SHANNON EQUATION REFORM AND APPLICATIONS

Chenguang Lu
Department of Computer Science
Changsha University, 410008, P.R.China

Abstract: Based on the distinction between two types of probabilities: normal and logical probabilities, Shannon entropy equation and Shannon cross-entropy equation are reformed. New equations do not require that a source or a destination of information must be a set of disjoint events, and, can be used to measure semantic and sensory information. The membership function of fuzzy sets and the discriminate probability of senses are brought into the new cross-entropy equation. Some examples of calculating semantic and sensory information are provided.

Keywords: Semantic information, Sensory information, Image information, Generalized information measure, Logical probability, Fuzzy sets, Fuzzy discrimination.

I. INTRODUCTION

The Shannon theory has made great successes in electronic communication[1]; yet, it is usually powerless in communications with language and with sensory organs. To resolve the problem of semantic information, researchers proposed many methods[2-6] that were inspiring more or less; however, these methods are difficult to be applied to practice. It seems to me that the main cause is that these methods still follow the forms of Shannon equations; yet, these forms are imperfect to wider extent and need to be improved.

II. REFORM OF SHANNON ENTROPY EQUATION

A. Distinction Between Two Types of Probabilities

There are two types of probabilities of a sentence or a proposition. One is the normal probability in which the sentence is selected. Another is the logical probability[2] in which the sentence, or more properly the proposition the sentence states, is judged true by many language users.

First we define the logical probability in terms of set theory. Suppose that X and Y are two random variables taking values x_i $(i=1,2,\ldots,N)$ in set A and $y_j(j=1,2,\ldots M)$ in set B; A is a set of disjoint events; B is a set of joint sentences and the disjunction of all sentences in B is true; $S(y_j)$ is a subset over A, every sentence it includes makes y_j be true. We define that the logical probability of y_j:

$$Q(y_j) = \sum_{x_i \in S(y_j)} P(x_i) \qquad (2.1)$$

Since some sentences in B may be joint, the sum of logical probabilities of all y_j in B may be greater than 1; in other words, the logical probabilities of sentences in B may not be normalized. However, their normal probabilities are normalized.

We use an example to explain distinction and relation between normal probability and logical probability. Let a set A = {non-rain(x_1), light-rain(x_2), heavy-rain(x_3)}, events in which are disjoint, and a set B = {"It doesn't rain"(NR), "It rains lightly"(LR), "It rains heavily"(HR), "It rains"(RA)}, some sentences in which are joint. Every Time, a man only selects one of four sentences in B to describe one of three different weathers. Hence y_j has normal probability $P(y_j)$ and logical probability $Q(y_j)$; and

$$\Sigma_j P(y_j) = 1 \quad \text{and} \quad \Sigma_j Q(y_j) \geq 1. \qquad (2.2)$$

Generally, $P(y_j) \leq Q(y_j)$; $P(y_j) = Q(y_j)$ only when y_j is disjoint with all other sentences. Some times, P(RA) may be less then P(LR) or P(HR) because the man selects "It rains" only when he does not know in detail a weather and therefore cannot determine which sentence, "It rains lightly" or "It rains heavily", is better.

B. Deducing Shannon Entropy Equation

Clearly, y_j being true or being selected means X happens in $S(y_j)$. By the definition of amount of information, the information transmitted by y_j is determined by prior probability, $Q(y_j)$, and posterior probability, which is equal to 1, of event that X takes place in $S(y_j)$, i.e.

$$I(y_j) = \log[1/Q(y_j)] = -\log Q(y_j). \qquad (2.3)$$

Here $I(y_j)$ shouldn't be $-\log P(y_j)$, otherwise there will be a wrong conclusion that RA may give more information then LR or HR. However, $P(y_j)$ is also useful. By the definition of average amount of information, we have

$$H(Y) = -\Sigma_j P(y_j)\log Q(y_j), \qquad (2.4)$$

If any two sentences in B are disjoint, then the above equation will revert to the Shannon entropy equation.

C. Generalized Shannon Entropy Equation for Sensory Information

We turn next to sensory information. Suppose there is a set, A = $\{x_1, x_2, \ldots, x_N\}$, of colours and the corresponding set, B, of sensations; y_i in B is the function of x_i, i.e. $y_i = y(x_i)$; a man cannot discern two colours that have adjacent subscripts because his visual discrimination is limited. Since y_j determines the set $S(y_j)$ that includes all x_i that cannot be distinguished from x_j by the man's eyes, sensation y_j also has two types of probabilities. One is the normal probability $P(y_j) = P(x_j)$; another is the logical probability $Q(y_j)$; for example,

$$Q(y_2) = P(x_1) + P(x_2) + P(x_3).$$

Similarly, using (2.4) we can turn out the average amount of information transmitted by Y.

Obviously, if the visual discrimination is high enough, Q(Y) is equal to P(Y) and H(Y) will revert to the Shannon entropy. So, (2.4) may be called a generalized Shannon entropy equation or a generalized entropy equation.

III. REFORM OF SHANNON CROSS-ENTROPY EQUATION

Actually, when x_i is given, logical value or logical condition probability of y_j is generally fuzzy, i.e. $Q(y_j|x_i)$ varies in interval [0,1]; the confusion probability of a sensation y_j with $y(x_i)$ is similar. Hence, we need a new cross-entropy equation.

A. Deducing Generalized Cross-entropy Equation

Suppose A is a set of disjoint events, such as, a set of some weathers; B is a set of joint sentences. When x_i in A(i=1,2,...,N) is given, the logical condition probability of y_j in B(j=1,2,...,M) is $Q(y_j|x_i)$, which is equivalent or similar to "membership function" in the fuzzy set theory[4]. By definitions in probability theory,

$$Q(j) = \Sigma_i Q(i)Q(j|i), \quad (3.1)$$
$$P(j) = \Sigma_i P(i)P(j|i), \quad (3.2)$$

where $Q(j)$, $Q(i)$, and $Q(j|i)$ are shots for $Q(y_j)$, $Q(x_i)$, and $Q(y_j|x_i)$; the others are similar. We also have

$$Q(j,i) = Q(i)Q(j|i), \quad (3.3)$$
$$P(j,i) = P(i)P(j|i). \quad (3.4)$$
$$Q(i|j) = Q(j,i)/Q(j), \quad (3.5)$$
$$P(i|j) = P(j,i)/P(j). \quad (3.6)$$

Now, y_j ascertains a fuzzy subset $S(y_j)$ over A; the logical probability $Q(y_j)$ is prior (normal) probability of the event that X happens in $S(y_j)$; the $Q(y_j|X)$ is posterior probability of the same event after y_j is judged true. By the classical definition, the amount of information transmitted by y_j from $X = x_i$ is

$$I(x_i;y_j) = \log[Q(y_j|X=x_i)/Q(y_j)] = \log[Q(j|i)/Q(j)]. \quad (3.7)$$

Since A is a set of disjoint events, $Q(i) = P(i)$; $Q(i|j) = P(i|j)$. Hence

$$Q(j|i)/Q(j) = Q(i|j)/Q(i) = P(i|j)/P(i); \quad (3.8)$$
$$I(x_i;y_j) = \log[P(i|j)/P(i)] \quad (3.9)$$

Equation (3.9) can be seen in the classical information theory. It means that $I(x_i;y_j)$ is determined by prior and posterior probabilities of x_i.

Further, the average amount of information transmitted by y_j from X is

$$I(X;y_j)= \Sigma_i P(i|j)I(x_i;y_j); \quad (3.10)$$

and the average amount of information transmitted by Y from X is

$$\begin{aligned} I(X;Y) &= \Sigma_j P(j)I(X;y_j) \\ &= \Sigma_j \Sigma_i P(j,i)\log\{Q(j,i)/[Q(j)Q(i)]\} \\ &= H(Y)-H(Y|X) \end{aligned}$$

$$= H(Y)-H(X|Y) \quad (3.11)$$

in which H(Y) and H(X) are determined by (2.2), and

$$H(Y|X) = \Sigma_j \Sigma_i P(j,i)\log Q(j|i), \quad (3.12)$$
$$H(X|Y) = \Sigma_j \Sigma_i P(j,i)\log Q(i|j). \quad (3.13)$$

These equations are similar to those in the classical information theory. Since A is a set of disjoint events, H(X) and H(X|Y) become the Shannon entropy and the classical condition entropy; (3.11) becomes

$$I(X;Y)= \Sigma_j \Sigma_i [P(j)/Q(j)]P(i)Q(j|i)\log[Q(j|i)/Q(j)] \quad (3.14)$$

which is a practical equation we will use later.

Actually, (3.9)-(3.11) are also suitable to those cases in which A is also a set of joint events. A possible case is that A is a set of sentences like "The temperature is about 10(or 11, 12, ...) degree Centigrade" and B is a set of sentences like "It is cold(or warm, hot, ...)". In this case, I(X;Y) is the cross-information between two sets of concepts or propositions instead of two sets of physical signals.

From (3.14), further, if y_j is also a set of disjoint events, then P(j) = Q(j), P(j|i) = Q(j|i); (3.14) will regress into Shannon cross-entropy equation. So, (3.11) may be called the generalized cross-entropy equation. The Shannon cross-entropy equation can be thought its special case as both the source and destination of information are sets of disjoint events so that two types of probabilities are equal.

B. Discussion about Fuzziness

The H(Y|X) is a generalized condition entropy and can also be called a fuzzy entropy. We can prove that if the any sentence in B has no fuzziness, i.e. Q(j|i) is equal to 0 or 1, then

$$H(Y|X) = 0 \text{ and } I(X;Y) = H(Y); \quad (3.15)$$

if the sentences in B are extremely fuzzy, i.e. Q(j|i) does not vary with i, then

$$H(Y|X) = H(X) \text{ and } I(X;Y) = 0. \quad (3.16)$$

These means that, for the given sentence entropy H(Y), the clearer the language, the more the information.

If B only includes two complementary sentences and all x_i in A have equal probability 1/N, then H(Y|X) will regress into Deluca and Termini's fuzzy entropy[5]. Here different from [6] is that H(Y|X) is posterior entropy and only decreases instead of yields information.

IV. APPLICATIONS OF NEW CROSS-ENTROPY EQUATION TO SEMANTIC INFORMATION

A. Calculation of Semantic Information

Now we use (3.14) to calculate cross-information between the set, A = $\{x_1, x_2, x_3\}$, of weathers and the set, B = $\{y_1(NR), y_2(LR), y_3(HR), y_4(RA)\}$, of sentences.

We use a matrix, Q, to represent the set $\{Q(j|i)\}$, an element at row i and column j of the matrix is $Q(j|i)(i=1,2,3;\ j=1,2,3,4)$. There are three possible $\{Q(j|i)\}$:

$$Q_1=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 1\end{pmatrix},\quad Q_2=\begin{pmatrix}0.9 & 0.1 & 0.0 & 0.1\\ 0.1 & 0.8 & 0.1 & 0.9\\ 0.0 & 0.1 & 0.9 & 1.0\end{pmatrix},\quad Q_3=\begin{pmatrix}0.8 & 0.2 & 0.0 & 0.2\\ 0.2 & 0.6 & 0.2 & 0.8\\ 0.0 & 0.2 & 0.8 & 1.0\end{pmatrix}.$$

They indicate that the coding and decoding in linguistic communication is clear, fuzzy, and very fuzzy in order. Let all x_i in A have equal probability 1/3. Then we calculate out $Q(j) = 1/3$ for $j=1,2,3$ and $Q(4) = 2/3$ for each $\{Q(j|i)\}$. About the normalized probabilities of all y_j, there are limitations:

$$P(y_1) = Q(y_1),\ P(y_j) \le Q(y_j),\ \text{for } j=2,3,4.$$

There are the values of I(X;Y) for different $\{Q(j|i)\}$ and P(RA).

P(RA)	I(X;Y)		
	Q_1	Q_2	Q_3
0	1.583	0.965	0.647
2/9	1.363	0.845	0.577
2/3	0.918	0.606	0.437

Note the maximal number 1.585 in the table is just $H(X) = \text{Log}_2 3$. The results tell us that the clearer the sentences, the better; the less the normal probability of sentence RA, which is implied by other sentences, the more the information.

B. Fuzzy Logic Operatons for Logical Condition Probabilities

In practice, we needn't use statistics to obtain all logical condition probabilities $Q(j|i)$. For example, in the above case, we can first obtain $Q(LR|x_i)$ and $Q(RA|x_i)$ by statistics and then reach $Q(NR|x_i)$ and $Q(HR|x_i)$ by calculations:

$$Q(NR|x_i) = 1 - Q(RA|x_i),$$
$$Q(HR|x_i) = Q(RA|x_i) - Q(LR|x_i).$$

I have ever used an analog logical algebra in establishing a symmetrically mathematical model of colour vision[7,8]. By this algebra, we can define a fuzzy set algebra in which the all laws in Boolean algebra are still tenable and calculate out the membership function of any fuzzy set function(expression) from the membership functions of some atomic fuzzy sets. If we calculate the condition probability $Q(j|i)$ of a compound sentences as the membership function of a fuzzy set function, the result will be approximate to that by statistics. This method can guarantee that the logical condition probability $Q(j|i)$ is equal to 1 when y_j is a true proposition formed by the disjunction of some propositions, and $Q(j|i)$ is equal to 0 when y_j is a false proposition formed by the conjunction of some prepositions. However, popular Zadeh algebra does not quarantee[9].

V. APPLICATIONS OF NEW CROSS-ENTROPY EQUATION TO SENSORY INFORMATION

A. New Cross-entropy Equation for Sensory Information

Suppose A is a set of disjoint events, such as, a set of colours, sounds, gray levels of a pixel of an image, or images; B is a set of sensations or perceptions; $y_i \in B$ is the function of $x_i \in A(i=1,2,\dots,N)$. Hence, $P(y_i) = P(x_i)$. All x_i that are confused with x_j by men's eyes forms a fuzzy set $S(y_j)$ over A. The logical condition probability or the confusion probability of y_j with $y(x_i)$ is $Q(j|i) = Q(y_j|x_i)$. The $Q(j|i)$ can also be called the membership function of $S(y_j)$ or the membership grade of x_i in $S(y_j)$. Imitating (3.1) and (3.14), we can calculate out $Q(j)$ and $I(X;Y)$. Different here from (3.14) is

$$P(j) = P(y_j) = P(x_i), \quad i,j = 1,2,\dots,N.$$

B. Calculation of Visual Information from a Pixel

Now we regard the visual information from a pixel of a white-black image in digits. The pixel has a series of gray levels $x_i \in A(i=1,2,\dots,N = 2^k)$. The B is a set of brightness sensations and can be described by one dimensional space [0,1].

Fig. 1 shows a possible confusion function $Q(j|i)$ of y_j. In facts, brightness discrimination varies with viewing conditions, such as, viewing distance and lasting time, area of a pixel, and which viewer; so, we use d as the parameter of discrimination. The less the d, the higher the discrimination.

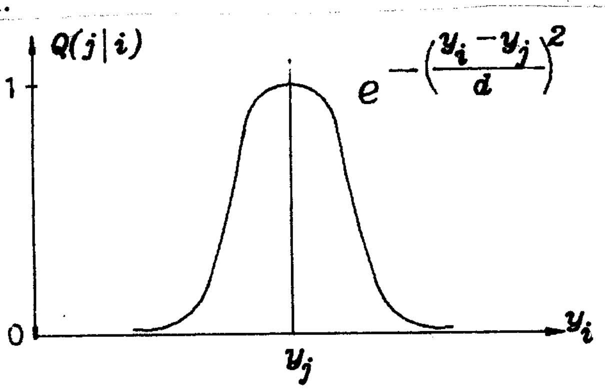

Fig.1 Confusion probability function of y_j with $y_i = y(x_i)$.

The relation between y_j and x_i we use is

$$y(x_i) = g(L_i) = g(h(x_i)), \qquad (5.1)$$

where L_i denotes the luminance of x_i. Function $g(L_i)$ and $h(x_i)$ are cited from CIE 1964[12] and [11]. We assume the brightness of original image has equal probability density distribution, and then calculate out all $P(x_i)$. In this way, $I(X;Y)$ can reach its maximum as d and k are given.

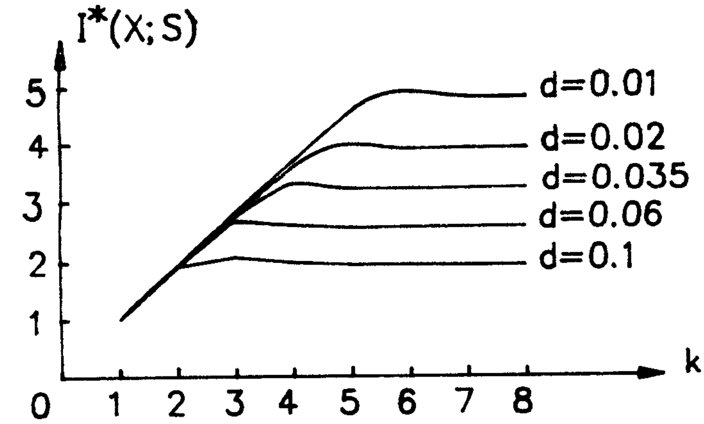

Fig.2. Cross-information I(X;Y)(bits) varies with number, k, of bits and discriminate parameter d.

Fig.2 shows that I(X;Y) varies with k and d. The result indicates that if the brightness discrimination is high enough, the amount of visual information is equal to that of physical information measured by the Shannon entropy; if the discrimination is limited, there exists the optimal number, $k = k_m$, of bits so that I(X;Y) reaches its maximum when {P(i)} and d are given. With k increasing from k_m, I(X;Y) will decrease down a little bit and then stop on a certain level. This result is somewhat beyond expectation. When $k = k_m$, some discerning work is done by the quantization of a machine; when $k > k_m$, this work need to be done by men's eyes, discrimination of which is fuzzy. So it is worse then before.

D. Discussion about Visual Information from Images

For an entire colour image in digits, although it is difficult to measure the visual information in practice as well as to measure the physical information, it is possible in theory.

Suppose an image has m rows and n columns of pixels; each pixel has an colour denoted by a vector (b,g,r) and $b,g,r = 1,2,\ldots,w = 2^k$. Then there are $N = w^{3mn}$ different images. we assume that A is a set of N possible images; B is a set of corresponding perceptions; y_j in B is the function of x_i in A. If we can obtain P(i) and Q(j|i) for each i,j(i,j=1,2,..N) by experiments, then we can use (5.1) to calculate out the average amount of visual information.

E. Generalized Cross-entropy Equation for Continuous Signals

We can extend the new cross-information equation to cases of continuous information sources and destinations.

Let A and B be sets of continuous signal, all x in A be disjoint and some y = y(x) in B be joint. Then the normal probability density of y is

$$p(y) = p(x)/y'(x), \tag{5.2}$$

The logical probability in stead of logical probability density of y is Q(y). The confusion probability of y with $y(x_0)(x_0 \in A)$ is $Q(y|x_0)$. Hence

$$Q(y) = \int_A Q(y|x_0)p(x_0)dx_0 \tag{5.3}$$

Similarly, we have

$$I(X;Y) = \int_B \int_A [p(y)/Q(y)]p(x_o)Q(y|x_o)\log[Q(y|x_o)/Q(y)]\, dx_o dy. \quad (5.4)$$

It is easy to prove that I(X;Y) must be finite as seen in Fig. 2 when k approaches infinity.

VI. CONCLUSION

In this paper, the logical probability is reduced into the normal probability based on set theory and fuzzy set theory; generalized entropy equation and generalized cross-entropy equation are deduced by the classical definition of amount of information instead of defined by one's willing. They surpass the Shannon theory, yet, can be understood on the base of the Shannon theory. Their applicability to semantic and sensory information has been shown by some examples. We can expect that the new theory is meaningful to the optimization of linguistic and sensory communications and to the research on some fields like philosophy and psychology.

References

[1] C. E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27, 379-429, 623-656(1948).

[2] Y. Bar-Hillel and R.Carnap, An outline of a theory of semantic information, Tech. Rep. No. 247, Research Lab. of Electronics, MIT(1952).

[3] H. W. Gottinger, Lecture notes on concepts and Measures of information, In 'Information Theory: New trends and Open Problems', G. Longo eds., Springer Verlag, CISM(1975).

[4] L. A. Zadeh, Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23,421-427(1968).

[5] A. De Luca and S. Termini, A definition of non-probabilistic entropy in the setting of fuzzy sets, Infor. Contr. 20, 201-312(1972).

[6] Y.-X. Zhong, United measure of information, IEEE ISIT 83, Session G3, Canada(1983).

[7] C. -G. Lu, Fuzzy switching quasi-Boolean algebra and vision simulation colour discernment, BUSEFAL 36, 132-138(1988).

[8] C. -G. Lu, Decoding model and its verification, ACTA OPTICA SINIC 9, 158-162(1989).

[9] L. A. Zadeh, Fuzzy sets, Infor. Contr. 8, 338-353(1965).

[10] W.-X. Xie and S. D. Bedrosian, An information measure for fuzzy sets, IEEE Trans. SMC SMC-14, 151-156(1984).

[11] W.-X. Xie and S. D. Bedrosian, Experimentally derived fuzzy membership function for gray level images, J. of Franklin Institute 325, 155-164(1988).

[12] M. Sato and R. Hioki, Modification of the(U*V*W*) system, J. Opt. Soc. Am. 59, 349-353(1969).