

認識対象毎に独立な構造を持つニューラル・ネットワークによる 7 B-2 単音節認識とその音声ワープロへの応用

吉野 敏樹 赤松 則男
徳島大学工学部

1. まえがき

現在、ユーザーインターフェースに関して、キーボードの代わりに、音声やペンを用いて入力する方法が注目されている。

音声入力の応用例として、音声ワープロが考えられる。日本語音声ワープロへの応用を考えた場合には、多様な文章を入力するために、単音節入力が不可欠である。我々の研究室では、特定話者の単音節入力に関して研究を行っている。

ニューラル・ネットワークを音声ワープロに使用する際には、認識率の他に、実時間処理、話者への適応等の課題がある。

本報告では、認識対象毎に独立な構造を持つニューラル・ネットワークを提案し、このネットワークを用いて日本語 68 種類の単音節を認識する方法を示す。さらに、本手法の特徴であるサンプルデータの追加学習方法を示し、これらの手法について検討する。

2. 認識対象毎に独立な構造を持つニューラル・ネットワーク

本報告では、図 1 に示すようなネットワーク構造を、認識対象毎に独立な構造を持つニューラル・ネットワーク (Independent Structure Neural Networks; ISNN) と呼ぶ。

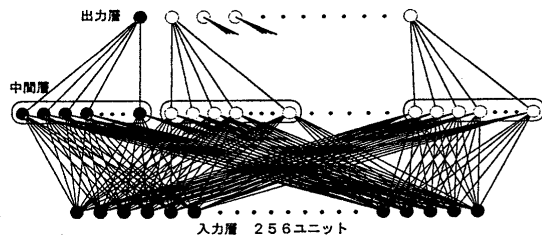


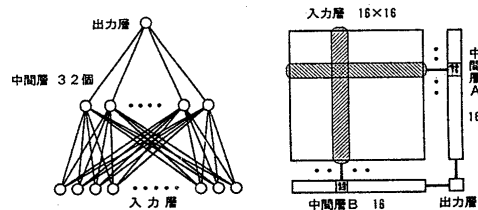
図 1. ISNN のネットワーク構造

ISNN とは、入力層のユニットは共通であるが、中間層、出力層のユニットは認識対象毎に用意し、他の

Syllable Recognition Using Independent Structure Neural Networks and its Application to Word-Processing
Toshiki YOSHINO, Norio AKAMATSU
University of TOKUSHIMA

ネットワークとは独立させる構造である。また、層間の結合もすべて認識対象毎に独立させる。従って、学習に関しても、他の認識対象を識別するネットワークの学習とは独立して行う。

本報告では、ISNN の構造を持つ全結合型 ISNN と結合制限型 ISNN について検討する。全結合型 ISNN とは、図 2 (a) に示すように、入力層と中間層、中間層と出力層のユニット同士を全て結合する構造である。結合制限型 ISNN とは、図 2 (b) に示すように、結合数を減少させ、中間層を A、B の 2 つのグループに分け、それぞれのグループ内のユニットは、入力層の縦あるいは横の 1 列のみに結合させる構造である。



(a) 3層全結合型 ISNN (b) 結合制限型 ISNN

図 2. 本報告で使用する 2 種類の ISNN 構造

3. 68 単音節の認識方法

3. 1 音声データ

学習に用いた音声データ (32kHz, 16bit) には、自動切り出しを行った後、人手により確認、一部修正されたデータを使用する。

ネットワークへの入力データは、以下のようにして求められる。音声データから、ハミング窓を用い、FFT によりパワースペクトルを求め、メルスケール 16 チャンネルの周波数データに変換する。正規化した周波数データを、16 フレーム分 (フレーム長 16ms, 分析周期 8ms) 使用する。従って、ネットワークへの入力は、16 ch × 16 フレームで 256 入力となる。

単音節の認識には、子音部と母音部の 2 つの入力データを用いる。それぞれの入力データを求める音声

データの区間を、図3に示す。

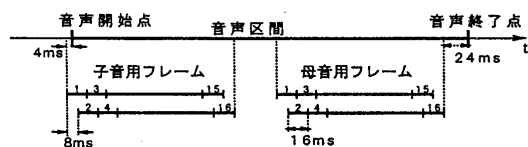


図3. 子音部, 母音部入力データの求める箇所

3.2 68 単音節の認識方法

母音部には定常性があり、母音部の認識は、子音部より容易に行うことができる。そこで、図4に示すように、6つの母音認識用 ISNN を用意し、全ての単音節の母音部の学習データを母音別に分類して、学習を行う。次に、“ん”を除く全ての単音節に対して子音認識用 ISNN を用意する。そして、母音部の共通する ISNN を1つのグループとし、そのグループに属するサンプルデータを用いて学習を行う。単音節の認識は、まず、母音部認識用 ISNN から母音を決定し、その母音のグループに属する子音部認識用 ISNN から、子音を決定する。ただし、“ん”に関しては、母音部のみを用いて判定する。

母音認識用	子音認識用
あ	あかきたなはまやらわがざだば
い	いきしちにひみりぎじびび
う	うくすつぬふむうるぐずぶぶ
え	えけせてねへめれげぜでべべ
お	おこそとのほもよろごぞどぼぼ
ん	

図4. 68 単音節を認識するために必要な ISNN と学習時のグループ分け

3.3 サンプルデータの追加学習

音声のように、多くの認識対象を扱う場合には、学習に用いるサンプルデータ数は、非常に多くなる。

ISNN のネットワーク構造の特徴は、認識対象毎に独立して学習する事にある。ISNN の独立性により、学習の際に他のネットワークへの影響はない。この特徴を利用して、サンプルデータの追加を行うことが可能である。

そこで、予め登録しておいたデータを用いて学習を行った後、話者が誤認識となったデータを追加、再学

習を行う。これにより、認識対象毎にデータを登録することが可能となり、不必要なデータの登録を防ぐことができる。

4. 認識実験

学習結果の評価には、未学習の音声データを単音節毎に 20 パターン用いる。

母音部の認識に関しては、68 種類の単音節毎に 5 パターンの音声データを用いて母音部の学習を行った結果、ほぼ 100 % の認識率が得られた。また、単音節毎に 10 パターンの音声データを用いて子音部の学習を行った結果、68 種類の単音節の認識率は、全結合型 ISNN を用いた場合には 95.6 %、結合制限型 ISNN を用いた場合には 94.9 % となった。

次に、あ段の子音部のデータを用いて、データの追加と認識率の関係を調べた。データの単音節毎に 10 個の学習データを用いて、ISNN の学習を行う。そして、実際に、そのネットワークを用いて認識を行い、誤認識となったデータを登録し、学習を行う。追加学習を繰り返し行い、あ段の子音部の認識率は 94.3 % から 97.8 % まで向上した。

さらに、これらの認識手法を用いて、コンピュータによるシミュレーションで音声ワープロを作成した。

5. まとめ

本報告では、ISNN を提案し、68 種類の単音節の認識方法を示し、良好な結果を得た。

さらに、ISNN の特徴を利用した、データの追加学習により、認識率が向上することが示された。従って、学習に必要なデータを、認識対象別に追加登録することができ、学習に必要なデータ数を少なく抑えることが可能となり、学習の高速化も期待できる。

参考文献

- 1) D.E. Rumelhart, J.L. McClelland and the PDP Research Group: "Parallel Distributed Processing", MIT Press (1986).
- 2) A. Waibel: "Phoneme recognition using time-delay neural networks", 信学技報, SP87-100 (1987).
- 3) 松岡達雄, 浜田洋, 中津良平: "グループ分割型ニューラルネットによる単音節音声認識", 信学論 (D-II), J72-D-II, 8, pp.1194-1199 (1989).