Foundations of Artificial Intelligence

47. Uncertainty: Representation

Malte Helmert and Gabriele Röger

University of Basel

May 24, 2017

Foundations of Artificial Intelligence

May 24, 2017 — 47. Uncertainty: Representation

47.1 Introduction

47.2 Conditional Independence

47.3 Bayesian Networks

47.4 Summary

Uncertainty: Overview

chapter overview:

- 46. Introduction and Quantification
- 47. Representation of Uncertainty

47.1 Introduction

Running Example

We continue the dentist example.

	toothache		$\neg$*toothache*	
	catch	$\neg$catch	catch	$\neg$catch
cavity	0.108	0.012	0.072	0.008
$\neg$cavity	0.016	0.064	0.144	0.576

Full Joint Probability Distribution: Discussion

Advantage: Contains all necessary information

Disadvantage: Prohibitively large in practice:
Table for n Boolean variables has size $O(2^n)$.

Good for theoretical foundations, but what to do in practice?

47.2 Conditional Independence

Reminder: Bayes' Rule

General version with multivalued variables and conditioned on some background evidence $\mathbf{e}$:

$$\mathbf{P}(Y \mid X, \mathbf{e}) = \frac{\mathbf{P}(X \mid Y, \mathbf{e})\mathbf{P}(Y \mid \mathbf{e})}{\mathbf{P}(X \mid \mathbf{e})}$$

Multiple Evidence

If we already know that the probe catches and the tooth aches, we could compute the probability that this patient has cavity from

$$\mathbf{P}(Cavity \mid catch \land toothache) = \alpha \mathbf{P}(catch \land toothache \mid Cavity)\mathbf{P}(Cavity).$$

Problem: Need conditional probability for $catch \land toothache$ for each value of $Cavity$.
$\rightsquigarrow$ same scalability problem as with full joint distribution

Conditional Independence: Example

	toothache		*¬toothache*	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Variables *Toothache* and *Catch* not independent but independent given the presence or absence of cavity:

$$\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\mathbf{P}(Catch \mid Cavity)$$

Conditional Independence

Definition

Two variables X and Y are conditionally independent given a third variable Z if

$$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z).$$

Conditional Independence and Multiple Evidence Example

Multiple evidence:

$$\begin{aligned}\mathbf{P}(Cavity \mid catch \land toothache) &= \alpha \mathbf{P}(catch \land toothache \mid Cavity)\mathbf{P}(Cavity)\\ &= \alpha \mathbf{P}(toothache \mid Cavity)\mathbf{P}(catch \mid Cavity)\mathbf{P}(Cavity).\end{aligned}$$

$\rightsquigarrow$ No need for conditional joint probabilities for conjunctions

Conditional Independence: Decomposition of Joint Dist.

Full joint distribution:

$$\begin{aligned}&\mathbf{P}(Toothache, Catch, Cavity)\\&\quad = \mathbf{P}(Toothache, Catch \mid Cavity)\mathbf{P}(Cavity)\\&\quad = \mathbf{P}(Toothache \mid Cavity)\mathbf{P}(Catch \mid Cavity)\mathbf{P}(Cavity)\end{aligned}$$

$\rightsquigarrow$ Large table can be decomposed into three smaller tables.

For n symptoms that are all conditionally independent given *Cavity* the representation grows as $O(n)$ instead of $O(2^n)$.

47.3 Bayesian Networks

Bayesian Networks

Definition

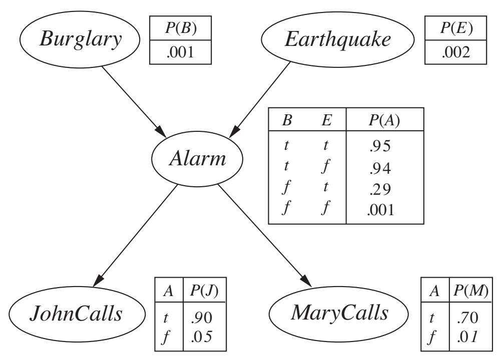
A Bayesian network is a directed acyclic graph, where

- each node corresponds to a random variable,
- each node X has an associated conditional probability distribution $\mathbf{P}(X \mid parents(X))$ that quantifies the effect of the parents on the node.

Bayesian networks are also called belief networks or probabilistic networks.

They are a subclass of graphical models.

Bayesian Network: Example



Semantics

The semantics for Bayesian networks expresses that

- the information associated to each node represents a conditional probability distribution, and that
- each variable is conditionally independent of its non-descendants given its parents.

Definition

A Bayesian network with nodes $\{X_1, \dots, X_n\}$ represents the full joint probability given by

$$P(X_1 = x_1 \wedge \dots \wedge X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid parents(X_i)).$$

Naive Construction

Order all variables, e.g.. as $X_1, \dots, X_n$.

For $i = 1$ to n do:

- Choose from $X_1, \dots, X_{i-1}$ a minimal set of parents of X_i such that $P(X_i \mid X_{i-1}, \dots, X_1) = P(X_i = x_i \mid parents(X_i))$.
- For each parent insert a link from the parent to X_i.
- Define conditional probability table $\mathbf{P}(X_i \mid parents(X_i))$.

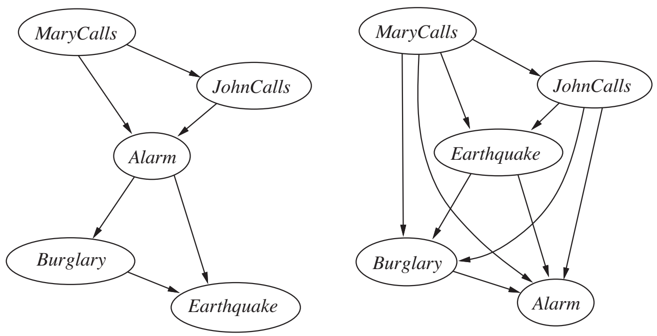
Compactness

Compactness of Bayesian networks stems from local structures in domains, where random variables are directly influenced only by a small number of variables.

- n Boolean random variables
- each variable directly influenced by at most k others
- full joint probability distribution contains 2^n numbers
- Bayesian network can be specified by $n2^k$ numbers

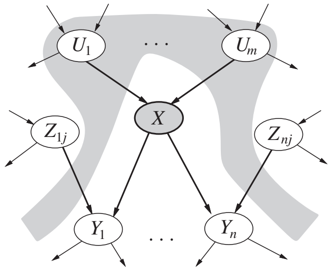
Influence of Node Ordering

A bad node ordering can lead to large numbers of parents and probabiliy distributions that are hard to specify.



Conditional Independence Given Parents

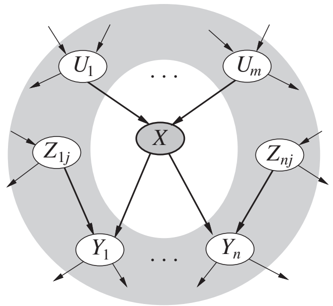
Each variable is conditionally independent of its non-descendants given its parents.



X is conditionally independent of the nodes Z_{ij} given $U_1 \dots U_m$.

Conditional Independence Given Markov Blanket

The Markov blanket of a node consists of its parents, children and children's other parents.



Each variable is conditionally independent of all other nodes in the network given its Markov blanket (gray area).

47.4 Summary

Summary & Outlook

Summary

- Conditional independence is weaker than (unconditional) independence but occurs more frequently.
- Bayesian networks exploit conditional independence to compactly represent joint probability distributions.

Outlook

- There are exact and approximate inference algorithms for Bayesian networks.
- Exact inference in Bayesian networks is NP-hard (but tractable for some sub-classes such as poly-trees).
- All concepts can be extended to continuous random variables.