深度学习——人工智能研究的新前沿

I.介绍

模仿人类的大脑表示信息的效率和鲁棒性一直是近几十年来人工智能研究的一个核心挑战。

一天中的每秒，人都在接收无数感官数据，并能在某种程度上通过允许在未来能以简洁方式使用的方式捕获数据的重要方面，通过允许在未来能以简洁方式使用的方式。50年多前，理查德·贝尔曼，一个提出动态规划理论和开创最优控制领域的人，断言数据的高维是在许多科学和工程应用中的基本障碍。主要的困难特别是在模式分类应用中，是学习复杂性随着数据的维数线性增加呈指数增长。他把这个现象叫做维度诅咒[1]。克服“诅咒”的主流方法是预先处理数据，通过这种方式将它的维数减少到它可以有效进行处理的维数，例如通过分类引擎。这个维数降低方案通常被称为特征提取。其结果，可以认为，后面许多模式识别系统的智能已转移到工程化的特征提取过程中，这有时是很有挑战性的，且高度依赖于应用[2]。此外，如果提取的特征是不完整的或错误的，分类过程就会天生的被限制性能。最近的神经科学的研究结果提供了在哺乳动物大脑对信息表示的管理原则，引出了设计代表信息的系统的新想法。一个重要的发现是，被认为与许多的认知能力有关的新皮层，并没有明确预处理感觉信号，而是允许它们通过一个复杂的层级传播[3]的模块，随着时间的推移，基于它们显示出[4]的规律学习表示观察结果。这一发现促使深度学习出现了一个子域，它侧重于用于与大脑皮层有相似特征的信息表达的计算模型。

除了现实生活中的数据的空间维度，时间分量也起着关键的作用。图案的观测序列往往传达一个意思给观察者，然而，该序列的独立片段将很难孤立地解读。含义往往是从在时间上最近收到事件或观察上推断出的[5] [6]。为达此目的，对观测物体的时间部分进行建模在有效的信息表示中的起着关键作用。基于观测的规律捕获时空相关性，被视为深度学习系统的一个基本目的。

假设健壮的深度学习已经被实现，它可能在一大组观察结果上训练出这样一个份层次的网络，然后从这个网络提取信号到一个相对简单的分类引擎以达到健壮的辨别图片的目的。稳健在这里指的是对各种各样的变换和扭曲，包括噪声，缩放，旋转，不同的光照等条件下，位移等表现出来的分类不变性。

本文提供了在过去十年的主流深度学习的方法和研究方向的概述。必须强调的是每种方法都有优点和缺点，这取决于它被应用到的应用程序和环境。因此，本文提出了对深度学习领域当前状态的总结和一些观点以推测到它可能如何演变。首先关注卷积神经网络（CNN）和深度信念网络（DBN）（及其各自的变化），因为它们都在深度学习领域建立的很好，并在今后的工作中大有希望。第二节介绍了CNN，并随后在第三节详细介绍了DBN。为更好的进一步深入了解这些技术的基础，读者可以参考文献[7]。第四节包含了近期提出的其他的深度架构。第五节包含一个简短的说明，这个说明关于这个研究是如何已经影响政府和行业活动的。结论提供了对深层次架构的潜在影响的看法，以及仍然存在需要回答的关键问题。

II. 卷积神经网络

CNN[8] [9]为二维数据如图像和视频特别设计的多层神经网络家族。CNN受早期的时间延迟神经网络（TDNN）影响。TDNN要求通过在时间维度共享权数降低学习计算需求，且旨在对语音和时间序列进行处理[53]。CNN是第一个真正成功的深度学习的方法，其中一个层级的许多层被成功地训练出一种健壮的方式。

A CNN is a choice of topology or architecture that leverages spatial and temporal relationships to reduce the number of parameters which must be learned and thus improves upon general feed-forward back propagation training.

CNN的是拓扑结构或架构，充分利用空间和时间的关系，以减少必须学会参数的数量，从而提高了在一般的前馈反向传播培训的选择。

CNN作为出于最小数据预处理要求的深度学习框被提出。在CNN中，所述图像（称为一个本地接受域）的一小部分被当作输入该分层结构的最低层。信息通常通过网络的不同层传播，由此在每一层都会应用数字滤波以获得所观察到的数据的显着特征。该方法提供一定程度的对平移、比例和旋转的不变性，同时本地接受域允许神经元或处理单元连接到基本特征如面向边或角。

其中一个关于这个专题的重要论文[8]介绍CNN对笔迹分析问题的应用。本质上，该输入图像和一组N小的过滤器进行卷积运算，其系数被训练，或使用一些标准预先确定。因此，网络的第一（或最低）层由“特征图”组成，这是卷积过程的结果，且具有正偏差以及可能的对特征的压缩或正常化。此初始阶段之后是子采样（通常为2x2平均运算），该阶段进一步降低维数，并提供一定的对空间变化的鲁棒性（见图2）。二次取样特征图然后接收加权和训练的偏差，最终通过激活功能传播。这方面有一些变体存在，为尽可能少每层一个图[13]或多个图的总和[8]。

当加权小时，激活函数是近似线性的，其结果是所述图像的模糊; 其他加权可导致激活的输出类似于一个AND或OR功能。这些输出形成一个新的特征图，然后新的特征图通过另一个卷积序列，子采样和激活功能，如图3所示。该过程可以重复任意次数。应当指出的是，随后的层可以结合一个或多个先前的层; 例如，在[8]中，最初六个特征图相结合以形成在后继层的16特征图。如在[33]中描述的，CNN通过被称为“特征池”的方法创建他们对对象的翻译不变（如图3中S层所示）。然而，特征池是网络组织者手工制作的，没有经过系统的培训或学习; 在CNN中，该池通过在学习过程中由参数进行“调整”，但基本的机构（例如，输入到S层的组合）由网络设计者设置。最后，在过程的最后阶段，激活输出被转发到产生该系统的最终输出的常规的前馈神经网络。

在CNN中，层次和空间信息之间的亲密关系使得它们非常适合于图像处理和理解，他们通常在从图像中自主提取显着特征方面表现良好。在某些情况下，Gabor滤波器已被用来作为初始的预处理步骤以模拟到人类对视觉激发的视觉响应[10]。在最近的工作中，研究人员已应用CNN到各种机器学习的问题上，包括人脸检测[11] [13]，文档分析[38]，和语音检测[12]。CNN最近[25]用时间一致的对象训练以在视频中能达到帧到帧的一致性，尽管这目标不必是特定于CNN的。

III. 深度信念网络

DBN，在最初在[14]中提出，与传统的神经网络的判别性质形成对比概率生成模型。生成模型提供观察数据和标签的联合概率分布， P(Observation|Label)和P(Label|Observation)的估计更容易，而判别模型受限于后者，P(Label|Observation)。DBN解决所遇到的问题时，采用传统的反向传播至深层次的神经网络，即：（1）用于训练的大幅标记的数据集的必要性，（2）慢的学习次数（即收敛），（3）导致不良局部最优的不足的参数选择技术。

DBN由几层波尔兹曼机构成，是一种神经网络（见图1）。这些网络被“限制”在层之间形成的（在层内单元未连接）一个可见层和一个隐藏层。隐藏的单元被训练用来捕捉那些在可见的单元观察到的高阶数据的相关性。最初，除了形成一个相联存贮器的顶部的两个层， DBN仅由直接自上而下生成的权重连接。RBM作为结构单元在更传统的和深层S形信念网络很有吸引力，因为它们易于学习这些连接权。为了获得生成的权重，在初始预训练以无监督的贪婪的层-层的方式进行，通过Hinton称作对比差异的东西使能[15]。在这个训练阶段，一个矢量v被呈现给可见单元并转发值到隐藏的单位。朝着相反的，在可见单元输入然后随机的在重建原始输入的尝试中找到。最后，这些新的可见神经元的激活被转发，使得重建隐藏单元的激活的步骤，h，可以实现。执行这些来回步骤被称为Gibbs抽样的方法，在隐藏激活和可见输入的相关的差异形成了权值更新的基础。训练时间显著降低，因为它可以证明，对于近似最大似然学习只有一个步骤是必要的。添加到网络中的每个层都改善了训练数据的对数概率，我们可以认为对数概率是增加真正代表性功率的。这项有意义的扩展，与未标记数据的利用相结合，在任何深度学习应用中都是重要组成部分。

在顶部两层，权重被连接在一起，以使得下层的输出为顶层联系其存储器内容提供参考线索或连结。我们经常遇到的问题中，判别性能的最终焦点，如在分类任务中。DBN可以通过反向传播利用标记的数据在预训练之后进行微调以提高判别性能。此时，一组标签被附着到顶层（扩大相联存储器），以澄清网络中类别的边界，通过这样，一组新的自下而上，识别权重被学习。[16]已经表明，这种网络经常表现的比那些训练有素只有反向传播的要好。这可以通过以下事实直观的解释。DBN相对于传统的前馈神经网络来说，反向传播时，才需要在权重（参数）空间中进行本地搜索，加快了训练和收敛时间。

运用DBN到MNIST手写字符识别任务时获得的性能结果表明显著改善了前馈网络。引入DBN之后不久，[17]中更透彻的分析固化了它们是无监督的任务以及连续值的输入的特点。[18] [19]进一步的测试表明DBN（以及其它深架构）面对变化不停增长的问题的恢复力。

通过引入卷积深层信念网络（CDBN）的概念，DBN的灵活性最近扩大[20]。DBN不固有地嵌入关于输入的图像的二维结构的信息，即输入是简单矢量格式的图像矩阵。与此相反，CDBN利用相邻像素的空间关系，引入被称作卷积RBM来提供一种具有高尺寸图像缩放以及平移不变生成模型的空间关系。DBN目前没有明确处理学习观测量之间的时间关系，虽然最近有工作在堆叠颞RBM [22]或这些的概括，被称为颞卷积机[23]，用于学习的序列。这样的序列学习者被应用到音频信号处理问题上，从而DBN最近已经取得了进展 [24]，提供了一个途径令人振奋的未来的研究方向。

用于DBN和CNN的静态图像测试中最常见的是手写数字的MNIST数据库[27]和加州理工学院-101数据库的各种对象（属于101类）[28]。每个体系结构的分类错误率可以在[19] [20] [21]中找到。[27]提供了应用到MNIST数据库的各种机器学习技术的全面且最新的性能比较。

最近有关DBN的作品包括使用堆叠自动编码器代替传统DBN下的RBM[17] [18] [21]。这种努力产生的深多层神经网络架构可以用和DBN相同的原则训练，但在对层的参数更不严格。不像DBN，自动编码器使用的是输入样本空间不能由结构进行采样的判别模型，使得它更难以解释在其内部什么网络被捕获。然而，[21]已经表明，去噪的自动编码器，在训练期间会随机失效，和传统DBN相比可能被堆叠以产生泛化性能（且在某些情况下，更好的是DBN）。对于单个去噪自动编码器训练过程对应于用于生成模型如RBMS的目标。

IV.最近提出的深度学习架构

有几个尝试模拟大脑皮层的计算结构。

这些模型已被灵感来源，如[42]，它试图在图像理解的各种计算相映射到皮层区域。