Reasoning Under Uncertainty Over Time

Alice Gao

Lecture 14

Readings: RN 15.1, 15.2.1, 15.2.2. PM 8.5.1 - 8.5.3.

Outline

Learning Goals

By the end of the lecture, you should be able to

- Construct a hidden Markov model given a real-world scenario.
- Perform filtering and smoothing by executing the forward-backward algorithm.

Inference in a Changing World

So far, we can reason probabilistically in a static world.
However, the world evolves over time.

In an evolving world, we have to reason about a sequence of events.

Applications of sequential belief networks:

- weather predictions
- stock market predictions
- patient monitoring
- robot localization
- speech and handwriting recognition

The Umbrella World

You are a security guard stationed at a secret underground installation.

You want to know whether it's raining today.

Unfortunately, your only access to the outside world occurs each morning when you see the director coming in with, or without, an umbrella.

States and Observations

- The world contains a series of time slices.
- Each time slice contains a set of random variables, some observable, some not.
  $\mathbf{X}_t$ the un-observable variables at time t
  $\mathbf{E}_t$ the observable variables at time t

What are the observable and unobservable random variables in the umbrella world?

X_t : $R_1, R_2, \ldots$ whether it rains.

E_t : $U_1, U_2, \ldots$ whether the director carries an umbrella.

The transition model

How does the current state depend on the previous states?

In general, every state may depend on all the previous states.

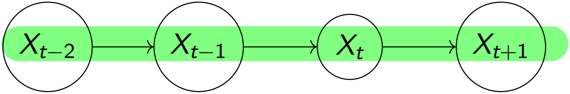
$$P(\mathbf{X}_t | \mathbf{X}_{t-1} \land \mathbf{X}_{t-2} \land \mathbf{X}_{t-3} \land \cdots \land \mathbf{X}_1)$$

Problem:
As t increases, the conditional probability distribution can be unboundedly large.

Solution:
Let the current state depend on a fixed number of previous states.

K-order Markov chain

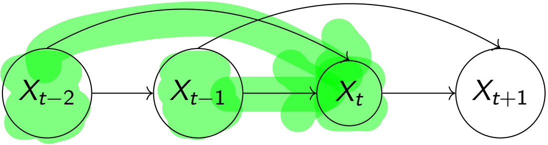
First-order Markov process:



The transition model:

$$P(\mathbf{X}_t | \mathbf{X}_{t-1} \wedge \mathbf{X}_{t-2} \wedge \mathbf{X}_{t-3} \wedge \cdots \wedge \mathbf{X}_1) = P(\mathbf{X}_t | \mathbf{X}_{t-1})$$

Second-order Markov process:



The transition model:

$$P(\mathbf{X}_t | \mathbf{X}_{t-1} \wedge \mathbf{X}_{t-2} \wedge \mathbf{X}_{t-3} \wedge \cdots \wedge \mathbf{X}_1) = P(\mathbf{X}_t | \mathbf{X}_{t-1} \wedge \mathbf{X}_{t-2})$$

Transition model for the umbrella world

The Markov assumption:

The future is independent of the past given the present.

$$\rightarrow R_{t-2} \rightarrow R_{t-1} \rightarrow R_t \rightarrow R_{t+1} \rightarrow$$

$P(R_{t-2}|R_{t-3})$ $P(R_{t-1}|R_{t-2})$ $P(R_t|R_{t-1})$ $P(R_{t+1}|R_t)$

$P(R_{t-2}|\neg R_{t-3})$ $P(R_{t-1}|\neg R_{t-2})$ $P(R_t|\neg R_{t-1})$ $P(R_{t+1}|\neg R_t)$

The transition model:

$$P(R_t|R_{t-1} \wedge R_{t-2} \wedge R_{t-3} \wedge \cdots \wedge R_1) = P(R_t|R_{t-1})$$

Stationary Process

Is there a different conditional probability distribution for each time step?

Stationary process:

- The dynamics does not change over time.
- The conditional probability distribution for each time step remains the same.

What are the advantages of using a stationary model?

Transition model for the umbrella world

$$\boxed{P(R_1) = 0.5}$$

$$\boxed{\begin{array}{l} P(R_t | R_{t-1}) = 0.7 \\ P(R_t | \neg R_{t-1}) = 0.3 \end{array}}$$

$$\rightarrow R_{t-2} \rightarrow R_{t-1} \rightarrow R_t \rightarrow R_{t+1} \rightarrow$$

Sensor model

How does the evidence variable $\mathbf{E}_t$ at time t depend on the previous and current states?

Sensor Markov assumption:
Each state is sufficient to generate its observations.

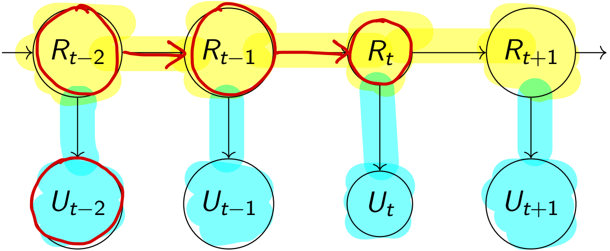
$$P(\mathbf{E}_t | \mathbf{X}_t \wedge \mathbf{X}_{t-1} \wedge \cdots \wedge \mathbf{X}_1 \wedge \mathbf{E}_{t-1} \wedge \mathbf{E}_{t-2} \wedge \cdots \wedge \mathbf{E}_1)$$
$$= P(\mathbf{E}_t | \mathbf{X}_t)$$

Complete model for the umbrella world

$$P(R_1) = 0.5$$

$$P(R_t | R_{t-1}) = 0.7$$
$$P(R_t | \neg R_{t-1}) = 0.3$$

$$P(U_t | R_t) = 0.9$$
$$P(U_t | \neg R_t) = 0.2$$



Hidden Markov Model

- A Markov process
- The state variables are unobservable.
- The evidence variables, which depend on the states, are observable.

Common Inference Tasks

- **Filtering:** Which state am I in right now?
  (today) $P(R_{10} | U_{1:10})$ $P(R_t | U_{1:t})$
- **Prediction:** Which state will I be in tomorrow?
  (future) $P(R_{15} | U_{1:10})$ $P(R_k | U_{1:t})$ $k > t$.
- **Smoothing:** Which state was I in yesterday?
  (past) $P(R_5 | U_{1:10})$ $P(R_k | U_{1:t})$ $1 \leq k < t$
- **Most likely explanation:** Which sequence of states is most likely to have generated the observations?

Algorithms for the inference tasks

a HMM is a Bayesian network.
We can perform inference using the variable elimination algorithm!

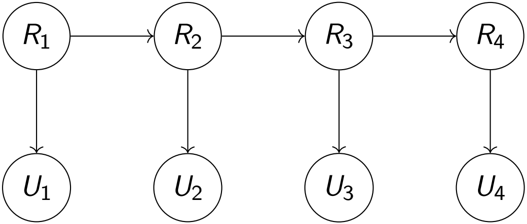
More specialized algorithms:

- The forward-backward algorithm: filtering and smoothing
- The Viterbi algorithm: most likely explanation

The umbrella world

$$\boxed{P(R_1) = 0.5}$$

$$\boxed{\begin{array}{l} P(R_t|R_{t-1}) = 0.7 \\ P(R_t|\neg R_{t-1}) = 0.3 \end{array}} \quad \boxed{\begin{array}{l} P(U_t|R_t) = 0.9 \\ P(U_t|\neg R_t) = 0.2 \end{array}}$$



Learning Goals

Filtering

Smoothing

Revisiting the Learning goals

Filtering

Given the observations up to today, which state am I in today?

Day 1: $P(R_1|u_1)$
Day 2: $P(R_2|u_{1:2})$
Day 3: $P(R_3|u_{1:3})$
...
Day t: $P(R_t|u_{1:t})$

R_t : we do not observe whether it rains (or not.)

$u_{1:t} = u_1 \wedge u_2 \wedge \cdots \wedge u_t$
we observe the value of u_i on day i.

Filtering (day 2)

How do we calculate $P(R_2|u_{1:2})$?

Filtering (day 2)

$$
\begin{aligned}
& P(R_2|u_{1:2}) \\
& = \alpha P(u_2|R_2 \wedge u_1) P(R_2|u_1) & (1) \\
& = \alpha P(u_2|R_2) P(R_2|u_1) & (2) \\
& = \alpha P(u_2|R_2) \sum_{r_1} P(R_2 \wedge r_1|u_1) & (3) \\
& = \alpha P(u_2|R_2) \sum_{r_1} P(R_2|r_1 \wedge u_1) P(r_1|u_1) & (4) \\
& = \alpha P(u_2|R_2) \sum_{r_1} P(R_2|r_1) P(r_1|u_1) & (5)
\end{aligned}
$$

Filtering (day 1)

CQ: Filtering (day 2) — step 1

CQ: Consider the first step of the derivation for filtering (day 2). What is the correct justification for this step?

$$P(R_2|u_{1:2})$$
$$= \alpha P(u_2|R_2 \wedge u_1)P(R_2|u_1)$$

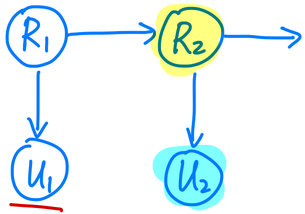
(A) The sum rule

(B) The chain rule

(C) Bayes' rule

(D) Markov assumption

(E) sensor Markov assumption



Prove that $P(R_2 \mid U_{1:2}) = \alpha P(U_2 \mid R_2 \wedge U_1) P(R_2 \mid U_1)$

Proof:

$$
\begin{aligned}
P(R_2 \mid U_{1:2}) &= P(R_2 \mid U_2 \wedge U_1) \\
&= \frac{P(R_2 \wedge U_2 \wedge U_1)}{P(U_2 \wedge U_1)} \\
&= \frac{P(U_2 \wedge R_2 \wedge U_1)}{P(U_2 \wedge U_1)} \\
&= \frac{P(U_2 \mid R_2 \wedge U_1) P(R_2 \mid U_1) \cancel{P(U_1)}}{P(U_2 \mid U_1) \cancel{P(U_1)}} \\
&= \frac{P(U_2 \mid R_2 \wedge U_1) P(R_2 \mid U_1)}{P(U_2 \mid U_1)} \\
&= \alpha P(U_2 \mid R_2 \wedge U_1) P(R_2 \mid U_1)
\end{aligned}
$$

QED

CQ: Filtering (day 2) — step 2

CQ: Consider the second step of the derivation for filtering (day 2). What is the correct justification for this step?

$$
\begin{aligned}
&= \alpha P(u_2 | \underline{R_2} \wedge u_1) P(R_2 | u_1) \\
&= \alpha P(u_2 | R_2) P(R_2 | u_1)
\end{aligned}
$$

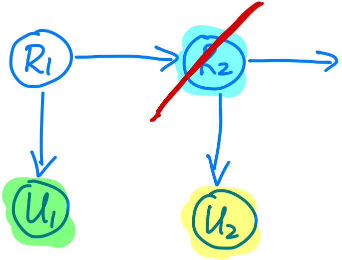
(A) The sum rule

(B) The chain rule

(C) Bayes' rule

(D) Markov assumption

(E) sensor Markov assumption

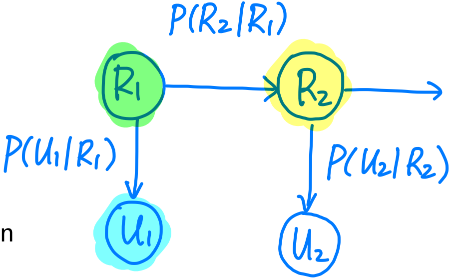


CQ: Filtering (day 2) — step 3

CQ: Consider the third step of the derivation for filtering (day 2). What is the correct justification for this step?

$$= \alpha P(u_2|R_2)P(R_2|u_1)$$
$$= \alpha P(u_2|R_2)\sum_{r_1} P(R_2 \wedge r_1|u_1)$$

(A) The sum rule
(B) The chain rule
(C) Bayes' rule
(D) Markov assumption
(E) sensor Markov assumption



CQ: Filtering (day 2) — step 4

CQ: Consider the fourth step of the derivation for filtering (day 2). What is the correct justification for this step?

$$= \alpha P(u_2|R_2) \sum_{r_1} P(R_2 \wedge r_1|\cancel{u_1})$$

$$= \alpha P(u_2|R_2) \sum_{r_1} P(R_2|r_1 \wedge \cancel{u_1}) P(r_1|\cancel{u_1})$$

(A) The sum rule

(B) The chain rule

(C) Bayes' rule

(D) Markov assumption

(E) sensor Markov assumption

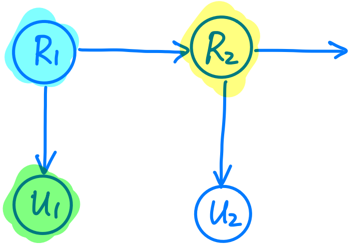
CQ: Filtering (day 2) — step 5

CQ: Consider the fifth step of the derivation for filtering (day 2). What is the correct justification for this step?

$$= \alpha P(u_2|R_2) \sum_{r_1} P(R_2|r_1 \wedge u_1) P(r_1|u_1)$$

$$= \alpha P(u_2|R_2) \sum_{r_1} P(R_2|r_1) P(r_1|u_1)$$

(A) The sum rule
(B) The chain rule
(C) Bayes' rule
(D) Markov assumption
(E) sensor Markov assumption



Filtering (day 2)

How do we calculate $P(R_2|u_{1:2})$?

$$P(R_2|u_{1:2}) = \alpha P(u_2|R_2) \sum_{r_1} P(R_2|r_1) P(r_1|u_1)$$

Where do we obtain the required probabilities?

- $P(u_2|R_2)$ sensor model
- $P(R_2|r_1)$ transition model
- $P(r_1|u_1)$ result of the filtering task on day 1.

CQ: Filtering (day 2)

CQ: For the umbrella world, what is the probability that it rains on day 2 given that the director brought an umbrella on both days 1 and 2?

That is, calculate $P(R_2 = true | u_1 = true \wedge u_2 = true)$.

$$P(R_2 | u_{1:2}) = \alpha P(u_2 | R_2) \sum_{r_1} P(R_2 | r_1) P(r_1 | u_1)$$

(A) 0.583
(B) 0.683
(C) 0.783
(D) 0.883
(E) 0.983

$P(R_1) = 0.5$

$P(U_1 = t | R_1 = t) = 0.9$

$P(U_1 = t | R_1 = f) = 0.2$

$P(R_2 = t | R_1 = t) = 0.7$

$P(R_2 = t | R_1 = f) = 0.3$

$$P(R_2|u_{1:2}) = \alpha P(u_2|R_2) \underbrace{\sum_{r_1} P(R_2|r_1) \underbrace{P(r_1|u_1)}_{①}}_{②}$$

$P(R_1) = 0.5$

$P(U_1=t|R_1=t) = 0.9$

$P(U_1=t|R_1=f) = 0.2$

$P(R_2=t|R_1=t) = 0.7$

$P(R_2=t|R_1=f) = 0.3$

① $P(R_1|u_1) = \alpha P(u_1|R_1) P(R_1)$ Bayes' rule

$$= \alpha \langle 0.9, 0.2 \rangle \langle 0.5, 0.5 \rangle$$

$$= \alpha \langle 0.45, 0.1 \rangle$$

$$\triangleq \langle 0.818, 0.182 \rangle$$

② $$\sum_{r_1} P(R_2|r_1) P(r_1|u_1) = \overbrace{\langle 0.7, 0.3 \rangle \times 0.818}^{r_1=t} + \overbrace{\langle 0.3, 0.7 \rangle \times 0.182}^{r_1=f}$$

$$\triangleq \langle 0.627, 0.373 \rangle$$

③ $$P(R_2|u_{1:2}) = \alpha P(u_2|R_2) \sum_{r_1} P(R_2|r_1) P(r_1|u_1)$$

$$= \alpha \langle 0.9, 0.2 \rangle \langle 0.627, 0.373 \rangle$$

$$= \alpha \langle 0.565, 0.075 \rangle \triangleq \langle 0.883, 0.117 \rangle$$

Filtering (day 3)

How do we calculate $P(R_3|u_{1:3})$?

Filtering (day 3)

$$
\begin{aligned}
& P(R_3|u_{1:3}) \\
&= \alpha P(u_3|R_3 \wedge u_{1:2}) P(R_3|u_{1:2}) && \text{Bayes' rule} && (6) \\
&= \alpha P(u_3|R_3) P(R_3|u_{1:2}) && \text{sensor Markov assumption} && (7) \\
&= \alpha P(u_3|R_3) \sum_{r_2} P(R_3 \wedge r_2|u_{1:2}) && \text{sum rule} && (8) \\
&= \alpha P(u_3|R_3) \sum_{r_2} P(R_3|r_2 \wedge u_{1:2}) P(r_2|u_{1:2}) && \text{chain rule} && (9) \\
&= \alpha \underbrace{P(u_3|R_3)}_{\text{sensor model}} \sum_{r_2} \underbrace{P(R_3|r_2)}_{\text{transition probability}} \underbrace{P(r_2|u_{1:2})}_{\text{Filtering (day 2)}} && \text{Markov assumption} && (10)
\end{aligned}
$$

Filtering (day t)

How do we calculate $P(R_t|u_{1:t})$?

Filtering (day t)

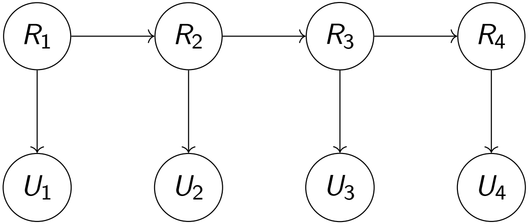
$$
\begin{aligned}
& P(R_t|u_{1:t}) \\
&= \alpha P(u_t|R_t \wedge u_{1:(t-1)}) P(R_t|u_{1:(t-1)}) && (11) \\
&= \alpha P(u_t|R_t) P(R_t|u_{1:(t-1)}) && (12) \\
&= \alpha P(u_t|R_t) \sum_{r_{t-1}} P(R_t \wedge r_{t-1}|u_{1:(t-1)}) && (13) \\
&= \alpha P(u_t|R_t) \sum_{r_{t-1}} P(R_t|r_{t-1} \wedge u_{1:(t-1)}) P(r_{t-1}|u_{1:(t-1)}) && (14) \\
&= \alpha P(u_t|R_t) \sum_{r_{t-1}} P(R_t|r_{t-1}) \underbrace{P(r_{t-1}|u_{1:(t-1)})}_{\text{Filtering (day t-1)}} && (15)
\end{aligned}
$$

Filtering through forward recursion

message $f_{1:k}$

$$P(R_t|u_{1:t}) = \alpha P(u_t|R_t) \sum_{r_{t-1}} P(R_t|r_{t-1}) P(r_{t-1}|u_{1:(t-1)})$$

$$\begin{array}{ccccccc} f_{1:1} & \longrightarrow & f_{1:2} & \longrightarrow & f_{1:3} & \longrightarrow & f_{1:4} \\ \| & & \| & & \| & & \| \\ P(R_1|U_1) & \longrightarrow & P(R_2|U_{1:2}) & \longrightarrow & P(R_3|U_{1:3}) & \longrightarrow & P(R_4|U_{1:4}) \end{array}$$



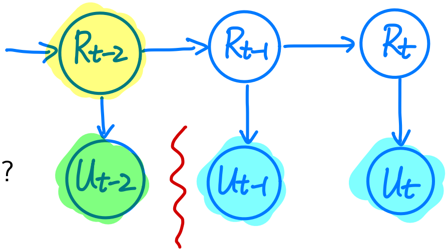
Smoothing

Given the observations up to today, which state was I in yesterday?

Day 1: $P(R_1|u_{1:t})$
Day 2: $P(R_2|u_{1:t})$
Day 3: $P(R_3|u_{1:t})$
...
Day t-1: $P(R_{t-1}|u_{1:t})$
Day t: $P(R_t|u_{1:t})$

Unsurprisingly, we will use another recursive procedure...

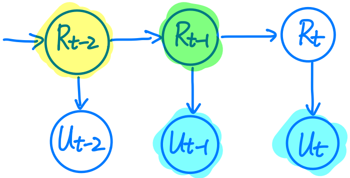
Smoothing (day t-2)



How do we calculate $P(R_{t-2}|u_{1:t})$?

$$
\begin{aligned}
&P(R_{t-2}|u_{1:t}) \\
&= P(R_{t-2}|u_{1:t-2} \wedge u_{t-1:t}) && (16)\\
&= P(R_{t-2}|u_{t-1:t} \wedge u_{1:t-2}) && (17)\\
&= \alpha P(R_{t-2}|u_{1:t-2}) P(u_{t-1:t}|R_{t-2} \wedge u_{1:t-2}) \quad \text{Bayes' rule} && (18)\\
&= \alpha \underbrace{P(R_{t-2}|u_{1:t-2})}_{f_{1:t-2}} \underbrace{P(u_{t-1:t}|R_{t-2})}_{b_{t-1:t}} \quad \text{conditional independence} && (19)
\end{aligned}
$$

Smoothing (day t-2) continued



How do we calculate $P(R_{t-2}|u_{1:t})$?

$b_{t-1:t}$

$$P(u_{t-1:t}|R_{t-2})$$

$$= \sum_{r_{t-1}} P(u_{t-1:t} \wedge r_{t-1}|R_{t-2}) \quad \text{sum rule} \tag{20}$$

$$= \sum_{r_{t-1}} P(u_{t-1:t}|r_{t-1} \wedge R_{t-2}) P(r_{t-1}|R_{t-2}) \quad \text{chain rule} \tag{21}$$

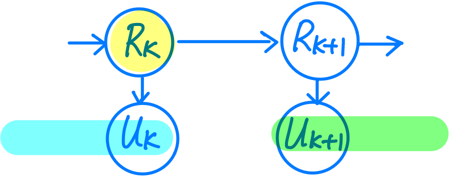
$$= \sum_{r_{t-1}} P(u_{t-1:t}|r_{t-1}) P(r_{t-1}|R_{t-2}) \quad \text{conditional independence} \tag{22}$$

$$= \sum_{r_{t-1}} P(u_{t-1} \wedge u_t|r_{t-1}) P(r_{t-1}|R_{t-2}) \quad \text{rewrite} \tag{23}$$

$P(U_{t:t}|r_{t-1})$

$$= \sum_{r_{t-1}} \underbrace{P(u_{t-1}|r_{t-1})}_{\text{sensor model}} \underbrace{P(u_t|r_{t-1})}_{b_{t:t}} \underbrace{P(r_{t-1}|R_{t-2})}_{\text{transition model}} \quad \text{cond. independence} \tag{24}$$

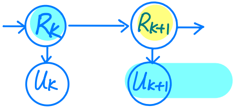
Smoothing (day k)



How do we calculate $P(R_k|u_{1:t})$ where $1 \leq k < t$?

$$
\begin{aligned}
& P(R_k|u_{1:t}) \\
&= P(R_k|u_{1:k} \wedge u_{k+1:t}) && \text{rewrite} \\
&= \alpha P(R_k|u_{1:k}) P(u_{k+1:t}|R_k \wedge u_{1:k}) && \text{Bayes' rule} \quad (25) \\
&= \alpha \underbrace{P(R_k|u_{1:k})}_{f_{1:k}} \underbrace{P(u_{k+1:t}|R_k)}_{b_{k+1:t}} && \text{conditional independence} \quad (26)
\end{aligned}
$$

Smoothing (day k) continued



How do we calculate $P(R_k|u_{1:t})$ where $1 \leq k < t$?

$$\overbrace{P(u_{k+1:t}|R_k)}^{b_{k+1:t}} \tag{27}$$

$$= \sum_{r_{k+1}} P(u_{k+1:t} \wedge r_{k+1}|R_k) \quad \text{sum rule} \tag{28}$$

$$= \sum_{r_{k+1}} P(u_{k+1:t}|r_{k+1} \wedge R_k)P(r_{k+1}|R_k) \quad \text{chain rule} \tag{29}$$

$$= \sum_{r_{k+1}} P(u_{k+1:t}|r_{k+1})P(r_{k+1}|R_k) \quad \text{conditional independence} \tag{30}$$

$$= \sum_{r_{k+1}} P(u_{k+1} \wedge u_{k+2:t}|r_{k+1})P(r_{k+1}|R_k) \quad \text{rewrite} \tag{31}$$

$$= \sum_{r_{k+1}} \underbrace{P(u_{k+1}|r_{k+1})}_{\text{sensor model}}\underbrace{P(u_{k+2:t}|r_{k+1})}_{b_{k+2:t}}\underbrace{P(r_{k+1}|R_k)}_{\text{transition model}} \quad \text{cond. independence} \tag{32}$$

Base cases for forward and backward recursion.

forward recursion

$P(R_1)$

$P(U_1 | R_1)$

$$\underbrace{P(R_1 | U_1)}_{\text{Bayes' rule}} \qquad P(R_2 | U_{1:2}) \qquad P(R_3 | U_{1:3}) \qquad P(R_4 | U_{1:4})$$

$$f_{1:1} \longrightarrow f_{1:2} \longrightarrow f_{1:3} \longrightarrow f_{1:4}$$

backward recursion (for day k, we need $P(\overbrace{U_{k+1:t} | R_k}^{b_{k+1:t}})$)

$$k = t-3 \qquad k = t-2 \qquad k = t-1 \qquad k = t$$

$$b_{t-2:t} \longleftarrow b_{t-1:t} \longleftarrow b_{t:t} \longleftarrow \underbrace{b_{t+1:t}}_{= \vec{1}}$$

Smoothing through backward recursion

$f_{1:k}$ $\quad$ $b_{k+1:t}$

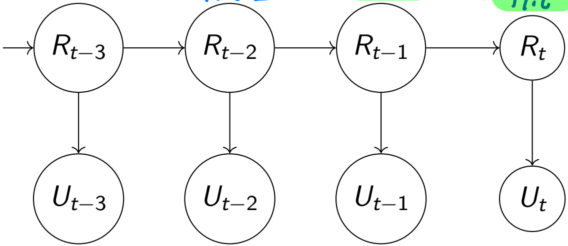
$$P(R_k|u_{1:t}) = \alpha P(R_k|u_{1:k})P(u_{k+1:t}|R_k)$$

$b_{k+1:t}$ $\quad$ $b_{k+2:t}$

$$P(u_{k+1:t}|R_k) = \sum_{r_{k+1}} P(u_{k+1}|r_{k+1})P(u_{k+2:t}|r_{k+1})P(r_{k+1}|R_k)$$

$\leftarrow b_{t-2:t} \leftarrow b_{t-1:t} \leftarrow b_{t:t} \leftarrow b_{t+1:t} = \vec{1}$

$\rightarrow f_{1:t-3} \rightarrow f_{1:t-2} \rightarrow f_{1:t-1} \rightarrow f_{1:t}$



Revisiting the Learning Goals

By the end of the lecture, you should be able to

- Construct a hidden Markov model given a real-world scenario.
- Perform filtering and smoothing by executing the forward-backward algorithm.