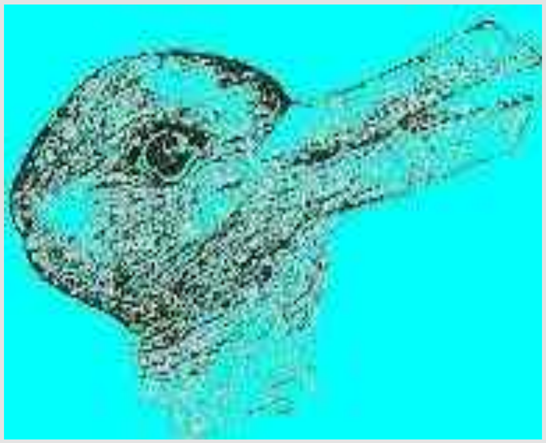
Probabilistic Graphical Models: Applications in Biomedicine

L. Enrique Sucar
Dept. of Computer Science
Instituto Nacional de Astrofísica, Óptica y Electrónica
Puebla, MEXICO

http://ccc.inaoep.mx/~esucar/

What do you see?



What we see depends on our previous knowledge (model) of the world and the information (data) form the images → Bayesian framework

Outline

- Introduction
  - Probability Theory
  - Probabilistic Graphical Models
- Bayesian Networks
  - Endoscopy assistant
- Temporal Bayesian Networks
  - Predicting HIV mutations
- Markov Decision Processes
  - User adaptation for rehabilitation
- Conclusions

What is Probability?

Two main interpretations:

- Objective (classical, frequency, propensity) – probabilities exist in the real world and can be measured
- Epistemological (logical, subjective) – probabilities have to do with human knowledge, degree of belief

Justifications of Probability

- Dutch book argument

  If someone bets without following the axioms of probability, he can loose always against an opponent

- Logical deduction

  From a series of basic requirements we can deduce the axioms of probability theory

Kolmogorov Axioms

- $0 <= P(A) <= 1$
- $P(S) = 1$
- $P(A \cup B \cup C \ldots) = P(A) + P(B) + P(C) + \ldots$

A, B, C … mutually exclusive

Conditional Probability

$$P(A \mid B) = P(A \cap B) / P(B)$$

- Probability of an event given that another event occurs:
  - What is the probability of obtaining a prime number in a die toss, given that it is a even number?
  - If someone has a headache, what is the probability the she has flu?

Bayes Rule

- From the definition of conditional probability we can deduce Bayes Rule:

$$P(B \mid A) = P(B)\, P(A \mid B) \,/\, P(A),\; P(A) > 0$$

- This allows us to “invert” the probabilities …

Independent Events

- Two events are independent if the occurrence of one event does not alter the probability of the other:

$$P(A \mid B) = P(A)$$

$$P(B \mid A) = P(B)$$

- Which is the same as:

$$P(A \cap B) = P(A)\, P(B)$$

Conditional Independence

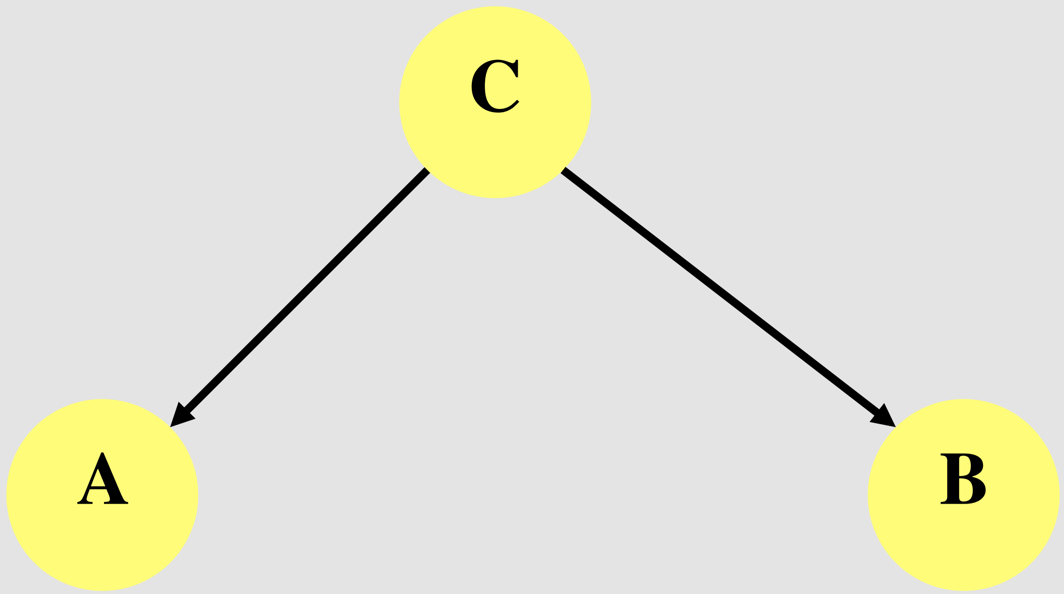
- *A* is conditionally independent of *B* given *C*, if knowing *C* makes *A* and *B* independent:

$$P(A \mid B,C) = P(A \mid C)$$

- Example:
  - A – water the garden
  - B – weather prediction
  - C – rain

Conditional Independence: Graphical Representation

- We can represent conditional independence relations using (directed or undirected) graphs



Bayesian Reasoning

- In the Bayesian approach we combine our previous knowledge (*priors*) with the evidence (*likelihood*) to arrive to conclusions (*posterior*):

$$P(H \mid E) \propto P(H)\, P(E \mid H)$$

Example: Bayesian perception

- The perception problem is characterized by two main aspects:
  - The properties of the world that is observed (prior knowledge)
  - The image data acquired by the observer (evidence)
- The Bayesian approach combines the two aspects which are characterized as probability distributions

Representation

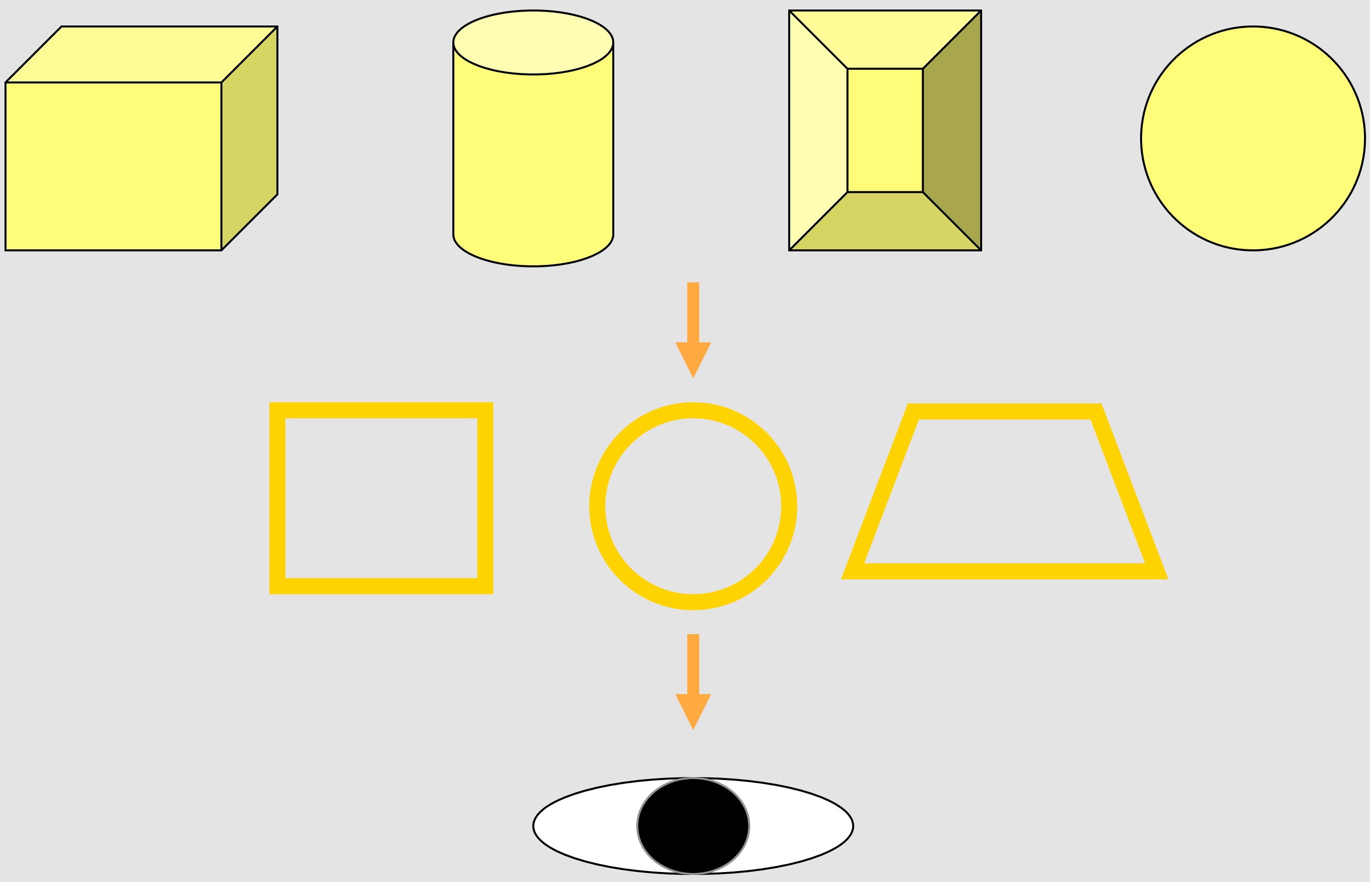
- Scene properties – S
- Model of the world – prior probability distribution – $P(S)$
- Model of the image – probability distribution of the image given de scene (likelihood) – $P(I|S)$
- The scene (object) is characterized by the posterior probability distribution – $P(S|I)$
- By Bayes theorem:

$$P(S|I) = P(S)\,P(I|S) \,/\, P(I)$$

- The denominator can be consider as a normalizing constant:

$$P(S|I) = k\,P(S)\,P(I|S)$$

Example



Example

- Prior distribution of objects – P(O)
  - Cube 0.2
  - Cylinder 0.3
  - Sphere 0.1
  - Prism 0.4

Example

- Likelihood function P(Silhouette|Object) – P(S|O)

	Cube	Cylinder	Sphere	Prism
Square	1.0	0.6	0.0	0.4
Circle	0.0	0.4	1.0	0.0
Trapezoid	0.0	0.0	0.0	0.6

Example

- Posterior distribution P(Object|Silhouette) – P(O|S)
- Bayes rule:

  $P(O|S) = k\, P(O)\, P(S|O)$

- For example, given S=square

  P(Cube | square)= k 0.2 * 1 = k 0.2 = 0.37

  P(Cylinder | square)= k 0.3 * 0.6 = k 0.18 = 0.33

  P(Sphere | square)= k 0.1 * 0 = 0

  P(Prism | square)= k 0.4 * 0.4 = k 0.16 = 0.30

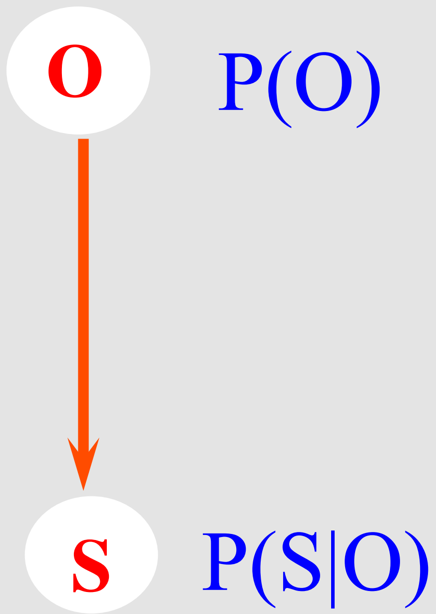
Probabilistic Graphical Models

- If we apply the Bayesian approach in naive way its complexity grows exponentially on the size (number of variables) of the model

- Probabilistic graphical models take advantage of the independence relations among the variables in a domain to develop more efficient representations as well as inference and learning techniques

Graphical Model

- We can represent the dependence relation in this simple example graphically, with 2 variables and an arc



Probabilistic Graphical Models

- Given a set of (discrete) random variables,

$$\boldsymbol{X} = X_1, X_2, \ldots, X_N$$

- The joint probability distribution,

$$P(X_1, X_2, \ldots, X_N)$$

- specifies the probability for each combination of values (the joint space). From it, we can obtain the probability of a variable(s) (marginal), and of a variable(s) given the other variables (conditional)

Probabilistic Graphical Models

- 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 build (from experts) or learn (from data)

Probabilistic Graphical Models

- A graphical model is specified by two aspects:
  - A Graph, *G(V,E)*, that defines the structure of the model
  - A set of local functions, $f(\mathbf{Y}_i)$, that defines the parameters (probabilities), where $\mathbf{Y}_i$ is a subset of $\mathbf{X}$
- The joint probability is defined by the product of the local functions:

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

Probabilistic Graphical Models

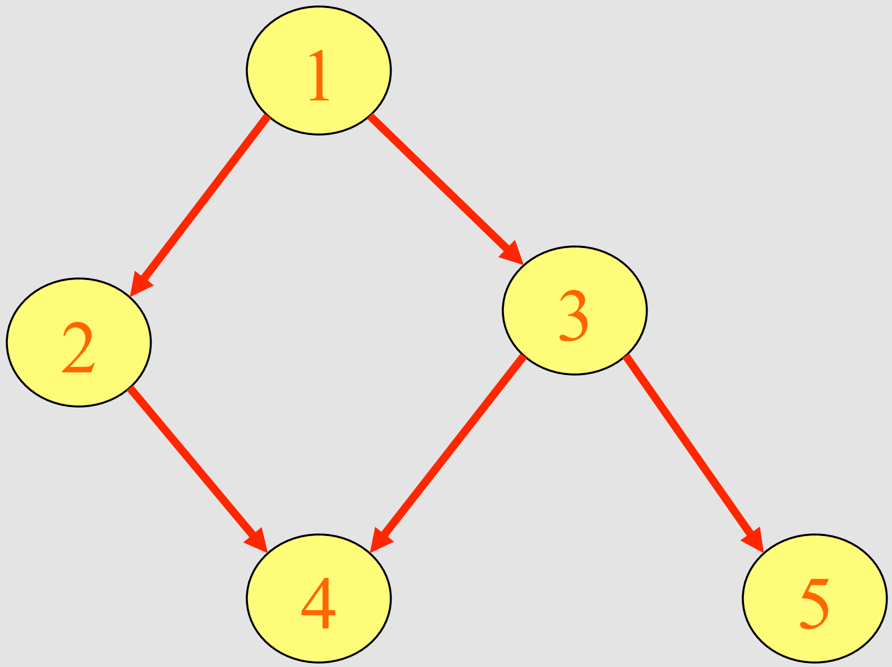
- This representation in terms of a graph and a set of local functions (called potentials) is the basis for *inference* and *learning* in PGMs
  - **Inference:** obtain the marginal or conditional probabilities of any subset of variables ***Z*** given any other subset ***Y***
  - **Learning:** given a set of data values for ***X*** (that can be incomplete) estimate the structure (graph) and parameters (local function) of the model

Probabilistic Graphical Models

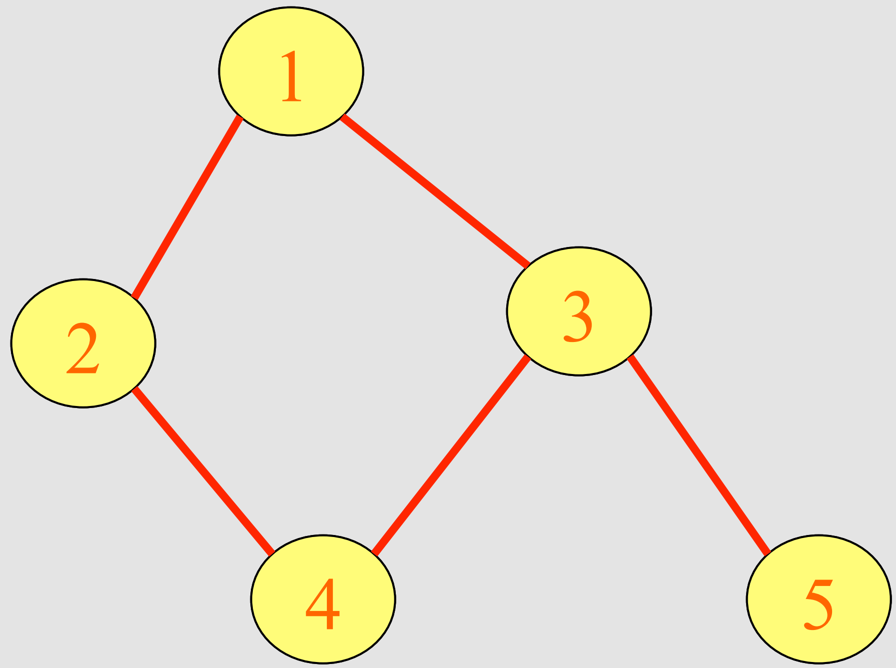
- We can classify graphical models according to 3 dimensions:
  - Directed vs. Undirected
  - Static vs. Dynamic
  - Probability vs. Decision

Probabilistic Graphical Models

- Directed

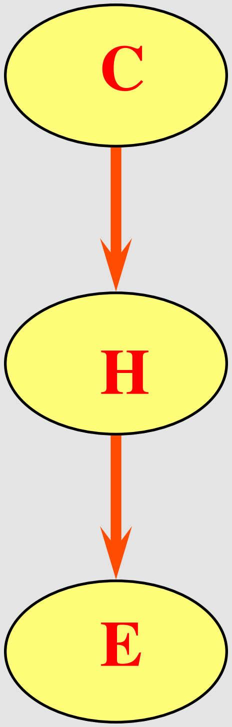


- Undirected

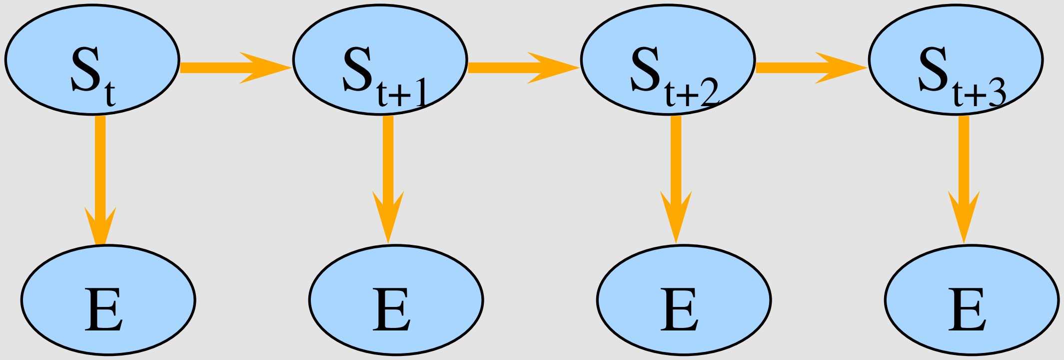


Probabilistic Graphical Models

- Static

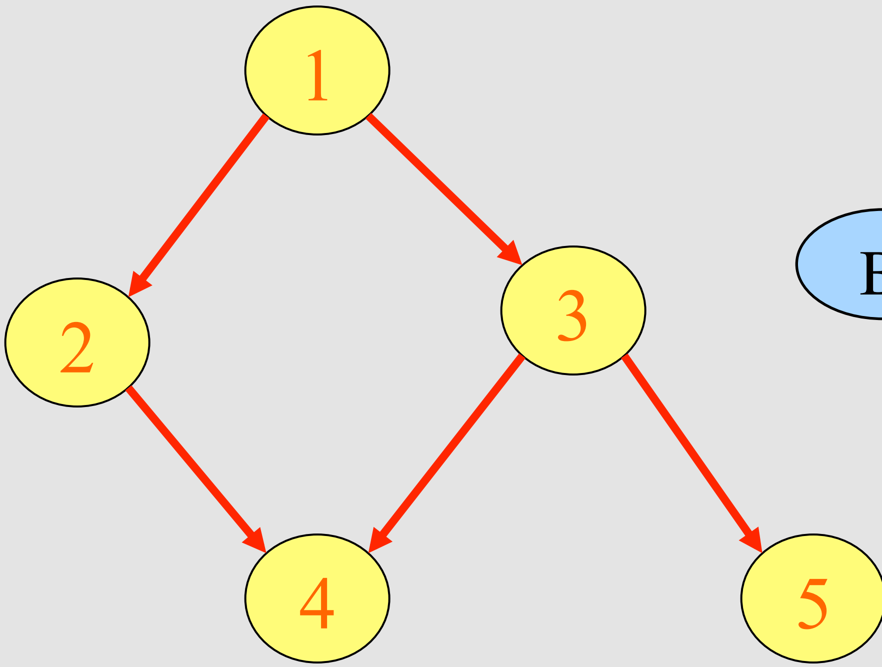


- Dynamic

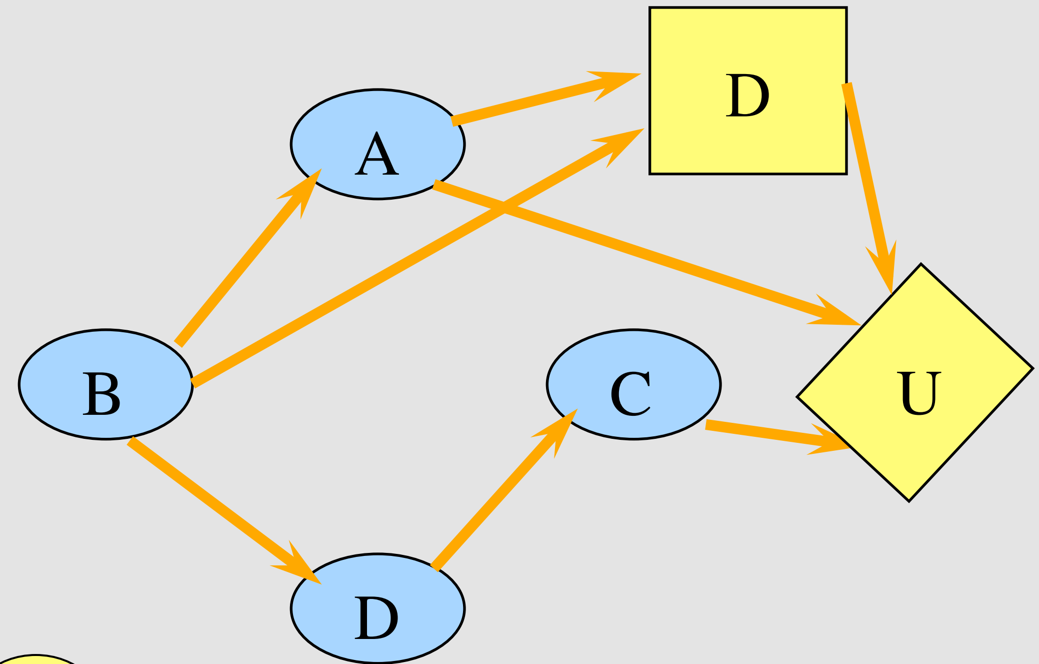


Probabilistic Graphical Models

- Only random variables



- Considers decisions and utilities



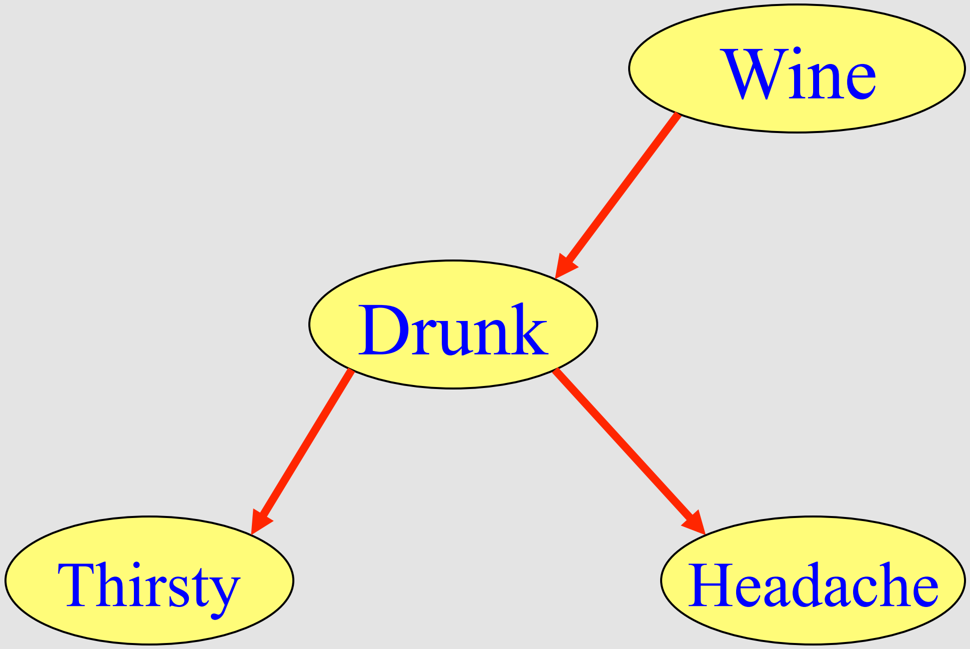
Types of PGMs

- There are different classes of PGMs:
  - Bayesian classifiers
  - Bayesian networks
  - Hidden Markov models
  - Dynamic Bayesian networks
  - Temporal Bayesian networks
  - Markov Random Fields
  - Influence diagramas
  - Markov decision processes

Bayesian Networks

- Bayesian networks (BN) are a graphical representation of dependencies between a set of random variables. A Bayesian net is a Directed Acyclic Graph (DAG) in which:
  - Node: Propositional variable.
  - Arcs: Probabilistic dependencies.
- An arc between two variables represents a direct dependency, usually interpreted as a *causal* relation.

An example of a BN



- Represents (in a compact way) the joint probability distribution:

$$P(W,D,T,H) = P(W)\, P(D|W)\, P(T|D)\, P(H|D)$$

Structure

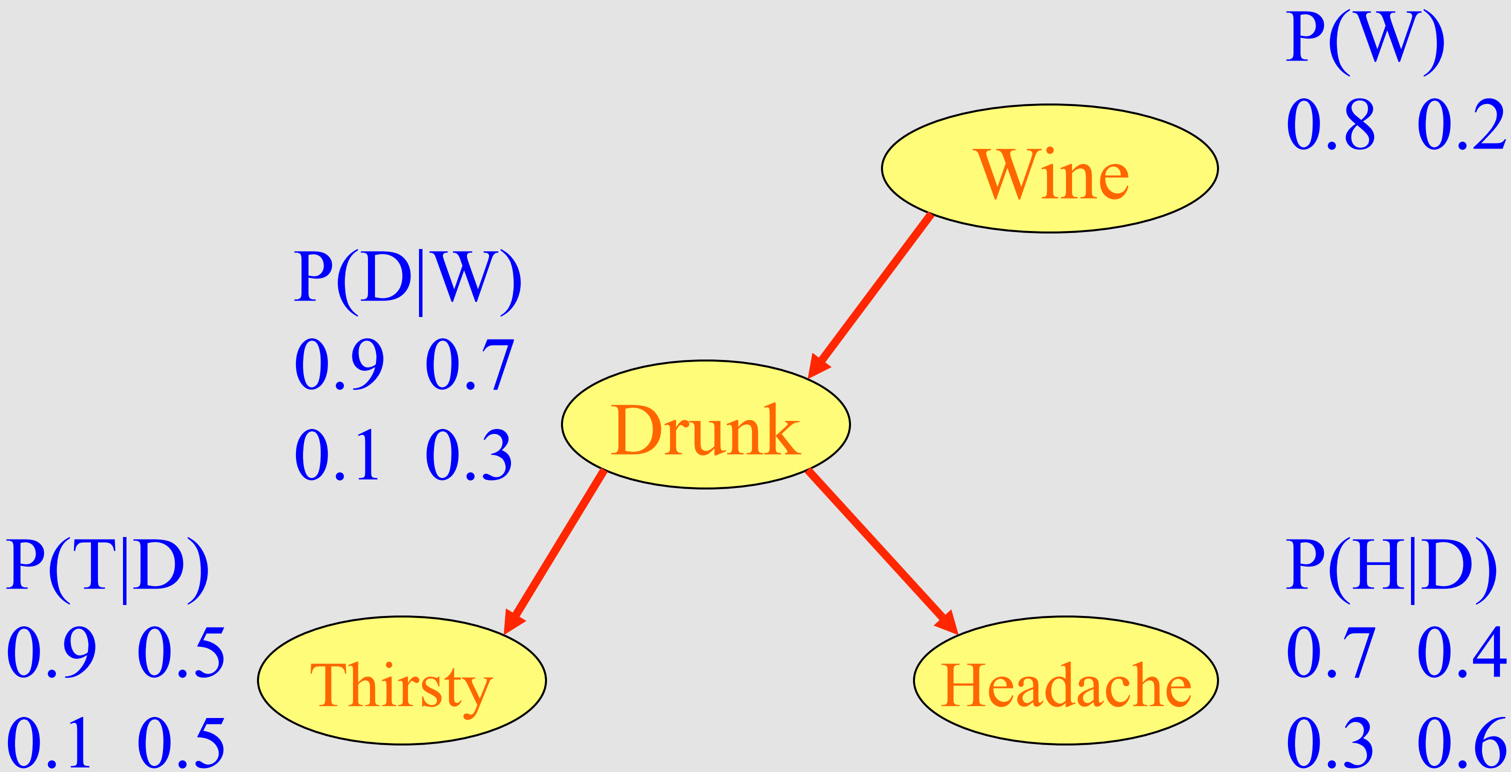
- The topology of the network represents the dependencies (and independencies) between the variables
- Conditional independence relations between variables or sets of variables are obtained by a criteria called *D-separation*

Parameters

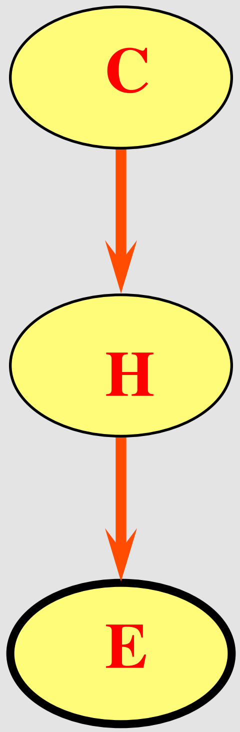
Conditional probabilities of each node given its parents.

- Root nodes: vector of prior probabilities
- Other nodes: matrix of conditional probabilities

For the example



Inference



Given certain evidence, E, estimate the posterior probabililty of the other variables, H, C

Inference

There are several inference algorithms:

- Variable elimination
- Message passing (Pearl's algorithm)
- Junction Tree
- Stochastic simulation
- …

- In the worst case it an NP-Hard problem, however given a sparse graph the state of the art algorithms are very efficient

Propagation Algorithm

Each node stores the vectors, π and λ, and the conditional probability matrix *P*

Probability propagation is done through a message passing mechanism in which each node *sends* messages to its parents and sons

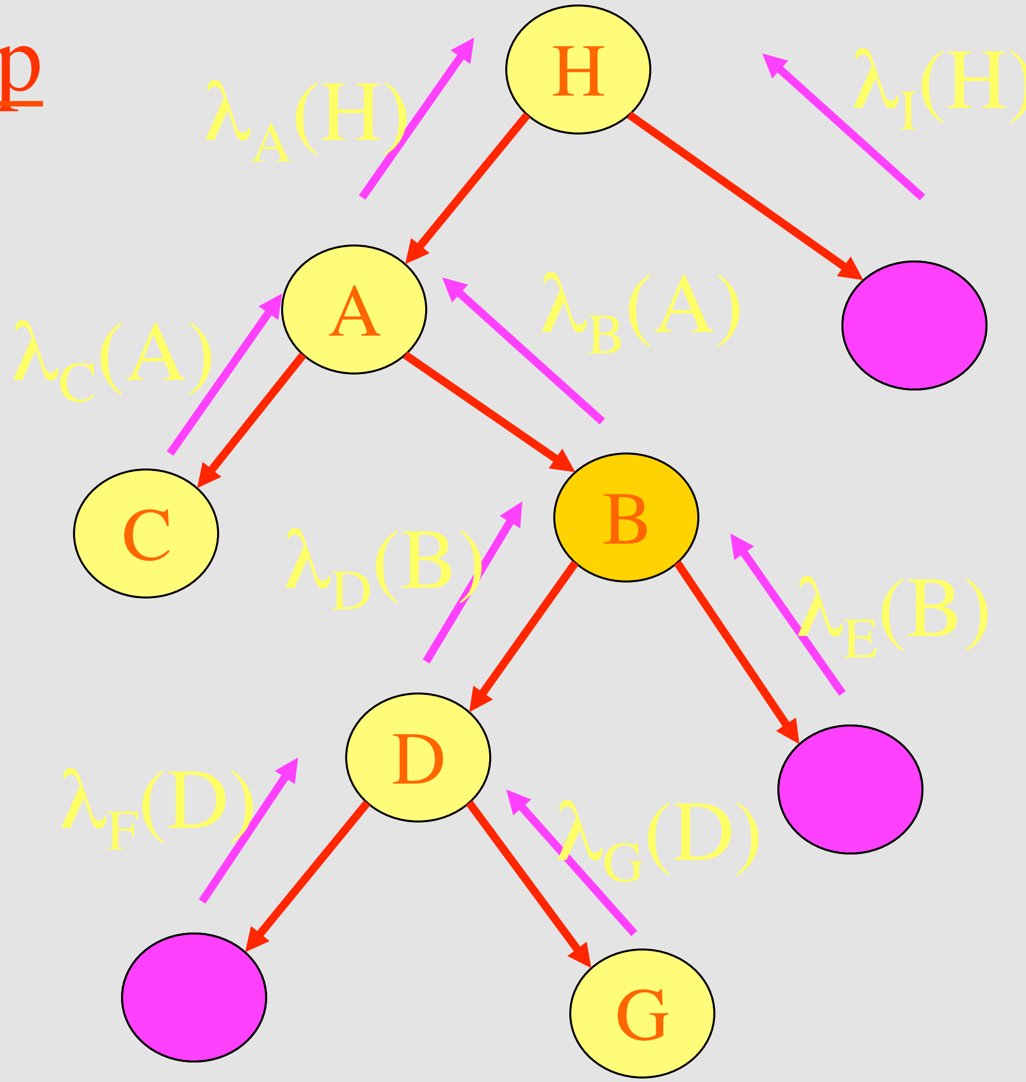
•Message to parent (upwards) -- node B to A:

$$\lambda_B(A_i) = \sum_j P(B_j|A_i)\lambda(B_j)$$

•Message to sons (downwards) -- node B to son S_k :

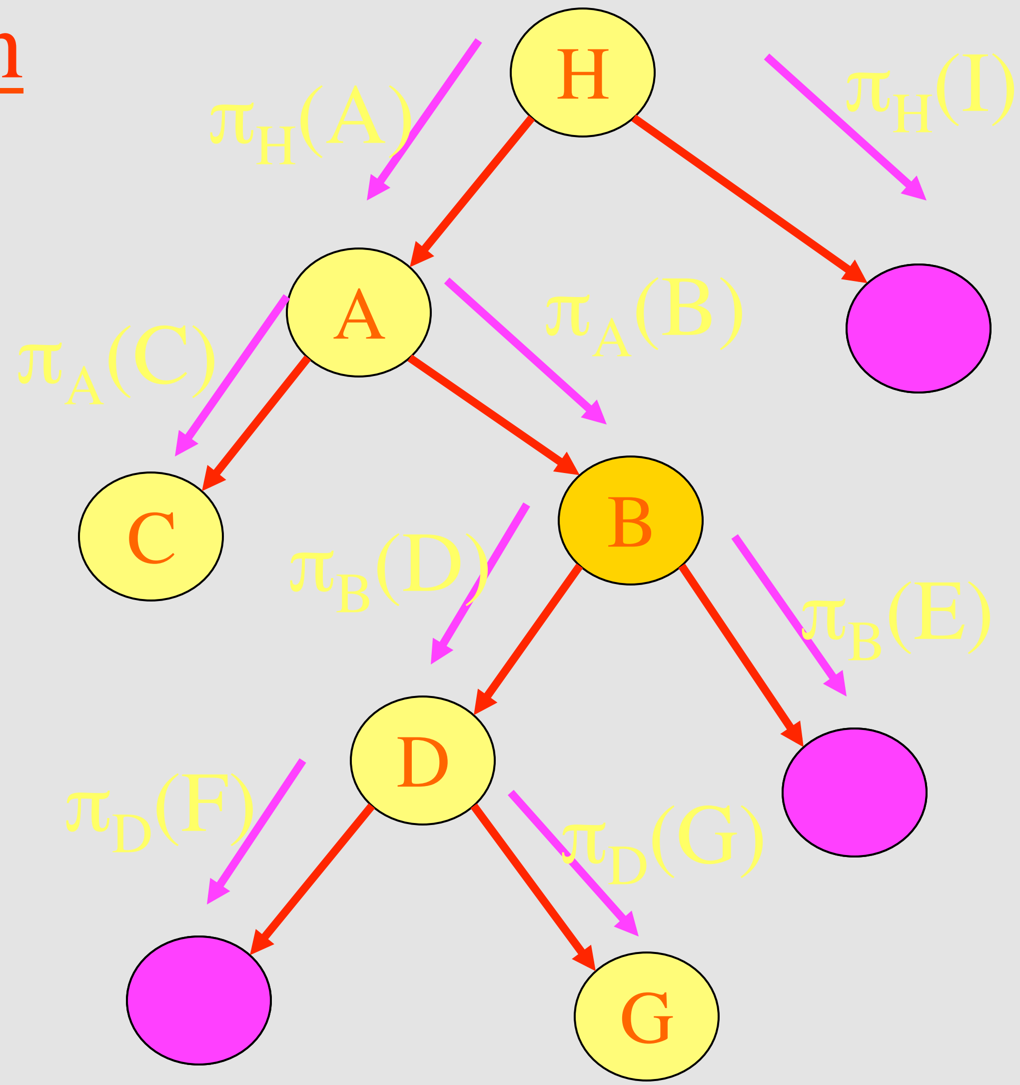
$$\pi_k(B_i) = \alpha\pi(B_j)\prod_{l\neq k}\lambda_l(B_j)$$

Bottom-up (λ)



H
A
B
C
D
G
$\lambda_A(H)$
$\lambda_I(H)$
$\lambda_C(A)$
$\lambda_B(A)$
$\lambda_D(B)$
$\lambda_E(B)$
$\lambda_F(D)$
$\lambda_G(D)$

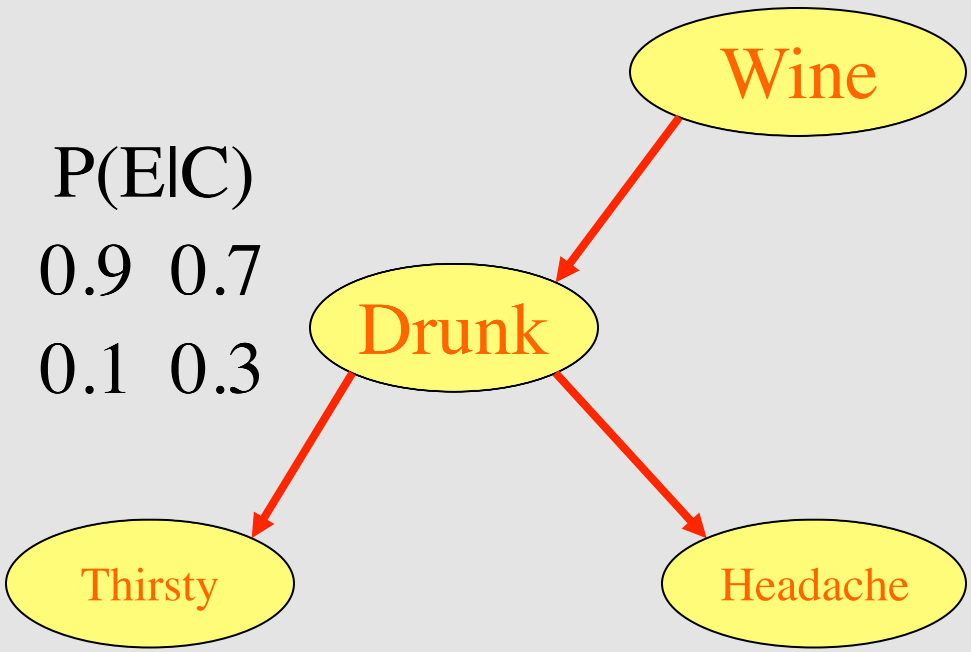
Top-down (π)



Example

P(C)
0.8 0.2

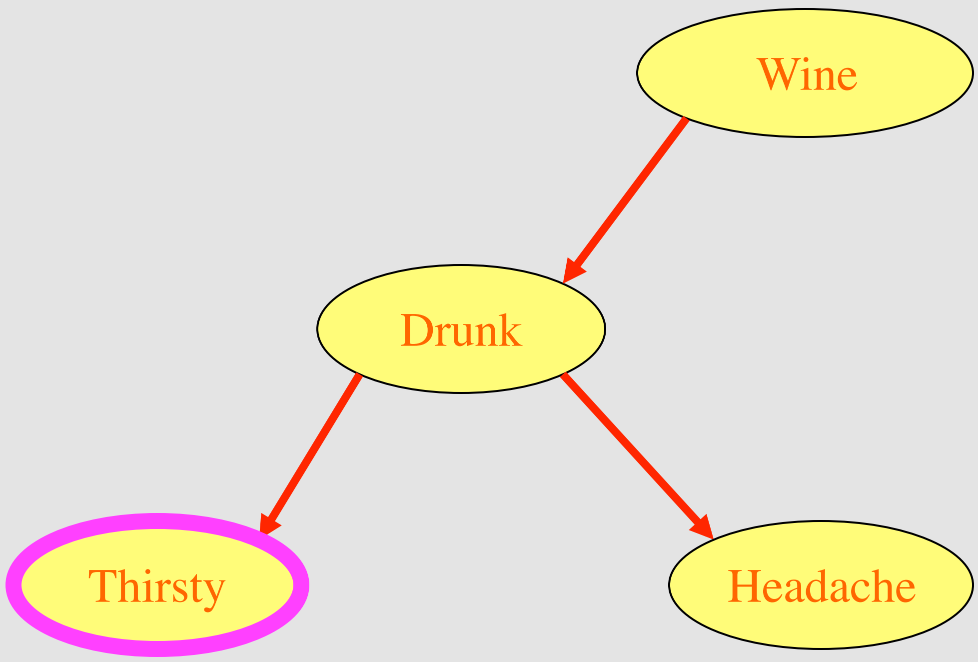
P(E|C)
0.9 0.7
0.1 0.3



P(F|E)
0.9 0.5
0.1 0.5

P(D|E)
0.7 0.4
0.3 0.6

Example



T=true

λ=[1,0]

λ=[1,1]

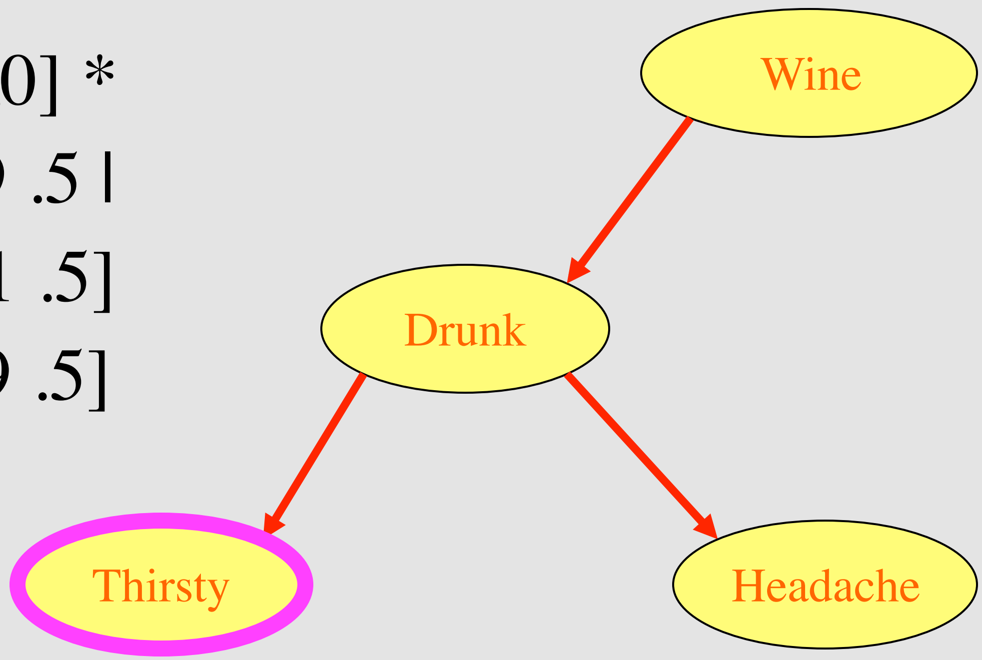
Example

λ_F= [1,0] *

[.9 .5 |

.1 .5]

= [.9 .5]



λ_D= [1,1] *

[.7 .4 |

.3 .6]

= [1 1]

P(F|E)

0.9 0.5

0.1 0.5

P(D|E)

0.7 0.4

0.3 0.6

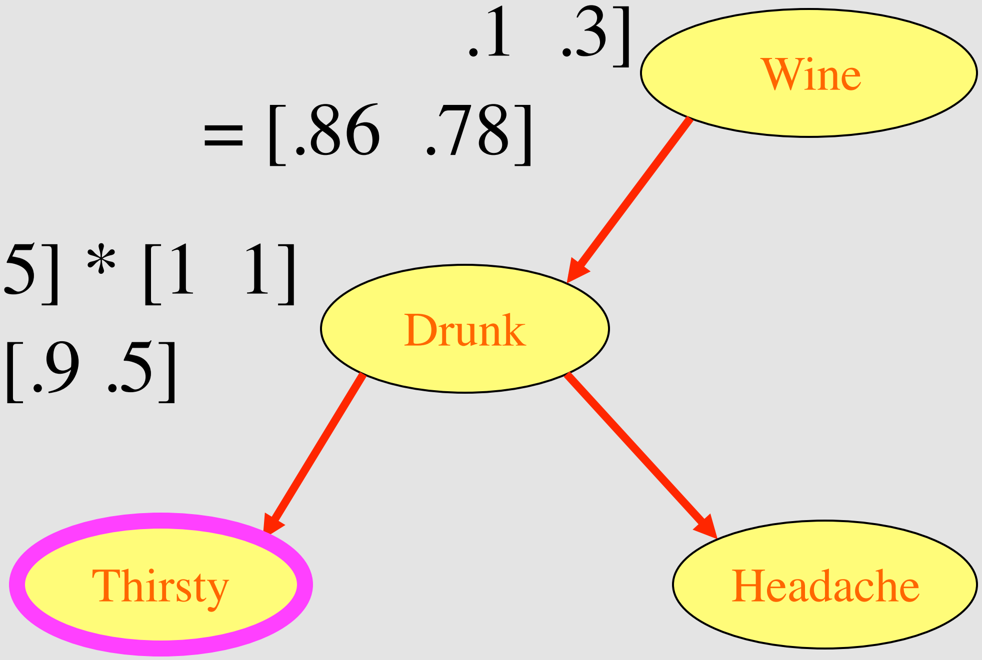
Example

$\lambda(C) = [.9 \;\; .5] * \begin{bmatrix} .9 & .7 \\ .1 & .3 \end{bmatrix}$

$= [.86 \;\; .78]$

$\lambda(E) = [.9 \;\; .5] * [1 \;\; 1]$

$= [.9 \;\; .5]$



P(E|C)

0.9	0.7
0.1	0.3

P(F|E)

0.9	0.5
0.1	0.5

P(D|E)

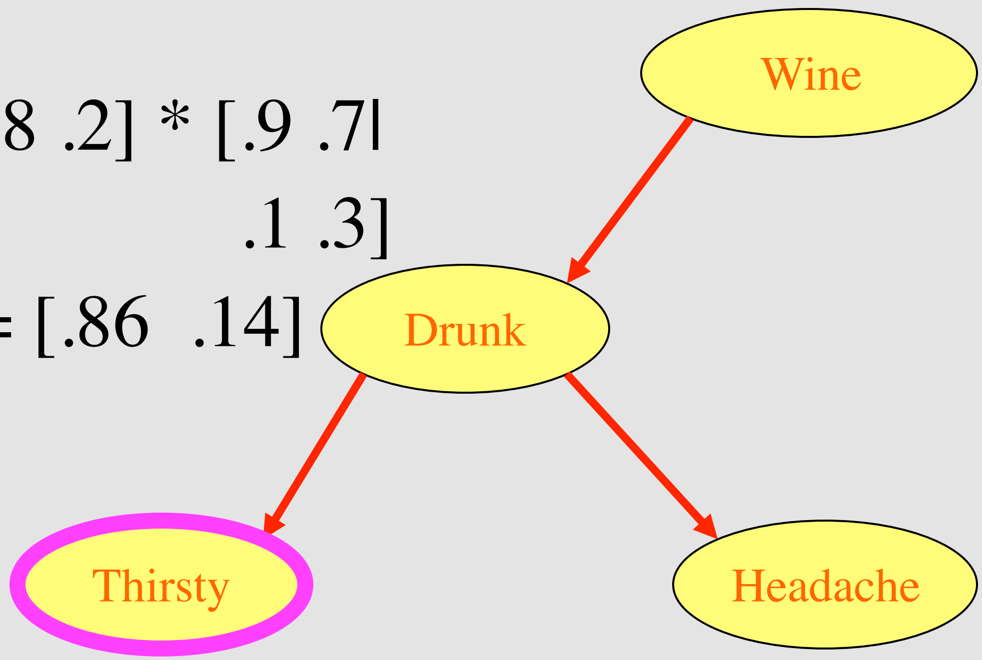
0.7	0.4
0.3	0.6

Example

$\pi(C) = [.8 \quad .2]$

$$\pi(E) = [.8 \;\; .2] * \begin{bmatrix} .9 & .7 \\ .1 & .3 \end{bmatrix}$$

$$= [.86 \quad .14]$$



P(E|C)

0.9	0.7
0.1	0.3

P(F|E)

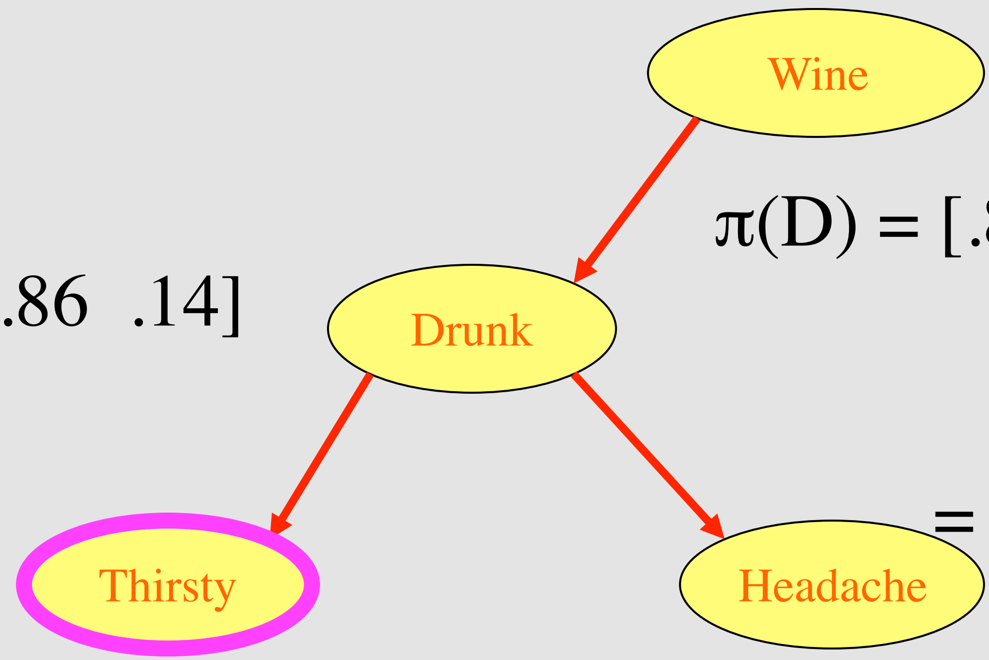
0.9	0.5
0.1	0.5

P(D|E)

0.7	0.4
0.3	0.6

Example

$\pi(C) = [.8 \quad .2]$



$\pi(E) = [.86 \quad .14]$

$\pi(D) = [.86 \ .14] * [.9 \quad .5]$

$[.7 \ \ .4|$

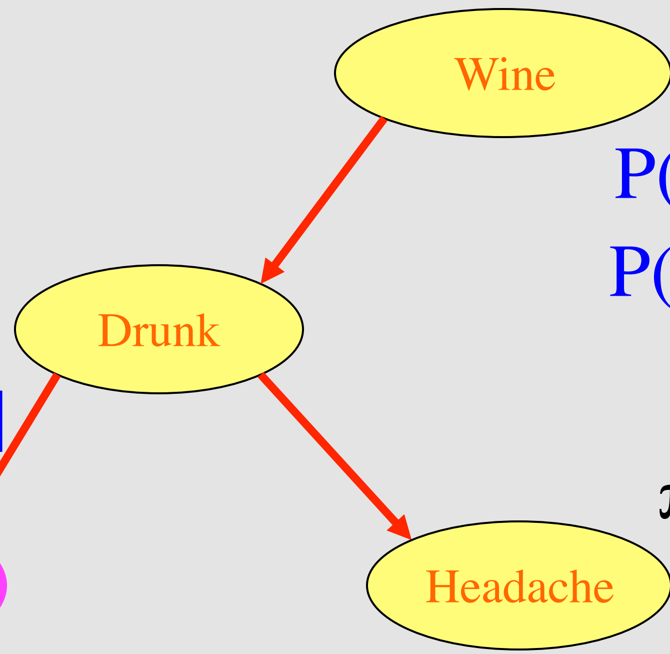
$.3 \ \ .6]$

$= [.5698 \quad .2742]$

$P(D|E)$

0.7	0.4
0.3	0.6

Example



$\pi(C) = [.8 \;\; .2]$

$\lambda(C) = [.86 \;\; .78]$

$P(C)=\alpha[.688 \;\; .156]$

$P(C)= [.815 \;\; .185]$

$\pi(E) = [.86 \;\; .14]$

$\lambda(E) = [.9 \;\; .5]$

$P(E)=\alpha[.774 \;\; .070]$

$P(E)= [.917 \;\; .083]$

$\pi(D) = [.57 \;\; .27]$

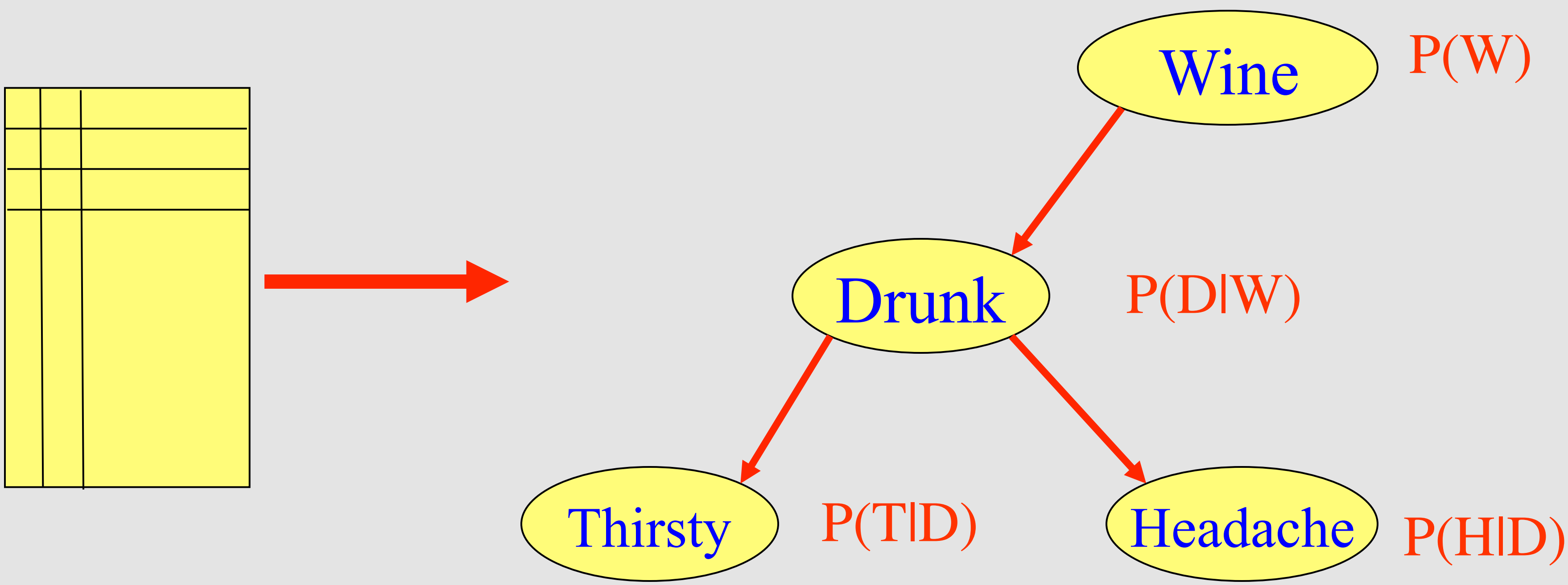
$\lambda(D)=[1,1]$

$P(D)=\alpha[.57 \;\; .27]$

$P(D)= [.67 \;\; .33]$

Learning

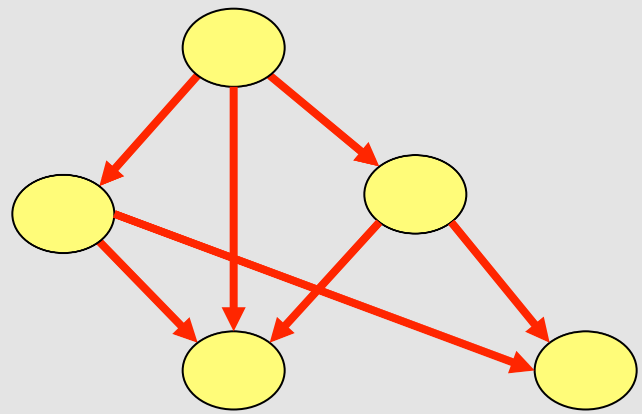
- Learning in Bayesian networks can be divided into two aspects:
  - Structure Learning
  - Parameter Learning



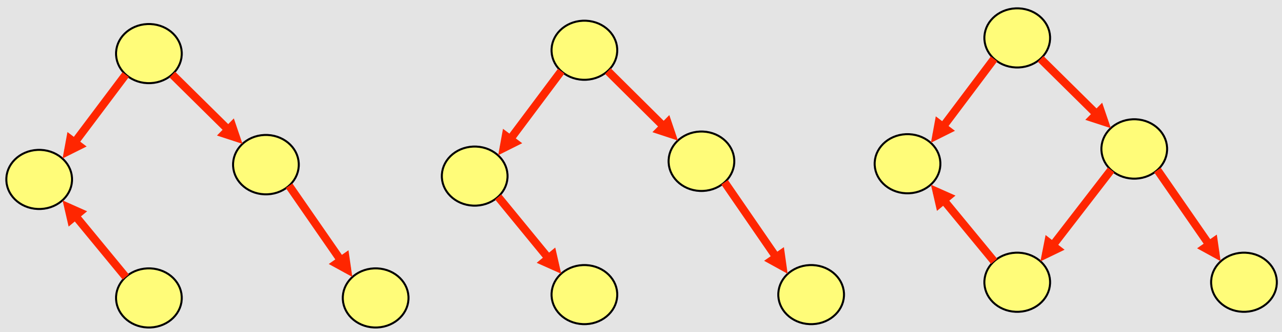
Structure Learning

Two general schemes:

Independence tests



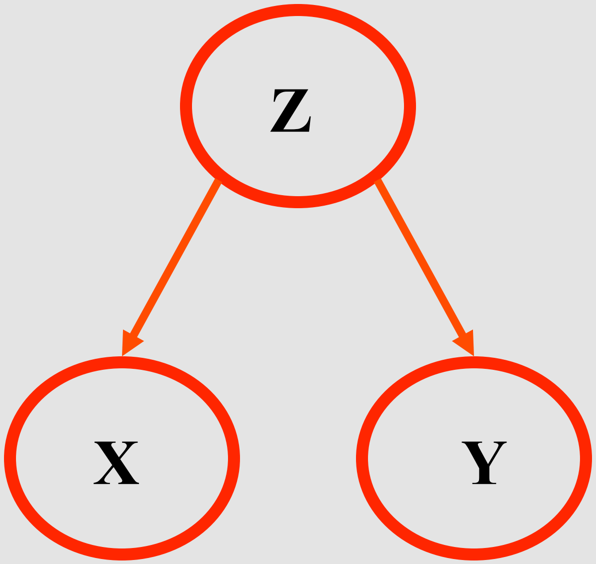
Search and score

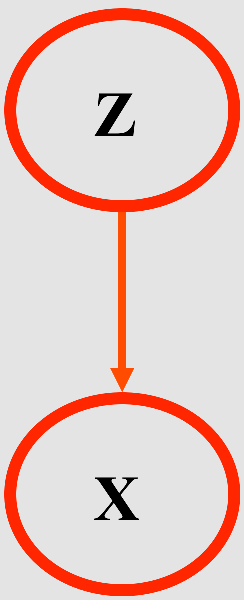


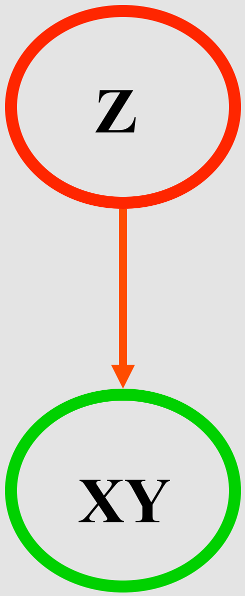
Structural Improvement

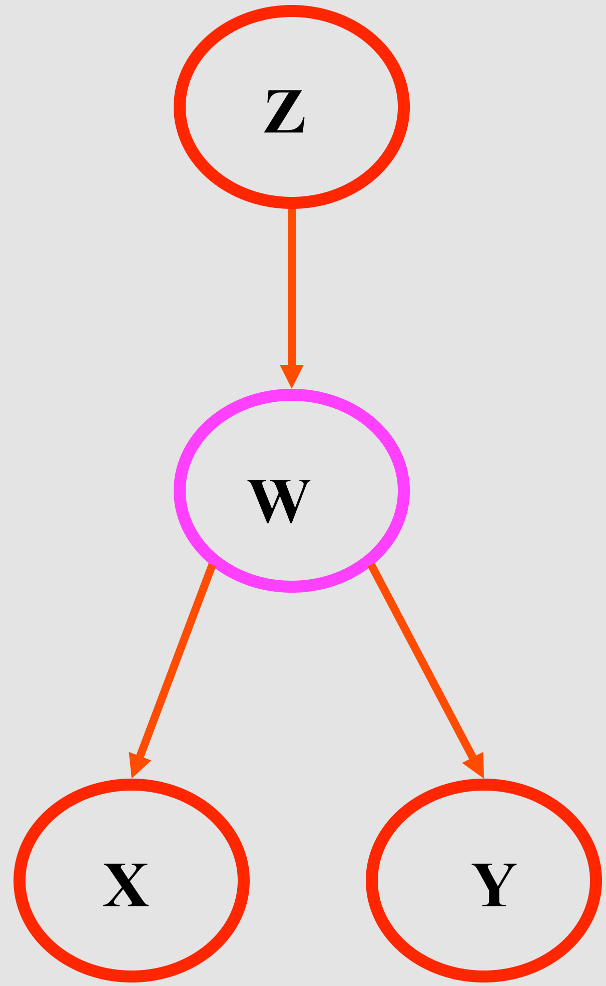
- Learning techniques require a large amount of data to obtain good models; an alternative is to combine expert knowledge and data
- We propose a method that starts from a subjective structure (given by an expert) and then improves it with data
- Assuming a tree structure, the conditional independence of child nodes given its parent are verified; if they are not independent there are 3 alternatives:
  - Node elimination
  - Node combination
  - Node insertion

Structural improvement









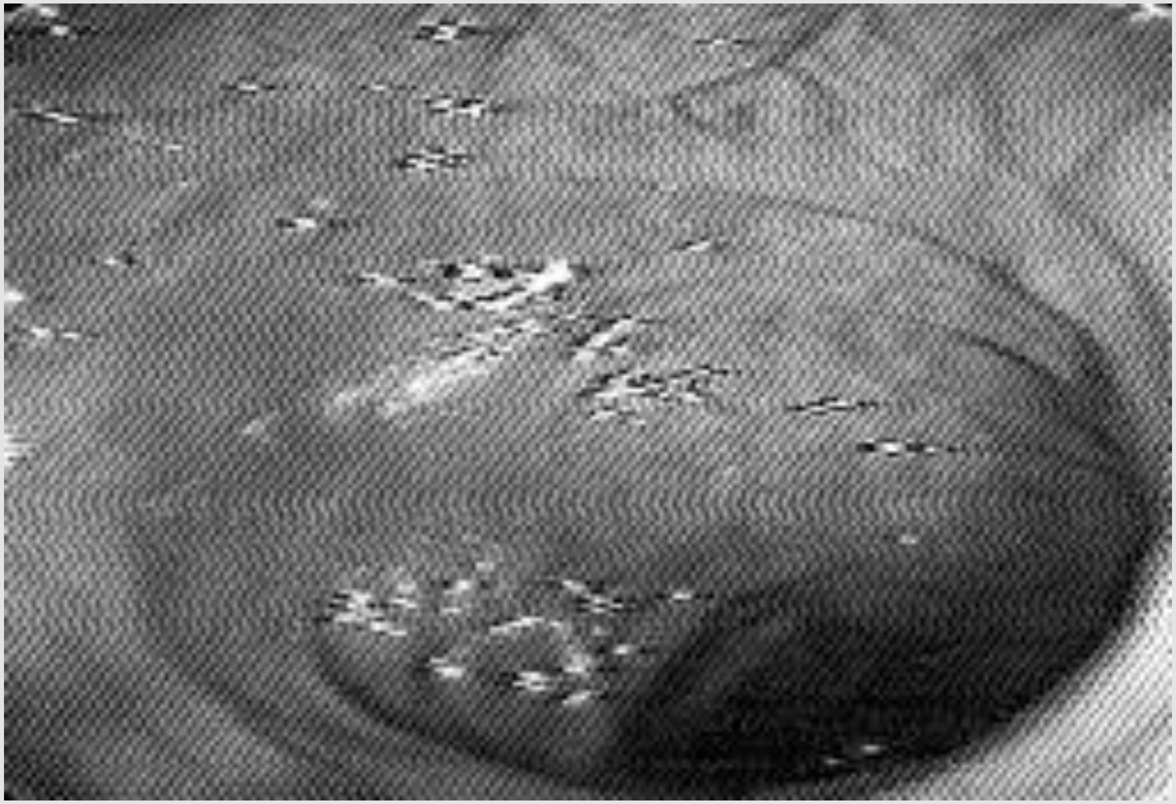
Algorithm

1. Build an initial BN structure based on expert knowledge
2. Repeat until the model can not be improved (based on the MDL principle):
   a. Eliminate redundant attributes
   b. Eliminate/Join dependant attributes
   c. Improve discretization of continuous attributes
3. Test on different data (cross validation)

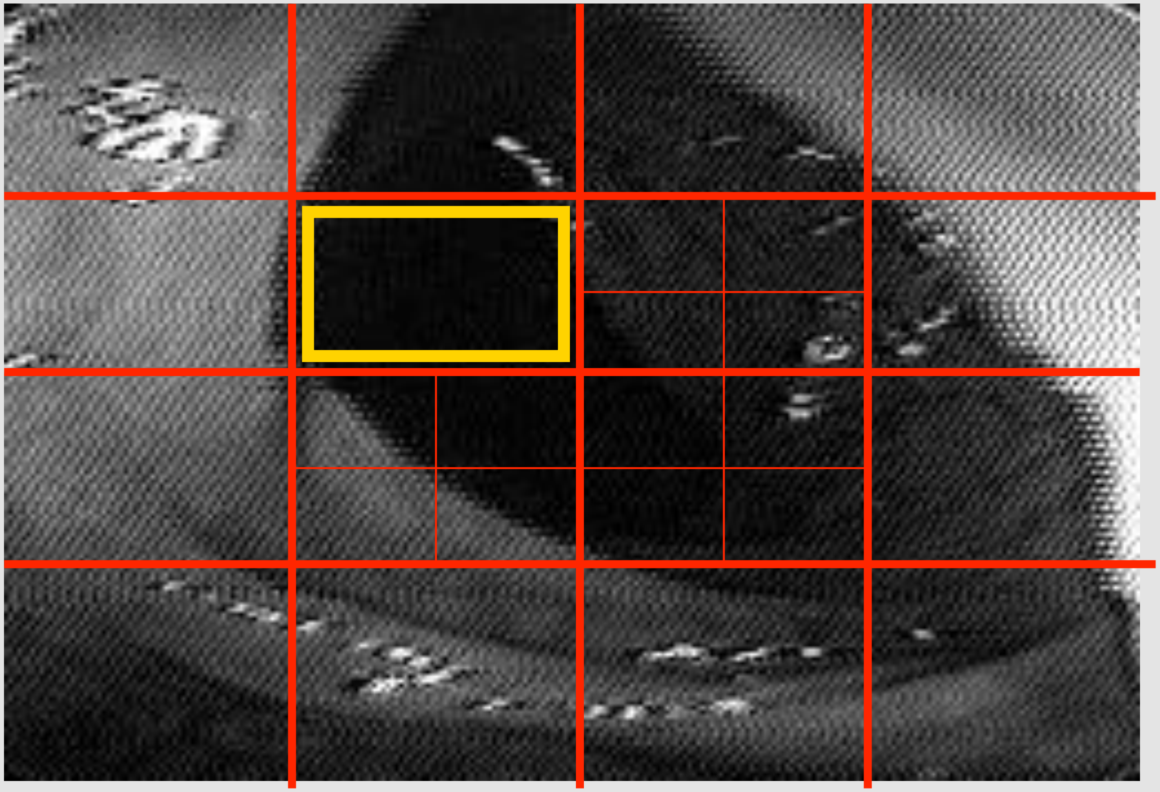
Endoscopy

- Endoscopy is a tool for direct observation of the human digestive system
- Navigating an endoscope is difficult due to the variability and dynamics of the human colon
- Thus, it is desirable to build a semi-automatic system that can assist an endoscopist
- The main challenge is to recognize the "objects" in endoscopy images which can be confused, such as "*lumen*" & "*diverticula*"
- The low-level vision algorithms can fail so we propose a Bayesian network that combines the information and arrives to final decisions

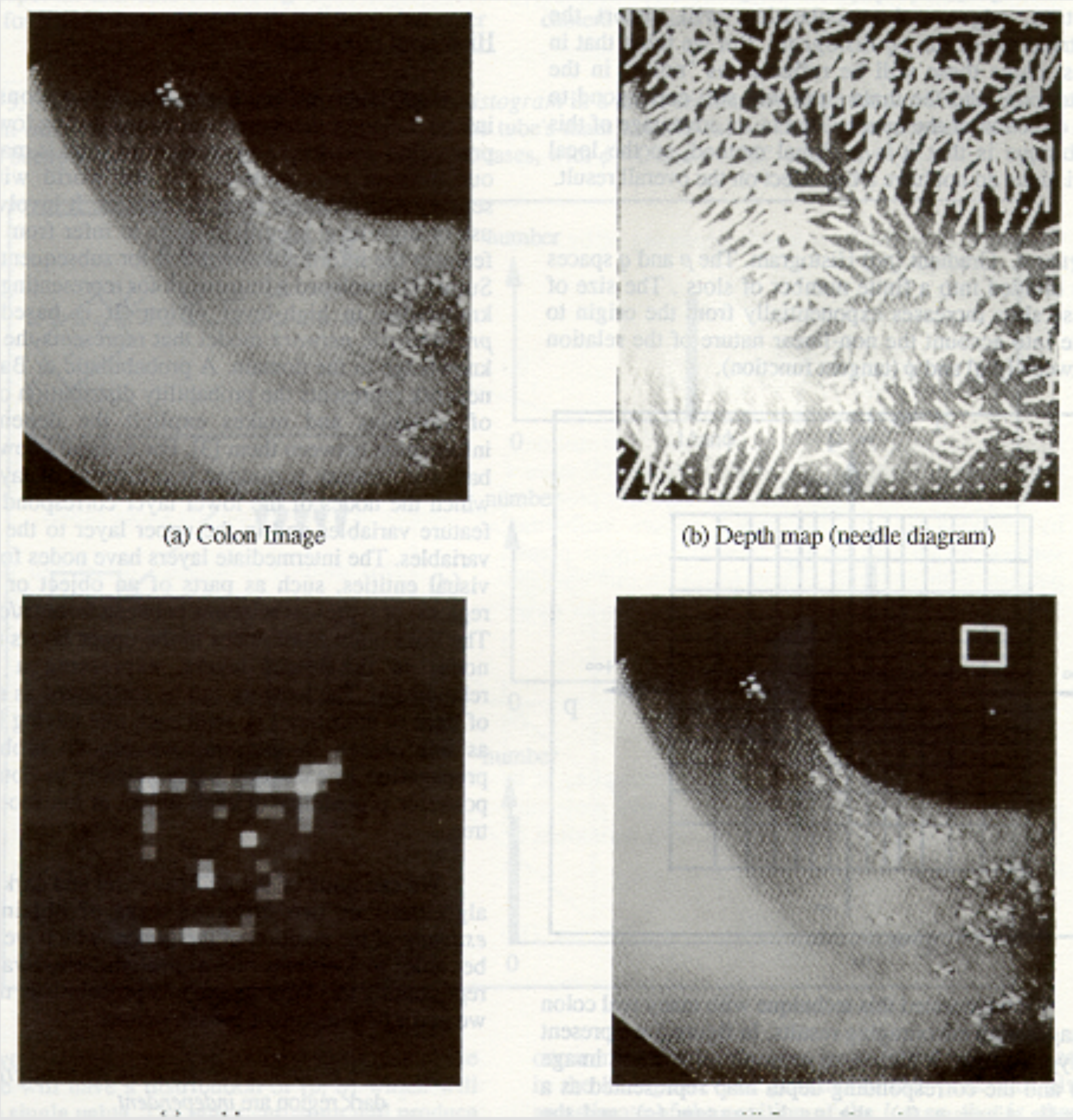
Colon Image



Low level features – dark region



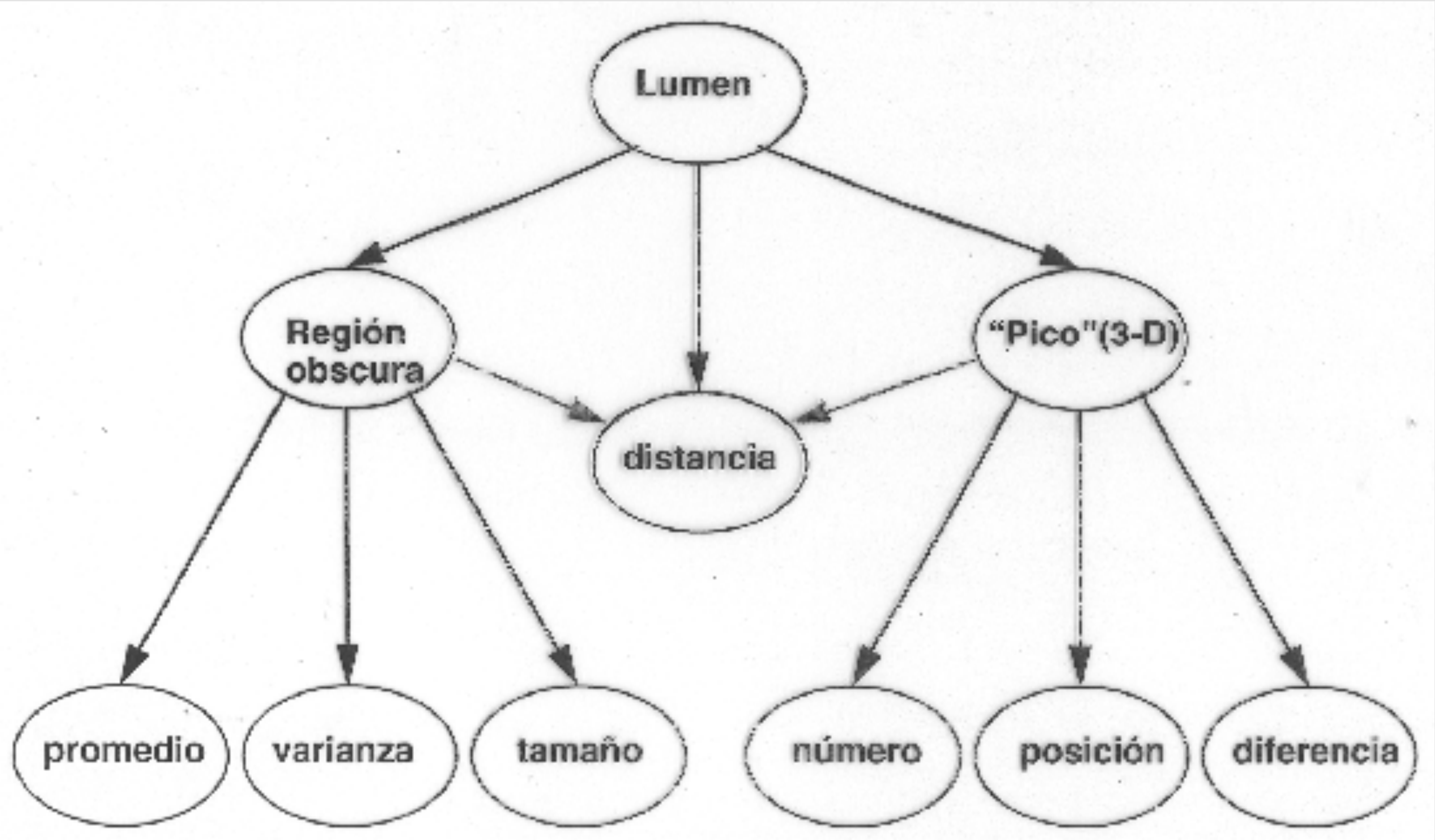
Low level features – shape from shading (pq histogram)



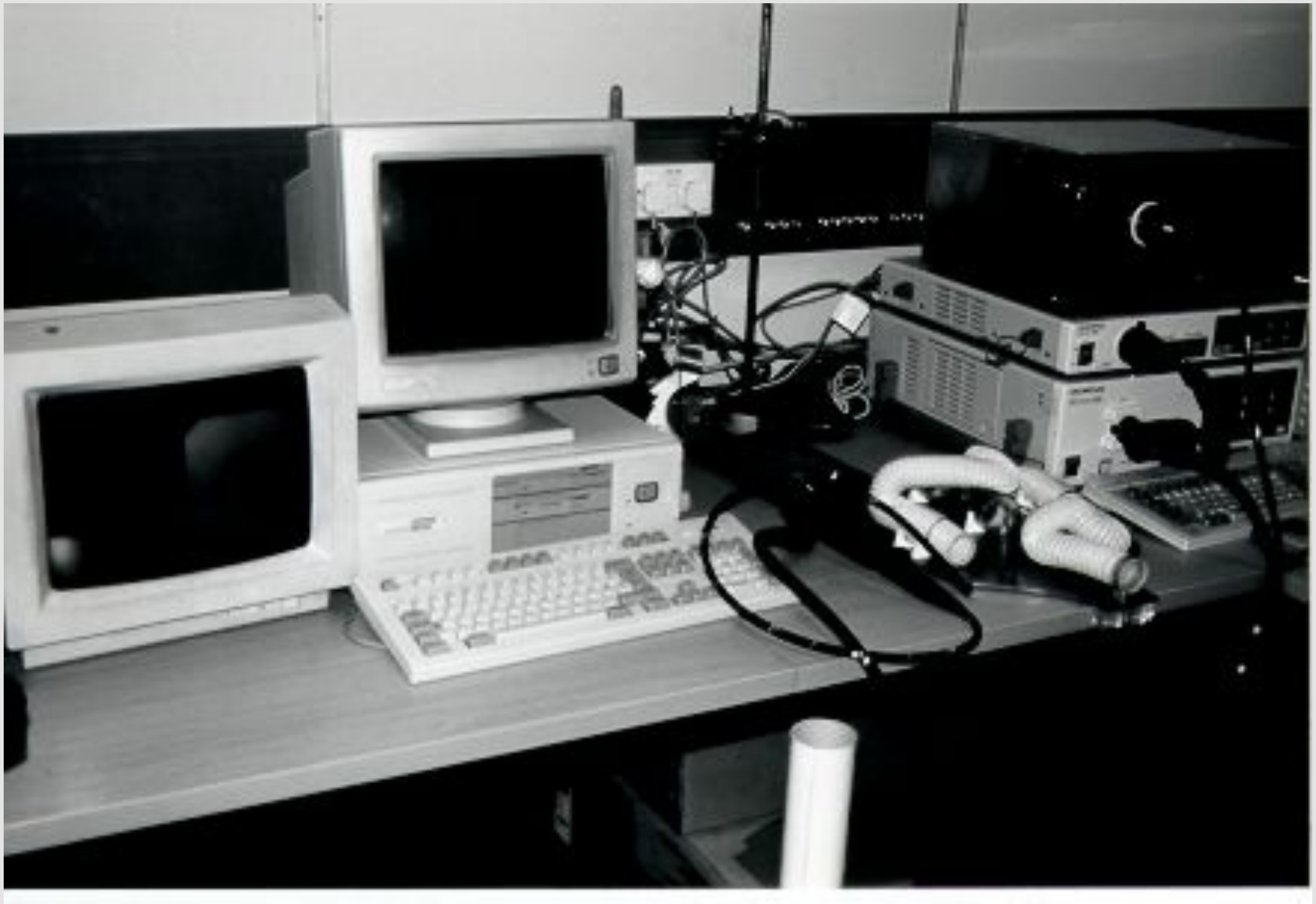
Model Construction

- The structure of the BN was built with the help of an expert endoscopist
- Later it was improved based on the structural improvement technique
- Parameters were learned from videos of *real* colonoscopy sessions

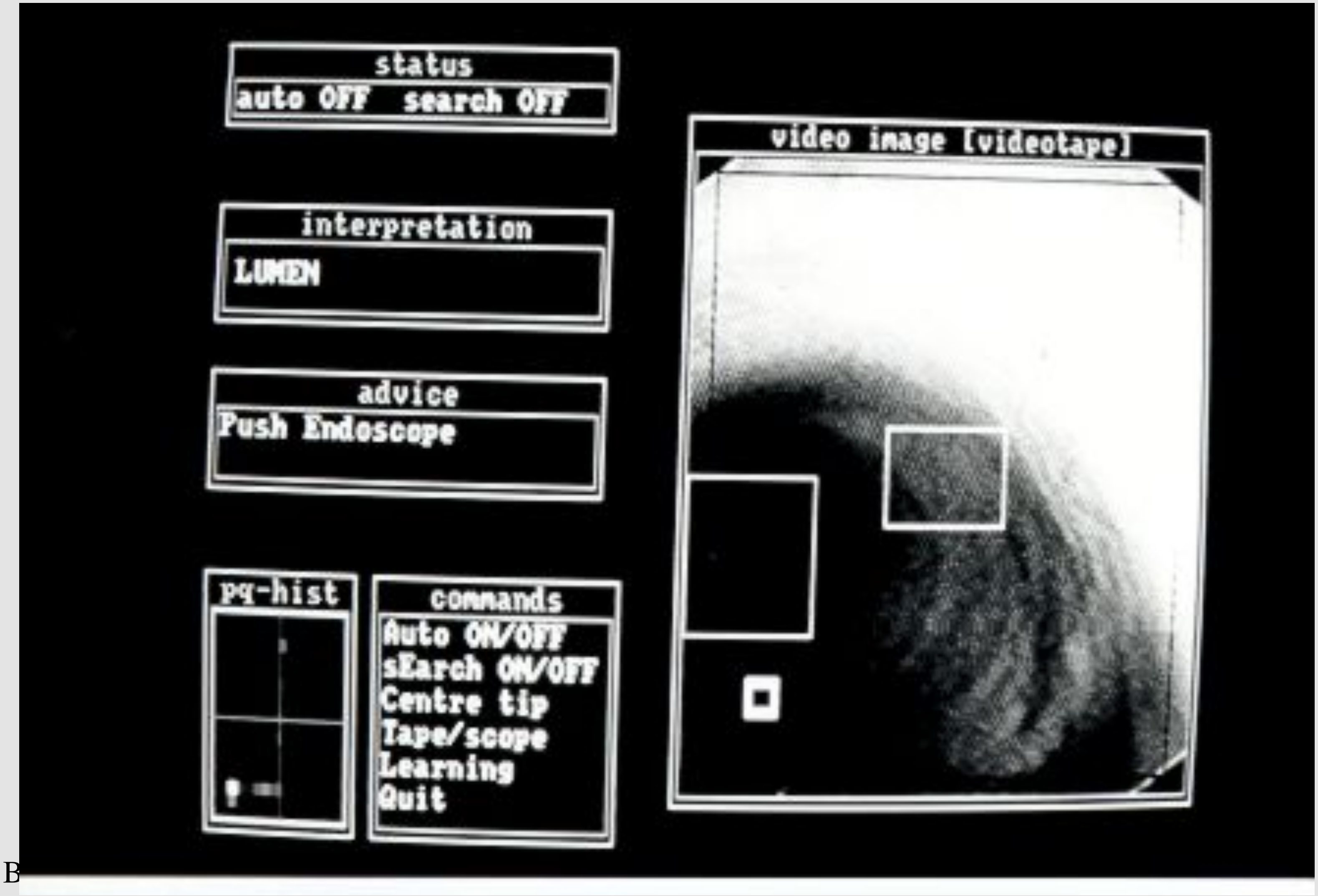
BN for endoscopy (partial)



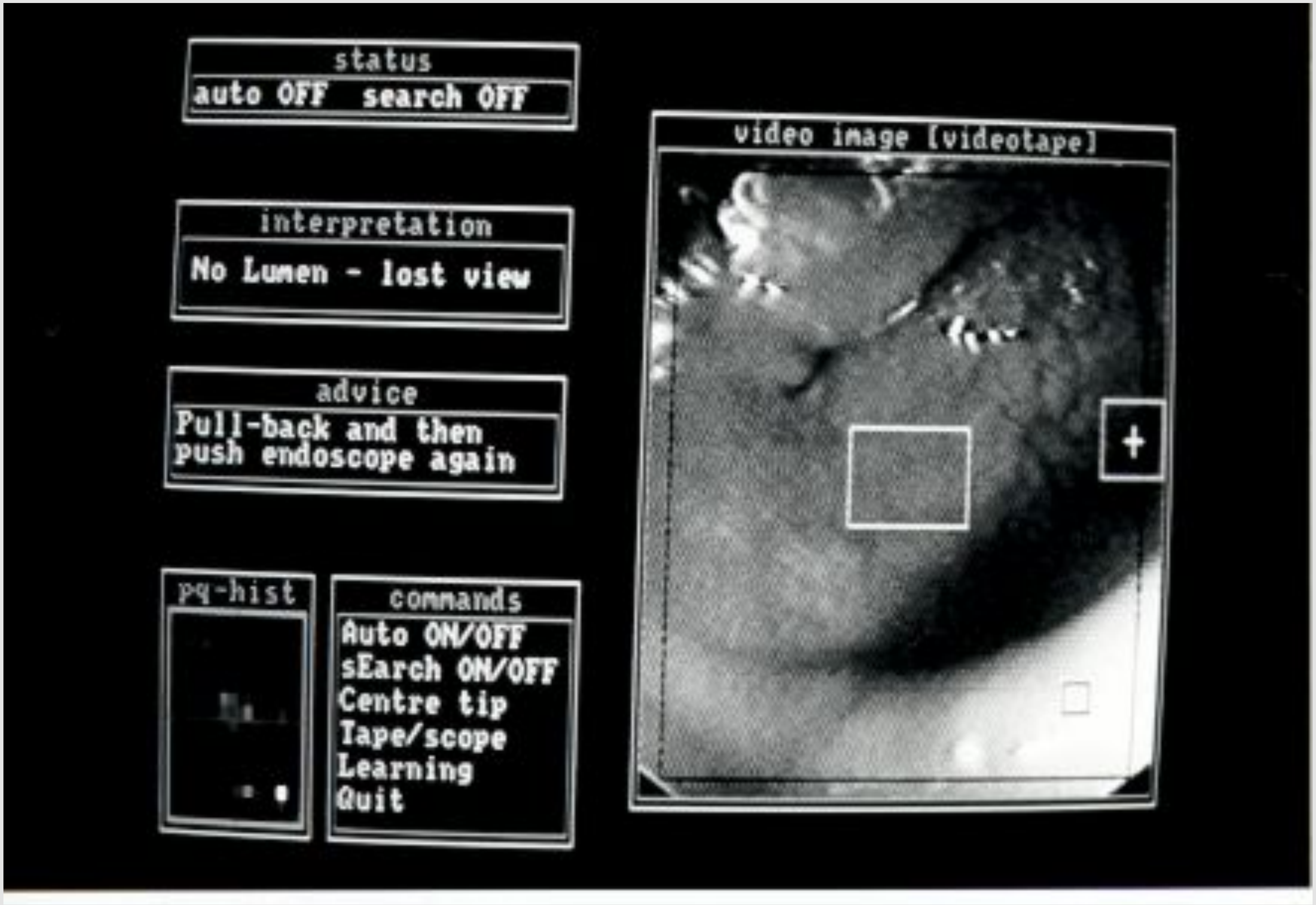
Semi-automatic Endoscope



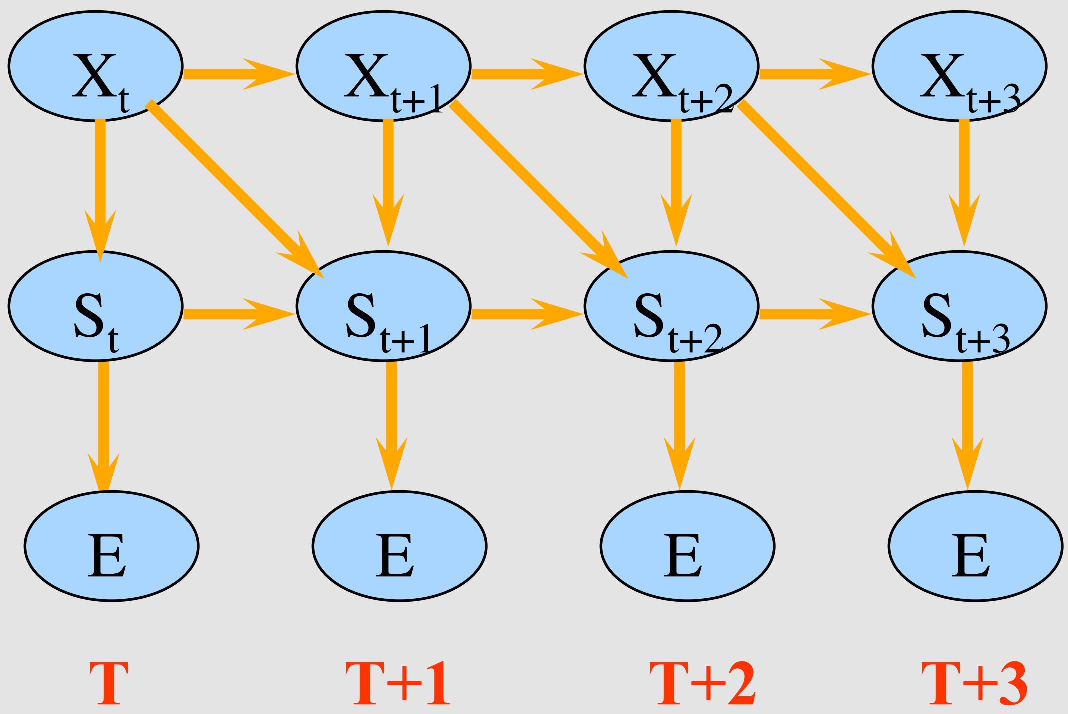
Endoscope navigation system: example 1



Endoscope navigation system: example 2



Dynamic Bayesian Networks



Temporal Nodes Bayesian Networks (RBNT)

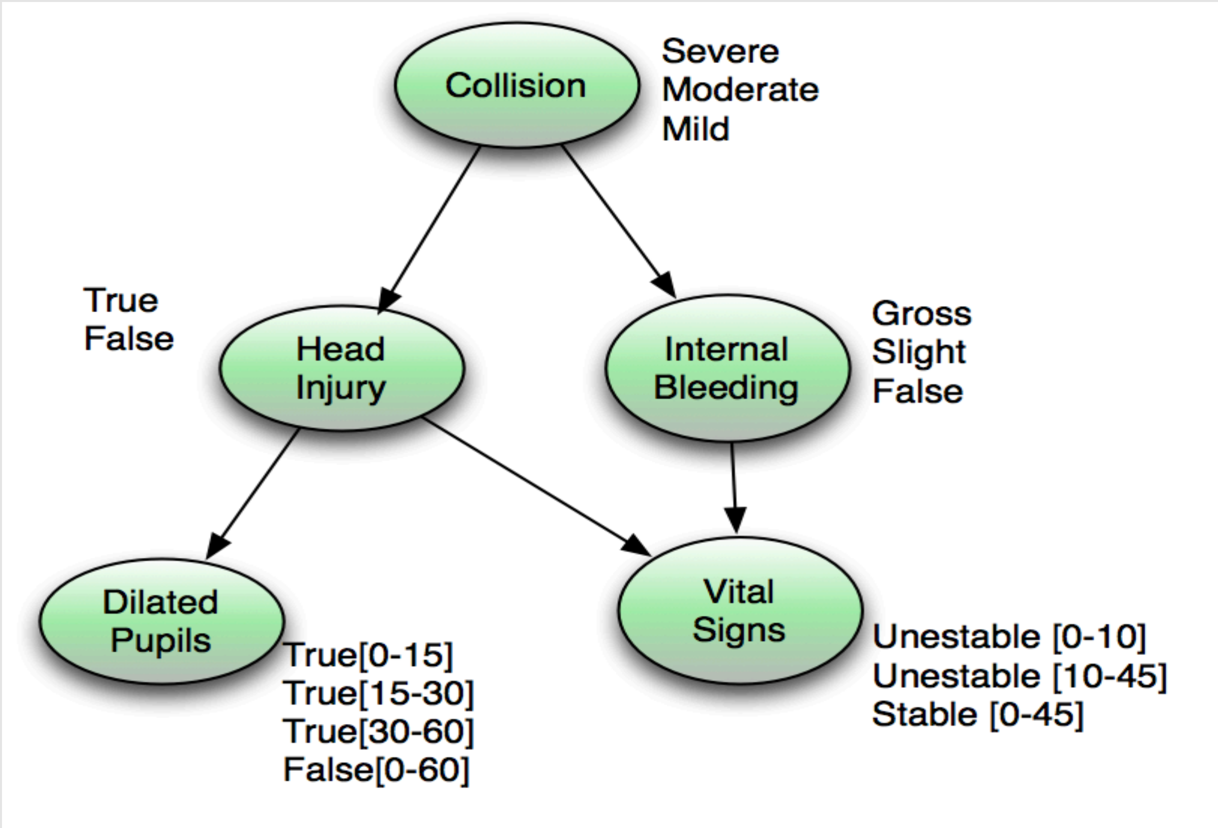
An alternative to Dynamic Bayesian Networks to model dynamic processes with uncertainty

Temporal information is within the nodes in the model, which represent the time of occurrance of certain events

The links represent temporal-causal relation

Adequate for applications in which there are few state changes in the temporal range

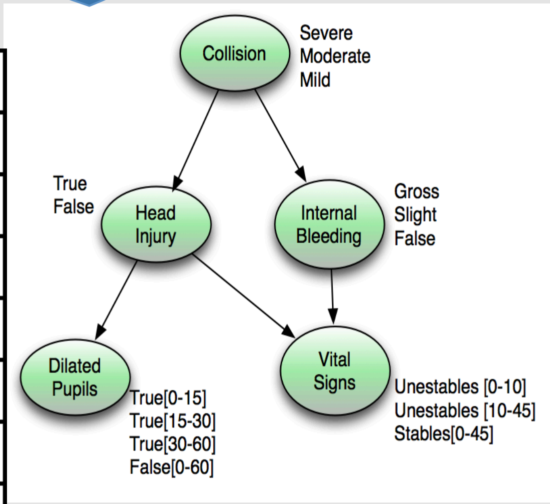
Example



Collision
Severe
Moderate
Mild
True
False
Head Injury
Internal Bleeding
Gross
Slight
False
Dilated Pupils
True[0-15]
True[15-30]
True[30-60]
False[0-60]
Vital Signs
Unestable [0-10]
Unestable [10-45]
Stable [0-45]

Learning TNBN

Collision	Head Inj.	Internal Ble.	Pupils Dil.	Vital Signs
severe	yes	gross	15	10
severe	yes	slight	25	20
mild	no	false	25	----
mild	no	false	21	----
moderate	yes	slight	----	20
moderate	yes	false	----	----
mild	no	false	----	----



Learning Algorithm

1. Define initial intervals for the temporal nodes
2. Learn the structure and parameters using standard techniques
3. Improve the temporal intervals using on clustering; selection based on predictions on validation data (Brier score)

Steps 2 and 3 can be repeated until convergance

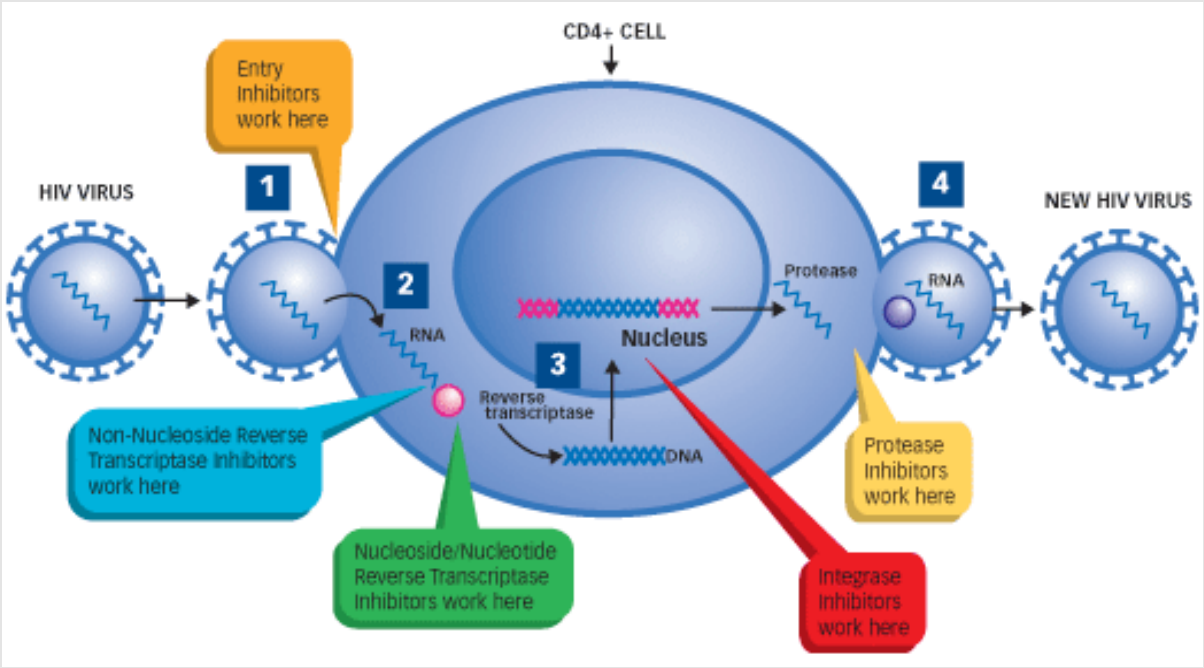
HIV

- HIV among fastest evolving organisms
- The HIV evolves (among other pressures) in response to antiretroviral therapy
- Although mutations conferring drug resistance are mostly known, the dynamics of the appearance chain of mutations remains poorly understood
- We use TNBN for modeling the relationships between antiretroviral drugs and HIV mutations, in order to analyze temporal occurrence of specific mutations in HIV that may lead to drug resistance.

Mutational Networks

- Mutational networks are "drug-associated mutational pathways in the protease gene, revealing the co-occurrence of mutations and its temporal relationships"
- If we could predict the most likely evolution of the virus in any host, then it would be plausible to select an appropriate antiretroviral regimen that prevents the appearance of mutations, effectively increasing HIV control.

Antiretrovirals



Antiretroviral therapy (ART) generally consists of well-defined combinations of three or four ARV drugs in order to reduce the possibility of development of drug resistance mutations.

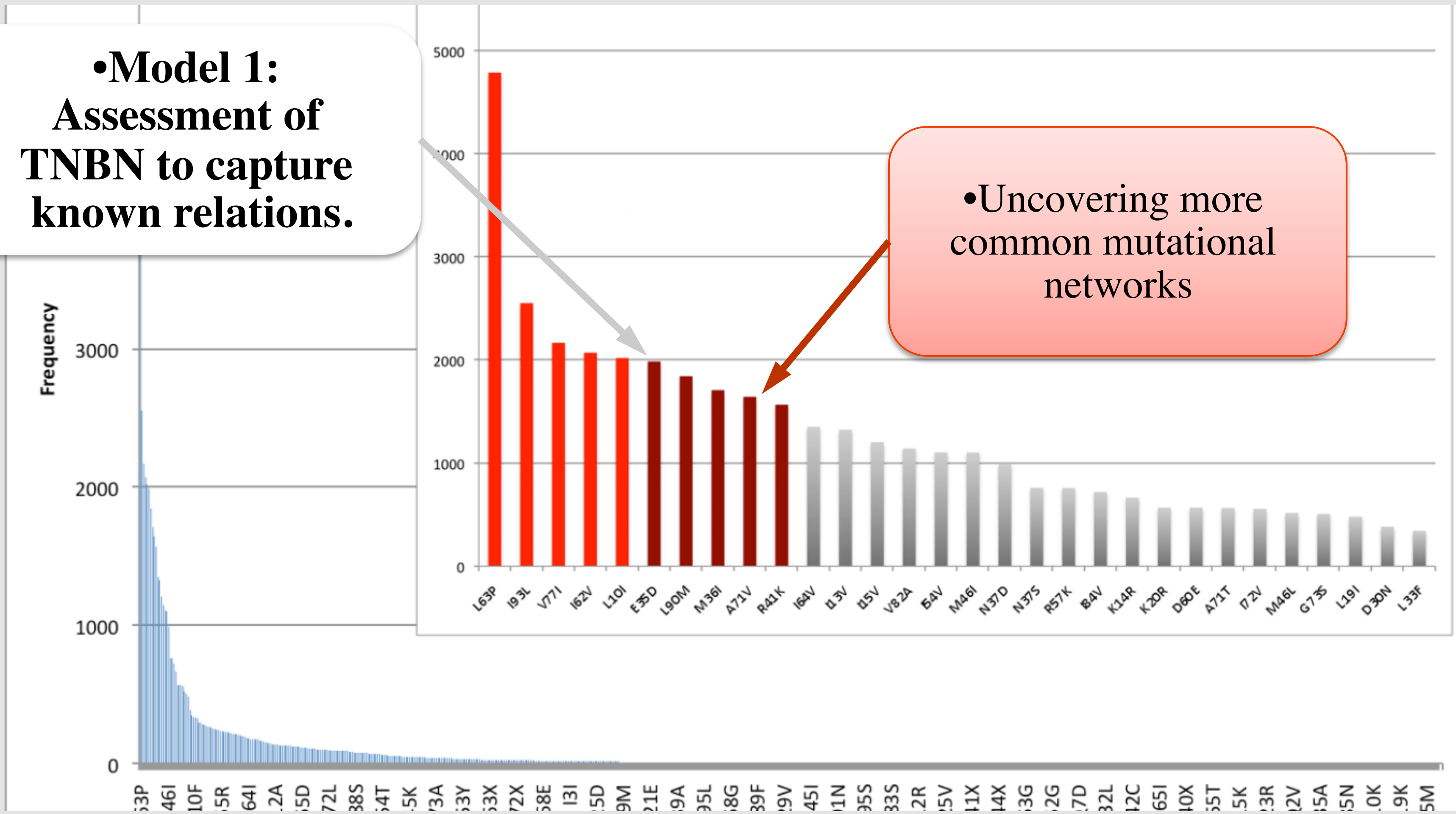
http://us.viramune.com/consumer/hiv-treatment

Experiments

- Data and preprocessing
  - HIV Stanford database (HIVDB) - HIV Drug Resistance Database
  - 2373 patients with subtype B was retrieved
  - Data retrieved contains a history consisting of a variable number of studies.

Patient	Initial Treatment	List of Mutations	Weeks
P1	LPV, FPV, RTV	L63P, L10I, V77I, I62V	15 30 10
P2	NFV, RTV, SQV	L10I V77I	25 45

Defining target mutations

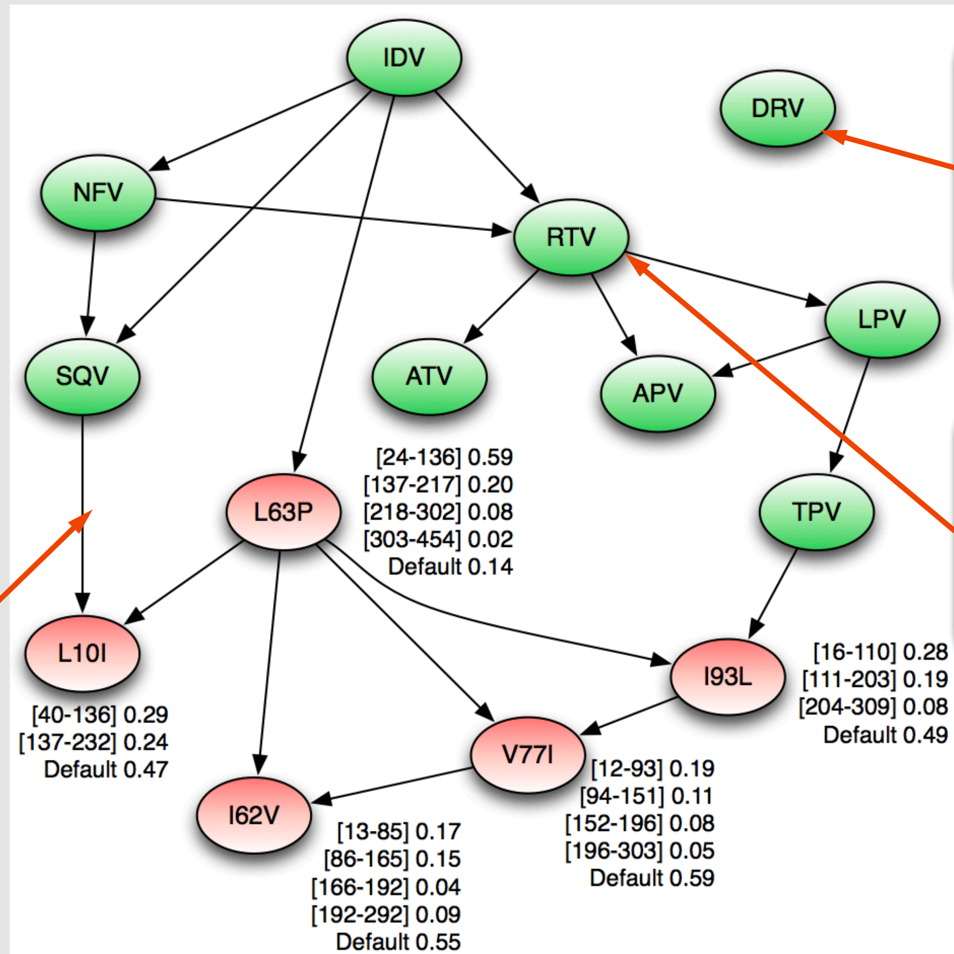


•Model 1: Assessment of TNBN to capture known relations.

•Uncovering more common mutational networks

Results: Model 1

Known relationships are appropiately captured



DRV isolation: New drug is hardly ever given as a first treatment

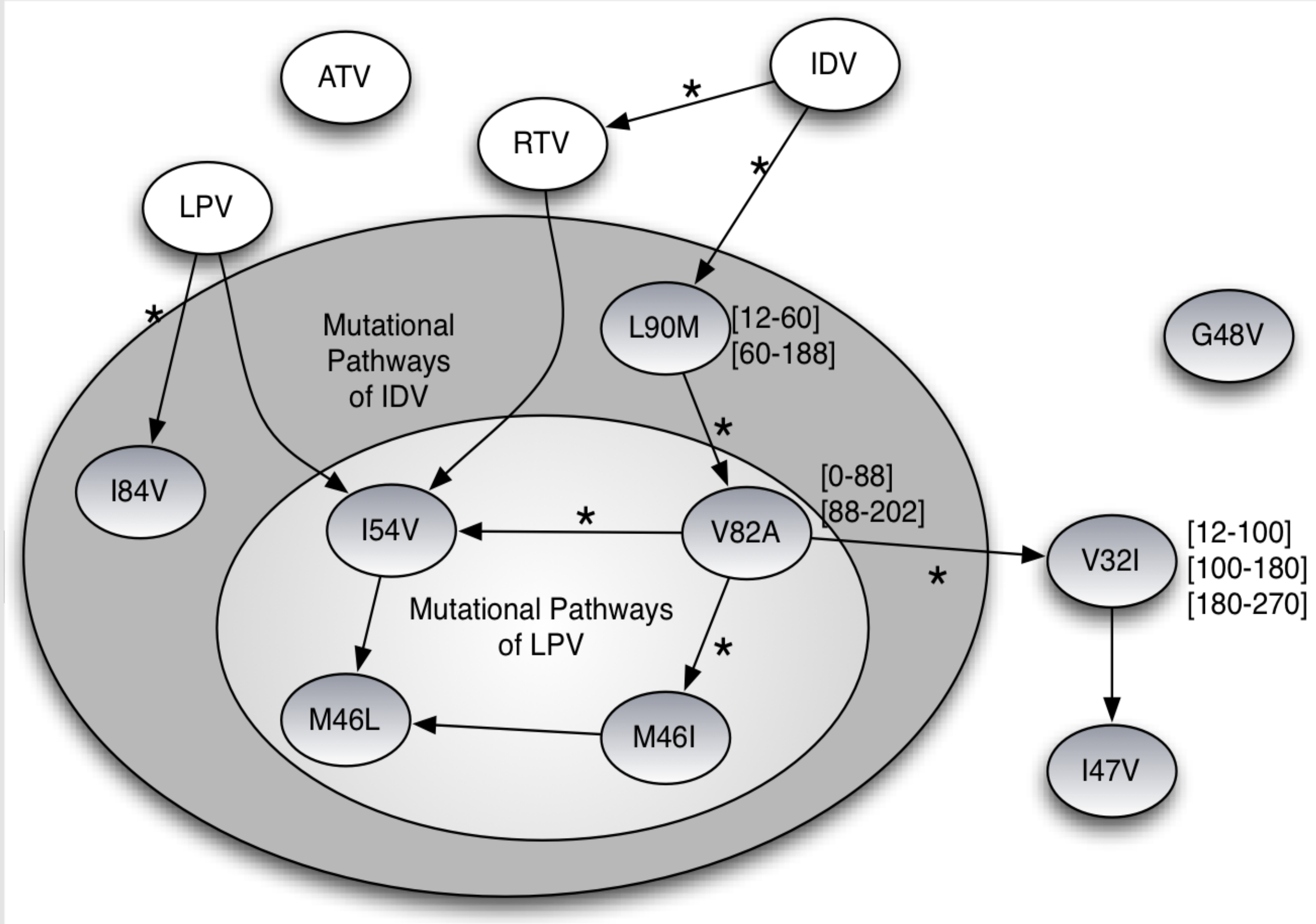
Role of RTV is often used in conjunction with other drugd (boosting)

Link between SQV and L10I

Model 2

- Use expert's information to select a subset of mutations and drugs of special importance.
- We used the Major HIV Drug Resistance Mutations and four drugs highly used in the past and nowadays.

Results



- The model was able to capture some mutational pathways already known (obtained by clinical experimentation).

LPV : M46I/L, I54V/T/A/S and V82T/F/S (Kempf et al., 2001) ,

IDV: V82A/T/F/S/M, M46I/L, I54V/T/A, I84V and L90M (Bélec et al., 2000; Descamps et al., 2005)

Markov Decision Processes

User Adaptation for Rehabilitation

Markov decision processes (MDPs)

- Ideal framework for planning under uncertainty.
- Main features:
  - Considers the uncertainty in the actions
  - Considers the utility of the plan
  - It allows to obtain optimal solutions
  - Considers uncertainty in the observations (POMDP)

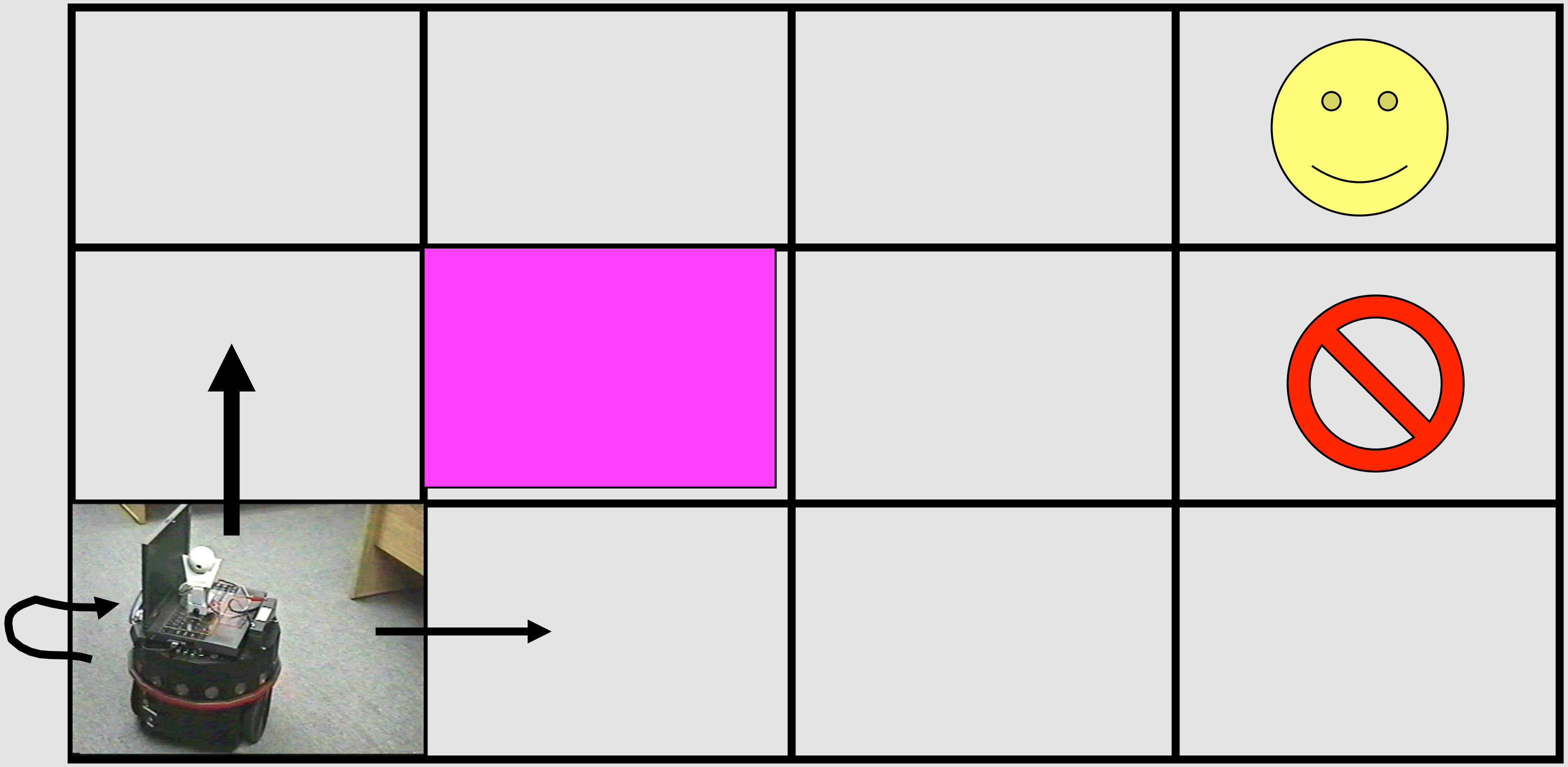
MDP

- Formally, a discrete MDP is defined by:
  - A finite set of states, S
  - A finite set of actions, A
  - A transition model, P (s’ | s, a)
  - A reward function for each state-action, r(s, a)

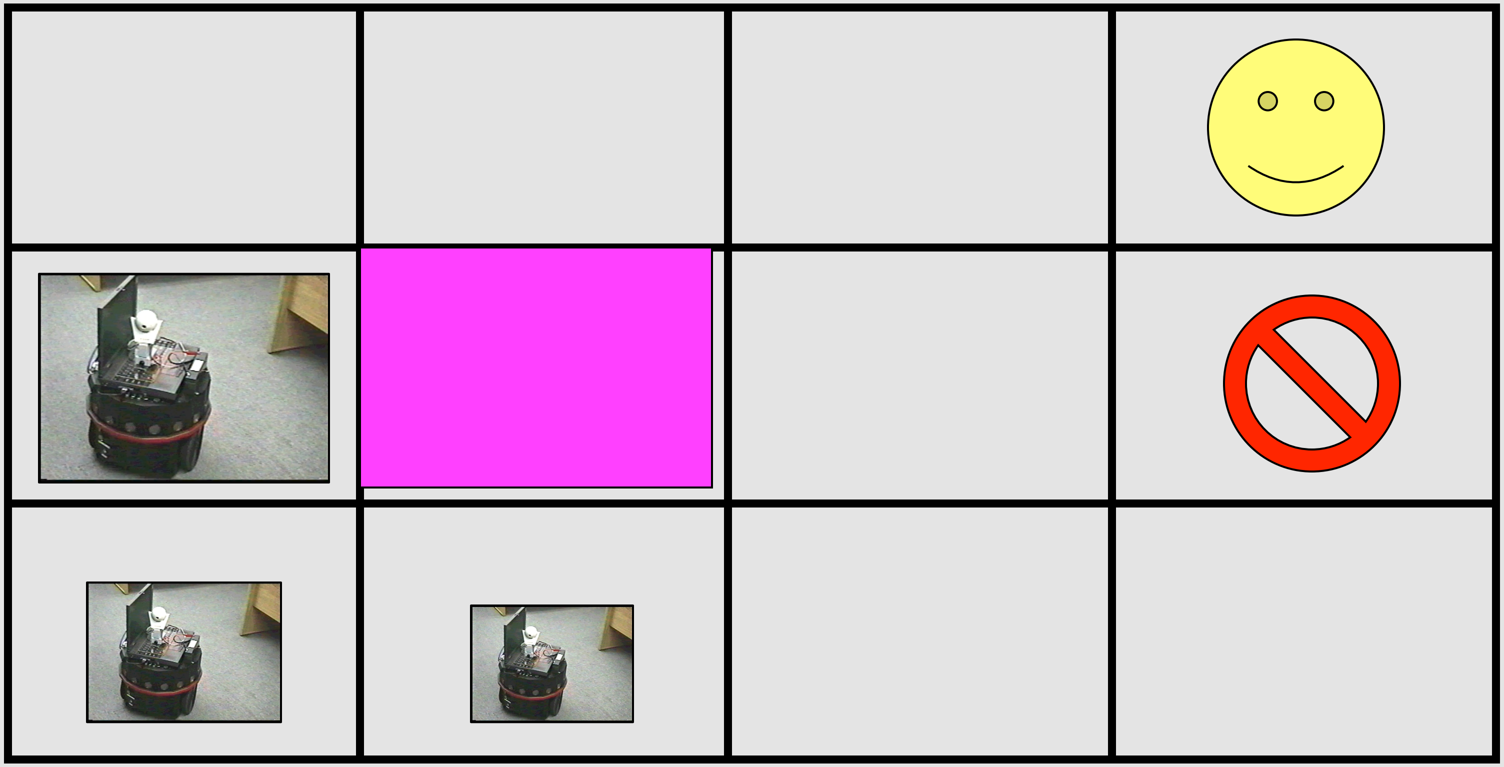
POMDP

- Besides the MDP model, a POMDP has:
  - An observation probability distribution, $P(O|S)$
  - An initial probability distribution, $P(S)$

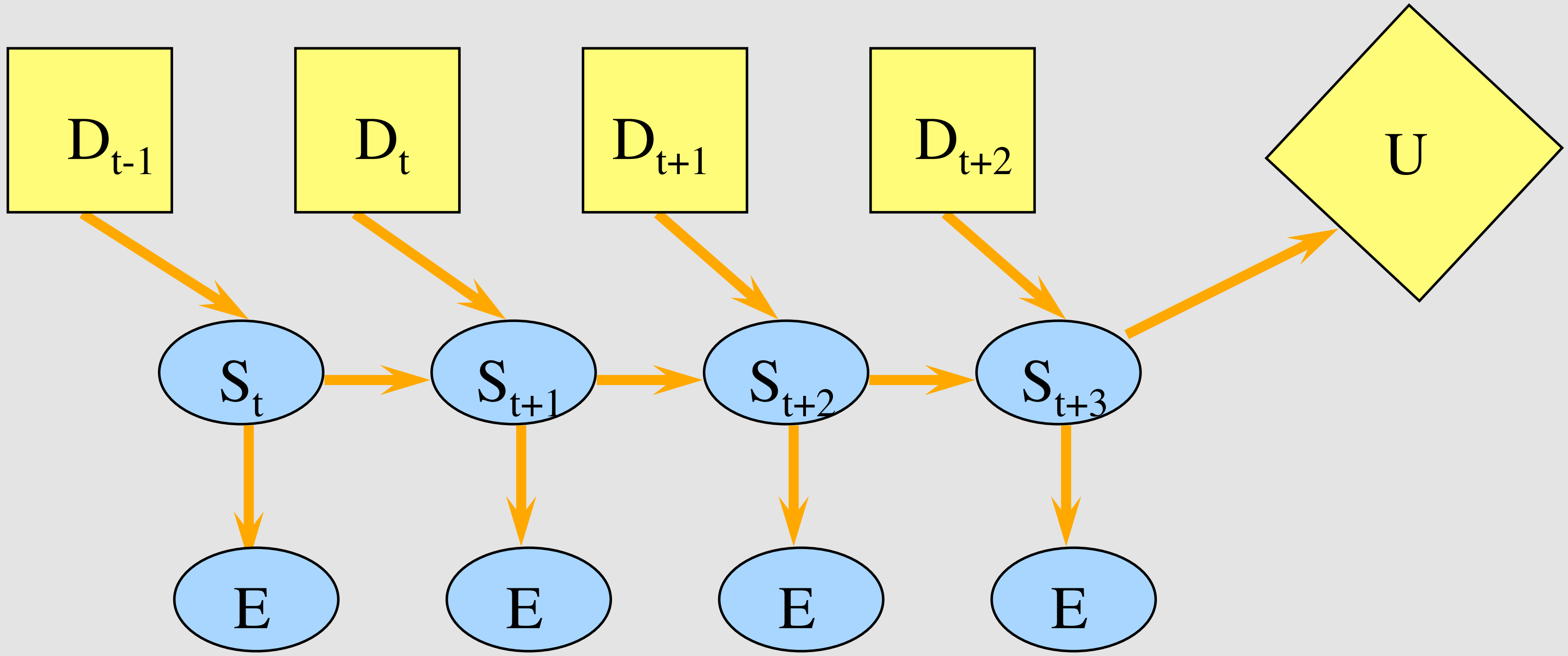
Uncertainty in the actions



Uncertainty in the state



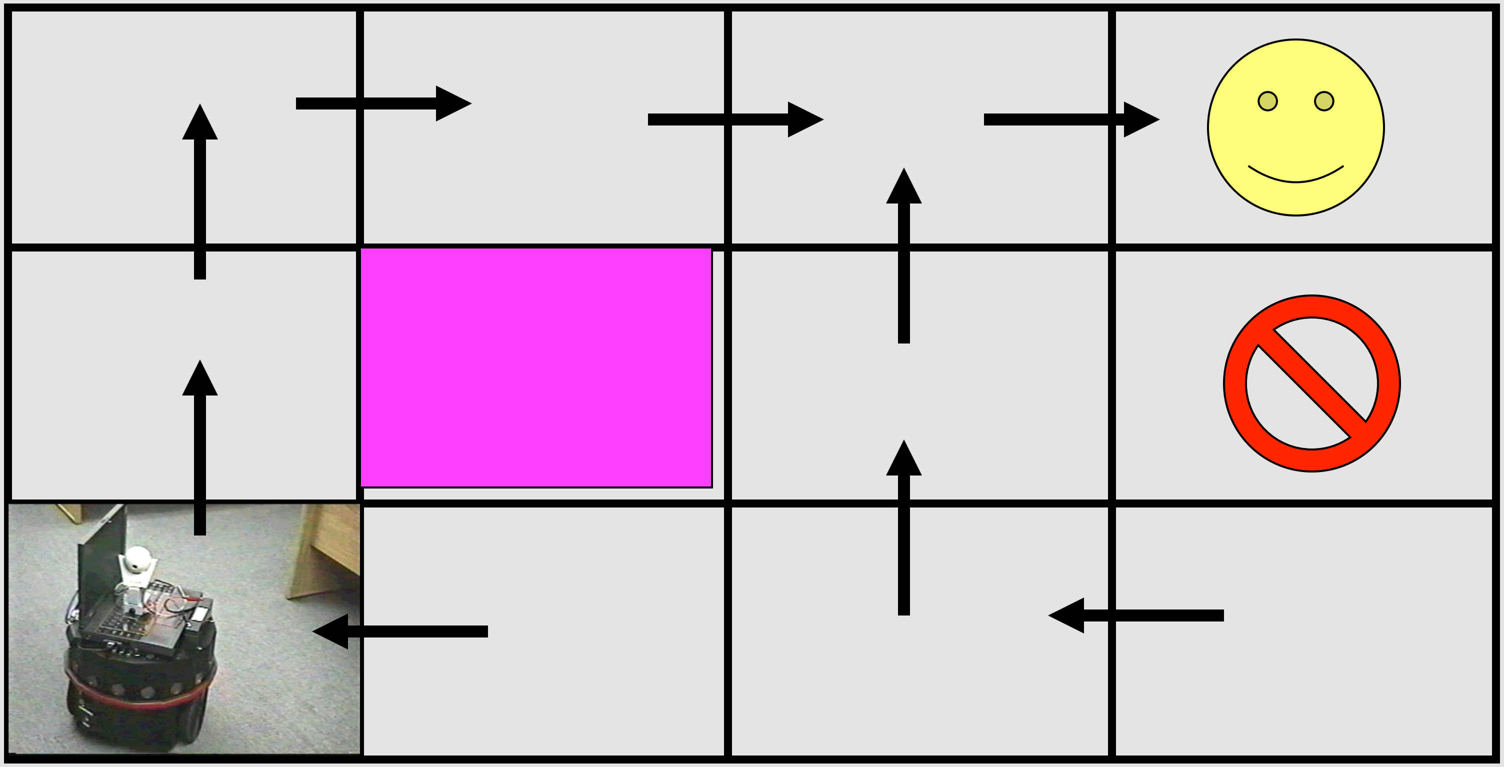
A POMDP as a Dynamic Decision Network



Basic solution techniques

- There are two main classes of algorithms:
  - Dynamic programming techniques: consider a known model (transition and reward functions) which is solved to obtain the optimal policy
  - Montecarlo and reinforcement learning: the model is not known, so the optimal policy is obtained by exploring the environment

Optimal *policy*



Initial position

Value function

- A policy for an MDP is an association $\pi: S \rightarrow A$ (action for each state).
- Given a policy, the value for finite horizon is $V_n^{\pi}: S \rightarrow \Re$

$$V^{\pi}(s) = R(s, a) + \sum P(s' \mid s,a)\, V(s')$$

- For infinite horizon, a *discount factor* is usually considered, $0<\gamma<1$:

$$V^{\pi}(s) = R(s, a) + \gamma \sum P(s' \mid s,a)\, V(s')$$

Optimal policy

- The solution for an MDP gives the optimal policy.
- That is, the policy that maximizes Bellman's equation :

$$V^*(s) = \max_a \{ R(s,a) + \gamma \sum_{s'} P(s' \mid s, a) V^*(s') \}$$

- Obtaining the optimal policy:

$$\pi^*(s) = \arg\max_a \{ R(s,a) + \gamma \sum_{s'} P(s' \mid s, a) V^*(s') \}$$

Value iteration

- For infinite horizon, we can obtain the utility and optimal policy with an iterative algorithm
- In each iteration (i+1), the utility of each state is estimated given the value in the previous stage (i):

$V_{i+1}(s) = R(s) + \max_a \Sigma_j P(s' \mid s,a) V_i(s')$

- When i → inf, the values converge and we obtain the optimal policy

Value iteration

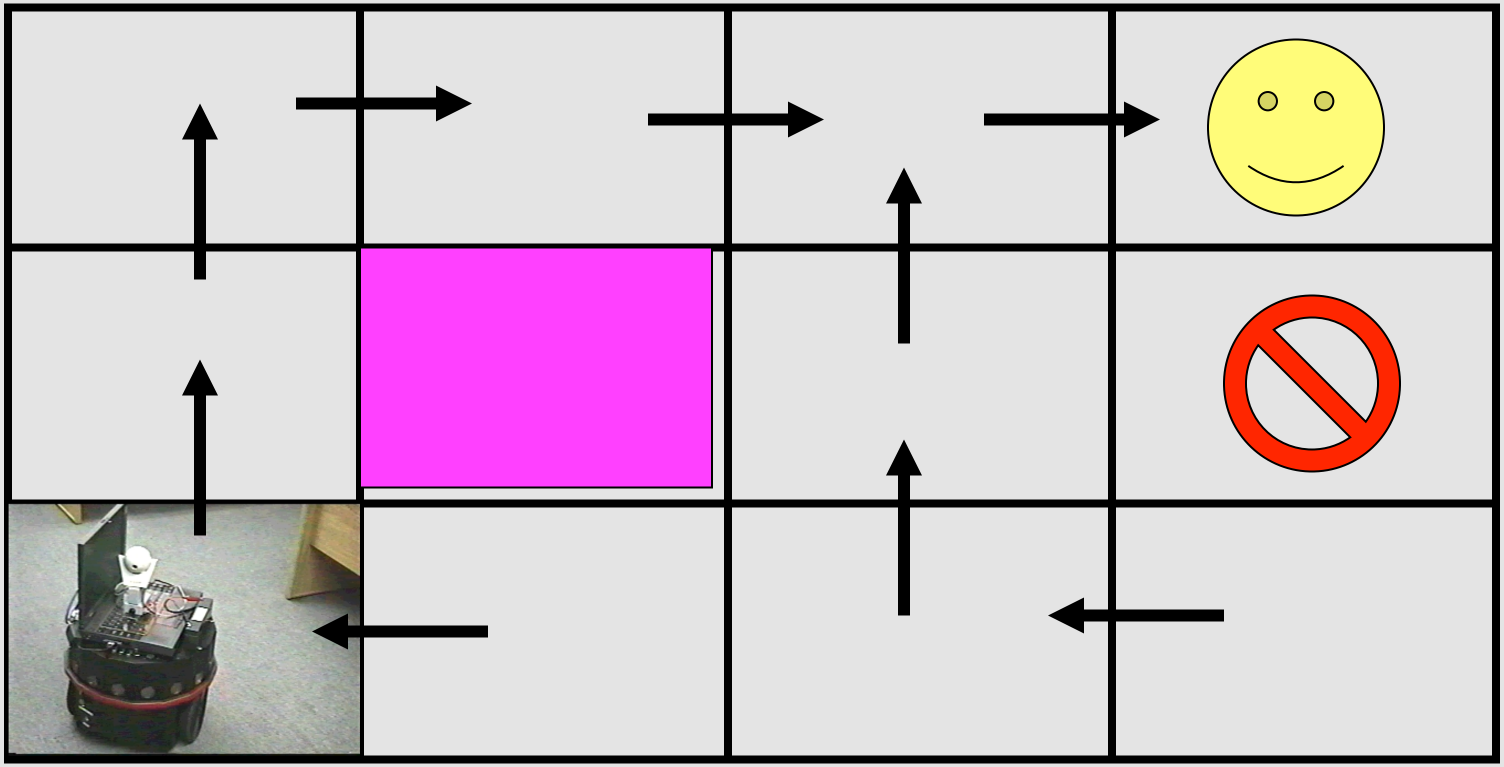
Algorithm:

- Initialization: $V_t = V_{t+1} = R$
- Repeat:
  - $V_t = V_{t+1}$
  - $V_{t+1}(s) = R(s) + \max_a \gamma \sum_j P(s' \mid s,a)\, V_t(s')$
- Until: $| V_t - V_{t+1} | < \varepsilon$

Example – utilities

0.812	0.868	0.912	
0.762		0.660	
0.705	0.655	0.611	0.338

Example – optimal policy



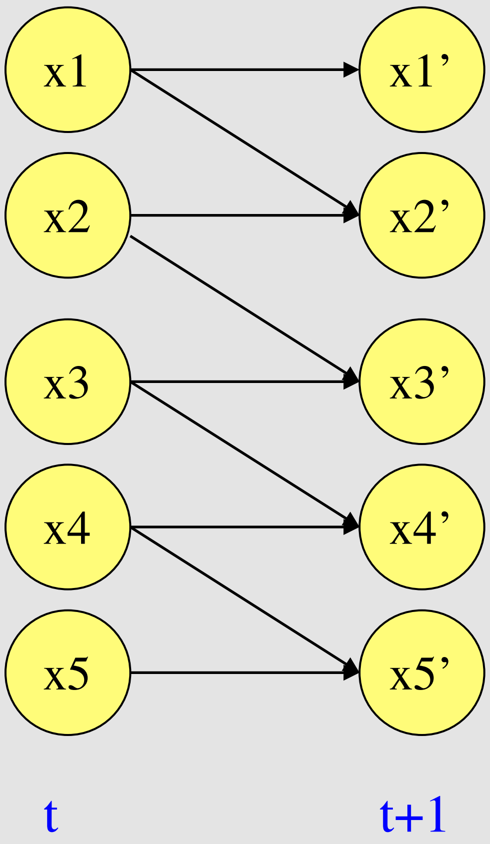
•Inicio

Factored MDPs

The state is decomposed in a set of factors or state variable:

$$X = \{x1, x2, x3, x4, x5\}$$

So the transition function is represented as a two-stage DBN per action

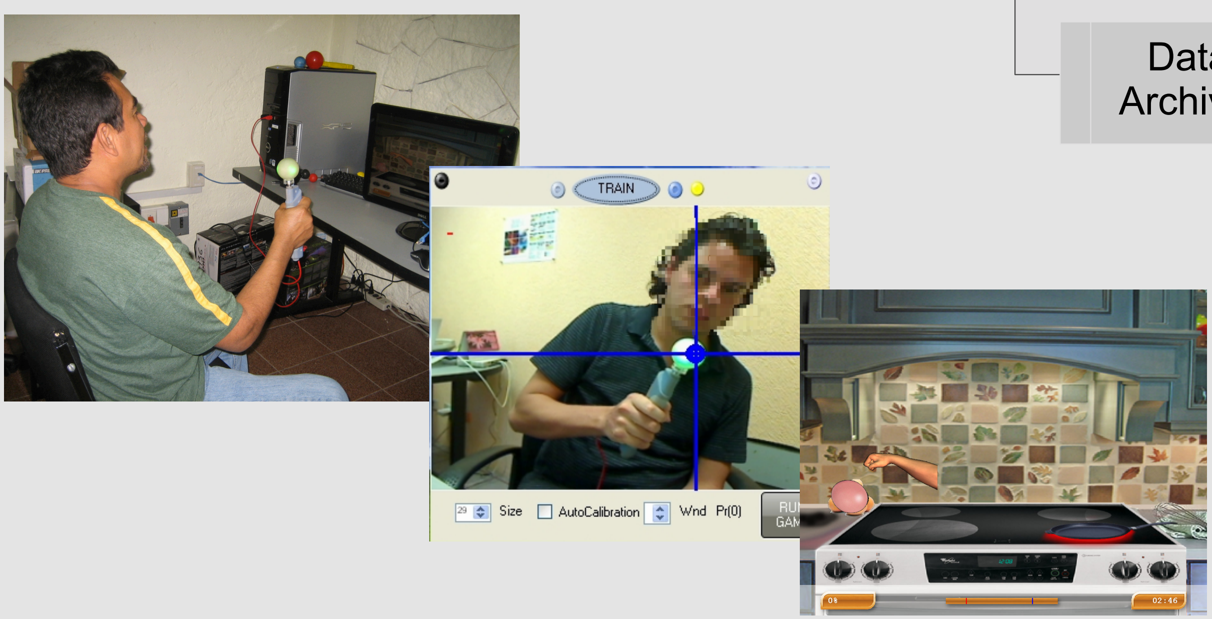


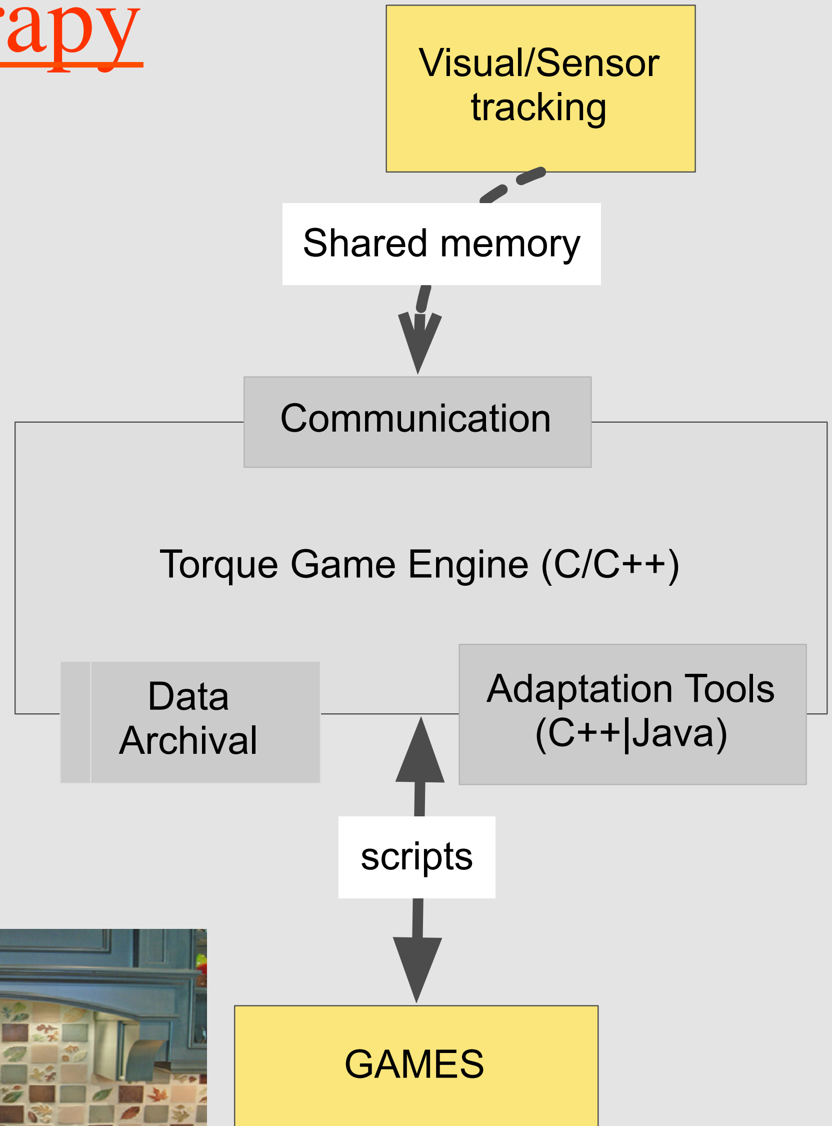
Gesture Therapy

- Many people suffer strokes (15 million worldwide per year)
- 80% lose arm and hand movement skills
- Physical and occupational therapy can help, but:
  - Expensive (requires a therapist)
  - Usually not enough
  - Patients loose motivation
- Robotic systems are too expensive for use at home or small clinics
- Develop low-cost technology that allows stroke patients to practice intensive movement training at home without the need of an always present therapist

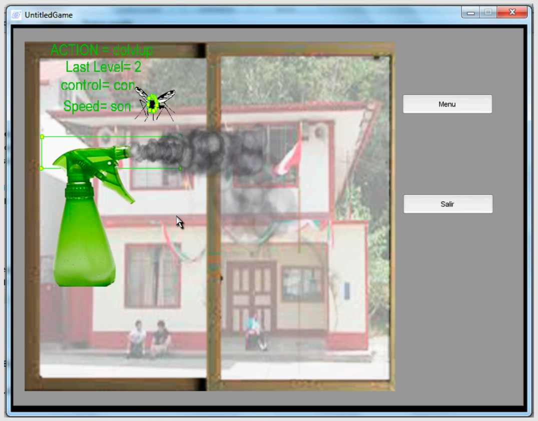
Gesture Therapy

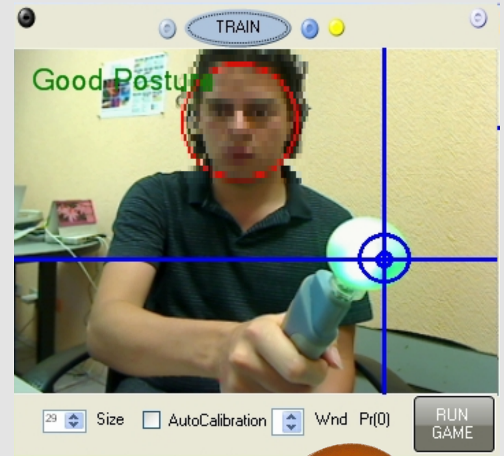
- Simulated environment
- Monocular tracker
- Gripper
- Trunk compensation detection
- Adaptation to the patient

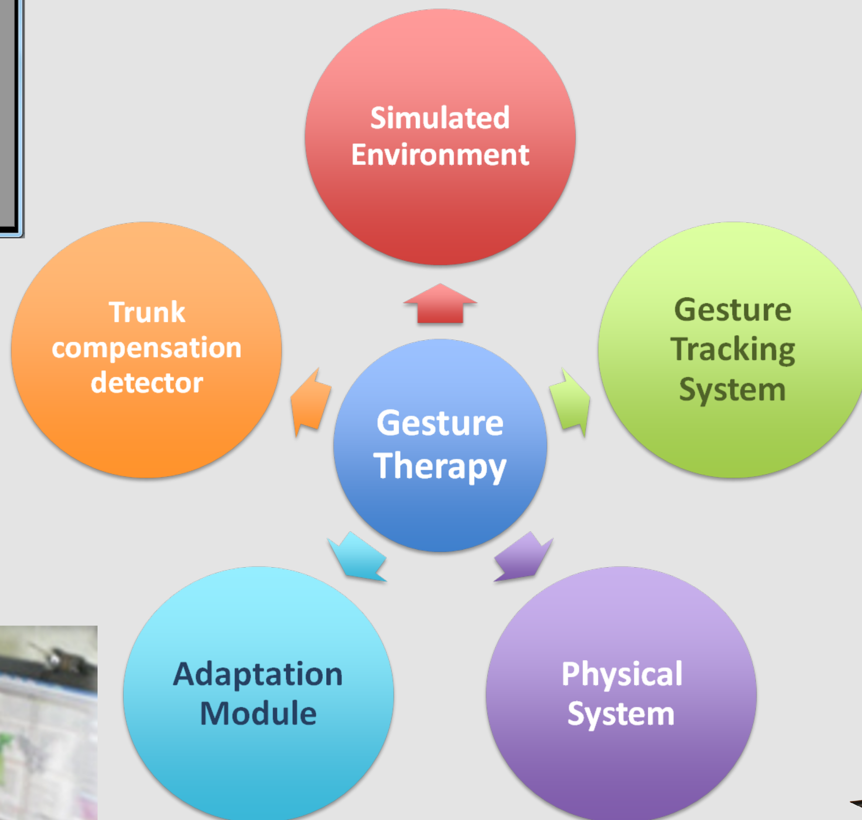




Gesture Therapy System







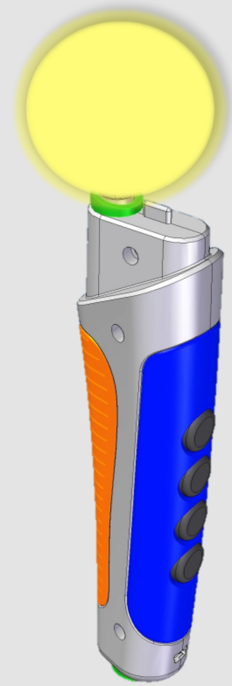
Visual Tracking System

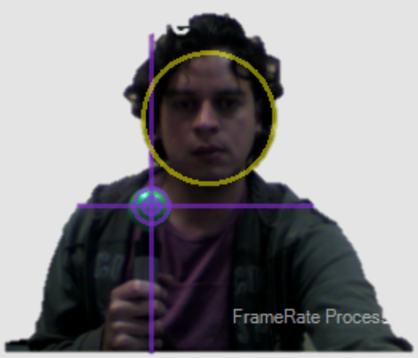
Orientación y textura

Plano H-S de HSV

Tracking (2Dbased on particle filters

Gripper

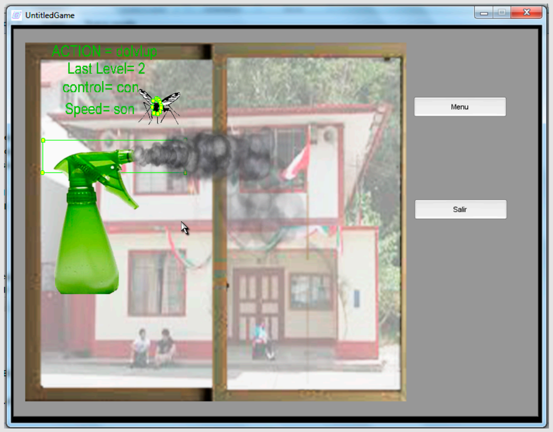




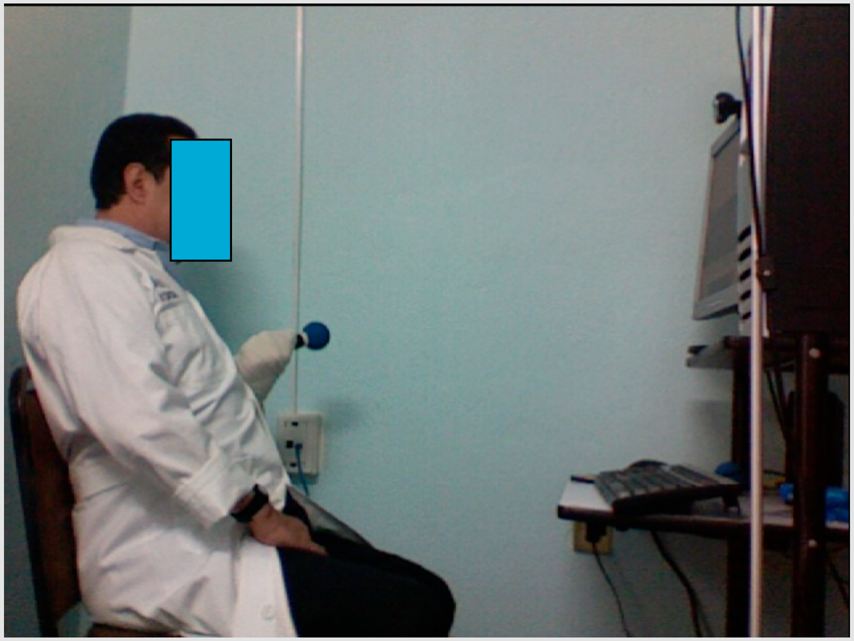
Compensation detection based on face recognition

Virtual Environment

- Serious games designed for stroke rehabilitation:
  - Simulate activities of daily leaving
  - Tailored for specific movements
  - Motivating

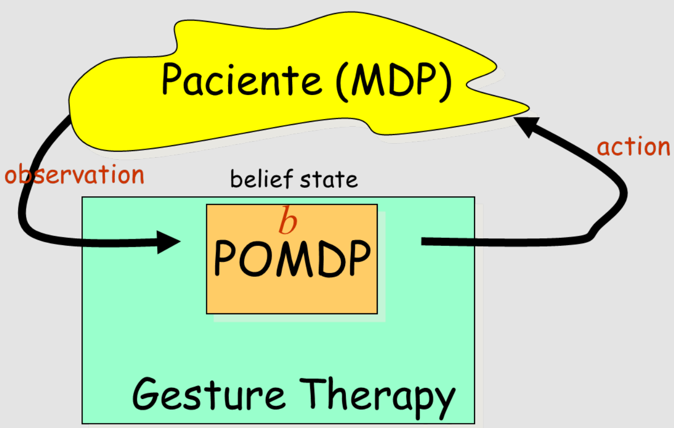


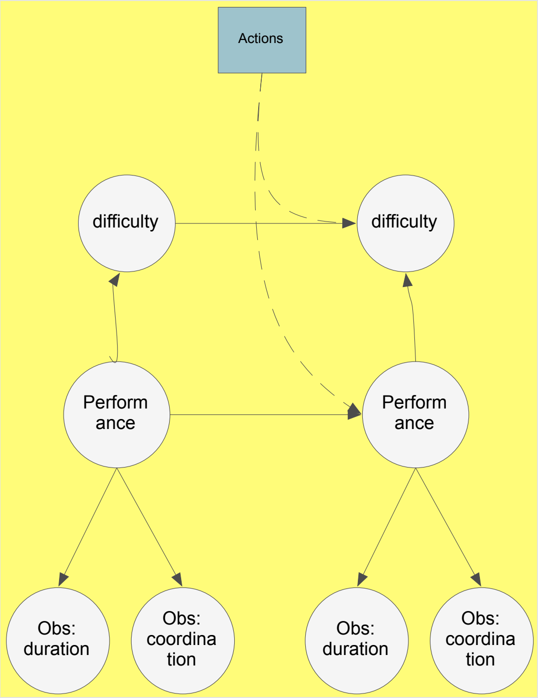
Prototype of the system at the INNN rehabilitation unit



Adptation to the patient

The system estimates the patient "state" based on observing its performance in the game (speed, control) and decides the game difficulty accordingly according to the policy dictated by an MDP



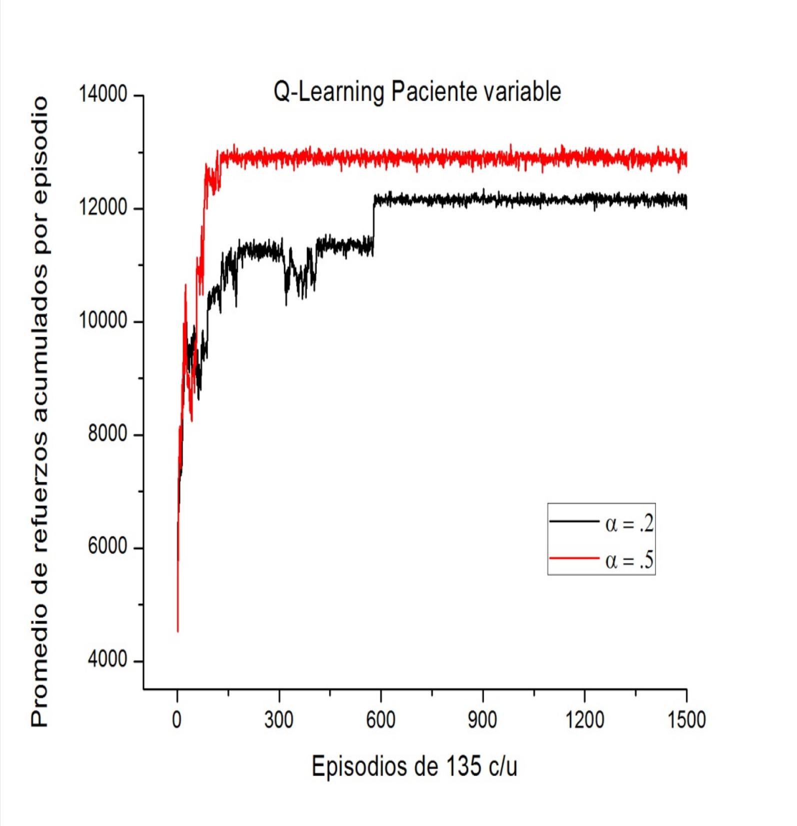


Policy adaptation

- The POMDP model could be *wrong* so the policy is not necessarily "optimal"
- Also, the best policy could depend on the patient
- We developed a policy adaptation algorithm based on RL+ reward shaping which improves an initial policy based on the therapist feedback

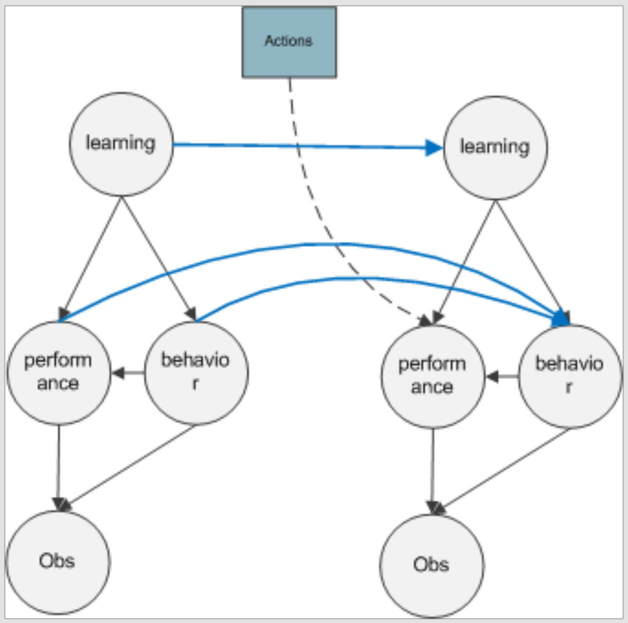
Initial results

- *Simulated therapist* – feedback based on the optimal policy

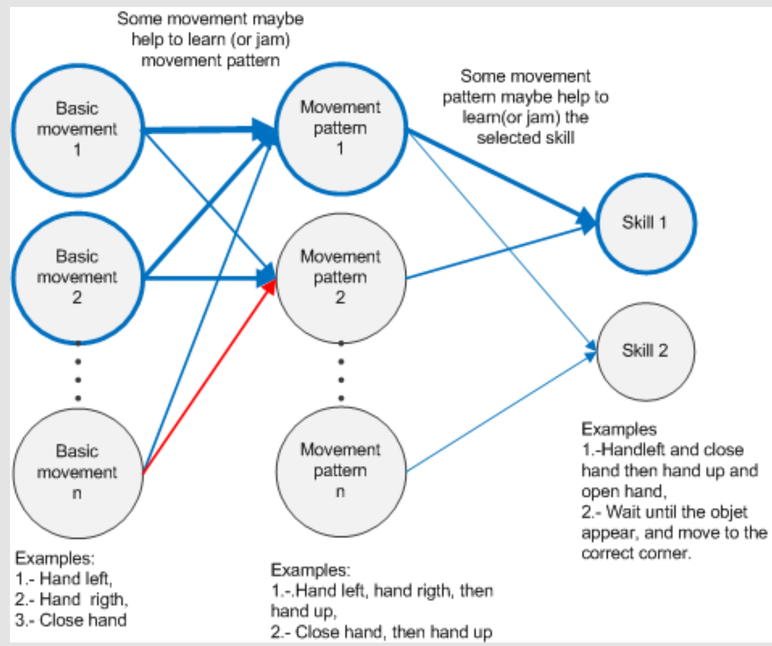


Adaptation at different levels

Withing game



Therapy planning



Conclusions

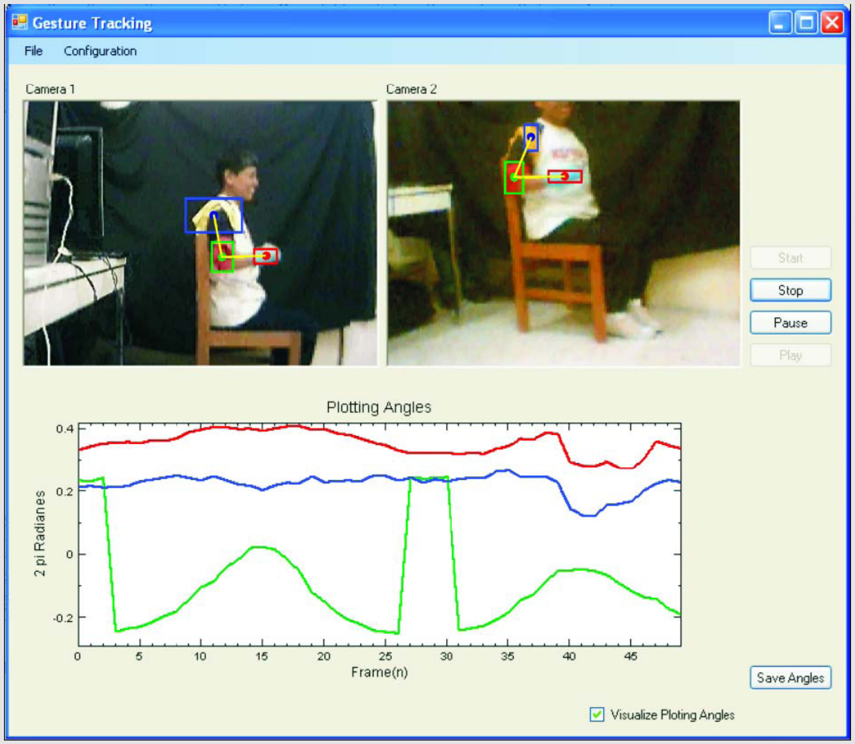
- The Bayesian approach combines prior knowledge (a priori probability) with evidence (likelihood) based on Bayes theorem
- Graphical models allow for an efficient and clear representation of probability distributions based on dependency & independency relations
- PGMs provide a set of techniques which can be applied to solve complex problems which require to model uncertainty, time and cost/utilities

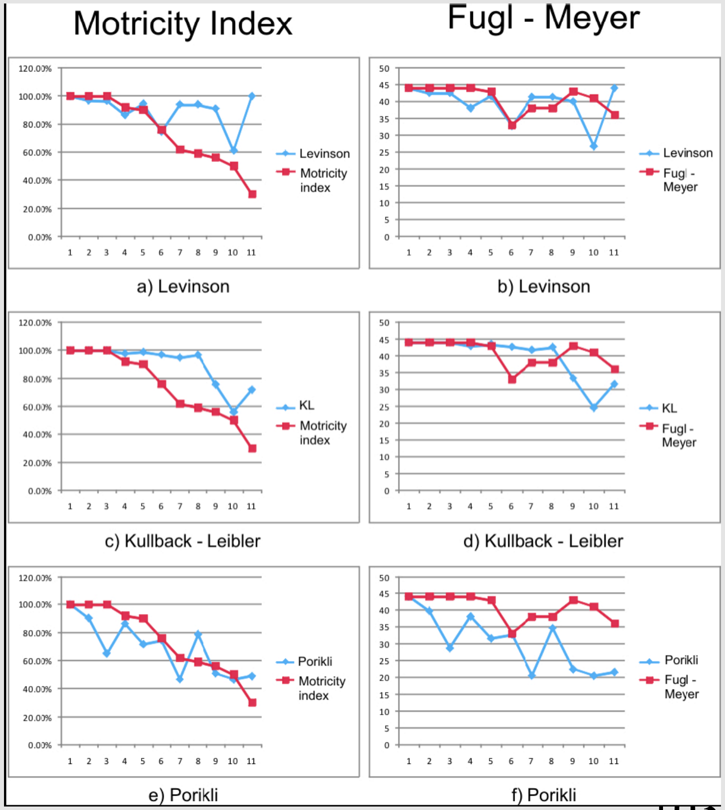
Current and Future Work

- HIV
  - Data analysis for Mexico and Central American
  - Consider genetic factors of the population
- Rehabilitation
  - Automatic evaluation – clinical scales
  - Analysis of affective/emotional aspects – consider in the adaptation process

Automatic Evaluation

Based on HMMs: measure of similarity against a “gold” standard using different metrics – comparision with clinical scales





Acknowedgements

- Collaborators:
  - Eduardo Morales, Felipe Orihuela, Hugo Escalante, Manuel Montes, Alma Rios (INAOE)
  - Santiago Ávila (INER)
  - Jorge Hernández (INNN)
  - Hesse Hoey (U. of Waterloo, Canada)
  - Ron Leder (UNAM)
  - David Reinkensmeyer (UC Irving, USA)
- Students:
  - Shender Ávila, Héctor Avilés, Pablo Hernández, Carlos Hernández
- Sponsors:
  - INAOE
  - CONACYT-Salud/SEP (Mexico)
  - FONCICYT (Mexico-EU)
  - RIC / Department of Education (USA)

Additional information …

- Gesture Therapy Blog: http://robotic.inaoep.mx/~foe/blog/
- L.E. Sucar, E. Morales, J. Hoey, *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions*, IGI-Global, 2012
- L. E. Sucar, *Probabilistic Graphical Models: Principles and Applications*, Springer-Verlag, 2014 (forthcoming)

- Course on Probabilstic Graphical Models:

http://ccc.inaoep.mx/~esucar/Clases-mgp/mgp.html

esucar@inaoep.mx