

Web ページに対するユーザの適・不適の評価を用いた 検索結果のリランキング

松岡 研二[†] 范 薇[‡] 小柳 佑介[‡] 渡邊 豊英[‡]

[†] 名古屋大学工学部電気電子・情報工学科 〒464-8601 名古屋市千種区不老町

[‡] 名古屋大学大学院情報科学研究科社会システム情報学専攻 〒464-8601 名古屋市千種区不老町

E-mail: [†] matsuoka@watanabe.ss.is.nagoya-u.ac.jp, [‡] {fan, koyanagi, watanabe}@watanabe.ss.is.nagoya-u.ac.jp

あらまし 検索エンジンを利用する際、膨大な Web ページから目的のページのみを抽出できるクエリ（検索語）を作成することは難しく、ユーザの意図に沿わないページを含むことも多い。この問題に対し、検索結果へのフィードバックによりユーザの意図を反映し、結果を改善する手法について多くの試みがなされてきた。しかし、従来の研究は検索過程で行われない判断・操作を要するフィードバックを用いており、ユーザに余分な負担を課している。我々は結果閲覧時にユーザが無意識に行うページの適・不適を抽出し、結果に反映させることが最も容易で、負荷が少ないと考えて、リランキングする手法を提案する。また、この評価から目的ページに含むべき語も推定できると考えて、その推定語を用いてクエリを修正する手法も検討する。

キーワード 情報検索, リランキング, 適合フィードバック

Dynamic Reranking of Web Search Results Based on User's Feedback on Result Pages

Kenji MATSUOKA[†] Wei FAN[‡] Yusuke KOYANAGI[‡] and Toyohide WATANABE[‡]

[†] Department of Information Engineering, Nagoya University

[‡] Graduate School of Information Science, Nagoya University

E-mail: [†] matsuoka@watanabe.ss.is.nagoya-u.ac.jp, [‡] {fan, koyanagi, watanabe}@watanabe.ss.is.nagoya-u.ac.jp

Abstract When we use Web search engines, we often face a situation that there are many undesirable pages in search results. This is because users take a lot of trouble with creating appropriate query. There have been many studies that tried to improve search results by using relevance feedback to the results. However, the studies employ feedbacks that users do not usually do in search, and do not consider burden on users well. We think that determination whether pages are appropriate or not is less burden on users as feedback, and propose a method of reranking search results by the determinations of users. We consider the possibility of query expansion, using significant words which are extrapolated from feedbacks.

Keyword Information retrieval, reranking, relevance feedback

1. はじめに

検索エンジンを用いて情報検索を行う際、目的のページを特定できる語を知っていれば、その語をクエリ（検索語）として検索をすることで、容易に目的のページを取り出すことができる。しかし、目的のページを特定できる語を知らない場合、目的のページを取り出すことは困難である。なぜならば、クエリ中の語を含む意図の異なるページが存在するためである。不要なページを含む結果から目的のページを探さなければならず、ユーザの大きな負担となっている。また、見つからない場合再検索をする必要があるが、閲覧結果を基にクエリを正しく修正することは難しく、閲覧と

クエリ修正を繰り返さなければ目的のページを見つけれない場合もある。我々は検索過程で行われるページの適・不適の評価に着目した。ページの評価から目的のページの特徴、不要なページの特徴を推測することができれば、結果から目的のページを取り出すことも可能であり、また再検索時のクエリ修正を支援することができる。

本稿の2章では関連研究について、3章では適・不適の評価を用いたリランキング手法について、4章ではクエリの修正手法について、5章ではまとめと今後の課題についてそれぞれ述べる。

2. 関連研究

検索結果の精度を高める研究は数多く存在するが、検索結果に対するフィードバックの有無により大きく二つに分けることができる。フィードバックを用いない研究では、ユーザが入力したクエリから検索意図を読み取る研究[1]や、履歴などからユーザのプロファイリングを行う研究[2]などが挙げられる。これらの研究は、ユーザに負担をかけないという利点がある一方、検索意図を読み取ることは容易ではないという欠点がある。一方、フィードバックを用いた研究では、キーワードの取捨選択やページの分類、ページの並替えなどを用いた手法がある。検索時の意図を汲み取ることが可能であり、結果に反映させることができる利点を持つ。本研究ではフィードバックなしに検索意図を推定することは困難であり、更なる精度改善にはユーザとシステムのインタラクションが不可欠であると考え、適合フィードバックを検討した。以下、本研究に関連する研究を示す。

山本ら[3]は検索結果中（タイトル・スニペット）のキーワードに対し追加・削除という操作を行うことで検索意図を伝え、検索結果をリランキングする手法を提案している。また頻出するキーワードをタグクラウドと呼ばれる語の集合としてユーザに提示することで、効率的なリランキングを可能としている。

軽部ら[4]はページに含まれる目的ページを表している文書片を選択することで、検索結果をリランキングする手法を提案している。キーワード単位では情報が少なく、ページ単位では雑多な情報が多く含まれるため、部分文章を選択している。

上記の既存研究と本研究との大きな違いは目的のページの特徴をユーザが推測するか、システムが推測するかにある。キーワードを選択する既存研究では、ノイズなく意図を伝えることができるが、目的のページに含まれるキーワードを推測しなければならない。また、複雑な意図を伝えることは難しいという欠点がある。部分文章を用いた方法では、ノイズなく柔軟に意図を伝えられる一方、目的のページを表しうる部分文章を探す必要がある。本手法では、目的のページの特徴と不要なページの特徴をページの評価を基にシステムが推測するため、検索過程においてユーザが判断する負担は既存手法と比べて少ないといえる。しかし、ページという様々な語を含み意図が明瞭でなくなる場合もあり、意図を推測するのが難しいといった欠点も

ある。本研究では評価されたページの特徴を比較することで意図を推測する。

3. Web ページに対するユーザの適・不適の評価を用いたリランキング

前章で述べたようにページには様々な語が含まれているため、一つのページから意図を推測することは難しい。本研究では、適・不適と評価されたページの特徴を比較することにより、目的のページを表す適基準、不適基準を推定し、結果をリランキングする手法を提案する。本手法で重要と考えられる事項は次の3点である。一つ目は重要語の抽出方法である。ページの特徴を表さない語をリランキングの評価に含めることは精度の低下に繋がるため、特徴を表す語を適切に選択しなければならない。二つ目はページの特徴をどのように決めるかである。ある語がページAに多く含まれていても、A以外のページでその語が多く含まれていれば、その語はページAの特徴を表しているとはいえない。ページの特徴は他のページとの差異とする必要がある。三つ目はリランキングの適・不適基準の推測手法である。ノイズの多いページの評価数回からユーザの意図を推測しなければならない。それぞれについて次節以降で述べる。

3.1 重要語の抽出

ページの特徴を表すのにページに含まれる語の特徴量を用いる方法は主流であり、TF・IDF が代表的である。TF・IDF は出現回数が多く、かつ特定のページに現れる語に高い重みをつける指標であり、多くの研究で採用されている。しかし、ある程度様々なページに含まれている語を選ばなければ、その語の特徴によってリランキングすることはできないため、出現するページが少なければ良いという IDF は本手法には不適切な基準である。我々は松尾ら[4]の語の共起の偏りから重要語を抽出する手法に着目した。語の共起に対する影響の大きい語は重要語である可能性が高いという仮定に基づき、語の重要度を測る手法であり、頻出回数の比較的少ない語も重要語として取り出せている。本手法では、結果内の上位ページの本文を一つの文章としてみなし、この手法を適用することで検索結果内での重要語を抽出する。式(1)が重要度を表す指標の導出式である。

$$X^2(w) = \sum_{g \in G} \frac{(\text{freq}(w, g) - n_w p_g)^2}{n_w p_g} \quad (1)$$

ここで、左辺 $X^2(w)$ はある語 w の X^2 値、 g は頻出語群 G の語 g 、 $\text{freq}(w, g)$ は語 w と語 g の行内での共起回数、 n_w は語 w が出現する文の語数の総和、 p_g は語 g が出現する語数の総和を全体の語数で割った値をそれぞれ表している。本手法ではこの値の高い語をページの特徴を表す語として用いる。

3.2 ページの特徴ベクトル

ある語がページ A に多く含まれていても、A 以外のほぼ全てのページがその語を同程度多く含んでいれば、その語が多く含まれていることから得られる情報は少なく、逆にその語があまり含まれていないことから得られる情報量は多い。本手法では重要語が他のページと比べて多いか、少ないかという語の量の偏差をページの特徴量として用いる。以下にページの特徴ベクトルを導出する手順を示す。

- (1) ページに含まれる重要語のそれぞれの TF・IDF 値を要素としたページベクトルを作成する。
- (2) 各ページベクトルをベクトルの大きさ 1 で正規化し、全ページベクトルの平均ベクトルを求める。正規化により、語の量ではなく、どのような語でページが構成されているかを特徴とすることが可能になる。
- (3) 各ページベクトルと平均ベクトルの差分ベクトルをそれぞれ求める。
- (4) 各ページに対応した差分ベクトルの各要素値を要素ごとの標準偏差で割ったベクトルをページベクトルとする。

3.3 リランキングアルゴリズム

本節では適と評価されたページの特徴と不適と評価されたページの特徴の差異から目的のページの基準、不適の評価から不要なページの基準をそれぞれ求め、リランキングを行う手法について述べる。本手法では、目的のページの基準と不要のページの基準を二つに分けている。これは不適と評価されたページの特徴の全てが不要なページの特徴とならない場合があるからである。例えば、適と評価されたページ A と不適と評価されたページ B があった場合、ページ A とページ B の類似度が低ければページ A の特徴は目的のページを表

しており、ページ B の特徴は不要ページを表している。しかし、ページ A とページ B の類似度が高い場合は、ページ B の特徴のほとんどは目的のページの特徴を表している。この場合、不要なページの特徴はページ B の特徴の中でページ A の特徴との差異が大きいものである。ゆえに適・不適のページ間の類似度の大きさにより、基準の求め方を変えるリランキングアルゴリズムを提案する。以下にその手順を示す。

- (1) 適ページベクトル群の総和、不適ページベクトル群の総和をそれぞれ求める。
- (2) 適の総和ベクトルとある不適ベクトルとのコサイン類似度を求め、類似度を超える要素の値を適基準のベクトルに加える。不適の場合も同様に行い、不適基準のベクトルに加える。これにより類似度が高い場合は差異の大きな特徴しか反映されなくなる。
- (3) 要素ごとに適ページベクトルと各ページベクトルと適基準のベクトルの値を掛け合わせ、その総和を求める。不適の場合も同様に計算する。前者の値から後者の値を引いたものをページごとのスコアとする。
- (4) スコア降順にリランキングし、結果を表示する。

3.4 評価実験

本手法の有効性を検証するため評価実験を行った。検索結果のページ数を 200 件とし、ページの評価はリランキング結果内の未評価のページ上位 5 件に適ページが含まれていれば、そのページに適と評価し、不適ページしかない場合は 1 位のページに不適と評価した。

実験に用いたクエリと適基準を表 1 に示す。本実験ではタスクを 2 種類用意した。一つ目は表 2 に示したように、結果に複数の意図のページが含まれており、語の特徴から意図別に分類可能なタスクである。二つ目は表 2 のように検索結果のページが類似していて分類が困難であるタスクである。それぞれに対して実験した。クエリ「東西線」の場合は、右に示す東京、仙台、札幌、京都のそれぞれに対して実験している。

評価の基準は、評価回数ごとの適合率と再現率とした。適合率とは上位 $N(1, 3, 5, 7, 10, 20)$ 件中の適ページの割合であり、再現率は結果 200 件中に含まれる適ページの総数の順位までに含まれる適ページの割合である。

図 1 の再現率では結果が分類可能な場合、5 回の評価で約 7 割の再現率が達成され、初期結果と比べると再現率を大幅に改善できている。しかし、分類困難な場合、5 回の評価でも再現率が約 4 割程度であり、目的ページを抽出できていない。その主な原因は、適と

評価する基準が二つ以上あったためである．本手法では適基準を一つとしているため，適と評価をされたページ同士の類似度が低い場合，共通する特徴が少なく，上手く目的ページの特徴を取り出せなかったためと考えられる．図2，図3の上位N件の適合率についても同様の傾向が見られ，改善する必要がある．

表 1：クエリと適基準（分類可能）

クエリ	適基準
東西線	東京，仙台，札幌，京都
ライオン	生活用品，動物
ジャガー	ミシン，動物，車，漫画
アマゾン	川，通販

表 2：クエリと適基準（分類困難）

クエリ	適基準
微積分	公式，教科書
世界遺産＋旅行	ツアー，旅行記
C 言語＋入門	初心者用解説

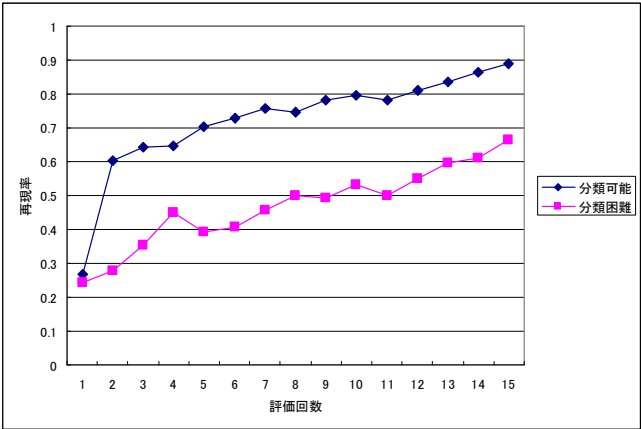


図 1：評価回数ごとの再現率

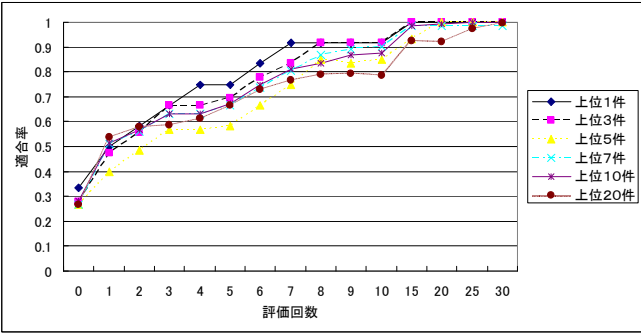


図 2：上位N件の適合率（分類可能）

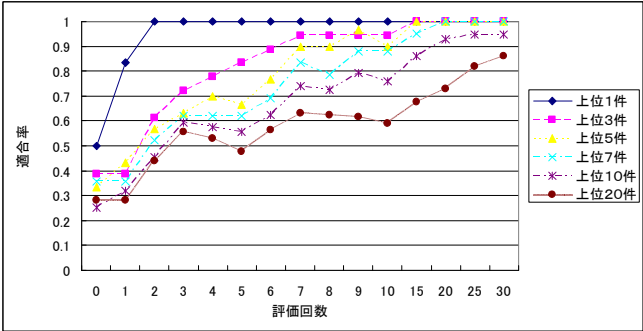


図 3：上位N件の適合率（分類困難）

4. クエリの修正

前章で述べたように，本手法では目的のページを表現しうる特徴的な語をシステムで推測している．本章では，これを用いたクエリ修正の可能性について検討する．

予備実験として，クエリ「世界遺産 旅行」の検索結果に対し，旅行記のページであるかどうかをそれぞれ2回ずつ評価し，得られた特徴語をクエリに加えて再検索した．表3はクエリに追加した語と，修正したクエリにより得られた初期結果上位200件に含まれていた適合ページ数である．

表 3：クエリ修正による再検索結果

順位	語	結果内の適ページ数
1	トラックバック	123
2	生活	79
3	バック	73
4	クリック	52
5	日記	38
	世界遺産＋旅行	31

世界遺産と旅行のみのクエリに比べると，推測した語を加えたクエリの精度が高いことが読み取れる．特にトラックバックを追加語とした場合は著しく良い結果となった．その理由として，トラックバックがブログにのみよく含まれる語であったことが考えられる．ブログは個人利用が多く，商用ではあまり利用されないため，目的ページを上手く抽出することができた．逆に，日記を追加した場合は検索結果を絞り込むことができなかった．本実験の結果は提案手法を用いたクエリの修正の可能性を明らかにしている．

5. まとめ

本稿では，ユーザが検索閲覧時に行っているページの適・不適の評価を用いて，検索結果をリランキン

推測した目的ページに含まれる語を用いて、クエリを修正する手法を検討した。

今後の課題として、初期検索結果の上位に、似たページばかりが含まれる場合や特徴の少ないページが含まれる場合に、ユーザの操作回数が増える問題がある。初期検索結果を予めクラスタリングし、各クラスタの代表ページを初期検索結果の上位として提示する手法が考えられる。

謝 辞

本研究を行うにあたり、小尻智子博士を初め、Paul Villavicencio 氏、木下 泰宏氏など、渡邊研究室の皆様
に感謝致します。

文 献

- [1] 田 馳，手塚太郎，小山聡，田島敬史，田中克己：“質問キーワードの近接性と密度分布に基づくウェブ検索の改善手法”，日本データベース学会論文誌，Vol.5，No.1，pp.113-116，2006.
- [2] 土方嘉徳：“情報推薦・情報フィルタリングのためのユーザプロファイリング技術”，人工知能学会誌，Vol.19，No.3，pp.365-372，2004.
- [3] 山本岳洋，中村聡史，田中克：“Rerank-by-example: 編集操作の意図伝播によるウェブの検索結果のリランキング”，情報処理学会論文誌(データベース)，Vol.49，No.SIG7(TOD 37)，pp.16-28，2008.
- [4] 軽部孝典，志築文太郎，田中二郎：“検索結果のインタラクティブな評価に基づくリランキングインタフェース”，第 70 回全国大会講演論文集，No.1，pp.489-490，2008.
- [5] 松尾豊，石塚満：“語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム”，人工知能学会誌，Vol.17，No.3，pp.364-371，2002.