ON NEGATIVE CHARACTER OF INFORMATION

BY G. K. AGGARWAL

(Assistant Professor, Institute of Armament Technology, Poona-12, India)

Received August 10, 1967

(Communicated by Prof. P. L. Bhatnagar, F.A.Sc.)

ABSTRACT

The article puts forth a hypothesis of negative information which identifies increase of uncertainty with negative character of information. An attempt is made to give a measure of the property without attaching significance to the meaning of the information. The significance of the concept is sought to be related to the field of artificial intelligence and Self-organising Systems within the broader field of computers.

BRILLOUIN¹ has pointed out that the present theory of communication does not take care of the negative character of information since the latter derives itself out of a consideration to the “meaning” or the “value” of information whereas under information theory we consider only the quantum of information. Distinction is made between free information, which occurs when the possible cases are regarded as abstract and have no specified physical significance, and bound information, which occurs when the possible cases can be interpreted as complexions of a physical system. Only in case of bound information, a connection between information and negentropy is considered. Initially, if the number of complexions is \( P_0 \), the bound information is \( I_{b0} = 0 \) and the entropy is \( S_0 = K \ln P_0 \). Finally, if the number of complexions is \( P_1 < P_0 \), the bound information is \( I_{b1} \neq 0 \) and the entropy is \( S_1 = K \ln P_1 \). Evidently, in this scheme, the system is not isolated, the entropy is decreased when information is obtained, reducing the number of complexions, and this information must be furnished by some external agent whose entropy will increase. The relation between the decrease in entropy of our system and the required information is

\[
I_{b1} = K (\ln P_0 - \ln P_1) = S_0 - S_1 \tag{1}
\]

\[
S_1 = S_0 - I_{b1} \tag{2}
\]
We can say that

\[ \text{Bound Information} = \text{Decrease in Entropy} \ S = \text{Increase in Negentropy} \ N \] (3)

We also like to put down here Shannow’s measure of information, according to which the information content of a message equals the amount of uncertainty removed by that message, thus if the message is very rare, the presence of it removes greater amount of uncertainty and its information content is greater than of a message that is less rare than this message. If the probability of this message is \( P \), its information content is

\[ I(x) = - \log_2 P \] (4)

and average information content per message out of several ones is

\[ H = - \sum_i P_i \log_2 P_i. \] (5)

Also, it is stipulated that

\[ I(x) = 0 \quad \text{if} \quad P = 0. \]

**Negative Information**

An example of negative information given by Brillouin\(^1\) does not appeal to me. According to that example if a professor, at the end of his one-hour talk, says, “Excuse me, this was all wrong,” this sentence is negative information because it destroys all the information the professor gave during the one hour of his talk. To my mind, this sentence is a positive information because it makes it certain that all the professor said earlier was definitely wrong. The last sentence leads towards certainty and is positive information since removal of uncertainty or addition of certainty is the essential character of positive information. Let me define Negative information as “increase of uncertainty” just as we define positive information as “decrease of uncertainty”. In other words, we may say that if an information increases the entropy of our system we call it negative information and its quantitative measure equals the increase in entropy of the system. Following the argument in terms of number of physical complexions, at the beginning let the number of physical complexions be \( P_0 \) and an incoming
information increase the number to $P_1$ where $P_1 > P_0$. The required information, then, is

$$I_{b1} = K (ln P_0 - ln P_1) = -K \log_e \frac{P_1}{P_0}$$

which is negative since $P_1 > P_0$.

Viewing negative information from another angle, let me examine the effect of a piece of information on the probability distribution of a group of other pieces of information. In other words, I am trying to enter the concept of negative information through its effect on probability of the given message and hence its information content. Suppose there were $X$ alternative methods to try a solution to some given problem and a professor picks up one of them successfully in his lecture. The information content is $-\log_e (x^{-1})$, assuming that he has no preference to any particular alternatives. If after the lecture he says that the alternative he picked up was not correct, the information content of this announcement is obviously $-\log_e (x - 1/x)$ which again is positive and not equal to the information it has supposedly destroyed. In this same illustration, suppose the professor initially picked up an alternative which, according to him, had a probability of success $1/y$, where $y < x$. The information content is clearly $-\log_e y^{-1} = \log_e y$. At the end of his lecture he discovers that the alternative picked up by him would be no better than any other. This announcement changes the uncertainty connected with his earlier choice and in fact with all or several other choices. The probability of success of his choice becomes $1/x$ and its information content is $\log_e x$. The announcement, therefore, increases uncertainty associated with that choice by an amount $\log_e (x/y)$ and is negative information of that magnitude. It is significant to point out here that we are not using this information to predict our future likely choice but that this information has only changed our confidence or the information content of our initial choice. We call it negative information since it increased the uncertainty associated with our initial choice. Had this additional information confirmed that our initial choice was definitely wrong, we would have called it positive information since it is confirmatory in nature, like the elimination tests in chemical analysis.

Let me take another example. Suppose messages are being received on four frequencies $f_1, f_2, f_3, f_4$ (carrier frequencies) and the reception is by way of random pick up of these frequencies at any instant. Let the frequency of interest be $f_1$. If no extraneous source is trying to confuse messages on
On Negative Character of Information

If, the probability of receiving a message on \( f_1 \) is \( \frac{1}{4} \) and the information content is \(- \log_2 \frac{1}{4} = 2 \) bits. However, if an extraneous source is also sending a message, different from the genuine message (in appearance and not attaching any importance to the meaning), on \( f_1 \), the probability of receiving a genuine message on \( f_1 \) becomes \( \frac{1}{8} \). (One may immediately point out that in so far as information theory is concerned the probability of receiving message on \( f_1 \) equals the probability of receiving \( f_1 \) itself or when broken down to bits, the information will be equal to sum of the bits in the two messages since no significance is attached to the sense or the meaning of the messages. My clarification is that we expect only one set of symbols in a time whereas we find that there are two sets of symbols being received on frequency \( f_1 \), both the sets are such as are admissible in the code we have chosen. Thus our original message to be received has been made ambiguous and doubtful by a probability factor of \( \frac{1}{8} \), we may call it a confusion-factor or ambiguity-factor). That means the amount of uncertainty associated with a message on \( f_1 \) has increased, increasing its information content from \(- \log_2 \frac{1}{4} \) to \(- \log_2 \frac{1}{8} \), \( i.e., \) from 2 bits to 3 bits. Let us, then, define the confusion-factor \( CF \) or the ambiguity-factor of decoy-messages as follows:

\[
CF = \frac{\text{Entropy of the message with decoy} - \text{Entropy of original message}}{\text{Entropy of original message}} = \frac{\text{Increase in Entropy of the message due to decoy}}{\text{Entropy of Original message}} = \frac{\text{Negative information content of decoy message}}{\text{Information content of original message}}.
\]

With reference to the above example, the information content of the decoy message is

\[
I_b = - \log_2 4 - \log_2 8 = -1 \text{ bit}
\]

or else,

\[
= k (\log_2 4 - \log_2 8) = -k \log_2 2.
\]

Still another example might be taken to illustrate that it is the addition to the uncertainty, rather than destruction or negation of a given information, that means negative information. Suppose a person shoots down a culprit and thinks him to be dead. After several days he is told that the culprit he shot at was not dead. This new information is not negative. If, however, after several days he comes across an identical face (with that
of the culprit) he is immediately in doubt, and the simplest calculation would show that this information is \(-k \log_2 n\), i.e., \(k (\log_e 1 - \log_e 2)\), or else \(-1\) bit since it increases the information content by an amount \(-\log_2 \frac{1}{2}\). I must remark that I am trying to avoid any consideration to the meaning of an information but how far I am successful in that attempt is for the reader to judge. This negative character of information enters by way of increasing the number of physical complexions in our system. Of course, taking into account the meaning also, the amount of uncertainty and ambiguity generated by the second information is enormous. A measure of this uncertainty taking meaning also into account could be visualised as the number of possible alternatives that need to be examined before finally giving a verdict on the validity or otherwise of the original information about which uncertainty has been generated. Alternately, the uncertainty may be measured as the number of alternatives to be examined or additional information required in order to modify the earlier information in the given context. At times, the context itself is changed as a result of some information. In this section, however, I have tried to consider the concept of negative information without regard to the meaning of it.

**NEGATIVE INFORMATION AND ARTIFICIAL INTELLIGENCE**

In the preceding section I have tried to bring out that negative information content of an information could be associated with what I have called “Confusion-factor” or “ambiguity-factor”. This aspect is quite relevant to the context of artificial intelligence. On the one hand, a machine (or a person) with fixed probability distribution over its various functional characteristics cannot generate fresh intelligence (i.e., fresh logical dexterity) and, therefore, introducing a certain amount of negative information or a confusion-factor into the machine (or the person) could result in fresh intelligence for the machine or the person since it will have to generate from within, or look up from without, additional information to dispel the ambiguity produced by the negative information. On the other hand, one could use negative information or the confusion-factor to test the intelligence stored in a machine since according to psychologists, there is a correlation between ambiguity-tolerance and intelligence. This suggests that if we can devise clear cut quantitative measure of negative information, we could measure quantitatively the amount of intelligence we have been able to store in an artificial-intelligence designed by us—at least we could compare quantitatively the intelligence stored in two or more artificial-intelligent machines.
ON NEGATIVE CHARACTER OF INFORMATION

CONCLUDING REMARKS

An excellent theory of information has been developed on the consideration of "removal of uncertainty" resulting in a quantitative measure of positive information. Probably, it might be possible to build up a theory of information based on "addition of uncertainty" resulting in a quantitative measure of "negative information" and an associated "confusion-factor". In any case, any attempt to do the latter would be worthwhile since to a machine that has an intelligence built into it, it is equally important, if not more, to assess its shortcomings (by facing confusion or ambiguity) as it is to assess its capabilities (to dispel uncertainty); Also the concept might lead to a measure of quantitative comparison of the intelligence stored into different artificial-intelligent machines. Who knows, we might be able to design "Confusion-generators" or "ambiguity-generators" to test artificially intelligent machines or to help them to acquire fresh intelligence which could be quantitatively assessed!

REFERENCE