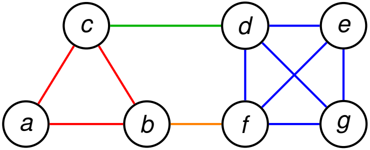
Lecture Pattern Analysis

Part 22: Markov Random Fields

Christian Riess
IT Security Infrastructures Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
June 25, 2021

Introduction

- A Markov Random Field (MRF) is just an undirected graphical model[1]
- The joint PDF factorizes as discussed in lecture 17:



$$p(a, b, c, d, e, f, g) = p(a, b, c)p(c, d)p(b, f)p(d, e, f, g) \quad (1)$$

- Note that these factors are maximal cliques (fully connected subgraphs)
- The factors in MRFs are typically strictly positive potential functions $\psi_C(\mathbf{x}_C)$ over a set of nodes $\mathbf{x}_C$ that form a maximal clique of the form

$$\psi_C(\mathbf{x}_C) = \exp(-E(\mathbf{x}_C)) \ , \quad (2)$$

where $E(\mathbf{x}_C)$ is an energy function (i.e., lower energy is closer to the goal)

[1] The literature source for this lecture is Bishop Sec. 8.3

Factorized MRF and the Partition Function

- The factorized MRF is hence

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{C} \psi_C(\mathbf{x}_C) \tag{3}$$

where $\mathbf{x}$ is a collection of all random variables, and

$$Z = \sum_{\mathbf{x}} \prod_{C} \psi_C(\mathbf{x}_\mathbf{C}) \tag{4}$$

is the **partition function** that normalizes the distribution by summing over all combinations of variable assignments

- Z enables us to use any positive function for the potentials, however, ...
- ...calculating Z is (almost always) intractable, which
  - prohibits us to learn the potential functions from data
  - still allows to perform inference for given potentials

MRF Inference

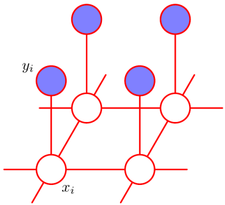
- MRFs are oftentimes used for labelling tasks
  - Each random variable is assigned one label from a **discrete** set of labels
  - We will look at Bishop's example of denoising a binary image: each pixel has a hidden variable, MRF inference decides for black or white
  - Another example is depth estimation from stereo vision: Here, the labels are depth values that are assigned to each pixel/variable
- The optimization task for inference is

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}}\, p(\mathbf{x}) = \underset{\mathbf{x}}{\operatorname{argmax}}\, \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C) \tag{5}$$

- Inference can be done with Gibbs Sampling:
  1. Select a node $\mathbf{x}_i$ (randomly, or following some pattern)
  2. Get the current labels from all neighbors (this "throws" a Markov blanket!)
  3. Assign a new label to $\mathbf{x}_i$ with minimal energy
  4. Goto 1 for a fixed number of iterations, or until convergence
- Gibbs sampling is only locally optimal
- Graph cuts find a globally optimal solution for binary labels ($\rightarrow$ next lecture)

Example: Denoising a Binary Image — Graph Structure

Figure 8.31 An undirected graphical model representing a Markov random field for image de-noising, in which x_i is a binary variable denoting the state of pixel i in the unknown noise-free image, and y_i denotes the corresponding value of pixel i in the observed noisy image.



- A typical MRF graph treats the unknown solution as hidden variables
- For denoising, the unknowns are the pixel values of the denoised image
- The hidden variables are connected in a grid (like the pixel grid)
- Each hidden variable has an associated observation, i.e., a noisy input pixel
- Hence, the maximal clique size is 2, and there are 2 types of connections (observations-hidden and hidden-hidden)
- Bishop models the binary pixels as $-1, +1$ values

Example: Denoising a Binary Image — Energy Functions

- Energy functions for observations-hidden connections are chosen as

$$E(x_i, y_i) = -\eta x_i y_i \quad , \tag{6}$$

with weight constant η, i-th input pixel y_i, and i-th hidden variable x_i

- Energy functions for hidden-hidden connections are chosen as

$$E(x_i, x_j) = -\beta x_i x_j \quad , \tag{7}$$

with a weight constant β

- Both functions provide lower energy for identical pixel signs: we encourage
    1. solutions that are close to the input
    2. solutions that are smooth (neighboring pixels are identical)
- To prefer a label, add also a bias term $E(\mathbf{x}_i) = h\mathbf{x}_i$ with suitable h

Example: Denoising a Binary Image — Whole Energy Function and Result

- The whole energy function is

$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{i,j \in \mathcal{N}(i)} x_i x_j - \eta \sum_i x_i y_i \tag{8}$$

- Top left: clean image
- Top right: noisy image
- Bottom left: Gibbs sampler output
- Bottom right: Graph cut output

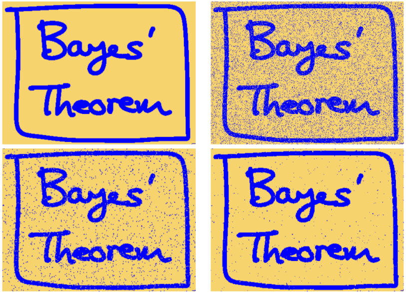


Figure 8.30 Illustration of image de-noising using a Markov random field. The top row shows the original binary image on the left and the corrupted image after randomly changing 10% of the pixels on the right. The bottom row shows the restored images obtained using iterated conditional models (ICM) on the left and using the graph-cut algorithm on the right. ICM produces an image where 96% of the pixels agree with the original image, whereas the corresponding number for graph-cut is 99%.

Remarks

- Potentials with energy terms $E(\mathbf{x}_i, \mathbf{y}_i)$ are also called **unary potentials**
- Potentials with energy terms $E(\mathbf{x}_i, \mathbf{x}_j)$ are also called **pairwise potentials**
- For anyone with an optimization background: one can identify unary potentials with the data term, and pairwise potentials with the regularizer
- The potentials allow more general functions than conditional probabilities. However, one could also write an MRF, e.g., as

$$p(x_1, \ldots, x_N, y_1, \ldots, y_N) = \prod_{i=1}^{N} p(y_i|x_i) \cdot \prod_{\substack{i=1 \\ x_j \in \mathcal{N}(x_i)}}^{N} p(x_i|x_j) \tag{9}$$

The Hammersley-Clifford Theorem shows the equivalence of this probabilistic formulation and the clique potential formulation