MTAT.05.113 Bayesian Networks

Construction of Bayesian Networks

Liina Kamm
Konstantin Tretjakov
University of Tartu
{kamm,kt}@ut.ee

Overview

- Causal Networks and reasoning under uncertainty
- Bayesian networks as causal networks
- Chain rule for Bayesian networks
- Considerations for determining the structure of a Bayesian network model
- Estimation of conditional probabilities and modeling methods

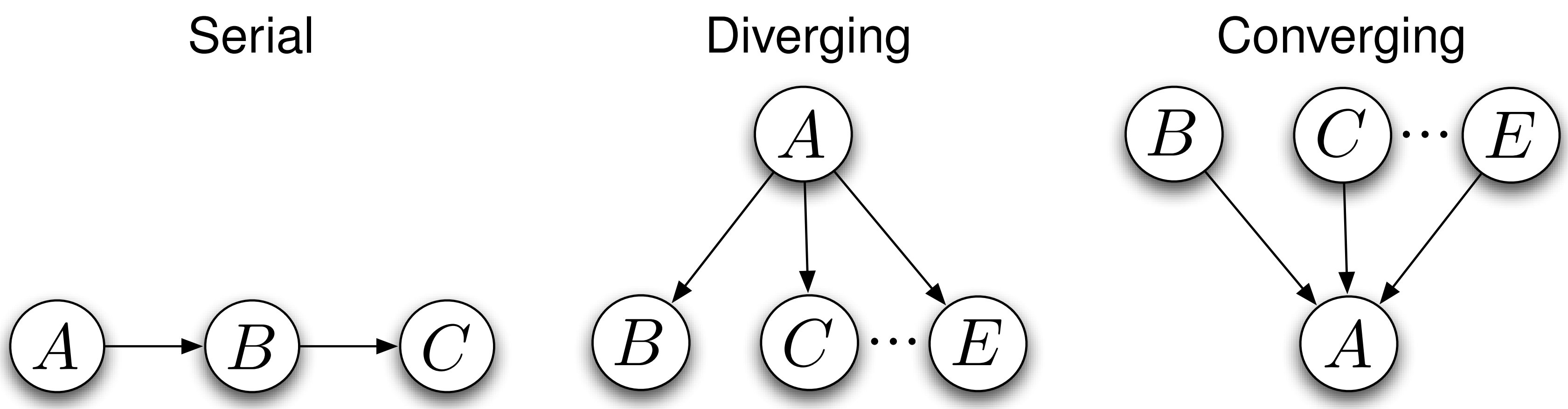
Reasoning Under Uncertainty

- Car start problem
- Logical reasoning
- Logical reasoning with certainties attached
- Combining certainties

Causal Networks

- Variables
- Directed links (arcs)
- Outcomes (states)
- Causal impact
- Direction of impact

Connections



- Instantiation, d-separation
- Evidence may be transmitted through a serial or converging connection unless the state of the variable in the connection is instantiated.

Converging Connections

- If nothing is known about the consequences, the parents are independent—evidence of one of them cannot influence the certainties of others.

- If anything is known about the consequences, information on one possible cause may tell us something about the other causes.

- Conclusion: Evidence may be transmitted through a converging connection only if either the variable in the connection or one of its descendants has evidence.

- Hard and soft evidence

d-Separation

Definition 1. *Two distinct variables A and B in a causal network are d-separated if for all paths between A and B there is an intermediate variable V (distinct from A and B) such that either*

- *the connection is serial or diverging and V is instantiated or*
- *the connection is converging, and neither V nor any of the descendants of V have received evidence.*

Definition 2. *The Markov blanket of a variable A is the set consisting of the parents of A, the children of A, and the variables sharing a child with A.*

When a Markov blanket is instantiated, then A is d-separated from the rest of the network.

How to Test d-Separation?

- Test whether A and B are d-separated given hard evidence on a set of variables $\mathcal{C}$.

1. Construct the *ancestral graph* consisting of A, B and $\mathcal{C}$ together with all nodes from which there is a directed path to either A, B or $\mathcal{C}$.

2. Insert an undirected link between each pair of nodes with a common child.

3. Make all links undirected obtaining a *moral graph* of the ancestral graph.

4. If all paths connecting A and B intersect $\mathcal{C}$ in the moral graph, then A and B are d-separated given $\mathcal{C}$.

Bayesian Networks

Definition 3. *A Bayesian network consists of the following:*

- *A set of variables and a set of directed edges between variables.*
- *Each variable has a finite set of mutually exclusive states.*
- *The variables together with the directed edges form an acyclic directed graph (DAG).*
- *To each variable* A *with parents* B_1, ..., B_n, *a conditional probability table* $P(A|B_1, \ldots, B_n)$ *is attached.*

The Chain Rule for Bayesian Networks

Theorem 1. *Let BN be a Bayesian network over $\mathcal{U} = \{A_1, \ldots, A_n\}$. Then BN specifies a unique joint probability distribution $P(\mathcal{U})$ given by the product of all conditional probability tables specified in BN:*

$$P(\mathcal{U}) = \prod_{i=1}^{n} P(A_i | pa(A_i)) \quad ,$$

where $pa(A_i)$ are the parents of A_i in BN, and $P(\mathcal{U})$ reflects the properties of BN.

If the variables A and B are d-separated in BN given the set $\mathcal{C}$, then A and B are independent given $\mathcal{C}$ in $P(\mathcal{U})$.

Inserting Evidence

Definition 4. *Let A be a variable with n states. A finding $\mathbf{e}$ on A is an n-dimensional table of zeros and ones.*

Theorem 2. *Let BN be a Bayesian network over the universe $\mathcal{U}$, and let $\mathbf{e}_1, \ldots, \mathbf{e}_m$ be findings. Then*

$$P(\mathcal{U}, e) = \prod_{A \in \mathcal{U}} P(A|pa(A)) \cdot \prod_{i=1}^{m} P(\mathbf{e}_i) \ ,$$

and for $A \in \mathcal{U}$ we have

$$P(A|e) = \frac{\sum_{\mathcal{U} \setminus \{A\}} P(\mathcal{U}, e)}{P(e)} \ .$$

Calculating Probabilities in Practice

- Variable elimination: We have a set $\mathcal{T}$ of tables, we want to marginalize variable X.

  1. Take from $\mathcal{T}$ all tables with X in their domains.
  2. Calculate the product.
  3. Marginalize X out.
  4. Place the resulting table in $\mathcal{T}$.

Building a Bayesian Network

- Hypothesis events—events whose probabilities are not directly observable.
- Hypothesis variables—sets of mutually exclusive events that hypothesis events are grouped into.
- Information variables—variables that group achievable information that may reveal something about the hypothesis variables.
- Directed links for a causal network.

Milk Test

- Hypothesis events: milk infected, milk not infected
- Hypothesis variables: infected? (yes, no)
- Information variables: test results (positive, negative)
- Directed links? Past knowledge?
- Hidden assumptions from d-separation properties
  - ⋆ Markov property: if we know the present, then the past has no influence on the future.
  - ⋆ Test nodes are d-separated given any infection node.

Naive Bayes Models

- The information variables are assumed to be independent given the hypothesis variable.
- Using this assumption, it is easy to calculate the conditional probability distribution for the hypothesis variable, given the information variables.
- In some areas (eg diagnosing), it has been shown to provide very good performance, even when the independence assumption is violated.
- If the conditional probability distribution does not change which state has the highest probability, then the naive Bayes model can be used without affecting the performance of the system.

Milk Test. Now with Probabilities!

- $P(Test|Infected?)$: false positives $P(Test = pos|Infected? = no)$, false negatives $P(Test = neg|Infected? = yes)$. Let both be 0.01.

- An estimate for the prior probability $P(Infected?)$ is the daily frequency λ of infected milk for each cow at a particular farm. 50 cows, milk infected one day a month, the probability that all the cows are clean is therefore $29/30$

- If we look at the outbreaks as independent, we also get that the probability of all 50 cows being clean on a given day is $(1-\lambda)^{50}$, and we get

$$(1-\lambda)^{50} = \frac{29}{30} \Rightarrow \lambda = 1 - \left(\frac{29}{30}\right)^{0.02} \approx 0.0007$$

Modeling Methods

Undirected Relations

- Problem: A relation has no direction attached, eg logical constraints.
- A configuration $R(A, B, C)$, where $R(A, B, C) = 1$ for all valid configurations of A, B and C.
- Add a new common child D (*constraint variable*) with two states y and n.
- The deterministic conditional probability table of D is given as $P(D = y|A, B, C) = R(A, B, C)$.
- Finally, enter the evidence $D = y$, forcing the relation/constraint to hold.

Noisy-Or

- Problem: A variable A has several parents, you must specify $P(A|\mathbf{c})$ for each configuration $\mathbf{c}$ of the parents. If you take the configurations from a database, they may be too specific for any expert.

- Let $A_1, \ldots, A_n$ be binary variables listing all the causes of the variable B. Each event $A_i = y$ causes $B = y$ unless an inhibitor prevents it, and the probability of the inhibitor is q_i ie $P(B = n|A_i = y) = q_i$.

- Then

$$P(B = n|A_1, A_2, \ldots, A_n) = \prod_{j \in Y} q_j \ ,$$

where Y is the set of indices for variables in the state y.

Divorcing

- Problem: There are too many parents to one variable with very different states (eg granting a loan).

- Let A_1, ..., A_n be parents of B. The set of parents A_1, ..., A_i is divorced from A_{i+1}, ..., A_n by introducing a mediating variable C, making C a child of A_1, ..., A_i and a parent of B.

- Noisy functional dependence, eg add the states by transforming them to numbers, adding and transforming back to states again.

Expert Disagreements

- Problem: Experts disagree on the conditional probabilities for a model.
- Take the mean or calculate the weighted average.
- Represent the experts in the model by adding a variable S with states for each expert. The variable has a link to the nodes, about whose tables the experts disagree.
- The model has now been prepared for adaptation—you can get an updated indication on which expert to believe.

Object Oriented Bayesian Networks 1

- Problem: Complex Bayesian networks often include copies of almost identical network fragments.
- Network fragment
  - ⋆ The conditional probability tables for the nodes in the fragment are the same for each fragment.
  - ⋆ The state spaces of the variables that are parents to corresponding objects in the fragments are the same.
- Construct a generic network fragment (*class*) that can be instantiated the required number of times (*object*).
- Input and output attributes (interface) and encapsulated attributes

Object Oriented Bayesian Networks 2

- Top down construction of object oriented Bayesian networks
- Subclasses and inheritance
- Transforming the OOBN into a BN

Dynamic Bayesian Networks 1

- Problem: Domains that evolve over time.
- Time slice, temporal links
- If the structures of the time slices and the conditional probabilities are identical and the temporal links are the same, we call the model a dynamic Bayesian network model.
- Hidden Markov model—dynamic Bayesian network model with the Markov property.
- Kalman filter—hidden Markov model where exactly one variable has relatives outside the time slice.

Dynamic Bayesian Networks 2

- Do the following:
  - ⋆ Specify the structure of a time slice.
  - ⋆ Specify the number of time slices.
  - ⋆ Specify the temporal links.
- Often yields calculational problems.

Dealing with Continuous Variables

- Problem: The states of a variable are continuous, we cannot use a conditional probability table.

- Specify a density function for each combination of states for the parent variables of the continuous variable.

- Constraints for hybrid Bayesian networks:

  ⋆ Each continuous variable is assigned a (linear) conditional Gaussian distribution.

  ⋆ No discrete variable has continuous parents.

Interventions

- Problem: You need to incorporate actions that change the state of some variables.
- Extend the model with a special variable.
- Introduce new nodes for the variables that may change state.
- Nonpersistent nodes are the descendants of the nodes affected by the intervention.

Thank You!