

ターゲット広告はどのペルソナで最も多いのか？

青山 綾佳 †

明治大学総合数理学部 先端メディアサイエンス学科 菊池研究室 †

1 はじめに

電通は 2021 年の日本の総広告費を通年で 6 兆 7,998 億円（前年比 110.4 %）と発表した。なかでもインターネット広告費は 2 兆 7,052 億円（前年比 121.4 %）と高い成長率を保っている。2021 年には、インターネット広告費が初めてマスコミ四媒体広告費（総計 2 兆 4,538 億円）を上回った [1]。

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。なぜならば、ユーザの閲覧履歴や検索結果をもとに、特定のユーザ層を標的としたターゲット広告には、自動的にプロファイリングされるアルゴリズムに透明性がなく、プライバシーの侵害の懸念があるからである。2018 年 EU で施行された一般データ保護規制（GDPR）にて、Cookie 等によるプロファイリングが規制されている [2]。日本でも 2020 年、個人情報保護法改正によって Cookie 情報等は個人関連情報として扱われることになった [3]。

そこで、本研究では Cookie 情報によって行われるターゲット広告の調査を目的とする。以下の問いに答えることを試みる。

1. ターゲット広告はどのペルソナで最も多いのか？
2. ターゲット広告はどの種類の Web サイトで最も多いのか？
3. ターゲット広告はどのくらいの期間表示されるのか？
4. ターゲット広告の自動観測は出来るか？

ここで、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像のことを指す。これらを調査するために、自動的に広告の URL を取得し、ペルソナごとのターゲット広告を判別するシステムを開発する。本研究では、Web アプリケーションをテストするためのポータブルフレームワークである Selenium [6] を用いて、自動で Web サイトを遷移し、そのサイトからターゲット広告の URL を取得、観測する

システムを構築する。開発したシステムとその観測結果を報告する。

2 先行研究

2.1 John Cook らの研究 [4]

2019 年 John Cook らはヘッダービidding (入札) を活用して広告主の入札行動を把握し、機械学習により広告主とトラッカー間のデータ共有関係を推論する方法を提案した [4]。ヘッダービidding は媒体者が複数の SSP (Supply Side Platform) やアドエクスチェンジに、同時に広告リクエストを提供する仕組みである。

John らは Alexa のカテゴリ別に 16 のペルソナを設定し、関連するブラウザの Cookie を保存した。そのペルソナを利用し、広告主にとってペルソナがどれだけ重要かを調査し、閲覧履歴無しの人ペルソナは有りよりも著しく広告の入札比率が低いことを示した。一方「健康」、「子供」、「スポーツ」のペルソナでは広告の平均落札率が高かった。これらのことから、ペルソナの Cookie 情報は広告主による入札に影響を与えることが分かった。すなわち、広告主によるユーザのプロファイリングは閲覧履歴に大きく依存していると言える。

2.2 Jose Estrada-Jime らの研究 [5]

2019 年 Jose Estrada-Jime らは広告プラットフォームにおける透明性を調査し、潜在的なプライバシー侵害を明らかにするため、エクアドルでのオンライントラッキングを測定した [5]。主要な Web サイトを自動でクロールし、リクエスト情報を取得した。それらの情報からサードパーティによって設定された Cookie を解析し、対応する文字列がユーザを追跡する過程で識別子として考慮される可能性があるかを判断した。

彼らの実験結果は収集された内の 66 % の Cookie が識別子とみなされ、分析した媒体者の 60 % 以上がユーザ関連情報を収集していることを示した。これはユーザ情報の漏洩を引き起こしている可能性があると言える。

†Kikuchi Laboratory, Department of Frontier Media Science, School of Interdisciplinary Mathematical Science, Meiji University.

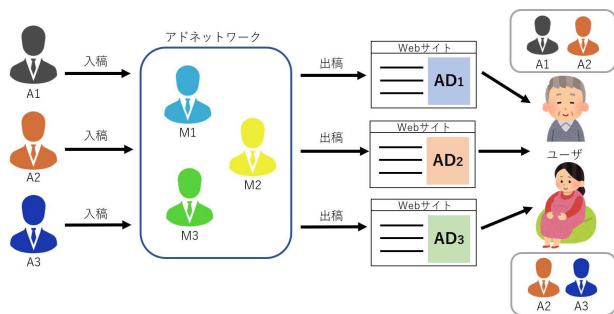


図1 アドネットワーク

3 アドネットワーク

広告配信の形態の1種であるアドネットワークは、媒体のWebサイトを多数集めて広告配信ネットワークを形成し、それらのWebサイト上にまとめてユーザに広告を配信する広告配信手法である。

アドネットワークの仕組みを図1を用いて説明する。まずアドネットワークに複数の媒体者が登録する。そして複数の広告主が広告を入稿し、アドネットワークを通して多数の媒体者が各自のWebサイトへ広告を出稿する。ユーザは自分の興味に関連する複数の広告主のWebサイトを閲覧し、その閲覧履歴をもとに媒体者のサイトにユーザのターゲット広告が届く。

アドネットワークを利用することで、様々なWebサイトに向けて広告出稿が出来るため、広告主はより多くのユーザに広告を届けることが出来る。また、媒体者にとってもアドネットワーク事業者に広告の受注や掲載の手続きを任せられるため、双方にとって業務効率化が可能になる。一方で、広告主は個別の媒体者が選べるわけではないため、広告の内容と関連のない媒体者のWebサイトで広告が配信されることもある。

4 提案システム

4.1 ペルソナ(興味の異なる架空のユーザ)

本研究では、ターゲット層ごとの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。例えば、「長寿」を興味とするペルソナは、老人ホームや健康グッズに興味を持つユーザと仮定する。

本研究ではアカウントごとに、興味の内容に関連したWebサイトから発行されるCookie情報等をブラウザに記録した。

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
10	シークレットアカウント	なし

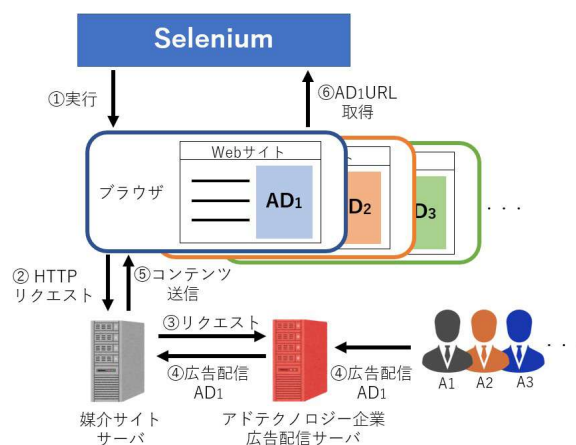


図2 システム構成図

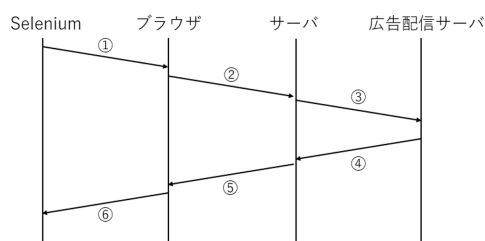


図3 タイムチャート

4.2 システム概要

本研究では、ターゲット広告の調査を目的とし、自動的に広告のURLを取得し、ペルソナごとのターゲット広告を判別するシステムを開発する。

本システムは、自動で複数の調査対象アカウント(4.1節にて前述)にログインし、調査対象のサイトを訪問しWebサイト内の広告URLを取得してファイルに出力する。システム構成図を図2に示し、タイムチャートを図3、出力例を表2、表3に示す。本システムは、Webサイトから自動で広告URLを取得するためにブラウザを操作するためのライブラリSelenium [6]を用いて実装した。

表2 システム出力結果例 (広告 URL)

実験日時	広告枠の番号	Web サイト URL
2022-10-24 14:26:58.295321	1	https://store.google.com/jp/product/pixel-7/
2022-10-24 14:26:58.295321	2	https://www.bellemaison.jp/shop/
2022-10-24 14:26:58.295321	3	https://cat.jp2.as.criteo.com/delivery/
2022-10-25 12:44:02.751080	1	https://www.uniqlo.com/jp
2022-10-25 12:44:02.751080	2	https://jp.shein.com/
2022-10-25 12:44:02.751080	3	https://pets-kojima.com/

表3 システム出力結果 (ターゲット広告数)

ターゲット広告数	35
全広告数	78
ターゲット広告率	0.448

本研究の提案システムの流れを示す。なお、本システムは以下の処理を自動的に行う。

1. 調査対象の各アカウントにログインした Chrome ブラウザを開く。
2. 対象 Web サイトにアクセスする。
3. Web サイトのソースより、iframe タグの広告 URL を取得する。
4. 実験日時、広告枠の番号、取得した広告 URL、全広告数、ターゲット広告数、ターゲット広告率を記録する。

ターゲット広告数を X 、対象ページ内の全広告数を Y とすると、ターゲット広告率 R を $R = \frac{X}{Y}$ とする。Selenium を用いて、JavaScript 実行後の Web サイト内の広告 URL を取得する。Selenium で Chrome ブラウザを開く際にプロファイルを指定することで、複数の対象アカウント (4.1 節にて前述) を切り替えて観測を行う。これによって、それぞれのアカウントで事前に関連した履歴に応じたターゲット広告を正しく判別する。

本研究でのターゲット広告の定義はペルソナの興味の内容に関連した媒体者から配信される広告とする。

4.3 実装の技術点

[7] において、2021 年に puppeteer を使って広告 URL を取得するシステムを実装した。そのシステムの問題点を示す。

1. 正しい結果が取得出来ない
実装したシステムは Web サイトにアクセスし iframe タグを取得した後に、iframe タグの広告 URL にアクセスすることで広告 URL を取得し

た。これにより 1 度目にアクセスしたサイトの広告と再アクセス時の広告とで差異が生じた。

2. ターゲット広告の判別をしない
昨年度は広告 URL を取得するのみで、ペルソナごとのターゲット広告の判別まで実装出来ず、手動でターゲット広告を記録していた。

これらの問題点を改善するべく、新たなシステムを実装する上で特に重要となったメソッドを下記に示す。

1. add_argument メソッド
Chrome のユーザー毎のデータ (履歴、ブックマーク、パスワード等) はユーザープロファイルに保存される。このメソッドを用いてユーザーアカウントごとのユーザープロファイルを Chrome 起動時に指定する。
2. find_element_by_id メソッド
Web サイトの HTML から広告 URL の記載されたタグの内容を取得する。
3. switch_to_frame メソッド
iframe タグにより配信される Web サイトの広告 URL をこのメソッドを用いてフレームを切り替えて取得する。

5 実験

5.1 実験目的

本研究では、ペルソナによるターゲット広告の影響を明らかにすることを目的とする。ターゲット広告を取得し、さらにシステムの精度を評価するために、手動と提案システムの両方を用いてターゲット広告を調査する。

5.2 実験方法

ペルソナごとの Cookie 情報や履歴の有無によって、表 4 の 10 の Web サイトで提示されるターゲット広告の差異を調べた。本実験ではまず 9 個の Google アカウントを作成した。そして表 1 に示すシークレットアカウント

表4 調査した10のWebサイト

サイト概要	サイト URL	サイト略称
Q & A	http://ja.uwenku.com/question/p-nrfjwnhr-nn.html	ExcelJS と Node.js で xlsx ファイルを修正
Web 制作	http://www.htmq.com/csskihon/	CSS の基本
開発	https://ichi.pro/node-js-de-no-puppeteer-niyoru-web-sukureipingu-81582238855555	Node.js での Puppeteer による Web スクレイピング
感想	https://sukiaraba-game.jp/?p=41542	2種のゲーム比較
漫画	https://yuublogkakutou.com/donfinish	漫画結末ネタバレ
ゲーム攻略	https://www.oi-mori.com/	どうぶつの森攻略
お菓子	https://myrecommend.jp/gifts-of-sweets-129/	お歳暮のお菓子まとめ
観光	https://yochi-orange.com/canada-eastside-trip/	カナダの観光モデルコース
家具	http://simplemodern-interior.jp/overseas-interior/scandinavian-style/	北欧インテリア 42 選
ブログ	https://www.marorika.com/entry/bootstrap-beginner	Bootstrap 入門

トを除く9種の興味から想定される検索ワードをアカウントごとに設定し、それぞれのアカウントで検索ワードに関連するWebサイトを複数閲覧した。なお、シークレットアカウントでは興味及び検索ワードを設定していない。

5.3 実験1 (システムを用いて長期的な日別ターゲット広告を観測)

表1の9のペルソナにおいて、表4の10のWebサイトのターゲット広告と全ての広告の数を観測する。実験は2022年7月28日から8月31日にかけて行った。

実験結果を表5に示す。ペルソナに注目すると、「食事」はターゲット広告率の平均が53%となり最も多い。「住宅」はターゲット広告率の平均が27%となり最も少ない。

Webサイトに注目すると、「Q & A」のサイトではサイト内の全広告数の60%がターゲット広告であり最も多い。一方で「家具」のサイトは16%で最も少ない。

ターゲット広告は、閲覧した日よりある日数が経過すると観測されなくなる。このターゲット広告が最後に観測された日付を表6に示す。×は実験で一度もターゲット広告が観測されなかったことを示す。ペルソナに注目すると、「長寿」は平均して8月28日までの32日間ターゲット広告が観測され最も長い。一方で「住宅」は平均22日間で最も短い。

Webサイトに注目すると、「お菓子」のサイトでは平均して8月27日まで31日間ターゲット広告が観測され最も長い。一方で「家具」、「ブログ」のサイトは平均22日間で最も短い。

表5は手動と自動でターゲット広告率に差異が生じている。手動の方がターゲット広告率が高いものの例を図4(開発)に示す。赤い枠で囲まれた広告がターゲット広告である。



図4 サイト「開発」(左:手動 右:自動)

ターゲット広告率に差異が生じた理由はWeb広告の性質に関連していると考えられる。Web広告はサイトを読み込むごとにアドサーバに再度リクエストされ、広告内容が一定しない。この性質により、結果に差異が生じたと考察する。

5.4 実験2 (5分ごとにターゲット広告の推移を観測)

表1の「育児」のペルソナについて、表4のWebサイトのターゲット広告と全ての広告の数を5分ごとに観測する。実験は2022年11月26日、12月3日、12月4日、12月14日から24時間計4日間行った。

4日間の実験結果の平均を図5に示す。図のx軸は実験開始からの経過時間、y軸は観測された1時間当たりのターゲット広告数を示す。どのWebサイトでも、実験開始から5時間以内で顕著にターゲット広告を観測出来た。特に「お菓子」、「攻略」、「Web制作」のサイトでは1時間当たり最大で20個以上のターゲット広告を観測した。一方、10時間を過ぎるとターゲット広告が全く観測されないサイトが現れた。これより、ターゲット広告は時間が経つと配信率が減ることが分かった。

表5 実験1 ターゲット広告率 (ターゲット広告数/総広告数)

サイト概要 ペルソナ	Q & A	Web 制作	開発	感想	漫画	攻略	お菓子	観光	家具	ブログ	平均 (自動)[%]
育児	2/3	1/3	1/4	1/3	1/4	1/2	1/6	0/7	1/5	0/5	27
勉強	2/3	1/3	3/4	1/3	1/4	1/2	1/6	2/7	0/5	2/5	37
食事	2/3	2/3	3/4	2/3	0/4	1/2	4/6	4/7	2/5	2/5	53
ファッション	2/3	2/3	2/4	1/3	3/4	1/2	2/6	3/7	1/5	1/5	46
長寿	2/3	1/3	3/4	2/3	3/4	1/2	3/6	5/7	0/5	1/5	51
住宅	2/3	1/3	1/4	0/3	0/4	1/2	3/6	0/7	1/5	0/5	25
電子機器	2/3	1/3	3/4	2/3	3/4	1/2	3/6	5/7	0/5	1/5	51
ダイエット	1/3	1/3	3/4	2/3	2/4	1/2	2/6	3/7	1/5	0/5	40
ゲーム	1/3	1/3	2/4	1/3	2/4	1/2	2/6	3/7	1/5	1/5	37
平均 (自動) [%]	60	40	58	43	43	50	38	40	16	18	
平均 (手動) [%]	70	40	71	30	43	50	27	29	9	18	

表6 ターゲット広告が最後に観測された日付

サイト概要 ペルソナ	Q & A	Web 制作	開発	感想	漫画	攻略	お菓子	観光	家具	ブログ	平均	日数 (日)
育児	8/25	8/17	8/16	8/25	8/26	8/22	8/27	8/15	8/7	×	8/20	24
勉強	8/31	8/31	8/25	8/26	8/10	8/31	8/23	8/29	8/27	8/27	8/26	30
食事	8/31	8/9	8/20	8/24	×	8/26	8/26	8/29	8/25	8/17	8/23	27
ファッション	8/21	8/18	8/20	8/24	8/26	8/24	8/24	8/17	8/19	8/9	8/20	24
長寿	8/26	8/31	8/31	8/30	8/26	8/31	8/31	8/29	8/22	8/21	8/28	32
住宅	8/26	8/14	8/14	8/25	8/3	8/26	8/26	8/23	8/9	8/9	8/18	22
電子機器	8/28	8/26	8/21	8/16	8/11	8/30	8/30	8/30	8/8	8/24	8/22	26
ダイエット	8/26	8/26	8/17	8/26	8/26	8/21	8/26	8/26	8/27	8/7	8/23	27
ゲーム	8/15	8/26	8/22	8/26	8/26	8/26	8/26	8/26	8/26	8/16	8/24	28
平均	8/25	8/22	8/21	8/25	8/19	8/26	8/27	8/25	8/18	8/18		
日数 (日)	29	26	25	29	23	30	31	29	22	22		

表7 自動観測システムの絶対誤差

サイト概要	Q & A	Web 制作	開発	感想	漫画	攻略	お菓子	観光	家具	ブログ	平均
絶対誤差度数 (件)	0.3/3.0	0.0/3.0	0.5/4.0	0.4/3.0	0.0/4.0	0.0/2.0	0.7/6.0	0.8/7.0	0.3/5.0	0.0/5.0	0.3/4.2
相対誤差率 (%)	10	0	13	13	0	0	11	11	7	0	7.1

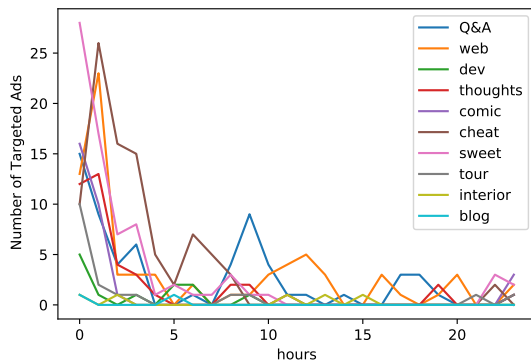


図5 実験2 時間ごとのターゲット広告数

率, \hat{R} をシステムでのターゲット広告率として, 絶対誤差 P を $P = |R - \hat{R}|$ とする.

表7より, 「Web制作」, 「漫画」, 「攻略」, 「ブログ」のサイトは絶対誤差が0%である. 一方で「開発」, 「感想」のサイトは絶対誤差が13%と最も大きい.

システムを使った場合と手動それぞれでWebサイトの広告の観測時間を表8に示す. システムを使うと手動の約2.3倍の速さで実験を行うことができる.

[7]において, 2021年にpuppeteerを使って実装したシステムで行った同様の実験と本実験の結果を表9, 表10と比較する. 表9より, ターゲット広告率が高かった「食事」と「長寿」は共通していた. 一方, 「育児」及び「住宅」は[7]に比べて約40%減少した. 表10よりWebサイトについてのターゲット広告率は変化がなかった.

5.5 観測システムの評価

手動の実験結果を真値とした時の自動の実験結果の絶対誤差を表7に示す. R を手動での平均ターゲット広告

表 8 1つの Web サイトの広告の観測時間

回数	システム [s]	手動 [s]
平均	16.6	37.7
標準偏差	3.3	3.2

表 9 ペルソナごとのターゲット広告率の比較 (高い順)

順位	2022 年	2021 年 [7]
1 位	「食事」 53 %	「長寿」 79 %
2 位	「長寿」 51 %	「食事」 72 %
3 位	「電子機器」 51 %	「ファッション」 71 %
4 位	「ファッション」 46 %	「育児」 63 %
5 位	「ダイエット」 40 %	「住宅」 62 %
6 位	「ゲーム」 37 %	「ダイエット」 53 %
7 位	「勉強」 37 %	「勉強」 46 %
8 位	「育児」 27 %	「電子機器」 40 %
9 位	「住宅」 25 %	「ゲーム」 10 %

表 10 ターゲット広告率の比較 (Web サイト)

	最も高い Web サイト	最も低い Web サイト
2022 年	「Q & A」 60 %	「家具」 16 %
2021 年 [7]	「Q & A」 56 %	「家具」 25 %

[7] のターゲット広告率の標準偏差は 19.9 であり、本実験よりもターゲット広告率の変動が大きい。

5.6 考察

[7] と大きく異なったターゲット広告率は「ゲーム」のペルソナであった。この理由として、「ゲーム」のアカウントで 1 年でなぜ大きくターゲット広告率が増加したかを考察する。

本実験では「ゲーム」ペルソナで原神というゲームアプリが最も多くターゲット広告として観測された。原神は [7] の実験を行った 2021 年に比べて 2022 年第一四半期の売上が前年比 44 % 増と好調で伸びている [8]。これより、本実験でターゲット広告率が増加したと考える。この様に、ターゲット広告率はその時の広告主の状況に大きく依存する。

6 おわりに

本研究では Cookie 情報によって行われるターゲット広告を調査した。自動的に広告 URL を判別するシステムを開発した。

実験の結果より以下のことがいえる。

1. ターゲット広告は「食事」のペルソナで最も多い。

2. ターゲット広告は「Q & A」の Web サイトで最も多い。
3. ターゲット広告は少なくとも 22 日間表示される。
4. ターゲット広告は自動観測出来る。

全てのサイトで広告 URL が取得出来るように改良することを今後の課題とする。

参考文献

- [1] 電通, “2021 年 日本の広告費 インターネット広告媒体費 詳細分析”, (閲覧日 2022/12/10, <https://www.dentsu.co.jp/news/release/2022/0309-010503.html>).
- [2] General Data Protection Regulation, 2016 (個人情報保護委員会による和訳あり), “GDPR (General Data Protection Regulation : 一般データ保護規則)”
- [3] 個人情報保護委員会, “令和 2 年 改正個人情報保護法について”, (閲覧日 2022/12/10, <https://www.ppc.go.jp/personalinfo/legal/kaiseihogohou/>).
- [4] John Cook et. al, ”Inferring Tracker-Advertiser Relationships in the Online Advertising Ecosystem using Header Bidding”, 2019, <https://www.petsymposium.org/2020/files/papers/issue1/popets-2020-0005.pdf>
- [5] Jose Estrada-Jim et. al, “Measuring Online Tracking and Privacy Risks on Ecuadorian Websites”, 2019 IEEE Fourth Ecuador Technical Chapters Meeting (ETCM), pp. 1-6, 2019.
- [6] Selenium, “The Selenium Browser Automation Project”, (閲覧日 2021/12/10, <https://www.selenium.dev/documentation/>).
- [7] 青山 綾佳, 梶間 大地, 菊池 浩明, ”Web サイト上のターゲット広告の観測システムの開発”, 情報処理学会第 84 回全国大会, pp.551-552, 2022.
- [8] gamebiz, “『原神』スマホ版の総売上が 30 億ドル (3916 億円) を突破 22 年 1Q の MAU は前年同期比で 44 % 増と好調…Sensor Tower 調査”, (閲覧日 2022/12/13, <https://gamebiz.jp/news/348632>).