Towards the Design of an Automated Physicist: A model (and program) for inferring the physical laws of a simplified world

Manuel Blum, Carnegie Mellon University

Abstract

Sloane's online Encyclopedia of Integer Sequences* contains the beginnings of over 200,000 integer sequences, each hand labeled, succinctly described, and carefully referenced. These include all algorithmically-generated, infinite, integer sequences that to my knowledge have ever been described in print in any field of mathematics, science, engineering, etc. A typical example is **A000796** = 3,1,4,1,5,9,2,6,5,3... = decimal expansion of pi. (The encyclopedia also contains 1/pi, pi^2, sqrt(pi), ln(pi), exp(pi), and over 1000 more sequences concerned principally with pi.) Sloane's Encyclopedia may be viewed as a collection of over 200,000 demons, each of which is on the lookout for the one sequence it knows.

The goal of the work described here has two parts:

1. define a large mathematically-natural efficiently-inferable class of sequences that includes a large fraction of Sloane sequences;
2. replace Sloane's enormous collection of sequences (viewed as demons) by a much smaller collection of more powerful demons that work together to efficiently infer all sequences in the class.

For example, Sloane's encyclopedia has roots of many integer polynomials (polynomials with integer coefficients), like **A002913** = root(x^2-2) = sqrt(2) = 1,4,1,4,2,1,3,5,6... and **A060006** = root(x^3-x-1) = 1,3,2,4,7,1,7,9,5..., but does not contain root(x^2-7x+1) = 1,4,5,8,9,8,0,3,3,7,5,0,3,1,5,4,5,.... Our automated physicist uses a lattice reduction demon to efficiently infer the coefficients of any integer polynomial from a modest number of digits (Big-Oh ((degree) x (length of largest coefficient))) of any (single) root of the polynomial.

I model the process of inferring the laws of the physical world (the inductive inference problem) as a game between a Deterministic Turing Machine (TM) whose input/output represents the physics of a deterministic world and an inference algorithm representing an automated physicist that seeks to infer the laws of that world. These TMs will be called the *world-TM* and the *physicist-TM* respectively. In general, the physicist-TM can experiment by feeding an input x to the world-TM, which computes away until at some point later in time, if ever, it outputs a response, y, and an upper bound, t, on its run time (the time it took to compute y from x). More generally, the physicist-TM feeds x1 and gets back <y1, t1>, then feeds x2 and gets back <y2, t2>, and so on. The goal of the physicist-TM is to converge (in *finite*(!) time) on a correct model of the world-TM, i.e. a model that on any input xi correctly computes and outputs yi <big>in (a predetermined polynomial of) the given time ti. (After converging on a correct model of the world-TM, the physicist-TM still does not know -- and in general cannot know -- that it has converged.)

Applications of such an inference algorithm vary widely from the automation of physics discovery to the construction of software that is sufficiently intelligent to learn from examples.

---

* The On-Line Encyclopedia of Integer Sequences, published electronically at http://oeis.org 2012