An Overview of Markov Random Field and Application to Texture Segmentation

Seong-Wook Joo

October 2003

1. What is MRF?

MRF is an extension of Markov Process

- MP (1D sequence of r.v.'s)

  unilateral (causal): $p(x_t | x_i,\ i<t) = p(x_t | x_j,\ t-m \leq j \leq t-1)$

  bilateral (non-causal): $p(x_t | x_i,\ i \neq t) = p(x_t | x_j,\ t-m \leq j \leq t+m,\ j \neq i)$

- MRF ((2D) field of r.v.'s)

  $p(x_{i,j} | x_{k,l},\ (k,l) \neq (i,j)\ ) = p(x_{i,j} | x_{k,l},\ (k,l) \in$ some neighborhood of $(i,j)\ )$

2. Notations

(1) Labeling problem: mapping sites to labels

a) sites: $\mathcal{S} = \{1,\ldots,m\}$

b) labels: $\mathcal{L} = \{l_1,\ldots,l_M\}$

c) labeling: each site is assigned a label, f_i, regarded as a function $f : \mathcal{S} \rightarrow \mathcal{L}$

- Sites can be regular/irregular, labels can be continuous/discrete. In the case of regular sites, the order of indices can be the raster scan order of $n \times n$ lattice and $m=n^2$. (Regular assumed here on)
- The "configuration" $\boldsymbol{f} \equiv \{f1,\ldots,f_m\}$ is an instance of labeling of the whole sites. $\mathbb{F} = \mathcal{L}^m$ is the set of all possible configurations.

(2) Neighborhood system & cliques

$\mathcal{N}_I = \{$ sites neighboring $I\ \} = \{\ j \in \mathcal{S} \mid \text{dist}(i,j)^2 \leq r\ \}$: r^{th} order

$\mathcal{C}_k = \{\ \{i_1,\ldots,i_k\} \mid$ all sites within are neighbors to every other $\}$

$\mathcal{C} = \cup_k \mathcal{C}_k$

* see fig 1.2 of [1]

3. Model Formulation

(1) MRF

Definition: MRF on $\mathcal{S}$ wrt $\mathcal{N}$ iff

$P(\boldsymbol{f}) > 0, \forall \boldsymbol{f} \in \mathbb{F}$ (positivity)

$P(f_i \mid \boldsymbol{f}_{\mathcal{S}-\{i\}}) = P(f_i \mid \boldsymbol{f}_{\mathcal{N}i})$ (Markoviantiy)

Specifying an MRF can be done either by (local) conditional probability $P(f_i \mid \boldsymbol{f}_{\mathcal{N}i})$ or by (global) joint probability $P(\boldsymbol{f})$

(2) Gibbs Random Field (GRF)

$p(\boldsymbol{f}) = \frac{1}{Z} e^{-\frac{1}{T}U(\boldsymbol{f})}$, partition function $Z = \sum_{\boldsymbol{f}=\mathbb{F}} e^{-\frac{1}{T}U(\boldsymbol{f})}$

$U(\boldsymbol{f}) = \sum_{c \in C} V_c(f)$: energy function

$V_C(f)$: potential function – depends on f_i's in the cliques

Often used expression for $U(\boldsymbol{f})$ is

$$U(f) = \sum_{\{i\} \in C_1 = S} V_1(f_i) + \sum_{\{i,j\} \in C_2} V_2(f_i, f_j) + \sum_{\{i,j,k\} \in C3} V_3(f_i, f_j, f_k) + \cdots$$

If we consider only the cliques of size one and two

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{j \in N_i} V_2(f_i, f_j)$$

Note $V_2(f_i,f_j) \neq V_2(f_j,f_i)$ is allowed.

(3) Markov-Gibbs Equivalence

"F is an MRF wrt $\mathcal{N}$ $\Leftrightarrow$ F is a GRF wrt $\mathcal{N}$"

($\Leftarrow$) Given GRF $p(\boldsymbol{f})$, then

$$p(f_i \mid \boldsymbol{f}_{S\text{-}\{i\}}) = \frac{p(f_i, \boldsymbol{f}_{S\text{-}\{i\}})}{p(\boldsymbol{f}_{S\text{-}\{i\}})} = \frac{p(\boldsymbol{f})}{\sum_{f_i' \in L} p(f_i', \boldsymbol{f}_{S\text{-}\{i\}})} = \frac{e^{-\sum_{c \in A} V_c(\boldsymbol{f}')}}{\sum_{f_i' \in L} e^{-\sum_{c \in A} V_c(\boldsymbol{f}')}} = \frac{p(f_i, \boldsymbol{f}_{N_i})}{p(\boldsymbol{f}_{N_i})} = p(f_i \mid \boldsymbol{f}_{N_i})$$

where $\boldsymbol{f}' = \{f_i', \boldsymbol{f}_{S\text{-}\{i\}}\}$. This is how we get conditional probabilities from joint probability.

($\Rightarrow$) proofs exist (complicated)

(4) Some examples of MRF models

a) auto-logistic model

$\mathcal{L} = \{0,1\}$

$$U(f) = \sum_{\{i\}\in C_1} \alpha_i f_i + \sum_{\{i,j\}\in C_2} \beta_{ij} f_i f_j \quad , \beta_{ij}\text{: interaction coefficients}$$

If 1st order neighborhood model (4-neighbors) is used this is called the Ising model.

b) auto-normal model (Gaussian MRF, [2])

Assume zero-mean r.v. $y_i \equiv f_i - \mu_i = \sum_{j\in NI} \beta_I \; y_i y_j + e$, where $e \sim N(0,\sigma^2)$. So f_i assumed continuous.

MRF representation: $P(f_i \mid f_{N_i}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \sum_{j\in N_i} \beta_{ij} y_j)^2}$

Gibbs representation: $V_1(f_i) = \dfrac{y_i^2}{2\sigma^2}$, $V_2(f_i) = \dfrac{-\beta_{ij} y_i y_j}{2\sigma^2}$

c) multi-level logistic model (MLL)

$\mathcal{L} = \{0,\ldots,M\}$. Considering clique size up to two ("pair-wise MLL"),

$$V_1(f_i) = \alpha_{f_i} \text{ (depends on label)}$$

$$V_2(f_i) = \begin{cases} \beta_c, & f_i = f_j \\ -\beta_c, & f_i \neq f_j \end{cases}$$

Note V_2 depends on the clique type $c\in\{1,2,3,4\}$. If $\beta_1=\beta_2=\beta_3=\beta_4$ i.e., isotropic, MLL produces blob-like regions and is used for modeling regions with same (discrete) label.

4. Problem Formulation: MAP-MRF labeling

- Problem: Given an MRF and observed data $\boldsymbol{d}$, find an optimal labeling solution $\boldsymbol{f}$. Parameters $\boldsymbol{\theta}$ for the MRF may not be given.
- For the optimality criterion, MAP (maximum *a posteriori*) is often used:

$$\begin{aligned} \boldsymbol{f}^* &= \arg\max_{\boldsymbol{f}\in\mathbb{F}} P(\boldsymbol{f} \mid \boldsymbol{d}) \\ &= \arg\max_{\boldsymbol{f}\in\mathbb{F}} P(\boldsymbol{d} \mid \boldsymbol{f}) P(\boldsymbol{f}) \\ &= \arg\min_{\boldsymbol{f}\in\mathbb{F}} \left\{ U(\boldsymbol{d} \mid \boldsymbol{f}) + U(\boldsymbol{f}) \right\} \end{aligned}$$

(1) MRF Parameter estimation

a) Maximum Likelihood

$$\arg\max_{\boldsymbol{f}} \left\{ p(\boldsymbol{f} \mid \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} e^{-U(\boldsymbol{f}|\boldsymbol{\theta})} \right\}$$

Intractable in general since $Z(\boldsymbol{\theta})$ is hard to evaluate and depends on $\boldsymbol{\theta}$.

b) Pseudo Likelihood: Assume some independence. Use $PL(\boldsymbol{f}|\boldsymbol{\theta}) = \prod_i P(f_i \mid \boldsymbol{f}_{Ni}, \boldsymbol{\theta})$

c) Coding Method: Partition $\mathcal{S}$ into disjoint sets $\mathcal{S}^{(k)}$. Within each $\mathcal{S}^{(k)}$ PL is the true likelihood.

d) Least Squares Method

Using potential function representation and estimated probability, solve for parameters.

$$P(f_i \mid f_{N_i}) = \frac{P(f_i, f_{N_i})}{P(f_{N_i})} = \frac{e^{-U_i(f_i, f_{N_i}, \boldsymbol{\theta})}}{\sum_{f_i \in L} e^{-U_i(f_i, f_{N_i}, \boldsymbol{\theta})}}$$

$P(f_i \mid \boldsymbol{f}_{Ni})$ can be estimated using histogram method. Take log, becomes linear in $\boldsymbol{\theta}$, use all i's to form over determined linear system, solve LS solution.

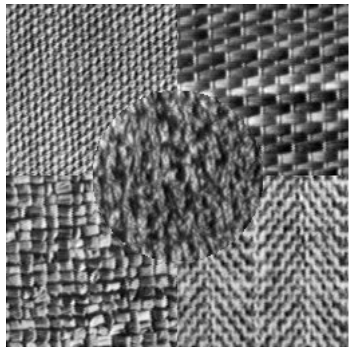
(2) Optimization

Local optimization: Iterated Conditional Modes, Relaxation Labeling, Dynamic Programming, …

Global optimization: Simulated Annealing, Mean Field Annealing, Graduated Non-Convexity, …

5. Application: Texture Segmentation [3,4]

(1) Problem



(2) Model

Two different MRF were used. A 4th order GMRF is used as the individual texture field and a simple 2nd order pair-wise MRF is used for the texture label field.

Texture field

For each texture type, an image block (64×64 pixels) is fitted to a 4th order GMRF model. The zero-mean (all images are assumed to be zero-mean) texture image region, $\{y(s) \mid s\in\Omega\}$, $\Omega=\{\ s=(i,j) \mid 1\leq i,j\leq M\}$ is obtained by subtracting from each pixel the average of the local (7×7) neighborhood. The 4th order neighborhood is defined by the following

			10	
			4	9
		s	1	5
		2	3	7
		6	8	

s denotes the site where the neighborhood is defined. The shaded region denotes the asymmetrical neighbor set N_s. The estimates of the parameters are obtained by the least squares method

$$\boldsymbol{\theta}^* = [\ \Sigma_{s\in\Omega}\, \boldsymbol{q}(s)\ \boldsymbol{q}\ (s)^{\mathrm{T}}\ ]^{-1}\ (\textstyle\sum_{s\in\Omega} \boldsymbol{q}\ (s)\ y(s))\ ,\ \boldsymbol{q}\ (s) = \mathrm{col}\ [y(s+r) + y(s\text{-}r)\ ,\ r\in N_s]$$
$$\mathbf{v}^* = 1/M^2 \textstyle\sum_{s\in\Omega} (\ y(s) - \boldsymbol{\theta}^{*\mathrm{T}}\ \boldsymbol{q}\ (s))^2$$

Ω is assumed to be wrapped around in a toroidal manner.

Texture label field

The label value can have {1,2,3,4,5}. The following model (energy function) is used

$$U(\ L(s) \mid L(s+r),\ r \in N_s\ ) = -\beta \sum_{r \in Ns} \delta(L(s) - L(s+r))$$

where N_s is the 2nd order neighborhood (8-neighbor) and $\delta(.)$ is the Kronecker delta. β is the parameter indicating the degree of clustering.

(3) Segmentation algorithm

The MRF labeling problem posed as a MAP (maximum *a posterior*) problem: Find L that maximizes

$$P(\boldsymbol{L} \mid \boldsymbol{y}) \propto P(\boldsymbol{y} \mid \boldsymbol{L}) P(\boldsymbol{L}).$$

The ICM method is used for the optimization. The main step of ICM is to maximize the (local) conditional posterior probability which satisfies

$$P(L(s) \mid L(s+r),\ y(s),\ r \in N_{Ls}\ ) \quad \propto \quad P(y(s) \mid L(s))\ P(L(s) \mid L(s+r),\ r \in N_{Ls}\ )$$

for each s. Here, $y(s)$ is assumed to be independent of $L(s+r)$, $r \in N_L$. Note that for each update, the only unknown is $L(s)$ – everything else is given. $P(y(s)|L(s))$ is actually $P(y(s)|L(s),\ y(s+r),\ r \in Ny_s)$. The probabilities on the right hand side are given by

$$P\ (y(s) \mid L(s)) = (1/(2\pi\ \nu_l)^{1/2})\ \exp[\ \{y_s - \textstyle\sum_{r \in Nys \cup -Nys} \theta_l(r)\ y(s+r)\ \}^2/2\nu_l]$$
$$P(L(s) \mid L(s+r),\ r \in N_{Ls}\ ) = \exp(\beta \textstyle\sum_{r \in NLs} \delta(L(s) - L(s+r))\ /\ Z$$

where Z is a normalizing constant, and subscript l denotes a parameter dependent of the label l. The algorithm is simple: (i) Initialize $\boldsymbol{y}$ with some values. (ii) For each s in the image, update $L(s)$ according to the above step (iii). Repeat step (ii) for max iteration.

Notes on ICM:

- ICM can be used when the unknown labeling is discrete.
- Initial labeling is usually assigned by the maximum likelihood values.
- Convergence is guaranteed to a *local* minimum.
- The final result tends to be highly dependent on the initial labeling.

References

[1] S. Z. Li, “Markov Random Field Modeling in Image Analysis”, Springer, 2001.

[2] R. Chellappa, “Two Dimensional Discrete Gaussian Markov Random Field Models for Image Processing”, *Progress in Pattern Recognition*, volume 2, pages 79-112. Elsevier Science Publishers B. V., 1985.

[3] B. S. Manjunath, T. Simchony, and R. Chellappa. “Stochastic and Deterministic Networks for Texture Segmentation”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-38:1039-1049, June 1990.

[4] R. C. Dubes and A. K. Jain. “Random Field Models in Image Analysis”, *Journal of Applied Statistics*, 16:131-164, 1989.