BITS Pilani

Pilani | Dubai | Goa | Hyderabad

Probabilistic Graphical Model

Session # 1 : Introduction

Seetha Parameswaran

seetha.p@pilani.bits-pilani.ac.in

The instructor is gratefully acknowledging the authors who made their course materials freely available online.

Table of Contents

Uncertainty

- We select a course of actions among many possibilities.
- Decisions may be based on the information obtained from the environment, previous knowledge and the objectives.
- Eg: It looks cloudy. Should I carry an umbrella?
- The information and knowledge is incomplete or unreliable. So the decisions made are not certain. We make decisions under uncertainty.
- One of the goals of AI is to develop systems that can reason and make decisions under uncertainty.

Uncertainty

- Complexity increases
  - Each piece of knowledge may not be independently used to arrive at decisions.
  - Deduced facts are maintained along with new facts. This increases the knowledge base.
- Examples
  - A medical doctor in an emergency.
  - An autonomous vehicle that detects what might be an obstacle in its way.
  - A financial agent needs to select the best investment.

correlation $\Rightarrow$ bad | independence $\Rightarrow$ good

Limitations of Traditional Approach

$P(A, B) \rightarrow 4 \times 4 = 16$ different values to store

too many entries

- Impractical for complex problems with many variables, as the size of the Joint probability table and the direct computation of Marginal and Conditional probabilities grow exponentially with the number of variables in the model. $f(x_1, x_2 \dots x_n)$
- Good estimates for the joint probabilities from data requires a very large database if there are many variables in the model.

is this a good thing? why

N variables $\rightarrow$ 2 values each (bi-valued)

$\boxed{P(A=x, B=y)}$

$= P(A=x)P(B=y)$

JPDF $\longrightarrow 2^N$ different values to store

Table of Contents

Probabilistic Graphical Models

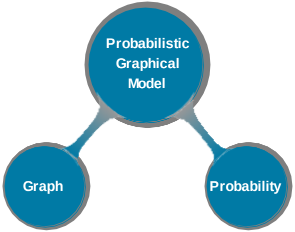
- Provide a framework for managing uncertainty based on probability theory in a computationally efficient way.
- Consider Independence relations that are valid for a certain problem, and use independence to reduce computational complexity and be memory efficient.
- Represent the dependence and independence relations between a set of variables using graphs.

Probabilistic Graphical Models

- Around 2000s
- Techniques based on probability and graphical representations were consolidated as powerful methods for representing, reasoning and making decisions under uncertainty.
- Bayesian networks, Markov networks, influence diagrams and Markov decision processes, among others.

Probabilistic Graphical Model

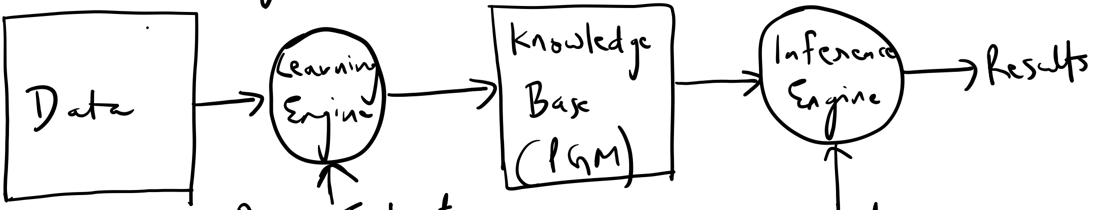
- Probabilistic Graphical Model combines probability theory and graph theory to deal with problems involving uncertainty and complexity and also in the design and analysis of machine learning algorithms.



Declarative Representation

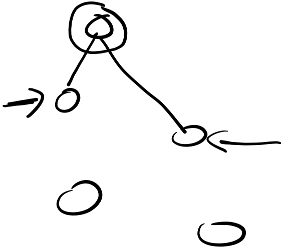
Separation of knowledge and reasoning — Encar, Chaffard

The representation has its own clear semantics, the reasoning algorithms are independent of this.



Need of Probabilistic Graphical Model

- Model uncertainty.
- Model complex structures with causal and spatial relationships.
- Model domain knowledge and prior knowledge.
- Draw inferences from the model.
- Learn the structure and parameters of the model.



Classification of Probabilistic Graphical Model

1. Direct or Undirected
   - Directed graphs represent parent-child relations or cause-effect relations.
   - Undirected graphs represent symmetric relations.
2. Static or Dynamic
   - Static – Model represents a set of variables at a certain point in time.
   - Dynamic – Model represents a set of variables across different times.
3. Probabilistic or Decisional
   - Probabilistic models include random variables.
   - Decisional models include random variables, decision and utility variables.

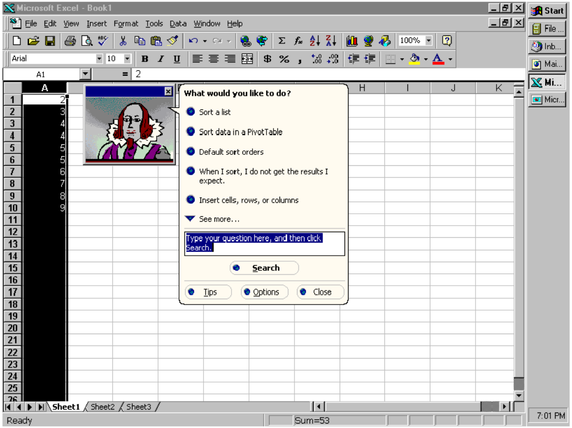
Common Probabilistic Graphical Models

Type	Directed / Undirected	Static / Dynamic	Probabilistic / Decisional
Bayesian Models	both	Static	Probabilistic
Markov Chains	Directed	Dynamic	Probabilistic
Hidden Markov Models	Directed	Dynamic	Probabilistic
Markov Random Fields	Undirected	Static	Probabilistic
Bayesian Networks	Directed	Static	Probabilistic
Dynamic Bayesian Networks	Directed	Dynamic	Probabilistic
Influence Diagrams	Directed	Static	Decisional
Markov Decision Processes	Directed	Dynamic	Decisional
Partially Observable MDPs	Directed	Dynamic	Decisional

L E Sucar

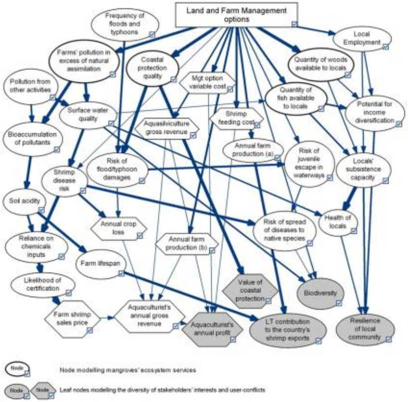
Table of Contents

Bayesian Models -Microsoft Lumiere Project



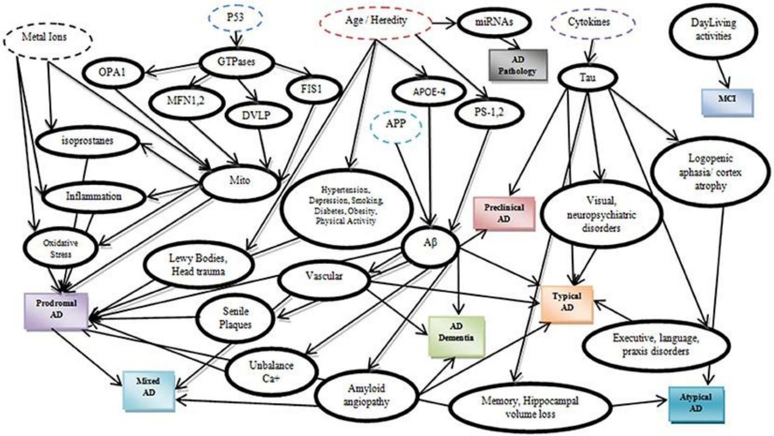
http://erichorvitz.com/lumiere.htm

Bayesian Models – Life Sciences



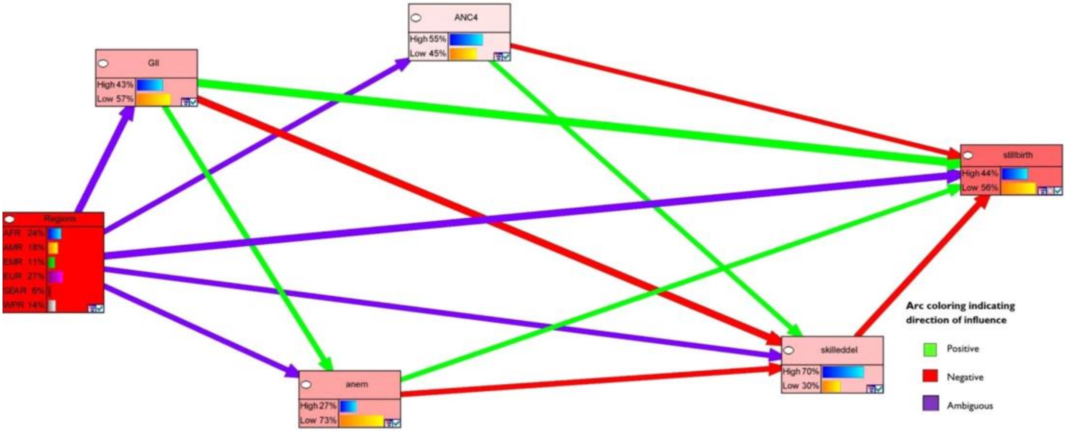
Schmitt, Laetitia & Brugere, Cecile. (2013). Capturing Ecosystem Services, Stakeholders' Preferences and Trade-Offs in Coastal Aquaculture Decisions: A Bayesian Belief Network Application.

Bayesian Models – Medical Diagnosis



https://doi.org/10.3389/fnagi.2017.00077

Bayesian Models – Spatial Relationship



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6930217/

Sucar's book

~~Hidden Markov~~ Model – Gesture Recognition

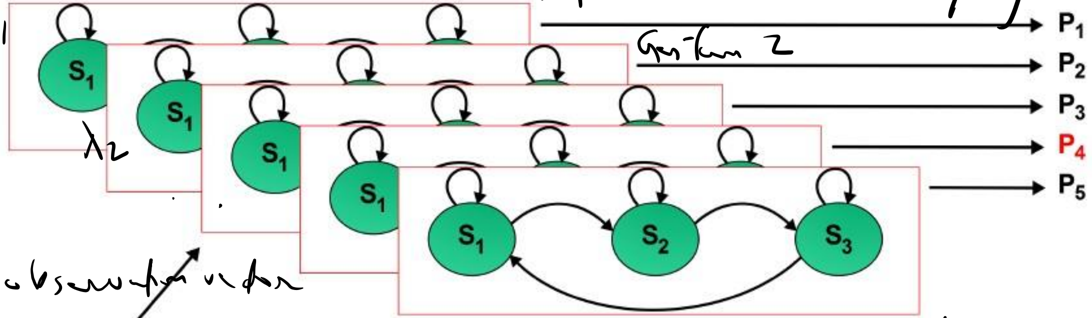
$P(\lambda_i / O)$

Gesture 1

Gesture 2

λ_1

λ_2



observation vector

∴ # of states (states are hidden)

observation probabilities $P(O/s_t)$

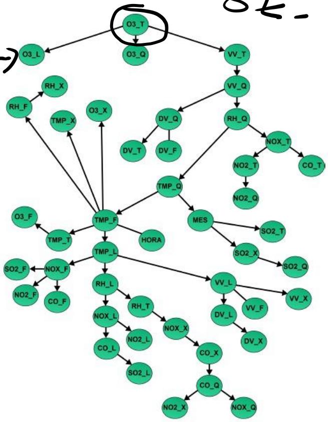
state transition probs $P(s_{t+1}/s_t)$

L E Sucar

Bayesian Networks – Ozone Prediction

Sucar's book

Tree



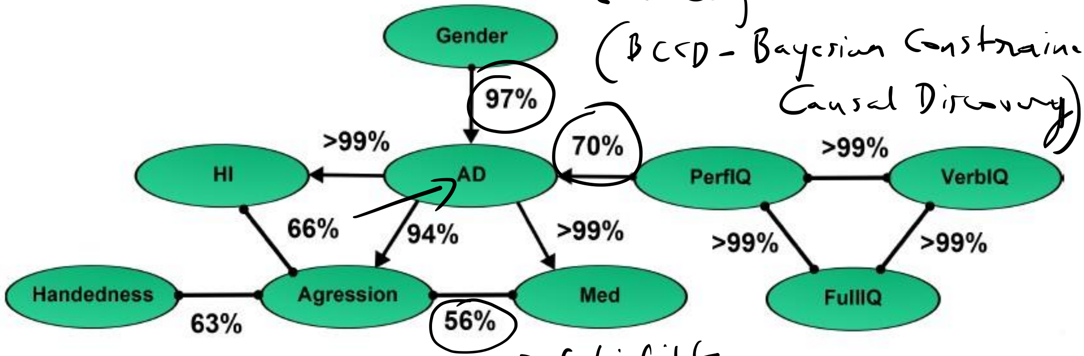
OZ_T

forecast pollution levels several hours in advance

Total of 47 variables
9 measurements each
for 5 stations + hour
+ month

L E Sucar

Bayesian Networks – Attention Deficit Model



Learning a causal model
(BCCD - Bayesian Constrained Causal Discovery)

→ reliability

L E Sucar

A Detailed Look

Consider a simple medical diagnosis problem - there are two diseases flu and hay fever. These diseases are not mutually exclusive

4-valued random variable → Season (Winter, Spring, Summer, Autumn)

2-valued symptoms → Congestion, Muscle pain

Total of 5 variables: Season, Congestion, Muscle pain, Flu, Hay Fever innovate achieve lead

A Detailed Look

What is the sample space size of this joint distribution?

$$4 \times 2 \times 2 \times 2 \times 2 = 64$$

(for season) (for Congestion) (for muscle pain) (for hay fever) (for flu)

<u>Question:</u> How likely is it that a patient has the flu given that it is <u>fall</u>, and the patient has <u>sinus congestion</u> but <u>no muscle pain</u>?

A Detailed Look

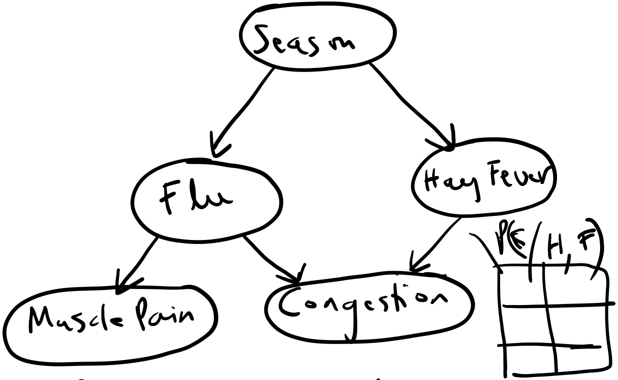
Daphne Koller

P(Flu = True | Season = Fall, Congestion = True, Muscle Pain = False)

- Already looks complicated even when there are only 64 points in the sample space
- What happens when we have hundreds of attributes?
- Fortunately there is a way out → can compactly encode a joint distribution over hundreds of variables using a graphical model.

$P(F/S, H) = P(F/S)$

A Detailed Look



Independencies

$(F \perp H \mid S)$ → is independent

$(C \perp S \mid F, H)$

$(M \perp H, C \mid F)$

$(M \perp C \mid F)$

$P(\text{Season}, \text{flu}, \text{hay fever} \ldots) = P(\text{Season})\, P\left(\frac{\text{Flu}}{\text{Season}}\right) P(HF/\text{Season})$

A Detailed Look

The graph can be viewed in two <u>perspectives</u>

The graph is a compact representation of a set of independencies

<u>for example</u>

$P(\text{Congestion} \mid \text{Flu}, \text{HayFever}, \text{Season})$
$= P(\text{Congestion} \mid \text{Flu}, \text{HayFever})$

Graph defines a skeleton for compactly representing a high-dimensional distribution

A Detailed Look

We can break up the distribution into smaller factors, each over a much smaller space of possibilities

The overall distribution is a product of these smaller factors

$P(\text{spring}, \text{no flu}, \text{hay fever}, \text{sinus congestion}, \text{muscle pain})$

$= P(\text{Season} = \text{spring})\, P(\text{Flu} = \text{false} \mid \text{Season} = \text{spring}) \times$

$P(\text{HayFever} = \text{true} \mid \text{Season} = \text{spring})\, P(\text{Congestion} = \text{True} \mid \text{HayFever} = \text{True}, \text{Flu} = \text{false})$

$P(\text{Muscle Pain} = \text{True} \mid \text{Flu} = \text{false})$

A Detailed Look

This parameterisation is significantly more compact

→ requires only 3 + 4 + 4 + 4 + 2 parameters or

17 parameters → why?

P(Season = spring), P(Season = fall), P(Season = winter), P(season = summer) → only 3 nonredundant parameters since if we know 3 of them, the 4th can be obtained from 1 − (Sum of the given three)

A Detailed Look

$P(Flu = False / Season = spring)$

$P(Flu = False / Season = summer)$

$P(Flu = False / Season = winter)$

$P(Flu = False / Season = Fall)$

Nonredundant

$P(Flu = True / Season = spring)$

$P(Flu = True / Season = summer)$

$P(Flu = True / Season = winter)$

$P(Flu = True / Season = Fall)$

Redundant

A Detailed Look

What about $P(\text{Congestion} \mid \text{Hay Fever}, \text{Flu})$?

$P(\text{Congestion} = \text{True} \mid \text{Hay} = T, \text{Flu} = F)$
$P(\text{Congestion} = \text{True} \mid \text{Hay} = T, \text{Flu} = T)$
$P(\text{Congestion} = \text{True} \mid \text{Hay} = F, \text{Flu} = T)$
$P(\text{Congestion} = \text{True} \mid \text{Hay} = F, \text{Flu} = F)$

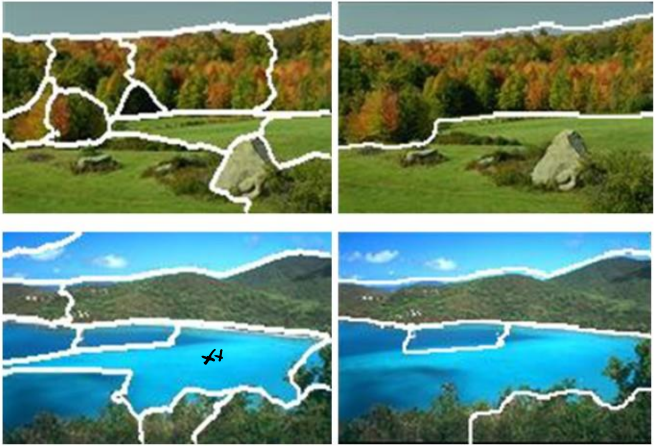
Nonredundant

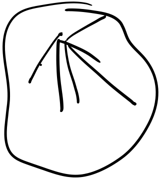
$P(\text{Congestion} = \text{False} \mid \text{Hay} = T, \text{Flu} = F)$
$P(\text{Congestion} = \text{False} \mid \text{Hay} = T, \text{Flu} = T)$
$P(\text{Congestion} = \text{False} \mid \text{Hay} = F, \text{Flu} = T)$
$P(\text{Congestion} = \text{False} \mid \text{Hay} = F, \text{Flu} = F)$

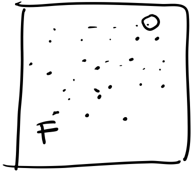
Redundant

$P(F/N, E) = P(F/N)$

Markov Random Fields – Image Segmentation

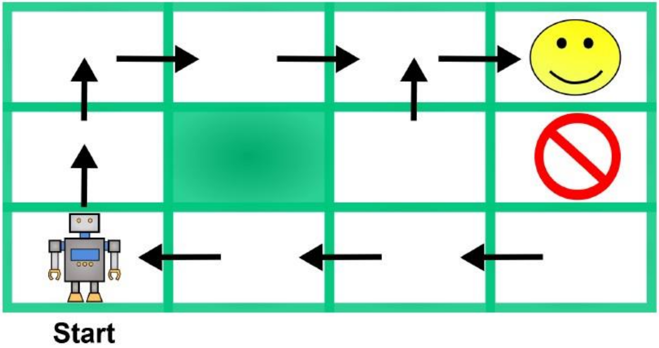






L E Sucar

Markov Decision Process – Robot Motion Planning



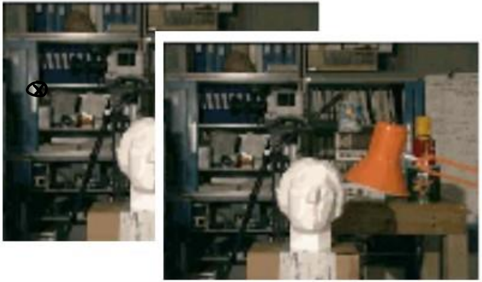
L E Sucar

Use Cases of PGM – Question Answer

D Sontag

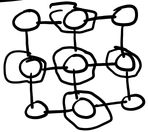
Use Cases of PGM – Stereo Vision

input: two images



output: disparity

Graphical Model



- node for each pixel
- infer depth for each pixel

D Sontag

MIT Ph.D

Google Translation Knowledge and Language Graph

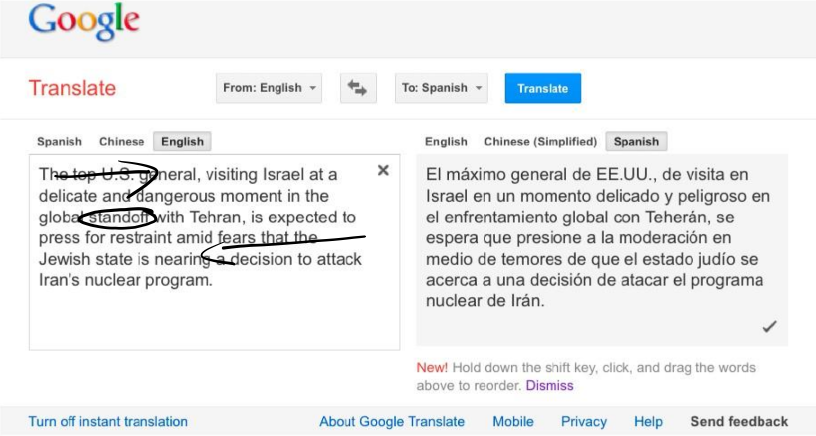
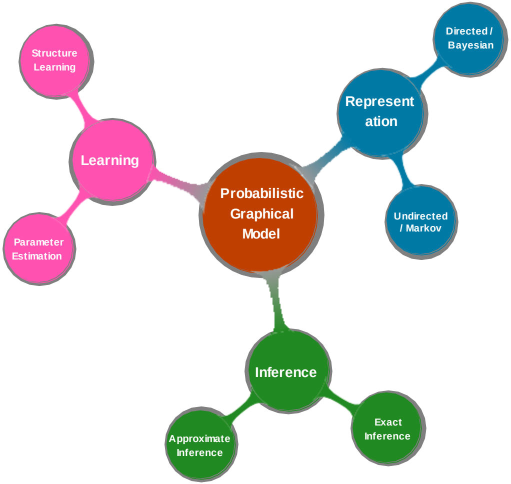


Table of Contents

Components of Probabilistic Graphical Model



Course Handout

M1 Introduction

M2 Mathematical Preliminaries

M3 Directed Graphical Models

M4 Undirected Graphical Models

M5 Exact Inference

M6 Approximate Inference

M7 Parameter Learning

M8 Structure Learning

M9 Models

Lab Sessions

1. Python
2. `pgmpy` Library

L1 Bayesian model representation

L2 Markov Model representation

L3 MAP on Bayesian model

L4 MLE on Bayesian Model

L5 MLE on Markov Model

L6 Learning Structure in Bayesian Model

Details will be posted in Canvas.

Evaluation Components

Component	Weightage
Assignments	20 %
Quiz	10 %
Mid Sem	30 %
Compre	40 %

Further announcements will be posted in Canvas.

References

1. Probabilistic Graphical Models: Principles and Techniques by Daphne Koller and Nir Friedman. MIT Press. 2009
2. Mastering Probabilistic Graphical Models using Python by Ankur Ankan, Abhinash Panda. Packt Publishing 2015.
3. Learning in Graphical Models by Michael I. Jordan. MIT Press. 1999

Thank You !!!