Lecture 2: Mathematical MRF Model

Anand Mishra

Center for Visual Information Technology
IIIT Hyderabad
http://researchweb.iiit.ac.in/∼anand.mishra/

March 2017

Objective of today's Tutorial

- Understanding the Maths behind MRF
- Basic terminologies
- Formulating MAP-MRF Estimation

This is where we stopped in the last lecture

- In MAP framework, We proved optimal configuration is:

$$f^* = \arg\max_f p(d|f)P(f)$$

- then we claimed all possible configuration are not of interest. Since Images are not complete random.
- Finally we posed a question how to model P(f) and said that an image can be treated as Markov Random Field to model P(f).
- So in this lecture we formally define MRF and formulate MAP-MRF framework.

Contents

Neighbourhood

The neighbourhood relationship has following property:

1. A site is not neighbouring to itself i.e. $i \notin N_i$
2. Neighbouring relationship is symmetric i.e.
   $i \in N_{i'} \implies i' \in N_i$

thus we can write neighbourhood set of i as:
$N_i = \{i' \in S : dist(i, i') <= r, i \neq i'\}$ where r is any positive real number. Obviously this gives us freedom to define different type of neighbourhood based on distance used and value of r.

Various type of neighbourhood

- First order neighbourhood(4-nbd)

	1	
1	X	1
	1	

- Second order neighbourhood(8-nbd)

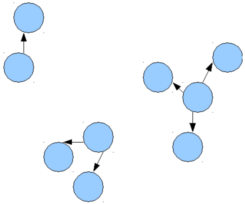
2	1	2
1	X	1
2	1	2

- Higher order neighbourhood

5	4	3	4	5
4	2	1	2	4
3	1	X	1	3
4	2	1	2	4
5	4	3	4	5

Various type of neighbourhood

- Regular neighbourhood: The neighbourhoods we discussed till now are regular neighbourhoods. As they are defined on regular sites and thus have regular shape and size.
- Irregular neighbourhood: But as there is a notion of irregular sites, we can also have irregular neighbourhood. Shape and sizes in irregular neighbourhood are irregular.

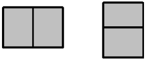


Clique

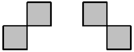
- Let us try to understand the concept of clique in our context.
- Let us construct a graph having set of sites as vertices and edges as neighbourhood system (i.e. vertex i and j share a link iff they are neighbour of each other)
- So G=(S,N) is a graph.
- A clique is a subset of sites in G.
- Every site is trivially a clique of order one.

Clique

- Set of sites (i,j) (where i and j are neighbours of each other) is is clique of order two. Bellow is the example of cliques in first order neighbourhood system.



  However, bellow are not clique in first order neighbourhood system.



- Similarly we can define Higher order cliques.

Markov Random Field

Let $S = \{1, 2, \cdots, m\}$ be a set of sites and $L = \{l_1, l_2, \cdots, l_n\}$ be a set of levels. Further suppose $F = \{F_1, F_2, \cdots, F_m\}$ be a family of random variables defined on a set S. F can take all possible configuration defined over sites S and Labels L. We define term $P(f)$ as the probability of a random vector F taking particular configuration f.

Then F is said to be a **Markov random field** on S with respect to neighbourhood $\mathcal{N}$ if and only if following two conditions are satisfied:

1. $P(f) > 0, \forall f \in F$
2. $P(f_i | f_{s-\{i\}}) = P(f_i | f_{\mathcal{N}_i})$

Conditionals are problematic

- There are two approaches to specify an MRF : First as Joint probability P(f) and second as conditional probability $P(f_i|f_{\mathcal{N}_i})$
- Unfortunately, writing down the conditional distributions does not work (most of the time, they contradict each other)
- Example: suppose that we have predict 2 $\times$ 2 patches based on other intensities

A	B
C	D

Suppose that A=C=D=1 implies B=1
Also, suppose A=B=D=0 implies C=0

Conditionals are problematic

- Now consider:

0	1	1
0	?	1
0	0	1

- So we need some mechanism to model joint probability $P(f)$.

Gibbs Random Field

A set of random variables F is said to be GRF on S with respect to $\mathcal{N}$ if and only if configuration obeys Gibbs distribution.

$$P(f) = Z^{-1} e^{\frac{-1}{T} U(f)}$$

where

$$U(f) = \sum_{c \in C} V_c(f)$$

and Z is the partition function. What is the relationship between MRF and GRF?

Hammersley Clifford theorem

Statement: F is an Markov random field on S with respect to $\mathcal{N}$ if and only if it is a Gibbs random field on S with respect to $\mathcal{N}$.

- Advantage:
  1. It provides simple way of specifying joint probability $P(f)$
  2. One can specify joint probability $P(f)$ by specifying clique potential functions $V_c(f)$ and choosing appropriate potential functions for desired system behavior.

Back to MAP Estimation

Once again we are back to our MAP estimation equation

$$f^* = \arg\max_f p(d|f)P(f)$$

Due to MRF-Gibbs equivalence we can write:

$$P(f) = Z^{-1} e^{\frac{-1}{T} U(f)}$$

where U(f) is known as prior energy (or clique potential or MRF prior). One example of clique potential is:
$U(f) = \sum_i \sum_{j \in N_i} (f_i - f_j)^2$

- Note that this is not the only type of clique potential.
- clique potential are modelled according to the problem.
- some time can also be learned from the training data.

Back to MAP Estimation

Observations are often noisy!! Let us assume the error (noise) e_i in observation d_i and actual configuration f_i is coming from Gaussian distribution of mean 0 and standard deviation σ. In other words, $e_i \sim N(0, \sigma^2)$, Hence

$$p(d|f) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(e_i - 0)^2/2\sigma^2}$$

or

$$p(d|f) = \frac{1}{\prod_{i=1}^{m} \sqrt{2\pi\sigma^2}} \prod_{i=1}^{m} e^{-(e_i)^2/2\sigma^2}$$

or

$$p(d|f) = \frac{1}{\prod_{i=1}^{m} \sqrt{2\pi\sigma^2}} e^{-} U(d|f)$$

where

$$U(d|f) = \sum_{i=1}^{m} (f_i - d_i)^2 / 2\sigma^2$$

since $e_i^2 = (f_i - d_i)^2$

Thus we can write:

$$f^* = \arg\max_f \{ \frac{1}{\prod_{i=1}^{m} \sqrt{2\pi\sigma^2}} e^{-U(d|f)} \times Z^{-1} e^{\frac{-1}{T} U(f)} \}$$

Back to MAP estimation

Taking negative logarithm both side we can rewrite above equation as:

$$f^* = \arg\min_f \{U(d|f) + U(f)\}$$

or

$$f^* = \arg\min_f \{\sum_{i=1}^m (f_i - d_i)^2 / 2\sigma^2 + U(f)\}$$

The right hand side of the above equation is popularly known as **posterior energy**. So our problem of finding optimal configuration becomes exactly equivalent to energy minimization.

Back to MAP Estimation

- So we have seen here how MAP estimation in MAP-MRF framework leads to energy minimization.
- The two terms of the posterior energy
    1. $\sum_{i=1}^{m}(f_i - d_i)^2$
    2. $U(f)$

  are also called Data term and Smoothness term respectively.

Properties of MRF prior

1. It should be function penalizing the penalizing the violation of smoothness caused by the difference between labels of neighbour's. i.e. $U(f) = g(f_i - f_j)$
2. The function should be **even**
$$g(-\eta) = g(\eta)$$
so that we have same penalty for negative or positive violation of smoothness.
3. It should be **non decreasing**
$$g^{'}(\eta) \geq 0$$
This is obviously required because we should have more penalty for more violation in smoothness.

4. It should be **Bounded**.

$$\lim_{\eta \to \infty} |g^{'}(\eta)| = C < \infty$$

This property is also known as discontinuity preserving property. This is required to avoid over smoothness and preserve edges.

One such function which satisfies all above properties is:

$$U(f) = min((f_i - f_j)^2, K)$$

Here f_i and f_j are labels at two neighbouring sites i and j. and K is a real constant.

References

- Stan Z. Li, Markov Random Field Modeling in Image Analysis , Spriger 3rd Ed. , chapter -2, 2009
- S. Geman and D. Geman, ”Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach. Intell, 6, 721-741, 1984