

逐次学習型時系列予測モデル

郷古 学^{†a)} 菅谷 至寛[†] 阿曾 弘具[†]

Time Series Prediction Model for Sequential Learning

Manabu GOUKO^{†a)}, Yoshihiro SUGAYA[†], and Hirotomoto ASO[†]

あらまし 本論文では、記憶容量制限を考慮した逐次学習型の時系列予測モデルとして、適応的逐次学習ネットワークモデル (Adaptive and Sequential Learning Network model: ASLN model) を提案する。提案モデルは入力として与えられる時系列情報を逐次的に記憶し、その記憶を用いて予測を行う。モデルは周囲の環境の変化に対し、記憶容量の有効利用を図りながら、自らの記憶を変化させることで追従していくことが可能である。モデルは入力時系列に含まれる要素と、その遷移に関する情報を記憶する。モデルは時系列の要素の中でも入力頻度の高いものを重要と判断し、優先的に記憶する。また入力される要素の遷移に関する情報を状態ベクトルという形で表現し、階層型ニューラルネットワークにより記憶する。状態ベクトルは過去の入力系列の情報を保持しており、文脈を表現することが可能である。計算機実験により、提案モデルが周囲の環境の変化に対し追従しつつ時系列を予測することを示す。また、ばらつきを含む手書き数字パターンを用いた数字列の学習実験を行い、パターン数を増やした場合の予測能力の検証を行った。更に、モデルが時系列の遷移情報を記憶する際に重要なパラメータについて、その設定指針を与えた。

キーワード 時系列予測, 逐次学習, 記憶容量, ニューラルネットワーク

1. ま え が き

複雑な思考の実現や忘却の防止、情報の共有化を目的とし、人間の記憶をサポートするシステムに関する研究が注目されている [1], [2]。このようなシステムでは、ユーザは記憶しておきたい情報を選択し、記憶モデルに登録（記憶）しておく。システムは外部からの入力に応じて、自動的に記憶情報を想起することでユーザをサポートする。しかし、扱う情報量が多いほど、ユーザが情報の選択に要する負荷も増大する。

近年、記憶すべき情報の選択をモデル自身が自律的に行う記憶モデルが提案されている [3] ~ [6]。そのような記憶モデルを応用し、次々と新しく入力される画像などの情報（パターン）を、ユーザの選択なしに記憶し、利用する記憶システムの構築が進められている [2]。動的に変化する環境からの情報を記憶・利用する場合、モデルの記憶容量（記憶可能なパターン数）

の制限を考慮する必要がある。すなわち、入力される膨大なパターンの中からユーザに必要なパターンと、それ以外のパターンを自律的に判断し、効率良く情報を記憶することが求められる [2]。

更に、これらのパターンが入力される順番ないし入力系列上の文脈を記憶することができれば、手順や経路のようなパターンの並びで表現される時系列情報を記憶することが可能になる。ここで、文脈とは同一の入力情報からでも、それより過去の入力系列に依存して異なる結果を得るような場合の過去の入力系列を意味する（文脈情報）。記憶した文脈を利用することで、現在の入力パターンをもとに、次に入力されるパターンを予測する時系列予測モデルを構築することができる。本論文の目的は、このような動的に変化する環境からの入力情報を記憶し予測を行う、時系列予測モデルを構築することである。

時系列予測モデルは時系列に含まれる要素とそれら要素の遷移の二つの情報を記憶しておき、各情報を利用することで実現できる。本論文では、時系列要素として環境から決まるパターンを対象とする。動的な環境の変化を考えた場合、記憶も環境の変化に追従する必要がある。このような記憶の実現方法として逐次学

[†] 東北大学大学院工学研究科電気・通信工学専攻, 仙台市
Department of Electrical and Communication Engineering,
Graduate School of Engineering, Tohoku University, 6-6-05
Aoba, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan
a) E-mail: gouko@aso.ecei.tohoku.ac.jp

習がある．逐次学習は入力を与えられるごとに学習（記憶の更新）を行う方法である．しかし，通常のモデルの記憶容量は有限であり，すべての入力情報を次々と学習・記憶し続けることは不可能である．すなわち，記憶すべき情報を選別する機能が必要となる．

パターン情報を扱うことを得意とし，また汎化性の高さや，システムの柔軟性という観点からニューラルネットワークを用いた時系列予測モデルの構築及び応用に関する研究が盛んに行われている [5]～[10]．しかし，Elman の提案したネットワーク（以下，Elman ネットと呼ぶ）[7] をはじめとしたニューラルネットワークを用いた時系列予測モデルの多くは，事前に学習すべき時系列を準備し，オフラインで教師あり学習を行うために，環境の変化への追従ができない．

文献 [9], [10] において提案されたモデルは，入力された時系列情報の教師なし学習が可能であり，逐次学習モデルへの拡張により，環境の変化への追従が可能であると考えられる．しかし，記憶容量の制限を考慮しておらず，容量を上回る数のパターンが入力された場合は学習がうまくいかない．

筆者らは，これまでに記憶容量の制限を考慮した逐次学習型の時系列予測モデルを提案してきた [5], [6]．これらのモデルでは，記憶すべき情報の選別の基準として時系列に含まれるパターンの出現頻度を用い，直前のパターンを記憶することができ，決定的単純マルコフで表される時系列を記憶し，予測できる．

本論文では記憶容量制限を考慮した逐次学習型の時系列予測モデルとして，適応的逐次学習ネットワークモデル (Adaptive and Sequential Learning Network model, 以下 ASLN model) を提案する．本モデルは決定的 N 重マルコフで表現される時系列を学習することが可能であり，使用環境の変化に追従する時系列予測システムを実現している．

以下，2. では提案モデルの構造，動作の定式化を行い，使用するパラメータの設定指針についての検討を行う．3. で計算機実験を行い提案モデルの動作及び性能に関して検証を行う．4. ではまとめを述べる．

2. 適応的逐次学習ネットワークモデル

提案モデルは，入力時系列に含まれるパターンを記憶する記憶層と，それらパターンの遷移情報を記憶する文脈ネットワークをもつ．パターンの記憶は芳澤ら [3] の提案した自己連想記憶モデルを用いており，モデルに入力される時系列情報に含まれるパターンの

ち，入力頻度の高いパターンを優先的に記憶することができる．文脈ネットワークでは記憶層で記憶されたパターンの時系列上での順序（遷移情報）を記憶する．

2.1 構造

提案モデルの構造を図 1 に示す．モデルは m_0 個のニューロンからなる入力層， m_1 個のニューロンからなる記憶層，階層型のニューラルネットワークである文脈ネットワーク及び m_0 個のニューロンからなる出力層により構成される．入力層の各ニューロンは記憶層のすべてのニューロンと全結合している．時刻 t における入力層の i 番目のニューロンと，記憶層の j 番目のニューロンとの結合重みを $W_{ji}^{pa}(t)$ と表す．記憶層の j 番目のニューロンがもつ入力層との結合重みベクトル $W_j^{pa}(t) = (W_{j1}^{pa}(t), W_{j2}^{pa}(t), \dots, W_{jm_0}^{pa}(t))$ は，一つの入力パターンを記憶したものである．したがって，本モデルの記憶容量は m_1 に等しい．文脈ネットワークは状態層，中間層，想起層からなる 3 層の階層型ニューラルネットワークであり，状態層，想起層のニューロン数は記憶層のニューロンと同数の m_1 個，中間層のニューロンの数は m_h 個である．モデルは記憶層から文脈ネットワークの状態層への結合を有し，その結合は記憶層の j 番目のニューロンと状態層の j 番目のニューロン間で 1 対 1 の結合であり，その結合重みはすべて 1 で固定である．この結合は記憶層のニューロンが発火したときのみ状態層ニューロンに影響を与え，発火しない場合は影響を与えないように働く．状態層のニューロン j の状態を $s_j(t)$ と表し，時刻 t における状態ベクトル $s(t) = (s_1(t), s_2(t), \dots, s_{m_1}(t))$ とする．文脈ネットワークは状態層-中間層間の結合重み，中間層-想起層間の結合重みを学習により更新する．想起層のニュー

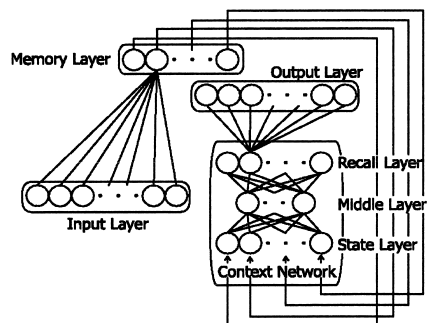


図 1 提案モデルの構造

Fig. 1 Structure of the proposed model.

ロン j から出力層のニューロン i への結合重みは、入力層のニューロン i から記憶層のニューロン j への結合重みに一致させる（記憶パターンの情報を実パターンに変換する機能を実現する）。

提案モデルは入力される時系列を学習する記憶処理と、学習した時系列を用いて、入力をもとに、次に入力されるパターンを出力（予測）する予測処理の二つの処理を同時に行う。

2.2 記憶処理

記憶処理では入力時系列に含まれるパターンを記憶するパターン記憶処理と、遷移情報を記憶する遷移記憶処理の二つの処理が同時に行われる。

2.2.1 パターン記憶処理

入力時系列に含まれるパターンの記憶は、芳澤らが提案した自己連想記憶モデル [3] を用いて行う（パターン記憶処理）。このモデルは入力パターンを正規化し、記憶パターンとの内積によりパターンを同定し、選択ニューロン j_{me} を発火させる処理を行う。記憶パターンとの内積が小さい場合などは、「入力パターンを識別できない」という情報を出力する。これは、 $j_{me} = 0$ の発火として表現する。学習では結合重みを更新するとともに内積値の有効性判定に用いる角度しきい値を更新する。この更新により、信頼度が高い記憶パターンは信頼度が低い記憶パターンと比較して小さな角度しきい値をもつため、新しい入力パターンに対して影響を受けにくく頑健になる。また結合重みを一定量だけ減少させる忘却処理も組み込まれており、入力頻度が低下したパターンの記憶は信頼度の低下という形で失われる。これらの処理の詳細は付録に示す。このモデルは、ある時間間隔 T 以内に含まれているパターンの頻度の高い方から順に m_1 個（記憶層ニューロン数）だけの入力パターンを記憶し、識別する機能をもつ。

2.2.2 遷移記憶処理

文脈ネットワークでは遷移記憶処理を行う。時刻 $t-1$ における状態層の状態ベクトル $s(t-1)$ は時刻 $t-1$ までの入力時系列に関するある情報を保持している。

時刻 t ではまず、状態層-中間層間の重み $W_{kj}^{co}(t-1)$ 、中間層-想起層間の重み $V_{lk}^{co}(t-1)$ の更新を行う。ここで、 $W_{kj}^{co}(t-1)$ は状態層の j 番目のニューロンと、中間層の k 番目のニューロン間の結合重みを、 $V_{lk}^{co}(t-1)$ は中間層の k 番目のニューロンと想起層の l 番目のニューロン間の結合重みを表している。これらの重みを記憶層の選択ニューロンの発火を教師信号とした

バックプロパゲーション法により学習更新する。教師信号を活性ベクトル $a(t) = (a_1(t), a_2(t), \dots, a_{m_1}(t))$ として次のように定義する。

$$a_j(t) = \begin{cases} 1 & \text{if } j = j_{me} \\ 0 & \text{otherwise} \end{cases} \quad (j = 1, 2, \dots, m_1) \quad (1)$$

学習は次式で定める。

$$\begin{aligned} V_{lk}^{co}(t) &= V_{lk}^{co}(t-1) + \eta_2 \delta_l^o(t-1) O_{hid\ k}^{co}(t-1) \\ \delta_l^o(t-1) &= O_{out\ l}^{co}(t-1)(1 - O_{out\ l}^{co}(t-1)) \\ &\quad \cdot (a_l(t) - O_{out\ l}^{co}(t-1)) \\ (l &= 1, 2, \dots, m_1, k = 1, 2, \dots, m_h) \quad (2) \\ W_{kj}^{co}(t) &= W_{kj}^{co}(t-1) + \eta_2 \delta_k^h(t-1) s_j(t-1) \\ \delta_k^h(t-1) &= O_{hid\ k}^{co}(t-1)(1 - O_{hid\ k}^{co}(t-1)) \\ &\quad \cdot \sum_{l=1}^{m_1} \delta_l^o(t-1) V_{lk}^{co}(t-1) \\ (k &= 1, 2, \dots, m_h, j = 1, 2, \dots, m_1) \quad (3) \end{aligned}$$

ここで、 η_2 はある定数で、学習係数である。 $O_{hid\ k}^{co}(t-1)$ は中間層の k 番目のニューロンの出力、 $O_{out\ l}^{co}(t-1)$ は想起層の l 番目のニューロンの出力である。この学習は $t-1$ までの状態ベクトル $s(t-1)$ に対して、時刻 t での入力に対応する活性ベクトル $a(t)$ を出力するようになされている。

続いて、状態 $s_j(t-1)$ を更新する。

$j_{me} \neq 0$ のとき

$$s_j(t) = \begin{cases} 1 & \text{if } j = j_{me} \\ \alpha s_j(t-1) & \text{otherwise} \end{cases}$$

$j_{me} = 0$ のとき

$$s_j(t) = s_j(t-1) \quad (j = 1, 2, \dots, m_1) \quad (4)$$

ただし α は $0 \leq \alpha < 1$ であり減衰定数と呼ぶ。初期状態においてすべての状態は 0 とする。選択ニューロン j_{me} が発火したという情報は、記憶層からの結合により伝達される。

遷移記憶処理についてまとめると、過去の時系列情報を記憶する状態ベクトルに対して、パターン記憶処理で得られる選択ニューロンから決まる活性ベクトルを出力するように学習がなされる。遷移記憶処理では教師信号を用いた学習がなされるが、教師信号はネットワークの内部で生成され、特別な教師データは不要であることから、ASLN model は全体として教師なし学習を行うニューラルネットワークモデルである。

2.3 予測処理

時刻 t において、あるパターンを入力として与えると、モデルはパターン記憶処理及び遷移記憶処理を行い、状態ベクトル $s(t)$ が式 (4) により定まる。この状態ベクトルから $O_{hid\ k}^{co}(t)$, $O_{out\ l}^{co}(t)$ を次の式で定める。

$$O_{hid\ k}^{co}(t) = g\left(\sum_{j=1}^{m_1} W_{kj}^{co}(t)s_j(t)\right) \quad (k = 1, 2, \dots, m_h) \quad (5)$$

$$O_{out\ l}^{co}(t) = g\left(\sum_{k=1}^{m_h} V_{lk}^{co}(t)O_{hid\ k}^{co}(t)\right) \quad (l = 1, 2, \dots, m_1) \quad (6)$$

g は次に示すシグモイド関数である。

$$g(x) = \frac{1}{1 + \exp(-\mu x)} \quad (7)$$

$O_{out\ l}^{co}(t)$ の最大値を与える l を求め、次式で決まる予測パターン $y_i(t) = (y_1(t), y_2(t), \dots, y_{m_0}(t))$ を出力する。

$$y_i(t) = W_{li}^{pa}(t) \quad (i = 1, 2, \dots, m_0) \quad (8)$$

ここで、 $W_{li}^{pa}(t)$ は想起層の l 番目のニューロンと出力層の i 番目のニューロンとの間の重みで、記憶層の l 番目のニューロンの記憶パターンと同じである。

提案モデルはあるパターンが入力されると、自らの記憶に基づき、そのパターンの後に入力されるはずのパターンを出力する。このような処理を予測処理と呼ぶ。予測処理は記憶処理に続いてすぐに行われる。例えばモデルに対して $A \rightarrow B \rightarrow C$ という時系列が繰り返し入力される場合、パターン A を入力するとパターン B を、パターン B を入力するとパターン C を出力する。また、時刻 t における出力を、次時刻の入力としてモデルにフィードバックする閉ループシステムを構築することにより、時系列の先頭パターンを入力として与えるだけで、その後続くパターンを次々と出力することも可能である。この動作をオフライン予測モードと呼ぶ。

ここで提案モデルの動作についてまとめる。時刻 t にモデルに対してパターンが入力されると、入力層から記憶層へ情報が送られる。記憶層においては入力層からの情報をもとに選択ニューロン j_{me} を決定し、選択ニューロンのもつ結合重みベクトル $W_{j_{me}}^{pa}(t)$ の更新、忘却処理及び角度しきい値の更新を行う（パターン記憶処理）。同時に活性ベクトル $a(t)$ を求め、文脈ネットワークの結合重みの更新を行う（遷移記憶処

理）。続いて状態ベクトルを $s(t)$ に更新し、それをもとに想起層のニューロンのもつ記憶パターンを出力する（予測処理）。この出力は次に入力されるはずのパターンを予測している。以上が時刻 t においてパターンが入力されたときのモデルが行う情報処理である。

選択ニューロンが決定されない場合には、文脈ネットワークの重みの更新は行われず、状態ベクトルも変化しない。この性質により、入力される時系列に含まれるパターン数よりも、記憶層のニューロンの数が少ない場合、すなわち記憶容量が制限された場合には、入力頻度の高いパターンがパターン記憶処理され、それらを用いた時系列の学習がなされる。また、入力パターン集合に対して、記憶容量が十分大きい場合には、すべてのパターンを記憶することができ、入力された時系列自体の学習がなされる。

2.4 減衰定数 α の決定法

状態ベクトル $s(t)$ は、記憶層のどのニューロンがどれだけ前に活性化したかという履歴の情報を表している。式 (4) により決まる $s_j(t)$ の値は、一般的に、 $s_j(t) = \alpha^k$ ($k \geq 0$) と表され、これは記憶層のニューロン j に対応する入力 k 時刻前、すなわち、 $t-k$ 時刻に入力されたことを表している。これは、パターン記憶処理を行う自己連想記憶モデルの出力系列に関して、決定的 N 重マルコフモデルを表現できることを意味する (N は計算精度に依存する)。すなわち、時系列の各要素の生起時刻を $s(t)$ という空間的な情報として保持している。

α が小さい値ならば α^k は急速に 0 に近くなり、遠い過去を忘れる。逆に α が 1 に近ければ、計算精度から決まる $\alpha^k \approx \alpha^{k+1}$ を満たす k 時刻前までの過去の状態を識別していることになる。以下では、入力時系列の学習に要した時間と減衰定数 α との関係を数値実験により調査し、その結果に基づき α の設定指針を述べるとともに、提案モデルにおける文脈情報の扱い方を従来モデルと比較する。実験ではモデルに入力する時系列として、図 2 に示す 2 種類を用いた。図中 P_* は時系列に含まれるパターンを意味している。* はパターン番号であり、番号が異なる場合はパターンもすべて異なる。図 2(a) の時系列を順列データ、図 2(b) の時系列を文脈データと呼ぶ。順列データ内の各パターンは、一つ前に入力されたパターンから一意に決まっている（決定的単純マルコフ）。一方、文脈データは P_{N-1} の後に P_N と P_{N+1} という異なるパターンが続く。どちらになるかは、 N 個前に入力されたパ

ターンが P_{N+1} だったか P_N だったかに依存して決まる (決定的 N 重マルコフ). 一般的に, n 時刻前までの情報が文脈情報となる時, その n を文脈長と呼ぶ. 図 2 (b) の文脈データの文脈長は N である.

実験は各パターンを順番にモデルに入力し記憶処理及び予測処理を行う. 時系列内のパターンを一通り入力したときに学習回数 1 回とする. 時系列データを繰り返し入力し, すべての入力パターンに対して正しい予測パターンが得られるまでに要した学習回数を調査した (本実験ではパターン記憶処理に要した学習回数は考慮せず, 遷移記憶処理に要した学習回数だけを考えて). 実験において使用したパラメータは $m_1 = m_h = N + 1, \eta_2 = 0.9, \mu = 1.0$ である.

順列データを学習させたときの α と学習回数の関係を図 3 に示す. 実験結果を見ると α が 0 に近いほど学習回数は少なく, α が 1 に近いほど学習回数は増加していく. このような傾向はパターン数 N によらず共通して見られる. これは α が 0 に近いほど各パターン入力時に得られる状態ベクトル同士の直交性が高くなり, 少ない学習回数での学習が可能となるためであると考えられる. 逆に α が 1 に近いほど状態ベクトル同士の直交性が低くなり, 多くの学習回数を要する.

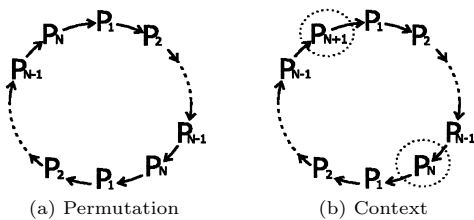


図 2 順列データと文脈データ

Fig. 2 Permutation data and context data.

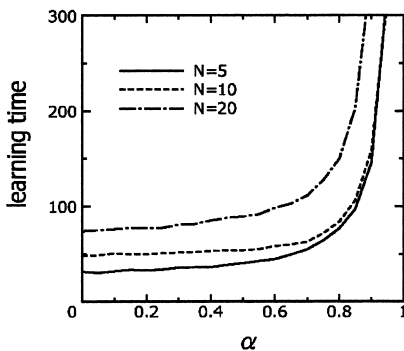


図 3 学習回数と減衰定数 α の関係 (順列データ)

Fig. 3 Relation between learning time and reduce parameter α . (permutation data)

次に文脈データを学習させたときの α と学習回数の関係を図 4 に示す. これを見ると, α の値が 1 に近い場合は順列データと同じ傾向であるが, α の値が小さい場合にも学習回数が増大し, 最小の学習回数を与える α の値が文脈長 N に応じて決まることを示している. また, この最適な α の値は文脈長 N とともに増大している. $m_1 = N + 1$ であるので, 状態ベクトルは $N + 1$ 次元であり, P_N の入力の直前では $s(t) = (\alpha^{N-2}, \alpha^{N-3}, \dots, \alpha^1, \alpha^0, \alpha^{2N-1}, \alpha^{N-1})$, P_{N+1} の入力の直前では $s(t) = (\alpha^{N-2}, \alpha^{N-3}, \dots, \alpha^1, \alpha^0, \alpha^{N-1}, \alpha^{2N-1})$ となっている (要素の並びは $P_1, P_2, \dots, P_{N-1}, P_N, P_{N+1}$ の順). したがって, 最初の $N - 1$ 個の要素は同じであるが, 最後の 2 個の要素が逆の値になっており, この違いが予測出力の違いとなるように学習されるはずである. すなわち, α^{N-1} と α^{2N-1} とは違う値でなければならないが, α が小さいと $\alpha^{N-1} \approx \alpha^{2N-1} \approx 0$ となり, 違いがなくなり, 学習できなくなる. α が 1 に近いと学習時間がかかるのは順列データと同じ理由による考えられる. つまり文脈情報という観点から見れば, α が大きいほど, 遠い過去に入力されたパターンが文脈情報として保持され, 予測に大きな影響を及ぼすと解釈できる. なお実験の結果から最小の学習回数を与える α の N 乗は, $N = 5, 10, 20$ について約 0.37, 0.31, 0.17 であった.

以上のことをまとめると, 順列データのような文脈長が 1 の時系列を扱う場合には, α を小さくすることで各状態ベクトルは互いに直交し, 速やかに学習ができる. また文脈データのように文脈長が N の時系列を扱う場合には, α をある程度大きくし, $\alpha^N = 0.2 \sim$

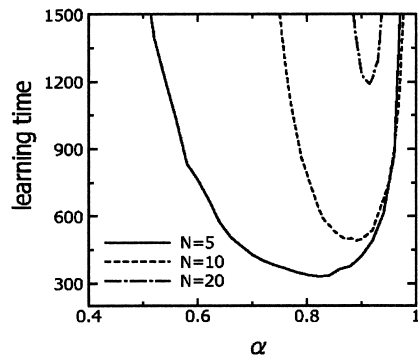


図 4 学習回数と減衰定数 α の関係 (文脈データ)

Fig. 4 Relation between learning time and reduce parameter α . (context data)

0.4 となる α を選べば、より以前に入力されたパターンによる文脈情報を用いた学習が可能となる。これらの傾向を利用し適用対象の特性から、文脈情報としての程度過去の情報までを扱う必要があるのかを考慮して α を決定することで、より少ない学習回数での時系列の記憶が実現可能となる。あるいは、逐次学習中に α の値を次第に大きくしていくことで、未知の N に対しても適切な値を選定することが可能であると思われるが、この点は今後の課題である。

なお文脈長の増大に伴い学習回数が増大する。しかし、実際の応用に関しては、それほど長い文脈長が必要とされない場合が多いと考えられ、その場合、現実的な学習回数での学習が可能である。

提案モデルと従来の研究における文脈情報の扱い方を比較する。Araújo ら [10] は、どれだけ過去の情報を文脈情報として扱うかをあらかじめ設定し、その情報を入力パターンに付加して用いる方法を提案している。彼らの方法では入力される時系列が設定よりも更に過去の情報を文脈情報として必要とする場合は学習ができず、ネットワーク構造そのものを再構築する必要がある。しかし提案モデルにおいては α の値の変更だけで対応でき、ネットワーク構造の変化を必要としない。

Elman ネット [7] では入力パターンと、入力履歴を表すパターンを単に連結するという方法を用いている。この方法は、入力パターン同士の直交性の影響が文脈表現能力の低下を招くことが指摘されている [11]。本モデルでは文脈情報を状態ベクトルとして表現し、遷移記憶処理に用いており、状態ベクトル同士の直交性は減衰定数の大きさのみ依存している。したがって、入力パターン同士の直交性により文脈表現能力が低下するという問題は生じない。また、Elman ネットを用いて文脈データの学習を行ったところ、 $N = 3$ で約 3000 回以上の学習を必要とし、ネットワーク構造（各層のニューロン数）を変化させると学習が収束しない場合が多く確認された。これらの理由からも、本モデルが対象とする、記憶すべき時系列が未知の場合には、Elman ネットは向いていないと考える。

3. 実験結果及び考察

計算機実験を行い、提案モデルが環境の変化に追従して、自らの記憶を変化させていく様子を確認した。また、ばらつきを含む手書き数字パターンを用いた数字列の学習実験を行い、パターン数を増やした場合の

予測能力の検証を行った。

3.1 記憶容量制限下における時系列の逐次学習

実験では提案モデルに図 5 に示す異なる 2 種類の時系列を入力する。図 5(a) (時系列 1) の場合はモデルに対して $A \rightarrow B \rightarrow (C \text{ or } D) \rightarrow E \rightarrow B \rightarrow F \rightarrow E \rightarrow (G \text{ or } H) \rightarrow F \rightarrow A \rightarrow \dots$ という順番でパターンを入力する、 $(C \text{ or } D)$ と $(G \text{ or } H)$ に関しては、どちらか一方のパターンを等確率でランダムに選択し入力する。図 5(b) (時系列 2) も同様である。実験では記憶処理、予測処理は各パターンの入力ごとになされ、すべてのパターンの入力が完了した（パターンが順番に 9 個入力された）ときを 1 ステップとする。実験で用いたパラメータは $m_0 = 49$ (7×7 の文字画像)、 $m_1 = 4$ 、 $m_h = 4$ 、 $\eta_1 = 0.01$ 、 $\eta_2 = 0.9$ 、 $\mu = 1.0$ 、 $\alpha = 0.7$ 、また、文献 [3] を参考に $R_{red} = 0.0003$ 、 $R_{min} = 0.2$ とした（付録参照）。

記憶層のニューロン数を $m_1 = 4$ とし、時系列 1 を繰り返し入力した場合、自己連想記憶モデルは 8 個のパターンのうち、頻度の高い B, E, F, A を記憶し、 C, D, G, H は記憶しないと予想される。したがって、文脈ネットワークには、 $A \rightarrow B \rightarrow E \rightarrow B \rightarrow F \rightarrow E \rightarrow F(\rightarrow A)$ という時系列が入力されるはずである。この系列は、例えば B の次に E と F がきており、これらを予測するには E の前の AB 、 F の前の EB という系列の遷移記憶が必要で、文脈長 2 の時系列である。同様に時系列 2 を入力した場合、 $A \rightarrow D \rightarrow E \rightarrow H \rightarrow D \rightarrow E \rightarrow H(\rightarrow A)$ という時系列が記憶されることが予想される。この系列では DEH という系列の次が、 D または A となっており、長さ 3 の記憶では両者を区別できず、 $ADEH$ 、 $HDEH$ という長さ 4 の記憶により初めて、次が D または A と予測できる、文脈長 4 の例である（ちなみに、時系列 2 自体は文脈長が 3 であるが、決定的三重マルコフ系列ではなく、 BDE, CDE の文脈の後、 F または G が確率的に生起する。 HDE の文脈で H

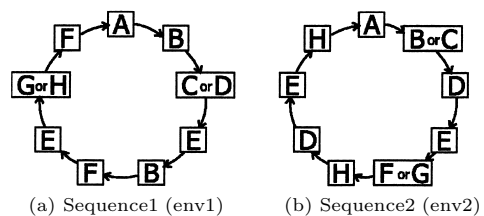


図 5 時系列データ
Fig. 5 Sequence data.

が続く)。

実験は 1200 ステップまで行くとし、前半の 600 ステップまでを環境 1 (env1) とし時系列 1 をモデルに繰り返し入力する。また後半の 601 から 1200 ステップまでを環境 2 (env2) とし、時系列 2 をモデルに繰り返し入力する。入力時系列の変化は環境の変化を表現している。モデルは環境の変化に追従して、自らが記憶している時系列を変化させることが期待される。すなわち、env1 では B, E, F, A が記憶され、env2 では D, E, H, A が記憶されると予想されるが、環境の変化直後、どのように記憶が更新されるかを確かめる。

モデルが記憶した時系列を確認するために、モデルにパターン A (初期パターン) を入力し、オフライン予測モードにより次々と出力されるパターンを確認した (図 6)。図中のパターンは予測パターンとして得られた出力層ニューロンの重みベクトルの各成分の大きさを、正規化して得られたものである。初期パターン A 以外のパターンは、モデル自身の予測により得られたものである

図 6(a), (b) はそれぞれ 600 ステップと 1200 ステップにおける出力結果である。これらを見ると、各環境において入力頻度の高いパターンを優先的に記憶し時系列を表現していることが分かる。また逐次学習により、環境の変化に追従し自らの記憶を変化させていることが分かる。

また、記憶層のニューロンを追加し $m_1 = 5$ として同様の実験を行った。記憶層に追加されたニューロンは、各環境において $m_1 = 4$ のときに記憶されない、入力頻度の低い 4 個のパターンが交ざったようなパターン (パターン X とする) を記憶することを確認した。このニューロンは入力頻度の低い 4 個のパターンすべてに発火する。そのため、記憶される系列は環

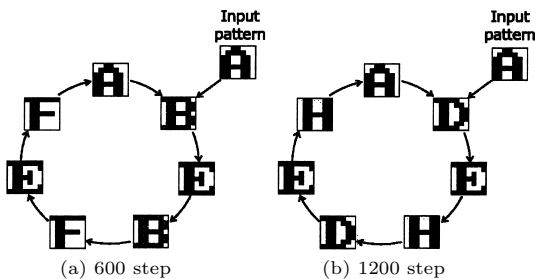


図 6 出力結果
Fig. 6 Output of sequences.

境 1 の場合、 $A \rightarrow B \rightarrow X \rightarrow E \rightarrow B \rightarrow F \rightarrow E \rightarrow X \rightarrow F (\rightarrow A)$ となる。ただし、記憶パターンは結合重みベクトル $W_j^{pa}(t)$ の初期値及び、入力パターンの確率に依存して形成されるため、必ずしもすべての低入力頻度のパターンに対し発火する記憶パターンが形成されるとは限らず、その場合は記憶される系列も異なる。

次に環境の変化に追従する際の、モデルの記憶する時系列の変化について着目する。図 7(a) は 1000 ステップにおける出力結果である。パターン A の入力をきっかけに $A \rightarrow E \rightarrow E \rightarrow \dots$ という時系列を繰り返し出力することを確認した。モデルは環境 1 においてはパターン A, B, E, F を記憶しているが、環境の変化に伴い入力頻度の下がったパターン B, F は徐々に忘却され、新たに環境 2 における高頻度入力パターン D, H の記憶の形成が始まる。忘却は徐々に進むため、環境が変化した直後は D, H が入力されても選択ニューロンは決定せず、文脈ネットワークの重みの更新は行われぬ。結果的に環境 1, 2 の両方で高頻度で入力されるパターン A, E のみで時系列が表現されたものと考えられる。

また、図 7(b) は 1100 ステップにおける出力結果であり、 $\dots \rightarrow D \rightarrow E \rightarrow H \rightarrow \dots$ という時系列を繰り返し出力することを確認した。この段階では環境 1 の記憶が忘却されて、新たにパターン D, H を記憶している。出力された時系列でパターン A が消えたのは以下の理由による。 A を予測するためには $HDEH$ という文脈の記憶が必要であり、それに対して A を出力するように学習されていない。また $ADEH$ の文脈のもとでは D を予測するはずである。文脈 $HDEH$ での状態ベクトルは $s(t) = (\alpha^6, \alpha^2, \alpha, 1)$ (要素の並びは A, D, E, H に対応), $ADEH$ の文脈は $s(t) = (\alpha^3, \alpha^2, \alpha, 1)$ であり、

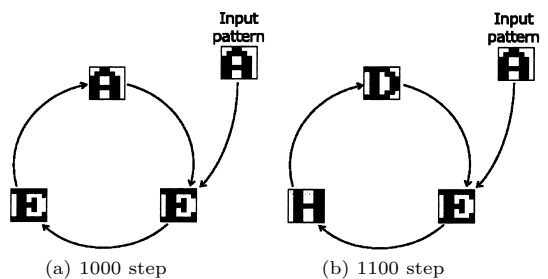


図 7 出力結果
Fig. 7 Output of sequences.

文脈ネットワークの結合重みの再更新が必要とされ、一時的にパターン A が出力されなくなったものと考えられる。更に学習を続けると図 6 (b) のように環境 2 (env2) で記憶することが期待される時系列の記憶を形成する。

本実験では、二つの環境とも入力パターン 8 個に対して 3 個が同一の高頻度、1 個 (A) がその半分の頻度、残りの 4 個が更にその半分の頻度となっており、 $m_1 = 4, m_1 = 5$ に対して、十分に一般的な実験条件といえる。

以上の結果、次のようなことが確認された。提案モデルは、環境が変化した場合、まず二つの環境で共通して入力頻度が高いパターンだけで、時系列を表現し、その後、新たな環境で入力頻度の高いパターンの記憶が形成されてくると文脈ネットワークの結合重みの更新がなされ、新たな時系列の記憶が形成される。ただし、各環境で共通のパターンが存在しない場合は、過去の環境で形成された記憶を忘却し新たな環境で入力されるパターンの記憶が形成されるまで、出力は抑制され、記憶形成後に改めて新しい時系列を学習する。

なお、提案モデルは、常に入力頻度の高い上位 m_1 のパターンの記憶を優先的に形成するが同一頻度のパターンのため、上位 m_1 が一意に決まらないときは $m_1 = 5$ の実験のように、入力パターン自体を記憶できないことがある。

3.2 手書き数字を用いた数字列の記憶

ばらつきを含む入力パターンを用い、パターン数を増やした場合の予測能力を検証するために、手書き数字を用いた数字列の学習実験を行った。実験では、図 8 (a) に示すように、手書きの数字を線形正規化し作成した 10×10 のドットパターン (数字パターン) をモデルへの入力とした。用いる数字は 0 から 9 までの 10 種類であり、ばらつきを考慮するため、各数字ごとに同一人物が 10 回ずつ手書き数字を書き、合計 100 個の数字パターンを用意した。記憶容量は 10

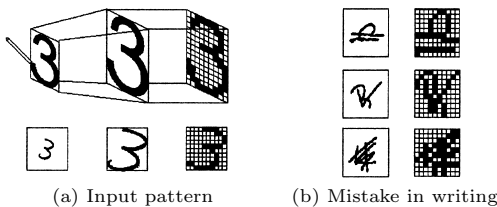


図 8 入力パターン
Fig. 8 Input patterns.

(0~9の数字に対応)とした。実験では0~9までの数字をランダムに並べて11けたの数字列を作成し、繰り返しモデルへ入力した。

一例として、数字列 35040683313 を用いた場合、モデルに対して $3 \rightarrow 5 \rightarrow 0 \rightarrow \dots$ と順番に数字パターンを繰り返し入力する。入力する数字パターンには、ばらつきが含まれており、10分の1の確率で間違いパターン (図 8 (b) に示す、10種類作成) を含んだ数字列を入力として与えた。モデルは高頻度入力パターンから優先的に記憶するため、このようなパターンを入力しても、数字パターンだけを記憶し、正しい時系列を記憶することが期待される。実験で用いたパラメータは $m_0 = 100, m_1 = 10, m_h = 10, R_{red} = 0.0001$ であり、その他のパラメータは 3.1 の実験と共通である。また、角度しきい値の最小値が 40 [degree] となるように設定した。

出力結果を図 9 に示す。各出力パターンは 3.1 と同様に、繰り返し入力を行ったモデルに対し、数字列の先頭の数字パターンを入力として与え、オフライン予測モードで出力させたものである。各出力パターンは、入力される各数字パターンのばらつきを平均化したパターンとなっているため、一部がぼやけているように見える。しかし、入力した時系列が正しく記憶されていることが確認できる。また種々の数字列で実験を行い、モデルの記憶パターンを確認したところ、数字パターンだけを記憶していることを確認した。

モデルに対し 13071425612 と 13003624537 という二つの数字列を交互に入力し、複数の数字列の記憶実験を行った、二つの数字列の最初の 3 けたは共通である。なお、各数字列の先頭のパターンを入力するたびに、状態ベクトルの初期化を行った。出力結果を図 10



図 9 出力結果
Fig. 9 Output of sequences.

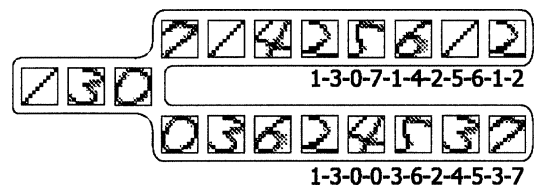


図 10 出力結果
Fig. 10 Output of sequences.

に示す．入力を繰り返したモデルに対し先頭の1を入力したところ，1307...と入力した数字列の片方を出力した．このとき，数字列の4番目の数字パターン7を出力する際の文脈ネットワークの想起層の出力値を確認したところ，記憶パターン0と7に対応するニューロンの出力が他のニューロンの出力に比べ大きいことが分かった．これは想起層のニューロンの出力があるしきい値を超えたものすべてを，その時点での次に入力される数字パターンの予測候補とできることを示唆している．次にモデルが先頭から3番目の0を出力した後に，外部から数字パターンの0を入力したところ，片方の数字列(13003...)を出力した．同様に外部から数字パターンの7を入力すると，もう一方の数字列を出力することが分かった．以上のことから，提案モデルは複数の数字列の記憶が可能であり，それらの数字列の先頭からの一部分が共通であっても，入力に応じて正しい数字列を予測することが可能であることが分かる．記憶すべき数字列の数の増加に対しては，中間層のニューロン数の増加で対処できると考えられるが，その詳細な検証は今後の課題である．

4. む す び

本論文では，記憶容量制限を考慮した逐次学習型の時系列予測モデルとして，適応的逐次学習ネットワークモデル(ASLN model)を提案した．ASLN modelは予測をするために必要となる時系列の記憶を，環境からの入力情報を逐次的に記憶することで獲得する．計算機実験により，提案モデルが記憶容量の有効利用を図りながら，自らの記憶を変化させることで周囲の環境の変化に対し追従していくことを示した．また，ばらつきを含む手書き数字パターンを用いた数字列の学習実験を行い，パターン数を増やした場合の予測能力の検証を行った．更に，モデルが時系列の遷移情報を記憶する上で重要となるパラメータについて実験を行い設定指針を与えた．

今回提案したモデルは，入力として時系列情報を繰り返し与える必要がある．そのため環境の変化に対する即応性という点では問題が残る．即応性の向上を実現するために，より少ない回数での学習が可能となるようにする必要があり，そのためには減衰定数の自動最適化等，文脈ネットワークの学習法について検討する必要がある．また応用に際して，より多くの時系列を同時に記憶できるようにするため，中間層のニューロン数を増加させればよいが，その定量的評価

は今後の課題である．今後は実データを用いた実験など，提案モデルの応用について，検証していく予定である．

謝辞 本研究の一部は東北大学 21 世紀 COE プログラム「新世代情報エレクトロニクスシステムの構築」(課題番号 14I70017)の援助により行われた．

文 献

- [1] 平田高志, 村上晴美, “携帯型外化記憶構築システムの試作,” 大阪市立大学学術情報総合センター紀要, vol.4, pp.45–50, March 2002.
- [2] 青木 猛, 公文俊朗, 安田 真, 中條直也, 中川秀一, 川島 毅, 芳澤伸一, 岩崎正宏, “創発型ソフトウェアの制御・スケジューリングと画像処理への応用,” 電学論 (D), vol.122, no.1, pp.4–6, Jan. 2002.
- [3] 芳澤伸一, 道木慎二, 大熊 繁, “記憶容量の制約下での追加学習を目的とする新しい連想記憶システムの提案,” 信学論 (D-II), vol.J82-D-II, no.6, pp.1072–1081, June 1999.
- [4] 長名優子, 萩原将文, “カオスニューラルネットワークにおける逐次学習,” 信学論 (D-II), vol.J82-D-II, no.1, pp.83–90, Jan. 1999.
- [5] 郷古 学, 菅谷至寛, 阿曾弘具, “記憶容量制限下で頻出系列を逐次学習するニューラルネットワーク,” 平 16 東北連大, 2F-10, p.223, Aug. 2004.
- [6] 郷古 学, 菅谷至寛, 阿曾弘具, “記憶容量制限下で手続的知識を獲得する連想記憶モデル,” 信学論 (D-II), vol.J88-D-II, no.7, pp.1304–1307, July 2005.
- [7] J. Elman, “Finding structure in time,” *Cognitive Science*, vol.14, pp.179–211, 1990.
- [8] G. Bugmann, K.L. Koay, N. Barlow, M. Phillips, and D. Rodney, “Stable encoding of robot trajectories using normalised radial basis functions: Application to an autonomous wheelchair,” *Proc. Int. Sym. Robotics (ISR '98)*, pp.232–235, April 1998.
- [9] G.A. Barreto and A.F.R. Araújo, “Unsupervised learning and recall of temporal sequences: An application to robotics,” *Int. J. Neural Syst.*, vol.9, no.3, pp.235–242, June 1999.
- [10] A.F.R. Araújo and G.A. Barreto, “Context in temporal sequence processing: A self-organizing approach and its application to robotics,” *IEEE Trans. Neural Netw.*, vol.13, no.1, pp.45–57, Jan. 2002.
- [11] 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,” 信学論 (D-II), vol.J85-D-II, no.10, pp.1602–1612, Oct. 2002.
- [12] J. Hertz, A. Krogh, and R.G. Palmer (著), 笹川辰弥, 吳 勇 (訳), ニューラルコンピュータ: 統計物理学からのアプローチ, トッパン, 東京, 1994.

付 録

自己連想記憶モデル[3]について説明する．時刻 t におけるモデルに対する入力パターンベクトルを

$I^{in}(t) = (I_1^{in}(t), I_2^{in}(t), \dots, I_{m_0}^{in}(t))$ とする。入力層では正規化処理を行い、出力ベクトル $O^{in}(t) = (O_1^{in}(t), O_2^{in}(t), \dots, O_{m_0}^{in}(t))$, $O_i^{in}(t) = I_i^{in}(t) / \|I^{in}(t)\|$ を出力する。記憶層のニューロン j の出力 $O_j^{me}(t)$ を次式で決める。

$$O_j^{me}(t) = \sum_{i=1}^{m_0} W_{ji}^{pa}(t) O_i^{in}(t) \quad (j = 1, 2, \dots, m_1) \quad (A.1)$$

$O^{in}(t)$ と $W_j^{pa}(t)$ のなす角度 $\theta_j(t)$ は $\theta_j(t) = \cos^{-1}(O_j^{me}(t)/R_j(t))$ と表される。ただし $R_j(t) = \|W_j^{pa}(t)\|$ (重みベクトル $W_j^{pa}(t)$ の大きさ) であり、記憶層のニューロン j のもつ記憶パターンの信頼度と呼ぶ。次に選択ニューロン j_{me} を次式で求める。

$$j_{me} = \begin{cases} j & \text{if } \theta_j(t) \leq \theta_{max,j}(t) \\ & \& \forall i; O_j^{me}(t) < O_i^{me}(t) \text{ について} \\ & \theta_i(t) > \theta_{max,i}(t) \\ 0 & \text{if } \forall i; \theta_i(t) > \theta_{max,i}(t) \end{cases} \quad (A.2)$$

ここで、 $\theta_{max,j}(t)$ は記憶層のニューロン j のもつ角度しきい値である。

選択ニューロン j_{me} について、 $j_{me} \neq 0$ のとき式 (A.3) に示す Oja 則 [12] を用い、重みベクトルを更新する。

$$W_{j_{me}i}^{pa}(t+1) = W_{j_{me}i}^{pa}(t) + \eta_1 O_{j_{me}}^{me}(t) (O_i^{in}(t) - O_{j_{me}}^{me}(t) W_{j_{me}i}^{pa}(t)) \quad (i = 1, 2, \dots, m_0) \quad (A.3)$$

ここで、 η_1 はある定数で、学習係数である。

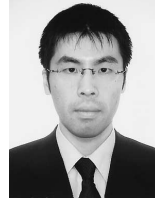
重みの更新後にすべての重みベクトル $W_j^{pa}(t)$ の大きさを次式に従い一定量だけ減ずる忘却処理を行う ($j_{me} = 0$ でも行う)。

$$W_j^{pa}(t+1) = \begin{cases} \frac{R_j(t) - R_{red}}{R_j(t)} W_j^{pa}(t) & \text{if } R_j(t) > R_{min} \\ \frac{R_{min}}{R_j(t)} W_j^{pa}(t) & \text{otherwise} \end{cases} \quad (j = 1, 2, \dots, m_1) \quad (A.4)$$

ここで、 R_{red} , R_{min} はある定数であり、信頼度の減衰量及び下限を意味する。

最後に、各記憶層のすべてのニューロンのもつ角度しきい値を $\theta_{max,j}(t+1) = 15 + 75 \exp(-7(R_j(t) - 0.2))$ ($j = 1, 2, \dots, m_1$) により更新する。このとき $15 \leq \theta_{max,j}(t+1) \leq 90$ である。

(平成 17 年 5 月 20 日受付, 9 月 30 日再受付, 11 月 10 日最終原稿受付)



郷古 学 (学生員)

平 11 山形大・工・機械システム卒・平 13 同大学院理工学研究科博士前期課程了。同年 (株) 富士通宮城エレクトロニクス (現, 富士通インテグレートドマイクロテクノロジー (株)) 入社。半導体生産技術開発に従事。平 15 東北大学大学院工学研究科博士後期課程編入学。平 16 同大 21 世紀 COE (新世代情報エレクトロニクスシステムの構築) リサーチアシスタント。ニューラルネットワーク, ファジィの研究に従事。情報処理学会会員。



菅谷 至寛 (正員)

平 7 東北大・工・通信卒。平 9 同大学院工学研究科博士前期課程了。平 14 同大学院工学研究科博士後期課程了。平 12 より同大学院工学研究科助手。博士 (工学)。シストリックアルゴリズム設計支援システム, タスク割当, 並列ソフトウェアなどの研究に従事。並列分散処理全般に興味をもつ。情報処理学会会員。



阿曾 弘具 (正員:フェロー)

昭 43 東北大・工・電気卒。昭 49 同大学院博士課程了。昭 48 東北大・工・助手, 昭 54 名大・工・講師。昭 57 同助教授, 昭 61 東北大・工・助教授を経て, 平 3 同教授。現在, 同大学院工学研究科教授。工博。その間, 学習オートマトン, セル構造オートマトン, 並行処理理論, シストリックアルゴリズム設計論, 文字認識, 音声認識, パターン認識, ニューラルネットワークなどの研究に従事。平 3 年度本会業績賞受賞。平 15 年度本会フェロー。IEEE, ACM, EATCS, 情報処理学会, 人工知能学会, 日本認知科学会, LA 各会員。