现在断输入是一组向量(a set of vectors): [] [] [] [] — [Model] — Scalors or Classes 13-15: [. 同于,将同个单词表示为veolin,一个同于就是"组vector (这个 set as 长度可能气发生变化)

3.图·自个节点是一个 vector (利用profile 来刻画)

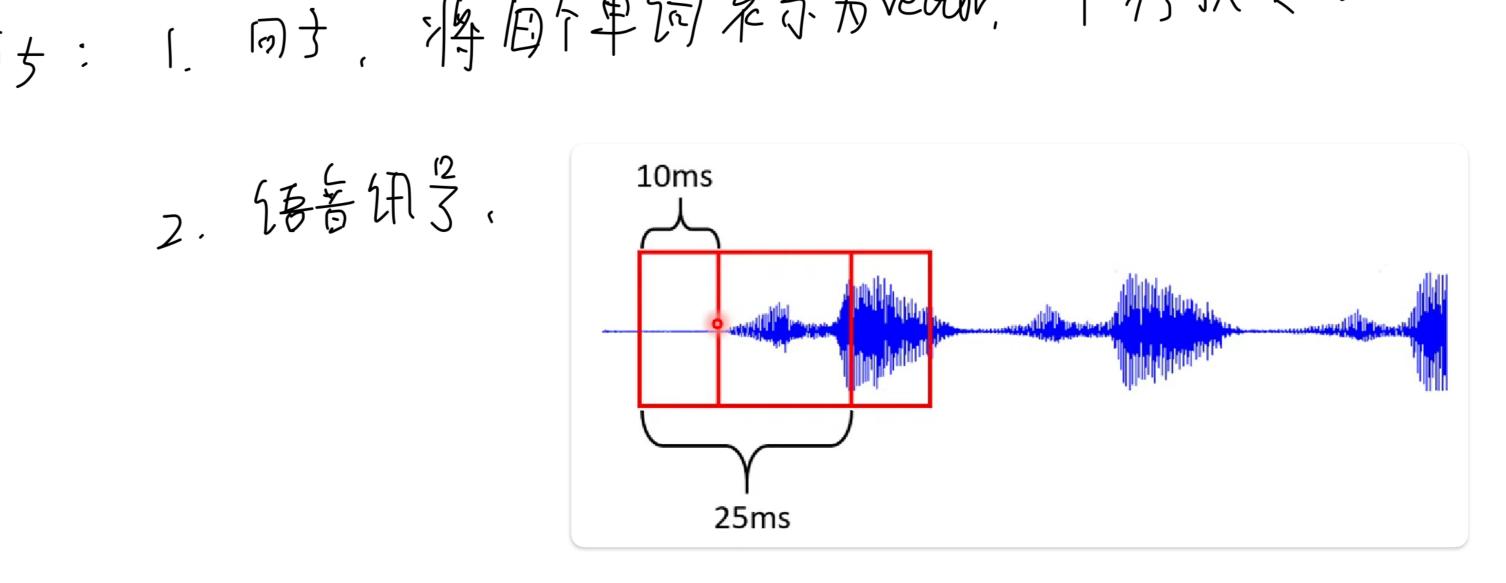


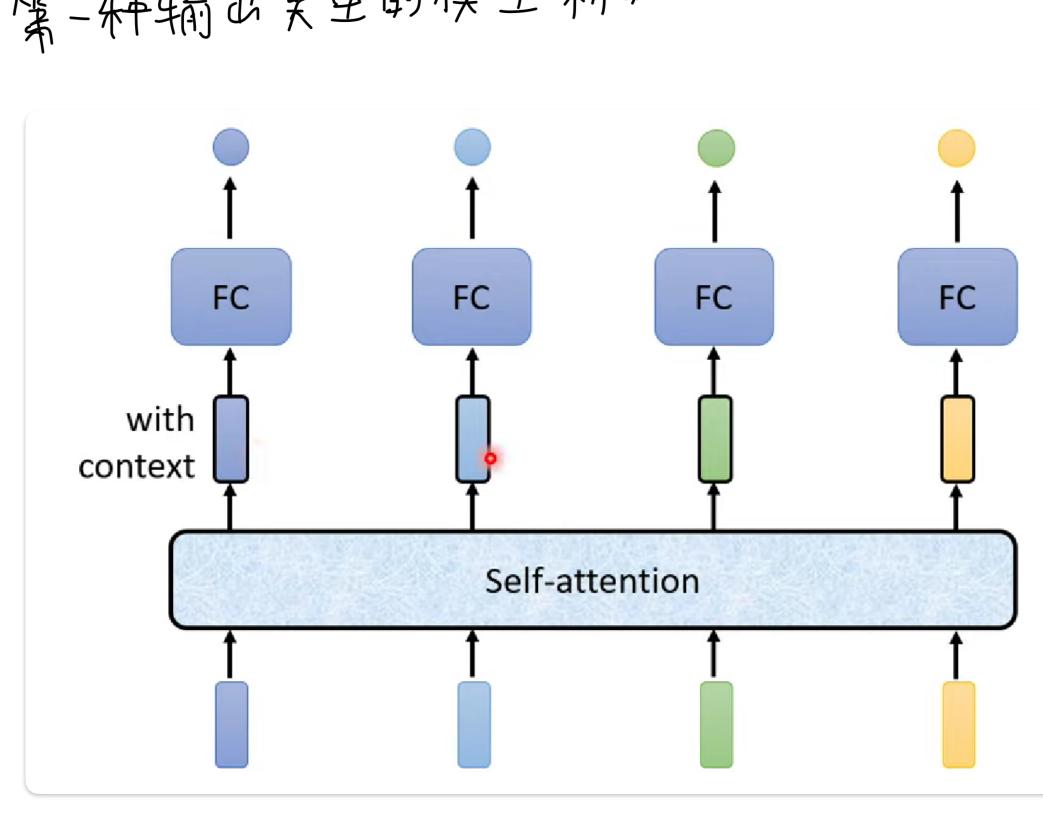
图 25 ms 形成-frame to即一个 vector 窗的移动10ms开作成另一个frame

输出: 1. Each vector has a lable, 例为: 饲性标注

2. The whole sequence has a lable . 例子: sentiment analysis Lith正版面)

3. Model decides the number of labels itelf, 1岁子: seq2 seq, 奮羽锋

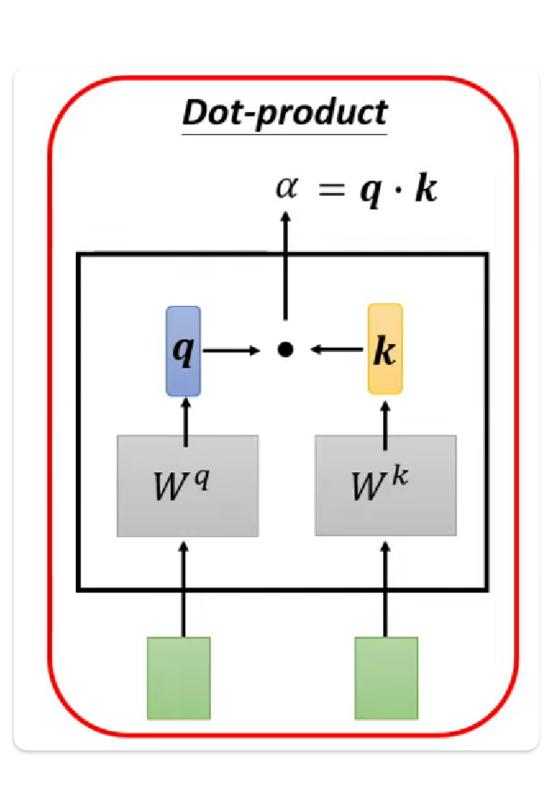
第一种输出类型的模型架构



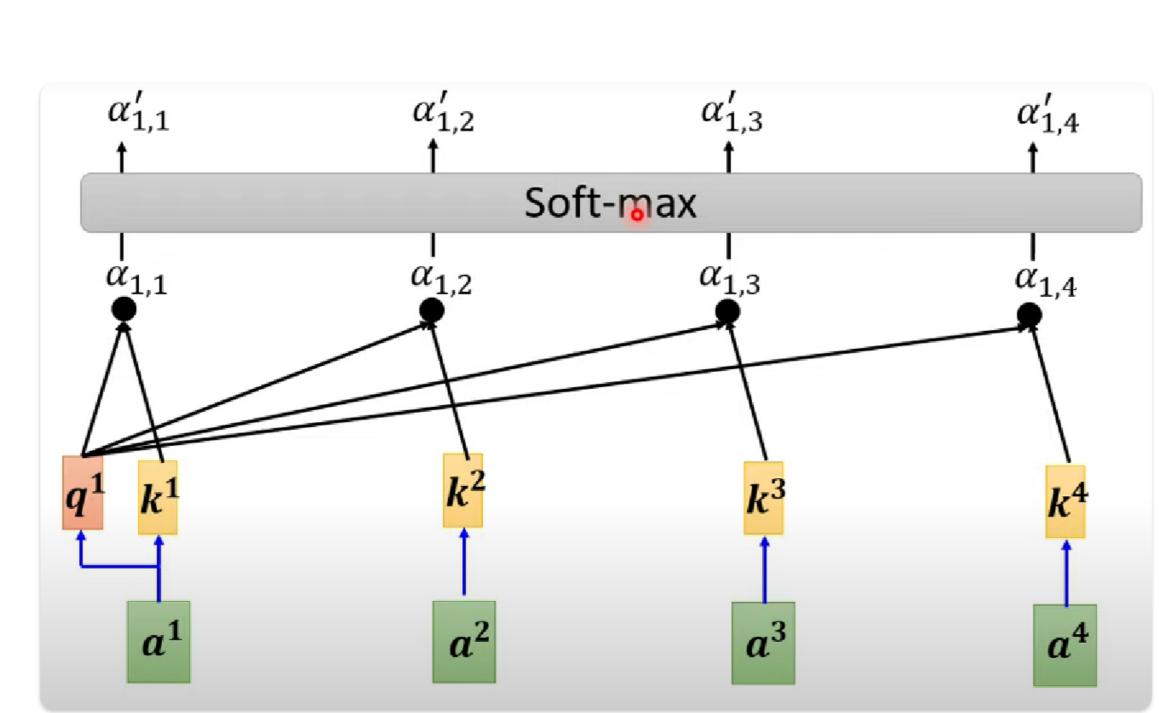
self-attention寿产整个上不多的资讯,

M输入. self-ortention 就有M输出【糖过上下负的】 然后每个下口再单独处理

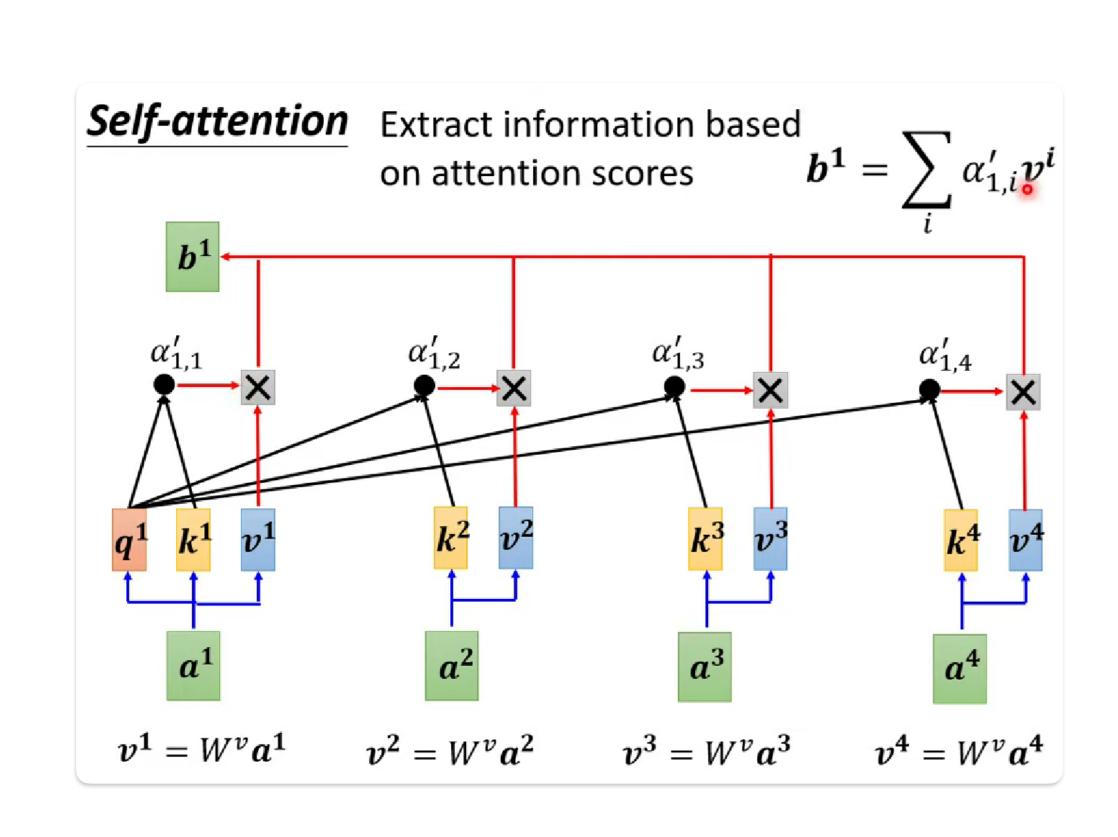
更深层没的架构可以将 self-attention 和FC交替叠加使用



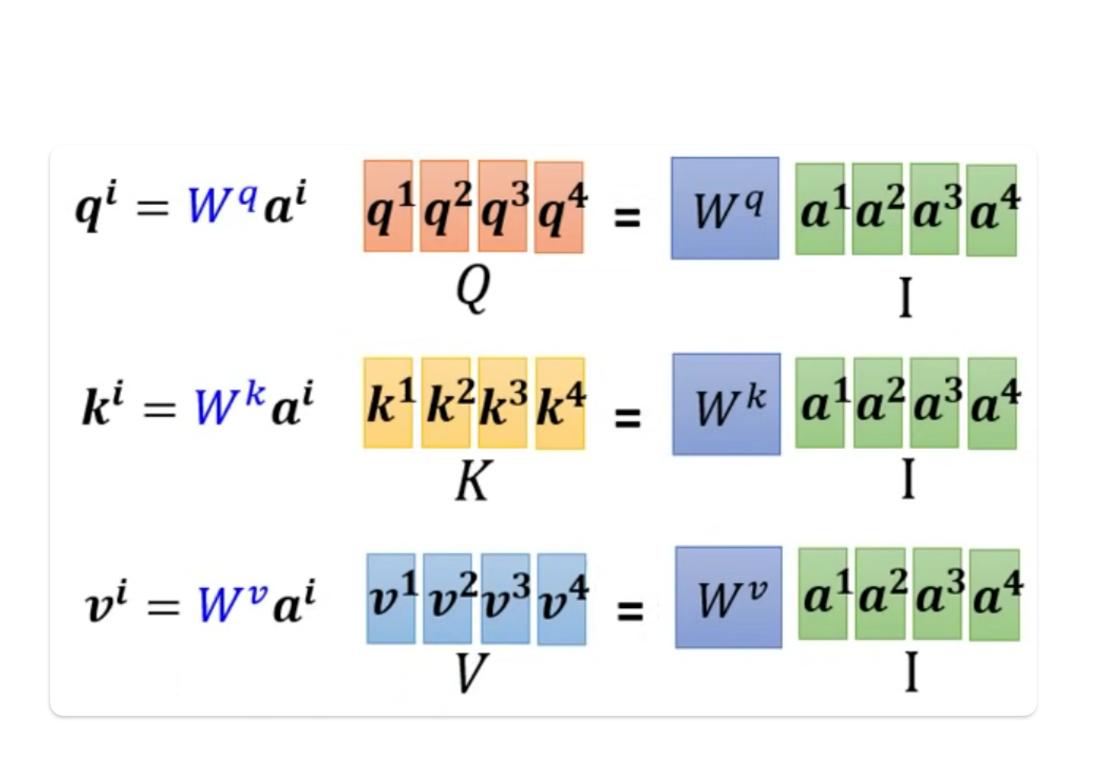
利用Dot-product来价算两个Vector的相似度. d.



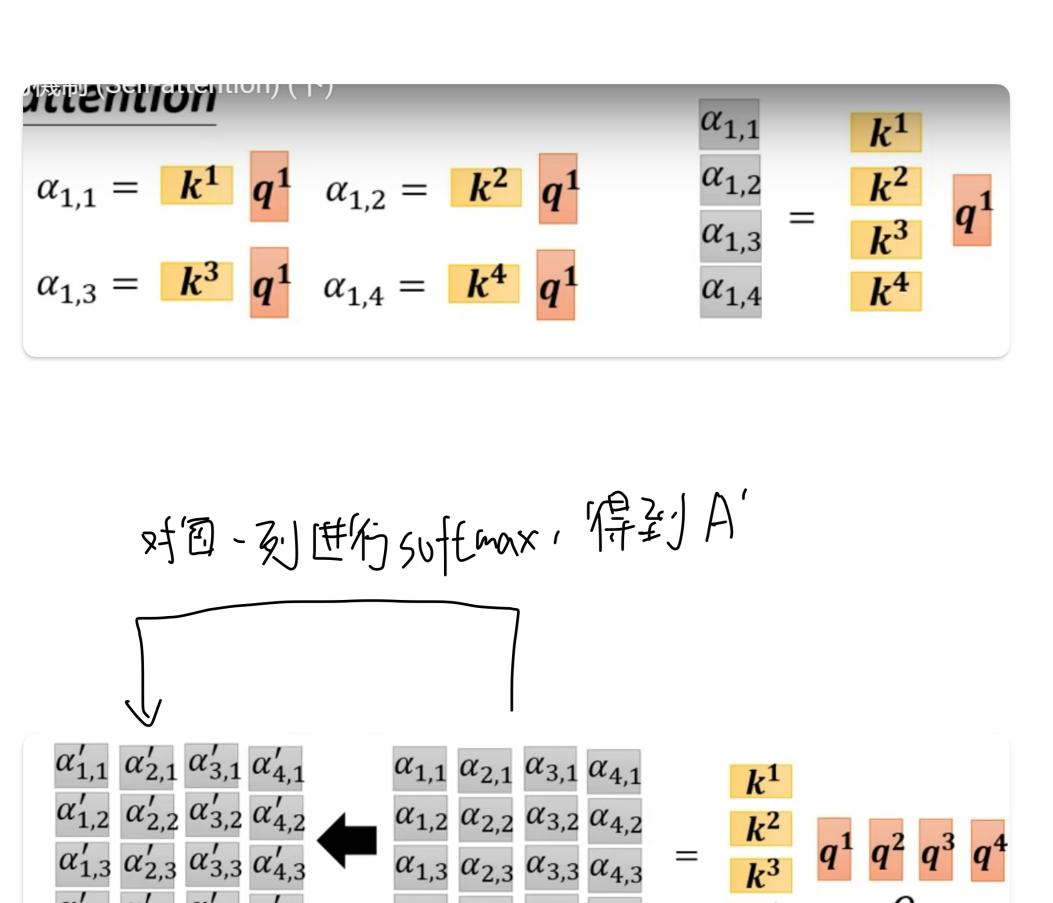
利用 quey, 和 key 2, 7, 4 好算和似度 d,, k.(b=1, 2,3) 面常还气用 Cluery 1 乐 key 计算确与自身的相似复 01/1 最后经过Softmax得到最终的%数(这里更有点像相似的概率)



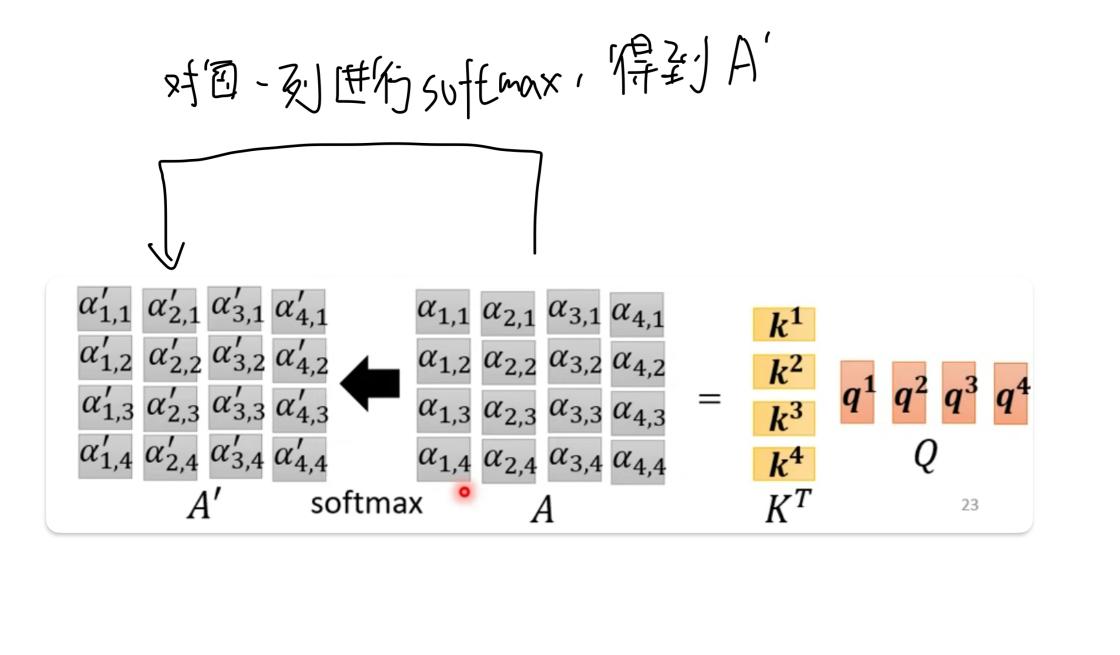
最后将自个Vector和新上来的年WV得到V. 再对每个V世份权重和,最终得到与,便即 01 对应的输出)



并行行算,得见下,V. 其中的每个列向量是Qi对应ESqi, Ki和Vi



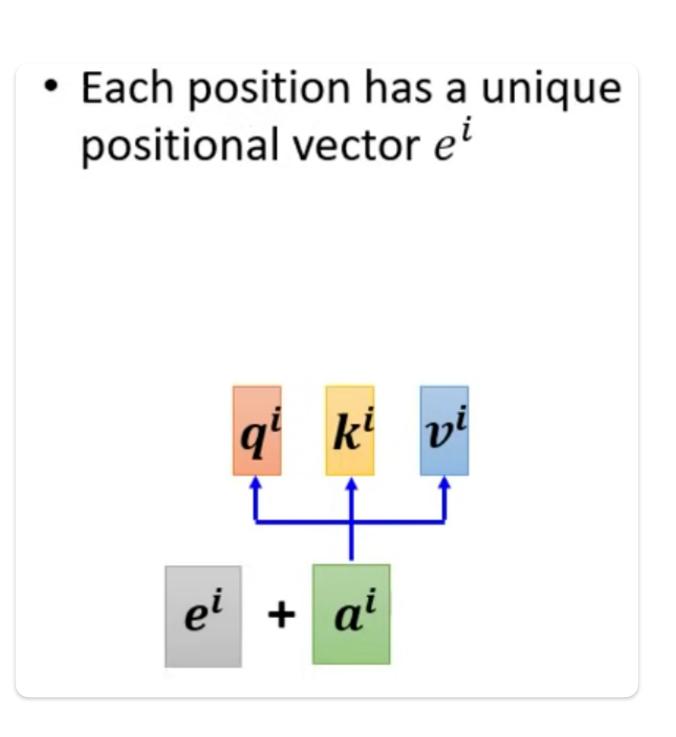
竹草相似友的数文同样是再行竹草



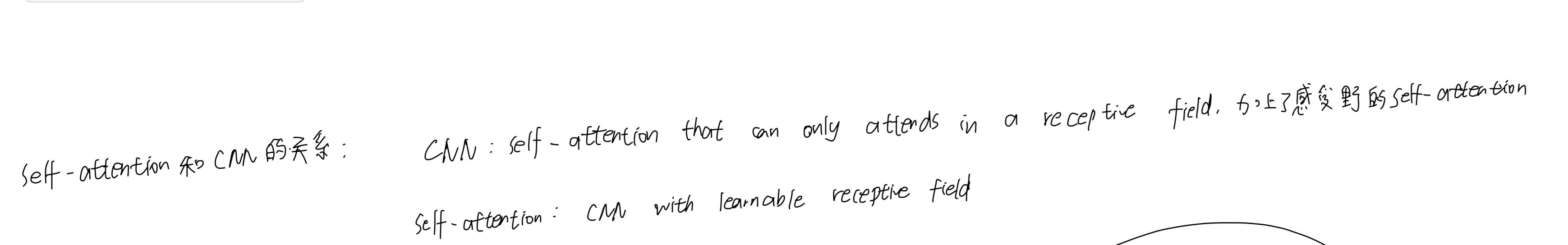
self-attention的复体: 2 heads self-attention (甚至更多head) 即将中,皮肤如其形积上上东医降,得到中",中",长"、长"之外。V"、V"上

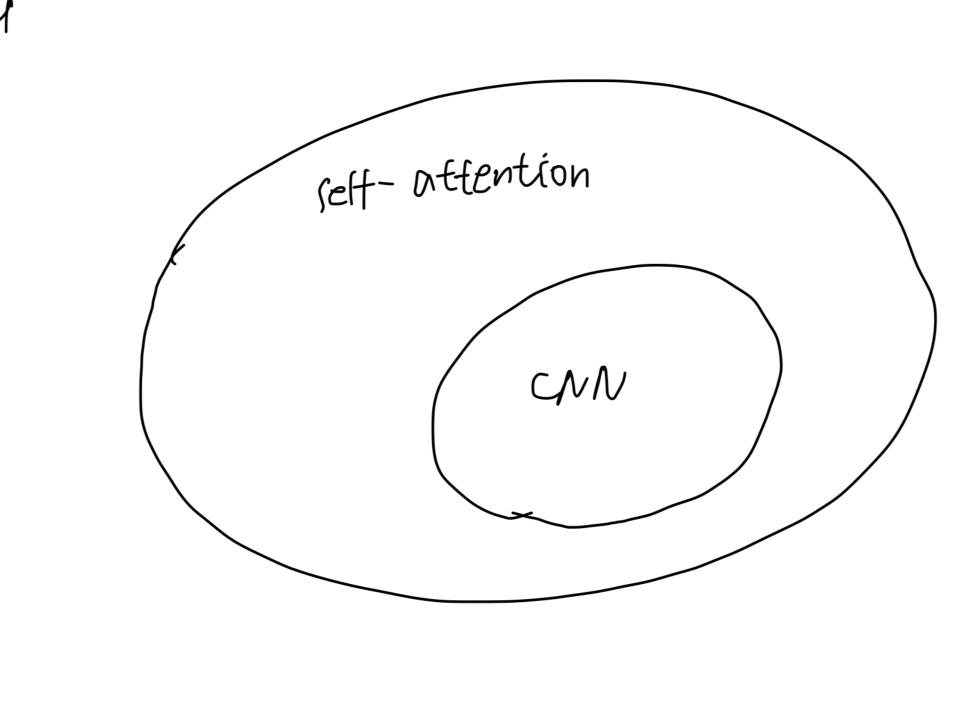
上述的 Self-Ottention 法人少了他爱的信息,

但有时候任置的信息比较重要, 比如:饲性标注任务中,第一个明初的的概率维小,这种信息对问练会很有帮助



给风力上巴,那位置信息





公关系: 1. RNN 最后的输出准 难考例最早的输入。断需要将其一直存储在 menory f Self-attention 能更加容易考虑。任意作置的输入 Self-onftonforRM的关系:

2. FMR nonparallel BS. a self-attention & parallel BS