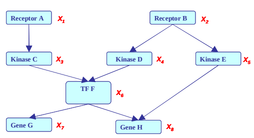
School of Computer Science
Carnegie Mellon

Some more Complex Graphical Models

Probabilistic Graphical Models (10-708)

Recitation 8, Nov 11, 2007

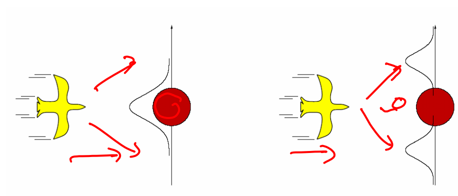


Hetunandan

Reading: J-Chap. 5,6, KF-Chap. 8

The need for complex dynamic models

- Complex dynamic systems:
  - Non-linearity
  - Non-Gaussianity
  - Multi-modality
  - ...



(a) (b)

- Limitation of LDS
  - defines only linearity evolving, unimodal, and Gaussian belief states
    - A Kalman filter will predict the location of the bird using a single Gaussian centered on the obstacle.
    - A more realistic model allows for the bird's evasive action, predicting that it will fly to one side or the other.

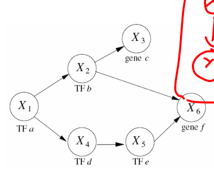
Representing complex dynamic processes

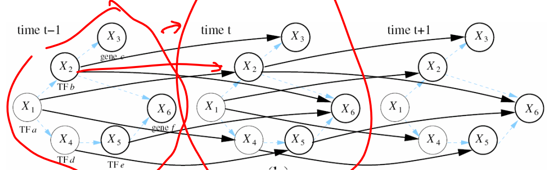
- The problem with HMMs
  - Suppose we want to track the state (e.g., the position) of D objects in an image sequence.
  - Let each object be in K possible states.
  - Then $X_t = (X_t^{(1)}, \ldots, X_t^{(D)})$ can have K^D possible values.

$\Rightarrow$ Inference takes time and space.

$\Rightarrow$ $P(X_t|X_{t-1})$ need parameters to specify.

Dynamic Bayesian Network





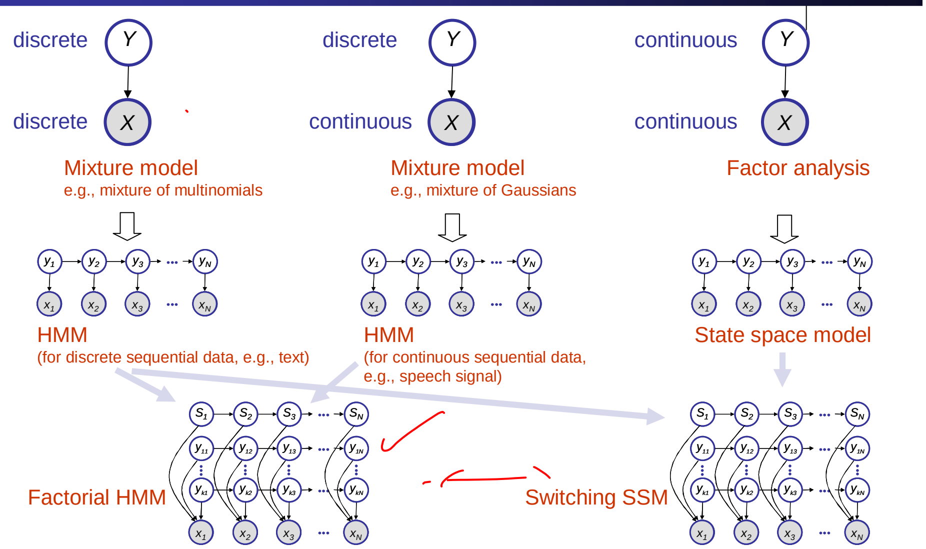
- A DBN represents the state of the world at time t using a set of random variables, $X_t^{(1)}, \ldots, X_t^{(D)}$ (factored/ distributed representation).
- A DBN represents $P(X_t|X_{t-1})$ in a compact way using a parameterized graph.

$\Rightarrow$ A DBN may have exponentially fewer parameters than its corresponding HMM.
$\Rightarrow$ Inference in a DBN may be exponentially faster than in the corresponding HMM.

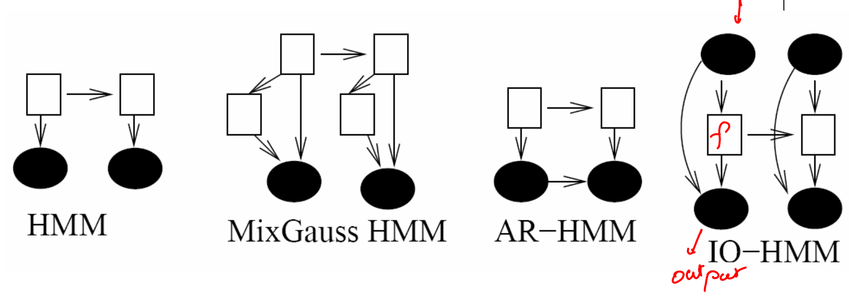
DBNs are a kind of graphical model

- In a graphical model, nodes represent random variables, and (lack of) arcs represents conditional independencies.
- DBNs are Bayes nets for dynamic processes.
- Informally, an arc from X_t^i to X_{t+1}^j means X_i "causes" X_j.
- Can "resolve" cycles in a "static" BN

A road map to complex dynamic models



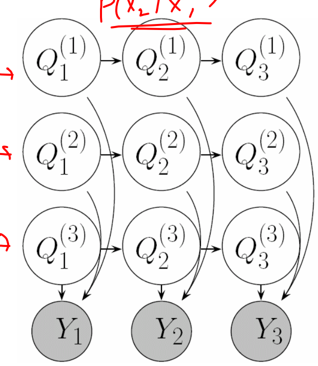
HMM variants represented as DBNs



- The same code (standard forward-backward, viterbi, and Baum-Welsh) can do inference and learning in all of these models.

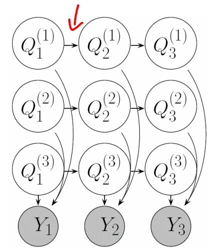
Factorial HMM

- The belief state at each time is $X_t = \{Q_t^{(1)}, \ldots, Q_t^{(k)}\}$ and in the most general case has a state space $O(d^k)$ for k d-nary chains
- The common observed child Y_t couples all the parents (explaining away).
- But the parameterization cost for fHMM is $O(kd^2)$ for k chain-specific transition models $p(Q_t^{(i)} | Q_{t-1}^{(i)})$ rather than $O(d^{2k})$ for $p(X_t | X_{t-1})$



Factorial HMMs vs HMMs

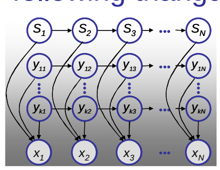
- Let us compare a factorial HMM with D chains, each with K values, to its *equivalent* HMM.

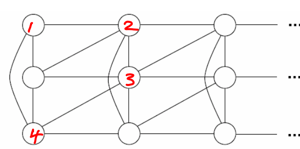


- Num. parameters to specify $p(X_t \mid X_{t-1})$
  - HMM: $t\,O(K^{2D})$
  - fHMM: $t\,O(K^2 D)$

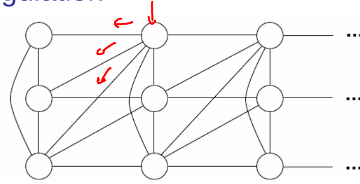
Triangulating fHMM

- Is the following triangulation correct?



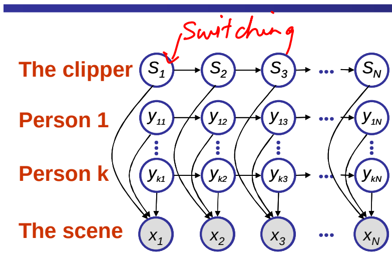


- Here is a triangulation



- We have created cliques of size k+1, and there are O(kT) of them. The junction tree algorithm is not efficient for factorial HMMs.

Special case: switching HMM

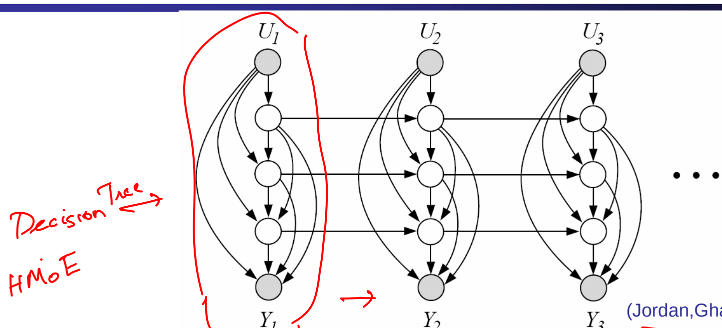


- Different chains have different state space and different semantics
- The exact calculation is intractable and we must use approximate inference methods

Multi-View Face Tracking with Factorial and Switching HMM

Peng Wang , Qiang Ji
Department of Electrical, Computer and System Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180

Hidden Markov decision trees

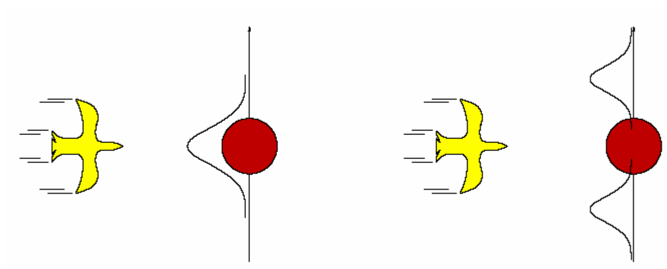


(Jordan,Ghahramani,&Saul,197)

- A combination of decision trees with factorial HMMs
- This gives a "command structure" to the factorial representation
- Appropriate for multi-resolution time series
- Again, the exact calculation is intractable and we must use approximate inference methods

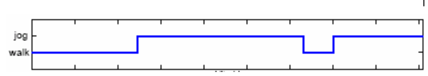
Problems with SSMs

- linearity
- Gaussianity
- Uni-modality

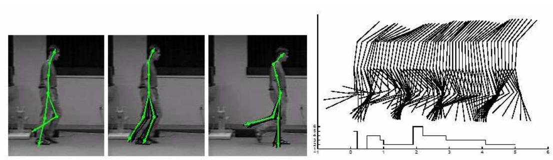


Switching SMM

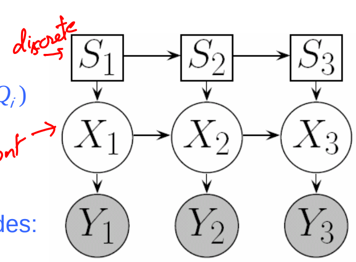
- Possible world:
  - multiple motion state:



- Task:
  - Trajectory prediction



- Model:
  - Combination of HMM and SSM



$$p(X_t = x_t \mid X_{t-1} = x_{t-1}, S_t = i) = \mathcal{N}(x_t; A_i x_{t-1}, Q_i)$$

$$p(Y_t = y_t \mid X_t = x_t) = \mathcal{N}(t_t; Cx_t, R)$$

$$p(S_t = j \mid S_{t-1} = i) = M(i, j)$$

  - Belief state has O(k^t) Gaussian modes: