

# コンテキスト付き検索ログを用いた要求パーティカルの分析

門田見侑大<sup>†</sup> 吉田 泰明<sup>††</sup> 藤田 澄男<sup>††</sup> 酒井 哲也<sup>†</sup>

<sup>†</sup> 早稲田大学大学院 基幹理工学研究科 情報理工・情報通信専攻

〒 169-9555 東京都新宿区大久保 3-4-1

<sup>††</sup> ヤフー株式会社

〒 107-6211 東京都港区赤坂 9-7-1 ミッドタウン・タワー

E-mail: <sup>†</sup>kdtm-783640@ruri.waseda.jp, <sup>††</sup>{yayoshid,sufujita}@yahoo-corp.jp, <sup>†††</sup>tetsuyasakai@acm.org

あらまし スマートフォンやタブレット機器などの小型デバイスの普及により、今日では場所や時間に依らず検索を行うことが可能である。ここで、情報要求はユーザーの置かれた状況に依存し、例えば PC と小型デバイスで行われる検索ではユーザーの求める情報の種類に差異がある可能性がある。しかし、現状の検索エンジンにおいては、検索結果にユーザー情報や利用デバイス等のコンテキスト情報が充分考慮されていないと思われる。本研究は、コンテキスト情報付きの検索ログデータを用いて、異なるコンテキストのユーザーが閲覧するパーティカルの差およびパーティカル間の遷移に着目した分析を行い、知見を共有することを目的とする。分析の結果、ユーザー情報やコンテキストによる利用パーティカルの傾向やクエリに含まれる品詞やターム数による利用されるパーティカルの差、時間による検索数の遷移、平日と休日の検索数の差について知見を得ることができた。

キーワード 情報検索, クエリログマイニング

## 1. はじめに

情報要求はいかなる時、いかなる場所でも発生する。情報要求を満たすため、日常的に検索行為が行われてきたが、スマートフォンやタブレット機器などのモバイルデバイスが普及する以前は PC による検索が主であった。しかし、モバイルデバイスの普及に伴い、それらを用いた検索件数が増加し、2015 年 5 月、ついに Google 検索数においてスマートフォンが PC を上回ったと Google の公式ブログ<sup>(注1)</sup>にて発表された。

モバイルデバイスを用いることで、我々は検索行為において場所や時間等の制限を受けることが少なくなった。これにより、情報要求が発生した瞬間に検索行為に移行できるため、検索行為が多様化していると想定される。しかし、現在の検索結果において個人の検索履歴による最適化は進んでいても、ユーザー情報や利用デバイス等のコンテキスト情報による最適化は進んでいない。

本論文では、コンテキスト情報による検索結果の最適化へ向けた知見を共有することを目的とし、多量の検索ログデータを用いて検索時のコンテキスト情報とユーザーが閲覧するページのパーティカルの関係を分析した結果を示す。

なお、Arguello ら [1] はパーティカルを「ニュース・旅行・ローカル検索のような特定の分野もしくは画像や動画のような特定のメディアのために特化された部分コレクション」と説明しており、我々もこの定義に従う。

## 2. 関連研究

クエリの分析とパーティカルに関する研究として、二項クエリモデルを用いた情報タイプ (本論文におけるパーティカル) の抽出がある [2]。検索システムの利用者はクエリを物事の下位範疇と情報の下位範疇の 2 項形式で表現すると考え、クエリログの集合から十分な頻度を持つ第 2 項の単語を収集し、情報要求の経年変化を分析している。この論文の考察では、出現頻度が高い主要な情報タイプは時間が経過しても出現頻度が高く、それらの情報タイプにはそれらを検索するための専用のサービスが存在することが示されている。

また、検索ログを用いた研究として、Jansen ら [3] [4] [5] は Dogpile.com や Excite のログを用いて絞り込みや汎化といったクエリの改変パターンの変遷やパーティカルからパーティカルへの遷移確率、クエリに含まれる単語数やページの閲覧数などを分析している。

パーティカルを用いたタスクに、TREC Federated Web Search Track [6] の Vertical Selection Task がある。これは、与えられたクエリに対し最適なパーティカルを推薦するタスクであり、参加者は与えられたクエリに対して適切なパーティカルを返すシステムを構築し、検索有効性を競う。参加チーム [7] [8] は、SVM, Random Forest といった機械学習手法や、TF-IDF (Term Frequency - Inverse Document Frequency) を拡張した TWF-IRF (Term Weighted Frequency - Inverse Resource Frequency), LSI (Latent Semantic Indexing) モデルといった言語モデルを用いるなど様々な手法でこのタスクに取り組んでいる。その中で、最も F 値の高かったチーム [9] は「あるパーティカルにおいて多く出現する単語はそのパーティカルを表す可能性が高い」という仮定のもと FTR (Frequent Term Rank)

(注1): "Building for the next moment", Google Inside AdWords, <http://adwords.blogspot.jp/2015/05/building-for-next-moment.html>.

という指標を用いている。これに、クエリ拡張や IR モデルを組み合わせた手法が最高の F 値を示している。

### 3. データ分析

#### 3.1 対象データ

Yahoo! JAPAN<sup>(注2)</sup>の2015年11月の検索ログから抽出されたデータを用いて分析を行った。これらのデータは1クリックにつき1レコードであり、表1に示すユーザー情報を含んでいる。アクセス先のページのパーティカル情報を表2に示す。加えて、デバイス情報、性別、生年の全てが含まれているレコードに関して、それぞれの割合を求めた結果を表3に示す。対象としたデバイス情報はスマートフォン端末、タブレット端末、パソコンであり、本論文では sp, tab, pc と表記する。

表1 データに含まれるユーザー情報

ブラウザの ID
ユーザーの性別
生年
職種
業種
検索を実行した UNIXTIME
IP からの位置情報
検索時のクエリ文字列
アクセスを行ったデバイス種別
アクセス先のパーティカルの種別
アクセス先の URL
クリックを実行した UNIXTIME

表2 パーティカル一覧

画像
知恵袋
動画
レシピ
地域情報 (Yahoo!JAPAN ロコ)
ニュース記事
ショッピング
オークション
リアルタイム検索
タレント情報
地図
辞書

デバイスの利用率に関しては、1章で紹介した Google の公式ブログの発表からスマートフォンと PC の検索差は僅かであると予測していたが、ヤフー株式会社の検索サービスを用いた検索ではスマートフォンが PC の約 1.44 倍にもなっていることが確認できた。女性の PC 利用率が少ないことがこの差を生んでいることが分かる。アプリやユーザー登録の有無によってこの割合は左右されると考えられる。世代の利用率に関しては、若い世代の利用率が高いのではないかと予測していたが、

(注2): <http://www.yahoo.co.jp>

10代以下と20代を合計しても30代、40代よりも利用率は低い結果となった。

#### 3.2 分析方法

本論文では以下の仮説に基づいて分析を行った。

- 使用するデバイスや世代、性別によって情報要求が異なる。
  - クエリの単語数や含まれる品詞によって利用されるパーティカルの割合が異なる。
  - 検索をする時間帯によって利用されるパーティカルに差がある。
  - 平日と休日とでは情報要求や検索をする時間、およびユーザーが異なり、利用されるパーティカルに差がある。
- 今回の分析ではこれらの仮説を検証するため、アクセス先のパーティカルが判明している検索ログデータを用いて以下の分析を行った。
- 各属性によるパーティカルのクリック頻度。
  - パーティカル毎の利用ユーザーの割合。
  - クエリのターム数別のパーティカルの割合。
  - クエリに含まれる品詞によるパーティカルの割合。
  - 平日と休日のクリックされるパーティカルの差。
  - 時間によるパーティカルのクリック数。
  - SERP (Search Engine Result Page) 内におけるパーティカルからパーティカルへの遷移確率。

品詞に関する分析では、IPAdic 辞書<sup>(注3)</sup>と形態素解析器 MeCab<sup>(注4)</sup>を使用し、クエリを形態素に分割した際に判定される品詞を用いた。

ここでは SERP を id と timestamp が同一のレコードの集合と定義する。そのうち、1ページのみ閲覧している SERP は 89.79%、2ページ閲覧している SERP は 7.99%、3ページ以上閲覧している SERP は 2.22%であった。

### 4. 分析結果

#### 4.1 パーティカルのクリック頻度分析

クリックされたパーティカルの割合を表4示す。

続いてデバイス毎、性別毎、世代毎に割合を算出した結果をグラフ化したものを図1, 2, 3にそれぞれ示す。ここではデバイス、性別、世代の各属性のデータが存在したレコード総数を全体数として割合を求めた。

結果から見える特徴的な点を以下にまとめた。

- PC は買い物系や地図の閲覧率が高く、スマートフォンでタレント関連のページを閲覧する頻度が他よりも高く、タブレットはレシピや動画の利用率が高い。
- 女性は男性よりもレシピサイトの利用率が高く、逆に男性は女性よりもオークションの利用率が高い。
- 年齢が若ければ若いほど知恵袋やリアルタイム検索を利用する割合が高く、逆にショッピングやオークションは年齢が高ければ高いほど利用する割合は高い。

(注3): <https://osdn.jp/projects/ipadic> .

(注4): <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> .

表 3 検索ログの各属性の割合

	男性			女性			合計
	sp	tab	pc	sp	tab	pc	
10代	0.87%	0.05%	0.42%	1.18%	0.05%	0.26%	2.84%
20代	6.82%	0.24%	3.39%	8.38%	0.17%	1.67%	20.67%
30代	8.71%	0.67%	5.24%	10.61%	0.60%	2.70%	28.53%
40代	7.02%	1.17%	7.87%	6.56%	0.82%	3.87%	27.31%
50代	2.50%	0.69%	5.63%	1.97%	0.43%	2.35%	13.57%
60代	0.81%	0.37%	4.14%	0.45%	0.19%	1.13%	7.08%
合計	26.73%	3.19%	26.69%	29.14%	2.26%	11.98%	100.000%
	56.62%			43.38%			

表 4 クリックされたパーティカルの割合

パーティカル	割合
辞書	0.71%
タレント情報	1.25%
地図	1.39%
オークション	2.50%
ショッピング	4.18%
ニュース	4.60%
地域情報	5.17%
リアルタイム検索	5.92%
レシピ	6.85%
動画	11.46%
知恵袋	15.44%
画像	40.52%

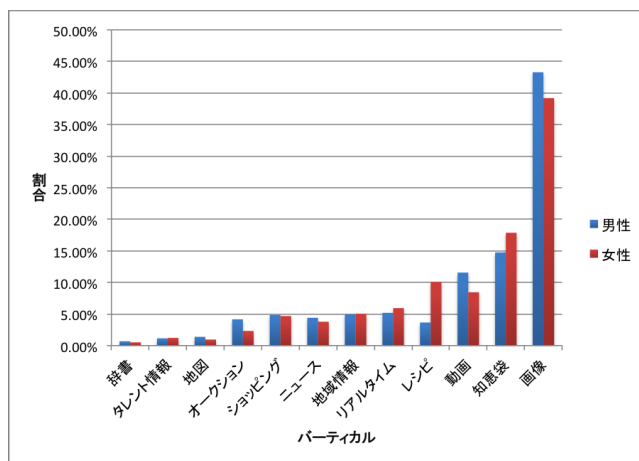


図 2 性別毎のクリックされたパーティカルの割合

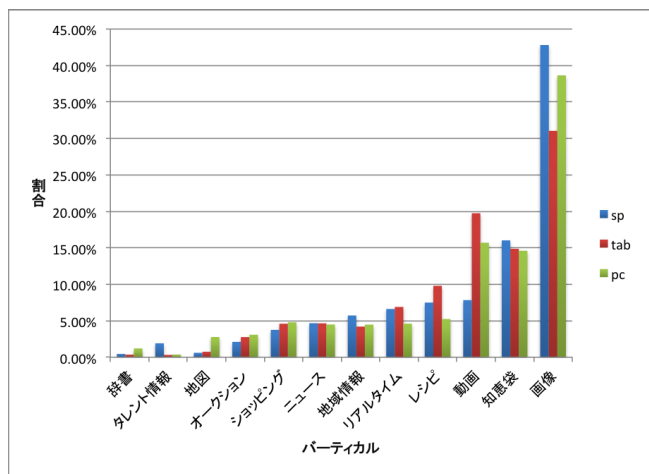


図 1 デバイス毎のクリックされたパーティカルの割合

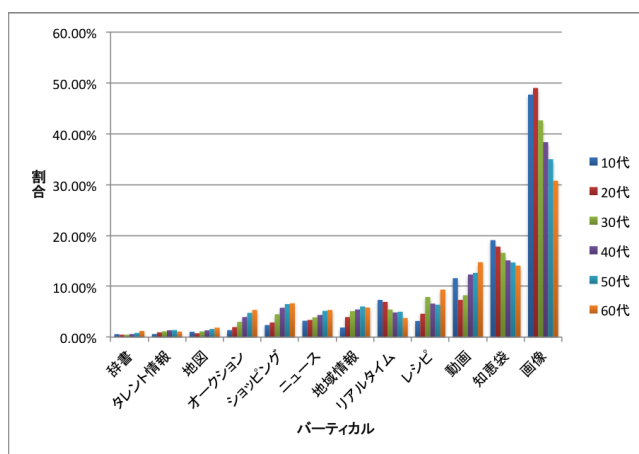


図 3 世代毎のクリックされたパーティカルの割合

デバイスごとの結果を見ると、少なからずデバイスの特徴が反映されていると考えられる。画面の大きなPCやタブレットで動画サイトを利用する割合が高く、防水機能付きも多く存在し、持ち運びが容易であるため、キッチンでも使いやすいスマートフォンやタブレットでレシピサイトの利用が多い、などがそれに当たる。

ショッピングやオークションは若い世代では少ない傾向にあるが、スマートフォンやタブレットの割合がPCよりも少ない事を考慮すると若い世代はそれらに特化したアプリを利用しているのではないかとの仮説も考えられる。しかし、レシピサイ

トなどは専用のアプリがあるにも関わらずそのような傾向は見られないため、さらなる分析が必要であると考えられる。男女の結果ではレシピにおける差は料理をする割合が女性のほうが多い<sup>(注5)</sup>ことから想像がつくが、オークションに限ってはなぜこれほど男女に差があるのか生のデータのクエリを分析しても傾向を掴むことはできなかった。

#### 4.2 クエリのターム数別のパーティカルの分析

クエリのターム数の割合を表5、ターム数別にクリックされ

(注5): “レシピに関する意識調査”, <http://release.center.jp/2009/03/0602.html> .

たパーティカルの確率を図4にそれぞれ示す。ここでクエリのターム数とは、クエリ文字列を全角、もしくは半角スペースによって区切った時の長さとする。結果から見える特徴的な点を以下にまとめた。

- 動画はターム数が多くなると検索される割合が増える。
- 知恵袋はターム数が1の場合利用される割合が低い。
- レシピはターム数が3の場合一番利用される割合が多く、5以上の場合最も少なくなる。

特徴が発生する具体的なクエリの傾向を分析するため、生のデータを見ると、以下の様な傾向が確認できた。

動画はターム数が1の場合はアーティスト名、楽曲名、番組名、アニメタイトルが多く、ターム数が多くなってもそれらに「動画」や「アニメ」、「CM」といった語が付加されたり、組み合わせたりして検索されている物が多く確認できた。また、ターム数が4以上では英字のバンドや楽曲名、洋画のタイトルでターム数が増えている物が多い傾向にあった。

知恵袋はターム数が1の場合は有名人や組織の名前、楽曲やアニメ、番組のタイトルが多い傾向が見られた。その他にも病名や職業名なども他のパーティカルに比べ多く確認できた。ターム数が増加すると、価格や商品の比較を目的としていると思われるクエリや、製品やwebサービスの使い方に関するクエリも見受けられる。その他にも悩みに関すると思われるクエリが増える傾向が見られるが、クエリの表層から判断するのは難しいと考えられる。

レシピサイトは「食材名(+食材名)+「レシピ」」や「料理名(+食材名)+「作り方」(オススメ)」といったクエリが多く、ターム数が1の場合の多くが料理名で検索している傾向にあった。ターム数の多いクエリは使いたい材料名を羅列しているものや「食材名+「レシピ」(料理名)+「人気」+「1」+「位」というクエリが頻出していることが確認できた。

ターム数が多くなっているものの中には、英字以外にも番組名やタイトルなどを形態素に分割して検索しているものなども存在した。

表5 クエリのターム数の割合

クエリのターム数	1	2	3	4	5以上
割合	47.01%	40.89%	10.54%	1.29%	0.27%

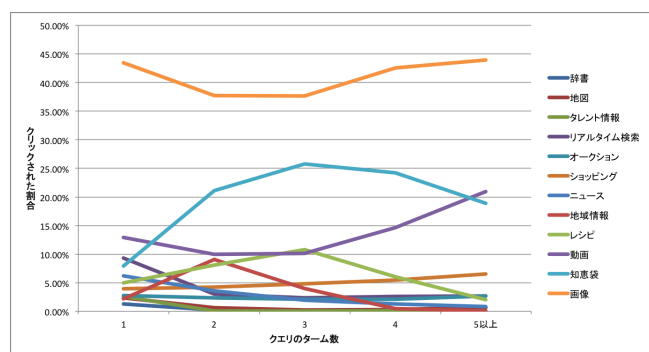


図4 クエリのターム数別のパーティカルの割合

#### 4.3 クエリに含まれる品詞によるパーティカルの分析

クエリ文字列を形態素解析し、その中に動詞、形容詞、助詞が含まれている場合にクリックされるパーティカルの割合を算出した。算出した結果を図5に示す。

結果から見える特徴的な点を以下にまとめた。

- 動詞が含まれるクエリにおいてレシピの割合が多い。
  - 形容詞が含まれるクエリにおいて画像、知恵袋の割合が多い。
  - 助詞が含まれるクエリにおいて辞書と動画の割合が多い。
- データを分析したところ、レシピにおいては「焼き方」、「取り方」といった調理方法や「～に合う」といった文脈で動詞が利用される割合が多いことが確認できた。辞書においては「～とは」というクエリが大多数を占めており、助詞の結果に大きく関わっていると考えられる。また、動画は番組や曲のタイトルで検索されることが多いため、助詞が含まれている割合が多いと考えられる。形容詞が含まれる割合が高い画像や知恵袋は他の品詞においても検索される割合が多いため、特徴的なクエリは確認できなかった。

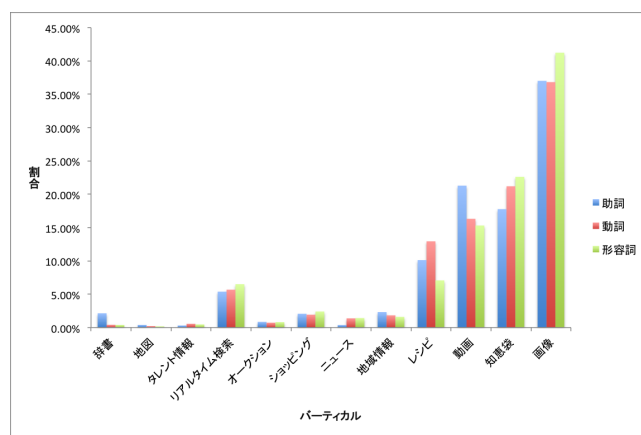


図5 クエリに含まれる品詞別のパーティカルの割合

#### 4.4 時間毎のクリック数分析

時間毎にレコードを集計し、全検索数のヒストグラムとパーティカル毎にまとめたヒストグラムを図6,7にそれぞれ示す。横軸は時間を表し、値の時間0分0秒から59分59秒までのクリック数を縦軸の値として取っている。

結果から見える特徴的な点を以下にまとめた。

- 全体の検索数は4時台にクリック数が最少となり、22時台に最大となる。
- 多くのパーティカルで全体の検索数と同様の推移をするが、レシピのみ他と大きく異なる。
- 全体の検索数と多くのパーティカルで12時台に山が存在する。
- レシピは夕食の準備時(15時以降)には増加するが、朝食昼食時にはあまり増加しない。

レシピ以外のパーティカルは仕事が終わる夜中や昼休みを迎える12時頃にクリック数が増える傾向にあることが確認できた。時間毎の特徴を得るため生のデータを分析したが、時間帯による特徴的な検索ワードを見つけることは出来なかった。

この結果から現段階で述べられる知見は、時間によってクリック数がグラフに示した様に変動する、ということだけである。

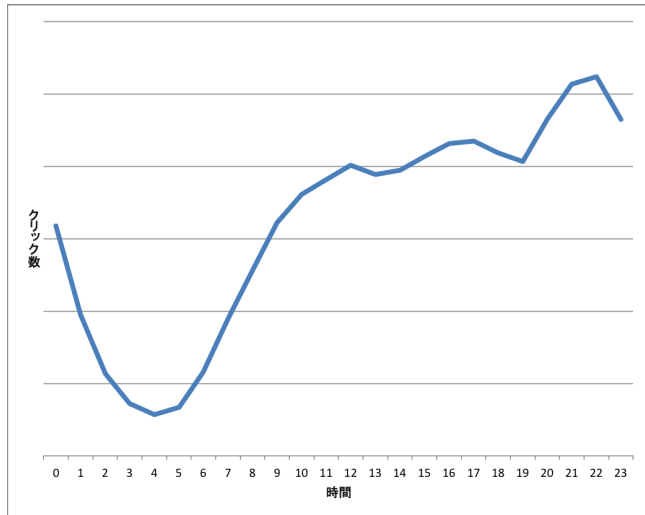


図 6 クリック数のヒストグラム

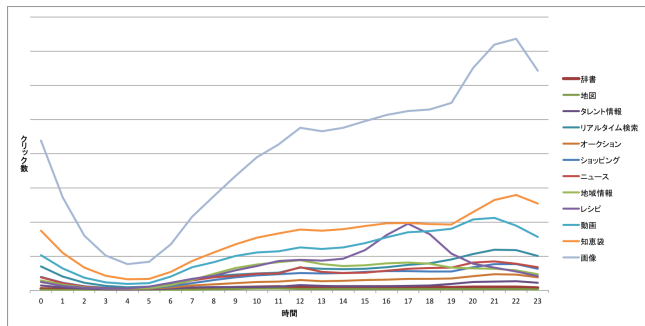


図 7 クリック数のパーティカル別のヒストグラム

#### 4.5 平日と休日の差の分析

平日と休日では情報要求や検索をするユーザーが異なるという仮定のもと分析を行った。クリックされたパーティカルを平日と休日に分けて集計した結果を表 6 に示す。一日あたりのクリック数は休日が平日の 1.18 倍であることが確認できた。

表 6 クリックされたパーティカルの割合の比較

パーティカル	平日の割合	休日の割合
辞書	0.76%	0.61%
タレント情報	1.24%	1.26%
地図	1.55%	1.02%
オークション	2.46%	2.60%
ショッピング	4.11%	4.34%
ニュース	4.88%	3.98%
地域情報	5.31%	4.86%
リアルタイム検索	5.86%	6.05%
レシピ	6.85%	6.85%
動画	11.04%	12.42%
知恵袋	15.42%	15.49%
画像	40.52%	40.52%

クリックされたパーティカルに関しては以下のことが確認で

きた。

- 多くのパーティカルでは大きな差はなく、レシピ、画像のレコードの割合はほぼ同じとなった。

- 平日に地図、ニュース、地域情報のクリック数が多くなり、その代わり休日には動画、リアルタイム検索、ショッピング、オークションのクリック数が多い。

生のデータを確認し、分析を行ったが、これらの差について特徴的な知見を得ることはできなかった。平日と休日のクリック率の差が有意なものであるか否かを検証するために有意水準 5%として統計的検定を行った。p 値と差の信頼区間を表 7 に示す。この結果からレシピ、画像以外のパーティカルにおいて平日と休日の差は有意であると言え、レシピと画像における平日と休日のクリック率の母集団に差があるとは言えないことが確認できた。

表 7 平日と休日のパーティカル毎のクリック率の差の検定結果

パーティカル	p 値	差の信頼区間
辞書	0	$[1.49 \times 10^{-3}, 1.56 \times 10^{-3}]$
地図	0	$[5.32 \times 10^{-3}, 5.40 \times 10^{-3}]$
タレント情報	0	$[-2.67 \times 10^{-4}, -1.81 \times 10^{-4}]$
リアルタイム検索	0	$[-1.98 \times 10^{-3}, -1.80 \times 10^{-3}]$
オークション	0	$[-1.56 \times 10^{-3}, -1.44 \times 10^{-3}]$
ショッピング	0	$[-2.40 \times 10^{-3}, -2.25 \times 10^{-3}]$
ニュース	0	$[8.94 \times 10^{-3}, 9.10 \times 10^{-3}]$
地域情報	0	$[4.39 \times 10^{-3}, 4.56 \times 10^{-3}]$
レシピ	0.676	$[-7.74 \times 10^{-5}, 1.19 \times 10^{-4}]$
動画	0	$[-1.39 \times 10^{-2}, -1.37 \times 10^{-2}]$
知恵袋	0	$[-8.55 \times 10^{-4}, -5.73 \times 10^{-4}]$
画像	0.684	$[-1.52 \times 10^{-4}, 2.31 \times 10^{-3}]$

次に、各属性におけるクリック数の割合を平日と休日に分けて集計した結果を表 8, 9 にそれぞれ示す。

利用するユーザーの属性においては世代、性別に関しては目立った差は確認できなかった。利用するデバイスにおいて休日のほうが小型デバイスを利用する傾向が見られた。外出先かどうかのデータも利用できるのであれば合わせて分析を行うのも今後の課題として上げられる。

最後に、時間毎のクリック数も平日と休日に分けて集計した。その結果を図 8, 9 からさらにパーティカル別に集計した結果を図 9, 10 に示す。ここでグラフに示したヒストグラムは一日あたりのクリック数を縦軸の値としている。

ヒストグラムの差について以下にまとめた。

- 全体的に休日のほうがクリック数が多いが 12 時台のみ平日が休日を上回っている。

- パーティカル毎の結果においては休日の 11 時台にレシピと知恵袋がピークを作っている。

12 時台のピークには会社や学校の昼休みが関連していると考えられる。休日の 11 時台にレシピと知恵袋がピークを迎えているので昼食のために利用するユーザーが多いと考えられる。

表 8 平日の検索ログの各属性の割合

	男性			女性			合計
	sp	tab	pc	sp	tab	pc	
10代	0.83%	0.05%	0.40%	1.12%	0.05%	0.24%	2.71%
20代	6.82%	0.24%	3.36%	8.34%	0.16%	1.70%	20.62%
30代	8.57%	0.64%	5.40%	10.57%	0.58%	2.87%	28.63%
40代	6.88%	1.10%	8.06%	6.47%	0.79%	4.02%	27.32%
50代	2.44%	0.65%	5.67%	1.93%	0.42%	2.40%	13.50%
60代	0.81%	0.36%	4.27%	0.45%	0.19%	1.16%	7.23%
合計	26.34%	3.03%	27.17%	28.89%	2.19%	12.38%	100.00%
	56.54%			43.46%			

表 9 休日の検索ログの各属性の割合

	男性			女性			合計
	sp	tab	pc	sp	tab	pc	
10代	0.97%	0.06%	0.45%	1.31%	0.06%	0.28%	3.14%
20代	6.84%	0.26%	3.44%	8.45%	0.17%	1.61%	20.77%
30代	9.02%	0.74%	4.89%	10.69%	0.62%	2.33%	28.30%
40代	7.34%	1.32%	7.44%	6.77%	0.89%	3.55%	27.30%
50代	2.63%	0.79%	5.55%	2.05%	0.46%	2.26%	13.74%
60代	0.82%	0.38%	3.84%	0.45%	0.20%	1.07%	6.75%
合計	27.61%	3.56%	25.62%	29.71%	2.40%	11.10%	100.00%
	56.78%			43.22%			

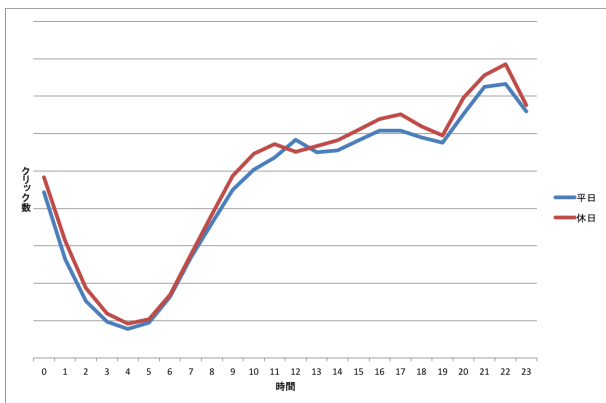


図 8 平日と休日のクリック数のヒストグラム

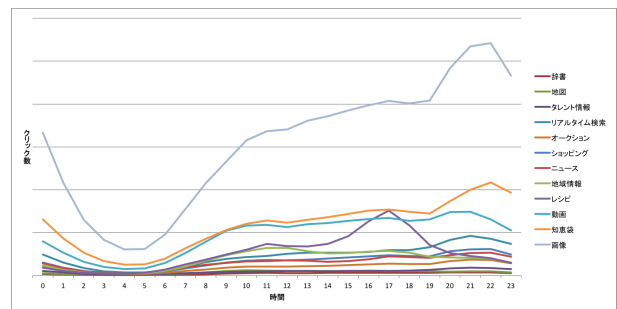


図 10 休日におけるパーティカル別クリック数のヒストグラム

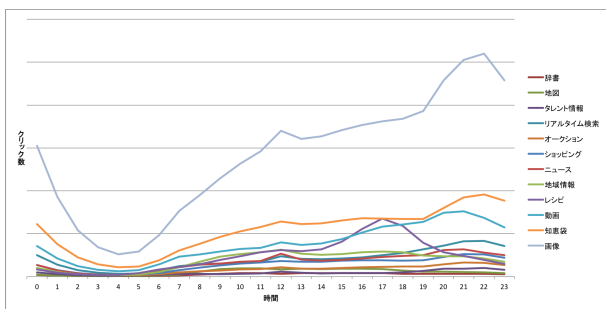


図 9 平日におけるパーティカル別クリック数のヒストグラム

#### 4.6 パーティカルの遷移確率

パーティカルからパーティカルへの遷移確率を表 10 に示す。この表は行のパーティカルから列のパーティカルへの遷移確率を表している。ここで「-」は一度も遷移しなかった事を表し、「0.00%」は遷移はしているが確率が 0.01%未満であることを表す。太字になっている部分が遷移確率が最大のパーティカルを表す。

辞書，リアルタイム検索，タレント以外のパーティカルでは前のパーティカルを続けてアクセスすることが多く，それ以外では知恵袋や画像検索に遷移する確率が高いことが確認できた。これらの他に SERP 内のクリック数やパーティカルの種類においても同様にデバイス，性別，世代による比較も行ったが，ほとんど差は出ず，2 ページ以上閲覧している場合の平均閲覧ページ数は約 2.3 ページでパーティカルの種類は約 1.2 種類であった。

表 10 SERP 内のパーティカルの遷移確率

	その他	辞書	地図	リアルタイム検索	タレント	オークション	ショッピング	ニュース	地域情報	レシピ	動画	知恵袋	画像
その他	85.22%	-	-	1.47%	0.10%	0.05%	-	6.88%	-	-	0.39%	1.47%	4.42%
辞書	-	4.72%	0.65%	0.58%	-	2.30%	2.74%	3.20%	-	0.84%	2.59%	43.43%	38.94%
地図	-	0.11%	82.63%	0.50%	-	0.04%	0.02%	1.54%	0.00%	0.00%	0.04%	3.30%	11.82%
リアルタイム検索	-	0.31%	0.14%	40.48%	0.81%	2.18%	0.74%	15.98%	0.05%	0.01%	6.47%	43.43%	10.02%
タレント	-	-	-	2.95%	18.32%	0.22%	0.01%	2.48%	-	0.01%	3.07%	2.89%	70.05%
オークション	-	0.07%	0.00%	0.21%	0.03%	70.09%	3.67%	0.55%	0.00%	0.07%	0.45%	17.51%	7.35%
ショッピング	-	0.03%	-	0.05%	0.00%	18.90%	67.53%	0.24%	0.01%	0.23%	0.11%	9.36%	3.54%
ニュース	-	0.18%	0.13%	9.38%	1.21%	0.77%	0.53%	61.19%	0.12%	0.10%	4.32%	12.28%	9.79%
地域情報	-	0.00%	-	0.01%	-	0.00%	0.02%	0.08%	96.51%	0.00%	0.00%	3.28%	0.09%
レシピ	-	0.02%	-	0.01%	0.00%	0.01%	0.20%	0.14%	0.02%	97.47%	0.00%	1.82%	0.30%
動画	-	0.04%	0.00%	0.69%	0.07%	0.50%	0.35%	1.55%	0.05%	0.02%	86.85%	4.51%	5.38%
知恵袋	-	0.05%	0.01%	0.21%	0.04%	0.20%	0.19%	0.59%	0.00%	0.01%	0.25%	97.33%	1.11%
画像	-	0.75%	0.10%	3.55%	0.17%	2.63%	3.71%	5.74%	0.02%	0.21%	5.65%	32.37%	45.10%

## 5. まとめと今後の課題

本論文ではコンテキスト付き検索ログデータの分析を行った結果について述べた。ユーザー情報やコンテキストに着目した分析の結果、利用されるサイトのパーティカルの順位を始め、デバイスの特徴を反映している知見が得られた。クエリに着目した分析では、クエリに含まれる品詞やクエリのターム数によるパーティカルの傾向も確認できた、時間に着目した分析では、時間によってクリック数は大きく変動し、4 時台に最小値、22 時台に最大値を出すことが確認できた。時間によるクリック数の推移については、絶対数の差はあるもののレシピ以外のパーティカルで同様な推移を示すことも確認できた。また、平日と休日に着目した分析も行った。大きな差は確認できなかったが、12 時台のみ休日より平日のほうが検索数が多いという興味深い結果が現れることが確認できた。最後に SERP 内におけるパーティカルの遷移についても分析を行ったが、多くのパーティカルでは遷移せずに同じパーティカルを利用する傾向が確認できた。

今後の課題としては、生のデータを分析し、さらに詳しい傾向の分析が挙げられる。また、ユーザーの位置情報登録の許可がない場合、許可を得た直近の位置情報がレコードに登録される仕様であったため正確性がないと判断し、今回の分析には位置情報を用いなかった。正確性が担保できる位置情報や職種、季節などに着目したクエリデータの分析や時間とユーザー情報やコンテキストを複合した状態での多面的な分析も今後の課題であると言える。さらに「ワンピース」という単語には衣服の種類と漫画の題名の 2 つの意味があり、それぞれでショッピングと動画、画像とユーザーの要求するパーティカルが異なる可能性がある。このような多義語がクエリに含まれている場合の要求パーティカルがコンテキストによって判定できるか否かも課題であると考えられている。

最終的には、これらの分析の知見を踏まえ、与えられたコンテキスト付きのクエリに対し、適切なパーティカルを推薦するシステムの構築を目指す。

## 文 献

- [1] Arguello, Jaime, et al. "Sources of evidence for vertical selection." Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009.
- [2] 中渡瀬 秀一, 大山 敬三, "サーチエンジンクエリ分析による情報タイプの抽出: Web 検索利用者の情報要求に即した Web 情報空間の再構成に向けて", 人工知能学会全国大会論文集 25, 2011.
- [3] Jansen, Bernard J., et al. "Real life information retrieval: A study of user queries on the web." ACM SIGIR Forum. Vol. 32. No. 1. ACM, 1998.
- [4] Jansen, Bernard J., Mimi Zhang, and Amanda Spink. "Patterns and transitions of query reformulation during web searching." International Journal of Web Information Systems 3.4 2007.
- [5] Jansen, Bernard J., Danielle L. Booth, and Amanda Spink. "Patterns of query reformulation during Web searching." Journal of the american society for information science and technology 60.7 2009.
- [6] Thomas Demeester, Dolf Trieschnigg, Dong Nguyen, Ke Zhou, Djoerd Hiemstra, "Overview of the TREC 2014 Federated Web Search Track", 2014.
- [7] Jin, Shan, and Man Lan. "Simple May Be Best-A Simple and Effective Method for Federated Web Search via Search Engine Impact Factor Estimation." TREC. 2014.
- [8] Emanuele Di Buccio, Massimo Melucci, "University of Padua at TREC 2014: Federated Web Search Track", 2014.
- [9] Feng Guan, Shuiyuan Zhang, Chunmei Liu, Xiaoming Yu, Yue Liu, Xueqi Cheng, "ICTNET at Federated Web Search Track 2014", 2014.