

# 携帯電話検索における検索語間の COS 関連度の可視化と分析

鎌形 孟<sup>†</sup> 徳永幸生<sup>†</sup> 杉山精<sup>‡</sup> 貝谷 實榮\* 木村 義彦\*

芝浦工業大学大学院 工学研究科<sup>†</sup> 東京工芸大学<sup>‡</sup> エフルート株式会社\*

## 1. はじめに

一般に検索行動においては、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、得られた検索結果を踏まえ、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ユーザの情報要求の生の声が反映されていると考えられる。PC を用いた Web 検索システムでは、検索ログデータを分析することで、検索行動の背景に潜む情報ニーズを探り、そこで得られた知見をもとにユーザの情報要求に関連する、他の選択肢を提示する等のサービスが試みられている[1]。

一方、近年では携帯電話の発展に伴い携帯電話から Web 検索を行うユーザの増加が目覚ましい。しかし、携帯電話は既存のコンピュータ機器と比較して性能や用途が大きく異なるメディアであり、携帯電話の検索ログデータに着目した研究は少ない。そこで、携帯電話検索で入力された検索語からユーザの情報要求を推測し、適切な支援方法を開発することを目的に、検索傾向による検索語のグループ化を行った[2]。本報告では、検索語間の関連性を可視化する手法を用い携帯電話検索全体の傾向、使用される検索語の種類や、検索語同士の関連性を分析した。

## 2. 携帯電話の特徴

携帯電話の特徴をパソコンと比較して考えた時、大きく 2 つの違いがある。1 つはデバイスに関わる違いである。画面の大きさや、キーボードやマウスの有無などの入力機器の違い、さらには扱えるデータの種類の違いなどがある。2 つ目は検索行動を行うシチュエーションの違いがある。携帯電話は屋内での使用にとどまらず、電車内や歩行中といった屋外での利用も日常化

Analyzing and Visualizing the Cosine-relationship of Keywords in Mobile Information Retrieval  
Hajime KAMAGATA<sup>†</sup>  
Yukio TOKUNAGA<sup>†</sup>  
Kiyoshi SUGIYAMA<sup>‡</sup>  
Jitsuei KAITANI<sup>†</sup> Yoshihiko KIMURA\*  
Graduate School of Engineering, Shibaura Institute of Technology<sup>†</sup>  
Tokyo Polytechnic University<sup>†</sup>  
froute Corporation\*

していること、片手での操作が容易であるということからテレビの視聴中のクロスメディアとして使用される。このように、携帯電話では、PC の利用形態とは異なる多種多様な利用形態がある。これらは検索行動の違いの要因となりその結果は検索ログデータに反映される。それらを分析することによって、携帯電話により適したサービス開発に向けた知見が得られると考えられる。

## 3. AND 検索ログデータの分析手法

本報告では 2010 年 1 月から 3 月までの携帯電話検索ログデータを対象に分析した。携帯電話の検索行動は一度の検索で終了することが多く、2 回目以降は直前の検索行動とは目的の異なることが多い[3]。そこで 1 クエリごとの検索語、特に AND 検索ログデータを取り上げる。

### 3.1 分析手法

ある検索語において、他の検索語との AND 検索が行われた回数を求める。例えば、検索語  $x$  と検索語  $m$  が AND 検索された回数を  $T_{xm}$  とし、それを他の検索語に関しても同様に求める。検索語  $x$  に着目したとき、AND 検索回数を要素数とした検索語  $x$  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  $W_y = (T_{y1}, \dots, T_{ym}, \dots, T_{yn})$  とすると検索語  $x$  と検索語  $y$  の検索傾向の類似度は以下の式を用いて算出することができる。これは一般に COS 関連度と呼ばれる。

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

### 3.2 結果の可視化

図 1 に 3.1 で得られた結果の可視化例を示す。

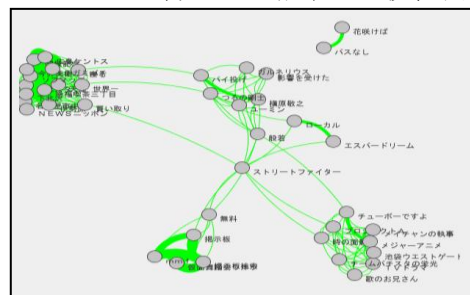


図 1 結果の可視化例

COS 関連度を用い、スプリングモデルでグラフ化し各検索語同士の関連を俯瞰的に表示した。グラフは各検索語をノードとし、エッジの太さを COS 関連度の強弱として表現する。各ノードは、COS 関連度の強い語同士が近くに表示される。

#### 4. 結果の分析と考察

##### 4.1. 軸となる検索語の存在

使用したログデータの時期に放映されたドラマ名や、楽曲名等の検索語 40 語のグラフの分析例を図 2 に示す。図 2 では検索傾向の異なる複数の検索語群があることが分かる。検索語群に含まれる検索語の特徴ベクトルをみると、ドラマやアニメのタイトル名の検索語群ではどの検索語も「主題歌」という検索語を特徴ベクトルの要素に含んでいた。同様に、歌手名や「サビ」といったような検索語群では「着うた」という検索語を特徴ベクトルの要素に持つという結果を得た。このことから、それぞれの検索語群には共通して検索される軸となる検索語が存在することが分かった。「主題歌」を軸とする検索語群では、ドラマやアニメ名の検索語が含まれていた。各検索語の特徴ベクトルの要素をみると、番組のオープニングやエンディングを示す「OP」や「ED」といった検索語が共通に含まれており、それらの検索語も関連の軸となっていた。このことから軸となる検索語は各検索語群の一つ以上存在することがわかった。軸となる検索語を用いることでグループ化された検索語群の検索傾向を把握でき、各検索語の情報要求の推測に有用と考えられる。



図 2 検索語群と軸語

「アルバム」を軸とする検索語群では「シングル」という検索語も軸となり、アーティストや楽曲名の検索語が関連しているが、他にも「オリコン」や「チャート」といった検索語が関連していた。携帯電話において CD アルバムに関する情報要求にはランキング等のエンターテインメントに関する情報要求が存在するという結果を得た。携帯電話検索では、テレビ等の他のメディアの影響が強く反映する傾向があるといえる。

##### 4.2. 軸となる検索語を用いた分析

次に軸となる検索語から、使用される検索語の種類を調べた。携帯電話において画像検索は情報要求が多いジャンルである。特徴ベクトルの要素に「画像」を含む検索語を抽出し、得られた検索語群中の細かな検索傾向の違いを分析した結果を図 3 に示す。

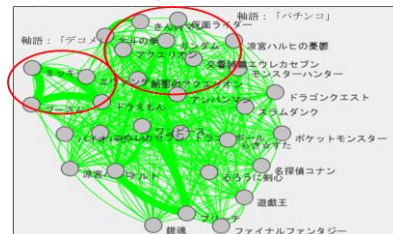


図 3 軸に「画像」を含む検索語群

携帯電話検索では漫画やアニメに関する検索語が多く検索される。しかし、キャラクター名の検索語同士でも「ミッキー」等の様にデコレーションアイテムとして利用されている物や「パチンコ」など検索語を軸として持つ様なキャラクターが二次利用されている検索語は、それぞれの検索傾向に対して軸となる検索語を持ち、画像に検索傾向を持つ検索語群の中でもパチンコやデコメといった複数の検索傾向を持つ検索語群が含まれることが確認できた。この結果から、それぞれの検索語に対して適切に情報要求を把握することができ、入力された検索語からユーザの求める情報を提示するのに役立つことができる。

#### 5. まとめ

本報告では、携帯電話検索において入力された検索語からユーザの情報要求を推測し、適切な支援方法を明らかにすることを目的に、検索語間の関連を可視化する手法を用いて、携帯電話検索全体の傾向を分析した。検索語ごとに検索傾向の似通う検索語群には軸となる検索語が一つ以上存在することが分かった。

今後は軸となる検索語を利用した情報提供手法や時間情報等のログデータに含まれる他の情報とも組み合わせる情報を提供する仕組みを検討する支援方法を考察していく。

#### 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).  
 [2] 鎌形孟, 徳永幸生, 杉山精, 貝谷實榮, 木村義彦, “携帯電話検索ログデータに基づく行動特性の分析”, FIT2010 第 9 回情報科学技術フォーラム, D-026, (Sep 2010)  
 [3] 藤村香央里, 浅野陽子, “携帯電話における検索行動特性”, ヒューマンインタフェースシンポジウム 2007, pp. 115-118