UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Improving Multi-Agent Cooperation using Theory of Mind

Permalink

https://escholarship.org/uc/item/5tj7v8zq

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Lim, Terence
Tio, Sidney
Ong, Desmond

Publication Date

2020

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Improving Multi-Agent Cooperation using Theory of Mind

Terence Lim
National University of Singapore, Singapore, Singapore

Sidney Tio
AI Singapore, Singapore, Singapore

Desmond Ong
National University of Singapore, Singapore, Singapore

Abstract

Recent advances in Artificial Intelligence have produced agents that can beat human world champions at games like Go, Starcraft, and Dota2. However, most of these models do not seem to play in a human-like manner: People infer others' intentions from their behaviour, and use these inferences in scheming and strategizing. Here, using a Bayesian Theory of Mind (ToM) approach, we investigated how much an explicit representation of others' intentions improves performance in a cooperative game. We compared the performance of humans playing with optimal-planning agents with and without ToM, in a cooperative game where players have to flexibly cooperate to achieve joint goals. We find that teams with ToM agents significantly outperform non-ToM agents when collaborating with all types of partners: non-ToM, ToM, as well as human players, and that the benefit of ToM increases the more ToM agents there are. These findings have implications for designing better cooperative agents.

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).