## Attention的通用计算方式

1. Attention一般都使用Q、K、V三个变量进行计算
2. 最终结果

## Soft Attentin

1. 三个全连接层

## Transformer

1. 三个全连接层

E是词向量，U是位置向量

## Transformer-XL

1. 三个全连接层

