

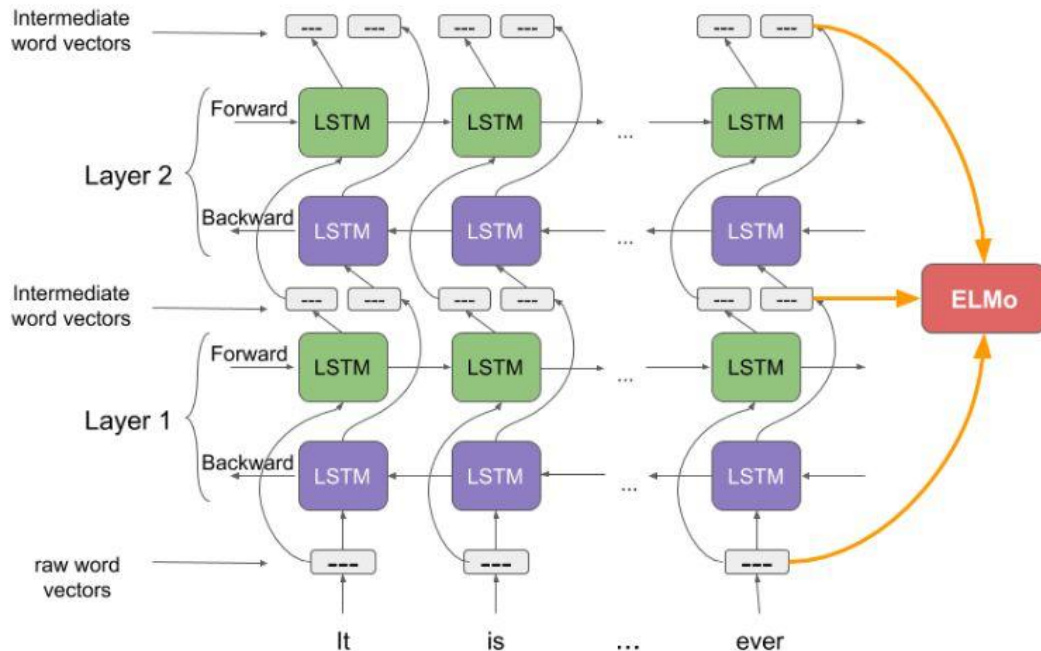
## Elmo

1. elmo 的网络结构为啥是这个样子？

1) raw word vectors 通过什么训练而来？

2) 虽然堆叠了两层的 lstm，但是双向并没有产生关系，只是在输出层做了拼接，为什么？如果双向产生了关系也并没有产生自己预测自己的问题

3) 多层不做 mask 会造成自己预测自己跟 elmo 这个有关系吗？



1. Raw word vectors 使用字符级卷积神经网络 (convolutional neural network, CNN) 来将文本中的词转换成原始词向量 (raw word vector)。这个 char-cnn 需要很大功夫学习，比较复杂。

2. 将这些原始词向量输入双向语言模型中第一层

3. 前向迭代中包含了该词以及该词之前的一些词汇或语境的信息

4. 后向迭代中包含了该词之后的信息

5. 这两种迭代的信息组成了中间词向量 (intermediate word vector)，这些中间词向量被输入到模型的下一层

6. 最终表示 (ELMo) 就是原始词向量和两个中间词向量的加权和

7. 加权的权值怎样确定？没找到

8. 为什么可以表示多义词。

动态 rnn，每个词由整句话的语义得到。当整个模型训练好了之后，输入一句话，会得到三层的词向量表示。因为对于不同的话，前后词是不一样的，所以词向量表示就是不一样的。