Improving sample efficiency with expert communication

Mathijs Mul, Thomas Unger, Gautier Dagan, Bence Keresztury, Elia Bruni

The BabyAI grid world (Chevalier-Boisvert et al., 2018) tests agents' ability to follow operational (e.g. 'put a ball next to a purple door') and navigational (e.g. 'go to the red ball') instructions expressed in natural language in a partially observable environment. So far, research using BabyAI has focused on the development of trainable agents, optimized through reinforcement or imitation learning. However, they are always on their own.

Here, we extend the paradigm with a neural expert agent, whose goal is to provide other agents with additional linguistic feedback to improve their learning efficiency.

We optimize the expert agent in a pre-training phase, where a discrete communication channel serves as a bottleneck between an input-processing module and a policy module. This requires no additional training signal. Next, when we train a new agent from scratch, we augment its regular sensory and linguistic input with the message produced by the expert agent at a given time step. We then investigate whether communication aids the learning process of a new agent and, if so, analyze the linguistic properties of the generated messages. We are interested in the correlations between emitted expressions by the expert agent and performed actions by the learner, the causal weight of these expressions in relation to the regular input, and the possible emergence of compositionality in the communication protocol.

Our first experiments on a basic BabyAI level have shown that access to the utterances of a pre-trained expert agent significantly improves the data efficiency of a new agent, while also smoothing the learning curve. Expert's messages strongly correlate with optimal actions when the expert is confident about which action to take. Our next steps are to check whether these findings hold for higher game levels, to perform a more extensive analysis of the developed communication protocol, allowing experts to communicate only every k timesteps and to add communication from the learner to the expert.

Furthermore, we want to incentivize the expert to produce more general, compositional messages. This can be done through a cultural evolutionary process (see, e.g. Smith and Kirby, 2012) where multiple experts and learners are trained and they are paired randomly at every run. By introducing new untrained agents into the existing culture, the newbie has to be able to pick up the language quickly. This will push the emergent language not only to be expressive but learnable as well, which could give rise to compositional properties.

Finally, we hope to streamline the expert pre-training by optimizing expert agents for multiple levels.

References

Chevalier-Boisvert, M., D. Bahdanau, S. Lahlou, L. Willems, C. Saharia, T. H. Nguyen, and Y. Bengio
2018. Babyai: First steps towards grounded language learning with a human in the loop. *arXiv preprint arXiv:1810.08272*.

Smith, K. and S. Kirby
2012. Compositionality and linguistic evolution. In *The Oxford handbook of compositionality*.