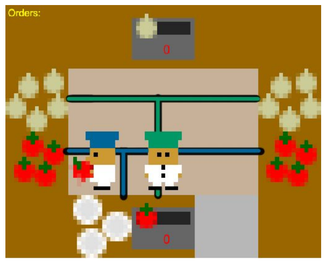
Adapting to Unknown Conventions in Cooperative MARL

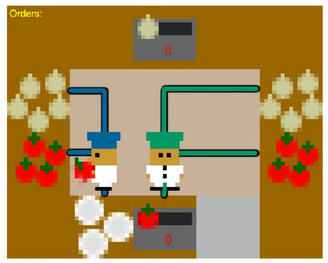
Bidipta Sarkar

Motivation

MARL agents often converge to arbitrary conventions.



Split by ingredient

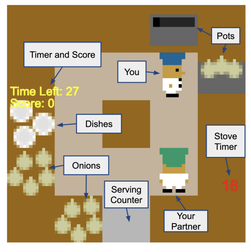


Split by side

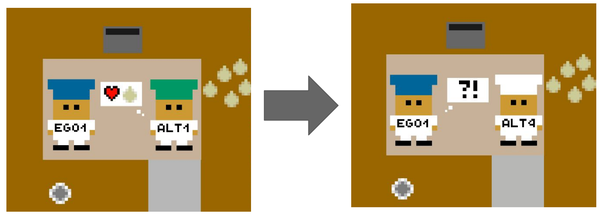
Can we adapt to new partners in real time?

Task Definition

Overcooked Game: Coordination Ring



Partner's convention defines the *Task*



Only the single-agent *transition* changes

Idea: Pre-train on a diverse set of partners to learn how to collaborate with others

Generating Conventions

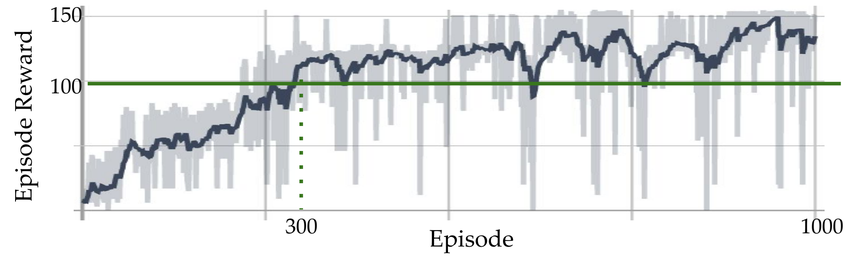
CoMeDi: generates diverse set of conventions

$$\mathcal{L}(\pi^n) = \mathcal{J}(\pi^n, \pi^n) - \beta\mathcal{J}(\pi^n, \pi^*) + \gamma\mathcal{J}_{MP}(\pi^n, \pi^*)$$

High Scoring (SP) — Semantically Different (XP) — Good Faith Actor (MP)

Generated 8 conventions where SP > 120 and XP < 20
(Meta-)Train on first 7 and (Meta-)Test on the last one

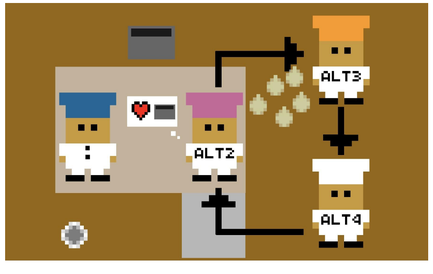
Training from Scratch



Need to play *300 games* for human-like performance (score=100)
- More than 3 hours of user data!

Common Best Response

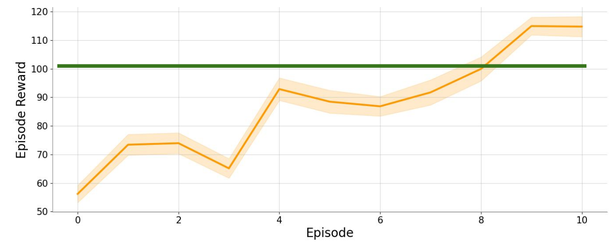
Train a single best response agent to training set.



Common technique for MARL Zero-shot coordination
- Score with test partner is 56

Few-shot Fine Tuning

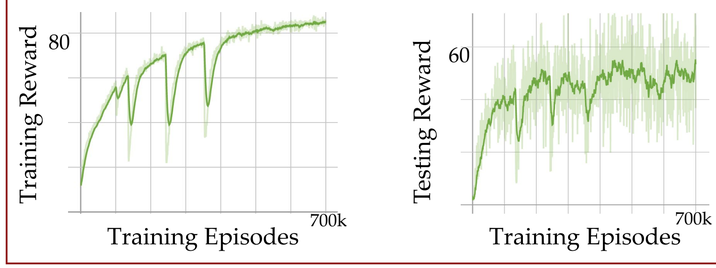
What if we *fine tune* the common best response model?



Human-like score in 10 episodes (7 mins of user data)

MAML

MAML with one episode as support data



Discussion

Using a diverse set of partners, few-shot fine tuning can adapt to *new partners*!

Efficient human-AI adaptation without datasets of human gameplay

Acknowledgements

Special thanks to Andy Shih and Dorsa Sadigh for collaborating on the CoMeDi algorithm.

References

[1] D Strouse, K McKee, M Botvinick, E Hughes, R Everett, Collaborating with humans without human data. *Adv. Neural Inf. Process. Syst.* **34**, 14502–14515 (2021).
[2] B Sarkar, A Talati, A Shih, S Dorsa, Pantheonrl: A marl library for dynamic training interactions in *Proceedings of the 36th AAAI Conference on Artificial Intelligence (Demo Track)*. (2022).
[3] C Yu, et al., The surprising effectiveness of ppo in cooperative, multi-agent games. *arXiv preprint arXiv:2103.01955* (2021).
[4]The garage contributors. Garage: A toolkit for reproducible reinforcement learning research. https://github.com/rlworkgroup/garage, 2019.