Group Discussion: Information Theory

Maths and Physics Club, IIT Bombay

Online cause 2020, 7/11

Table of Contents

Group Discussion: Information Theory

Without *much* of the math…

Can anyone come up with a word starting with the letter Q?

Queer

Question

Quantum

Quote

Quarantine

Queen

Quick

Quality

Quiet

Quilt

Quite

Quint

If you like Scrabble: quintessentially

Another Brainteaser

Wierldy enuogh, I’m prtety srue taht you mnagaed to raed tihs whtoiut filncinhg

Myabe ecxept taht lsat wrod

What is Information?

Words

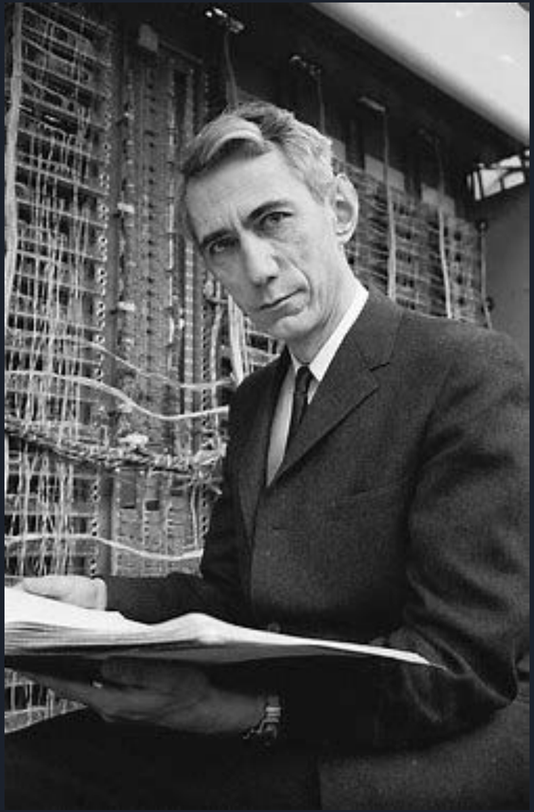
9 letter word - Gibberish

Coins

9 coin flips

The Father of Information Theory

Claude Elwood Shannon was an American mathematician, electrical engineer, and cryptographer known as "the father of information theory". Shannon is noted for having founded information theory with a landmark paper, "A Mathematical Theory of Communication", which he published in 1948



Entropy

Measure of Information in terms of Uncertainty

Higher Uncertainty implies Higher Entropy

Higher the Entropy, more the amount of Information contained in the system

_A__L_ T_N__S

7 blank spaces, 26^7 different words possible

Information present in coins?

Say, I have a double headed coin

What information does it give you?

Before I flip?

After I flip?

Quantifying Information

To know the state of a system, you need to ask questions about it

I'll throw in a constraint that you can only ask yes/no type of questions

And define entropy as the least number of questions you can ask (on average) before you know the state of a system

The more questions you are **forced** to ask, the more states the system could be in, implying, there's more uncertainty in what the state of the system could actually be

Information == Entropy?

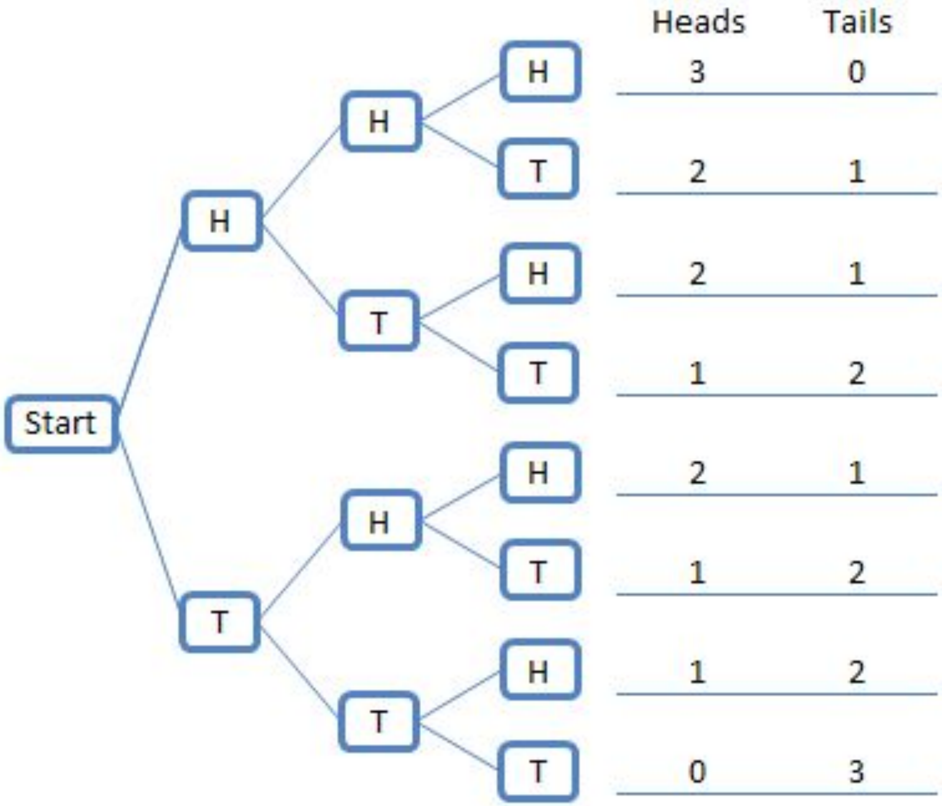
Related, but not the same

Entropy: The measure of uncertainty **before** the flip

Information: The knowledge available **after** the flip

A regular coin

... and the **BIT**



4.7 bits

For 9 letters, it will be 9 x 4.7 = 42.3 bits

If you are too into formulae: $2^{\#\text{ questions}}$ = *message space* or the possible # values each position could take

How many questions do you need to ask before you know what card has been picked from a deck?

But does it really contain that much information?

Wierldy enuogh, I’m prtety srue taht you mnagaed to raed tihs whtoiut fiIncinhg
Myabe ecxept taht lsat wrod

First line - original text

Second line - Participant guesses

Dashes represent the letters correctly guessed in the first try

[Here](#) is the link to the complete paper on "Prediction and Entropy of Printed English"

case will be $\frac{4.5}{5.5}$ or .818 of F_N for the 26-letter alphabet when N is reasonably large.

3. PREDICTION OF ENGLISH

The new method of estimating entropy exploits the fact that anyone speaking a language possesses, implicitly, an enormous knowledge of the statistics of the language. Familiarity with the words, idioms, cliches and grammar enables him to fill in missing or incorrect letters in proof-reading, or to complete an unfinished phrase in conversation. An experimental demonstration of the extent to which English is predictable can be given as follows; Select a short passage unfamiliar to the person who is to do the predicting. He is then asked to guess the first letter in the passage. If the guess is correct he is so informed, and proceeds to guess the second letter. If not, he is told the correct first letter and proceeds to his next guess. This is continued through the text. As the experiment progresses, the subject writes down th correct text up to the current point for use in predicting future letters. Th result of a typical experiment of this type is given below. Spaces were included as an additional letter, making a 27 letter alphabet. The first line is the original text; the second line contains a dash for each letter correctly guessed. In the case of incorrect guesses the correct letter is copied in th second line.

```
(1) THE ROOM WAS NOT VERY LIGHT A SMALL OBLONG
(2) ----ROO------NOT-V-----I------SM----OBL----
(1) READING LAMP ON THE DESK SHED GLOW ON
(2) REA----------O------D----SHED-GLO--O--
(1) POLISHED WOOD BUT LESS ON THE SHABBY RED CARPET
(2) P-L-S------O---BU--L-S--O-------SH-----RE--C-----
```

An example I used in a preceding slide:

_A__L_ T_N__S

Queer

Quantum

Quote

Quarantine

Question

Queen

Quick

Quality

Quiet

Quilt

Quite

Quint

HOWEASYDOYOUFINDTHISTOREAD

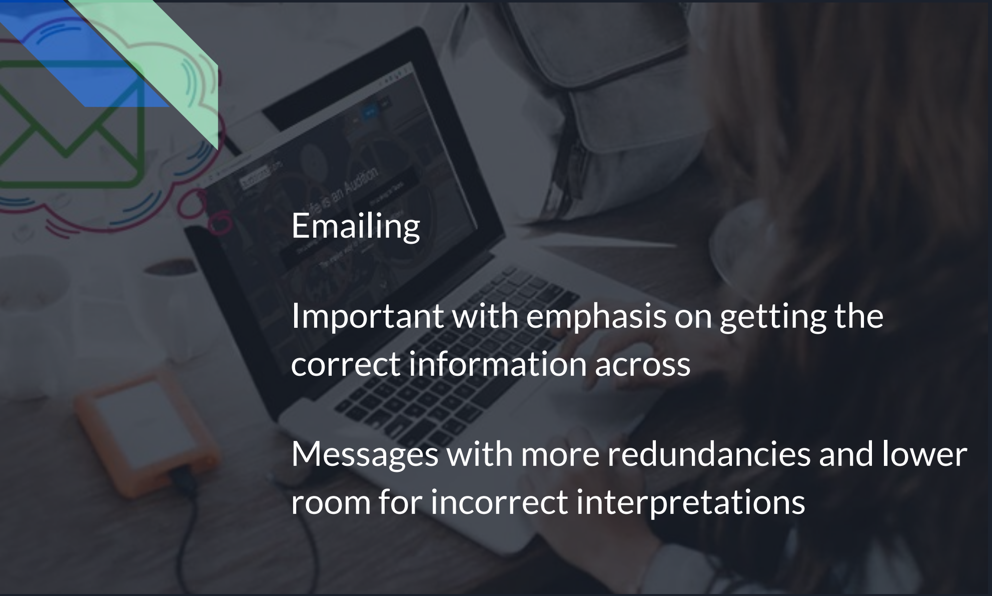
Wrdly ngh, m prtty sr y mngd t rd ths wtht flnchng

Myb xcpt tht lst wrd

WRDLYNGHMPRTTYSRYMNGDTRDTHSWTHTFLNCHNGMYBXCPTTHTLSTWRD

English hmm...

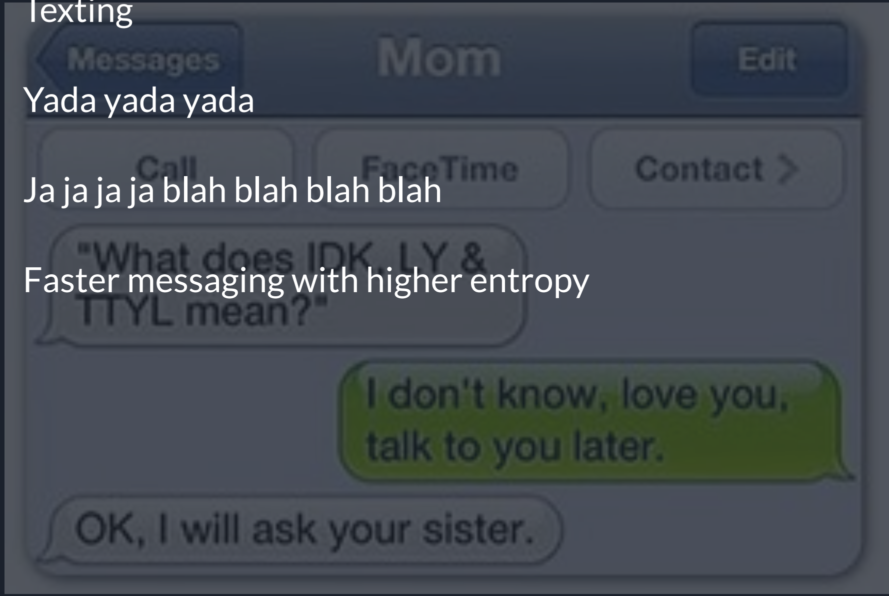
Calendar	Kalender
Licence	Lizenz
Protocol	Protokoll
Yogurt	Joghurt



Emailing

Important with emphasis on getting the correct information across

Messages with more redundancies and lower room for incorrect interpretations



Texting

Yada yada yada

Ja ja ja ja blah blah blah blah

Faster messaging with higher entropy

Also you could delve deeper into such mathematical entropy when we share the slides with you LaTeR. And do read the original paper by Claude Shannon

To summarise: Entropy is maximum when all outcomes are equally likely and anytime you move away from equally likely outcomes or introduce predictability, the entropy must go down

Data Compression

Exploiting Redundancies

As a reminder, the most random systems have the highest entropies

Exploiting Redundancies

Say, you need to transmit letters ABCD each occurring with different (but known) probabilities:

A=50%	B=12.5%	C=12.5%	D=25%

One straightforward method will be to correspond:

A=00	B=01	C=10	D=11

But this doesn't exploit the statistical fact you know about the system - that A occurs half of the time. If we could workaround by using only one bit to transmit A...

Binary Decision Diagrams come into play. In fact, they were literally invented for this

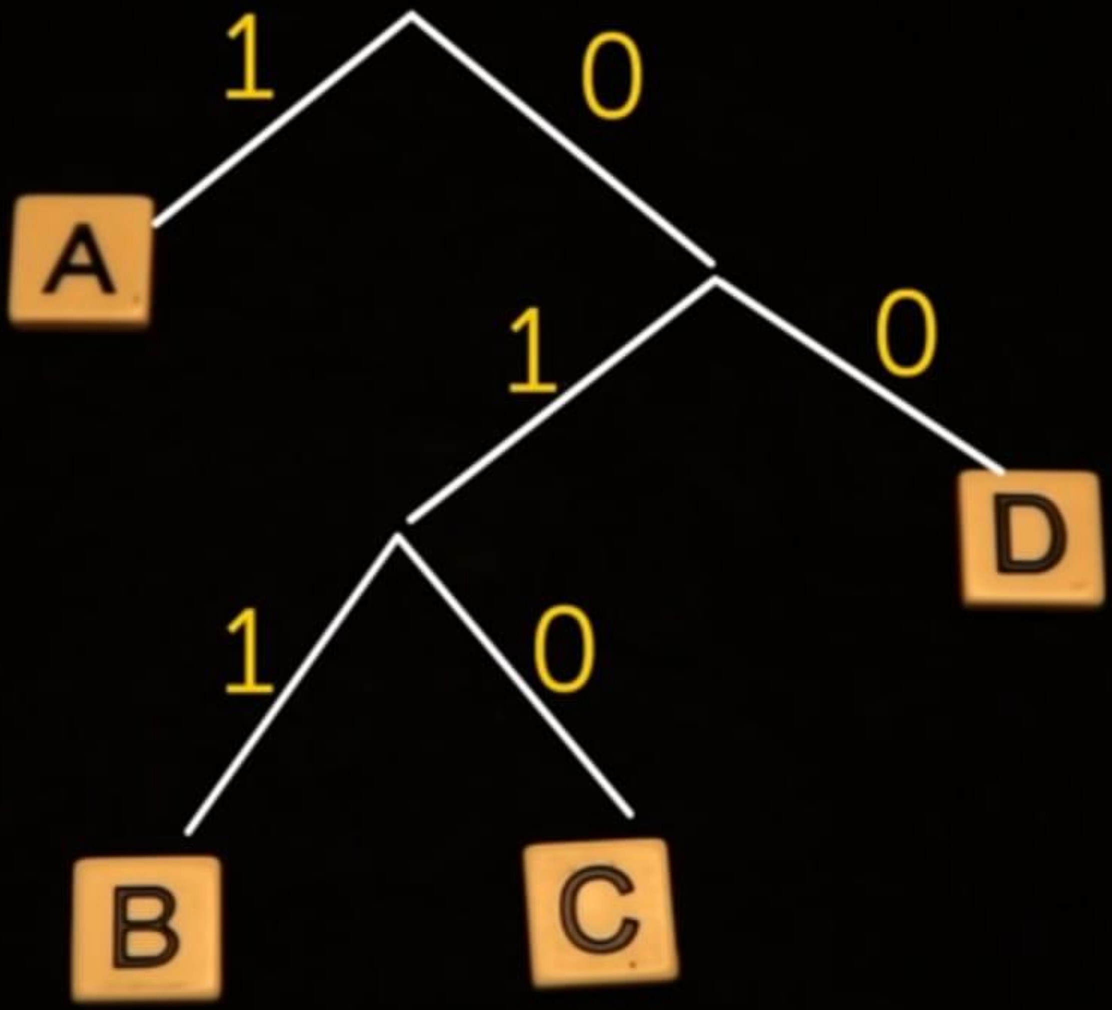
And making the Reduced Order BDD, gives a method of compression

This method of BDDs and the relevant Huffman coding gives you the mathematically most efficient algorithm

A=1	B=011	C=010	D=00

1011010000110101 = ABCDBCA

Note that in the above example it does take more bits than the straightforward correspondence. But the whole point is that over large data transmissions, A occurs more often - and the reduction in data transmission is 12.5%


1
0
A
1
0
D
1
0
B
C

And Claude Shannon was the first to claim that the limit of compression will always be the entropy of the message source

As the uncertainty of the source decreases, the ability to compress increases

Silicon Valley anyone?

A simple follow up question:

How do you minimise the cost to transmit the sum from the roll of two dice?

Extraterrestrial Intelligence

The search for extraterrestrial intelligence began in 1959

What does an intelligent signal look like?
In short, we don’t know. But as a starting point, we look for a signal which nature will not produce
Nature’s communication systems like dolphins have a structured ‘talk’. There is a decrease in entropy, as they say more things

A narrow band radio signal + higher order Shannon entropies, we most probably got it from an intended alien civilisation

Without even understanding the language, Claude Shannon’s entropy is a unit of measure that can allow us to detect the presence of structural rules regardless of meaning

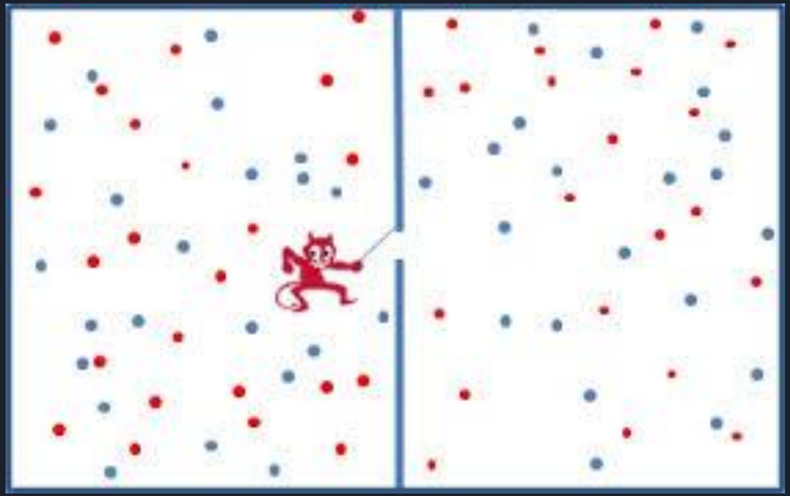
Maxwell’s Demon

Maxwell’s Demon

In 1871, James Clerk Maxwell proposed the existence of a machine that apparently violated the second law of thermodynamics. He envisioned a miniature little ‘demon’, which could reduce the entropy of a gas cylinder initially at equilibrium by individually separating the fast and slow molecules into the two halves of the cylinder.

Maxwell's Demon

When a fast molecule approaches from the left side the demon opens a door, allowing the molecule through, and then closes the door. By doing this many times the total entropy of the cylinder can be decreased, an apparent violation of the second law of thermodynamics.



Resolution to this paradox

We often fail to consider the subject: the demon.

In order to identify faster/hot molecules, it has to perform an observation (measurement) of the molecules in order to determine their velocities.

The result of this measurement must be stored in the demon's memory.

Because any memory is finite, the demon must eventually begin erasing information from its memory, in order to have space for new observations.

Landauer’s principle

Any logically irreversible manipulation of information, such as the erasure of a bit or the merging of two computation paths, must be accompanied by a corresponding entropy increase in non-information-bearing degrees of freedom of the information-processing apparatus or its environment.

Maxwell’s Demon

The act of erasing information increases the total entropy of the combined system – demon, gas cylinder, and their environments. Thus the entropy of the combined system is increased at least as much by this act of erasing information as the entropy of the combined system is decreased by the actions of the demon, ensuring that the second law of thermodynamics is obeyed.

“If your theory is found to be against the second law of thermodynamics, I give you no hope; there is nothing for it but to collapse in deepest humiliation.”

- Arthur Eddington

Entropy and its generalisation

Shannon Entropy

Shannon entropy for a random variable is given by:

$$H = -\sum_i P_i \log P_i$$

Where the summation is over all possible outcomes of the random variable.

For a coin toss we have the entropy as 1 bit (taking log base 2 gives the entropy in bits).

Von Neumann Entropy

A generalisation of Shannon entropy to the field of quantum mechanics.

For any quantum system we have a density matrix of the system which describes it statistical state.

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

Where p_i is the probability that the pure state $|\psi_i\rangle$ occurs.

The Von Neumann entropy is given by: $S = -\mathrm{tr}(\rho \ln \rho)$

Different interactions and entropy changes

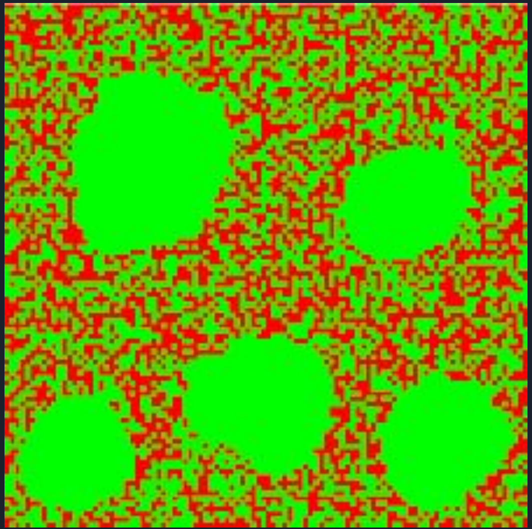
Interactions	Effect	Entropy
Object - Object	Unitary Evolution	Unchanged
Object - Subject	Observation	Decreases
Object - Environment	Decoherence	Increases

How inflation explains low entropy (10^123 vs 10^89)

Inflation is a theory of exponential expansion of space in the early universe.

2 points in the inflationery (green) region are strongly correlated.

In general, entropy decreases linearly with observations. But for this case entropy is decreasing exponentially with observations.



Disclaimer

Too many postulates and facts at face-value

Additional topics that might interest you:

- Source Encoding
- Error Correction
- Etymology
- Vexicography (Sheldon Cooper anyone?)
- Markov Chains
- Morse Code
- Hamming Code

Throughout tonight’s discussion, I never directly answered “What is Information Theory?”

Information is pretty much anything and there are so many different ways to share it. And at the end of the day, it comes down to one thing - THE BIT

Also, if you like, you could stick around and we could take up a tangential topic about “Evolutionary Linguistics and Etymology”

Hope you liked today’s GD. We would soon be floating a feedback form which we encourage you to fill as it will help us tailor our future events. :)