The Representational Capacity of Action-Value Networks for Multi-Agent Reinforcement Learning

Extended Abstract

Jacopo Castellini
Dept. of Computer Science
University of Liverpool
Liverpool, United Kingdom
J.Castellini@liverpool.ac.uk

Rahul Savani
Dept. of Computer Science
University of Liverpool
Liverpool, United Kingdom
rahul.savani@liverpool.ac.uk

Frans A. Oliehoek
Interactive Intelligence Group
Delft University of Technology
Delft, The Netherlands
F.A.Oliehoek@tudelft.nl

Shimon Whiteson
Dept. of Computer Science
University of Oxford
Oxford, United Kingdom
shimon.whiteson@cs.ox.ac.uk

ABSTRACT
Recent years have seen the application of deep reinforcement learning techniques to cooperative multi-agent systems, with great empirical success. In this work, we empirically investigate the representational power of various network architectures on a series of one-shot games. Despite their simplicity, these games capture many of the crucial problems that arise in the multi-agent setting, such as an exponential number of joint actions or the lack of an explicit coordination mechanism. Our results quantify how well various approaches can represent the requisite value functions, and help us identify issues that can impede good performance.

KEYWORDS
multi-agent systems; neural networks; decision-making; action-value representation; one-shot games

ACM Reference Format:

1 INTRODUCTION
In this paper, we focus on value-based multi-agent reinforcement learning (MARL) [2, 5, 6, 14, 20, 24] approaches for cooperative multi-agent systems (MASs) [10, 12, 22, 26, 29]. Value-based single-agent RL methods use (deep) neural networks to represent the action-value function \( Q(s, a; \theta) \) to select actions directly [17] or as a ‘critic’ in an actor-critic scheme [13, 16]. Current deep MARL approaches are either based on the assumption that the joint-action value function \( Q(s, a) \) can be represented efficiently by neural networks (when, in fact, the exponential number of joint actions usually makes a good approximation hard to learn and scales poorly in the number of agents [4]), or that it suffices to represent individual action values \( Q_i(s_i, a_i) \) [15, 25], that are known to be hard to learn because of non-stationarities from the perspective of a single agent due to the simultaneous learning of the others [4, 25, 28].

To overcome these difficulties and be able to learn useful representations while not incurring excessive costs, a middle ground is to learn factored \( Q \)-value functions [7, 8], which represent the joint value but decompose it as the sum of a number of local components, each involving only a subset of the agents.

This paper examines the representational capacity of these approaches by studying the accuracy of the learned \( Q \)-function approximations \( \hat{Q} \), as recently factored approaches have shown some success in deep MARL [20, 23]. We consider the optimality of the greedy joint action, which is important when using \( \hat{Q} \) to select actions, and the distance to optimal value \( \Delta Q = |Q - \hat{Q}| \). Minimising \( \Delta Q \) is important for deriving good policy gradients in actor-critic architectures and for sequential value estimation in any approach (such as \( Q \)-learning) that relies on bootstrapping.

To minimise confounding factors, we focus on one-shot (i.e., non-sequential) problems [19] that require a high level of coordination. Despite their simplicity, these one-shot games capture many of the crucial problems that arise in the multi-agent setting, such as an exponential number of joint actions. Thus, assessing the accuracy of various representations in these games is key step towards understanding and improving deep MARL techniques.

2 ACTION-VALUE FUNCTIONS FOR MARL
In many problems, the decision of an agent is influenced by those of only a small subset of other agents [8, 9] and thus the joint action-value function \( Q(a) \) can be represented as a factorization, i.e. a sum of smaller action-value functions \( Q_e(a_e) \) defined over a coordination graph [7, 11, 21] describing these influences. However, there are many cases in which the problem itself is not perfectly factorized according to such a graph, or the underlying factorization may be unknown beforehand. In these cases, however, it can still be useful to resort to an approximate factorization [8]:

\[
Q(a) \approx \hat{Q}(a) = \sum_e \hat{Q}_e(a_e),
\]
obtained by considering a decomposition of the original function into a desired number of local approximate terms $Q_e(\alpha_e)$, thus forming an approximation $\tilde{Q}$ of the original action-value function.

We investigate the representations obtained with the proposed \textit{fac-tored Q-function} approach. There, we also investigate another approach called the \textit{mixture of experts} approach [1], which, in the case of one agent per factor, corresponds to independent learners [25].

![Figure 1: Example coordination graphs for: (a) random partition, (b) overlapping factors, (c) complete factorization.](image)

We investigate four different coordination graph structures used to approximate the action-value function:

- \textit{Single agent decomposition} [9]: each agent $i$ is represented by an individual neural network and computes its own individual action-values $Q_i(a_i)$, one for each output unit, based on its local action $a_i$. This corresponds to the value decomposition networks from [23].

- \textit{Random partition}: agents are randomly partitioned to form factors of size $f$, with each agent $i$ in the team $D$ involved in only one factor, resulting in $\frac{|D|}{f}$ factors.

- \textit{Overlapping factors}: a fixed number of factors is picked at random from the set of all possible factors of size $f$.

- \textit{Complete factorization}: each agent $i$ is grouped with every possible combination of the other agents in the team $D \setminus i$ to form factors of size $f$, resulting in $\prod_{i=1}^{\left|D\right|} |D| - 1$ factors.

In each of the investigated factorizations, each factor is represented by an individual neural network that represents local action-values $Q_e(\alpha_e)$, using an output unit for each one, for a certain factor $e$, where $\alpha_e$ is the local joint action of agents in factor $e$. We consider factors of size $f \in \{2, 3\}$. Also, a \textit{joint learner} with an exponential number of output units is used as a baseline.

### 3 EXPERIMENTS

We investigate the representations obtained with the proposed factorizations on a series of challenging one-shot coordination games with $n = 6$ agents that do not present an explicit decomposition of the reward function $Q(\alpha)$ (non-factored games), and then on two factored games. Complete results and plots for all the proposed games are presented in the full-length paper available online [3].

Here, we are reporting results for only two of them: the non-factored \textit{Climb Game} [27], known to enforce a phenomenon called \textit{relative overgeneralization} that pushes the agents to underestimate a certain action even if it is optimal when perfectly coordinating on it, and a one-shot version of the factored game \textit{Aloha} [18], in which neighbouring agents in the coordination graph can interfere with agents’ actions.

Table 1 presents the accuracy using various measures of the investigated representations on these two problems in terms of reconstruction error, action ranking and action selection. We report mean values and standard errors across 10 runs. Some of our main findings are:

- There are pathological examples where all types of factorization result in selecting the worst possible joint action. Given that only joint learners seem to be able to address such problems, currently no scalable deep RL methods for dealing with those seem to exist.

- Beyond those pathological examples, ’complete factorizations’ of modest factor size yield near perfect reconstructions and rankings of the actions, also for non-factored action-value functions, while exhibiting better scaling behaviour.

- For these more benign problems, random overlapping factors also achieve excellent performance.

### 4 CONCLUSIONS

In this work, we investigated how well neural networks can represent action-value functions arising from multi-agent systems. This is an important question since accurate representations can enable taking (near-) optimal actions in value-based approaches, and computing good gradient estimates in actor-critic methods. In this paper, we focused on one-shot games as the simplest setting that captures the exponentially large joint action space of MASs. We compared a number of existing and new action-value network factorizations and learning approaches.

Our results highlight the difficulty of compactly representing action values in problems that require tight coordination, but indicate that using higher-order factorizations with multiple agents in each factor can improve the accuracy of these representations substantially. We also demonstrate that there are non-trivial coordination problems - some without a factored structure - that can be tackled quite well with simpler factorizations. Intriguingly, incomplete, overlapping factors perform very well.

### ACKNOWLEDGEMENTS

This research made use of a GPU donated by NVIDIA. F.A.O. is funded by EPSRC First Grant EP/R001227/1. This project received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 758824 – INFLUENCE).
REFERENCES


