Simulator and Some Applications for Potential Game Theoretic Control

Yasuaki WASA
FL 11_10_01
8, July, 2011

(Recap.) Potential Game and Objective

Game Theoretic Approach

- Agents are "self-interested" ⇒ Non-Cooperative Game
- The solution to the problem = the equilibrium of the game

Advantages

- Robustness to failures and environmental disturbances
- Scalability and adaptability in real time

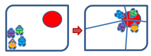


Fig.1 Coverage

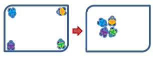


Fig.2 Consensus

Potential Game

- Nash Equilibrium (NE) exists
- Design: 1. Utility Design (ϕ, U_i) 2. Learning Algorithm (decision-making rule)

Objective of this work

- To verify the validity of the theoretic result [1] by way of experiment
- To consider the applicability of PIPIP to other scenarios
- To consider the applicability to an environmental change

[1] 後藤, "ポテンシャルゲーム理論的協調制御における学習アルゴリズムの提案," 東京工業大学修士論文, 2011.

Outline

- Potential Game and Objective
- Simulator : CCPGS
  - Structure / Elements
  - Let's try !!
- Simulation / Experiment
  - Coverage
  - Consensus
  - Attitude Coordination
- Summary

Motivation (Simulator)

Motivation

- many proposals for the Learning Algorithms on Potential Game

But… ⇒ few applications with them

Therefore… ⇒ Really, are their Learning Algorithms useful?

- environmental change
- scalability
- field size
- detailed settings

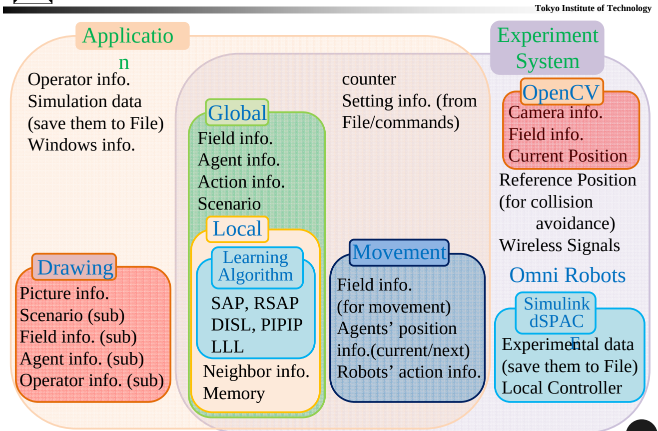
So… ⇒ Need a simulator we can easily verify them with.

CCPGS
(Cooperative Control and Potential Game Simulator)

Input: Mouse only (intuitive operation) NEW LOAD

⇒ Apply the application to operation with using "touch panel" ?!

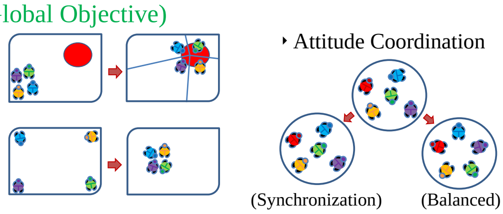
Class Overview



Application on CCPG Simulator

Scenario (Global Objective)

- Coverage
- Consensus
- Attitude Coordination



Learning Algorithm (Local Controller)

Finite Memory

Virtual Payoff-based

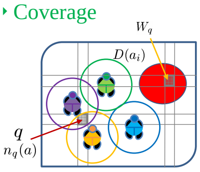
- (Restrictive) Spatial Adaptive Play (SAP/RSAP)

Payoff-based

- Payoff-based Log-Linear Learning (LLL)
- Distributed Inhomogeneous Synchronous Learning (DISL)
- Payoff-based Inhomogeneous Partially Irrational Play (PIPIP)

Utility Design

	Potential Function	Utility Function
Coverage	$\phi(a) = \sum_{q \in Q} \sum_{l=1}^{n_q(a)} \frac{W_q}{l}$	$U_i(a) = \sum_{q \in D(a_i) \cap Q} \frac{W_q}{n_q(a)}$



$q \in Q$: Mission Space and its Point
W_q : Reward (density) of a point q
$D(a_i)$: the agent i's sensing area
$n_q(a)$: The number of agents who can sense a point q

- Consensus [2] $\phi(a) = -\sum_{i \in V} \sum_{j \in N_i} \frac{\|a_i - a_j\|}{2}$ $U_i(a) = -\sum_{j \in N_i} \|a_i - a_j\|$
- Attitude Coordination (Circular Formation) [5]

$$W_1(a) = \frac{1}{2N} \sum_{i=1}^{N} \sum_{j \in N_i} (1 - \cos(a_i - a_j)) \quad W_2(a) = \frac{1}{N} \sum_{j \in N_i} \cos(a_i - a_j)$$

(Synchronization) $\phi_s(a) = -W_1(a)$ $U_i(a) = W_2(a)$

(Balanced) $\phi_b(a) = W_1(a)$ $U_i(a) = -W_2(a)$

Learning Algorithms

LA.	Information	1trial	Target Area	Features
SAP [2]	Finite memory Virtual Payoff-based	1 agent	All actions	Opt. NE (high prob.) Heavy Calculation Relatively [bad] no reality
RSAP [2]	Finite memory Virtual Payoff-based	1 agent	Restricted (neighbor)	Opt. NE (high prob.) [good] small field [good] no environmental info.
DISL [3]	Finite memory Payoff-based	All agents	Restricted (neighbor)	NE / Opt. NE (not always) [good] large field [good] coverage / plain field
PIPIP [1]	Finite memory Payoff-based	All agents	Restricted (neighbor)	Irrational Action Opt. NE (prob. Conv.) [good] large field [good] coverage / obstacle field
LLL [5]	Finite memory Payoff-based	All agents	All actions	Opt. NE (prob. Conv.) [bad] no reality

Simulation Settings

Simulation Parameters

Agent Relations
- The number of Agents
- Initial Positions
- Connection

[option] Restricted Action Set
[option] Sensing Area

Game Design Relations
- each exploration parameter

[option] scaling

Environmental Relations
- Field Size
- Obstacles

[option] Density Distribution

Measurement Relations
- movement style
- maximum trial step
- drawing span

[option] the amount of skips

Agent's behavior ⇒
- Exploration rate (parameter) β, ε
- Scaling of agent's utility functions

Outline

- Potential Game and Objective
- Simulator : CCPGS
  - Structure / Elements
  - Let's try !!
- Simulation / Experiment
  - Coverage
  - Consensus
  - Attitude Coordination
- Summary

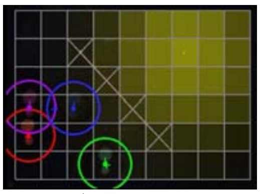
(Recap.) Coverage : Settings

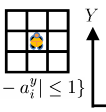
Area $Q = \{1, \cdots, 9\} \times \{1, \cdots, 6\}$

Sensing Area $D(a_i) = \{q \in Q \mid \|q - a_i\|_2 \le 1\}$

Obstacle $\mathcal{O} = \{(x, y) \mid x + y = 8,\ 3 \le x \le 6\}$

(substantially non-convex)





Action Set $\mathcal{A}_i = \{Q \backslash \mathcal{O}\}$

Restricted Action Set

$R_i(a_i) = \{(x, y) \mid |x - a_i^x| \le 1 \wedge |y - a_i^y| \le 1\}$

Agents $N = 4$

Initial Position (1,1), (1,2), (2,1), (2,2)

Skips (Experiment mode) 40 steps

Maximum Steps

(Simulation) 5000 steps (Experiment)2000 steps

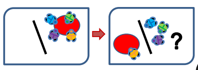
Reward (Gaussian Distribution)

Learning Algorithm DISL, PIPIP

$\varepsilon = 0.15, k = 0.5$

Potential Function $\phi(a) = \sum_{q \in Q} \sum_{l=1}^{n_q(a)} \frac{W_q}{l}$

Utility Function $U_i(a) = \sum_{q \in D(a_i) \cap Q} \frac{W_q}{n_q(a)}$



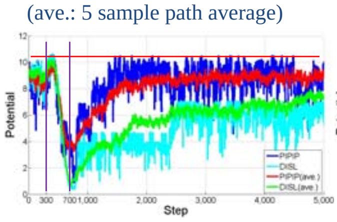
Coverage : case 1 (movable density)

Density Distribution

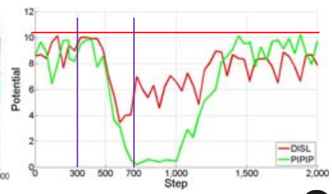
$$W_q(q) = \exp\left\{\frac{(q_x - c_x)^2 + (q_y - c_y)^2}{4}\right\}$$

$$(c_x, c_y) = \begin{cases} (2, 2) & \text{if} \quad k \in [0, 300] \\ (2\frac{700-k}{400} + 7\frac{k-300}{400}, 2\frac{700-k}{400} + 5\frac{k-300}{400}) & \text{if} \quad k \in (300, 700) \\ (7, 5) & \text{if} \quad k \in [700, \infty) \end{cases}$$

Simulation
(ave.: 5 sample path average)



Experiment



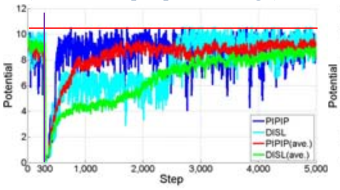
Coverage : case 2 (teleport density)

Density Distribution

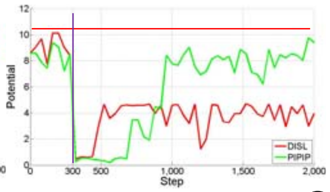
$$W_q(q) = \exp\left\{\frac{(q_x - c_x)^2 + (q_y - c_y)^2}{4}\right\}$$

$$(c_x, c_y) = \begin{cases} (2,2) & \text{if} \quad k \in [0, 300) \\ (7,5) & \text{if} \quad k \in [300, \infty) \end{cases}$$

Simulation
(ave.: 5 sample path average)

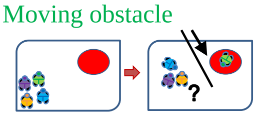


Experiment



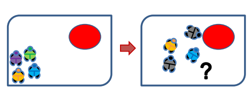
Coverage : Other cases

Moving obstacle



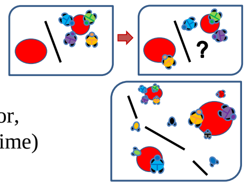
Setting is hard
(collision avoidance between agent and obstacles)

Out of order



Setting is hard
1. Who, and When or Where?
2. Sensing is available?
3. Same as obstacle?

Complex density distribution
Non-convex obstacle
Huge field



Their settings are possible with Simulator, but they are omitted. (Due to Seminar Time)

Consensus : Settings

Potential Function $\phi(a) = -\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \frac{\|a_i - a_j\|}{2}$

Utility Function $U_i(a) = -\sum_{j \in \mathcal{N}_i} \|a_i - a_j\|$

Experiment Real measurement

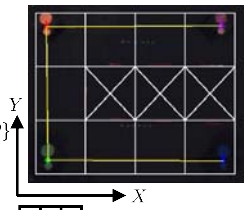
Area $Q = \{1, \cdots, 4\} \times \{1, \cdots, 3\}$

Obstacle $\mathcal{O} = \{(x, 2) | 2 \le x \le 4\}$

Agents $N = 4$

Action Set case1 $\mathcal{A}_i = \{Q\}$ case2 $\mathcal{A}_i = \{Q \backslash \mathcal{O}\}$

Experimental Field
(Initial Positions)
Communication $\mathcal{N}_i$

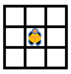


Simulation

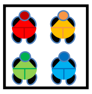
Learning Algorithm RSAP, PIPIP

Restricted Action Set

$R_i(a_i) = \{(x, y) | |x - a_i^x| \le 1 \wedge |y - a_i^y| \le 1\}$



Stay Position

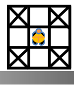


Experiment (for movement model)

Learning Algorithm RSAP

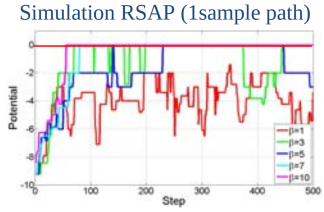
Restricted Action Set

$R_i(a_i) = \{(x, y) | |x - a_i^x| + |y - a_i^y| \le 1\}$

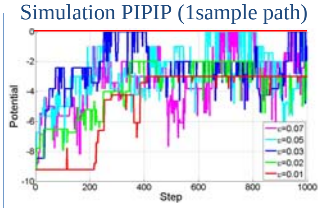


Consensus: case1

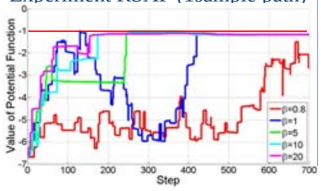
Simulation RSAP (1sample path)



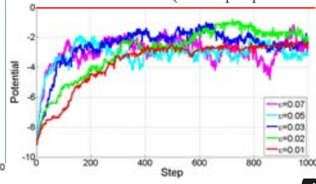
Simulation PIPIP (1sample path)



Experiment RSAP (1sample path)

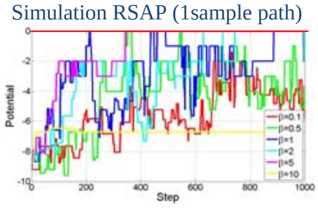


Simulation PIPIP (10sample paths ave.)

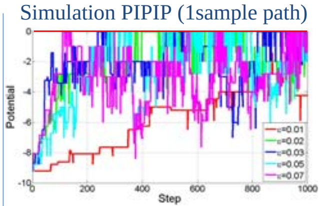


Consensus: case2

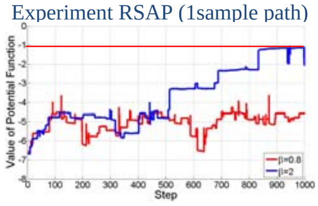
Simulation RSAP (1sample path)



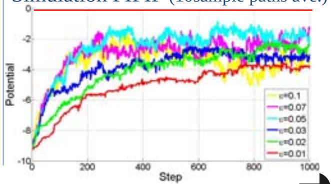
Simulation PIPIP (1sample path)



Experiment RSAP (1sample path)



Simulation PIPIP (10sample paths ave.)



Outline

- Potential Game and Objective
- Simulator : CCPGS
  - Structure / Elements
  - Let's try !!
- Simulation / Experiment
  - Coverage
  - Consensus
  - Attitude Coordination
- Summary

Summary

Objective of this work

- To verify the validity of the theoretic result [1] by way of experiment
- To consider the applicability of PIPIP to other scenarios
- To consider the applicability to an environmental change

Summary

- Compose Movement Model and Simulator
- PIPIP is very useful under obstacles and environmental changes
- SAP is useful under no environmental information, but no reality
- RSAP is useful under no environmental information, but limitation
- DISL is useful under no obstacle
- LLL is tricky

Future Work

- Vision-based Coverage Problem and its experiment
- Human-Robotic Decision-making with Payoff-based LA

[1] 後藤, "ポテンシャルゲーム理論的協調制御における学習アルゴリズムの提案," 東京工業大学修士論文, 2011.

Reference

[1] 後藤, "ポテンシャルゲーム理論的協調制御における学習アルゴリズムの提案," 東京工業大学修士論文, 2011.

[2] J. R. Marden, G. Arslan and J. S. Shamma, "Cooperative Control and Potential Games," *IEEE Transacions on Systems, Man and Cybernetics*, Vol. 39, No. 6, pp. 1393-1407, 2009.

[3] M. Zhu and S. Martinez, "Distributed Coverage Games for Mobile Visual Sensors (i): Reaching the Set of Nash Equilibria," *Proc. of the 48th IEEE Conference. on Decision and Control and 28th Chinese Control Conference*, pp. 169-174, 2009.

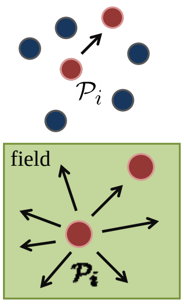
[4] J. R. Mardenand and J. S. Shamma, "Revisiting Log-linear Learning: Asynchrony, Completeness and Payoff-based Implementation," *Games and Economic Behavior*, 2008.

[5]畑中, 後藤, 藤田, "ポテンシャルゲーム理論的姿勢協調:同期・平衡の達成," システム制御情報学会論文誌, Vol. 24, No. 7, 2011 (to appear).

APPENDIX

- Learning Algorithms (SAP/LLL)
- Convergence

Spatial Adaptive Play (SAP)



Step1 Randomly choose one player $\mathcal{P}_i$

Step2 $\mathcal{P}_i$ selects one action a_i

$p_i(t) \in \Delta(\mathcal{A}_i)$: Set of probability distribution

$$p_i^{a_i}(t) = \frac{\exp\{\beta U_i(a_i, a_{-i}(t-1))\}}{\sum_{\bar{a}_i \in \mathcal{A}_i} \exp\{\beta U_i(\bar{a}_i, a_{-i}(t-1))\}}$$

($\beta > 0$: exploit parameter)

Assumption (Reversibility) $\forall i \in \mathcal{V}\ \ \forall a_i^1, a_i^2 \in \mathcal{A}_i$
$a_i^2 \in R_i(a_i^1) \Leftrightarrow a_i^1 \in R_i(a_i^2)$

Assumption (Feasibility) $\forall i \in \mathcal{V}\ \ \forall a_i^{k-1}, a_i^k \in \mathcal{A}_i$
$a_i^k \in R_i(a_i^{k-1})$

Proposition The unique stationary distribution

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in \mathcal{A}} \exp\{\beta\phi(\bar{a})\}}$$

Opt. NE (high prob.)

Problem Mobility Limitation

Payoff-based Log-Linear Learning (LLL) [4]

I.C. $a_i(0),\ \forall i,\ U_i(a(0)),\ a(1) = a(0),\ U_i(a(1)) = U_i(a(0)),\ x_i(1) = 0$

Action Selection a_i^{tp} $(t \geq 2)$

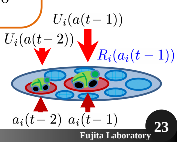
case1 $x_i(t-1) = 0$ (experimented in period t) $\varepsilon \in (0, 1)$

probability	selection	experimental flag
ε	$a_i^{tp} \in \mathcal{A}_i$ (randomly,	$x_i(t) = 1$
$1-\varepsilon$	$a_i^{tp} = a_i(t-1)$ormly)	$x_i(t) = 0$

case2 $x_i(t-1) = 1$ (not experimented in period t) $\beta > 0$: exploit parameter

probability	selection	experimental flag
$\frac{e^{\beta U_i(a(t-2))}}{e^{\beta U_i(a(t-1))} + e^{\beta U_i(a(t-2))}}$	$a_i^{tp} = a_i(t-1)$	$x_i(t) = 0$
$\frac{e^{\beta U_i(a(t-1))}}{e^{\beta U_i(a(t-1))} + e^{\beta U_i(a(t-2))}}$	$a_i^{tp} = a_i(t-2)$	$x_i(t) = 0$

Theorem $\varepsilon = (e^{-\beta})^m$, m : sufficiently large
The stochastically stable states are contained in the set of potential maximizers. $\lim_{t \to \infty} P(z(t) \in \text{diag}(\zeta(\Gamma))) = 1$



Convergence

The (weak) law of large numbers $\bar{X}_n$ converge in probability on μ

X_i :independent $\bar{X}_n$:average
$\mu = E(X_i), \sigma_i^2 = V(X_i) \leq \sigma^2$

$\forall \varepsilon > 0,\ P(|\bar{X}_n - \mu| < \varepsilon) \to 1\ (n \to \infty)$

Interpretation: If a lot of samples are taken, the average of samples may be regarded as the true (expected) value

Proof: $E(\bar{X}_n) = \mu, V(\bar{X}_n) = \frac{\sigma^2}{n}$ ➡ $P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2}$ ➡ 0 ($n \to \infty$)

cf. Chebyshev's inequality: $P(|X - E(X)| \geq k) \leq \frac{V(X)}{k^2}$

Central Limit Theorem: $Z_n = \frac{X_n - \mu}{\sigma/\sqrt{n}},\ E[X_n] = 0,\ V[X_n] = 1$

$$P(a \leq Z_n \leq b) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\ (n \to \infty)$$

The (strong) law of large numbers

X_i :independent, same distribution $\mu = E(X_i), \sigma^2 = V(X_i), \nu_4 = E(X_i - \mu)^4$

$P(\lim_{n \to \infty} \bar{X}_n = \mu) = 1$ $\bar{X}_n$ converge with probability 1 on μ