

# 内部状態と行動をベースとした ロボットの内部状態予測手法の提案

克蘭工業大学 情報電子工学系専攻 認知ロボティクス研究室 杉本 大志

## 1. はじめに

動的な環境下で適切に動作するロボットの研究は、古くから数多く行われている<sup>[1,2,3]</sup>。例として家庭、工場等とあらゆる作業空間内で人間の代わりに活躍する、様々なロボットの開発が行われている<sup>[4]</sup>。しかしながら、動的な環境下に於いてロボットを人間の様に行動させる事は容易ではない<sup>[5]</sup>。我々人間は或る環境下で作業をする際、周囲の環境の変化や、自身の次の状態を予測し、それに合わせた行動を無意識の内に選択している。具体的には、過去の経験と記憶から作業をスムーズに行う為に、自分自身の姿勢や力の入れ具合を予測して、無意識の内に制御をしながら操作を行っている<sup>[6]</sup>。これは、あまりに多くの事を意識して行動していると、即座の対応が難しくなり、目的の行動を達成できず、『損をする』場合がある。例えば人間の歩行は、環境及び身体に関する感覚入力に応じて適切な調節がなされる結果、常にリズムカルで安定した動作となる。この時『脳』が意識すべきは歩行と環境であり、その環境に合わせて『個々の関節をどの程度変化させるのか』は無意識で行われる。この予測によって、普段の社会生活に於いて人間はボディバランスを制御し、転倒や接触等のリスクを低減している<sup>[6]</sup>。

同様にロボットの場合は、予測を行わない場合、行動選択に対する制御処理の負荷が大きくなる事が考えられる。特に、有限の計算資源や、ハードウェアが持つ物理的制約が存在する事から、ロボットが動作する際は、現状態に対し、制御遅れによってロボットの行動は1ステップ遅れる事になる。この為、人間の作業の代行として、ロボットが動的な環境下にて或る作業を実現する場合に於いても、人間同様に『状態予測からの行動決定』が重要である。これに関連した研究のアプローチとして、ロボットに対し機械学習を適用する事で、環境に適した行動を自律的に学習させるものが在る。

従来『予測』を取り扱う研究としては、株価に対して時系列予測を適用した株価予測の研究<sup>[11]</sup>や、飛行ロボットやロボットアームをプラントと見做した予測制御<sup>[12,13]</sup>、複数の状態予測モデルと報酬予測モデルの組合せによって出力が決定されるモジュール強化学習方式<sup>[14]</sup>が挙げられる。

また、“Online SVR”と呼ばれる、イタリアの Parrella が提案したバッチ SVR を発展させた手法も存在する<sup>[15]</sup>。これは従来使われてきたバッチ SVR<sup>[16,17]</sup> が持っていた問題点である、訓練データを全て揃えなければ学習出来ない(バッチ学習)、と言う点をオンライン学習により解決したものである。

但し、ここで挙げた先行研究の内、文献[11]は純粋に株価を時系列データとして捉え、それが未来でどのような状態に変化するかを予測している。また、文献[12,13]については、制御量と操作量について着目し、これらの状態の変化を予測し、フィードフォワード制御として予測結果を用いてプラントの制御を行っている。そして、文献[14]では、変化する環境や報酬に合わせて複数の状態と報酬の予測モデルを適宜切り替えながら学習制御を行っている。最後に文献[15]では、時系列データとして与えた一次元データを逐次与える事で、そのデータに対する回帰分析を行っている。つまり、これらの先行研究では、予測を間接的に用いた制御を行うか、純粋に予測そのものについて検討が為されているが、何れも状態に対する予測に焦点を当てたもので

はない。従って、従来の手法によって『状態予測からの行動決定』を実現する事は出来ない。

その理由として、ロボットは現在の環境の状態を感知し、一つの行動を実行する。状態と行動によって環境は新しい状態に移り、それに対して報酬をロボットに渡す。これらの相互作用を通して、ロボットは与えられたタスクを遂行する目的行動を学習する<sup>[18]</sup>。この為、学習で得られた経験を利用して行動するロボットの行動は、決定論に加えて、ロボットの行動が環境に作用する因果律にも影響される。例えば、自動車の動きそのものは物理法則に従うが、交差点で直進するか曲がるかを決めるのは運転手の都合だ。つまり運転手が選んだ行動が、次の状態及び環境に影響を与える事になる。この様にロボットは、状況に応じて複数の選択肢から一つを選択する。

以上の様に、ロボットの場合、自身が選択した行動によって、未来に決定される状態が無数に存在する為、未来の状態を予測する事が難しい。従って、ロボットの採った行動が未来に影響を与える事になると言う、ロボット特有の問題が存在する。この様な理由から、ロボットの為の『状態予測』の考え方が必要となる。

## 2. 提案手法

本研究では、現在の内部状態と行動から、未来の内部状態を予測する手法を提案する。これによって、自身の採る行動によって変化する内部状態を予測する事が可能となる(図1)。

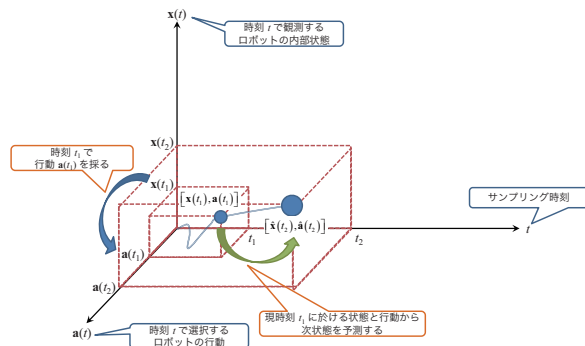


図1 内部状態と行動から未来を予測する考え方

### 2.1 内部状態予測の概要

この内部状態の予測を実現する上で、本研究では文献[18]で述べられている様に、状態(内部状態)と行動によって新たな状態に移る点を踏まえ、ロボットの現在の内部状態と採る行動から、次にロボットが遷移する内部状態を予測する。また、未来の内部状態は、現在の内部状態と行動の他にも、過去に経験した行動や内部状態にも依存するものと考えられる。そこで、内部状態の予測を行う為に、Online SVRを『学習器』として採用する。この“Online SVR”に対して現状態と行動を訓練セットとして与える事で、未来の内部状態の予測を実現する。

#### 2.1.1 逐次内部状態予測部

逐次内部状態予測部は、文献[19]によって定義されている逐次予測に、内部状態と行動の考え方を導入したものである。内部状態を  $\mathbf{x}_t$ 、行動を  $a_t$  とした時、 $\mathbf{z}_t = [\mathbf{x}_t \mid a_t]$  と訓練セット

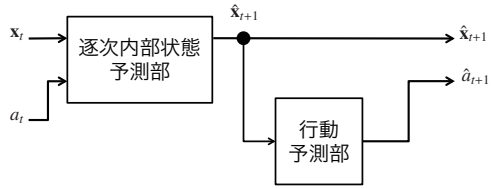


図2 内部状態予測システムの構成

を定義する事によって

$$\hat{x}_{t+1,i} = \begin{cases} 0 & \text{if } t = 0 \\ \Delta\theta & \text{if } t = 1 \\ \mathbf{k}_{sv}(\mathbf{z}_t)^\top (\mathbf{K}_{sv} + \lambda \mathbf{I}_l)^{-1} \mathbf{z}_{sv} + b'_i & \text{otherwise} \end{cases} \quad (2.1)$$

when  $i \in \dim \mathbf{x}_t$

で示される。これによって、次時刻に於ける内部状態  $\hat{\mathbf{x}}_{t+i}$  を予測する事が可能となる。

### 2.1.2 行動予測部

行動予測部は、逐次内部状態予測部で得られた、未来の内部状態を用い、この未来の内部状態に対して採るべき行動を決定するものである。予測対象となるモデルの動的挙動が、適切に離散時間での非線型状態空間モデルとして記述する事が出来るとするなら、予測した内部状態を、行動を決定する為の状態フィードバックゲイン  $\mathbf{k}_f$  に入力する事によって、新たに未来の行動  $\hat{\mathbf{a}}_{t+1}$  をも予測出来る様になる。つまり、

$$\hat{\mathbf{a}}_{t+1} = \mathbf{k}_f \hat{\mathbf{x}}_{t+1} \quad (2.2)$$

である。この結果を用いる事によって、任意の  $n$  サンプル時刻先までの内部状態の行動を予測出来る。

## 3. 検証実験

本研究では、自走式二輪型倒立振子をモデルとして、この制御応答に対し提案手法を適用した。図3は逐次内部状態予測によって予測した内部状態  $\psi$  (車体傾斜角)、図4は長期内部状態予測によって予測した内部状態  $\psi$  の予測結果である。尚、黒の実線が実データである。長期内部状態予測に於いては、行動

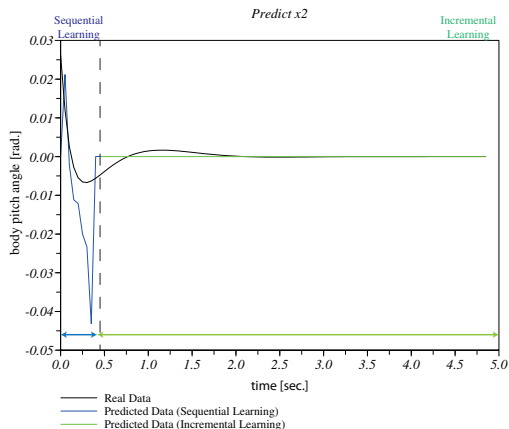


図3 車体傾斜角  $\psi$  に対する逐次内部状態予測

予測部に於ける  $\mathbf{k}_f$  を、最適レギュレータにより決定して倒立振子の内部状態が安定化される様にした<sup>[20]</sup>。図3では、外乱が混入しない場合は系が0に収束する様子予測している。図4からは、行動予測部に於ける  $\mathbf{k}_f$  によって、予測可能な外乱である正弦波が与えられても、未来に於いては安定化される様子予測し、鉛直近傍で振動している事が判る。

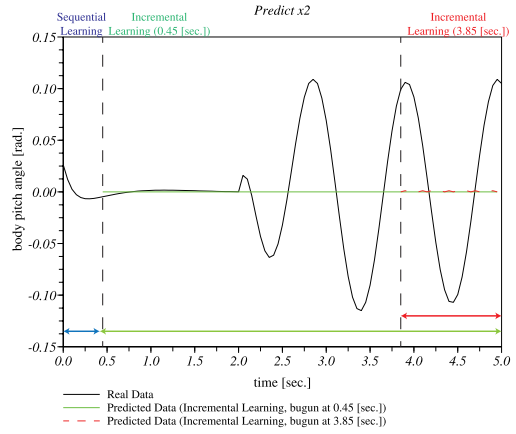


図4 車体傾斜角  $\psi$  に対する長期内部状態予測

## 4. 結論

本研究では、ロボットが採る動作に遅れが生じる点に注目し、内部状態と行動から未来の内部状態を予測する手法を提案した。また、実験によりその有用性を示した。

## 参考文献

- [1] Sebastian Thrun, Wolfram Burgard, Dieter Fox : Probabilistic Robotics (Intelligent Robotics and Autonomous Agents series), The MIT Press (2005)
- [2] 浅香俊一, 石川繁樹 : 動的な環境における自律移動ロボットの行動制御, 日本ロボット学会誌 Vol.12 No.4, pp.583-589 (1994)
- [3] 神田崇行, 石黒浩, 小野哲雄, 今井倫太, 前田武志, 中津良平 : 研究用プラットフォームとしての日常活動型ロボット "Robovie" の開発, 電子情報通信学会論文誌 D-I Vol.J85-D-I No.4, pp.380-389 (2002)
- [4] 国際ロボット連盟 : 産業用ロボットの明るい展望 ~ ロボットが百万人以上の雇用を創出 ~ (2011)
- [5] 十河卓司, 木元克美, 石黒浩, 石田亨 : 分散視覚システムによる移動ロボットの誘導, 日本ロボット学会誌 Vol.17 No.7, pp.1-7 (1999)
- [6] 西岡基夫, 岡田明, 宮野道雄, 森一彦, 山下久仁子, 中山圭以 : 繰り返し作業における操作方略の変化と動作特性に関する基礎的研究 - パフォーマンスデータと生理的データからみた評価方法の検討 -, 生活科学研究誌 Vol.7 居住環境分野, pp.45-55 (2008)
- [7] Thomas Mitchell : Machine Learning, The McGraw-Hill Companies (1997)
- [8] 銅谷賢治 : 臨時別冊数理学 SGC ライブラリ 60 「計算神経科学への招待」 脳の学習機構の理解を目指して, サイエンス社 (2007)
- [9] M.O. Abu-Shawiesh, F.M. Al-Athari, H.F. Kittani : Confidence Interval for the Mean of a Contaminated Normal Distribution, Journal of Applied Sciences, pp.2835-2840 (2009)
- [10] Teuvo Kohonen : Self-Organizing Maps, Springer (2000)
- [11] 植田英三郎 : 時系列解析による株価予測, 大阪府立大学経済研究 53(3), pp.95-111 (2007)
- [12] Jongho Shin, H. Jin Kim, Sewook Park and Youdan Kim : Model predictive flight control using adaptive support vector regression, Neurocomputing 73(4-6): pp.1031-1037 (2010)
- [13] Younggeun Choi, Shin-Young Cheong and Nicolas Schweighofer : Local Online Support Vector Regression for Learning Control, Proceedings of the 2007 IEEE International Symposium on Computational Intelligence in Robotics and Automation Jacksonville, FL, USA (2007)
- [14] 杉本徳和, 鮫島和行, 銅谷賢治, 川人光男 : 複数の状態予測と報酬予測モデルによる強化学習と行動目標の推定, 電子情報通信学会誌 D-II, 87(2), pp.683-694 (2004)
- [15] Francesco Parrella : Online Support Vector Regression, A Thesis presented for the degree of Information Science, Department of Information Science, University of Genoa, Italy (2007)
- [16] Corinna Cortes and V. N. Vapnik : Support-Vector Networks, Machine Learning, 20 (1995)
- [17] V. N. Vapnik : The Nature of Statistical Learning Theory, Springer, New York (1995)
- [18] 高橋泰岳, 浅田稔 : 実ロボットによる行動学習のための状態空間の漸次的構成, 日本ロボット学会誌 Vol.17 No.1, pp.118-124 (1999)
- [19] 杉本大志, 倉重健太郎 : オンラインサポートベクトル回帰におけるリアルタイム予測の検討, 2011年度精密工学会北海道支部学術講演会講演論文集, pp.59-60 (2011)
- [20] 杉本大志, 吉村斎, 阿部司, 大村功 : Scilab/Scicos による組み込みシステムのモデルベース開発に関する研究, 2010年度精密工学会春季大会学術講演会講演論文集, pp.343-344 (2010)