

インタラクティブシーケンスに着目した商品検索目的抽出エージェントの開発

著者	川村 隆浩, ワ コラ , 中川 博之, 田原 康之, 大須賀 昭彦
雑誌名	電子情報通信学会論文誌. D, 情報・システム
巻	J94-D
号	11
ページ	1783-1790
発行年	2011-11-01
URL	http://id.nii.ac.jp/1438/00009039/

インタラクションシーケンスに着目した商品検索目的抽出エージェントの開発

川村 隆浩^{†,††} ワ コラ[†] 中川 博之[†] 田原 康之[†]
大須賀昭彦[†]

Development of Web Agent to Extract Product Searching Intention according to Interaction Sequence

Takahiro KAWAMURA^{†,††}, Kora VA[†], Hiroyuki NAKAGAWA[†], Yasuyuki TAHARA[†], and Akihiko OHSUGA[†]

あらまし e-Commerceの急速な発展を背景に、Web検索結果への検索連動型広告（スポンサーリンクなど）が広告配信手段として注目されている。しかし、ユーザの検索意図と異なる広告の配信は、長期的に見てユーザの不快感を招き、逆効果となるリスクがある。そこで、ユーザの検索行動から検索の目的を推測し、適切な情報を提供することが重要となってきている。これまでも検索結果に対するユーザのブラウザ側の操作（インタラクション）に注目した研究は行われているが、一連のインタラクションの前後関係を考慮した研究は行われていない。そこで、本研究では商品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、検索の目的を購入目的か、情報収集目的かに分類するエージェントを提案する。実験により、先行手法と提案手法を組み合わせることにより、商品検索目的の抽出精度を10%以上向上できることを確認した。

キーワード 検索目的, ユーザ行動, 目的抽出, ユーザインタラクション, 行動モデル

1. ま え が き

近年、e-Commerceの急速な発展を背景に、Web上に様々な広告手法が生まれている。その中でもスポンサーリンクなどと呼ばれる検索連動型広告（検索エンジンで検索したキーワードに関連した広告を検索結果画面に表示する広告）の認知度は高く、ある調査[20]によればクリックしたことのある人は約70%に及んでいる。普及が進んだ主な理由としては、広告配信側はターゲットとしたいユーザを絞って広告を表示でき、ユーザ側は入手したい商品の情報などに簡単にたどりつけるといったメリットが挙げられるだろう。しかし、同時にユーザの検索目的と異なる広告を配信することでスポンサーリンクを無視するようになったり、過度

に表示することで不快に感じてしまうリスクもある。現に上記調査ではクリック率の伸び悩みが指摘されている。そのため、広告配信側にとって、ユーザの検索目的を推測して適切な情報を提供するなど、より興味を引くことが重要となってきている。

しかし、ユーザは様々な目的でWeb検索を行う。例えば、話題のテレビ番組の放送日時を調べたり、訪問先の会社の場所を確認したり、オンラインでチケットを予約する場合などである。そのため、ユーザの検索行動から、そうした目的をどのような場合においても抽出できるようにすることは非常に困難である。そこで、本研究では広告配信側にとって最も直接的ともいえる商品名での検索に着目した。商品検索の場合、ユーザは大きく二つの目的をもっていると推測できる。一つは、単にその商品に関する情報収集である（以下、情報収集目的と呼ぶ）。もう一つは、特定の商品を購入することを目的としている場合である（以下、購入目的と呼ぶ）。例えば、“iPad”で検索する場合、ユーザの目的はiPadを含むタブレットPCに関する情報を収集することか、Apple社のiPadを購入すること

[†] 電気通信大学大学院情報システム学研究科, 調布市
Graduate School of Information Systems, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, 182-8585 Japan

^{††} (株) 東芝研究開発センター, 川崎市
Research & Development Center Toshiba Corp., 1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki-shi, 212-8582 Japan

か、の二つに分類できると仮定する。そして、前者であれば、Android など iPad 以外のタブレット PC に関する最新機種などの広告も積極的に出すことが有効と考えられるが、後者であれば、むしろ iPad の安売り情報や iPad 用の周辺機器、サブライ商品の広告を出す方が効果的であるだろう。そのため、こうしたユーザの検索目的が抽出できれば、スポンサーリンクの動的な再配信や、直後のセッションのスポンサーリンクの調整に利用でき、EC サイト・検索サイトとしてはリンク枠を広告主に売りやすい、クリック課金型や保証型として売っているならば少しでもクリック数を上げるのに利用できるなどの理由から有用である。そこで、本研究では広告配信側のメーカや EC サイトに代わって、ユーザ（消費者）が商品の検索を行う場合の検索目的（情報収集目的、または購入目的）を抽出するエージェント機能を開発することを目的とする。

本研究と同様に、ユーザの商品検索目的を抽出する先行研究として Gou らの研究 [1] が挙げられる。ここでは、ユーザの検索インタラクションに注目し、ユーザの目的 (purchase or research) を抽出するために、検索エンジンの結果ページ (Search Engine Result Page, 以下 SERP と呼ぶ) に対するユーザのインタラクション、クエリに関する特徴量、検索結果の質などの六つのグループの特徴量を使用している。しかし、SERP に対するユーザのインタラクションの前後関係 (系列) は考慮されていない。通常、ユーザは SERP に対して連続的にインタラクションを起こし、ユーザのある時点でのインタラクションは、その直前に行ったインタラクションから影響を受け、その直後のインタラクションに影響を与えると考えられる。また、そうしたユーザの一連のインタラクションは検索目的によって異なってくると考えられる。例えば、情報収集目的のユーザは、SERP から参考になりそうなものを選択し、そのリンク先のページ (詳細ページ) を訪れ、詳細ページで様々なインタラクションを起こし、SERP に戻る、といった一連の行動を行うことが考えられる。一方で、購入目的のユーザは上記のパターンよりも、既に知っている EC サイトを訪れ、目的の商品を探索し、価格比較サイトに訪れる、といった行動を行うだろうなどと考えられる。そこで、我々はユーザの一連のインタラクションの系列 (インタラクションシーケンス) に着目し、ユーザの商品検索の目的 (情報収集目的か、購入目的か) を抽出する手法を提案する。

本論文の構成は次のとおりである。まず 2. では商

品検索目的抽出エージェントの概要について述べる。次に、3. で検索目的の抽出手法について述べた後、4. で評価実験の結果と考察を示す。そして、5. で関連研究との比較について述べ、6. でまとめとする。

2. 商品検索目的抽出エージェント

本エージェントは、ユーザが検索エンジンを利用して商品検索を行う際の目的 (情報収集目的、購入目的) を抽出するものであり、主に JavaScript で実装されている (図 1)。ここでは前述した目的に照らし、実用性の観点から特別なインストールをせずに JavaScript から取れるインタラクションのみを対象とする。ただし、実験・評価を行うにあたり、EC サイト・検索サイトのページに我々のスクリプトを埋め込むことはできないため、今回は Firefox のアドオンとして開発した。将来的には、JavaScript は EC サイト・検索サイトの検索結果ページに埋め込まれ、意図抽出はサイト側のサーバによってユーザのインタラクション中にリアルタイムで行われることを想定している。なお、今回、分類器には機械学習ツール Weka [14] に実装された SVM (Support Vector Machine) を利用した。ユーザが任意のキーワードで商品の検索を行うと、インタラクション取得部はユーザの SERP に対するインタラクションを取得し、出力する。分類器は出力されたインタラクション (特徴量) を用いて、ユーザの検索目的 (購入目的か、情報収集目的か) を判定する。

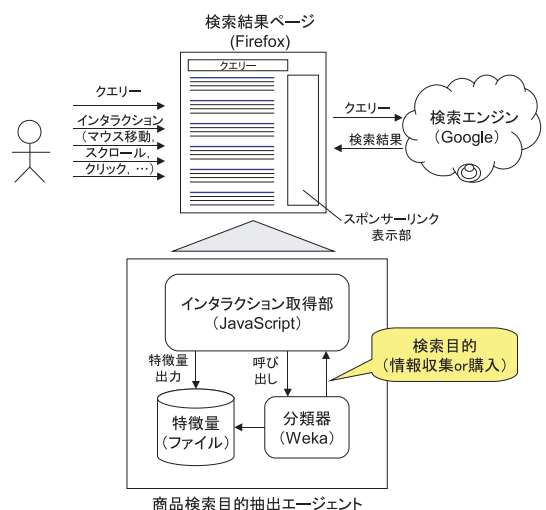


図 1 システム構成
Fig. 1 System architecture.

3. 検索目的の抽出手法

3.1 セッションの定義

本論文では、ユーザの一連の検索行為をセッションと呼ぶ(図2)。1セッション S は複数のクエリ q から構成され、各セッションには一つの目的があるものとする。

3.2 インタラクションシーケンスのモデル

本論では、ユーザの一連のインタラクションの系列(インタラクションシーケンス)を特徴量として扱うため、まずインタラクションシーケンスのモデル化を行う。SERPに対して、ユーザが必ず行う基本的なインタラクションとしては、マウス移動(Mousemove)、スクロール(Scroll)、クリック(Click)、ホバー(Hover、リンクの上でマウスが停止すること)、及びWebページを読む(Read)という動作が挙げられるだろう。ここでは、任意の二つのインタラクション動作間の間隔が一定時間(1秒)以上、一定時間以下(3分)である場合にWebページを読んでいたと認識し、Readと判定している。そこで、我々はこの五つのインタラクション動作を中心にインタラクションシーケンスをモデル化した。

図3は、簡単なモデル例である。上部のラベルはユーザのインタラクション動作を、縦線は時間を、縦

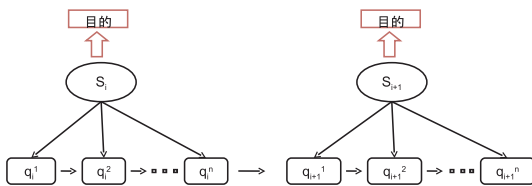


図2 セッションの定義
Fig.2 Session definition.

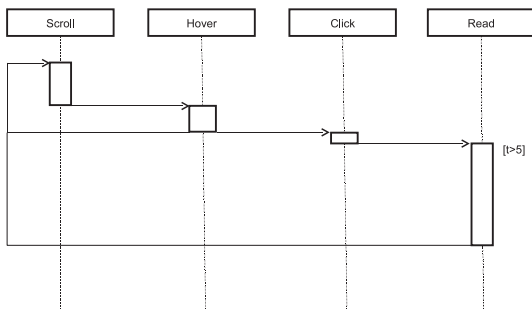


図3 インタラクションシーケンスのモデル
Fig.3 Interaction sequence model.

長方形の長さはそれぞれの動作の時間を表す。ただし、Clickの場合は便宜上、1クリックに単位時間かかるものとし、方形の長さでクリック数を表すものとする。同様に、スクロールは連続スクロールの時間、またはスクロールの回数を表す。図3は、ユーザがスクロールして、マウスをリンクに当て、クリックし、ページを読む動作を表している。それぞれのインタラクション動作間を結ぶ右向きの矢印は、動作が連続していることを表している。動作の終了と開始は基本的に同期しているが、ここでは厳密に区別してはいない(検索意図抽出に重要ではないと思われるため)。ま

表1 特徴量一覧
Table 1 Table of features.

No.	特徴量	値
1	Is(ScrlUpnDown_HoverLT2s_ClkSERP)	{T, F}
2	Is(ScrlUpnDown_HoverLT2s_ClkRK)	{T, F}
3	Is(ScrlUpnDown_HoverLT2s_ClkNext)	{T, F}
4	Is(ScrlUpnDown_ReadSERPMT5s_ClkSM)	{T, F}
5	Is(ClkECsites_ReadMT30s)	{T, F}
6	Is(ClkPCsites_ReadMT30s)	{T, F}
7	Is(ClkNextOfSERP_ClkOnSERP)	{T, F}
8	NumOf(ClkOnECsites_SAT)	{0,...,N}
9	NumOf(ClkOnCPsites_SAT)	{0,...,N}
10	NumOf(ClkOnECsites_DSAT)	{0,...,N}
11	NumOf(ClkOnCPsites_DSAT)	{0,...,N}
12	NumOf(NotECorCPsiteClk_ReadLT15s)	{0,...,N}
13	NumOf(NotECorCPsiteClk_ReadMT30s)	{0,...,N}
14	NumOf(SERP_Clk_ReadMT5s_FirstClk)	{0,...,N}
15	NumOf(SERP_Clk_Scrl_ReadMT5s)	{0,...,N}
16	NumOf(Scrl_ReadSERPMT5_ClkSERP)	{0,...,N}
17	NumOf(Scrl_ReadSERPMT5_ClkSERP)	{0,...,N}
18	NumOf(Scrl_HoverAdsMT5s_SERP_Clk)	{0,...,N}
19	NumOf(Scrl_HoverAdsMT5s_AdsClk)	{0,...,N}
20	NumOf(Scrl_HoverSERPMT5s_ClkSERP)	{0,...,N}
21	NumOf(ScrlUpnDown_ClkLast5ofSERP)	{0,...,N}
22	NumOf(ClkSERP_ReadMT30s)	{0,...,N}
23	NumOf(ScrlUpnDown_Clk_ReadMT15s)	{0,...,N}
24	NumOf(ReadTillBottomOfPage_Clk)	{0,...,N}
25	NumOf(ScrollI15_ClkOnLast5Results)	{0,...,N}
26	NumOf(SERP_Clk_ReadLT15s)	{0,...,N}
27	NumOf(Clk_ReadMT5sOnDetailPage)	{0,...,N}
28	NumOf(ClkOnImage_ReadMT5s)	{0,...,N}
29	NumOf(Clk_ReadMT15sOnDetailPage)	{0,...,N}
30	NumOf(ScrollUpDown_Clk_ReadMT15s)	{0,...,N}
31	NumOf(Clk_ReadLT15sOnDetailPage)	{0,...,N}
32	NumOf(ReadTillTheBottomOfPage_Clk)	{0,...,N}
33	NumOf(ClkNextOfSERP)	{0,...,N}
34	AvgOf(ScrollSpeedOfDetailPage)	{0,...,N}
35	AvgOf(ReadTimeOnDetailPage)	{0,...,N}
36	AvgOf(MousemoveOnDetailPage)	{0,...,N}
37	AvgOf(ClkOnDetailPage)	{0,...,N}
38	MaxOf(ReadTimeOnDetailPage)	{0,...,N}
39	MaxOf(MousemoveOnDetailPage)	{0,...,N}
40	MaxOf(ClkOnDetailPage)	{0,...,N}

た、SERP をスクロールしながら検索結果を閲覧する場合、Hover と Scroll は交互に繰り返す。Read から Scroll なども同様である。Hover, Read の終了から左、上に伸びる矢印は、こうした繰り返し動作を意味している。なお、5 秒以上など補足的情報を $[t > 5]$ のように括弧内に付記している。

3.3 特徴量

我々はまず SERP に対してユーザがどのようなインタラクションシーケンスを行うかを検証するため、研究室のメンバ 5 名で種々の検索行動を行い、それらを動画でキャプチャした。そして、検索行動を上モデルに照らして形式化し、表 1 のようなインタラクションシーケンスを特徴量として抽出した。なお、表 1 内の記号の意味を表 2 に示す。また、表 2 における満足・不満足の評定は、Fox らの研究 [6] に基づいてユーザが 30 秒以上ページを読んでいた場合、ユーザはそのページに満足していると判定する。また、ユーザが 15 秒以下でページを読むことをやめた場合、ユーザはそのページに満足していないと判定している。

各特徴量の名前は、Is, NumOf, AvgOf, MaxOf から始まり、それぞれの値は Bool 値、または整数値である。例えば、Is(ScrollUpnDown_HoverLT2s.ClkSERP) (表 1 の一番上の特徴量) は、ユーザが検索エンジンの結果ページを上下にスクロールしてから、検索結果一覧の各項目にマウスを 2 秒以下当てた後、検索結果をクリックする、という一連のインタラクションをしたかどうかを意味する。同様に、

NumOf(SERPClick_ReadMT5s.FirstClick) (表 1 の 14 番目の特徴量) は、検索エンジンの結果ページをクリックしてから、詳細ページを 5 秒以上読んで、その後詳細ページに貼ってあるリンクを初めてクリックする、という一連のインタラクションが何回行われたかを意味する。

4. 評価実験

4.1 実験方法

本手法の有効性を評価するために、7 名の被験者が情報収集目的の検索を 2 セッション、購入目的の検索を 2 セッションの計 4 セッションを実施した。実験では可能な限りユーザの自然な検索を再現するために、検索する商品は指定せず、ユーザが個人的に欲しいと思う商品 (調べたいと思う商品) を、ユーザのプライベートな時間に検索した。セッションに時間的な制限も設けていない。実際は、被験者とセッションの目的によって異なるが 10~30 分程度であった。また、購入目的の場合に、セッション中に実際に購入したかどうかは調査していない。購入目的の検索と情報収集目的の検索の実行順序も自由とし、実験後にそれぞれのセッションデータに正解ラベルとして目的を付与した。

また、比較のため、前述の先行研究で用いられた特徴量 32 種も同時に取得した (表 3)。なお、先行研究では筆者らが独自に改良したライブラリを組み込んだブラウザを用いて 127 の特徴量を取得しているが、我々は前述した目的に照らし、実用性の観点から JavaScript で取得できる 32 種を対象とした。

4.2 データセット

7 名の被験者から集めたデータ (表 4) をデータセットとし、9 割を学習データ、残りの 1 割をテストデータとした。

4.3 実験結果

実験では、提案手法が推測する目的と、被験者の目的がどの程度一致するかを、正確度、適合率、再現率の観点から評価した (表 5)。それぞれの定義は以下のとおりである。

- 正確度は、購入目的のラベルが付いたクエリを正しく購入目的として推測した数 (trueBuy) と、情報収集目的のラベルが付いたクエリを正しく情報収集目的として推測した数 (trueResearch) の和を、データセットの全クエリの数 (allQueriesInDataSet) で割ったものである。

表 2 記号一覧
Table 2 Table of symbols.

記号	意味
-	動作遷移
SERP	検索エンジンの結果ページ
RK	検索エンジンが提供する関連キーワード (Related Keywords)
Next	検索エンジンの Next ボタン
SM	検索エンジンの左側のカテゴリー検索の各メニューリスト (Side Menu)
Ads	検索エンジンのスポンサーリンク (Advertisement)
SAT	満足である場合 (Satisfied)
DSAT	不満足である場合 (Dissatisfied)
s	秒 (Seconds)
LT	より小さい (Less Than)
MT	より大きい (More Than)
DetailPage	SERP の結果をクリックしたときのリンク先のページ
ECsites	電子商取引サイト (E-Commerce Sites)
PCsites	商品価格比較サイト (Price Comparison Sites)

表 3 本実験で取得した先行研究の特徴量
Table 3 All features of previous work.

特徴量のグループ	特徴量の数	特徴量
Query	3	QueryLengthChars
		QueryLengthWord
		IncludesTLD
Result Quality	7	TotalAds
		NorthAds
		EastAds
		SnippetOverlap
		SnippetOverlapNorm
		AdOverlap
		AdOverlapNorm
Interaction	8	TotalMouse
		TotalScroll
		TotalKeypress
		SERP dwellTime
		DeliberationTime
		HoverEastAds
		HoverNorthAds
		HoverOrganic
		ClickUrl
Click	7	NumBrowseAfterClick
		AverageDwellTime
		TotalDwellTime
		SAT
		DSAT
		ClickType
Context	7	ClickUrl
		NumBrowseAfterClick
		AverageDwellTime
		TotalDwellTime
		SAT
		DSAT
		ClickType

表 4 データセット
Table 4 Dataset.

目的	セッション数	クエリ数
購入	14	29 クエリ
情報収集	14	18 クエリ

表 5 商品検索目的の抽出の結果
Table 5 Experimental results of the extraction of searching intention.

	正確度		情報収集目的		購入目的	
	Acc.	Prec.	Recall	Prec.	Recall	
提案手法	72.34	66.7	55.6	75.0	82.8	
先行研究	61.70	50.0	38.9	66.7	75.9	
提案手法+先行研究	74.4	68.8	61.6	77.4	82.8	

$$Acc. = \frac{trueBuy + trueResearch}{allQueriesInDataSet}$$

• 適合率は、購入目的の適合率 (PrecOfBuy) と情報収集目的の適合率 (PrecOfResearch) の二つがある。情報収集目的のラベルが付いたクエリを誤って購入目的として推測した数を falseResearch とし、購入

目的のラベルが付いたクエリを誤って情報収集目的として推測した数を falseBuy とした場合、購入目的の適合率と情報収集目的の適合率は以下の式で表すことができる。

$$PrecOfBuy = \frac{trueBuy}{trueBuy + falseResearch}$$

$$PrecOfResearch = \frac{trueResearch}{trueResearch + falseBuy}$$

• 再現率は、適合率と同様に、購入目的の再現率 (RecOfBuy) と情報収集目的の再現率 (RecOfResearch) の二つがある。それぞれ以下の式で表すことができる。

$$RecOfBuy = \frac{trueBuy}{trueBuy + falseBuy}$$

$$RecOfResearch = \frac{trueResearch}{trueResearch + falseResearch}$$

4.4 考察

4.4.1 商品検索目的の抽出

表 5 の結果から、全ての評価項目 (正確度, 適合率, 再現率) において、提案手法は先行研究の特徴量 (の一部) を用いた場合よりも高い精度を得ていることが分かる。特に、購入目的の再現率は高い精度が得られた。これは、提案手法が購入目的の検索を推測するのに十分なインタラクションシーケンスを特徴量として追加できていることを意味していると思われる。それらはいわば、比較的迷いのない検索行動のパターンであり、それを特徴として捉えることで推測が容易になったものと考えられる。また、表 5 における「提案手法+先行研究」とは先行研究の特徴量と提案手法の特徴量 (インタラクションシーケンス) を分類器への同レベルの入力として合わせて学習を行ったものである。この結果から、インタラクションを単独で用いた場合よりも、インタラクションシーケンスを加えた場合、全ての評価項目について精度を改善できたことが確認できる。

4.4.2 データセットの分析

表 4 より、計 28 セッション (7 人 × 4 セッション) の中に 47 のクエリが存在した。つまり、1 セッション当りのクエリ数は 1.67 であり、ユーザは一つの目的を実現するために 1 クエリ以上検索したことが分かった。また、購入目的のクエリ数が 29 クエリあったが、情報収集目的のクエリ数は 18 クエリしかなく、購入

目的の場合は被験者がより多くの検索したと考えられる。それに対し、情報収集目的の場合は被験者は多くのクエリを発行するよりも検索結果をブラウズすることが多かったと考えられる。更に詳しく見ると、被験者は、購入目的の場合、一つの目的を達成するために平均で 2.07 クエリ (29/(7 × 2)) 検索し、情報収集目的の場合は一つの目的を達成するために平均で 1.28 クエリ (18/(7 × 2)) 検索している。この結果から、ユーザは購入目的のときは調査目的のときより慎重になり複数のクエリ (2 クエリ以上) を発行し、検索を行う傾向があると考えられる。これは直感にも合致している。したがって、これを両者の違いとして特徴量に加えることによって、更なる抽出精度の向上が期待できるだろう。

5. 関連研究

まず、前述した先行研究 [1] と本研究との相違について述べる。先行研究の目的は、我々と同様に検索エンジンの結果に対するユーザのインタラクションを用いて、ユーザの商品検索の目的を抽出することである。ここでは、六つの特徴グループ (Query, SERP Content, Result Quality, Interaction, Click, Context) から、SVM と CRF (Conditional Random Fields) を用いて、ユーザの商品検索の目的を抽出している。しかし、例えば HoverOrganic (SERP にマウスを移動する時間) や、DeliberationTime (SERP が表示されてから最初のクリックが発生するまでの時間) など非常に細かな特徴が取得されている一方で、それぞれは単独で表現され、その前にユーザがどのような動作をしたか、または、その後どのような動作をしたか等のインタラクションシーケンスは考慮されていない。そこで、我々はユーザのある時点でのインタラクションは、その直前または直後のインタラクションと関係が深く、それらは検索目的により異なると仮定し、インタラクションシーケンスに着目した特徴量を提案した。そして、実験によりインタラクションを単独で扱う場合と比べて、検索目的の抽出精度を向上できる可能性を示した。結果的に精度向上できた理由を定性的に分析すると、インタラクションシーケンスはインタラクションを複数組み合わせたものであり、いわば SVM 分類器にとって幾つかの入力変数を関数化したものとなっている。そのため、高次元データを低次元に縮約する主成分分析等と同様の効果が得られたのではないかと考えている。しかし、前述したように本研究では

広告配信への応用を目的としてインストール不要な JavaScript で取得できる特徴に限定しているため、本研究が先行研究よりも精度が良いことを主張するものではない。あくまでもインタラクションを単独で扱うだけでなく、シーケンスとして扱うことの有用性を示したものである。

なお、先行研究では SERP におけるユーザの行動モデルのみを対象としており、SERP からクリックされたリンク先のページにおけるユーザの行動は考慮されていない。それに対して、本研究ではリンク先のページにおけるユーザの閲覧行動やインタラクションも考慮している点にも相違がある。

ユーザの検索目的と検索行動の関係を明らかにすることは、本研究で対象としている広告配信に限らず、Web 検索の有用性を向上させるのに重要と考えられる。しかし、ユーザの検索目的を抽出するためには、ユーザの行動モデルやユーザの嗜好モデルを作成し、これらのモデルを用いて、ユーザがどのような目的で検索したか、または、これからどのような目的で検索するかを推測する必要がある。

そのため、Broder ら [3] はまずユーザの検索目的を分類するために、Web 上での情報探索タスクを情報指向 (informational)、ナビゲーション指向 (navigational)、トランザクション指向 (transactional) の 3 カテゴリーを提案している。

また、ユーザの行動履歴やインタラクションからユーザの嗜好を抽出し、ユーザに有益なページを推薦する研究もある。Claypool ら [4] は、Web ページに対するユーザの滞在時間、クリック、スクロールなどといったユーザインタラクションからユーザの嗜好を抽出している。それに対し、White ら [2] はユーザの嗜好を、時間の観点から 3 種類 (short, medium, long) の嗜好に分類している。そして、これら 3 種類の嗜好を効果的に抽出するために、五つのコンテキストモデル (Interaction, Task, Collection, Historic, Social) を提案している。しかし、両研究ともユーザがブラウジングした履歴やユーザが起こしたインタラクションからユーザの嗜好を抽出している。

一方、Cheng ら [5] は、あるセッションの直後のユーザ行動は、その直前と現在のセッションの行動から影響を受けるという仮定に基づき、ユーザのブラウジングから検索に移り変わるパターンに着目し、そのセッションの直前の行動やセッション中の行動から重要なキーワードを抽出して、その直後の検索でユーザにク

エリを推薦している。

また, Fox ら [6] は, ユーザのセッション中のインタラクションを利用し, 検索結果に対するユーザの満足度を三つのレベル, 満足 (Satisfied), 部分的に満足 (Partially Satisfied), 不満足 (Dissatisfied) に分類している. Fox らの目的は, ユーザの明示的な評価 (explicit rating) と暗黙的な評価 (implicit rating) の関係を明らかにすることである. ベイズモデルを用いて, 検索結果に対するユーザの満足度を分類するために最も貢献した特徴量を探し, 各特徴量の関係の比較を行っている. このようにユーザの嗜好・目的を抽出するためには, 抽出の対象となる目的を明確にし, そのための特徴量とタスクの設定を決めることが重要である. また, 特徴量については, ユーザが発行したクエリ, ユーザが見た (見ている) ページの内容, ユーザが行った閲覧行動など, 様々な観点から分析し, モデルを作成することも重要であると考えられる.

ほかにも, 発行されたクエリに対する検索結果の中からユーザがどの結果に対して, より興味をもっているかをユーザのクリックモデルから推測する手法も提案されている. Agichtein ら [7] の研究では, ユーザが検索結果一覧の中の 2 番目と 5 番目の結果をユーザがクリックした場合に, どちらがよりユーザの興味に近いかをユーザの行動モデルから推測している. 更に, Murdock ら [8] は大規模な検索データからユーザの Click-through データとユーザモデルを用い, Web 検索エンジンの性能評価を行っている.

6. む す び

本論文では, 効果的な広告配信を目的として, 広告配信側に代わってユーザの商品検索意図を推定するエージェントシステムを開発した. そして, ユーザのインタラクションシーケンスを特徴量として扱う手法を提案し, 実験によりインタラクションを単独で扱う場合と組み合わせることで抽出精度を 10%程度向上できることを示した.

今後は実際に EC サイトと協働し, 本手法を EC サイト上の検索システムへ適用していきたい. また, 特徴量を追加し, 抽出精度の更なる向上を図っていきたい.

謝辞 本研究を遂行するにあたり, 研究の機会と議論・研鑽の場を提供して頂き, 御指導頂いた国立情報学研究所/東京大学本位田真一教授をはじめ, 活発な議論と貴重な御意見を頂いた研究グループの皆様

に謝致します.

文 献

- [1] Q. Gou and E. Agichtein, "Ready to buy or just browsing? Detecting web searcher goals from interaction data," Proc. 33rd ACM SIGIR Conf., 2010.
- [2] R.W. White, P. Baily, and L. Chen, "Predicting user interests from contextual information," Proc. 32nd ACM SIGIR Conf., 2009.
- [3] A. Broder, "A taxonomy of Web search," SIGIR Forum, vol.36, no.2, pp.3-10, 2002.
- [4] M. Claypool, D. Brown, P. Le, and M. Waseda, "Inferring User Interest," IEEE Internet Computing, pp.32-39, Dec. 2001.
- [5] Z. Cheng, B. Gao, and T.-Y. Liu, "Actively predicting diverse search intent from user browsing behaviors," Proc. 19th World Wide Web Conf. (WWW), 2010.
- [6] S. Fox, K. Karnawat, and M. Mydland, "Evaluating implicit measures to Improve Web search," ACM Trans. Information System, pp.147-168, USA, April 2005.
- [7] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning user interaction models for predicting Web search result preferences," Proc. 29th ACM SIGIR Conf., 2006.
- [8] V. Murdock and B. Piwowarski, "Web search engine evaluation using clickthrough data and a user model," Proc. 16th World Wide Web Conf. (WWW), 2007.
- [9] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey, "An experimental comparison of click position-bias models," Proc. 1st Intl. Conf. Web Search and Data Mining (WSDM), 2008.
- [10] J. Teevan, S.T. Dumais, and D.J. Liebling, "To personalize or not to personalize," Proc. 31st SIGIR Conf., 2008.
- [11] M. Holub and M. Bielikova, "Estimation of user interest in visited Web page," Proc. 19th World Wide Web Conf. (WWW), 2010.
- [12] Y. Liu, J. Bian, and E. Agichtein, "Predicting information seeker satisfaction in community question answering," Proc. 31st SIGIR Conf., 2008.
- [13] E. Agichtein, "Recent progress on inferring Web searcher intent," Proc. 19th World Wide Web Conf. (WWW), 2010.
- [14] Weka, <http://www.cs.waikato.ac.nz/ml/weka/>
- [15] H. Terai, H. Saito, Y. Egusa, M. Takaku, M. Miwa, and N. Kando, "Differences between informational and transactional tasks in information seeking on the Web," Proc. Information Interaction in Context, 2008.
- [16] F. Qiu and J. Cho, "Automatic identification of user interest for personalized search," Proc. 15th World Wide Web Conf. (WWW), 2006.
- [17] M. Kellar, C. Watters, and M. Shepherd, "A field study characterizing Web-based information seeking

task,” J. American Society for Information Science and Technology (JASIST), pp.999–1018, 2007.

- [18] B. Piwowarski and H. Zaragoza, “Predictive user click models based on click-through history,” Proc. Conf. Information and Knowledge Management (CIKM), 2007.
- [19] D.E. Rose and D. Levinson, “Understanding user goals in web search,” Proc. 13th World Wide Web Conf. (WWW), 2004.
- [20] 第 14 回 ネット広告に関する調査,
<http://japan.internet.com/research/20110325/1.html>

(平成 23 年 1 月 13 日受付, 5 月 23 日再受付)



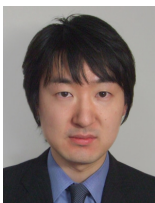
川村 隆浩

1992 早大・理工・電気卒. 1994 同大学院理工学研究科電気工学専攻修士課程了. 同年, (株) 東芝入社. 現在, 同社研究開発センター主任研究員. 工博. 2001~2002 米国カーネギーメロン大学ロボット工学研究所客員研究員. 2003 より電気通信大学大学院情報システム学研究科客員准教授. 2007 より大阪大学大学院工学研究科非常勤講師. 主としてマルチエージェントシステム, セマンティック Web の研究・開発に従事. 人工知能学会, 情報処理学会各会員.



ワ コラ

2011 電気通信大学大学院情報システム学研究科社会知能情報学専攻修士課程了. 同年, 楽天 (株) 入社.



中川 博之 (正員)

1997 阪大・基礎工・情報工学卒. 同年, 鹿島建設 (株) 入社. 2007 東京大学大学院情報理工学系研究科修士課程了. 2008 同大学院博士課程中退. 同年より電気通信大学助教. 主としてエージェント及び自己適応システム開発手法の研究に従事. 情報処理学会, IEEE CS 各会員.



田原 康之

1991 東京大学大学院理学系研究科数学専攻修士課程了. 同年, (株) 東芝入社. 1993~1996 情報処理振興事業協会に外向. 1996~1997 英国 City 大学客員研究員. 1997~1998 英国 Imperial College 客員研究員. 2003 国立情報学研究所入所. 2008 より電気通信大学准教授. 博士 (情報科学) (早稲田大学). 主としてエージェント技術, 及びソフトウェア工学などの研究に従事. 情報処理学会, 日本ソフトウェア科学会各会員.



大須賀昭彦 (正員)

1981 上智大・理工・数学卒. 同年, (株) 東芝入社. 同社研究開発センター, ソフトウェア技術センターなどに所属. 1985~1989 (財) 新世代コンピュータ技術開発機構 (ICOT) 外向. 2007 より電気通信大学大学院情報システム学研究科教授. 工学博士 (早稲田大学). 主としてソフトウェアのためのフォーマルメソッド, エージェント技術の研究に従事. 1986 年度情報処理学会論文賞受賞. 情報処理学会, 日本ソフトウェア科学会, 人工知能学会, IEEE CS 各会員.