假期总结

第一周

神经网络基本概念，学习方法及各种激活函数

1. **前馈神经网络　(　Feedforward Neural Networks )**

这种网络只在训练过程会有反馈信号，而在分类过程中数据只能向前传送，直到到达输出层，层间没有向后的反馈信号，因此被称为前馈网络。感知机( perceptron)与BP神经网络就属于前馈网络。

1. **反馈神经网络　(　Feedback Neural Networks )**

反馈型神经网络是一种从输出到输入具有反馈连接的神经网络，其结构比前馈网络要复杂得多。典型的反馈型神经网络有：Elman网络和Hopfield网络。

1. **自组织网络 ( SOM ,Self-Organizing Neural Networks )**

自组织神经网络是一种无导师学习网络。它通过自动寻找样本中的内在规律和本质属性，自组织、自适应地改变网络参数与结构。

学习方法：

监督学习：BP算法，delta学习http://pic002.cnblogs.com/images/2011/63234/2011030722150043.png

其中Wij表示神经元j到神经元i的连接权，di是神经元i的期望输出，yi是神经元i的实际输出，xj表示神经元j状态，若神经元j处于激活态则xj为1，若处于抑制状态则xj为0或－1（根据激活函数而定）。a是表示学习速度的常数。假设xi为1，若di比yi大，那么Wij将增大，若di比yi小，那么Wij将变小。

无监督学习：Hebb学习http://pic002.cnblogs.com/images/2011/63234/2011030722144483.png

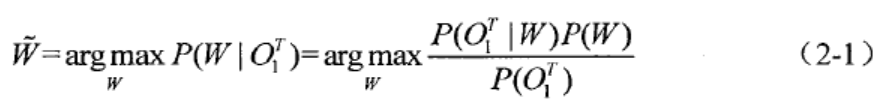
其中wij表示神经元j到神经元i的连接权，yi与yj为两个神经元的输出，a是表示学习速度的常数。若yi与yj同时被激活，即yi与yj同时为正，那么Wij将增大。若yi被激活，而yj处于抑制状态，即yi为正yj为负，那么Wij将变小

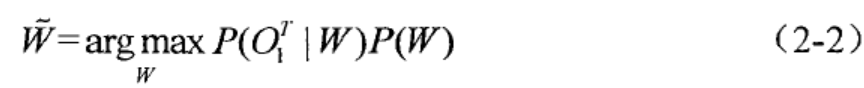
半监督学习

第二周

阅读了**《automatic speech recognition a deep learning approach》**

**神经网络在语音识别上的应用**

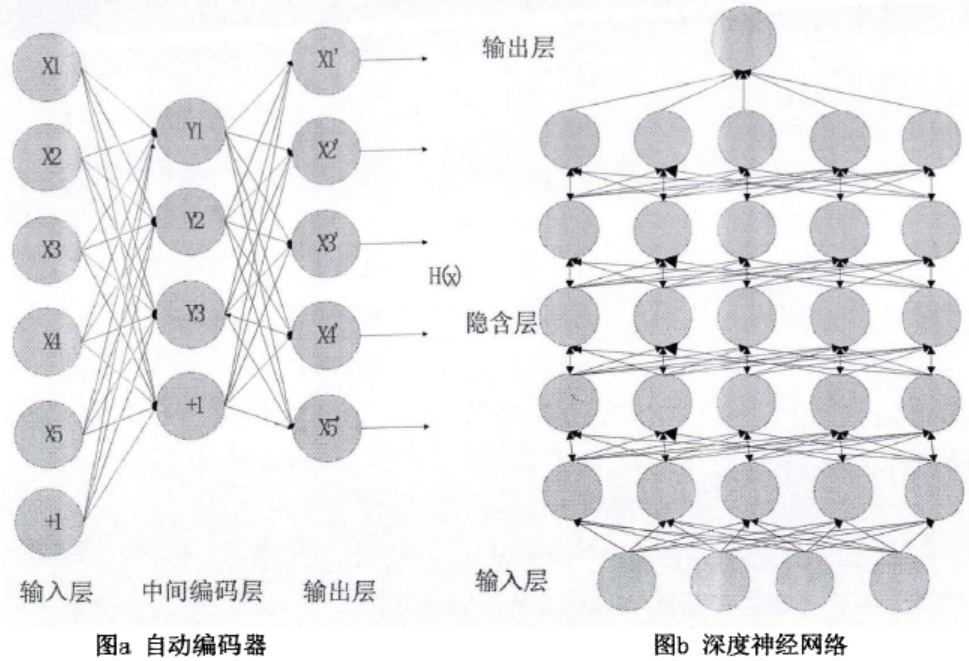
大词汇量连续语音识别系统都是在隐马尔科夫模型（HMM）为框架进行统计训练，在给定语音特征序列后，结合声学模型和语言模型，根据最大后验概率算法输出词序列，其数学表示如下式所示

上式中，P(W)是语言模型，代表特定词序列W出现的先验概率; 为声学模型，表示给定词序列为W时，输出的声学特征时的概率; 是声学特征的概率，与词序列W的无关，所以可以忽略掉。因而，式2-1可变为: 

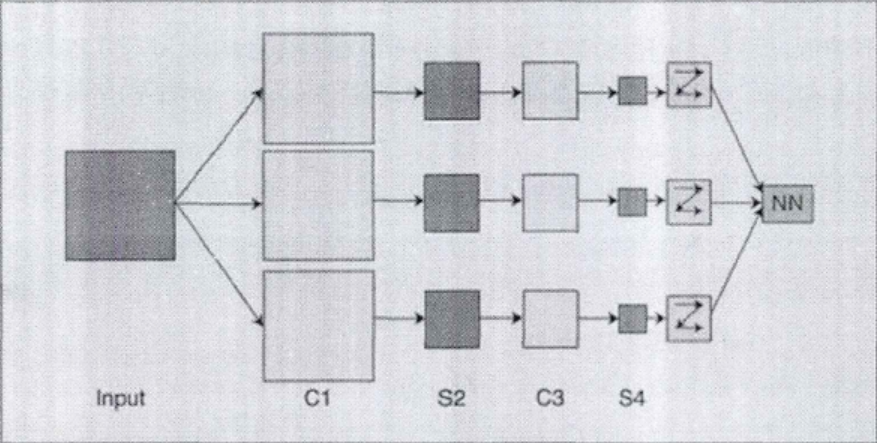
深度学习的模型：

1. 自动编码机：

自动编码器的原理是将原始输入数据x经过多个隐含层的映射计算转换为编码c(x)输出，用同样方法将c(x)经过多个隐含层的逆向映射计算还原为最幵始的网络输入数据，即自动编码器的目标输出数据就是输入数据本身。当隐含层节点的输入和输出间呈线性关系，误差准则函数采用最小均方训练，这种编码模型可以看做为主成分分析,若隐含层映射关系为非线性映射，则是自动编码器



1. 深度卷积网络CNN



1. 递归神经网络RNN

深度学习的模型构成

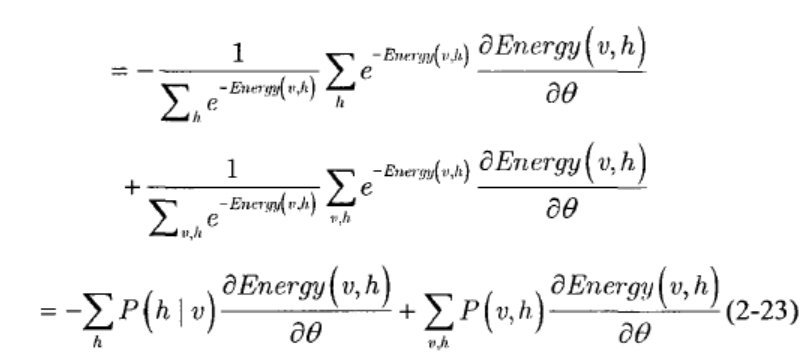
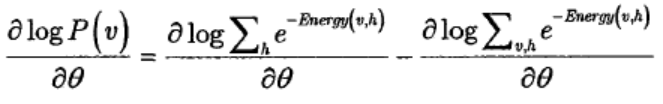
1. RBM模型

学习RBM本质是寻找合适的参数θ来拟合给定的训练数据的分布，通过最大

化可观测数据概率分布的对数似然函数来确定θ，公式如下：



使用随机梯度上升法来确定的最大值，其中的偏导如下：

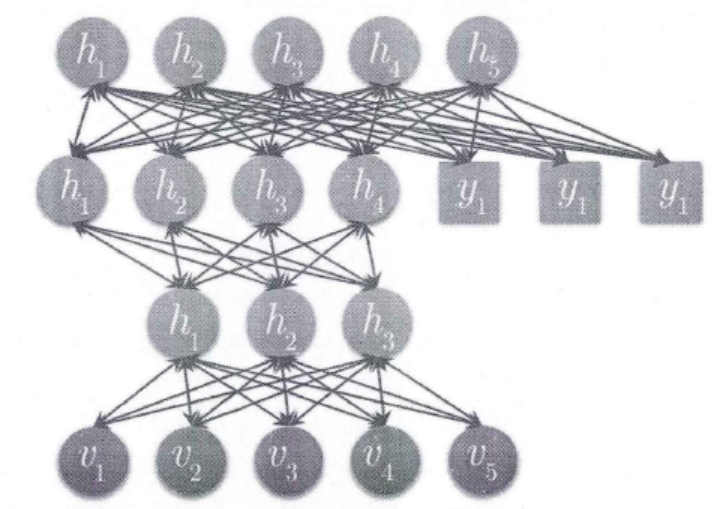


因为RBM是对称的模型结构，以及层中个各单元是相互独立的，所以可以根据Gibbs采样的方法得到服从RBM分布的随机样本。

1. DBN模型

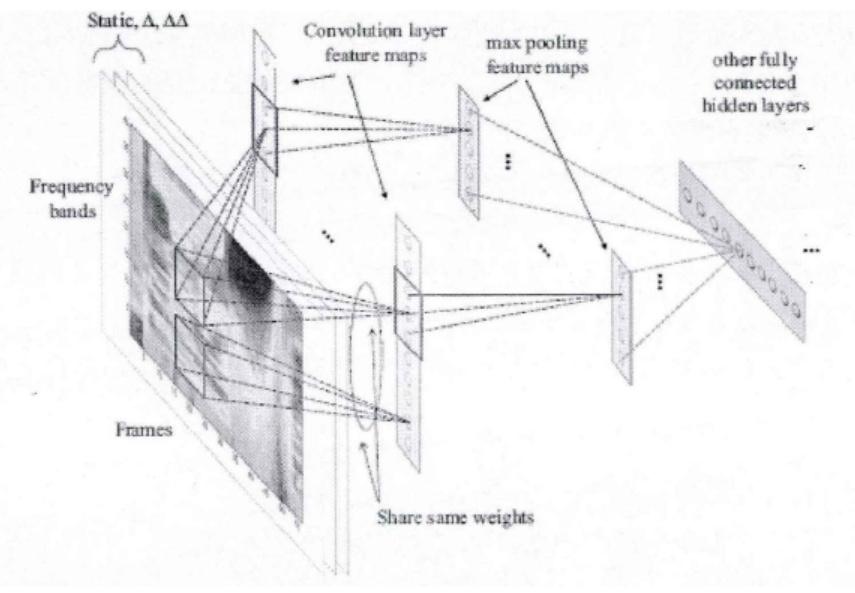
DBN模型是由多层RBM叠加形成，下层各部分是单向传播，顶层是一个全联通的网络，所以既可以被看作生成模型，也可以被当作判别模型，训练方法是，使用无监督的贪婪逐层预训练方法获得初始权重

训练好的DBN模型如下图所示，图中的方块形状的单元就是顶层中的RBM标签数据。



1. CNN模型

CNN本来是被用作图像识别的，我在许多其他的论文中都有CNN在声场识别方向的应用，每一篇侧重点都不相同，本文是用作语音识别，因为CNN 对语音信号具有特殊的优势，



深度神经网络训练中用到的技术

1. 随机梯度下降算法的应用
2. 网络的GPU训练，这篇文中就有用到

图形处理器（GPU）多核体系结构含有几千个流处理器，可并行化执行矢量运算，计算时间将大幅缩短。并行程序在系统中的运行速度相比于单核系统往往能提升几十倍乃至上千倍。

实验采用863中文语音库用作深度神经网络的训练数据。整个语音库由训练集数据和测试集数据两大块组成，总共有166个说话人，男女所占比例为1:1。朗读语料选取从1993-1994年间的《人民日报》共计1560句文本，总共分三类，说话人选取一类文本朗读采样，采样频率定为16khz。本文中训练集数据为原数据中所有的训练数据共计42638条，测试集数据是对原数据中的测试数据随机抽取2000条。语言模型的训练语料为，部分863文本和新闻联播语料还有北方语料库的总和。

特征处理： 本文所采用的特征处理方法用Kaldi中特征提取工具以帧长25ms、帧移10ms提取原始数据生成39维MFCC 特征（12维滤波器输出值加上1维对数能量，以及其一阶差分和二阶差分。）

标签选择： 标签主要是指特征块所对应的状态，这里所指的状态是特征块的中心倾所对

应的音素状态，有单音素和三音素两类，其中单音素有200种标签，三音素有1399种标签，本实验用到的是三音素标签。音素状态获得主要是通过常规的GMM-HMM先训练单音素模型后绑定训练三音素模型，待识别率达到一定标准，用训练好的三音素模型去强制对齐得到三音素标签，单音素标签可以通过三音素标签转换得到。

深度网络的结构： 深度神经网络结构参数，主要是隐含层层数、每个隐含层所包括的单元数，还有隐层单元类型的选择。本文深度神经网络主要由一层输入层、五层输隐含层

和一层输出层构成

在实验中网络结构有好多种，输入数据也不局限一种，输出层维数与标签数量对应，标签数定为。标签是通过传统方法训练的声学模型对标签文件做强制对其得到的，实验中只考

虑三音素标签。输出层用Softmax函数，惩罚函数用交叉熵

对深度神经网络的训练，首先是无监督逐层训练DBN网络，再用训练好的DBN网络去初始化DNN网络，然后通过标签对DNN网络做全局的误差反向传播，最后截取训练好的DNN中的一部分作为生成网络，生成新的语音特征。

实验结果：随着深度神经网络输入数据左右绑定倾数的增加，提取的新特征在逐渐变好，可以看出顿目前是最优值。深度神经网络提取的特征最好的效果是比特征在词错误率和句错误率上分别有1.98%和4.21%。从句识别错误率下降较大可以看出深度神经网络用在语音识别有效的提取了语音帧间的相关信息，在连续语音识别上有一定的优势。

第三周第四周

《基于深层神经网络的音频特征提取及场景识别研究》

该论文的独特之处：

首先把特征分析角度从时域或者频域特征分析转变为基于频域的长时分析；深度神经网络在图像分析领域有着非常好的识别性能（CNN），所以借助其特征分析的能力对语谱图进行深层特征分析提取；然后使用分类器进行最后的分类。

第五周

学习了mlp算法，并找到了一些神经网络的程序运行并理解

阅读了论文Remote Sensing Image Classification Model Based on Artificial Neural Network and Multi-layer Sensor\_MLP\_

