Collaborative Learning in Strategic Environments

Akira Namatame, Noriko Tanoura, Hiroshi Sato
Dept. of Computer Science, National Defense Academy
Yokosuka, 239-8686, JAPAN
E-mail: { nama, hsato}@cc.nda.ac.jp

Abstract

There is no presumption that collective behavior of interacting agents leads to collectively satisfactory results. How well agents can adapt to their social environment is different to how satisfactory a social environment they collectively create. In this paper, we attempt to probe a deeper understanding of this issue by specifying how agents interact by adapting their behavior. We consider the problems of asymmetric coordination, which are formulated as minority games, and we address the following question: how do interacting agents realize an efficient coordination without any central authority through self-organizing macroscopic orders from bottom up? We investigate several types of learning methodologies including a new model, give-and-take learning, in which agents yield to others if they gain and they randomize their actions if they lose or do not gain. We show that evolutionary learning is the most efficient in asymmetric strategic environments.

Keyword: *aymmetric coordination, social efficiency, evolutionary learning, give-and-take learning*

1. Introduction

There are many situations where interacting agents can benefit from coordinating their actions. Social interactions pose many coordination problems for individuals. Individials face problems of sharing and distributing limited resources in an efficient way. Consider a competitive routing problem of networks in which the paths from sources to destination have to be established by multiple agents. In the context of traffic networks, for instance, agents have to determine their route independently, and in telecommunication networks, they have to decide on what fraction of their traffic to send on each link of the network.

Coordination implies that increased effort by some agents leads the remaining agents to follow suit, which gives rise to multiplier effects. We classify this type of coordination as symmetric coordination [3]. Coordination is also necessary to ensure that their individual actions are carried out with little conflicts. We classify this type of coordination as asymmetric coordination [7]. Consider the following situation: A collection of agents have to travel using one of the route A or B. Each agent gains a payoff if he chooses the same route what the majority does. This type of coordination is classified as symmetric coordination. On the other hand, each agent gains a payoff if he chooses the opposite route what the majority does. This type of coordination is classified as asymmetric coordination.

Coordination problems are characterized with many equilibria, and they often face the problem of coordination failure resulting from their independent inductive processes [1][4]. An interesting problem is then under what circumstances will a collection of agents realizes a particular stable situation, and whether they satisfy the conditions of social efficiency? In recent years, this issue has been addressed by formulating minority games (MG)[2][10]. However, the growing literature on MG treats agents as automata, merely responding to changing environments without deliberating about individuals' decisions [13]. There is no presumption that the self-interested behavior of agents should usually lead to collectively satisfactory results [8][9]. How well each agent does in adapting to its social environment is not the same thing as how satisfactory a social environment they collectively create for themselves. An interesting problem is then under what circumstances will a society of rational agents realize social efficiency? Solutions to these problems invoke the intervention of an authority who finds the social optimum and imposes the optimal behavior to agents. While such an optimal solution may be easy to find, the implementation may be difficult to enforce in practical situations. Self-enforcing solutions, where

agents achieve optimal allocation of resources while pursing their self-interests without any explicit agreement with others are of great practical importance.

We are interested in the bottom-up approach for leading to more efficient coordination with the power of more effective learning at the individual levels [11]. Within the scope of our model, we create models in which agents make deliberate decisions by applying rational learning procedures. We explore the mechanism in which interacting agents are stuck at an inefficient equilibrium. While agents understand that the outcome is inefficient, each agent acting independently is powerless to manage decisions whci reflect collective activity. Agents also may not know about what to do and also how to make a decision. The design of efficient collective action is crucial in many fields. In collective activity, two types of activities may be necessary: Each agent behaves as a member of society, while at the same time, it behaves independently by adjusting its view and action. At the individual level, it learns to improve its action based on its own observation and experiences. At the same level, they put forward their learnt knowledge for consideration by others. An important aspect of this coordination is the learning rule adapted by individuals.

2. Formalism of Asymmetric Coordination and Minority Games

The EL Farol bar problem and its variants provide a clean and simple example of asymmetric coordination problems [1][4]. Brian Arthur used a very simple yet interesting problem to illustrate effective uses of inductive reasoning of heterogeneous agents. There is a bar called El Farol in the downtown of Santa Fe. In Santa Fe, there are agents interested in going to the bar each night. All agents have identical preferences. Each of them will enjoy the night at El Farol very much if there are no more than the threshold number of agents in the bar; however, each of them will suffer miserably if there are more than the threshold number of agents. In Arthur's example, the total number of agents is *N=100*, and the threshold number is set to 60. The only information available to agents is the number of visitors to the bar in previous nights.

What makes this problem particularly interesting is that it is impossible for each agent to be perfectly rational, in the sense of correctly predicting the attendance on any given night. This is because if most agents predict that the attendance will be low (and therefore decide to attend), the attendance will actually be high, while if they predict the attendance wto be high (and therefore decide not to attend) the attendance will be low. Arthur investigated the number of agents attending the bar over time by using a diverse population of simple rules. One interesting result obtained was that over time, the average attendance of the bar is about 60. Agents make their choices by predicting ahead of time whether the attendance on the current night will exceed the capability and then take the appropriate course of action. Arthur examined the dynamic driving force behind this equilibrium.

The Arthur's "El Farol" model has been extended in the form as Minority Games (MG), which show for the first time how equilibrium can be reached using inductive learning [2]. The MG is played by a collection of rational agents $G = \{A_i : 1 \le i \le N\}$. Without losing the generarity, we can assume N is an odd number. On each period of the stage game, each agent must choose privately and independently between two strategies $S = \{S_1, S_2\}$. We represent the action of agent A_i at the time period t by $a_i(t) = 1$ if he chooses S_1, and $a_i(t) = 0$ if he chooses S_2. Given the actions of all agents, the payoff of agent A_i is given by

$$\text{(i) } u_i(t) = 1 \text{ if } a_i(t) = 1 \text{ and } p(t) = \sum_{1 \le i \le N} a_i(t)/N < 0.5$$
$$\text{(ii) } u_i(t) = 0 \text{ if } a_i(t) = 0 \text{ and } p(t) > 0.5 \qquad (2.1)$$

Each agent first receives aggregate information $p(t)$ which represents all agents' actions, and then he decides whether to choose S_1 or S_2. Each agent is rewarded with a unitary payoff whenever the side he chooses happens to be chosen by the minority of the agents, while agents on the majority side get nothing. All agents have access to public information on the record of past histories on $p(\tau), \tau \le t$. The past history available at the time period t is represented by $\mu(t)$. How do agents choose actions under the common information $\mu(t)$? Agents may behave differently because of their personal

beliefs on the outcome of the next time period $p(t+1)$, which only depends on what agents do at the next time period $t+1$, and the past history $\mu(t)$ has no direct impact on it.

We analyze the structure of the MG to see what we should expect. The social efficiency can be measured from the average payoff of one agent over a long-time period. Consider the extreme case where only one agent take one side, and all the others take the other side at each time period. The lucky agent gets a reward, nothing for the others, and the average payoff per agent is *1/N*. Equally extreme situation is that when *(N-1)/2* agents on one side, *(N+1)/2* agents on the other side where the average payoff is about 0.5. From the society point of view, the latter situation is preferable.

The MG game is characterized with many solutions. It is easy to see that this game has $\binom{N}{(N-1)/2}$ asymmetric Nash equilibria in pure strategies in the case where exactly *(N-1)/2* agents choose either one of the two sides. The game also presents a unique symmetric mixed strategy Nash equilibrium in which each agent selects the two sides with un equal probability. With this mixed strategy, each agent can expect the payoff 0.5 on each time period, and the society payoff follows a binomial distribution with the mean equal to *N/2* and the variance *N/4*. The variance is also a measure of the degree of social efficiency. The higher the variance, the higher the magnitude of the fluctuations around *N/2* and the corresponding aggregate welfare loss. Several learning rules have been found to lead to an efficient outcome when agents learn from each other [2][15].

How exactly does an agent's utility depend on the number of total participants? We now show the MG can be represented as 2x2 games in which an agent play with the aggregate of the society with payoff matrix in Table 1. Let suppose each agent plays with all other agents individually with the payoff matrix in Table 1. The average payoffs of agent A_i from the play S_1 and S_2 with one agent are given:

$$U_i(S_1) = 1 - \sum_{1 \le i \le N} a_i(t)/N$$
$$U_i(S_2) = \sum_{1 \le i \le N} a_i(t)/N \qquad (2.2)$$

$p(t) = \sum_{1 \le i \le N} a_i(t)/N$ represents the proportion of agents who choose S_1 at the time period t.

Table 1 The payoff matrix of the minority games

Own's strategy \ The other's strategy	S_1 (go)	S_2 (stay)
S_1 (go)	0 / 0	1 / 1
S_2 (stay)	1 / 1	0 / 0

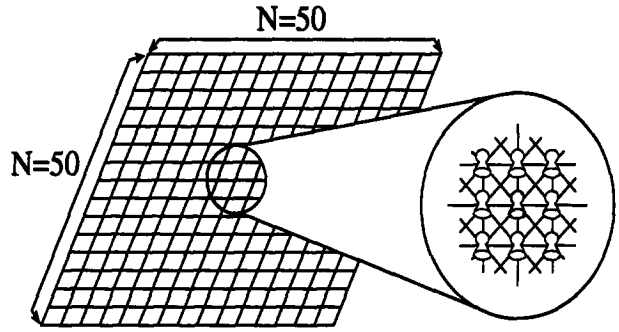
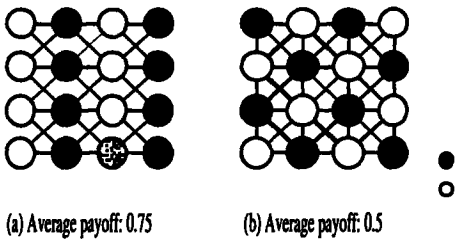
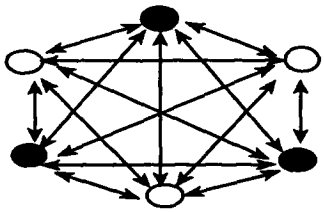


Figure 1 Local Matchings with 8 neighbours

The matching methodology also plays an important role in the outcome of the game. The uniform matching, random matching, or local matching are used as the matching methodologies in many literatures [3][6]. Agents interact with all other agents, which is known as the uniform matching. Agents are not assumed to be knowledgeable enough to correctly anticipate all other agents' choices, however they can only access information about the aggregate behavior of the society. The optimal solution of the MG depends on the matching methodology. As shown in Figure 2, if each agent is matched with all other agents (uniform matching), he receives 0.5 as the optimal payoff. As shown in Figure 2, if each agent is matched with their eight neighbors (local matching), he receives 0.75 or 0.5 as the optimal payoff, depending on how they split into two groups of choosing S_1 (dark circle) or S_2 (white circle).





(c) average payoff: 0.5

Figure.2: The optimal payoffs per agent with different matchings: (a) (b) Local matching with 8 neighbours, (c) Uniform or random matching

3. Learning Models in Strategic Environments

Game theory is typically based upon the assumption of rational choice. In our view, the reason for the dominance of the rational-choice approach is not because scholars think it is realistic. Nor is game theory used solely because it offers good advice to a decision maker, because its unrealistic assumptions undermine much of its value as a basis for advice. The real advantage of the rational-choice assumption is that it often allows deduction. The main alternative to the assumption of rational choice is some form of adaptive behavior. Adaptation may be expected at the individual level through learning, or it may be at the population level through differential survival and reproduction of the more successful individuals. Either way, the consequences of adaptive processes are often very hard to deduce when there are many interacting agents following rules that have nonlinear effects.

We specify how agents adapt their behavior in response to others' behavior in strategic environments. Among the adaptive mechanisms that have been discussed in the learning literature are the following [5][6][12][14]. An important issue in strategic environment is the learning strategy adapted by each individual.

(1) Reinforcement learning

Agents tend to adopt actions that yielded a higher payoff in the past, and to avoid actions that yielded a low payoff. Payoff describe choice behavior, but it is one's own past payoffs that matter, not the payoffs of the others. The basic premise is that the probability of taking an action in the present increases with the payoff that resulted from taking that action in the past [6].

(2) Best response learning

Agents adopt actions that optimize their expected payoff given what they expect others to do. In this learning model, agents choose best replies to the empirical frequencies distribution of the previous actions of the others.

(3) Evolutionary learning

Agents who use high-off payoff strategies are at a productive advantage compared to agents who use low-payoff strategies, hence the latter decrease in frequency in the population over time (natural selection). In the standard model of this situation agents are viewed as being genetically coded with a strategy and selection pressure favors agents that are fitter, i.e., whose strategy yields a higher payoff against the population.

(4) Social learning

Agents learn each from other with social learning. For instance, agents may copy the behavior of others, especially behavior that is popular to yield high payoffs (imitation). In contrast to natural selection, the payoffs describe how agents make choices, and agents' payoff must be observable by others for the model to make sense. The crossover strategy is also another type of social learning.

These learning models can be represented on the spectrum in Figure 3. The reinforcement learning and social learning based on give-and-take take limiting cases representing at the right-most and left-most points of the spectrum. models.

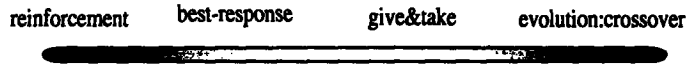


Figure.3 : The spectrum of learning

4. Best-Response Learning

With the assumption of rationality, agents are assumed to choose an optimal strategy based on a sample of information about what others agents have done in the past. Agents are able to calculate best replies and learn the strategy distribution of play in a society. Gradually, agents learn the strategy distribution in the society. With best-response learning, each agent calculates his best strategy based on information about the current distributional patterns of the strategies [5]. At each period of time, each agent decides which strategy to choose given the knowledge of the aggregate behavior of the population. Each agent thinks strategically, knowing that everyone else is also making a rational choice given its own information.

An important assumption is how agents receive knowledge of the current strategy distribution. For simplicity we assume that there is no strategic interaction across time; an agent's decision depends only on current information and not on any previous actions. The dynamics for collective decision of agents are described as follows: Let $p(t)$ be the proportion of agents who have chosen S_1 at time t. Let $U_i(S_k)$ be the expected payoff to A_i when he chooses S_k, $k=1,2$. The best-response of agent is then given as follows:

$$\text{If } U_i(S_1) > U_i(S_2) \text{ then choose } S_1$$
$$\text{If } U_i(S_1) < U_i(S_2), \text{ then choose } S_2 \qquad (4.1)$$

The expected payoffs of agent A_i are obtained as

$$U_i(S_1) = \theta(1-p(t)), \quad U_i(S_2) = (1-\theta)p(t) \quad (4.2)$$

The best-response adaptive rule of agent A_i is then obtained as follows:

$$\text{(i) If } p(t) < q, \text{ then choose } S_1$$
$$\text{(ii) If } p(t) > q, \text{ then choose } S_2 \qquad (4.3)$$

Aggregate information $p(t)$, the current status of the collective decision, has a significant effect on agents' rational decisions.

The result of the learning with the global best-response strategy is simple. Starting from any initial condition $p(0)$, it cycles between the two extreme situations where all agents choose S_1 or S_2. Under this cyclic behavior, no agent gains resulting in a huge waste. This result has a considerable intuitive appeal since it displays situations where rational individual action, in pursuit of well-defined preferences, lead to undesirable outcomes.

5. Collaboraive Learning with Give-and-Take

In this section, we propose the give-and-take learning which departs from the conventional assumption such that agents update their behaviors in order to improve their measure functions such as payoffs. It is commonly assumed that agents tend to adopt actions that yield a higher payoff in the past, and to avoid actions that yield a low payoff. With the give and take learning, on the contrary, agents are assumed to yield to others if they receive a payoff by taking the opposite starategy at the next time period, and they choose randomly if they do not gain the payoff. Each agent gets the common information $p(t)$ which aggregate all agents' actions of the last time period, and then he decides whether to choose S_1 or S_2 at the time period $t+1$ by considering whether he is rewarded at time t: He is rewarded a unitary payoff whenever the side he chooses happens to be chosen by the minority of the agents.

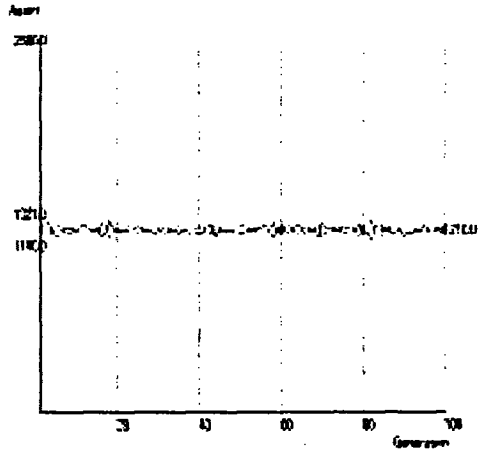


Figure 4 (a) The number of agents to choose S_1 and S_2 with the mixed strategy.

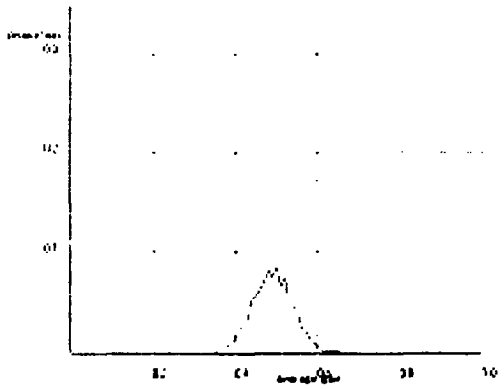


Figure 4 (b) The proportion of agents with the same average payoff with the mixed strategy

We formalize give-and-take learning as folows: The action $a_i(t+1)$ of agent A_i at the next time period $t+1$ is determined by the following rule:

(i) $a_i(t+1)=0$ (Choose S_2) if $a_i(t)=1$ and $p(t)\leq 0.5$ (Gain)

(ii) $a_i(t+1)=1$ (Choose S_1) if $a_i(t)=0$ and $p(t)>0.5$ (Gain)

(iii) $a_i(t+1)=RND(x)$ if $a_i(t)=1$ and $p(t)>0.5$ (No gain)

(iv) $a_i(t+1)=RND(x)$ if $a_i(t)=0$ and $p(t)\leq 0.5$ (No gain) (5.1)

where *RND(x)* represents the mixed starategy *x=(x, 1-x)* of chossing S_1 with the probability *x* and S_2 with 1-*x*.

We evaluate the performace by comparing the average payoff of an individual. The expected payoff of agents who choose S_1 is given *1-p*, and those of who choose S_2 is *p*, where p denotes the proportion of agents who choose S_1 .Therefore the average payoff oer individual is given by $2p(1-p)$, which takes the maximum value 0.5 at *p=0.5*. If the exact number *N=50* attends the bar , they receive 1 as the payoff and the average payoff is 0.5, which is the maximum payoff. With the corraborative learning, every agent receive the payoff and the majority of agents receive 0.35..

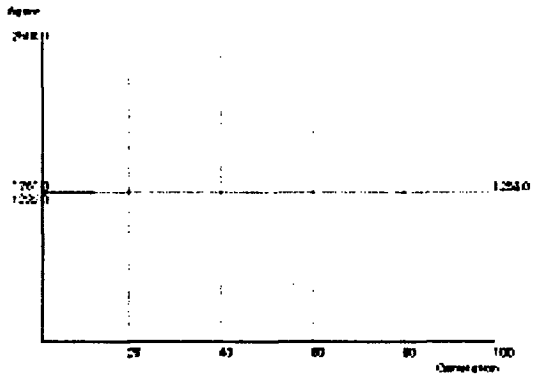


Fig.5(a): The number of agents to choose S_1 and S_2, with give-and-take learning

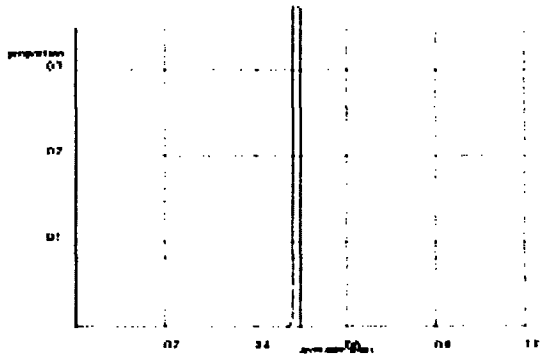


Fig.5(b): The proportion of agents with the same average payoff with give-and-take learning

6. Evolutionary Learning with Local Matching

In this section, we investigate evolutionary learning where agents learn from the most successful neighbours, and they co-evolve their strategies over time. Each agent adapts the most successful strategy as guides for their own decision (individual learning). Hence their success depends in large part on how well they learn from their neighbours. If the neighbour is doing well, its strategy can be imitated by all others (collective learning). In an evolutionary approach, there

is no need to assume a rational calculation to identify the best strategy. Instead, the analysis of what is chosen at any specific time is based upon an implementation of the idea that effective strategies are more likely to be retained than ineffective strategies [15]. Moreover, the evolutionary approach allows the introduction of new strategies as occasional random mutations of old strategies. The evolutionary principle itself can be thought of as the consequence of any one of three different mechanisms. It could be that the more effective individuals are more likely to survive and reproduce. A second interpretation is that agents learn by trial and error, keeping effective strategies and altering ones that turn out poorly. A third interpretation is that agents observe each other, and those with poor performance tend to imitate the strategies of those they see doing better.

In this section we consider the local matching as shown in Figure 1, where each agent is modeled to be matched with his 8 neighbours. Each agent is modeled to be matched several times with the same neighbour, and the rule of the strategy selection is coded as the list as shown in Figure 6. A part of the list is replaced with that of the most successful neighbour. An agent's decision rule is represented by the N binary string. At each generation gen, $gen \in [1,..,lastgen]$, agents repeatedly play the game for T iterations. An agent A_i, $i \in [1...N]$, uses a binary string i to make a decision about his action at each iteration t, $t \in [1...T]$. A binary string consists of 22 positions (genes). Each position p_j, $j \in [1,22]$, is represented as follows. The first and second position, p_1 and p_2, encodes the action that the agent takes at iteration $t = 1$ and $t = 2$. A position p_j, $j \in [3,6]$, encodes the history of mutual hands (cooperate or defect) that agent i took at iteration $t-1$ and $t-2$ with his neighbor (opponent). A position p_j, $j \in [7,22]$, encodes the action that agent i takes at iteration $t > 2$, corresponding to the position p_j, $j \in [3,6]$.

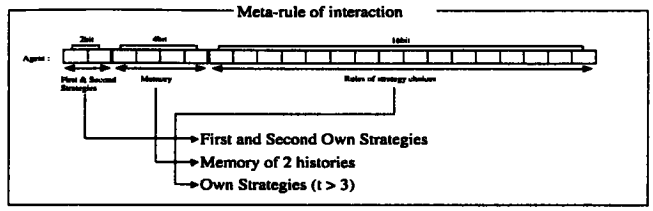


Figure 6: The representation of the meta-rule as the list of strategies

We consider two types of evolutionary learning, mimicry and crossover. Each agent interacts with the agents on all eight adjacent squares and imitates the strategy of any better performing one. In each generation, each agent attains a success score measured by its average performance with its eight neighbours. Then if an agent has one or more neighbours who are more successful, the agent converts to the strategy of the most successful of them or crosses with the strategy of the most successful neighbour. Neighborus also serve another function as well. If the neighbor is doing well, the behaviour of the neighbour can be shared, and successful strategies can spread throughout a population from neighbour to neighbour [11].

Significant differences were observed between the mimicry and the crossover. As shown Figure 7, the case with the mimicry strategy, each agent acquires a payoff of approximately 0.35, and with the cross-over, each agent acquires a payoff of approximately 0.7. Consequently, we can conclude that evolution learning leads to a more efficient situation in the strategic environments.

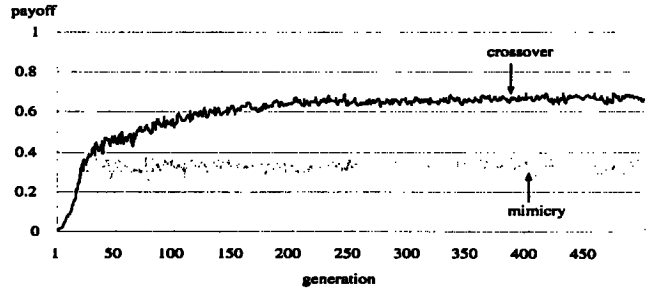


Figure 7: The average payoff with the evolutionary learning: crossover and mimicry

7. Conclusion

The interaction of heterogeneous agents produces some kind of coherent, systematic behavior. We investigated the macroscopic patterns arising from strategic interactions of heterougeneous agents who behave based on the local rules. In this paper we address the questions such as: 1) how a society of selffish agents self-organizes, without a central authority, their collective behavior to satisfy the constraints? 2) How does learning at individual levels generate more efficient collective behavior? 3) How does co-evolution in a society put its invisible hands to promote self-oranization of emerging collective bahaviors?

In previous work on collective behavior, the standard assumption was that agents use the same kind of adaptive rule. In this paper, we departed from this assumption by considering a model of heterogeneous agents with respect to their meta-rule of making decisons at each time period. Agents use ad hoc meta-rules to make their decision based on the past performance. We also consider several types of learning rules agents use to update their meta-rule for making decisions. We consider specific strategic environments in which a large number of agents have to choose one of two sides indepenently and those on the minority side win, which is knon as a minority game. A rational approach is helpless in our minority game as it generates large-scale social inefficiency. We intrduced a new learning model at the individual level, give-and-take learning where every agent should make his decison based on the past history of collective behavior. It is shown that emegent collective behavior is more efficient than that generated from the mixed Nash equilibrium strategies. We also proposed collaborative learning based on a Darwinian approach.. It is shown that in strategic environments where every agent has to keep improving his meta-rule for making decisons in order to survive, if agents learn from each other, then the social efficiency is realized without a central of authority.

8. References

[1]Arthur W.B. "Inductive reasoning and Bounded Rationality", *American Economic Review*, Vol.84, pp406-411, 1994.

[2] Challet,D& Zhang.C.," Emergence of Cooperation and Organization in an Evolutionary Game, Physica A246, 1997.

[3] Cooper, R: *Coordination Games*, Cambridge Univ. Press, 1999.

[4] Fogel B., Chellapia K., "Inductive Reasoning and Bounded Rationality Reconsidered", *IEEE Trans. of Evolutionary Computation*, Vol.3, pp142-146, 1999.

[5] Fudenberg, D., and Levine, D., *The Theory of Learning in Games*, The MIT Press, 1998

[6] Hart,S.,and Mas-Colell, *A*, Simple Adaptive Procedure Leading to Correlated Equilibrium, *Econometrica*, 2000.

[7] Iwanaga S., Namatame. A: "Asymmetric Coordination of Heterogeneous Agents", IEICE Trans. on Information and Systems, Vol. E84-D,, pp.937-944, 2001.

[8] Iwanaga S., Namatame. A: "The Complexity of Collective Decision, Journal of Nonlinear Dynamics and Control", (to appear).

[9]Kaniovski, Y., Kryazhimskii, A., & Young, H. Adaptive Dynamics in Games Played by Heterogeneous Populations. *Games and Economics Behavior, 31*, 50-96,2000.

[10] Marsili Matteo.,"Toy Models of Markets with Heterogeneous Interacting Agents", Lecture Notes in Economics and Mathematical Systems, Vol.503, pp161-181, Springer, 2001

[11] Namatame, A,and Sato, H., "Collective Learning in Social Coordination Games", Proc. of AAAI Spring Symposium, SS-01-03,pp.156-159, 2001.

[12] Samuelson,,L., *Evolutionary Games and Equilibrium Selection*, The MIT Press, 1998.

[13] Sipper M., *Evolution of Parallel Cellular Machines*: The Cellular Programming Approach, Springer, 1996.

[14] Young, H.P., *Individual Strategy and Social Structure*, Princedom Univ. Press, 1998.

[15] Weibull J., *Evolutionary Game Theory*, The MIT Press, 1996.