ロボットブレインのための戦略学習モデルの考案とそのLSI化

小野　将寛（先端研量子物質科学専攻　D1),
佐々木　守（先端研半導体集積科学専攻　助教授),
岩田　穆（先端研半導体集積科学専攻　教授）

1. 研究目的

近年のロボットの進歩は目覚しく，今後ロボットが社会において担う役割はより拡大し，その需要も増加することが期待される．特に，期待される分野の一つとして，福祉，コミュニケーションなどの対人型タスクが挙げられる．人間は学習するので，ロボットが認識する環境や状況（癖や戦略など人間の特徴）は頻繁に変化する．そのような条件下でタスクを遂行させるためには，従来よりもロバストな知能を実現する必要がある．具体的には，学習によって，素早く相手の特徴（癖，戦略）を把握し，それに基づいて適切な戦略を選択する柔軟な制御機構を持つロボットブレインが必要になる．そこで本研究では，ロボットブレインを実現するための戦略学習モデル（制御機構）を考案し，その有効性を検証し，モデルに適した小型で低消費電力の LSI で実装することを目的とする．

2. 研究成果概要

提案した戦略学習モデル（Fig. 1）は三つのセクションから構成される．まず，予め記憶された代表的な特徴に対する戦略の中から現在の相手に適切な戦略の選択を行い(セクション 1)，それをベースに学習によって相手に有効な戦略を構築し（セクション 2)，必要に応じて戦略の追加/削除を行う（セクション 3)．これまでの研究でセクション 1，2 を考案し，その有効性をシミュレーション実験により確認した．

各セクションのアルゴリズム概要を説明する．まず，セクション 2 は，強化学習[1]（QPSP-Learning[2]）を用いて構成した．戦略を「各状況での行動の集合」と定義し，時刻 i での状況 $\mathbf{s}_i$ に対して行動 $\mathbf{a}_i$ の適切度を計る関数 $Q(\mathbf{s}_i,\mathbf{a}_i)$を用意し，Q 関数に基づいて行動を決定する．そして，その結果の良し悪しを Q 関数に反映する形で相手の特徴に合った戦略学習を行う．次にセクション 1 のアルゴリズムを説明する． Q 関数が相手の特徴を反映する点に着目し，新たに相手の特徴検出用の Q 関数を用意し，記憶中の Q 関数の中でそれに最も類似する戦略を選択することにより構成した．

実験は，対人型タスクとしてエアホッケーを例にとり，ロボットと対戦プレイヤーをコンピュータ上に作成して行った．実験内容は，プレイヤーに複数の特徴（癖，戦略）を持たせ，ロボットにはそのうち数種類に対しての有効な戦略を予め記憶させた上で 1 セット 20 ポイント先制ルールの下に対戦させた．

Fig. 3 に実験結果の一例を示す．ここでは，プレイヤーの特徴を A～E とし，そのうちロボットは E 以外に対して有効な戦略（戦略 A, B, C，D）を記憶させた上で試合を行った．なお，本例でのプレイヤーの特徴はスティックを構える位置と設定（Fig. 2）した．この結果から分かるように，記憶している戦略の選択が素早くほぼ正常に動作しており，また，Fig. 5 より戦略が記憶されていない未知の特徴(E)に対しても，記憶中の既存戦略の中から適切なもの選択し，それをベースとして学習を行うことで，有効な戦略を構築していることを確認した．ただし，相手の特徴が B の時に戦略 C を選択している点については， Fig. 4 より，プレイヤーに 1 点も許すことなく 1 ゲームを終了していることから適切な選択であると考えられる．以上より，セクション 1，2 に対しての有効性が示された．

3. COE プログラムと成果の関連

COE プログラムの目的は「シリコンナノデバイス・回路・アーキテクチャの融合による高度認識・学習機能集積化システムの基盤構築」である．そこで，本研究成果である戦略学習モデルを今後 LSI 実装することにより，COE プログラムの目的の一つである学習機能を有する LSI チップとして提供することができる．具体的には，戦略の選択，戦略学習と追加・削除機能有するチップと戦略データベースとして複数のチップを用意し，各チップ間で無線通信を行うことで戦略学習モデルを実現する．

4. まとめと今後の予定

対人型タスクを処理するためのロボットブレインを実現するため，強化学習を用いた戦略学習モデルを提案した．そして，そのモデルが相手の特徴に応じた戦略を素早く選択し，そこから更に有効な戦略を構築することをシミュレー

ション実験により確認した．今後は，実際に人間を相手に本モデルを適用し，その有効性を確認したのち，本モデルの LSI 化へ研究を進展させる．

5. これまでの研究発表

1. 小野将寛, 汐崎充, 佐々木守, 岩田穆, 強化学習を用いた対戦相手適応型戦略モデル, 電子情報通信学会信学技報, NC2003-44, pp. 61-66
2. 小野将寛, 汐崎充, 佐々木守, 岩田穆, 強化学習による対戦相手適応型戦略モデル, FIT2003 第 2 回情報科学技術フォーラム論文集, G-008, pp. 287-288
3. 小野将寛, 汐崎充, 佐々木守, 岩田穆, ロボットブレインのための戦略学習モデル, 第5回 IEEE 広島支部学生シンポジウム論文集, D-44, pp. 274-276

参考文献

[1] Richard S.Sutton and Andrew G. Barto, Reinforcement Learning, MIT Press, 1998.

[2] Horiuchi, T., Fujino, A., Katai, O. and Sawaragi, T. Q-PSP Learning: An Exploitation-Oriented Q-Learning Algorithm and Its Applications. SICE(in Japanese), 35(5), 645-653(1999).

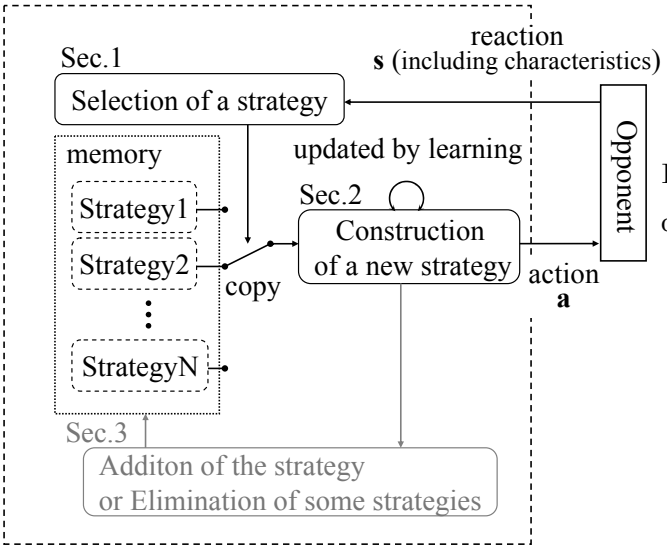

Fig. 1: A proposed model for Robot brain

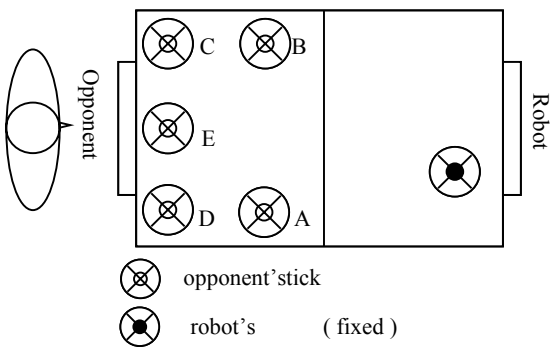

Fig. 2: An example of the opponent's characteristics

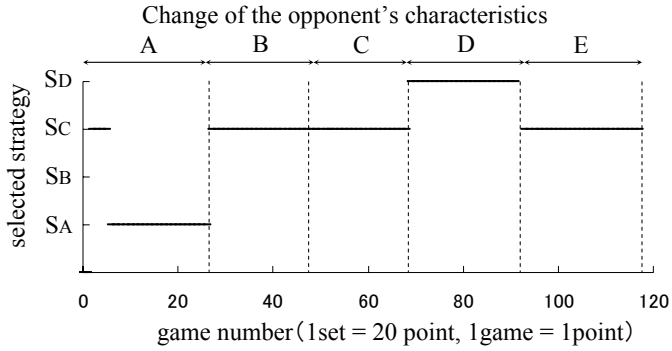

Fig. 3: The result of the experiment where the opponent's stick position changes A from E by one set(= twenty points)

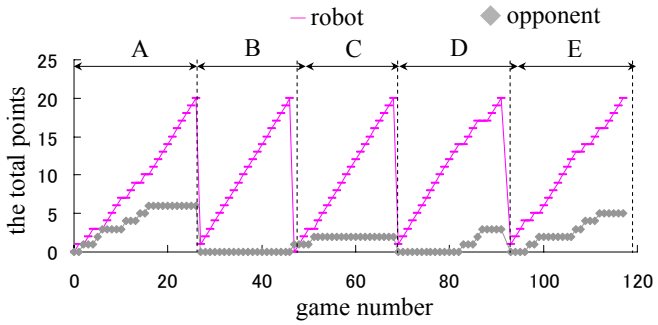

Fig. 4: The total points that the robot and the opponent got

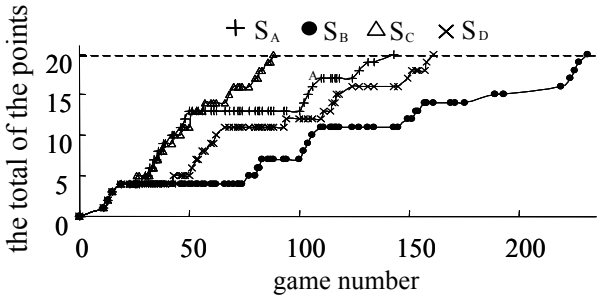

Fig. 5: The total points given to the robot by using each strategy during position E