Stigmergic Independent Reinforcement Learning for Multiagent Collaboration

Xing Xu, Rongpeng Li, Zhifeng Zhao, and Honggang Zhang

***Abstract*— With the rapid evolution of wireless mobile devices, there emerges an increased need to design effective collaboration mechanisms between intelligent agents to gradually approach the final collective objective by continuously learning from the environment based on their individual observations. In this regard, independent reinforcement learning (IRL) is often deployed in multiagent collaboration to alleviate the problem of a nonstationary learning environment. However, behavioral strategies of intelligent agents in IRL can be formulated only upon their local individual observations of the global environment, and appropriate communication mechanisms must be introduced to reduce their behavioral localities. In this article, we address the problem of communication between intelligent agents in IRL by jointly adopting mechanisms with two different scales. For the large scale, we introduce the stigmergy mechanism as an indirect communication bridge between independent learning agents, and carefully design a mathematical method to indicate the impact of digital pheromone. For the small scale, we propose a conflict-avoidance mechanism between adjacent agents by implementing an additionally embedded neural network to provide more opportunities for participants with higher action priorities. In addition, we present a federal training method to effectively optimize the neural network of each agent in a decentralized manner. Finally, we establish a simulation scenario in which a number of mobile agents in a certain area move automatically to form a specified target shape. Extensive simulations demonstrate the effectiveness of our proposed method.**

***Index Terms*— Artificial intelligence, collective intelligence, multiagent collaboration, reinforcement learning, stigmergy.**

I. Introduction

WITH the rapid development of mobile wireless communication and Internet of things technologies, many scenarios have arisen in which collaboration between intelligent agents is required, such as in the deployment of unmanned aerial vehicles (UAVs) [1]–[3], distributed control in the field of industry automation [4]–[6], and mobile crowdsensing and computing (MCSC) [7], [8]. In these scenarios, traditional centralized control methods are usually impractical due to limited computational resources and the demand for ultralow latency and ultrahigh reliability. As an alternative, multiagent collaboration technologies can be used in these scenarios to relieve the pressure on the centralized controller.

Guiding autonomous agents to act optimally through trial-and-error interaction with the corresponding environment is the primary goal in the field of artificial intelligence and is regarded as one of the most important objectives of reinforcement learning (RL) [9]–[11]. Recently, deep RL (DRL), which combines RL and deep neural networks, has improved the ability to obtain information from high-dimensional input, such as high-resolution images, and has demonstrated extraordinary learning ability across a wide array of tasks [12]. There are a number of advanced DRL algorithms that can direct a single agent to improve its behavioral policy by continuously learning from the environment [13], [14]. However, the extension of single-agent DRL to multiagent DRL is not straightforward, and many challenging problems remain to be solved in the application of multiagent RL (MARL) [15], [16]. In particular, in a completely distributed multiagent system (MAS), each agent is usually limited to partially observe the global environment, and its learning process following this local observation can thus be nonstationary, as other agents' behavioral policies may change temporally. In addition, the assignment of an individual reward is another challenging problem, as there is only one global reward for feedback in most cases. As an alternative, independent RL (IRL) has been proposed to alleviate the problem of a nonstationary learning environment, where each agent undergoes an independent learning process with only self-related sensations [17].

In IRL, most behavioral policies learned by intelligent agents are self-centered, aiming to maximize their own interests. Thus, the target of mutual communication is to integrate these individual behaviors effectively for the same task. Several studies have attempted to solve the problem of mutual communication in IRL. Foerster *et al.* [18] proposed differentiable interagent learning (DIAL), in which an additional communication action is added to the action set of each agent. In addition to the selection of the current action, a piece of interagent message is also generated and sent to other agents through a specified communication channel.

Manuscript received November 1, 2019; revised September 29, 2020 and December 29, 2020; accepted January 27, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1804804, in part by the National Natural Science Foundation of China under Grant 61731002 and Grant 62071425, in part by the Zhejiang Key Research and Development Plan under Grant 2019C01002 and Grant 2019C03131, in part by the Huawei Cooperation Project, the Project sponsored by Zhejiang Lab under Grant 2019LC0AB01, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY20F010016. *(Corresponding author: Rongpeng Li.)*

Xing Xu, Rongpeng Li, and Honggang Zhang are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: hsuxing@zju.edu.cn; lirongpeng@zju.edu.cn; honggangzhang@zju.edu.cn).

Zhifeng Zhao is with Zhejiang Lab, Hangzhou, China (e-mail: zhaozf@zhejianglab.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3056418.

Digital Object Identifier 10.1109/TNNLS.2021.3056418