Problem Solving with AI Techniques
Bayesian Networks

Paul Weng

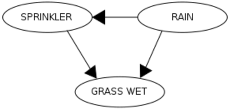
UM-SJTU Joint Institute

VE593, Fall 2018

Bayesian Network: Definition

- Bayesian Network: directed acyclic graph (DAG) where
  - a node represents a random variable X
  - a directed edge $\sim$ "directly influences"
  - each node has a conditional probability distribution $P(X \mid Parents(X))$
- For discrete random variables with finite support, the conditional distribution is represented as a conditional probability table (CPT)

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



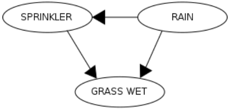
RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Bayesian Network: Definition

- Bayesian Network: directed acyclic graph (DAG) where
  - a node represents a random variable X
  - a directed edge $\sim$ "directly influences"
  - each node has a conditional probability distribution $P(X \mid Parents(X))$
- For discrete random variables with finite support, the conditional distribution is represented as a conditional probability table (CPT)

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

- How about continuous variables?

Examples of Continuous Cases

Toy examples:

- Gaussian variables
- Discrete variables and Gaussian variables

Some known cases:

- Unknown parameter of a Bernoulli distribution
- Unknown parameter of Categorical distribution
- Mixture model
- Linear regression

Bayesian Network: Implications

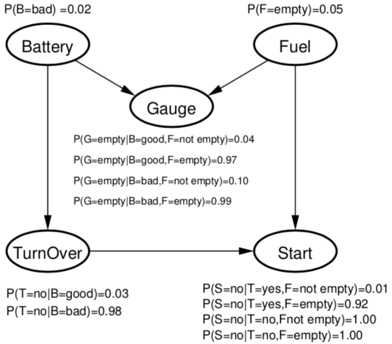
- DAG: nodes can be sorted such that edges only go from lower to higher indices (topological order)
- Bayesian network provides a factorization of the joint distribution:

$$P(X_{1:n}) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

- Missing links imply conditional independence
- Generative model: Bayesian network provides a way to generate samples (from ancestors to children)

Quiz: Car Diagnostic

(Heckermann 1995)



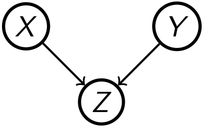
- Given this Bayes net, how can $P(S, T, G, F, B)$ be written?
- How many values are needed if stored as a whole table?
- How many values are needed if stored as this Bayes net?

What can we do with Bayes Nets?

- Conditional Independence: Given Z, are X and Y independent?
- Inference: Given some pieces of information (e.g., prior, observed variables), what is the implication (e.g., posterior) on a non-observed variables?
- Decision-making: If utilities are also provided, how to make optimal decisions (w.r.t. expected utility)?
- Learning:
  - Fully Bayesian learning: inference over parameters
  - Maximum likelihood training
- Structure learning: Find the Bayes net model (i.e., graph structure) fits the data best; thereby uncovering conditional independencies in the data

Bayes Networks & Conditional Independence

- Independence $X \perp\!\!\!\perp Y \Leftrightarrow P(X, Y) = P(X)P(Y)$
- Conditional Independence
  $X \perp\!\!\!\perp Y \mid Z \Leftrightarrow P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$

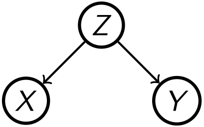


(convergent)

$X \perp\!\!\!\perp Y$

$X \not\perp\!\!\!\perp Y \mid Z$

Common effect

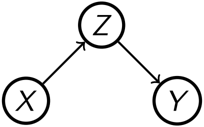


(divergent)

$X \not\perp\!\!\!\perp Y$

$X \perp\!\!\!\perp Y \mid Z$

Common cause



(serial)

$X \not\perp\!\!\!\perp Y$

$X \perp\!\!\!\perp Y \mid Z$

Intermediate cause

Proofs

- Convergent: $X \perp\!\!\!\perp Y$

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

- Divergent: $X \perp\!\!\!\perp Y \mid Z$

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

- Serial: $X \perp\!\!\!\perp Y \mid Z$

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

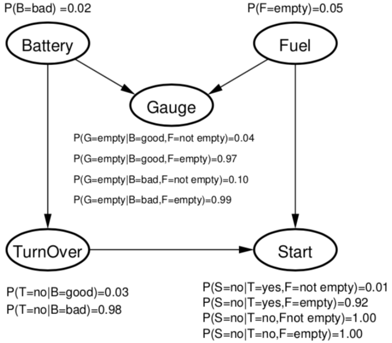
D-Separation: Rule for Determining Conditional Independence

D-separation: Given three groups of random variables X, Y and Z

- $X \perp\!\!\!\perp Y \mid Z \Leftrightarrow$ every path from X to Y is "blocked" by Z
- A path is "blocked" by $Z \Leftrightarrow$ on this path...
    - $\exists$ a node in Z that is divergent w.r.t. the path, or
    - $\exists$ a node in Z that is serial w.r.t. the path, or
    - $\exists$ a node A that is convergent w.r.t. the path and neither A nor any of its descendants are in Z

Example: Car Diagnostic (contd.)

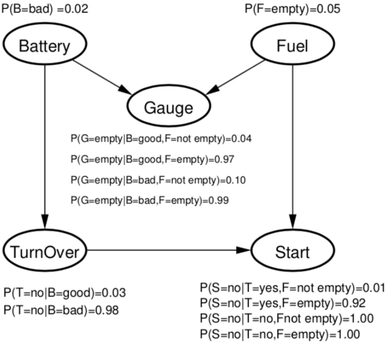
(Heckermann 1995)



$T \perp\!\!\!\perp F$?

Example: Car Diagnostic (contd.)

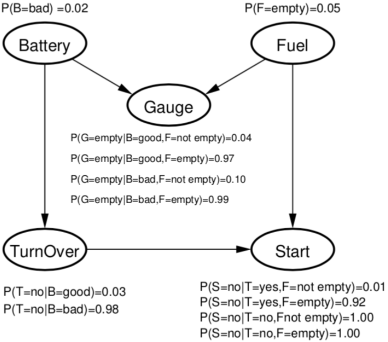
(Heckermann 1995)



$T \perp\!\!\!\perp F$? Yes

Example: Car Diagnostic (contd.)

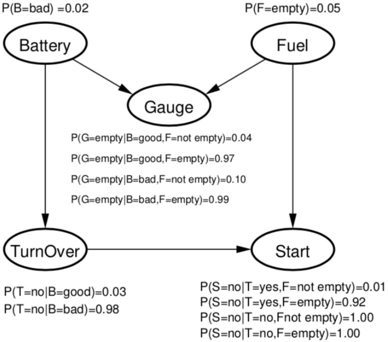
(Heckermann 1995)



$T \perp\!\!\!\perp F$? Yes $\quad B \perp\!\!\!\perp F \mid S$?

Example: Car Diagnostic (contd.)

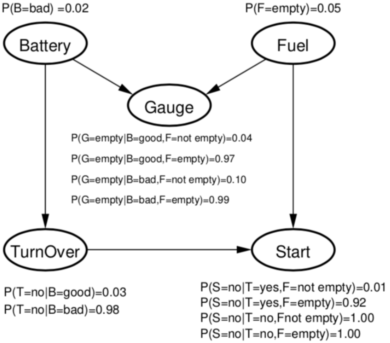
(Heckermann 1995)



$T \perp\!\!\!\perp F$? Yes $\quad B \perp\!\!\!\perp F \mid S$? No

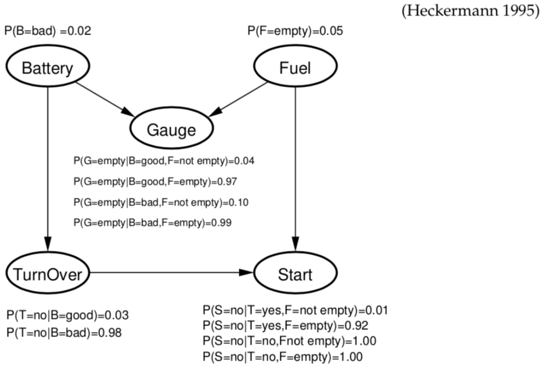
Example: Car Diagnostic (contd.)

(Heckermann 1995)



$T \perp\!\!\!\perp F$? Yes $\quad B \perp\!\!\!\perp F \mid S$? No $\quad B \perp\!\!\!\perp S \mid T$?

Example: Car Diagnostic (contd.)



$T \perp\!\!\!\perp F$? Yes $\quad B \perp\!\!\!\perp F \mid S$? No $\quad B \perp\!\!\!\perp S \mid T$? Yes

Implications of D-Separation

- A node is conditionally independent to its non-descendants given its parents

- Markov blanket of a node: its parents, its children and "spouses" (i.e, parents of common children)
- A node is conditionally independent of ***all*** other nodes given its Markov blanket

How to Check for D-Separation?

More operational definition of d-separation:

- $X \perp\!\!\!\perp Y \mid Z \Leftrightarrow$ there is no path from X to Y in the undirected ancestral moral graph with Z removed
- Algorithm:
  - Ancestral graph: keep only X, Y, Z and their ancestors
  - Moral graph: add edge between all pairs of parents if not connected
  - Undirected graph: remove direction of all edges
  - Remove nodes from Z
  - Check if there is a path from X to Y