Statistical Methods for NLP

Semantics, Brief Introduction to Graphical Models

Sameer Maskey

Week 7, March 2010

Topics for Today

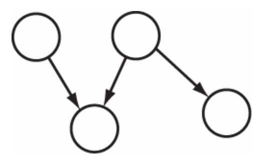
- Brief Introduction to Graphical Models
- Discussion on Semantics and its use in Information Extraction, Question Answering
- Programming for text processing

Graphical Models

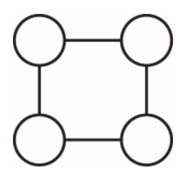
- “Graphical Model is a family of probability distributions defined in terms of a directed or undirected graph [1]” Michael Jordan
- Nodes of the graph represent random variables and lack of arcs represent conditional independence

Graphical Models

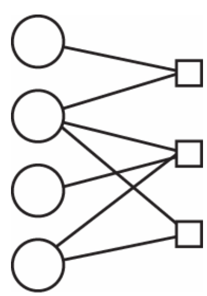
- Joint probability distributions can be computed by taking products over functions defined on connected subset of nodes [1]
- Many different types of graphical models
  - Bayesian Networks



  - Markov Random Fields

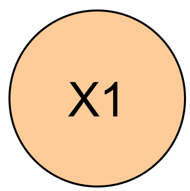


  - Factor Graphs

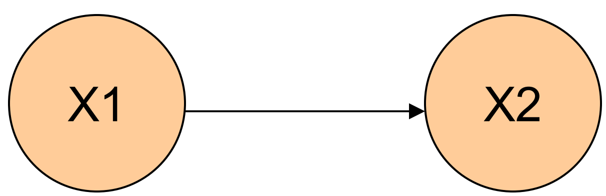


Graphical Models

- Given a random variable X we can represent it as a node in a graph



- We can represent relationships between random variables by arcs



Joint Probability

- Computing joint probability may not be feasible in many cases, too many parameters to estimate

$$P(Y_k, X_1, X_2, ..., X_N)$$

- For N binary variable joint probability table will have 2^N parameters to estimate

Conditional Independence

- Given random variables X, Y,Z, X is conditionally independent of Y given Z if and only if

$$P(X|Y,Z) = p(X|Z)$$

$$\begin{aligned} P(X|Y) &= P(X_1, X_2|Y) \\ &= P(X_1|X_2, Y)P(X_2|Y) \\ &= P(X_1|Y)P(X_2|Y) \end{aligned}$$

Conditional Independence in Naïve Bayes Classifier

$$P(y_k, X_1, X_2, ..., X_N) = P(y_k)\Pi_i P(X_i|y_k)$$

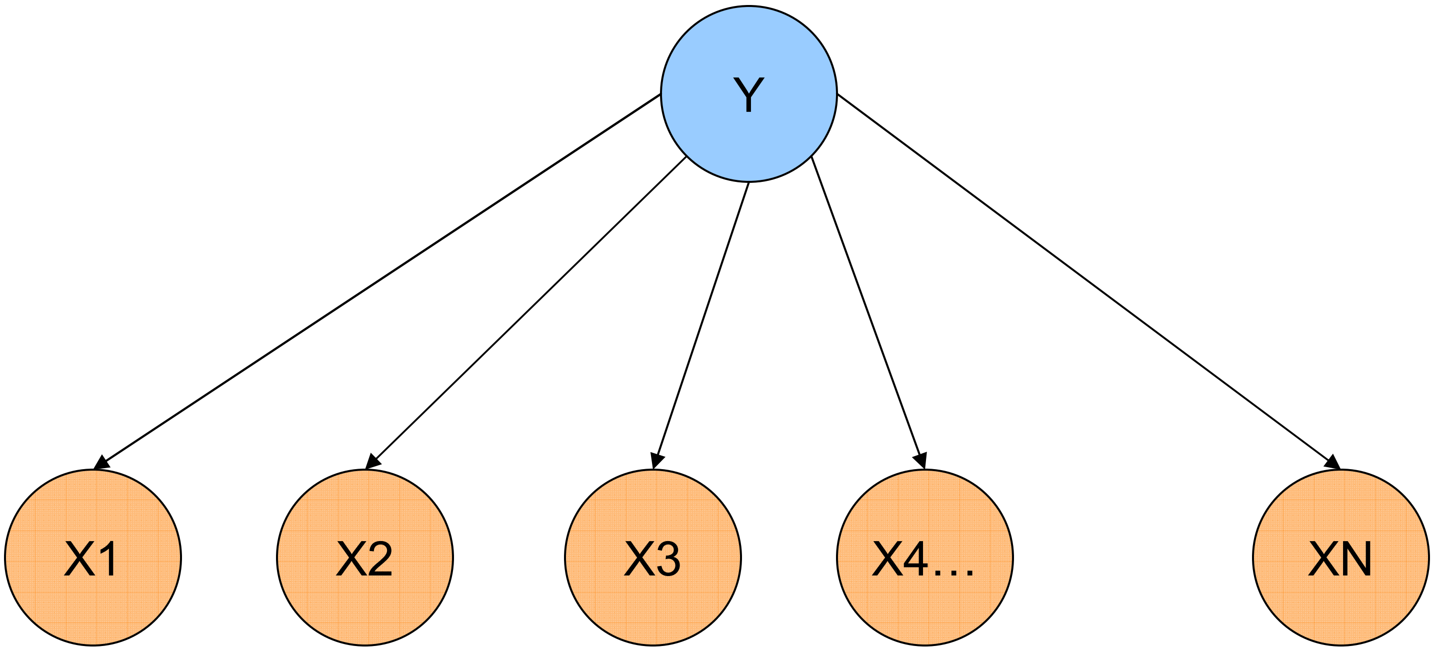
Prior Probability of the Class

Conditional Probability of feature given the Class

Conditional Independence assumption lowered the number of parameters to estimate

Graphical Models Can Efficiently Represent Conditional Independence

- Graphical Model for Naïve Bayes Classifier



Questions We Ask When Modeling

- How can we find distributions that satisfy given independence property?
  - Representation
- How do we use independence property to efficiently make an inference
  - Inference
- How can we find independence properties in data
  - Learning

Bayesian Networks

- Type of graphical model
- Network structure G is Directed acylic graph
- Model represents factorization of the joint probability of all random variables
- Parent-Child relationship represented by an arrow starting at parent with destination to the child

$$p(x_1, x_2) = p(x_1)p(x_2)$$

$$p(x_1, x_2) = p(x_2|x_1)p(x_1)$$

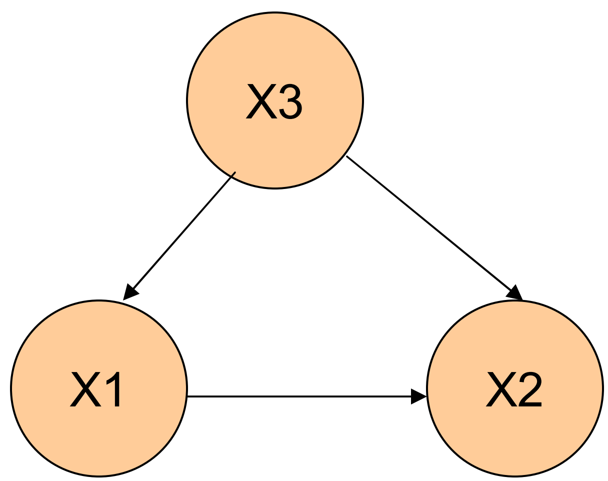
Bayesian Networks

- Graph G where G is acyclic and is defined as follows

$$G = (V, E)$$

$$V = X_1, X_2, ..., X_N$$

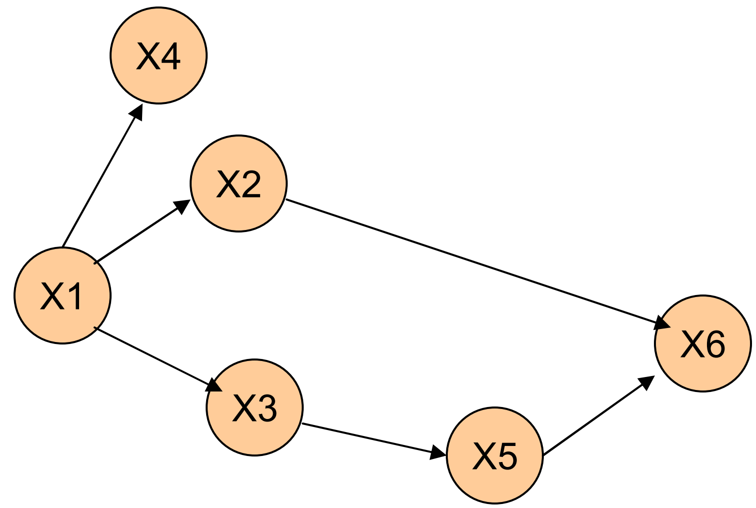
$$E = (X_i, X_j) : i \neq j$$



- Each node has a set of parents

Factorization of Joint Probability

- Factorization of joint probability reduces the number of parameters to estimate



$$P(x_1, ..., x_n) = \prod_{i=1}^{N} p(x_i|\pi_i)$$

$$p(x_1, x_2, x_3, x_4, x_5, x_6) =$$

2^6

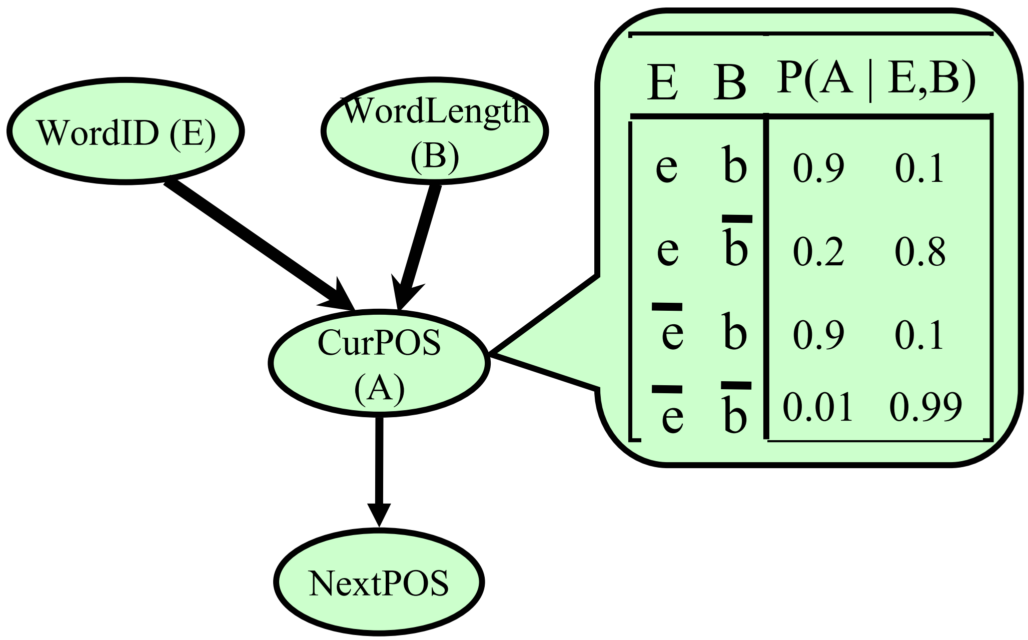
$$p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_1)p(x_5|x_3)p(x_6|x_2, x_5)$$

2^1 2^2 2^2 2^2 2^2 2^3

- Conditional Probability Tables in each node are smaller

Normalization of Probability Tables

- Normalization: Sum of conditional probabilities equals 1 for each setting of parents



Bayesian Network Example

Class Discussion

Semantics and Its Use in IE and QA

References

- [1] Jordan, M., “Graphical Models,” Statistical Science, 2004