# Why self-attention?

①每一层的计算复杂度

假设输入序列长度为n，序列维度为d，k为卷积操作卷积核的尺寸。则对于self-attention来说，