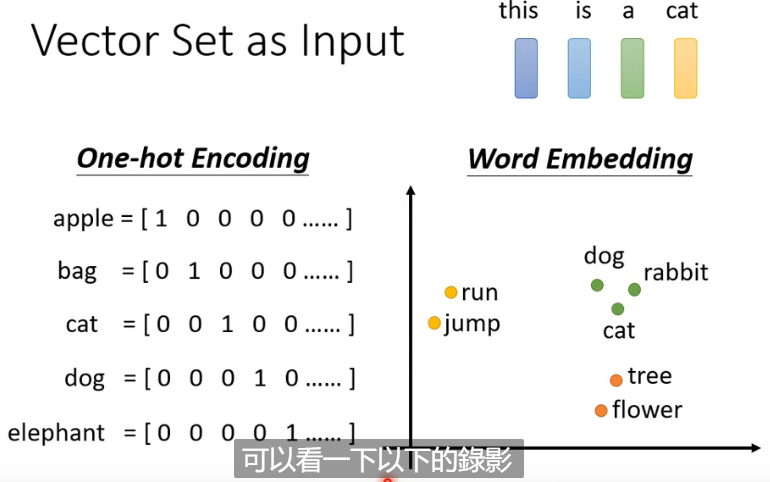
将词表示成向量

One-hot Encoding

通过一个很长的向量其中每一个词对应一个位置的值设为1，使得每一个维度对应一个词。但这会使得每一个词之间都看不出它们之间的关系。

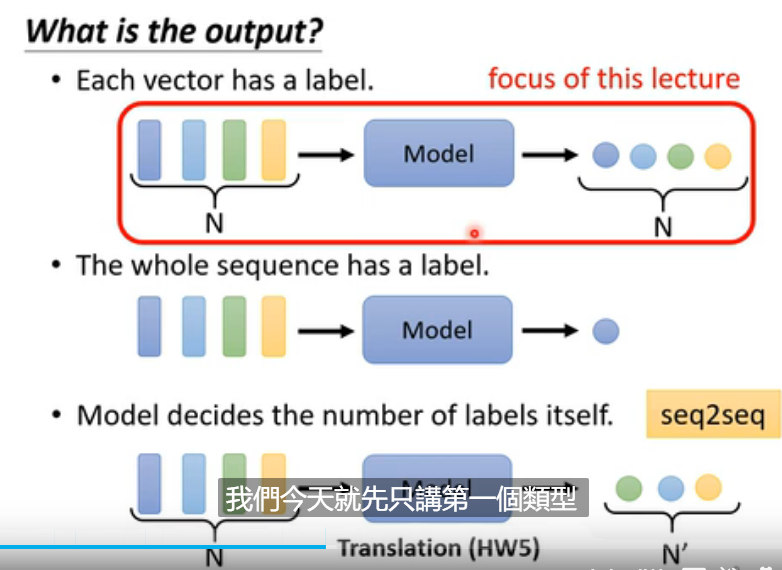
Word Embedding

将每个词对应着一个向量。



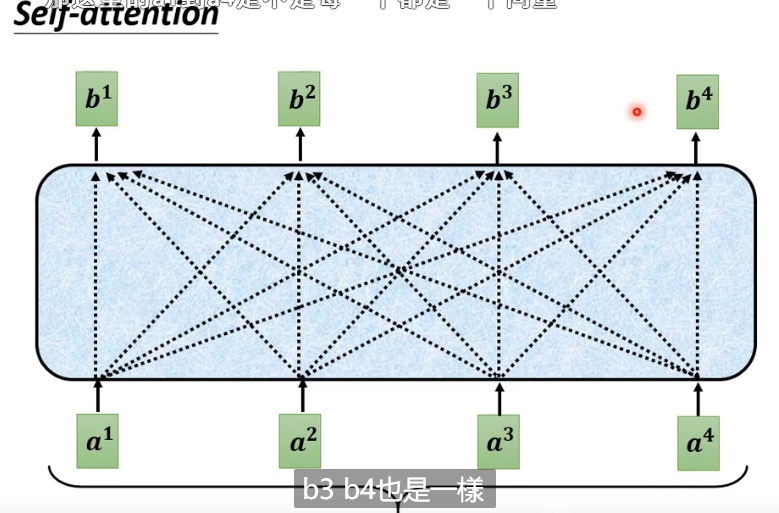
模型的输入为一堆向量，这些向量可以是语音，图，句子

输出就有三种可能性，每一个向量都对应一个标签、整个输入对应着一个输出结果、多个输出对应着多个输出。

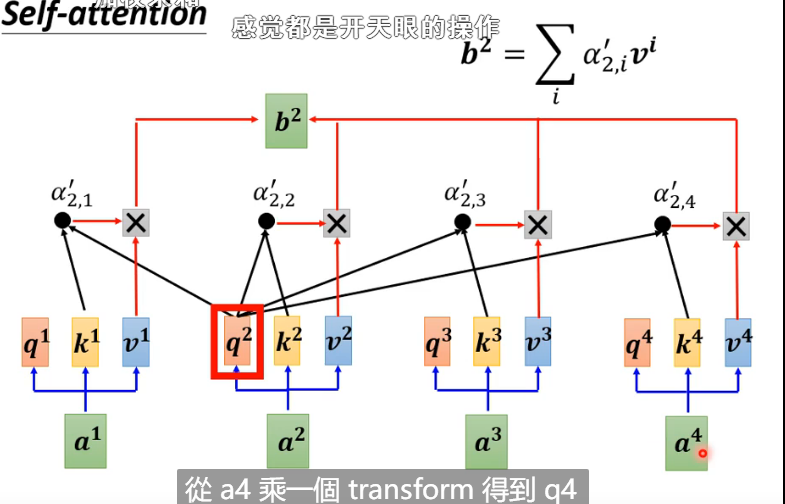


Self-attention

自注意力机制的输入可能是未处理过的input也可能是某个隐藏层的输出，都是一排的向量，通过考虑计算这一排所有的向量之后得到对应的输出。

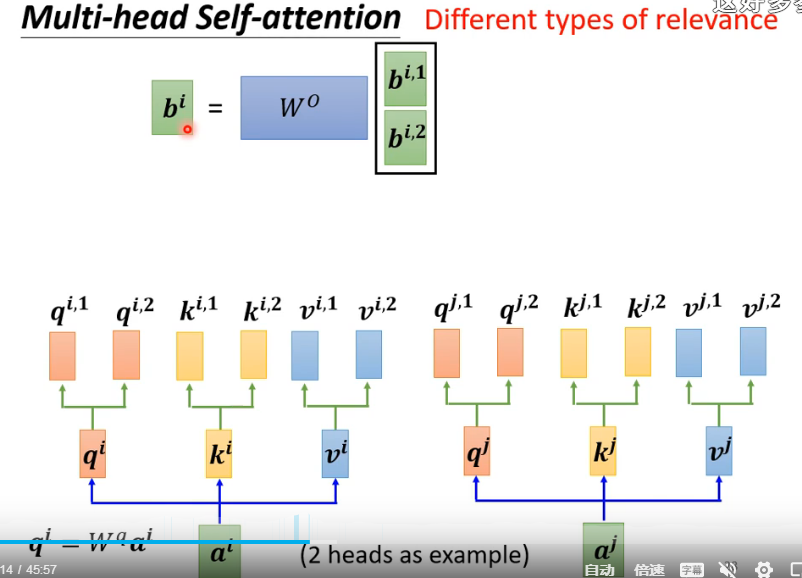


通过计算每个向量相乘不同的矩阵w得到q和k，然后将q和k相乘得到两个向量的相关性。当一个向量和其他向量包括自己算完相关性后，通过一层softmax将这些值归一化，通过计算w和向量相乘，得到v，然后再和归一化后的相关性的值相乘，然后全部加强了就得到了输出结果



Multi-head Self-attention

通过w得到q，k，v后再乘多个矩阵得到多个不同的q，k，v，并且q，k，v的数量要相同并且对应，然后同一类的q，k，v进行注意力机制的运算。最后将所有结果进行拼接再乘上一个矩阵就得到最后的结果。

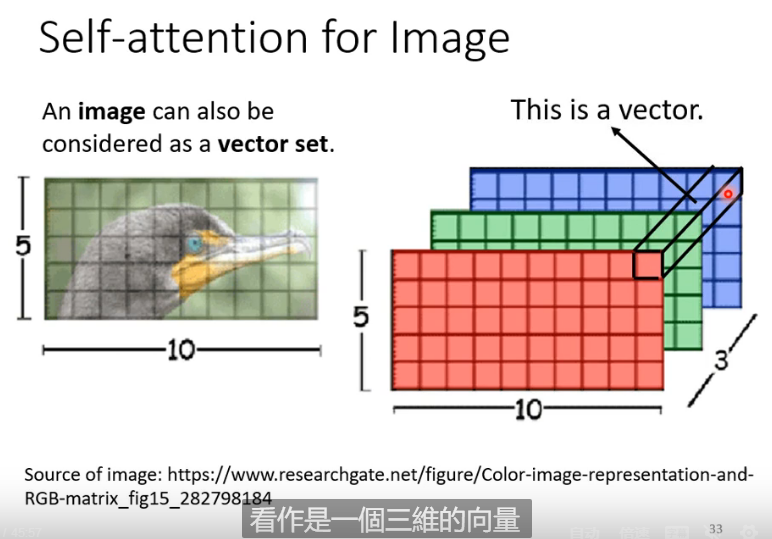


Positional Encoding

通过位置编码解决在注意力机制计算中无法区别每一个词之间的距离远近无法区别的问题。

Self-attention for image

将图片看成一个多维的向量，然后就可以通过self-attention来处理图片。



Rnn

在通过神经网络的输入多了一个state，这个state是保存的上一个向量输入的内容，每一次的输出都会有一个state。因此Rnn是都有考虑了上一个的输出的情况，而self-attention则是考虑的整个句子的输入。

