Viterbi Algorithm

Natalie Parde
UIC CS 421

Decoding

- Given an observation sequence and an HMM, what is the best hidden state sequence?
  - How do we choose a state sequence that is optimal in some sense (e.g., best explains the observations)?
- Very useful for sequence labeling!

Naïve Approach:

- For each hidden state sequence Q, compute P(O|Q)
- Pick the sequence with the highest probability

However, this is computationally inefficient!

- $O(N^T)$

Decoding

How can we decode sequences more efficiently?

- **Viterbi Algorithm**
  - Another dynamic programming algorithm
  - Uses a similar trellis to the Forward algorithm
- Viterbi time complexity: $O(N^2T)$

Viterbi Intuition

- **Goal:** Compute the joint probability of the observation sequence together with the best state sequence
- So, **recursively compute the probability of the most likely subsequence of states** that accounts for the first t observations and ends in state q_j.
  - $v_t(j) = \max_{q_0,q_1,\ldots,q_{t-1}} P(q_0, q_1, \ldots, q_{t-1}, o_1, \ldots, o_t, q_t = q_j|\lambda)$
- Also **record backpointers** that subsequently allow you to backtrace the most probable state sequence
  - $bt_t(j)$ stores the state at time t-1 that maximizes the probability that the system was in state q_j at time t, given the observed sequence

Formal Algorithm

create a path probability matrix Viterbi[N+2,T]

for each state q in [1,…,N] do:

Viterbi[q,1] ← $a_{0,q} * b_q(o_1)$

backpointer[q,1] ← 0

for each time step t in [2,…,T] do:

for each state q in [1,…,N] do:

$$viterbi[q,t] \leftarrow \max_{q' \in [1,\ldots,N]} viterbi[q', t-1] * a_{q',q} * b_q(o_t)$$

$$backpointer[q,t] \leftarrow \operatorname{argmax}_{q' \in [1,\ldots,N]} viterbi[q', t-1] * a_{q',q} * b_q(o_t)$$

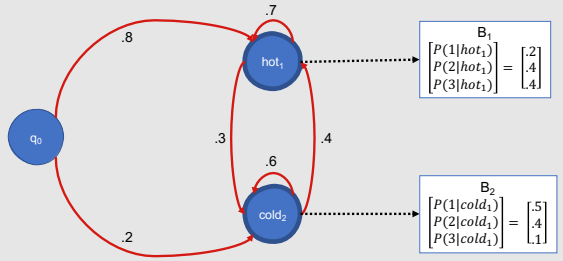
$$bestpathprob \leftarrow \max_{q' \in [1,\ldots,N]} viterbi[q', T]$$

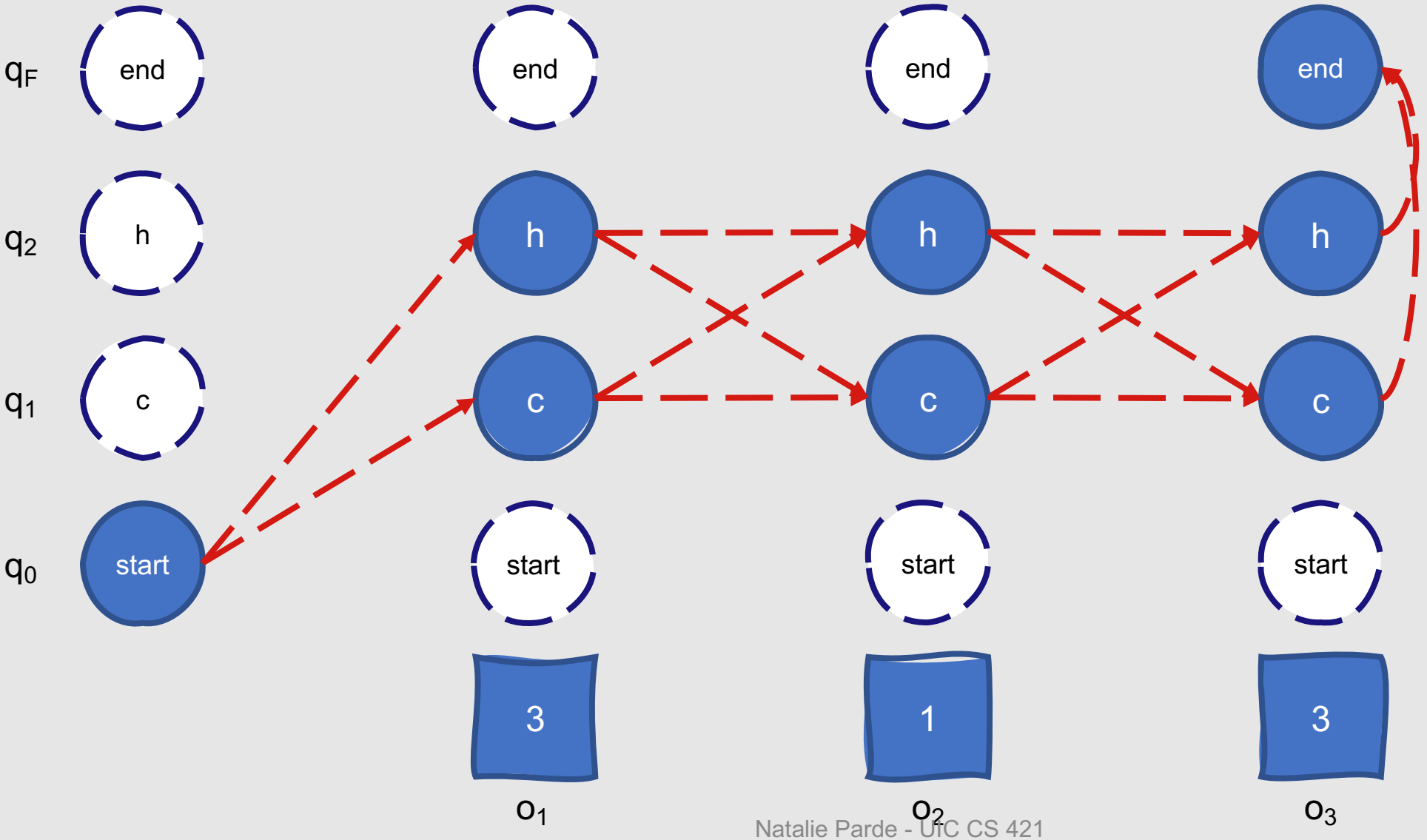
$$bestpathpointer \leftarrow \operatorname{argmax}_{q' \in [1,\ldots,N]} viterbi[q', T]$$

Seem familiar?

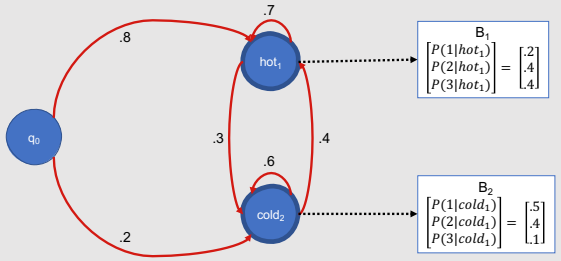
- Viterbi is basically the forward algorithm + backpointers, and substituting a max function for the summation operator

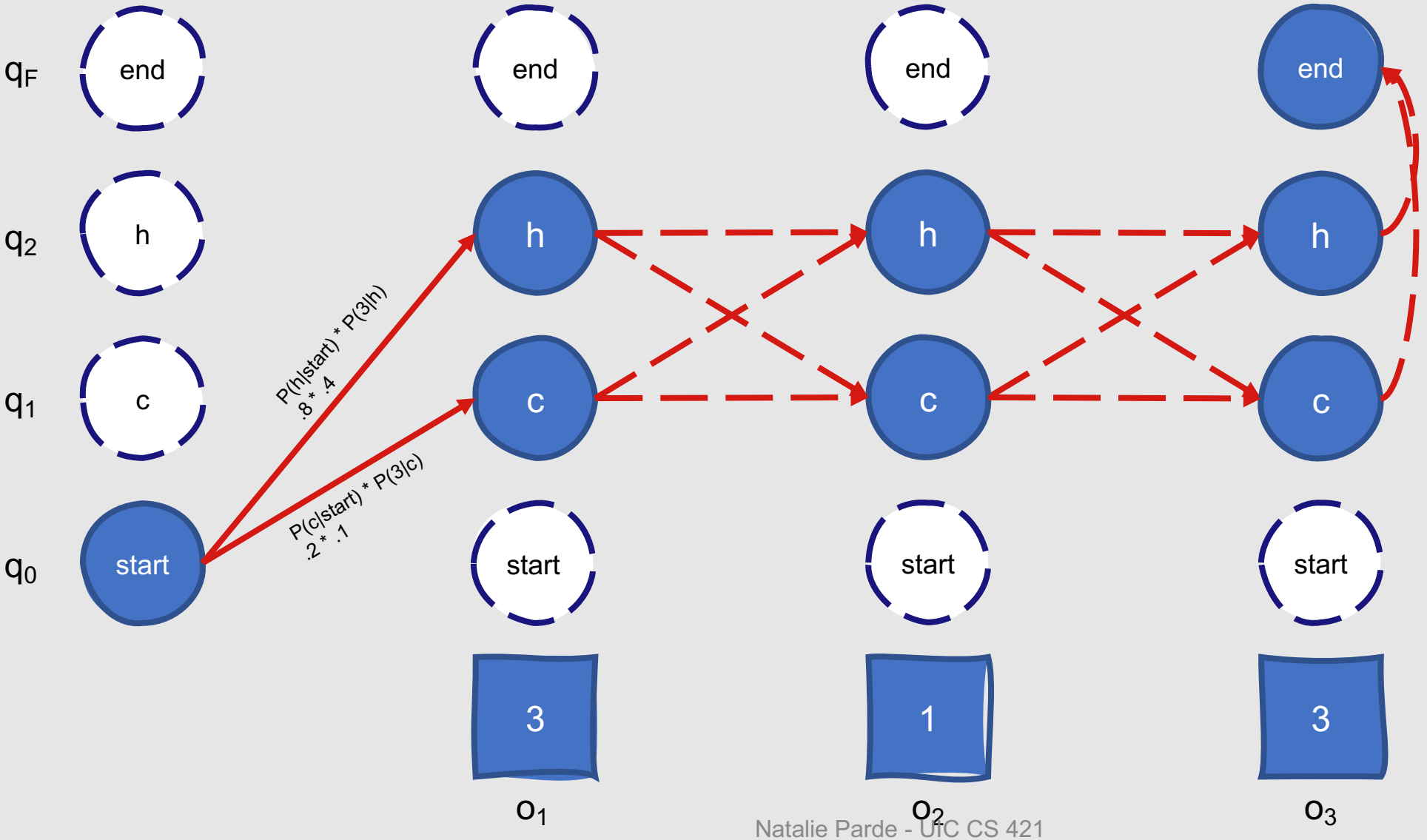
Viterbi Trellis



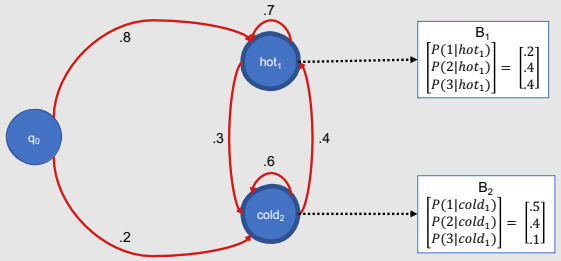


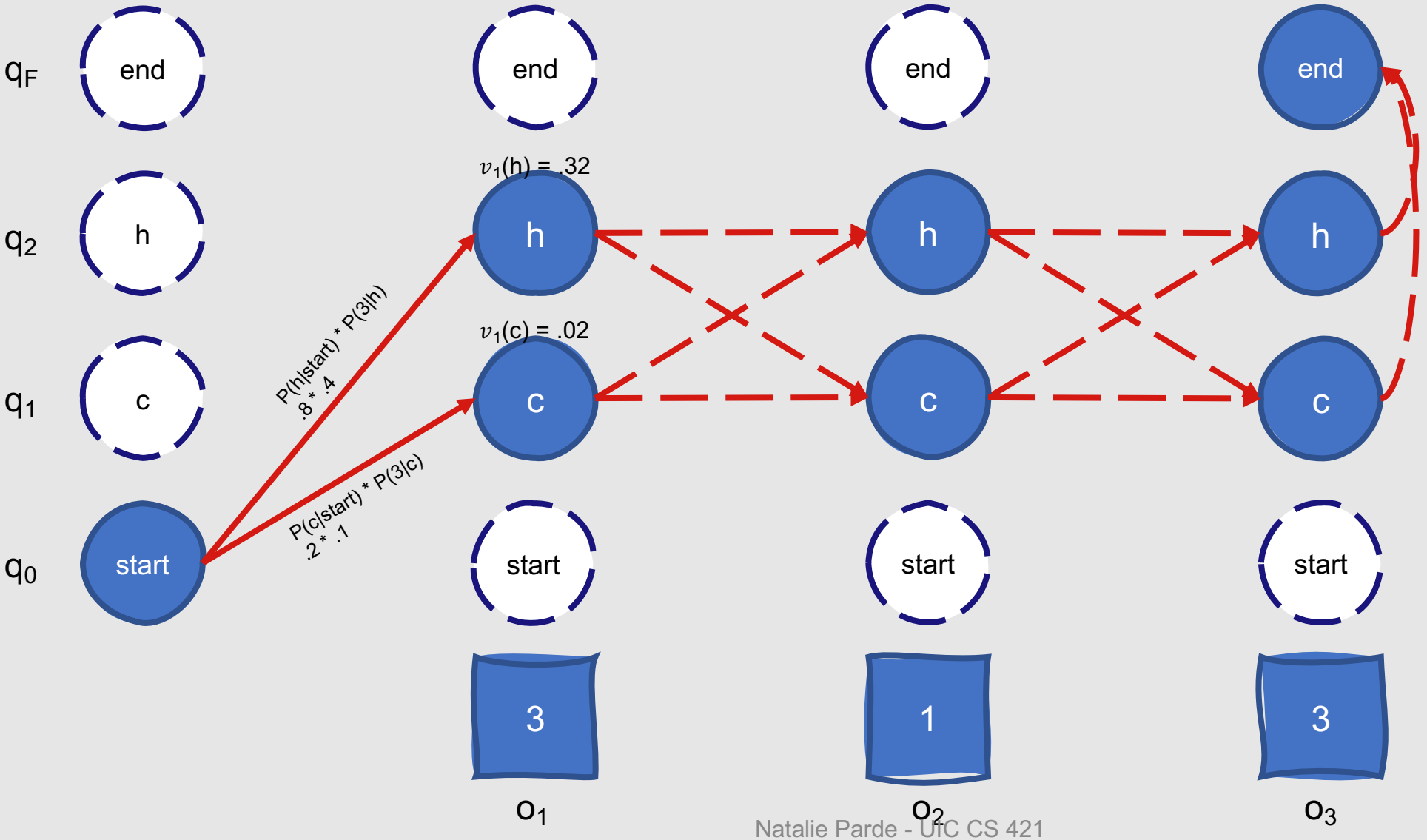
Viterbi Trellis



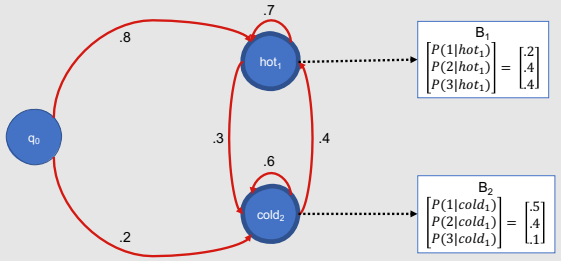


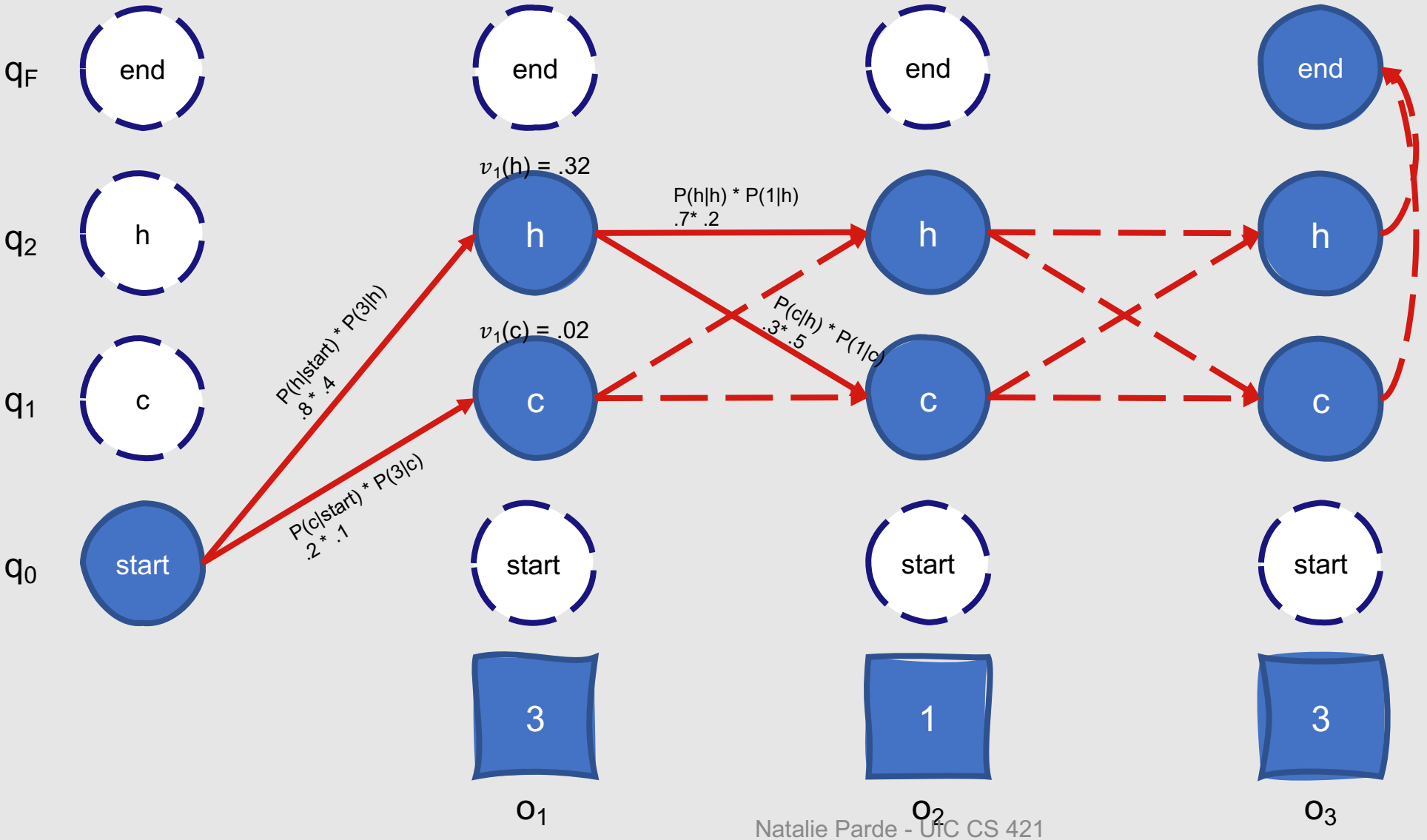
Viterbi Trellis



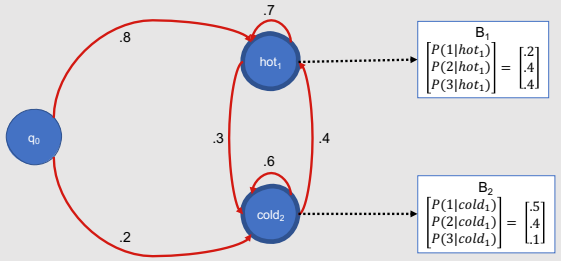


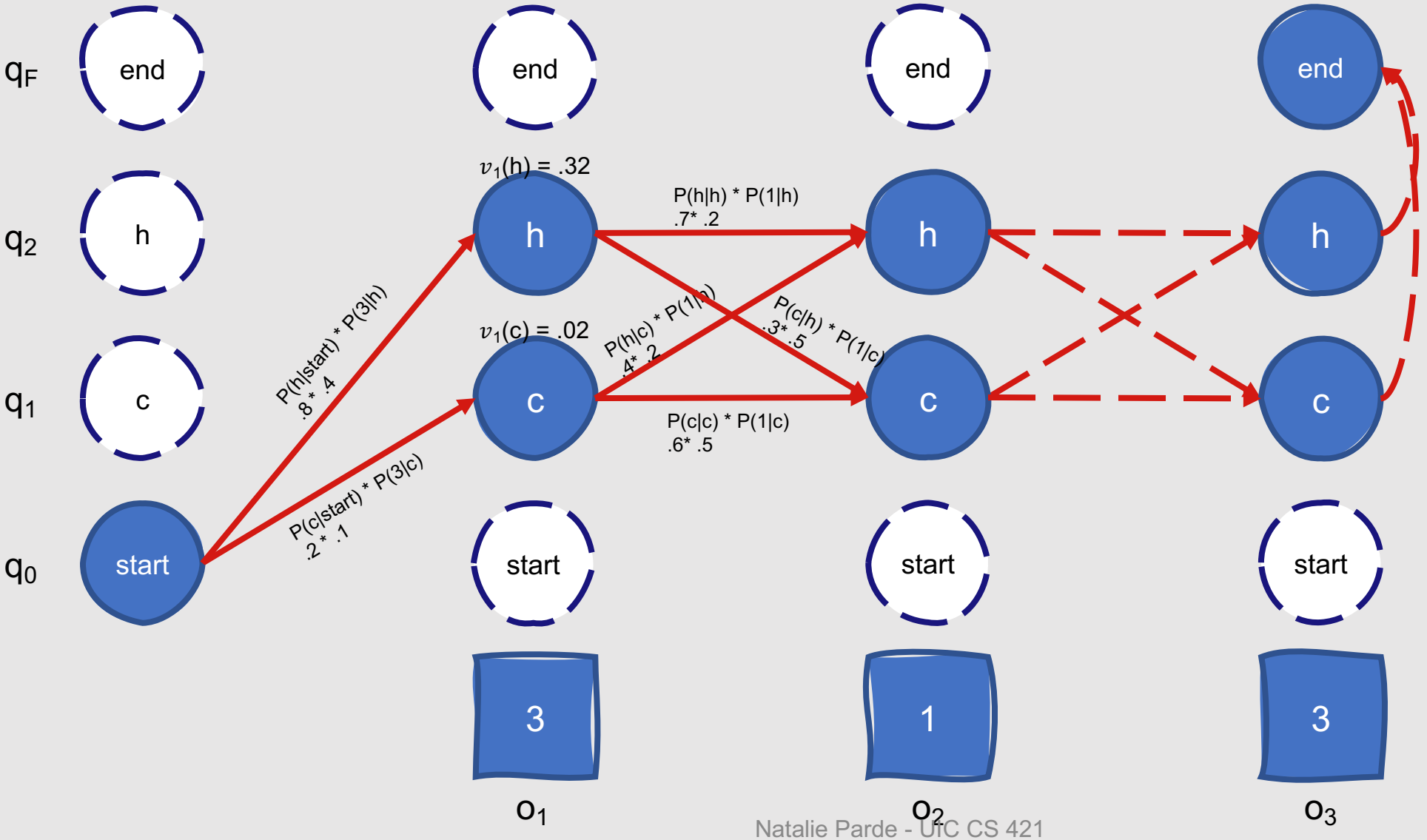
Viterbi Trellis



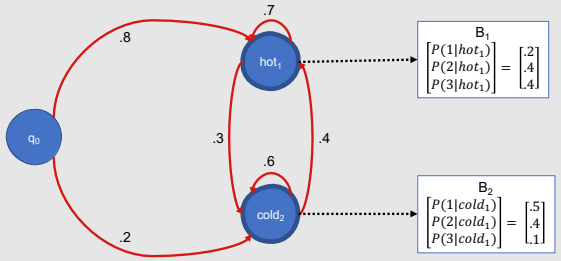


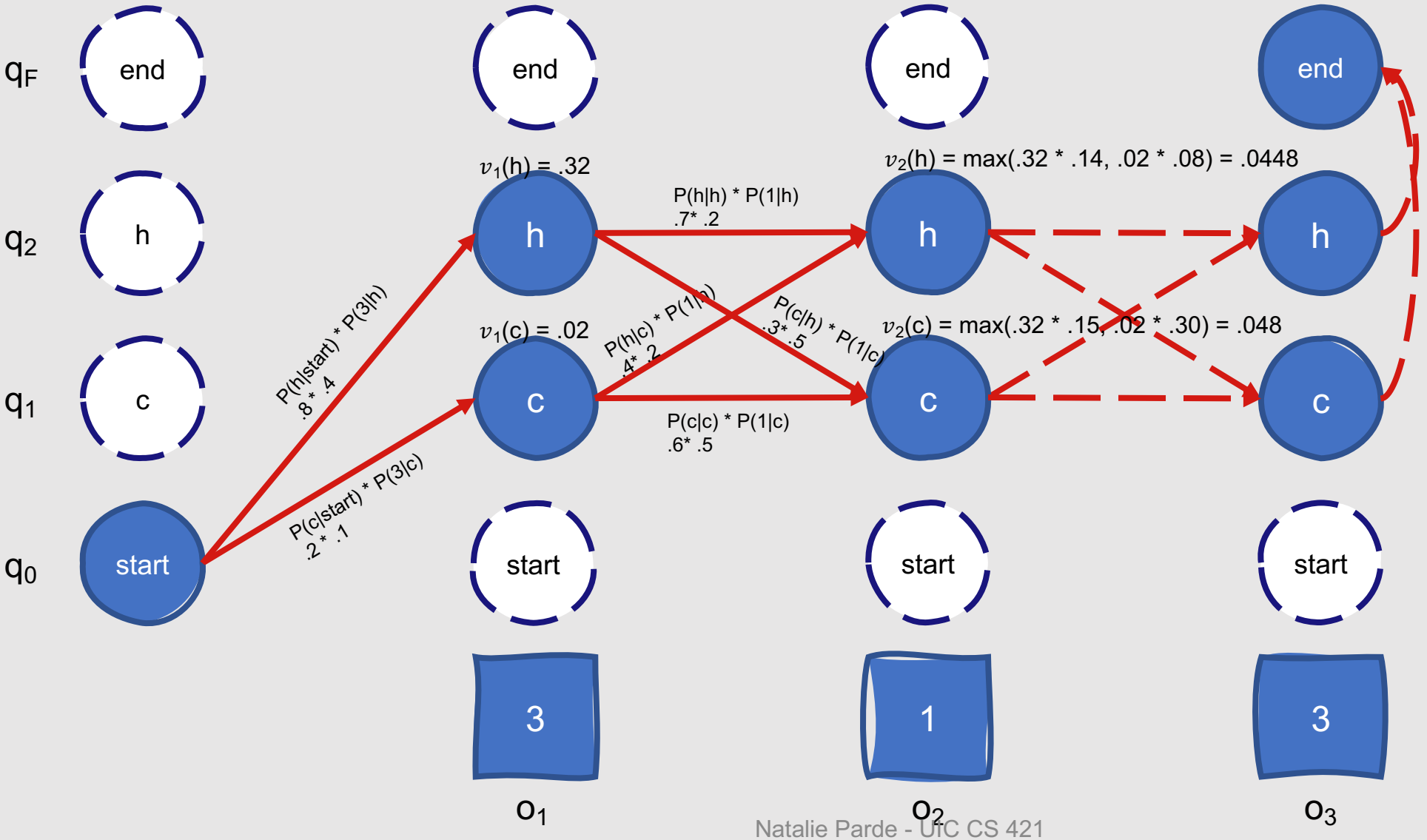
Viterbi Trellis



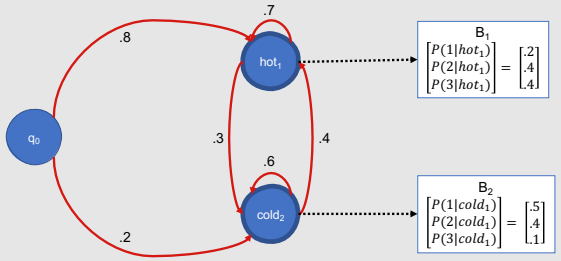


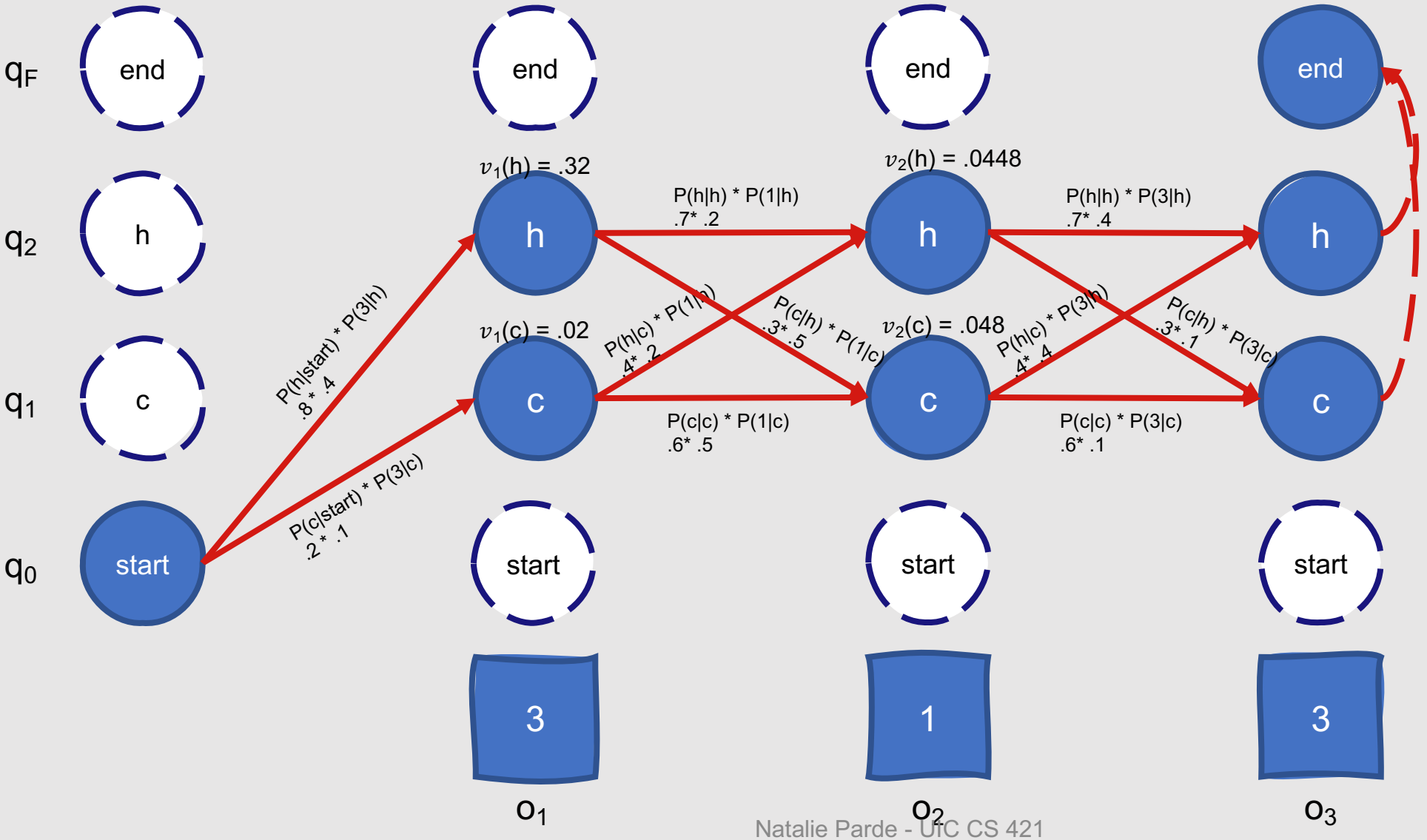
Viterbi Trellis



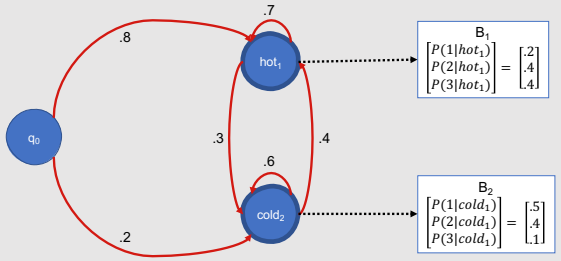


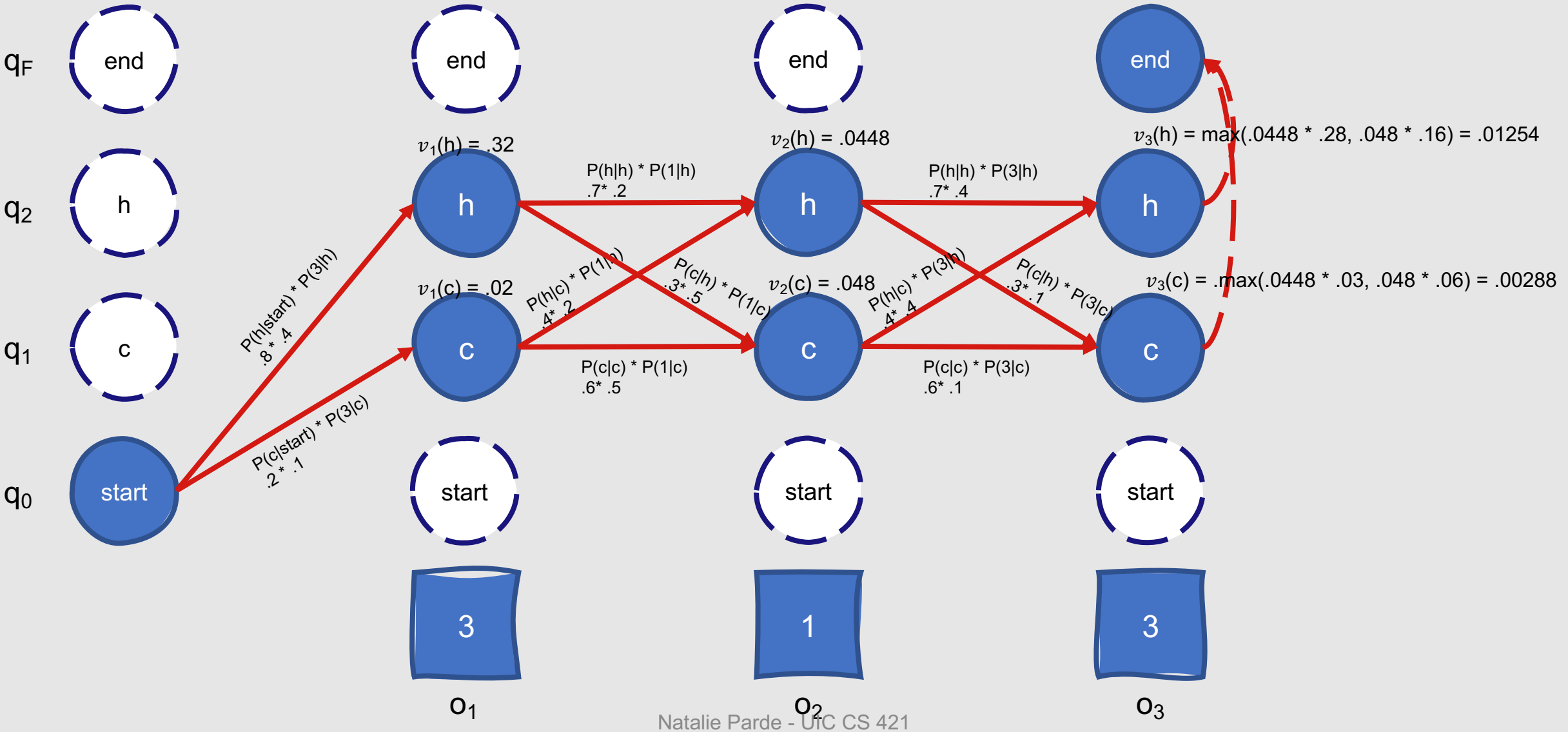
Viterbi Trellis



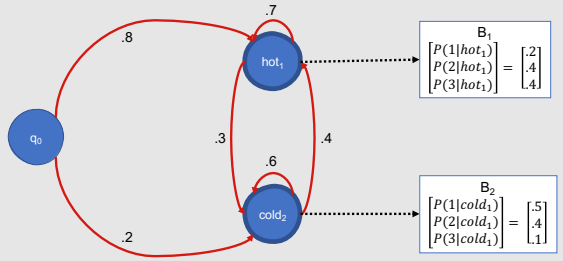


Viterbi Trellis





Viterbi Trellis



$bestpathprob$ = max(.01254, .00288) = .01254

q_F end

q_2 h

q_1 c

q_0 start

v_1(h) = .32

v_2(h) = .0448

v_3(h) = max(.0448 * .28, .048 * .16) = .01254

v_1(c) = .02

v_2(c) = .048

v_3(c) = .max(.0448 * .03, .048 * .06) = .00288

P(h|start) * P(3|h)
.8 * .4

P(c|start) * P(3|c)
.2 * .1

P(h|h) * P(1|h)
.7* .2

P(h|c) * P(1|h)
.4* .2

P(c|h) * P(1|c)
.3* .5

P(c|c) * P(1|c)
.6* .5

P(h|h) * P(3|h)
.7* .4

P(h|c) * P(3|h)
.4* .4

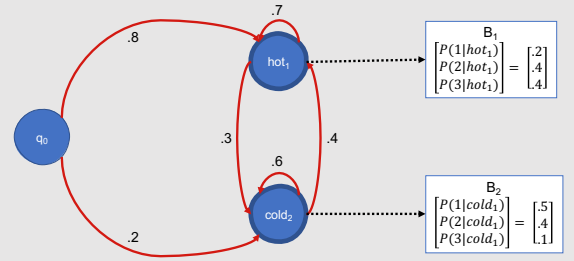
P(c|h) * P(3|c)
.3* .1

P(c|c) * P(3|c)
.6* .1

3 1 3

o_1 o_2 o_3

Viterbi Backtrace



$bestpathprob$ = max(.01254, .00288) = .01254

q_F q_2 q_1 q_0

end end end end

h h h h

c c c c

start start start start

v_1(h) = .32

v_2(h) = .0448

v_3(h) = max(.0448 * .28, .048 * .16) = .01254

v_1(c) = .02

v_2(c) = .048

v_3(c) = .max(.0448 * .03, .048 * .06) = .00288

P(h|start) * P(3|h)
.8 * .4

P(c|start) * P(3|c)
.2 * .1

P(h|h) * P(1|h)
.7* .2

P(h|c) * P(1|c)
.4* .2

P(c|h) * P(1|c)
.3* .5

P(c|c) * P(1|c)
.6* .5

P(h|h) * P(3|h)
.7* .4

P(h|c) * P(3|c)
.4* .4

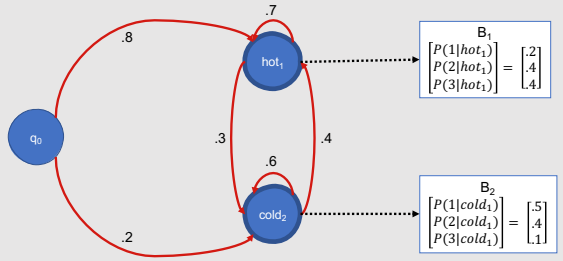
P(c|h) * P(3|c)
.3* .1

P(c|c) * P(3|c)
.6* .1

3 1 3

o_1 o_2 o_3

Viterbi Backtrace



$bestpathprob$ = max(.01254, .00288) = .01254

q_F q_2 q_1 q_0

end end end end

h h h h

c c c c

start start start start

v_1(h) = .32

v_1(c) = .02

P(h|start) * P(3|h)
.8 * .4

P(c|start) * P(3|c)
.2 * .1

P(h|h) * P(1|h)
.7* .2

P(h|c) * P(1|c)
.4* .2

P(c|h) * P(1|c)
.3* .5

P(c|c) * P(1|c)
.6* .5

v_2(h) = .0448

v_2(c) = .048

P(h|h) * P(3|h)
.7* .4

P(h|c) * P(3|c)
.4* .4

P(c|h) * P(3|c)
.3* .1

P(c|c) * P(3|c)
.6* .1

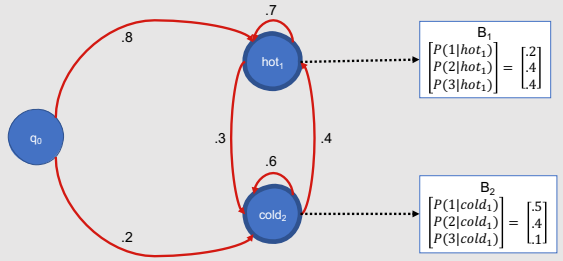
v_3(h) = max(.0448 * .28, .048 * .16) = .01254

v_3(c) = .max(.0448 * .03, .048 * .06) = .00288

3 1 3

o_1 o_2 o_3

Viterbi Backtrace



$bestpathprob$ = max(.01254, .00288) = .01254

q_F q_2 q_1 q_0

end end end end

h h h

c c c

start start start

3 1 3

o_1 o_2 o_3

v_1(h) = .32

v_2(h) = .0448

v_3(h) = max(.0448 * .28, .048 * .16) = .01254

v_1(c) = .02

v_2(c) = .048

v_3(c) = .max(.0448 * .03, .048 * .06) = .00288

P(h|start) * P(3|h)
.8 * .4

P(c|start) * P(3|c)
.2 * .1

P(h|h) * P(1|h)
.7* .2

P(h|h) * P(3|h)
.7* .4

P(h|c) * P(1|h)
.4* .2

P(c|h) * P(1|c)
.3* .5

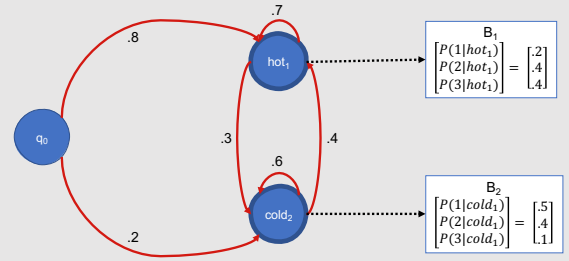
P(h|c) * P(3|h)
.4* .4

P(c|h) * P(3|c)
.3* .1

P(c|c) * P(1|c)
.6* .5

P(c|c) * P(3|c)
.6* .1

Viterbi Backtrace



$bestpathprob$ = max(.01254, .00288) = .01254

q_F q_2 q_1 q_0

end end end end

h h h

c c c

start start start

v_1(h) = .32 v_2(h) = .0448 v_3(h) = max(.0448 * .28, .048 * .16) = .01254

P(h|h) * P(1|h) .7* .2

P(h|h) * P(3|h) .7* .4

P(h|start) * P(3|h) .8 * .4

P(c|start) * P(3|c) .2 * .1

v_1(c) = .02 v_2(c) = .048 v_3(c) = .max(.0448 * .03, .048 * .06) = .00288

P(h|c) * P(1|c) .4* .2

P(c|h) * P(1|c) .3* .5

P(h|c) * P(3|c) .4* .4

P(c|h) * P(3|c) .3* .1

P(c|c) * P(1|c) .6* .5

P(c|c) * P(3|c) .6* .1

3 1 3

o_1 o_2 o_3