強化学習を用いたBCI協調学習の有用性の検討

An Analysis of Usefulness of Collaborative Learning Using Reinforcement Learning in BCI

林　勲　福島 邦彦
I.Hayashi　K.Fukushima
関西大学　関西大学
Kansai University　Kansai University

Abstract Recently, BCI(Brain-computer interface) and BMI(Brain-machine interface) come into the research limelight. However, nonsynchronous spontaneous action potentials and evoked action potentials exist contain brain signal, and we need an interface model between brain and machine for control and stability. We have already proposed collaborative learning system consisting of reinforcement learning and brain signal. Brain signal is interpreted as a deliberate assignment of the subject, and we utilize reinforcement learning in control and stability for BCI. In this paper, we discuss the usefulness of collaborative learning for BCI using reinforcement learning. We first design the collaborative learning system with near-infrared spectroscopy (NIRS), and apply it to maze problem. In addition, we discuss the comprehensive evaluation of collaborative learning system in terms of difficulty of the problem, precision of the problem and the mental load to the subject, and show the usefulness of the proposed system.

1. はじめに

近年，BMI(Brain-machine interface) や BCI(Brain-computer interface) [1] と呼ばれる研究が注目されている．BMI/BCI では，脳と機械・コンピュータとを相互に接続し，脳から外部機械に信号を出力するトップダウン処理と外部機械から脳へ信号を入力するボトムアップ処理とを安定良く制御する必要がある [2]．しかし，脳の信号反応は非同期の自発的反応や反射反応，行動反応などが混在し，これらの反応を種別ごとに分類することは困難である．そこで，脳と外部機械との中間にインタフェースモデルを介在させることにより，制御や信号の安定性を確保させる．その一手法として，BCI に強化学習 [3, 4] を用いた協調学習が提案されている [5, 6]．協調学習では，強化学習を用いて外部機器を安定的に制御することができ，BCI による脳示唆を強化学習に与えて，機器をより人間の意図に沿うよう制御することができる．

本論文では，強化学習を用いた BCI 協調学習の有用性について議論する．具体的には，協調学習の効果を課題回答の精度，被験者への負荷，課題の困難性の 3 つの観点から総合的に評価する．課題回答の精度とは，協調学習の課題達成に関する結果の精度である．また，被験者への負荷とは，被験者が課題達成に際して強いられる負担量を表す．課題の困難性とは，与えられた課題に対する難易度である．これらの 3 つの評価値による加重平均を用いて総合評価指標を定義し，迷路探索問題を例として評価指標を議論する．迷路探索問題では，6×6 の合計 36 マスからなる迷路を構成した．迷路には危険地帯を設け，スタートからゴールまでの経路を強化学習で探索する．しかし，BCI による脳示唆によって，大きな負の報酬を伴う危険地帯を避けるように効率的に最適経路を学習する．脳示唆は，近赤外分光法 (NIRS) により被験者の指示行動として計測した．

2. 強化学習を用いた BCI 協調学習

図 1 に協調学習の概要を示す．破線内は強化学習である．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．協調学習では，脳信号による示唆 (脳示唆)$su(t)$ を行う．脳示唆により内部報酬が与えられるので，学習効率が向上し，強化学習によるインタフェースを実現しているので，安定かつ高精度な制御を実現することができる．

ここでは，協調学習の評価を次の観点から議論する．まず，強化学習は教師なし学習を実現するが，学習過程では最適性を満足しているとはいえない．また，BCI による脳示唆では，常時，被験者は脳信号が計測されるので，その観測負担は大きい．そこで，協調学習の総合評価として，与えられた課題に精度とその課題に対する困難性，及び，脳示唆による被験者への負担の 3 種類の評価から総合的に定義する．

F_1：課題に対する精度

F_2：被験者への負担

F_3：課題の困難性

ここで，F_1 は協調学習の課題達成に関する結果の精度である．F_2 は被験者が課題達成に際して強いられる負担量を表す．F_3 は与えられた課題に対する難易度である．

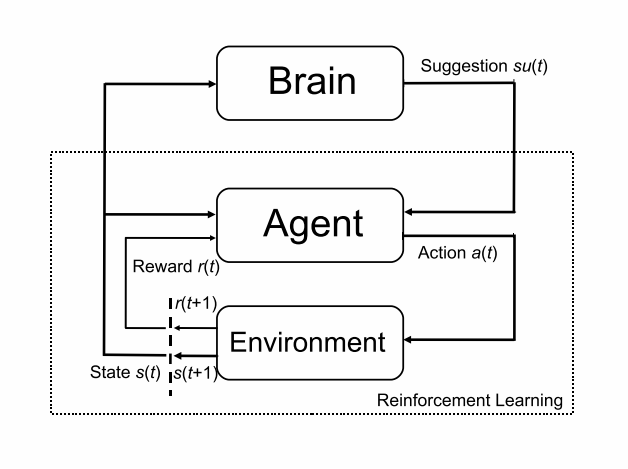


Fig. 1: Proposed Collaborative Learning System

ここでは，F_1, F_2, F_3 の 3 評価の総合評価を F とし，評価式を次のように定義する．

$$F = w_1F_1 + w_2F_2 + w_3F_3$$

ただし，w_1, w_2, w_3 は重み係数である．

3. 迷路探索実験

協調学習システムの例として，迷路探索問題を取り上げる．被験者は探索中の迷路を固視し，強化学習による探索に意図を介入させる場合には，その指示行動を BCI の脳示唆として与える．

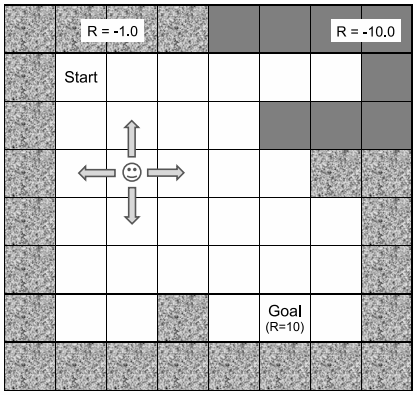


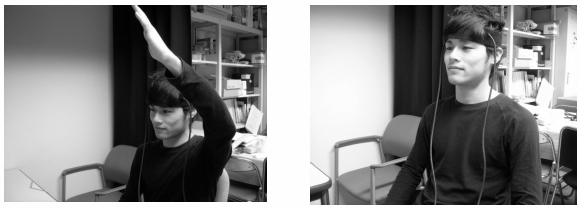
Fig. 2: Maze

被験者は 20 歳代の 3 名である．迷路はモニターに表示され，被験者との距離を 150cm に固定した．図 2 に迷路を示す．迷路は，6×6 の合計 36 マスから構成され，脳示唆を与えない場合には，Q 学習により，スタート (S) からゴール (G) までの経路を探索する．エージェントは，[上，下，左，右] の 4 方向から 1 方向を選択して行動し，ゴールに到達した場合には報酬「10」を得る．ただし，壁に衝突した場合には「-1」，危険地帯に侵入した場合には「-10」を得て，1 ステップ前の位置から再探索する．被験者が探索に意図を介入させたい場合には，被験者の脳信号が近赤外分光法 (NIRS) により計測され，その指示行動を BCI の脳示唆として与える．具体的には，被験者が探索前に記録した各マスでの意図行動に従って被験者の行動を決定する「指示行動」か，あるいは，Q 学習の ε-greedy 法による「非指示行動」かを計測する．計測は，国際 10-20 法による $FP1$ と $FP2$ での酸化ヘモグロビン変化量 (ox) と還元ヘモグロビン変化量 ($deox$) を用いた．

指示行動：被験者の行動指示

非指示行動：ε-greedy 法による Q 学習

まず，実験前に，規範行動として被験者の指示行動と非指示行動のそれぞれの計測値を 10sec 間で 5 回計測し平均値として算出した．被験者の「指示行動」と「非指示行動」の例を図 3 に示す．左部 (a) は指示行動を示し，右部 (b) は非指示行動を示す．



(a) Supervisor Instruction　(b) Non-supervisor Instruction

Fig. 3: Supervisor Instruction and Non-supervisor Instruction

探索中において，脳示唆による介入を行う際には，被験者は，規範行動と同様な「指示行動」もしくは「非指示行動」を行う．その際の被験者の $FP1$ と $FP2$ の酸化ヘモグロビン変化量 (ox) と還元ヘモグロビン変化量 ($deox$) をそれぞれ 10sec 間で計測する．被験者の「指示行動」か「非指示行動」の判定は次式を用いる．

$$SI: \sum_{h=\{ox,deox\}} |E(h) - E(si)| \leq \sum_{h=\{ox,deox\}} |E(h) - E(ns)|$$

$$NS: \sum_{h=\{ox,deox\}} |E(h) - E(si)| > \sum_{h=\{ox,deox\}} |E(h) - E(ns)|$$

ただし，$E(h)$ は探索中に観測した酸化ヘモグロビン変化量 (ox)，及び，還元ヘモグロビン変化量 ($deox$) の平均値を示し，$E(si)$，$E(ns)$ は，それぞれ，規範行動としての「指示行動 (SI)」と「非指示行動 (NS)」を示す．

ここでは，協調学習の有用性を検討することが目的であるので，脳示唆は，探索ステップの間隔を変化させ，15, 30, 45, 60, 75, 90, 105, 120, 200, 300, 400, 500, 600 回ごとに与えるものとして，探索ステップ間隔の変化に対する評価値 F を算出して，協調学習の有用性を検討した．

4. 探索結果

4.1 総合評価 F

実験では，3 名の被験者に対して合計 7 回の計測を行った．総合評価 F を算出する各評価 F_1，F_2，F_3 は，それぞれ，探索効率評価，脳示唆行動評価，危険地帯到達評価として，$[0,1]$ に変換した．各評価を次のように定義する．

$$F_1 = \frac{P_1 + P_2 + P_3}{3} \quad (1)$$

$$P_1 = \frac{x_1 - min(x_1)}{max(x_1) - min(x_1)} \quad (2)$$

$$P_i = \frac{max(x_i) - x_i}{max(x_i) - min(x_i)}, \quad i = 2,3 \quad (3)$$

$$F_i = \frac{max(x_{i+2}) - x_{i+2}}{max(x_{i+2}) - min(x_{i+2})}, \quad i = 2,3 \quad (4)$$

ただし，F_1 は，評価 P_1，P_2，P_3 から成り立つとし，x_1 は収益，x_2 は試行数，x_3 はステップ数，x_4 は脳試行回数，x_5 は危険地域到達回数とする．

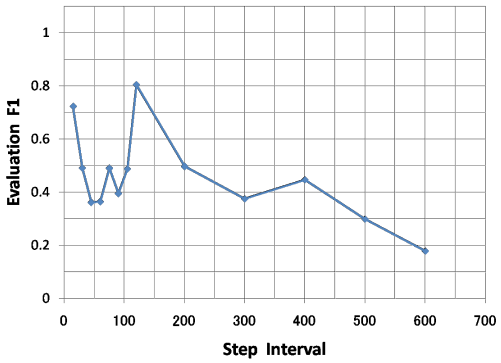


Fig. 4: Estimation of Search Efficiency

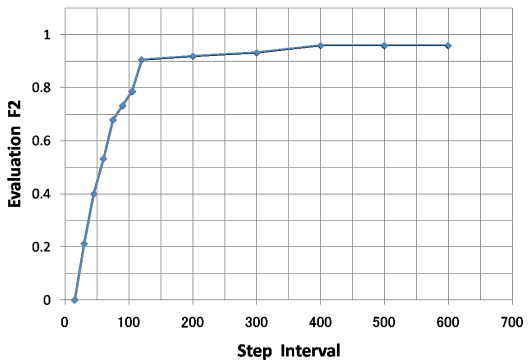


Fig. 5: Estimation of Supervisor Instruction

探索効率評価 F_1 の結果を図 4 に示す．脳示唆ステップ間隔が上昇すると，探索効率評価 F_1 が降下していることがわかる．脳示唆を与えない，つまり，強化学習のみによる評価値は 0.43 として得られた．したがって，脳示唆ステップ間隔が 250 回以内では，協調学習が強化学習よりもほぼ効率が良く，250 回以上では，強化学習のみの方が効率的であるといえる．

次に，脳示唆行動評価 F_2 の結果を図 5 に示す．脳示唆ステップ間隔が上昇すると，脳示唆行動評価 F_2 が上昇し，ステップ間隔が 400 回以上では，ほぼ 1 を満足している．

次に，危険地帯到達評価 F_3 の結果を図 6 に示す．脳示唆ステップ間隔が上昇すると，被験者が介入する機会が増えるので，危険地帯到達評価 F_3 が降下している．強化学習のみによる評価値は 0.51 として得られた．脳示唆ステップ間隔が 200 回以内では，協調学習が強化学習よりも危険地帯に到達する回数が少なく，脳示唆ステップ間隔を 200 回以上にすると，危険地帯に到達する回数が強化学習の場合よりも多くなる．

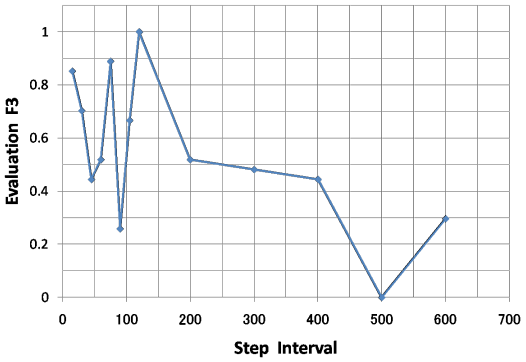


Fig. 6: Estimation of Reaching Dangerous Area

F_1 から F_3 の評価値を $w_i = 1/3$, $i = 1,2,3$ として加重平均した総合評価 F の結果を図 7 に示す．脳示唆ステップ間隔が 120 回の探索のとき総合評価 F は最大評価値を示している．すなわち，協調学習は脳示唆ステップ回数が 120 回の場合に，最も効率の良い探索を行っているといえる．

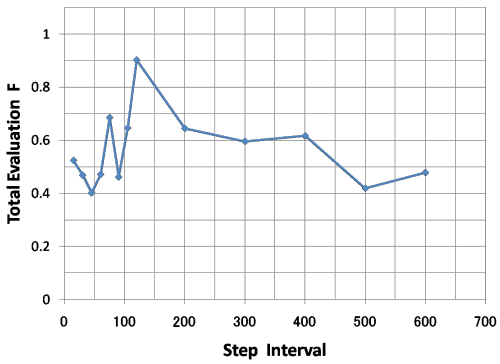


Fig. 7: Total Evaluation

4.2 壁衝突回数の比較

脳示唆ステップ回数が 120 回の場合の協調学習と強化学習とを壁衝突回数で比較した．図 8 に試行ごとの壁衝突回数の比較を示す．協調学習の壁衝突回数を実線で，強化学習の壁衝突回数を点線で示す．協調学習は，3 名の被験者に対して，脳示唆ステップ回数を 120 回として 15 回の連続試行を行った．強化学習は，Q 値を保持しながら 15 回の連続試行を行い，5 回の実験の平均値である．協調学習は強化学習と比較して，ほとんどの試行で強化学習よりも低い衝突回数を示し，さらに，試行を繰り返すことによって，衝突回数が低下していることがわかる．

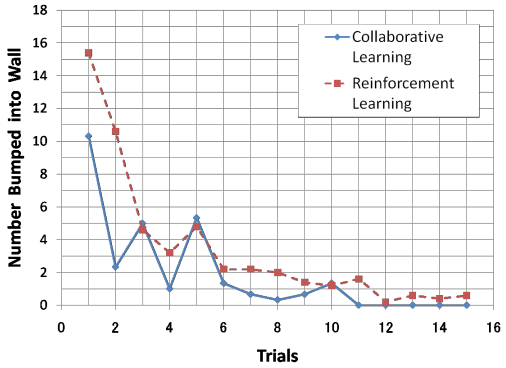


Fig. 8: Number Bumped Wall

4.3 Q 値の比較

3 名の被験者の協調学習と強化学習との Q 値を比較した．結果を図 9 に示す．協調学習は，脳示唆ステップ間隔を 120 回とし，15 回の連続試行の Q 値の平均値である．強化学習は，15 回の連続試行で 5 回実験を行った Q 値の平均値である．各セルでは，上移動 (左上に表示)，下移動 (右下に表示)，左移動 (左下に表示)，右移動 (右上に表示) を表し，上部の数値が協調学習の Q 値であり，下部の数値が強化学習の Q 値である．協調学習は，危険地帯付近である $D1$，$D2$，$E3$ で危険地帯を避ける方向に Q 値が更新されていることがわかる．

5. おわりに

本論文では，強化学習を用いた BCI 協調学習の有用性を課題回答の精度，被験者への負荷，課題の困難性の 3 つの観点から総合的に議論した．迷路探索問題を例として協調学習の有用性を示した．今後，迷路探索問題だけではなく他の多くの応用問題に適用して，協調学習の有用性を議論する必要がある．

なお，本研究の一部は「文部科学省私立大学戦略的研究基盤形成支援事業（平成 20 年度〜平成 24 年度）」によって行われた．

	A	B	C	D	E	F
1	-0.79 0.0 0.02 -0.06 S -0.83 0.01 0.01 -0.1	-0.67 0.0 0.06 -0.18 0.0 -0.79 0.0 -0.03	-0.81 0.0 -0.02 -0.22 0.0 0.01 0.0 -0.25	-6.67 0.0 1.36 0.0 0.0 0.0 0.0 -0.37	-7.4 0.0 0.65 0.0 0.0 -5.83 0.0 0.42	-7.5 -8.75 -0.09 -2.5 0.0 -6.25 0.0 0.18
2	0.0 -0.67 0.0 0.12 -0.67 0.08 0.09 -0.26	Wall	0.0 0.0 -0.01 0.0 -0.67 0.21 -0.02 -0.5	0.0 -6.67 0.0 0.18 0.0 0.0 -0.01 -0.83	Wall	
3	0.0 -0.5 -0.02 0.34 -0.63 0.38 0.15 -0.45		0.0 0.0 -0.03 -1.06 -0.5 0.81 0.19 -0.27	0.0 0.0 0.0 -0.77 0.0 0.0 0.0 -2.39	0.0 -0.5 5.86 0.08 0.0 0.34 -0.43 -0.86	
4	0.0 0.97 -0.01 -0.22 -0.58 0.0 0.05 -0.39	-0.58 0.0 0.08 -1.66 0.0 1.62 -0.01 1.33	0.0 0.0 0.0 -2.84 0.0 2.57 0.0 0.74	0.0 0.0 -0.66 -1.74 0.0 1.27 -0.51 -2.92	0.0 0.0 -0.08 0.0 0.46 2.25 -0.67 -0.99	-0.5 -0.5 0.1 0.18 0.0 0.0 -2.48 0.0
5	0.0 0.0 -0.02 -0.71 0.0 0.0 0.71 0.0	0.44 2.59 0.44 0.65 0.0 0.0 0.0 -0.06	0.42 6.41 -0.03 1.93 0.0 -0.75 0.0 -0.08	0.0 8.53 -0.6 6.90 1.20 0.0 0.81 -6.4	0.0 0.0 -2.27 -0.99 0.0 9.93 0.0 1.71	0.0 -0.5 -0.23 0.0 1.13 0.75 0.48 -1.22
6	0.0 0.0 0.0 0.0 -0.5 0.0 0.08 0.10	0.0 -0.5 -0.46 0.24 0.0 0.0 0.0 0.63	Wall	3.61 5.0 2.52 -4.67 0.0 -0.5 0.53 0.09	Goal	0.0 0.0 0.0 0.5 7.5 0.0 0.89 0.5

Fig. 9: Comparison of Q Values

参考文献

[1] 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)

[2] 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)

[3] 伊藤：ロボットインテリジェンス，オーム社 (2007)

[4] 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)

[5] 林，三輪，福島，堀：脳信号と強化学習による災害時経路探索の基礎的研究，安全工学シンポジウム *2009* 講演論文集，pp.172-173 (2009)

[6] 林，三輪，仙浪：強化学習と脳信号による BCI 協調学習の基礎的研究，第 *25* 回ファジィシステムシンポジウム講演論文集，1B1-02 (2009)

[連絡先]

林　　勲　関西大学 総合情報学部
〒 569-1095 大阪府高槻市霊仙寺町 2-1-1
tel. 072-690-2448
fax. 072-690-2491
e.mail ihaya@cbii.kutc.kansai-u.ac.p