Review-18

Weighted QMIX: Expanding Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning

Suitability of QMIX in Multi-Agent Reinforcement Learning (MARL) has been pivotal in training decentralized policies in the centralized setup. While QMIX aids in an expressive joint Q-function utilizing monotonic constraints in mixing, it often fails to recover representations wherein the ordering of actions may depend on other agent's actions. This requires one to rethink the mapping between Q-space and Q_{mix}-space from an operational point of view. The work throws light on this aspect by formulating the QMIX operator which demonstrates nuances in its projections. In order to address these nuances, two novel weighting schemes are proposed which place more importance on better joint actions. Centrally-Weighted QMIX (CW-QMIX) and Optimistically-Weighted QMIX (OW-QMIX) demonstrate improved performance on StarCraft II micromanagement scenarios.

The incorrect $argmax$ estimations of QMIX projections demonstrate its limitation to recover optimal policies in case of non-monotonic scenarios. The problem is further highlighted by observing that QMIX restricts function class approximations which fail to suitably project Q-values in the set of all factorizations, hence resulting in a gap (action gap) between Q_{tot} and Q^*. This is evident from the formulation of QMIX as an operator which does not possess a unique fixed point. Additionally, the operator underestimates the value of optimal joint action which is a direct consequence of inaccurate $argmax$ approximations in projection space. To this end, the work introduces weighting in the original QMIX objective which places importance on favourable joint actions. The initial QMIX objective comprises of a unit weighting. On the other hand, proposed weighting schemes downweight Q-values which either represent underestimations or sub-optimal actions. While CW-QMIX strictly weights the importance of joint action w.r.t optimal approximates $\hat{Q}^*$, OW-QMIX optimistically weights Q_{tot} exactly.

CW-QMIX and OW-QMIX demonstrate improved performance on SMAC StarCraft II micromanagement scenarios and Predator Prey which require greater collaboration per timestep. In addition to performance, the weighting schemes depict robustness to increased exploration indicating suitability of weighting in retrieving expressive joint action representations. While the weighting scheme stabilizes the performance and expressivity of mixing, it presents two shortcomings. Firstly, in case of scenarios with large number of agents and difficuly dynamics, CW-QMIX and OW-QMIX require extensive exploration to converge to an optimal policy. This may indicate that weighting Bellman updates leads to sample-efficient learning. Lastly, the work highlights the limitation of learning $\hat{Q}^*$ as a potential limitation since it aggregates complexity in the network architecture. This may be addressed using a more sophisticated robust weighting scheme which does not affect Q with the increasing number of agents.

Introduction of novel weighting schemes for improving joint action projections of QMIX are an apt suggestive for collaboration in MARL. The work hints at two new directions for future research. Firstly, suitability of CW-QMIX and OW-QMIX still remain an open problem on temporally-extended and extensive MARL problems consisting of large number of agents. These could be tackled by combining weighting with other RL methods such as Hierarchical RL. Lastly, introduction of alternate weighting schemes which eliminate learning of $\hat{Q}^*$ could be studied for improved performance of QMIX architecture.