speech and language processing 第七章报告

09017244 郑健雄

1. 自己提出的问题的理解：
2. 比起logistic regression等传统模型，神经网络模型的优势体现在哪里？

讨论后的理解：无论是logistic regression还是svm实际上对于复杂的输入形式的表达都比较困难，即使可以使用一些办法来处理这些问题。而神经网络的优势在于其可以自动学习出一种表示的形式，比如异或问题的处理。神经网络可以更好地表达复杂的输入，比普通模型更优，同时其具有较好的抽象性，随着网络层数的增多，其学习到的特征可以进一步加深，从而让学习结果可以泛化。

1. 基于神经网络的LM比起传统的基于贝叶斯的LM的优点是什么？为什么会有好的效果？

讨论后的理解：一方面，神经网络使用了词向量的形式，从而和之前说到的dense vector联系起来，从而可以在训练过程中考虑到词汇本身的语义问题，此外，其可以很好地克服贝叶斯LM下受到数据影响的部分，比如说词频为0等情况。

1. 别人提出的问题的理解
2. P15页7.5中说neural language model不需要smothing是指什么吗？handle much longer histories是具体什么意思？

自己的理解：考虑到统计词频等问题，n-gram需要进行一些加一的操作来让计算结果不要出现0或者除以0的情况，神经网络可以不考虑这个问题。longer histories应该是说其可以考虑更多的前文内容，而不是像n-gram只能使用一种局部的前后几个词之间的关联。

1. P15页为什么神经网络学习中是初始化权重用一些小的随机数？以及为什么输入值均值为0，用有统一的方差效果较好？

自己的理解：权值过大的话，往往会出现影响的不均衡的情况，使得某个参数的影响变得很大，但实际的情况并不是这样，比如说词频等等，所以从小的随机值开始训练并且限制可以减少这种情况，减少过拟合。统一化的方差以及正则化也是这样，为了减少不同维度造成的尺度差异，而将其正则化，从而可以更好地进行训练。

1. 读书计划

1、本周完成的内容章节：阅读第8章。

2、下周计划：阅读完speech and processing第9章内容。

四、读书摘要总结：

1. Neural Network：神经网络是模仿人的神经组织结构提出的一种数学模型，但是其本质上和人类神经系统并不相同，单层神经网络相当于感知机模型。其基本单元是神经元，神经元的工作是对于输入的值乘以一个权值向量然后加上bias，并且使用激活函数例如sigmoid，tanh等。

2. 在一个全连接，前向的网络中，第i层的每一个神经元都会和第i+1层的每一个神经元相连，并且没有环路。其训练思路也是使用梯度下降来优化每个神经元的参数，从而让其产生最好的效果。由于其可以接受很多输出并且使用了激活函数，其可以构造出弯曲的边界，在处理XOR这类问题中有着显著的优势。

3. 前馈神经网络一般是多层神经网络，包括隐藏层，输入层和输出层，而loss function一般是来衡量标注结果和预测结果之间的距离，并以此为基础进行优化。通过计算其gradient并且使用反向传播的方法，其可以迭代的对每一层的权值进行更新。反向传播是基于计算图的一种计算方法，其本质上还是链式法则在矩阵中的应用。

4. 在实际使用中，神经网络可以用在LM，标签分类等问题上，并且有着很好的效果。其训练可以使用已经给出的embedding作为输入，也可以自学习embedding，一般word2vec的训练结果使用较多。