Solomonoff's theory of inductive inference proves that, under its common sense assumptions (axioms), the best possible scientific model is the shortest algorithm that generates the empirical data under consideration. In addition to the choice of data, other assumptions are that, to avoid the post-hoc fallacy, the programming language must be chosen prior to the data [1] and that the environment being observed is generated by an unknown algorithm. This is also called a theory of induction . Due to its basis in the dynamical (state-space model) character of Algorithmic Information Theory , it encompasses statistical as well as dynamical information criteria for model selection . It was introduced by Ray Solomonoff , based on probability theory and theoretical computer science . [2] [3] In essence, Solomonoff's induction derives the posterior probability of any computable theory, given a sequence of observed data. This posterior probability is derived from Bayes' rule and some universal prior, that is, a prior that assigns a positive probability to any computable theory.

Solomonoff proved that this induction is incomputable (or more precisely, lower semi-computable), but noted that "this incomputability is of a very benign kind", and that it "in no way inhibits its use for practical prediction" (as it can be approximated from below more accurately with more computational resources). [2] It is only "incomputable" in the benign sense that no scientific consensus is able to prove that the best current scientific theory is the best of all possible theories. However, Solomonoff's theory does provide an objective criterion for deciding among the current scientific theories explaining a given set of observations.

Solomonoff's induction naturally formalizes Occam's razor [4] [5] [6] [7] [8] by assigning larger prior credences to theories that require a shorter algorithmic description.

Origin

Philosophical

The theory is based in philosophical foundations, and was founded by Ray Solomonoff around 1960. [9] It is a mathematically formalized combination of Occam's razor [4] [5] [6] [7] [8] and the Principle of Multiple Explanations . [10] All computable theories which perfectly describe previous observations are used to calculate the probability of the next observation, with more weight put on the shorter computable theories. Marcus Hutter 's universal artificial intelligence builds upon this to calculate the expected value of an action.

Principle

Solomonoff's induction has been argued to be the computational formalization of pure Bayesianism . [3] To understand, recall that Bayesianism derives the posterior probability $\mathbb{P}[T|D]$ of a theory T given data D by applying Bayes rule, which yields

where theories A are alternatives to theory T . For this equation to make sense, the quantities $\mathbb{P}[D|T]$ and $\mathbb{P}[D|A]$ must be well-defined for all theories T and A . In other words, any theory must define a probability distribution over observable data D . Solomonoff's induction essentially boils down to demanding that all such probability distributions be computable .

Interestingly, the set of computable probability distributions is a subset of the set of all programs, which is countable . Similarly, the sets of observable data considered by Solomonoff were finite. Without loss of generality , we can thus consider that any observable data is a finite bit string . As a result, Solomonoff's induction can be defined by only invoking discrete probability distributions.

Solomonoff's induction then allows to make probabilistic predictions of future data F {\displaystyle F} , by simply obeying the laws of probability. Namely, we have $P[F|D] = E_T[P[F|T,D]] = \sum_T P[F|T,D]P[T|D]$ {\displaystyle \mathbb {P} [F|D]=\mathbb {E} _{T}[\mathbb {P} [F|T,D]]=\sum _{T}\mathbb {P} [F|T,D]\mathbb {P} [T|D]} . This quantity can be interpreted as the average predictions $P[F|T,D]$ {\displaystyle \mathbb {P} [F|T,D]} of all theories T {\displaystyle T} given past data D {\displaystyle D} , weighted by their posterior credences $P[T|D]$ {\displaystyle \mathbb {P} [T|D]} .

Mathematical

The proof of the "razor" is based on the known mathematical properties of a probability distribution over a countable set . These properties are relevant because the infinite set of all programs is a denumerable set. The sum S of the probabilities of all programs must be exactly equal to one (as per the definition of probability) thus the probabilities must roughly decrease as we enumerate the infinite set of all programs, otherwise S will be strictly greater than one. To be more precise, for every ■ {\displaystyle \epsilon } > 0, there is some length l such that the probability of all programs longer than l is at most ■ {\displaystyle \epsilon } . This does not, however, preclude very long programs from having very high probability.

Fundamental ingredients of the theory are the concepts of algorithmic probability and Kolmogorov complexity . The universal prior probability of any prefix p of a computable sequence x is the sum of the probabilities of all programs (for a universal computer) that compute something starting with p . Given some p and any computable but unknown probability distribution from which x is sampled, the universal prior and Bayes' theorem can be used to predict the yet unseen parts of x in optimal fashion.

Mathematical guarantees

Solomonoff's completeness

The remarkable property of Solomonoff's induction is its completeness. In essence, the completeness theorem guarantees that the expected cumulative errors made by the predictions based on Solomonoff's induction are upper-bounded by the Kolmogorov complexity of the (stochastic) data generating process. The errors can be measured using the Kullback–Leibler divergence or the square of the difference between the induction's prediction and the probability assigned by the (stochastic) data generating process.

Solomonoff's uncomputability

Unfortunately, Solomonoff also proved that Solomonoff's induction is uncomputable. In fact, he showed that computability and completeness are mutually exclusive: any complete theory must be uncomputable. The proof of this is derived from a game between the induction and the environment. Essentially, any computable induction can be tricked by a computable environment, by choosing the computable environment that negates the computable induction's prediction. This fact can be regarded as an instance of the no free lunch theorem .

Modern applications

Artificial intelligence

Though Solomonoff's inductive inference is not computable , several AIXI -derived algorithms approximate it in order to make it run on a modern computer. The more computing power they are given, the closer their predictions are to the predictions of inductive inference (their mathematical limit is Solomonoff's inductive inference). [11] [12] [13]

Another direction of inductive inference is based on E. Mark Gold 's model of learning in the limit from 1967 and has developed since then more and more models of learning. [14] The general scenario is the following: Given a class S of computable functions, is there a learner (that is,

recursive functional) which for any input of the form (f (0), f (1),..., f (n)) outputs a hypothesis (an index e with respect to a previously agreed on acceptable numbering of all computable functions; the indexed function may be required consistent with the given values of f). A learner M learns a function f if almost all its hypotheses are the same index e , which generates the function f ; M learns S if M learns every f in S . Basic results are that all recursively enumerable classes of functions are learnable while the class REC of all computable functions is not learnable. [citation needed] Many related models have been considered and also the learning of classes of recursively enumerable sets from positive data is a topic studied from Gold's pioneering paper in 1967 onwards. A far reaching extension of the Gold's approach is developed by Schmidhuber's theory of generalized Kolmogorov complexities, [15] which are kinds of super-recursive algorithms .

See also

Algorithmic information theory

Bayesian inference

Inductive inference

Inductive probability

Mill's methods

Minimum description length

Minimum message length

For a philosophical viewpoint, see: Problem of induction and New riddle of induction

References

Sources

Angluin, Dana; Smith, Carl H. (Sep 1983). "Inductive Inference: Theory and Methods" . Computing Surveys . 15 (3): 237– 269. doi : 10.1145/356914.356918 . S2CID 3209224 .

Burgin, M. (2005), Super-recursive Algorithms , Monographs in computer science, Springer. ISBN 0-387-95569-0

Burgin, M., "How We Know What Technology Can Do", Communications of the ACM , v. 44, No. 11, 2001, pp. 82–88.

Burgin, M.; Eberbach, E., "Universality for Turing Machines, Inductive Turing Machines and Evolutionary Algorithms", Fundamenta Informaticae , v. 91, No. 1, 2009, 53–77.

Burgin, M.; Eberbach, E., "On Foundations of Evolutionary Computation: An Evolutionary Automata Approach", in Handbook of Research on Artificial Immune Systems and Natural Computing: Applying Complex Adaptive Technologies (Hongwei Mo, Ed.), IGI Global, Hershey, Pennsylvania, 2009, 342–360.

Burgin, M.; Eberbach, E., "Evolutionary Automata: Expressiveness and Convergence of Evolutionary Computation", Computer Journal , v. 55, No. 9, 2012, pp. 1023–1029.

Burgin, M.; Klinger, A. Experience, Generations, and Limits in Machine Learning, Theoretical Computer Science , v. 317, No. 1/3, 2004, pp. 71–91

Davis, Martin (2006) "The Church–Turing Thesis: Consensus and opposition]". Proceedings, Computability in Europe 2006. Lecture Notes in Computer Science, 3988 pp. 125–132.

Gasarch, W. ; Smith, C. H. (1997) "A survey of inductive inference with an emphasis on queries". Complexity, logic, and recursion theory , Lecture Notes in Pure and Appl. Math., 187, Dekker, New York, pp. 225–260.

Hay, Nick. " Universal Semimeasures: An Introduction ," CDMTCS Research Report Series, University of Auckland, Feb. 2007.

Jain, Sanjay; Osherson, Daniel; Royer, James; Sharma, Arun, Systems that Learn: An Introduction to Learning Theory (second edition), MIT Press , 1999.

Kleene, Stephen C. (1952), Introduction to Metamathematics (First ed.), Amsterdam: North-Holland
.

Li Ming; Vitanyi, Paul, An Introduction to Kolmogorov Complexity and Its Applications , 2nd Edition, Springer Verlag, 1997.

Osherson, Daniel; Stob, Michael; Weinstein, Scott, Systems That Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists , MIT Press , 1986.

Solomonoff, Ray J. (1999). "Two Kinds of Probabilistic Induction" (PDF) . The Computer Journal . 42 (4): 256. CiteSeerX 10.1.1.68.8941 . doi : 10.1093/comjnl/42.4.256 .

Solomonoff, Ray (March 1964). "A Formal Theory of Inductive Inference Part I" (PDF) . Information and Control . 7 (1): 1– 22. doi : 10.1016/S0019-9958(64)90223-2 .

Solomonoff, Ray (June 1964). "A Formal Theory of Inductive Inference Part II" (PDF) . Information and Control . 7 (2): 224– 254. doi : 10.1016/S0019-9958(64)90131-7 .

External links

Algorithmic probability – Scholarpedia