Unified Information Gain Measures for Inference Processes

Jose Hernandez-Orallo[1]

Universitat Politècnica de València, Dep. de Sistemes Informàtics i Computació.
Camí de Vera 14, Aptat. 22.012 E-46071, València, Spain. E-mail:jorallo@dsic.upv.es
Also at the Dep. of Logic of the University of Valencia. KGS Member.

Abstract. We present a definition of information gain based on descriptional complexity, which applies to different inference processes, either deductive or inductive, and evaluates the relative value of new inference results, weighting the convenience of leaving them implicitly or explicitly.

Concretely, processes which are apparently so different as induction and deduction can be explained in a computational framework as inference processes that *just* generate an output from an input, which must follow some semantical restrictions and/or selection criteria, widely studied in philosophy of science and mathematical logic, respectively. The term information is seen as the result of a computational effort, analogically to the way energy is seen as the result of a physical work. This suggests many questions, especially how to measure this computational effort. The *answer* was given by Levin in the seventies, proving that the weighting $LT(x) = length(x) + logCost(x)$ between space and time was optimal in the sense of universal search problems. Given two objects, the effort from x to y in a computational system ϕ is then measured as the relative Levin-descriptional complexity $Kt(y|x) = min\{LT(p) : \phi(< p, x >) = y\}$.

The Information Gain of object y wrt. object x is then defined as the quotient between the effort which is necessary to describe y from x and the effort which is necessary to describe y alone. More formally, $G(y|x) = Kt(y|x)/Kt(y)$. Some properties of this measure are shown before applying it to inference processes. In the case of induction, information gain represents how informative is the hypothesis wrt. the evidence (in Popper's sense) and it is compared with other selection criteria, especially simplicity. This leads to the notion of *authentic learning*, quite different from Gold's identification. In the case of deduction, the measure also represents how informative is the conclusion from the premises. This establishes a generic measure of the gain which is obtained from making explicit something that was implicit, provided that the system is not omniscient and resource limited, where there is a clear difference between the explicit or surface information, and implicit or depth information, as it was highlighted by Hintikka for first-order logic.

We study optimal compromises between the size of a theory and its explicitness, formalising the necessity of lemmata and the use of extensional properties for mathematical practice, in order to avoid difficult derivations that were already done (while still maintaining under control the whole size of the theory).

References

1. Bar-Hillel, Y.; Carnap, R. "Semantic Information". *British Journal for the Philosophy of Science* 4, 1953, 147-157.
2. Bibel, Wolfgang "Deduction. Automated Logic" *Academic Press Limited* 1993.
3. Blum, M. "A machine-independent theory of the complexity of recursive functions". *J. ACM 14, 4*, 322-6, 1967.
4. Blum, L.; Blum, M. "Towards a mathematical theory of inductive inference". *Inform. and Control 28*, pp. 125-155, 1975.
5. Board, R.; Pitt, L. "On the necessity of Occam algorithms". *In Proc., 22nd ACM Symp. Theory of Comp.*, 1990.
6. Carnap, R. "The Continuum of Inductive Methods" University of Chicago, Chicago, IL, 1952.
7. Chaitin, Gregory J. "Information-theoretic limitations of formal systems" *Journal of the ACM*, 21, 403-424., 1974.
8. Chaitin, G.J. "Algorithmic Information Theory". *Fourth printing, Cambridge University Press*, 1992.
9. Freivalds, R.; Kinber, E.; Smith, C.H. "On the Intrinsic Complexity of Learning". *Inf. and Control 123*, pp. 64-71, 1995.
10. Gold, E. M. "Language Identification in the Limit". *Inform and Control, 10*, pp. 447-474, 1967.
11. Hernandez-Orallo, J. ; Garcia-Varea, I. "Distinguishing Abduction and Induction under Intensional Complexity" in Flach, P.; Kakas, A. (eds.) European Conference of Artificial Intelligence (ECAI'98) Workshop on Abduction and Induction in AI, pp. 41-48, Brighton 1998. .
12. Hernandez-Orallo, J. ; Garcia-Varea, I. "On Autistic Interpretations of Occam's Razor" S.Rini, G.Poletti (eds.) Proceedings of the 1998 International Conference on Model Based Reasoning (MBR'98), pp. 25-27, Pavia 1998.
13. Hintikka, Jaakko; Suppes, Patrick "Surface Information and Depth Information" in Hintikka, Jaako; Suppes, Patrick *Information and Inference* Reidel Pub. Company 1970.
14. Hintikka, Jaakko "Logic, Language-Games and Information" *The Calrendon Press*, Oxford University Press 1973.
15. Kolmogorov, A.N. "Three Approaches to the Quantitative Definition of Information" *Problems Inform. Transmission* , 1(1):1-7, 1965.
16. Levin, L.A. "Universal search problems". *Problems Inform. Transmission, 9*, pp. 265-266, 1973.
17. Li, M.; Vitanyi, P. "An Introduction to Kolmogorov Complexity and its Applications" *2nd Ed. Springer-Verlag*, 1997.
18. Popper, K.R. "Conjectures and Refutations" London 1969.
19. Rissanen, J. "Modelling by the shortest data description" *Automatica-J.IFAC, 14*, pp. 465-471, 1978.
20. Solomonoff, R.J. "A formal theory of inductive inference". *Inf. Control.* Vol. 7, 1-22, Mar., pp. 224-254, June 1964.
21. Wolff, J.G. "Computing as Compression: An Overview of the SP Theory and System". *New Gen. Computing 13*, pp. 187-214, 1995.

This article was processed using the LaTeX macro package with LLNCS style