

# NLP Project

POS tagging using CRF tool for Maithili Language

*Jyoti Jha & Pushyami Rachapudi*

# Aim of the project

To run CRF and HMM on a given data of Maithili Language and compare their outputs.

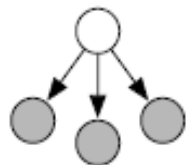
# CRF

- Conditional Random Fields are a probabilistic framework for labeling and segmenting structured data
- CRFs are a type of discriminative undirected probabilistic graphical model.

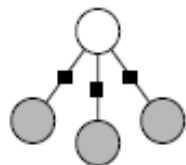
# CRF Vs HMM

- CRF is a discriminative model which outputs a confidence measure. This is really useful in most cases because we want to know how sure the model is about the label at that point.
- HMM is a generative model and it gives the output directly by modeling the transition matrix based on the training data.
- The primary advantage of CRFs over HMMs is their conditional nature, resulting in the relaxation of the independence assumptions required by HMMs. A CRF can be considered as a generalization of HMM or we can say that a HMM is a particular case of CRF where constant probabilities are used to model state transitions. CRFs outperform HMMs on a number of real-world sequence labeling tasks.

# CRF Vs HMM



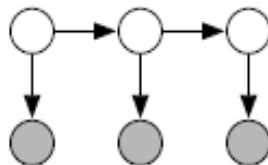
Naive Bayes



Logistic Regression



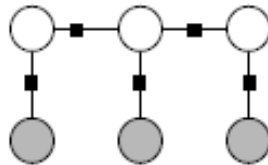
**SEQUENCE**



HMMs



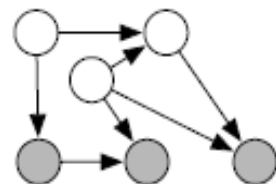
**CONDITIONAL**



Linear-chain CRFs



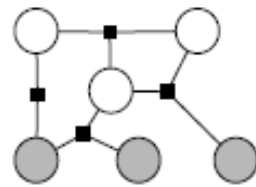
**GENERAL  
GRAPHS**



Generative directed models



**CONDITIONAL**



General CRFs

# Observations

- CRF's accuracy 97%
- For less data, accuracy for HMM is 88%