

グラフデータベースを用いた Web 閲覧者の行動分析手法の提案

A Proposal of Behavior Analysis Method of Web Browsers Using Graph Database

椋本大輔[†], 森山真光[†]

Daisuke Mukumoto[†], and Masamitsu Moriyama[†]

[†]近畿大学大学院 総合理工学研究科

[†]Graduate School of Science and Engineering Research, Kindai Univ.

要旨

近年, Web サイトによる情報発信は重要な位置を占めている. Web 閲覧者の動向を調査する上で有効な手法にマウスカーソルの軌跡を用いた分析がある. 既存の手法では, 時間毎, 空間毎に分析する手法が多く存在するが, 膨大な軌跡から Web 閲覧者毎の特徴を掴むのは困難である. そこで本研究では, Web ページの遷移とページ内の要素の遷移を組み合わせたグラフデータベースによる Web 閲覧者の行動分析手法の提案をする.

1. はじめに

近年, インターネットやスマートフォンの普及により個人から企業まで幅広く Web サイトを運用, 管理をしている [1]. BtoB や BtoC において情報収集方法, 情報発信方法としても Web サイトの利用割合は高いものとなっている. そのため企業が Web サイトに取り組む事は重要な位置を占めており, アクセス数の増加, Web 閲覧者の媒体に合わせた Web サイトのデザイン変更などを行う事が必要になっている. その方法の一つとして Web サイトのアクセスログを用いたアクセスログ解析が挙げられる. 一般的なアクセスログには, アクセスされた時刻や URL などクライアントからサーバへアクセスした際のデータが蓄積されており, これらから Web 閲覧者のページ遷移や検索エンジンからの様な単語でその Web サイトに行き着いたのかを分析する研究が行われている. また Web 閲覧者が Web サイト内でどのページに興味を示したのか, どのコンテンツに興味を持ったのかといった Web 閲覧者の動向を調査する方法としてマウスカーソルの動きを捉えたマウスログから軌跡として可視化, 分析する手法も存在する [2] [3] [4] [5]. この手法によってページ上の座標といった空間的要素毎, 及びアクセス時間といった時間的要素毎にマウスログを可視化することは可能になるが, それでも膨大なマウスカーソルの軌跡から Web 閲覧者毎の特徴を掴むのは困難である. そこで本研究では, Web 閲覧者の行動分析手法として, Web ビーコン型のマウスログ取得スクリプトを用いることで得られたログから, Web 閲覧者がアクセスした Web ページから別の Web ページへの遷移と, Web ページ内の要素から別の要素への遷移を組み合わせたグラフデータベースを構築, 可視化を行う.

2. 研究手法

Web 閲覧者の行動分析手法として, Web ページから別の Web ページへの遷移について, また, Web ページ内の要素から別の要素への遷移について考える.

2.1. Web ページの遷移の有向グラフ

Web 閲覧者による Web ページの遷移の有向グラフについて考える. Web ページをノードとし, その集合を V_a とする. Web 閲覧者がアクセスした Web ページから別の Web ページへの遷移を表現したものをエッジとし, その集合を E_a とする. このとき, Web ページの遷移の有向グラフ G_a は次のようになる.

$$G_a = (V_a, E_a). \quad (1)$$

ノードの集合である V_a の要素として $v \in V_a, v' \in V_a$ が存在し, エッジの集合である E_a の要素として $e = (v, v') \in E_a$ が存在する場合は, ノード v から v' へのページの遷移が存在している.

2.2. Web ページ内の要素の遷移の有向グラフ

Web ページ内に存在している要素の遷移による有向グラフについて考える. Web ページ内の要素をノードとし, その集合を V_b とする. Web 閲覧者が Web ページ内の要素から別の要素への遷移を表現し

たものをエッジとし、その集合を E_b とする。このとき、Web ページ内の要素の遷移の有向グラフ G_b は次のようになる。

$$G_b = (V_b, E_b). \tag{2}$$

ノードの集合である V_b の要素として $v \in V_b, v' \in V_b$ が存在し、エッジの集合である E_b の要素として $e = (v, v') \in E_b$ が存在する場合は、ノード v から v' への要素の遷移が存在している。すなわち、本研究で用いるグラフデータベース D は (1), (2) より次のようになる。

$$D = \{G_a, G_b\}.$$

2.3. グラフデータベース Neo4j

グラフデータモデルとは、関係性を表すグラフのことであり、ソーシャルネットワークやパナマ文書の複雑な解析などに使用されるグラフ専用のデータベースのことである。今回、グラフデータベース Neo4j¹を用いて、Web ページの遷移、Web ページ内での要素の遷移を可視化する。Neo4j は Cypher 言語と呼ばれるクエリ言語を用いて操作する。これは SQL ライクなクエリ言語であるため、構築したグラフデータベースに対して、追加、変更、検索などのデータ操作が可能になる。

図1に構築するデータモデルを示す。Neo4j のデータモデルは、ノード (Node)、関係 (Relationship)、プロパティ (Propaty) で構成される。ノードには、Web ページの情報を持った Pages、Web 閲覧者の情報を持った Users、Web ページ内の要素の情報を持った Elements がある。Users には ID, cookieID, 使用している OS やブラウザのプロパティ、Pages には ID, URL のプロパティ、Elements には ID, 要素名のプロパティを持つ。関係には、Web 閲覧者が Web ページにアクセスしたことを示す ACCESS, Web ページの遷移を示す P_MOVE, Web ページ内の要素の遷移を示す E_MOVE がある。ACCESS には、ID, アクセスした時間プロパティ、P_MOVE には、ID, 遷移した時間, アクセスした Web 閲覧者の cookieID プロパティ、E_MOVE には、ID, 遷移した時間, その要素を持つ Web ページの URL, アクセスした Web 閲覧者の cookieID プロパティを持つ。

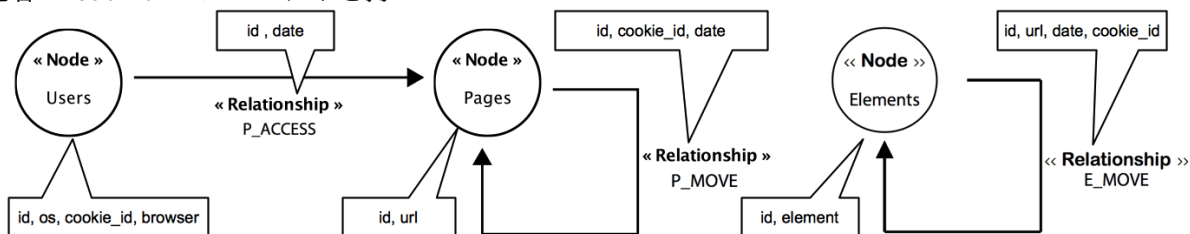


図1: データモデル図

2.4. マウスログの取得方法と可視化までの流れ

本研究ではマウスカーソルの動作の傾向を調べるためにマウスカーソルの動きを捉えたマウスログを用いる。マウスログの取得には Web ビーコンと Web サーバを用いた Web ビーコン型マウスログ取得方法を用いる。図2に使用する Web ビーコンによって Web サーバに送信されるマウスログのフォーマットを示す。フォーマットは Web 閲覧者の IP アドレスやアクセスされた時刻 (Unix time), リファラなどの情報に加え、マウスカーソルの X 座標, Y 座標や、マウスカーソル上の要素名などを JSON 形式にまとめている。

図3にマウスログ取得から可視化までの流れを示す。Web ビーコンを埋め込んだ Web ページに Web 閲覧者がアクセスする (3-1)。アクセスすると Web ビーコンが図2のフォーマットに沿ったデータをログ用サーバに送信する (3-2)。Web ページ毎に分割し (3-3), Neo4j に保存する (3-4)。Neo4j に Cypher 言語を用いて検索を行い、ブラウザで描写し可視化する (3-5)。

3. 実験結果・考察

3.1. マウスログの集計結果

本実験には予め可視化対象に Web ビーコンを配備し、一定期間内に取得したマウスログを用いて実験を行った。可視化対象には、我々が管理している研究室の Web サイト²と Wiki³を使用し、各 Web ペー

¹Neo4j Graph Platform The Leader in Graph Databases : <https://neo4j.com/>

```

<request time> mouse
{
  "time":<UNIX time>,"ri":<IP address>,"v":<script version>,"c":<site code>,"h":<host>,
  "l":<request URI>,"o":<OS>,"b":<browser>,"t[]":<page stay time>,"iw[]":<inner Width>,
  "ih[]":<inner Height>,"px[]":<X coordinate>,"py[]":<Y coordinate>,"eN[]":<element name>,
  "tL[]":<tag list>,"id":<CookieID>
}
    
```

図 2: マウスログのフォーマット

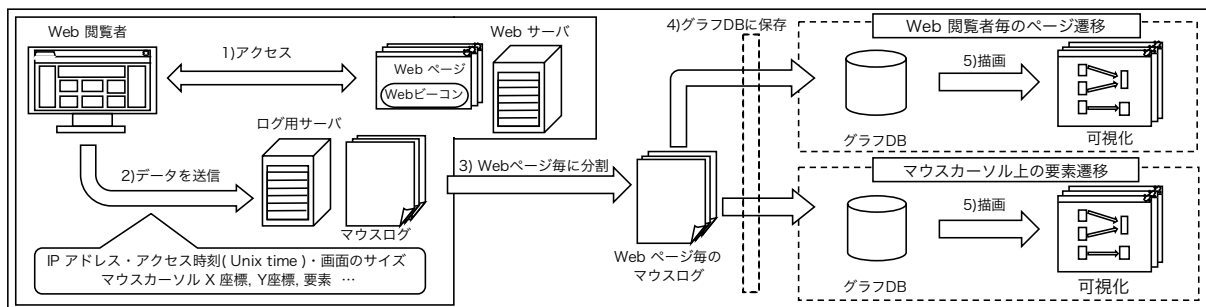


図 3: マウスログ取得から可視化までの流れ

に Web ビーコンを配備した。表 1 に 2018 年 09 月 01 日から 09 月 30 日の 1 ヶ月間のマウスログの集計結果を示す。

表 1: 2018 年 09 月のマウスログの集計結果

Pages (件)	160
Users (件)	196
Elements (件)	3,507
P_ACCESS (件)	2,340
P_MOVE (件)	2,340
E_MOVE (件)	43,528

3.2. 可視化

Web ページの遷移の可視化手法として、サンキー・ダイアグラム (Sankey-diagram) を用いた。サンキーダイアグラムとはフロー・ダイアグラムの一種であり、ノードとパスを用いて、工程間の流用を表現する場合や時間やプロセスの進行に伴う物事の流れを視覚化するための表現手法である。

図 4 にサンキーダイアグラムによる Web ページの遷移図を示す。これにより、Web 閲覧者の Web ページがどの Web ページに遷移したのかを視覚的に表現した。図 4(a)、図 4(b)、図 4(c) はそれぞれ別の Web 閲覧者の Web ページの遷移図である。図 4(a) や図 4(b) の場合では Web 閲覧者は 1 つの Web ページから複数の Web ページに多く遷移しているのではないかと読み取れる。しかし、図 4(c) のような遷移図は視認性が悪く、一部の Web ページの遷移が読み取りづらいなど改善する必要がある。

図 5 に Web ページ内の要素の遷移図を示す。Web ページ内の要素の遷移の可視化には、HTML5 の Canvas API による可視化手法を用いて視覚的に表現した。図 4(a)、図 4(b)、図 4(c) はそれぞれ別の Web 閲覧者であり、可視化対象として 1 つの Web ページ内の要素の遷移を可視化した図である。始点に S と表示 (青丸) し、終点には E と表示 (赤丸) している。図 4(a) の場合では、左上に遷移し、別の Web ページに遷移したのではないかと読み取れる。図 4(b) や図 4(c) の場合では、それぞれ別の記事に遷移したことが読み取れる。

²電子商取引研究室 : <https://ecl.info.kindai.ac.jp/>

³電子商取引研究室 Wiki : <https://ecl.info.kindai.ac.jp/>

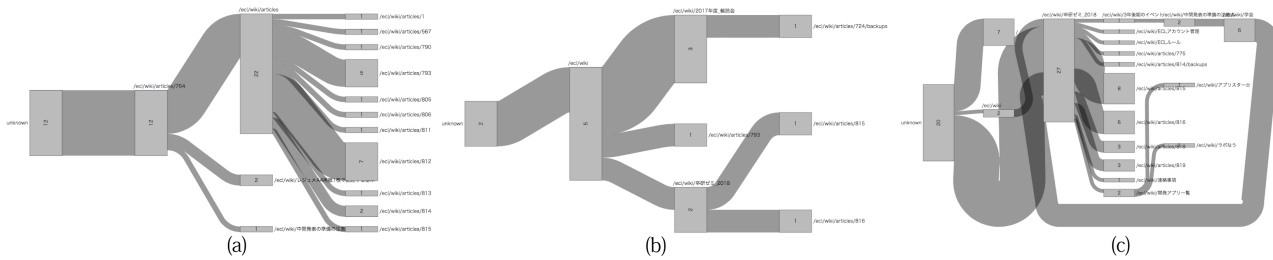


図 4: サンキーダイアグラムによる Web ページの遷移図

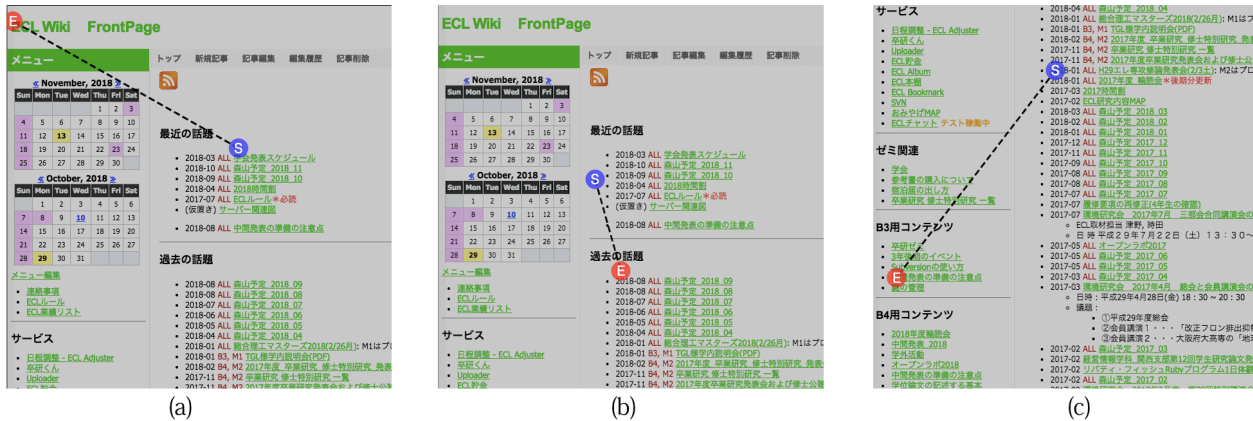


図 5: Web ページ内の要素の遷移図

4. おわりに

本研究では、Web 閲覧者の行動分析手法として、Web ビーコン型のマウスログ取得スクリプトを用いることで得られたマウスログから、Web 閲覧者がアクセスした Web ページから別の Web ページへの遷移と、Web ページ内の要素から別の要素への遷移を組み合わせたグラフデータベースを構築、可視化を行った。グラフデータベースには Neo4j を使用し、可視化手法には、サンキー・ダイアグラムや HTML5 の Canvas API による可視化手法を用いた。今後の課題としては、本実験では小規模な Web サイトを対象としたが、大規模な Web サイトにも適用させることが挙げられることや、どの要素から別の Web ページに遷移したのかを把握するために、Web ページを跨いでの要素の遷移を可視化を行うなど、可視化手法の利用者によって理解しやすいようにする、といった点が挙げられる。

参考文献

- [1] 経済産業省, ”平成 29 年度我が国におけるデータ駆動型社会に係る基盤整備 (電子商取引に関する市場調査)”, 2018.
- [2] Luis A. Leiva, Je Huang, “Building a better mousetrap: Compressing mouse cursor activity for web analytics”, Information Processing & Management, Vol. 51, Issue 2, pp. 114-129, 2015.
- [3] Shakti Kundu, L Garg, ”Web log analyzer tools: A comparative study to analyze user behavior”, Cloud Computing, Data Science & Engineering - Confluence, 2017 7th International Conference on, 12-13 Jan. 2017.
- [4] Clemens Schefels, Sven Eschenberg, Christian Schneberger, ”Behavioral Analysis of Registered Web Site Visitors with Help of Mouse Tracking”, Commerce and Enterprise Computing (CEC), 2012 IEEE 14th International Conference on, 9-11 Sept. 2012.
- [5] Nada Elgendy, Ahmed Elragal, ”Big Data Analytics in Support of the Decision Making Process”, Procedia Computer Science, Volume 100, Pages 1071-1084, 2016.