

Sequence Labeling & Conditional Random Field (CRF)

References:

- https://cocoxu.github.io/CS7650_fall2021/slides/lec6-seq2.pdf
- <https://www.youtube.com/watch?v=2KTeXhfsc-k>

Sequence Labeling Use Cases



When Sebastian Thrun **PERSON** started working on self - driving cars at Google **ORG** in 2007 **DATE** , few people outside of the company took him seriously . “ I can tell you very senior CEOs of major American **NORP** car companies would shake my hand and turn away because I was n't worth talking to , ” said Thrun **PERSON** , in an interview with Recode **ORG** earlier this week **DATED** .

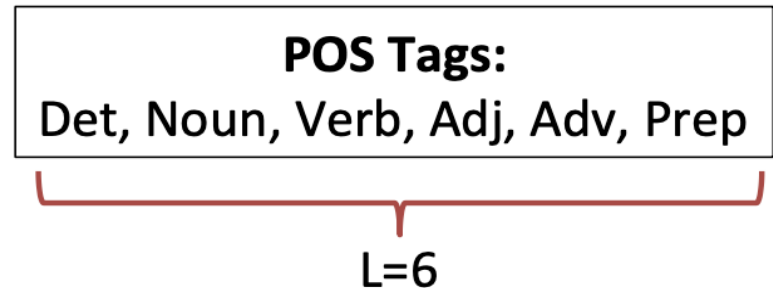
Q: Who is the first wife of Albert Einstein ?

E: Einstein/O married/O his/O first/O wife/O
eva/B Marić/I in/O 1903/O

A: Mileva Marić

Sequence Labeling: POS

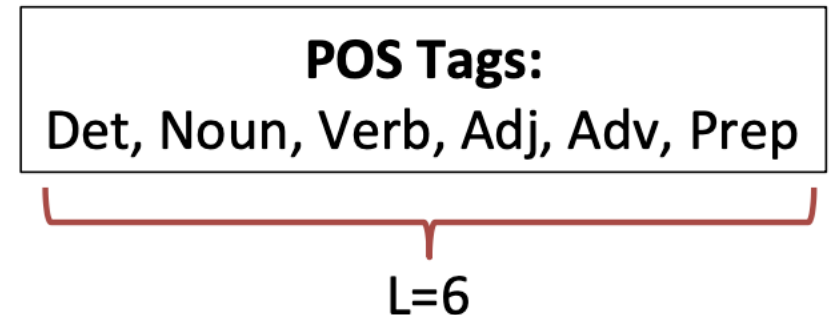
- Input: $x = (x^1, \dots, x^M)$
- Predict: $y = (y^1, \dots, y^M)$
 - Each y^i one of L labels.



- $x = \text{"Fish Sleep"}$
- $y = (N, V)$
- $x = \text{"The Dog Ate My Homework"}$
- $y = (D, N, V, D, N)$
- $x = \text{"The Fox Jumped Over The Fence"}$
- $y = (D, N, V, P, D, N)$

General Multi-class

- $x = \text{"Fish sleep"}$
- $y = (N, V)$
- Multiclass prediction:
 - All possible length- M sequences as different class
 - $(D, D), (D, N), (D, V), (D, Adj), (D, Adv), (D, Pr)$
 $(N, D), (N, N), (N, V), (N, Adj), (N, Adv), \dots$
- **L^M classes!**
 - Length 2: $6^2 = 36!$



General Multi-class

- $x = \text{"Fish sleep"}$
- $y = (N, V)$

POS Tags:
Det, Noun, Verb, Adj, Adv, Prep

$L=6$

- M

Can Model Everything!
(In Theory)

ss

- L

Exponential Explosion in #Classes!
(Not Tractable)

Independent Multiclass

x="I fish often"

POS Tags:
Det, Noun, Verb, Adj, Adv, Prep

L=6

- Treat each word independently (assumption)
 - Independent multiclass prediction per word
 - Predict for x="I" independently
 - Predict for x="fish" independently
 - Predict for x="often" independently
 - Concatenate predictions.

Independent Multiclass

$x = \text{"I fish often"}$

POS Tags:
Det, Noun, Verb, Adj, Adv, Prep

L=6

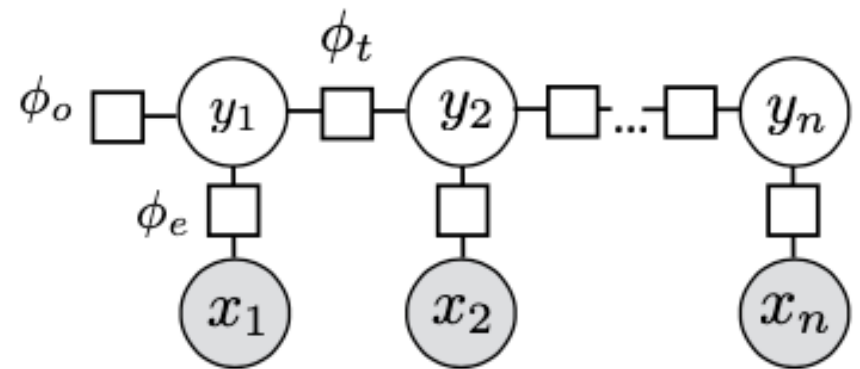
#Classes = #POS Tags
(6 in our example)

Solvable using standard multiclass prediction.
But ignores context!

Sequential Conditional Random Fields (CRF)

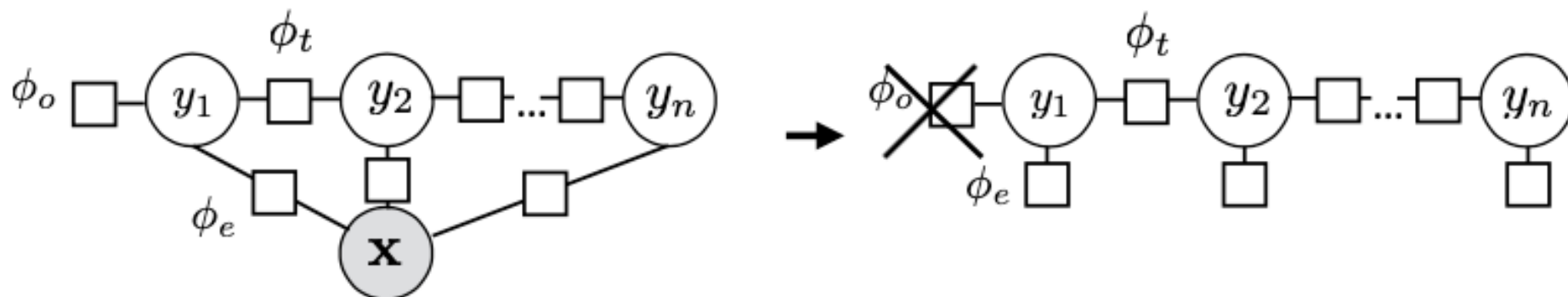
► CRFs:

$$P(\mathbf{y}|\mathbf{x}) \propto \prod_k \exp(\phi_k(\mathbf{x}, \mathbf{y}))$$



$$P(\mathbf{y}|\mathbf{x}) \propto \exp(\phi_o(y_1)) \prod_{i=2}^n \exp(\phi_t(y_{i-1}, y_i)) \prod_{i=1}^n \exp(\phi_e(x_i, y_i))$$

Sequential CRFs

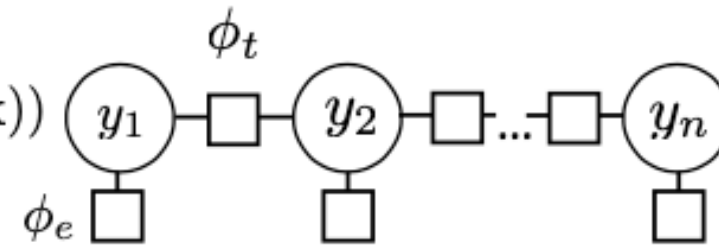


- ▶ Notation: omit \mathbf{x} from the factor graph entirely (implicit)
- ▶ Don't include initial distribution, can bake into other factors

Sequential CRFs:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \prod_{i=2}^n \exp(\phi_t(y_{i-1}, y_i)) \prod_{i=1}^n \exp(\phi_e(y_i, i, \mathbf{x}))$$

Feature Functions

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \prod_{i=2}^n \exp(\phi_t(y_{i-1}, y_i)) \prod_{i=1}^n \exp(\phi_e(y_i, i, \mathbf{x}))$$


The diagram illustrates a Markov chain structure. It consists of a sequence of nodes y_1, y_2, \dots, y_n represented by circles. These nodes are connected by a chain of square nodes. Specifically, there is a square node between y_1 and y_2 , and another between y_2 and y_3 , and so on, up to y_n . Additionally, there is a square node directly below each y_i . The top square node (between y_1 and y_2) is labeled ϕ_t , and the bottom square node (below y_1) is labeled ϕ_e .

- This can be almost anything! Here we use linear functions of sparse features

$$\phi_e(y_i, i, \mathbf{x}) = w^\top f_e(y_i, i, \mathbf{x}) \quad \phi_t(y_{i-1}, y_i) = w^\top f_t(y_{i-1}, y_i)$$

$$P(\mathbf{y}|\mathbf{x}) \propto \exp w^\top \left[\sum_{i=2}^n f_t(y_{i-1}, y_i) + \sum_{i=1}^n f_e(y_i, i, \mathbf{x}) \right]$$

- Looks like our single weight vector multiclass logistic regression model

Features for NER

- ▶ Word features (can use in HMM)
 - ▶ Capitalization
 - ▶ Word shape
 - ▶ Prefixes/suffixes
 - ▶ Lexical indicators
- ▶ Context features (can't use in HMM!)
 - ▶ Words before/after
 - ▶ Tags before/after
- ▶ Word clusters
- ▶ Gazetteers

Leicestershire

Boston

Apple released a new version...

According to the New York Times...