CSCI6505 2021W Project Proposal

Jasper Dupuis
Mani Teja Varma Kucherlapati

February 12, 2021

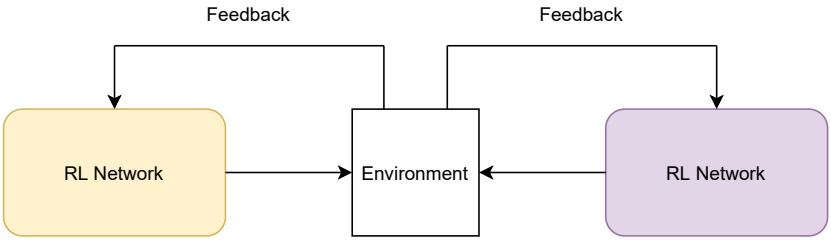
1 Project overview and identifying success

Topic: Iterated social dilemmas with imperfect communication using reinforcement learning models.

Our proposal is to look at the effect of communication on multi-agent iterated social dilemmas populated by reinforcement learning (RL) models.

Figure 1 shows the simplest RL problem for the sequential social dilemma. Two agents make a blind decision, input it to the environment, and receive as feedback their reward[1]. It is this reward that allows for optimization by the two agents. This work has been done for RL models in simple environments, offering a way to begin our work.

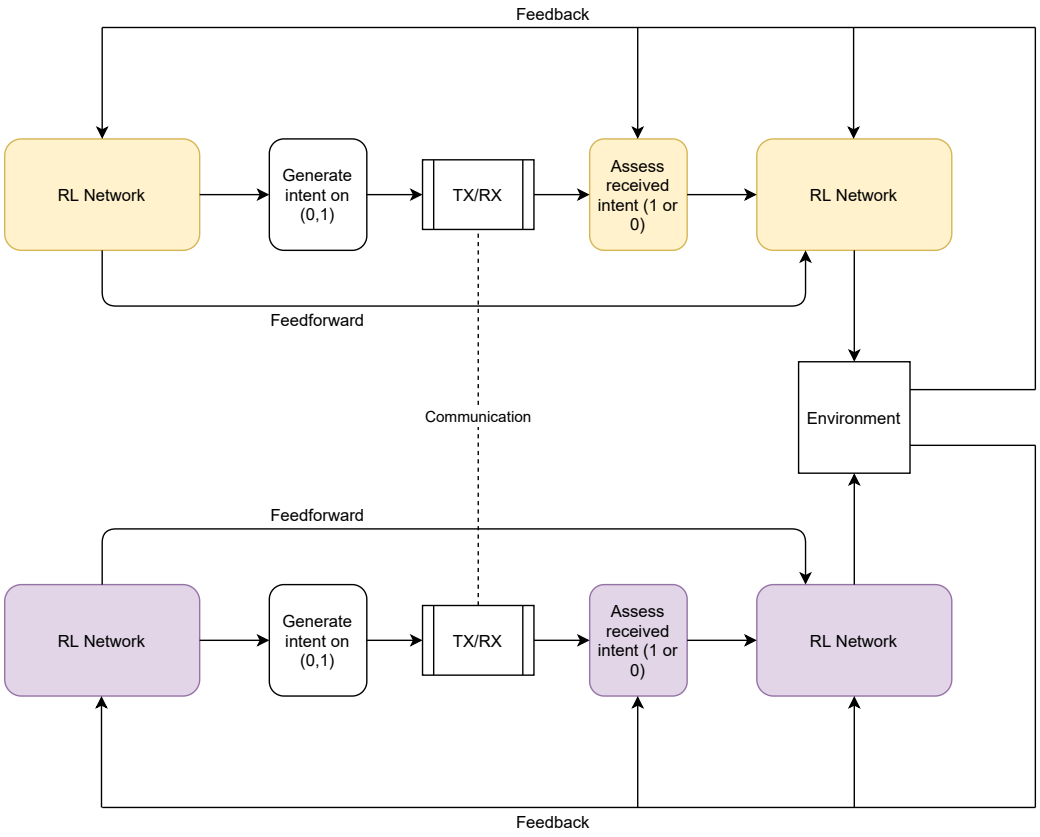
Figure 1: Simple representation of two-agent reinforcement learning.



We propose to investigate the effect of communications on the above framework. Figure 2 shows a schematic of what this network may look like. The major difference here is the two agents exchanging a number on (0,1) representing intent with associated truthfulness. That is, an agent may plan to act nefariously but will still try to fool the other agent. The number may be derived from an article on truthfulness in the game of Diplomacy, in which humans detect deception about 80% of the time [2]. Note this use of a float sidesteps earlier concerns about generating and assessing natural language.

The inclusion of communication introduces two more opportunities for learning in addition to the original strategy from Figure 1. These are assessing the received intent, and devising a new strategy based on original strategy and assessment of received intent. To use human terms, these might be named "gut feeling" for the first and "conviction" for the second. Other research has considered this effect as noise, where trained or deterministic decisions are flipped by a rate up to 5% [3].

Figure 2: Proposed architecture for experimentation.



If we can succeed at generating interesting results using just RL models, we expect the work to this point to represent a good course project. Failure to successfully train models, or failure to train differentiable ones, could be acceptable if we could explain why our addition of "gut feeling" and "conviction" as learnable parameters fails. We could then investigate replacing these learnt parameters with hyperparameters of our choosing as a method of exploring weight space.

This project could also fail to produce a model that performs substantially differently from other RL or deterministic models in the literature. This is not the expected outcome. There is an appealing intuition that the known deterministic models would perform worse in an environment in which they were forced to communicate - and therefore occasionally fail to lie - with our RL-trained agents.

Ultimately it would be interesting to tackle this problem with three or more trained agents. This is very much a stretch goal.

2 Explicit proposal deliverables

Major parts of project with their "value" if they're the end state:

1. Reproduce simple RL models (Low success project, reading week),
2. Introduce communication and train, as in Figure 2, (Successful project, Feb 26),
3. Introduce communiction to benchmark agents and compare results (Super successful project with possibility of publishing, March 23).

There is no dataset, just feedback from a prisoner's dilemma model of the environment. This is already available in a Python library[4].

We think training basic RL agents will be straightforward, as this has already been done. This will let us figure out how to perform RL training.

The hard part will be implementing the training of multi-block agents as in Figure 2. It's not immediately clear how best to approach this right now.

Our first step for all of us is to make sure we're conversant with all the references to this proposal, and to begin trying to reproduce RL models on the basic version of this task. We've both agreed to do this in parallel to maximize learning and utility in the future.

3 References

References

[1] D. R. Hofstadter, *Metamagical themas : questing for the essence of mind and pattern*. Basic Books New York, 1985.

[2] D. Peskov, B. Cheng, A. Elgohary, J. Barrow, C. Danescu-Niculescu-Mizil, and J. Boyd-Graber, "It takes two to lie: One to lie and one to listen," in *Association for Computational Linguistics*, 2020.

[3] M. Harper, V. Knight, M. Jones, G. Koutsovoulos, N. E. Glynatsi, and O. Campbell, "Reinforcement learning produces dominant strategies for the iterated prisoner's dilemma," *PLOS ONE*, vol. 12, pp. 1–33, 12 2017.

[4] The Axelrod project developers, "Axelrod: Tournament repository," Apr. 2016.