Review-26

"Other-Play" for Zero-Shot Coordination

While Self-Play (SP) methods demonstrate improved payoffs in centralized training with decnetralized control, these algorithms suffer at the hands of zero-hot coordination. In order to make agents cooperate with novel players at test time, one needs to construct novel policies which are robust to invariant symmetries in the MDP of agents. The work introduces "Other-Play" (OP) which yields robust policies that enhance SP as a result of robustness to unknown symmetries in the underlying MDP. OT provisions the usage of joint policies which are invariant under symmetries and provide emta-equilibrium points correspondng to all agents i nthe multi-agent setup. Empiricial evaluation of OT when agents are paired with novel agent and human partners demonstrates the suitabiity of the proposed approach.

OT utilizes coordinated symetry breaking to receive higher payoffs. Unknown symmetries present int eh MDP prohibit collaboration between novel agents as policies are prone to their variations. This is proved using a small lever coordination game wherein SP agents fail to coordinate. OT on the otehr hand, succesfully breaks symmetries and allows agents to make suitable choices while collaborating. ymmetries in the MDP are constructed as bijections which comprise of equivalence mappings of its various components such as state and action spaces. The OT objective function maximizes payoffs when randomly matched with symmetry-equivalent policy of its partner agent. As a result, resulting policies are robust to symmetries and comprise of the fixed point of meta-equilibrium with the joint olicy being the Best Response (BR) to the mixture of symmetry permutations. OT is implementted using Deep Reinforcement Learning (DRL) wherein agents having different architectures are randomly paired with each ohter at execution time. Additionally, agents are paired with human players to collaborate and yield suitable results for human-AI collaboration.

OT demonstrates suitable results in Hanabi in comparison to conventional state-of-the-art SP benchmarks. Suitability of coordination is established by observing that OT agents collaborate well and select actions which are interpretable to their partners. In addition to inter-agent collaboration, experiments comprise of human-AI collaboration wherein OT coordinates well with human players and demonstrates success on 15 out of 20 Hanabi per-seed games. While the work demonstrates sufficient promise for human-AI and zero-shot coordination, it lacks a solid definition of the symmetry problem. Symmetry matching does not naturally appear as a significant problem in practical scenarios. For instance, the example provided of a driver driving a car on left or right side of the road is invariant in the MDP and does not demonstrate why symmetries persist as a problem. Furthermore, it would be interesting if more insights into the structure and occurence of symmetries were provided which would depict their presence as a hindrance to coordination.

OT presents a zeros-shot