

Web 検索時の行動情報を用いたクエリ修正タイプの予測

梅本 和俊^{1,2,a)} 中村 聡史^{3,4,b)} 山本 岳洋^{1,c)} 田中 克己^{1,d)}

受付日 2012年12月20日, 採録日 2013年4月7日

概要: 旅行に関する情報を Web 検索エンジンを通じて収集する場合, 観光地や宿泊先などさまざまな観点から検索を行う必要がある. このように, いくつかのサブタスクから構成される検索タスクでは, 1つの検索クエリのみですべての情報を得ることは難しく, クエリの修正をともなう検索が反復的に行われることによって, タスクが実行されることが多い. 本稿では, こうした検索クエリの修正を, 事前に予測する手法を提案する. 提案手法は, 現在の検索行動を特徴量とする分類器を構成することで, 次の検索におけるクエリ修正タイプの予測を行う. 実際の検索行動のログデータを用いた評価実験の結果, 提案手法は5種類の検索クエリ修正タイプを約41%の精度で分類可能なことが明らかになった. また, 分類器構成に用いる特徴量を変化させることで, 修正タイプの予想に有用な特徴量の検証も行った. さらに今後の展望として, 次の検索において実際に入力されるクエリの予想可能性についても考察する.

キーワード: クエリ修正, 検索行動, ログ分析, 視線情報, 情報検索

Predicting Query Reformulation Type from Web Searcher Behavior

KAZUTOSHI UMEMOTO^{1,2,a)} SATOSHI NAKAMURA^{3,4,b)}
TAKEHIRO YAMAMOTO^{1,c)} KATSUMI TANAKA^{1,d)}

Received: December 20, 2012, Accepted: April 7, 2013

Abstract: In search tasks composed of multiple sub-tasks (*e.g.*, trip planning task), it is difficult for searchers to obtain information that satisfies their information needs completely with a single search query. In these tasks, they usually have to search iteratively by reformulating the query. This paper addresses the problem of predicting searchers' query reformulations beforehand. The proposed method tries to predict which category the following reformulation belongs to by constructing a classifier from search behavior data. As a result of evaluation based on the real search log data, we found that our method can predict query reformulation types with about 41% accuracy. We also analyze which and to what extent the user's behavior data is useful for predicting query reformulations. In addition to that, we discuss the predictability of the query itself issued in the following the search as the next step of this work.

Keywords: query reformulation, search behavior, log analysis, eye tracking, information retrieval

¹ 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University, Kyoto
606-8501, Japan

² 日本学術振興会特別研究員 (DC1)
JSPS Research Fellow (DC1), Chiyoda, Tokyo 102-0083,
Japan

³ 明治大学総合数理学部
School of Interdisciplinary Mathematical Sciences, Meiji
University, Nakano, Tokyo 164-8525, Japan

⁴ 科学技術振興機構 CREST
JST CREST, Chiyoda, Tokyo 102-0076, Japan

a) umemoto@dl.kuis.kyoto-u.ac.jp

b) satoshi@snakamura.org

c) tyamamot@dl.kuis.kyoto-u.ac.jp

d) tanaka@dl.kuis.kyoto-u.ac.jp

1. はじめに

Web 検索で入力される検索クエリのうち, 約28%は以前の検索において利用されたクエリが修正され再利用されている [18]. このことから, 多くの検索ユーザは, 検索クエリの修正を繰り返すことで, 求める情報にたどり着こうとしていると考えられる. たとえば, 京都旅行に関する情報を Web 検索で調べる場合, ユーザは訪問する観光地から宿泊するホテルまで, さまざまな観点から検索を行う必要がある. そのため, “京都 観光地 おすすめ” という検索クエリで人気の観光地をある程度探した後で, クエリを “京

都 ホテル 激安” に修正するなどして、検索対象となる情報を変更する。

クエリ修正が起こる別の要因として、検索ドメインに対する、ユーザの知識量の影響があげられる。医療分野に対してあまり知識のないユーザがアルツハイマー病の治療薬について調べる際に、最初の検索の段階で、具体的な治療薬の名前を検索クエリに投入することは難しいと考えられる。こうしたユーザは“アルツハイマー病 治療薬”という一般的な語をクエリとして検索を行い、その検索を通してアルツハイマーの治療薬に「ドネベジル」というものがあることを知るだろう。その後で、ようやく“アルツハイマー病 ドネベジル”というクエリを用いることで、もともとの検索よりもさらに専門的な情報を得ることができるようになる。

このようなユーザの検索クエリの修正は5種類のカテゴリに分類されるといわれている [1]。たとえば、上述の京都の観光地から、京都のホテルへの検索の遷移については、京都の観光に関する検索タスクという意味では変わらないが、検索の前後で対象となる検索の観点が変わっている。こうしたクエリ修正は *Parallel Move* とよばれている。一方、“アルツハイマー病 治療薬” から“アルツハイマー病 ドネベジル”という検索クエリの修正では、ユーザは現在の検索結果よりも、より詳細な情報に絞り込もうとしているといえる。このようなクエリの修正を *Specialization* とよぶ。他のクエリ修正も含めて、詳しくは3章で述べる。

複数回のクエリ修正をともなう検索タスクで、毎回適切なクエリを考えることは容易ではない。そのため、このようなユーザに対して、検索エンジン側がクエリの推薦をすることは、大きな検索支援になると考えられる。しかし、既存のクエリ推薦は、ユーザが検索エンジンに何らかの検索クエリを入力してはじめて行われるものであるため、彼らを支援することは難しい。こうした問題に対する1つの解決方法として、ユーザがクエリの入力始める前に、次にどのようなクエリが入力されるかを予測することで、その予測に基づいたクエリ推薦を行うということが考えられる。

もし、次の検索対象を事前に知ることができれば、検索エンジンはユーザの行動に応じて、次に推薦する検索クエリを動的に切り替えることが可能になる。たとえば、検索行動中にあるトピックに関する情報に絞り込みたいとユーザが感じた場合にはトピックの関連語を、別の観点で情報を調べたいと思ったユーザには現在のクエリの同位語を、次の検索クエリの入力前に推薦することができる。また、検索クエリだけでなく、次に検索してアクセスするページまでも予測し、そのページを推薦するといったことも考えられる。

こうした検索クエリの事前の推薦を最終目標として、本稿ではその第一歩として、次に行われる検索において、そ

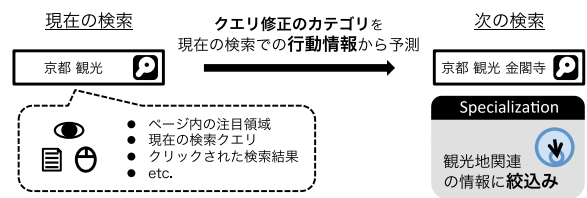


図 1 本研究で取り組む問題の概要。現在の検索時におけるさまざまな行動情報から、次の検索でのクエリ修正のカテゴリを推定する

Fig. 1 Overview of the addressed problem. Proposed method predicts the category of following query from user behavior in the current search.

のクエリ修正がどのカテゴリに属するかを予想することに取り組む。本稿では、次の検索におけるクエリ修正タイプの推定にあたって、「現在の検索時におけるユーザの行動情報は、その後で行われる検索に対して大きな影響を与える」という仮定を置く。我々のアプローチの概要は図1のとおりである。検索修正における次のカテゴリの推定にあたって我々は、検索時におけるユーザのさまざまな行動情報をログデータとして収集し、特徴量の計算を行う。具体的には、Web ページ上の視線位置やカーソルの移動速度といったデータに加えて、現在の検索クエリやユーザが閲覧したページの URL、クリックされた検索結果の順位のような、既存のユーザ行動分析に関する研究 [10], [24] で用いられてきたデータについても、推定のための特徴量として採用する。これらの特徴量をもとに素性を計算し、サポートベクタマシン (SVM) を用いた分類器を構成することで、ある検索に関するユーザの行動情報が与えられた際に、その次の検索でのクエリ修正のカテゴリの予測を実現する。我々は、実際の検索行動から構成した分類器を用いて、提案手法によるクエリ修正の予測精度の評価を行う。また、その予測の際に、どういった行動に関する素性が有効に働いているのかを調べることで、検索時のユーザ行動とクエリ修正との関係性を明らかにする。

本研究の情報検索分野における貢献を以下に列挙する。

- 現在の検索が次の検索に与える影響を考慮することで、次の検索クエリの修正の種類を、ユーザの行動情報を用いて事前に推定する手法を提案し、その精度の評価を行った点
- 収集したユーザの行動ログと提案手法による分類結果を詳細に分析することで、視線情報をはじめとする種々のユーザ行動とクエリ修正との間の関係性の解明に取り組んだ点

2. 関連研究

近年では、Web を通じてさまざまな情報を手に入れることができるため、ユーザの検索意図は多様化してきている。たとえ検索クエリが同じであっても、実際に求められ

ている情報は、ユーザごとに異なることがある [17]。そのため、Web 検索エンジンがユーザを効果的に支援するためには、検索時のユーザの意図をリアルタイムに推定し、その意図に応じた情報を提供することが重要となる。こうした背景から近年では、ユーザの検索意図の推定に関する研究がさかんに行われている。

2.1 行動情報を用いたユーザ行動分析と応用

ユーザの検索意図推定には大きく分けて次の2つのアプローチが存在する。一方はクエリログやクリックスルーログのようなサーバ側で取得可能な大規模ログデータを分析することで、典型的な検索意図を推定するというものである。もう一方は、クライアント側によって記録された詳細な行動情報を利用することで、より精細な検索意図を発見しようとするものである。

検索時のクエリや検索結果に対するクリックといったデータから、Web 検索においてどんな情報が求められているかや、検索結果ページに対するユーザのインタラクションを知ることができる。こうしたデータは、検索エンジン側で容易に取得可能なため、検索におけるユーザのフラストレーション検出 [10] や、検索タスクに対する満足度を推定するためのユーザ行動のモデリング [16]、検索エンジンの切替えの予測 [15], [24] といったような、さまざまな目的で利用されている。

Boldi ら [1] は、検索クエリのログデータから機械学習により分類器を構成することで、連続する2つの検索クエリが与えられた際に、その修正の種類に対してカテゴリ分類を行っている。彼らの手法では、修正前後の2クエリを分類器の素性とするすることで、カテゴリ推定を実現している。しかし、実際の Web 検索において、次に入力されるクエリを事前に知ることは困難であるため、この手法を検索タスク中のクエリ修正のオンライン推定に適用することはできない。本稿では、検索クエリの修正カテゴリの推定において、現在の検索クエリおよびそのときの行動情報のみを利用する手法を提案するため、既存手法に比べて適用範囲が広いと考えられる。

Guo らは、既存の研究で広く使われてきたクエリログやクリックスルーデータに加え、スクロールやホバー位置のようなマウス操作に関するデータの有用性を提唱している [13]。彼らはこうした複数の情報源を利用することで、ユーザの検索タスクが情報収集指向であるのか、購買指向であるのかを分類する手法を提案している [11]。本研究は、ユーザの行動情報をもとにユーザの検索行動を分類するという点では、彼らの研究と類似しているが、情報収集や購買に比べより一般的な検索意図の遷移という問題を扱う。また、マウス操作に加えて、ページ中のどこを見ているかという特徴量も用いることで、視線情報の有用性についても考察する。

2.2 視線検出技術の Web 検索への利用

上述したように、これまでの研究では、クエリログやクリックスルーデータ、あるいはマウスカーソル操作情報を分析することで、Web 検索時におけるユーザの行動分析やそれに基づく支援手法の提案が行われてきた。近年では、粒度のさらに細かな分析が可能になるという点で、視線検出技術に研究者の注目が集まっており、その技術はさまざまな目的で利用されるようになってきている。

その主たる目的の1つに、Web 検索におけるユーザ行動の理解があげられる。Cutrell ら [8] は、Web 検索結果の表示内容や量を変更することで、ユーザの検索スタイルに変化があるかを分析するために視線検出器を用いている。また、Web ページ上での閲覧行動の特性分析のために、ユーザの視線情報を用いた研究も存在する [3]。こうした研究と同様に、本研究でもクエリ修正におけるユーザ行動の分析のために、視線検出技術を利用する。

視線検出技術の別の用途として、視線情報を利用した Web 検索支援という研究も行われている。Buscher ら [4] は、ユーザの視線位置を利用してクエリ拡張のためのキーワードを抽出する手法を提案した。彼らはこの手法の中で、リーディングやスキミングといったユーザの閲覧パターンを判定し、これらの特徴をキーワード抽出に利用している。また、我々もこれまでに、視線検出装置を用いた情報探索支援システムを提案してきた [23]。提案システムは、Web 検索時にユーザが注目している語から検索意図を推定し、その意図に応じて検索結果の再ランキングや、Web ページ中の適合情報の強調を動的に行うことで、ユーザのリアルタイムな情報要求を支援する。他にも、ユーザの読み方に応じて文書を要約するという研究 [25] にも、視線検出技術が利用されている。

現状では、視線検出器の普及率は決して高いものとはいえない。しかし最近になって、検出機能を搭載したノート PC のプロトタイプの公開^{*1}や、市販の Web カメラを利用した視線検出技術の開発^{*2}が話題となっている。また、マウス座標からの視線位置の推定に関する研究 [12] がさかんに行われていることから、多くの研究者が検索ユーザの視線の動きに興味を示していることが考えられる。そのため、今後の視線検出に関する研究の重要性は高まっていくものと思われる。

3. 問題定義

本研究では、現在の検索におけるユーザの行動情報から、検索クエリがどう修正されるかを推定するという問題に取り組む。本稿では、検索クエリ修正の種類をクエリ修正タ

*1 <http://www.tobii.com/en/group/news-and-events/press-room/#/video/view/tobii-presents-the-eye-controlled-laptop-prototype-at-cebit-2011-4864>

*2 <http://www.inference.phy.cam.ac.uk/opengazer/>

イブ (query reformulation type; QRT) とよぶ。本章ではまず、各クエリ修正タイプを紹介する。その後、本稿で取り組む問題の定式化を行う。

3.1 クエリ修正タイプ

Boldi らの研究 [1] では、連続する 2 つの検索クエリに対する修正タイプとして (1) *Generalization*, (2) *Specialization*, (3) *Parallel Move*, (4) *Mission Change*, および (5) *Error Correction* の 5 種類をあげている。

Generalization とは、ユーザが検索空間を広げる際に生じるクエリ修正タイプである。たとえば、検索クエリが「京都 和食 おすすめ」から「京都 ごはん おすすめ」に修正されたときは、「京都でおすすめの和食」という条件が「京都でおすすめのごはん」へと緩められたと考えられる。このような修正のカテゴリが *Generalization* である。*Specialization* は *Generalization* と対をなす修正タイプであり、ユーザが現在の検索範囲をさらに絞り込む際に生じる。検索クエリが「京都 観光 お寺」から、より具体的な「京都 観光 清水寺」に変化したときなどがこれに相当する。*Parallel Move* は、検索のコンテキストは変えないまま、別の観点から検索を行うといった際に生じる。このクエリ修正タイプが生じる検索の例としては、旅行の計画のために宿泊先のホテルについて調べていたユーザが、続いて観光地を調べるといった状況があげられる。*Mission Change* は、*Parallel Move* よりも広い概念であり、移動後の検索空間が現在とは大きく異なる際の修正タイプと定められる。閲覧中のページで現在の検索対象とは関係のない広告の内容が気になり、それに関する情報を検索してしまうという行為は *Mission Change* の一例である。最後に、*Error Correction* は、スペルミスの修正や別表現への書き換えのように、検索クエリは変わるもののユーザの検索意図そのものは変わらないようなクエリの修正を指す。

3.2 問題の定式化

本稿では、上述のクエリ修正タイプ集合を $T = \{G, S, P, M, E\}$ によって表す (各クエリ修正タイプをその名称の 1 文字目で記している)。また、ユーザ u が入力する検索クエリ系列の集合を $Q_u = \{q_i^u\}_{i=1}^{|Q_u|}$ 、クエリ q_i^u による u の検索時の行動情報に関するログデータを $BL(q_i^u)$ 、 q_i から q_{i+1} にクエリが修正された際の修正タイプを $QRT(q_i^u, q_{i+1}^u) \in T$ と定義する。本稿で扱う問題はこれらの定義に従うと、ユーザ u の次の検索クエリの修正タイプ $t = QRT(q_c^u, q_n^u)$ を、 q_c^u をクエリとする現在の検索に関する行動ログデータ $BL(q_c^u)$ から推定することと定式化できる。ここで、 q_n^u は、次の検索において u が入力する検索クエリであり、本研究ではその具体的な値を用いずに修正タイプ t の推定に取り組むことによって、応用として次の検索を行う前の時点でのユーザ支援が可能となる。

3.3 問題解決のための基本的アイデア

1 章でも述べたように、我々はクエリ修正タイプの予測に取り組むにあたって、「現在の検索におけるユーザの行動情報をもとに機械学習を用いてクエリ修正タイプの分類器を構築する」というアプローチを採用する。2 章において、ユーザの検索行動の収集が可能な情報源には、クエリログやクリックスルーデータ、視線情報などさまざまなものがあるが、我々は各情報源からそれぞれが得意とする特徴量を抽出する。たとえば、ユーザが入力した検索クエリからは、現在の検索対象の範囲を推定することができる。また、ページ閲覧時のユーザの視線の動きから、現在ユーザがどんな情報に興味を示しているかといったことも分かる可能性がある。そこで、各情報源の性質を考慮したうえで、分類器構築に用いる素性の検討を行う。ここで述べたアイデアをもとに、次章では提案手法によって実際に分類器を構築するまでの流れを詳細に説明する。

4. クエリ修正タイプの予測

4.1 行動ログデータの収集方法

クエリ修正タイプの予測に向けて、まず Web 検索時のユーザの行動ログの収集を行う必要がある。1 章でも述べたように、本研究では通常のクエリログやクリックスルーデータに加えて、ユーザのマウス操作や視線情報などの多様な情報源から、ユーザの行動情報を抽出する。記録するログの一覧を表 1 に示す (各行動情報は、その発生時刻を示すタイムスタンプとともに記録される)。今回は、ユーザの視線情報を取得するにあたって、Tobii 社^{*3}の視線計測装置 Tobii T60 およびソフトウェア開発キット Tobii SDK を用いる。また、マウスおよび視線のモニタ上における座標位置に関しては、既存研究 [19] を参考にそれぞれ 100 ms, 25 ms 間隔で取得する。

4.2 収集したログデータからの訓練データの生成

次に、このようにして収集したユーザの行動ログから、クエリ修正タイプを予測する分類器構築のための訓練データを生成する。その生成にあたって、まずはじめに図 2 に示すように、各ログデータを検索クエリごとのグループへの分割を行う。すなわち、ユーザ u があるクエリ q^u を用いて検索を行った際の一連の行動情報 $BL(q^u)$ が 1 つの

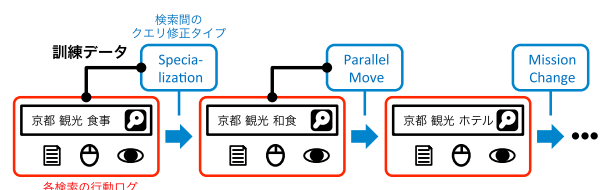


図 2 行動情報からの訓練データの構成

Fig. 2 Constructing training data from behavior log data.

*3 <http://www.tobii.com>

表 1 クエリ修正タイプの事前予測に取り組むにあたって検索時に収集する行動情報
 Table 1 Behavior data collected during Web search, which is used for predicting QRTs in advance of the following search.

情報源	記録する行動情報
検索結果ページ	検索クエリ, 総検索結果数 (ヒットカウント), 検索結果オフセット, ページ HTML, ページサイズ, 出現語集合, 検索結果リスト, 検索結果領域, クリックされた検索結果, 推薦クエリリスト, クエリ推薦領域, クリックされた推薦クエリ, 広告リスト, 広告領域, クリックされた広告
非検索結果ページ	ページタイトル, ページ URL, ページ HTML, ページサイズ, 出現語集合
マウス操作	ブラウザのタブ切替え, カーソル位置, クリックイベント, スクロールイベント, カーソル上の単語 (100 ms 間隔で取得)
視線情報	モニタ上の視線位置, 注目された語 (25 ms 間隔で取得)



図 3 訓練データ構成のためのラベリングシステムのインタフェース
 Fig. 3 Labeling system interface for constructing training data.

訓練データに該当する. このようにして構成された各グループに対して, 分類のための正解ラベルを付与する. あるグループ $BL(q_i^u)$ に対する正解ラベルには, そのグループにおける検索クエリと, その次の検索に関するグループ $BL(q_{i+1}^u)$ におけるクエリ間での修正タイプ $QRT(q_i^u, q_{i+1}^u)$ を割り当てる. たとえば図 2 のように, “京都 観光 食事” というクエリによる検索に続いて, “京都 観光地 和食” というクエリで検索が行われている場合には, 前者のグループに対して Specialization が正解ラベルとなる. ただし, 以降の検索クエリが存在しないグループに対しては, 今回は, その正解ラベルを Mission Change として扱う.

ここで, 第三者の検索行動から実際に正解ラベルを付与することは, ときとして簡単ではない場合がある. そのため, 我々は, 図 3 に示すようなプログラムを実装し, 実際に検索を行ったユーザ自身にラベリングを行ってもらうという方法を選んだ. クエリ修正タイプのラベリングを行うには, 検索のときの状況をユーザに可能な限り正確に思い出してもらう必要があるため, このプログラムでは, 検索時に入力されたクエリに加えて, 閲覧された Web ページやクリックされた検索結果といったさまざまな行動情報がその発生時刻とともに提示されるようになっている.

またラベリングの際には, 修正前後での検索意図や目的の変化を基準にしてもらった. 具体的には, 3 章で述べた各クエリ修正タイプの性質に従って,

- G 現在の検索よりも広い条件で検索を行う場合
- S 現在の検索に対し, さらに絞り込んで情報を探す場合

- P 同一タスクの中で, 別の観点から検索を行う場合
 - M 現在とは, 別のタスクに移り変わる場合
 - E クエリ修正の前後で求める情報に変化がない場合
- という判断基準を与え, これに従ってクエリ修正タイプの判定を行ってもらった.

被験者に検索行動の収集およびクエリ修正タイプのラベリングを行ってもらう際には, 事前に数回の検索を通してプログラムを利用してもらい, 普段の検索時と行動が大きく異ならないように配慮した.

4.3 分類器構築に用いる素性

本稿では, クエリ修正タイプの予測のアプローチとして, 分類問題に対して一般的に用いられている機械学習アルゴリズムである SVM を採用する. 分類器を構成するためには, 収集した行動情報のログデータから, どういった特徴量を素性に利用するかを考察する必要がある. 我々は, 以下に示す 5 つのカテゴリから表 2 に示す種々の素性を計算することで, クエリ修正タイプの予測を行う分類器を構築する. ここで表中の各素性の左側には素性番号 f_i が表記されている. 以降の表では, 素性を表す際にこの番号を用いる.

4.3.1 遷移カテゴリ

クエリ修正の分析に関する Boldi らの先行研究 [1] では, Specialization と Generalization の 2 つの修正タイプが交互に起きやすい傾向にあると報告されている. たとえば, ある情報について詳しく調べようと思いきやクエリを詳細化してみたものの, 絞り込みすぎて適合結果が少なすぎた場合には, ユーザはクエリの条件を緩和することで検索結果の再現率を上げようとすると考えられ, この場合は Specialization に続いて Generalization が生じる. 先行研究では英語圏のデータに対する分析を行っているため, 日本語の場合においても同様の傾向があるとは限らないが, 我々は現在の検索行動の情報に加えて, 前回から今回の検索へ移行する際のクエリ修正タイプについても, 分類器の素性に利用する.

4.3.2 クエリカテゴリ

検索エンジンに対して入力されたクエリは, ユーザの検索意図を多かれ少なかれ表現しているといえる. たとえば, 入力クエリ中の単語数や文字数といった情報から, 現

表 2 クエリ修正タイプ分類器の構成に用いる素性
Table 2 Features used for constructing classifier.

遷移カテゴリ	
[f ₁]	直前のクエリ修正タイプが Generalization か
[f ₂]	直前のクエリ修正タイプが Specialization か
[f ₃]	直前のクエリ修正タイプが Parallel Move か
[f ₄]	直前のクエリ修正タイプが Mission Change か
[f ₅]	直前のクエリ修正タイプが Error Correction か
クエリカテゴリ	
[f ₆]	クエリに含まれる単語数 ^{*4}
[f ₇]	クエリ中の文字数
[f ₈]	検索結果のヒットカウント
[f ₉]	検索結果ページ中のクエリ推薦数
[f ₁₀]	検索結果ページ中の広告数
クリックカテゴリ	
[f ₁₁]	検索結果に対する総クリック数
[f ₁₂]	推薦クエリに対する総クリック数
[f ₁₃]	広告に対する総クリック数
[f ₁₄]	$f_{11} + f_{12} + f_{13}$
[f ₁₅]	検索結果ページと非検索結果ページの切替数
[f ₁₆]	f_{15}/f_{14}
[f ₁₇]	1つ以上の検索結果がクリックされたか
[f ₁₈]	最初にクリックされた検索結果の順位
[f ₁₉]	クリックされた検索結果の順位の平均
[f ₂₀]	推薦クエリがクリックされたか
[f ₂₁]	クリックされた推薦クエリの順位
[f ₂₂]	1つ以上の広告がクリックされたか
[f ₂₃]	最初にクリックされた広告の順位
[f ₂₄]	クリックされた広告の順位の平均
従事カテゴリ	
[f ₂₅]	閲覧された検索結果ページの最大オフセット
[f ₂₆]	閲覧された検索結果ページ数
[f ₂₇]	検索結果ページ上での総滞在時間数
[f ₂₈]	各検索結果ページ上での平均滞在時間
[f ₂₉]	閲覧された非検索結果ページ数
[f ₃₀]	非検索結果ページ上での総滞在時間
[f ₃₁]	各非検索結果ページ上での平均滞在時間
視線カテゴリ	
[f _{32,33,34}]	検索結果領域に対する注目割合の平均値, 最大値, 最小値
[f _{35,36,37}]	クエリ推薦領域に対する注目割合の平均値, 最大値, 最小値
[f _{38,39,40}]	広告領域に対する注目割合の平均値, 最大値, 最小値
[f _{41,42,43}]	検索結果ページにおける視線移動速度の平均値, 最大値, 最小値
[f _{44,45,46}]	非検索結果ページにおける視線移動速度の平均値, 最大値, 最小値
[f _{47,48,49}]	非検索結果ページの左部に対する注目割合の平均値, 最大値, 最小値 ^{*5}
[f _{50,51,52}]	非検索結果ページの右部に対する注目割合の平均値, 最大値, 最小値 ^{*5}
[f _{53,54,55}]	非検索結果ページの上部に対する注目割合の平均値, 最大値, 最小値 ^{*5}
[f _{56,57,58}]	非検索結果ページの下部に対する注目割合の平均値, 最大値, 最小値 ^{*5}
[f _{59,60,61}]	非検索結果ページの中央部に対する注目割合の平均値, 最大値, 最小値 ^{*5}
マウスカテゴリ	
[f _{62,63,64}]	検索結果ページにおけるカーソル移動速度の平均値, 最大値, 最小値
[f _{65,66,67}]	非検索結果ページにおけるカーソル移動速度の平均値, 最大値, 最小値

在の検索対象の広さというものを推定することができる。ここで、検索対象が広すぎる場合には、ユーザは検索結果をもう少し絞り込みたいと考えるはずである。そのため、こうした特徴量は次の検索クエリ修正の予測に有効かもしれない。

ユーザの行うクエリ修正の多くは Parallel Move や Specialization であるということと、クエリ推薦がクエリログ

^{*4} 実験の被験者は、キーワード集合を空白記号で区切ったものを検索クエリとして用いていたので、今回は形態素解析による単語数の計算を行わず、空白記号によるクエリの分割を行った際の分割数を、 f_6 の値として用いた。

^{*5} ページ幅を w 、ページ高を h としたときに、 x 座標が 250 未満の領域を左部、 $w - 250$ を超えるものを右部とし、それ以外の領域については、 y 座標が 250 未満の領域を上部、 $h - 250$ を超えるものを下部とし、これらのどの領域にも含まれない範囲を中央部として、注目割合の計算を行った。

をもとに行われるということ [1] から、検索エンジンが推薦するクエリは、この2種類が大部分を占めると考えられる。そこで、推薦クエリに関する素性は、これらの推定に貢献することが期待できる。

ここで、Boldi ら [1] が用いているクエリペア間の Jaccard 係数を利用すれば、高い精度でクエリ修正タイプの予測が可能になるであろう。しかし、本研究の目的は、次の検索に関する知識を用いることなく修正タイプを予測することであるため、今回はこうしたクエリペア間から得られる特徴量は利用しない。

4.3.3 クリックカテゴリ

クリックスルーに関する既存研究 [20] で述べられているように、検索結果のクリック数は、その検索に対するユーザの満足度の指標に使えるかもしれない。そのため、ハイパーリンクに対するクリック情報からもさまざまな特徴量を抽出する。その例として、「検索結果を1つ以上クリックしたかどうか」などがあげられる。ユーザが現在の検索において、どの検索結果もクリックすることなく次の検索に移動した際には、ユーザはより良い情報を求めてクエリ修正を行ったと考えるのが妥当である。そのためこのケースでは、まったく別の検索への移行である Mission Change は発生しにくくなると予想される。

検索結果以外にも、広告に対するクリック情報も考えられる。この特徴量を、ユーザの興味の移り変わりの指標とすることで、Mission Change や Parallel Move の予想に使うことができるかもしれない。

4.3.4 従事カテゴリ

現在の検索に対して、ユーザがどれだけ深く関わっているかに関する素性がこのカテゴリに属する。「検索結果を何ページ目まで閲覧したか」は、検索タスクの難易度に暗に示唆している。Boldi らが行った英語圏でのクエリログの分析 [1] によると、仕事や勉強のような必要に迫られて行われる検索タスクの後には、娯楽関連の情報収集のようにリラックスを目的とした検索が行われやすいといわれている。そのため、こうした素性を用いることで、現在の検索に対するユーザの従事度を把握できれば、Mission Change のように別トピックへの遷移の推定が可能になるかもしれない。

Guo ら [13], [14] は、dwell time とよばれる「検索結果をクリックしてある Web ページを訪れてからそのページの閲覧を終えるまでの時間」についての特徴量を用いることで、検索結果の適合性や検索に対するユーザの満足度の推定を試みている。検索結果が適合しているか、あるいは検索に満足しているかによって、次に行われる検索の種類は異なることが予想されるため、こうした素性は本研究においても利用可能である。

4.3.5 視線カテゴリ

同じ Web ページを閲覧していても、どの情報に興味を

いだろうか。ユーザの検索意図に大きく依存すると考えられる。また、検索タスクの性質というのも、少なからず影響するであろう。実際に既存研究 [11] において、商品購入に関する検索に代表されるトランザクショナル検索 [2] では、他のタスクに比べて広告のクリック率が上昇するという実験結果が報告されている。このようなケースでは、検索結果ページから遷移したページの内容が、次の検索に大きな影響を及ぼすことが考えられる [6]。たとえば、ある商品の購入のために訪れた通販サイトで、いつの間にかまったく別の商品に興味が移ってしまった際には、ユーザはその商品に関する検索を次に調べようとするかもしれない。ページの左右に広告を配置している Web サイトは多いため、Web ページ中での注目領域の同定も、クエリ修正タイプの推定に役立つ可能性があるといえる。

4.3.6 マウスカテゴリ

マウスの操作情報からは、カーソル位置やページのスクロール速度など、さまざまな特徴量を計算できる。こうしたデータは、Web ページ中でのユーザの興味を示唆しているといえる [11]。これらもクエリ修正タイプの推定に利用可能と考えられる。これまで本稿では、ユーザの行動情報の情報源として、視線情報を利用するという何を何度か述べてきた。我々は、マウス操作から得られる特徴量の多くは、ユーザの視線情報を利用することで補完可能であり、また視線情報の方がより詳細な特徴をとらえることができると考えている。マウスカーソルはユーザが意識的に操作するものであるのに対して、視線は気になる対象に対して半意識的にも反応しやすい。たとえば、レポート課題の情報を調べているときに、自分の好きな芸能人の画像がページ中に表示されていれば、ついそこに目がいかってしまうかもしれない。その際には Mission Change が起きやすくなることが考えられるが、視線情報を使うことでその予想が可能になると我々は予想する。

このような考察をもとに、今回はマウス素性の利用は最小限にとどめ、ページ中での移動速度に関するものだけを採用する。しかし近年では、Web ページ中での視線とカーソルの関係分析や、カーソル操作からの視線位置の推定に関する研究はさかに行われている [5], [12], [22]。そのため、一般の検索ユーザの行動情報のように、現状では視線情報に関するデータが得られないケースであっても、このカテゴリの素性を使うことで、視線カテゴリ内のいくつかの素性を代用でき、クエリ修正タイプの推定を実現できると我々は考える。

4.4 ユーザの特性が素性に与える影響

前節で述べたクエリ修正タイプ分類タイプの素性には、ユーザの性格や検索の行い方に影響を受けるものとそうでないものが存在する。たとえばクエリカテゴリには、検索結果のヒットカウントや、クエリ推薦数、広告数といった、

ユーザの行動に依存しない特徴量が多く含まれている。そのため、ユーザの特性による影響は比較的少ないことが予想される。一方で、クリックカテゴリに属する「検索結果中でのどの結果をクリックするか」といった特徴や、マウスカテゴリおよび視線カテゴリに関連する「ページ中でどういった情報に興味を持つか」といった特徴は、ユーザごとに異なる可能性が高いはずである。そのため、こうした性質を持つ素性が多く含まれるクリックカテゴリや、視線カテゴリ、マウスカテゴリは、ユーザの特性に関する依存性が特に高いと思われる。同様に、従事カテゴリの特徴である「ページの閲覧にどの程度、時間をかけるか」という点に関しても、その値のとりうる範囲には、ユーザごとに差異があると考えられる。しかし、どのユーザに対しても、ページの閲覧に時間をかける場合とかけない場合とでは、それぞれの特徴量の値域に開きがあることが予想される。そのため、従事カテゴリに関しても、ユーザ特性による依存性は存在するものの、その影響はクリックカテゴリや視線カテゴリ、マウスカテゴリの依存性に比べると、比較的低いと考える。

このように、クエリ修正タイプの分類に素性には、ユーザの特性に影響を受ける可能性のあるものが存在するため、本研究では、クエリ修正タイプの分類器をユーザごとに構成するというアプローチをとる。

5. 評価

提案手法によるクエリ修正タイプの予測可能性を評価するため、前章で述べた方法をもとに Web 検索の行動情報を収集し、クエリ修正タイプの分類器を構成した。以下では、行動情報の収集によって得られたデータを概観した後で、分類器の構築の具体的な手続きを示し、その分類器に対して行った評価について述べる。

5.1 収集した行動情報の概観

4章で述べた手続きに従って、クエリ修正タイプのラベルを含む行動情報の収集を行った。データの収集は、第1著者の2012年4月4日、25日、5月4日、6日、7日の計5日間にわたる Web 検索を対象とした*6。なお、検索結果の取得に関しては、Yahoo! JAPAN の検索 API を利用した*7。また、検索行動の収集の際に、被験者に明示的な検索タスクを与えることはせず、普段と同様に検索を行ってもらうことにした。

上記5日間にわたって被験者の検索を記録し続けた結果、183の検索クエリからなる行動情報を得ることができた。実際に収集したログデータに含まれていた検索タスクや、

*6 被験者の個人的な都合のため、最初の収集日とそれ以外の収集日には開きがあるが、どの収集日におけるクエリ修正タイプの分布も表4に示した結果と大きな差異はなかった。そのため、収集日の偏りによる実験結果への影響は少ないものと考えられる。

*7 <http://developer.yahoo.co.jp/webapi/search/>

表 3 収集した行動情報に含まれていたタスクにおける検索クエリの修正例。クエリ間の矢印の上の記号は、その修正タイプを表している

Table 3 Examples of search tasks and query reformulations in the collected data. A symbol on an arrow means a QRT between two queries before and after it.

検索タスク	クエリ遷移例
就職活動の情報収集	“NTT ドコモ グループディスカッション” \xrightarrow{E} “NTT ドコモ グループ 面接” \xrightarrow{P} “NTT データ 二次面接”
博士課程への進学	“ドクター 進学” \xrightarrow{S} “ドクター 進学 就職” \xrightarrow{P} “ドクター 進学 決めて” \xrightarrow{E} “ドクター 進学 決め手”
ヨーロッパ旅行	“イタリア カプリ島” \xrightarrow{S} “イタリア カプリ島 青の洞窟” \xrightarrow{G} “イタリア カプリ島 観光スポット”
PC トラブル	“パソコン 排熱” \xrightarrow{S} “パソコン 温度 高い” \xrightarrow{P} “パソコン 温度 speedy” \xrightarrow{E} “パソコン 温度 speccy”
プログラミング言語	“型付け” \xrightarrow{S} “動的型付け 強い 弱い” \xrightarrow{P} “python オブジェクト指向” \xrightarrow{S} “python オブジェクト指向 継承”

表 4 収集した行動ログに含まれていたクエリ修正タイプの分布。クエリ修正タイプ5種類の分布とセッション内 (Mission Change 以外) の分布を、参考として Yahoo! UK および Yahoo! US のクエリログ内での分布 [1] とともに示している

Table 4 QRT distributions in the collected data. The left side is about two kinds of distributions: all the five QRTs and the four QRTs except Mission Change. The distributions in query logs of Yahoo! UK and Yahoo! US [1] are shown on the right side for reference.

クエリ修正タイプ	本研究での分布		既存研究 [1] での分布	
	すべて	セッション内	UK	US
Generalization	13.1%	15.1%	4.4%	9.5%
Specialization	36.6%	42.1%	37.5%	30.1%
Parallel Move	29.5%	34.0%	47.7%	55.5%
Error Correction	7.7%	8.8%	10.4%	5.0%
Mission Change	13.1%	—	—	—
データサイズ	—	0.2k	6.5M	10.5M

そのタスクにおけるクエリ修正の例を表 3 に示す。また、ログデータ中における各クエリ修正タイプの分布を表 4 に示す。なお、同表には参考として、既存研究 [1] によって調査された、Yahoo! UK と Yahoo! US の 2 つの検索エンジンのクエリログ中におけるクエリ修正タイプの分布も示している。表中では、2 種類の方法で修正タイプの分布を計算している。一方は、3 章で述べた 5 種類すべての発生割合である。そしてもう一方は、Mission Change を除いた 4 種類に対して、その発生割合を計算するというものであり、この値から検索セッション中における各クエリ修正タイプの分布というものを知ることができる。

今回収集したログデータ中では、Specialization または Parallel Move の頻度が高く、その 2 つが全体において占める割合は 66.1%であった。さらにセッション中での分布に限定した際には、その値は 76.1%に上昇し、この結果は表の右側に示した既存研究におけるそれぞれの値 85.2%および 85.6%に近いものとなった。今回収集したデータは日本語による検索のログであり、既存研究のログは英語圏のものであるため、単純に比較することは難しいが、これらの 2 つのクエリ修正タイプが多数を占めているという結果

については、両者のデータ間で一貫しているといえる。

今回検索ログを収集した被験者の特徴として、1 つ 1 つの検索タスクを時間をかけて慎重に行うということがあげられる。被験者が行った 183 回の検索中には、24 回の Mission Change が含まれていた。ここから、このユーザは 1 タスクあたり平均して 7 回程度の検索を行っていることが分かる。こうした検索ユーザの特性が実験結果に与える影響に関しては、6.1 節で考察を行う。

5.2 分類器の構成

3 章で述べたように、本稿ではクエリ修正タイプの予測を多クラス分類問題として扱う。その分類器の構成には、分類問題に対して広く扱われる SVM ライブラリである LIBSVM^{*8}を採用した。SVM のカーネルについては RBF カーネルを使用し、パラメータ値に関してはデフォルト値をそのまま用いることとした。本稿では、5 種類のクエリ修正タイプ全体に対する分類という観点と、各修正タイプに対する (2 値) 分類という 2 つの観点から、提案手法の有用性を検証するために、以下のそれぞれのアプローチ [9] で分類器を構成し、それぞれについて、分類精度の評価を行うことにした。

- 1 対 1 方式：2 クラスの各組合せに対して、与えられたデータがどちらのクラスに属するかを判定する分類器を構成し、それぞれの分類器の出力結果に対して多数決によって最終的な分類クラスを決定する方式。クラス数 n の問題に対して、 $nC_2 = n(n-1)/2$ 個の分類器を構成することで、多クラス分類を実現する。
- 1 対他方式：各クラスに対して、与えられたデータがそのクラスに属するかを判定する分類器を構成し、それぞれの分類器が出力した decision value や所属確率の値に応じて、分類クラスを決定する方式。この場合は、クラス数と同数の分類器が構成されることになる。

5.3 クラス集合全体に対する多クラス分類

まずは、1 対 1 方式を用いて多クラス分類器を構築することで、クエリ修正タイプ全体での分類に対する評価を

^{*8} <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 5 各手法におけるクエリ修正タイプの分類精度

Table 5 QRT classification accuracy for each method.

SVM	RAND	DOM
0.410	0.200	0.366

行った。

5.3.1 分類精度

分類精度の評価のために、5.1 節で述べた行動情報のログデータに対して、leave-one-out 交差検定を行った。ここで leave-one-out 交差検定とは、1つのデータをテストデータとし、残りのデータで学習するという手続きを全組合せ数行う方法である。提案手法（以降、SVM と表記する）の分類精度の比較のためのベースラインとして、RAND と DOM の 2 種類を用意した。RAND 手法は与えられた行動情報に対して、次の検索に対するクエリ修正タイプを、全タイプの中から等確率で毎回ランダムに決定する。一方 DOM 手法は、発生頻度の最も高いクエリ修正タイプをつねに出力するものであり、今回の場合は 5.1 節で示したとおり、つねに Specialization を正解と見なす。

分類器の精度の評価尺度としては、Accuracy を用いる。ここで、Accuracy はテストデータのうち、正解ラベルであるクエリ修正タイプを正しく分類することのできた割合と定義される。RAND 手法の分類精度は、その性質上 0.2 であり、DOM 手法については、今回収集したログデータの分布から 0.366 とした。各手法の分類精度をまとめたものを表 5 に示す。

上表に示すとおり、我々の提案である SVM 手法では、0.41 の精度でクエリ修正タイプを分類することができた。これは、ベースラインよりも高い精度ではあるが、特に DOM 手法との精度差はわずかであった。その大きな原因として、今回分類器の構築に用いたデータサイズが少なかったことが考えられる。これについては、6 章で考察する。

5.3.2 分類貢献度の高い素性

ここでは、分類器の構成に用いた素性のうち、クエリ修正タイプの推定に貢献しているものを示す。表 2 に示した各素性に対する分類貢献度を知るために、その素性を除いて分類器を構成した場合の精度と、もともとの分類精度との比較を行った。ここで、除いたことによって大きく精度が下がる素性ほど、クエリ修正タイプの分類に貢献していると考えられる。一方で、除去後の分類精度が上昇する場合については、その素性は今回採用した中では比較の有用ではないものといえる。各素性に対して、この手続きを行った際の結果を表 6 に示す。なお、この表中に存在しない素性については、除去の前後において分類精度が変化しなかった。

上表より、「1 つ以上の検索結果がクリックされたか」や「非検索結果ページ中での左部に対する注目割合」、「クエ

表 6 分類に有用/非有用な素性およびその素性除去後の分類精度。除外後の精度が高い素性ほど、分類貢献度が高い

Table 6 Useful/useless features for classification with accuracy after ablation. A feature that has low accuracy makes a great contribution to the classification.

クエリ修正タイプの分類に対する貢献度の高い素性		
1: f_{17} (0.393)	2: f_{48} (0.399)	3: $f_{3,6,31}$ (0.404)
クエリ修正タイプの分類に対する貢献度の低い素性		
1: f_1 (0.432)	2: $f_{9,60}$ (0.421)	3: $f_{5,10,62,63,64,44,59}$ (0.415)

りに含まれる単語数」、「各非検索結果ページ上での平均滞在時間」といった素性については、除去後の分類精度が減少していることが分かる。そのため、これらの素性は、次の検索におけるクエリ修正タイプの推定に有用であると考えられる。

今日の Web サイトでは、ページが複数ペインで構成されており、左や右といった周辺部に広告情報が表示されていることがよくある。今回の結果から、こうした領域へのユーザの注目に関する情報が、クエリ修正タイプの予想の手がかりとなることが分かる。この素性の貢献度が高かった理由としては、前章の素性検討の項目で述べたように、ページ周辺部への注目度が高いほど、次の検索では Parallel Move や Mission Change といった別の観点・トピックへのクエリ修正が起きやすくなることがあげられる。

また、クエリ中の単語数もクエリ修正タイプの推定に貢献していた。ユーザは検索を始める際に、最初は一般的で短い検索クエリで検索結果を得て、その内容をもとにクエリを具体化することで、情報を手に入れる傾向がある。そのため、この素性は Specialization の推定にうまく働いていたと考えられる。実際に、収集したログデータにおいて、クエリ語数が 3 語未満の場合は 60% の割合で次の検索クエリ修正が Specialization であったのに対し、現在の検索クエリが 3 語以上からなる場合は、次の検索におけるクエリ修正タイプのうち Specialization が占める割合はわずか 18% であった。

非検索結果ページでの滞在時間が短い場合は、ユーザはそのページに満足しておらず、再び同じ意図で検索を続けることが多いといえる。そのため、この素性もクエリ修正タイプの推定に有効に働いたものと我々は考える。

このような分類寄与度の高い素性とは対照的に、除去によって分類精度が元の値よりも上昇する素性も多く存在した。こうした素性の実例として、「非検索結果ページ中での中央部に対する注目割合」や「直前のクエリ修正タイプが Generalization か」などがあげられる。これらの素性については、今回の結果からは、5 種類のクエリ修正タイプの分類の妨げになるといえる。

Web ページの左側への注目度がクエリ修正タイプの分類に貢献していたのに対し、中央部への注目度は、その分類精度に悪影響となった。ある Web ページを訪れたユーザ

は、その意図がどんなものであったとしても、主要な情報が含まれやすいページ中の中央部を多く見るものと考えられる。そのため、Web ページの中央部への注目からは、次の検索クエリ修正の種類を絞り込むことが困難であったものと予想される。

既存研究 [1] において、Generalization と Specialization が交互に起きるという報告があったこともあり、我々は直前のクエリ修正が Generalization であるという情報は、次のクエリ修正が Specialization であるかどうかの推定に有効に働くことを期待していた。しかしながら、この結果からはその傾向は見られなかった。この結果に対する理由は明らかではないが、言語や文化の違いがユーザの検索行動に影響を与えている可能性も考えられる。一方で、今回の結果を前向きにとらえることもできる。直前のクエリ修正タイプに関する素性が分類精度上昇に貢献しないということは、見方を変えれば、その情報が存在しない検索セッションの開始時においても、提案手法は同程度の精度で次の検索クエリの修正タイプを予想可能であるといえる。

ここでは、分類器構成に用いた各素性が、5 種類すべてのクエリ修正タイプの推定にどの程度有用かを検証した。その結果、多くの素性については、5 種類の分類には大きく貢献していないことが分かった。しかしながら、そのような素性の中には、次の検索クエリが特定の修正タイプに属するかどうかの推定には有効に働くものも存在する可能性がある。そこで以降では、各クエリ修正タイプに対する 2 値分類について、提案手法の評価を行う。

5.4 各クラスに対する 2 値分類

1 対他方式を用いることで、ある検索行動に関するデータが与えられた際に、次の検索クエリの修正に対する各クエリ修正タイプの発生尤度を推定することができる。その情報から得られる、次のクエリ修正タイプに関するランキングに対して評価を行った。

5.4.1 分類精度

各クラスに対する分類精度の評価のために、以下に示す手続きを行った。

- (1) 収集したログデータをタイムスタンプ順に前半部と後半部の 2 グループに分割して、一方を訓練データ、他方をテストデータとする。
- (2) (a) 訓練データを用いて 1 対他方式によって分類器を構築する。
(b) テストデータ中の各ログに対して、分類器の推定結果をもとにクエリ修正タイプを順位付けし、得られたランキングを評価する。
- (3) 訓練データとテストデータを入れ替えて、前ステップを同様に行う。

ここで、得られたランキングに含まれる正解は 1 つであるため、その評価尺度としては、MRR (Mean Reciprocal

表 7 予測されたクエリ修正タイプのランキングに対する MRR

Table 7 MRR score for predicted QRT ranking.

クエリ修正タイプ	前半の ログデータ	後半の ログデータ	Ave.
Generalization	0.75	0.46	0.60
Specialization	0.40	0.67	0.54
Parallel Move	0.61	0.55	0.58
Error Correction	0.59	0.64	0.61
Mission Change	0.55	0.48	0.51
Macro Average	0.58	0.56	0.57
Micro Average	0.53	0.54	0.54

Rank) を用いた。

表 7 は、各グループをテストデータとした場合の MRR 値を示している。前半部分のログをテストデータとした場合は、Generalization に対する分類器の精度が最も高く、Specialization の分類結果は最も悪かったのに対し、後半部のログに対するテスト結果では、それとは反対の結果となった。その原因は明らかではないが、グループ間でのクエリ修正タイプの分布が異なることの影響もあると考えられる。実際に、前半のロググループには Error Correction が 13% の割合で含まれているのに対して、後半でのその割合は 2% にすぎなかった。

このような差異はあるものの、どちらのグループにおいても一貫性のある結果も存在する。Error Correction の MRR 値は、全体のマクロ平均およびマイクロ平均の値よりも高くなっている。一方で、Mission Change については、どちらのグループでテストした場合においても、平均値に比べその値が低いことが分かる。そのためこの結果から、Mission Change は他のクエリ修正タイプに比べて、予想を行うことが難しいといえる。対照的に、与えられた検索行動から、次のクエリ修正が Error Correction を推定することは比較的容易であるということはこの結果は示している。このような結果が生じた要因として、それぞれの修正タイプが起こりうる範囲が異なることがあげられる。Mission Change は、検索の対象が現在と大きく異なるクエリ修正がすべて含まれるのに対し、Error Correction は次の検索が現在の検索意図とまったく同じ場合にのみ生じる狭いカテゴリである。これらを考慮すると、精度上昇のためには、クエリ修正タイプごとに分類に貢献しそうな素性を検討し、それぞれに対して分類器を構成する必要があると思われる。

5.4.2 分類貢献度の高い素性

5.3 節で述べたように、ある素性を除くことで分類器の分類精度が低下する場合は、その素性がクエリ修正タイプの分類に貢献していると考えられる。そこで、各クエリ修正タイプに対する、各素性の分類貢献度を調べるために、素性除去による精度比較を再帰的に行った。具体的には、各クエリ修正タイプ $t \in T$ に対して、収集した行

Algorithm 1 IterAblation(F, n_{abl}, D_t)**Input:** 素性集合 F , 除去素性数 n_{abl} ($\leq |F|$), 評価データ集合 D_t **Output:** 除去素性集合 F_{abl} (除去順)

```

1:  $F_{abl} \leftarrow \emptyset$ 
2: while  $|F_{abl}| < n_{abl}$  do
3:   //  $\text{Accuracy}_{SVM}(\cdot)$  は評価データ集合から
4:   // 素性を計算することで構成される SVM の分類精度
5:    $f_{abl} \leftarrow \arg \min_{f \in F} \text{Accuracy}_{SVM}(F \setminus \{f\}, D_t)$ 
6:    $F_{abl} \leftarrow F_{abl} \cup \{f_{abl}\}$ 
7:    $F \leftarrow F \setminus \{f_{abl}\}$ 
8: end while
9: return  $F_{abl}$ 

```

動情報のログデータ集合 D から評価データ集合 D_t を作成し、これと全素性集合 F を用いて、Algorithm 1 に示す IterAblation($F, |F|, D_t$) を適用することで、クエリ修正タイプ t に対する各素性の分類貢献順位を計算した。

評価データ集合 D_t については、 D に含まれるログデータのうち、正解ラベルが t であるものの集合 (正例) と、正解ラベルが $t' \in T \setminus \{t\}$ であるものを正例と同数だけサンプリングした集合 (負例) から構成した。ただし、サンプリングされる負例は一意には定まらないため、今回はランダムサンプリングを 10 回繰り返し、各反復における素性除去結果を平均することで、各素性の貢献順位を算出した。

ここで、2つの素性 f_i と f_j に対して、クエリ修正タイプ t に対する分類貢献度の値に、(f_i の貢献度) $>$ (f_j の貢献度) という関係が成り立つ場合、 f_i の方が t の分類に対する有用性が高いといえる。また、 t に対する貢献度が等しい2つの素性に対して、前者についてはどのクエリ修正タイプにおいても貢献度の値が同程度であり、後者についてはこのタイプの貢献度が他のタイプと比べて高い、という場合には、 t の分類に対する相対的な貢献度という意味では、後者の方が有用と考えられる。そこで本研究では、クエリ修正タイプ t の分類にとりわけ有用である素性は、「 t の分類に対する貢献度が高い」と「全タイプの分類に対する貢献度に比べ、 t の分類に対する貢献度が相対的に高い」という2つの性質を満たすと考える。

1つ目の性質に関する指標 r の値には、上述の素性除去手続きによって得られる、クエリ修正タイプ t に関する素性の貢献順位の順序値を用いた。また、2つ目の性質に関する指標 d を導入するために、素性の各タイプに対する貢献順位を平均することで、全クエリ修正タイプに対する貢献順位を求め、その順序値を r_{whole} とした。そして、 $d = r - r_{\text{whole}}$ によって、2つ目の指標値を計算した。どちらの指標についても、その値が小さいほど、それぞれの性質をよく満たすことになる。また、一方の性質だけを満たす素性よりも、両者の性質をともに満たす素性の方が、特定のクエリ修正タイプに対する有用性が高いと考えられる。そこで今回は、両者の指標値の線形和である $r + d$ を、各クエリ修正タイプに対する素性の有用性を表す指標とし

て採用する。この値が小さい素性ほど、特定のクエリ修正タイプの分類に有用な素性と考えられる。

各クエリ修正タイプに対する有用性が高かった素性 f のうち上位 20 個を、上述の指標値とともに表 8 に示す。この結果から、各クエリ修正タイプの分類に有用な素性いくつかの傾向があることが分かる (それらの素性を表中では太字で示している)。Generalization に関しては、「検索結果のヒットカウント」に関する素性が上位に現れている。検索結果のヒットカウントが 1 件もない、あるいはあまりにも小さいときは、ユーザは検索クエリの内容を緩和することで、適合情報を得ようとすると考えられる。この結果は、そのような状況では Generalization が起きやすいということを示している。「検索クエリ中の単語数」に関する素性は、Specialization の分類に対する貢献度が高かった。この素性は、5.3 節での検証においても、分類貢献度が高いことが分かっていたが、この結果から、特に Specialization の予測に有用であると考えられる。Parallel Move や Mission Change といった修正タイプに関しては、「クリックされた広告の順位」や「検索結果ページ中の広告数」のような広告に関連する素性の分類貢献度が高い傾向にあることが表中より分かる。また、「非検索結果ページにおける周辺部に対する注目度」に関する素性もいくつか出現している。広告情報は、視覚的にも内容的にもユーザの興味をひきつけるように作られることが多いと考えられる。これらの素性の大きさは、広告に対するユーザの興味の高まりを表しており、今回の結果から、そのようなケースでは検索意図の大きな遷移が起きやすくなることが分かる。最後の Error Correction に関しては、5.3 節で全修正タイプの分類に有用であった「検索結果に対するクリック」に関する素性が上位 20 件のうちに含まれていた。この素性や「クエリ推薦数」、「検索結果のヒットカウント」といった素性は、現在のクエリに入力ミスが含まれる際に、その値の変化量が大きくなると考えられる。また、「検索クエリの文字数」という素性も上位に位置していた。長い単語を入力するときには、タイピングミスをしやすくなるため、この素性の Error Correction に対する分類貢献度が高くなったものとして予想される。

6. 考察

6.1 一般的なユーザに対する提案手法の有用性

本稿では、第 1 著者の検索行動を記録したログデータのみを対象とし、提案手法に対する評価実験を行った。そのため得られた実験結果は、被験者の特性に類似する人へのみ、有用な可能性がある。

6.1.1 同様の結果が得られるユーザ層

そこでまず、被験者の特性や検索タスクの性質が検索行動に与える影響について考察を行った。その考察のために、Feildら [10] が既存研究において行ったユーザ実験の

表 8 各クエリ修正タイプの 2 値分類器に対する素性の除去結果. 修正タイプ t について, $r+d$ の値が高い素性ほど, t に対する分類貢献度が高いといえる. 修正タイプごとに傾向が見られた素性については太字で示してある

Table 8 Ablated features for binary classifier of each QRT. A feature that has a high value for $r+d$ in QRT t makes a great contribution to classification of t . Features that have unique characteristics in a certain QRT are indicated in boldface.

Generalization				Specialization				Parallel Move				Mission Change				Error Correction				全体	
f	r	d	$r+d$	f	r	d	$r+d$	f	r	d	$r+d$	f	r	d	$r+d$	f	r	d	$r+d$	f	r_{whole}
f_{48}	28	-33	-5	f_6	4	-23	-19	f_{67}	20	-25	-5	f_{17}	6	-14	-8	f_4	5	-19	-14	f_{13}	1
f_{24}	10	-12	-2	f_{43}	28	-30	-2	f_{56}	19	-22	-3	f_{10}	24	-29	-5	f_9	14	-18	-4	f_{12}	2
f_{12}	1	-1	0	f_{36}	19	-19	0	f_{63}	12	-14	-2	f_{19}	13	-17	-4	f_8	3	-6	-3	f_{22}	3
f_{21}	4	-2	2	f_{65}	14	-14	0	f_{25}	3	-4	-1	f_7	8	-9	-1	f_{14}	1	-3	-2	f_{14}	4
f_{14}	3	-1	2	f_{22}	2	-1	1	f_{24}	11	-11	0	f_{16}	5	-6	-1	f_5	16	-17	-1	f_{23}	5
f_{13}	2	1	3	f_{13}	1	0	1	f_{62}	8	-7	1	f_{12}	1	-1	0	f_{15}	12	-11	1	f_{21}	6
f_{20}	7	-3	4	f_{31}	23	-21	2	f_{22}	2	-1	1	f_{14}	3	-1	2	f_{16}	6	-5	1	f_{25}	7
f_{26}	12	-7	5	f_{41}	10	-8	2	f_{13}	1	0	1	f_{13}	2	1	3	f_{12}	2	0	2	f_{35}	8
f_8	8	-1	7	f_{12}	3	1	4	f_{35}	5	-3	2	f_{20}	7	-3	4	f_2	17	-14	3	f_8	9
f_{23}	6	1	7	f_{23}	5	0	5	f_{23}	4	-1	3	f_{22}	4	1	5	f_{20}	7	-3	4	f_{20}	10
f_{22}	5	2	7	f_{17}	13	-7	6	f_{32}	9	-5	4	f_{30}	19	-10	9	f_{11}	9	-4	5	f_{16}	11
f_{66}	22	-12	10	f_{27}	9	-3	6	f_{38}	14	-7	7	f_{11}	11	-2	9	f_{13}	4	3	7	f_{27}	12
f_{37}	23	-12	11	f_{21}	6	0	6	f_{21}	7	1	8	f_8	9	0	9	f_7	13	-4	9	f_{11}	13
f_{16}	11	0	11	f_{32}	11	-3	8	f_{39}	24	-14	10	f_4	18	-6	12	f_{26}	15	-4	11	f_{32}	14
f_{25}	9	2	11	f_{35}	8	0	8	f_{58}	23	-13	10	f_{25}	10	3	13	f_{22}	8	5	13	f_{62}	15
f_{54}	31	-19	12	f_{62}	12	-3	9	f_{12}	6	4	10	f_{55}	28	-14	14	f_{24}	18	-4	14	f_{36}	16
f_{41}	16	-2	14	f_{14}	7	3	10	f_{47}	29	-17	12	f_{52}	35	-20	15	f_{37}	25	-10	15	f_7	17
f_{32}	14	0	14	f_{15}	17	-6	11	f_{28}	25	-12	13	f_{35}	12	4	16	f_6	21	-6	15	f_{41}	18
f_{18}	33	-18	15	f_{38}	16	-5	11	f_{44}	27	-13	14	f_{36}	17	1	18	f_{23}	10	5	15	f_{26}	19
f_{29}	29	-14	15	f_{59}	31	-18	13	f_{65}	21	-7	14	f_9	26	-6	20	f_{21}	11	5	16	f_{17}	20

表 9 情報検索への慣れや事前知識の違いによる検索タスク中のクエリ数の変化

Table 9 Change of the number of queries in search tasks under the different factors.

	FAMILIAR	UNFAMILIAR	KNOWN	UNKNOWN	全体
平均	2.65	2.10	1.83	2.28	2.23
標準偏差	1.62	1.10	0.92	1.30	1.27

公開ログデータ^{*9}を分析した. この実験は, 検索時におけるユーザのフラストレーション予測のために実施されたものであり, 実験には大学に所属する 30 人が被験者として参加している. 各被験者の専攻は計算機科学, 工学, 運動生理学, 経済学, 文学など多岐にわたっており, 情報検索に馴染みのあるユーザ層 (FAMILIAR) とそうでないユーザ層 (UNFAMILIAR) の両方が参加している. また, 実験で行われた各タスクに対して, ユーザの事前知識の有無 (KNOWN/UNKNOWN) に関する情報が記録されている.

これらの条件によって検索行動に違いがあるのかを確かめるために, 全体および各条件の下での, 検索タスク中の平均クエリ数を調べた. その結果を表 9 に示す. 表より, 情報検索に馴染みのある人の方がタスク中に多くのク

エリを入力することが分かる. また, タスクに関する事前知識がない場合ほど, 多くのクエリがタスク中に入力されている. この実験は国外で行われたものであるため, 日本語での検索にもあてはまるとは必ずしも限らないが, このような検索の慣れやタスクに関する事前知識といった要因は, ユーザの検索行動に少なからず影響を及ぼしているものと考えられる.

今回の実験の被験者である本稿の第 1 著者は, 情報学研究科の大学院に在籍する学生であるため, 一般の検索ユーザに比べて情報検索に関する多くの知識を有していると考えられる. また表 3 にも示したように, 実験で被験者が行ったタスクは, 事前に明確な答えが存在せず, 検索の過程でその答えを見つけるといったものが高い割合を占めていた. そのため, 表 9 の分類の中では, FAMILIAR や UNKNOWN という性質が, 被験者や検索タスクの特性にあてはまるものと考えられる. 実際に 5.1 節で述べたように, 被験者は 1 セッション中に平均して 7 個の検索クエリを入力しており, その値は高い傾向にある.

今回の実験と Feild らの実験とは, 実験設定や被験者の言語圏および文化圏が異なるため, 両者の結果を単純に比較することはできない. しかし, 上述の考察をふまえる

^{*9} <http://ciir.cs.umass.edu/~hfeild/downloads.html>

と、情報検索に不慣れなユーザよりも、馴染み深いユーザを対象とする方が、今回の実験と同様の結果が得られる可能性が高いと考えられる。特に、後者のユーザ層に対して、事前知識の少ない状況下で検索タスクを行ってもらったことで、同様の結果が得られる可能性が高くなると考えられる。

6.1.2 実験結果のユーザ依存性

ここでは4.4節で述べた、素性とユーザ特性の関係性を考慮したうえで、今回の実験から得られた結果のユーザ依存性を考察する。

1対1方式によるクエリ修正タイプ全体に対する分類に関しては、評価の結果、 f_{17} (1つ以上の検索結果がクリックされたか)や f_{48} (非検索結果ページ中での左部に対する注目割合)、 f_6 (クエリに含まれる単語数)、 f_{31} (各非検索結果ページ上での平均滞在時間)といった素性の分類貢献度が高いことが判明した。これらのうち、 f_6 はクエリカテゴリの素性であるため、4.4節で述べたように、そのユーザ依存性は高くないと思われる。また f_{17} は、検索結果ページ中に適合結果が存在するかを示す素性であるため、その値はクエリ依存であり、クリックカテゴリの中では比較的ユーザ依存性の低い部類に入るものと思われる。その一方で、視線カテゴリに属する f_{48} は、ユーザ間での差異が大きくなる素性の1つであると考えられる。また、 f_{31} については、とりうる値の範囲についてはユーザごとで開きがあるものの、滞在時間が長い場合と短い場合とに二分されることが予想される。そのため、 f_{48} に比べるとユーザ依存性は低くなるかもしれない。

そのため、被験者以外のユーザに対しては、 f_6 や f_{17} といった素性は今回同様、分類に貢献することが期待できるが、 f_{48} のような素性についてはユーザごとに有用性が異なる可能性がある。また、 f_{31} に関しては、今回の被験者のように慎重に情報検索を行うユーザに対しては、同様の有用性があるものと思われるが、他のユーザに対しては分類精度に悪影響を及ぼす可能性もある。

1対他方式による各クエリ修正タイプの分類に関する評価では、Generalizationには「検索結果のヒットカウント」、Specializationには「検索クエリ中の単語数」、そして、Error Correctionには「クエリ推薦数」や「検索クエリの文字数」などといった素性が分類に貢献しているという結果が得られた。これらは、先に述べたように、ユーザ依存性の低い素性と考えられるため、他のユーザに対しても、同様の有用性があることが期待できる。その一方で、Parallel MoveやMission Changeといった、ユーザの検索意図に大きな変化の生じやすいクエリ修正タイプに関しては、「クリックされた広告の順位」や「ページ中での周辺部に対する注目」など、ユーザ依存性の高い素性が分類に貢献しているという結果が得られた。そのため、これらのクエリ修正タイプに関する分類精度や素性貢献度は、他のユーザの場合には大きく異なる可能性もある。

表 10 異なる訓練データ数に対する分類精度の比較

Table 10 Classification accuracy among different sampling sizes.

試行回数	訓練データ数				
	30	60	90	120	150
1回目	0.33	0.39	0.39	0.44	0.44
2回目	0.33	0.33	0.33	0.33	0.44
3回目	0.33	0.33	0.44	0.33	0.44
4回目	0.28	0.33	0.33	0.39	0.38
5回目	0.33	0.33	0.44	0.44	0.44
平均	0.32	0.34	0.39	0.39	0.43

今後は、今回の被験者とは特性の異なるユーザ層に対して、同様の実験を行うことで、各素性のユーザ依存性を評価し、提案手法のユーザ間での適用可能範囲について考察する必要がある。

6.2 分類に必要な訓練データの数

5.3節で述べたように、全クエリ修正タイプに対する提案手法の分類精度は0.410であり、ベースライン2つのそれを上回る結果となった。しかしながら、DOM手法との差は大きいものではなかった。考えられる原因のうち、ログデータ数の少なさの影響度に関する検証を行った。具体的には、収集するログデータ数の増加による分類精度向上の有無を確かめることを目的として、異なる訓練データ数(30から150まで30刻み)に対する分類器の精度を比較した。その結果を表10に示す。なお、サンプリングデータの選び方によって、分類精度が大きく異なる可能性があるため、各サンプルサイズに対して、サンプリングを5回繰り返し、それぞれに対して精度を計算した。

この結果から、サンプルデータによって多少のばらつきはあるものの、一般的にはサンプルサイズが大きいときほど、分類精度が高いことが分かる。また、少なくとも150までのサンプリングでは、訓練データ数の増加に対する分類精度の収束は見受けられない。これらより、ログデータのサイズを増やすことで、分類精度の上昇を期待できる。そこで、今後も行動情報の記録を続けることで、精度変化の検証を行っていきたい。

6.3 現在以前の検索の影響

本稿では、次に起こる検索でのクエリ修正タイプの予測を実現するにあたって、現在の検索におけるユーザの行動情報に着目した。すなわち、現在以前の行動情報については考慮しておらず、ユーザのクエリ修正に対する行動を1階マルコフ過程でモデル化しているといえる[7]。しかし、評価実験における分類精度の結果からも分かるように、このモデルはクエリ修正タイプの予測という問題に対しては、制約が少し強すぎると考えられる。たとえば、検索中に自分の知らない単語に出会い、その意味を調べるとい

うケースを考える。この場合、ユーザはその単語の意味を知った後で、もとの検索タスクに戻ってくるのが予想される。そのため、より正確なクエリ修正タイプの予想のためには、現在の検索以前の情報も考慮する必要がある。このことに配慮した手法の改善案として、検索セッションに関する素性を追加することや、分類器の構築にSVMでなくCRF (Conditional Random Field) [21]を用いることなどが考えられる。

6.4 検索行動を利用した後続クエリの予測可能性

本稿では、ユーザの検索行動を用いることで、次の検索におけるクエリ修正タイプの予測に取り組んだ。しかしこの内容は、後続する検索を支援するための第一歩といえる。たとえば、検索クエリが入力される前に次の検索に関する支援を行うためには、次のクエリ修正タイプだけでなく、検索クエリについても予測する必要がある。

そこでまず、ユーザが閲覧したページに由来する検索がどの程度存在するのかを調べるために、対象とする検索ログの条件を、全体 (条件なし)、正解の語が閲覧ページ中に含まれているもの (条件1)、正解の語をユーザが実際に注目しているもの (条件2) と変化させた際の、クエリ修正タイプの分布を計算した。ここで正解とは、ユーザが次に入力する検索クエリに含まれる単語のうち、現在の検索クエリ中には含まれない語の集合である。その結果を表 11 に示す。この表より、今回収集した行動情報のログデータのうち、131/183 ≃ 72%の検索ログに関しては、次に入力されるクエリに新しく含まれる単語が、現在の検索時に閲覧されたページ中に存在していることが分かる。さらに、94/183 ≃ 51%の検索ログに関しては、ページ中に出現するその語をユーザが実際に注目していることが分かる。これらの結果は、現在の検索におけるユーザの行動が、次の検索に与える影響が大きいことを示唆している。

表中において、クエリ修正タイプごとの変化に注目すると、Specialization については他のタイプに比べて、条件1や条件2を満たすログ数が、すべてのSpecializationに関する検索ログ数と大きく変化していないことが分かる。この結果から、ユーザが次の検索において情報を絞り込む際には、現在の検索中に閲覧したページに出現していた単語

を用いる事例が多いものと考えられる。

次にこれらの各条件の下で、閲覧ページ中でのユーザの注目単語を用いることで、次に入力される検索クエリがどの程度予測可能であるのかを検証した。その方法を以下に示す。

- (1) 各検索における行動情報のログデータから、閲覧ページにおけるユーザの注目語を文献 [23] の手法によって抽出し、注目度によって順位付けする (GAZE 手法)。
- (2) 得られたランキングを、MAP (Mean Average Precision) によって評価する。

ここで、MAP 値を計算するにあたっては、次の検索でのクエリに含まれる単語のうち、現在の検索クエリには含まれないものを正解とした。GAZE 手法の比較として、語の出現頻度を利用したランキングも行い (TF 手法)、それぞれの手法の評価を行った。その結果を表 12 に示す。

対象とする問題の難しさのため、どちらの手法に関してもMAP 値は決して高くない。しかし、どの条件においても、Specialization と Error Correction に関しては、GAZE 手法によって推定精度が大きく改善されていることが分かる。先に述べたように、次のクエリ修正タイプがSpecialization である場合は、ユーザが現在の検索において、次の検索クエリに利用する単語を注目している確率が高いといえる。そのため、GAZE 手法によって比較的高い精度で、その推定が可能になったものと思われる。Error Correction については、表 11 より、注目語がクエリ修正に利用される頻度はあまり高くないといえる。それにもかかわらず、GAZE 手法による推定精度が高くなった理由について、以下で考察を行った。

Error Correction に分類されるクエリ修正タイプが行われる際のユーザの状況の例として、次の2つが存在する。

- 検索ボタンを押した直後に、クエリ中にタイプミスが存在することに気づき、即座にクエリ修正を行う。
- 検索実行直後は、クエリのタイプミスに気付かず (あるいは、適切な言い換え表現を思いつかず)、ページを閲覧する過程において修正候補を発見する。

1つ目の例の場合、ユーザはクエリ候補語を注目することなく、次のクエリ修正を行うため、GAZE 手法では適切な推定が行えない。一方で、2つ目の例の場合、候補語

表 11 対象の検索ログを変化させた際のクエリ修正タイプの分布

Table 11 QRT distribution under the different conditions.

	条件なし	条件1	条件2
Generalization	24 (13.1%)	7 (5.3%)	4 (4.3%)
Specialization	67 (36.6%)	61 (46.6%)	46 (48.9%)
Parallel Move	54 (29.5%)	42 (32.1%)	29 (30.9%)
Error Correction	14 (7.7%)	7 (5.3%)	4 (4.3%)
Mission Change	24 (13.1%)	14 (10.7%)	11 (11.7%)
総数	183	131	94

表 12 次の検索クエリの推定ランキングに対する MAP 値
Table 12 MAP score for predicted subsequent queries.

		Genera- lization	Specia- lization	Parallel Move	Error Correction	Mission Change
条件なし	TF	0.03	0.04	0.03	0.04	0.01
	GAZE	0.02	0.09	0.03	0.10	0.01
条件1	TF	0.09	0.04	0.03	0.09	0.02
	GAZE	0.09	0.10	0.04	0.19	0.02
条件2	TF	0.15	0.05	0.04	0.12	0.02
	GAZE	0.15	0.14	0.06	0.34	0.03

がユーザの目にとまることが予想されるため、GAZE手法によって高い精度で次の検索クエリ推定を行うことが可能になる。ここで条件2を満たすのは、2つ目の状況の場合のみであるため、その条件の下では、GAZE手法によるError Correctionタイプのクエリ推定が有効に働いたものと思われる。

しかしながら、今回検証に用いた検索ログの数は少ないため、検索行動の収集を継続し、より大規模なデータに対して、再度検証を行う必要がある。

7. まとめ

本稿では、Web検索時における5種類のクエリ修正を推定する手法を提案した。提案手法は、現在の検索におけるユーザの検索行動をもとに、機械学習によって分類器を構成することで、その推定を行う。我々は、実際のユーザの検索行動を記録することで分類器を構成し、提案手法の評価を行った。評価実験の結果から、5種類の修正タイプに対して提案手法は約40%の精度でその予測が可能であることが明らかになった。また、各修正タイプに対して、「ページ中の周辺領域への注目」や「広告情報への注目」、「検索クエリの長さ」などに関するものが、その推定に有用であることが分かった。しかしながら、実験で用いたログデータの規模は、一般的な行動分析で用いられているものと比較すると少ないといえる。

そこで今後は、行動ログの収集数を増やすことで、クエリ修正タイプの分類に必要な訓練データの数に関する検証を行う必要がある。また、今回の手法で機械学習に用いた素性は、クエリの単語数やページ閲覧時間、ページ中の注目位置など、計算は容易ではあるが検索時の表層的な行動情報に関するものが多かった。しかしながら、本研究において情報源の1つとして利用しているユーザの視線情報からは、注目語とクエリの関係や、ページを読む速度、視線の動きのパターンなど、より意味的な特徴量も計算することができる。そこで、こうした特徴量の追加による分類精度の向上に努めていく予定である。さらに、クエリ修正タイプの予測に用いる各素性が、ユーザ依存であるのか非依存であるのかを評価することで、提案手法の適用範囲についても詳しく分析していきたい。

また今回は、既存のカテゴリ [1] をクエリ修正タイプの分類クラスとして用いたが、各タイプの境界部分の曖昧性などが原因で、ラベリングを行うものが難しい検索ログも存在した。そのため、より適切なクエリ修正のカテゴリについても、今後検討の余地がある。

謝辞 本研究の一部は、文科省科研費基盤研究(A)「ウェブ検索の意図検出と多元的検索意図指標にもとづく検索方式の研究」(研究代表者: 田中克己, 課題番号: 24240013), 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, 課題番号: 23680006), 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, 課題番号: 22650018)によるものです。ここに記して謝意を表します。

参考文献

- [1] Boldi, P., Bonchi, F., Castillo, C. and Vigna, S.: From “Dango” to “Japanese Cakes”: Query Reformulation Models and Patterns, *Proc. 2009 IEEE/WIC/ACM International Conference on Web Intelligence*, pp.183-190 (2009).
- [2] Broder, A.: A taxonomy of web search, *SIGIR Forum*, Vol.36, No.2, pp.3-10 (2002).
- [3] Buscher, G., Cutrell, E. and Morris, M.R.: What do you see when you're surfing?: Using eye tracking to predict salient regions of web pages, *Proc. 27th International Conference on Human Factors in Computing Systems*, pp.21-30 (2009).
- [4] Buscher, G., van Elst, L. and Dengel, A.: Segment-level display time as implicit feedback: A comparison to eye tracking, *Proc. 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.67-74 (2009).
- [5] Chen, M.C., Anderson, J.R. and Sohn, M.H.: What can a mouse cursor tell us more?: Correlation of eye/mouse movements on web browsing, *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, pp.281-282 (2001).
- [6] Cheng, Z., Gao, B. and Liu, T.-Y.: Actively predicting diverse search intent from user browsing behaviors, *Proc. 19th International Conference on World Wide Web*, pp.221-230 (2010).
- [7] Chierichetti, F., Kumar, R., Raghavan, P. and Sarlos, T.: Are web users really Markovian?, *Proc. 21st International Conference on World Wide Web*, pp.609-618 (2012).
- [8] Cutrell, E. and Guan, Z.: What are you looking for?: An eye-tracking study of information usage in web search, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.407-416 (2007).
- [9] Duan, K.-B. and Keerthi, S.S.: Which Is the Best Multi-class SVM Method? An Empirical Study, *Multiple Classifier Systems*, Vol.3541, pp.732-760 (2005).
- [10] Feild, H.A., Allan, J. and Jones, R.: Predicting searcher frustration, *Proc. 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.34-41 (2010).
- [11] Guo, Q. and Agichtein, E.: Ready to buy or just browsing?: Detecting web searcher goals from interaction data, *Proc. 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.130-137 (2010).
- [12] Guo, Q. and Agichtein, E.: Towards predicting web searcher gaze position from mouse movements, *Proc. 28th International Conference Extended Abstracts on Human Factors in Computing Systems*, pp.3601-3606 (2010).
- [13] Guo, Q. and Agichtein, E.: Beyond dwell time: Estimating document relevance from cursor movements and other post-click searcher behavior, *Proc. 21st International Conference on World Wide Web*, pp.569-578 (2012).
- [14] Guo, Q., Lagun, D. and Agichtein, E.: Predicting web

- search success with fine-grained interaction data, *Proc. 21st ACM International Conference on Information and Knowledge Management*, pp.2050-2054 (2012).
- [15] Guo, Q., White, R.W., Zhang, Y., Anderson, B. and Dumais, S.T.: Why searchers switch: Understanding and predicting engine switching rationales, *Proc. 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.335-344 (2011).
- [16] Hassan, A., Jones, R. and Klinkner, K.L.: Beyond DCG: User behavior as a predictor of a successful search, *Proc. 3rd ACM International Conference on Web Search and Data Mining*, pp.221-230 (2010).
- [17] Hu, B., Zhang, Y., Chen, W., Wang, G. and Yang, Q.: Characterizing search intent diversity into click models, *Proc. 20th International Conference on World Wide Web*, pp.17-26 (2011).
- [18] Huang, J. and Efthimiadis, E.N.: Analyzing and evaluating query reformulation strategies in web search logs, *Proc. 18th ACM Conference on Information and Knowledge Management*, pp.77-86 (2009).
- [19] Huang, J., White, R.W. and Dumais, S.: No clicks, no problem: Using cursor movements to understand and improve search, *Proc. 2011 Annual Conference on Human Factors in Computing Systems*, pp.1225-1234 (2011).
- [20] Joachims, T.: Optimizing search engines using click-through data, *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.133-142 (2002).
- [21] Lafferty, J.D., McCallum, A. and Pereira, F.C.N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proc. 18th International Conference on Machine Learning*, pp.282-289 (2001).
- [22] Rodden, K., Fu, X., Aula, A. and Spiro, I.: Eye-mouse coordination patterns on web search results pages, *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, pp.2997-3002 (2008).
- [23] Umemoto, K., Yamamoto, T., Nakamura, S. and Tanaka, K.: Search intent estimation from user's eye movements for supporting information seeking, *Proc. International Working Conference on Advanced Visual Interfaces*, pp.349-356 (2012).
- [24] White, R.W. and Dumais, S.T.: Characterizing and predicting search engine switching behavior, *Proc. 18th ACM Conference on Information and Knowledge Management*, pp.87-96 (2009).
- [25] Xu, S., Jiang, H. and Lau, F.C.: User-oriented document summarization through vision-based eye-tracking, *Proc. 14th International Conference on Intelligent User Interfaces*, pp.7-16 (2009).



梅本 和俊

京都大学大学院情報学研究科博士後期課程在学中。日本学術振興会特別研究員 (DC1)。2013年京都大学大学院情報学研究科修士課程修了。主に情報検索におけるユーザ行動の分析と応用に関する研究に従事。日本データベース

学会学生会員。



中村 聡史 (正会員)

1976年生。2004年大阪大学大学院工学研究科博士後期課程修了。同年独立行政法人情報通信研究機構専攻研究員。2006年京都大学大学院情報学研究科特任助手, 2009年同特定准教授, 2013年明治大学総合数理学部准教授,

現在に至る。博士 (工学)。サーチとインタラクションや、情報曖昧化技術、ソーシャルアノテーション分析等の研究活動に従事。ヒューマンインタフェース学会等各会員。



山本 岳洋 (正会員)

1984年生。2011年京都大学大学院情報学研究科博士後期課程修了。同年日本学術振興会特別研究員 (PD), 2012年京都大学大学院情報学研究科特定研究員, 現在に至る。博士 (情報学)。

主に情報検索, 特に情報検索におけるユーザインタラクションに関する研究に従事。日本データベース学会会員。



田中 克己 (フェロー)

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。博士 (工学)。主にデータベース, マルチメディアコンテンツ処理, ウェブ検索の研究に従事。

IEEE Computer Society, ACM, 人工知能学会, 日本ソフトウェア科学会, 日本データベース学会各会員。

(担当編集委員 鈴木 伸崇)