Reading Assignments

AIMA (Russell and Norvig)

- Ch 15.1-.3, 20.3 — State Estimation and Hidden Markov Models

From Monday:

- Ch 13 — Review of Probabilities
- Ch 14.1-4 — Probabilistic Reasoning

- No homework due this week.

Outline

- Overview
- Belief State Update
  - Sequential, Model-based Diagnosis
  - Robot Localization

Multiple Faults Occur

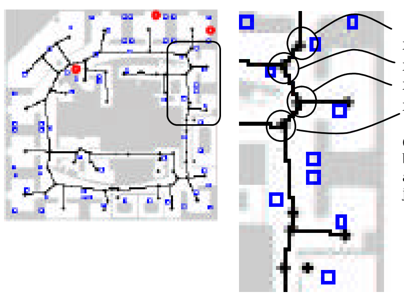
APOLLO 13

- three shorts, tank-line and pressure jacket burst, panel flies off.

➔ How do we rank order a large set of consistent diagnoses?

9/13/00 4

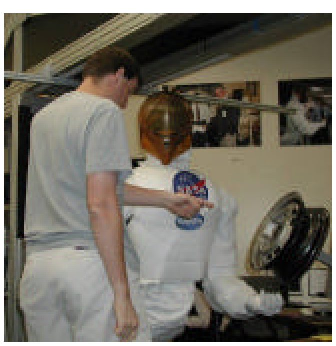
How does a robot determine its position from noise data, as it moves around?



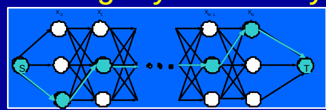
x_1
x_2: $p(x_2|x_1,a)= .9$
x_3: $p(x_3|x_1,a)=.05$
x_4: $p(x_4|x_1,a)=.05$

Observations can be features such as corridor features, junction features, etc.

How does a robot understand what is being said to it?



Estimating Dynamic Systems

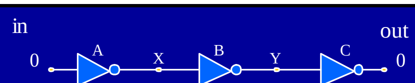


Given a sequence of observations and commands:

- What is the likelihood of a particular state?
  - ⇨ **Belief State Update**: (filtering and smoothing)
- What is the most likely sequence of states that got me here?
  - ⇨ Decoding: (Viterbi Algorithm)
- What is the most likely sequence of observations generated?
  - ⇨ Evaluation/Prediction:
- What HMM most likely generated these observations?
  - ⇨ Learning: (Baum-Welch Algorithm, Expectation-Maximization Algorithm)

Outline

- Overview
- Belief State Update
  - Sequential, Model-based Diagnosis
  - HMMs and Belief State Update
  - Robot Localization



Diagnoses: (42 of 64 candidates)

Fully Explained Failures

- [A=G, B=G, C=S0]
- [A=G, B=S1, C=S0]
- [A=S0, B=G, C=G]

. . .

Partial Explained

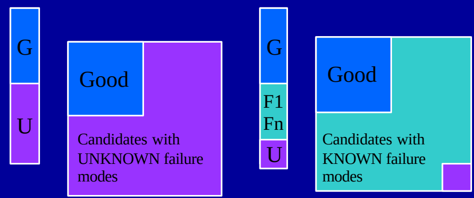
- [A=G, B=U, C=S0]
- [A=U, B=S1, C=G]
- [A=S0, B=U, C=G]

. . .

Fault Isolated, But Unexplained

- [A=G, B=G, C=U]
- [A=G, B=U, C=G]
- [A=U, B=G, C=G]

Due to the unknown mode, there tends to be an exponential number of diagnoses.



But these diagnoses represent a small fraction of the probability density space.

⇨ Most of the density space may be represented by enumerating the few most likely diagnoses

Sequential Model-based Diagnosis

Input:
- Set of component mode variables M, with finite domains.
- Set of observables X, with finite domains.
- Device model F over M and X, in propositional logic.
- Prior distribution $P(M_i)$ of mode assignments for each component i.
- Observation sequence $X_{1,n} = x_{1,n}$ provided dynamically.

Output:
- P(M) Prior Probability of Failure
- $P(M \mid X_{1,n} = x_{1,n})$ Posterior Given Observation updated after each observation is received.

Assume:
- Independence of component mode prior distribution.
- Conditional independence of observations given candidate (Naïve Bayes).
- Uniform distribution of observables, given candidate.

Candidate (Prior) Initial Probabilities

$$P(M) = P(M_n, M_{n-1,}, \ldots, M_1)$$
$$= P(m_n \mid m_{n-1} \ldots m_1) \ldots P(m_2 \mid m_1) P(m_1)$$

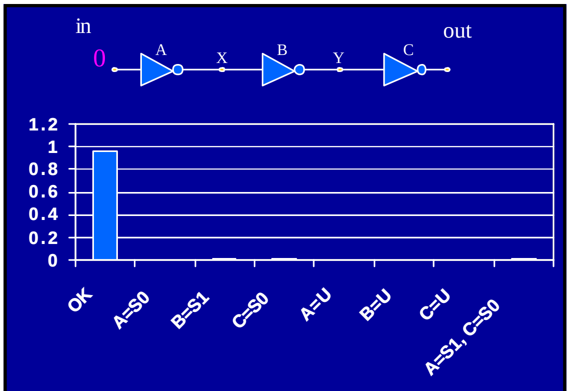
$$P(M) = \prod_{M_i \in M} P(M_i)$$

Assume Independence Of Initial Mode

	A	B	C
P(G)	.99	.99	.99
P(S1)	.008	.008	.001
P(S0)	.001	.001	.008
P(U)	.001	.001	.001

P(A=G, B=G, C=G) = .97
P(A=S1, B=G, C=G) = .008
P(A=S1, B=G, C=S0) = .00006
P(A=S1, B=S1, C=S0) = .0000005

Example: P(A = S1, B=G, C=S0) = .008 x .99 x .008 = .00006



9/13/00 13

Posterior Probability, after Observations $X_{1,n} = x_{1,n}$

$$P(M \mid x_1) = \frac{P(x_1 \mid M)P(M)}{P(x_1)}$$

Bayes' Rule

Normalization Term

$$= a\, P(x_1 \mid M) P(M)$$

For n > 1:

$$P(M \mid x_{1,n}) = a\, P(x_n \mid x_{1,n-1}, M) P(M \mid x_{1,n-1})$$

Observations are conditionally independent

$$= a\, P(x_n \mid M) P(M \mid x_{1,n-1})$$

9/13/00 14

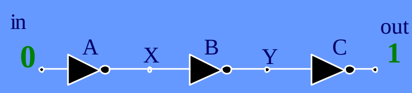
Estimating the Observation Probability $P(x_i \mid M)$

Assumption: All consistent observations for X_i are equally likely

$P(x_i \mid M)$ is estimated using model, F, according to:

- **If** previous observations $X_{1,i-1} = x_{1,i-1}$, M and F entails $X_i = x_i$
  Then $P(x_i \mid M) = 1$
- **If** previous observations $X_{1,i-1} = x_{1,i-1}$, M and F entails X_i ? v_i
  Then $P(x_i \mid c) = 0$
- **Otherwise**, Assume all consistent assignments to X_i are equally likely observations:
  let $D_{ci} \equiv \{x_c \in D_{Xi} \mid c, F$ is consistent with $X_i = x_c\}$
  Then $P(x_i \mid M) = 1/|D_{ci}|$

9/13/00 15

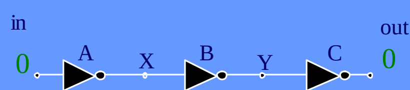


$$P(M \mid x_{1,n}) = a\, P(x_n \mid M)\, p(M \mid x_{1,n-1})$$

Observe out = 1:

- m = <A=G, B=G, C=G>
- Prior: P(m) = .9x.9x.9 = .97
- P(out = 1 | m) = ?
- = 1 (model entails out =1)
- P(m | out = 0) = ?
- = a x 1 x .97 = .97a

9/13/00 copyright Brian Williams, 2000 16

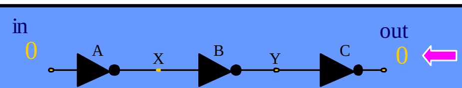


$$P(M \mid x_{1,n}) = a\, P(x_n \mid M)\, P(M \mid x_{1,n-1})$$

Observe out = 0:

- m = <A=G, B=G, C=G>
- P(m) = .97
- P(out = 0 | m) = ?
- = 0 (model inconsistent w out = 0)
- P(m | out = 0) = ?
- = 0 x .97 x a = 0

9/13/00 copyright Brian Williams, 2000 17



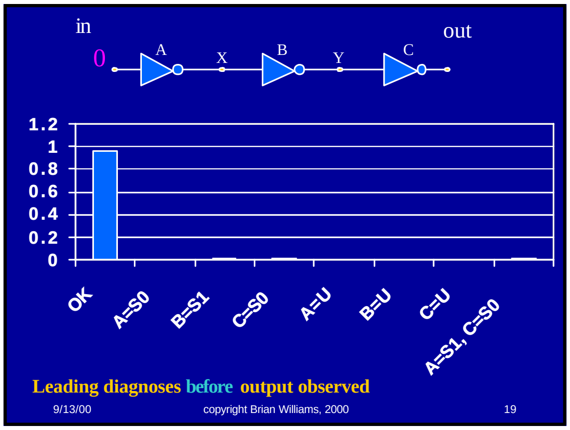
Example: Track single fault probabilities as observations are made:

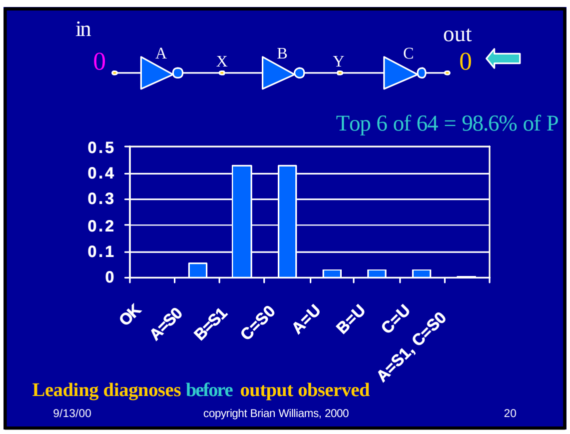
- which are eliminated?
- which predict observations?
- Which are agnostic?

Priors for Single Fault Diagnoses:

	A	B	C
P(S1)	~~.008~~	$\boxed{.008}$	~~.001~~
P(S0)	$\boxed{.001}$	~~.001~~	$\boxed{.008}$
P(U)	.001	.001	.001

9/13/00 copyright Brian Williams, 2000 18

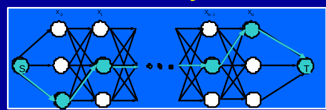




Outline

- Overview
- Belief State Update
  - Sequential, Model-based Diagnosis
  - HMMs and Belief State Update
  - Robot Localization

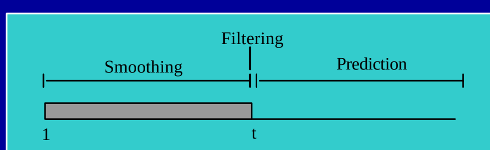
Estimation For Dynamic Systems



Given a sequence of observations and commands:

- What is the likelihood of a state?
  - ⇨ **Belief State Update**: (filtering, smoothing, prediction)
- What is the most likely sequence of states that got me here?
  - ⇨ Decoding: (Viterbi Algorithm)
- What is the most likely sequence of observations generated?
  - ⇨ Evaluation:
- What HMM most likely generated these observations?
  - ⇨ Learning: (Baum-Welch Algorithm, Expectation-Maximization Algorithm)

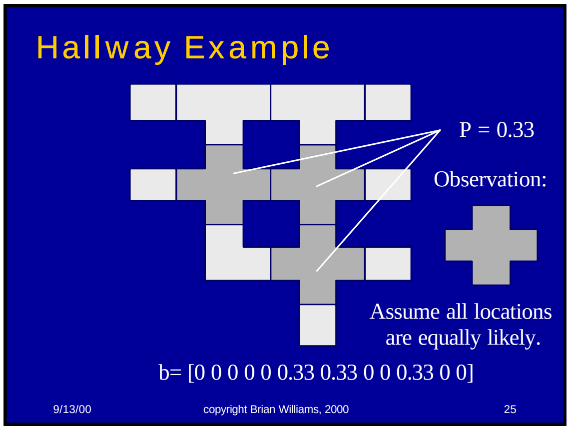
What is the likelihood of a state?

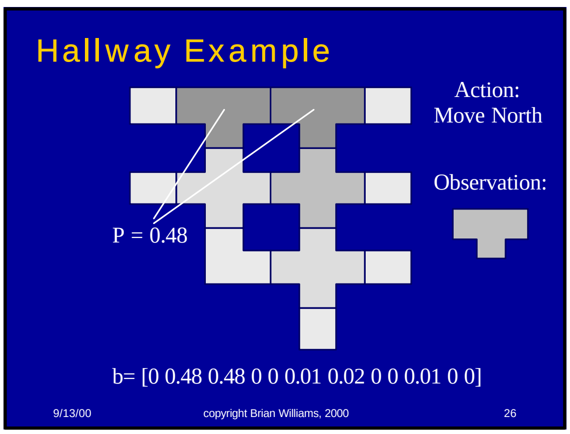


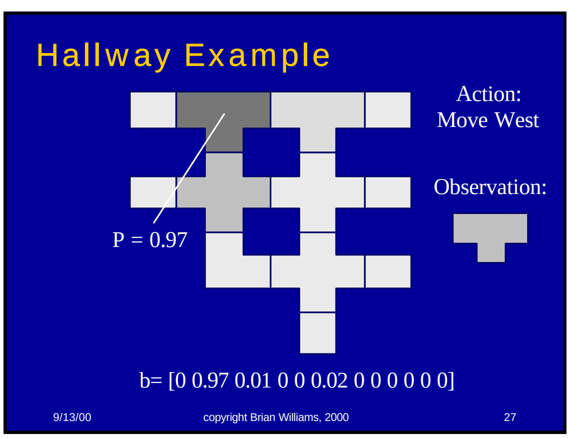
- Filtering
  - Probabilities of current states
- Prediction
  - Probabilities of future states
- Smoothing
  - Probabilities of past states

Notation

- **S^{t+1}**: set of hidden variables in the t+1 time slice
- **s^{t+1}**: set of values for those hidden variables at t+1
- **x^{t+1}: set of observations at time t+1**
- **$x^{1:t}$: set of observations from all times from 1 to t**
- **a: normalization constant**


Hallway Example
P = 0.33
Observation:
Assume all locations are equally likely.
b= [0 0 0 0 0 0.33 0.33 0 0 0.33 0 0]
9/13/00
copyright Brian Williams, 2000
25


Hallway Example
Action: Move North
Observation:
P = 0.48
b= [0 0.48 0.48 0 0 0.01 0.02 0 0 0.01 0 0]
9/13/00
copyright Brian Williams, 2000
26


Hallway Example
Action: Move West
Observation:
P = 0.97
b= [0 0.97 0.01 0 0 0.02 0 0 0 0 0 0]
9/13/00
copyright Brian Williams, 2000
27

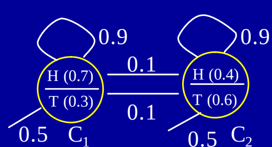
Hidden Markov Models

- Finite States S, Actions A & Observations Ω
- State transition function
  $T(S^i, A^i, S^{i+1}) \equiv P(S^{i+1} | S^i, A^i)$
- Observation function
  $O(S^i, \Omega^i) \equiv P(\Omega^i | S^i)$
- Initial state distribution
  $\Theta(S)$: $P(S^1)$

Notation:
$\Pi(S)$ denotes all subsets of S

Hidden Markov Models

- $S = \{Coin_1, Coin_2\}$
- $\Omega = \{H,T\}$
- $A = \{\}$
- T, O, Θ:



- Observed sequence:
  H, T, H, H, H, H, T, H, T, T, H, H, H, H, H
- Hidden sequence:
  $C_1, C_1, C_1, C_1, C_1, C_2, C_2, C_2, C_2, C_1, C_1, C_2, C_2, C_2, C_2$

Markov Assumptions

Given a distribution over the current state, the future states and current and future observations are independent of the past.

- **First-order Markov process**
  - $P(S^t|S^{0:t-1}) = P(S^t|S^{t-1})$
- **Markov assumption of evidence**
  - $P(X^t|S^{0:t}, X^{0:t-1}) = P(X^t|S^t)$

HMM Belief State Update

$$b(S^{i+1}) \equiv P(S^{i+1} \mid a^{1,i}\ x^{1:i+1})$$

$b(S^{i+1})$ denotes a belief distribution over S at time i+1:

$$b(S^{i+1}) = a\,P(x^{i+1} \mid S^{i+1}, a^{1:i}, x^{1:i})P(S^{i+1} \mid a^{1:i}, x^{1:i})$$

by Bayes' Rule on S^{i+1} and x^{i+1}

$$= a\,P(x^{i+1} \mid S^{i+1})P(S^{i+1} \mid a^{1:i}, x^{1:i})$$

Markov Assumption: Present and future behavior is independent of the past, given the current state.

9/13/00 31

HMM Belief State Update (cont.)

$$b(S^{i+1}) = a\,P(x^{i+1} \mid S^{i+1})P(S^{i+1} \mid a^{1:i}, x^{1:i})$$

$$= a\,P(x^{i+1} \mid S^{i+1})\sum_{s^i \in S^i} P(S^{i+1}, S^i \mid a^{1:i}, x^{1:i})$$

by marginalizing S^i

$$= a\,P(x^{i+1} \mid S^{i+1})\sum_{s^i \in S^i} P(S^{i+1} \mid S^i, a^{1:i}, x^{1:i})P(S^i \mid a^{1:i}, x^{1:i})$$

by chain rule

$$= a\,P(x^{i+1} \mid S^{i+1})\sum_{s^i \in S^i} P(S^{i+1} \mid S^i, a^i)P(S^i \mid a^{1:i-1}, x^{1:i})$$

by Markov assumption

9/13/00 32

HMM Belief State Update (cont.)

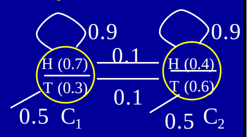
$$b(S^{i+1}) = a\,P(x^{i+1} \mid S^{i+1})\sum_{s^i \in S^i} P(S^{i+1} \mid S^i, a^i)P(S^i \mid a^{1:i-1}, x^{1:i})$$

$$\boxed{b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1})\sum_{s^i \in S^i} T(S^i, a^i, S^{i+1})b(S^i)}$$

9/13/00 33

Belief Update Example

$$b(S^{i+1}) = aO(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$



Observed sequence:

H T H H H H T H

C_1 0.5 a x 0.7 x [0.9 x 0.5 + 0.1 x 0.5] = 0.35a = 0.64

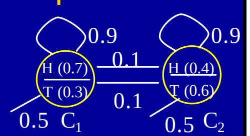
C_2 0.5 a x 0.4 x [0.1 x 0.5 + 0.9 x 0.5] = 0.20a = 0.36

$\sum_{s_i \in S} P(s_i) = 1$, hence 0.35a + 0.20a = 1 ⇨ a = 1.82

9/13/00 copyright Brian Williams, 2000 34

Belief Update Example

$$b(S^{i+1}) = aO(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$



Observed sequence:

H T H H H H T

C_1 0.5 0.64 a x 0.3 x [0.9 x 0.64 + 0.1 x 0.36] = 0.18a = 0.44

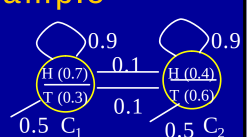
C_2 0.5 0.36 a x 0.6 x [0.1 x 0.64 + 0.9 x 0.36] = 0.23a = 0.56

$\sum_{s_i \in S} P(s_i) = 1$, hence 0.18a + 0.23a = 1 ⇨ a = 2.44

9/13/00 copyright Brian Williams, 2000 35

Belief Update Example

$$b(S^{i+1}) = aO(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$



Observed sequence:

H T H H H H T

C_1 0.5 0.64 0.44 a x 0.7 x [0.9 x 0.44 + 0.1 x 0.56] = 0.32a = 0.37

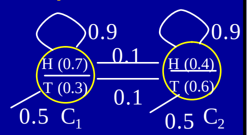
C_2 0.5 0.36 0.56 a x 0.4 x [0.1 x 0.44 + 0.9 x 0.56] = 0.55a = 0.63

$\sum_{s_i \in S} P(s_i) = 1$, hence 0.32a + 0.55a = 1 ⇨ a = 1.15

9/13/00 copyright Brian Williams, 2000 36

Belief Update Example

$$b(S^{i+1}) = \alpha O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$



Observed sequence:

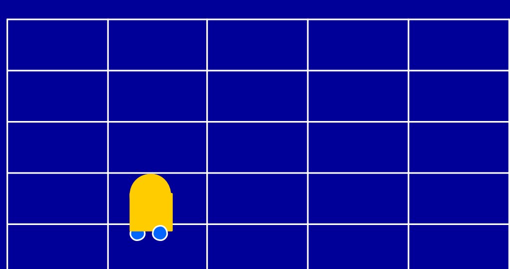
		H	T	H	H	H	H	T
C_1	0.5	0.64	0.44	0.37				
C_2	0.5	0.36	0.56	0.63				

$\sum_{s_i \in S} P(s_i) = 1$, hence 0.32a + 0.55a = 1 ⇨ a = 1.15

Outline

- Overview
- Belief State Update
  - Sequential, Model-based Diagnosis
  - HMMs and Belief State Update
  - Robot Localization

Localization Example: Where is Robbie?



Localization Example: Where is Robbie?

- Actions: North, South, East, West
- Observations: Wall, No-Wall
- Transition Function $T(S^i, A^i, S^{i+1}) \equiv P(S^{i+1} | S^i, A^i)$:

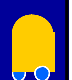
  If action pushes against wall,
  Then action succeeds with P = 0.0
  remains in place with P = 1.0

  Else action succeeds with P = 0.75
  remains in place with P = 0.25

9/13/00 40

Localization Example: Where is Robbie?

- Observation Function $O(S^i, \Omega^i) \equiv P(\Omega^i | S^i)$:



If next to wall,
Then observe Wall with P = 1.0
No-Wall with P = 0.0

If one square from wall,
Then observe Wall with P = 0.25
No-Wall with P = 0.75

Otherwise,
observe Wall with P = 0.0
observe No-Wall with P= 1.0

9/13/00 41

Example: Where is Robbie? Just Turned on, could be anywhere

4%	4	4	4	4
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4

Initial state distribution $\Theta(S) \equiv P(S^1)$,
assume uniform distribution.

9/13/00 42

Example: Where is Robbie? Moves North

4%	4	4	4	4
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4

.75 x 4 + 1 x 4 = 7
.75 x 4 + .25 x 4 = 4
.25 x 4 = 1

$$b(S^{i+1}) = aO(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

9/13/00 copyright Brian Williams, 2000 43

Example: Where is Robbie? Moves North

7%	7	7	7	7
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4
1	1	1	1	1

$$b(S^{i+1}) = aO(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

9/13/00 copyright Brian Williams, 2000 44

Example: Where is Robbie? Observes No-Wall

7%	7	7	7	7
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4
1	1	1	1	1

Column along vertical wall

0 x 7 = 0
0 x 4 = 0
.75 x 4 = 3 Interior
1 x 4 = 4 columns
0 x 1 = 0

$$b(S^{i+1}) = aO(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

9/13/00 copyright Brian Williams, 2000 45

Example: Where is Robbie? Observes No-Wall

0	0	0	0	0
0	3	3	3	0
0	3	4	3	0
0	3	3	3	0
0	0	0	0	0

Numbers not renormalize by a

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

Example: Where is Robbie? Moves North

0	0	0	0	0
0	3	3	3	0
0	3	4	3	0
0	3	3	3	0
0	0	0	0	0

Numbers not renormalize by a

.75 x 3 + 1 x 0 = 2.25
.75 x 3 + .25 x 3 = 3
.75 x 4 + .25 x 3 = 3.75
.75 x 3 + .25 x 4 = 3.25
.75 x 0 + .25 x 3 = .75

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

Example: Where is Robbie? Moves North

0	2.25	2.25	2.25	0
0	3	3.75	3	0
0	3	3.25	3	0
0	2.25	2.25	2.25	0
0	0	0	0	0

Numbers not renormalized by a

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

Example: Where is Robbie? Observes Wall

0	2.25	2.25	2.25	0
0	3	3.75	3	0
0	3	3.25	3	0
0	2.25	2.25	2.25	0
0	0	0	0	0

Column along vertical wall

Numbers not renormalized by a

1 x 2.25 = 2.25
.25 x 3 = .75
.25 x 3.75 = 0.94
0 x … = 0
.25 x 2.25 = .56

Interior columns

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1})\, b(S^i)$$

Example: Where is Robbie? Observes Wall

0	2.25	2.25	2.25	0
0	.75	.94	.75	0
0	0	0	0	0
0	.56	.56	.56	0
0	0	0	0	0

Numbers not renormalized by a

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1})\, b(S^i)$$

Example: Where is Robbie? Moves East

0	2.25	2.25	2.25	0
0	.75	.94	.75	0
0	0	0	0	0
0	.56	.56	.56	0
0	0	0	0	0

Numbers not renormalized by a

.75 x 2.25 + 1 x 0 = 1.69
.75 x 2.25 + .25 x 2.25 = 2.25
.75 x 0 + .25 x 2.25 = .56
.75 x .75 + 1 x 0 = .56
.75 x .94 + .25 x .75 = .89
.75 x .75 + .25 x .94 = .80
.75 x 0 + .25 x .75 = .19

.75 x .56 + 1 x 0 = .42
.75 x .56 + .25 x .56 = .56
.75 x 0 + .25 x .56 = .14

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1})\, b(S^i)$$

Example: Where is Robbie? Moves East

0	.56	2.25	2.25	1.69
0	.19	.80	.89	.56
0	0	0	0	0
0	.14	.56	.56	.42
0	0	0	0	0

Numbers not renormalized by a

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

9/13/00 copyright Brian Williams, 2000 52

Example: Where is Robbie? Observes Wall

0	.56	2.25	2.25	1.69
0	.05	.20	.22	.56
0	0	0	0	0
0	.03	.14	.14	.42
0	0	0	0	0

Numbers not renormalized by a

1 x P = P along wall
.25 x P = one step from wall

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

9/13/00 copyright Brian Williams, 2000 53

Example: Where is Robbie? Observes Wall

0	6.6%	26.6	26.6	19.8
0	.6	2.3	2.6	6.6
0	0	0	0	0
0	.4	1.6	1.6	4.9
0	0	0	0	0

Normalizing in terms of %

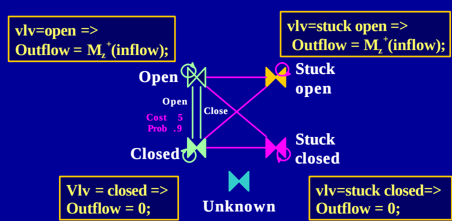
a = 11.75 %

$$b(S^{i+1}) = a\,O(x^{i+1}, S^{i+1}) \sum_{s^i \in S^i} T(S^i, a^i, S^{i+1}) b(S^i)$$

9/13/00 copyright Brian Williams, 2000 54

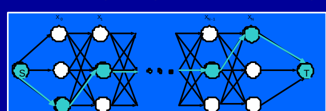
Diagnosing Dynamic Systems: Via Probabilistic Constraint Automata

- Devices modes
- Probabilistic transitions between modes
- State constraints for each mode
- One automata per component



9/13/00 copyright Brian Williams, 2000 55

Estimating Dynamic Systems



Given a sequence of observations and commands:

- What is the likelihood of a particular state?
  - ⇨ **Belief State Update**: (filtering, smoothing, prediction)
- What is the most likely sequence of states that got me here?
  - ⇨ **Decoding**: (Viterbi Algorithm)
- What is the most likely sequence of observations generated?
  - ⇨ **Evaluation**:
- What HMM most likely generated these observations?
  - ⇨ **Learning**: (Baum-Welch Algorithm, Expectation-Maximization Algorithm)

9/13/00 copyright Brian Williams, 2000 56