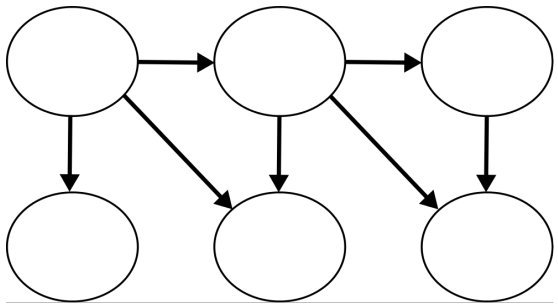
Probabilistic Graphical Models: Principles and Applications

Chapter 9: DYNAMIC AND TEMPORAL BAYESIAN NETWORKS

L. Enrique Sucar, INAOE



Outline

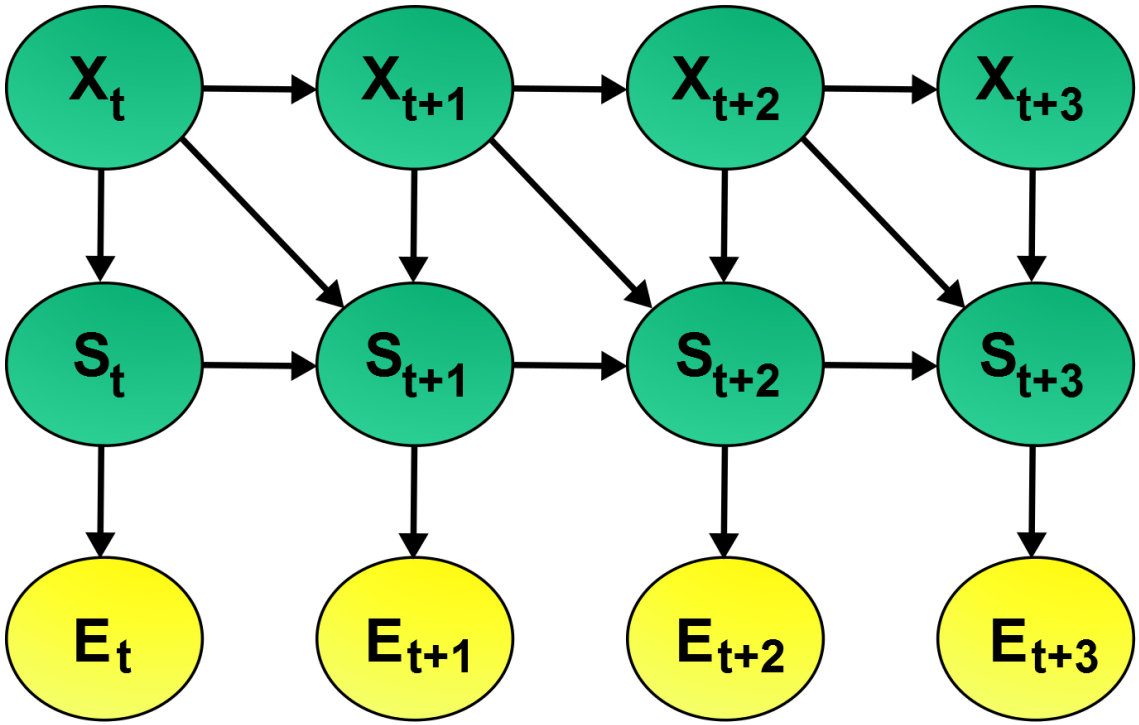
Introduction

- There are two basic types of Bayesian network models for dynamic processes: state based and event based
- State based models represent the state of each variable at discrete time intervals, so that the networks consist of a series of *time slices*, where each time slice indicates the value of each variable at time t – *Dynamic Bayesian Networks*
- Event based models represent the changes in state of each state variable; each temporal variable will then correspond to the time in which a state change occurs – *Event Networks or Temporal Networks*.

Dynamic Bayesian networks

- *Dynamic Bayesian networks* (DBNs) are an extension of Bayesian networks to model dynamic processes
- A DBN consists of a series of *time slices* that represent the state of all the variables at a certain time, t
- For each temporal slice, a dependency structure between the variables at that time is defined, called the *base network*
- Additionally, there are edges between variables from different slices, with their directions following the direction of time, defining the *transition network*

DBN - example



Assumptions

- First order Markov model. The state variables at time t depend only on the state variables at time $t-1$ (and other variables at time t).
- Stationary process. The structure and parameters of the model do not change over time.

Particular cases: MCs, HMMs, KFs

- DBNs can be seen as a generalization of Markov chains (MCs):

$$P(X_1, X_2, ..., X_T) = P(X_1)P(X_2 \mid X_1)...P(X_T \mid X_{T-1}) \quad (1)$$

- And hidden Markov models (HMMs):

$$P(\{S_{1:T}, Y_{1:T}\}) =$$

$$P(S_1)P(Y_1 \mid S_1) \prod_{t=2}^{T} P(S_t \mid S_{t-1})P(Y_t \mid S_t) \quad (2)$$

- Kalman Filters (KFs), also have one state and one observation variable, but both variables are continuous. The basic Kalman filter assumes Gaussian distributions and linear functions for the transitions and observations

Types of Inference

- *Filtering*. Predict the next state based on past observations: $P(X_{t+1} \mid Y_{1:t})$.
- *Prediction.* Predict future states based on past observations: $P(X_{t+n} \mid Y_{1:t})$.
- *Smoothing.* Estimate the current state based on past and future observations (useful for learning): $P(X_t \mid Y_{1:T})$.
- *Decoding.* Find the most likely sequence of hidden variables given the observations: $\text{ArgMax}(X_{1:T})\ P(X_{1:T} \mid Y_{1:T})$.

Inference

- Efficient inference methods have been developed for particular types of models, such as HMMs
- For more complex models, inference becomes computationally intractable
- In these cases we can apply approximate methods based on sampling – Markov chain Monte Carlo, such as Particle Filters

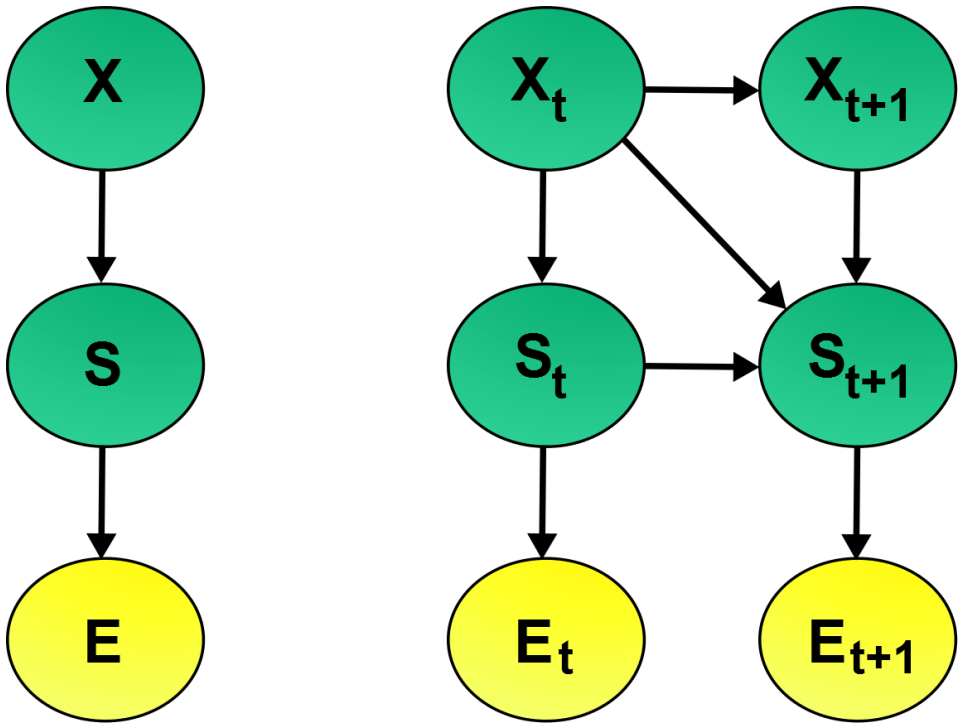
Learning DBNs

- There are 4 basic cases for learning DBNs
- For all the cases we can apply extensions of the methods for parameter and structure learning for BNs

Structure	Observability	Method
Known	Full	Maximum likelihood estimation
Known	Partial	Expectation–maximization (EM)
Unknown	Full	Search (global) or tests (local)
Unknown	Partial	EM and global or local

Unknown Structure and Full Observability

- Assuming that the DBN is stationary (time invariant), we can consider that the model is defined by two structures: (i) the base structure, and (ii) the transition structure
- Divide the learning of a DBN into two parts, first learn the base structure, and then, given the base structure, learn the transition structure



Structure Learning: base and transition networks

- For learning the base structure we can use all the available data for each variable, ignoring the temporal information. This is equivalent to learning a BN
- For learning the transition network we consider the temporal information, in particular the data for all variables in two consecutive time slices, $\mathbf{X_t}$ and $\mathbf{X_{t+1}}$
- Considering the base structure, we can then learn the dependencies between the variables at time t and $t+1$

Temporal Event Networks

- Temporal event networks (TENs) are an alternative to DBNs for modeling dynamic processes
- In a temporal event network, a node represents the *time* of occurrence of an event or state change of certain variable
- For some problems, in which there are few state changes in the temporal range of interest, event networks provide a simpler and more efficient representation
- We will focus on one variant known as Temporal Nodes Bayesian Networks

TNBNs: Definition

- A Temporal Nodes Bayesian Network (TNBN) is composed of a set of Temporal Nodes (TNs)
- TNs are connected by edges, where each edge represents a causal-temporal relationship
- There is at most one state change for each variable (TN) in the temporal range of interest
- The value taken by the variable represents the interval in which the event occurs – time is discretized in a finite number of intervals, allowing a different number and duration of intervals for each node
- Each interval defined for a child node represents the possible delays between the occurrence of one of its parent events (cause) and the corresponding child event

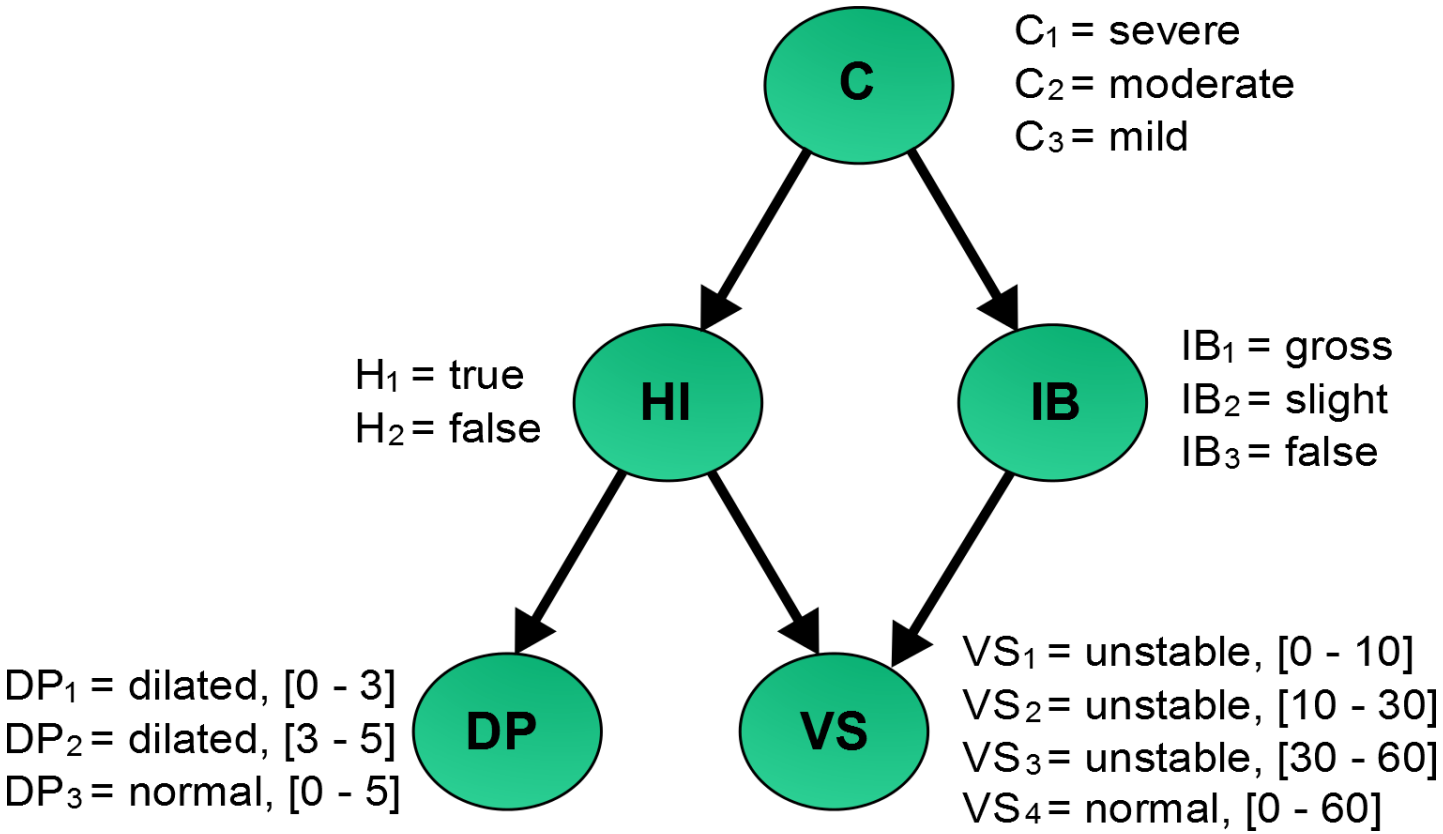
Car accident example

- Assume that at time $t = 0$, an automobile accident occurs, that is, a Collision. This kind of accident can be classified as *severe*, *moderate*, or *mild*. To simplify the model, we will consider only two immediate consequences for the person involved in the collision: Head Injury and Internal Bleeding. A Head Injury can bruise the brain, and chest injuries can lead to Internal Bleeding. These are all instantaneous events that may generate subsequent changes, for example the Head Injury event might generate Dilated Pupils and unstable Vital Signs

Car accident example

- Additionally, a physician domain expert provided some important temporal information: If a head injury occurs, the brain will start to swell and if left unchecked the swelling will cause the pupils to dilate within 0 to 60 minutes. If internal bleeding begins, the blood volume will start to fall, which will tend to destabilize vital signs. The time required to destabilize vital signs will depend on the severity of the bleeding: if the bleeding is gross, it will take from 0 to 15 minutes; if the bleeding is slight it will take from 15 to 45 minutes. A head injury also tends to destabilize vital signs, taking from 0 to 15 minutes to make them unstable

TNBN: car accident example



TNBN: inference

- A TNBN allows for reasoning about the probability of occurrence of certain events, for diagnosis (i.e., finding the most probable cause of a temporal event) or prediction (i.e., determining the probable future events that will occur given a certain event)
- Standard probability propagation techniques for standard BNs can be applied
- However, given that a TNBN represents relative times between events, the cases of prediction and diagnosis have to be differentiated

Prediction

- In the case where at least one of the root (instantaneous) nodes of the TNBN is part of the evidence, then the time reference for the model is fixed and probability propagation can be performed directly, obtaining the posterior probability of the subsequent events
- For temporal nodes, the inference procedure will determine the probability of occurrence for each temporal interval, or of not occurrence

Diagnosis

- In the case where none of the instantaneous nodes are known, and the evidence is given only for temporal nodes, then several *scenarios* need to be considered
- In this case, all the n possible intervals for the TN have to be considered, performing inference n times, one for each interval
- The results for each scenario have to be maintained, until there is additional evidence, such as the occurrence of another event, that allows for discarding some scenarios

TNBNs: learning

- Learning a TNBN involves 3 aspects: (i) learning the temporal intervals for the temporal nodes, (ii) learning the structure of the model, (iii) learning the parameters of the model
- As these three components are interrelated, an iterative procedure is required, that learns an initial estimate of one (or more) of these aspects, and then improves these initial estimates iteratively
- Next we present the LIPS algorithm

Algorithm

- First, it performs an initial discretization of the temporal variables, for example using an Equal-Width discretization – obtains an initial approximation of the intervals for all the Temporal Nodes
- Then it performs standard BN structural and parameter learning
- The interval learning algorithm refines the intervals for each temporal node (TN) by means of clustering. For this, it uses the information of the configurations of the parent nodes
- It obtains different sets of intervals that are merged and combined, this process generates different interval sets that will be evaluated in terms of the predictive accuracy
- Finally, the parameters (CPTs) are updated according to the new set of intervals for each TN

Data to Learn the Accident TNBN

- Top: original data showing the time of occurrence of the temporal events.
- Bottom: temporal data after the the initial discretization

Collision	Head Injury	Internal Bleeding	Dilated Pupils	Vital S.
severe	true	gross	14	20
moderate	true	gross	25	25
mild	false	false	-	-
...	...	...	...	...

Collision	Head Injury	Internal Bleeding	Dilated Pupils	Vital S.
severe	true	gross	[10-20]	[15-30]
moderate	true	gross	[20-30]	[15-30]
mild	false	false	-	-
...	...	...	...	...

Learning example – accident TNBN

- After the initial discretization, the initial structure and parameters are learned
- Then the intervals for each temporal node are refined
- Finally the parameters are re-estimated
- Intervals obtained for the node *Dilated Pupils* – the temporal data is divided into two partitions, one for each configuration of the parent node

Partition	**Intervals**
Head Injury=true	$[11-35]$ $[11-27][32-53]$ $[8-21][25-32][45-59]$
Head Injury=false	$[3-48]$ $[0-19][39-62]$ $[0-14][28-40][47-65]$

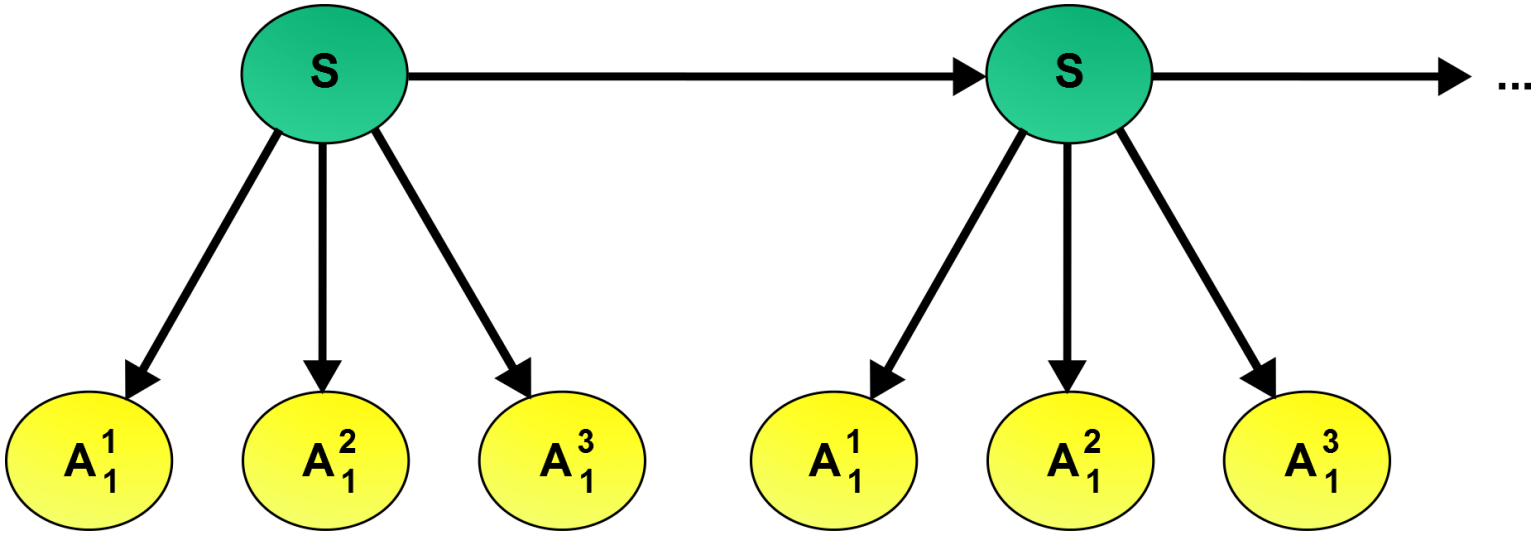
Applications

- Dynamic Bayesian networks are used for dynamic gesture recognition
- Temporal event networks are used for predicting HIV mutational pathways

Gesture Recognition

- Dynamic Bayesian networks provide an alternative to HMMs for dynamic gesture recognition
- Greater flexibility in terms of the structure of the models
- We consider a particular type of DBN known as *dynamic Bayesian network classifier* (DBNC)
- A DBNC has a hidden state variable for each time instant, S_t; however, the observation variable is decomposed into m attributes, $A_t^1, \ldots A_t^m$, which are assumed to be conditionally independent given S_t

DBNC



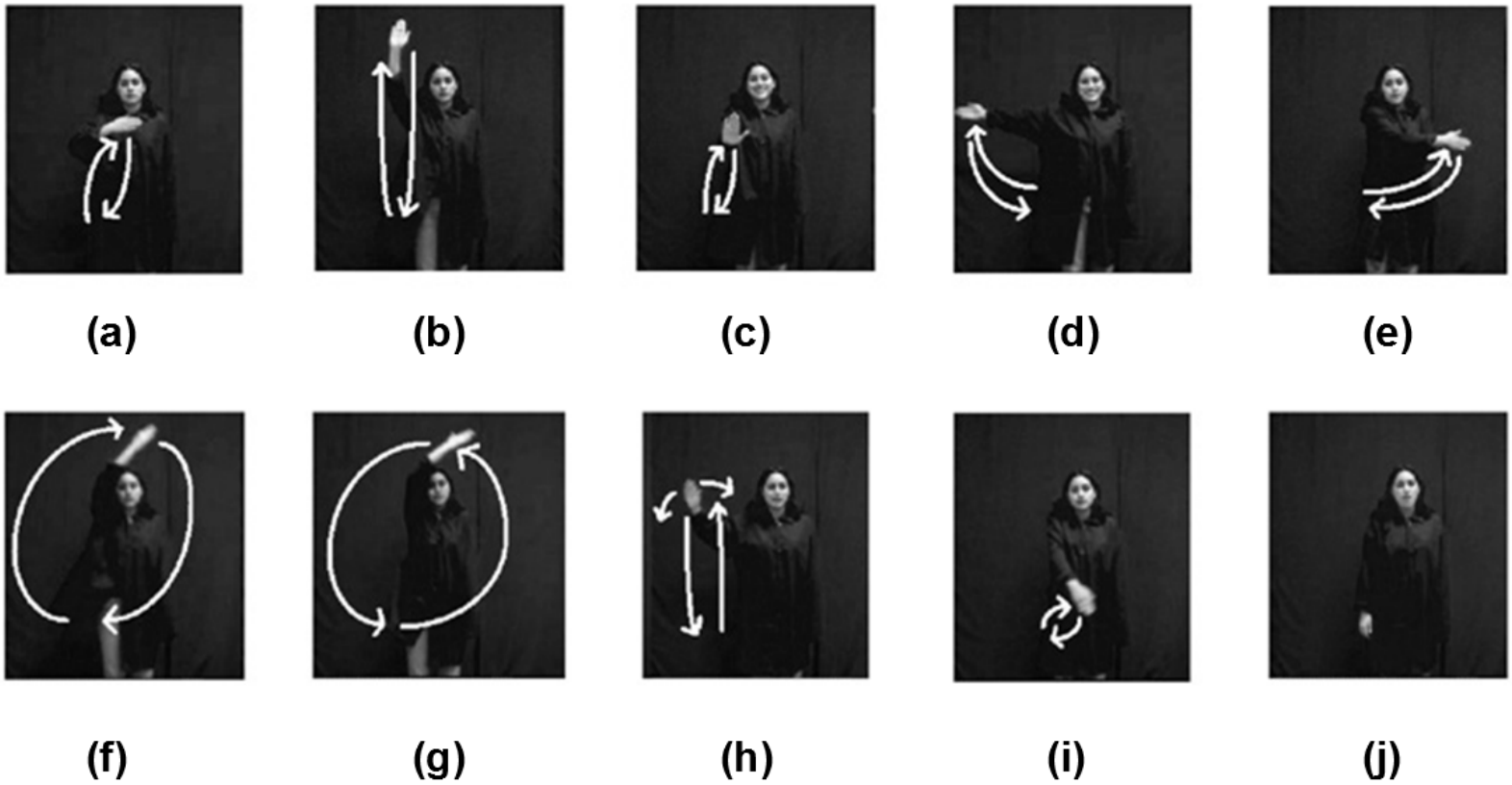
- The joint probability of a DBNC can be factored as:

$$P(\{S_{1:T}, \mathbf{A}_{1:T}\}) =$$

$$P(S_1)[\prod_{m=1}^{M} P(A_1^m \mid S_1)] \prod_{t=2}^{T} P(S_t \mid S_{t-1})[\prod_{m=1}^{M} P(A_t^m \mid S_t)] \quad (3)$$

Gesture recognition with DBNCs

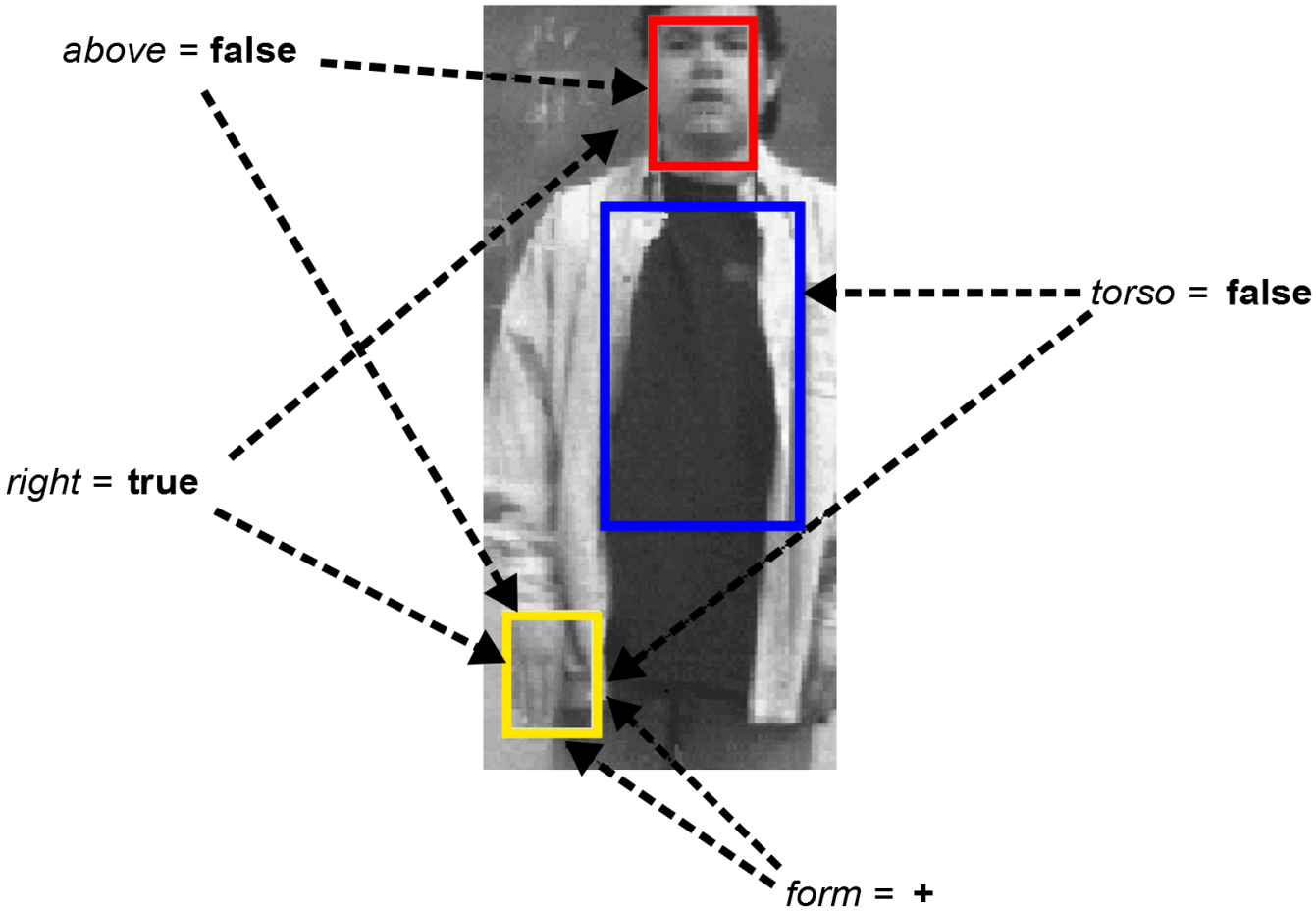
- 9 hand gestures oriented to command a mobile robot



Features

- The features include motion and posture information, including in total 7 attributes
- Motion features are Δarea - or changes in the hand area-, Δx and Δy -or changes in hand position on the XY-plane of the image
- Posture features named *form, right, above, and torso* describe hand orientation and spatial relations between the hand and other body parts, such as the face and torso

Posture features



Experiments

- Three experiments were done to compare classification and learning performances of DBNCs and HMMs
- An important difference between the DBNCs and the HMMs is the number of parameters that are required for each model
- The number of parameters to specify state observation distributions of HMMs with posture-motion features is 648 and with only motion data is 27
- With DBNCs, parameters are 21 in the former case, and 12 in the latter case. This significant reduction in the number of parameters for the DBNC has an important impact in the training

Results

- In the case of the experiments with the same person using motion and posture attributes, both HMMs and DBNCs obtain a very good performance, with recognition rates between 96 and 99%
- DBNCs obtain slightly better recognition results, but with a significant reduction in training time, about ten times faster than HMMs
- For the experiments with multiple people, as expected the performance of both, HMMs and DBNCs decreases, to about 86% recognition with motion-posture attributes. If only motion attributes are used, the performance is in the order of 65%
- In the case of the third experiments, variations in distance and orientation also have an impact in the recognition rates, with the second aspect having a greater effect

Predicting HIV Mutational Pathways

- The human immunodeficiency virus (HIV) is one of the fastest evolving organisms on the planet
- In the case of antiretroviral therapy (ART), a strong selective pressure acting on HIV that, under suboptimal conditions, readily selects for mutations that allow the virus to replicate even in the presence of highly potent antiretroviral drug combinations
- We address the problem of finding mutation-mutation and drug-mutation associations in individuals receiving antiretroviral therapy – focused on protease inhibitors (PIs)
- Historical data from HIV patients was used to learn a TNBN, and then this model was evaluated in terms of the discovered relationships

Data

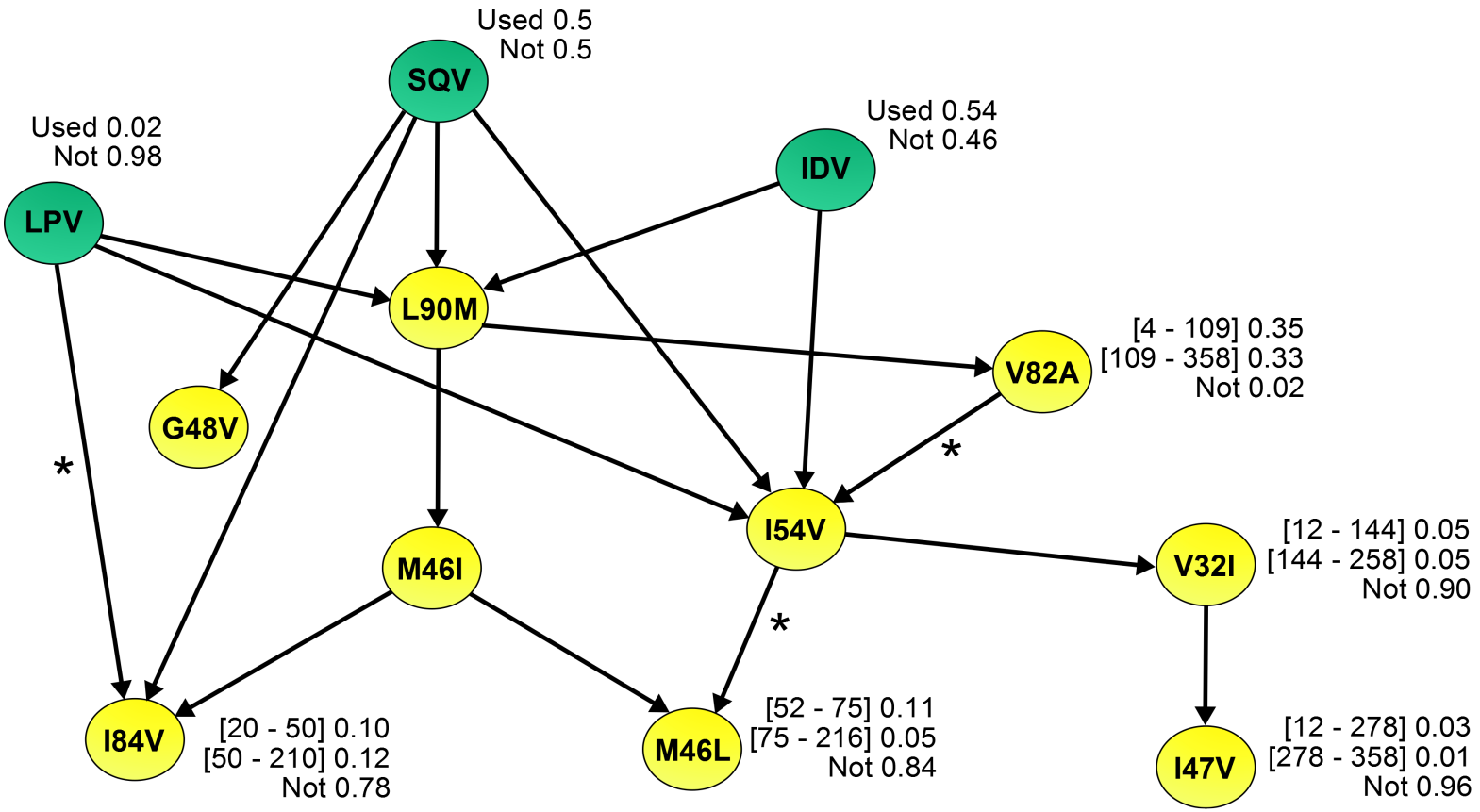
- Clinical data from 2373 patients with HIV subtype B was retrieved from the HIV Stanford Database
- For each patient, data consisted of an initial treatment (a combination of drugs) administered to the patient and a list of the most frequent mutations in the viral population at a specific time after the initiation of treatment

Patient	Initial Treatment	List of Mutations	Time (Week
Pat_1	LPV, FPV, RTV	L63P, L10I	15
		V77I	25
		I62V	50
Pat_2	NFV, RTV, SQV	L10I	25
		V77I	45

Learning the model

- A TNBN was learned from a reduced HIVDB data set that focused on certain drugs and part of the virus
- Two modifications to the original algorithm were made, one to measure the strength of the temporal-probabilistic relations, and another to vary the variable order given to the structure learning algorithm (K2), so the results are not biased by a particular predefined order
- In order to evaluate the models and to measure the statistical significance of edge strengths, non-parametric bootstrapping was used (obtaining several models)
- A strong relation was defined as one that appeared in at least 90% of the graphs, and a suggestive relation was defined as one that occurred with values between 70% and 90%

Learned TNBN



* strong relation.

Analysis

- The model was able to predict clinically relevant associations between the chosen drugs and mutations
- A strong association between SQV, G48V, and I84V was predicted in the model
- Two possible mutational pathways for LPV resistance were predicted:
  - I54V $\rightarrow$ V32I $\rightarrow$ I47V
  - L90M $\rightarrow$ M46IL $\rightarrow$ I84V
- Also, the shared mutational pathway between IDV and LPV was observed, involving mutations L90M, M46IL, I54V, V82A and I84V

Book

Sucar, L. E, *Probabilistic Graphical Models*, Springer 2015 – Chapter 9

Additional Reading - DBNs

 N. Friedman, N., Murphy, K., Russell, S.: Learning the Structure of Dynamic Probabilistic Networks. In: Proceedings of the Fourteenth Conference on Uncertainty in Artificial (UAI), Morgan Kaufmann Publishers Inc., pp. 139–147 (1998)

 Ghahramani, Z.: Learning Dynamic Bayesian Networks. Lecture Notes in Computer Science. 1387, 168–197 (1998)

 Murphy, K.: Dynamic Bayesian Networks: Representation, Inference and Learning. Dissertation, University of California, Berkeley (2002)

Additional Reading - Event Networks


Arroyo-Figueroa, G., Sucar, L.E.: A Temporal Bayesian Network for Diagnosis and Prediction. In: Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI), Morgan-Kaufmann, San Mateo, pp. 13–20 (1999)


Galán, S.F., Arroyo-Figueroa, G., Díez, F. J. Sucar, L.E.: Comparison of Two Types of Event Bayesian Networks: A Case Study. Applied Artificial Intelligence. 21(3), 185–209 (2007)


Hernández-Leal, P., González, J.A., Morales, E.F., Sucar, L.E.: Learning Temporal Nodes Bayesian Networks. International Journal of Approximate Reasoning. 54(8), 956–977 (2013)

Additional Reading - Applications


Avilés-Arriaga, H.H., Sucar, L.E., Mendoza-Durán, C.E., Pineda-Cortés, L.A.: Comparison of Dynamic Naive Bayesian Classifiers and Hidden Markov Models for Gesture Recognition. Journal of Applied Research and Technology. 9(1), 81–102 (2011)


Hernández-Leal, P., Rios-Flores, A., Ávila-Rios, S., Reyes-Terán, G., González, J.A., Fiedler-Cameras, L., Orihuela-Espina, F., Morales, E.F., Sucar, L.E.: Discovering HIV Mutational Pathways using Temporal Bayesian Networks. Artificial Intelligence in Medicine. 57(3), 185–195 (2013)