CS 343: Artificial Intelligence

Bayes Nets: Representation & Independence

Profs. Peter Stone and Yuke Zhu — The University of Texas at Austin

[These slides based on those of Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]
Announcements

- **Past Dues**
  - HW1-4: Search, CSPs, Games, MDPs
  - 8 reading responses: AI100 report; 7 Textbook readings
  - P0, 1, 2: tutorial, Search, Games

- **Upcoming EdX Homeworks**
  - HW5: RL – due Monday 3/22 at 11:59 pm
  - HW6: Bayes Nets – due Monday 4/5 at 11:59 pm

- **Upcoming Programming Projects**
  - P3: RL – due Wednesday 3/31 at 11:59pm

- **Midterm Exam**
  - Thursday 3/25 or so (Materials up to this week)
We’re done with Part I: Search and Planning!

We’ve seen how AI methods can solve problems in:
- Search
- Constraint Satisfaction Problems
- Games
- Markov Decision Problems
- Reinforcement Learning

Next up: Uncertainty and Learning!
CS343 Outline

- We’re done with Part I: Search and Planning!

- Part II: Probabilistic Reasoning
  - Diagnosis
  - Speech recognition
  - Tracking objects
  - Robot mapping
  - Genetics
  - Error correcting codes
  - … lots more!

- Part III: Machine Learning
Probability Recap

- **Conditional probability**
  \[ P(x|y) = \frac{P(x, y)}{P(y)} \quad \text{and} \quad P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

- **Product rule**
  \[ P(x, y) = P(x|y)P(y) \]

- **Chain rule**
  \[ P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \]
  \[ = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \]

- **X, Y independent if and only if:**
  \[ \forall x, y : P(x, y) = P(x)P(y) \quad X \independent Y \]

- **X and Y are conditionally independent given Z if and only if:**
  \[ \forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \quad X \independent Y | Z \]
Probabilistic Models

- Models describe how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - “All models are wrong; but some are useful.”
    – George E. P. Box

- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information
Probabilistic Models

- A probabilistic model is a joint distribution over a set of random variables.

- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called outcomes
  - Joint distributions: say whether assignments (outcomes) are likely
  - Normalized: sum to 1.0
  - Ideally: only certain variables directly interact

- Constraint satisfaction problems:
  - Variables with domains
  - Constraints: state whether assignments are possible
  - Ideally: only certain variables directly interact

### Distribution over $T, W$

<table>
<thead>
<tr>
<th>$T$</th>
<th>$W$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Constraint over $T, W$

<table>
<thead>
<tr>
<th>$T$</th>
<th>$W$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>$T$</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>$F$</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>$F$</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>$T$</td>
</tr>
</tbody>
</table>
Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For about 10 min, we’ll be vague about how these interactions are specified
Example Bayes Net: Insurance
Example Bayes Net: Diagnosis

- **battery age**
  - leads to **battery dead**
- **battery dead**
  - leads to **battery meter**
  - leads to **lights**
  - leads to **oil light**
  - leads to **gas gauge**
- **battery meter**
  - leads to **battery flat**
  - leads to **no oil**
  - leads to **no gas**
  - leads to **car won’t start**
  - leads to **dipstick**
- **alternator broken**
- **fanbelt broken**
- **no charging**
  - leads to **no oil**
  - leads to **no gas**
  - leads to **fuel line blocked**
  - leads to **starter broken**
  - leads to **dipstick**

The diagram represents a Bayesian network for diagnosing car issues, where each node is connected by directed edges indicating causal relationships.
**Graphical Model Notation**

- **Nodes**: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

- **Arcs**: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)

- For now: imagine that arrows mean direct causation (in general, they don’t!)
Example: Alarm Network

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
Bayes Net Semantics
Bayes Net Semantics

- A set of nodes, one per variable $X$
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
    $$P(X|a_1 \ldots a_n)$$
  - CPT: conditional probability table
  - Description of a noisy “causal” process

A Bayes net = Topology (graph) + Local Conditional Probabilities
Probabilities in BNs

- Bayes nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

\[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i)) \]

- Example:

\[ P(+\text{cavity}, +\text{catch}, -\text{toothache}) \]
Probabilities in BNs

- Why are we guaranteed that setting
  \[ P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i)) \]
  results in a proper joint distribution?

- Chain rule (valid for all distributions):
  \[ P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|x_1 \ldots x_{i-1}) \]

- Assume conditional independences:
  \[ P(x_i|x_1, \ldots x_{i-1}) = P(x_i|\text{parents}(X_i)) \]
  \[ \Rightarrow \text{Consequence:} \quad P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i)) \]

- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies
Example: Alarm Network

P(+b, -e, +a, -j, +m) = ?
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) = P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) = 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Example: Traffic

- Causal direction

![Diagram with tables]

<table>
<thead>
<tr>
<th>$P(R)$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>3/4</td>
</tr>
</tbody>
</table>

| $P(T | R)$ |     |
|----------|-----|
| +r       | +t  | 3/4 |
| +r       | -t  | 1/4 |
| -r       | +t  | 1/2 |
| -r       | -t  | 1/2 |

<table>
<thead>
<tr>
<th>$P(T, R)$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
</tr>
</tbody>
</table>
Example: Reverse Traffic

- Reverse causality?

\[
P(T) \\
\begin{array}{c|c}
+t & 9/16 \\
-t & 7/16 \\
\end{array}
\]

\[
P(R|T) \\
\begin{array}{c|c|c}
+t & +r & 1/3 \\
 & -r & 2/3 \\
-t & +r & 1/7 \\
 & -r & 6/7 \\
\end{array}
\]

\[
P(T, R) \\
\begin{array}{c|c|c}
+r & +t & 3/16 \\
+r & -t & 1/16 \\
-r & +t & 6/16 \\
-r & -t & 6/16 \\
\end{array}
\]
Causality?

- **When Bayes nets reflect the true causal patterns:**
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts

- **BNs need not actually be causal**
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables Traffic and Roof Drips
  - End up with arrows that reflect correlation, not causation

- **What do the arrows really mean?**
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence
    \[ P(x_i|x_1, \ldots x_{i-1}) = P(x_i|\text{parents}(X_i)) \]
Bayes Nets

- So far: how a Bayes net encodes a joint distribution
  - Next: how to answer queries about that distribution
    - Today:
      - First assembled BNs using an intuitive notion of conditional independence as causality
      - Then saw that key property is conditional independence
    - Main goal: answer queries about conditional independence and influence

- Thursday: how to answer numerical queries (inference)
Size of a Bayes Net

- How big is a joint distribution over N Boolean variables?
  - $2^N$

- How big is an N-node net if nodes have up to k parents?
  - $O(N \times 2^{k+1})$

- Both give you the power to calculate
  $$P(X_1, X_2, \ldots X_n)$$

- BNs: Huge space savings!

- Also easier to elicit local CPTs

- Also faster to answer queries (coming)
Bayes Nets: Assumptions

- Assumptions we are required to make to define the Bayes net when given the graph:

\[ P(x_i|x_1 \cdots x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

- Beyond above “chain rule \(\Rightarrow\) Bayes net” conditional independence assumptions:
  - Often additional conditional independences
  - They can be inferred from the graph structure

- Important for modeling: understand assumptions made when choosing a Bayes net graph
Conditional independence assumptions directly from simplifications in chain rule:

Standard chain rule: \[ p(x, y, z, w) = p(x)p(y|x)p(z|x, y)p(w|x, y, z) \]

Bayes net: \[ p(x, y, z, w) = p(x)p(y|x)p(z|y)p(w|z) \]

Since: \[ z \perp x \mid y \quad \text{and} \quad w \perp x, y \mid z \quad \text{(cond. indep. given parents)} \]

Additional implied conditional independence assumptions?

\[ p(w|x, y) = \frac{p(w, x, y)}{p(x, y)} = \frac{\sum_{z} p(x)p(y|x)p(z|y)p(w|z)}{p(x)p(y|x)} = \sum_{z} p(z|y)p(w|z) = \sum_{z} p(z|y)p(w|z, y) = \sum_{z} p(z, w|y) = p(w|y) \]
Independent in a BN

- **Important question about a BN:**
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:

  ![Diagram](image)

  - Question: are X and Z necessarily independent?
    - Answer: no. Example: low pressure causes rain, which causes traffic.
    - X can influence Z, Z can influence X (via Y)
    - Addendum: they *could* be independent: how?
Bayes Nets

- Representation
- Conditional Independences
  - Probabilistic Inference
    - Enumeration (exact, exponential complexity)
    - Variable elimination (exact, worst-case exponential complexity, often better)
    - Probabilistic inference is NP-complete
    - Sampling (approximate)
  - Learning Bayes’ Nets from Data