

研究種目：基盤研究(C)

研究期間：2007～2010

課題番号：19500214

研究課題名(和文) 古文献の時代定位システムの開発に関する研究

研究課題名(英文) On developing age-identifying system of Japanese historical documents

研究代表者

中田 充 (NAKATA MITSURU)

山口大学・教育学部・准教授

研究者番号：60304466

研究代表者の専門分野：情報工学，人文社会情報学

科研費の分科・細目：情報学・図書館情報学・人文社会情報学

キーワード：文字認識，グラフ理論，古文献

1. 研究計画の概要

- (1) グラフ理論を用いた一連の文字の構造情報に関するモデル化手法を提案する.
- (2) 原文画像からつづけ字で書かれた一連の文字のグラフを構成する技術を確立する.
- (3) 仮名の字母情報データベースを構築する.
- (4) グラフによりモデル化された仮名の構造を用いた仮名字母変換技術を確立する.
- (5) 字母出現頻度情報データベースを構築.
- (6) 古文献の時代定位や筆者を推定する技術の確立する.

2. 研究の進捗状況

(1) 平成 19 年度(2007 年度)

崩し字・つづけ字で書かれた一連の文字の構造を，グラフ理論を用いて表現するモデル化手法を以下のような手順で提案した.

- ① 手書きの文字を構成する，直線・曲線等の筆使いの情報を抽出し，その情報に基づいて文字の一部分をグラフ化した.
- ② ①のグラフを用いて，変体仮名辞典や代表的な古文献影印本に書かれている文字の一字ずつの構造をグラフ化した.

③ 一文字の画像から文字の形状を表現する特徴グラフを生成するアプリケーションを作成し，実際の古文献の文字画像に適用して，その有用性を確認した.

(2) 平成 20 年度(2008 年度)

- ① 1 文字の構造を表現する特徴グラフの生成アルゴリズムを改良し，文字の形状をより厳密に表現できるようにした.
- ② 2 つの特徴グラフの類似性の尺度である適合度を定義し，その算出アルゴリズムを提案し実装した. さらに，同字種の複数文字について適合度がどのようになるか評価実験を行った.

③ 認識用辞書のスキーマを設計し，データの登録，検索，変更等が行えるシステムを設計し，変体仮名辞典に掲載されている変体仮名の画像データから生成した特徴グラフとその他の文字情報を登録した.

(3) 平成 21 年度(2009 年度)

- ① 前年度までに実現してきた文字画像から特徴グラフを生成するプログラムの改良を行

い、より正確な文字の構造を表現できる特徴グラフを生成可能とした。

② 認識用辞書に対して、その形式を見直すとともに、実際に認識実験を行うことができるように多数の文字を格納するなどの処理を行った。

③ 文字認識アルゴリズムとそれを採用した認識プログラムを改良し、認識率の評価実験を行った。

3. 現在までの達成度

③やや遅れている。

(理由) 古文書中の文字には異体字、字の崩れなどの様々な特徴があり、一文字を対象とした文字認識の精度を必要なレベルまで高めることに時間がかかったため。

4. 今後の研究の推進方策

(1) 昨年度までに確立した一文字を対象とした文字認識技術とグラフ理論を応用して、続け字を対象とした文字認識技術を提案する。

(2) 年代や真偽が明らかになっている代表的な古文書で使用されているつづけ字や崩し字で書かれている仮名を調査することで、字母情報データベースを充実させる。

(3) 対象とする古文書の字母情報の抽出と管理の枠組みを検討する。

(4) (2)と(3)の技術を前提として、古文書の時代定位の仕組みをデータマイニングの技法を応用して検討する。

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計3件)

① Shuich Nishida, Mitsuru Nakata, Qi-Wei Ge, Makoto Yoshimura, Discussion on Handwritten Character Recognition by Feature Graphs. the Fifth International Conference on Information, International Information Institute, 128-133, 2009 査読有り

② Shuich Nishida, Mitsuru Nakata, Qi-Wei Ge, Makoto Yoshimura, A Method of Handwritten Character Recognition by Feature Graphs. 24th International Technical Conference on Circuits/Systems, Computers and Communications, IIEE, 863-866, 2009. 査読有り

③ Masaki Hayashi, Shuichi Nishida, Mitsuru Nakata, Qi-Wei Ge, Makoto Yoshimura, A Method of Generating Feature Graph for Handwritten Character Recognition of Japanese Historical Documents. The 23rd International Technical Conference on Circuits/Systems, Computers and Communications, IEICE, 305-308, 2008. 査読有り

[学会発表] (計5件)

① 林 正紀, 西田秀一, 倉持真理子, 二宮亜佐美, 中田 充, 葛 崎偉, 吉村 誠, 手書き文字認識のための特徴グラフ生成アルゴリズムの改善, 電子情報通信学会技術報告, 108, 415, 39-44, 2009年1月29日, 神奈川県産業振興センター会議室.

② 西田秀一, 林正紀, 倉持真理子, 二宮亜佐美, 中田充, 葛崎偉, 吉村 誠, 手書き文字認識のための特徴グラフの類似性判定アルゴリズム, 電子情報通信学会技術報告, 108, 278, 95-100, 2008年11月7日, 大阪大学.

③ 西田秀一, 林正紀, 中田充, 葛崎偉, 吉村誠, 手書き文字認識のための特徴グラフのマッチングアルゴリズムの提案, 電子情報通信学会技術報告, 108, 79, 25-30, 2008年6月3日, 名古屋大学.

④ 林正紀, 齋藤名美, 中田充, 葛崎偉, 吉村誠, 古文書を対象とした文字認識のための文字構造グラフの生成法, 電子情報通信学会技術報告, 107, 472, 35-40, 2008年1月29日, 徳島大学工業会館.

⑤ 崔日男, 齋藤名美, 中田充, 葛崎偉, 吉村誠, 古文書時代定位システムのためのグラフ理論による文字認識モデルの考察, 電子情報通信学会技術報告, 107, 203, 13-18, 2007年8月31日, 島根大学総合理工学部

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

○取得状況 (計0件)

[その他]

該当無し