A Standard Model of Mind:
AAAI Technical Report FS-17-05

Toward a General Model of Human-Like General Intelligence

Ben Goertzel, Eddie Monroe
Hanson Robotics & OpenCog Foundation
7th Floor, Bioinformatics Bldg., Science Park
Sha Tin, N.T., Hong Kong

Abstract

It is argued that any real-world, limited-resources general intelligence is going to manifest a mixture of general principles such as Solomonoff induction and complex self-organizing adaptation, with specific structures and dynamics that reflect corresponding structures and dynamics in the tasks and environments in whose context it was created. This interplay between the general and the specific will play out differently in each type of intelligent system. A number of ideas drawn from previous publications are reviewed here – e.g. cognitive synergy, PGMC and the Mind-World Correspondence Principle – which formalize aspects of this perspective, and provide guidance on how to use it to analyze and create general intelligences.

Introduction

What is general intelligence? And what, specifically, characterizes human-like general intelligence? What are the necessary and/or sufficient structures and dynamics and properties of general intelligence in general, and/or of the sort of general intelligence that human beings display?

It is not clear how fully we need to understand these questions in order to build working AGI systems. Often, in history, people have built artifacts that worked perfectly well, long before the underlying principles were well understood. However, there seems little doubt that more fully understanding the principles of general intelligence would help with building AGI systems; and of course, understanding the nature of intelligence is a valuable intellectual pursuit in its own right.

During the last decade or so, the author has published a series of books and papers exploring the notion of a "general theory of general intelligence" in its various aspects. This brief paper comprises an extended abstract summarizing some of the key points made in this body of work, and giving the relevant references for the reader who wishes to dig deeper.

Many of the ideas presented here were developed in the course of work on the OpenCog AGI architecture (Goertzel, Pennachin, and Geisweiller 2013a) (Goertzel, Pennachin, and Geisweiller 2013b); however, as theoretical notions they are not restricted to OpenCog, and indeed among their implications is that there are many different workable approaches to AGI, all of which are likely to manifest some similar emergent structures and dynamics.

Cognitive Systems As Self-Organizing, Complex Adaptive Systems

"Artificial General Intelligence" is a valuable concept sociologically due to the emphasis it places on systems that can generalize and carry out a wide variety of tasks, rather than being narrowly constrained to individual tasks or contexts. However, as argued in(Goertzel. 2015), the three concepts of "artificial", "general" and "intelligence" are all somewhat problematic. As hybrid bio-cybernetic systems are explored, and self-modifying machines are constructed, the "artificial" nature of engineered intelligent systems will become fuzzier. Generality of intelligence is also quite fuzzy; for instance we humans are much smarter in 3D space than we would be in 17D space, and most of us are much cleverer at seducing members of the opposite sex than at factoring large numbers. And "intelligence" itself, with its focus on problem-solving and goal-achievement, arguably captures only part of what is interesting about clever systems like, say, humans, dolphins and crows.

An alternative is to think about humans as a particular sort of Self-Organizing, Complex Adaptive System (SCADS). This ties in with Weaver and Veitas' notion of "Open-Ended Intelligence" (Weinbaum and Veitas 2015) – in which intelligence is conceived as a capability for self-transcendence via coupling with the wider world. In this view, the crux of intelligence is not striving and succeeding to achieve known goals, but rather fundamentally self-modifying so that one comes to understand the world in whole new ways, including new goals and new vistas.

In (Goertzel 2017b) this perspective is explored mathematically; a hierarchy of increasingly specialized models is created, beginning with a quite generic model of a self-modifying system comprising elements that modify each other to produce new elements, and culminating with more specific self-modifying system models consisting of hypergraphs some of whose subhypergraphs encode rules for mapping subhypergraphs into subhypergraphs. These hypergraph-based models have close connections to probabilistic and category-theoretic models of intelligence, and also to hypergraph-based AGI systems such as OpenCog.

Cognition as Composed of Forward and Backward Growth Processes and Diffusion Processes

An extension of this line of thinking is to model cognitive processes as forward and backward growth processes, and diffusion processes (Goertzel 2006). A forward growth process is one in which elements of a system combine with each other to produce new elements – and so on recursively. A backward growth process is one in which an element x of a system elicits the creation of other elements y and z of the system so that the combination of y and z can produce x – and so on recursively. A diffusion process is one in which some quantity or dynamic spreads from system elements to neighboring elements, and so on recursively.

These sorts of processes exist throughout nature and in the human body and in the human mind. We can easily see them in common AI algorithms as well: in logic we have forward and backward chaining inference; in neural nets we have backpropagation (which is a backward growth process on connection weights) and activation spreading (which is a diffusion process), and in some neural nets we have Hebbian learning or neurogenesis or synaptogenesis as forward growth processes.

One can view these as processes that occur in complex, self-organizing systems and help them to adapt to their environments.

Modeling Cognition as Reinforcement and Optimization

There is a robust tradition modeling intelligence as optimization of utility functions defined over computable environments (Hutter 2005). The core ideas here are extensions of Solomonoff induction – of the notion that the best way to predict the future is to use the simplest accurate model of the past. This idea can be used to guide an intelligent system's actions, because action selection can be done by figuring out which action would give the most reward in the future, based on simple probabilistic extrapolation from the simplest accurate model of the past.

This approach leads to some elegant mathematics; for instance one can show that, given near-infinite computing resources, one can make a nearly-optimally-intelligent agent using a very brief computer program. The relation of this mathematics to general intelligence or complex system dynamics in the real world remains a subject of debate.

Shane Legg and Marcus Hutter define general intelligence as, roughly speaking, the average complexity of the utility functions that a system can optimize, averaged over computable environments (Legg and Hutter 2007). They measure complexity in terms of the Solomonoff universal prior, which says that shorter programs (in some assumed universal programming language) are less complex. In (Goertzel 2010) I modified this definition to look at more general complexity measures (not just the Solomonoff prior, which always exponentially decays in the large), and to look at both the complexity of the goal and the complexity of the environment in a more balanced way.

One weakness of this sort of approach, though, is that is always depends on who is defining the complexity measure. In the large, the specifics of the complexity measure don't matter; but intelligence in the real world is always dealing with the medium scale. One could argue that much of intelligence is actually about figuring out what is the right language to use to model reality, and what is the right way to measure complexity – and that then optimizing things relative to a given language and complexity measure is just a "mathematical crunching" problem rather than the cruz of intelligence.

The Specific Evolutionary and Ecological Pressures Shaping Human-Like Minds

Any system with finite resources is going to be better at doing some things than others. And if current physics is at all on target, any real system is going to have finite resources (due to quantum theory, special relativity, and so forth placing limitations on the amount of information processing that can be done with a given amount of space, time and energy). In this sense, a totally generally intelligent system is a fiction (though perhaps a useful mathematical fiction).

Human beings, as specific types of systems with very finite resources, are especially biased to be good at doing certain sorts of things. In (Goertzel 2009b) it is argued that human intelligence evolved largely out of the need for humans to interact socially with other humans in a shared perceptual and motoric context. It is argued that the prior distribution over "program space" implied by this sort of focus on embodied communication and interaction within a community, naturally assigns a high degree of simplicity to programs that deal with episodic memory, declarative memory, procedural memory and sensorimotor memory in a roughly humanlike way.

In (Goertzel, Ikle', and Wigmore 2012), I attempted to draw a sort of "unified cognitive architecture diagram", with boxes for the standard types of memory and standard cognitive processes as reviewed in cognitive psychology and cognitive science textbooks. I connected this with a variety of different AGI architectures. According to the "embodied communication prior" idea, this sort of breakdown into types of memory and processes constitutes a computational model that assigns the greatest simplicity to those processes that happen to be useful for survival as an agent with needs to communicate with other agents in a community with which a sensorimotor environment is shared.

Probability and Intelligence

Probability theory has played an increasingly prominent role in AI in recent decades; it also plays a significant role in neuroscience via concepts like probabilistic population coding. There is reason to believe this has fundamental explanations rather than just being a trend.

Logic, Probability, and Emergent Symmetries in Cognitive Systems

When one digs deep into the foundations of probability theory as e.g. Knuth and Skilling (Knuth and Skilling 2012) have done, one sees that probability theory is a reflection of certain basic mathematical symmetries. These symmetries are present in the physical universe in various ways, and they also arguably are present in the nature of many communication processes, even viewed on an abstract level. It seems likely that, in order to survive effectively in a universe that is simply described in terms of the symmetries from which probability theory derives, a system should manifest some of these same symmetries; and carrying out probabilistic reasoning is one way of doing this.

It also seems possible that standard Kolmogorovian probability theory is not the best model of probability theory as carried out in real-world intelligent systems. In (Goertzel 2017a) I have given an argument for intuitionistic probability theory instead, which seems to emerge more naturally from hypergraphs which are natural models of many aspects of intelligent systems. One is then led into fairly subtle issues of exactly what are the relevant symmetries of the perceived and enacted world as well as of the inferring mind.

Cognitive Processes as Probabilistic Growth and Mining of Combinations

The general notion of forward and backward growth processes as summarized above may also be modeled probabilistically. One can view a growth process as a series of choices: which among many pathways for growth to take next? Each of these choices may be considered as sampling from a probability distribution. This leads to a model of cognitive processes in terms of Probabilistic Growth and Mining of Combinations or PGMC ((Goertzel 2016b)). With a bit of effort, for instance, all the key cognitive processes in the OpenCog AI system can be expressed either as forward or backward PGMC processes or as diffusion processes (Goertzel 2016a).

Cognitive Synergies

A hallmark of complex, self-organizing systems is the phenomenon of ”emergence” – static or dynamic properties of a whole system that are very hard to deduce from properties of the parts. Cognitive systems display these in spades; and they are especially prone to display a particular kind of emergence I have labeled ”cognitive synergy” (Goertzel 2009a). This is a phenomenon in which, within a set of cognitive processes, whenever one of them tends to get plagued with combinatorial explosions (meaning, high entropy in the probability distribution among choices at its disposal), others are generally able to come in and solve the problem at hand without combinatorially exploding. This can be formulated mathematically in an abstract way (Goertzel 2017b); and prior papers have explored its manifestations in specific aspects of cognition, such as the synergy between procedural and declarative learning (Goertzel et al. 2011) , and the synergy between probabilistic logic and attentional diffusion (Harrigan et al. 2014).

It has also been argued that ”cognitive synergy is tricky,” in the sense that a highly synergetic set of 10 processes may not have any highly synergetic subset of, say, 5 processes (Goertzel and Wigmore 2011).

The Mind-World Correspondence Principle

Pulling together a number of these ideas, in (Goertzel 2011b) I have argued for a ”Mind-World Correspondence Principle,” according to which: For a complex system with limited resources to display general intelligence in a certain complex environment, the internal dynamics of the system must display many of the same high-level dynamical patterns as the environment does. This is modeled using category theory, but in many ways is a very simple concept. The prominence of hierarchy in the human mind-brain is one example; these hierarchical dynamics are useful to the human mind-brain because our region of the physical universe is marked by so many hierarchical structures and dynamics. The ”Embodied Communication Prior” mentioned above is a subtler example of the same principle.

Hypergraphs as a General Cognitive Modeling Framework

To get beyond highly general and abstract principles, one needs to make some specific formal assumptions about what kind of intelligent system one is going to explore. Linas Vepstas, who for many years has been the lead AI engineer on the OpenCog AI system, has argued that modeling intelligent systems in terms of hypergraphs is a generally valuable strategy (Brainwave 2013). Declarative, procedural and episodic knowledge can all be expressed elegantly in terms of hypergraphs, as can goals and many other aspects of human-like intelligence.

Whether hypergraphs form a valuable way to model all human-like intelligent systems is not yet clear, but they are certainly a useful way to model OpenCog whose Atomspace knowledge store is explicitly constructed as a weighted, labeled hypergraph. In OpenCog, cognitive synergy is manifested via multiple cognitive processes that are implemented as dynamics for hypergraph rewriting (and many of these dynamics are implemented themselves as hypergraphs within the same Atomspace).

It is especially unclear whether hypergraphs are a good way to model sensorimotor knowledge; however, one can handle this case via hybridizing hypergraphs with hierarchical continuous-valued functions such as deep neural nets. In (Goertzel 2017c) (Goertzel 2011a) and other recent work I have taken a neural-symbolic approach to perception processing, in which surprising patterns in the structure of deep neural nets are represented in a symbolic hypergraph, where they can be reasoned on and can then feed back into the neural nets providing symbolic guidance. This may reflect a more general mathematical approach in which hypergraphs are linked to hierarchical constructs built on spaces of continuous functions. From a Mind-World Correspondence Principle perspective, one might view this sort of mathematical hybridization as corresponding to an exploration of the intersections and interactions between different aspects of the tasks and environments with which human-like minds are confronted.

Conclusion

A comprehensive, detailed model of general intelligence, or human-like general intelligence, is not yet at hand; but nor are we completely clueless. It is clear that any real-world, limited-resources general intelligence is going to manifest a mixture of general principles such as Solomonoff induction and complex self-organizing adaptation, with specific structures and dynamics that reflect corresponding structures and dynamics in the tasks and environments in whose context it was created. Exactly how this interplay between the general and the specific plays out in the human mind, and exactly how it should or must play out in roughly but not exactly human-like AGI systems that we may create, is something we are in the midst of discovering. Ideas reviewed here such as cognitive synergy, PGMC and the Mind-World Correspondence Principle may perhaps be of aid in this discovery process.

References

Brainwave, O. 2013. Why hypergraphs? http://blog.opencog.org/2013/03/24/why-hypergraphs/.

Goertzel, B., and Wigmore, J. 2011. Cognitive synergy is tricky. *Chinese Journal of. Mind and Computation*.

Goertzel, B.; Pitt, J.; Wigmore, J.; Geisweiller, N.; Cai, Z.; Lian, R.; Huang, D.; and Yu, G. 2011. Cognitive synergy between procedural and declarative learning in the control of animated and robotic agents using the opencogprime agi architecture. In *Proceedings of AAAI-11*.

Goertzel, B.; Ikle', M.; and Wigmore, J. 2012. The architecture of human-like general intelligence. In *Foundations of Artificial General Intelligence*.

Goertzel, B.; Pennachin, C.; and Geisweiller, N. 2013a. *Engineering General Intelligence, Part 1: A Path to Advanced AGI via Embodied Learning and Cognitive Synergy*. Springer: Atlantis Thinking Machines.

Goertzel, B.; Pennachin, C.; and Geisweiller, N. 2013b. *Engineering General Intelligence, Part 2: The CogPrime Architecture for Integrative, Embodied AGI*. Springer: Atlantis Thinking Machines.

Goertzel, B. 2006. A system-theoretic analysis of focused cognition, and its implications for the emergence of self and attention. *Dynamical Psychology*.

Goertzel, B. 2009a. Cognitive synergy: A universal principle of feasible general intelligence? In *Proceedings of ICCI 2009, Hong Kong*.

Goertzel, B. 2009b. The embodied communication prior. In *Proceedings of ICCI-09, Hong Kong*.

Goertzel, B. 2010. Toward a formal definition of real-world general intelligence. In *Proceedings of AGI-10*.

Goertzel, B. 2011a. Integrating a compositional spatiotemporal deep learning network with symbolic representation/reasoning within an integrative cognitive architecture via an intermediary semantic network. In *Proceedings of AAAI Symposium on Cognitive Systems,*.

Goertzel, B. 2011b. A mind-world correspondence principle: Toward a general theory of general intelligence. *Dynamical Psychology*.

Goertzel., B. 2015. *The AGI Revolution*. Humanity+ Press.

Goertzel, B. 2016a. Opencoggy probabilistic programming. http://wiki.ogoertzel2016probabilisticpencog.org/w/OpenCoggy_Probabilistic_Programming.

Goertzel, B. 2016b. Probabilistic growth and mining of combinations: A unifying meta-algorithm for practical general intelligence. In *International Conference on Artificial General Intelligence*, 344–353. Springer.

Goertzel, B. 2017a. Cost-based intuitionist probabilities on spaces of graphs, hypergraphs and theorems.

Goertzel, B. 2017b. From abstract agents models to real-world agi architectures: Bridging the gap. In *Proceedings of AGI-17, submitted for publication*. Springer.

Goertzel, B. 2017c. The role of deep learning in human-level agi. *submitted for publcation*.

Harrigan, C.; Goertzel, B.; Ikle', M.; Belayneh, A.; and Yu, G. 2014. Guiding probabilistic logical inference with nonlinear dynamical attention allocation. In *Proceedings of AGI-14*.

Hutter, M. 2005. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer.

Knuth, K. H., and Skilling, J. 2012. Foundations of inference. *Axioms* 1(1):38–73.

Legg, S., and Hutter, M. 2007. A collection of definitions of intelligence. In *Advances in Artificial General Intelligence*. IOS.

Weinbaum, D. W., and Veitas, V. 2015. Open-ended intelligence. http://arxiv.org/abs/1505.06366.