Task 2.1 (2019-20)

Information State Embedding in Partially Observable Cooperative Multi-Agent Reinforcement Learning

Weichao Mao, Kaiqing Zhang, Erik Miehling, and Tamer Başar
University of Illinois at Urbana-Champaign

Motivation

There is no known fixed-domain belief state in partially observable cooperative multi-agent reinforcement learning problems.

Goal: Can we compress the history into a fixed-domain (approximate) belief state and obtain provable performance guarantees when decisions are made according to this compression?

Model

State evolves stochastically; each agent receives noisy observations of the underlying state:

$$X_{t+1} = f_t\left(X_t, A_t^{1:N}, W_t^x\right),$$
$$Y_{t+1}^i = l_{t+1}^i\left(X_{t+1}, A_t^{1:N}, W_{t+1}^i\right),$$
$$Z_{t+1} = l_{t+1}^c\left(X_{t+1}, A_t^{1:N}, W_{t+1}^c\right),$$

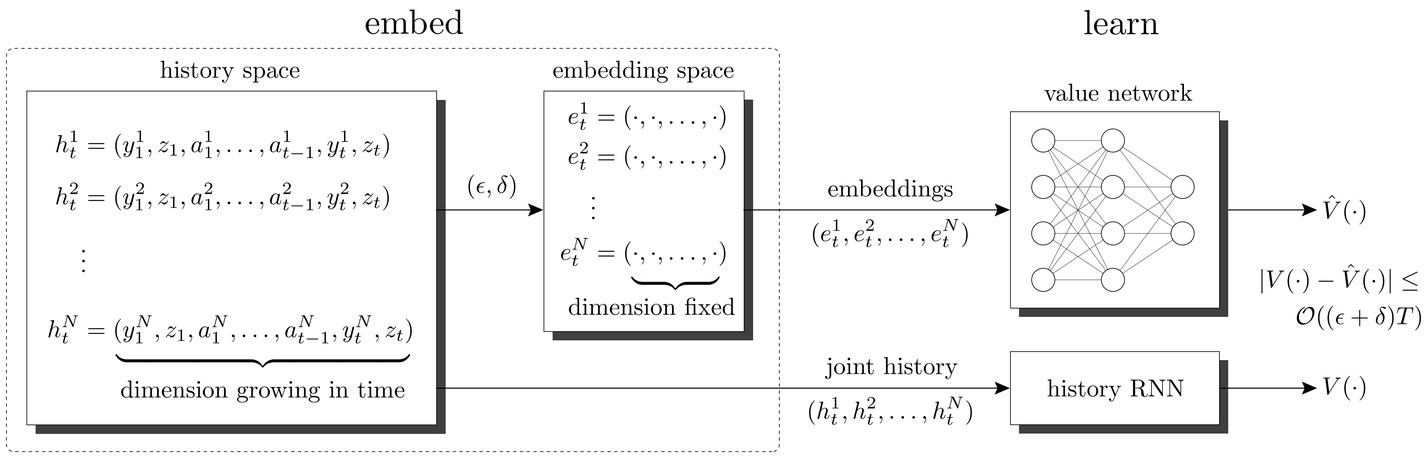
Each agent receives a (joint) reward at each time-step: $R_t(X_t, A_t^{1:N})$

Contribution

We propose the notion of an (ε, δ) **- information state embedding** quantified in terms of two factors:

i) The embedding is approximately sufficient to ***predict future rewards***.
ii) The embedding is approximately sufficient to ***predict future beliefs***.

Furthermore, we evaluate how the compression error of the embedding translates into an **error in the value functions and policies**.



Results

Our main result states that the error of the value function obtained under the information state embedding compared to the value function under history-based policies obeys:

$$|V(\cdot) - \hat{V}(\cdot)| \leq \mathcal{O}((\varepsilon + \delta)T)$$

Compared to existing state-of-the-art algorithms, our approach adopts a less stringent form of information decentralization (parameter sharing) while still achieving competitive performance with centralized learning approaches such as oSARSA (Dibangoye & Buffet, 2018), FB-HSVI (Dibangoye et al., 2016).

	Parameter Sharing			Centralized Learning	
T	RNN-E	FM-E	PCA-E	oSARSA	FB-HSVI
Grid3x3corners					
6	0.86	0.83	0.26	1.49	1.49
7	1.41	1.30	0.51	2.19	2.19
8	1.94	1.93	0.72	2.95	2.96
9	2.69	2.53	1.01	3.80	3.80
10	3.47	3.25	1.30	4.69	4.68
Dectiger					
3	4.58	4.89	0.06	5.19	5.19
4	2.97	3.78	1.00	4.80	4.80
5	1.46	2.02	0.65	6.99	7.02
6	2.50	2.95	0.71	2.34	10.38
7	0.85	1.89	0.53	2.25	9.99
Boxpushing					
3	12.63	64.92	17.81	65.27	66.08
4	65.06	76.83	17.76	98.16	98.59
5	81.51	94.22	34.28	107.64	107.72
6	91.00	97.03	34.65	120.26	120.67
7	106.76	143.53	34.23	155.21	156.42

Future Directions

1) **Development of a general theory of feature extraction for decision/control problems**

2) **Compute embeddings that explicitly minimize the compression error**

3) **Incorporate battlefield constraints into the theory:** distributed decision systems in the battlefield environment consist of devices that may have strict memory or energy limitations; features of the environment such as geographical separation/obfuscation may result in communication delays, e.g., due to the need for line-of-sight communication.

Publications

Mao, W., Zhang, K., Miehling, E., & Başar, T. (2020, December). **Information State Embedding in Partially Observable Cooperative Multi-Agent Reinforcement Learning**. In 2020 59th IEEE Conference on Decision and Control (CDC) (pp. 6124-6131).