Graphical Models

Lecture 5:
Template-Based Representations

Andrew McCallum
mccallum@cs.umass.edu

Thanks to Noah Smith and Carlos Guestrin for some slide materials.
Administration

• Homework #3 won’t go out until early March. Push back HW#2 due date?

• Lagrange Multipliers?

• Calendar.
BN with Repeated Structure
Plate Model
“Unrolled” Ground Network

Ground network
Students and their Grades

Example: A = student, B = grade
Student, Course, Grade, Difficulty

Each student takes only one course

Example: $A_1 =$ course difficulty, $A_2 =$ student aptitude for the area, $B =$ grade
Student, Course, Grade, Difficulty

Multiple courses per student

Example: $A_1 = \text{assignment difficulty}, A_2 = \text{intelligence}, B = \text{grade}$
Plate Models: Limitations and Alternatives

• Limitations:
  – can’t have edges between two “copies” of the same variable, (e.g. \textit{position} a time \(t\) depends on \textit{position} at time \(t-1\))
  – can’t have edges between particular pairs selected by some other relation, (e.g. \text{Genotype}(U_1) \text{ depends on Genotype}(U_2)\), where \(U_2\) is mother of \(U_1\).

• Alternatives
  – Dynamic Bayesian Networks (DBNs)
    • Specific to repetitions over time
  – Probabilistic relational models
    • More flexible; see K&F 6.4.2.
Temporal Models

• \( \mathbf{X} \) takes different values at each (discrete) time step.
  – \( \mathbf{X}^{(t)} \) is the random variable at time \( t \)

• Markov Assumption:
  \( \mathbf{X}^{(t+1)} \perp \{ \mathbf{X}^{(0)}, ..., \mathbf{X}^{(t-1)} \} \mid \mathbf{X}^{(t)} \)

• Stationary Assumption (aka \textit{time invariant} or \textit{homogeneous})
  \( P(\mathbf{X}^{(t+1)} \mid \mathbf{X}^{(t)}) \) is the same for all \( t \).

• Can use \textit{conditional Bayesian network} to define
  \( P(\mathbf{X}^{(t+1)} \mid \mathbf{X}^{(t)}) \)
Hidden Markov Model

2-time-slice conditional BN
unrolled or ground Bayesian network
Dynamic Bayesian Network

- Bayesian network over $X^{(0)}$, conditional Bayesian network for $X^{(t+1)}$ given $X^{(t)}$ (2-time-slice)
  - HMM is a special case.
  - Kalman filter (linear dynamical system) is a special case.
Example: DBN for vehicle position

Time slice $0$

Time slice $t$

Time slice $t+1$
Example: DBN for vehicle position

Unrolled over 3 steps

Factor template
Dynamic Bayesian Networks

Factorial HMM

Coupled HMM

Pair HMM
Probabilistic Relational Models

• Contingent Dependency
  – specifies the context in which some dependency holds, with a “guard”—a formula that must hold for the dependency to be applicable.
  – e.g. Location(V) depends on Location(U) contingent on Precedes(U,V)
  – e.g. Genotype(V) depends on Genotype(U) contingent on Mother(U,V)

• Relational Uncertainty (one kind of structural uncertainty)
  – The “guard” predicates are random variables!
Object Uncertainty

- The set of objects is not predetermined.
  - Get list of authors in 100 BibTeX files.
    “Stuart Russell” “Stuart Rusell” “S. Russell”
    How many people are mentioned?

- Introduce
  *person-objects* (represents entity)
  *person-reference objects* (represents mention)
  *refers-to(m,o)* relation

- Model generates (a) # of people, (b) person objects, (c) their reference objects.

[Milch et al “BLOG”]
Directed Factor Graph Notation

[Laura Dietz 2010]
**Variables and Constants**

<table>
<thead>
<tr>
<th>Latent variable / latent parameter</th>
<th>Directed factor graph</th>
<th>Pseudocode</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="symbol" alt="var" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed variable</td>
<td><img src="symbol" alt="obs" /></td>
<td></td>
</tr>
<tr>
<td>Constant / hyper parameter</td>
<td>const</td>
<td></td>
</tr>
</tbody>
</table>
Factors and Densities

<table>
<thead>
<tr>
<th>Factor with one input parameter</th>
<th>Directed factor graph</th>
<th>Pseudocode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image" alt="Factor Diagram" /></td>
<td>1: \texttt{draw} \hspace{5pt} \text{out} \sim \text{Density}(\text{in})</td>
</tr>
</tbody>
</table>

| Example: Gaussian               | ![Gaussian Diagram](image) | 1: \texttt{draw} \hspace{5pt} x \sim \mathcal{N}(\mu, \sigma) |

**Directed factor graph**

- **in**
- **Density**
- **out**

**Pseudocode**

1. \texttt{draw} \hspace{5pt} \text{out} \sim \text{Density}(\text{in})
### Replication with Plates

<table>
<thead>
<tr>
<th>Plate</th>
<th>Directed factor graph</th>
<th>Pseudocode</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
<td><strong>with one</strong></td>
<td><strong>Density</strong> out</td>
</tr>
<tr>
<td>global</td>
<td></td>
<td>var</td>
</tr>
<tr>
<td>$\forall i \in {1..N}$</td>
<td></td>
<td>1: <strong>for all</strong> $\forall i \in {1..N}$ do 2: <strong>draw</strong> $\text{var}_i \sim \text{global}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Example:</strong> repeated Gaussian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$\mu$</td>
<td>$\mathcal{N}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Nested Plates

<table>
<thead>
<tr>
<th>Directed factor graph</th>
<th>Pseudocode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nested plates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1: for all $\forall k \in \{1..K\}$ do
2: for all $\forall i \in \{1..N\}$ do
3: draw $x_{k,i} \sim \mathcal{N}(\mu_k, \sigma)$
Conditioning with *Gates*

Minka & Winn 2008

<table>
<thead>
<tr>
<th>Directed factor graph</th>
<th>Pseudocode</th>
</tr>
</thead>
</table>
| ![Unrolled boolean gate](image) | 1: if $c = 1$ then  
2: `draw x ~ f(\theta_1)`  
3: else  
4: `draw x ~ g(\theta_2)` |
Plates & Gates (and implicit combo)

<table>
<thead>
<tr>
<th></th>
<th>Directed factor graph</th>
<th>Pseudocode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicated gate</td>
<td><img src="image1" alt="Directed factor graph" /></td>
<td>1: <strong>draw</strong> ( x \sim \text{Multi}(\theta_c) )</td>
</tr>
<tr>
<td>Implicit notation for replicating gates</td>
<td><img src="image2" alt="Directed factor graph" /></td>
<td>1: <strong>draw</strong> ( x \sim \text{Multi}(\theta_c) )</td>
</tr>
</tbody>
</table>

Algorithm 1

Generative process of latent Dirichlet allocation.

1: for all \( M \) documents do
2: draw \( \theta \sim \text{Dir}(\alpha) \)
3: for all of the \( N \) words \( w_n \) do
4: draw \( \text{topic } z_n \sim \text{Multi}(\theta) \)
5: draw \( \text{word } w_n \sim \text{Multi}(\beta_{z_n}) \)

For brevity, the plate around the single-box gate may be dropped when the iteration range is obvious. In cases, where the control variable itself is used to index the input variables, the gate condition description may be dropped as well.
Latent Dirichlet Allocation

[Blei, Ng, Jordan]

cf. [7], Figure 1