2.1 CNN介绍

2.2.1 CNN介绍

卷积神经网络（Convolutional Neural Network）最早是应用在计算机视觉当中，而如今CNN也早已应用于自然语言处理的各个任务中；

介绍卷积神经网络的各个重要环节：卷积，池化，激活函数；

首先介绍卷积神经网络中卷积的运算，可以简单的将卷积运算理解为作用于矩阵的一个滑动窗口函数，滑动窗口也可以称作卷积核、滤波器或特征检测器，在图像中滤波器通常是对图像的一小块区域进行计算，而在文本中，是以一句话所构成的词向量作为输入，每一行代表一个词的词向量，所以在处理文本时，滤波器通常覆盖上下几行的词，滤波器的宽度与输入词向量的宽度相同，通过这种方式，多个连续词之间的特征就能够被捕捉到，并且能够在同一类特征计算时共享权重，下图（表示的是一个输入矩阵经过卷积运算时后得到的结果）

卷积神经网络的另一个重要概念是池化层，一般是在卷积层之后，池化层对输入做降采样，池化的过程实际上是对卷积层分区域求最大值或者对每个卷积层求最大值，在自然语言处理中，通常是对整个输出层做池化，每个卷积层只有一个输出值，（下图所示为池化操作的过程）

池化首先是可以输出一个固定大小的矩阵，这对于自然语言处理当中输入句子的长度不一有非常大的作用。例如，如果使用200个卷积核，并对每个输出使用最大池化，那么无论卷积核的尺寸是多大，也无论输入数据的维度或者单词个数如何变化，都将得到一个200维的输出，这可以应对不同长度的句子和不同大小的卷积核，但总是得到一个相同维度的输出结果，用作最后的分类。

另外池化层在降低数据维度的同时还能够保留显著的特征。每一种卷积核都是用来检测一种特定的特征。在以句子分类中，每一种卷积核可以用来检测某一种含义的词组，如果这种类型的含义的词语出现了，该卷积核的输出值就会非常大，通过池化过程就能够尽可能地将该信息保留下来。

2.2 LSTM介绍

人类并不是每时每刻都从一片空白的大脑开始他们的思考，在阅读一篇文章的时候，都是基于自己已经拥有的对先前所见词的理解来推断当前词的真实含义，而不会将所有的东西都全部丢弃，然后用空白的大脑进行思考，也就是说人类的思想具有持久性，传统的神经网络并不能做到这点，例如，假设希望对电影中的每个时间点的事件类型进行分类，传统的神经网络很难来对这个问题进行处理，很难使用电影中先前的事件推断后续的事件；

RNN解决了这个问题，RNN不仅会学习当前时刻的信息，也会依赖之前的序列信息，由于其特殊的网络模型结构决定了其信息保存的能力，所以RNN对处理时间序列和语言文本序列问题有独特的优势，递归神经网络都具有一连串重复神经网络模块的形式，每个神经网络会把消息传递给下一个，如果将这个循环展开，则如下图所示：



图1：RNN网络的链式结构图

在上图所示的神经网络中，输入为，输出为，上的环允许将每一步产生的信息传递到下一步中，这种链式的特征揭示了RNN本质上是与序列和列表相关的，它们是对于这类数据最自然的神经网络架构；在标准的RNN中，这个重复的模块只有一个非常简单的结构，例如一个tanh层，如下图所示：



图2：标准RNN网络中重复模块的内部结构

长短期记忆网络（Long Short Term Memory Network,LSTM)，是一种改进之后的循环神经网络，可以解决RNN无法处理长距离的依赖所导致的梯度消失问题，以及由于依赖激活函数和网络参数而可能导致的梯度爆炸问题，近几年来被广泛应用，LSTM通过可能的设计来避免长期依赖问题，记住长期的信息在实践中是LSTM的默认行为，而非需要很大的代价才能获得的能力。

 图3：LSTM网络的内部结构

LSTM的关键点是细胞状态，也叫单元状态，就是穿过图中的水平线，单元状态类似传送带，它贯穿整个链条，只有一些线性的相互作用，这使得信息很容易的能以不变的方式向下流动；LSTM有能力向单元状态中移除或添加信息，通过门结构来管理，包括“遗忘门”，“输出门”，“输入门”，通过门让信息选择性通过，来去除或增加信息到单元状态，模块中sigmoid层输出0到1之间的数字，描述了每个成分应该通过门限的程度，0表示不让任何成分通过，而1表示让所有成分通过。

“遗忘门”决定从“单元状态”中丢弃什么信息；比如完形填空中填“他”或者“她”的问题，细胞状态可能包含当前主语的类别，当看到新的代词，希望忘记旧的代词。如上图所示，左边第一个即为遗忘门，将上一时刻的输出和这一时刻的输入进行拼接，然后判断以多大的程度来保留这部分信息，即得到概率值。，如下式所示，









“输入门”决定放什么信息到单元状态中；如图，第二个sigmoid层决定什么值需要更新，如式（2）所示，代表的是以多大的概率来更新信息；tanh层创建一个新的候选值向量，如式（3）所示，表示现在的全部信息，这两个操作都是为状态更新做准备；接着更新“细胞状态”；更新为，如式（4）所示，把旧状态与相乘，就丢弃掉了确定需要丢弃的信息，然后将乘以，以确定要更新的信息，通过相加操作得到新的细胞状态；

接着根据“细胞状态”得到输出，首先运行一个sigmoid层来确定细胞状态在哪个部分将输出，如式（5）所示，接着用tanh处理细胞状态，得到一个在-1~1之间的值，再将它和sigmoid门的输出相乘，如式（6）所示，得到确定输出的部分。





2.3 注意力机制介绍