Compression Progress, Pseudorandomness, & Hyperbolic Discounting

Moshe Looks

Google, Inc.

1600 Amphitheatre Pkwy, Mountain View, CA 94043

madscience@google.com

Abstract

General intelligence requires open-ended exploratory learning. The principle of compression progress proposes that agents should derive intrinsic reward from maximizing "interestingness", the first derivative of compression progress over the agent's history. Schmidhuber posits that such a drive can explain "essential aspects of ... curiosity, creativity, art, science, music, [and] jokes", implying that such phenomena might be replicated in an artificial general intelligence programmed with such a drive. I pose two caveats: 1) as pointed out by Rayhawk, not everything that can be considered "interesting" according to this definition is interesting to humans; 2) because of (irrational) hyperbolic discounting of future rewards, humans have an additional preference for rewards that are structured to prevent premature satiation, often superseding intrinsic preferences for compression progress.

Consider an agent operating autonomously in a large and complex environment, absent frequent external reinforcement. Are there general principles the agent can use to understand its world and decide what to attend to? It has been observed going back to Leibniz that understanding is in many respects equivalent to compression.[1] To understand its world, a competent agent will thus attempt, perhaps implicitly, to compress its history through the present, consisting of its observations, actions, and external rewards (if any). Any regularities that we can find in our history through time t, $h(\leq t)$, may be encoded in a program p that generates the data $h(\leq t)$ as output by exploiting said regularities.

Schmidhuber has proposed the *principle of compression progress* (Sch09): long-lived autonomous agents that are computationally limited should be given intrinsic reward for increasing subjective "interestingness", defined as the first derivative of compression progress (compressing $h(\leq t)$). Agents that are motivated by compression progress will seek out and focus on regions of their environment where such progress is expected. They will avoid both regions of the world which are entirely predictable (already highly compressed), and entirely unpredictable (incompressible and not expected to yield to compression progress).

---

[1] Cf. (Bau04) for a modern formulation of this argument.

A startling application of the principle of compression progress is to explain "essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes", as attempted in (Sch09). The unifying theme in all of these activities, it is argued, is the active process of observing new data which provide for the discovery of novel patterns. These patterns explain the data as they unfold over time by allowing the observer to compress it more and more. This progress is explicit and formal in science and mathematics, while it may be implicit and even unconscious in art and music. To be clear, engaging in these activities often provides external rewards (fame and fortune) that are not addressed here; we consider only the intrinsic rewards from such pursuits.

Rayhawk (Ray09) criticizes this attempt with a gedankenexperiment. First, generate a (long) sequence of 2^n bits with a psuedorandom number generator (PRNG) using an unknown but accessible random seed, n bits long. Assuming that the PRNG is of high quality and our agent is computationally limited, such a sequence will require $\Theta(2^n)$ bits to store. Access the random seed, and use it to recode the original 2^n bits in $\Theta(n)$ space by storing just the seed and the constant-length PRNG code. This will lead to compression progress, which can be made as large as we would like by increasing n. Of course, such compression progress would be very uninteresting to most people!

The applicability of this procedure depends crucially on two factors: 1) how the complexity of compression programs is measured by the agent, namely the tradeoff between explanation size (in bits) and execution time (in elementary operation on bits); and 2) which sorts of compression programs may be found by the agent. Consider an agent that measures compression progress between times t and $t+1$ by $C(p(t), h(\leq t+1)) - C(p(t+1), h(\leq t+1))$ (see (Sch09) for details). Here $p(t)$ is the agent's compression program at time t, and $C(p(t), h(\leq t+1)$ is the cost to encode the agent's history *through time $t+1$*, with $p(t)$. If execution time is not accounted for in C (i.e. cost is simply the length of the compressor program), and p may be any primitive recursive program, the criticism disappears. This is because even without knowing the random seed, $O(n)$

bits are sufficient to encode the sequence, since we can program a brute-force test of all possible seeds without incurring any complexity costs, while storing only a short prefix of the overall sequence. Thus, the seed is superfluous and provides no compression gain. If execution time has logarithmic cost relative to program size, as in the speed prior (Sch02), then learning the seed will provide us with at most a compression gain logarithmic in n. This is because testing all random seeds against a prefix of the the sequence takes $O(n2^n)$ time, so $C(p(t), h(\leq t+1))$ will be about $n + log(n)$, while $C(p(t+1), h(\leq t+1))$ will be about n.

Thus, such pathological behavior will certainly not occur with a time-independent prior. Unfortunately, the compression progress principle is intended for precisely those computationally limited agents with time-dependent priors, that are too resource-constrained to brute-force random seeds. A reasonable alternative is to posit an *a priori* weighting over data that would assign zero utility to compression progress on such a sequence, and nonzero utility to compression of e.g. knowledge found in books, images of human faces, etc. This gives a principle of *weighted* compression progress that somewhat less elegant, but perhaps more practical.

A very different theory that also addresses the peculiar nature of intrinsic rewards in humans is hyperbolic discounting, based on long-standing results in operant conditioning (Her61). In standard utility theory, agents that discount future rewards against immediate rewards do so exponentially; an expected reward occurring t units of time in the future is assigned utility $r\gamma^t$ relative to its present utility of r, where γ is a constant between 0 and 1. The reason for the exponential form is that any other function leads to inconsistency of temporal preferences; what the agent prefers now will not be what it prefers in the future. However, considerable empirical evidence (Ain01) shows that humans and many animals discount future reward not exponentially, but hyperbolically, approximating $r(1+t)^{-1}$. Because of the hyperbolic curve's initial relative steepness, agents discounting according to this formula are in perpetual conflict with their future selves. Immediately available rewards can dominate decision-making to the detriment of cumulative reward, and agents are vulnerable to self-induced "premature satiation", a phenomenon that is nonexistent in exponential discounters (Ain01). While an exponential discounter may prefer a smaller sooner reward (when $\gamma < 1$), this preference will be entirely consistent over time; there will be no preference reversal as rewards become more imminent.

Hyperbolic discounting and the compression progress principle intersect when we consider activities that provide time-varying intrinsic rewards. *They conflict when rewards may be consumed at varying rates for varying amounts of total reward.* Consider an agent examining a complex painting or sculpture that is not instantaneously comprehensible, but must be understood sequentially through a series of attention-shifts to various parts. Schmidhuber (Sch09) asks: "Which sequences of actions and resulting shifts of attention should he execute to maximize his pleasure?" and answers "According to our principle he should select one that maximizes the quickly learnable compressibility that is new, relative to his current knowledge and his (usually limited) way of incorporating / learning / compressing new data." But a hyperbolically discounting agent is incapable of selecting such a sequence voluntarily! Due to temporal skewing of action selection, a suboptimal sequence that provides more immediate rewards will be chosen instead. I posit that the experiences humans find most aesthetically rewarding are those with intrinsic reward, generated by weighted compression progress, *that are structured to naturally prevent premature satiation.*

In conclusion, I posit two major qualifications of the applicability of the principle of compression progress to humans. First, that the value of compression progress is weighted by the *a priori* importance of the data that are being compressed. This is most obvious in our interest in faces, interpersonal relations, etc. Even more abstract endeavors such as music (Mit06) and mathematics (LN01) are grounded in embodied experience, and only thus are such data worth compressing to begin with. Second, that experiences that intrinsically limit the "rate of consumption" of compression progress will be preferred to those requiring self-regulated consumption, even when less total reward is achievable by a rational agent in the former case than in the latter. AGI designers should bear these caveats in mind when constructing intrinsic motivations for their agents.

References

G. Ainslie. *Breakdown of Will.* Cambridge University Press, 2001.

E. B. Baum. *What is Thought?* MIT Press, 2004.

R. Herrnstein. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1961.

G. Lakoff and R. Núñez. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being.* Basic Books, 2001.

S. J. Mithen. *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body.* Harvard University Press, 2006.

S. Rayhawk. Personal communication, 2009.

J. Schmidhuber. The speed prior: a new simplicity measure yielding near-optimal computable predictions. In *Conference on Computational Learning Theory*, 2002.

J. Schmidhuber. Driven by compression progress. In G. Pezzulo, M. V. Butz, O. Sigaud, and G. Baldassarre, editors, *Anticipatory Behavior in Adaptive Learning Systems.* Springer, 2009.