

修士学位論文

文字認識アルゴリズムの
高精度化に関する研究

東北大学大学院工学研究科
情報工学専攻

越後 和 徳

目次

1	序論	1
1.1	本研究の背景	1
1.1.1	パターン認識	1
1.1.2	文字認識	1
1.2	本研究の目的	2
1.3	本論文の構成	4
2	特徴量空間分析	5
2.1	前書き	5
2.2	文字認識の方法	5
2.3	パターン整合法の定式化	5
2.3.1	文字入力	6
2.3.2	前処理	6
2.3.2.1	ノイズ除去・スムージング	6
2.3.2.2	正規化	6
2.3.2.3	細線化	7
2.3.2.4	線素化	7
2.3.3	特徴抽出	7
2.3.3.1	方向線素特徴量	7
2.3.3.2	方向密度特徴量	8
2.3.3.3	メッシュ特徴量	8
2.3.4	認識	8
2.3.4.1	辞書作成	8
2.3.4.2	評価値計算	9
2.3.5	候補出力	9
2.4	特徴量に関する一般的性質	17
2.5	後書き	24
3	特徴抽出法の性能評価	25
3.1	前書き	25
3.2	実験の目的	25
3.3	実験の対象	26

3.3.1	サンプル集合	27
3.4	実験結果と考察	29
3.5	パターン整合法への応用	35
3.6	後書き	35
4	パターン整合法への適用	36
4.1	前書き	36
4.2	実験の目的	36
4.3	実験方法	37
4.4	特徴領域整合法	37
4.5	実験結果	39
4.6	考察	40
4.7	後書き	42
5	標本領域半径の改善	43
5.1	前書き	43
5.2	実験の目的	43
5.3	実験の準備	44
5.3.1	標本領域半径	44
5.4	実験の方法	46
5.5	実験結果と考察	46
5.6	後書き	50
6	多特徴領域によるパターン整合法	52
6.1	前書き	52
6.2	実験の目的	52
6.3	実験の方法	53
6.3.1	K -means 法	53
6.4	テンプレート数 2 による実験結果と考察	54
6.5	テンプレート数可変による実験結果と考察	56
6.6	後書き	57
7	N 次元 2^N 面体型特徴領域によるパターン整合法	58
7.1	前書き	58
7.2	実験の目的	58
7.3	実験の方法	59
7.4	N 次元空間における多面体と超楕円	60
7.5	実験結果と考察	61
7.6	認識結果の正しさの保証について	61
7.7	後書き	63

8 結論	64
8.1 前書き	64
8.2 本研究の成果	64
8.3 今後の課題	69
8.4 後書き	69
 謝辞	 70
 参考文献	 71
 公表目録	 72

目次

1.1	現状の文字認識システムにおける正しい認識内容を得るまでの課程	3
2.1	文字認識アルゴリズム	10
2.2	前処理後のイメージ	11
2.3	ノイズ除去・スムージングのためのマスク	12
2.4	方向線素特徴量	13
2.5	方向線素特徴量の抽出例	14
2.6	方向密度特徴量	15
2.7	メッシュ特徴量	16
2.8	領域半径の例	20
2.9	誤認識文字集合の例	21
2.10	安全認識半径の例	22
2.11	評価値の最小値が必ずしも正解を与えない例	23
3.1	標本領域半径の例	27
3.2	誤認識文字集合の大きさの度数分布	34
4.1	特徴領域整合法	38
4.2	特徴領域整合法による認識率低下の原因	41
5.1	標本領域半径の改善	45
5.2	領域半径標準測度と認識率、誤認識文字数の関係	48
5.3	領域半径標準測度と判断不能文字数、計算対象文字数及び誤認識文字集合の 大きさの関係	49
5.4	特徴領域の変更前と変更後の例	51
6.1	シングルテンプレートの領域半径とテンプレート数2による誤認識文字集合 の大きさの変化の関係	55
8.1	本研究における認識率と誤認識文字数の推移	66
8.2	本研究における判断不能文字数と誤認識文字集合の大きさの推移	67
8.3	本研究における計算対象文字数の推移	68

表 目 次

3.1	実験用文字データ (1)	28
3.2	領域半径	31
3.3	安全認識半径	31
3.4	領域半径と安全認識半径の関係	32
3.5	安全認識領域に属するパターン数とその割合	32
3.6	誤認識文字集合の大きさ及び平均+標準偏差までの文字数	33
3.7	誤認識文字パターン数と特徴領域外となった文字パターン数及び認識率	33
4.1	実験用文字データ (2)	37
4.2	特徴領域整合法によるクローズ実験の結果	39
4.3	特徴領域整合法によるオープン実験の結果	40
6.1	多特徴領域によるパターン整合法の結果	56
7.1	N 次元空間内の多面体	61
7.2	N 次元 2^N 面体型特徴領域によるパターン整合法の結果	61
7.3	命題 2 による認識結果の正しさの保証についての文字数とその率	62

第 1 章

序論

1.1 本研究の背景

1.1.1 パターン認識

人間同士が情報の交換や記録に用いている情報の媒体(メディア)には、文字、音声、図形、画像などがある。人間の文化が伝承され、学問が発達し、日常生活が滞りなく営まれていくには、これらの情報媒体が大きな役割を果たしている。人間社会に計算機などの情報処理機械が数多く利用され、機械と人間とが共生する社会になってくるにつれて、人間が用いている情報媒体と、機械が用いているコード化されたデジタル情報との間を、相互に交換する必要が生じてきた。

その中でも、機械の方から人間に理解できるように情報を出力する技術は、文字を印刷したり、ディスプレイ装置に表示したりする技術、音声を合成する技術、計算機の中に存在する情報を人間にとって認識しやすい形にグラフ化したり、図形、透視図、動画などで表示するコンピューターグラフィック技術、更に、普通では直接肉眼では見ることのできない脳の断層画像を再構成して表示する技術などが発達しており、それぞれ大きな応用分野をもっている。

一方、人間が用いている情報の媒体を機械のコード情報に変換する技術は、パターン認識の技術と呼ばれており、情報科学の分野で最も困難な基本的課題の一つであると考えられている。人間は本能や幼時からの長年の学習によって文字、音声や画像などを認識する能力は高度に発達している。しかしながら、人間が脳の中でどのように認識を行っているのかについては、生理学的にも心理学的にも解明されていないので、機械にパターン認識させるには情報科学的な独自のモデルを作り、実験的に一步一步技術を発達させて行かなければならない。

1.1.2 文字認識

パターン認識は文字を読み音を聞き分けるといった人間の能力を機械で実現するための要求から出発し、これまで多くの研究が行われてきた。とりわけ文字認識は最も歴史が古く、かつ実用化が進んだ分野である。文字は身近な存在であり、記録性に優れ、再現性が

良く、人間の直感を生かしやすく、1文字に一つの概念が対応するなど、パターン認識研究の素材としては格好の性質を備えている。また、文字を読む技術は、コンピューターへのデータ入力をキーボードを経ずに直接行う省力装置の開発に直結することから、強い社会的ニーズに支えられてこれまで発展してきたともいえる。世界的に見て文字認識及び文字読取装置の研究・実用化が進んでいるのは日本と米国である。タイプライターの普及している米国では、読取り文字種が活字の英数字記号に限られることもあって、文字認識そのものの研究は現在では一段落している感がある。ところが、我が国では数千種に及ぶ漢字を常用しており、そのうえタイプよりも手書きの比率が多い。米国に比べて比較にならぬほど複雑で高度である。このため過去30年にわたる技術開発にもかかわらず、字種や字体に関する制約を完全に取り除ける段階にはない。つい最近までの光学式文字読取装置(Optical Character Reader, 略してOCR)は、読取り文字種を英数字片仮名などに制限し、装置を使う人が守れる最大限の規約(人間と機械間の妥協点…用紙、記入枠、字体、印字様式等)を定め、これを拠り所にして大量のデータを取り扱う業種に主として用いられているというのが実状である。

1.2 本研究の目的

今、社会の情報化は、文字を介してコンピューターシステムへ直接アクセスする用例、例えば金融機関や官公庁での窓口処理、更には膨大な量の日本語データをOCRで読取り、データベースを構築する処理(特許、文献データベース等)の必要性を生み出している。将来倍増するデータベースに対して、漢字認識の社会的要請はますます高まっている。しかし現状では、図1.1のように認識結果に対し正しい認識内容を得るためには対話的な修正が多く必要とされ、漢字認識の完全な自動化には到っていない。実際には点字翻訳する場合など、誤認識の修正に手間がかかり過ぎるために多くの労力が必要とされる。将来的には大量の文章を読みとらせる場合、完全な自動化(図1.1において「修正」を除去すること)は不可欠であり、労力の省力化に大変期待が寄せられている。

完全な自動化を達成するには正確な認識が必要である。即ち、辞書にない文字や変形の著しい文字はそのことを指示し、それ以外の正しいと認識した文字については誤りがあってはならないことである。従来は、認識アルゴリズムの良さを認識率で評価した。しかし、認識率では上に述べたような個々の文字の正確さを判定できない。即ち、ある文字に類似しているとは判定するが、その判定の正しさは確率論的に保証され、決定論的には保証されない。ある認識アルゴリズムを用いて得られた結果は最終的には人間がその正しさを判定している。

本研究では、正確な認識を達成することを究極の目的とし、パターン整合法の認識アルゴリズムにより、識別可能なこととは何か、認識結果の正しさの保証をどう与えるかについて考えたものである。そしてその実験及び検討していくことは、必然的に認識率の向上にも繋がって行く。故に準目的としてその高精度化の達成も念頭に置きながら研究を進めて行く。

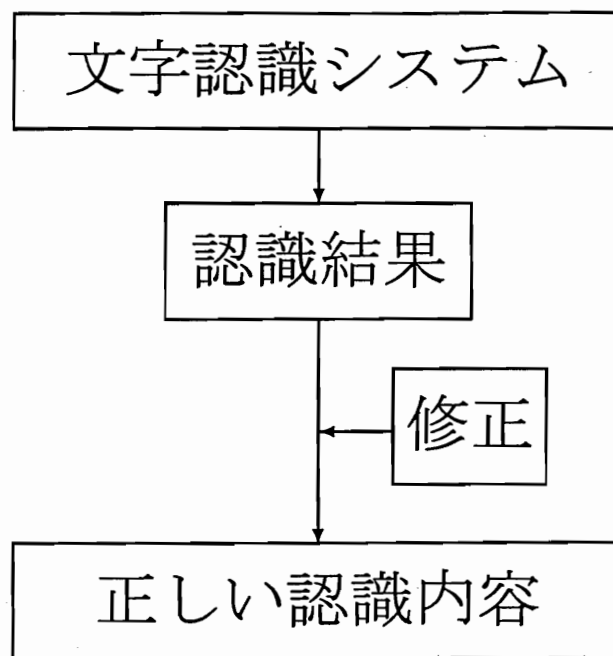


図 1.1: 現状の文字認識システムにおける正しい認識内容を得るまでの課程

1.3 本論文の構成

本論文の構成は次の通りである。

- 第 1 章 序論であり、本研究の背景と目的を述べる。
- 第 2 章 パターン整合法の理論的分析について述べる。併せて認識結果の正しさの保証をどう与えるかについて命題として提案する。
- 第 3 章 第 2 章の理論的結果に対し具体的な特徴抽出法の実験によって、現実的な妥当性について調べ考察する。
- 第 4 章 パターン整合法の理論的な分析により考えられた特徴領域整合法を提案し、全数整合法と比較実験をする。これは、各文字に領域半径、誤認識文字集合の情報を与え、利用するものである。
- 第 5 章 第 4 章の第 1 改良案として、標本領域半径の大きさを可変することによる真の領域半径の選出について、パターン整合法を行いながら検討する。
- 第 6 章 第 4 章の第 2 改良案として、マルチテンプレート法によるパターン整合法に特徴領域整合法を適用する。即ち、多特徴領域の効果について述べる。
- 第 7 章 第 4 章の第 3 改良案として、特徴領域を、サンプルパターンの特徴量の各成分の標準偏差により変形 (N 次元 2^N 面体型) することによる効果について述べる。
- 第 8 章 結論であり、本研究の成果と今後の課題について述べる。

第 2 章

特徴量空間分析

2.1 前書き

本章ではまず文字認識の手法について述べ、パターン整合法を定式化するとともに平行して処理課程も説明する。その後、距離尺度により定められる特徴量空間の性質を分析する。

2.2 文字認識の方法

文字認識の方法は大きく分けて次の 2 つがある。

- パターン整合法
- 構造解析法

前者のパターン整合法は、パターン同士の重なり具合から類似度を基に認識を行う。類似文字の識別は難しいが文字の変形やノイズに強く、また、計算機上で簡単に実現できる。後者の構造解析法は、線分の接続関係や位置関係などの文字構造に着目し構造の類似性で認識を行う。類似文字の多い漢字の認識などでは有効であるが、計算機に線分を正しく抽出させ、線分関係を定義させるのは大変難しい。認識の高精度化が期待できる反面、処理量が膨大で実用化にはかなり問題がある。

活字認識においては、実用上特に高速性が要求されるため計算量の少ないパターン整合法を用いたほうが有利である。本研究では、活字文字を扱うためパターン整合法を用いて研究を進めることにする。

2.3 パターン整合法の定式化

文字パターンは認識対象文字のイメージ (2 次元画像) である。イメージのサイズを $m \times m$ ドットとして、 $M = \{1, 2, \dots, m\}$ とおき、イメージの各ドットの明るさの値の集合を B とおく。イメージ P は、写像 $P : M \times M \rightarrow B$ として定義できる。この写像 P の集合を \mathcal{P} とおく。

認識対象の全ての文字の集合を K とおく。各文字 $k \in K$ に対し、その文字パターンは多数ある。その集合を $\mathcal{K}(k)$ とおく。即ち、 $\mathcal{K}(k) = \{P_i^k | P_i^k \text{ は } i \text{ 番目のパターン}\} \subset \mathcal{P}$ である。また $\mathcal{K} = \bigcup_{k \in K} \mathcal{K}(k)$ とおく。 $\mathcal{K} \subset \mathcal{P}$ である。 \mathcal{K} の要素を文字パターンと呼ぶ。 $\mathcal{K}(k)$ は原理的には有効であるが、その要素数が多いため現実的にはその要素全てをリストすることはできず、明示的に定めることにも成功していないと言える。

文字パターン P に対して、 $P \in \mathcal{K}(k)$ となる k を一意に定めることが文字認識である。この一意性が成立するためには、各 $k, j \in K$, $k \neq j$ について、

$$\mathcal{K}(k) \cap \mathcal{K}(j) = \phi \quad (2.1)$$

が成立している必要がある。これは例えば平仮名の「へ」と片仮名の「へ」を同一視することを意味する。

本研究で使用するパターン整合法のシステムは図 2.1 のような処理課程を行う。以降は各課程を説明しながら定式化していくことにする。

2.3.1 文字入力

認識される文字データはイメージスキャナによって入力される。出力されたイメージデータは、切り出しの処理により各文字毎に切り出される。イメージデータを図 2.2(a) に示す。

2.3.2 前処理

切り出された文字データは、特徴量を求めるために次のような前処理が行われる。前処理は、ノイズ除去・スムージング、正規化、細線化、線素化の 4 つの処理によって構成されている。

2.3.2.1 ノイズ除去・スムージング

イメージデータは印刷時のリボンの状態やイメージスキャナなどの性質によって、ノイズや線分上の凹凸が少なからず発生している。これらのものは、認識精度に悪い影響を及ぼすため、まずノイズ除去を行う。本研究では、図 2.3(a) のように 2×2 ドット以下の孤立点をノイズと見なし除去することにする。スムージング (線分の円滑化) は、 3×3 ドットのマスクを用いて行う。図 2.3(b) は、凸となっている 3×3 マスクの 1 番上を除去する例で、図 2.3(c) は、凹となっている中心のドットを埋める例である。ノイズ除去・スムージング後のイメージを図 2.2(b) に示す。

2.3.2.2 正規化

活字文字は一般的に 4 倍角、縦倍角、全角などさまざまである。正規化は、切り出した文字の大きさや位置の違いによる影響を吸収するために、もとのイメージデータを一定の大きさ (本研究では 64×64 ドット) に拡大又は縮小する。正規化後のイメージを図 2.2(c) に示す。

2.3.2.3 細線化

正規化後の画像に対して、文字線幅の違いを吸収するために細線化を行う。細線化は、太さを持った図形を幅 1 の図形に変換する処理である。本研究では Hilditch の方法^[7]を用いる。この方法は、各画素においてその周りの 3×3 の 8 近傍に着目し、8 連結数をもとに図形を 1 ドットずつ削っていく方法である。図形の変化がなくなるまでこれを繰り返す。すなわちパターンの線分の太さの半分の回数を繰り返すことにより、最終的に線幅 1 の連結図形が得られる。細線化後のイメージを図 2.2(d) に示す。

2.3.2.4 線素化

線素化は、細線化された図形の各画素について、黒画素であればその周囲 3×3 の小領域を参照し最も自然な方向と考えられる 4 方向 (縦「|」、横「—」、斜め 45° 「/」、斜め 135° 「\」) のうちの 1 つの線素に対応させる。線素化後のイメージを図 2.2(e) に示す。

2.3.3 特徴抽出

パターン整合法では、文字パターンの持つ冗長度の削減と処理の高速化、パターン分離の効率化などのために、文字パターンを特微量という数値ベクトルに変換する。この課程を特徴抽出という。まず、特微量は N 次元の整数値ベクトルである。特徴抽出は \mathcal{P} から Z^N への写像 F であると定める。但し、 Z は整数の集合で、 Z^N を特微量空間と呼ぶ。各文字 $k \in K$ に関する特微量の集合を $\mathcal{F}(k)$ とおく。即ち、

$$\mathcal{F}(k) = F(\mathcal{K}(k)) \quad (2.2)$$

さらに $\mathcal{F} = \bigcup_{k \in K} \mathcal{F}(k)$ とおく。

この定義から、 $P \in \mathcal{K}(k)$ ならば、 $F(P) \in \mathcal{F}(k)$ である。逆に、文字パターン $P \in \mathcal{K}$ の特微量 $u = F(P)$ が $u \in \mathcal{F}(k)$ を満たすとき、 $P \in \mathcal{K}(k)$ であると決定したいわけである。これが可能であるためには特徴抽出 F において各 $k, j \in K$, $k \neq j$ について、

$$\mathcal{F}(k) \cap \mathcal{F}(j) = \phi \quad (2.3)$$

の条件を満たす必要がある。これは特徴抽出 F が \mathcal{P} からの写像として 1 対 1 であれば満たされるが、ある文字の異なるパターンが同一の特微量を持つこと (多対 1) も可能で、この要請さえ満足されれば、原理的にはある文字の特微量を他の文字の特微量と区別できる。つまり、 $u \in Z^N$ について、 $u \in \mathcal{F}(k)$ のとき、 u は k であるということにする。

以降は本研究で使用する方向線素特微量^[1]、方向密度特微量^[2]、メッシュ特微量^{[3][4]} について説明する。

2.3.3.1 方向線素特微量

方向線素特微量^[1]の抽出法は図 2.4 のように、まず、 64×64 ドットの線素化文字を 8 ドット間隔に縦横を分割する。次に、左上から 16×16 ドットを半分ずつ重複させて 49 領域*がで

*左上を第 1 領域とし、左から右、上から下へ順に領域番号が付けられている。

きる。1領域のベクトルは4方向のカウンタからなる4次元ベクトルでできており、1領域内の重みは図2.4の下図のようになっている。よって1文字当りの次元数は $N = 49 \times 7 = 196$ 次元のベクトルとなる。図2.5に「千」をサンプルにした第4領域の抽出例を示す。

2.3.3.2 方向密度特徴量

方向密度特徴量^[2]は、図2.6のように1文字を 70×70 ドットにする。それを 5×5 の25領域(1領域は 14×14 ドット)に分割する。各々の領域には4つの方向線素に左右上下のカウンタが存在し、計16個のカウンタが存在する。このカウンタは1つの方向線素と位相情報によってカウントされる。

実際にカウントの方法を図2.6を使って説明する。今1つの縦の線素が存在したとする。この線素を中心に上下左右7ドットの位置に、 7×14 ドットのウィンドを図のようにおく。このウィンドは領域のカウンタのカウントを決める。カウントさせる量はウィンドが領域にかかる面積に比例している。つまり、ウィンドは一定のカウント量を持っていて、それをウィンドにかかるいくつかの領域に分配する。例えば、UP WINDOWは2つの領域にかかり、左と右の面積比が1:3程度なので右の領域に大きくカウントされる。

このような操作を全ての方向線素に行い、25領域の各16要素のカウンタに次々とカウントさせる。よって1文字当りの次元数は $N = 25 \times 16 = 400$ 次元のベクトルとなる。

2.3.3.3 メッシュ特徴量

メッシュ特徴量^{[3][4]}は、正規化した文字を使って分割した各領域内の黒画素の割合を計測することにより抽出する。抽出法を図2.7を使って説明する。まず 64×64 ドットの全体の領域を8ドット間隔に分割することにより、 8×8 ドットの1領域が64個できる。次に、各領域で黒画素を計測することにより、 $N = 64$ 次元の特徴量となる。

特徴としては、

- 細線化をしないため、潰れによる影響が比較的少ない。
- 次元数が少ないため、距離計算が速い。

等が挙げられる。

2.3.4 認識

ここでは最も一般的な手法である全数整合法を説明する。全数整合法は、各文字の標準特徴量と特徴抽出で得られた特徴量で評価値を求め、近いものから順に識別する方法である。アルゴリズムの簡潔性やある程度の高い認識率が得られることから、一般的に用いられている。

2.3.4.1 辞書作成

パターン整合法は u が k であることを判定するために、あらかじめ文字 k に対して標準特徴量 $v^k = (v_1^k, \dots, v_n^k, \dots, v_N^k) \in Z^N$ を選んでおく。標準特徴量 $v^k \in Z^N$ の集合を辞書とい

う。辞書作成は一般に $D: K \rightarrow Z^N$ と表せる。辞書は $D = \cup_{k \in K} D(k)$ と記すことにする。辞書作成法としては、

- 文字 k のサンプルパターンの集合 $S(k)$ の重心ベクトル
- 重心ベクトル付近の $F(k)$ の元

等の例がある。本研究では前者の方法を使用した。即ち、集合 S の大きさを $s = |S|$ として、

$$v^k = D(k) = \frac{1}{s} \sum_{i=1}^s F(P_i^k) \quad (2.4)$$

とした。

2.3.4.2 評価値計算

未知入力文字の特徴量 u と各標準特徴量との類似性や整合性の評価は、写像 $E: Z^N \times Z^N \rightarrow Z$ でなされ、 u に対する文字 k に関する評価値を $E(u, v^k)$ と表す。

評価値の例としてよく使用されるものとして

$$E(u, v^k) = \sqrt[d]{\sum_{n=1}^N |u_n - v_n^k|^d} \quad (2.5)$$

において、 $d=1$ のシティブロック距離[†]と $d=2$ のユークリッド距離があげられる。これらは距離の公理を満たし、距離尺度と呼ぶ。本研究ではこれらを主に使用することにする。その他の距離尺度をあげると $d=\infty$ のとき、即ち $E(u, v^k) = \max_{n=1}^N |u_n - v_n^k|$ のチェス盤距離[‡][9][10] や、分布の中心からの共分散行列を考えるマハラノビス距離^{[9][11]}等がある。

その他の評価値としては、2つの特徴量の内積から定まる類似度^[9]が使われている。定義は、

$$E(u, v^k) = \frac{u \cdot v^k}{|u||v^k|} = \cos \theta \quad (2.6)$$

である。

距離尺度は $[0, \infty)$ の値を取り、 $u = v^k$ の時に 0 の値となる。類似度は u と v^k のなす角 θ に等しく、 $[-1, 1]$ の値を取り、 $u = v^k$ の時に 1 の値となる。距離尺度は 0、類似度は 1 に近いほうが類似しているとみなされる。

2.3.5 候補出力

評価値計算をした後評価値の小さい順に、第 1 位候補、第 2 位候補、…、第 N 候補を出力する。一般には、第 c 候補文字出力を写像 $C: Z^N \times \{\text{順位}\} \rightarrow K$ とおくと、次のことが言える。

$$C(u, c) = k \iff E(u, v^k) = \min_{j \in K} (c) E(u, v^j) \quad (2.7)$$

但し、 $\min(c)$ は第 c 最小値を意味する。

[†]市街区距離、4 連結距離とも言う。

[‡]8 連結距離とも言う。

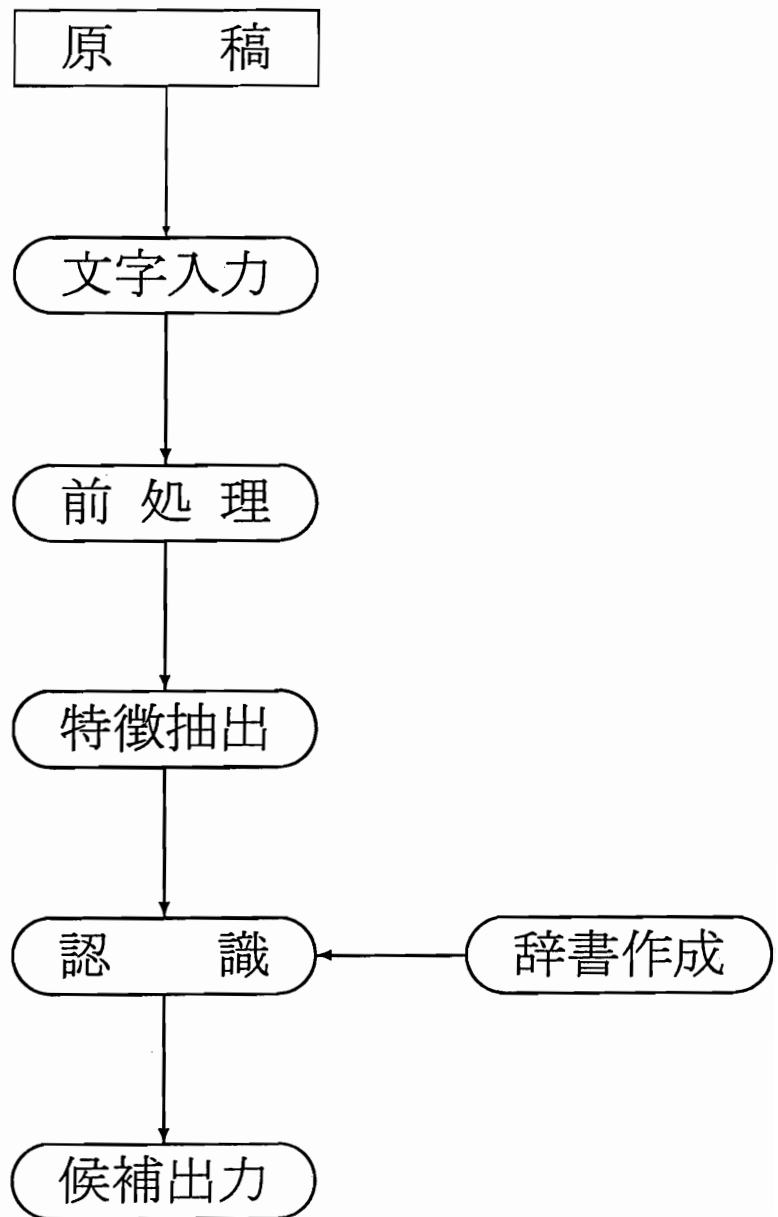
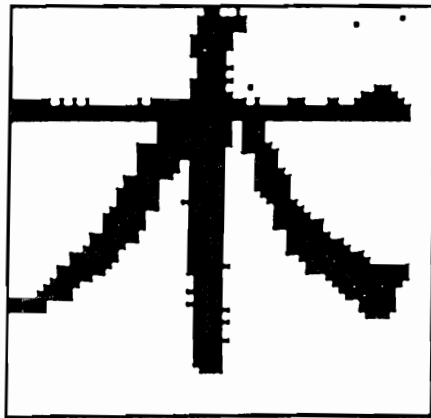
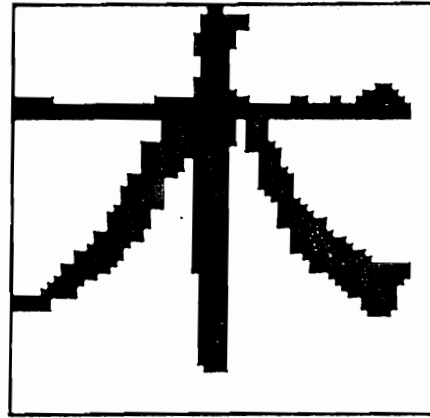


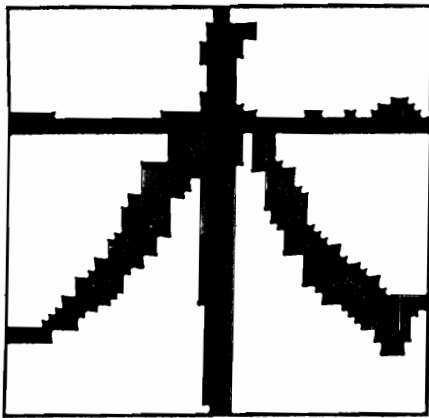
図 2.1: 文字認識アルゴリズム



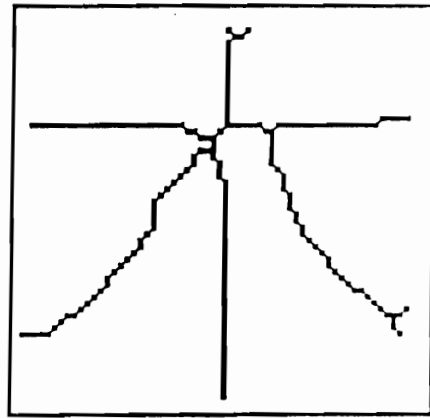
(a) イメージデータ



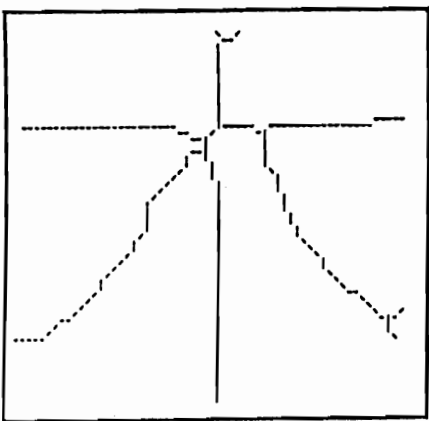
(b) ノイズ除去・スムージング後



(c) 正規化後

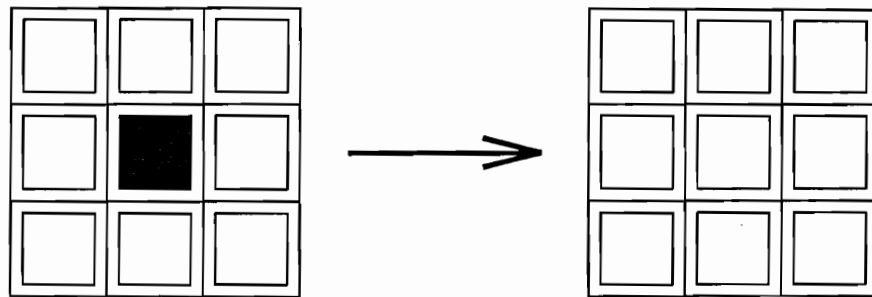


(d) 細線化後

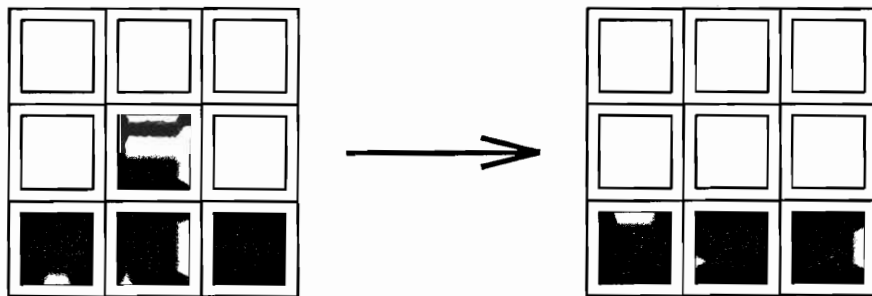


(e) 線素化後

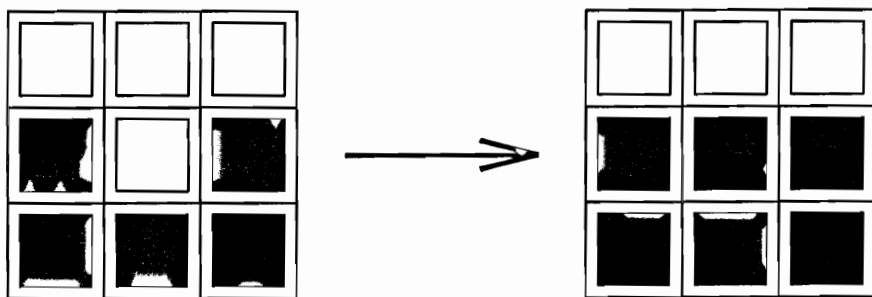
図 2.2: 各処理後のイメージ



(a)

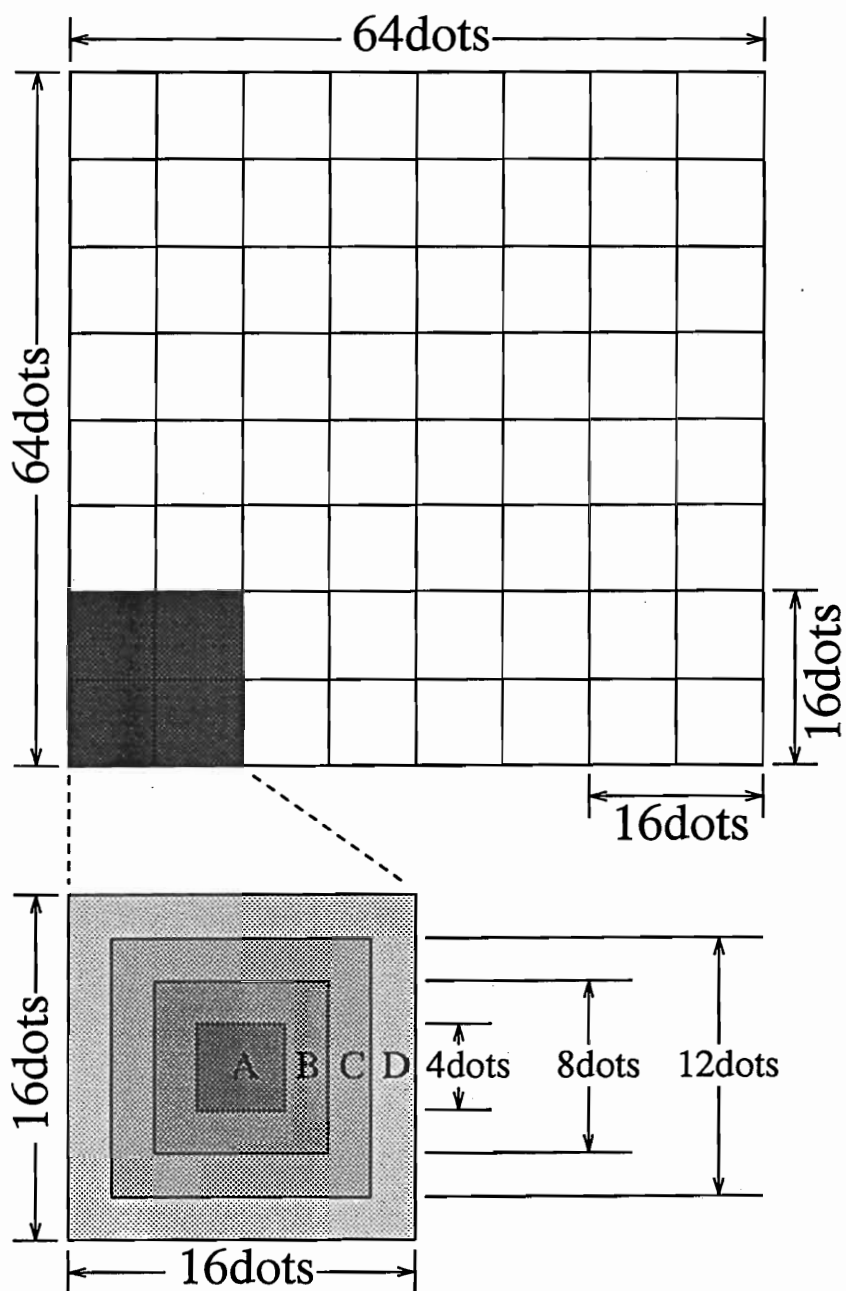


(b)



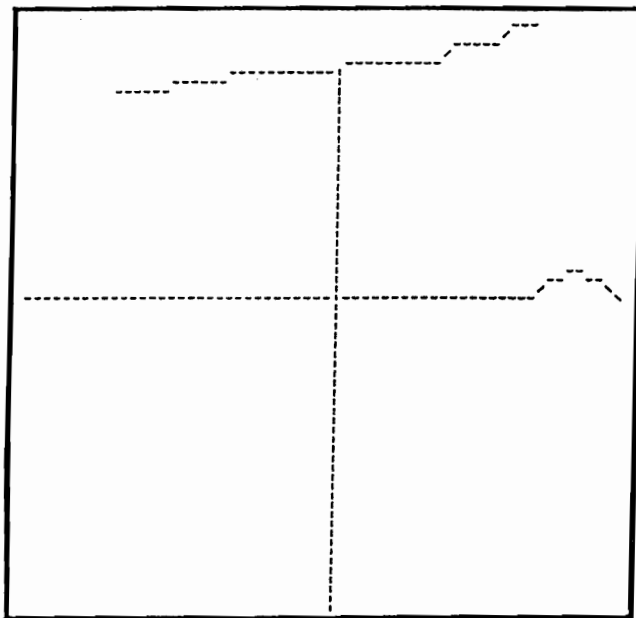
(c)

図 2.3: ノイズ除去・スムージングのためのマスク

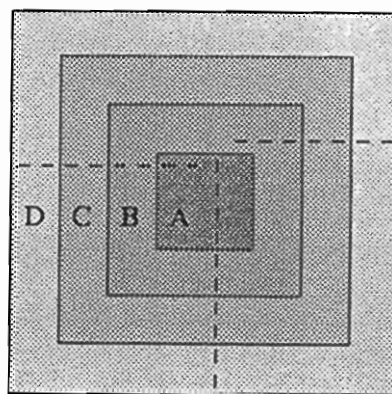


- 方向線素：“|”，“—”，“/”，“\”
- 1 領域：16 × 16 ドット
- 領域数：7 × 7 = 49 個
- 次元数：4 × 49 = 196 次元
- 重みのかけ方： 枠 A → 4
 枠 B → 3
 枠 C → 2
 枠 D → 1

図 2.4: 方向線素特徴量



第 4 領域



「千」

$$\begin{aligned}
 F(P^{\text{千}}) &= \left(F(P^{\text{千}})_{\text{第 1 領域}}, F(P^{\text{千}})_{\text{第 2 領域}}, \dots, F(P^{\text{千}})_{\text{第 49 領域}} \right) \\
 &= \left(F_{1|}, F_{1-}, F_{1/}, F_{1\setminus}, F_{2|}, F_{2-}, F_{2/}, F_{2\setminus}, \dots, F_{49|}, F_{49-}, F_{49/}, F_{49\setminus} \right) \\
 &= \left(F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8, \dots, F_{193}, F_{194}, F_{195}, F_{196} \right)
 \end{aligned}$$

第 4 領域の抽出例

$$\begin{aligned}
 F_{13} = F_{4|} &= \overbrace{4 \times 4}^{\text{A}} + \overbrace{2 \times 3}^{\text{B}} + \overbrace{2 \times 2}^{\text{C}} + \overbrace{2 \times 1}^{\text{D}} = 28 \\
 F_{14} = F_{4-} &= 2 \times 4 + 5 \times 3 + 4 \times 2 + 4 \times 1 = 35 \\
 F_{15} = F_{4/} &= 0 \times 4 + 0 \times 3 + 0 \times 2 + 0 \times 1 = 0 \\
 F_{16} = F_{4\setminus} &= 0 \times 4 + 0 \times 3 + 0 \times 2 + 0 \times 1 = 0
 \end{aligned}$$

$$\underline{F_{\text{第 4 領域}}} = (28, 35, 0, 0)$$

図 2.5: 方向線素特徴量の抽出例

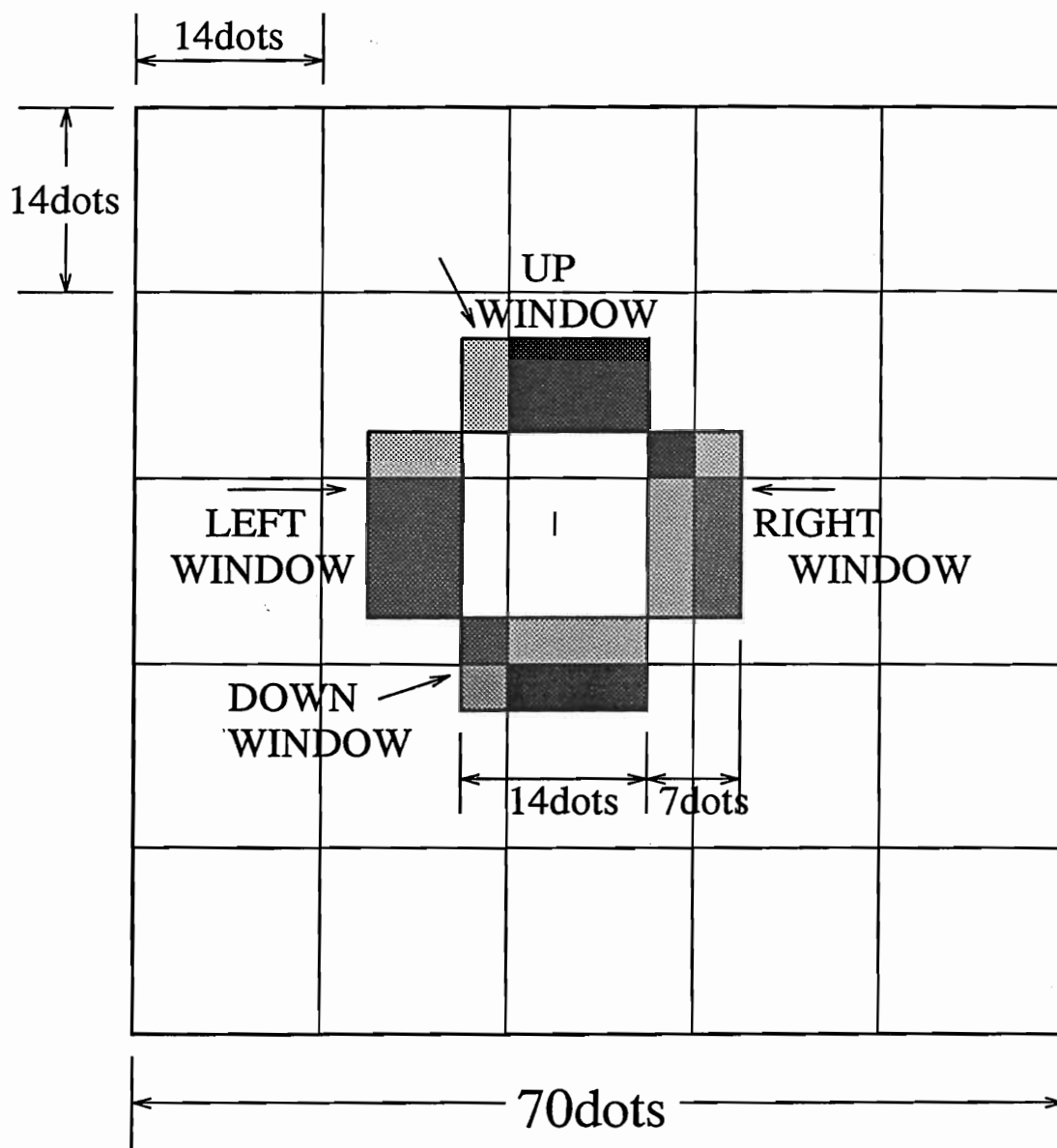
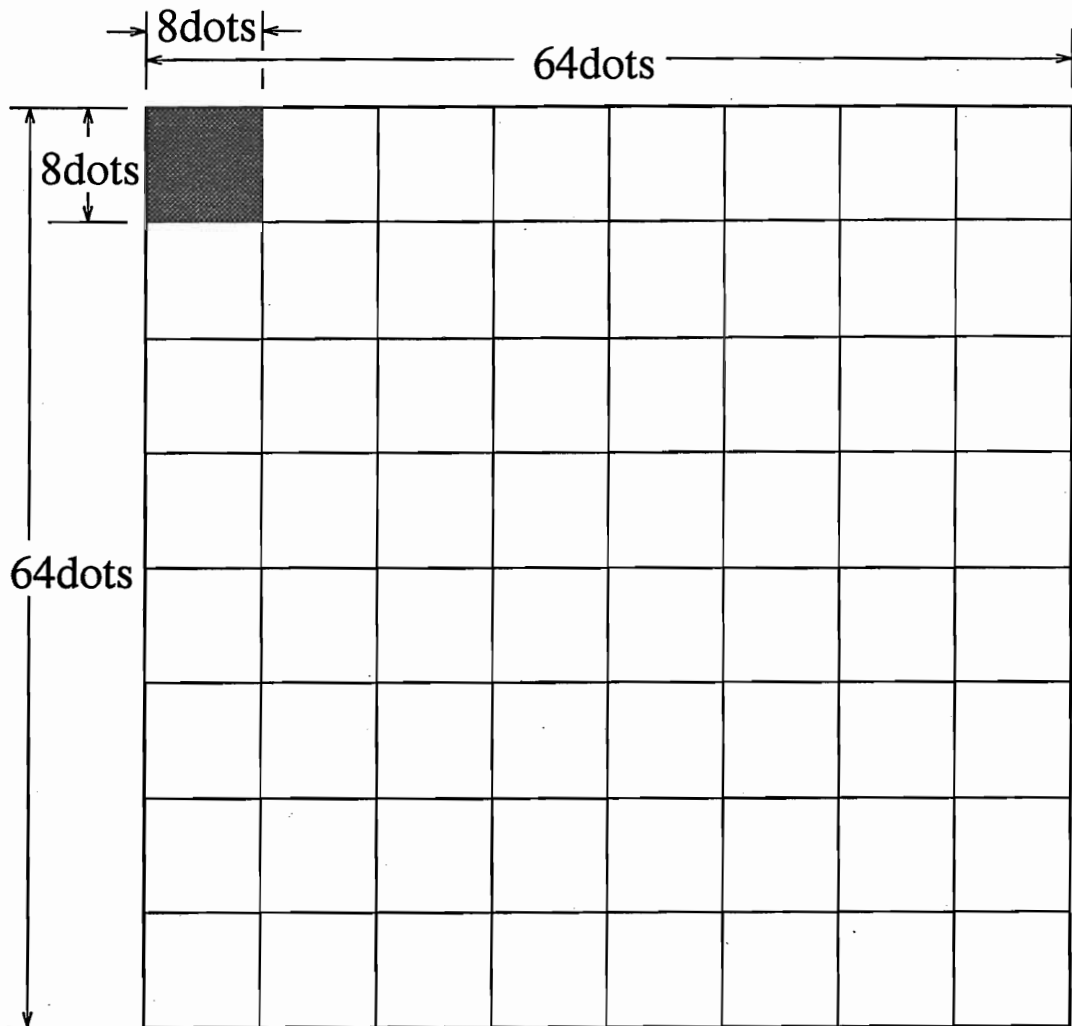


图 2.6: 方向密度特征量



- 1 要素 : 領域内の黒画素数
- 1 領域 : 8×8 ドット
- 領域数 : $8 \times 8 = 64$ 個
- 次元数 : 64 次元

図 2.7: メッシュ特徴量

2.4 特徴量に関する一般的性質

パターン整合法を特徴づける特徴抽出と距離尺度から得られる情報だけで、個々の文字パターンの認識が可能かどうか、またその方法について、特徴量空間がもつ構造をもとに考察する。

次の r_k を文字 k の領域半径という。

定義 1 (領域半径)

$$r_k = \max_{P \in \mathcal{K}(k)} E(F(P), v^k) \quad (2.8)$$

図 2.8 は領域半径の例である。図 2.8 のように文字 k の全てのパターン P の特徴量が特徴量空間内に分布している場合、 v^k から最遠の $F(P)$ までの距離を r_k とする。 r_k は実線で描かれた超球の半径を表す。次はこの領域半径の定義から明らかに成立する。

命題 1 $E(u, v^k) > r_k$ ならば、 u は k でない。

これは、図 2.8 の例に示すように実線で描かれた超球の外部に、未知文字文字の特徴量 u が位置したならば、 u は k でないという保証を与えることである。

文字 k のある特徴量の k に関する整合度は r_k 以下であるので、 u に対して、それを与える候補文字の集合を次のように定めることができる。

定義 2 (候補文字集合)

$$\text{Cand}(u) = \{k \mid E(u, v^k) \leq r_k\} \quad (2.9)$$

と定め、 u に対する候補文字集合と呼ぶ。 $\text{Cand}(u)$ に属するどの文字 k に対しても、 u は k に関する整合度の必要条件を満たしている。従って、 $|\text{Cand}(u)| \geq 2$ のときは、この整合度だけからは u が $\text{Cand}(u)$ のどの文字であるかは決定できない。しかし、命題 1 は $j \notin \text{Cand}(u)$ のとき、 u が j でないことを保証している。従って、次が成立する。

命題 2 $u \in \mathcal{F}$ かつ $|\text{Cand}(u)| = 1$ ならば、 u は k である。ただし、 $\text{Cand}(u) = \{k\}$ 。

$\text{Cand}(u)$ を知るには、 u と全字種の標準特徴量との整合度が必要である。しかし、あらかじめ文字パターンの集合を分析しておくことにより、字種のある部分集合に対してのみ整合度を求めることで、 $\text{Cand}(u)$ を求めることができる。そのためにまず次の定義をする。

文字 k の特徴領域 $C(k)$ を次のように定める。

定義 3 (特徴領域)

$$C(k) = \{u \mid E(u, v^k) \leq r_k\} \subset Z^N \quad (2.10)$$

特徴領域の例としては、図 2.8 において実線で描かれた超球の内部が文字 k の特徴領域を表す。明らかに、 $\mathcal{F}(k) \subset C(k)$ である。文字 k の特徴領域 $C(k)$ は、整合度のみを用いて定まる $\mathcal{F}(k)$ の近似であるといえる。この定義と命題 1 より、 $u \notin C(k)$ ならば、 u は k でない。しかし、 $u \in C(k)$ だからといって、 u が k であるとは判定できない。

文字 k の誤認識文字集合 $\text{Er}(k)$ を次のように定める。

定義 4 (誤認識文字集合)

$$\text{Er}(k) = \{j \mid j \neq k, C(j) \cap C(k) \neq \emptyset\} \subset K \quad (2.11)$$

図 2.9 に、誤認識文字集合の例を示す。文字 k の誤認識文字集合はその特徴領域と重なりがある特徴領域をもつ文字の集合 (図 2.9 では j_1 と j_2) で、文字 k のあるパターンを整合度だけで認識しようとするとき間違っ認識されてしまうかも知れない文字の集合を表している。文字 k と j の分離度 d_{kj} を $d_{kj} = E(v^k, v^j)$ と定める。以下では E は距離の公理を満たすものとする。

命題 3 $\text{Er}(k) = \{j \mid j \neq k, d_{kj} \leq r_k + r_j\}$

(証明) 任意の u について、距離の公理から、 $E(v^k, v^j) \leq E(u, v^k) + E(u, v^j)$ これと特徴領域の定義から明らかである。 \square

これは、各文字の領域半径と文字間の分離度から誤認識文字集合を求めることができることを示している。さらに次が成立する。

命題 4 任意の $k \in \text{Cand}(u)$ について、 $\text{Cand}(u) \subseteq \text{Er}(k) \cup \{k\}$ 。

(証明) $\forall j \in \text{Cand}(u)$, $E(u, v^j) \leq r_j$, $j \neq k$ のとき、 $d_{kj} \leq E(u, v^k) + E(u, v^j) \leq r_k + r_j$ 。従って、 $j \in \text{Er}(k)$ 。 \square

従って、ある文字 k について、 $u \in C(k)$ が判れば、各 $j \in \text{Er}(k)$ について、 $u \in C(j)$ かどうかを判定することで、 $\text{Cand}(u)$ が決定される。この方法は全字種について調べることに較べてかなり効率的になる可能性がある。また、逆に考えると、 $u \in C(k)$ が判れば、 u が $\text{Er}(k) \cup \{k\}$ に属さない文字であることはないことがわかる。命題 2 は次のようにも表現できる。

命題 5 u が文字パターンの特微量であるとする。 $u \in C(k)$ 、かつ、任意の $j \in \text{Er}(k)$ について $u \notin C(j)$ ならば、 u は k である。特に、 $\text{Er}(k) = \emptyset$ ならば、 $u \in C(k)$ は k である。

(証明) 命題 2、4 から明らかである。 \square

この前半は、 $\text{Er}(k) \neq \emptyset$ であっても、ある $u \in C(k) \cap \mathcal{F}$ に対して成立することがあり意味をもつ。これは誤認識する可能性のある文字に対して、消去法で正しい文字を確認することを意味する。この確認中にある j で $u \in C(j)$ ならば、 u がどの文字であるかは特定できなくなる。 $\text{Er}(k) = \emptyset$ の場合は、 $u \in C(k)$ を確認さえできれば、 u が k であることを保証することを意味する。 $u \in C(k)$ が $u \notin C(j)$ であるひとつの十分条件は次のように与えられる。

命題 6 $E(u, v^k) < d_{kj} - r_j$ ならば、 u は j でない。

(証明) $E(u, v^j) \geq E(v^k, v^j) - E(u, v^k) > d_{kj} - (d_{kj} - r_j) = r_j$ 。命題 1 から成立する。 \square

そこで次のように定義する。次の値 s_k を文字 k の安全認識半径という。

定義 5 (安全認識半径)

$$s_k = \begin{cases} \min_{j \in \text{Er}(k)} (d_{kj} - r_j), & \text{Er}(k) \neq \phi \\ r_k, & \text{Er}(k) = \phi \end{cases} \quad (2.12)$$

図 2.10 は安全認識半径の例である。図 2.10 より、 $\text{Er}(k) = \{j_1, j_2\}$ において $d_{kj_1} - r_{j_1} < d_{kj_2} - r_{j_2}$ なので $s_k = (d_{kj} - r_j)|_{j=j_1}$ となる。 s_k は、細実線で描かれた超球の半径を表す。次はこの命名の理由である。

命題 7 $E(\mathbf{u}, \mathbf{v}^k) < s_k$ ならば、 \mathbf{u} は k である。

(証明) 命題 5、6 と定義から明らかである。 □

これは図 2.10 において、細実線で描かれた半径 s_k の超球内に \mathbf{u} が位置したならば、 \mathbf{u} は k であるという保証を与えることである。

次の A を文字 k の安全認識領域と呼ぶ。

定義 6 (安全認識領域)

$$A(k) = \begin{cases} \{\mathbf{u} \mid E(\mathbf{u}, \mathbf{v}^k) < s_k\}, & s_k > 0 \\ \phi, & s_k \leq 0 \end{cases} \quad (2.13)$$

$A(k) \subseteq C(k)$ であり、 $\mathbf{u} \in A(k)$ ならば、 $|\text{Cand}(\mathbf{u})| = 1$ である。安全認識領域の例としては、図 2.10 において細実線で描かれた超球の内部が文字 k の安全認識領域を表す。安全認識半径は分離度と領域半径から決まり、それは一般には正負の値をとる。負の場合は次の意味をもつ。これは安全認識半径の定義から明らかである。

命題 8 (a) $s_k \leq 0$ ならば、ある j が存在し、 $\mathbf{v}^k \in C(j)$ 。

命題 8 (b) さらに、 $s_k \leq -r_k$ ならば、ある j が存在し、 $C(k) \subseteq C(j)$ 。

この命題で、(a) は標準特徴量自体が他の文字の特徴領域に属することを示し、(b) は文字 k の特徴領域が文字 j の特徴領域に含まれてしまい、 $\mathbf{u} \in C(k)$ ならば、 $|\text{Cand}(\mathbf{u})| \geq 2$ が必ず成立することを意味する。このとき、整合度だけでは \mathbf{u} を識別できないことになる。

次に、パターン整合法において 1 位候補が正しいという保証を与えることが、一般的にはできないということを示す。図 2.11(a) のように $E(\mathbf{u}, \mathbf{v}^k) < E(\mathbf{u}, \mathbf{v}^j)$ かつ $\mathbf{u} \notin C(k), \mathbf{u} \in C(j)$ のとき、 k の領域半径がかなり小さく (文字 k のパターンのぼらつきが小さい)、 j の領域半径がかなり大きい (j のパターンのぼらつきが大きい) 場合には、ユークリッド距離などを距離尺度とすると、未知入力文字の特徴量 \mathbf{u} の整合度で最もよい文字 k が必ずしも正しい認識を与えない場合がある。さらに (b) のように、その文字 k の誤認識文字集合にも正解が入っていないことがある。即ち、パターン整合法で整合度の最適値を与える候補が正解であると判断することに何の保証もなく、かえって間違ってしまうことがあることを示している。このことを示す命題として次が成立する。

命題 9 $j \notin \text{Er}(k)$ とする。ある \mathbf{u} について、 $E(\mathbf{u}, \mathbf{v}^k) < E(\mathbf{u}, \mathbf{v}^j)$ かつ $\mathbf{u} \notin C(k), \mathbf{u} \in C(j)$ ならば、 $r_k < \frac{d_{kj}}{2} < r_j$ 。 E が線形で平行移動不変であれば、この逆も成立する。

(証明) $r_k < E(\mathbf{u}, \mathbf{v}^k) < E(\mathbf{u}, \mathbf{v}^j) < r_j$ が成立している。 $d_{kj} \leq E(\mathbf{u}, \mathbf{v}^k) + E(\mathbf{u}, \mathbf{v}^j) < 2r_j$ 。

一方、仮定 $j \notin \text{Er}(k)$ と条件より、 $d_{kj} > r_k + r_j > 2r_k$ 。故に結論が成立する。逆は、 $\lambda = \frac{r_j - \varepsilon}{d_{kj}}$ ($\varepsilon > 0$) とおいて、 $\mathbf{u} = \lambda \mathbf{v}^k + (1 - \lambda) \mathbf{v}^j$ と \mathbf{u} を定めると、 $\mathbf{u} \in C(j)$ が成立する。仮定から、 $\mathbf{u} \notin C(k)$ 。条件から、 $r_j - \frac{d_{kj}}{2} > \varepsilon > 0$ となる ε を選ぶことができる。このとき、 $E(\mathbf{u}, \mathbf{v}^j) - E(\mathbf{u}, \mathbf{v}^k) = 2r_j - d_{kj} - 2\varepsilon > 0$ 。従って、逆が証明された。□

この命題の条件が成立する場合は証明に示されているように、1 位候補 k に対して、 $r_k < E(\mathbf{u}, \mathbf{v}^k)$ が成立しているから、命題 1 より \mathbf{u} は k でないことが保証され、 k を候補から棄却できる。これは、パターン整合法において、領域半径を求めておくことで認識結果をより確かなものにできることを意味する。

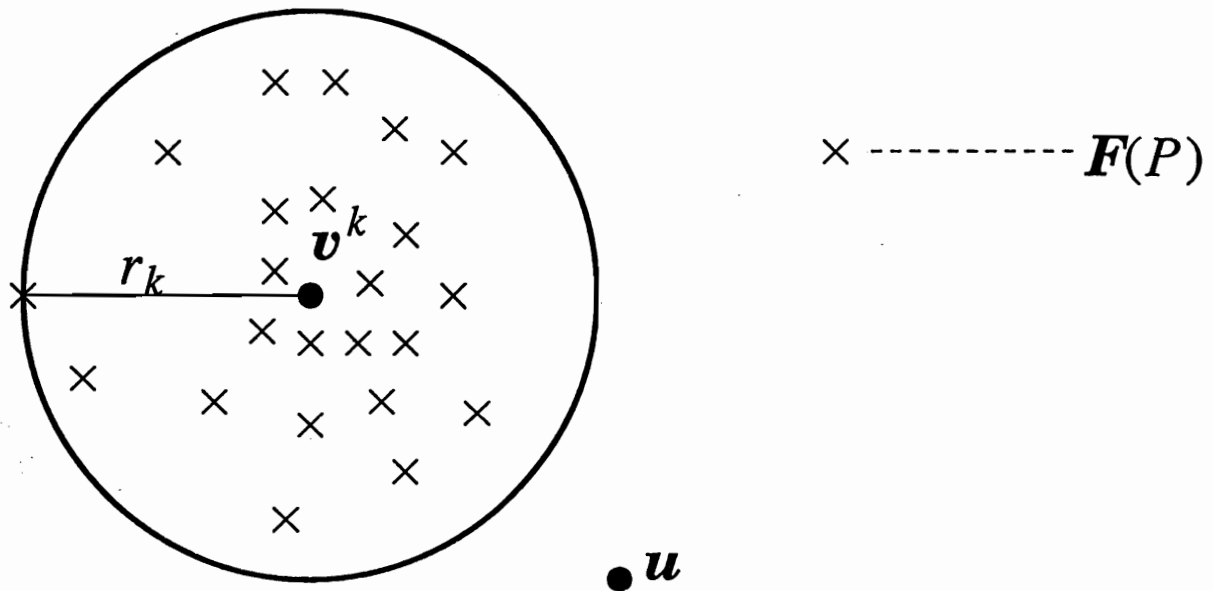


図 2.8: 領域半径の例

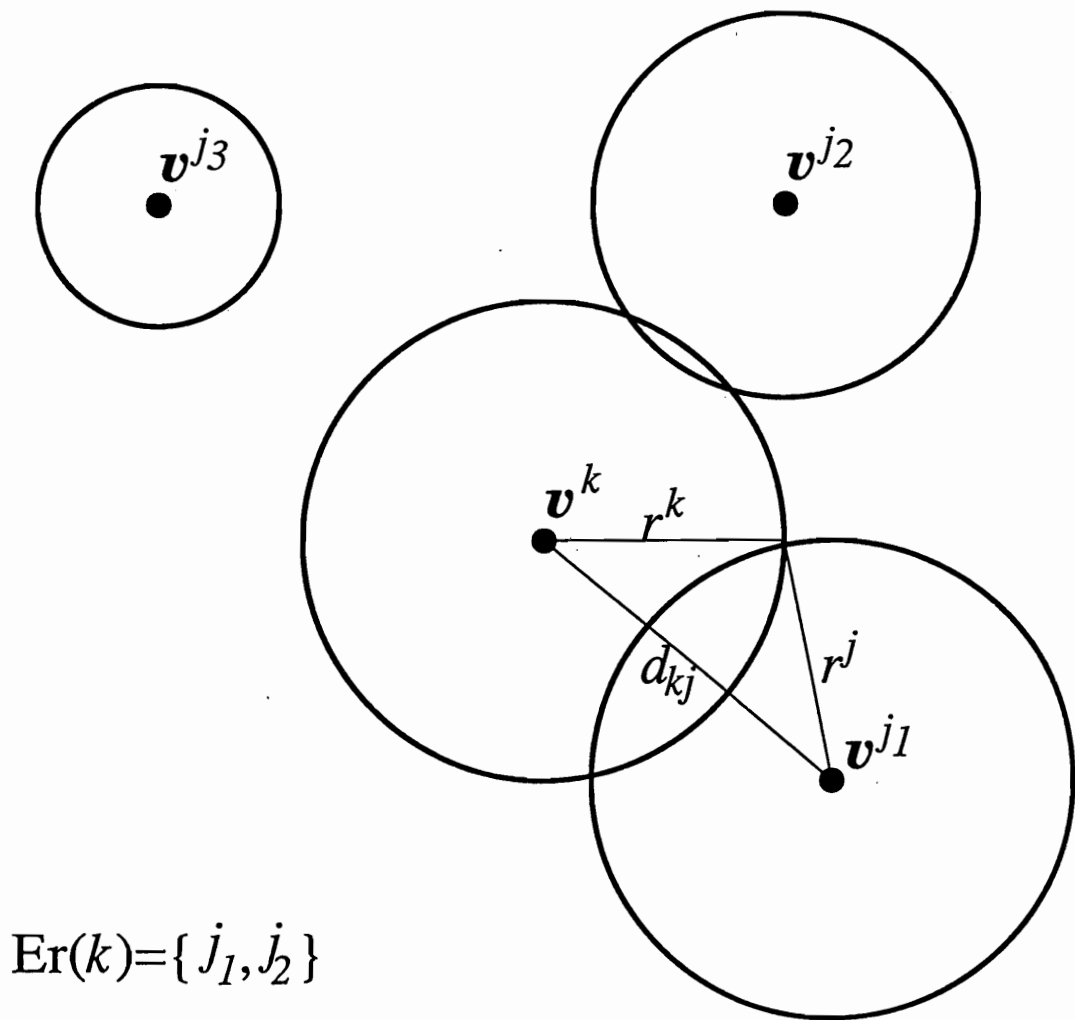


図 2.9: 誤認識文字集合の例

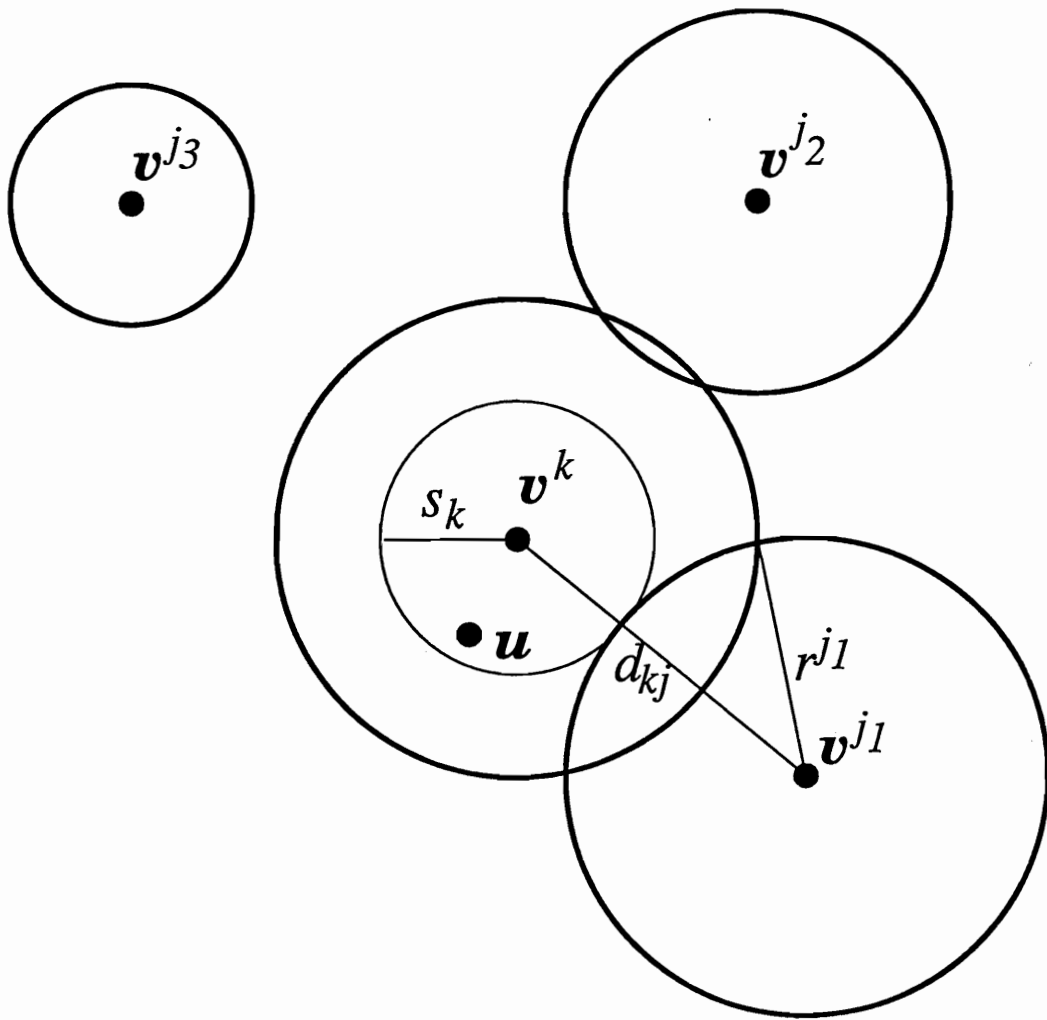
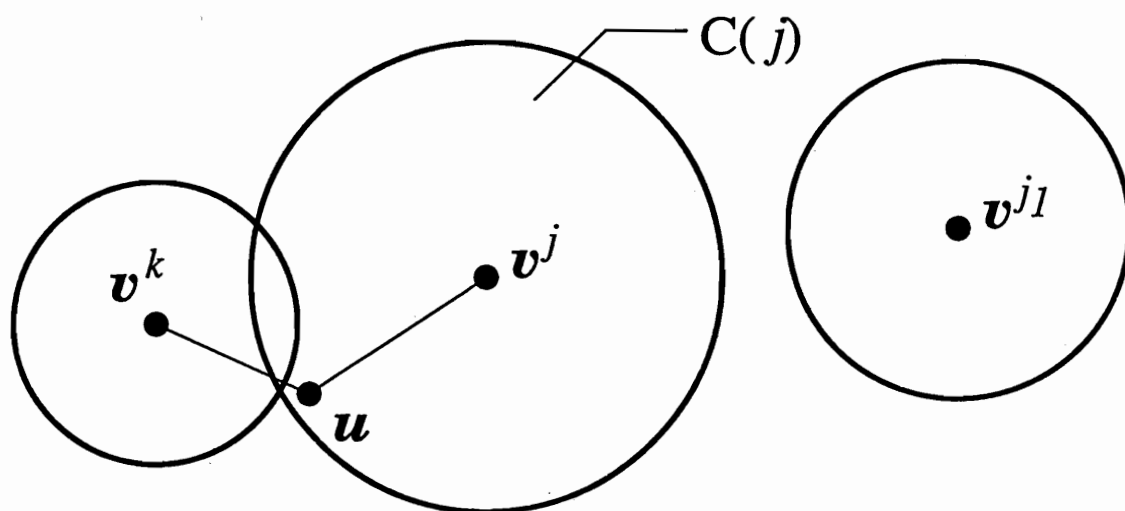
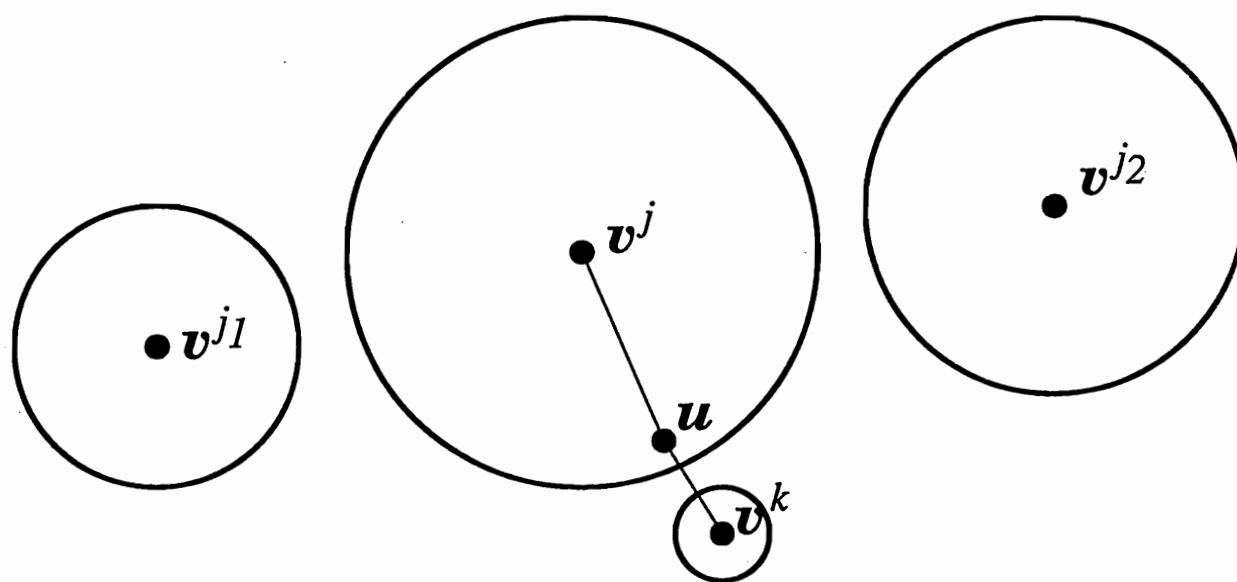


図 2.10: 安全認識半径の例



(a) $E(u, v^k) < E(u, v^j)$, $u \notin C(k)$, $u \in C(j)$ のとき



(b) さらに $j \notin \text{Er}(k)$ のとき

図 2.11: 評価値の最小値が必ずしも正解を与えない例

2.5 後書き

この章では、距離尺度を用いたパターン整合法に焦点をあて、与えられた特徴抽出法に対してその特徴量空間を分析し、認識結果の正しさを保証できるための理論的根拠を導いた。また、領域半径や安全認識半径が正しさの判定基準であることを示し、パターン整合法による認識が必ずしも正しくはないことを明らかにしている。

第 3 章

特徴抽出法の性能評価

3.1 前書き

本章では、具体的な文字パターン集合および具体的な特徴抽出法に対して、第 2 章での理論的な議論は有効かどうかを調べる。まず、実験の目的では第 2 章で定義された命題や定義に対し、それぞれについて検討すべきことを具体的にあげ、まとめておく。次に実験の対象となるものについての説明をし、実験に入る。実験結果から第 2 章の議論における妥当性を述べ、併せて以後の利用方法についても述べる。

3.2 実験の目的

距離尺度を用いたパターン整合法に、その妥当性は一般的にはないことがわかったが実際には検討しなければならない。これらを実験的に評価するため、以下のことを考える。個々の文字特徴量 u に関して、候補文字集合 $\text{Cand}(u)$ の大きさが 1 であることが常に成立すれば、正確に認識することを保証できる。これは実際に成立するだろうか。各特徴量 u に対してこの問いに答えることは、 u が非常に多いことから現実的には不可能である。そこで、命題 4 から、誤認識文字集合の大きさ $|\text{Er}|$ が重要となる。 $\text{Er}(k) = \phi$ となることが理想的であるが、 $|\text{Er}(k)|$ が小さければ各 u において $|\text{Cand}(u)| = 1$ となる可能性もある。実際の文字パターン集合および特徴抽出法に対して、それらから決まる $|\text{Er}|$ はどの程度であろうか。

さらに、各 u が安全認識領域内に入るならば、正確に認識できることが保証された。これはどの程度成立するであろうか。安全認識半径自体は正の値をとるであろうか。領域半径に比べて十分に大きいであろうか。定性的には、 $|\text{Er}(k)|$ が小さいためには領域半径は小さく、正確に識別するためには安全認識半径が大きいことが望ましい。

また、パターン整合法で第 1 位候補になった文字に対して、未知入力文字の特徴量はその特徴領域内に実際に入っているであろうか。明らかに、第 1 位候補が正しい文字であれば、未知入力文字の特徴量は特徴領域内に存在する。それが、さらに安全認識領域内に存在することが確認できれば、正しい認識であるという保証を与えることができる。この保証を与えることができる文字パターンはどの程度存在するであろうか。

第1位候補が間違っている場合は、その候補の特徴領域内に入る場合と入らない場合がある。特徴領域内に存在しなければ、その文字は正しくないという保証を与えることができ、正解は第2位以下にあることがわかる。第1位候補が間違っていてかつその特徴領域内にある場合は、第2位以下の候補を調べる必要があるが、第2位以下の候補文字の安全認識領域内には入らないはずであり、整合度だけからは正しいという保証を与えることはできない。それぞれ、どの程度成立するだろうか。

具体的に与えられた特徴抽出法(特徴量の定め方)に対して、以上の問題に答えることは、その性能の評価になると考えられる。勿論、以上の問題への答えは各文字パターンに依存する。従って、ある固定された文字パターン集合に対して、上に述べたことに関する統計量(分布の形、平均値、分散など)が、特徴抽出法の性能を示すと考えられる。特徴抽出法に関する従来の定量的評価は認識率だけであったと言えるが、上に述べたことを実験で定量的に評価することにより、個々の文字パターンの認識結果の正しさを保証するという観点からの評価が可能になると思われる。

実際の文字パターン集合と特徴抽出法に対して実験を行い、併せて特徴抽出法の性能評価法として実際に役立つかどうかを検討する。

3.3 実験の対象

文字パターンは 64×64 ドットの正規化された2値画像を考える。特徴抽出法は抽出された特徴量の名前によって区別されているが、ここでは、方向線素特徴量、方向密度特徴量、メッシュ特徴量を選び*、この3つについて、領域半径、安全認識半径、誤認識文字集合の大きさなどに関して統計をとった。方向線素特徴量は活字文字に対して高い認識率を与えている。方向密度特徴量は手書き文字に対して方向密度ベクトルとして定義されたものである。メッシュ特徴量は古くから大分類のために使われているものであるが、ここでは 8×8 の小領域の黒画素数を数えて64次元のベクトルとしたものを採用した。

標準特徴量の定め方は各種あるが、本実験ではパターン整合法の多くの認識実験で使われている重心ベクトルを考える。文字 k の文字パターンの適当なサンプル集合 $S(k) \in \mathcal{K}(k)$ を与え、その各パターンの特徴量の平均を v^k とする。 $(v^k \in \mathcal{F}(k))$ とは限らない。)整合性の距離尺度としてはユークリッド距離およびシティーブロック距離の2種類を用いて、それぞれについて実験をした。

各文字 k の実際の文字パターン集合 $\mathcal{K}(k)$ を明示的に定めることができないので、領域半径を現実を求めることはできない。そこで、標準特徴量を求めたサンプル集合 $S(k)$ をもとに近似することにより、標本領域半径 r_k^S として次のように定める。

定義 7 (標本領域半径)

$$r_k^S = \max_{P \in S(k)} E(F(P), v^k) \quad (3.1)$$

図 3.1に標本領域半径の例を示す。特徴量空間内に文字 k のサンプルの特徴量 $F(P)$ が有限個図 3.1のように分布した場合、実線で描かれた超球の半径が r_k^S を表わしている。点線で

*これらを選んだ理由は、著者らのまわりで提案されたり、使用されているものであるということが大きい。方向線素特徴量の性能を知ることが第1の目的で他は比較のために選んだ。

描かれた超球の半径は r_k を表わしている。明らかに、 $r_k^S \leq r_k$ である。安全認識半径についても真の値を知ることはできず、この標本領域半径を用いて同様に定義し、標本安全認識半径 s_k^S と呼ぶ。 $s_k \leq s_k^S$ である。誤認識文字集合についても同様に、標本領域半径を用いて定めるが、真のものより小さくなっている。 $S(k)$ は $\mathcal{K}(k)$ を適切に反映していることが期待されるが、実際の認識で $S(k)$ をもとに定めたこれらの値や集合を使うことの妥当性については実験的に確認すべきことである。しかし、この確認は今後の問題であり、本章ではまず、第 2 章で述べた理論が具体的な文字パターン集合に対して、どの程度有効に働くかを確認するというを主目的として、 $S(k)$ 自体を文字パターンの世界とする、いわば $S(k)$ を $\mathcal{K}(k)$ とみなす立場で実験をした。従って、以下では、標本領域半径、標本安全認識半径を単に、領域半径、安全認識半径と呼ぶ。

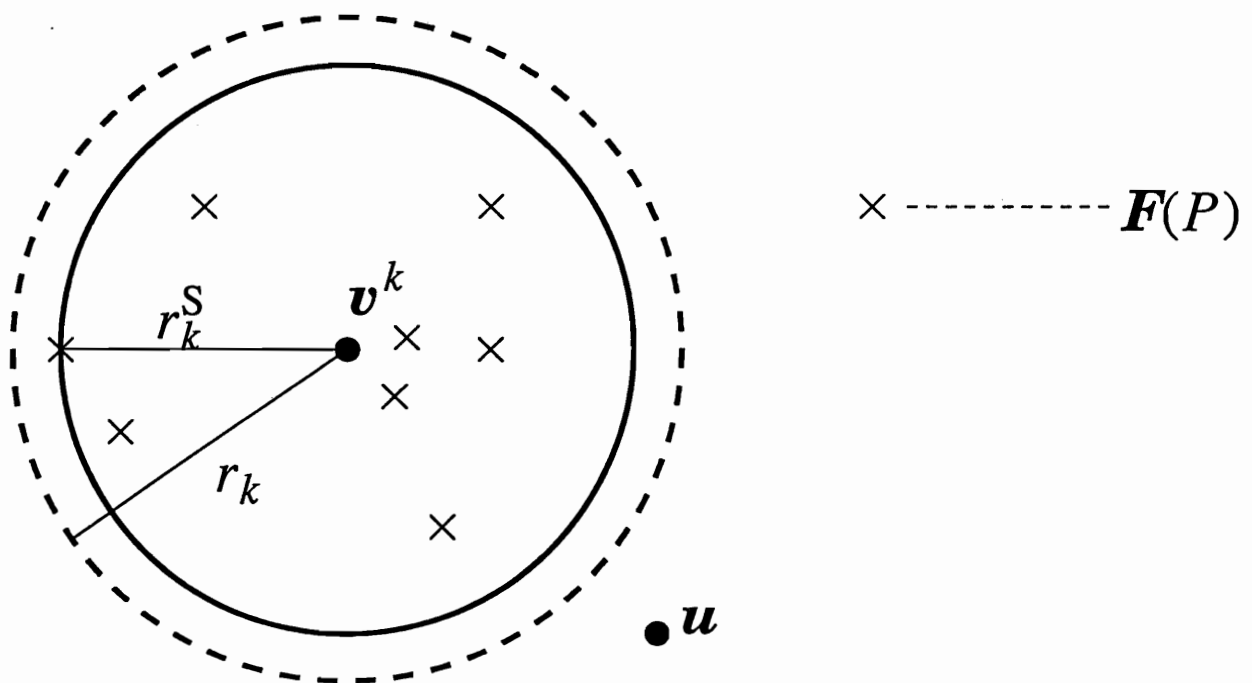


図 3.1: 標本領域半径の例

3.3.1 サンプル集合

実験で用いた文字パターンのサンプル集合は、潰れ等のない本来の文字パターンを対象とできるように、4 倍角および縦倍角で印字したものを対象とする。表 3.1 に示すように JIS 第 1 水準漢字 2965 字種の 4 倍角パターンを、3 種類のドットプリンタと 1 台のレーザープリンタを用いて印字したもの計 7 セット (1 台のドットプリンタによる印字だけ 1 回で、他は 2 回の印字) とそのうちの 3 セットをコピーしたものとの計 10 セットのデータセット甲とコピーの 3 セットを縦倍角で印字したものに置き換えた計 10 セットのデータセット乙をサンプル集合とした。甲は、各文字に関するパターンの違いが小さく、特徴量の分散が

小さいことが期待される。乙は、大きさの違う文字パターンが加わって、特徴量の分散は大きくなるはずである。

甲、乙のデータセットに関して独立に、それぞれの文字の領域半径、安全認識半径、誤認識文字集合の大きさを調査する。2つのデータセットから、各文字について2つのサンプル集合が得られ、それらについて独立に実験する理由は、サンプル集合の選び方で領域半径などの統計量に違いがどの程度であるかを知るためである。サンプル集合による違いを知ることによって、実際の認識に使うためのサンプル集合の選択のめやすが得られると考えられる。

各文字 k について、 $S(k)$ の大きさ $s = |S|$ は 10 で、標準特徴量は 10 セットのパターンの特徴量の平均である。

表 3.1: 実験用文字データ

甲			
1	エプソンプリンター印刷文字	4 倍角	No. 1
2	エプソンプリンター印刷文字	4 倍角	No. 2
3	エプソンプリンター印刷文字	4 倍角	No. 1 のコピー
4	富士通プリンター印刷文字	4 倍角	No. 1
5	富士通プリンター印刷文字	4 倍角	No. 2
6	富士通プリンター印刷文字	4 倍角	No. 2 のコピー
7	キャノンレーザープリンター印刷文字	4 倍角	No. 1
8	キャノンレーザープリンター印刷文字	4 倍角	No. 2
9	キャノンレーザープリンター印刷文字	4 倍角	No. 2 のコピー
10	NEC プリンター印刷文字	4 倍角	No. 1

乙			
1	エプソンプリンター印刷文字	4 倍角	No. 1
2	エプソンプリンター印刷文字	4 倍角	No. 2
3	富士通プリンター印刷文字	4 倍角	No. 1
4	富士通プリンター印刷文字	4 倍角	No. 2
5	富士通プリンター印刷文字	縦倍角	No. 1
6	キャノンレーザープリンター印刷文字	4 倍角	No. 1
7	キャノンレーザープリンター印刷文字	4 倍角	No. 2
8	キャノンレーザープリンター印刷文字	縦倍角	No. 1
9	NEC プリンター印刷文字	4 倍角	No. 1
10	NEC プリンター印刷文字	縦倍角	No. 1

3.4 実験結果と考察

まず、各文字の領域半径を求めた。その度数分布は正規分布に近いものであった。その統計量を表 3.2 に示す。ただし、整合度(距離値)は、3つの特徴抽出法の比較のため、各特微量と距離尺度ごとにデータセット乙の平均値を100とした相対的な値で示した。対応する実際の距離値も示してある。この平均値が分布のピークを与える半径にほぼ等しい。甲のデータだけを用いた領域半径が乙のそれより小さくなっているが、それはデータセットの選び方から予測されることである。その領域半径の差は比較的小さく、ある意味で2つのサンプル集合それぞれが真の文字パターン集合における分布を適切に反映しているとも解釈できる。各特微量ごとの最小に対する最大の比は、ユークリッド距離で5~11倍であり、シティーブロック距離では6~23倍となっている。各文字の領域半径の大きさは、各文字パターンの変形の程度を示すものとも考えられ、文字パターンの変形が文字毎にかなり異なっていることを示している。3種類の特微量がその異なり方をどう適切に吸収しているかを、この結果から相対的に解釈すれば、メッシュ特微量について比が最も小さく変形を吸収しているといえる。方向線素特微量、方向密度特微量では、その倍くらいになっており変形がそのまま特微量に反映していると言えるかもしれない。

次に安全認識半径について調べた。各文字の安全認識半径の度数分布は、方向線素特微量、メッシュ特微量についてはほぼ正規分布に近く、方向密度特微量に関してはほぼ正規分布に近いが幾分大きい方に偏っていた。その統計量を表 3.3 に示す。ただし、表 3.2 と同様に乙の領域半径の平均値を100とした相対的な値で示している。定義からはすべて100以下のはずであるが、最大値が100以上になっているのは分離度と領域半径の差の最小値を求めるとき $Er(k)$ の代わりに全字種 K を用いて K に関する最小値を計算したためである。(100との正の差は、その特徴領域と隣の特徴領域との境界のギャップを示している。) 表 3.3 の平均値もそれらの値の平均として求めている。表 3.3 より、方向線素特微量を用いたものは平均-標準偏差がおよそ0に近いことから、大部分の文字が正の安全認識半径を持っていることが分る。また、領域半径の $\frac{1}{4} \sim \frac{1}{5}$ の所にピークがある。これに対して、方向密度特微量については、平均+標準偏差さえ負になっており(データ乙)、大部分が有効な安全認識半径を持っていないことを示している。これは特徴領域が重なり過ぎていることを意味する。メッシュ特微量については、1つを除いて平均値は正であるが小さく、有効な安全認識半径をもつ文字が半数程度であることを意味する。

具体的に領域半径と安全領域半径との比較による場合分けで全文字を分類したものが表 3.4 である。上段が全2965字種に対する文字数、下段がその率である。 $s > r > 0$ の場合が $Er(k) = \phi$ を与える文字 k の数である。方向線素特微量では2桁の数が正確に認識できる保証が与えられたが、全体に占める割合は小さい。方向密度特微量ではほとんどの文字で保証できない。メッシュ特微量で10文字程度が正確に識別可能である。表 3.4 で $s > r > 0$ となる文字が、個々のパターンによっては安全認識半径を用いて認識結果の正しさの保証を与えることができるものである。

実際の文字パターンが安全認識領域内に属する数とその率を表 3.5 に示す。上段が全29650パターンに対するパターン数、下段がその率である。方向線素特微量については、表 3.4 から安全認識半径が有効となる文字種の占める割合が最も大きい、表 3.5 から実際のパター

ンで安全認識領域に属するものの割合は小さいことがわかる。これは、表 3.3の s の値が r に比べて小さいことから予想がつくことである。他の2つについてもその割合は小さく、正確な認識のために安全認識半径は大きくは役立たないことがわかる。しかし、方向線素特徴量については、安全認識領域にあるパターンの占める割合が4~8%となっており、パターン整合法で得られた第1位候補に対して安全認識半径内かどうかを調べるだけで、それだけの文字パターンの認識結果の正確さを保証できることを意味し、正確さの保証として一定の役割は果たせるように思われる。メッシュ特徴量については2%前後のパターンに有効であるが、方向密度特徴量ではほとんど役に立たない。表 3.5は3つの特徴量について、各文字パターンの特徴領域の分離に関する性質の差異を現わしていると言える。

次に、各文字の誤認識文字集合の大きさ $|Er(k)|$ について調べた。その統計量を表 3.6(明かに、 $0 \leq |Er(k)| < 2965$ である)に示す。この大きさは各文字の特徴領域の重なりを示し、 $|Er(k)| + 1$ は、 $|Cand(u)|$ の上限を与えており、小さい方がよい。特徴抽出法に関する差異は平均値において、方向線素特徴量、メッシュ特徴量、方向密度特徴量の順に桁が違っており、方向線素特徴量が最もよいことがわかる。

特に、方向線素特徴量でデータ甲にユークリッド距離を用いた場合の $|Er(k)|$ の度数分布を図 3.2に示す。横軸の階級の幅は10字であるが、1番左は $|Er(k)| = 0$ のみの度数であり、 $Er(k) = \phi$ を満たす度数を意味するので特別に図示した。図 3.2は指数分布の形であるといえる。他の場合の分布も例外を除いて同様であった。例外は、方向密度特徴量でデータセット乙を用いたもので、双峰性のなだらかな分布であった。図 3.2は約470文字について誤認識文字集合の大きさが10以下で、大きさが200(平均+標準偏差に近い値である)以下である文字が大部分を占めていることを示している。他の場合も、大きさが平均+標準偏差以内である文字(表 3.6の最右欄)が大部分を占めている。特徴領域の分布状況を反映する指標として、誤認識文字集合の大きさの平均と標準偏差が適切であるように思われる。

最後に、パターン整合法による認識率と第1位候補が間違った場合(誤認識数)に特徴領域外であった文字パターン数を表 3.7に示す。全文字パターンの数は29650であり、その内誤認識されたものが、方向線素特徴量では2桁、他の2つは3桁の数であった。回復率は、誤認識文字パターンのうち特徴領域外であった文字パターンの占める割合であり、領域半径を用いることで第1位候補を棄却できる文字パターンの割合である。第2位以下に正解があるとわかるということで回復率と称した。これは56%以上となっており、パターン整合法での1位認識率がほぼ99%以上という高い状況でさらに認識率を上げるために領域半径を有効に生かせることを示している。特に、シティーブロック距離を用いる場合は有効であるといえる。

表 3.2: 領域半径

			最小	平均	標準偏差	最大	*の実距離
方向線素 特徴量	ユークリッド距離	甲	24.2	95.5	19.9	282.4	86.8
		乙	41.5	100.0*	21.3	292.7	
	シティーブロック距離	甲	11.1	95.7	22.8	255.9	714.7
		乙	24.1	100.0*	24.0	264.3	
方向密度 特徴量	ユークリッド距離	甲	23.0	79.7	20.3	232.1	343.4
		乙	22.1	100.0*	30.2	236.5	
	シティーブロック距離	甲	13.4	77.2	19.4	210.4	3619.1
		乙	12.6	100.0*	33.4	206.8	
メッシュ 特徴量	ユークリッド距離	甲	37.9	93.7	17.6	172.6	71.3
		乙	43.5	100.0*	17.4	193.6	
	シティーブロック距離	甲	25.9	93.3	19.3	171.8	437.1
		乙	29.7	100.0*	19.5	179.8	

表 3.3: 安全認識半径

			最小	平均	標準偏差	最大
方向線素 特徴量	ユークリッド距離	甲	-101.4	24.3	25.7	103.7
		乙	-110.6	16.0	25.7	93.4
	シティーブロック距離	甲	-68.3	26.5	25.2	120.7
		乙	-71.2	20.6	25.2	102.7
方向密度 特徴量	ユークリッド距離	甲	-104.5	-11.2	20.5	43.7
		乙	-140.7	-53.0	16.5	18.3
	シティーブロック距離	甲	-88.5	-2.1	22.1	76.5
		乙	-144.8	-49.4	20.9	36.3
メッシュ 特徴量	ユークリッド距離	甲	-77.2	8.0	21.7	117.9
		乙	-61.7	7.8	20.1	103.8
	シティーブロック距離	甲	-83.5	0.6	23.3	127.7
		乙	-78.7	-1.9	22.8	103.9

表 3.4: 領域半径と安全認識半径の関係

			$s > r > 0$	$r \geq s > 0$	$0 \geq s > -r$	$0 > -r \geq s$
方向線素 特徴量	ユークリッド距離	甲	41 字 1.38%	2424 字 81.75%	500 字 16.86%	0 字 0.00%
		乙	19 字 0.64%	2199 字 74.17%	747 字 25.19%	0 字 0.00%
	シティーブロック距離	甲	57 字 1.92%	2477 字 83.54%	431 字 14.54%	0 字 0.00%
		乙	30 字 1.01%	2330 字 78.58%	605 字 20.40%	0 字 0.00%
方向密度 特徴量	ユークリッド距離	甲	0 字 0.00%	947 字 31.94%	2011 字 67.82%	7 字 0.24%
		乙	0 字 0.00%	1 字 0.03%	2835 字 95.62%	129 字 4.35%
	シティーブロック距離	甲	7 字 0.24%	1476 字 49.78%	1476 字 49.78%	6 字 0.20%
		乙	0 字 0.00%	30 字 1.01%	2773 字 93.52%	162 字 5.46%
メッシュ 特徴量	ユークリッド距離	甲	10 字 0.34%	1925 字 64.92%	1030 字 34.74%	0 字 0.00%
		乙	1 字 0.03%	1951 字 65.80%	1013 字 34.17%	0 字 0.00%
	シティーブロック距離	甲	11 字 0.37%	1521 字 51.30%	1432 字 48.30%	1 字 0.03%
		乙	8 字 0.27%	1389 字 46.85%	1568 字 52.88%	0 字 0.00%

表 3.5: 安全認識領域に属するパターン数とその割合

距離尺度	ユークリッド距離		シティーブロック距離	
	甲	乙	甲	乙
データ				
方向線素 特徴量	2305 字	1284 字	2498 字	1636 字
	7.77%	4.33%	8.42%	5.52%
方向密度 特徴量	54 字	0 字	474 字	0 字
	0.18%	0.00%	1.60%	0.00%
メッシュ 特徴量	799 字	482 字	752 字	495 字
	2.69%	1.63%	2.54%	1.67%

表 3.6: 誤認識文字集合の大きさ及び平均+標準偏差までの文字数

			最小	平均	標準偏差	最大	平均+標準偏差 までの文字数
方向線素 特徴量	ユークリッド距離	甲	0	84.5	128.0	2118	2716
		乙	0	165.2	212.7	2427	2619
	シティーブロック距離	甲	0	55.0	77.4	1197	2669
		乙	0	102.3	134.0	1378	2624
方向密度 特徴量	ユークリッド距離	甲	1	529.3	449.3	2784	2514
		乙	17	1695.1	739.0	2964	2301
	シティーブロック距離	甲	0	200.7	224.5	2563	2592
		乙	11	1262.4	752.8	2954	2319
メッシュ 特徴量	ユークリッド距離	甲	0	442.4	464.6	2828	2510
		乙	0	598.3	539.8	2731	2456
	シティーブロック距離	甲	0	370.1	397.7	2557	2531
		乙	0	507.5	472.4	2516	2470

表 3.7: 誤認識文字パターン数と特徴領域外となった文字パターン数及び認識率

距離尺度		ユークリッド距離		シティーブロック距離	
データ		甲	乙	甲	乙
方向線素 特徴量	領域外	13 字	15 字	21 字	29 字
	誤認識数	23 字	26 字	25 字	34 字
	回復率	56.5%	57.7%	84.0%	85.3%
	認識率	99.922%	99.912%	99.916%	99.885%
方向密度 特徴量	領域外	162 字	620 字	86 字	639 字
	誤認識数	233 字	994 字	106 字	830 字
	回復率	69.5%	62.4%	81.1%	77.0%
	認識率	99.214%	96.648%	99.642%	97.201%
メッシュ 特徴量	領域外	98 字	87 字	159 字	139 字
	誤認識数	114 字	111 字	188 字	182 字
	回復率	86.0%	78.4%	84.6%	76.4%
	認識率	99.616%	99.626%	99.366%	99.386%

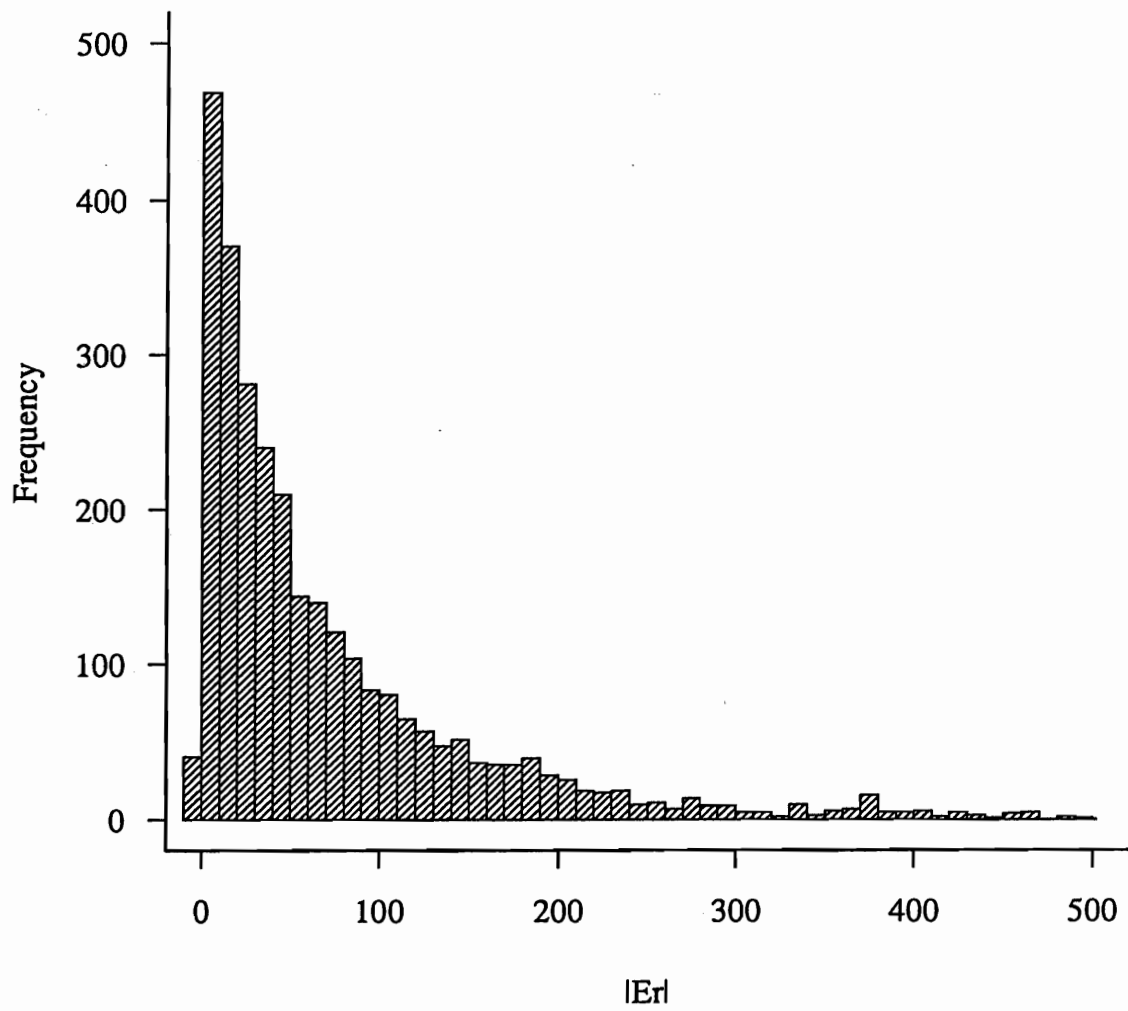


図 3.2: 誤認識文字集合の大きさの度数分布

3.5 パターン整合法への応用

パターン整合法において、整合度を用いるだけでは厳密な正確さの保証を一般的には実現しえないことが明らかになったが、本章の実験結果は、より効率的でかつ正確さを部分的に保証する(命題 1により保証する)方法があることを示唆している。そのいくつかを示す。

まず、パターン整合法における辞書に加えて、各文字に関する領域半径、安全認識半径、誤認識文字集合の情報を求めておく。

未知入力文字の特徴量が第 1 位候補の特徴領域内にあるか(距離値が領域半径以内か)どうかを調べることで、入っていなければそれを棄却する。さらに、特徴量が第 1 位候補文字の安全認識領域内にあるかどうかを調べることで、入っていればその正しさを保証できる。方向線素特徴量の場合、これらが有効に働くことは実験結果が示している。

誤認識文字集合は次のように利用することで整合度の計算を高速化できる。整合度の計算において、その値が領域半径以下であるかどうかを調べ、ある文字 k で半径以下であることが確認されたら、以後は文字 k の誤認識文字集合に属する文字だけを対象に整合度を計算する。この方法でも全字種を対象とする整合度の計算による認識結果と同一の結果を得ることを、命題 4 で保証している。これらは第 4 章で特徴領域整合法を提案し検討する。

また、ある特徴抽出法に関する各文字の誤認識文字集合は絶対的である。即ち、命題 1、4 が、その元以外の文字は正しくないという保証を与えている。従って、複数の特徴抽出法に関する誤認識文字集合を求めると、その共通部分と自分自身との和集合だけに真の候補が存在することになる。これは性質の異なる特徴量を組み合わせることでより正確に候補を絞り込めることを意味し、厳密な正確さの保証を与える方法の一つの可能性を示していると思われる。

それから、候補文字集合 $\text{Cand}(u)$ の元が 2 以上のとき、整合度だけでは正しい認識を保証しえない。これは、別な観点から $u \in \mathcal{F}(k)$ を検証する必要があることを意味する。従って、複数の特徴抽出を行なうこと、あるいは、構造分析法などを併用して検証すべきことの根拠を与えている。

3.6 後書き

実際の文字パターン集合を対象に、具体的に、方向線素特徴量、方向密度特徴量、メッシュ特徴量を用いて実験的に検討し、ある程度有効であることを明らかにした。この実験結果は、領域半径、安全認識半径、誤認識文字集合の大きさなどが具体的な特徴抽出法の性能評価に使えることも示している。また、この結果から、パターン整合法に正確さの検証法を部分的に付加できることや整合度の計算を効率化できることなどが明らかになった。

また、今回の実験はクローズ実験であり、標本領域半径を領域半径とみなしたが、実際の認識に適用する場合は、標本領域半径を求めたサンプル集合以外の文字パターンも識別しなければならず、真の領域半径をいかに推定するかが重要になる。これらの検討は第 4、5 章で検討する。

第 4 章

パターン整合法への適用

4.1 前書き

第 3 章 3.5 節で領域半径、安全認識半径、誤認識文字集合の情報を用いることにより、効率的でかつ正確さを部分的に保証する (命題 1 を保証する) 方法があることを幾つか示した。これらの中で本章では未知入力文字の特徴量 u が特徴領域内に属するか否かという事と、誤認識文字集合による整合計算の効率化の 2 点に注目する。そしてこれらを利用した特徴領域整合法を提案し、実験検討を行う。

4.2 実験の目的

未知入力文字の特徴量が、第 1 位候補の特徴領域内にあるかどうか調べることで、誤認識文字のうち約 56% 以上が領域外であることが表 3.7 より明らかになった。つまり、回復率 (誤認識文字のうち特徴領域外であった文字数の割り合い) 約 56% 以上ということは、誤認識文字のうち約 56% 以上の文字が第 1 位候補が間違っているという保証を与えることができ、第 2 位以下に正解があるということである。

本章は、命題 1 を基に第 2 位以下の候補についても (2.8) 式で定義された領域半径を適用し、特徴領域内に属するか否かを調べ認識を行う。これは従来の全数整合法に比べどれ程の効果があるかを検討する。また、パターン整合法は全字種について評価値を計算するが、ここではある文字 k で一度領域半径内に入ることが確認されたら、以後は文字 k の誤認識文字集合に属する文字を対象に評価値を計算することで良い。このように計算対象を絞り込むことで効率化が期待できる。予め (2.11) 式で定義される誤認識文字集合を用意し、評価値計算の対象とした文字数についても検討し考察する。

第 3 章の実験はクローズ実験であるが、本章ではさらにオープン実験について一連の内容を行うことにより、どれだけ標本領域半径が妥当であるかについても検討し考察する。

4.3 実験方法

使用する文字パターンを表 4.1 に示す。第 3 章と同様に潰れのない本来の文字パターンを対象とするために、4 倍角及び縦倍角で印字したものを対象とする。実験では、JIS 第 1 水準漢字 2965 字種のパターンを 3 種類のドットプリンタと 1 台のレーザープリンタを用いて印字したものを、辞書用及びクローズ実験用データとして 4 倍角 7 セット、縦倍角 3 セット、計 10 セット用意し、オープン実験用データとして 4 倍角 3 セット、縦倍角 1 セット、計 4 セット用意した。辞書用及びクローズ実験用は第 3 章で使用した表 3.1 の乙と同じである。

特徴量は第 3 章の結果より評価された方向線素特徴量を使用し、距離尺度は、認識率が従来から比較的良いと言われるユークリッド距離と回復率が良いシティーブロック距離の 2 種類を用いる。辞書は 10 セットの平均を標準特徴量とする重心ベクトルとした。

以上を用いて 4.1 節を踏まえた特徴領域整合法 (詳細は 4.4 節) を適用する。比較のために全数整合法も同時に併用しクローズ実験とオープン実験の両方を行う。

表 4.1: 実験用文字データ

辞書作成用及びクローズ実験用			
1	エプソンプリンター印刷文字	4 倍角	No. 1
2	エプソンプリンター印刷文字	4 倍角	No. 2
3	富士通プリンター印刷文字	4 倍角	No. 1
4	富士通プリンター印刷文字	4 倍角	No. 2
5	富士通プリンター印刷文字	縦倍角	No. 1
6	キャノンレーザープリンター印刷文字	4 倍角	No. 1
7	キャノンレーザープリンター印刷文字	4 倍角	No. 2
8	キャノンレーザープリンター印刷文字	縦倍角	No. 1
9	NEC プリンター印刷文字	4 倍角	No. 1
10	NEC プリンター印刷文字	縦倍角	No. 1

オープン実験用			
1	エプソンプリンター印刷文字	4 倍角	No. 1 のコピー
2	富士通プリンター印刷文字	4 倍角	No. 2 のコピー
3	キャノンレーザープリンター印刷文字	4 倍角	No. 2 のコピー
4	エプソンプリンター印刷文字	縦倍角	No. 1

4.4 特徴領域整合法

本章の文字認識実験のアルゴリズムに使用する特徴領域整合法を図 4.1 を例に説明する。

$$K = \{\text{亜}, \text{唾}, \dots, k', \dots, k, \dots, j_1, \dots, \text{腕}\}, \quad (k' \leq k)$$

$$\text{Er}(k') = \{k, j_1\}$$

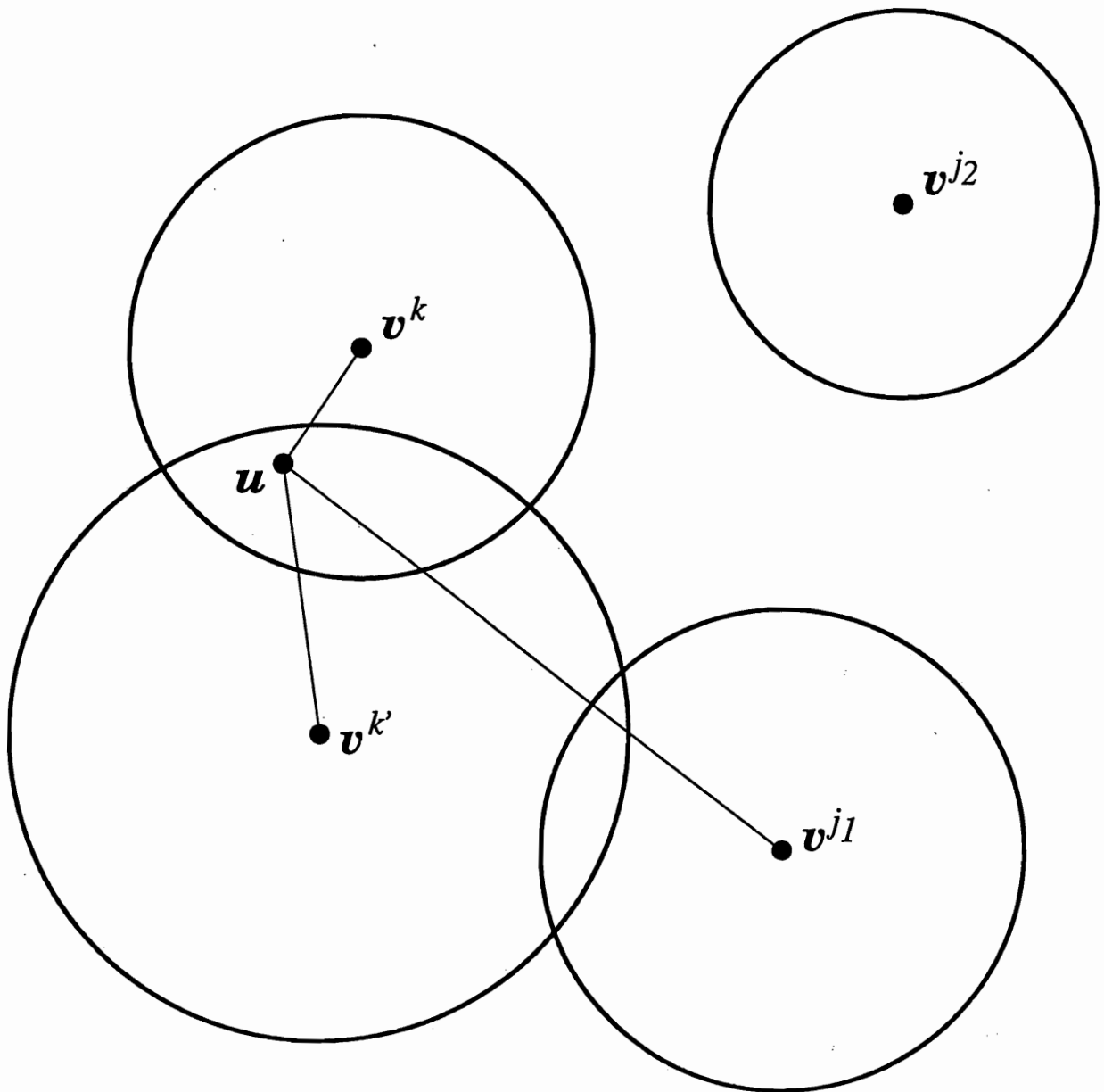


図 4.1: 特徴領域整合法

- 認識実験の前に、あらかじめ各文字について領域半径と誤認識文字集合を求めておく。
- 未知入力文字の特徴量 u と辞書を用い全数整合法で計算し始める。
- u が最初に特徴領域内に属する文字 k' を探す。
- 文字 k' の既に計算の対象とした文字を除いた誤認識文字集合に属する文字 (図 4.1 では k と j_1) を対象に距離値を計算する。
- 候補文字の特徴領域内に u が入り、かつ距離値の最小のものを候補出力する。

計算対象文字数は全数整合法が 2965 字種全てを対象とするのに対し、この特徴領域整合法アルゴリズムは $\{(k' \text{を見つけるまでの文字数}) + |\text{Er}(k')|\}$ 字以下となる。

4.5 実験結果

クローズ実験の結果を表 4.2 に示す。入力データは辞書に用いた 10 セットを使用し、表の数字は 1 セット (JIS 第 1 水準漢字 2965 字種) 当りの平均値である。横の項目には距離尺度がユークリッド距離、シティーブロック距離における結果の順である。計算対象文字数とは、最終候補を出力するまでに辞書を参照した文字数のことで、必然的に全数整合法は辞書に用意された全ての文字を参照する。

オープン実験の結果を表 4.3 に示す。入力データは辞書の他に新たに用意した 4 セットを使用し、表の数字は 1 セット当りの平均値である。判断不能文字数とは、オープン実験なので未知入力ベクトルが辞書にある全ての文字の特徴領域外に位置し候補が定まらなかった文字数のことである。誤認識文字数は、入力した 2965 字種から判断不能文字数を引いた、候補を判定した文字数のうちの誤認識した文字数を表わす。認識率は、その率であることに注意しておく。

表 4.2: 特徴領域整合法によるクローズ実験の結果

距離尺度	ユークリッド距離		シティーブロック距離	
	全数整合法	特徴領域整合法	全数整合法	特徴領域整合法
認識率	99.912%	99.966%	99.885%	99.983%
誤認識文字数	2.60 字	1.00 字	3.40 字	0.50 字
計算対象文字数	2965 字	1562.69 字	2965 字	1526.61 字

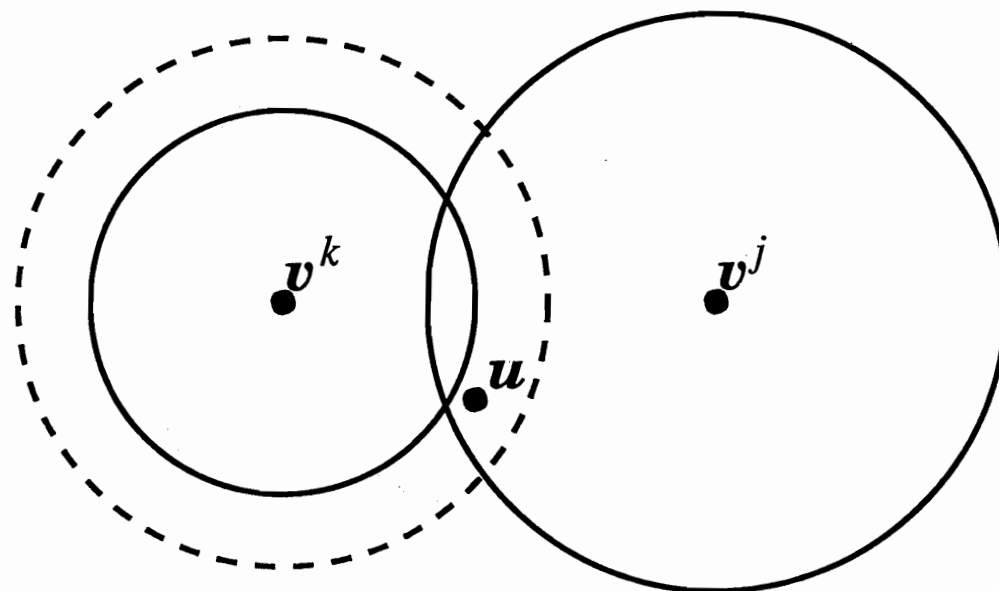
表 4.3: 特徴領域整合法によるオープン実験の結果

距離尺度	ユークリッド距離		シティーブロック距離	
	全数整合法	特徴領域整合法	全数整合法	特徴領域整合法
認識率	99.907%	98.915%	99.889%	99.868%
誤認識文字数	2.75 字	29.25 字	3.25 字	3.50 字
判断不能文字数	—	269.00 字	—	307.25 字
計算対象文字数	2965 字	1707.92 字	2965 字	1680.27 字

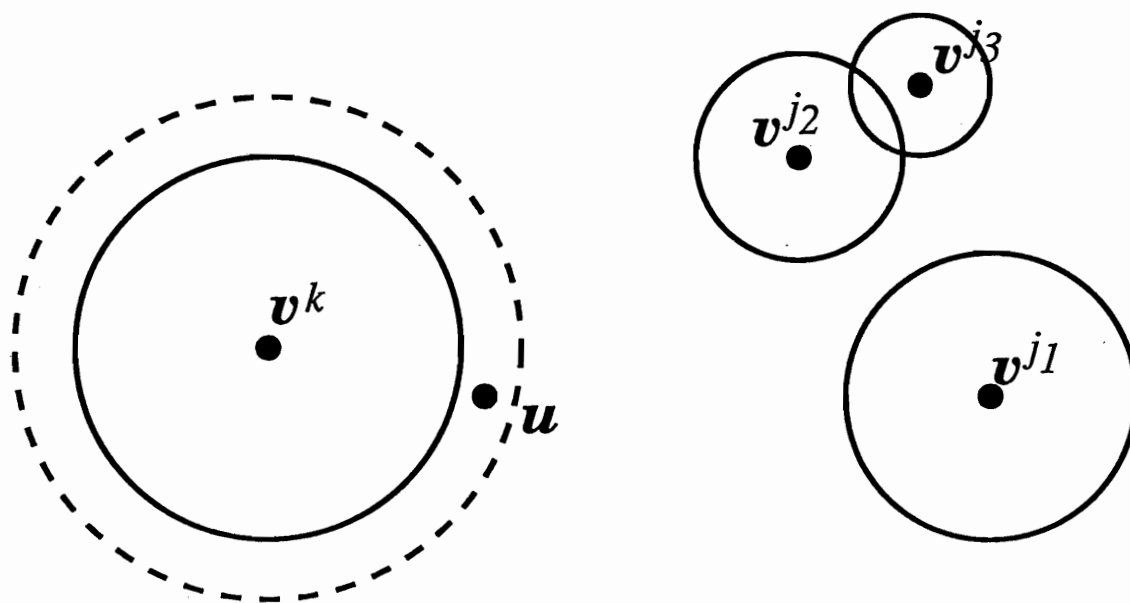
4.6 考察

クローズ実験では、全数整合法より特徴領域整合法が認識率が良い。特に、シティーブロック距離の例で言うと誤認識した 3.4 文字の 85.3% に当たる 2.9 文字が、認識した候補の特徴領域内に存在していないことから領域半径を利用したことが非常に有効だったことがわかる。以前から言われているように、全数整合法で認識率を比較するとシティーブロック距離よりもユークリッド距離の方が認識率は良いのだが、本章の実験で提案した特徴領域整合法で認識を行うと、シティーブロック距離の認識率の方が良い。これは、回復率がユークリッド距離は 61.5% であるのに対し、シティーブロック距離が 85.3% であり、領域半径を登録しておくことで、シティーブロック距離には特に有効に働くためである。評価値計算の対象とした文字数については、全数整合法はアルゴリズムから 2965 字種全てについて対象とするが、特徴領域整合法で行うことによりユークリッド距離では 47.3%、シティーブロック距離では 48.5% 計算対象文字数を減らすことができた。総合的には特徴領域整合法とシティーブロック距離を用いる組み合わせが認識率と距離尺度の定義と評価値計算の対象文字の減少率から一番評価できる組み合わせと言える。

オープン実験では、特徴領域整合法より全数整合法が認識率がよい。この原因の第 1 は、領域半径の求め方にあると考えられる。今回の実験では領域半径 r_k の代わりに (3.1) 式の標本領域半径 r_k^S (図 3.1) を使用しているのだが、データに形の整った 4 倍角を多く使用しているために $\frac{r_k^S}{r_k}$ が、予想よりも小さい値になっていると考えられる。そのために、図 4.2(a) のように正しく認識されるべき文字の特徴領域内に、未知入力ベクトル u が存在しないことが比較的多くなっている。それに、図 4.2(b) のように文字であるか否かの判定が厳しくなり、判断不能文字数が多い結果となっていることが分かる。距離尺度別では、全数整合法でユークリッド距離、特徴領域整合法ではシティーブロック距離で認識率が良く、クローズ実験と同様の結果となる。評価値計算の対象とした文字数については、特徴領域整合法で行うことによりユークリッド距離では 42.4%、シティーブロック距離では 43.3% 計算対象文字数を減らすことができ、クローズ実験と同程度となり、今後高速化に期待できそうである。



(a) 認識結果 $\rightarrow j$ (誤認識)



(b) 認識結果 \rightarrow 判断不能

図 4.2: 特徴領域整合法による認識率低下の原因



4.7 後書き

距離尺度を用いたパターン整合法における文字認識に焦点を当て、その認識結果の正しさの保証を与えるために特徴量の空間の性質を利用しながら、認識実験を行った。従来特徴量空間の分析は認識手法の分類で言えば、厳密な分割 ($\forall k, E_r(k) = \phi$) を求めるパターン空間分析法^[8] であるのに対し、ここでは特徴量空間の厳密な分割は諦め、重複を許しながら明らかにしていくことに特徴がある。

実験により、従来のパターン整合法に各文字に関して領域半径や誤認識文字集合の情報を与えておくことで、より効率的に候補出力を求めることができたと考えている。しかし、命題1を保証する方法に関しては、もう少し検討することが必要であり今後の課題である。

今後の課題の具体的方法として、命題1を保証するために改良案を3つあげる。

まず第1の改良は、標本領域半径の値の求め方である。本章の実験では領域半径として、サンプルパターンの特徴量の重心から最遠のサンプルパターンの特徴量までの距離を用いたが、今後は標準偏差などを考慮しながら領域半径を吟味し、真の値に近似する必要がある。調整された r_k は、比較的形の整った4倍角程度の入力データの場合常に何らかの特徴領域内に存在し、変形の著しい入力データはそのことを積極的に指示していく必要がある。この標本領域半径の検討については第5章で行う。

第2の改良は、マルチテンプレート化の利用である。本章の実験では文字の標準特徴量として1つだけを考えてが、各文字に複数の標準特徴量を用意するマルチテンプレート法もよく知られた方法である。この方法は、特徴領域を1つの超球または正多面体として把握する考えに対して、複数の超球または正多面体で覆われたものとして考えるということであり、領域半径に相当するパラメータの値が小さくなり、安全認識領域をより大きくし、誤認識文字集合を小さくする効果があることが期待される。

第3の改良は、重み付きの距離尺度の導入である。これは特徴領域を超球型または正多面体型として把握する考えに対して、特徴領域を超楕円体型または多面体型とすることに相当する。領域半径に相当するパラメータの数は増えるが、各文字のサンプルを効率よく包囲することができる。これも安全認識領域をより大きくし、誤認識文字集合を小さくする効果があることが期待される。

これらの3つの改良案についてはそれぞれ第5、6、7章で検討を行う。



第 5 章

標本領域半径の改善

5.1 前書き

本章は第 4 章の第 1 改良案である標本領域半径の値の求め方の検討について述べる。

まず実験の目的では、第 4 章までの標本領域半径についての問題点をあげ、新しく改善する標本領域半径にはどのようなことに着目しながら大きさを定める必要があるかについて、いくつかの項目をあげる。次に標本領域半径について新しく定義をし、実験では大きさを变化させながら各項目に最適な標本領域半径について調べる。最後に総合的な評価により最適な標本領域半径を定め、最終的には超球 (正多面体) 型の特徴領域について考察をする。

5.2 実験の目的

第 4 章までは、標本領域半径としてサンプルパターンの特徴量の重心から最遠のサンプルパターンの特徴量までの距離とし、それを領域半径と見なした。これは文字 k のサンプル集合 $S(k)$ が文字 k の文字パターン集合 $K(k)$ を適切に反映していると期待したものであり、第 3 章ではある程度適切な領域半径として役目を果たした。しかし第 4 章で実際にパターン整合法に適用すると、真の領域半径から見た標本領域半径の小ささが浮き彫りとなり、結果として図 4.2 ような誤認識や判断不能文字が比較的多く出力された。

これを補正する方法は、単に以前よりも標本領域半径を大きめにすることである。今までの標本領域半径を少し大きくすることで、第 4 章のオープン実験の結果より向上することは予想される。しかし、標本領域半径を大きくしすぎると全数整合法の実験結果にだんだん近付き、最終的には同じ結果になる。これは特徴領域整合法としての利用価値を失っていると言える。

本章では、標本領域半径としてサンプル集合の標準偏差を用いて新たに定義し、それを実数倍し大きさを变化させることにより、どの位の大きさが標本領域半径として相応しいか検討する。標本領域半径に対する大きさの検討内容としては、

- 命題 1 を保証するにはどの位の大きさが必要か。



- 認識率に関して最適な領域半径はどの位の大きさか。
- 計算対象文字数を最小にするにはどの位の大きさか。
- 判断不能文字数は標本領域半径を大きくするに伴いどのような減少傾向をみせるか。

などがあげられる。

これら全てに関し最適な状態を満たす標本領域半径が存在しないことは予想される。最終的には、これらの項目を総合的に判断した結果より標本領域半径を定め、以後の実験などに利用する考えである。

5.3 実験の準備

実験に用いる文字パターンは第 4 章に使用した表 4.1 のものを、辞書用、オープン実験の入力用共に同様に本章でも使用する。また特徴量、距離尺度についてはそれぞれ第 3、4 章で評価の良かった方向線素特徴量とシティーブロック距離を使用する。辞書も 10 セットの平均を標準特徴量とする重心ベクトルとする。

本章以降は第 4 章のパターン整合法の結果に対する比較検討を行い、容易に向上が理解できるように、文字パターン、特徴抽出法、距離尺度、辞書作成に関しては上記のようになるべく統一して実験検討していくことにする。

5.3.1 標本領域半径

本章で使用する新標本領域半径 r_k^S を以下のように定義する。

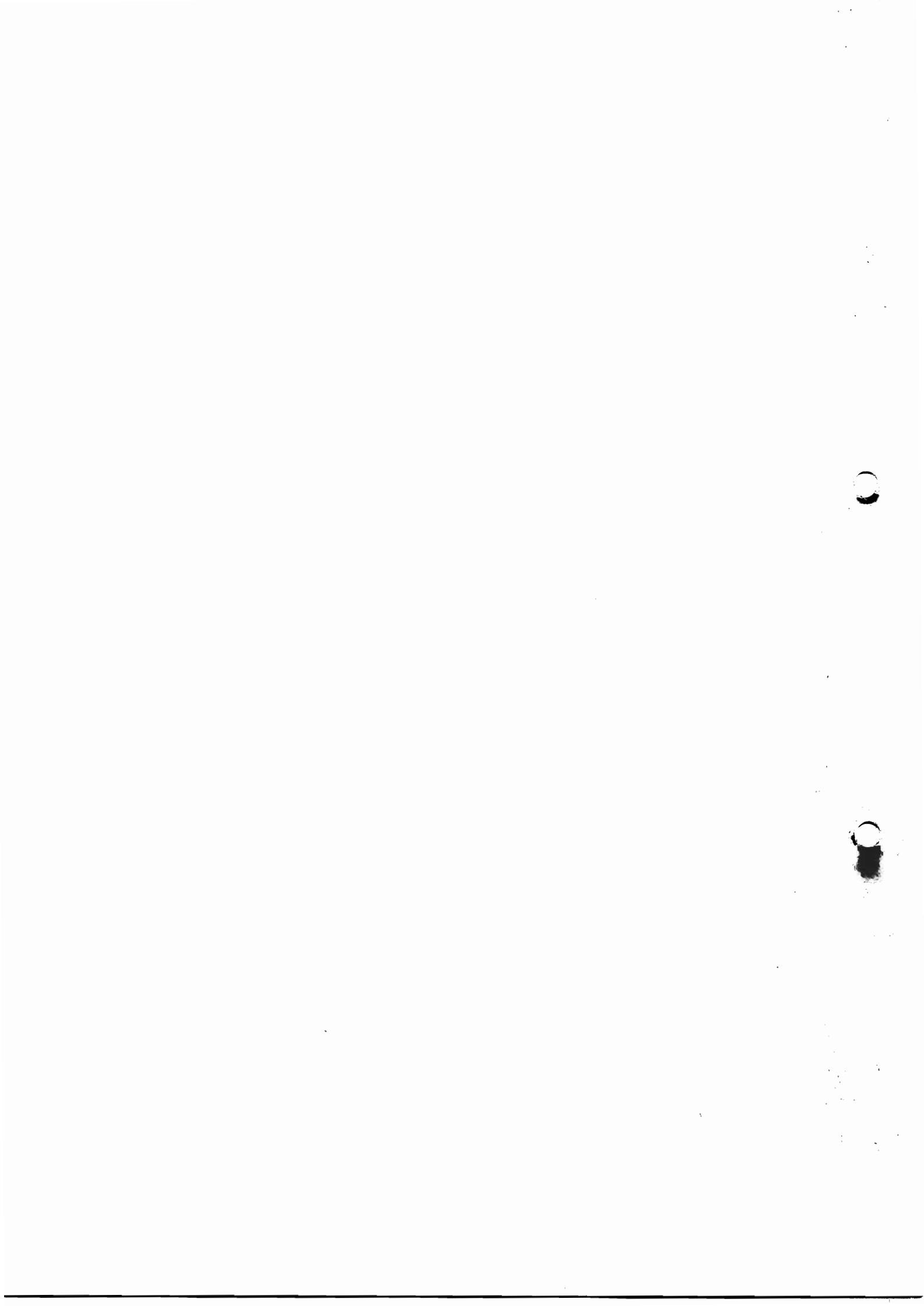
定義 8 (標本領域半径)

$$\sigma_k = \sqrt{\frac{1}{s} \sum_{i=1}^s \left\{ E(F(P_i^k), v^k) \right\}^2} \quad (5.1)$$

$$r_k^S = r_r \sigma_k \quad (5.2)$$

特徴量空間内に文字 k のサンプルパターンの特徴量 $F(P)$ が、図 5.1 のように分布した場合の新標本領域半径を説明する。(5.1) 式の σ_k は、文字 k の辞書 v^k から各サンプルパターンの特徴量 $F(P)$ までの距離値に対して、 v^k を距離値の平均とみなしたときの標準偏差である。(5.2) 式により、この σ_k の r_r 倍を新標本領域半径 r_k^S とする。 r_r を領域半径標準測度と呼ぶ。また、新標本領域半径を以後は単に標本領域半径と呼ぶことにする。

本章の実験では領域半径標準測度 r_r を可変することにより、標本領域半径の大きさを変化させる。図 5.1 の点線で描かれた超球の半径は内側から $r_r = 0.5, 1.0, 1.5, 2.0$ における標本領域半径 r_k^S で、単に目盛的な意味として書いておく。 r_k^S の例として $r_r = 1.3$ のときにできる超球を実線で描いてある。この大きさは、(3.1) 式で定められる標本領域半径の大きさと概ね同じである。厳密には定義が違うので、文字によってそれぞれの標本領域半径には大小の差位がある。



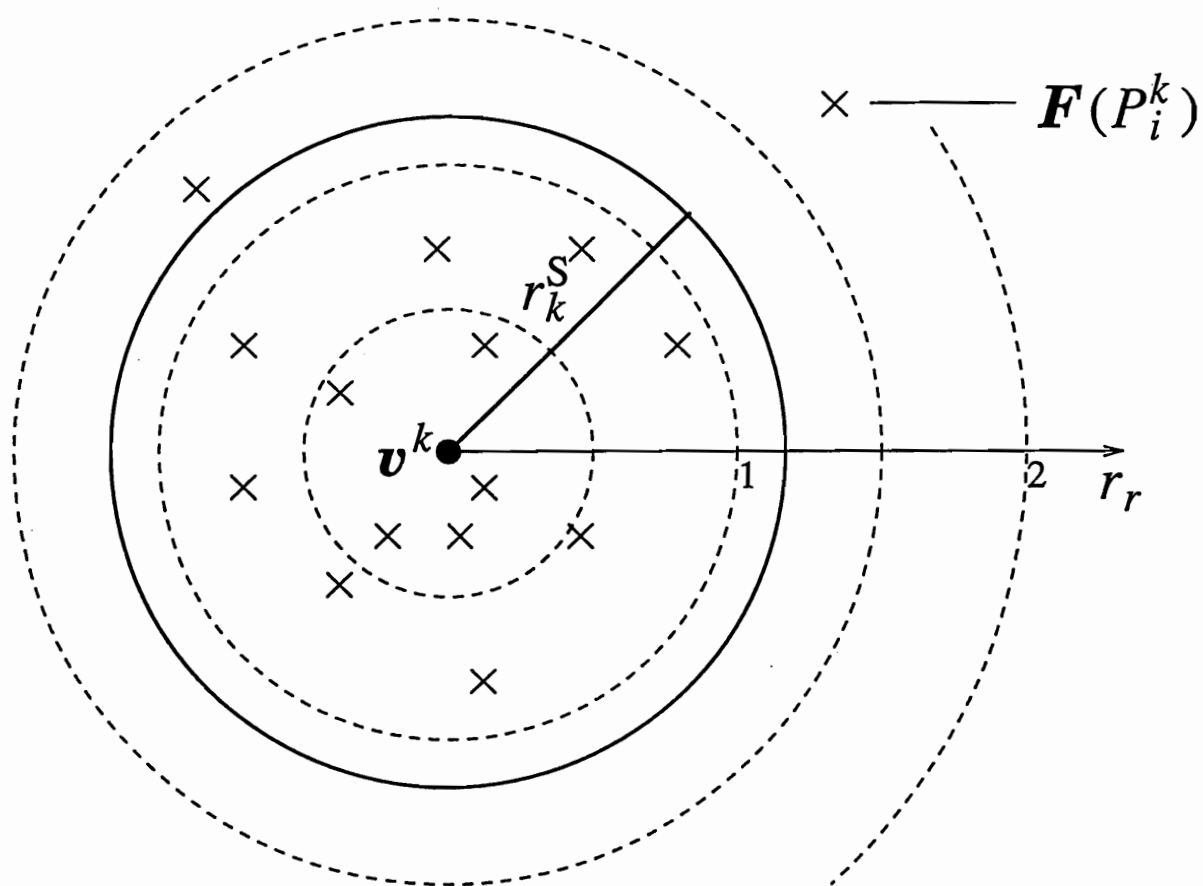


図 5.1: 標本領域半径の改善



5.4 実験の方法

実験の方法は第 4 章と同様に特徴領域整合法を用いてパターン整合法を行う。但し、オープン実験のみ行い、クローズ実験は本章以降行わない。

領域半径標準測度 r_r は 1.0 から 3.0 まで 0.1 刻みに離散的に大きさを変え、計 21 回のパターン整合法を行う。1 回のパターン整合法について以下の項目について調べる。理想的には無数のサンプル文字集合を用いて r_r を連続的に大きさを変化させた結果が最適なのだが、本章の実験の検討内容から 0.1 刻みの r_r で十分と考えられる。参考までに誤認識文字集合の大きさも項目の末尾に加えておく。

- 認識率
- 誤認識文字数
- 判断不能文字数
- 計算対象文字数
- 誤認識文字集合の大きさ

各文字数に関しては同時に全文字種に対する率も求める。それぞれの検討の意味は実験の目的で述べた通りである。

5.5 実験結果と考察

領域半径標準測度と認識率、誤認識文字数の関係を図 5.2 に示す。横軸に領域半径標準測度 r_r 、それに対する左縦軸が認識率、右縦軸が 1 セット当り (全 2965 字種) の誤認識文字数である。右縦軸の誤認識文字数 1 (文字/セット) の幅に対し、左縦軸の幅を全 2965 字種に対する 1 文字の率 (約 0.034%) としてある。○印が認識率、△印が誤認識文字数を表わしているが、○印と△印が左右の縦軸の縮尺を同様にしているにもかかわらず重ならない場合が存在する理由は、第 4 章と同じく、誤認識文字数には入力した全 2965 字種から判断不能文字数を引いた、候補を判定した文字数のうちの誤認識した文字数を表し、認識率はその率であるためである。つまり、全 2965 字種から判断不能文字数を引いた文字を対象に認識率と誤認識文字数を求めているためである。

図 5.2 より $r_r \geq 2.8$ になると表 4.3 の全数整合法と同様の結果 (誤認識数が 3.25 字) となる。これは特徴領域整合法の効果が殆ど無く、標本領域半径が大きすぎることを意味する。図 5.2 より全数整合法の誤認識文字数 3.25 字よりも少なくなるのは、 $r_r < 1.7$ であることが分る。このときに特徴領域整合法の効果が表われていると言える。特に $r_r = 1.2$ のとき認識率が最高となる。

次に領域半径標準測度と判断不能文字数、計算対象文字数、誤認識文字集合の大きさの関係を図 5.3 に示す。横軸に領域半径標準測度 r_r 、それに対する左縦軸が 1 セット当り (全 2965 字種) の文字数、右縦軸がその率であり左縦軸と対応している。×印が判断不能文字数、◇印が計算対象文字数、▽印が誤認識文字集合の大きさを表している。

図 5.3 より r_r が大きくなるにつれて、全数整合法の結果 (計算対象文字数、誤認識文字集合の大きさは 2965 字、判断不能文字数は 0 字) に近付いていく様子が分る。計算対象文字数は、 $r_r = 1.4$ のとき最も効率良く候補出力することができ、うまく標本領域半径を定めれば最大で 50% 近くまで減少できることが分った。判断不能文字数は $r_r \geq 1.5$ のとき大幅な減少が殆ど無くなり、 $r_r \geq 2.4$ のときは完全に判断不能文字数が無くなった。図 5.2 で $r_r = 1.2$ のとき認識率が一番良かったが、図 5.3 のときの $r_r = 1.2$ を調べると、判断不能文字数を約 500 字出力し候補出力の約 20% を占めていることから、標本領域半径が小さすぎることを示している。また、誤認識文字集合の大きさから (3.1) 式定義の旧標本領域半径が r_r が約 1.3 付近であることが分った。

命題 1 を保証する標本領域半径はどの位の大きさが必要かについて、認識結果の文字種と図 5.2、5.3 を使って調べると r_r が 2.6 を越える程度が必要なことが分った。また、認識率の精度と効率的に候補出力することを示す計算対象文字数及び判断不能文字数の少なさから総合的に判断して $r_r = 1.5$ が適当な大きさと言える。

最後に $r_r = 1.5$ とする新標本領域半径を用いた特徴領域整合法と第 4 章の旧標本領域半径を用いた特徴領域整合法のオープン実験の結果を比べて見る。認識率と誤認識文字数の結果は図 8.1 に示す。この図は第 4 章以降のオープン実験において、章を追うごとにどのように向上していくかを一目瞭然にするために本章以降利用する。横軸は各章で行われるオープン実験について、そのときの章で表し、1 つの章で複数の実験が行われた場合は、節で表す。例として、横軸の 6.4 は第 6 章の第 4 節のオープン実験の結果を表している。また 0 は全数整合法によるオープン実験の結果である。横軸の 0 及び 4 は表 4.3 の全数整合法、特徴領域整合法の結果を表している。本章の $r_r = 1.5$ とする新標本領域半径を用いた特徴領域整合法の結果は、当然横軸の 5 を参照すれば良い。図 8.1 より、誤認識文字数 (Δ 印) が 1.50 字となり、第 4 章に比べて大幅に認識率 (\circ 印) が上昇したことが分る。また第 4 章の特徴領域整合法は全数整合法によるオープン実験よりも認識率の精度が落ちた。これは旧標本領域半径に原因があった。しかし本章の提案した標本領域半径の改善により、全数整合法の結果 (横軸の 0) を認識率が大きく上回るすることができた。このことは、新標本領域半径の定義が文字パターン集合 $K(k)$ をより効果的に包囲していることを意味し、旧標本領域半径からの改善がある程度成功したことを示している。

判断不能文字数と誤認識文字集合の大きさについては図 8.2、計算対象文字数については図 8.3 に同様の結果を示す。本章の $r_r = 1.5$ とする新標本領域半径を用いた特徴領域整合法により、判断不能文字数 (\times 印) が 31.50 字と大幅に縮小することができたため、何らかの出力結果が多く得られるようになった。また計算対象文字数 (\diamond 印) から第 4 章の実験結果とあまり変化させることなく*、効率的に候補出力することを維持したまま認識率がより高精度になったことが分る。

*厳密には少し減少することができた。

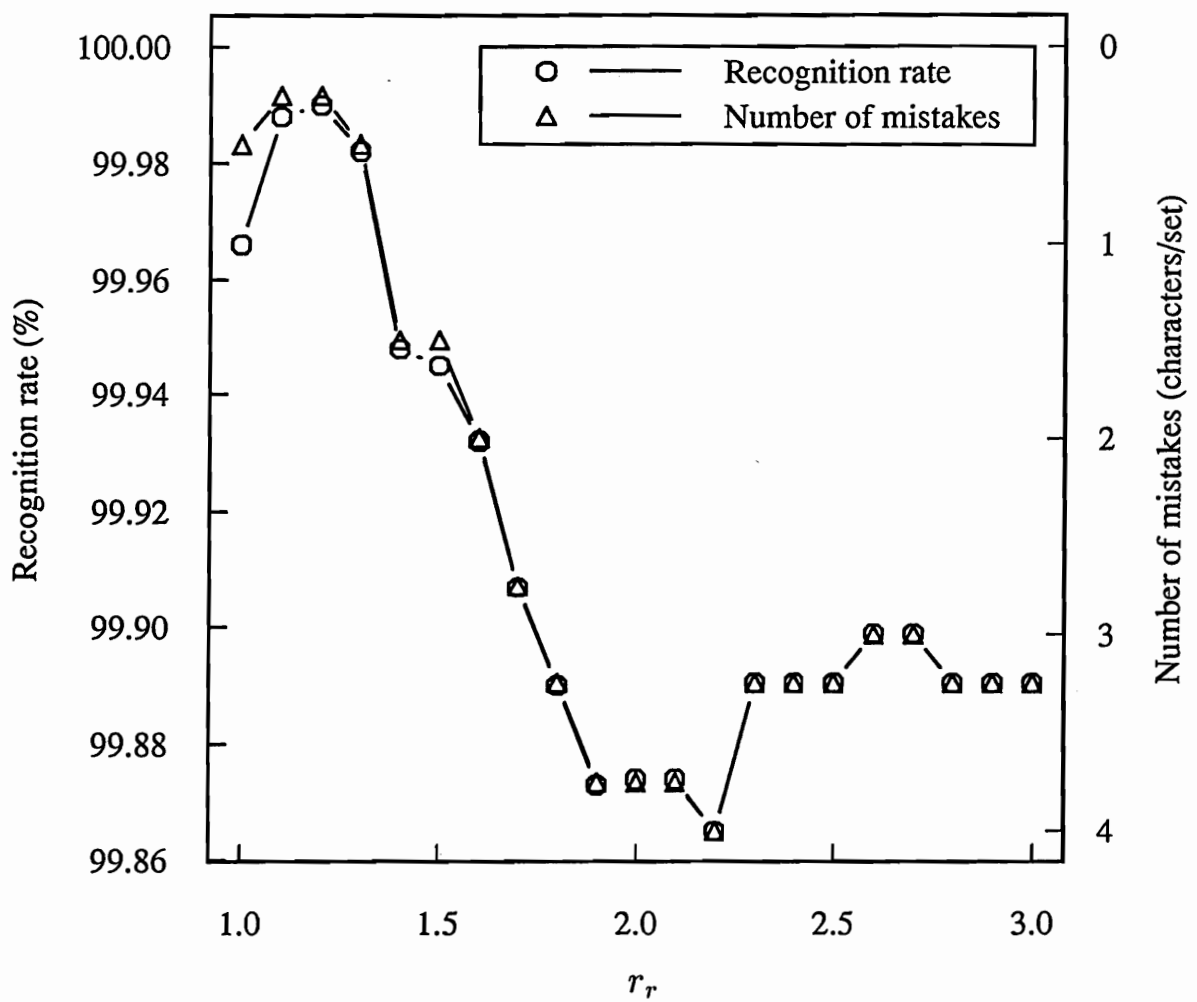


図 5.2: 領域半径標準測度と認識率、誤認識文字数の関係

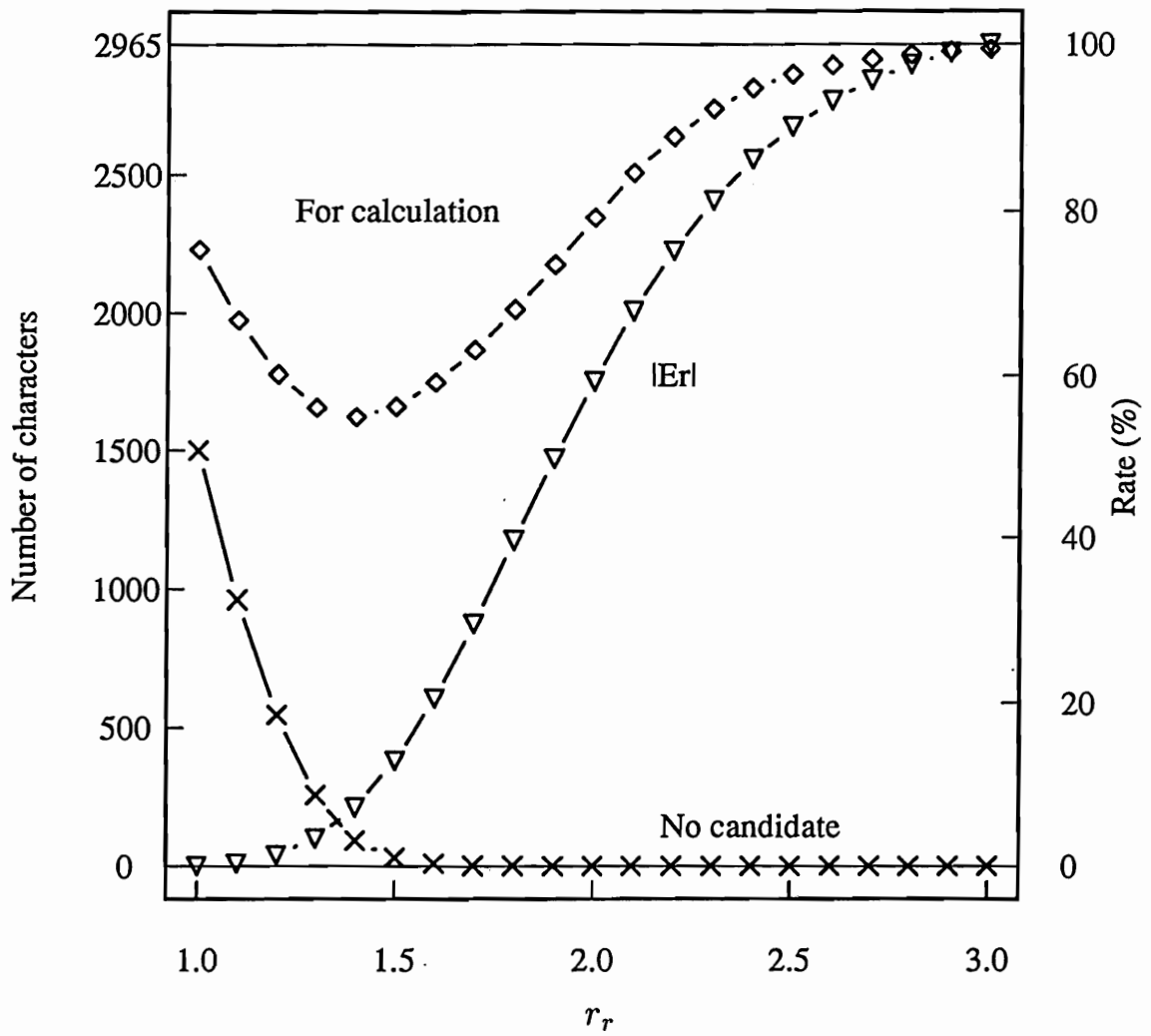


図 5.3: 領域半径標準測度と判断不能及び計算対象文字数、誤認識文字集合の大きさの関係



5.6 後書き

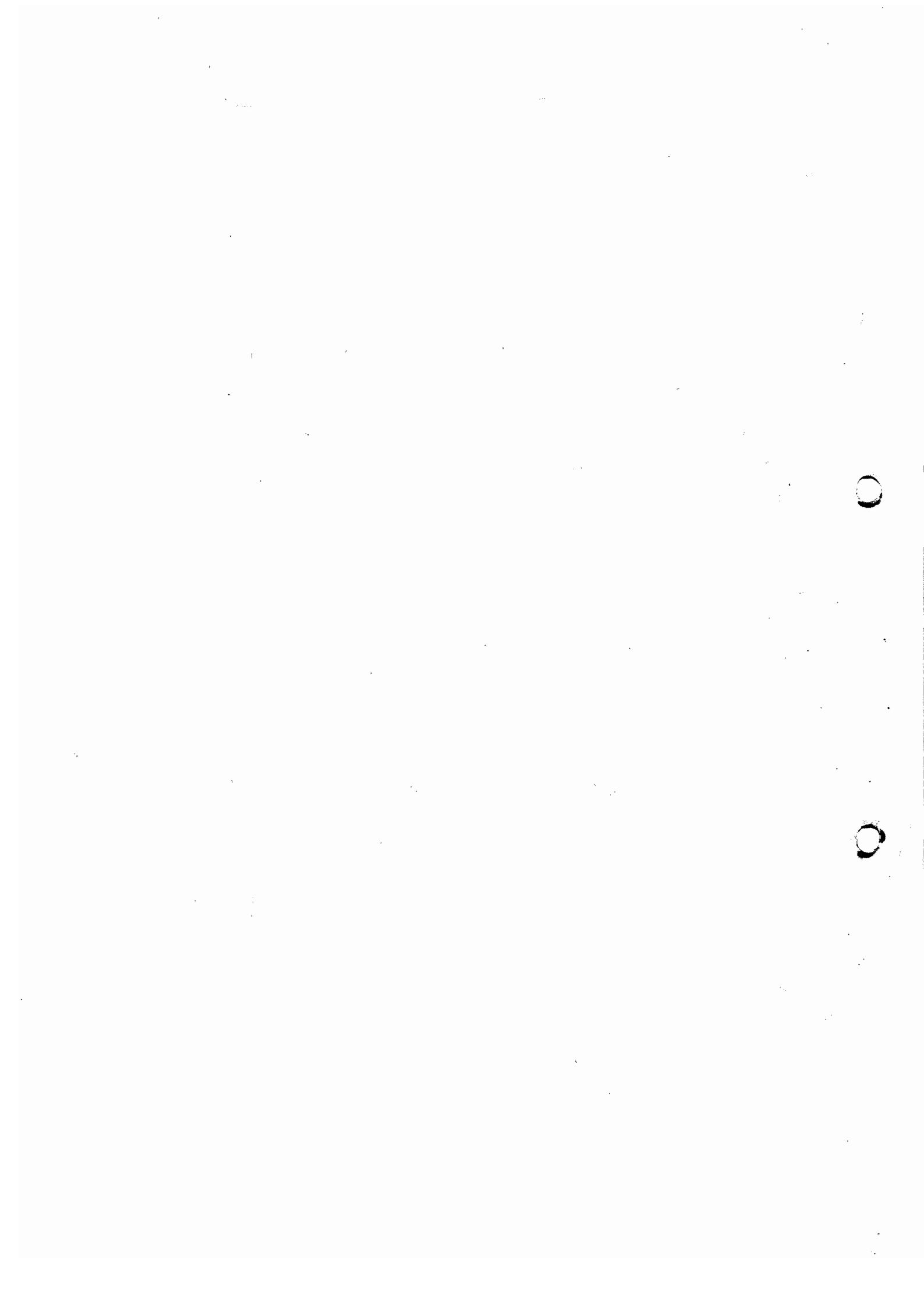
本章は第 4 章の第 1 改良案である標本領域半径の値の求め方について、旧標本領域半径の問題点をあげ、新標本領域半径にはどのようなことに着目しながら大きさを定める必要があるかについて述べ、実験を行った。

その結果、命題 1 を保証する標本領域半径は r_r が 2.6 を越える程度が必要であるが、候補出力のための総合的な評価から考えると $r_r = 1.5$ が十分に適当な大きさであると結論を下した。

これらの領域半径標準測度 r_r の 2.6 と 1.5 の差は、超球 (正多面体) 型特徴領域による整合法の限界を示唆していると考えられる。最終的目標である候補出力の文字に対し正確な保証を与えるには、この差があり続ける限り達成できない。超球 (正多面体) 型特徴領域であると N 次元のパターン特徴量空間の全ての軸方向を包囲し、図 8.2 の第 4、5 章の結果や、図 5.3 に見られるように判断不能文字数 (×印) が多くなるかまたは誤認識文字集合の大きさ (▽印) が大きくなるかのどちらかになる。これは、文字パターン集合 $K(k)$ が超球 (正多面体) 型でないために、1 つの超球 (正多面体) 型の特徴領域で囲もうとするとどうしても無理が出てしまうためである。そのためにこの差が存在すると考えられる。今後の課題としてはこの差を縮めることである。この差が無くなることは判断不能を除く候補を効率的に定めることができると同時に、正確な保証をある程度与えることができることを意味している。そのためには今までの 1 文字に対し 1 つの超球 (正多面体) 型特徴領域という考え方に対して、必然的に変更が要求される。

容易に考えられる変更の方法は、第 4 章の今後の課題でも述べたように、1 文字に対し複数の特徴領域にする方法と特徴領域を変形する方法である。前者はマルチテンプレート化を応用することであり、標準特徴量を複数にする意味がある。この検討は第 6 章で行う。後者は各文字のサンプルパターンの特徴量の各成分の標準偏差を利用することで超球 (正多面体) 型から超楕円 (多面体) 型へ特徴領域を変形する方法である。この検討は第 7 章で行う。

これらの例を図 5.4 に示す。文字 k のサンプルパターンの特徴量が図 5.4 のように位置したとき、図上部が本章の実験における特徴領域であり、シティーブロック距離を距離尺度としたときの特徴領域の変更前の例である。図下部矢印先がそれぞれの特徴領域の変更後の例である。左がマルチテンプレート化を応用し複数の特徴領域を用いてサンプルパターンの特徴量を包囲する例である。右が 2^N 面体型特徴領域によるサンプルパターンの特徴量を包囲する例である。これらの方法は、各文字のサンプルパターンの特徴量のさらなる効果的な包囲が期待される。



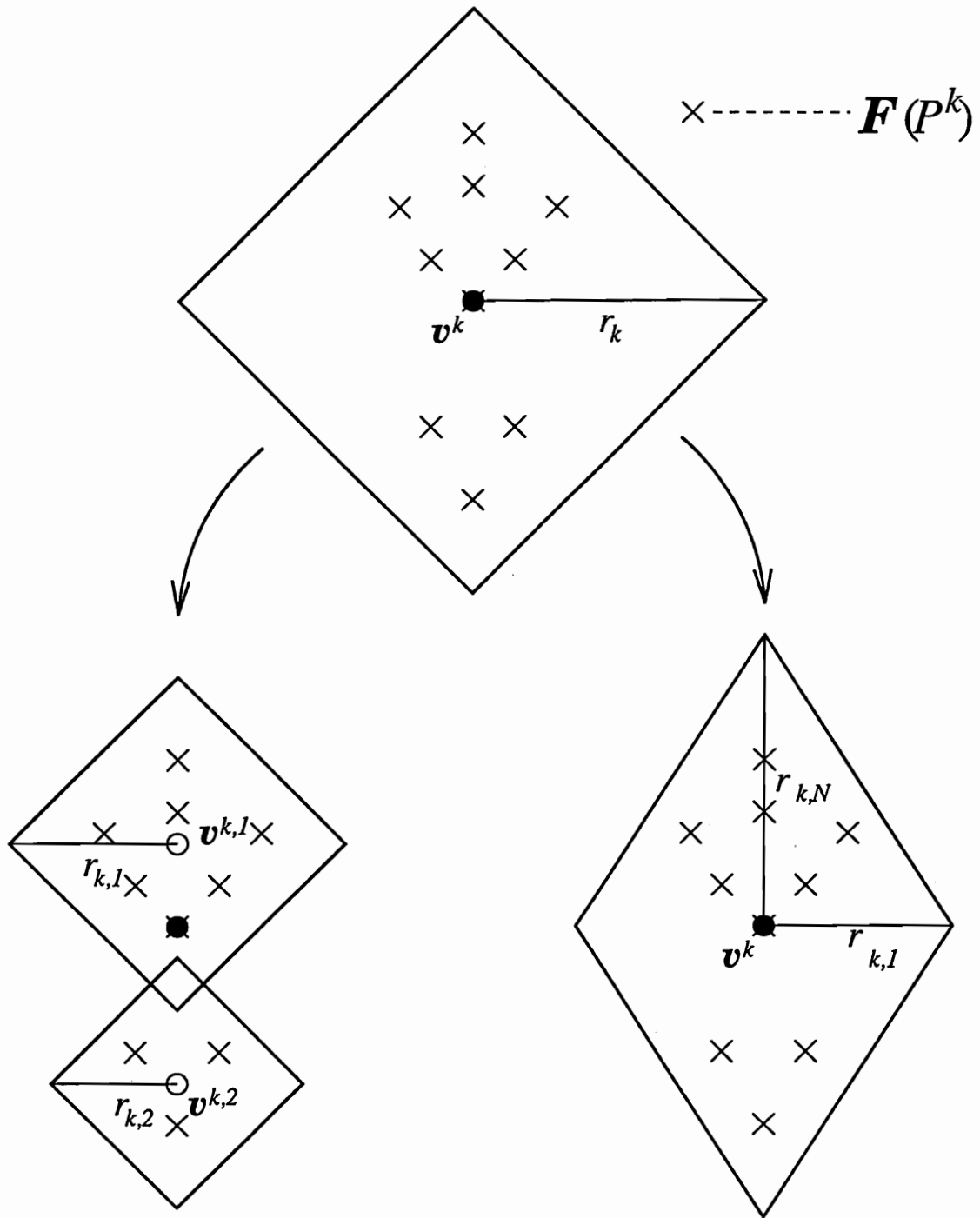


図 5.4: 特徴領域の変更前と変更後の例



第 6 章

多特徴領域によるパターン整合法

6.1 前書き

本章は第 4 章の第 2 改良案であるマルチテンプレート化の適用について述べる。

まず実験の目的では、マルチテンプレート化を適用することは、パターン特徴量空間においてどのような意味があるかについて述べる。そしてそれを応用することでどのような効果が期待されるかについて、その項目を列記する。実験よりその効果について確かめ、結果について考察を行う。

6.2 実験の目的

第 5 章の結論として、正多面体 (以後超球と書かない) 型特徴領域のまま特徴領域整合法を行うことは、最終目的である認識結果に正しさの保証を与えることに対し、第 5 章の結果以上の効果が望めないことを述べた。そこで容易に考えられる特徴領域の変更方法の 1 つとしてマルチテンプレート化の利用を提案した。

マルチテンプレート化を利用することは、図 5.4 左に示すようにパターン特徴量空間内では 1 文字に対し 1 つの特徴領域であったものが、1 文字に対し複数の特徴領域を割り当てることを意味する。

多特徴領域に変更することにより、特徴領域を表すパラメータはテンプレートの数と同じになり増加する反面次のような効果が期待される。

- 領域半径*の縮小
- 誤認識文字集合の縮小
- 特徴量空間内におけるサンプルパターンの特徴量の効果的包囲

これらはいずれも、認識結果に正しさの保証を与えるための条件であることは第 3 章でも述べている。特徴量空間内におけるサンプルパターンの特徴量が効果的に包囲できれば、特徴領域整合法によるパターン整合法を行うことにより、認識率の高精度化に繋がる。

*定義 8 の標本領域半径のことである。以後も単に領域半径と呼ぶ。

○

○

6.3 実験の方法

使用する文字パターン、特徴抽出法、距離尺度は第 4 章以降のオープン実験と比較できるように、本章においてもそれぞれ同様に表 4.1、方向線素特徴量、シティーブロック距離を使用する。

マルチテンプレートの実験に先立ち、まずマルチテンプレート化の対象になる文字の選定を行わなければならない。本章では文字 k の領域半径 r_k がある閾値 r_θ より大きいとき、文字 k をマルチテンプレート化する。この理由は、領域半径はサンプルの散らばりの度合を含むので、領域半径が大きい程マルチテンプレート化による効果が大きくなると考えられるためである。

実験は 2 つに大別する。第 1 の実験では、 r_θ とテンプレート数の最大値を決定するために、まず単純に全ての字種においてテンプレート数を 2 に定める。これは辞書作成において 1 文字あたり 2 つの標準特徴量を用意をすることである。2 つの標準特徴量の選出の仕方は、各文字のサンプルパターンを K -means 法^{[9][10]} を用いて 2 つにクラスタリングし、それぞれのクラスタ (テンプレート) の重心ベクトルを標準特徴量とする。各クラスタにはそれぞれに領域半径を定める。領域半径は第 5 章の結果より (5.2) 式で $r_r = 1.5$ としている。通常より 2 倍の数になった辞書と領域半径を用いて特徴領域整合法でオープン実験を行う。比較のために全数整合法によるオープン実験も行う。その結果から r_θ とテンプレート数の最大値を決定し、それに基き第 2 の実験としてテンプレート数可変による特徴領域整合法を用いたオープン実験を行う。同じく比較のために全数整合法によるオープン実験も行う。

6.3.1 K -means 法

サンプルパターンの特徴量は、パターン特徴量空間において多くの場合幾つかの部分的なクラスタ (塊) から形成されているのが普通である。そのクラスタの相対的位置関係や分布がどのようになっているかを解析し分類することを、クラスタリングという。

その代表的なアルゴリズムである K -means 法を、本章の実験で利用する。これは次のような手順から成り立つ。

1. K 個のクラスタの中心位置 (重心ベクトル) の初期値を、任意に選んだ K 個のサンプルパターンの特徴量の位置に置く。
2. すべてのサンプルパターンの特徴量について、どのクラスタの重心ベクトルに近いかで分類する。
3. あるクラスタに属すると判定されたサンプルパターンの特徴量全体の平均ベクトルを求め、これを次の試行のクラスタの中心位置とする。
4. 平均ベクトルの位置が変化しなくなるまで 2. 及び 3. の手順を繰り返す。

この方法は K の取り方に依存するので、種々の K について何度も試みる必要がある。しかし本章の実験では厳密なクラスタリングが目的ではなく、特徴領域を複数にするための利用であるので、 $K = 2$ についての実験のみである。



6.4 テンプレート数 2 による実験結果と考察

最初に領域半径と誤認識文字集合の大きさに関して調べた。領域半径は、第 5 章のシングルテンプレートと比べて、平均で 31.2% 減少できた。その結果 1 文字当り平均の誤認識文字集合の大きさは 40.7 字であった。これを今までの結果と比較すると、図 8.2[†] から大幅に減少していることが分り、マルチテンプレート化による効果を示すことができた。

オープン実験の結果を表 6.1 の中段に示す。比較のために上段にシングルテンプレートの結果を載せておく。これは第 5 章の $r_r = 1.5$ の結果と同じである。テンプレート数を 2 にすることにより、誤認識文字数が 0.75 字となり認識率が向上したことが分る。

また計算対象文字数はテンプレート数が 2 であることから、シングルテンプレートより増加することは否めない事だが、結果的にシングルテンプレートの全数整合法より少ない 2437.06 字であったことは、評価できる結果であると考えられる。全数整合法の計算対象文字数に対する特徴領域整合法の計算対象文字数を考えると、シングルテンプレートが約 56% に対し、テンプレート数 2 の方は約 41% となり、特徴領域整合法がマルチテンプレート化を行うことにより候補出力を求めることに対して非常に有効に働くことが分った。図 8.3 において実線で表されているのが計算対象文字数の実数値 (目盛は左側)、点線で表されているのが先程の率 (目盛は右側) である。この図 8.3 から明らかである。特徴領域整合法において、マルチテンプレート化により効率的に候補が定まる理由は、整合の順番を一通り全字種に対し行うので、全数整合法のテンプレート数分の 1 の段階で未知入力文字 u が最初に特徴領域に属する文字 k' を見つけることが比較的簡単にできるためである。

判断不能文字数については 221.25 字とシングルテンプレートの 31.50 字よりも増加した。これは、1 文字当りのサンプルパターンが 10 セットということで、テンプレート数 2 であると 1 テンプレートにつき平均 5 セットで領域半径を定めていることになり、サンプル数が少なすぎることが主な原因であると思われる。極端な例として 1 つのクラスタに属するサンプルパターンが 1 つだけのものもあり、そのため領域半径が 0 という場合もあった。これは今後セット数を増加することにより、ある程度解消できると考えられる。

最後にテンプレート数可変の実験に使用する閾値 r_θ を決めるために、各文字について、シングルテンプレートの領域半径 r に対するシングルテンプレートの誤認識文字集合の大きさとテンプレート数 2 の誤認識文字集合の大きさの差 (減少数) $\Delta(|Er|)$ の関係を図 6.1 にプロットした。この図 6.1 より、領域半径が比較的小さいとき $\Delta(|Er|)$ があまり文字によって大差なく、0 に近い値をとることが分る。これはマルチテンプレート化による効果があまり無いことを意味している。領域半径が比較的大きくなると、文字によって違いはあるがマルチテンプレート化による効果が大いことが分る。この図 6.1 から r_θ を 600 とした。またテンプレート数の最大値は、サンプルパターンが 10 セットということで 3 以上の実験は理想の特徴領域との隔たりがさらに大きくなることが予想される。故に最大値はそのまま 2 とした。

[†]本節における実験は横軸の 6.4 である。以下同じ。



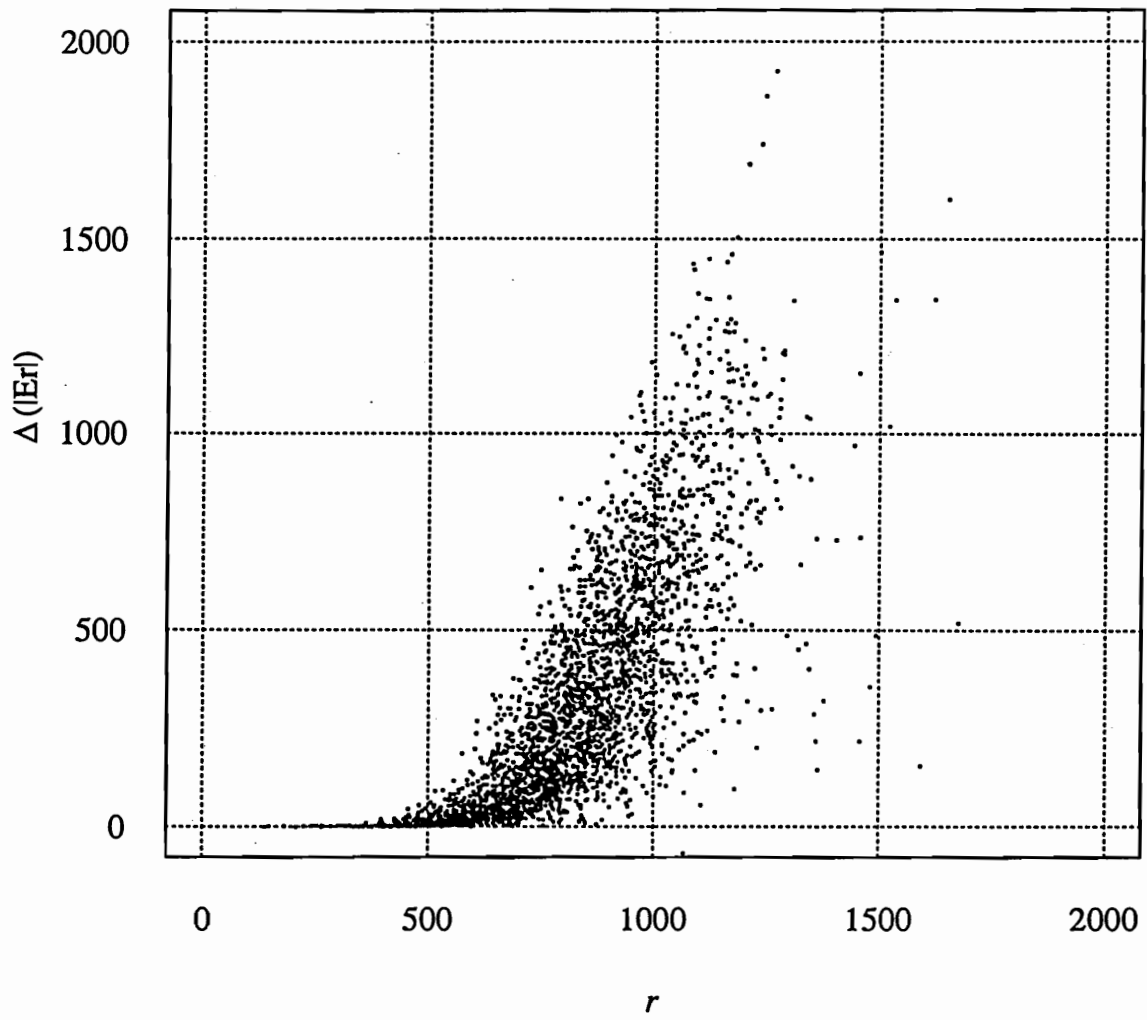


図 6.1: 領域半径とテンプレート数 2 による誤認識文字集合の大きさの変化の関係

0

0

6.5 テンプレート数可変による実験結果と考察

シングルテンプレートの領域半径の段階で r_θ を下回ったのは、378 字種で全体の 12.75%であった。残りの字種はマルチテンプレート化の対象となる。テンプレート数可変によるオープン実験の結果を、表 6.1 下段に示す。認識率を高精度に保ちながら、テンプレート数 2 のときよりも判断不能文字数が 14.58%、計算対象文字数は 7.03%減少させることができた。

図 8.3[†]から分るように、全数整合法の計算対象文字数に対する特徴領域整合法の計算対象文字数の率をテンプレート数 2 とあまり変化させることなく、即ち、効率的に候補を求められるように保ちながら、実際の計算対象文字数を減少させることができた。この結果は今後サンプルパターンの充実しだいではさらに良い効果が期待できることを示唆している。

また図 8.2 から分るように、テンプレート数 2 と比べて誤認識文字集合の大きさを保ちながら、判断不能文字数を減少させることができた。

表 6.1: 多特徴領域によるパターン整合法の結果

シングルテンプレート		
認識方法	全数整合	特徴領域
認識率	99.561%	99.949%
誤認識文字数	3.25 字	1.50 字
判断不能文字数	—	31.50 字
計算対象文字数	2965 字	1661.53 字

テンプレート数 2		
認識方法	全数整合	特徴領域
認識率	99.975%	99.973%
誤認識文字数	0.75 字	0.75 字
判断不能文字数	—	221.25 字
計算対象文字数	5930 字	2437.06 字

テンプレート数可変		
認識方法	全数整合	特徴領域
認識率	99.966%	99.973%
誤認識文字数	1.00 字	0.75 字
判断不能文字数	—	189.00 字
計算対象文字数	5552 字	2265.69 字

[†]本節における実験は横軸の 6.5 である。以下同じ。

0

0

6.6 後書き

本章は第 4 章の第 2 改良案であるマルチテンプレート化の適用について実験を行った。

その結果、多特徴領域に発展することにより、シングルテンプレートの全数整合法と比べて、計算対象文字数を減少させかつ認識率を高められた。またマルチテンプレート化による特徴領域整合法は、その全数整合法に対する計算対象文字数からの減少率が高いことが分った。これはマルチテンプレート化の特徴領域整合法に対する有効性の 1 つであると言える。

最終目標に対して、判断不能文字という結果はある程度許そうというものの、サンプルパターンの不十分さが主な原因で、判断不能文字数が多く存在する結果であったことから、サンプルパターンの充実や特徴領域の改良しだいでは、さらに判断不能文字数を減少させられることを示唆しているように思われる。

0

0

第 7 章

N 次元 2^N 面体型特徴領域によるパターン 整合法

7.1 前書き

本章は第 4 章の第 3 改良案である N 次元 (以後 N 次元は省略する) 2^N 面体型特徴領域によるパターン整合法について述べる。但し、 N はパターン特徴量空間の次元数*である。

まず実験の目的では 2^N 面体型に特徴領域を変形することは、パターン特徴量空間においてどのような意味があるかについて述べる。そしてそれを応用することでどのような効果が期待されるかについて、その項目を列記する。実験では、比較のためにその他の型の特徴領域にも注目しその効果について確かめ、結果について検討と考察を行う。

7.2 実験の目的

第 5 章の結論として、正 2^N 面体型特徴領域のまま特徴領域整合法を行うことは、最終目的である認識結果に正しさの保証を与えることに対し、第 5 章の結果以上の効果が望めないことを述べた。そこで容易に考えられる特徴領域のもう 1 つの変更方法として 2^N 面体型特徴領域によるパターン整合法を提案した。

2^N 面体型特徴領域の利用は、図 5.4 右に示すようにパターン特徴量空間内で距離尺度にシティーブロック距離を用いているために正 2^N 面体型特徴領域であったものが、認識の段階で距離尺度に重み付きのシティーブロック距離を利用することにより、 2^N 面体型特徴領域に変更しようという試みである。これにより、特徴領域を表すパラメータは次元数の数と同じになり増加する反面次のような効果が期待される。

- 特徴量空間内におけるサンプルパターンの特徴量のさらなる効果的包囲
- 安全認識領域の拡大

*特徴抽出に方向線素特徴量を使用しているため $N = 196$ である。故に、 $2^{196} \doteq 1.0043 \times 10^{59}$ 面体型特徴領域となる。

†第 6 章までは (正) 多面体と表記したが、本章以降は具体的な面数を記すことにより (正) 多面体の種類を区別する。

0

0

- 判断不能文字数の縮小

これらの上 2 つは、認識結果に対して正しさの保証を与えることが実現するための条件であることは第 3 章でも述べている。判断不能文字数の縮小とは、第 6 章ではサンプル数の少なさが原因で判断不能文字数が比較的多くなったが、 2^N 面体型特徴領域により特徴量空間内におけるサンプルパターンの特徴量が効果的に包囲できれば、特徴領域整合法によるパターン整合法を行うことにより、認識率の高精度化に繋がると共に判断不能文字数の縮小が期待できるということである。これは出力結果として何らかの候補が出力されやすくなる意味がある。

また、本章の実験では他の特徴領域の効果も調べるために、超楕円型と直 $2N$ 面体型についても調べる。これらは重み付きのユークリッド距離、重み付きのチェス盤距離と見なすことができ、 2^N 面体型、超楕円型、直 $2N$ 面体型の順に距離尺度の次元数が高くなる。

7.3 実験の方法

使用する文字パターン、特徴抽出法、辞書作成は第 4 章以降のオープン実験と比較できるように、本章においてもそれぞれ同様に表 4.1、方向線素特徴量、重心ベクトルを使用する。

2^N 面体型特徴領域の作成及び本章の実験方法について説明する。各文字の各成分において予めサンプルパターンの特徴量 $F(P_i^k) = (F_1(P_i^k), \dots, F_n(P_i^k), \dots, F_N(P_i^k))$ と辞書 $v^k = (v_1^k, \dots, v_n^k, \dots, v_N^k)$ から、標準偏差 $\sigma_{k,n}$ を求めておく。即ち、

$$\sigma_{k,n} = \sqrt{\frac{1}{s} \sum_{i=1}^s F_n^2(P_i^k) - (v_n^k)^2} \quad (7.1)$$

と表される。2965 字種に対し 196 次元なので 581140 個の標準偏差 $\sigma_{k,n}$ を求めることになる。

(7.3) 式の 2^N 面体内部を表す式において、各文字 k に対し標準特徴量を中心と考え、 $r_{k,n}$ は $\sigma_{k,n}$ に概ね比例した値をとるようにする。即ち、

$$r_{k,n} = a(\sigma_{k,n} + b) \quad (7.2)$$

とする。比例定数 a は、第 5 章の実験の $r_r = 1.5$ をもとに算出した。未知入力文字の特徴量 u が単純に辞書 v^k から標準偏差 $(\sigma_{k,1}, \dots, \sigma_{k,N})$ の 1.5 倍の位置に存在した場合、 $a = 1.5N$ で算出した大きさの特徴領域でないと包囲できない。実際には全成分において均等に辞書 v^k から離れることはないので、方向線素特徴量の場合 $a < 1.5N = 1.5 \times 196$ という条件になる。第 5 章の $r_r = 1.5$ の実験において、1 位候補の認識距離とその領域半径の関係から、 2^N 面体型特徴領域の場合 $a = 180$ と決めた。また標準偏差が 0 である場合、(7.3) 式の左辺が不定になるのを防ぐために、バイアス b を加えておく。本実験に使用するバイアス b を決めるために、予備実験としてバイアスを変化させたときの認識率を調べた結果から $b = 2.0$ とした。

これらにおいて特徴領域整合法でパターン整合法を行う。但し、誤認識集合は求めないため全字種に対して特徴領域内か否かを調べる。比較のために超楕円型と直 $2N$ 面体型特徴

0

0

領域についても同様に行う。相違点は当然特徴領域の作成時で、それぞれ(7.4)式、(7.6)式において r_n を求める際、(7.2)式で $a = 20$ 、 $a = 6.23$ とした。

最後に認識結果に対して高精度な結果が得られた場合、正しさの保証をどの程度与えることができるかについて検討を行う。

7.4 N 次元空間における多面体と超楕円

本章の実験で特徴領域の型として使用する N 次元空間における多面体と超楕円について説明する。この場合、面とは1次元低い空間における物体を指す。 N 次元空間における代表的な多面体は、表7.1に示すように、 $N+1$ 面体、 2^N 面体、 $2N$ 面体がある。

以降は何れも未知入力文字の特徴量 \mathbf{u} に関するパターン特徴量空間に合わせて考え、 $r_{k,n} > 0$ 、中心は \mathbf{v}^k とする。文字 k の 2^N 面体型特徴領域内部は次式で表される。

$$\sum_{n=1}^N \frac{|u_n - v_n^k|}{r_{k,n}} \leq 1 \quad (7.3)$$

N 次元 2^N 面体は表7.1に示すように、 $N-1$ 次元 N 面体が 2^N 個集まってできる。頂点の数は $2N$ 個である。実際の評価値計算では距離尺度として左辺を計算し、1を越えた時点で特徴領域外と判断し、1以内の最小値を与える文字を候補として出力する。

文字 k の超楕円型特徴領域内部は次式で表される。

$$\sum_{n=1}^N \frac{(u_n - v_n^k)^2}{r_{k,n}^2} \leq 1 \quad (7.4)$$

(7.3)式の定義における次元数2への拡張である。これも 2^N 面体型のときと同様に、実際の評価値計算では距離尺度として左辺を計算し、1を越えた時点で特徴領域外と判断し、1以内の最小値を与える文字を候補として出力する。

文字 k の直 $2N$ 面体型特徴領域内部は次式で表される。

$$\lim_{d \rightarrow \infty} \sum_{n=1}^N \frac{|u_n - v_n^k|^d}{r_{k,n}^d} \leq 1 \quad (7.5)$$

(7.3)式の定義において次元数無限大への極限として表される。 $2N$ 面体は表7.1に示すように、 $2N-2$ 面体が $2N$ 個集まってできる。頂点の数は 2^N 個である。 $2N$ 面体の中で、隣面どうしが直交しているものを特に直 $2N$ 面体と言う。この(7.5)式は単純に表すと次式のようになる。

$$\left\{ (u_1, \dots, u_N) \mid \frac{|u_n - v_n^k|}{r_{k,n}} \leq 1, \quad (n = 1, \dots, N) \right\} \quad (7.6)$$

実際の評価値計算では距離尺度として

$$E(\mathbf{u}, \mathbf{v}^k) = \max_{n=1}^N \frac{|u_n - v_n^k|}{r_{k,n}} \quad (7.7)$$

を用い、以下 2^N 面体型のときと同様に1を越えた時点で特徴領域外と判断し、1以内の最小値を与える文字を候補として出力する。

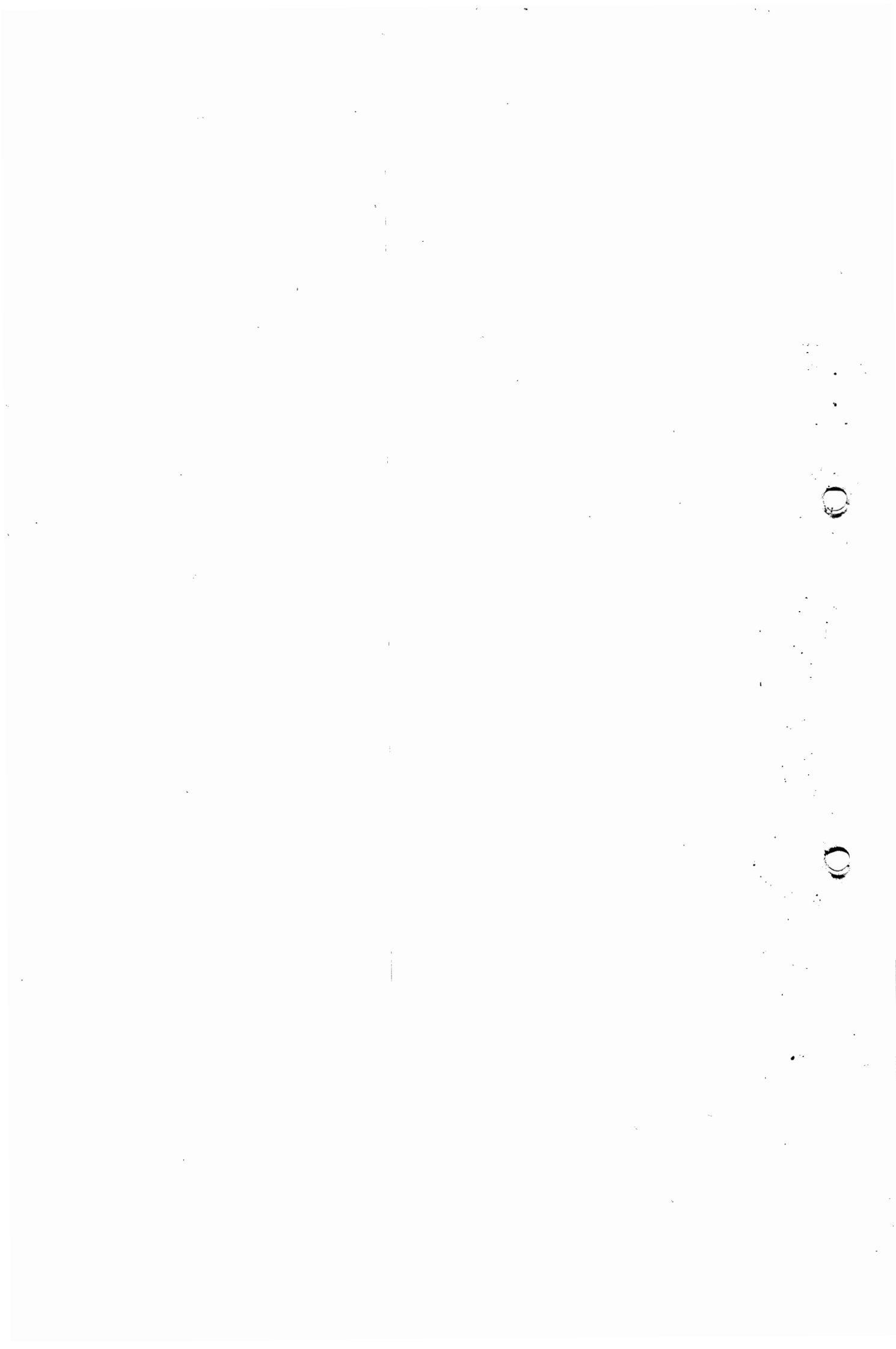


表 7.1: N 次元空間内の多面体

	$N - 1$ 次元多面体の		頂点の数
	種類	個数	
$N + 1$ 面体	N 面体	$N + 1$	$N + 1$
$2N$ 面体	$2N - 2$ 面体	$2N$	2^N
2^N 面体	N 面体	2^N	$2N$

7.5 実験結果と考察

実験の結果を表 7.2 に示す。数字は今までと同様 1 セット (JIS 第 1 水準 2965 字種) 当りの平均値である。 2^N 面体型特徴領域によるパターン整合法が認識率、判断不能文字数ともに他型特徴領域より良い結果となったのが分る。これは、第 4 章でも示したように、距離尺度の定義における次元数が小さい程、サンプルパターンの特徴量を効果的に包囲していることが明らかになった。

2^N 面体型特徴領域に注目すると、4 セットの未知入力文字のうち、誤認識したのは表 4.1 のオープン実験用 3 番の「土」を「土」と判定した 1 例だけであり、認識率 99.992% という高精度な結果が得られた。第 4 章以降のパターン整合法実験と比べても、図 8.1 の横軸の 7 より、認識率 (○印) の精度がさらに高くなったことが分る。

また判断不能文字数は、表 4.1 のオープン実験用 2 番の「誉」を判断不能と判定した 1 例だけである。図 8.2 の 6 の多特徴領域によるパターン整合法と比べ、判断不能文字数 (×印) が大幅に減少できたことが分る。

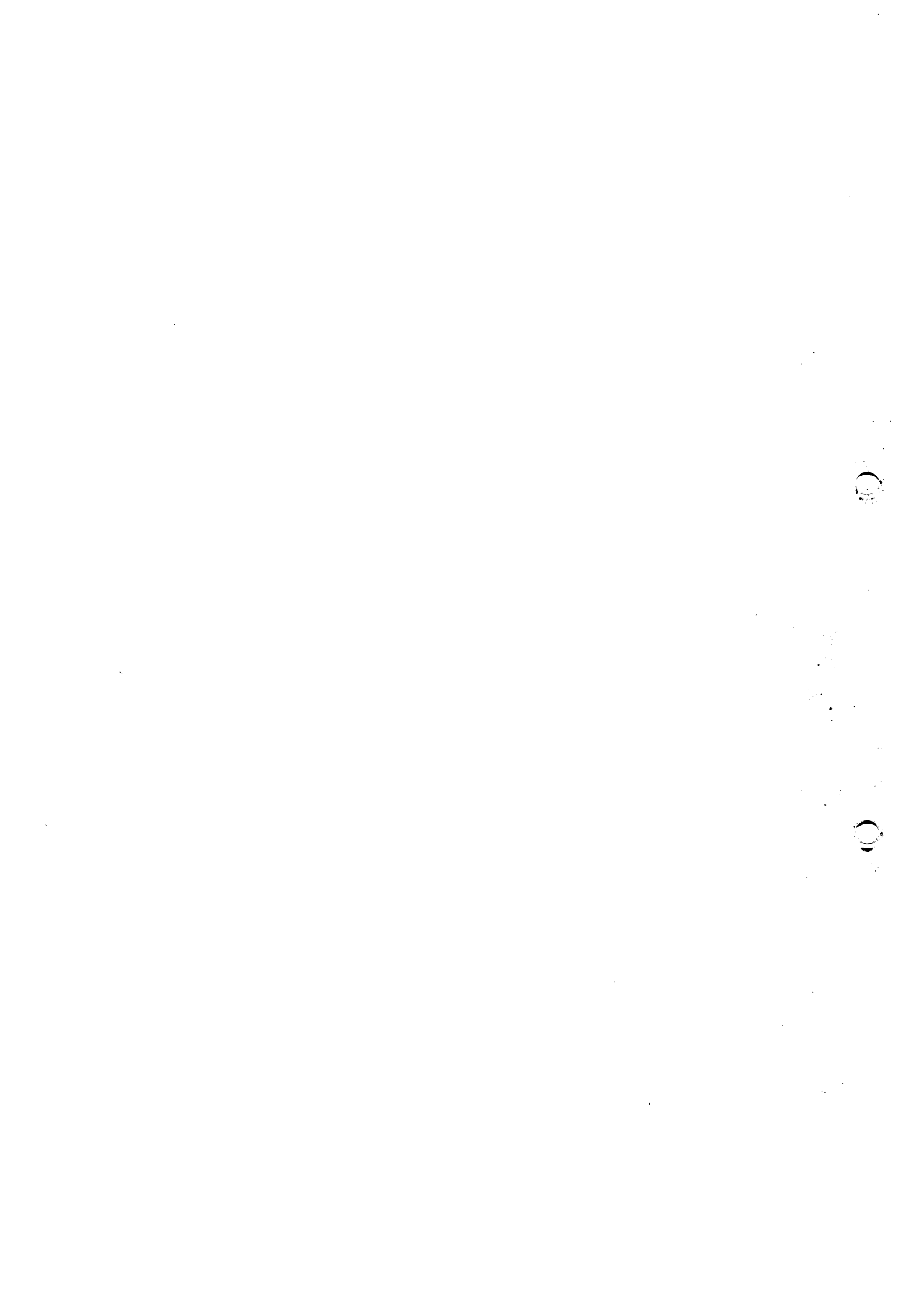
計算対象文字数については、今回の実験では誤認識文字集合を求めていないためにその見解は述べない。

表 7.2: N 次元 2^N 面体型特徴領域によるパターン整合法の結果

特徴領域の型	2^N 面体	超楕円	直 $2N$ 面体
認識率	99.992%	99.797%	93.384%
誤認識文字数	0.25 字	6.00 字	194.25 字
判断不能文字数	0.25 字	5.25 字	29.00 字

7.6 認識結果の正しさの保証について

ある程度の高精度化が達成し、また特徴領域整合法により多くの文字に何らかの判定をすることができたので、以降は候補出力された文字に対して、 2^N 面体型特徴領域によるパターン整合法がどの程度正しさの保証を与えられるかについて調べてみた。



まず命題1の

$$u \notin C(k)^{\dagger} \text{ならば、} u \text{は} k \text{でない。}$$

について、候補出力されたもの全部に対し調べた。「土」の未知入力文字の特徴量は、「土」という結果を出力したが、「土」の特徴領域 $C(\text{土})$ には属している。そのため誤認識した全ての文字は、正解すべき文字の特徴領域に存在することになり、候補出力し正解だったものは特徴領域整合法のアルゴリズム上必ずその特徴領域に属するので、命題1による保証は、本章の実験により完全に与えたことになる。

誤認識文字集合は処理時間の関係で求めなかったが、代用として候補文字集合を求めた。その結果候補文字集合の大きさ $|\text{Cand}|$ の1文字当りの平均は2.27字であった。そこで命題2の

$$u \in \mathcal{F} \text{かつ} |\text{Cand}(u)| = 1 \text{ならば、} u \text{は} k \text{である。但し、} \text{Cand}(u) = \{k\}。$$

を用いて認識結果の正しさの保証を与えるとすると、表7.3のようになる。数字は1セット当たり(全2,965字種)の平均値である。正しさの保証無しは欄は2段になっており下段は内訳である。内訳の「正」、「誤」、「判断不能」の意味は、それぞれ正しさの保証を与えられないが、結果として正解だった文字、誤認識した文字、判断不能文字を表す。その表7.3から分るように、約半分の候補に正しさの保証を与えることができた。

さらに厳密な正確の判定について、正しさの保証を与えるには命題7の

$$u \in A(k)^{\S} \text{ならば、} u \text{は} k \text{である。}$$

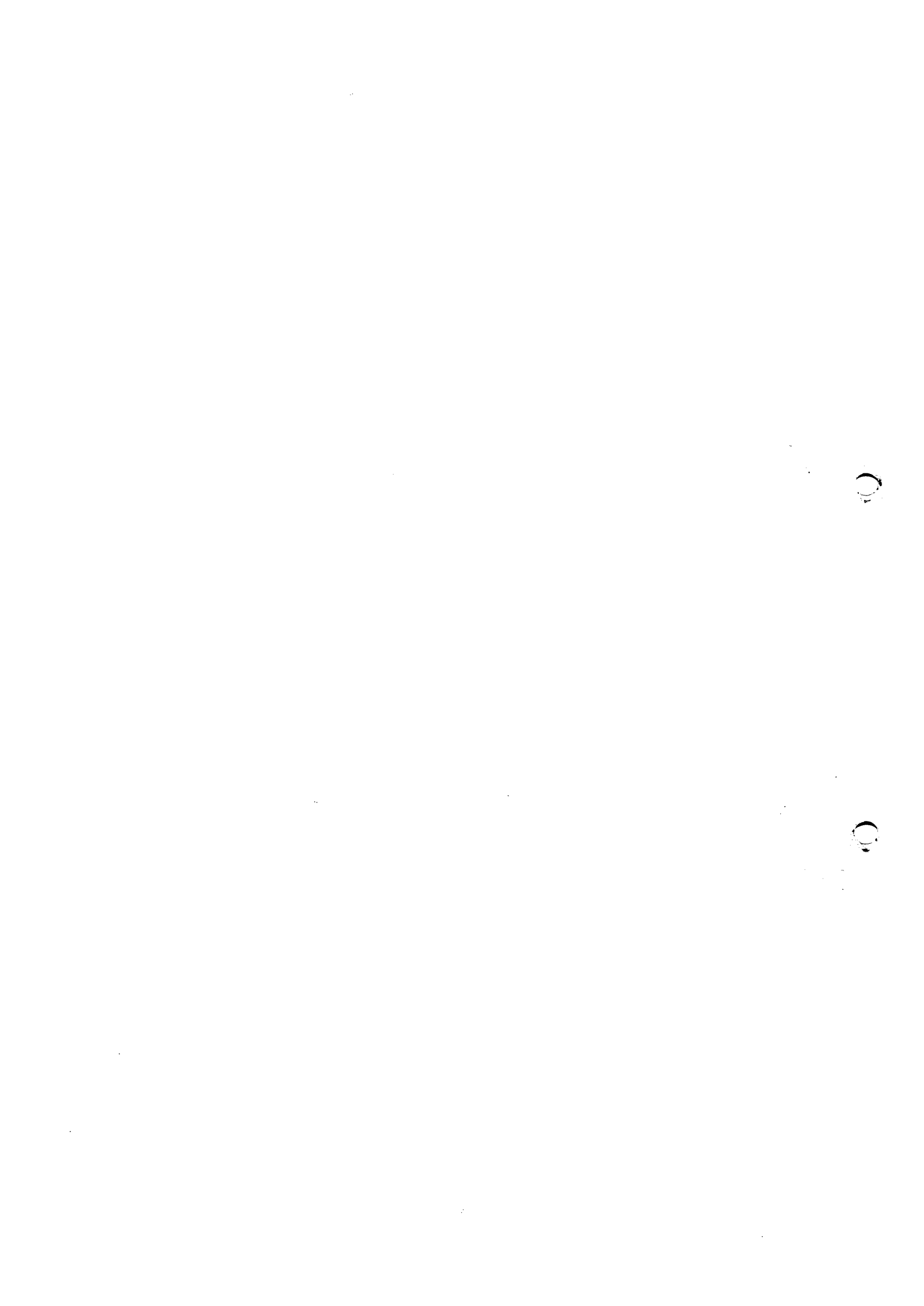
を用いる。しかし本章では誤認識文字集合を求めておらず、安全認識領域は定まらない。そのため結果及び考察は割愛する。

表7.3: 命題2による認識結果の正しさの保証についての文字数とその率

	正しさの保証有り	正しさの保証無し		
		正	誤	判断不能
文字数	1421.25 字	1543.75 字		
		1543.25 字	0.25 字	0.25 字
率	47.934%	52.066%		
		52.049%	0.008%	0.008%

[†]第2章の命題1では、「 $E(u, v^k) > r_k$ ならば、 u は k でない。」であるが、本章では各文字毎に(7.3)式左辺より距離尺度を変更しているため、領域半径は全て1になる。よって現段階において同義である特徴領域 C を用いて表記しておく。

[§]第2章の命題7では、「 $E(u, v^k) < s_k$ ならば、 u は k である。」であるが、これも現段階において同義である安全認識領域 A を用いて表記しておく。



7.7 後書き

本章は第4章の第3改良案である 2^N 面体型特徴領域によるパターン整合法について実験を行った。

その結果、距離尺度の定義における次元数が小さい程、サンプルパターンの特徴量を効果的に包囲していることが明らかになった。また図8.2より、多特徴領域によるパターン整合法と比べ、判断不能文字数(\times 印)を大幅に減少できた。本章の誤認識文字集合の大きさ(∇ 印)は候補文字集合からの推定である。判断不能文字数及び誤認識文字集合の大きさが共に小さくなることは、本章の実験結果以外例がなく、 2^N 面体型に特徴領域を変更したことは大変効果的であると言える。さらに認識率に関しては第4章以降の実験で図8.1に示すように最高の精度を得ることができた。計算対象文字数は本章の実験では割愛したため、図8.3には示されていない。

最後に認識結果の正しさの保証についてどれだけ与えられるか調べた。その結果、 2^N 面体型特徴領域によるパターン整合法により、命題1による保証を完全に与えることができた。さらに命題2により、正しさの保証を半分近く与えることができた。本章の実験では処理時間等の諸理由により、誤認識文字集合を求めなかった。そのために、命題7による正確な保証をどれだけ与えるかは今後の課題である。これは表3.5の結果からどれだけ成果が得られているかが注目される。

00

00

第 8 章

結論

8.1 前書き

本研究では認識結果に正しさの保証を与える正確な認識を究極の目的とし、それに準ずる目的である認識率の高精度化の達成ために様々な方法を提案し実験をした。本章ではその成果について纏め、最後に今後の課題を述べる。

8.2 本研究の成果

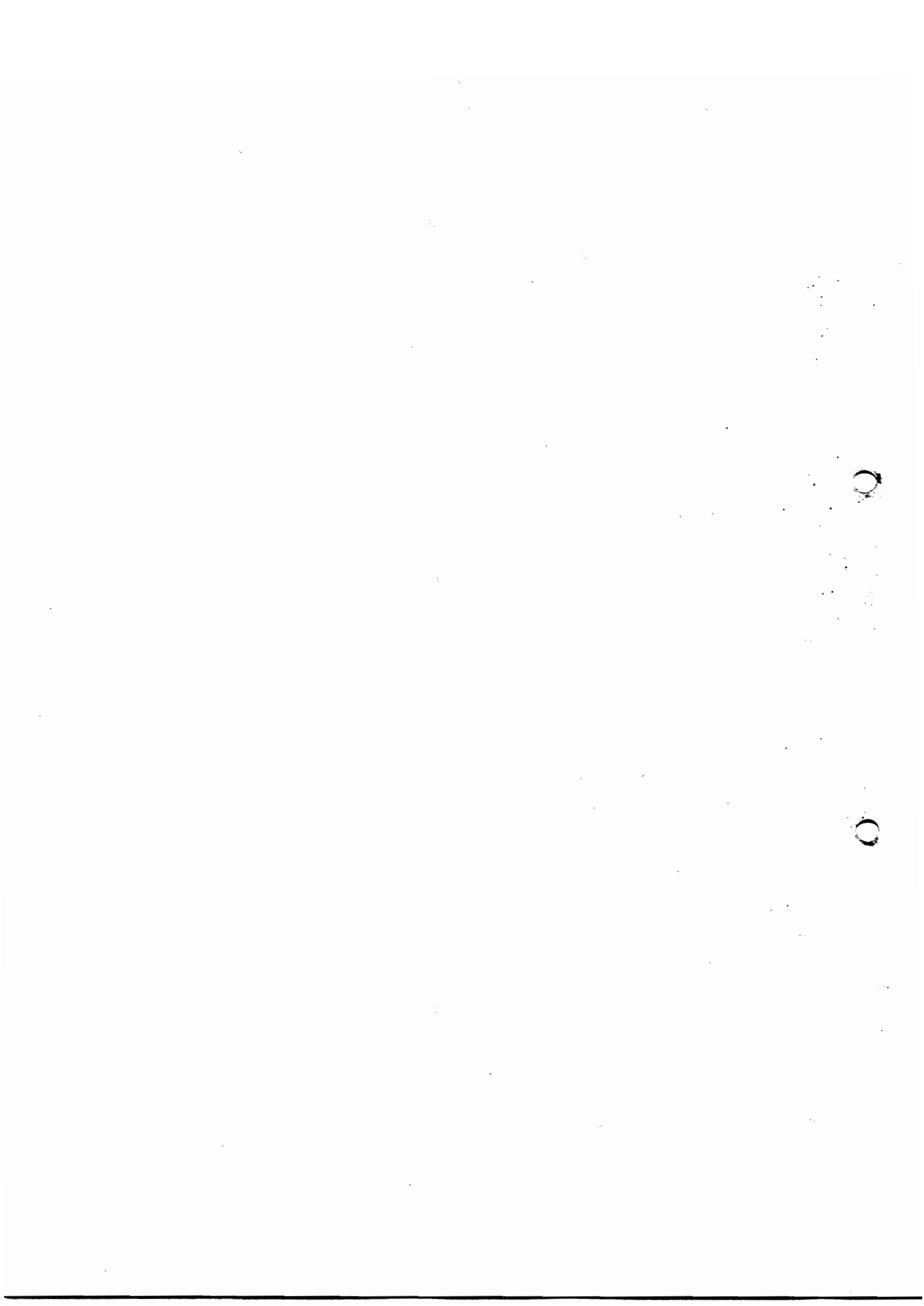
まず最初に各章毎の成果について纏める。

第 2 章では、パターン整合法の理論的分析について述べ、認識結果の正しさの保証をどう与えるかについて提案した。そして種々の定義や命題を表記し、パターン整合法において評価値を用いるだけでは厳密な正確さの保証を一般的には実現しえないことを明らかにした。

第 3 章では、第 2 章の理論的結果に対し具体的な特徴抽出法に注目し実験を行った。その結果、安全認識領域に属するパターンの特徴量は少なく、認識結果の正しさを保証するためには、現状のパターン整合法のままでは安全認識半径は大きく役立たず、ある程度の有効性しか示されなかった。しかしパターン整合法の理論的な分析は、特徴領域整合法等の幾つかの応用例を生み出した。また領域半径、安全認識半径、誤認識集合を求めることで特徴抽出法の評価に使用できることが明らかになった。

第 4 章では、特徴領域整合法を提案し全数整合法と比較実験をした。実験から、特徴領域整合法に距離尺度としてシティーブロック距離を用いる組み合わせが一番有効であることを示した。また特徴領域整合法は全数整合法に比べ、効率的に候補出力できることが示された。しかしこの実験で定義した領域半径は真の領域半径より小さいことが明らかになり、次章以降への課題となった。

第 5 章では、第 4 章の第 1 改良案として、標本領域半径の大きさを可変することによる真の領域半径の選出について、パターン整合法を行い実験をした。その結果、候補出力のための総合的な評価から新しい標本領域半径を定めた。また認識結果に対し、命題 1 により保証する最小の標本領域半径も定めた。そしてこれらの標本領域半径の差が、超球(正



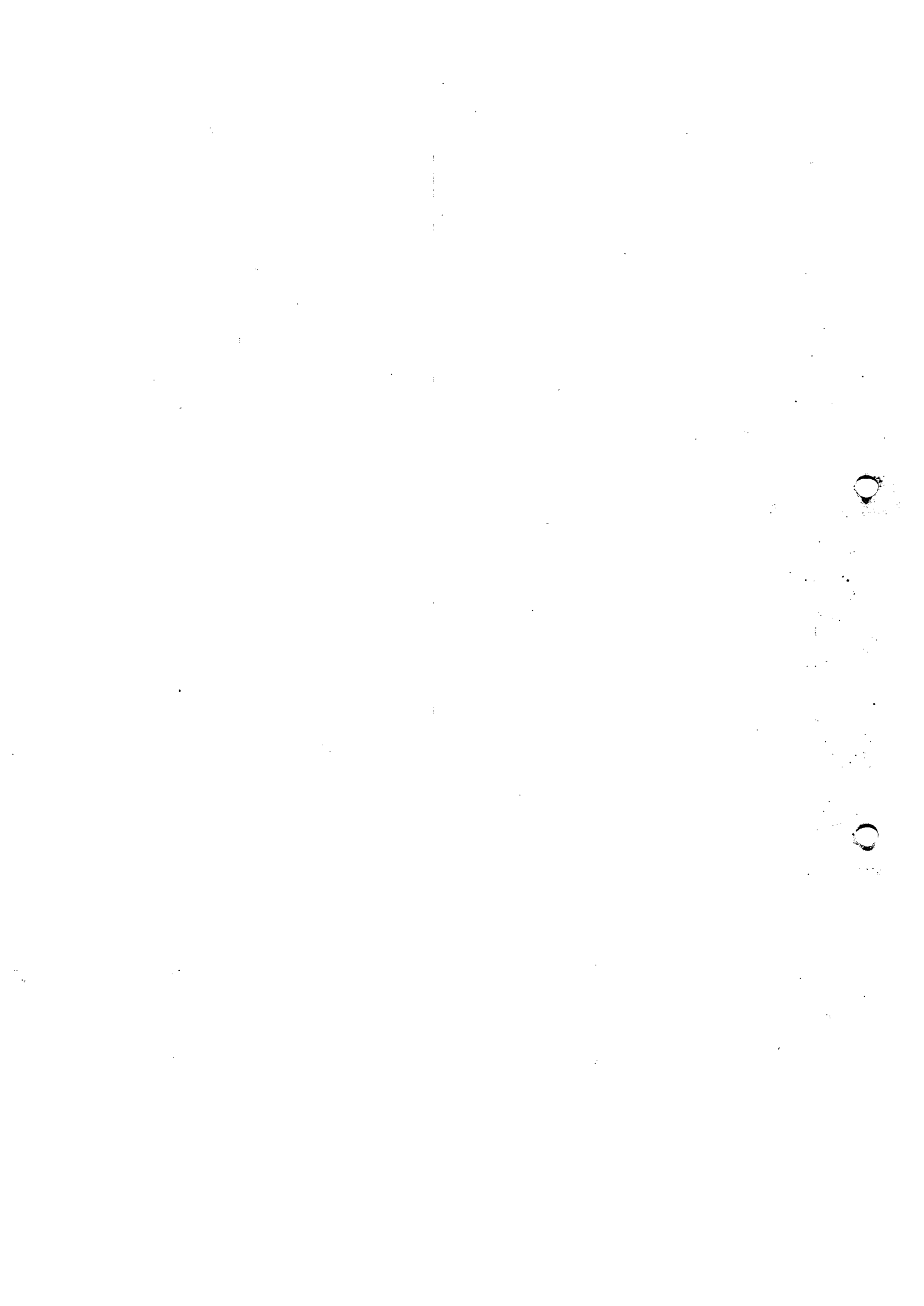
2^N 面体)型特徴領域による整合法の限界を示唆していることを明らかにした。

第 6 章では、第 4 章の第 2 改良案として、マルチテンプレート法を利用した特徴領域整合法によるパターン整合法を行った。その結果、シングルテンプレートの全数整合法と比べ、計算対象文字数を減少させかつ認識率を高めることができた。また、全数整合法に対する計算対象文字数の減少率が高いことが分り、多特徴領域による有効性を示した。

第 7 章では、第 4 章の第 3 改良案として、特徴領域をサンプルデータの特徴量の各成分の標準偏差により変形 (2^N 面体型) することによる効果について調べた。その結果、距離尺度の定義において次元数が小さい程、サンプルパターンの特徴量を効果的に包囲していることが明らかになった。また、多特徴領域によるパターン整合法と比較すると、判断不能文字数を大幅に減少できそれと共に認識率の向上も達成できた。

最後に総合的な成果として結論づける。本研究はパターン整合法に焦点をあて、特徴量空間を分析し、最終目的である認識結果に正しさの保証を与えるための理論を導いた。そして領域半径、安全認識半径、誤認識文字集合等が正しさの判定基準であることを示し、同時に認識率の高精度化を目指した。各章においてその度毎に改良を加えながら実験を行った結果、章の進行における認識率 (○印)、誤認識文字数 (△印)、判断不能文字数 (×印)、誤認識文字集合の大きさ (▽印)、計算対象文字数 (◇印) は、図 8.1、8.2、8.3 のように推移した。最終的には認識結果に対し、命題 1 による保証は完全に与えられ、命題 2 により正しさの保証を半分近く与えることができた。以前までは確率的に候補を定めていたのに対し、本研究により初めて候補が決定的に定められる結果が生まれた。以上をもって本研究の目的は達成できた。

それから認識結果に対する正しさの保証の有無にかかわらず、文字認識アルゴリズムの評価方法としてこれまで君臨してきた認識率を考えた場合、最終的に 99.992% という結果を得ることができ、従来の文字認識アルゴリズムである全数整合法と比較して 0.103% 上昇することができた。99% を越える認識率にありながら、0.1% 以上の精度を高められたことは、本研究による特徴領域整合法が大変有効だったことを意味している。



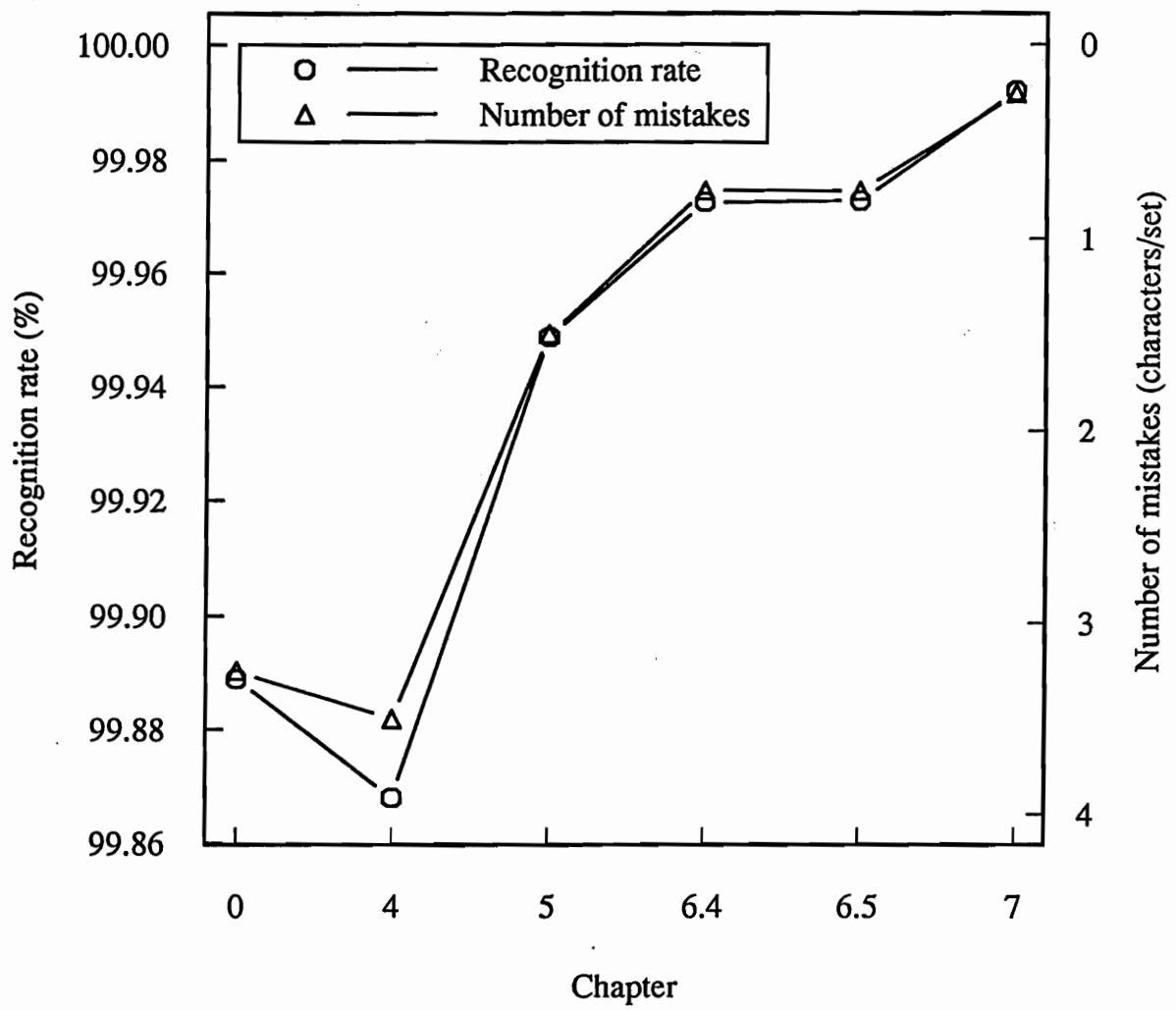


図 8.1: 本研究における認識率と誤認識文字数の推移

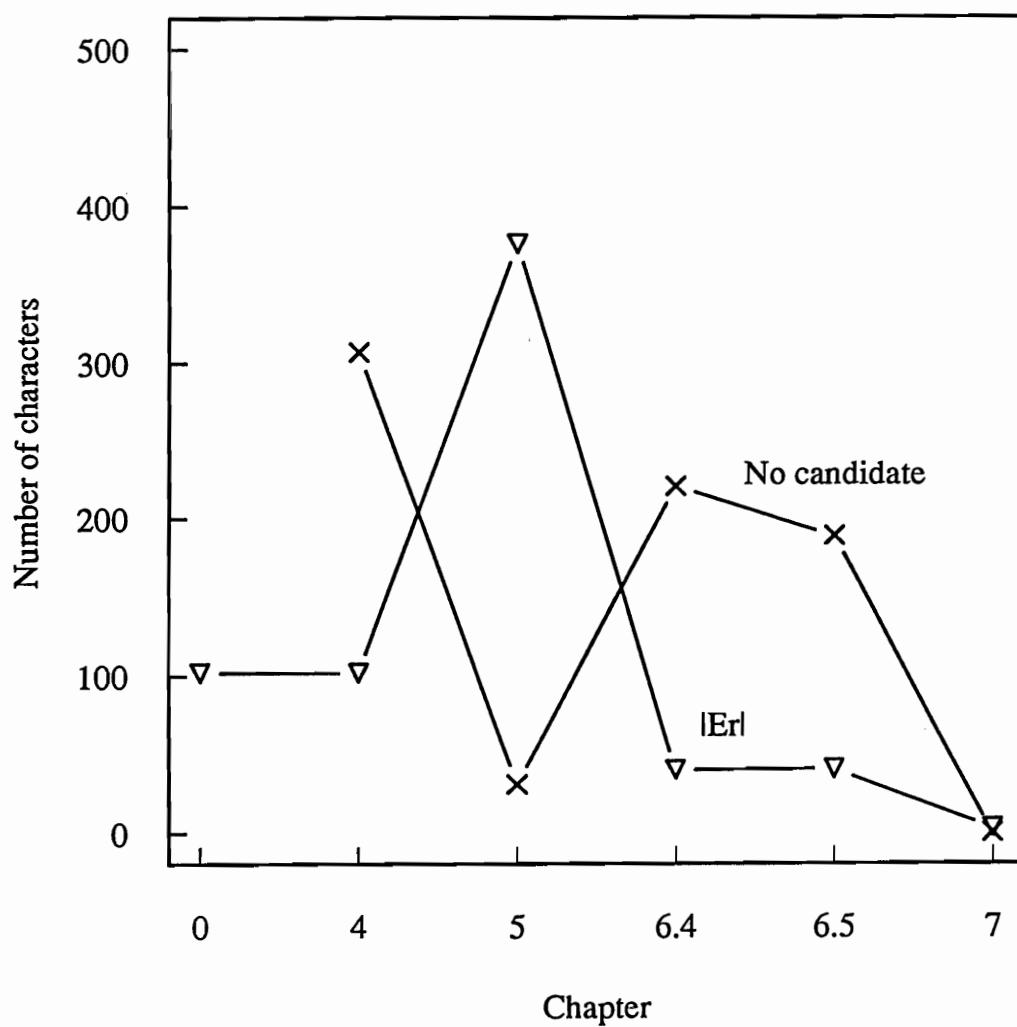
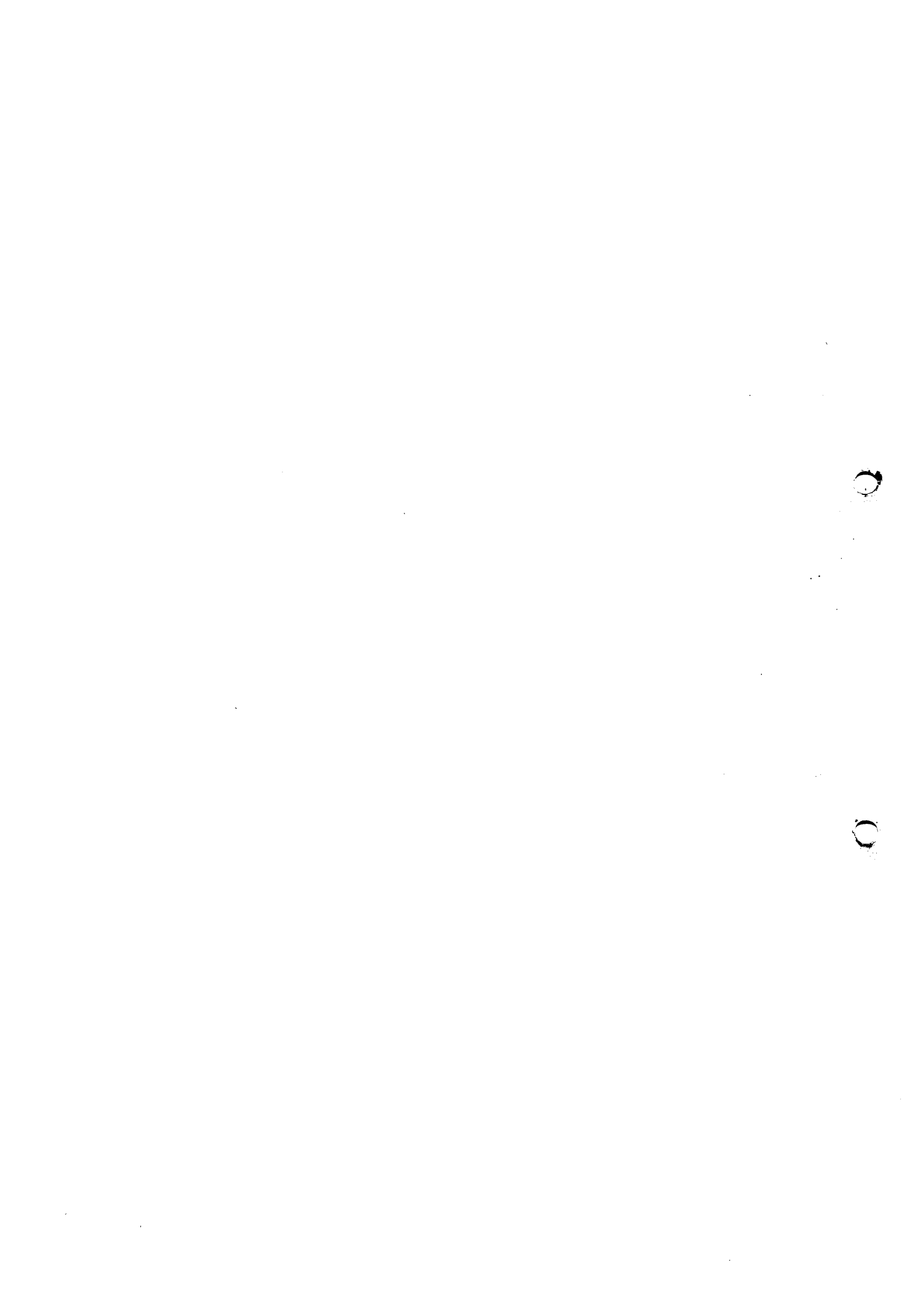


図 8.2: 本研究における判断不能文字数と誤認識文字集合の大きさの推移



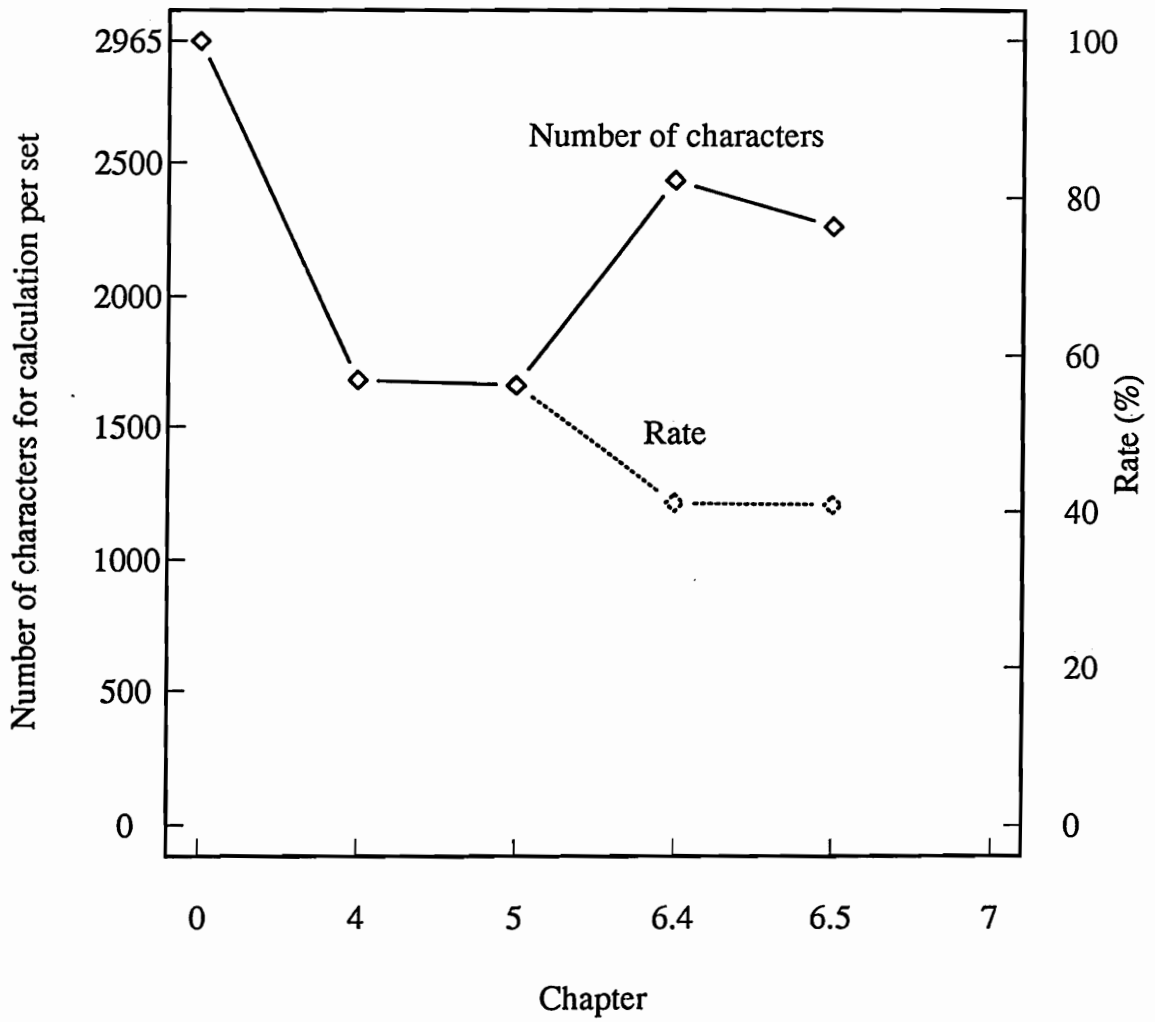


図 8.3: 本研究における計算対象文字数の推移

8.3 今後の課題

本研究の目的は達成できたものの、このまま保証の有無を用いた文字認識システムを開発したとすると、認識結果に対し正しいという保証が有るものが半分弱で、その他は全て間違えているかもしれないという曖昧な結果を出力する。認識率は100% *になるものあまり実用的ではない。今後は正しさの保証を多数の候補に与えられるように、保証率[†]の向上が課題である。

その方法としてまず考えられるのが 2^N 面体型多特徴領域による方法である。これは第6章と第7章の組み合わせで、マルチテンプレート法とサンプルパターンの特徴量の各成分の標準偏差を用いることにより、さらなる向上が期待できる。但し、マルチテンプレートの実験ではサンプルパターンのさらなる充実が不可欠である。

次に考えられる方法として主成分分析を用いる方法である。第7章の 2^N 面体の軸は、パターン特徴量空間において座標軸に対し平行であった。これに主成分分析を適用することにより、 2^N 面体が特徴量空間内で回転し、サンプルパターンの特徴量のさらなる効果的包囲が予想され、それと同時に保証率の向上が期待できる。

本研究では正確さの保証を命題2で与えたが、安全認識領域で判断する命題7により、さらに厳密に保証を与えられるようにすることも今後の課題である。そしてそれは、表3.5からの向上より成果が伺える。そのためには誤認識文字集合を求めなければならない。今後誤認識文字集合を求めることは特徴領域の複雑化に伴い、困難になっていくことが予想される。その解法も今後の課題である。誤認識文字集合からは計算対象文字数が求められる。 2^N 面体型特徴領域によるパターン整合法の場合、誤認識文字集合の大きさから推定して約1500前後の計算対象文字数で候補を出力できることが期待できる。

最後に本研究は個別漢字認識であり、実際の文章に利用することは言及しなかった。そのため認識の整合の順序をJIS第1水準のコードの順に合わせた。しかし、仮に実際の文章に利用した場合は文字の頻度に合わせて整合することにより、計算対象文字数がさらに縮小できることが期待できる。実際の文章における利用も今後の大きな課題である。

8.4 後書き

文字認識についての研究が本格的に始まってから30数年たった現在、文字認識技術がこれからますます向上する中、それを処理する計算機も平行して進展し続けている。そして両者相互が発展し続ける中で、近い将来認識結果に対する厳密な正確さの保証は完全に与えられるであろう。

* 認識率は候補を判定しないものは含めず、確実に出力されたものを計算の対象にするので結果として認識率100%になる。

[†] 未知入力文字の総数に対する、認識結果の正しさの保証を得た文字数の割合。

謝辞

本研究を進めるにあたり全般的な御指導を賜りました北陸先端科学技術大学院大学の木村正行教授、多大な御指導、御助言を賜りました東北大学工学部の阿曾弘具教授に心から感謝致します。

また本論文を纏めるに際して貴重な御意見を戴いた東北大学工学部の伊藤貴康教授、東北大学電気通信研究所の佐藤雅彦教授に深く感謝致します。

北陸先端科学技術大学院大学の堀口進教授、同じく下平博助教授、東北大学情報処理教育センターの阿部亨助手の方々には、日頃より御助言を戴いたり、パターン認識全般に渡り御指導して戴き誠に有り難う御座いました。

さらに東北大学大型計算機センターの孫寧助手、東北大学工学部の瀧本英二助手、セイコーエプソン株式会社の内山喜照氏、東北大学大学院工学研究科情報工学専攻博士課程在学の大町真一郎氏、同じく成富敬氏、同じく後藤英昭氏の方々には、熱心な御討論、有益な御教示を戴き、また併せて計算機環境の整備等にも力を添えて戴き厚く感謝致します。

最後に多面に渡り御意見、御協力を戴き、また日頃の生活において御世話になった研究室の皆様、並びに同期修了生各位に感謝致します。

参考文献

- [1] 孫寧、田原透、阿曾弘具、木村正行：「方向線素特徴量を用いた高精度文字認識」
電子情報通信学会論文誌 (D-II), Vol.J74-D-II, No.3, pp. 330-339 (平成3年3月)
- [2] 江島俊朗、中村洋介、木村正行：「構造情報を含む手書き漢字認識のための特徴量」
電子情報通信学会論文誌 (D), Vol.J68-D, No.4, pp. 789-796 (昭和60年4月)
- [3] 梅田三千雄：「マルチフォント印刷漢字の分類」
電子情報通信学会論文誌 (D), Vol.J62-D, No.2, pp. 133-140 (昭和54年2月)
- [4] 進藤宣博：「低品質印刷文字の認識に関する研究」
東北大学大学院工学研究科情報工学専攻 修士学位論文 (平成2年2月)
- [5] 阿曾弘具、越後和徳：「正確な個別文字認識の検討」
電子情報通信学会技術研究報告, PRU91-68, pp. 47-54 (平成3年10月)
- [6] 孫寧、阿曾弘具、木村正行：「連想整合法に基づく高速文字認識アルゴリズム」
情報処理学会論文誌, Vol.32, No.3, pp. 404-413 (平成3年3月)
- [7] C. J. Hilditch : "Linear Skeleton from Square Cupboards"
In Machine Intelligence 6, B. Meltzer & D. Michie, Eds., Univ. Press, Edinburgh,
pp. 403-420 (1969)
- [8] 飯島泰蔵：「パターン認識理論」
森北出版 (平成元年5月)
- [9] 森健一：「パターン認識」
電子情報通信学会編 (昭和63年11月)
- [10] 長尾真：「パターン情報処理」
コロナ社 (昭和58年3月)
- [11] 有馬哲、石村貞夫：「多変量解析のはなし」
東京図書 (昭和62年10月)

公表目録

- [1] 越後和徳、阿曾弘具、木村正行：「文字認識アルゴリズムにおける正確な判定について」
平成3年度電気関係学会東北支部連合大会, 1H-1, (平成3年8月)
- [2] 阿曾弘具、越後和徳：「正確な個別文字認識の検討」
電子情報通信学会技術研究報告, PRU91-68, pp. 47-54 (平成3年10月)
- [3] 越後和徳、阿曾弘具：「パターン特徴量空間における性質を利用した文字認識」
平成4年度第1回情報処理学会東北支部研究会, (平成4年6月)
- [4] 越後和徳、阿曾弘具：「パターン特徴量空間の性質を利用した効率的な文字認識」
平成4年度電気関係学会東北支部連合大会, 2C-3, (平成4年8月)
- [5] 阿曾弘具、越後和徳、木村正行：「文字特徴量空間の性質と特徴抽出法の性能評価法」
電子情報通信学会論文誌 (D-II), (掲載予定)