Problem Solving with AI Techniques
Reasoning under Uncertainty

Paul Weng

UM-SJTU Joint Institute

VE593, Fall 2018

1. Reasoning under Uncertainty
2. Example
3. Graphical Models
4. Applications

Why Reasoning under Uncertainty?

- Previous search algorithms requires full observability of states and generally deterministic actions

Why Reasoning under Uncertainty?

- Previous search algorithms requires full observability of states and generally deterministic actions

- Knowledge is generally uncertain:

Why Reasoning under Uncertainty?

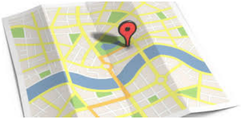
- Previous search algorithms requires full observability of states and generally deterministic actions

- Knowledge is generally uncertain:
    - Partial/non observable world

Why Reasoning under Uncertainty?

- Previous search algorithms requires full observability of states and generally deterministic actions

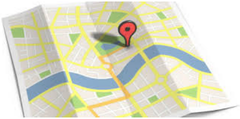
- Knowledge is generally uncertain:
  - Partial/non observable world
  - Observation comes from imperfect sensors



Why Reasoning under Uncertainty?

- Previous search algorithms requires full observability of states and generally deterministic actions

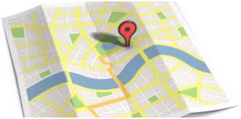
- Knowledge is generally uncertain:
    - Partial/non observable world
    - Observation comes from imperfect sensors
    - Data often imprecise, missing or contradictory



Why Reasoning under Uncertainty?

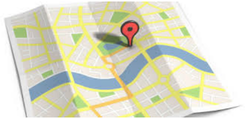
- Previous search algorithms requires full observability of states and generally deterministic actions

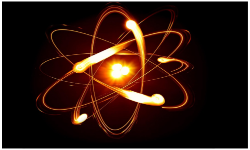
- Knowledge is generally uncertain:
  - Partial/non observable world
  - Observation comes from imperfect sensors
  - Data often imprecise, missing or contradictory
  - World is stochastic?



Why Reasoning under Uncertainty?

- Previous search algorithms requires full observability of states and generally deterministic actions

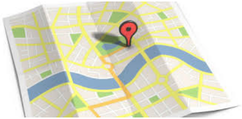


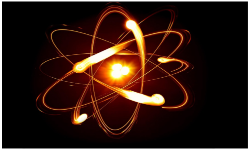


- Knowledge is generally uncertain:
  - Partial/non observable world
  - Observation comes from imperfect sensors
  - Data often imprecise, missing or contradictory
  - World is stochastic?
- Uncertainty can be handled with probability distribution

Why Reasoning under Uncertainty?

- Previous search algorithms requires full observability of states and generally deterministic actions





- Knowledge is generally uncertain:
  - Partial/non observable world
  - Observation comes from imperfect sensors
  - Data often imprecise, missing or contradictory
  - World is stochastic?
- Uncertainty can be handled with probability distribution
- Previous search algos are therefore not directly applicable anymore

What is Reasoning under Uncertainty?

- A state is an assignment of $X_{1:n} = (X_1, \ldots, X_n)$
- Probability distribution $\mathbb{P}(X_{1:n})$ represents uncertainty about possible states
- Given $\mathbb{P}$, we can obtain the probability of any event

What is Reasoning under Uncertainty?

- A state is an assignment of $X_{1:n} = (X_1, \ldots, X_n)$
- Probability distribution $\mathbb{P}(X_{1:n})$ represents uncertainty about possible states
- Given $\mathbb{P}$, we can obtain the probability of any event
- We can also compute the conditional probability of any event, i.e., we can update our belief after some observation

What is Reasoning under Uncertainty?

- A state is an assignment of $X_{1:n} = (X_1, \ldots, X_n)$
- Probability distribution $\mathbb{P}(X_{1:n})$ represents uncertainty about possible states
- Given $\mathbb{P}$, we can obtain the probability of any event
- We can also compute the conditional probability of any event, i.e., we can update our belief after some observation
    - Diagnostic: Given some effect, what are the probable cause?

What is Reasoning under Uncertainty?

- A state is an assignment of $X_{1:n} = (X_1, \dots, X_n)$
- Probability distribution $\mathbb{P}(X_{1:n})$ represents uncertainty about possible states
- Given $\mathbb{P}$, we can obtain the probability of any event
- We can also compute the conditional probability of any event, i.e., we can update our belief after some observation
    - Diagnostic: Given some effect, what are the probable cause?
    - Prediction: Given some cause, what are the probable consequence?

What is Reasoning under Uncertainty?

- A state is an assignment of $X_{1:n} = (X_1, \ldots, X_n)$
- Probability distribution $\mathbb{P}(X_{1:n})$ represents uncertainty about possible states
- Given $\mathbb{P}$, we can obtain the probability of any event
- We can also compute the conditional probability of any event, i.e., we can update our belief after some observation
  - Diagnostic: Given some effect, what are the probable cause?
  - Prediction: Given some cause, what are the probable consequence?
  - Intercausal: Given some cause & effect, what about other causes?

What is Reasoning under Uncertainty?

- A state is an assignment of $X_{1:n} = (X_1, \ldots, X_n)$
- Probability distribution $\mathbb{P}(X_{1:n})$ represents uncertainty about possible states
- Given $\mathbb{P}$, we can obtain the probability of any event
- We can also compute the conditional probability of any event, i.e., we can update our belief after some observation
  - Diagnostic: Given some effect, what are the probable cause?
  - Prediction: Given some cause, what are the probable consequence?
  - Intercausal: Given some cause & effect, what about other causes?
  - Mixed: Combination of two or more of the above

What is Reasoning under Uncertainty?

- A state is an assignment of $X_{1:n} = (X_1, \ldots, X_n)$
- Probability distribution $\mathbb{P}(X_{1:n})$ represents uncertainty about possible states
- Given $\mathbb{P}$, we can obtain the probability of any event
- We can also compute the conditional probability of any event, i.e., we can update our belief after some observation
  - Diagnostic: Given some effect, what are the probable cause?
  - Prediction: Given some cause, what are the probable consequence?
  - Intercausal: Given some cause & effect, what about other causes?
  - Mixed: Combination of two or more of the above
- Issue: memory requirement to store $\mathbb{P}$

1. Reasoning under Uncertainty
2. Example
3. Graphical Models
4. Applications

Medical Diagnosis of Lung Cancer

Variables	Type	Values
Pollution (P)	Binary	{Low, High}
Smoker (S)	Boolean	{True, False}
Cancer (C)	Boolean	{True, False}
Dyspnoea[1] (D)	Boolean	{True, False}
X-ray (X)	Binary	{Pos., Neg.}

- Assume we have probability distribution $\mathbb{P}(P, S, C, D, X)$ (e.g., obtained from data).

[1] Short breath

Medical Diagnosis of Lung Cancer

Variables	Type	Values
Pollution (P)	Binary	{Low, High}
Smoker (S)	Boolean	{True, False}
Cancer (C)	Boolean	{True, False}
Dyspnoea[1] (D)	Boolean	{True, False}
X-ray (X)	Binary	{Pos., Neg.}

- Assume we have probability distribution $\mathbb{P}(P, S, C, D, X)$ (e.g., obtained from data).
- We can then easily compute probabilities for any event: e.g., $\mathbb{P}(C = True), \mathbb{P}(D = True), \mathbb{P}(C = True, S = F)$...

[1] Short breath

Medical Diagnosis of Lung Cancer

Variables	Type	Values
Pollution (P)	Binary	{Low, High}
Smoker (S)	Boolean	{True, False}
Cancer (C)	Boolean	{True, False}
Dyspnoea[1] (D)	Boolean	{True, False}
X-ray (X)	Binary	{Pos., Neg.}

- Assume we have probability distribution $\mathbb{P}(P, S, C, D, X)$ (e.g., obtained from data).
- We can then easily compute probabilities for any event: e.g., $\mathbb{P}(C = True), \mathbb{P}(D = True), \mathbb{P}(C = True, S = F)$...
- Even better, we can compute conditional probabilities: e.g., $\mathbb{P}(C = True \,|\, D = True)$, $\mathbb{P}(C = False \,|\, X = True)$, $\mathbb{P}(C = True \,|\, S = True, X = False)$...

[1] Short breath

Memory Requirement

- How many values do we need to store $\mathbb{P}(P, S, C, D, X)$?

Memory Requirement

- How many values do we need to store $\mathbb{P}(P, S, C, D, X)$?
- In general, $2^n - 1$ for n boolean variables

Memory Requirement

- How many values do we need to store $\mathbb{P}(P, S, C, D, X)$?
- In general, $2^n - 1$ for n boolean variables
- We can exploit (conditional) independences to reduce this requirement e.g., Assume:

$$\mathbb{P}(P, S, C, D, X) = \mathbb{P}(P) \times \mathbb{P}(S) \times \mathbb{P}(C \mid P, S) \times \mathbb{P}(D \mid C) \times \mathbb{P}(X \mid C)$$

what is the space requirement now?

Memory Requirement

- How many values do we need to store $\mathbb{P}(P, S, C, D, X)$?
- In general, $2^n - 1$ for n boolean variables
- We can exploit (conditional) independences to reduce this requirement
  e.g., Assume:

$$\mathbb{P}(P, S, C, D, X) = \mathbb{P}(P) \times \mathbb{P}(S) \times \mathbb{P}(C \mid P, S) \times \mathbb{P}(D \mid C) \times \mathbb{P}(X \mid C)$$

  what is the space requirement now?
- Huge saving in large domains with many (conditional) independences

1 Reasoning under Uncertainty

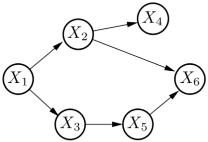
2 Example

3 Graphical Models

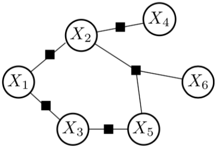
4 Applications

Graphical Models

- What is a graphical model?
  - Graphical notation to express all the (random) variables and their (conditional) (in)dependences
  - (Compact) representation of a joint distributions over the variables
- Two basic variants:
  - Directed model: Bayesian network or belief network
  - Undirected model: factor graph or Markov random field



$P(x_{1:6}) = P(x_1)\ P(x_2|x_1)\ P(x_3|x_1)\ P(x_4|x_2)\ P(x_5|x_3)\ P(x_6|x_2, x_5)$



$P(x_{1:6}) = f_1(x_1, x_2)\ f_2(x_3, x_1)\ f_3(x_2, x_4)\ f_4(x_3, x_5)\ f_5(x_2, x_5, x_6)$

From Marc Toussaint

Why are Graphical Models Useful?

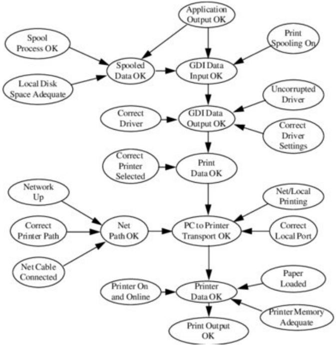
- Why do we need graphical models?
  - Real problems have many variables
  - Real problems have some structure (i.e., conditional probabilistic independences)

- What can we do with graphical models?
  - Are two variables independent given a third one?
  - What is the probability of some event?
  - What is the probability of some event given some observation?

1. Reasoning under Uncertainty
2. Example
3. Graphical Models
4. Applications

Diagnostics

- Medical diagnosis
- Printer troubleshooting

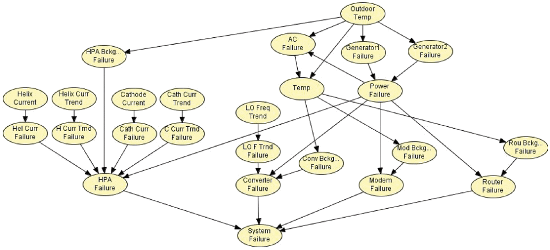


Heckerman and Breese, 1996

- Equipment failure
- Intelligent tutoring

Prediction and decision-making

- Failure prediction in satellite



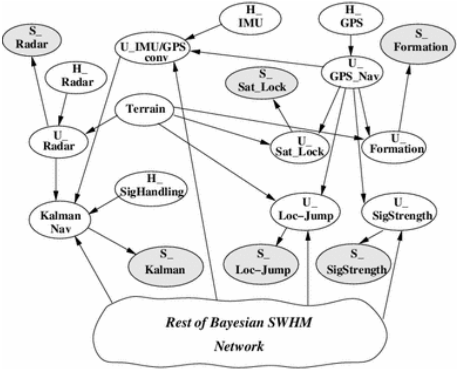
Bottone et al., 2008

- Disease prediction
- Price evolution in stockmarket

- With utilities, optimal decisions with influence diagrams
- More about this, in the last part of this course

Anomaly Detection

- Software health management in UAVs

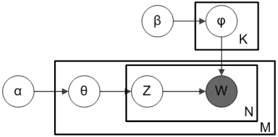


Schumann et al., 2013

- Breakdown detection
- Fraud detection in banking and finance
- Spam detection

Machine Learning Models

- Any probabilistic ML model could exploit Bayes net
- Many applications in computer vision, speech recognition, bioinformatics...
- e.g., Latent Dirichlet Allocation in NLP



α parameter of Dirichlet prior of per-document topic distribution
β parameter of Dirichlet prior of per-topic word distribution
θ_m topic distribution for document m
ϕ_k word distribution for topic k
z_{mn} topic for for n-th word for document m
w_{mn} n-th word for document m