

修士学位論文

時系列を学習する  
確率的単純再帰ネットワークモデル  
に関する研究

東北大学大学院工学研究科  
電気・通信工学専攻  
竹谷 仁志

# 目次

第1章	序論	1
1.1	はじめに	1
1.2	目的	2
1.3	本論文の構成	3
第2章	単純再帰ネットワーク	4
2.1	はじめに	4
2.2	準備	4
2.2.1	しきい値素子型モデル	5
2.2.2	バックプロパゲーション法	6
2.3	単純再帰ネットワーク	7
2.4	実験	9
2.4.1	問題設定	10
2.4.2	評価方法	11
2.4.3	結果	12
2.5	まとめ	12
第3章	系列の区切りを考慮した 単純再帰ネットワークモデル	13
3.1	はじめに	13
3.2	Elmanの言語獲得過程のモデル	13
3.2.1	問題設定	14
3.2.2	評価方法	14
3.3	系列の区切りを考慮したモデル	16
3.4	実験	18
3.4.1	実験条件	18
3.4.2	評価方法	19
3.4.3	結果と考察	19
3.4.4	COSINE	20
3.4.5	文脈層の評価	22
3.5	未解決の問題点	24
3.5.1	ダイバージェンスによる評価	24

3.6	まとめ	25
<b>第4章 時系列を学習する</b>		
	確率的単純再帰ネットワークモデル	26
4.1	はじめに	26
4.2	単純再帰ネットワークの学習方法についての検討	26
4.2.1	3層パーセプトロンによる検討	27
4.2.2	隠れ層パターンの類似による学習能力の低下	28
4.2.3	クラスター分析による評価	29
4.3	文脈層パターンの平均化を防ぐ新学習法の提案	30
4.3.1	文脈層パターンを利用した新学習法	31
4.3.2	実験による評価	32
4.4	既存の時系列学習方法との比較	35
4.4.1	慣性項を考慮したバックプロパゲーション法	35
4.4.2	実験による評価	35
4.5	提案学習法への慣性項の追加	37
4.5.1	慣性項の追加	37
4.5.2	実験による評価	37
4.6	追加実験	40
4.6.1	どのような問題を扱うか	40
4.6.2	実験	41
4.6.3	考察	43
4.6.4	残された課題	43
4.7	まとめ	44
<b>第5章 結論</b>		
5.1	本論文の成果	45
5.2	課題	45
謝辞		47
参考文献		48
付録 A		50
A.1	評価値 (COSINE, ダイバージェンス)	50
A.1.1	主語	51
A.1.2	目的語	52
A.1.3	動詞	54

# 目次

2.1	ニューロン	5
2.2	しきい値素子	5
2.3	時系列の取り扱い	8
2.4	単純再帰ネットワーク	8
2.5	次の入力値の予測	11
2.6	使用するネットワーク	11
2.7	$E(t)$ の時間変化	12
3.1	1991 年モデルの系列の生成規則	17
3.2	Elman のモデル	18
3.3	区切りを考慮したモデル	18
3.4	period なし	18
3.5	period あり	18
3.6	COSINE の推移	20
3.7	book_S	23
3.8	book_O	23
3.9	period と book_O	23
3.10	verb_tran	23
3.11	ダイバージェンス	25
3.12	停滞がおこる単語	25
4.1	3 層パーセプトロンの構成	27
4.2	2 層パーセプトロンの構成	28
4.3	クラスター分析	30
4.4	period なし	31
4.5	period あり	31
4.6	クラスター分析 (提案学習法)	33
4.7	COSINE(主語 boy)	34
4.8	D(主語 boy)	34
4.9	COSINE(目的語 boy)	34
4.10	D(目的語 boy)	34
4.11	COSINE(chase)	34

4.12	D(chase)	34
4.13	COSINE(主語 boy)	36
4.14	D(主語 boy)	36
4.15	COSINE(目的語 boy)	36
4.16	D(目的語 boy)	36
4.17	COSINE(chase)	36
4.18	D(chase)	36
4.19	COSINE(主語 boy)	38
4.20	D(主語 boy)	38
4.21	COSINE(目的語 boy)	38
4.22	D(目的語 boy)	38
4.23	COSINE(chase)	38
4.24	D(chase)	38
4.25	慣性項のみの学習法	39
4.26	慣性項を追加した提案学習法	39
4.27	中央はめ込み文の一例	40
4.28	単純再帰ネットワークの構成	41
4.29	深さ 3 の場合の入力系列	41
4.30	隠れ層の状態 (深さ 4)	42
4.31	隠れ層の状態 (深さ 9)	42
4.32	深さ 11 の場合における予測精度の推移	44
A.1	COSINE(主語 noun_human)	51
A.2	D(主語 noun_human)	51
A.3	COSINE(主語 noun_animal)	51
A.4	D(主語 noun_animal)	51
A.5	COSINE(主語 noun_inanimal)	51
A.6	D(主語 noun_inanimal)	51
A.7	COSINE(主語 noun_agpat)	52
A.8	D(主語 noun_agpat)	52
A.9	COSINE(目的語 noun_human)	52
A.10	D(目的語 noun_human)	52
A.11	COSINE(目的語 noun_animal)	52
A.12	D(目的語 noun_animal)	52
A.13	COSINE(目的語 noun_inanimal)	53
A.14	D(目的語 noun_inanimal)	53
A.15	COSINE(目的語 noun_fragile)	53
A.16	D(目的語 noun_fragile)	53
A.17	COSINE(目的語 noun_food)	53

---

A.18 D(目的語 noun_food)	53
A.19 COSINE(1)	54
A.20 D(1)	54
A.21 COSINE(2)	55
A.22 D(2)	55
A.23 COSINE(3)	55
A.24 D(3)	55
A.25 COSINE(4)	55
A.26 D(4)	55
A.27 COSINE(5)	56
A.28 D(5)	56
A.29 COSINE(6)	56
A.30 D(6)	56
A.31 COSINE(7)	56
A.32 D(7)	56
A.33 COSINE(8)	57
A.34 D(8)	57
A.35 COSINE(9)	57
A.36 D(9)	57
A.37 COSINE(10)	57
A.38 D(10)	57
A.39 COSINE(11)	58
A.40 D(11)	58
A.41 COSINE(12)	58
A.42 D(12)	58

# 第1章

## 序論

### 1.1 はじめに

現在人類は、人類が創り出した高度な科学技術によって自身の機能を拡張している。しかし、拡張した機能によって何をして良いか、何をしてはいけないか、また何をすべきかということについての議論は十分であるとは言えない。金銭的に恵まれ、安全で生存すること自体に非常な困難を伴わなくなった社会では、どう人間らしく生きるかにその目的が移るものと考えられる。その目的を実現するためには異なる思想や宗教、民族の人々が相互にその違いを認めて良好な関係を持つことが必要であり、そのためにもその議論は必要なことである。それにも関わらずその議論が進まないのは、人自体についての理解が十分ではないためと考えられる。人がどのような存在であるかが客観的に明らかになっていなければ、人がどのような存在であるべきかについての議論を収束に向かわせることは困難である。勿論人がどのような存在であるかについての解釈は、各時代に生きる人達によって与えられるべきものである。しかし、科学によって論拠となる客観的な事実を用意することは可能であり、それは客観的であるが故に時代によって変化しない意義のあるものである。

ここで言う人自体の理解とは人の知性を支える高度な認知機能についての理解である。人は肉体にある感覚器官によって得た外界からの情報と肉体の内部に保持された情報を基に、脳でなんらかの情報処理を行うことでその認知機能を実現しているものと考えられている。従って脳の情報処理のメカニズムの解明を一つの目標として据えられる。現在脳の解明を目的として、医学、生理学、心理学などの多くの学問分野の研究者達が所属する分野の枠を越えて研究を行っている。本論文の著者は脳の情報処理のメカニズムの解明を大きな目標とし、他の研究者の研究結果の補間や研究手法の改善といった形での貢献をすることとした。本研究では、最初に述べた問題意識をもとに他の脳機能の解明を目的とした研究との関係を持つことを目指す。

## 1.2 目的

人の優れた脳機能の一つとして、言語機能が挙げられる。現在、言語機能の獲得（言語獲得）のメカニズムはいまだに明らかになっておらず、言語学、心理学、生理学、工学などの複数の学問分野をまたいで活発に研究が行われている。

言語獲得は言語機能に必要な知識（言語知識）の獲得であると考えられる。その知識がそもそもどのようなものなのかということについては大きく分けて三つの分野から研究されている。統語論、意味論、音韻論である。統語論における知識、統語的知識の一つとして文法規則が挙げられる。文法規則には人称、数、格、性、時制などについての規則の他に単語がどのように並ぶのかの語順についての規則がある。

Elman はニューラルネットワークの一種である単純再帰ネットワークを使用してモデルを構築し、計算機実験を通して予測によって語順についての規則が獲得され得ることを示した [1]。従来の言語獲得における議論では、言語知識を生まれつき持つと考える生得性を強く主張する立場と生まれつきの知識ではなく、環境からの学習によって知識を獲得すると考える経験を重視する立場に分かれていた。しかし近年様々な学問分野をまたいで研究されるようになるにつれてこの状況は変わってきた。工学的手法としてのニューラルネットワークを使用したモデルで研究するコネクショニストと呼ばれる研究者は両者の立場の中間の立場をとるようになってきた。その立場ではネットワークの構造や学習規則、与えられる入力を生得的制約と考え、その制約の元に学習することによって経験的に言語獲得がなされると考えるのである。

Elman の研究は、フィードバックの付加と予測するという性質を生得的制約、単語系列を入力とした学習を経験として言語知識の獲得可能性を示したという点でコネクショニストの立場からの結果を先駆的に示した価値あるものと言える。現在でも Elman の研究結果をもとにした研究は行われている。例えば、語順だけでなく格関係も同時に獲得可能であることを示した玉森、乾の研究 [9] や、PATON に単純再帰ネットワークを付加することで意味と語順を考慮した言語モデルを提案した大森らの研究 [6],[7] などがある。

しかし Elman による研究にも幾つか考慮すべき点があることがわかってきた。一つはモデルとして不自然と考えられる点、もう一つはモデルに使用している単純再帰ネットワークの計算能力における問題点である。前者については Elman の後の研究 [2],[3] で使用しているモデルでは改善されているが、その理由については論文中で触れられてはならず、問題点の一つと考えられる。後者については以前からその要因が研究されてきており、森田らは単純再帰ネットワークの問題点を非直交性の問題、文脈の平均化の問題に分けて説明している [4],[5]。この問題は当然のことであるが学習後のモデルの評価にも関わってくるため、改善が必要である。

そこで本研究ではこれらの二つの問題点について扱うことにする。一つ目の問題点についてはその理由の説明を目的とした評価実験を行い、二つ目の問題点については単純再帰ネットワークのための新学習法を提案し、改善を図った。



## 1.3 本論文の構成

本論文の構成は以下の通りである.

### 第 1 章 序論

本研究の背景, 目的, 及び論文の構成を述べる.

### 第 2 章 エルマンネット

本研究で用いる単純再帰ネットワークに関する基礎知識について説明する.

### 第 3 章 系列の区切りを考慮した単純再帰ネットワークモデル

Elman(1990) のモデルに系列の区切りの情報を加え, 従来のモデルと区切りを考慮したモデルとで質的な違いが見られるかを数値シミュレーションを通して検証する.

### 第 4 章 確率的単純再帰ネットワーク

種々の検討から単純再帰ネットワークの学習性能に問題があることがわかってきた. 本章ではまずその問題点について述べ, 問題解決を目的として考案した新学習法について説明する. その後提案学習法の効果の検証のために数値シミュレーションを行い, 得られた結果について評価, 検討する.

### 第 5 章 結論

本研究の成果についてまとめ, 今後の課題について述べる.

## 第2章

# 単純再帰ネットワーク

### 2.1 はじめに

本章では、本研究で用いる単純再帰ネットワークについて説明する。まず、単純再帰ネットワークについて説明するための基礎知識として、人間の脳の構造とその情報処理の仕組みについて簡単に触れる。その後、脳の数学的モデルであるしきい値回路について説明する。そして、ニューラルネットワークモデルの学習則の一つであるバックプロパゲーション法について説明する。最後に、これらの基礎知識をもとにして、単純再帰ネットワークについて説明する。

### 2.2 準備

人間の脳は、数百億ともいわれるニューロン（神経細胞）が結合することにより構成されている。ニューロンは図 2.1 に示すように、細胞体、樹状突起、軸索、シナプスの 4 つの部分から構成される。細胞体はニューロンの本体であり、細胞核を持っている。ニューロンの入力端は樹状突起であり、ニューロンはシナプスと呼ばれる接合部を介して他のニューロンからの信号を受け取る。シナプスでの信号伝達は、神経伝達物質により化学的に行われる。シナプスの結合の強さはそれぞれ異なっている。ニューロンは、他の多数のニューロンから入力を加え合わせ、この和（刺激の総和）がある一定の値を越えると発火する。発火したニューロンは発火状態にあり、発火していないニューロンは静止状態にある。発火したニューロンはパルス電位を発生し、このパルスは軸索上を伝搬し、シナプスを介して次のニューロンに伝わる。

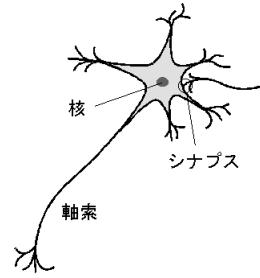


図 2.1: ニューロン

また, 脳の情報処理の特徴として,

- 1 ニューロンとニューロン間の相互作用が基本となっている.
- 2 並列分散処理を行う, 大規模階層システムである.
- 3 学習能力, 自己組織能力を持っている.

がある. これらの特徴を満たすモデルとして, しきい値回路が挙げられる.

### 2.2.1 しきい値素子型モデル

ニューロンの動作を模倣するモデルとして, 1943年に, W. S. McCulloch と W. S. Pitts は, しきい値素子 (図 2.2) を提案した. しきい値素子とは, しきい値関数を計算するブール素子である. こ

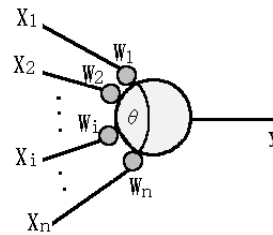


図 2.2: しきい値素子

こで,  $X = (x_1, \dots, x_n) \in \{0, 1\}^n$  としたとき, しきい値関数  $f(X) : \{0, 1\}^n \rightarrow \{0, 1\}$  は, 以下に示すブール関数である.

$$f(X) = \begin{cases} 1 & (\sum_{i=1}^n w_i \cdot x_i \geq \theta \text{ のとき}) \\ 0 & (\sum_{i=1}^n w_i \cdot x_i < \theta \text{ のとき}) \end{cases} \quad (2.1)$$

式 (2.1) において,  $w_i (i = 1, 2, \dots, n)$  は結合強度と呼ばれる実数であり,  $\theta$  はしきい値と呼ばれる実数である. 従って, 式 (2.1) は, あるニューロンに伝わる刺激の総和がしきい値以上であれば, そのニューロンが発火することを表している. その意味では, 重み  $w_i$  はシナプスの結合の強さに対応している. また,  $w_i > 0$  は興奮性のシナプスを表し,  $w_i < 0$  は抑制性のシナプスを表している.

しきい値型モデルは、しきい値素子が結合することで構成される。このことは、人間の脳がニューロンを結合することで構成されることに対応している。

しきい値型モデルにおける学習は、辺の重みが増加することによって表現される。このことは、人間の脳の学習が脳の可塑性によること、すなわちシナプス結合の強さが変化することによって行われることに対応している。

しきい値回路は、ニューロンの発火、静止の二つの状態をそれぞれしきい値論理素子の 1, 0 の値に対応させたモデルであると言える。しかし、このモデルでは式 (2.1) の階段関数のような不連続関数を用いているため、解析的に問題を解くことが出来ない。解析的な取扱いをすることが出来るモデルとしてアナログしきい値素子モデルが挙げられる。このモデルでは、ある時刻  $t$  でのニューロンへの入力信号  $x_i(t)$  を、 $i$  番目のシナプスの、時刻  $t$  付近での興奮伝達の平均ととらえる。そして、最高頻度の出力を 1, 最低頻度の出力を 0 と規格化できると考えて、ニューロンの出力を  $z$  を  $0 \leq z \leq 1$  の連続値とする。入出力関係は次式で表される。

$$z = f\left(\sum_i w_i x_i(t) - h\right) \quad (2.2)$$

関数  $f(v)$  には次式で表されるものが使われる。

$$f(v) = \frac{1}{1 + e^{-v}} \quad (2.3)$$

この関数を用いるのは、 $f(v)$  の  $v$  についての微分の結果が  $\frac{df}{dv} = f(v)(1 - f(v))$  と簡単な形で表せるからである。

### 2.2.2 バックプロパゲーション法

本節では、ネットワークの学習則 (結合強度の更新) の一つである、バックプロパゲーション法 (以下 BP 法) について説明をする。まず、以下の多層フィードフォワードネットワークを考える。 $m$  層のネットワークを考え、 $k$  層の  $i$  番目のノードへの総入力を  $x_i^k$ 、このノードの出力を  $y_i^k$ 、 $k-1$  層の  $j$  番目のノードから  $k$  層の  $i$  番目のユニットへの結合強度を  $w_{ij}^{k-1,k}$  と表す。各ノードの出力は、

$$y_i^k = f(x_i^k) = \frac{1}{1 + e^{-x_i^k}} \quad (2.4)$$

$$x_i^k = \sum_j w_{ij}^{k-1,k} y_j^{k-1} \quad (2.5)$$

と表される。

入力  $x$  に対する教師信号  $t$  が与えられた場合、BP 法では、次式のように誤差関数  $E$  を定義する。

$$E = \frac{1}{2} \sum_i (y_i^m - t_i)^2 \quad (2.6)$$

この  $E$  の値が最小となるように、結合強度の更新量を次式を用いて算出する。

$$\Delta w_{ij}^{k-1,k} = -\epsilon \frac{\partial E}{\partial w_{ij}^{k-1,k}} \quad (2.7)$$

まず,  $k = m$  の場合の  $w_{ij}^{m-1,m}$  の結合強度の更新量を次式によって算出する.

$$\begin{aligned}\Delta w_{ij}^{m-1,m} &= -\epsilon \frac{\partial E}{\partial w_{ij}^{m-1,m}} \\ &= -\epsilon \frac{\partial E}{\partial y_j^m} \frac{\partial y_j^m}{\partial x_j^m} \frac{\partial x_j^m}{\partial w_{ij}^{m-1,m}} \\ &= -\epsilon (y_j^m - t_j) y_j^m (1 - y_j^m) y_i^{m-1} \\ &= -\epsilon \delta_j^m y_i^{m-1}\end{aligned}\tag{2.8}$$

ただし,  $\delta_j^m = (y_j^m - t_j) y_j^m (1 - y_j^m)$  とする.

次に,  $k = m - 1$  以下の  $k = n$  の場合の更新には次式を誤差信号  $\delta$  として用いる.

$$\begin{aligned}\frac{\partial E}{\partial y_j^n} &= \sum_k \frac{\partial E}{\partial y_k^{n+1}} \frac{\partial y_k^{n+1}}{\partial x_k^{n+1}} \frac{\partial x_k^{n+1}}{\partial y_j^n} \\ &= \sum_k \frac{\partial E}{\partial y_k^{n+1}} \frac{\partial y_k^{n+1}}{\partial x_k^{n+1}} \frac{\partial}{\partial y_j^n} \sum_i w_{ik}^{n,n+1} y_i^n \\ &= \sum_k \delta_k^{n+1} w_{kj}^{n,n+1}\end{aligned}\tag{2.9}$$

このようにして, 結合強度の更新を出力層から入力層に向かって再帰的に行う.

まとめると, 結合強度の修正量  $\Delta w_{ij}^{k-1,k}$  は以下のようにして算出される.

$$\Delta w_{ij}^{k-1,k} = -\epsilon \delta_j^k y_i^{k-1}\tag{2.10}$$

ただし,

$$\delta_j^k = \begin{cases} (y_j^m - t_j) y_j^m (1 - y_j^m) & k = m \text{ のとき} \\ \sum_l \delta_l^{k+1} w_{lj}^{k,k+1} & k \neq m \text{ のとき} \end{cases}\tag{2.11}$$

## 2.3 単純再帰ネットワーク

人間の言語処理や運動などの高次の脳処理をモデル化して検討する際に, 時間の概念が重要となってくる. 言語処理を例にすると, 人は単語の系列である文を聞いて理解することが出来る. 理解には単語がどのような順番で並んでいたかという文脈の情報を用いていると考えられる. この場合, ある時刻にある単語を聞くと考えると単語の系列は時系列データであると言える. 従来のフィードフォワードネットワークでは入力層の各ノードを各時刻に対応させることで時間を表現する. しかし, この表現方法ではノードの数で, 扱える時刻の幅が決ってしまう. また, この方法は扱う時刻の幅の情報を一時的に保存しておくシフトレジスタの存在を仮定している (図 2.3). シフトレジスタの役割を担う機能が人間の脳にあるという事実は確認されていない. これらの点からフィードフォワードネットワークによる時間の表現は適切ではないと考えられる. シフトレジスタのような特定の機能の存在を仮定せずに時系列データを扱うために, J. Elman は単純再帰ネットワークを提案した (図 2.4). このネットワークは提案者の名前を取って, Elman ネットとも呼ばれる. このネットワークは隠れ層にフィードバックを持ち, 入力されたデータがどのように並んでいたかという文脈の情報を考慮した出力を返すことが出来るようになっている.

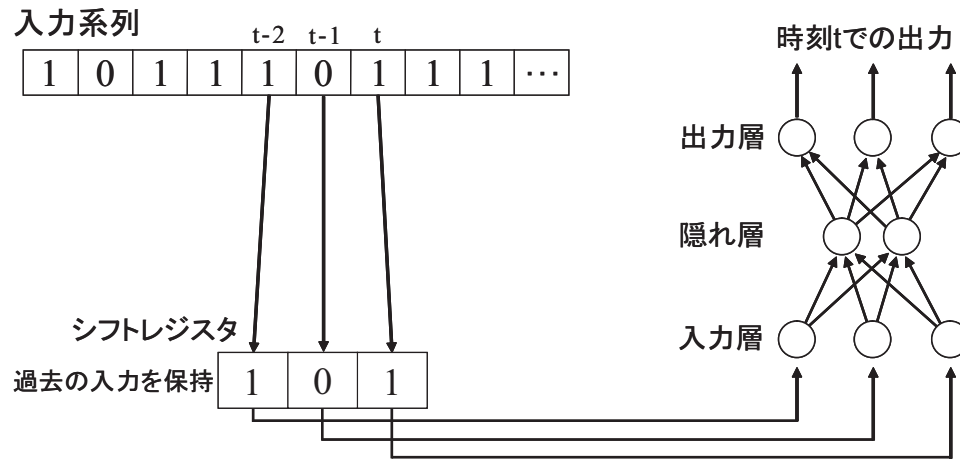


図 2.3: 時系列の取り扱い

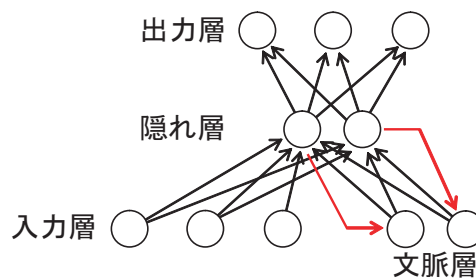


図 2.4: 単純再帰ネットワーク

次に、単純再帰ネットワークの具体的な構成について説明する。単純再帰ネットワークは、入力層、隠れ層、出力層、文脈層からなり、通常の三層パーセプトロンに、隠れ層の各ノードの値をフィードバックして一時的に保存する層（文脈層）を付加した構造となっている。文脈層に保存された隠れ層の各ノードの値は次の時点での入力となる。このフィードバックの結合は1に固定されている。そのため、学習時にはフィードフォワードネットワークにおけるBP法を用いることができる。このネットワークでは入力層から出力層への1回の伝搬を時間の単位と考える。入力系列としては時系列データを用意し、扱う時系列データの最初のデータの時点をも  $t = 0$  とする。

このことを数式を用いて説明をする。時点  $t$  での入力層、隠れ層、出力層、文脈層の各ノードの値をそれぞれ、 $x_i(t)$ ,  $y_j(t)$ ,  $z_k(t)$ ,  $c_l(t)$  とする。入力層のノード  $i$  から隠れ層のノード  $j$  への結合強度を  $w_{ij}$ 、文脈層のノード  $l$  から隠れ層のノード  $j$  への結合強度を  $w_{lj}$ 、隠れ層のノード  $j$  から出力層のノード  $k$  への結合強度を  $w_{jk}$  とする。これらを用いて入力から出力までの関係は次のように表される。

$$y_j(t) = \phi\left(\sum_i w_{ij}x_i(t) + \sum_l w_{lj}c_l(t) - \theta_j\right) \quad (2.12)$$

$$z_j(t) = \phi\left(\sum_j w_{jk}y_j(t) - \theta_j\right) \quad (2.13)$$

$$\phi(v) = \frac{1}{1 + e^{-v}} \quad (2.14)$$

出力層の各ノードが0から1までの連続値を出力する。隠れ層と文脈層のノードの数は等しく、隠れ層の各ノードの値が文脈層の各ノードに保存される。保存された値は次の時点での文脈層から隠れ層への伝搬の際に用いられる。式で表すと次のようになる。

$$c_l(t+1) = y_l(t) \quad (2.15)$$

学習にはBP法を用いる。教師信号を  $d_k(t)$  とすると、誤差関数は次式で表される。

$$E = \sum_k (z_k(t) - d_k(t))^2 \quad (2.16)$$

この誤差関数を最小化するように、次式によって結合強度を更新する。

$$\begin{aligned} \Delta w_{jk} &= -\epsilon \delta_k y_j \\ \Delta w_{ij} &= -\epsilon \delta_j x_i \\ \Delta w_{lj} &= -\epsilon \delta_j c_l \end{aligned} \quad (2.17)$$

$\delta_k, \delta_j$  は次式によって算出する。

$$\begin{aligned} \delta_k &= (z_k - d_k) z_k (1 - z_k) \\ \delta_j &= \sum_k \delta_k w_{jk} \end{aligned} \quad (2.18)$$

次に、文脈層の挿入による効果について説明する。ある時点  $t$  での文脈層の各ノードの値に注目すると、次式のように表すことができる。

$$\begin{aligned} y(t) &= x(t) + w y(t-1) \\ &= x(t) + w x(t-1) + w y(t-1) \\ &\vdots \\ &= \sum_t w^t x(t) \end{aligned} \quad (2.19)$$

このように、時点  $t$  での文脈層の各ノードは過去の入力の全履歴に依存した値を出力することが分かる。この性質によって、単純再帰ネットワークは入力系列の文脈を反映した出力が出来るものと考えられる。

## 2.4 実験

前節で述べた単純再帰ネットワークに簡単な系列データを学習させ、文脈を反映した出力が可能かどうかを確認する。具体的には elman の行った XOR(排他的論理和) 問題の学習について追実験する。

### 2.4.1 問題設定

本節では、問題設定について説明する。

#### 入力系列

扱う時系列は、2つの論理変数に対する XOR からなるものとする。真理値表は表 3.1 のようになる。表 3.1 の、四つのパターンを並べたものを入力系列とする。具体的には、000, 011 というよう

表 2.1: 真理値表

	$p$	$q$	$p \oplus q$
パターン 1	0	0	0
パターン 2	0	1	1
パターン 3	1	0	1
パターン 4	1	1	0

に、最初の 2bit を二つの論理変数の値とし、残りを XOR の値として一つのパターンを表す。四つのパターンをランダムに選び、000011..., というように並べた系列をネットワークへの入力とする。入力系列の最初の bit を時刻  $t = 0$  での入力、その次の bit を時刻  $t = 1$  での入力というように考えて、入力系列を時系列データとして扱う。

#### ネットワークに与える課題

ネットワークに与える課題は、各時点で、入力される値に対してその次の時点で入力される値を予測させることとする。教師信号は次の時点での入力の値ということになる。ネットワークの学習には、BP 法による教師あり学習を用いる。本実験では、入力系列の大きさを 3000bit、学習回数を 100 回とした。

#### 扱う問題の性質

この実験において、まず、入力系列は 3bit ごとに 4 つの XOR のパターンがランダムに並べられたものである。また、各パターンにおいて 3bit のうちの最初の 2bit から最後の bit の値が決まる (XOR)。そのため最初の 1bit から次の 1bit を予測することは出来ない。さらに、最後の 1bit の値から次のパターンの最初の 1bit の値を予測することも出来ない (図 2.5)。そのため、学習が進んでも予測不可能な bit におけるネットワークの出力と教師との誤差は収束しないものと考えられる。そこでこの実験の評価方法として各時点での誤差を用いる。予測不可能な bit での誤差に比べ、予測可能な bit での誤差が小さい値となればネットワークが望ましい学習を行っていると言える。



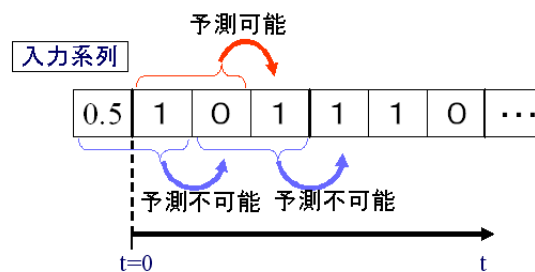


図 2.5: 次の入力値の予測

### 単純再帰ネットワークの構成

使用する単純再帰ネットワークの構成を説明する. 入力ノードの数は 1 つ, 隠れ層, 文脈層のノードの数はそれぞれ 2 つ, 出力ノードの数は 1 つとする (図 2.6). ネットワークの結合の強さの初期値は,  $-0.1$  から  $0.1$  の値でランダムに決める. ただし, 前章でも述べたように, 隠れ層から文脈層の結合の強さは 1 とする. 下図に使用するネットワークを示す.

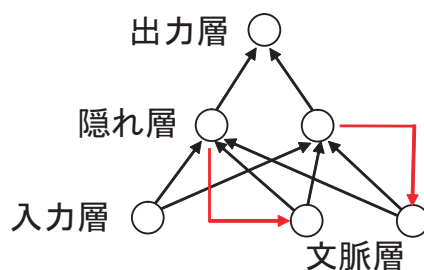


図 2.6: 使用するネットワーク

### 2.4.2 評価方法

学習後のネットワークの評価方法について説明する. まず, 学習用の系列とは別に, 3000bit の系列を 100 個用意して, 100 個の系列それぞれについて, 各時点ごとに, 教師  $d(t)$  とネットワークの出力  $z(t)$  との誤差  $e(t) = d(t) - z(t)$  を算出する. そして, その誤差の平均  $E(t)$  を次のように計算する.

$$E(t) = \frac{e_1(t) + e_2(t) + \cdots + e_{100}(t)}{100} \quad (2.20)$$

この  $E(t)$  の値により, 単純再帰ネットワークが入力系列の 3bit ごとに XOR のパターンが並んでいるという構造を予測学習によって獲得できているかを評価する.

### 2.4.3 結果

$t = 2050$  から  $t = 2064$  の範囲での  $E(t)$  を, 図 2.7 に示す. 図から, 3bit ごとに誤差が小さくなっ

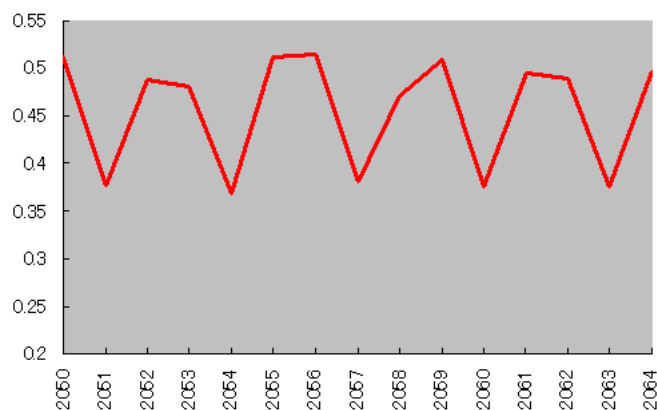


図 2.7:  $E(t)$  の時間変化

ていることがわかる. 誤差が小さくなっている時点  $t_a$  は予測可能な bit と対応しており, 単純再帰ネットワークが  $t_a - 1$  での入力  $I_{t_a-1}$  と  $t_a$  での入力  $I_{t_a}$  との XOR を出力するように学習していると言える. この結果から, 単純再帰ネットワークが予測学習を通して入力系列の持つ特徴を獲得可能である事が確認された.

## 2.5 まとめ

本章では本論文に必要とされる基礎的な知識について説明し, 本研究で使用する単純再帰ネットワークの時系列データの学習能力を確認するため, 簡単な実験を行った. 実験により, 単純再帰ネットワークが予測学習により系列の持つ構造に応じた動作をするようになることが確認された. 次章では単純再帰ネットワークを使用して言語知識の学習による獲得可能性を示した Elman のモデルについて, Elman のモデルを紹介し, 説明不足である点について述べ, 不足した説明を補うことを目的として評価を行っていく.

## 第3章

# 系列の区切りを考慮した 単純再帰ネットワークモデル

### 3.1 はじめに

Elman(1990)は前章で説明した単純再帰ネットワークを用いて言語獲得の過程をモデル化し、実験により環境から得られるデータ(文)を利用した学習による言語知識の獲得可能性を示した[1]. Elmanは後に関係詞を含むより複雑な文をデータとして加えても学習可能であることを示し[2], さらに、文脈層のノードの数を学習回数の増加に応じて増やすなどいくつかの条件を加えることで学習結果が改善されるという結果を示し、乳児の持つ生得的制約と関連付けて考察する[3]など1990年のモデルを元として研究を進めてきた。しかし、1990年の最初のモデルにはモデルとして不自然と考えられる点がある。ネットワークに与えられる情報に系列の区切りとなる情報が与えられていないという点である。1991年の研究ではこの点が改善されているものの、改善に至った経緯等については説明されていない。

そこで本章では1990年のElmanのモデルを元に、系列の区切りの有無が学習結果にどのような影響を及ぼすのか、学習後のネットワークに質的な違いが生じるかを調査することを目的とする。まず始めにElmanのモデルについて説明し、次に系列の区切りを追加したモデルについて説明する。その後に両モデルについて計算機シミュレーションを行い、その結果について評価、検討する。

### 3.2 Elmanの言語獲得過程のモデル

本節ではElmanの用いた言語獲得過程のモデルについて説明する。

Elmanのモデルでは入力系列として2語または3語からなる簡単な文を多数連結した単語系列を単語ごとにネットワークに与える。学習課題は前章のXOR課題と同様次の時点での入力の予測である。このモデルでは単純再帰ネットワークを乳児(学習者)、入力系列を乳児の周囲でなされる

会話, 学習課題を乳児の予測を行うという性質と考える. まず具体的な問題設定について説明し, その後に系列の区切りの情報がネットワークに与えられていないことがどのような意味を持つかを説明する.

### 3.2.1 問題設定

ネットワークに与える入力系列は文からなる単語系列とする. 単語は 29 種類用意し, それぞれ 29bit で符号化する (表 3.1). 各単語は予め 12 個のカテゴリに分けられている (表 3.2). 単純再帰ネットワークに与えられるのは表 3.1 の符号化された 29bit の値でしかないため, カテゴリの情報は明示的には単純再帰ネットワークに与えられない. 文の作成には 16 種類のテンプレートを用いる (表 3.3). 生成する文は, SV, SVO の簡単な文法に基づいており, 1 つの文の長さは, 2 語, あるいは 3 語となる. 時制などによる動詞の語尾変化や冠詞の付加は考えない. 表 3.3 に示されたテンプレートの 1 つをランダムに選択し, 次に WORD1, WORD2, WORD3 のそれぞれについて該当するカテゴリに属する単語をランダムに選択して文を作成する. 次に, ネットワークに与える課題と学習方法について説明する. ネットワークに与える課題は入力される単語に対して次の時点で入力される単語の予測をすることとする. 前章と同じように教師は次の時点での入力値とする. 学習には BP 法を用いた教師あり学習を用いる.

### 3.2.2 評価方法

次に, 学習後のネットワークの評価方法について述べる. Elman のモデルで扱う入力系列は先に述べた様に複数の文を繋げて作成した単語系列である. 本論文で扱う文は, ある単語の次に続く単語は一意に決まらない. 例えば, 主語の boy に続く単語としては, eat, like, move などがあり, boy の後にこれらのどの単語が続くかは予測することが出来ない. 学習後のネットワークは, ある単語の入力に対してどれか一つの単語を予測するのではなく, 入力された単語の, 次に続き得る単語を複数出力するようになる. 具体的には, 入力された単語に対してその単語の次に続き得る単語を表す複数のノードがある程度の発火をする. 出力ノードの発火の程度は, 入力された単語の次に, どの程度の割合でその出力ノードの表す単語が続いたかによると考えられる. つまり, ネットワークの出力は, 入力された単語の次に続く単語の確率を表しているものと考えられる. そのため, 学習に用いた単語と予測した単語との平均二乗誤差による評価ではなく, 文生成の規則から算出される理想の確率と出力から算出される確率の類似度で評価することとする. 具体的には, 文生成の規則から算出される次に続く単語のカテゴリの確率を要素としたベクトル  $V$  と学習後のネットワークの出力から算出されるベクトル  $O$  との類似度をネットワークの学習の評価値とする.

まず, ベクトル  $O$  の算出方法について説明する. ネットワークの出力は 29bit であり, 各 bit が単語に対応している. 各単語は 12 種類のカテゴリのいずれかに属している. そこで, 各カテゴリごとに, そのカテゴリに属する単語に対応するノードの出力の和を算出する. このようにして算出された 12 次元のベクトルを  $O$  とする. 例として入力 eat に対するベクトル  $O^{\text{eat}}$  の要素である  $O_{N\_human}^{\text{eat}}$  の算出の仕方を示す.

$$O_{N\_human}^{\text{eat}} = O_{\text{boy}}^{\text{eat}} + O_{\text{girl}}^{\text{eat}} + O_{\text{man}}^{\text{eat}} + O_{\text{woman}}^{\text{eat}} \quad (3.1)$$



表 3.2: カテゴリー

カテゴリー	カテゴリーに属する単語
noun_human	boy, girl, man, woman
noun_animal	cat, dog, lion, mouse
noun_inanimal	book, car, rock
noun_agress	dragon, monster
noun_fragile	glass, plate
noun_food	bread, cookie, sandwich
verb_intran	exist, sleep, think
verb_tran	chase, like, see
verb_agpat	break, move
verb_percept	see, smell
verb_destroy	break, smash
verb_eat	eat

同様にして  $O_{N_{\text{animal}}}^{\text{eat}}$ ,  $O_{N_{\text{inanimal}}}^{\text{eat}}$  など 12 種類の要素を算出して  $O^{\text{eat}}$  を算出する.  $V$  と  $O$  の二つのベクトルの類似度は次式に示す, COSINE とする.

$$\text{COSINE} = \frac{V \cdot O}{\|V\| \|O\|} \quad (3.2)$$

この COSINE を各単語ごとに算出し, ネットワークの学習の評価として使用する. COSINE が大きければ, 次に続き得る語を予測できていると言える. 各単語ごとの COSINE が学習により上昇し, 次に続き得る語を予測できていると判断される場合に語順に関する言語知識を獲得しているとして評価する.

### 3.3 系列の区切りを考慮したモデル

前節で説明した Elman のモデルでは系列の区切りとなる情報は明示的にはネットワークに与えられていない. 現実の乳児がおかれている環境について考えた場合, 乳児が聞く他者の発話には文と文の間には沈黙の区間があり, 系列の区切りの情報は与えられているとみなす方が自然だと思われる. 実際, 1991 年の Elman のモデルでは系列の区切りの情報が単語 period として与えられている (図 3.1, 図中赤の下線が period を表す). こちらのモデルでは関係詞節を含むため, 関係詞との違いを明確にするために区切りの情報を追加したとも考えられるが, その意図については明確には述べられていない.

1990 年のモデルでは単純再帰ネットワークの入力系列は図 3.2 のようになる. 図中では単語で表されているが, ネットワークに与えられる入力とは符号化された 29bit のベクトルであり, 動詞や目的語, その他単語間の関係などの情報は与えられない. そのため, ネットワークは文生成の規則から算出された確率を学習するためには 2 語, または 3 語で文が区切れているということも学習し

表 3.3: 文のテンプレート

WORD 1	WORD 2	WORD 3
noun_human	verb_eat	noun_food
noun_human	verb_percept	noun_inainmal
noun_human	verb_destroy	noun_fragile
noun_human	verb_intran	
noun_human	verb_tran	noun_human
noun_human	verb_agpat	noun_inanimal
noun_human	verb_agpat	
noun_animal	verb_eat	noun_food
noun_animal	verb_tran	noun_animal
noun_animal	verb_agpat	noun_inanimal
noun_animal	verb_agpat	
noun_inanimal	verb_agpat	
noun_agress	verb_destroy	noun_fragile
noun_agress	verb_eat	noun_human
noun_agress	verb_eat	noun_animal
noun_agress	verb_eat	noun_food

なければならない。具体的には主語となる単語が入力された場合の予測ために、その 1 時点前までの情報を切り捨てるという動作を獲得することが期待される。それは、今回の入力系列の作成方法では文と文の連結はランダムに行われるため、主語となる単語の予測には直前の情報は必要ないからである。直前の文脈の影響を受けず、一定の確率を出力するために文脈の情報を切り捨てる必要がある。系列の区切りのある 1991 年のモデルでは、period の入力によって文脈の情報の切り捨てが行われることが期待される。

1990 年の Elman の研究によって系列の区切りの情報がないモデルでも COSINE がある程度上昇することは確認されている。しかし、追実験により COSINE が 1 にはならず、ある程度の値で停滞してしまう単語や減少してしまう単語があることが確認された。そこで次節では Elman と同様図 3.3 に示すように系列の区切りを表す単語を新たに加えた入力系列を作成し、系列の区切りの情

```

S → NP VP ""
NP → PropN | N | N RC
VP → V (NP)
RC → who NP VP | who VP (NP)
N → boy | girl | cat | dog | boys | girls | cats | dogs
PropN → John | Mary
V → chase | feed | see | hear | walk | live | chases |
    feeds | sees | hears | walks | lives

```

図 3.1: 1991 年モデルの系列の生成規則

報の追加によって学習結果に差が現れるかを比較する事とする。

まずは類似度 COSINE により系列の区切りのない場合とある場合との比較をする。次に学習後のネットワークの動作に違いが現れるかを特に文脈層に注目して評価する。

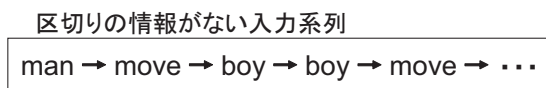


図 3.2: Elman のモデル

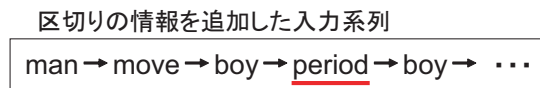


図 3.3: 区切りを考慮したモデル

### 3.4 実験

本節では、従来の Elman のモデル (以下 period なしのモデルと呼ぶ) と系列の区切りの情報を追加したモデル (以下 period ありと呼ぶ) との比較実験を行う。まず始めに実験条件について説明し、続いて評価方法について説明する。その後に実験結果を示し、考察を行う。

#### 3.4.1 実験条件

使用する単純再帰ネットワークの構成を説明する。period なしの方では、入力ノードの数は 29 個、隠れノード、文脈ノードの数はそれぞれ 150 個、出力ノードの数は 29 個とする (図 3.4)。period ありの方では、単語 period を追加するため入力ノードの数と出力ノードの数を 30 個とする。隠れ層と文脈層については period なしのモデルと同じノード数とした。(図 3.5)。結合強度の初期値は、-0.1 から 0.1 までの値でランダムに決める。隠れ層から文脈層の結合強度は 1 とする。

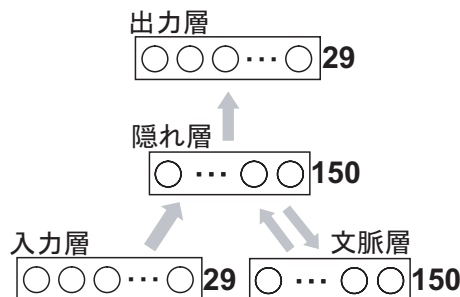


図 3.4: period なし

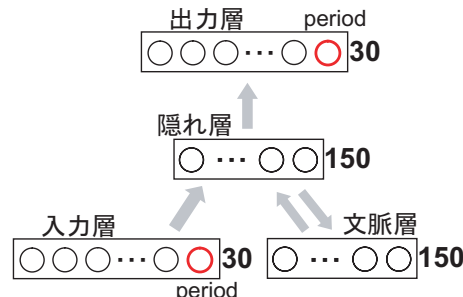


図 3.5: period あり

period なしの方で使用する入力系列は表 3.3 に従い 60000 個の文を作成し、連結したものである。period ありの方は表 3.3 の全ての文例の最後に period が続くようにし、40000 個の文を作成して連結したものである。



### 3.4.2 評価方法

学習後のネットワークの評価方法について説明する。period なし、ありのどちらも単語が入力されるごとにバックプロパゲーション法で学習を行う。注目する学習語数ごとに、得られた結合強度を固定した状態で入力に対する出力を算出して評価に使用する。出力の算出には表 3.3 に基づいて新たにテスト用の入力系列を用意してこれを使用する。

#### COSINE による評価

具体的な COSINE の算出法について述べる。単語  $x$  が入力されたときの出力層の各ノードの出力を  $x$  の次に入力される単語の生起確率の予測値と考える。 $x$  は入力系列中に複数回現れるので、予測値の平均を求める。これをさらにカテゴリごとに平均したものを 12 次元ベクトル  $O_x$  とする。この  $O_x$  と文の生成規則から算出される予測値の理想値  $P_x$  の類似度を表す  $\text{COSINE}_x$  を  $\text{COSINE}_x = \frac{O_x \cdot P_x}{|O_x| |P_x|}$  と定義する。

noun\_human, noun\_animal, noun\_inanimal のカテゴリに属する単語は主語と目的語の二つの場合があり、目標となる確率の分布が異なるのでこれらは分けて評価をする。

#### 文脈層に注目した評価

過去にどのような入力があったかの文脈の情報は文脈層の各ノードの値を要素とした 150 次元のベクトルで表される。単純再帰ネットワークは文脈層の状態によって同じ単語が入力された場合にでも出力を変えることが出来る。本研究で扱うモデルでは主語の単語が入力された時点で直前の文脈の情報が切り捨てられることと同じ単語が入力された場合でも文脈によって出力を変えて理想の確率を出力することが求められる。事前に知識が与えられていないネットワークが予測という課題を通してこれらの動作をした場合に、言語知識に対応した知識が得られたと評価するためである。そこでまず、系列の区切りの部分での文脈層の状態に注目して評価をする。単語  $x$  が入力された時の文脈層の状態を  $C_x$  とする。period が入力された時の  $C_{\text{period}}$  とその平均  $\bar{C}_{\text{period}}$  との平均二乗誤差  $\text{MSE}_{\text{period}}$  を算出し、period 以前の文脈情報によらず一定の値になるかを評価する。比較のため、30 個の単語の平均からなる  $C_{\bar{x}}$  と  $\bar{C}_{\bar{x}}$  との平均二乗誤差  $\text{MSE}_{\bar{x}}$  を算出する。 $\text{MSE}_{\bar{x}}$  に比べて  $\text{MSE}_{\text{period}}$  の方が小さい値となれば、period が他の単語に比べて文脈層を一定にしていると評価出来る。また、period なしの入力系列についても文末の単語の入力時の文脈層の値を系列の区切りの部分として、平均二乗誤差による評価値を  $\text{MSE}_{\text{period}}$  とした。

次に、主語と目的語の両方となる単語についてそれぞれの場合の平均二乗誤差  $\text{MSE}_{x,S}$  と  $\text{MSE}_{x,O}$  を算出して比較する。系列の区切りが獲得されている場合、文脈の情報が出力に影響を及ぼす目的語の方の値にばらつきがあらわれ、平均二乗誤差は大きくなると考えられる。

### 3.4.3 結果と考察

period なし、ありのモデルについて実験を行い、前節で説明した評価方法により評価する。period なし、period ありの二つのモデルについて、先に説明した文生成の規則に従って入力系列を作成し

学習させた。period なしの方は 60000 個の文を連結した 165075 語の単語系列を入力系列として使用した。period ありの学習では 40000 個の文を連結した 150040 語の単語系列を入力系列として使用した。評価は 0 語学習から 150000 語学習時まで注目して行う。

### 3.4.4 COSINE

まず、全体としてどちらのネットワークの学習がより進んだかを確認するため、1000 語学習するごとに単語全体の COSINE の平均値を算出し、学習回数に対してどう推移したかを示す (図 3.6)。period ありの学習では period なしに比べ学習時に早い段階から COSINE が高い値になっており、系列の区切りを表す単語を入力系列に加えた period ありのモデルの方が学習が速く進むことが確認された。

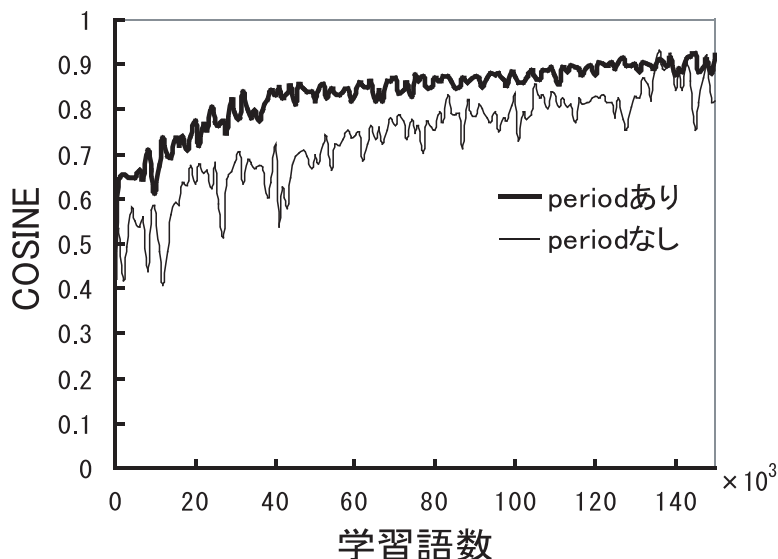


図 3.6: COSINE の推移

period なしと period ありのそれぞれのモデルについて学習が十分に進んだ 150000 語学習時点での各単語ごとの COSINE を示す。period なしの結果とありの結果をそれぞれ表 3.4 と表 3.5 に示す。period ありのモデルでは、eat を除く全ての単語における COSINE が period なしのモデルより高い値となっていることが確認出来る。

次に主語と目的語の両方となる単語について分けて評価した場合の COSINE を示す (表 3.6, 表 3.7)。noun\_human, noun\_animal, noun\_inanimal のどのカテゴリーに属する単語においても period ありのモデルの方が COSINE が高くなっている。特に表 3.7 の目的語の COSINE は全ての単語でほぼ 1 となった。これは、今回の文の生成規則では目的語の後には必ず系列の区切りを表す単語 period のみが続いたため、目的語が入力された場合の教師が必ず period となり、他の単語より学習しやすかったためと考えられる。しかし 3.3 節で述べたように、ネットワークには入力された単語が主語なのか目的語なのかの情報は与えられない。そのため、文脈層のパターンを利用してネットワーク自身が出力を変えなければならない。period ありのモデルでは主語の COSINE も高い値と

表 3.4: COSINE(period なし)

入力	COSINE		入力	COSINE		入力	COSINE
book	0.902		eat	0.980		plate	0.921
boy	0.818		exist	0.757		rock	0.908
bread	0.932		girl	0.831		sandwich	0.936
break	0.902		glass	0.924		see	0.434
car	0.905		like	0.656		sleep	0.776
cat	0.893		lion	0.897		smash	0.988
chase	0.666		man	0.831		smell	0.222
cookie	0.938		monster	0.962		think	0.762
dog	0.897		mouse	0.849		woman	0.822
dragon	0.962		move	0.647			

表 3.5: COSINE(period あり)

入力	COSINE		入力	COSINE		入力	COSINE
book	0.983		eat	0.969		plate	0.999
boy	0.878		exist	0.996		rock	0.986
bread	0.999		girl	0.881		sandwich	0.999
break	0.801		glass	0.999		see	0.847
car	0.986		like	0.668		sleep	0.996
cat	0.933		lion	0.941		smash	0.988
chase	0.678		man	0.883		smell	0.651
cookie	0.999		monster	0.969		think	0.996
dog	0.941		mouse	0.940		woman	0.882
dragon	0.968		move	0.934		period	0.988

っており、文脈層のパターンを利用して出力を変えるよう動作していることがわかる。

一方、period なしのモデルでは主語の COSINE が高い値となっていない。それに関わらず、表 3.4 から主語と目的語の COSINE の平均値で一つの単語として評価した場合には高い値となっている。これは period なしのモデルでは文脈層のパターンによって出力を大きくは変えておらず、文脈によらず平均的なパターンを出力していることを示唆している。このことは後の文脈層が一定のパターンになっているかを評価をすることで確認出来る。次に、以上の結果を踏まえつつ、次節の文脈層パターンの平均二乗誤差による評価に移る。

表 3.6: COSINE(period なし)

human	COSINE	animal	COSINE	inanimal	COSINE
boy_S	0.724	cat_S	0.840	book_S	0.763
girl_S	0.725	dog_S	0.832	car_S	0.749
man_S	0.726	mouse_S	0.834	rock_S	0.757
woman_S	0.709	lion_S	0.852		
boy_O	0.965	cat_O	0.969	book_O	0.964
girl_O	0.964	dog_O	0.975	car_O	0.963
man_O	0.964	mouse_O	0.968	rock_O	0.963
woman_O	0.967	lion_O	0.962		

表 3.7: COSINE(period あり)

human	COSINE	animal	COSINE	inanimal	COSINE
boy_S	0.868	cat_S	0.923	book_S	0.954
girl_S	0.859	dog_S	0.950	car_S	0.950
man_S	0.871	mouse_S	0.952	rock_S	0.954
woman_S	0.870	lion_S	0.954		
boy_O	0.999	cat_O	0.999	book_O	0.999
girl_O	0.999	dog_O	0.999	car_O	0.999
man_O	0.999	mouse_O	0.999	rock_O	0.999
woman_O	0.999	lion_O	0.999		

### 3.4.5 文脈層の評価

次に、period なしの学習と period ありの学習のそれぞれについて  $MSE_{\bar{x}}$  と  $MSE_{\text{period}}$  の 10000 回学習時の結果を示す (表 3.8)。period ありの方が period なしの場合に比べて  $MSE_{\bar{x}}$  の値が大きくなっている。これは、同じ単語を入力した場合でも以前に入力された単語の違いによって、文脈層の各ノードの値を大きく変えていることを示す。つまり、period ありの方が過去の文脈の情報を強く出力に反映した学習結果を与えるモデルだと言える。次に  $MSE_{\text{period}}$  に注目すると、 $MSE_{\bar{x}}$  に比べて小さい値となっていることがわかる。特に、period ありの方は非常に小さな値となっている。これは period の入力により、それ以前の入力の違いに関わらず文脈層の状態を一定とし、文脈層のリセットの効果が現れていることを示す。つまり、入力系列中に区切りとなる部分があることを SRN が学習できていると評価出来る。

次に主語と目的語の両方となる単語の入力に注目する。今回はその一つとして boy に注目し、 $MSE_{\text{boy}_S}$  と  $MSE_{\text{boy}_O}$  の値を、period ありと period なしのそれぞれについて示す (表 3.9)。どちらにおいても、主語に比べて目的語の方が平均二乗誤差の値は大きくなっている。ただし、値の比を考えると、period ありの方が大きくなっており、過去の入力情報を反映させて出力を変え

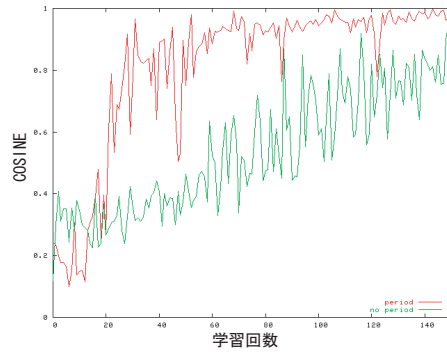


図 3.7: book\_S

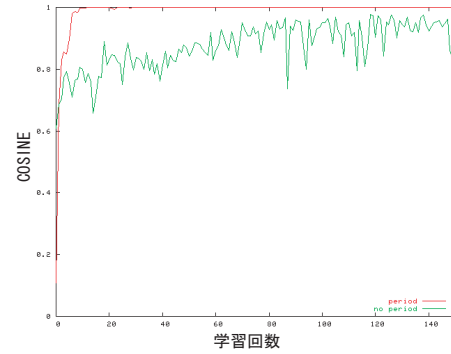


図 3.8: book\_O

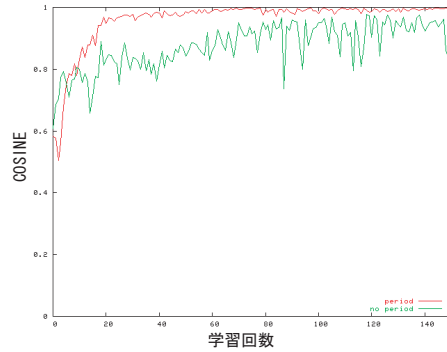


図 3.9: period と book\_O

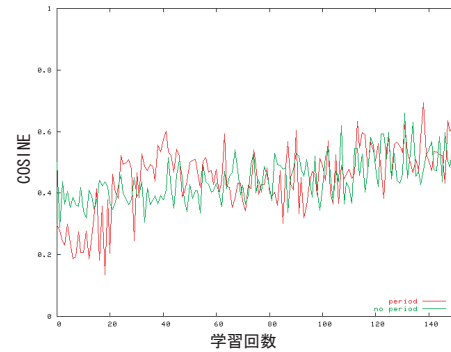


図 3.10: verb\_tran

る傾向が period なしより強くなっているものと考えられる。

表 3.8: 平均と period

	$MSE_{\bar{x}}$	$MSE_{\text{period}}$
periodあり	$1.08 \times 10^{-2}$	$5.28 \times 10^{-4}$
periodなし	$3.76 \times 10^{-3}$	$1.74 \times 10^{-3}$

表 3.9: 主語と目的語

	$MSE_{\text{boy}_S}$	$MSE_{\text{boy}_O}$
periodあり	$1.20 \times 10^{-4}$	$5.38 \times 10^{-4}$
periodなし	$4.23 \times 10^{-4}$	$6.81 \times 10^{-4}$

また、period ありで  $C_x$  の大きさに注目したところ、 $\bar{C}_{\text{boy}_S}$  と  $\bar{C}_{\text{boy}_O}$  とで大きさが異なることが確認された (表 3.10、表 3.11)。period ありにおいては  $|C_{\text{boy}}|$  の平均値に差があり、分散を考慮に入れると主語の場合の boy と目的語の場合の boy とを分類して扱っていると考えられる。目的語の方の分散が大きくなっているのは、過去の入力の違いにより値が変わるためと考えられる。一方 period なしについては平均値の差は小さくなっている。また、分散の方にも大きな差は見られない。これは、period なしでは主語と目的語の違いを SRN が明確に区別できていないことを示唆する結果である。

### 3.5 未解決の問題点

前節までの議論により、系列の区切りの情報を与えて学習させることが学習結果の改善に繋がる  
ことが確認された。系列の区切りを表す period の入力により文脈層のパターンが一定となること  
と、同じ単語の入力によっても主語と目的語で文脈層のパターンは異なっていることがネットワー  
クの特徴として現れていた。前者は不要な文脈情報を切り捨てていること、後者は同じ入力でも過  
去の入力情報である文脈情報を利用して大きく出力値を変えていることを示している。学習結果  
の改善はこれらが有効に働いているためと考えられる。

しかし、period なし、period ありのどちらのモデルにも見うけられる問題があることが分かって  
きた。本節ではその問題点について説明する。

#### 3.5.1 ダイバージェンスによる評価

今回扱ってきたモデルでは、ネットワークが規則から算出された確率分布を出力するようになる  
かどうかをベクトルの類似度として COSINE により算出して評価した。しかし、確率分布として  
類似したかの評価にはより適切な評価方法と考えられるダイバージェンスがある。ダイバージェ  
ンス  $D$  は二つの確率分布  $P_M(x)$  と  $P(x)$  がどの程度離れているかを示す距離尺度であり、 $D$  が 0 に  
近い程確率分布が類似していると評価出来る。ダイバージェンス  $D$  は以下の式で定義される。

$$D(x) = \sum_i P(C^{(i)}|x) \log_2 \frac{P(C^{(i)}|x)}{P_M(C^{(i)}|x)} \quad (3.3)$$

$P_M(C^{(i)}|x)$  と  $P(C^{(i)}|x)$  はそれぞれ  $P_M(x)$  と  $P(x)$  のある  $x$  に対する確率変数  $C$  の実現値  $C^{(i)}$  の  
発生する確率を表している。上式は言語モデルの評価としても使用されており、本研究におけるモ  
デルの評価としてより適切なものと考えられる。そこで本研究の period なしとありのモデルにつ  
いてダイバージェンスによる評価を行った。言語モデルの評価として使用される場合には  $P_M(x)$   
を目標とする言語が従う確率分布、 $P(x)$  を言語モデルが実現した確率分布として  $D$  を算出する。  
本研究ではそれぞれ  $x$  が入力された単語、 $P_M(x)$  が規則から算出された確率、 $P(x)$  が学習後の出  
力から算出された確率に対応する。前節で使用した period なし、ありのモデルについて  $D$  を各単  
語ごとに算出した。学習回数に対する推移を図 3.11, 図 3.12 に示す。図 3.11 のように値が 0 に近  
付き、目標の確率分布に近付いている単語 (図は verb\_agpat に属する単語) の他に、図 3.12 のよう  
にある値で停滞が起り減少しない単語 (図は noun\_human に属する単語が主語として入力され  
た場合) があることが確認された。

ダイバージェンス  $D$  は定義の式から  $P_M(C^{(i)}|x)$  が 0 ではない場合に  $P(C^{(i)}|x)$  が 0 になる、或  
は  $P_M(C^{(i)}|x)$  が 0 である場合に  $P(C^{(i)}|x)$  が 0 にならないなど、あるカテゴリー  $C^{(i)}$  の確率が

表 3.10: period なし

	$E[\bar{C}_{\text{boy}}]$	$V[\bar{C}_{\text{boy}}]$
主語	3.20	$1.66 \times 10^{-3}$
目的語	3.28	$1.92 \times 10^{-2}$

表 3.11: period あり

	$E[\bar{C}_{\text{boy}}]$	$V[\bar{C}_{\text{boy}}]$
主語	3.22	$1.66 \times 10^{-3}$
目的語	3.44	$1.92 \times 10^{-2}$

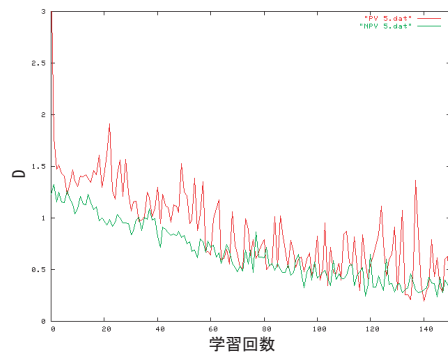


図 3.11: ダイバージェンス

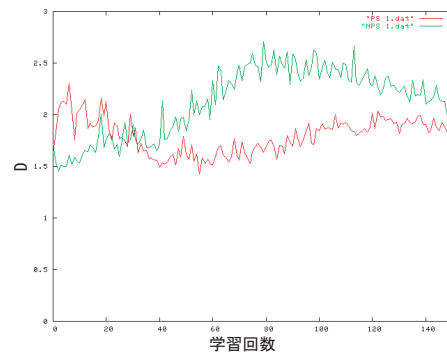


図 3.12: 停滞がおこる単語

抜け落ちるようになった場合に大きい値となる評価値である。停滞が起こった単語は主語としての `noun_human` である。表 3.3 を見ても分かるように `noun_human` は最も多くの種類の文を作る。そのため確率が 0 とならないカテゴリーは多く、いずれかのカテゴリーの確率が抜け落ちてしまった場合には  $D$  は大きくなってしまふ。その結果を反映したのが図 3.12 だと考えられる。これは正確な確率分布が出力されていないことを示し、停滞は学習が進んでいないと評価されるため、問題であると考えられる。

### 3.6 まとめ

系列の区切りの考慮がなかった Elman のモデルと文の区切りの情報を追加したモデルとを評価実験を通して比較検討した。系列の区切りを加えた入力系列による学習では学習結果が改善され、区切りを表す単語が入力されると文脈層の状態が一定となるという不要な文脈情報の切り捨てを単純再帰ネットワークがするようになることが確認された。Elman は学習の際に文脈層の値を決められた語数ごとに強制的にリセットしつつ学習させることで学習結果が改善されると述べている。そのため、今回は系列の区切りの情報を単語として加えることで単純再帰ネットワークに与えたが、区切りの情報を文脈層のパターンのリセット（区切りの部分で文脈層の各ノードの値を一定にする）として与えて学習させれば本研究の結果と同様、学習結果に改善が見られるかもしれない。

しかし、学習結果に改善が見られた `period` ありのモデルにおいても、ダイバージェンスによる評価では学習が進んでいない単語が確認された。この結果が本モデルの不完全さを要因とする妥当な結果か否かを議論する必要がある。そこで次章ではこれがどのような要因から起こっているかについて検討し、その後に考えられる要因の改善のための方法を提案していく。

## 第4章

# 時系列を学習する 確率的単純再帰ネットワークモデル

### 4.1 はじめに

前章の議論から、系列の区切りの情報を学習の際に与えることによって学習結果が改善されることを確認した。しかし、ダイバージェンスによる評価によって改善されたモデルにおいても系列の作成規則から算出される確率分布を正確には出力出来ていないことがわかってきた。簡単なモデルによる検討を通して、これは period あり、なしなどと言ったモデルにおける問題があるのではなく、単純再帰ネットワークの学習方法自体に問題があるのではないかと考えられた。この学習方法における問題の改善は、単純再帰ネットワークを用いたモデルを構築して議論している他の研究においても望ましいものだと考えられる。なぜならば今回の問題が単純再帰ネットワークの計算能力の低下に関わってくるからである。モデル研究で得られた結果の妥当性を高めるためにも計算能力を十分に引き出す必要がある。

そこで本章では予測学習により次に続く入力の確率分布を正確に出力するための単純再帰ネットワークの学習方法を提案する。この学習方法を適用した単純再帰ネットワークによるモデルを確率的単純再帰ネットワークモデルと呼ぶこととする。本研究ではそのモデルとして前章までで扱った Elman の問題を考え、提案学習方法の有効性を確かめるための評価実験を行う。

### 4.2 単純再帰ネットワークの学習方法についての検討

本節では単純再帰ネットワークの学習方法について、確率分布を正確に出力するような学習方法であるかを検討する。これまで単純再帰ネットワークの学習方法には2章で説明したとおりバックプロパゲーション法を使用してきた。しかし、バックプロパゲーション法による教師あり学習は入力されたパターンに対してある決まったパターンを出力するように学習するための方法である。



本研究で扱っているモデルでは、3章でも述べたようにある入力に対する出力は決定的には決まらない。結果的に確率分布に類似した出力をするようになるが、確率分布を獲得するための学習方法とは言えないのではないかと考えられた。そこで、まずはバックプロパゲーション法により確率分布が正確に出力されるようになるかを検討する。

#### 4.2.1 3層パーセプトロンによる検討

バックプロパゲーション法は多層パーセプトロンの学習法として広く使用されている。本研究で使用している単純再帰ネットワークは3層パーセプトロンに文脈層を付加したものである。この文脈層に保持された直前の隠れ層のパターンは次の時点での入力として扱われる。そのため、文脈層を入力層と考えれば各時点で通常の3層パーセプトロンと同じものであると考えることが出来る。そこでバックプロパゲーション法について評価を行うために3層パーセプトロンを使用して簡単な実験を行うことにする。3層パーセプトロンの構成は以下のようである(図4.1)。このネット

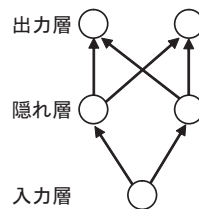


図 4.1: 3層パーセプトロンの構成

ワークには常に入力として1を与え、教師として(1,0)と(0,1)の二つのパターンをそれぞれ0.5, 0.5の確率で与えてバックプロパゲーションにより十分な回数学習させた。その結果、ネットワークは確率を正確に出力するようになった。二つのパターンを与える確率を変えて学習させてもこの結果は変わらなかった。

この結果についての考察を行う。出力層の二つのノードの出力値をそれぞれ $o_1, o_2$ 、対応する教師を $t_1, t_2$ とする。バックプロパゲーション法で使用される誤差関数 $E$ は以下の式で表される。

$$E = \frac{1}{2}((o_1 - t_1)^2 + (o_2 - t_2)^2) \quad (4.1)$$

今、 $o_1$ と $o_2$ の和が1であるとする、 $o_2 = 1 - o_1$ であり、上式は以下の様に変えられる。

$$E = \frac{1}{2}((o_1 - t_1)^2 + (1 - o_1 - t_2)^2) \quad (4.2)$$

学習の際に与えた各入力と教師を $N$ 回与えることにより誤差関数の平均を算出すると、

$$\begin{aligned} \frac{1}{N} \sum E &= \frac{1}{N} \sum \frac{1}{2}((o_1 - t_1)^2 + (1 - o_1 - t_2)^2) \\ &= \frac{1}{2N} \sum (o_1^2 - 2o_1t_1 + t_1^2 + (1 - o_1)^2 - 2(1 - o_1)t_2 + t_2^2) \end{aligned} \quad (4.3)$$

今、教師が(1,0)を $p$ 、(0,1)を $1-p$ の確率で与えたとする、十分 $N$ を大きく取れば $\frac{1}{N} \sum t_1 = p$ ,

$\frac{1}{N} \sum t_1^2 = p$ ,  $\frac{1}{N} \sum t_2 = 1 - p$ ,  $\frac{1}{N} \sum t_2^2 = 1 - p$  となるので最終的に以下の様にまとめられる.

$$\begin{aligned} \frac{1}{N} \sum E &= \frac{1}{2}(o_1^2 - 2o_1p + p + (1 - o_1)^2 - 2(1 - o_1)(1 - p) + (1 - p)) \\ &= (o_1 - p)^2 + p(1 - p) \end{aligned} \quad (4.4)$$

上式 (4.4) を  $o_1$  の関数と考えると誤差関数最小となるのは  $o_1 = p$  の場合である. 3層パーセプトロンが正確に確率を出力するようになるのはこれを反映したものだと考えられる.

以上の議論からバックプロパゲーション法による学習によって, ある入力に対する教師が確率的に変わる場合に確率を出力することが可能であると考えられる.

#### 4.2.2 隠れ層パターンの類似による学習能力の低下

前節の議論からバックプロパゲーション法により確率を出力可能であると考えられる. しかし, 前章のモデルではその目的となる出力が実現されてはいない. 前章のモデルと前節の3層パーセプトロンによるモデルとの違いとして学習する入力パターンが複数あるかないかという点が挙げられる. 単純再帰ネットワークでは文脈層のパターンも入力として考えられ, 文脈によって出力すべき確率も変わるため, 学習すべき入力パターンが非常に多いと言える. 文脈層パターンを加えた入力パターンを分類して表現するのは隠れ層である. 隠れ層パターンが類似した場合には隠れ層パターンから算出される出力層パターンも類似したものとなると考えられ, 学習結果が悪化することの一つの説明となる.

確認のため, 前節扱ったネットワークをさらに単純化させ, 2層のネットワーク (図 4.2) により実験を行った. 入力層を3層パーセプトロンの隠れ層と考え, 二つの類似したパターンを入力としてそれぞれ別の確率で  $(0, 1)$ ,  $(1, 0)$  の教師を与え, それぞれその確率を出力するように学習出来るかを確認した. 具体的には入力パターン  $A$  と  $B$  について  $A = (0.6, 0.5)$ ,  $B = (0.5, 0.6)$  とし, 教師

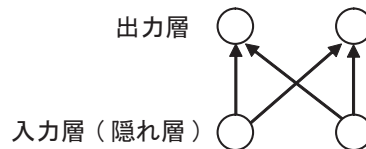


図 4.2: 2層パーセプトロンの構成

として  $(0, 1)$  と  $(1, 0)$  を  $A$  にはそれぞれ 0.2 と 0.8 の確率で与え,  $B$  にはそれぞれ 0.5 と 0.5 の確率で与える. 十分な学習回数学習させた結果,  $A$  の入力に対しては  $(0.28, 0.72)$ ,  $B$  の入力に対しては  $(0.43, 0.57)$  を出力するようになった. この結果から入力パターンが類似している場合には正しく学習出来ない場合があることが確認された.

以上の検討から前章で扱ってきたモデルが正確に確率を出力するようにならなかったのは, 異なる確率を出力すべきパターン同士が隠れ層で類似してしまっているためだと考えられる. この考えの妥当性を裏付ける既存の議論として単純再帰ネットワークにおける計算能力の低下の問題についての議論がある. 森田ら ([4], [5]) は単純再帰ネットワークの学習能力における問題点についてまとめており, 文脈の平均化の問題を問題点の一つとして挙げている. 単純再帰ネットワーク

における文脈層のノードの数を多数にすると、文脈層で表されるパターンが平均的なものとなるという説明をしている。文脈層が平均的なパターンとなるということは分離すべきパターンが類似してしまうということである。

前章でも既に何度も述べたが、入力と同じでも文脈層に保存された文脈情報によって異なる出力が可能であるのが単純再帰ネットワークの特徴であり、本研究で扱う問題においても主語と目的語で同じ単語の入力でも出力を変えるとといった動作のために必要とされる。Elmanをはじめとして、単純再帰ネットワークを使用したモデルによる研究のために、単純再帰ネットワークの学習能力を低下させる文脈層の平均化の問題の改善は必要である。

### 4.2.3 クラスタ分析による評価

実際に文脈層の平均化の問題が起きているかを確認するため、学習後のネットワークについて各単語に対応する文脈層パターンの類似性によって評価する。具体的には、入力された各単語の隠れ層での高次元表現をクラスタ分析することで、ネットワークがどの単語とどの単語を類似したものとして扱っているかを評価する。前章では period なしのモデルよりも period ありのモデルの方に改善が見られた。この結果が平均化の問題の改善によるものであれば、異なる確率を出力する文脈層パターン同士はその類似性が低くなるものと考えられる。クラスタ分析においてはクラスタ間距離が大きくなるかどうかでそれが判断出来る。そこでこれら二つのモデルについてクラスタ分析の結果を比較し、period ありのモデルの方が異なる出力をするパターン同士が離れたクラスタに分類されるかに注目する。

#### クラスタ分析

まずクラスタ間距離について説明する。クラスタ分析は多変量解析の一手法であり、個体どうしの類似性を評価するために用いられる。類似性の評価には取扱の容易なユークリッド距離を用いた。クラスタ分析には、距離の再定義の方法によって、最近隣法、最遠隣法、重心法、群平均法、Ward 法などがあるが、今回は Ward 法を用いた。

$n$  個の個体が  $p$  個の特徴  $X_{i1}, X_{i2}, \dots, X_{ip}$  ( $i = 1, 2, \dots, n$ ) を持つ場合の、Ward 法を用いたクラスタ分析の手順を説明する。初期状態として  $n$  個のクラスタがあるとする (各クラスタは 1 個体ずつを含むと考える)

##### 手順 1

クラスタ間のユークリッド平方距離  $d_{ij}^2$  を算出する。

$$d_{ij}^2 = \sum_{k=1}^p (X_{ik} - X_{jk})^2 \quad (i, j = 1, 2, \dots, n) \quad (4.5)$$

##### 手順 2

ユークリッド平方距離の最も近いクラスタを併合して、1つのクラスタとする。今、クラスタ  $a$  とクラスタ  $b$  が併合されてクラスタ  $c$  が作られるとする。  $d_{ab}$ ,  $d_{xa}$ ,  $d_{xb}$  をクラスタが併合する前の各クラスタ間の距離とした場合、併合後のクラスタ  $c$  と  $a, b, c$  以

外のクラスター  $x$  とのユークリッド平方距離は次式によって算出される.

$$d_{xc}^2 = \frac{n_x + n_a}{n_x + n_c} d_{xa}^2 + \frac{n_x + n_b}{n_x + n_c} d_{xb}^2 - \frac{n_x}{n_x + n_c} d_{ab}^2 \quad (4.6)$$

ただし,  $n_x, n_a, n_b, n_c$  はそれぞれのクラスターに属する個体の数とする.

### 手順3

手順2によって2個のクラスターが1個のクラスターにまとめられ, 総クラスター数が1個減る. クラスター数が1個になるまで手順2を繰り返す.

例として2次元の変数をもつ6個体の場合のクラスタリングの様子を示す(図4.3). クラスタ分析の結果は樹形図で表される. この図の, 個体どうしの類似性, およびクラスターどうしの類似性が評価できる.

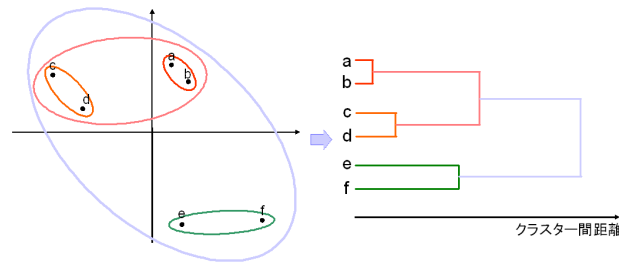


図 4.3: クラスタ分析

### 結果の比較

単純再帰ネットワークの学習に用いた系列を入力として与え, 各入力に対する隠れ層の値を計算し, 入力した単語ごとに平均を取ったものを入力に対する隠れ層の各ノードの値として用いる. そして, 各単語を個体, 単語に対する隠れ層の各ノードの値を特徴として用いる. つまり, 150次元の変数を持つ29個体(periodありは30個体)についてクラスタ分析をする. periodなしのモデルとperiodありモデルの結果をそれぞれ図4.4, 図4.5に示す. 二つの結果を比較すると, periodありのモデルでのクラスター間距離が非常に大きくなっていることが分かる. これは, 異なる確率を出力する文脈層パターン同士の類似性が低くなり, 平均化の問題が改善されていることを示している. このことから, 文脈層パターンの平均化を防ぐことによって学習結果が改善される可能性があると考えられる.

## 4.3 文脈層パターンの平均化を防ぐ新学習法の提案

前節で行った検討により, 学習の際の文脈層の平均化が正確な確率の出力を阻害している一因だと考えられた. 本節では文脈層の平均化を防ぐ新たな学習法を提案する. 具体的には, 従来のバックプロパゲーション法をベースとして, 新学習法では文脈層から隠れ層への結合の更新則に改変を

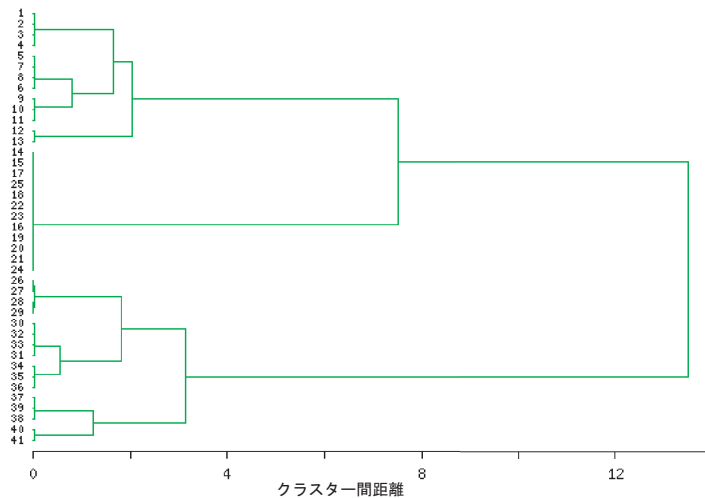


図 4.4: period なし

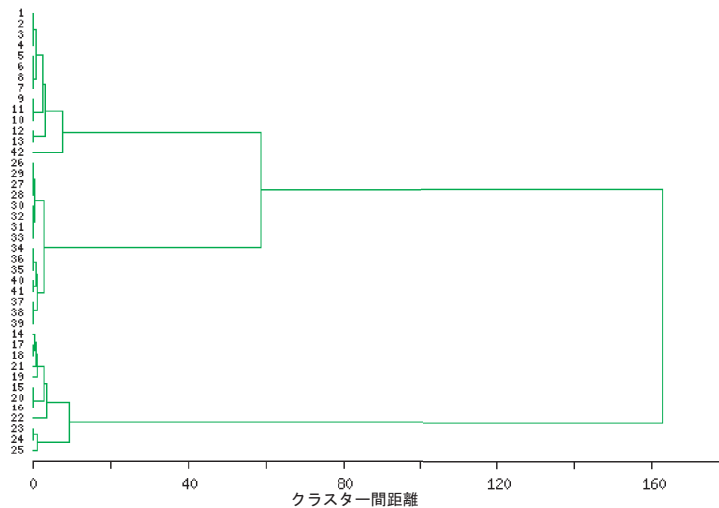


図 4.5: period あり

加えた。この改変にはネットワーク自身の持つパラメータを利用しており、単純再帰ネットワークの構造や機能に新たな要素は追加していない。

#### 4.3.1 文脈層パターンを利用した新学習法

本節では提案する学習方法について説明する。2章で説明したようにある時点  $t$  での単純再帰ネットワークの文脈層パターン  $C(t)$  は直前の隠れ層パターン  $H(t-1)$  そのものである。そのため、文脈層が平均的なパターンとならないようにするという事は、文脈層パターンから隠れ層パターンを算出する際に出来るだけ異なるパターンとなるようにすれば良いと言える。つまり、一時点前の隠れ層パターン  $H(t-1)$  と  $H(t)$  とが大きく異なる様にすれば良いということである。その

実現のためには文脈層から隠れ層への結合の更新則に改変を加えれば良い。そこで、文脈層ノード  $l$  から隠れ層ノード  $j$  への結合  $w_{lj}$  の更新量  $\Delta w_{lj}(t)$  を決定する更新式を以下のように改変した。

$$\Delta w_{lj}(t) = -\epsilon \alpha^{(1 - (H_j(t) - C_j(t))^2)} \delta_j C_l(t) \quad (4.7)$$

$C_l(t)$  は文脈層ノード  $l$  の出力値、 $H_j(t)$  は隠れ層ノード  $j$  の出力値、 $\epsilon$  は学習率、 $\alpha$  は正の定数である。上式の特徴は  $(1 - (H_j(t) - C_j(t))^2)$  を利用した更新量の重み付けである。この重み付けでは  $H_j(t) = C_j(t)$  の場合に最も大きくなる。ここで、 $C_j(t) = H_j(t-1)$  であり、各隠れ層のノードが直前の値からは離れた値になるように学習させるということである。その結果として、文脈層パターンの平均化の問題を改善しつつ正確に確率を出力するようになることを期待する。

### 4.3.2 実験による評価

本節では提案学習法の有効性の確認を目的として、実験による評価を行う。実験には前章で使用した period ありのモデルを使用する。前章行った文脈層の評価をもとにした議論で、period ありのモデルの方が period なしのモデルよりも文脈情報の違いによって大きく動作を変えられていたためである。比較には通常のバックプロパゲーション法を使用している前章の period ありのモデルの結果を用い、類似度 COSINE とダイバージェンス  $D$  に注目する。実験に使用する入力系列、ネットワークの構造などは前章と同じものを使用する。

実験結果を以下の図に示す。図中赤線が提案学習法、緑線がバックプロパゲーション法に対応している。6つの図はそれぞれ主語としての `noun_human` が入力された場合、`noun_human` が目的語として入力された場合、`verb_agpat` が入力された場合である。図は横軸が学習回数であり、0回学習から150000回学習までに注目する。

図 4.7, 図 4.9, 図 4.11 から Elman が使用した従来の評価方法 COSINE においてはほぼ同等の結果と評価される。一方ダイバージェンスを評価方法として使用した場合には、図 4.8, 図 4.10, 図 4.12 から提案学習方法の方がより正確に目標の確率を出力するようになったという評価を得た。特に主語として `noun_boy` を入力した場合の結果が大きく改善されており、文脈層パターンの平均化を防ぐことを目的とした提案学習法が有効に働いていると言える。

### クラスター分析による比較

提案学習法についてクラスター分析による評価は以下ようになった (図 4.6)。提案学習法の適用によって文脈層パターン同士が period ありのモデルの場合 (図 4.5) よりクラスター間距離が離れるように学習が進んでいることを確認した。

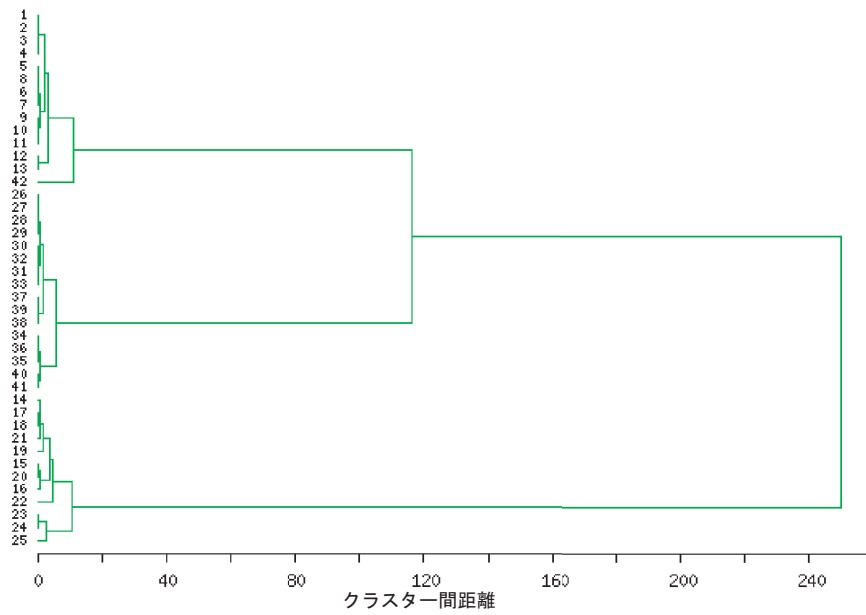


図 4.6: クラスタ分析 (提案学習法)

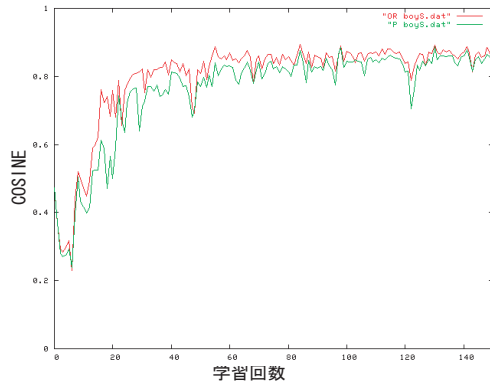


図 4.7: COSINE(主語 boy)

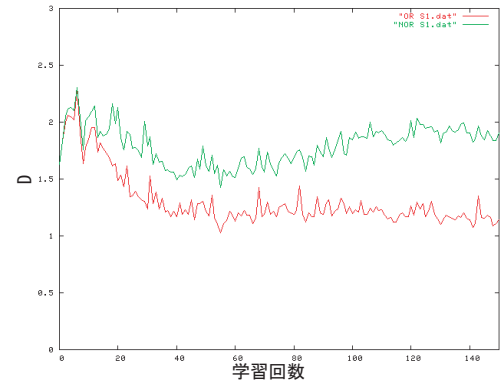


図 4.8: D(主語 boy)

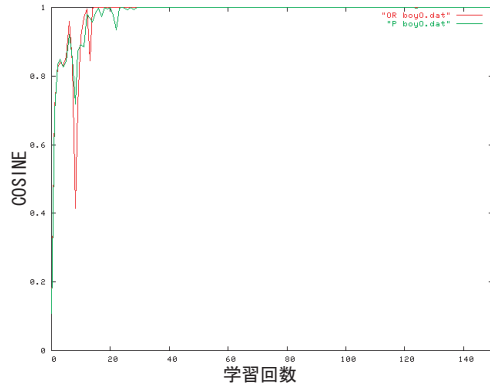


図 4.9: COSINE(目的語 boy)

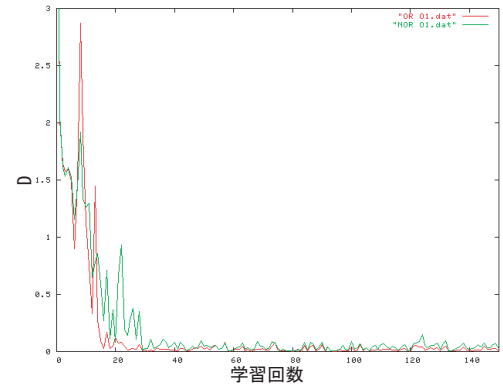


図 4.10: D(目的語 boy)

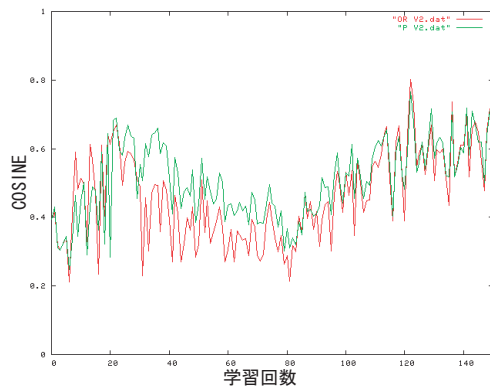


図 4.11: COSINE(chase)

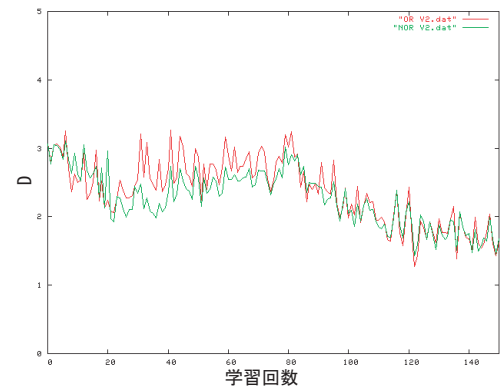


図 4.12: D(chase)



## 4.4 既存の時系列学習方法との比較

本章で提案した学習方法は、式(4.5)の重み付けの部分に注目し、 $C_j(t) = H_j(t-1)$ であることを踏まえると、過去のネットワークのパラメータを利用する学習方法だと言える。過去のネットワークの情報を利用する学習方法による学習は時系列学習と呼ばれ、提案手法もその一つだと見なすことが出来る。既存のニューラルネットワークの時系列学習法の一つとして慣性項を考慮したバックプロパゲーション法 [16] が挙げられる。そこで本節では既存の時系列学習法である慣性項を考慮したバックプロパゲーションと提案手法との比較を行うこととする。

### 4.4.1 慣性項を考慮したバックプロパゲーション法

慣性項を考慮した学習法は2で説明した通常のバックプロパゲーション法を以下の更新式に変更する。

$$\begin{aligned}\Delta w_{jk}(t) &= -\epsilon \delta_k y_j + a \Delta w_{jk}(t-1) \\ \Delta w_{ij}(t) &= -\epsilon \delta_j x_i + a \Delta w_{ij}(t-1)\end{aligned}\tag{4.8}$$

各結合の更新量を算出する際、直前の更新量を考慮するのが慣性項を考慮した学習法の特徴である。この学習方法は直前と同方向の更新が加速され、逆方向の更新は減速される。 $a$ は0から1までの正の定数で、この値が大きい程直前の更新量の影響を受けるようになる。この学習法は局所解からの脱出、学習速度の上昇など収束性の向上のために使用される。次節で提案学習法との比較を行う。

### 4.4.2 実験による評価

慣性項を考慮したバックプロパゲーション法と提案学習法との比較のための実験を行った。実験条件は4.3節での実験と同様に前章で使用した period ありのモデルを使用した。学習回数の増加に対する COSINE, ダイバージェンスの両評価値の推移を図4.13から図4.18に示す。図中赤線が慣性項を考慮した学習法、緑線が提案学習法に対応している。主語、目的語それぞれの場合の noun\_boy については慣性項を考慮したバックプロパゲーション法と提案学習法とはほぼ同等の結果となった。一方、verb\_chase については慣性項を考慮したバックプロパゲーション法の方が収束性が向上している。最終的な評価値の値が提案学習法より良好であり、さらに特徴的なのは学習回数の増加に従ってよりスムーズに評価値が上昇(ダイバージェンスは減少)していることが確認出来る。これは慣性項の導入の目的の一つである学習速度の上昇を良く反映した結果である。この結果だけを見ると提案学習法の有用性が否定されているように見える。しかし、提案学習法の目的は文脈層パターンの平均化を防ぐことにあり、収束性の向上を目的としたものではない。同じ時系列学習でもその目的が異なる。むしろ、慣性項の導入による収束性の向上が学習結果を改善した事は、両学習法を組み合わせることでそれぞれの学習法の利点を同時に得ることが期待出来るという点で望ましい結果だと言える。そこで次節では実際に提案学習法に慣性項を追加し、提案学習法によって文脈層パターンの平均化を防ぎつつ慣性項によって収束性を向上させられるかを確認する。

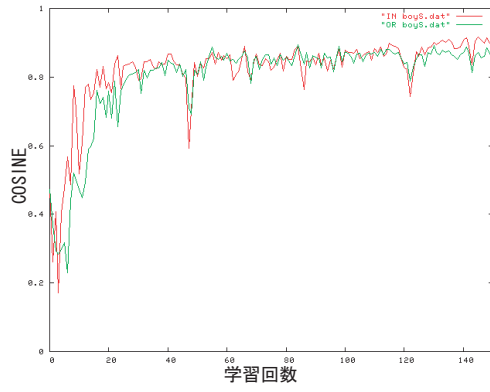


図 4.13: COSINE(主語 boy)

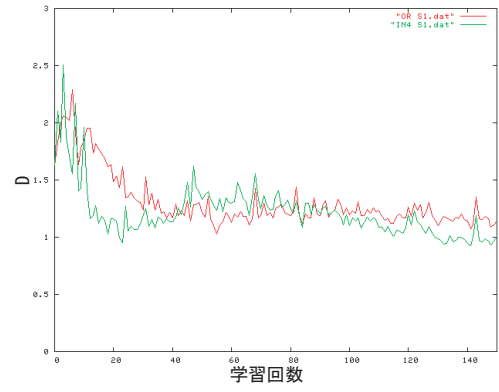


図 4.14: D(主語 boy)

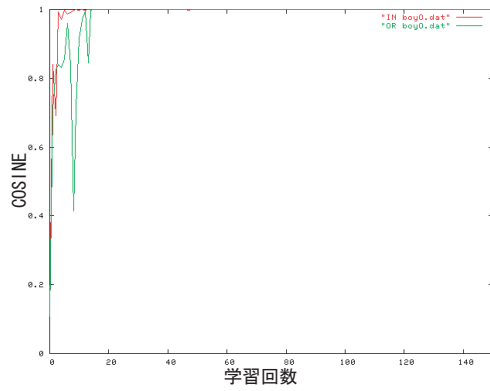


図 4.15: COSINE(目的語 boy)

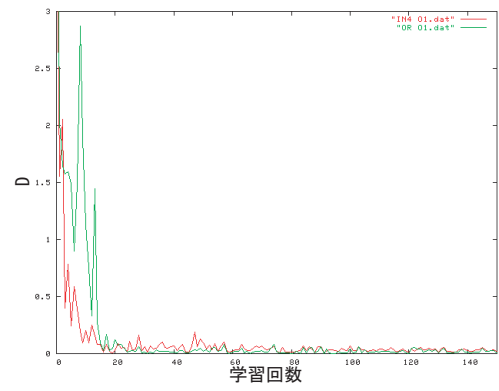


図 4.16: D(目的語 boy)

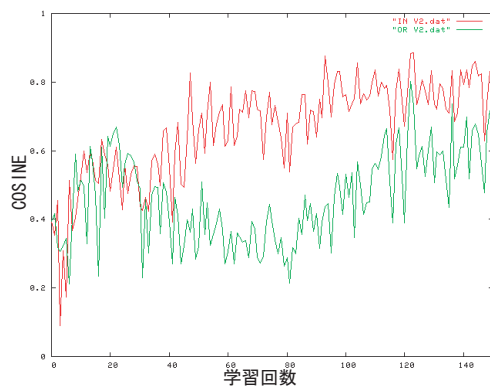


図 4.17: COSINE(chase)

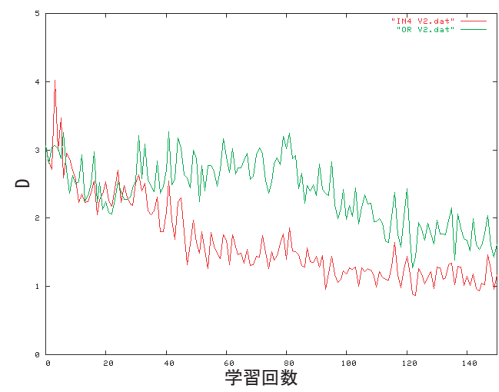


図 4.18: D(chase)

## 4.5 提案学習法への慣性項の追加

慣性項を考慮した学習法は局所解からの脱出, 学習速度の上昇など, 収束性の向上を目的としたものである. 前節で行った実験からその有効性が確認されたため, 本節では文脈層の平均化を防ぐための提案学習法に慣性項を追加し, さらなる改善がなされるかを確認する.

### 4.5.1 慣性項の追加

以下のように慣性項を追加する.

$$\begin{aligned}\Delta w_{jk}(t) &= -\epsilon \delta_k y_j + a \Delta w_{jk}(t-1) \\ \Delta w_{ij}(t) &= -\epsilon \delta_j x_i + a \Delta w_{ij}(t-1) \\ \Delta w_{lj}(t) &= -\epsilon \alpha^{(1-(H_j(t)-C_j(t))^2)} \delta_j c_l + a \Delta w_{lj}(t-1)\end{aligned}\tag{4.9}$$

$\delta_k, \delta_j$  は次式によって算出する.

$$\begin{aligned}\delta_k &= (z_k - d_k) z_k (1 - z_k) \\ \delta_j &= \sum_k \delta_k w_{jk}\end{aligned}\tag{4.10}$$

次節では上記の慣性項を追加した提案学習法を使用し, その有効性を確かめるための実験を行う.

### 4.5.2 実験による評価

慣性項を考慮しない提案学習法, 慣性項を考慮した学習法の二つと慣性項を追加した提案学習法とを比較する. 実験には前節と同様のモデルを使用した. 学習率  $\epsilon$  は 0.1,  $\alpha$  は 4.0,  $a$  は 0.7 とした. これらは実験的に決定した値である. 結果を図 4.19 から図 4.23 に示す. 図中それぞれ赤線が慣性項を追加した提案学習法, 緑線が慣性項を考慮したバックプロパゲーション法, 水色線が慣性項の無い提案学習法に対応している.

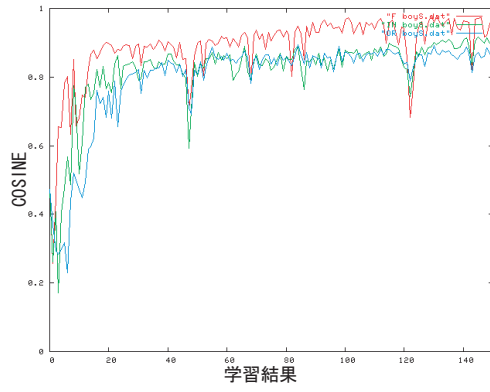


図 4.19: COSINE(主語 boy)

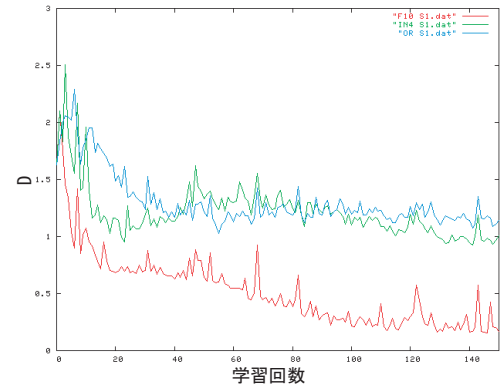


図 4.20: D(主語 boy)

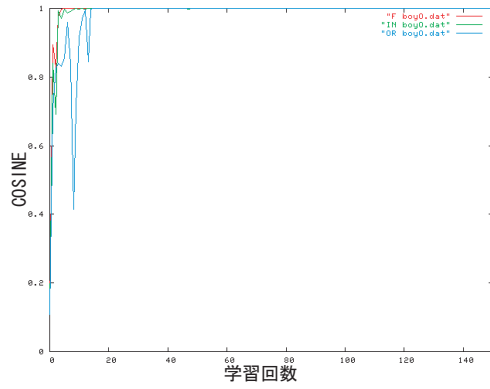


図 4.21: COSINE(目的語 boy)

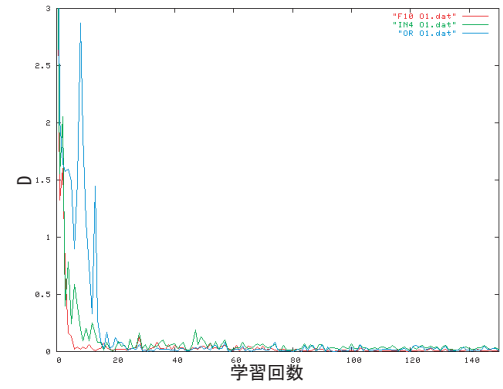


図 4.22: D(目的語 boy)

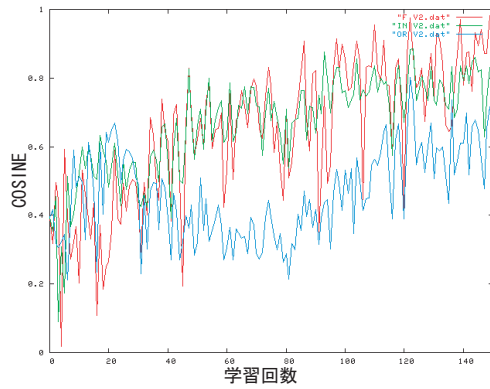


図 4.23: COSINE(chase)

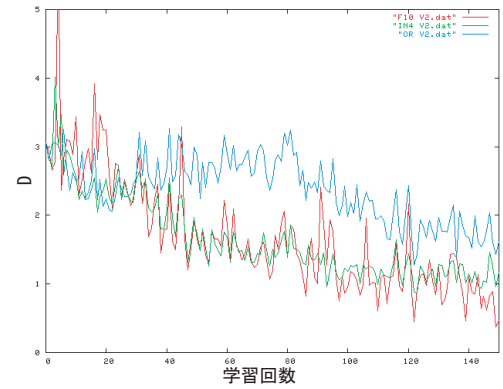


図 4.24: D(chase)

クラスター分析による比較

慣性項を考慮したバックプロパゲーション法は平均化の問題の解決を目的とした方法ではない。前節で示した結果では提案手法に慣性項を追加することによってさらなる改善がなされることを確認したが、これは平均化の問題の改善によって実現されたものと考えられる。そこで、本節ではクラスター分析によって文脈層パターンの類似性を慣性項を考慮したバックプロパゲーション法と慣性項を追加した提案学習法とで比較を行う。150000回学習時点での結果をそれぞれ図4.25、図4.26に示す。慣性項のみの場合に比べ、提案学習法の場合では確かにクラスター間距離が大きく

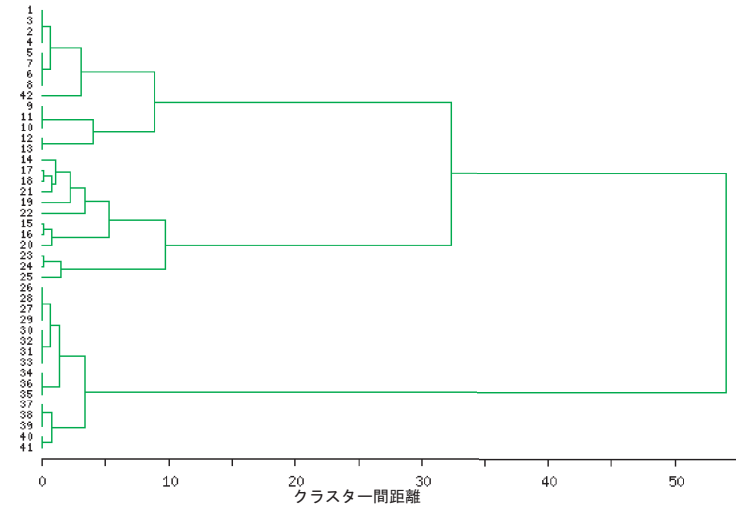


図 4.25: 慣性項のみの学習法

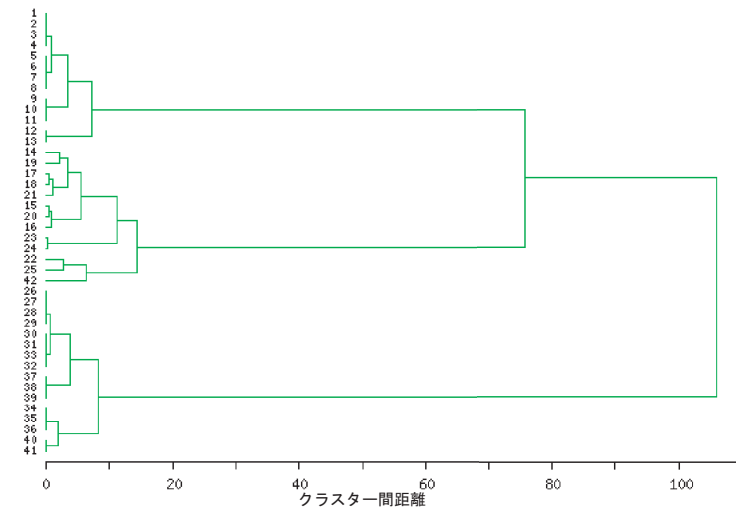


図 4.26: 慣性項を追加した提案学習法

なっていることが分かる。今回は150000回学習時の結果のみ載せたが、提案学習法では同じクラスター同士でまとまるのがより学習初期の段階からおこっており、学習速度の上昇を反映した結果

も得られている。

以上より提案学習法が本モデルにおいて有効であることを確認した。次節では提案学習法が本モデルだけに特化した学習法ではないかを確認するために新たに実験を行い評価する。

## 4.6 追加実験

本節では提案学習法が本研究で使用したモデル以外に対しても有効かを調査するため、簡単なモデルを使用した実験を行い評価をする。

### 4.6.1 どのような問題を扱うか

本研究で使用したモデルの性質の一つとして、同じ入力でも文脈(過去の入力)に応じて出力を変化させる必要がある、という点が挙げられる。今回使用した入力系列は2語、または3語からなる文が連結されたものであり、文と文の連結はランダムに行っており、何らの依存関係もないため目標の確率を出力には一時点前の情報のみで十分である。Elman は後に、単純再帰ネットワークを使用したモデルで関係詞節を含んだより複雑な文を学習させている。関係詞節を含んだ文の一例として、中央はめ込み型の文が挙げられる(図 4.27)。この文の特徴としては、人称による動詞の変

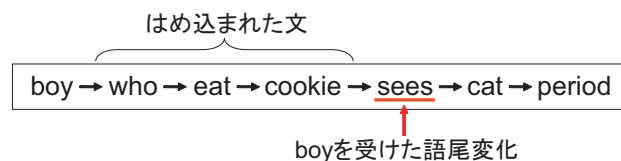


図 4.27: 中央はめ込み文の一例

化(例: boy eats の”s”)が考慮されており、中央はめ込み型の文において最後の単語を正しく予測するためには一時点前以上過去の情報が必要とされる。これは、単純再帰ネットワークがどれだけ過去の情報(何時点前までの入力)を考慮出来るか、という問題だと言える。そこで本研究で提案した学習法が、より過去の情報を考慮した出力を可能にするかどうかを調査することとする。

どれだけ過去の情報を扱えるかに関連する研究として Stephan らの研究が挙げられる [15]。Stephan らは自然言語が文脈依存言語と文脈自由言語の中間に属することを述べた後、自然言語の持つ特徴を反映した簡単な記号列を単純再帰ネットワークがどれだけ学習出来るかを調査している。使用している記号は a, b, c の三つであり、aaabbbccc などのように a と b と c が同数続く(一つの記号が続く数を「深さ」と呼ぶ)記号系列を扱っている。Stephan らは、この記号系列をより少ない学習回数でより深さのある記号系列を学習するための学習法の提案や単純再帰ネットワークの構造の改良などを行っている。

Stephan らが扱った記号系列で特徴的なのは、a が複数回連続して続いた後に b を予測する必要がある、という点である。同じ a の入力でもその回数に応じて異なる文脈層パターンをとらなければ a から b へ切り替わる部分での b の予測は困難になるものと考えられる。提案学習法が本研究

で扱ったモデルでたまたま動作したのではなく、文脈層の平均化を防ぐ効果があるならば、より深さのある記号系列を学習可能だと考えられる。

そこで今回は a が複数回続き、その後に b が続く記号系列を予測学習させる実験を行う。提案学習法がどれだけの深さの系列まで学習可能かを調査し、通常のバックプロパゲーション法、慣性項を考慮したバックプロパゲーション法との比較も行う。

#### 4.6.2 実験

##### 実験条件

使用する単純再帰ネットワークは図 4.28 の構成のものを使用する。

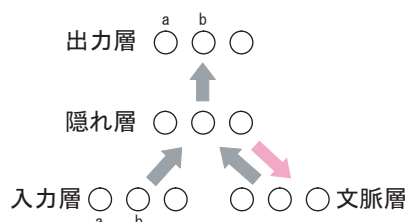


図 4.28: 単純再帰ネットワークの構成

記号 a を  $(1, 0, 0)$ 、記号 b を  $(0, 1, 0)$  として表す。a が深さの分だけ続いた後に b が現れるという記号系列を 1 セットとし、それを多数連結したものを入力系列とする。予測学習であるので教師系列はこれまで行ってきた実験と同様、次の時点での入力教師となる。Stephan らは様々な深さの記号系列のセットを混ぜたものを入力系列としていたが、ここでは一つの入力系列には一つの深さの記号系列のセットのみということにした。例として図 4.29 に深さ 3 の場合の入力系列を示す。

入力系列

a → a → a → b → a → a → a → b → …

図 4.29: 深さ 3 の場合の入力系列

1000 回学習するごとに正しく予測が出来るかを評価する。ここで言う 1 回とは 1 つの記号が入力されて学習することとする。扱う入力系列には確率的な要素はなく必要な過去の情報さえあれば決定的に次の記号を予測可能であるので確率分布としての評価ではなく、どの記号が選択されたかに注目する。出力層の a を表すノードと b を表すノードを比較し、より大きい値を取った方をネットワークが選択した記号と考える。次の時点で入力される記号が選択されていれば予測が正しく出来ていることになる。学習用より短い記号数のテスト用の入力系列を用意し (およそ 10000 個の記号) 学習後のネットワークに入力し、正しく予測出来た場合に予測可能であると評価する。今回はおよそ 400000 個の記号からなる系列を入力系列として 400000 回まで学習させ、それでも正しく予測が出来なかった場合にはその深さの記号系列は学習不可能であると判断する。

## 実験結果

各学習法での学習法ごとに、どれだけの深さの記号系列まで学習可能であったかを示す(表 4.1). パラメータ  $\epsilon$ ,  $\alpha$ ,  $a$  はいくつかの値を試し、最も良かった場合の結果を表 4.1 に示してある.

表 4.1: 各学習法との比較

学習法	学習可能な深さ	最も大きい深さで要した学習回数
バックプロパゲーション法	4	13000 回
慣性項を考慮したバックプロパゲーション法	5	11000 回
提案学習法	9	377000 回
慣性項を追加した提案学習法	11	140000 回

提案学習法, 慣性項を追加した提案学習法の二つの場合が他の既存の学習法と比較してより深い記号系列まで学習可能であった. これは提案学習法の直前の隠れ層のパターンとの引き離しの効果が有効に働いたためと考えられる. 最初の  $a$  から  $b$  に切り替わる直前の  $a$  までのそれぞれの  $a$  について, その入力の結果得られる隠れ層のパターンを分離することによって初めて  $b$  の予測が可能となるからである.

今回使用したネットワークの構成では隠れ層(文脈層)のノード数が3つであるため, 各ノードの値を1つの次元として3次元プロットを行い, 先に述べた隠れ層パターンの分離が行われているかを確認した. バックプロパゲーション法でも学習可能であった深さ4の場合と提案学習法で学習可能であった深さ9の場合の結果を示す(図 4.30, 図 4.31). 図中赤線が慣性項を考慮した提案学習法, 黒線が提案学習法, 紫線が慣性項を考慮したバックプロパゲーション法, 青線が通常のプロパゲーション法に対応している. どちらの場合も提案学習法ではより広く隠れ層の3次元の空

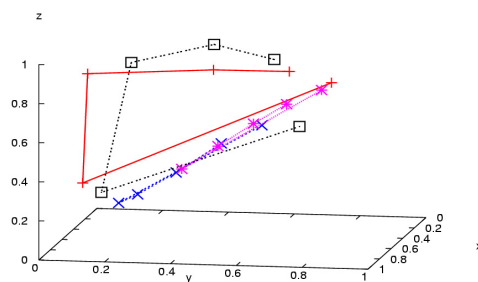


図 4.30: 隠れ層の状態(深さ 4)

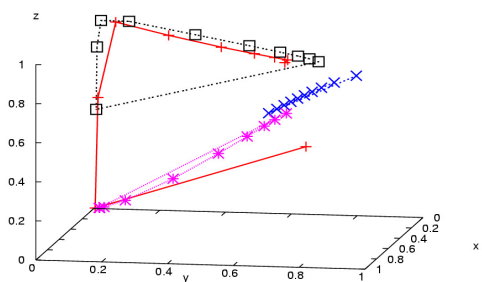


図 4.31: 隠れ層の状態(深さ 9)

間を使用していることが確認出来る. また, 通常のプロパゲーション法では他の学習法の場合に比べて深さ9の場合に9つの  $a$  で取るパターン同士が近くなっていることが確認出来る. この場合のネットワークは常に  $a$  を選択するようになっており, 9つのパターンを明確に分離出来ないことによる予測結果の悪化を反映している. これは平均化の問題そのものだと考えられる. 一



方慣性項を考慮したバックプロパゲーション法では隠れ層の空間を広く使用しているように見える。しかし、 $a$ から $b$ に切り替わる付近でのパターンに注目すると、互いに類似したパターンを取っていることが分かる。この場合のネットワークは、 $b$ を全く選択しなくなるか、まだ $a$ を予測すべきところで $b$ を選択する(例えば、aaaaaaaaabではなくaaaaaabbbを予測する)様になる。二つの提案学習法ではそれぞれ8つ目の $a$ と9つ目の $a$ に対応するパターンが大きく分離されていることが確認出来る。正しく $b$ を予測するという動作を可能にしているのはこのパターンの分離が行われているためと考えられる。

#### 4.6.3 考察

実験により、提案学習法を使用することにより過去の情報を考慮出来るようになることを確認した。隠れ層に注目した評価により、提案学習法によってより広く隠れ層の空間を使用し、各時点でのパターンを分離していることを確認した。

本節での実験結果により、提案学習法が今回扱ったモデルでのみ有効に働くのではなく、過去の情報を必要とするような課題において有効に働くということが主張出来る。より過去の情報を必要とする、後のElmanの関係詞節を含む文の学習においても提案学習法は有効に働くものと考えられる。

#### 4.6.4 残された課題

前節で提案学習法によってより過去の情報を考慮する必要がある課題を学習出来ることが確認された。しかし、今回扱った課題により新たな問題点が発見された。その問題点とは、学習回数によって結果が大きく変わってしまうこととパラメータによってかなり結果が変わってくるという2点である。

前者について説明するために学習結果に対する予測精度の推移を示す。図4.32は慣性項を考慮した提案学習法において、深さ11の入力系列を学習した場合である。100000回~150000回までの学習回数に対する予測精度の推移が示してある。図中赤線が全体の予測精度、青線が $b$ の予測精度である。およそ10000のテスト用の系列において、全ての時点の入力に対して次の時点で入力される記号が出力層で選択され、正しい予測が出来ていれば予測精度は1の値をとる。

図より、少し学習回数が変わっただけで予測精度が大きく変わることが分かる。これは学習結果が安定しないことを意味しており、大きな問題であると言える。実際、本研究で使用したElmanのモデルでも評価値が大きく振動しているものがあり(結果は付録を参照のこと)、学習の安定性という観点からは問題があると言わざるを得ない。

また、記号系列の学習が可能であるかどうかはパラメータ $\epsilon$ ,  $\alpha$ ,  $a$ の値に大きく依存していることも実験を通して確認された。記号の深さによって適切なパラメータが変化していたため、これも大きな問題点だと言える。

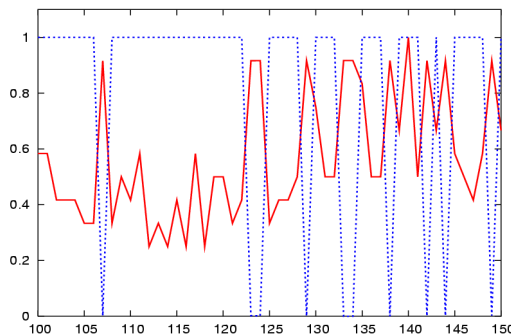


図 4.32: 深さ 11 の場合における予測精度の推移

## 4.7 まとめ

本節では確率を正確に出力することを目標とし、文脈層パターンの平均化の問題を改善することで目標の達成を目指した。提案学習法はネットワークに新たな機能を付加させるなどの恣意的な変更を加えるような学習法ではなく、ネットワーク自体が持つ値を利用した学習法である。実験の結果から提案手法の有効性を確認し、さらに慣性項を追加することによってさらなる改善がなされることを確認した。

単純再帰ネットワークを使用した言語獲得のモデル研究では確率を出力するようになることを一つの目標としている。その際、モデル自体に不足する点があるからではなく、学習能力が十分に引き出せていないために結果が悪くなることが考えられる。そのため、ネットワークの学習能力を向上し、モデル自体に見られる不足点等を明確にするためにも提案手法は有効であると言える。

今回問題として扱ったのは Elman(1990) のモデルのみであり、提案手法の有効性がこのモデルに特化したものではないかという懸念があった。

そこで Stephan らの研究を参考にした簡単なモデルでの実験を行い、正しい出力のためにより過去の情報を必要とする場合についてもその学習のために提案学習法が有効であることを確認した。この結果から、後に Elman が行った関係詞節を含むより複雑な文の学習のためにも提案学習法は有効に働くものと思われる。

# 第5章

## 結論

### 5.1 本論文の成果

本論文では, Elman の 1990 年のモデルを基に大きく分けて二つの研究について説明した. 一つは系列の区切りの情報の追加がネットワークにどのような影響を及ぼすかの検討, もう一つは単純再帰ネットワークにおける文脈層パターンの平均化の問題についての検討とそれを防ぐ新学習法の提案である.

系列の区切りの情報の追加によって, ネットワークは文脈情報をより考慮するようになった. 具体的には, 文脈層に保持された過去の情報により, 同じ名詞の単語が入力された場合でも主語としての出力と目的語としての出力とで大きくその出力を分けて動作するようになった. また文脈層に注目した評価により, 文の区切りを表す period が入力されることによって文脈層のパターンが一定となり, 予測に不必要な情報が消去されるようになることを確認した.

単純再帰ネットワークには文脈層パターンの平均化が起こるという問題があった. これは過去の入力に応じて出力を変える必要がある本研究で使用した課題においては大きな問題であった. 通常のバックプロパゲーション法を基にし, 文脈層パターンと隠れ層パターンが離れるように学習させる新学習法を使用することでこの問題が改善されることを確認した. さらに, 提案学習法に慣性項を付加させることによってさらなる改善を実現した.

### 5.2 課題

提案学習法が本モデルに特化したものではないことを確認するため, 簡単な記号系列による実験は行ったが, さらに他のモデルでも実験を行う必要がある. 具体的には, Elman の 1991 年のモデルや乾らのモデルが考えられる.

また, 従来の学習法でも提案した学習法においても未解決の問題として更新が収束しないということが挙げられる. これは, 各時点ごとで算出される誤差関数が本質的に収束しないようになって

いるためである。簡単な記号系列による実験結果でも確認されたが、これは学習回数のわずかな変化に対して学習結果を不安定にする要因となるものと思われる。

この問題の解決には「印象」の定義によって解決が図れるかもしれない、ネットワーク内のパラメータにより「印象値」なるものを定義して、何度もネットワークが学習した文脈についてはその結合強度の更新量を弱めるようにするというのが基本的な考えである。これは新奇の文や単語の学習の際に更新量が大きくなり、相対的に他の既知の文より学習が進むといったことも期待できる。ネットワークの構造や恣意的なパラメータの導入などを行わずに「印象」が定義出来たらさらに良い研究になると考えられる。

Elman はより複雑な、より過去の情報を必要とするような文の学習には文脈層の強制的なリセット（隠れ層ノードの値を強制的に一定値 2 にする）、文脈層ノードの個数を学習回数に対して徐々に増加させる一種の構造学習が有効に働くと述べ、これらを乳児の脳の発達と関連付けて議論している。しかし、前者については文の生成規則に意味を反映した条件を付加させることによって必要が無くなるという報告もある (Rohde, 1999)。いずれにせよ、ネットワークに新たな機能を付加させる前に学習方法自体について見直す必要がある。誤差関数や学習課題を改良することで、研究者がネットワークの機能として付加させたいような機能を自律的に発現させられたら面白いと思う。高次の認知過程である言語において、モデル研究に求められるものはおそらく言語学等における既存の知見との精緻な対応付けではない。言語においては意味、統語、音韻に分けられて研究されているが、これらが明確には分離して議論が可能なものとは限らない。簡単なネットワーク構成、簡単な学習規則、簡単な入力によって、これら 3 つに対応するような知識が獲得されるようなモデルが出来れば逆に言語をどのような要素に分類して研究すべきかも議論可能になるのではないだろうか。既存の知見との精緻な議論の前に、きっかけは適当でも言語のようなものが発現するモデルを作る事やその方法を考える事がモデル研究に求められていることであるように思う。

## 謝辞

本研究を進めるにあたり、多大な御指導とともにこの研究の機会を与えて下さった東北大学大学院工学研究科 阿曾弘具教授に心より感謝致します。

本研究について貴重な御意見、御指導を賜りました東北大学大学院工学研究科 吉澤誠教授、東北大学大学院情報科学研究科 中尾光之教授に深く感謝致します。

研究室ゼミ等を通じ、本研究についての御意見、御指導を多数賜りました東北大学大学院工学研究科大町真一郎助教授に深く感謝致します。

本研究全般に渡り専門的な御意見を頂いた東北大学大学院工学研究科三宅章吾助教授に心より感謝致します。

研究を進めるにあたり専門的かつ具体的な御助言、御意見を頂いた東北福祉大学 佐藤俊治講師に深く感謝致します。

常日頃から研究についての御意見、御指導を多数頂き、この研究の方針を与えて下さった、東北大学大学院工学研究科 菅谷至寛助手に深く感謝致します。

また、東北大学大学院工学研究科 電気・通信工学専攻 阿曾研究室の皆様には日頃から研究に限らず多岐に渡ってお世話になりました。深く感謝致します。

## 参考文献

- [1] J. L. Elman. : “Finding Structure in Time,” *Cognitive Science*, vol.14,pp.179-211,1990.
- [2] J. L. Elman. : “Distributed representations, simple recurrent networks, and grammatical structure,” *Machine Learning*, vol.7,pp.195-224,1991.
- [3] J. L. Elman. : “Learning and development in neural networks: the importance of starting small,” *Cognition*, vol.48,pp.71-99,1993.
- [4] 森田昌彦, 村上聡. : “非単調神経回路網による時系列パターンの認識,” 電子情報通信学会論文誌 (*D-II*), vol.J81-D-II,pp.1679-1688,1998.
- [5] 森田昌彦, 松沢浩平, 諸上茂光. : “非単調神経素子の選択的不感化を用いた文脈依存的連想モデル” 電子情報通信学会論文誌 (*D-II*), vol.J85-D-II,pp.1602-1612,2002.
- [6] 望月彰子, 大森隆司. : “PATON : 文脈依存性を表現する動的神経回路網モデル,” 日本神経回路学会誌, vol.3,pp.81-89,1996.
- [7] 大森隆司, 下斗米貴之. : “文法を利用した語彙獲得における加速現象の脳モデル化,” 認知科学, vol.7,pp.223-235,2000.
- [8] 遠山修治, 下斗米貴之, 大森隆司. : “語彙獲得における即時学習則の適用条件モデル,” 電子情報通信学会 信学技報, NC-2001-91, 2002.
- [9] 玉森彩弥香, 乾敏郎. : “Elman ネットによる統語範疇の配列と格関係の学習,” 認知科学 vol.6,pp.359-368,1999.
- [10] 波多野誼余夫. : “Connectionist infants は統語規則を獲得しうるか,” 心理学評論, vol.40,pp.319-327,1997.
- [11] J. L. Elman. E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. : *Rethinking Innateness.*, MIT Press, Cambridge, 1996. 乾敏郎, 今井むつみ, 山下博志 (訳)(1998). 認知発達と生得性, 東京 : 共立出版.
- [12] 馬場則男, 小島史男, 小澤誠一 : “ニューラルネットの基礎と応用,” 1994, 東京 : 共立出版.

- [13] 守一雄, 都築誉史, 楠見孝: “コネクショニストモデルと心理学,” 北大路書房, 2001.
- [14] Douglas L.T. Rohde, David C. Plaut. : “Language acquisition in the absence of explicit negative evidence: important is starting small?, ” *Cognition*, Vol.72,pp.67-109,1999.
- [15] Stephan K. Chalup, Alan D. Blair. : “Incremental training of first order recurrent neural networks to predict a context-sensitive language, ” *Neural Networks*, Vol.16,pp.955-972,2003.
- [16] Robert A. Jacobs. : “Increased rates of convergence through learning rate adaptation, ” *Neural Networks*, Vol.1, pp.295-307,1988.

# 付録 A

## A.1 評価値 (COSINE, ダイバージェンス)

第4章で行った実験で得られた COSINE, ダイバージェンス  $D$  の学習回数に対する推移を以下に全て示す. まず始めに名詞の単語が入力された場合の結果について示す. 本研究では名詞はに属する単語は主語と目的語の二つに分けられる (表 A.1). 主語の場合は動詞が続き, 目的語の場合には period が続くという違いがある. 主語については4つのカテゴリーによって次に続く動詞のカテゴリーの確率が異なるため目標となる確率分布 (規則から算出される確率分布) が異なる. まず表 A.1 の9つの場合それぞれについて, 評価値の推移を示す (図 A.1 から図 A.18). 図中青線が通常のバックプロパゲーション法, 茶線が提案学習法, 紫線が慣性項を考慮したバックプロパゲーション法, 赤線が慣性項を追加した提案学習法に対応する.

表 A.1: 主語となるカテゴリーと目的語となるカテゴリー

主語	目的語
noun_human	noun_human
noun_animal	noun_animal
noun_inanimal	noun_inanimal
noun_agress	noun_fragile
	noun_food



## A.1.1 主語

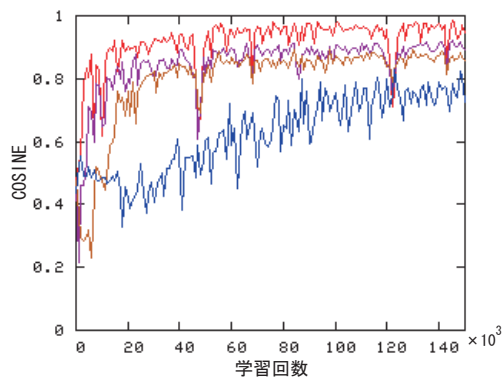


図 A.1: COSINE(主語 noun\_human)

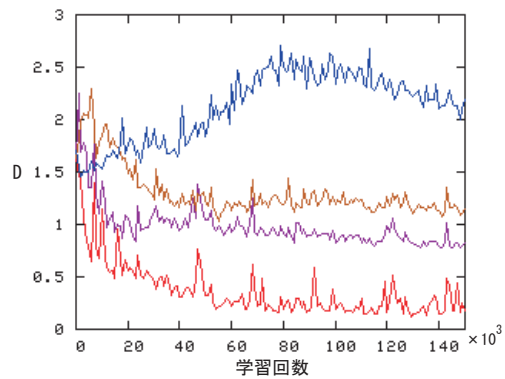


図 A.2: D(主語 noun\_human)

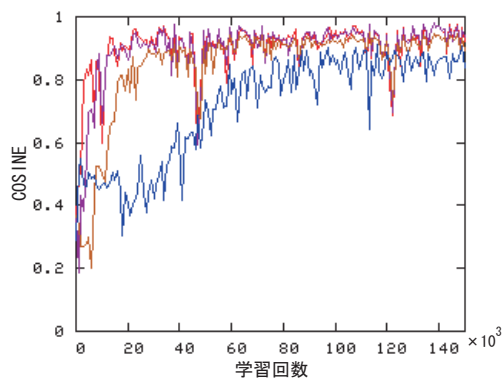


図 A.3: COSINE(主語 noun\_animal)

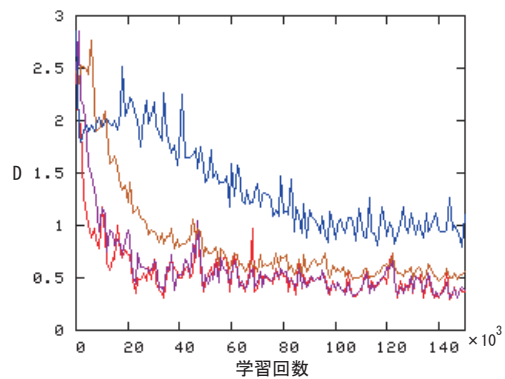


図 A.4: D(主語 noun\_animal)

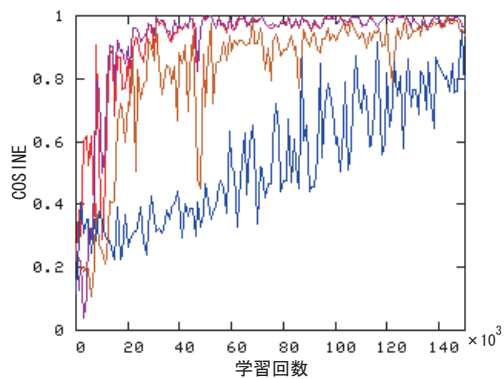


図 A.5: COSINE(主語 noun\_inanimal)

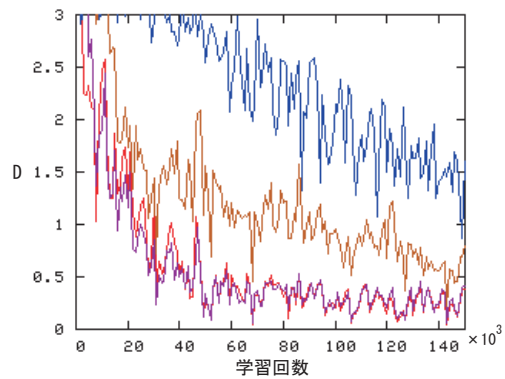


図 A.6: D(主語 noun\_inanimal)

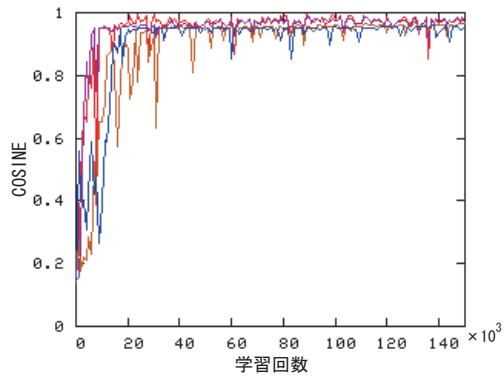


図 A.7: COSINE(主語 noun\_agpat)

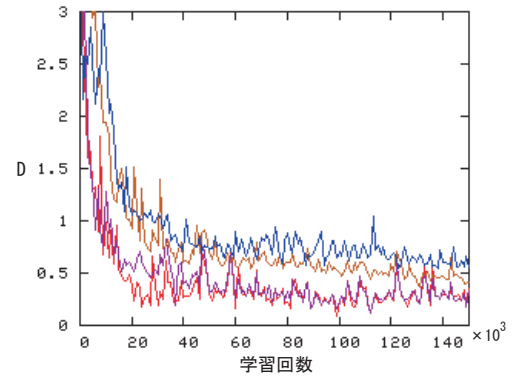


図 A.8: D(主語 noun\_agpat)

### A.1.2 目的語

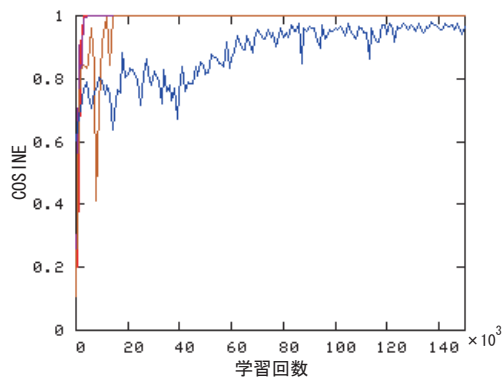


図 A.9: COSINE(目的語 noun\_human)

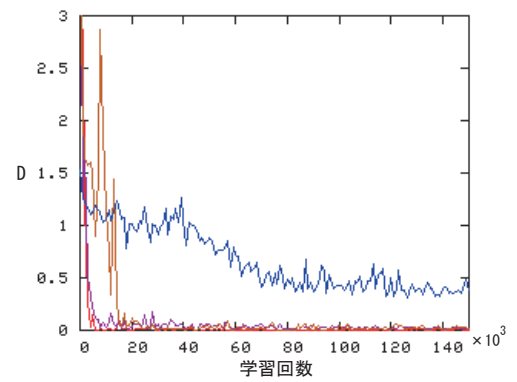


図 A.10: D(目的語 noun\_human)

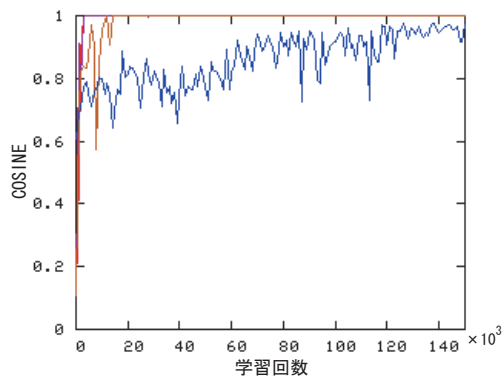


図 A.11: COSINE(目的語 noun\_animal)

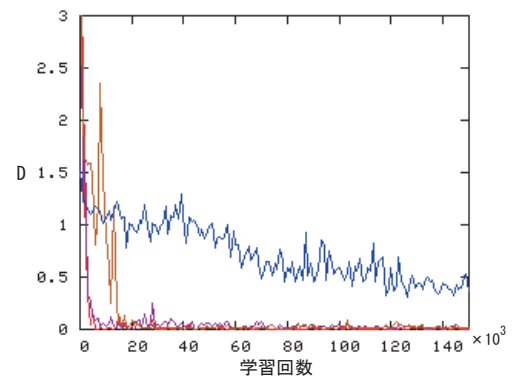
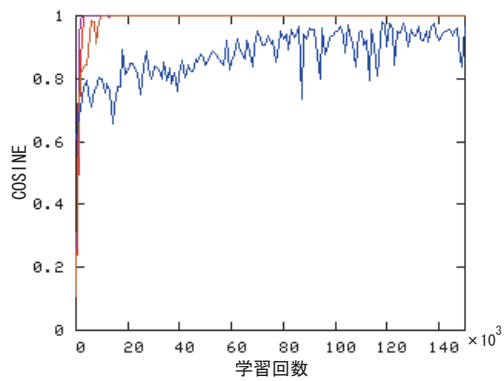
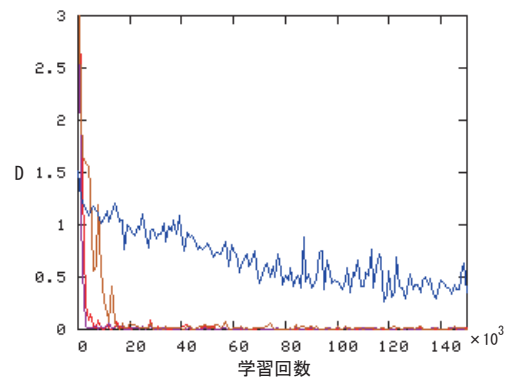


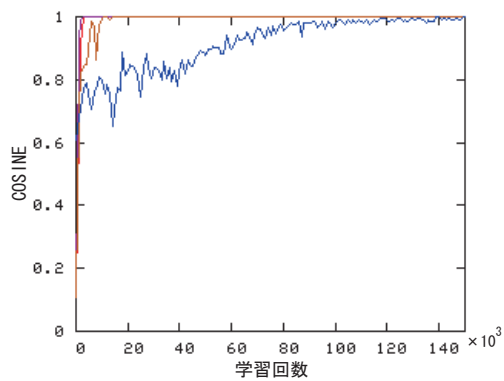
図 A.12: D(目的語 noun\_animal)



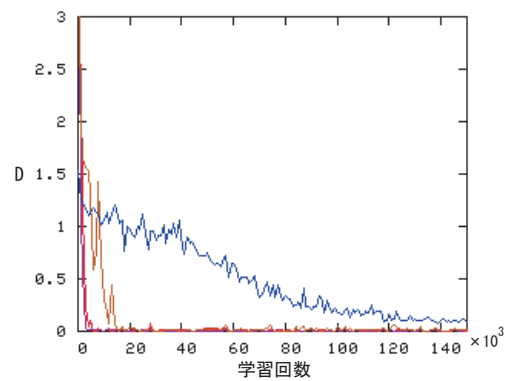
☒ A.13: COSINE(目的語 noun\_inanimal)



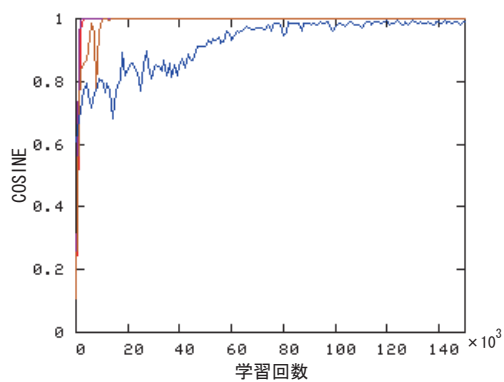
☒ A.14: D(目的語 noun\_inanimal)



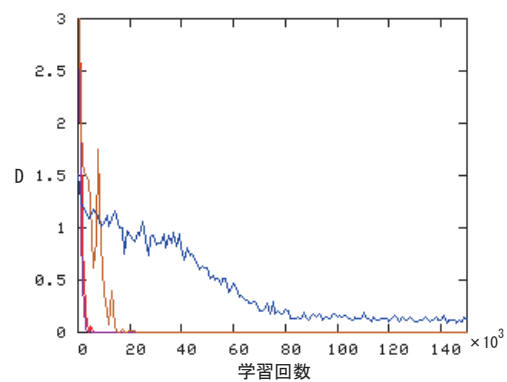
☒ A.15: COSINE(目的語 noun\_fragile)



☒ A.16: D(目的語 noun\_fragile)



☒ A.17: COSINE(目的語 noun\_food)



☒ A.18: D(目的語 noun\_food)

次に動詞の単語が入力された場合の結果を示す。動詞に属するカテゴリーは次に目的語が period が続く。動詞の場合は、同じカテゴリーでも直前の主語によって目標となる確率分布が異なる。本研究ではこれらを分けて評価した。具体的には表 A.2 に示すように 12 通りの場合に分けて評価した。それぞれについての COSINE と D の結果を示す (表中のインデックスが図の題の括弧内の数字と対応している)。

表 A.2: 動詞の入力

インデックス	直前の入力	動詞
1	noun_human	verb_intran
2	noun_human	verb_tran
3	noun_animal	verb_tran
4	noun_human	verb_agpat
5	noun_animal	verb_agpat
6	noun_inanimal	verb_agpat
7	verb_human	verb_percept
8	verb_human	verb_destroy
9	verb_agress	verb_destroy
10	verb_human	verb_eat
11	verb_animal	verb_eat
12	verb_agress	verb_eat

### A.1.3 動詞

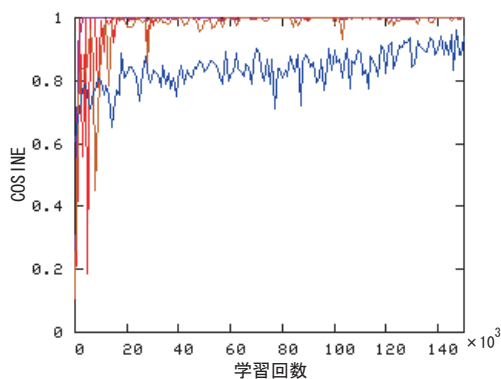


図 A.19: COSINE(1)

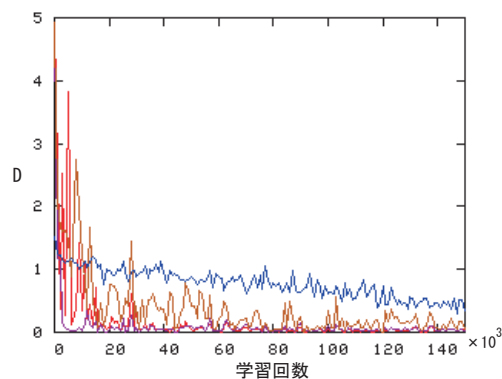
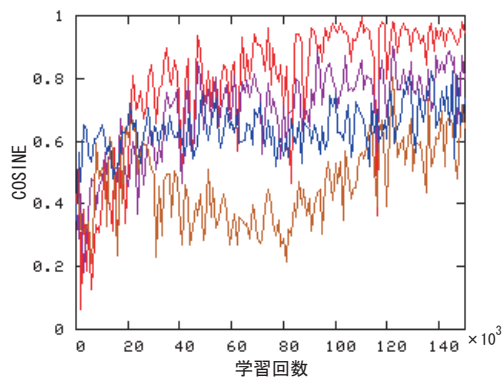
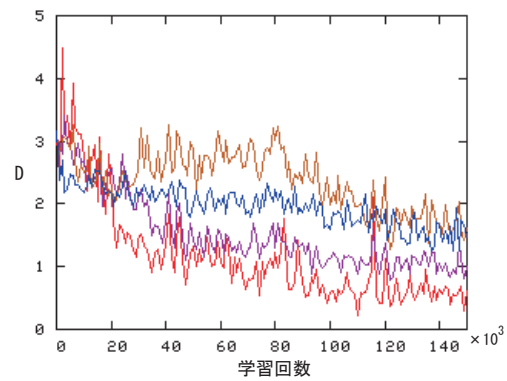


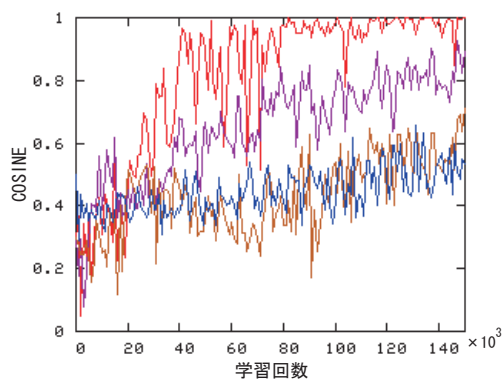
図 A.20: D(1)



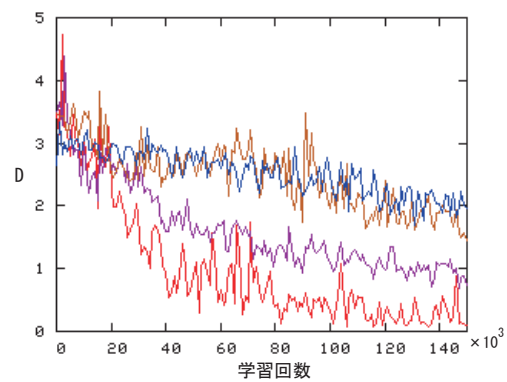
☒ A.21: COSINE(2)



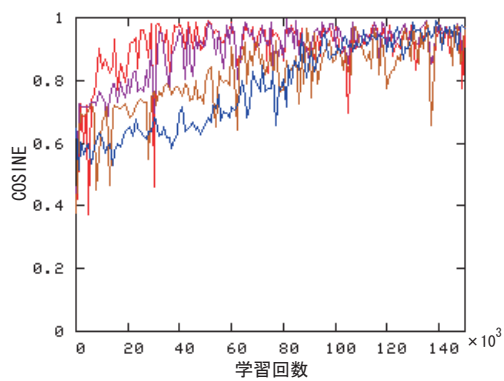
☒ A.22: D(2)



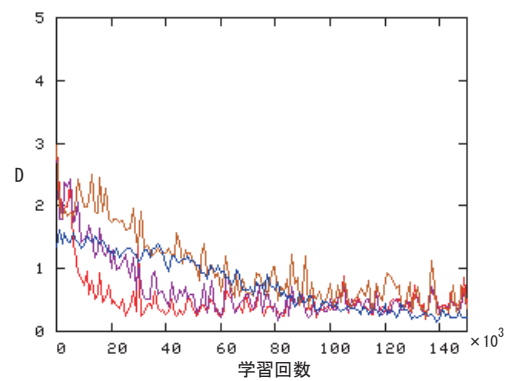
☒ A.23: COSINE(3)



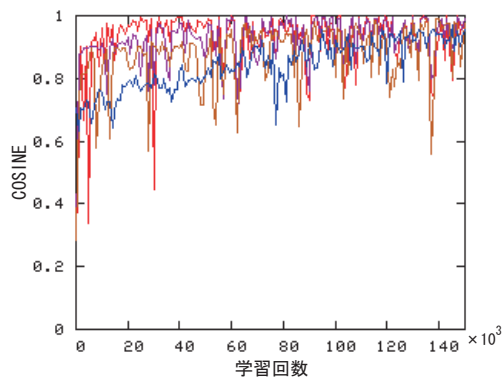
☒ A.24: D(3)



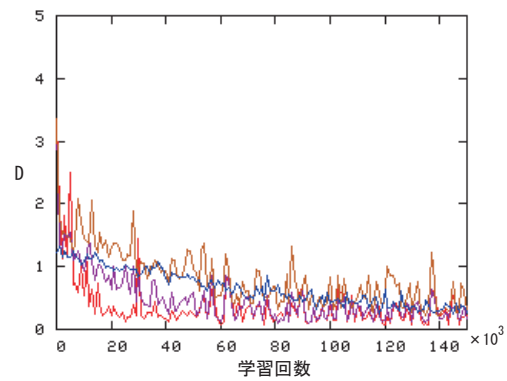
☒ A.25: COSINE(4)



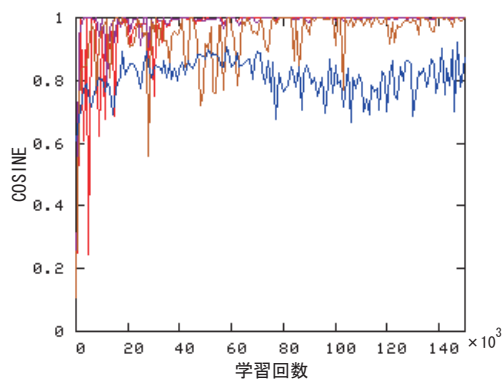
☒ A.26: D(4)



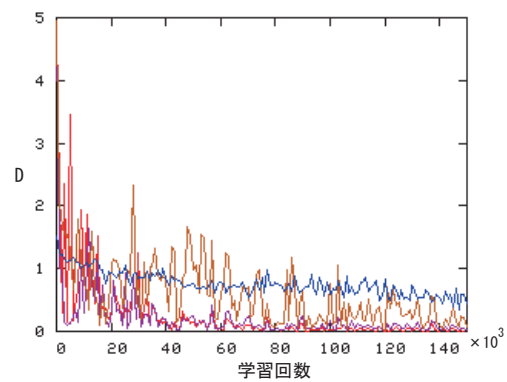
☒ A.27: COSINE(5)



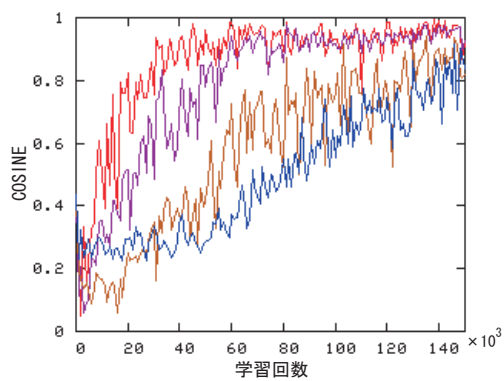
☒ A.28: D(5)



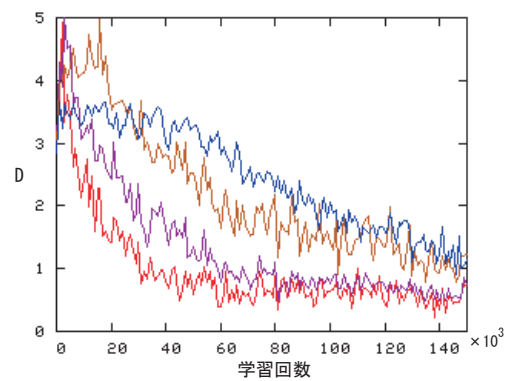
☒ A.29: COSINE(6)



☒ A.30: D(6)



☒ A.31: COSINE(7)



☒ A.32: D(7)

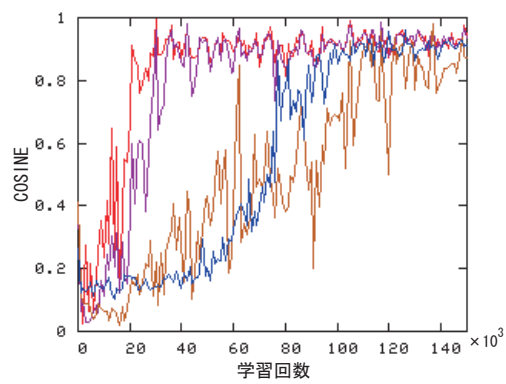


图 A.33: COSINE(8)

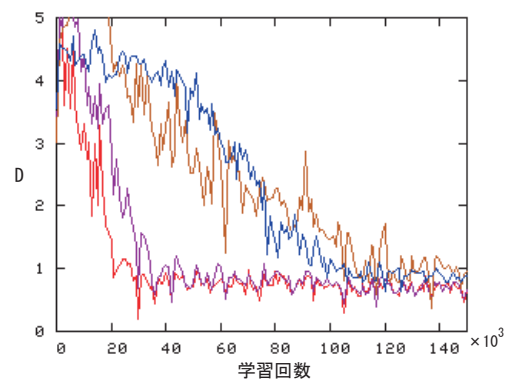


图 A.34: D(8)

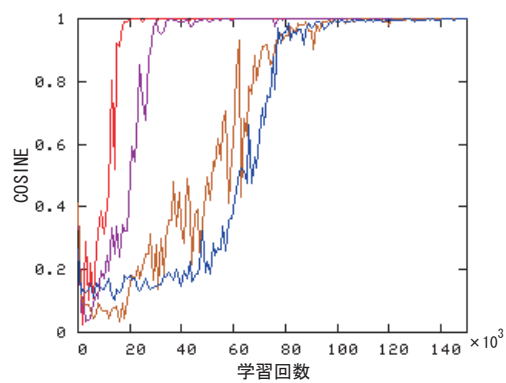


图 A.35: COSINE(9)

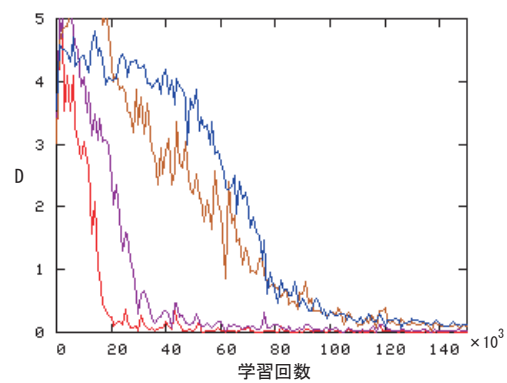


图 A.36: D(9)

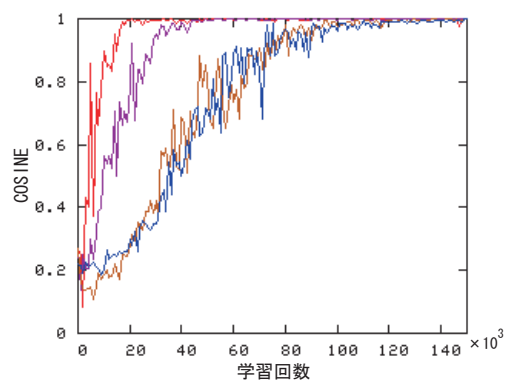


图 A.37: COSINE(10)

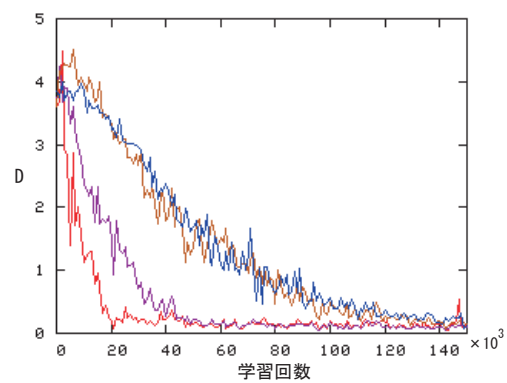


图 A.38: D(10)

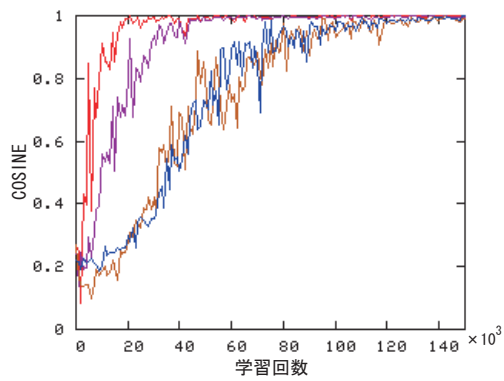


図 A.39: COSINE(11)

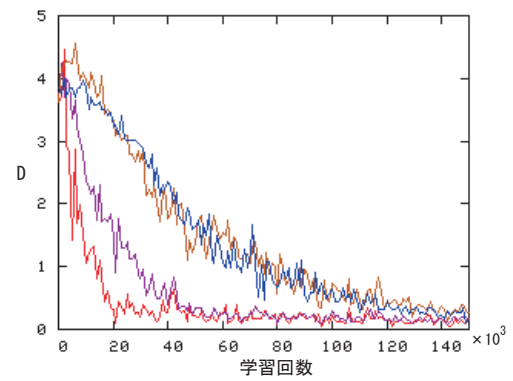


図 A.40: D(11)

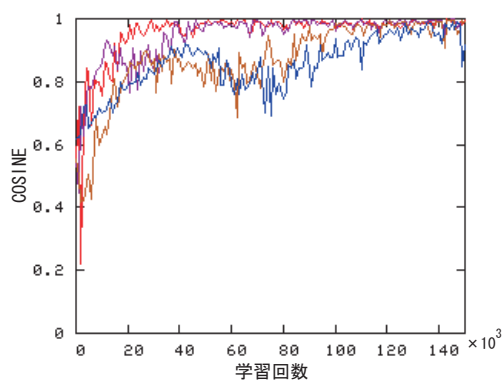


図 A.41: COSINE(12)

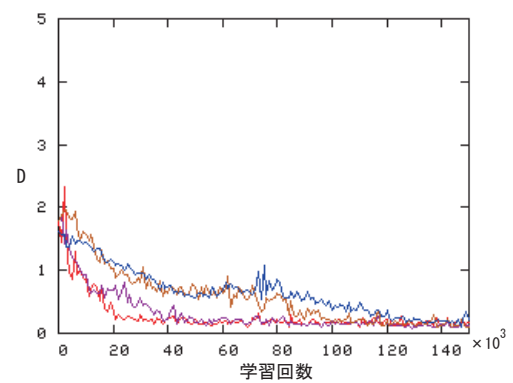


図 A.42: D(12)