

# 读书报告

09118119 黄一凡

2020 年 8 月 17 日

## 1 别人提出的问题

### 1. Neural Language Model 为什么不需要 smoothing?

考虑到统计词频时有可能出现 0 的情况，n-gram 需要进行正则化来保证运算结果中不会出现除以 0 的情况。而在神经网络中则不需要考虑这个问题。Long history 指的是更多的考虑前文内容，而不是像 n-gram 一样只能考虑前后几个词的关联。

### 2. 为什么神经网络的初始权重通常会采用一些较小的随机数?

因为如果不是使用一些较小的权值，而是会存在较大的权值，那么很有可能会出现一种不均衡的情况，且可能导致过拟合的情况。

### 3. 为什么神经网络的输入值要求均值为 0，方差为规定值?

这一步骤通常被称为 normalize，也就是正则化。正则化的目的是为了减少一些较为特殊的值，比如这个值很大或者与其他值的 scale 不同来影响模型的准确度和避免过拟合。

4. 比起传统模型，神经网络模型的优势体现在哪里？神经网络可以自动学习出一种对输入的表现形式，比如解决异或问题，而这点是传统的模型所不能达到的。而且，神经网络模型较为复杂，因此也就可以学习到更为复杂的知识。神经网络的鲁棒性较好，对于不同的任务需要调整的参数较少。最后，神经网络非常适合于迁移学习，因为它可以很好地提取出事物的抽象特征，这一点对于基本上所有的学习任务都是适用的，因此其可以很方便地用于迁移学习。

## 2 读书计划

本周所读: 《Speech and Language Processing》 Chapter 7

下周计划: 《Speech and Language Processing》 Chapter 8