Leibo Hughes Lanctot Graepel – Autocurricula and the Emergence of Innovation from Social Interaction: A Manifesto for Multi-Agent Intelligence Research (2019)

Summary

This position paper makes the case that multi-agent learning provides a setting to study how the social dynamics resulting from interaction between agents leads to discovery of intelligent policies.

The paper contrasts this with the setting of a single agent, where without constant interventions by creating of more challenging environments, the agent's intelligence hits a ceiling. Multi-agent settings circumvent this problem due to the non-stationary, and often adversarial, environment, which forces each agent to constantly keep evolving.

Strengths

- The viewpoint proposed in the paper is very visionary, and provides a fertile ground for research.

- The problem of curriculum design is important and yet understudied. This paper provides valuable insights on how multi-agent settings can induce autocurriculum.

- The paper uses the settings of biological organisms and societies as inspirations for many of its ideas. This cross-pollination of ideas can provide enlightening perspective.

- The paper points out some of the challenges, like forgetfulness, inherent in multi-agent learning; the setting of human societies is unique and the paper briefly discusses how culture and language prevents us from them.

- The paper is very clearly written and provides many references to explore further.

Weaknesses

- For a manifesto, the paper is too short, and doesn't discuss any setting in enough detail for the case to be fully convincing.

- While the paper collects a lot of ideas spread throughout the literature, it doesn't attempt to make any of them concrete. This leaves the task of formally studying the phenomenon of autocurriculum and the resultant innovations to follow-up work.
For example, the setting of multi-agent reinforcement learning is probably too inflexible to study the settings described in the paper. Humans often broadcast the rewards they receive, and in response others recalibrate their reward functions. Modeling such flexible rewards is an open problem.

- The paper conflates exploration with learning. While learning entails exploration, it also involves an intelligent use of the observed data.

Final Decision?

Accept