

ユーザの利用履歴に基づく携帯端末上でのブラウジング効率化*

深澤 佑介 長沼 武史 黒宮 寧 倉掛 正治†

NTT ドコモネットワーク研究所‡

1. はじめに

近年、携帯端末からインターネットに接続し、インターネット上のサイトを閲覧する機会が増えている。現状、インターネット上のサイトへたどり着くための手段としては、ポータルサイト（複数の異なるドメインのリンク先を一覧として有するサイト）を利用する場合が一般的である。一般に、このようなポータルサイトは多くのリンクを有しており、ユーザは多数のリンクの中から目的とするリンクを探し出さなければならない。さらに、携帯端末を利用する場合には、画面の大きさの制限上、一度に閲覧可能なリンク数が少なく、目的のリンクが下位に存在する場合には画面スクロールさせなければならない場合があるため、リンク選択の支援がより重要となる。

1.1 関連研究と研究の目的

複数のリンクを有するサイトからリンク選択を支援する手法としては、ユーザの利用履歴を用いてクリックされる確率の高いリンク先を予想し、ユーザに提示する手法が多く提案されている。Internet Explorer[1]などの Web ブラウザではユーザが過去に閲覧したリンク先の色を変更することでユーザにリンク先の候補を提示している。また、Rameshらは、ユーザのリンクをたどる様子をマルコフ連鎖として扱い、ユーザが次に閲覧する可能性が高いリンク先を予測している[2]。Jiyangらユーザの閲覧履歴からユーザが興味をもつトピックをデータマイニング手法により抽出し、同様に閲覧する可能性が高いリンク先の予測を行っている[3]。

しかしながら、携帯端末の場合にはユーザのクリック履歴を用いるなどして予測したリンクをサイト内で上位に表示するだけでは不十分である。なぜなら、携帯端末を用いたブラウジング操作においては、リンクのフォーカス遷移とページ全体の画面スクロールを融合させた操作方式が一般的であり、かつ、厳しい画面制約があるため「選択されたリンク」の情報だけでは、本当は必要としているが画面制約上見ていなかったリンクを「不要なリンク」と判定してしまう可能性があるためである。

そこで本稿では、サイト内のすべてのリンク先について、クリックされる確率（以下、被クリック確

率）を計算し、確率の高いリンクから順に並び替え表示するシステムを提案する。具体的にリンクの被クリック確率は、クリックされたリンクに関しては従来手法と同様に高く設定し上位に表示する。さらにここでは、携帯端末のリンクにユーザがフォーカスを遷移しながらスクロールをするという特徴を生かし、クリックされなかったリンクを「フォーカス遷移された」リンクと「フォーカス遷移されなかった」リンクの2つに分けて処理する。「フォーカス遷移された」にもかかわらずクリックされなかったリンクはユーザにとって不要なリンクと考えられ被クリック確率を低く設定する。これにより、画面制約上見ていなかったリンク（フォーカス遷移されなかったリンク）を上位に表示させることが可能となる。詳細を3章で述べる。

2. システム構成

Fig.1 に提案システムの構成を示す。提案システムは、携帯端末、ポータルサイトを提供する外部のWEBサーバとそれらを仲介するリンクデータ管理サーバからなる。リンクデータ管理サーバは、携帯端末からポータルサイトに送信されたリンクの利用情報（リンク情報）を記録して保持する。また、ユーザがリンクデータ管理サーバに登録されているサイトにアクセスした場合には、同サイト内のリンクの被クリック確率から表示するリンクの順序を決定し、ページ情報を更新した上で携帯端末に送信する。さらに、リンクデータ管理サーバでは、ユーザが登録されたサイト内のリンクをクリックするたびに同サイトのリンクの被クリック確率を更新する。

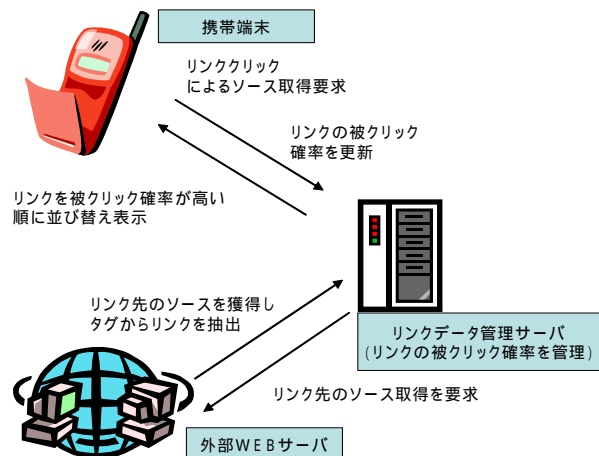


Fig. 1 提案システムの構成

3. リンクの並べ替え方式

3.1 各リンクの被クリック確率の計算法

サイト内で上から i 番目のリンクを L_i ($0 < i < N+1$)

*Effective Web Browsing on Mobile Phone Referring User History

†Yusuke Fukazawa, Takefumi Naganuma, Shizuka Kuromiya, Shoji Kurakake

‡Network Laboratories, NTTDoCoMo, Inc.

とし、リンク L_i の被クリック確率を $P(i)$ とする。 N はサイト内のリンクの総数である。 サイトへのアクセス総数を K 回とすると被クリック確率 $P(i)$ は(1)式のように更新される。

$$P(i) = \frac{(K-1) \times P(i) + \Delta P(i)}{K} \quad (1)$$

ここで、 $P(i)$ はリンク L_i の被クリック確率の更新量を表す。 上から j 番目のリンク L_j がクリックされたときの $P(i)$ を式(2)のように表す。

$$\Delta P(i) = \begin{cases} 0 & (i < j) \\ 1 & (i = j) \\ 1/2 & (i > j) \end{cases} \quad (2)$$

ここで、サイト内においてリンク L_j より上に位置するリンク L_i ($i > j$) はフォーカス遷移したがその後選択されていないことからユーザにとって必要のないリンクと考えられ、被クリック確率を 0 としている。 また、サイト内においてリンク L_j より下に位置するリンク L_i ($i < j$) はフォーカス遷移されていない可能性が高く被クリック確率を 1/2 とする。

3. 2リンクの並べ替え

サイト内のリンクを被クリック確率 $P(i)$ が高いリンクから順に並べ替えて表示する。 並べ替えを行ったサイトにユーザがアクセスし、リンクをクリックした際には、並べ替えられた順番でリンクの被クリック確率を更新する。

例を Table1 に示す。リンク総数 N が 6 のサイトにおいて L_4 がクリックされたとする。まずリンクの被クリック確率の更新量 $P(i)$ を計算する (Table1 (左) $P(i)$ 列)。次回アクセス時にはリンクの被クリック確率をもとに並び替え表示する (Table1 (中) Ordered Link 列)。ここで、 L_4 がクリックされた場合には並び替えられたリンクの順番で更新量 $P(i)$ を計算する (Table1 (中) $P(i)$ 列)。更新後の $P(i)$ の計算結果を Table1 (右) $P(i)$ 列に示す。

Table1 リンクの被クリック確率 $P(i)$ の計算例

Ordered Link	$P(i)$	$P(i)$	Ordered Link	$P(i)$	$P(i)$	Ordered Link	$P(i)$	$P(i)$
1	0	0	4	1	0	6	1/2	0
2	0	0	5	1/2	0	4	1/4	0
3	0	0	6	1/2	1	5	3/4	0
4	0	1	1	0	1/2	1	1/4	0
5	0	1/2	2	0	1/2	2	1/4	1
6	0	1/2	3	0	1/2	3	1/4	1/2

4. 提案システムの実装例

リンクごとの被クリック確率の管理を SQL データベース MySQL により、また、リンクの被クリック確率の計算、リンクの並べ替えなどの処理を JAVA サブレットによりデータ管理サーバとして実装した。ここでは、ポータルサイトのように複数の異なるドメインのリンク先を有するサイトを対象として提案システムの検証を行った。

Fig.2 (左) に対象とするサイトを示す。この画面は、ユーザが初めてこのサイトを閲覧した際に表示

される。

ユーザが Fig.2 (左) に示すサイトから“懸賞”で表されるリンクをクリックしたとする。同ユーザが再度このサイトを閲覧した際、被クリック確率が高いリンク順に Fig.2 (中) に示す画面が表示される。ここで Fig.2 (左) に示す画面で“懸賞”リンクより下に位置するリンクの被クリック確率は 1/2、上に位置するリンクの被クリック確率は 0 となっている。クリックされなかったリンクの被クリック確率をすべて 0 または 1/2 とした場合には Fig.2 (右) のような画面が表示される。

3. 1 節で述べたように、Fig.2 (左) における“ニュース”から“学習”までのリンクは、フォーカス遷移したがその後選択されていないことからユーザにとって必要のないリンクと考えられる。提案システムでは、“ニュース”から“学習”までのリンクの被クリック確率を低く設定することで再閲覧時に下位に表示されていることが分かる。さらに、“求人”から“クーポン”までのリンクは画面制約上見えていなかったリンクである可能性があり際閲覧時に上位に表示することが可能となっている。

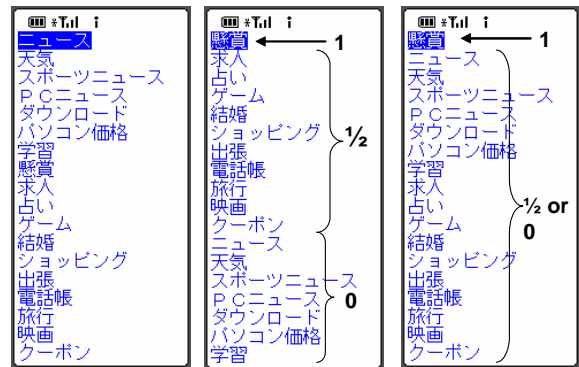


Fig.2 提案システムの実装例

5. おわりに

本稿では、ユーザのリンククリック履歴からリンクの被クリック確率を決定し、リンクを並べ替え表示するシステムを提案した。提案システムをアプリケーションサーバ上に実装し、ポータルサイトを提供する Web サーバを対象として有効性を検証した。今後は、リンクのクリック履歴と同時にクリックされたリンクのセマンティクスを考慮し推薦するリンクのさらなる高精度化を目指したい。

参考文献

- [1] <http://www.microsoft.com/windows/ie/default.mspx>
- [2] Ramesh R. Sarukkai, Link prediction and path analysis using Markov chains, computer networks 33,377-386, 2000.
- [3] Jiyang Chen, Lisheng Sun, Osmar R. Zaiane, Randy Goebel: Visualizing and Discovering Web Navigational Patterns. WebDB 2004.