Bayesian Networks and Image Processing

Review: Markov Random Fields

- models **symmetric** relationships, e.g., among pixels
- examples:
Bayesian Networks for Scene Interpretation

- models *asymmetric* relationships
- example: scene interpretation

![Diagram of a Bayesian network showing relationships between 'motorcycle', 'bike', 'exhaust', 'engine', 'wheel', and 'pedals'.]
Bayesian Networks (cont’d)

- Directed Acyclic Graph (DAG) \(\rightarrow\) no feedback loops

- Example (Visit to Asia):

```
Visit to Asia?
\rightarrow
Has Tuberculosis
\rightarrow
Tuberculosis or Cancer
\rightarrow
Positive X-Ray?
```
```
Smoker?
\rightarrow
Has Lung Cancer
\rightarrow
Tuberculosis or Cancer
\rightarrow
Dyspnoea
```
```
Has Bronchitis
\rightarrow
Tuberculosis or Cancer
```

Graph Semantics in Detail

- graph captures **independencies** among the variables

- the independences can be read off from the graph based on some notion of **graph separation**

![Graph](image)

conditional Independence
Graph Semantics in Detail (cont’d)

- this is the interesting case: **v-structure / collider**

![Diagram of v-structure and collider]

- **induced dependences**: when we learn more, previously independent variables can become conditionally dependent

- ’explaining away’: when conditioned on ’Alarm’, evidence concerning ’Earthquake’ affects our belief in ’Burglary’

- **graph separation** criterion for Bayesian networks is more complicated than the one for MRFs.
D-Separation Criterion

In a DAG, two disjoint sets of variables $\mathcal{A}$ and $\mathcal{B}$ are d-separated (conditionally independent) given a third set $S \subseteq \mathcal{V} \setminus (\mathcal{A} \cup \mathcal{B})$, denoted as $\mathcal{A} \perp \mathcal{B} | S$, if and only if along every path between a variable in $\mathcal{A}$ and a variable in $\mathcal{B}$ there is a variable $Z$ satisfying one of the following two conditions:

- $Z$ has converging edges and none of $Z$ or its descendants are in $S$, or
- $Z$ does not have converging edges and $Z \in S$. 

Image Analysis with Statistical Models HS 2007. Buhmann & Einhäuser. 239
D-Separation Criterion: Example I

Are $X$ and $Y$ independent when $Z$ is given?
D-Separation Criterion: Example II

Are $X$ and $Y$ independent when $Z$ is given?
D-Separation Criterion: Example III

\[
\begin{align*}
B \perp D | A, C & \quad A \perp C | B \\
\text{but:} & \quad B \not\perp D | C & \quad A \not\perp C | B, E
\end{align*}
\]
Review: Separation Criteria in MRF

- pairwise Markov Condition
- local Markov Condition
- global Markov Condition
• review: The Markov blanket of a variable $X_i$ contains those variables that render $X_i$ conditionally independent of the “rest”.

• cf. MRF: Markov blanket of $X_i$ is the set of neighbors of $X_i$
Markov Blanket (cont’d)

- Which variables are contained in the Markov blanket of variable \( I \) in the Bayesian network?
Markov Blanket (cont’d)

- Which variables are contained in the Markov blanket of variable $I$ in the Bayesian network?

- The Markov blanket in a Bayesian network contains
  - the parents,
  - the children,
  - AND the parents of the children
Moral Graph of a BN

- Another way to determine the Markov blanket in a Bayesian network: construct the **moral graph** and apply the rule for undirected graphs (MRF).

- A BN can be transformed into a special **undirected** graph called the **moral graph**.

- the moral graph is obtained in two steps
  - marry the parents of each node in the BN
  - then drop the edge orientations
Moral Graph of a BN (cont’d)

- property of the moral graph: it is the minimal undirected graph that reflects all the (conditional) dependences implied by the BN.
Markov-Equivalence

- do different DAGs always imply different cond. independences?

- Example:

  - \[ m_0 \]
    - \( A \) and \( B \) dependent (marginally)
    - \( A \) and \( B \) conditionally independent given \( C \)
  - \[ m_1 \]
    - \( A \) and \( B \) independent
  - \[ m_2 \]
    - \( A \) and \( B \) dependent given \( C \)
  - \[ m_3 \]
    - \( A \) and \( B \) dependent given \( C \)

- DAGs can be partitioned into equivalence classes that represent the same conditional independences.
Markov-Equivalence (cont’d)

• Two DAGs are Markov-equivalent iff they have
  – the same edges when ignoring their orientations
  – and the same v-structures (↘↙)

Example: 2 Markov-equivalent DAGs
Toward a Quantitative Model

- so far: qualitative reasoning $\rightarrow$ conditional (in)dependences

- desirable: a quantitative description of the interrelations among the variables $X_i$ in the graph
Marginal and Conditional Probabilities

- joint probability distribution over $A, B$: $p(A, B)$
- marginal probability of $A$: $p(A) = \sum_B p(A, B)$
- conditional probability ('given that $B$ is known ...'):

$$p(A|B) = \frac{p(A, B)}{p(B)}, \quad p(B) \neq 0$$

- Bayes’ rule:

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}, \quad p(B) \neq 0$$
Marginal and Conditional Independence:

- (marginal) independence: \( p(A, B) = p(A) \cdot p(B) \)
- conditional independence:

\[
A, B \text{cond. indep. given } C \iff p(A, B|C) = p(A|C) \cdot p(B|C)
\]
\[
\iff p(A|B, C) = p(A|C)
\]
Joint Probability Distribution of BN

**Factorization Theorem:** The most general form of the probability distribution that is consistent with the graph factors according to 'node given its parents':

\[
p(X) = \prod_{i=1}^{n} p(X_i | \Pi_i)
\]

where \( \Pi_i \) is the set of parents of \( X_i \); \( n \) is the number of nodes (variables) in the graph.

Note: the parameters of a Bayesian network are **conditional probabilities**

Note: \( p(X) \) is NOT required to be strictly positive for Bayesian networks (this is different from MRFs).
Example of a Bayesian Network

- network structure: directed acyclic graph (DAG)

- parameters: conditional probabilities

- joint distribution:

\[ p(A, B, C, D) = p(D|C, X, A) \cdot p(C|B, X) \cdot p(B|A) \cdot p(A) \]
How to model the conditional distributions?

- discrete variables: tables

\[
\begin{align*}
P(x_2): & \quad \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix} \\
P(x_3): & \quad \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} \\
P(x_1|x_2,x_3): & \quad \begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.1 \\ 0.9 & 0.8 & 0.6 & 0.9 \end{bmatrix}
\end{align*}
\]

- the model parameters are the values in the tables.

- How quickly does the size of the table (i.e., the number of independent parameters) grow with the number of parents?
cf. MRF: Joint Distribution

- graph: undirected edges
- parameters: clique potentials $\Psi(X^C)$
- joint distribution: $p(X) = \frac{1}{Z} \prod_C \Psi(X^C)$ where $X^C$ is a set of variables that make up a clique
- example:

\[
p(A, B, D, F) \propto \Psi(A, B, F) \cdot \Psi(A, B) \cdot \Psi(A, F) \cdot \Psi(B, F) \cdot \Psi(D, F) \cdot \Psi(A) \cdot \Psi(B) \cdot \Psi(D) \cdot \Psi(F)
\]
Graphical Models: General Properties

- combination of graph theory and probability theory

<table>
<thead>
<tr>
<th>graph</th>
<th>probability distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>set of nodes $i = 1, \ldots, n$</td>
<td>set of random variables $X_1, \ldots, X_n$</td>
</tr>
<tr>
<td>edges (between nodes)</td>
<td>conditional (in)dependences (between random variables)</td>
</tr>
<tr>
<td></td>
<td>quantitative description: joint distribution factories</td>
</tr>
</tbody>
</table>

- advantages
  - visualization
  - simple language for describing complex interdependencies
modular description entails computational simplification

- main variants of graphical models:
  - Markov Random Field
  - Bayesian Network
  - Chain Graph
Special Cases: Naive Bayes Network

- graph

\[
\begin{array}{c}
\text{hidden class variable} \\
\downarrow \\
X_1 \quad X_2 \quad X_3 \quad \ldots \quad X_n
\end{array}
\]

- assumption: feature variables \(X_1, \ldots, X_n\) are independent conditional on the class variable

- useful for clustering the feature variables \(X_1, \ldots, X_n\)

- the class variable is hidden

- when graph is given, the parameters are learned from given data (e.g., EM algorithm)
Special Cases: Tree

- example

- graph does not contain loops

- special case of decomposable model

- factorization of joint distribution:

\[
p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | \Pi_i) = \prod_{i=1}^{n} \frac{p(X_i, \Pi_i)}{p(\Pi_i)}
\]
Inference Problems

- example: scene interpretation

- (Given the evidences) what are the marginal probabilities over each of the object variables (here: motorcycle, bike)?
Naive Approach

\[
p(A, B, C, D, E) = p(A)p(B)P(C|A, B)P(D|C)p(E|C)
\]

- goal: when \( E = 1 \), compute marginal distribution \( P(C, E = 1) \)
- summation: \( P(C, E = 1) = \sum_{A, B, D} P(A, B, C, D, E = 1) \)
- exploiting conditional independences:

\[
P(C, E = 1) = \left\{ \sum_{A, B} p(A)p(B)P(C|A, B) \right\} \left\{ \sum_{D} p(D|C) \right\} p(E = 1|C)
\]
Inference Techniques

• exact methods
  – belief propagation (on polytrees)
  – junction tree algorithm (for general DAGs)
  – ...

• approximations
  – sampling
  – loopy belief propagation
  – ...