BMT22
4th International Conference
Business Meets Technology
Ansbach, 7th – 9th July 2022

ON THE INTELLIGENCE OF INTERACTING AUTONOMOUS ROBOTS AND VIRTUAL AGENTS

Geißelsöder, Stefan [a] **and Narovlyanskyy, Andriy**

[a] *University of Applied Sciences Ansbach, Germany (stefan.geisselsoeder@hs-ansbach.de)*

Abstract: This work explains some aspects why it is hard to pinpoint what intelligence is and more specifically, how to assess the intelligence of AI. It motivates a setup that is designed to foster the investigation of this question using reinforcement learning agents as complex AI systems. Such a setup can be used in an attempt to sidestep theoretical considerations on the cognitive power of Machine Learning algorithms. Instead, an example is given how the well-established experimental testing of intelligence in animals could be translated to the described AI system. While the published work-in-progress state of the implementation allows similar experiments of multiple interacting virtual robots to be conducted and a theoretical outline for future tests is sketched, a lot of further research will be required before a robot can demonstrably recognize itself in a mirror.

Keywords: *Artificial Intelligence; Reinforcement Learning; Intelligence Test*

1. INTRODUCTION

In recent years the field of artificial intelligence has seen a resurge in popularity initiated by the vast success of deep neural networks. However, the definition of AI has become very broad, often encompassing algorithms from mathematical optimization, database lookups or even the control flow of programs as hardcoded rule systems. While all of these aspects are important factors for AI systems, the naive understanding of the term “intelligence” implies different expectations. These different capabilities are frequently referred to as weak and strong, a.k.a. general AI (Russel & Norvig, 2003). Most typical current AI applications can be categorized as weak AI, focused on solving a particular problem. To investigate how far the abilities of current deep neural network architectures

How to cite: Geißelsöder, Stefan and Narovlyanskyy, Andriy. 2022. On the intelligence of interacting autonomous robots and virtual agents. In Proc.: 4th International Conference Business Meets Technology. Ansbach, 7th – 9th July 2022. Doi: https://doi.org/10.4995/BMT2022.2022.15555

2022, Editorial Universitat Politècnica de València

can reach, we need to define what we are looking for and consider a problem complex enough for any difference to be distinguishable.

2. INTELLIGENCE

The term intelligence itself eludes a rigorous definition. Common among many definitions is the ability to memorize and learn from information and to adapt to new circumstances (Webster, 2022). The issue doesn't resolve itself when it comes to artificial intelligence. Over the course of history, there have been many attempts to make intelligent algorithms. An example that was convincing for its time would be the chatbot ELIZA (Weizenbaum, 1966). While it did not incorporate any learning component, it was able to mimic intelligent behavior well enough to fool several humans that interacted with it. The famous Turing Test (Turing, 1950) builds upon the idea of an algorithm conversing with humans in such a way, that humans cannot reliably distinguish it from other humans. While the details required by different versions of the test vary significantly, a common criticism is, that it relies on the ability to fool human, not necessarily true intelligence. For example, a hypothetical machine that memorizes all human responses to all statements ever uttered would likely pass the test without any intelligence besides a giant database lookup.

One of the common requirements besides a working memory is the ability to generalize knowledge or to form new ideas based on memorized ones. While this has been achieved in academic settings by rule-based systems decades ago (Dai et al., 1990), these attempts did not continue to see wide success in most real-word applications. Deep neural networks do exhibit some implicit form of generalization when they create concepts that generalize their training data (Madan et al., 2021), but it is unclear if this can be considered an abstraction or any form of comprehension of the data. Due to the highly complex but essentially deterministic nature of the training process, it can be argued that all forms of machine learning are merely a mathematical optimization. Considering published literature it remains unclear if these processes can at some point in time yield aspects of strong AI such as high-level intelligence or even self-aware systems through emergence. To this day any such claim the author is aware of, e.g. (Sutskever, 2022), remains without corroborating evidence.

3. INVESTIGATION USING REINFORCEMENT LEARNING

To be able to distinguish "more intelligent" capabilities, the task needs to be able to prefer complex but beneficial behaviors over simple solutions. It is also helpful if the optimal solution is highly dynamic and can be changed over time, since this benefits the ability to adapt already learned knowledge.

A training scenario that fits these requirements is reinforcement learning for robots or virtual agents. In a nutshell, in a given situations a so-called agent is tasked to decide upon its next action(s) based on its current state and perception (Kaelbling et al., 1996). Virtual agents are more convenient for this type of study than actual robotic agents, as a virtual environment in which agents learn to act can be designed to focus on important aspects and it can be simulated faster than real-time data recording. The logic used for this decision is called a policy. More favorable policies do not just exploit the current sensorial input but rely on an internal memory of key aspects of their environment. These allow for more complex strategies to arise during training. Especially when long-term causal connection has to be exploited, these more strategic policies outperform short sighted systems. A memory of previous states can be explicitly given to an agent, but deep neural networks, especially with recurrent architectures, can also learn to remember relevant aspects of previous states. To be able to estimate the expected outcome of an action, an agent must not only be able to act upon key aspects of its current state (and potentially previous states), but it also needs to predict the effect of any known action on the environment. This can be achieved using a world model. Sometimes this is explicitly implemented as a simulation of the environment, but again, deep neural networks have been shown to learn predictions for actions in arbitrary environments (Dashkina et al., 2021).

There is a selection of published frameworks for simulations coupled with reinforcement learning (Brockman et al., 2016) (Beattie et al., 2016) (Körber et al., 2021), but besides the benefit of a simulated environment tailored to answer a specific question, a simplified simulation makes it easier for students to participate with own experiments. The current state of this framework can be found at (Geißelsöder, 2021). It simulates a basic world where agents need to rely on certain actions to survive. Coupled with the ability to reproduce, this setup creates a selection pressure favoring fit agents which maximize the chance of survival for their entire species. The setup already contains several different simple agents that follow fixed policies which can serve as interaction partners or additional selection pressures. To allow different roles as well as a comparison of different policies for the same role, the actions an agent is capable of (e.g. templates for wolf, sheep, ape, ...) are split from the implementation of their policies (brains). With these basics set up, a new policy may be implemented as the “brain” in class Ape. For educational purposes it is foreseen that many students can implement their own policies and compete against each other.

3.1. Survival

Due to an intrinsic pressure for survival of agents caused by slowly diminishing health over time, any policy has to identify and favor actions that are beneficial for the health of the individual agent. The goal of reaching a high health level doesn't have to be set explicitly for the training of the RL policy (these functions are called reward functions), but can be intrinsic to the environment, as policies that do not fullfil this criterion do not manage to persist in the environment. However, setting this as an explicit reward function simplifies training.

3.2. Cooperation

By default, an environment contains multiple individuals of a given policy. The simulation is designed such, that the policy is identifiable in the sensorial perception of every agent, theoretically allow them to recognize each other. There are also dedicated actions available for every agent to try communication with any object in its vicinity and arbitrary as well as fixed alphabets. This is designed with the intention of fostering cooperation between individuals that share the same (or compatible) policies to further enhance the survival probabilities of policies that incorporate actions benefitting its entire population. An example for such an action is described in the next section.

4. ADAPTING FROM BIOLOGY AND PSYCHOLOGY

Policies that are entirely created based on learning in an environment and that are intelligent enough to keep agents alive and allow its population to thrive through cooperation must be recognized as some kind of intelligence. However, the counterargument that the entire process can be seen as a complex mathematical optimization of an implicit reward function cannot be invalidated.

With this perspective in mind, we might want to turn our attention to domains that have more experience with the struggle to find empirical tests for the intelligence of beings. While the gap to and specialization of human level IQ tests seems a bit challenging, many years of testing the intelligence of animals have proven to yield well established and reproducible results. Besides communication (Seyfarth & Cheney, 2003), counting, abstract thinking etc., the famous tests of animals recognizing themselves in mirrors comes to mind (Gallup, 1970).

To test this scenario, we need to add mirrors to the simulation that alter the perception of agents accordingly. Similar to the markings applied to the foreheads of animals in the mirror tests, we also need to add a property to each individual that encodes a marking, which must be perceivable for the agents. And lastly, the agents need to be able to perform the action of removing the mark. If we were to now set an explicit reward function that directly rewards the removal of a mark, the setup would not be comparable to the same test with animals. The way chosen for this implementation considers some important details that are meant to motivate the behavior but also to discern altruism from self-recognition. To motivate the behavior, the mark has been assigned with a long delayed, but severe drop in health of the agent if it is not removed by any agent. This is to motivate the behavior as an altruistic act. However, it is also penalized by a prolonged time of forced inaction for the removing individual. This is meant to inhibit frequent but casual and inconsiderate removal actions to minimize the chance of the mark removal action being triggered without clear motivation. Furthermore, the action of removal requires the concept of a mirror to be at least partially understood by the agents, as it requires a target to be specified relative to the location of the acting agent. If the agent tries to remove the mark from another individual at twice the distance to the mirror instead of itself, the concept is not yet understood (and of course the mark is not removed).

The status of the ongoing research revolves around options how to discern an altruistic action intended to help another agent that happens to be perceived at the same location as the acting agent from an action that is deliberately performed on itself. Introducing yet another action to clear the mark from oneself currently is the preferred option, but this introduces the additional complexity that the connection between the two actions (removing a mark on another agent versus removing the mark for the acting agent itself) needs to be learned on top of all previously required learning. However, once successfully designed and implemented, this approach could allow the assessment of the intelligence of AI agents with similar justification as intelligence research in animals established over decades.

5. SUMMARY AND OUTLOOK

This work mentions some aspects why it is hard to pinpoint how intelligent AI actually is. It briefly describes a setup that is designed to allow the investigation of some of the tangible aspects and it motivates what could be suited aspects. In its current state, the environment is suited to perform basic RL experiments. In the future this setup is intended to be used for the more detailed investigation of the intelligence of AI as it tries to recreate situations similar to tests of intelligence in animals. The next big step is to perform training of agents that can go beyond pure survival. Shooting for the moon would be to find ways to train AI agents that can demonstrably recognize themselves in a mirror.

REFERENCES

Russel & Norvig (2003). Artificial Intelligence: A Modern Approach, Pearson, ISBN 0130803022

Webster (2022). https://www.merriam-webster.com/dictionary/intelligence, last accessed May 2 2022

Joseph Weizenbaum (1966). ELIZA – A Computer Program For the Study of Natural Language Communication Between Man And Machine. in: Communications of the ACM., ISSN 0001-0782

Alan M. Turing (1950). Computing Machinery and Intelligence, in Mind. Band LIX, Nr. 236, ISSN 0026-4423, p. 433–460

Dai et al. (1990). Issues of Real-Time Expert Systems, in AI and Cognitive Science'89, pp179-197, Springer

Madan et al. (2021). When and how CNNs generalize to out-of-distribution category-viewpoint combinations, Nature Machine Intelligence volume 4, pages 146–153 (2022), https://arxiv.org/abs/2007.08032

Ilya Sutskever (2022). Twitter, February 9 2022,

https://twitter.com/ilyasut/status/1491554478243258368

L. P. Kaelbling, M. L. Littman and A. W. Moore (1996), Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research, Vol 4, (1996), 237-285, https://arxiv.org/abs/cs/9605103

Alexandra Dashkina et al. (2021). Neural Network Modeling as a Method for Creating Digital Twins: From Industry 4.0 to Industry 4.1, SPBPU IDE '20: Proceedings of the 2nd International Scientific Conference on Innovations in Digital Economy

Marian Körber, Johann Lange, Stephan Rediske, Simon Steinmann, Roland Glück (2021). Comparing Popular Simulation Environments in the Scope of Robotics and Reinforcement Learning, https://arxiv.org/abs/2103.04616

G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang and W. Zaremba (2016). OpenAI Gym, https://arxiv.org/abs/1606.01540 and https://gym.openai.com, last accessed May 2 2022

C. Beattie et al. (2016). DeepMind Lab, https://arxiv.org/abs/1612.03801,

https://www.deepmind.com/blog/open-sourcing-deepmind-lab

Stefan Geißelsöder (2021). https://github.com/sgeisselsoeder/erkenntnis

R. Seyfarth & D. Cheney (2003). Signalers and Receivers in Animal Communication,

Annual Review of Psychology. 54 (1): 145–173

GG Jr. Gallup (1970). Chimpanzees: Self recognition, in Science. 167 (3914): 86–87

