

题目：Learned in Translation Contextualized Word Vectors

主要讲了基于背景的词向量自然语言处理。

这篇论文的灵感启发来源于类别图像识别。在图像识别中，图像中像素点之间的关系与图像的展示方式有关，也就是背景 context，因此联想到可以将词向量建立在背景的基础之上，由此获得更高的准确率。

文中用到了一种类似于 word2vec 的方法，叫 GloVe，除了像 word2vec 一样认为一个 word 与其周围的 word 有关之外，还加入了先验知识，也就是 word 与其他 Word 出现的概率。

模型的框架大概有如下演变过程：

1. 这是最基本的 LSTM encoder 框架，输入词向量，中间神经网络学习过程叫 encoder，得到的输出叫做 hidden vectors

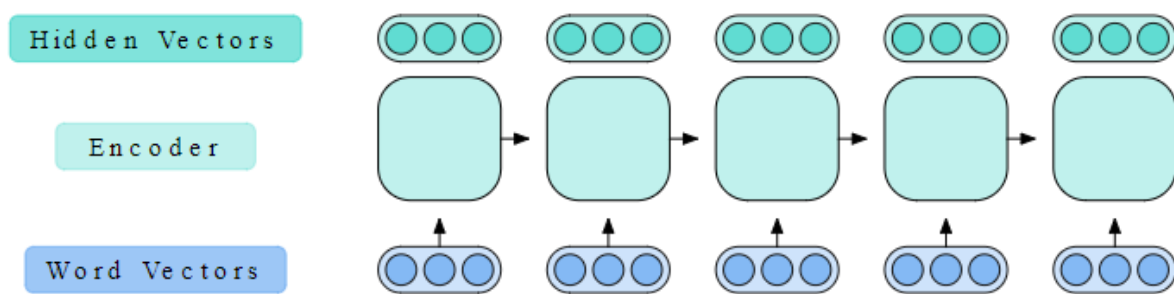


Figure 6: An LSTM encoder takes in a sequence of word vectors and outputs a sequence of hidden vectors.

2. 鉴于各个词向量之间的学习序列只能单向传播，为了克服这个问题，将其改为双向传播，形成如下双向 encoder

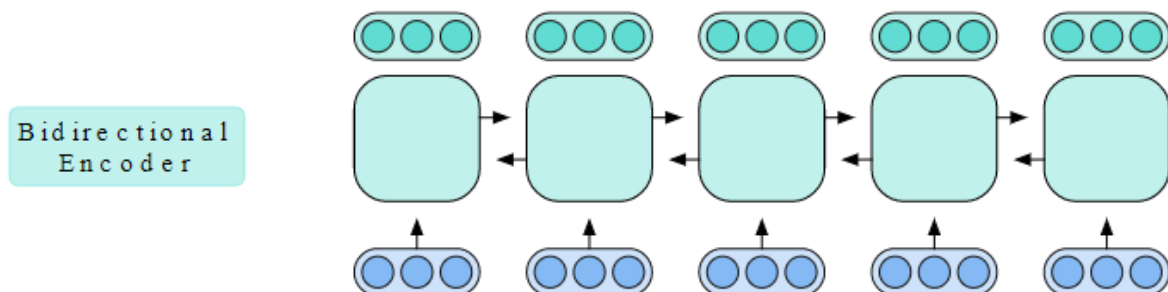


Figure 7: A bidirectional encoder incorporates information that precedes and follows each word.

3. 习得 encoder 之后，将其用于目标语言处理。这篇文章探讨的核心任务是德语和英语的翻译。先前的输入词向量为英语，得到对应 encoder。现在将德语词向量输入进去，输入到 encoder 中，获得对应 decoder 后的结果。

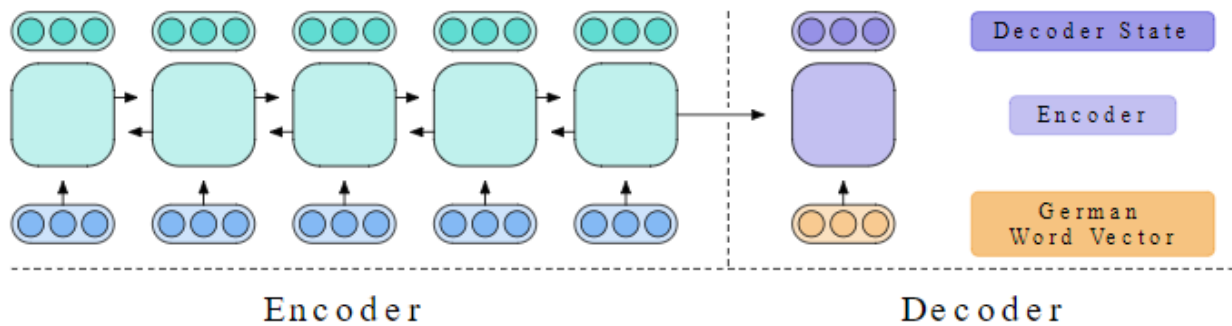


Figure 8: The decoder uses a unidirectional LSTMs to create the decoder state from input word vectors.

4. 为了决定下一步该翻译什么，加入 attention 模块，也就是每一步对所有隐藏节点进行权重判断，然后不断更新向量，这也是状态在不断变化的过程。

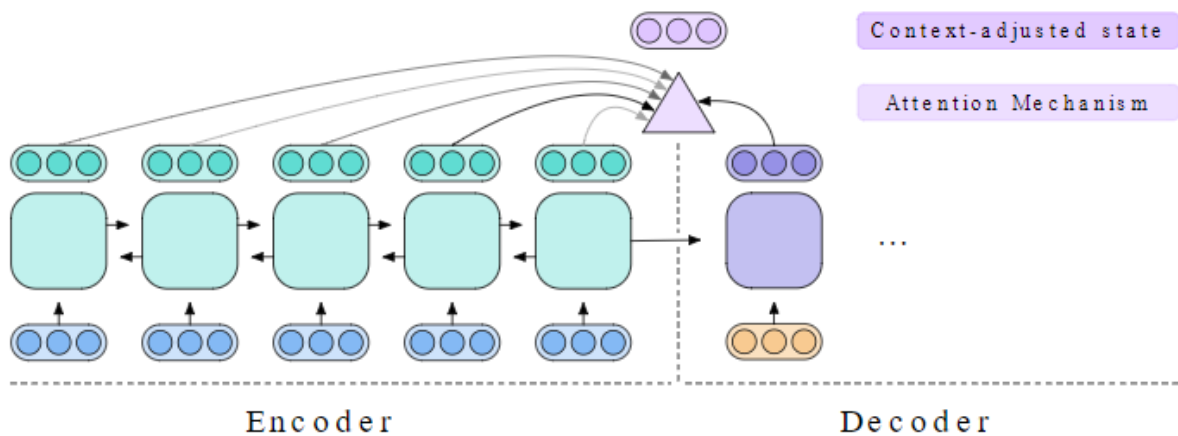


Figure 9: The attention mechanism uses the hidden states and decoder state to produce a context-adjusted state.

5. 最后，加上输出模块，就是翻译模型了

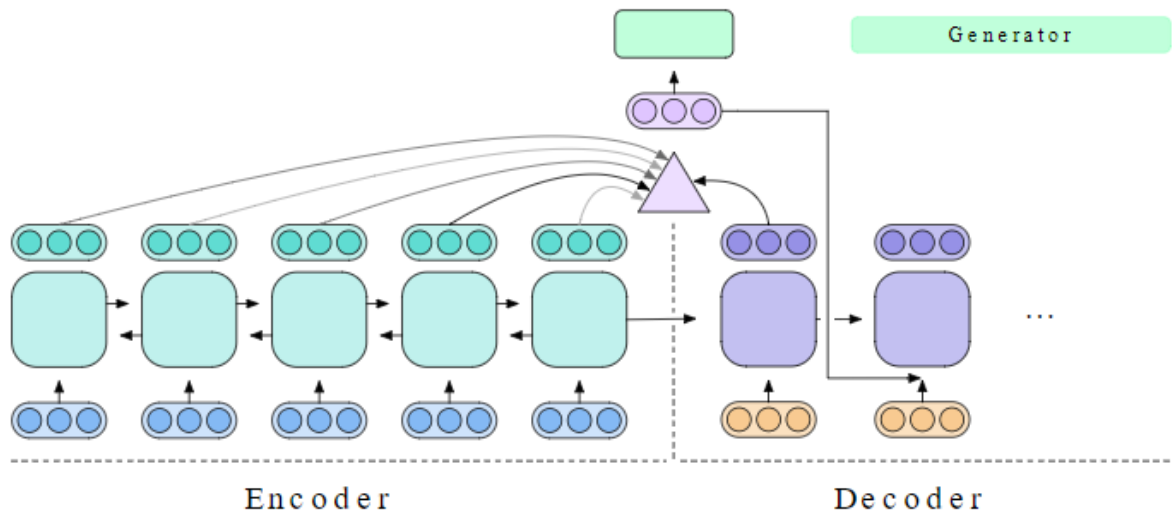


Figure 10: The generator uses the context-adjusted state to select an output word.

在实际应用中，使用词向量加上当前词向量训练出来的 encoder 当做 context，共同作为输入，来训练模型。

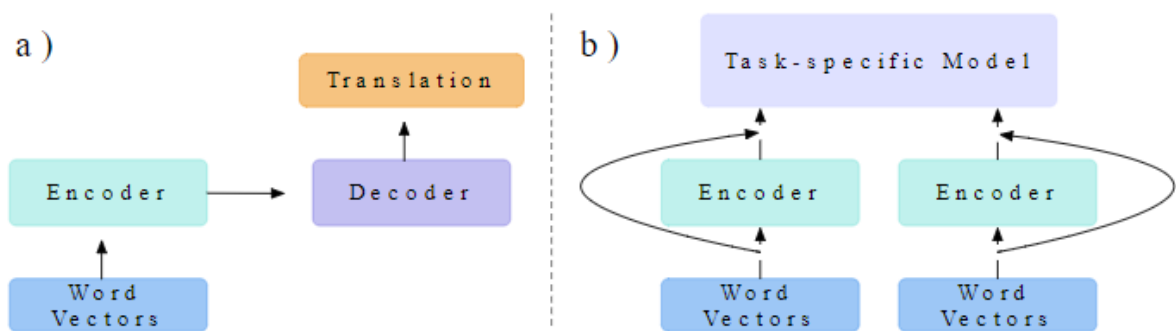


Figure 11: A general overview of how we a) train an encoder and b) reuse it as part of a new model.