一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
DE2018-16 (2018-09)

オンラインショッピングにおける商品選択プロセスのモデル化手法の提案

渡辺　郁弥†　野崎　祐里††　佐藤　哲司†††

† 筑波大学 情報学群 知識情報・図書館学類
†† 筑波大学 図書館情報メディア研究科
††† 筑波大学 図書館情報メディア系

あらまし　近年，オンラインショッピングサイト上で買い物を行う機会が増えてきている．オンラインショッピングサイトの特徴として，検索フォームにクエリを入力して商品を検索できることが挙げられる．ユーザの検索行動を分析することは，適切な検索支援や売上向上の点で重要である．そこで本稿では，検索クエリの移り変わりに注目し，系列パターンマイニングの手法を用いて検索行動の分析を行う．具体的には，頻出する検索パターンの抽出と，初期の検索パターンと検索行動の長さの関係を評価実験により明らかにする．評価実験の結果，クエリを変更しないで検索結果の別のページに移動する行動が検索行動の大半を占め，また，この行動が検索初期に続くと検索行動が長くなることが知見として得られた．

キーワード　EC サイト，情報検索，クエリ変更，系列パターンマイニング

Modeling Item Selection Process on Online Shopping

Fumiya WATANABE†, Yuri NOZAKI††, and Tetsuji SATOH†††

† College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba
†† Graduate School of Library, Information and Media Studies, University of Tsukuba
††† Faculty of Library, Information and Media Science, University of Tsukuba

Abstract　In recent years, opportunities to purchase products on online shopping sites are increasing. One of the features of online shopping sites is to search for products by entering queries on search forms. Analyzing the user's search behavior is important for appropriate search support and sales improvement. In this paper, we focus on the change of the search query and analyze the search behavior using the sequence pattern mining method. We extracted frequent search patterns and analyze the relationship between the initial search pattern and the length of the search. As a result of the evaluation experiment, it was found that the behavior which moves to another page of the search result without changing the query occupies the majority of the search behavior. It was also found that the search behavior becomes longer when that action follows the initial stage of the search.

Key words　EC site, information retrieval, change query, sequence pattern mining

1. 背　　景

近年，Amazon や楽天，メルカリなどの Web サービスの登場により，インターネット上で買い物を行う機会が増えてきている．オンラインショッピングのメリットとして，実店舗での買い物と比較して，店舗に出向く手間が省けること，複数の店舗に訪れる必要がないこと，購入した商品を運ばなくてよいことなどが挙げられる．さらに，オンラインショッピングサイトの機能として，検索フォームにクエリを入力して商品を検索することができる．検索機能を使いこなせば，利用者は目的の商品のページに到達することが容易に可能になる．一方でオンラインショッピングサイトの運営者は，自社のサービスを利用して商品を購入する回数が増えることを期待しており，ユーザがどのような行動をしているのかを把握することは重要である．

そこで本稿では，オンラインショッピングサイト上のユーザ行動として検索クエリの遷移を分析する．検索ログを用いてユーザのセッションを抽出し，セッション内において前後のクエリの変化を比較しコード化する．コード化したシークエンスを用いて，頻出パターンやパスの長さを分析することで有用な知見を発見する．

本稿の構成は以下の通りである．まず，第 2 章で既存の検索行動の研究を示し，本研究の位置づけを明らかにする．第 3 章で提案手法として，検索ログからセッションを抽出する方法，セッション内のクエリ変化をコード化する手法および，コード化したシークエンスの分析手法について説明する．第 4 章で実データを用いて評価実験を行い，第 5 章で結果の考察を行う．第 6 章で本稿のまとめと今後の課題について述べる．

2. 関 連 研 究

クエリの変更に関する研究として，関口ら [1] や Umemoto [2] らはクエリ変更意図を絞込，汎化，関連，修正，新規の 5 クラスを定義し機械学習による変更意図の予測を行っている．Jiang ら [3] は，クエリの変更要因をセッション，クエリ，キーワードの観点から分析をしている．Boldi ら [4] や Bordino ら [5] はクエリログからクエリフローグラフの構築を行っている．これはクエリ間の遷移確率をグラフで表現したもので，クエリの系列からクエリ候補の推薦手法を提案している．また，関口ら [6] は，シード語句に対して絞込に使用している語句群の類似度を計算することで，同一の属性語の抽出を行っている．深澤ら [7] は，レシピ検索サイトにおいて共起する食材の頻度を時系列に追い，その分散の変化を見ることで検索語の意味変化を分析している．

オンラインショッピングサイトのユーザ行動を分析している研究として，Moe [8] の研究がある．Moe はオンラインショッピングサイトのユーザの閲覧行動として，目標志向購買，探索・熟慮，快楽的閲覧，知的構築の 4 種類を定義し，セッションをいずれかの行動に分類をしている．また，Nozaki ら [9] は，セッションにおけるクエリの変更とページアクセスの完了率の推移を商品カテゴリごとに分析を行っている．

本研究は，セッションにおけるクエリの変更パターンを分析するので，関口ら [1] や Umemoto [2] らの研究と近い．これらの研究では，クエリの変更パターンを詳細に定義し予測タスクを行っているため，前後のクエリ間にのみ焦点を当てているが，本研究は前後のクエリにとどまらず，系列データとしてまとまった単位で分析することに違いがある．

3. 提 案 手 法

3.1 セッションの抽出

検索ログを用いてユーザのセッションを抽出する．セッションとは，目的のページを探し出すための一連の探索行動のことである．本稿で用いる検索ログには，ユーザの ID，タイムスタンプ，検索クエリのカラムを持つことを想定している．セッションの抽出の手順は以下のステップで行う．まず，検索ログをユーザの ID ごとに分割する．次にユーザ ID ごとに分けた検索ログデータをタイムスタンプをキーとして昇順に並び替える．並び替えたログを順に走査し，直前とのログの時間差が 30 分未満の場合同一のセッションとし，30 分以上の場合は別々のセッションとして切り出す．セッションの切り出しの基準を 30 分に設定したのは，Boldi [4] らの先行研究を参考にした．セッションの抽出例を表 1 に示す．

3.2 クエリ遷移のコード化

抽出したセッションにおいて，前後のクエリの変化に基づきコードを割り振る．クエリの変化を比較する際，クエリからキーワードの抽出を行う．キーワードとは，クエリを半角または全角スペースで分割した際の各文字列のことである．前のクエリを $Q_{pre} = \{w_1, w_2, ...w_n\}$，後のクエリを $Q_{post} = \{w_1, w_2, ...w_m\}$ で表す．w_i はクエリに出現するキーワードである．本稿はクエリの変化として以下の 5 種類を定義する．

全部置換（R）

前後のクエリ間に共通するキーワードが 1 つも存在しないとき，全部置換のコード R を割り振る．つまり，前後のクエリ間で以下の関係が成り立つ．

$$Q_{pre} \wedge Q_{post} = \emptyset \tag{1}$$

一部置換（M）

前後のクエリ間に共通するキーワードが 1 つ以上存在し，なおかつ前後のクエリのキーワード間に部分集合の関係がないとき，一部置換のコード M を割り振る．よって，前後のクエリ間で以下の関係が成り立つ．

$$Q_{pre} \wedge Q_{post} \neq \emptyset \tag{2}$$

$$Q_{pre} \neq Q_{post} \tag{3}$$

$$Q_{pre} \not\subset Q_{post} \tag{4}$$

$$Q_{pre} \not\supset Q_{post} \tag{5}$$

追 加（A）

前のクエリのキーワードが後のクエリのキーワードの部分集合で，なおかつキーワード数が後のクエリの方が大きいとき，追加のコード A を割り振る．よって，前後のクエリ間で以下の関係が成り立つ．

$$Q_{pre} \subset Q_{post} \tag{6}$$

$$Q_{pre} \neq Q_{post} \tag{7}$$

削 除（D）

後のクエリのキーワードが前のクエリのキーワードの部分集合で，なおかつキーワード数が前のクエリの方が大きいとき，削除のコード D を割り振る．ゆえに，前後のクエリで以下の関係が成立する．

表 1　セッション抽出例

userID	タイムスタンプ	クエリ	セッション
1	2016-09-05 19:37:41	usb	1
1	2016-09-05 19:37:48	usb 64gb	1
1	2016-09-05 21:58:25	ノートパソコン	2
1	2016-09-05 21:58:34	ノートパソコン	2
1	2016-09-05 22:41:44	花 プレゼント	3
1	2016-09-05 22:53:40	薔薇	3

表 2　クエリ変更パターンのコード化例

前のクエリ	後のクエリ	コード
水	お茶	R
お茶	お茶 500ml	A
お茶 500ml	お茶 500ml	C
お茶 500ml	お茶 12 本	M
お茶 12 本	お茶	D

$$Q_{pre} \supset Q_{post} \tag{8}$$

$$Q_{pre} \neq Q_{post} \tag{9}$$

継 続（C）

前後のクエリのキーワードが完全に一致するとき，継続のコード（C）を割り振る．つまり，前後のクエリで以下の関係が成立する．

$$Q_{pre} = Q_{post} \tag{10}$$

これは同一のクエリで，検索結果の 2 ページ目，3 ページ目へと移動している行動であることが考えられる．

クエリ遷移をコード化した例を表 2 に示す．コード化した系列データを本稿ではシークエンスと呼ぶことにする．

3.3　クエリ遷移の分析

抽出したシークエンスを分析することで，検索行動の特徴を明らかにする．分析内容は，以下の 3 種類である．

3.3.1　頻出パターンの抽出

どのような検索パターンが検索行動で出現しやすいのか分析することは重要である．そのため，多くのシークエンスに出現するパターンの抽出を行う．パターンの抽出アルゴリズムとして，PrefixSpan を用いる．これは，深さ優先型の探索で頻出系列パターンを抽出する手法である．本手法では，系列間のギャップ数は考慮せず，隣接するパターンのみを数え上げる．また，1 つのシークエンスに特定のパターンが複数回出現しても 1 回のみの出現として数える．

3.3.2　頻出パターンの出現割合の分析

前節で抽出した頻出パターンについて，シークエンス中に出現する割合をパス長ごとに比較する．これにより，パス長が長くなるほど検索パターンに差が生じるかを明らかにすることができる．パス長とは，セッション中のログ数のことであり，(シークエンスのコード数 +1) となる．特定のパターンについて，シークエンスの長さが 5，10，20，50 のときの出現割合を算出する．長さ n のシークエンスにおける長さ k のパターンが i 回出現したときの出現割合 $rate$ は以下の式で求められる．

$$rate = \frac{i}{n-k+1} \tag{11}$$

3.3.3　検索開始パターンの分析

ユーザがオンラインショッピングサイト上で買い物をするとき，あらかじめ購入したい商品が決まっている場合とそうでない場合がある．購入したい商品が決まっている場合は，数回の行動で検索が終了するが，そうでない場合は試行錯誤を行いながら探索を行う．これらの行動の違いは，検索行動の開始パターンによって把握することができるのではないかと考える．そのため，シークエンスの開始パターンととパス長との関係を分析する．

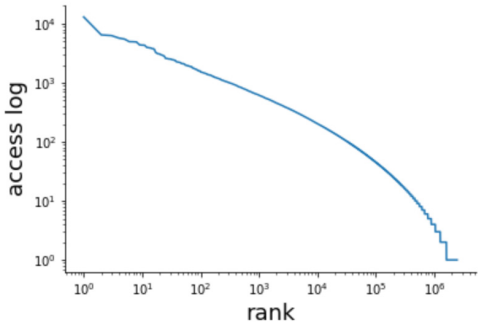

図 1　ユーザのログ数のランキング

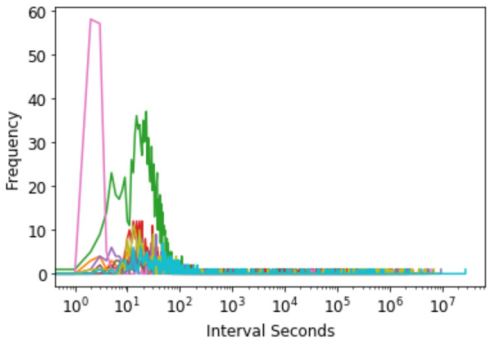

図 2　ログ間の時間差の分布

4.　評 価 実 験

4.1　実 験 準 備

実データを用いて提案手法の有効性を確認する．データセットとして，株式会社リクルートテクノロジーズから提供を受けたポンパレモール(注1)の検索ログデータを使用する．ログは 2016 年 6 月〜2017 年 12 月に記録されたもので，総数は 24,582,912 件である．

図 1 はデータセット中のログをユーザごとに集計し，ログの数が大きい順に並び替えたものである．ログ数のランキングが 100 位から 10,100 位までの 10,000 ユーザを実験用データとして用いる．

図 2 は実験用データのユーザからランダムに 10 名抽出し，ログ間の時間差の分布を示したものである．ログ間の時間差の多くは 100 秒以内に集まっており，セッションの時間差の区切りを 30 分に設定するのは十分であることが確認できる．

4.2　頻出パターンの結果

シークエンス中に出現する割合が高い頻出パターンを表 3 に示す．C のみのパターンの出現割合がかなり高くなっている．図 3 はシークエンスの長さが 1 のパターンが各シークエンス長においてどれだけの割合で出現しているのか平均をとったものである．また，図 4 では，シークエンス長が増加したときの

(注1)：https://www.ponparemall.com/

表 3　頻出パターン上位 5 件
(括弧内の値は出現割合)

rank	k=1	k=2	k=3
1	C(0.97)	C,C(0.738)	C,C,C(0.62)
2	R(0.368)	C,R(0.311)	C,R,C(0.271)
3	M(0.126)	R,C(0.311)	C,C,R(0.261)
4	A(0.104)	R,R(0.156)	R,C,C(0.259)
5	D(0.053)	M,C(0.11)	R,C,R(0.156)

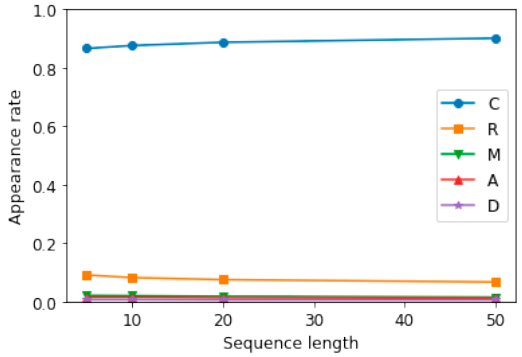

図 3　特定のシークエンス長におけるパターンの平均出現割合

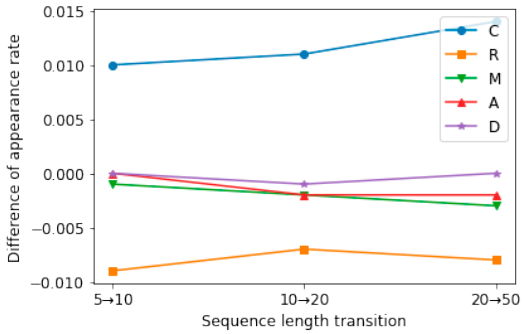

図 4　パターンの平均出現割合の増減

出現割合の増減をグラフにしている．グラフから，どの長さのシークエンスにおいても C が大部分を占めていること，シークエンスの長さが大きくなると C の割合が増加し，その他のコードの割合が低下することがわかる．

4.3　検索開始時の行動とパス長の関係

シークエンスの先頭の出現パターンにおけるシークエンスの長さの平均値との関係を図 5 に示す．分析対象パターンは，長さが 3 以下でシークエンスの先頭に出現する回数が 100 回以上のパターンのみである．図 5 から，コード C が続くと検索が長くなる傾向があることがいえる．

5.　考　　察

評価実験の結果から，クエリの変更パターンの大半が変更を行わない行動であることが明らかになった．つまり，ユーザは最初にクエリを入力したら，それ以降クエリの変更をほとんど行わず，ページの遷移を行うだけで検索を終了してしまうのである．このような行動が考えられる要因として，特定の商品名を入力せず，「ゲーム」「プレゼント」「訳あり」など広義な意味

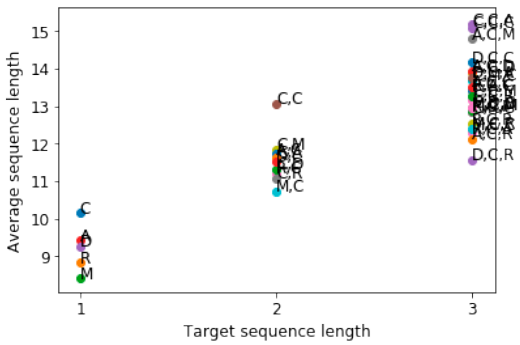

図 5　初期パターンとシークエンスの長さの関係

のとれるクエリを入力して，検索結果のページを移動しながらお得な商品や面白い商品がないかを探していることが考えられる．シークエンスの初期パターンが C のみだと検索行動が長くなるのも，このことが理由であることが示唆される．

6.　お わ り に

本稿では，オンラインショッピングサイトの検索ログの遷移の分析を行った．ユーザのログ群からセッションを抽出し，セッション内の前後のログの変化をコード化したシークエンスを構築した．評価実験では，シークエンスの頻出パターンと，シークエンスの初期のパターンと検索行動の長さの関係を分析した．実験の結果から，大半のクエリの推移パターンは「変更なし」であること，シークエンスの初期パターンが変更なしであることが続くと長い検索行動になることが明らかになった．

今後の課題として，今回は取り扱わなかったクエリの内容に着目してみること，行動パターンに基づくユーザのクラスタリングなどが挙げられる．

謝辞　本研究は JSPS 科研費 JP16H02904 の助成を受けたものである．また，DBSJ Data Challenge プログラムに参加し，株式会社リクルートテクノロジーズから提供を受けたポンパレモールのデータを利用している．ここに記して謝意を示す．

文　　献

[1] 関口裕一郎, 杉崎正之, 内山匡, 藤村滋, 望月崇由. 検索クエリログを用いたクエリ変更意図の自動推定. 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM2011), 2011.

[2] Kazutoshi Umemoto, Takehiro Yamamoto, Satoshi Nakamura, and Katsumi Tanaka. Predicting query reformulation type from user behavior. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pp. 894–901, 2013.

[3] Jiepu Jiang and Chaoqun Ni. What affects word changes in query reformulation during a task-based search session? In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, CHIIR '16, pp. 111–120, 2016.

[4] Paolo Boldi, Francesco Bonchi, Carlos Castillo, Debora Donato, Aristides Gionis, and Sebastiano Vigna. The query-flow graph: Model and applications. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pp. 609–618, 2008.

[5] Ilaria Bordino, Carlos Castillo, Debora Donato, and Aristides Gionis. Query similarity by projecting the query-flow graph. In *Proceedings of the 33rd International ACM SIGIR*

Conference on Research and Development in Information Retrieval, SIGIR '10, pp. 515–522, 2010.

[6] 関口裕一郎, 田中智博, 内山匡, 藤村滋, 望月崇由. 検索クエリログのセッション情報を利用した属性語句抽出. 第 2 回データ工学と情報マネジメントに関するフォーラム (DEIM2010), 2010.

[7] 深澤祐援, 原島純. 料理レシピサービスにおける検索語の意味変化に関する分析. 研究報告自然言語処理（NL）, Vol. 2016-NL-228, No. 2, pp. 1–8, 2016.

[8] Wendy W. Moe. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology*, Vol. 13, No. 1, pp. 29 – 39, 2003. Consumers in Cyberspace.

[9] Yuri Nozaki and Tetsuji Satoh. Category classification methods reflecting item search behaviors on online shopping sites. In *Proceedings of 7th IIAI International Congress on Advanced Applied Informatics*, pp. 32–37, 2018.