MODELING OF COOPERATIVE BEHAVIOR IN MULTIAGENT SYSTEMS

UDC 004.94, 004.852

DOI: https://doi.org/10.35546/2313-0687.2018.24.57-66

ASIEIEV Vladislav
Master of Computer Science, Petro Mohyla Black Sea National University,
Nikolaev, Ukraine, aseev.vlad@gmail.com , ORCID ID: 0000-0002-3081-1651

KULAKOVSKA Inessa
Ph.D. in Physics and Mathematics, associate Professor of the Intellectual Information Systems Department of the Petro Mohyla Black Sea National University, Nikolaev, Ukraine, inessa.kulakovska@chmnu.edu.ua, ORCID ID: 0000-0002-8432-1850

Abstract. Cooperative behavior is understood as a community of agents who decide to cooperate to reduce the average weighted fines to solve a task or achieve a certain goal, in our case, synchronize the lightning.

The problem of forming cooperative behavior is intensively investigated in modern scientific literature on the use of multi-agent systems, for example, for distance learning, management of organizational systems, the construction of various virtual organizations and communities, management of distributed computing, management of public institutions and socio-economic processes, and others.

In this paper an actual theme of optimal policies in games with local interaction is considered, the stimulating training of multiagent systems in gaming is considered.

The purpose of this work is to consider the method of constructing a system with local interaction of agents based on the task of "synchronization" with the help of the Markov model of stochastic game.

The research method is a computer program for modeling a task using the Q-method of training. Formation of coalitions in multiagent systems is formulated as a competitive or cooperative task of assigning an object to one of the clusters. The problem of solving such problems is the theory of games, and in the conditions of uncertainty the theory of stochastic games. In this regard, from the scientific and practical point of view, the use of stochastic game methods for the formation of coalitions under conditions of incomplete information is relevant. The decision of the stochastic game is to find policies for agents that maximize their winnings to provide a certain collective balance of interests for all players. To find optimal players' policies under uncertainty, we will use the method of stimulating learning.

The result is a developed game model that provides a dynamic MAS self-organization, which manifests itself in the rhythmic change of pure agent policies that mimic the light effects of the colony of fireflies. A characteristic feature of the considered game self-organization is locally defined information about the policies of the behavior of neighboring agents, which as a result of learning leads to global coordination. policy of all agents.

The scientific novelty of the obtained results consists in the development of a gaming model, the effectiveness of the game self-organization of the MAS policies for solving the decision-making problem in systems with co-operative behavior of agents under uncertainty has been determined. The repetition of the values of the characteristics of the game in various experiments with unique sequences of random variables confirms the reliability of the results.

Results can be used in practice to model the dynamics of social processes, management of social Internet services in the Internet and others.

Key words: *multiagent system, stochastic game, adaptive gaming method, Q-method.*

Introduction. Core functionality of most modern information systems (IS) is based on strictly programmed algorithms. Due to unforeseen environmental influences on such systems, stability of operating modes may be affected, which may lead to various types of emergency situations. In order to prevent critical states, distributed IS software should consist of interacting autonomous modules, be intelligent, able to independently monitor changes in the state of the environment and make right in-time decisions. The IS agent is an autonomous software module with elements of artificial intelligence that is able to make decisions on its own, interact with the environment, other agents and user while solving the task. Interaction of IS agents is carried out within the computer network. A population of computer network agents that solve a common task is called the multi-agent system (MAS).

The work of MAS [2, 3], as a rule, is carried out in a state of uncertainty about the state of the environment, decision-making and the actions of other agents[4]. In connection with this policy of conduct, agents must be adaptive due to the ability of agents to self-study. Among methods of teaching in conditions of uncertainty methods based on incentives [5, 6] showed the best results since they do not require a mathematical model of the environment and provide a possible decision directly in the learning process. The basis of encouraging training include mechanisms of reflex behavior of living organisms with the nervous system. An effective method of encouraging learning is the marking of Q-learning [7], which performs the numerical identification of the characteristic system of a dynamic system in the state of action space. As the characteristic feature used, uses the functional expected remuneration.

Compared to single-agent systems, the structure, operation and research of multi-agent Q-learning methods are significantly complicated. Due to the collective interaction of agents, the stationary environment is translated into non-stationary. States changes of the environment and the value of the wins of each agent depends on the actions of other agents. In the general case, the MAS agent can not achieve the maximum winnings which is equal to its winnings in a single-agent system. The optimal benefits of agents must be balanced and meet the criteria of benefit, equity, and equilibrium. So, instead of scalar maximizing the benefits of a single-agent system, the criteria for vector maximization of MAS wins are introduced, for example Nash equilibrium, Pareto efficiency or others.

Under conditions of use of Q-learning methods MAS there is an iterative construction of the system of characteristic Q-functions in the space of state-action, and the growth of the elements of these functions is carried out in the direction of achieving their collective equilibrium.

The purpose of the work is to construct an iterative method of incentive training for solving a stochastic MAS game under uncertainty. In order to achieve the goal, it is necessary to develop a multi-agent stochastic game model, determine the criteria for collective equilibrium, the method and algorithm for solving the game problem.

In order to create an MAS it is necessary to perform preliminary research on the basis of mathematical models that enable to study the dynamics of the system under uncertainty conditions, to construct the policies of the behavior of agents that provide optimal parameters for the functioning of the system. Taking into account the peculiarities of the subject area, namely, multi-agency, uncertainty of the decision-making environment, antagonism or competition of goals, communicative, coordination of actions, adaptive strategies of agents, for the construction of MAS models we use the mathematical apparatus of the theory of stochastic games [6, 7].

The solution to a stochastic game is to seek out strategies of agents that maximize their winnings so as to provide a certain collective balance of interests for all players. To search optimal strategies of players in conditions of uncertainty we will use the method of encouraging learning.

The object of research is the processes of self-organization of multiagent systems under uncertainty, aimed at achieving the coordinated work of the constituent elements of multiagent systems due to the properties of self-study and adaptation, which results in the fact that the distributed system of elements functions as a coherent harmonious organism.

The subject of this work is a stochastic game model of the self-organization of multi-agent systems, which provides a balance of the values of payment functions of the team of players and is manifested in the achievement of coordinated strategies of agents.

The purpose of the work is to build a gaming model of self-organization of multi-agent systems to support decision-making under uncertainty. This goal is achieved by solving next problems: developing a mathematical model of multi-agent stochastic game; development of self-learning method and algorithm for solving stochastic game; development of software for simulation of stochastic game; analysis of the results and recommendations for their practical application.

To achieve the goal, it is necessary to analyze and solve the following problems: collective development and decision making; ensuring coordination and cooperation in the IAU; exploration of the states of the MAS functioning environment; definition of optimal structural organization of MAS; development of methods and means of multi-agent training; development of methods, languages and means of communication agents. The research method is a computer program for modeling a task. The research method is a computer program for modeling a task.

The ideas of the article P.O. Kravets [1] "The game model of self-organization of multi-agent systems" were used, which considers the main properties of the MAS and the connection of the task of "simulating the synchronized rhythmic glow of colony of fireflies" from MAS. The purpose of modeling is to determine the conditions and mechanisms of local coordination of agents, for the self-organization of MAS. To do this, we need to solve the following tasks: build a model of the game, develop a method and algorithm for solving, and execute computer simula tion software to identify the coordination and self-organization of the MAS.

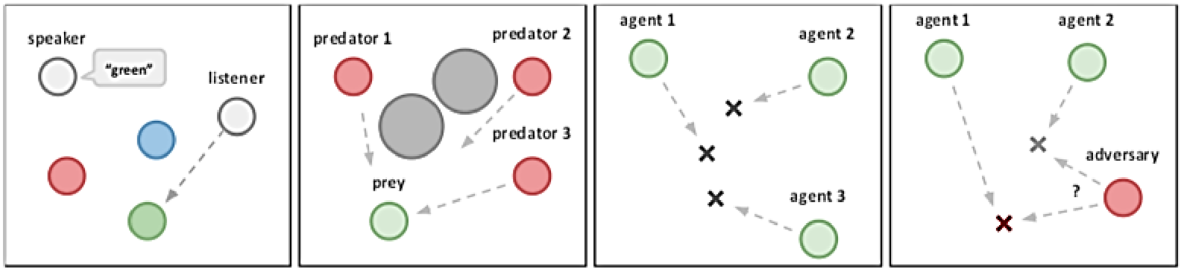


Figure 1. Illustrations of the experimental environment and some of the considered tasks, including: a) cooperative communication, b) predator-victim, c) cooperative navigation, d) physical deceit.

Cooperative behavior. To carry out experiments, we use a well-grounded communication environment consisting of N agents and L milestones that inhabit a two-dimensional world with continuous space and discrete time. Agents can do physical actions in the environment and communication actions that are passed to other agents. We do not assume that all agents have the same space of action and observation, or act on the same policy π. We also consider games that are cooperative (all agents must maximize joint return) and competitive (agents have opposite goals). Some environments require explicit communication between agents to achieve the best rewards, while in other environments, agents can only perform physical activity. Information about each environment is listed below.

For the construction of MAS it is necessary to perform preliminary research on the basis of mathematical models that enable to study the dynamics of the system under uncertainty conditions, to construct policy of agents behavior, which provide optimal technical and economic parameters of the system's operation. Taking into account the peculiarities of the subject area, namely, multiagency, uncertainty of the decision-making environment, antagonism or competition of goals, communicative, coordination of actions, adaptive strategies of agents, for constructing MAS models we will use the mathematical apparatus of the theory of stochastic games. The solution to a stochastic game is to seek out strategies of agents that maximize their winnings so as to provide a certain collective balance of interests for all players. To search optimal strategies of players in conditions of uncertainty we will use the method of encouraging learning.

Mathematical model of stochastic game. Average loses of agents

$$\theta_n^i\left(\{\tau, u_n^{Di}\}\right) = \frac{1}{n}\sum_t^n \xi_t^i\,, \forall i \in D \qquad (1)$$

characterize the quality of the game at time n.

The purpose of each agent is to minimize their own average loses function:

$$\lim_{n\to\infty} \theta_n^i \to \min_{\{u_n^i\}} \forall i \in D. \qquad (2)$$

The task of the stochastic game lies in the fact that agents are based on observing current losses $\{\xi_t^i\}$, must learn to choose a clean policy $\{u_n^i\}$so that with the course of time n = 1,2, ... ensure fulfillment of the system of criteria (2).

To solve the problem (2), it is necessary to determine the method of forming sequences of clean policies $\{u_n^i\}$ in time that ensure the fulfillment of condition (4) due to the asymptotic adequacy of the functions of average payoffs (3).

The value of the function θ_n^i for the average player losses is approaching the values of the V^i functions of the average loses of the matrix game:

$$\forall i \in D\ \lim_{n\to\infty} n^k\, M\left\{\left[\theta_n^i - V^i(p^{Di^*}\right]^2\right\} < \infty. \qquad (3)$$

The condition of complementary non-rigidity, weighed by the elements of the vectors of mixed strategies, describes the game's solutions in both mixed and pure strategies:

$$diag(p_i)\left(\nabla_{pi}V_i(p) - e^{N_i}V_i(p)\right) = 0, \forall i \in D, \qquad (4)$$

where $diag(p_i)$ is a square diagonal matrix of order N_i, constructed of elements of the vector p_i.

The current losses of agents will be determined as a penalty for violating the spatial and temporal coordination of strategies:

$$\xi_n^i = \lambda \sum_{s\epsilon D}\frac{|u_n^i - u_n^s|}{L_i} + (1-\lambda)\left|\overline{u_n^i} - u_{n-1}^i\right| + \mu_n, \qquad (5)$$

where $\xi_n^i \epsilon R^1$; $\lambda\epsilon[0;1]$ - weight coefficient; Di is the set of neighboring agents corresponding to the Figure 2 bonds; L_i=|Di| - number of neighboring agents; u_n^i- clean policy with binary meaning; $\overline{u_n^i}$- inversed value of pure strategy; $\mu_n \sim Normal(0,d)$ - white Gaussian noise, normally distributed random value with zero expectation and dispersion d> 0.

The first component of expression (5) defines a penalty for violating the spatial (reciprocal) coordination of player strategies within a subset of D_i; the second component is a fine for violation of time coordination in two consecutive moments of time, and the third component defines the action of random noise in the form of white noise.

Taking into account the magnitude of the period $\tau = N_i$ of the dynamic self-organization of the MAS, we construct the sequences of pure strategies with the desired properties on the basis of the matrices of the probabilities of transitions between the pure policies $\{u_n^i\}$ of the agents:

$$p_n^i = \left[\begin{pmatrix} p_n^i(1.1) & \cdots & p_n^i(1,N_i) \\ \vdots & \ddots & \vdots \\ p_n^i(N_i,1) & \cdots & p_n^i(1N_i,N_i) \end{pmatrix}\right], \forall i \in D. \qquad (6)$$

The matrix lines p_n^i are the mixed strategies of the i player if he chose a clean policy $u_n^i \in U^i$. Lines elements $p_n^i(j.k)$ are conditional probabilities of choosing clean strategies depending on the current version of action u_n^i and the resulting loss ξ_n^i. Let's assume that the selected clean strategies match the agent's states. Then p_n^i (6) is a matrix of probabilities of changing agent states.

The game begins with uninformed mixed strategies $p_n^i(j.k) = \frac{1}{N_i}$ j, k $= 1 \ldots N_i$. To adapt the formation of the distribution of random strategies that minimizes the average losers (1) of all players, the probability of choosing strategies with less losers should increase over time *n=1,2,...*

Taking into account (4), we obtain the following recursive method of changing the vectors of mixed strategies:

$$p_{n+1}^i = \pi_{\varepsilon_{n+1}}\{p_n^i(u_n) - \gamma_n \xi_n^i [e(u_n^i) - p_n^i(u_n)]\}, \forall i \in D, \quad (7)$$

where $p_n^i(u_n)$ - mixed policy of i-player in a state $u_n \in U^i$; $\pi_{\varepsilon_{n+1}}$ - design operator for a single ε -simplex $S_\varepsilon^{N_i} \subseteq S^{N_i}$(11), which is a subset of the unit simplex S^{N_i}; $\gamma_n > 0$ - a monotonically decreasing sequence of positive quantities that regulates the size of the step of the method; ερ> 0 is a monotonically decreasing sequence of positive quantities that governs the expansion rate of the ε -simplex.

The study of the convergence of the method (11) will be performed in the class of monotone sequences $\{\gamma_n\}$ and $\{\varepsilon_n\}$.

$$\gamma_n = \gamma(n+\alpha)^{-\alpha};\ \alpha > 0;\ \varepsilon_n = \varepsilon(n+\beta)^{-\beta};\ \beta > 0. \quad (8)$$

The convergence of the method (11) is observed:

1) with probability 1, if α ∈ (0.5; 1); β> 0;

2) in the rms, if α ∈ (0,1); β> 0

Method (7) provides an adaptive selection of agents for clean strategies through the dynamic rebuilding of mixed strategies based on the processing of current losses.

Based on the current distribution of probabilities $p_n^i(u_n{}^i)$, the agent carries out a random selection of a clean policy $\forall i \in D$.

$$u_n^i = \{\frac{u^i(l)}{l} = arg\min_l \sum_{k=1}^{l} p_n^i(j,k) > \omega(j,l = 1..N_i)\}, \quad (9)$$

where w∈ [0,1] is a random variable with uniform distribution.

So, if at time n the agent is in a state u_n^i then based on a mixed strategies $p_n^i(u_n)$ it choose clean policy $u_n^{i\prime}$ according to (3), for which, by the time *t +1*, the current loss is received ξ_n^i which uses to calculate the mixed policy $p_{n+1}^i(u_n)$ according to (7), after which it becomes a new state $u_{n+1}^i = u_n^{i\prime}$.

Evaluating the effectiveness of the game self-organizing MAS will perform on the following indicators:

1) Average loss function or game price:

$$\theta_n = \frac{1}{L}\sum_{i=1}^{L}\theta_n^i,$$

where *L*=|*D*|- number of players;

2) spatial coordination coefficient of player strategies:

$$K_n = \frac{1}{nL}\sum_{t=1}^{n}\sum_{i=1}^{L}\chi(\sum_{s\epsilon D_i}|u_t^i - u_t^s|=0),$$

where $\chi \epsilon \{0,1\}$ - indicator function of the event;

Example modeling. For an example, consider a stochastic game model of self-organizing fireflies from the Lampyridae family that lead a nightlife in tropical regions of the world. Males of these insects for attracting females launch a mechanism of luminescent radiation of their abdomen. Self-organization is manifested in the emergence of the phenomenon of rhythmic synchronized glow throughout the colony of males.

Modeling the behavior of fireflies will be done using a stochastic game of agents, each of which can be in one of two states $u_n^i \epsilon \{0,1\}$, where 0 indicates absence, and 1 - presence of glow.

We solve the stochastic game of two agents with two clean strategies in a two-state environment. The matrices of the average winnings of such a game are given in the table.

Table 1.

Matrices of the winnings agents

state	strategie	agents 1		agents 2	
s_1	-	$\pi_2(s_1, u_2[0])$	$\pi_2(s_1, u_2[1])$	$\pi_2(s_1, u_2[0])$	$\pi_2(s_1, u_2[1])$
	$\pi_1(s_1, u_1[0])$	0.5	0.2	0.4	0.1
	$\pi_1(s_1, u_1[1])$	0.6	0.7	0.1	0.9
s_2	-	$\pi_2(s_2, u_2[0])$	$\pi_2(s_2, u_2[1])$	$\pi_2(s_2, u_2[0])$	$\pi_2(s_2, u_2[1])$
	$\pi_1(s_2, u_1[0])$	0.9	0.2	0.4	0.6
	$\pi_1(s_2, u_1[1])$	0.2	0.9	0.6	0.8

Each agent can observe the states of neighboring agents and change their own state so that the actions can be as close as possible to their neighbors. The structure of the relationships between agents is depicted in Figure 2.

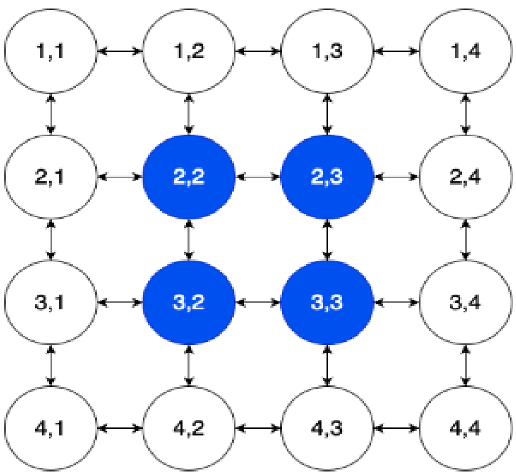


Figure 2. Model of stochastic game "Firefly"

The regular structure of the game is given by the number of agents $L = m*m$, $m \geq 2$, subsets of the neighboring agents D_i and the number of clean strategies $N_i = N = 2$, $i = 1..L$.

The dynamics of the process of self-organization consists of spatial and temporal coordination of the strategies of agents. Spatial coordination is to observe the ratio of agent strategies in locally defined areas of D_i, N_i as shown in Figure 2. Time coordination is determined by observing the ratio of agent strategies over time $\tau = 2$.

In game terminology, spatial coordination will consist in choosing the same values of players' pure strategies at fixed moments of time (agents try to repeat each other's actions), and time coordination - in changing binary strategies to opposite values at two consecutive moments of time. The result of the self-organizing of agents is the inverse change of the matrices of binary clean strategies $[0]_{m*m}$-$[1]_{m*m}$-$[0]_{m*m}$-$[1]_{m*m}$ - in time that simulates the rhythmic glow of the colony of fireflies:

$$p^{i^*}(u_n^i) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Such a change is provided by the matrix of trained mixed strategies agents with identical initial states.

Algorithm for solving a stochastic game

1. Set the initial values of the parameters:

- $n = 0$ - the initial time;
- $L = |D|$ - number of players;
- N - the number of players' pure strategies;
- $U_i = (u^i[1], u^i[2], ..., u^i[N])$, $i = 1..L$ - vectors of pure player strategies;
- $p_0^i = (1 / N, ..., 1 / N)$, $i = 1..L$ - initial mixed player strategies;
- $\gamma > 0$ is the parameter of the learning step;
- $\alpha \in (0,1)$ - the order of the learning step;
- ε is the parameter of the ε-simplex;

– β> 0 is the order of the expansion rate of the ε - symplex;

– d> 0 - dispersion of noise;

– max n - maximum number of steps of the method.

2. Select action options $u^i_n \in U_i$, i = 1..L according to (9).

3. Get the value of current losses ξ^i_n, i = 1..L according to (5). Current values of Gaussian white noise are calculated by the formula, where ω ∈ [0,1] is a real random number with uniform distribution law.

4. Calculate the values γ_n, ε_n of the parameters γn, εn according to (8).

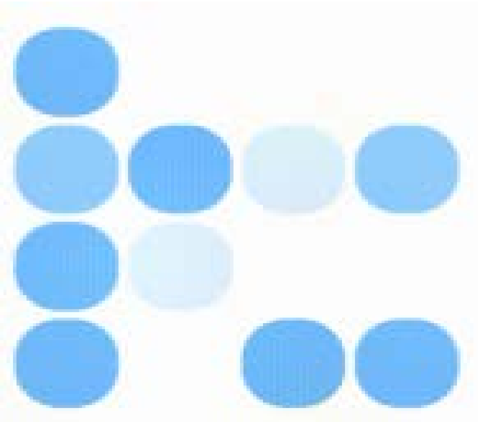
5. Calculate the elements of the vector of mixed strategies pin, i = 1..L according to (7).

6. Calculate the quality characteristics of the decision making θn (10), Kn (11).

7. Specify the next time point n: = n +1.

8. If n <nmax, then go to step 2, otherwise - end.

Thanks to the local coordination of the strategy of agents, this solution ensures the self-organization of the MAS "firefly". Each player watches the actions of neighbors and gets their own losses through non-matching, which forces him to dynamically choose strategies with less fines. The dynamic selection of strategies transforms locally coordinated actions of players into global coordination of the game, when the team of players behaves as a holistic organism.


a)

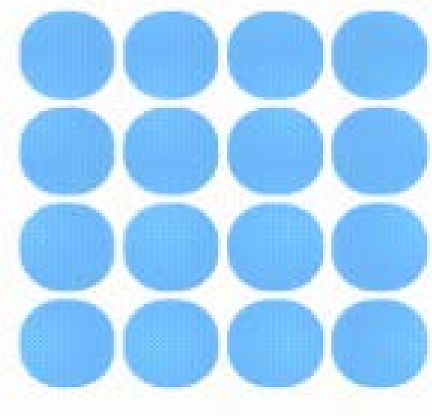

b)

Figure 3. a) in the learning process b) self-organized agents

Conclusions. The developed game model provides a dynamic self-organization of the MAS, which manifests itself in the rhythmic change of the pure agent strategies that simulates the light effects of the colony of fireflies. A characteristic feature of the considered game self-organization is the locally determined collection of information about the strategies of the behavior of neighboring agents, which as a result of learning leads to global coordination. strategies of all agents.

The generation of sequences of clean strategies with the required properties is provided by random distribution, built on dynamic mixed player strategies. The calculation of mixed strategies is carried out using the adaptive recurrent method obtained on the basis of the stochastic approximation of the complementary non-rigidity condition, which describes the collective solutions of the game satisfying the equilibrium condition according to Nash.

The effectiveness of the game self-organization of MAS strategies was studied with the help of the functions of average losses, co-ordinates and norms of deviation of dynamic mixed strategies from optimal values. The decline of the function of average losers and the function of rejection of mixed strategies, the growth of co-ordination factors indicate the convergence of the game method and the entry of MAS into self-organization. Repeating the values of the game's characteristics in various experiments with unique sequences of random variables confirms the reliability of the results.

REFERENCES:

1. Kravec P. O. (2015). *Igrova model samoorganizaciyi multiagentnih system.* Lviv: Visnik Nacionalnogo universitetu "Lvivska politehnika". Seriya: Informacijni sistemi ta merezhi : zbirnik naukovih prac, 829, 161–176. http://ena.lp.edu.ua:8080/handle/ntb/31452 [in Ukrainian].
2. Wooldridge M. (2002). *An introduction to multiagent systems*. John Wiley&Sons. [in English].
3. Weiss G. (1996). *Multiagent Systems. A modern approach to distributed artificial intelligence.* Berlin: Springer Verlag.
4. Musiyenko M., Zhuravska I., Kulakovska I., Kulakovska A. (2016). Simulation the behavior of robot sub swarm in spatial corridors. *2016 IEEE 36th International Conference on ELNANO. 382-387*. DOI: 10.1109/ELNANO.2016.7493090 http://ieeexplore.ieee.org/document/7493090/ [in English].
5. Kaelbling L., Michael L. L., Moore A. W. (1996). *Reinforcement learning: a survey.* Journal of Artificial Intelligence Research, 4, 237–285. [in English].
6. Hu J., Wellman M. P.(2003). *Nash Q-learning for general-sum stochastic games.* Machine Learning Research. 4, 1039–1069. [in English].
7. Gorodetsky V.I. (2012). *Self-organization and multi-agent systems. I. Models of multi-agent self-organization.* Izv. RAS. Theory and control systems. 2. 92–120. [in Russian].

АСЄЄВ Владислав Дмитрович

магістр комп'ютерних наук, Чорноморський національний університет ім. Петра Могили, Миколаїв, Україна, aseev.vlad@gmail.com, ORCID ID: 0000-0002-3081-1651

КУЛАКОВСЬКА Інесса Василівна

кандидат фізико-математичних наук, в.о.доцента кафедри інтелектуальних інформаційних систем Чорноморського національного університету імені Петра Могили, Миколаїв, Україна, inessa.kulakovska@chmnu.edu.ua, ORCID ID: 0000-0002-8432-1850

МОДЕЛЮВАННЯ КООПЕРАТИВНОЇ ПОВЕДІНКИ В МУЛЬТИАГЕНТНИХ СИСТЕМАХ

Анотація. Під кооперативною поведінкою розуміють спільноту агентів, які для зменшення середньозважених штрафів вирішують співпрацювати, щоб розв’язати поставлену задачу або досягти певної мети, в нашому випадку синхронізувати світіння.

Проблема формування кооперативної поведінки інтенсивно досліджується у сучасній науковій літературі із застосування мультиагентних систем, наприклад, для дистанційного навчання, керування організаційними системами, побудови різноманітних віртуальних організацій та спільнот, керування розподіленими обчисленнями, керування суспільними інститутами та суспільно-економічними процесами та інших.

У даній роботі розглянуто актуальну тему оптимальних стратегій в іграх з локальною взаємодією, розглянуто стимулююче навчання мультиагентних систем у ігровій постановці.

Метою даної роботи є розгляд методу побудови системи з локальною взаємодією агентів на основі завдання «синхронізації» за допомогою марковської моделі стохастичної гри.

Метод дослідження - комп'ютерна програма для моделювання завдання з використанням Q-методу навчання. Формування коаліцій у мультиагентних системах формулюється як конкурентна або кооперативна задача зарахування об’єкта до одного із кластерів. Розв’язування подібних задач вивчає теорія ігор, а в умовах невизначеності – теорія стохастичних ігор. У зв’язку з цим з наукової та практичної позицій актуальне застосування методів стохастичних ігор для формування коаліцій в умовах неповноти інформації. Рішення стохастичної гри полягає в пошуку стратегій агентів, які максимізують свої виграші, щоб забезпечити певний колективний баланс інтересів для всіх гравців. Для пошуку оптимальних стратегій гравців в умовах невизначеності будемо використовувати метод стимулювання навчання.

Результатом є розроблена ігрова модель, яка забезпечує динамічну самоорганізацію МАС, що проявляється в ритмічній зміні чистих стратегій агентів, які імітують світлові ефекти колонії світлячків. Характерною особливістю

розглянутої самоорганізації гри є локально визначена інформація про стратегії поведінки сусідніх агентів, які в результаті навчання призводить до глобальної координації. стратегії всіх агентів.

Наукова новизна отриманих результатів полягає в розробці ігрової моделі, визначені ефективності ігрової самоорганізації стратегій МАС для розв'язування задачі прийняття рішень в системах з кооперативною поведінкою агентів в умовах невизначеності. Повторність значень характеристик гри в різних експериментах з унікальними послідовностями випадкових величин підтверджує достовірність результатів.

Результати можуть бути використані на практиці для моделювання динаміки соціальних процесів, керування соціальними інтернет-сервісами у мережі інтернет та інших.

Ключові слова: *мультиагентна система, стохастична гра, адаптивний ігровий метод, Q-метод.*

АСЕЕВ Владислав Дмитриевич
магистр компьютерных наук, Черноморский национальный университет имени Петра Могилы,
Николаев, Украина, aseev.vlad@gmail.com, ID ORCID: 0000-0002-3081-1651

КУЛАКОВСКАЯ Инесса Васильевна
к. ф.-м. наук, в.о.доцента кафедры интеллектуальных информационных систем Черноморского национального университета
имени Петра Могилы, Николаев, Украина, inessa.kulakovska@chmnu.edu.ua, ID ORCID: 0000-0002-8432-1850

МОДЕЛИРОВАНИЕ КООПЕРАТИВНОГО ПОВЕДЕНИЯ В МУЛЬТИАГЕНТНЫХ СИСТЕМАХ

Под кооперативным поведением понимают деятельность множества агентов, которые для уменьшения средневзвешенных штрафов решают сотрудничать для решения поставленной задачи или достичь определенной цели, в нашем случае синхронизировать свечение. Проблема формирования кооперативного поведения интенсивно исследуется в современной научной литературе по применению мультиагентных систем, например, для дистанционного обучения, управления организационными системами, построения различных виртуальных организаций и сообществ, управления распределенными вычислениями, управления общественными институтами и общественно-экономическими процессами и другое.

В данной работе рассмотрена актуальная тема оптимальных стратегий в играх с локальным взаимодействием, рассмотрено стимулирующее обучение мультиагентных систем в игровой постановке.

Целью данной работы является рассмотрение метода построения системы с локальным взаимодействием агентов на основе задания «синхронизации» с помощью марковской модели стохастической игры.

Метод исследования - компьютерная программа для моделирования задачи с использованием Q-метода обучения. Формирования коалиций в мультиагентных системах формулируется как конкурентная или кооперативная задача присоединения объекта к одному из кластеров. Решение подобных задач изучает теория игр, а в условиях неопределенности - теория стохастических игр. В связи с этим с научной и практической позиций актуально применение методов стохастических игр для формирования коалиций в условиях неполноты информации. Решение стохастической игры заключается в поиске стратегий агентов, которые максимизируют свои выигрыши, чтобы обеспечить определенный коллективный баланс интересов для всех игроков. Для поиска оптимальных стратегий игроков в условиях неопределенности будем использовать метод стимулирующего обучения.

Результатом является разработанная игровая модель, которая обеспечивает динамическую самоорганизации МАС, что проявляется в ритмической смене чистых стратегий агентов, которые имитируют световые эффекты колонии светлячков.

Научная новизна полученных результатов заключается в разработке игровой модели, определении эффективности игровой самоорганизации стратегий МАС для решения задачи принятия решений в системах с кооператив-

ным поведением агентов в условиях неопределенности. Повторность значений результатов самообучения агентов в игре со случайными последовательностями начальных стратегий подтверждает достоверность результатов.

Результаты могут быть использованы на практике для моделирования динамики социальных процессов, управления социальными интернет-сервисами в сети интернет и других.

Ключевые слова: *мультиагентная система, стохастическая игра, метод адаптивных игр, Q-метод.*

ЛІТЕРАТУРА:

1. Kravets P.O. Gaming model of self-organization of multi-agent systems. Bulletin of the National University "Lviv Polytechnic". Series: Information systems and networks: a collection of scientific works. 2015. N829. P. 161-176. URL: http://ena.lp.edu.ua:8080/handle/ntb/31452
2. Wooldridge M. An Introduction to Multiagent Systems. John Wiley&Sons, 2002. 366 pp.
3. Weiss, G. Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence. Springer Verlag, Berlin, 1996. 643 pp.
4. Musiyenko M., Zhuravska I., Kulakovska I., Kulakovska A. Simulation the behavior of robot sub swarm in spatial corridors. 2016 IEEE 36th International Conference on ELECTRONICS and NANOTECHNOLOGY (ELNANO). April 19-21, 2016. Kyiv, Ukraine. Page(s) 382-387. DOI: 10.1109/ELNANO.2016.7493090 http://ieeexplore.ieee.org/document/7493090/
5. Leslie Kaelbling, Michael L. Littman, Andrew W. Moore. Reinforcement learning: A survey. Journal of Artificial Intelligence Research. 1996. N.4. P. 237–285.
6. J. Hu, M. P. Wellman. Nash Q-learning for general-sum stochastic games. Machine Learning Research. 2003. N. 4. P. 1039 – 1069.
7. Gorodetsky V.I. Self-organization and multi-agent systems. I. Models of multi-agent self-organization. Izv. RAS. Theory and control systems. 2012. N. 2. P. 92–120.