強化学習と脳信号によるBCI協調学習の基礎的研究

Fundamental Study of Collaborative Learning Consisting of Reinforcement Learning and Brain Signal in BCI

林　勲　三輪 亮太　仙浪 雅典
I.Hayashi　R.Miwa　M.Sennami
関西大学　関西大学　関西大学
Kansai Univ.　Kansai Univ.　Kansai Univ.

Abstract Reinforcement learning has been studied as an unsupervised learning method. Alternatively, BCI(Brain computer interface) comes into the research limelight recently. However, non-synchronous spontaneous action potentials and evoked action potentials exist noisy brain signal, and we need an interface model which exists between brain and machine for control and stability. In this paper, we propose a collaborative learning system consisting of reinforcement learning and brain signal in BCI. Brain signal is interpreted as a deliberate assignment of the subject, and we utilize reinforcement learning in control and stability for BCI. We apply the collaborative learning to maze problem and show the usefulness of the proposed system.

1. はじめに

最近，教師なし学習の一手法として強化学習 [1] が注目されている．エージェントが環境状態を観測し，行動を決定して，報酬が最大となるように目標を達成する．しかし，最適解を得るには学習プロセスにおいて内部報酬が必要となる [2]．一方，脳信号を用いて機械を制御するBCI(Brain computer interface) [3] の研究が推進されている．しかし，脳信号には非同期の自発的活動や反射反応などのノイズ的信号が混在し，信号を種別ごとに分類することが困難で，脳と機械にインタフェースモデルを介在させて安定性を確保させる必要がある [4].

本研究では，BCI と強化学習とを融合させることによって，強化学習の内部報酬として脳からの信号反応による示唆を行い，BCI における制御安定インタフェースとして強化学習システムを用いる協調学習システムを提案する．BCI における脳示唆は脳学習の結果として得られ，また，制御や行動決定のために強化学習を用いるので，本システムは脳示唆と強化学習による協調学習を構成しているといえる．ここでは迷路探索を用いて，協調学習の有用性を検討する

2. 協調学習システム

図 1 に協調学習システムを示す．破線内は従来の強化学習の概要である．エージェントは時刻 t に環境の状態 $s(t)$ を観測して行動 $a(t)$ を決定し，それに応じた報酬 $r(t)$ を得る．一方，協調学習では，強化学習に加えて，脳信号による示唆 (脳示唆)$su(t)$ を行う．脳示唆を内部報酬とすることで，学習が効率よく行われ，また，脳と機械の中間インタフェースを得ることにより，どのような環境においても安定かつ正確な制御を行うことが可能になる．

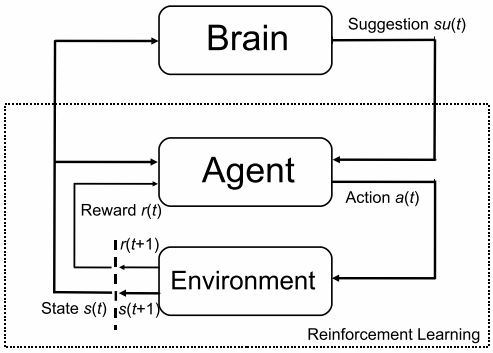

Fig. 1: Proposed Collaborative Learning System

3. 探索問題への応用

図 2 に探索に用いる迷路を示す．ここでは，2 種類の探索問題について，近赤外分光法 (NIRS) により，被験者の規範行動の脳示唆を脳信号として計測し，一定ステップ数ごとに $greedy$ 法による [指示] か $\epsilon - greedy$ 法による [非指示] かの示唆を判別し，ゴールにたどり着くまで経路を探索する．

探索 1:　迷路の一部を開示して，協調学習による探索

探索 2:　迷路の全体を開示して，協調学習による探索

探索 1 では 50 ステップか 100 ステップごとに，探索 2 では 50 ステップごとに示唆を得た．脳示唆の判別は次式を用いる．

$$SU \quad : \quad \sum_{\{ox,deox\}} |E(l) - E(su)| \leq \sum_{\{ox,deox\}} |E(l) - E(ns)|$$

$$NS \quad : \quad \sum_{\{ox,deox\}} |E(l) - E(su)| > \sum_{\{ox,deox\}} |E(l) - E(ns)|$$

ただし，探索中に観測した酸化ヘモグロビン量 (ox) と還元ヘモグロビン量 ($deox$) の平均値を $E(l)$ で示し，規範行動としての [指示行動 (SU)] と [非指示行動 (NS)] を $E(su)$，$E(ns)$ で表す．

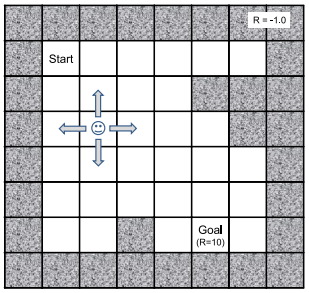

Fig. 2: Maze

4. 探索結果

4.1 試行回数の比較

探索 (1) 行動では，3 名の被験者に対して 13 回の探索から解析を行い，探索 (2) 行動では，2 名の被験者に対して 8 回の探索から解析を行った．一方，強化学習では， 25 回の探索を行った．1 回の探索における平均試行回数の比較では，強化学習が 119.3 回に対して，探索 (1) 行動は 113.0 回，探索 (2) 行動は 118.9 回となった．したがって，協調学習は強化学習と比較して平均試行回数が少なく，効率的な探索が行われている．

4.2 無更新回数の比較

ある探索において，36 個のセルの 4 個における上下左右移動の Q 値の無更新回数を比較した (図 3 参照)．36 セルを 9 セルにまとめ，各セルの Q 値を上移動，下移動，左移動，右移動別に示した．探索の最短ルートから外れる方向への移動，例えば，C1 の上移動と右移動，C2 の上移動，B3 の下移動と左移動の比較では，協調学習の無更新回数が多く，協調学習が最短ルート付近に集中して Q 学習を行っていることがわかる．

4.3 Q 値の比較

36 個のセルにおける強化学習の 25 回の Q 値の平均と協調学習の Q 値の平均との偏差を図 4 に示す．各セルの Q 値を上移動 (左上に表示)，下移動 (右下に表示)，左移動 (左下)，右移動 (右上) 別に示した．例えば，B1 の下移動では，脳示唆の判断回数が最も高いグレードを示し，強化学習の平均 Q 値は-0.94，協調学習との偏差は 0.16 である．脳示唆の高回数に応じて Q 値が良く更新されている．一方，B6 の右移動と F1 の右移動では，強化学習と協調学習の平均 Q 値には差があり，脳示唆

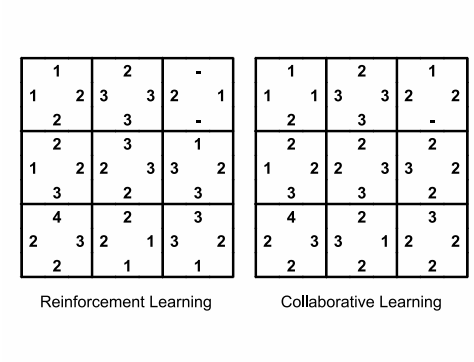

Fig. 3: Comparison of Learning Times

の低回数がこの差に影響を与えている．したがって，強化学習と比較して，協調学習が最短ルート付近に集中して Q 学習を行っていることがわかる．

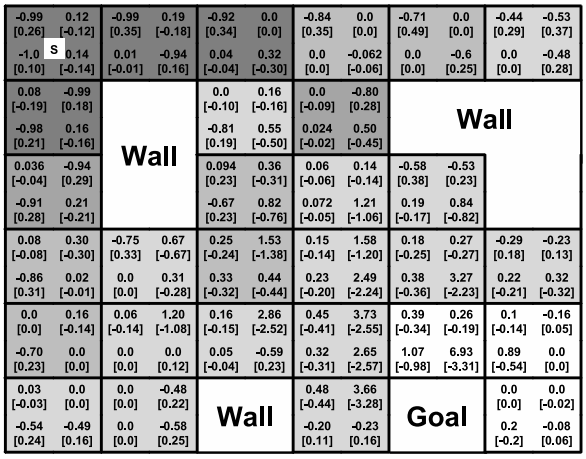

Fig. 4: Comparison of Q Values

5. おわりに

今後，脳示唆の頻度を変えて協調学習の有用性を検討する必要がある．

参考文献

[1] 伊藤：ロボットインテリジェンス，オーム社 (2007)

[2] 石川：知能の根源としての分節化と好奇心，第 5 回脳と知覚研究部会ワークショップ，BP08-01 (2008)

[3] 櫻井，八木，小池，鈴木：ブレインマシン・インタフェース最前線，工業調査会 (2007)

[4] 林，徳田，清原，田口，工藤：生態表現システム：ファジィ推論を用いた培養神経細胞における適応学習の解析, 第 23 回ファジィシステムシンポジウム講演論文集, pp.565-570 (2007)

[**連絡先**]
林　　勲　　関西大学 総合情報学部
e.mail ihaya@cbii.kutc.kansai-u.ac.p