TR-I-0249

Radial Basis Function を適用した Fazzy Partition Model による音声認識
Radial Basis Function-Based Fuzzy Partition Models Aiming at Speech Recognition

安岐聡† 加藤喜永 杉山雅英
Satoshi AKI, Yoshinaga KATO and Masahide SUGIYAMA

概要

本研究では Fuzzy Partition Model (ＦＰＭ) に Radial Basis Function (ＲＢＦ) を適用し、日本語６子音と１８子音の認識実験、および連続音声認識実験を行った。本論文で用いるＲＢＦとは素子からの出力と結合重みとのユークリッド距離を求める関数である。また、ＲＢＦを用いた時にはニューラルネットワークの活性化関数にガウシアン関数を用いる。従ってこのネットワークによる空間は超球面上に形成される。一方、ＦＰＭは多入力多出力素子で構成されるニューラルネットワークで、連続音声認識認識においてＴＤＮＮより高い認識率が得られている。音素認識実験を行った結果、従来のＦＰＭとＲＢＦを適用したＦＰＭとは同等の認識率であった。また、連続音声認識実験に対しては、従来のＦＰＭに比較してＲＢＦを適用したＦＰＭの方は認識率が 2.2 ポイント 低下したがＴＤＮＮより 3.6 ポイント 高い認識率が得られた。さらに比較実験の結果、ＲＢＦを出力層に適用したＦＰＭと、出力層、中間層共にＲＢＦを適用したＦＰＭとでは前者の方が認識率が高かった。

†豊橋技術科学大学

目次

1 はじめに

ニューラルネットワークの素子は二つの部分に分解できる。一つは素子への入力関数、もう一つは素子の活性化関数である。前者は普通、内積タイプかユークリッド距離タイプであり、後者の代表的な例としてはシグモイド関数があげられる。素子への入力関数にユークリッド距離タイプを用いたものは *radial basis function* (RBF)[1, 7] と呼ばれる。 RBF に対する活性化関数には通常ガウシアン関数が適用される。

音声認識に対する RBF の応用として、ニューラルネットワークに統計的識別理論を導入するために用いられた例がある [2] 。この例で RBF はニューラルネットネットワークの入力部に用いられ、入力パラメータを統計的に扱いやすい特徴パラメータに変換する役目をしている。この RBF は従来の Back-Propagation (BP)[3] などによる学習は行われていない。本論文では 識別器の一部として RBF を使用し、音響空間をより滑らかに識別させることを試みた。また RBF 素子の重みも BP により学習する。

一方、Fuzzy Partition Model (FPM)[4, 5] は多入力多出力素子で構成される内積タイプのニューラルネットワークであるが、この FPM を用いた連続音声認識において TDNN に比較して高い認識率が得られている [6] 。

そこで本研究では FPM に RBF を適用し、日本語 6 子音と 18 子音の認識実験、および連続音声認識実験を行った。

2 Radial Basis Function

ニューラルネットワークの素子への入力の方法としては式 (1) のような内積タイプのものと式 (2) のようなユークリッド距離タイプのものがある。但し、d_i は素子 i の入力、x_j は素子 j の出力、w_{ij} は素子 j から素子 i への重みである。また、N は素子 j がある層の素子数である。

$$d_i = \sum_{j=1}^{N} w_{ij} x_j \tag{1}$$

$$d_i = \sum_{j=1}^{N} (w_{ij} - x_j)^2 \tag{2}$$

内積タイプのニューラルネットワークは超平面で境界面が形成される。一方、ユークリッド距離タイプのニューラルネットワークは超球 (円) 面で境界面が形成される。よって内積タイプのものより滑らかに境界が形成される。

RBF はユークリッド距離タイプに次式のガウシアン活性化関数

$$exp(-\beta d) \tag{3}$$

を用いたものであるから、RBF を用いた素子の出力は超球の中心で 1 となりその中心から離れるにつれて指数関数的に小さくなる。

ここで視覚的に比較するために素子への入力が x_1, x_2 という 2 つの場合を考える。図 1(a) に内積タイプの場合を図 1(b) にユークリッド距離タイプの場合を示す。図 1(a) はカテゴリ A とカテゴリ B とが超平面で切られている状態である。内積タイプで重み w を学習していくということは、この超平面の傾きを変えていきカテゴリをうまく分割できるようにすることである。図 1(b) にはユークリッド距離タイプの 2 次元のときの状態を表しているので円で示される。ユークリッド距離タイプで重み w を学習するということはこの円の中心の位置を決めることになる。高次元だと超球の中心の位置を決めることになる。したがって認識カテゴリの中心ベクトルを決めるものと考えられる。RBF を用いた素子の出力はこの円の中心で値 1 をとり、中心から離れるにつれて指数関数的に値が小さくなる。さらに RBF を FPM に適用すると 1 素子内の出力値の合計が 1 になるように正規化される。

3 Fuzzy Partition Model

3.1 Fuzzy Partition Model

FPM は内積タイプのフィードフォワードニューラルネットワークであるが、図 2 に示すように多入出力素子で構成される点が従来の BP モデル [3] と異なる。以下 N 個の出力をもつ FPM 素子を N 次元の素子と呼ぶことにする。第 m 層第 s 素子の中の k 番目の入力、出力をそれぞれ $u_{km}^{(s)}$, $a_{km}^{(s)}$ とし、$a_{jm-1}^{(g)}$ と $u_{km}^{(s)}$ とを介する重みを $w_{km\,jm-1}^{(sg)}$ とおけば素子の入出力関係は次式で表すことができる。但し、第 m 層の素子数を M^m と表す。

$$a_{km}^{(s)} = \frac{\exp(u_{km}^{(s)})}{\sum_{j=1}^{N^m}\exp(u_{jm}^{(s)})} \qquad (k=1,\ldots,N) \tag{4}$$

$$u_{km}^{(s)} = \sum_{g=1}^{M^{m-1}}\sum_{j=1}^{N^{m-1}} w_{km\,jm-1}^{(sg)} a_{jm-1}^{(g)} \qquad (k=1,\ldots,N) \tag{5}$$

式 (4) は、$\exp(u_{km}^{(s)})$ を正規化することを意味するから、素子内の出力は常に正でその総和は 1 となり、1 素子内の出力群の関係は次式で制限される。

$$\sum_{j=1}^{N^m} a_{jm}^{(s)} = 1 \tag{6}$$

$$0 \le a_{jm}^{(s)} \le 1 \qquad (\forall j) \tag{7}$$

次に FPM の学習アルゴリズムについて述べる。学習には最急降下法を用い、出力層の教師と出力との誤差が最小になるように素子間の重みを逐次的に変化させてモデルを収束させる。出力層の誤差評価関数には Kullback ダイバージェンス D を用いた。学習率を η とすれば重み修正量は次式で与えられる。

$$\Delta w_{km\,jm-1}^{(sv)} = -\eta \frac{\partial D}{\partial w_{km\,jm-1}^{(sv)}} \tag{8}$$

$$D = \sum_{g=1}^{M^m}\sum_{j=1}^{N^m} t_j^{(g)} \log \frac{t_j^{(g)}}{a_{jm}^{(g)}} \tag{9}$$

式 (9) を式 (8) に代入し、過去に変更した修正量の影響量 α も考慮すれば次式になる。

$$\Delta w_{km\,jm-1}^{(sv)} = \eta \delta_{km}^{(s)} a_{jm-1}^{(v)} + \alpha \Delta w_{km\,jm-1}^{(sv)} \tag{10}$$

$\delta_{km}^{(s)}$ は層によって異なり、m が出力層の場合には式 (11) 、それ以外の場合には式 (12) となる。

$$\delta_{km}^{(s)} = t_{km}^{(s)} - a_{km}^{(s)} \tag{11}$$

$$\delta_{km}^{(s)} = a_{km}^{(s)}(\sigma_{km}^{(s)} - \sum_{j=1}^{N^m} a_{jm}^{(s)} \sigma_{jm}^{(s)}) \tag{12}$$

但し、

$$\sigma_{km}^{(s)} = \sum_{g=1}^{M^{m+1}}\sum_{i=1}^{N^{m+1}} w_{im+1\,km}^{(gs)} \delta_{im+1}^{(g)} \tag{13}$$

学習アルゴリズムの詳細は文献 [5] で報告されている。

ここで説明したものはオリジナルの FPM とは多少異なっている。オリジナルの FPM は素子の N 番目の入力が 0 で固定されているが、今回実験に用いた FPM は RBF と整合性をとるために N 番目の入力に式 (5) を用いている。

3.2 Radial Basis Functions の適用

FPM に RBF を適用するとその素子の入出力関係は次式のようになる。

$$a_{km}^{(s)} = \frac{\exp(-u_{km}^{(s)})}{\sum_{j=1}^{N^m}\exp(-u_{jm}^{(s)})} \qquad (k=1,\ldots,N) \tag{14}$$

$$u_{km}^{(s)} = \sum_{g=1}^{M^{m-1}}\sum_{j=1}^{N^{m-1}} (w_{km\,jm-1}^{(sg)} - a_{jm-1}^{(g)})^2 \qquad (k=1,\ldots,N) \tag{15}$$

表 1: 入力音声の分析条件

サンプリング周波数	12kHz
シフト幅	5ms
窓関数	21.3ms ハミング窓 (256 点)
周波数分析	256 点 FFT メルスケール 16 チャネル
入力パラメータ	10ms 毎にまとめ 0.0~1.0 の間に 平均値 0.5 として正規化 112 次元

また学習アルゴリズムは以下のようになる。前節と同様、誤差評価関数に Kullback ダイバージェンス D を用い、学習率を η とし、過去に変更した修正量の影響量 α を考慮すれば重み修正量は次式になる。

$$\Delta w^{(sv)}_{k_m l_{m-1}} = \eta \delta^{(s)}_{k_m} (w^{(sv)}_{k_m l_{m-1}} - a^{(v)}_{l_{m-1}}) + \alpha \Delta w^{(sv)}_{k_m l_{m-1}} \tag{16}$$

$\delta^{(s)}_{k_m}$ は層によって異なり、m が出力層の場合には式 (17) 、それ以外の場合には式 (18) となる。

$$\delta^{(s)}_{k_m} = -2(t^{(s)}_{k_m} - a^{(s)}_{k_m}) \tag{17}$$

$$\delta^{(s)}_{k_m} = 2a^{(s)}_{k_m}(\sigma^{(s)}_{k_m} - \sum_{j=1}^{N^m} a^{(s)}_{j_m} \sigma^{(s)}_{j_m}) \tag{18}$$

但し、

$$\sigma^{(s)}_{k_m} = \sum_{g=1}^{M^{m+1}} \sum_{i=1}^{N^{m+1}} (w^{(gs)}_{i_{m+1} k_m} - a^{(s)}_{k_m}) \delta^{(g)}_{i_{m+1}} \tag{19}$$

学習アルゴリズムの詳細な導出は付録 A に示す。音声認識実験においては FPM の全ての層に RBF を適用したものと FPM の出力層だけに RBF を適用したものを用いている。この節では前者の方の学習アルゴリズムを述べてある。後者の学習アルゴリズムは出力層においてはここで述べたものと同じであるが、中間層では式 (16) の代わりに式 (10) を、式 (18) の代わりに式 (12) を使う。

4 音素認識実験

内積タイプ および RBF タイプ を適用した FPM を用いて日本語 6 子音 (/b, d, g, m, n, N/) と日本語 18 子音 (/b, d, g, p, t, k, ch, ts, s, sh, h, z, m, n, N, r, w, y/) の音素認識実験を行った。

4.1 実験条件

学習には、ATR データベース (男性話者一人) の 5240 単語の偶数番単語 (5.7 モーラ / 秒) からラベルを用いて切り出した音素を用いた。学習サンプルは、6 子音 /b,d,g,m,n,N/ の学習では 1 音素クラス当たり 500 (重複を含む) の計 3000 サンプル、18 子音の学習では 1 音素クラス当たり 1200 (重複を含む) の計 21600 サンプルを用いた。評価には、ATR データベースから 5240 単語の残りの奇数番単語 (5.7 モーラ / 秒) 、文節発声 (7.7 モーラ / 秒) 、短い文節発声 (7.1 モーラ / 秒) 、および文発声 (9.6 モーラ / 秒) のデータからラベルを用いて切り出した音素を用いた。音素は 1 フレーム (10ms) 当たりメルスケール 16 チャネルで周波数分析し、7 フレーム (70ms) 分を 0.0 から 1.0 の間で正規化し、モデルの入力パラメータとした。入力音声の分析条件を表 1 に示す。

実験モデルは 中間層に内積タイプ、出力層だけに RBF タイプを適用した FPM (以下 IP-RBF と呼ぶことにする) 、および出力層、中間層共に RBF を適用した RBF (以下 RBF-RBF と呼ぶことにする) である。モデルの概略図を図 3 に示す。さらに比較のため、中間層は従来の BP モデル、出力層は RBF を適用した FPM という構造のモデル (以下 BP-RBF と呼ぶことにする) でも実験を行った。また各モデルは 3 層構造であり、実験結果を TDNN[8]

表 2: 各モデルのパラメータ数 (6 子音認識の場合)

モデル	入力層素子数 (次元、素子)	中間層素子数 (次元、素子)	出力層素子数 (次元、素子)	パラメータ数
TDNN	112	第 1 層 50 第 2 層 24	6	628
FPM	(112,1)	(2,3)	(6,1)	708
IP-RBF	(112,1)	(2,3)	(6,1)	708
RBF-RBF	(112,1)	(2,3)	(6,1)	708
BP-RBF	112	5	(6,1)	595

表 3: 各モデルのパラメータ数 (18 子音認識の場合)

モデル	入力層素子数 (次元、素子)	中間層素子数 (次元、素子)	出力層素子数 (次元、素子)	パラメータ数
TDNN	112	第 1 層 200 第 2 層 54	18	3454
FPM	(112,1)	(2,13)	(18,1)	3380
IP-RBF	(112,1)	(2,13)	(18,1)	3380
RBF-RBF	(112,1)	(2,13)	(18,1)	3380
BP-RBF	112	26	(18,1)	3406

の結果と比較するため、過去に同様の実験に用いられた TDNN のパラメータ数 [10, 11] とほぼ同等のパラメータ数とした。各モデルのパラメータ数を表 2 、表 3 に示す。

ネットワークの学習は TDNN の実験結果と比較するため、6 子音認識実験では学習データに対して第 1 位の認識率が 98.9% になるまで、または学習回数が 1000 回まで学習を繰り返し、18 子音認識実験では学習データに対して第 1 位の認識率が 97.5% になるまで学習を繰り返した。学習率 η の初期値は 0.001 とし、振動しはじめると値をその 1/10 に変更していった。 α は 0.9 とした。また式 (3) の β は 1 とした。

4.2 実験結果

各評価データ (単語発声、文節発声、短い文節発声、文発声データ) に対する 6 子音の認識結果を表 4 に、18 子音の認識結果を表 5 に示す。

- 6 子音の認識結果
  - 第 1 位、累積認識率共に BP-RBF の結果が最も高い。また第 1 位の認識率では TDNN もかなり良い結果がでているが累積認識率は最も低い。
  - FPM、IP-RBF、RBF-RBF はほぼ同等の結果が得られた。
- 18 子音の認識結果
  - 学習データと発話速度に差がある文節発声、文発声データに対しては FPM、IP-RBF の結果が最も良い。また両者の差はほとんどない。
  - 学習データと発話速度に差がない単語発声データに対しては BP-RBF、TDNN の認識率が高い。しかし文節発声、文発声データに対しては認識率が低い。特に累積認識率が低い。

5 連続音声認識実験

各 FPM と LR パーサ [9] を統合した認識システムを用いて連続音声認識実験を行った。

表 4: 6 子音認識結果 (単位 %)

モデル	単語発声	文節発声	短い文節発声	文発声
TDNN	95.2(99.8)	84.8(96.5)	84.5(95.3)	77.8(92.2)
FPM	94.0(99.9)	79.3(97.8)	80.3(97.4)	76.1(96.5)
IP-RBF	94.3(99.8)	80.1(98.0)	82.2(97.6)	75.7(97.2)
RBF-RBF	94.0(99.9)	79.7(97.6)	81.1(96.7)	75.0(96.1)
BP-RBF	96.3(99.9)	84.5(98.5)	85.7(97.9)	80.2(98.0)

(括弧内は第 3 位までの累積認識率)

表 5: 18 子音認識結果 (単位 %)

モデル	単語発声	文節発声	短い文節発声	文発声
TDNN	92.0(99.3)	73.7(89.2)	—	62.4(78.5)
FPM	90.5(99.4)	76.5(95.8)	76.5(95.1)	67.6(91.0)
IP-RBF	91.0(99.5)	76.7(95.7)	78.5(95.4)	67.9(91.9)
RBF-RBF	89.6(99.2)	74.8(94.7)	75.1(93.5)	64.6(89.2)
BP-RBF	92.3(99.5)	70.9(91.6)	70.8(91.5)	58.8(83.8)

(括弧内は第 3 位までの累積認識率)

表 6: 各モデルのパラメータ数

モデル	入力層素子数 (次元、素子)	中間層素子数 (次元、素子)	出力層素子数 (次元、素子)	パラメータ数
TDNN	112	第 1 層 1250 第 2 層 100	25	24825
FPM	(112,1)	(2,78) 2 層	(25,1)	24804
IP-RBF	(112,1)	(2,52) 2 層	(25,1)	25064
RBF-RBF	(112,1)	(2,52) 2 層	(25,1)	25064

5.1 実験条件

音素認識実験と同様 ATR データベース 5240 単語の偶数番単語からラベルを用いて切り出した 25 音素を用い、学習サンプルには 1 音素クラス当たり 2000 (重複を含む) の計 50000 サンプルを用いた。入力音声の分析条件も表 1 と同様である。

実験モデルはオリジナル FPM、IP-RBF、RBF-RBF である。モデルの概略図を図 4 に示す。各モデルは 4 層構造であり、実験結果を TDNN の結果と比較するため、過去に同様の実験に用いられた TDNN[11] のパラメータ数とほぼ同等のパラメータ数とした。各モデルのパラメータ数を表 6 に示す。

ネットワークの学習は TDNN の実験と比較するため、学習データに対して第 1 位の認識率が 98.6% になるまで学習を繰り返した。学習率 η の初期値は 0.001 とし、振動しはじめると値をその 1/10 に変更していった。α は 0.9 とした。また式 (3) の β は 1 とした。

5.2 実験結果

各 FPM と LR パーサを統合した連続音声認識システムを用いた認識率を表 7 に、また累積認識率を図 5 に示す。

音素認識実験では FPM と IP-RBF の認識率は同等であったが、連続音声認識では IP-RBF は TDNN よりは認識率が高いが従来の FPM より第 1 位の認識率で 2.2%、第 3 位までの累積認識率で 2.5% 低下した。

RBF-RBF の認識率はさらに低く第 1 位の認識率では TDNN と同等であった。第 3 位までの累積認識率では TDNN

表 7: 連続音声認識結果 (単位 %)

モデル	認識率
TDNN	66.9(82.4)
FPM	72.7(90.3)
IP-RBF	70.5(87.8)
RBF-RBF	66.2(83.8)

(括弧内は第 3 位までの累積認識率)

表 8: テンプレートマッチングによる音素認識率 (単位 %)

モデル	学習データ	単語発声	文節発声	短い文節発声	文発声
平均	59.1(84.8)	57.9(84.7)	44.9(81.6)	46.5(82.7)	42.6(79.7)

(括弧内は第 3 位までの累積認識率)

より 1.4% 高かった。

6 考察

認識実験から考えられることを以下に記す。 IP-RBF と FPM は音素認識実験では同等の結果が得られたが、連続音声認識実験では IP-RBF の認識率が第 1 位認識率で 2.2%、第 3 位までの累積認識率で 2.5% 低下した。現在 IP-RBF では重みの値に制限を設けていないので負値もとりうる。HMM をはじめとする確率的手法の有効性を考慮すると、素子の出力は正値に制限されているから重みの値も正値になるよう制限を加えた方がよいように思われる。

RBF-RBF は 6 子音認識実験では FPM, IP-RBF と同等の結果が得られたが 18 子音認識実験では認識率が低かった。RBF-RBF は出力層、中間層共に RBF を適用したモデルだが、認識率の低下は中間層の素子数に原因があるように思う。内積タイプでは学習入力パターンで形成される空間を超平面で切ることにより境界を分割するのでその特徴領域が広いが、 RBF では学習入力パターンで形成される空間を超球面で切ることにより境界を分割するのでその特徴領域が超球の中心付近だけであると考えられ、内積タイプに比較して狭い。したがって中間層の素子数を多くし、特徴点を増やす必要があると考える。もしくは式 (3) の β を 1 より小さくすることによって特徴領域を広げることが考えられる。

BP-RBF は学習データと発話速度に差がない単語発声データに対しては他のモデルに比較して高い認識率が得られている。しかし発話速度に差がある文節発声データ、文発声データに対しては音素カテゴリが少ないときは高い認識率が得られているが、音素カテゴリが多いときは認識率が低い。IP-RBF と BP-RBF のモデルの違いは中間層における構造が BP 素子か FPM 素子かの違いだけであるから発話速度に差がある評価データに対しては、中間層も FPM 素子にするのが有効である。

ここで RBF を適用したニューラルネットワークの有効性を示すため、18 子音のカテゴリ毎に学習データの特徴パラメータの平均ベクトルを計算し、評価データとテンプレートマッチングにより音素認識を行った結果を報告する。未知パタンはユークリッド距離が最小になった平均パタンのカテゴリに分類されるものとする。表 8 に実験結果を示す。実験結果から RBF を適用したニューラルネットワークはテンプレートマッチングより有効であることがわかる。従って特徴パラメータから直接平均ベクトルを求めるよりも、RBF タイプのニューラルネットワークにより平均ベクトルを変換した方が高性能な認識を行うことができる。

7 むすび

多入力多出力素子で構成される FPM に RBF を適用し、6 子音音素認識実験、18 子音認識実験、および連続音声認識実験を行った。音素認識実験では従来の FPM と RBF を適用した FPM は同等の認識率であった。連続音声認識実験の結果は、従来の FPM に比較して RBF を適用した FPM は認識率が低下したが TDNN より高い認識率が得られた。RBF を出力層に適用した FPM と、出力層、中間層共に RBF を適用した FPM では前者の方が認識率が高かった。また、中間層における FPM 素子の有効性が示された。今後以下のことについてさらに調べる必要がある。

- RBF を入力層に使用した場合の認識性能を評価する。
- 素子間の重みにも FPM の出力のように制約条件を導入した場合の認識性能を評価する。
- RBF を超楕円で空間分割できるように拡張した場合（β を可変にする）の認識性能を評価する。
- 内積タイプと RBF タイプとの違いを理論的に説明する。

RBF 素子の重みは内積タイプに比べて解析がしやすく扱いやすい。従って、話者間の適応に応用する場合にも RBF 素子の重みを操作することによって比較的容易にネットワークの構築ができると考える。

謝辞

研究の機会を与えて頂いた樽松明社長と豊橋技術科学大学 中川聖一教授に感謝します。また親切にご助言、ご討論下さった嵯峨山茂樹室長をはじめとする音声情報処理研究室の皆様に感謝します。

参考文献

[1] D.S.Broomhead and D.Lowe, "Multivariable functional interpolation and adaptive networks," Complex Systems, Vol.2, pp. 321-355, 1988.

[2] H.Ney, "Speech recognition in a neural network framework: Discriminative training of gaussian models and mixture densities as radial basis functions," Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 573-576, May 1991.

[3] D.E.Rumelhart, G.E.Hinton and R.J.Williams, " Learning internal representations by error propagation," in Parallel Distributed Processing,Vol.1, Cambridge, MA, MIT Press, 1986

[4] Y.Tan and T.Ejima, "A network with multipartitioning units", Proc. Int. Joint Conf. Neural Networks, Washington D.C., pp. II 439-442, June 1989.

[5] 丹 康雄、江島 俊朗：" 多入 / 出力素子を用いたネットワーク Fuzzy Partition Model の提案とその基本的性質 ", 信学技報, PRU89-45 (1989-09).

[6] 加藤 喜永、杉山 雅英：" 多入出力素子をもつニューラルネットワークを用いた連続音声認識 ", 信学「マルコフモデル・ニューラルネットワークを包含する新しい音声認識手法」資料, SPREC91-2, pp. 47-48 (1992-02).

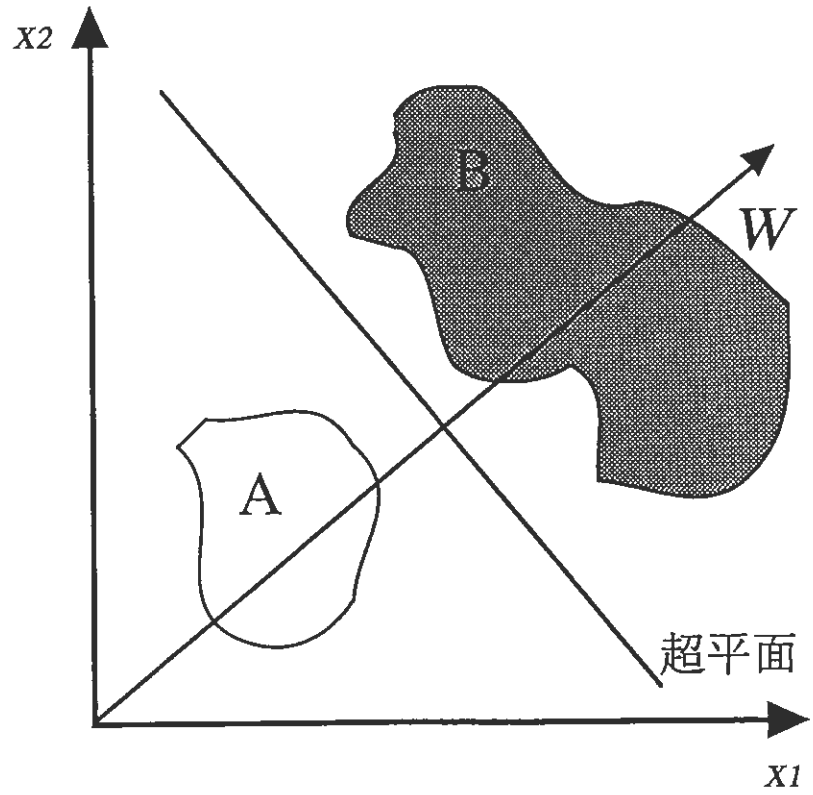
[7] D.P.Morgan, C.L.Scofield, "Neural Networks and Speech Processing," Kluwer Academic Publishers, 1991.

[8] A.Waibel, T.Hanazawa, G.Hinton, K.Shikano and K.Lang, "Phoneme recognition using Time-Delay Neural Networks," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-37, pp. 328-339, Mar 1989.

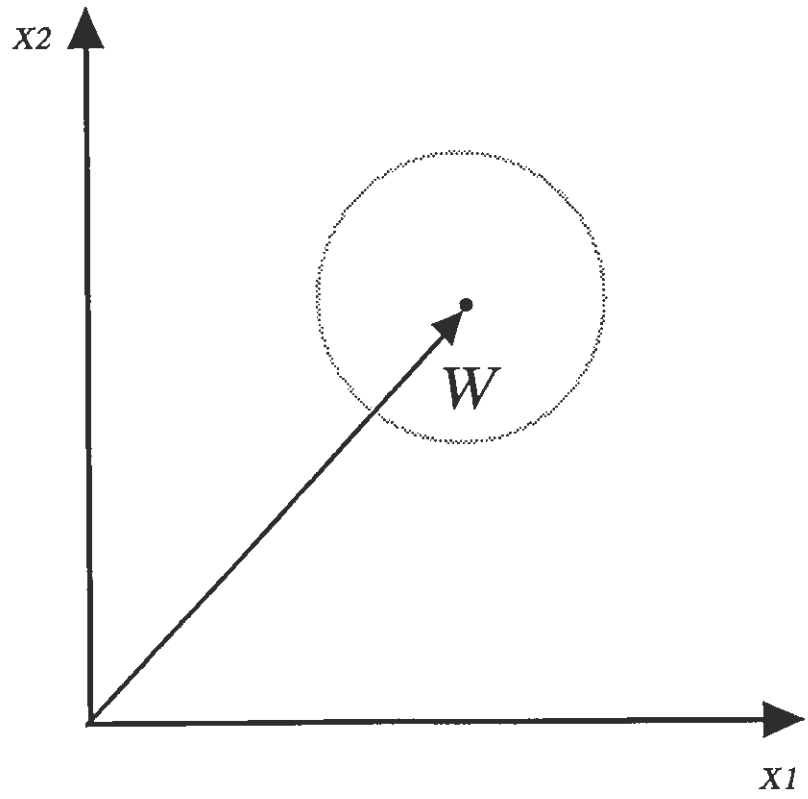
[9] 南 泰浩、沢井 秀文、宮武 正典：" 時間遅れ神経回路網 (TDNN) による音韻スポッティング法と予測 LR パーザを用いた大語い単語音声認識 ", 信学論 (D-II), J73-D-II, 6, pp. 788-795 (平 2-06).

[10] J.Takami and S.Sagayama, "Phoneme recognition by Pairwise Discriminant TDNNs," Proc. Int. Conf. Spoken Lang. Processing, pp.677-680, Nov. 1990.

[11] 小森 康弘、嵯峨山 茂樹、アレックス ワイベル：" ニューラル・ファジー学習法による TDNN-LR 連続音声認識システムの性能向上 ", 信学技法, SP91-18, pp. 49-56 (1991-06).



(a) 内積タイプ



(b) RBF タイプ

図1 ニューラルネットワークが分割する空間

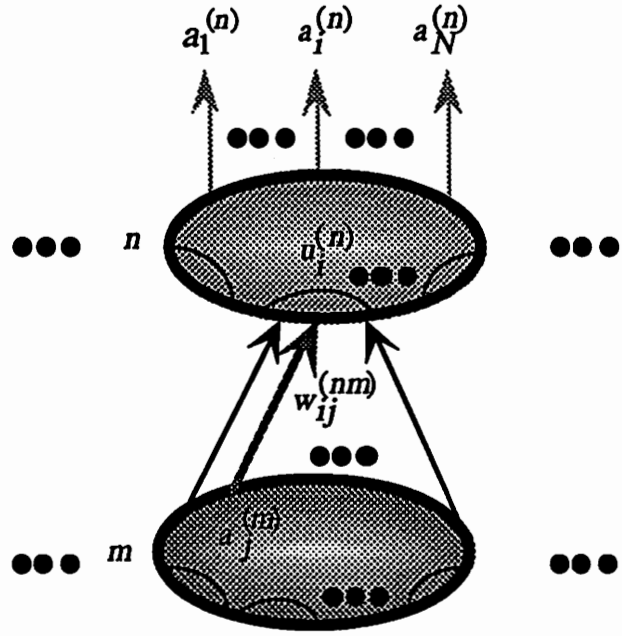


図2 N 次元のFPM 素子

$$\sum_{i=1}^{N} a_i^{(n)} = 1, 0 \le a_i^{(n)} \le 1, \text{for all } i$$

・内積タイプ

$$a_i^{(n)} = \frac{\exp(u_i^{(n)})}{\sum_k^N \exp(u_k^{(n)})}, \quad (i = 1, \ldots, N)$$

$$u_i^{(n)} = \sum_m \sum_j w_{ij}^{(nm)} a_j^{(m)}, \qquad (i = 1, \ldots, N).$$

・RBF タイプ

$$a_i^{(n)} = \frac{\exp(-u_i^{(n)})}{\sum_k^N \exp(-u_k^{(n)})}, \quad (i = 1, \ldots, N)$$

$$u_i^{(n)} = \sum_m \sum_j (w_{ij}^{(nm)} - a_j^{(m)})^2, \quad (i = 1, \ldots, N).$$

$a_i^{(n)}$: n 番目の素子, i 番目からの出力
$u_i^{(n)}$: n 番目の素子, i 番目への入力
$w_{ij}^{(nm)}$: m 番目の素子, j 番目と n 番目の素子, i 番目との間の重み

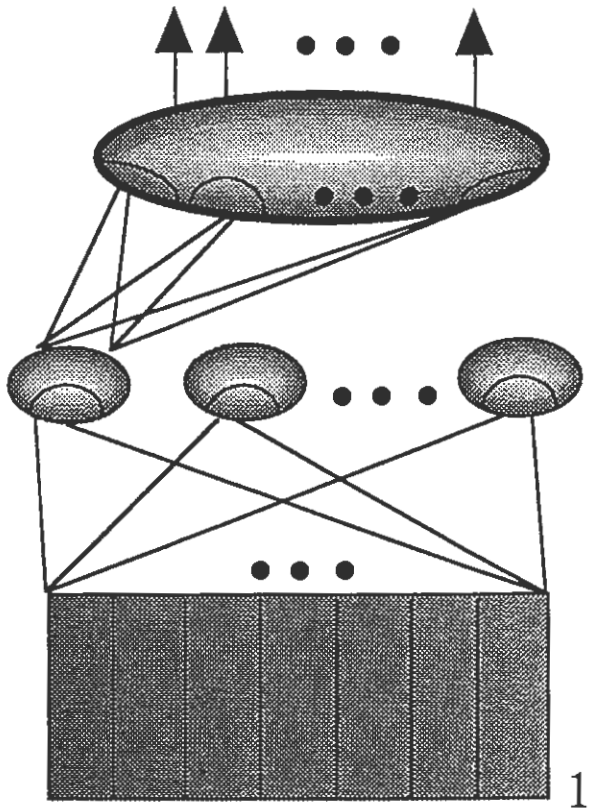


図3 音素認識に用いる FPM

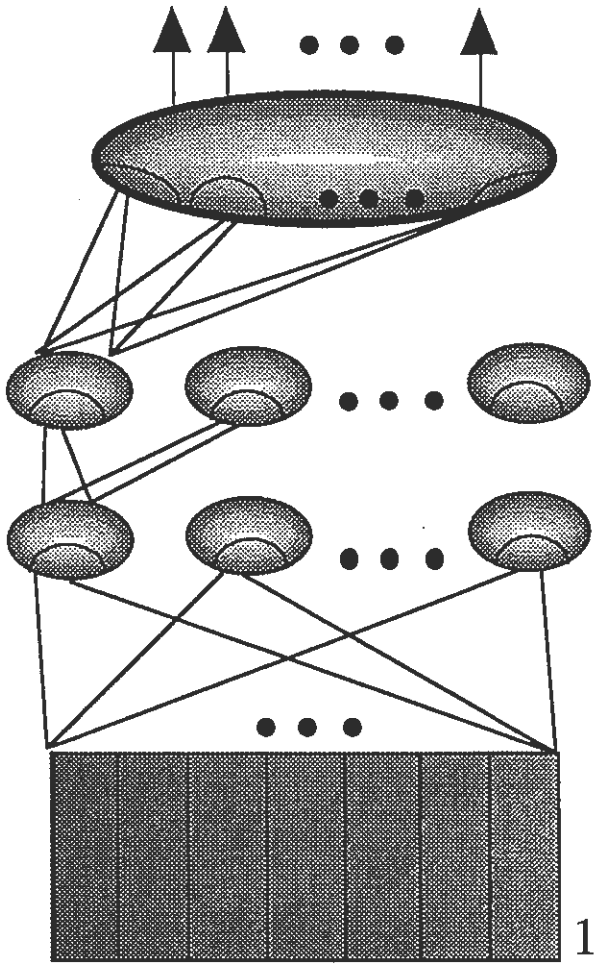


図4 連続音声認識に用いる FPM

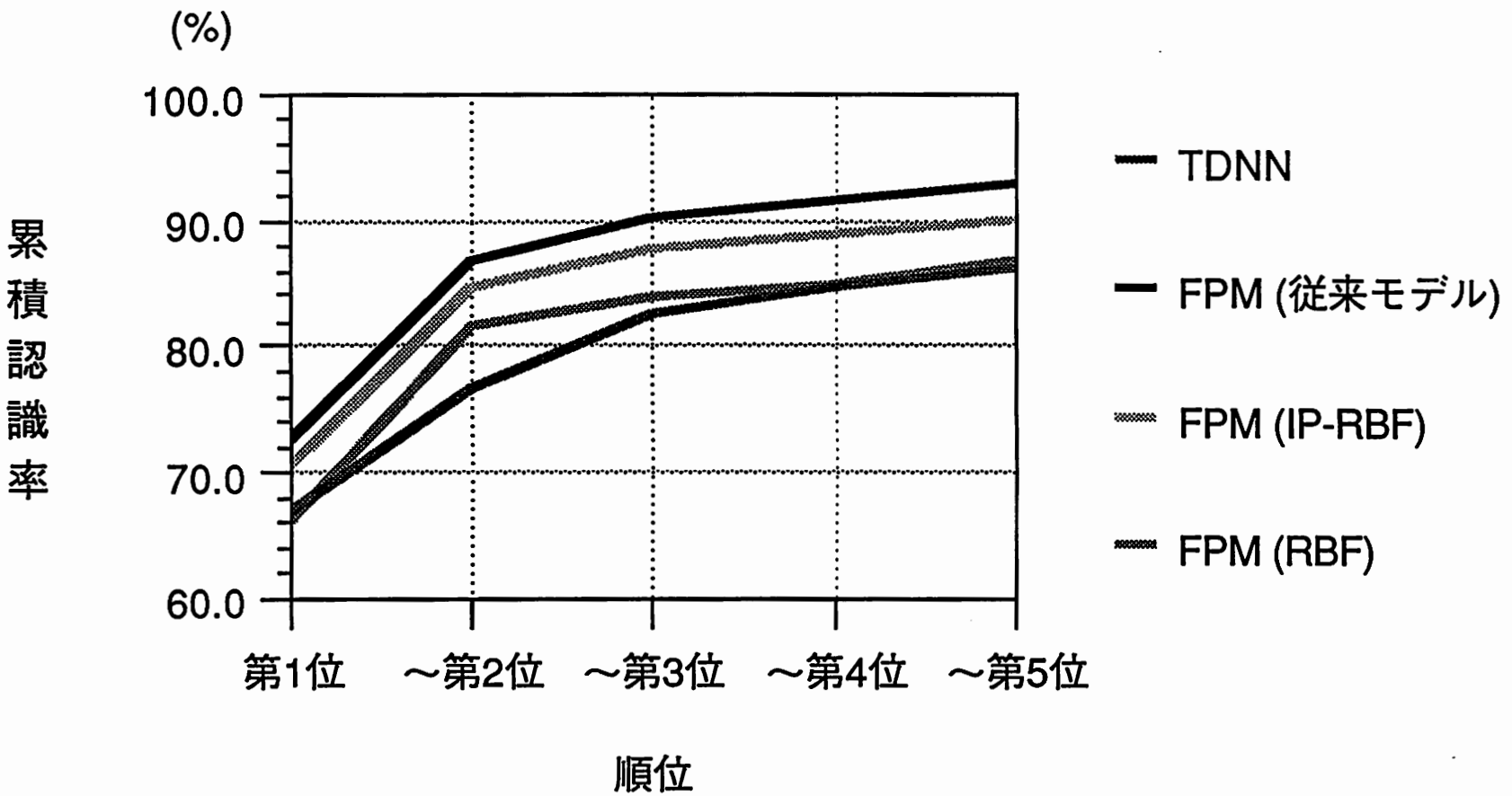

(%)
100.0
90.0
80.0
70.0
60.0
累積認識率
第1位
～第2位
～第3位
～第4位
～第5位
順位
TDNN
FPM (従来モデル)
FPM (IP-RBF)
FPM (RBF)

図5 文節認識率

付録

A　RBF を適用した FPM の学習アルゴリズム

第 m 層 → 第 (m-1) 層

$$D = \sum_{r}^{M^m} \sum_{i}^{N^m} t_{im}^{(r)} \log \frac{t_{im}^{(r)}}{a_{im}^{(r)}}$$

$$a_{im}^{(s)} = \frac{\exp(-u_{im}^{(s)})}{\sum_{q}^{N^m} \exp(-u_{qm}^{(s)})} \qquad (i = 1, \ldots, N)$$

$$u_{im}^{(s)} = \sum_{g}^{M^{m-1}} \sum_{j}^{N^{m-1}} \left(w_{im\,jm-1}^{(sg)} - a_{jm-1}^{(g)} \right)^2 \qquad (i = 1, \ldots, N)$$

$$-\frac{\partial D}{\partial w_{km\,lm-1}^{(sv)}} = -\sum_{r} \sum_{i} \frac{\partial D}{\partial a_{im}^{(s)}} \frac{\partial a_{im}^{(s)}}{\partial w_{km\,lm-1}^{(sv)}}$$

$$= \sum_{i} \frac{t_{im}^{(s)}}{a_{im}^{(s)}} \frac{\partial a_{im}^{(s)}}{\partial u_{km}^{(s)}} \frac{\partial u_{km}^{(s)}}{\partial w_{km\,lm-1}^{(sv)}}$$

$$= \left(\sum_{i \neq k} \frac{t_{im}^{(s)}}{a_{im}^{(s)}} \frac{\partial a_{im}^{(s)}}{\partial u_{km}^{(s)}} + \frac{t_{km}^{(s)}}{a_{km}^{(s)}} \frac{\partial a_{km}^{(s)}}{\partial u_{km}^{(s)}} \right) \frac{\partial u_{km}^{(s)}}{\partial w_{km\,lm-1}^{(sv)}}$$

$$= \left[\sum_{i \neq k} \frac{t_{im}^{(s)}}{a_{im}^{(s)}} \frac{\exp(-u_{im}^{(s)}) \exp(-u_{km}^{(s)})}{\left\{ \sum_{q} \exp(-u_{qm}^{(s)}) \right\}^2} - \frac{t_{km}^{(s)}}{a_{km}^{(s)}} \frac{\exp(-u_{km}^{(s)}) \sum_{i} \exp(-u_{im}^{(s)}) + \left\{ \exp(-u_{km}^{(s)}) \right\}^2}{\left\{ \sum_{q} \exp(-u_{qm}^{(s)}) \right\}^2} \right] \frac{\partial u_{km}^{(s)}}{\partial w_{km\,lm-1}^{(sv)}}$$

$$= \left\{ \sum_{i \neq k} \frac{t_{im}^{(s)}}{a_{im}^{(s)}} a_{im}^{(s)} a_{km}^{(s)} - \frac{t_{km}^{(s)}}{a_{km}^{(s)}} a_{km}^{(s)} \left(1 - a_{km}^{(s)} \right) \right\} \frac{\partial u_{km}^{(s)}}{\partial w_{km\,lm-1}^{(sv)}}$$

$$= 2 \left\{ \sum_{i \neq k} t_{im}^{(s)} a_{km}^{(s)} - t_{km}^{(s)} \left(1 - a_{km}^{(s)} \right) \right\} \left(w_{km\,lm-1}^{(sv)} - a_{lm-1}^{(v)} \right)$$

$$= 2 \left\{ \left(1 - t_{km}^{(s)} \right) a_{km}^{(s)} - t_{km}^{(s)} \left(1 - a_{km}^{(s)} \right) \right\} \left(w_{km\,lm-1}^{(sv)} - a_{lm-1}^{(v)} \right)$$

$$= -2 \left(t_{km}^{(s)} - a_{km}^{(s)} \right) \left(w_{km\,lm-1}^{(sv)} - a_{lm-1}^{(v)} \right)$$

第 $(m\text{-}1)$ 層 → 第 $(m\text{-}2)$ 層

$$-\frac{\partial D}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}} = -\sum_{r,i}\frac{\partial D}{\partial a^{(r)}_{i^m}}\sum_{j}\frac{\partial a^{(r)}_{i^m}}{\partial a^{(s)}_{j^{m-1}}}\frac{\partial a^{(s)}_{j^{m-1}}}{\partial u^{(s)}_{k^{m-1}}}\frac{\partial u^{(s)}_{k^{m-1}}}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}}$$

$$= -2\sum_{r,i}\frac{t^{(r)}_{i^m}}{a^{(r)}_{i^m}}\sum_{j}\left[\begin{array}{l} -\exp(-u^{(r)}_{i^m})\sum_{q}\exp(-u^{(r)}_{q^m})\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right) \\ \qquad +\exp(-u^{(r)}_{i^m})\sum_{q}\exp(-u^{(r)}_{q^m})\left(w^{(rs)}_{q^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right) \end{array}\right]$$

$$\cdot\frac{1}{\left\{\sum_{q}\exp(-u^{(r)}_{q^m})\right\}^2}\frac{\partial a^{(s)}_{j^{m-1}}}{\partial u^{(s)}_{k^{m-1}}}\frac{\partial u^{(s)}_{k^{m-1}}}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}}$$

$$= -2\sum_{r,i}\frac{t^{(r)}_{i^m}}{a^{(r)}_{i^m}}\sum_{j}\left\{-a^{(r)}_{i^m}\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right) + a^{(r)}_{i^m}\sum_{q}a^{(r)}_{q^m}\left(w^{(rs)}_{q^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right)\right\}\frac{\partial a^{(s)}_{j^{m-1}}}{\partial u^{(s)}_{k^{m-1}}}\frac{\partial u^{(s)}_{k^{m-1}}}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}}$$

$$= -2\sum_{r,i}\left\{\sum_{i}-t^{(r)}_{i^m}\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right) + \sum_{i}t^{(r)}_{i^m}\sum_{q}a^{(r)}_{q^m}\left(w^{(rs)}_{q^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right)\right\}\frac{\partial a^{(s)}_{j^{m-1}}}{\partial u^{(s)}_{k^{m-1}}}\frac{\partial u^{(s)}_{k^{m-1}}}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}}$$

$$= 2\sum_{r,i}\sum_{j}\left\{t^{(r)}_{i^m} - a^{(r)}_{i^m}\right\}\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right)\frac{\partial a^{(s)}_{j^{m-1}}}{\partial u^{(s)}_{k^{m-1}}}\frac{\partial u^{(s)}_{k^{m-1}}}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}}$$

$$= 2\sum_{r,i}\left\{t^{(r)}_{i^m} - a^{(r)}_{i^m}\right\}\left\{\sum_{j\neq k}\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right)a^{(s)}_{j^{m-1}}a^{(s)}_{k^{m-1}} - \left(w^{(rs)}_{i^m k^{m-1}} - a^{(s)}_{k^{m-1}}\right)a^{(s)}_{k^{m-1}}\left(1 - a^{(s)}_{k^{m-1}}\right)\right\}$$

$$\cdot\frac{\partial u^{(s)}_{k^{m-1}}}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}}$$

$$= 2\sum_{r,i}\left\{t^{(r)}_{i^m} - a^{(r)}_{i^m}\right\}\left\{\sum_{j}\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right)a^{(s)}_{j^{m-1}}a^{(s)}_{k^{m-1}} - \left(w^{(rs)}_{i^m k^{m-1}} - a^{(s)}_{k^{m-1}}\right)a^{(s)}_{k^{m-1}}\right\}\frac{\partial u^{(s)}_{k^{m-1}}}{\partial w^{(sv)}_{k^{m-1}l^{m-2}}}$$

$$= -2\left\{\sum_{r}^{M^m}\sum_{i}^{N^m}2\left(t^{(r)}_{i^m} - a^{(r)}_{i^m}\right)\left(w^{(rs)}_{i^m k^{m-1}} - a^{(s)}_{k^{m-1}}\right) - \sum_{r}^{M^m}\sum_{i}^{N^m}2\left(t^{(r)}_{i^m} - a^{(r)}_{i^m}\right)\sum_{j}^{N^{m-1}}\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right)a^{(s)}_{j^{m-1}}\right\}$$

$$\cdot\, a^{(s)}_{k^{m-1}}\left(w^{(sv)}_{k^{m-1}l^{m-2}} - a^{(v)}_{l^{m-2}}\right)$$

$$= 2\left\{\sum_{r}^{M^m}\sum_{i}^{N^m}\delta^{(r)}_{i^m}\left(w^{(rs)}_{i^m k^{m-1}} - a^{(s)}_{k^{m-1}}\right) - \sum_{r}^{M^m}\sum_{i}^{N^m}\delta^{(r)}_{i^m}\sum_{j}^{N^{m-1}}\left(w^{(rs)}_{i^m j^{m-1}} - a^{(s)}_{j^{m-1}}\right)a^{(s)}_{j^{m-1}}\right\}a^{(s)}_{k^{m-1}}\left(w^{(sv)}_{k^{m-1}l^{m-2}} - a^{(v)}_{l^{m-2}}\right)$$

where $\quad \delta^{(r)}_{i^m} = -2\left(t^{(r)}_{i^m} - a^{(r)}_{i^m}\right)$

$$= 2\left(\sigma^{(s)}_{k^{m-1}} - \sum_{j}^{N^{m-1}}\sigma^{(s)}_{j^{m-1}}a^{(s)}_{j^{m-1}}\right)a^{(s)}_{k^{m-1}}\left(w^{(sv)}_{k^{m-1}l^{m-2}} - a^{(v)}_{l^{m-2}}\right)$$

where $\quad \sigma^{(s)}_{k^{m-1}} = \sum_{r}^{M^m}\sum_{i}^{N^m}\delta^{(r)}_{i^m}\left(w^{(rs)}_{i^m k^{m-1}} - a^{(s)}_{k^{m-1}}\right)$

第 $(m\text{-}2)$ 層 → 第 $(m\text{-}3)$ 層

$$-\frac{\partial D}{\partial w^{(sv)}_{k^{m-2}l^{m-3}}} = -\sum_{r,i}\frac{\partial D}{\partial a^{(r)}_{i^m}}\sum_{p,j}\frac{\partial a^{(r)}_{i^m}}{\partial a^{(p)}_{j^{m-1}}}\sum_{f}\frac{\partial a^{(p)}_{j^{m-1}}}{\partial a^{(s)}_{f^{m-2}}}\frac{\partial a^{(s)}_{f^{m-2}}}{\partial u^{(s)}_{k^{m-2}}}\frac{\partial u^{(s)}_{k^{m-2}}}{\partial w^{(sv)}_{k^{m-2}l^{m-3}}}$$

$$= -2\sum_{p,j}\left\{\sum_{r,i}\left(t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\left(w^{(rp)}_{i^m j^{m-1}}-a^{(p)}_{j^{m-1}}\right)-\sum_{r,i}\left(t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\sum_{q}\left(w^{(rp)}_{i^m q^{m-1}}-a^{(p)}_{q^{m-1}}\right)a^{(p)}_{q^{m-1}}\right\}a^{(p)}_{j^{m-1}}$$

$$\cdot\sum_{f}\frac{\partial u^{(p)}_{j^{m-1}}}{\partial a^{(s)}_{f^{m-2}}}\frac{\partial a^{(s)}_{f^{m-2}}}{\partial u^{(s)}_{k^{m-2}}}\frac{\partial u^{(s)}_{k^{m-2}}}{\partial w^{(sv)}_{k^{m-2}l^{m-3}}}$$

$$= -2\sum_{p,j}\left\{\sum_{r,i}\left(t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\left(w^{(rp)}_{i^m j^{m-1}}-a^{(p)}_{j^{m-1}}\right)-\sum_{r,i}\left(t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\sum_{q}\left(w^{(rp)}_{i^m q^{m-1}}-a^{(p)}_{q^{m-1}}\right)a^{(p)}_{q^{m-1}}\right\}a^{(p)}_{j^{m-1}}$$

$$\cdot(-2)\sum_{f}\left(w^{(ps)}_{j^{m-1}f^{m-2}}-a^{(s)}_{f^{m-2}}\right)\frac{\partial a^{(s)}_{f^{m-2}}}{\partial u^{(s)}_{k^{m-2}}}\frac{\partial u^{(s)}_{k^{m-2}}}{\partial w^{(sv)}_{k^{m-2}l^{m-3}}}$$

$$= -2\sum_{p,j}\left\{\sum_{r,i}2\left(t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\left(w^{(rp)}_{i^m j^{m-1}}-a^{(p)}_{j^{m-1}}\right)-\sum_{r,i}2\left(t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\sum_{q}\left(w^{(rp)}_{i^m q^{m-1}}-a^{(p)}_{q^{m-1}}\right)a^{(p)}_{q^{m-1}}\right\}a^{(p)}_{j^{m-1}}$$

$$\cdot\left\{\left(w^{(ps)}_{j^{m-1}k^{m-2}}-a^{(s)}_{k^{m-2}}\right)-\sum_{f}\left(w^{(ps)}_{j^{m-1}f^{m-2}}-a^{(s)}_{f^{m-2}}\right)a^{(s)}_{f^{m-2}}\right\}a^{(s)}_{k^{m-2}}\frac{\partial u^{(s)}_{k^{m-2}}}{\partial w^{(sv)}_{k^{m-2}l^{m-3}}}$$

$$= -2\left[\begin{array}{l}\sum_{p,j}\left\{\begin{array}{l}\sum_{r,i}\left(2t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\left(w^{(rp)}_{i^m j^{m-1}}-a^{(p)}_{j^{m-1}}\right)\\ -\sum_{r,i}\left(2t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\sum_{q}\left(w^{(rp)}_{i^m q^{m-1}}-a^{(p)}_{q^{m-1}}\right)a^{(p)}_{q^{m-1}}\end{array}\right\}a^{(p)}_{j^{m-1}}\left(w^{(ps)}_{j^{m-1}k^{m-2}}-a^{(s)}_{k^{m-2}}\right)\\ -\sum_{p,j}\left\{\begin{array}{l}\sum_{r,i}\left(2t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\left(w^{(rp)}_{i^m j^{m-1}}-a^{(p)}_{j^{m-1}}\right)\\ -\sum_{r,i}\left(2t^{(r)}_{i^m}-a^{(r)}_{i^m}\right)\sum_{q}\left(w^{(rp)}_{i^m q^{m-1}}-a^{(p)}_{q^{m-1}}\right)a^{(p)}_{q^{m-1}}\end{array}\right\}a^{(p)}_{j^{m-1}}\sum_{f}\left(w^{(ps)}_{j^{m-1}f^{m-2}}-a^{(s)}_{f^{m-2}}\right)a^{(s)}_{f^{m-2}}\end{array}\right]$$

$$\cdot 2a^{(s)}_{k^{m-2}}\left(w^{(sv)}_{k^{m-2}l^{m-3}}-a^{(v)}_{l^{m-3}}\right)$$

$$= 2\left[\begin{array}{l}\sum_{p,j}\left\{\begin{array}{l}\sum_{r,i}\delta^{(r)}_{i^m}\left(w^{(rp)}_{i^m j^{m-1}}-a^{(p)}_{j^{m-1}}\right)\\ -\sum_{r,i}\delta^{(r)}_{i^m}\sum_{q}\left(w^{(rp)}_{i^m q^{m-1}}-a^{(p)}_{q^{m-1}}\right)a^{(p)}_{q^{m-1}}\end{array}\right\}a^{(p)}_{j^{m-1}}\left(w^{(ps)}_{j^{m-1}k^{m-2}}-a^{(s)}_{k^{m-2}}\right)\\ -\sum_{p,j}\left\{\begin{array}{l}\sum_{r,i}\delta^{(r)}_{i^m}\left(w^{(rp)}_{i^m j^{m-1}}-a^{(p)}_{j^{m-1}}\right)\\ -\sum_{r,i}\delta^{(r)}_{i^m}\sum_{q}\left(w^{(rp)}_{i^m q^{m-1}}-a^{(p)}_{q^{m-1}}\right)a^{(p)}_{q^{m-1}}\end{array}\right\}a^{(p)}_{j^{m-1}}\sum_{f}\left(w^{(ps)}_{j^{m-1}f^{m-2}}-a^{(s)}_{f^{m-2}}\right)a^{(s)}_{f^{m-2}}\end{array}\right]$$

$$\cdot 2a^{(s)}_{k^{m-2}}\left(w^{(sv)}_{k^{m-2}l^{m-3}}-a^{(v)}_{l^{m-3}}\right)$$

$$= 2\left[\begin{array}{l}\sum_{p,j}2\left(\sigma^{(p)}_{j^{m-1}}-\sum_{q}\sigma^{(p)}_{q^{m-1}}a^{(p)}_{q^{m-1}}\right)a^{(p)}_{j^{m-1}}\left(w^{(ps)}_{j^{m-1}k^{m-2}}-a^{(s)}_{k^{m-2}}\right)\\ \quad -\sum_{p,j}2\left(\sigma^{(p)}_{j^{m-1}}-\sum_{q}\sigma^{(p)}_{q^{m-1}}a^{(p)}_{q^{m-1}}\right)a^{(p)}_{j^{m-1}}\sum_{f}\left(w^{(ps)}_{j^{m-1}f^{m-2}}-a^{(s)}_{f^{m-2}}\right)a^{(s)}_{f^{m-2}}\end{array}\right]$$

$$\cdot a^{(s)}_{k^{m-2}}\left(w^{(sv)}_{k^{m-2}l^{m-3}}-a^{(v)}_{l^{m-3}}\right)$$

$$= 2\left[\sum_{p,j}\delta^{(p)}_{j^{m-1}}\left(w^{(ps)}_{j^{m-1}k^{m-2}}-a^{(s)}_{k^{m-2}}\right)-\sum_{p,j}\delta^{(p)}_{j^{m-1}}\sum_{f}\left(w^{(ps)}_{j^{m-1}f^{m-2}}-a^{(s)}_{f^{m-2}}\right)a^{(s)}_{f^{m-2}}\right]a^{(s)}_{k^{m-2}}\left(w^{(sv)}_{k^{m-2}l^{m-3}}-a^{(v)}_{l^{m-3}}\right)$$

$$\text{where}\qquad \delta^{(p)}_{j^{m-1}} = 2\left(\sigma^{(p)}_{j^{m-1}}-\sum_{q}^{N^{m-1}}\sigma^{(p)}_{q^{m-1}}a^{(p)}_{q^{m-1}}\right)a^{(p)}_{j^{m-1}}$$

$$= 2\left(\sigma^{(s)}_{k^{m-2}}-\sum_{f}^{N^{m-2}}\sigma^{(s)}_{f^{m-2}}a^{(s)}_{f^{m-2}}\right)a^{(s)}_{k^{m-2}}\left(w^{(sv)}_{k^{m-2}l^{m-3}}-a^{(v)}_{l^{m-3}}\right)$$

where $\sigma_{k^{m-2}}^{(s)} = \sum_{p}^{M^{m-1}} \sum_{j}^{N^{m-1}} \delta_{j^{m-1}}^{(p)} \left(w_{j^{m-1}k^{m-2}}^{(ps)} - a_{k^{m-2}}^{(s)} \right)$

以上の計算より $\delta_{k^{m-1}}^{(s)}$ は再帰的に計算でき、m が出力層の場合には

$$\delta_{k^m}^{(s)} = -2 \left(t_{k^m}^{(s)} - a_{k^m}^{(s)} \right)$$

それ以外の場合には

$$\delta_{k^m}^{(s)} = 2 \left(\sigma_{k^m}^{(s)} - \sum_{q}^{N^m} \sigma_{q^m}^{(s)} a_{q^m}^{(s)} \right) a_{k^m}^{(s)}$$

$$\sigma_{k^m}^{(s)} = \sum_{g}^{M^{m+1}} \sum_{i}^{N^{m+1}} \delta_{i^{m+1}}^{(s)} \left(w_{i^{m+1}k^m}^{(gs)} - a_{k^m}^{(s)} \right)$$

となる。