明治大学大学院　先端数理科学研究科

2022 年度

修士学位請求論文

Header Bidding を用いたオンラインターゲティング広告の観測に関する研究

学位請求者　先端メディアサイエンス専攻

柴山りな

あらまし

ターゲティング広告とは，ユーザー属性や Web サイトの閲覧履歴などの条件を指定することで，ユーザーに適した内容を掲載する広告手法である．ターゲティング広告により，広告主は効率的にコンバージョン（購入・資料請求など）につながる広告を掲載できるというメリットがある． 一方，我々ユーザは自分のウェブサイト閲覧履歴がどの程度取得され，広告配信に利用されているかを知らない．

そこで本研究では，HB が使用されているサイトの割合や特徴を明らかにすることおよび，ヘッダービッディングの入札価格に影響を与える要因を明らかにすることを目的として調査を行う．自動的に観測するためのプログラムを Python により開発し，Header Bidding の入札情報を取得する Javascript を実行することにより実験を行い，結果を分析する．

目 次

第1章　序論

1.1　背景

ターゲティング広告とは，ユーザー属性や Web サイトの閲覧履歴などの条件を指定することで，ユーザーに適した内容を掲載する広告手法である．ターゲティング広告により，広告主は効率的にコンバージョン（購入・資料請求など）につながる広告を掲載できるというメリットがある．一方，我々ユーザは自分のウェブサイト閲覧履歴がどの程度取得され，広告配信に利用されているかを知らない．Olejnik ら [6] は，異なる条件（ユーザ属性，閲覧履歴，地域等）のユーザに描画された広告の入札額および落札額を比較することで，それぞれの情報が広告配信に利用されているかどうかを調査している．Cook ら [2] は，Header Bidding に着目し，広告の入札額と利用者のプロファイルとの関係を明らかにしている．

リアルタイムビッディング (RTB) とは，オンライン広告で最も広く使われているシステムである．ヘッダービッディングは，新しい RTB システムのひとつであり，従来の RTB と比較してパブリッシャーの収量を増加が見込めるため急速に人気を集めている．また，透明性の観点では，RTB は落札者とその入札価格が公開されるのに対し，HB で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などが HTTP リクエストヘッダーで公開される．RTB や HB を利用して，ターゲティング広告の傾向や業者間でのデータ共有関係を明らかにするための研究が行われている [5, 4, 3]．早朝の時間帯は，落札価格が高くなる傾向や，健康に興味のあるユーザは入札額が高い傾向があることなどが分かっている．

1.2　ユーザのプライバシーに対する懸念

広告から検索結果まで，インターネットを閲覧しているユーザーが目にするものの多くは，そのユーザーについて推論したアルゴリズムによってターゲット化またはパーソナライズされている．
ユーザーがインターネットを閲覧すると，閲覧しているウェブサイトだけでなく，HTTP クッキーからブラウザのフィンガープリントに至るまで様々な技術を使用して，第三者の広告会社や分析会社によって，ユーザーのオンライン活動が追跡される [18]．これらの閲覧行動のログを利用して，ユーザーの興味，嗜好，および属性を推測する [19].
Blase[20] らによると，ユーザーはこのようなターゲティングを便利であると同時に，見えないところで何が行われているのかが明らかになっていないことに恐怖やて気味の悪さを感じていることが報告されている．
オンラインターゲティング広告で効果的な広告掲載基準について，Avi ら [24] は，サイトのコンテンツに即した広告が効果的であると報告している．また，明らかに目立つ位置に広告を掲載することで

効果が上がるとしている．しかしながら，広告をウェブサイトのコンテンツにマッチさせることと，広告の目立たせることの両方を行ってしまうと，どちらか一方のみの広告よりも購買意欲を高める効果が低い．これは，プライバシーを懸念するユーザによる行動によるとの考察がなされている．

1.3 データトラッキング

よりユーザの嗜好や属性に即した広告を配信するため，広告配信業者はCookieを利用することが一般的であった．しかしSafariとFirefoxのような主流のブラウザが厳しいサードパーティのクッキーポリシーを実装し始めたり，ヨーロッパでGDPRが施行されたりとユーザのプライバシー保護の観点からCookieの第３者による取得が制限され始めた．現在ではCookieに依存しないトラッキング形態として，ファーストパーティID リーク，IDシンクロ，ブラウザフィンガープリントなどが利用されている [21, 22] .

それに対して，業者によるユーザのデータの取得を明らかにする手法として，クライアントサイドでクッキーの同期を検出することが挙げられます．しかしこの方法は，さまざまな経験や環境に依存することや，クッキーの同期などのクライアントサイドの分析では，サーバサイドのデータ共有は検出できないなどの欠点がある [23].

広告枠に対するリアルタイムオークションを利用して業者によるデータ取得を明らかにする手法がある．この手法では，複数のユーザや条件下での広告枠の落札額，入札額を取得し，それらを比較することで広告業者のトラッキングの有無を調査する．つまり任意のウェブページを閲覧し，閲覧をしていないユーザと比べて入札額が高かったとすると，ユーザの閲覧履歴はなんらかの業者により取得されているということができる．逆に言うとユーザの閲覧履歴情報を持っている入札者は，その情報に応じて入札額を決定することが前提となる．

1.4 RTB と HB

リアルタイムビッディング (RTB) とは，オンライン広告で最も広く使われているシステムである．一つの広告枠に対して，リアルタイムで広告業者が入札を行い，一番高い値段で入札した者が落札者となり広告を表示できるという仕組みである．

RTBを利用して，ターゲティング広告の傾向や業者間でのデータ共有関係を明らかにするための研究が行われている [5, 4, 3]．特にOlejnikら [6] は，異なる条件（ユーザ属性，閲覧履歴，地域等）のユーザに描画された広告の入札額および落札額を比較することで，それぞれの情報が広告配信に利用されているかどうかを調査している．結果，早朝の時間帯は落札価格が高くなる傾向傾向があることなどが分かっている．

ヘッダービッディング (HB) は，新しいRTBシステムのひとつであり，従来のRTBと比較してパブリッシャーの収量を増加が見込めるため急速に人気を集めている．また，透明性の観点では，RTBは落札者とその入札価格が公開されるのに対し，HBで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などがHTTPリクエストヘッダーで公開される．Cookら [2] は，Header Biddingに着目し，広告の入札額とユーザの属性との関係を明らかにしている．

1.5 本研究の目的

オンラインターゲティング広告は，効率的に消費者に商品をアピールできるため人気である一方，社会的な問題も存在する．それは主に広告を見るユーザのプライバシーに関する問題と，広告の枠組みの不透明さである．

オンラインターゲティング広告は，広告主のサービスに興味のあるユーザに効果的に広告を表示するため，ブラウザ閲覧履歴，地域，時間，広告が表示されているサイトなどの情報が利用されている．ユーザのプライバシーに関わりうる情報が，業者によって取得されることがあってはならない．しかし閲覧者は，Cookie などの情報がどの程度取得され，広告出稿を判断するデータとして利用されているのかを知らない．また，仮に取得されている情報がユーザの重要なプライバシーに関わるものでないとしても，ユーザが情報を取得されていること自体を知らないことは社会的な問題である．よって本研究では，どんな情報を取られているか明らかにすることが，ユーザのプライバシーを守るための第一歩につながるのではないかと考えた．そこで閲覧履歴やその時間に応じて，ユーザの広告枠の価値がどのように変動するかを調査する．

また，オンライン広告には，表示される広告がどのように決定されるのかが不透明であるという問題点もある．いつ，どこに，どのようなメカニズムで広告が表示されるのかを広告主は知らない．この不透明性は広告配信業者による広告不正に繋がる恐れがある．広告配信のメカニズムの一部を明らかにすることは，広告不正を防止することに繋がるのではないかと考えられる．

本研究の目的は，オンラインターゲティング広告でユーザのプライバシーに関わる情報がどれほど利用されているのかを明らかにすることで，ユーザのプライバシーを守ることと，そして表示される広告が決定されるデータの要素を少しでも明らかにすることで広告の枠組みの不透明さの解消へと繋げることである．

1.6 本研究の貢献

本論文では次のことを明らかにする．

- Header Bidding を使用するウェブサイトの概要
- Header Bidding で行われる広告枠の入札の入札額に影響を与える要因

1.7 本論文の位置づけ

Olejnik ら [6] は，ユーザ属性，閲覧履歴，地域等などの条件が違う場合の入札額および落札額を比較することで，それぞれの要素が広告配信に利用されているかどうかを調査している．また Cook ら [2] は，Header Bidding に着目し，広告の入札額とユーザの属性とのより詳しい関係を明らかにしている．しかし，これらの従来研究では個々の条件が入札額に与える影響については明らかにしているものの，どの要素の影響が強いのかは明らかになっていない．また，従来研究では，広告が表示されるパブリッシャーサイトの違いによる入札価格の変化の詳細は明らかになっていない．そこで本研究

表 1.1: ステークホルダーの役割とサービス例

	ステークホルダー	役割	例
1	入札者	広告在庫を持ち広告枠に入札する	Yahoo SSP, ix, Yandex
2	媒体サイト，パブリッシャー	ウェブサイトやアプリを運営し広告枠を提供する	nifty.com, kakaku.com
3	広告主サイト	入札者に広告の出稿を依頼する	G-star.com, baitoru.com
4	アドエクスチェンジ	入札者を束ねるサービス	DoubleClick
5	Header Bidding	アドエクスチェンジや入札者を束ねるサービス	Prebid.js
6	閲覧者	ウェブサイトやアプリを訪れ，広告を閲覧する	-

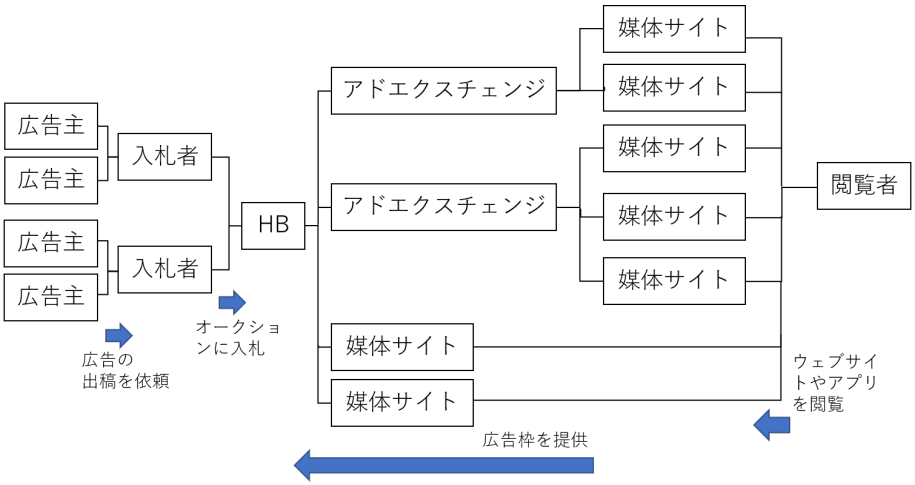

図 1.1: ステークホルダーの関係

では，入札額に影響を与える要素として，ユーザ属性，閲覧履歴，時間に加えて，パブリッシャーサイトの特徴を調査し，これらの要素の中でどの要素が最も入札価格に影響を与えるのかを明らかにすることを試みる．

1.8 ステークホルダーの定義と関係

ヘッダービッディングを使用するオンライン広告には，5 つの重要なステークホルダーが存在する．それぞれのステークホルダーの役割は表 1.1 に示すとおりである．また，ステークホルダー間の関係を図 1.1 に示す．

1.9 論文構成

本論文は 4 章で構成される．第 1 章では，まず本研究の背景と目的を述べ，次に本研究の概要を示し，その着想と貢献を述べる．第 2 章では，関連研究の概要を述べ，本研究の位置づけを明らかにする．第 3 章では，ヘッダービッディングでの入札状況を調査するにあたって，まずは HB が使用されているサイトの割合や特徴を，実験により調査する．第 4 章では，広告の入札額が，表示されるブラウザの閲覧履歴や地域，時間，広告が表示されているサイトなどの要素にどの程度の影響を受けているのかを，実験により調査し結果を分析する．

第2章　従来研究と本研究の位置づけ

2.1　Olejnik ら [6]

Olejnik ら [6] は，ユーザのプライバシー情報の価値を分析するため，閲覧履歴，時刻，位置情報，ユーザのプロファイルが落札価格に与える影響を調査している．実験の方法としては，同僚や友達に RTB での落札価格を取得する Firefox プラグインを配布して 1 時間に 1 回の間隔で 1 か月の間計測を行った．そして記録が 70 サイト以下だったものを除いた 100 人のデータについて分析を行った．
その結果，0〜8 時の間が一日の中で入札額が有意に高いとしている．また，広告配信地域に関しても，アメリカ，フランス，日本の順で高く有意差があるという結果であった．また，ユーザの意図についても調査を行っている．通常の商品とは関係のないウェブページを閲覧した場合と比べて，特定の商品やサービスを閲覧した場合は，落札価格が高い傾向がある．

2.2　Cook ら [2]

Cook ら [2] は，複数の興味対象の違うペルソナ間での HB での入札価格の違いを調べることで，広告業者間でのユーザのプライバシー情報の共有関係を明らかにしている．OpenWPM というオープンソースを用いて，16 のペルソナを作成し，それぞれ 50 ページの興味対象に関するページを訪問し，その後，HB を利用するサイトで入札価格を取得するという手順でデータの取得を行った．これを各ペルソナに対し 10 回ずつ行った．
結果，健康ペルソナは人気だが，それ以外のペルソナは入札者によって人気・不人気があったこと，興味履歴がないペルソナよりも低いペルソナもあったこと，購買意欲があるユーザの広告枠は，総じて人気であること，Rubicon という入札業者は他の入札者よりも総じて入札額が高いことなどが分かった．彼らは，あるサイトをブロックした際に，ブロックしていないユーザと比べて，価格が大幅に下がった場合，データの共有が行われているという前提のもと実験を行った．結果，いくつかのデータ共有関係が明らかになった．

2.3　本研究の位置づけ

しかし，これらの従来研究では個々の条件が入札額に与える影響については明らかにしているものの，どの要素の影響が強いのかは明らかになっていない．また，従来研究では，広告が表示されるパブリッシャーサイトの違いによる入札価格の変化の詳細は明らかになっていない．そこで本研究では，入札額に影響を与える要素として，ユーザ属性，閲覧履歴，時間に加えて，パブリッシャーサイトの

表 2.1: 従来手法との比較

		本研究	従来研究 A[2]	従来研究 B[6]
データの収集方法		自動プログラム	自動プログラム	ユーザ実験
対象システム		HB	HB	RTB
媒体サイト調査数		100	25	70 以上
調査した要因	(閲覧履歴)	○	○	○
	(位置情報)	×	×	×
	(時刻)	○	×	○
	(経過時間)	○	×	×
	(媒体サイトの情報)	○	×	×

特徴を調査し，これらの要素の中でどの要素が最も入札価格に影響を与えるのかを明らかにすることを試みる．先行研究との比較を表 2.1 に示す．

第3章　ヘッダービッディングを使用するウェブサイトの調査

3.1　目的

ヘッダービッディングでの入札状況を調査するにあたって，まずはHBが使用されているサイトの割合や特徴を明らかにすることを目的とする．

3.2　方法

2023年1月にAhrefs Rank[1]の上位16,000のウェブサイトに対して，調査を行う．Ahrefsとは，SEOやウェブマーケティングに使用するツールを提供するサービスであり，ランキングなどのデータの集計対象としてGoogleだけでなくYouTube，Amazon，Bing，Yahoo，Yandexなど様々な検索エンジンを総合的に取得している有料サービスである．Ahrefsランクの上位16,000サイトは，2023年1月時点でのランキングを使用する．調査した16,000のウェブサイトのうち上位20件を表3.1に示す．16,000のサイトに対して，次の3つの項目を調査する．

- **ヘッダービッディングを使用しているかどうか** Header Biddingを提供するサービスはいくつかあるが，本調査では最も使用されている`Prebid.js`を対象とする．`Prebid.js`を呼び出すJavascriptのコマンド`pbjs.getBidResponses()`を実行してオブジェクトが返されたページをヘッダービッディングを使用するサイトと判断する．また，URLが存在しない，読み込めないなどのエラーが発生した場合はエラーとして分類する．
- **サイトの使用言語** サイトのHTMLに含まれるタイトル，見出し，Metadescriptionを取得し，Pythonの言語判定ライブラリ`translate`で判定を行う．
- **サイトで使用されているキーワード** HTMLから取得したタイトル，見出し，Metadescriptionの中に特定のキーワードが存在するかどうかで，サイトのカテゴリーを判定する．比較のため，サイトのカテゴリーの判定には英語のサイトのみを使用した．キーワードとしては，先行研究[15]で使用されていたAlexa Top Sitesのカテゴリーの一部である`business`，`news`，`tool`，`product`，`health`，`arts`，`sports`，`science`，`games`，`communication`，`kids`，`shopping`を使用した．

これらの調査項目を，Pythonの Seleniumモジュールを使用して，16,000サイトを自動で判定する．ブラウザとしてはChromeを使用した．

自動プログラムを実行するためにSeleniumで主に３つの設定を行った．まず，閲覧履歴が少ないとHBが行われない可能性があるため，履歴の多い通常のユーザでログインを行い実験を行った．ログ

表 3.1: Ahrefs Rank Top 20 サイト

Ahrefs Rank	URL
1	facebook.com
2	instagram.com
3	twitter.com
4	youtube.com
5	linkedin.com
6	google.com
7	wordpress.org
8	pinterest.com
9	plus.google.com
10	whatsapp.com
11	miit.gov.cn
12	apple.com
13	goo.gl
14	qq.com
15	policies.google.com
16	youtu.be
17	microsoft.com
18	maps.google.com
19	play.google.com
20	wa.me

インは Selenium のオプションでユーザプロファイルの存在するパスを指定することで実現した．また，自動プログラムがセキュリティのせいでブラウザが停止してしまうのを防ぐために，いくつかのセキュリティを緩くするための Selenium オプションを設定した．最後に，HB のオークション情報を取得するのに Javascript が必要であるため Javascript を有効化するオプションを設定した．python のプログラムで利用した Selenium のオプションを表 3.3 に示す．

16000 のサイトに対して自動プログラムにより以下の手順を行う．

1. サイトにアクセスする．
2. ランダムで 2〜7 秒待機する．
3. 10000 ピクセルのスクロールを行う．
4. Javascript で `pbjs.getBidResponses()` を実行し，何らかのオブジェクトが返って来たら HB を利用するサイトだと判断する．エラーの場合は，HB を使用しないサイトだと判断できる．
5. Python の request メソッドで HTML を取得し，サイトの使用言語の判定とサイトで使用されているキーワードの判別を行う．

表 3.2: Selenium Chromedriver で指定したオプション一覧

変数	機能	目的
profile-directory	ユーザを指定	プラットフォーム (Google) にログインし，プロファイルを指定する．
user-data-dir	ユーザのフォルダを指定	
no-sandbox	sandbox モードを解除する	セキュリティを緩くし，クラッシュを回避する
disable-setuid-sandbox	同上	
disable-gpu	GPU ハードウェアアクセラレーションを無効にする	
enable-javascript	Javascript を有効化	Javascript の使用
user-agent	ユーザエージェントを指定	通常のアクセスに似せる
intlàccept_languages	言語を指定	

表 3.3: HB を使用するサイトの割合

	サイト数
HB	1,010
HB でない	13,129
エラー	1,894

3.3 結果と考察

ヘッダービッディングを使用するサイト数の調査結果を表 3.3 に示す．16,000 サイト中，HB を使用するサイトは 1,010，使用しないサイトは 13,129 であった．ただし HB ではないサイトには，そもそも広告枠が存在しないサイトも含まれる．

また，サイトランクごとの HB サイトを使用する割合を図 3.1 に示す．1000 位ごとの HB を使用するサイトの割合に，ランクによる違いはほぼなかった．

16,000 のウェブサイトの言語の内訳を図 3.2 に示す．英語のサイトが 7 割，次いでドイツ語，日本語，フランス語が上位であった．サイトの言語 (上位 40 言語) ごとの HB を使用する割合を図 3.3 に示す．ベトナム語 (n = 36) で特に HB の割合が高かったほか，ポルトガル語 (n = 97) やイタリア語 (n = 127)，スペイン語 (n = 156)，ノルウェー語 (n = 32)，フランス語 (n = 360) のサイトで HB の割合が 10 %を超えていた．日本語に関しては 36/444 サイトで HB が使用されていた．また，デンマーク語 (n = 55) やハンガリー語 (n = 19) など HB が使用されていない言語も存在した．この結果より国によって，HB が使用されているかどうかには差があると言える．

16,000 サイトのうち，英語のサイトは 10,314 サイトであった．英語サイトを対象として，カテゴリーごとの割合を調査するために，取得した文書の中に特定のキーワードが含まれるか判定を行った結果を図 3.4 に示す．スポーツというワードを含むサイトでは，HB を使用するサイトが 5 割を超えて最も多かった．このことから，スポーツに関するコンテンツを提供するサイトが HB を使用する傾向が高いことや，そもそも広告を配信している割合が高いことが分かる．また，コミュニケーション

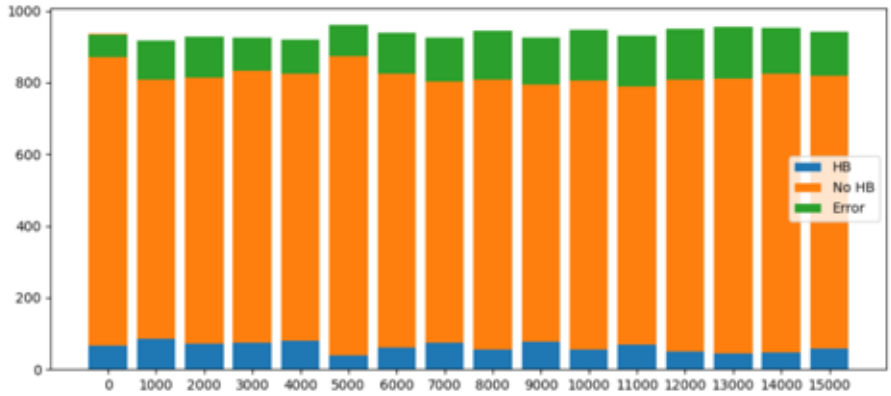

図 3.1: サイトランクごとの HB サイトを使用する割合

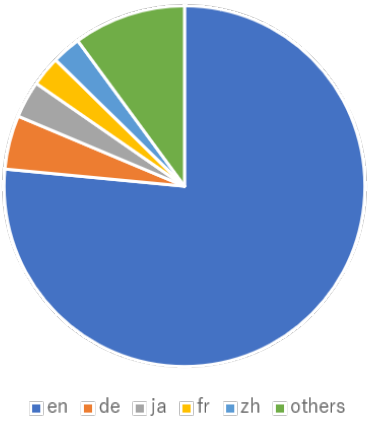

図 3.2: Ahrefs トップサイトでの使用言語の分布

に関するサイトでは，HB を使用する割合が低かった．このことからコミュニケーションに関するサイトでは HB があまり使用されていないことや，そもそも広告が配信されていないと考えられる．

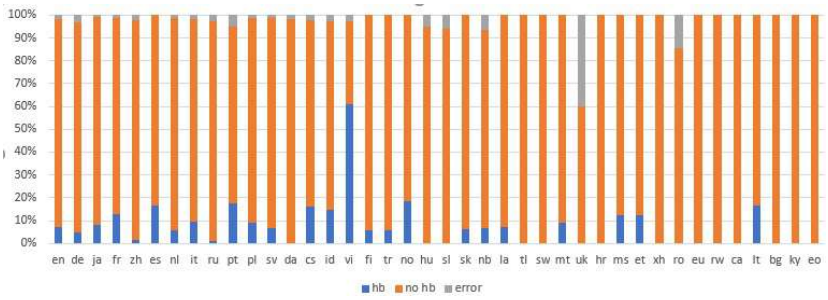

図 3.3: サイトの言語ごとの HB を使用する割合

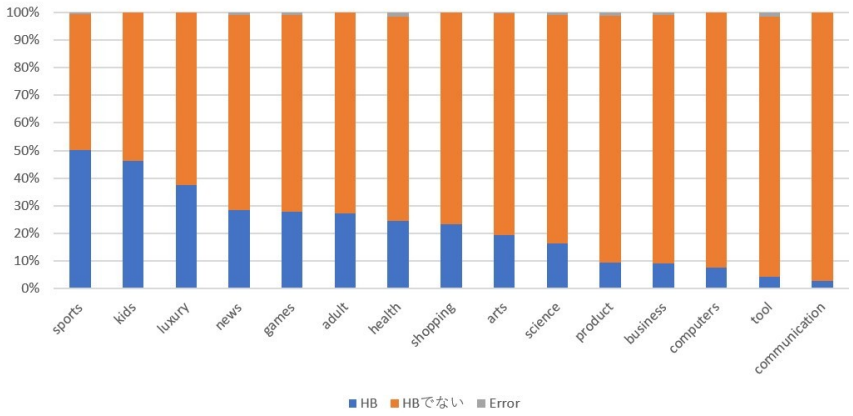

図 3.4: サイトのカテゴリごとの HB を使用する割合

第4章　ヘッダービッディングの入札価格に影響を与える要因の調査

4.1　目的

オンライン広告は，広告が表示されるブラウザの閲覧履歴や地域，時間，広告が表示されているサイトなどにより変動する．しかし我々は広告がこれらの要因にどの程度の影響を受けているのかを正確に知らない．よって本実験では，それぞれの要因が広告枠の入札価格に与える影響を明らかにする．

4.2　方法

3 章で調査した HB が使用されている 1,010 のウェブサイトに対して，時刻による入札価格の違い，閲覧履歴（ユーザ）による入札価格の違い，広告が表示されるサイト（パブリッシャー）による入札価格の違い，商品閲覧後 48 時間の入札額の推移の 4 つの項目についてそれぞれ Python の Selenium を用いた自動プログラムでデータを取得，分析を行った．詳しい方法をそれぞれの節で述べる．

4.2.1　時刻による違い

時刻による入札額の違いを調査するため，HB が利用されているサイトの中でも入札が多いサイトを 9 つ選び，調査対象とした．サイトのリストを表 4.2 に示す．2023 年 1 月に自宅で 48 時間 1 時間に 1 回の間隔で行った．事前準備として，Chrome のブラウザの 10 のアカウントを新設し，アカウントログインはせずに使用する．閲覧履歴が少ない場合，HB でのオークションが行われなかったり入札がなかったりする可能性がある．そこで表 4.1 に示す広告キャンペーンを行っている広告主のウェブページを訪れる．10 のアカウントをグループに分け，3 つの広告主に対してそれぞれ 3 ユーザずつ割り当て，残りの 1 ユーザは履歴のない対照ユーザとして実験を行う．実験は以下の手順で行う．

1. 閲覧履歴作成のため，広告主のサイトのトップページに存在するリンクをランダムで 15 ページ閲覧する．
2. パブリッシャーサイトにアクセスする．
3. ランダムで 2〜7 秒待機する．
4. 10000 ピクセルのスクロールを行う．
5. Javascript で `pbjs.getBidResponses()` を実行し，HB のオークション情報を json ファイルでローカル環境に保存する．

表 4.1: HB 観測対象の商品サイト

サービス名	サービスカテゴリ	サイト URL
G-ster	衣服	https://www.g-star.com/ja_jp
バイトル	求人	https://www.baitoru.com/
スカイスキャナー	航空券	https://www.skyscanner.jp/

表 4.2: 対象パブリッシャーサイト

	URL
1	thetimes.co.uk
2	timeout.com
3	thesun.co.uk
4	kitco.com
5	timesunion.com
6	appledaily.com
7	wickedlocal.com
8	kakaku.com
9	speedtest.net

6. Cookie を削除する．

7. 手順 2 6 を 9 つの HB を使用するパブリッシャーサイトに対して 1 回ずつ行う．

8. 手順 7 を 10 のユーザに対して行う．

4.2.2 ユーザによる違い

Python による自動プログラムで大学の PC で 2023 年 1 月に調査を行った．HB が利用されているサイトの中でも入札が多いサイトを 130 サイト選び，調査対象とした．閲覧履歴（ユーザ）による入札額の違いとして，通常のユーザ（多くの閲覧履歴を有する），健康に興味のあるユーザ，技術に興味のあるユーザ，履歴のないユーザの 4 種類を調査する．それぞれの Windows アカウントで，Chrome のユーザを 1 つずつ新設し，ログインせずに使用する．まず，Windows のユーザアカウントを 7 つ新設しログインせずに使用する．以下の手順でデータの取得を行う．

1. HB のオークション情報を取得する前に，履歴の作成を行う．通常のユーザの閲覧履歴は，もとより 2296 ドメイン，41501 ページのアクセスがあった．健康に興味のあるユーザでは，Alexa Top Sites の健康カテゴリのトップ 50 ページを閲覧する．技術に興味のあるユーザに関しても同様である．対照としての履歴のないユーザに関してはここでは何も行わない．

2. 対象のパブリッシャーサイトにアクセスする．

3. ランダムで 2〜7 秒待機する．
4. 10000 ピクセルのスクロールを行う．
5. Javascript で `pbjs.getBidResponses()` を実行し，HB のオークション情報を json ファイルでローカル環境に保存する．
6. Cookie を削除する．
7. 手順 2 5 を 130 の HB を使用するパブリッシャーサイトに対して 1 回ずつ行う．

各ユーザ 3 回ずつ HB が行われているウェブサイトにアクセスし，入札価格の平均値を比較する．

4.2.3 パブリッシャーによる違い

Python による自動プログラムで大学の PC で 2023 年 1 月に調査を行った．HB を使用する 1,010 のサイトのうち英語の 749 サイトを対象とした．ユーザとしては履歴の多い通常のユーザのみを使用した．これらのサイトでの入札価格とサイトの AhrefsRank の関係を分析する．また，入札価格の高い特定のサイトカテゴリがあるかを調査するため，サイトのタイトル，Metadescription，h2，h3，h4 のいずれかに選んだキーワードが 1 回以上出現するかで判断したカテゴリについて入札額の分布の違いを分析する．

4.2.4 閲覧後の推移

閲覧後の推移の調査は，時刻の調査と全く同じ方法で調査を行う．加えて，手動での調査も行う．手動により商品を閲覧した後の入札価格の推移の調査を行った．商品ページを閲覧した後，`kakaku.com` で入札情報の取得を 1 6 時間に 1 回ので 18 時間行う．

4.3 結果

4.3.1 時刻による違い

入札価格を時間で比較した結果を図 4.1 に，外れ値を除外した結果を図 4.2 に示す．14〜18 時にかけて入札額が高く，朝の 4〜9 時にかけて入札額が低い．

4.3.2 ユーザによる違い

ユーザごとの入札額の分布を図 4.3 に示す．履歴の多い通常のユーザでは，高い価格での入札が多かった．また，外れ値を除いたユーザごとの入札額分布を図 4.4 に示す．通常のユーザはそれほど高くなく，健康に興味のあるユーザは入札の中央値が 4 種類のユーザの中で最も高かった．一番入札額が低いのは，閲覧履歴のない対照ユーザであった．

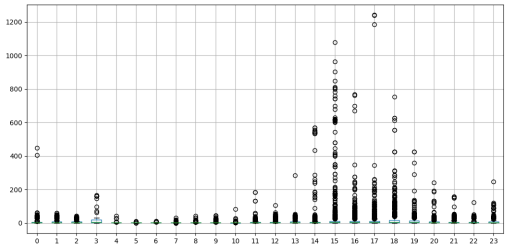

図 4.1: 時刻ごとの入札額分布

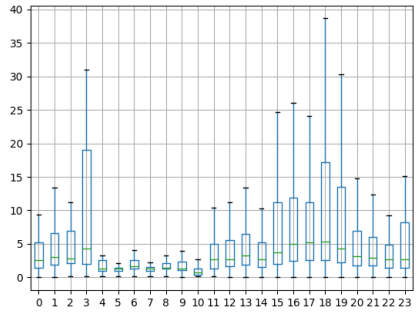

図 4.2: 時刻ごとの入札額分布 (外れ値除く)

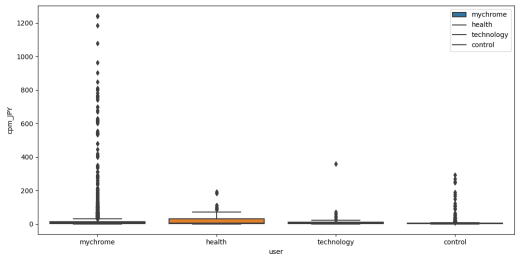

図 4.3: ユーザごとの入札額の分布

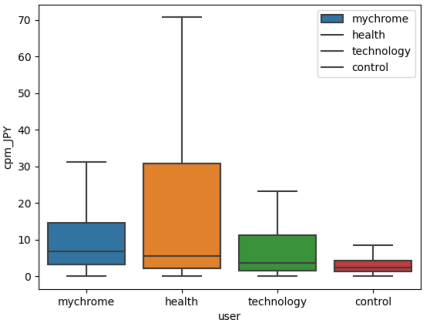

図 4.4: ユーザごとの入札額の分布 (外れ値を除く)

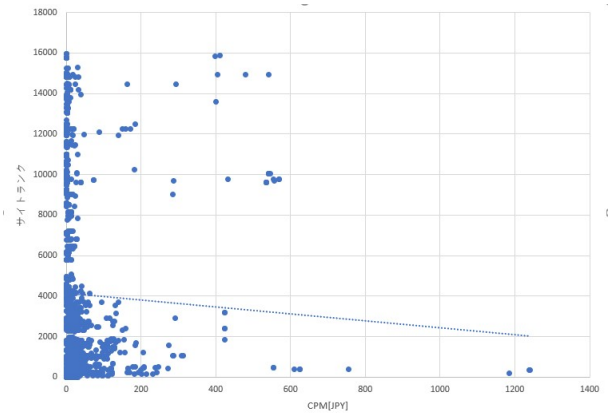

図 4.5: Ahrefs Rank と入札額の散布図

4.3.3 パブリッシャーによる違い

サイトの AhrefsRank と入札価格の散布図を図 4.5 に示す．$R2 = 0.0012$ で，線形回帰による有意な相関関係は見られなかった．また，指定キーワードが 1 回以上出現するかしないかによる入札額の分布の違いを図 4.6〜4.19 に示す．`0` は該当キーワードが存在せず，`1` は存在することを意味する．`shopping`，`comunication` が存在するサイトは，存在しないサイトに比べて，入札価格が低い傾向にあった．`art` や `health` が含まれるかどうかで入札価格に大きな違いは見られなかった．一方，`luxury`，`computers`，`games`，`science` が含まれるサイトでは，入札価格が高かった．

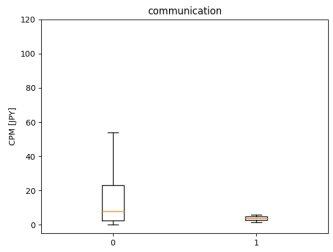

図 4.6: Communication

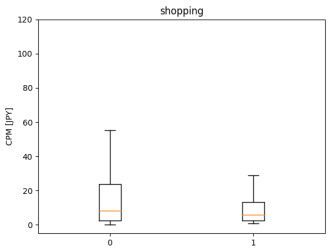

図 4.7: Shopping

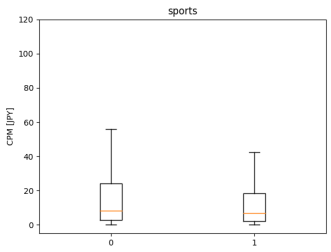

図 4.8: Sports

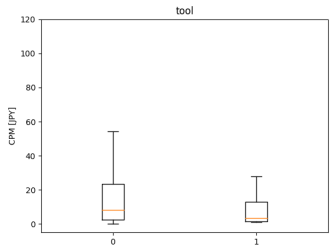

図 4.9: Tool

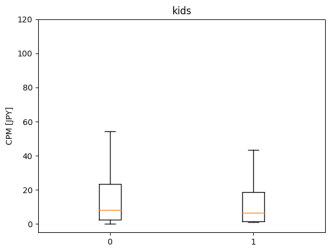

図 4.10: Kids

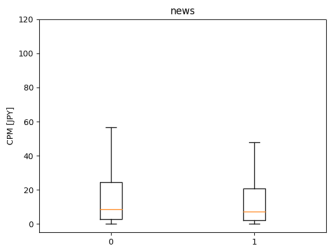

図 4.11: News

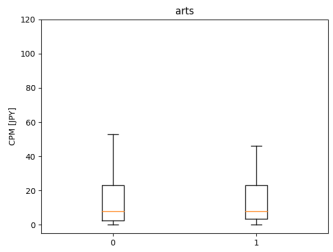

図 4.12: Arts

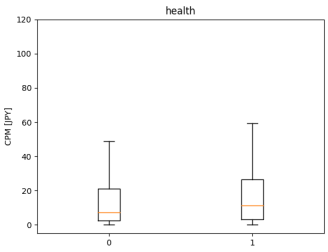

図 4.13: Health

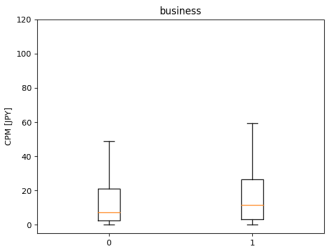

図 4.14: Business

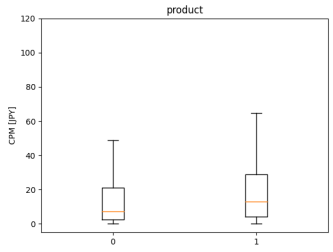

図 4.15: Products

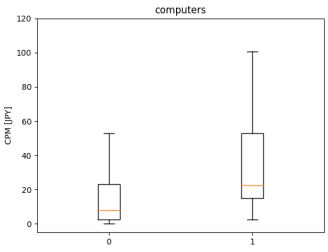

図 4.16: Computers

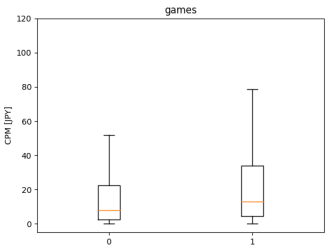

図 4.17: Games

図 4.18: Luxury

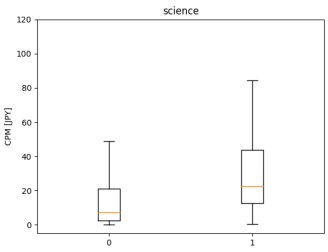

図 4.19: Science

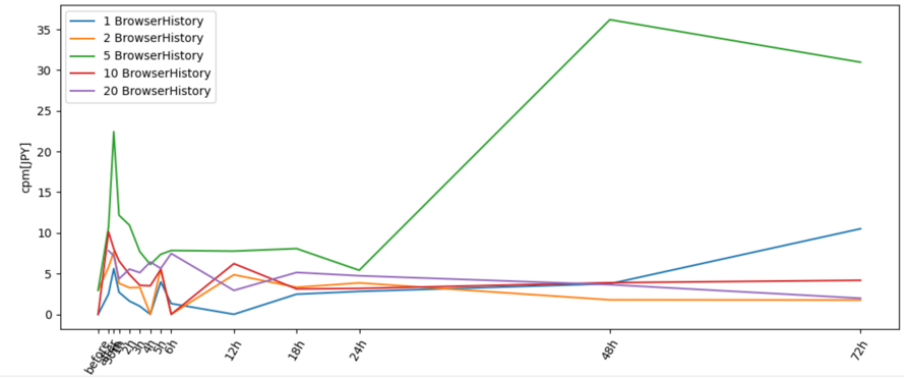

図 4.20: 商品閲覧後の最高入札額の推移

4.3.4 閲覧後の推移

商品を閲覧した後の入札価格の推移を図 4.20 に示す．ばらつきはあるものの商品の閲覧から直後には価格が上がり，約 6 時間後には元の価格まで下がっている．

手動で得られた結果をより一般的に表現するために，一つの商品につきユーザを 3 つ用意し，それぞれのユーザで同じ条件で閲覧を行った結果の中央値を採用したグラフを図 4.21 に示す．各ユーザは，表 4.1 に示す 3 の商品を 15 ページずつランダムに閲覧した後，各パブリッシャーサイトでの入札価格の推移を計測した．どのパブリッシャーサイトのどの商品を閲覧した場合も，手動で得られたような傾向は見られず，またサイト間での似た推移も見られなかった．

4.3.5 重回帰分析

それぞれの情報が入札価格に与える影響を明らかにするため重回帰分析を行った．目的変数を入札価格，説明変数を商品閲覧履歴からの経過時間，閲覧商品（商品 A，商品 B，商品 C，閲覧履歴なし），時刻，サイトランクとした．ここで用いたデータでは，商品 A のウェブページを閲覧したユーザは，商品 B，C は閲覧しておらず，他の商品に関しても同様である．結果を表 4.3 に示す．重回帰分析の結果から，サイトランクが一番関係が強いと考えられる．サイトランクと入札価格の散布図を図 4.22 に示す．サイトランクが高い程，入札価格が高い傾向が認められた．ただデータのばらつきが大きいため，重回帰分析からこれ以上の有意な結果は導き出せない．

4.4 考察

4.4.1 時刻による違い

図 4.2 と図 4.1 で，グラフの形に大きな差がなかったことから，少数の入札業者が特定の時間帯を狙って入札しているなどの行為がなかったことが分かった．

先行研究 [3] では朝の時間帯（0 8 時）で入札価格が高かったのに対し，図 4.2 より，夕方から夜にかけての時間帯で入札価格が高いという結果であった．従来研究による考察では，朝の時間帯はパブリッシャーサイトを閲覧するユーザが少なく，少ない枠に入札が殺到するためではないかとされてい

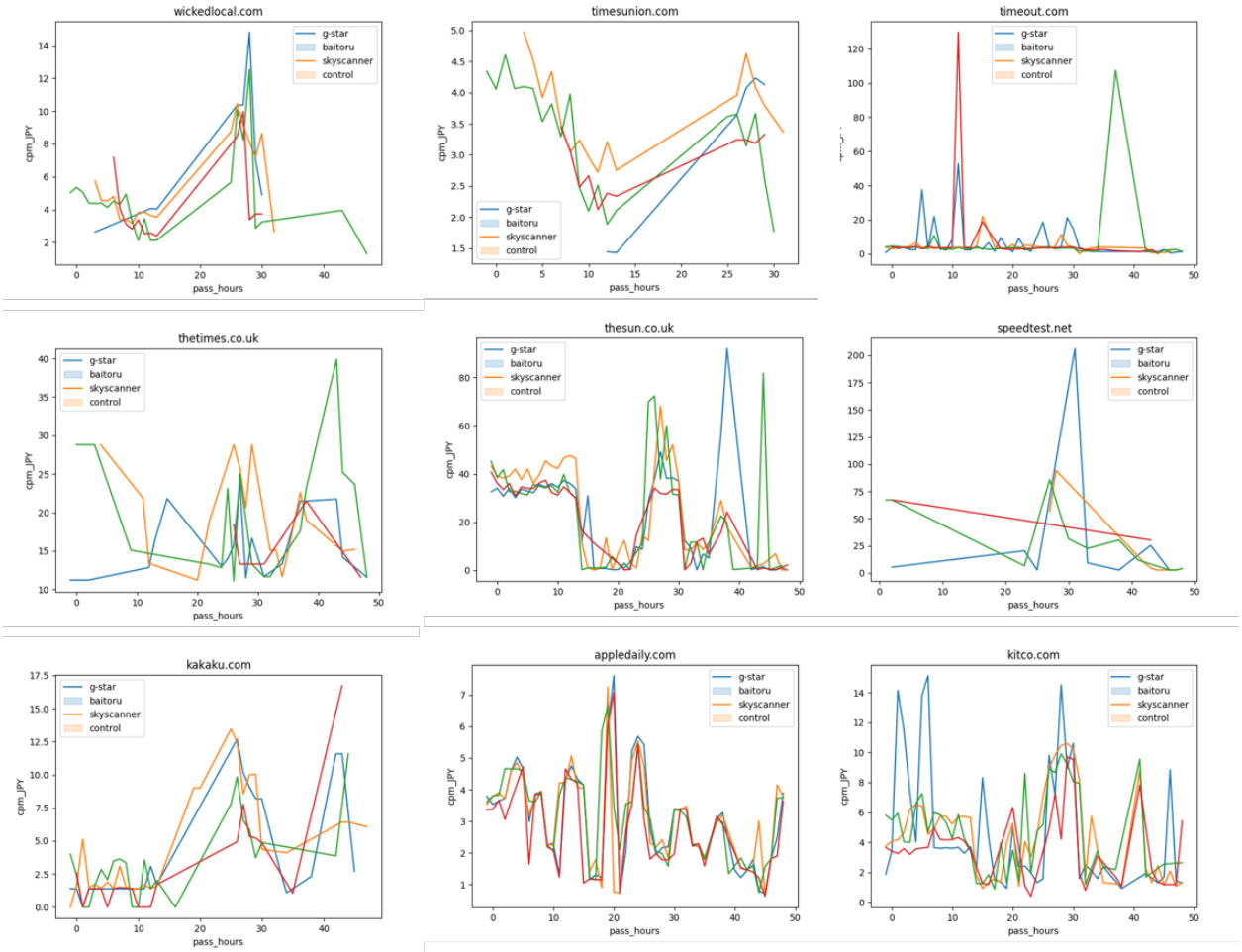

図 4.21: パブリッシャーサイトごとの最高入札額の推移

る．先行研究との結果が大きく異なった原因として考えられるのは，従来研究 [3] やその他の朝の時間帯が入札価格が高いという結果は，RTB から得られたものであり，HB では違う傾向にある可能性がある．予想としては，夕方から夜にかけてが最も活動的なユーザが多く，活発に広告クリックなどを行う傾向にあるのではないかと予想していた．ある広告主による効果計測によると [25]，夕方の時間帯ではクリック率が高いのに対し，早朝の時間帯ではコンバージョン率（商品の購入，無料体験の予約など広告主が目標とするユーザのアクションのこと）が高いという結果がある．目標とするユーザのアクションは広告主，業界や商品の種類などによって異なるため一概には言えないが，夕方の入札ではクリックを得ること，早朝の入札ではコンバージョンを得ることがそれぞれ目的とされていることが考えられる．

4.4.2 ユーザによる違い

図 4.3 と図 4.4 より，入札業者はユーザごとに違う振る舞いをしていると読み取れる．どの入札業者にも総じて人気であったのが健康に興味のあるユーザであり，一方履歴の多い通常のユーザは一部の入札業者からは他のユーザより圧倒的な入札価格での入札も行われていたもののそれ以外の入札，

表 4.3: 重回帰分析

	係数	標準誤差	t	p 値
(Intercept)	0	N/A	N/A	N/A
経過時間	-0.125060136	0.029578876	-4.228021866	2.61912E-05
商品 A	20.30309534	1.491545735	13.6121172	3.07413E-38
商品 B	20.54499301	1.506645529	13.63624861	2.34633E-38
商品 C	20.93063567	1.516246496	13.80424339	3.54791E-39
閲覧履歴なし	20.26322385	1.578508422	12.83694377	1.53491E-34
時刻	0.032976034	0.062606813	0.526716381	0.59853112
サイトランク	-0.001870851	0.000132536	-14.11581043	1.02831E-40

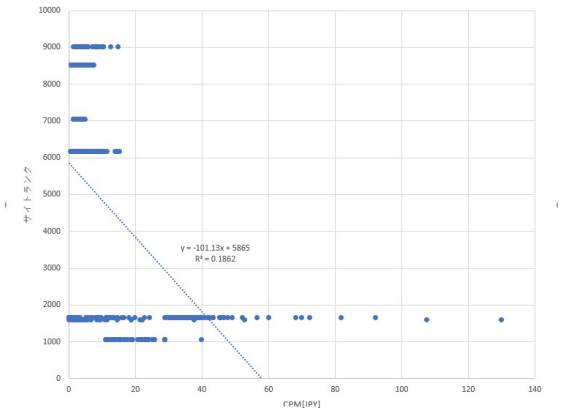

図 4.22: サイトランクと入札価格の散布図

入札業者による価格は平均的に高いというわけではなかった．原因として考えられるのは，健康に興味のあるユーザのアカウントでは健康に関する普通の記事を中心に閲覧しているのに対し，通常のユーザでは商品のページなども多く閲覧した履歴がある．よって 1 度自社サイトを訪れたユーザーを，サイト離脱後も追跡するリターゲティング広告が行われているため，価格が高くなっていることが考えられる．また，技術に興味のあるユーザと閲覧なしのユーザは，他の 2 つのユーザに比べて入札価格が低かった．価格が低い理由としては，技術に関する商品が多くないことや技術に興味のあるユーザによる広告のクリック率が低いことなどが考えられる．また，閲覧履歴がない，または少ないユーザの入札価格が低いことは，先行研究と同様の結果であり，ターゲティング広告として興味のあるユーザにリーチしたほうが広告効果が高いということがうかがえる．

4.4.3 パブリッシャーによる違い

ランクの高いサイトは，ユーザの滞在時間が長い傾向やユーザがコンテンツをしっかり見ている傾向などがあると考えられる．そのためランクの高いサイトの広告枠ほど入札額が高いのではないかと予想した．サイトの Ahrefs ランクと入札額の関係を示した図 4.5 より，有意な相関関係は認められなかったが，これは外れ値が大きいことが原因として考えられる．外れ値を除いて見ると，サイトのラ

ンクが 1〜4000 位のサイトでは，4000〜16000 位のサイトと比べて比較的入札価格が高い傾向が散布図から読み取れる．このことから，高い価格での入札をするかの判断には表示されるパブリッシャーサイトはあまり関係しないが，一般的な価格での入札ではサイトのランクが高いほど入札額も高くなりやすい傾向がある．比較的高い価格での入札がリターゲティング広告であるならば，どのサイトに表示されるかはあまり問題ではないためこのような傾向が見られたと考えられる．

また，サイトのカテゴリによる入札価格の違いを明らかにするため，指定キーワードがサイトのトップページに 1 回以上出現するかしないかによる入札額の分布の違いを図 4.6〜4.19 に示した．luxury，computers，games，science が出現するサイトでは入札額が高く，shopping，comunication が存在するサイトは，入札価格が低かった．また art や health が含まれるかどうかで入札価格に大きな違いは見られなかった．luxury，computers，games，science が含まれるサイトを閲覧するユーザは，購買意識が高いと判断されている可能性がある．一方，shopping，comunication は，何か他の商品を探していたり，人とコミュニケーションをとるという別の目的でインターネットを利用していることから，入札額が低くなっているのではないかと考えられる．

4.4.4 閲覧後の入札価格推移

商品閲覧後の入札価格の推移を示した図 4.21 と図 4.20 より，商品を閲覧したすぐ後に価格が上がることはなかった．また，閲覧する商品による大きな違いはなく，商品を閲覧しないユーザもおおむね入札価格の推移は一致していた．このことから，商品を閲覧した後，入札の価格が上がることがあるが，必ずしも上がるとは限らないことが分かった．原因として考えられるのは，リターゲティング広告が行われるためには，一定の条件があることが考えられる．例えば，閲覧履歴のある期間が短いユーザよりも，長い期間の閲覧履歴があるユーザのほうが入札価格が高い傾向にあることは，先行研究でも明らかにされている．

4.4.5 重回帰分析

表 4.3 より，４つの要素，商品閲覧履歴からの経過時間，ユーザ（商品 A，商品 B，商品 C，閲覧履歴なし），時刻，サイトランクが入札価格に与える影響を比較した結果，サイトランクが一番入札価格との相関が強かった．オンライン広告は，ユーザにあった広告を配信出来ることがメリットではあるが，やはり広告枠自体の性質も広告配信のメカニズムに大きく関わってると言うことができる．

第5章　結論

本研究では，HB の入札価格に影響を与える要因を明らかにすることでユーザのプライバシーの状況を明らかにすること，そして広告配信のメカニズムを少しでも解明することを目的として調査を行った．Header Bidding の入札情報を取得する実験を行い，以下のことが明らかとなった．

- リターゲティング広告を行う商品のサイトを閲覧したからと言って，必ずしも入札価格が上がるとは限らない．商品の閲覧からの経過時間が入札価格に影響を与えるには一定の条件がある．
- 広告が表示される時刻によって入札価格は変動する．夕方の時間帯で入札価格が高い傾向が認められた．その時間帯で広告のクリック率が高いと感じている広告主や広告配信業者が多いことが理由として考えられる．
- ユーザの閲覧履歴によって入札価格は変動する．リターゲティング広告を行う商品ページを閲覧したことがあるユーザには，高い価格で入札がされる．また，ユーザの興味対象によっても入札価格は変動し，健康に興味のあるユーザは入札価格が高かった．一方，履歴が多いだけで興味対象が広告主が求めるものと離れているユーザでは，入札価格は上がらない可能性がある．
- 広告が表示されるパブリッシャーサイトよって入札価格は変動する．サイトのランキングが高いほど入札価格も高い傾向にあり，また，特定のキーワードを含むサイトにおける入札価格は高く，特定のキーワードを含むサイトにおける入札価格は低かった．
- 経過時間，時刻，ユーザの興味対象，パブリッシャーサイトの特徴のうち，入札価格に最も大きな影響を与えるのはパブリッシャーサイトのランクであった．ユーザの興味対象の違いや時間軸など広告配信に影響を与える要素がいくつかある中で，サイトのランクが重要であるという広告配信の傾向が明らかとなった．

本研究の実験では研究目的のうち，経過時間や時刻，ユーザの閲覧履歴などの情報が入札価格にどれほどの影響を与えているのかを調査することでユーザのプライバシー問題を明らかにすることや，ユーザの閲覧履歴とパブリッシャーサイトの特徴がそれぞれどれほど影響を与えているのかを重回帰分析を用いて比較することで広告配信の傾向の一部を明らかにするはできた．しかし，ユーザに関するどんなデータがどれほど利用されているかや広告配信のメカニズムは依然としてあいまいであり，数理モデルを用いて説明したり，はっきりとしたデータを示すことはできなかった．今後の課題として，ユーザの履歴に関するより多くの要素を調査することやデータ数を増やすこと，そしてデータの取得方法として自動プログラムの影響を受けないために通常のユーザのブラウザに拡張機能を利用して取得するなどの工夫をすることが考えられる．

謝辞

本研究を行うにあたり，多くの方より御指導いただきました．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，菊池浩明教授に深く感謝申し上げます．修士論文審査にて有益なご助言を頂いた，斉藤裕樹教授，中村聡史教授，荒川薫教授並びに予備実験等に協力してくださった菊池研究室の皆様，先端メディアサイエンス学科の方々に深く感謝の意を表するとともに，謝辞とさせていただきます．

参考文献

[1] Ahrefs - SEO Tools & Resources To Grow Your Search Traffic, https://ahrefs.com/

[2] John Cook, Rishab Nithyanand, and Zubair Shafiq，”Inferring Tracker-Advertiser Relationships in the Online Advertising Ecosystem using Header Bidding”，Proceedings on Privacy Enhancing Technologies, 2020 (1):1–18

[3] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson，”Tracing Information Flows Between Ad Exchanges Using Retargeted Ads”，25th USENIX Security Symposium, 2016

[4] Papadopoulos, Panagiotis and Kourtellis, Nicolas and Rodriguez, Pablo Rodriguez and Laoutaris, Nikolaos, and Christo Wilson，”If you are not paying for it, you are the product: How much do advertisers pay to reach you?”，IMC ’17: Proceedings of the 2017 Internet Measurement Conference

[5] José González Cabañas, Ángel Cuevas, Rubén Cuevas，”FDVT: Data Valuation Tool for Facebook Users”，CHI ’17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems

[6] Lukasz Olejnik, Tran Minh-Dung, Claude Castelluccia, ”Selling Off Privacy at Auction”, 2014

[7] 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.

[8] 2019 年日本の広告費，(https://www.dentsu.co.jp/knowledge/ad_cost/2019/，2020/12/16 参照．)

[9] ADSTAGE，(https://www.adstage.io/，2020 年 12 月参照．)

[10] NHK,「もうけは誰の手に？闇に消えるネット広告費」，(https://www3.nhk.or.jp/news/special/net-koukoku/article/article_05.html，2020 年 6 月参照．)

[11] Md Shahrear Iqbal，Mohammad Zulkernine，Fehmi Jaafar，Yuan Gu，“ Protecting Internet users from becoming victimized attackers of click-fraud ”，WILEY，Journal of Software Evolution and Process，2018.

[12] Metwally A, Agrawal D, Abbadi AE. Using association rules for fraud detection in web advertising networks. Proceedings of the 31st International Conference on Very Large Databases, VLDB Endowment, Trondheim, Norway, 2005.

[13] Immorlica N, Jain K, Mahdian M, Talwar K. Click fraud resistant methods for learning click-through rates. Proceedings of the Internet and Network Economics: Springer, Hong Kong, China, 2005.

[14] Xu H, Liu D, Koehl A, Wang H, Stavrou A. Click fraud detection on the advertiser side. Proceedings of the 19th European Symposium on Research in Computer Security: Springer, Wroclaw, Poland, 2014.

[15] 金井 文宏 ほか，広告ネットワーク上で観測されたユーザアクティビティの分析による広告不正の実態調査，情報処理学会 研究報告セキュリティ心理学とトラスト (SPT)， pp. 1-6，No. 17，Vol. 2018-SPT-27，2018.

[16] 総務省,「平成 29 年版 情報通信白書 主なメディアの利用時間帯」， (https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc262520.html，2020 年 12 月参照．)

[17] 石川 善一郎，奥 牧人，河野 崇，”Web リスティング広告における基本広告データを用いたコンバージョン分析と予測”，DEIM Forum，2017

[18] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In Proc. USENIX Security

[19] Jonathan R. Mayer and John C. Mitchell, ”Third-party web tracking: Policy and technology.”, 2012

[20] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor,Richard Shay, and Yang Wang. 2012. ”Smart, useful,scary, creepy: Perceptions of online behavioral advertising”, In Proc. SOUPS.

[21] Papadogiannakis, Emmanouil and Papadopoulos, Panagiotis and Kourtellis, Nicolas and Markatos, Evangelos P., ”User Tracking in the Post-cookie Era: How Websites Bypass GDPR Consent to Track Users”, Association for Computing Machinery, 2021

[22] Yamato Hojyo, Yuta Saito, Takamichi Saito, ”Passive Fingerprinting Enforced with Deep Neural Network”, Computer Security Symposium, 2019

[23] P. Papadopoulos, N. Kourtellis, and E. Markatos. ”Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask”, In The Web Conference (WWW), 2019

[24] Avi Goldfarb, Catherine Tucker, ”Online Display Advertising: Targeting and Obtrusiveness”, Marketing Science, Vol. 30, No. 3, May–June 2011, pp. 389–404

[25] カルテットコミュニケーションズ, ”《リスティング広告》一歩踏み込んだ分析を〜曜日×時間編〜”, (https://quartet-communications.com/info/listing/technique/24615，2023 年 2 月参照．)