脳信号と強化学習による災害時経路探索の基礎的研究

○ 林　勲 (関西大学 総合情報学部)　三輪 亮太 (関西大学 総合情報学部)
福島 邦彦 (関西大学 総合情報学部)　堀 雅洋 (関西大学 総合情報学部)

1. はじめに

安全・安心な地域生活 (セキュアライフ) を創出するには，安全を脅かす兆候を検知し，その状況を回避するための知識 (安全知) を獲得する必要がある．そのためには，状況に応じて必要な情報を収集する環境モニタリング技術の開発とモニタリングによって収集された環境情報と当該地域の地理情報とを共有するシステムの開発が不可欠である [1]．災害時の避難路確保を例にとれば，火災や黒煙によってその経路を確保することが困難な場合，避難者は，構造的が有する窓や扉，通路などの設計的環境情報と避難者の経験に基づいた危険回避のための知識行動とを共有的に考慮して安全知を獲得し，最適行動を決定する．すなわち，避難者は建造物に関する学習性と経験に裏打ちされた学習性によって協調的規範を構成する．

本研究では，簡易な迷路探索問題を取り上げ，環境情報に対する強化学習と人間の脳信号とを協調して学習する協調学習システムにより，災害時における避難者の避難行動の特性や規範について議論する．強化学習 [2] とは教師なし学習法の一手法である．エージェントは環境状態を観測し，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言われている [3]．一方，人間の脳信号を検知してコンピュータを制御する BCI(Brain computer interface) [4] の研究が推進されている．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，脳とコンピュータとの間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 [5]．

ここでは，強化学習と BCI 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し，迷路の経路探索に適用して，協調学習の有用性を議論する．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．

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，脳信号による脳示唆 $su(t)$ を行う．脳示唆を内部報酬とすることで効率の良い学習が実現され，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．

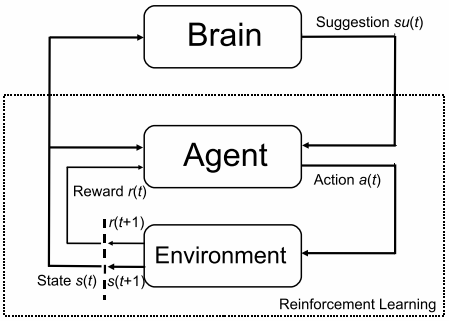

Fig. 1: Proposed Collaborative Learning System

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．

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

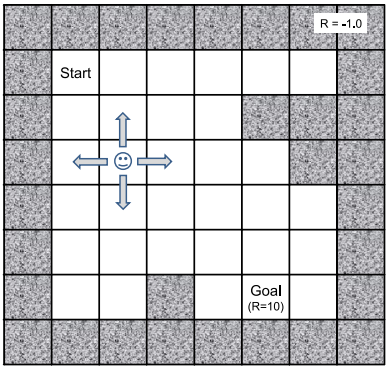

Fig. 2: Maze

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．

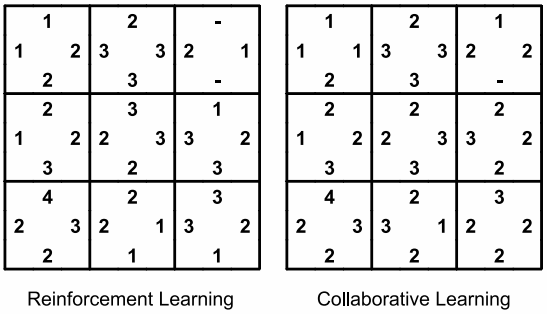

Fig. 3: Comparison of Learning Times

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．

以上の結果から，協調学習は強化学習と比較して最短ルールを効率的に探索するように経路を探索しており，災害時における避難路確保には協調学習が不可欠であるといえる．

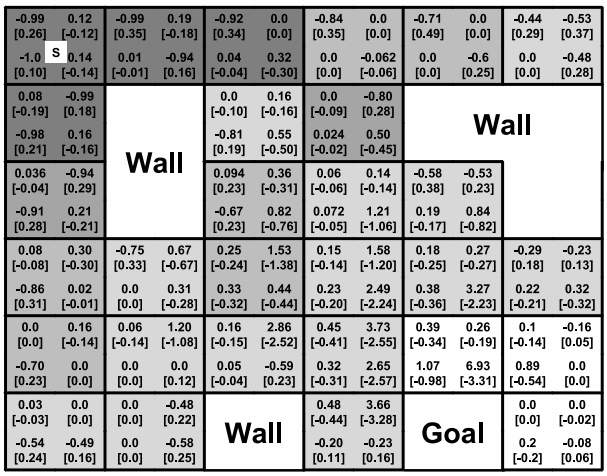

Fig. 4: Comparison of Q Values

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議論し，脳示唆の自動検出により最適経路を自動決定するシステムの実現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．

参考文献

[1] 関西大学 総合情報学研究センター：セキュアライフ創出のための安全知循環ネットワークに関する研究，平成 20 年度私立大学戦略的研究基盤形成支援事業 (2008)

[2] 伊藤：ロボットインテリジェンス，オーム社 (2007)

[3] 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)

[4] 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)

[5] 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)