

TR-I-0176

ニューラルネットによる英単語品詞列予測モデル  
*English Word Category Prediction Based on  
Neural Network*

中村雅己、鹿野清宏<sup>†</sup>

*M. Nakamura, K. Shikano*

1990.9

概要

文章中の単語の音声認識結果の誤りを訂正する方法として、Trigram等の確率モデルによる方法があるが、有効な統計量を獲得するために膨大なデータを必要とし、単語組数を多くすれば、計算機の記憶容量が指数的に増加するという問題がある。そこで我々は、文章中の単語間に内在する局所的な言語情報を内部に獲得し、少ないパラメータで次の単語を予測する構造を持つ、ニューラルネットによる単語品詞列予測モデル(NETgram)を提案する。本論文では、NETgramの構成、英文テキストデータ(Brown Corpus)を用いた予測実験結果および英単語音声認識の改善実験結果について述べる。予測実験の結果、Trigram予測においてNETgramはTrigram確率モデルより優れた予測的中率を示し、Bigramで補間処理された確率モデルに匹敵する予測性能を有することがわかった。また、中間層の出力パターンを解析した結果、単語品詞が言語的に意味のある5つのグループに分類された。

Abstract

Word category prediction is used to implement an accurate word recognition system. Traditional statistical approaches require considerable training data to estimate the probabilities of word sequences, and many parameters to memorize probabilities. To solve this problem, NETgram, which is the neural network for word category prediction, is proposed. Training results show that the performance of the NETgram is comparable to that of the statistical model although the NETgram requires few parameters than the statistical model. Also the NETgram performs effectively for unknown data, i.e., the NETgram interpolates sparse training data. Results of analyzing the hidden layer show that the word categories are classified into some linguistically significant groups.

ATR 自動翻訳電話研究所

ATR Interpreting Telephony Research Laboratories

© ATR 自動翻訳電話研究所

© ATR Interpreting Telephony Research Laboratories

<sup>†</sup> 現在、NTTヒューマンインタフェース研究所勤務

## 目次

1. まえがき	3
2. NETgram(単語列予測ニューラルネットモデル)の設計	4
2.1 Bigramネットワークモデル	5
2.2 N-gramネットワークモデルへの拡張	6
3. NETgramの学習方法	7
3.1 Bigramネットワークモデルの学習	7
3.2 Trigramネットワークモデルの学習	7
4. NETgram学習結果	8
4.1 Bigram予測実験結果	8
4.2 Trigram予測実験結果	9
4.3 NETgramと従来の確率モデルの比較	10
4.4 4-gramネットワークモデルについて	12
4.5 英単語音声認識の改善結果	12
5. むすび	14
図	16
表	24
参考文献	25

謝辞

研究の機会を与えて下さったATR自動翻訳電話研究所樽松社長に、感謝します。また、貴重な御助言、御討論をして頂いた音声情報処理研究室、言語情報研究室、データ処理研究室の皆様に、感謝します。

## 1. まえがき

現在、ATRでは自動翻訳電話システムのための基礎研究を行っているが、その実現のためには精度の高い音声認識システムが必要とされる。音響情報のみで、ある程度精度の良い音声認識は可能であるが、より高い精度を追求した場合、言語情報の利用が不可欠なものとなる。たとえば、文章中の単語の音声認識結果の誤りには英語の場合、“I”と“by”、“the”と“be”等の1シラブルのものも多くあるが、これらの半分以上が局所的な構文知識を用いることで、訂正することができる。

この局所的な構文知識を統計的に扱う方法として、次に続く単語の出現確率を用いて単語列を予測する方法があり、単語の音声認識結果の誤りの訂正に効果があることが報告されている<sup>(1)(2)</sup>。このような単語列予測の場合、過去いくつまでの単語情報を利用するかが、予測能力を向上させるためのキーポイントとなるが、確率情報を用いる従来の方法では、有効な統計量を獲得するために膨大なデータを必要とし、単語組数(gram数)を多くすれば、計算機の記憶容量が指数的に増加するという問題があった。

一方、ある程度限られたデータ量から一般的な特徴、法則を学習する可能性のある手法として、ニューラルネットが注目されている。ニューラルネットは特にパターン認識においてその能力の有効性が確認されているが<sup>(3)</sup>、記号処理においても、英文テキストから音素記号を生成するNETtalk<sup>(4)</sup>のような成功例が報告されている。

そこで著者らは単語列予測にニューラルネットを適用する研究を行った。ねらいは、テキストデータをニューラルネットに学習させることにより、連続する単語間に内在する局所的な言語情報をニューラルネット内部に獲得させ、少ないパラメータで次の単語を予測する構造をニューラルネッ

トにもたせることにある。また、限られた学習データから獲得した特徴がどの程度一般性を持っているか、言い換えると、このニューラルネットが未学習のデータに対してどの程度、予測能力を保持しうるかも興味のあるところである。

本論文ではBigram予測(2つ組予測)モデルからN-gram予測(N組予測)モデルまで容易に拡張が可能で、かつ計算時間、記憶容量も大幅に増加しないような2つのニューラルネットワークモデル(以後NETgramと呼ぶ)を提案し、英文テキストデータベースを用いた学習実験を行い、予測性能の評価を確率モデルとの比較により行う。

## 2. NETgram(単語列予測ニューラルネットワークモデル)の設計

NETgramの設計において考慮した条件を次に示す。

(1) 学習データとして、英文で書かれたBrown Corpusテキストデータベース<sup>(6)</sup>の文章中の各単語に与えられている品詞(以後、カテゴリとよぶ)を用いる。

(2) ニューラルネットの入力層はカテゴリ数(89個)×入力単語数のユニットを持ち、入力単語カテゴリ番号に対応するユニットのみ1の値を持ち、他はすべて0とする局所的表現を用いる。

(3) 出力は予測される次の単語のカテゴリとする。出力もそれぞれのユニットが1つのカテゴリに対応する局所的表現を用いる。よって、ユニット数は89個である。

(4) 学習にはBack-Propagation法<sup>(6)</sup>を用いる。

以上が設計の基本条件であるが、さらに次の条件も考慮に入れる。

(5) 中間層は学習後、入力や出力の特徴をアナログ符号化するような構造とする。

(6) 2組(Bigram)予測から3組(Trigram)、…N組(N-gram)予測まで容易に拡張でき、しかもパラメータ数(リンクウェイト数)がそれにともない大きく増加しないようなネットワーク構造とする。

条件(1)は、学習データが単語レベルだとタスクが限定されないと語彙が多くなり、現実的でなくなるからである。Brown Corpus は文章中の各単語に品詞(89個カテゴリ)が与えられており、データも豊富(約100万語)にあり、同様の問題に過去、取り上げられた実績がある<sup>(2)</sup>。

ニューラルネットの入出力データの表現形式を符号化することにより、入出力ユニットの数を小さくすることができるが、そのためにはカテゴリ間の距離を考慮せざるを得なくなり、何らかの言語情報を用いる必要が生じる。ここでは、それよりはむしろ、条件(2)、(3)のように入出力ユニット1つずつにカテゴリ番号を対応させる局所的表現(Local Representation)をとり、言語知識の無い状態から条件(5)に示すように学習の過程でニューラルネットにカテゴリの特徴を獲得させる方を選んだ。

以上の設計条件を考慮して、次のようなNETgramを提案する。

## 2.1 Bigramネットワークモデル

図2に示すような4層のフィードフォワードネットワーク構造とした。2層の中間層としたのは、それぞれの中間層が入力データ、出力データの特徴のアナログ符号化を学習することを期待するためである。入力に対し出力は次の単語のカテゴリを学習させるため、中間層は単なる単語カテゴリの符号化ではなく、文章内で隣接する単語カテゴリ間の言語情報を持った符号化がなされることが期待できる。

## 2.2 N-gram ネットワークモデルへの拡張

N-gramへの拡張に対し、2つのモデルを提案する。

モデル1は図3に示すように、Bigram ネットワークモデルを横に複数個並べた構造とし、上下方向の結合はBigram ネットワークモデルの全結合リンクウェイトセット(図中 $w_1, w_2, w_3$ )の値をそのまま継承する。横方向の結合は中間層2(以後HL2と呼ぶ; Hidden Layer 2)同士の結合のみとする(リンクウェイトセットは $w_4$ )。

モデル2は図4に示すように、Bigram ネットワークモデルに対し入力単語の増加分だけ入力層を増やしてHL1に結合させてTrigram ネットワークモデルを作る。さらに、Trigram から4-gram への拡張の際はHL1を新たに追加してHL2に結合させる。

これらのモデルの特長は、Bigram ネットワークモデルからN-gram ネットワークモデルへの拡張が容易であることと、Bigram ネットワークモデルのリンクウェイトセット $w_1, w_2, w_3$ をそのままTrigram ネットワークモデルの初期値として用いることができる点である。さらに、モデル1はBigram ネットワークモデルを並列にHL2同士の結合のみを介して構成されたものであるため、入力、出力の単語それぞれに対して独立に特徴のアナログ符号化を学習することができる。また、モデル1はHL2に $w_4$ という自己ループをもったリカレントネットの有限展開型<sup>(7)</sup>と等価である。ただ、モデル1はgram数が多くなると先頭単語(1st word)の入力層から出力層までの層数が多くなり、この間のリンクウェイトの学習が困難となる可能性がある。一方、モデル2ではHL1は隣り合う2つの入力単語の情報を符号化したものとなるが、入力層から出力層へのネットワーク層数は4層一定となる。

### 3. NETgramの学習方法

NETgramの学習方法を図5に示す。入力にはBrown Corpusの文章の先頭の単語から句読点の1つ前までを順に与えていく。出力には入力単語の次の単語を教師データとして順に与え、その誤差(ここでは誤差関数として、NETgram出力値と教師データ値の差の2乗和の1/2を用いる)を最小にするようにBack-Propagation法により、リンクウェイトを修正する。なお、入力データ、教師データの表現形式は、単語カテゴリ番号に対応するユニットのみ1で他はすべて0の局所的表現とする。

#### 3.1 Bigramネットワークモデルの学習

Bigramネットワークモデルの学習予備実験として、3センテンス(約72単語)の学習を行ったが、ネットワークの出力パターンはBigramの確率分布に収束することを確認した。

しかし、データセットが多くなればなるほど収束が極端に遅くなった。これは本問題が多対多の写像問題であるため例外データが多く、リンクウェイトの修正方向が振動し易い構造を持っているからであると考えられる。このため、Bigramネットワークモデルの学習では、予め1024センテンスについて求めておいた次の単語カテゴリの出現確率を教師データとして与えることにする。ただし、Trigramネットワークモデル以降の学習は生データ(0、1の値)を教師データとする。

#### 3.2 Trigramネットワークモデルの学習

モデル1では、Trigramネットワークモデルへの拡張時は、Bigramネットワークモデルから継承したリンクウェイトセット $w_1, w_2, w_3$ のうち、 $w_1, w_2$ を最初は固定とし、新たに追加された横方向のリンクウェイトセット $w_4$ と $w_3$ のみを学習する。ある程度学習が進めば、他のリンクウェイトセッ

ト  $w_1, w_2$  を自由にし、全ての学習を行う。以降、N-gramへの拡張時は(N-1)-gramで学習したリンクウェイトセット  $w_1, w_2, w_3, w_4$  をそのまま継承して学習を行う。

モデル2では、Bigramネットワークモデルで学習したリンクウェイトセット  $w_1, w_2, w_3$  をそのまま継承し、左側(入力単語列の古い方)に入力層を新たに増やし、そのリンクウェイトセット  $w_1'$  の初期値に0を与える。こうすることにより、Bigramネットワークモデルで学習した出力と同じ値から学習を始めることができる。Trigramから4-gramへの拡張の際は、新たにHL2と入力層を増やし、そのリンクウェイトセット  $W_1, W_1'$  の初期値はTrigramネットワークからコピーし、 $W_2'$  の初期値には0を与える。

#### 4. NETgram 学習結果

##### 4.1 Bigram予測実験結果

予め1,024センテンス(1センテンス約24単語)について求めておいた、次の単語のカテゴリの出現確率を教師データとした学習した結果、ネットワークの出力値の大きいユニットから数えてk個の候補の中に、次の単語カテゴリ番号が存在する率、すなわち予測的中率を計算すると、学習用(Closed)データ、評価用(Open)データとも従来の確率モデルによるものと比較し、NETgramの予測精度は同じ程度であった。

Bigram学習のあと、89個の単語カテゴリに対するそれぞれのHL1の出力パターンを分類するために類似度を計算した。2つの入力単語カテゴリ  $c_1, c_2$  間の類似度Sは次式により計算した。

$$S(c_1, c_2) = \frac{(M(c_1), M(c_2))}{\|M(c_1)\| \|M(c_2)\|} \quad (1)$$

ここで $M(c1), M(c2)$ は入力単語カテゴリ $c1, c2$ に対するHL1の出力ベクトルである。類似度0.985以上でクラスタリングした結果を図6に示す。HL1はHAVE動詞グループ、BE動詞グループ、主格代名詞グループ、名詞に続くカテゴリグループ(形容詞、冠詞、数詞、動詞、名詞、他)の5つのグループにきれいに分類された。すなわち、NETgram(Bigram予測モデル)はHL1に単語カテゴリーに内在する特徴を自然に学習している。

#### 4.2 Trigram予測実験結果

1,024センテンスについて学習したTrigram予測のNETgramおよびTrigram確率モデルの、学習用データとは別の評価用データ512センテンスに対する単語カテゴリの1位から5位までの累積予測的中率を図7に示す。NETgramはモデル1、モデル2ともTrigram確率モデルに比べ、同等以上の予測的中率を示している。

さらに、候補順位30位までのNETgramモデル2とTrigram確率モデルの評価用データ512センテンスに対する累積予測的中率を図8にそれぞれ○印、●印で示す。やはり、NETgramはTrigram確率モデルより優れている。候補数を多くしてもTrigram確率モデルの性能があまり向上しないのは、1,024センテンスの学習データがTrigram確率値を計算するために十分な統計量ではないことに起因する。例えば、Trigram確率モデルの場合、学習データに含まれていない未知データの出力に対する確率値はすべて0となり、予測することは出来ない。しかし、NETgramは未知データに対しても図8の□印に示すように妥当な予測結果を得ている。これはNETgramの学習はBigram予測モデルから始めていることによる。すなわち、Trigramデータとしては未知であっても、Bigramデータとしては学習データに含まれてい

る場合が多いため、NETgramはBigram予測モデルのときに学習した知識(構造)により、妥当な予測結果を出力しているものと考えられる。

確率モデルの場合、Trigramとして十分な統計量を得られていないデータに対して、Bigram予測モデルの予測結果により補間する方法がある<sup>(8)</sup>。Bigramにより補間処理された確率モデルの予測的中率を図8に×印で示す。削除補間法<sup>(8)</sup>で求めたBigram、Trigramの重み係数はそれぞれ0.62、0.38である。NETgramは補間処理された確率モデルの性能に匹敵していることがわかる。すなわち、NETgramはまさにネットワーク構造およびBigramからTrigramへの学習過程によりその補間能力を得ていると考えられる。補間処理された確率モデルの未知データに対する予測的中率を比較のために図8に+印で示す。

#### 4.3 NETgramと従来の確率モデルの比較

従来の確率モデルは基本的にはテーブルルックアップである。すなわち、Trigramに関していうと、先行する2つの単語のカテゴリに対して次に続く単語のカテゴリの出現頻度を学習データすべてについてカウントし、得られた頻度分布から確率分布を計算し、 $89 \times 89 \times 89$ のテーブルにそれぞれの確率値を入れる。予測の際は特定された選考する2つの単語カテゴリ番号に対して、次に続く単語のカテゴリ89個全ての出現確率を $89 \times 89 \times 89$ のテーブルから引く。よって、確率モデルは $89 \times 89$ の記号の順列から89個の確率分布をテーブルルックアップにより得ることになる。すなわち、

$$B^{89 \times 89} \Rightarrow R^{89} \quad (B; \text{バイナリ空間、} R; \text{実数空間})$$

一方、NETgramは $89 \times 89$ の記号の順列を16次元のアナログコードに符号化し、さらに予測出力値の16次元のアナログコードに変換し、そのコード

によって89個の予測出力値を得る。その様子を次のような写像としてとらえることができる。

$$B_{89 \times 89} \Rightarrow R_{16} \Rightarrow R_{16} \Rightarrow R_{89}$$

各空間の値はNETgramの各層の出力値として現れ、その写像はNETgramのリンクウェイトの値により一意に計算できる。すなわち、ネットワーク各層のユニットの出力とリンクウェイトの積和にシグモイド関数を通したものが、次の層のユニット出力となる。

この2種のアプローチを実現するための記憶容量(パラメータの数)を表1に示す。すなわち、確率モデルはテーブルルックアップであるため、Trigramの場合、パラメータは $89 \times 89 \times 89$ の確率値そのものであり、gram数の増加にともない指数的に増加する。実際には $89 \times 89 \times 89$ の確率値の中には0となる順列が多くあるので、ある種のテクニックにより、確率モデルの表の大きさを削減することはできるが、それは扱うタスクの種類や学習データ数に依存するため、ここでは最大値を示している。それに対し、NETgramはパラメータとしてリンクウェイト値を覚えておけば、 $89 \times 89 \times 89$ の確率値を計算により一意に求めることができ、gram数の増加に対してリンクウェイトの数はリニアに増加するだけである。

次にデータ表現という意味で両者を比較してみると、確率モデルでは入力記号列から直接確率値を得る1対1のテーブルルックアップであり、データ構造は何も考慮していない。それに対し、NETgramは入力記号列を16次元のアナログコード列に変換する時点で、4.1節に示したように単語カテゴリー間の距離を考慮したような、データに内在する構造の特徴抽出を行って、それを予測出力値に変換している。そのため、Bigram、Trigramと学習が進む中で、スパースなデータ間の補間が行われ、Trigramとしては未知

のデータでも効果的に予測を行うことができる。また、データの符号化という意味では、NETgramは出力のR89の空間をR16の中間形式に圧縮していると言える。

#### 4.4 4-gramネットワークモデルについて

Trigram予測モデルの学習データは1,024センテンスであったが、4-gram予測モデルの学習は2,048センテンスのデータを用いて行った。その結果、テストデータに対する予測的中率はTrigramネットワークモデルのとあまりかわらなかった。確率モデルによる学習データと予測的中率の関係を予備実験として計算すると、1,024センテンスのTrigram予測結果と同等の4-gram予測結果を得るためには、6,000センテンス以上のデータが必要であることがわかった。しかも、Trigram予測の場合、16,000センテンスまで予測的中率の上昇が続く。NETgramは前述の補間効果により、2,048センテンスの4-gramデータの統計量不足を補っていることになるが、やはり、Trigram予測結果を大きく上回るためには、16,000センテンスの数十倍以上の学習データが必要になることは、統計的にみても明らかである。Brown Corpusのデータ量、学習計算時間の問題から、今回はこれ以上の実験は行えず、4-gramモデルの評価ができなかった。

#### 4.5 英単語音声認識の改善結果

NETgramを用いて英単語音声認識の改善実験を行った<sup>(9)</sup>。音声認識はHMMによる手法を用いた<sup>(10),(11)</sup>。また、特別の発声訓練を受けていない米国人男性1人により発声された、国際会議の問合せやホテルの予約等の電話英会話文(377文、2834単語)を単語に区切って発声した単語データ中の1487語(190文)をHMMの学習に用い、残りの1347語(187文)を評価データとして単語認識に用いた。

実験結果を表2に示す。HMMのみによる単語認識率は81.0%であるのに対し、1,024センテンスで学習したNETgramを用いることにより認識率が86.9%に向上した。Trigram確率モデルを用いた場合よりもNETgramを用いる方が高い認識率を得ている。4.2節、4.4節で言及したように、512、1,024センテンスはTrigramの確率値を計算するには、まだ不十分なデータであるが、NETgramはこのように少ない学習データでも効率よく学習を行っていることがわかる。

## 5. むすび

文章中の単語の音声認識結果を改善する方法として、ニューラルネットによる単語列予測モデル(NETgram)の構成方法、実験結果について述べた。

NETgramの特徴は次のとおりである。

- (1) Bgram予測(2つ組予測)モデルからN-gram予測(N組予測)モデルまで拡張が容易
- (2) 予測計算をするためのパラメータの数が確率モデルよりも少ない。
- (3) NETgram内部(中間層)に入力データや出力データの特徴をアナログ符号化する構造を持つ。

英文テキストデータ(Brown Corpus)を用いて学習を行い、予測実験をして次のような結果を得た。

- (1) Trigram予測において、統計的に十分でない学習データを用いた場合、NETgramはTrigram確率モデルより優れた予測的中率を示した。
- (2) NETgramはTrigram予測において、未知データに対しても効果的に予測することが確認され、Bigramで補間処理された確率モデルに匹敵する予測性能を有することがわかった。
- (3) Bigramネットワークモデルの中間層HL1の出力パターンをクラスタリング分析した結果、言語的に意味のある5つのグループに分類され、NETgramは単語カテゴリに内在する特徴を学習していることが確認された。

実際問題として、Trigram等の予測モデルを構築する際に必要とされる学習用データは有限であり、正確な確率モデルを得ることができない場合が多い。このような場合、Bigram予測学習からTrigram予測学習を行うNETgramは、Trigramとしては未知データでもBigram予測学習で得た構造

により、効果的に予測を行うことができる。また、NETgramは計算機の記憶容量(パラメータの数に対応)の節約にも有効である。

今後の課題は、実際の音声認識への適用と、4-gram予測以上のモデルを学習するのに要するデータ量、計算時間を、実際に計算できる程度に圧縮する方法を考えることである。

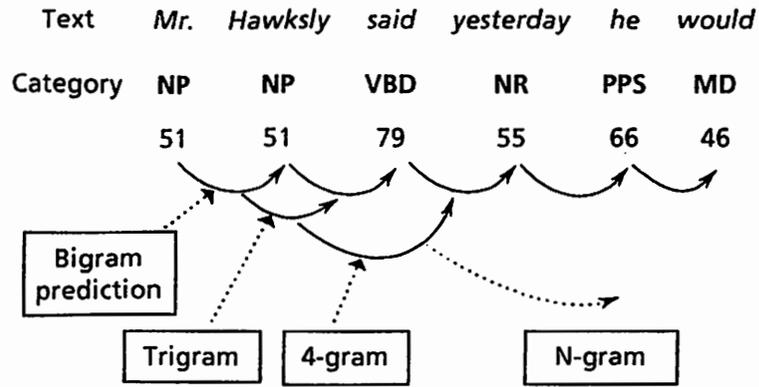


図1 Brown Corpus テキストデータによる単語品詞予測

Fig.1 Word Category Prediction

Using Brown Corpus Text Data

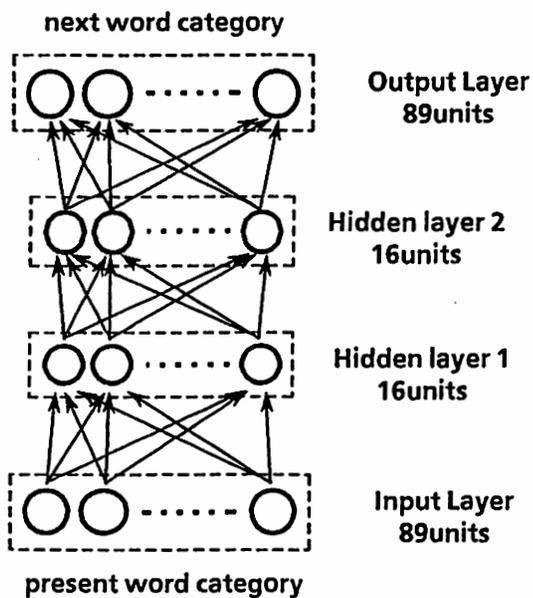


図2 Bigram ネットワークモデルの構成  
 Fig.2 Bigram Network for Word Category Prediction

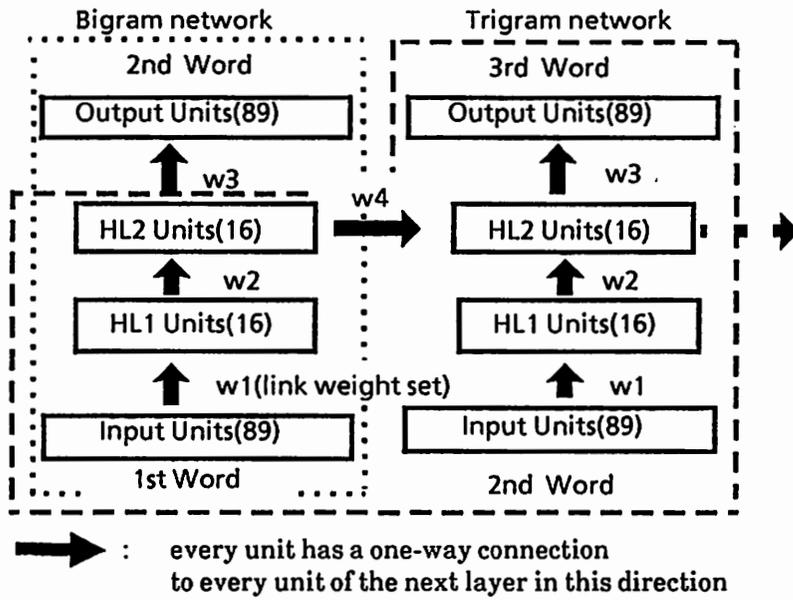


図3 NETgramモデル1の構成

Fig.3 NETgram Model 1 for Word Category Prediction

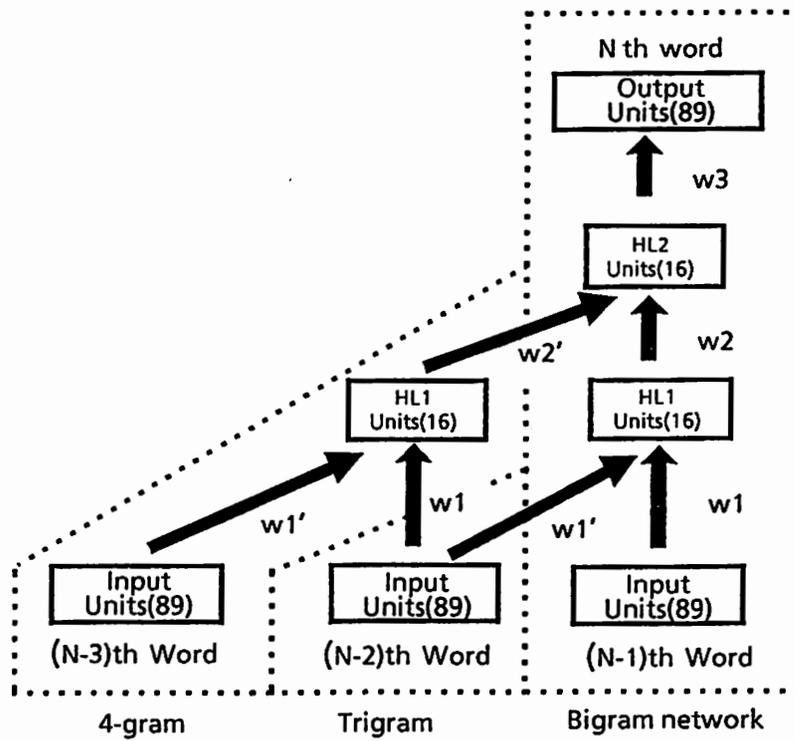


図4 NETgramモデル2の構成

Fig.4 NETgram Model 2 for Word Category Prediction

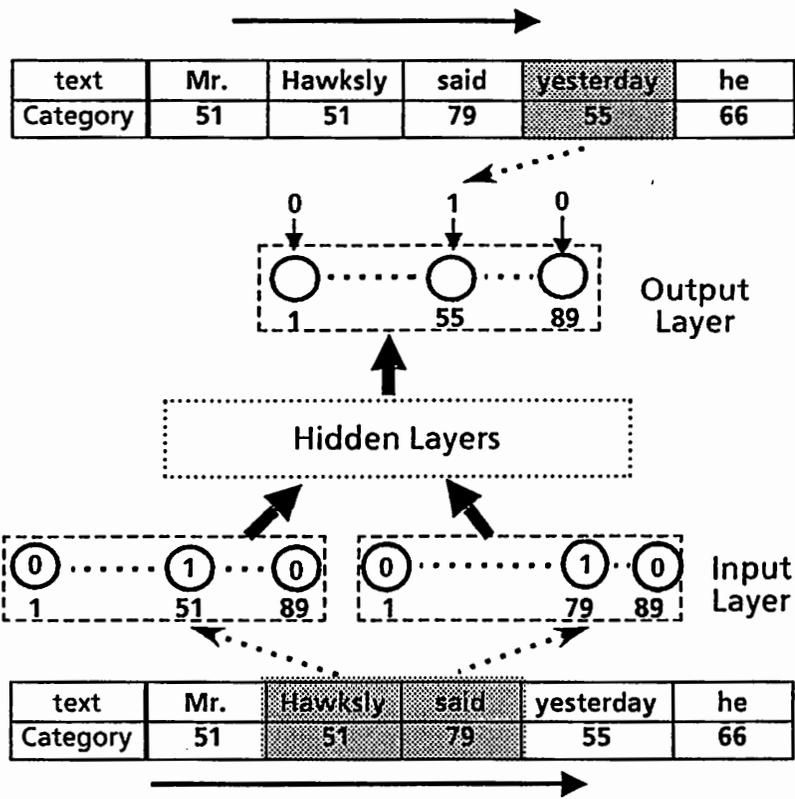


図5 NETgramの学習方法

Fig.5 How to Train NETgram (Trigram Model)

CATE- GORY	EXAMPLE (part of speech)	Threshold of Similarity				
		1.000	0.995	0.990	0.985	0.980
36 HV 37 HVD 40 HVZ	have had has					
15 BED 16 BEDZ 20 BER 21 BEZ 19 BEN 14 BE 17 BEG 38 HVG	were was are is been be being having					
66 PPS 86 WPS 67 PPSS	he, it who, which I, we, they					
29 DT 45 JJT 58 OD 42 JJ 48 NN\$ 61 PP\$ 52 NP\$ 13 AT 78 VB 80 VBG 06 , 11 AP 22 CC 09 ABN 10 ABX 75 RP 81 VBN 89 DUM 23 CD 43 JJR 47 NN 55 NR 79 VBD 82 VBZ 32 DTS 65 PPO 49 NNS 70 RB 50 NNS\$ 51 NP	this, that biggest first, 2nd (adjective) dog's my, our ATR's a, the (verb, base) (verb, -ing) , many, next and, or half, all both about, off (verb, -ed) (dummy) one, 2 (comp.adj.) (noun, single) home, west (verb, past) (verb, -s, -es) these me, him, it (noun, plural) (adverb) men's ATR, Tom					
others	others					

図6 各々のカテゴリに対するNETgram(Bigram)のHL1の出力ベクトルの  
クラスタリング結果

Fig.6 Clustering Result of HL1 Output Vectors of NETgram ( Bigram )

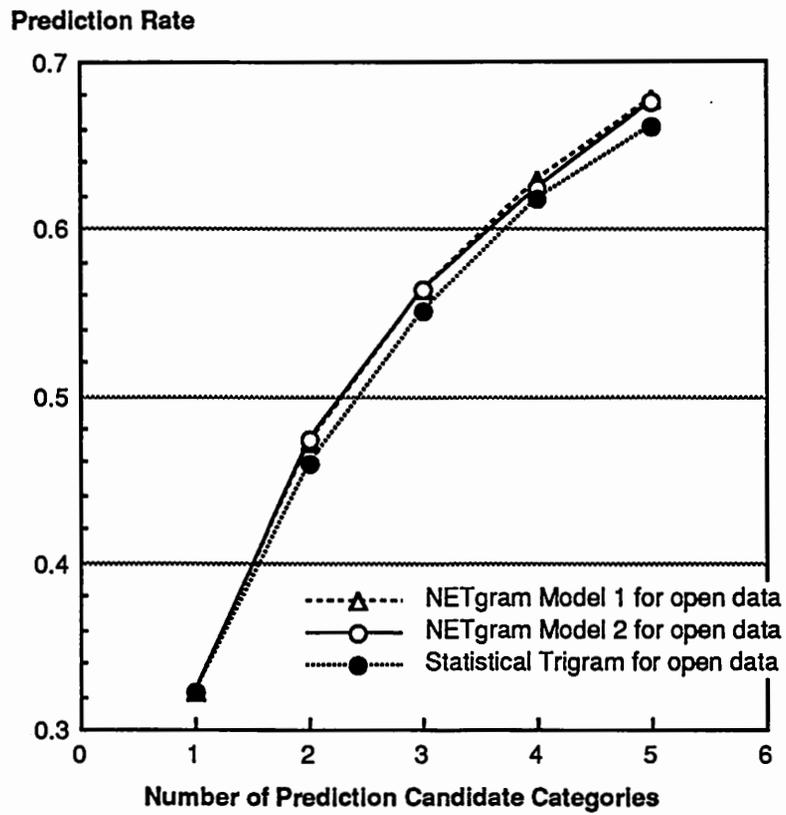


図7 NETgramおよび確率モデルの予測実験結果  
Fig.7 NETgram and Statistical Trigram Prediction Rates

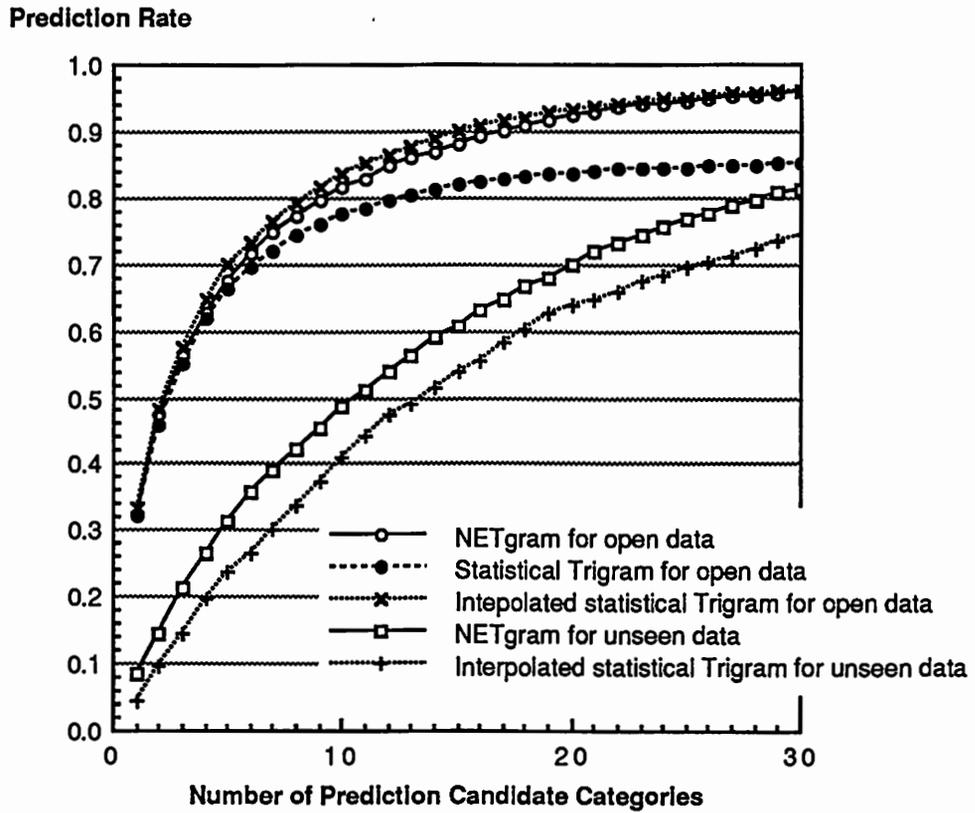


図8 NETgramおよび確率モデルの予測実験結果(30位まで)  
 Fig.8 NETgram and Statistical Trigram Prediction Rates (to top 30)

表1 確率モデルとNETgramのパラメータの数

	Statistical Model	NETgram Model 1	NETgram Model 2
Bigram	7,921 = 89 <sup>2</sup>	3,225	3,225
Trigram	704,969 = 89 <sup>3</sup>	5,193	4,649

表2 単語予測モデルによる  
認識率の改善結果(%)

Training Sentences	NETgram	Statistical Trigram Model
512	86.3	85.5
1,024	86.9	85.4

## 参考文献

- (1) F. Jelinek : "Continuous Speech Recognition by Statistical Methods",  
Proceedings of the IEEE, Vol.64, No.4 (1976 - 4).
- (2) 鹿野清宏 : "Trigram Modelによる単語音声認識結果の改善", 信学技  
報, SP87-23 (1987 - 6).
- (3) A. Waibel : "時間遅れ神経回路網(TDNN)による音声認識", 信学技報,  
SP87-100 (1987 - 12).
- (4) T.J.Sejnowski and C.R.Rosenberg : "NETtalk : A Parallel Network  
that Learn to Read Aloud", Tech. report, The Johns Hopkins  
University EECS-86-01 (1986).
- (5) Brown University : "Brown corpus", Tech. report, Brown university  
(1967).
- (6) D.E.Rumelhart, G.E.Hinton and R.J.Williams : "Learning Internal  
Representations by Error Propagation" in Parallel Distributed  
Processing, 1, pp.318-330, The M.I.T.Press (1986).
- (7) D.E.Rumelhart, G.E.Hinton and R.J.Williams : "Learning Internal  
Representations by Error Propagation" in Parallel Distributed  
Processing, 1, pp.354-361, The M.I.T.Press (1986).
- (8) F.Jelinek and R.Mercer, "Interpolated Estimation of Markov Source  
Parameters from Sparse Data", Pattern Recognition in Practice,  
pp.381-397, ed. E.S. Gelsema and L. N. Kanal, North-Holland Publish  
Company (1980).

- (9) 丸山活輝、中村雅己、川端豪、鹿野清宏：“HMM音韻連結学習とNETgramを用いた英単語音声の認識”、信学技報、SP89-89(1989-12).
- (10) 丸山活輝、花沢利行、川端豪、鹿野清宏：“HMM音韻連結学習を用いた英単語音声の認識”、信学技報、SP88-119(1989-1).
- (11) 花沢利行、川端豪、鹿野清宏：“HMM音韻認識におけるセパレートベクトル量子化の検討”、音講論集2-P-8(1988-10).