Algorithms **2010**, *3*, 260-264; doi:10.3390/algor3030260

OPEN ACCESS

algorithms

ISSN 1999-4893

www.mdpi.com/journal/algorithms

Obituary

Ray Solomonoff, Founding Father of Algorithmic Information Theory

Paul M.B. Vitanyi

CWI, Science Park 123, Amsterdam 1098 XG, The Netherlands; E-Mail: Paul.Vitanyi@cwi.nl

Received: 12 March 2010 / Accepted: 14 March 2010 / Published: 20 July 2010

Ray J. Solomonoff died on December 7, 2009, in Cambridge, Massachusetts, of complications of a stroke caused by an aneurism in his head. Ray was the first inventor of Algorithmic Information Theory which deals with the shortest effective description length of objects and is commonly designated by the term "Kolmogorov complexity."

In the 1950s Solomonoff was one of the first researchers to treat probabilistic grammars and the associated languages. He treated probabilistic Artificial Intelligence (AI) when "probabilistic" was unfashionable, and treated questions of machine learning early on. But his greatest contribution is the creation of Algorithmic Information Theory.

Already in November 1960, Solomonoff published [1] that presented the basic ideas of Algorithmic Information Theory as a means to overcome serious problems associated with the application of Bayes's rule in statistics.

Ray Solomonoff was born on July 25, 1926, in Cleveland, Ohio, United States. He studied physics during 1946–1950 at the University of Chicago (he recalls the lectures of E. Fermi and R. Carnap) and obtained an M.Sc. from that university. From 1951–1958 he worked in the electronics industry doing math and physics and designing analog computers, working half-time. His scientific autobiography up to 1997 is published as [2].

Solomonoff's objective was to formulate a completely general theory of inductive reasoning that would overcome shortcomings in Carnap's [3]. Following some more technical reports, in a long journal paper in two parts he introduced 'Kolmogorov' complexity as an auxiliary concept to obtain a universal a priori probability and proved the invariance theorem governing Algorithmic Information Theory [4]. The mathematical setting of these ideas is described in some detail below.

Solomonoff's work has led to a novel approach in statistics [5] leading to applicable inference procedures such as the minimal description length principle. J.J. Rissanen, credited with the latter, explicitly relates that his invention is based on Solomonoff's work with the idea of applying it to classical statistical inference [6].

Since Solomonoff is the first inventor of Algorithmic Information Theory one can raise the question whether we ought to talk about 'Solomonoff complexity.' However, the name 'Kolmogorov complexity' for shortest effective description length has become well entrenched and is commonly understood. Solomonoff's publications apparently received little attention until the great Soviet mathematician A.N. Kolmogorov started to refer to them from 1968 onward. Says Kolmogorov, "I came to similar conclusions [as Solomonoff], before becoming aware of Solomonoff's work, in 1963–1964" and "The basic discovery, which I have accomplished independently from and simultaneously with R. Solomonoff, lies in the fact that the theory of algorithms enables us to eliminate this arbitrariness by the determination of a 'complexity' which is almost invariant (the replacement of one method by another leads only to the supplement of the bounded term)."

Solomonoff's early papers contain in veiled form suggestions about randomness of finite strings, incomputability of Kolmogorov complexity, computability of approximations to the Kolmogorov complexity, and resource-bounded Kolmogorov complexity. Marvin Minsky referred to Solomonoff's work already in the early 1960s. To our knowledge, these are the earliest documents outlining an algorithmic theory of descriptions.

A.N. Kolmogorov's later introduction of complexity was motivated by information theory and problems of randomness. Solomonoff introduced algorithmic complexity independently and earlier and for a different reason: inductive reasoning. Universal a priori probability, in the sense of a single prior probability that can be substituted for each actual prior probability in Bayes's rule was invented by Solomonoff with Kolmogorov complexity as a side product, several years before anybody else did.

R.J. Solomonoff obtained a Ph.B. (bachelor of philosophy) and M.Sc. in physics. He was already interested in problems of inductive inference and exchanged viewpoints with the resident philosopher of science at the University of Chicago, Rudolf Carnap, who taught an influential course in probability theory. In 1956, Solomonoff was one of the 10 or so attendees of the Dartmouth Summer Study Group on Artificial Intelligence, at Dartmouth College in Hanover, New Hampshire, organized by M. Minsky, J. McCarthy, and C.E. Shannon, and in fact stayed on to spend the whole summer there. (This meeting gave AI its name.) There Solomonoff wrote a memo on inductive inference. McCarthy had the idea that given every mathematical problem, it could be brought into the form of "given a machine and a desired output, find an input from which the machine computes that output". Solomonoff suggested that there was a class of problems that was not of that form: "given an initial segment of a sequence, predict its continuation." McCarthy then thought that if one saw a machine producing the initial segment, and then continuing past that point, would one not think that the continuation was a reasonable extrapolation? With that the idea got stuck, and the participants left it at that.

Later, Solomonoff presented the paper "An Inductive Inference Machine" at the IEEE Symposium on Information Theory, 1956, describing a program to unsupervisedly learn arithmetic formulas from examples. At the same meeting, there was a talk by N. Chomsky based on his paper [7]. The latter talk started Solomonoff thinking anew about formal machines in induction. In about 1958 he left his half-time position in industry and joined Zator Company full time, a small research outfit located in some rooms at 140 1/2 Mount Auburn Street, Cambridge, Massachusetts, which had been founded by Calvin Mooers around 1954 for the purpose of developing information retrieval technology. Floating mainly on military funding, Zator Co. was a research front organization, employing Mooers,

Solomonoff, Mooers's wife, and a secretary, as well as at various times visitors such as Marvin Minsky. It changed its name to the more martial sounding Rockford Research (Rockford, Illinois, was a place where Mooers had lived) around 1962. In 1968 the US Government reacted to public pressure (related to the Vietnam War) by abolishing defense funding of civil research, and Rockford floundered. Being out of a job, Solomonoff left and founded his own (one-man) company, Oxbridge Research, in Cambridge in 1970, and has been there ever since, apart from spending nine months as research associate at MIT's Artificial Intelligence Laboratory, 1990–1991 at the University of Saarland, Saarbruecken, Germany, and a more recent sabbatical at IDSIA, Lugano, Switzerland.

In 1960 Solomonoff published [1] in which he gave an outline of the notion of universal a priori probability and how to use it in inductive reasoning (rather, prediction) according to Bayes's rule. This was sent out to all contractors of the Air Force who were even vaguely interested in this subject. In [4] Solomonoff developed these ideas further and defined the notion of enumeration of monotone machines and a notion of universal a priori probability based on the universal monotone machine. In this way, it came about that the original incentive to develop a theory of algorithmic information content of individual objects was Solomonoff's invention of a universal a priori probability that can be used instead of the actual a priori probability in applying Bayes's rule.

In Solomonoff's original approach he used Turing machines with markers that delimit the input. In this model, the a priori probability of a string x is the sum of the uniform probabilities of inputs from which the Turing machine computes output x. To obtain the universal a priori probability one uses a universal Turing machine. But this is improper since the universal probabilities do not converge, and the universal probability is not a proper probability at all. For prediction one uses not the universal a priori probability which is a probability mass function, but a semimeasure which is a weak form of a measure. Even using the mathematical framework developed by L.A. Levin [8] the problem remains that the universal a priori probability for prediction is a semimeasure but not a measure. (the probability concentrated on all extensions of the empty string is less or equal to 1, and the probability concentrated on all extensions of a nonempty string x is less or equal to the probability concentrated on x.) To solve this problem Solomonoff in 1964 suggested, and in 1978 exhibited, a normalization. However, the resulting measure is not even lower semicomputable like the original a priori probability. According to Solomonoff this is a small price to pay. In fact, in some applications we may like the measure property and do not care about semicomputability at all. The universal measure or semimeasure has remarkable properties and applications which we will not go into here.
Leonid A. Levin [8] in 1970 gave a mathematical expression of the universal a priori probability as a universal (that is, maximal) lower semicomputable semimeasure M, and showed that the negative logarithm of M(x) coincides with the Kolmogorov complexity of x up to an additive logarithmic term. Solomonoff's version of a machine which by L.A. Levin's work was transformed in the mathematically sound version we now call "monotonic", which name we will use henceforth.

One of Solomonoff's striking contributions is the simple-looking inductive inference formula M(xy)/M(x) (of course, with the caveat of incomputability). Here M(z) stands for the universal a priori probability concentrated on z. An example is as folows. Suppose we have an infinite binary sequence every odd bit of which is uniformly random and every even bit is a bit of pi = 3.1415... written in binary. If x is a growing initial segment of this sequence and y is the next bit, then M(xy)/M(x)

eventually predicts the even bits (those of pi) almost certainly and the odd bits with probability going to 1/2.

In 1978, Solomonoff [9] proved the one major result justifying the use of the universal a priori probability: If we predict the next elements of an infinite binary sequence drawn from any computable measure using the single universal lower semicomputable semimeasure M, then the total expected squared error (of predicting 0 instead of 1 or vice versa) over all infinitely many predictions is bounded by a constant related to the Kolmogorov complexity of the computable measure. The expectation is taken over the computable measure. This means that if the expected squared error in the n-th prediction is a monotonic nonincreasing function of n, then it goes down faster than 1/n. This is better than any classical predictive strategy can do. Since the universal a priori distribution is a weighted sum of all lower semicomputable distributions, and is itself one, this theorem in effect states a convergence result about "expert learning" now fashionable in computational machine learning.

The topic has spawned an elaborate theory of prediction in both static and reactive unknown environments, based on universal distributions with arbitrary loss bounds (rather than just the logarithmic loss) using extensions and variations of the proof method, inspiring information theorists such as T.M. Cover. An example is the textbook of M. Hutter [10]. Computable approximations are investigated by V. Vovk among others.

In his later years Solomonoff tried to find a practical learning algorithm. He experimented much with "training sequences" using "conceptual jump size" (now called "chunking" in AI). He continued to believe in the existence of a learning algorithm that one should find and found the approach used for example in practical speech recognition misguided, since there the algorithm may have 2000 tuneable real number parameters. In the 1990s Ray Solomonoff started a company to predict stock performance on a scientific basis provided by his theories. Eventually, he dropped the venture claiming that “convergence was not fast enough.”

It is unusual to find a productive major scientist that is not regularly employed at all. But from all the elder people (not only scientists) I know, Ray Solomonoff was the happiest, the most inquisitive, and the most satisfied. He continued publishing papers right up to his death at 83.

The question of normalization of the universal a priori semimeasure continued to haunt Solomonoff. In about 1992 R.M. Solovay proved that every normalization of the universal a priori semimeasure to a measure would change the relative probabilities of extensions by more than a constant (even incomputably large) factor. In a recent paper with a clever and appealing proof, Solomonoff [11] proved that if we predict a computable measure with a the universal a priori semimeasure normalized according to his prescription, then the bad changes a la Solovay happen only with expectation going fast to 0 with growing length of the predicted sequence, the expectation taken with respect to the computable measure.

In 2003 he was the first recipient of the Kolmogorov Award by The Computer Learning Research Center at the Royal Holloway, University of London, where he gave the inaugural Kolmogorov Lecture. Solomonoff was a visiting Professor at the CLRC. A list of his publications (published and unpublished) is at http://world.std.com/~rjs/pubs.html.

Ray Solomonoff is survived by his wife, Grace Morton, 72 Winter Street, Arlington, MA 02474, and by his nephews, Alex Solomonoff, of Somerville, and David Solomonoff.

References

1. Solomonoff, R.J. *A Preliminary Report on a General Theory of Inductive Inference*. Technical Report ZTB-138; Zator Company: Cambridge, MA, USA, November 1960.
2. Solomonoff, R.J. The discovery of algorithmic probability. *J. Comput. Syst. Sci.* **1997**, *55*, 73–88.
3. Carnap, R. *Logical Foundations of Probability*; Univ. Chicago Press: Chicago, IL, USA, 1950.
4. Solomonoff, R.J. A formal theory of inductive inference: Parts 1 and 2. *Inf. Contr.* **1964**, *7*, 1–22, 224–254.
5. Fine, T.L. *Theories of Probability*; Academic Press: San Diego, CA,USA, 1973.
6. Rissanen, J.J. A universal prior for integers and estimation by minimum description length. *Ann. Stat.* **1983**, *11*, 416–431.
7. Chomsky, N. Three models for the description of language. *IRE Trans. Inf. Theory* **1956**, *2*, 113–126.
8. Zvonkin, A.K.; Levin, L.A. The complexity of finite objects and the development of concepts of information and randomness by means of the theory algorithms. *Russian Math. Surveys* **1970**, *25*, 83–124.
9. Solomonoff, R.J. Complexity-based induction systems: comparisons and convergence theorems. *IEEE Trans. Inf. Theory* **1978**, *24*, 422–432.
10. Hutter, M. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*; Springer-Verlag: Berlin, Germany, 2005.
11. Solomonoff, R.J. The probability of "undefined" (non-converging) output in generating the universal probability distribution. *Inf. Proc. Lett.* **2008**, *106*, 238–240.

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).