第五章课后答案

1、基于规则的方法依赖于人工提取的规则所建立的各类语言知识库，这不仅需要大量繁杂的人工工作，还需要一定的领域知识；而基于统计的方法不需要语言知识库，减少了人工工作，但其依赖于大量的语料库。如果语料库很小，并不能得到很好的效果。

2、统计方法的基础是数据的概率分布。预训练模型学习到了数据的先验分布。从模型参数的角度来讲，预训练为神经网络设置了一个好的初始权重，这样模型可以收敛到更优的点。

3、自然语言理解：词性标注与分词、语法解析、语言模型、情感分析、命名实体识别；自然语言生成：对话系统、机器翻译、文本摘要。

第六章课后答案

1、神经网络每一层都经过激活函数输出，通过反向传播来优化参数。反向传播实际上就是对参数求导，然后使用梯度下降法来调整每一层参数。一个常见的激活函数是，它的梯度不会超过0.25。这样如果“无限”的加深网络，就会发现梯度更新信息将会以指数形式衰减，即发生了梯度消失。越靠近输入层时，由于梯度消失现象，会导致靠近输入层的隐藏层权值更新缓慢或者更新停滞。这就导致在训练时，只等价于后面几层的浅层网络的学习。

2、循环神经网络某一时刻的输出依赖于上一时刻，因此无法并行化。对很长的序列来说循环神经网络相当于一个深层的神经网络，训练时很容易发生梯度消失现象，因此无法解决长时依赖问题。

3、激活函数的主要作用是为网络引入了非线性，一个非线性的函数适合作为激活函数。此外由于神经网络使用梯度信息更新参数，激活函数的梯度不能趋于无限，例如就不适合作为激活函数。

4、不可以。将参数全部初始化为0会导致反向传播更新参数时，对每一层参数的更新都是相同的。这意味这每一层学习到的特征都是相同的，很可能导致模型失效，无法收敛。

5、(1)数据增强，复制现有数据并加随机噪声，重采样。

(2)参数正则化（权值衰减）,在损失函数和模型复杂度之间进行折中，可以采用L1和L2，其中L1采用的是拉普拉斯先验，倾向于聚焦网络的权值在相对少量的高重要连接上，而其他权重就会趋向于0，而L2正则采用的是高斯先验。

(3) Dropout，随机地删除网络中的一般隐藏的神经元，并且让输入层和输出层的神经元保持不变。每次使用梯度下降时，只使用随机的一般神经元进行更新权值和偏置，因此我们的神经网络是在一半的隐藏神经元被丢弃的情况下学习的。

(4) 早停法，在初始化网络的时候一般初始化较小的权值，训练时间越长，部分网络权值可能越大。如果我们在合适的时间停止训练，就可以将网络的能力限制在一定的范围内。

82页：6.7引言

深度神经网络、卷积神经网络、循环神经网络是目前最常见的、主流的神经网络。除了这些算法外，本节简要介绍了一些其他常见的神经网络。

第五章拓展阅读

世界上第一台聊天机器人——Eliza。

人工智能历史上最为著名的软件——Eliza，也是最早的与人对话程序，是由系统工程师约瑟夫·约瑟夫·魏岑鲍姆和精神病学家肯尼斯·科尔比在20 世纪60 年代共同编写的。是世界上第一个真正意义上的聊天机器人。他们将程序命名为Eliza，灵感来自于英国著名戏剧家肖伯纳的戏剧《偶像》中的角色，它能够使计算机与人用英语谈话。在自然语言理解技术尚未真正取得突破性进展时，这是一个令人费解的现象。

魏岑鲍姆在他的《计算机能力与人类理性》一书中记述了一次Eliza典型的对话交流。

“男人都是一样的。”一位年轻女子给伊丽莎发出了这样一句话。

“这是从何说起呢？”伊丽莎问。

“他们总是用这样那样的事来烦我们。”

“你能想出一个具体事例吗？”

“比如我男朋友非让我来这里。”

“是你男朋友让你来这儿的？ ”

“是，他说我总是很沮丧。”

“听到你不开心，我很难过。”

Eliza得到了最简单形式的自然语言——模式匹配的支持。

而魏岑鲍姆同时也以“Eliza”质疑图灵所提出的“图灵测试 (Turing test)”，原因在于“Eliza”程序运作建立在以人为主的互动模式，亦即针对人类提问内容分析主词关联，并且找到其中关键字词，做出相应回答。

其中更加入对话引导的心理应用，让“Eliza”能依循提问内容重复说词，或是针对关键字词进行回答，借此满足提问者内心预期听到答案，进而达成让提问者认为对话对象是真人的目的。

因此魏岑鲍姆将上述情况命名为“Eliza”效应，认为并非人工智能理解人类想法，而是在与人类互动过程中所展现反应，让人类更愿意与其互动，甚至相信“Eliza”是真实存在人类。

第六章拓展阅读

杰弗里·辛顿（Geoffrey Hinton），被称为“神经网络之父”、“深度学习鼻祖”，他曾获得爱丁堡大学人工智能的博士学位，并且为多伦多大学的特聘教授。在2012年，辛顿还获得了加拿大基廉奖(Killam Prizes，有“加拿大诺贝尔奖”之称的国家最高科学奖)。2013年，辛顿加入谷歌并带领一个AI团队，他将神经网络带入到研究与应用的热潮，将“深度学习”从边缘课题变成了谷歌等互联网巨头仰赖的核心技术，并将反向传播算法应用到神经网络与深度学习。

早在1960年代，杰弗里·辛顿还在读高中，就有一个朋友告诉他，人脑的工作原理就像全息图一样，创建一个3D全息图，需要大量的记录入射光被物体多次反射的结果，然后将这些信息存储进一个庞大的数据库中。大脑储存信息的方式居然与全息图如此类似，大脑并非将记忆储存在一个特定的地方，而是在整个神经网络里传播。这是杰弗里第一次真正认识到大脑是如何工作的，对他来说，这是人生的关键引导，也是他成功的起点。

为了对神经网络刨根问底，杰弗里·辛顿进阶求学期间，在剑桥大学以及爱丁堡大学继续他的神经网络研究。在剑桥大学的心理学专业的本科学习当中，杰弗里·辛顿发现，科学家们并没有真正理解大脑，人类大脑有数十亿个神经细胞，它们之间通过神经突触相互影响，形成极其复杂的相互联系，然而科学家们并不能解释这些具体的影响和联系，神经到底是如何进行学习以及计算的，对于杰弗里，正是他所关心的核心问题。

1986年，杰弗里·辛顿联合同事大卫·鲁姆哈特（David Rumelhart）和罗纳德·威廉姆斯（Ronald Williams），发表了一篇突破性的论文，详细介绍了一种叫作反向传播算法。通过推导人工神经网络的计算方式，反向传播可以训练多层神经网络。

然而提出反向传播算法之后，杰弗里·辛顿并没有迎来事业的蓬勃发展，1980年代末期，第二波人工神经网络热潮带来大量投资，然而因为1987年全球金融危机和个人计算机的发展，人工智能不再是资本关注的焦点。同时，当时的计算机硬件无法满足神经网络需要的计算量，也没有那么多可供分析的数据，辛顿的理论始终无法得到充分实践。到1990年代中期，神经网络研究一度被打入冷宫，辛顿的团队在难以获得赞助的情况下挣扎。

2004年，学术界对他们的研究仍兴趣不大，而这时距离他们首次提出“反向传播”算法已经过了近20年。但也就在这一年，靠着少量的来自CIFAR以及LeCun和Bengio的资金支持，杰弗里·辛顿创立了Neural Computation and Adaptive Perception（NCAP，神经计算和自适应感知）项目，该项目邀请了来自计算机科学，生物，电子工程，神经科学，物理学和心理学等领域的专家参与，当时应该算是一个创举，因为此前，科学家和工程师们各自为政，很少交叉合作。

2011年，NCAP研究成员同时也是斯坦福大学的副教授Andrew Ng在Google创立并领导了Google Brain项目。 2012 年，计算机硬件的性能大幅提高，计算资源也越来越多，他的理论终于能在实践中充分发展。他带领两个学生利用卷积神经网络 (CNN) 参加了“ImageNet大规模视觉识别挑战”比赛。比赛的其中一项是，让机器辨认每张图像中的狗是什么类型，从而对100多种狗进行分类。在比赛中，杰弗里·辛顿带着他的学生以16%的错误率获胜——这个错误率低甚至低于人眼识别的错误率18%，并且远低于前一年 25%的获胜成绩，这让人们见识了深度学习的潜力，从此，深度学习一炮而红。