Learning Policy Representations in Multiagent Systems

Aditya Grover¹, Maruan Al-Shedivat², Jayesh K. Gupta¹, Yura Burda³, Harrison Edwards³
¹Stanford University  º²Carnegie Mellon University  º³OpenAI

ABSTRACT

Modeling agent behavior is central to understanding the emergence of complex phenomena in multiagent systems (MAS). We propose a general learning framework for modeling agent behavior in any multiagent system using only a handful of interaction data.

Our framework casts agent modeling as a representation learning problem. Consequently, we construct a novel objective inspired by imitation learning and agent identification and design an algorithm for unsupervised learning of representations of agent policies.

LEARNING OBJECTIVE

We propose two desiderata for the learned representations in multiagent systems.

1. Generative. The representation should be useful for simulating the agent’s policy. → Conditional Imitation Learning
2. Discriminative. The representation should be able to distinguish the agent’s policy with the policies of other agents. → Agent Identification using Triplet Loss

ALGORITHM

Algorithm 1 Learn Policy Embedding Function (fθ)

input \( (E_i)_{i=1}^{\lambda} \) = interaction episodes, \( \lambda \) = hyperparameter.
1: Initialize \( \theta \) and \( \phi \)
2: for \( i = 1, 2, \ldots, \lambda \) do
3: Sample a positive episode \( e_i \) \( \sim \) \( E_1 \)
4: Sample a reference episode \( r_e \) \( \sim \) \( E_0 \)
5: Compute \( \text{Emb} \) \( \approx \sum_{a_i} \log p_{\theta}(a_i|e_i) \)
6: for \( j = 1, 2, \ldots, \lambda \) do
7: if \( j \neq i \) then
8: Sample a negative episode \( n_e \) \( \sim \) \( E_j \)
9: Compute \( \text{IDemb} \approx \langle \theta(a_e|e), \phi(r_e) \rangle \)
10: Set \( \text{Loss} = \text{Emb} + \alpha \cdot \text{IDemb} \), \( \text{Update} \ \theta \) and \( \phi \) to minimize \( \text{Loss} \)
12: end if
13: end for
14: end for
output \( \theta \)

REFERENCES