Chapter 28. Bayesian Networks

Lecture Notes on Artificial Intelligence

Summarized by Kim, Byoung-Hee and Lim, Byoung-Kwon

Biointelligence Laboratory
School of Computer Science and Engineering
Seoul National University

http://bi.snu.ac.kr
Contents

28.1 Representing Probabilities in Networks
28.2 Automatic Construction of Bayesian Networks
28.3 Probabilistic Relational Models
28.4 Temporal Bayesian Networks

Appendix
Overview of Chapter 28

- Probabilistic graphical models (PGMs)
  - AI systems need to be able to deal with uncertain information
  - Judea Pearl suggested using graphical structures to encode probabilistic information

- Bayesian Networks
  - The representative of PGMs (another member is MRF)
  - Learning Bayesian networks from data
  - Inference with the established BN

- Two variations of Bayesian Networks
  - Probabilistic relational models
  - Temporal Bayesian networks: HMM, DBN,
Chapter 28. Bayesian Networks

28.1 Representing Probabilities in Networks
Reasoning about Uncertain Beliefs

- Uncertain nature of knowledge
  - Provisional: subject to change
  - Qualified: having various levels of confidence

- Tools of probability theory
  - Much of the knowledge needs to be qualified by probability values

- Difficulties involved in reasoning about uncertain beliefs
  - AI’s old nemesis: the combinatorial explosion
  - Interdependencies among variables results in large representations and intractable computations

- Probabilistic graphical model as a power tool that simplifies both the representational and computational problems.
Reasoning about Uncertain Beliefs

- An example: ‘Car Start’
  - set of four propositions about an automobile engine

- Problems and Cases
  - Two extreme cases: considering whole the joint cases vs. complete independence of propositions
  - “in-betweeness”: realistic problems are in between

- Judea Pearl
  - Developed some of the main representational and computational methods of probabilistic graphical models
  - Key insight
    - beliefs about propositions and other quantities could often be regarded as “direct causes” of other beliefs
    - these causal linkages could be represented in graphical structures that encode simplifying assumptions about relationships among probabilities.
Bayesian (Belief) Networks

- **Components**
  - Graph: node + arrow (influence, probabilistic independence). DAG (directed acyclic graph) structure.
  - Probability: conditional probability table (CPT)

- **Conditional independence**
  - Knowing P1 does not tell us anything new about P4 if we already know P2 and P3 \(\Rightarrow\) CPT for P4 is represented as \(p(P4|P2, P3)\)
  - Taking account of conditional independence reduces the complexity of probabilistic reasoning

Figure 28.2: A network representation
Probabilistic Inference in Bayesian networks

- **Probabilistic inference**
  - Computations about how the probabilities of some nodes in the network are affected by the probabilities of others

- **Three main styles of inference**
  - Causality reasoning
  - "evidential" or "diagnostic" reasoning
  - Explaining away

- **Approximate inference**
  - For larger, more complex BN exact probabilistic inference is computationally intractable
  - Tools from statistics and control engineering are used

(shade means observed)
28.2 Automatic Construction of Bayesian Networks
“Learning” Bayesian Networks

Synopsis

- Bayesian networks can be “learned”
- The learned versions can be used for reasoning

Process overview

- Learning the structure of the network
  - Disposition of nodes and links (example: ...
- Learning network’s CPT
  - Ex) CPT for P4
    - $p(P4 | P2, P3)$,
    - $p(P4 | P2, \neg P3)$,
    - $p(P4 | \neg P2, P3)$,
    - $p(P4 | \neg P2, \neg P3)$.

“sample statistics”
(the greater #samples
The better the estimates)
Chapter 28. Bayesian Networks

28.3 Probabilistic Relational Models
Probabilistic Relational Models

- Elaboration of Bayesian Networks
  - Integrate probability with predicate calculus
  - “template” is used to make different subnetworks

- Example of a PRM
  - A person's blood type and chromosomal information depends on chromosomes inherited from his or her parents

- Construction and usage of PRMs
  - The structure and CPTs of a PRM can either be specified by a designer or learned from data
Applications of Bayesian Networks

- Too many, many applications
- Genomic studies
- Automobile traffic forecasting and routing
- Modeling disease outbreaks
- Guessing at a computer user’s next actions
  - Enable the Windows OS to “prefetch” application data into memory before it is demanded
- Some companies sell knowledge-capturing and reasoning systems based on Bayesian networks

* Importance of massive amounts of data: major theme of modern AI system
Chapter 28. Bayesian Networks

28.4 Temporal Bayesian Networks
Temporal Bayesian Networks

- Static vs. temporal Bayesian Networks
- Hidden Markov models (HMMs)
- Temporal inference tasks
  - Filtering: calculating $p(x_i | y_{1:i})$
  - Prediction: calculating $p(x_{i+t} | y_{1:i}), t > 0$
  - Smoothing: calculating $p(x_{i-t} | y_{1:i}), t > 0$
  - Most-likely-state sequence: $\arg\max_{x_{1:i}} p(x_{1:i} | y_{1:i})$
- Dynamic Bayesian networks
- Markov random fields
Chapter 28. Bayesian Networks

Appendix
Reasoning about Uncertain Beliefs – an Example

**Propositions**

P1: The starter motor is ok.
P2: The starter motor cranks the engine when the starter switch is turned on.
P3: The fuel system is ok.
P4: The car starts when the starter switch is turned on.

**Relations**

P4 depends on P1, P2, P3
P1 and P2 are obviously related

**Full account of dependencies**

<table>
<thead>
<tr>
<th>P1, P2, P3, P4</th>
<th>̈P1, P2, P3, P4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1, P2, P3, ̈P4,</td>
<td>̈P1, P2, P3, P4,</td>
</tr>
<tr>
<td>P1, P2, ̈P3, P4,</td>
<td>̈P1, P2, ̈P3, P4,</td>
</tr>
<tr>
<td>P1, ̈P2, P3, P4,</td>
<td>̈P1, ̈P2, P3, P4,</td>
</tr>
<tr>
<td>P1, ̈P2, ̈P3, P4,</td>
<td>̈P1, ̈P2, ̈P3, P4.</td>
</tr>
</tbody>
</table>

\[ p(P1, P2, P3, P4) = 0.999 \]
\[ p(P1, P2, P3, P4) = 0.0001 \]

... 2^4 - 1 = 15

Probability values should be specified

We can calculate other probabilities, i.e. can do inference

\[ p(P4 \mid P3) \quad p(\neg P4 \mid P1) \]
Reasoning about Uncertain Beliefs – Problems and Cases

Two extreme cases

\[ 2^{30} - 1 = 1,073,741,823 \]

If we have 30 propositions, or ‘binary’ variables, and much more if any of variables has more states than 2

\[ p(P1, P2, P3, P4) = p(P1)p(P2)p(P3)p(P4) \]

Only 4 values are required

More realistic problems are in between these two extremes: “in-betweeness”

⇒ the key to making probabilistic reasoning more tractable

**Judea Pearl** (1936- ; Fig. 28.1) developed some of the main representational and computational methods

“using graphical structures to encode probabilistic information”
Pearl’s seed conjectures

- The consistent agreement between plausible reasoning [by humans] and probability calculus could not be coincidental but strongly suggests that human intuition invokes some crude form of probabilistic computation.

- In light of the speed and effectiveness of human reasoning, the computational difficulties that plagued earlier probabilistic systems could not be very fundamental and should be overcome by making the right choice of simplifying assumptions which humans store in their head.
Probabilistic Inference in Bayesian networks

- **Probabilistic inference**
  - Computations about how the probabilities of some nodes in the network are affected by the probabilities of others
  - In Fig. 28.2, if we know all the CPTs, we can compute all sixteen joint probabilities

- **Three main styles of inference**
  - **Causality reasoning** – “migrating” known probability values downward in the network
  - “evidential” or “diagnostic” reasoning - “migrating” known probability values upward in the network
  - **Explaining away** – the confirmation of one cause of an observed or believed event reduces the need to invoke alternative causes
Explaining Away

In Fig. 28.3

- After observing that the car does not start, the probability that the starter motor is the problem is computed to be 0.023.
- and the probability that the fuel system is the problem is computed to be 0.283.
- upon additionally observing that the starter motor has failed, the probability that the fuel system is the cause of the problem drops by more than half to 0.1.

Figure 28.3: A Bayesian network from an interactive Web site.
Approximate Inference

- For larger, more complex BN
  - Exact probabilistic inference is computationally intractable
  - Simplifications that permit further reductions of parameters
  - Approximate inferences are practical

448 nodes and 908 arrows
Full CPT - 133,931,430 probabilities with simplifications
Reducing to 8,254 values

Figure 28.4: A large medical Bayesian network.
Structure Learning of Bayesian Networks

- Start with some basic candidate structure, such as one that has no connections between nodes, and use the data collection to estimate its CPTs.
- Calculate a “goodness measure” for this network.
- Begin a “hill-climbing” search process by evaluating “nearby” networks that differ from the previous one by small changes.
  - Settle on that changed network with the best improvement in goodness.
- Continue the hill-climbing process until no more improvements can be made.

* Sometimes, installing “hidden nodes”, which are attributes that do not occur in the data set, results in simplified structure.
Alarm network: 37 variables for a problem involving an alarm system in a hospital intensive-care unit.

Random sample generation (10,000 samples) \(\rightarrow\) training set for the no-edge network.

Learned network
Note the very close similarity in structure.

Figure 28.5: Learning a Bayesian network.
Probabilistic Relational Models (PRMs)

- **General Concept**
  - Inventors: Daphne Koller, Avi Pfeffer, Lise Getoor, Nir Friedman
  - Elaboration of Bayesian networks
    - Integrate probability with predicate calculus
    - Exploiting the fact that some nodes might share the same attributes except for the values of variables internal to those attributes
  - “template” is used to make different subnetworks whose attribute variables are instantiated to different individuals
Repeated subnetworks

- A person's blood type and chromosomal information depends on chromosomes inherited from his or her parents

Using PRMs makes the design of Bayesian networks much more efficient than would be the process of having to design each (only slightly different) subnetwork separately
Construction and Usage of PRMs

- The structure and CPTs of a PRM can either be specified by a designer or learned from data

- **Added benefit**
  - Objects resulting from instantiating template variables can be linked in the resulting Bayesian network
  - Relationships among these objects can be specified by hand or learned

- **Probabilistic inference** procedures can be used to answer queries (as with any Bayesian networks)
Static vs. Temporal Networks

- **Static vs. temporal**
  - Static Bayesian networks: propositions and quantities represented by the nodes and CPTs are timeless
  - Temporal probabilistic networks involve quantities at different times (ex) hidden Markov models (HMMs, Section 17.3.2)

![Figure 17.3: Two hierarchical levels in speech generation](image)
Hidden Markov Models

- **Elements**
  - State variable $x_i$. Values are unknown (hidden)
  - Observable variable $y_i$. Can be measured and thus known
  - Emission/transition probability matrix: probabilities of the observables/the next state given the value of a state

- **Example**
  - $x_i$: airport weather. foggy or not
  - $y_i$: trasmitter signal: observed
  - $p(x_i = a|x_{i-1} = a) = 0.75, a = 0$
  - $p(y_i = a|x_i = b) = 0.05, a \neq b$

Figure 28.8: A hidden Markov model.
Temporal Inference Tasks

- **Filtering**: calculating $p(x_i | y_{1:i})$
  - Ex) Prob. That the landing strip is foggy right now based on a sequence of previous observations up to and including the present one

- **Prediction**: calculating $p(x_{i+t} | y_{1:i}), t > 0$

- **Smoothing**: calculating $p(x_{i-t} | y_{1:i}), t > 0$

- **Most-likely-state sequence**: $\arg\max_{x_{1:i}} p(x_{1:i} | y_{1:i})$

- **Inference algorithms**
  - All the above can be performed using Bayesian network inference procedures
  - Forward-backward algorithm, the Viterbi algorithm, Kalman filtering
Dynamic Bayesian networks

- More general framework that includes HMM as a special case
  - More state variables at every time instant
  - All of which affect each other, the observations, and subsequent state variables

- Computation in DBNs
  - Exact computations are intractable
  - Approximate methods are required, i.e. particle filtering
  - Learning DBNs is possible from databases containing information about temporal processes

- Applications
  - Recognizing and tracking moving objects
  - Certification of collision avoidance systems for aircraft
Markov Random Fields

- **Markov random field (MRF) or Markov networks**
  - A class of probabilistic graphical models
  - Links between nodes are non-directional
  - Originally developed to deal with problems in statistical physics
  - **Boltzmann machines** (Section 25.5.1) are instances of MRFs

- **Applications of MRFs**
  - Image processing
  - Sensory perception
  - Brain modeling