深度学习

背景

介绍AE，SAE，CNN

简单介绍其应用

正文：

（引入）

80~90年代，大量的研究者投身于神经网络的研究，然而那是的计算机速度慢，数据量也少，研究因而受到了很大的约束。

2000~2010左右的很长一段时间里，神经网络消失在机器学习的领域中。

而近几年来，神经网络似乎又回到了机器学习研究者的视线中来。

首先是09年的语音识别，紧接着12年的计算机视觉和14年机器翻译，让人们重新认识了神经网络，不过这次有了新的名字—深度学习。

那么是什么让深度学习再一次回归呢？答案是大量的数据和价格便宜且越来越快的GPU。

现在回到深度学习的概念

* 深度学习：一种基于无监督特征学习和特征层次结构的学习方法
* 可能的的名称：
  + 深度学习
  + 特征学习
  + 无监督特征学习

**传统的模式识别方法：**

**原始数据经过预处理、特征提取、特征选择最后进行预测或识别。**

* + **良好的特征表达，对最终算法的准确性起了非常关键的作用；**
  + **识别系统主要的计算和测试工作耗时主要集中在特征提取部分；**
  + **特征的样式目前一般都是人工设计的，靠人工提取特征。**

（AE）

AE通过将可视层的输入变换到隐藏的输出层，然后通过隐藏层进行重构使得自动编码器的目标输出与原始输入自身几乎相等。，如图 3a所示。AE的目标函数为



损失函数L可以是连续值的传统的方差损失函数或者是二值的交叉熵损失函数

（SAE）

SAE则是在AE基础上进行扩展，具体方法（贪心逐层预训练算法）如下：

1. 先训练第一层AE
2. 将该AE抽取的特征作为下一层AE的输入
3. 重复2操作
4. 将最后一层AE的特征作为有监督层的输入，进行有监督学习
5. 最后可以选择对最高层（监督层）进行微调，也可以对全部层进行微调。

（CNN）

卷积操作

（con.gif）

*f* = σ(*W*(1)*xsmall* + *b*(1))

Pooling操作

（pooling.gif）

Maxpool、Meanpool

深度学习的应用

***语音识别***

在过去几十年中，语音识别领域的研究者们都把精力用在基于 HMM-GMM的系统，而忽略了原始语音数据内部原有的结构特征。深度神经网络DNN在 2010 年开始被引入处理语音识别问题，因为 DNN对数据之间的相关性有较大的容忍度，使得当 GMM被 DNN替换时效果明显有了飞跃。2012年，微软公司一个基于深度学习的语音视频检索系统（Microsoft Audio Video Indexing Service, MAVIS）成功问世，将单词错误率降低了 30%（从27.4%到 18.5%）。2014 年 IBM的沃森研究中心的T.N.Sainath的工作结果显示 DNN比以往过去的GMM-HMM模型有 8% ~ 15% 的提升,而 CNN相比于一般 DNN来说能对数据间强烈的相关性有更强的适应力，同时足够深的网络还有对数据的平移不变性的特性。

***计算机视图***

深度学习在计算机视觉上的成功应用，主要体现在对象识别和人脸识别领域上。过去很长一段时间，机器视觉中的对象识别一直依赖于人工设计的特征，例如尺度不变特征转换（Scale Invariant Feature Transform SIFT）和方向梯度直方图,（Histogram of Oriented Gradients，HOG），然而像 SIFT和 HOG这样的特征只能抓取低等级的边界信息。

针对以往小规模样本所无法表现的真实环境中更复杂的信息，2010年人们引入了更大的数据集，例如 ImageNet 数据集中有着 1500万的标记高分辨率图像和超过 2 万 2 千个类别。A.Krizhevsky等在 2012 年通过训练一个大的深度神经网络来对ImageNet LSVRC-2010中包含着1000个不同类别的120万个高分辨率图像进行分类。在测试数据中，他们在 top-1 , 和top-5上的错误率是 37.5%和 17.5%，刷新了这个数据集的最好记录。

2014 年 Sun Yi 等提出了深度隐藏身份特征（Deep Hidden Identity Feature， DeepID）的方法去学习高等级特征表征来进行人脸识别。通过将人脸部分区域作为每个卷积网络的输入，在底层中提取局部低等级特征，并在深度卷积网络的最后一层隐藏层的神经元激活值中形成 DeepID特征，试验结果显示 Yi 等在 LFW上获得了 97.45%的准确度。

***自然语言处理***

自然语言处理（Natural Language Processing，NLP）意在将人类语言转换到能够容易地被计算机操作的表征的过程。大多数的研究者将这些问题分离式考虑，例如词性标注、分块、命名实体识别、语义角色标注、语言模型和语义相关词等，而没有注重到整体性，使得自然语言处理领域中的进展不是很乐观。具体来说现有的系统有3个缺陷：①它们都是浅层结构，而且分类器通常是线性的；② 对于一个效果好的线性分类器来说，它们必须事先用许多人工特征来预处理；③ 从几个分离的任务中进行串联特征以至于误差会在传播过程中增大。

2008 年 R.Collobert 等通过将一个普通的深度神经网络结构用于NLP，在“学习一个语言模式”和“对语义角色标签”任务上通过将重点关注到语义角色标签的问题上进行了没有人工设计特征参与的训练，其错误率为14.3%的结果刷新了最好记录。

信息检索

信息检索（Information Retrieval，IR）就是用户输入一个查询到一个包含着许多文档的计算机系统，并从中取得与用户要求所需最接近的文档[2]。深度学习在 IR上的应用主要是通过提取有用的语义特征来进行子序列文档排序，由 R.Salakhutdinov等在 2009 年提出，他们针对当时最广泛被使用在文档检索上的系统 TF-IDF上的分析，认为 TF-IDF系统有着以下的缺陷：在词计数空间中直接计算文档的相似性，这使得在大词汇量下会很慢；没有使用词汇间的语义相似性。因为在DNN模型的最后一层中的隐藏变量不但在使用基于前向传播的训练后容易推导，而且在基于词计数特征上给出了对每个文档更好的表征，他们使用从深度自动编码器得到的紧凑的编码，使得文档能够映射到一个内存地址中，在这个内存地址中语义上相似的文档能够被归类到相近的地址方便快速的文档检索。从词计数向量到紧凑编码的映射使得检索变得高效，只需要更便捷的计算，更少的时间。

2014年 Shen Yelong等提出了卷积版的深度结构语义模型（Convolutional Deep-Structured Semantic Modeling，C-DSSM），C-DSSM 能将上下文中语义相似的单词通过一个卷积结构投影到上下文特征空间向量上，从之前 43.1% 的准确率提高到了44.7%。

总结

介绍了AE、SAE、CNN，

接着介绍了四大应用领域分别为：语音识别、计算机视图、自然语言处理、信息检索。