ADVANCES IN CYBER-PHYSICAL SYSTEMS
Vol. 5, No. 2, 2020

DECENTRALIZED CONTROL OF ADAPTIVE DATA COLLECTION PROCESSES BASED ON EQUILIBRIUM CONCEPT AND REINFORCEMENT LEARNING

Alexey Botchkaryov

Lviv Polytechnic National University, 12, S. Bandery Str., Lviv, 79013, Ukraine
Authors' e-mail*: alb@lp.edu.ua*

https://doi.org/10.23939/acps2020.02.050

Submitted on 01.11.2020

***Abstract*: The model of decentralized control of adaptive data collection processes has been developed based on the equilibrium concept, which is used to study the problem of coordinating joint collective actions from the point of view of finding an effective scheme for complementing the individual actions of data collection processes in the absence of a control center. The method of decentralized control of adaptive data collection processes in autonomous distributed systems based on the equilibrium concept and reinforcement learning by the method of normalized exponential function (softmax) has been developed. The method allows one to organize autonomous distributed exploration under the conditions of dynamic changes in the number of data collection processes and unreliable local information interaction between them. As a result of research and modeling of the developed method of decentralized control, it has been established that the use of the reinforcement learning (normalized exponential function method) provides more effective search for a solution compared to the method of adaptive random search (by an average of 28.3 %). Using the efficiency retention rate, an estimate was obtained for the dependence of the work of the developed decentralized control method on the change in the number of adaptive data collection processes and the change in the information interaction channels between adaptive data collection processes.**

***Index Terms:* decentralized control, adaptive data collection process, reinforcement learning**

I. INTRODUCTION

The relevant issue of using multi-agent systems technologies [1–7] and machine learning methods, in particular learning automata [8–11], reinforced learning [12–17] and multi-agent reinforced learning [18–23] has been considered in the paper in order to solve the problem of organizing adaptive data collection processes (DC-processes) in autonomous distributed systems [24–34]. In the context of this approach, an adaptive data collection process implements the behavior of the corresponding autonomous agent and data collection in the autonomous distributed system is generally considered as the behavior of the corresponding multi-agent system in the absence of a control center.

To solve optimization problems for the organization of adaptive data collection processes, in particular structural adaptation problems [30, 31], effective methods of coordination of joint actions of agents of the corresponding multi-agent system are required. One of the main concepts of designing coordination methods in multi-agent systems is the concept of equilibrium. Within this concept, the situation when agents reach some kind of equilibrium is interpreted as a solution to the problem of coordination. Coordination methods developed on the basis of this concept are based on different interpretations of the concept of "equilibrium", including the concept of equilibrium in game theory (Nash equilibrium and its variants) [1, 2, 5]. In this paper, the concept of equilibrium is interpreted in the physical aspect, in particular as an analogue of mechanical equilibrium.

The second section presents a model of decentralized management of data collection processes based on the concept of equilibrium. The purpose of using this model is to study the problem of coordination of joint collective actions of agents in terms of finding an effective scheme of their complementarity in the absence of a control center.

The third section presents the method of decentralized control of adaptive data collection processes based on the concept of equilibrium and reinforcement learning by the method of normalized exponential function (softmax). The proposed method allows to organize autonomous distributed explorations under conditions of dynamic changes in the number of data collection processes and unreliable communication channels between them.

The fourth section presents the results of the simulation of the developed method of decentralized control. The simulation results show that the use of the normalized exponential function method provides more efficient solution of the coordination problem than the adaptive random search methods. In addition, the simulation evaluates the dependence of the developed method on changes in the number of data collection processes and changes in the communication channels between them in terms of maintaining the efficiency of the corresponding multi-agent system.

II. THE MODEL OF DECENTRALIZED CONTROL OF ADAPTIVE DC-PROCESSES

To study the problem of coordinating joint collective actions from the point of view of finding an effective scheme for complementing the actions of individual DC-processes in

the absence of a control center, let us consider the following model of decentralized control of adaptive DC-processes based on the equilibrium concept [35, 36]:

$$M_D=\langle A, G(a,t), X, C, q(t)=f(F_q,\{x_t(a)\}_{N(t)})\rangle, \quad (1)$$

where $A=\{a\}_{N(t)}$ is the collective of DC-processes (agents) in the amount of N(t), which are located in some space X along the coordinates $\{x_t(a)\}_{N(t)}$, G(a,t) is a scheme of information interaction of DC-processes, *C* is a set of restrictions that are imposed on the coordinates of agents in the space *X*, q(t) is an indicator of the state of equilibrium and distance to it, such as

$$q(t) = \sum_{i=1}^{N(t)} (F_q - x_t(a_i)) , \quad (2)$$

where F_q is the parameter of the equilibrium condition, the value of which is unknown to DC-processes.

The collective of DC-processes is tasked with finding locations $\sigma(A)=\{x_t(a)\}_{N(t)}$ in the space *X* for which q(t)=0. The value of q(t) or the sign (+/ -) of its value is communicated to agents at each step. An agent can only change its own coordinate $x_t(a)$. Therefore, using the model M_D, the search for an effective scheme for complementing the individual actions of DC-processes in the course of solving a certain problem is modeled as a search for the equilibrium state implementing the equilibrium concept.

The search efficiency is defined as

$$w(T)=T_{min}/T, \quad (3)$$

where T is the time during which the collective found the equilibrium state, T_{min} is the minimum possible time to find the equilibrium state.

The model allows one to study the influence of the dynamics of changes in the number of DC-processes N(t) and the parameters of the information interaction scheme G(a,t) on the collective search speed for the equilibrium state.

III. THE METHOD OF DECENTRALIZED CONTROL OF ADAPTIVE DC-PROCESSES

Within the framework of the model M_D, a decentralized control method (DCM) for DC-processes has been developed. The convergence of DC-processes' actions to the equilibrium state is ensured by the use of the reinforcement learning method for stationary random environment (a binary version of a Multi-armed bandit), in particular the method of normalized exponential function (softmax).

The DCM implements the following principles (Fig. 1):

1) the maximum step size $\Delta x_{max}(a)$ is proportional to the speed and magnitude of the change of q(t): $\Delta x_{max}(a)=f_v(\{q(t)\}_{\Delta t})$,

2) the action value $V_t(a,d)$ is modified according to the corresponding action values $\{V_t(a_i,d)\}_{K(t)}$, which at step t were obtained from other agents under the current condition of the information interaction *G*(*a*,*t*).

According to the method of normalized exponential function, the next action $d\in\{-\Delta x_{max}(a),...,0,...+\Delta x_{max}(a)\}$ is chosen with the probability

$$p_t(d) = \frac{e^{V_t(A,d)/\mu}}{\sum_{D_a} e^{V_t(A,b)/\mu}}, \quad (4)$$

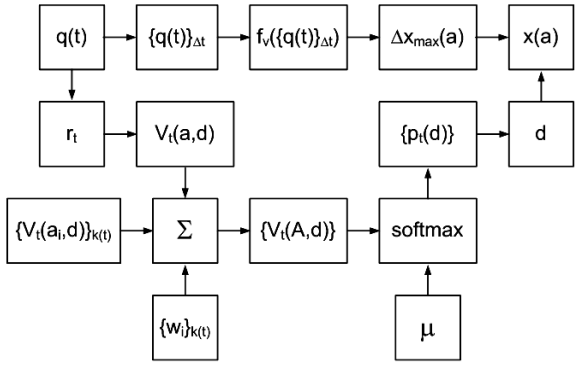

Fig. 1. The outline of decentralized control of adaptive DC-processes

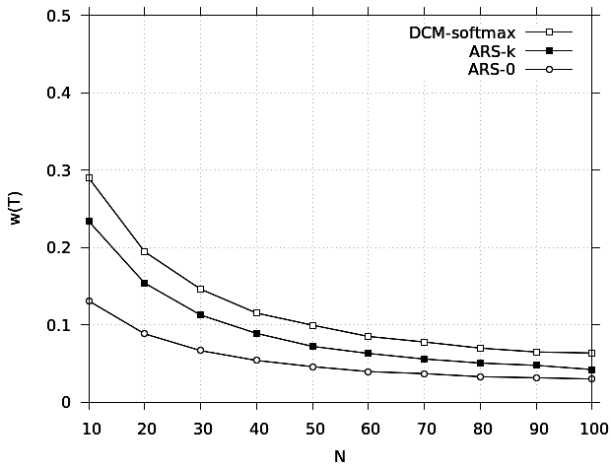

Fig. 2. Simulation results, X=1000, F_q=500, k=2, n=10000

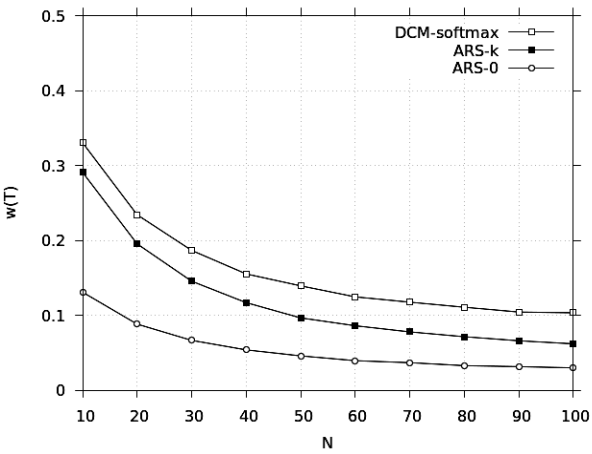

Fig. 3. Simulation results, X=1000, F_q=500, k=10, n=10000

Where μ is the scaling factor (μ> 0, μ = const), $V_t(a,d)$ is a modified action value for the action d:

$$V_t(A,d) = \sum_{i=1}^{k(t)} w_i V_t(a_i,d) , \quad (5)$$

where $\{w_i\}_{k(t)}$ are weight coefficients such as

$$\sum_{i=1}^{k(t)} w_i = 1. \quad (6)$$

The local values of $V_t(a,d)$ change according to the obtained rewards r(t)=q(t)-q(t-1):

$$V_{t+1}(a,d) = V_t(a,d) + \alpha(r_t - V_t(a,d)), \quad (7)$$

where $\alpha \in (0,1]$ is a learning step.

IV. SIMULATION RESULTS

Simulation of the proposed decentralized control method (DCM-softmax) (Fig. 2, Fig. 3) showed its advantage over the adaptive random search methods [37, 38]. Two variants of the adaptive random search are considered:

1) ARS-0 – adaptive random search without information exchange,

2) ARS-k – adaptive random search with information exchange, in which k is the number of neighbors, k < N).

On average, for various combinations of simulation parameters (X = {500, ..., 10000}, N (t) = {10, ..., 100}, k = {2, ..., 10}) DCM-softmax is better by 28.3% in the search efficiency w(T).

Using the efficiency retention rate K_T, an estimate was obtained of the dependence of DCM-softmax on changes in the number of DC-processes (in the form of a stationary Poisson failure flow of DC-processes with a flow rate λ_a at the beginning) (Table 1) and changes in the information interaction scheme G(a,t) (in the form of a stationary Poisson flow of failures of communication channels between DC-processes with a flow rate λ_g at the beginning) (Table 2).

The efficiency retention rates $K_T(\lambda_a)$ and $K_T(\lambda_g)$ were determined by simulating the operation of DCM-softmax. Here, $K_T(\lambda_a)=w(T)/w(T_n)$, where w(T) is the equilibrium search efficiency for failures at the first T_1 steps, and $w(T_n)$ is the nominal equilibrium search efficiency without failures at the first T_1 steps. According to the results obtained (Table 1, Table 2), DCM-softmax allows one to organize autonomous distributed exploration under conditions of dynamic changes in the number of DC-processes and unreliable local communications between them.

Table 1

Efficiency retention rate KT(λ_a), X=1000, Fq=500, N=50, k=2, T1=100, T2=1000, n=10000

	λ_a								
	0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045
ARS-0	0.765	0.485	0.334	0.245	0.197	0.156	0.181	0.135	0.115
ARS-k	0.798	0.509	0.367	0.276	0.221	0.185	0.165	0.140	0.128
DCM-softmax	0.841	0.590	0.459	0.385	0.331	0.301	0.281	0.261	0.247

Table 2

Efficiency retention rate KT(λ_g), X=1000, Fq=500, N=50, k=(N-1), T1=100, T2=1000, n=10000

	λ_g								
	0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045
ARS-k	0.796	0.510	0.362	0.279	0.228	0.186	0.164	0.140	0.126
DCM-softmax	0.867	0.612	0.480	0.404	0.355	0.319	0.303	0.283	0.272

V. CONCLUSION

The model of decentralized control of adaptive data collection processes has been developed based on the equilibrium concept, which is used to study the problem of coordinating joint collective actions from the point of view of finding an effective scheme for complementing the individual actions of data collection processes in the absence of a control center.

The method of decentralized control of adaptive data collection processes in autonomous distributed systems based on the equilibrium concept and reinforcement learning by the method of normalized exponential function (softmax) has been developed. The method allows one to organize autonomous distributed exploration under the conditions of dynamic changes in the number of data collection processes and unreliable local information interaction between them.

As a result of research and modeling of the developed method of decentralized control, it has been established that the use of the reinforcement learning (normalized exponential function method) provides more effective search for a solution compared to the method of adaptive

random search (by an average of 28.3 %). Using the efficiency retention rate, an estimate was obtained for the dependence of the work of the developed decentralized control method on the change in the number of data collection processes and the change in the information interaction channels between data collection processes.

REFERENCES

[1] *Multiagent Systems*, by Gerhard Weiss (Editor), 2nd edition, The MIT Press, 2013. 920 p.

[2] Michael Wooldridge, *An Introduction to MultiAgent Systems*, 2nd edition, Wiley, 2009. 484 p.

[3] Yoav Shoham, Kevin Leyton-Brown, *Multiagent Systems: Algorithmic,Game-Theoretic, and Logical Foundations*, Cambridge University Press, 2008. 504 p.

[4] *Multi-Agent Systems: Simulation and Applications*, by Adelinde M. Uhrmacher (Editor), Danny Weyns (Editor), CRC Press, 2009. 566 p.

[5] Stuart Russell, Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th edition, Pearson, 2020. 1136 p.

[6] Jacques Ferber, Multi-agent systems: An introduction to distributed artificial intelligence, Addison-Wesley Professional, 1999. 528 p.

[7] Serge Kernbach, Structural Self-Organization in Multi-Agents and Multi-Robotic Systems, Logos Verlag, 2008. 250 p.

[8] Tsetlin M. L. *Finite automata and models of simple forms of behaviour*, Russian Mathematical Surveys, 18(4):112, 1963, 3–28 p.

[9] Tsetlin, M. L. (1973). *Automaton Theory and Modeling of Biological Systems*. Academic Press, New York. 288 p.

[10] Varshavsky V. I. *Collective behavior of automata*, Moscow, Nauka, 1973. 408 p. (in Russian)

[11] Narendra, K. and Thathachar, M. A. L., *Learning Automata: An Introduction*, 2nd ed., Dover Publications, 2013. 496 p.

[12] Richard S. Sutton, Andrew G. Barto, *Reinforcement Learning: An Introduction*, 2nd Ed., A Bradford Book, 2018. 532 p.

[13] *Reinforcement Learning: Theory and Applications*, ed. by C. Weber, M. Elshaw, N. M. Mayer, I-Tech Education and Publishing, Vienna, 2008. 424 p.

[14] Kaelbling L. P. , Michael L. Littman, and Andrew W. Moore. *Reinforcement learning: A survey*. Journal of AI Research, No. 4, 1996. pp. 237–285

[15] Csaba Szepesvari, *Algorithms for Reinforcement Learning*, Morgan and Claypool Publishers, 2010. 104 p.

[16] *Reinforcement Learning: State-of-the-Art*, ed. by Marco Wiering, Martijn van Otterlo, Springer, 2012. 672 p.

[17] Dimitri Bertsekas, *Reinforcement Learning and Optimal Control*, Athena Scientific, 2019. 388 p.

[18] John Wilbur Sheppard, *Multi-Agent Reinforcement Learning in Markov Games*, D.Ph. Thesis, Johns Hopkins University, Baltimore, Maryland, 1997.

[19] Thomas Gabel, Learning in Cooperative Multi-Agent Systems: Distributed Reinforcement Learning Algorithms and their Application to Scheduling Problems, Südwestdeutscher Verlag für Hochschulschriften, 2009. 192 p.

[20] Buşoniu L., Babuška R., and B. De Schutter, "*Multi-agent reinforcement learning: An overview*", Chapter 7 inInnovations in Multi-Agent Systems and Applications – 1 (D. Srinivasan and L. C. Jain, eds.), vol. 310 ofStudies in Computational Intelligence,Berlin, Germany: Springer, 2010. pp. 183–221

[21] Howard M. Schwartz, Multi-Agent Machine Learning: A Reinforcement Approach, Wiley, 2014. 256 p.

[22] Felipe Leno Da Silva, Anna Helena Reali Costa, *A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems*, Journal of Artificial Intelligence Research (JAIR), Vol. 64, 2019. pp. 645–703

[23] Arup Kumar Sadhu, Amit Konar, Multi-Agent Coordination: A Reinforcement Learning Approach, Wiley, 2020. 320 p.

[24] Botchkaryov A., Golembo V., *Distributed contact sensing system based on autonomous intelligent agents*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, 2001, No. 437. pp. 14–20 (in Ukrainian)

[25] Botchkaryov A., Golembo V., *Models of collective behavior of measuring agents*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, No. 463, 2002. pp. 19–27 (in Ukrainian)

[26] Melnyk A., Golembo V., Botchkaryov A., *The new principles of designing configurable smart sensor networks based on intelligent agents*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, No. 492, 2003. pp. 100–107 (in Ukrainian)

[27] Botchkaryov A., *Collective behavior of mobile intelligent agents solving the autonomous distributed exploration task*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, No. 546, 2005. pp. 12–17 (in Ukrainian)

[28] Melnyk A., Golembo V., Botchkaryov A., *Multiagent approach to the distributed autonomous explorations*, Proceedings of NASA/ESA Conference on Adaptive Hardware and Systems (AHS-2007), Edinburgh, UK, 2007. Pp. 606–610

[29] Golembo V., Botchkaryov A. Applying the concepts of multi-agent approach to the distributed autonomous explorations // International Book Series "Information Science and computing", No. 13" "Intelligent Information and Engineering Systems", supplement to International Journal "Information Technologies and Knowledge", Vol. 3, 2009. pp. 136–142

[30] Botchkaryov A., *Structural adaptation of the autonomous distributed sensing and computing systems*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, No. 688, 2010. pp. 16–22 (in Ukrainian)

[31] Botchkaryov A., *The problem of organizing adaptive sensing and computing processes in autonomous distributed systems*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, No. 745, 2012. pp. 20–26 (in Ukrainian)

[32] Botchkaryov A., Golembo V., *Applying intelligent technologies of data collection to autonomous cyber-physical systems*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, No. 830, 2015. pp. 7–11 (in Ukrainian)

[33] Botchkaryov A., *Organization of adaptive processes of information collection in mobile cyberphysical systems*, Proceedings of the Second Scientific Seminar "Cyberphysical Systems: Achievements and Challenges", Lviv Polytechnic National University, Lviv, June 21–22, 2016. pp. 62–67 (in Ukrainian)

[34] *Cyber-physical systems: data collection technologies*, by A. Botchkaryov, V. Golembo, Y. Paramud, V. Yatsyuk

(authors), A. Melnyk (editor), Lviv: Magnolia 2006, 2019. 176 p. (in Ukrainian)

[35] Opoitsev V. I. *Equilibrium and stability in collective behavior models*, Moscow, Nauka, 1974. 245 p. (in Russian)

[36] Botchkaryov A., *Solving the problem of mechanical balancing by collective of mobile agents*, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, No. 463, 2002. pp. 14–18 (in Ukrainian)

[37] Rastrigin L. A., Ripa K. K., Tarasenko G. S., *Random search adaptation*, Riga, Zinatne, 1978. 239 p. (in Russian)

[38] Rastrigin L. A. *Adaptation of complex systems*, Riga, Zinatne, 1981. 375 p. (in Russian)

Alexey Botchkaryov was born in 1975 in Lviv, Ukraine. He received the B.S. and the M.S. degrees in computer engineering at Lviv Polytechnic National University in 1998 and Ph.D. degree in computer systems and components at Lviv Polytechnic National University in 2019. He has been doing scientific and research work since 1994. He is an author of a book (with coauthors) and more than 40 articles. Currently, he is an Associate professor at the Computer Engineering Department of Lviv Polytechnic National University. His research interests include self-organization in complex systems, structural adaptation, intelligent information-gathering agents and multi-agent systems.