

Web サイト上のターゲット広告の観測システムの開発

青山 綾佳† 梶間 大地† 菊池 浩明†

明治大学総合数理学部†

1 はじめに

近年, Web 広告は広告市場の中でも著しく成長している. Web 広告の目覚ましい発展の一方で, Web 広告における個人情報の取り扱いに対する規制強化が検討されている. 中でも, ユーザの閲覧履歴や検索結果をもとに特定のユーザー層を標的としたターゲット広告は, 自動的に送信される仕組みに透明性がなく, プライバシーの侵害の懸念がある.

そこで, 本研究では, 閲覧履歴によって行われるターゲット広告の頻度や分野の調査を行うことを目的とする. しかし, ターゲット広告はブラウザの情報をもとに広告のリアルタイム入札が行われてから配信されるため, 通常のクローラーでは観測が困難である. 本研究では, Node.js ライブラリの一種である Puppeteer を用いて, 自動で Web サイトを遷移し, そのサイトから広告の URL を観測するシステムを構築した. 開発したシステムとその観測結果を報告する.

2 提案システム

ターゲット広告を効率良く観測するために, 自動的に広告の URL を取得するシステムを Puppeteer を用いて提案する [1]. システム構成図を図 1 に示し, タイムチャートを図 2 に示す.

1. 調査対象のペルソナにログインする.
2. 対象 Web サイトのサーバにアクセスする.
3. 調査したい Web サイトを開いた後, ブラウザで JavaScript を実行する.
4. 実行結果のソースから iframe タグの広告 URL を取得して, ブラウザからアクセスする.
5. 広告 URL を取得した日時, 調査した Web サイトの URL, 取得した広告 URL を Excel ファイルに書き込み出力する.

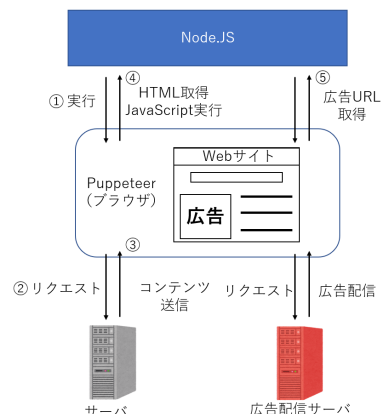


図 1 システム構成図

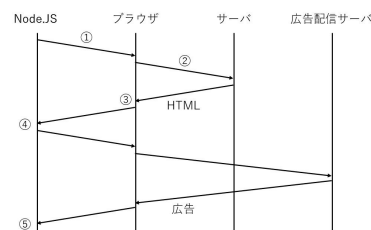


図 2 タイムチャート

表 1 ペルソナ (自動, 手動)

興味	検索ワード
1 育児	ベビー用品, ベビーグッズ
2 勉強	受験, 塾, 予備校
3 食事	宅配サービス, 生協
4 ファッション	コーデ, パンプス
5 長寿	健康サプリ, 老人ホーム
6 住宅	賃貸, 住まい
7 電子機器	iphone, mac
8 ダイエット	ダイエットサプリ, ダイエットグッズ
9 ゲーム	ゲームソフト, ゲームアプリ

3 実験

3.1 ペルソナ (興味の異なる仮定のユーザ)

ターゲット広告の差異を確認するために 9 種のペルソナを設定した. ペルソナとはマーケティングにおいて活用される概念でサービスや商品の典型的なユーザー像のことを指す.

本研究では作成した 9 個の Google アカウントで, 表 1 に示す 9 種の興味から想定される検索ワードに関連した Web サイトを複数閲覧した.

Development of system observing targeted advertisements on websites

†Ayaka Aoyama, Daichi Kajima, Hiroaki Kikuchi, School of Interdisciplinary Mathematical Science, Meiji University.

表2 システムで調査したターゲット広告率（ターゲット広告数/総広告数）と相対誤差

サイト概要 ペルソナ	Q & A	Web 制作	開発	ブログ	感想	漫画	攻略	お菓子	観光	家具	平均
育児	0/3	1/2	2/4	1/3	1/1	1/2	0/3	0/1	0/2	0/1	0.6/2.2
勉強	3/3	3/3	1/4	1/3	1/1	2/2	0/3	1/1	2/2	0/1	1.4/2.3
食事	2/3	3/3	2/3	3/3	1/1	2/2	2/3	1/1	0/1	3/3	1.9/2.3
ファッション	3/3	2/3	3/3	3/3	1/1	2/2	2/3	1/1	2/2	1/1	2.0/2.2
長寿	4/4	0/3	4/5	0/3	1/1	1/2	1/2	1/1	2/2	0/2	1.4/2.5
住宅	0/3	3/3	1/3	1/3	0/1	0/2	0/3	0/2	0/2	0/1	0.5/2.3
電子機器	4/4	1/3	1/3	2/2	1/1	0/2	0/3	0/2	3/3	0/1	1.2/2.4
ダイエット	0/3	0/2	3/5	0/3	0/1	0/2	0/3	0/2	0/2	0/1	0.3/2.4
ゲーム	0/3	0/3	0/4	0/3	0/1	0/2	0/3	0/2	0/2	0/3	0/2.6
平均 (自動)	1.8/3.2	1.4/2.8	1.9/3.8	1.2/2.9	0.7/1.0	0.9/2.0	0.6/2.9	0.4/1.4	1.0/2.0	0.4/1.6	
平均 (手動)	3.9/4.1	2.4/3.0	4.1/7.3	2.7/5.9	2.0/3.0	3.4/6.9	3.2/5.0	2.0/5.3	6.9/13.8	1.0/3.0	
相対誤差 (%)	22.0	6.7	48.0	50.8	66.7	71.0	42.0	73.6	85.5	46.7	

表3 調査した 10 の Web サイト

サイト概要	Web サイト URL
Q & A	http://ja.uwaku.com/question/p-nrfjwnhr-nn.html
Web 制作	http://www.htmq.com/csskihon/
開発	https://ichi.pro/node-js-de-no-puppeteer-niyoru-web-sukureipingu-81582238855555
ブログ	https://www.marorika.com/entry/bootstrap-beginner
感想	https://sukiaraba-game.jp/?p=41542
漫画	https://yuuublogkakutou.com/donfinish
攻略	https://xn--zck9awe6d820vk6qg9be46k.com/kouryakumenyu/111016/
お菓子	https://myrecommend.jp/gifts-of-sweets-39/
観光	https://yochi-orange.com/canada-eastside-trip/
家具	http://simplemodern-interior.jp/overseas-interior/scandinavian-style/

表4 1つの Web サイトの広告の観測時間

実験	時間	比
自動で行った実験	28 秒	1.0
手動で行った実験	1 分 4 秒	2.3

3.2 実験概要

本研究では、表1のペルソナについて表3の10の対象Webサイトを開発したシステムを使って自動で広告URLを取得し、ターゲット広告と全ての広告の数を観測する。評価のために、同じ実験を手動でも行った。

3.3 実験結果

自動の実験結果と平均、及び手動の実験結果の平均を表2に示す。自動と手動それぞれでWebサイトの広告の観測時間を測定し、さらに自動の実験の観測時間を1.0とした手動の観測時間の比を表4にまとめた。表4より、自動の実験は手動の実験よりも2.3倍の速さで実験を行うことが出来た。

一方、自動の実験は、手動の実験に比べて取得できる広告数が少なく、ターゲット広告数も少なかった。手動の実験結果を真値とした自動の実験結果の相対誤差を表2の下部に示す。 X を手動の実験で取得した総広告数、 Y を自動の実験で取得した総広告数として、相対誤差 ΔA を

$$\Delta A = \frac{|X-Y|}{X} \text{ とする.}$$

3.4 考察

「観光」のサイトで誤差が大きくなった原因として、Webサイト内の動画広告が多かったことが挙げられる。本システムでは動画広告を取得出来ないため、観測可能な広告数が少なく、誤差が大きくなったと思われる。さらに、広告URLを取得せずに処理を終えてしまう場合もあった。これは、実行中にシステムがタイムアウトしたためではないかと考えられる。自動でのページ遷移は回線やサーバの混雑状況によって動作に時間がかかることがあるため、対象Webサイトまでたどり着いても、全ての広告を取得せずに実行を終えたのではないかと考えた。

4 おわりに

本研究では、閲覧履歴によって行われるターゲット広告の調査を行うことを目的とし、広告のURLを取得するスクレイピングを行うシステムを開発した。

実験結果より、ペルソナごとのターゲット広告を自動で観測し、さらに手動よりも効率良く行うことが出来た。システムの精度向上を今後の課題とする。

参考文献

- [1] GitHub, “Puppeteer”, 閲覧日 2021/12/10, <https://github.com/puppeteer/puppeteer>
- [2] 美崎他, ”Puppeteer 入門 スクレイピング + Web 操作 自動処理プログラミング”, 秀和システム, 2018.
- [3] J. Estrada-Jimenez, et al. “Measuring Online Tracking and Privacy Risks on Ecuadorian Websites”, IEEE, pp. 1-6, 2019.