GHENT UNIVERSITY

FACULTY OF ENGINEERING AND ARCHITECTURE

E016350: Artificial Intelligence

Lecture 15

Reasoning under Uncertainty & Bayesian ML

Temporal probability models
Part 2
Viterbi Algorithm

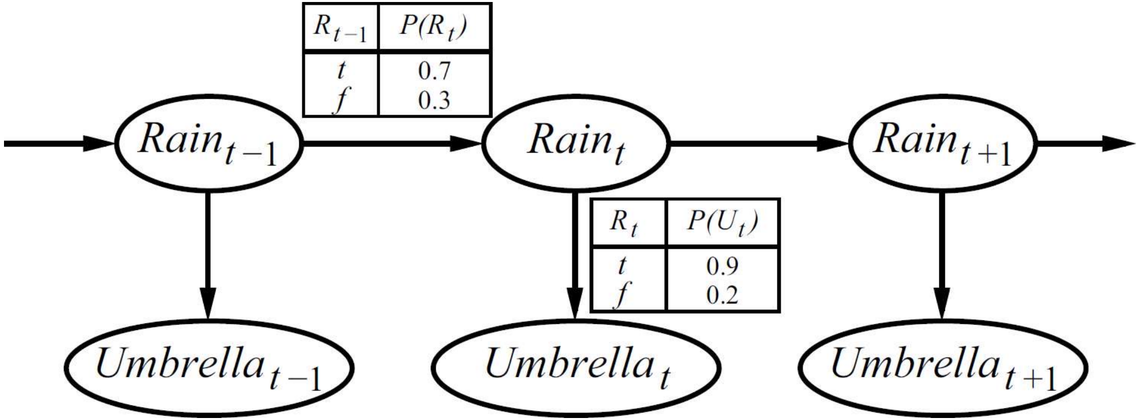
Aleksandra Pizurica

Ghent University

Fall 2024

Umbrella example

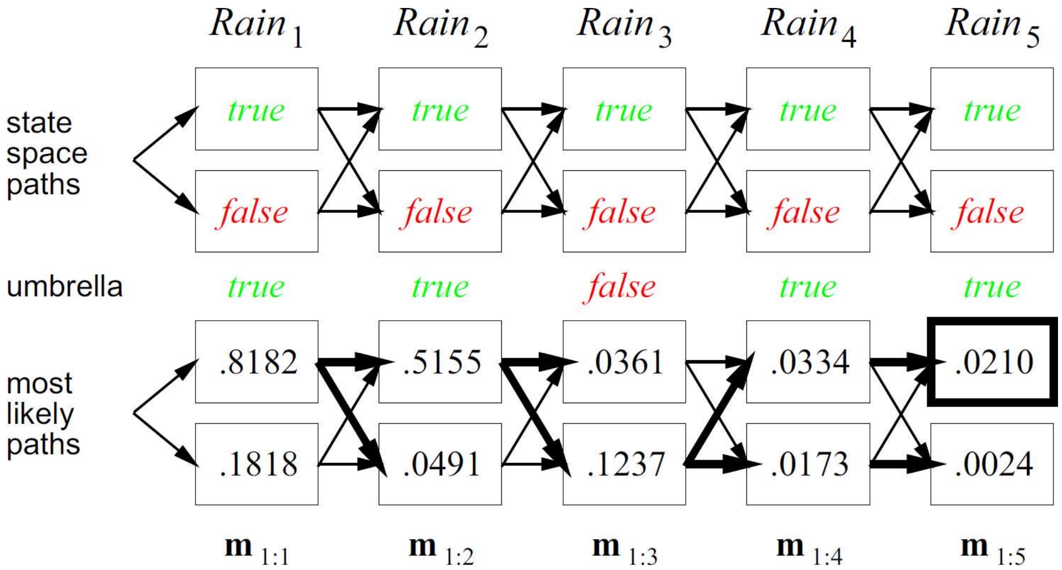
Umbrella Example



The Viterbi algorithm finds the most likely sequence of states for a given sequence of evidence variables.

Now we explain the algorithm in detail on the “umbrella” example given above and taken from [R&N, ch 14].

Umbrella example: Viterbi solution



The solution of the Viterbi algorithm for the "umbrella" example as was shown in the slides on temporal probability models, Part 1.

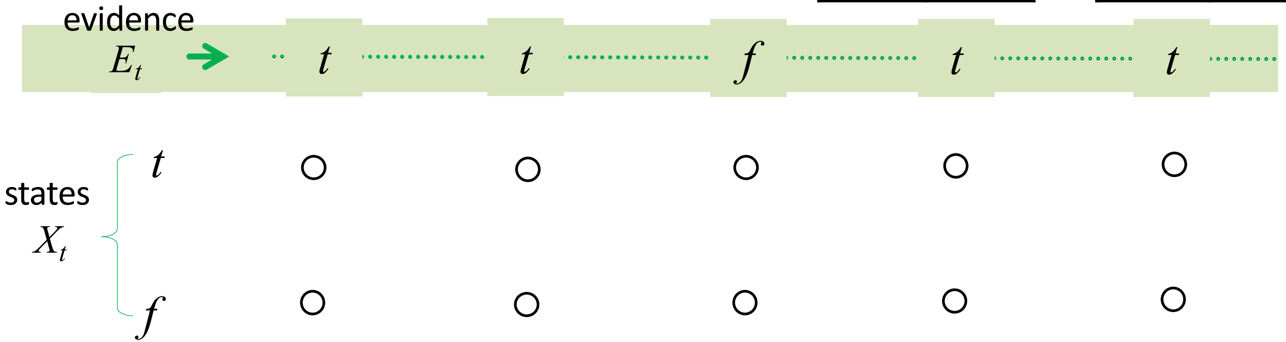
In the following slides, a detailed explanation follows.

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



We use here the following notation for the "umbrella" example from [R&N, ch 14]:

$X_t\!=\!t$: it rains on day t ($R_t\!=\!true$); $X_t\!=\!f$: no rain on day t ($R_t\!=\!false$);

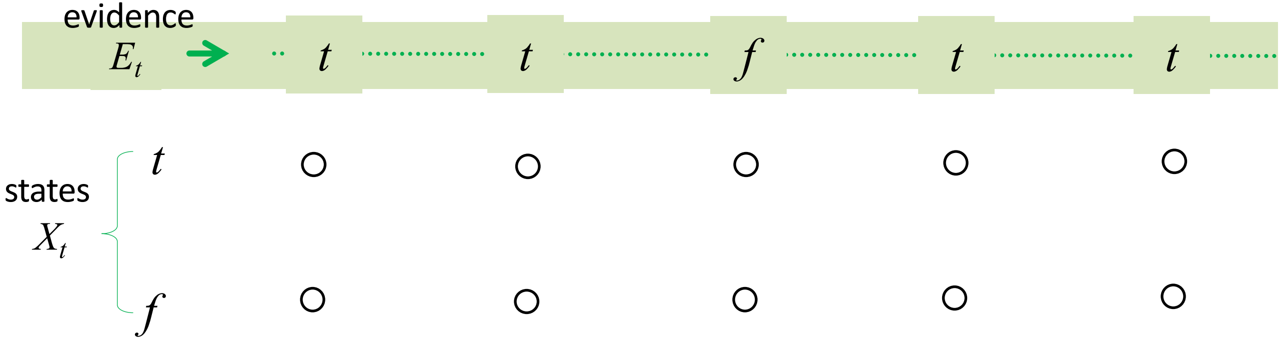
$E_t\!=\!t$: umbrella on day t ($U_t\!=\!true$); $E_t\!=\!f$: no umbrella on day t ($U_t\!=\!false$);

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



We use here the following notation for the "umbrella" example from [R&N, ch 14]:

X_t=t : it rains on day t (R_t=*true*); X_t=f : no rain on day t (R_t=*false*);

E_t=t : umbrella on day t (U_t=*true*); E_t=f : no umbrella on day t (U_t=*false*);

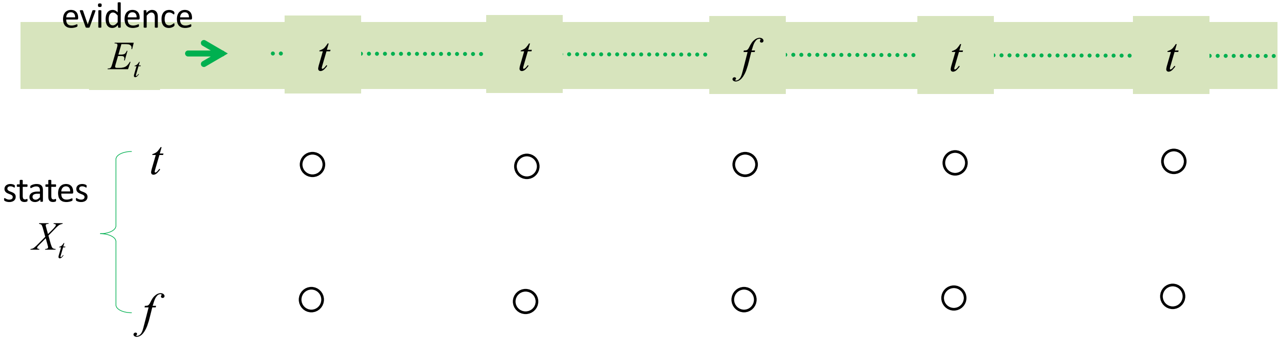
Task: given the evidence sequence $\mathbf{e}=\{e_1,\ldots,e_6\}$, find the most likely $\mathbf{x}=\{x_1,\ldots,x_6\}$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



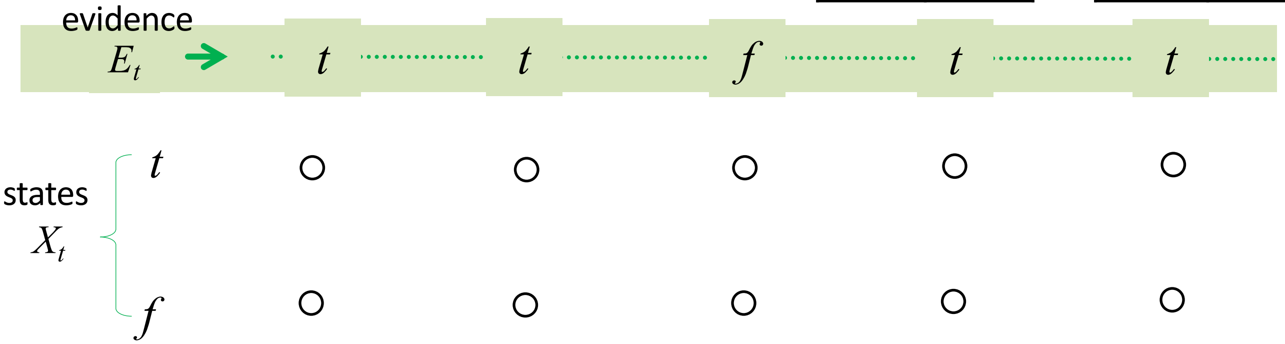
Main observation: **there is a recursive relationship between the most likely path to each state x_{t+1} and most likely paths to each previous state x_t**

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



Main observation: **there is a recursive relationship between the most likely path to each state x_{t+1} and most likely paths to each previous state x_t**

$$\max_{x_1,\dots,x_t} \mathbf{P}(x_1,\dots,x_t,X_{t+1} \mid \mathbf{e}_{1:t+1})$$

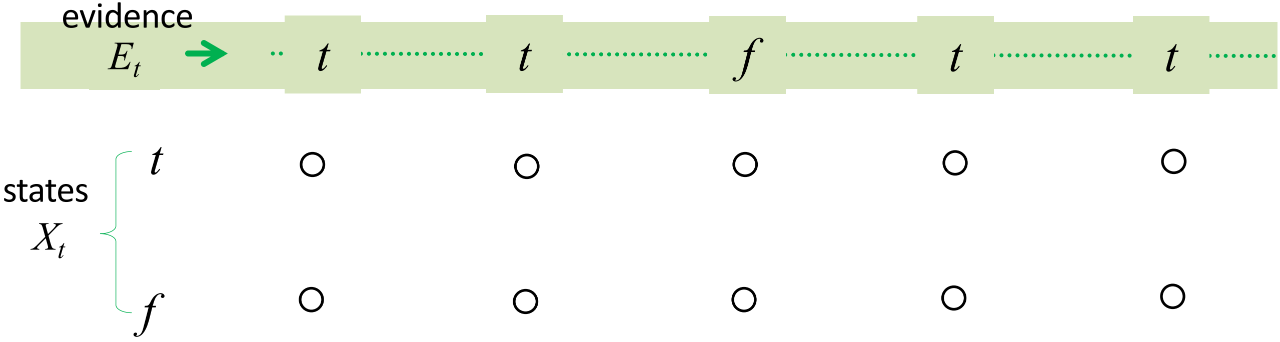
$$= \alpha \mathbf{P}(e_{t+1} \mid X_{t+1}) \max_{x_t} \left(\mathbf{P}(X_{t+1} \mid x_t) \max_{x_1,\dots,x_{t-1}} P(x_1,\dots,x_{t-1},x_t \mid \mathbf{e}_{1:t}) \right)$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



Define **message** as

$$\mathbf{m}_{1:t} = \max_{x_1,\ldots,x_{t-1}} \mathbf{P}(x_1,\ldots,x_{t-1},X_t \mid \mathbf{e}_{1:t}) = \begin{bmatrix} M^1_{1:t} \\ M^2_{1:t} \end{bmatrix}$$

$M^1_{1:t}$ for $X_t = t$

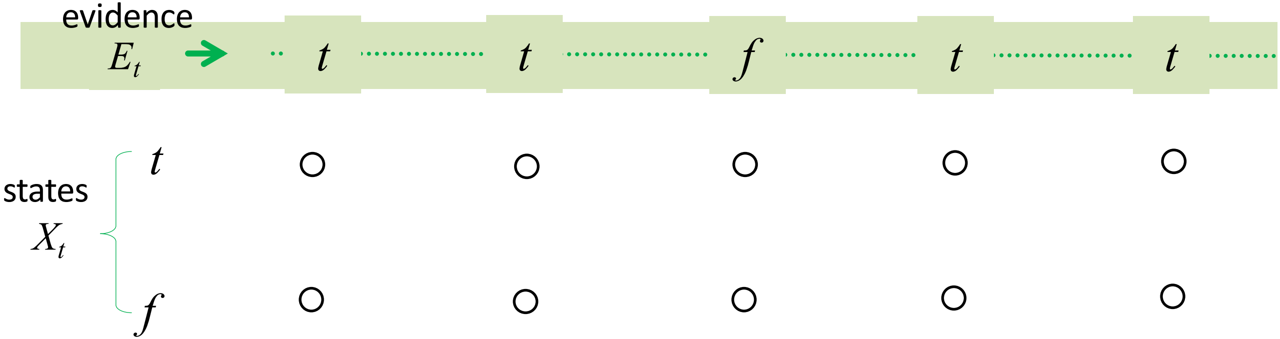
$M^2_{1:t}$ for $X_t = f$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



With this notation, for our example the recursion becomes:

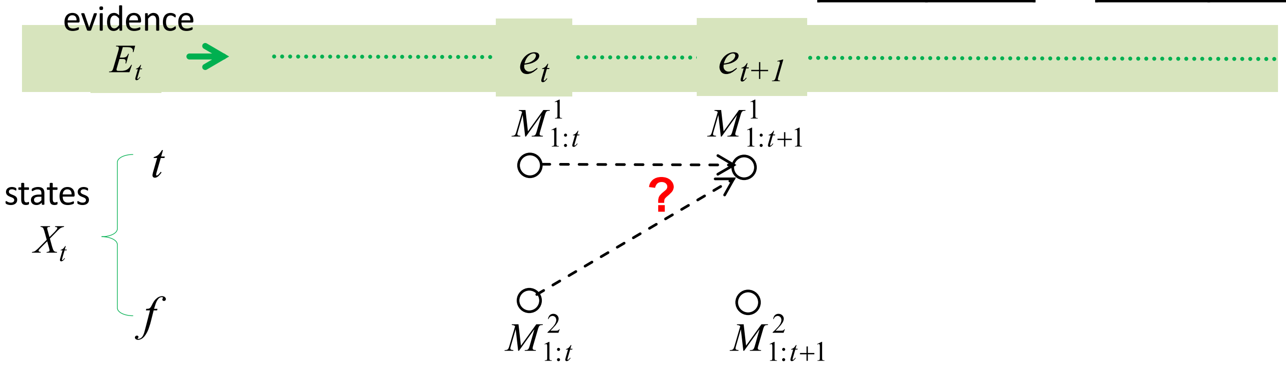
$$\mathbf{m}_{1:t+1} = \alpha \mathbf{P}(e_{t+1} \mid X_{t+1}) \max\left(\mathbf{P}(X_{t+1} \mid X_t = t) M^1_{1:t},\ \mathbf{P}(X_{t+1} \mid X_t = f) M^2_{1:t}\right)$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



So, for time instant t+1, we need to compute

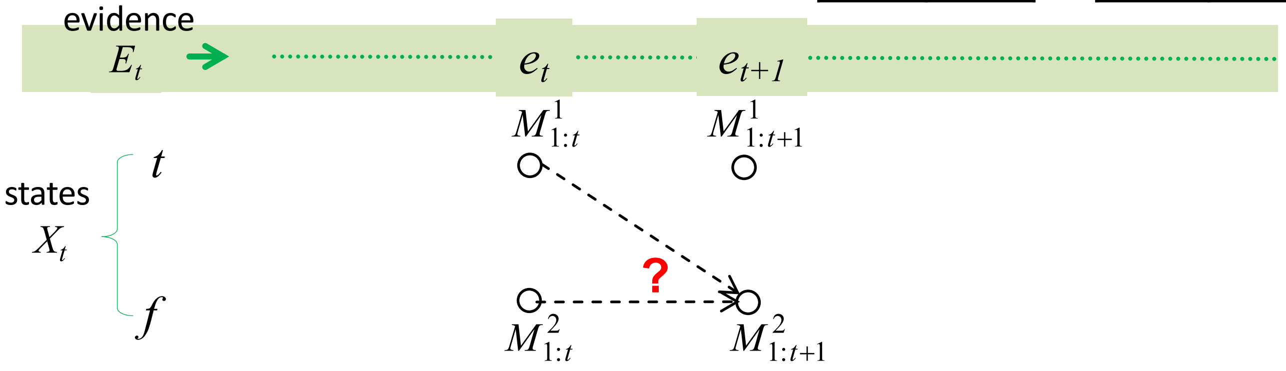
$$M^1_{1:t+1} = P(e_{t+1} \mid X_{t+1} = t)\max\left\{ \underbrace{P(X_{t+1} = t \mid X_t = t)M^1_{1:t}}_{\text{for the path from } X_t=t},\ \underbrace{P(X_{t+1} = t \mid X_t = f)M^2_{1:t}}_{\text{for the path from } X_t=f} \right\}$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



and, equivalently, the second component

$$M^2_{1:t+1} = P(e_{t+1} \mid X_{t+1} = f)\max\{P(X_{t+1} = f \mid X_t = t)M^1_{1:t}, P(X_{t+1} = f \mid X_t = f)M^2_{1:t}\}$$

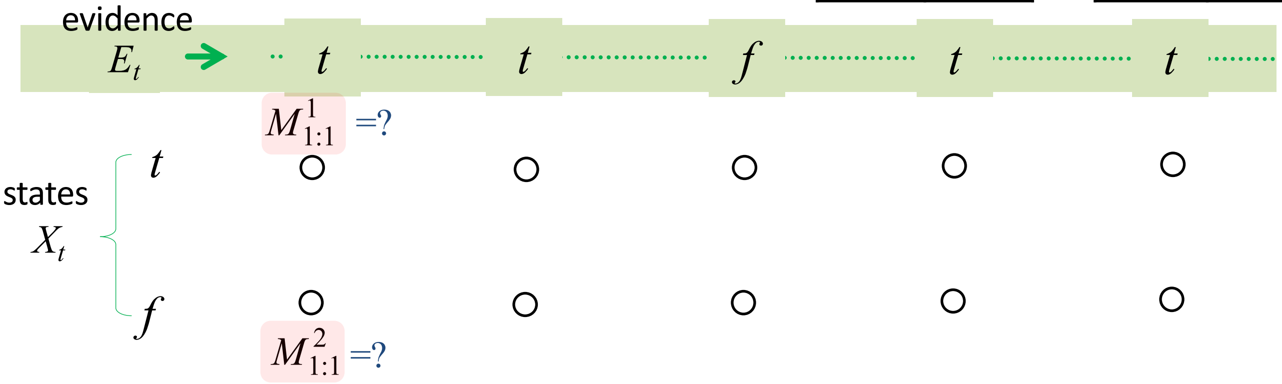
for the path from $X_t=t$ for the path from $X_t=f$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



Initialization (note: this is actually **filtering**)

$$\mathbf{m}_{1:1} = [M^1_{1:1} \;\; M^2_{1:1}]^{\mathsf{T}} = \mathbf{P}(X_1 \mid \mathbf{e}_{1:1}) = \mathbf{P}(X_1 \mid e_1) = \alpha \mathbf{P}(e_1 \mid X_1)\mathbf{P}(X_1)$$

$$= \alpha \langle P(e_1 \mid X_1 = t),\; P(e_1 \mid X_1 = f)\rangle \langle P(X_1 = t),\; P(X_1 = f)\rangle$$

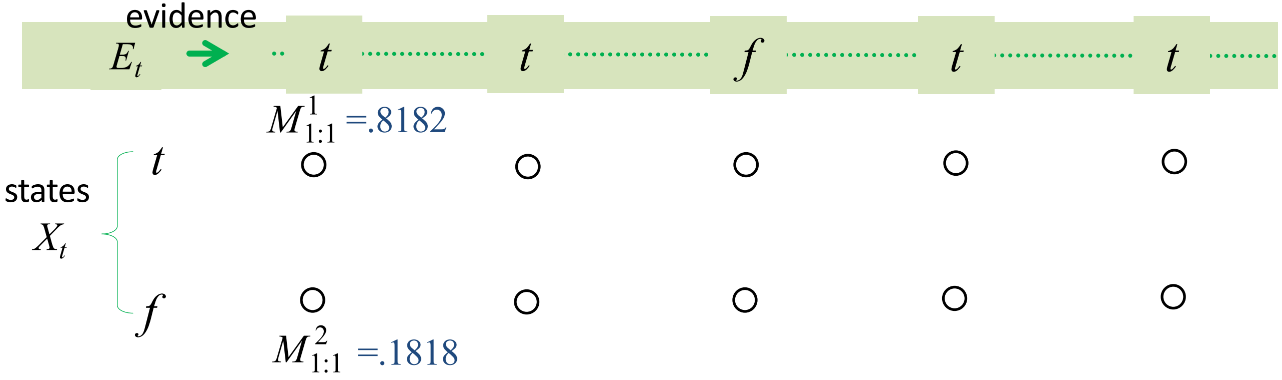
$$= \alpha \langle 0.9,\; 0.2\rangle \langle 0.5,\; 0.5\rangle = \alpha \langle 0.45,\; 0.1\rangle = \langle 0.8182,\; 0.182\rangle$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2

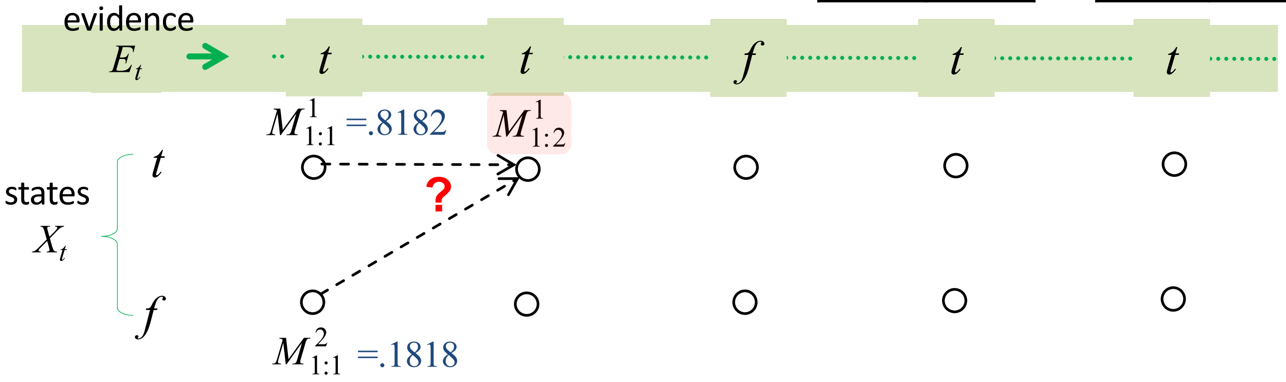


Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



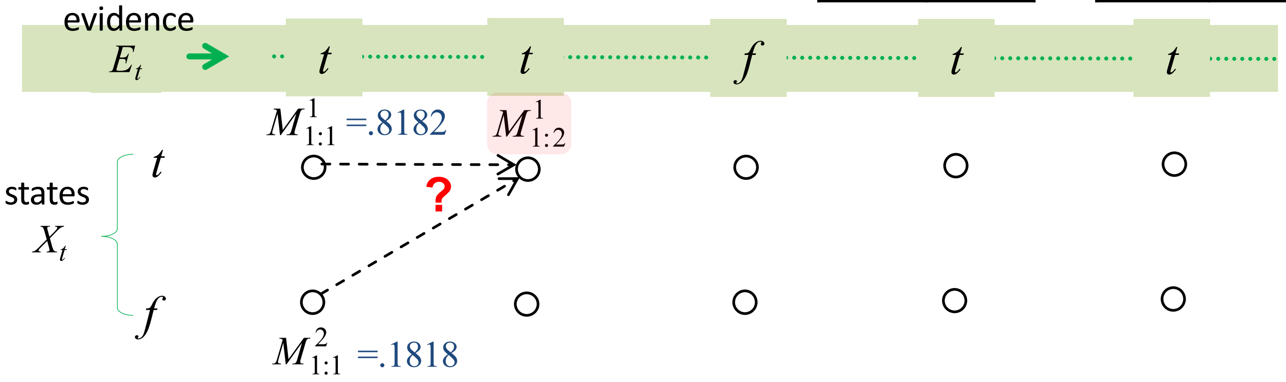
$$M^1_{1:2} = P(e_2 \mid X_2 = t) \max\left\{ P(X_2 = t \mid X_1 = t) M^1_{1:1},\ P(X_2 = t \mid X_1 = f) M^2_{1:1} \right\}$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



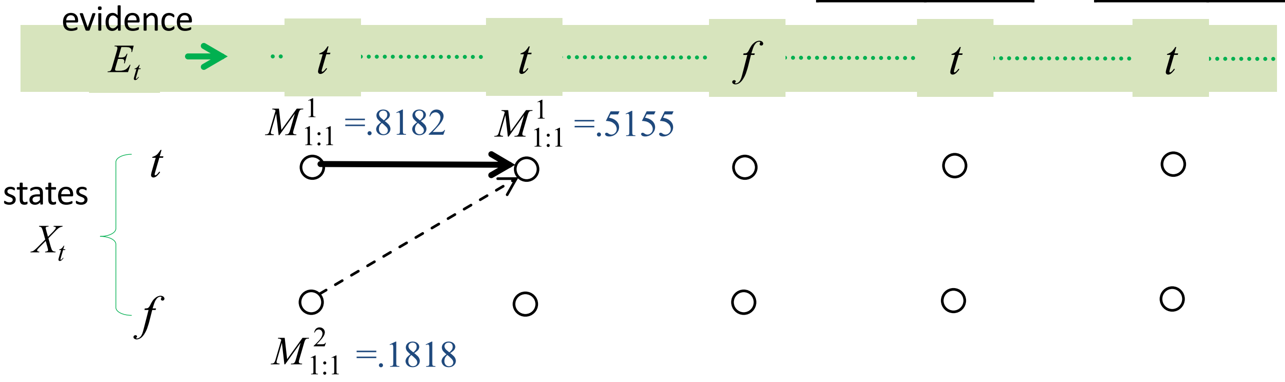
$$M^1_{1:2} = \underbrace{P(e_2 \mid X_2 = t)}_{0.9} \max\left\{ \underbrace{P(X_2 = t \mid X_1 = t)}_{0.7} \underbrace{M^1_{1:1}}_{0.8182},\ \underbrace{P(X_2 = t \mid X_1 = f)}_{0.3} \underbrace{M^2_{1:1}}_{0.1818} \right\}$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^1_{1:2} = \underbrace{P(e_2 \mid X_2 = t)}_{0.9} \max\left\{ \underbrace{P(X_2 = t \mid X_1 = t)}_{0.7}\underbrace{M^1_{1:1}}_{0.8182},\ \underbrace{P(X_2 = t \mid X_1 = f)}_{0.3}\underbrace{M^2_{1:1}}_{0.1818} \right\}$$

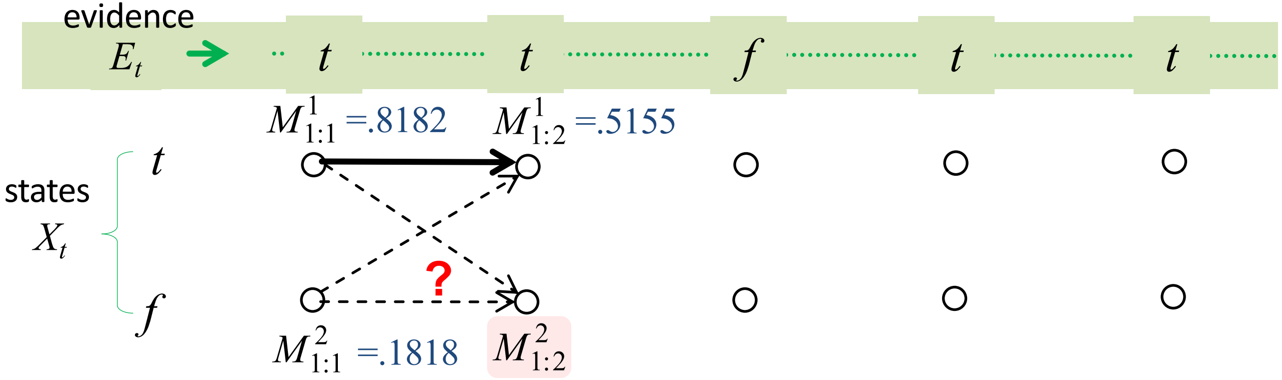
$$= 0.5155$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



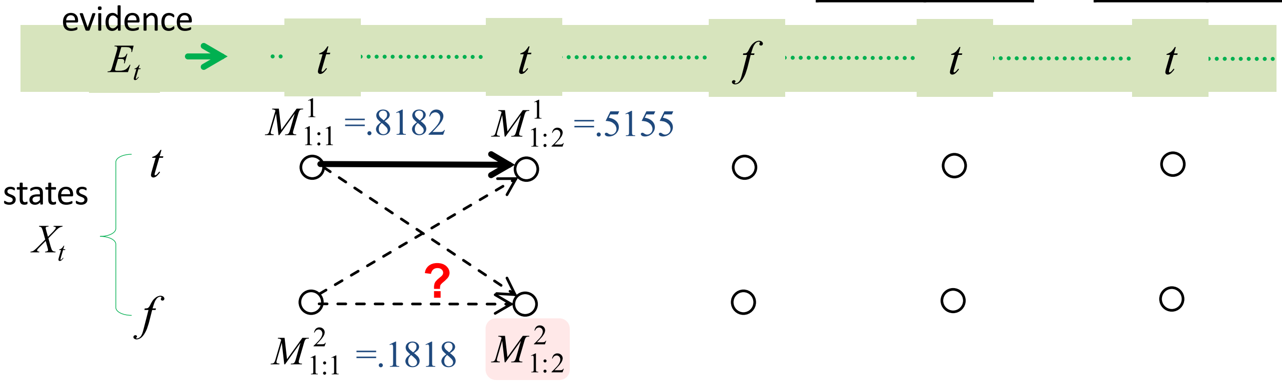
$$M^2_{1:2} = P(e_2 \mid X_2 = f)\max\left\{P(X_2 = f \mid X_1 = t)M^1_{1:1},\ P(X_2 = f \mid X_1 = f)M^2_{1:1}\right\}$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



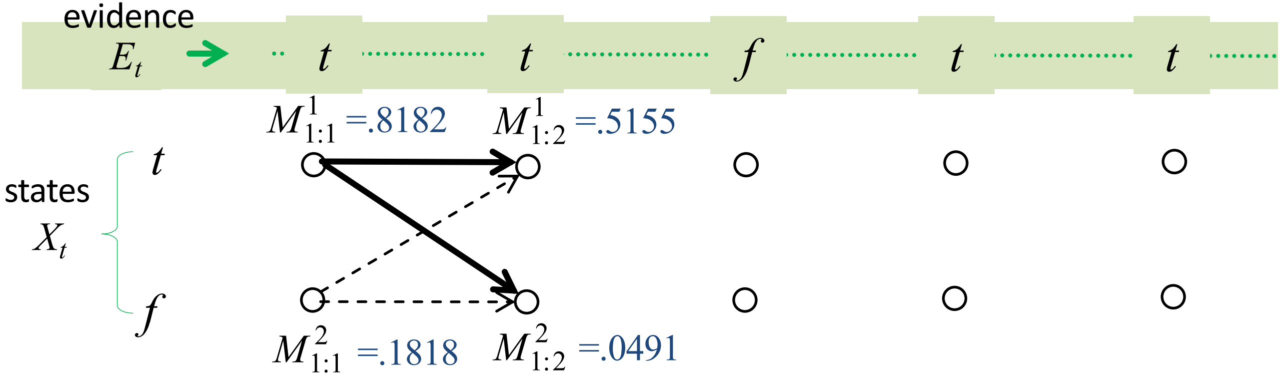
$$M^2_{1:2} = \underbrace{P(e_2 \mid X_2 = f)}_{0.2} \max\left\{ \underbrace{P(X_2 = f \mid X_1 = t)}_{0.3} \underbrace{M^1_{1:1}}_{0.8182},\ \underbrace{P(X_2 = f \mid X_1 = f)}_{0.7} \underbrace{M^2_{1:1}}_{0.1818} \right\}$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^2_{1:2} = \underbrace{P(e_2 \mid X_2 = f)}_{0.2} \max\left\{ \underbrace{P(X_2 = f \mid X_1 = t)}_{0.3}\underbrace{M^1_{1:1}}_{0.8182},\ \underbrace{P(X_2 = f \mid X_1 = f)}_{0.7}\underbrace{M^2_{1:1}}_{0.1818} \right\}$$

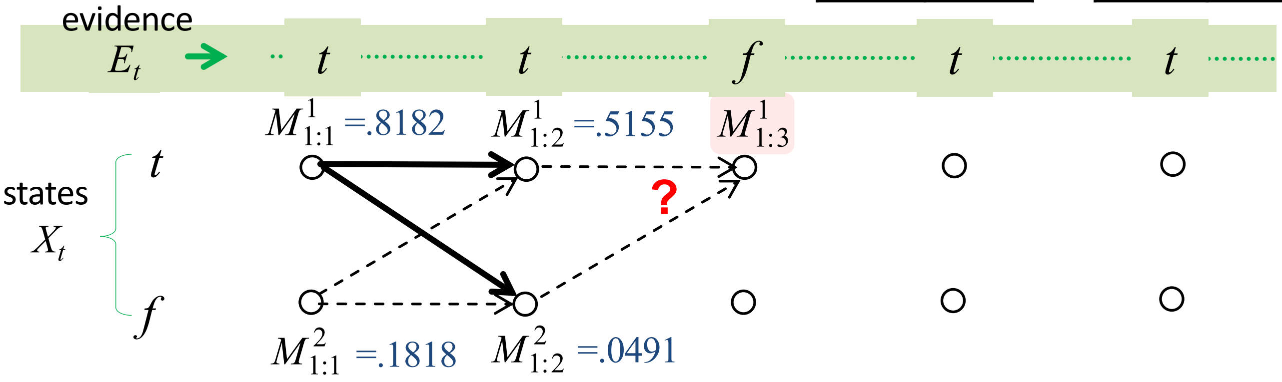
$$= 0.0491$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^1_{1:3} = \underbrace{P(e_3 \mid X_3 = t)}_{0.1} \max\left\{ \underbrace{P(X_3 = t \mid X_2 = t)}_{0.7}\underbrace{M^1_{1:2}}_{0.5155},\ \underbrace{P(X_3 = t \mid X_2 = f)}_{0.3}\underbrace{M^2_{1:2}}_{0.0491} \right\}$$

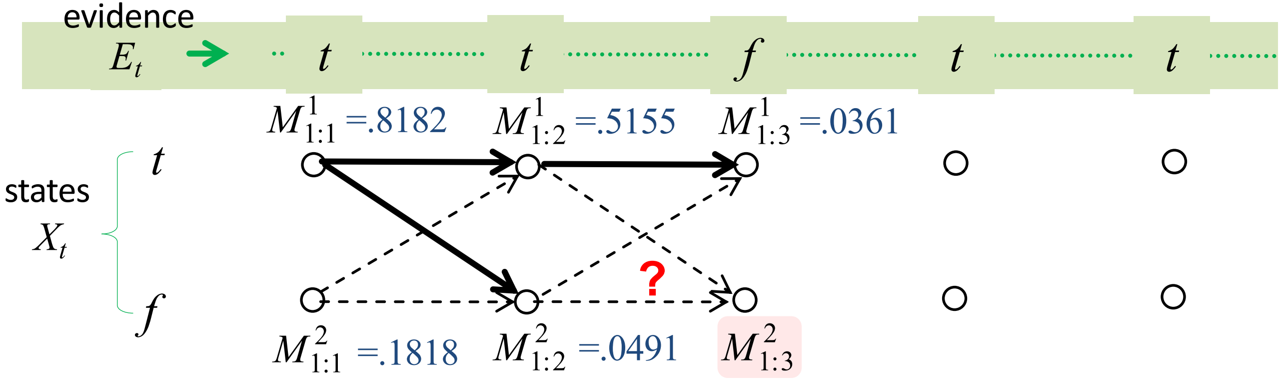
$$= 0.0361$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^2_{1:3} = \underbrace{P(e_3 \mid X_3 = f)}_{0.8} \max\left\{ \underbrace{P(X_3 = f \mid X_2 = t)}_{0.3} \underbrace{M^1_{1:2}}_{0.5155},\ \underbrace{P(X_3 = f \mid X_2 = f)}_{0.7} \underbrace{M^2_{1:2}}_{0.0491} \right\}$$

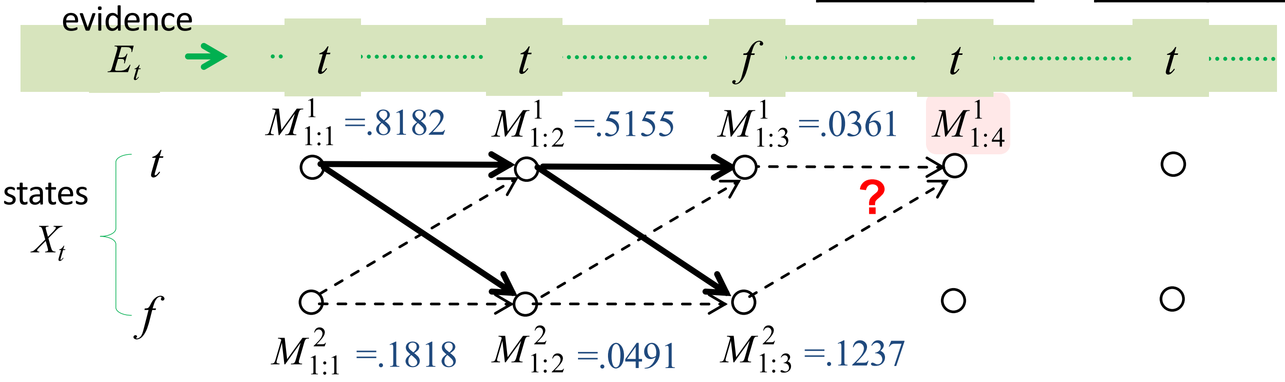
$$= 0.1237$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^1_{1:4} = \underbrace{P(e_4 \mid X_4 = t)}_{0.9} \max\left\{ \underbrace{P(X_4 = t \mid X_3 = t)}_{0.7} \underbrace{M^1_{1:3}}_{0.0361},\ \underbrace{P(X_4 = t \mid X_3 = f)}_{0.3} \underbrace{M^2_{1:3}}_{0.1237} \right\}$$

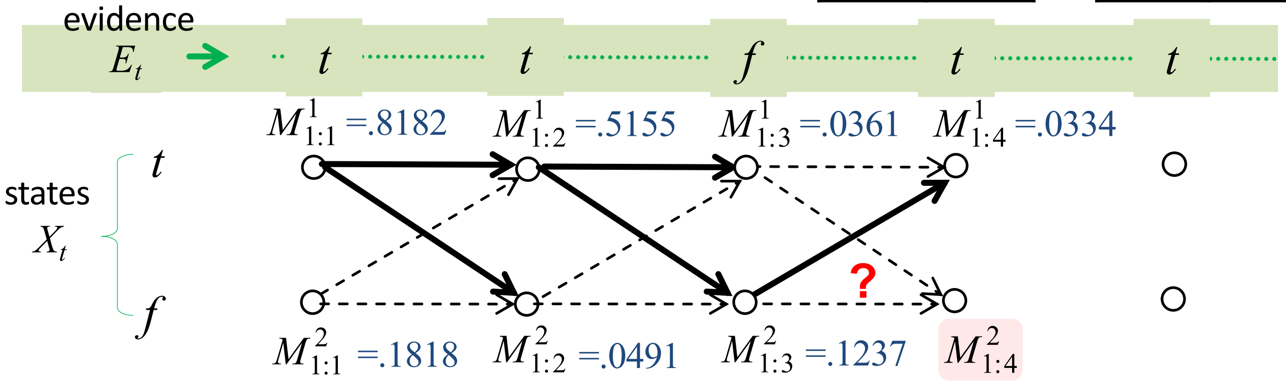
$$= 0.0334$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^2_{1:4} = \underbrace{P(e_4 \mid X_4 = f)}_{0.2} \max\Big\{ \underbrace{P(X_4 = f \mid X_3 = t)}_{0.3} \underbrace{M^1_{1:3}}_{0.0361},\ \underbrace{P(X_4 = f \mid X_3 = f)}_{0.7} \underbrace{M^2_{1:3}}_{0.1237} \Big\}$$

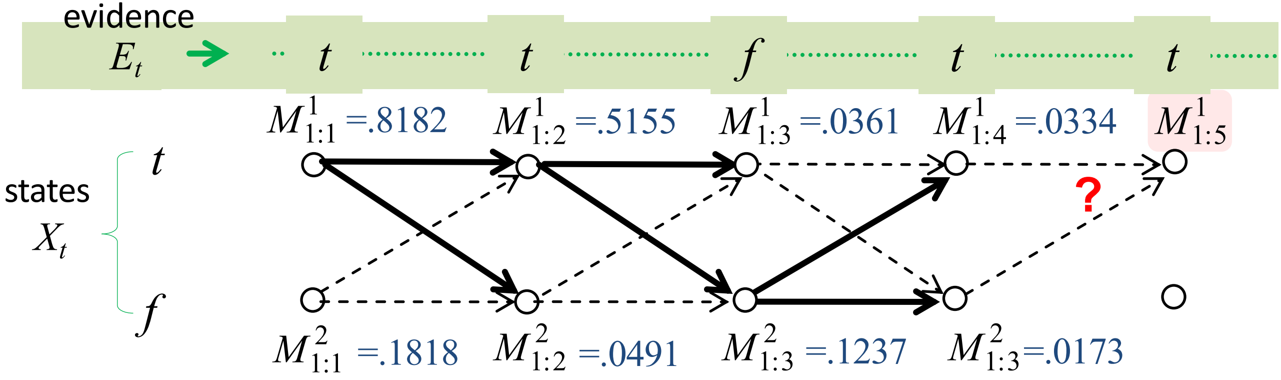
$$= 0.0173$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^1_{1:5} = \underbrace{P(e_5 \mid X_5 = t)}_{0.9} \max\left\{ \underbrace{P(X_5 = t \mid X_4 = t)}_{0.7} \underbrace{M^1_{1:4}}_{0.0334},\ \underbrace{P(X_5 = t \mid X_4 = f)}_{0.3} \underbrace{M^2_{1:4}}_{0.0173} \right\}$$

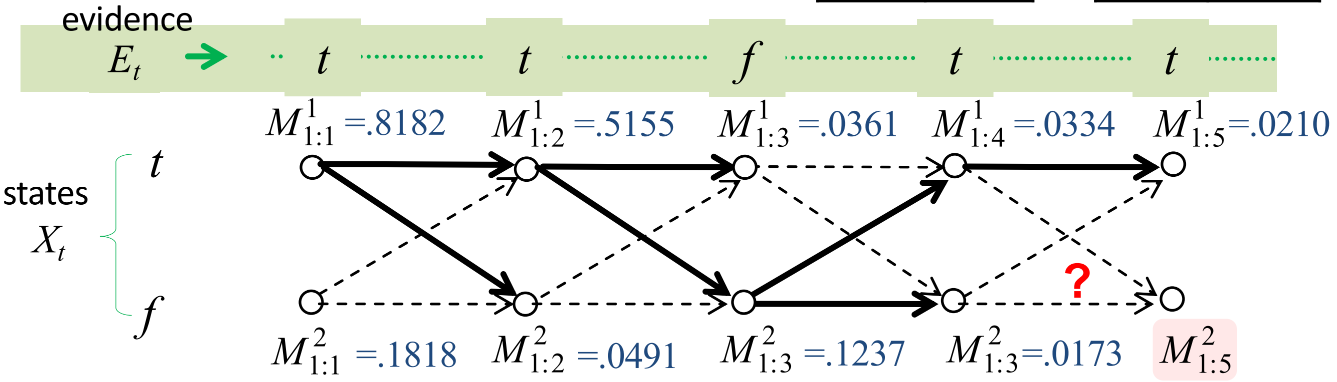
$$= 0.0210$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

X_{t-1}	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2



$$M^2_{1:5} = \underbrace{P(e_5 \mid X_5 = f)}_{0.2} \max\left\{ \underbrace{P(X_5 = f \mid X_4 = t)}_{0.3} \underbrace{M^1_{1:4}}_{0.0334},\ \underbrace{P(X_5 = f \mid X_4 = f)}_{0.7} \underbrace{M^2_{1:4}}_{0.0173} \right\}$$

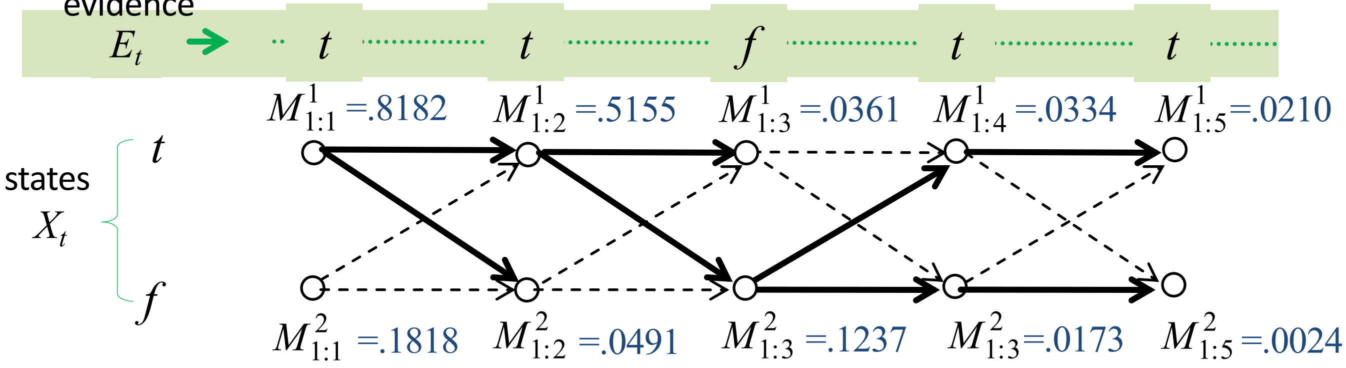
$$= 0.0024$$

Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2

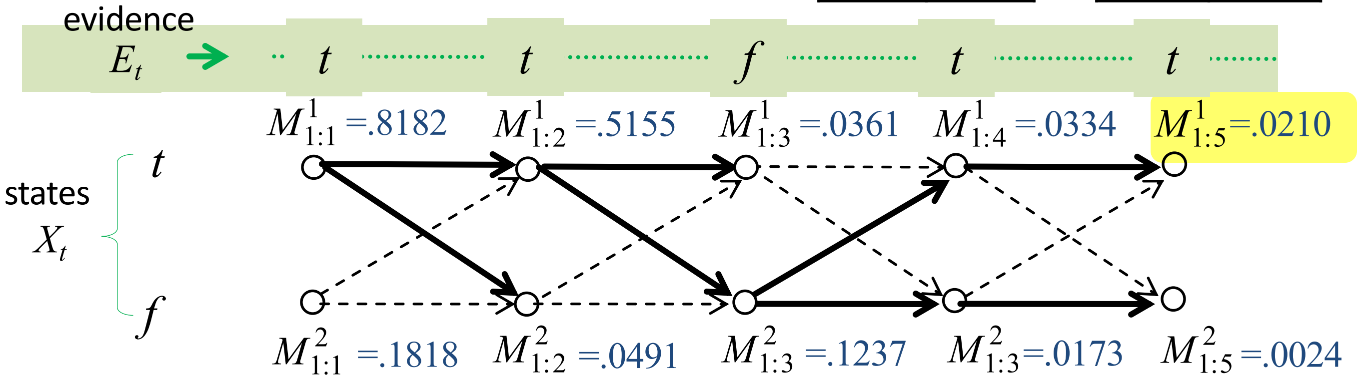


Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2

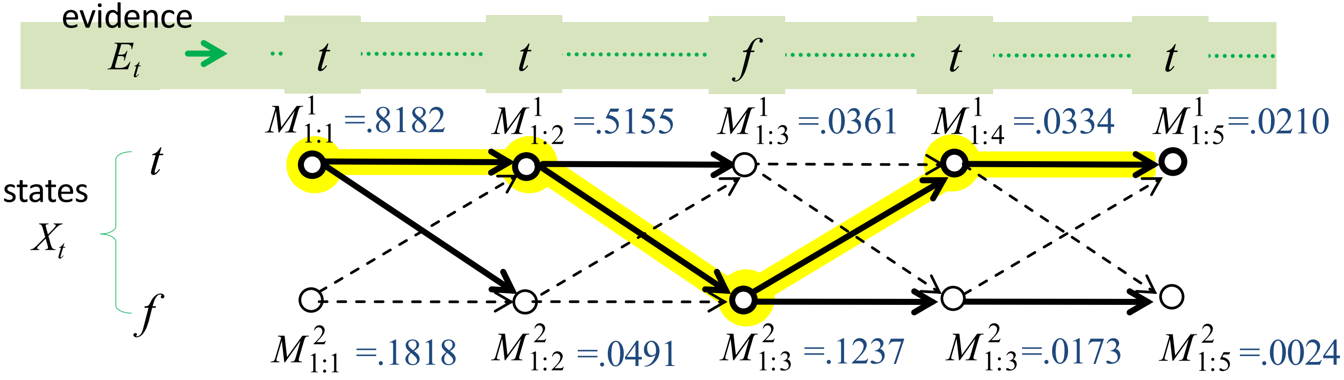


Viterbi Algorithm

Example with 2 states $X_t = \{t, f\}$

$X_{t\text{-}1}$	$P(X_t)$
t	0.7
f	0.3

X_t	$P(E_t)$
t	0.9
f	0.2

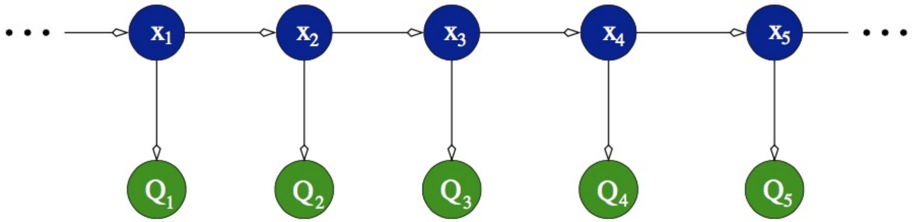


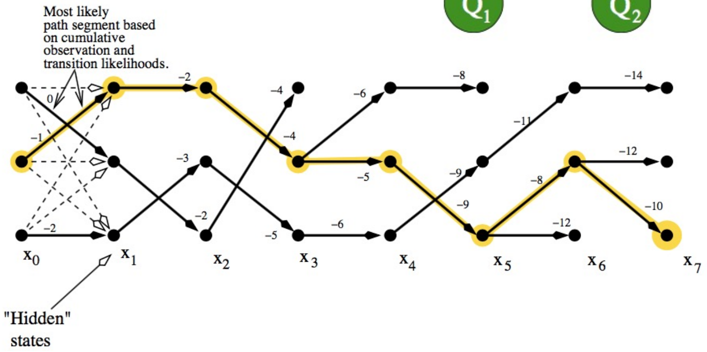
The most likely sequence is $\mathbf{x} = \{t, t, f, t, t\}$

i.e., in our concrete example: $\{Rain, Rain, NoRain, Rain, Rain\}$

A slightly more complex example

$$O_{ij} = \begin{array}{r} \\ \text{Walk in park} \\ \text{Shopping} \\ \text{Clean flat} \\ \text{Visit girlfriend} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rainy} & \text{Sunny} & \text{Windy} \end{array} \\ \begin{bmatrix} 0.05 & 0.7 & 0.25 \\ 0.05 & 0.15 & 0.4 \\ 0.25 & 0.05 & 0.05 \\ 0.65 & 0.1 & 0.3 \end{bmatrix} \end{array}$$





Some notes on the Viterbi algorithm

- This example explained finding the most likely sequence using the Viterbi algorithm
- Note that this is a variant of the "max-product" product algorithm with back-tracking (see the slides on inference in graphical models). The algorithm maximizes over the incoming messages to each node.
- Remember that "sum-product" algorithm gives marginal probabilities in the graph nodes, while "max-product" gives the maximum joint probability of the network configuration (or in the case of temporal models, the maximum probability of a sequence of states).
- In practice, computing products of small numbers may become instable, so often the problem is reformulated as finding the maximum of the logarithm of the joint probability
  - This results in replacing the products by summations, and yields the "max-sum" algorithm.
  - Back-tracking gives the same sequence with the "max-product" and "max-sum" variants.