Working Toward Pragmatic Convergence:
AGI Axioms and a Unified Roadmap

Itamar Arel

Machine Intelligence Lab
Department of Electrical Engineering and Computer Science
The University of Tennessee
Knoxville, TN 37996, USA
Email: itamar@ieee.org

Abstract

A primary goal of the AGI community should be to promote and encourage researchers to work toward coherent and broadly-acknowledged objectives. Perhaps one of the most important achievements that can be aspired to at this point in time is a framework agreement pertaining to what AGI systems should possess. Such an agreement can serve as basis for defining a roadmap of key challenges that would help the community at large progress toward the ultimate goal of achieving human-level intelligence. This paper proposes several fundamental *AGI axioms*, which are core functional attributes argued mandatory for a system to be considered as having AGI.

Introduction

Despite early efforts to collaborate and organize as a community, AGI researchers continue to diverge on disparate trajectories with no clear, common and well-defined end goal This is particularly true due to a lack of understanding or consensus on what AGI really is. Although promoting a wealth of ideas is a positive aspect of any healthy research discipline, particularly a relatively new one, it is desirable to hold a unified view of what the overarching goals of the community may be. This paper attempts to address this issue by offering an initial framework to facilitate consensus of short-term research focus in the AGI community.

The author argues that the community must strive to identify a set of *AGI axioms,* which are defined as core functional and behavioral attributes without which a system would not be considered to have AGI. The advantages of doing so are threefold. First, such axioms would clarify the otherwise nebulous words used in many AGI research papers, thus making claims, questions, and results concrete and testable Second, it will guide the community as a whole toward a unified vision, thereby promoting progress altogether. Third, it will facilitate the formation of a benchmarks roadmap that will improve credibility and indirectly enhance funding availability. This paper argues that the notion of a roadmap of benchmark tasks is vital for guaranteeing scientific progress and is further addressed later in this paper.

AGI Axioms

One can conceive of numerous attributes of a "true" AGI system, without which human-level intelligence would not be attainable. The following are five of what the author considers to be fundamental prerequisites for an AGI system, which may be viewed as elemental AGI axioms.

Axiom #1: Observability

The first AGI axiom may appear obvious: an AGI system must have the ability to continuously collect observations pertaining to the environment with which it interacts. Such observability is required at a rate which would suffice in the sense that receiving observations at any higher rate would not yield a significant increase in the informational content obtained. For example, if an AGI system is to interact with a physical environment in a manner similar to that with which humans do, it is expected that somewhere around 200 visual observations per second should suffice. That is not to claim that a higher rate of visual information capturing would not help the system perform tasks that humans are incapable of, but if it is human capability that we seek to match (at least initially), then human rate of information gathering is arguably adequate. To that end, these observation acquisition rates will strongly depend on applications targeted and nature of the particular sensory data considered (e.g., visual, auditory, etc.).

Needless to say, human-level intelligence implies partial observability in the sense that information delivered by the sensors provide only partial representation of the true "state" of the environment. To that end, as will be elaborated on later, an AGI system must be able to construct a model which complements the observations to form a consistent and coherent internal representation of the environment.

Axiom #2: Actuation Capability

Another AGI axiom is the ability to affect the environment in some desired manner. This implies a control mechanism that an AGI would be in command of. Without the ability to impact the environment, no system can be truly called AGI. Such passive intelligence does not appear to support a framework which can lead to an incremental intelligence capacity. To that end, an AGI system must be able to control some physical actuators applied to its environment.

Axiom #3: Ability to Process High-Dimensional Signals

Mammal brains are concurrently exposed to enormous amounts of information at any given moment, originating from their millions of sensors. This implies the existence of an efficient mechanism for processing super high-dimensional signals at very fast speeds. Limiting AGI systems to a small number of inputs would thus be unreasonable. Stated differently, an AGI architecture would not be considered such if it does not scale with respect to the ability of processing high-dimensional inputs (e.g. large images).

Axiom #4: Capturing Spatio-Temporal Dependencies

In addition to the ability of an AGI system to handle high-dimensional sensory inputs, it should also be able to capture various dependencies that may exist within the data. This is particularly true when viewed in the context of temporal information processing. We as humans have the inherent ability to represent spatio-temporal dependencies that span a large scale. We learn to anticipate certain events as a consequence of other events. In some cases the latter would have taken place seconds ago, while in other cases trigger events occurred years before. Thus, it is argued that an AGI system must be able to capture and internally represent dependencies pertaining to its sensory data in both space and time. The precise mechanisms for achieving such capabilities are at the core of numerous research studies at present time.

Axiom #5: Utility Function

Mammals have a functional goal at any given time. Such goals can be vague and abstract at times, yet they always exist. Working toward a goal is what drives the action-selection process, targeted at shaping the environment in some desired manner. Recent neuroscientific studies continue to support the notion that action selection in the brain is driven by a utility function. In other words, actions are selected so as to maximize some predefined notion of long-term reward.

The particular nature and definition of the utility function (or *value function*, as it is referred to in the reinforcement learning literature) are far from being understood. Humans inherently solve what is commonly referred to as the *credit assignment problem*. Stated simply, the credit assignment problem refers to the agent's ability to correctly assess both the short and longer term impact an action it takes would have on its environment. Many behavioral characteristics of mammals are related to this characteristic. Strategic thinking, in the context of predicting future outcomes resulting from current actions, is argued to be at the core of intelligence. It is something that humans do rather easily and current "AI" programs only aspire to.

Working toward a Roadmap

The above AGI axioms are most-likely a (small) subset of those truly required. However, it is important to initiate a discussion on them and eventually make progress in identifying those that are missing. The growing community of researchers interested in constructing general intelligence must create a roadmap. Such a map should reflect basic notions including what behavior one would expect from an AGI system. It would focus the energies of those working toward a common goal, while clarifying what precisely are the targets. Moreover, a common roadmap encourages active external verification of research procedures and results in the literature by posing benchmark problems or concrete progressive steps. A recent example of a similar effort is the Virtual Worlds Roadmap (http://www.virtualworldsroadmap.org/), which attempts to accelerate the progress of massively multiplayer online worlds.

Of course, calling for a roadmap is much easier than actually proposing one, and there will surely be much initial disagreement. The first step in creating a roadmap and defining progressively difficult benchmark tasks is broad agreement on AGI axioms.

A key aspect of a useful roadmap will be clearly defined steps toward the ultimate goal of human-level intelligence. Well recognized and respected problems should be introduced as parts of milestones that are progressively difficult. Solving each one of these milestone challenges represents a substantial achievement in itself and moves the state of the art forward toward strong artificial general intelligence. However, an important pitfall to avoid is the introduction of problems that can be solved by known "narrow" AI techniques. Such benchmarks would fail to distinguish general (or strong) AI systems from existing narrow ones. As with the RoboCup competition, posing several significant and complex problems could stimulate and unify much research aimed at constructing broadly competent and generally intelligent systems. Informally, it creates a common stage for sharing and proving the value of one's work.

The 10-IQ Fallacy

Some of the difficulty associated with defining and proving the existence of true AGI technology lies in what can be called the *10 IQ fallacy*. The idea suggests that even if scalable AGI technology were to exist today, proving it by solving small-scale problems would be extremely difficult. There are countless problems, such as face or speech recognition, that would otherwise be good to demonstrate an AGI system; however, they often can be solved with narrow AI techniques. To that end, proving that a system is truly generally intelligent on problems which would require the equivalent of, say 10 IQ, is challenging as there may be narrow AI solutions that can solve the problem at a similar level of performance. It may very well be the case that there is an intelligence threshold, only beyond which one could truly prove the existence of AGI.

Conclusions

We have passed the year 2001 without the formidable Hal, but not for lack of human potential. By designing a roadmap to artificial general intelligence and creating important benchmark problems that define steps on the path to this "grand dream," the original goal of AI may yet be achieved. But it will not happen if this nascent community fails to act and divides into factions with no shared path or vision. Will the common goal of AGI be achieved? The answer depends at least in part if not fully upon whether AGI axioms and a quality roadmap are formed.

Acknowledgement

The author would like to thank Scott Livingston for his valued insight and endless hours of discussions on this topic.