

## 単純再帰結合型ネットワークを拡張した 語系列予測モデルの提案

篠 沢 佳 久<sup>†1</sup>

エルマンは単純再帰結合型ネットワーク（エルマンネット）を用いて言語の獲得過程のモデルを提案した。これは文章を単語単位に区切って現時点の単語を提示し、次単語を予測する学習（語系列予測課題学習）をエルマンネットを用いて行うことによって、文法特徴を獲得することが可能なことを示唆した。エルマンネットによる文法や語彙の獲得の学習は文法規則に制約を設けた人工文を対象とし、単語数についてもその個数が限定されている。また学習対象が長文の場合、過去に入力された単語情報が蓄積されていないといった指摘もされている（Low Memory Accuracy）。そこで本論文においては語系列予測課題学習において、文法獲得の精度の向上および Low Memory Accuracy の課題に対処するため、構造上、エルマンネットを拡張した語系列予測モデルを提案する。これは複数のネットワークを用いて分散処理を図り、学習用の例文中に表出する単語の接続情報（バイグラム）を用いて、各ネットワークの入出力構造を決定する。すなわち学習中、新たな単語が表出するたびに、対応したネットワークの入力層のニューロンを追加することによって実現する。さらにネットワークの内部構造の局所結合化を図り、中間層に蓄積される情報を自己フィードバックと文脈フィードバックの2種類のフィードバックにより、過去の単語の履歴情報と文脈情報に分けて学習する方法を提案する。そして語系列予測課題学習を通じた実験の結果、提案モデルにおいては次単語予測、文法規則の獲得の精度の向上および Low Memory Accuracy の改善を示すことができた。

### A Model of Word Sequence Prediction by Improved Simple Recurrent Networks

YOSHIHISA SHINOZAWA<sup>†1</sup>

Elman proposed a simple recurrent network which is a model of language acquisition. Elman showed that a SRN learned to predict the word to come next in the sentences and could acquire grammatical concepts and meanings. It is difficult for a SRN to learn the sentences which contain a number of word and complicated grammatical structure. Especially, it is difficult to memorize

input words in earlier step (low memory accuracy). In this paper, we propose a model of word sequence prediction by improving a SRN. We propose how to learn to predict the next words by distributed networks, whose structure is decided by using relationship between words in the sentences, bigram. In learning of proposed networks, a neuron of the input layers is added when a new word emerges in the sentence. Proposed networks have two loopbacks from hidden layers to context layers and local connections to learn information about words and grammatical features stored in the hidden layers separately. We examine effectiveness of our model with learning of next word prediction. Obtained results show that our model can acquire grammar features and improve low memory accuracy.

#### 1. はじめに

現在、人間の言語の理解や獲得についてのモデルを構築する研究がさまざまな分野で行われている。特にエルマンの提案するニューラルネットワークを用いた言語情報の獲得モデル<sup>1)–3)</sup>は、計算機科学の分野においてこれまで多くの研究がされてきた。エルマンは多層パーセプトロン型のニューラルネットワーク（以下ネットワークと略す）を拡張した単純再帰結合型ネットワーク（本論文においてはエルマンネットと呼ぶ）を用いて、学習による文法の獲得過程のモデルを提案した。これは英語文を単語単位に区切り、文頭から順にエルマンネットに1単語ごとに入力し、次単語を予測する学習（語系列予測課題学習とも呼ぶ）を行った。その結果、学習後のネットワーク内の中間層における活性パターンが統語範疇（名詞や動詞といったカテゴリ）や意味範疇（人や物といったカテゴリ）を表すクラスタに分かれていることが判明した。すなわち文法情報や単語の意味的な知識を与えず文章中の単語を先頭から順に提示し、次単語を予測する学習を繰り返すことによって文法特徴を獲得することが可能なことを示したのである。これまでにエルマンネットを基礎として、さまざまな文法<sup>4)–5)</sup>や語彙<sup>6)–7)</sup>についての言語獲得モデルが提案されている。特に文法特徴の獲得については関係詞を含む複雑な文章の構造<sup>2)</sup>や疑問詞を文頭にした疑問文<sup>4)</sup>、さらには前置詞を含む副詞句<sup>5)</sup>などの学習が可能なことが示されている。

エルマンネットによる語系列予測課題学習の詳細については2章において述べるが、文法規則に制約を設けた人工文を対象とし、特に含まれる単語についてはその個数も限定さ

<sup>†1</sup> 慶應義塾大学理工学部管理工学科

Faculty of Science and Technology, Keio University

れている．そして新たな単語が表出した際に，どのようにネットワークの構造を対応させていくかといった問題も生じる．これはエルマンネットを語彙獲得モデルとして拡張していくうえで解決しなければならない課題である．また文法獲得モデルとして構築するうえでも，汎化能力の向上の困難さや学習対象が長文の場合，過去に入力された単語情報が蓄積されていかない (Low Memory Accuracy と呼ぶ<sup>8)</sup>) といった指摘もされている．

そこで本論文においては語彙数や文法規則の複雑化に対応が容易で，かつ Low Memory Accuracy の課題に対処するため，構造上，エルマンネットを拡張した語系列予測モデルおよびその学習方法を提案する．これは単一のエルマンネットで処理する機能を，複数のネットワークを用いて分散処理を図る．各ネットワークの入出力構造については，学習用の例文中に表出する単語の接続情報 (バイグラム) を用いて決定する．またエルマンネットにおいては，直前までのネットワークの状態を文脈層に内部表現として蓄積していくが，文法規則の獲得の精度を向上させるため，文脈層に蓄積される情報を過去の単語の履歴情報と文脈情報に分割し，学習できるようにする．これについては中間層から文脈層へのフィードバックを 2 種類 (自己フィードバック，文脈フィードバック) 設け，各ネットワークの内部構造の局所結合化を図ることによって実現を試みる．そして語系列予測課題学習を通して，提案モデルによる文法規則の獲得の精度の向上および Low Memory Accuracy の改善について検証する．

## 2. エルマンネットによる次単語予測学習

### 2.1 エルマンネットの構造と学習方法

エルマンネットは入力層，中間層 (複数層存在する場合もある)，出力層から構成された多層パーセプトロン型のネットワークを基本構造としている (図 1)．さらに入力層は外部からの入力値を処理する入力層と直前の中間層の値がフィードバックされる文脈層の 2 つから構成される．文脈層は直前の中間層の値を保持することから，直前のネットワークの状態 (過去に入力された情報) を記憶することになり，この性質を利用してエルマンネットは言語情報の獲得モデルだけでなく時系列処理にも利用されてきた．

エルマンネットによる文法特徴の獲得は次単語予測学習 (語系列予測課題学習) を通じて行われる．次単語予測学習とは文頭から順に 1 単語のみを提示し，次に表出する単語を予測する学習のことである．たとえば *Mary sees dogs.* という文を学習する場合，文頭の *Mary* を提示し次単語として *sees* を予測する．次に *sees* を提示し *dogs* を予測する．そして *dogs* を提示し文末「。」を予測するといったようにである．

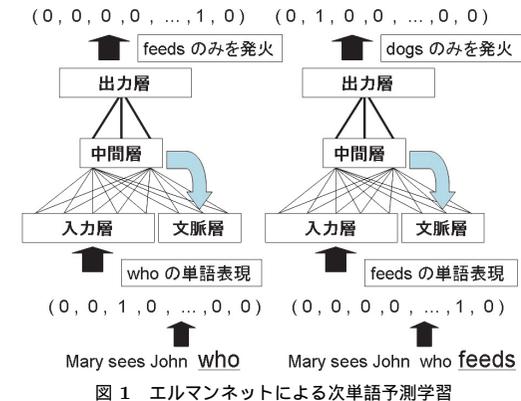


図 1 エルマンネットによる次単語予測学習

Fig. 1 Learning of word sequence prediction by a simple recurrent network.

通常，次単語予測学習をエルマンネットによって行う場合，学習用の例文中に表出する任意の 1 単語と入力層および出力層中の特定の 1 ニューロンとを対応させる．そのため学習に用いる例文中に表出する単語数を  $N$  個とすると，入力層および出力層はそれぞれ  $N$  個のニューロンが必要となり，その個数分のニューロンをあらかじめ配置しておく．そして文章を単語ごとに区切って先頭の単語から順次入力層に入力していく．その際現時点に入力する単語と対応する入力層のニューロンには 1 を，それ以外は 0 というような単語表現で入力する．たとえば図 1 において *who* の単語表現は  $(0, 0, 1, 0, \dots, 0, 0)$ ，*feeds* は  $(0, 0, 0, 0, \dots, 1, 0)$ ，そして *dogs* は  $(0, 1, 0, 0, \dots, 0, 0)$  といったようにである．したがって各単語は互いに直交した  $N$  次元ベクトルで表現され，単語間の距離は等しいためどの単語が類似しているかといった情報をあらかじめ与えない．

文脈層については中間層と同じ個数のニューロンを用意しておき，直前の中間層の値 (内部表現と呼ぶ) をそのままフィードバックして入力する．すなわち入力層には現時点の単語情報を単語表現という形式で入力し，文脈層には直前のネットワークの状態を中間層に蓄積された内部表現という形式で再入力する．このような 2 種類の入力値をネットワークに同時に入力することにより，上位部に位置する中間層には現時点までに入力された文章の情報が蓄積されることになり，この中間層の情報をもとに次単語の予測を行う仕組みとなっている．

学習を行う場合，まずこの 2 種類の入力値をもとにネットワークを稼働させ，出力層の値を求める．そして次単語と対応する出力層のニューロンのみが発火し，それ以外は発火しな

いように教師信号（次に表出する単語の単語表現）を与え、出力値がその教師信号と一致するようにネットワーク内の結合係数を修正する．これを文頭の単語から文末まで、複数の学習例文を用いてネットワークが適切に次単語を予測するようになるまで繰り返す．学習には多層パーセプトロン型のネットワークに適用が容易な誤差逆伝播則（バックプロパゲーションアルゴリズム<sup>9)</sup>）が利用される．

そして学習後のネットワーク内の中間層の活性パターンのクラスタ分析を行ったところ、動詞や名詞といった品詞ごとのカテゴリや名詞においては人や物といったカテゴリ、さらには動詞においても3単元の動詞と原型動詞といったカテゴリに分かれており、階層型の単語クラスタが形成されていることが判明した<sup>1)</sup>．すなわち文頭から現時点の単語のみを示し、次単語を予測する学習のみで単語の統語範疇や意味範疇といった文法特徴を獲得できることを示したのである．これは N-gram といった統計的手法とは異なり、過去の単語の入力情報を直接用いていないことに大きな特徴を持っている．

## 2.2 エルマンネットによる次単語予測学習の課題

以上のようなエルマンネットを用いた文法や語彙獲得の学習は文法規則に制約を設けた人工文（英語）を対象とし、特に含まれる単語数についてはその個数も限定されている．これはネットワークの構造および機能上、入力層および出力層において各単語とニューロンを1対1に対応づけするため、単語数分のニューロンを必要とするためである．したがって、単語数が増加した場合あるいは新たな単語が表出した場合、以下の課題が生じる．

### 課題（1）

単語数が多くなった場合、ネットワークの規模が大きくなり単一のネットワークでは学習が困難なことが予想される．

### 課題（2）

学習中、未知の単語が表出した場合、どのように構造を対応させるのかということが問題となる．

単体のエルマンネットを用いて数千語の単語を対象とした自然言語の文章を解析するという研究事例は少なく、一般的ではない．むしろエルマンネットを用いて言語情報の解析を行う場合、語彙や文法特徴の獲得モデルの構築に用いられるため課題（2）の方が解決すべき問題となる．これについては次の2通りの解決方法が考えられる．

- 学習開始前にネットワークの構造をあらかじめ決めておき、表出すると予測される単語と対応したニューロンを入力層および出力層に配置しておく<sup>1)</sup>．
- 新たな単語が表出するたびに、新規に入力層と出力層のニューロンを増やしていく<sup>7)</sup>．

未知の単語を学習する場合、表出すると予想される単語に対応したニューロンをあらかじめ入力層と出力層に用意し、その単語の単語表現を固定して決めておくのは語彙獲得モデルとしては柔軟ではなく、新たな単語が表出するたびにニューロンを新規に増やしていくことが望まれる．しかしこの場合、新規にニューロンを追加することによってネットワーク内部の結合が新たに生成され、さらにはすべての単語表現が変わってしまうため、これまでの学習結果に影響を及ぼしてしまうことが予想される．すなわち新たな単語を追加するたびに、これまでの学習結果に影響を及ぼしてしまうのは言語獲得モデルとしてエルマンネットを拡張していくうえで解決しなければならない課題である．そして学習中においては、直前の中間層の活性状態のみを文脈層にフィードバックしているため下記の課題も生じる．

### 課題（3）

直前もしくは数ステップ前に入力された単語情報のみしか蓄積されないため、長文の場合、過去の入力情報が蓄積されていない（Low Memory Accuracy）．

エルマンネットを用いて関係節を含んだ複雑な文章構造の学習が可能であることが示されているものの、長期にわたる入力情報が正確に蓄積されていかず、古い過去の入力情報が必要な場合の予測が困難であることが指摘されている．これについては中間層（文脈層）のニューロン数を増やし、直前の中間層の値だけでなく、過去の時点における中間層の値も記憶しておきフィードバックさせる<sup>10)</sup>、あるいは現時点の入力情報だけでなく、過去の入力情報も利用する<sup>11)</sup>といったように過去の情報のバッファを増やすことで対処している研究事例もある．また Cascade Correlation<sup>14)-15)</sup>のように、学習過程において段階的に中間層の階層を増やしていくことによって、解決すべき問題に対して適切な中間層の階層を持つネットワークの構造を決定する学習方法も提案されている．

そこで本論文においては構造上、課題（1）（2）に対応が容易な、すなわち新たな単語が表出した際に局所的な構造の変更で済むようにエルマンネットを構造上の点から拡張し、文法獲得の精度を向上させる工夫をすることによって、特に課題（3）の Low Memory Accuracy の問題に対応した語系列予測モデルおよびその学習方法を提案する．これはエルマンネットの構造において下記の3点を拡張することで実現を試みる．

- （1）単一のエルマンネットの機能を複数のネットワークを用いて分散処理を図ることによって単一のネットワークの規模を小さくし、単語数の増加に対応が容易なものとする．
- （2）学習で用いる例文中に表出する単語の接続情報（バイグラム）を利用し、各ネットワークの入出力構造を決定することによって、新たな単語が出現した場合においても

局所的な構造での変更のみで対応でき、他のネットワークの構造に影響を与えないようにする。

- (3) 本論文においては、直前の中間層の値のみをフィードバックするという制約の下で、中間層のニューロン数（過去の情報のバッファ）や階層の増加によって、上記の問題を解決するのではなく、蓄積される入力情報を単語の履歴情報と文脈情報に分けて学習することによって解決を試みる。これは中間層から文脈層へのフィードバックについて単語の履歴情報をフィードバックする自己フィードバックと文脈情報をフィードバックする文脈フィードバックに分け、さらにネットワークの内部構造の局所結合化を図り、2つの情報を別々に学習することによって実現し、文法獲得の精度の向上および Low Memory Accuracy の改善を試みる。

そして新たな単語が表出するごとにネットワークを逐次増やし、ネットワークごとで協調しながら学習を行う手法を提案する。

### 3. 提案モデルの構造

#### 3.1 ネットワークの分散処理

エルマンネットを用いて次単語予測学習を行う場合、構造上、出力層と入力層には総単語数分のニューロンが必要となる。そこで単一のネットワークでその機能処理するのではなく、複数個のネットワークにおいて分散処理を行い、1つのネットワークの規模を小さくすることを試みる。

まず出力層についてはニューロンを1個と制限し、特定の1単語と対応づけ、特定のネットワークは対応した1単語のみの予測を行うことを試みる（図2、分散型モデルと呼ぶ）。したがって、次単語予測学習において  $N$  個の単語を含む例文を対象とする場合、各単語を処理するために  $N$  個のネットワークが必要となる。入力層、中間層、文脈層についてはエルマンネットの構造と同様である。このように多出力のネットワークの機能を出力層が1つしかない複数のネットワーク群に分散し協調処理するモデル<sup>12)-13)</sup> はパターン認識の分野では実現され、その有効性も確認されている。

分散型モデルを用いて次単語予測学習を行う場合、エルマンネットと同様に文頭から順に単語を入力していく。その際、単語を  $N$  個のネットワークに同一の単語表現で同時に入力する。そして次単語と対応するネットワークの出力層のニューロンのみを発火させ、それ以外は発火させないように教師信号を与え、各ネットワークごとで結合係数の修正を行う。たとえば図2は *Mary sees John* まで入力した時点を示しているが、次には *who* の単語表

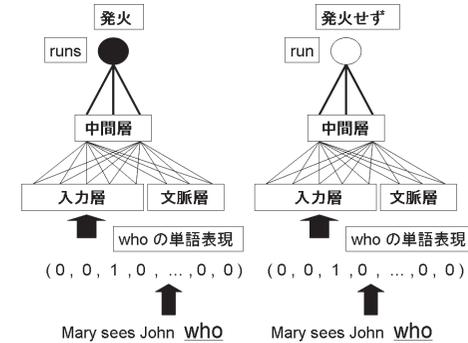


図2 分散型モデル

Fig. 2 A model of distributed networks.

現を  $(0, 0, 1, 0, \dots, 0, 0)$  としてすべてのネットワークに入力し *runs* を出力層の対象としたネットワークのみを発火させ、それ以外は発火させないように学習する。直前の中間層の値を文脈層へ戻すフィードバックについては3.3節で述べる。

分散型モデルにおいて、新たな単語が表出した場合、その単語を処理するネットワークを新たに生成することになる。この場合、入力層についてはエルマンネットと同じ構造のため、すべてのネットワークの入力層において新たな単語を対象としたニューロンを1個ずつ増やさなければならず、すべての単語表現が変わることになる。そこでさらに入力層についても入力する単語情報を制約することによってこの問題の解決を試みる。

#### 3.2 単語の接続情報の利用

次に入力層については学習例文中の単語の接続情報（バイグラム）を利用し、その対象数を制限することを試みる。学習例文中に表出する単語の接続情報より、単語  $i$  の直前に出現する単語集合  $S_i$  を調べる。そして単語  $i$  を出力層の対象とするネットワークの入力層は単語集合  $S_i$  のみを入力対象とする（図3、Bigramモデルと呼ぶ）。したがって、同じ単語を入力しても、各ネットワークにおいて入力層の対象となる単語は異なるため、その単語表現はそれぞれ異なる。

たとえば図3において *runs* を出力層の対象としたネットワーク（図3左図）では、入力の対象は *boy, girl, dog, cat, Mary, John, who* であり、*who* を入力する場合は  $(0, 0, 0, 0, 0, 0, 1)$  という単語表現となる。一方で *run* を対象としたネットワーク（図3右図）では入力の対象は *boys, girls, dogs, cats, who* であり、*who* は  $(0, 0, 0, 0, 1)$  という

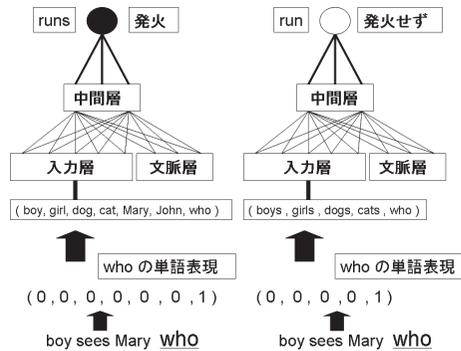


図 3 Bigram モデル  
Fig. 3 A model of distributed networks by bigram.

単語表現となる。

Bigram モデルにおいて次単語予測学習を行う場合、まず現時点の単語を  $N$  個のネットワークに入力する。その際、これまでのモデルとは異なり各ネットワークごとに単語表現に変換し入力する。そして分散型モデルと同様に次単語に対応した出力層を持つネットワークのみを発火させ、その他のネットワークは発火させないように学習を行う。

また学習中において新たな単語が表出した場合、たとえば *Mary sees* から新たな単語として *John* が表出し、これを予測するものとする。この場合、*John* を出力層の対象とし *sees* を入力層の対象としたネットワークを新規に生成する。そしてさらに次単語予測を進めて *Mary sees John* から既知の単語 *who* が表出し、これを予測するものとする。この場合、*who* を出力層の対象としたネットワークの入力層には *John* を対象としたニューロンは存在しないため、入力層のニューロンを 1 つ増やしその対象を *John* とする。このような構造とすることによって、新たな単語が表出した場合、新たな単語および接続する単語を対象としたネットワークのみに影響が生じるだけで済むものと考えた。さらに単語数が増加した場合においても、入力層のニューロン数が制限できるという利点も持つ。

一方でエルマンネットや分散型モデルと比較して、入出力構造に制約を設けており単語特徴に関する入力情報をより与えていると考えられるが、実際には欠落した単語情報を入力していることになる。たとえば図 3 左図の *runs* を対象としたネットワークにおいて、*boys*, *girls*, *dogs*, *cats* を入力した場合、これらの単語を入力層での対象としていないため、単語表現は  $(0, 0, 0, 0, 0, 0, 0)$  とすべて 0 ベクトルになってしまう。Bigram モデルにおいて通常

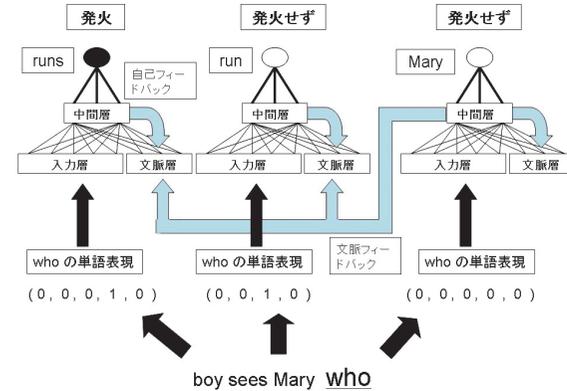


図 4 文脈層へのフィードバック  
Fig. 4 Loopbacks from hidden layer to context layer.

のエルマンネットと同様に中間層から文脈層へのフィードバックのみを考慮し、個々のネットワークを独立して学習するだけでは上記のように各ネットワークごとに、入力層で対象としていない単語を入力した際、その単語情報が蓄積されていない。そこで中間層から文脈層へのフィードバックについてはネットワークの協調学習と関連させて解決を試みる。

### 3.3 文脈層へのフィードバック

エルマンネットは直前の中間層の状態を文脈層にフィードバックさせることによって、過去の入力情報を蓄積している。本論文で提案するネットワークにおいて、中間層から文脈層へのフィードバックについては次の 2 種類を考案する (図 4)。

- 自己フィードバック

通常のエルマンネットと同様に各ネットワークにおいて、それぞれの直前の中間層の活性状態を文脈層にフィードバックさせる。このフィードバックにより過去に入力された単語情報は蓄積されていく。しかし前節で述べたように Bigram モデルの場合、各ネットワークの入力層はすべての単語を対象としておらず、対象としていない単語を入力した場合、その単語表現が 0 ベクトルになってしまうため、自己フィードバックのみでは過去に入力された情報が欠落するものと考えた。

- 文脈フィードバック

提案モデルにおいては、直前の単語を処理する (出力層を発火させる) ネットワークの中間層の活性状態を伝えていくことによって、各ネットワークの中間層にはその時点ま

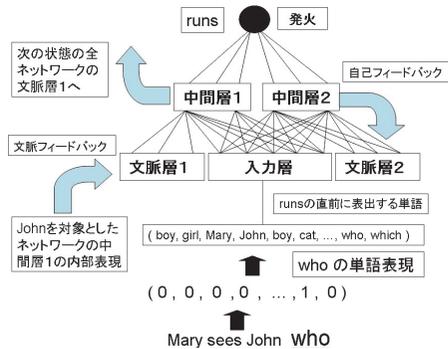


図5 提案モデルの構造

Fig. 5 A proposed model which improves simple recurrent networks.

でに入力された文章の状態が内部表現として保存されていくことになる。そこで直前の単語を処理するネットワークの中間層の活性状態を、すべてのネットワークの文脈層にフィードバックする。たとえば図4においては、*boy sees Mary* から *who* を次単語として予測(学習)する場合、直前の *Mary* を対象とするネットワークの中間層の値をすべてのネットワークの文脈層へフィードバックする。

しかしこの場合、中間層と文脈層には単語の履歴情報と文脈情報という2種類の情報が混在し、自己フィードバック、文脈フィードバックによって2種類の情報を別々に蓄積していくことができない。そこで文脈層と中間層を2つに分割し、自己フィードバックと文脈フィードバックにより、2つの中間層と文脈層をそれぞれ別々に結合した構造とする(図5)。

すなわち図5に示すように、入力層はすべての中間層と結合しているが、2つの文脈層はそれぞれ対応した中間層とのみ結合した局所結合型の構造とする。このような構造とすることにより中間層1に蓄積される文脈情報は文脈フィードバックによって伝達され、中間層2に蓄積される過去の単語の履歴情報は自己フィードバックによって別々に伝達される。このように蓄積する情報を区分し、2種類の情報を混在させずに学習することによって文法獲得の精度を向上させ、過去の入力情報をより長く蓄積することができるのではないかと考えた。

#### 4. 学習アルゴリズム

提案モデル(図5)におけるネットワークの次単語予測学習アルゴリズムは下記のとおり

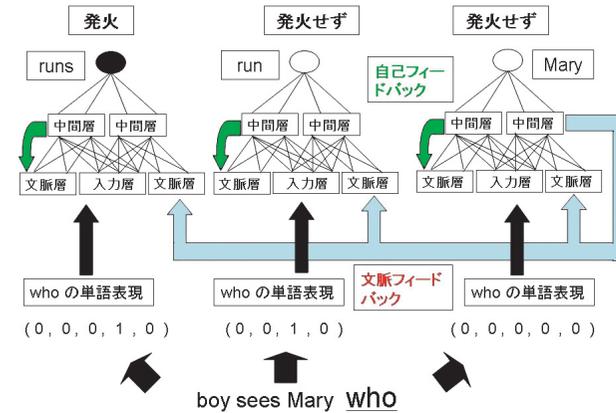


図6 提案モデルによる次単語予測学習

Fig. 6 Learning of word sequence prediction by proposed networks.

である(図6)。

##### ● 各ネットワークの構造

総単語数  $N$  個の文章を学習対象とした  $N$  個のネットワークの状態を想定する(初期状態は  $N = 0$  から開始する)。単語  $i$  ( $i = 1, 2, \dots, N$ ) を対象とした出力層のネットワーク  $i$  において、入力層の対象を単語  $i$  の直前に表出する集合  $S_i$  とする。したがって入力層のニューロン数は集合  $S_i$  の要素数である。またすべてのネットワークにおいて出力層のニューロン数は1個であり、2種類の中間層と文脈層のニューロン数はそれぞれ  $M$  個とする(合計で  $2M$  個)。

##### ● 各ネットワークの学習

(1) 学習に用いる例文数は  $J$  個とする。入力文章

$L_j = \{w_j(0), w_j(1), \dots, w_j(t), \dots, w_j(T)\}$  (例文数  $j = 1, 2, \dots, J$ ) を単語単位  $w_j(t)$  に区切る。 $j = 0$  として例文の個数分だけ以下の(2)から(8)を繰り返す。

(2) 初期状態を  $t = 0$  とする。すべてのネットワークの中間層の値を初期化する(0とする)。

(3) 現時点の単語  $w_j(t)$  および次単語  $w_j(t+1)$  が新たな単語かどうかを調べる。

- 次単語  $w_j(t+1)$  を対象とした出力層のネットワークが存在しない場合、単語  $w_j(t+1)$  を出力層の対象とするネットワークを1個生成する( $N = N + 1$  と

する). そのネットワークの入力層には単語  $w_j(t)$  を対象としたニューロンを 1 個生成する. ネットワーク内の結合係数は乱数によって初期化する.

- また次単語  $w_j(t+1)$  を対象とした出力層のネットワークにおいて, 入力層に単語  $w_j(t)$  を対象としたニューロンが存在しない場合は, 単語  $w_j(t)$  を対象としたニューロンを入力層に追加する. 新たに追加された結合係数は乱数によって初期化する.
- (4) 学習対象とするネットワーク  $k$  ごとに単語  $w_j(t)$  を単語表現  $V_{kj}(t)$  に変換し, 入力層に入力する.
  - (5) 文脈フィードバックによって, 直前の単語  $w_j(t-1)$  に対応したネットワーク  $m$  の中間層 1 の値  $H_m(t-1)$  をすべてのネットワークの文脈層 1 にそのままコピーする.
  - (6) 自己フィードバックによって, 各ネットワーク  $k$  ごとに直前の中間層 2 の値  $H_k(t-1)$  を文脈層 2 にそのままコピーする.
  - (7) 各ネットワーク  $k$  を稼働させ, 出力値  $O_k(t)$  を求める. 出力層は次単語  $w_j(t+1)$  に対応したネットワークは発火し, それ以外は発火しないように各ネットワークごとに教師信号  $T_k(t)$  を与える. そして出力値との誤差の自乗和  $E_{kj}(t) = (O_k(t) - T_k(t))^2$  を求め, この値が小さくなるようにネットワーク内の結合係数の修正を行う. 学習には誤差逆伝播則 (バックプロパゲーションアルゴリズム) を利用する.
  - (8) 単語  $w_j(t+1)$  が文末でなければ  $t = t+1$  とし, (3) へ戻る. 文末であれば  $j = j+1$  とし (2) へ戻り, 別の例文で学習する.

## 5. 次単語予測

次単語予測の方法は以下のとおりである. 学習後の  $N$  個のネットワークを用いて行う.

- (1) 入力文章  $L = \{w(0), w(1), \dots, w(t), \dots, w(T)\}$  を単語単位  $w(t)$  に区切る.
- (2) 初期状態を  $t = 0$  とする. すべてのネットワークの中間層の値を初期化する (0 とする).
- (3) 現時点の単語  $w(t)$  を各ネットワーク  $k$  ( $k = 1, 2, \dots, N$ ) ごとに, 単語表現  $V_k(t)$  に変換し入力層に入力する. そして学習アルゴリズムの (5) および (6) と同様に自己フィードバック, 文脈フィードバックによりそれぞれの中間層の値を対応した文脈層へコピーする.
- (4) 各ネットワーク  $k$  を稼働させ, 出力値  $O_k(t)$  を求める. そして最大発火しているネットワークの対象としている単語を予測結果とする.

- (5) 次単語  $w(t+1)$  が文末でなければ  $t = t+1$  とし, (3) へ戻る. 文末であれば終了する.

## 6. 評価実験

### 6.1 学習例文の生成規則

語系列予測課題学習を通して提案モデルの評価を行う. 本論文においては, 学習対象の例文中において, 単語数が増加した場合そして特に文法構造が複雑になった場合, エルマンネットと提案モデルとの文法規則獲得の精度の比較を目的としている. そこで, エルマンのシミュレーションで使用された人工文の生成規則<sup>2)</sup> を下記の点で拡張した.

- do などの助動詞および what などの疑問詞で始まる疑問文
  - 関係詞 which, who, whom による関係節
  - 単語数の増加
  - 単語間の共起関係 (意味的制約) の複雑化
- 人工文の生成規則を表 1, 表 2, 表 3 に示す.

生成規則は, 単語系列の語順を決める統語規則 (表 1) と単語の出現確率を決める語彙規則 (表 2) そして文法, 意味的制約 (表 3) から構成される. 表 1 の生成規則中, 品詞や句の生成において, その構成要素の出現確率も同じとする. ただし名詞の後に関係節 (RC)

表 1 学習例文の生成規則  
Table 1 Grammers.

S	→	S1   S2   S3
S1	→	NP VP .
S2	→	Aux NP VP ?
S3	→	WH NP VP ?   WH Aux NP VP ?
RC	→	(who   which) VP   (which   whom) NP VP
NP	→	N   N RC
VP	→	V1   V2 OP
OP	→	O   O RC
N	→	N1   N2
O	→	N1   N2   N3
N1	→	Human   Animal
N2	→	Humans   Animals
N3	→	Drink   Food1   Food2   Book   Fragile
V1	→	Intrans_v   Intrans_vs
V2	→	Trans_v   Trans_vs

表 2 語彙規則  
Table 2 Meta-words.

メタワード	単語数	メタワード中に含まれる単語の例
Human	8	boy girl PronN1
Humans	8	boys girls PronN2
PronN1	4	John Mary
PronN2	4	John&Mary
Animal	4	cat dog
Animals	4	cats dogs
Drink	4	milk water
Food1	4	cake fruits
Food2	4	pet_food fruits
Book	4	book newspaper
Fragile	4	dish box
Intrans_v	12	walk work
Intrans_vs	12	walks works
Trans_v	16	see hear feed drink eat read break think
Trans_vs	16	sees hears feeds drinks eats reads breaks thinks
Aux	2	do does
WH	5	who what whom where when

表 3 文法・意味的制約の一部（上表は動詞，下表は助動詞）  
Table 3 Grammatical and semantic constraints.

動詞における制約	動詞	主語	目的語
原型動詞	walk	N1	
三単元動詞	walks	N2	
自動詞での制約	walks	N1	
	works	Human	
	thinks	Human	
他動詞での制約	sees	N1	N1 N2
	feeds	Human	Animal Animlas
	drinks	N1	Drink
	eats	Human	Food1
	eats	Animal	Food2
	reads	Human	Book
	breaks	N1	Fragile
	thinks	Human	N1 N2

助動詞の使い分け	助動詞	主語	動詞
単数名詞	does	N1	Intrans_v Trans_v
複数名詞	do	N2	Intrans_v Trans_v

が接続する確率は 20%としている。表 2 の語彙規則について、各メタワード中に含まれる単語は等確率で表出する。単語数は文頭を示す (HEAD) と文末の (平叙文の場合), ? (疑問文の場合) そして関係詞 (which, whom, who) を含めて 111 個である。

文法的に簡単な例文から学習を行うようにするため、下記に示す順序でネットワークへ提示した。

- (1) 単文 (S1)
- (2) 単文 (S1), 疑問文 (S2)
- (3) 単文 (S1), 疑問文 (S3)
- (4) 単文 (S1), 疑問文 (S2, S3)
- (5) 制約を設けた関係節 (RC)<sup>\*1</sup>を含む文 (S1)
- (6) 関係節 (RC) を含む文 (S1)
- (7) 制約を設けた関係節 (RC) を含む文 (S1), 疑問文 (S2)
- (8) 制約を設けた関係節 (RC) を含む文 (S1), 疑問文 (S3)

\*1 目的語のみ関係節が接続し、主語には接続しない。

(9) 制約を設けた関係節 (RC) を含む文 (S1), 疑問文 (S2, S3)

(10) 制約を設けない文 (S1, S2, S3)

(1)~(4) の例文は関係節を含まない単文のみで構成されており、それ以外は関係節を含む複文も学習例文に含まれる。(10) は生成規則によって生成されるすべての文が含まれる。(1)~(8) まではそれぞれ 20,000 文, (9), (10) についてはそれぞれ 30,000 文を作成し、合計 220,000 文を学習用の例文とする。そして (10) について別に 30,000 文を用意し、評価用の例文とする。学習例文の 1 文あたりの平均単語数は 4.63 語、評価例文の場合は 5.80 語であった。また例文の長さは最短が 3, 最長が 18 であった。

## 6.2 評価方法

次単語予測において、次に提示される単語を予測するということは、次に表出すると考えられる単語の出現確率を予測することになる。本論文においては、以下の指標を用いて学習後のネットワークの評価を行う。

- cosine による評価
- 予測順位による評価

cosine はエルマンによって提案されている評価指標である。まず学習例文より、単語  $x$

の次に表出する単語の出現確率ベクトル $v_x$  (111次元)を求めておく。また単語 $x$ を学習後のネットワークに入力したとき、全ネットワークからの出力ベクトル $y_x$  (111次元)を次に表出する単語の出現確率の予測値とする。そして全ネットワークの出力ベクトル $y_x$ が、学習例文における単語の出現確率ベクトル $v_x$ にどの程度近似できているかを $COS_w$ を用いて調べる(式(1))。

$$COS_w(x) = \frac{y_x^t v_x}{|y_x| |v_x|} \quad (1)$$

また単語 $x$ の次に出現する単語を予測する場合、特定の1個の単語を予測するのではなく、その単語と同じく出現する単語カテゴリを予測することができればよいことになる。そこで、単語をカテゴリ(カテゴリ数はメタワードをさらに細かく分け37個とした)に分け、 $y_x$ からカテゴリごと出現確率をまとめて算出した $p_x$  (37次元)を次に表出する単語カテゴリの出現確率の予測値とする。そして学習例文より、単語 $x$ の次に表出する単語カテゴリの出現確率ベクトル $c_x$  (37次元)を求めておき、ネットワークの予測値 $p_x$ が、単語カテゴリの出現確率ベクトル $c_x$ にどの程度近似できているかを $COS_c$ を用いて調べる(式(2))。

$$COS_c(x) = \frac{p_x^t c_x}{|p_x| |c_x|} \quad (2)$$

以上2つの指標は学習例文における次単語の出現確率にいかに正しく近似できているかを調べる指標である。一方で語系列予測課題学習の目的は、学習例文中における次単語の出現確率を獲得することよりも、学習例文を生成する文法規則(生成規則)を獲得することにある。したがって、単語 $x$ の次に出現する単語は、生成規則のうえで正しい単語であれば正しく予測していることになる。たとえば *Mary who* と入力した場合、次に続く単語は三人称単数の動詞(Intrans\_vs と Trans\_vs)の中に含まれる単語であればどの単語でも生成規則のうえでは正しいことにもなる。そこで学習例文の生成規則より、単語 $x$ の次に表出する生成確率ベクトル $g_x$  (111次元)を求めておく。生成確率ベクトル $g_x$ は、生成規則のうえで単語 $x$ の次に表出する単語に該当するすべての要素の値は1とし、それ以外は0として作成する。前述した *Mary who* と入力した場合、生成確率ベクトル $g_{who}$ は Intrans\_vs と Trans\_vs に該当する単語の要素のみ1となり、それ以外は0となる。そしてネットワークの予測値 $y_x$ が、単語の生成確率ベクトル $g_x$ にどの程度近似できているかを $COS_g$ を用いて調べる(式(3))。

$$COS_g(x) = \frac{y_x^t g_x}{|y_x| |g_x|} \quad (3)$$

各 *cosine* の指標は1に近い値ほど、近似できていることを示す。すなわち $COS_w$ と $COS_c$ によって次単語予測学習の精度を、 $COS_g$ によって生成規則の獲得の精度を調べる。

また任意の単語 $x$ から次単語として単語 $y$ が出現する予測を行う場合、実際にネットワークが単語 $y$ を次単語として予測する順位(予測順位と呼ぶ)も調べる。

### 6.3 次単語予測学習の結果

評価実験では提案モデルとエルマンネット(図1)の比較を行う。提案モデルにおいては、ネットワークは単語数分生成される。各ネットワークは4層型の構造とした。出力層のニューロン数は1個、圧縮層のニューロン数は20個、中間層(文脈層)のニューロン数は変化させる。また入力層のニューロン数は、学習例文の単語の接続関係より決定される。一方でエルマンネットの構造も同様に4層型とした。入力層と出力層のニューロン数は単語数分必要のため111個、圧縮層のニューロン数は20個、中間層(文脈層)のニューロン数は同じく変化させる。提案モデルおよびエルマンネットの学習にはバックプロパゲーションアルゴリズムを用いた(学習係数は0.5とし、さらにモーメント法<sup>16)</sup>を利用(モーメント係数は0.9)した)。

まず提案モデルおよびエルマンネットの中間層と文脈層のニューロン数を変化させ、生成規則より作成した220,000文の例文を学習させた。提案モデルについては、各ネットワークにおいて2種類の間層を合わせて、ニューロン数を32個、64個、128個と変化した(2種類の間層のニューロン数は同数としている)。エルマンネットについては、64個、96個、128個、160個、192個、224個、256個と変化した。220,000文の学習例文を用いて学習を行った後、学習例文および30,000文の評価用の例文を用いて、各指標の値を求めた。これを1サイクルとして続けて20回繰り返した。各指標については、各単語ごとで求め、その平均値を求めた。さらに各ネットワークの結合係数は乱数(-0.5~0.5)によって初期化しており、その初期化の仕方によっては学習後のネットワークの挙動に差異が生じる。そのため、上記の手順を独立に5回行い、各指標は5回分の平均値を求めた。その結果を表4、表5に示す。

表4、表5には提案モデル、エルマンネットにおいて、中間層(文脈層)のニューロン数を変化させた場合の学習例文、評価例文に対する予測順位、 $COS_w$ 、 $COS_c$ 、 $COS_g$ の平均値を示す。予測順位については標準偏差も示す(標準偏差についても5回ごとの学習でそれぞれ求め、その平均値を示している。括弧内の値が標準偏差を示す)。予測順位の値が7を示すということは、任意の文において次単語を予測した場合、その次単語の表出順位が第7位であったことを意味する。たとえば *Mary* の後に動詞が続く場合、3単元動詞であれば、

表 4 提案モデルによる学習結果

Table 4 Results of prediction by proposed neural networks.

中間層のニューロン数	学習例文			
	予測順位	$COS_w$	$COS_c$	$COS_g$
64	6.71(9.18)	0.791	0.870	0.877
96	6.53(8.92)	0.775	0.862	0.889
128	6.32(8.40)	0.766	0.855	0.891
中間層のニューロン数	評価例文			
	予測順位	$COS_w$	$COS_c$	$COS_g$
64	8.38(13.04)	0.764	0.796	0.808
96	7.98(12.55)	0.737	0.768	0.827
128	7.58(11.44)	0.722	0.754	0.832

表 5 エルマンネットによる学習結果

Table 5 Results of prediction by a simple recurrent network.

中間層のニューロン数	学習例文			
	予測順位	$COS_w$	$COS_c$	$COS_g$
64	6.99(10.61)	0.734	0.755	0.806
96	6.91(10.30)	0.727	0.745	0.799
128	7.19(11.32)	0.726	0.745	0.803
160	7.01(10.55)	0.718	0.736	0.800
192	6.89(10.56)	0.718	0.738	0.804
224	7.20(11.65)	0.718	0.738	0.797
256	7.14(11.24)	0.720	0.741	0.803
中間層のニューロン数	評価例文			
	予測順位	$COS_w$	$COS_c$	$COS_g$
64	8.62(14.75)	0.691	0.713	0.782
96	8.14(14.01)	0.683	0.702	0.782
128	8.46(14.80)	0.675	0.695	0.787
160	8.09(13.56)	0.673	0.692	0.789
192	8.01(13.78)	0.671	0.692	0.794
224	8.67(15.54)	0.672	0.694	0.783
256	8.55(15.17)	0.671	0.693	0.787

どの単語でもよい場合、予測順位の値が大きいからといって予測できていないわけではない。

表 4 より、提案モデルにおいては、中間層のニューロン数を増やすと予測順位の平均値と  $COS_g$  は上昇し、 $COS_w$  と  $COS_c$  は下降する傾向にあることが分かる。また予測順位の標準偏差も小さくなることから、予測精度が安定する傾向にある。前節でも述べたように語系列予測課題学習の目的は、現在の単語を提示し、次単語を予測する学習を通して文法規則を獲得することにある。したがって文法獲得の指標である  $COS_g$  の値が高いことが本研究においては重要となる。エルマンネット、提案モデル共にすべての結果において、

$COS_w < COS_c < COS_g$  となっており、次単語の出現確率よりも単語カテゴリの出現確率を、そして単語カテゴリの出現確率よりも文法規則を獲得していることが分かる。

表 5 より、エルマンネットにおいても、中間層のニューロンを増やすと予測順位と  $COS_g$  は上昇し、 $COS_w$  と  $COS_c$  は下降する傾向にあることが分かる。ただし中間層のニューロン数が 192 個以上においては、精度に変化は見られない（予測順位は平均値、標準偏差とも低下する傾向にある）。このようにエルマンネットにおいて、次単語予測の精度を上げるため、中間層のニューロンを単純に増やしていても精度は飽和状態となってしまう（今回の実験においては、中間層のニューロンが 192 個以上の場合）。

表 4、表 5 より、エルマンネットと提案モデルを比較すると、提案モデルにおいて中間層のニューロン数が最も少ない 64 個の場合においても、エルマンネットでも最も指標の高い 192 個の場合と比較して提案モデルの優位性が見てとれる（評価例文における予測順位以外、すべての指標が向上している）。特に提案モデルにおいて中間層のニューロン数が 128 個の場合では、エルマンネット（192 個）と比較して、評価例文を対象とした予測順位は 8.01 から 7.58 に、 $COS_w$  は 0.671 から 0.722 に、 $COS_c$  は 0.692 から 0.754 に、そして  $COS_g$  は 0.794 から 0.832 へと次単語予測および文法規則の獲得の精度を高めることができた。

本論文においては単一のネットワークの規模を小さくした複数のネットワークによる分散処理を行っているが、関連研究として Mixture of Experts<sup>17)-18)</sup> に代表されるアンサンブル学習<sup>19)</sup> があげられる。本論文においては、こうした学習方法とは異なり、各ネットワークごとの出力を制御する統合部（Mixture of Experts の場合、Gating Network）を設けず、各ネットワークが独立に学習を行い、各ネットワークごとの出力値をそのまま最終的な出力結果としている。これは、各ネットワークを統合する制御は、文脈フィードバックによって直前までに入力された文章の状態を内部表現として、各ネットワークに渡すことにより実現しているためである。

以上、提案手法であるネットワークの分散化によって次単語予測が可能であり、さらには精度の向上につながることを示すことができた。また今回の学習例文を用いた場合、学習後のネットワーク 1 個あたりの入力層の平均ニューロン数は 46 個であった。提案モデルのように、入力層にすべての単語に対応したニューロンを配置しなくても、次単語予測が可能であること、単語数が増加した場合においても、入力層のニューロン数の制限が可能であることを示すことができた。また表 6 に、実験に用いたエルマンネット、提案モデルにおける単一のネットワークの結合数を示す（提案モデルにおいては、入力層のニューロン数を平均個数の 46 個として求めている）。提案モデルの場合、ネットワークの分散化を試みている

表 6 単一のネットワークにおける結合数

Table 6 The number of connections in a SRN and a proposed network.

中間層のニューロン数	エルマンネット	提案モデル
32	-	2,644
64	14,700	6,292
96	22,736	-
128	35,372	16,660
160	48,780	-
192	64,236	-
224	81,740	-
256	101,292	-

ため、ネットワークは単語数分生成され、総結合係数は単一のエルマンネットと比較して多くなるが、1個あたりのネットワークの結合数を少なくすることが可能となる。単一のネットワークの規模を大きくしていくだけでは精度の向上には限界があるが、単一のネットワークの規模を小さくしたうえで、ネットワークの分散化を図ることによって精度の向上を図ることができた。

ところで表 4, 表 5 より、中間層のニューロン数が増えると次単語予測の指標である  $COS_w$  と  $COS_c$  が下がる傾向にあることが分かる。そこで、単語  $x, y$  の順番の次に続く単語  $z$  もしくは単語  $z$  が属する単語カテゴリの出現確率の近似  $COS'_w, COS'_c$  を同様に調べた。その結果を表 7, 表 8 に示す。

表 7, 表 8 より、中間層のニューロン数を増やすと、表 4, 表 5 と同様に  $COS'_w, COS'_c$  の指標は下がる傾向にあること、エルマンネットと比較して提案モデルの精度が高いことが分かる。この場合、文法的な制約が強くなるため、いずれの中間層のニューロン数においても  $COS_w < COS'_w, COS_c < COS'_c$  という結果になっている。中間層のニューロン数を増やすと、次単語予測の精度が向上するのではなく、文法規則の獲得の精度が向上していくことが分かる。以後、エルマンネットにおいては、最も精度の高い中間層のニューロン数が 192 個の場合、提案モデルにおいては 128 個の場合について解析を進める。次にエルマンネットおよび提案モデルについて、学習回数ごとの評価例文に対する各指標の推移を図 7 に示す。

図 7 (左上, 右下) より、予測順位および  $COS_g$  についてはエルマンネット、提案モデルとも学習回数が増すにつれ、向上していく傾向にあることが分かる。一方で図 7 (右上, 左下) より、 $COS_w$  と  $COS_c$  については、エルマンネットにおいては微増し、提案モデルにおいては下降していく傾向にある。学習を繰り返すたびに  $COS_g$  が向上することから、2 つ

表 7 エルマンネットによる次単語予測の分析

Table 7 Analysis of prediction by a simple recurrent network.

中間層のニューロン数	学習例文		評価例文	
	$COS'_w$	$COS'_c$	$COS_w$	$COS_c$
64	0.768	0.796	0.722	0.753
96	0.761	0.786	0.717	0.745
128	0.763	0.788	0.713	0.741
160	0.749	0.774	0.703	0.731
192	0.752	0.779	0.702	0.732
224	0.751	0.778	0.703	0.735
256	0.754	0.783	0.701	0.733

表 8 提案モデルによる次単語予測の分析

Table 8 Analysis of prediction by proposed neural networks.

中間層のニューロン数	学習例文		評価例文	
	$COS'_w$	$COS'_c$	$COS_w$	$COS_c$
64	0.844	0.870	0.810	0.844
96	0.836	0.862	0.796	0.828
128	0.829	0.855	0.785	0.818

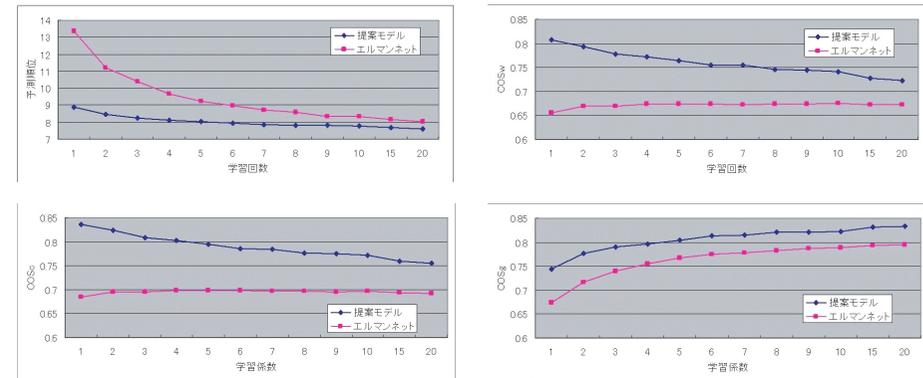


図 7 学習回数ごとの指標の変化 (左上: 予測順位, 左下:  $COS_c$ , 右上:  $COS_w$ , 右下:  $COS_g$ )  
Fig. 7 Results in increasing the number of learning.

表 9 疑問文を対象とした次単語予測の結果  
Table 9 Results of prediction for question sentences.

	予測順位	$COS_w$	$COS_c$	$COS_g$
エルマンネット	7.53	0.669	0.684	0.796
提案モデル	7.03	0.725	0.753	0.828

のモデルとも学習を繰り返すにつれて文法規則を獲得していくことが分かる．提案モデルはエルマンネットと比較して，表 4，表 5 に示したように学習回数が 20 回目の時点で，各指標値が高いたくだけでなく，1 回目の学習終了時においても各指標値が高いことから文法規則の学習速度も速いことが分かる．

#### 6.4 Low Memory Accuracy についての分析

次に Low Memory Accuracy の改善について調べるため，文章が長くなった際の次単語予測の精度についての分析を行う．これについては疑問文中の文末「？」の予測精度を調べることにする．疑問文において文末の「？」を予測するためには，文頭が助動詞もしくは疑問詞であることを文脈層に最後まで記憶しておかなければならないからである．

そこで，生成規則より疑問文（制約は設けない）のみを新たに 30,000 文，評価用として作成し，前節で述べた学習済みのネットワーク（提案モデルについては，中間層のニューロン数は 128 個，エルマンネットは 192 個の結果を利用）を用いて次単語予測を行った．生成した疑問文 1 文あたりの平均単語数は 6.23 語であり，長さは最短が 3，最長が 17 であった．表 9 に，疑問文を対象とした次単語予測における各指標値を示す．

表 9 より，これまでの結果と同様に評価例文の対象が疑問文のみとなっても，提案モデルの方がエルマンネットよりも次単語予測および文法獲得の精度が高いことが分かる．次に疑問文の文末「？」の予測順位を調べた．この結果を図 8 に示す．

図 8 は，文末の位置ごとに出現する「？」の予測順位を示している．図 8 より，提案モデルにおいては，文末が長くなっても予測順位は 2 未満となっており，予測精度は安定していることが分かる．予測順位が 2 未満となるのは，疑問文において「？」を予測する際，その候補は「？」もしくは名詞の後に関係代名詞が接続するかのどちらかになるためである．エルマンネットは文末の長さが 9 までは正しく予測できているが，それ以上の長さになると，予測順位は悪化し正しく予測できないことから，文脈層には正しく文頭の情報を蓄積できていないことが分かる．一方で提案モデルにおいては，文章が長くなった場合においても，正しく予測できていることから，文頭の単語情報を正しく文末まで蓄積できていることが分かる．中間層および文脈層を 2 つに分け，自己フィードバック，文脈フィードバックにより

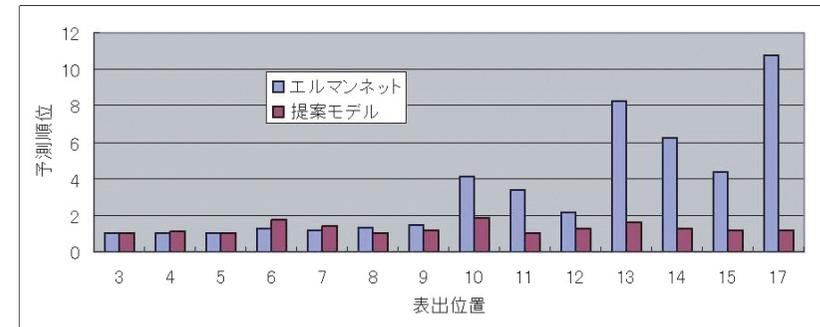


図 8 疑問文における文末の予測結果

Fig. 8 Results of prediction for '?' in question sentences.

単語の履歴情報と文脈情報を混在せずに学習させる方法をとった提案モデルにおける Low Memory Accuracy の改善を示すことができた．

## 7. む す び

本論文においてはエルマンネットを拡張した語系列予測モデルおよびその学習方法を提案した．これは単一のエルマンネットの機能を分散処理することによって実現を試みた．すなわちネットワークの出力を 1 単語として制限し，特定のネットワークは対応した 1 単語の処理を行うことにした．さらに学習例文中における単語の接続情報を利用してネットワークの入力構造を決定した．そしてネットワークの内部構造の局所構造化を図り，単語の履歴情報と文脈情報を自己フィードバック，文脈フィードバックによって区分して学習する方法を提案した．語系列予測課題学習による実験の結果，エルマンネットと比較して提案モデルにおいては次単語予測および文法獲得の精度が高く，さらに過去の入力情報をより蓄積できることを示すことができた．今後は提案モデルにおいて，中間層の内部表現の分析を行い，文法規則の獲得や汎化についての解析をすすめていく予定である．

## 参 考 文 献

- 1) Elman, J.L.: Finding Structure in Time, *Cognitive Science*, Vol.14, pp.179–211 (1990).
- 2) Elman, J.L.: Distributed representations, simple recurrent networks and grammat-

- ical structure, *Machine Learning*, Vol.7, pp.195–224 (1991).
- 3) Elman, J.L.: Learning and development in neural networks: The importance of starting small, *Cognition*, pp.71–99 (1993).
  - 4) 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, No.3, pp.359–368 (1999).
  - 5) 上原謙吾, 原田賢一: 後続要素の予測を利用する系列予測モデル, 第4回情報科学技術フォーラム (FIT2005) 論文集, H-014 (2005).
  - 6) 森藤大地, 乾 敏郎: 自己組織化マップを含む Elman 型ネットワークを用いた文法知識の獲得モデル, *認知心理学会第1回大会* (2003).
  - 7) 下斗米貴之, 遠山修治, 大森隆司: 文法メタ知識による語彙学習加速のコネクショニストモデル, *認知科学*, Vol.10, No.1, pp.104–117 (2003).
  - 8) Mayberry, M.R. and Mikkilainen, R.: SARSDRN: A Neural Network Shift-Reduce Parser, *Proc. 16th Annual International Joint Conference on Artificial Intelligence*, pp.820–827 (1999).
  - 9) Rumelhart, D.E. and McClelland, J.L.: *Parallel Distributed Processing*, MIT Press, Vol.1 (1986).
  - 10) 島田 智, 原田耕治, 白鳥則郎: 遅延型再起的ニューラルネットにおける学習過程の研究, *電子情報通信学会技術研究報告*, NLP2002-113, pp.37–42 (2003).
  - 11) 菊地進一, 中西正和: 短期記憶を用いたリカレントニューラルネットワークと高速な構造学習法, *電子情報通信学会論文誌*, Vol.J84-D-II, No.1, pp.159–169 (2001).
  - 12) Saruta, K., Kato, N., Abe, M. and Nemoto, Y.: High Accuracy Recognition on ETL9B Using Exclusive Learning Neural Network-II (ELNET-II), *IEICE Trans.*, Vol.E79-D, No.5, pp.516–522 (1996).
  - 13) 石井 力, 和泉勇治, 加藤 寧, 根元義章: ニューラルネットワークを用いた二者択一方式の手書き文字認識システム, *電子情報通信学会論文誌*, Vol.J83-D-II, No.3, pp.988–995 (2000).
  - 14) Fahlman, S.E. and Lebiere, C.: The cascade correlation learning architecture, *Advances in Neural Information Processing Systems 2*, Touretzky, D.S. (Ed.), pp.524–532, Morgan Kaufman (1990).
  - 15) Fahlman, S.E.: The Recurrent Cascade-Correlation Architecture, *Advances in Neural Information Processing Systems 3*, Touretzky, D.S. (Ed.), pp.190–196, Morgan Kaufmann (1991).
  - 16) 中野 馨: 入門と実習 ニューロコンピュータ, 技術評論社 (1988).
  - 17) Jacobs, R.A., Jordan, M.I., Nowlan, S.J. and Hinton, G.E.: Adaptive mixtures of local experts, *Neural Computation*, Vol.3, pp.79–87 (1991).
  - 18) Jordan, M.I. and Jacobs, R.A.: Hierarchical mixtures of experts and the EM algorithm, *Neural Computation*, Vol.6, pp.181–214 (1994).
  - 19) 上田修功: アンサンブル学習の新展開, *電子情報通信学会技術研究報告*, NC2002-49, pp.31–46 (2002).

(平成 20 年 9 月 5 日受付)

(平成 21 年 5 月 13 日採録)



篠沢 佳久 (正会員)

1994 年慶應義塾大学理工学部管理工学科卒業。1996 年同大学大学院理工学研究科管理工学専攻前期博士課程修了。1999 年同大学院後期博士課程修了。博士(工学)。機械学習, パターン認識に興味を持つ。現在, ニューラルネットワークを用いた言語獲得の研究に従事, 情報システム学会, 電子情報通信学会各会員。