Markov Random Fields

Leonid Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

21.11.2013

Conditional probabilities

- Joint probability distribution
  $P(x_0, ..x_n) = P(X_0 = x_o, ..., X_n = x_n) = P(\mathbf{x})$
  random variable X_i takes value x_i
- Chain rule

$$P(x_0, ...x_n) = P(x_0)P(x_1|x_0)P(x_2|x_1, x_0).. = P(x_0)\prod_{i=1}^{n} P(x_i|x_0, ..x_{i-1})$$

- Markov property

$$P(x_0, ...x_n) = P(x_0)P(x_1|x_0)P(x_2|x_1).. = P(x_0)\prod_{i=1}^{n} P(x_i|x_{i-1})$$

- Independent variables

$$P(x_0, ...x_n) = P(x_0)P(x_1)P(x_2).. = \prod_{i=0}^{n} P(x_i)$$

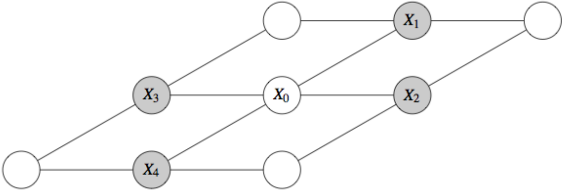
Markov Random Fileds

- A Markov random field is set of random varibles X_i on the finite set of sites $S = \{1..N\}$ (lattice/network) with probability function satisfying Markovian property relative to the neighborhood

$$P(X_i = x_i | X_j = x_j, j \neq i) = P(X_i = x_i | X_j = xj, j \in \mathcal{N}(i))$$

- Undirected graphical model: nodes represent variables, edges represent the dependence srtucture between random variables
- Markov network represents joint probability distribution
- Only neighboring sites interact with each other (local Markov property)

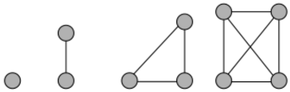
Markov Random Fileds



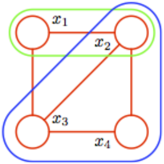
$$\mathcal{N}(X_0) = \{X_1, X_2, X_3, X_4\}$$

Cliques amd maiximal cliques

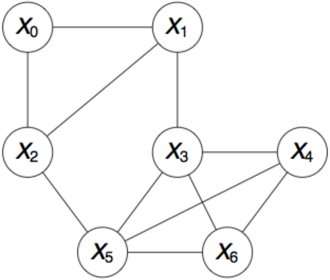
A clique of a graph is its complete subgraph



A maximal clique is a clique that is not a subset of another clique (can not be extended by adding another vertex)



Click factorizaton



$$\{X_0, X_1, X_2\}, \{X_1, X_3\}, \{X_2, X_5\}, \{X_3, X_4, X_5, X_6\}$$

Click factorization

Probability distribution factorizes with respect to a given undirected graph if it can be written as

$$P(X) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(x_c)$$

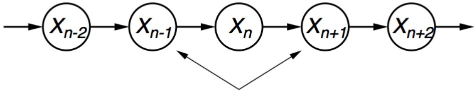
$\mathcal{C}$ - set of maximal cliques, c-maximal clique
ψ_c - factor potentials / clique potentials, real valued function
Z - normalization factor, partition function

$$Z = \sum_{x \in X} \prod_{c \in \mathcal{C}} \psi_c(x_c)$$

if $V_c(x) = -\log \psi_c(x)$

$$P(X) = \frac{1}{Z} e^{-\sum_{c \in \mathcal{C}} V_c(x_c)}$$

Markov chains as MRF



- neighbours $n-1, n+1$
- cliques = pairs
- Joint probability

$$P(x) = P(x_0) \prod_{n=1}^{N} P(x_n | x_n - 1) = P(x_0) \exp \left\{ \sum_{n=1}^{N} \log P(x_n | x_{n-1}) \right\}$$

- Potential $V(x_n, x_{n-1}) = -\log P(x_n | x_{n-1})$

Gibbs distribution

- Gibbs distribution (measure)

$$P(X) = \frac{1}{Z} e^{-\beta U(x)}$$

$U(x)$ - energy function, β - spatial smoothness parameter

- partition function

$$Z = \sum_x e^{-\beta U(x)}$$

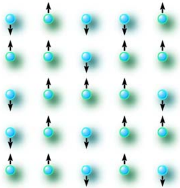
- In statistical mechanics $\beta = 1/kT$, inverse temperature
- Boltzman distribution

Ising Model

- Classical model of magnetizm (1D model solved by Ernst Ising in 1925)
- one or two-dimensional lattices, spin takes values $\sigma_i = \pm 1$
- $$U = -J \sum_{<i,j>} \sigma_i \sigma_j - H \sum_i \sigma_i$$
- Configuration probability

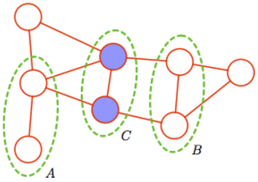
$$P(\sigma) = \frac{1}{Z} e^{-\beta U}$$

- $J > 0$ - ferromagnetic, $J < 0$ - antiferromagmetic



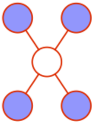
Graph separation

In an underected graph with disjoint subsets of nodes $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ if every path from $\mathcal{A}$ to $\mathcal{B}$ inlcudes at least one node from $\mathcal{C}$, then $\mathcal{C}$ separates $\mathcal{A}$ from $\mathcal{B}$.



Markov blanket

- Markov blanket of a node is the (minimal) set of variables making the given node independent of all the remaining nodes in the model;
- The local Markov property referred to in the Hammersley-Clifford theorem states that the neighbors of a node in a Markov random field are a Markov blanket of that node in the graph. That is, $MB(Xi) = N(Xi)$;



Hammersley - Clifford theorem

Given random vector X and underited graph with positive probability distribution $P(X) > 0$ for any realization of X, the following conditions are equivalent:

- $P(X)$ is a Gibbs distribution that factorizes according to the maximal cliques in the graph
- local Markov property:
  $P(X_i = x_i | X_j = x_j, j \neq i) = P(X_i = x_i | X_j = xj, j \in \mathcal{N}(i))$
- global Markov property:
  If $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ are three disjoint subsets of X and $\mathcal{C}$ separates $\mathcal{A}$ from $\mathcal{B}$, then $P(\mathcal{A}|\mathcal{B}, \mathcal{C}) = P(\mathcal{A}|\mathcal{C})$

Log-linear model

- Any MRF can be written as log-linear model (logistics model)
- For each state of each maximal clique introduce $f(X_c) = \{0, 1\}$
- Potential function

$$\psi_c(x_c) = \exp\left\{\sum_{j=1}^{m_c} w_j f_j(x_c)\right\}$$

m_c number of features in the clique

$$P(x) = \frac{1}{z}\prod_C \exp\left\{\sum_{j=1}^{m_c} w_j f_j(x_c)\right\} = \frac{1}{z}\exp\left\{\sum_C \sum_{j=1}^{m_c} w_j f_j(x_c)\right\}$$