

検索結果を統合するための情報量の概念を考慮したスコア正規化手法

鈴木 優[†] 波多野 賢治[†]
吉川 正俊^{††} 植村 俊亮[†]

複数の検索システムへ問合せを行い、それぞれの結果を統合して精度を向上させる手法であるメタ検索エンジンが、Web 文書や画像の検索などに広く活用されている。この手法では、1 つの検索対象オブジェクトに対して複数の検索システムがそれぞれスコアを付与し、関数を用いてそれらを統合する。メタ検索エンジンの各検索システムが計算するスコア群の平均、分散、総和などはそれぞれ異なるため、統合する前に正規化する必要がある。ところが従来のスコア正規化手法では、各検索システムが検索対象オブジェクトに付与したスコアの分布に偏りがあるようなスコア群を正規化することができないため、最適な正規化ができない場合がある。本論文では、各検索システムが計算したスコアに対する検索対象オブジェクト数の分布を考慮したスコアの正規化手法を提案する。本提案では、ある検索対象オブジェクトのスコアが同じであっても、そのスコア周辺の検索対象オブジェクトの数が多い場合は少ない場合に比べて相対的に低い値となるべきであると考えた。つまり、検索対象オブジェクト数が多いスコアの範囲では正規化後のスコアを低くし、検索対象オブジェクト数が少ないスコアの範囲では正規化後のスコアを高くする。その結果、高い値を持つスコアが多い検索結果と低い値を持つスコアが多い検索結果を、それぞれ相互に統合可能なスコアへ正規化を行うことが可能となる。

A Relevance Score Normalizing Method Using Shannon's Information Measure

YU SUZUKI,[†] KENJI HATANO,[†] MASATOSHI YOSHIKAWA^{††}
and SHUNSUKE UEMURA[†]

In metasearch engines, many individual retrieval systems calculate raw relevant scores to a retrieval target respectively, and the metasearch engine directly combines these raw relevant scores into the similarities between the user's query and the retrieval target. At this time, the metasearch engine should normalize the raw relevant scores to be equivalent with each other, because the raw relevant scores are not always adequate to be combined. That is to say, the same normalized relevant scores should indicate the same similarities between the user's query and the retrieval targets, even if the normalized relevant scores are calculated by different retrieval systems. Many normalization methods have been proposed so far, but these normalization methods are not always sufficient for normalizing the raw relevance scores. In this paper, we propose a method for normalizing raw relevant scores using Shannon's information measure. By applying Shannon's information measure to our proposed normalization method, the high raw relevant scores are converted to low normalized relevant scores if a retrieval system calculates many high raw relevant scores. On the contrary, the high raw relevant scores are converted to high normalized relevant scores if a retrieval system calculates only a few number of high raw relevant scores. We assume that the retrieval targets are relevant for the user's query if both the raw relevance scores and the Shannon's information measure are high. From our experimental result, we confirmed that the accuracy of our normalization method is better than others. Consequently, we confirmed that the proposed assumption is correct for relevant score normalization method.

1. はじめに

情報検索システムの検索精度向上に関する研究がさかんに行われている。検索精度を向上させるためには様々な方法が考えられるが、現在では1つの閉じた検索システムの検索精度を向上させる方法には限界があ

[†] 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute
of Science and Technology

^{††} 名古屋大学情報連携基盤センター
Information Technology Center, Nagoya University

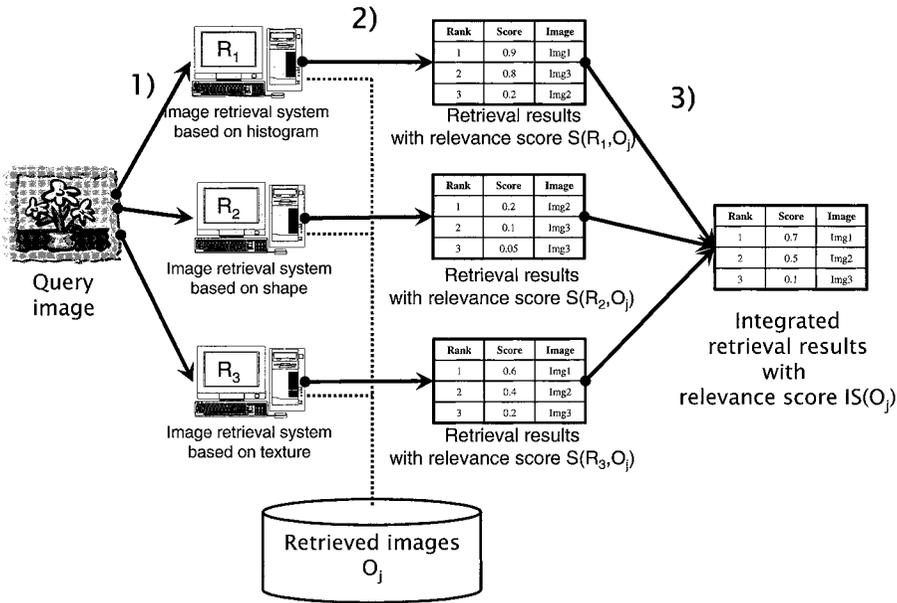


図 1 複数の検索システムを統合したメタ検索エンジンの例
Fig. 1 Integration of multiple information retrieval systems.

と考えられており、複数の検索システムによる検索結果を統合することによって、検索精度を向上させる研究、つまりメタ検索エンジンに関する研究が注目されている。たとえば、画像検索システムにおけるメタ検索エンジンの例として、図 1 のようなシステムがあげられる。図 1 は、利用者が問合せとして 1 つの画像を入力したときに、問合せ画像に類似した画像を検索するためのシステムであり、1) メタ検索エンジンが利用者の問合せ画像を複数の検索システムへ入力し、2) それぞれの検索システムが順位、スコア付き検索結果を出力し、3) 2) で計算されたスコアを統合し検索結果を出力する、という 3 つの手順で問合せ処理が行われる。本論文では、検索を行うためのシステム全体を“メタ検索エンジン”と呼び、メタ検索エンジンを構成している個々の検索を行うためのシステムを“検索システム”と呼んでいる。

このようなメタ検索エンジンによる検索精度の向上に関する研究は、現在までに数多く行われているが、それらの課題を列挙すると次の 5 点に集約することができる。

- (1) 検索対象オブジェクト（先の例では画像に相当する）からの特徴量の抽出方法。
- (2) 各検索システムにおけるスコアの計算方法。
- (3) 算出されたスコアの正規化方法。
- (4) 正規化されたスコアの統合方法。
- (5) 利用者への検索結果の提示方法。

我々は、複数の検索システムが出力する検索結果を統合することによる検索精度の向上を研究の目的としてきた。先行研究¹⁷⁾ではこれらの 5 つの問題点のうち 4 番目の手法、すなわちスコアの統合方法に関する研究を行ったが、スコアを統合するにはスコアの正規化が検索精度に大きく反映されることが評価実験によって判明したため、本論文では 3 番目の課題、つまりスコアの正規化の問題に着目した。

メタ検索エンジンを構成している各検索システムは、それぞれ色ヒストグラムを用いた方法や模様情報を用いた方法などそれぞれ各々の方法を用いてスコアを計算するため、スコアの分布、範囲はそれぞれ異なり、そのまま統合することはできない。そこで、複数の検索システムが出力したスコアをそれぞれ統合可能な値へ変換するために、スコアの正規化を行う。従来からスコアを正規化するための様々な方法が提案されているが^{6),8),10)}、これらの方法ではスコアの中に極端に大きい値や小さい値がある場合、またスコアの分布に偏りがある場合には、正しく正規化を行うことができない場合があるという問題点がある。この問題点は、従来の手法は正規化のための関数が線形関数を基にしている点から生じていると考えられる。

本論文では、Shannon の情報量⁹⁾の概念を利用したスコアの正規化手法を提案する。ある検索システムが検索対象オブジェクト群に対して低いスコアを付与したにもかかわらず、ある検索対象オブジェクトに対

してだけ高いスコアを付与したとき、その検索対象オブジェクトは利用者にとって適合している可能性が非常に高いと考えられる。そこで我々は、複数のスコアのうちのきわめて高いスコアが付与された検索対象オブジェクトは、検索対象オブジェクト群の中で特に目立った検索対象オブジェクトであるために、利用者によってその検索対象オブジェクトは適合していると判断されると仮定した。この仮定に従って、検索対象オブジェクトに付与された正規化前のスコアがきわめて高い値であるかを判断するために、Shannon の情報量を用いている。なぜなら情報量の概念は、数値の集合のうちのある 1 つの数値が他の数値群と比較してどの程度目立った値であるかを測定するために用いることができるからである。

提案手法の有効性を確かめるために、評価実験を行った。評価実験では、色ヒストグラム、形状、模様情報それぞれを用いた 3 つの画像検索システムを統合した類似画像メタ検索エンジンを構築し、以下に述べる 3 つの点においての考察を行う。1) 提案手法による場合と従来の手法による場合における正規化後のスコアの分布を比較することにより、正規化後のスコアの値が相互に比較可能であるかを確かめる。2) 提案手法におけるパラメータである、正規化前スコアの範囲を分割する数に関する考察も行っている。3) 構築した類似画像メタ検索エンジンにおける検索精度を再現率・適合率によって比較し、提案手法の有効性を確かめる。

本論文は次のような構成である。まず 2 章では我々の提案手法と関連する手法との差異について述べる。次に 3 章では、Shannon の情報量の概念を考慮したスコアの正規化手法について述べる。4 章では評価実験を行い提案手法の有効性を確かめる。最後に 5 章では、本論文のまとめを行い、さらに今後の研究課題について述べる。

2. 関連研究

本章では従来提案されているメタ検索エンジンに関する研究と本提案手法との差異について述べる。

メタ検索エンジンと単一の検索システムを比較して、検索精度がなぜ高くなるかについて、Diamond の研究結果³⁾で示されており、次の 2 つの点が理由としてあげられている。

- (1) 多くの検索システムによって高いスコアを得られた検索対象オブジェクトは、利用者の検索要求に適合していることが多いため。
- (2) 1 つの検索システムでは検索対象オブジェクト

に対して誤ったスコアが付与され、適合、不適合の判断を誤る可能性があるが、複数の検索システムを用いることによってその誤りが減少するため。

我々は、メタ検索エンジンの検索精度を向上させるために、スコアを統合する際の前処理としての正規化が必要であると考えている。なぜなら、スコアはすべて統一された基準で計算されているわけではなく、それらを統合することによって検索システムを平等に評価できないからである。一方、小作らによる実験結果¹⁶⁾では、複数の検索システムによるスコアを統合する場合に、正規化を行う必要がないことが示されている。しかし、小作らのメタ検索エンジンの前提条件は、すべての検索システムが同じアルゴリズムで動作すること、検索対象オブジェクトが異なることであるのに対して、我々の提案するメタ検索エンジンの前提条件はすべての検索システムが異なるアルゴリズムで動作すること、検索対象オブジェクトが同一であることである。そのため、必ずしもスコアの正規化が不要であるとはいえないことが分かる。

スコアの値を統合することを考えた場合、統合されるスコアは比較可能でなければならない。ここでスコアが比較可能であるとは、3.2 節で述べるとおり、複数の検索システムが計算したスコアがすべて統一された基準で付与されていることである。そこで、スコア相互を正規化するための手法が提案されている。たとえば、Web 文書を検索するためのメタ検索エンジンに使用するための正規化として、Montague らは 3 つの手法、*Standard*、*Sum*、*ZMUV* を用いている⁷⁾。これらの手法は、正規化を行わない場合と比較して有用であることが彼らの実験によって示されているが、論理的な根拠が乏しく、正規化を行う利点が評価実験だけでしか示されていない。我々は、提案手法を提案するうえで正規化手法の妥当性について 3.2 節において議論しており、Montague らの手法と比較してより妥当な方法であるといえる。

一方、利用者のフィードバックを用いて各検索システムで計算されたスコアを正規化を行う手法が Vogt ら^{11)~14)}、Bartell²⁾によって提案されている。これらの手法では、利用者の検索対象オブジェクトに対する適合、不適合の判断を正規化の手法に反映させることができるため、その点では非常に理想的なシステムであるといえる。ところが、適合フィードバックによる方法は、利用者にとって非常に手間がかかる点が問題である。提案手法は適合フィードバックによる方法を採用していないため、利用者が検索に必要な手間が

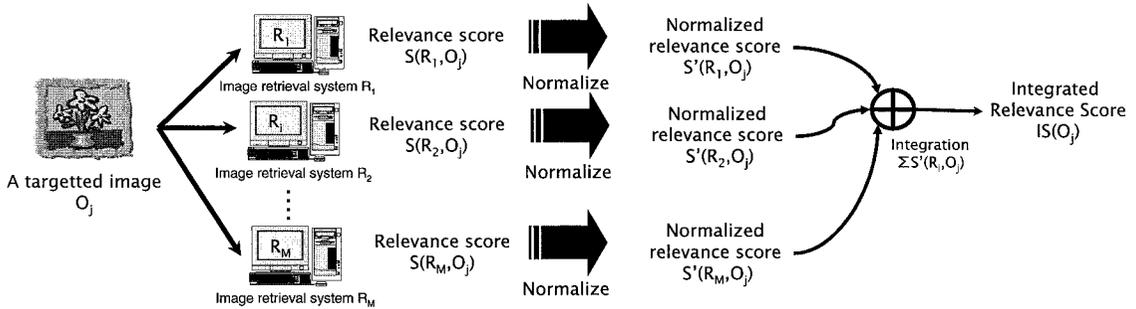


図 2 スコアの正規化とその統合

Fig. 2 Score normalization and integration.

軽減される点も特徴である。

3. スコアの正規化手法

複数の検索システムが計算したスコアをどのようにして正規化するかは、本論文の研究課題である。なぜなら、検索システムがすべて統一された基準でスコアを算出されているとは限らないためである。そこで本章では、検索対象オブジェクトの数の分布を考えた正規化を行う手法について述べる。

3.1 本論文で用いる用語の定義

図 2 において、本論文で使用する変数、定数の定義を行う。まず、メタ検索エンジンを構成している検索システムをそれぞれ R_i ($i = 1, 2, \dots, M$) とおく。すべての検索システムでは同じ検索対象オブジェクト集合を検索対象としていると仮定し、それぞれの検索対象オブジェクトを O_j ($j = 1, 2, \dots, N$) とする。検索システムはすべての検索対象オブジェクトと問合せとの類似度を計算し、スコアを付ける。ここで、検索システム R_i によって付与された検索対象オブジェクト O_j のスコアを $S(R_i, O_j)$ とする。最終的には統合関数を使って、検索対象オブジェクト O_j に対して式 (1) を用いて 1 つの統合スコア $IS(O_j)$ を計算する。

$$IS(O_j) = \oplus_{i=1}^M S'(R_i, O_j) \quad (1)$$

ここで、 $S'(R_i, O_j)$ は $S(R_i, O_j)$ を提案手法によって正規化したスコアであり、3.3.3 項の式 (6) によって定義される値である。

3.2 スコアの正規化のための直感的な考え方

本節では、まず提案手法における前提条件、次に正規化後のスコアが満たさなければならない条件について述べ、さらにそれらの条件を満たした正規化手法について述べる。

提案手法を考える上で、次の 2 つの前提条件を仮定している。

(1) 各検索システムはすべて同程度の検索精度で

ある。

(2) 検索システムをすべて平等に評価することによって、メタ検索エンジンの検索精度は向上する。

1 つ目の条件を設定した理由は、メタ検索エンジンを構築する際に検索精度が悪い検索システムを基にメタ検索エンジンを構築した場合に、その検索システムの影響でメタ検索エンジンの検索精度が低下してしまうためである。また、2 つ目の条件を設定した理由は、すべての検索システムを平等に評価していない場合、すなわち一部の検索システムの検索結果を無視している場合には、すべての検索システムの検索結果が統合後の結果へ反映されていないため、メタ検索エンジンの検索精度が下がるためである。Montague らの実験⁷⁾ によって、統合する検索システムの数が多ければ多いほどメタ検索システムの検索精度が向上することが示されているため、本論文では以上に示した 2 つの前提条件を仮定した。

我々は、正規化後のスコアはすべて統一された基準で付与されていると考える。まず、ある検索対象オブジェクト O_1 に対して検索システム R_1 が付与したスコアが $S(R_1, O_1)$ であり、同様に検索オブジェクト O_2 に対して検索システム R_2 が付与したスコア $S(R_2, O_2)$ であったとする。さらに、 $S(R_1, O_1)$ と $S(R_2, O_2)$ が同じ値であったとする。このとき O_1 と O_2 が利用者にとって同程度適合していると判断した場合、検索システム R_1, R_2 で計算されたスコア相互は統合可能であるといえる。このように、複数の検索システムによって計算されたスコアが統合可能である状態であることを、本論文では「相互に比較可能である」と呼ぶ。

一方、Montague ら⁷⁾ が使用した正規化手法(表 1)では、以下のような仮定がなされている。

(1) Standard

すべての検索システムで出力されたスコアの最

表 1 Montague らが提案したスコア正規化手法⁷⁾Table 1 Normalization algorithms designed by Montague and Aslam⁷⁾.

正規化手法	説明
Standard	スコアの最小値を 0 に, 最大値を 1 に揃える
SUM	スコアの最小値を 0 に, 総和を 1 に揃える
ZMUV	スコアの平均を 0 に, 分散を 1 に揃える

大値, 最小値が同じであるならば, スコアは相互に比較可能である.

(2) Sum

すべての検索システムで出力されたスコアの総和が同じであるならば, スコアは相互に比較可能である.

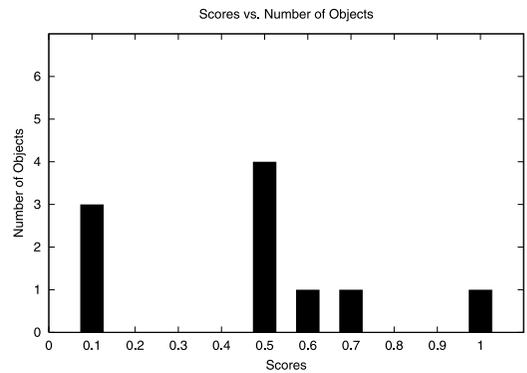
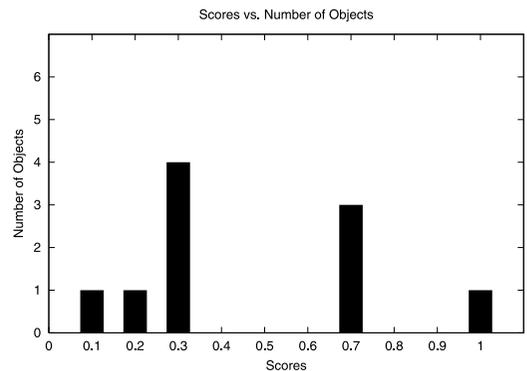
(3) ZMUV

すべての検索システムで出力されたスコアの平均, 分散が同じであるならば, スコアは相互に比較可能である.

これらの仮定を見ると, Montague らの仮定では正規化後のスコアが必ずしも相互に比較可能とはならないことが分かる.

たとえば, 検索システム R_1, R_2 があり, それらのスコアに対する画像の数は図 3 のように分布していると仮定する. これらのスコアの分布を見ると, 2 つのスコアは最大値, 最小値, 総和, 平均, 分散が双方同じである. つまり, Standard, Sum, ZMUV におけるすべての仮定は満たされている.ところが, 同じスコアを付与されている検索対象オブジェクトは同じ程度適合しているとはいえない.なぜなら, 検索システム R_1 においてはスコア 0.7 の検索対象オブジェクトが 1 つであることにに対して, R_2 では 3 つであるからである.この例の場合では, R_2 では比較的高い値のスコアが算出されやすい点と比較して, R_1 では 0.5 に近い値が算出されやすいことが分かる.つまり, もし利用者がスコア 0.7 以上を適合と判断すると仮定した場合, R_1 と比較して R_2 は相対的に多くの検索対象オブジェクトを不適合と判断したことが分かる.ところが, 2 つの検索システムのどちらも適合している検索対象オブジェクトの数は同じである.つまり, R_1 と R_2 2 つの検索システムが計算したスコアは相互に比較可能ではないことが分かる.

スコアを相互に比較可能とするためには, 利用者がどの検索対象オブジェクトに対して適合すると判断するのかを正規化手法において推測しなければならない

(a) 検索システム R_1 (b) 検索システム R_2 図 3 検索システム R_1, R_2 を用いた場合のスコアの分布図Fig. 3 The number of objects for each score in the retrieval systems R_1 and R_2 .

い.そこで本論文では, 利用者は検索対象オブジェクト群に付与されたスコアのうち目立って高い値であるものを適合していると判断すると仮定した.この仮定に従って, スコアの正規化の際に, スコアが他の検索対象オブジェクト群と比較してどの程度目立っているかを判定し, そのための指標として Shannon の情報量⁹⁾ の概念を用いた.

3.3 Shannon の情報量の概念を利用したスコアの正規化手法

本節では, 情報量に基づいたスコアの正規化手法の具体的な方法について述べる.我々の提案する手法は, 次の 3 つの手順で計算される.

- (1) 最大値, 最小値を用いた正規化を行い, スコアを $[0, 1]$ の範囲に収める.
- (2) 計算された各スコアに対して情報量を求める.
- (3) スコアと情報量を統合し, 正規化されたスコアを導出する.

以下では, これら 3 つの手順について詳細を述べる.

3.3.1 正規化のための準備

まず, 正規化後のスコアの値の範囲を $[0, 1]$ に収め

最大値は 1, 最小値は 0.1, 総和は 4.6, 平均は 0.46, 分散は 0.27 である.

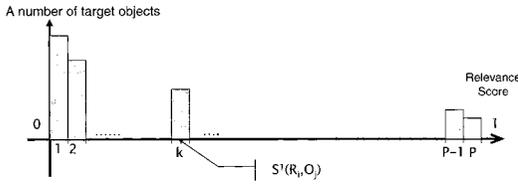


図4 スコアに対する情報量の計算

Fig. 4 Calculation of information measure for relevance scores.

るため、従来の正規化手法である *Standard* を用いてスコアを $[0, 1]$ へ正規化する。 $S_{max}(R_i)$ をスコアの最大値、 $S_{min}(R_i)$ をスコアの最小値とした場合に、正規化後のスコア $S^*(R_i, O_j)$ を次のように定義する。

$$S^*(R_i, O_j) = \frac{S(R_i, O_j) - S_{min}(R_i)}{S_{max}(R_i) - S_{min}(R_i)} \quad (2)$$

3.3.2 Shannon の情報量の計算

情報量を求めるために、正規化前のスコアをそのまま用いず、スコアの範囲 $[0, 1]$ を区切り、その範囲のスコアを付与された検索対象オブジェクトの個数から計算する。なぜなら、正規化前のスコアは離散値であり、ほぼすべての値は異なる値であると考えられるため、情報量を正しく求めることができないためである。

まず、スコアの範囲 $[0, 1]$ を 0 から L ずつ分割する。つまり、各フィールドにおいてとりうるスコアの最大値と最小値の差は L となる。ここで、スコアの範囲 $[0, 1]$ を等間隔に分割するために、 $L = \frac{1}{P}$ を満たす L のみを用いることができることとする。 P は $[0, 1]$ を分割した後のフィールドの個数である。

図4に示すように、スコア $S^*(R_i, O_j)$ が先頭のフィールドから数えて k ($k = 1, 2, \dots, P$) 番目のフィールドに入ったとする。ここでスコア $S^*(R_i, O_j)$ の情報量 $I(R_i, O_j)$ を求める。まず、 k 番目のフィールドに区分されたオブジェクトの数 $F(k)$ を計算する。さらに、検索対象オブジェクトの総数 N と $F(k)$ を用いることによって、オブジェクト O_j の情報量 $I(R_i, O_j)$ は次の式で計算できる。

$$I(R_i, O_j) = -\log_2 \left(\frac{F(k)}{N} \right) \quad (3)$$

この式は、Shannon の情報量の定義⁹⁾ に基づくものであり、フィールドに区分されるスコアの割合を生起確率として考えている。たとえば、あるスコアが区分されたフィールドに他のスコアがほとんど区分されていない場合には、そのスコアは情報量が高いと考えられる。

ところが、 $F(k)$ をそのまま正規化のために用いた場

合、正規化後のスコアが逆転してしまう可能性がある。たとえば、検索対象オブジェクト O_1 は正規化前のスコアが $S(R_j, O_1) = 0.1$ であり、情報量 $I(R_j, O_1)$ は 5 となったとする。また、検索対象オブジェクト O_2 は正規化前のスコアが $S(R_j, O_2) = 0.4$ であり、情報量 $I(R_j, O_2)$ は 1 となったとする。この場合には、正規化前のスコアが O_1 より O_2 のほうが小さいにもかかわらず、正規化後のスコアが O_1 よりも O_2 のほうが大きくなってしまふ。この問題は、 $F(k)$ が単調減少とならない場合に発生する問題である。そこで、 $F(k)$ を基に単調減少関数 $G(k)$ を次のように定義する。

$$\begin{aligned} G(P) &= F(P) \\ G(k-1) &= \max(F(k-1), G(k)) \\ &\quad (k = 2, 3, \dots, P) \end{aligned} \quad (4)$$

以上より、単調減少である $G(k)$ を用いた情報量 $J(R_i, O_j)$ は次のように定義される。

$$J(R_i, O_j) = -\log_2 \left(\frac{G(k)}{N} \right) \quad (5)$$

3.3.3 正規化後のスコアの導出

Standard で正規化したスコアとその情報量を用いて、正規化後のスコア $S'(R_i, O_j)$ を次のように計算する。

$$S'(R_i, O_j) = S^*(R_i, O_j) \cdot J(R_i, O_j) \quad (6)$$

4. 評価実験

本論文における提案手法が有効であることを確かめるために、評価実験を行った。評価実験における目的は3つであり、1) 正規化手法において必要なパラメータ L の最適な値を求め、2) 実際に正しく正規化されていることをスコアの分布から確かめ、3) 従来の正規化手法と比較し、提案手法の有効性を確かめた。

以下では、まず評価実験に用いた類似画像メタ検索エンジンの概要を示し、次に実験の手順について説明を行い、最後に実験結果、考察を示した。

4.1 評価に用いた画像検索システム

本評価実験では画像を検索することのできるメタ検索エンジンを構築した。メタ検索エンジンは3つの画像検索システムから構成され、それぞれ、色ヒストグラム、画像に含まれる物体の形状情報、その模様情報を用いて検索するシステムである。すべての画像検索システムは画像を問合せとして入力とし、出力として検索対象画像群のうち問合せ画像に類似している画像を出力する。検索結果は、検索対象画像それぞれに、3.3.3 項における検索システムが付与したスコア $S'(R_i, O_j)$ を基にした順位とともに、ランクが上位のものから順に出力される。

4.1.1 比較に用いた3つの正規化手法

我々は Montague らの提案した正規化手法を、提案手法の比較対象として用いた。Montague らの提案する手法は3つであり、それらの定義は2章の表1に示したとおりである。

統合関数は、Fox らの提案した CombSUM や CombMNZ を使用しており、それぞれ次の式(7)、(8)で定義されている⁵⁾。

$$IS(O_j) = \sum_{i=1}^N S'(R_i, O_j) \quad (7)$$

$$IS(O_j) = \sum_{i=1}^N (S'(R_i, O_j) \cdot K(R_i)) \quad (8)$$

ここで $K(R_i)$ は検索システム R_i におけるスコアが0より大きい検索対象オブジェクトの数である。

Montague らの実験結果では、SUM と CombSUM を組み合わせた場合、もしくは ZMUV と CombMNZ を組み合わせた場合に最も検索性能が高いことが判明している⁷⁾。ところが、Montague らの実験はテキスト文書検索における実験結果であり、我々の実験環境と同一ではないため、我々の実験結果と Montague らの実験結果が一致しない場合も考えられる。そこで、Montague らの用いた3つの正規化手法をすべて比較対象とした。

4.2 実験方法

我々は、次のような手順で評価実験を行った。

- (1) 4人の被験者によって、15組の問合せ画像、正解画像の組を作成する。
- (2) 問合せ画像をそれぞれ入力し、画像検索システムを用いて検索結果を得る。検索結果はスコアによって順位付けされた画像の集合である。
- (3) 被験者によって作成された正解画像と、画像検索システムによって検索された画像の順位付きリストから、その平均適合率を計算する。平均適合率に関しては、4.3節の式(9)で述べる方法を用いて計算を行う。

使用した画像はおよそ30,000枚であり、Photodisc社から刊行されているCD-ROMに収められているものである。本論文で構築した画像検索システムにおける問合せの検索結果は、適合する画像かどうかという評価が主観的なものとなってしまう、被験者によって異なるものである。つまり、1人の評価者によって作成されたテストコレクションでは、正当な評価を行うことができないと考えられる。また、我々は画像検

索システムの検索精度を評価するためのテストコレクションを発見できなかったため、独自のテストコレクションを構築した。

4.3 評価の指標

情報検索システムの検索精度を計測する目的で用いられる指標の1つに、再現率、適合率がある¹⁾。これら2つの尺度を用いて、ランキングされた検索結果の精度を測定することができる。この場合、1つの問合せに対して複数の再現率、適合率を計算することができ、それらをグラフとして表現したものは“再現率-適合率グラフ”と呼ばれている。ところが、再現率-適合率グラフを用いてメタ検索エンジンの検索精度を測定することは困難である。なぜなら、本実験では正規化手法と統合手法の組合せの数、問合せの数を考えると、非常に多くの再現率-適合率グラフを描く必要がある。ところが、グラフの数が多ければ多いほど、検索システム間の検索精度を比較することは難しい。そこで、1つのシステムから得られた複数個の再現率、適合率を“平均適合率”と呼ばれる評価指標で表現することがしばしば行われる¹⁸⁾。平均適合率の定義は次の式による定義される。

$$\nu = \frac{1}{\sum_{i=1}^N x(O_i)} \sum_{j=i}^N \frac{x(O_i)}{j} \left(1 + \sum_{k=1}^{j-1} x(O_k) \right) \quad (9)$$

ここで、 N を出力画像の総数、 O_i ($i = 1, 2, \dots, N$) を出力順位第 i 位の文書の適合/不適合を表す変数とする。ただし、第 i 位の文書が適合している場合は $x(O_i) = 1$ 、適合していない場合は $x(O_i) = 0$ となる。評価実験においては、平均適合率も再現率-適合率グラフとあわせて用いることによって、メタ検索エンジンの検索精度を測定した。

4.4 実験結果

4.4.1 提案手法における最適なパラメータ

3.3.2項で示したとおり、提案手法ではパラメータとして L を設定する必要がある。我々は、 L の値を0.001から0.5まで変化させ、平均適合率の変化を調査した。統合関数によって平均適合率が異なるため、2つの統合関数(CombSUM, CombMNZ)それぞれの場合について、平均適合率とパラメータとの相関関係を調べた。結果を図5に示す。

これらの図から、 L の値として0.2を用いた場合に、最も平均適合率が高いことが分かった。つまり、スコアの範囲 $[0, 1]$ を5等分にした場合に、情報量が適切な値となることが分かった。

4.4.2 正規化後のスコア分布の変化

次に、提案した手法を用いることによって、スコア

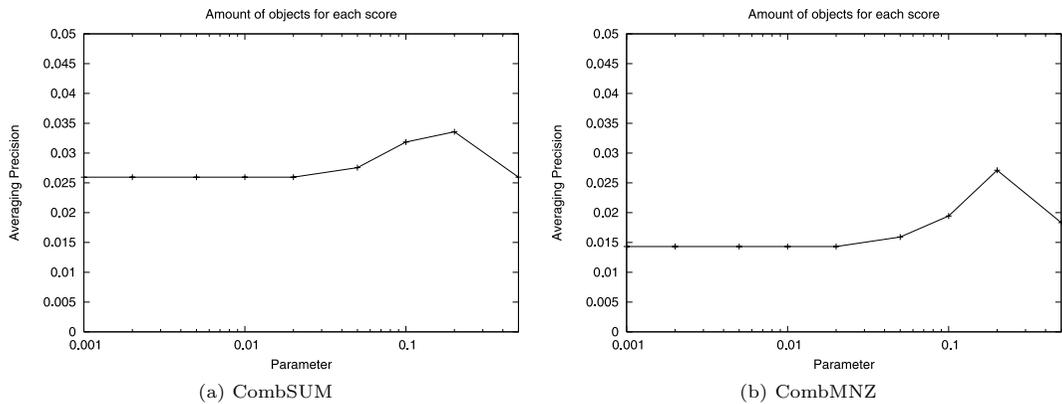


図 5 パラメータ L を変化させた場合の平均適合率の変化
 Fig. 5 Averaging precision by changing the parameter L .

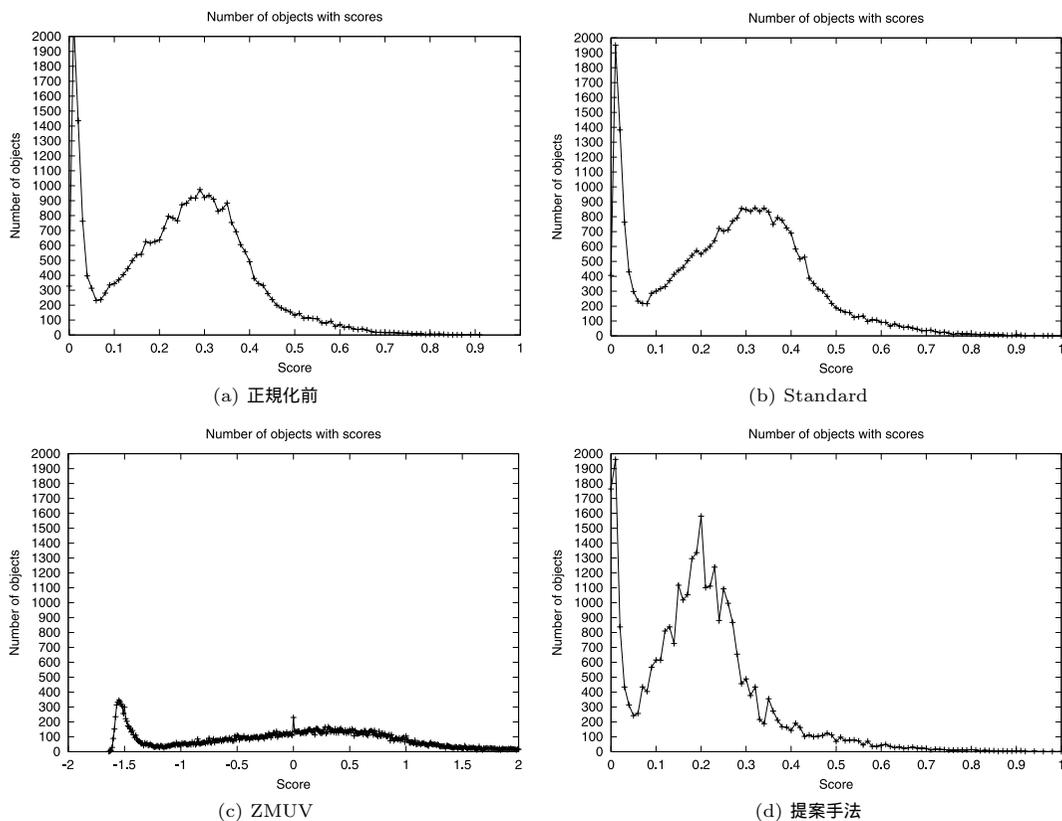


図 6 色ヒストグラムによる検索システムの検索結果の正規化前と正規化後のスコアの分布
 Fig. 6 The distribution of color histogram relevance scores at a query.

の分布がどのように変化したのかを実際に確認した。問合せは、実際に被験者が設定した入力画像のうちの 1 つを用いた。図 6, 図 7, 図 8 に、正規化を行う前のスコアの分布, Standard, ZMUV, 提案手法それぞれで正規化を行ったときのスコアの分布をそれぞれ示す。正規化手法として SUM を用いた場合のスコアの

分布は, Standard を用いた場合のグラフと同じグラフとなるため, 省略した。また, 提案手法におけるパラメータ L の値として 0.2 を用いた。これは, 4.4.1 項における実験において最適なパラメータ L の値が求められているからである。

ここで, 図 6 ~ 図 8 に示した正規化前, 正規化後の

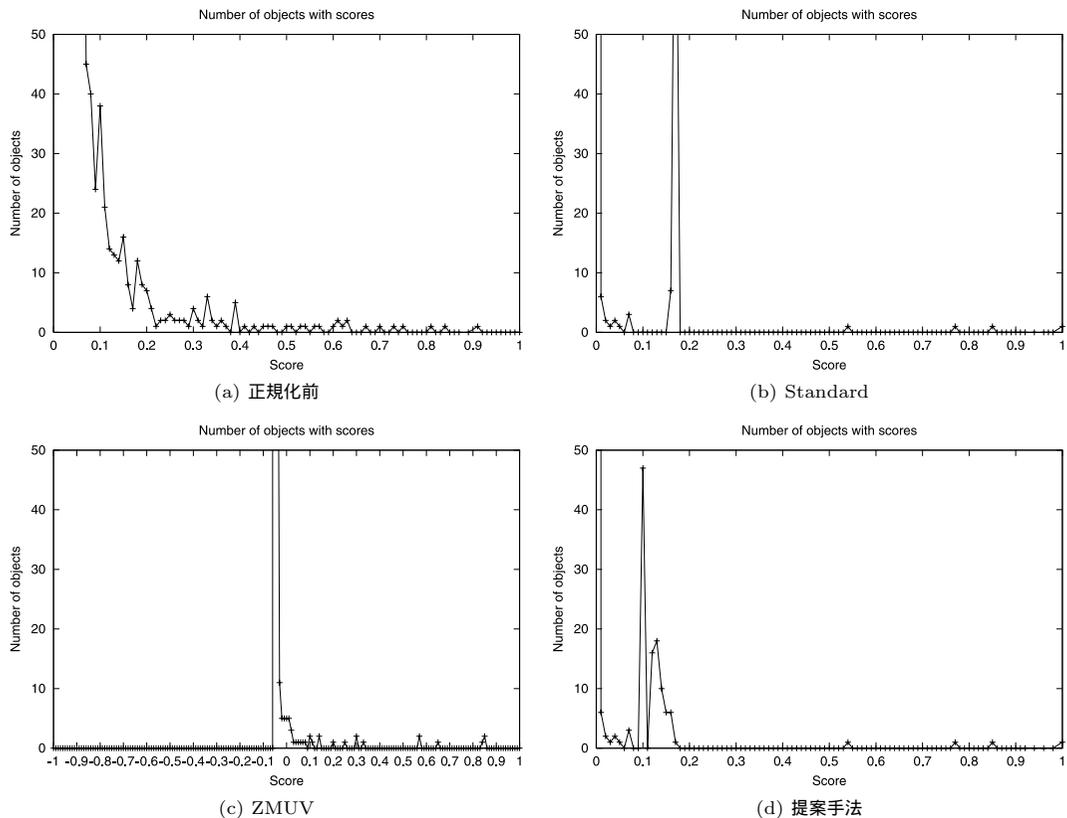


図 7 形状情報による検索システムの検索結果の正規化前と正規化後のスコアの分布

Fig. 7 The distribution of shape relevance scores at a query.

スコアの分布を比較する。まず、正規化前のスコアの分布を見ると、図 6 ではスコア 0 が付与された画像が多く、次に 0.2 から 0.3 のスコアが与えられた画像が多いことが分かる。図 7 でも図 6 における分布と同様に、スコアの値が 0 から 0.1 が付与された画像が多いことが分かる。一方、図 8 ではスコア 0.9 に近い画像が多く、スコア 0 が付与された画像はほとんどない。このような場合、統合関数を用いてこれら 3 つのスコアを統合してしまうと、平均的にスコアの低い検索結果、つまりこの場合では模様情報を基にした検索システムによる検索結果のみが、統合後の検索結果に反映されてしまうことは明らかである。

正規化手法として Standard、もしくは SUM を選択した場合のグラフを見ると、正規化前のグラフとほとんど変化がないことが分かる。つまり、Standard を用いて正規化した場合でも、平均的に模様情報を基にした検索システムによる検索結果におけるスコアは、他の検索結果におけるスコアと比較してもやはり相対的に高いと考えられるため、正規化前と同様、模様情報を基にした検索システムによる検索結果のみが統合

後の検索結果に反映されてしまう問題は解決されていない。

一方、ZMUV を選択した場合のグラフを見ると、平均や分散は統一されているが、同じスコアを持つ検索対象オブジェクト群は、やはり類似したスコアの値として正規化されてしまう。たとえば、模様情報においてスコアの値が 0.9 付近であったものが ZMUV における正規化でやはり 0.4 付近となってしまふ。この場合では、形状情報におけるスコアの値が 0 付近であるものが多いことから、これらのスコアは相互に比較可能であるとはいえない。

しかし、提案手法を用いてスコアを正規化した場合、模様情報におけるスコア分布のように、正規化前のスコアが高い値に集中した場合であっても、正規化後のスコアの値は 0.3 から 0.8 と比較的大きな範囲に分散することが分かる。また、正規化前のスコアの分布が比較的異なっている 3 つのグラフが、正規化を行うことによって比較的スコアの分布している範囲が類似していることが分かる。これは、画像に付与されたスコアに対して求められた情報量が、模様情報におけるス

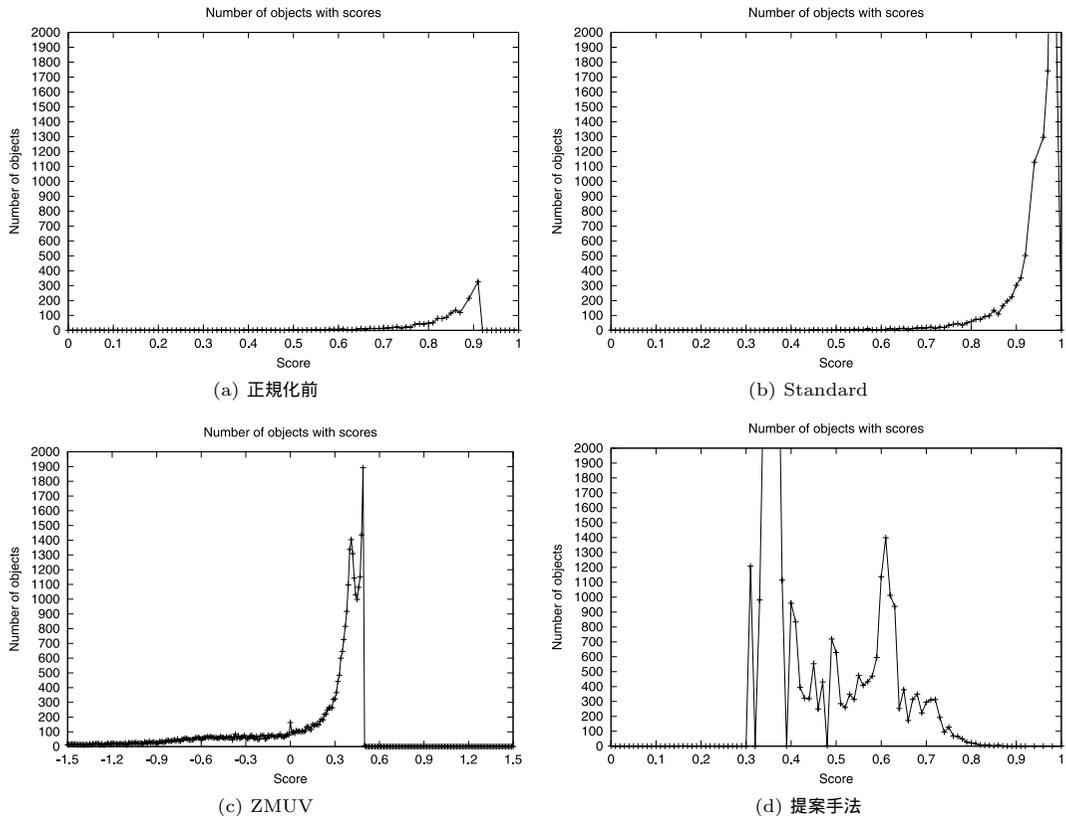


図 8 模様情報による検索システムの検索結果の正規化前と正規化後のスコアの分布

Fig. 8 The distribution of texture relevance scores at a query.

コア分布の場合は比較的低い値となり、他の 2 つの情報におけるスコア分布の場合は比較的高い値となるためである。つまり、模様情報におけるスコアの分布が 0.9 付近に集中しているために、スコアの値が高い点では情報量が低下するためである。

3.2 節における正規化後のスコアの要件から、すべてのスコアの分布の範囲が同じとなる場合に最も良い正規化を行うことができると考えられる。この観点から図 6～図 8 を見ると、提案手法を用いた場合に最もスコアの分布が類似したグラフに変換できていると考えられ、最も良い正規化手法であるといえる。

以上のことから、正規化後のスコアの分布は提案手法を用いることによって、複数の検索システムによって出力されたスコアが、互いに比較可能な値となるように正規化されていることを示すことができた。ところが、以上の実験だけでは正規化を行ったことによる有効性を示したことはならない。そこで、実際に各正規化手法を適用したメタ検索エンジンの検索精度を比較することによって有効性の確認を行う。

4.4.3 正規化を行ったことによる適合率の変化

最後に、4.3 節において実際に正規化を行った場合の再現率-適合率グラフを、図 9 において示し、Standard, ZMUV, SUM を用いた場合と比較した。ここで、4 つの正規化手法と 2 つの統合関数の組合せすべてについて実験を行った。凡例では正規化手法と統合関数の組を表記している。つまり、“Proposed-SUM” という表記は、提案手法を正規化手法として用い、CombSUM を統合関数として用いた場合であることを示している。また、4.3 節で述べたように、図 9 で示した再現率-適合率グラフのみを用いて検索性能を評価することは困難である。そのため、各問合せごとの平均適合率を図 10 で示した。図 10 においては、統合前の検索システムの検索結果、つまり 1 つの特徴量だけを用いた場合の平均適合率に関してもあわせて表記した。

まず、1 つの特徴量だけを用いて検索結果を得た場合と、複数の特徴量による検索結果との比較を述べる。図 10 から、正規化手法として Sum を用いた場合を除くと平均適合率は向上していることが分かる。この

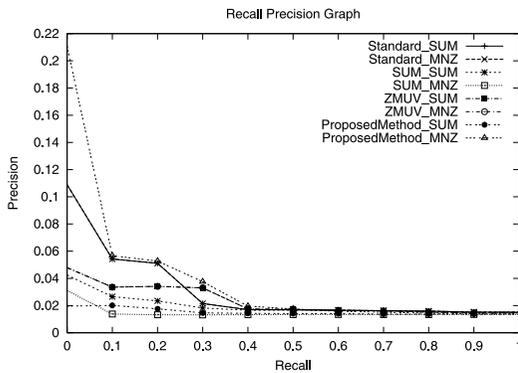


図 9 再現率-適合率グラフ

Fig. 9 Recall precision graph.

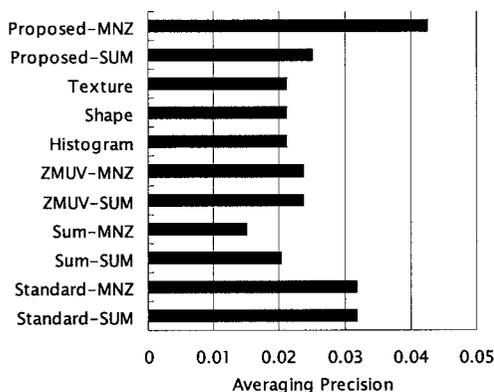


図 10 平均適合率

Fig. 10 Averaging precision.

理由としては、Sum を用いることによってスコアの最大値と最小値の差が小さくなってしまいうため、被験者が適合していると判断した検索対象オブジェクトのスコアと適合していないと判断した検索対象オブジェクトのスコアとの差が小さくなってしまいう点があげられる。

次に、我々の提案手法と、比較対象した正規化手法との平均適合率を比較する。図 10 から、2 つの統合関数のうちの 1 つを用いた場合には最も高い検索精度であることが分かったため、提案手法と CombMNZ を用いた場合の有効性を示すことができた。ところが、提案手法と統合関数 CombSUM を用いた場合には、Standard などと比較して悪い検索精度となってしまうことも分かった。この理由として、我々の正規化手法の目的はスコアのうち目立って高いものに注目している点に対し、CombSUM は平均を求めるための関数であるため、目立って高いものがあまり重要視されない関数である点が考えられる。

以上のように、正規化手法とともに統合関数は非常

に重要な概念であると考えられるが、統合関数によっては我々の提案手法は他の手法よりも劣る場合があると考えられる。つまり、提案手法を用いる場合には統合関数の選択が重要であることが分かった。

5. おわりに

本論文では、メタ検索エンジンにおいて、スコアによって順位付けされた複数の検索結果を統合する際に課題となる、スコアを正規化するための方法として、Shannon の情報量の概念を考慮したスコアの正規化手法についての提案を行った。我々は、スコアが他のスコアと比較して目立って高い値である場合に、利用者がそのスコアが付与された検索オブジェクトを適合していると判断すると仮定した。その仮定に基づいて、ある検索対象オブジェクトのスコアの分布からスコアの情報量を求め、その情報量を用いたスコアの正規化を行った。評価実験を行った結果、提案手法を用いることによって確かに検索精度が向上することが分かった。提案手法は、正規化手法の妥当性について議論した点が特徴である。

今後の研究課題として、次のような点があげられる。

- (1) 今回は、統合関数として CombSUM や CombMNZ を用いた。これらの統合関数はテキスト検索においては良い関数であることが示されているからである。ところがこれらの関数は、画像検索手法やマルチメディア文書検索では適した統合関数であることが示されていない。一方、最適な統合関数は問合せによって変化すると考えられるため、問合せによって自動的に最適な関数を選択する方法について考える必要がある。
- (2) 今回想定している類似画像メタ画像検索エンジンでは、すべての検索システムが同じ画像群を検索対象としている。ところが、たとえば Web 文書を検索するためのメタ検索エンジンのように、それぞれの検索システムは異なる画像群を検索対象とする場面も考えられる。このような場合、1 つの画像に対してすべての検索システムがスコアを付与できるわけではない。この場合の解決方法として、複数の検索システムに共通している検索対象オブジェクトに付与されたスコアを用いて、検索対象となっていないオブジェクトに対するスコアを推定する方法などが考えられる。
- (3) メタ検索システムで正規化を用いない理由として、スコアを統合する際の計算速度が遅い点が

あげられる。本論文では、計算速度についての考察は行わなかったが、従来のテキスト検索において tf-idf を高速に計算する手法¹⁵⁾ や、メタ検索エンジンの検索結果のうち上位 k 件のみを高速に計算する手法⁴⁾ などを用いて、計算速度を向上させる必要があると考えられる。

謝辞 本研究の一部は、文部科学省科学研究費(課題番号: 14780325, 15200010), 文部科学省 21 世紀 COE「フロンティアバイオサイエンスへの展開—細胞機能を支える動的分子ネットワーク」によるものである。ここに記して謝意を表します。また、本論文を査読していただいた編集委員の春本先生、査読者の方々に深く感謝いたします。さらに、テストコレクションの作成に協力していただいた奈良先端科学技術大学院大学植村研究室の皆様にも感謝いたします。

参 考 文 献

- 1) Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, ACM Press (1999).
- 2) Bartell, B.T.: Optimizing Ranking Functions, Ph.D. Thesis, University of California, San Diego (1994).
- 3) Diamond, T.: Information retrieval using dynamic evidence combination, Ph.D. Thesis, Syracuse University (1996).
- 4) Fagin, R.: Combining Fuzzy Information from Multiple Systems, *Journal of Computer and System Sciences*, Vol.58, pp.83–99 (1998).
- 5) Fox, E.A. and Shaw, J.A.: Combination of Multiple Searches, *The 2nd Text REtrieval Conference (TREC2)*, pp.243–252 (1993).
- 6) Lee, J.H.: Analyzing the Effectiveness of Extended Boolean Models in Information Retrieval, Technical Report TR95-1501, Cornell University (1995).
- 7) Montague, M. and Aslam, J.: Relevance Score Normalization for Metasearch, *Proc. 10th ACM International Conference on Information and Knowledge Management (CIKM01)*, pp.427–433 (2001).
- 8) Montague, M. and Aslam, J.: Condorcet Fusion for Improved Retrieval, *Proc. 11th ACM International Conference on Information and Knowledge Management (CIKM02)*, pp.538–548 (2002).
- 9) Shannon, C.E.: A Mathematical Theory of Communication, *Bell System Technical Journal*, Vol.27, pp.379–423 (1948).
- 10) Suzuki, Y., Hatano, K. and Uemura, S.: A Calculation Method of Document Scores for Multimedia Document Retrieval, *IASTED International Conference on Information Systems and Databases (ISDB 2002)*, pp.178–183 (2002).
- 11) Vogt, C.C.: Adaptive Combination of Evidence for Information Retrieval, Ph.D. Thesis, University of California, San Diego (1999).
- 12) Vogt, C.C.: How Much More is Better? Characterizing the Effects of Adding More IR Systems to a Combination, *Content-Based Multimedia Information Access (RIAO)*, pp.457–475 (2000).
- 13) Vogt, C.C. and Cottrel, G.W.: Fusion via a Linear Combination, *Information Retrieval*, Vol.1, No.3, pp.151–173 (1999).
- 14) Vogt, C.C., Cottrel, G.W., Belew, R.K. and Bartell, B.T.: Using Relevance to Train a Linear Mixture of Experts, *The 5th Text REtrieval Conference (TREC5)*, pp.503–515 (2001).
- 15) Witten, I.H., Moffat, A. and Bell, T.C.: *Managing Gigabytes: Compressing and Indexing Documents and Images*, Morgan Kaufmann Publishers (1999).
- 16) 小作浩美, 内山将夫, 井佐原均, 河野恭之, 木戸出正継: WWW 検索における複数検索結果の統合処理とその評価, 情報処理学会論文誌: データベース, Vol.44, No.SIG8 (TOD 18), pp.78–91 (2003).
- 17) 鈴木 優, 波多野賢治, 吉川正俊, 植村俊亮: 複数のメディアで構成された電子文書の検索手法, 情報処理学会論文誌: データベース, Vol.42, No.SIG10 (TOD 11), pp.11–21 (2001).
- 18) 岸田和明: 検索実験における評価指標としての平均精度の性質, 情報処理学会論文誌: データベース, Vol.43, No.SIG2 (TOD 13), pp.11–26 (2002).

(平成 15 年 9 月 20 日受付)

(平成 16 年 1 月 7 日採録)

(担当編集委員 春本 要)



鈴木 優 (学生会員)

1999 年神戸大学工学部情報知能工学科卒業。2001 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。同年同大学院情報科学研究科博士後期課程入学、現在に至る。マルチメディア電子文書検索、携帯端末における情報配信に関する研究に従事。ACM, 日本データベース学会各会員。



波多野賢治(正会員)

1995年神戸大学工学部計測工学科卒業。1999年同大学大学院自然科学研究科博士後期課程修了。博士(工学)。同年から奈良先端科学技術大学院大学情報科学研究科助手。情報検索システム、データベースシステムに関する研究に従事。電子情報通信学会、ACM、IEEE Computer Society、日本データベース学会各会員。



吉川 正俊(正会員)

1980年京都大学工学部情報工学科卒業。1985年同大学大学院工学研究科博士後期課程修了。工学博士。同年京都産業大学計算機科学研究所講師。同大学工学部助教授、奈良先端科学技術大学院大学情報科学研究科助教授を経て、2002年より名古屋大学情報連携基盤センター教授、現在に至る。1989年～1990年南カリフォルニア大学客員研究員、1996年～1997年ウォータルー大学客員准教授。XMLデータベース、多次元空間索引等の研究に従事。電子情報通信学会、ACM、IEEE Computer Society 各会員。日本データベース学会理事。



植村 俊亮(フェロー)

1964年京都大学工学部電気工学科卒業。1966年同大学大学院工学研究科修士課程修了。同年電気試験所(産業技術総合研究所)。1970年マサチューセッツ工科大学電子システム研究所客員研究員、1981年ソフトウェア部プログラム研究室長、1988年東京農工大学教授を経て、1993年から奈良先端科学技術大学院大学情報科学研究科教授。データ工学、データベースシステムの研究に従事。工学博士。IEEE Fellow、電子情報通信学会フェロー。現在、情報処理学会理事、日本情報考古学会理事、データベース振興センター評議員等。