数, 以及训练网络, 通过对最终的识别结果的分析, 证明了基于代数算法的神经网络应用于语音识别中的有效性。

4.1 实验设计

本实验所用的语音样本以 0-9 这十个数字作为实验对象, 取自 5 个人, 3 男 2 女。训练和测试数据均在普通实验室中录制, 采样率设置为 8kHz, 量化精度 16bit, 单声道。每人

将各个数字重复读 6 遍,前 3 次作为训练样本,后 3 次作为测试样本。这样便构成了各有 150 个样本的训练样本集和测试样本集。再将获取的语音信号样本,提取出 12 阶美尔倒谱系数和 12 阶的一阶差分参数,并合并成 24 阶的特征参数,最后用 K 均值算法将帧数规整到 4 帧,作为神经网络的输入,输入神经网络的特征矢量维数为 96。

4.2 实验结果与分析

实验中将基于代数算法的神经网络和基于 BP 算法的神经网络分别进行非特定人孤立词语音识别实验。从识别的精确度上对两者进行比较。

(1) 基于代数算法的神经网络:输入层 96 个神经元,输出层 10 个神经元。隐层 150 个神经元。学习误差为 0。隐层神经元的激励函数为 Delta 函数。

(2) 基于 BP 算法的神经网络:输入层 96 个神经元,输出层 10 个神经元。隐层 60 个神经元。学习误差 0.01,学习速率为 0.1。训练的最大步长 5000。隐层神经元的激励函数选为 Sigmoid 函数。

实验结果如表 1、表 2 所示。

表 1 采用 BP 算法的孤立词语音识别结果

识别的数字	测试样本数	正确识别数	识别率(%)
0	15	11	73
1	15	10	66
2	15	13	86
3	15	14	93
4	15	13	86
5	15	14	93
6	15	10	66
7	15	11	73
8	15	13	86
9	15	14	93

表 2 采用代数算法的孤立词语音识别结果

识别的数字	测试样本数	正确识别数	识别率(%)
0	15	13	86
1	15	12	80
2	15	14	93
3	15	13	86
4	15	15	100
5	15	15	100
6	15	14	93
7	15	14	93
8	15	15	100
9	15	14	93

从表 1 的结果可以看出,BP 算法对各个数字识别率的差异较大,最高可达 0.93,最低只有 0.73。正因为对“0”和“6”,“1”和“7”,这些易混淆数字的识别能力较弱,造成系统的性能不稳定,总体识别率偏低,基于 BP 算法的语音系统的平均识别率为 0.815。

表 2 的结果表明,代数算法对各个数字的识别率差异较小,最高的为 1,最低的为 0.8。由于基于代数算法的网络对样本的学习是无误差的精确映射,能够在网络中更加精确的记忆每个语音的特征,因而在对网络的训练上优于 BP 网络,对易混淆词的识别率有明显改善,基于代数算法的语音识别系统的平均识别率为 0.924。

综合以上实验结果可以看出,对于 0-9 这 10 个数字,代数算法应用于语音识别的识别效果优于 BP 算法。但众所周知,神经网络隐层节点数对网络的性能有极大的影响,在上面的实验中,我们将基于代数算法和 BP 算法的语音识别系统模型的隐层神经元个数分别设置成 150 和 60,从数值上看,两者的隐层结点数相差较大。因此接下来我们用以下实验,通过隐层神经元个数的不同设置来对两种算法在语音识别中的性能做进一步分析。

两种神经网络的输入输出,以及相关参数都同前一个实验一致,隐层神经元的个数分别取 60,80,100,120,150,200。

实验结果如表 3 所示。

表 3 两种算法识别率对比图

隐层神经元个数	BP 算法识别率(%)	代数算法识别率(%)
60	81.5	81.6
80	83.6	85.0
100	84.9	89.1
120	85.0	90.5
150	85.7	92.4
200	85.7	93.0

从表 3 我们可以看出,隐层神经元个数对代数算法的识别率的影响是巨大的,但是当隐层数大于理论值之后,其识别率的增长微弱,然而 BP 算法的神经元个数在按照先验知识设定之后,再逐渐增加,这对识别率的影响较小,识别率的增长明显小于代数算法。从实验结果中两者的识别率比对照来看,当代数算法中隐层数降到 60 的时候,其识别率与 BP 算法相当。从其隐层数的减少和识别率降低的比率来看,不难推知,如果代数算法中的隐层神经元个数降到 60 以下,识别率很可能会大幅度降低,而低于 BP 算法的识别率。从总体上看,如果两种算法的隐层神经元个数均取其最佳理论值,在语音识别的应用中代数算法是优于 BP 算法的。

结束语 本文将一种新型的前馈网络学习算法——代数算法引入到语音识别当中,通过代数算法来改进语音识别中神经网络的学习方式,使其在训练阶段实现无误差的精确映射,而达到一种理想的训练效果。

在非特定人孤立字识别的实验中将代数算法与目前应用最广泛的 BP 算法相比较,从统计结果来看,代数算法的确在对神经网络的训练上达到了良好的效果,识别率明显优于 BP 算法。但代数算法要实现神经网络对神经网络的无误差的精确映射,要求隐层神经元个数等于训练样本数目,在样本数目过大的情况下,会造成神经网络结构的复杂化。

由于本实验的目的主要是验证引入了代数算法的语音识别系统的性能,因此将其与传统 BP 算法相比较,在语音特征参数提取上没有做过多的优化,以及训练和测试样本集的规模偏小。在实际应用中,为了有更高的识别率,可以通过优化语音特征参数和增大训练样本集,以及适当增加测试样本集来实现。

参考文献

- [1] 张有为,等. 人机自然交互[M]. 北京:国防工业出版社,2004;1-3,86-89,98-99
- [2] 张代远. 神经网络新理论与方法[M]. 北京:清华大学出版社,2006;28-30,45-49
- [3] 褚蕾蕾,陈绥阳,周梦. 计算机智能的数学基础[M]. 北京:科学出版社,2002;69-72
- [4] Lexington M, Darpa. Neural Network Study. MIT Lincoln Laboratory, 1988
- [5] (美)Hagan M T, et al. 神经网络设计[M]//戴葵,等译. 北京:机械工业出版社,2002;227-232
- [6] 周开利,康耀红. 神经网络模型及其 MATLAB 仿真程序设计[M]. 北京:清华大学出版社,2005
- [7] 张代远,虞厥邦,邱玉辉. 能实现精确映射的前馈神经网络快速算法与结构设计[J]. 计算机科学,1999,26(4):63-65
- [8] He Q, Zhang Y W. On Prefiltering and Endpoint Detection of Speech Signal[J]. ICSP98, 1998, 1: 749-752. Baldi P. Gradient Descent learning algorithm overview; General dynamical systems perspective. Control Theory Appl, 1995, 6(1):182-195
- [9] Baldi P. Gradient Descent learning algorithm overview; General dynamical systems perspective. Control Theory Appl, 1995, 6(1):182-195