

検索質問に対する検索エンジンの動的選択手法に関する有用性評価

水野 広大[†] 鈴木 優[†] 川越 恭二[†]

[†] 立命館大学大学院理工学研究科

〒 525-8577 滋賀県草津市野路東 1-1-1

E-mail: [†]mizuno@coms.ics.ritsumei.ac.jp, ^{††}suzuki@ics.ritsumei.ac.jp, ^{†††}kawagoe@is.ritsumei.ac.jp

あらまし 本稿では、利用者の検索質問に対して動的に検索エンジンを選択する手法の提案を行い、その有用性を検証する。利用者がある特定の目的で検索を行うとき、その目的のために開発された検索エンジンを選択する。ところが、利用者が全ての検索エンジンの特徴を把握することは困難である。そのため、特定の目的のために開発された検索エンジンを適切に選択することができず、適切な検索結果が得られない場合がある。そこで本研究では、検索質問に対して検索エンジンが出力する検索結果数に着目し、検索質問に対する検索エンジンの適合度を算出する。具体的には、検索質問に対する検索結果数の期待値の差および検索エンジンが他の検索質問に対して出力する検索結果数の期待値との相対的な差を用いて、検索エンジンの適合度算出を行う。評価実験ではまず、この二つの差を用いる比率について最適値を求めた。最後に、提案手法と手動選択による検索エンジン選択を比較して、平均精度の検証を行った。検証結果によって、特定の目的のために開発された検索エンジンを適切に選択することが可能となり、適切な検索結果を得ることが可能となることを示した。

キーワード 情報検索, 検索エンジン選択, 検索結果数

A Search Engine Selection Method for User's Information Need

Kodai MIZUNO[†], Yu SUZUKI[†], and Kyoji KAWAGOE[†]

[†] Graduate School of Science and Engineering, Ritsumeikan University

Nojihigashi 1-1-1, Kusatsu, Shiga, 525-8577 Japan

E-mail: [†]mizuno@coms.ics.ritsumei.ac.jp, ^{††}suzuki@ics.ritsumei.ac.jp, ^{†††}kawagoe@is.ritsumei.ac.jp

Abstract In this paper, we propose a dynamic search engine selection method for users' information need. When users retrieve on the Internet, the users can select the best search engine for their information need. However, one of the most important problem is that the users cannot understand features of all search engines. Consequently, because the users cannot select the best search engine, the users cannot get the best retrieval results. In this paper, we focus that the search engine not matching user's information need outputs less retrieval result than the best one, and we calculate search engine's matching score for the user's information need. As a result, users can select the best search engine using the score that our system calculated, and users can get the best retrieval result.

Key words Information Retrieval, Search Engine Selection, Number of Retrieval Result

1. はじめに

利用者が Web ページや画像といったコンテンツを検索するために、様々な検索エンジンが開発されている。このとき、利用者が利用する全ての検索エンジンの特徴を把握することが可能であれば、利用者はその目的、意図に応じて適切な検索エンジンを選択する。ここで適切な検索エンジンとは、全ての検索エンジンと比較して利用者の検索意図を満たす検索エンジンである。例えば、あるアーティストの画像を必要とする利用者が適切な検索エンジンを選択するとき、適切な検索結果はあるアー

ティストの画像であると考えられる。そのため、他の検索エンジンと比較して画像に特化した検索結果を出力する点において、利用者にとって適切な検索エンジンは画像検索エンジンであることが考えられる。つまり、利用者はあるアーティストの画像に対して適切な検索エンジンとして、画像検索エンジンを選択することができる。また、検索エンジン選択において、利用者は検索質問に対してある検索エンジンは適切なのかといった特徴把握を用いて、適切な検索エンジンを選択することが考えられる。

ところが、利用者が全ての検索エンジンの特徴を把握するこ

とは困難である。そのため、適切な検索エンジンを選択することができず、適切な検索結果が得られない場合がある。例えば、カレーの献立に関するコンテンツを必要とする利用者が適切な検索エンジンを選択するとき、適切な検索結果はカレーの献立に関するコンテンツであることが考えられる。ところが、Web ページ検索エンジンや献立検索エンジンといった検索エンジンは、カレーの献立に関するコンテンツを検索対象としていると考えられる。この場合、利用者は適切な検索エンジンの選択を行うために、カレーの献立に関するコンテンツの検索を行う検索エンジン全ての特徴を把握する必要があることが考えられる。ところが、利用者が検索エンジン全ての特徴を把握することは困難である。そのため、利用者はカレーの献立に関する適切な検索エンジンを選択することができず、適切な検索結果を得られない場合が考えられる。

そこで本研究では、利用者が適切な検索結果を得るために、利用者の検索質問に応じて動的に検索エンジンを選択する手法を提案する。提案手法では、検索質問に対して検索エンジンが出力する検索結果数に着目し、検索質問に対する検索エンジンの適合度を算出する。具体的には、検索質問に対する検索結果数における二つの差を用いて、各検索エンジンが検索質問に対して適している度合いを示す適合度を算出する。一つは、検索質問に対する検索結果数の期待値の差である。この差を用いて、対象検索エンジンにおける検索結果数の期待値の最大値と各検索エンジンの検索結果数の期待値との割合算出を行う。もう一つの差は、検索エンジンが他の検索質問に対して出力する検索結果数の期待値との相対的な差である。この差を用いて、各検索エンジンにおける検索結果数の期待値の最大値と各検索エンジンの検索結果数の期待値との割合算出を行う。算出した二つの割合を用いて、検索質問に対する各検索エンジンの適合度を算出する。算出された検索エンジンの適合度によって、利用者は他の検索エンジンにおける検索結果数および他の検索質問における検索結果数と比較したある検索エンジンの検索結果数の割合を把握することが可能となる。そのため、利用者は算出した適合度を利用することによって、適切な検索エンジンを選択することが可能となり、適切な検索結果を得ることが可能となる。

2. 関連研究

本研究に関連する研究として、複数のデータベースから検索質問に適切なデータベースの検索を行う研究がある。Gravano ら [1] [2] は、検索質問に適切な文書データベースの検索を行うために、検索質問に対する文書データベースの適合度を算出している。Gravano らの手法では、検索質問を構成する各キーワードの出現期待値を用いて検索質問に対する文書データベースの適合度を算出して、算出した適合度を用いて検索質問に適切な文書データベースの検索を行っている。ただし、文書データベースの検索を行うために、検索質問を構成する各キーワードの出現期待値を用いていることから、他の検索質問に対する文書データベースの特徴を考慮していない。そのため、他の検索質問に対する文書データベースの特徴と比較して、検索質問に対する文書データベースの特徴を把握することは困難である

と考えられる。一方で、本研究は検索質問に対する検索エンジンの適合度を算出するために、検索エンジンが他の検索質問に対して出力する検索結果数の期待値との相対的な差を用いている。そのため、本研究では検索結果数の期待値を用いて他の検索質問に対する検索エンジンの特徴を把握を行うことを可能としている。

一方で、検索エンジンの検索結果のリファインを行うことによって、利用者の検索質問に対して適切な検索結果を与える研究がある [3] [4]。高木ら [5] は、フルテキスト検索において、検索質問に適合した各文書に対して文書重要度の付与を行い、利用者の検索質問に対して適切な検索結果の出力を行っている。利用者の検索質問を用いて、検索エンジンの出力する検索結果が検索質問に適しているかの判断を行う点において、本研究と類似している。ただし、検索エンジンが出力した適合文書に対して、文書重要度によって順序付けを行っている。そのため、複数の検索エンジンが存在した場合に、文書重要度の付与を行うことに適している検索エンジンを判断することを考慮していない。この点において、検索質問に対して適切な検索エンジンの選択を行い、検索質問によって動的にクローラ範囲の変化する本研究と異なっている。

3. 検索質問に対する検索エンジンの動的選択手法

利用者が検索を行うとき、利用者は様々な検索エンジンから、利用者の検索意図を満たす検索エンジンの選択を行う。ところが、検索エンジンの特徴はそれぞれ異なっていることが考えられる。このことに対し、利用者が全ての検索エンジンの特徴を把握することは困難である。そのため、利用者が適切な検索エンジンを選択することができず、適切な検索結果を得られない問題が考えられる。

例として、カレーの献立に関する検索結果を必要とする利用者が適切な検索エンジンを選択する場合を考える。ここで、利用者の選択対象となる検索エンジンは、Web ページ検索エンジンである Google および献立検索エンジンである COOKPAD とする。この場合、Web ページ検索エンジンや献立検索エンジンといった検索エンジンは、カレーの献立に関するコンテンツを検索対象としていると考えられる。そのため、利用者は適切な検索エンジン選択のために、2 個の検索エンジンの特徴を把握する必要があることが考えられる。ところが、Web ページ検索エンジンおよび献立検索エンジンは、カレーの献立に関するコンテンツを検索対象としていると考えられるため、Google および COOKPAD のうち、どちらの検索エンジンが適切な特徴をもつのかを利用者が把握することは困難である。そのため、利用者が適切な検索エンジンを手動選択した場合は、利用者はカレーの献立という検索質問に対して、適切な検索エンジンを選択することができない場合が考えられる。

そこで本研究では、利用者が適切な検索エンジンを選択するために、検索質問に対する検索エンジンの適合度を算出する。提案手法を用いたシステムが算出した適合度を利用することによって、利用者は検索質問に対して適切な検索エンジンの選択を

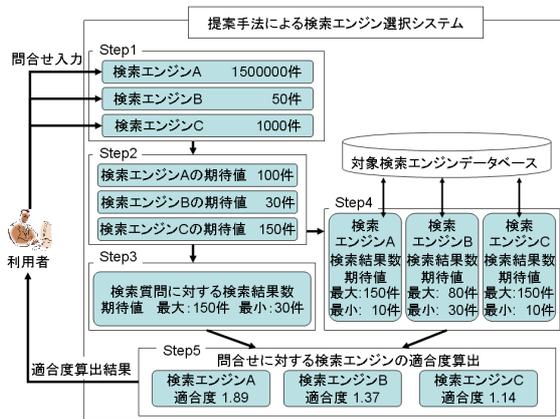


図 1 提案手法による検索エンジン選択システムの概要

行う．例として，前段落で述べた例を用いる．利用者が「カレー 献立」という検索質問に適していない Google を選択した場合，適している COOKPAD を選択した場合と比較して，検索結果数の期待値は少なくなることが考えられる．このことにより，検索質問「カレー 献立」に対して，2 個の検索エンジンにおける検索結果数の期待値の最大値と COOKPAD における検索結果数の期待値との割合算出を行う．また，COOKPAD が「カレー 鍋 献立」という検索質問に適していない場合，COOKPAD が「カレー 献立」という検索質問に適している場合と比較して，検索結果数の期待値は少なくなることが考えられる．このことにより，COOKPAD の検索結果数における期待値の最大値と検索質問「カレー 献立」に対する COOKPAD の検索結果数の期待値との割合算出を行う．算出した二つの割合を用いて，検索質問「カレー 献立」に対する COOKPAD の適合度を算出する．同様の割合および適合度算出により，検索質問「カレー 献立」に対する Google の適合度を算出する．利用者は算出した適合度を利用することによって，適切な検索エンジンを選択することが可能となり，適切な検索結果を得ることが可能となる．

3.1 処理の概要

提案手法による検索エンジン選択システムの概要を図 1 に示す．本システムでは，対象検索エンジンが検索質問に対して出力する検索結果数の期待値における最大値および最小値を用いて，検索エンジンの適合度を算出する．同様に，各検索エンジンにおける検索結果数の期待値における最大値および最小値を用いて，検索エンジンの適合度を算出する．利用者は，システムが算出した適合度を利用することによって，適切な検索エンジンの選択を行う．

適合度算出のために，利用者が過去に行った全ての検索質問に対する各検索エンジンの検索結果数を前処理によって取得する．そして，対象検索エンジンデータベースの構築を行う．対象検索エンジンデータベースには，利用者が過去に行った全ての問合せと，検索質問に対する各検索エンジンの検索結果数によって得られた検索結果数の期待値を格納する．

利用者は，検索意図を表すキーワードを検索質問として入力する．システムは，Step1 として各検索エンジンが検索質問に

対して出力する検索結果から，各検索エンジンの検索結果数を抽出する．次に，Step2 では検索質問に対する各検索エンジンの検索結果数を用いて，各検索エンジンの検索結果数の期待値を算出する．そして，Step3 では検索質問に対する対象検索エンジンの期待値の最大値および最小値を抽出する．一方，Step4 として，対象検索エンジンデータベースに格納された各検索エンジンの検索結果数の期待値により，各検索エンジンにおける検索結果数の期待値の最大値および最小値を抽出する．検索質問に対する対象検索エンジンの検索結果数の期待値および各検索エンジンにおける検索結果数の期待値を取得すると，システムは Step5 として検索質問に対する各検索エンジンの適合度を算出する．最後に，システムは利用者に対して適合度算出結果を出力する．

3.2 検索質問に対する検索エンジンの適合度算出

提案手法では，検索質問に対して検索エンジンが出力した検索結果数によって，検索質問に対する検索エンジンの検索結果数の期待値を算出する．検索質問 Q_h が a 個の単語で構成されるとき，検索質問 Q_h を (1) 式に示す．

$$Q_h = (q_1, q_2, \dots, q_a) \quad (1)$$

検索質問 Q_h ($h=1, 2, \dots, M$) に対して検索エンジン E_i ($i=1, 2, \dots, N$) は $E_i(Q_h)$ 件の検索結果数を得る．また，検索エンジン E_i が検索を行う総コンテンツ数を D_{E_i} 件とする．このとき，検索質問 Q_h に対する検索エンジン E_i の検索結果数の期待値 e_{Q_h, E_i} を (2) 式に示す．

$$e_{Q_h, E_i} = D_{E_i} \times \left(\frac{E_i(q_1)}{D_{E_i}} \times \frac{E_i(q_2)}{D_{E_i}} \times \dots \times \frac{E_i(q_a)}{D_{E_i}} \right) \quad (2)$$

次に，提案手法は，検索質問に対する検索エンジンの適合度を算出する際に，検索結果数の期待値から得られる四つの値を用いる．一つは，検索質問に対する対象検索エンジンの期待値における最大値および最小値である．この二つの値によって，対象検索エンジンにおける期待値の最大値と検索質問における各検索エンジンの期待値との割合を算出する．算出する割合の正規化を行うために，二つの期待値から対象検索エンジンにおける期待値の最小値を引くことを行う．ここで，検索結果数の増加に伴って，1 件の検索結果が検索結果数に占める割合が小さくなることが考えられるため，割合算出に用いる全ての期待値に対数进行することとする．検索質問 Q_h に対して最大検索結果数を出力した検索エンジンを E_{\max} ，最小検索結果数を出力した検索エンジンを E_{\min} とする．また，対象検索エンジンにおける最大検索結果数の期待値を $e_{Q_h, E_{\max}}$ 件，最小検索結果数の期待値を $e_{Q_h, E_{\min}}$ 件とする．このとき，検索エンジン E_{\max} と検索エンジン E_i ($i=1, 2, \dots, N$) との検索結果数の期待値の割合 $R_i(Q_h)$ の算出式を (3) 式に示す．

$$R_i(Q_h) = \frac{\log(e_{Q_h, E_i} + 1) - \log(e_{Q_h, E_{\min}} + 1)}{\log(e_{Q_h, E_{\max}} + 1) - \log(e_{Q_h, E_{\min}} + 1)} \quad (3)$$

もう一つは，対象となる各検索エンジンにおける期待値の最大値および最小値である．この二つの値によって，各検索エンジンの期待値における最大値と検索質問における各検索エンジ

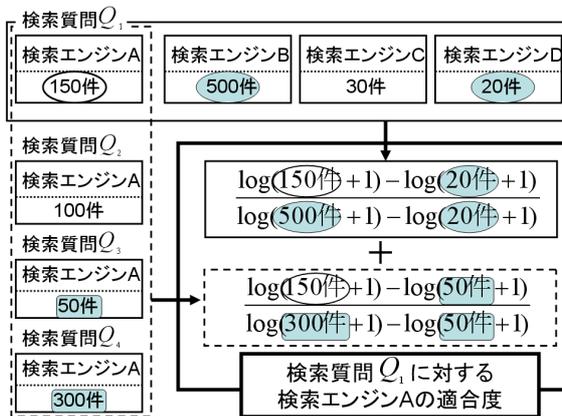


図2 検索結果数による検索エンジンの適合度算出手法

の期待値との割合を算出する．この割合は (3) 式と同様に，二つの期待値から検索質問における検索エンジンの期待値の最小値を引くことを行い，全ての期待値に対数をかけるものとする．検索エンジン E_i は， M 件の検索質問において，検索質問 Q_{\max} に対して最大検索結果数，検索質問 Q_{\min} に対して最小検索結果数を出力する．また，検索質問 Q_{\max} における検索エンジン E_i の期待値を e_{Q_{\max}, E_i} 件，検索質問 Q_{\min} における検索エンジン E_i の期待値を e_{Q_{\min}, E_i} 件とする．このとき，検索エンジン E_i における期待値の最大値と検索質問 Q_h に対する検索エンジン E_i の期待値との割合 $r_i(Q_h)$ の算出式を (4) 式に示す．

$$r_i(Q_h) = \frac{\log(e_{Q_h, E_i} + 1) - \log(e_{Q_{\min}, E_i} + 1)}{\log(e_{Q_{\max}, E_i} + 1) - \log(e_{Q_{\min}, E_i} + 1)} \quad (4)$$

算出された二つの割合を用いて，システムは検索質問に対する各検索エンジンの適合度を算出する．検索質問 Q_h に対する検索エンジン E_i の適合度 $S_i(Q_h)$ の算出式を (5) 式に示す．ここで， α ($0 \leq \alpha \leq 1$) は二つの割合の比率を示すパラメータとする．

$$S_i(Q_h) = \alpha R_i(Q_h) + (1 - \alpha)r_i(Q_h) \quad (5)$$

本研究における検索エンジンの適合度算出の具体例を図2に示す．図2の例では (5) 式のパラメータ α は考慮しないものとする．図2では，検索質問 Q_1 に対する検索エンジン A の適合度を算出している．まず，システムは検索質問 Q_1 に対して検索エンジン A, B, C, D が出力する検索結果数の期待値の最大値および最小値を抽出する．図2により，検索エンジン A, B, C, D における期待値の最大値は 500 件，最小値は 20 件となる．そして，この二つの値を用いて，検索エンジン A, B, C, D における検索結果数の最大値と検索質問 Q_1 に対する検索エンジン A における検索結果数の期待値との割合を算出する．図2により，検索質問 Q_1 における対象検索エンジン A が出力した検索結果数の期待値の割合は， $\frac{\log(150+1) - \log(20+1)}{\log(500+1) - \log(20+1)} = 0.622$ となる．次に，システムは検索質問 Q_1, Q_2, Q_3, Q_4 に対して検索エンジン A が出力した検索結果数の期待値における最大値および最小値を算出する．図2により，検索質問 Q_1, Q_2, Q_3, Q_4 における検索エンジン A の期待値の最大値は 300 件，

最小値は 50 件となる．そして，この二つの値を用いて，検索質問 Q_1, Q_2, Q_3, Q_4 における検索エンジン A の検索結果数における最大値と問合せ Q_1 における対象検索エンジン A の検索結果数の期待値との割合を算出する．図2により，検索質問 Q_1 における対象検索エンジン A が出力した検索結果数の期待値の割合は， $\frac{\log(150+1) - \log(50+1)}{\log(300+1) - \log(50+1)} = 0.611$ となる．最後に，算出した二つの値を用いて，システムは検索質問 Q_1 に対する検索エンジン A の適合度を算出する．図2により，検索質問 Q_1 における対象検索エンジン A が出力した検索結果数における期待値の割合は 0.622，検索質問 Q_1 における対象検索エンジン A が出力した検索結果数における期待値の割合は 0.611 となる．この二つの値によって，検索質問 Q_1 に対する検索エンジン A の適合度は， $0.622 + 0.611 = 1.23$ となる．

図2と同様の適合度算出によって，システムは検索質問 Q_1 に対する検索エンジン B, C, D の適合度をそれぞれ算出する．算出した適合度を利用することにより，利用者は検索質問 Q_1 に対して適切な検索エンジンの選択を行うことが可能となる．

4. 評価実験

4.1 実験方法

提案手法によってシステムが算出した適合度 1 位の検索エンジンにおける検索結果上位 30 件の再現率および精度と，既存手法における検索結果上位 30 件の再現率および精度を比較することによって，提案手法の評価を行う．また，提案手法が再現率および精度の高い検索エンジンを選択していることを評価するために，実験に用いた検索エンジンにおいて最も平均精度の高かった検索エンジンとの比較を行う．提案手法として，以下の三つの手法を用いる．

- 検索エンジンの適合度算出に，対象検索エンジンにおける期待値の最大値と検索質問における各検索エンジンの期待値との割合 (3) 式だけを用いた手法
 - 検索エンジンの適合度算出に，各検索エンジンの期待値における最大値と検索質問における各検索エンジンの期待値との割合 (4) 式だけを用いた手法
 - 検索エンジンの適合度算出に，(3) 式および (4) 式によって算出された二つの割合を用いた (5) 式を用いた手法
- また，利用者が検索を行うときに，様々な検索エンジンの特徴

表1 各検索エンジンの総文書数

検索エンジン	総文書数 (件)
Google	100000000000
COOKPAD	481801
e-words	9490
Wikipedia	562398

表2 各検索エンジンの最大および最小検索結果数

検索エンジン	最大検索結果数 (件)	最小検索結果数 (件)
Google	10450000000	14800
COOKPAD	212319	8
e-words	762	8
Wikipedia	422041	1

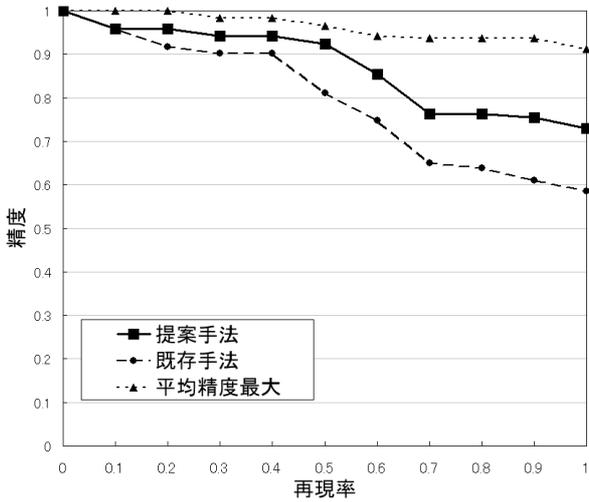


図 3 提案手法および既存手法による再現率精度曲線

を利用者が把握することによって、検索エンジンの選択を行うことが考えられる。そのため、既存手法として、問合せに対して適切な検索エンジンを利用者が手動選択する手法を用いる。

本実験では、まず(5)式におけるパラメータ α の最適値の検証を行う。パラメータ α の最適値は、パラメータ α を 0.1 ずつ変化させた場合に、適合度 1 位に選択された検索エンジンの平均精度が最大値を示した場合のパラメータ α の値とする。次に、既存手法の再現率および精度と比較して、提案手法によってシステムが選択した検索エンジンの再現率および精度が高い値を示すことによって、提案手法の有用性を示す。ここで、提案手法はパラメータ α の最適値を用いることとする。最後に、提案手法が検索質問に対して最適な検索エンジンを選択していることを検証するために、検索質問に対して最も高い平均精度を示した検索エンジンと提案手法によって選択された検索エンジンとの比較を行う。比較を行うために、検索質問に対して最も高い平均精度を示した検索エンジンおよび提案手法によって選択された検索エンジンの再現率および精度を用いる。

本実験では、検索エンジン 4 個を、本実験の対象となる検索エンジンとして用いた。検索質問には、豚肉の Pasta に関する検索質問である「豚肉 Pasta」、vista の LAN の設定に関する検索質問である「vista LAN 設定」、マシ・オカの wiki に関する検索質問である「マシ・オカ wiki」、定額給付金の所得制限に関する検索質問である「定額給付金 所得制限」、カレーの献立に関する検索質問である「カレー 献立」、そして、卵と豆腐とキャベツに関する検索質問である「卵 豆腐 キャベツ」を用いた。

本実験において用いる検索エンジンの総文書数を表 1 に、各検索エンジンの最大および最小検索結果数を表 2 に示す。また、本実験における検索質問に対して各検索エンジンが出力した検索結果上位 30 件について正解集合を作成した。本実験の正解集合は、3 人の人手により作成した。検索質問に対する正解の基準として、3 人の被験者のうち 3 人が検索質問に対して適切であると判断していることを正解の基準とした。

★ お豆腐とキャベツのお好み焼き風 by なお★nao 🍴
お好み焼きのような豆腐ハンバーグのような...ちょっとヘルシーなメニューだけど食べ応え...
材料:木綿豆腐、キャベツ、玉子、顆粒昆布出汁、塩コショウ、薄力粉、@刻みネギ、@ポン酢
公開:08/01/09 更新:08/09/05 レシビID:485609

★ 豆腐とキャベツでダイエットお好み焼き♡ by ハチ子 🍴
キャベツ、豆腐、卵でフワリしたお好み焼きが完成(๑_๑)大阪人の私も満足の間!...
材料:豆腐(木綿)、キャベツ、卵、塩
公開:07/03/08 更新:07/03/24 レシビID:341080

★ 栄養満点♪豆腐チャンプルー☆ by ミルチャンママ 🍴
沖縄料理の定番!豆腐チャンプルーです☆ちぎった豆腐とたっぷり野菜、仕上げに卵でとじて〜...
材料:豚肉、木綿豆腐(魚豆腐)、キャベツ、玉ねぎ、にんじん、もやし、顆粒だし、醤油、ごま油、醤油、卵
公開:08/09/20 更新:09/02/16 レシビID:649169

● 博多名物●スープがうま〜い☆もつ鍋 by ちさぶー 🍴
あっさりしたしょう油味がけど、にんにくとニラの風味でとってもこくのある、美味しいスープ...
材料:もつ(生でもゆでもつでも)、●水、●だし、●鶏ガラスープの素、●しょう油、●酒、●みりん、赤唐辛子、にんにく、キャベツ、ニラ、もやし、豆腐、まいたけ(他のきのこでも)、ご飯・卵・香辛料
公開:07/01/29 更新:08/01/09 レシビID:323732

お豆腐 de お好み焼き by Lantana 🍴
お豆腐のかりずでほわんほわん。たっぷり食べても、ヘルシーだから安心♪...
材料:木綿豆腐、キャベツ、たまご、小麦粉、だしの素、しょうゆ、ひじき、マヨネーズ、お好み焼きソース、植物油、オリーブ油
公開:07/09/10 更新:08/06/02 レシビID:427481

居酒屋再現メニュー♪手ごねつくね風〜改良 by つしぐみ 🍴
焼き鳥屋さんで食べた「手ごねつくね」、美味しかった〜♪家族4人でお気に入り。その美味し...
材料:高野豆腐(※700g)、牛乳、合いびき肉、長ネギ(白い部分)、しいたけ、卵、酒、塩、ごしょう、☆しょうゆ、☆砂糖、☆みりん、☆酒、☆付け合せ用>卵黄、☆付け合せ用>キャベツ
公開:06/02/15 更新:06/03/27 レシビID:239038

キャベツと豆腐でヘルシーお好み焼き by たんくす 🍴
見た目も味もまるでお好み焼き(๑_๑)超ふわわりヘルシーです、ヘルシーにいきた...
材料:キャベツ、豆腐(木綿・絹ごしでも可)、卵、☆お好み焼きソース、☆マヨネーズ、☆香のり、☆かつおぶし、☆削り粉
公開:08/08/25 更新:08/09/05 レシビID:632783

簡単!月末どんぶり☆天キャベ豆腐卵とじ丼 by ジョシロクッカー 🍴
冷蔵庫の残り物をかき集めて(笑)甘辛でどんどんいけちゃう丼ですよ〜♪お財布にも体...
材料:ごま、絹ごし豆腐、キャベツ、天カス、卵、ごま油、好みでもみのりなど、水、本だし、しょうゆ、みりん、砂糖
公開:08/11/05 更新:08/11/05 レシビID:676800

卵豆腐とキャベツの簡単スープ☆ by kurukurumin 🍴
すばやく出来て、美味しいキャベツスープ☆ニンニクが効いてて元気になれるスープです...
材料:キャベツ、卵豆腐、コンソメ、水、ニンニク、生姜、塩・胡椒
公開:07/09/27 更新:08/12/09 レシビID:435707

コッペリ旨ウマ♪肉味噌丼 by とーみママ 🍴
簡単でとってもおいしい肉味噌丼の肉味噌が余ったら、豆腐にのせたり、レタスに包んでも♪...
材料:合いびき肉(又は豚挽き肉)、玉ねぎ、生姜、ごま、☆酒、☆砂糖、☆しょうゆ、☆みりん、☆味噌、レタス(又はキャベツなど)、ごま
公開:08/06/12 更新:09/01/28 レシビID:587540

図 4 検索質問「卵 豆腐 キャベツ」に対する COOKPAD の検索結果上位 10 件

4.2 実験結果

まず、提案手法の(5)式におけるパラメータ α の最適値検証結果を表 4 に示す。表 4 から、パラメータ $\alpha=0$ すなわち検索質問に対する検索結果数の期待値の差を考慮しない場合に平均精度の低下がみられることがわかった。しかし、本実験のデータではパラメータ α による平均精度に大きな変化はみられなかった。そのため、本実験ではパラメータ α は考慮しないものとして、以降の実験を行った。

次に、提案手法によって選択された検索エンジン、既存手法によって選択された検索エンジンおよび検索質問に対して最も高い平均精度を示した検索エンジンによる再現率精度曲線について、図 3 に示す。ここで、提案手法および既存手法の各曲線において 11 点平均精度を算出すると、それぞれ 0.87, 0.79 といった値を得られた。これらの値により、提案手法では、既存手法の 11 点平均精度と比較して 0.08 の向上を得ることができ、提案手法は既存手法と比較して有用性が高いといえる。同様に、

表 3 パラメータ α の最適値の検証結果

検索質問	11 点平均精度											
	$\alpha=$ なし	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
豚肉 パスタ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
vista LAN 設定	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
マシ・オカ wiki	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82
定額給付金 所得制限	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63
カレー 献立	0.92	0.87	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
卵 豆腐 キャベツ	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
平均精度	0.87	0.86	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87

表 4 各手法によって選択された検索エンジン

検索質問	選択された検索エンジン		
	提案手法	既存手法	平均精度最大
豚肉 パスタ	COOKPAD	COOKPAD	Google, COOKPAD
vista LAN 設定	Google	Google	Google
マシ・オカ wiki	Google	Google	Wikipedia
定額給付金 所得制限	Google	Google	Wikipedia
カレー 献立	Google	COOKPAD	Google
卵 豆腐 キャベツ	COOKPAD	Google	COOKPAD

提案手法および最も高い平均精度の各曲線において 11 点平均精度を算出すると、それぞれ 0.87, 0.96 といった値を得られた。これらの値により、提案手法では、最も高い平均精度を示した検索エンジンの 11 点平均精度と比較して 0.09 低下することがわかった。このことより、検索エンジンが出力する検索結果数が少ない場合であっても、利用者の検索質問に対して最適な検索結果を出力する可能性があることが考えられる。

最後に、検索質問「卵 豆腐 キャベツ」に対する COOKPAD の検索結果上位 10 件を図 4 に示す。図 4 より、卵、豆腐およびキャベツを主材料とする献立が上位検索結果に出力されるという結果が得られた。また、提案手法によって選択された COOKPAD および既存手法によって選択された Google の検索結果における 11 点平均精度を算出すると、それぞれ 0.98, 0.57 といった値を得られた。これらの値により、提案手法では、既存手法の 11 点平均精度と比較して 0.41 の向上を得ることができ、提案手法では検索対象文書の数少なく精度の高い検索エンジンの選択が可能といえる。

5. おわりに

本研究では、利用者が適切な検索結果を得るために、利用者の検索意図に応じて動的に検索エンジンを選択する手法の提案を行った。検索エンジンを選択するために、検索質問に対して検索エンジンが出力する検索結果数に着目した。具体的には検索質問に対する検索結果数における二つの差に着目し、各検索エンジンの検索質問に対する適合度を算出した。一つは、検索質問に対する検索結果数の期待値の差である。この差を用いて、対象検索エンジンにおける検索結果数の期待値の最大値と各検索エンジンの検索結果数の期待値との割合を算出した。もう一つの差は、検索エンジンが他の検索質問に対して出力する検索結果数の期待値との相対的な差である。この差を用いて、各検索エンジンにおける検索結果数の期待値の最大値と各検索エンジン

の検索結果数の期待値との割合を算出した。算出した二つの割合を用いて、検索質問に対する各検索エンジンの適合度を算出した。システムが算出した適合度を利用することによって、適切な検索エンジンを選択することが可能となり、適切な検索結果を得ることが可能となることを評価実験によって示した。また、提案手法では、既存手法では困難であった検索エンジンの特徴把握に対して有用な手法であることを評価実験によって示した。

今後の課題として、二つの事項について検討する。一つは、検索エンジンの適合度算出に最適なパラメータの与えることである。本実験では、問合せに対する検索エンジンの適合度算出のために特別なパラメータを与えなかった。しかし、検索質問のデータ数が膨大であった場合には、パラメータの変化によって平均精度に変化が生じることが考えられる。今後は、検索エンジンの適合度のさらなる精度向上のために、二つの割合に対してパラメータを与え、最適なパラメータについて検討を行うことによって改善を図ることを考える。

もう一つは、複数のキーワードによって構成された問合せに対する適合度の精度向上を検討する。現段階では、問合せに対する各検索エンジンの検索結果数によって適合度算出を行うのみで、問合せを構成するキーワードのうち、どのキーワードが利用者の検索意図にあたるのかを考慮できていない。この課題に対して、今後は複数のキーワードによって構成された問合せにおける各キーワードの関係性に着目し、どのキーワードが他のキーワードと比較して利用者の検索意図を表現しているのかを考慮すること [5] を行う。

文 献

- [1] L. Gravano, H. Garcia-Molina and A. Tomasic: "The effectiveness of gloss for the text database discovery problem", ACM SIGMOD Record, **23**, 2, pp. 126-137 (1994).
- [2] L. Gravano, H. Garcia-Molina and A. Tomasic: "Gloss: text-source discovery over the internet", ACM Transactions

on Database Systems (TODS), **24**, 2, pp. 229–264 (1999).

- [3] J. LIU, H. CHEN, K. FURUSE and N. OHBO: “Supporting web search by using zoomable interface to present relativity”, 電子情報通信学会第 19 回データ工学ワークショップ (DEWS2008) (2008).
- [4] 山田剛一, 森辰則, 中川裕志: “複合語マッチングと共起情報を併用する情報検索”, 情報処理学会論文誌, **39**, 8, pp. 2431–2439 (1998).
- [5] T. Takaki and T. Kitani: “Relevance ranking of documents using query word co-occurrence”, Transaction of Information Processing Society of Japan, **24**, SIG8(TOD4), pp. 74–84 (1999).