Lecture 06. Measures of information in the system

The purpose of the lecture: to consider an introduction to various ways of setting measures for measuring the amount of information, their critical comparative analysis, the main connections of information and the entropy of the system.

Lecture plan:
Introduction
1 Amount of information
2 Measure of R. Hartley
3 K. Shannon's measure
4 Thermodynamic measure
5 Energy-informational (quantum mechanical) measure
6 Other measures of information.
Conclusion

Keywords: information, software, amount of information, estimation problem, measure, function, costs, R. Hartley's measure, system state, set of states, base, bit, word, RAM, ordered systems, chromosomes, K. Shannon's measure, Shannon's formula , probability, Shannon's measure, Hartley's measure, universality, mean time, arbitrary, value, alphabet symbol, alphabet, combinatorics, expression, entropy, definition, Information theory, Boltzmann's formula, connection, thermodynamic measure, energy-informational, energy, resource, named set , energy-informational measure, ratio, sets, vocabulary, transmission rate, evolution, thesaurus, length, variable, index, component.

Lecture content:

Introduction

In the previous lecture, it was noted that information can be understood and interpreted in different problems, subject areas in different ways. As a consequence, there are different approaches to defining the measurement of information and different ways to introduce a measure of the amount of information.

1 Amount of information

The amount of information is a numerical value that adequately characterizes the updated information in terms of diversity, complexity, structuredness (orderliness), certainty, and the choice of states of the displayed system.

If some system is considered that can take one of n possible states, then the actual task is the task of assessing this choice, outcome. Such an assessment can be a measure of information (events).

The measure, as mentioned above, is a continuous real nonnegative function defined on the set of events and is additive (the measure of a sum is equal to the sum of measures).

Measures can be static and dynamic, depending on what information they allow you to evaluate: static (not updated; in fact, messages are evaluated without taking into account resources and the form of updating) or dynamic (updated, i.e. the costs of resources for updating are also estimated information).

Below we will not always, mainly for greater persuasiveness and greater meaningful understanding, draw clear mathematical boundaries between the concepts of "amount of information" and "measure of the amount of information", but a strict reader should always ask quite important questions: about the amount of information or about the least information in the specific sequence of events in question? are we talking about deterministic or stochastic information? what is the measure of the amount of information and how adequate is it?

2 Measure of R. Hartley

Let there be N states of the system S or N experiments with different, equally possible, sequential states of the system. If each state of the system is encoded, for example, with binary codes of a certain length d, then this length must be chosen so that the number of all different combinations is at least N. The smallest number for which this is possible is called the measure of the diversity of the set of states of the system and is given R. Hartley's formula: $H = k\log_a N$, where k is the coefficient of proportionality (scaling, depending on the chosen unit of measure), *a* is the base of the system of measure.

If the measurement is carried out in the exponential system, then $k = 1$, $H = \ln N$ (nature); if the measurement was made in the binary system, then $k = 1/\ln 2$, $H = \log_2 N$ (bit); if the measurement was made in decimal system, then $k = 1/\ln 10$, $H = \lg N$ (dit).

Example. To find out the position of a point in a system of two cells i.e. to get some information, you need to ask 1 question ("Left or right cell?"). Having learned the position of the point, we increase the total information about the system by 1 bit ($I = \log_2 2$). For a four-cell system, you need to ask 2 similar questions, and the information is 2 bits ($I = \log_2 4$). If the system has n different states, then the maximum amount of information will be determined by the formula: $I = \log_2 n$.

Hartley's statement is valid: if in some set $X = \{x_1, x_2, ..., x_n\}$ it is necessary to select an arbitrary element $x_i \in X$, then in order to select (find) it, it is necessary to obtain at least $\log_a n$ (units) information.

If N is the number of possible equally probable outcomes, then $k\ln N$ is a measure of our ignorance about the system.

According to Hartley, for a measure of information to be of practical value, it must be such as to reflect the amount of information in proportion to the number of choices.

Example. There are 192 coins. It is known that one of them is fake, for example, lighter in weight. Let's determine how many weighings need to be done to identify it. If we put an equal number of coins on the scales, we get 3 independent possibilities: a) the left cup is lower; b) the right cup is lower; c) the cups are balanced. Thus, each weighing gives the amount of information $I = \log_2 3$, therefore, to determine a counterfeit coin, at least k weighings must be made, where the smallest k satisfies the condition $\log_2 3^k \geq \log_2 192$. Hence, $k \geq 5$ or, k = 4 (or k = 5 – if we count as one weighing and the last one, which is obvious for determining the coin). So, it is necessary to do at least 5 weighings (5 is enough).

Example. Human DNA can be thought of as a word in a four-letter alphabet, where each letter marks a DNA chain link or nucleotide. Let us determine how much information (in bits) DNA contains if it contains approximately 1.5×10^{23} nucleotides (there are other estimates of this volume, but we will consider this option). One nucleotide contains $\log_2(4) = 2$ (bits) of information. Therefore, the structure of DNA in the human body can store 3×10^{23} bits of information. This is all information, including redundant information. There is much less information actually used – structured in the memory of a person. In this regard, we note that a person uses about 5-6% of neurons during an average life span (nerve cells in the brain - "human RAM cells"). The genetic code is an extremely complex and orderly system for recording information. The information contained in the genetic code (according to Darwin's teaching) has been accumulating for many millennia. Chromosome structures are a kind of cipher code, during cell division, copies of the cipher are created, each chromosome is doubled, each cell has a cipher code, while each person usually receives his own set of chromosomes (code) from the mother and from the father. The cipher code unfolds the process of human evolution. All life, as E. Schrödinger noted, "is the ordered and regular behavior of matter, based ... on the existence of order, which is maintained all the time."

Hartley's formula is abstracted from the semantic and qualitative, individual properties of the system under consideration (the quality of information in the manifestations of the system using the considered N states of the system). This is the main and positive side of the formula. But there is its main and negative side: the formula does not take into account the distinguishability and difference of the considered N states of the system.

A decrease (increase) in H may indicate a decrease (increase) in the variety of states N of the system. The converse, as follows from the Hartley formula (since the base of the logarithm is greater than 1!), is also true.

3 K. Shannon's measure

Shannon's formula evaluates information independently, abstracted from its meaning:

$$I = -\sum_{i=1}^{n} p_i \log_2 p_i,$$

where n is the number of states of the system; p_i is the probability (or relative frequency) of the system transition to the i-th state, and the sum of all p_i is equal to 1.

If all states are equally probable (i.e., $p_i = 1/n$), then $I = \log_2 n$.

K. Shannon proved the uniqueness theorem for the measure of the amount of information. For the case of a uniform probability density distribution, the Shannon measure coincides with the Hartley measure. The validity and sufficient versatility of the Hartley and Shannon formulas is confirmed by the data of neuropsychology.

Example. The response time t of the subject to the choice of an object from the available N objects linearly depends on $\log_2 N$: $t = 200 + 180\log_2 N$ (msec). The time of transmission of information in a living organism also changes according to a similar law. One of the experiments to determine the psychophysiological reactions of a person consisted in the fact that one of n lamps was lit in front of the subject a large number of times, to which he had to point during the experiment. It turned out that the average time required for the subject's correct answer is proportional not to the number n of bulbs, but to the value of I, determined by Shannon's formula, where p_i is the probability to light bulb i

It is easy to see that in the general case

$$I = -\sum_{i=1}^{n} p_i \log_2 p_i \leq \log_2 n.$$

If the choice of the i-th option is predetermined (in fact, there is no choice, $p_i = 1$), then $I = 0$.

A message about the occurrence of an event is less likely to carry more information than a message about the occurrence of an event with a higher probability. The message about the occurrence of a reliably upcoming event carries zero information (and this is quite clear: the event will still happen someday).

Example. If the position of a point in the system is known, in particular, it is in the k-th cell, i.e. all $p_i = 0$, except $p_k = 1$, then $I = \log_2 1 = 0$ and we do not get new information here (as it should be expected).

Example. Let's find out how many bits of information carry an arbitrary two-digit number with all significant digits (apart from its specific numerical value, i.e. each of the possible digits can appear at a given place, in a given bit with the same probability). Since there can be only 90 such numbers (10-99), the amount of information will be $I = \log_2 90$ or approximately $I = 6.5$. Since in such numbers the significant first digit has 9 values (1-9), and the second has 10 values (0-9), then $I = \log_2 90 = \log_2 9 + \log_2 10$. The approximate value for $\log_2 10$ is 3.32. So, a message of one decimal unit carries 3.32 more information than one binary unit (than $\log_2 2 = 1$), and the second digit, for example, in the number aa, carries more information than the first one (if the digits and both digits are unknown; if these digits a are known, then there is no choice and the information is zero).

If we denote $f_i = -n\log_2 p_i$ in Shannon's formula, then we get that I can be understood as the arithmetic mean of f_i.

Hence, f_i can be interpreted as the informational content of the alphabet symbol with index i and value p_i of the probability of this symbol appearing in the message transmitting information.

Example. Let us consider the alphabet of two characters of the Russian language – "к" and "а". The relative frequencies of these letters in the frequency dictionary of the Russian language are, respectively, $p_1 = 0.028$, $p_2 = 0.062$. Take an arbitrary word p of length N from k letters "к" and m (k + m = N) letters "a" over this alphabet. The number of all such possible words, as follows from combinatorics, is n = N!/(k!m!). Let's estimate the amount of information in such a word: $I = \log_2 n = \ln n/\ln 2 = \log_2 e\,[\ln N! - \ln k! - \ln m!]$. Using the well-known Stirling formula (this formula, as is known from mathematical analysis, is quite accurate for large N, for example, for N > 100) - $N! \approx (N/e)^N$, or rather, its important consequence, – $\ln N! \approx N(\ln N - 1)$), we get an estimate of the amount of information (in bits) per 1 character of any word:

$$I_1 = I/N \approx (\log_2 e/N)[(k + m)(\ln N - 1) - k(\ln k - 1) - m(\ln m - 1)] =$$

$$= (\log_2 e/N)[k\ln(N/k) - m\ln(N/m)] =$$

$$= -\log_2 e[(k/N)\ln(k/N) + (m/N)\ln(m/N)] \leq$$

$$\leq -\log_2 e[p_1\ln p_1 + p_2\ln p_2]=$$

$$= -\log_2 e[0{,}028\ln 0{,}028 + 0{,}062\ln 0{,}062] \approx 0{,}235.$$

Example. The message contains 4 letters "а", 2 letters "б", 1 letter "и", 6 letters "р". Let's determine the amount of information in one such (of all possible) messages. The number N of different messages with a length of 13 letters will be equal to: $N = 13!/ (4!\times 2!\times 1!\times 6!) = 180180$. The amount of information I in one message will be equal to the value: $I = \log_2(N) = \log_2 180180 \approx 18$ (bit).

If k is the Boltzmann coefficient, known in physics as $k = 1.38 \times 10^{-16}$ erg/deg, then the expression

$$S = -k\sum_{i=1}^{n} p_i \ln p_i,$$

in thermodynamics known as entropy, or measure of chaos, disorder in a system. Comparing expressions I and S, we see that I can be understood as information entropy (entropy due to lack of information about/in the system).

L. Boltzmann gave a statistical definition of entropy in 1877 and noticed that entropy characterizes the missing information. 70 years later, K. Shannon formulated the postulates of information theory, and then it was noticed that Boltzmann's formula is invariant to information entropy, and their systemic connection, the consistency of these fundamental concepts was revealed.

It is important to note the following.

The maximum information corresponds to zero entropy. The basic relationship between entropy and information:

$$I + S\,(\log_2 e) / k = \text{const}$$

or in differential form

$$dI/dt = -((\log_2 e)/k)\, dS/dt.$$

When passing from state S_1 with information I_1 to state S_2 with information I_2, the following cases are possible:

a) *$S_1 < S_2$ ($I_1 > I_2$) – destruction (reduction) of old information in the system;*
b) *$S_1 = S_2$ ($I_1 = I_2$) – saving information in the system;*
c) *$S_1 > S_2$ ($I_1 < I_2$) – the birth of new (increase) information in the system.*

The main positive side of Shannon's formula is its abstraction from the semantic and qualitative, individual properties of the system. Unlike Hartley's formula, it takes into account the difference, the different probability of states – the formula has a statistical nature (takes into account the structure of messages), which makes this formula convenient for practical calculations. The main negative side of Shannon's formula is that it does not distinguish between states (with the same probability of reaching, for example), cannot evaluate the states of complex and open systems, and is applicable only for closed systems, abstracting from the meaning of information. Shannon's theory was developed as a theory of data transmission over communication channels, and Shannon's measure is a measure of the amount of data and does not reflect the semantic meaning.

An increase (decrease) in the Shannon measure indicates a decrease (increase) in the entropy (organization) of the system. In this case, entropy can be a measure of the disorganization of systems from complete chaos ($S = S^{max}$) and complete information uncertainty ($I = I^{min}$) to complete order ($S = S^{min}$) and complete information certainty ($I = I^{max}$) in the system.

4 Thermodynamic measure

The information-thermodynamic approach connects the value of the entropy of the system with the lack of information about the internal structure of the system (not replenished in principle, and not simply unrecorded). In this case, the number of states determines, in essence, the degree of incompleteness of our information about the system.

Let a thermodynamic system (process) S be given, and H_0, H_1 be the thermodynamic entropies of the system S in the initial (equilibrium) and final states of the thermodynamic process, respectively. Then the thermodynamic measure of information (negentropy) is determined by the formula:

$$H(H_0, H_1) = H_0 - H_1.$$

This formula is universal for any thermodynamic system. A decrease in $H(H_0, H_1)$ indicates the approach of the thermodynamic system S to a state of static equilibrium (given the resources available to it), while an increase indicates removal.

Let us pose a question about the state of a thermodynamic system. Let, before the start of the process, one can give p_1 equally probable answers to this question (none of which is preferable to the other), and after the end of the process – p_2 answers. Information change in this case:

$$\Delta I = k\ln(p_1/p_2) = k\,(\ln p_1 - \ln p_2).$$

If $p_1 > p_2$ ($\Delta I > 0$) – there is an increase in information, i.e. information about the system has become more definite, and for $p_1 < p_2$ ($\Delta I < 0$) – less definite. It is universal that we did not explicitly use the structure of the system (the mechanism of the process flow).

Example. Suppose there is a developing socio-economic system with 10 states, which as a result of evolution has developed to a system with 20 states. We are interested in the state of some component of the system (for example, an enterprise). At the beginning we knew the answer to the question and therefore $p_1 = 1$ ($\ln p_1 = 0$). The number of responses was proportional to [ln10]. After development, we already know the microeconomic state, i.e. the change in information about the state of the system is $\Delta I = -k\ln(20/10) = -k\ln 2$ (nat).

Example. Suppose that there is a thermodynamic system – a gas in a volume V, which expands to a volume of 2V (Fig. 6.1).

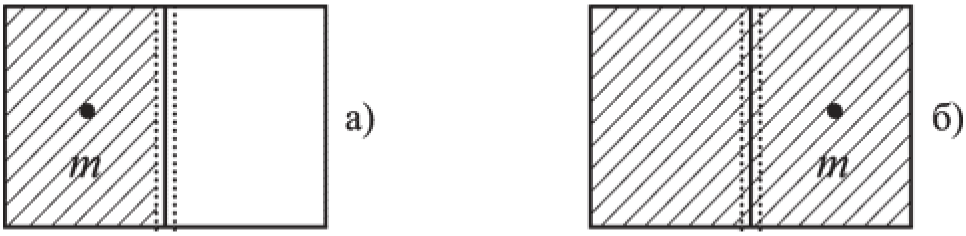

Figure 6.1. Gas volume V (a) expandable to 2V (b)

We are interested in the question of the coordinate of the gas molecule m. At the beginning (a) we knew the answer to the question and therefore $p_1 = 1$ ($\ln p_1 = 0$). The number of responses was proportional to lnV. After raising the damper, we already know this coordinate (microstate), i.e. change (decrease) in information about the state of the system will be equal to

$$\Delta I = -k\ln(2V/V) = -k\ln 2 \text{ (nat)}.$$

We have obtained an expression known in thermodynamics for the increase in entropy per molecule, and it confirms the second law of thermodynamics. Entropy is a measure of the lack of information about the microstate of a static system.

The value ΔI can be interpreted as the amount of information required to move from one level of system organization to another (with $\Delta I > 0$ – a higher, and with $\Delta I < 0$ – a lower level of organization).

The thermodynamic measure (entropy) is applicable to systems in thermal equilibrium. For systems that are far from thermal equilibrium, for example, living biological systems, the entropy measure is less suitable.

5 Energy-informational (quantum mechanical) measure

Energy (resource) and information (structure) are two fundamental characteristics of systems of the real world, connecting them material, spatial, temporal characteristics. If A is a named set with a carrier of the so-called "energetic origin", and B is a named set with a carrier of "informational origin", then we can define an energy-informational measure f: $A \rightarrow B$, for example, we can take a naming relation for a named set with a carrier (many names) A or B. The naming relation should reflect the mechanism of interconnections between physical-informational and material-energy structures and processes in the system.

Note that it is now more relevant to talk about bioenergy-informational measures reflecting the mechanism of interconnections between biophysical-informational and material-energy structures and processes in the system.

Example. The process of cell division is accompanied by the emission of energy quanta with frequencies up to approximately $N = 1.5\times10^{15}$ Hz. This spectrum can be perceived as a spectrum of functioning of the cell vocabulary as a bioinformation system. Using this spectrum, it is possible to encode up to 10^{15} different biochemical reactions, which is approximately 10^{7} times more than the number of reactions actually occurring in the cell (their number is approximately 10^{8}), i.e. the cell's vocabulary is redundant for effective recognition, classification, regulation of these reactions in the cell. The amount of information per 1 quantum of energy: $I = \log_2 10^{15} \approx 50$ bits. When cells are dividing, the amount of energy spent on transmitting 50 bits of information is equal to the energy of a quantum (h is Planck's constant, n is the radiation frequency):

$$E = h\nu = 6.62\times10^{-27} \text{ (erg/sec)} \times 0.5 \times 10^{15}(\text{sec}^{-1}) = 3.3\times10^{-12} \text{ (erg)}.$$

In this case, for 1W of "transmitter" power or $\mu = 10^{7}$ erg/sec. the number of quanta can be transferred:

$$n = \mu/E = 10^{7} \text{ (erg/sec)} / (3.3\times10^{-12} \text{ (erg)}) \approx 3.3\times10^{18} \text{ (quantum)}.$$

The total information transfer rate per 1 W of the power expended by the cell is determined by the number of different states of the cell N and the number of quanta (radiation) m:

$$V = n\log_2 N = 3.3\times10^{18}\times50 \approx 1.6\times10^{20} \text{ (bits / sec)}.$$

Any information is updated in a certain system. The material carrier of any system is a message, a signal. Any actualization is accompanied by a change in the energy properties (change of state) of the system. Our knowledge (and, consequently, the evolution of society) extends to the extent that information deepens and the possibility of its actualization improves.

6 Other measures of information.

Many authors have recently been considering various quantitative measures for measuring the meaning of information, for example, a measure based on the concept of a goal (A. Kharkevich and others); a measure based on the concept of thesaurus T = <X, Y, Z>, where X, Y, Z are sets, respectively, of names, meanings and meanings (pragmatics) of this knowledge (Yu. Schreider and others); a measure of the complexity of recovering binary words (A. Kolmogorov and others); measures of a posteriori knowledge (N. Wiener and others); the measure of the success of decision-making (N. Moiseev and others); measures of information similarity and diversity and other methods, approaches to the consideration of measures of information.

Example. As a measure (Kolmogorov) of recovering a binary word y from a given mapping f and given binary words x from a non-empty set X, we can take $H(f, y) = \min|x|$, $x \in X$, $f(x) = y$. Here $|x|$ is the length of the binary word x.

Example. If a priori it is known that some variable lies in the interval (0; 1), and a posteriori that it lies in the interval $(a; b) \subset (0; 1)$, then as a measure (Wiener) of the amount of information extracted from the posterior knowledge, we can take the ratio of the measure (a; b) to the measure (0; 1).

Example. In biological sciences, the so-called index measures, measures of species diversity, are widely used. The index is a measure of the state of the main biological, physicochemical, and other components of the system, which makes it possible to assess the strength of their impact on the system, the state and evolution of the system. Indexes should be relevant, general, interpretable, sensitive, minimally sufficient, high quality, widely used, rational. For example, an indicator of species diversity in a forest can be

$$\nu = \sqrt{p_1} + \sqrt{p_2} + \cdots + \sqrt{p_n},$$

where $p_1, p_2, ..., p_n$ are the frequencies of the community species inhabiting the forest, n is the number of species.

Conclusion

We considered various ways of introducing a measure for measuring the amount of information, their positive and negative sides, the relationship with changes in information in the system, examples.

Control questions

See the manual on the organization of students' independent work.