读书报告 7.27-8.2

61518426 周之遥

·情况说明

由于之前两周在学习LSTM、Transformer、BERT等模型，读书中断了一段时间，所以较进度慢一章，之后会跟上。

·读书内容

学习第3章N-gram Language Models。

语言模型（Language Model）的目标是根据已有的文本预测接下来的词汇，以及为句子赋予概率。语言模型在各方面都有很大作用。（语音识别、拼写纠错、语法纠错、机器翻译等）

**N-gram**是一种典型的简单的语言模型。为了根据已有的文本求出接下来词汇的概率或是整个文本序列的概率，将概率分解为条件概率之积。显然不能直接求（否则需要大量的数据和计算量），N-gram做出的假设为当前单词由之前的n-1个单词所决定，特别的，bigram中只考虑前一个单词。

那么这些条件概率怎么求呢？N-gram的办法是极大似然估计，也就是简单的频率估计。用前后组合出现的次数除以前面单独出现的次数，即为条件概率。注意在一个句子开头或结尾需要加上相应数量的特殊标识。

根据实验，一般N越大模型的效果就越好。

N-gram有两个问题，一个是会出现很多不合理的概率0的组合（因为在训练集中没有出现），这就需要进行平滑，即将已有的概率分给概率为0的组合，常见的平滑方法有**Laplace Smoothing**、**Add-k Smoothing**、**backoff**、**interpolation**。**backoff**的方法是当出现概率0时，使用较小的N-gram模型；另外**interpolation**则是将不同的N-gram模型组合起来。

**Kneser-Ney Smoothing**是一种平滑方法，在discount已有的概率时，该方法采取的办法是discount固定的次数。（absolute discounting），同时与较小的N-gram模型结合时，较小的N-gram模型采用的不是简单的频次计数，而是组合的种类计数，以找到更广泛适用而非更经常出现的单词。

**Stupid Backoff**是简单但有效的平滑方法，当出现概率0时，直接使用较小N-gram的模型而不考虑所有概率之和为1。

一般在语言模型的训练中，将数据集分为8：1：1，分别用于训练、调整参数和最终测试。困惑度是语言模型的常用评估标准，为概率倒数的N次开根号，其中N为长度。