

NAIST-IS-MT9651022

修士論文

運動パターンの視覚情報に基づく学習・予測・認識
- リカレントネットワークを用いた実験的検討 -

小川 忠

1998年3月20日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学)授与の要件として提出した修士論文である。

小川 忠

指導教官： 横矢 直和 教授
東倉 洋一 教授
鹿野 清宏 教授
竹村 治雄 助教授
銅谷 賢治 助教授

運動パターンの視覚情報に基づく学習・予測・認識

- リカレントネットワークを用いた実験的検討 -*

小川 忠

内容梗概

物体や生体の運動を視覚的情報より学習・予測・認識するモデルをエルマン型ネットワークにより構成し、複雑な軌道を持つ複数の運動パターンを用いて計算機実験を行なった。特に、本研究では人間の動作から得られた時系列データを用いて、ネットワークの持つ短期予測能力、長期予測能力、動作クラスの識別能力、視点方向変化に対する汎化能力、複数の視点から見た運動パターンの学習・予測能力、時間変化に対する適応性を分析検討した。

キーワード

リカレントネットワーク, 視点汎化性, 運動予測, 動作識別, 時間適応性

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9651022, 1998年3月20日.

Visual Learning and Prediction of Motion Patterns*

Tadashi Ogawa

Abstract

We studied Elman-type recurrent neural networks for predicting and classifying spatio-temporal visual patterns. Computer experiments, using the complex temporal data of human arm movements, demonstrated that the network model has the following abilities; 1)short-term prediction, 2)long-term prediction, 3)learning multiple patterns from different viewpoints, 4)motion pattern classification, 5)view generalization, and 6)temporal adaptation for time scaling.

Keywords:

recurrent neural networks, view generalization, spatio-temporal prediction, motion pattern classification, temporal adaptation

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9651022, March 20, 1998.

目次

1. はじめに	1
2. 研究の背景	2
2.1 Biological Motion	2
2.2 View-based アプローチ	3
2.3 運動パターンの認識：関連研究	5
2.4 学習に基づく運動パターンの認識と予測	6
3. ニューラルネットワークを用いた時系列予測	7
3.1 ニューラルネットワーク	7
3.1.1 学習モデル：誤差逆伝播法	10
3.1.2 汎化能力	12
3.1.3 Quickprop 法：誤差逆伝播法の拡張	14
3.2 ニューラルネットワークと時系列処理	15
3.3 エルマン型ネットワーク	16
3.3.1 エルマン型ネットワークでの時系列処理	17
3.3.2 エルマン型ネットワークの学習	18
4. 計算機実験 1：単振り子・二重振り子	20
4.1 実験の目的	20
4.2 計算機実験	21
4.2.1 予測方法	21
4.2.2 単振り子・中間層の素子数の決定	22
4.2.3 単振り子・3層ネットワークとの比較	23
4.2.4 二重振り子	26
5. 計算機実験 2：人間の日常動作	30
5.1 実験データ	30
5.2 単一視点方向の予測実験	33
5.2.1 短期予測	33

5.2.2	長期予測	37
5.3	複数視点方向に対する予測実験	39
5.3.1	短期予測	39
5.3.2	長期予測	41
5.4	識別実験	43
5.5	コンテキスト層の自己回帰係数	45
5.5.1	自己回帰係数と短期・長期予測	45
5.5.2	パターン途中からの入力	48
6.	まとめと今後の課題	50
6.1	計算機実験のまとめ	50
6.2	今後の課題	51
6.2.1	濃淡画像を用いた計算機実験	51
6.3	モジュール構造化	52
	謝辞	53
	参考文献	55
	付録	63
	A. 単振り子・二重振り子の軌道算出	63

目 次

1	MLD: Moving Light Displays	2
2	McCulloch-Pitts の形式ニューロン	7
3	シグモイド関数	8
4	3 層ネットワーク	9
5	ニューラルネットワークの汎化性	13
6	エルマン型・ジョーダン型ネットワーク	17
7	エルマン型ネットワークの学習	19
8	単振り子・二重振り子	20
9	短期予測・長期予測	21
10	エルマン型ネットワーク：中間層の素子数と予測誤差	22
11	単振り子 (中間層素子数：4,8 個，学習：2000 回)	23
12	3 層ネットワーク・エルマン型ネットワーク	24
13	3 層ネットワークの学習曲線	24
14	単振り子テスト結果 (学習：2000 回，短期予測)	25
15	単振り子テスト結果 (学習：2000 回，長期予測 (9 ~ 16))	26
16	エルマン型ネットワーク (入出力層 4 個，中間・コンテキスト層 8 個)	27
17	二重振り子テスト結果 (エルマン型ネットワーク，学習：2000 回) .	28
18	二重振り子長期予測 (エルマン型ネットワーク，学習：2000 回) . .	29
19	動作クラス 1(コップの水を飲む)	30
20	動作クラス 2(ドアをロックする)	31
21	視点方向 (View Angel)	31
22	実験データ：動作クラス 1(90 度方向)， $(0 \leq x, y \leq 1)$	32
23	視点方向の変化：動作クラス 1	33
24	学習曲線：動作クラス 1(90 度)	33
25	動作クラス 1(90 度) の短期予測	34
26	時間方向に対する予測誤差 (動作クラス 1：40 度，90 度)	35
27	視点方向と時間方向に対する予測誤差 (動作クラス 1：90 度)	36
28	視点方向と時間方向に対する予測誤差 (動作クラス 1：40 度)	36
29	視点方向に対する平均予測誤差 (動作クラス 1：0・40・90 度) . . .	37

30	動作クラス 1 : 90 度の長期予測 ($0.3 \leq x, y \leq 1.0$)	38
31	時間方向に対する長期予測誤差 (動作クラス 1 : 90 度)	38
32	視点方向と時間方向に対する短期予測誤差 (40 度・90 度同時学習) .	39
33	時間方向に対する短期予測誤差 (40 度・90 度同時学習)	40
34	視点方向に対する平均短期予測誤差 (0・40, 40・90 度同時学習) .	40
35	視点方向と時間方向に対する長期予測誤差 (40 度・90 度同時学習) .	41
36	時間方向に対する長期予測誤差 (40・90 度同時学習)	42
37	視点方向に対する平均長期予測誤差 (0・40, 40・90 度同時学習) .	42
38	識別実験 (学習 : クラス 1(90 度) / テスト : クラス 2(90 度))	43
39	時間方向に対する予測誤差 (クラス 1(90 度) / クラス 2(90 度)) . . .	43
40	視点方向と時間方向に対する予測誤差 (クラス 1(90 度) / クラス 2) .	44
41	$y = \alpha^t$, $\alpha = 0.25, 0.5, 0.75$	45
42	自己回帰係数の違いによる短期予測誤差	46
43	動作クラス 1(90 度) の軌道	46
44	コンテキスト層の効果 : パターン 17 の短期予測	47
45	自己回帰係数の違いによる長期予測誤差	47
46	パターンの途中からの入力・短期予測 (40 パターン)	48
47	パターンの途中からの入力・短期予測 (20 パターン)	49
48	32 × 32 画素の濃淡画像データ	51
49	識別システム (モジュール構造)	52

1. はじめに

複雑な運動パターンを視覚情報から認識し対象の動きを予測することは、生物にとって必要不可欠な能力である。例えば、敵への攻撃、獲物の捕獲、さらに敵からの逃走の際には、相手の動きを予測してそれに備える必要がある。また、我々の日常生活においても、スポーツでのボールの軌道予測、対戦相手の動きの認識・予測、車の運転における状況予測、コミュニケーション時の相手のジェスチャー認識など、視覚情報に基づき複雑な動きパターンの認識・予測は繁栄に行なわれている。

本論文では、以下において生物にとって必要不可欠な視覚情報に基づいた対象の動き(運動パターン)の認識・予測能力を、学習によって獲得するニューラルネットによる計算モデルを提案し、さらに計算機実験により能力を確認し、その有用性を示す。

各章の概略は以下の通りである。次章において本研究の研究背景を示す。ここでは運動パターンの認識に関する関連研究を紹介するとともに、本研究のアプローチ方法を述べ本研究の位置付け・研究方針を明確にする。さらに第3章では、はじめにニューラルネットワークの概略を紹介し、本研究で採用したエルマン型ネットワークの詳細を解説する。計算機実験は二つの段階に分かれる。第4章では基礎実験として、人為的に作成された行きと帰りの経路が同じ振り子軌道を用いて実験を行なう。特に、ここでは、エルマン型ネットワークと一般に広く知られている層構造のネットワークとの能力を比較する事を目的とする。続く第5章では、実際の人間の動作から得られたデータを用いてエルマン型ネットワークの各種能力について実験を行ない、その能力を考察する。最後の第6章では、本研究で得られた知見をまとめ今後の課題について述べる。

2. 研究の背景

本章では、本研究の研究背景を示す．ここでは運動パターン(人間の動作を時系列で記録したもの)の認識に関する関連研究を紹介するとともに、本研究のアプローチを述べ本研究の位置付け・研究方針を明確にする．

2.1 Biological Motion

人間の運動に関する視覚認識について、Johanssonらにより1970年代に特徴的な心理実験が提案されている [26][27]．

初めに、人間の関節など動きの特徴を示すと思われる部位に点光源を装着し、暗闇の中で実験目的にかなう動作をさせ、点光源の軌跡を記録する．これは、暗闇の中で点光源の記録なので、奥行き情報が欠如した2次元の視覚情報である．この点光源の軌跡はMLD: Moving Light Displays と呼ばれる運動パターンである．このようにして作成された視覚情報を用いて、生体の運動 (Biological Motion) の知覚・認識についての心理実験が、現在までに各種行なわれている．

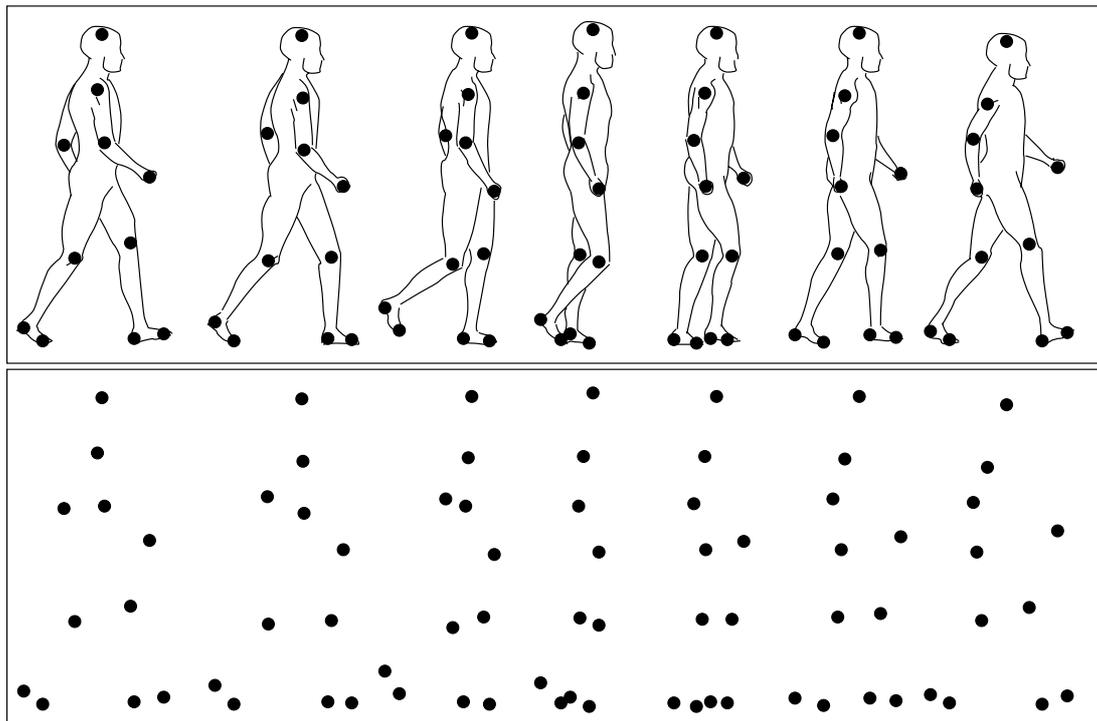


図1 MLD: Moving Light Displays

図1において，上段は点光源を装着した人間の左から右への歩行運動，下段は対応する点光源を示す(実際のMLDは点光源の軌跡である)。

このような，奥行き情報を排除した2次元の運動パターンからでも，生体か否か[26]，性差[36]，上下の違い[58]，動作(歩行[26]，ダンス[27])，感情表現[12][13]等の対象の特徴を人間は知覚することが可能と報告されている。さらに，Biological Motionの知覚能力は生後5ヶ月の幼児においても報告されており，人間に生得的に備わっていて学習により強化されていく考える事もできる[6]¹。

しかしながら，上に挙げたように数多くの心理実験が行なわれているが，ほとんどは観察される心理現象としての報告であり，計算論的なモデルを提案し，人間の能力として解明している研究例はあまり多くない[24][21]²。

2.2 View-based アプローチ

一般に，物体や運動の認識モデルを構成する場合，

- 3次元構造の記述を用いる手法
- 与えられた2次元情報に，より近い表現を直接的に使用する手法

の二つに大きく分ける事が出来る。

前者としては，対象に関する視点方向(見る方向)に依存しない3次元構造表現をあらかじめ内部にもち，与えられた2次元網膜像(場合によっては複数)から3次元構造を再構築し，それらの構造記述に基づいて物体認識を行なう方法が，Marrらによって静止した3次元物体認識について提案され，物体認識理論の大きな柱となっている[34]。この場合，視点方向に対して独立な3次元構造表現を内部にもつので，回転などでの見え方の変化に対して容易に対処出来るという利点がある。しかしこの手法では，対象に関して視点に不変な3次元構造の記述が必要であり，それが可能な場合のみ有効であるが，現在のところこのような表現方法を構築する普遍的な方法は存在しない。特に対象が人間の場合，人間は剛体ではなく手を動かしたりしたときにある種の拘束条件を伴って変形する。そのため，この認識手法に適した人間の動作の完全な3次元構造の記述は一層困難となる。また，この手法で

¹猫も Biological Motion を知覚するという報告もあるので [5]，生体全体に備わっている能力かもしれない。

²運動認識研究の実験データとしてMLDを使用しているのは沢山ある ([9] 等参照)

は与えられた 2 次元網膜像 (2 次元画像情報) から 3 次元構造を再構築する必要がある。しかしながら、2 次元画像情報から 3 次元特徴を精度良く抽出し、再構築する方法は容易ではなく、コンピュータビジョン研究の大きなテーマの一つでもある。

後者の、与えられた情報により近い表現を直接的に使用する方法は、見かけに基づく手法 (Appearance-Based アプローチ) とよばれ、近年研究が活発に行なわれている [61][40][41]。視点方向による見え方の違いに特に注目した認識手法を、View-based アプローチとよぶことがある [50][62][59][20]³。

近年、心理・生理実験により、人間やサルの物体認識は視点方向に依存すると報告されており [17][31]、視点依存表現を直接的に学習する View-based アプローチによる認識モデルが生理・心理学的にも支持を得ている。特に、Oram らによるサルの生理学実験において、サルの上側頭溝多種感覚野 (Superior Temporal Polysensory area: STPa) に、人間の MLD の見え方・動きの種類・動きの方向に選択的に反応する細胞が存在することが報告されている [44]。

脳の皮質での視覚関連領野に関する最近の生理学の研究において、視覚情報の経路として腹側経路 (ventral stream) と背側経路 (dorsal stream) の二つの存在が考えられている [18][67]。視覚情報は、網膜から外側膝状態を経て大脳皮質の後頭葉にある第一次視覚野 (V1) に伝えられ、この二つの経路に分れる。腹側経路は V1 を出発して側頭葉に至る経路であり「見ている対象物が何であるか」を知るための情報処理をしていると考えられ物体視経路ともよばれる。それに対し、背側経路は V1 を出発して頭頂葉に至る経路で「見ている対象の位置や動き」を知るための情報処理をになっていると考えられ空間視経路ともよばれる [63]。STPa はこのどちらの経路からの入力を受けており、形に関する情報と動きに関する情報を統合する領野として注目されている。このような領野で、MLD の見え方の違いに対して選択的に反応する細胞が確認された事は非常に大きな成果であり、MLD を用いた生体の視覚研究の妥当性が得られたも言える。

また、MLD を用いた歩行認識の心理実験 [64] により、人間の運動認識においての視点依存性も報告されている。

これらの成果をふまえて、本研究では運動パターンの視点依存表現を学習によって獲得し、認識を行なう新しい計算モデルを提案する。

³ただし、これら二つを明確に分離する事はできない。

2.3 運動パターンの認識：関連研究

ここで、関連する運動パターンの認識研究を簡単に紹介する⁴。前節の認識モデルでの3次元構造の記述を行なう手法は、対象に関しての知識を積極的に用いて、モデルを内部にもつのでモデルベース的でトップダウンな手法ともよばれ、運動パターンの認識でも数多くの研究がなされている [45][23][57][2]。

しかし運動パターンの場合、与えられる情報は連続する画像系列(動画像)であり、画像ごとに特徴点を抽出し、かつ、系列内でそれらの一貫性を保ち、対象の3次元構造を再構築することは、対象が移動し変形する可能性がある⁵、静止物体の場合より格段と難しくなる。そこで、運動パターンの認識においても普遍的な3次元構造表現を使用しない、見かけに基づく手法(Appearance-Based Model)が使用されてきている。これらの中には、2次元の画像データから特徴点を抽出しないで、動作パターンを認識させる方法が幾つか提案されている。Murase等は、静止物体の認識において高い評価を得ているパラメトリック固有空間法 [40][41]を動画像に適用し、歩行の認識を行なっている [42]。Polana等は画像データをメッシュ状に分割し(例えば16分割)、分割された区域の平均値の変化を動作パターンの特徴ベクトルとし、繰り返し動作の認識を実現している [51]。

さらに、時系列情報を扱うという点での共通性がある、音声認識の研究で幅広く採用されている隠れマルコフモデル(Hidden Markov Model)を適応した研究も提案されている [65][66]。同様なモデルとして、Bobickらは状態の遷移による認識システムを提案している [7]。

GoddardはMLDを用いて、コネクショニスト的モデルにより人間の動作パターン認識を行なっている [21]。このモデルでは、動作パターンから *scenarios* と呼ばれる高次表現を抽出し、それをを用いて認識・注意のメカニズムをモデル化している。

視点依存表現された画像に対しての予測モデルとしては、カルマンフィルターを使用した認識予測モデルが最近になって提案されている [56]。

⁴運動パターン認識一般についてのサーベイは [1][10][46] を参照。

⁵本研究で対象とする運動パターンは、人間(非剛体)の動作を時系列で記録したものである。

2.4 学習に基づく運動パターンの認識と予測

2.3節で紹介した既存の関連研究に対しての、本研究の特徴をここで述べる。本研究の大きな枠組は人間に学ぶというものである。我々人間は、柔軟性のあるさまざまな能力を学習によって獲得している。例えばスポーツなどは、体験し学習するを通して楽しむ事が出来るようになる。

計算機科学の分野において、積極的に学習を取り込んでいる領域として、ニューラルネットワーク (Nerual Networks) がある。学習によってネットワークを所望のシステム (振舞い) に仕上げるので、ニューラルネットワークと学習は不可分な関係といえる。

これまで、時系列データの認識・予測する方法は各種提案されてきているが [71] , ここでは系列データの取り扱いの為に Elman によって提案された、エルマン型ネットワーク [16] を使用した、運動パターンの視点依存表現を学習によって獲得し、認識・予測する計算モデルを提案する。

特に本研究では、以下の2つのアプローチを採用する：

- 人間の身体を再構築する，
モデルベース的なトップダウン手法ではない。
- 入力表現の要素間に明示的な拘束条件を設けない
(問題領域に関しての前提知識を必要としない)。

また、2.2で紹介した Oram らの研究より、MLD の刺激に対して選択的に反応する細胞が存在することが示された。これは生体の運動パターンの認識研究において、MLD を刺激として使用する事の妥当性が与えられたとも言えるので、本研究では、実験刺激 (実験データ) として、MLD を用いる。

さらに、2.3節で紹介した諸研究の主目的が、運動パターンの識別であるのに対し、本研究では識別のみならず動作の予測を目標にしている。

3. ニューラルネットワークを用いた時系列予測

本章では、はじめにニューラルネットワークの概略を紹介し、最後に本研究で採用したエルマン型ネットワークの詳細を解説する。

3.1 ニューラルネットワーク

我々人間の脳には 100 億個以上のニューロン (神経細胞) があり、それらが相互作用する事により複雑な情報処理を行なっている。このような人間の仕組みを模倣し人間の持つ柔軟な情報処理を実現しようとする試みはコンピュータの実現と同じ位に古く 1940 年代から行なわれてきている。

1943 年に McCulloch と Pitts はニューロンの単純なモデル (McCulloch-Pitts の形式ニューロン) を提案し、これによりいかなる有限論理表現が実現可能ということを示した [32]。

このモデルでは、入力 $x_1, \dots, x_n \in \{1, 0\}$ に対してニューロンの出力 z を

$$z = 1 \left[\sum_{i=1}^n a_i x_i - \theta \right] \quad (1)$$

とする。ただし、

$$1(u) = \begin{cases} 1, & u > 0 \text{ のとき} \\ 0, & u \leq 0 \text{ のとき} \end{cases} \quad (2)$$

ここで a_i は各入力 i の重み係数、 θ はニューロンの閾値を示す実数である (図 2 に模式的に示す)。

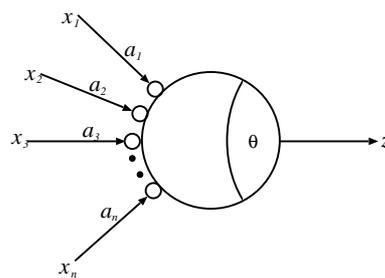


図 2 McCulloch-Pitts の形式ニューロン

このモデルは計算可能性を示したものにすぎないが、神経細胞が興奮したとき、それに刺激を与えた入力結合部 (シナプス) の結合が強化され、さらに刺激が伝え

られるやすくなる，という仮説が 1949 年に Hebb により提唱された [22]．この仮説 (Hebb のシナプス強化法則) は，複数のニューロンが結合されたネットワーク構造における，学習モデルの初期のアイデアとなり，現在でも多くの学習モデルに影響を与えている．

McCulloch-Pitts の形式ニューロンは離散時間・離散情報モデルであるが，連続情報モデルに拡張する為にニューロンを以下のように一般化する．あるニューロン j において，他のニューロン i からの出力 $y_i (i = 1, \dots, n)$ に対応する重み係数 w_{ji} をかけた値の総和を u_j とする

$$u_j = \sum_{i=1}^n w_{ji} y_i \quad (3)$$

ニューロン出力 y_j は，入力の総和 u_j に単調増加関数 f を施したもので表わす．

$$y_j = f(u_j - \theta_j) \quad (4)$$

ここで，関数 f としてステップ関数 (式 2) を用いると，McCulloch-Pitts のモデルとなる． f として，図 3 に示すシグモイド関数を用いた場合，この関数は微分可能なので解析的に問題を解く事が可能となる．

$$f(u) = \frac{1}{1 + e^{-u}} \quad (5)$$

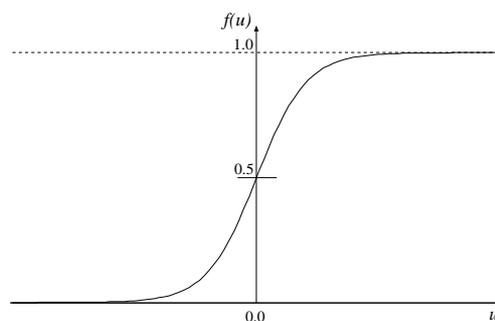


図 3 シグモイド関数

以上に示したニューロンが複数個結合したネットワークの例として，3 層構造のネットワークを図 4 に示す．工学的見地から，以下はニューロンの代わりに素子という言葉を用いる．パターン (多次元ベクトル) が入力される素子を入力層，ネットワークの出力を出力層，入力層と出力層とに挟まれた層を中間層とよぶ

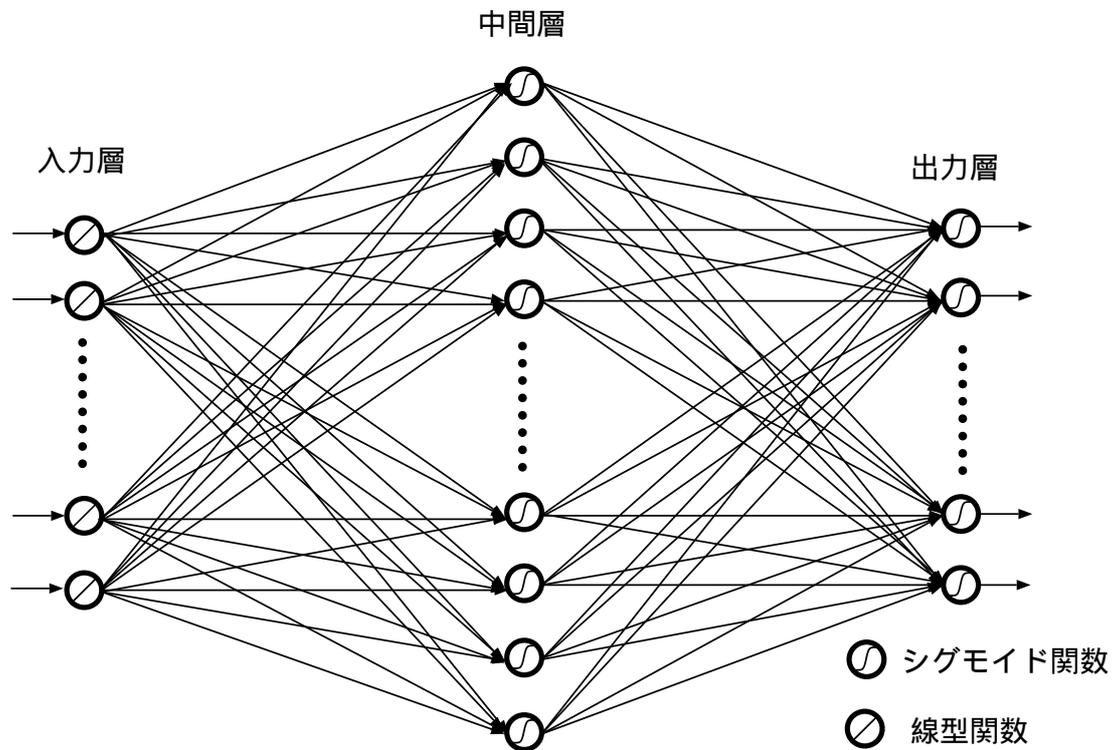


図 4 3 層ネットワーク

一般的な 3 層ネットワークにはフィードバック結合は存在せず，情報は入力層から出力層方向への一方向（前向き結合）しか伝わらない．さらに通常，入力層各素子は中間層素子全てに結合し，中間層各素子は全ての出力層素子に結合している．入力層素子数と出力層素子数が同じである必要はなく，これは対象とする問題に依存する．

層構造をとるネットワークは 3 層とは限らず，中間層を複数層持つ多層ネットワークや中間層のないネットワークも可能である．入出力層の素子数は対象によって拘束されるが，中間層の層の数や各層での素子数は自由に決定する事が可能であるが，それらの決定方法は現在においても理論的に確定してなく，ニューラルネットワーク研究の大きな課題でもある⁶．その為，本研究では中間層の素子数は計算機実験により決定する．

⁶例えば，Net News : comp.ai.neural-nets の FAQ(Frequently Asked Questions) に，“How many hidden layers should I use?”，“How many hidden units should I use?”というセクションがある (<ftp://ftp.sas.com/pub/neural/FAQ.html> 参照)．

3.1.1 学習モデル：誤差逆伝播法

入力ベクトルは中間層を前向きに伝播されながら出力層にベクトルとして出力させる（どちらもスカラの場合もあり得るが以下はベクトルと仮定する）。ある入力ベクトルに対して望ましい出力ベクトルが与えられている場合は教師あり学習であり、この望ましい出力ベクトルは教師信号とされる。本研究ではこの入出力ベクトルのペアをパターンとよぶ⁷。ある入力に対して、ネットワークから望ましい出力が得られた場合は学習しないが、それ以外の場合は出力された結果と教師信号と差を減らすように、ネットワークの結合の重み係数を調整する。これによりネットワークの学習が実現される。

まずはじめに、Rumelhart 等によって提案され広く普及した誤差逆伝播法 (Error Back Propagation) : 略して BP 法を紹介する [53][54]⁸。

はじめに、つぎのような二乗誤差関数を考える。

$$E_p = \frac{1}{2} \sum_{j \in \text{output}} (t_{pj} - y_{pj})^2 \quad (6)$$

t_{pj} は出力層素子 j のパターン p での教師信号、 y_{pj} はそのときの実際出力である。

出力 y_{pj} はその時の重み係数 w_{ji} で決定されるので、各重み係数の軸で張られる空間内で、 E_p を高さと考えればこの二乗誤差は誤差曲面となる。任意の重みの状態からこの誤差曲面の極小値 (最小値とは限らない) に到達させるには、最急降下法により、重み係数 w_{ji} を

$$\Delta_p w_{ji} = -\eta \frac{\partial E_p}{\partial w_{ji}} \quad (7)$$

のように変化させれば良い (η : 学習係数は 0 に近い正の定数)。誤差信号 δ_{pj} を

$$\delta_{pj} = -\frac{\partial E_p}{\partial u_{pj}} \quad (8)$$

⁷望ましい出力が与えられない場合は、教師なし学習となる。

⁸基本的な考え方は、1974年の Werbos の博士論文で既に示されていた事が、1980年中頃に明らかになった [72]。

とすれば (u_{pj} は式 (3) 参照), 合成関数の微分公式より

$$\delta_{pj} = -\frac{\partial E_p}{\partial u_{pj}} = -\frac{\partial E_p}{\partial y_{pj}} \frac{\partial y_{pj}}{\partial u_{pj}} \quad (9)$$

となり, δ_{pj} は誤差の出力についての微分と, 出力の入力総和についての微分の積に分解出来る. 式 (4), (5) より 2 番目の微分は

$$\frac{\partial y_{pj}}{\partial u_{pj}} = f'(u_{pj})$$

となる. しかし, 一つ目の微分には場合分けがある.

出力層の素子の場合, 式 (6) より

$$\frac{\partial E_p}{\partial y_{pj}} = -(t_{pj} - y_{pj})$$

だから, 式 (9) に代入すると出力層に関しての δ_{pj}

$$\delta_{pj} = (t_{pj} - y_{pj}) f'(u_{pj}) \quad (10)$$

が得られる.

それ以外の場合, ふたたび合成関数の微分を用いて,

$$\begin{aligned} \sum_k \frac{\partial E_p}{\partial u_{pk}} \frac{\partial u_{pk}}{\partial y_{pj}} &= \sum_k \frac{\partial E_p}{\partial u_{pk}} \frac{\partial}{\partial y_{pj}} \sum_i w_{ki} y_{pi} \\ &= \sum_k \frac{\partial E_p}{\partial u_{pk}} w_{kj} \\ &= -\sum_k \delta_{pk} w_{kj} \end{aligned}$$

となり, この場合の δ_{pj} は

$$\delta_{pj} = f'(u_{pj}) \sum_k \delta_{pk} w_{kj} \quad (11)$$

と計算できる. ここで k はこの素子の出力を受けている全素子を示す.

全てのパターンのついで二乗誤差の総和

$$E_{net} = \sum_{p \in \text{patterns}} E_p \quad (12)$$

に対して、

$$\frac{\partial E_{net}}{\partial w_{ji}} = \sum_p \frac{\partial E_p}{\partial w_{ji}}$$

であるから、全パターンでの一度ずつの重み係数の変化の総和は、 $\partial E_{net}/\partial w_{ji}$ に比例する⁹、よって、以下は E_{net} を使用する。

実際の学習では、初めにネットワークに与えられた入力ベクトルが前向きに伝播、出力結果が得られる。得られた結果と教師信号との差から、式 (10) を用いて誤差信号 δ_{pj} を計算し、式 (7) により、出力層の素子への重み係数の変更量が決定される。中間層の素子では、その素子が出力を送っている一つ後ろの層の素子の式 (11) より、誤差信号 δ_{pj} を求める事ができる。これを再帰的に繰り返す事により、各素子の結合の重み係数を更新することができる。この演算はネットワークを後ろ向きに伝播するので、提案された学習方法は誤差逆伝播法とよばれる。

誤差逆伝播法では、全ての重み係数を同じ値にして学習を開始した場合、全ての重み係数が同じように変化してしまう。このとき、中間層の全ての素子が同じような振舞いをし、非対称なベクトルをもつパターンの学習は出来ない。これを避けるため、それぞれの重み係数の初期値は小さな乱数で与える。

3.1.2 汎化能力

学習過程で用いられたパターンの関係のみの獲得だけではなく、与えられたパターンの背後に存在するだろう規則性を探り出し、未知の入力に対しても正しく出力することが出来るようにならなければ、人間の柔軟な情報処理能力を模倣した学習システムとは言い難い。このような能力を汎化能力とよぶ。

図5に示す連続な関数 $y = f(x)(X_1 \leq x \leq X_6)$ の学習を例とする。本研究で使用しているニューラルネットワークは離散時間・連続情報モデルなので x 軸方向を時間軸とすると、学習パターンは有限離散個 $((X_1, f(X_1)), \dots, (X_6, f(X_6)))$ となる。

⁹各パターンの度に重み係数を更新する場合、はじめのほうのパターンでの変更が後のほうのパターンの重み係数の変更に影響するので、厳密には一致しないが、差は僅差であり η を十分小さくとれば無視できる。

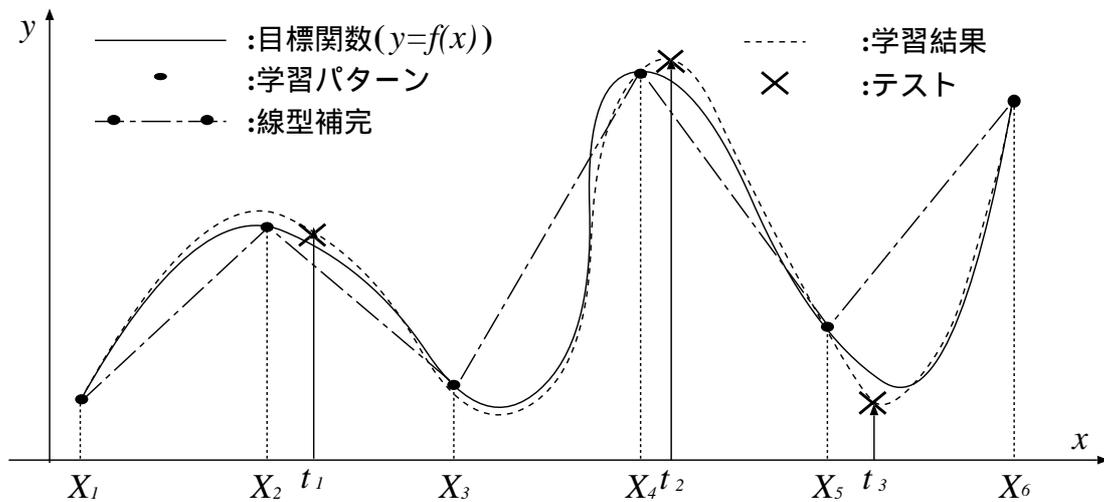


図5 ニューラルネットワークの汎化性

与えられたパターン $((X_1, f(X_1)), \dots, (X_6, f(X_6)))$ に対して、ニューラルネットワークが点線で描かれるような曲線を学習できたとすれば、学習パターン以外において、線型補完より目標関数に近似した結果を出力する事になる。ニューラルネットワークの学習においては、未知な入力にも確からしさをもって対処できるように、つまり汎化能力を高めるようにネットワークの重み係数を決めていく必要がある。

そこで、最小化する二乗誤差関数 (12) にペナルティ関数 C_{decay} を追加し、新たなコスト関数を導入する [49]

$$C_{decay}(W) = \frac{1}{2} \lambda \sum_i w_i^2$$

$$C_{total} = E_{net} + C_{decay}$$

この時、式 (7) より

$$\Delta w_{ji}(t) = -\eta \frac{\partial E_{net}}{\partial w_{ji}}(t) - \lambda w_{ji}(t)$$

$$w_{ji}(t+1) = (1 - \lambda)w_{ji}(t) - \eta \frac{\partial E_{net}}{\partial w_{ji}}(t)$$

となり， λ は重み係数の過度な変更 (学習したパターンに対する過学習) を抑制する係数であることが解る．これより λ は重み変更抑制係数 (Weight Decay) とよばれる¹⁰．

3.1.3 Quickprop 法：誤差逆伝播法の拡張

誤差逆伝播法は，いかなる結合係数の初期値からでも誤差が極小値になる事 (最小値とは限らない) が保証されているが，一般に誤差曲面は極小値の近くでは緩やかな谷底をもつため，学習は非常に遅くなる．式 (7) の η (学習係数) の値を大きくすると，1 回の重み係数の変化量が大きくなり学習速度は増すが，大きくし過ぎると振動してしまう場合がある．

そこで，学習係数をできるだけ大きくしながら，かつ，重み係数の振動を抑えるために，時刻 (t) での重み係数の変化量 (誤差曲面での傾き：式 (7)) に，次のような慣性項 α を加える [54]．

$$\Delta w_{ji}(t) = -\eta \frac{\partial E_{net}}{\partial w_{ji}}(t) + \alpha \Delta w_{ji}(t) \quad (13)$$

慣性項により 1 時刻前の重み変化の影響が加味され，それが一種の慣性を生じさせ，誤差曲面の細かい凸凹を無視する効果があられる．本研究では，この考え方をさらに拡張し，誤差逆伝播法の高速度手法の一つとして Fahlman によって提案された Quickprop 法を採用する [19]．

この手法では式 (13) を次のように拡張する．

$$\begin{aligned} \Delta w_{ji}(t) &= -\eta^{(t)} S_{ji}(t) + \alpha_{ij}^{(t)} \Delta w_{ji}(t-1) \\ S_{ji}(t) &:= \frac{\partial E_{net}}{\partial w_{ji}}(t) + \lambda w_{ji}(t) \end{aligned} \quad (14)$$

λ は式 (13) で導入した重み変更抑制係数であり，学習係数 $\eta^{(t)}$ は

$$\eta^{(t)} = \begin{cases} \eta, & \Delta w_{ji}(t-1) = 0.0 \text{ または } S_{ji}(t) \Delta w_{ji}(t-1) > 0.0 \\ 0, & \text{それ以外} \end{cases} \quad (15)$$

¹⁰Mackey によりベイズ理論に基づく理論的解釈がなされている [33]

慣性項 $\alpha_{ij}^{(t)}$ は

$$\alpha_{ij}^{(t)} = \begin{cases} \mu, & \tilde{\alpha}_{ij}^{(t)} > \mu \text{ または } S_{ji}(t)\tilde{\alpha}_{ij}^{(t)}\Delta w_{ji}(t-1) < 0.0 \\ \tilde{\alpha}_{ij}^{(t)} = \frac{S_{ji}(t)}{S_{ji}(t-1)-S_{ji}(t)}, & \text{それ以外} \end{cases} \quad (16)$$

とする．さらに，重み係数の急激な変更による振動を避けるために，最大変化量 μ を追加導入する．

以上より，本研究でのニューラルネットワークの学習には

η : 学習係数 (式 15)

μ : 最大変化量 (式 16)

λ : 重み変更抑制係数 (式 14)

の 3 係数が必要となる (以下，学習条件とよぶ) ．

しかし，最大変化量は誤差曲面での振舞いを，重み変更抑制係数は汎化能力を調整する係数であり，どちらもある種の補助的な項とも考えられる．そこで，ネットワークの学習という観点からみれば，学習係数の設定が非常に大切といえる．

3.2 ニューラルネットワークと時系列処理

前節迄で解説した誤差逆伝播法およびその拡張である Quickprop 法は，各パターンでの入出力関係を学習させるのであり，各パターン間の順序関係はその対象となっていない．また図 4 に示したような層構造のネットワークには，構造自体に時系列性は陽に組み込まれてはいない．そのため，時系列パターン (順序付きパターン) を効率良く学習させるためにさまざまな工夫が提案されてきている．

信号処理，特に音声認識の分野では時系列データを扱ってきいるので，数多くの研究がなされている [38][25] ．そのなかでも Waibel らによって提案された時間遅れニューラルネットワーク： *Time Delay Neural Network* (TDNN) は，音韻の識別特徴の検出を目指しながら時間軸方向での不変性を目指したものである [68] ．学習方法としては既存の誤差逆伝播法を用いているが音韻識別の先験的な知識を層構造ネットワーク拡張に取り入れたことにより，高い認識率を達成している¹¹ ．

¹¹音韻認識テストにおける隠れマルコフモデル (Hidden Markov Model: HMM) の認識率平均：93.7%(1989 年当時) に対して，同じデータでの TDNN の平均は 98.5% であった ．

一方、今迄述べてきた層状ネットワーク（および、その拡張である TDNN）以外に、各素子の自由な結合を許したネットワークとしてリカレントネットワーク：*Recurrent Network* がある。このネットワークには後ろ向きの結合も含まれるので、時系列パターンを陽に取り扱うことも可能である。しかし、自由な結合を許すためその構造は複雑となり、学習方法にも工夫が必要とされる。リカレントネットワークには、時系列パターンを明示的に対象とした学習方法が各種提案されている [70][47]。代表的なものとして、誤差逆伝播法を時間方向に拡張した *Back Propagation Through Time*(BPPT) ¹² [54][73]、ネットワークを動作させながら同時に結合係数を変更する *Real-Time Recurrent Networks*[74]、連続時間への拡張 [48][14][15] 等も提案されている。

3.3 エルマン型ネットワーク

全ての素子間での結合を許すリカレントネットワークにたいして、階層構造のネットワークに後ろ向き方向の結合：フィードバック結合を部分的に追加したネットワークを部分結合リカレントネットワーク (*Partially Recurrent Network*) とよぶ ¹³。

特に時系列パターンの取り扱いを目的として提案されたものとして、Elman によるエルマン型ネットワーク [16]、Jordan によるジョーダン型ネットワーク [29] がある (図 6)。

エルマン型ネットワークは、層構造のネットワークにコンテキスト層 (Context Layer) を追加し、中間層の出力がコンテキスト層への入力となり、その出力がふたたび中間層の入力になるというフィードバック結合をもつ。一方、ジョーダン型ネットワークでは、層構造のネットワークに状態層 (State Layer) を追加し、ネットワークの出力が状態層への入力となり、その出力が中間層へと繋がっている。特に、ジョーダン型の状態層の各素子は、自分自身へのフィードバック結合と同層の素子間の結合が存在する。ジョーダン型では入力層をプラン層 (Plan Layer) とよび、状態層とプラン層を合わせてネットワークへの入力層としている。

¹²BPPT 法はリカレントネットワークに限らず、層構造ネットワークでも学習が可能である。また、基本的な考え方は誤差逆伝播法と同様に 1974 年の Werbos の博士論文で示されていた [72]。

¹³自由な結合を許すリカレントネットワークを全結合リカレントネットワーク (*Fully Recurrent Network*) とよび明示的に区別する場合もある。

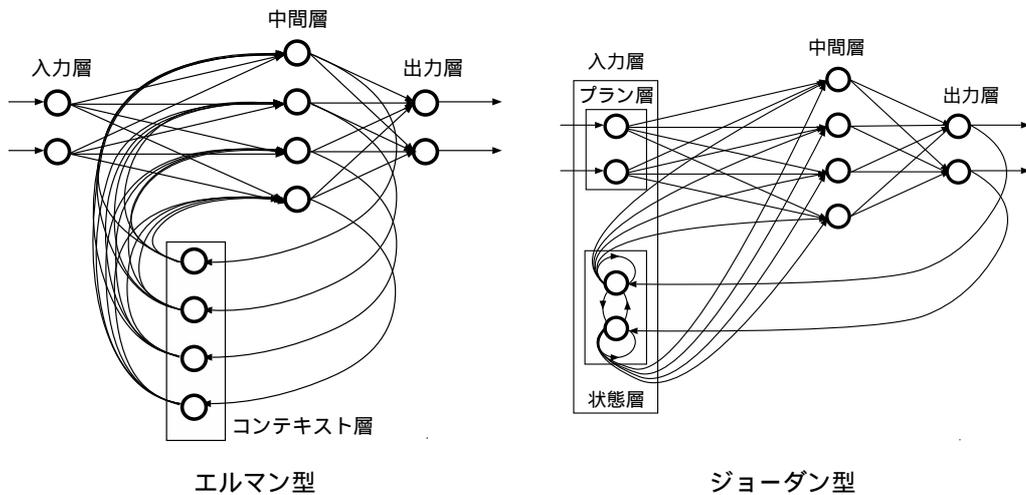


図 6 エルマン型・ジョーダン型ネットワーク

エルマン型の場合，中間層の素子とコンテキスト層の素子は一対一に対応しているため，コンテキスト層の素子数は中間層のそれと等しく，ジョーダン型の場合は出力層の素子数と状態層の素子数が等しくなる．そのため，ジョーダン型の場合は，問題対象の出力次元によって状態層の素子数が拘束されてしまう．一方エルマン型の場合，中間層の素子数が問題対象の出力次元によって拘束されることはないため，ジョーダン型に比べてコンテキスト層の素子数は自由に設定できる．

本研究では，画像情報を直接学習パターンに使用する計算機実験も行なう．この場合入出力の次元は非常に大きくなる（例えば 1024 次元）．したがって，ジョーダン型を使用した場合，ネットワーク構造が巨大となり，学習時間が非常に長くなる可能性がある．そのため，本研究ではエルマン型ネットワークを採用する．

3.3.1 エルマン型ネットワークでの時系列処理

エルマン型ネットワークでは，中間層からコンテキスト層への結合部分の重み係数は 1 に固定してある．時刻 (t) において，コンテキスト層には時刻 $(t-1)$ での中間層の出力が入力され，その出力は時刻 (t) での入力ベクトルと共に中間層への入力となる．時刻 $(t-1)$ でのコンテキスト層には，時刻 $(t-2)$ 中間層の出力が入力され，以下同様に続いていく．つまり，ある時刻でのネットワークの状態は，それ以前の過去の時刻でのネットワークの状態を反映している．このように，エルマン型ネットワークの場合は，フィードバック結合とコンテキスト層の効果により，時

系列データを効果的に処理することが可能となっている¹⁴。

ここで、ジョーダン型ネットワークで導入された状態層での自分自身へのフィードバック結合を考える。状態層の素子が線型関数、フィードバック結合の重み係数が α で固定されているとすると、時刻 (t) での状態層の素子出力： $S(t)$ は、時刻 $(t-1)$ での中間層の出力を $H(t-1)$ とすれば

$$\begin{aligned} S(t) &= \alpha S(t-1) + H(t-1) \\ &= \sum_{r=1}^t \alpha^{r-1} H(t-r) + \alpha^t S(0) \end{aligned}$$

となる。ただし $H(0)$ は中間層素子、 $S(0)$ は状態層素子の初期値とする。 $S(t)$ の無限大への発散を防ぐために、 $(0 \leq \alpha < 1)$ とすれば、 $S(t)$ には過去の情報が α 倍で足し込まれていると解釈できる。ここでは素子が線型関数と仮定したが、シグモイド関数のような非線型関数でも同様に考えることができる。本研究では、コンテキスト層の素子は線型関数とする。

Elmanによって提案されたオリジナルのエルマン型ネットワークでは、コンテキスト層の素子には自分自身へのフィードバック結合は存在しないが[16]、本研究ではこの結合を追加することにし、以降はこの拡張されたエルマン型ネットワークをエルマン型ネットワークとし¹⁵、 α を自己回帰係数とよぶことにする。

3.3.2 エルマン型ネットワークの学習

エルマン型ネットワークのようにフィードバック結合がある場合、3.1.3節で紹介したQuickprop法はそのままでは適用できない。そこで、図7のように考える。

先述した通り、エルマン型ネットワークでは、中間層からコンテキスト層への結合係数は1、コンテキスト層の自己回帰係数は α で固定されている。そのため、時刻 (t) でのネットワークのデータの流れは、このように層構造のネットワークとして表現できる。よって、時刻 (t) での重み係数をQuickprop法で学習させることが可能となる。

¹⁴このような構造を時系列データの短期記憶機構と捉える考え方もある[39]。

¹⁵拡張されたエルマン型ネットワークをエルマン型ネットワークと紹介しているものもある。

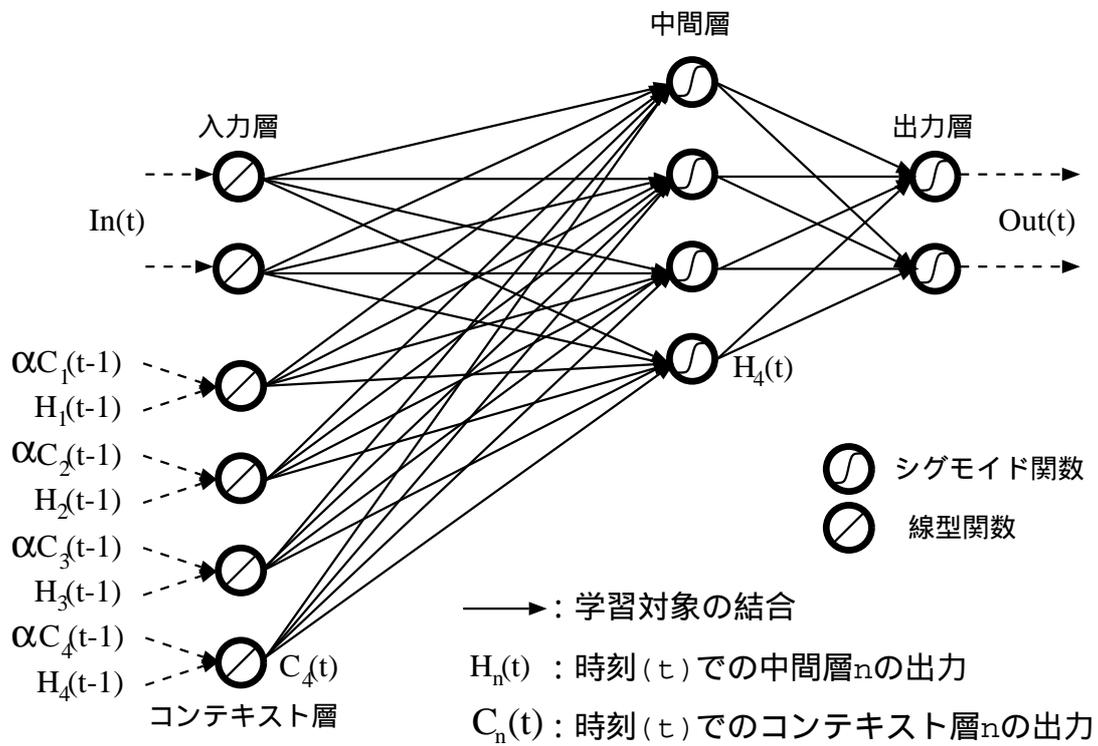


図 7 エルマン型ネットワークの学習

次章では、人為的に作成した時系列データに対して、エルマン型ネットワークを用いた計算機実験について議論を進める。

4. 計算機実験 1 : 単振り子・二重振り子

本章では基礎実験として，人為的に作成された，行きと帰りの経路が同じ振り子軌道を用いて実験を行なう．特にここでは，エルマン型ネットワークと一般に広く知られている層状ネットワークとの比較を目的とする．

4.1 実験の目的

Mother らの，MLD を用いた人間の歩行についての心理実験より，認識過程において手先・足先端点の動きが，非常に重要であると報告されている [37]．ここで使用した実験刺激は重心を固定した歩行データであり，手先の動きは単振り子軌道，手先と肘の動きは二重振り子軌道に似ている．そこで，ここでは単振り子・二重振り子の軌跡を例にして，計算機実験を行なう．

特に，振り子の軌道は半周期毎に重なっていて，次の座標値を決定するには過去の座標値系列 (履歴情報) が必要とされる．よって，振り子軌道予測は，単なる位置予測ではなく，時系列予測の問題と捉える事も出来る．計算機実験に用いた 2 つの振り子の軌道を図 8 に示す (軌道の算出は付録 A を参照) ．

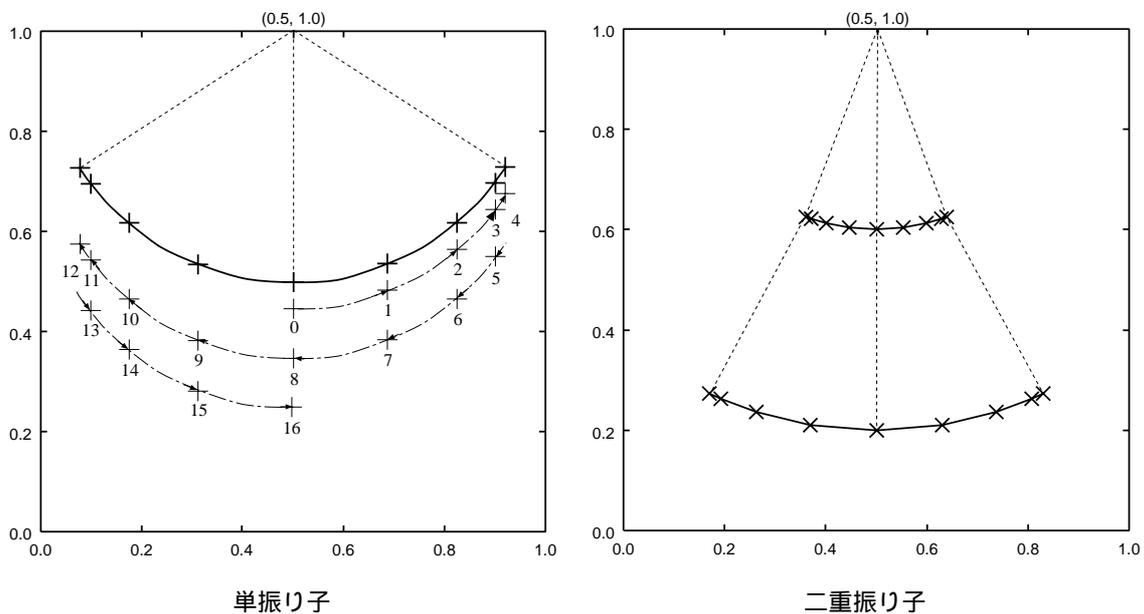


図 8 単振り子・二重振り子

どちらの軌跡も最下点を開始位置とし左向きに振り上げる方向に進み、左端で折り返し、同じ軌道を通り最下点を經由して右向きに振り上り、最後に最下点に戻るものとし、これを1周期とする。この1周期を時間方向に17等分しそれぞれの時点での (x,y) 座標値をデータとする(データ: 0 ~ 16)。よって学習のペアは、パターン1: (入力: データ0・教師信号: データ1), パターン2: (入力: データ1・教師信号: データ2), ..., パターン16: (入力: データ15・教師信号: データ16)まで全部で16個となる。学習時は、パターン1, パターン2, ..., パターン16までを連続して学習させ、この一連のパターンを系列とよぶ。ここでの振り子系列は16組のパターンから成り立つ。

4.2 計算機実験

以下、本研究の計算機実験で用いる2通りの予測方法について説明、中間層の素子数の決定、エルマン型ネットワークと3層ネットワークとの比較、エルマン型ネットワークの能力の検討を進めていく。

4.2.1 予測方法

対象の動きを予測する場合、刻々と与えられる情報に対して次の時刻の情報を予測する状況と、ある時刻までの入力情報だけを用いて、それ以降の情報を全て予測する状況が考えられる。

本研究では、入力は常に与えられ、次のデータのみを予測する場合(これは、学習パターンの再現でもある)を、短期予測(図9左図)とよぶ。また、ある時刻まではデータが与えられ、それ以降はネットワーク自身の予測結果を次の入力として繰り返し用いて、最後まで予測していく方法を、長期予測(図9右図)とよぶ。

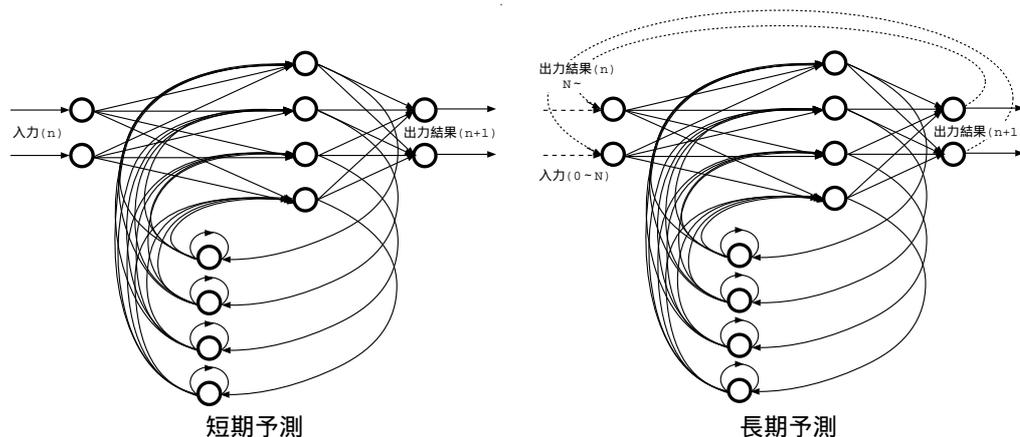


図9 短期予測・長期予測

4.2.2 単振り子・中間層の素子数の決定

単振り子の場合，軌跡は一つであるから，入力層・出力層共に一組の (x,y) 座標値を表現する 2 個となる．中間層の素子数決定方法は，現段階では確立してなく，経験的・実験的にならざる負えない．まず，エルマン型ネットワークの中間層の素子数による，出力誤差 (式 12) の違いを比較検討し，必要な素子数を決定する．

計算機実験条件は，学習対象として 16 組のパターン (入力・教師信号の組) からなる単振り子系列，予測方法として短期予測とする．図 10 に，中間層の素子の数を 2,4,8 個と変えた場合での，学習回数 (横軸) と予測誤差 (縦軸) の関係を示す．予測誤差は，式 (12) で定義した誤差： E_{net} とする (1 学習回数は 1 系列の学習を意味する)．以下特にことわらない限り，学習方法は 3.3.2 節で示した Quickprop (学習条件： $\eta : 0.03$ ， $\mu : 2.0$ ， $\lambda : 0.0005$) を使用し，中間層の初期値は 0，コンテキスト層の初期値は 0.5，コンテキスト層の自己回帰重み係数は 0.5 とする．

図 10 より，中間層素子が 2 個の場合では，学習回数を増やしても E_{net} はある値より下らない傾向が解る．4 個と 8 個の場合，どちらも学習回数 2000 回迄は下る傾向にあり，2000 回での予測誤差は，4 個：0.004011，8 個：0.000654 である．その時の短期予測の結果を図 11 に示す．

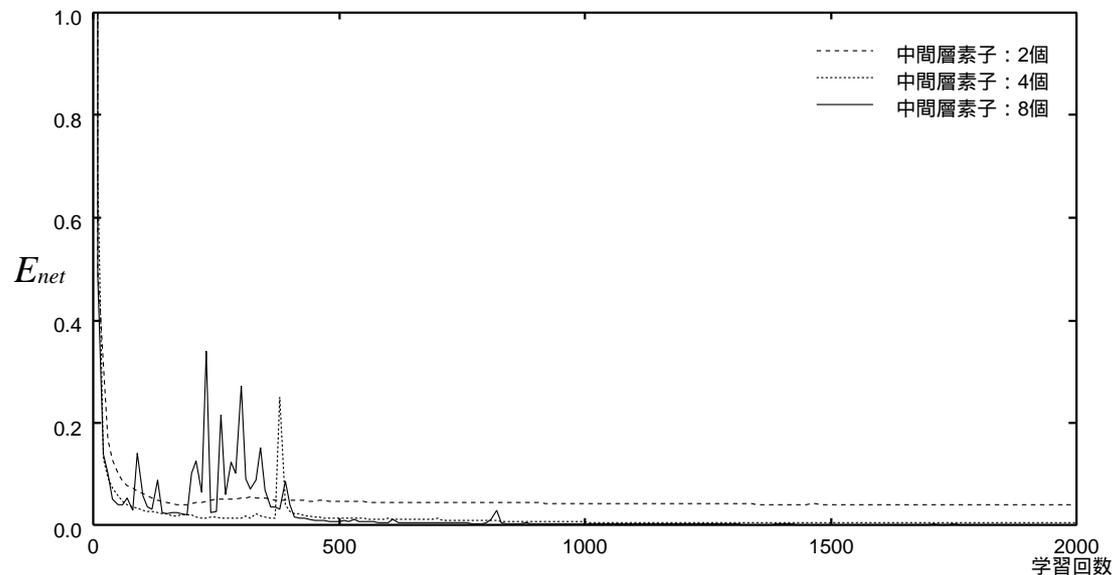


図 10 エルマン型ネットワーク：中間層の素子数と予測誤差

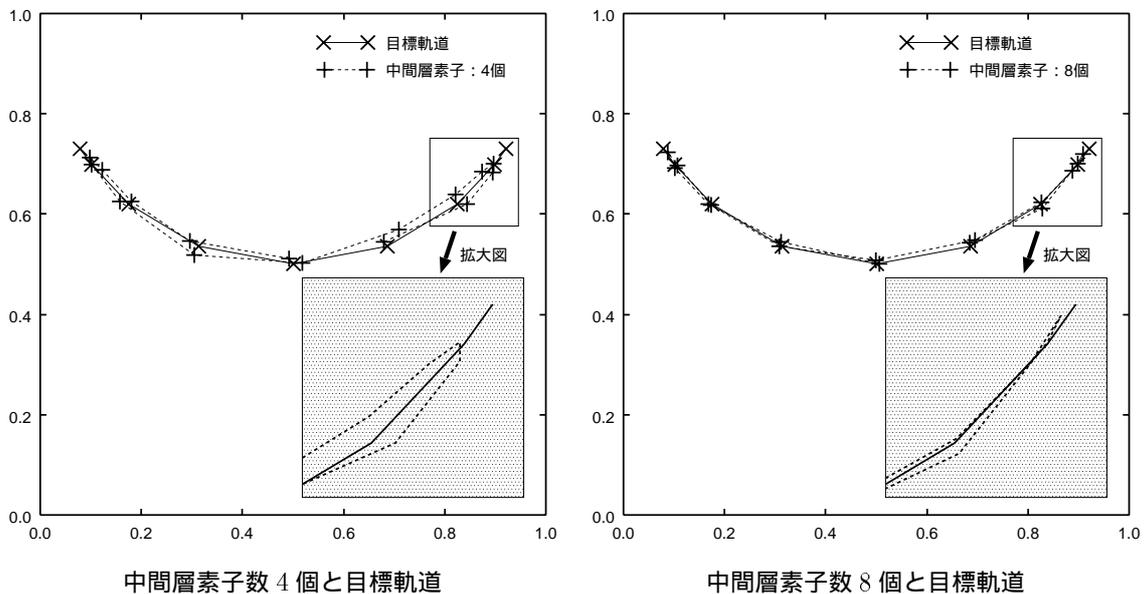


図 11 単振り子 (中間層素子数：4,8 個, 学習：2000 回)

拡大図から解るように，中間層素子が 4 個では折り返し近傍の予測が上手くいっていない．この場合，学習回数を 8000 回迄増やしても， E_{net} は 2000 回時の値より下ならず，予測結果も向上しない．さらに，中間層素子数を 12 個にした場合でも， E_{net} は 8 個の場合とほとんど変わらず，予測結果も同等である．

以上の結果をふまえて，これ以降の計算機実験では中間層素子 8 個のエルマン型ネットワークを用いる．

4.2.3 単振り子・3 層ネットワークとの比較

ここでは，単振り子の軌道予測について，エルマン型ネットワークと 3 層ネットワークとの比較を行なう．入力層・出力層素子各 2 個，中間層素子 8 個のエルマン型ネットワークは，コンテキスト層を含めると，全部で 20 個の素子及び 112 本の結合 (112 個の重み係数) がある． (ただし，中間層からコンテキスト層への結合の重み係数，コンテキスト層の自己回帰係数は固定) ．

ここでは，結合数を揃えた 3 層ネットワークと比較する．この場合入出力層の素子数は同じなので，中間層の素子は 28 個となる．図 12 に比較対象の 2 種類のネットワークを示す．

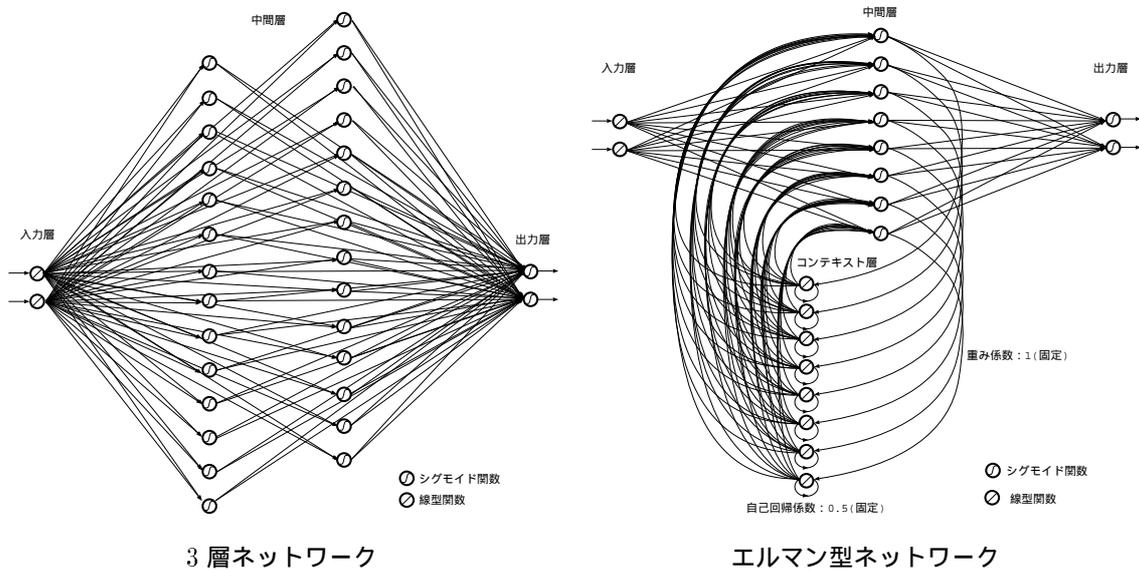


図 12 3層ネットワーク・エルマン型ネットワーク

3層ネットワークに Quikprop 法を用いて、学習係数のみを 0.3, 0.03, 0.003 と変えた場合の、学習回数に対する E_{net} 変化を図 13 に示す。エルマン型ネットワークを図中に破線で示す。

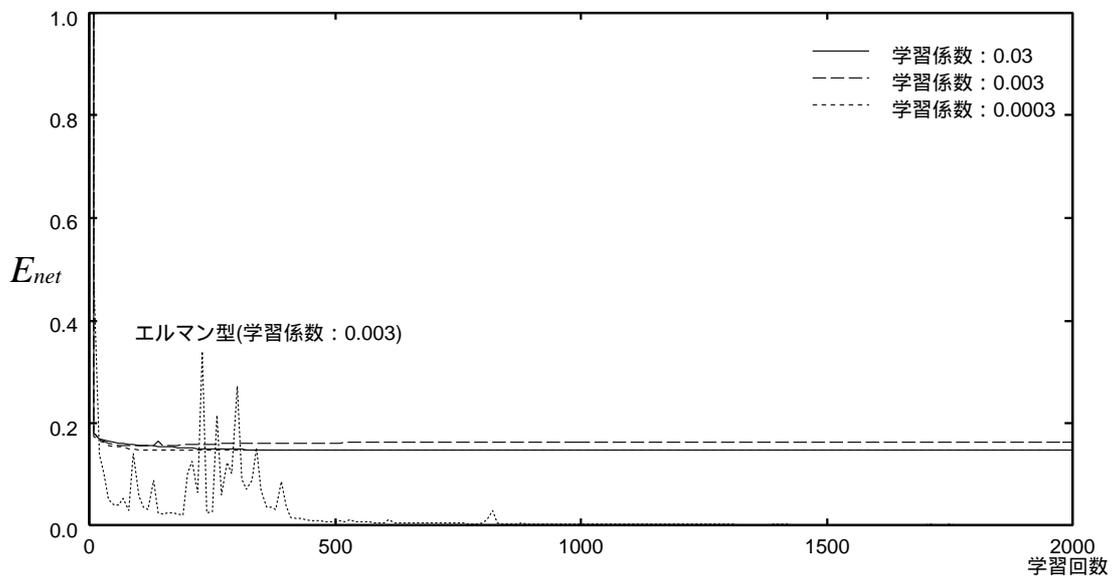


図 13 3層ネットワークの学習曲線

エルマン型ネットワークの場合、学習回数が500回位迄は振動が見られるが、それ以降は低い値で落ち着いている。それに対し、3層ネットワークでは、学習係数を変えても E_{net} にはほとんど差が見られず、500回以上学習を継続しても E_{net} は低下しない(2000回終了時のエルマン型ネットワークの E_{net} は0.000654、3層ネットワークでは0.146)。3層ネットワークでは、他の学習条件(最大変化量・重み変化係数)を変化させた場合でも改善されない。

学習回数2000回(学習条件: 0.03, 2.0, 0.0005)での、短期予測のテスト結果を図14に示す。図中、斜体数字(0, 1, 2, 3)は目標軌道のデータ番号、ゴシック数字(1, 2, 3)は予測結果を示す

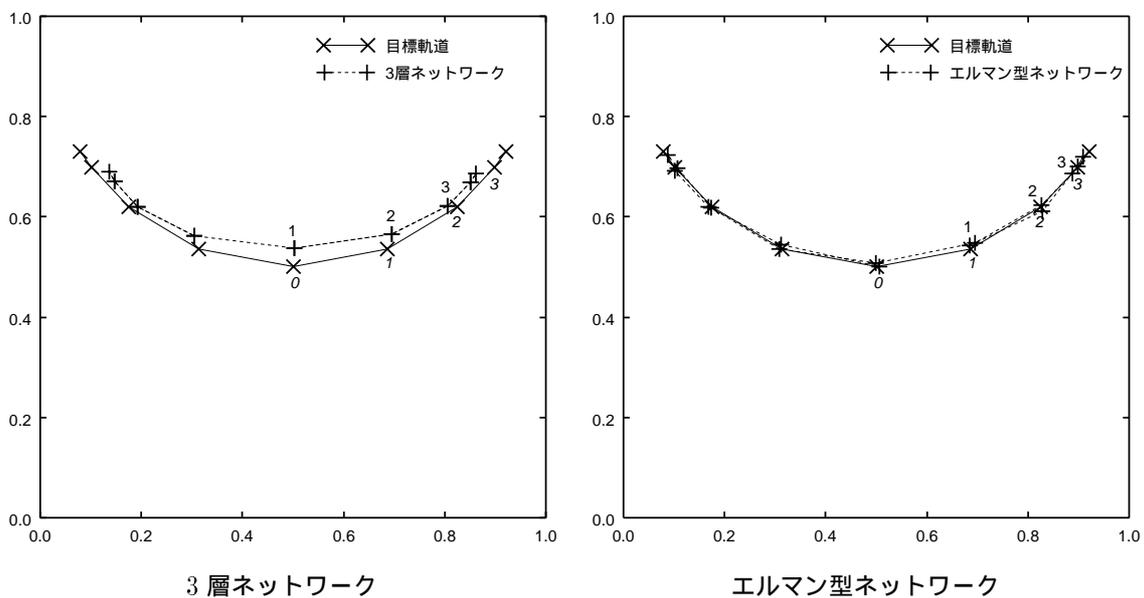


図14 単振り子テスト結果(学習: 2000回, 短期予測)

3層ネットワークの場合(左図), 半周期毎に軌道は完全に重なる(結果1と結果9は完全に一致する)。しかし、データ0入力に対しての予測結果としてデータ1を出力する様に学習させた場合、データ0近くの値が結果1として出力されてしまう。以下のデータにおいても同様な傾向が続き、軌道の時系列予測の学習が出来ていない。学習回数を10倍にしたり、前述したように各種学習条件を変えても、結果は改善されない。

3層ネットワーク, エルマン型ネットワーク(学習回数: 2000回)の長期予測の

結果を図 15 に示す．ここでは，データ 0 からデータ 7 迄は正しい座標値を与え：短期予測（結果 1 ～ 結果 8），それ以降は自分の出力結果を入力して次の座標値を予測する（結果 9 ～ 結果 16）．

エルマン型ネットワークの場合，図に示すように，結果 9 以降の予測結果が目標軌道に類似して得られ，長期予測が出来ていると言える．しかし，3 層ネットワークの場合，他の時点で長期予測に切り替えた場合でも（例えばデータ 9 まで短期予測），最終的にはこの図と同じ様に，結果 16 辺りに収束してしまい，短期予測と同様に予測不可能と言わざる負えない．

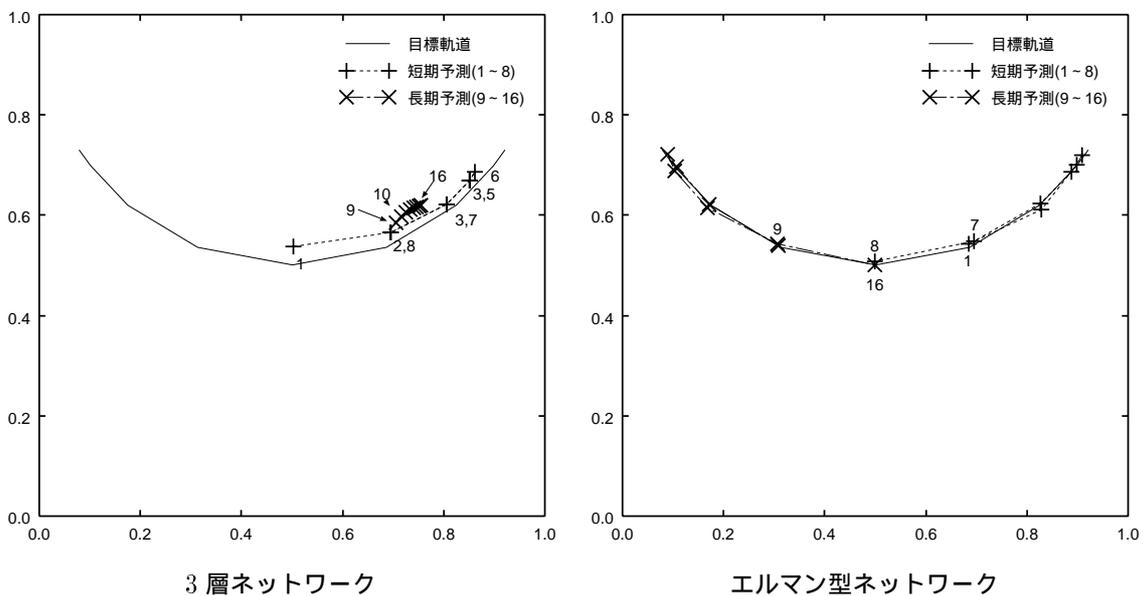


図 15 単振り子テスト結果 (学習：2000 回，長期予測 (9 ~ 16))

以上の計算機実験より，単振り子軌道の時系列予測問題においては，エルマン型ネットワークで可能な軌道の予測が，結合数が同じ 3 層ネットワークでは不可能な事が示された．

よって，次節以降はエルマン型ネットワークを用いた計算機実験とする．

4.2.4 二重振り子

二重振り子の場合，独立した軌跡が二本あるので，ネットワークの入出力層の素子数は各々 4 個となる．単振り子と同様に，中間層・コンテキスト層は 8 素子とす

る (図 17: 数字は素子番号) . ここでは, 上側軌跡の (x, y) 座標を素子 (1, 2), 下側軌跡の (x, y) 座標を素子 (3, 4) に入力し, 出力はそれぞれ (13, 14)(15, 16) としている . 本研究では, 素子 (1, 2) と素子 (3, 4) は別の軌跡等の座標データ内での依存関係は組み込まず, 対称構造を持つネットワークを使用している¹⁶ .

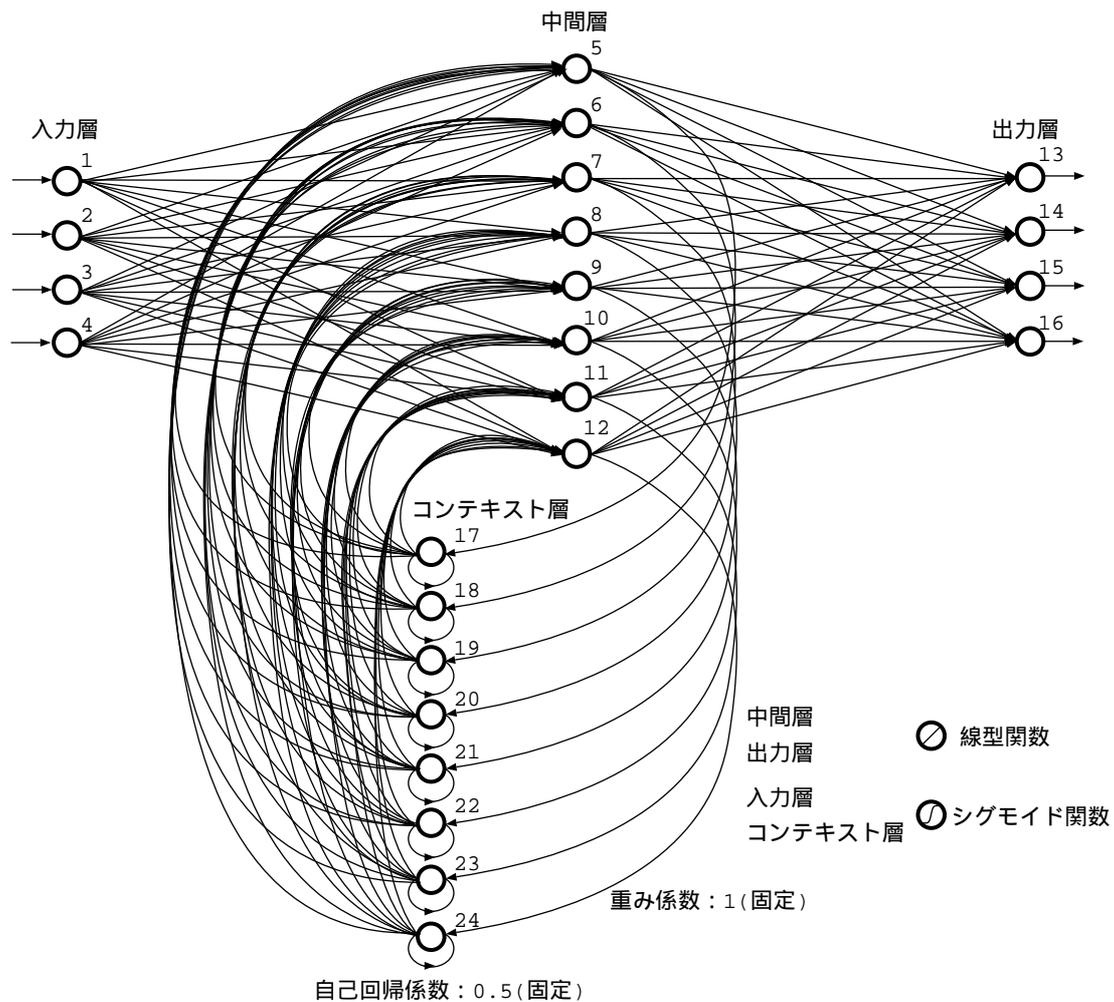


図 16 エルマン型ネットワーク (入出力層 4 個, 中間・コンテキスト層 8 個)

¹⁶ 上側軌跡の入力を (2, 4), 出力を (15, 16) としても同様な結果が得られる (当然ながら, 学習・テスト系列でのパターンに一貫性は必要である) .

ここでも，中間層からコンテキスト層への結合の重み係数は1に固定（パターン (t) での中間層の出力は対応するコンテキスト層のパターン $(t+1)$ での入力），コンテキスト層の自己回帰係数は0.5に固定である（コンテキスト層のパターン (t) での出力は0.5倍されて自分自身のパターン $(t+1)$ の入力となる）。

単振り子と同一な条件：Quickprop(学習条件：0.03, 2.0, 0.0005)，学習回数2000回での，エルマン型ネットワークが二重振り子軌道に対して，短期・長期予測可能である事を図17に示す。

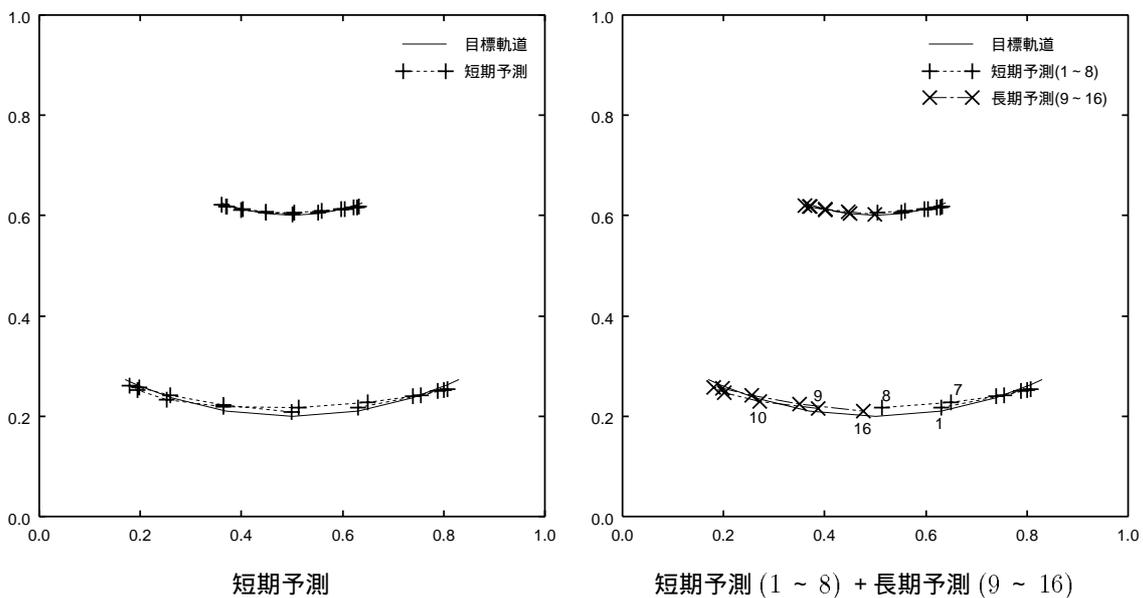
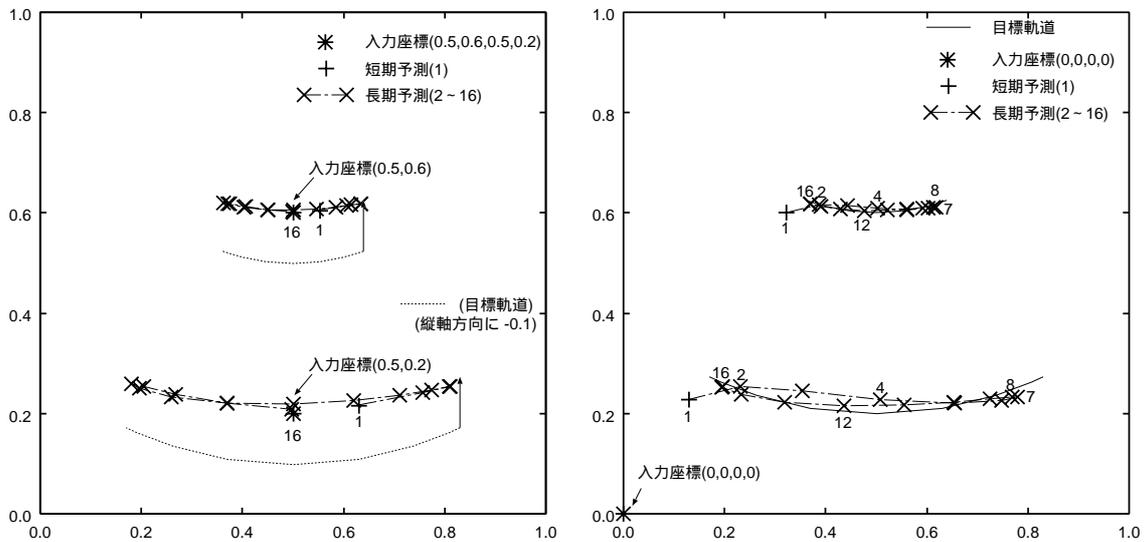


図17 二重振り子テスト結果(エルマン型ネットワーク, 学習: 2000回)

さらに，学習済みのネットワークに，初期値であるデータ 0 だけ入力し，それ以降を長期予測にした場合，初めに，軌道に無関係なデータ値 $(0, 0, 0, 0)$ を入力し，それ以降を長期予測にした場合の結果を，図18に示す(左図では目標軌道は縦軸方向に0.1下げて描画してある)。



初めだけデータ $0(0.5, 0.6, 0.5, 0.2)$ を入力

初めだけ $(0, 0, 0, 0)$ を入力

図 18 二重振り子長期予測 (エルマン型ネットワーク, 学習: 2000 回)

データ 0 だけ入力した場合でも (左図), 予測される軌跡はデータ $0 \sim$ データ 7 まで入力した時の長期予測の結果 (図 17: 右図) とほとんど変わらない. 学習とは無関係なデータ値 $(0, 0, 0, 0)$ を入力した場合の結果 (右図) でも, 学習させた軌跡のおおまかな形状を示している.

以上の様に, 3 層ネットワークでは, 履歴情報が必要な場合の軌道予測は不可能であるが, 時系列処理の為に提案された, エルマン型ネットワークの場合, 短期予測・長期予測も可能である事が確認できた.

次の章では, 実際の人間の動作から得られたデータを用いての計算機実験とその考察を行なう.

5. 計算機実験 2：人間の日常動作

ここでは、実際の人間の動作から得られたデータを用いてエルマン型ネットワークの各種能力について実験を行ない、その能力を考察する。

5.1 実験データ

人間の日常的な動き方は、同じ目的の動作でも個人差があり、同じ人間であっても、全く同じ動きを再現する事は難しい。この為、時間軸方向での処理を考えた場合、人間の動きの時系列データを時間方向である種の動作単位に文節化 (セグメンテーション) する事は非常に困難である。

一方、この動作のセグメンテーションは、人間の動作認識の研究において重要な課題でもある [60][43][11]。隠れマルコフモデル (Hidden Markov Model) を使用した大和らの動画データからの人物認識についての先駆的研究においても、事前に人間によってセグメンテーションする必要がある [65][66]。本研究では、事前のセグメンテーションは行なわずに、システムが自動的にこの問題に対処出来る可能性を持った枠組を提案する。

本研究では、3次元計測システム (OPTOTRACK, Northern Digital 製) を使用して計測された、実際の人間の動作の3次元時系列情報を元データとする¹⁷。人間の動きの特徴を示す、頭頂・右肩・右肘・右手首・右手の甲 (2個) 計6個所にマーカー (赤外線発光ダイオード) を装着し、3台の校正済みカメラにより各マーカーの3次元位置情報を周期: 120Hz で計測する¹⁸。

ここでは日常的な一つの動作を動作クラスとよぶ。今回は、次の二つの動作クラスを対象とした (図は人物を右側面から見た場合を、左から右に模式的に示した)。

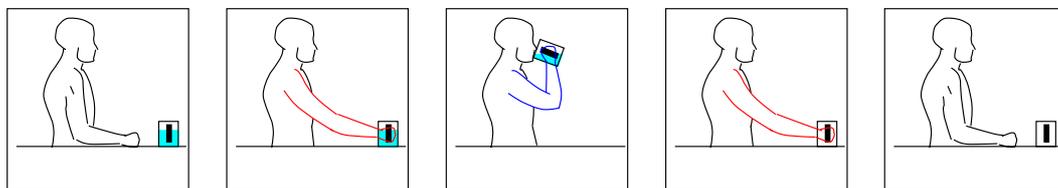


図 19 動作クラス 1(コップの水を飲む)

¹⁷実際には、Bruderlin, Armin 博士 (元 ATR 知能映像通信研究所研究員) の提供によるデータを使用 ([4] を参照)。

¹⁸この場合の精度は 0.5mm 以内 [4]。

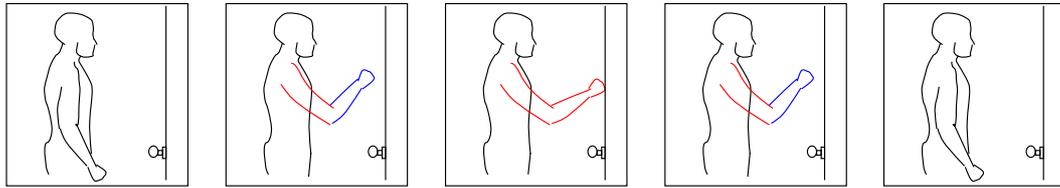


図 20 動作クラス 2(ドアをロックする)

動作クラス 1 は (図 19), 座っている姿勢で, 前方にあるコップを取り, 口に運び, 元の場所にコップを置き, 手を元の状態に戻す一連の動作である.

動作クラス 2 は (図 20), 立った状態で, ドアの頭の高さ程の位置を 3 回手でロックして, 手を元の状態に戻すという一連の動作である.

得られた 3 次元時系列データを計測開始と終了のデータを含む様に時間軸方向に等間隔にサンプリングする. 本研究では事前のセグメンテーションは行なわないので, 動作が計測時間途中で終了している場合でも, その時点でサンプリングを終了することはしない. 次に, 3 次元位置情報を視点方向に依存した表現に変換するために, サンプリングされた各データに対して視点位置を水平方向無限遠とした正射影変換を行ない, 視点方向に依存した 2 次元情報 (x, y) が得られる. ここでの視点方向 (View Angel) は, 図 21 に示すように頭頂を中心に真上から見て正面を 0 度として時計周り方向に値が増えていく角度で表現された値である.

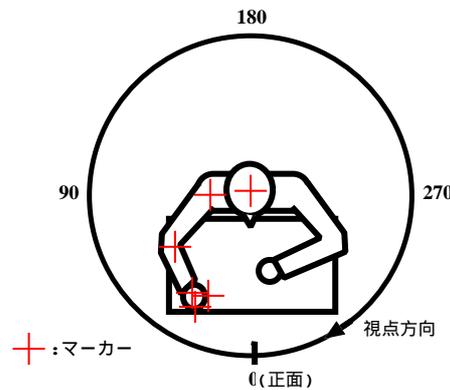


図 21 視点方向 (View Angel)

正射影変換の後に, (x, y) データを $(0 \leq x, y \leq 1)$ に正規化する. 本研究ではニューラルネットワークの出力層 (中間層・コンテキスト層) の素子にシグモイド

関数を使用する．図 3 に示したように，この関数の出力は $(0 \leq f(x) \leq 1)$ なので，学習データをこの範囲に収める必要がある． x 軸， y 軸方向を同じ比率で縮小し，全データの頭頂マーカの位置平均が $(0.5, 0.5)$ となるように並行移動し，大きさと位置の正規化を行なう．

以上のようにして，6 箇所の特徴点の視点方向に依存した 2 次元時系列データが得られる．時間軸方向にサンプリングされ，視点方向に依存して表現された 6 組の 2 次元データをフレームとよぶ．1 つのパターン (Pattern) は入力フレームと教師信号フレームからなる．

図 22 に，動作クラス 1 を 90 度の視点方向から変換した場合のフレームの模式図を，一部時系列で示す (右上の数字はフレーム番号)．ここでは，3 次元計測システムで得られたデータを，20 組のパターンとなるように 21 枚のフレーム (フレーム 0 ~ 20) にサンプリングしている．図では，模式的に頭頂以外の，肩から手先迄直線で繋げて示しているが，実際に使用するのは 6 組の (x, y) の実数値 $(0.0 \leq x, y \leq 1.0)$ である．

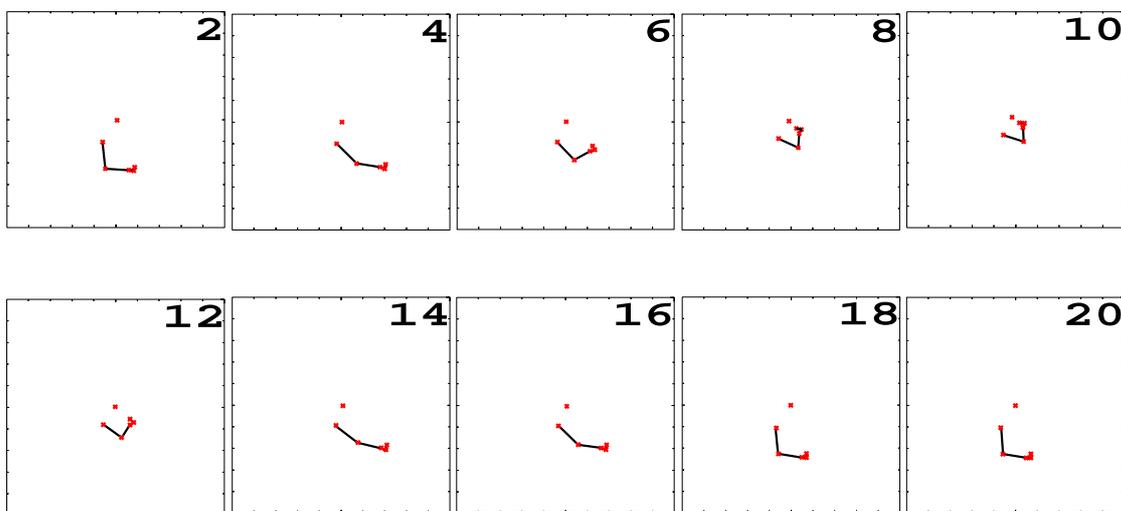
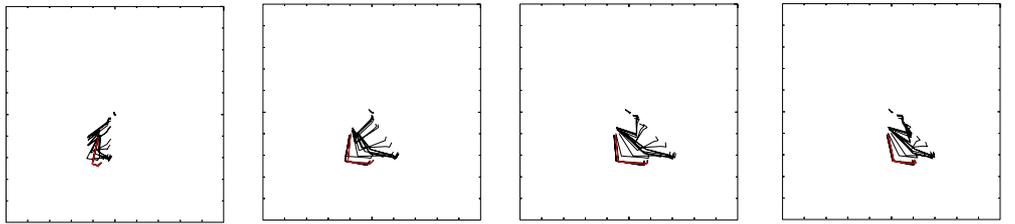


図 22 実験データ：動作クラス 1(90 度方向)， $(0 \leq x, y \leq 1)$

視点方向を変化させた場合での見え方の違いを図 23 に示す．



0 度方向：正面 40 度方向：斜め前方 90 度方向：真横 120 度方向：斜め後方

図 23 視点方向の変化：動作クラス 1

これらは、全てのフレーム (0 ~ 20 : 全 21 組) を重ね書きしている。

5.2 単一視点方向の予測実験

5.2.1 短期予測

初めに、単一視点方向での短期予測能力を検証する。学習・テスト系列として、動作クラス 1 の 90 度のデータを用いた。1 フレームあたりのデータが 6 組の (x,y) 座標なので、エルマン型ネットワークの入力層・出力層は 12 素子、ここでは中間層・コンテキスト層は 48 素子とした。コンテキスト層の初期値は 0.5、コンテキスト層の自己回帰係数 α は 0.5 と固定してある。学習条件を 4 章と同じく ($\eta : 0.03$, $\mu : 2.0$, $\lambda : 0.0005$) した場合、図 24 に示すように発振が起こる。

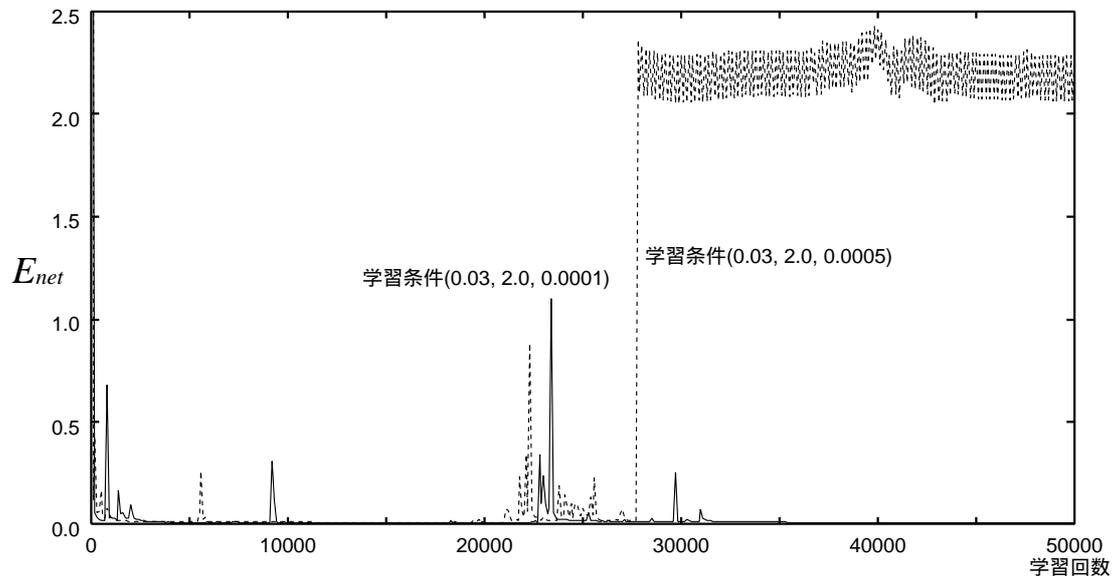
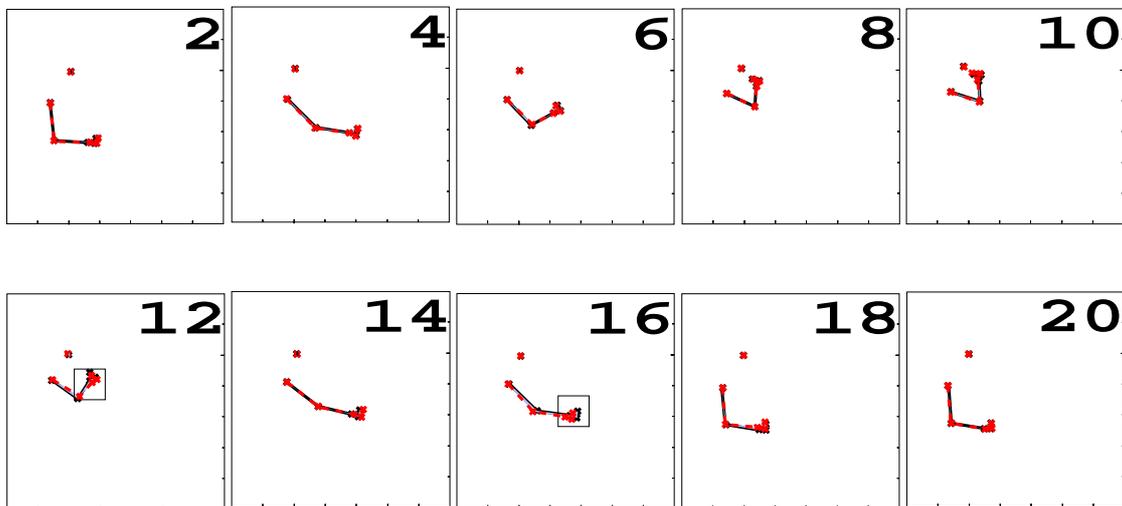


図 24 学習曲線：動作クラス 1(90 度)

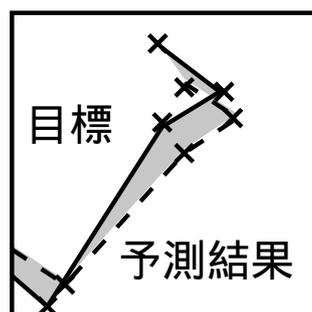
重み変更抑制係数： λ を 0.0001 に下げると発振はおさまり，学習は安定して行なわれた．4章で使用した振り子系の動きに比べ，このデータの方が直前の学習で得られた重み係数に依存した，緩やかな学習が必要と考えられる．これは学習回数が 50000 回必要という事からも推測出来る．

これ以降の計算機実験では，学習条件を ($\eta : 0.03$, $\mu : 2.0$, $\lambda : 0.0001$) , 学習回数は 50000 回とする．

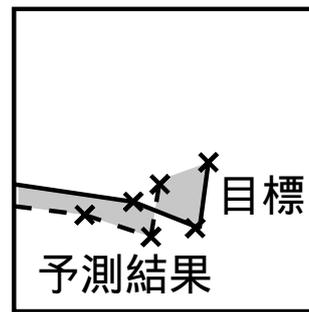
上の条件で学習したネットワークを用いた短期予測の結果を，頭部を除いた腕から手先の 5 点を繋げた模式図にして，図 25 に示す．



動作クラス 1(90 度)の短期予測 ($0.3 \leq x, y \leq 1.0$)



フレーム 12 (拡大図)



フレーム 16 (拡大図)

図 25 動作クラス 1(90 度)の短期予測

上2段は、目標と予測結果を重ね(実線：目標，点線：予測結果)，時系列状に並べたものである(ただし，縦横軸とも $0.3 \leq x, y \leq 1.0$ の範囲を切り出し，右上を(1.0, 1.0)に揃えている)。目標・予測結果がほぼ一致しているため，最下段に最も予測誤差が大きいと考えられるパターン12，パターン16の結果の一部を拡大して示す。

縦軸を6点全ての E_p (式6：以下予測誤差)とし，横軸をパターン(時間方向)として，図26に示す(図には視点方向：40度の結果も描画)。

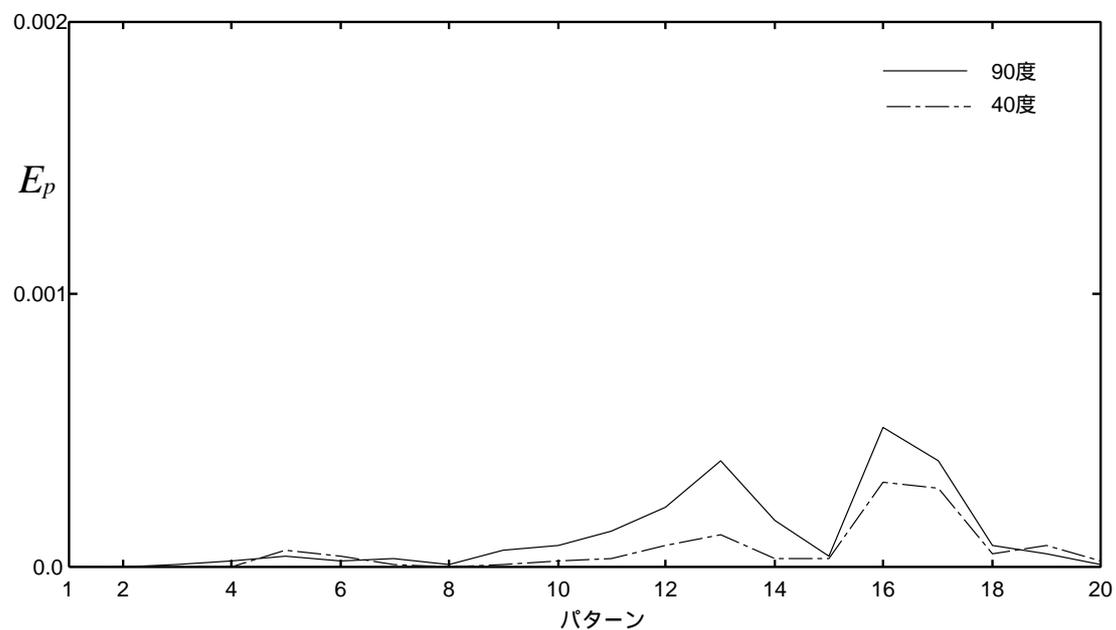


図 26 時間方向に対する予測誤差 (動作クラス 1：40 度，90 度)

90 度方向を学習したネットワークに，0 度から 350 度迄 10 度ごとの視点方向から作成したデータにより短期予測させた予測誤差の，3 次元グラフを図 27 に示す(横軸：視点方向，縦軸： E_p ，奥行き方向軸：パターン)。図 28 には，40 度方向を学習したネットワークの場合を示す。

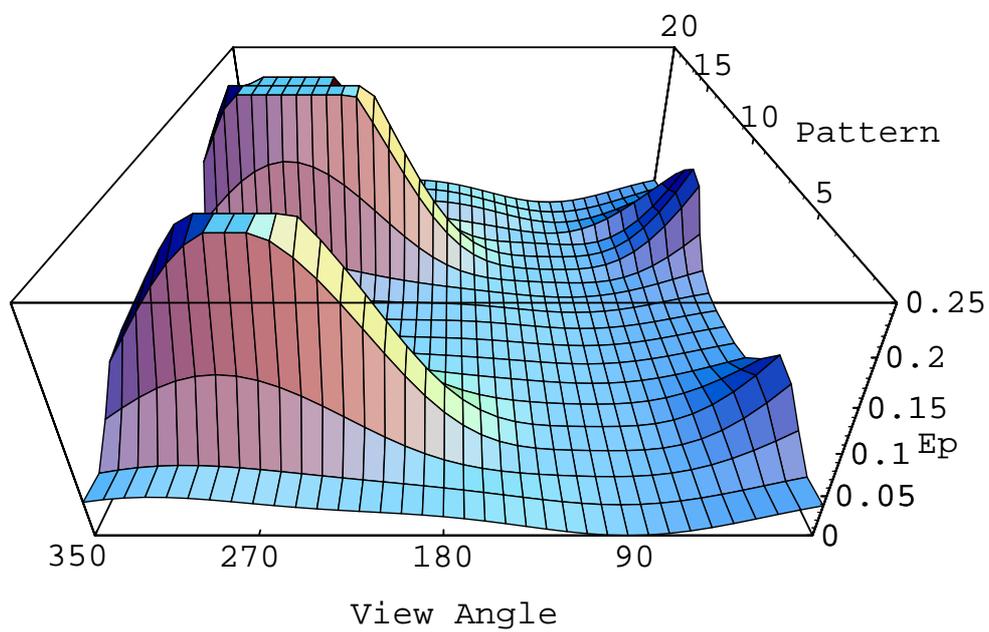


図 27 視点方向と時間方向に対する予測誤差 (動作クラス 1: 90 度)

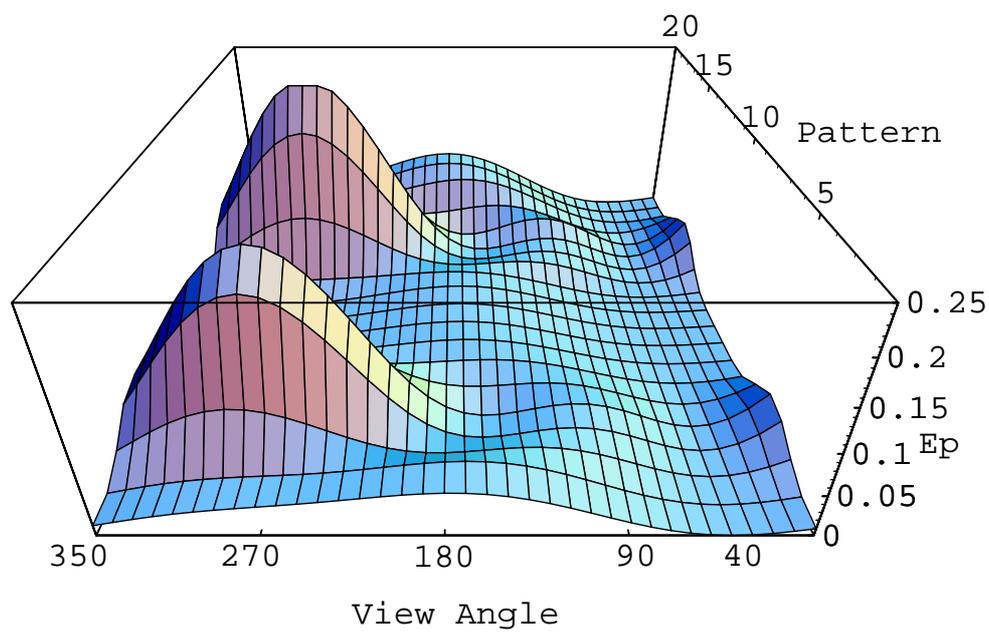


図 28 視点方向と時間方向に対する予測誤差 (動作クラス 1: 40 度)

これらの3次元グラフより，学習した視点方向から離れるに従い(= 入力するデータが徐々に変化)，予測誤差が徐々に増加していくというネットワークの視点方向変化に対する特性が確認できる．

0度，40度，90度の視点方向データで各々学習した，ネットワークの視点方向変化に対する，予測誤差のパターンでの平均を図29に示す．

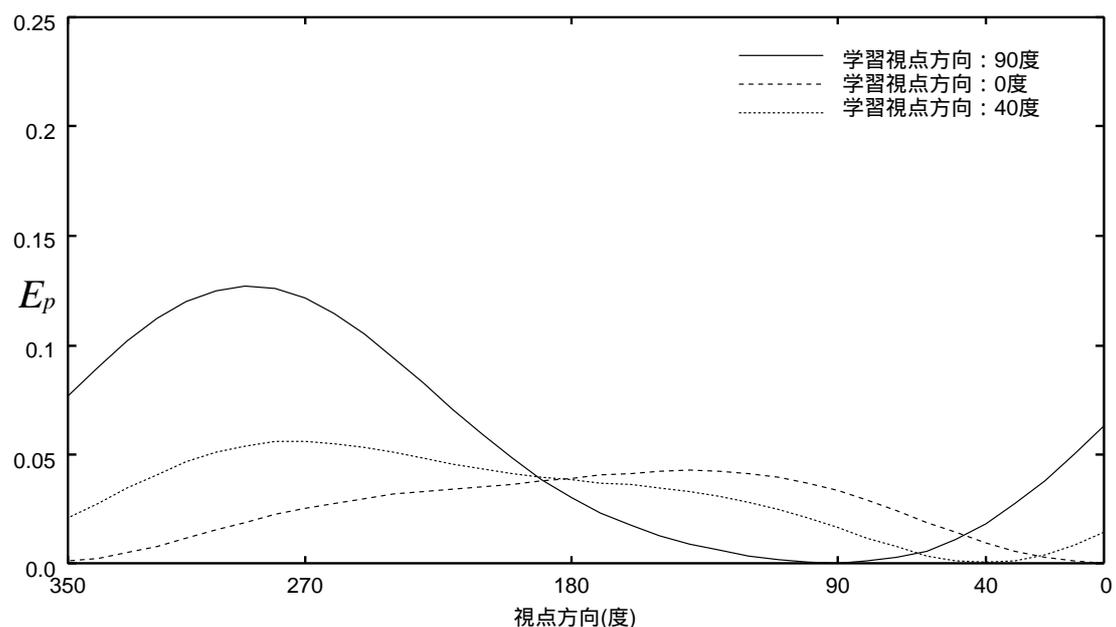


図 29 視点方向に対する平均予測誤差 (動作クラス 1: 0・40・90 度)

5.2.2 長期予測

ここでは，エルマン型ネットワークにおける長期予測能力を検討する．

前節の学習済みネットワーク (動作クラス 1: 90 度) に対して，パターン 2 迄データを与え¹⁹，その出力 (パターン 2 の短期予測結果) を次の予測の為のとし，以下続けた結果を図 30 に，時系列で示す (ただし短期予測の結果と同様に，縦横軸とも $0.3 \leq x, y \leq 1.0$ の範囲を切り出し，右上を (1.0, 1.0) に揃えている)．

¹⁹実際はデータ 0，データ 1 を入力．

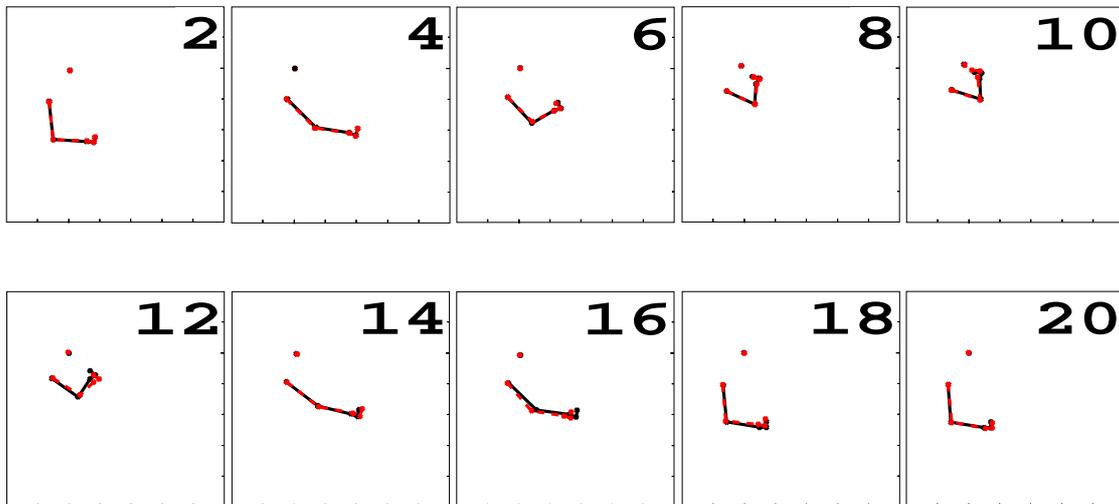


図 30 動作クラス 1 : 90 度の長期予測 ($0.3 \leq x, y \leq 1.0$)

図 31 に示すように，極めて限られた入力情報だけからでも，短期予測とほぼ同様な予測結果が得られている．

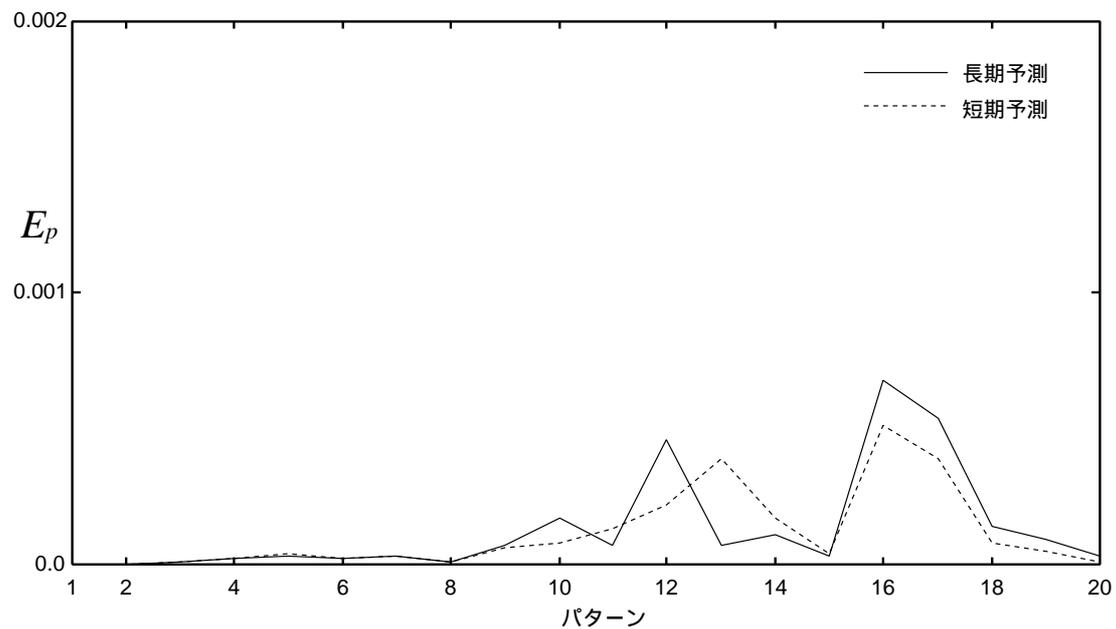


図 31 時間方向に対する長期予測誤差 (動作クラス 1 : 90 度)

5.3 複数視点方向に対する予測実験

5.3.1 短期予測

ここでは、同じ動作クラスに対する複数の視点方向からの学習を検討する。

本研究では、二つの異なる視点方向の系列を同時に学習させるために、二つの系列を2回ずつ交互に学習させた(以下同時学習とよぶ)。ネットワーク構成・学習条件等は単一視点方向での実験と同一とした²⁰。

図32に、動作クラス1の0度と40度方向からび系列を、同時学習にしたネットワークの短期予測結果を、3次元グラフを示す。

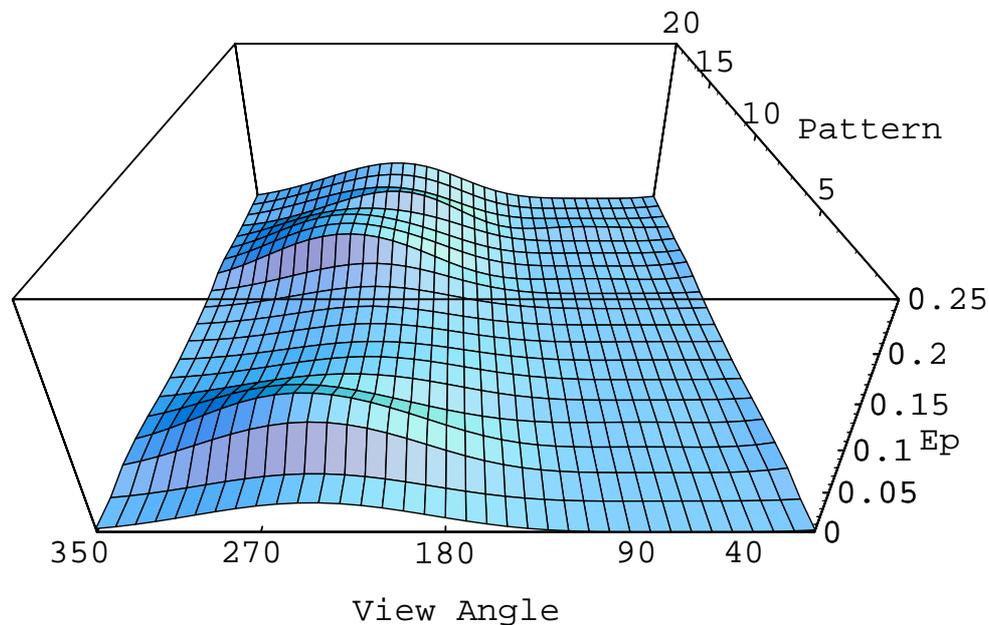


図 32 視点方向と時間方向に対する短期予測誤差 (40 度・90 度同時学習)

3次元グラフより、学習対象の40度と90度の間での視点方向の変化に対しては、予測誤差は低くなっているのが読み取れる。

学習対象の40度・90度、およびその間の60度と70度の視点方向での、時間軸方向に対する短期予測誤差を、図33に示す。

²⁰入・出力層：12素子，中間・コンテキスト層：48素子。
全学習回数：50000回，(η :0.03， μ :2.0， λ :0.0001)

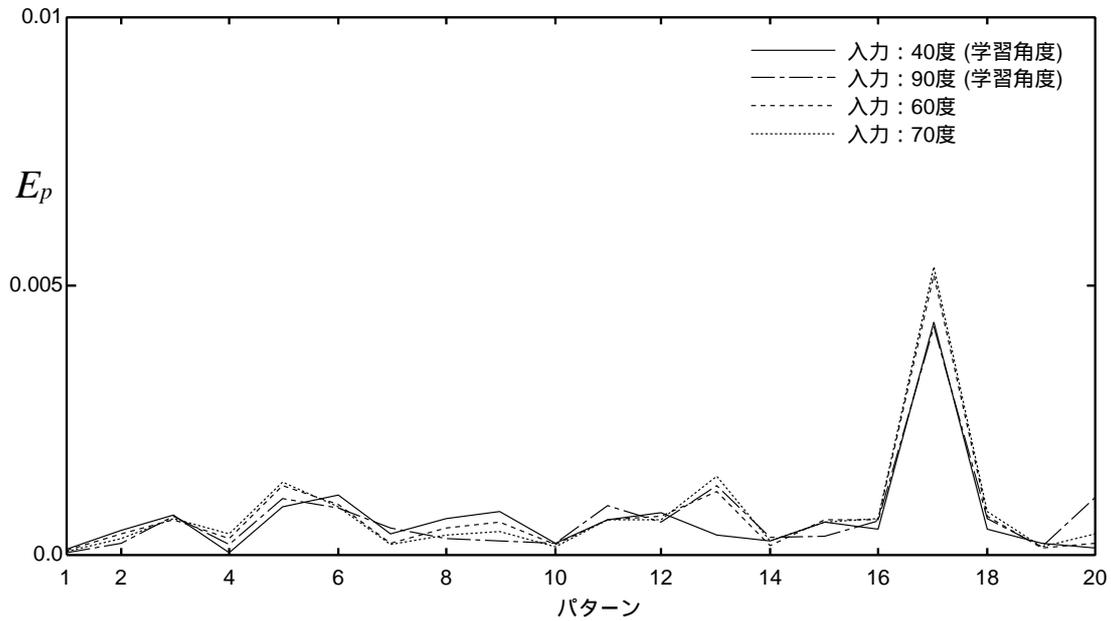


図 33 時間方向に対する短期予測誤差 (40 度・90 度同時学習)

縦軸を予測誤差の時間平均とした視点方向に対する予測誤差を図 34 に示す。40 度・90 度の同時学習以外に 0 度・90 度の結果もあわせて示す。これからも、同時学習間では、予測誤差が低く抑えられているのが解る。

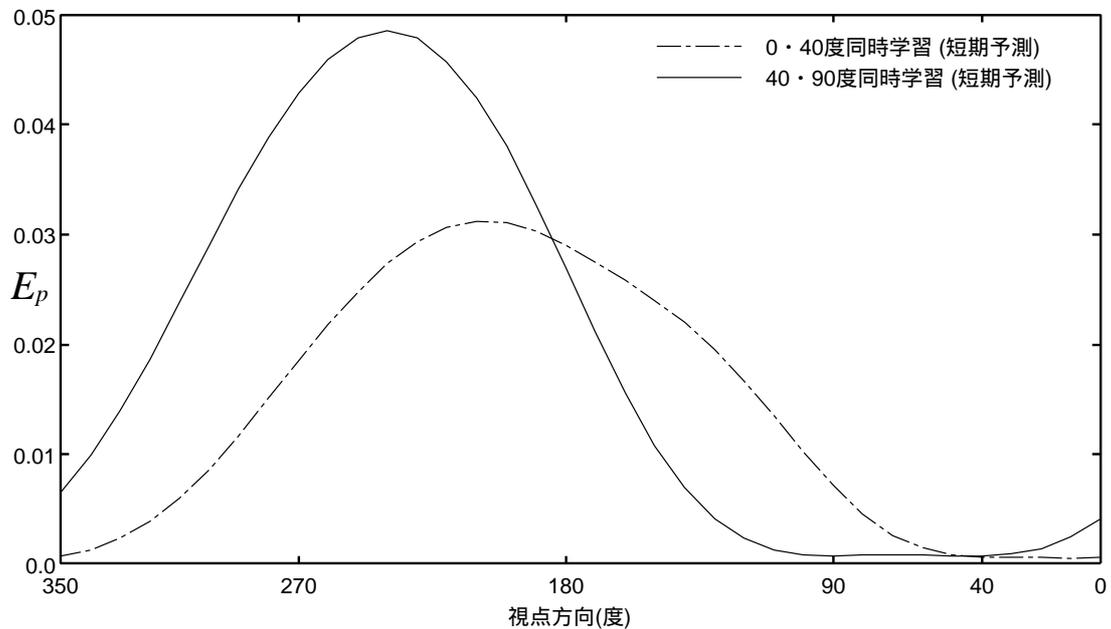


図 34 視点方向に対する平均短期予測誤差 (0・40, 40・90 度同時学習)

5.3.2 長期予測

複数視点方向の系列を，同時学習したネットワークでの長期予測能力を確認する．前節で使用した，40度・90度同時学習したネットワークで，5.2.2節と同様な長期予測²¹．を行なった結果を，図35に示す．

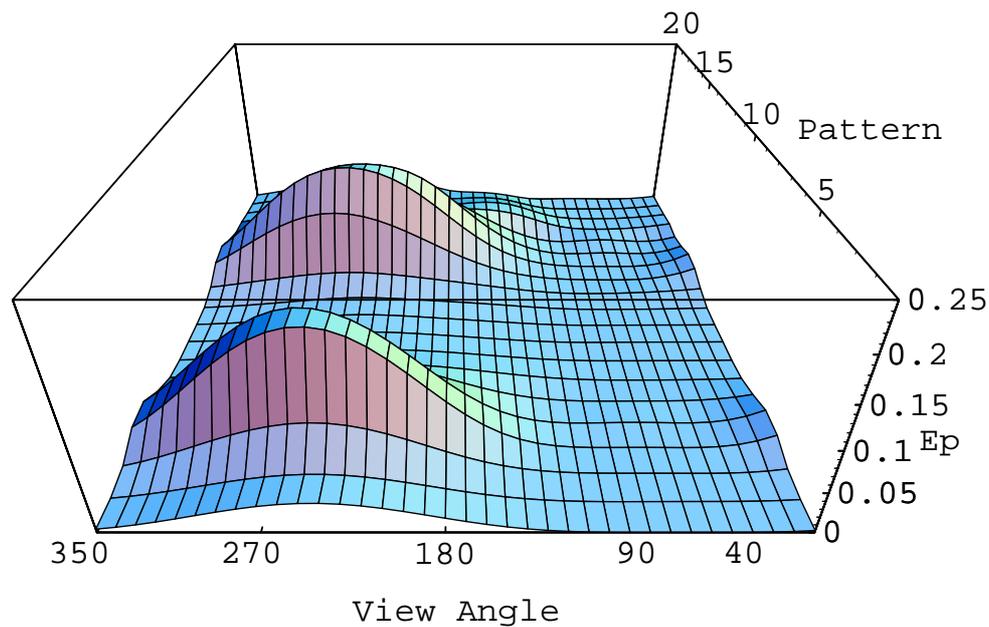


図 35 視点方向と時間方向に対する長期予測誤差 (40度・90度同時学習)

複数視点の短期予測と同様に，40度・90度同時学習の時間方向に対する長期予測誤差を，図36に示す．短期予測の場合と比べると，後半での予測誤差が若干大きくなっている．

図37に，同時学習での，視点方向に対する平均予測誤差を，短期予測と長期予測の結果を重ねて示す(ただし，縦方向の表示領域は図34の2倍)．

²¹初めの2パターンだけ正しいフレームを与える．

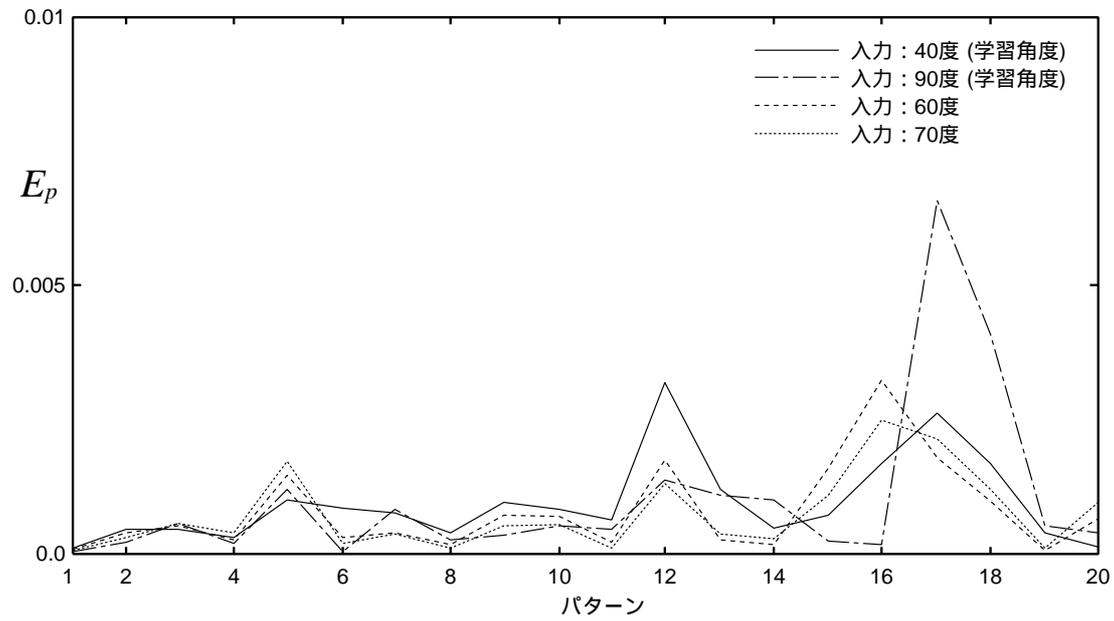


図 36 時間方向に対する長期予測誤差 (40・90度同時学習)

平均長期予測誤差の場合，0・40度での長期予測での誤差の拡大が目立つ．しかし，どちらも学習した区間での誤差はさほど拡大はしていない

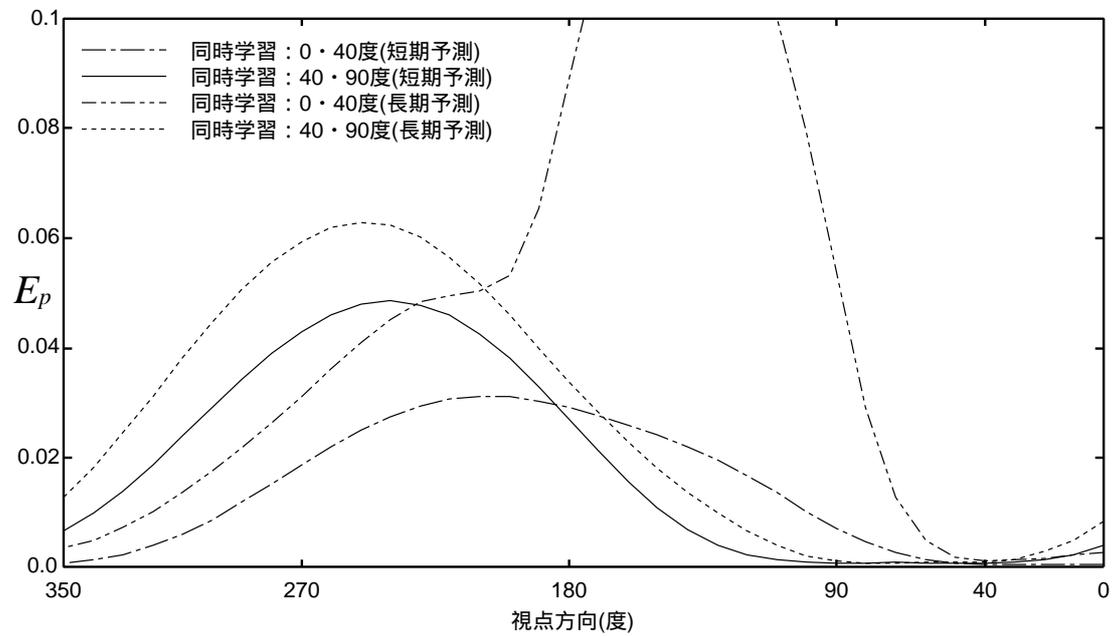


図 37 視点方向に対する平均長期予測誤差 (0・40, 40・90度同時学習)

5.4 識別実験

ここでは、異なる動作クラスの識別能力について検討する。

動作クラス1(90度)で学習したネットワークに、動作クラス2(90)0度を入力した結果を図38に時系列で示す(目標を実線、予測結果を点線で描画する。また前の結果と同様に、縦横軸とも $0.3 \leq x, y \leq 1.0$ の範囲を切り出し、右上を(1.0, 1.0)に揃えている)。また、図39に時間方向での予測誤差を示す。

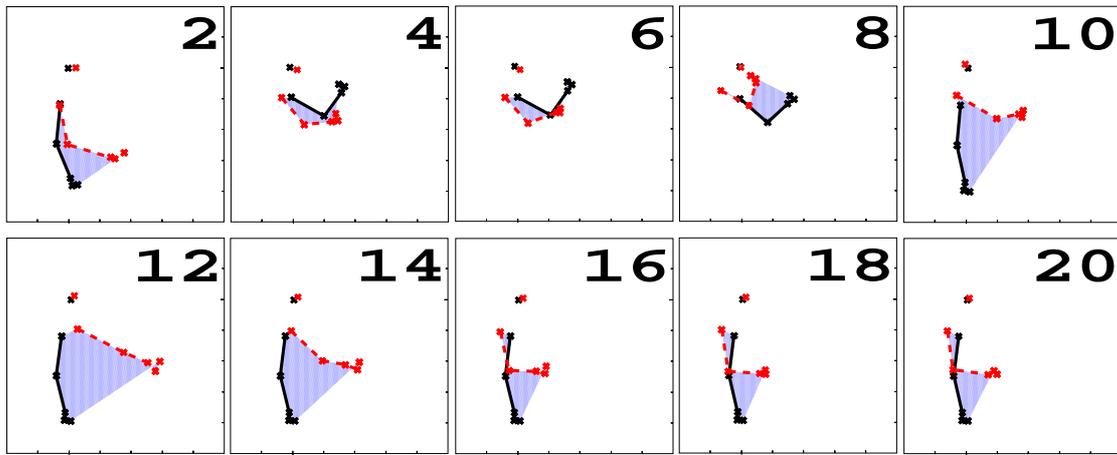


図 38 識別実験 (学習 : クラス 1(90 度) / テスト : クラス 2(90 度))

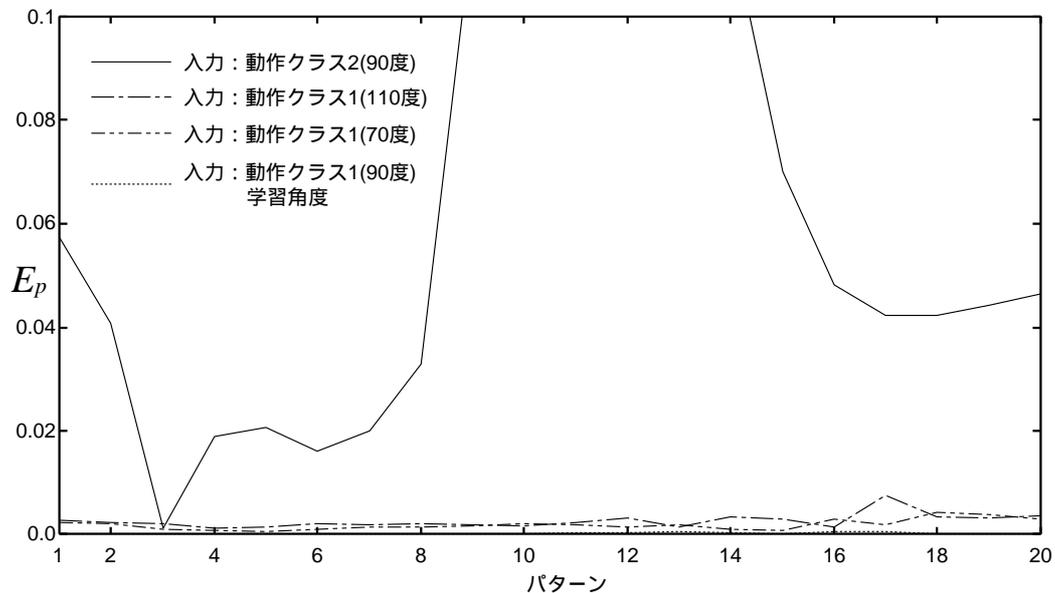


図 39 時間方向に対する予測誤差 (クラス 1(90 度) / クラス 2(90 度)) .

図 38より，目標 (実線) と予測結果 (点線) の違い (網かけ領域で強調) は非常に大きいことが解る．

動作クラス 1(90 度) を学習したネットワークに，動作クラス 1(70 度・90 度・110 度)，動作クラス 2(90 度) を入力した場合での時間方向に対する短期予測誤差のグラフを図 39に示す．学習したパターンの視点方向変化による誤差に比較して，異なる動作クラスのパターンを入力した場合での予測誤差の方が大きい事がわかる．これより，予測誤差に対して適切な閾値を設定すれば，異なる動作クラスの識別が可能であると言える．

図 39に，視点方向と時間方向に対する予測誤差の 3 次元グラフを示す．図 27と比較して，全体に渡って誤差が大きい事は明らかである．

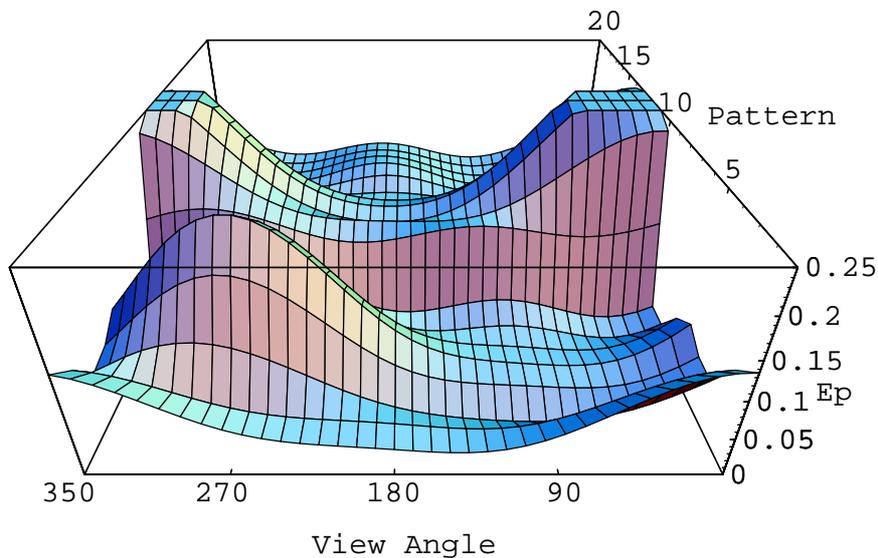


図 40 視点方向と時間方向に対する予測誤差 (クラス 1(90 度)/ クラス 2)

図 39のグラフにおいて，動作クラス 2 を入力した場合でもフレーム 3 での誤差がかなり低くなっているが，これは，フレーム 3 での両者の動作が動作クラス 1 の場合には手を延ばす途中で，また，動作クラス 2 の場合には手を上げる途中でフレームが酷似しているためである

5.5 コンテキスト層の自己回帰係数

5.5.1 自己回帰係数と短期・長期予測

エルマン型ネットワークの特徴であるコンテキスト層での、自己回帰係数の値の違いによる予測への影響を調べる。

3.3.1節で説明したように、時刻 (t) でのコンテキスト層の素子出力： $S(t)$ は時刻 $(t-1)$ での中間層の出力を $H(t-1)$ とすれば

$$\begin{aligned} S(t) &= \alpha S(t-1) + H(t-1) \\ &= \sum_{r=1}^t \alpha^{r-1} H(t-r) + \alpha^t S(0) \end{aligned}$$

と表わされる。

図 41 に、 $y = \alpha^t$ ($\alpha = 0.25, 0.5, 0.75$) のグラフを示す ($0 \leq t \leq 16$)。

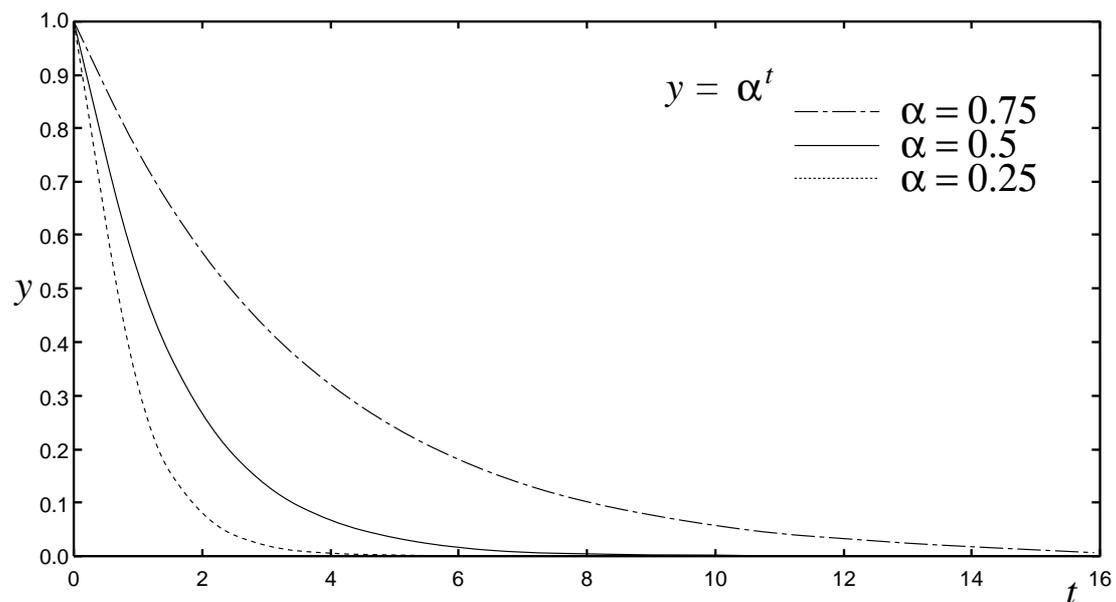


図 41 $y = \alpha^t$, $\alpha = 0.25, 0.5, 0.75$

このように α の値により、減衰にかなりの差がみられる。

ここでは、 α の値を 0.25, 0.5(標準), 0.75 および 0[?]Elman オリジナル: Elman:90) とした場合の、予測能力を比較検討してみる。動作クラス 1(90 度)の短期予測での、時間方向に対する予測誤差を、図 42 に示す。

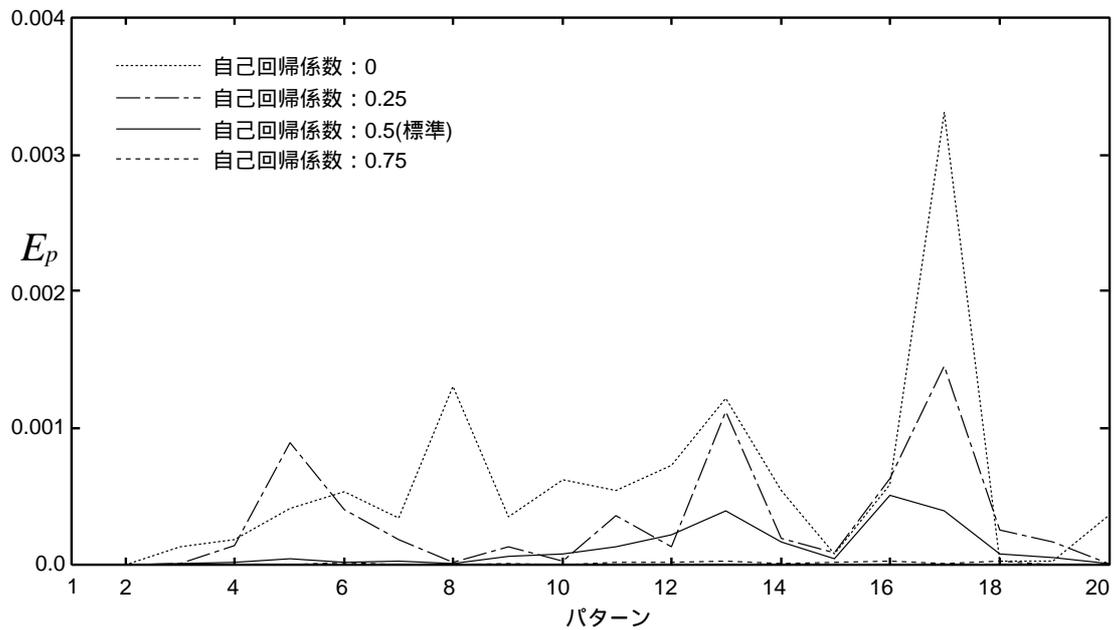
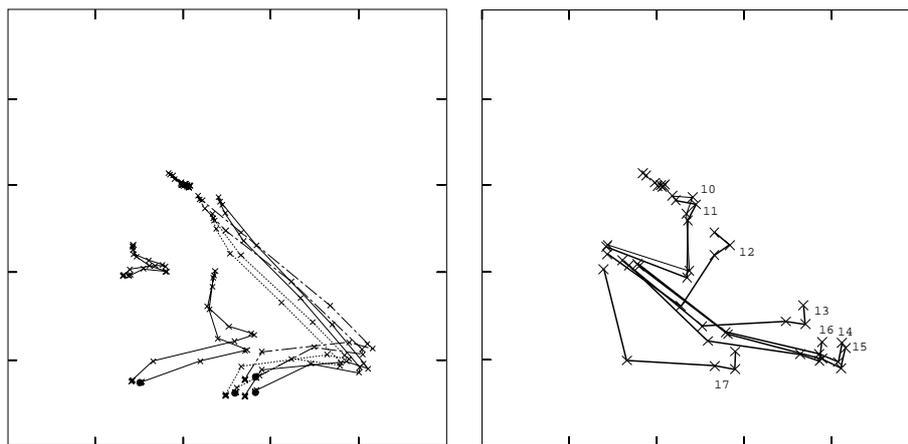


図 42 自己回帰係数の違いによる短期予測誤差

特に、フレーム 13, 17 での各係数での予測誤差の隔差が大きい。フレーム 13 では、口元での動作 (フレーム 8 ~ 12) の直後のコップを台に置く動作開始、フレーム 17 は、コップを台に起きしばらく静止 (フレーム 14 ~ 16) してから、急に動き出す部分である。どちらも、動き初めのタイミングを得るためにも過去の履歴情報が必要である。コンテキスト層の自己回帰の役割が大きく、この値の変化が予測誤差に与える影響が大きいと考えられる。(図 43, ただし $(0.2 \leq x, y \leq 0.7)$)。



6 点の軌道 (パターン 0 は)

フレーム 11,12,13, および 15,16,17

図 43 動作クラス 1(90 度)の軌道

図 44に、自己回帰係数：0，0.25 の場合での、パターン 17 の短期予測の結果を破線で示す ($0.2 \leq x, y \leq 0.7$)。どちらも、目標まで達っていないのがわかる。

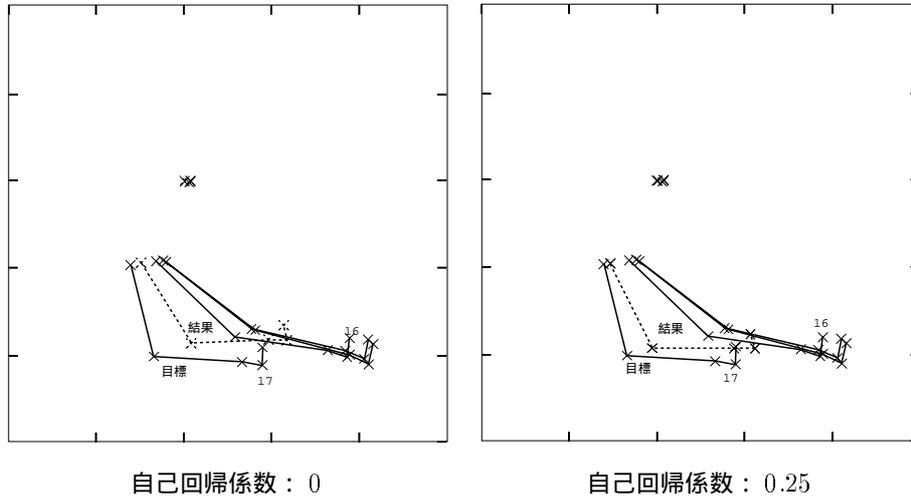


図 44 コンテキスト層の効果：パターン 17 の短期予測

長期予測の結果を、図 45に示す (縦軸は短期予測：図 42の 10 倍)。自己回帰係数が 0 の場合、予測誤差が飛び抜けて大きい。この結果より、長期予測の場合においてはコンテキスト層での自己回帰が必要不可欠なのがわかる。

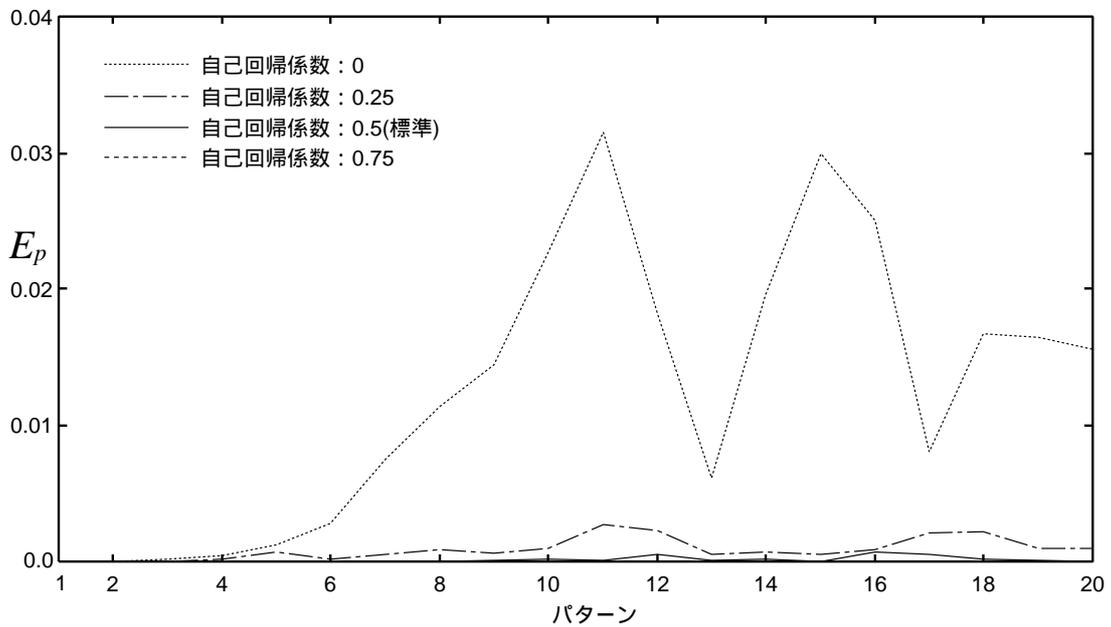


図 45 自己回帰係数の違いによる長期予測誤差

5.5.2 パターン途中からの入力

時系列データの場合，実際に入力されるデータが，学習したものと同じくフレーム 0 から入力されるとは限らない．そこで，フレーム 0 から学習したネットワークに，途中からのフレームを入力してみる（これは，その時点でコンテキスト層の全素子の出力値を強制的に 0.5(初期値)にするのと同じ効果がある）．ここでは，動作クラス 1(90 度)を，パターン数：40 で学習したネットワークを使用する²²．

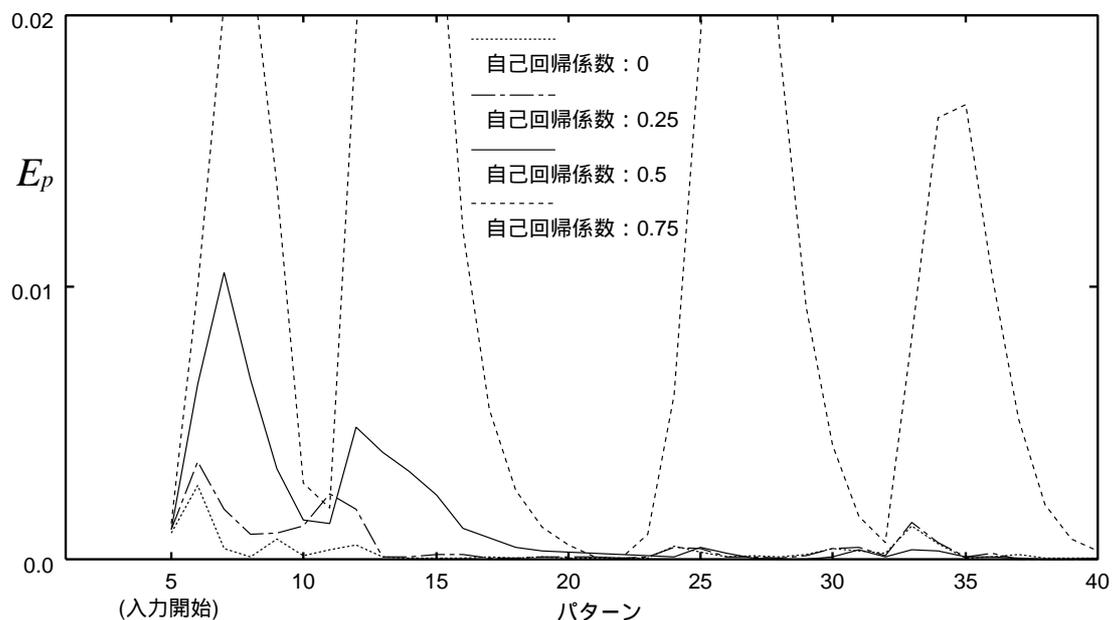


図 46 パターンの途中からの入力・短期予測 (40 パターン)

この場合，短期予測 (図 42) で最も予測誤差が小さかった自己回帰係数：0.75 の値が非常に大きく，逆に，自己回帰係数：0 の予測誤差が一番小さくなっている．また，自己回帰係数：0, 0.25, 0.5 の場合，時間が経つ (パターンが進む) 程，予測誤差が低下していく．これらの場合，コンテキスト層の各素子が徐々に正しい値に近づいていくからと推測できる．0.75 の場合は，逆に，過去の間違った情報に囚

²²今迄使用してきたデータのサンプリング間隔を半分にしたもの．
自己回帰係数：0, 0.25での短期予測誤差の最大値は0.001程度，自己回帰係数：0.5では0.0003(共にパターン数：20でのフレーム17に対応)，自己回帰係数：0.75では0.0001．

わてしまい，予測結果が悪くなると考えられる．当然，フレーム 0 から入力した場合，フレーム 5 の時点で，コンテキスト層の全ての素子の出力値が 0.5 ではない²³．

パターン数：20 の場合では，図 47 に示すように予測誤差の低減はあまりみられない（縦軸は，図 46 の 2 倍）．これより，コンテキスト層の各素子が通常の短期予測の値に近づくには，ある一定個の正しい入力値が必要であると推測できる．

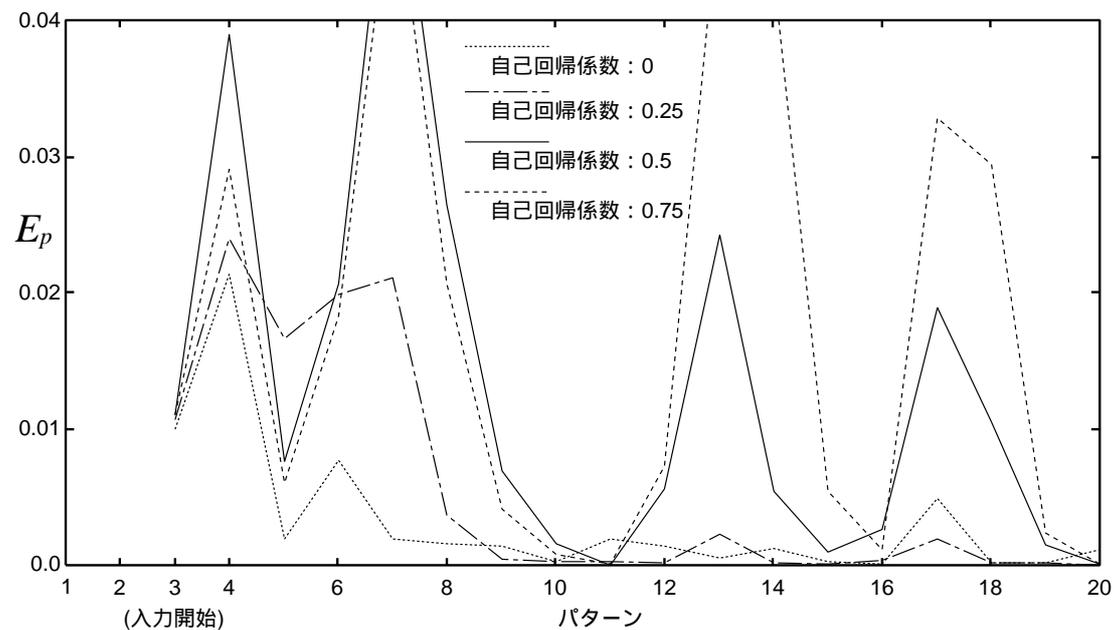


図 47 パターンの途中からの入力・短期予測 (20 パターン)

本節の二つの計算機実験より，コンテキスト層の自己回帰係数は 0.5 で適切と結論できる．

²³自己回帰係数：0.75 の場合，24 素子中 0.4 ~ 0.6 の値は 6 素子，1.0 以上が 14 素子，2.0 以上が 7 素子ある（コンテキスト層の各素子は線型関数）．

6. まとめと今後の課題

6.1 計算機実験のまとめ

エルマン型ネットワークを用いて、以下の2つのアプローチ

- 人間の身体を再構築する、
モデルベース的なトップダウン手法ではない。
- 入力表現の要素間に明示的な拘束条件を設けない
(問題領域に関しての前提知識を必要としない)。

による計算機実験を行なった。

その結果、以下のことが示された、

1. 過去の履歴情報が必要な軌道に関しての、
短期予測・長期予測での3層ネットワークに対する優位性
2. 人間の日常動作から作成された時系列パターンに対しては、
以下の4つの予測が可能である：
 - 短期予測
 - 長期予測
 - 複数視点方向の系列に対しての同時学習と短期予測
 - 複数視点方向の系列に対しての同時学習と長期予測
3. 予測誤差を適切に行うことにより、異なる動作クラスの識別が可能である。
4. コンテキスト層の自己回帰は、以下の場合に必要である：
 - 短期予測
 - 長期予測
 - パターンの途中からの入力

6.2 今後の課題

ニューラルネットワークによる研究一般に言える事であるが，問題対象に必要な中間層の素子数，学習条件 (学習係数 (η)・最大変化量 (μ)・重み変更抑制係数 (λ))，学習回数は，本研究で行なったように，実験結果に基づくしかない．これらの項目に関する指標が得る為には，現時点では，さらなる計算機実験を行い，経験的にデータを集めるしか方法がない．

特に，複数視点方向に対しての学習 (5.3節) の場合，どのような組み合わせ (範囲) で学習する事が出来るか (複数視点方向の学習可能を条件) を確かめる必要がある．例えば，本研究では，視点方向 40 度と 90 度の同時学習，0 度と 40 度の同時学習が良い結果を出しているが，今のネットワーク構造で，0 度・40 度・90 度の 3 系列の学習の可能性を調べる事は，有意義な研究と言える．

今後の拡張性として次の二点が考えられる．

- 濃淡画像を用いた計算機実験
- モジュール構造化

6.2.1 濃淡画像を用いた計算機実験

本研究で使用したエルマン型ネットワークは対称構造であり，対象に対しての特別な先験的知識は入れていない．そのため，濃淡画像を直接ネットワークに入力・出力する事も可能である．32 × 32 画素の濃淡画像を用いた場合，ネットワークの入出力層の素子数は各 1024 個となる．この場合のデータを図 48 に示す．



図 48 32 × 32 画素の濃淡画像データ

6.3 モジュール構造化

図 49に識別システムの為のモジュール構造を示す．これは，Jordan らによって提案された，階層構造を持つモジュール型のネットワーク (Hierarchical Mixture of Expert)[30] を，エルマン型ネットワークに拡張したものである．下段の網かけの四角形は，それぞれ本研究で述べたエルマン型ネットワークである．これらのネットワークは，個別に各クラスで学習したものである（それ故，エキスパートネットワークとも呼ばれる）．これらエキスパートの出力とテストデータから，ゲートネットワークが適当にこれら二つのネットワークの出力を調整し，システム全体として予測誤差を低下させようという考えである．

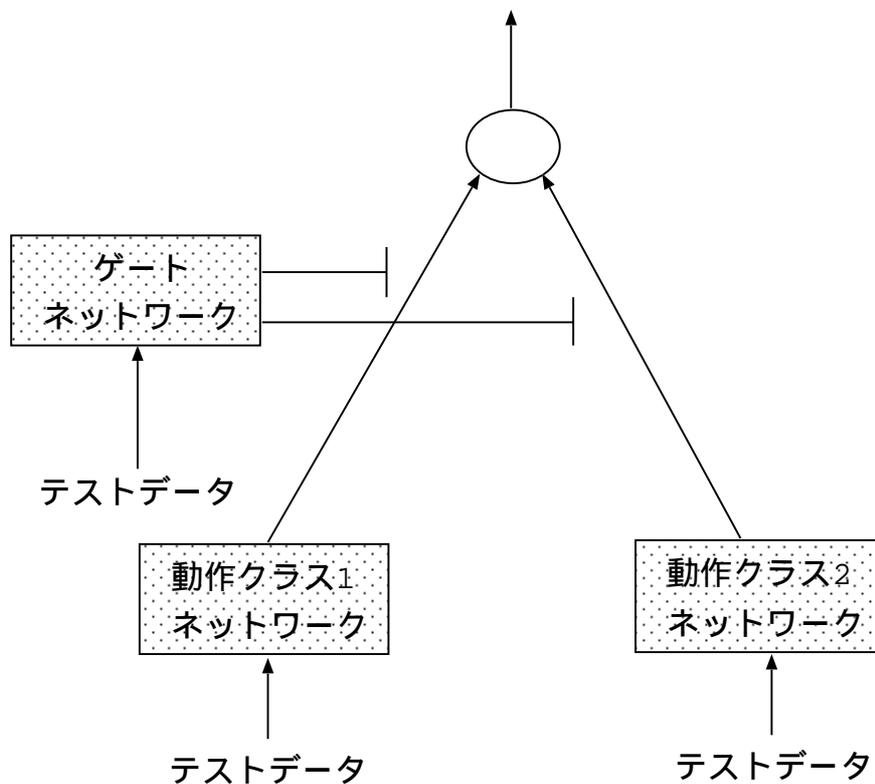


図 49 識別システム (モジュール構造)

謝辞

本研究を遂行するにあたり，終始ご指導を頂きました，本学基幹講座：ソフトウェア基礎講座 横矢直和 教授に厚く御礼申し上げます．

本研究は，本学および(株)ATR 人間情報通信研究所との協力により開設された，連携講座：人間情報処理学で行われたものです．非常に恵まれた環境で研究する機会を与えて下さり，さらに，研究を進めるにおいて終始適切かつ親身なご指導と助言を頂きました，東倉洋一 連携講座教授((株)ATR 人間情報通信研究所)に心から深く感謝いたします．

副審査委員として助言を頂きました，本学 音情報処理講座 鹿野 清宏 教授に深く感謝致します．基幹講座においてまた副審査委員としてご指導を頂きました，ソフトウェア基礎講座 竹村治雄 助教授に厚く御礼申し上げます．

銅谷賢治 連携講座助教授からは，研究進行・論文作成等において大変有益なアイデアを頂きました，ここに感謝いたします．

安藤広志 ATR 人間情報通信研究所主任研究員からは，本研究についての基本アイデア，さらに研究進行について 2 年間に渡り終始親身な助言を頂きました，ここに厚く御礼申し上げます．

一ノ瀬 裕 (株)ATR 人間情報通信研究所社長には，真に刺激的で素晴らしい研究環境を提供して頂きました，ここに厚く御礼申し上げます．

本研究で使用した実験データを，快く提供して下さった Bruderlin, Armin 博士(元 ATR 知能映像通信研究所研究員)に，この場をかりて深く感謝致します．

ソフトウェア基礎講座 岩佐英彦 助手，山澤一誠 助手には，研究への助言，本学での研究環境の整備等に協力して頂きました，深く感謝致します．

連携講座関係研究員の皆様からは，数限りない貴重なアドバイスを頂き本当に有り難う御座いました．(株)ATR 人間情報通信研究所第 3 研究室 川人光男 室長には，研究への情熱さらに真摯に打込むその姿勢から，研究する事の素晴らしさと厳しさを学ぶことが出来，大変感謝しております．また，(株)ATR 人間情報通信研究所研究員皆様，および科学技術振興事業団 ERATO 川人学習動態脳プロジェクト研究員皆様，日々の討論・助言は研究進行においての貴重な道案内となりました，心から感謝いたします．

共に学び，苦しい時には励ましてくれた，連携講座：人間情報処理学所属の学生

諸兄，本学基幹講座であるソフトウェア基礎講座所属の学生諸兄には，感謝の念にたえません．

さらに，日々の研究活動を支えて下さった，ソフトウェア基礎講座事務補佐員 福永博美 女史，(株)ATR 人間情報通信研究所第3研究室秘書 難波美恵子 女史に心より感謝申し上げます．また，連携講座の活動を支えて下さった，(株)ATR 人間情報通信研究所企画課およびテクニカル・サポート・グループ (TSG) の皆様には大変御世話になり本当に有り難う御座いました．

最後になりましたが，いつまでも学ぶ私を暖かく見守って下さった両親に，心からの御礼を述べたいと思います．

参考文献

- [1] Aggarwal, J.K. and Cai, Q., “Human Motion Analysis: A Review”, In *Proceedings of IEEE Computer Society Workshop on Motion of Non-Ridged and Articulated Objects*, pp. 90-102 (1997).
- [2] 天谷, 原, 青木, “逆解析手法による 3 次元人体運動の再構成”, 日本機械学会論文誌 C 編, Vol. 63, No. 608, pp. 1167-1171 (1997).
- [3] Anderson, J.A. and Rosenfeld, E. (Eds.), “*Neurocomputing: Foundations or Research*”, MIT Press (1988).
- [4] Amaya, K., Bruderlin, A. and Calvert, T., “Emotion from Motion”, In *Proceedings of Graphics Interface '96*, pp. 222-229 (1996).
- [5] Blake, R., “Cats Perceive Biological Motion”, *Psychological Science*, Vol. 4, No. 1, pp. 54-57 (1993).
- [6] Bertenthal, B.I., “Infants' Perception of Biological Motions: Intrinsic Image and Knowledge-based Constraints”, In *23rd Carnegie Symposium on Cognition: Visual Perception and Cognition in Infancy*, pp. 175-214 (1993).
- [7] Bobick, A.F., and Wilson, A.D., “A State-Based Approach to the Representation and Recognition of Gesture”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 12, pp. 1325-1337 (1997).
- [8] Campbell, L. and Bobick, A., “Recognition of Human Body Motion Using Phase Space Constraints”, In *Proceedings of Fifth International Conference on Computer Vision*, pp. 624-630 (1995).
- [9] Cédras, C. and Shah, M., “A Survey of Motion Analysis from Moving Light Displays”, In *Proceedings of 1994 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 214-221 (1994).
- [10] Cédras, C. and Shah, M., “Motion-Based Recognition: A Survey”, *Image and Vision Computing*, Vol. 13, No. 2, pp. 129-155 (1995).

- [11] Darrell, T.J., Essa, I.A. and Pentland, A.P., “Task-Specific Gesture Analysis in Real-Time Using Interpolated Views”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 12, pp. 1236-1242 (1996).
- [12] Dittrich, W.H., “Action Categories and the Perception of Biological Motion”, *Perception*, Vol. 22, pp. 15-22 (1993).
- [13] Dittrich, W.H., Troscianko, T., Lea, S.E.G. and Morgan, D., “Perception of Emotion form Dynamic Point-Light Displays Represented in Dance”, *Perception*, Vol. 25, pp. 727-738 (1996).
- [14] Doya, K. and Yoshizawa, S., “Adaptive Neural Oscillator Using Continuous-time Backpropagation”, *Neural Networks*, Vol. 2, No. 5, pp. 375-385 (1989).
- [15] 銅谷, 吉沢, “時間パターンの記憶の神経回路モデル”, *電子情報通信学会論文誌 D-II*, Vol. J73-D-II, No. 8, pp. 1150-1157 (1990).
- [16] Elman, J.L., “Finding Structure in Time”, *Cognitive Science*, Vol. 14, pp. 179-211 (1990).
- [17] Edelman, S. and Bühlhoff, H.H., “Orientation Dependence in the Recognition of Familiar and Novel Views of Three-dimensional Objects”, *Vision Research*, Vol. 32, No. 12, pp. 2385-2400 (1992).
- [18] Felleman, D.J. and Van Essen D.C., “Distributed Hierarchical Processing in the Primate Cerebral Cortex”, *Cerebral Cortex*, Vol. 1, No. 1, pp. 1-47 (1991).
- [19] Fahlman, S.E., “*An Empirical Study of Learning Speed in Back-Propagation Network*”, Technical Report CMU-CS-88-162, Carnegie-Mellon University, Computer Science Dept., (1988).
- [20] Fujita, T., Suzuki, S. and Ando, H., “3D Object Recognition by Coupling Mixtures of Autoencoders and Dynamic Matching”, In *Proceedings of the*

- International Conference on Neural Information Processing (ICONIP)*, pp. 377-382 (1996).
- [21] Goddard, N., “*The Perception of Articulated Motion: Recognizing Moving Light Display*”, PhD thesis, The University of Rochester, (1992).
- [22] Hebb, D.O., “*The Organization of Behavior*”, Wiley (1949). Reprinted partially in [3].
- [23] Hogg, D., “Model-based Vision: a Program to See a Walking Person”, *Image and Vision Computing.*, Vol. 1, No. 1, pp. 5-20 (1983).
- [24] Hoffman, D.D. and Flinchbaugh, B.E., “The Interpretation of Biological Motion”, *Biological Cybernetics*, Vol. 42, pp. 195-204 (1982).
- [25] “Special Issue on Neural networks”, *IEEE Transaction on Signal Processing*, Vol. 45, No. 11 (1997).
- [26] Johansson, G., “Visual Perception of Biological Motion and a Model for Analysis”, *Perception & Psychophysics*, Vol. 14, No. 2, pp. 201-211 (1973).
- [27] Johansson, G., “Visual Motion Perception”, *Scientific American.*, pp. 76-88 (1975).
- [28] Jordan, M.I., “*Serial Order: A Parallel Distributed Processing Approach*”, Technical Report ICS-8604, University of California at San Diego, Institute for Cognitive Science., (1986).
- [29] Jordan, M.I., “Attractor Dynamics and Parallelism in a Connectionist Sequential Machine”, In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pp. 531-546 (1986).
- [30] Jordan, M. and Jacobs, R.A., “Hierarchical Mixture of Expert and the EM Algorithm”, *Neural Computation*, Vol. 6, No. 2, pp. 181-214 (1994).

- [31] Logothetis, N.K., Pauls, J. and Poggio. T., “*Spatial Reference Frames for Object Recognition Tuning for Rotations in Depth*”, A.I. Memo No. 1533, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, (1995).
- [32] McCulloch, W.S. and Pitts, W., “A Logical Calculus of the Ideas Immanent in Nervous Activity”, *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115-133 (1943). Reprinted in [3].
- [33] MacKay, D.J.C., “A Practical Bayesian Framework for Backprop Network”, *Neural Computation*, Vol. 4, No. 3, pp. 448-472 (1992).
- [34] Marr, D. and Nishihara, H.K., “Representation and Recognition of the Spatial Organization of Three-dimensional Shapes”, *Proceedings of the Royal Society of London series B*, Vol. 200, pp. 269-294 (1978).
- [35] Massone, L., Bizzi, E., “A Neural Network Model for Limb Trajectory Formation”, *Biological Cybernetics*, Vol. 61, pp. 417-425 (1989).
- [36] Mather, G., Murdoch, L., “Gender Discrimination in Biological Motion Display based on Dynamic Cues”, *Proceedings of the Royal Society of London series B*, Vol. 258, pp. 273-279 (1994).
- [37] Mather, G., Radford, K. and West, S., “Low-level Visual Processing of Biological Motion”, *Proceedings of the Royal Society of London series B*, Vol. 249, pp. 149-155 (1992).
- [38] Morgan, D.P. and Scofield, C.L. (Eds.), “*Neural Networks and Speech Processing*”, Kluwer Academic Publishers (1991).
- [39] Mozer, M.C., “Neural Net Architecture for Temporal Sequence Processing”, In [71], pp. 243-264 (1994).
- [40] Murase, H. and Nayar, S.K., “Learning Objects Models from Appearance”, In *Proceedings of AAAI-93*, pp. 836-843 (1996).

- [41] 村瀬, S. K. Nayar: “2次元照合による3次元物体認識 - パラメトリック固有空間法 -,” 電子情報通信学会論文誌 *D-II*, Vol. J77-D-II, No. 11, pp. 2179-2187 (1994).
- [42] Murase, H. and Sakai, R., “Moving Object Recognition in Eigenspace Representation: Gait Analysis and Lip Reading”, *Pattern Recognition Letters*, Vol. 17, pp. 155-162 (1996).
- [43] 西村, 岡, “ジャスチャー動画の逆時間動作や静止動作をスポッティング認識認識するための Non-monotonic 連続 DP”, 電子情報通信学会 信学技法, PRMU-96-32, pp. 49-56 (1996).
- [44] Oram, M.W. and Perrett, D.I., “Responses of Anterior Superior Temporal Polysensory (STPa) Neurons to “Biological Motion” Stimuli”, *Journal of Cognitive Neuroscience*, Vol. 6, No. 2, pp. 99-116 (1994).
- [45] O’Rourke, J. and Badler, N.I., “Model-Based Image Analysis of Human Motion Using Constraint Propagation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 6, pp. 522-536 (1980).
- [46] Pavlovic, V.I, Sharma, R. and Huang, T.S., “Visual Interpretation on Hand Gestures for Human-Computer Interaction: A Review”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 677-695 (1997).
- [47] Pearlmutter, B.A., “Gradient Calculations for Dynamic Recurrent Neural Networks : A Survey”, *IEEE Transactions on Neural Networks*, Vol. 6, No. 5, pp. 1212-1228 (1995).
- [48] Pineda, F.J., “Generalization of Backpropagation to Recurrent Neural Networks”, *Physical Review Letters*, Vol. 19, No. 59, pp. 2229-2232 (1987).
- [49] Plaut, D.C., Nowlan, S.J. and Hinton, G.E., “*Experiments on Learning by Propagation*”, Technical Report CMU-CS-86-126, Carnegie-Mellon University, Computer Science Dept., (1986).

- [50] Poggio, T. and Edelman, S., “A Network that Learns to Recognize Three-dimensional Objects”, *Nature*, Vol. 343, pp. 263-266 (1990).
- [51] Polana, R. and Nelson, R., “Recognizing Activities”, In *Proceedings of 1994 IEEE Conference on Computer Vision and Pattern Recognition*, Vol. A, pp. 815-820 (1994).
- [52] Psarrou, A., Gong, S. and Buxton, H., “Modelling Spatio-Temporal Trajectories and Face Signatures on Partially Recurrent Neural Networks”, *Proceedings of the IEEE International Conference on Neural Networks*, (1995).
- [53] Rumelhart, D.E., Hinton, G.E. and Williams, R.J., “Learning Internal Representations by Back-Bropagation Errors”, *Nature*, Vol. 323, pp. 533-536 (1986). Reprinted in [3].
- [54] Rumelhart, D. E., Hinton, G. E. and Williams, R. J., “Learning Internal Representations by Error Propagation”, In [55], Vol. 1, pp. 318-362,(1988).
- [55] Rumelhart, D.E., McClelland, J.L. and tne PDP Research Group, “*Parallel Distributed Processing, Volume1, 2*”, MIT Press (1988).
- [56] Rao, R.P.N., “*Robust Kalman Filters for Prediction, Recognition, and Learning*”, Technical Report 645, The University of Rochester, Computer Sceinece Dept., (1996).
- [57] Rohr, K., “Towards Model-Based Recognition of Human Movements in Image Sequences”, *CVGIP: IMAGE UNDERSTANDING*, Vol. 59, pp. 94-115 (1994).
- [58] Sumi, S., “Upside-down Presentation of the Johansson Moving Light-spot Pattern”, *Perception*, Vol. 13, pp. 283-286 (1984).
- [59] Suzuki, S. and Ando, H., “Unsupervised Classification of 3D Objects from 2D Views”, In Tesauro, G., Touretzky, D. and Leen, T. (Eds), *Advances in Neural Information Processing Systems 7*, pp. 949-956, MIT Press (1995).

- [60] 高橋, 関, 小島, 岡, “ジェスチャー動画像のスポットニング認識”, 電子情報通信学会論文誌 *D-II*, Vol. J77-D-II, No. 8, pp. 1552-1561 (1994).
- [61] Turk, M.A. and Pentland, A.P., “Face Recognition Using Eigenfaces,” In *Proceedings of 1991 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591 (1991).
- [62] Ulman, S. and Basri, R., “Recognition by Linear Combinations of Models”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 10, pp. 992-1006 (1991).
- [63] Underleider, L.G. and Mishikin, M., “Two Cortical Visual Systems”, In Ingle, D.J., Goodale, M.A. and Mansfield, R.J.W. (Eds), *Analysis of Visual Behavior*, pp. 549-586, MIT Press (1982).
- [64] Verfaillie, K., “Orientation-Dependent Priming Effects in the Perception of Biological Motion”, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 19, No. 5, pp. 992-1013 (1993).
- [65] Yamato, J., Ohaya, J. and Ishii, K., “Recognizing Human Action in Time-Sequential Images using Hidden Markov Model”, In *Proceedings of 1992 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 379-385 (1992).
- [66] 大和, 大谷, 石井, “隠れマルコフモデルを用いた動画像からの人物の行動認識”, 電子情報通信学会論文誌 *D-II*, Vol. J76-D-II, No. 12, pp. 2556 – 2563 (1993).
- [67] Yang, M.P., “Objective Analysis of the Topological Organization of the Primate Cortical Vision System”, *Nature*, Vol. 358, pp. 152-155 (1992).
- [68] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K. and Lang, K.J., “Phoneme Recognition Using Time Delay Neural Networks”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 37, No. 3, pp. 328-339 (1989).

- [69] Wang, D., Liu, X. and Ahalt, S.C., “On Temporal Generalization of Simple Recurrent Networks”, *Neural Networks*, Vol. 9, No. 7, pp. 1099-1118 (1996).
- [70] 渡辺, 郷原, 内川, “リカレントニューラルネットワークの各学習則に関する検討および学習曲面の形状”, 電子情報通信学会論文誌 *D-II*, Vol. J74-D-II, No. 12, pp. 1776-1787 (1991).
- [71] Weigend, A.S. and Gershenfeld, N.A. (Eds.), “*Time Series Prediction: Forecasting the Future and Understanding the Past*”, Addison-Wesley Publishing Company (1994).
- [72] Werbos, P.J., “Backpropagation Throught Time: What it Does and How to Do it”, *Proceedings of the IEEE*, Vol. 78, No. 10, pp. 1550-1560 (1990).
- [73] Williams, R.J. and Peng, J., “An Efficient Gradient-based Algorithm for On-line Training of Recurrent Network Trajectories”, *Neural Computation*, Vol. 2, No. 3, pp. 490-501 (1990).
- [74] Williams, R.J. and Zipser, D., “A Learning Algorithm for Continually Running Fully Recurrent Networks”, *Neural Computation*, Vol. 1, No. 3, pp. 270-280 (1989).

付録

A. 単振り子・二重振り子の軌道算出

単振り子・二重振り子どちらも座標 (0.5,1.0) を振り子の支点として
単振り子：

$$\begin{cases} x(t) = l \sin(\theta(t)) + 0.5 \\ y(t) = -l \cos(\theta(t)) + 1.0 \end{cases}$$

ただし

$$\begin{aligned} \theta(t) &= \sin(\sqrt{\frac{g}{l}}t), 0 \leq t < 2\pi\sqrt{\frac{g}{l}} \\ l &= 0.5, g = 9.8 \end{aligned}$$

二重振り子：

$$\begin{cases} x_1(t) = l \sin(\theta_1) + 0.5 \\ y_1(t) = l \cos(\theta_1) + 1.0 \\ x_2(t) = l \sin(\theta_2) + x_1(t) \\ y_2(t) = l \cos(\theta_2) + y_1(t) \end{cases}$$

ただし

$$\begin{aligned} \theta_1(t) &= \frac{1}{2\sqrt{2}} \{A \cos(\omega_1 t) + B \cos(\omega_2 t)\} \\ \theta_2(t) &= \frac{1}{2} \{A \cos(\omega_1 t) - B \cos(\omega_2 t)\} \\ \omega_1 &= \sqrt{(2 - \sqrt{2})\frac{g}{l}}, \omega_2 = \sqrt{(2 + \sqrt{2})\frac{g}{l}} \\ 0 &\leq t < 2\pi\sqrt{\frac{g}{l}} \\ l &= 0.4, g = 9.8, A = 1.0, B = 0.0 \end{aligned}$$