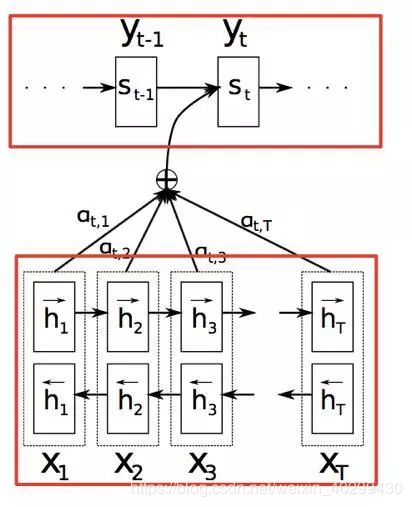
* 基本的Attention原理。
* HAN的原理（Hierarchical Attention Networks）。
* 利用Attention模型进行文本分类。

Attention原理

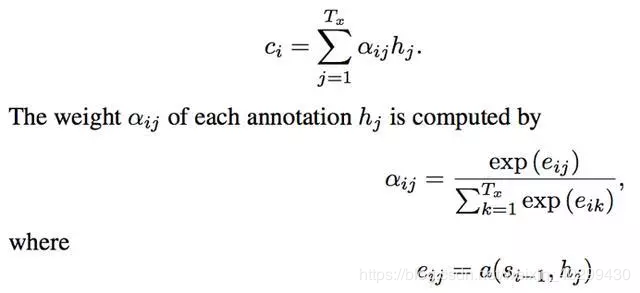
目前大多数的注意力模型都是依附在 Encoder-Decoder 框架下，但并不是只能运用在该模型中，attention机制作为一种思想可以和多种模型进行结合，其本身不依赖于任何一种框架。Encoder-Decoder 框架是深度学习中非常常见的一个模型框架，例如在 Image Caption 的应用中 Encoder-Decoder 就是 CNN-RNN 的编码 - 解码框架；在神经网络机器翻译中 Encoder-Decoder 往往就是 LSTM-LSTM 的编码 - 解码框架，在机器翻译中也被叫做 Sequence to Sequence learning 。

所谓编码，就是将输入的序列编码成一个固定长度的向量；解码，就是将之前生成的固定向量再解码成输出序列。这里的输入序列和输出序列正是机器翻译的结果和输出。

为了说明 Attention 机制的作用，以 Encoder-Decoder 框架下的机器翻译的应用为例，该框架的抽象表示如下图：



Attention Model其中下半部分为Encoder结构，这里采用双向RNN构成，前向RNN顺序输入单词，后向RNN反序输入单词。将同一时刻的两个RNN单元的隐状态做拼接形成最终的隐状态输出ht，这样ht既包含当前单词前一个时刻的信息，也包含后一个时刻的信息。上半部分为Decoder结构，为一个单向的RNN。中间部分就是Attention，采用如下公式计算：

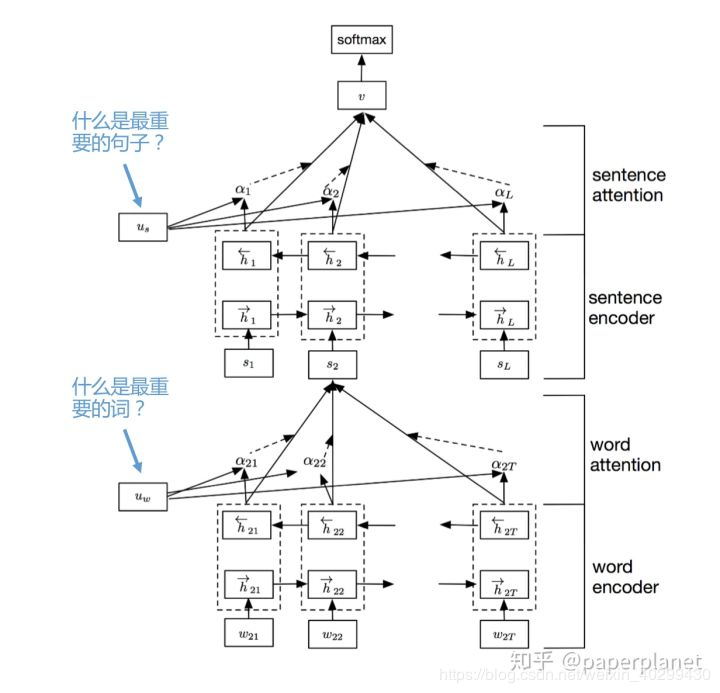


其中，si-1为Decoder上一个时刻的隐状态，hj为j时刻Encoder隐藏层输出状态。使用一个网络结构a训练，得到的分值eij表示j时刻输入与i时刻输出之间的匹配程度。之后用一个softmax函数归一化，得到的标准概率表示alpha ij即为hj在翻译yi中的重要性表示。最后以对应的alpha作为权值，加权计算每一个时刻的输入语义向量ci，即体现了每一个输入单词，在翻译不同输出单词中的重要性。

HAN的原理（Hierarchical Attention Networks）

参考链接 https://zhuanlan.zhihu.com/p/44776747

在深度学习文本分类模型中，HAN（Hierarchical Attention Network）是一个非常有意思也值得深入研究的模型，不仅解决了TextCNN丢失文本结构信息的问题，在长文本上有不错的分类精度，更为重要的事，在现代模型中，他的可解释性非常强。



以上图的第二句为例，输入词向量序列 [公式] ，通过词级别的Bi-GRU后，每个词都会有一个对应的Bi-GRU输出的隐向量 [公式] ，再通过 [公式] 向量与每个[公式]向量点积得到attention权重，然后把[公式]序列做一个根据attention权重的加权和，得到句子summary向量 [公式] ，每个句子在通过同样的Bi-GRU结构再加attention得到最终输出的文档特征向量v向量，然后根据v向量通过后级dense层再加分类器得到最终的文本分类结果。模型结构非常符合人的从词->句子->再到篇章的理解过程。

利用Attention模型进行文本分类

参考文献 https://www.cnblogs.com/jiangxinyang/p/10208227.html

Bi-LSTM + Attention 就是在Bi-LSTM的模型上加入Attention层，在Bi-LSTM中我们会用最后一个时序的输出向量 作为特征向量，然后进行softmax分类。Attention是先计算每个时序的权重，然后将所有时序 的向量进行加权和作为特征向量，然后进行softmax分类。在实验中，加上Attention确实对结果有所提升。其模型结构如下图：

