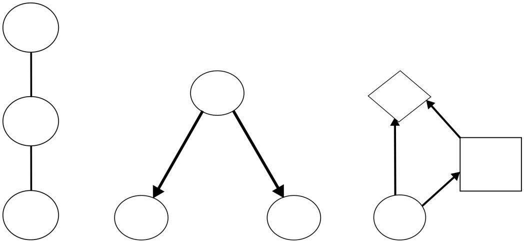
Introduction

Probabilistic Graphical Models

L. Enrique Sucar, INAOE



Outline

Intelligent Agents

- Intelligent agents, natural or artificial, have to select a course of actions among many possibilities
- They make decisions based on the information they can obtain from their environment, their previous knowledge and their objectives
- Information and knowledge is incomplete or unreliable, and the results of their decisions are not certain, they have to make *decisions under uncertainty*: medical doctor in an emergency, an autonomous vehicle that detects what might be an obstacle in its way, financial agent needs to select the best investment
- One of the goals of artificial intelligence is to develop systems that can reason and make decisions under uncertainty

Effects of Uncertainty

- Early artificial intelligence systems were based on classical logic, in which knowledge can be represented as a set of logic clauses or rules. These systems have two important properties:
  - *modularity*, each *piece* of knowledge can be used independently to arrive to conclusions
  - *monotonicity*, knowledge always increases monotonically: any deduced fact is maintained even if new facts are known by the system
- If we have uncertainty these two properties are not true in general, which makes a system that has to reason under uncertainty more complex

Beginnings (1950's and 60's)

- Artificial intelligence (AI) researchers focused on solving problems such as theorem proving, games like chess, and the "blocks world" planning domain, which do not involve uncertainty
- The symbolic paradigm dominated AI in the beginnings

Ad hoc techniques (1970's)

- The development of expert systems for realistic applications such as medicine and mining, required the development of uncertainty management approaches, ad hoc techniques were developed for specific expert systems:
  - MYCIN's certainty factors
  - Prospector's pseudo-probabilities
- Later it was shown that these techniques had a set of implicit assumptions
- Alternative theories were proposed to manage uncertainty: fuzzy logic and the Dempster-Shafer theory

Resurgence of probability **(1980's)**

- Probability theory was used in some initial expert systems, however it was later discarded as its application in naive ways implies a high computational complexity
- New developments, in particular Bayesian networks, make it possible to build complex probabilistic systems in an efficient manner

Diverse formalisms (1990's)

- Bayesian networks continued and were consolidated with the development of efficient inference and learning algorithms
- Other techniques such as fuzzy and non-monotonic logics were considered as alternatives

Probabilistic graphical models **(2000's)**

- Several 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

The basic approach

- If we apply probability in a naive way to complex problems, we are soon deterred by computational complexity
- We can model a problem using a naive probabilistic approach based on a *flat* representation; and then we can use this representation to answer some probabilistic queries
- This will help to understand the limitations of the basic approach, motivating the development of probabilistic graphical models

Model Formulation

- Many problems can be formulated as set of variables, $X_1, X_2, ...X_n$ such that we know the values for some of these variables and the others are unknown. In medical diagnosis, the variables might represent certain symptoms and the associated diseases
- Let us consider that the set of possible values is finite; for example, $X = \{x_1, x_2, ..., x_m\}$ might represent the m possible diseases in a medical domain
- Each value of a random variable will have a certain probability associated in a context, we can represent a domain as:
  1. A set of random variables, $X_1, X_2, ...X_n$.
  2. A joint probability distribution associated to these variables, $P(X_1, X_2, ...X_n)$.

Probabilistic queries (1)

- Marginal probabilities: the probability of one of the variables taking a certain value. This can be obtained by summing over all the other variables of the joint probability distribution. In other words, $P(X_i) = \sum_{\forall X \neq X_i} P(X_1, X_2, ...X_n)$. This is known as *marginalization*.
- Conditional probabilities: the conditional probability of X_i given that we know X_j is $P(X_i \mid X_j) = P(X_i, X_j)/P(X_j), P(X_j) \neq 0$. $P(X_i, X_j)$ and $P(X_j)$ can be obtained via marginalization, and from them we can obtain conditional probabilities.

Probabilistic queries (2)

Total Abduction or MPE: given that a subset (E) of variables is known, abduction consists in finding the values of the rest of variables (J) that maximize their conditional probability given the evidence, $maxP(J \mid E)$. That is $Arg_{X \in J}[maxP(X_1, X_2, ...X_n)/P(X \in E)]$.

Partial abduction or MAP: in this case there are 3 subsets of variables: the evidence, E, the query variables that we want to maximize, J, and the rest of the variables, K, such that we want to maximize $P(J \mid E)$. This is obtained by marginalizing over K and maximizing over J, that is $Arg_{X \in J}[max \sum_{X \in K} P(X_1, X_2, ...X_n)/P(X \in E)]$.

golf example

- In this problem we have 5 variables: *outlook, temperature, humidity, windy, play*.
- All variables are discrete so they can take a value from a finite set of values, for instance Outlook could be sunny, overcast or rain

A sample data set for the golf example

Outlook	Temperature	Humidity	Windy	Play
sunny	high	high	false	no
sunny	high	high	true	no
overcast	high	high	false	yes
rain	medium	high	false	yes
rain	low	normal	false	yes
rain	low	normal	true	no
overcast	low	normal	true	yes
sunny	medium	high	false	no
sunny	low	normal	false	yes
rain	medium	normal	false	yes
sunny	medium	normal	true	yes
overcast	medium	high	true	yes
overcast	high	normal	false	yes
rain	medium	high	true	no

Reduced Example

Simplify the example using only two variables, Outlook and Temperature:

Table: Joint probability distribution for Outlook and Temperature.

	Temp.		
Outlook	H	M	L
S	0.143	0.143	0.071
O	0.143	0.071	0.071
R	0	0.214	0.143

Marginal probabilities

- If we sum per row (marginalizing Temperature) then we obtain the marginal probabilities for Outlook, $P(Outlook) = [0.357, 0.286, 0.357]$
- If we sum per column we obtain the marginal probabilities for Temperature, $P(Temperature) = [0.286, 0.428, 0.286]$
- From these distributions, we obtain that the most probable Temperature is M and the most probable values for Outlook are S and R.

Conditional Probabilities

$P(Temp. \mid Outlook = R) = P(Temp. \wedge Outlook = R)/P(Outlook = R)$

= [0,0.6,0.4]

$P(Outlook \mid Temp. = L) = P(Outlook \wedge Temp. = L)/P(Temp. = L)$

= [0.25,0.25,0.5]

- The most probable Temperature given that the Outlook is Rain is Medium, and the most probable Outlook given that the Temperature is Low is Rain
- The most probable combination of Outlook and Temperature is $\{Rain, Medium\}$, which in this case can be obtained directly from the joint probability table

Limitations

- Becomes impractical for complex problems with many variables, as the size of the table and the direct computation of marginal and conditional probabilities grow exponentially with the number of variables in the model.
- Another disadvantage is that to have good estimates for the joint probabilities from data, we will require a very large database if there are many variables in the model
- The joint probability table does not say much about the problem to a human; so this approach also has cognitive limitations.

Probabilistic Graphical Models (PGMs)

- Provide a framework for managing uncertainty based on probability theory in a computationally efficient way
- Consider only those independence relations that are valid for a certain problem, and include these in the probabilistic model to reduce complexity in terms of memory requirements and also computational time
- Represent the dependence and independence relations between a set of variables using graphs

Representation

- A Probabilistic Graphical Model is a compact representation of a joint probability distribution, from which we can obtain marginal and conditional probabilities. It has several advantages over a flat representation:
  - It is generally much more compact (space).
  - It is generally much more efficient (time).
  - It is easier to understand and communicate.
  - It is easier to learn form data or to construct based on expert knowledge.

Specification

A PGM is specified by two aspects: (i) a graph, $G(V, E)$, that defines the structure of the model; and (ii) a set of local functions, $f(Y_i)$, that define the parameters, where Y_i is a subset of X. The joint probability is obtained by the product of the local functions:

$$P(X_1, X_2, \ldots, X_N) = K \prod_{i=1}^{M} f(Y_i) \tag{1}$$

Inference and learning

Inference: obtain the marginal or conditional probabilities of any subset of variables Z given any other subset Y.

Learning: given a set of data values for X (that can be incomplete) estimate the structure (graph) and parameters (local functions) of the model.

Classification of PGMs

1. Directed or Undirected – Undirected graphs represent symmetric relations, while directed graphs represent relations in which the direction is important
2. Static or Dynamic – Defines if the model represents a set of variables at a certain point in time (static) or across different times (dynamic)
3. Probabilistic or Decisional – Probabilistic models only include random variables, while decisional models also include decision and utility variables

Common PGMs

Type	**D**ir/**U**ndir	**S**tatic/**D**yn	**P**rob./**D**ec
Bayesian Classifiers	D/U	S	P
Markov Chains	D	D	P
Hidden Markov Models	D	D	P
Markov Random Fields	U	S	P
Bayesian Networks	D	S	P
Dynamic Bayesian Networks	D	D	P
Influence Diagrams	D	S	D
Markov Decision Processes	D	D	D
Partially Observable MDPs	D	D	D

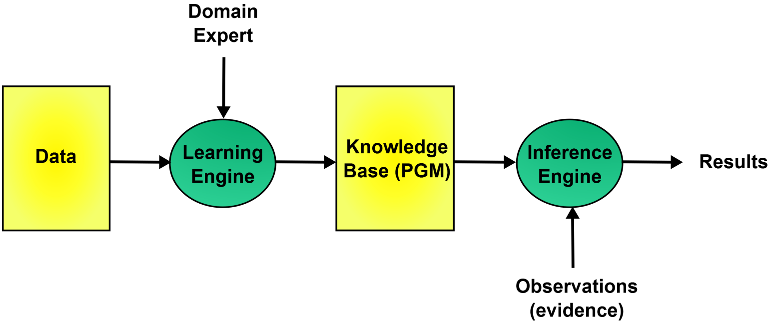
Representation, Inference and Learning

Uncertainty
A brief history
Basic probabilistic models
Example
Probabilistic graphical models
Applications
Overview
References

- The representation is the basic property of each model, and it defines which entities constitute it and how these are related.
- Inference consists in answering different probabilistic queries based on the model and some evidence
- To construct these models there are basically two alternatives: to build it “by hand” with the aid of domain experts or to induce the model from data.

General Paradigm

- Separate the inference and learning techniques from the model –the reasoning mechanisms are general and can be applied to different models



App Examples

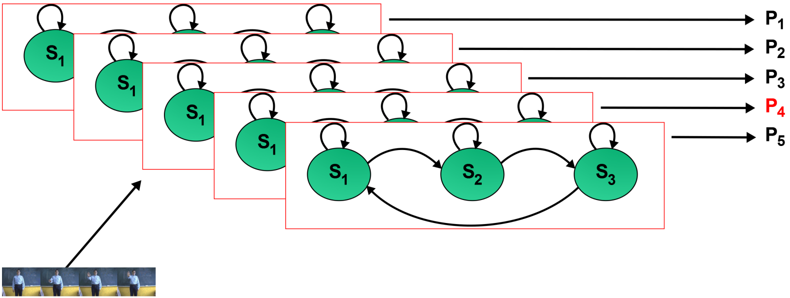
- Medical diagnosis and decision making.
- Mobile robot localization, navigation and planning.
- Diagnosis for complex industrial equipment such as turbines and power plants.
- User modeling for adaptive interfaces and intelligent tutors.
- Speech recognition and natural language processing.
- Pollution modeling and prediction.
- Reliability analysis of complex processes.
- Modeling the evolution of viruses.
- Object recognition in computer vision.
- Information retrieval.
- Energy markets.

Bayesian classifiers - skin detection

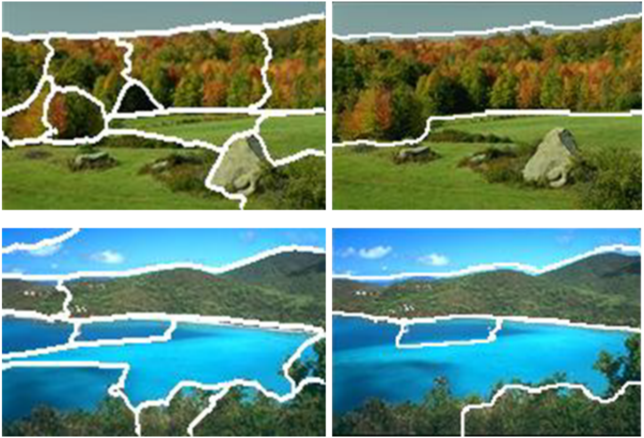
(a) Original image

(b) Image with skin pixels detected

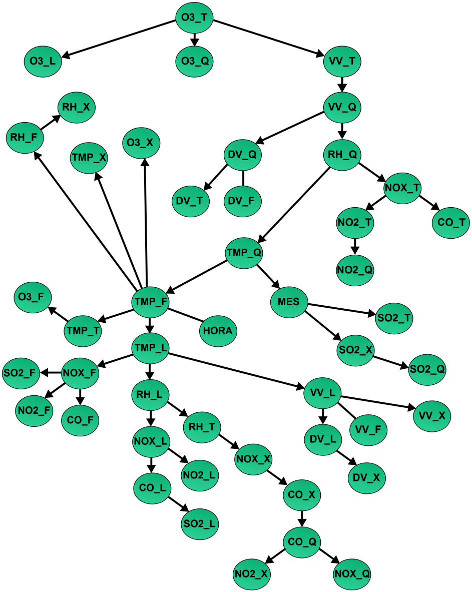
HMMs - gesture recognition



MRFs - image segmentation

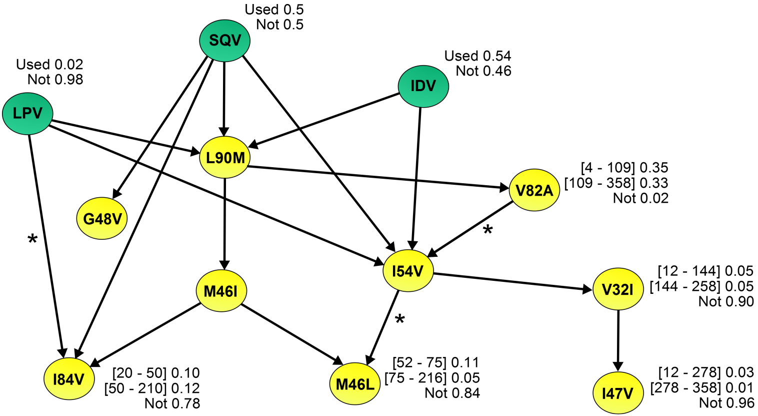


BNs - ozone prediction

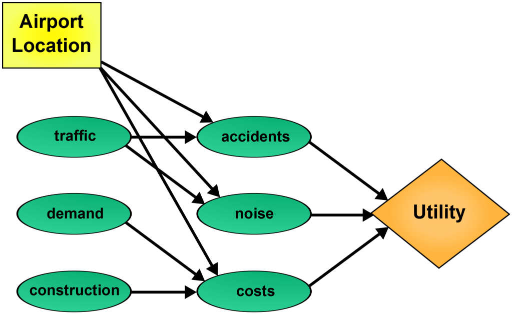


Uncertainty
A brief history
Basic probabilistic models
Example
Probabilistic graphical models
Applications
Overview
References

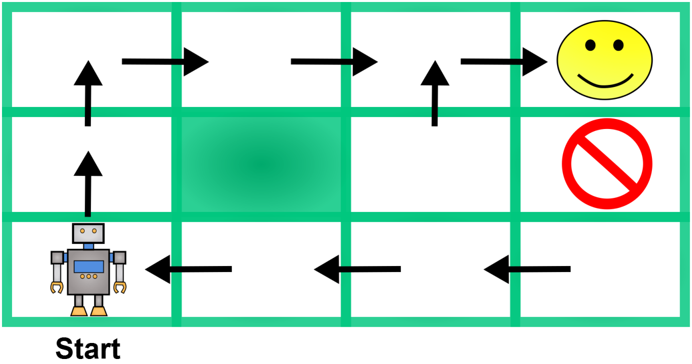
TNBNs - HIV mutation prediction



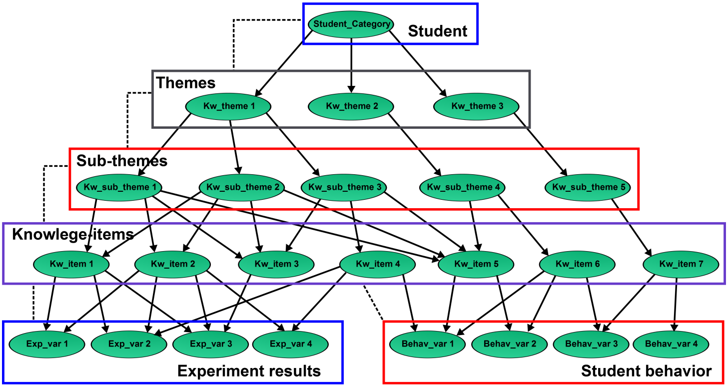
Decision net - airport location



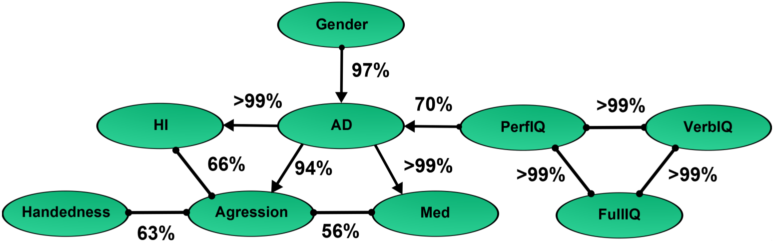
MDP - robot motion planning



PRMs - student modeling



Causal BN - attention deficit model



Overview (1)

- Part I provides the mathematical foundations: probability theory and graph theory
- Part II covers the probabilistic models that only have random variables
  - Bayesian classifiers
  - Markov chains and hidden Markov models
  - Markov random fields
  - Bayesian networks
  - Dynamic Bayesian networks and temporal networks

Overview (2)

- Part III presents those models that consider decisions and utilities: Decision Diagrams and Markov Decision Processes
- Part IV considers alternative paradigms that can be thought as extensions to the traditional probabilistic graphical models:
  - relational probabilistic models,
  - causal graphical models

Book

Sucar, L. E, *Probabilistic Graphical Models*, Springer 2015 – Chapter 1

Additional Reading (1)

- Bessiere, Mazer, E., P., Ahuactzin, J.M., Mekhnacha, K.: Bayesian Programming. CRC Press, Boca Raton, Florida (2014)
- Darwiche, A.: Modeling and Reasoning with Bayesian Networks. Cambridge University Press, New York (2009)
- Duda, R.O., Hart, P.A., Nilsson, N.L.: Subjective Bayesian Methods for Rule-Based Inference Systems. In: Proceeding of the National Computer Conference, vol. 45, pp. 1075–1082 (1976)
- Heckerman, D.: Probabilistic Interpretations for MYCIN's Certainty Factors. In: Proceedings of the First Conference on Uncertainty in Artificial Intelligence (UAI), pp. 9–20 (1985)

Additional Reading (2)

- Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer-Verlag, New York (2001)
- Daphne Koller, D., Friedman. N.: Probabilistic Graphical Models: Principles and Techniques. MIT Press. Cambridge (2009)
- Lauritzen, S.L.: Graphical Models. Oxford University Press, Oxford(1996)
- Neapolitan, R. E.: Probabilistic Reasoning in Expert Systems. John Wiley & Sons, New York (1990)
- Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco (1988)

Additional Reading (3)

- Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
- Shortliffe, E.H., Buchanan B.G.: A Model of Inexact Reasoning in Medicine. Mathematical Biosciencies, 23, 351–379 (1975)
- Zadeh, L.A.: Knowledge Representation in Fuzzy Logic. IEEE Trans. on Knowledge and Data Engineering, 1(1), 89–100 (1989)