**第十章读书报告**

09118223 吴亦珂

读书进度：第十章读完一半

1. **问题列表**

（我提出）P5页Decoder中提到a weakness是具体指什么？解决方法说把它当做一个参数加入进计算是指什么？

讨论结果：context vector会随着循环神经网络的迭代，影响会逐渐减弱。而如果把其变为一个参数输入，就相当于固定了这个内容，影响不会减弱。从而解决了这个问题。

（别人提出）Encoder-Decoder网络的优点和缺点有哪些？

讨论结果：最大的局限性就在于编码和解码之间的唯一联系就是一个固定长度的语义向量C。也就是说，编码器要将整个序列的信息压缩进一个固定长度的向量中去。但是这样做有两个弊端，一是语义向量无法完全表示整个序列的信息，还有就是先输入的内容携带的信息会被后输入的信息稀释掉，或者说，被覆盖了。输入序列越长，这个现象就越严重。这就使得在解码的时候一开始就没有获得输入序列足够的信息，那么解码的准确度自然也就要打个折扣了。

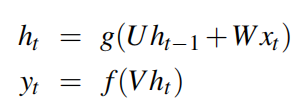
（别人提出）Encoder-Decoder网络在哪些NLP工作中用得比较广泛？

讨论结果：机器翻译、语音识别、文本摘要、问答系统。

1. **读书收获**

**10.1 Neural Language Models and Generation Revisited**

1. 简单的循环神经网络



V，W是权重矩阵，在训练过程中会发生变化，g是激活函数，f是softmax函数。

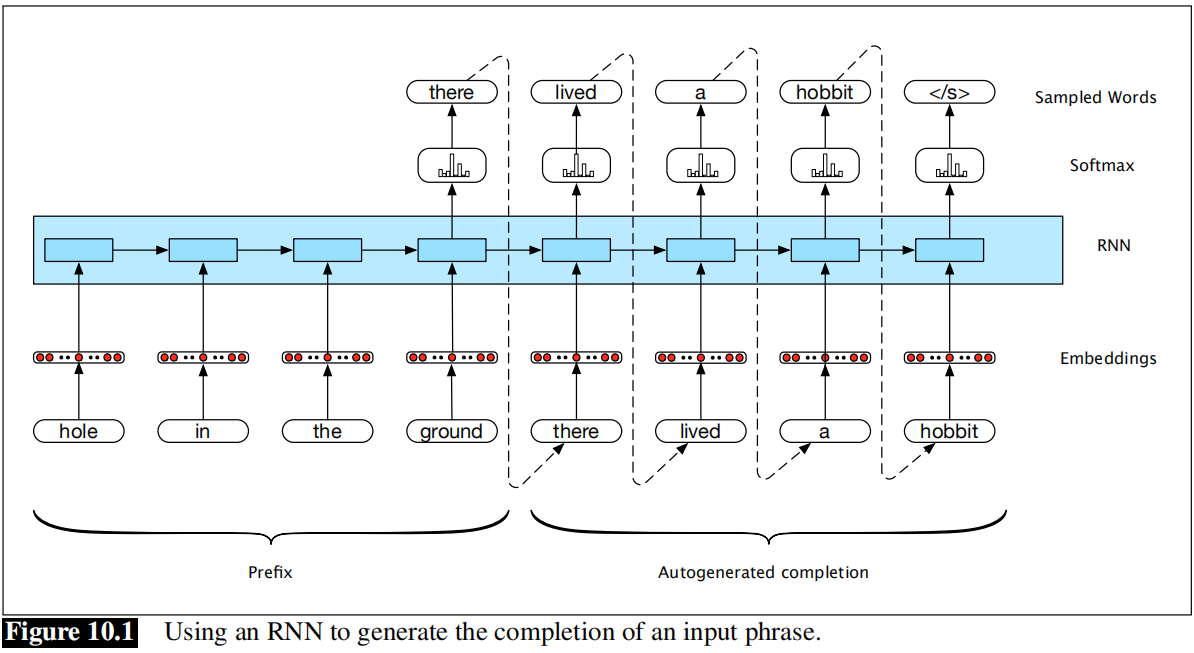
RNN-based language model：

首先任意选择一个合适的词作为句子的开头。

根据上一步的隐藏状态和上一步产生的词嵌入来产生当前的词。

Simple variation：

使用prefix产生隐藏状态，然后使用最后的隐藏状态作为起点，可以生成对于prefix的一个补全。



机器翻译：

训练翻译模型的材料：parallel text或bitext

需要翻译的文字是source，输出是target。

为了将上述模型用于机器翻译，需要在每一个bitext source句子末尾增加一个句子结束的标志，并且简单地将对应的标签连接在上面。

进行翻译时，首先通过网络使用前向推理的方法生成隐藏状态直到source的结尾，之后开始autoregressive generation,首先将最后一个单词的隐藏状态以及句子结束的标志作为输入，随后的单词都取决于前一个隐藏状态以及上一个词的embedding。

**10.2 Encoder-Decoder Networks**

Encoder-decoder networks有三部分组成：

一个encoder接受输入的句子，产生其语义表示。

可以使简单的网络，也可以是一个堆叠的网络，此时最上层的网络作为最终的表示形式。

一个context vector，是encoder产生的语义表示的函数，将其本质传递给decoder层。

一个decoder，接受context vector作为输入并且产生一个序列的隐藏状态，最终得到输出。

用于产生序列，每次输出一个，直到句子终止符。一个典型的方法是使用一个LSTM,GRU-based RNN。生成过程取决于之前的隐藏层以及之前状态的输出。

