Anales PANEL'81/12 JAIIO
Sociedad Argentina de informática
e Investigación Operativa. Buenos Aires, 1981

ALGORITHMIC INFORMATION THEORY

TEORIA DE LA INFORMACION ALGORITMICA

Gregory J. Chaitin
IBM Thomas J. Watson Research Center
P. O. Box 218
Yorktown Heights, New York 10598, U.S.A.

Abstrac: We give a brief introduction to algorithmic information theory, a new field which combines ideas from information theory and from the theory of algorithms. We also briefly discuss the new light thrown by algorithmic information theory on foundational issues in the theories of probability and mathematical logic.

Resumen: Damos aquí una breve introducción a la teoría de la información algorítmica, un nuevo campo de estudios que combina ideas de la teoría de la información y de la teoría de los algorítmos. También nos referimos brevemente a la manera en que la teoría de la información algorítmica aclara ciertos problemas de base de las teorías de la probabilidad y de la lógica matemática.

The Shannon entropy concept of classical information theory [1] is an ensemble notion; it is a measure of the degree of ignorance concerning which possibility holds in an ensemble with a given a priori probability distribution:

$$H(p_1,\ldots,p_n) \equiv -\sum_{k=1}^{n} p_k \log_2 p_k.$$

In algorithmic information theory the primary concept is that of the *information content* of an individual object, which is a measure of how difficult it is to specify or describe how to construct or calculate that object. This notion is also known as *information-theoretic complexity*. For introductory expositions, see [2-4]. For the necessary background on computability theory and mathematical logic, see [5-7]. For a more technical survey of algorithmic information theory and a more complete bibliography, see [8]. See also [9].

The original formulation of the concept of algorithmic information is independently due to R. J. Solomonoff [10], A. N. Kolmogorov [11], and G. J. Chaitin [12]. The information content $I(x)$ of a binary string x is defined to be the size in bits (binary digits) of the smallest program for a canonical universal computer U to calculate x. (That the computer U is universal means that for any other computer M there is a prefix μ such that the program μp makes U do exactly the same computation that the program p makes M do.) The *joint information* $I(x, y)$ of two strings is defined to be the size of the smallest program that makes U calculate both of them. And the *conditional* or *relative information* $I(x \mid y)$ of x given y is defined to be the size of the smallest program for U to calculate x from y. The choice of the standard computer U introduces at most an $O(1)$ uncertainty in the numerical value of these concepts. ($O(f)$ is read "order of f" and denotes a function whose absolute value is bounded by a constant times f.)

With the original formulation of these definitions, for most x one has

$$I(x) = |x| + O(1) \tag{1}$$

(here $|x|$ denotes the length or size of the string x, in bits), but unfortunately

$$I(x, y) \leq I(x) + I(y) + O(1) \tag{2}$$

only holds if one replaces the $O(1)$ error estimate by $O(\log I(x)I(y))$.

Chaitin [13] and L. A. Levin [14] independently discovered how to reformulate these definitions so that the subadditivity property (2) holds. The change is to require that the set of meaningful computer programs be an instantaneous code, that is, that no program be a prefix of another. With this modification, (2) now holds, but instead of (1) most x satisfy

$$\begin{aligned} I(x) &= |x| + I(|x|) + O(1) \\ &= |x| + O(\log |x|). \end{aligned}$$

Moreover, in this theory the decomposition of the joint information of two objects into the sum of the information content of the first object added to the relative information of the second one given the first, has a different form than in classical information theory. In fact, instead of

$$I(x, y) = I(x) + I(y \mid x) + O(1), \tag{3}$$

one has

$$I(x, y) = I(x) + I(y \mid x, I(x)) + O(1). \tag{4}$$

That (3) is false follows from the fact that $I(x, I(x)) = I(x) + O(1)$ and $I(I(x) \mid x)$ is unbounded. This was noted by Chaitin [13] and studied more precisely by R. M. Solovay [13, p. 339] and P. Gač [15].

Two other concepts of algorithmic information theory are *mutual* or *common information* and *algorithmic independence*. Their importance has been emphasized by T. L. Fine [9, p. 141]. The mutual information content of two strings is defined as follows:

$$I(x : y) \equiv I(x) + I(y) - I(x, y).$$

In other words, the mutual information of two strings is the extent to which it is more economical to calculate them together than to calculate them separately. And x and y are said to be algorithmically independent if their mutual information $I(x : y)$ is essentially zero, that is, if $I(x, y)$ is approximately equal to $I(x) + I(y)$. Mutual information is symmetrical, i.e., $I(x : y) = I(y : x) + O(1)$. More important, from the decomposition (4) one obtains the following two alternative expressions for mutual information:

$$\begin{aligned} I(x : y) &= I(x) - I(x \mid y, I(y)) + O(1) \\ &= I(y) - I(y \mid x, I(x)) + O(1). \end{aligned}$$

Thus this notion of mutual information, although it applies to individual objects rather than to ensembles, nevertheless shares many of the formal properties of the classical version of this concept.

Up to this time there have been two principal applications of algorithmic information theory: (a) to provide a new conceptual foundation for probability theory and statistics by making it possible to rigorously define the notion of a *random sequence,* and (b) to provide an information-theoretic approach to metamathematics and the limitative theorems of mathematical logic. A possible application to theoretical mathematical biology is also mentioned below.

A random or patternless binary sequence x_n of length n may be defined to be one of maximal or near maximal complexity, that is, one whose complexity $I(x_n)$ is not much less than n. Similarly, an infinite binary sequence x may be defined to be random if its initial segments x_n are all random finite binary sequences. More precisely, x is random if and only if

$$\exists c \forall n [I(x_n) > n - c]. \tag{5}$$

In other words, the infinite sequence x is random if and only if there exists a c such that for all positive integers n, the algorithmic information content of the string consisting of the first n bits of the sequence x, is bounded from below by $n - c$. Similarly, a *random real number* may be defined to be one having the property that the base-two expansion of its fractional part is a random infinite binary sequence.

These definitions are intended to capture the intuitive notion of a lawless, chaotic, unstructured sequence. Sequences certified as random in this sense would be ideal for use in Monte Carlo calculations [16], and they would also be ideal as one-time pads for Vernam ciphers or as encription keys [17]. Unfortunately, as we shall see below, it is a variant of Gödel's famous incompleteness theorem that such certification is impossible. It is a corollary that no pseudo-random number generator can satisfy these definitions. Indeed, consider a real number x such as $\sqrt{2}$, π or e which has the property that it is possible to compute the successive binary digits of its base-two expansion. Such x satisfy

$$I(x_n) = I(n) + O(1) = O(\log n),$$

and are therefore maximally non-random. Nevertheless, most real numbers are random. In fact, if each bit of an infinite binary sequence is produced by an independent toss of an unbiased coin, then the probability that it will satisfy (5) is one. We shall now consider a particularly interesting random real number, Ω, discovered by Chaitin [13, p. 336].

A. M. Turing's theorem that the halting problem is unsolvable is a fundamental result of the theory of algorithms [4]. Turing's theorem states that there is no mechanical procedure for deciding whether or not an arbitrary program p eventually comes to halt when run on the universal computer U. Let Ω be the probability that the standard computer U eventually halts if each bit of its program p is produced by an independent toss of an unbiased coin. The unsolvability of the halting problem is intimately connected to the fact that the halting probability Ω is a random real number, i.e., its base-two expansion is a random infinite binary sequence in the very strong sense (5) defined above. From (5) it follows that Ω is normal (a notion due to É. Borel [18]), that Ω is a kollectiv with respect to all computable place selection rules (a concept due to R. von Mises and A. Church [19]), and it also follows that Ω satisfies all computable statistical tests of randomness (this notion being due to P. Martin-Löf [20]). An essay by C. H. Bennett on other remarkable properties of Ω, including its immunity to computable gambling schemes, is contained in [3].

K. Gödel established his famous incompleteness theorem by modifying the paradox of the liar: instead of "This statement is false" he considers "This statement is unprovable." The latter statement is true if and only if it is unprovable; it follows that not all true statements are theorems and thus that any formalization of mathematical logic is incomplete [5-7]. More relevant to algorithmic information theory is the paradox of "the smallest positive integer which cannot be specified in less than a billion words." The contradiction is that the phrase in quotes only has fourteen words even though at least a billion should be necessary. This is a version of the Berry paradox, first published by B. Russell [6, p. 153]. To obtain a theorem rather than a contradiction, one considers instead "the binary string s which has the shortest proof that its complexity $I(s)$ is greater than a billion." The point is that this string s cannot exist. This leads one to the metatheorem that although most bit strings are random and have information content approximately equal to their lengths, it is impossible to prove that a specific string has information content greater than n unless one is using at least n bits of axioms. See [4] for a more complete exposition of this information-theoretic version of

Gödel's incompleteness theorem, which was first presented in [21]. It can also be shown that n bits of assumptions or postulates are needed to be able to determine the first n bits of the base-two expansion of the real number Ω.

Finally, it should be pointed out that these concepts are potentially relevant to biology. The algorithmic approach is closer to the intuitive notion of the information content of a biological organism than is the classical ensemble viewpoint, for the role of a computer program and of DNA are roughly analogous. [22] discusses possible applications of the concept of mutual algorithmic information to theoretical biology; it is suggested that a living organism might be defined as a highly correlated region, one whose parts have high mutual information. See also [23].

REFERENCES

General References

[1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, 1949. (The first and still one of the very best books on classical information theory.)

[2] G. J. Chaitin, "Randomness and mathematical proof," *Scientific American* **232**, No. 5 (May 1975), pp. 47-52. (An introduction to algorithmic information theory emphasizing the meaning of the basic concepts.)

[3] M. Gardner, "The random number Ω bids fair to hold the mysteries of the universe," Mathematical Games Dept., *Scientific American* **241**, No. 5 (Nov. 1979), pp. 20-34. (An introduction to algorithmic information theory emphasizing the fundamental role played by Ω.)

[4] M. Davis, "What is a computation?" in *Mathematics Today: Twelve Informal Essays,* L. A. Steen (ed.), Springer-Verlag, New York, 1978, pp. 241-267. (An introduction to algorithmic information theory largely devoted to a detailed presentation of the relevant background in computability theory and mathematical logic.)

[5] D. R. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid,* Basic Books, New York, 1979. (The longest and most lucid introduction to computability theory and mathematical logic.)

[6] J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931,* Harvard University Press, Cambridge, 1977. (This book and the next together comprise a stimulating collection of all the classic papers on computability theory and mathematical logic.)

[7] M. Davis (ed.), *The Undecidable – Basic Papers on Undecidable Propositions, Unsolvable Problems And Computable Functions,* Raven Press, Hewlett, 1965.

[8] G. J. Chaitin, "Algorithmic information theory," *IBM Journal of Research and Development* **21** (1977), pp. 350-359, p. 496. (A survey of algorithmic information theory.)

[9] T. L. Fine, *Theories of Probability: An Examination of Foundations,* Academic Press, New York, 1973. (A survey of the remarkably diverse proposals which have been made for formulating probability mathematically. Caution: the material on algorithmic information theory contains some inaccuracies, and it is also somewhat dated due to recent rapid progress in this field.)

Technical References

[10] R. J. Solomonoff, "A formal theory of inductive inference," *Information and Control* **7** (1964), pp. 1-22, pp. 224-254.

[11] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," *Problems of Information Transmission* **1** (1965), pp. 1-7.

[12] G. J. Chaitin, "On the length of programs for computing finite binary sequences," *Journal of the ACM* **13** (1966), pp. 547-569; **16** (1969), pp. 145-159.

[13] —, "A theory of program size formally identical to information theory," *Journal of the ACM* **22** (1975), pp. 329-340.

[14] L. A. Levin, "Laws of information conservation (nongrowth) and aspects of the foundation of probability theory," *Problems of Information Transmission* **10** (1974), pp. 206-210.

[15] P. Gač, "On the symmetry of algorithmic information," *Soviet Mathematics – Doklady* **15** (1974), pp. 1477-1480.

[16] G. J. Chaitin and J. T. Schwartz, "A note on Monte Carlo primality tests and algorithmic information theory," *Communications on Pure and Applied Mathematics* **31** (1978), pp. 521-527.

[17] H. Feistel, "Cryptography and computer privacy," *Scientific American* **228**, No. 5 (May 1973), pp. 15-23.

[18] M. Kac, *Statistical Independence in Probability, Analysis and Number Theory,* Mathematical Association of America, 1959.

[19] A. Church, "On the concept of a random sequence," *Bulletin of the AMS* **46** (1940), pp. 130-135.

[20] P. Martin-Löf, "The definition of random sequences," *Information and Control* **9** (1966), pp. 602-619.

[21] G. J. Chaitin, "Information-theoretic computational complexity," *IEEE Transactions on Information Theory* **IT-20** (1974), pp. 10-15.

[22] —, "Toward a mathematical definition of 'life'," in *The Maximum Entropy Formalism,* R. D. Levine and M. Tribus (eds.), MIT Press, Cambridge, 1979, pp. 477-498.

[23] F. Papentin, "Complexity of snowflakes," *Naturwissenschaften* **67** (1980), pp. 174-177.