Tamino Jakobs

Theory of Mind

Source: http://proceedings.mlr.press/v80/rabinowitz18a.html

1. What actually was done?

The scientists created a neural network called “ToMnet” that should be able to learn from experience, such as our brain does. The “ToMnet” comprises three neural networks, the first that should analyze the current actions, the second analyzes the current beliefs, the third one collects the results of the two and should prognosis the next steps of the machines with these characteristics. There were three characters, one machine that couldn’t see, the second couldn’t remember the recent steps and the third couldn’t see or remember. The “ToMnet” could identify after a few steps from the machines with different characters, which character they were and could predict the future steps.

2. What was the experiment (with numbers and results)?

After a few steps “ToMnet” could identify which character the machines were. There is no exact number of attempts. At the end, “ToMnet” could identify when the machines had a wrong belief such as being nearsighted. They trained a ToMnet to observe Npast $\sim$ U{0, 5} full trajectories of randomly-selected agents before making its behavioural prediction. But finally, “ToMnet” is still not on the level as human children.

3. What ideas from ToM were used?

To predict the future behavior of machines based on their past behavior and the current belief in order let machines better work together. The basic idea they used is to be aware of the beliefs, desires and intention of other’s.

4. What is the solution architecture?

Its the ToMnet architecture, which is comprised on three modules, the character net, the mental state net and the prediction net. Character net characterises the agent by parsing the previously perceived episode. The goal of the mental net is to mentalise about the presented agent during the current episode. Lastly the prediction net leverages the character net and the mental net to predict the subsequent moves of the agent.