UNIT-1

INFORMATION THEORY

Information theory is a branch of science that deals with the analysis of a communications system. We will study digital communications – using a file (or network protocol) as the channel Claude Shannon Published a landmark paper in 1948 that was the beginning of the branch of information theory. The messages will be a sequence of binary digits Does anyone know the term for a binary digit.

Fig 1.1 Block Diagram of a typical communication system

(Source:https://www.google.com/search?q=Block+Diagram+of+a+typical+communication+system)

One detail that makes communicating difficult is noise noise introduces uncertainty Suppose I wish to transmit one bit of information what are all of the possibilities tx 0, rx 0 - good tx 0, rx 1 - error tx 1, rx 0 - error tx 1, rx 1 - good Two of the cases above have errors – this is where probability fits into the picture In the case of steganography, the noise may be due to attacks on the hiding algorithm. Claude Shannon introduced the idea of self-information.

$$I(Xj) = \log\frac{1}{p(xj)} = \log\frac{1}{p(j)} = -\log pj$$

Suppose we have an event X, where Xi represents a particular outcome of the

Consider flipping a fair coin, there are two equiprobable outcomes: say X0 = heads, P0 = 1/2, X1 = tails, P1 = 1/2 The amount of self-information for any single result is 1 bit. In other words, the number of bits required to communicate the result of the event is 1 bit. When outcomes are equally likely, there is a lot of information in the result. The higher the likelihood of a particular outcome, the less information that outcome conveys However, if the coin is biased such that it lands with heads up 99% of the time, there is not much information conveyed when we flip the coin and it lands on heads. Suppose we have an event X, where Xi represents a particular outcome of the event. Consider flipping a coin, however, let‘s say there are 3 possible outcomes: heads (P = 0.49), tails (P=0.49), lands on its side (P = 0.02) – (likely much higher than in reality).

Information

There is no some exact definition, however Information carries new specific knowledge, which is definitely new for its recipient; Information is always carried by some specific carrier in different forms (letters, digits, different specific symbols, sequences of digits, letters, and symbols , etc.); Information is meaningful only if the recipient is able to interpret it. According to the Oxford English Dictionary, the earliest historical meaning of the word information in English was the act of informing, or giving form or shape to the mind. The English word was apparently derived by adding the common "noun of action" ending "-action the information materialized is a message.

Information is always about something (size of a parameter, occurrence of an event, etc). Viewed in this manner, information does not have to be accurate; it may be a truth or a lie. Even a disruptive noise used to inhibit the flow of communication and create misunderstanding would in this view be a form of information. However, generally speaking, if the amount of information in the received message increases, the message is more accurate. Information Theory How we can measure the amount of information? How we can ensure the correctness of information? What to do if information gets corrupted by errors? How much memory does it require to store information? Basic answers to these questions that formed a solid background of the modern information theory were given by the great American mathematician, electrical engineer, and computer scientist Claude E. Shannon in his paper ―A Mathematical Theory of Communication‖ published in ―The Bell System Technical Journal in October, 1948.

A noiseless binary channel 0 0 transmits bits without error, What to do if we have a noisy channel and you want to send information across reliably? Information Capacity Theorem (Shannon Limit) The information capacity (or channel capacity) C of a continuous channel with bandwidth BHertz can be perturbed by additive Gaussian white noise of power spectral density N0/2, C=B log2(1+P/N0B) bits/sec provided bandwidth B satisfies where P is the average transmitted power P = Eb Rb (for an ideal system, Rb= C). Eb is the transmitted energy per bit, Rb is transmission rate.

ENTROPY:

Entropy is the average amount of information contained in each message received.

Here, message stands for an event, sample or character drawn from a distribution or data stream. Entropy thus characterizes our uncertainty about our source of information. (Entropy is best understood as a measure of uncertainty rather than certainty as entropy is larger for more random sources.) The source is also characterized by the probability distribution of the samples drawn from it.

Formula for entropy:

Information strictly in terms of the probabilities of events. Therefore, let us suppose that we have a set of probabilities (a probability distribution) P = {p1, p2, . . . , pn}. We define entropy of the distribution P by

$$H(P) = \sum_{i=1}^{n} pi * log(1/pi).$$

Shannon defined the entropy of the a discrete random variable X with possible values {x1, ..., xn} and probability mass function P(X) as: Here E is the expected value operator, and I is the information content of X. I(X) is itself a random variable. One may also define the conditional entropy of two events X and Y taking values xi and yj respectively, as

$$H(X|Y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(y_j)}{p(x_i, y_j)}$$

where $p(x_i,y_j)$ is the probability that $X=x_i$ and $Y=y_j$.

Properties:

- If *X* and *Y* are two independent experiments, then knowing the value of *Y* doesn't influence our knowledge of the value of *X* (since the two don't influence each other by independence):

$$H(X|Y) = H(X).$$

- The entropy of two simultaneous events is no more than the sum of the entropies of each individual event, and are equal if the two events are independent. More specifically, if*X* and *Y* are two random variables on the same probability space, and *(X,Y)* denotes their Cartesian product, then

$$H[(X, Y)] \leq H(X) + H(Y).$$