H-032

古文書画像に対するワードスポッティングにおける適合判定基準の検討

A thresholding method for the word spotting on historical documents

高橋 渉† 寺沢 憲吾‡
Wataru Takahashi Kengo Terasawa

1． 背景

近年，歴史的資料や貴重な文化財などがディジタルアーカイブ化され， 様々な場面で活用されることが多くなってきている．これらの歴史的文書を博物館や資料館などで冊子状のまま一般の来客者などに公開することは， 一般の来客者によるなんらかの事故によってそれらを破損してしまう恐れがあり，危険である．このような問題に対し，ディジタルアーカイブ化されたものをオンラインで公開するという方法は非常に有益な方法であると言える．この方法は近年活発に行われており．例えば，北海道函館市中央図書館のホームページには明治 17 年の函館新聞や亜国来使記といった歴史的文書を閲覧するとともに、全文検索もできる文書画像検索システムがある（図 1）．

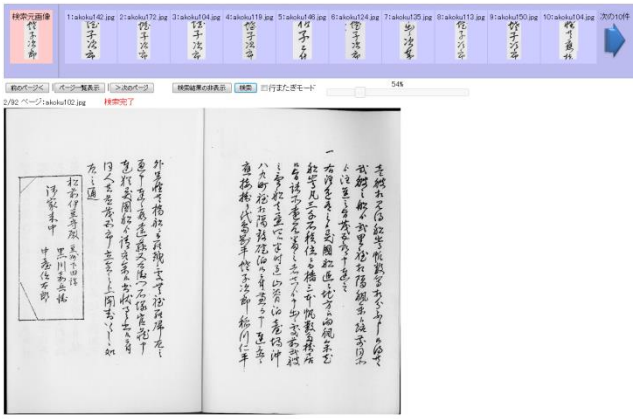

図 1: 文書画像検索システム

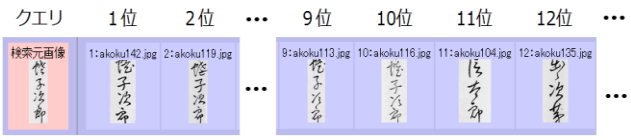

図 2: 検索結果の拡大図

図 1 は，「蛯子次郎」を検索クエリとして全文検索を行った例である．クエリ画像と見かけが似ている画像を似ている順に表示している．ただしこのシステムでは，画像を似ている順に表示することはできるが，何位までがクエリ文字列と同一文字列なのかを知ることはできない．たとえば図 2 の例では，10 位までがクエリ画像と同一文字列であり，11 位以降はクエリ画像と異なる文字列であるが，システム側でそれを検知することはできない．もしこれを自動的に判定することができれば，検索結果を利用するユーザーの利便性が高まるとともに，画像技術によるキーワード抽出などといったさらなる応用に役立つ可能性もある．

本研究は，検索結果の何位までが同じ文字列であるのかを判定するための適合判定基準を考案することを目的とする．

2. 関連研究

2.1 ワードスポッティング

文書画像から文字列検索を行うにあたり，一般的な手法としては光学式文字読取装置(OCR)を用いる手法がある．これは，手書き文字や印字された文字を光学的に読み取り 1 文字単位で切り出しを行い，前もって記憶されているパターンとの照合によって文字を特定し，文字データを入力する装置である．従って，全文検索は文字コード同士の比較によって行われることになる．ところが，一般の OCR ソフトウェアは現代の活字文書を対象としているので，歴史的文書などに多い連綿文字の手書き文書に用いる場合の適応性は決して高くはない．つまり，それらを対象とした文字列検索の精度も決して高くはないものとなることを意味する．

そこで，文字画像をテキスト形式に変換するのではなく，画像形式のままで検索を行う手法が既に提案されている．この手法がワードスポッティング(Word Spotting)である．ワードスポッティングは Manmatha ら[1]によって初めに考案された，手書き文書から文字列探索を行う手法である．文書画像中の一部の文字列である部分画像をクエリ画像とし，文書画像中からクエリ画像と類似度の高い部分を抽出する．Manmatha らが考案したワードスポッティングは英語の手書き文書を対象としている．

日本語の毛筆手書き文書を対象としたワードスポッティングは Terasawa ら[2]が提案している．これは文字列をスリット状に切り出して固有空間法を適用させることによって各スリット画像ごとに特徴量ベクトルを計算する．この特徴量ベクトルを用いて，クエリ画像と各画像のベクトル間の距離を計算する際に，DTW (dynamic time warping)を適用することで文字列の伸縮に対するロバスト性を付加させながら類似度を求める．この手法は，江戸末期の毛筆文書画像を対象にキーワードの検索を行った実験で，平均適合率 73～93%を示し，実用性の高さが確認されている．前章で紹介した文書画像検索システムは，この手法を用いて作成されたものである．

† 公立はこだて未来大学大学院システム情報科学研究科
‡ 公立はこだて未来大学システム情報科学部

2.2 既存の適合判定基準

Terasawa ら[3]は，文書画像からの自動キーワード抽出に取り組む研究の中で，ワードスポッティングによる検索結果の適合判定基準についても検討している．

まず，対象とする文書画像内からワードスポッティングに用いるクエリ画像となる文字列を指定する．指定されたクエリ画像と，文書画像内の文書となる部分の全画像領域との特徴量ベクトル間の距離を計算する．この距離が近ければ近いほど類似していると考えることができる．この段階で，クエリ画像に類似している文書画像内の文字列画像を，類似している順にデータとして得ることができる．このクエリ画像の文字列が頻出語かどうか判断するには，文書内にその単語がいくつ存在するのか知る必要がある．そのためには，ワードスポッティングで得られたデータを見た時，どれほど類似していればクエリ画像と同じ文字列であるのか，という適合判定基準を設定する必要がある．

図 3 は，「亜国来使記」（図 4）という文書画像内から「伊豆守」，「勘解由」，「も下り」をそれぞれクエリ画像としたときのワードスポッティングの結果である．「伊豆守」，「勘解由」は全文中に 10 回程度出現する文字列である．「も下り」は全文中に 1 回しか出現しない文字列である．図中の白丸はクエリ画像と同じ文字列，黒丸はクエリ画像と異なる文字列を意味する．縦軸はクエリ画像との特徴量ベクトル間の距離，横軸は順位を意味する．グラフの結果を見ると，クエリ画像によってそれぞれ適切な閾値が異なっており，どの文字列をクエリ画像としても単純な方法では対応する閾値を設定することができないことがわかる．Terasawa らの手法では，クエリ画像に応じて閾値を変化させる適合判定基準を用いている．クエリ画像と白色画素との間の距離を「クエリ画像のエネルギー」とし，その 0.20 倍を閾値とする（図 3 には赤線で表示されている）．この閾値以下のものをクエリ画像と同じ文字列と判断する．図 3 からわかるようにこの方法は適合文字列（白丸）と不適合文字列（黒丸）を最適に分離できているとは言い難い．本研究では， さらに高精度な適合判定基準の考案する．

3. 新しい適合判定基準

既存の手法に用いられている適合判定基準によって求められた図 3 の赤線の閾値では，クエリと同じ文字列であるにもかかわらず拾うことの出来ていない文字列がいくつも存在している．つまり，適合判定基準に関してはまだ改善の余地があると考えることができる．

既存の適合判定基準は，画像間の距離の値に着目したものであるが，本研究ではクエリ画像を微小に変化させた際の検索結果の変化の様子にも着目した．そして，その他にも複数の適合判定基準を考案し，それらの結果を組み合わせて分類を行うアンサンブル学習を用いることで，更に高精度化を目指すことにした．アンサンブル学習とは，複数の適合判定基準による結果を組み合わせて，それらを統合することによって，個々の適合判定基準よりも高い精度にする方法である．本研究では，3 つの適合判定基準のうち，2 つ以上の適合判定基準にて分類されたものを，クエリと同じ文字列であると判断する．

以下では，考案した 3 つの適合判定基準について説明する．

3.1 グローバルな閾値

1 つ目は，ワードスポッティングでの検索結果に対して，あらかじめグローバルな閾値を定めておき，その値以下のものをクエリ画像の文字列と同じ文字列であると判断するものである．本研究では閾値を 0.6 とする．この値はいくつかのクエリに対して，グローバルな閾値を用いた適合判定基準の精度検証実験を行った際，最も高精度な結果となった値として実験的に定めた．なお，現在の文書画像検索システムは Terasawa ら[3]の研究当時とは異なる特徴量を用いている．そのため，[3]のようにクエリ画像のエネルギーに比例させた閾値を用いる方法よりも，グローバルな閾値のほうが精度が高いという結果となった．

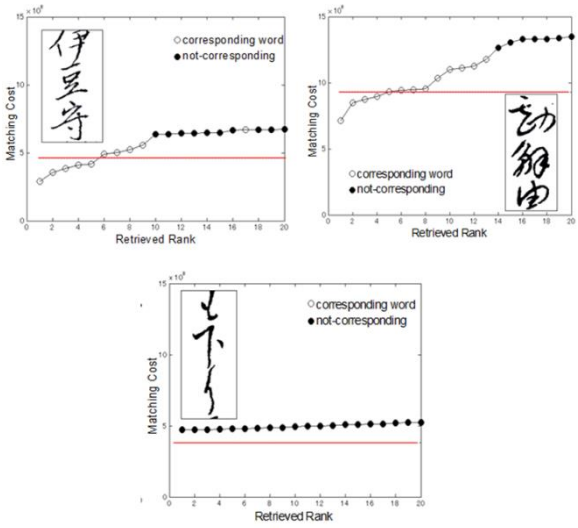

図 3: 検索結果

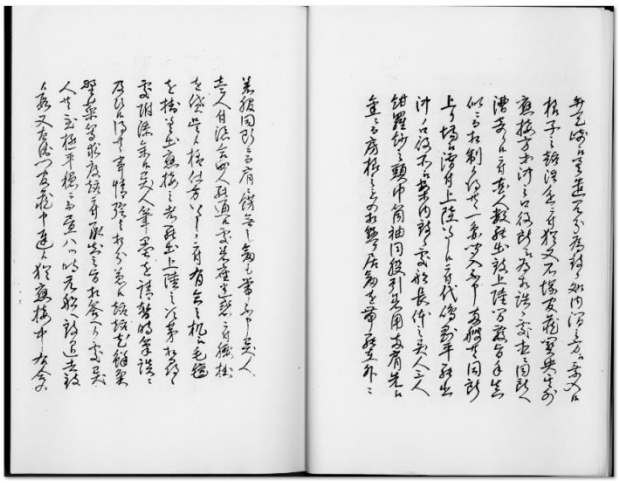

図 4: 亜国来使記

3.2　検索結果 1 位の値を定数倍した閾値

2 つ目は，ある文字列をクエリ画像としてワードスポッティングで検索した時の，検索結果 1 位との画像間の距離を基準として，それを定数倍したものを閾値とするものである．様々なクエリによるワードスポッティングの検索結果を考察したところ，検索結果 1 位との画像間の距離の値が小さいと，クエリに適した閾値も小さくなる傾向の結果が多く見られた．逆に，検索結果 1 位との画像間の距離の値が大きいと，クエリに適した閾値も大きくなる結果も同様に多く見られた．このことから，検索結果 1 位との画像間の距離の値に比例して，閾値も変動するのではないかと考え，2 つ目の適合判定基準とした．

対象となるクエリの検索結果 1 位の値を d，閾値を定める係数を c とした時，対象となるクエリでの検索結果に用いる閾値 t の値を求める式は

$$t = cd \tag{1}$$

となる．いくつかのクエリに対して，検索結果 1 位の値を定数倍した閾値を用いた適合判定基準の精度検証実験を行った際，c の値を 1.5 としたものが最も高精度な結果となったため，本研究では c の値を 1.5 とする．

なお，この方法は図 2 の「も下り」のように適合する文字列が 1 つもない場合（検索結果 1 位の文字列が不適合である場合）必ず誤った結果を出力する．そのため，他の手法と組み合わせて用いる必要がある．

3.3　スリット数の増減による順位変動

3 つ目は，著者らが先行研究[4]において開発した，クエリ画像を微小に変化させた際の検索結果の変化の様子に着目する方法である．これは，クエリと真に一致する文字列は理由があって上位に検出されているのに対し，一致していない文字列は上位に検出されたとしてもそれはたまたまのものであるから，クエリ画像の微小な変化によって順位が大きく変わるだろうという予見に基づくものである．本研究では，クエリ画像の微小な変化として画像の終端位置（スリット数）を微小に変化させるという方法を採用した．

図 5 は，ある文書から「試し書き」という文字列をクエリ画像とした場合の検索結果を，左からクエリ画像の終端のスリットの位置を短くして検索した時，終端の位置をずらさないで検索した時，終端のスリットの位置を伸ばして検索した時の順に並べたものである．終端の位置をずらさない場合での検索結果の 1 位から 13 位までをそれぞれ①から⑬までの記号に置き換えた．14 位以降のものはランク外と示した．それぞれの結果を比較すると，クエリ画像の終端のスリットの位置を短くした場合での 3 位から 5 位までの結果は，ずらさない場合での 3 位から 5 位までの結果が入れ替わっただけの結果となっている．7 位と 8 位も同様に入れ替わっただけの結果である．つまり，短くした場合での 1 位から 8 位までの検索結果は，ずらさない場合での 1 位から 8 位までの結果が入れ替わっただけの結果であると言える．8 位以降の結果も比較してみる．短くした場合での 9 位にはずらない場合での 11 位のものがランクインしている．短くした場合での 10 位，11 位にはずらさない場合でランク外だったものがランクインしてきている．つまり，短くした場合での 8 位以降の結果は，ずらさない場合での 8 位以降の結果が大きく順位変動した結果であると言える．上述の結果は，終端の位置を伸ばした場合での検索結果と終点の位置をずらさない場合での検索結果を比較しても同じようなことが言える．

この結果に対して本研究では次のような適合判定基準を考案した．それは「クエリ画像のスリットの終点の位置をずらした時の検索結果が，元のクエリ画像で検索した時にランクインした順位内のものが入れ替わっただけの結果となる最大範囲を境目とする」というものである．上述の結果に対してこの適合判定基準を用いると，1 位から 8 位までがクエリ画像と同じ文字列だと自動的に判断することができる．

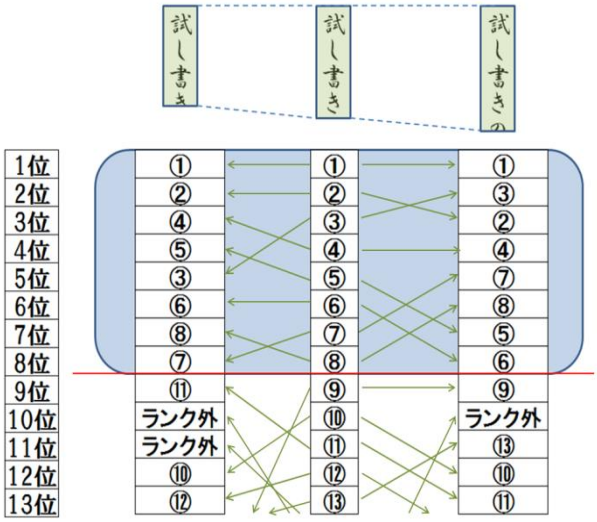

図 5: 順位変動の予想結果

3.4　アンサンブル学習

これら 3 つの適合判定基準をそれぞれ適用させた結果に対して 2 つ以上の基準で適合と判定されたものを適合文字列と判定することにする．アンサンブル学習を用いた時の精度が、既存の適合判定基準、新しく考案した適合判定基準の個々の精度よりも高いものであるのか確かめるための検証実験を次章で説明する．

4.　実験

全画像ページ数 92 の手書き文書「亜国来使記」を対象とする．文書から出現回数が 10～20 回程度の文字列を中心に，それより出現回数が多いものや少ないものも含めて適当に選定し，クエリ画像とした（表 1)．それらのクエリ画像の検索結果に対して 3 章で考案した 3 つの適合判定基準で分類された結果を，アンサンブル学習を用いて統合する．統合する際には，3 つの適合判定基準のうち，2 つ以上の適合判定基準にて適合と判定されたものを，クエリと同じ文字列であると判断する．同じ文字列であると判断された画像をそれぞれについて，正解データを用いて正誤を判定し，結果をまとめる．

結果をまとめる際，精度比較のため，再現率と適合率と F 値をそれぞれ求めた．再現率は拾うべき文字列がいくつ拾えたか，適合率は拾った文字列全ての中にクエリ画像と

同じ文字列はいくつあるかを意味する．F 値は適合率と再現率の調和平均であり，高ければ高いほど性能が良いものと判断することができる．アンサンブル学習を用いた結果を，新しく考案した 3 つの適合判定基準をそれぞれ適用した時の結果，既存の手法の適合判定基準を適用した時の結果と比較することで，どれほどの精度であるのか確かめる．

既存の適合判定基準による結果，アンサンブル学習を用いた結果を表 2 に示す．グローバルな閾値を用いた適合判定基準，検索結果 1 位を定数倍した閾値による適合判定基準，順位変動に着目する適合判定基準のそれぞれの結果を表 3 に示す．

表 1: クエリ画像一覧（括弧内の数字は出現回数）

藤原主馬(10)　石塚官蔵(25)　平山謙二郎(11)
ホハタン(19)　工藤茂五郎(11)　蛯子次郎(13)
警固之物(11)　船江罷歸(11)　二番入津(12)
安間純之進(14)　富左右江(19)　四ツ時頃(20)
段見廻方(23)　稲川仁平(24)　ヘンテウリヤムス(24)
井上富左右(25)　勘解由(14)　二番入津之異船(10)
異国船(30)　本船江引取候(21)　沖ノ口役所(7)
別紙應接書(22)　亀田濱(14)　応接所(16)
上陸いたし(17)　中台信太郎(3)　松前伊豆守(6)
マシトニアン(7)　ハンテリヤ(8)

表 2: 既存の手法とアンサンブル学習を用いた結果

	既存の適合判定基準	アンサンブル学習
再現率	0.451	0.924
適合率	1	0.837
F 値	0.622	0.808

表 3: 3 つの適合判定基準の個々の結果

	グローバル	定数倍の閾値	順位変動
再現率	0.934	0.932	0.84
適合率	0.794	0.815	0.869
F 値	0.864	0.873	0.854

既存の適合判定基準の結果に比べ，新しく考案した 3 つの適合判定基準の個々のほうが F 値が高い．アンサンブル学習を用いた結果はそれらの個々の F 値をさらに上回る結果となった．このことから，新しく考案した 3 つの適合判定基準での結果をアンサンブル学習で統合させた方法が，最も高精度であるということがわかった．

5. まとめと今後の展望

本研究で考案した 3 つの適合判定基準は既存のものと比べると，どれも性能の高いものであるということがわかった．また，アンサンブル学習を用いることでそれぞれの適合判定基準を用いた場合よりもさらに高精度な結果を出すことができた．しかし，今回の実験は対象画像が「亜国来使記」に限られているため，今後はさらにほかの歴史的文書に対しても同様の実験を行い，提案手法の適用可能性を検証する必要がある．また，この精度で実際にキーワード抽出の自動化を行ったとき，十分な精度が得られるかについての検証も今後の課題である．

参考文献

[1] R. Manmatha, C. Han, and E. M. Riseman, “Word spotting: a new approach to indexing handwriting,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 631-637, 1996.

[2] K. Terasawa, T. Nagasaki, and T. Kawashima, “Eigen-space Method for Text Retrieval in Historical Document Images,” Proc. Int. Conf. on Document Analysis and Recognition, pp. 437-441, 2005.

[3] K. Terasawa, T. Nagasaki, and T. Kawashima, “Automatic Keyword Extraction from Historical Document Images,” 7th IAPR Workshop on Document Analysis Systems, DAS2006, LNCS 3872, pp. 413-424, Nelson, New Zealand, Feb. 13-15, 2006.

[4] 高橋 渉，寺沢憲吾，“手書き文書画像からの自動キーワード抽出の高精度化”，平成 24 年度電気・情報関係学会北海道支部連合大会, (2012.10).