ANKARA YILDIRIM BEYAZIT UNIVERSITY

FACULTY OF ENGINEERING AND NATURAL SCIENCE

ELECTRICAL & ELECTRONICS ENGINEERING

EEE 306 TELECOMMUNICATIONS II

“A Method for the Construction of Minimum-Redundancy Codes”

Lecturer: Dr. Mustafa Anıl Reşat

Final

Anıl Erman

16050511039

JUNE 2021

ANKARA

Transmitting a signal a message is as important today as it was in 1952. While transmitting a message, both the receiver and the transmitter must do this job in the fastest way with the least error. In this research, I examined a method that was introduced in the 1950's. And DAVID A. HUFFMAN proposed a method to convey this problem. For the transmission of messages, Huffman put forward the idea that instead of transmitting the symbol sequences to describe the message. Some messages may need to use more than one symbol if there are many types of symbols to convey the message at hand. Assuming that it takes the same time of a message is directly proportional to the symbol lenght of that message.

According to Huffman, the symbol or set of symbols associated with a given message will be called a "message code". The entire number of messages that can be forwarded. will be reffered to as the "message ensemble". The mutual agreement between the transmitter and receiver about the meaning of the code for each messages of the community is called the "ensemble code". The mutual agreement between the transmitter and receiver about the meaning of the code for each message is called the "ensemble code".

The number of messages is called N. P(i) i. is the probability of message. And the sum of all these probabilities gives the result 1.

$$\sum_{i=1}^{N} P(i) = 1$$

The lenght of a message, $L(i)$, is the number of encoding digits. assigned to it. Therefore, the average message lenght:

$$L_{av} = \sum_{i=1}^{N} P(i) \cdot L(i)$$

It is expressed by the formula.

With this method, it will be defined as an ensemble code that gives the lowest possible average message lenght for a message ensemble consisting of a finite number of N and a certain number of encoding digits.

The following basic restrictions should apply to a message code:

- Either two messages will consist of the same encoding digits.
- Message code, once the starting point of a message thread is known, no additional indication will be required to indicate where a message code begins and ends.
- Either message will be set as the first part of any message code of greater lenght.

To explain this:

In order to transmit a message consisting of A-B-C-D characters, the frequency of using these characters in the message to be sent must be calculated first. We will cover this in depth later in the article. Now back to the explanation. After the frequency is calculated, a symbol must be assigned for the characters A B C D. We will hypothetically specify a random symbol.

A = 01

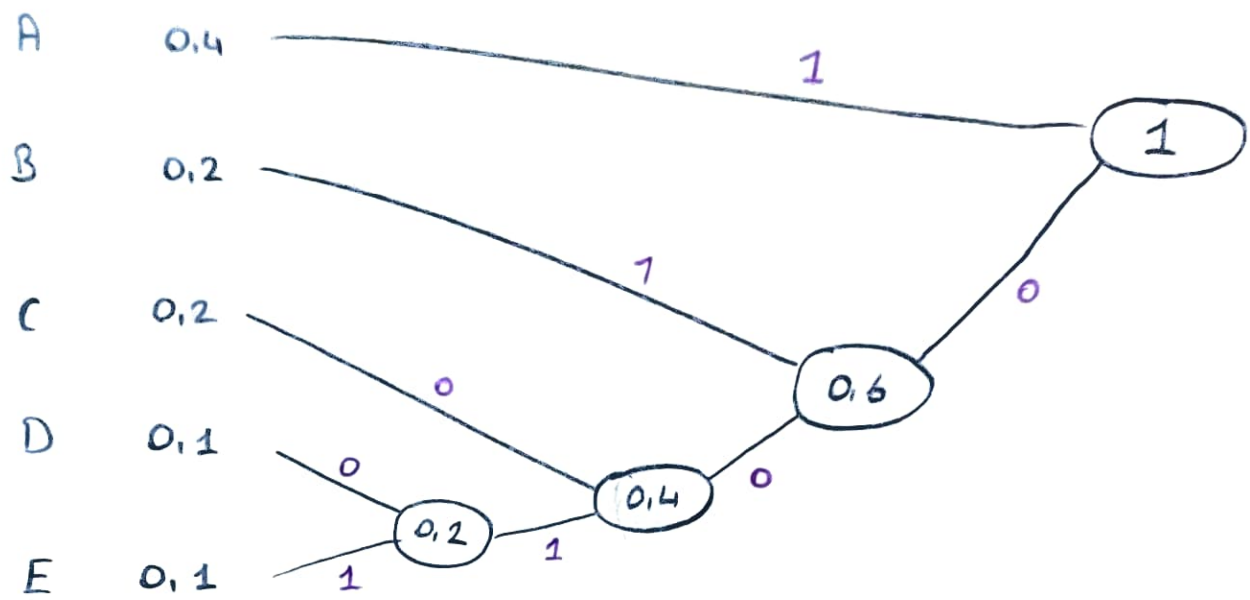
B = 102

C = 111

D = 202

Let is defined it. Let our message be "ABACCD". As a result, 01 1 02 0 1 1 1 1 1 1 202 will be the message to be transmitted. The only thing the recipient needs to know for this message is the ensemble code. The receiver can use this ensemble code to split the message into 01 - 102 - 01 - 111 - - 11 - 202 parts and go backwords. As a result, it reaches the "ABACCD" message again.

One of the most important things in this article is for an optimal code, the lenght of a given message code can never be less than the lenght of a more likely message code. Suppose an optimum code in wich neither of the two messages encoded with lenght $L(N)$ has the same initial signal of sequence $L(N)-1$. For an optimal code, none of these messages of lenght $L(N)$ can have the beginning of any code corresponding to other codes. In this case, the last digit of the entire massage group becomes significant.



I will try to explain this technique with an example so that it is easier to understand. We rank the probability (frequency) of the known message from most likely to least likely. The bottom two probabilities are added together and a balloon is made and according to Shannon, 0 is written on the top and 1 on the bottom. Then the bubble is added with the other possibility (C). Write 0 on the top and 1 on the bottom. Again, with another probability (B), the bubble is added and then the one with the highest probability, that is, 0 is written on the top, and the one with the lowest probability, that is, 1 on the bottom

Then with the other possibility (A), if the bubbles are added, 1 is obtained.

Now we can assign the symbols of the characters in the message to be transmitted. For this, you should start from 1 and go towards the character. For example, 0-0-1-0 path is followed for the D character.

A → 1

B → 01

C → 000

D → 0010

E → 0011

The result is reached. And as seen above, it also provides all the restrictions. For examples, no symbol represent the beginning of another symbol. Now we can calculate the Lav value of our sample. For this, we will multiply and add the number of characters with probabilities.

$$0.4 \times 1 + 0.2 \times 2 + 0.2 \times 3 + 0.1 \times 4 + 0.1 \times 4 \approx 2.2$$

While we can transmit a 5-character message with 3 bits under normal conditions, we have now reduced this bit value to 2.2.

Now let is compare the message transmitted with these symbols and the standart message. For standart symbols we need 3 bits.

Provisions:

A → 001

B → 010

C → 011

D → 100

E → 101

★ Let our message be "ABAABCBACDE".

Standart method = 0010 1000 1001 0100 1101 0001 0111100101

This method = 1011 10 100 001 1000 0010 0011

The result of the article is, as seen in the example given, our message has been shortened, wich reduces errors and makes it easier to transmit the messages. This technique is used in many fields today. The most important of these is the data compression area. It is also used in file formats such as png, jpeg, mp3, as well as HTTP's standart compresion method.