Review Section

Variable Elimination, Particle Filtering and Gibbs Sampling

Hao Su

October 26, 2012

Outline

First Example of VE in CSP Problems

Assume we have the following equation system

$$\begin{cases} 2x_1 + x_2 - x_3 = 0 \\ x_2 + x_3 = 1 \end{cases}$$

- Q1: Draw the factor graph
- Q2: Reformulate the problem as a CSP problem
  - Use indicator functions to represent constraints
  - Syntax of indicator function: [statement]
- Q3: Show how to eliminate x_2

A More Interesting Example of VE in CSP Problems

Assume we have the following equation system

$$\begin{cases} x_1 + 2x_2 & = 1 \\ 5x_2 + x_4 & = 3 \\ x_2 - x_3 & \leq 4 \\ 3x_3 + x_4 & \leq 5 \\ x_1 + x_3 & \geq 0 \end{cases}$$

- Q1: Draw the factor graph
- Q2: Reformulate the problem as a CSP problem
- Q3: Eliminate x_3

General Principle

CSPs: $f_{new}(x) = [\exists x_i : f_j(x) = 1 \textit{ for all } j = 1, \ldots, k]$

A More Interesting Example of VE in CSP Problems

- Q4: Rewrite the CSP formulation in factor product form
- Q5: Check if solution exists (max-product)
- Q6: Count number of solutions (sum-product)

	Unweighted	**Weighted**
max	CSPs (exist solution?)	weighted CSPs
sum	CSPs (# solutions)	Markov networks

Weighted CSP (Assign Weights to Each Constraint)

- Replace indicator functions ($f(x) \in \{0, 1\}$) by real functions ($f(x) \in \mathbb{R}$)
- Q1: Find the highest weight assignment
- Q2: Make an equivalent problem that does not contain x_3
  - Solution: VE by max-product
  - Note that the number of variables is reduced by 1

General Principle

Weighted CSPs: $f_{new}(x) = \max_{x_i} \prod_{j=1}^{k} f_j(x)$

Variable Elimination in MRF

- Assume we have four random variables X_1, X_2, X_3, X_4, the factor graph forming a chain
  - i.e., $P(X_1, X_2, X_3, X_4) \propto f(X_1, X_2)f(X_2, X_3)f(X_3, X_4)$
- We want to compute $P(X_2)$
- How? We eliminate X_1, X_3, X_4.

Basic Concepts

- Random Variable: a variable whose value is subject to variations due to chance
- Example: $X \sim Bernoulli(p)$
- Sample: observations of some random variable (x_n)
- Density: the chance of a value, i.e., $P(X = 1) = p$.

Multivariate distribution

- Joint probability: $P(X, Y)$
- Conditional probability: $P(Y|X)$
- Chain rule: e.g.
  $P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)$

Lecture Revisited

- Particles are samples. Why particles?
    - The full parameterization of joint distribution may be costly
        - e.g., N discrete random vector, each takes M values, then we need $\mathcal{O}(M^N)$ space to store the density of joint distribution
    - Use samples to represent the probability distribution
        - 1-D example
- Basic setting
    - Progressively solve larger and larger problems
    - Use samples to represent the density estimation for the current partial assignments
    - Each sample has a weight

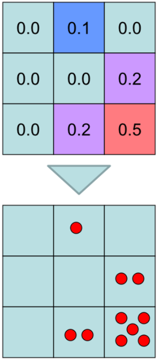
A Simple Particle Filtering Example

Assume we have a Pacman moving in a maze. This maze has radars installed, so that the location of the Pacman can be observed at each time point. However, the observations are contaminated by random noises.

- Q1: Draw the factor graph

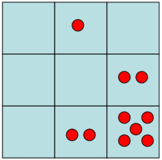
A Simple Particle Filtering Example (cont)

- Assume we have a prior Pacman location at time t_1
- Generate initial samples with equal weights w_i for time t_1



A Simple Particle Filtering Example (cont)

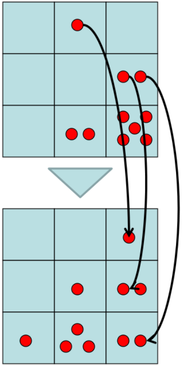
- Assume we have a prior Pacman location at time t_1
- Generate initial samples with equal weights for time t_1



Particles:
(3,3)
(2,3)
(3,3)
(3,2)
(3,3)
(3,2)
(2,1)
(3,3)
(3,3)
(2,1)

At iteration n,

- For each sample x_i^{n-1} in the previous iteration, draw a new sample x_i^n using the proposal distribution $\pi(x_i^{n-1}, x_i^n)$
- Good $\pi(x_i^{n-1}, x_i^n)$ approximates the probability of transition

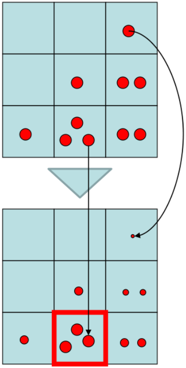


At iteration n,

- For each sample x_i^{n-1} in the previous iteration, draw a new sample x_i^n using the proposal distribution $\pi(x_i^{n-1}, x_i^n)$
- Good $\pi(x_i^{n-1}, x_i^n)$ approximates the probability of transition
- Reweight each sample by

$$w_i \leftarrow \frac{w_i[t(x_i^{n-1}, t_i^n)o(t_i^n, e_i^n)]}{\pi(x_i^{n-1}, x_i^n)}$$

- Normalize w_i by $w_i \leftarrow \frac{w_i}{\sum_i w_i}$



A Simple Gibbs Sampling Example

- A Chain Model