



# **SNS COLLEGE OF ENGINEERING**



**Kurumbapalayam(Po), Coimbatore – 641 007**

**Accredited by NAAC-UGC with 'A' Grade**

**Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai**

## **Department of Artificial Intelligence and Data Science**

**Course Name – 19AD601 – Natural Language  
Processing**

**III Year / VI Semester**

**Unit 2 – WORD LEVEL ANALYSIS**

**Topic 9- Conditional Random Fields**





# Conditional Random Fields

Conditional Random Fields are a discriminative model, used for predicting sequences. They use contextual information from previous labels, thus increasing the amount of information the model has to make a good prediction.

Assuming we have a sequence of input words  $X = x_1 \dots x_n$  and want to compute a sequence of output tags  $Y = y_1 \dots y_n$ . In an HMM to compute the best tag sequence that maximizes  $P(Y|X)$  we rely on Bayes' rule and the likelihood  $P(X|Y)$ :

$$\begin{aligned} \hat{Y} &= \operatorname{argmax}_Y p(Y|X) \\ &= \operatorname{argmax}_Y p(X|Y)p(Y) \\ &= \operatorname{argmax}_Y \prod_i p(x_i|y_i) \prod_i p(y_i|y_{i-1}) \end{aligned}$$

# Conditional Random Fields

However, the CRF does not compute a probability for each tag at each time step. Instead, at each time step the CRF computes log-linear functions over a set of relevant features, and these local features are aggregated and normalized to produce a global probability for the whole sequence.

$$p(Y|X) = \frac{\exp\left(\sum_{k=1}^K w_k F_k(X, Y)\right)}{\sum_{Y' \in \mathcal{Y}} \exp\left(\sum_{k=1}^K w_k F_k(X, Y')\right)}$$

$$p(Y|X) = \frac{1}{Z(X)} \exp\left(\sum_{k=1}^K w_k F_k(X, Y)\right)$$
$$Z(X) = \sum_{Y' \in \mathcal{Y}} \exp\left(\sum_{k=1}^K w_k F_k(X, Y')\right)$$



# Conditional Random Fields

$$F_k(X, Y) = \sum_{i=1}^n f_k(y_{i-1}, y_i, X, i)$$

Each of these local features  $f_k$  in a linear-chain CRF is allowed to make use of the current output token  $y_i$ , the previous output token  $y_{i-1}$ , the entire input string  $X$  (or any subpart of it), and the current position  $i$ .

This constraint to only depend on the current and previous output tokens  $y_i$  and  $y_{i-1}$  are what characterizes a linear chain CRF.



**THANK YOU**