A BRIEF HISTORY OF INFORMATION THEORY

Source: Khan Academy – Information Theory (15 Parts)

https://www.youtube.com/watch?v=69-YUSazuic&list=PLP6PHJ8SLR6D4ytpHhZBdylPNcazU5m7o&index=2

Compression by HV Gupta (07/17/23)

The Development of Written Communication from Pictorial Representations to Alphabets

- ***Information*** is a *message* that is *stored* and/or *transmitted* using some ***medium***.
- Any ***Language*** consists of *messages* formed by arranging ***symbols*** (which encode meaning) into specific ***patterns***.
  - The first *written* form of *language* ***was pictorial representations***, that enabled ***symbolic visualizations*** of messages, and spanned from:
    - ***detailed Artistic renditions***
    - to ***simplified Pictogram*** symbols that resemble natural forms, and eventually
    - ***Ideograms*** to visualize abstract concepts that are hard to draw "*realistic*" pictures
  - ***Messages*** consisting of distinct mental objects could be constructed by combining individual symbols in terms of their meaning, called ***Protowriting***.
- ***Symbols in early written languages*** (3000 BC; hieroglyphics, cuneiform) were broadly of two types, representing
  - Single concepts (***word/meaning signs***) – larger fraction
  - Chunks of sound (***sound/phonetic signs***) – smaller fraction
    - of which even fewer represented distinct consonants.
- The invention of ***Papyrus*** as a cheap/portable medium for ***sending messages across space***, made writing more accessible.
  - This led to *hieratic* and then *demotic* writing using fewer and more simplified symbols, with a shift towards phonetic symbols (*letters*).
- This eventually led to ***Alphabets of ~22 consonant symbols*** that could be fitted to diverse tongues.
  - ***Alphabets*** *are powerful* because the actual choice of *symbols* (encoding method) is unimportant – Information is conveyed simply by selecting from a set of <u>possible</u> *symbols*.

Development of Methods for Transmitting Information Across Large Distances

- One of the oldest technologies was the '***signal fire***'
  - the presence/absence of fire indicates a single binary symbol that can convey $s = 2$ possible messages ($On/Off, Y/N, A/B$, etc).
  - Using n such '*torches*', $s^n = 2^n$ possible "messages" can be sent, and with $n = 5$ one can send all 26 letters of the English alphabet and more ($2^5 = 32$).
- The *'six-shutter telegraph'* (Murray 1795) could transmit up to $2^6 = 64$ possible symbols/messages.

Development of Methods for Transmitting Information using Electricity

- The discoveries of ***electricity***, ***conductors, insulators*** and ***capacitors***, and how to generate discharges on demand, led to systems for sending ***flows of electricity*** down long wires.
  - The discovery (1819) that a ***current-carrying wire creates a magnetic field*** led to development of the ***galvanometer*** for measuring the strength of electrical current
  - This led to the ***electromagnet***, making it possible to create magnetic fields that could move needles with precision and force using electric current applied at a distance.
- Gauss and Weber (1832) coupled a ***galvanometer*** with a ***switch*** to ***reverse the direction of current/magnetic field*** (deflecting the needle to left or right)
  - This was a precursor to the ***needle telegraph***, enabling transmission of ***all letters of the alphabet using one circuit***.

- Morse and Vail (1938) develop the ***spring-loaded lever (key)*** to enable controlled pulses of electric current to be transmitted by the tap of a finger
    - the length of a pulse could be varied to be short (called a 'dot') or long (called a 'dash').
- In the late 19th century, ***machines*** were developed for automatically converting higher-level symbol inputs (letters) into a precise and rapid stream of lower-level symbols (electrical pulses).
    - The ***Quadruplex Telegraph*** (Edison) combined signals of two different *strengths* (produced using weak and strong batteries) with two current *directions* to increase the ***number of different signaling events*** (*symbol space*) to $s = 4$
    - This increased the symbol transmission rate (Baud Rate).
    - Using strings of n symbols, $s^n = 4^n$ possible messages (letters) could be transmitted.
- It became clear that an information transmission system could be primarily thought of in terms of its ***capacity***.
    - The important thing was *how many* messages could be transmitted, and how efficiently (cheaply), rather than the actual *meaning* assigned to each symbol string.

Coding Strategies & Transmission Rate

- ***Information*** can be sent at a certain ***rate*** (called the ***channel capacity*** or ***Baud Rate***) using:
    - ***sequences of transmission "pulses"***, and a
    - ***coding strategy*** that uniquely associates each possible message (from a finite predetermined set) with a different number of transmission pulses
- For faster transmission rates, the ***coding strategy*** should assign fewer transmission pulses to more frequent/probable messages.
    - The "***Morse Code***" uses a coding strategy that assigns shorter symbol sequences to more probable letters, based on known letter frequencies.

Development of a Mathematical Understanding of Information

- '***Information'*** is an elastic term, because entire messages can be mapped to single symbols/words.
- ***Hartley*** (1928), building on ideas of ***Nyquist***, defined ***'Information'*** as the logarithm of the message space $h = \log(s^n) = n \times \log(s)$.
    - Assuming *random* symbol selection, this is the minimum number of questions to correctly determine a message from a string of n symbols drawn from an alphabet of size s
    - In reality most communication is a mix of predictability and surprise, which knowledge can used to reduce the length of a transmission (required number of yes-or-no questions).
- ***Bernoulli*** (1700's) showed that, in a sequence of repeated *independent* trials, the expected ratio of two possible events will converge on the actual ratio (*Weak Law of Large Numbers*).
    - Extensions of his work showed that the probability of variation away from averages also follows a familiar underlying shape/distribution (*Central Limit Theorem*).
    - ***Markov*** (1906) invented the concept of modeling sequences of random events using states and state-transitions (***Markov Chain***) and extended Bernoulli's results to *dependent* events.
- Shannon (1949) used ***Markov models*** as the basis for how we can think about communication
    - he showed how to ***design a Machine to generate English-like text*** (with the same statistical structure) using a Markov chain with conditional/dependent probabilities of selecting the next letter (word).
    - Shannon realized that that ***the amount of Information in some Message is tied up in the design of the Machine*** that could be used to generate similar looking sequences.
- Shannon reframed "*Which machine produces more information?*" as "*To predict the next symbol from each machine, what is the minimum number of yes or no questions you would expect to ask?*"
    - For more predictable systems (non-equal message probabilities), the machine produces less information (output is less uncertain), and so the expected number of questions is smaller.
    - Shannon defined a measure of average uncertainty/surprise "***Entropy***" $H = \sum_{i=1}^{n} p_i \times \log_2\left(\frac{1}{p_i}\right)$, measured in "bits", where p_i is the probability of a symbol

Data Compression via Huffman Coding

- ***Huffman Coding*** (1952) provides the ***Optimal Compression Strategy*** when the symbol probabilities are different and known.
  - The ***Limit of Compression*** will always be the ***Entropy*** of the message source.
  - The ability to compress increases as the source entropy decreases due to ***Statistical Structure***.
  - To compress beyond *Entropy*, we must *throw away information* in our messages.

Error Correction via Hamming Coding

- ***Transmission of messages across distance*** encounters the ***problem of Signal Corruption***.
  - To communicate error free as the channel noise increases, we must increase the amount of ***Redundancy***, which decreases the effective amount of information sent per unit time.
- ***Hamming Coding*** (1940s) is a coding method for ***detecting and correcting single bit errors***, at the expense of slightly increasing the size (redundancy) of the source message.
  - Hamming's ***'seven-four' code***, adds three parity bits to each string of four data bits

The Structure of Intelligent Signals

- ***Zipf's Law*** indicates that as humans develop from babies to adults, the ranked graph of sounds they produce changes from having a slope that is nearly level (all produced sounds occur randomly), to having a *negative-one slope on a log-log chart*.
- Due to the linguistic structure of human language, the *predictability* of the next word in a sequence increases (*information entropy* decreases) when conditioned on longer word strings.
  - Because this pattern emerges in both human and non-human communication, it has been suggested that this decreasing of entropy is essential for the transmission of knowledge.
- Shannon's entropy can be used to detect the presence of structural rules, regardless of meaning.