Discrete Markov models

Harald Nautsch

Markov sources

Assume that we have a source that gives a sequence $\{x_n\}$ as output. If the conditional probabilities for the outcome x_n at time n only depend on the previous k outcomes

$$p(x_n|x_{n-1}x_{n-2}x_{n-3}\ldots) = p(x_n|x_{n-1}\ldots x_{n-k})$$

the source is said to be a *Markov source of order* k. That means that a Markov source has a limited memory k steps back in time.

If the alphabet has the size N the Markov source can be described by a state model, where we have N^k states $(x_{n-1}\ldots x_{n-k})$ and where we go from state $(x_{n-1}\ldots x_{n-k})$ to state $(x_n\ldots x_{n-k+1})$ with the probability $p(x_n|x_{n-1}\ldots x_{n-k})$. These probabilities are called *transition probabilities*. We can call the states s_i, where $i = 1\ldots N^k$.

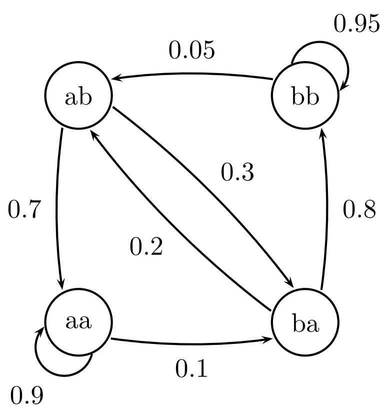
Example

Suppose we have a Markov source of order 2 with the alphabet $\mathcal{A} = \{a, b\}$ and the transition probabilities

$$p(a|aa) = 0.9\ ;\ p(b|aa) = 0.1\ ;\ p(a|ab) = 0.7\ ;\ p(b|ab) = 0.3$$

$$p(a|ba) = 0.2\ ;\ p(b|ba) = 0.8\ ;\ p(a|bb) = 0.05\ ;\ p(b|bb) = 0.95$$

Note that the transition probabilities for each state must sum to 1. We can draw the Markov source as the graph below:



□

Stationary distribution

We want to be able to calculate the probability that we at any give time are standing in a certain state, the *stationary distribution*.

The Markov source can be described by its *transition matrix* $\mathbf{P}$. This matrix contains on row i and column j the probability of going from state s_i to state s_j.

Example, cont.

For our example source we have, setting $s_1 = aa, s_2 = ab, s_3 = ba$ and $s_4 = bb$:

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0 & 0.1 & 0 \\ 0.7 & 0 & 0.3 & 0 \\ 0 & 0.2 & 0 & 0.8 \\ 0 & 0.05 & 0 & 0.95 \end{pmatrix}$$

□

Suppose that we at time n are standing in state s_i with the probability p_i^n. The distribution at time $n+1$ can then be calculated as

$$[p_1^{n+1}\ p_2^{n+1}\ \ldots\ p_{N^k}^{n+1}] = [p_1^n\ p_2^n\ \ldots\ p_{N^k}^n] \cdot \mathbf{P}$$

If we now let n approach infinity, we get the stationary distribution. We denote the probability of standing in state s_i as w_i and the stationary distribution as $\bar{w} = [w_1\ w_2\ \ldots\ w_{N^k}]$. The stationary distribution can then be found by solving the equation system

$$\bar{w} = \bar{w} \cdot \mathbf{P}$$

or

$$\bar{w} \cdot (\mathbf{P} - \mathbf{I}) = \bar{0}$$

This equation system is underdetermined. To be able to solve it, we have to remove one of the equations and replace it with $\sum_{i=1}^{N^k} w_i = 1$ (since w_i are probabilities, they must sum to 1).

Example, cont.

Solve

$$\bar{w} \cdot (\mathbf{P} - \mathbf{I}) = \bar{w} \cdot \begin{pmatrix} -0.1 & 0 & 0.1 & 0 \\ 0.7 & -1 & 0.3 & 0 \\ 0 & 0.2 & -1 & 0.8 \\ 0 & 0.05 & 0 & -0.05 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \end{pmatrix}$$

Replace the third equation with $\sum_{i=1}^{N^k} w_i = 1$ (it doesn't matter which equation we choose). So instead we solve the equation system

$$\bar{w} \cdot \begin{pmatrix} -0.1 & 0 & 1 & 0 \\ 0.7 & -1 & 1 & 0 \\ 0 & 0.2 & 1 & 0.8 \\ 0 & 0.05 & 1 & -0.05 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\Longrightarrow \quad \bar{w} = \begin{pmatrix} 0.28 & 0.04 & 0.04 & 0.64 \end{pmatrix}$$

I.e., we're standing in state *aa* with the probability 0.28, in state *ab* or *ba* with the probability 0.04 each and in state *bb* with the probability 0.64. □

Probabilities of symbol sequences

By calculating the probabilities for the different states we have also calculated the probabilities for k-tuples of source symbols. From this distribution we can of course calculate the probabilities for shorter sequences by calculating marginal distributions, i.e. if we want to know the probabilities for n-tuples, $n < k$ we just sum the probabilities where the first n symbols are the same.

Example, cont.

$$p(a) = p(aa) + p(ab) = 0.28 + 0.04 = 0.32$$

$$p(b) = p(ba) + p(bb) = 0.04 + 0.64 = 0.68$$

□

We can also calculate the probabilities for longer symbol sequences by using the transition probabilities. For example, for triplets we get
$p(x_{i+2}x_{i+1}x_i) = p(x_{i+1}x_i) \cdot p(x_{i+2}|x_{i+1}x_i)$.

Example, cont.

$$\begin{aligned}
p(aaa) &= p(aa) \cdot p(a|aa) = 0.28 \cdot 0.9 = 0.252 \\
p(aab) &= p(ab) \cdot p(a|ab) = 0.04 \cdot 0.7 = 0.028 \\
p(aba) &= p(ba) \cdot p(a|ba) = 0.04 \cdot 0.2 = 0.008 \\
p(abb) &= p(bb) \cdot p(a|bb) = 0.64 \cdot 0.05 = 0.032 \\
p(baa) &= p(aa) \cdot p(b|aa) = 0.28 \cdot 0.1 = 0.028 \\
p(bab) &= p(ab) \cdot p(b|ab) = 0.04 \cdot 0.3 = 0.012 \\
p(bba) &= p(ba) \cdot p(b|ba) = 0.04 \cdot 0.8 = 0.032 \\
p(bbb) &= p(bb) \cdot p(b|bb) = 0.64 \cdot 0.95 = 0.608
\end{aligned}$$

□

Entropy rates for Markov sources

The entropy rate of a Markov source is given by

$$\lim_{n \to \infty} H(X_n|X_1X_2 \ldots X_{n-1}) = H(X_n|X_{n-1} \ldots X_{n-k})$$

because of the limited memory. The entropy rate can also be calculated as the average of the entropies for the different states.

$$\sum_{j=1}^{N^k} w_j \cdot H(S_{i+1}|S_i = s_j)$$

S_i is a random process that describes the state sequence. $H(S_{i+1}|S_i = s_j)$ is the entropy of the transition probabilities in state s_j.

Example, cont.

$$
\begin{aligned}
H(X_n|X_{n-1}\ldots X_{n-k}) &= -0.252 \cdot \log 0.9 - 0.028 \cdot \log 0.7 \\
&\quad -0.008 \cdot \log 0.2 - 0.032 \cdot \log 0.05 \\
&\quad -0.028 \cdot \log 0.1 - 0.012 \cdot \log 0.3 \\
&\quad -0.032 \cdot \log 0.8 - 0.608 \cdot \log 0.95 \\
&\approx \underline{0.3787}
\end{aligned}
$$

$$
\begin{aligned}
\sum_{j=1}^{N^k} w_j \cdot H(S_{i+1}|S_i = s_j) &= 0.28 \cdot (-0.1 \cdot \log 0.1 - 0.9 \cdot \log 0.9) + \\
&\quad 0.04 \cdot (-0.7 \cdot \log 0.7 - 0.3 \cdot \log 0.3) + \\
&\quad 0.04 \cdot (-0.2 \cdot \log 0.2 - 0.8 \cdot \log 0.8) + \\
&\quad 0.64 \cdot (-0.05 \cdot \log 0.05 - 0.95 \cdot \log 0.95) \approx \\
&\approx \underline{0.3787}
\end{aligned}
$$

□