

音声認識併用による手書き文章処理アルゴリズム

1. 背景

紙状の文字を、検索やデータマイニング等、電子データのように扱いたい。例えば「昨日の会議の書類」など、特定の書類を探したい場合、大量の書類の中から一つ一つ目で確認しながら探す必要があるなど、現状の紙ベースのメディアのユーザビリティは高いとはいえない。紙状の文字を電子データ化できれば、キーワード検索機能による文書探索の大幅な高速化が可能となり、ユーザビリティの向上が期待できる。紙状の文字を電子化する手法として光学文字認識(OCR)手法があり、近年では高い認識率を実現している。しかし、現状の認識率では認識誤りの修正に多くの時間がかかるため、広く普及するには至っていない。そこで、既存のOCR技術に音声認識を併用し、それぞれの認識結果を統合することで、OCRの問題点を改善するシステムの構築を目指す。

2. 目的

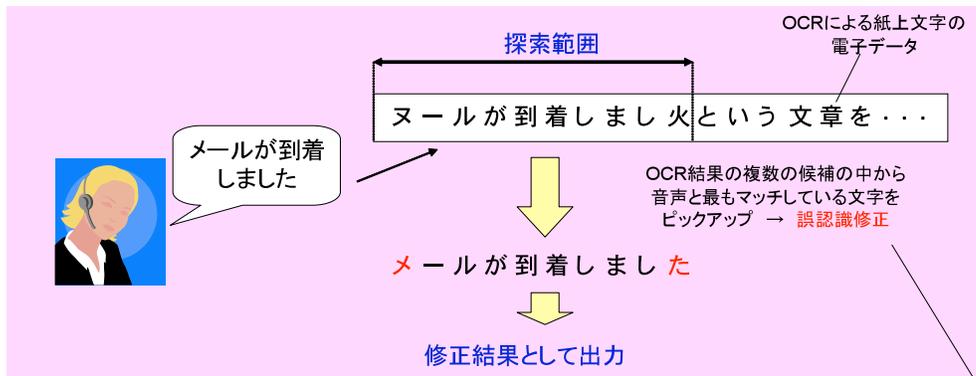
OCRの問題点を改善するシステムの構築を目指し、音声認識とOCRの認識結果を効率的に組み合わせるアルゴリズム、LiveBandsアルゴリズムを構築する。さらに提案アルゴリズムのアプリケーションとして、OCRの誤認識を含む単語を読み上げることで誤認識箇所を自動で検索、修正できるアプリケーションを開発する。本アルゴリズム及びシステムにより、従来ユーザビリティの低さよりあまり重要視されていなかった紙ベースメディアを最大限に有効利用できる環境を作る。

3. 開発の内容

音声情報と文字情報を統合するシステムとして、文章認識による文章書き起こしの他に、キーワード検索が考えられる。ここで、文章書き起こし問題は、部分空間(単語)ごとの最適筆記候補を探索し、文意に沿ってそれらを組合せる問題といえる。キーワード探索問題は、広い探索空間の中からキーワードにマッチする筆記を探索する問題と言える。つまり、図1に示すように、これらの問題は探索空間が異なるだけで、根本は「最適筆記を探索する問題」という1つの問題を共有する。しかし従来は、このような点に注目したシステムは無く、一般性に欠けていた。

本プロジェクトでは、そのような汎用性を持った、「Live Bands を用いた最適筆記アルゴリズム」を確立した。

文章書き起こし問題



キーワード検索問題

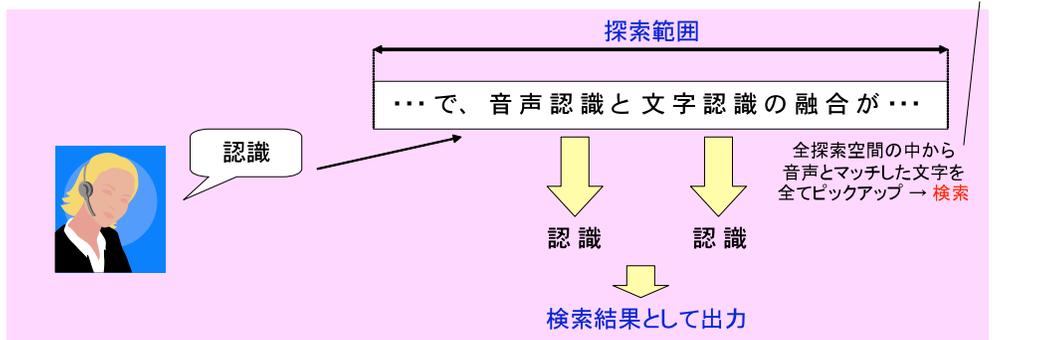


図 1. 文章書き起こし問題とキーワード検索問題の共通の問題

提案アルゴリズムは、音声認識のビーム探索時の適切な枝刈りや、検索時の途中計算結果の再利用性を用い、精度、速度の向上だけでなく、日本語特有の漢字や平仮名の混合文字(例：大おか山など、以下混ぜ書きとする)の正確な筆記書き起こし、キーワード検索への応用を可能とする。探索幅は任意に設定でき、探索幅を狭めて逐次的に探索を行うことで、混ぜ書きの正確な書き起こしが可能となる。探索幅を広げれば、文中から最も発音に近いキーワード句を検索することが出来る。リアルタイム適用が可能なほど計算量が小さく、音声と文字を統合したシステムのスタンダードなアルゴリズムとなる可能性がある。

さらに提案アルゴリズムのアプリケーションとして、OCR の誤認識を含む単語を読み上げることで誤認識箇所を自動で検索、修正できるアプリケーションを開発した。従来のOCRでは誤認識は全てキーボードによるリタイピングによって修正していたため、時間がかかってしまいがちだった。本アプリケーションによって、音声を用いることによる早く、容易でかつキーボードレスな誤認識修正が可能となる。

図 2 に本プロジェクトで作成した GUI を示す。本アルゴリズム及びシステムにより、従来ユーザビリティの低さよりあまり重要視されていなかった紙ベースメディアを最大

限に有効利用できる環境を作る。

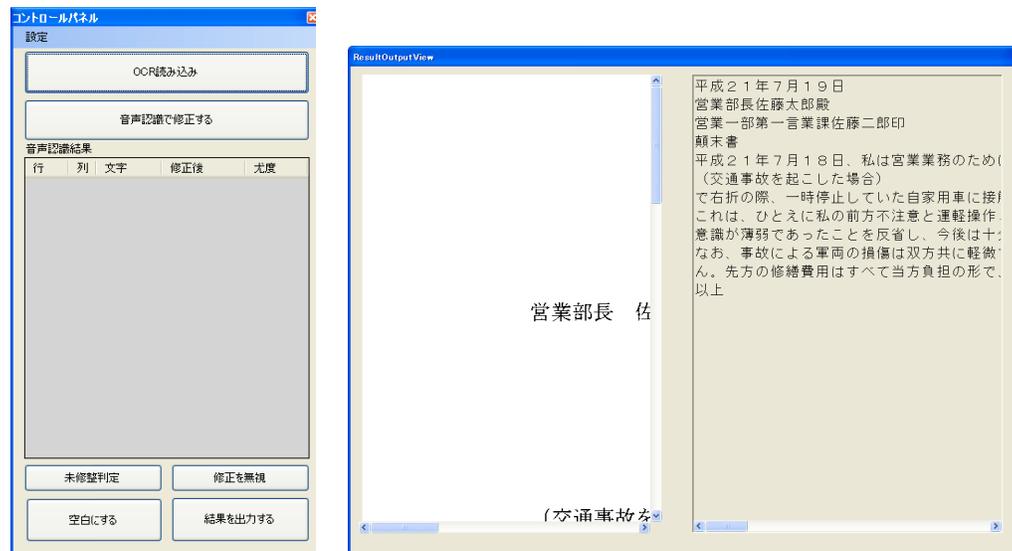


図2. 音声によりOCR結果を自動で修正するGUI

4. 従来技術との相違点

認識精度向上の手法として、文字認識と音声認識の結果を統合し最終的な結果を出力するマルチモーダル認識システムが提案されている[1][2]。しかし、日本語のように漢字、カタカナなど種類の多い言語への適用は困難であったり、OCRのような既に書かれている文字に対しては適用できなかつたり等の問題点があった。音声情報と文字情報を統合するシステムとして、文章書き起こし(つまり文章認識)の他に、キーワード検索が例として考えられる。ここで、文章書き起こし問題は、部分空間(単語)ごとの最適筆記候補を探索し、文意に沿ってそれらを組合せる問題といえる。キーワード探索問題は、広い探索空間の中からキーワードにマッチする筆記を探索する問題と言える。つまり、これらの問題は探索空間が異なるだけで、根本は「最適筆記を探索する問題」という1つの問題を共有する。しかし従来には、このような点に注目したシステムは無く、一般性に欠けていた。さらに、本成果物はノートやメモ帳等に書かれた手書き文字検索に発展できる。カメラやスキャナによって取り込まれたノート等内の手書き文字情報を、キーワードを読み上げるだけで素早く検索する。従来、手書きのノート等から欲しい情報を瞬時に探すことは困難であった。ポメラ[3]等の電子メモも出てきているが、図を自由にかけない、レイアウトに制限がある等の問題があり、大学の講義等ではこれからもノートが使われることが多いと考えられ、手書き文字検索のニーズは高いだろう。本システムにより、ノート等に書かれた特定手書き文字情報を瞬時に検索可能となる。従来、重要度は高いが自由度が低い、と考えられがちだった手書き文字データを、電子データのように自在に扱えるための素地を、本システムにより築くことができる。

5. 期待される効果

従来では、OCRの誤認識はキータイプにより手動で修正していた。しかし、本システムでは誤認識文字を含む単語を読み上げるだけで誤認識の自動修正が可能となる。例えば同じ単語が複数誤認識をしていた場合、従来ではひとつひとつキータイプで直す必要があったため時間がかかってしまっていたが、単語を一度読み上げるだけで複数の同じ文字の単語を同時に修正可能なため、修正にかかる時間の大幅な削減が期待できる。

さらにキーボードレスでOCRの誤認識を修正可能なため、キータイピングが遅い、さらには手が不自由等の理由によりそもそもキータイプができない、といった人でも容易にOCR結果の修正が可能となり、OCRの更なる普及につながる可能性がある。OCRの誤認識を音声により容易に修正できるシステムである。従来では、OCRの誤認識はキータイプにより手動で修正していた。しかし、本システムでは誤認識文字を含む単語を読み上げるだけで誤認識の自動修正が可能となる。例えば同じ単語が複数誤認識をしていた場合、従来ではひとつひとつキータイプで直す必要があったため時間がかかってしまっていたが、本手法では単語を一度読み上げるだけで複数の同じ文字の単語を同時に修正可能なため、修正にかかる時間の大幅な削減が期待できる。

さらにキーボードレスでOCRの誤認識を修正可能なため、キータイピングが遅い、手が不自由等の理由によりそもそもキータイプができない、といった人でも容易にOCR結果の修正が可能となり、OCRの更なる普及につながる可能性がある。

6. 普及の見通し

従来ではキータイプにより手動で修正していたOCRの誤認識を、本システムにより誤認識文字を含む単語を読み上げるだけで誤認識の自動修正が可能となる。さらにキータイピングが遅い、あるいは手が不自由等の理由によりそもそもキータイプができない、といった人でも容易にOCR結果の修正が可能となり、OCRの更なる普及につながる可能性がある。経費削減等の理由により、従来手動で行っていたことを自動化する動きが近年増えているように感じる。OCRを用いれば、従来手作業による集計が主だったアンケート集計などを自動化することができるため、OCRは自動化、コスト削減のニーズに応えることができるため、今後普及していくものと考えられる。提案システムは、OCRのユーザビリティを更に向上させるものであり、コストの削減にも貢献できる。そのため提案システムが今後普及することに期待が持てると考えている。

7. 開発者名(所属)

太田悠平 (東京工業大学 大学院総合理工学研究科 物理情報システム専攻)
馬越健治 ((株)日本電信電話)

参考文献

- [1] S. SUK, H. JUNG, and H. CHUNG, "A Successive State and Mixture Splitting Algorithm for an On-line Speech and Character Combined Recognition System," Technical report of IEICE. EA, vol. 103, no. 24, pp. 25–30, 2003.
- [2] 渡邊康司, 岩田憲治, 中川竜太, 篠田浩一, 古井貞熙, "音声とペン入力の同時入力に対する認識方式の検討(認識・理解・対話・一般)," 電子情報通信学会技術研究報告. SP, 音声, vol. 106, no. 123, pp. 49–54, 2006.
- [3] デジタルメモ「ポメラ」KING JIM, <http://www.kingjim.co.jp/pomera/index.html>