

文字認識精度推定による確認修正手段の選択とその効果

松村 季樹[†] 小林 公知^{††}
中村 修^{†††} 小倉 健司[†]

文字認識技術は、手書き文書を入力するための有効な方法として期待されている。しかし、従来の入力方式のように、認識誤りの多い文書に対して確認と修正を繰り返すのみでは、入力に要する作業の効率は低下する。一方、確認修正手段の適用法の改善等により、作業効率の向上が可能となる場合もある。そこで本論文では、文字認識率等の認識精度と、文字認識後の文書を修正する場合の作業効率との関係に着目し、認識精度に応じて最も効率の良い確認修正手段を、複数の手段の中から自動選択する入力方式を提案する。本方式により、任意の認識精度の文書に対して効率的な入力が可能であることを示す。まず、認識精度ごとに最大入力効率を示す確認修正手段を、測定によりあらかじめ認識精度と対応づけておく。入力効率として、たとえば、単位時間あたりの入力可能文字数を表す入力速度を想定する。次に、得られた対応関係を参照して、文書の文字認識結果に基づいて推定された認識精度から、入力速度が最大となる確認修正手段を自動選択する。入力作業は、提示された確認修正手段に従って行う。手書き文書を対象とした実験により、従来のような単一の確認修正手段を用いた入力方式と比較して、およそ 30% の入力速度向上という結果が得られ、提案方式の有効性を確認した。

Selection of Correction Workflow Modes Using Assessment of Character Recognition Accuracy and Effect of Selection

SUEKI MATSUMURA,[†] TOMONORI KOBAYASHI,^{††}
OSAMU NAKAMURA^{†††} and KENJI OGURA[†]

This paper proposes a document-input method that selects the optimum workflow mode on the basis of character recognition accuracy. This method automatically selects the optimum workflow mode using the relation between recognition accuracy and input efficiency and can increase overall input speed (the number of inputted characters per unit time) for any accuracy. Experimental results using handwritten documents show that with this method input speed increases as a whole in comparison with input methods using a single workflow mode.

1. ま え が き

近年は、オフィスへのパーソナルコンピュータの普及が著しく、従来のような紙を媒体とする文書と並んで、電子ファイルとしての文書が一般的になってきた。しかし現状では、手書き文書やファクシミリ (FAX) で受信した文書のように、紙による文書の利用の割合も依然として高いといえる。これらの文書は、受注業務等においては、何らかの入力手法により、コンピュータにとって可読な形態へと変換することによって初めて有効な利用が可能となる。現時点で最も効率の良

い入力手法として、文字認識技術を応用したシステム (OCR) の利用があげられる。そこで、OCR の性能向上を目的に、従来から様々な研究開発が行われてきており、現在もおお継続して進められている¹⁾。特に、その中核をなす文字認識アルゴリズムに関して、文字形状の変形の整形²⁾に関する課題等をはじめとして、認識率向上のために様々な手法が提案されている³⁾。

一方、このような文字認識に関する要素技術の向上とは異なる視点からの研究も行われている。すなわち、現段階での文字認識技術をそのまま適用し、文書入力に要する作業の削減に重点を置いた研究開発である⁴⁾。文書入力は、通常スキャナーあるいは FAX を用いて文書を取り込み、文字認識により文字をコード化することにより行う。認識結果の誤りに対しては、現状の文書入力システムでは、確認と修正を繰り返す方法による入力 (認識結果修正型入力) が一般的である。ところがこの方法では、入力作業の削減効果が低下して

[†] NTT インテリジェントテクノロジー株式会社
NTT Intelligent Technology Co. Ltd.

^{††} NTT アドバンステクノロジー株式会社
NTT Advanced Technology Corporation

^{†††} NTT サイバーソリューション研究所
NTT Cyber Solution Laboratories

しまう場合がある。すなわち、認識誤りが増加するにつれて、修正のための作業が増大し、文字認識適用の効果が半減してしまうのである。

このような問題に対し、文書の入力効率の向上をねらいとして、特に字体の変形に起因する文字認識結果の確からしさの変動に着目することにより、問題解決をねらった関連研究として以下があげられる。まず、手書き文字のかすれやつぶれ等の品質に対する評価尺度を導入し、定量化を行う方法が報告されている⁵⁾。また、入力イメージの特性に基づいたOCRの認識性能の予測に関する手法があげられる^{6),7)}。文献6),7)の提案では、印刷文書における入力文字イメージの劣化特性を定量化することにより、認識結果とは独立にOCRの認識性能の予測/分類を実現している。

以上の研究成果は、現状の文字認識技術の要素技術向上とは異なる視点から効率の良い文書入力技術を確立するうえで、重要な意義を持つと考えられる。そこで、文字認識関連要素技術そのものの向上とは異なる方向に目を向け、現状の文字認識技術を有効に活用する立場に立って高効率文書入力の実現を目指すのが、本研究の目的である。具体的には、OCR出力に誤りが含まれることを前提として、認識精度（認識結果を定量的に表現する尺度で、たとえば第1位文字候補が正解である率を表す正読率）の変化に柔軟に対応する方式を提案する。すなわち、様々な認識精度に対応して全体として最も効率が良い確認修正手段を選択する基本メカニズム^{8),9)}を一般化し、確認修正手段のより柔軟な選択法を示す。また、提案方式を実現する主要な構成要素として、OCRから得られる結果の認識精度を推定する手法、複数の確認修正手段を選択する条件の導出法、推定された文書の認識精度に基づく最適な確認修正手段の選択法について述べる。

本論文では、2章で認識精度と文書の入力効率の関係について述べ、その関係に基づいて、認識精度に対応して適切な確認修正手段を選択する方式の基本的な考え方を示す。3章ではOCRの出力結果から認識精度を推定する手法について、4章では認識精度に応じた最適な確認修正手段切替のメカニズムについて述べる。5章ではそれらの技術に基づくシステム構成例ならびに文書入力時間短縮の実験から、本方式の有効性について述べる。

2. 文字認識結果と文書入力

本章では、文書の認識精度と、確認と修正の繰返しに基づく入力効率の関係について述べる。

受注業務等で入力の対象となる手書き文書/文字

は、たとえば電子技術総合研究所の漢字データベース ETL9B のような文字とは大きく異なっており、大多数が低品質である。したがって、これら低品質文字を認識対象とした場合、認識精度は大きく低下することが予想される。実際、郵便はがきの宛名の町域相当部分を認識対象とすると、第1位文字候補の正読率は、57.5%という結果が得られている¹⁰⁾。

また、我々が行った手書き文書の出現率の分析についても同様の結果が得られた。出現率は、まず第1位認識候補の正読率を文書ごとに求め、次にそれらの正読率が等しい文書の出現頻度を率として表す。認識対象は、実際の受注業務で得られた手書き文書に含まれる漢字である。これらは不特定多数の記入者によって記入されており、また、楷書・行書、JIS第1・第2水準等といった区別はなく、様々な種類・品質で構成されている。正読率の導出には、外郭方向寄与度特徴に基づく認識アルゴリズム¹¹⁾を用いた。この結果から、平均±標準偏差の範囲に含まれる68.3%の文書は、正読率56%から85%の間に分布していることが分かった。すなわち、これら中程度の正読率を持つ文書は高い割合で出現し、かつ多くの誤り文字を含む。一方、正読率が85%以上の文書については、修正箇所は少なく、かつ出現頻度は低い。したがって、出現頻度の高い文書ほど多くの入力作業を必要とする。

次に、従来から行われている一般的な認識結果修正型入力による入力速度の特性を分析した。ここで、認識結果修正型入力および入力速度をそれぞれ以下のように定義する。まず、認識結果修正型入力では、OCR出力に対し、文書の投入者（オペレータ）が目視で誤り文字の有無を確認し、誤りが見つかった場合に、元の文書と比較しながら正しい文字となるよう修正・入力していく確認修正手段のことである。また入力速度は、認識結果の正誤の確認時間も含めて、単位時間あたりの入力文字数で定義する。すなわち、認識結果が正しく、修正されなかった文字も考慮されることになる。また、漢字の場合には、ローマ字（またはかな）入力によって入力した文字を変換し、確定するまでの時間と確定後の文字数とを採用する。入力速度の測定には、前述した受注業務で得られた手書き文書のうち、同一の書式100枚を用い、文書内の1つの特定記入欄のみ対象とした。これは、記入欄間のカーソル移動に要する時間を測定から除外するためである。この分析から、正読率がおおよそ90%以上の文書に対しては、高い入力速度で文書の確認修正が行えることが分かった。ところが、平均±標準偏差の範囲に含まれる68.3%の文書に対しては、認識結果修正型入力の効果

はきわめて低い。このため、認識誤りの修正に多くの時間を必要とする。したがって、この分析で用いた文書の確認修正には、たとえば正読率 50% から 90% の領域で、認識結果修正型入力より入力速度が優位となるような確認修正手段を用いる方が効率的であるといえる。

さらに、オペレータが認識結果を目視で確認した結果、誤りが多いと分かれば、通常は認識結果修正型入力を行うことはない。すなわち、認識結果を一括削除して、ワープロのように最初からかな漢字変換による入力（非認識型入力）を行うことになる。ここで、前述の文書データを用いて、正読率 50% 以下の文書に対して認識結果修正型入力を行った場合と、非認識型入力によりすべてを入力し直した場合のそれぞれの入力時間を分析した。その結果、認識結果修正型入力を行う場合の方が多くの時間を要することが分かった。したがって、文書の入力時間を短縮するためには、従来のような認識誤りに対する認識結果修正型入力に依存する、画一的な入力方式を改善しなければならない。すなわち、正読率のような認識精度ごとに、文書全体としての入力時間が最短となる確認修正手段を適切に用いる必要がある。特に FAX を利用した受注業務等においては、記入者ごとに受注伝票の文字の品質が大きく異なり、オペレータの目視による品質の見極めは、負担の増加を招くこととなる。すなわち、ある程度のまとまった量の認識結果を確認した後に確認修正手段選択の判断を行う場合、すでに確認修正を行っているのと等価となる。それゆえ、一括削除後に認識結果を用いずかな漢字変換機能のみに依存して入力を行う非認識型入力は、実は余分な作業を強いられていることとなり、入力時間の増大すなわち入力効率の低下につながる。

したがって、オペレータの負担を軽減し、かつ効率的な文書入力を行うためには、(1) すべての認識精度にわたって入力特性（たとえば入力速度）が既知である確認修正手段を数種類用意しておき、(2) 取り込まれた文書の文字認識結果に基づいて認識精度を自動推定し、(3) 得られた認識精度の値から最も高い入力速度となる確認修正手段を自動選択することが必要であると考える。

以上の結果を基に、本提案方式の概略を図 1 に示す。確認修正手段 M_1 , M_2 , M_3 を選択対象とし、あらかじめ入力速度を測定しておく。簡単のため、確認修正手段数を 3 つに限定する。 M_1 , M_2 , M_3 の入力速度を認識精度 z (%) の関数として $V_1(z)$, $V_2(z)$, $V_3(z)$ と定義する。図 1 に示すように $z_k \leq z$ におい

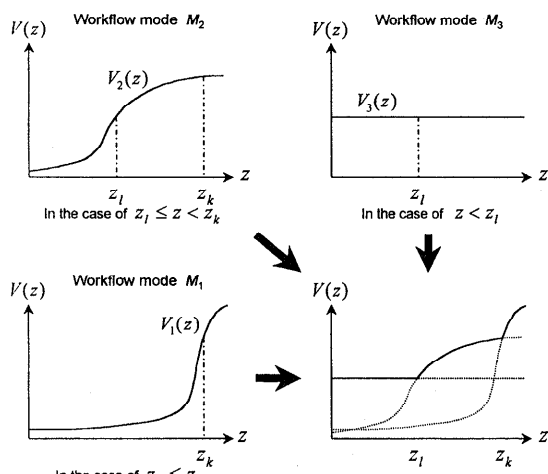


図 1 提案方式の概念

Fig. 1 The concept.

て、 $V_1(z)$ が最大入力速度を示す場合、この領域では確認修正手段 M_1 を選択する。ところが、 $z_1 \leq z < z_k$ では、 $V_1(z)$ と $V_2(z)$ の優劣が逆転し、 $V_2(z)$ が最大となるので、確認修正手段 M_2 を選択する。同様の考え方により、 $z < z_1$ では確認修正手段 M_3 を選択する。このように、入力速度の優劣の変化を選択条件として、入力速度が最大となるように確認修正手段を選択すれば、たとえば確認修正手段 M_1 のみによって修正する場合と比較して、 $z < z_1$ の領域での入力速度が向上することになる。したがって、認識精度 z の領域すべてにわたり、入力速度を向上させることが可能となる。

3. 認識精度の推定

本提案では、確認修正手段の選択に必要な認識精度を、文字認識結果を用いて推定する。

本来、認識精度は認識結果候補と修正後に得られる正解文字とを照合することによって求まる。ところが本方式では、認識精度に応じて修正のための方法を提供することが目的であるため、修正に先立ってあらかじめ認識精度を知っておく必要がある。そこで、これらを推定によって求めることとする。

本章では、推定法について、その考え方を簡単に述べる。

3.1 可搬性を考慮した推定法

本論文で提案している確認修正手段の切替手法およびそれを適用した文書入力システムは、文字認識アルゴリズムやその出力結果に依存しないことが望ましい。特に認識精度の推定法については必須の条件といえる。なぜなら、仮に推定法が認識アルゴリズムに強く依存

表 1 閾値の設定条件
Table 1 The condition of threshold.

正解・誤り文字の出現数 (左辺) と 0 (右辺) との関係		閾値の更新計算
正解文字	誤り文字	
等しい (=)	等しい (=)	更新せず (既定値のまま)
大なり (>)	等しい (=)	正解文字の距離値の最大値 $\times 1.1$
等しい (=)	大なり (>)	誤り文字の距離値の最小値 $\times 0.9$
大なり (>)	大なり (>)	正解文字の距離値の平均値と誤り文字の距離値の平均値との平均

する場合、正読率向上にともなってアルゴリズム変更を行うたびに、システムの内部変更が必要となり、コストの増大を招くからである。

そこで本節では、認識アルゴリズムの変更に対して柔軟に対応可能な認識精度推定法について説明する。推定のための情報源として、認識結果の距離値¹²⁾や入力イメージの外郭形状¹³⁾が考えられる。ここでは、多くの認識手法にとって出力が可能である距離値を用いることとする。

以下に述べる推定法は、距離値の範囲には依存しない。また、記号や JIS 第 2 水準に含まれる一部の文字等で、認識辞書に含まれていないような文字種に対しても依存しない構成になっている。したがって、距離値を出力する認識手法であれば、容易に変更可能となる。

推定の基本的考え方は、まず 1 文字単位でその確からしさ (正解または誤り) を推定し、次に注目しているブロック単位で集計して推定認識精度を求めることである。ブロックとは、段落やページ、帳票ならばその記入欄等である。確からしさの推定は、認識結果候補文字の距離値と、あらかじめ求めておいた閾値とを比較することにより行う。すなわち、認識結果候補文字の距離値が、閾値以下であれば正解とする。逆に、閾値を超えていれば誤りとする。また、正解の文字数をブロックに記入されている文字数で割った値 ($\times 100$) を推定認識精度とする。後述するが、本論文では特に、実際の受注業務等で需要の多い枠ありの場合を対象としているため、文字数のずれの問題は生じない。枠なしの場合については今後の課題である。

以下に推定の基本である、文字ごとの確からしさの導出法について示す。

上述した閾値を求めるため、まず認識結果距離値に基づいてテーブルを作成する。次にこのテーブルを基に、全文字コードと閾値とを対応させた辞書を作成する。本論文で述べる文書入力システムにより実際の入力作業を行う際には、この閾値辞書のみを用いる。前述のテーブルには、認識辞書の文字種に含まれていない場合でも、全文字種 (記号、英数字、かな、JIS 第

1・2 水準漢字等)、およびそれらに対応して、文字コード、閾値、正解および誤り文字のそれぞれの出現数、正解文字の距離値の最大値および平均値、誤り文字の距離値の最小値および平均値が含まれる。閾値は、あらかじめ大雑把に求めておき、初期値として格納しておく。この閾値を、実漢字データの認識結果を用いて最適化していく。新規に閾値を最適化する場合は、出現数を 0 に設定しておく。

閾値の最適化は、以下のように行う。まず、候補文字の認識結果と正解文字とを照合し、正解・誤りに応じて、正解および誤り文字出現数をそれぞれ増加させる。また、正解文字であった場合には、それまでに出現した距離値の最大値を求めてテーブルの値を更新し、同様に誤り文字であった場合には、距離値の最小値を求めて値を更新していく。さらに、正解文字または誤り文字の出現数が 1 増えるごとに、それぞれ対応する候補文字について距離値のその時点までの平均値を求めて更新する。これら一連の更新処理は 1 文字ごとに行う。次に、このテーブルの出現数に応じて閾値の設定を行う (表 1)。最後にこれら閾値を、各文字コードと対応させて辞書とする。

本節の最初に述べたように、認識アルゴリズムを変更したことにより距離値の値やその範囲等の特性が変化する場合には、上記の処理を経て閾値辞書を作成し直すことになるが、システム内部に立ち入らずに済むため、容易かつ柔軟に対応可能である。

3.2 評価結果と考察

前節で述べた推定法を用いて、推定精度の評価を行った結果を図 2 に示す。評価に用いたデータは、実際の受注業務において、複数の FAX によって送信されてきた手書き文書 500 枚 (80 文字/枚) 中の漢字である。これらは不特定多数の記入者によって記入されており、また、楷書・行書、JIS 第 1・第 2 水準等といった区別はなく、様々な種類・品質で構成されている。まず、これら文書のうち 450 枚を用いて、外郭方向寄与度特徴に基づく認識アルゴリズム¹¹⁾により認識処理を行う。そして、これらの結果を用いて、前節で述べた処理に基づいて閾値辞書を作成する。次に、

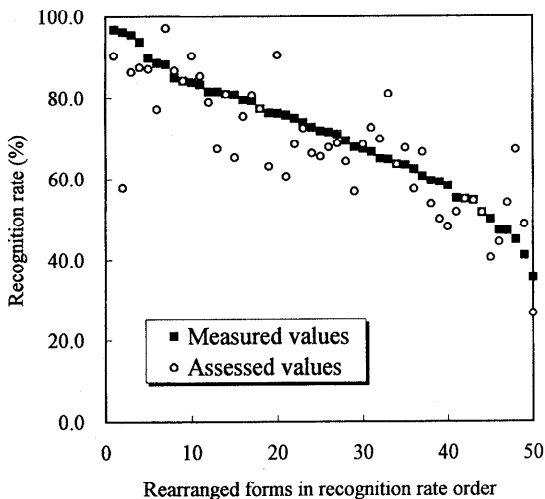


図 2 実測および推定認識率の比較

Fig. 2 Measured values and assessed values.

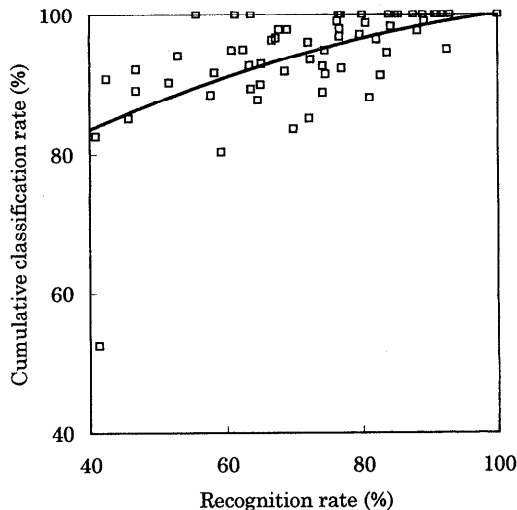


図 3 文字認識率と正解候補包含率の関係

Fig. 3 The relationship between recognition rate and cumulative classification rate.

残り 50 枚に含まれる漢字を未知データとして入力し、前節で述べた推定法による推定正読率、および正解文字と比較して求めた実正読率を求める。図 2 では、これら実正読率と推定正読率を縦軸にとり、実正読率の高い順に文書を並べ替えて横軸とした。図 2 には、第 1 位の正読率の評価結果のみ示したが、第 1 位の正読率と第 n 位までに正解候補が含まれる率を表す正解候補包含率 (第 n 位包含率) の関係を用いて、第 n 位候補文字の推定認識精度を求めることも可能である。図 3 では、1 文書に対して第 1 位実正読率と第 20 位実包含率をそれぞれ求め、全文書に対して求めた値をプロットしてある。これらを曲線近似することにより、正読率と包含率の変換を行うことが可能となる。

ここで推定精度および実正読率と推定正読率の一致度を求める。推定精度は、正しく認識できた文字を正しいと推定した場合の文字数、および誤って認識した文字を誤りであると推定した場合の文字数の和を求め、この和の認識対象文字数に占める割合として定義する。また、一致度は、実正読率に対する推定正読率の割合として定義する。得られた認識結果に基づいて全文書データの平均を計算すると、推定精度は 76.3%、一致度は 96.4% である。推定正読率には、実際には誤り文字であるにもかかわらず、正しいと推定する場合がわずかながら含まれるため、一致度もわずかながら低下する。推定精度は必ずしも高いとはいえないが、一致度は十分な精度が得られている。本提案では、確認修正手段は推定正読率の値によって選択されるため、この値が実正読率の値に近ければ近いほど良い。すなわち、高い一致度が得られていれば、高精度に確認修正

手段を選択することが可能となる。実際の文書入力システムを構築する場合は、後述するように、確認修正手段は 2 つから 3 つ程度が適切であり、そのようなシステム構成にすれば、それぞれの確認修正手段に対応する認識精度の領域はそれぞれある程度の幅を持つため、推定正読率の誤差に対する実用上の影響は少ないと考えられる。

4. 確認修正手段の切替

図 4 に本方式の実現例を示す。すなわち、確認修正手段の特性を評価 (4.1 節) して選択条件を設定し、この条件と推定認識精度に基づいて確認修正手段を選択 (4.2 節) する。

4.1 確認修正手段の特性評価

選択対象となる確認修正手段の特性は、正読率や包含率等の認識精度を表す尺度に依存する。そのため、特性を評価するには、確認修正手段ごとに尺度を決めておく必要がある。尺度が 1 つである場合は、2 章で述べた方法がそのまま適用できる。ところが、異なる複数の尺度上で特性が得られた場合には、そのままでは単純な比較はできない。そこで、本節では異なる尺度上の特性を同じ尺度上の特性に変換する方法を述べる。

説明のため、尺度として正読率と包含率を、特性として入力速度を想定する。表 2 に、これらの尺度の依存性によって確認修正手段を分類した例を示す。I 型の例として、認識結果修正型入力あげられる。II 型の場合、認識候補の第 1 位が正解でなくとも、第 2 位

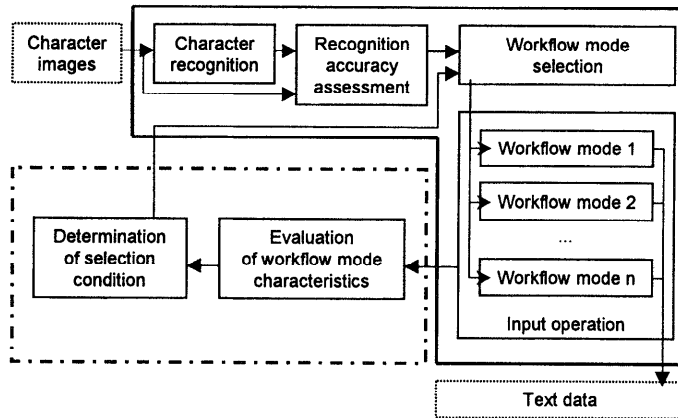


図4 文書入力方式の構成例
Fig. 4 An example of document-input system.

表2 確認修正手段の分類
Table 2 The classification of workflow modes.

確認修正手段の型	認識精度を表す尺度の依存性
I	文字認識率
II	正解候補包含率
III	(依存性なし)

以下に正解が含まれる率が高いことに注目した知識処理型の確認修正手段が考えられる。III型の例として、かな漢字変換によって直接入力する通常のワードプロセッサ型の確認修正手段があげられる。

選択対象とする確認修正手段を M_1, M_2, \dots, M_n とし、I, II および III 型に属する確認修正手段ごとに入力速度を測定する。以下簡単のため、正読率 x の全領域で最も入力速度の大きいI型の確認修正手段を M_1 とする。同様に、包含率 y の全領域で最も入力速度の大きいII型の確認修正手段を M_2 とする。また、 x, y の全領域で入力速度が一定であるIII型の確認修正手段を M_3 とする。それぞれの入力速度が $V_1(x), V_2(y)$ および V_3 と求められたとすれば、これらの優劣を議論すればよい。

$V_1(x)$ と $V_2(y)$ を同一の尺度上で比較するために、 $V_2(y)$ を x の関数 $V_2(x)$ に変換する。そのために x と y の関係を用いる。この関係は、実際の手書き文書を用いて正読率と包含率を求めることにより得られる(図3)。本方式では、図3で示す近似値に基づいて、異なる認識精度間の変換が近似的に行えればよい。仮にその変換時に誤差が存在したとしても、入力作業中に動的に補正可能である。この関係を用いて、 $V_2(y)$ から $V_2(x)$ が得られれば、同じ尺度上で $V_1(x)$ と $V_2(x)$ を比較することが可能となる。 $V_2(y)$ と V_3 および V_3 と $V_1(x)$ については、 V_3 が x にも y にも

依存しないため、比較は容易である。

尺度が3つ以上存在する場合に対しても、 $V_1(x)$ と $V_2(y)$ の比較と同様に尺度間関係を調べることにより、同じ尺度上で比較することが可能になる。

4.2 確認修正手段選択法

確認修正手段の選択条件は入力特性を比較して設定する。説明のため、確認修正手段を前節の M_1, M_2 および M_3 に限定する。特性として、入力速度を想定し、上記の確認修正手段に対応してそれぞれ $V_1(x), V_2(y)$ および V_3 とする。

4.1節で求めた入力速度 $V_1(x)$ と $V_2(x)$ の優劣が、正読率 x 上の値 x_i によって逆転し、以下の関係が得られたとする。

$$V_2(x) \leq V_1(x) \quad (x_i \leq x)$$

$$V_1(x) < V_2(x) \quad (x < x_i)$$

この場合、確認修正手段 M_1 と M_2 の選択は、値 x_i によって決まる。同様に、入力速度 $V_2(y)$ と V_3 の優劣が、包含率 y 上の値 y_j によって変化し、以下の関係が得られたとする。

$$V_3 \leq V_2(y) \quad (y_j \leq y)$$

$$V_2(y) < V_3 \quad (y < y_j)$$

この場合、確認修正手段 M_2 と M_3 の選択は、値 y_j によって決まる。以上の方法により、選択条件は認識精度上の値 x_i, y_j として求められる。

なお、ここで $y_j < x_i$ が成り立っていれば、入力速度 $V_1(x)$ と V_3 の比較を行う必要はなく、3つの確認修正手段 M_1, M_2 および M_3 すべての選択が可能となる。逆に、 $x_i < y_j$ の場合は、 $V_1(x)$ と V_3 も比較する。これは次の理由による。 $x_i < y_j$ が成立している場合、 M_1 と M_3 の入力速度の優劣は、 x_i と y_j の間の値 (u_k とする) で逆転することになるため、 u_k

表3 確認修正手段の選択
Table 3 The workflow mode selection.

確認修正手段	選択尺度	
	文字認識率 (%)	正解候補包含率 (%)
M_1	$p \geq x_i$	-
M_2	$p < x_i$	$q \geq y_j$
M_3	-	$q < y_j$

より小さい x_i で切り替える M_2 は、つねに M_3 より入力速度が小さくなる。したがって、 u_k 以上の領域で M_1 、 u_k 未満の領域で M_3 を選択することになる。このとき、 M_2 は不要となる。

確認修正手段は、推定によって得られた認識精度と、前述した選択条件に基づいて選択される。説明のため、認識精度として正読率、包含率を想定する。推定により、正読率 $p\%$ 、包含率 $q\%$ が得られたとする。この場合の選択条件の値 x_i 、 y_j に基づく確認修正手段選択法の例を表3に示す。

推定値のうち、まず正読率に着目する。正読率が非常に高い場合は、第1位候補のほとんどが正解であるため、わずかな誤りを修正すればよい。そこで、確認修正手段として M_1 を選択する。ところが、正読率が低い場合には、 M_1 による入力には認識誤りの修正に多くの時間を費やすのみであるから、他の確認修正手段を選択することになる。

正読率が低くとも、包含率が高い場合には、第2位以下に含まれている正解候補を利用することが可能であるため、知識処理型の確認修正手段 M_2 を選択する。正読率、包含率ともに低い場合には、認識結果候補を利用した入力の効果はきわめて低いため、認識結果に依存しない確認修正手段 M_3 を選択する。

以上の説明から、確認修正手段の選択条件は、 x_i と y_j という2つの値を求めさえすればよいことになる。他の方法として、ある推定認識精度の値に対して $V_1(x)$ 、 $V_2(y)$ および V_3 の最大値を与える確認修正手段を選択することも考えられるが、この場合は、認識精度の値と、最大入力速度を示す確認修正手段を対応づけた表を、あらかじめ作成しておく必要がある。しかも、すべての認識精度に対して必要となる。そこで本提案では、表3に示すように、より単純で少数の条件で選択できるようにした。さらに、このような条件を用いることにより、実際の確認修正作業中に、時間帯やオペレータのスキル等により変化する入力速度に応じて、選択条件を動的に補正することも容易となる。

ここで、認識精度の推定精度が問題になるが、実際には高精度な推定結果が必要とされることは少ないと考えられる。すなわち、認識精度の全領域にわたって、

たとえば選択対象となる確認修正手段が数種類ある場合、それぞれの確認修正手段が割り当てられる認識精度の領域は、ある一定の幅を持つため、ある程度の認識精度推定誤差、場合によっては数十%の誤差であっても正しい確認修正手段選択が可能である。

以上説明したように、2つの尺度と2つの選択条件の設定値により、3つの確認修正手段の選択が可能となる。ここでは尺度を2つ用いたが、さらに増やすことも可能である。例として、尺度を1つ追加する場合を考える。ただし、説明のため、追加する尺度 w を縦軸にとり、包含率 y を横軸にとった場合に、図3と同様な関係 ($y \leq w$) があるとする。この場合、表3の包含率 y による確認修正手段 M_2 と M_3 の選択と同様に、 y より w の方が高いことに注目した確認修正手段を選択することが可能となる。すなわち、新たな確認修正手段 M_4 を追加し、 w 上で選択条件の設定値を求めれば、確認修正手段 M_3 と M_4 の選択が可能となる。

上記の例では、1つの尺度の追加にともない、確認修正手段も1つ追加したが、尺度を追加せずに確認修正手段を追加することも可能である。すなわち、たとえば、正読率のような1つの尺度に対して選択条件の設定値を2つ決定すれば、3つの確認修正手段を対応づけることができる。

以上の考察から、一般に確認修正手段数 n_M 、選択条件の設定値の数 n_S および尺度数 n_C との間に以下の関係式が成立することが分かる。

$$n_M = n_S + 1 \geq n_C + 1.$$

したがって、この式を満たす範囲内で認識精度の尺度や確認修正手段を追加することができるため、より柔軟な確認修正手段選択を可能としている。

5. システムの構成および評価実験

本方式における入力時間短縮の効果を明らかにするため、手書き文書の入力実験を行った。以下に、その概要、評価方法および結果を示す。

5.1 実験システムの構成

本提案方式に基づいて、評価モデルを構築する。そのために、選択のための尺度および確認修正手段を具体的に設定する。尺度として、表3に示した正読率および正解候補包含率を採用する。また、確認修正手段として3つ採用する。すなわち、(1) 第1位認識候補に対して確認を行いながら修正していく認識結果修正型入力 (M_1)、(2) 認識候補の第1位が正解でなくとも、第2位以下に正解が含まれている率が高いことに注目した同音語補正入力 (M_2)、(3) 候補中に正解が

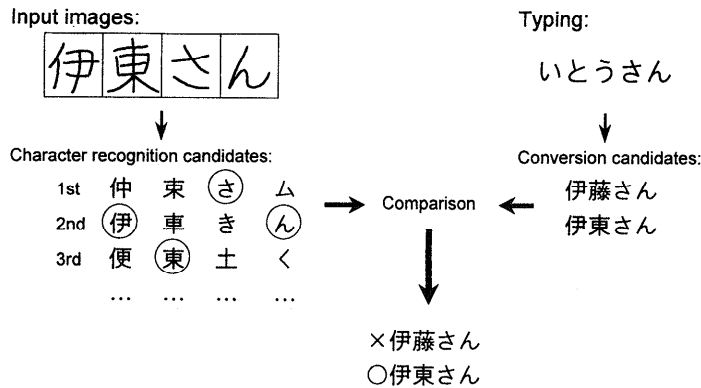


図 5 同音語補正

Fig. 5 The homonym correction.

含まれている率が低い場合に、認識結果を用いずに入力し直す非認識型入力 (M_3)、である。

ここで、同音語補正入力について概略を説明しておく。この確認修正手段は、現状のかな漢字変換入力における同音語の選択に関して、入力時間の大幅な増大という問題を解決する確認修正手段の1つである。かな漢字変換の結果、複数の同音語候補が現れた場合、これらの候補と文字認識候補との照合を行う。読みが同じでも字形が異なる文字が認識候補中に正解として含まれていれば、複数の同音語候補から正しい結果を自動選択することが可能となる。同音語補正の例を図5に示す。入力イメージとして入ってきた「伊東」の認識結果として、図5に示す文字候補が得られたとする。一方、オペレータによりかな入力された「いとう」に対してかな漢字変換を行った結果、「伊藤」および「伊東」という同音語候補が得られたとする。そこで、これらと認識結果候補との照合を行う。候補中には「東」の文字が含まれているため、正しく「伊東」を得ることができる。

前述の3つの確認修正手段を用いて、それぞれ独立に入力時間を測定する。その結果、それぞれの確認修正手段が切り替わる認識精度の値が存在すれば、これらを選択条件として自動選択が可能となる。さらに、この場合に選択される確認修正手段は、他の確認修正手段より入力時間が短いため、結果的に入力時間短縮の効果として現れる。したがって、独立に測定を行えば十分である。

また、確認修正手段の自動選択を行うためには、認識精度の推定値に基づいた確認修正手段選択が精度良く行われる必要がある。したがって、確認修正手段選択精度の評価もあわせて行う。

評価には、3.1節で述べたデータ、すなわちFAXで

表 4 評価データ諸元
Table 4 The data features.

評価対象	FAX 注文受付票 100 枚
記入欄項目 (平均記入文字数)	氏名 (5.1) 会社名 (14.2) 本文 (55.9)
記入枠	あり (10 mm × 10 mm)

受信した実際の受注業務で得られた手書き文書の漢字データを用いた。また、適用認識方式も3.2節と同様、外郭方向寄与度特徴に基づく認識アルゴリズム¹¹⁾を用いた。漢字データの特性等もすでに3.1節で述べたとおりである。データ諸元を表4に示す。

以下に具体的な評価方法を示す。

(a) 確認修正手段選択精度の評価

認識精度の推定値および実測値に基づいて、表3から求めた確認修正手段をそれぞれ比較する。

(b) 入力時間短縮効果の評価

3つの確認修正手段による入力時間測定は、次のとおり行う。まず、受注伝票の記入欄に到達するのに要するカーソル移動時間は、どの確認修正手段でも同一記入欄に対しては同一である。そこで、あらかじめ求めておいた平均カーソル移動時間を測定値から除外する。次に、同一受注伝票に対して各確認修正手段を用いてそれぞれ2回入力を行い、2回目の入力時間を採用する。1回目の入力作業でオペレータが文章を覚えたとすれば、2回目の作業では入力時間の短縮化がかなり限界に近づいていると考えられる。そこで、この2回目の値をそれぞれの確認修正手段のデータとして採用することで、手段相互間の学習効果のばらつきが抑制され、相

表 5 確認修正手段選択の精度

Table 5 The accuracy of workflow mode selection.

ID No. の差	-2	-1	0	1	2
頻度 (%)	0.00	3.80	84.81	11.39	0.00

(注) ID No. 1: 認識結果修正型入力, 2: 同音語補正入力
3: 非認識型入力

対的に打ち消されたことと等価になると考えられる。本評価では、確認修正手段相互で入力時間を比較することに重点を置き、あえてこの方法を用いた。なお、実験に携わったオペレータは1人であるが、入力を行う時間帯や日によって、入力時間にかなりのばらつきが生じるため、時間帯を変えながらかつ数カ月にわたって実験を行ったことで、ある程度は多数のオペレータによる入力を再現できたと考える。また、オペレータの入力速度は、平均40字/分である。多数のオペレータによる評価は今後行っていく予定である。

5.2 評価結果および考察

表5に、確認修正手段選択の結果を示す。確認修正手段に割り当てられた番号(ID No.)を基に、推定認識精度から得られた確認修正手段のID No.から、実測認識精度から得られた確認修正手段のID No.を引いたものをID No.の差としてある。また、それぞれの場合に得られた選択度数を、全体の度数に対する割合として示した。表5から、確認修正手段はほとんど正しく選択できていることが分かる(ID No.の差が0の場合)。ここで問題となるのは、番号の差が非零の領域、すなわち、高い認識精度であるにもかかわらず、低い認識精度向けの確認修正手段が選択される場合、およびその逆である。前者の場合、正しく認識できている文字に対しても改めて入力し直すため、認識結果の有効性が損なわれる可能性がある。また、後者の場合、同音語補正入力とうまく機能しない、認識結果修正型入力を行うには文字数が多すぎる等の問題により、入力時間は短縮どころか、かえって増大してしまうことが予想される。しかし、番号の差が非零の領域が全体に占める割合は非常に小さいため、文書入力の効果全体に与える影響は少ないと考えられる。

図6に、入力時間短縮効果の評価例を示す。各確認修正手段ごとの実測入力時間を基に入力速度を計算し、非認識型入力に対する相対値で表現してある。具体的には、得られた入力時間からカーソルの平均移動時間を除外し、逆数を求めることで、単位時間あたりの入力文字数(入力速度)を求める。図6から、正読率80%近傍で、同音語補正入力に対して認識結果修

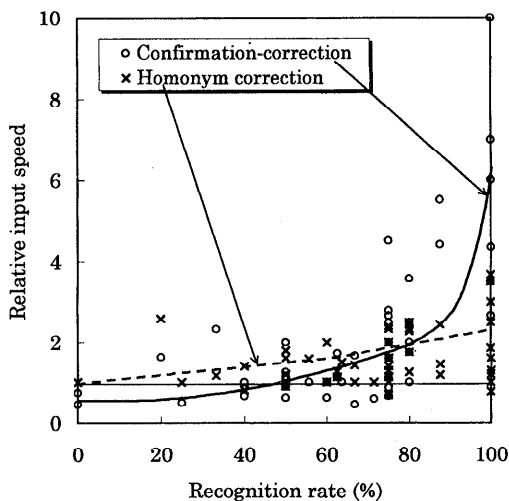


図 6 相対入力速度

Fig. 6 The relative input speed.

表 6 入力速度の比較

Table 6 The comparison of input speed.

	入力速度 (入力文字数/秒)
認識結果修正型入力のみ	0.744
3 確認修正手段切替え	0.971

正型入力の入力速度が逆転することが分かる。すなわち、実測値から選択条件の設定値を求めることにより、認識精度に応じた確認修正手段の最適な選択が可能となる。ここで、認識精度の推定結果を用いなくて入力を行う認識結果修正型入力のみにより入力を行った場合と、認識精度に基づいて3確認修正手段を適切に切り替えて入力を行った場合の入力速度を表6に示す。認識結果修正型入力は、OCR出力に対し、確認と修正を繰り返しながら入力する方法としてはきわめて一般的な入力形態であり、通常はこの確認修正手段により入力が行われる。表6より、認識結果修正型入力のみの入力方式に対して、本方式により3確認修正手段を適切に選択した場合、30.5%の入力速度向上という結果が得られた。本来、認識精度の低い領域においては、認識結果を利用しない非認識型入力が選択されるのであるが、図6では、同音語補正入力が選択されている。これは、試行回数が少ないことに起因していると考えられる。なお、この場合は、同音語補正入力と非認識型入力のどちらが選択されても問題ない。

また、3.2節で説明したように、確認修正手段選択において、認識精度の高精度な推定結果が必要とされることは少ない。しかし、確認修正手段が切り替わる認識精度点の近傍では、精度が要求される場合がある。すなわち、確認修正手段切替えの近傍で、認識精

度（たとえば正読率）に対する入力速度の変化が大きい場合、推定認識精度のわずかな誤差により入力速度に大きく影響を与えることとなる。この点に関しては、推定精度を向上させる等の対策が必要となり、今後の課題の1つとしたい。一方、認識精度に対する入力速度変化の割合が小さい場合には、精度推定の誤差から受ける影響も小さい。したがって、本評価で用いた3種類の確認修正手段を用いる場合には、認識精度の広範囲で入力速度変化の割合は小さいため、システム全体のパフォーマンスを低下させる原因とはなりえないといえる。

6. む す び

文書入力時間短縮を目的として、認識精度に対応して最適な確認修正手段を選択することにより、全体として入力時間を短縮させる文書入力方式を提案した。また、本提案方式の評価モデルを用いて、入力時間短縮の効果を検証する実験を行った。従来のような単一の確認修正手段のみを用いる入力方式と比較すると、本方式では、およそ30%の入力速度向上という結果が得られ、入力時間短縮に有効であることが確認できた。

本論文で提案した方式では、確認修正手段に対応する認識精度の領域に幅が存在するため、推定精度にある程度の誤差が存在するにもかかわらず、高精度に確認修正手段選択を行うことが可能である。

さらに、文字認識技術の向上によるアルゴリズム変更にも柔軟に対応できる。この場合、認識精度の高精度化はそのまま入力時間短縮に反映される。一方、低い認識精度の文書に対する入力の必要性は依然として存在するため、確認修正手段選択の有効性は維持される。本方式の他の特徴を以下に要約する。

- (a) 認識精度の関数として入力特性をあらかじめ測定可能な確認修正手段であれば、任意の確認修正手段を自動選択の対象として導入することができる。
- (b) 選択対象となる確認修正手段相互の選択条件を任意に設定できるため、導入した確認修正手段の性能改善に柔軟に対応することができる。
- (c) 使用する文字認識アルゴリズムとして、任意のアルゴリズムを用いることができる。この場合、認識結果に対する確からしさを表す量（たとえば、標準パターンとの距離値を表現する数値）が出力できればよい。

今後、低い認識精度を持つデータを用いた評価を継続し、この精度での効果を明確にする必要がある。また、オペレータの人間の側面にも目を向け、特にオペ

レータのスキルに応じて確認修正手段選択の誤差を動的に補正する場合の効果等を、多数のオペレータによって明らかにしていく予定である。

謝辞 本研究の遂行にあたりご指導、ご討論いただいたNTTヒューマンインタフェース研究所の中野博隆部長をはじめとするメディア応用システム研究部の各位、また、データ収集にご協力いただいたNTT電報事業本部システム部の堂領伸一郎氏、NTTアドバンステクノロジ株式会社の橘美雪氏、岩崎尚子氏に深謝する。

参 考 文 献

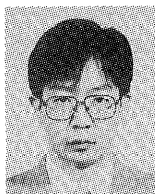
- 1) 萩田紀博：手書き文字認識の研究動向，*NTT R&D*, Vol.45, No.11, pp.1189-1194 (1996).
- 2) Wakahara, T. and Odaka, K.: Adaptive Normalization of Handwritten Characters Using Global/Local Affine Transformation, *Proc. 4th Int. Conf. Document Anal. Recognition*, pp.28-33 (1997).
- 3) Hildebrandt, T.H. and Liu, W.: Optical Recognition of Handwritten Chinese Characters: Advances Since 1980, *Pattern Recognition*, Vol.26, No.2, pp.205-225 (1993).
- 4) 野地幸一, 竹内充弥, 渡辺 敏, 中村 修: サービスオーダ投入の効率化と品質向上を目指して, *NTT 技術ジャーナル*, Vol.4, No.8, pp.14-16 (1992).
- 5) 加藤隆仁, 横澤一彦: 手書き文字品質の定量評価, *信学論 (D-II)*, Vol.J75-D-II, No.9, pp.1573-1581 (1992).
- 6) Blando, L.R., Kanai, J. and Nartker, T.A.: Prediction of OCR Accuracy Using Simple Image Features, *Proc. 3rd Int. Conf. Document Anal. Recognition*, pp.319-322 (1995).
- 7) Gonzalez, J., Kanai, J. and Nartker, T.A.: Prediction of OCR Accuracy Using a Neural Network, *Proc. Int. Assoc. Pattern Recognition Workshop Document Anal. Syst.*, pp.323-337 (1996).
- 8) 松村季樹, 倉掛正治, 木村義政, 小林公知, 中村修, 稲垣充廣: 文字認識精度に対応した日本語文書入力方式, *信学総大*, D-4721, p.260 (1996).
- 9) 松村季樹, 小林公知, 橘 美雪, 中村 修: 文字認識精度に対応した文書入力手段の選択方式, *信学総大*, D-345, p.348 (1996).
- 10) 堤田敏夫, 城戸 賛, 太田一浩: 手書き漢字画像データベース「IPTP CD-ROM2」とその諸性質の分析結果, *信学総大*, D-12-971, p.304 (1997).
- 11) 萩田紀博, 内藤誠一郎, 増田 功: 外郭方向寄与度特徴による手書き漢字の識別, *信学論 (D)*, Vol.J66-D, No.10, pp.1185-1192 (1983).
- 12) 梅田三千雄: 手書き漢字認識研究の歩み, *信学*

技報, PRU95-120 (1995).

- 13) 萩田紀博, 増田 功: 大局的・局所的線密度を併用した手書き漢字の分類方式, 信学論 (D), Vol.J65-D, No.6, pp.734-741 (1982).

(平成 10 年 6 月 29 日受付)

(平成 10 年 12 月 7 日採録)



松村 季樹 (正会員)

昭和 63 年東北大学理学部物理学科卒業。平成 2 年同大学大学院理学研究科博士課程前期課程修了。同年日本電信電話 (株) (現 NTT) 入社。以来, 人間の認知特性を利用したマンマシンインタフェースの研究, 文字認識応用システムの研究開発, 映像情報構造化の研究に従事。現在, NTT インテリジェントテクノロジー (株) 勤務。



小林 公知

昭和 49 年東京理科大学理 II・物理学科卒業。昭和 43 年日本電信電話公社 (現 NTT) 入社。以来, 図形発生, ファクシミリ通信網におけるセンタエンド形メディア変換処理装置の研究・実用化, ファクシミリ入力による文字認識応用システムの研究開発に従事。昭和 60 年オーム技術賞受賞。現在, NTT アドバンステクノロジー (株) 勤務。電子情報通信学会会員。



中村 修 (正会員)

昭和 49 年青山学院大学工学部電気電子工学科卒業。同年日本電信電話公社 (現 NTT) 入社。以来, 記憶制御方式, 並列処理方式, 文字認識応用システムの研究開発に従事。現在, NTT サイバーソリューション研究所勤務。工学博士。電子情報通信学会, IEEE 各会員。



小倉 健司

昭和 49 年千葉大学工学部電子工学科卒業。同年日本電信電話公社 (現 NTT) 入社。以来, ビデオテックス等画像通信システムの研究開発, 画像符号化の標準化, エレクトロニックコマース等の各種マルチメディアシステムの開発, 文字認識応用システムの研究開発に従事。現在, NTT インテリジェントテクノロジー (株) 勤務。電子情報通信学会, 映像情報メディア学会各会員。