

# インタラクション・シーケンスに着目したユーザの商品検索の目的の抽出に関する検討と実装

ワ コラ<sup>†</sup> 川村 隆浩<sup>†</sup> 中川 博之<sup>†</sup>  
田原 康之<sup>†</sup> 大須賀 昭彦<sup>†</sup>

ユーザの検索目的、検索結果、および検索行動の関係を明らかにすることは、Web 検索の有効性を向上させるために非常に重要である。最近のユーザの検索行動マイニングに関する研究は、主にユーザの検索行動履歴やサーバ側のログなどに焦点を当ててきた。また、ブラウザ側からのユーザの検索結果に対するユーザのインタラクションに注目した研究もあったが、インタラクション間の関係を考慮する研究はまだほとんど行われていない。そこで、我々は検索エンジンの検索結果に対するユーザのインタラクション・シーケンスに着目して、ユーザの商品検索目的（購入目的、情報収集目的か）を抽出する手法を提案する。本手法は、ユーザのクリック動作(Click)、スクロール動作(Scroll)、マウスオーバー動作(Hover)と読む動作(Read)の順序と時間を利用し、ユーザの商品検索の目的を抽出するためのユーザの行動シーケンスモデルを用いた。実験により、本研究の手法は先行研究の手法より 10%以上の高精度の結果が得られ、本手法の有効性を示した。

## Detecting User Product Search Intention from Sequential Behavior of User Interaction

KORA VA<sup>†</sup> TAKAHIRO KAWAMURA<sup>†</sup> HIROYUKI NAKAGAWA<sup>†</sup>  
YASUYUKI TAHARA<sup>†</sup> AKIHIKO OHSUGA<sup>†</sup>

The relationship between user searching behavior and search result in Web search system is very important to improve the effectiveness of search engine. There are various studies on mining user search behavior which focused on server-side log or user searching history. Furthermore, there are also many researches which are interested in user's interaction with search result on web browser. But most of them lack of consideration of the relationship between the user's interactions within a session. Therefore, in our proposed method, we take advantage of interaction sequence of user with search engine results page in order to detect user intention(purchase or research) of product searching within the session. Our method uses primarily the sequences of user-click behavior (Click), scrolling behavior (Scroll), mouse-over behavior(Hover) and read behavior (Read) for interaction features. Using our method, we could extract effectively the intention of each query within the session of the user's product searching. The experimental results showed that our proposed method performed more than 10% better than the previous method .

### 1. はじめに

近年、インターネットの急速な発展を背景に、ビジネス販促手法として使われてきた従来の広告手法も変わりつつある。これまでは新聞、雑誌やテレビといったメディアが広告手段として使用されてきたが、最近ではそれらに加え、インターネットも広告発信メディアとして使用されるようになってきている。広告の配信方法には様々な形式があるが、その中でも Web 検索結果としてのスポンサーリンクが最も使用されている。その主な理由として、広告主側は、自分がターゲットとしたユーザに効果的に広告発信できる利点が挙げられる。また、検索ユーザ側は、入手したい情報を

効率的に取得できるメリットがある。しかし、その一方でデメリットもある。例えば、ユーザの検索目的と異なる情報を過度に提示してしまうことにより、ユーザがスポンサーリンクを無視するようになる恐れがある。また、検索結果のスポンサーリンクは不要な情報であると、検索ユーザに感じさせてしまう危険性も考えられる[1]。

以上のような状況を防ぐために、Web 検索やユーザの行動から、ユーザの目的や嗜好を抽出する手法が多く研究されている。幅広い範囲のユーザの嗜好を抽出する研究[2,12,15]もあれば、目的を限定してユーザの嗜好を抽出する研究もある[1,6,8]。特に[1]では、ユーザの Web 検索インタラクションに注目し、ユーザの目的(buy or search)を抽出する。[1]では Web 検索エンジン(Google)の検索結果に対するユーザのインタラ

<sup>†</sup> 電気通信大学 大学院情報システム学研究科  
University of Electro-Communication

クシオンの他に、クエリに関する特徴量、検索結果の質などの6つのグループの特徴量が使用されている。だが、特徴量として用いる検索結果に対するユーザのインタラクシオン間の関係が考慮されていない。一般的にユーザは、検索結果に対して連続的にインタラクシオンを起こすと考えられる。この事実を踏まえ、我々はユーザのインタラクシオン間の関係を考慮した上で、インタラクシオン・シーケンスに着目したユーザの商品検索目的（情報収集目的か、購入目的か）を抽出する手法を提案する。

## 2. 本研究のシステム

本研究は、検索ユーザが行うインタラクシオンを取得するために、独自に開発した Firefox ブラウザに実装したアドオンを使用する。また、ユーザの検索目的を抽出するために、機械学習のツールとして研究者に普及している Weka に実装された SVM(Support Vector Machine)を利用する。

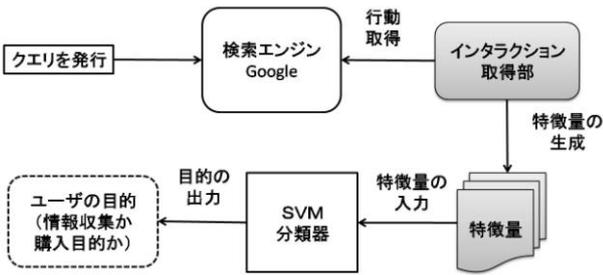


図1 本研究のシステムの構成

まず、ユーザがインストールした Firefox のアドオンをキャプチャー開始状態にすると、システムは新たなタブを開いて自動的に Google の検索ページに移動する。その後、ユーザが好きなキーワードで商品の検索を行うと、特徴量の取得部（本研究のアドオン）がユーザの検索結果ページに対するインタラクシオンを取得する。そして取得した特徴を特徴量としてファイルに書き出す。また、書き出された特徴量を SVM 分類器に入力すると、出力としてユーザの検索目的（購入目的か、情報収集目的か）が出力される。

アドオンのインストール後に、Tools メニューに 1つのメニュー(Capture Interaction)が追加される。ユーザは、Google 検索エンジンに対するインタラクシオンをキャプチャーしたい場合、メニューの Start Session を選択して、セッションを開始する。また、セッション終了時に同じメニューから Finish Session を選択すると取得対象とする全ての特徴量がファイルに書き出される。

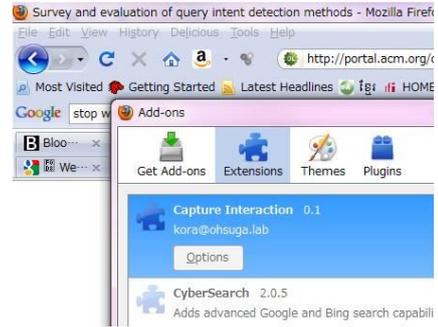


図2 アドオンをインストール後の状態

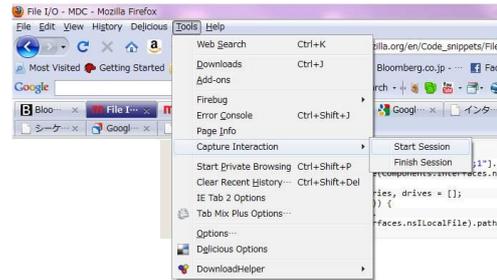


図3 本研究のアドオンの使い方

## 3. 検索目的の抽出手法

### 3.1 セッションの定義

本手法の目的は、ユーザの一連の検索行為における各クエリに潜在する目的、つまり購入が目的であるのか、あるいは情報収集が目的であるのかを抽出することである。この一連の検索行為をここではセッションと呼ぶ。セッションに関する考え方は、研究の目的によって様々であるが、本研究では、1つのセッションには複数のクエリから構成され、各セッションには1つの目的があるものとする。セッション内の各クエリの目的は、共通してそのセッションの目的が割り当てられ、新しいセッションが始まるのは、ユーザが明示的に Start Session をクリックした時点である。

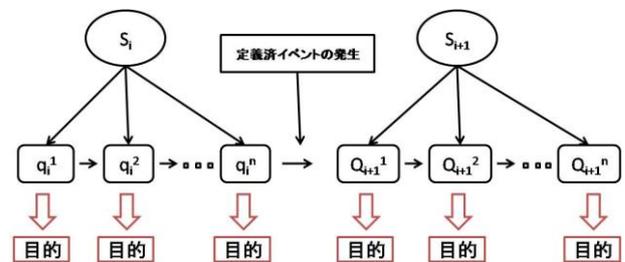


図4 セッションの定義

### 3.2 シーケンシャル・インタラクシオンモデル

通常、検索ユーザは検索結果のページに対して、様々なインタラクシオンを起こすため、そのパターン

は多様なものとなる。本手法では、ユーザが行う一連のインタラクション（インタラクション・シーケンス）を用いるため、それをモデル化し、明確に考えなければならない。したがって、以下で本研究におけるユーザのインタラクションモデルについて詳細に述べる。

ユーザは検索エンジンの結果ページに対して連続的にインタラクションを起こすと考えられる。また、検索結果に対して、ユーザが必ず行うインタラクションとしてはクリック(Click)、スクロール(Scroll)、マウスオーバー(MouseOver)と読む(Read)という動作がある。そのため、我々はこの4つの動作を軸にシーケンシャルなインタラクションモデルを提案する。

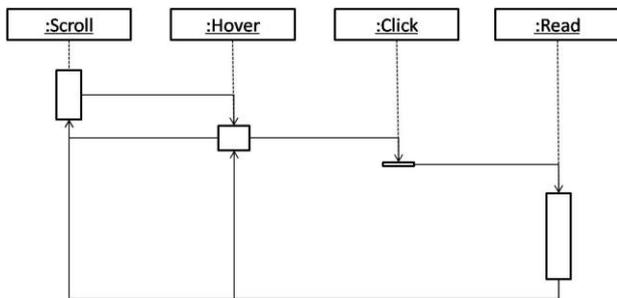


図5 シーケンシャル・インタラクションモデル

図5の横長矩形（横棒）はユーザのインタラクション動作ノードを、縦長矩形（縦棒）はそれぞれの動作に対する時間または回数（Clickの場合）を表す。また、スクロール動作は連続スクロールである可能性もあり、その場合縦棒はスクロールの回数または連続スクロールの時間を表すことになる。野線は、それぞれのノードから動作が始まる場合の特徴量を考えるために必要である。例えば、Click動作は必ずHover動作を介して発生するが、特徴量としては、ClickThenReadMoreThan5s（クリックしてから5秒以上にページの内容を読む場合）といったものも考えられる。同様にRead動作には必ずClickが必要であり、HoverまたはScrollにも必ずReadが必要である。また、ユーザは検索結果の各項目に上にスクロールしながら検索結果を閲覧する場合もあるので、HoverとScrollは相互に動作を発生させることが可能である。ただし、図5ではReadから矢印が上方向に向かっているが、これはRead動作からHoverまたはScrollに遷移するものであり、時間が戻ることを意味するものではない。

### 3.3 特徴量

本研究では検索結果に対するユーザのインタラクシ

ョン・シーケンスに注目する。ユーザが行った操作行動の直前または直後のインタラクションまで考慮することが、本研究で提案する特徴量の独自性と言える。

表1 本研究の特徴量の一部

特徴量	値
Is_ScrollUpandDown_HoverLessThan2s_ClickSERP	{ "T", "F" }
Is_ScrollUpandDown_HoverLessThan2s_ClickRK	{ "T", "F" }
Is_ScrollUpandDown_HoverLessThan2s_ClickNext	{ "T", "F" }
Is_ScrollUpandDown_ReadSERPMoreThan5_ClickSM	{ "T", "F" }
Is_ClickECsites_ReadMoreThan30s	{ "T", "F" }
Is_ClickPCsites_ReadMoreThan30s	{ "T", "F" }
Is_ClickNextOfSERP_ClickOnSERP	{ "T", "F" }
NumOf_LastScroll_ReadSERPLessThan5_ClickSERP	{ 0, ..., N }
NumOf_LastScroll_HoverAdsMoreThan5s_SERPClick	{ 0, ..., N }
NumOf_LastScroll_HoverSERPMoreThan5s_ClickSERP	{ 0, ..., N }
NumOf_ScrollUpandDown_ClickLast5ofSERP	{ 0, ..., N }
NumOf_ClickSERP_ReadMoreThan30s	{ 0, ..., N }
NumOf_ScrollUpAndDown_Click_ReadMoreThan15s	{ 0, ..., N }
NumOf_ReadTillTheBottomOfPage_ClickOnLink	{ 0, ..., N }
AvgOf_ReadTimeOnDetailPage	{ 0, ..., N }
AvgOf_MousemoveOnDetailPage	{ 0, ..., N }
AvgOf_ClickOnDetailPage	{ 0, ..., N }
MaxOf_ReadTimeOnDetailPage	{ 0, ..., N }
MaxOf_MousemoveOnDetailPage	{ 0, ..., N }
MaxOf_ClickOnDetailPage	{ 0, ..., N }
その他の特徴量	20 個
合計	40 個

各特徴量はそれぞれ Bool 値か整数値の値をとる。記号 "\_" は、動作の遷移を意味する。SERP(Search Engine Result Page)は検索エンジンの結果ページ、RK(Related Keywords)は検索エンジンが提供する関連キーワード、Next は検索エンジンの Next ボタン、SM(Side Menu)は検索エンジンの左側のカテゴリ検索の各メニューリスト、Ads(Advertisement)は、検索エンジンのスポンサーリンク、s(Seconds)は時間の単位、ECsites(E-Commerce Sites)は EC サイト、PCsites(Price Comparison Sites)商品の価格比較サイトを表す。

## 4. 評価実験

### 4.1 実験方法と結果

本手法の有効性を評価するために、7人の被験者に、

本研究で実装したアドオンを Firefox ブラウザにインストールしてもらい、情報収集目的の検索を2セッション、購入目的検索を2セッションの計4のセッションの検索を実施してもらった。購入目的のセッションと情報集目的のセッションの検索順序は被験者が決めることとし、被験者には実験後、それぞれのセッションの目的をラベルとして記述させた。その後、提案手法が推測する目的と、被験者の目的とがどの程度一致するかを、正解率、適合率、再現率の観点から評価した(表3)。

表2 実験データ

クエリの数 (47)	目的
29 クエリ	購入目的
18 クエリ	情報収集目的

表3 実験結果

手法	正解率	情報収集目的		購入目的	
	Acc.	Prec.	Recall	Prec.	Recall
本手法	<b>74.46</b>	<b>71.4</b>	<b>55.6</b>	<b>75.8</b>	<b>86.2</b>
既存手法[1]	61.70	50	38.9	66.7	75.9

#### 4.2 考察

表3の結果を見ると、すべての項目(正解率、適合率、再現率)において、本手法の方が高い精度を出力できていることが分かる。特に、購入目的の再現率では非常に高い数値が得られた。これは、提案手法が購入目的の検索を同定することに対して有効であることを示している。その一方で、先行研究の結果より10%以上の精度の向上を得られた。これは、提案手法が商品検索の目的を抽出することにおいても有効であることを示している。

#### 5. おわりに

本稿では、Web 検索エンジンにおけるユーザのインタラクシオン・シーケンスに着目したユーザの商品検索目的(購入目的か、情報収集目的か)抽出手法を提案した。また、実験により、提案手法は先行研究の手法より10%以上の高い精度を示しており、本手法が有効であることが示された。

今後、我々は本手法をECサイトの商品検索システムに適用して結果を検証する予定である。また、特徴量を追加することで、さらに高い精度の抽出結果を挙げることも今後の課題である。

#### 参考文献

- 1) Q. Gou, E. Agichtein: Ready to Buy or Just Browsing? Detecting Web Searcher Goals from Interaction Data. In Proc. of SIGIR (2010)
- 2) R. W. White, P. Baily, L. Chen: Predicting User Interests from Contextual Information. In Proc. of SIGIR( 2009)
- 3) E. Agichtein : Recent Progress on Inferring Web Searcher Intent. Tutorial, WWW(2010)
- 4) A. Broder: A Taxonomy of Web Search. SIGIR Forum, Vol. 36, No.2, pp.3-10(2002)
- 5) Claypool, M. et al: Inferring User Interest. IEEE Internet Computing, pp. 32-39(2001)
- 6) Z. Cheng, B. Gao, Tie-Yan Liu: Actively Predicting Diverse Search Intent from User Browsing Behaviors. In Proc. of WWW(2010)
- 7) S. Fox, K.Karnawat, M. Mydland: Evaluating Implicit Measures to Improve Web Search. ACM Transactions on Information System. pp. 147-168(2005)
- 8) E. Agichtein, E. Brill, S. Dumais, R. Ragno. Learning User Interaction Models for Predicting Web Search Result Preferences. In Proc. of SIGIR(2006)
- 9) G Dupret,
- 10) V. Murdock, B. Piwowarski: Web Search Engine Evaluation Using Clickthrough Data and a User Model. In Proc. of WWW(2007)
- 11) N. Craswell, O. Zoeter, M. Taylor, B. Ramsey: An Experimental Comparison of Click Position-Bias Models. ACM, WSDM(2008)
- 12) J. Teevan, S. T. Dumais, D. J. Liebling: To Personalize or Not to Personalize. In Proc. of SIGIR(2008)
- 13) M. Holub, M. Bielikova: Estimation of User Interest in Visited Web Page. In Proc. of WWW(2010)
- 14) Y. Liu, J. Bian, E. Agichtein: Predicting Information Seeker Satisfaction in Community Question Answering. In Proc. Of SIGIR(2008)
- 15) <http://www.cs.waikato.ac.nz/ml/weka/>
- 16) Terai, H. et al: Differences Between Informational and Transactional Tasks in Information Seeking on the Web. In Proc. Of IIX, pp. 152-259(2008)
- 17) Qiu, F. & Cho, J. : Automatic Identification of User Interest for Personalized Search. In Proc. Of WWW(2006)