A Mathematical Theory of Communication

By C. E. Shannon

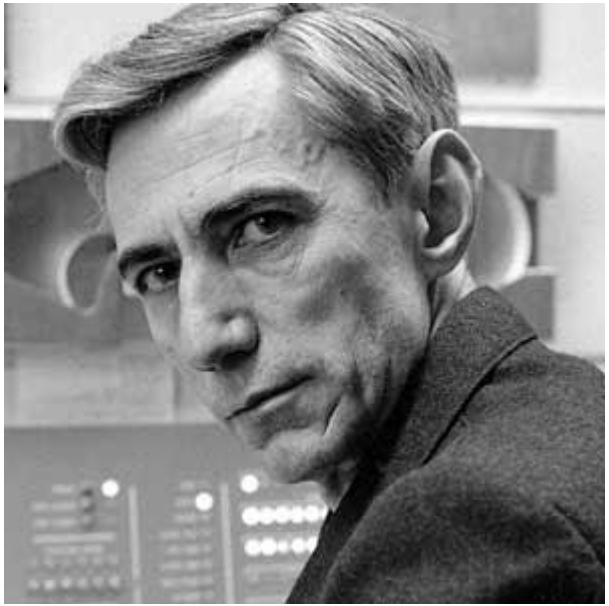
Agenda

- Introduction
  - History, Who is Shannon?
- Basic Information Theory
  - What is information?
  - How to measure it?, Entropy
- Grams
  - Stochastic Process for generating symbols / information
  - n-gram

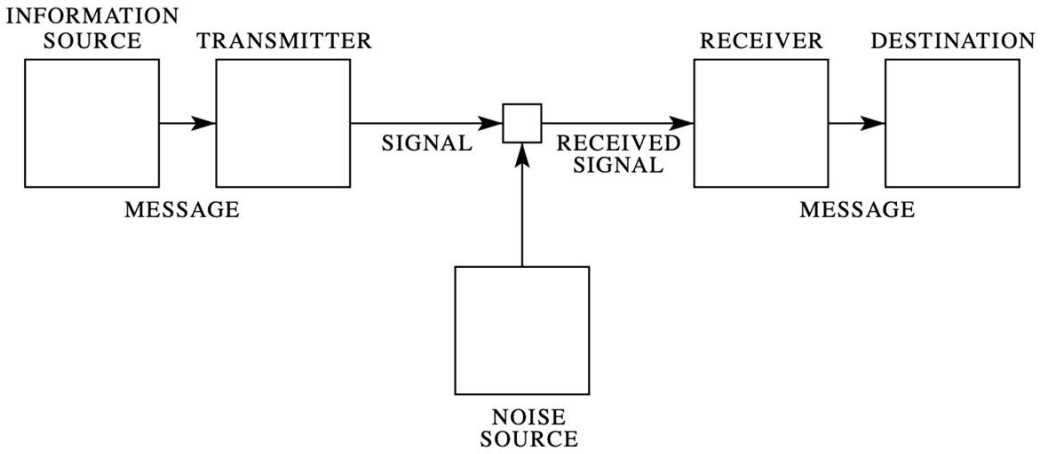
History

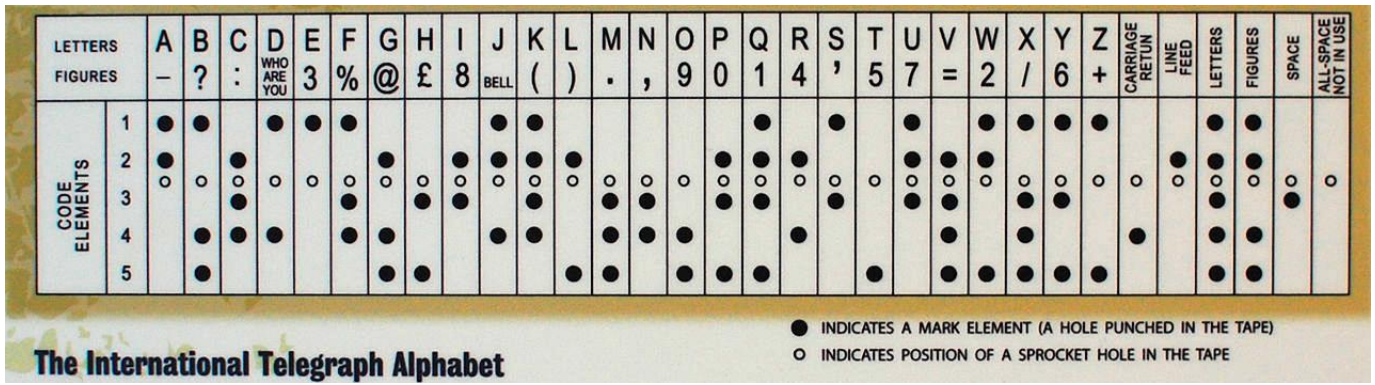
Claude Elwood Shannon, often referred to as the "father of modern digital communication and information theory," was a pioneering American mathematician, electrical engineer, and cryptographer. He made significant contributions to various fields, including information theory, cryptography, and computer science.

- Shannon's most famous work, "A Mathematical Theory of Communication" (1948), laid the foundation for information theory.



Claude Elwood Shannon


INFORMATION SOURCE
TRANSMITTER
SIGNAL
RECEIVED SIGNAL
RECEIVER
DESTINATION
MESSAGE
MESSAGE
NOISE SOURCE


LETTERS
FIGURES
CODE ELEMENTS
A –
B ?
C :
D WHO ARE YOU
E 3
F %
G @
H £
I 8
J BELL
K (
L)
M .
N ,
O 9
P 0
Q 1
R 4
S ’
T 5
U 7
V =
W 2
X /
Y 6
Z +
CARRIAGE RETUN
LINE FEED
LETTERS
FIGURES
SPACE
ALL-SPACE NOT IN USE
● INDICATES A MARK ELEMENT (A HOLE PUNCHED IN THE TAPE)
○ INDICATES POSITION OF A SPROCKET HOLE IN THE TAPE
The International Telegraph Alphabet

Information Theory

What is information ?

- Semantic meaning of a message is irrelevant to its transmission. A message should be conceived as a sequence with statistical properties. It is the message's statistics that could be captured and its coding minimized to allow for effective transmission.

- Information is measured in *bits*, and one bit of information allows you to choose between two equally alternatives.

Basic laws of information

1. Upper limit for channel capacity
2. Noise
3. Encoding of data

Measure of information

$$H = -\sum_{i}^{n} p_i log(p_i)$$

Consider a,b,c,

$$p(a) = 2/4 = 0.5$$

$$p(b) = 1/4 = 0.25$$

$$p(c) = 1/4 = 0.25$$

When, $p(c) = 1/4$,we need the following number of bits to represent them all.

$$log_2(4) = log_2(1/p) bits = 2 bits$$

Now, to determine the overall storage required

- a occurs half and need 1 bit, so 0.5*1
- b occurs quarter of the time and needs 2 bits, so 0.25*2
- c occurs quarter of the time and needs 2 bits, so 0.25*2

In total (0.5*1) + (0.25*2) + (0.25*2) = 1.5 bits

Stochastic Process for generating symbols / information

- Consider 5 letters A B C D E, let the probabilities be 0.4,0.1,0.2,0.2,0.1.

- A typical message constructed from this source is
  "AAACDCBDCEAADADACEDA E A D C A B E D A D D C E C A A A A A D."

- Transition Probabilities $p_i(j)$
  - The probability of i is followed by letter j

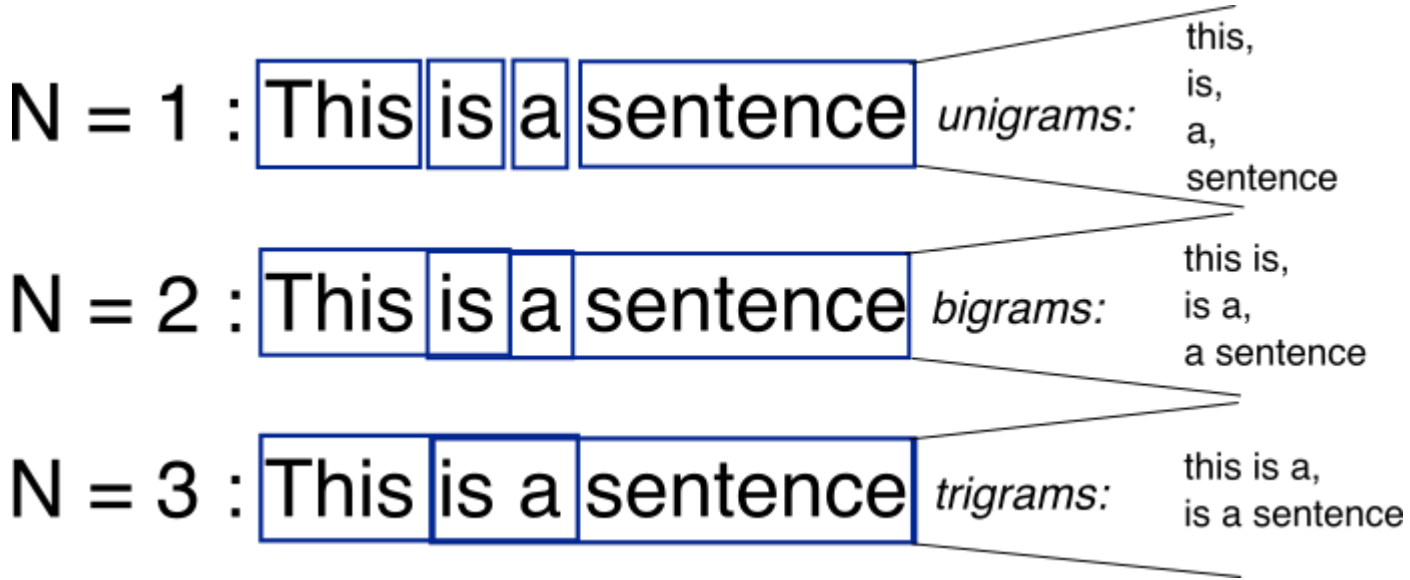


Image Source : Stackoverflow

n - gram

- n- grams is a sequence of n adjacent words or letters from the text corpus

- <s> I am Sam </s>
- <s> Sam I am </s>
- <s> I do not like green eggs and hams </s>

Here are the calculations for some of the bigram probabilities from this corpus

$P(I\|<s>) = 2/3 = 0.67$	$P(Sam\|<s>) = 1/3 = 0.33$	$P(am\|I) = 2/3 = 0.67$
$P(</s>\|Sam) = 1/2 = 0.5$	$P(Sam\|am) = 1/2 = 0.5$	$P(do\|I) = 1/3 = 0.33$

- For the general case of MLE n-gram parameter estimation:

$$P(w_n|w_{n-N+1:n-1}) = \frac{C(w_{n-N+1:n-1}\ w_n)}{C(w_{n-N+1:n-1})}$$

n - gram Applications

- Sentiment analysis
- Text classification
- Text generation

Thank You