TR-I-0227

話者写像ニューラルネットの識別誤り最小基準による最適化

An Optimization of Speaker Mapping Neural Network Using Mimimal Error Criterion

栗並 賢太郎　杉山 雅英

Kentaro KURINAMI　Masahide SUGIYAMA

内容梗概

これまでに、話者間の写像にニューラルネットワークを適用することの有効性が確認されている。また最近認識モデルのパラメータの最適化手法として、識別誤り最小の基準に基づく最適化手法が提案されている。本報告ではこの識別誤り最小の基準に基づく最適化手法を用いて、話者間写像ニューラルネットワークを最適化する方法について検討する。始めに話者間写像ニューラルネットワークの概念、及び構造を述べる。次に一般的な識別誤り最小の基準に基づく最適化手法について述べ、つづいてそれを話者間写像ニューラルネットワークに適用した場合の最適化手法を示す。最後にそれを用いて行なった実験の結果を示し、本手法の評価、検討を行なう。

目次

表目次

図目次

1 はじめに

本報告では話者写像ニューラルネットワークの識別誤り最小化基準を用いた最適化の手法について述べる。近年、ニューラルネットワークが音声認識などの諸分野で活発に研究されている。これらのニューラルネットは主に識別タイプのネットワークであり、著者等は写像タイプのニューラルネットワークを用いた話者間の写像の検討を行ないその有効性を明らかにしてきた [1][2] [3]。一方認識モデルのパラメータの最適化の手法として、識別誤り最小の基準に基づく最適化手法が提案されその有効性が検討されている [4][5] [6]。本報告では識別誤り最小化基準による認識モデル最適化手法を用いた話者間写像ニューラルネットワークの最適化について検討する。

2 話者間写像ニューラルネットワークについて

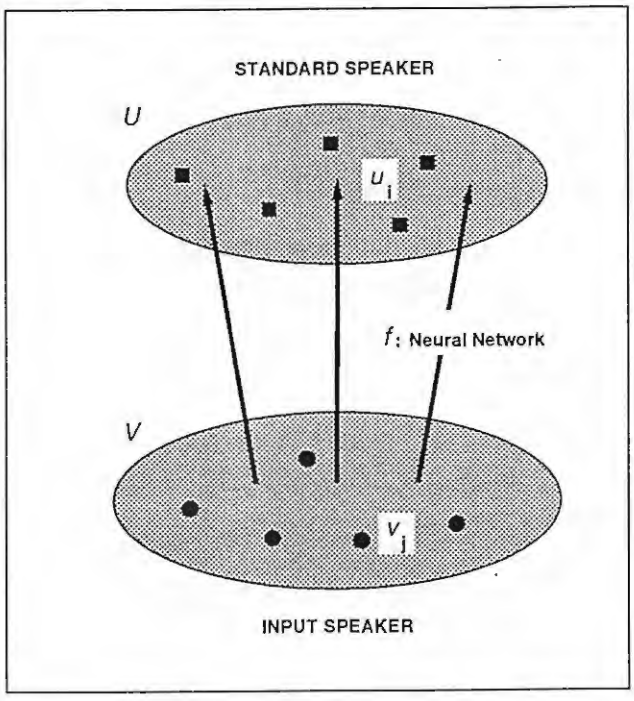

図 1: 話者間写像の概念図

計算量の削減のために標準話者の発声データをベクトル量子化することによって、各カテゴリ毎にいくつかの代表ベクトルを定める。このベクトルの集合を $U = \{u_i\}$ とする。また未知話者の発声データの集合を $V = \{v_j\}$ とする。ここで、集合 V から集合 U への写像 f を考える。

$$f : V \rightarrow U \tag{1}$$

写像 f は、未知話者のベクトルの集合 V の各要素が、集合 U の中のその要素と同じカテゴリに属するベクトルへ写像されるように定める。この f を用いて、未知話者のベクトルを標準話者のベクトルに写像させることによって、話者適応を行なう。この様子を図 1に示す。

この未知話者ベクトルから標準話者ベクトルへの写像をニューラルネットワークを使って実現するのが、話者間写像ニューラルネットワークである。用いたニューラルネットワークの構造を図 2に示す。入力層と出力層は各々 16 ユニット、隠れ層 1 と隠れ層 2 は各々 80 ユニットを持つ 4 層フィードフォワード型ニューラルネットワークである。入力層及び隠れ層ではシグモイド関数による非線形ユニットを用いているが、最終層（出力層）のユニットではシグモイド関数による非線形効果を入れていない。隠れ層のユニット数を入出力層のユニット数より大きくすることによって、十分な写像性能を得ることをめざしている。

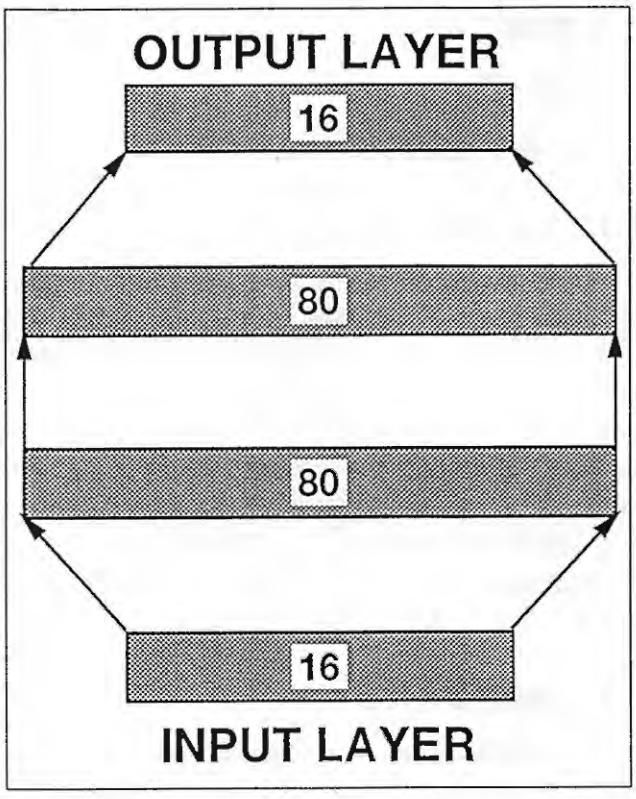

図 2: 話者間写像ニューラルネットワーク

実際に V から U への写像を求めるために、未知話者の入力ベクトルを v、ニュラールネットワークによる v の写像を $f(v)$、標準話者のベクトルを u としたとき、$f(v)$ が v の属するカテゴリと同じカテゴリに属する u に最も近くなるように、ネットワークを学習させる。ここで、従来の学習では $f(v)$ と望ましい u との距離を誤差関数とし、この誤差関数を最小にするように、学習を行なっていた。この場合、他のカテゴリとの関係についてはまったく考慮されていない。しかし学習の目的は、与えたベクトルが誤ったカテゴリのベクトルよりも正しいカテゴリのベクトルにより近いように写像されることである。これをより直接的に

実現するために、本研究では識別誤りを誤差関数とし、この誤差関数を最小にするように学習を行なった。

3 識別誤り最小基準による最適化手法について

学習パターンの組が与えられて、これらが K 個のカテゴリのどれかに属する場合を考える。識別誤り最小基準で最適化するということは、誤識別率が最小になるように、認識システムを学習させることである。

識別誤り最小基準で最適化を行なうために、つぎの4つの関数を定義する。

1. 識別関数
2. 誤差関数
3. 損失関数
4. 損失の和

これらの意味と定義を次に示す。

3.1 識別関数の定義

まず始めに識別関数 $g_k(v,\Lambda)$ を定義する。Λ は認識システムのパラメータの状態を表すベクトルで、例えばHMMでは、モデルの出力確率と遷移確率を表す。また、標準パターンを用いるDTWの場合にはパターンのパラメータを表すことになる。識別関数とは、この Λ の元で入力ベクトル v が k 番目のカテゴリに属する度合を表す。入力ベクトルに対してこの識別関数が最小（最大）になったカテゴリが、識別されたカテゴリとして選ばれることになる。以下では識別関数が最小になる場合に限定して議論を進めることにするが、一般性は失われることはない。

3.2 誤差関数の定義

次に誤差関数 $d_\alpha(v,\Lambda)$ を定義する。誤差関数とは入力ベクトル v が他のカテゴリと比べて、どの程度正しいカテゴリに識別されているかを表す。誤差関数は一般的に次のように表される。

$$d_\alpha(v,\Lambda) = g_\alpha(v,\Lambda) - \left[\frac{1}{K-1}\sum_{k,k\neq\alpha} g_k(v,\Lambda)^\zeta\right]^{\frac{1}{\zeta}} \quad (2)$$

ここで α は入力ベクトル v の属するカテゴリである。また ζ は正の実数で、この ζ を変化させることにより、各カテゴリの影響の大きさを変化させることができる。$\zeta\to\infty$ で $d_\alpha(v,\Lambda)$ は次のように近似できる。

$$d_\alpha(v,\Lambda) = g_\alpha(v,\Lambda) - g_\beta(v,\Lambda) \quad (3)$$

ただし β は α 以外で誤差関数が最小になるクラスタである。誤差関数が負の場合、その入力は正しいクラスタに識別され、誤差関数が正の場合、その入力は誤ったクラスタに識別されたことになる。

3.3 損失関数の定義

次に損失関数 $l_\alpha(v,\Lambda)$ を定義する。損失関数は入力ベクトルが、正しいカテゴリに識別されたかどうかを表す。$l_\alpha(v,\Lambda)$ は理想的には次のようになる。

$$l_\alpha(v,\Lambda) = \begin{cases} 0 & if\ d_\alpha(v,\Lambda)\le 0 \\ 1 & if\ d_\alpha(v,\Lambda) > 0 \end{cases} \quad (4)$$

つまり、入力が正しく識別された場合は0、誤識別された場合は1となる。しかしこのままでは最適化が困難なため実際には次のように表す。

$$l_\alpha(v,\Lambda) = \begin{cases} 0 & if\ d_\alpha(v,\Lambda)\le 0 \\ l(d_\alpha) & if\ d_\alpha(v,\Lambda) > 0 \end{cases} \quad (5)$$

あるいはシグモイド関数等で表すことも良く行なわれる。

3.4 損失の和の定義

最後に全入力ベクトルの損失関数の和として、$L(\Lambda)$ を定義する。$L(\Lambda)$ は次のように表される。

$$L(\Lambda) = \sum_\alpha \sum_j l_\alpha(v_j,\Lambda) \quad (6)$$

この $L(\Lambda)$ は識別誤りの合計にあたるわけであるから、この $L(\Lambda)$ を目的関数として、$L(\Lambda)$ を最小にすることによって、識別誤りを最小にすることができる。

$L(\Lambda)$ 最小化のために、最急降下法を用いる。ある時点 t でのパラメータの状態を $\Lambda(t)$ とすると学習による変化は次のようになる。

$$\Lambda(t+1) = \Lambda(t) + \delta\Lambda(t) \quad (7)$$

ここで Λ の変化量 $\delta\Lambda(t)$ は次のように表される。

$$\delta\Lambda(t) = -\varepsilon\nabla L(\Lambda) \quad (8)$$

ε は正の実数で Λ の変化量を制御する変数である。このようにして識別誤りを最小にするように最適化を行なうことができる。

4 話者間写像ニューラルネットワークへの適用

4.1 話者間写像における定式化

前節で識別誤り最小基準による最適化の一般的な方法を述べた。この節ではそれを話者間写像ニューラルネットワークに適用できるように定式化する。

さてここで一般性を持たせるために、m 層のフィードフォワード型ニューラルネットワークについて考える。ただし、最終層のみ線形ユニットであり、他の層は非線形ユニットであるとする。n 層の第 y ユニットへの入力の総和を i_y^n 、n 層の第 y ユニットからの出力を o_y^n とする。また、$n-1$ 層の第 x ユニットから n 層の第 y ユニットへの結合の重みを $w_x^{n-1}{}_y^n$ とする。m 層のユニットの出力が、ネットワークの出力となる。入出力関数の関係を以下の式 (9), (10) に示す。

$$o_x^n = \sigma(i_x^n) \quad (n = 1, \ldots, m) \tag{9}$$

$$i_x^n = \sum_{x'} w_{x'}^{n-1}{}_x^n o_{x'}^{n-1} \tag{10}$$

ただし、式 (9) において $n = m$ の時、σ は恒等写像とする。識別誤り最小化のための Λ はネットワークの結合の重み $w_x^{n-1}{}_y^n$ を表し、ネットワークの出力は入力と Λ によって決まる。

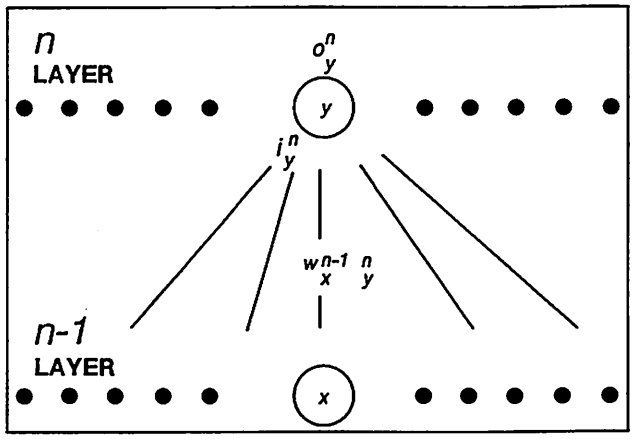

図 3: ニューラルネットワークの重み係数の定義

カテゴリ k に属する i 番目の標準話者ベクトルを u_i^k とする。($u_i^k{}_y$: u_i^k の y 番目の要素) また j 番目の未知話者入力ベクトルを v_j 、入力 v_j に対するネットワークの出力を $f(v_j)$ とする。このとき識別関数をネットワークの出力ベクトルと、標準話者ベクトルとのユークリッド距離を用いて、次のように定める。

$$g_k(v, \Lambda) = \left\{ \sum_i (\|f(v) - u_i^k\|^2)^{-\zeta} \right\}^{-\frac{1}{\zeta}} \tag{11}$$

ここで $\zeta \to \infty$ とすると式 (11) は次のように近似できる。

$$g_k(v, \Lambda) = \min_i \|f(v) - u_i^k\|^2 \tag{12}$$

誤差関数は式 (2) となる。損失関数はシグモイド関数を用いることを考える。

$$l_\alpha(v, \Lambda) = \sigma(d_\alpha) = \frac{1}{1 + e^{-d_\alpha}} \tag{13}$$

損失の和は式 (6) となる。

パラメータの変化量を求めるために,最小にすべき目的関数である $L(\Lambda)$ を、変化させるパラメータである結合の重みで偏微分する。なお今回のタスクはそれほど複雑では無いため、識別関数および誤差関数は $\zeta \to \infty$ として近似した値を用いる。また i_1, i_2 を次のように定める。また、α, β は前節で定義したものと同様である。

$$i_1 = \arg\min_i \|f(v_j) - u_i^\alpha\|^2 \tag{14}$$

$$i_2 = \arg\min_i \|f(v_j) - u_i^\beta\|^2 \tag{15}$$

4.2 $\nabla L(\Lambda)$ の算出

4.2.1 出力層に対する計算式

始めに出力層での結合の重みに対する偏微分係数を算出する。

$$\begin{aligned}
\frac{\partial L}{\partial w_x^{m-1}{}_y^m} &= \frac{\partial}{\partial w_x^{m-1}{}_y^m} \sum_\alpha \sum_j l_\alpha(v_j, \Lambda) \\
&= \sum_\alpha \sum_j \frac{\partial}{\partial w_x^{m-1}{}_y^m} \sigma(d_\alpha) \\
&= \sum_\alpha \sum_j \sigma(d_\alpha)\{1 - \sigma(d_\alpha)\} \\
&\quad \times \frac{\partial}{\partial w_x^{m-1}{}_y^m} \{g_\alpha(v_j, \Lambda) - g_\beta(v_j, \Lambda)\} \\
&= \sum_\alpha \sum_j \sigma(d_\alpha)\{1 - \sigma(d_\alpha)\} \\
&\quad \times \frac{\partial}{\partial w_x^{m-1}{}_y^m} \{\|f(v_j) - u_{i_1}^\alpha\|^2 \\
&\qquad - \|f(v_j) - u_{i_2}^\beta\|^2\} \\
&= \sum_\alpha \sum_j \sigma(d_\alpha)\{1 - \sigma(d_\alpha)\} \\
&\quad \times \left[\{2(o_y^m - u_{i_1}^\alpha{}_y) \frac{\partial}{\partial w_x^{m-1}{}_y^m} o_y^m \} \right. \\
&\qquad \left. - \{2(o_y^m - u_{i_2}^\beta{}_y) \frac{\partial}{\partial w_x^{m-1}{}_y^m} o_y^m \} \right] \\
&= \sum_\alpha \sum_j 2\sigma(d_\alpha)\{1 - \sigma(d_\alpha)\} \\
&\quad \times \{(o_y^m - u_{i_1}^\alpha{}_y) - (o_y^m - u_{i_2}^\beta{}_y)\} \\
&\quad \times \frac{\partial}{\partial w_x^{m-1}{}_y^m} \sum_{x'} w_{x'}^{m-1}{}_y^m o_{x'}^{m-1} \\
&= \sum_\alpha \sum_j 2\sigma(d_\alpha)\{1 - \sigma(d_\alpha)\} \\
&\quad \times (u_{i_2}^\beta{}_y - u_{i_1}^\alpha{}_y) o_x^{m-1}
\end{aligned} \tag{16}$$

ここで誤差信号として

$$\delta_{my} = 2\sigma(d_\alpha)\{1 - \sigma(d_\alpha)\}(u_{i_2}^\beta{}_y - u_{i_1}^\alpha{}_y) \tag{17}$$

と定めると、式(16)は次のようになる。

$$\frac{\partial L}{\partial w_{x\ y}^{m-1\ m}} = \sum_{\alpha}\sum_{j} \delta_{my} o_x^{m-1} \tag{18}$$

4.2.2 中間層に対する計算式

次に第 n 層への結合の重みに対する偏微分係数を算出する。

$$\begin{aligned}
\frac{\partial L}{\partial w_{x\ y}^{n-1\ n}} &= \frac{\partial}{\partial w_{x\ y}^{n-1\ n}} \sum_{\alpha}\sum_{j} l_\alpha(v_j, \Lambda) \\
&= \sum_{\alpha}\sum_{j} \sigma(d_\alpha)\{1-\sigma(d_\alpha)\} \\
&\quad \times \sum_{z} \Big[\{2(o_z^{n+1} - u_{i_1\ z}^{\alpha}) \\
&\qquad -2(o_z^{n+1} - u_{i_2\ z}^{\beta})\} \frac{\partial}{\partial w_{x\ y}^{n-1\ n}} o_z^{n+1}\Big] \\
&= \sum_{\alpha}\sum_{j}\sum_{z} \delta_{(n+1)z} \frac{\partial}{\partial w_{x\ y}^{n-1\ n}} w_{y\ z}^{n\ n+1} o_y^n \\
&= \sum_{\alpha}\sum_{j}\sum_{z} \delta_{(n+1)z} w_{y\ z}^{n\ n+1} \\
&\quad \times \frac{\partial}{\partial w_{x\ y}^{n-1\ n}} \sigma\Big(\sum_{x'} w_{x'\ y}^{n-1\ n} o_{x'}^{n-1}\Big) \\
&= \sum_{\alpha}\sum_{j}\sum_{z} \delta_{(n+1)z} w_{y\ z}^{n\ n+1} o_y^n (1-o_y^n) o_x^{n-1}
\end{aligned} \tag{19}$$

誤差信号として

$$\delta_{ny} = \sum_{z} \delta_{(n+1)z} w_{y\ z}^{n\ n+1} o_y^n (1-o_y^n) \tag{20}$$

と定めると、式(19)は次のようになる。

$$\frac{\partial L}{\partial w_{x\ y}^{n-1\ n}} = \sum_{\alpha}\sum_{j} \delta_{ny} o_x^{n-1} \tag{21}$$

式(7)、(8)および式(18)、(21)よりネットワークの結合の重みの学習規則は、次のようになる。

$$w_{x\ y}^{n-1\ n}(t+1) = w_{x\ y}^{n-1\ n}(t) - \varepsilon(t) \sum_{\alpha}\sum_{j} \delta_{ny} o_x^{n-1} \tag{22}$$

4.3 学習アルゴリズムの解釈

さてここで、識別誤り最小基準によって求められた学習アルゴリズムの意味を考えてみよう。式(20)を見ると、δ_{ny} の誤差信号は $\delta_{(n+1)z}$ を含み、いわゆるバックプロパゲーションとまったく同じ形となっている。これは、重みの修正のための誤差を出力側から入力側へ逆に伝搬していくという点で、この手法が普通のバックプロパゲーションと同じであることを表している。その違いであるが、普通のバックプロパゲーションでは出力での誤差として、出力値と望ましい値との差 $(o_y^m - u_{i1\ y}^{\alpha})$ を用いるのに対して、今回の手法では式(17)より $(u_{i2\ y}^{\beta} - u_{i1\ y}^{\alpha})$、即ち、誤った標準ベクトルと正しい標準ベクトルとの差を用いているのがわかる。また普通のバックプロパゲーションでは正しく識別されたかどうかは学習の際には問題にしないのに対して、今回の手法では式(17) $(\sigma(d_\alpha)\{1-\sigma(d_\alpha)\})$ より、d_α が0に近いほど、つまり識別誤りが発生するカテゴリの境界付近ほど良く学習が行なわれることがわかる。

表 1: 実験条件

話者		標準話者： 男性1名(MAU) 未知話者： 男性1名(MHT)
認識タスク		5母音 /a,i,u,e,o/
特徴パラメータ		16次LPCケプストラム
分析条件	LPC分析次数	14次
	標本化周波数	12kHz
	窓長	256点 (21.3ms)
	窓関数	Hamming
VQコードブックサイズ		40[8×5母音] (母音カテゴリごとに作成)
サンプル	コードブック作成	500samples[100×5母音] (標準話者5240単語偶数番)
	適応学習サンプル	500samples[100×5母音] (未知話者5240単語偶数番)
	テストサンプル	500samples[100×5母音] (未知話者5240単語奇数番)

5 母音識別による評価実験

5母音を用いた話者適応化の実験により、話者写像ニューラルネットワークの識別誤り最小基準による最適化法の評価を行なった[7][8]。実験条件を表1に示す。各サンプルは母音中心部より抽出した。

5.1 損失関数の選択

今回損失関数として式(13)のシグモイド関数と式(23)の2通りを試してみた。

$$l_\alpha(v, \Lambda) = \begin{cases} 0 & if\ d_\alpha(v,\Lambda) \le 0 \\ (d_\alpha)^\xi & if\ d_\alpha(v,\Lambda) > 0 \end{cases} \quad (0 < \xi \ll 1) \tag{23}$$

しかし式(23)を使った場合には、ある程度の認識率の改善は行なわれるものの、認識率が安定しないという欠点があった。その理由としては、ξ の値を適切な値に設定するのが難しいためと考えられる。図5及び図6に学習が進むにつれての、誤認識の数の変化を示す。図5は $\xi = 0.01$、図6は $\xi = 0.001$ の場合である。$\xi = 0.01$ の場合、誤認識数が大きく変動しているのがわかる。また $\xi = 0.001$ の場合に

は、$\xi = 0.01$ の時よりは安定しているが、それでもかなりの変動がみられる。図7は損失関数にシグモイド関数を使った場合であるが、細かな変動はあるものの、全体として良く収束しており、また認識率も良いことがわかる。そこで以後の実験は損失関数にシグモイド関数を用いて行なった。

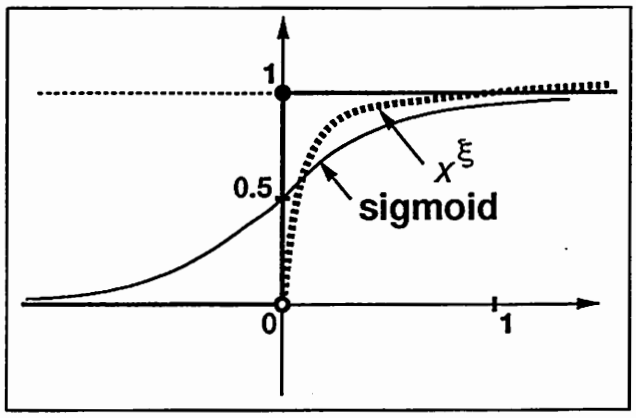

図 4: 種々の損失関数の定義関数の特性

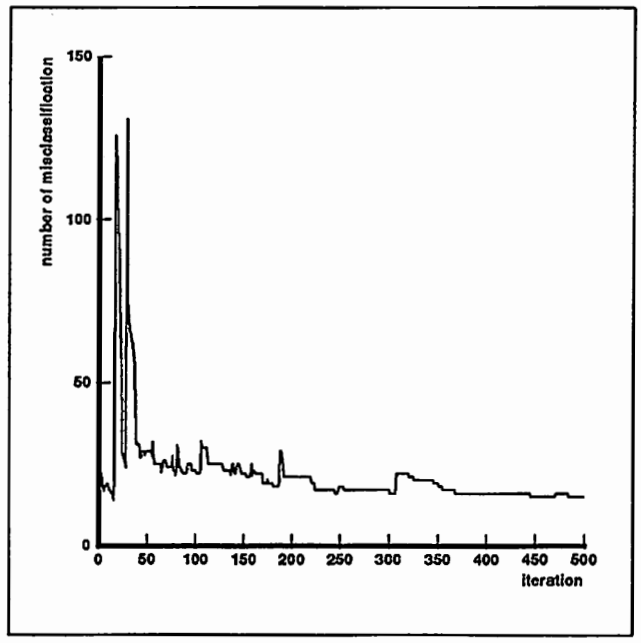

図 6: 学習の収束 (巾関数 $x^\xi, \xi = 0.001$)

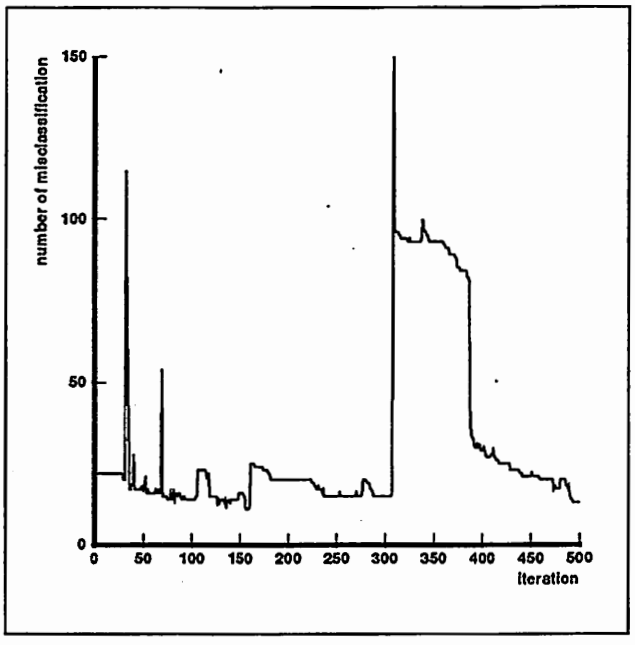

図 5: 学習の収束 (巾関数 $x^\xi, \xi = 0.01$)

5.2 パラメータの更新ステップ幅

パラメータの更新幅 ϵ は次のようにして決定した。始めに単調減少関数として、$\epsilon(t) = 1/t$ を用いて実験を行なった。t は学習の繰り返し回数である。しかしこの場合、始めの数回の学習で値（認識率）が発散してしまった。値が発散してしまうというのは、出力値が望ましい値とはかけはなれた値になってしまうということである。学習アルゴリズムからわかるように、認識誤りが起こってもそれが望ましい値から大きくはなれている場合、学習はほとんど行なわれない。従って、この場合それ以上学習が進まなくなってしまうのである。この原因はパラメータの更新量が大きすぎたためと考えられる。そこで更新幅 $\epsilon(t)$ を抑制して、更新量をおさえることを考える。そのために、次のような方法を用いた。適当な値 a を定め、それを式 (24) に示すように $\epsilon(t)$ の式の分母に加えるのである。このようにして $\epsilon(t)$ の値を抑制することによって、学習の発散を防ぐことができる。実際の実験では a の値を 10 以上に設定することによって、うまく最適化が実現できた。a を大きくすると収束が遅くなった。

$$\epsilon(t) = \frac{1}{t+a} \tag{24}$$

5.3 ニューラルネットの重みの初期値

ネットワークの重みの初期値に関しては、ランダムな値と、あらかじめ教師なしで学習を行なったネットワークの重みの、2通りについて実験を行なった。教師なし学習後の重みを用いたのは、初期値をある程度正しい重みに設定することによって、学習が確実に収束するのではないかと考えたからである。

5.4 識別実験結果

実験結果として、学習データに対する認識率を表2、テストデータに対する認識率を表3に示す。認識率（R[%]）

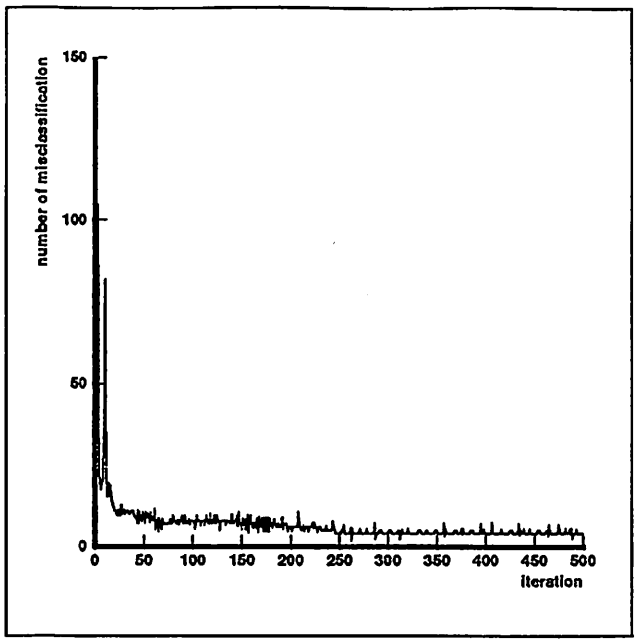

図 7: 学習の収束 (シグモイド関数)

表 2: 母音識別実験結果 [学習データ]

(認識率 : %)

初期値	最適化前	最適化後
ランダム	20.0	99.6
教師なし学習後	95.6	99.2

は、N_a : 全サンプル数、N_s : $g_\alpha < g_\beta$ となるサンプル数、として次の式に従って算出した。

$$R = \frac{N_s}{N_a} \times 100 \qquad (25)$$

実験の結果、話者写像ニューラルネットワークを識別誤り最小基準で最適化することによって、学習データに対して 99.6%、テストデータに対して 97.4% の認識率を得ることができた。初期値の設定の変更による認識率の変化はほとんど見られなかった。本報告で提案した教師あり学習は、教師なし学習より認識率の点で改善されているのがわかる。また従来の DTW による単語単位での対応関係を用いる教師あり学習の結果は、25 単語による話者適応で 95.8% の認識率となっている。タスクが異なるので単純に比較はできないが、25 単語による教師あり適応よりも良い認識率が得られている。

表 3: 母音識別実験結果 [テストデータ]

(認識率 : %)

初期値	最適化前	最適化後
ランダム	20.0	97.4
教師なし学習後	94.4	97.2

6 むすび

話者間での写像を行なう写像タイプのニューラルネットワークである話者写像ニューラルネットワークに、識別誤り最小基準で最適化を行なう手法を適用し、その定式化を行なった。また、5 母音を用いた話者適応化実験を行ない、その結果学習データに対して 99.6%、テストデータに対して 97.4% の認識率を得、本手法が有効であることを確認した。本手法は NN を特徴抽出器と捉えると、特徴抽出器の最適設計問題に対する 1 つの解決方法を与えたことになる。今後の課題を以下に示す。

- 一般の ζ に対する認識性能の比較
- 他の音素カテゴリへの拡張
- DTW, HMM を認識系とする話者写像 NN の学習法への拡張
- 一般化した NN の学習法への拡張

謝辞

研究の機会を与えて頂いた樽松社長、貴重な御助言、御検討頂いた嵯峨山研究室長、dcp, nnwb などのニューラルネットワークソフトウェアツールを御教示頂いた福沢氏をはじめとする音声情報処理研究室の皆様、視聴覚機構研究所の片桐氏に感謝します。

参考文献

[1] 福沢, 沢井, 杉山, ニューラルネットワークによる恒等写像を用いた話者適応, 音響学会講演論文集, 1-8-16, pp.31-32 (1990-09).

[2] K.Fukuzawa, H.Sawai, M.Sugiyama, Segment-based Speaker Adaptation by Neural Network, IEEE Workshop on Neural Networks for Siganl Processing (1991-10).

[3] 福沢, 小森, 沢井, 杉山, セグメントベース話者適応ニューラルネットワークを用いた文節音声認識, 音響学会講演論文集, (1991-10).

[4] Shigeru Katagiri, Chin-Hui Lee, Biing-Hwang Juang, A generalized probabilistic decent method, 音響学会講演論文集, 2-P-6, pp.141-142 (1990-09).

[5] Erik MacDermott, Shigeru Katagiri, Discriminative Training for Various Speech Units, 音声研究会資料, SP91-12, pp.47-54 (1991-06).

[6] Shigeru Katagiri, Chin-Hui Lee, Biing-Hwang Juang, Discriminative Multi-Layer Feed-Forward Networks, IEEE Workshop on Neural Networks for Siganl Processing (1991-10).

[7] 杉山, 福沢, 沢井, 嵯峨山, ニューラルネットによる集合間写像の教師なし学習, 音響学会講演論文集, 2-P-10, pp.149-150 (1990-09).

[8] 福沢, 杉山, ニューラルネットワークによる教師なし話者適応法とその評価, 音響学会講演論文集, (1991-10).