被覆ゲームに対する最適均衡解の探索 利得に基づく学習アルゴリズム設計*

和佐 泰明[†]・後藤 達彦[†]・畑中 健志[†]・藤田 政之[†]

Seeking Optimal Equilibria for Coverage Games —Payoff-based Learning Approach*

Yasuaki WASA[†], Tatsuhiko GOTO[†], Takeshi HATANAKA[†] and Masayuki FUJITA[†]

In this paper, we investigate the game theoretic coverage control whose objective is to lead agents to optimal configurations over a mission space. In particular, the objective of this paper is to achieve the control objective (i) in the absence of the perfect prior knowledge on importance of each point and (ii) in the presence of the action constraints. For this purpose, we first formulate coverage problems with two different global objective functions as so-called potential games. Then, we present a payoff-based learning algorithm determining actions based only on the past actual outcomes. The feature of the present algorithm is to allow an agent to take an irrational action. We also clarify a relation between a design parameter of the algorithm and the probability which agents take the optimal actions and prove that the probability can be arbitrarily increased. Then, we demonstrate the effectiveness of the present algorithm through experiments on a testbed.

1. はじめに

放射性物質の拡散や汚染地域の特定, 地球温暖化など現在の環境問題において, 効率的なモニタリング技術の必要性が高まっている. これらの課題の計測範囲は広域にわたるため, 無線通信機能をもつ複数の小型センサを活用したセンサネットワーク [1] が有用である. さらに, 計測対象が時々刻々変化する場合や放射線計測などの計測自体に危険が伴う場面においては, 個々のセンサに移動能力を課すモバイルセンサネットワーク [2] の利用が望まれる.

モバイルセンサネットワークの最適配置問題は被覆制御問題とよばれ, 代表的な協調制御問題としてこれまで数多くの制御法が提案されてきた [3–8]. 一方, 上記の現実の社会問題には, (i) 計測点の重要度に関する先見情報がない, あるいは不足している, (ii) 障害物の存在や移動体の性能に起因する移動制限が存在する, という課題があるが, 従来手法はこれらの問題に対処する方法論を与えているとはいえない. たとえば, 文献 [3–8] は

すべて各地点の重要度が密度関数とよばれる関数によって事前に与えられることを前提としており, (i) のシナリオは想定されていない. また, (ii) の課題は文献 [7] などにおいて部分的に考慮されているが, 系統的方法論を与えるものとはいえない.

他方, 協調制御の分野では, 問題を非協力ゲームとして捉え, 特定の Nash 均衡解への収束を制御目的の達成と同一視するゲーム理論的アプローチが注目されている [9–11]. 文献 [9] では, 多くの代表的な協調制御問題とともに被覆制御問題がポテンシャルゲーム [12,13] とよばれるゲームに帰着できることを示している. また, 文献 [10,11] においても文献 [9] とは異なる被覆ポテンシャルゲームが提案されている.

さらに, 文献 [9] は行動制限を考慮したエージェントの行動決定アルゴリズム (以下, 学習アルゴリズム) を提案しており, これは被覆制御の文脈において課題 (ii) を解決する. しかし, 文献 [9] の学習アルゴリズムは行動決定前に行動選択後の利得関数値が計算できることを仮定しており, これは被覆制御において密度関数の事前知識を前提とすることにほかならず, 課題 (i) を考慮していない. 一方, 文献 [11,14] では過去の利得の履歴のみを基準に行動選択を行う “利得に基づく学習アルゴリズム” が提案されており, これは課題 (i) を解決する. しかし, 文献 [11] はある均衡解への収束を保証するが, それが全

* 原稿受付 2012 年 3 月 12 日

[†] 東京工業大学 大学院 理工学研究科 Graduate School of Science and Engineering, Tokyo Institute of Technology; 2-12-1, Ookayama, Meguro-ku, Tokyo 152-8550, JAPAN

Key Words: coverage control, potential game.

Fig. 1 Coverage control: Field and sensing (The darker: higher density)

体最適であることは保証されない．他方，文献 [14] は十分な時間の経過後に高確率で全体最適を達成するが，課題 (ii) を考慮していない．

そこで本論文では，文献 [11] の枠組みに基づいて，課題 (i), (ii) を同時に解決し，かつ重要度を加味した最適配置にセンサを誘導する，利得に基づく学習アルゴリズムを提案する．具体的には，文献 [11] の学習アルゴリズムに対して過去の履歴において低い利得を得た行動を選択する非合理的決定の確率を設けた学習アルゴリズム Payoff-based Inhomogeneous Partially Irrational Play (PIPIP) を提案する．つぎに，提案アルゴリズムを適用したときの設計パラメータと最適均衡解を選択する確率の関係を明らかにし，パラメータ値次第ではこの確率を任意に高められることを証明する．また，従来研究では行われていない被覆制御問題の実験的検証を通して提案アルゴリズムの有効性を検証する．

2. 問題設定

本節で示す問題設定は文献 [11] に準拠する．

本論文では，2次元空間に存在する n 台のエージェント $\mathcal{V} := \{1, \dots, n\}$ から構成されるロボティックネットワークを考える．エージェントは有限離散の対象領域 $\mathcal{Q} = \{1, \dots, m_x\} \times \{1, \dots, m_y\}$, $m_x, m_y \in \mathbb{N}$ 間を移動する．ここで，エージェント $i \in \mathcal{V}$ の位置を $p_i \in \mathcal{Q}$ と表記する．対象領域内の地点 $q \in \mathcal{Q}$ はそれぞれ計測重要度 $W_q \in \mathbb{R}_+$ をもつ．エージェントは重要度 W_q に関する事前知識を有しておらず，対象地点 $q \in \mathcal{Q}$ を計測したときに限り，その地点の重要度 W_q をフィードバックできるものとする (Fig. 1)．

各エージェントの位置推移は運動学モデル (1) に従うものとする．

$$p_i(k+1) = a_i(k), \quad a_i(k) := p_i(k) + u_i(k) \quad (1)$$

ここで， $u_i(k)$ はエージェント i に対する速度入力であり，ハードウェア制約や障害物の存在ゆえに生じる可動範囲の制限を考慮して一定範囲の入力値とする．また，

Fig. 2 Sensing and communication

時刻 $k \in \mathbb{Z}_+$ の位置 $p_i(k)$ から入力 $u_i(k)$ で移動可能な範囲を $\mathcal{R}_i(p_i(k))$ と表記する．エージェント i は位置 p_i から半径 r_m の円内に含まれる対象領域 $q \in \mathcal{Q}$ の環境情報を計測できるとし，計測対象範囲 $\mathcal{D}(p_i)$ を次式で定義する．

$$\mathcal{D}(p_i) = \{q \in \mathcal{Q} : \|q - p_i\|_2 \leq r_m\} \quad (2)$$

このとき，対象地点 $q \in \mathcal{Q}$ を計測しているエージェントが複数存在する場合，互いに通信を必要とする．したがって，エージェント $i \in \mathcal{V}$ は半径 $2r_m$ の円内に含まれる他のエージェント $j \in \mathcal{V}$ と互いに通信可能とする (Fig. 2)．すなわち，エージェント $i \in \mathcal{V}$ が通信可能な近傍 $\mathcal{N}_i$ は次式で与えられる．

$$\mathcal{N}_i := \{j \in \mathcal{V} : \|p_j - p_i\|_2 \leq 2r_m\} \quad (3)$$

つぎに，全体の目的関数を定義する．エージェントの位置 p_i を集めて p と表記し，対象地点 $q \in \mathcal{Q}$ を観測しているエージェント数を $n_q(p)$ とする．本論文では計測重要度を利益と捉え，二つの制御目的表現を定式化する．まず，対象地点 $q \in \mathcal{Q}$ を観測することで得られる利益 W_q が観測しているエージェント数 $n_q(p)$ に応じて増加する場合，全体の目的関数は，たとえば (4) 式で表現できる．

$$W^1(p) = \sum_{q \in \mathcal{Q}} \sum_{l=1}^{n_q(p)} \frac{W_q}{l} \quad (4)$$

一方，対象地点 $q \in \mathcal{Q}$ を観測する利益 W_q がエージェント数 $n_q(p)$ には関係しない場合，全体の目的関数は (5) 式で表現できる．

$$W^2(p) = \sum_{q \in \mathcal{Q} : n_q(p) \neq 0} W_q \quad (5)$$

本論文では $a_i(k) \in \mathcal{Q}$ を設計対象とし (これを“行動”とよぶ)，全体の目的関数 (4), (5) を最大化するようにこれを決定する問題を考える．また，集合 $\mathcal{A}_i \subseteq \mathcal{Q}$ は行動集合とよばれ，すべてのエージェントの行動集合を集めて $\mathcal{A} := \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ と定義する．さらに，すべてのエージェントの行動をまとめて $a = (a_1, \dots, a_n) \in \mathcal{A}$ と表す． i 以外の行動群を $a_{-i} := (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$ と表記するとき，行動群 a は $a = (a_i, a_{-i})$ と表現する

Fig. 3 Payoff and global objective (Assumption 1)

ことができる。

エージェント $i \in \mathcal{V}$ は時刻 k において、移動可能な範囲 $\mathcal{R}_i: \mathcal{A}_i \rightarrow 2^{\mathcal{A}_i}$ の中から行動 $a_i(k)$ を選択する。ここで、集合 $\mathcal{R}_i(\cdot)$ は以下の仮定を満たすとする。

(仮定 1) [9] 集合 $\mathcal{R}_i(\cdot)$ は以下の条件を満たす。

可逆性：任意の $i \in \mathcal{V}$ および $a_i^1, a_i^2 \in \mathcal{A}_i$ に対して $a_i^2 \in \mathcal{R}_i(a_i^1) \leftrightarrow a_i^1 \in \mathcal{R}_i(a_i^2)$ が成立する。

可解性：任意の $i \in \mathcal{V}$ および $a_i^1, a_i^m \in \mathcal{A}_i$ に対して、すべての $l \in \{2, \dots, m\}$ において $a_i^l \in \mathcal{R}_i(a_i^{l-1})$ を満たす列 $(a_i^1, \dots, a_i^m)$ が存在する。

可逆性はある時刻に行動を変化させた場合、それ以降の時刻において元の行動に戻ることができることを意味する。可解性は行動集合 $\mathcal{A}_i$ 内のいかなる行動にも数回の行動変化を繰り返せば到達できることを意味する。

以上より、各エージェントがもつ、利得関数とよばれる局所的な目的関数 $U_i: \mathcal{A} \rightarrow \mathbb{R}$ を定義すれば、被覆制御問題はエージェント群 $\mathcal{V}$ 、行動群集合 $\mathcal{A}$ 、利得関数群 $\{U_i(\cdot)\}_{i \in \mathcal{V}}$ 、拘束行動集合群 $\{\mathcal{R}_i(\cdot)\}_{i \in \mathcal{V}}$ で構成される拘束付戦略型ゲーム $\Gamma = (\mathcal{V}, \mathcal{A}, \{U_i(\cdot)\}_{i \in \mathcal{V}}, \{\mathcal{R}_i(\cdot)\}_{i \in \mathcal{V}})$ に帰着できる。とくに、全体の目的関数が (4), (5) 式で明示されているため、被覆ゲーム Γ は利益 W_q を加えた社会厚生ゲーム $\Gamma_w = (\Gamma, W_q)$ なるクラスに属する [15]。

3. 利得関数設計

本論文では、次に定義されるゲームに着目する。

【定義 1】 (拘束付ポテンシャルゲーム [9,11]) 被覆ゲーム Γ を考える。このとき、ある関数 ϕ が存在して任意の $i \in \mathcal{V}$, $a_i' \in \mathcal{A}_i$, $a_i'' \in \mathcal{R}_i(a_i')$ に対して

$$U_i(a_i'', a_{-i}) - U_i(a_i', a_{-i}) = \phi(a_i'', a_{-i}) - \phi(a_i', a_{-i}) \quad (6)$$

を満たすとき、 Γ は拘束付ポテンシャルゲームを構成する。

本論文では、ポテンシャル関数 ϕ が全体の目的 (4), (5) と一致するように利得関数を設計する [15]。このとき定義 1 は、唯一行動変化したエージェントの利得関数の増分と全体の目的関数の増分が一致することを意味する。

まず、対象地点 $q \in \mathcal{Q}$ が生み出す利益 W_q はその地点を観測しているエージェントに均等配分されるとする。このとき、エージェント i が地点 q を観測している場合、エージェント i が地点 q から得る利益は $W_q/n_q(a)$ となる。したがって、エージェント i が計測範囲 $\mathcal{D}(a_i)$ 内から得る利益は (7) 式で表現される (Fig. 3)。

Fig. 4 Payoff and global objective (Assumption 2)

$$U_i^1(a) = \sum_{q \in \mathcal{D}(a_i) \cap \mathcal{Q}} \frac{W_q}{n_q(a)} \quad (7)$$

このとき命題 1 が示される。

【命題 1】 [11] 利得関数を $U_i = U_i^1$ とおくと、被覆ゲーム Γ は $\phi = W^1$ に対するポテンシャルゲームを構成する。

一方、対象地点 $q \in \mathcal{Q}$ の利益 W_q はその地点を観測しているエージェントが一台であるときのみ、そのエージェントに利益 W_q が配分されるとする。このとき、エージェント i が計測範囲 $\mathcal{D}(a_i)$ 内から得る利益は (8) 式で表現される (Fig. 4)。

$$U_i^2(a) = \sum_{q \in \mathcal{D}(a_i): n_q(a)=1} W_q \quad (8)$$

このとき命題 2 が示される。

【命題 2】 利得関数を $U_i = U_i^2$ とおくと、被覆ゲーム Γ は $\phi = W^2$ に対するポテンシャルゲームを構成する。

(証明) $n_q(a_i', a_{-i}) = 1$ かつ $n_q(a_i'', a_{-i}) = 0$ を満たす対象地点の集合を $\mathcal{Q}_1$, $n_q(a_i', a_{-i}) = 0$ かつ $n_q(a_i'', a_{-i}) = 1$ を満たす集合を $\mathcal{Q}_2$ とおく。このとき行動を a_i' から a_i'' に変化したときのポテンシャル関数の変化は

$$\begin{aligned} \phi(a_i'', a_{-i}) - \phi(a_i', a_{-i}) &= - \sum_{q \in \mathcal{Q}_1} W_q + \sum_{q \in \mathcal{Q}_2} W_q \\ &= U_i(a_i'', a_{-i}) - U_i(a_i', a_{-i}) \end{aligned}$$

となる。定義 1 より命題が示される。

(注意 1) (7) 式の構成は Equally Shared Utility (Shapley Value Utility), (8) 式の構成は Wonderful Life Utility とよばれる形式である [15]。

(注意 2) 利得関数 U_i^1, U_i^2 とともに、計測地点 $q \in \mathcal{D}(a_i)$ に対応する重要度 W_q および観測しているエージェント数 $n_q(a)$ がフィードバックできれば計算できる。ここで、 $n_q(a)$, $q \in \mathcal{D}(a_i)$ は $\mathcal{N}_i$ 内のエージェント $j \in \mathcal{V}$ からその位置 a_j を通信できれば取得可能である。したがって、利得関数 U_i^1, U_i^2 は分散計算可能である。

基本的に、各エージェントは自身の利得関数 U_i を最大化するように行動 $a_i \in \mathcal{A}_i$ を選択する。ここで、ゲーム理論の目的は与えられた利得関数に対して以下の Nash 均衡解 a^* を求めることである。

【定義 2】 (拘束付 Nash 均衡解 [9]) すべてのエージェント $i \in \mathcal{V}$ に対して

$$U_i(a_i^*, a_{-i}^*) = \max_{a_i \in \mathcal{R}_i(a_i^*)} U_i(a_i, a_{-i}^*) \quad (9)$$

を満たすとき、行動群 a^* は拘束付 Nash 均衡解である。すなわち、Nash 均衡解 a^* からエージェント i のみが行動を変化しても利得を増加させることができない。ポテンシャルゲームは Nash 均衡解が必ず存在する点で扱いやすいクラスの問題である [12]。とくに、ポテンシャル関数 ϕ の最大点は Nash 均衡解である。

これまで、ポテンシャルゲームの Nash 均衡解の計算アルゴリズムが数多く提案されている [9, 11, 14]。しかし、被覆ゲーム Γ の目的は全体の目的であるポテンシャル関数 ϕ を最大化する特定の Nash 均衡解への到達である。そこで、最終的にポテンシャル関数を最大化するように行動 a_i を決定する学習アルゴリズムを設計する必要がある。

4. 学習アルゴリズムの設計

本節では、(i) 計測点の重要度に関する先見情報を用いず、(ii) 移動制限のもとで全体の目的を達成する学習アルゴリズムを提案する。文献 [9] は (i) を満たさず、文献 [14] は (ii) を満たさない。文献 [11] の学習アルゴリズム Distributed Inhomogeneous Synchronous Learning (DISL) は条件 (i)、(ii) を満たし、Nash 均衡解に確率収束するが、全体の目的を達成する特定の Nash 均衡解に収束することは保証されていない。そこで本論文では、すべての課題を達成するために、DISL を基調とした学習アルゴリズム Payoff-based Inhomogeneous Partially Irrational Play (PIPIP) を提案する。

各エージェント $i \in \mathcal{V}$ は時刻 k において、直近の過去 2 回の自分の行動 $a_i(k-2)$ 、 $a_i(k-1)$ とそのときに得られた利得 $U_i(a(k-2))$ 、 $U_i(a(k-1))$ だけを記憶しているものとする。また、パラメータ κ は

$$\kappa \in \left(\frac{1}{C-1}, \frac{1}{2} \right], \quad C := \max_{i \in \mathcal{V}} \max_{a_i \in \mathcal{A}_i} |\mathcal{R}_i(a_i)| \quad (10)$$

を満たすものとする。ここで、 $|\mathcal{R}_i(a_i)|$ は $\mathcal{R}_i(a_i)$ の要素数である。このとき、エージェントは一定の探索度 $\varepsilon(k) \in (0, 1)$ を用いて、つぎの規則に従って $\mathcal{R}_i(a_i(k-1))$ の中から $a_i(k)$ を決定する。

過去 2 回の利得 $U_i(a(k-2))$ 、 $U_i(a(k-1))$ を比較し、 $\Delta_i := U_i(a(k-2)) - U_i(a(k-1)) \leq 0$ の場合、

- 確率 ε で $a_i(k-1)$ 以外からランダム選択
- 確率 $1-\varepsilon$ で $a_i(k) = a_i(k-1)$ 。

一方 $\Delta_i > 0$ の場合、

- 確率 ε で $a_i(k-1)$ 、 $a_i(k-2)$ 以外からランダム選択
- 確率 $(1-\varepsilon)\kappa\varepsilon^{\Delta_i}$ で $a_i(k) = a_i(k-1)$
- 確率 $(1-\varepsilon)(1-\kappa\varepsilon^{\Delta_i})$ で $a_i(k) = a_i(k-2)$ 。

このとき、探索行動集合 $\mathcal{R}_i(\cdot) \setminus \{\cdot\}$ が空集合である場合は探索行動を除いた知識利用だけを用いて選択する。すなわち、提案アルゴリズム PIPIP は一定確率で探索し、

Algorithm 1: Payoff-based Inhomogeneous Partially Irrational Play (PIPIP)

Initialization: Action a is chosen randomly from $\mathcal{A}$. Set $a_i^1 \leftarrow a_i$, $a_{-i}^1 \leftarrow a_{-i}$, $U_i^1 \leftarrow U_i(a)$, $U_{-i}^1 \leftarrow U_{-i}(a)$, $\Delta_i \leftarrow 0$ for all $i \in \mathcal{V}$ and $k \leftarrow 2$.

Step 1: Update $\varepsilon(k)$.

Step 2: If $U_i^1 \geq U_{-i}^1$, then set (i) $\mathcal{R}_i(a_i^1) \setminus \{a_i^1\} \neq \emptyset$;

$$a_i^{tmp} \leftarrow \begin{cases} \text{rnd}(\mathcal{R}_i(a_i^1) \setminus \{a_i^1\}), & \text{w.p. } \varepsilon \\ a_i^1, & \text{w.p. } 1-\varepsilon \end{cases},$$

(ii) otherwise; $a_i^{tmp} \leftarrow a_i^1$.

If $U_i^1 < U_{-i}^1$, then set (i) $\mathcal{R}_i(a_i^1) \setminus \{a_i^1, a_{-i}^1\} \neq \emptyset$;

$$a_i^{tmp} \leftarrow \begin{cases} \text{rnd}(\mathcal{R}_i(a_i^1) \setminus \{a_i^1, a_{-i}^1\}), & \text{w.p. } \varepsilon \\ a_i^1, & \text{w.p. } (1-\varepsilon)(\kappa \cdot \varepsilon^{\Delta_i}) \\ a_{-i}^1, & \text{w.p. } (1-\varepsilon)(1-\kappa \cdot \varepsilon^{\Delta_i}) \end{cases},$$

(ii) otherwise;

$$a_i^{tmp} \leftarrow \begin{cases} a_i^1, & \text{w.p. } \kappa \cdot \varepsilon^{\Delta_i} \\ a_{-i}^1, & \text{w.p. } 1-\kappa \cdot \varepsilon^{\Delta_i} \end{cases}.$$

Step 3: Execute the selected action a_i^{tmp} and receive $U_i^{tmp} \leftarrow U_i(a^{tmp})$.

Step 4: Set $a_i^2 \leftarrow a_i^1$, $a_{-i}^1 \leftarrow a_i^{tmp}$, $U_i^2 \leftarrow U_i^1$, $U_{-i}^1 \leftarrow U_i^{tmp}$ and $\Delta_i \leftarrow U_i^2 - U_i^1$.

Step 5: $k \leftarrow k+1$ and go to Step 1.

残りの確率で直近 2 行動のうち、高い利得を得た行動を優先的に選択する。ただし直近の行動の利得が低い場合、その行動を非合理的選択する可能性も残すという非常に単純な規則から構成される。

以上の行動決定則を整理したものを Algorithm 1 に示す。従来手法である DISL [11] は PIPIP の行動選択の規則において、 $\kappa=0$ の場合に対応する。つまり、DISL は非合理的選択をしないアルゴリズムである。

被覆ゲーム Γ において、全体の目的を達成する特定の Nash 均衡解の集合を $\zeta(\Gamma)$ と定義する。このとき、仮定 1, 2 のもとで定理 1 が導かれる。証明は付録 2. に記す。

(仮定 2) $a'_i \in \mathcal{R}_i(a_i)$, $a_{-i} = a'_{-i}$ となるすべての $a, a' \in \mathcal{A}$ に対して $U_i(a') - U_i(a) < 1$ が成立する。

仮定 2 は全利得をスケールリングすれば満たされる。

【定理 1】 仮定 1, 2 を満たす被覆ゲーム Γ を考える。すべてのエージェントが学習アルゴリズム PIPIP を用いるとき、任意の確率 $p < 1$ に対してある十分小さい探索度 ε が存在し、十分大きな時刻 $k \in \mathbb{Z}_+$ に対して、次式を満たす。

$$\Pr[z(k) \in \text{diag}(\zeta(\Gamma))] > p \quad (11)$$

ただし、 $z(t) := (a(k-1), a(k))$, $\text{diag}(\mathcal{A}') = \{(a, a') \in \mathcal{A} \times \mathcal{A} : a \in \mathcal{A}' \subseteq \mathcal{A}\}$ とする。

定理 1 は文献 [14] と同様の表現形式であり、PIPIP の

Fig. 5 Experimental system

Fig. 6 Experimental field and initial positions (Black rectangle: Obstacle)

探索度が十分小さく、十分時間経過したとき、既知情報だけを利用し、移動制限のもとで最も効率的な行動が高確率で選択されることを保証している。したがって、提案アルゴリズム PIPIP は重要度 W_q を用いずに行動選択し、要求 (i), (ii) がともに満たされる中で全体の目的を達成するアルゴリズムである。

ただし、探索度を大きく設定すれば、全体目的を達成しない非最適 Nash 均衡解に留まる時間は短縮されるが、全体目的を達成する最適 Nash 均衡解から脱する確率も高まる。他方、探索度を小さく設定すれば、最適 Nash 均衡解に留まる確率は高まる一方で、最適均衡解への到達時間は増大するというトレードオフの関係がある [9]。探索度の設計として、徐々に探索度を小さくするように設定する手法が提案されているが [11]、系統的な設計方法は今後の課題である。

5. 学習アルゴリズムの有効性検証

本節では、被覆制御問題の実験的検証を通して提案学習アルゴリズム PIPIP の有効性を検証する。

5.1 実験環境

実験環境の概念図を Fig. 5 に示す。全方向移動ロボットは 22cm 4 車輪ウレタンオムニキット (土佐電子社)、ロボットの位置姿勢取得は 30[fps] の USB カメラ FMVU-03MTC-CS (ViewPLUS 社) を用いる。まず、天井に取り付けた USB カメラから計算機に取り込んだ画像を画像処理ライブラリ OpenCV 2.0 で処理し、特徴点から各

Fig. 7 Time evolution of potential function for $\varepsilon = 0.15$ Fig. 8 Time evolution of potential function for $\varepsilon = 0.30$

ロボットの位置姿勢を抽出する。つぎに、抽出した各ロボットの位置姿勢情報に基づき、学習アルゴリズムを用いた更新則の適用および目標移動位置の更新を判断する。そして、Simulink (The Math Works 社)、DSP ボード DS1104 (dSPACE 社) により各ロボットへの速度制御入力を計算し、各モータの PWM Duty 比を取得する。最後に、各モータの Duty 比を計算機から無線ポート XBee を介して各ロボットに送信し、各ロボットは受信した Duty 比を実行する。計算機上の処理は C 言語で記述し、Visual Studio 2005 (Microsoft 社) 上で動作させる。この操作を約 70[ms] 周期で繰り返し、実験を行う。

台数を $n = 4$ 、領域を $Q = \{1, \dots, 9\} \times \{1, \dots, 6\}$ 、障害物を $O = \{(3,5), (4,4), (5,3), (6,2)\}$ とし、行動集合を $\mathcal{A}_i = Q \setminus O$ とする。また、計測範囲を $\mathcal{D}(a_i) = \{q \in Q : \|q - a_i\|_2 \leq 1\}$ 、拘束行動集合を $\mathcal{R}_i(a_i) = \{a_i + (b_1, b_2) \in \mathcal{A}_i : b_j \in \{-1, 0, 1\}, j = 1, 2\}$ 、初期位置をそれぞれ (1,1), (1,2), (2,1), (2,2) とする (Fig. 6)。さらに、全体の目的関数は (4) 式、利得関数は (7) 式を用いる。

バッテリー消費などの実験可能時間を考慮し、エージェントは α 回に 1 度だけ実際に動作させる。このとき、エージェントは衝突回避を考慮してゲームで導かれた配置に出来るだけ近い配置に移動する。衝突回避移動モデルは本論文の目的である学習アルゴリズムの振舞いに関して本質的ではなく、考察に影響を及ぼさないで、本論文では省略する。

5.2 有効性の検証

複数の Nash 均衡解が存在する環境の中で、有限時間内で全体の目的を達成する最適 Nash 均衡解への到達を検証する。また、全体の目的を達成する Nash 均衡解への収束性を保証していない DISL と比較する。密度分布は

Fig. 9 Configurations by (a)DISL and (b)PIPIP for $\varepsilon = 0.15$ ((a-3): Nash equilibrium, (b-3): Optimal Nash equilibrium)

$$W_q = \exp \left\{ -\frac{\|q - \mu\|^2}{8} \right\}, \quad \mu = (7, 5) \quad (12)$$

とし、実際の動作回数の離散値は $\alpha = 10$ とする．探索度 $\varepsilon = 0.15, 0.30$ ，パラメータ $\kappa = 0.5$ における実験結果を Figs. 7~9 に示す．

Figs. 7, 9 より探索度 $\varepsilon = 0.15$ において、PIPIP は最適 Nash 均衡解 (Fig. 9 (b-3)) に到達しているが、DISL は障害物により構成された非最適 Nash 均衡解 (Fig. 9 (a-3)) に留まり、一定確率で探索行動を選択するにもかかわらず、そこから脱出できない様子がわかる．また、Fig. 8 より、探索度 $\varepsilon = 0.30$ は探索行動を高確率で選択するため、PIPIP は一度最適 Nash 均衡解に到達しているが、一部のエージェントが障害物を再び乗り越える行動を示している．一方、DISL は最適 Nash 均衡解に到達していないことがわかる．したがって、少なくとも本環境設定では PIPIP は DISL より優れており、実用的にも好まれることがわかる．

以上の結果から、提案アルゴリズム PIPIP は有限時間内で最適 Nash 均衡解への到達が示された．また、探索度 ε を大きくするにつれて Nash 均衡解から脱出して、未知の最適 Nash 均衡解に移移する可能性が大きくなることを示唆している．

5.3 環境適応性の検証

つぎに、計測対象が時々刻々変化する場合において、環境変化を知らないエージェントがどのような振舞いをするかを検証する．密度分布は

Fig. 10 Configurations (Adaptability to environmental change)

Fig. 11 Time evolution of potential function (Adaptability to environmental change)

$$W_q = \exp \left\{ -\frac{\|q - \mu(k)\|^2}{4} \right\}, \quad (13)$$

$$\mu(k) = \begin{cases} (2, 2) & \text{if } k \in [0, 300] \\ \left(\frac{k}{80} - \frac{7}{4}, \frac{3k}{400} - \frac{1}{4} \right) & \text{if } k \in (300, 700) \\ (7, 5) & \text{if } k \in [700, \infty) \end{cases}$$

とする．これは密度関数が連続的に変化していることを意味する．本検証では障害物は設置しないこととする ($\mathcal{O} = \emptyset$)．動作頻度 $\alpha = 20$ ，探索度 $\varepsilon = 0.10$ ，パラメータ $\kappa = 0.5$ における実験結果を Figs. 10, 11 に示す．

Fig. 10 よりエージェント群は未知の環境変化を計測し、重要度の高い領域を被覆し続けていることがわかる．また Fig. 11 より、ポテンシャル関数値がその時刻における最適 Nash 均衡解の値に近い値で推移していることがわかる．よって、提案アルゴリズム PIPIP は環境変化に対する追従適応性を有することがわかる．なお、この

ような挙動は密度関数をあらかじめ指定する [3–8] のような手法では実現できないものである。

6. おわりに

本論文では被覆制御問題にポテンシャルゲームを適用した。まず、全体の目的関数を定義し、適切なポテンシャル関数と個々の目的関数を用いることにより、被覆制御問題をポテンシャルゲームとして表現できることを示した。また、現実の問題設定を考慮した移動制限や不完全な環境情報のもとでも、得られた情報だけを用いて全体の目的を達成する特定の均衡解に高確率で選択する学習アルゴリズム PIPIP を提案した。最後に、提案アルゴリズムを被覆問題に適用し、アルゴリズムの有効性を実験的に検証した。

参考文献

- [1] 飯野, 畑中, 藤田: センサネットワークと制御理論; 計測と制御, Vol. 47, No. 8, pp. 649–656 (2008)
- [2] N. E. Leonard, D. A. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni and R. E. Davis: Collective motion, sensor networks and ocean sampling; *Proc. of the IEEE*, Vol. 95, No. 1, pp. 48–74 (2007)
- [3] S. Martinez, J. Cortes and F. Bullo: Motion coordination with distributed information; *IEEE Control Systems Magazine*, Vol. 27, No. 4, pp. 75–88 (2007)
- [4] J. Cortes, S. Martinez, T. Karatas and F. Bullo: Coverage control for mobile sensing networks; *IEEE Trans. on Robotics and Automation*, Vol. 20, No. 2, pp. 243–255 (2004)
- [5] W. Li and C. G. Cassandras: A minimum-power wireless sensor network self-deployment scheme; *Proc. of 2005 IEEE Wireless Communications and Networking Conference*, pp. 2542–2547 (2005)
- [6] C. H. Caicedo-N and M. Zefran: Balancing sensing and coverage in mobile sensor networks: A min-max approach; *Proc. of the 46th IEEE Conference on Decision and Control*, pp. 6124–6129 (2007)
- [7] F. Bullo, J. Cortes and S. Martinez: *Distributed Control of Robotic Networks*, Series in Applied Mathematics, Princeton University Press (2009)
- [8] R. Votal, D. A. W. Barton, T. Goto, T. Hatanaka, M. Fujita and J. Moehlis: Equilibrium configurations for a territorial model; *SIAM Journal on Applied Dynamical Systems*, Vol. 8, No. 3, pp. 1234–1260 (2009)
- [9] J. R. Marden, G. Arslan and J. S. Shamma: Cooperative control and potential games; *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 39, No. 6, pp. 1393–1407 (2009)
- [10] N. Hayashi, T. Ushio and T. Kanazawa: Potential game theoretic approach to power-aware mobile sensor coverage problem; *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E94-A, No. 3, pp. 929–936 (2011)
- [11] M. Zhu and S. Martinez: Distributed coverage games for mobile visual sensor networks; *SIAM Journal on Control and Optimization*, revised (available at arXiv:1002.0367) (2011)
- [12] D. Monderer and L. Shapley: Potential games; *Games and Economic Behavior*, Vol. 14, No. 1, pp. 124–143 (1996)
- [13] 小木曾: 研究活動紹介: ジョージア工科大学 Shamma 研究室における研究活動; システム/制御/情報, Vol. 55, No. 8, pp. 345–346 (2008)
- [14] J. R. Marden and J. S. Shamma: Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation; *Games and Economic Behavior*, accepted (2011)
- [15] J. R. Marden and A. Wierman: Distributed welfare games; *Operations Research*, submitted (2008)
- [16] H. P. Young: *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton University Press (2001)
- [17] T. Goto, T. Hatanaka and M. Fujita: Payoff-based inhomogeneous partially irrational play for potential game theoretic cooperative control of multi-agent systems; (available at arXiv:1107.4838) (2011)

付 録

付録 1. 抵抗木の基礎

有限状態空間 $\mathcal{X}$ 上のマルコフ過程 $\{P_t^0\}$ を考える。摂動的マルコフ連鎖 $\{P_t^\varepsilon\}$, $\varepsilon \in [0, 1]$ は、確率 $1 - \varepsilon$ で斉時的マルコフ連鎖 $\{P_t^0\}$, 確率 ε で $\{P_t^0\}$ に従わないマルコフ連鎖の遷移行列に従うものとする。

【定理 2】 [16] 以下の条件を満たすならば、確率論的過程 $\{P_t^\varepsilon\}$ の族は $\{P_t^0\}$ の標準摂動という。

(R1) $\varepsilon^* > 0$ が存在して、すべての $\varepsilon \in (0, \varepsilon^*]$ に対して、 $\{P_t^\varepsilon\}$ は既約で非周期的である。

(R2) マルコフ過程 $\{P_t^\varepsilon\}$ 上の $x \in \mathcal{X}$ から $y \in \mathcal{X}$ までの遷移確率を P_{xy}^ε と表記する。このとき、すべての $x, y \in \mathcal{X}$ において、 $\lim_{\varepsilon \rightarrow 0} P_{xy}^\varepsilon = P_{xy}^0$ を満たす。

(R3) ある ε が存在して、 $P_{xy}^\varepsilon > 0$ ならば、次式を満たす $\chi(x \rightarrow y) \geq 0$ が存在する。 $\chi(x \rightarrow y)$ は状態 x から y までの遷移の抵抗とよばれる。

$$\lim_{\varepsilon \rightarrow 0} \frac{P_{xy}^\varepsilon}{\varepsilon \chi(x \rightarrow y)} \in (0, \infty) \quad (\text{A1})$$

つぎに、遷移 $x = x^{(0)} \rightarrow x^{(1)} \rightarrow \dots \rightarrow x^{(m)} = x'$ 上の状態 $x \in \mathcal{X}$ から状態 $x' \in \mathcal{X}$ までの道 r に対し、(A2) 式を満たす抵抗 $\lambda(r)$ を考える。

$$\lim_{\varepsilon \rightarrow 0} \frac{P^\varepsilon(r)}{\varepsilon \lambda(r)} \in (0, \infty) \quad (\text{A2})$$

ただし $P^\varepsilon(r)$ は状態遷移確率とする。このとき、 $\lambda(r)$ は次式で与えられる。

$$\lambda(r) = \sum_{i=0}^{m-1} \chi(x^{(i)} \rightarrow x^{(i+1)}) \quad (A3)$$

$x \rightsquigarrow y$ は状態 $x \in \mathcal{X}$ から状態 $y \in \mathcal{X}$ に到達可能, つまり, ある状態からの遷移確率が 0 ではないことを表すとする. このとき, $x \rightsquigarrow y$ かつ $y \rightsquigarrow x$ ならば x は y と連結である. 再帰的連結クラスはすべての状態が同一クラス内の状態と連結であり, クラス外の状態と連結ではないクラスである. $H_1, \dots, H_J$ をマルコフ連鎖 $\{P_t^0\}$ における再帰的連結クラスとする. 定義から各クラス H_l 内では, すべての状態に関して他の状態に抵抗 0 で行ける道が存在する. また摂動的マルコフ過程 $\{P_t^\varepsilon\}$ の場合, 二つの再帰的連結クラス H_l, H_k に対して, H_l 内の状態から H_k 内の状態への道はいくつか存在し, その最小の抵抗を χ_{lk} と表記する.

また, 再帰的連結クラス $\mathcal{H} = \{H_1, \dots, H_J\}$ 上の重み付き完全有向グラフを $\mathcal{G} = (\mathcal{H}, \mathcal{H} \times \mathcal{H}, \mathcal{W})$ と定義する. ここで, 各再帰的連結クラス H_l の任意の状態を頂点 l とし, 辺 (l, k) の重み $w_{lk} \in \mathcal{W}$ はその最小の抵抗 χ_{lk} と一致する. つぎに, $\mathcal{G}$ 上の各頂点 k から l に唯一の道がある全域木を l -tree と定義し, $\mathcal{G}$ 上の l -tree の集合を $\mathcal{G}(l)$ と表記する. さらに, l -tree の抵抗を l -tree の辺の抵抗を足し合わせたものと定義する. 最後に, 再帰的連結クラス H_l の確率論的ポテンシャルを $\mathcal{G}(l)$ に含まれる木の抵抗の中で最も低い抵抗と定義する.

【定義 3】(確率論的安定状態) 状態 $x \in \mathcal{X}$ が $\lim_{\varepsilon \rightarrow 0+} \mu_x(\varepsilon) > 0$ を満たすならば, x は確率論的安定であるという. ただし, $\mu_x(\varepsilon)$ は定常分布 $\mu(\varepsilon)$ の状態 x に対応する要素の値である.

【命題 3】[16] $\{P_t^\varepsilon\}$ を $\{P_t^0\}$ の標準摂動とする. このとき, $\lim_{\varepsilon \rightarrow 0+} \mu(\varepsilon)$ は存在し, $\mu(0)$ は $\{P_t^0\}$ の定常分布となる. さらに, 確率論的安定状態は最小の確率論的ポテンシャルをもつ再帰的連結クラスに含まれる.

付録 2. 定理 1 に用いる命題, 補題

提案アルゴリズム PIPIP の探索度 $\varepsilon(k) = \varepsilon \in (0, 0.5]$ が定数の場合は Payoff-based Homogeneous Partially Irrational Play (PHPIP) とよばれる. 本論文では以下の補題 1~3 を証明なしで用いる. これらの証明に関しては文献 [17] を参照されたい.

【補題 1】被覆ゲーム Γ のもとで PHPIP で導出される摂動的マルコフ連鎖 $\{P_t^\varepsilon\}$ は, 仮定 1 のもとで斉時的マルコフ連鎖 $\{P_t^0\}$ の標準摂動となる.

【補題 2】被覆ゲーム Γ のもとで PHPIP で導出されるマルコフ過程 $\{P_t^\varepsilon\}$ を考える. このとき, $\text{diag}(\mathcal{A}) = \{(a, a) \in \mathcal{A} \times \mathcal{A} : a \in \mathcal{A}\}$ の要素により, 非摂動的マルコフ過程 $\{P_t^0\}$ の再帰的通信クラス $\mathcal{H}_i = \{(a_i, a_i)\}, i \in \{1, \dots, |\mathcal{A}|\}$ が与えられる.

【定義 4】(直行ルート) 集合 $\mathcal{E}_s$ を

$$\mathcal{E}_s := \{(z = (a, a), z' = (a', a')) \in \text{diag}(\mathcal{A}) \times \text{diag}(\mathcal{A}) :$$

$$\exists i \in \mathcal{V} \text{ s.t. } a_i \in \mathcal{R}_i(a'_i), a_i \neq a'_i \text{ and } a_{-i} = a'_{-i}\}$$

とする. $(z^0, z^1) \in \mathcal{E}_s$ である $\text{diag}(\mathcal{A})$ 内の任意の 2 状態 $z^0 = (a^0, a^0), z^1 = (a^1, a^1)$ 間のルートのうち, はじめの反復で 1 台のエージェント i だけが現在の行動 a_i^0 から行動を a_i^1 に変え, 次の反復においても探索した同一のエージェント i が同じ行動 a_i^1 をとり, 他のエージェントが探索しない ($a_{-i}^0 = a_{-i}^1$) ルートを直行ルートという.

【補題 3】被覆ゲーム Γ のもとで PHPIP で導出されるマルコフ過程 $\{P_t^\varepsilon\}$ 上の $(z^0, z^1) \in \mathcal{E}_s$ を満たす任意の 2 状態, $z^0 = (a^0, a^0)$ から $z^1 = (a^1, a^1)$ への遷移を考える. 仮定 2 のもとで z^0 から z^1 までの直行ルート r の抵抗 $\lambda(r)$ は 2 未満であり, z^0 から z^1 までのすべての抵抗の中で最小である.

【定義 5】(m -直行ルート) $\text{diag}(\mathcal{A})$ 内に m 個の頂点をもち, 頂点間のルートが直行ルートであるルートを m -直行ルートという.

【補題 4】被覆ゲーム Γ のもとで PHPIP で導出されるマルコフ過程 $\{P_t^\varepsilon\}$ を考える. (A4) 式を満たす状態 $z^0 = (a^0, a^0) \in \text{diag}(\mathcal{A})$ から $z^1 = (a^1, a^1) \in \text{diag}(\mathcal{A})$ までの $\{P_t^\varepsilon\}$ 上の m -直行ルート r を考える.

$$z^0 = z^{(0)} \Rightarrow z^{(1)} \Rightarrow \dots \Rightarrow z^{(m-2)} \Rightarrow z^{(m-1)} = z^1 \quad (A4)$$

ただし, $z^{(i)} = (a^{(i)}, a^{(i)}) \in \text{diag}(\mathcal{A}), i \in \{0, \dots, m-1\}$ であり, 状態間の矢印は直行ルートを表す. また, (A5) 式に示す逆ルート r' を考える.

$$z^0 = z^{(0)} \Leftarrow z^{(1)} \Leftarrow \dots \Leftarrow z^{(m-2)} \Leftarrow z^{(m-1)} = z^1 \quad (A5)$$

r' も z^1 から z^0 までの m -直行ルートである. このとき, 仮定 2 のもとで $\phi(a^0) > \phi(a^1)$ ならば $\lambda(r) > \lambda(r')$ である.

(証明) 抵抗が 1 より大きい p 個の直行ルートを含むルート r と q 個の直行ルートを含むルート r' を考える. ルート $z^{(i)} \Rightarrow z^{(i+1)}, z^{(i)} \Leftarrow z^{(i+1)}$ における探索エージェントをそれぞれ j_i, j'_i と表記する. 補題 3 より, ルート $z^{(i)} \Rightarrow z^{(i+1)}$ の抵抗は $U_{j_i}(a^{(i+1)}) \geq U_{j_i}(a^{(i)})$ のとき 1, $U_{j_i}(a^{(i+1)}) < U_{j_i}(a^{(i)})$ のとき $1 + \Delta_{j_i} \in (1, 2)$ である.

(9) 式より次式が成立する.

$$\begin{aligned} \Delta_{j_i} &= U_{j_i}(a^{(i)}) - U_{j_i}(a^{(i+1)}) = \phi(a^{(i)}) - \phi(a^{(i+1)}) \\ &= U_{j'_i}(a^{(i)}) - U_{j'_i}(a^{(i+1)}) = -\Delta_{j'_i} \end{aligned} \quad (A6)$$

すなわち, 直行ルート $z^{(i)} \Rightarrow z^{(i+1)}, z^{(i+1)} \Leftarrow z^{(i)}$ の一方の抵抗は 1 であり, 他方は 1 より大きい. ただし, $U_i(a^{(i+1)}) = U_i(a^{(i)})$ の場合は両抵抗ともに 1 である. したがって, 不等式 $p + q \leq m - 1$ が成立する. ここで, $z^{(i)} \Rightarrow z^{(i+1)}$ の抵抗が 1 より大きい Δ_{j_i} を抽出し, $\Delta_1, \Delta_2, \dots, \Delta_p$ と表す. 同様に逆ルート r' についても $\Delta'_1, \Delta'_2, \dots, \Delta'_q$ を定義する. (A6) 式より次式が示される.

$$\Delta_1 + \dots + \Delta_p - (\Delta'_1 + \dots + \Delta'_q) = \phi(a^0) - \phi(a^1) \quad (A7)$$

Fig. A1 Image of Kruskal's algorithm

また, (A7) 式は $U_{ji}(a^{(i+1)}) = U_{ji}(a^{(i)})$ となる $(a^{(i)}, a^{(i+1)})$ も考慮されている. 抵抗の定義から $\Delta_1 + \dots + \Delta_p = \lambda(r) - (m-1)$, $\Delta'_1 + \dots + \Delta'_q = \lambda(r') - (m-1)$ であり, 次式が成立する.

$$\lambda(r) = \lambda(r') + \phi(a^0) - \phi(a^1) \quad (\text{A8})$$

(A8) 式より補題 4 が示された.

【補題 5】被覆ゲーム Γ のもとで PHPIP で導出されるマルコフ過程 $\{P_t^\varepsilon\}$ から構成される重み付有向グラフ G を考える. l -tree を $\mathcal{T} = (\text{diag}(\mathcal{A}), \mathcal{E}_l, \mathcal{W})$ と表記し, $z^l \in \text{diag}(\mathcal{A})$ の確率論的ポテンシャルを与える. 仮定 1, 2 を満たすとき, $\mathcal{E}_l \subset \mathcal{E}_s$ となる.

(証明) グラフ G の辺 $\mathcal{E}$ を $\mathcal{E}_s$ と $\mathcal{E}_d := \mathcal{E} \setminus \mathcal{E}_s$ の二つのクラスに分割する. 補題 3 より $\mathcal{E}_s$ に含まれる辺の重みは 2 未満である. ここでは, $\mathcal{E}_d$ に含まれる辺の重みを考える. PHPIP の性質から $z(t) \in \text{diag}(\mathcal{A})$ であるとき, いかなるエージェントも探索せずに他の行動に変化することはできない. したがって, 探索行動は $\mathcal{E}_d$ 内の辺における遷移をする必要があるので, 2 回以上実行しなければならない. これは $\mathcal{E}_d$ 内の辺の重みは 2 より大きい.

以後, $\mathcal{E}_s$ 内の辺の重みを $w_s (< 2)$, $\mathcal{E}_d$ 内の辺の重みを $w_d (\geq 2)$ と表現し, この簡易化されたグラフ上のルート z^l の最小抵抗木を構築する. ただし, この簡易化は $\mathcal{E}_l$ の要素を変化しない. また仮定 1 より, すべての再帰的通信クラス ($\text{diag}(\mathcal{A})$) は直行ルートだけで連結している. Kruskal のアルゴリズムの手順より, Fig. A1 に示す抵抗 w_d の辺は最小木の辺として選択されない. このように, 確率論的ポテンシャルを与える木は $\mathcal{E}_s$ に含まれる辺だけで構成される.

【命題 4】被覆ゲーム Γ のもとで PHPIP で導出されるマルコフ過程 $\{P_t^\varepsilon\}$ を考える. 仮定 1, 2 を満たすとき, 確率論的安定状態は集合 $\text{diag}(\zeta(\Gamma))$ に含まれる.

(証明) 命題 3, 補題 1, 2 より, G 上の最小確率論的ポテンシャルを含む $\text{diag}(\mathcal{A})$ 内の状態は $\zeta(\Gamma)$ に含まれるという証明の十分性を示せた. ここで, 非最適行動 a_{nop} の状態 $z_{nop} = (a_{nop}, a_{nop}) \in \text{diag}(\mathcal{A})$ と最適 Nash 均衡解 a_{op} の状態 $z_{op} = (a_{op}, a_{op}) \in \text{diag}(\mathcal{A})$ を与える. z_{nop} が $\mathcal{T}$ の根である場合, $\mathcal{T}$ 上で z_{op} から z_{nop} までの唯一のルートが存在する. 補題 5 より, ルート r は m -直行ルートである. いま, ルート r が逆ルート r' に変更された z_{op} の根をもつ $\mathcal{T}'$ を構成する. このとき, $\phi(a_{op}) > \phi(a_{nop})$ より補題 4 から $\lambda(r) > \lambda(r')$ である. したがって, $\mathcal{T}'$ の抵抗は $\mathcal{T}$ の抵抗より小さく, z_{opt} の確率論的ポテンシャルは $\mathcal{T}'$ の抵抗より小さい. これは a_{nop} の選択に関係なく成立する.

命題 3, 4, 補題 1 より, 定理 1 が証明される.

著者略歴

和 佐 やす あき
泰 明

2011 年東京工業大学工学部制御システム工学科卒業, 同年同大学大学院理工学研究科機械制御システム専攻博士前期課程入学, 現在に至る. ゲーム理論を用いた協調制御, 環境・エネルギーの研究に従事. 計測自動制御学会の会員.

後 藤 たつ ひこ
達 彦

2009 年東京工業大学工学部制御システム工学科卒業, 2011 年 同大学大学院理工学研究科機械制御システム専攻博士前期課程修了. 同年 (株) 東芝に入社, 現在に至る. 計測自動制御学会の会員.

畑 中 たけ し
健 志 (正会員)

論文誌 Vol. 25, No. 6, p. 144 参照

藤 田 まさ ゆき
政 之 (正会員)

論文誌 Vol. 25, No. 6, p. 144 参照