状態行動地図の組み合わせによる移動障害物の存在下でのロボットの行動生成

—Decision Making of Robots under Existence of Moving Objects by Combination of State-Action-Maps—

1xxxxxxx

研究代表者　　東京大学大学院工学系研究科　助教　　上　田　隆　一

[研究の目的]

実世界で知的にふるまうロボットは，工場のように固定されていない不確定な環境に対してロバストである必要がある．他の移動体は，環境の不確定要素として代表的なものである．他の移動体をタスク中に考慮することは，ロボットの行動決定の研究で困難な課題の一つである．

本研究では，障害物の存在する状況下での行動決定則を状況に応じて作成するための一般的な対処方法を提案する．提案手法の特徴は，障害物の無い状況での行動決定則を事前知識として，障害物を回避しつつタスクを達成するための行動を算出することにある．

提案手法の流れは，1) 他の移動体や障害物がない状態を想定し，あるタスクに対する基本的なロボットの行動決定則を作成，2) 移動体が存在する場合を想定した行動決定則をいくつか作成し，3) 1,2 の行動決定則を事前知識とみなして組み合わせ，タスクを達成するための行動シーケンスを生成するものである．

[研究の内容・成果]

本年度は，提案手法中で最も重要な，事前知識（行動決定則）の組み合わせ手法を重点的に研究し，学会で発表するために予稿を執筆した．本稿では例題として，ロボットサッカー（ロボカップ 4 足リーグ）でのロボットの行動決定を扱い，シミュレーション・実機で評価を行った結果を示す．シミュレーション中では，移動障害物として敵ロボットを考慮した．敵ロボット（移動障害物）を考慮した行動決定則として，守備行動の行動決定則を作成し，他の行動決定則と組み合わせた．

Fig.1 RoboCup environment

Fig.2 ERS-7

1.　実験環境と問題設定

ロボカップ 4 足リーグ [3] では，Fig.2 のロボットが Fig.1 に示すサッカーフィールド上でサッカーを行う．本稿で扱うタスクは，3 台のロボットが敵のゴールを避けて相手ゴールにシュートすることとする．ロボットは仲間ロボットとの無線による通信を行いながら，自律的に行動する．ロボットが獲得できる情報は不完全，不確実であるが，ロボットはサッカーにおいて実時間で行動決定しなければならない．

2.　方策の連鎖

2·1　状態空間の分割

本稿では，試合の状態が変数の組 $\{x_1, x_2, \ldots, x_n\}$ で記述されると仮定する．これらの変数は状態変数と呼ばれ，状態変数で張られる空間 X は状態空間と呼ばれる．Fig.3 に示すように，この問題で最低限考慮すべき状態変数は各ロボットの姿勢（位置と向き）3×3 変数とボールの位置 2 変数の合計 11 変数で

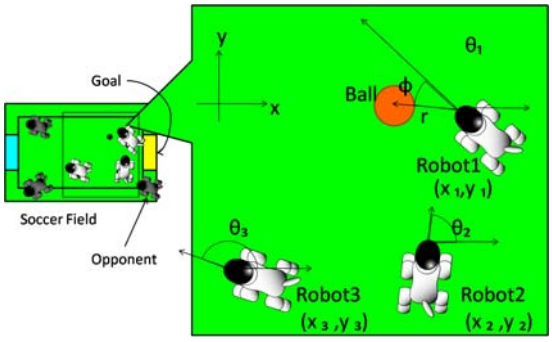


Fig.3 Coordinate system of Robots and the Ball

ある．

ロボットにあるタスクが与えられたとき，ロボットはタスクに対して最善な行動をとるべきである．タスクを評価関数により定義すると，ロボットの挙動は評価関数により定量的に評価できる．最適制御理論をここでは完全に説明しないが，以降の議論のために以下の式を定義する．

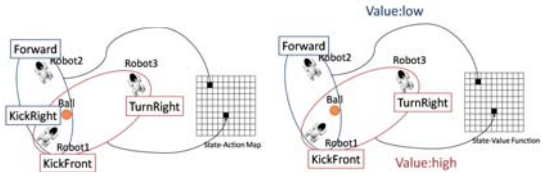
$$\pi : X \rightarrow A \text{ and } V : X \rightarrow \Re \qquad (1)$$

π は状態空間 X 内の状態から行動空間 A 内の行動への写像である．本稿での行動とは，ロボット 3 台の行動の組を表す．この写像は方策と呼ばれる．状態価値関数 V はタスクにおける状態の価値を表す．方策が多次元のルックアップテーブルで作成される場合，方策は状態行動地図と呼ばれる．この場合，状態空間は多次元格子で離散化される．

2·2　状態行動地図

文献 [2] において，ロボット 2 台とボール 1 個の 8 個の状態変数を考慮した状態行動地図が作成されている．作成には動的計画法 [1] が用いられている．この状態行動地図は約 6 億個の離散状態に対して作成されている．

本稿では，さらに多くの離散状態を含む状態行動地図を計算する．フィールドのサイズは縦 4[m] 横 6[m] であるため，状態行動地図と状態価値関数は 30 億状態となり，計算には 1 か月を要した．この地図により，ロボットは 14 種類の歩行行動と 4 種類のキック行動を選択できる．この部分空間にはロボットが 2 台存在するので，行動



(a) confliction of two partial policies　　(b) comparison of values

Fig.4 Confliction and its resolve

空間 A 内の行動の数は行動数の平方となる．しかし，本稿では計算コストを抑えるため，74 通りの行動の組のみを考慮している．この地図を 8D 攻撃地図と呼ぶ．

同様に，次の状態行動地図を作成した．

- 5D 攻撃地図：ロボット 1 台，ボール 1 個を考慮．ロボットのタスクは，ボールを敵ゴールまで，できるだけ少ないステップ数で運ぶ
- 3D 守備地図：ロボット 1 台を考慮．ロボットのタスクは自ゴール前の定位置へ移動し，待機する

これらの 2 つの地図はロボット 1 台のみを考慮している．これらの地図は，8D 攻撃地図の補助的なものである．

2·3　状態価値関数を組み合わせた行動決定

提案手法の概念を Fig.4 に示す．ロボットは前節で作成した地図をさまざまに組合せて使用する．Fig.4(a) は，ロボットがすべて 8D 攻撃地図により行動決定する場合を表す．ここでは，1 台のロボットが両方の部分空間に属する．この場合，各空間での行動決定が互いに対立してしまう．提案手法は，各空間を状態価値関数で比較することでこの問題を解決する．この手法では，ロボットの機能の価値が評価されており，より価値の高い行動が選択される．この手順は Fig.4(b) に示される．

他の状態行動地図を使用すると，価値の比較はより複雑になる．各部分空間のタスクや規模が異なるため，各部分空間を単純に比較することは望ましくない．

ここで，筆者らは以下に述べるヒューリスティックを設計する．しかし，提案手法の基本的な方

針は価値の比較であり，ここでも地図の組合せを価値関数によって比較を行う．

3 台のうちボールに最も近いロボットが攻撃地図を使用すると仮定すると，残り 2 台が攻撃または守備を行うことになる．この場合，役割の組合せは以下の式により選択される．

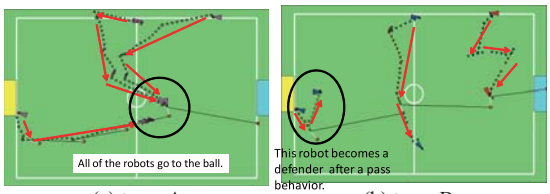
$$\max \begin{cases} V_{\text{offense5D}}(s_1) + \alpha\{V_{\text{defense}}(s_2) + V_{\text{defense}}(s_3)\} \\ 2V_{\text{offense8D}}(s_{12}) + \alpha V_{\text{defense}}(s_3) \\ 2V_{\text{offense8D}}(s_{13}) + \alpha V_{\text{defense}}(s_2) \\ \max\{V_{\text{offense8D}}(s_{12}), V_{\text{offense8D}}(s_{13})\} \\ \quad + V_{\text{offense8D}}(s_{12}) + V_{\text{offense8D}}(s_{13}) \end{cases} \tag{2}$$

s_1 はボールとボールに最も近いロボットの離散状態を表す．以降，ボールに最も近いロボットを Robot1 と呼ぶ．タスク中，この名前が与えられるロボットは環境の変化に応じて変化する．Robot1 以外のロボットは Robot2，Robot3 と呼ぶ．s_{12} は 8 次元空間におけるロボット 2 台とボール 1 台の離散状態である．$V_{\text{offense8D}}, V_{\text{offense5D}}, V_{\text{defense}}$ は各地図の状態価値関数を表す．パラメータ α は正の値であり，攻撃，守備の相対的な重要度を決定する．この式を最大化するロボットの役割の組合せが選択される．役割の組合せが選択されると，ロボットは各組合せに対応する地図を使用し行動決定する．

3. シミュレーション

シミュレーションにより提案手法の有効性を評価する．検証に用いるシミュレータは大域的な行動決定のために作成されており，以下のような点で実環境と異なる．

- ロボットは自身の正確な姿勢を獲得できる．また，他機の正確な姿勢も通信により獲得できる．
- 敵ロボットの姿勢，ボールの位置は，視野内にあるときに正確に獲得できる．
- ボールの把持などの局所的行動を考慮していない．ロボットはボールとの距離が 150[mm] から 250[mm]，角度が −95[deg] から 95[deg] のときにキック可能．



(a) team A (b) team B

Fig.5 Typical Trials on Simulation

3 台のロボットが行うタスクは，敵ゴールへボールを運ぶことである．ボールがフィールド外に出た場合，敵チームがゴールに成功した場合は，タスクが失敗したとみなす．

また，評価に使用するロボットのチーム（3 台）は以下の 3 種類である．

- Team A：5D 攻撃地図のみ使用
- Team B：8D 攻撃地図のみ使用
- Team C：3 種類の地図を使用

敵ロボットは Team A と同種であり，5D 攻撃地図のみを使用する．すなわち，各敵ロボットは仲間へパスすることなく，自分だけで得点を目指して行動する．このチームとの対戦を各チーム 10000 回行い，ゴールの成功数を集計した．

はじめに，Team A と Team C の挙動を Fig.5 に示す．ここでは，説明のため敵ロボットの軌跡を表示していない．Team A のロボットはすべてボールに接近している．一方，Team C のロボットはパスを駆使してボールを移動させている．さらに，守備地図を使用することで，左のロボットがゴール前に待機している．Team C においてロボットの役割分担に成功していることが分かる．

シミュレーションの結果を Table 1 に示す．8D 攻撃地図の使用により，Team B は Team A に比べてゴール数が 5.5[%] 増加した．さらに，3 種類の地図を使用により，Team C では Team A に比べてゴール数が 7.4[%] 増加した．以上より，提案手法によりタスク達成の効率が改善したと言える．8D 攻撃地図のみを使用した Team B に対して，3 種類の地図を用いた Team C の改善が小さい理由として，守備地図の性能が悪い可能性がある．

Table 1 The result of simulation

team	goals
Team A(uncooperative)	5228
Team B(cooperative, only 8D map)	5516
Team C(cooperative, three maps)	**5613**

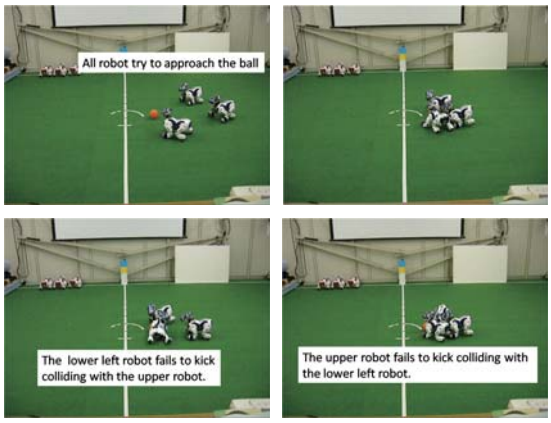


Fig.6 A trial of the uncooperative team (Team A)

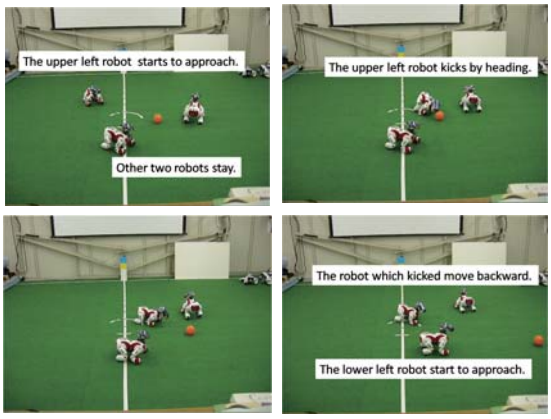


Fig.7 A trial of the uncooperative team (Team B)

4. 実機実験

提案手法を ERS-7 へ実装し，実環境での検証を行う．本節では，非協調の Team A，8D 攻撃地図のみを用いる Team B の 2 チームの挙動を比較する．本実験では，敵チームを使用せず，両チームの挙動のみを確認する．

両チームの挙動を，Fig.6 と Fig.7 に示す．Team A のロボットは互いに協調していないため，ロボット間で衝突が発生している．一方，提案手法による Team B のロボットは互いに距離をとっており，衝突は発生していない．

5. おわりに

本稿では，ロボット 3 台の準最適な協調行動を実現する手法を提案した．ロボットの作業前に動的計画法によりロボット 2 台用の最適方策を作成し，それらを組み合わせて 3 台に適用した．各ロボットがどの方策を用いるかを，状態価値関数により役割の組合せを比較することで決定した．ロボカップ 4 足歩行リーグをモデル化したシミュレーションにおいて，提案手法により 3 台の協調に成功し，ゴールまでの時間の短縮に成功した．また，実機実験において，提案手法によるチームではロボットが互いに距離をとって行動する様子が観測された．

[今後の研究の方向，課題]

以上のアルゴリズムに，敵ロボットの位置を価値関数計算に考慮して行動決定する機能を追加した評価が完了している．今後，論文等で順次発表する予定である．

[成果の発表，論文等]

1) 上田隆一，小林圭祐，新井民夫：“価値反復による 2 台のサッカーロボットの協調動作計画 —30 億状態を有する状態行動地図の生成と評価”，自律分散システムシンポジウム，pp.229-234，2008.

2) 小林佳祐，上田隆一，新井民夫：二台用最適方策を用いた複数ロボットの協調行動，日本機械学会ロボティクス・メカトロニクス講演会（ROBOMEC），長野（2008）

3) K. Kobayashi, R. Ueda and T. Arai：“Cooperative Behavior of Multiple Robots by Chain of Monolithic Policies for Two Robots,” The 10th International Conference on Intelligent Autonomous Systems, Baden Baden, Germany（2008）

[参考文献]

1 R. E. Bellman:Dynamic Programming, Princeton University Press, 1957.

2 R. Ueda, *et al.*:“Dynamic Programming for Creating Cooperative Behavior of Two Soccer Robots — Part 1: Computation of State-Action Map,” Proc. of IEEE ICRA, pp.1-7 (2007)

3 H. Kitano, *et al.*:“RoboCup: The Robot World Cup Initiative,” Proc. of International Conference on Autonomous Agents, pp.340-347 (1997)

4 P. Stone and M. Veloso:“Task Decomposition, Dynamic Role Assignment, and Low-Bandwidth Communication for Real-Time Strategic Teamwork,” Artificial Intelligence, vol.110, pp.241-273 (1999)