Date.

& & chap-9 125 to truto	5月堂江北	, , , , , , , , , , , , , , , , , , , ,	0 3 6 18
接到这条度	一点生资色	विध्वत विदेश्य	· · · · · · · · · · · · · · · · · · ·
•		2135.1270	- 教養 - 3
网络京曼	2 265 65		
	P. Carperton 1254	4	87.1 HE pho
进力: 面出自上面	下的信息出择机制来区	过滤技术全元美色	٤
			: एट में हो
外部记忆: 依依在时间的	经的记忆生物本 超高度	见发生.	- 18°
	神性团交机	这种的种子人	X.v.
型注意力和到		一九十七十萬五十二十十十十十十十十十十十十十十十十十十十十十十十十十十十十十十十十十	9 -
[854310 11.0]	端飞路 化机网络.	1 1 2 5 1 1	8
	2027212121212	是月季行病力	do
自注意力:可以作为不	中层网层中的一层末使用	一方纹地建筑北	定会依赖问题。
记礼等元	(18.	Ex) 2) months =	
0.5	接多量等量过接空性	得提作.	
七十二月期	神遥网络。	/80	
层部	北芝(神经及治生)	V 范克和斯克利	(5,302
中共的	到到记忆		
长基A	9年7岁数	8.332 F 313	
Mari	S. R. san		
板站方	内部定寻处为主	1.000	al ra
108 Wo n 4	DE KIME ED .	\$ 12 S.A. 8	2 (* (*)

8.1 注意力

联集寸:自己而下看的公主动 (主要)

是考隆: 舒和上天意识、被动

8.7.1 注流为机制

面下却是是卷1光 [i] th

Xiii 八午輪刀傷魚 茎的向草

2 注意力变量,包示神选择信息的等引位置.

di 这样争行新力信息的概率, 这意为各市

di = p(8=21 XIIN, q)

= softmax (5 (xi, g))

S(xi,至) 产意为打节之英

trul生程型 S(xx,)=vitanh(wxi+vg)

点歌模型 S(花夏) = 花着

能放 怎般核型 SCX. (推荐) (推荐)

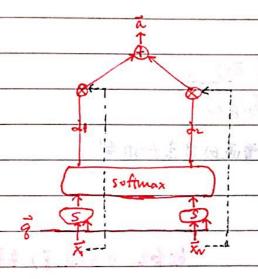
双线维模型 SCE_系)= X: W家

注意力多年上、可解科为上下文查询分时,多自	注意力	当年	Li	り解	科为	とでえ	查詢	द धार	. 浄	i
-----------------------	-----	----	----	----	----	-----	----	-------	-----	---

软性生态力 小别

at (x, 8) = 2 2; xi

= Expresize, ZX7



X1 29 => S(x, 7) DP 3 = 5

S(x, 8) (2tt softnox =) di

Ji 和 X ź É き ⇒ att (X, 至)

3.1.3 注意力机制的变体

ヨる更性注意方

&B1/:

吕美注到某一位置之的任息,却不是注意的各本下的共同望

1.能取最高水配等的输力信息

att $(X \vec{g}) = \vec{X}$; $j = \arg\max_{i=1}^{N} \alpha_i$

2. 适色力多车至6000金机条埠

Date.

3 键值对话意为

な: け真注意からる しょ

值: 计异聚全线泵

(k,v) = [(k,v), ..., (k, v)]

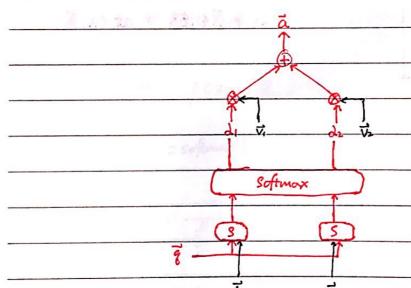
N午輪刀信息

对战、th: 同意 1713 4 新新

att ((K,N), g) = \(\overline{2}\) dive

= $\frac{\sqrt{2}}{2} \exp(s(\vec{k}_1, \vec{k}_2)) = \vec{v}_1$

专人=10时, 健估对色是为案价于普通的区意力机制



記事を 3 5 (h, を) 3数

我+3月至日本山斯

Schi, & ZEI Softmax > di

di 私びと信气

· 三十年女一位首立的信息

=>多关注意/

利用多个支的 Q = C克,…克加了来平约地计年从输力信息中处现的个信息。 每个注意力关注不同部分。

at (ch. 10), (e) = at (ch, v), (e), (om Gat (ch, v), (m)

		No.
		Date
与 造构化准备	, Rasa	0
	4	
是次化的往至为		N 18 11 1 2
	and the second	
	1 639	C 3 11 64
	A	
The state of the s		