Two Parts

(1) Markov Random Fields for Spatial Context

(2) Mean Field Theory for Inference on Markov Random Fields

JOHNS HOPKINS UNIVERSITY

Markov Random Fields (MRFs)

We introduce MRFs to model specified context of semantic segmentation

They capture the intuition that if one pixel is an airplane then neighborhood pixels are likely to also be airplanes

Energy Function $$E(\mathbf{x}) = \sum_{i \in W} \psi_i(x_i) + \sum_{j \in N(i), i \in W} \psi_{ij}(x_i, x_j)$$

$i \in W$: The pixels of the image

x_i : The Pixel label, e.g. x_i={Airplane, Sky, …}

$\psi_i(x_i)$: Unary evidence for pixel i to have label x_i

Can be learned by a Deep Network

$\psi_{ij}(x_i, x_j)$: Context terms

$N(i)$: The neighborhood of pixel i

JOHNS HOPKINS UNIVERSITY

Markov Random Fields (MRFs)

The goal is to find the spatial configuration $\hat{\mathbf{x}} = \{\hat{x}_i : i \in W\}$

such that $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} E(\mathbf{x})$

This gives the spatial configuration which best combines the unary and the binary terms

Problem: Estimating $\hat{\mathbf{x}}$ is typically very hard. Often an NP complete problem
If each x_i takes M states (airplane, sky …), then there are $M^{|W|}$ possible configurations of $\mathbf{x}$ ← Too many to enumerate

Probabilistic formulation

Gibbs Distribution $P(\mathbf{x}) = \dfrac{e^{-E(\mathbf{x})}}{Z}$ where $Z = \sum_{\mathbf{x}} e^{-E(\mathbf{x})}$

Note: typically impossible to compute Z

JOHNS HOPKINS UNIVERSITY

Markov Random Fields (MRFs)

Solving $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} E(\mathbf{x})$ ➡ equivalent to finding $\hat{\mathbf{x}} = \arg\max P(\mathbf{x})$

There are several algorithms for estimating $\hat{\mathbf{x}}$ from the probabilistic formulations

(1) Sampling by Markov Chain Monte Carlo (MCMC)

(2) Mean Field Theory (MFT)

(3) Belief Propagation (BP)

This course will describe MFT & BP.

A handout (Yuille chapter) will describe all three.

MCMC is almost always much slower than MFT & BP.

JOHNS HOPKINS UNIVERSITY

Mean Field Theory (MFT)

Instead of directly estimating $\hat{\mathbf{x}} = \arg\max P(\mathbf{x})$, try to find a distribution $Q(\mathbf{x})$ which approximates $P(\mathbf{x})$ and, from which, $\hat{\mathbf{x}} = \arg\max Q(\mathbf{x})$ can be estimated

$$Q(\mathbf{x}) = \prod_{i \in W} q_i(x_i)$$

Factorizable, so easy to estimate arg max Q

Then $\hat{\mathbf{x}} = \arg\max Q(\mathbf{x})$ can be computed by $\hat{x}_i = \arg\max q_i(x_i)$

$$\hat{\mathbf{x}} = \left\{ \hat{x}_i : i \in W \right\}$$

Select $\hat{Q}(\mathbf{x}) = \arg\min_{Q} \sum_{\mathbf{x}} Q(\mathbf{x}) \log \dfrac{Q(\mathbf{x})}{P(\mathbf{x})}$

Kullback-Leibler divergence between $Q(\mathbf{x})$ & $P(\mathbf{x})$: Relates to cross entropy

Free energy $F[Q] = \sum_{\mathbf{x}} Q(\mathbf{x}) \log Q(\mathbf{x}) - \sum_{\mathbf{x}} Q(\mathbf{x}) \log P(\mathbf{x})$

Mean Field Theory (MFT)

Minimizing $F[Q]$ with respect to Q is difficult, often impossible, but there exist algorithms which often find good approximate solutions.

Note: $F[Q]$ is a differentiable function of a continuous variable Q

So we have replaced a combinatorial optimization problem $\mathbf{x} = \arg\min E[\mathbf{x}]$

With **x** discreate by an optimization problem in continuous var

$$\hat{Q} = \arg\min F[Q]$$

Example of MFT

Simplify by making x_i a binary variable $x_i = \{0, 1\}$

Replace $\mathrm{E}[\mathbf{x}]$ by $\sum_{i \in W} x_i \psi_i + \sum_{i \in W, j \in N(i)} x_i x_j \psi_{ij}$

Use the identity $\psi(x_i) = x_i \psi(x_i = 1) + (1 - x_i)\psi(x_i = 0)$

$$P(\mathbf{x}) = \frac{1}{Z} e^{-\sum_i x_i \psi_i - \sum_{i,j} x_i x_j \psi_{ij}}, Q(\mathbf{x}) = \prod_i q_i(x_i)$$

(**Note**: drop $\in W, \in N(i)$ from $\sum_i, \sum_j$ for simplicity)

Let $q_i(x_i = 1) = q_i,\ q_i(x_i = 0) = 1 - q_i$

Then $\sum_{\mathbf{x}} Q(\mathbf{x}) \log Q(\mathbf{x}) = \sum_i \{q_i \log q_i + (1 - q_i) \log(1 - q_i)\}$

$$\log P(\mathbf{x}) = -\sum_i x_i \psi_i - \sum_{i,j} x_i x_j \psi_{ij} - \log Z$$

$$\Rightarrow \sum_{\mathbf{x}} Q(\mathbf{x}) \log P(\mathbf{x}) = -\sum_i q_i \psi_i - \sum_{i,j} q_i q_j \psi_{ij} - \log Z$$

Example of MFT

$$F[Q] = \sum_{\mathbf{x}} Q(\mathbf{x}) \log Q(\mathbf{x}) - \sum_{\mathbf{x}} Q(\mathbf{x}) \log P(\mathbf{x})$$

$$= \sum_i \{q_i \log q_i + (1-q_i)\log(1-q_i)\} - \sum_i q_i \psi_i - \sum_{i,j} q_i q_j \psi_{ij} - \log Z$$

Note: the $\log Z$ term is independent of the q's. so we do not need to know it when we minimize $F[Q]$

Can minimize $F[Q]$ by steepest descent

$$q_i^{t+1} = q_i^t - \eta \frac{\partial}{\partial q_i} F(\mathbf{q}^t)$$

Better is to use a discrete optimization algorithm

Discrete optimization algorithm $q_i^{t+1} = \dfrac{e^{\left\{-\psi_i - \sum_j q_j^t \psi_{ij}\right\}}}{1 + e^{\left\{-\psi_i - \sum_j q_j^t \psi_{ij}\right\}}}$

Convergence can be guaranteed by variational bounding and, in particular by Concave-Convex procedure (CCCP) provided certain ψ_{ij} conditions apply.

Derivative

$$E_{\text{convex}}[q] = \sum_i q_i \log q_i + (1-q_i)\log(1-q_i) \Rightarrow \frac{\partial E_{\text{convex}}}{\partial q_i} = \log \frac{q_i}{1-q_i}$$

$$E_{\text{concave}}[q] = -\sum_i q_i \psi_i - \sum_{i,j} q_i q_j \psi_{ij} \Rightarrow \frac{\partial E_{\text{concave}}}{\partial q_i} = -\psi_i - \sum_j q_j \psi_{ij}$$

Setting $\dfrac{\partial E_{\text{concave}}}{\partial q_i^{t+1}} = -\dfrac{\partial E_{\text{convex}}}{\partial q_i^t}$ (CCCP)

Note: Concavity of E_{concave} depends on ψ_{ij} .

Can add extra term $\lambda \sum q_i^2$ to ensure convexity

The algorithm converges to a local minimum of $F[Q]$.
This is often a good approximate solution.

Continuum Method

Can smooth the probability distribution by introducing a temperature T

$$P[\mathbf{x};T] = \frac{e^{-E[\mathbf{x}]/T}}{Z[T]}$$

Free energy $F[Q,T] = \sum_{\mathbf{x}} Q(\mathbf{x}) \log Q(\mathbf{x}) - \frac{1}{T} \sum_{\mathbf{x}} Q(\mathbf{x}) E(\mathbf{x}) - \log Z$

Minimize $F[Q, T]$ for large T

Use as initialization to minimize $F[Q, T]$ for small T

➡ Deterministic Annealing