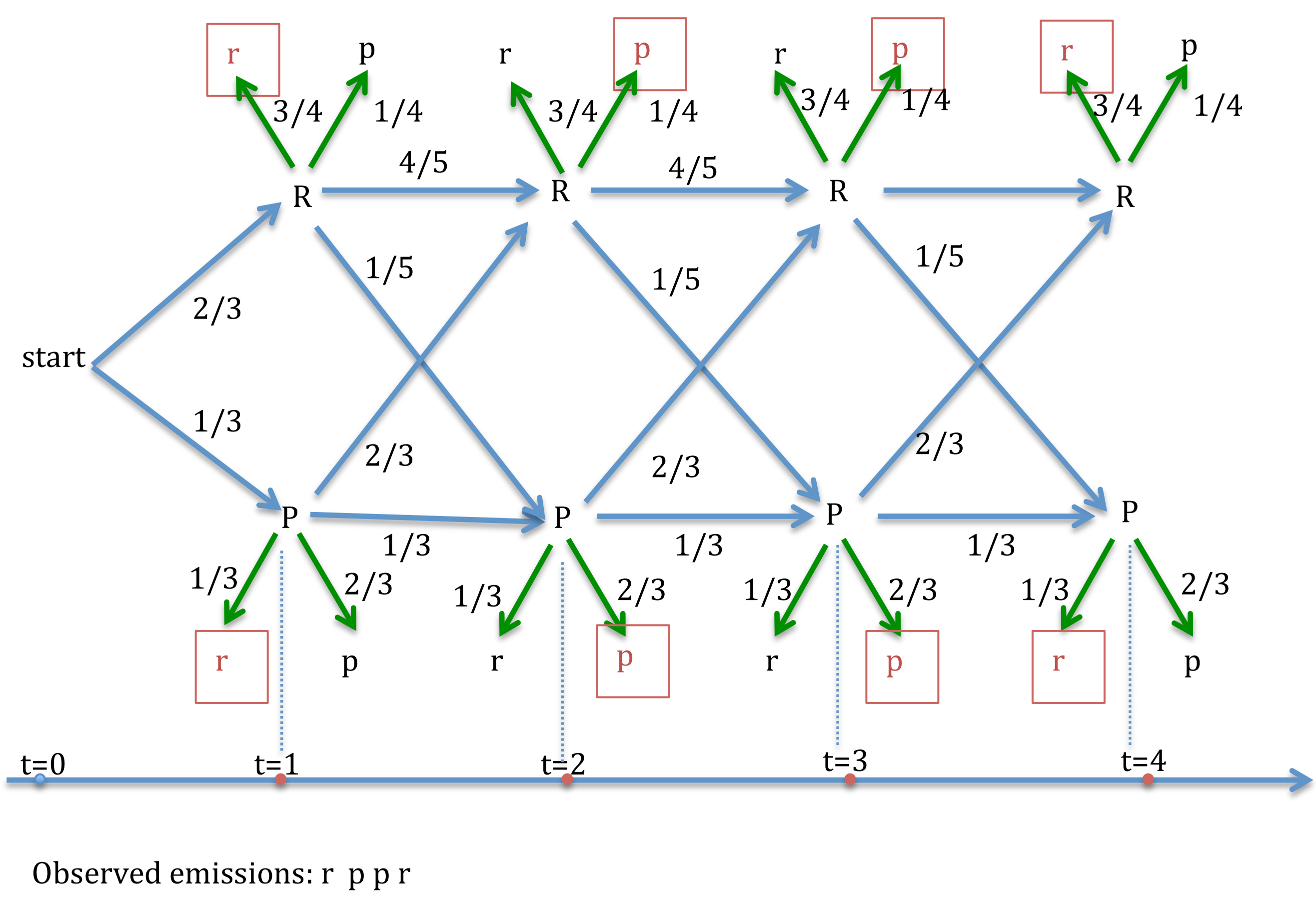
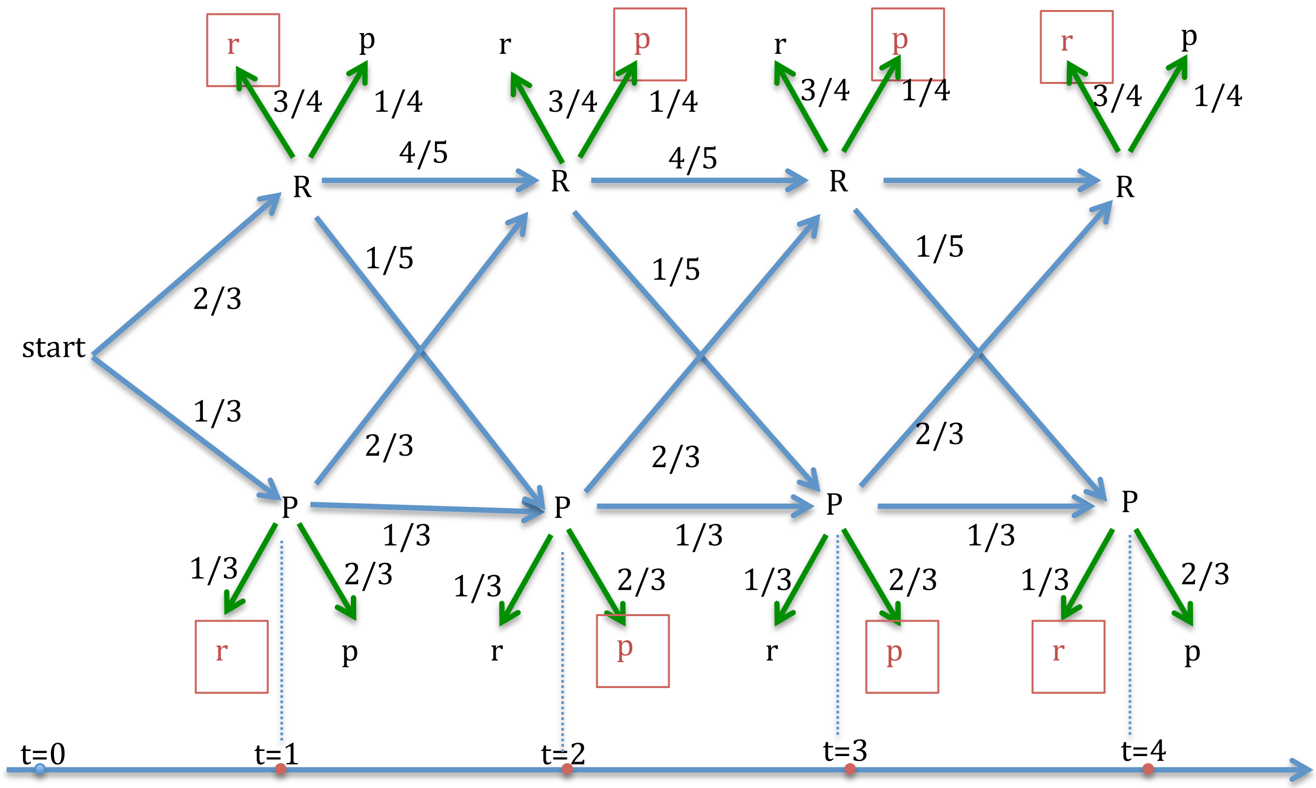
COMP4121

An example of the Viterbi algorithm application

- In California there are twice as many raccoons as possums. Having gotten a job with Google, you are in California observing your back yard. It is dusk, so the probability that you think you saw a raccoon when you are actually looking at a possum at a distance is $1/3$; the probability that you think you saw a possum while you are actually looking at a raccoon at a distance is $1/4$. Raccoons move in packs; so if a raccoon comes to your back yard the probability that the next animal to follow will also be a raccoon is $4/5$. Possums are solitary animals, so if a possum comes to your back yard, this does not impact the probabilities what the next animal to come will be (a possum or a raccoon, but recall in California there are twice as many raccoons as possums!) You believe that you saw four animals coming in the following order: a raccoon, a possum, a possum, a raccoon ($rppr$). Given such a sequence of observations, what actual sequence of animals is most likely to cause such a sequence of your observations?
- probabilities of initial states: $\pi(R) = 2/3$; $\pi(P) = 1/3$.
- transition probabilities:
  $\mathcal{P}(R \to R) = 4/5$; $\mathcal{P}(R \to P) = 1/5$; $\mathcal{P}(P \to P) = 1/3$; $\mathcal{P}(P \to R) = 2/3$.
- emission probabilities:
  $O(r|P) = 1/3$; $O(p|P) = 2/3$; $O(p|R) = 1/4$; $O(r|R) = 3/4$.





Observed emissions: r p p r

- The sequence of observations is $rppr$.
- probabilities of initial states: $\pi(R) = 2/3$; $\pi(P) = 1/3$.
- transition probabilities:
  $\mathcal{P}(R \to R) = 4/5$; $\mathcal{P}(R \to P) = 1/5$; $\mathcal{P}(P \to P) = 1/3$; $\mathcal{P}(P \to R) = 2/3$.
- emission probabilities:
  $E(r|P) = 1/3$; $E(p|P) = 2/3$; $E(p|R) = 1/4$; $E(r|R) = 3/4$.
- The states of the Markov Chain states: $s_1 = R$; $s_2 = P$. (actual animals)
- The observations: $o_1 = r$; $o_2 = p$. (observations you make, i.e., what you believe that you have seen)
- The initialisation and the recursion:

$$L(1,k) = \pi(s_k) \cdot E(o_1|s_k) \tag{1}$$

$$L(i,k) = \max_{m \in S} \left(L(i-1,m)\, \mathcal{P}(s_m \to s_k)\, E(o_i|s_k) \right); \tag{2}$$

$$\sigma(i,k) = \arg\max_{m \in S} \left(L(i-1,m)\, \mathcal{P}(s_m \to s_k)\ E(o_i|s_k) \right). \tag{3}$$

Here $\sigma(i,k)$ stores the value of m for which $L(i-1,m)\,\mathcal{P}(s_m \to s_k)\,E(o_i|s_k)$ is the largest which allows us to reconstruct the sequence of states which maximise the probabilities we are tracking. We now obtain the solution for our original problem by backtracking:

$$x_T = \arg\max_{m \in S}(L(T,m))$$

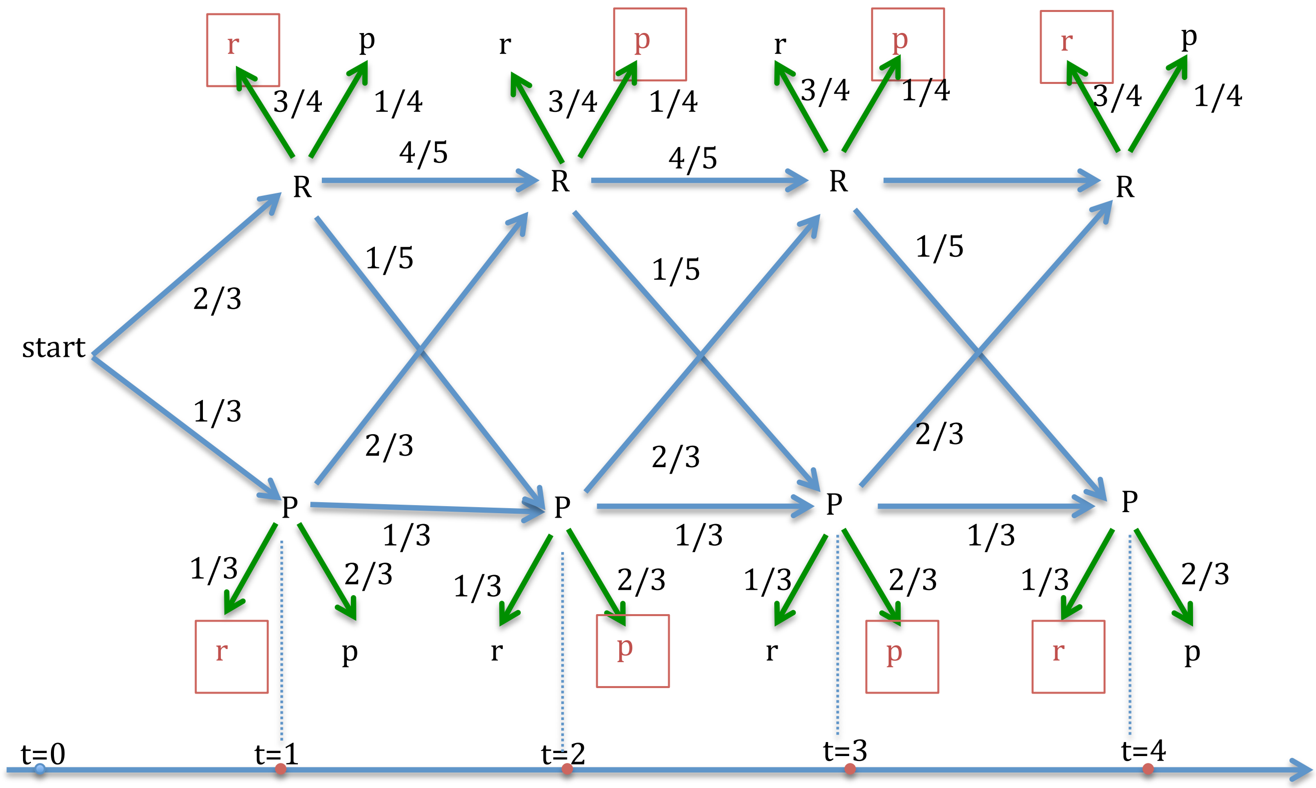
$$x_{i-1} = \sigma(i, x_i), \quad i = T, T-1, \ldots, 2.$$

Initialisation $t = 1$ with our first observation $o_1 = r$, using 1 applied to state $s_1 = R$ and then state $s_2 = P$:

$$L(1,R) = \pi(s_1)E(o_1|s_1) = \pi(R)E(r|R) = 2/3 \times 3/4 = 1/2;$$

$$L(1,P) = \pi(s_2)E(o_1|s_2) = \pi(P)E(r|P) = 1/3 \times 1/3 = 1/9.$$

Note that, generally, likelihoods do not define probability spaces, because they need not sum up to 1!



Observed emissions: r p p r

$$
\begin{aligned}
L(2,R) &= \max\{L(1,R)\,\mathcal{P}(R\to R)\,E(p|R),\;\; L(1,P)\,\mathcal{P}(P\to R)\,E(p|R)\}\\
&= \max\{1/2\times 4/5\times 1/4,\;\; 1/9\times 2/3\times 1/4\} = \max\{1/10,\; 1/54\} = 1/10;\\
\sigma(2,R) &= R;\\
L(2,P) &= \max\{L(1,R)\,\mathcal{P}(R\to P)\,E(p|P),\;\; L(1,P)\,\mathcal{P}(P\to P)\,E(p|P)\}\\
&= \max\{1/2\times 1/5\times 2/3,\;\; 1/9\times 1/3\times 2/3\} = \max\{1/15,\; 2/81\} = 1/15;\\
\sigma(2,P) &= R;
\end{aligned}
$$

$$
\begin{aligned}
L(3,R) &= \max\{L(2,R)\,\mathcal{P}(R\to R)\,E(p|R),\;\; L(2,P)\,\mathcal{P}(P\to R)\,E(p|R)\}\\
&= \max\{1/10\times 4/5\times 1/4,\;\; 1/15\times 2/3\times 1/4\} = \max\{1/50,\; 1/90\} = 1/50;\\
\sigma(3,R) &= R;\\
L(3,P) &= \max\{L(2,R)\,\mathcal{P}(R\to P)\,E(p|P),\;\; L(2,P)\,\mathcal{P}(P\to P)\,E(p|P)\}\\
&= \max\{1/10\times 1/5\times 2/3,\;\; 1/15\times 1/3\times 2/3\} = \max\{1/75,\; 2/135\} = 2/135;\\
\sigma(3,P) &= P;
\end{aligned}
$$

$$
\begin{aligned}
L(4,R) &= \max\{L(3,R)\,\mathcal{P}(R\to R)\,E(r|R),\;\; L(3,P)\,\mathcal{P}(P\to R)\,E(r|R)\}\\
&= \max\{1/50\times 4/5\times 3/4,\;\; 2/135\times 2/3\times 3/4\} = \max\{3/250,\; 1/135\} = 3/250;\\
\sigma(4,R) &= R;\\
L(4,P) &= \max\{L(3,R)\,\mathcal{P}(R\to P)\,E(r|P),\;\; L(3,P)\,\mathcal{P}(P\to P)\,E(r|P)\}\\
&= \max\{1/50\times 1/5\times 1/3,\;\; 2/135\times 1/3\times 1/3\} = \max\{1/750,\; 2/1215\} = 2/1215;\\
\sigma(4,P) &= P.
\end{aligned}
$$

Since $L(4,R) > L(4,P)$ we conclude that the event ends with R,. and this events' likelihood is $3/250 \approx 0.012$. To retrieve the whole most likely sequence we now use σ function to backtrack. Since $\sigma(4,R) = R$;

$\sigma(3, R) = R$, $\sigma(2, R) = R$ we conclude that the most likely sequence is actually $RRRR$, despite the fact that our observations were $rppr$!

Let us now calculate the likelihood that it was sequence $RPPR$ which caused our observations $rppr$:

$$\underbrace{\pi(R)\mathcal{P}(R \to P)\mathcal{P}(P \to P)\mathcal{P}(P \to R)}_{\text{likelihood of run } RPPR} \quad \times \quad \underbrace{E(r|R)E(p|P)E(p|P)E(r|R)}_{\text{likelihood of emissions } rppr \text{ for run RPPR}} \quad =$$

$$= (2/3 \times 1/5 \times 1/3 \times 2/3) \times (3/4 \times 2/3 \times 2/3 \times 3/4)$$

$$= 4/135 \times 1/4$$

$$= 1/135 \approx 0.0074$$

Thus, it is more likely that the sequence $RRRR$ caused the sequence of observations $rppr$ (likelihood ≈ 0.012) then that the sequence $RPPR$ caused the same observations (likelihood ≈ 0.0074). Without the Viterbi algorithm we would conclude that the sequence is $RPPR$ which is likely incorrect. So you can appreciate why this algorithm is a powerful tool for solving many problems related to noisy signals and thus noisy observations, such as for speech recognition, decoding convolutional codes in telecommunications and in many other applications in diverse fields such as bioinformatics and many others.