Iterated Communication Through Negotation

Michael Noukhovitch
MILA
michael.noukhovitch@umontreal.ca

Abstract

Recently, [1] investigated the emergence of language in a negotiation game setting. We take a critical look at the setup, execution, and conclusions of this work. We note several drawbacks and missteps, and extend the work by correcting for them and updating the conclusions of this very interesting direction.

1 Introduction

One of the first philosophers of language, Ludvig Wittgenstein, posited that "language is use" [8]. This idea, that the use of language is what gives it its meaning, is a profound statement that also has consequences for how we think of language. Wittgenstein saw language as wholly tied to its use, there could be no language separate from reality or possible use. To this end, he defined language games as games with simpler forms of language "consisting of language and the actions into which it is woven".

Recently, the AI community has taken this philosophy of language and sought to use it as the basis for the communication of autonomous agents [7]. The field of "emergent communication" seeks to understand language starting from the most basic of language games; the goal is to teach agents to communicate with each other by grounding their language and interactions in a simpler world described by some "game." This game can be as simple as two agents meeting at some point in a gridworld [2] to something as complicated as self-driving car interactions [6].

In their ICLR paper, [1] look at a negotiation game as the basis of their investigation of emergent communication. The game involves two agents A, and B who seek to divide a pool of items p. Each agent has a hidden utility over the items u and the agents take turns creating proposals P of how to split the pool between themselves (e.g. if there are 5 of item 1, an agent can suggest to split the pool 4 to themselves and 1 to the opponent). Once an agent accepts their opponents proposal, agents recieve a reward given their type: selfish agents i only care about themselves and recieve a reward of $u_i \cdot P$, whereas prosocial agents care equally about their opponent and recieve reward $0.5(u_A \cdot P + u_B \cdot P)$. Agents can communicate across two possible channels during the game, the *proposal channel* where they can inform their opponent about their proposal, and the *linguistic channel* where agents can send a sequence of discrete symbols. [1] find that prosocial agents learn to use the linguistic channel and perform better, whereas selfish agents do not and perform worse. They posit that sociality may be a necessary trait for communication to emerge.

2 Related Work

The negotiation game is a classical example of a zero-sum game from game theory [3] but the approach to solving it is based off of recent work on cooperative multi-agent reinforcement learning (MARL) [5]. This MARL approach test the linguistic idea that cooperation is necessary for language emergence [4] in the familiar context of emergent language RL [7], but augmented with deep RL [**?**].

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

3 Reproduction and Criticism

3.1 Reproduction

The reproduction code is an extension of Hugh Perkins' attempt at reproducing the paper `https://github.com/ASAPPinc/emergent_comms_negotiation` with many notable fixes and extensions. The paper itself is relatively reproducible with prosocial agent results being stronger than selfish agent results. And prosocial agents using language as well as proposal being more successful than agents just using proposals. The exact numbers achieved were not replicated, neither were the results that prosocial agents with just a linguistic channel achieved the highest score. Selfish agents in general tended to take longer to converge and also had higher variance in their end result.

Reproducibility was not greatly hampered by resources as the code is not very intensive but each full 500k episode run (with batch size 128) does take at least 10 hours on a reasonable GPU. This means that reproducing the 10 run average result and verifying any changes in a resilient way is probably only feasible with access to a GPU cluster though there is quite a bit of room for speedups.

3.2 Criticism

After reproducing results, there are still questions that are raised not about the results themselves but about methodology and approach. We investigate some of these questions.

3.3 Prosocial Agents Don't Negotiate

A negotiation game where the interests of two parties do not clash is not as much about negotiation as it is just sharing preferences and calculating the optimum. This is exactly what is seen in case of prosocial agents with only a linguistic channel: one agent communicates their preferences and the other agent finds the optimal division and proposes it. In this way, the goal and even the whole task is completely different for prosocial agents compared to selfish agents whose selfish opponents do not have their best interests in mind. In this way, selfish agents have no incentive to communicate their preferences as it could potentially weaken their negotiating ability. In this way, we see that it is not selfish agents failing to learn to communicate but learning to not communicate not because they are selfish but because the situation does not incentivise them to in a one-shot environment. This is explored in 4.1

3.4 Unmotivated Agents

The paper and unfinished repository both allow for sampling situations that give no possible reward to an agent, either through sampling a pool with 0 objects of any kind or sampling a utility for an agent such that there are no object for which they have non-zero utility (e.g. [1,0,1]). In this case, any rational agent will accept any opponent's offer and there is no motivation to negotiate. As well, if both agents are unmotivated then *joint reward optimality* of [1] will be $\frac{0}{0}$.

3.5 Fractional Reward is Biased

The metric used by the paper "fraction of joint reward" compares the reward achieved by each agent (utility over the accepted proposal $u_i \cdot P$) with the total reward possible. In the case of prosocial agents the metric measures the optimality as both agents' rewards are the same (an equally weighted sum of their individual rewards) and they seek to optimize the total possible reward.

$$FR_{\text{prosocial}} = \frac{u_A \cdot P + u_B \cdot P}{\max u_A \cdot P + u_B \cdot P} \tag{1}$$

where u is utility

p is the accepted proposal

This reward scheme is extended to selfish agents in [1]. But in the case of selfish agents, the total possible reward is the not the objective the agent is optimizing. The agents optimize their own reward and using the total possible reward metric is therefore a bad way of judging performance, and inaccurate for judging an agent's negotiation ability. Instead, you could measure how well each agent does at maximizing their own reward, average across agents, and divide by the total possible reward

$$FR_{\text{selfish}} = 0.5 * \frac{u_A \cdot P}{\max u_A \cdot P} + 0.5 * \frac{u_B \cdot P}{\max u_B \cdot P} \tag{2}$$

But this is also wrong because the maximum possible reward for one agent will usually decrease the reward for the opponent. In this way, the only situation where $FR_{\text{selfish}} = 1$ is possible is if there do not exist any items for which both agents have a non-zero utility. Since this situation is quite rare (the whole point of negotiation is to resolve conflicting preferences) this metric is unfair to selfish agents by usually being lower than for prosocial agents and artificially makes them appear worse. Total possible reward is therefore a bad metric for selfish agents and using a different metric is proposed in 4.2

4 Exploratory Extensions

4.1 Utility Channel

One way to test whether selfish agents are learning to communicate or whether they are specifically learning not to communicate is to simulate what information an agent could possibly communicate. Since the prosocial agents learn to communicate their hidden utilites, it seems fitting to test if giving full information of hidden utilities helps achieve better results. To this end, we experiment with giving an agent information about the opponent's utility similar to a communication channel and plot agents' rewards in fig **??**.

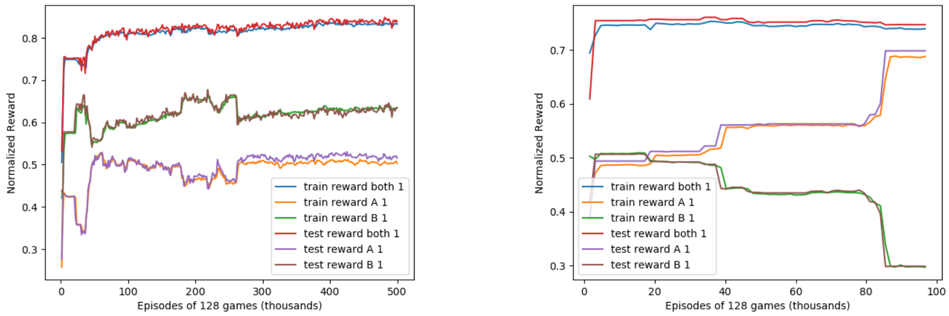

Figure 1: The agent that knows the other's utilities, dominates

Agent B Agent A	Communicate	Don't Communicate
Communicate	Equal	B dominates
Don't Communicate	A dominates	Equal

Table 1: payoff matrix for communicating utility as a selfish agent

For selfish agents we find that an agent tends to dominate its opponent if it knows the opponent's preferences, and outcomes are relatively similar if neither agent knows the opponent's preferences. This means that for selfish agents communicating their utility as prosocial agents do can be described with the payoff matrix in table 1. In this way, selfish agents learn the optimal choice to not communicate their utilities because they are disincentived to in this specific situation.

4.2 Equitable Reward Metric

Since fractional reward is biased as explained in 3.5, a better metric for selfish agents is proposed to measure the performance of their negotiation: equitable reward. Since each agent's preferences may be at odds with the other, it is better to measure the agents' performances taking into account a comparison of their respective utilities. We consider optimality if each agent recieves a proportion of the items in the pool corresponding to the fraction of their utility over their utility and their opponents. Essentially, if each agent recieves items according to how much they care about them:

$$ER_{\text{selfish}} = \frac{u_A \cdot P}{(u_A + u_B) \cdot P} \tag{3}$$

Since for any given P, for any selfish agents, $ER(A) + ER(B) = 1$, and since we expect two random uniform policies to be equal in expectation, therefore we expect both agents to perform with ER 0.5 in expectation for them to be performing fully equitably. This result is found in running the selfish agent with just a proposal channel in fig 2 comparing the ER rewards (e.g. train reward A) with the *joint reward optimality* (e.g. train reward both)

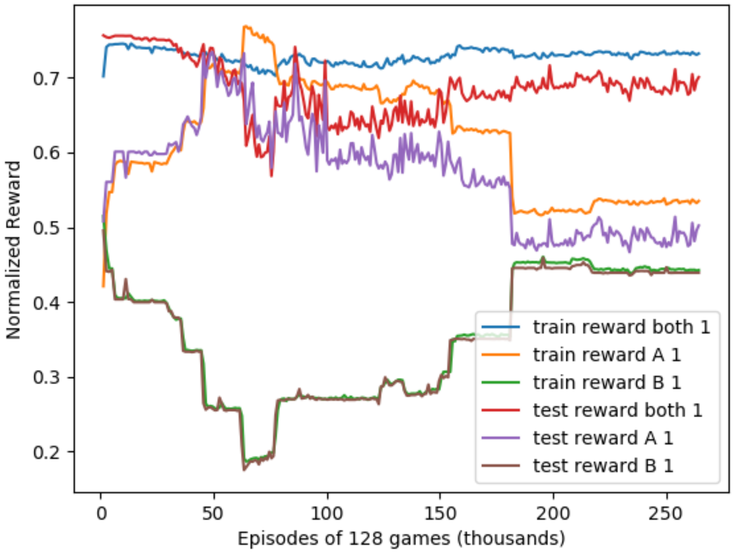

Figure 2: Selfish agents Using proposal channel

4.3 Utility Sampling

One consequence of randomly sampling utilities is that there is no guarantee on the clash of utilities in negotation as the players could have non-zero utility only for the items that their opponent has zero utility and negotiation is simplified. To combat this issue, as identified in 3.4 it is proposed to guaranteee non-zero utility to every item.

Another issue is that in the social case, one player's utilities could dominate the other's $u_j^1 > u_j^2 \forall j$. In such a case, the optimal strategy for both players is to give all items to the player with the dominating utility, and again the player. The final split is therefore pareto optimal, but doesn't feel "fair" for the side of the dominated player. A similar situation for a selfish agent would generally lead to a more even split with a smaller total reward. For this reason, we can experiment with avoiding domination situations by normalizing the utilities so that each agents utilities all sum to 15.

5 Conclusion

After thoroughly investigating the paper, we find certain flaws and extend the work to cover them. Still, the main crux of the paper is that cooperation is necessary for linguistic communication to emerge, and this thesis is not well enough defended. We find that the experiments are biased towards prosocial agents and through the setup deter selfish agents from communicating. In this way, we find that it is not selfish agent failing to communicate but learning specifically not to. For this reason, the idea that cooperation is necesary for language is not thoroughly tested. Instead, it makes sense to reduce our requirements and compare the idea of [8] that it is not cooperation but opponent awareness that facilitates communication. For this reason, we propose future work to investigate a model of selfish agents that are incentivized to communicate to better understand their opponent, without having the need to communicate with them.

References

[1] Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Z Leibo, Karl Tuyls, and Stephen Clark. Emergent communication through negotiation. In *International Conference on Learning Representations*, 2018.

[2] Claudia V Goldman, Martin Allen, and Shlomo Zilberstein. Learning to communicate in a decentralized environment. *Autonomous Agents and Multi-Agent Systems*, 15(1):47–90, 2007.

[3] John F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[4] Martin A Nowak and David C Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999.

[5] Liviu Panait and Sean Luke. Cooperative multi-agent learning: The state of the art. *Autonomous Agents and Multi-Agent Systems*, 11(3):387–434, November 2005.

[6] Cinjon Resnick, Ilya Kulikov, Kyunghyun Cho, and Jason Weston. Vehicle community strategies. *arXiv preprint arXiv:1804.07178*, 2018.

[7] Kyle Wagner, James A Reggia, Juan Uriagereka, and Gerald S Wilkinson. Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37–69, 2003.

[8] Ludwig Wittgenstein. *Philosophical investigations*. John Wiley & Sons, 2009.