

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 16 日現在

機関番号：24402

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24700148

研究課題名(和文)人の行動決定モデルに基づくエージェントを介した遠隔コミュニケーション

研究課題名(英文)Human Action Model Based on Multiple Agent for Remote Communication

研究代表者

田窪 朋仁(Takuto, Tomohito)

大阪市立大学・大学院工学研究科・准教授

研究者番号：80397695

交付決定額(研究期間全体)：(直接経費) 3,400,000円、(間接経費) 1,020,000円

研究成果の概要(和文)：本研究では、ネットワーク越しのコミュニケーションにおける情報伝達の遅延問題に対し、ユーザのコミュニケーション動作を一時的に代替することが可能な人の行動決定モデルを提案する。行動決定モデルは、人の行動理念に基づき短期報酬を期待した反射行動、累積報酬を期待した予測行動の2層構造で再現される。PC上に構築した簡易なゲームを対象に人の操作特性を獲得した結果、提案する2層構造で操作特性をよく獲得できていることを確認した。また、学習したモデルが独立してゲームを継続できていることから、人の特性を模倣したゲーム操作が獲得できており、一時的な時間遅れに対して人の操作を代替することが可能であることが確認された。

研究成果の概要(英文)：To solve the communication delay via network, the human action model which can replace a human is proposed. The model is structured by the two-layer. The first layer assumes reactive behavior, and it is modeled by the multiple agents. The agent consists of a linear quadratic controller and a linear predictive function. The second layer assumes predictive action with accumulative reward. The second layer indicates commitments to agents on the first layer to act strategical. The proposed human action model is implemented to simple games on PC and the control habit of the human skilled player is modeled well. Furthermore, the acquired model can play the same game instead of the human player. The human operation model is expected to make the communication via internet smooth.

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：強化学習 データマイニング 行動予測

1. 研究開始当初の背景

近年の通信インフラの整備により、情報を通信しゲームを行うネットワークゲームや遠隔地のロボットをコントロールする遠隔コミュニケーションの需要が高まっている。このような需要に対し、遠隔地との情報通信には、情報量の制限や通信時間の遅れなどがあり、円滑なコミュニケーションを行うことができない問題がある。例えば、通信を介した挨拶が時間遅れにより表示の遅れや動作の遅れとして相手に伝わることで、違和感のある動きがコミュニケーションを妨げたり、意図したとおりの動作が遠隔地のロボットで実行できないことで事故を起こす場合が考えられる。円滑なコミュニケーションをする場合、無数に存在する行動の選択肢の中から適切な動作を選び行動に移す必要があり、これらはコミュニケーションや遠隔地での動作応答の許容できる時間以内で実行されなければならない。情報通信の観点からリアルタイム性を維持するための研究が行われているが、プロトコルの変更だけでは根本的な遅れに対応したコミュニケーションの問題解決は実現することができない。そこで、本研究では、ユーザの行動選択を学習する「エージェント」を介した通信を行うことで、ユーザの次の行動選択を予測し、事前にエージェントが実際のユーザに代わり表示や動作を先行出力するシステムを提案する。このシステムは、学習したエージェントが操作者に代わり一時的にコミュニケーションの代行をするシステムであり、送信された情報を適切に選択し出力側の代替モデルが予測行動を実行する機能を実現する必要がある。

2. 研究の目的

ユーザの行動選択を学習し、ユーザの次の行動選択を予測することのできる行動決定モデルを確立することを目的とする。行動学習のモデルとして、人間の行動学習の基本機能をモデル化した階層型の強化学習を採用する。また、ユーザの入力情報の癖を取得した学習後のシステムが、操作者の入力情報から次の入力を予測可能であることを明らかにし、システムの操作が代替可能であることを示す。

3. 研究の方法

本研究では、操作における人間の行動理念は、短期報酬、累積報酬、行動の変容・収束という3者に分けることができると考え、図1のように、2層の階層モジュールに強化学習を当てはめて人間の行動決定を推定するシステムを構築することを考える。この構成により学習した行動分岐の判断・出力は、人間の考え方に近い形で過去の操作情報から次の操作を予測できると考えられる。

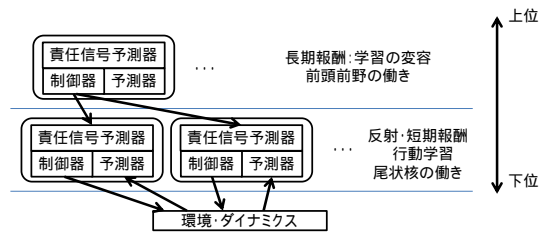


図1 人の行動決定を模倣した階層モデル

反射行動を学習する第1層は、人の判断を高速にシステムに反映する必要が有る。そこで、線形制御器と予測器を組み合わせたモジュールを配置し、第1層のみで強化学習を行うことで、人の反射行動を獲得させる。次に、第1層のモジュールを制御する第2層は長期報酬に基づきモジュールを学習させる。

4. 研究成果

モジュール切り替えにより人の反射的入力特性を再現する第1層のコントローラを学習により獲得する。本研究では各モジュールの予測器に線形予測、制御器に最適フィードバック制御を用いる。一般的に人間の行動特性は状況に応じて刻一刻と変わるものであり非線形な性質を示すが、これを線形の組み合わせにより再現することを考える。

図2にモジュール構成を示す。各モジュールはそれぞれ現在の状態に対する次の行動を予測する。現在観測される状態から各モジュールに対する適応度を示す責任信号を決定し、コントローラが出力する最終的な値に対する比率とする。各モジュールが予測した値を足し合わせたものが行動予測となる。

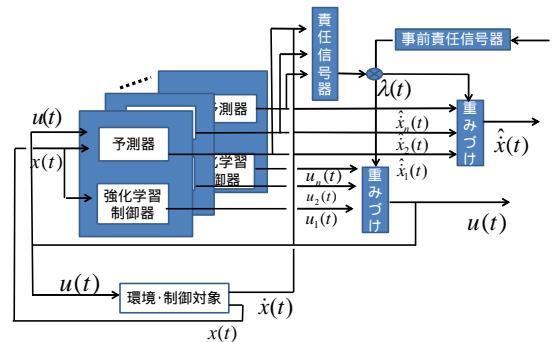


図2 第1層の反射行動予測モジュール

人の操作モデルは環境に依存した反射行動の組み合わせであるという仮定のもと、線形予測モデルの組み合わせにより人の入力特性を再現する。各モジュールで利用する予測-制御対のモジュールは線形制御理論を用いることで、高速な学習と予測を可能とし、複数のモジュールの組み合わせにより、単体では予測不可能な人の非線形な入力を再現する。

一方、人の入力を含めた最適制御によるコントローラを導出するためには、人の入力を含めた状態変数のみを含んだ状態方程式を

構築し、可制御性を保つシステム構成にすることで予測器と制御器の設計を行う必要がある。観測状態に関連する項目は複数モジュールの利点を利用して、担当範囲を分けることで状況に応じた人の操作を担当範囲毎に線形近似し特性を学習させる。人の非線形な行動特性はモジュールの個数と担当範囲、および、モジュールごとの制御出力の重ねあわせで再現されることを期待する。モジュールの領域変化についても、人の入力を含めた状態変数のみを更新の対象とすることでモジュールの担当領域の学習を行う。各モジュールの予測器、制御器の学習は実際の人の入力をもとに逐次最小2乗法を用いて行う。

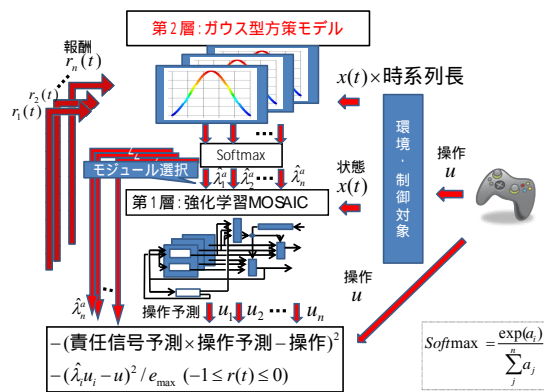


図3 2層の長期報酬を含んだ学習モデル

第1層で獲得した複数モジュールによる反射行動を長期的な視点で操作を最適化し、人の行動モデルを再現するための第2層を図3のように構成する。本研究では、第2層の学習モデルとして自然政策勾配法を採用し、多峰性ガウス分布により報酬の期待値を最大化する方策を強化学習により求める。

提案手法をピンポンゲームに当てはめ、操作する人間の特性を獲得する実験を行った。図4にゲーム環境の概要を示す。観測できる変数は人が操作するパドルの位置とボールの位置・速度とする。ボールがy軸方向に上昇してから壁にあたって戻ってくるまでの往復距離800の間に第1層のモジュールを8個均等配置し、各モジュールに人の反射行動を学習させる。第1層を学習し終えた操作モデルは反射行動のみの学習であるため、人を代替してゲームを行うと、ゲームの継続が困難であることが確認された。

次に、第2層の自然政策勾配法により、第1層の反射行動モジュール群を利用するためのタイミングを学習させる。第2層のガウス政策モデルは8個設定し、ボールが100往復するまでのゲーム操作のデータから人の操作モデルを学習する。学習させたモデルは300往復分のデータでの入力の正答率により評価した。その結果、2層のモデルで学習させた人の操作予測の正答率は60.3%であった。高い正答率ではないが、人に代替して学習後のモデルでゲームを行わせると長時間のゲーム継続が可能であることが確認され

た。また、特徴的な行動として、壁に反射して戻ってくるボールの落下点を予測して待機するなどの高度な動きが学習されていることが確認された。

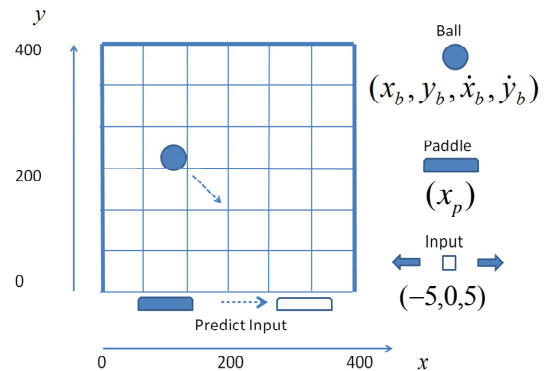


図4 ピンボールゲームの実験設定

以上の結果より、提案する2層の階層モジュールによる強化学習を当てはめた人間の行動決定モデルは、人を代替してエージェントが一時的に操作を補助するシステムが構築可能になる可能性を示すことができた。

今後は、人の操作の正答率を高めるための研究や、時間の遅延を考慮した人の操作を代替するタイミングを学習する研究を実施することで実用的な分野に発展していくことが考えられる。

5. 主な発表論文等

〔学会発表〕(計 4 件)

田窪 朋仁、石田 玄也、上野 敦志、過去の操作系列と状況変化の履歴に基づく隠れマルコフモデルを用いた人の操作予測、日本機械学会ロボティクス・メカトロニクス講演会、2A1-S6、2014年5月27日、富山

Genya Ishida, Tomohito Takubo, Atsushi Ueno, Human's Operation Modeling based on Hidden Markov Model with Consideration for the Operation Change and Timing, IEEE/SICE International Symposium on System Integration, 査読有、pp.210-215、2013年12月15日、神戸

稲野 拓巳、田窪 朋仁、上野 敦志、領域可変複数線形モジュールによる人の操作特性の獲得、計測自動制御学会システムインテグレーション部門講演会、1A1-5、2012年12月18日、福岡

稲野 拓巳、田窪 朋仁、上野 敦志、複数モジュールによる人の入力特性の獲得、日本機械学会ロボティクス・メカトロニクス講演会、1A2-S11、2012年5月28日、浜松

6. 研究組織

(1) 研究代表者

田窪 朋仁 (TAKUBO, Tomohito)

大阪市立大学・大学院工学研究科・准教授

研究者番号：80397695