28/1/24 - Friday Transposiner Dew der Architecture Self attention and wors output probabilities attention - presug Softman Devodo Encolor Trans former linear Encoder - emodes input Devoder - duode the envodel Add a Norm K Feed forward (Devoder) 1 Eniodes Add & Norm (December) 2 2 (Eurodu Devoler ? multi head 3 (Ehwoler) Devoder 4 attention 4 (Emoder) Decode 5 (Emoder) from (necoda) 6 emode 6 Eneoch Add a Norm + Enwoodn and Dewoden are Wall up of 6 blocks lack marked multi head attents. Feed found NN Enerden Suf Attenton 2) Positional Enoding in can anodu, its made up offer 2 blocky Output Embedding

I tak emoder block has some blocks invide ther but parameters change. output Add & Narm Feed Forward NN Red person Suf attention Add 4 Norm Multi Head attent Encoder So flur av 6 fuer blocks the output been this Encoder will be given to Feed forward NN duoder. Cross Attention Devoder Marked Self Attention

feed found un Masked Mulh Multi Head Attents. like this there are 6 dusde black in Add & Norm of Marked Multiflead the overall decoder ta) what happene in a bryle duoder? (A) As all the devoter tilveter are some to it you understand one decode block , aust all are Some. output blocs 1 dewda blow Steps: Input block Machine Translahm English to Hindi English = we are friends Hind; = Hum dost Hair

English Sentance goer into Encoder generater contintual Embedalig for duoder to Oinput block there are a speakon in this; 1- Shiffing 2. Tokunjahan 7. Enbedding I the work of input block is to devotors first block.

Sentance (hindi sentance) and precure it first block.

My form to lend it to devotors first block. adde a foken start hefore proven Hum dost How Oright Shift hand Jokans Pentances 18tat 7 Hum doct Hai (O) Tokenizalam my of eq. 1 gram

(Hai) into tokens [dos-] to numbers/ siz din 3 Embeddings 512 din here we don't know when word is before like the position of the world so I fackle that The we see this @ - positional Encodings. 512 dm 12

Now we are done with input block. the final 21, 12, 21, 94 vector are the wints for devolu 1 Devoder Ploch. 1 jet duoden belock has 3 tasks in it [Masked Multi Head Attention)

2. Multi hear Attention (Cross-attention)

marked multihear

ottente 1) 3. Feed Farward Network OMasked Multi Head Attention here markedy is implimental for every imput vertor, corresponds I that in contentual vector il generated, 7, 72, 7, 71 2. 2 Churt 7

to enal this marked contentual vectors we
only consider 2, verber, 1.e., (Stort) and we do not compair rust. 2 2 2,
Is not consider next. 2 2
and we do not start >
The same of the same of
To general the mod hun
\sim χ η_2
7 2 × 1, 1/2
Po generate 2, -> we only consider (8tout) [18tout)
To gentran 27 Hun (doct)
(1 × 1 × 1 × 1 × 1 × 1 × 1 × 1 × 1 × 1
2 % 4, 12 19
23'
only worder per
To gownt tu - we only worder previous one word
Mai)
To garden and work Mai
~ 1. 1/2 17
Zy ~
7
20 (1111)
21
the output of
This are it he
Li Hear
madera Mult
14

Now perfor Add a Norm du duodes Add = toutputs of marked mulhi head atter N, 72 73 24 21 22 37 24 Now poten Normalization = layer Norm Layer Normalization for each veutr, the will calculate a, in and with that help it normalizer, that is able from to make training pursuits stable Now were attention come into field, here, the connection between thinds and English Sentance will happen. Add a Norm x Multi head (conoci) attento Zirim, Ernom, Zinom, Zinon t input from poort 1 of decoder

Cross attention 1 2 seg [. 13 Lo 2 seg [Hindi] fuor previous step of devoder I this is done as we need . I vector ley, value and Quers Erwele 3 Army bey one Valer is your from Durde - Query verber is gran and Cross attention | Mulli-head attention tcr tc1 there are the output from cross attention where the contentual vector information is attached

for each token of the output verlow (Z, Zz, Zz, tu) then will be a contented enhedding vertor 263 712 そら Perenous et es output from marked multi had atti (ZI Norm, 22 Nom, ZINON Ay Nom) Zynim 21 Norm ZNorm 74 24 762 200 7C. Lager Normalizeha Zu Nich ZerNorn Zegnira Done with I'm block of durden

Add a Norm feed foward NN input from person blow 7CINOM, ZCINOM, BUNOM, ZCYNOM 2Cu Non Zanom 3CL Norm 20, Norm Feed for and 512 Same and testral B 2048 his 20484512 512 num BL SIZ Hay output 3

(W1.2 + b1) ~ Ux 20UF] d'menerun 5/x × 2041 Relu (w, t +hi) Shage = 4 x 20 48 in nent layer We [Reln (W1.2+h1)] + h2 · 20/18 . 512 = 415/2) dimenus (ux wys linear [Wz [Reh (w1.2 +h1)] +hz] U x 512 + Add FNorm) + ZC3Nim ZG Norm

Normalizah. yendos Y2Nom Y2Nom YINOrm ther y vertons are the first output of 512 diment duoder 1 3 Output block Then y porm, y morn --- yunon lagars are but I more dewder totals and Same operatione will the executed. with slight change in the parameter the output from 6 th dewoler layer the output are called as yt, Norm, Yto Norm - Yfywon final outerst

Outgat block Softman made of 2 layer, linear and softman Can be compared to feel fromward No Can he compared output layer ytun .-Yty Jorn ytzNom yf, Norn L Hou's (doy.) (Han) 1Stort > 1 × 512 512 110000 (normely J Soffman 1.5 0.1 6.1] Now Lo weresponding to when (Start >) 4 f. Norm, which word has high pueds ability from the neural retwo to the neural network now send ytz Norm onel see when word

how high perchability, that is the hind. Homelation of that english words.