**读书报告**

王贵涛

自然语言处理的第八章主要介绍了编解码器模型、注意力和上下文嵌入。

**重视神经语言模型与生成（Neural Language Models and Generation Revisited）**

在简单递归神经网络中，一个特定点的隐藏状态由之前的隐藏状态和当前输入决定，网络输出则是这个新的隐藏状态的函数。

为了建立一个基于神经网络的语言模型，训练网络使用语料库预测序列中的下一个单词，以这种方法训练的语言模型称为自回归模型。如果是给定前缀完成序列，首先使用前向推理产生一个以最后一个单词隐藏状态结束的隐藏状态序列。

机器翻译将一种语言翻译成另一种语言，训练文本主要是平行文本或比特文本，由不同彼此翻译的句子组成的大型文本集合。被翻译的文本称为源，翻译的输出称为目标。为了将语言模型和自回归拓展到机器翻译，在源句子的末尾添加结束标记，然后将对应的目标连接到它，作为组合语言的训练数据，对网络进行自回归训练。

**编解码器网络（Encoder-Decoder Networks）**

网络元素处理输入序列并组成编码器，生成输入的上下文表示，体现在最终隐藏状态，后者将其馈入第一个隐藏状态，解码器接受这种状态并进行一系列输出。

编码器-译码器由三个组件组成：编码器接受输入序列并生成上下文化表示，上下文向量（context vector）作用于上下文表示传给译码器，译码器以上下文向量（context vector）为输入生成任意长度隐藏状态序列得到输出状态序列。

简单RNNs，LSTMs，GRUs，卷积网络和transform网络都可以用作编码器。

译码器典型的方法是使用LSTM或基于GRU的RNN，其中上下文由编码器的最终隐藏状态组成，用于初始化译码器第一个隐藏状态。生成每个隐藏状态都以前一个隐藏状态为条件，并输出。这种方法缺点是上下文向量只在过程开始时使用，其影响会随着输出序列的生成而减弱。一种解决方案是在每一步都提供上下文向量。