

モバイル協調検索のためのクエリ 推薦・提示手法

Query Suggestion and Presentation for Mobile Collaborative Search

大重 智志・中村 聡史・田中 克己・

Satoshi OHSHIGE Satoshi NAKAMURA
Katsumi TANAKA

本稿では、モバイルを使用した複数人での協調検索を効率的に行うための手法を提案する。モバイル協調検索では結論を求めるために意思決定を行う必要があるが、スムーズに進まないことが多い。その理由として、複数ユーザが重複した検索行動を行うため網羅的に検索できていないという点と、ユーザが各自バラバラに検索を行うことでの的を絞れていないという点の2つの問題がある。この2つは、探索範囲が狭くなってしまっていることと、探索範囲が広がってしまっていることとで、トレードオフの関係にある。そこで本稿では、グループ全体として検索の話題がどれほど発散しているか、もしくはどれほど収束しているかということ推測し、状況に応じて話題の発散・収束を促すクエリを生成して推薦する手法を提案する。また、提案した手法を実装し、有効性を実験により明らかにする。

In this paper, we propose a method for supporting efficient collaborative search using smartphone. In mobile collaborative search, group members have to make decision to reach satisfactory conclusion, but they tend not to do smoothly. We think that there are two points as the cause. The first point is that group members tend to do similar searches and they cannot find various pages. The second point is that group members tend to search freely according to his or her own inclination and they cannot focus on common purpose as the group. There is trade-off relationship between these two points, that is narrow search range and wide one. For this reason, in this paper, we estimate how divergent topic is and how convergent a topic is. Then, queries for promoting divergence and convergence of a topic are generated and suggested. Additionally, we implement a prototype system and confirm usefulness of our method through experiment.

・ 京都大学大学院情報学研究科 修士課程
ohshige@dl.kuis.kyoto-u.ac.jp

・ 正会員 京都大学大学院情報学研究科
{nakamura, tanaka}@dl.kuis.kyoto-u.ac.jp

1. はじめに

爆発的な人気を博すスマートフォンによって、人々は時と場所を選ばずにWeb検索を行うことができるようになった。総務省によると2011年9月時点での携帯電話の人口普及率は96.2%に達し[1]、携帯電話所有者のうち8.9%がスマートフォン所有者、11.5%が所有意向者である。特に、20代のスマートフォン所有者は23.1%、所有意向者は16.2%にもなる[2]。

Morris[3]は、複数のユーザが協調してWeb上で情報検索を行うという行為が多く行われていることを明らかにしている。例えば、グループ旅行の計画のために必要な情報を皆で検索したり、ネットショッピングで様々な候補の中から最善のものを選択し購入したりするなどの状況が挙げられる。本稿では、このような複数ユーザでの検索行動を協調検索と呼ぶ。スマートフォンの普及により、時と場所を選ばずにWeb検索を行うことができるようになったため、PCを利用した協調検索だけではなく、小型のモバイル端末を用いて状況に応じた協調検索(モバイル協調検索)を行う機会も増えてきた。例えば、グループでの旅行中にその周辺の観光スポットを探すような状況、パーティーの後で急遽行くこととなった2次会の店を探す状況、災害直後に被災したメンバーと帰路を探す状況などがある。一般的なPCを用いた協調検索とモバイル協調検索の大きな違いは、後者は突発的に検索を開始することが多く、前者に比べて情報検索に十分な時間を掛けることができない点である(理由については3章で述べる)。また、後者はディスプレイサイズが小さいため表示および共有できる情報量にも制限がある。

モバイル協調検索では、グループとして満足する結論に達する必要があり、意思決定が行われる。しかし、モバイル協調検索では意思決定までスムーズに進めることは容易ではない。この原因として、我々は大きく分けて2つあると考えている。1つ目は、グループの複数ユーザが重複およびとても類似した検索行動を行ってしまう点である。グループとしての検索行動の方針が決まっていない検索開始初期などにおいて、似たようなクエリで検索すると、グループ全体での探索範囲が狭まってしまう。そのために、網羅的な探索を行うことができず、グループ全体として質の高い良いページに辿り着けない場合がある。結論として、満足のいく結論に達しないこともある。2つ目は、各ユーザが個人の好みに合わせすぎて自由に探索を行ってしまうという点である。各ユーザの好みだけで行動すると、いつまで経っても探索範囲がバラバラでグループとしての的を絞れなくなってしまふ。結果として、最終的なゴールに辿り着けず目的を果たせなくなってしまううえ、時間の浪費となってしまう。また、この両者は探索範囲が狭くなっていることと探索範囲が広がっていることとでトレードオフの関係にある。

そこで本稿では、上記の問題を解決し、モバイル協調検索における円滑な意思決定を支援する手法を提案する。ここでは、色々な候補を検討しつつ意思決定までスムーズに進めるため、グループ全体の話題の発散と収束をシステムによりコントロールすることを考える。ここでの「話題」とは、ある時点におけるグループ全体の検索目的に関するトピックやテーマのことである。本手法では、話題を発散させることで探索範囲を広げてグループ内での重複した検索行動を防ぎ、話題を収束させることでグループ内での探索範囲を狭めて的を絞った情報検索を行えるようにする。具体的には、話題

の発散を促すクエリと話題の収束を促すクエリを推薦することで、グループ全体としての話題の発散・収束をコントロールし、意思決定における問題の解決を図る。ここでは、発散を促すクエリを「発散クエリ」、収束を促すクエリを「収束クエリ」と呼ぶ。また、システムが各ユーザの入力したクエリからグループ全体としてのおおまかな意図を推定し、その推定された意図に基づいて発散および収束クエリを生成・推薦する。なおここで、発散クエリとは、ある入力されたクエリとは関連があるが別の観点によるものことである。また、収束クエリとは、ある入力されたクエリと関連があり、さらに詳細なものを指すものとする。

2. 関連研究

SearchTogether[4]は、遠隔地にいるユーザがPCを用いて同期的・非同期的に協調検索を行うためのシステムである。各ユーザがそれぞれのPCで入力したクエリが他の協調ユーザと共有されたり、気に入ったページを推薦したり、ページに対して評価やコメントなどを付加できるなどの機能がある。ユーザはこうした機能を利用することで、効率的に協調検索を行える。

CoSearch[5]では、複数のユーザが1台のPCと複数のマウス・モバイル端末を用いて協調検索を行う。各ユーザにはカーソルが1つずつ割り当てられ、ディスプレイ上に色で区別されたカーソルが表示されている。各ユーザは自身のカーソルをマウスで操作し、クエリや閲覧ページをキューに貯めて協調検索を行う。モバイル端末では、マウスの機能に加え、クエリ入力やページの閲覧、ページに対するメモなどが行える。これにより、1台のPC上で効率的に協調検索を行える。

伊豆らの研究[6],[7]では、ユーザの探索に同期してシステムが有用なWebページをグループの共有データ内から選び提示することで、探索を支援する。検索結果に至るプロセスやその行為を含めて共有することで、ユーザの探索目的や探索範囲を把握する。ユーザの行動に沿ったページを提示するために、ユーザの検索行動を、いろいろなページを見ている発散フェーズ、類似ページを続けて見ている収束フェーズ、ユーザが考える範囲を調べ終えた補完フェーズに分け、ユーザがどのフェーズにいるかを判定し、それに応じて何を提示すべきかを判断している。それらのフェーズは、探索目的に沿ったページをどれ程多く閲覧したかを表す指標である達成度と、どの程度他の類似Webページと吟味されて推薦ページは選ばれたのかを表す指標である精選度の2つの尺度を用いて決定している。

これらの研究は1台以上のPCが使えることを想定している。本稿は、モバイル端末のみが使えるPCは使えない状況を想定しているという点でこれらの研究と異なる。また、伊豆らの研究は既に関連されたページを提示するのにに対し、本稿ではクエリを推薦するという点で異なる。さらに、本稿でも発散・収束という語を用いているが、発散・収束がどの程度の強さであるかという考え方をしている我々とは異なる。

奥らの研究[8]では、ユーザの意見を反映したウェブコンテンツを共有し、それらをユーザの評価で整理および提示することで、ウェブコンテンツの比較・検討を支援するインタフェースを提案している。

小谷らの研究[9]では、気に入ったウェブページを他のユーザの検索結果一覧に割り込ませることにより通知する機

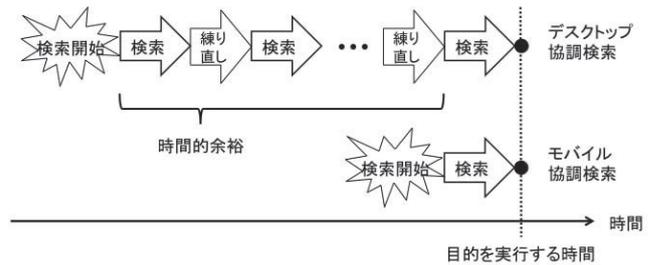


図1 協調検索による時間的余裕の違い

Fig. 1 Difference between temporal spare of collaborative search

能や、各ユーザの探索空間を共有する機能により、ユーザ間の情報共有を支援している。

これらの研究は我々の研究同様モバイル端末を用いた状況を想定しており、情報の比較・共有に注目している。本稿では比較や共有という観点ではなく、話題の発散や収束を促し、意思決定をスムーズに進めるよう支援するものである。

3. 提案手法

3.1 モバイル協調検索の問題点と解決案

モバイル協調検索はグループ単位で行う検索であるので最終的にグループ全体として満足する必要があり、全員が満足する意思決定のためのページを発見しなければならない。しかし、モバイル協調検索では意思決定までスムーズに進めることは容易ではない。1章で述べた通り、我々はグループの重複した検索行動と目的を絞れない検索の仕方が原因であると考えている。

一般的な協調検索でも意思決定までスムーズに進まないという問題はある。しかし、一般的な協調検索では意思決定まで時間を掛けることが可能であるが、モバイル協調検索は突発的に行うことが多く検索行動に掛けられる時間が限られているため、意思決定までスムーズに進めることが特に重要となる。図1は、デスクトップを用いた一般的な協調検索に掛けられる時間と、モバイル協調検索に掛けられる時間の差を模式的に表したものである。一般的な協調検索では、ある目的を実行する時間まで時間的な余裕があるのが普通であり、検索や繰り返しを繰り返し行うことで全員が満足する結論を得ることができる。しかし、モバイル協調検索では目的を実行する時間までに余裕がなく、意思決定までスムーズに進めることがより重要となる。

そこで我々は、話題の発散と収束という概念を導入し、話題の発散を促すクエリと話題の収束を促すクエリをユーザに推薦することで問題の解決を図る。また、協調検索のある段階において話題がどれほど発散・収束しているかという割合を推定する。ここでは、特に、どれほど話題が収束しているかに着目し、その割合を話題の収束度と呼ぶ。話題の収束度が高いほど話題は収束し、低いほど発散しているということになる。そして、話題の収束度から、話題をどれほど発散・収束すべきかという方針を考える。また、話題の収束度に基づいて、推薦する発散クエリと収束クエリの数の割合を変更する。すなわち、ある時点での話題の発散・収束具合を動的に推定し、それに応じて推薦するクエリの内容も動的に変更するというものである。

3.2 話題の収束度

グループの話題の収束度を求めるため、ユーザの閲覧したページの類似度と選択した推薦クエリの2つに注目する。

3.2.1 閲覧ページの類似度の利用

ある時点における、閲覧しているページと今までに閲覧した各ページとの類似度の重み付き平均 SimilarityScore を、話題の収束度を決定するための一要素とする。

まず、ユーザが新規ページ p を閲覧するたびにそのページに含まれている語を抽出し、その語の頻度を利用した特徴ベクトル $v_p = (v_{p,t_1}, v_{p,t_2}, \dots, v_{p,t_n})$ を求める。 v_{p,t_i} はここで、

$$v_{p,t_i} = \frac{\text{freq}(p, t_i)}{\sum_{t \in T_p} \text{freq}(p, t)}$$

であり、 $\text{freq}(p, t)$ はページ p における単語 t の出現頻度、 T_p はページ p に出現する全単語集合である。

また、最近閲覧したページと協調検索開始直後に閲覧したページとでは重要度が異なり、最近閲覧したページの方がグループの話題の収束具合を表していると考えられる。そこで、あるページ p の重み w を、検索行動開始からページ p を閲覧し始めるまでに経過した時間 s_p を用いて、 $w_p = \text{normalize}(s_p)$ とする。なお、 $\text{normalize}(s_p)$ は s_p を 0 から 1 までの値に正規化するための関数であり、0 が最も古いものを表し、1 が最も新しいものを表す。

そして、あるユーザがページ p_{now} を見ているとき、その時点での SimilarityScore を、

$$\text{SimilarityScore}(p_{now}, P) = \frac{\sum_{p \in P} \text{Sim}(p_{now}, v_p) \cdot w(p)}{\sum_{p \in P} w(p) + 1} \quad (1)$$

として求める。ここで、 P はグループ全体の既閲覧ページ集合、 $\text{Sim}(v_{p_1}, v_{p_2})$ はページ p_1 と p_2 の特徴ベクトル v_{p_1} と v_{p_2} のコサイン類似度である。これにより、ある時点での、最近閲覧したページと今までに閲覧した各ページとの類似度の重み付き平均が求まる。この値が低い時、グループで類似していないページを閲覧していることになり、話題は発散の段階にある。この値が高い時、グループで類似したページを閲覧していることになり、話題は収束の段階にある。

伊豆らの研究[6],[7]でも、先述の達成度と精選度を求めるために閲覧ページの特徴ベクトルの類似度を用いている。しかし、伊豆らは達成度を検索キーワードとクリックアンカー文字列による特徴ベクトルと各閲覧ページの特徴ベクトルの類似度の総和から求めており、本稿がページ同士の類似度を用いている点で異なる。また、伊豆らは精選度は推薦ページの特徴ベクトルと各閲覧ページの特徴ベクトルの類似度の総和から求めており、本稿がより最近閲覧したページが類似しているほどより収束していると考えている点で異なる。

3.2.2 選択した推薦クエリの利用

推薦クエリのうち発散クエリが多く選択されている場合、話題が発散の段階にあり、一方、収束クエリが多く選択されている場合、話題が収束の段階にあると考えられる。

そこである時点における、各ユーザの選択した直近の推薦クエリのうちで収束クエリの割合 ConQueryScore を、話題の収束度を決定するための一要素とする。

ある時点における直近の各ユーザの入力クエリ集合 $Q = (q_{u_1}, q_{u_2}, \dots, q_{u_n})$ のうち、発散クエリとして推薦されたものの集合を $Q_{div}(Q_{div} \subseteq Q)$ 、収束クエリとして推薦されたものの集合を $Q_{con}(Q_{con} \subseteq Q)$ として、以下で求める。

$$\text{ConQueryScore} = \frac{|Q_{con}|}{|Q_{div}| + |Q_{con}|} \quad (2)$$

3.2.3 話題の収束度の計算

以上が話題の収束度を求めるための要素を求める方法であり、そのページの類似度による式(1)とクエリの種類による式(2)を用いて、話題の収束度を求める。

ある時点における話題の収束度 ConvergenceScore は、

$$\text{ConvergenceScore} = \alpha \cdot \text{SimilarityScore}(p_{latest}, P) + \beta \cdot \text{ConQueryScore}(Q) \quad (3)$$

として求める。ここで、 p_{latest} は最も最近閲覧されたページ、 P はグループ全体の既閲覧ページ集合、 Q は直近の各ユーザの入力クエリ集合、 α と β は重みである。

3.3 推薦クエリの生成

各ユーザの入力したクエリからグループ全体としてのおおまかな検索の背景を推定し、その背景に基づいて発散クエリと収束クエリを生成する。ここで背景とは、検索の文脈を表すもので、そのような背景を表した語を背景語と呼ぶ。

3.3.1 発散クエリ

発散クエリとしては、あるクエリをシードとし、そのクエリと関連はあるが別の観点であるものを生成する。以降、シードとなるクエリのことをシードクエリと呼ぶ。

例えば、“京都 観光”をシードクエリとして考える。このクエリには2つの背景が考えられ、それは「京都へ旅行するときに、観光地について検索する」という背景と「関西を観光したいので、特に京都について検索する」という背景である。前者の場合、京都で観光だけでなくショッピングのことも考えられるため、“京都 ショッピング”が発散クエリとして考えられる。後者の場合、関西での観光地は他にも大阪があるため、“大阪 観光”が発散クエリとして考えられる。

ここで、各クエリに含まれる語に関して注目すると、前者の場合「京都」、後者の場合「観光」という語が共通して現れており、それぞれが背景に対応している。また、前者の場合、「観光」と「ショッピング」という語が、後者の場合、「京都」と「大阪」という語が、それぞれ同位概念のような関係になっている。以降、このように同位概念のような関係になる語を兄弟語と呼ぶ。

そこで、あるシードクエリ“ $t_1 t_2$ ”から背景を表す語 t_1 を選び、残りの語 t_2 の兄弟語 t_2' を求め、新しい発散クエリ“ $t_1 t_2'$ ”を求める。このようにして発散クエリを生成する。

3.3.2 収束クエリ

収束クエリとしては、あるクエリをシードとし、そのクエリと関連があり、さらにそれを詳細化するものを生成する。

以下、“京都 観光”をシードクエリとして考える。“京都 観光”で「観光」に注目して詳細化する場合、「京都での観光、特に紅葉が見れるところ」という意味で“京都 観光 紅葉”が発散クエリとして考えられる。また、「京都」に注目すると、「京都、特に嵐山での観光」という意味で“京都 嵐山 観光”などが収束クエリにあたりと考えられる。ここで、クエリに含まれる語に注目すると、「紅葉」は「京都」という背景で「観光」を詳細にするもので、「嵐山」は「観光」という背景で「京都」を詳細にするものである。

また、グループ全体としての的を絞るという意味で、シードクエリ自身もそのシードクエリを入力したユーザ以外のユーザにとっては収束クエリとなりうる。

そこで、あるシードクエリ“ $t_1 t_2$ ”から背景を表す語 t_1 を選び、残りの語 t_2 をより詳細にする語 t_3 (詳細語) を求め、収束クエリ“ $t_1 t_2 t_3$ ”を求める。また、シードクエリ自身も収束クエリとする。このようにして生成する。

3.3.3 背景語の発見

クエリに多く現れる語はそれだけ背景を意味していると考えられるので、背景語は全クエリに含まれる単語のうち頻出のものとする。頻出の語はひとつとは限らないので、背景語は集合として表す。求める背景語集合を $T_{context}$ とすると、

$$T_{context} = \operatorname{argmax}_{t \in T} \operatorname{freq}(t) \quad (4)$$

で求められる。ただし、 T は全クエリ中の単語集合で、 $\operatorname{freq}(t)$ は単語 t の出現頻度である。

3.3.4 シードクエリの発見

発散・収束クエリを生成するためのシードクエリは、クエリ履歴の中から、式(4)で求めた背景語集合の要素が含まれるクエリを取得する。ただし、簡単化のために、シードクエリはある単語と背景語の2つだけからなるものに限る。このようにしてシードクエリを発見する。

ここで、取得したシードクエリ候補集合 Q_{base} の要素 q の重み $w_{base}(q)$ を求める。そのクエリ q に含まれる背景語ではない語を t 、単語 t の出現頻度を $\operatorname{freq}(t)$ 、検索行動を開始してからそのクエリを入力するまでに経過した時間を s_q とすると、

$$w_{base}(q) = \alpha \cdot \operatorname{freq}(q) + \beta \cdot \operatorname{normalize}(s_q) \quad (5)$$

で求める。ただし、 $\operatorname{normalize}(s_q)$ は式(1)で使用したのと同じで、 α と β は重みである。この式は、グループとしてよく検索されている単語を含み、その語を含むクエリが最近検索されているほど重みが大きくなるものである。検索されていないグループの意図を反映できず、より最近のクエリほど最近の意図を表していると考えられるからである。

3.3.5 兄弟語と詳細語の発見

クエリ“ $t_1 t_2$ ”から背景を表す語 t_1 を選んだ後は、残りの語 t_2 の兄弟語もしくは詳細語を見つける必要がある。兄弟語や詳細語を求める研究は数多くあるが、本稿では大島ら[10]、野田ら[11]の研究に着目した。

大島らは本稿での「兄弟語」を「同位語」と呼んでいる。ある語に対する同位語は並列助詞「や」で接続されることを利用して、Web 検索エンジンに対するクエリを作成し、その検索結果から同位語を得るというものである。また、野田らは本稿での「詳細語」を「話題語」と呼んでいる。ある語に対する話題語は連体助詞「の」で接続されることを利用して、Web 検索エンジンに対するクエリを作成し、検索結果のウェブページ群から話題語を得るというものである。

これら同位語と話題語を取得する手法が実装された API を利用し、本稿では兄弟語と詳細語を取得する。

3.4 クエリの推薦

ある状況において、発散クエリと収束クエリをそれぞれどのくらいの割合で推薦するかは、式(3)で求めたスコアを使用する。推薦できるクエリの最大数を $n_{all} (> 0)$ としたとき、ある時点の収束クエリの数 n_{con} は、

$$n_{con} = \operatorname{round}(n_{all} \cdot \operatorname{ConvergenceScore})$$

で求める。ここで、 $\operatorname{round}(n)$ は n を四捨五入した値を求め

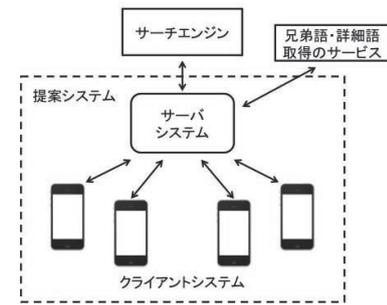


図2 システムの構成

Fig. 2 Construction of our system

る関数である。また、発散クエリの数 n_{div} は $n_{div} = n_{all} - n_{con}$ で求める。ただし、発散クエリも収束クエリも最低1つは推薦できるように、 $n_{con} = 0$ のときは n_{con} を1とし、 $n_{con} = n_{all}$ のときは n_{con} を $n_{con} - 1$ とする。

モバイル協調検索では、ディスプレイサイズの制約があり一度に推薦できるクエリの数には制限があるため、3.3節で生成したクエリのうちどのクエリを推薦するかが重要になる。そこで、生成した各クエリに重み w を与え、その値の上位から n_{div} もしくは n_{con} の分だけ推薦する。クエリ q の重み w は、式(5)で求めた w_{base} と、そのクエリを構成する兄弟語・詳細語の順位 r (前述のAPIから取得する際に付加されているメタデータ) を用いて、以下の式で求める。

$$w(q) = \begin{cases} w_{base}(q) & (\text{シードクエリ自身が収束クエリの場合}) \\ \alpha \cdot w_{base}(q) + \beta \cdot \frac{1}{r} & (\text{それ以外}) \end{cases}$$

4. システムの実装

我々は、3章の手法を基にサーバ/クライアントモデルを採用したプロトタイプシステムを設計した。サーバはPythonを用いてCGIとして動作し、クライアントはObjective-Cを用いてiOS5.0を搭載した端末上で動作する。システムの構成は図2のようになっており、サーバは検索エンジンや兄弟語・詳細語を取得するためのサーバとやりとりしている。

図3は検索を行うメイン画面である。上部にクエリを入力するサーチバーがあり、それを選択すると図4のようにキーボードが現れる。キーボードの検索ボタンを押すと、中央に検索結果がランキングで表示される。また、サーチバーと検索結果の間に、推薦クエリを表示するエリアがある。本システムでは、6つのクエリを推薦する。

5. 評価実験

提案手法の有用性を示すため、協調検索によって被験者にあるタスクを達成してもらうという実験を実施した。実験によって、発散クエリによってグループとして網羅的に検索し様々な候補を検討できているか、収束クエリによってグループとしての絞った検索ができているか、それらによって意思決定がスムーズに進んでいるか、これらを明らかにする。

5.1 比較システム

我々は、提案手法との比較のために、4章のプロトタイプとほぼ同じであるが、一部機能が異なるシステムを作成した。

提案手法によるプロトタイプシステム(以下、提案システム)と比較のためのシステム(以下、比較システム)とでは、

¹ <http://www.dl.kuis.kyoto-u.ac.jp/~ohshima/wiki/index.php?YayalaAPI>

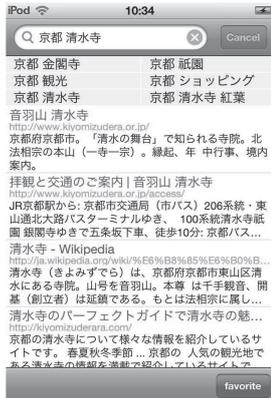


図 3 メイン画面



図 4 メイン画面のキーボード

Fig. 3 Main screen Fig. 4 Keyboard in main screen

推薦されるクエリの生成・推薦の仕方のみが異なっている。提案システムでは、推薦されるクエリを3章の手法により生成・推薦しているが、比較システムでは、Yahoo!の関連検索ワードの API²を用いている。ユーザの入力したクエリから API によって関連検索ワードを取得・推薦する。比較システムは、一般的な関連キーワードによるクエリ推薦を模したものであり、システムを通して協調することは想定していない。

5.2 実験内容

我々は、大きく分けて2つの実験を行った。1つ目は協調検索による実験あり、2つ目は全ての協調検索の実験が終わった後の客観的な評価である。

我々はまず、システムを利用しモバイル協調検索のシチュエーションを模した実験を行った。この実験では先述の通り、網羅的に検索できているか、的を絞った検索ができていないか、意思決定がスムーズに進んでいるか、これらを明らかにする。

1 グループにつき各手法で1回ずつ実験を行い、被験者は日常的に会話を行うメンバー4人を1つのグループとした、合計8グループ、32名の男女で、年齢は18~33歳、平均年齢は22.3歳である。我々が与えたタスクを、提案・比較システムでそれぞれ1タスク10分程度で行ってもらった。

そして、全実験後、推薦された全てのクエリの精度を評価した。各クエリに対して、そのクエリが推薦されるクエリとして適切であるかという尺度と、そのクエリがより広範囲を検索するのに役立つかまたは的を絞って狭い範囲を検索するのに役立つかという尺度で、それぞれ5段階で評価した。

被験者にはグループの検索意図(タスク)を見せ、各時点でユーザが入力したクエリを提示した。それらを考慮し、先述の2つの尺度で、推薦されたクエリを評価してもらった。

6. 実験結果

どのタスクについてもほぼ10分程度で結論に達した。実験の様子を撮影した映像と操作ログ、アンケート、クエリの精度の評価について分析を行った。

グループ毎の結論に達するまでに掛かった時間と発話数の平均を表1に示す。結論に達するまでに掛かった時間は提案システムの方が比較システムよりも短くなっており、10分を切っている。つまり、提案システムの方が短時間でス

表 1 掛かった時間と発話数の平均

Table 1 Average of spent time and # of conversation

	結論に達するまでの秒数	発話回数
提案システム	589.1 ($\sigma = 54.5$)	25.2 ($\sigma = 12.9$)
比較システム	604.3 ($\sigma = 37.4$)	29.8 ($\sigma = 14.7$)

表 2 クエリ数など

Table 2 # of queries and so on

	ユニーククエリ数		入力クエリのうち推薦クエリの割合
	クエリ数	クエリ数	
提案システム	18.6 ($\sigma = 6.0$)	14.0 ($\sigma = 3.2$)	31.5%
比較システム	19.6 ($\sigma = 4.2$)	16.0 ($\sigma = 3.5$)	13.4%

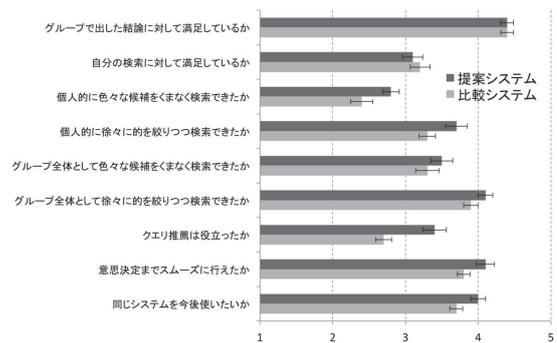


図 5 アンケート結果

Fig. 5 Results of questionnaire

ズに意思決定に至ったことがわかる。また、グループ毎の平均発話数は、提案システムの方が比較システム比べて少なくなっている。

操作ログからの入力クエリ数などを表2に示す。提案システムでは各ユーザに同じ時点では同じクエリを推薦するので、クエリ数もユニーククエリ数も比較システムの方がわずかに多くなっている。しかし、入力クエリのうち推薦クエリから選択したものの占める割合は、提案システムで31.5%、比較システムでは13.4%と、提案システムの方が圧倒的に多く、より興味を示すクエリが提示されたということがわかる。

アンケートの結果を図5に示す。「グループで出した結論に対して満足しているか」、「意思決定までスムーズに行えたか」といった質問に対しては、両システムに有意差はなかった。しかし、「クエリ推薦は役立ったか」という質問に対しては、両システムで有意な差が見られ ($p < 0.05$)、提案システムの方が良い結果となった。

クエリの精度を2つの尺度で評価した結果を表3に示す。1つ目の尺度「適合度」は、そのクエリが推薦されるクエリとして適切であるかという尺度で、不適切:1~適切:5で評価してもらった。2つ目の尺度「収束度」は、そのクエリがより広範囲を検索するのに役立つか(1)または的を絞って狭い範囲を検索するのに役立つか(5)という尺度で、5段階で評価してもらった。全推薦クエリについての平均、発散クエリとして推薦したクエリについての平均、収束クエリとして推薦したクエリについての平均を載せている。結果としては、全体の適合度はやや適合の傾向にあるが、収束クエリについてはわずかに高くなっている。収束度については、発散クエリの収束度は低く、収束クエリの収束度は高くなっており、つまり発散クエリは発散クエリとして適切で、収束クエリは収束クエリとして適切であったということがわかる。

² <http://developer.yahoo.co.jp/webapi/search/assistsearch/v1/webunitsearch.html>

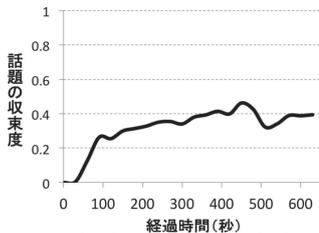


図 6 全グループの収束度の平均

Fig. 6 Average of convergence degree of all groups

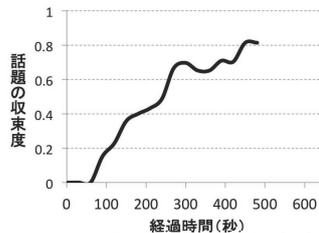


図 7 グループ 8 の収束度

Fig. 7 Convergence degree of 8th group

また、協調検索を行った際にシステムが推定した各グループの収束度の平均の推移 (図 6) と、特に変化が顕著であったグループ 8 の収束度の推移 (図 7) の 2 つを示す。どちらのグラフも右肩上がりであり、網羅的に検索し、徐々に絞っていったということがわかる。

7. まとめ

本稿では、複数ユーザがモバイルを用いて協調検索を行う際に、満足度を高めつつ意思決定をスムーズに進めるための手法を提案した。具体的には、各ユーザが重複した検索行動をとることでグループとして網羅的に検索できないという問題を解決するために話題の発散を促すようなクエリを推薦し、各ユーザが自由に検索することでグループとしての絞った検索ができないという問題を解決するために話題の収束を促すようなクエリを推薦するという手法を提案した。

また、評価実験によって、色々な候補をくまなく検索できたか、クエリ推薦は役に立ったかというような尺度では提案手法の効果は示せた。特に、提案手法によるクエリ推薦については、実際の撮影映像や操作ログからもアンケートからも有用性があるということが判明した。

今後の課題としては、グループの意図をより粒度を細かく推定できるようにしたり、兄弟語・詳細語だけでない発散・収束クエリについて検討したりするなど、生成の仕方を改善する必要がある。また、ユーザ毎に検索を分担していたり、先行知識が違ったりと、全ユーザが平等であるとは限らないので、ユーザ毎に推薦するクエリを変更するという点についても考えたいと思っている。

[謝辞]

本研究の一部は、文部科学省科学研究費補助金挑戦的萌芽研究「モバイル協調検索に関する研究」(研究代表者: 中村聡史, #22650018), 若手研究 (A) 「インタラクティブな再ランキング・再サーチを可能とする次世代検索に関する研究」(研究代表者: 中村聡史, #23680006), グローバル COE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」(拠点リーダー: 田中克己), 文科省科研費基盤 (A) 「ウェブ検索の意図検出と多角的検索意図指標にもとづく検索方式の研究」(24240013, 研究代表者: 田中克己) によるものです。ここに記して謝意を表すものとします。

[文献]

[1]総務省.情報通信統計データベース携帯電話・PHS の加入契約数の推移.

<http://www.soumu.go.jp/johotsusintokei/field/data/gt01020101.xls>.

[2]社団法人中央調査社.2011 年パーソナル先端商品の利用状況 (第 26 回) .

<http://www.crs.or.jp/data/pdf/ptg2011.pdf>.

[3]Meredith Ringel Morris. A survey of collaborative web search practices. In Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI'08, pages 1657-1660. ACM, 2008.

[4]Meredith Ringel Morris and Eric Horvitz. Searchtogether: an interface for collaborative web search. In Proceedings of the 20th annual ACM symposium on User interface software and technology, UIST'07, pages 3-12. ACM, 2007.

[5]Saleema Amershi and Meredith Ringel Morris. Cosearch: a system for co-located collaborative web search. In Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI'08, pages 1647-1656. ACM, 2008.

[6]伊豆陸, 中島伸介, 田中克己. グループ支援型 web 閲覧における閲覧履歴の視覚化と共有. 日本データベース学会 letters, 3(1):121-124, 6 2004.

[7]伊豆陸, 中島伸介, 田中克己. グループ支援型 web 探索におけるナビゲーションのための既閲覧ページ群の同期化提示. 日本データベース学会 letters, 3(2):13-16, 9 2004.

[8]奥梓, 小牧大治郎, 荒瀬由紀, 原隆浩, 上向俊晃, 服部元, 西尾章治郎. 携帯端末を用いた協調 web 検索におけるコンテンツ比較支援インタフェース. In DEIM Forum 2010 D9-4, 2010.

[9]小谷大祐, 中村聡史, 田中克己. モバイル協調検索におけるユーザ間の情報共有支援. In DEIM Forum 2011 A6-6, 2011.

[10]大島裕明, 小山聡, 田中克己. Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見. 情報処理学会論文誌. データベース, 47(19):98-112, 12 2006.

[11]野田武史, 大島裕明, 小山聡, 田島敬史, 田中克己. 主題語からの話題語自動抽出とこれに基づく web 情報検索(情報抽出). 電子情報通信学会技術研究報告. DE, データ工学, 106(149):239-244, 7 2006.

大重 智志 Satoshi OHSHIGE

京都大学大学院情報学研究科修士課程在学中.

中村 聡史 Satoshi NAKAMURA

京都大学大学院情報学研究科社会情報学専攻特定准教授, JST CREST. 2004 年大阪大学大学院博士後期課程修了. 博士(工学). 主にヒューマンコンピュータインタラクション, ウェブ検索の研究に従事. 情報処理学会, 日本データベース学会各会員.

田中 克己 Katsumi TANAKA

京都大学大学院情報学研究科社会情報学専攻教授. 1976 年京都大学大学院修士課程修了. 京大工博. 主にデータベース, Web 情報検索, Web マイニング, マルチメディアコンテンツ処理の研究に従事. IEEE Computer Society, ACM, 人工知能学会, 日本ソフトウェア科学会, 情報処理学会, 日本データベース学会等各会員.