A Study of AI Population Dynamics with Million-agent Reinforcement Learning

Extended Abstract

Yaodong Yang
University College London
London, U.K.
yaodong.yang@cs.ucl.ac.uk

Lantao Yu
Shanghai Jiaotong University
Shanghai, China
yulantao@apex.sjtu.edu.cn

Yiwei Bai
Shanghai Jiaotong University
Shanghai, China
bywbilly@gmail.com

Ying Wen
University College London
London, U.K.
ying.wen@cs.ucl.ac.uk

Weinan Zhang
Shanghai Jiaotong University
Shanghai, China
wnzhang@sjtu.edu.cn

Jun Wang
University College London
London, U.K.
jun.wang@cs.ucl.ac.uk

ABSTRACT

We[1] conduct an empirical study on discovering the ordered collective dynamics obtained by a population of intelligence agents, driven by million-agent reinforcement learning. Our intention is to put intelligent agents into a simulated natural context and verify if the principles developed in the real world could also be used in understanding an artificially-created intelligent population. To achieve this, we simulate a large-scale predator-prey world, where the laws of the world are designed by only the findings or logical equivalence that have been discovered in nature. We endow the agents with the intelligence based on deep reinforcement learning (DRL). In order to scale the population size up to millions agents, a large-scale DRL training platform with redesigned experience buffer is proposed. Our results show that the population dynamics of AI agents, driven only by each agent's individual self-interest, reveals an ordered pattern that is similar to the *Lotka-Volterra* model studied in population biology. We further discover the emergent behaviors of collective adaptations in studying how the agents' grouping behaviors will change with the environmental resources. Both of the two findings could be explained by the *self-organization* theory in nature.

KEYWORDS

Multi-agent reinforcement learning; population dynamics

ACM Reference Format:
Yaodong Yang, Lantao Yu, Yiwei Bai, Ying Wen, Weinan Zhang, and Jun Wang. 2018. A Study of AI Population Dynamics with Million-agent Reinforcement Learning. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

By employing the modeling power of deep learning [4], reinforcement learning (RL) has endowed AI agents with human-level intelligence on certain tasks [6, 11]. In the real world, the theory of *self-organization* suggests that the ordered global dynamics that live populations show, no matter how complex, are induced from repeated interactions between local individuals, without external supervisions or interventions. Ancient philosopher *Lucretius* once said: "*A designing intelligence is necessary to create orders in nature.*" [8], an interesting question for us is to understand what kinds of ordered macro dynamics, if any, that a community of artificially-created agents would possess when they are together put into the natural context.

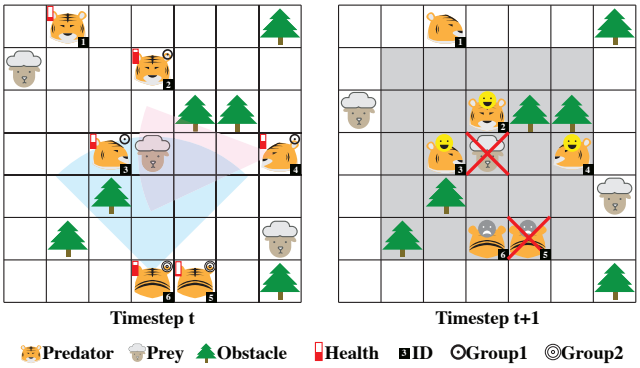

Figure 1: In the 2-D world, there exist preys, predators, and obstacles. Predators hunt the prey so as to survive from starvation. Each predator has its own health bar and limited eyesight view. Predators can form a group to hunt the prey so that the chance of capturing can increase, but this also means that the captured prey will be shared among all group members. When there are multiple group targeting the same prey, the largest group within capture radius will win. In this example, predators $\{2, 3, 4\}$ form a group and win the prey over the group $\{5, 6\}$. Predator 5 soon dies because of starvation.

2 DESIGN OF THE PREDATOR-PREY WORLD

We conduct an empirical study in a AI-powered predator-prey world. To avoid introducing any specific rules that could harm the generality of the observed results, we design the laws of the world

[1] First three authors contribute equally. Full paper can be found [15].

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

© 2018 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

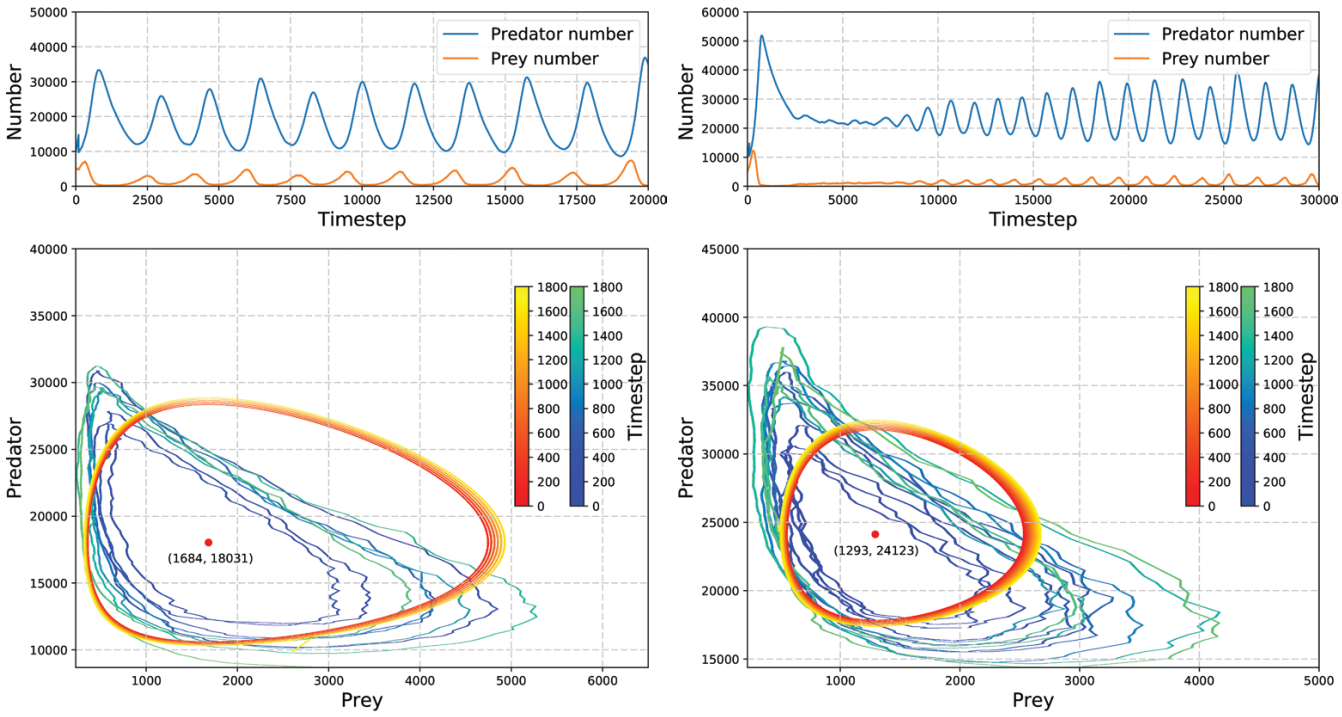

Figure 2: Population dynamics in both the time space (1^{st} row) and the phase space (2^{nd} row). The orange circles denote the theoretical solutions to the *Lotka-Volterra* equation, with the red spot as the equilibrium. The green-blue circles denote the simulation results. a): The simulated birth rate of preys is 0.006. b): The simulated birth rate of preys is 0.01.

(see Fig. 1) by only considering those real findings or logical equivalence that have been observed in the natural system; they include *Positive Feedback* [1, 14], *Negative Feedback* [2], *Individual Variation* [3, 9], *Response Threshold* [13], *Redundancy* [10], *Synchroni-sation* [7], *Selfishness* [12].

3 THE POPULATION DYNAMICS

We find that the AI population reveals an ordered pattern when measuring the population dynamics. As shown in Fig. 2, the population sizes of both predators and preys reach a dynamic equilibrium where both curves present a wax-and-wane shape, but with a 90° lag in the phase, *i.e.*, the crest of one is aligned with the trough of the other. The underlying logic of such ordered dynamics could be that when the predators' population grows because they learn to know how to hunt efficiently, as a consequence of more preys being captured, the preys' population shrinks, which will later cause the predators' population also shrinks due to the lack of food supply, and with the help of less predators, the population of preys will recover from the shrinkage and start to regrow. Such logic drives the 2-D contour of population sizes (see the green-blue traits in the 2^{nd} row in Fig. 2) into harmonic cycles, and the circle patterns become stable with the increasing level of intelligence agents acquire from the reinforcement learning. As it will be shown later in the ablation study, enabling the individual intelligence is the key to observe these ordered patterns in the population dynamics.

In fact, the population dynamics possessed by AI agents are consistent with the *Lotka-Volterra* (LV) model studied in biology (shown by the orange traits in Fig. 2). In population biology, the LV model [5] describes a *Hamiltonian* system with two-species interactions, *e.g.*, predators and preys. In the LV model, the population size of predators q and of preys p change over time based on the following pair of nonlinear differential equations:

$$\frac{1}{p}\frac{dp}{dt} = \alpha - \beta q, \quad \frac{1}{q}\frac{dq}{dt} = \delta p - \gamma. \tag{1}$$

The preys are assumed to have an affluent food resource and thus can reproduce exponentially with rate α, until meeting predation, which is proportional to the rate at which the predators and the prey meet, represented by βq. The predators have an exponential decay in the population due to natural death denoted by γ. Meanwhile, they can also boost the population by hunting the prey, represented by δp. The solution to the equations is a harmonic function (wax-and-wane shaped) with the population size of predators lagging that of preys by 90° in the phase. On the phase space plot, it shows as a series of periodical circle $V = -\delta\, p + \gamma\, \ln(p) - \beta\, q + \alpha\, \ln(q)$, with V dependent on initial conditions. In other words, which equilibrium cycle to reach depends on where the ecosystem starts. Similar patterns on the population dynamics might indicate that the orders from an AI population is induced from the same logic as the ecosystem that LV model describes. However, the key difference here is that, unlike the LV equations that model the observed macro dynamics directly, we start from a microcosmic point of view – the AI population is only driven by the self-interest (powered by RL) of individual agent, and then reaching the macroscopic principles.

REFERENCES

[1] Eric Bonabeau, Guy Theraulaz, Jean-Louls Deneubourg, Serge Aron, and Scott Camazine. 1997. Self-organization in social insects. *Trends in Ecology & Evolution* 12, 5 (1997), 188–193.

[2] Simon Goss, Serge Aron, Jean-Louis Deneubourg, and Jacques Marie Pasteels. 1989. Self-organized shortcuts in the Argentine ant. *Naturwissenschaften* 76, 12 (1989), 579–581.

[3] Robert L Jeanne. 1988. *Interindividual behavioral variability in social insects*. Westview Press.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.

[5] Alfred J Lotka. 1925. Elements of physical biology. (1925).

[6] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.

[7] Zoltán Néda, Erzsébet Ravasz, Tamás Vicsek, Yves Brechet, and Albert-Lázló Barabási. 2000. Physics of the rhythmic applause. *Physical Review E* 61, 6 (2000), 6987.

[8] Ada Palmer. 2014. *Reading Lucretius in the Renaissance*. Vol. 16. Harvard University Press.

[9] Tanya Pankiw and Robert E Page Jr. 2000. Response thresholds to sucrose predict foraging division of labor in honeybees. *Behavioral Ecology and Sociobiology* 47, 4 (2000), 265–267.

[10] Thomas D Seeley. 2009. *The wisdom of the hive: the social physiology of honey bee colonies*. Harvard University Press.

[11] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.

[12] Randy Thornhill. 1976. Sexual selection and paternal investment in insects. *The American Naturalist* 110, 971 (1976), 153–163.

[13] Anja Weidenmüller. 2004. The control of nest climate in bumblebee (Bombus terrestris) colonies: interindividual variability and self reinforcement in fanning response. *Behavioral Ecology* 15, 1 (2004), 120–128.

[14] Edward O Wilson et al. 1971. The insect societies. *The insect societies*. (1971).

[15] Yaodong Yang, Lantao Yu, Yiwei Bai, Jun Wang, Weinan Zhang, Ying Wen, and Yong Yu. 2017. An Empirical Study of AI Population Dynamics with Million-agent Reinforcement Learning. *CoRR* abs/1709.04511 (2017). arXiv:1709.04511 http://arxiv.org/abs/1709.04511