

変動吸収特性核を用いた 手書き平仮名文字認識

HANDPRINTED "HIRAGANA" CHARACTERS RECOGNITION BY A PATTERN MATCHING METHOD
USING VARIATION ABSORBABLE AUTOCORRELATION MATRICES

村瀬 洋[†] 木村 文隆[†] 吉村 ミツ[†] 三宅 康二^{††}

Hiroshi MURASE Fumitaka KIMURA Mitsu YOSHIMURA Ysuji MIYAKE

[†] 名古屋大学工学部 ^{††} 三重大学工学部

Faculty of Engineering, Nagoya University Fac. of Eng., Mie Univ.

This paper describes a method and experiments of machine recognition of handprinted Japanese alphabetic characters called HIRAGANA by a "soft pattern matching". Specially at learning step, we made an autocorrelation matrix from preprocessed character vectors for each character category and gave this matrix some modification representing typical mathematical deformations of characters. As a result, we obtained a high recognition rate of 98.1% and the effect of the modification of the matrix was sustained even when the number of learning sample was decreased. In addition to this result, this paper refers to an experiment of a two stage decision and a dimension decreasing by KL expansion.

1 まえがき

筆者らは、これまで認識アルゴリズムが単純で並列処理による高速認識装置の可能性を持つパターン整合法を用いて手書き文字の認識実験を行ってきた。複合類似度法を代表とするこれらの手法は、印刷文字に対しては優れた性能を示すが、前報²⁾で述べたように手書き文字に対しても比較的良好な認識率を与える。しかし変形の大きい手書き文字の認識にはまだ多くの問題点が残されている。本報告も前報と同様にカタカナに比較し認識が困難とされている手書き平仮名文字を対象としている。

前報²⁾であげた問題点の一つは未学習文字の認識率が学習文字のそれと比べてかなり低いことである。その原因としては学習文字数の不足を考へることもできるが、未知パターンからのわずかな位置ずれで距離値が大きくなるという本手法の特徴もその理由に考へられ、少数の学習サンプルで未知パターンを類推する人間のすぐれた認識能力に欠けているというこ

ともできる。

その対策には文字の位置、大きさの正規化に加え、部分的ゆがみ、ストロークの凹凸等のより丁寧な修正が考へられるが、後者については処理が簡単でない。そこで今回は学習文字集合の特徴を記述している特性核に部分的ゆがみに対応する雑音を加えて、手書き文字の部分間の位置関係の変動を吸収し、より柔軟な整合を試みたところ、98%以上の認識率が得られたので報告する。

更に、学習文字数を減少させたときの本手法の効果の確認に加え、類似文字の混同状況とその対策の一部および計算量に直接関係する特徴次元数をKL変換により低下したときの効果についても述べる。

データは電総研常用手書き平仮名データベースのうち、学習文字として9600文字を、認識文字として4600文字を使用した。

2 学習および認識システム概要

処理の手順を図1に示す。

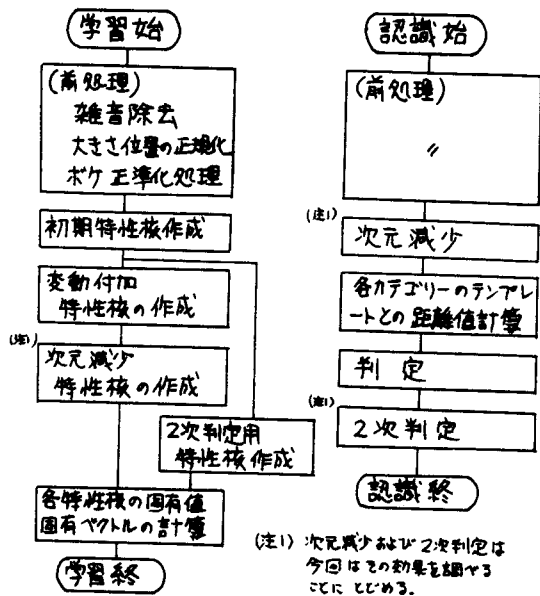


図1 システム概要

3 実験に使用した距離値

3.1 主認識部における距離値

カテゴリ- l のサンプルパターンベクトル (m 次元) を $g_j^{(l)}$ ($j=1, \dots, m$) とすると, 特性核 $K^{(l)}$ は式(1)によって表される。

$$K^{(l)} = \sum_{j=1}^m g_j^{(l)} g_j^{(l)T} \quad (1)$$

$K^{(l)}$ の固有ベクトルを固有値の大きさ順に長短とって得られる関数系 $\varphi_k^{(l)}$ ($k=1, \dots, L$) ($1 \leq k \leq m$) は KL 展開の結果が示すように, そのカテゴリ-に属するパターン-の展開時の誤差 (平均2乗誤差) を最小とする関数系である。ここでは入力パターン-とカテゴリ- l との距離 $d^{(l)}(f)$ をその関数系で展開したときの近似程度 (2乗誤差) により定義した (式(2))。なお結果は飯島の複合類似度と順序関係を逆にしたものと同等である。

$$d^{(l)}(f) = f^T f - [Bf]^T [Bf] \quad (2)$$

$$\text{但し } B = [\varphi_1 \varphi_2 \dots \varphi_L]^T \quad (3)$$

3.2 2次判定部での距離値

ここでは2次判定部として2クラスの場合だけ考える。2クラスの判定には KL 展開の拡張手法から得られる最適な関数系が存在する³⁾。そこでこの性質を利用し各のクラス1及び2との

距離を式(4)及び(5)で定義する。

$$d_2^{(1)}(f) = \sum_{i=1}^2 [\{f^T [K^{-1}]^T e_i\}^2 - \lambda_i^{(1)}]^2 \quad (4)$$

$$d_2^{(2)}(f) = \sum_{i=1}^2 [\{f^T [K^{-1}]^T e_i\}^2 - 1 + \lambda_i^{(2)}]^2 \quad (5)$$

但し e_i は式(7)の i 番目の固有ベクトルであり, その固有値を $\lambda_i^{(1)}$ とする ($\lambda_1^{(1)} > \lambda_2^{(1)} \rightarrow \dots > \lambda_m^{(1)}$)

$$K_2 = K^{(1)} + K^{(2)} \quad (6)$$

$$S_1 = K_2^{-\frac{1}{2}} K^{(1)} K_2^{\frac{1}{2}} \quad (7)$$

但し $K^{(1)}, K^{(2)}$ はそれぞれ別のクラスの特性核である。なお特性核作成時のベクトルを $g_j^{(1)}, g_j^{(2)}$ とすると式(8)が成立することが分っている。

$$\left. \begin{aligned} \lambda_i^{(1)} &= \sum_j [g_j^{(1)T} [K^{-1}]^T e_i]^2 \\ 1 - \lambda_i^{(2)} &= \sum_j [g_j^{(2)T} [K^{-1}]^T e_i]^2 \end{aligned} \right\} \quad (8)$$

4 文字変動

4.1 分類と方針

手書き文字は位置, 大きさ, ストロークの方向, 右より・右下り等印刷文字に見られない多様な変動を持っている。その各変動要因は筆記条件, 筆記者の癖等で説明されるが, 文字認識においてはその変動を極力吸収しなければならぬ。ここでは認識に有利な変動として次の3種のパラメータを考えた。

(1) 各文字自身の持つパラメータ

(例) 文字の大きさ, 位置。

(2) カテゴリ-集合の持つパラメータ

(例) 平仮名"す"の下端付近のストロークの方向 (図4-A参照)。

(3) 文字集合の持つパラメータ

(例) 手書き平仮名文字集合の左上方の複座値の分散 (図4-B参照)。

合理的な変動吸収の手段は, (1) は各文字一様に施す前処理により, (2) は下降式の正規化により, (3) は判定時に考慮することにより実現できる。今回は, (1) については位置および大きさの正規化を行った。又(2)(3)については学習テンプレート側すなわち特性核に正規化変形の逆変形を加えることがある程度の解決となる。

つまり入力パターン-に $m^{(2)}$ 種類の正規化変換 $N_k^{(2)}$ を考慮して第 l カテゴリ-のテンプレ-

トベクトル $g^{(i)}$ との内積値を計算する際、そのつど入力パターンに対し $m^{(i)}$ 回の変換操作 [式(9)] が必要となる。しかしこの計算は式(10)の

$$f_i^{(i)*} = N_i^{(i)} f \quad (i=1, 2, \dots, m^{(i)}) \quad (9)$$

式変形により、あらかじめ用意した式(11)の $g_i^{(i)*}$ と入力パターンとの内積値を求めることと同

$$(f_i^{(i)*}, g_i^{(i)*}) = f^t N_i^{(i)t} g_i^{(i)} = (f, g_i^{(i)*}) \quad (10)$$

$$g_i^{(i)*} = N_i^{(i)t} g_i^{(i)} \quad (i=1, 2, \dots, m^{(i)}) \quad (11)$$

じである。これより入力パターンに下降式に行う必要のある $\sum_{i=1}^m m^{(i)}$ 回の正規化は完全に学習サンプルセットに含まれることができ、認識処理の単純化にもつながる。

これは従来行われた整合法において位置による重みづけ整合法の際にテンプレートだけを重みづけた場合と等価である事実、およびぼかして整合する際一方だけをより大きくぼかした場合に等しい事実と同質のことである。

4.2 変動の例

整合の際の問題の一つは、微小な位置ずれ変形で距離値が極端に大きくなることである。すなわち位置に対する感度が極めて高いことである。この問題は複合類似度のような多次元での整合によりある程度改善されるが、学習段階ですべての変形パターンを用意しておかないと前報のように学習サンプルと未学習サンプルで認識率がかなり異なることになる。解決策としては位置感度を低下させる処理が有効と思われる。ぼけを増大させること、部分パターンみずみの雑音を加えること等を試みた。

図2の十字形で部分パターンの位置ずれとし x を変化させ、 $x=21$ の標準パターンとの距離値を図3に示す。5章で示すみずみ雑音を加える、あるいはぼけを増大することにより位置感度が低下することが分かる。

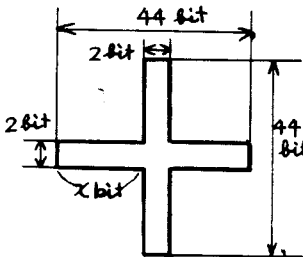


図2 テストパターン

る、あるいはぼけを増大することにより位置感度が低下することが分かる。

又、文字パターンの各メッシュの濃度についてのカテゴリ内分散 y_{ji}

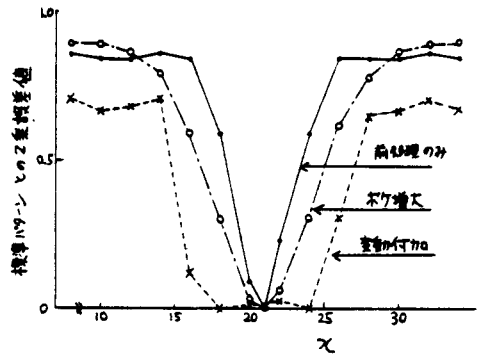


図3 部分要素位置ずれ感度曲線

級内分散 y_1 、級間分散 y_2 、F比 y_3 は式(12)~(15)で与えられる。但し $x_j^{(i)}$ はカテゴリ j の第 i 番目のサンプルの第 i メッシュの濃度値である。

$$y_{ji} = \frac{y_j}{m_j} \sum_{i=1}^{m_j} (x_{ij}^{(i)} - \bar{x}_i^{(j)})^2 \quad (12) \quad y_{2i} = \frac{y_2}{m_2} \sum_{i=1}^{m_2} (\bar{x}_i^{(j)} - \bar{x}_i)^2 \quad (14)$$

$$y_{1i} = \frac{y_1}{m_1} \sum_{i=1}^{m_1} (x_{ij}^{(i)} - \bar{x}_i^{(j)})^2 \quad (13) \quad y_{3i} = y_{2i} / y_{1i} \quad (15)$$

図4に実例を示す。特にDに示されるF比は特徴量選択の1つの手段として用いられる量で文字の周辺部が文字の分離に特に重要であることを示している。これは印刷漢字認識における4辺コード採用とも一致する結果である。これはクラス内では全体に変動しているがクラス間では特に周辺部でよく変動しているためである。

A:分散 y_1 (%)	B:級内分散 y_2 (%)	C:級間分散 y_3 (%)	D: F比 y_3

図4 パターンの各点での分散等

5 変動吸収を考慮した特性核と次元減少

5.1.1 変動吸収を考慮した特性核

カテゴリ j の文中ベクトル $g_j^{(i)}$ ($j=1, \dots, m$) により得られる特性核 $K^{(j)}$ [式(16)] を初期特性核と呼ぶことにする。

$$K^{(j)} = \sum_{i=1}^m g_j^{(i)} g_j^{(i)t} \quad (16)$$

この $K^{(j)}$ を用いて変形を含んだ変動付加特性核 $K^{(j)*}$ を作成する。ここでカテゴリ j に加え

るし番目の変形マトリクスを $N_i^{(i)}$ ($i=1, \dots, m^{(i)}$) とすると変形を加えた文字集合より作成される特性核は式(19)で与えられる。

$$K^{*(i)} = \sum_{j=1}^{m^{(i)}} \sum_{l=1}^{m^{(i)}} N_i^{(i)} g_j^{(i)} [N_i^{(i)} g_l^{(i)}]^t \quad (17)$$

$$= \sum_{j=1}^{m^{(i)}} N_i^{(i)} \left\{ \sum_{l=1}^{m^{(i)}} g_j^{(i)} g_l^{(i)t} \right\} N_i^{(i)t} \quad (18)$$

$$= \sum_{j=1}^{m^{(i)}} N_i^{(i)} K^{(i)} N_i^{(i)t} \quad (19)$$

この式(17)~(19)の変型により1つのカテゴリ-集合に同様の変型を加える操作が容易になり多種の変形を加えることが可能となる。今回は次に示す各種変動を全カテゴリ-共通に加えた(図5)。

(1) 位置変動(9種)

文字の突起部分の長さ等前処理における位置の正規化で吸収されなかった変動を吸収する目的で上下左右斜め方向へ確率密度が正規分布になるように与えた。

(2) 傾き変動(4種)

右上り右下り等の変形

(3) 中央部の変動(4種)

前処理で行われるわく合せの正規化だけでは中央部の位置合せ精度が悪くなるため、この変動を与える。又中央部は図4で示されるように縦間変動に比べて縦内変動の大きい部分である。

(4) 辺上の変動(辺上のわん曲)(4種)

わく合せによる大きさ・位置の正規化では吸収しきれない変動の1つである。

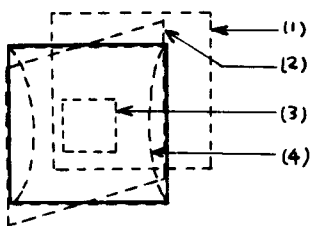


図5 変形例

5.1.2 ぼけの増大

5.1.1と同様の目的で、ぼけを増大した特性核として式(20)を用いた。

$$K^{**i} = \left(\sum_{j=1}^{m^{(i)}} N_i^{(i)} \right) K^{(i)} \left(\sum_{l=1}^{m^{(i)}} N_i^{(i)} \right)^t \quad (20)$$

5.2 K L展開を用いた次元減少⁵⁾

2次元画像では分解能を上げようとしてメッシュ数を細くすると次元数が膨大になり計算が

困難となる。そこで今回用いている整合法と密接に関連しているKL展開を用いた次元減少の効果を図べる。

式(21)の行列Mにより与えられる大きな固有値n個に対する固有ベクトルを縦ベクトル成分とする行列Tにより式(22)で次元を減少させ、文字パターンfからベクトルyを作る。又このときの次元減少特性核 $K^{(i)}$ は式(23)で与えられる。すなわち、ここで用いられる方式は入力

$$M = \sum_{i=1}^k K^{*(i)} \quad (21)$$

$$y = T^t f \quad (22)$$

$$K^{(i)} = T^t K^{*(i)} T \quad (23)$$

ターンとして $T^t f$ を、カテゴリ-別の標準パターン集合として $T N_j^{(i)} g_j^{(i)}$ ($i=1, \dots, m; j=1, \dots, m^{(i)}$) を用いたときのパターン整合法に等しい。

7 実験結果および検討

7.1 位置感度低下の効果

文字ベクトルfとして210次元を選択し、全次元を使用し認識実験を行った。表1に初期特性核を使用した場合、ぼけを増大させた場合、変動付加特性核を使用した場合の認識率を示す。

表1 認識率

	学習サンプル(2300)	非学習サンプル(400)
(A) 初期特性核使用	100.0%	97.3%
(B) ぼけを増大	99.9%	96.9%
(C) 変動付加特性核使用	99.5%	98.1%

今回の手法(C)を用いたことにより、学習サンプルについてはやや認識率は下が、たものの、未学習サンプルについては向上し、認識率は98.1%に達した。又学習サンプルで饒めなくな、た文字も7.8で示すように極度に汚ない文字ばかりである。(A)での認識原因は入力文字が他カテゴリ-文字に似ている場合、変形がやや大きい場合が考えられるが、手法(C)で改善されたのは後者である。ぼけを増大させたときに認識率が低下したのは位置感度の低下に伴って類似カテゴリ-との混同が増大したためである。

そこで以下の実験は手法(C)で行った。

7.2 文字品質との関係

電通研常用平仮名データベースでは、文字シ

トごとに3段階(0,1,2)の文字品質記号が与えられている。品質別に本手法を用いたときの認識率を表2に示す。低品質文字についてはまだ問題を残しているが、中程度以上の文字については高い正確率が得られることが分かる。

表2 文字品質に関する認識率

文字品質	0	1	2
認識率	99.0%	98.2%	92.4%

7.3 候補カテゴリ-抽出能力

整合法は、従来候補をしぼる、すなわち大分類をするのに有効とされてきた。そこで何位までに正しい候補が選ばれるかを調べたところ図6の示す結果を得た。

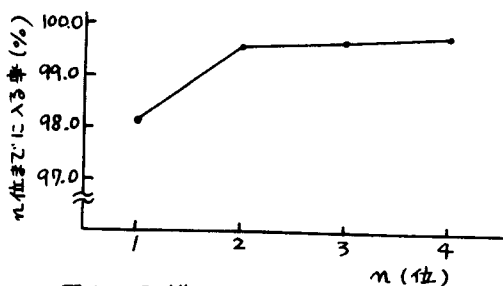


図6 候補カテゴリ-抽出能力

図6から分かるように、2位までに正解が99.6%含まれるというすぐれた能力が確認できた。後段で候補間を詳細に分離する処理が可能となれば、更に高い認識率が可能となる。

7.4 学習数と認識率の関係

本手法の目的の一つに、学習数を低下させても認識率をなるべく下らないようにすることを目指したが、ここではその効果を前報の手法と比較して図7に示す。

初期特性核を使用したときに比較し、学習数が低下するに伴って顕著な効果が表われないことがわかる。

以下に、今後の展望を考え、次元減少および2次判定等の効果の例を示す。

7.5 KL展開による次元数低下の実験

認識段階の計算量は次元のほぼ2乗に比例する。7.4までは全ての次元を用いて実験を行ってきたが、ここでKL展開で低下したときの効果を調べる。図8に次元数に対する認識率、

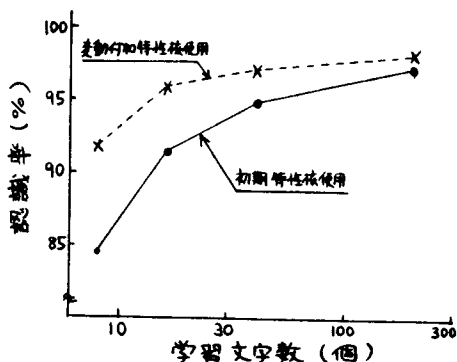


図7 学習文字数と認識率の関係

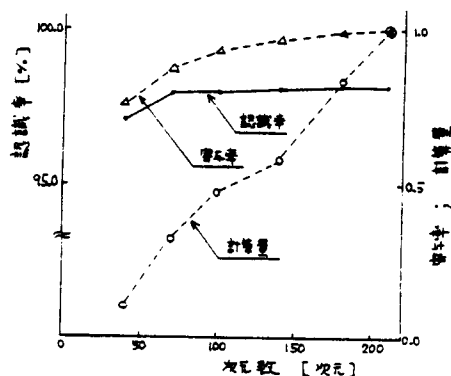


図8 次元数に対する認識率、精度および計算量の変化

率および大よその計算量を示す。但し式(3)中の丸の値はそのつど最適なものを選んだ。次元数を半分程度低下させても認識率の変化はほとんどなくKL展開の有効性が示された。

7.6 類似カテゴリ-2分木

7.5までは、全てのカテゴリ-の認識を問題にしてきたが、実際は認識しやすいカテゴリ-や、混同しやすいカテゴリ-関係が存在する。カテゴリ-間の類似性を式(25)で定義し、階層的クラスタ分析の手法を用いて、全てのカテゴリ-からなる2分木を作成したところ、混同文字関係が実験結果と比較的一致していた。この2分木は、2次判定の目安に使用できる。得られた2分木リストの一部を図10に示す。

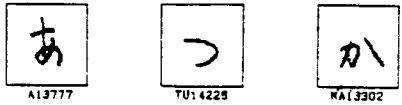
$$S(l, l') = \sum_i \sum_j [g_i^{(l)T} g_j^{(l')}]^2 \quad (24)$$

$$\doteq \sum_{i=1}^{10} \sum_{j=1}^{10} \lambda_i^{(l)} \lambda_j^{(l')} [\varphi_i^{(l)T} \varphi_j^{(l')}]^2 \quad (25)$$

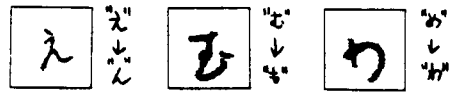
但し、カテゴリ- l の特性核の固有値 $\lambda_i^{(l)}$ ($\lambda_1^{(l)} < \dots < \lambda_n^{(l)}$)に対する固有ベクトルを $\varphi_i^{(l)}$ とする。

ノードリスト (カタゴト) : ノードに属すカタゴト

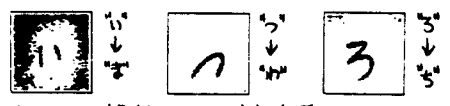
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	



(a) 本方式により正読した文字



(b) 誤読した文字



(c) まだに誤読した文字 (学習パターン)

図12 文字例

8 むすび

特性板に変動を加えることにより、学習サンプルと少しずれている文字等が認識できるようになり、その結果手書き平仮名文字4600文字に於して98.1%の認識率が得られた。今回は特性板すべてに同様の変動を加えたが、カテゴリごとにその文字に適した変動を收拾選択することにより、更に改善できるものと思われる。

少数学習サンプルでの有効性は、学習サンプル数を多く集めることが困難な場合大きな意味を持つ。例えば個性的な文字を直前に登録したのち読み取らせるなどの認識装置の実現も可能となる。

今後の課題としては、K次元減少法での効果の確認を基にして、入力段階でメッシュ数を増大させ、より分解能を高めた時の効果を調べることで、及び詳細な2次判定手法の確立等がある。

最後に、常用手書き平仮名データベース使用に際しお世話になった電子技術総合研究所森俊二室長はじめ関係諸氏、日頃御指導を賜る本学工学部本多竣雄教授、ならびに研究室の皆様へ感謝いたします。

文献

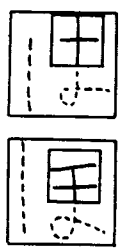
- 1) 飯島ほか：日経エレクトロニクス 1972-5-22 P66
- 2) 村瀬ほか：信学会 研資 PRL 79-2
- 3) T.Pavlidis：“Structural Pattern Recognition”
- 4) 森ほか：日経エレクトロニクス 1979-10-31 P102
- 5) 村瀬ほか：東海支部連大 319
- 6) 奥野ほか：“多変量解析法” 日科技連

⑩ノードに属すカタゴト一組でよく境用をおこす
〇 混同をおこしやすい

図10 カテゴリ間の類似性を考慮した2分木リストの一部

7.7 2次判定

図10に示されるように、特定の2カテゴリ間の分離が極めて悪い(“は”↔“ほ”、“き”↔“せ”等)。従ってそれらが候補となったとき、図10のそれを含む適当なノードに再入力させ、2次判定させることが有効である。今回は、“は”↔“ほ”の場合だけについて、式(8)の距離値により部分パターン重畳を用いて判定実験を試みた。図11にその手順を、表3に2次判定を行った場合の認識率を示す。



(1)手順1
特定の大きさの窓を用意し初期位置を右上にとり、窓内の重心が中央にくるよう位置調整を数回行う。

(2)手順2
窓内のパターンを特徴と考え、2.2で示した距離値を用いて判定する。

図11 手順(“は”と“ほ”の判別)

表3 “は”と“ほ”の認識率(100文字)

2次判定をしないとき	94.0%
2次判定をしたとき	97.0%

7.8 誤読文字例

図12に示す。