INCENTIVE LEARNING OF MULTIAGENT SYSTEMS

P.O. Kravetz
Lviv Polytechnic National University, Information Systems and Networks Department

The functioning of majority of modern information systems (IS) is based on hard-coded algorithms. In conditions of unpredictable environmental effects in such systems may disrupt the stability of the system modes that can lead to all sorts of emergency situations. To prevent the critical states software of distributed IS should be composed of interacting autonomous modules, be intelligent, flexible, able independently to track the change of states of the external environment and to take timely and appropriate decisions. Otherwise, such systems should be based on the principles of agent-oriented methodology. The agent of IS is a autonomous software module with elements of artificial intelligence, able to make decisions independently, to interact with the environment, other agents and human being in the course of solving the task. The interaction of agents of IS is implemented within a computer network. The population of agents in a computer network that solve the joint problem is called multi-agent system (MAS).

The functioning of MAS is usually carried out under conditions of a priori uncertainty of the information about the statuses of environment of decision making and the actions of other agents. In this regard, the strategic behavior of agents should be adaptive due to the ability of agents to the self-teaching. Among the methods of learning under uncertainty practical appeal is received by methods based on stimulation, as they do not require a mathematical model of the environment and ensure the possibility of decision-making directly in the learning process. The mechanisms of reflex behavior of living organisms with a developed nervous system make the basis of stimulation learning. Effective method of encouraging learning is Markiv Q-learning, which provides a numeric identification of the characteristic function of dynamical system in the space of state-action. The function of total expected reward of the agent is typically in the form of the characteristic function.

In comparison with monoagent systems, structure, functioning and research of methods of multiagent Q-learning is much more complicated. Due to the collective interaction of agents stationary environment is transformed into a class of non-stationary. The change of environmental states and the value of the winnings of each agent depend on actions of other agents. In general in MAS the agent cannot reach the maximum winning equal to its winning in monoagent system. Optimal winnings of agents must be balanced and meet the criteria of benefit, justice, equilibrium. So, instead of the criterion of scalar maximization of the winnings of monoagent system, the criteria of vector maximization of MAS winnings, for example, equilibrium according to Nesh, Pareto optimality, etc are introduced.

When using the method of MAS Q-learning occurs iterative construction of a system of characteristic Q-functions in the space of state-action, with the growth of the elements of these functions is made in the direction of achievement of their collective balance.

To build MASS, it ia needed to perform a preliminary studies on the basis of adequate mathematical models that allow to study the dynamics of the system under uncertainty, build strategic behavior of the agents that provide the best technical and economic parameters of operation of the system. Taken into consideration the characteristics of the subject area, namely, multiagent form, the uncertainty of the environment of decision making, antagonism, or competition of goals, communication, action coordination, adaptability of strategies of the behavior of agents to build models of MAS is used the mathematical apparatus of the theory of stochastic games. The solving of stochastic games lies in the search for such strategies of agents that maximize their winnings to provide certain collective balance of interests of all players. The search for optimal strategies of players in the conditions of uncertainty is workable based on the method of incentive learning.

The aim of this work is to build an iterative method of encouraging learning for solving stochastic games of MAS in the conditions of uncertainty.The goal should be reached based on the development of a model of multi-agent stochastic game, determination of the criteria of a collective equilibrium, method and algorithm for solving game problems.

The given method of incentive learning in determined variant requires knowledge of each agent in Q – functions of all other agents. These functions are used by agents to identify strategies that ensure the dynamic method in the direction of points of collective equilibrium.

Function values can be obtained as a result of the exchange of information between agents. If integrated information about the Q-functions are not available to the agent, then the agent must determine

their values independently in the learning process, observing the current winnings of other agents and performing evaluation of Q-functions. If such observations are not possible, then agents can perform a reflexive evaluation of Q-functions of other agents.

Another way of constructing algorithms of incentive learning of agents under conditions of uncertainty lies in the application of the method of stochastic approximation for the corresponding conditions of the collective equilibrium.

Practical use of methods of gaming incentive learning requires a preliminary analysis to determine the conditions of convergence to the collective state of equilibrium. Such studies are carried out on the basis of estimation of sequences of random variables that characterize the current deviation strategies of players from their optimal values.

The speed of convergence of Q-method of incentive method is determined by the parameters α_t and T. Parameter α_t must satisfy the general conditions of the stochastic approximation. The value of the parameter T depends on the absolute values of the elements of the Q-matrices. It was established experimentally that given in the work matrices of average winnings which elements take values from the interval [0, 1], the convergence of the gaming Q-method is provided when $T \in (0,\ 0.2]$ in the range of values of the parameter $\alpha_t = t^{-\kappa}$, $\kappa \in (0, 1]$. The highest speed of convergence of game method of incentive learning is achieved when $T = 10^{-2}$.

The article the problem of incentive learning of multi-agent systems in a playful formulation is described. The Markiv model of stochastic game is made, criteria of game-based learning are formulated, Q-method and the corresponding algorithm for solving stochastic games are described, the results of a computer implementation of the Q-method is analyzed.

Keywords - multi-agent system, stochastic game, incentive learning, Q-method.