

When they ask me how deep learning is going



Her: I'm a deep learning practitioner
Me: [Thinking of something to impress her] I don't know math either



Agenda

1. Classical Approaches
2. Sequence Labelling Approaches
3. CRF
4. BI-LSTM + CRF
5. Stanza (Post-Read)

Before Transformation

SOTA

load entire data \rightarrow RAM



Batch-size \downarrow

Reduce-batch size
more len
Bottle neck
Training will happen

GPU-RAM

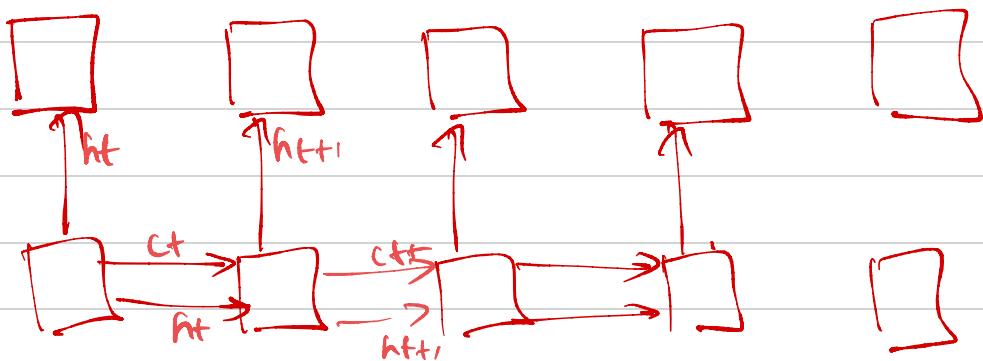
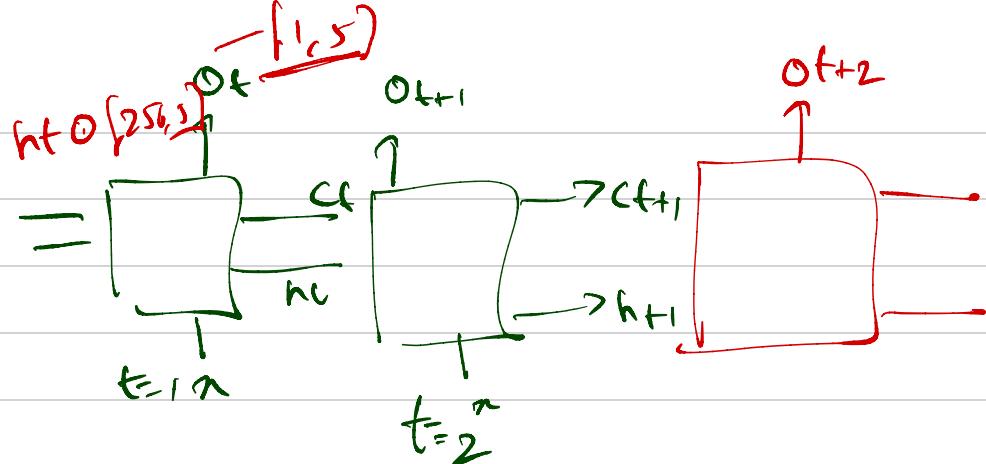
GPT-4 \rightarrow

25000 \rightarrow A 100

80 GB VRAM

2 months

100 million dollar +



NER -

Given a sentence, I want you to find out entities among these

1. Person
2. Location - role - from/to
3. time

Type	Tag	Sample Categories	Example sentences
People	PER	people, characters	Turing is a giant of computer science.
Organization	ORG	companies, sports teams	The IPCC warned about the cyclone.
Location	LOC	regions, mountains, seas	The Mt. Sanitas loop is in Sunshine Canyon.
Geo-Political	GPE	countries, states, provinces	Palo Alto is raising the fees for parking.
Entity			
Facility	FAC	bridges, buildings, airports	Consider the Golden Gate Bridge.
Vehicles	VEH	planes, trains, automobiles	It was a classic Ford Falcon.

J am Shinon Brosod

B. Person P. Person

Beginnin Inside

→ Based on what we've learned so far
what can we use??.

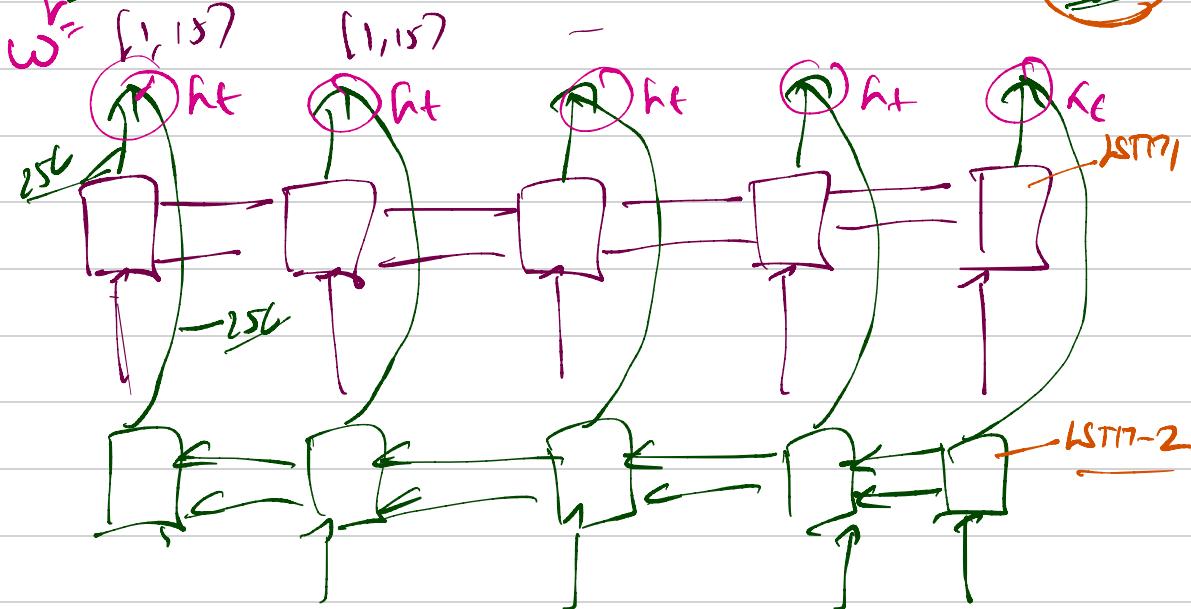
I love visiting bank near ganga river.

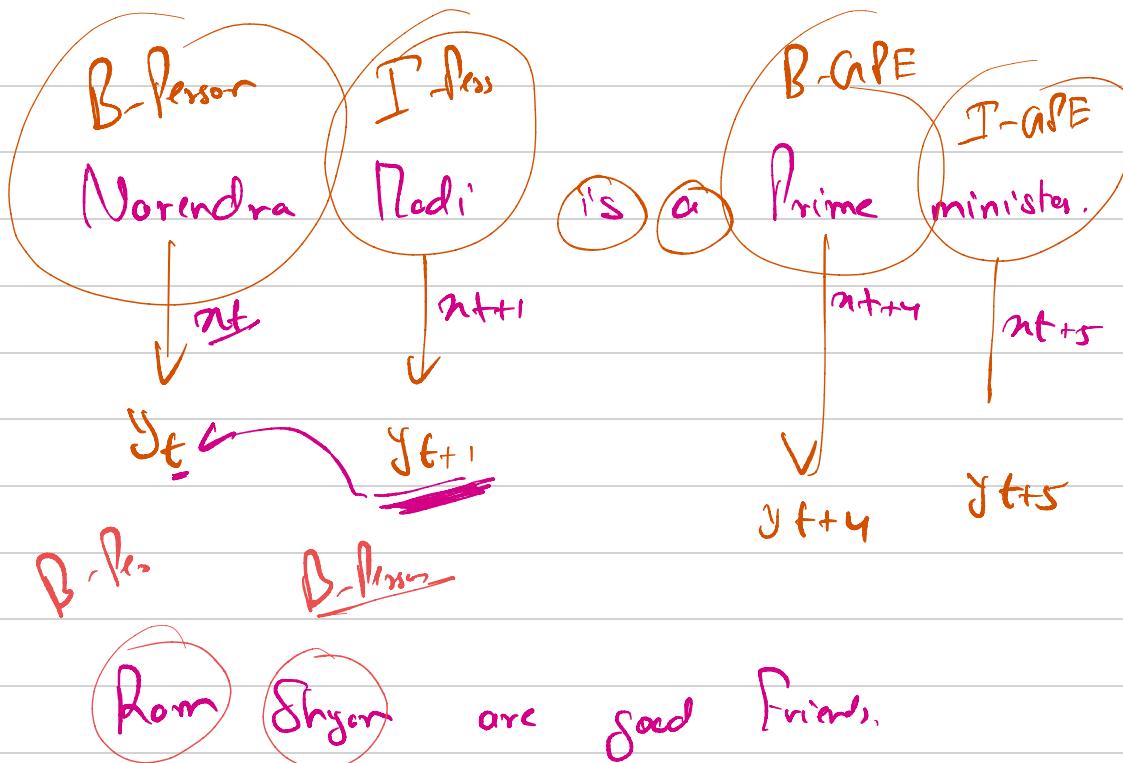


Feature extraction \rightarrow inf is short, Posterior factor
decision making step \rightarrow " " not short

256, 256

(512, 15)





CONDITIONAL RANDOM FIELDS - CRF

$$\text{Softmax} = \left[\begin{matrix} ? & 8, & 9, & 6 \end{matrix} \right]$$

$$\left[\frac{e^7}{e^7 + e^8 + e^9 + e^6}, \frac{e^8}{e^7 + e^8 + e^9 + e^6}, \frac{e^9}{e^7 + e^8 + e^9 + e^6}, \frac{e^6}{e^7 + e^8 + e^9 + e^6} \right]$$

CRF \rightarrow "g love meditation"
[O O] $B: 1 \text{ hobby} \rightarrow$ Score

Generate all probable entities seq, calculate score
for each seq.

① O O O \rightarrow Score₁

② O B: 1 hobby O \rightarrow Score₂

③ B: 1 hobby O O \rightarrow Score₃

Sample Feature functions

$f(y^{-1}, y, m, i)$

Prev. label
Current label
all the
wo.
index

S - diff - enti

$\frac{LR}{DT}$

$O \quad O$

O

B:GPE I:GPE

Sep. B:Person I:Person

Narendra modi

B:GPE I:GPE

prime minister

$f(O_{nt} | O, O)$

$O-1 \quad O \quad 'O \quad O$

DT $f(O, nt, O, O)$ Score

$O-1$

Consider the below formula:

000, D.M.I.a determining train

$$CRF: p_{\theta}(y|x) = \frac{\exp(\sum_j w_j F_j(x, y))}{\sum_{y'} \exp(\sum_j w_j F_j(x, y'))}$$

, where $F_j(x, y) = \sum_{i=1}^L f_j(y_{i-1}, y_i, x, i)$

Here,

$p_{\theta}(y|x)$ refers to the probability of calculating a Label sequence(y) given a word sequence(x).

→ If Score → Calculated Score For the entire
GF correct +
Score → high

Score → low

○ ○ ○ ○ B: Pers low

○ ○ ○ B: Pe I: Pe

○ ○ ○ ○ ○

B: Perso I: Perso

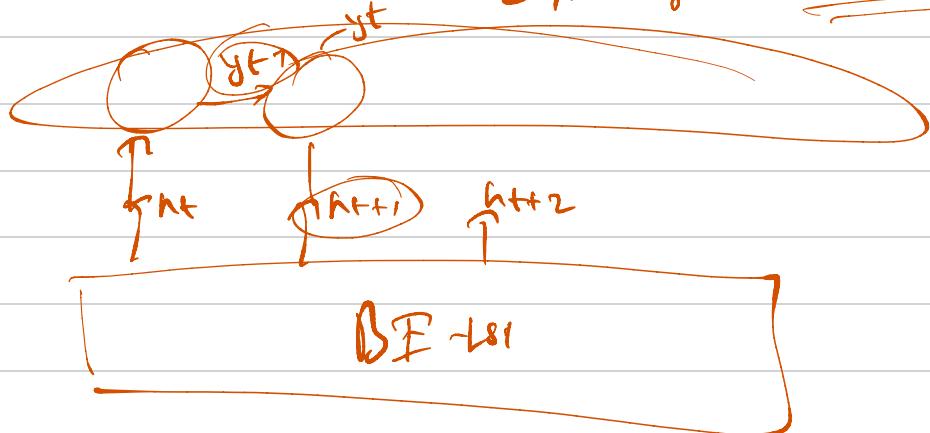
Narendra modi

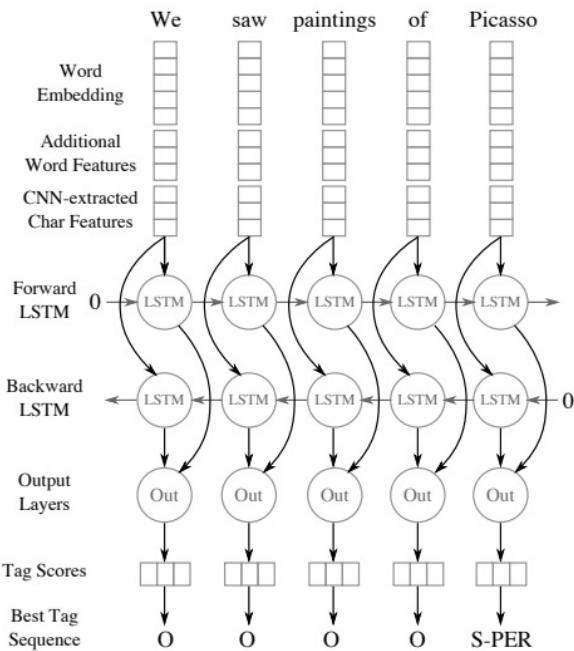
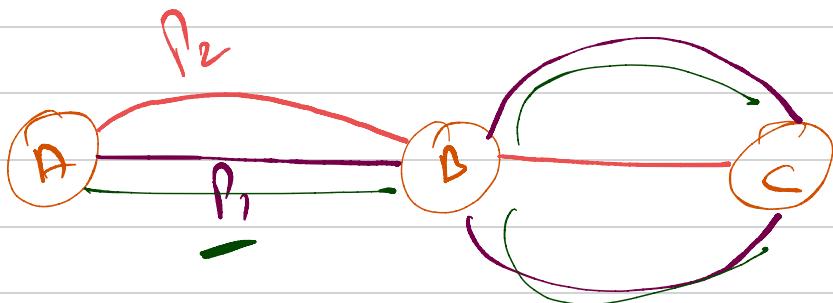
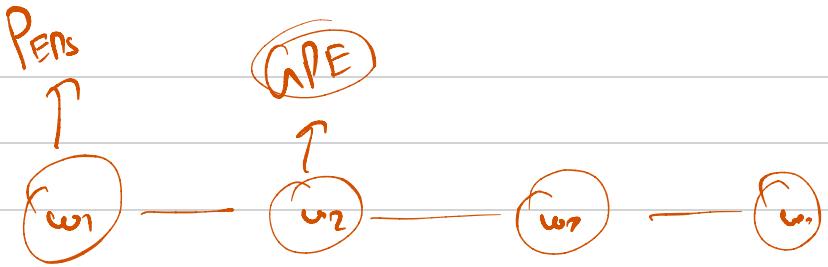
is D: GPE I: GPE \rightarrow his So

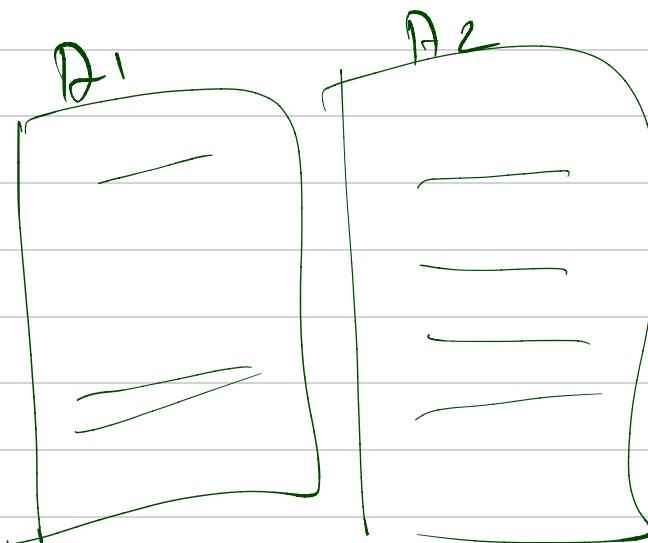
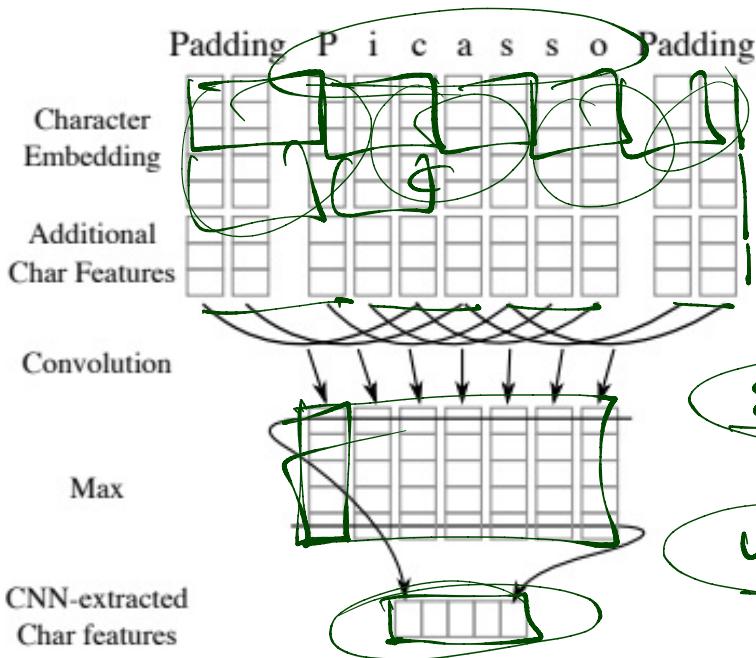
$$LR(s_1 / O_{de} / N_{er}, o) = \underline{v-low}$$

$$LR(s_1 / B'Nom / N_{er}, o) = \underline{v-high}$$

Sig - high \Rightarrow Predictor







$w^1 \rightarrow \text{APE} / \text{PERS}$

w_1 w_2 w_3 w_4 w_5

Grf

Perls

bank



word level

1 (x5) \rightarrow dominant.

River
ben¹¹

bank near river

bank across road.