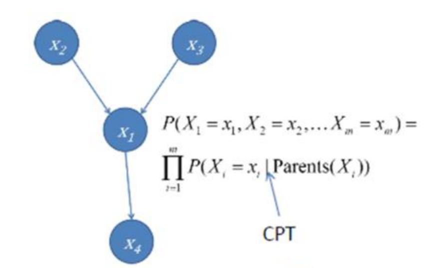
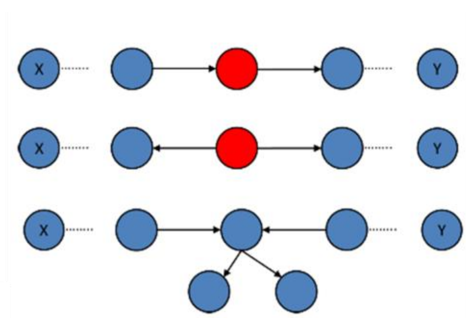
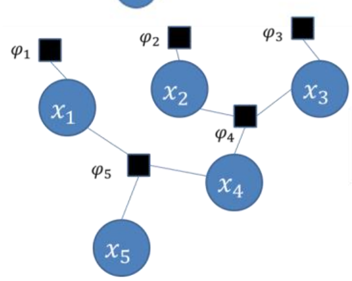
Reasoning with uncertainty III







$$m_{ji}(x_i) = \max_{x_j} f_{ij}(x_i, x_j) \prod_{N(j)\setminus i} m_{kj}(x_j)$$
$$m_{ji} = \sum_{x_j} f_{ij}(x_i, x_j) \prod_{N(j)\setminus i} m_{kj}(x_j)$$

Sum- and max-product, belief propagation
Exact on trees
Junction trees
Complexity wrt treewidth

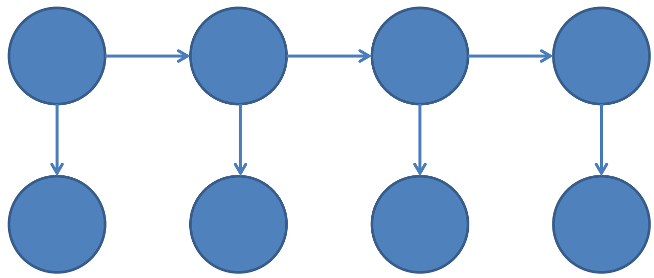
Temporal Models

- X_t = state of the world at time *t*
- Y_t = Observation (evidence, measurement,..) at time *t*
- Examples:
- *X* = types of actions, *Y* = observations from video
- *X* = location/orientation in the world, *Y* = observations

Assumptions

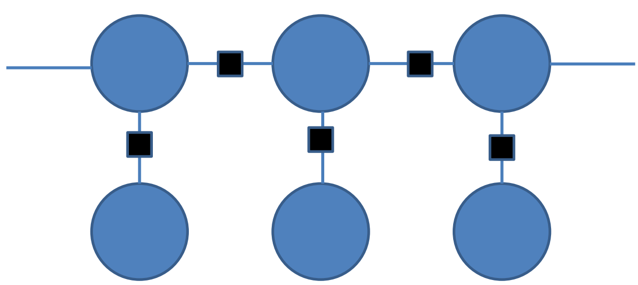
- Stationarity
  - Technically, infinite set of variables
  - Assume that process does not change over time: the conditional probabilities that define the model do not change over time
- Markov
  - State and measurements depend only on variables within a bounded time window
  - Conditionally independent from all the other variables conditioned on that time window
  - $P(X_t | X_{[1:t-1]}) = P(X_t | X_{[t-d:t-1]})$
  - $P(Y_t | X_{[1:t]}, Y_{[0:t-1]}) = P(Y_t | X_t)$

Example



- First-order:
- From d-separation, X_{t+1} is conditionally independent of X_{t-1} given X_t
- No independence of measurements Yt

Alternate view

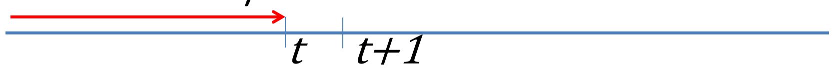


- Polytree in directed representation
- Tree in factored graph

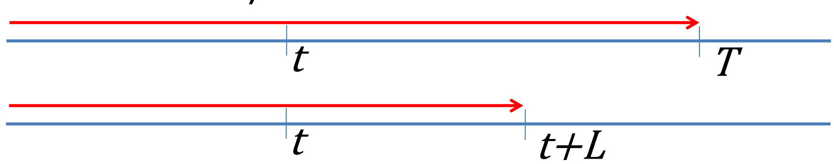
Operations

- Filtering: P(current state given all previous observations)

- Prediction: P(future states given all previous observations)



- Smoothing: P(past state given all observations)



Operations

- Filtering: P(current state given all previous observations)

$$P(X_{t+1}|y_{1:t+1})\alpha P(y_{t+1}|X_{t+1})\sum_{x_t} P(X_{t+1}|x_t)P(x_t|y_{1:t})$$

- Prediction: P(next state given all previous observations)

$$P(X_{t+1}|y_{1:t})\alpha \sum_{x_t} P(X_{t+1}|x_t)P(x_t|y_{1:t})$$

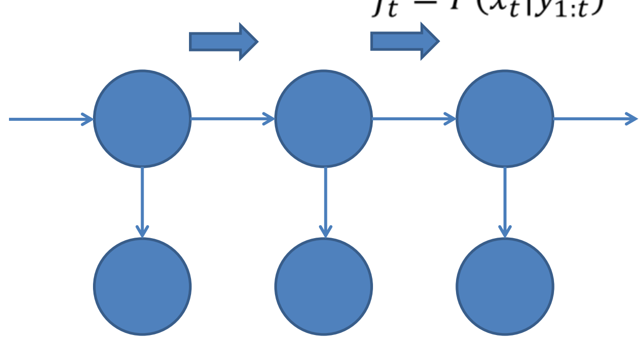
- Smoothing: P(past state given all observations)
- $P(X_k|y_{1:t})\alpha P(X_k|y_{1:k})P(y_{k+1:t}|X_k)$
- $P(y_{k+1:t}|X_k)$= $\sum_{x_{k+1}} P(y_{k+1}|x_{k+1})\, P(x_{k+1}|X_k)P(y_{k+2:t}|x_{k+1})$

- $P(X_{t+L+1}|y_{1:t}) \alpha \sum_{x_t} P(X_{t+L+1}|x_{t+L}) P(x_{t+L}|y_{1:t})$

- Further prediction?
- No added evidence
- Stationary distribution

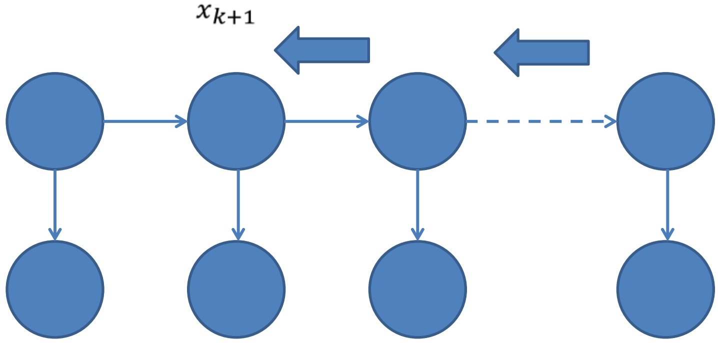
$$P(X_{t+1}|y_{1:t+1}) \alpha P(y_{t+1}|X_{t+1}) \sum_{x_t} P(X_{t+1}|x_t) f_t$$

$$f_t = P(x_t|y_{1:t})$$



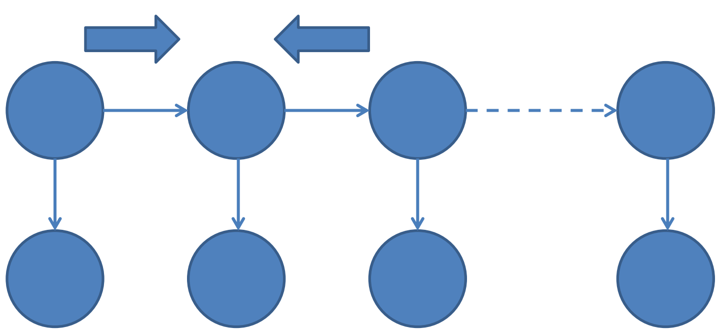
- Estimate from past: Propagate forward in time

$$b_{k+1:t} = P(y_{k+1:t} | X_k)$$
$$b_{k+1:t} = \sum_{x_{k+1}} P(y_{k+1} | x_{k+1})\, P(x_{k+1} | X_k) b_{k+2:t}$$



- Estimate from future: Propagate   backward in time

$$P(X_k | y_{1:t}) \alpha f_k b_{k+1:t}$$



- Estimate from both: Combine forward and backward terms
- Inference over 1:t linear in number of states
- (Polytree case in which inference is efficient; sum-product)

Inferring the most likely set of states

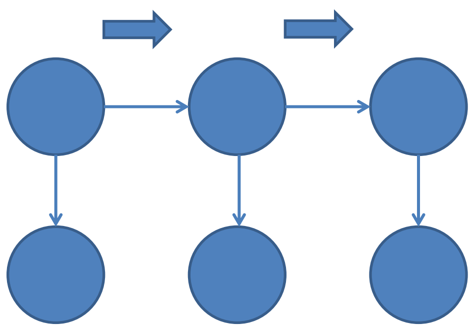
Same as max-product vs. sum-product

$$\max_{x_1,..,x_t} P(x_1,..,x_t,X_{t+1}|y_{1:t+1})\alpha$$

$$P(y_{t+1}|X_{t+1})\max_{x_t}(P(X_{t+1}|x_t)\max_{x_1,..,x_{t-1}} \mathrm{P}(x_1,..,x_{t-1},x_t|y_{1:t})$$

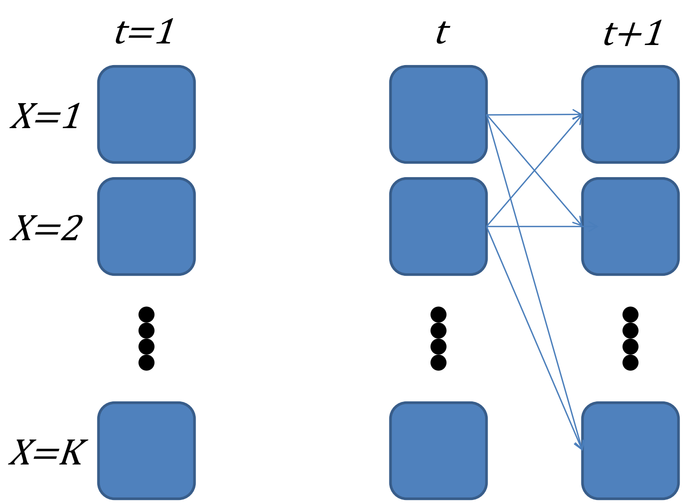
$$m_{1:t+1}\alpha$$

$$P(y_{t+1}|X_{t+1})\max_{x_t}(P(X_{t+1}|x_t)m_{1:t})$$



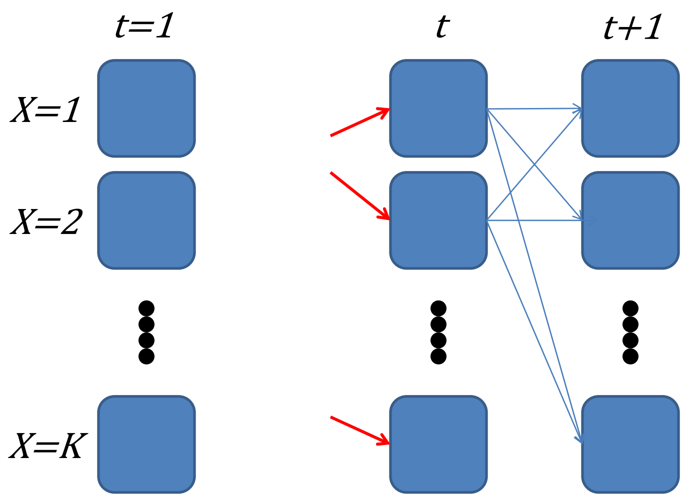
Special case: HMM

- X discrete with K states
- K^2 transitions: $P(X_{t+1}|X_t)$
- Emission probabilities $P(Y_t|X_t)$



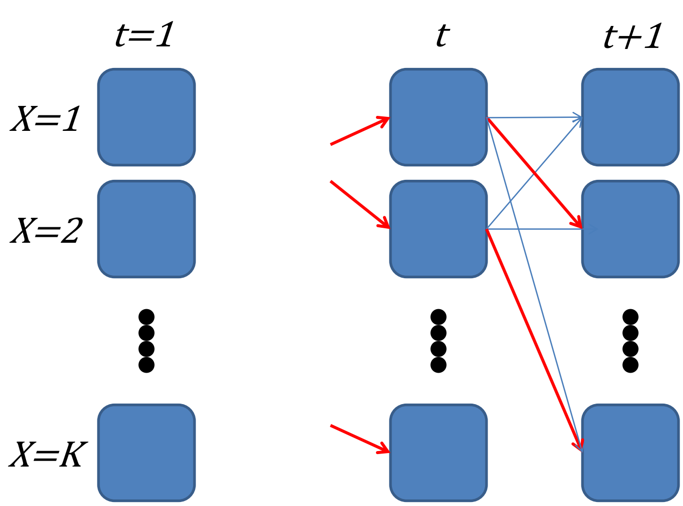
Special case: HMM

- $m_{1:t}(x_t) = \max_{x_1,..,x_{t-1}} P(x_1, .., x_t | y_{1:t})$



Special case: HMM

$$m_{1:t+1}(x_{t+1}) = P(y_{t+1}|x_{t+1}) \max_{x_t}(P(x_{t+1}|x_t)m_{1:t})$$



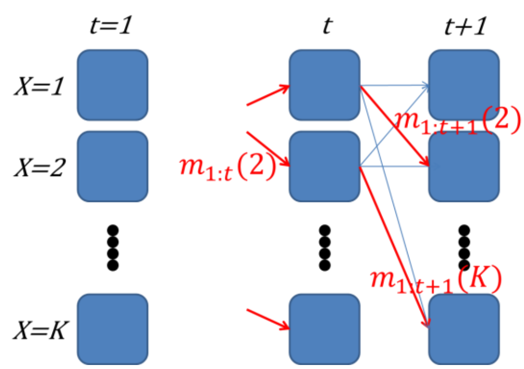
Special case: HMM

( )

(|) (|)

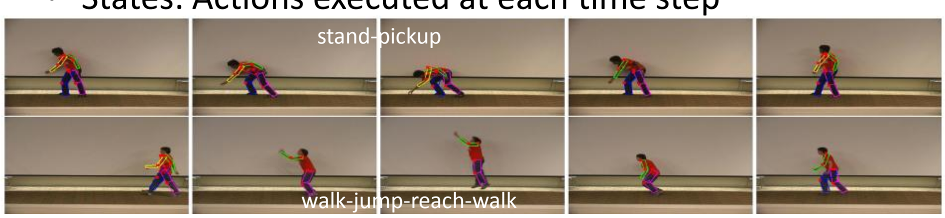
→ DP

- Special case of max-product (same as tree example the last time)
- time in general (excepting special form of (|)



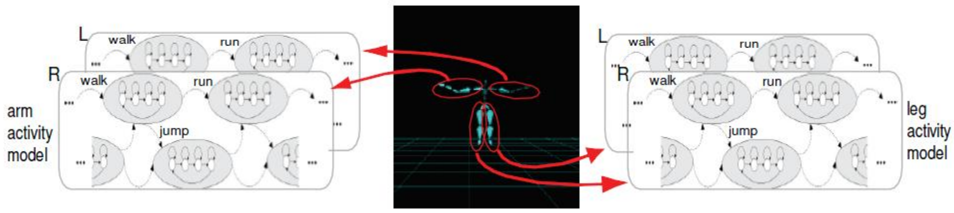
Example

- Observations: (Noisy) estimates of link locations
- States: Actions executed at each time step

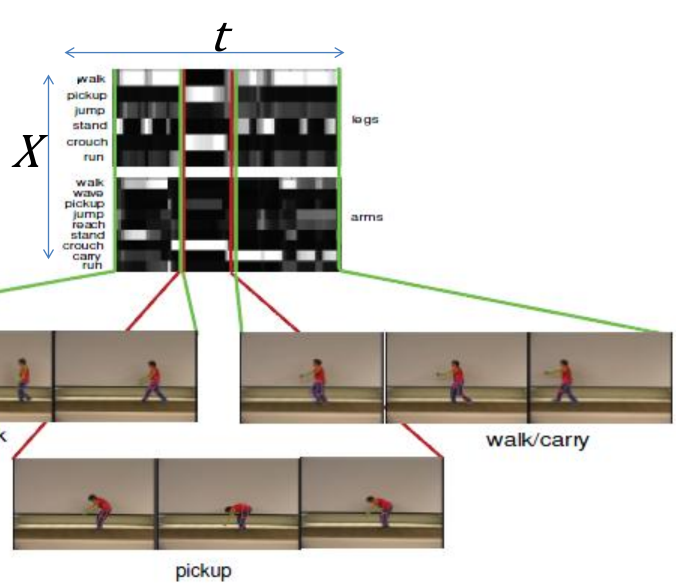


Noisy!

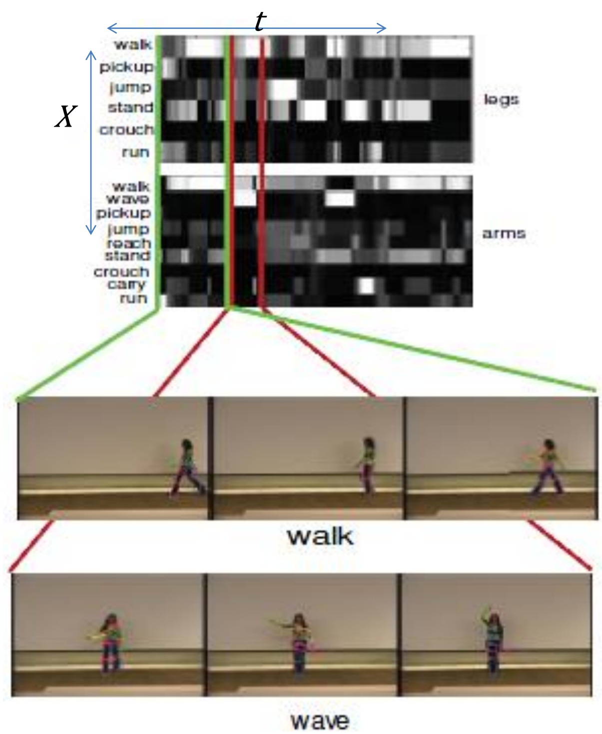
Nazli Ikizler and David Forsyth, "Searching video for complex activities with finite state models" IEEE Conference on Computer Vision and Pattern Recognition, 2007

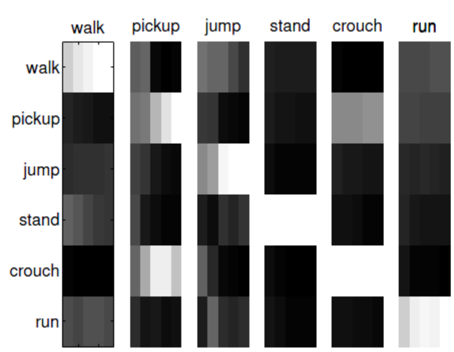


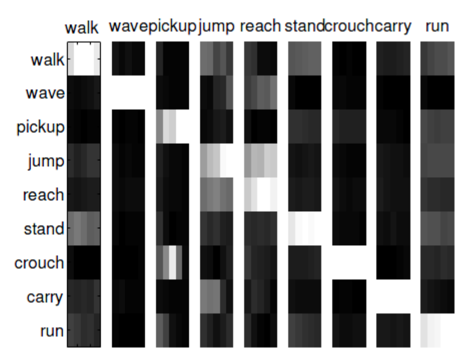
- Different action model for each part (learned from motion capture data)



Nazli Ikizler and David Forsyth, "Searching video for complex activities with finite state models" IEEE Conference on Computer Vision and Pattern Recognition, 2007

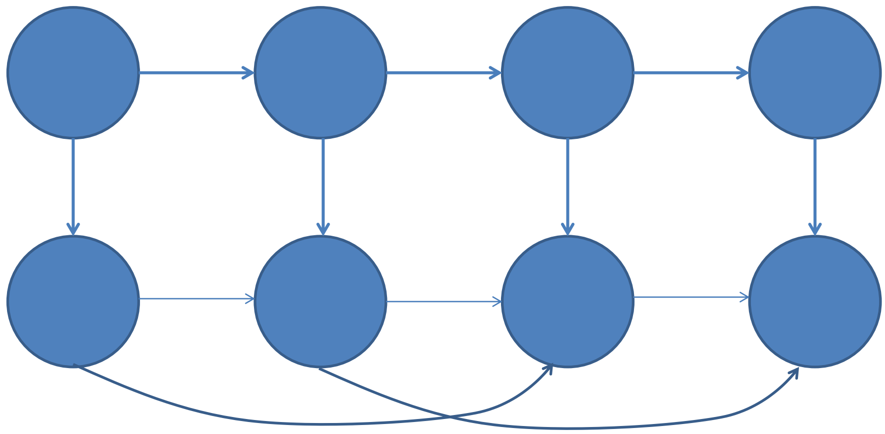






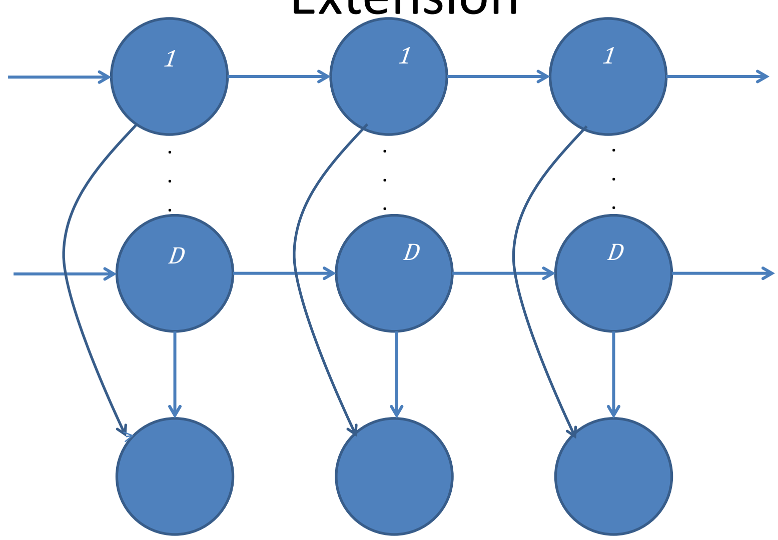
Nazli Ikizler and David Forsyth, “Searching video for complex activities with finite state models” IEEE Conference on Computer Vision and Pattern Recognition, 2007

Extension

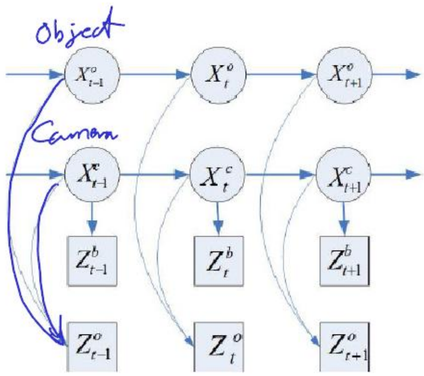


- Longer-range connections (e.g., tracking, ..)
- No problem (in principle): Update representation of

Extension



- X^i and X^j are not separated
- Need to represent all states
- Grouping states does not solve the problem: Standard HMM back/forward is NK^{2D}



Mei and Porikli ‘08

Linear dynamical models

- Canonical case: $x = [u, v, \dot{u}, \dot{v}]$
- What parametric representations?
- Desirable property (closure):

$$P(X_{t+1}|y_{1:t+1}) \alpha P(y_{t+1}|X_{t+1}) \sum_{x_t} P(X_{t+1}|x_t) P(x_t|y_{1:t})$$

and

$$P(X_t|y_{1:t})$$

must be of the same form (only parameters change)

- Need to use exponential family
- Gaussian model $N(\mu, \Sigma)$ satisfies closure
- Other models grow in complexity without bounds

- Linear transform of variable $A \sim N(\mu, \Sigma)$ to $MA + U$ $U \sim N(0, \Gamma)$ yields $\mathrm{N}(M\mu, M\Sigma M^T + \Gamma)$

Linear dynamical models

$$P(X_{t+1}|X_t) = N(AX_t, \Sigma)$$
$$P(Y_t|X_t) = N(BX_t, \Gamma)$$

- Equivalent to:

$$X_{t+1} = AX_t + w \quad w \sim N(0, \Sigma)$$
$$Y_t = BX_t + v \quad v \sim N(0, \Gamma)$$

E.g., constant velocity:

- $x = [u, v, \dot{u}, \dot{v}]$
- $A = [I\ \Delta t; 0\ I]$

Linear dynamic systems

Gaussian model propagates through the chain:

$$P(x_{t+1}|y_{1:t+1}) \alpha P(y_{t+1}|x_{t+1}) \sum_{x_t} P(x_{t+1}|x_t) P(x_t|y_{1:t})$$

$$P(x_{t+1}|y_{1:t+1}) \alpha P(y_{t+1}|x_{t+1}) \int P(x_{t+1}|x_t) P(x_t|y_{1:t}) dx_t$$

$P(x_t|y_{1:t}) \sim N(\mu_t, U_t)$

$$\mu_{t+1} = A\mu_t + K(y_{t+1} - BA\mu_t)$$
$$V_t = AU_tA^T + \Sigma$$
$$K = V_tB^T(BV_tB^T + \Gamma)^{-1}$$
$$U_{t+1} = (I - KB)V_t$$

Linear dynamic systems

Gaussian model propagates through the chain:

$$P(x_{t+1}|y_{1:t+1})\alpha P(y_{t+1}|x_{t+1})\sum_{x_t} P(x_{t+1}|x_t)P(x_t|y_{1:t})$$

$$P(x_{t+1}|y_{1:t+1})\alpha P(y_{t+1}|x_{t+1})\int P(x_{t+1}|x_t)P(x_t|y_{1:t})dx_t$$

$$P(x_t|y_{1:t}) \sim N(\mu_t, U_t)$$

Innovation

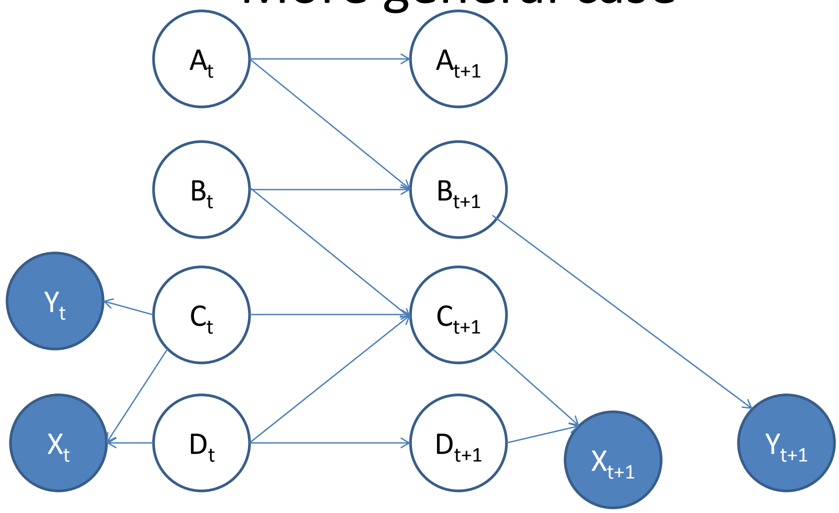
Gain =
B^{-1} if trust new observation
0 if trust X_t

$$\mu_{t+1} = A\mu_t + K(y_{t+1} - BA\mu_t)$$
$$V_t = AU_tA^T + \Sigma$$
$$K = V_tB^T(BV_tB^T + \Gamma)^{-1}$$
$$U_{t+1} = (I - KB)V_t$$

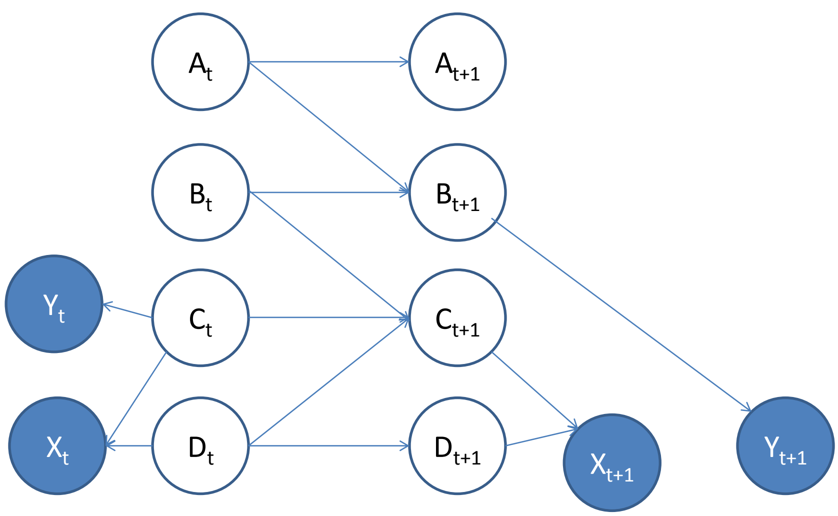
Predicted covariance without any measurements

- Non linear case:
  - $X_{t+1} = f(X_t) + w \quad w \sim N(0, \Sigma)$
- First-order approximation $f(X) \sim f(\mu) + J(X - \mu)$
- Local approximation
  - JX_t
- Local linear approximation can be very bad (e.g, X = [x y θ])
- Other possibility: Sample U_t, transform the samples, estimate V_t from the samples

More general case

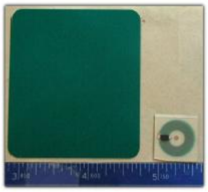


- Arbitrary connections between state and observation variables at any time *t*
  1. Replicate over time (unroll) → General graph, can't do exact inference (in general)
  2. Collapse state variables wrt observed → K^D state tables in general
- In the discrete case, DBN <=> HMM but note the complexity issue



- Arbitrary connections between state and observation variables at any time *t*
  1. Replicate over time (unroll) → General graph, can't do exact inference directly (in general)
  2. Collapse state variables wrt observed → K^D state tables in general
- In the discrete case, DBN <=> HMM but note the complexity issue
- Alternative
  - Sampling
  - Assumed density

Example

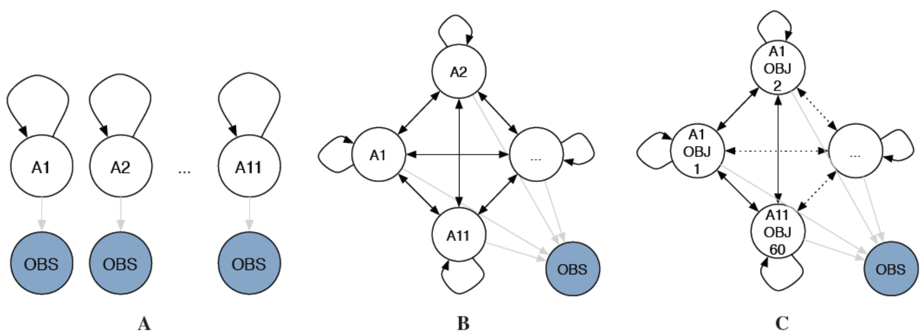


1	Using the bathroom
2	Making oatmeal
3	Making soft-boiled eggs
4	Preparing orange juice
5	Making coffee
6	Making tea
7	Making or answering a phone call
8	Taking out the trash
9	Setting the table
10	Eating breakfast
11	Clearing the table

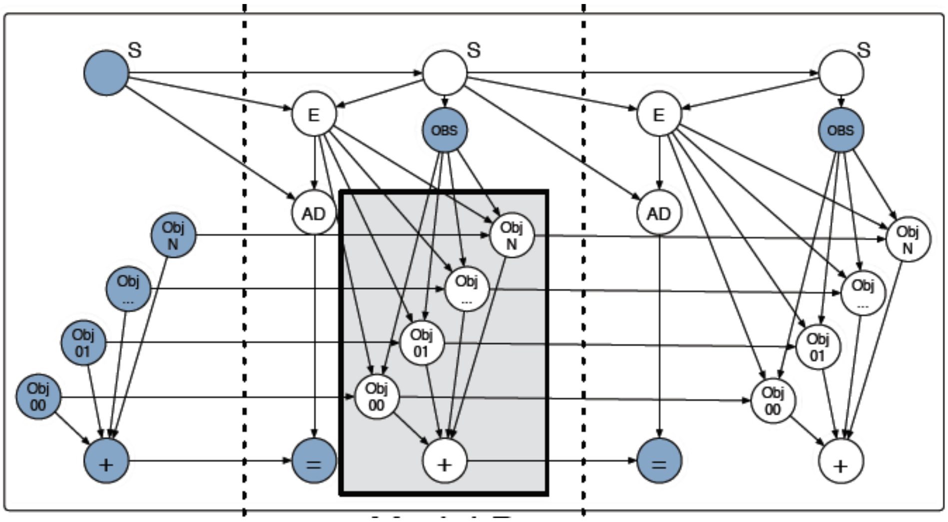
bowl, coffee container, coffee grinder, coffee tamper, cupboard(6), dishwasher, door(2), drawer(2), egg carton, espresso cup(2), espresso handle, espresso steam knob, espresso switches, faucet(2), freezer, milk, hand soap, juice, juice pitcher, kettle, measuring cup-half, measuring cup-one, measuring scoop, milk steaming pitcher, mug, oatmeal, refrigerator, salt, saucepan, cooking spoon, stove control(2), sugar, table cup(4), table plate(4), table spoon(4), tea bag, tea box, telephone, toilet flush handle, toilet lid, vanilla syrup

- Observations: IDs of (60) objects manipulated (RFID tags)
- State: Activity performed (11 fine-grained activities requiring extensive observations)
- Hypothesis: "invisible human hypothesis"

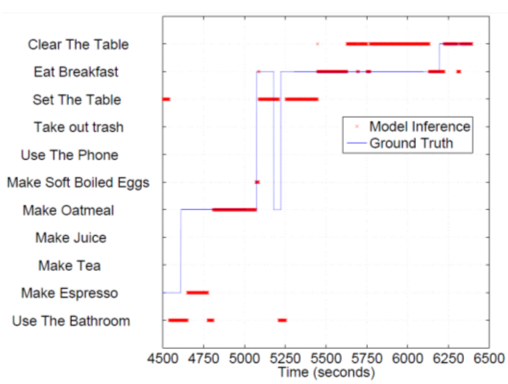
Patterson, D.J.; Fox, D.; Kautz, H.; Philipose, M. Fine-grained activity recognition by aggregating abstract object usage. Intern. Symp. Wearable Computers.



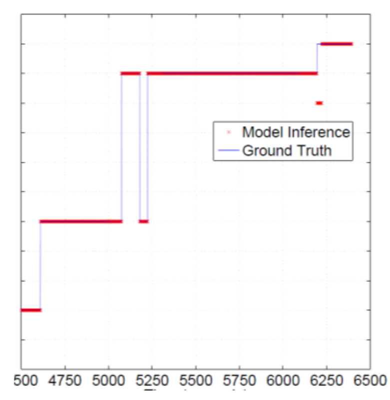
- Baselines:
  - A. Each activity has its own HMM (11 HMMs) → take the best
  - B. A single HMM for all the activities (11-valued states)
  - C. A single HMM with state = activities x objects (660-valued states)

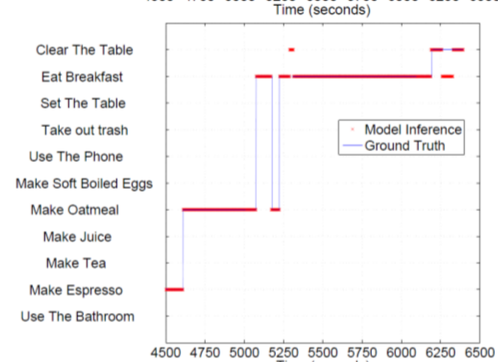


- More complicated relations:
  - The number of objects is an indication of the type of activities (setting the table vs. eating breakfast)
  - E node (Exit) indicates end of previous activity
  - AD node (Aggregate Distribution)

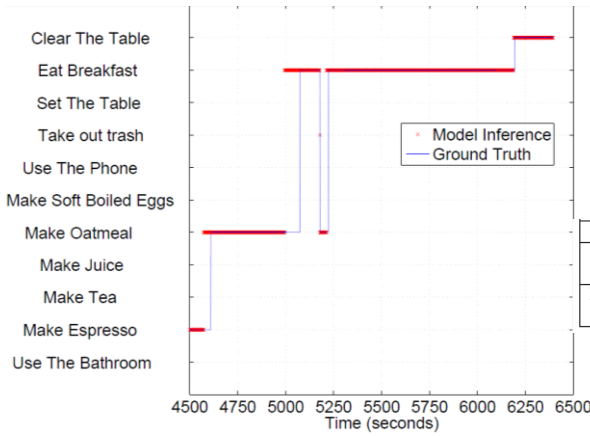


A





B



		Accuracy			
Time-Slice	μ	68%	88%	87%	88%
	(σ)	(5.9)	(4.2)	(9.3)	(3.1)
Edit Distance	μ	12	9	14	7
	(σ)	(2.9)	(6.2)	(10.4)	(2.2)

- Does not scale well (660^2 tables) for C
- Better representation of relations in D
- Hierarchical representation of object list to address robustness issues?

Patterson, D.J.; Fox, D.; Kautz, H.; Philipose, M. Fine-grained activity recognition by aggregating abstract object usage. Intern. Symp. Wearable Computers.