

RNN 予測器を用いた人物行動パターンの自動分節化

牛 童[†] 岩井 儀雄[†] 谷内田正彦^{††}

[†] 大阪大学大学院基礎工学研究科

〒 560-8531 大阪府豊中市待兼山町 1-3

E-mail: niu@yachi-lab.sys.es.osaka-u.ac.jp, iwai@sys.es.osaka-u.ac.jp

^{††} 大阪工業大学情報科学部情報メディア学科

あらまし 近年, 犯罪の多様化により, セキュリティの強化に対する意識が高まっている. しかし, 一般的な防犯カメラ監視システムでは, 監視者の負担による作業ミスや高コストなどといった問題点がある. そこで, 人の行動を認識し, 未知の行動を例外行動として自動検出するようなシステム開発が求められている. 例外行動を検出するためには, まず, 正常の行動パターンを学習し, 行動モデル構築する. 次にモデルの適合度を評価し, 例外行動検出を行う. 行動モデルを構築する際, 行動パターンが多岐に渡り, 行動軌跡全体を学習するには学習コストが大きいという問題点があった. そこで, 本研究ではより効率的に分割を行うため, RNN(Recurrent Neural Network) を用いて人物行動モデルを構築する方法を用いる. RNN の予測可能性に基づき, 分節点を切り出す手法を提案する.

Automatic Segmentation of Human Motion Patterns using Predictor RNN

Tong NIU[†], Yoshio IWAI[†], and Masahiko YACHIDA^{††}

[†] Graduate School of Engineering Science, Osaka University

1-3 Machikaneyama-cho, Toyonaka, Osaka, 560-8531 Japan

E-mail: niu@yachi-lab.sys.es.osaka-u.ac.jp, iwai@sys.es.osaka-u.ac.jp

^{††} Faculty of Information Science and Technology, Osaka Institute of Technology

Abstract Recently, diversity of crimes has become a major social problem, calling for reinforcements in security systems. The objective of the author's research is to create a system that automatically detects unknown behaviors as exceptions. Such systems require retraining of the whole system when supplementing a new behavior into the model. In this research, the authors aim to segment the behaviors, based on predictability of the recurrent neural network, for retraining only the sub-model for supplementation. In this paper, the authors present the results of the segmentation experiment using proposed method.

1. はじめに

近年, 防犯に対する意識が高まり, カメラによる監視システムが広く利用されている. しかし, 人による動画からの異常検出は目視という単純作業による疲労や

怠惰などから起こる人為的なミスや, コスト, 効率の面で問題がある. そのため, より監視効果を高め, 人件費コストを削減するため, 異常行動を自動的検出し, 通報するシステムの開発が求められている [1], [2]. このような自動監視システムには一般的に学習段階と認識段階の

二つから構成される。学習段階では、正常行動を学習し、行動モデルを構築する。認識段階では、観測した行動に対して、行動モデルの適合度を計算し、適合度が低いモデルに不適合な行動を例外行動として検出する。

行動モデルを構築する際、学習データに対して、意味のあるまとまりを時間軸に区切る必要があり、これは分節化と呼ばれる。一連の行動を部分行動に分割し、類似行動を統合して学習することで、学習コストを減らすことができる。一般的に行動の分節箇所を手動で設定する場合が多く、システム構築者にとって、手間のかかる作業である。分節する箇所は静止点で設定する研究[3]があったが、人物の行動は静止点で分節化できるものと、静止点ではなくても分節化できるものがある。より効率的に学習を行うため、行動の時系列データの自動的に分節化の技術が重要である。自動分節化の研究に関して、体の特徴点の速度変化、空間的位置関係によって分節点を切り出す手法は一般的である。指揮の手の運動に対して、速度の極小点と方向変化の極大点を基準に分節する手法[5]、舞踊動作における両手、両足の重心の速さと音楽リズムから分節の候補点を絞り込む方法[6]が提案されている。しかし、これらの手法は反復動作や周期的な運動にしか適応しないという欠点がある。

行動の自動分節化に関する研究は、ロボティクスの分野でも行われており、時系列学習モデルであるリカレントニューラルネットワーク (Recurrent Neural Network, 以下 RNN と略記) によるロボットの部分行動分節化が提案されている[9]。分節化・予測機構を統合的に学習させ、現在の観測情報及び予測情報による行動生成を行われた。本研究ではこのモデルに基づき、RNN の予測(出力)誤差をもとに人物行動パターンの自動分節化を目標とする。RNN を学習器と予測(認識)器の両方として用い、行動モデルを構築する。予測可能性に基づき分節の指標を算出し、追加学習行動を既知のパターン(予測可能)と新しいパターン(予測不可能)の二つに分ける。予測不可能部分について RNN で学習を行うことで、学習コストの削減が期待できる。

2. 提案システム

本研究では RNN 予測器を用いて行動モデルを構築する。RNN の予測可能性に基づき行動パターンの自動分節化を行う。以下に本研究で目標とするシステムの構築方法について述べる。まず、RNN 予測器を説明し、次に自動分節化の手法と行動モデルの構築について述べる。

2.1 RNN 予測器

RNN はパターン識別器だけでなく、パターン予測器としても用いることができる。時系列 $S(t)$ 番目の

入力に対して $S(t+1)$ 番目を予測するようにサンプルデータから学習を行う。学習方法としては BPTT(Back Propagation Through Time)[12]を用いて、入力に対して出力誤差が小さくなるように、各層の結合重みを修正していく。重みの更新はニューラルネットワークと同様に最急降下法によって行われる。RNN を予測器として用いた認識手法として、音声認識[10]やロボットの行動生成[11]などが挙げられる。ロボットの行動生成においては、Haruno ら[11]がモデル学習時に行動軌跡を分割し、各分割軌跡を独立に予測器で学習し、各予測器を上位モジュールで制御することで追加学習を容易に行えるシステムを提案している。

本研究は Jordan 型 RNN[13]を用いて人物行動パターン学習を行う。RNN は階層型ニューラルネットワークに回帰構造を付け加えたものであり、回帰結合の値によってそれまでの入力の履歴を代表させることができる。そのため、回帰結合のない階層型ニューラルネットワークとは異なり、時系列入力を自然な形で扱うことができる。

図1に示すように、入力層 (Input Layer)、中間層 (Middle Layer)、出力層 (Output Layer)、コンテキスト層 (Context Layer) の4つの層から構成されている。観測特徴量は入力層に入力され、この入力値によって、どの時系列特徴量を生成するのが選ばれる。コンテキスト層へは出力層からの特徴量がフィードバックされている。コンテキスト層のニューロンの特性は出力値の1次遅れ、あるいは2次遅れのフィルタになっているので、過去の出力履歴をこの層で帰国することができる。入力層とコンテキスト層の出力は中間層、出力層へと伝わった出力が得られる。さらに、この出力は再びコンテキスト層へフィードバックされる。このように、RNN の内部にコンテキスト層 → 中間層 → 出力層 → コンテキスト層のループをもつので、出力層に次々と時系列観測特徴量を出力することができる。

2.2 自動分節化

本節では RNN の予測可能性に基づく分節化について

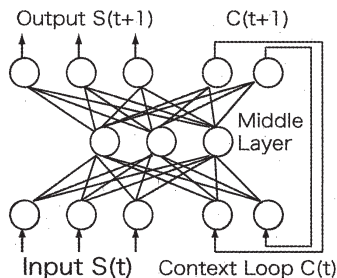


図1 RNN の構成

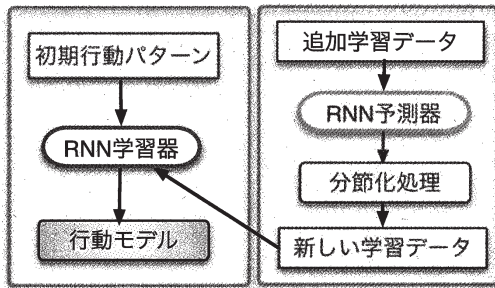


図2 システムの処理流れ

て説明する。行動モデルを構築する際、逐次的に新しい行動パターンを追加学習する必要がある。そこで、追加学習データに対して予測を行い、これまで構築した行動モデルにどこまで適応するかを測る。予測誤差が閾値を越えるところを検出し、そこで追加学習データを前半（予測可能）と後半（予測不可能）の二つに分ける。予測不可能部分を新しい RNN で学習し、予測可能部分はもとの RNN 学習器に入れ、再学習を行い、新しい行動モデルを作る。追加学習を含めた行動モデル構築の全体的な流れは図2に示す。処理の手順は主に以下の4ステップとなっている。

- (1) 学習：初期行動（ n パターン）を RNN で学習し、初期行動モデル M_n を構築する。
- (2) 予測：追加学習データ $NewData$ に対して、予測を行い、分節点を検出する。
- (3) 分節：追加学習データ $NewData$ を予測可能な $NewData_{known}$ と予測不可能な $NewData_{unknown}$ の二つに分ける。
- (4) 再学習：予測不可能の $NewData_{unknown}$ に対して RNN で学習し、新しい行動モデル M_{n+1} を構築する。

本研究では予測可能性を定量的に評価することで自動分節化を行う。観測された未知の行動パターンを予測する時、RNN に現在の観測特徴 $S(t)$ を入力することで次時刻の予測観測特徴量 $S'(t)$ が出力される。この予測値と次の観測特徴量 $S(t+1)$ を比較することで予測可能性を判定する。精度を高めるため、数ステップ分の予測値と観測特徴量を(1)式で評価し、閾値より大きくなると予測不可能とする。ここで、 f は評価関数、 T_e は閾値、 h は評価するステップ数、 $S(t)$ は観測特徴量、 $S'(t)$ は RNN の出力である。

$$f(S(t+1-h), S'(t-h), \dots, S(t+1), S'(t)) > T_e(1)$$

本稿では簡単のため、評価関数は各ステップにおける予測値と観測特徴量の差の絶対値を取る。また、これらの差があらかじめ設定したステップ分、閾値 T_e より大きくなると予測不可能として判定する。すなわち、式(2)が満たされるところを分節点とする。

$$\lim_{h \rightarrow 0} |S(t+1-h) - S'(t-h)| > T_e \quad (2)$$

2.3 行動モデル構築

分節化された行動パターンに対して、本研究では郷古ら[9]が用いる Mixture of experts (以下 MoE と略記) モデルに基づき行動モデルの構築を行う。MoE は階層型のネットワークであり、入力データ空間をいくつかの小領域に分割し、その分割された各領域に対してひとつのニューラルネットワークを割り当てることによって複雑な問題の解を求めるため手法である。複雑な問題を分割して小領域に区切ることによって、1つの大きなニューラルネットワークを学習させるよりも効率の良い学習をさせようというモデルである。提案手法では一連の行動系列をいくつかの部分行動に分節し、その分節された各部分行動にひとつの RNN を割り当てモデル学習する。各行動パターンは部分行動間の遷移によって表現される。

人物の行動モデルを考える際、複雑な一連の行動も、単純な部分的パーツの組み合わせから構成されると考えられる。例えば、食べ物を口に入れるという行為は、箸を持つ、食べ物を掴む、食べ物を口に入れるといった三つのパーツに分解できると考えられる。これらのパーツが部分行動にあたる。多様で複雑な行動は、手持ちのいくつかの部分行動をうまく組み合わせることにより、構成できると考えられ、これを行動の分節化という。行動を分節化し、類似行動を同一の部分行動として統合、それぞれを独立に別々の RNN で学習する。ここで、RNN で学習された部分行動のことをサブモデル (Sub-Model, SM) と呼ぶ。一連の行動を部分行動に分割し、学習することで学習コストを削減することができる。例えば、図3では、サブモデルが3つあり、行動パターンの数は2である。サブモデル SM_0 で分節化を行い、 SM_0 から SM_{1a} へ遷移するような行動がパターン1であり、 SM_0 から SM_{1b} へ遷移するような行動はパターン2である。このように、行動パターン1とパターン2で共通のサブモデル SM_0 を用いることにより、学習コストを減らすことができる。

RNN で学習し、新しい行動モデルを構築する自動分節化の例を図4、5に示す。これまで学習済みのモデル構造図4に示すように、 SM_0 から SM_{1a} , SM_{1b} に分節し、 SM_{1b} から SM_{2a} , SM_{2b} を分節し、5つのサブモデルが3つの行動パターンを表す。この行動モデルに

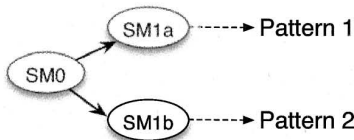


図3 サブモデルの遷移

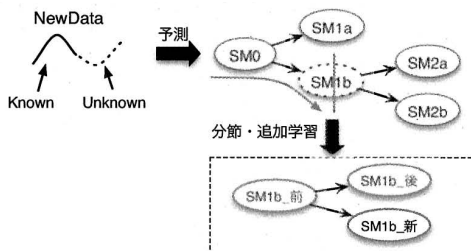


図4 分節化・追加学習処理の例

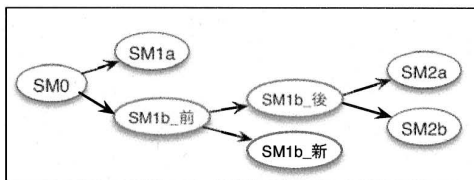


図5 新しい行動モデル

対して、新しい学習データが入るとする。この行動モデルで予測を行うと、サブモデル $SM1b$ の中間で予測不可能になった、つまり予測値と実際の学習データの差が大きくなったとする。この分節点で新しい学習データを予測可能な部分と予測不可能の部分に分ける。予測不可能の部分进行学习し、新しいサブモデル $SM1b_{\text{新}}$ を作製する。それと同時に学習済みの $SM1b$ を $SM1b_{\text{前}}$ と $SM1b_{\text{後}}$ に分割し、分割された $SM1b_{\text{前}}$ と $SM1b_{\text{後}}$ について再学習する。最後に新しいサブモデル $SM1b_{\text{新}}$ と $SM1b$ の後半部を $SM1b$ の前半部につける。すると新しい行動モデル (図5) はサブモデルが7つ、4つの行動パターンを表すことができる。このように分割された箇所のみ再学習自動分節をすることで、分節に関係ない $SM0$, $SM2a$ と $SM2b$ を考慮しなくてもいいため、学習コストを削減した追加学習ができる。

3. 分節化実験

以上に述べた方法を用いて、実際の動きデータを分節化した結果を示す。今回は人がオフィス内の行動を想定し、3種類の行動パターンを収集した。動きデータは

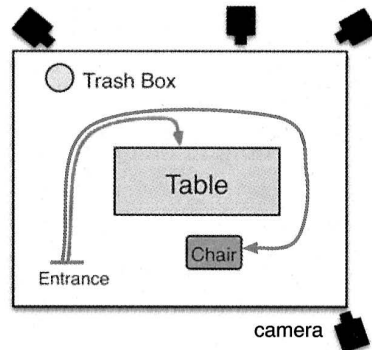
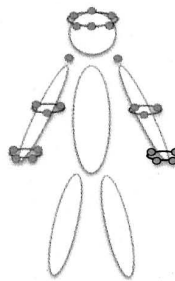


図6 実験室のレイアウト



(a) マーカの位置



(b) 実際の装着

図7 マーカ装着の様子

モーションキャプチャーにより計測され、各部位のマーカ一群の中心点を求め、その中心点を各部位の特徴点のモーションデータとする。特徴点として、右肘、左肘と頭の3点を選び、合計9次元をRNNの入力とする。学習段階では、まず $P1$ を初期行動パターンとして学習し、次に、 $P2$, $P3$ の順に追加学習を行い、図8に示すサブモデル $SM0 \sim SM4$ が5つを含む行動モデルを構築する。モデルの認識精度を評価するため、テストデータに対する予測誤差を測った。

同じ行動パターンであっても、被験者によってモーションデータに個人差 (身長、体型など) がある。さらに、同じ被験者でも、試行ごとに差が生じてしまう。そこで、被験者毎に行動モデルを構築し、未学習の試行に対して、どれくらい正しく認識されるかを調べた。次に、一人の被験者のモーションデータを用いて、学習を行い、行動モデルを作り、他の被験者のモーションデータに対して予測を行った。未学習の被験者行動の予測誤差を測ることによって、分節化により構築された行動モデルの予測性能と汎化性を評価する。

3.1 モーションデータの取得

PhaseSpace [14] という光学式モーションキャプチ

P1:お茶を運ぶ行動



P2:新聞を読む行動



P3:ノートを整理する行動



図 8 3種類の行動パターン

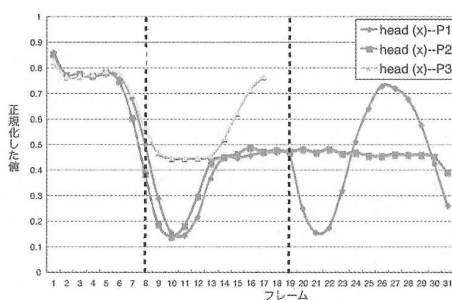


図 9 モーションデータ (頭の x 座標)

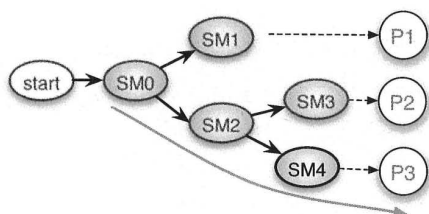


図 10 行動モデル

ャーシステムを用いて、時間軸に沿った動きのモーションデータを取得した。被験者の上半身にマーカーを 22 点 (頭:6 点, 肩:2 点, 両肘:6 点, 両手首:8 点) を付け、マーカー装着の様子を図 7 に示す。各マーカーの位置情報 (x,y,z) を実験室に設置させた 4 台のカメラから読み取る。カメラの測定範囲は 4m×4m である。

人がオフィス内での行動を想定し、実験室を図 6 のように設置した。お茶を運ぶ行動 P1, 新聞を読む行動 P2 とノートを整理する行動 P3, 3 種類の行動パターンを取得した。実験の様子を図 8 に示す。図 7 のグラフは P1,P2,P3 のモーションデータのうち、頭の x 座標値をプロットしたものである。グラフからみると、フレーム

8 まで 3 つの行動は同じような動きをすることがわかる。フレーム 9 からは P1 は P2 と P3 から分岐され、フレーム 9 を分節点とする。また、フレーム 19 において、P2 と P3 を分節することができる。このように、2 回の分節処理により、3 つの行動を 5 つのパーツに分けられる。これにより、行動パターン P1, P2,P3 の順に学習する際、図 8 に示すサブモデル SM0~SM4 が 5 つを含む行動モデルを構築することができる。

P1 (お茶を運ぶ行動): 入室 → お茶をテーブルの上に置く → 退室

P2 (新聞を読む行動): 入室 → 椅子に座る → 新聞をめくる → 手元のゴミをゴミ箱に投げる

P3 (ノートを整理する行動): 入室 → 椅子に座る → ノートを整理する → 一冊のノートをゴミ箱に捨てる

3.2 実験結果

3.2.1 同一人物による未学習同一行動の予測精度

被験者毎に行動モデルを構築し、モデルの認識精度を測った。被験者は女性 2 名, 男性 4 名, 計 6 名である。各被験者は、各行動パターンにつき実験を 4 回ずつ行った。その 4 回のうち、1 つを学習データとし、次に残りの 3 つをテストデータとする。行動パターン P1, P2,P3 の順に学習を行い、図 8 に示す行動モデルを構築する。構築した行動モデルに対して、未学習の試行をテストデータとして入力し、行動モデルの予測誤差を調べ、その結果を表 1 に示す。表 1 によると、被験者によって、モデルの認識精度がばらつきがあることがわかる。同一人物の同一行動に対して、予測誤差が最大 17cm であることが示された。

3.2.2 異なる人物による未学習同一行動の予測精度

本稿では、未学習の被験者の動きを未学習行動と呼ぶ。Jack-Knife 法を用いて、未学習行動に対して予測を行い、予測誤差を調べた。各行動パターンの平均予測誤差を表 2 に示す。図 11 のグラフは未学習行動が P3 のときの予

表 1 同一人物の予測誤差

被験者 (性別)	P1(cm)	P2(cm)	P3(cm)
1 (男)	8.65	8.09	12.04
2 (男)	12.53	6.50	10.59
3 (男)	9.03	12.28	12.30
4 (女)	9.50	8.08	11.82
5 (男)	8.87	7.05	12.02
6 (女)	11.59	10.81	17.24
平均	10.03	8.80	12.67
分散	2.64	5.11	5.37

表 2 異なる人物の予測誤差

未学習の被験者	P1(cm)	P2(cm)	P3(cm)
1	9.47	14.03	13.64
2	8.05	11.25	11.52
3	8.63	8.15	11.07
4	11.08	12.50	15.19
5	8.04	9.52	16.60
6	10.87	7.99	13.03
平均	9.36	10.57	13.51
分散	1.85	5.95	4.51

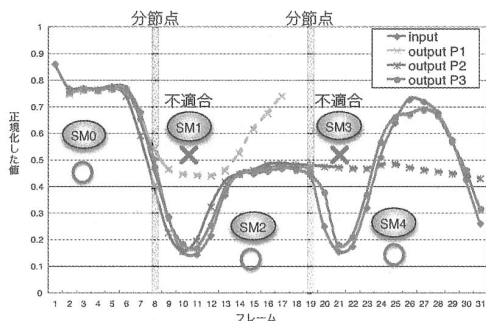


図 11 予測結果の例

測結果を示している。図 11 により、本手法によってモデルの分節化が可能であることが分かる。このグラフは input (頭の x 座標値) に対して、各行動パターンでの予測値 (output P, output P2, output P3) をプロットしたものである。SM0 で予測する際、すべての output と input の差が小さく、P1, P2, P3 全てのモデルに適合する。フレーム 9 にて、SM1 の出力値は input から大きくはずれ、P1 は不適合になる。また、同様にフレーム 19 において、P2 は不適合となる。input に適合するのは SM0→SM2→SM4 のサブモデル系列であり、つまり、その未学習行動は行動 P3 と認識される。これにより、未学習の行動に対しても正しく予測していることがわかる。

3.2.3 考察

同一人物の同一行動で未学習行動については、表 1 のような予測誤差がある。一方、違う人による同じ行動では、表 2 のような予測誤差になる。表 1 と表 2 の予測誤差の平均と分散を比べてみると、これらの誤差の間にはほとんど差がないことがわかる (予測誤差の平均の差: 0.67cm ~ 1.77cm)。これによりモデルの予測性能は個体差 (身長や行動の違いなど) に影響を受けないことが示された。

4. まとめと今後の課題

本研究では人物行動モデルの自動分節化を目標としている。そのアプローチとして、RNN を用いて人物行動モデルを構築する方法を用い、RNN の予測可能性に基づき、分節点を切り出す手法を提案した。実験により、3 種類の行動パターンに対して自動的に分節化することができた。

今後はより複雑な行動パターンに対する本手法の有効性を確認し、追加学習の性能評価を行う予定である。

文 献

- [1] 中村元美, 河野賢治, 三好雅則, 五十嵐芳治, 山口伸一郎, 会田敬一, “昇降機の安全・安心を提供する遠隔監視システム”, 人と地球のためのまちづくりと日立グループの都市開発技術, Vol. 90, No.09, pp. 742-745.
- [2] 南里卓也, 大津展之, “複数人動画像からの異常動作検出”, 情報処理学会論文誌, Vol. 46, No.SIG 15.
- [3] 北橋忠宏, 福永邦雄, 小島篤博, 長田典子, 樋口未来 “人間行動とその対象物体との機能に着目した協調的認識機構と認識結果の自然言語表現”, 「特定領域研究」17 年度報告書.
- [4] 青木茂樹, 岩井嘉男, 大西正輝, 小島篤博, 福永邦雄: “人物の位置・姿勢に注目した行動パターンの学習・認識と非日常状態検出への応用”, 電子情報通信学会論文誌 (D-II), Vol.J87-D-II, No.5, pp.1 083-1093, May 2004.
- [5] Tian-Shu Wang, Heung-Yeung Shum, Ying-Qing Xu, and Nan-Ning Zheng, “Unsupervised Analysis of Human Gestures”, *Proceedings of the Second IEEE Pacific-Rim Conference of Multimedia*, no.10, pp. 174-181, 2001.
- [6] 白鳥貴亮, 中澤篤志, 池内克史. “モーションキャプチャと音楽情報を用いた舞踊動作解析手法”, 電子情報通信学会 D-2, Vol.J 88-D-2 No. 8, pp. 1583-1590
- [7] 大川和宏, 中村仁彦, “統計的モデルを用いた運動認識用階層構造の自動構築”, 第 12 回ロボティクスシンポジウム, pp 418-423, 2007.
- [8] 門根秀樹, 中村仁彦, “パターンの相関と連想記憶に基づく運動パターンの分節化・記憶・抽象化”, 第 24 回日本ロボット学会学術講演会, 1D32, 2006.
- [9] 郷古学, 伊藤宏司, “環境の分節化と予測に基づく内部モデルの獲得,” 第 19 回自律分散システム・シンポジウム資料, pp. 43-48 2007
- [10] 内山 徹, 高橋治久: “リカレントニューラル予測モデルを用いた不特定話者単語音声認識”, 信学論 D, Vol. J83-D2, No.2, pp. 776-783, 2000.
- [11] M. Haruno, D. Wolpert, and M. Kawato: “MO-SAIC Model for Sensori-motor Learning and Control”, *Neural Computation*, 13, pp. 2201-2220, 2001.
- [12] D. Rumelhart, G. Hinton, and R. Williams: Learning internal representation by error propagation, In D.E.Rumelhart and J.L.McClelland, eds., *Parallel Distributed Processing*, MIT Press, Cambridge, MA, 1986.
- [13] M. Jordan: “Attractor Dynamics and Parallelism in a Connectionist Sequential Machine”, Eighth Annual Conference of the Cognitive Science Society, pp. 513-546, 1986.
- [14] <http://www.nihonbinary.co.jp/Virtual/PhaseSpace/index.I>