

Societies of Intelligent Agents

Bill Hibbard

SSEC, University of Wisconsin-Madison,
1225 W. Dayton St., Madison, WI 53706, USA
test@ssec.wisc.edu

Abstract. A definition of language is proposed in which language is a low-bandwidth channel that can increase agent rewards in a reinforcement learning setting, and in which agents can learn to produce language and teach it to other agents. Societies of agents are being modeled by economists to understand economic instability and other non-equilibrium phenomena. I hypothesize a divergent distribution of intelligence in societies of agents when rewards can be exchanged for increases in agent information processing capacity.

Introduction

This paper poses questions about societies of intelligent agents using a model based on reinforcement learning (RL) [1, 2]. The paper proposes a definition of language using the RL model. The model is adapted to societies of agents and used to pose questions about the evolution of society as intelligence increases beyond the plateau defined by humans.

In a simple reinforcement learning model, an agent interacts with its environment at a sequence of discrete times, sending action a_i to the environment and receiving observation o_i and reward r_i from the environment at time i . These are members of finite sets A , O , and R respectively, where R is a set of rational numbers between 0 and 1.

Language

Agents use language to communicate information to other agents. Language consists of strings of symbols, where the symbols generally represent objects recognized in observations of the environment, as well as properties and actions of objects also recognized in observations. In order to maximize rewards from the environment, agents try to predict the rewards they will get in response to various choices of actions they may make. Language is a valuable tool to help agents learn to predict the behavior of the environment from other agents (e.g., "If you poke a wasp nest, expect a large negative reward"). An important property of language is that a small amount of information in a language string (i.e., the number of bits required to encode the language string) can represent a much larger amount of information in observations of the environment, to agents who can recognize objects, properties and actions in those observations and know how they map to language symbols.

To describe this in the RL model, define a finite set L of language sentences. These sentences are strings over a finite set S of symbols, with length bounded by some constant (human agents have no need for sentences of more than one million symbols) and including the null string of length zero. In the definition of agent, add a new kind of observation from L (hearing a language sentence) and a new kind of action to L (speaking a language sentence). An agent π has mappings:

$$R \times O \times L \rightarrow \pi \rightarrow A \times L \quad (1)$$

where π includes some internal state that is updated at each time step.

Because language is specialized to particular environments, rather than trying to apply Legg's and Hutter's measure of agent intelligence [2], we will use a simple measure of agent success as the expected value of the sum of rewards during a time interval $(t1, t2)$:

$$V(\pi; t1, t2) = \mathbf{E}(\sum_{t=t1}^{t2} r_t) \quad (2)$$

where r_t is the reward agent π receives from the environment at time t .

The simplest social setting for language consists of two agents $\pi1$ and $\pi2$, with the language input of each connected to the language output of the other, and both receiving the same observations from the environment. We set up two scenarios for agent $\pi2$ over a time interval $(t1, t2)$, labeled $\pi1$ and *null*. In the $\pi1$ scenario the language inputs and outputs of $\pi2$ are connected to $\pi1$ and in the *null* scenario the language input of $\pi2$ receives only the null string at each time step. We require that language be a compressed description of environment observations, so if s_{t1}, \dots, s_{t2} are the sentences sent from $\pi1$ to $\pi2$ over the time interval then (many of them may be null):

$$(\sum_{t=t1}^{t2} \text{length}(s_t)) \log(|S|) \ll (t2 - t1 + 1) \log(|O|) \quad (3)$$

Let $V_{\pi1}(\pi2; t1, t2)$ and $V_{null}(\pi2; t1, t2)$ be the successes of agent $\pi2$ over the time interval $(t1, t2)$ in these two scenarios.

Definition 1. In the situation as described in the previous paragraphs, the effectiveness of $\pi1$ as a language teacher is $E(\pi1; \pi2, t1, t2) = V_{\pi1}(\pi2; t1, t2) / V_{null}(\pi2; t1, t2)$. $\pi1$ is an *effective language teacher* if $E(\pi1; \pi2, t1, t2) \geq C$, where $C > 1.0$ is a constant of the definition.

Language ability can be passed from agent to agent. To describe this, at time step $t2$ let $\pi2$ switch its language connections from $\pi1$ to a different agent $\pi3$ and continue to time step $t3$.

Definition 2. In the situation as described in the previous paragraphs, $\pi2$ has *learned language* if $E(\pi2; \pi3, t2, t3) \geq C$, where C is the constant from Definition 1.

The two-way language channels between agents provide a means for the learner to practice and get feedback from the teacher. This feedback is outside an agent's normal reward channel, but since learning language will increase the learner's rewards the learner should value feedback from the teacher. As we discuss in the next section, if agents are able to exchange reward in an economy, the learner may even pay reward to the teacher.

The notion that a language learner can become a language teacher disallows the possibility that language is simply a magic oracle predicting future rewards. Rather, language transmits knowledge that agents may use to predict rewards on their own.

Societies of Intelligent Agents

Now consider a set of agents $\{\pi_i, i \in I\}$. As an agent has only a single language input it must have a way to select which other agent to listen to. So add another action to the agent model of the previous section, a value in I to select which agent's language output is connected to this agent's language input. Now we can define the success of the society of agents over time interval $(t1, t2)$ as:

$$V(t1, t2) = \sum_{i \in I} V(\pi_i, t1, t2) \quad (4)$$

where $V(\pi_i, t1, t2)$ is defined in equation (2).

Consider two scenarios, labeled *lang* and *null*. In the *lang* scenario, the agents communicate via language and in the *null* scenario the language inputs to all agents are forced to null strings at every time step. Let $V_{lang}(t1, t2)$ and $V_{null}(t1, t2)$ be the successes of the society of agents over the time interval $(t1, t2)$ in these two scenarios.

Definition 3. In the situation as described in the previous paragraphs, the effectiveness of language for the society of agents is $E(t1, t2) = V_{lang}(t1, t2) / V_{null}(t1, t2)$. Language is *effective* for this society if $E(t1, t2) \geq C$, where $C > 1.0$ is a constant of the definition.

Mathematical models of societies of agents are an important new trend among economists, helping them to overcome limitations of the dynamic stochastic general equilibrium models they have long employed [3]. Agent-based economic models can help understand market instabilities and other non-equilibrium phenomena [4]. These models sometimes include agents that learn by reinforcement from the results of their behavior [5]. This suggests equating money with reward to create a society in which agents exchange reward for goods and services. Baum created such a society of agents to solve the credit assignment problem in reinforcement learning [6]. His society of agents learned to solve a block stacking problem and Rubik's cube, reinforced for successful solutions. Agents exchanged reward for computation, where the combined computations of many agents solved the problems.

Humans form a society of intelligent agents, but they are not collaborating on a simple common problem such as Rubik's cube. Each human agent has their own internal reward system, although these have enough in common that humans can form orderly markets for goods and services. Money is a first approximation to reward, although examples demonstrate the futility of absolutely equating money with reward.

The previous section offers a definition of language as information that can help agents obtain reward. If an agent can learn to produce language then it can provide it as a service to another agent, in exchange for reward. Thus our language definitions provide a rudimentary model of social knowledge. As social knowledge increases and is incorporated into language, agent intelligence increases.

Humans do increase their intelligence by working in social groups. They also manipulate their environment to create tools, some of which are used with language. Writing, books, newspapers, telegraph, telephone and audio recordings are tools for storing and transmitting language. Computers and networks are tools capable of more sophisticated language processing, epitomized by our host for this conference.

Although humans can augment their information processing capacity via tools and by working in social groups, their most valuable information processing capacity is in their physical brains. And all human brains have roughly equal information processing capacity and intelligence. Biotechnology, nanotechnology, and information technology are likely to enable humans to transcend the physical limits on their brains during the Twenty First Century. The application of these technologies to increase human intelligence will be a service of great value to individual humans, enabling them to obtain more reward. Thus humans will be willing to exchange reward for this service.

This suggests an economic model of agents that learn by reinforcement, in which agents exchange reward for information and also for increased information processing capacity. Information processing capacity and intelligence would be an attribute of each agent in such a model. Buchanan discusses the utility of agent-based models for understanding economic instability [4]. I hypothesize that a free market economy in which agents can exchange reward for increased information processing capacity will be unstable, with a large divergence in the intelligence of individual agents. There is some evidence for this in my work with adversarial sequence prediction, in which agents are rewarded with increased or decreased information processing capacity [7]. The ability to exchange wealth for increased intelligence, and to use that intelligence to increase wealth, will create a positive feedback loop amplifying differences among the intelligence of different humans. This could create differences in language processing capacities such that the most intelligent humans will speak languages that less intelligent humans can never learn.

A society of agents of such unequal intelligence will be fundamentally different from the society of agents of roughly equal intelligence that we are used to. Humans of different intelligence levels may have different legal rights, and humanity may essentially divide into multiple species. This is an issue that we as a society should consider carefully before we actually create these technologies. My prescription is that humans or machines with greater-than-natural-human-intelligence should require a license, granted under the condition that their values satisfy certain altruistic standards [8]. In any case, people should be informed about AI and transhumanist technologies and given a chance to debate and democratically decide whether and how they should be regulated.

References

1. Hutter, M.: *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Springer, Berlin (2004)
2. Legg, S., Hutter, M.: A Formal Measure of Machine Intelligence. In: 15th Annual Machine Learning Conference of Belgium and The Netherlands (Benelearn 2006), pp. 73-80. Ghent (2006) <http://www.idsia.ch/idsiareport/IDSIA-10-06.pdf>

3. Farmer, J. D. and Foley, D. The Economy Needs Agent-based Modeling. *Nature*. **460**, 685-686. 2009.
4. Buchanan, M. Meltdown Modeling. *Nature*. **460**, 680-682. 2009.
5. Nanduri, V. and Das, T. K. A Reinforcement Learning Model to Assess the Market Power Under Auction-Based Energy Bidding. *IEEE Trans. On Power Systems*. **22**, 85-95. 2007.
6. Baum, E. *What is Thought?* MIT Press, Cambridge, Mass. 2004.
7. Hibbard, B.: Adversarial Sequence Prediction. In: *The First Conference on Artificial General Intelligence (AGI-08)*. pp. 399-403. IOS Press, Amsterdam (2008)
http://www.ssec.wisc.edu/~billh/g/hibbard_agi.pdf
8. Hibbard, B.: The technology of mind and a new social contract. *Journal of Evolution and Technology* **17**(1), 13-22. 2008.