

検索結果を統合するための関数選択手法

鈴木 優[†] 波多野 賢治^{††} 吉川 正俊^{†††}
植村 俊亮^{††} 川越 恭二[†]

現在、多くの検索システムを組み合わせることによって、検索システムの精度を向上させる方法が注目されている。このような検索システムでは、統合関数と呼ばれる、複数のスコアを統合するための関数を変更することによって、精度が向上することが知られている。ところが、最適な統合関数は利用者の問合せによって異なると考えられるため、あらかじめ一意に統合関数を定めることができない。ここで我々は、統合後のスコアの分布から、最適な統合関数を推定することができるのではないかと考えた。そこで本稿では、最適な統合関数を問合せごとに推定する方法として、シャノンの情報量の概念を援用した尺度を用いて、スコアの分布から統合関数の適合度を測定する方法の提案を行う。本提案では、利用者にとって必要な検索対象の数が検索対象全体の数と比較してきわめて少なく、検索結果に含まれる高いスコアの数が少ないとき、その検索結果は十分に正解集合を絞り込んでいると考えることができるため、利用者の検索目的に適した統合関数であると仮定した。評価実験を行うことによって、実際に適した統合関数を選択することができることを示した。

An Automatic Selection Method of Integration Function

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

Currently, many information retrieval systems that combine multiple retrieval results have been proposed to improve the accuracy of these information retrieval systems. In these systems, selections of the combination functions, the functions for combining multiple relevance scores, are important to improve the accuracies. We seem that the most suitable combination function should change by the users' queries. In this paper, we propose an automatic selection method of combination functions. In our method, we automatically select using the distributions of relevance scores. When a retrieval system calculates a small number of high relevance scores, the system can identify relevant retrieval targets. Therefore, we suppose that the sureness of the combination functions should depend on the number of high relevance scores. Then, we propose a calculating method of the sureness of combination function using Shannon's information measure. From our experimental results, we find out that our proposed method can select better combination functions.

1. はじめに

現在、コンピュータネットワークが普及すると共に大量のテキストや画像などが流通しているため、それらを検索するためのシステムが数多く提案されている。例えば、利用者がテキスト中の単語の出現頻度を基準とした検索を行う必要がある場合、利用者の目的に合致した検索システムを用いることによって、最も精度の高い

検索を行うことができると考えられる。ところが、検索目的が明確でない場合や、利用者が全ての検索システムにおいて高精度に検索できるような問合せの種類を把握していない場合、利用者が目的に応じた検索システムを選択することができない。そのため、検索精度が下落すると考えられる。これは、検索システムと利用者の検索目的の相違が原因であると考えられる。

そこで、利用者の検索目的に応じた検索を容易に行うために、複数の異なる検索システムを統合する方法を考える。この検索システムでは、あらかじめ利用者にとって必要であると考えられる複数の検索システムを検索サブシステムとして用意し、各々の検索サブシステムから出力された検索結果を統合する方法である。利用者の

[†] 立命館大学 情報理工学部

College of Information Science and Engineering, Ritsumeikan University

^{††} 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

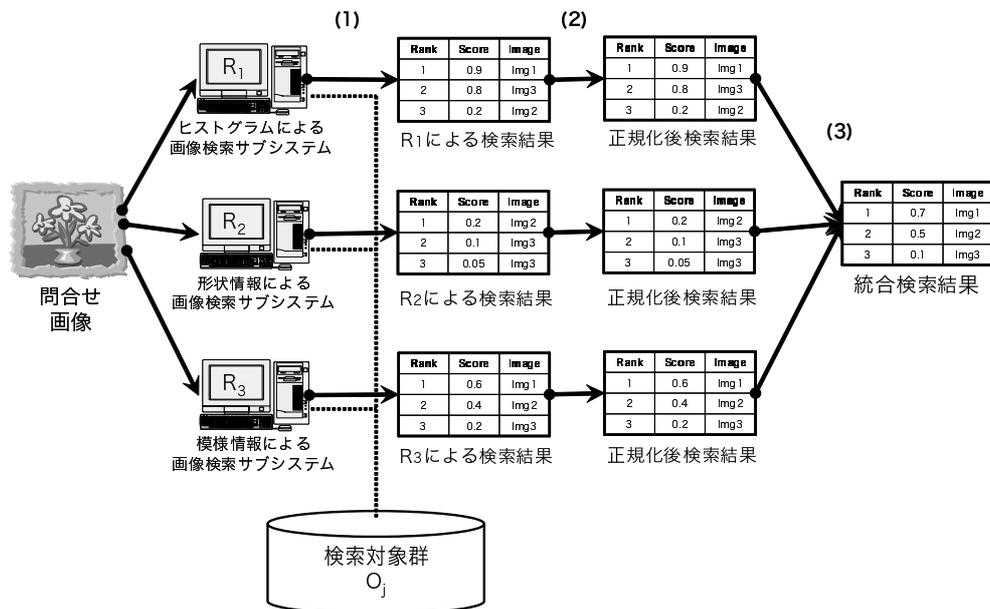


図 1 複数の検索サブシステムを組み合わせた検索システム

Fig. 1 An image retrieval system with multiple information retrieval subsystems

検索目的に適合した検索サブシステムが検索システムに含まれている場合、有効となる検索サブシステムが出力した検索結果が統合後の検索結果に反映されるため、検索精度が上昇すると考えられる。

本研究で想定している検索手法は、World Wide Web (WWW) 上のメタ検索エンジンと類似した検索手法であるといえる。メタ検索エンジンに関する従来の研究から、検索精度を向上させるためには、検索サブシステムの検索結果をどのように統合するかという点に注目する必要があることが知られている¹⁾。ここで、本研究で想定している検索サブシステムとして、全ての検索対象に対してスコアと呼ばれる、問合せに対する検索対象の適合度を計算し、スコアの昇順に出力するシステムを想定している。つまり、検索サブシステムが出力するスコアをどのように統合するかが検索システムの精度に大きな影響を与えると考えられる。

ここで、我々の以前の研究²⁾における仮定と同様の仮定を行う。我々は以前、複数の検索システムを考えたとき、ある検索対象に付与されたスコアの値が高く、他の検索対象群に付与されたスコアの値に低いものが多いとき、高いスコアが付与された検索対象は利用者の検索目的に適

合している可能性が高いと考えられることを示した。なぜなら、一般的に検索対象群の数と比較して利用者の検索目的に適合した検索対象の数は非常に小さいと考えられるため、検索システムが少ない検索対象に高いスコアの値を付与しているということは、その検索システムが十分に適合検索対象を絞り込んでいると考えられるためである。つまり、検索システムが出力したスコアの値の分布から、その検索システムの精度を推定できることを示した。本研究ではこの仮定を、統合関数の自動選択のために用いる。

本稿では、スコアを統合するために最適な統合関数を自動的に選択するための方法として、統合後のスコアの分布から推定する方法についての提案を行う。統合後のスコアに高い値が少なければ少ないほど、その統合関数は検索精度を上昇させることを示すと考えられるため、この性質を利用することによって検索システムの精度を向上させることができると考えられる。

2. 最適な統合関数の自動選択手法

2章では、検索システムの全体像を示し、統合関数を自動的に選択するための詳細を示す。提案手法は、統合後のスコア分布が本稿で定義する理想的なスコアの分布に近い場合に、その統

合関数が適していると考えられる方法である。そのため、どのようなスコアの分布を理想とすべきかを定義する必要がある。そこで、理想的なスコア分布についての議論を行う。さらに、実際の統合後のスコア分布が、本稿で定義した理想的なスコア分布とどの程度類似しているかを測定する必要がある。そこで本稿では、ある統合関数が利用者の目的に適しているかを判断する尺度として、シャノンの情報量³⁾の概念を援用した尺度を考える。

2.1 検索システムの全体図

本稿では、検索サブシステムを組み合わせた検索システムを考える。図1に、本研究で想定している検索システムを画像検索に用いた場合の全体図を示す。検索システムは、大きく分けて三つの部分に分けることができ、それぞれ次のような機能で構成されている。

(1) 検索サブシステムによる検索

利用者が検索システムに問合せを入力した時、その問合せを検索サブシステムに再び入力する。次に、検索サブシステムが、それぞれの検索手法を用いて検索結果を出力する。ここで検索結果は、一つの検索対象に対して、複数の検索サブシステムが計算したスコアを付与している。

(2) スコアの統合

スコアの正規化手法を用いて、各々の検索サブシステムが出力したスコアを正規化する。さらに、統合関数を用いてそれぞれのスコアを統合し統合スコアを計算する。最後に統合スコアを基に検索対象を昇順に並べる。つまり、統合関数の数と同じ数の統合検索結果が得られる。

(3) 統合関数の選択

統合検索結果群におけるスコアの分布から、最も検索精度が良い統合関数を推定し、その統合関数による統合検索結果を利用者に提示する。

以下では、三つの検索プロセスのうち統合関数を選択する部分、つまり(3)の部分を中心に説明する。

2.2 基本的な考え方

本稿で提案する方法は、統合検索結果におけるスコアの分布から検索結果を出力する方法である。そこで、まず検索精度が高いと考えられる

場合のスコア分布、つまり理想的なスコア分布について考える。

利用者が検索システムを用いて検索を行うとき、正解検索対象の数が検索対象全体の数と比較して比較的少ない場合と多い場合の二つが考えられる。また、それぞれの場合によって最適なスコア分布は異なると考えられる。そこで、それぞれの場合における理想的なスコアについての議論を行う。

2.2.1 適合検索対象数が少ない場合

一般に、利用者が入力する問合せに適合する検索対象、つまり正解検索対象集合の数は、検索対象集合全体の数と比較すると非常に少ないと考えられる。例えば、TREC⁴⁾やNTCIR⁵⁾等に代表される、情報検索システムを評価するためのテストコレクションには、問合せとそれに対する正解検索対象集合が示されている。これらのテストコレクションにおいても、やはり正解検索対象集合は検索対象集合全体の数と比べて非常に少ない。

このような条件下では、ある検索システムによって高いスコアを付与された検索対象の数が小さいほうが、その検索システムは十分に検索対象群から正解検索対象を見つけ出すことができているのではないかと考えられる。つまり、検索システムが十分少ない検索対象集合を抽出することができた場合には、その検索システムは検索対象集合のうち的一部分に利用者の検索意図と合致した特徴を発見できていると推定できる。この場合、その検索システムの検索精度が高いことが推定できる。

以上の考え方を、スコアの分布から推定する方法に応用する。ここで、二つの統合関数 F_1 、 F_2 を用いた検索システム $R(F_1)$ 、 $R(F_2)$ を考える。それぞれの検索システムは、統合関数の違いを除いて全く同一のシステムであるとする。さらに、ある問合せをそれぞれの検索システムに入力した場合のスコアの分布は図2(a)、(b)に示すように表現できたとする。これらの図の x 軸はスコアの値を表しており、 y 軸は x 軸で表示されているスコアの値と1との間のスコアが付与されている検索対象の数を表している。例えば、 x 軸が0.1の部分の y 軸の値は、検索システムによって付与されたスコアの値が0.1以上1以下である検索対象の数を表している。

これらの図から、 $R(F_1)$ におけるスコア分布では高いスコアが付与された検索対象の数が $R(F_2)$ の場合よりも少ないことが分かる。つまり、前節で述べた仮定から、 $R(F_1)$ は利用者の検索意図と合致した検索対象だけ対して高いスコアを付与している可能性が高いと考えられる。以上の議論から、 $R(F_1)$ は $R(F_2)$ よりも良い検索システムであると考えられるため、検索システムは統合関数 F_1 を選択する。

2.2.2 正解検索対象数が多い場合

前節では、正解検索対象の数が全検索対象の数と比べて比較的少ない場合の議論であるが、実際には正解検索対象集合の数が非常に多い場合も存在する。例えば、データベースに関連する論文群だけを検索対象とした場合、利用者が検索キーワードとして“データベース”と入力したときには、多くの検索対象を正解とすべきであると考えられる。そこで、このような場合には、どのようなスコア分布が良い分布であるかを考える必要がある。

ここで、前節と同様、統合関数 F_1, F_2 をそれぞれ用いた検索システム $R(F_1), R(F_2)$ のスコア分布が図2であったとする。ところが以前の条件と異なり、正解検索対象数が多いという条件から、検索システムは検索対象群から少数の正解検索対象へ絞り込む必要がない。つまり、検索システムは多くの検索対象に高いスコアを付与した場合(図2(b))であっても、それらの検索対象は全て正解である可能性がある。同様に、少数の検索対象に高いスコアを付与した場合(図2(b))には、低いスコアが付与された検索対象が正解である可能性がある。つまり、利用者の検索意図に合致した統合関数をこれら二つの統合関数から選択することは難しい。

一方、利用者はスコアが高い順に高々数百件だけを閲覧するだけである点に注目する。つまり利用者が閲覧する検索対象は、二つの検索システムのうち少なくともどちらか一方に高いスコアを付与された検索対象だけである。そのため、二つの検索システムのうちどちらを用いた場合であっても、正解検索対象を閲覧できる可能性が高い。つまり、どちらの統合関数を検索システムが選択した場合でも、検索精度は低下しないと考えられる。

以上の議論から、正解検索対象の数の大小にか

かわらず、本提案手法を利用することによって検索精度が向上すると考えられることを示した。次節では、これらの手法を定式化し、実際に統合関数を選択するためのアルゴリズムを示す。

2.3 統合関数の自動選択手法

2.2節で述べた、統合関数の自動選択手法に関する議論を基に、統合関数の自動選択手法を示す。提案手法は次に示す三つの手順で表すことができる。

(1) スコアの分布を求める

まず、統合後のスコアを $[0, 1]$ の範囲となるように正規化を行う。次に、スコアの範囲を等しく分割し、それぞれのスコアの範囲に含まれる検索対象の数を求める。

(2) 各々の検索対象に対して情報量を求める

検索対象集合全体の数に対する各々のスコアの範囲における検索対象の数を、各々の検索対象がそのスコアを付与される確率と考えられる。そこで、そのスコアの値に対する情報量をシャノンの情報量³⁾を用いて定める。

(3) 利用者の検索意図への統合関数の適合度を計算する

それぞれの検索対象に付与された情報量を統合して、ある統合関数が利用者の検索意図にどの程度適合しているかを計算する。

以下、それぞれの手順について説明する。

2.3.1 スコアの分布

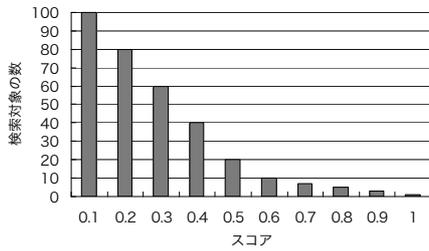
まず、統合関数 F_i ($i = 1, 2, \dots, M$)を用いた検索システム $R(F_i)$ を用いて、検索対象 O_j ($j = 1, 2, \dots, N$)に対してスコア $S(F_i, O_j)$ を計算する。次に、これらのスコア群から、スコアの分布 $G(F_i, k)$ ($k = 0.1, 0.2, \dots, 1$)を計算するために、次式を用いる。

$$G(F_i, k) = |\{O \mid 0 \leq S(F_i, O_j) \leq k\}| \quad (1)$$

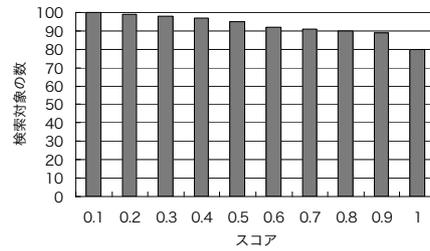
ここで、 O は検索対象集合であり、スコアが0以上 k 以下のものである。さらに、 $|\{O\}|$ は O の要素数、つまり条件を満たした検索対象集合の数である。つまり、 $G(F_i, k)$ はスコアが0以上 k 以下である検索対象数を表している。

2.3.2 検索対象における情報量の計算

次に、スコアの分布を基に、それぞれの検索対象に対して情報量を計算する。統合関数 F_i を用いた場合の、検索対象 O_j の情報量 $I(F_i, O_j)$ を次



(a) 検索システム $R(F_1)$



(b) 検索システム $R(F_2)$

図2 検索システム $R(F_1)$ と $R(F_2)$ を用いた場合のスコア分布

式で示す.

$$I(F_i, O_j) = -\log \frac{G(F_i, k)}{N} \quad (2)$$

つまり、検索対象 O_j が区分されている $G(F_i, k)$ の値を用いて情報量を計算する.

2.3.3 統合関数への適合度の計算

最後に、全ての検索対象に対して計算された情報量を統合して、統合関数への適合度を計算する. 統合関数 F_i への適合度 $T(F_i)$ は、次式を用いて計算する.

$$T(F_i) = \sum_{j=1}^M I(F_i, O_j) \quad (3)$$

以上に示した適合度を、あらかじめ準備した全ての統合関数に対して計算する. そして、適合度が最も高い統合関数は利用者の検索意図に最も適合すると考えられるため、利用者にはその統合関数を用いて計算された検索結果を提示する.

次に、提案手法を用いることによって検索精度が上昇することを、実際の検索システムを用いて示す.

3. 予備実験

本稿で提案した提案手法が有効であることを確かめるために、小規模な予備実験を行った. 本実験では、提案手法を用いた統合関数を用いた場合とランダムに統合関数を用いた場合の比較を行い、提案手法を用いることによって検索精度が向上することを示すことを目的としている.

3.1 実験方法

本実験では、複数の検索システムを比較するために、複数の画像検索サブシステムを組み合わせた検索システムを用いる. 実験の手順を以下に示す.

- (1) 問合せを検索システムに入力する.
- (2) 検索システムはあらかじめ用意した複数の統合関数を用い、検索結果を出力する.
- (3) 提案手法を用いて、利用者の検索意図に合致した統合関数を求める.
- (4) すべての検索結果から、再現率と適合率を計算する.
- (5) 提案手法によって選択された統合関数による再現率、適合率を他の統合関数の場合と比較する.

再現率、適合率を求める場合には、テストコレクションと呼ばれる、問合せとそれに対応する正解検索対象の組を用意する必要がある. ところが、我々は画像検索システムに適用することができなかった. そのため、本実験では第一著者を含む数人の研究協力者の協力を得て画像検索エンジンのためのテストコレクションを作成した. 検索対象となる画像の数は約 30000 枚であり、問合せと正解検索対象の組の数は 15 個である.

本実験で用いた検索サブシステムは三つであり、それぞれ画像の色ヒストグラム、画像に写っている対象物の模様情報と形状情報の三つの特徴量を用いた検索システムである. これは、現在提案されている主な画像検索エンジンは、これら三つの種類の特徴量のうちのどれかが用いられていることが多いためである.

3.2 実験に用いた統合関数

本実験で用いた統合関数は CombSUM, CombMNZ, CombANZ の三つであり、それぞれ次の式 (4),

(5), (6) で定義されている¹⁾.

$$IS(O_j) = \sum_{i=1}^N S(R_i, O_j) \quad (4)$$

$$IS(O_j) = \sum_{i=1}^N (S(R_i, O_j) \cdot K(R_i)) \quad (5)$$

$$IS(O_j) = \sum_{i=1}^N \left(\frac{S(R_i, O_j)}{K(R_i)} \right) \quad (6)$$

3.3 評価尺度

現在、情報検索システムを評価するための尺度として、再現率-適合率グラフ⁶⁾が用いられる。ところが、複数のグラフを比較することが困難であるため、平均適合率⁷⁾を用いて再現率、適合率を一つの値に統合することが多い。平均適合率とは、複数の再現率における適合率の平均を求めたものである。つまり、正解検索対象が検索結果の上位に出現した場合と下位に出現した場合は同様に扱われる。ところが、2.2.2節で述べたように、利用者は検索結果の上位となる一部分だけを閲覧すると考えている。そのため、正解検索対象が検索結果の上位に出現するかどうかを検索結果の下位よりも重視した評価尺度が必要になる。そこで、本実験では評価尺度として、再現率によって重み付けを行う重み付き平均適合率を評価尺度として用いた。

以下では、本稿で用いた重み付き平均適合率について述べる。まず、一つの検索システムに対して11組の再現率、適合率を求める。ここで、再現率は0, 0.1, 0.2, ..., 1となっている。ここで、再現率0を1番目としたときのx番目の再現率を Re_x 、この再現率に対応する適合率を Pr_x とする場合、重み付き平均適合率Aを次式によって計算する。

$$A = \sum_{i=1}^{11} (Re_x \cdot Pr_x) \quad (7)$$

3.4 実験結果および考察

図3に、各々の問合せを用いた場合の適合率を示す。左から順に、三つの統合関数を用いた検索システムの平均適合率のうち、最も高いもの、平均、最も低いもの、提案手法によって選択されたものをそれぞれ表している。

これらの実験結果から、11個の問合せにおいて最も平均適合率が高い統合関数を選択していることが分かる。

本実験で用いた問合せでは、形状情報に注目した問合せが多い。このような場合、形状情報を用

いた検索サブシステムによる検索結果が統合検索結果に反映されている統合関数が良い傾向があることが予期できる。我々は、これらの問合せの傾向からCombMNZを用いることが最も利用者の検索意図に適していると考えていた。そのため、もし提案手法を用いてCombMNZを用いた場合には検索精度が向上すると予想された。実験の結果、9個の問合せにおいてCombMNZが選択されていることが分かった。つまり、利用者の検索意図に合致した統合関数が選択されている可能性が高いことが分かる。

以上の結果から、確かに本提案手法によって検索精度が高いと考えられる統合関数を選択できることが分かった。

4. 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研究との関連について述べる。

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⁸⁾で示されており、次の二つの点が理由として挙げられている。

- (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は、利用者の検索要求に適合していることが多いため。
- (2) 一つの検索システムでは検索対象に対して誤ったスコアを付与され、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。

本研究の目的を実現する方法には、提案手法のようにスコアの分布から最適な統合関数を推定する方法の他に、関連フィードバックを用いる方法や機械学習による方法など様々な方法が考えられる。そこで、それぞれの方法と我々の手法との差異について述べる。

関連フィードバックによる方法とは、一度何らかの検索結果を利用者に提示した後、利用者によって必要であると考えられる検索対象を利用者自身によって選択することによって、統合関数を選択する方法である。利用者自身の検索対象の選択を統合関数の選択に直接反映させることができるため、利用者の検索意図が反映されやすく、検索精度は大幅に向上すると考えられる。ところが、この方法では検索システムが最初に提

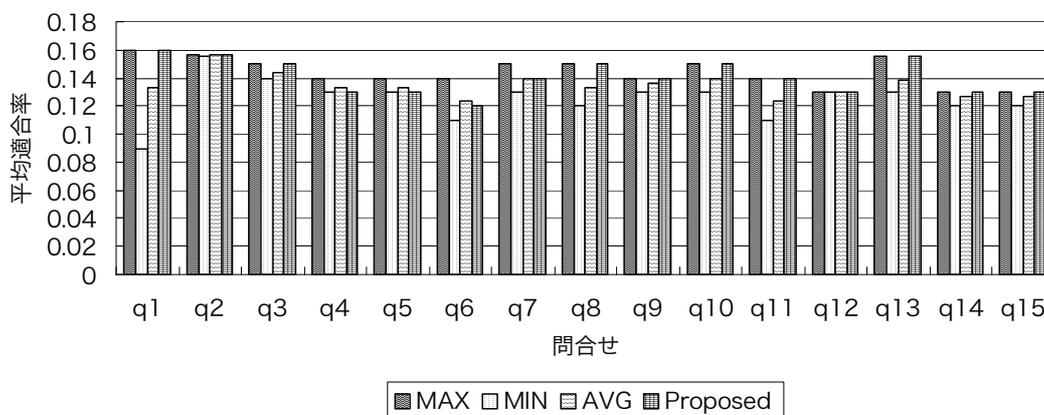


図3 重み付き平均適合率の比較
Fig.3 Comparison of weighted average precision ratio.

示する検索結果に利用者の検索意図を反映させることができない。さらに、最初に提示する検索結果の上位に正解検索対象が含まれていない場合には、利用者が必要であると考えている検索対象を選択することができず、検索精度が低下する場合があると考えられる。つまり、提案手法を用いた後の検索結果に対して関連フィードバックを用いることによって、さらに検索精度を向上することも考えられる。

一方、ある利用者にとって特定の統合関数を用いることが良いと推定される場合、利用者による複数の問合せ、正解検索対象の入力を用いて、利用者ごとに特定の統合関数を推定することが機械学習などの方法を使って可能であると考えられる。ところが、我々は利用者の問合せごとに最適な統合関数は異なると考えており、利用者が連続して複数の検索を行った場合であっても、それらの検索意図は相互に異なると考えている。そのため、利用者の検索意図に対する傾向だけを用いて統合関数を定めた場合に、検索意図の差異が統合関数の選択に反映されず、利用者の検索意図に合致した統合関数を選択することができないと考えた。

以上の議論から、提案手法では、利用者によるフィードバックを行わない方法であり、しかも利用者の検索意図の長期的傾向に基づかない手法を用いた。

5. おわりに

本稿では、検索システムの精度を向上させるために、統合関数を問合せごとに自動的に選択する方法を提案した。ここで、検索対象に付与されたスコアから統合関数ごとにスコア分布を求め、統合関数に対して利用者の検索目的への適合度を計算した。さらに、スコア分布から適合度を求めるためにシャノンの情報量を用いた。最後に、予備実験を行うことによって、提案手法を用いて選択した統合関数は確かに利用者の検索意図に適合していると考えられることが分かった。今後の研究課題は次の通りである。

- 本稿では予備実験として、三つの統合関数から一つの統合関数を選択する実験を行った。ところが、情報検索の分野では今まで29種類の統合関数が提案されていることがJoon⁹⁾によって示されている。そこで、これら29種類の統合関数を用いた実験を行って、提案手法が実際に有効であることを示す必要があると考えられる。
- 本実験では三つの検索サブシステムを用いたが、これらの検索サブシステムが検索精度の向上に寄与していない場合もあると考えられる。つまり、これらの検索サブシステムのうち二つを組み合わせただけの場合や一つだけを用いた場合には、三つ全てを用いた場合よりも検索精度が向上する場合があると考えられる。このような場合、提案手法を拡張し、二つの検索サブシステムを組み合わせただけの場合の検索結

果からスコア分布を求め、適合度を求めることによって、さらに検索システムの検索精度が向上すると考えられる。そこで、提案手法による特徴量の選択手法を実験によって示すことを考えている。

- 提案手法では、検索システムは一つの間合せについて複数の検索結果を計算しなければならず、検索対象の数や統合関数の数、検索サブシステムの組合せの数などを考慮すると莫大な計算量が必要となることが予想される。ここで、複数の検索結果を算出する部分は複数の計算機で並列に計算することが可能であるため、統合関数を選択する部分の高速化が必要になると考えられる。そこで、検索システムの高速化手法について考えなければならない。

参 考 文 献

- 1) Fox, E. A. and Shaw, J. A.: Combination of Multiple Searches, *The Second Text REtrieval Conference (TREC2)*, pp. 243 – 252 (1993).
- 2) 鈴木優, 波多野賢治, 吉川正俊, 植村俊亮: 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, *情報処理学会論文誌: データベース*, Vol. 45, No. SIG 4 (TOD 21), pp. 1 – 2 (2004).
- 3) Shannon, C. E.: A Mathematical Theory of Communication, *Bell System Technical Journal*, Vol. 27, pp. 379 – 423 (1948).
- 4) of Standards, N. I. and Technology: Text REtrieval Conference (TREC). <http://trec.nist.go.jp/>.
- 5) 国立情報学研究所: NTCIR 情報検索システム評価用テストコレクション構築プロジェクト. <http://research.nii.ac.jp/ntcir/>.
- 6) Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, ACM Press (1999).
- 7) 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(TOD13), pp. 11 – 26 (2002).
- 8) Diamond, T.: *Information retrieval using dynamic evidence combination*, PhD Thesis, Syracuse University (1996).
- 9) Lee, J. H.: Analyzing the Effectiveness of Extended Boolean Models in Information Retrieval, Technical Report TR95-1501, Cornell University (1995).