GAME DECISION MAKING IN MULTI AGENT SYSTEMS

Temani Moncef\textsuperscript{1,1}, Rekik Ali\textsuperscript{2}, and Gabsi Mounir\textsuperscript{3}

\textsuperscript{1}Department of Computer Sciences, Faculty of Sciences, Tunis, Tunisia
monc.tem@gmail.com\textsuperscript{1}

\textsuperscript{2}Department of Computer Sciences, Higher Institute of Technological Studies, Sfax, Tunisia
alirekik1@yahoo.com\textsuperscript{2}

\textsuperscript{3}Department of Computer Sciences, Higher Institute of Technological Studies, Nabeul, Tunisia
mounirgabsi@yahoo.fr\textsuperscript{3}

Abstract: The system multi-agent (SMA) is a compound system of a set agents, situated in a certain environment and interacting to certain relations. The theory of the games is a formalism which aims at studying the interactions between agents knowing that such interactions can extend from the cooperation to the conflict. The game decision making in a system multi-agent require the players to manage a high number of units placed in a very sophisticated environment where complex real-world problems are posed and state of the artificial intelligence techniques are ineffective.

This article describes the problem of game decision making in multi-agent systems. Then we present the game model. After that, we propose the formalization of the game problem. Finally, we quote the methods for solving the game problem.

INTRODUCTION

The problem of effective decision-making is one of systems of command the most using in the artificial intelligence \cite{1}. The manufacturing process and the decision-making are formulated according to the purposes decided by the interaction of the intellectual in agent's environment of problem \cite{2}, under the decision in discreet systems understand(include) the choice of an agent with alternative actions based on control of appropriate environmental effects and the knowledge a priori of an agent on sectors of problem.

In distributed systems, the process of decision-making is the game of agents \cite{3}. It complicates the choice of effective solutions because their combination of options grows exponentially with the increasing number of agents. In this respect, the construction of the models and the methods based on the elements of the theory of artificial intelligence makes the reduction of space to find effective solutions for the scientific problems and the real practical cases.

Using multi-agent systems provide several advantages, which include the following:

To speed operations working by parallelism in systems that allow decomposition global problem of subtasks that can be performed by individual agents.

- Increase the stability of the system in case of failure of individual agents through their replacement by other agents.
- Scaling the system by introducing new agents, possibly with new features.
- The best technical and economic performance compared with systems based on full-blow agents, because the cost of individual agents is usually less due to their specific functions.

The agents of decision-making for the systems to multi-agent are made individually and maybe independently the current reaction of the means which is decided by the collective actions of many agents. The game full of options is decided by the collective of the Cartesian product of agents' individual actions. In the process of interaction with the agents of environment can resell their own versions having many decisions.

In practice, we choose options in the manufacturing of a priori uncertain environment as the free game of the actions of individual agents \cite{4}. In such a circumstances, the system to multi-agent has no necessary data to calculate and implement the solutions in the optimal stage. Agent's coherent interaction with the environment chooses then finally a protocol, which is based put a mechanism of probability decision-making. Agents' effective method of decision-making in front of the uncertainty is the used in the rules of Markov to treat the current information \cite{5, 6}. In agents of system Markov, ignoring the context, we calculate the solution following the base of the data in the current moment realizing such an environment of effects, which will optimizes finally the edition of the means of reaction, as the maximization of the average return we shall reduce at least secondary losses.

The global purpose of multi-agent's system is probably the coordination of the agents \cite{7}. The coordination is the process of determination, which supplies a global purpose of the system to multi-agent in the functions of objective, concerned the optimization of local agents. In multi-agent's optimization the global solution, as a general rule, is a compromise on the best local solutions. The coordination extends in the imposition of limitation for the agents' possible actions. The Known manners to resolve such of problems of coordination: communication, collective agreements and agents training using the communication which allows every agent to inform the other agents of possible actions, so limiting their choice. Collective agreements have to determine the rules of

© JGRCS 2010, All Rights Reserved
common behavior of agents. If the agents reveal their intentions for the new action, in every situation, the rule (ruler) of priority chose some actions from the others.

The problem of coordination is compounded by the exponential expansion of space combined options multi-agent systems and collective decision-making under conditions of parametric or structural and functional uncertainty. In practice, as a rule, agents are characterized by locally-defined connections, and therefore there is no need to optimize their objective functions for all space solutions. Uncertainties in decision-making system are compensated using adaptive self-learning methods.

In a priori uncertainty for the efficiency and time-consuming decision-making to a large extent depend on the ability of agents to learn and adapt to the uncertainty of decision-making [8 - 10]. Education agents are to establish restrictions on ineffectiveness actions can be performed calculation of dynamic weights (or priorities) options. Obtained by the reaction medium on completed actions, the agents perform recursive computation of new weights of discrete options, taking into account the current value of prizes. At this point in time, the option is implemented with a maximum weight. Implementation of options may be done by chance, based on the probability choices, which can be calculated valuation weights of possible actions of agents. Education agents are based on models, methods and algorithms, control theory and artificial intelligence - machines with constant and variable structure [11], networks of M-automata [12], models of adaptive random search [13], Markov decision-making processes [6], various variants of Q-learning [14], artificial neural networks [15], selective and genetic algorithms [16], Bayesian networks [17], heuristics [18], models of stochastic games [19, 20], etc.

The Study’s coordination of agents in an uncertain environment appropriate to perform model-based stochastic games, which make it possible to investigate the processes of competition and cooperation in a team of agents to identify compromise solutions, decentralized system.

The aim is to construct stochastic game models of decision making in multi-agent systems, development of methods for solving stochastic games, and determining conditions of their convergence to equilibrium states of the Collective.

GAME MODEL OF MULTI-AGENT SYSTEM

In a game of interpretation agents are players who are characterized by vectors of pure and mixed strategies. Elements of mixed strategies determine the probability of selecting the appropriate pure strategies in discrete moments of time. Players choose pure strategies, regardless of time and independently of each other in accordance with a probabilistic mechanism, built on the basis of mixed strategies. After the implementation of collective strategies players receive a random realization of this winning with a priori unknown distribution law [21]. Winning strategies for each player is determined by the local subset of its neighboring players. In the game without the exchange of information the players do not notify each other of the realized strategy and the value obtained payoff. In the game with the exchange of information, each player said the players from the local subset of the importance of this win. In the game, players pass the current failures moves with a certain probability. As a result, each player will receive information from many players, winnings, which depend on its strategies. The convolution of the current win-weighted positive coefficients is an assessment of the strategy chosen by the player in the current moment. Sequences of selected variants estimated averaged over the prehistory of the functions of the average payoff of players. The aim of each player in the asymptotic time to maximize the average payoff function. Interchanges vector optimization problem looking at the set of Nash equilibrium points (for the game without the exchange of information) or Pareto-optimality (for the game with the exchange of information) [22].

Asymptotic goals are achieved through self-learning of Markov recursion methods for forming vectors of mixed strategies [23], which can be obtained by the method of stochastic approximation [24] on the basis deterministic formulation of adequate matrix game problem [22]. Sufficient conditions for convergence of the method concerning the game are based on upper bounds for the conditional expectation of error for the current the problem’s solution of averaging the estimates obtained for the realization of the events and results concerning the theorems on recursive numerical inequalities [21, 25].

FORMALIZATION OF THE GAME PROBLEM

Assume there is no empty set of players D Assume there is no empty set of players D, each \( i \in D \) performs at discrete time’s \( n = 1,2,... \) independent choice of one of their own pure strategies \( U_i^i = u_i \in U_i = (u_i(1),u_i(2),...,u_i(N_i)) \) and by the time \( n + 1 \) observed random current win \( \xi_n^i = g_n(u_n^D_i) \), that is a function of joint strategies \( u_n^D_i \in U_i^D = \bigotimes_{j \in D} U_j \) players from the local subset \( D_i \subseteq D, D_i \cap D_j = \emptyset \) \( \forall i \in D \). It is believed that the sequence of random variables \( \{ \xi_n^i \}_{n=1}^\infty \) independent \( \forall u_n^D_i \in U_i^D, \forall i \in D, \forall n = 1,2,..., \) and their mathematical expectations

\[
M \{ [g_n(u_n^D_i)] \} = \sigma^2(u_n^D_i) \quad \infty \text{.}
\]

Matrix expectation of winning \( [V^i(u_n^D_i) \forall i \in D \) called the game environment. If \( \forall u_n^D_i \in U_i^D, \forall i \in D, v(u_n^D_i) \} \) \( \rho \) that environment is a positive sign, if \( v(u_n^D_i) \} \) \( - \rho \) - negative sign, otherwise - a general form.

We consider such models play: no sharing and the exchange of information, without failures and failures of the players. In the game without the exchange of information the players do not notify each other of the realized strategy and the value received winning \( g_n^i \).

In the game players are characterized by a failure probability \( \eta^i \in [0,1] \). Let \( \psi^i \in [0,1] \) - sign of participation of \( i \) - player in the game. If \( \psi^i = 0 \), then the player refuses the
current status of the game with probability $\eta^i$, if $\psi^i = 1$ - participates in the game with probability $1 - \eta^i$. At each stage of the game featured an exchange of current states $\psi^i$ between neighboring sets of players from $D_i \forall i \in D$. Waivers players lead to a change in the composition of sets of $D_i \forall i \in D$, and consequently, to change the current win. Enter the complete group of events $\psi^i = 2^{\left| D^i \right|}$, associated with failures of players with sets $D_i$. Let $D_i(\omega) \subseteq D_i$ - subset of players who remain in the game for the event $\omega \in \psi^i$.

Then the current gain of i-th player equal

$$\zeta^i_n = \chi(\prod_{j=1}^{n} \psi^j_n \geq \psi^i) \xi^i_n(\hat{u}^i_n(D^i(\omega)))$$

where $\chi(\cdot) \in \{0,1\}$ - indicator function of event, $\hat{\mathcal{Y}} \prod 0$ - threshold games, or minimal amount of players who have not denied. Given the failures of the players selected sequence variants $\{u^i_n(D^i(\omega)) \mid t = 1,n\}$ estimated current average yield

$$\phi^i_n(\{u^i_n(D^i(\omega))\}) = \frac{1}{n} \sum_{t=1}^{n} \zeta^i_t \forall i \in D$$

The aim of each player is to maximize the function of the average payoff

$$\lim_{n \to \infty} \phi^i_n(\{u^i_n(D^i(\omega))\}) \to \max \forall i \in D$$

Finding solutions to the vector optimization problem (2) is carried out in the set of Nash equilibrium points (for the game without the exchange of information)

$$\forall i \in D \lim_{n \to \infty} \phi^i_n(\{u^i_n(D^i(\omega))\}) - \phi^i_n(\{\hat{u}^i_n(D^i(\omega))\}) \geq 0$$

or Pareto optimality (for the game with the exchange of information)

$$\forall i \in D \lim_{n \to \infty} \left[ \phi^i_n(\{u^i_n(D^i(\omega))\}) - \phi^i_n(\{\hat{u}^i_n(D^i(\omega))\}) \right] \geq 0$$

where inequality (3) and (4) are satisfied with probability 1, and $u^i_n(D^i(\omega)), \hat{u}^i_n(D^i(\omega)) \in U(D^i(\omega))$; $\hat{u}^i_n = u^i_n \setminus \hat{u}^i_n \in U(D^i(\omega))$; $u^i_n, \hat{u}^i_n \in U^i$.

METHODS OF SOLVING THE GAME PROBLEM

Asymptotic goals (3) or (4) are achieved through self-learning recursive methods for forming vectors of mixed strategies $p^i_n$, elements of which are conditional probabilities of selecting the appropriate pure strategies, $p^i_n(u^i_n) = P[u^i_n \mid \psi^j_n \geq \psi^i \mid (t = 1,n - 1)] \forall u^i_n \in U^i, \forall i \in D$

$$V^i = \prod_{\psi^i \in \psi^j} \prod_{u^i_n \in U^i} \{(1 - \eta^i)\chi(\psi^i = 1) + \eta^i \chi(\psi^i = 0)\} V^i(\omega)$$

Where $V^i(\omega) = \prod_{u^i_n \in U^i} \prod_{u^i_n \in U^i} p^i_n(u^i_n)$ - function of average yield, determined for one of the situations of the game with failures.

Pareto-optimal decoupling matrix game is defined by the method of conditional maximization of the local bundle features the average yield of $W^i = \prod_{\psi^i \in \psi^j} \chi(k)W^i \chi(k)$ convex unit simplex $S^{N_i} \forall i \in D$.

For the differentiated $S^{N_i}$ functions of $W^i$ optimal mixed strategy are provided with complimentary no rigidity.

$$\nabla_p W^i = W^i e^{N_i}, p^i \in S^{N_i}, \forall i \in D$$

where $\nabla_p W^i$ - gradient function $W^i; e^{N_i}$-vector consisting of $N_i$ units.

Condition (5) defines a smoothing of the Nash strategies for coalitions of players $D^i \forall i \in D$. In general, these strategies are local $\epsilon$-optimal solutions on Pareto-base coalition-free games. Leveling a Nash strategy coalition-free games derived from (5) for $W^i = V^i \forall i \in D$.

The current decision clarifies the players with the help of recursion method [21] $p^i_{n+1} = \pi^N_{\epsilon^i} \{p^i_n - R(x^i_n, p^i_n, \xi^i_n)\}$.

where $\gamma_n$ - step method; $R(x^i_n, p^i_n, \xi^i_n)$ - motion vector method; $\pi^N_{\epsilon^i}$ - projector on a single $\epsilon$-simplex [21], necessary for the normalization of $p^i_n$ and the accumulation of complete statistical information on a random environment. Motion vector method to ensure compliance with conditions (2). For this large value of current gains $\xi^i_n(x^i_n)$ must comply with high values of conditional probabilities $p^i_n(x^i_n)$.

Asymptotic decoupling vector optimization problem (2) must be sought in the set of points in the Nash equilibrium, Pareto optimality, etc. [21, 22]. Achieving a specific solution is determined by the type of method (5) and a way to change its adjustable parameters. Based on the language game problem under uncertainty and a deterministic matrix game problem, using stochastic approximation conditions are not complimentary stiffness (5) is constructed such recurrent Markov game methods:

$$p^i_{n+1} = \pi^N_{\epsilon^i} \{p^i_n - \gamma \xi^i_n \epsilon^N - \frac{e(u^i_n)}{e^T(u^i_n)p^i_n} \epsilon^T(u^i_n)p^i_n \}$$

$$p^i_{n+1} = \pi^N_{\epsilon^i} \{p^i_n - \gamma \xi^i_n(p^i_n - e(u^i_n)) \}$$
where \( \pi_{\varepsilon}^{N_i} \) - operator design for \( \varepsilon \) - simplex \( S_{\varepsilon}^{N_i} \subseteq S_{\varepsilon}^{N_i} \); 
\( \gamma_n \geq 0 \) - parameter, which governs the step size method; 
g \in \{0,1\}; \( e(u_n^i) \) - unit vector indicator chose \( u_n^i \).

Method (6) obtained on the basis of the conditions are not complimentary terms of rigidity (5). Method (7) obtained on the basis of the component weighting vectors conditions (5)

\[
Z^i = \text{diag}(p_n^i)(e^{N_i}W^i - \nabla \rho W^i)
\]  
(8)

where \( Z^i \in R^{N_i}; \) \( \text{diag}(p_n^i) \) - square diagonal matrix of order \( N_i \), composed of the elements of the vector \( p_n^i \).

Weigh-in (8) gives an opportunity to consider decoupling the game problem in completely mixed strategies, which are placed on the boundary of the unit simplex.

A number of modifications of (6) and (7): the gradient method (6) with \( g = 0 \); without projection method (7) under constraints \( \gamma_n \in [0,1]; \) methods without the exchange of information with \( \zeta_n^i = \xi_n^i \); information-sharing with \( \zeta_n^i = \xi_n^i \); with failures of players in the \( \xi_n^i = \chi(\prod_{p \in D} u_{n}^{p}) \); regulated methods to \( \zeta_n^i = \xi_n^i + \delta_n^T(u_n^i)p_n^i \).

Determine sufficient conditions for the convergence of gaming practices (6) and (7) to the asymptotic optimal solutions in certain environments and sign the general form of media. Based on the upper estimates of the conditional mathematical expectation of the current error of \( \Delta_n = \prod_{i \in D} \| Z^i \|^2 \) conditions are not complimentary to the rigidity of a fixed background of events and consequences of theorems on recurrent numerical inequalities [21] the conditions of convergence with probability 1. Based on estimates obtained by averaging realizations of events obtained conditions of convergence in the quadratic medium. Estimation of the asymptotic order of convergence holds for sequences of \( \gamma_n^\varepsilon \alpha_n^\varepsilon \beta_n^\varepsilon \); \( \varepsilon_n = \Theta_n^\varepsilon \beta_n \); \( \lim_{n \to \infty} n^\varepsilon M \{ \Delta_n \} \leq \vartheta \)

where \( \Theta \) - order parameter, \( \vartheta \) - value of the rate of convergence. Large \( \Theta \) and \( \vartheta \) corresponds to a lower rate of convergence is a big game method.

Found that as a sign of positive environment the maximum order average quadratic speed of convergence of method (6) is \( n^{-1/2} \), which is due to \( \alpha \in (1/2,1], \beta = \alpha - 1/2 \).

Maximum asymptotic order of the mean velocity convergence of the method (7) is \( n^{-1} \), which is achieved when \( \alpha = 1, \beta \geq 1 \).

Among the general form of method (6) maximizes the asymptotic order of convergence rate equal to \( n^{-1/3} \) and is achieved at \( \alpha = 2/3, \beta = 1/3 \). For the method (7) the maximum order equals \( n^{-1/2} \), attained for \( \alpha = 2/2, \beta \geq 1/2 \).

From these results it follows that under no certain method (7) provides a higher order rate of convergence to the set of optimal solutions than the method (6) - as a sign of certain environments and in environments of general form. According to the results of additional studies found that the failures of the players lead to the \( \varepsilon \)-optimal solutions of problem gaming and to slow the rate of convergence. With a decrease in the threshold game \( \vartheta \) rate of convergence playing techniques is growing. Methods without sharing and information exchange have the same asymptotic order of convergence rate \( \Theta \) in the affine-equivalent environments. Exchange of information between players leads to an increase of the speed of convergence (reducing the value \( \vartheta \)). These methods provide stability of solutions to the game problem in mixed strategies.

**CONCLUSION**

Developed game models make it possible to investigate the processes of coordination of collective interaction in multi-agent system decision-making. In the game formulation of coordination is the process of harmonizing the actions of agents to achieve compromise collective decisions. Coordination is achieved by the imposition of restrictions on the actions of rational agents in order to implement effective solutions. Construction of the compromise solutions by coordinating the actions of agents carried out the methods of learning and adapting to the uncertainty of decision-making.

The model of game problems of adaptive selection of options in making a priori no certainty to the exchange of information and denial allow players to expand applications of the theory of stochastic games class of problems of distributed control of random processes with local interaction.

For models built on the basis of supplementary conditions not rigidity and properties of equalizing strategies using stochastic approximation synthesized a new class of recurrent playing methods, which helped with the uniform system position examine the conditions of their disability. Established that the exchange of information between players increases the rate of convergence of gaming practices and it is a prerequisite for finding the Pareto-optimal solutions of games with no a priori certainty of payoff functions. Sufficient conditions for convergence are determined by the structure of this exchange, the type of recurrent practices and restrictions on their parameters.

Found that restorative failure of players lead to slower rate of convergence of gaming practices and the the asymptotic solutions rejection of the game from the optimal values obtained at the absence of failures players.

The overall analysis of the game selecting the choices made in conditions of certainty indicates that further theoretical and empirical research in this area should focus on building and setting terms of efficiency as educational gaming techniques with elements of cognitive processing.
ACKNOWLEDGMENT

The authors gratefully acknowledge the research in Tunisia was conducted under scientific agreements between the Research Group SOIE for (ISG Tunis). We thank our colleagues Khaled Ghedira, Dmytro Peleshko, for their inputs, in the unnumbered footnote on the first page.

REFERENCES