

携帯電話検索ログデータに基づく行動特性の分析 Analysis of User's Behavior Based on Log Data of Their Internet Searches Through Mobile Phones

鎌形 孟[†] 徳永 幸生[†] 杉山 精[‡] 貝谷 實榮* 木村 義彦*
Hajime Kamagata Yukio Tokunaga Kiyoshi Sugiyama Jitsuei Kaitani Yoshihiko Kimura

1. はじめに

一般に検索行動において、ユーザは求める情報に関する検索語を Web 検索システムに入力し、その検索結果を踏まえ、再度検索語の追加・変更による試行錯誤を繰り返しながら求める情報を探す。従って、ユーザの検索行動の履歴である検索ログデータには利用者の情報要求の生の声が反映されていると考えられる。

パソコンによる Web 検索行動に関しては、Web 検索システムから未知の情報を検索する時の行動を、情報を得るために使用された検索語間の使用時間間隔に基づいて分析し、検索語間の関連性を抽出する研究がある。この研究では検索語間の関連度から、検索行動の背景に潜む情報ニーズを探る試みがなされている[1]。

近年では携帯電話の発展・普及に伴い携帯電話から Web 検索を行うユーザも増えてきた。検索行動をみると、10年前は 1 語での検索が主流であったのに対し、現在では 2 語以上を組み合わせた AND 検索が多数行われるようになっている。携帯電話における検索行動を考えた時、パソコンと大きく異なる点として検索行動を行うシチュエーションの違いがある。それは携帯電話には屋内での使用にとどまらず、電車内や、待ち合わせの間、歩行中などといった多様な利用シーンがあることや、テレビの視聴中のクロスメディアとしても使用されていることである。

本報告ではパソコンとの違いを念頭に置きつつ、携帯電話検索ログデータからユーザの検索の使用時間間隔の分布特性を基に AND 検索ログデータと時間差のあるログデータに分け、それぞれにおいて携帯電話における検索語間の関連を分析した。

2. 携帯電話における使用時間間隔と検索回数

2009 年 2 月の 1 ヶ月分の携帯電話検索のログデータを対象に分析を行った。ユーザごとに検索が行われてから次の検索までの時間差を求めた。その回数分布を図 1 に示す。

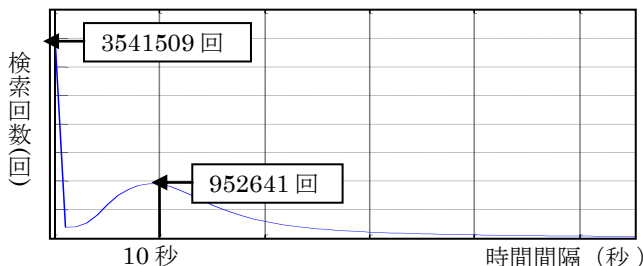


図 1 携帯電話における使用時間間隔と検索回数のグラフ

[†] 芝浦工業大学, Shibaura Institute of Technology

[‡] 東京工芸大学, Tokyo Polytechnic University

* エフルート株式会社, froute Corporation

図 1 において AND 検索は 0 秒として考え、それ以外は前後の検索行動の時間差ごとに、検索行動が行われた回数を表示している。図 1 からは 10 秒付近で 2 回目以降の検索行動のピークが見られる。

3. 分析手法

3.1 AND 検索ログデータにおける関連度分析手法

携帯電話検索において AND 検索は全体(21477859 件)の 16%(3541509 件)を占める。パソコンにおける AND 検索の回数は近年増加していることから、携帯電話検索においても今後 AND 検索回数は増加していくと考えられる。AND 検索ログデータの中のある検索語について、他の検索語との AND 検索回数を求める。検索語 x と検索語 m が AND 検索された回数を T_{xm} とし、他の検索語に関しても同様に求める。それらを用いて要素 n 個の検索語 x の特徴ベクトルを $W_x=(T_{x1}, \dots, T_{xm}, \dots, T_{xn})$ と定義する。検索語 y についても同様に $W_y=(T_{y1}, \dots, T_{ym}, \dots, T_{yn})$ とすれば両者の検索語の類似度は cos 関連度を用いて算出できる。

$$\cos(x, y) = \frac{T_{x1} \times T_{y1} + T_{x2} \times T_{y2} + \dots + T_{xn} \times T_{yn}}{\sqrt{T_{x1}^2 + T_{x2}^2 + \dots + T_{xn}^2} \times \sqrt{T_{y1}^2 + T_{y2}^2 + \dots + T_{yn}^2}}$$

3.2 時間間隔に基づく関連度分析手法

図 1 の 0 秒を除く検索行動について、時間差に基づく分析を行った大久保らの手法を適用する。大久保らはまず、パソコンにおける検索行動を図 2 のように 3 段階の行動モデルで表した。

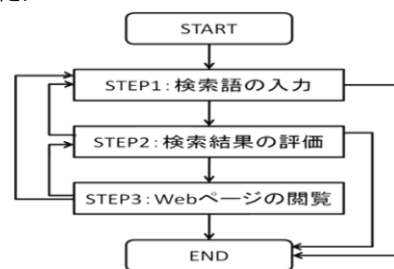


図 2 パソコンにおける検索行動モデル

STEP2 で評価する検索結果には、Web ページのタイトルや概要が含まれ、実際に Web ページを閲覧する前にある程度内容が推測できる。このため STEP1 - STEP2 間は短い時間間隔で頻繁に検索行動が繰り返される。STEP3 に至ると実際に Web ページを閲覧するため長い時間間隔での検索行動となる。これらを踏まえて時間間隔に基づく assoc 関数を定義して検索語同士の関連度を算出した。

本報告では assoc 関数を携帯電話用に再定義した。但し、0 秒は AND 検索のため除外した。20 秒までは検索語の追加変更による検索行動、すなわち同じ目的での検索行動で

ある。20秒から30秒までは目的に対する検索語の比重が小さい。30秒以降は別の目的で検索が行われる(図3)。

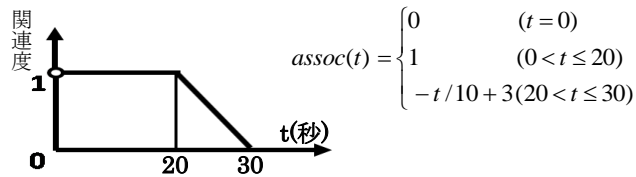


図3 携帯電話用 assoc 関数

4. 結果と考察

4.1 AND 検索ログデータにおける関連度分析

AND 検索ログデータにおける検索語「芝浦工業大学」の結果を示す。表1は「芝浦工業大学」と cos 関連度が高い検索語3語の特徴ベクトルの表である。

表1 「芝浦工業大学」と cos 関連度の高い検索語 (回)

検索語要素	芝浦工業大学	東京理科大	女子美術短期大学	愛知大学
入試	7	7	4	8
合格発表	4	1	0	5
後期日程	1	0	0	0
解答	0	1	0	0
合格	0	1	0	0
倍率	0	0	0	8
120億	0	0	0	1
移転	0	0	0	1
cos 関連度	-	0.904	0.861	0.744

「芝浦工業大学」と cos 関連度が高い検索語には「東京理科大」や「女子美術短期大学」、「愛知大学」といった学校名が得られた。表1からどの検索語も「入試」という検索語と AND 検索されていることが分かる。これらの検索語の示す学校は、2月に試験日があり「入試」や「合格発表」のような検索語とともに AND 検索されている。

次に AND 検索ログデータにおける「九州大学」の分析結果を表2に示す。

表2 「九州大学」と cos 関連度の高い検索語 (回)

検索語要素	九州大学	徳島大学	京都大学	信州大学
柔道部	2	0	0	0
学費	1	0	0	0
出願状況	1	7	2	1
出願	0	1	0	0
理学部	0	1	0	0
芝浦	0	1	0	1
cos 関連度	-	0.408	0.384	0.288

「九州大学」はログデータに残る検索回数が少ない。よって特徴ベクトルの要素が少なく、値も小さいために、「芝浦工業大学」と比較すると、cos 関連度が一番高いものであっても0.4程度の低い値が算出されていた。表2からは「九州大学」の関連語は「出願状況」と AND 検索されているという検索傾向が分かった。これは、国立大学は試験日が私立大学と比較して遅いため、この時期に学生が受験する学校の情報を検索しているためだと考えられる。

今回、AND 検索ログデータの分析結果から検索傾向の似ている検索語同士の関連が得られた。学校名の例に示し

たように、学校名というグループの中の検索語であっても、期間によっては検索傾向がそれぞれ異なり、グループの中でも検索傾向により、さらに細分化できると考えられる。

一方で、関連語の中には検索回数が1回のように極端に少ないものが存在する。この時、cos 関連度は特徴ベクトルの要素の一部が合致した場合でも高く算出される。より精度の高いグループ分けには、より長い期間のログデータで検索傾向を分析する必要がある。

4.2 時間間隔に基づく関連度分析

検索語「無料」に関して、時間間隔に基づく分析結果と AND 検索ログデータを用いた分析結果の関連語上位10語を対比させたものを表3に示す。

表3 「無料」の関連語

AND 検索		時間間隔関連度	
検索語	検索回数(回)	検索語	関連度
デコメール	139	浜崎あゆみ	33.2
HP	91	相武紗季	30
素材	89	木下優樹菜	26
携帯	65	女子高生	25.5
ダイエット	62	綾瀬はるか	23.6
着うたフル	56	谷村奈南	22.7
着うた	54	中学生	21.9
画像	51	長澤まさみ	21.4
動画	46	優木まおみ	20.8
フル	44	ほしのあき	20.5

「無料」について、AND 検索ログデータを用いた分析結果では特徴ベクトルの要素に、「画像」や「着うた」といった検索語が含まれる。この結果から、AND 検索において「無料」を入力するユーザは、「無料」の「デコメール」や「着うた」といった携帯電話向けのコンテンツやアイテムを探す目的を持っていると推定される。AND 検索において共に用いられる検索語は、求める情報を具体的に表しており、入力された検索語からユーザの検索行動の目的を直接くみ取ることができる。

一方、時間間隔に基づく結果では、歌手やアイドルのような人名の検索語が「無料」の特徴ベクトルの要素に含まれた。この結果は、ユーザが一度の入力で検索行動を終了し、その後異なる目的をもった行動、すなわち、画面内のリンクを辿るような行動の結果であると推定される。時間間隔に基づく分析の結果から、携帯電話における10秒付近の検索行動では、アイドル名を入力し画像を眺めるような、散漫とした行動が多く行われると考えられる。

5. おわりに

本報告では携帯電話におけるユーザの検索行動を、ログデータを用いて分析した。AND 検索ログデータの分析からは検索傾向の類似している検索語同士の関連を得ることができた。また、前後の検索行動の時間間隔に基づく分析からは携帯電話における10秒付近での検索行動には目的を持たない散漫とした行動が行われることが示唆された。

参考文献

- [1] 大久保雅且, 井上孝史, 杉崎正之, 田中一男, “www 検索ログに基づく情報ニーズの抽出”, 情報処理学会論文誌, vol.39, No.7, (1997).