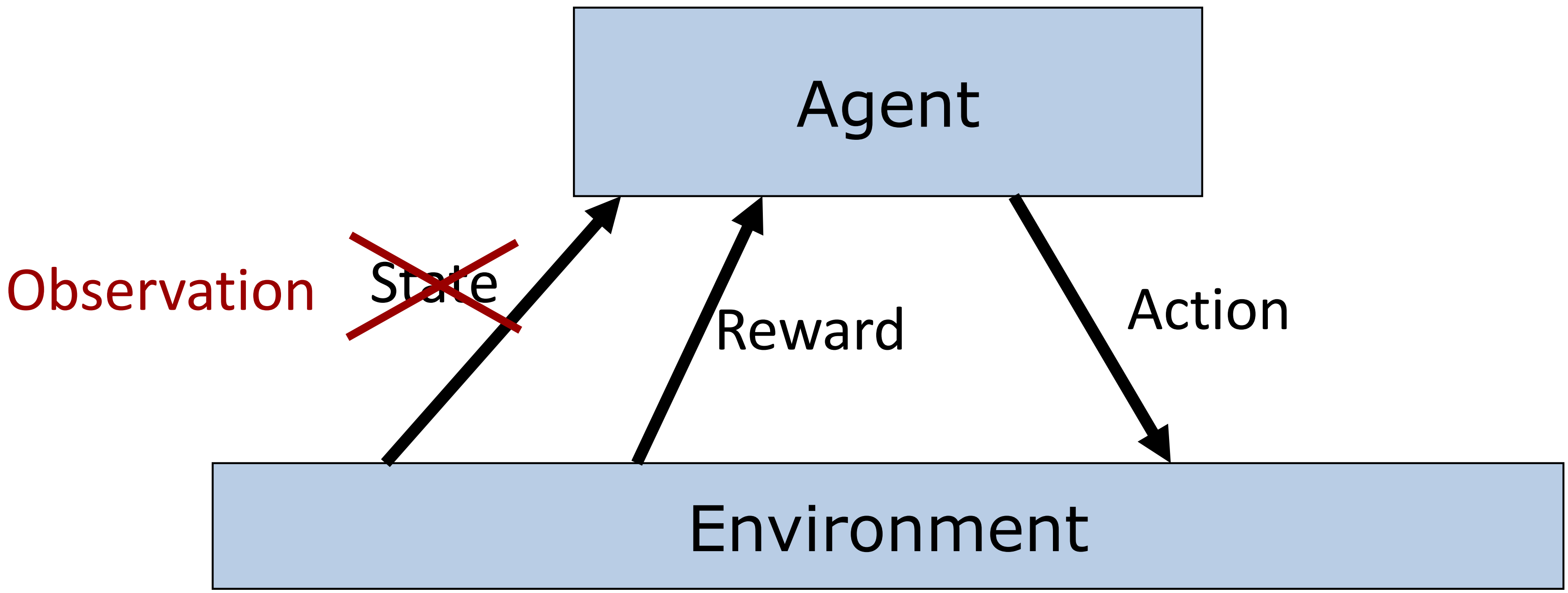
CS885 Reinforcement Learning
Lecture 11a: June 6, 2018

Hidden Markov Models

[RusNor] Sec. 15.3 [SutBar] Sec. 17.3

Reinforcement Learning Problem



Goal: Learn to choose actions that maximize rewards

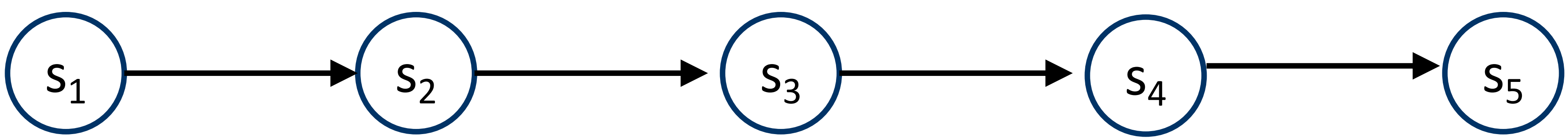
Markov Process

- Assumptions:
  - (first-order) Markovian:

$$\Pr(s_t | s_{t-1}, \dots, s_0) = \Pr(s_t | s_{t-1})$$

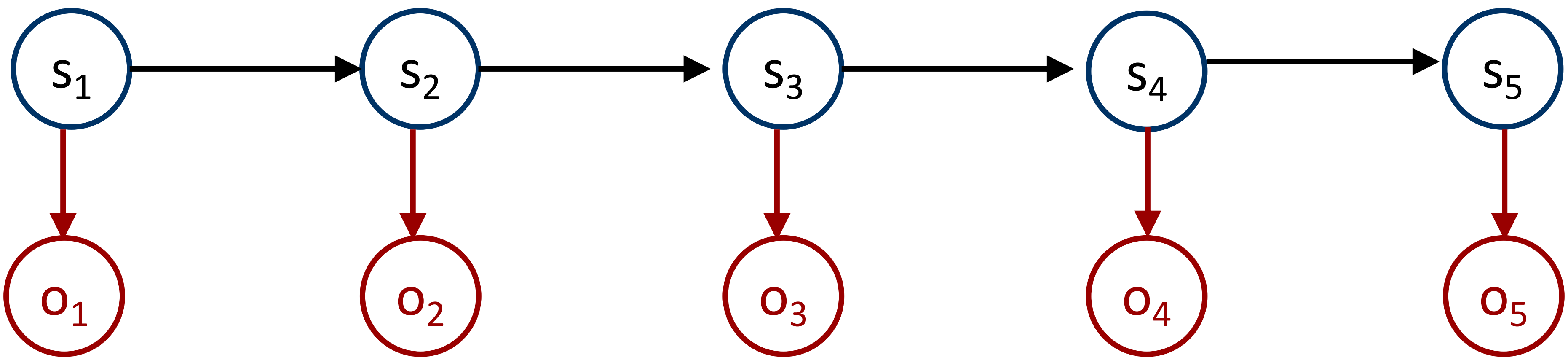
  - Stationary:

$$\Pr(s_t | s_{t-1}) = \Pr(s_{t+1} | s_t) \; \forall t$$

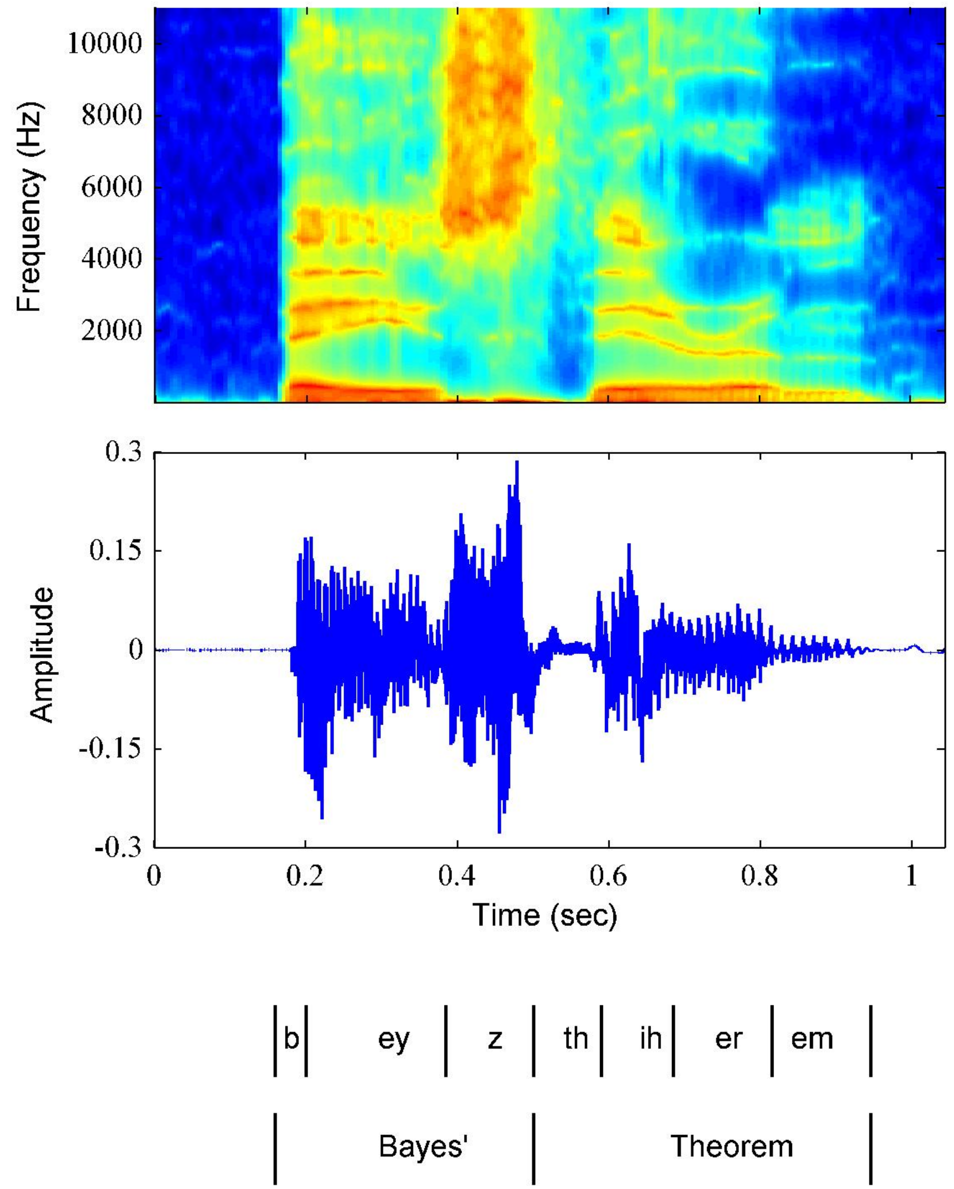


Hidden Markov Model

- Assumptions:
  - (first-order) Markovian:
    $$\Pr(s_t|s_{t-1},\ldots,s_0) = \Pr(s_t|s_{t-1})$$
  - Stationary:
    $$\Pr(s_t|s_{t-1}) = \Pr(s_{t+1}|s_t)\ \forall t$$
    $$\Pr(o_t|s_t) = \Pr(o_{t+1}|s_{t+1})\ \forall t$$

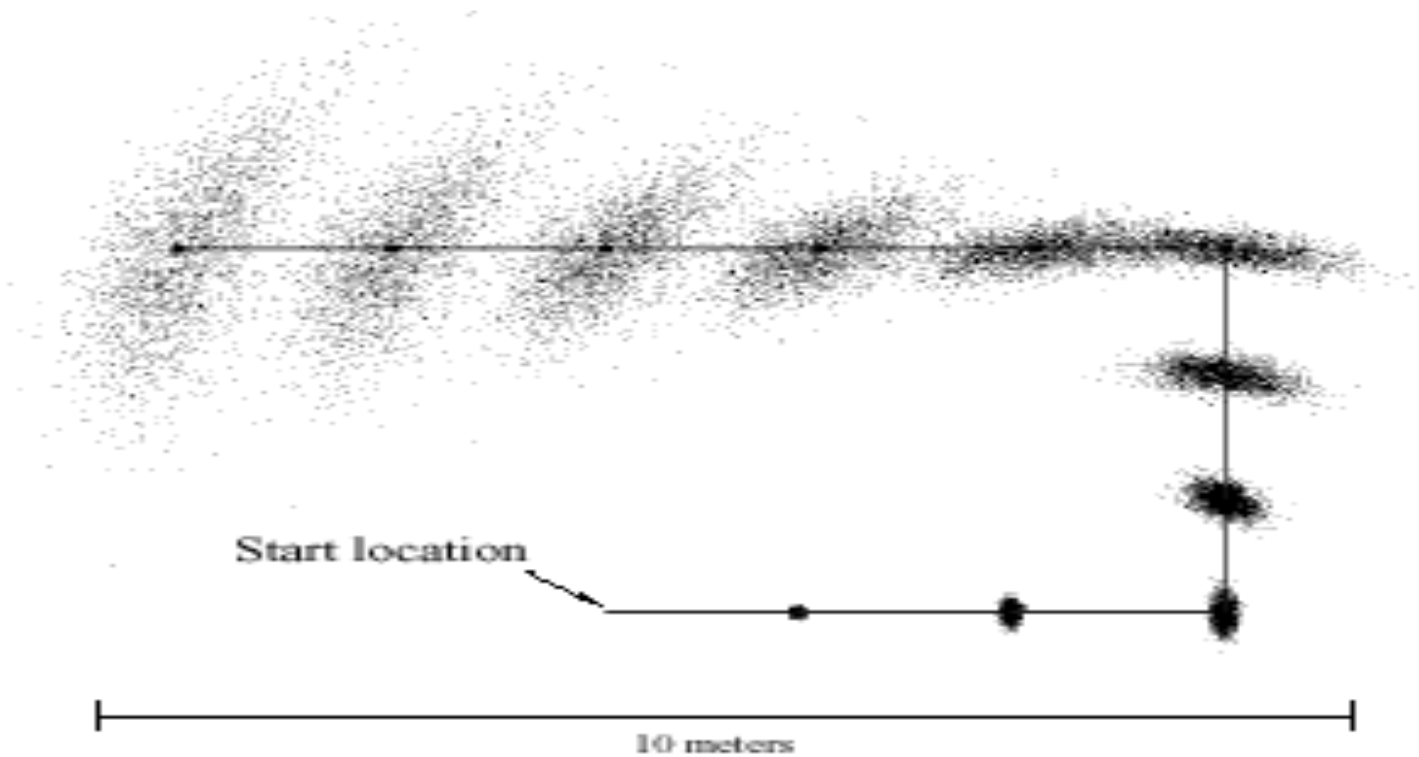


Speech Recognition



Mobile Robot Localisation

- Example of a Markov process



- Problem: uncertainty grows over time…

Mobile Robot Localisation

- Hidden Markov Model:

  $\boldsymbol{s}$: coordinates of the robot on a map

  $\boldsymbol{o}$: distances to surrounding obstacles (measured by laser range finders or sonars)

  $\mathbf{Pr}(\boldsymbol{s_t}|\boldsymbol{s_{t-1}})$: movement of the robot with uncertainty

  $\mathbf{Pr}(\boldsymbol{o_t}|\boldsymbol{s_t})$: uncertainty in the measurements provided by laser range finders and sonars

- **Localisation:** $\Pr(s_t|o_t, \dots, o_1)$?

Inference in temporal models

- Four common tasks:
  - **Monitoring:** $\Pr(s_t|o_{1..t})$
  - **Prediction:** $\Pr(s_{t+k}|o_{1..t})$
  - Hindsight: $Pr(s_k|o_{1..t})$ where $k < t$
  - Most likely explanation:

$$argmax_{s_1,\dots,s_t} \Pr(s_{1..t}|o_{1..t})$$

- What algorithms should we use?

Monitoring

- $\Pr(s_t|o_{1..t})$: distribution over current state given observations
- Examples: robot localisation, patient monitoring
- Recursive computation:

$\Pr(s_t|o_{1..t}) \propto \Pr(o_t|s_t, o_{1..t-1})\Pr(s_t|o_{1..t-1})$ by Bayes' thm

$= \Pr(o_t|s_t)\Pr(s_t|o_{1..t-1})$ by conditional independence

$= \Pr(o_t|s_t)\sum_{s_{t-1}}\Pr(s_t, s_{t-1}|o_{1..t-1})$ by marginalization

$= \Pr(o_t|s_t)\sum_{s_{t-1}}\Pr(s_t|s_{t-1}, o_{1..t-1})\Pr(s_{t-1}|o_{1..t-1})$

by chain rule

$= \Pr(o_t|s_t)\sum_{s_{t-1}}\Pr(s_t|s_{t-1})\Pr(s_{t-1}|o_{1..t-1})$ by cond. ind.

Forward Algorithm

- Compute $\Pr(s_t|o_{1..t})$ by forward computation

$\Pr(s_1|o_1) \propto \Pr(o_1|s_1)\Pr(s_1)$

For $i = 2$ to t do

$\quad \Pr(s_i|o_{1..i}) \propto \Pr(o_i|s_i) \sum_{s_{i-1}} \Pr(s_i|s_{i-1}) \Pr(s_{i-1}|o_{1..i-1})$

End

- Linear complexity in t

Prediction

- $\Pr(s_{t+k}|o_{1..t})$: distribution over future state given observations
- Examples: weather prediction, stock market prediction

- Recursive computation

$\Pr(s_{t+k}|o_{1..t}) = \sum_{s_{t+k-1}} \Pr(s_{t+k}, s_{t+k-1}|o_{1..t})$ by marginalization

$= \sum_{s_{t+k-1}} \Pr(s_{t+k}|s_{t+k-1}, o_{1..t}) \Pr(s_{t+k-1}|o_{1..t})$ by chain rule

$= \sum_{s_{t+k-1}} \Pr(s_{t+k}|s_{t+k-1}) \Pr(s_{t+k-1}|o_{1..t})$ by cond. ind.

Forward Algorithm

1. Compute $\Pr(s_t|o_{1..t})$ by forward computation

   $\Pr(s_1|o_1) \propto \Pr(o_1|s_1)\Pr(s_1)$
   For $i = 1$ to t do
   $$\Pr(s_i|o_{1..i}) \propto \Pr(o_i|s_i) \sum_{y_{i-1}} \Pr(s_i|s_{i-1}) \Pr(s_{i-1}|o_{1..i-1})$$
   End

2. Compute $\Pr(s_{t+k}|o_{1..t})$ by forward computation

   For $j = 1$ to k do
   $$\Pr(s_{t+j}|o_{1..t}) = \sum_{s_{t+j-1}} \Pr(s_{t+j}|s_{t+j-1}) \Pr(s_{t+j-1}|o_{1..t})$$
   End

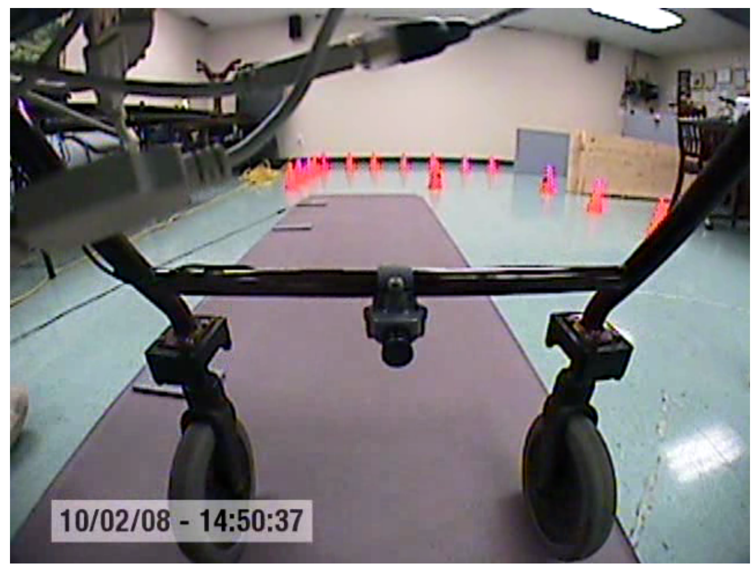
- Linear complexity in $t + k$

Case Study: Activity Recognition

- Task: infer activities performed by a user of a smart walker
  - Inputs: sensor measurements
  - Output: activity

Backward view

Forward view



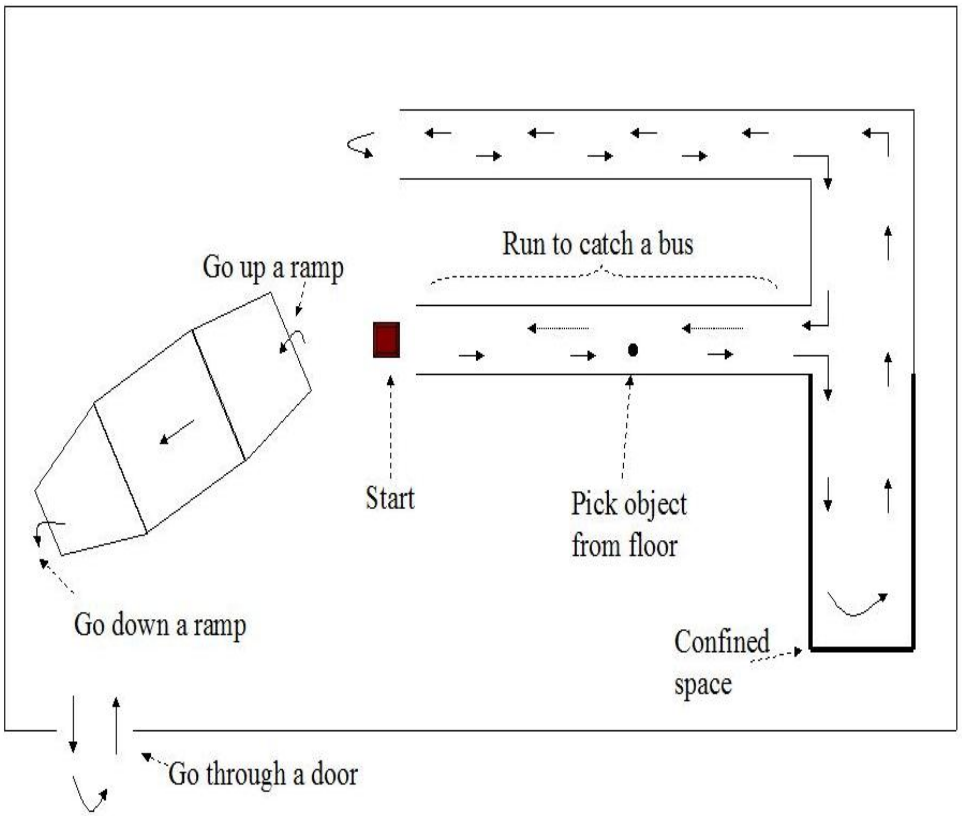
Inputs: Raw Sensor Data

- 8 channels:
  - Forward acceleration
  - Lateral acceleration
  - Vertical acceleration
  - Load on left rear wheel
  - Load on right rear wheel
  - Load on left front wheel
  - Load on right front wheel
  - Wheel rotation counts (speed)

- Data recorded at 50 Hz and digitized (16 bits)

Data Collection

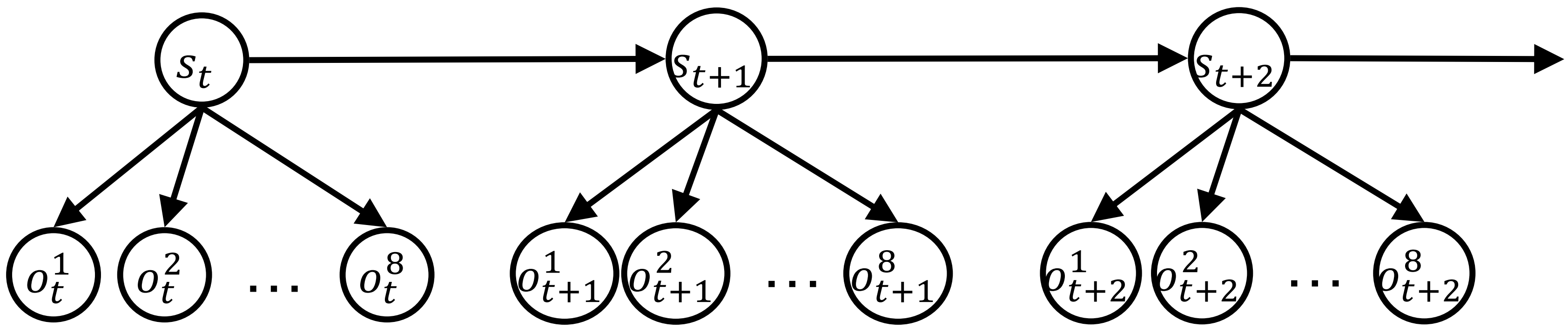
- 8 walker users at Winston Park (84-97 years old)
- 12 older adults (80-89 years old) in the Kitchener-Waterloo area who do not use walkers



Output: Activities

- Not Touching Walker (NTW)
- Standing (ST)
- Walking Forward (WF)
- Turning Left (TL)
- Turning Right (TR)
- Walking Backwards (WB)
- Sitting on the Walker (SW)
- Reaching Tasks (RT)
- Up Ramp/Curb (UR/UC)
- Down Ramp/Curb (DR/DC)

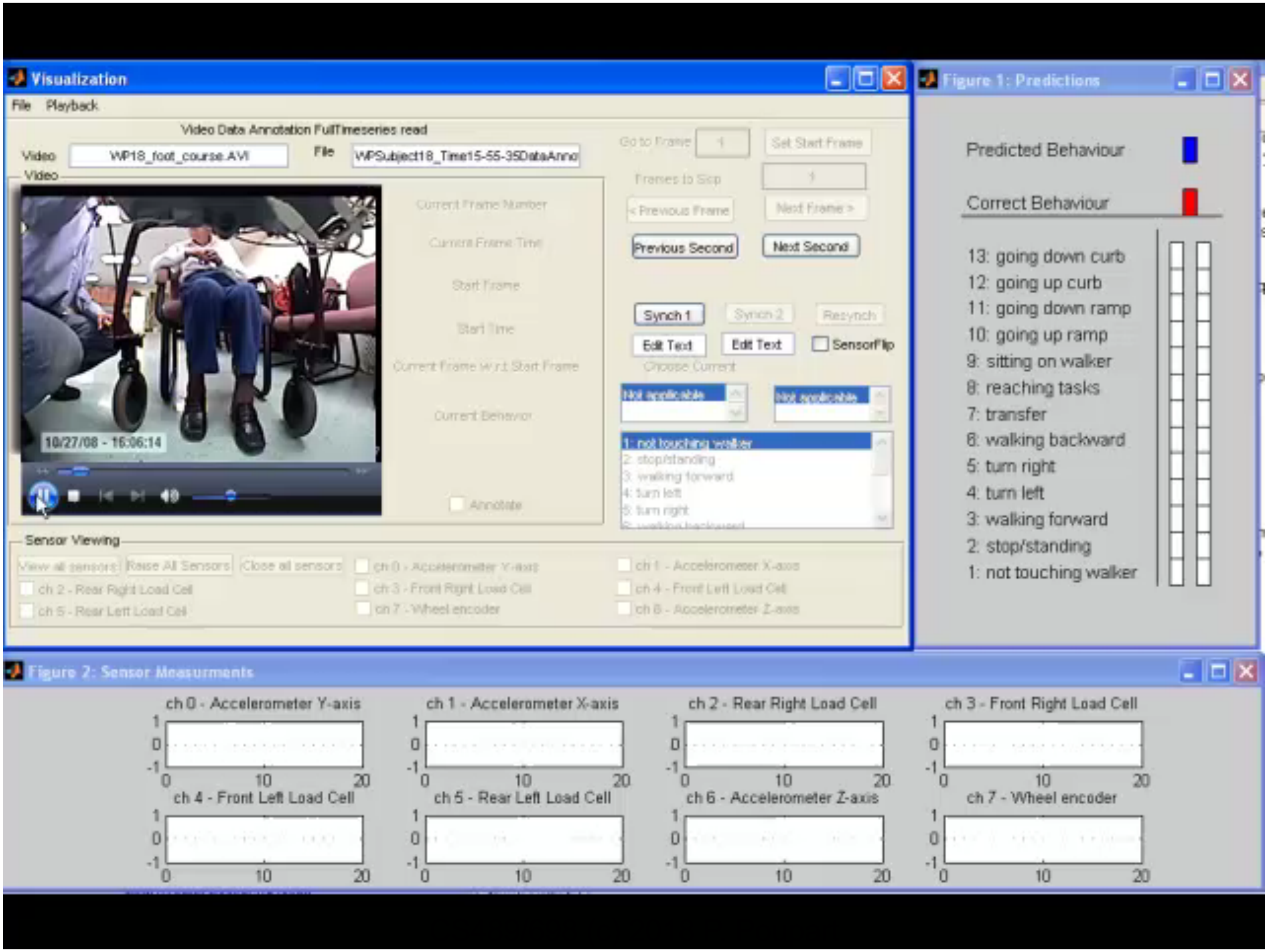
Hidden Markov Model (HMM)



- Parameters
  - Initial state distribution: $\psi_{class} = \Pr(s_1 = class)$
  - Transition probabilities: $\theta_{class'|class} = \Pr(s_{t+1} = class' | s_t = class)$
  - Observation probabilities: $\phi^i_{val|class} = \Pr(o_t^i = val | o_t = class)$

    or $N(val|\mu^i_{class}, \sigma^i_{class}) = \Pr(o_t^i = val | o_t = class)$
- Maximum likelihood:
  - Supervised: $\psi^*, \theta^*, \phi^* = argmax_{\psi,\theta,\phi} \Pr(s_{1:T}, o_{1:T} | \psi, \theta, \phi)$
  - Unsupervised: $\psi^*, \theta^*, \phi^* = argmax_{\psi,\theta,\phi} \Pr(o_{1:T} | \psi, \theta, \phi)$

Demo


Visualization
File
Playback
Video Data Annotation FullTimeseries read
Video
WP18_foot_course.AVI
File
WPSubject18_Time15-55-35DataAnno
10/27/08 - 16:06:14
Previous Second
Next Second
Synch 1
Edit Text
Edit Text
SensorFlip
Not applicable
Not applicable
1: not touching walker
Sensor Viewing
Figure 1: Predictions
Predicted Behaviour
Correct Behaviour
13: going down curb
12: going up curb
11: going down ramp
10: going up ramp
9: sitting on walker
8: reaching tasks
7: transfer
6: walking backward
5: turn right
4: turn left
3: walking forward
2: stop/standing
1: not touching walker
Figure 2: Sensor Measurments
ch 0 - Accelerometer Y-axis
ch 1 - Accelerometer X-axis
ch 2 - Rear Right Load Cell
ch 3 - Front Right Load Cell
ch 4 - Front Left Load Cell
ch 5 - Rear Left Load Cell
ch 6 - Accelerometer Z-axis
ch 7 - Wheel encoder