Coding for Constrained Markovian Sources

Course of Stochastic Processes and Performance Modeling - 2021

Stefano Della Fiore

1 Introduction

Definition 1.1. *An information source* $\boldsymbol{X}$ *is an infinite sequence of random variables* $X_1, X_2, X_3, ...$ *taking values in a finite alphabet* $\mathcal{X}$. *The source* $\boldsymbol{X}$ *is said to be memoryless if* $X_1, X_2, ...$ *are independent and identically distributed (i.i.d.). So we say that the source has memory if the random variables* $X_1, ..., X_n$ *are not independent.*

Definition 1.2. *A variable-length code for a random variable* X *is a map from the alphabet* $\mathcal{X}$ *to* $\mathcal{D}^*$, *where* $\mathcal{D}^*$ *is the set of finite length sequences of symbols from a D-ary alphabet. For each possible symbol* $x \in \mathcal{X}$ *the codeword associated to* x *is* $C(x)$ *and with* $l(x)$ *we identify its length.*

Definition 1.3. *A code is said to be non-singular if:*

$$\forall x_i, x_j \in \mathcal{X},\ \ x_i \neq x_j \ \textit{implies}\ C(x_i) \neq C(x_j). \tag{1}$$

Definition 1.4. *The extension* C^* *of a code* C *is the mapping from finite length sequences of* $\mathcal{X}$ *to finite length sequences of* $\mathcal{D}$ *defined by:*

$$C(x_1 x_2 ... x_n) = C(x_1)C(x_2)...C(x_n) \tag{2}$$

where the strings represent the concatenation of different source symbols and coding symbols.

Definition 1.5. *In the classic sense a code* C *is said uniquely decodable if its extension is non-singular.*

Definition (1.5) makes the assumption that all the combinations of symbols can be produced by the source with positive probability. It means that if we represent the source with a first-order Markov chain the transition probability matrix P has all its entries positive meaning that all the transitions between different states are always possible. When we are in this configuration (when all combinations of symbols are possible) then we call these sources: ***unconstrained sources***; at the other end if some combinations are impossible to obtain we call these sources: ***constrained sources***.

Definition 1.6. *Let $\boldsymbol{X}$ be an information source with alphabet $\mathcal{X}$. A code C is said to be uniquely decodable for the source $\boldsymbol{X}$ if no two different finite sequences of source symbols producible by $\boldsymbol{X}$ have the same codeword.*

With this definition (1.6) we have a complete coverage of all the possible sequences of symbols outcome from a source. All the uniquely decodable sources in the "*classic sense*" and also codes for constrained sources are incorporated in the new definition.

Definition 1.7. *A code is called a prefix-code (prefix-free) if no codeword is a prefix of any other codeword.*

Under prefix-free conditions the code is said to be *instantaneous*: this means that the decoder can decode the received messages in linear time. Given a sequence of code symbols the decoder can identify a unique sequence of the received symbols in order to reconstruct the message correctly.

Theorem 1.8 (Kraft Inequality)**.** *Let $l_i, i = 1, ..., n,$ be the lengths of the codewords of a prefix code and let $D = |\mathcal{D}|$ be the size of the code alphabet $\mathcal{D}$. It states that*

$$\sum_{i=1}^{n} D^{-l_i} \leq 1. \tag{3}$$

Conversely, if a set of integers $l_i, i = 1, ..., n$ satisfy the Kraft inequality then a prefix-free code can be constructed with those codeword lengths.

Theorem 1.9. *It follows from the Kraft inequality that if a prefix code is used for encoding a random variable X, then the expected length of the codeword generated is always greater or equal to the entropy of X.*

$$E[l(X)] \geq H(X). \tag{4}$$

Corollary 1.10. *For every prefix code the expected length of the code for n symbols of a source $\boldsymbol{X}$ satisfies:*

$$E[l(X_1, X_2, ..., X_n)] \geq H(X_1, X_2, ..., X_n) \tag{5}$$

This corollary on prefix-code is a strong result with respect to the asymptotic lower bound given by Shannon (infinite number of symbols) because it is applied to finite sequences of symbols.

In the previous Theorem we consider prefix-codes or non-constrained uniquely decodable codes. For these codes McMillan [4] proved that both uniquely decodable and prefix-free codes satisfy the Kraft inequality, meaning that there is no advantage in using a uniquely decodable code instead of a prefix-free because both have the same lower bound on the expected length seen in Corollary 1.10. Using a uniquely decodable code only makes the decoder more complex. As we said before there are some sources that do not produce some of the possible sequences of symbols; in these cases the Kraft inequality does not hold anymore [2] and this brought to a new theorem (Theorem 2.1).

2 Constrained Markovian Source

Theorem 2.1. *(Dalai [3]) There exists at least one source* $\boldsymbol{X} = (X_1, X_2, ...,$ $X_n)$ *and a uniquely decodable code for* $\boldsymbol{X}$ *such that, for every* $n \geq 1$*:*

$$E[l(X_1, X_2, ..., X_n)] < H(X_1, X_2, ..., X_n) \tag{6}$$

Proof. Let us see an example with a first order Markovian source, considering a source $\mathbf{X}$ generating symbols $X_1, X_2, ...$ where each $X_i \in \mathcal{X} = \{A, B, C, D\}$. The sequence of generating symbols is based on the following transition probability graph:

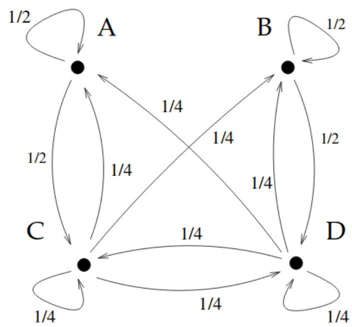

Figure 1: *Graph related to a Markov source with some impossible transitions.*

The associated transition probability matrix is

$$\mathbf{P} = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}. \tag{7}$$

It can be easily shown that the unique stationary distribution is the uniform distribution over the set of source symbols, that is $\boldsymbol{\pi} = (1/4, 1/4, 1/4, 1/4)$.

If we set the distribution $P_{X_1}(x_1) = \boldsymbol{\pi}$ then the source is stationary. The Markov source is also irreducible and aperiodic and so the source is ergodic.

We consider a classic encoding algorithm (Huffman) that generates the optimal code in the classic sense for a sequence of symbols. In this particular case all the entries of $\mathbf{P}$ are powers of the coding dictionary cardinality $|\mathcal{D}| = 2$ and this implies that the $H(X_1^n)$ can be reached with a prefix-free code. The Huffman code is constructed by encoding the first symbol independently and all the successive symbols are encoded using the transition probability matrix $\mathbf{P}$ creating different Huffman codes for all the $\mathbf{P}$-rows.

Example 2.2. *An Huffman code for the stationary Markov chain based on the transition probability matrix $\boldsymbol{P}$ in eq.* (7) *can be constructed as follows*

$$\begin{aligned} &P_{X_1}(x_1) = (1/4, 1/4, 1/4, 1/4), \\ &\mathcal{C}_1 = (A \to 00, B \to 01, C \to 10, D \to 11), \end{aligned}$$

where $\mathcal{C}_1$ is the Huffman code for the first symbol. Then for all the possible state-transitions an Huffman code is constructed:

$$\begin{aligned} &P_{X_2|X_1}(x_2|x_1 = A) = (1/2, 0, 1/2, 0), \\ &\mathcal{C}_A = (A \to 0, -, C \to 1, -), \end{aligned}$$

where $\mathcal{C}_A$ is the Huffman code related to transitions which start from symbol A. With the same procedure we carry out

$$\begin{aligned} &\mathcal{C}_B = (-, B \to 0, -, D \to 1), \\ &\mathcal{C}_C = (A \to 00, B \to 01, C \to 10, D \to 11), \\ &\mathcal{C}_D = (A \to 00, B \to 01, C \to 10, D \to 11), \end{aligned}$$

which are the codes for the remaining state-transitions.

The expected length of the Huffman code for the first n symbols reaches the entropy of the sequence which can be expressed as

$$\begin{aligned} H(X_1^n) &= H(X_1) + \sum_{i=2}^{n} H(X_i|X_1^{i-1}) \\ &= H(X_1) + (n-1)H(X_2|X_1) \\ &= 2 + \frac{3}{2}(n-1), \end{aligned}$$

for all $n \geq 1$. This result is achieved using the chain rule of the entropy and the fact that the source is stationary.

As we said before since the entries of the $\mathbf{P}$ matrix are power of 2 and $P_{X_1}(x_1) = (1/4, 1/4, 1/4, 1/4)$, then the expected length of the codeword equals the entropy of the sequence:

$$E[l(X_1^n)] = H(X_1^n) = 2 + \frac{3}{2}(n-1). \tag{8}$$

In [2], a different code is proposed: a fixed map from $\mathcal{X} \to \mathcal{D}^*$ such that $A \to 0$, $B \to 1$, $C \to 01$, $D \to 10$. This code is not uniquely decodable in the classic sense because the concatenation BA produces 01 that is also produced by symbol D. The advantage of this coding scheme is that the impossible transitions are exploited. In fact, the transition $A \to B$ and vice versa are not allowed by the Markovian source, so this encoding scheme does not produce ambiguity and the source sequence can be correctly reconstructed at the decoder (making it uniquely decodable code in this sense).

Evaluating the expected length of the code, we see that:

$$E[l(X_1^n)] = \sum_{i=1}^{n} E[l(X_i)] = \frac{3}{2}n. \tag{9}$$

$\square$

Thinking about the asymptotic equipartition (AEP) for ergodic sources (McMillan) shown in Gallager's book [1], intuitively, the minimum expected length per symbol for both constrained and unconstrained Makovian ergodic sources is the entropy rate of the source. So a natural conjecture arises.

Conjecture 2.3. *Given a stationary Markovian source* $\mathbf{X} = X_1, X_2, \ldots, X_n$*, then for every uniquely decodable code (that works with concatenation) for the source* $\mathbf{X}$*, we have that*

$$E[l(X_1^n)] \geq nH(\mathcal{X}) \text{ for all } n \geq 1,$$

where $H(\mathcal{X}) = \lim_{n\to\infty} H(X_1, X_2, \ldots, X_n)/n$*.*

With the fix-mapped code the gain obtained with respect to the Huffman code is only at the first symbol because $H(X_2|X_1) = 3/2$ and so, there is no gain when we have state transitions. The Huffman code is more expensive in terms of computational complexity because the decoder must have stored all the codes for each possible transition, while for the custom-code the decoder needs to know the impossible transitions but the matching between encoding bits and symbols is faster.

Proving or disproving the Conjecture 2.3 for all stationary markovian sources is not an easy task, so we restrict our attention to the following problem.

Problem 2.4. *Given the stationary markovian source* $\mathbf{X} = X_1, \ldots, X_n$*, described by the transition probability matrix* $\mathbf{P}$ *given in eq. (7), prove or disprove there exists a map* $\mathcal{C} : \mathcal{X}^* \to \{0,1\}^*$ *from sequences of letters (of any length), where* $\mathcal{X} = \{A, B, C, D\}$*, to sequences of bits (of any length) such that:*

1. $\mathcal{C}$ *is invertible;*

2. $E[l(\mathcal{C}(X_1, X_2, \ldots, X_n))] \le 3/2n$ *for all* $n \ge 1$*;*

3. $E[l(\mathcal{C}(X_1, X_2, \ldots, X_n))] < 3/2n$ *for at least one* $n \ge 1$*.*

Problem 2.4 raises the question of weather it exists a code that performs better, in terms of expected length, than the one given in Theorem 2.1. We believe that such code does not exist and to corroborate this claim we show that using a greedy code constructed in the next paragraph we cannot satisfy all the hypotheses in Problem 2.4.

We construct iteratively (in a greedy way) the code for the stationary markovian source described by $\mathbf{P}$ in eq. (7) as follows.

- For $n = 1$, using the distribution of the first symbol $P_{X_1}(x_1) = (1/4, 1/4, 1/4, 1/4)$ we encode the source output sequences of length 1, that are, $\{A, B, C, D\}$. We are constrained to use 2 codewords of 1 bit and 2 codewords of 2 bits. Then

$$E[l(\mathcal{C}(X_1))] = 3/2.$$

- For $n = 2$, we have the following sequences $\{AA, AC, BB, BD, CA, CB, CC, CD, DA, DB, DC, DD\}$ with $P_{X_1,X_2}(x_1, x_2) = (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16})$. In this case we assign the remaining 2 codewords of 2 bits to the sequences $\{AA, AC\}$, then all the 8 codewords of 3 bits to $\{BB, BD, CA, CB, CC, CD, DA, DB\}$ and finally 2 codewords of 4 bits to $\{DC, DD\}$. Then

$$E[l(\mathcal{C}(X_1, X_2))] = 23/8 < 3/2 \cdot 2 = 3.$$

- $\cdots$

So, iteratively, at step n we assign the shortest unused (in the first $n-1$ steps) sequences of bits to the source output sequences following a descending probability order. It can be easily shown that at each step k ($k \ge 1$) we have to encode $4 \cdot 3^{k-1}$ sequences produced by the markovian source under consideration. Then, after n steps we have encoded

$$\sum_{k=1}^{n} 4 \cdot 3^{k-1} = 2 \cdot (3^n - 1)$$

sequences. This implies that at each step n we are using all the sequences of bits of length smaller than $\lfloor \log_2(3^n - 1) \rfloor - 1$ to encode our source output sequences. Then, for n large enough it can be easily verified that the expected length of the iterative code increases as $n \log_2 3$. Therefore, there exists an n for which the expected length of the code is strictly greater than $3/2n$ since $\log_2 3 > 3/2$. So, the second hypothesis in Problem 2.4 is not satisfied.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication. Wiley, New York, 1968.*

[2] M. Dalai and R. Leonardi, *On Unique Decodability, McMillan's Theorem and the Expected Length of Codes, Technical Report R.T., 2008.*

[3] M. Dalai, *New Techniques for Signal Representation and Coding, Ph.d thesis, 2006.*

[4] B. McMillan, *Two inequalities implied by unique decipherability. IEEE Trans. Inform.Theory, 1956.*