

明治大学大学院 先端数理科学研究科

2024年度

修士学位請求論文

ユーザ閲覧行動履歴によるターゲティング広告の入
札モデルに関する研究

学位請求者 先端メディアサイエンス専攻

田口 勇翔

目次

第1章 序論	1
1.1 本研究の背景	1
1.2 本研究の目的	2
1.3 本研究の構成	2
第2章 基本定義と従来研究	3
2.1 基本定義	3
2.1.1 ターゲティング広告	3
2.1.2 ステークホルダー	3
2.1.3 Real Time Bidding(RTB)	4
2.1.4 Header Bidding(HB)	4
2.1.5 HB 情報	4
2.1.6 OpenWPM	5
2.1.7 ペルソナ	7
2.1.8 HBCollector	7
2.2 従来研究	8
2.2.1 閲覧履歴が広告に与える影響および広告主とトラッカーのデータ共有に関する研究	8
2.2.2 実ユーザのブラウザ環境を用いた HB の入札情報の収集と分析	8
2.2.3 RTB におけるターゲティングの持続期間に関する調査	8
2.2.4 本研究の位置付け	8
第3章 Header Bidding 情報を用いたペルソナに応じたターゲティング広告の調査	9
3.1 目的	9
3.2 提案方式	9
3.3 実験概要	9
3.3.1 実験1：手動によるペルソナの影響の観測	10
3.3.2 実験2：広告クリックによる広告価格への影響調査	10
3.3.3 実験3：自動によるペルソナの影響の観測実験	11
3.3.4 実験結果	11
3.3.5 考察	12
3.4 まとめ	15

第 4 章	インプレッション数による入札額の変化	16
4.1	目的	16
4.2	アクティブなインターネットユーザ数	16
4.3	インプレッション数と入札額の仮説を検証するための調査	17
4.3.1	実験概要	17
4.3.2	実験条件	17
4.3.3	実験結果	17
4.3.4	考察	20
4.4	まとめ	20
第 5 章	オンラインターゲティング広告におけるユーザ閲覧履歴入札の数理モデル	21
5.1	目的	21
5.2	DS.INSIGHT を用いた実践的ペルソナの設計	21
5.3	リターゲティングによる数理モデルの提案	21
5.3.1	実験結果	23
5.3.2	線形重回帰による数理モデルの提案	26
5.4	時間経過による入札額, 入札数の変化	28
5.4.1	実験方法	28
5.4.2	実験結果	29
5.5	リターゲティング広告の影響検証実験	31
5.5.1	実験概要	31
5.5.2	実験方法	31
5.5.3	実験結果	31
5.6	まとめ	32
第 6 章	結論	34
	参考文献	34
	謝辞	37
付 録 A	実験結果 4.3.3 各広告枠に対する入札の入札額の分布	39

第1章 序論

1.1 本研究の背景

近年オンライン広告市場は大きく成長している。例えば、電通グループは、デジタル広告の高い市場成長率予測と、総広告費に置けるデジタル広告の割合が2026年に60%に達する見込みを報告している [1]。とりわけ、広告の効果を高めるために、ユーザの趣味・趣向をユーザのWebサイトの閲覧履歴などから分析し、広告効果の期待できるユーザ層に配信するターゲティング広告が主流である。

ターゲティング広告はユーザと広告業者の双方に利点があり、広く認められている。一方で、Webサイトにおいて広告業者やトラッキング業者により、ターゲティング広告を決めるために用いられる情報が、ユーザが気が付かない間に収集されることに対して、プライバシーに対する懸念が生じている。例えば、Blaseら [12] による米国の48人に対するインタビュー調査によると、85%のユーザはオンライン上のユーザの行動のトラッキングに対して、見えないところで何が行われているか分からない事に対する恐怖や不快感を覚えると回答している。

広く普及している Real Time Bidding (RTB) では、Data Management Platform (DMP) と呼ばれるシステムで、Web上のユーザの行動データをトラッキングし集約管理することで、ユーザに見えないところでターゲットのセグメントが決められていた [23]。こうした仕組みは、ユーザに透明性を欠いた形で広告が表示される一因となっている。

ユーザトラッキングに用いられる 3rd Party Cookie はプライバシー上の懸念がたびたび指摘されてきた [3][4]。このため、Firefox や Safari といったブラウザでは、3rd Party Cookie がデフォルトで無効化されている。しかし、最もシェア率の高い [2] Chrome ブラウザを提供し、かつ広告業者である Google LLC は、3rd Party Cookie の廃止方針を撤回し [6]、さらに、Fingerprinting [5] によるトラッキングを許可する方針を示している [7]。3rd Party Cookie の廃止計画の撤回について、プライバシー保護の取り組みを遅らせるものとして批判されている [8]。また、Fingerprinting によるトラッキングはプライバシーに敏感なユーザでさえブロックすることが難しいとしてさらなる批判を招いている [9]。

ユーザトラッキング技術の悪用例として、「リクナビ事件」が挙げられる。Web上でのユーザのトラッキングが2019年に株式会社リクルートサービス(現・株式会社リクルート)が運営する新卒向けの就職活動サービスのリクナビで提供されていたサービスであるリクナビ DMP フォローは、就活生のリクナビ上での閲覧履歴と、企業のアンケートを Cookie により紐づけ、就活生の内定辞退率の予測データを企業に販売していた [10]。この問題により、リクルートキャリア社およびリクルート社、サービスの利用企業38社が行政指導の対象となった [11]。この事例は、ユーザトラッキング技術の悪用リスクを広く社会に認識させ、個人関連情報(個人情報ではないが、受信者が個人情報と紐づけられる情報)が導入されるきっかけとなった。

1.2 本研究の目的

本研究では、オンライン広告の入札がどのように決定されるか、広告配信のロジックを明らかにすることを目的とし、オンライン広告業界の不透明さを改善することを試みる。

特に、どのようなユーザの属性や行動がトラッキングされ広告配信に影響を与えるのかを定量的に分析する。また、パブリッシャーや、時間帯によるインプレッション数の違いによる影響といった、環境的な要因が広告配信に影響を与える影響を検討する。この研究により、オンライン広告業界の不透明さを改善し、ユーザのプライバシーを向上させることを目指す。

具体的には、

1. ユーザの閲覧履歴による入札額, 入札数の変化 (第3章)
2. サイト (パブリッシャー) による入札額, 入札数の変化 (第3章)
3. インプレッション数と入札額の関係 (第4章)
4. 訪問履歴 (リターゲティング) による入札数, 入札価格の変化 (第5章)

を明らかにすることを試みる。

そのために、Header Bidding(HB)[13]に着目する。HBにおいては、RTBと異なり、閲覧者に対する全ての入札値がブラウザに送信され、ブラウザ上で最高額の広告が落札される。従って、ブラウザ側で全ての入札を観測する事ができる。

本研究では、このHBの仕組みを活用し、入札情報を分析することで、ユーザの属性や行動、時間帯などの環境的な要因が広告価値にどのように影響を与えるかを解明する。

1.3 本研究の構成

本論文は6章により構成されている。

第1章では、まず本研究の背景と目的を述べる。ターゲティング広告に用いられるユーザトラッキング技術の、プライバシー上の問題を、実例を交えながら示し、HBの特性及びその解析の利点を述べる。

第2章では、広告業界で用いられる用語の定義や、オンライン広告のオークションの仕組みを示す。次に、関連研究の概要を述べ、本研究の新規性を明確にする。

第3章では、ユーザの閲覧履歴や行動、パブリッシャーによって広告単価が変化することを示す。また、時間帯により広告単価が変化することを示す。

第4章では、第3章で得られた時間帯ごとの広告単価の変化を基に、時間帯によるアクティブなインターネットユーザ数の変化によって発生するインプレッション数の増減によって、広告価格が変化すると仮説を立て、検証を行う。

第5章では 実際のユーザの検索行動をシミュレートし、広告の入札を観測することで、性別、収入、サイト訪問履歴の有無を説明変数として、ウェブサイトを訪れるユーザについての広告価格を定量化することを試みる。また、特定のユーザの属性が広告単価に大きな影響を及ぼすことを示す。

第6章では、本論文のまとめを述べ、本研究について結論付ける。

第2章 基本定義と従来研究

2.1 基本定義

2.1.1 ターゲティング広告

ターゲティング広告とは、Cookie 等でユーザの Web 上の行動履歴をトラッキングし、ユーザの属性を推測することで、最適化された広告を配信する手法であり、オークションを通じて広告枠を自動的に取引する運用型広告の一種である [23]。ユーザがパブリッシャーを訪問すると、ユーザに配信する広告を決めるためのオークションが、アドエクスチェンジ上で SSP, DSP により行われる。

2.1.2 ステークホルダー

ターゲティング広告には、次のステークホルダーがいる [23]。

広告主

広告主は、広告を掲載したい企業や個人、あるいはその代理店のことである。広告主は DSP やアドネットワークに依頼することでターゲット広告を配信する。

パブリッシャー

Web サイト等に広告を表示することで収益を得ることを目的に、広告枠を提供する。(例：jmtty.jp, eiga.com, cnn.com)

インプレッション・CPM

インプレッションは、広告が表示された回数である。Cost Per Mille(CPM) は、1000 インプレッションあたりの広告単価である。

例えば、CPM が 100 円であった場合、1 インプレッションあたりの広告価格は 0.1 円となる。

SSP・DSP

Supply Side Platform(SSP) は、パブリッシャーが広告収入を最大化させるためのプラットフォームである。

(例：Criteo, AppNexus, Ad Generation)

Demand Side Platform(DSP) は, 広告主の利益を最大化させるためのプラットフォームである. 広告主と契約を結び, 要望する広告配信先情報を収集する.

(例: Amazon DSP, MicroAd, MediaMath)

2.1.3 Real Time Bidding(RTB)

RTB は, パブリッシャーが単価の高い広告を得るための広告オークションの入札方式の1つである. インプレッションが発生するたびに広告枠に対するセカンドプライスオークション, すなわち, 最も高い価格での入札者が二番目に高い金額で落札し, パブリッシャーによって決められた最低価格を上回っていた場合, 広告を配信する.

2.1.4 Header Bidding(HB)

HB は, 新しい広告オークションの一つであり, 従来の RTB と違い, ファーストプライスオークションで落札する. 複数の SSP に対し同時にリクエストを送るため, パブリッシャーの収益の増加を見込むことができる. RTB は落札者とその価格のみが公開されるのに対し, HB は落札者を含めた入札者全員の入札額や広告主のドメイン等がブラウザに送信される.

HB の処理フローを図 2.1 に示す.

1. ユーザがパブリッシャーのサイトにアクセスする.
2. パブリッシャーがユーザに, コンテンツと HB wrapper(JavaScript による function) をレスポンスする
3. ユーザは HB wrapper を実行する
4. HB wrapper は複数の Demand partners(SSP, DSP など) に同時に入札リクエストを送る
5. SSP は DSP に入札リクエストを送る
6. DSP は SSP に入札する
7. Demand partners は HB wrapper に全ての入札を送る
8. HB wrapper はアドサーバにすべての入札を送る
9. 最終的に表示する広告をユーザに送る

2.1.5 HB 情報

本研究では, 最も利用されている, OSS の HB ソリューションである Prebid.js[16] を対象にする. Prebid.js では, JavaScript のメソッド `pbjs.getBidResponses()` を実行することで, 入札に関するオ

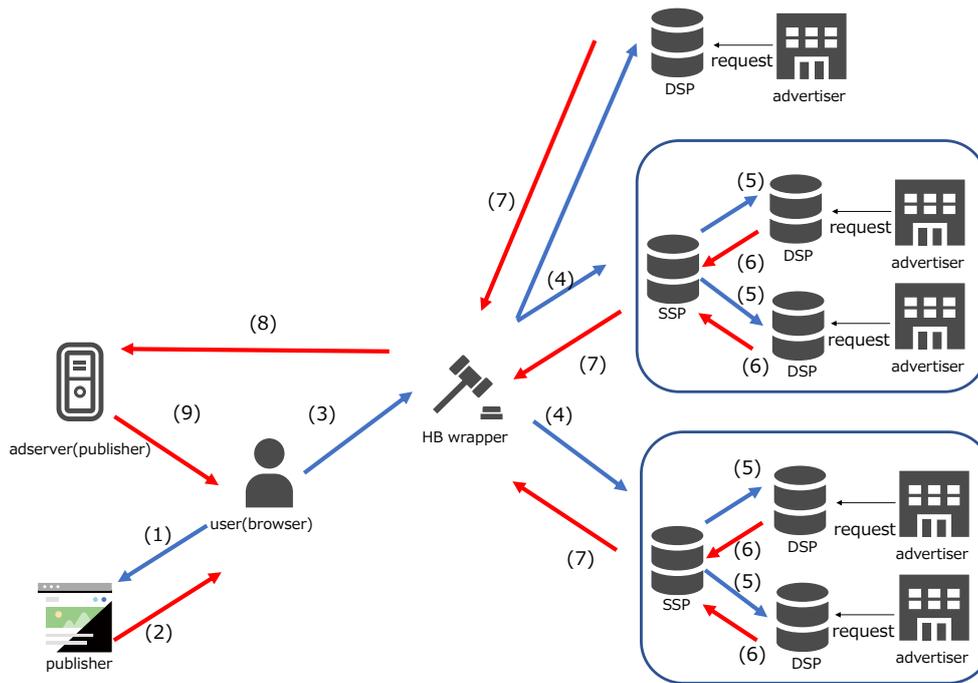


図 2.1: HB のフロー

プロジェクトを得る. 本研究では, このオブジェクトから入札額 (cpm), 入札者 (bidderCode) の値を参照する. Prebid.js のメソッドと, 実行することで得られるオブジェクトの一部の例を図 2.2 に示す.

また, `pbjs.getAllWinningBids()` を実行することで, HB の落札者かつ広告が正しく表示された入札のオブジェクトを得る. `pbjs.AllPrebidWinningBids()` を実行することで HB の落札者であるが最終的に広告が表示されなかった入札のオブジェクトを得る. この場合, ネイティブアドや RTB など, 他の方式で決まった広告が優先して表示された可能性がある.

表 2.1: Prebid.js の各メソッドで得られるオブジェクト

メソッド	得られるオブジェクト
<code>pbjs.getBidResponses()</code>	全ての入札情報
<code>pbjs.getAllWinningBids()</code>	HB で勝利しページに表示された広告の入札情報
<code>pbjs.getAllPrebidWinningBids()</code>	HB で勝利したがページに表示されなかった広告の入札情報

2.1.6 OpenWPM

OpenWPM[17] は, Python ライブラリの Selenium をベースに Firefox の上で動作するウェブサイト測定自動化フレームワークである. 自動で複数のブラウザの制御をすることができ, ウェブ上の観測に有用なツールであるが, Benjamin ら [18] によると, バニラの OpenWPM では `navigator.webdriver` プロパティや画面解像度などのフィンガープリント等を通じて Bot 検知可能であること, および広告業界で多く Bot 検知ツールが使われていること, Bot 検知を対策することでオンライン上のトラッキング目的の cookie の量が 50%程度増加する事が知られている. そのため, 本研究では, Benjamin らによる Bot 検知対策を施した `OpenWPMhide`[21] を用いた.

```

{
  "bids": [
    {
      "bidderCode": "pubmatic",
      "width": 300,
      "height": 250,
      "statusMessage": "Bid available",
      ...
      "mediaType": "banner",
      "source": "client",
      "cpm": 3.1148,
      "creativeId": "413049021",
      "currency": "JPY",
      ...
      "meta": {
        ...
        "advertiserDomains": [
          "kirinbeer-en.co.jp"
        ],
        ...
      },
      "adapterCode": "pubmatic",
      "originalCpm": 0.02,
      "originalCurrency": "USD",
      ...
      "bidder": "pubmatic",
      "adUnitCode": "div-gpt-ad-1498036666257-1",
      "timeToRespond": 211,
      ...
      "size": "300x250",
      "status": "rendered",
      ...
    }
  ]
}

```

図 2.2: HB オブジェクトの例

2.1.7 ペルソナ

ペルソナとは、特定の領域に関心を持つ架空の人物像のことである。広告主は自社のサービスのターゲットとなるペルソナを決め、そのペルソナと一致するユーザへの広告の配信を広告業者に依頼する。広告業者は、閲覧履歴などからユーザの関心を推定し、そのユーザへ広告を配信する。

本研究では、複数のペルソナを仮定し、ペルソナが Web 上で行う検索ワードを設定した。章 3 では、「食事に関心のある人」などの 5 つのペルソナを、章 5 では性別 (男, 女), 世帯年収 (600 万~1000 万, 1000 万~1500 万, 1500 万以上) の組み合わせで、 $2[\text{sex}] * 3[\text{income}] = 6[\text{persona}]$ のペルソナを仮定し、それぞれの検索ワードを設定した。

2.1.8 HBCollector

HB 観測のための Chrome ブラウザの拡張機能 HBCollector¹を開発した。アクセスしたサイトに Prebid.js が採用されているか確認し、採用されていた場合 HB 情報を自動で記録する。収集された HB 情報は CSV ファイルとして出力が可能である。これにより、手動での HB 入札情報の収集を効率的に行うことができる。また、サイト上の HB による広告について、入札額と、広告枠に対する入札数を可視化する機能を持つ。実行例を図 2.3 に示す。章 5 の手動観測にて使用した。



図 2.3: HBCollector による HB 可視化の実行例

¹ <https://www.kikn.fms.meiji.ac.jp/~yuto/code/HBcollector.zip>

2.2 従来研究

2.2.1 閲覧履歴が広告に与える影響および広告主とトラッカーのデータ共有に関する研究

Cook ら [13] は, OpenWPM を用いて HB の入札情報を解析し, 閲覧履歴が入札価格に与える影響を調査した. また, 広告を出稿する広告業者とトラッキング業者のデータ共有関係を明らかにした. ペルソナ, 購買意欲によって入札額が変化することが分かった. ただし, 閲覧行動を Alexa Top Sites の各カテゴリ上位サイトをクローリングすることで再現しており, 実際のユーザの検索行動を直接反映していなかった. さらに, 入札の時間的な変化を考慮しておらず, OpenWPM による測定結果と実際のユーザが利用するブラウザでの入札結果の違いについても明確にしていなかった.

2.2.2 実ユーザのブラウザ環境を用いた HB の入札情報の収集と分析

Eric ら [14] は, Chrome, Edge を利用している年齢, 性別, 民族等に統計的な偏りのない 186 人のアメリカ人参加者のブラウザ上で, HB 情報を収集し分析することで, ユーザの属性と広告の内容・価格の関係を調査した. その結果, 性別や人種, サイトのカテゴリによって広告のカテゴリが変化することが分かった. ただし, 実ユーザのブラウザを利用したことで, ユーザのどの行動が広告に影響を与えているか, 理想的な環境で厳密に観測することは困難であった. そのため, リターゲティング広告の影響に関する分析は, 定性的な評価にとどまった.

2.2.3 RTB におけるターゲティングの持続期間に関する調査

青山ら [15] は RTB によるターゲット広告のターゲティングが維持される期間について調査を行った. ターゲット広告は 22 日間以上表示されることが分かったが, ユーザの行動が広告の価格や入札数などへ与える影響は定量的に示されていなかった.

2.2.4 本研究の位置付け

これらの研究では広告のクリックなどユーザの行動や時間帯による影響が示されていなかった. 一般に入札値はユーザの活動に大きく依存し, アルゴリズムが秘匿されており, それが不安を増幅させているため, ユーザの行動に関する広告内容決定のアルゴリズムを明らかにする必要がある.

本研究では, ブラウザのプロファイル機能による理想的な環境で, ユーザの検索行動を広告主向けのデスクリサーチツールである DS.INSIGHT[20] を利用することで再現し, ユーザの属性による入札内容の変化を観測した. また, 自動観測の可能性について議論し, ターゲティングの継続期間について入札額, 入札数の変化を観測することで定量的に示した.

第3章 Header Bidding情報を用いたペルソナに応じたターゲティング広告の調査

3.1 目的

本章では、ユーザの行動が広告へ与える影響を明らかにするために、下記の問題を考える。

RQ1：ユーザの閲覧履歴が広告に与える影響は？

RQ2：ユーザの広告クリックが広告に与える影響は？

また、クローラによる自動観測では、大規模に広告の情報を収集できる一方で、広告業者による Bot 検知などの影響により、特定の広告が配信されない可能性がある [18]。その結果、収集されたデータが実際のユーザ環境における広告配信状況を正確に反映していない可能性を考慮する必要がある。そのため、本章では OpenWPM で収集された入札情報がどの程度手動観測と一致しているかを検証する。

RQ3：広告配信の自動観測は可能か？

3.2 提案方式

HB を用いて、次の手順で観測を行う。まず、特定の趣味・趣向を持った複数の仮定のペルソナを用意し、該当する Web サイトにアクセスして履歴を残す。作成したペルソナで HB を用いている Web サイトにアクセスし、趣味に応じた広告と入札値の分布を観測し分析する。更に、OpenWPM を用いて、これらの観測を自動的に行い、手動による結果と比較する。

これにより、広告を掲載している Web サイトを対象に、広告クリックなどのユーザの行動、ユーザの趣向、サイトの内容、サイトの人気度や広告業者の種類、時間帯などの要素が広告の内容・価格にどの程度影響を与えているかを明らかにすることを試みる。

3.3 実験概要

本研究ではユーザの閲覧履歴や広告クリックが広告に与える影響を定量的に示し、自動観測の妥当性を検証することを目的に、3つの観測実験を行った。

3.3.1 実験1：手動によるペルソナの影響の観測

閲覧履歴による広告の単価及び入札数に対する影響を調べるために、表 3.1 に示す 5 種類のペルソナを作成し、それぞれのペルソナで HB を利用している表 3.2 に示すサイトにアクセスし、広告単価や入札数を調査する。青山らによる先行研究 [15] と同様の手法でペルソナを作成した。実験は 2024 年 4 月 25 日の 12 時から、Windows11 の Chrome ブラウザで行った。

- (1) Chrome ブラウザのプロファイル機能を用いて、4 つの趣味・趣向に関するペルソナと閲覧履歴を持たないコントロールペルソナの、計 5 つのペルソナと、HB を使用しているパブリッシャーの 5 サイトについての 25 のプロファイルを作成する。 ($5[\text{persona}] \times 5[\text{site}] = 25[\text{profile}]$)
- (2) 各プロファイルにおいて、コントロールを除く各ペルソナに関連するワードを Google.com で検索し、上位 5 サイトにアクセスすることで閲覧履歴を持たせる。検索ワードを表 3.1 に示す。ただし、Amazon や楽天などの総合オンラインショッピングサイトは除いた。
- (3) HB を利用している表 3.2 のサイトにアクセスし、全入札の数、入札値、広告枠名、広告業者、広告主のドメインを収集する。
- (4) (3) を 3 回繰り返す。

表 3.1: ペルソナの検索ワード

ペルソナ	検索ワード
meal	宅配サービス, 生協
device	iphone, mac apple
life	健康サプリ, 老人ホーム
fashion	コーデ, パンプス
control	検索ワードなし

表 3.2: 実験の対象サイト (パブリッシャー)

実験 \ サイト	jmtj.jp ¹	eiga.com ²	jiji.com ³	kakaku.com ⁴
実験 1	○	○	○	○
実験 2	○	○	×	○
実験 3	○	○	○	○

3.3.2 実験2：広告クリックによる広告価格への影響調査

広告クリック前後の広告単価への影響を明らかにするために、次の手順で観測を行う。実験は 2024 年 4 月 29 日の 13 時から、Windows11 の Chrome ブラウザで行った。なお、jiji.com では落札されたに

¹ <https://jmtj.jp/> ² <https://eiga.com/> ³ <https://www.jiji.com/> ⁴ <https://kakaku.com/>

もかかわらず広告が表示されなかったため実験の対象外とした。

- (1) 閲覧履歴を持たないコントロールペルソナで、HB が有効に機能している 3 サイト分のプロフィールを作成する。
- (2) 表 3.2 のサイトにアクセスし、HB による広告が表示されるまでリロードを繰り返し。広告が表示された際の、入札数と、各入札の入札値、広告枠名、広告業者、広告主のドメインを収集する。
- (3) 表示された広告をクリックする。
- (4) 広告クリック直後の各入札を同様に収集する。
- (5) 以降 30 分おきに各入札を同様に収集する。これを 3 時間行う。

3.3.3 実験 3：自動によるペルソナの影響の観測実験

Bot 検知対策を施した、OpenWPM_{hide} で、実験 1 と同様の手法で、HB 情報を 3 分に 1 度収集し、広告価格とペルソナの影響を調べた。実験は 2024 年 5 月 3 日の 13 時から 24 時間行い、OS は Ubuntu 22.04 LTS を使用した。

3.3.4 実験結果

実験 1 の統計量を表 3.3, 表 3.4, 入札額と入札数の分布を図 3.2, 図 3.1 に示す。ペルソナについて、入札額の中央値は fashion ペルソナが最も低く 3.10 円 (CPM), life ペルソナが最も高く 15.86 円 (CPM) であり、最大で 5.1 倍の違いが見られた。入札数の平均はコントロールペルソナが最も少なく平均 5.00 件, life ペルソナが最も多く平均 23.00 件であり、最大で 16.6 倍の違いが見られた。

サイトについて、入札額の中央値は kakaku.com が最も低く 2.65 円 (CPM) で、eiga.com が最も高く 9.88 円 (CPM) であり、最大で 3.7 倍の違いが見られた。入札数の平均は kakaku.com が最も少なく 1.20 件, jiji.com が最も多く 19.60 件であり、最大で 16.33 倍の違いが見られた。ただし、各サイトが契約している SSP により違いが生じている可能性がある。また、入札数に関しては、各サイトの広告枠の数を考慮する必要がある。

図 3.3 に、広告業者ごとに集計した入札数と入札額の散布図を示す。入札価格が大きく異なり、入札数が少ないが入札額が高い、入札数が多いが入札額が低いなど、業者によって入札傾向の違いがみられた。入札数と入札額 (CPM) の間には緩やかな正の相関がある。相関係数は 0.26 である。

実験 2 の広告クリック前後の CPM の変化と時間変化を、図 3.4, 図 3.6 に示す。図 3.5, 図 3.7 は、入札額と入札数の分布を示す。広告クリックにより CPM が平均 15 倍、入札数は平均 1.4 倍と大きく上昇した。1 時間程度で元の入札額に戻ることが分かった。

実験 3 の 2024 年 5 月 3 日の 13 時から 24 時間の全ペルソナの CPM および入札数を 図 3.8, 図 3.9 に示す。時間帯によって差が出ることが分かった。図 3.11, 図 3.10 に、手動と OpenWPM による自動観測の入札額と入札数の比較を示す。手動での観測値を真値としたときの、OpenWPM の観測値の相対誤差率を求める。相対誤差率を

$$error = \frac{manual - OpenWPM}{manual}$$

と定義する。相対誤差率を表 3.5 に示す。

表 3.3: 実験 1 の入札額の統計量

	persona					site			
	fashion	device	meal	life	control	jmtj.jp	eiga.com	jiji.com	kakaku.com
mean	11.50	7.15	6.35	45.13	6.57	10.04	17.95	38.79	6.64
std	20.05	8.83	6.90	92.70	5.51	11.65	17.37	91.45	7.96
min	1.04	0.71	0.94	1.52	0.12	0.12	1.67	0.71	1.67
25%	1.67	1.67	1.67	4.82	2.91	1.67	6.20	3.10	2.41
50%	3.10	3.92	4.01	15.86	4.80	3.43	9.88	7.88	2.65
75%	11.92	8.27	8.02	34.37	8.80	14.74	28.88	24.26	6.99
max	92.79	33.78	29.30	586.60	22.10	57.43	79.42	586.60	22.10

表 3.4: 実験 1 の入札数の統計量

	persona					site			
	fashion	device	meal	life	control	jmtj.jp	eiga.com	jiji.com	kakaku.com
mean	10.00	6.00	6.25	23.00	5.00	9.60	6.60	19.60	1.20
std	9.54	5.57	9.84	16.21	4.32	11.72	7.83	12.05	0.45
min	1.00	1.00	1.00	2.00	1.00	2.00	1.00	11.00	1.00
25%	5.00	3.00	1.00	15.50	2.50	3.00	1.00	12.00	1.00
50%	9.00	5.00	1.50	25.00	4.00	4.00	5.00	14.00	1.00
75%	14.50	8.50	6.75	32.50	6.50	9.00	6.00	21.00	1.00
max	20.00	12.00	21.00	40.00	11.00	30.00	20.00	40.00	2.00

3.3.5 考察

各 RQ に対して、以下のように考察する。

ユーザの閲覧履歴が広告に与える影響は？

各ペルソナ間で、入札額の中央値は最大 5.1 倍、入札数平均が最大 16.6 倍異なることが確認された。よって、閲覧履歴が広告に影響を与えると考えられる。また、サイトによって入札額の中央値は最大 3.7 倍、入札数平均は最大 16.33 倍異なることから、広告枠のサイズやサイトの内容・人気度などにより入札価格が変化すると考えられる。ただし、広告主の予算状況、時間帯による閲覧者の変動、閲覧者による閲覧頻度の変動などの様々な影響があり、広告は時間帯によって変動するので、それらの外乱に対して、観測された入札額や入札数の変化が有意かどうかを検討する必要がある。

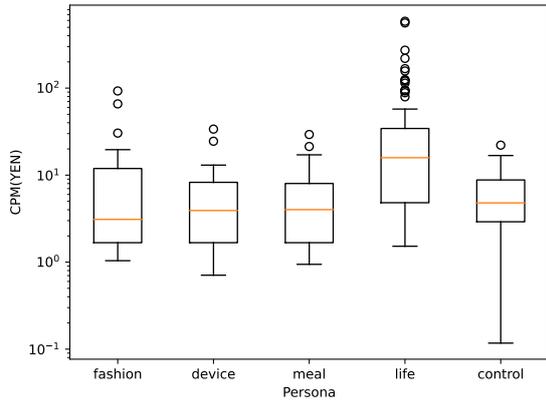


図 3.1: 各ペルソナの入札額の分布

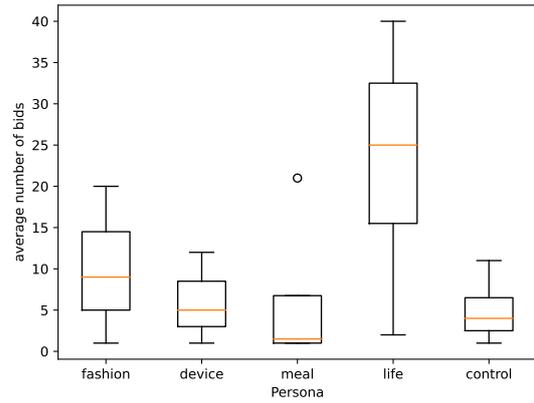


図 3.2: 各ペルソナの入札数の分布

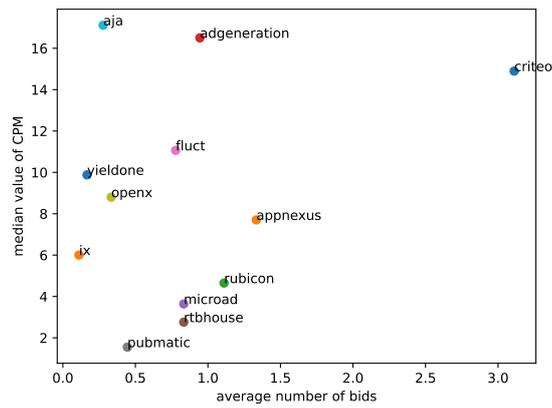


図 3.3: 広告業者の入札回数と入札額

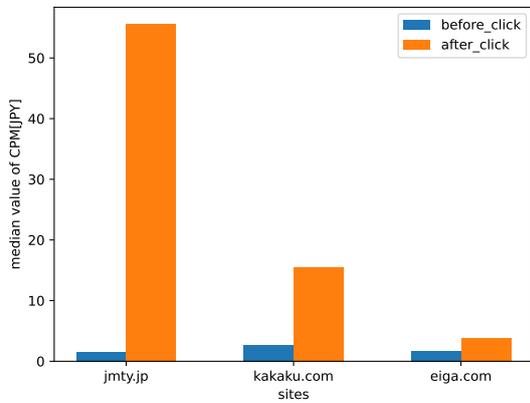


図 3.4: 広告クリック前後の入札額の変化

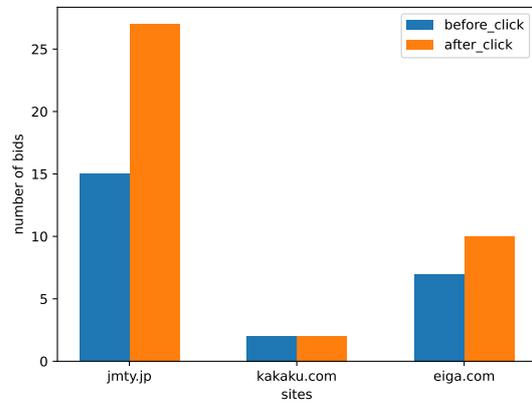


図 3.5: 広告クリック前後の入札数の変化

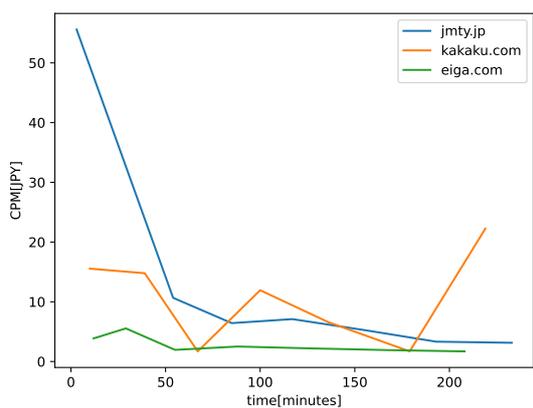


図 3.6: 広告クリック後の入札額の時間的変化

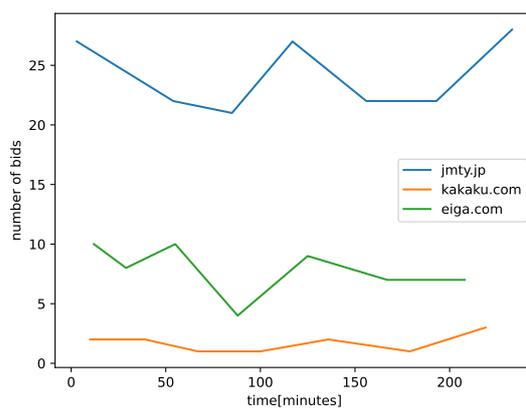


図 3.7: 広告クリック後の入札数の時間的変化

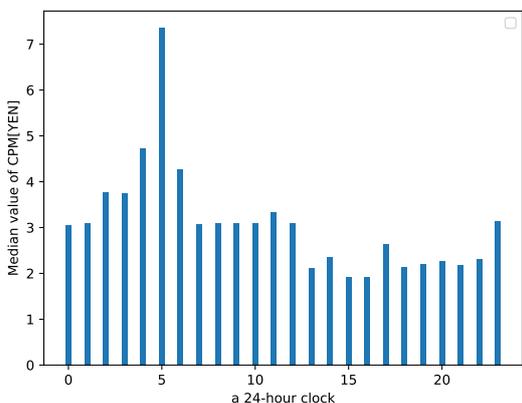


図 3.8: OpenWPM における 2024 年 5 月 3 日の 24h の入札額の中央値

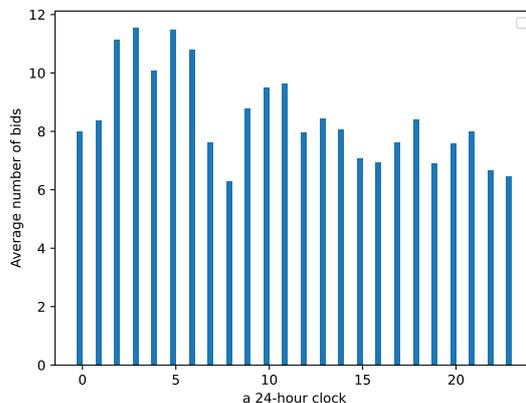


図 3.9: OpenWPM における 2024 年 5 月 3 日の 24h の入札数平均

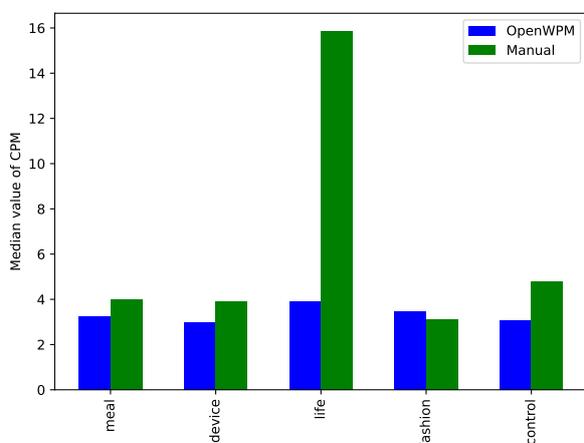


図 3.10: OpenWPM における各ペルソナの入札額中央値

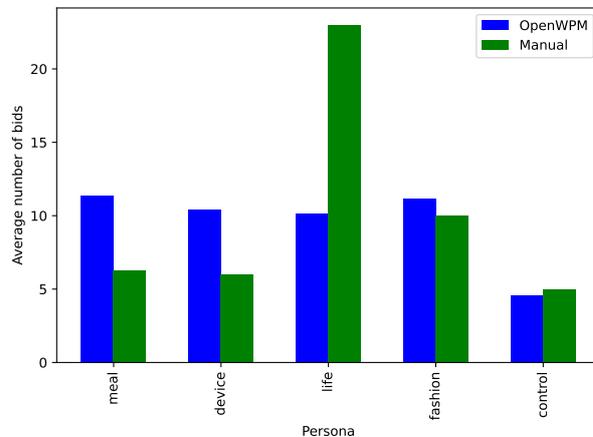


図 3.11: OpenWPM における各ペルソナの入札数平均

表 3.5: OpenWPM の誤差率

persona	cpm	bid
meal	-0.19	0.82
device	-0.24	0.73
life	-0.75	-0.56
fashion	0.12	0.12
control	-0.36	-0.08

ユーザの広告クリックが広告に与える影響は？

広告クリックにより入札額は平均 15 倍上昇し、入札数は 1.4 倍上昇したが、サイトにより上がり幅に差があり、業者によってユーザの行動に対する評価が異なることが分かった。

広告のクリックという動作を見ているのではなく、サイトに訪問したユーザに対するリターゲティング広告の可能性がある。広告クリックによる影響による入札の差分を確認することができた。

広告配信の自動観測は可能か？

OpenWPM を用いた観測の自動化は、手動の観測と入札数、入札価格に多少の誤差が生じた。原因として、User-Agent が異なること、Bot 検知により入札がされていない可能性が考えられる [18]。そのため、自動観測は可能だが、多少の誤差が生じることを考慮する必要があると言える。しかし、図 3.11 によると、OpenWPM の方が手動よりも入札数が多いサイトが多い。Bot と検知されたら、入札数が減るはずなので、更なる調査が必要である。また、24 時間観測した結果、入札額、入札数ともに時間帯によって変化する現象が観測され、特に入札額は深夜から早朝の時間帯に高騰した。

3.4 まとめ

本章では各ペルソナによって入札数、入札額が異なることから、閲覧履歴が広告に影響を与えることを定量的に示した。また、サイトによって広告価格が変わること、広告をクリックにより大きく広告価格や入札に影響が出ることが分かった。

入札額が深夜から早朝にかけ高騰した現象について、その時間帯はインターネットのアクティブなユーザ数が減少し、インプレッション数が減少することで少ない枠を巡って広告オークションが激化することで入札額が高騰することが原因であると考え、第 4 章で検証する。

より実際のユーザの行動に近いように、ペルソナの作り方を改善する必要がある。また、オンライン広告市場は日々変化するために、今後より大規模な観測実験を行う必要がある。そのため、ペルソナを性別や世帯年収とすることで実ユーザに近づけ、より現実的なワードを用いて大規模な実験を行う。また、入札者の振る舞いを表す数理モデルにより、ユーザのどの要素がどれほど広告価格に影響を与えているかを定量的に示す。本実験については第 5 章で述べる。

第4章 インプレッション数による入札額の変化

4.1 目的

第3章で、図3.8が示すように、HBの入札価格が時間帯により変動する現象が観測された。特に、3時から5時の時間帯に入札額が高騰する現象が観測された。これは、深夜のユーザ数が少なくインプレッション数に限りがある時間帯ではインプレッションの価値が上がり、入札額が高騰するのではないかと考えた。この仮説を検証するために、まず深夜に入札額が高騰する現象が偶然によるものではないことを確かめるために、時期を変え、長期間観測しても同様の現象が確認できるかを確かめた。

2024年6月3日0時～2024年6月10日0時に、閲覧履歴を持たないコントロールペルソナで、3.3.3節と同様のサイトを対象に、3分に1度HB情報を観測した。図4.1、図4.2に入札額、入札数を示す。深夜の時間帯の入札額が高くなる現象は同様に確認された。そのため、本章ではユーザ数の減少が広告枠の競争を高める要因であるという仮説を検証することを目的とする。

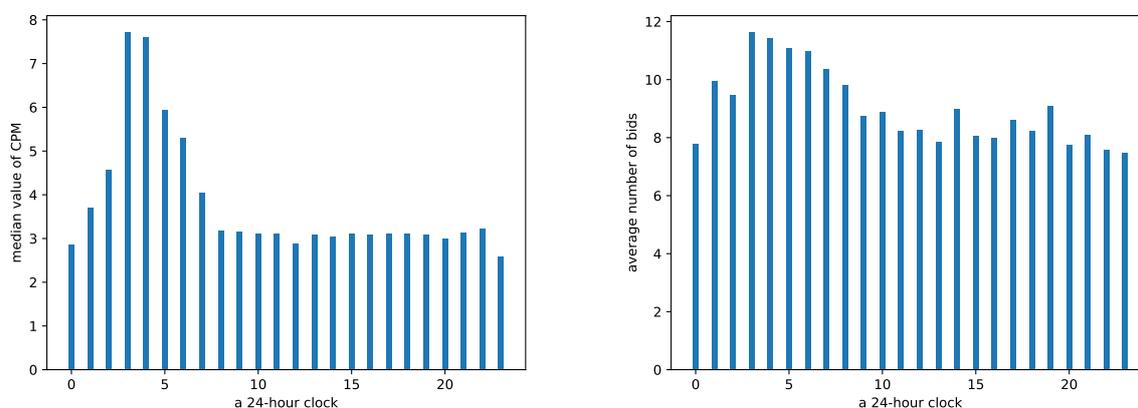


図 4.1: 2024年6月3日～6月10日の時間帯ごとの入札額中央値

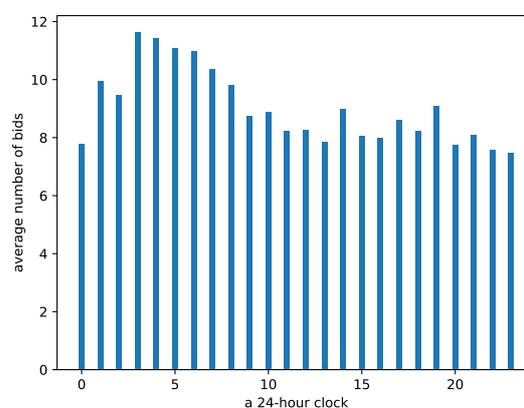


図 4.2: 2024年6月3日～6月10日の時間帯ごとの入札数平均

4.2 アクティブなインターネットユーザ数

総務省情報通信政策研究所 [19] は、13歳から69歳までの男女1500人を対象に、インターネット等のメディアの利用時間や時間帯等の調査を行った。時間帯によるインターネットの閲覧の行為者率の変化を図4.3に示す。

図3.8の入札額とインターネットの閲覧の行為者率のピアソン相関係数は-0.711、-0.762と算出され、アクティブユーザが少ないほど入札額が高くなる負の相関が観測された。この結果から、アクティブユーザの減少、つまりインプレッション数の減少が入札額の上昇に寄与している可能性が示唆される。

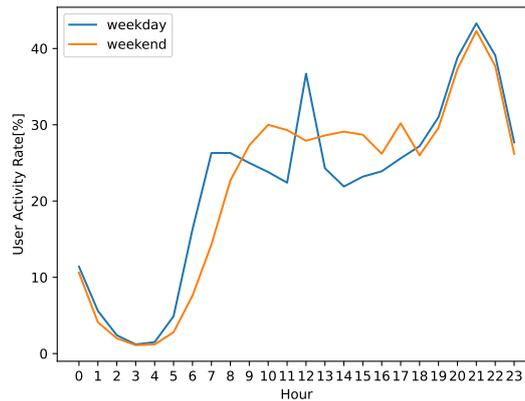


図 4.3: 時間帯ごとのインターネットの閲覧の行為者率の割合の変化 (出典: 総務省 情報通信政策研究所 [19] より筆者作成)

4.3 インプレッション数と入札額の仮説を検証するための調査

4.3.1 実験概要

サイトを閲覧しているユーザ数（インプレッション数）と入札値の関係を評価したいが、実運用しているサイトのアクセス履歴を入手するのは困難である。そこで、人工的に閲覧数を増加して、入札額の変化を観測する。ただし、機械的に大量のアクセスを引き起こすのは、倫理上問題である。そこで、実験の影響が小さいと考えられるユーザ数の少ない深夜4時に、短時間でアクセスする頻度が増加すると、HBの入札額に有意な変化が生じると仮定し、次の条件で比較を行う。実験には $\text{OpenWPM}_{\text{hide}}$ を用いた。

4.3.2 実験条件

深夜の4時から下記の手順で実験を行った。図4.4のタイムチャートで示される閲覧パターンを加えた。

- (1) 5秒に1度、パブリッシャーのサイトに c_0 回 (少量) のアクセスを行う。
- (2) 20分間待機する。
- (3) 5秒に1度、パブリッシャーのサイトに c_1 回 (多量) のアクセスを行う。
- (4) (1)-(3) を3回繰り返す。

条件を変え、4回の観測実験を行った。実験条件を表4.1に示す。

4.3.3 実験結果

図4.5, 4.7, 4.9, 4.11に各条件の入札額の分布を示す。ただし、少数アクセスは $num = 0, 2, 4$, 多数アクセスは $num = 1, 3, 5$ である。図には $Q3 + 1.5 * IQR$ より大きい値または $Q1 - 1.5 * IQR$ 未満の値は表示しない。多数アクセス、少数アクセスの結果を集計し、図4.6, 4.8, 4.10, 4.12に示す。

表 4.1: 実験条件

	日時	c_0	c_1	対象サイト
条件 1	2024/12/16	5	60	eiga.com
条件 2	2024/12/17	5	60	jmtj.jp
条件 3	2024/12/19	50	200	jiji.com
条件 4	2024/12/20	20	200	jiji.com

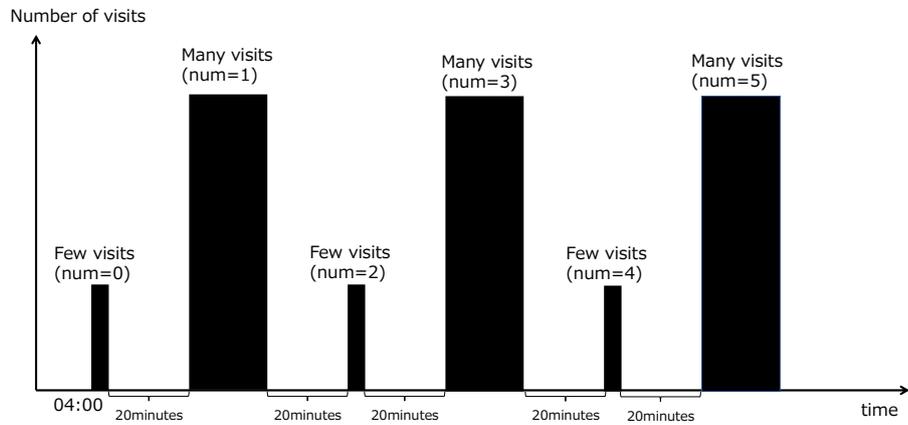


図 4.4: 実験の流れ

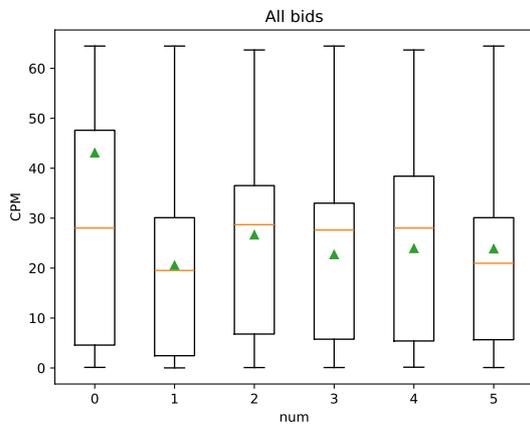


図 4.5: 条件 1 の入札額の分布

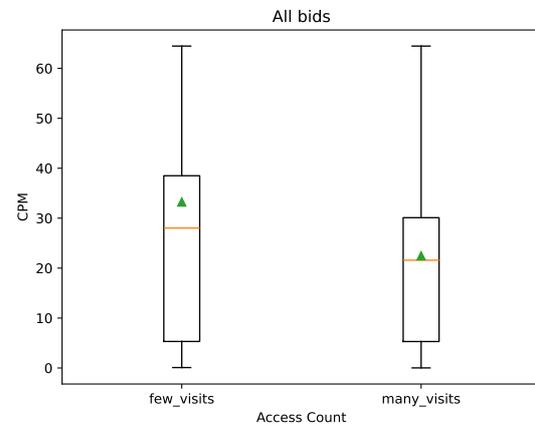


図 4.6: 条件 1 の入札額の分布 集計

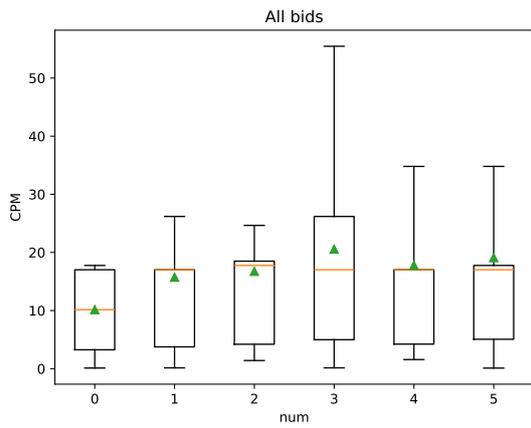


図 4.7: 条件 2 の入札額の分布

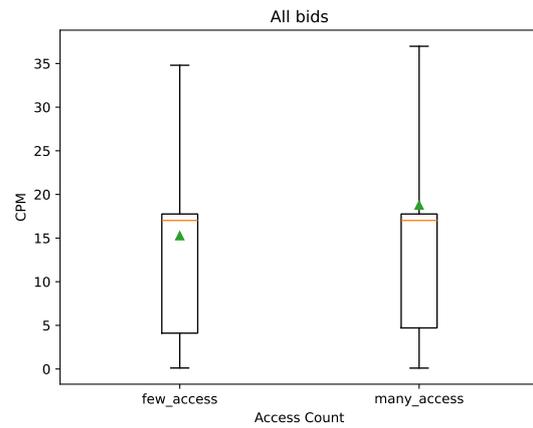


図 4.8: 条件 2 の入札額の分布

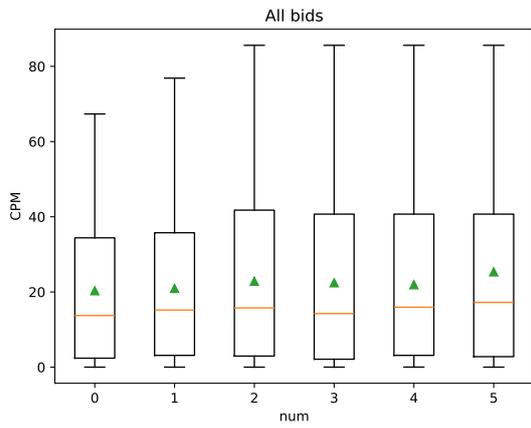


図 4.9: 条件 3 の入札額の分布

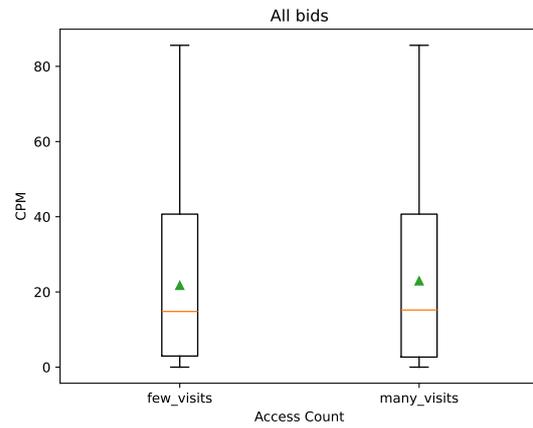


図 4.10: 条件 3 の入札額の分布 集計

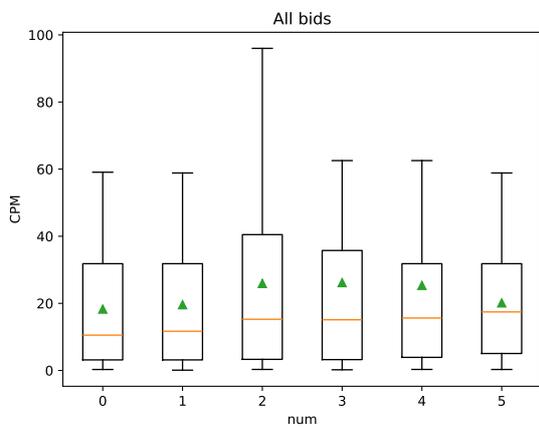


図 4.11: 条件 4 の入札額の分布

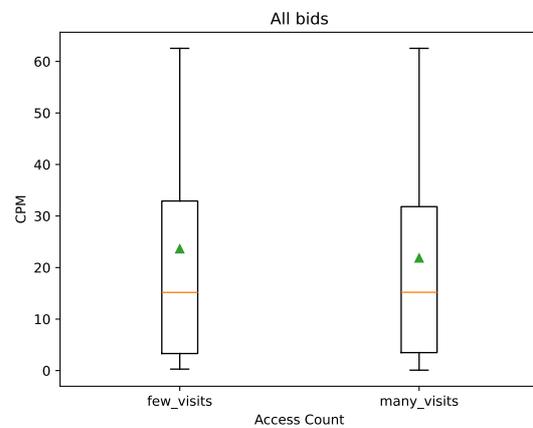


図 4.12: 条件 4 の入札額の分布 集計

少数アクセス, 多数アクセスの入札の入札額の平均値に有意な差があるかを確認するために, 独立な2群のt検定を行った. 有意水準5%で分散の等質性の検定を行い, 全ての組み合わせで同質性が確認できなかったため Welch の t 検定を実施した. 検定の結果を表 4.2 に示す. 有意水準5%を**, 1%を***で記す.

表 4.2: 少数, 多数アクセスの入札額の平均値に対する t 検定の結果

	少数アクセス平均値	多数アクセス平均値	P 値	
条件 1	33.232	22.455	0.0083	***
条件 2	15.269	18.783	0.0064	***
条件 3	21.743	22.940	0.021	**
条件 4	23.625	21.838	0.13	

4.3.4 考察

条件 1 では少数アクセスの入札額の平均値が高く, 差は統計的に有意であった. 従って, 仮説は検証出来たと考える. しかし, 一方, 条件 2, 3 では多数アクセスの入札額の平均値が有意に低かった. 条件 4 では有意差は見られなかった. そのため, 少数アクセスと多数アクセスに差が見られなかったと言える. そのため, インプレッション数と入札額の間関係を確かめるためには, 実験条件を変えるか, あるいはパブリッシャーを運用しているサイトとの協力による別のアプローチが必要であると考えられる.

4.4 まとめ

本章では, インプレッション数が少ない時間帯に広告価格が高騰するという仮説の検証を行った. インターネットのアクティブユーザ数と入札額間には強い負の相関が見られたものの, 少数アクセスと多数アクセスの間に差は見られなかった. 他の要因 (例えば, 広告主の予算が 0 時にリセットされることで, 深夜帯にオークションが激化し早期に予算が消費されるため, 午後には入札価格が落ち着いている可能性) を考慮しながら, シミュレーションを含むより一層の分析が必要である.

第5章 オンラインターゲティング広告における ユーザ閲覧履歴入札の数理モデル

5.1 目的

本章では、性別、収入、サイト訪問履歴の有無を説明変数として、ウェブサイトを訪問するユーザについての入札を定量化することを試みる。各属性が入札額に与える影響を明確にすることで、ユーザが自身の属性に基づく広告価値を把握し、広告配信の透明性を向上させることが期待される。

5.2 DS.INSIGHT を用いた実践的ペルソナの設計

DS.INSIGHT は、LINE ヤフー社が提供するデスクリサーチツールであり、Yahoo! JAPAN のユーザから収集された検索データ等を基に、オンライン上のユーザの興味・関心の分析を行うためのツールである [20]。本ツールには Persona 機能があり、設定したペルソナ群が頻繁に検索するワードを抽出することができる。本研究ではペルソナを性別(男, 女)、世帯年収(600万~1000万, 1000万~1500万, 1500万以上)の組み合わせとして定義し、6種類のペルソナに基づいて関連するワードを抽出した。各ペルソナの検索キーワードを表 5.1 に示す。

5.3 リターゲティングによる数理モデルの提案

実験方法

ユーザの属性別入札データを収集する目的で、表 5.1 に示す 7 種類のペルソナを作成し、それぞれのペルソナで HB を利用している表 5.2 にアクセスし、HB 情報を集めた。なお、実ユーザの振る舞いに近づけるために HBCollector による手動観測を行った。実験は 2025 年 1 月 13 日に、Windows11 の PC で、HBCollector を追加した Chrome ブラウザで行った。

- (1) Chrome ブラウザのプロファイル機能を用いて、性別(男, 女)、世帯年収(600-1000万, 1000万-1500万, 1500万以上)の組み合わせの 6 種類のペルソナと、閲覧履歴を持たないコントロールペルソナの、計 7 つのプロファイルを作成する。
- (2) 各プロファイルにおいて、コントロールを除く各ペルソナに関連するワードを Google.com で検索し、上位 5 サイトにアクセスすることで閲覧履歴を持たせる。検索ワードを表 5.1 に示す。Cookie 同意バナーが表示された場合同意する。

表 5.1: ペルソナの検索ワード

ペルソナ	検索ワード
女性, 世帯年収 1500 万以上	ヴァンクリーフアーペル, 一休 レストラン予約, パレスホテル東京, マックスマーラ, エルメス バッグ, ブシュロン, 一休 ホテル, マンダリンオリエンタル東京, tasaki, ショーメ, ヴアレクストラ, ペニンシュラ東京, アレクサンドルドゥパリ, 近沢レース オンライン, ポメラート
女性, 世帯年収 1000 万~1500 万	ヴァンクリーフアーペル, フェイラー, 一休 レストラン予約, アンテプリマ, ボールアンドチェーン, トゥモローランド, 日本橋高島屋, 日本橋三越, 一休レストラン, 伊勢丹新宿, 銀座三越, エルベシャプリエ, パレスホテル東京, 伊勢丹オンライン, ルミネ 10%オフ
女性, 世帯年収 600 万~1000 万	英検, 英検 日程, プティマイン, スナップスナップ, 英検 合格発表, バースデイ オンライン, 読書感想文 書き方, スマイルゼミ, 英検 3 級, 出産予定日 計算, キッザニア甲子園, 英検準 2 級, アカチャンホンポ ネット, チャレンジタッチ, ファミリア オンライン
男性, 世帯年収 1500 万以上	pubmed, m3, 食べログ 東京, 日本オープンゴルフ選手権 2024, lpga 日本女子プロゴルフ協会, m3.com, 太平洋クラブ, skyscanner, ゴルフ会員権, omakase 予約, 東京ゴルフ倶楽部, ゴル天, 大利根カントリークラブ, google scholar, ヴァッシュロンコンスタンタン
男性, 世帯年収 1000 万~1500 万	ゴルフパートナー, テーラーメイド, キャロウェイ, ゴルフ 5, zozo チャンピオンシップ 2024, gdo, タイトリスト, 楽天 gora, アコーディアゴルフ, 楽天ゴルフ, スコッティキャメロン, gdo ゴルフ予約, 大学ラグビー選手権 2024, つるやゴルフ, jlpga
男性, 世帯年収 600 万~1000 万	エックスビデオ, ランクル 250, ワンピース ネタバレ, ドラクエウォーク, ランドクルーザー 250, デリカ d5, ウインターカップ 2024, 辻井ほのか, ブレイキングダウン 14, キングダム 最新刊, プロスピ a, メンズ 髪型 ショート, ランクル 300, ランクル 70, インターハイ 2024 サッカー
コントロール	検索ワードなし

(3) HB を利用している表 5.2 のサイトに 3 回ずつアクセスし、各入札の入札値、広告主のドメインを収集する。

表 5.2: 実験の対象サイト

	URL
1	kakuyomu.jp ¹
2	trilltrill.jp ²
3	weathernews.jp ³
4	natalie.mu ⁴
5	eiga.com ⁵
6	jmtj.jp ⁶
7	full-count.jp ⁷
8	jprime.jp ⁸
9	fashion-press.net ⁹
10	jiji.com ¹⁰

5.3.1 実験結果

手動観測について、ターゲティング、特にリターゲティング(自社サイトに訪問履歴のあるユーザに対する広告)は時間経過あるいは広告表示回数により影響が大きく変化するため、ペルソナ作成時から最も近い 100 件の入札情報を抽出し、ペルソナ別の入札額の分布を図 5.1 に示す。ただし、図には $Q3 + 1.5 * IQR$ より大きい値または $Q1 - 1.5 * IQR$ 未満の値は表示しない。この結果を属性別に分け、表 5.2、表 5.3 に、性別、世帯収入ごとの分布を示す。

なお、性別については以下のように表記する：

- 女性：female
- 男性：male

世帯年収については、以下のように表記する：

- 1500 万以上：high
- 1000 万～1500 万：medium
- 600 万～1000 万：low

¹ <https://kakuyomu.jp/> ² <https://trilltrill.jp/> ³ <https://weathernews.jp/> ⁴ <https://natalie.mu/>
⁵ <https://eiga.com/> ⁶ <https://jmtj.jp/> ⁷ <https://full-count.jp/> ⁸ <https://www.jprime.jp/> ⁹ <https://www.fashion-press.net/> ¹⁰ <https://www.jiji.com/>

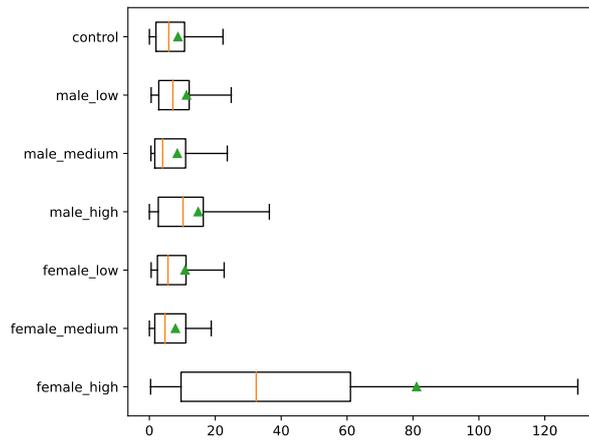


図 5.1: 各ペルソナの入札額

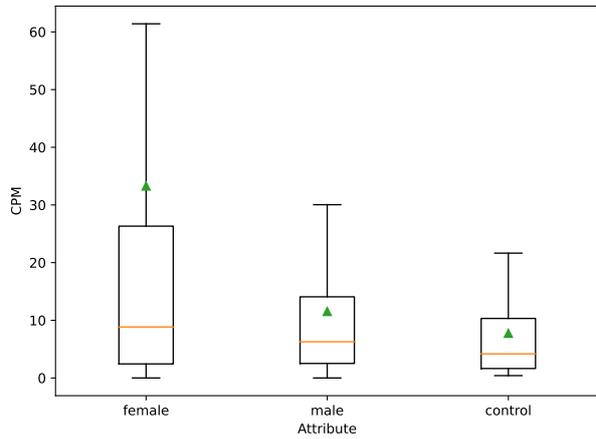


図 5.2: 性別ごとの入札額

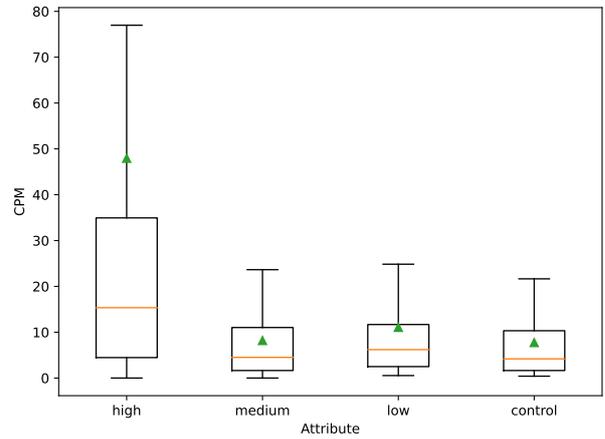


図 5.3: 世帯年収ごとの入札額

表 5.3: ペルソナごとの入札額の統計量

	female_high	female_medium	female_low	male_high	male_medium	male_low	control
count	100.00	100.00	100.00	100.00	100.00	100.00	100.00
mean	81.09	7.91	10.83	14.82	8.48	11.29	7.75
std	127.91	9.12	17.07	19.04	12.44	15.87	10.74
min	0.38	0.00	0.54	0.00	0.48	0.55	0.41
25%	9.61	1.65	2.44	2.71	1.65	2.86	1.65
50%	32.42	4.74	5.66	10.22	4.04	7.14	4.19
75%	61.03	11.06	11.13	16.36	11.01	12.08	10.33
max	510.48	50.58	99.56	95.46	73.52	133.70	72.91

表 5.4: 性別ごとの入札額の統計量

	female	male	control
count	300.00	300.00	100.00
mean	33.27	11.53	7.75
std	81.79	16.17	10.74
min	0.00	0.00	0.41
25%	2.44	2.52	1.65
50%	8.85	6.28	4.19
75%	26.33	14.08	10.33
max	510.48	133.70	72.91

表 5.5: 収入ごとの入札額の統計量

	high	medium	low	control
count	200.00	200.00	200.00	100.00
mean	47.95	8.20	11.06	7.75
std	97.07	10.88	16.44	10.74
min	0.00	0.00	0.54	0.41
25%	4.47	1.65	2.50	1.65
50%	15.35	4.51	6.20	4.19
75%	34.94	11.03	11.68	10.33
max	510.48	73.52	133.70	72.91

性別, 世帯年収間の入札額の平均値に, コントロールペルソナとの有意な差があるかを確認するために, 独立な2群のt検定 [24] を行った. まず, t検定を実行するにあたり, 各属性とコントロールペルソナ間の分散の等質性の検定を有意水準5%で行った. その結果, mediumのみ分散の等質性が確認された. そのため, mediumとcontrolの比較ではStudentのt検定を, その他とcontrolの比較ではWelchのt検定を実施した. 検定の結果を表5.6に示す. 有意水準5%を**, 1%を***で記す. medium以外でcontrolとの有意な差が確認された.

表 5.6: 各属性とコントロールペルソナの平均値に関する t 検定の結果

属性	p 値	
性別	female	2.5×10^{-7} ***
	male	0.0083 ***
世帯年収	high	2.6×10^{-8} ***
	medium	0.74
	low	0.037 **

ペルソナ作成時にアクセスしたドメインからの入札をリターゲティング広告であると定義し, その数を表5.7に示す. また, リターゲティング広告による入札と, それ以外の入札の入札額の分布, 統計量を図5.4, 表5.8に示す.

リターゲティング広告と, それ以外の入札の入札額の平均値に有意な差があるかを確認するために, 独立な2群のt検定を行った. 有意水準5%で分散の等質性の検定を行い, 同質性が確認できなかったためWelchのt検定を実施した. $p = 1.5 \times 10^{-8}$ (***) であり, 平均値の差に有意な差が確認された.

表 5.7: 各ペルソナのリターゲティング広告による入札数

persona	retargeting ads
female_high	88/100
female_medium	44/100
female_low	47/100
male_high	67/100
male_medium	48/100
male_low	45/100
全体	349/600

5.3.2 線形重回帰による数理モデルの提案

ユーザの属性や訪問履歴 (リターゲティング) が入札額にどの程度影響を与えているかを示すために, 線形重回帰モデルを提案する.

入札額を目的変数に, 性別, 世帯年収, リターゲティングを説明変数に重回帰分析を適用し, 求められた次の線形モデルを提案する. ただし性別, 世帯年収, リターゲティングはカテゴリ変数であるた

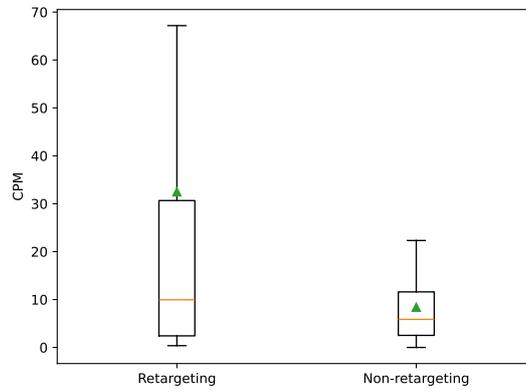


図 5.4: リターゲティングとそれ以外の入札の入札額の分布

表 5.8: リターゲティング広告とそれ以外の入札の入札額の統計量

	Retargeting	Non-Retargeting
count	349.00	251.00
mean	32.47	8.40
std	76.43	11.30
min	0.38	0.00
25%	2.43	2.52
50%	9.95	5.87
75%	30.66	11.60
max	510.48	133.70

め, ダミー変数に変換する. 世帯年収の基準カテゴリは medium とする.

$$\log(CPM) = \beta_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5$$

係数, P 値, 定義域を表 5.9 に示す.

表 5.9: モデルの係数, p 値, 定義域

係数		Estimate	p 値	定義域
β_1		1.013	9.2×10^{-15} ***	
性別	female β_2	0.304	0.0081 **	0, 1
世帯年収	low β_3	0.370	0.0084 **	0, 1
	high β_4	1.039	1.1×10^{-12} ***	0, 1
Retargeting β_5		0.516	2.1×10^{-5} ***	0, 1

5.4 時間経過による入札額, 入札数の変化

5.4.1 実験方法

入札額, 入札数の推移を長期間観測し, ユーザに対する広告主の関心が時間経過によりどのように変化するかを確かめる目的で, 表 5.1 に示す 7 種類のペルソナを作成し, それぞれのペルソナで HB を利用している表 5.2 にアクセスし, HB 情報を集めた. なお, 長期間高頻度に観測するために OpenWPM_{hide} による自動観測を行った. 2025 年 1 月 15 日にペルソナを作成し, 2025 年 1 月 16 日 0 時~2025 年 1 月 17 日 0 時, 2025 年 1 月 18 日 0 時~2025 年 1 月 25 日 0 時に観測を行った. 表 5.10 に実験の工程と日時を示す.

- (1) OpenWPM のプロファイル機能を用いて, 手動観測と同様にペルソナを作成する.
- (2) 各プロファイルで表 5.2 のサイトにランダムにアクセスする.
- (3) 1 分間待機する.
- (4) (2)-(3) を実験期間中繰り返す.

表 5.10: 実験の日程

内容	日時
ペルソナ作成	2025 年 1 月 15 日 18 時~2025 年 1 月 15 日 22 時
観測 1	2025 年 1 月 16 日 0 時~2025 年 1 月 17 日 0 時
観測 2	2025 年 1 月 18 日 0 時~2025 年 1 月 25 日 0 時

5.4.2 実験結果

手動観測では female_high ペルソナが最も閲覧履歴による影響を受けた。そのため、female_high ペルソナの入札額、入札数の変化を図 5.5, 5.6, 図 5.7, 図 5.8 に示す。ただし、データのノイズが大きいため、移動平均による平滑化を行った。移動平均の定義を下記に示す。

$$SMA_n = \frac{1}{size} \sum_{i=0}^{size-1} y_{n+i} (1 \leq n \leq N - size + 1)$$

ここで、 y_n は n 回目のアクセス時の観測値、 $size$ はウィンドウサイズである。観測 1 の結果には $size = 5$ 、観測 2 の結果には $size = 20$ を適応した。

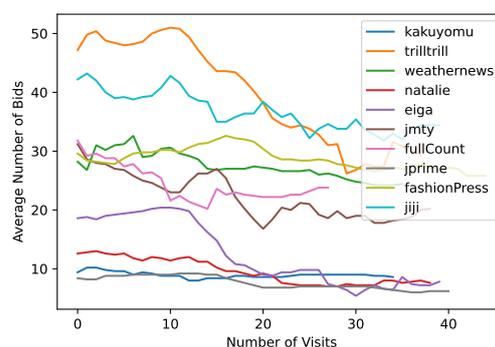
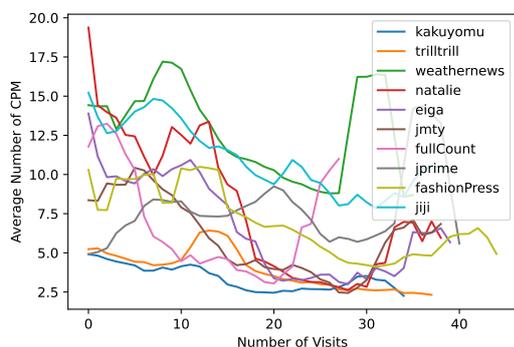


図 5.5: female_high ペルソナの入札額の変化: 観測 1 図 5.6: female_high ペルソナの入札数の変化: 観測 1

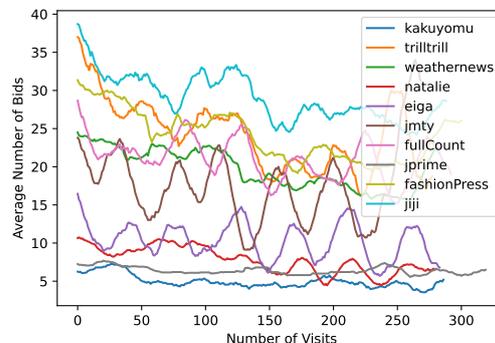
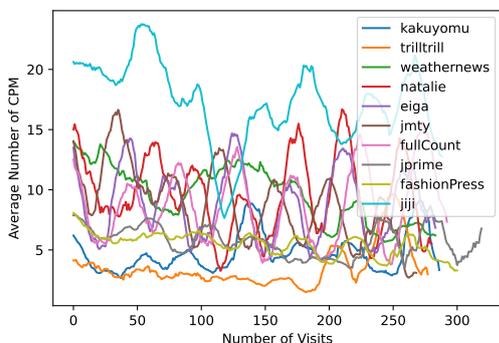


図 5.7: female_high ペルソナの入札額の変化: 観測 2 図 5.8: female_high ペルソナの入札数の変化: 観測 2

アクセス回数による入札額、入札数の変動を定量的に示すために、アクセス回数を説明変数、目的変数を入札額、入札値とした単回帰分析を行った。観測 1 および観測 2 における切片、係数、決定係数 R^2 を表 5.11, 5.12 に示す。回帰直線を図 5.9 および図 5.10, 図 5.11, 図 5.12 に示す。入札額において、観測 1 の傾きの平均は-0.128, 観測 2 の傾きの平均は-0.0099 であり、入札数において、観測 1 の傾きの平均は-0.252, 観測 2 の傾きの平均は-0.0159 であることから、ペルソナ作成から 24 時間程度までは広告主の関心が急速に低下し、それ以降は大きな変化見られないことが示唆される。

表 5.11: 観測 1 の入札額, 入札数の線形単回帰分析

サイト	CPM			入札数		
	傾き	切片	決定係数 R^2	傾き	切片	決定係数 R^2
kakuyomu.jp	-0.059	4.408	0.608	-0.022	9.320	0.202
trilltrill.jp	-0.090	5.590	0.637	-0.738	53.387	0.880
weathernews.jp	-0.114	14.742	0.177	-0.185	30.760	0.750
natalie.mu	-0.309	13.841	0.630	-0.174	12.867	0.849
eiga.com	-0.205	10.700	0.600	-0.427	20.912	0.829
jmtj.jp	-0.129	8.492	0.382	-0.289	27.767	0.745
full-count.jp	-0.175	9.273	0.177	-0.279	28.229	0.550
jprime.jp	0.083	6.210	0.179	-0.082	9.271	0.782
fashion-press.net	-0.129	9.798	0.631	-0.089	30.692	0.410
jiji.com	-0.155	14.057	0.694	-0.242	41.466	0.770

表 5.12: 観測 2 の入札額, 入札数の線形単回帰分析

サイト	CPM			入札数		
	傾き	切片	決定係数 R^2	傾き	切片	決定係数 R^2
kakuyomu.jp	0.006	3.844	0.111	-0.006	5.937	0.436
trilltrill.jp	0.009	2.264	0.186	-0.037	29.276	0.425
weathernews.jp	-0.021	12.693	0.591	-0.025	23.348	0.650
natalie.mu	-0.047	10.284	0.015	-0.016	10.141	0.639
eiga.com	-0.002	9.114	0.003	-0.007	11.399	0.073
jmtj.jp	-0.025	12.198	0.413	0.001	17.474	0.0003
full-count.jp	0.005	7.934	0.020	-0.012	22.972	0.173
jprime.jp	-0.004	6.434	0.109	-0.002	6.806	0.229
fashion-press.net	-0.007	6.603	0.489	-0.027	28.144	0.520
jiji.com	-0.013	19.231	0.121	-0.028	33.225	0.621

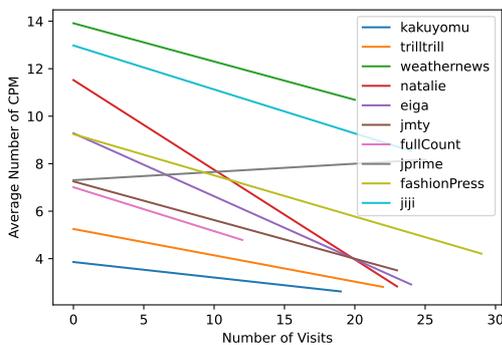


図 5.9: female_high ペルソナの入札額の回帰直線: 観測 1

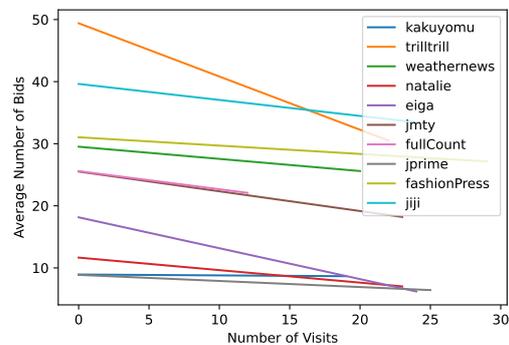


図 5.10: female_high ペルソナの入札数の回帰直線: 観測 1

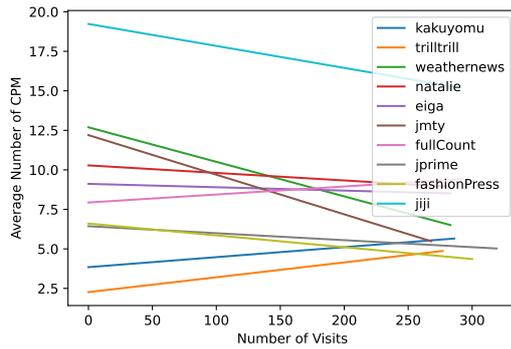


図 5.11: female_high の入札額の回帰直線: 観測 2

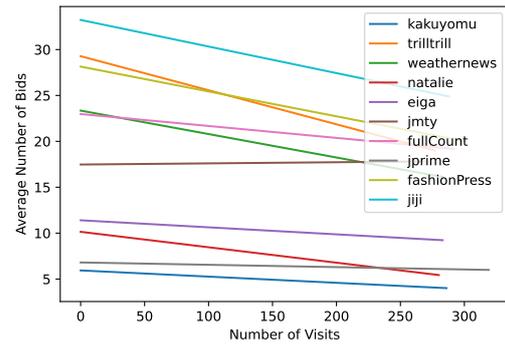


図 5.12: female_high の入札数の回帰直線: 観測 2

5.5 リターゲティング広告の影響検証実験

5.5.1 実験概要

5.3 節の観測結果の中で、入札の一部がリターゲティング広告によるものであることを確かめるために実験を行った。

female_high ペルソナでは広告主 maxmara.com¹¹，male_medium ペルソナでは広告主 callaway-golf.jp¹²からの入札が多数観測された。そのため、ペルソナ作成時にこれらのドメインのサイトにアクセスした場合、アクセスしなかった場合を比較し、リターゲティングによる影響を検証した。

HB 情報を収集するパブリッシャーは、広告枠の数が多いい trilltrill.jp, jiji.com とした。

5.5.2 実験方法

- (1) Chrome ブラウザで4つのプロファイルを作成する。(2[persona]*2[access, not access] = 4[profile])
- (2) 各ペルソナにおいて、プロファイル P1 では表 5.1 のペルソナに関連するワードを Google.com で検索し、上位 5 サイトにアクセスする。プロファイル P2 ではワードを検索するが、検索結果の中で指定した広告主のドメイン以外のサイトのみにアクセスする。
- (3) 6 時間待機する。
- (4) trilltrill.jp, jiji.com にアクセスする。
- (5) (4) を 5 回繰り返す。

5.5.3 実験結果

各プロファイルの広告主の分布を図 5.13, 5.14, 5.15, 5.16 に示す。

いずれも、広告主のドメインのサイトにアクセスしたプロファイルは、そのドメインからの広告が入札されていた。一方で、広告主のドメインのサイトにアクセスしなかったプロファイルでは、そのド

¹¹ <https://jp.maxmara.com/> ¹² <https://www.callawaygolf.jp/>

メインからの入札は確認されなかった。したがって、5.3節で観測された入札の一部はリターゲティング広告による影響であると考えられる。

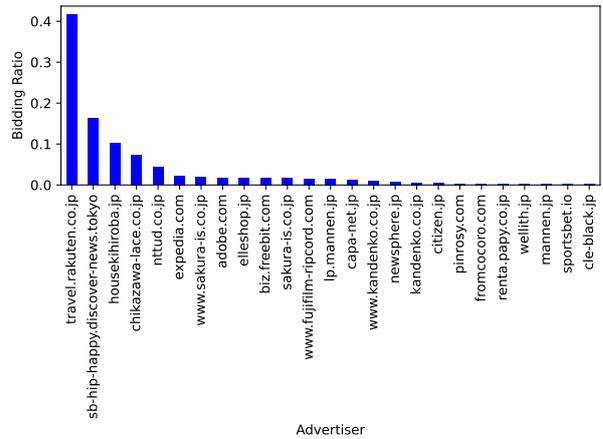
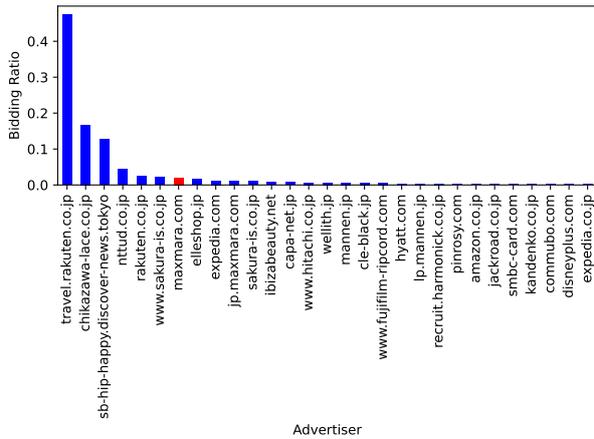


図 5.13: female_high, サイトアクセス有の広告主の分布 図 5.14: female_high, サイトアクセス無の広告主の分布

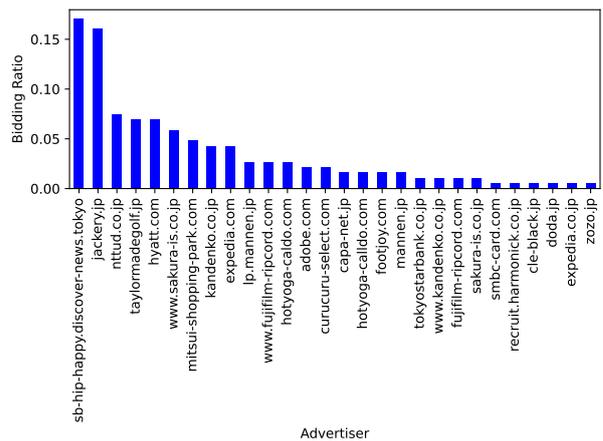
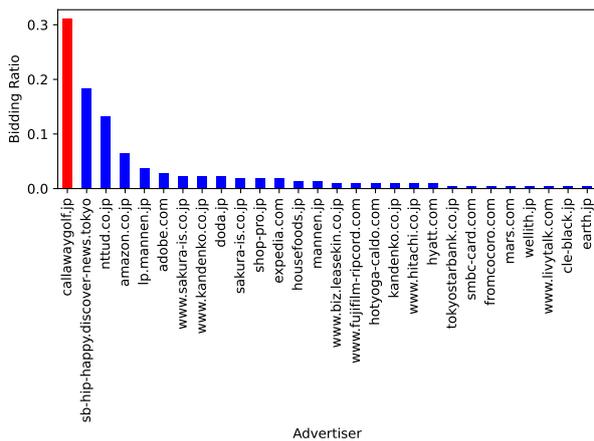


図 5.15: male_medium, サイトアクセス有の広告主の分布 図 5.16: male_medium, サイトアクセス無の広告主の分布

5.6 まとめ

本章では、ユーザの属性や訪問履歴の価値を定量的に示すために、線形重回帰による数理モデルの提案を行った。数理モデルに使用する入札データの精度を高めるために、DS.INSIGHT を利用した検索ワードによるペルソナを作成した。その結果、各属性の平均値とコントロールペルソナの間に有意な差がみられ、精度の高いモデルを提案した。また、リターゲティング広告による入札が数多く観測され、その影響が大きいことを確認した。そのため、より正確にユーザの属性の価値を評価するためには、純粋なユーザのセグメントに対する入札を検証する必要がある。

また、ユーザに対する広告主の関心がどのように変化するかを確かめる目的で、自動観測により入札を長期間高頻度に観測した。その結果、ペルソナ作成から 24 時間程度は急速に関心が薄れ、それ以

降はあまり変化しないことが分かった。ただし、観測 1, 2 の間で 24 時間経過していることに留意する必要がある。また、単純な線形回帰による比較だけでなく、より当てはまりのよいモデルを検討する必要がある。

第6章 結論

本研究はオンライン広告の配信ロジックを明らかにすることで、ユーザが、自身のオンライン上での価値がどのように決定されるかを認識することで、プライバシーの向上へつなぐことへ取り組んだ。

第3章では、異なるペルソナ間では入札額が最大で5.1倍、入札数が16.6倍変化することが示された。また、サイト(パブリッシャー)によって入札額が最大3.7倍、入札数が16.33倍変化することが示された。

第4章では、インプレッション数が少ないと入札額が高騰する可能性が示された。ただし、これを証明するには別のアプローチが求められる。

第5章では、ユーザの属性や広告主サイトへの訪問の価値を定量的に示すための、数理モデルの提案を行った。その中で、ユーザに対する広告の入札は、リターゲティング広告の影響が大きいことを示した。

ただし、第3章および第5章では観測が一回のみであった。広告主のキャンペーンや時期により、入札内容が変化する可能性が考えられる。そのため、再現性を確保するためには、同じ実験を違う日時に複数回行い、その差分を確かめる必要がある。

また、ペルソナの作成時にアクセスしたサイトからのリターゲティング広告が多く観測された。そのため、リターゲティング広告だけでなく、ユーザの趣味趣向にあった他の広告主からの広告を観測するために、リターゲティング広告の影響を極力排除した観測をする必要がある。

広告配信のロジックを明らかにするためには、入札の観測実験だけでなく、オンライン広告オークションにおけるDSPなどのステークホルダーの振る舞いをシミュレーションなどで理論的に示す必要があると考える。

ユーザのプライバシーを守るためには、プライバシーに疎いユーザでもトラッキングからプライバシーを保護できる仕組みや、そもそものオンライン広告のターゲティングの仕組みを見直す必要があると考える。

参考文献

- [1] 株式会社電通グループ, 2023, 電通グループ、「世界の広告費成長率予測（2023～2026）」を公表, (2024年3月取得, <https://www.group.dentsu.com/jp/news/release/001091.html>).
- [2] StatCounter, Browser Market Share Worldwide, (2025年1月取得, <https://gs.statcounter.com/>).
- [3] 若江雅子, 森亮二, 吉井英樹, “オンライン広告におけるトラッキングの現状とその法的考察—ビッグデータ時代のプライバシー問題にどう対応すべきか”, 総務省 学術雑誌『情報通信政策研究』, 第2巻第2号, pp. 1-21, 2019.
- [4] Balachander Krishnamurthy, Craig Wills, “Privacy diffusion on the web: a longitudinal perspective”, Proceedings of the 18th international conference on World wide web, pp.541-550, 2009.
- [5] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, Giovanni Vigna, “Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting”, 2013 IEEE Symposium on Security and Privacy, pp. 541-555, 2013.
- [6] Google LLC, A new path for Privacy Sandbox on the web(2025年1月取得, <https://privacysandbox.com/news/privacy-sandbox-update/>).
- [7] Forbes, Google’s Fingerprinting Returns In 8 Weeks And It Will Track Your Devices, (2025年1月取得, <https://www.forbes.com/sites/zakdoffman/2024/12/21/forget-chrome-google-will-start-tracking-you-and-all-your-smart-devices-in-8-weeks/>).
- [8] World Wide Web Consortium, Third-party cookies have got to go, (2025年1月取得, <https://www.w3.org/blog/2024/third-party-cookies-have-got-to-go/>).
- [9] 英国データ保護機関 (ICO), Our response to Google’s policy change on fingerprinting, (2025年1月取得, <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2024/12/our-response-to-google-s-policy-change-on-fingerprinting/>).
- [10] 株式会社リクルート, 『リクナビ DMP フォロー』の問題点と再発防止策について (2025年1月取得, <https://www.recruit.co.jp/r-dmpf/>).
- [11] 日本経済新聞, リクナビ問題、国の調査終了、経済団体に要望書、「情報利用、就活生に説明を」., 2019-12-14, 朝刊, 38p.

- [12] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. “Smart, useful, scary, creepy: Perceptions of online behavioral advertising”, Proceedings of the Eighth Symposium on Usable Privacy and Security, pp. 1-15, 2012.
- [13] John Cook, Rishab Nithyanand, and Zubair Shafiq, “Inferring Tracker-Advertiser Relationships in the Online Advertising Ecosystem using Header Bidding” , Proceedings on Privacy Enhancing Technologies, pp.65-82, 2020.
- [14] Eric Zeng, Rachel McAmis, Tadayoshi Kohno, Franziska Roesner, “What Factors Affect Targeting and Bids in Online Advertising? A Field Measurement Study” , Proceedings of the 22nd ACM Internet Measurement Conference, pp. 210-229, 2022.
- [15] Kikuchi Hiroaki, and Ayaka Aoyama. “Targeted Ads Analysis: What are The most Targeted Personas?” , 2023 IEEE International Conference on Big Data (BigData), pp. 5512-5518, 2023.
- [16] Prebid, What is Prebid.js?, (2024年5月取得, <https://docs.prebid.org/prebid/prebidjs.html>).
- [17] Steven Englehardt and Arvind Narayanan, “Online tracking: A 1-million-site measurement and analysis” , Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp.1388-1401, 2016.
- [18] Krumnow, Benjamin, Hugo Jonker, and Stefan Karsch. “How gullible are web measurement tools? a case study analysing and strengthening OpenWPM’s reliability” , Proceedings of the 18th International Conference on emerging Networking EXperiments and Technologies, pp. 171-186, 2022.
- [19] 総務省 情報通信政策研究所, 令和5年度情報通信メディアの利用時間と情報行動に関する調査報告書, pp.16, 2024.
- [20] LINE ヤフー株式会社, ヤフー・データソリューション - 行動ビッグデータの活用サービス, (2025年1月取得, <https://ds.yahoo.co.jp/>).
- [21] Krumnow, Benjamin, Hugo Jonker, and Stefan Karsch, OpenWPM_{hide}[21], (2024年5月取得 https://github.com/bkrumnow/OpenWPM/tree/stealth_extension).
- [22] 柴山 りな, 草野 蘭之介, 菊池 浩明, “アドネットワークにおける広告効果指標の調査” , マルチメディア, 分散, 協調とモバイル (DICOMO2021) シンポジウム, 2021.
- [23] 広瀬信輔, アドテクノロジーの教科書, 翔泳社, pp. 2-59, 2019.
- [24] 山田剛史, 杉澤武俊, 村井潤一郎, Rによるやさしい統計学, オーム社, pp. 109-158, 2014.

謝辞

本論文は筆者が明治大学先端数理科学研究科先端メディアサイエンス専攻博士前期課程に在学中の研究成果をまとめたものである。本研究を遂行するにあたり多くの方々から多大なるご指導とご援助を賜りました。

特に、明治大学総合数理学部先端メディアサイエンス学科の菊池浩明教授には、研究のディスカッションに大変多くの時間を割いていただき、また的確な助言をしていただいたことで、本論文を完成に導くことができました。深く感謝申し上げます。

さらに、菊池研究室の皆様には、研究の議論や雑談に付き合っていていただき、研究活動の精神的な支えとなりました。深く感謝申し上げます。

最後に、大学4年間だけでなく、博士前期課程まで進学する機会を与えてくださった家族に深く感謝申し上げます。

研究業績

国内研究会

- (1) 田口勇翔, 菊池浩明, “Header Bidding 情報を用いたペルソナに応じたターゲティング広告の調査”, マルチメディア, 分散, 協調とモバイル (DICOMO2024) シンポジウム, pp.642-648, 2024.

付録A 実験結果4.3.3 各広告枠に対する入札の入札額の分布

パブリッシャーのウェブサイトでは、広告を表示するための広告枠が複数設定されている。ここでは、4.3.3節の実験結果を広告枠ごとに分類し示す。

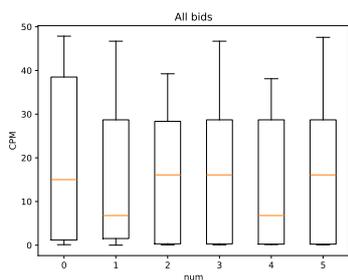


図 A.1: 条件1 広告枠 1

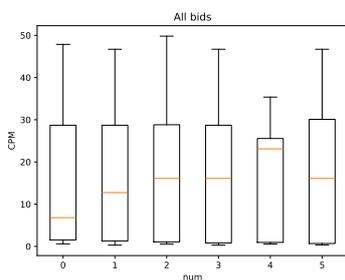


図 A.2: 条件1 広告枠 2

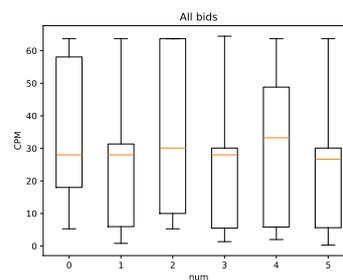


図 A.3: 条件1 広告枠 3

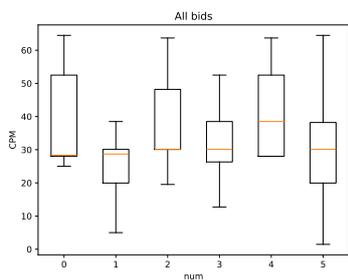


図 A.4: 条件1 広告枠 4

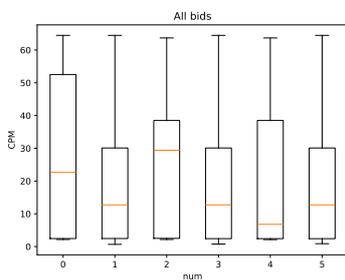


図 A.5: 条件1 広告枠 5

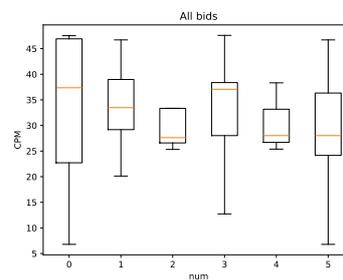


図 A.6: 条件1 広告枠 6

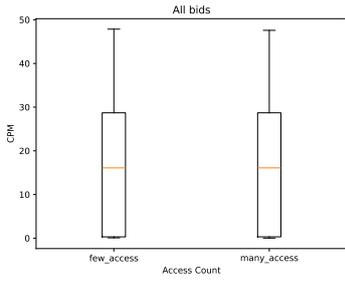


図 A.7: 条件 1 広告枠 1 集計

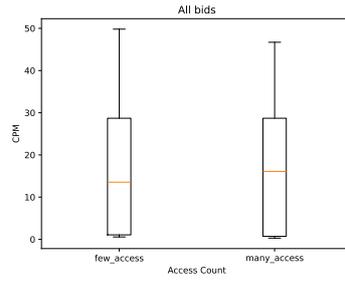


図 A.8: 条件 1 広告枠 2 集計

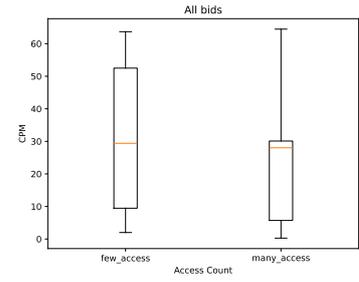


図 A.9: 条件 1 広告枠 3 集計

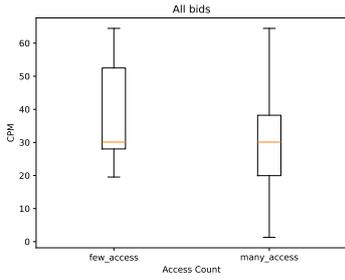


図 A.10: 条件 1 広告枠 4 集計

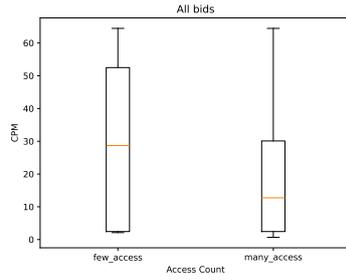


図 A.11: 条件 1 広告枠 5 集計

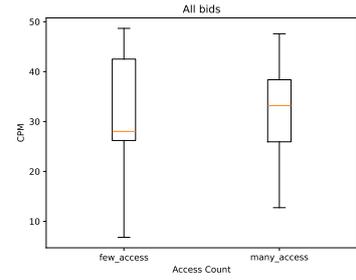


図 A.12: 条件 1 広告枠 6 集計

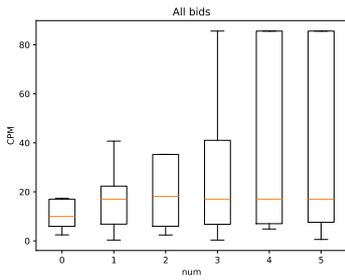


図 A.13: 条件 2 広告枠 1

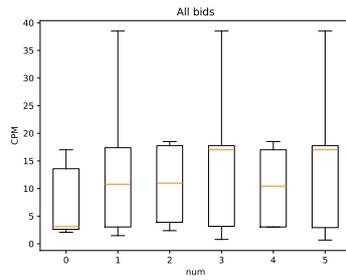


図 A.14: 条件 2 広告枠 2

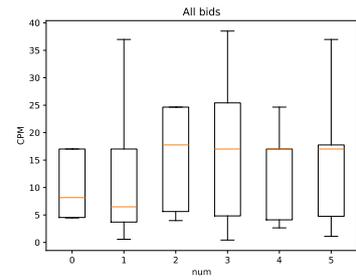


図 A.15: 条件 2 広告枠 3

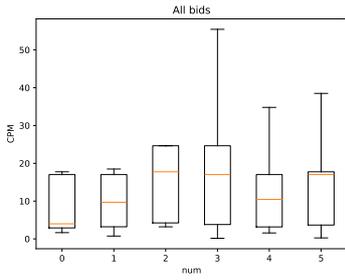


図 A.16: 条件 2 広告枠 4

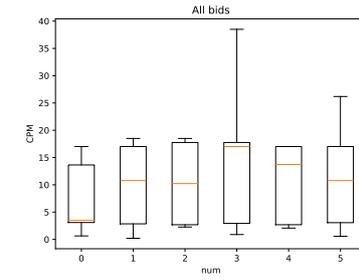


図 A.17: 条件 2 広告枠 5

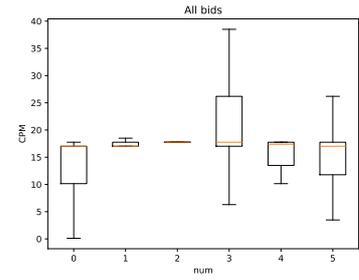


図 A.18: 条件 2 広告枠 6

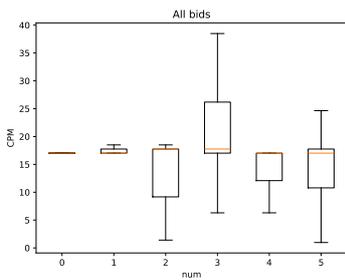


図 A.19: 条件 2 広告枠 7

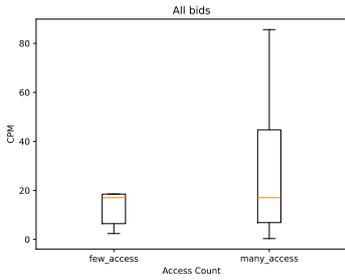


図 A.20: 条件 2 広告枠 1 集計

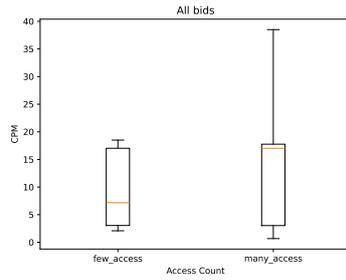


図 A.21: 条件 2 広告枠 2 集計

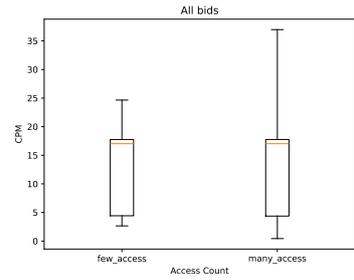


図 A.22: 条件 2 広告枠 3 集計

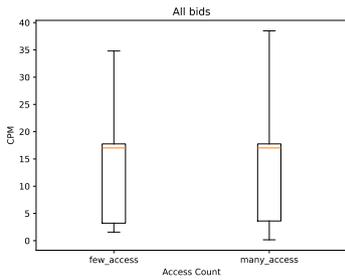


図 A.23: 条件 2 広告枠 4 集計

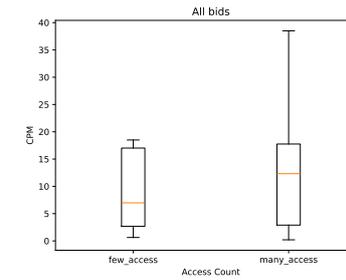


図 A.24: 条件 2 広告枠 5 集計

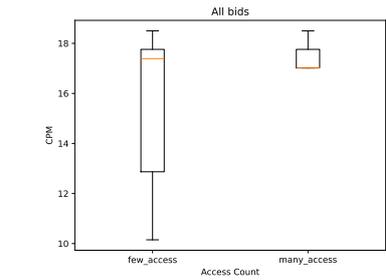


図 A.25: 条件 2 広告枠 6 集計

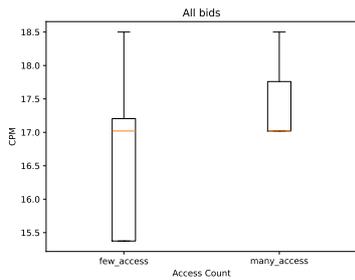


図 A.26: 条件 2 広告枠 7 集計

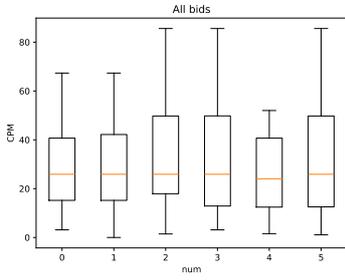


图 A.27: 条件 3 广告 1

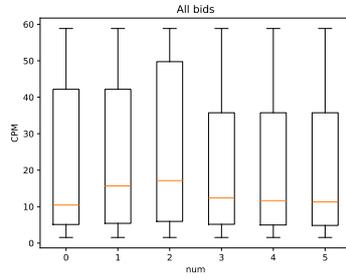


图 A.28: 条件 3 广告 2

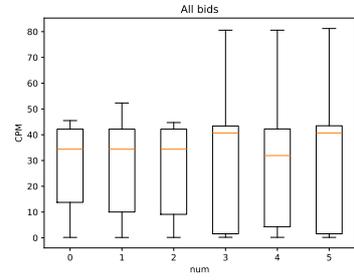


图 A.29: 条件 3 广告 3

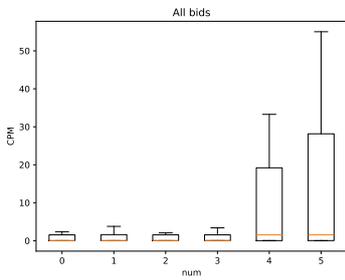


图 A.30: 条件 3 广告 4

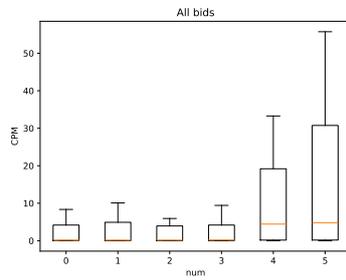


图 A.31: 条件 3 广告 5

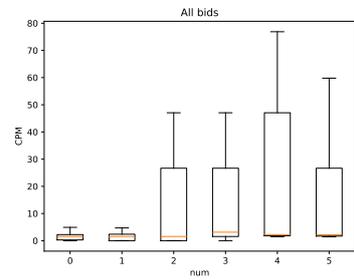


图 A.32: 条件 3 广告 6

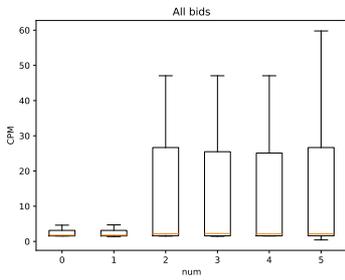


图 A.33: 条件 3 广告 7

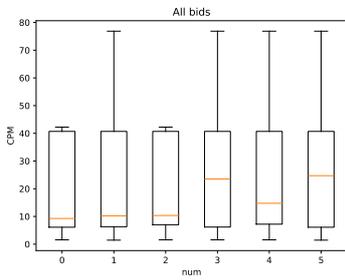


图 A.34: 条件 3 广告 8

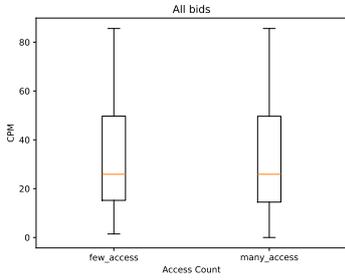


図 A.35: 条件 3 広告枠 1 集計

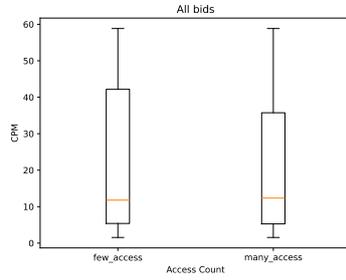


図 A.36: 条件 3 広告枠 2 集計

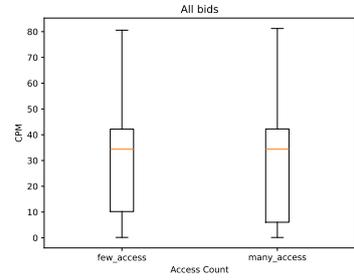


図 A.37: 条件 3 広告枠 3 集計

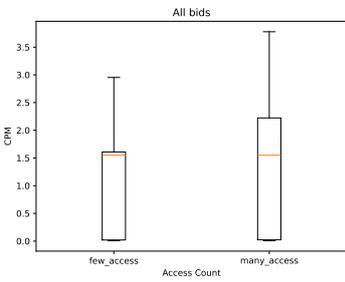


図 A.38: 条件 3 広告枠 4 集計

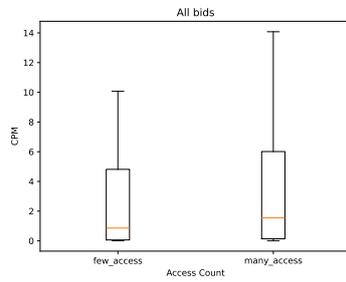


図 A.39: 条件 3 広告枠 5 集計

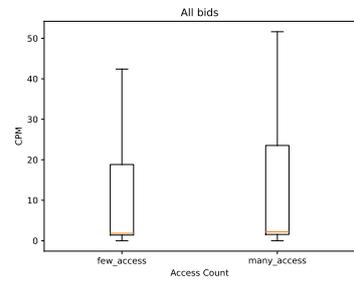


図 A.40: 条件 3 広告枠 6 集計

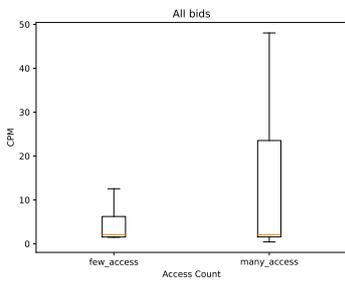


図 A.41: 条件 3 広告枠 7 集計

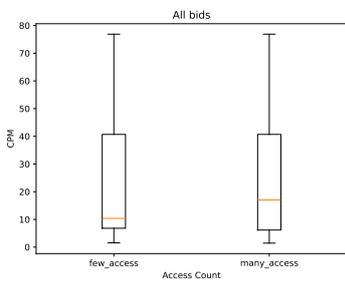


図 A.42: 条件 3 広告枠 8 集計

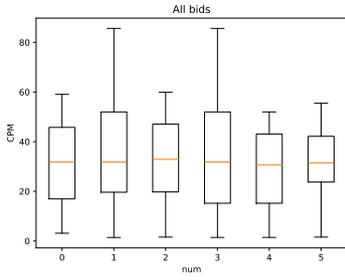


图 A.43: 条件 4 广告 1

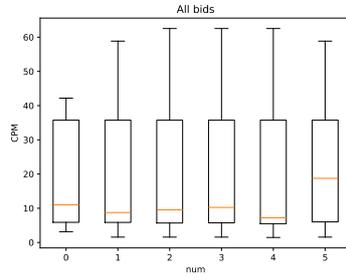


图 A.44: 条件 4 广告 2

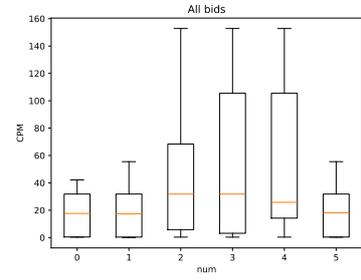


图 A.45: 条件 4 广告 3

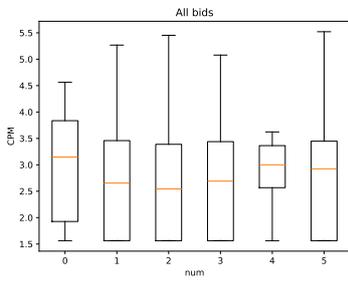


图 A.46: 条件 4 广告 4

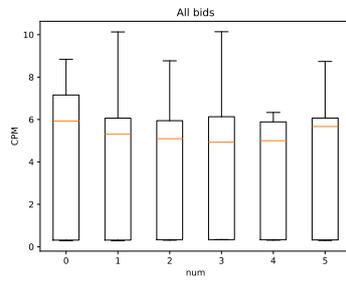


图 A.47: 条件 4 广告 5

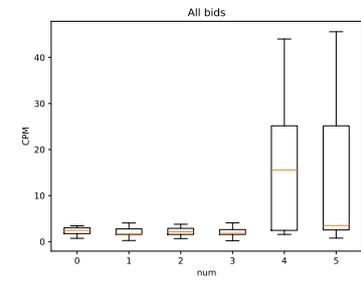


图 A.48: 条件 4 广告 6

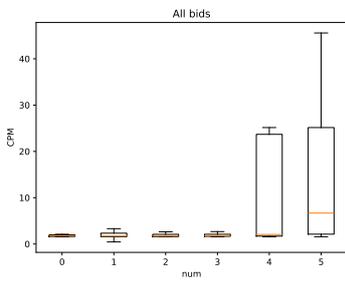


图 A.49: 条件 4 广告 7

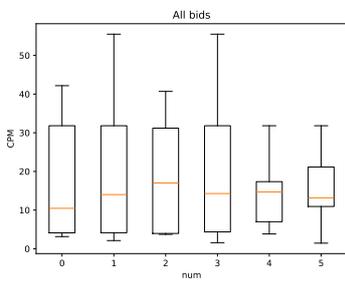


图 A.50: 条件 4 广告 8

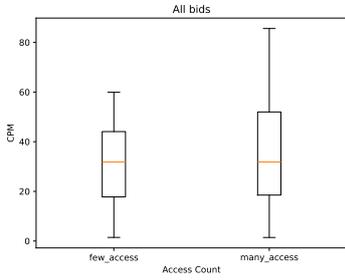


図 A.51: 条件 4 広告枠 1 集計

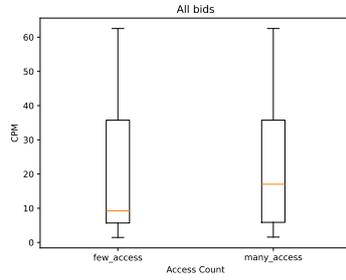


図 A.52: 条件 4 広告枠 2 集計

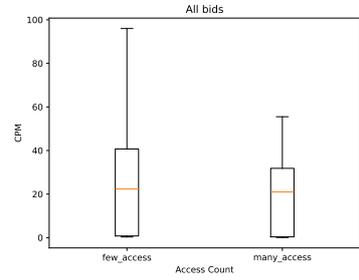


図 A.53: 条件 4 広告枠 3 集計

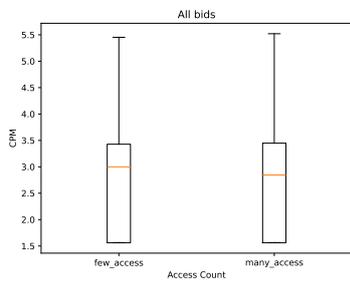


図 A.54: 条件 4 広告枠 4 集計

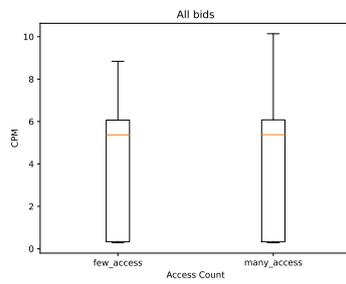


図 A.55: 条件 4 広告枠 5 集計

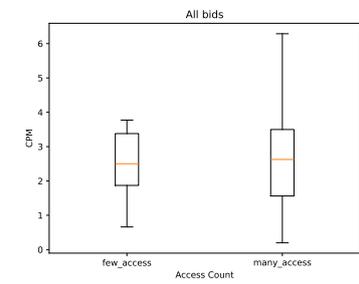


図 A.56: 条件 4 広告枠 6 集計

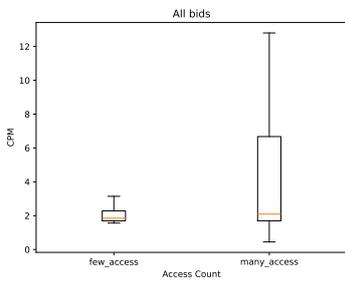


図 A.57: 条件 4 広告枠 7 集計

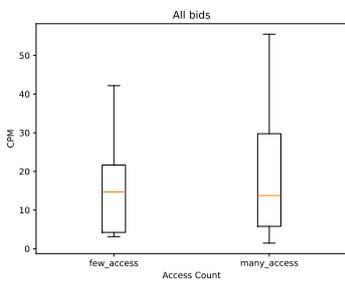


図 A.58: 条件 4 広告枠 8 集計