

Distinguished Lecture 2002

Department of Computer Science

The Similarity Metric



Paul Vitanyi

Centrum Voor Wiskunde en Informatica,
Amsterdam

December 10, 3:30 p.m.
1414 Molecular Biology

A new class of metrics appropriate for measuring effective similarity relations between sequences, say one type of similarity per metric, is studied. We propose a new "normalized information distance", based on the noncomputable notion of Kolmogorov complexity, and show that it minorizes every metric in the class (that is, it is universal in that it discovers all effective similarities). We demonstrate that it too is a metric and takes values in $[0,1]$; hence it may be called the **similarity metric**. This is a theory foundation for a new general practical tool. We give two distinctive applications in widely divergent areas (the experiments by necessity use just computable approximations to the target notions). First, we computationally compare whole mitochondrial genomes and infer their evolutionary history. This results in a first completely automatic computed whole mitochondrial phylogeny tree. Secondly, we give fully automatically computed language tree of 52 different language based on translated versions of the "Universal Declaration of Human Rights".

This is joint work with Ming Li, Xin Chen, Xin Li, Bin Ma, and will be presented at the 14th ACM-SIAM Symposium on Discrete Algorithms, Baltimore, 2003. It can be accessed at <http://www.cwi.nl/~paulv/selection.html>

Paul M.B. Vitanyi received his Ph.D. from the Free University of Amsterdam (1978) and he holds positions at the national CWI Research Institute in Amsterdam and is professor of Computer Science at the University of Amsterdam. He serves on the editorial boards of Distributed Computing, Information Processing Letters, Theory of Computing Systems, Parallel Processing Letters, International journal of Foundations of Computer Science, Journal of Computer and Systems Sciences (guest editor), and elsewhere. He has worked on cellular automata, computational complexity, distributed and parallel computing, machine learning and prediction, physics of computation, Kolmogorov complexity, quantum computing. Together with Ming Li they pioneered applications of Kolmogorov complexity and co-authored "An Introduction to Kolmogorov Complexity and its Applications," Springer-Verlag, New York, 1993 (2nd Edition 1997), parts of which have been translated into Chinese, Russian and Japanese. Web page: <http://www.cwi.nl/~paulv/>

