

CTR 向上を目的とした  
WEB ページ上でのオンライン広告配置位置推定  
An Online AD placement on WEB pages for improving CTR  
大谷 一善<sup>†</sup> 滝川 真弘<sup>†</sup> 堀田 弘明<sup>†</sup> 山名 早人<sup>†</sup>  
Kazuyoshi Ootani Masahiro Takigawa Hiro-Aki Hotta Hayato Yamana

## 1. はじめに

スマートフォン等の携帯端末の普及に伴い、人々が容易にインターネットへ接続できる環境となった。北原ら<sup>1</sup>によると「広告の掲載手段としてインターネットが年々増加の一途を辿り、2015 年におけるオンライン広告費は総広告費の 18.8%と約 2 割を占める」に至っており、広告ビジネスにおいてオンライン広告が重要な媒体となっていることが分かる。オンライン広告を掲載する目的は、広告主が設定したコンバージョン(会員登録や商品購入など)をユーザが行うこと、または広告の内容をユーザに認知させることである。そのため、広告の表示数に対するコンバージョン数やクリック数は広告効果を測る上で重要な指標であり、それぞれコンバージョン率(CVR: Conversion Rate)、クリック率(CTR: Click Through Rate)が評価において用いられる。また、オンライン広告は一般に広告媒体を通じて配信されるため、広告がユーザの端末に表示される、もしくはユーザが広告をクリックすることで広告媒体への収益となる。以上、CTR・CVR の向上は、広告主だけでなく、広告媒体の収益向上に大きな影響を与える。

広告主の目標は、最終的に CVR を向上させることである。しかし、CVR は一般的に小さく、コンバージョンを起こしたユーザと広告との関係はスパース性が高いため、コンバージョンを指標として各種手法を提案することは困難である。そこで、本研究では CVR の代わりに CTR に着目し、CTR の向上を目的とする。CTR 向上には、ユーザに対して最適化した広告配信が有効となる。ここで、広告配信サービス提供者が検索エンジンなどを並行して提供している場合、検索履歴などの付加情報をもとに広告配信のチューニングを行うことができる。しかし、一般的な広告配信サービスでは、こうしたユーザの詳細な特徴量を用いることができない。

こうした状況のもと、本稿では、ユーザの詳細な特徴量を用いないことを前提とした上で CTR を向上させるための手法として「広告の WEB ページ上での配置」に取り組む。Chen ら[1]は、オンライン広告を WEB ページ中に配置する際、WEB ページの上部と下部では CTR が最大で約 25 倍の差があると報告している。このことから、オンライン広告の WEB ページ中での配置位置は CTR に影響を与える一要因であることがわかる。

オンライン広告の WEB ページ中での最適な配置に関する研究は大きく 2 つに分類できる。(1) 予め被広告掲載 WEB ページ中の広告枠が複数与えられ、CTR・CVR の高い広告枠を選択する手法[1][2][3][4]、(2) 予め被広告掲載 WEB ページは与えられるが広告枠の位置は決定しておら

ず、ページのレイアウト変更により、当該 WEB ページ中の広告枠配置を変更し CTR・CVR を向上させる手法[5]である。(1)の手法は与えられた複数の広告枠から実際に広告を表示する枠を選択する手法であり、WEB ページのレイアウト変更は行わない。一方、(2)の手法は、掲載広告に応じて WEB ページのレイアウト変更を行う手法であり、同一の WEB ページであっても掲載される広告に依存してレイアウトを変更する。これは、ユーザビリティの観点から適切ではなく、多くのページに汎用的に適用することはできない。よって本稿では、(1)の手法を前提とする。

(1)に関する研究は、さらに 2 つに分類できる。(a)ユーザが検索エンジンを使用した際のクエリやユーザの詳細情報を付加情報として用いた研究[1][2]と(b)クエリやユーザの詳細情報を用いずに広告のクリック履歴等のみを用いた研究[3][4]である。(a)では、ユーザと広告を紐付ける要素として、ユーザの検索クエリや性別や年齢層等のユーザの詳細情報をユーザの特徴として利用している。しかし、一般的な広告配信サービスでは、検索クエリやユーザ情報等のユーザと広告を紐付ける情報は持ち得ない。このため一般的な広告配信サービスでは、(b)の手法を用いて広告配信の最適化が必要となる。

以上述べたように、本稿では、広告の WEB ページ上での最適配置、すなわち CTR を最大化する配置について、上記の(b)を制約条件とした上での最適化に取り組む。具体的には、あるオンライン広告のある被掲載 WEB ページが持つ何れかの広告枠に掲載する際に、広告 CTR が最大となる広告枠の推定を行う。ユーザの詳細情報を用いず、過去のユーザのクリック履歴をもとにユーザを分類すると共に、被クリック数が少ない広告に対しても推定できるように、広告を分類しまとめることで推定する。ユーザ及び広告の分類では LDA を用いる。ユーザがクリックした広告が属する分野、クリック時間などを元にして LDA によりユーザを分類する。次に、各広告の被クリックユーザ集合を元に、LDA により広告を分類する。これらの分類されたユーザと広告との関係から CTR が最大となる広告配置位置を推定する。本稿の貢献は、ある広告が与えられた際に、WEB ページ内での広告枠選択をユーザに対して最適化することにある。

本稿では 2 節で関連研究について紹介し、3 節では実験で用いるログデータの解析を行う。その結果を元に、4 節で提案手法を示す。5 項では提案手法を評価し、6 節にてまとめる。

<sup>†</sup>早稲田大学

‡株式会社ジーニー

<sup>1</sup>株式会社 電通, “日本の広告費”, <http://www.dentsu.co.jp/news/release/pdf/cms/2016022-0223.pdf>

## 2. 関連研究

CTR 向上を目的としたオンライン広告の適切な配置に関する研究は大きく 2 つに分類できる。適切な広告枠を選択する手法とページレイアウトを比較・選択する手法である。2.1 では前者の研究として、Schroedl らの手法[2]、Chen らの手法[1]、Yuan らの手法[3]、Koutsopoulos らの手法[4]を挙げる。また 2.2 では後者の研究として、Cheng らの研究[5]を挙げる。

### 2.1 広告 CTR が最大となる広告枠を選択する手法

広告 CTR が最大となる広告枠を選択する手法は、WEB ページにおいて設定された複数の広告枠に対し、広告の CTR が最大となる広告枠へ入稿する手法である。このため、CTR 向上を目的として、WEB ページ上の広告枠の位置やコンテンツを表示する位置の変更は行わない。本項では GBDT(Gradient Boosted Decision Tree)[6]を用いた手法と EM アルゴリズム[7]を用いた手法、品質スコアを用いた手法、グラフ最適化を用いた手法を説明する。

#### 2.1.1 Schedule らの手法

Schroedl ら[2]は、収益向上を目的とし、Yahoo!<sup>1</sup>のデータを用いて WEB ページ上の広告表示の最適化を行った[2]。具体的には、WEB ページにおける複数の広告枠に入稿する広告の CTR 予測を行い、算出された CTR に対し入札額  $bid$ 、位置バイアスを考慮し当該 WEB ページに掲載される広告から得られる収益  $revenue$  を予測する。最後に、収益の予測値と予め定めた目標値  $\theta$  を比較し、少なくとも目標収益が得られるように広告枠を選択する広告 CTR を算出する方法として、広告の掲載位置に依存しない、位置について正規化された正規化 CTR を算出する。正規化 CTR に対し、広告に依存せず WEB ページ上の特定位置に広告が掲載されることによる CTR のバイアス乗じることで広告 CTR 予測値とした。さらに、各ユーザのクリック・検索履歴から、ユーザ  $u$  の固有の特徴量を示す  $ucp_u$  を新たに定義した。 $ucp_u$  は、ユーザ  $u$  とある位置に表示される広告間のバイアスを示し、広告の種類やページに依存しない値である。

先に示した広告 CTR 予測値に、 $ucp_u$  を乗じた値を、ユーザ  $u$  に対する当該広告枠集合における CTR とし、この結果を最大化するような、広告の表示数と配置位置の推定を行った。推定には 45 個のパラメータを使用し、GBDT により推定を行った。WEB ページ中の広告配置を最適化した結果、WEB ページ中の広告数を約 9%減らした場合も、減らす前と同等の CTR が得られ、さらにページ全体の収益が約 1%向上したと報告している。

#### 2.1.2 Chen らの手法

Chen ら[1]は、Microsoft Ad center<sup>2</sup>のデータを用いて CTR 予測を行う際、広告配置によるバイアスを排除する方法を提案した[1]。すなわち、広告が特定の位置に配置されることにより CTR が上がる／下がるといった現象(バイアス)を考慮した CTR を提案した。Chen らは CTR を 2 種類の確率として捉え、個々を推定し、掛け合わせることで CTR 予測を行った。2 種類の確率とは、位置に依存せず、

広告  $ad$  とクエリ情報  $query$  の組み合わせからクリックされる確率  $p(\text{click} | ad, query)$ 、位置  $position$  に広告  $ad$  が配置される確率  $p(ad | position)$  である。

また、広告がクリックされる事象は、クリックされる／されないの 2 値で表現され、二項分布に基づく事象とした。また、一般的に CTR は非常に低い値であり、ページ内の位置によりインプレッションとクリックデータ量にばらつきがある。以上から、Chen らは広告がクリックされる事象は Gamma-Poisson 分布と仮定し、近似した。

以上を踏まえて、EM アルゴリズムを用いて CTR 予測を行った。素性には広告の表示位置、クエリとの適合度、同一広告枠集合中の広告掲載数を用いた。先に示した Gamma-Poisson 分布を考慮せず、位置に関する特徴として位置のみを使用した場合との比較を行った。その結果、クリック予測の AUC が 0.8225 から 0.8750 に改善したと報告した。さらに、訓練期間を 2 倍にあたる 2 週間に延長することで、AUC が 0.8809 に向上すると報告している。

#### 2.1.3 Yuen らの手法

Yuan ら[3]は、検索結果に表示される広告の品質スコア (Quality Score) に注目し、収益が最大となるような広告の配置を行なった[3]。品質スコアとは、Google AdWords<sup>3</sup> や Yahoo Search Marketing<sup>4</sup> などが設定している検索広告の評価指標であり、同一ページに掲載される広告同士を比較するために用いられる。品質スコアは、当該広告の過去の CTR に大きく影響を受け、品質スコアと CTR の間には正の相関があるとされている。また、品質スコアにより広告の表示位置が変化する。

Yuan らは品質スコアを元に広告の配置位置を決定する方法を提案した。具体的には、広告の品質スコアを維持しつつ広告配置位置を適宜変化する方法(以下、方法 3-1)、一定期間において過去の品質スコアから最大収益をとる配置位置を決定し広告配置位置を固定する方法(以下、方法 3-2)の 2 つを提案し、比較した。方法 3-1 と方法 3-2 のそれぞれにおいて、次にインプレッションがあった際の収益が最大となる広告配置位置を 1000 回 RTB に対し行った。RTB(Real Time Bidding)とは、WEB ページ上の広告枠に掲載する広告を決定する手続きである。また、方法 3-1 が 30%の RTB において方法 3-2 よりも高収益であった。なお、10%の RTB は方法 3-2 が方法 3-1 よりも高収益であった。以上から方法 3-1 が方法 3-2 に比べ、性能が良いことがわかった。しかし、方法 3-1 は RTB 毎に広告配置位置を計算するため、1,000 回 RTB を行う際に、広告の配置位置を固定する方法 3-2 に比べて計算量が多い。さらに、方法 3-2 では広告の配置位置決定が簡単な一方で、動的に大きく変化する広告市場に対応しきれない可能性がある。以上のことを考慮し、2 つの方法には性能的な差はないと報告している。

<sup>1</sup> <https://advertising.yahoo.com/>

<sup>2</sup> <https://secure.azure.bingads.microsoft.com/>

<sup>3</sup> Google AdWords, <http://www.google.co.jp/adwords/>

<sup>4</sup> Yahoo!マーケティングソリューション, <http://marketing.yahoo.co.jp/>

### 2.1.4 Koutsopoulos らの手法

Koutsopoulos ら[4]は The Guardian<sup>1</sup> のデータを用いて、フィード中に表示される広告を収益が最大となる広告配置方法の提案をした[4]。フィードとはニュースやブログサイトにおいてコンテンツの見出しや概要を簡潔に配信するに適したフォーマットである。Koutsopoulos らは、複数の広告を同時に複数の広告枠へと掲載する方法として、グラフ最適化を用いた 2 つの手法を提案した。具体的には、最短経路問題利用した方法(以下、方法 4-1)と 2 部グラフにおける最大マッチング問題を利用した方法(以下、方法 4-2)である。

方法 4-1 は、頂点を広告配置位置と広告のペアとして、その頂点を結ぶ最短経路問題である。2 頂点を結ぶ辺は有向であり、始点が示す広告の次に、終点が示す広告を掲載した際の収益の期待値を辺の重みとして設定した。経路はフィード中に掲載される広告とその広告配置位置を表し、フィード内全体の収益を表す。また、同じ位置には複数の広告を掲載できないこと、同じ広告を複数位置に掲載できないことを制約条件とした。以上に示した最短経路問題を解き、フィード内の収益が最大となる広告とその配置位置を決定する。

方法 4-2 は、2 部グラフにおける 2 つの頂点集合を、広告の集合と広告配置位置の集合とした。頂点集合同士を結ぶこと、すなわち広告を表す頂点と広告配置位置を表す頂点を結び、辺を作ることで、ある広告のフィード中の広告掲載位置を定義できる。辺には重みとして、頂点が表す広告配置位置に広告を掲載した際の収益の期待値を設定した。以上から、2 部グラフマッチングを行うことでフィード内全体の収益を最大化する各広告のフィード内における配置位置が決定する。

以上から、方法 4-1 と方法 4-2 におけるフィード内における収益を最大化する広告の配置方法を求める。なお、経験的に各広告において最も収益が高くなる広告の配置法を比較手法とした。

方法 4-1 は比較手法に比べ最大 8%収益が向上したが、方法 4-2 は比較手法に比べ、収益が最大 20%低下した。しかし広告数が 1 つの場合、提案した 2 手法と比較手法を用いた収益は近い値であった。また方法 4-2 は方法 4-1 に比べ計算量が少ない。以上から、N が小さい場合は方法 4-1 を用いた上で、広告数が多い場合は方法 4-2 を用いることで、最大収益となる広告のフィードへ配置が可能となると報告した。

### 2.2 ページレイアウトを比較選択する手法

ページレイアウトを比較・選択する手法は、広告の CTR や収益の向上を目的として、WEB ページ上の広告枠や記事などの構成要素の配置(レイアウト)を変更する手法である。本項では、GBDT を用いた Cheng らの手法[5]を取り上げる。

Cheng らは Yahoo! のデータを用いて、WEB ページにおけるレイアウト選択の最適化を行なった。Cheng らは、5 つの WEB ページのテンプレートを用いた。各テンプレートについて、広告を掲載する際の eCPM (1,000 回表示する

ために必要な広告掲載費)を予測し、得られた数値を用いて、レイアウト比較を行った。eCPM の予測には、サイト ID や独自のカテゴリ ID などを素性として、GBDT を用いて eCPM を予測した。その結果、動的にレイアウトを変更しない場合に比べ 45.87%収益が向上したと報告している。

### 3. 実験データと予備実験

本研究では GenieeDSP<sup>2</sup> において、PC サイト上に掲載されるバナー広告のログデータを使用した。2016 年 11 月 1 日から 2016 年 11 月 28 日の 4 週間におけるサンプルクリックデータ 3,386,930 件とサンプルインプレッションデータ 1,715,027,298 件を使用した。また、解析の精度を高めるため、ユーザの本意ではない意図的なクリックをするユーザを排除するための前処理として、経験則に基づき 4 週間で 50 回以上クリックしたユーザを排除した。結果を表 1 に示す。

以下では、各種素性が CTR に与える影響を調査するために予備実験を行い、その結果について述べる。

表 1 データクリーニング前後でのデータ数

	クリック数	インプレッション数	ユーザ数
除外前	3,386,930	1,715,027,298	45,347,239
除外対象*	973,879	174,779,544	8,206
(全体に占める割合[%])	28.75(%)	10.19(%)	0.18(%)
除外後	2,416,051	1,540,247,754	45,339,033

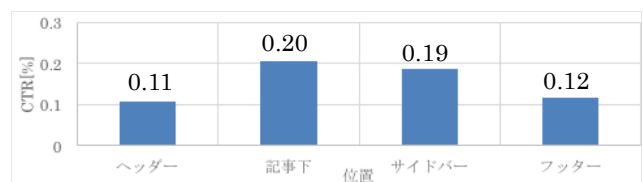


図 1 各位置における CTR

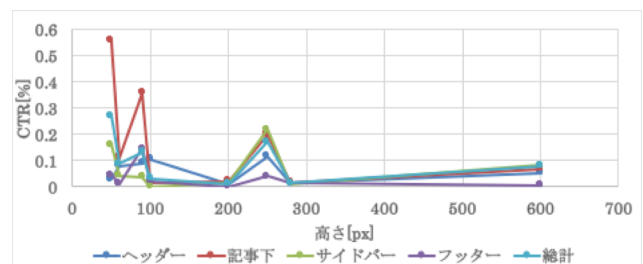


図 2 各位置における広告サイズ(高さ)と CTR [%] の関係

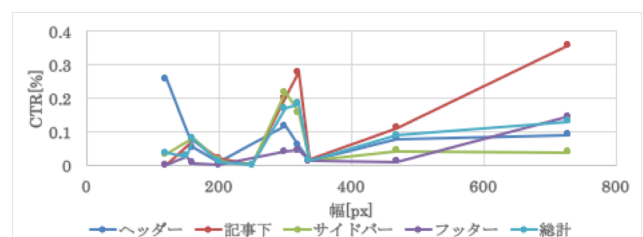


図 3 各位置における広告サイズ(幅)と CTR [%] の関係

<sup>1</sup> <https://www.theguardian.com/international>

<sup>2</sup> GenieeDSP, " <http://geniee.co.jp/products/dsp.php> "

### 3.1 予備実験1：広告表示位置によるCTRの違い

Schroedl らの報告[2]にあるように、広告は一般的に、その配置位置がページ最上部の時に最もクリックされ最下部になるにつれて低下していく。この性質を確認するために、実験データにおいて広告の配置位置毎における CTR を算出し、図 1 に示す。図 1 からページ最上部であるヘッダーが必ずしも CTR が最高でないことが分かる。具体的には CTR は最上部が最も低く、記事下が最も高い。また、ページ中央から外れる(サイドバー)ことで CTR が低下する。

### 3.2 予備実験2：広告サイズによるCTRの違い

広告の形状による CTR の影響を調べるために、広告サイズと広告面積ごとの CTR をそれぞれ調べた。広告サイズは 12 種類観測された。広告面積と CTR との相関係数は 0.47 となり、弱い正の相関がある。サイドバーについて、広告面積と CTR の相関係数は 0.46 と弱い相関があった。サイドバー以外の各広告配置位置においては、CTR と広告面積の間に相関はほぼ無いことが分かった。

次に、広告における高さ(図 2)、幅と CTR の関係(図 3)を調べた。記事下とフッターにおいて、高さ(幅)と CTR の相関係数はそれぞれ -0.42 と -0.37 であり、共に弱い負の相関がある。一方、幅と CTR の相関係数はそれぞれ 0.74 と 0.82 であり、正の相関がある。これより、記事下とフッターには幅が長い広告を配置することで CTR が向上すると推測される。ヘッダーとサイドバーでは高さ・幅において CTR との相関はないことが分かった。以上から広告面積に比べ広告の幅・高さが広告 CTR に影響を及ぼしていることが分かる。

### 3.3 予備実験3：時間帯によるCTRの違い

曜日・時間帯による CTR の変動を調べた。1 日を 6 時間に分割した CTR を図 4 に示す。また、人が活動する時間に合わせた時間帯区分の例として電気料金時間帯区分<sup>1</sup>を用いた時間帯を区分した場合の結果を図 5 に示す。図 4 と図 5 より、CTR の遷移に大きな差はなく、夕方～夜の時間帯に CTR が向上することが分かる。

### 3.4 予備実験4：バナー・ゾーンカテゴリによるCTRの違い

広告の分類カテゴリが CTR に与える影響を調べるために、バナーカテゴリ毎の CTR について解析した。バナーカテゴリは 504 分類存在し、このうち 65 分類が本研究の解析期間において出現した。バナーカテゴリは IAB[8]が定めた分類であり、全 504 分類から上位分類である 26 分類を解析では用いた。最終的に、バナーカテゴリ 26 分類とバナーを表示するゾーン(広告位置)の種類(ゾーンカテゴリ) 26 分類の間でクロス集計を行った。

結果を図 6 に示す。図 6 は色が濃いほど CTR が高いことを示している。なお、図 6 作成にあたっては、CTR の外れ値検定の結果、11.10%を超えるものは外れ値として黄色で、インプレッションがない場合はグレーで示している。なお、バナーカテゴリは 19 分類出現している。図 6 より

バナーとゾーンカテゴリの両者を比較したときに、カテゴリ別 CTR は共に傾向が異なっていることが分かる。

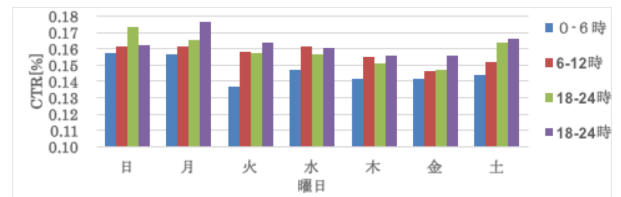


図 4 曜日別・時間帯別 CTR(1 日を均等に 4 分割)

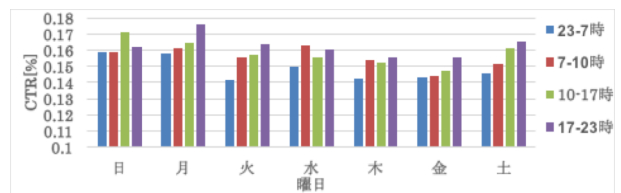


図 5 曜日別・電機料金区分時間帯別 CTR

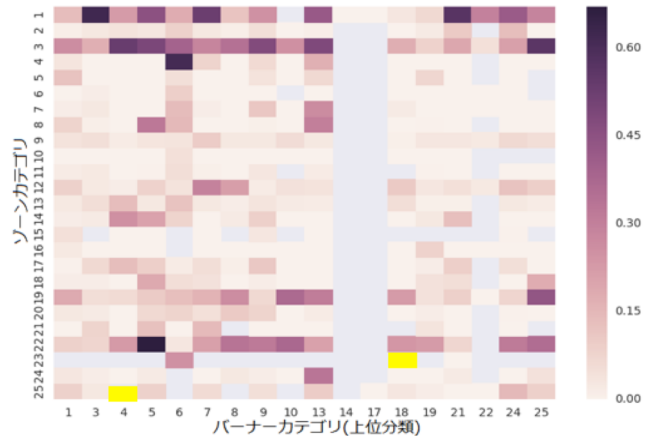


図 6 ゾーン・バナーカテゴリ別 CTR のヒートマップ

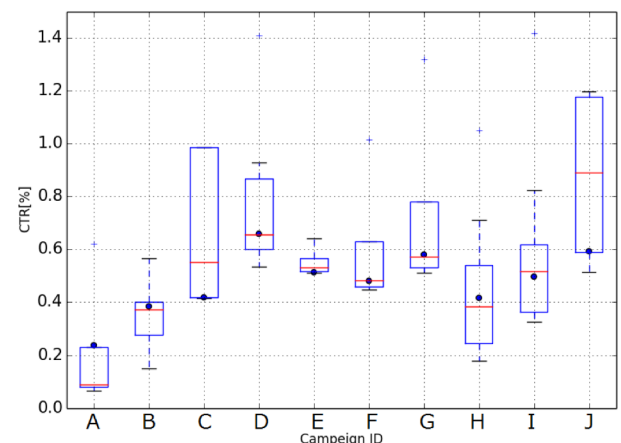


図 7 キャンペーン別 CTR  
(バナーカテゴリ 5・ゾーンカテゴリ 3)

<sup>1</sup>東京電力エナジーパートナー, “電化上手-電気料金プラン”,  
“<http://www.tepco.co.jp/ep/private/plan2/old03.html>”

### 3.5 予備実験 5 : キャンペーンによる CTR の違い

本項では、1つのバナーカテゴリに複数存在するキャンペーンに着目し、キャンペーンが CTR に影響を与える影響を調べる。同一バナーカテゴリかつ、同一ゾーンカテゴリにおける各キャンペーン別 CTR の分析を行う。キャンペーンとは、宣伝活動の単位である。図 7 にバナーカテゴリ 5, ゾーンカテゴリ 3 の各キャンペーンのバナー別 CTR を示した。なお、本組み合わせは全てのデータの中から最もクリック数が多い組み合わせを抽出した。

図 7 より、キャンペーンごとに CTR の差があることが分かる。また同一キャンペーン内でも、バナー毎で CTR に差がある。キャンペーン A, C, J は中央値と平均値の差が大きい。この原因として、特定の広告によりキャンペーン別 CTR が引き上げられている。もしくは、引き下げられていると推測される。逆に、キャンペーン E, F は中央値と平均値の差が小さく、同一キャンペーンの広告であれば広告 CTR がお互いに近い値である。前項を踏まえて、同一バナー・ゾーンカテゴリにおいては、キャンペーンごとに分散が大きい。しかし、同一キャンペーン内では広告 CTR の分散が大きいキャンペーンが存在するが、分散の小さいキャンペーンも複数確認できる。よって、広告の CTR をある広告集合単位でまとめ上げること考える場合、キャンペーンを特徴量のひとつとして利用できる可能性がある。

## 4. CTR 向上を目的とした広告配置位置推定方法

### 4.1 概要

本研究では、ユーザの広告クリック情報のみを用いて、広告被掲載 WEB ページにおける広告の配置位置を推定し、CTR を向上させる方法を提案する。3 節における解析の結果、広告の位置、サイズ、時間、被広告 WEB ページが広告 CTR に大きく影響を及ぼす要素として確認できた。よって以下では、これらの特徴量を用いて広告配置位置を求める。手順は以下の通りである。(1)ユーザのクリック情報を用いたユーザ分類を行い、(2) (1)の結果を用いて広告を多クラス分類する。そして、(3)広告の多クラス分類結果を用いて、CTR 向上を目的とした広告の WEB ページにおける配置位置を推定する。

- (1) 一般的な広告配信サービスではユーザの詳細な情報は持ち得ない。代わりにユーザの広告クリック情報を用いてユーザ分類を行うことで、ユーザの特徴を抽出し利用する。その際、CTR に大きな影響を及ぼすクリック情報として、キャンペーン、広告の被掲載 WEB ページ、時間、広告サイズ、広告の位置を用いる。
- (2) 各広告において、クリックしたユーザの集合を元に広告に対し多クラス分類を適用する。その際、各ユーザの特徴として、(1)におけるユーザ分類結果を用いる。
- (3) 各広告に付与された分類を用いて、CTR が最大となる広告の配置位置を推定する。

時間は曜日と時間帯を用いる。時間帯は電気料金の時間区分を用いて 1 日を 4 分割する。1 週間単位で扱うため、

時間は 28 分割される。また、本研究の貢献として 2 点挙げられる。

- 検索クエリやユーザの詳細な情報は持ち得ない一般的な広告配信サービスにおいて、クリック情報からユーザの特徴を抽出し反映する手法の提案。我々が知る限り、広告の最適な配置位置推定で用いた事例は存在していない。
- 各広告をクリックしたユーザ集合を元に多クラス分類し、その分類結果を用いて各広告における CTR が最大となる広告の位置を推定することにより、ユーザ毎に CTR 最大配置位置を求める手法と比較してコールドスタート問題に強く、かつ、複数の広告の情報を利用できるために、広告の WEB ページにおける配置位置推定の精度を向上が可能となる点。

比較手法となるベースラインとして、次の 4 手法を用いる。

- (i) WEB ページ別広告配置手法  
広告に関係なく、WEB ページ単体で最大 CTR となる位置を推定する手法。(ユーザ情報を用いない手法)
- (ii) 広告別広告配置手法  
WEB ページに関係なく、広告単体で最大 CTR となる位置を推定する手法。(WEB ページ情報を用いない手法)
- (iii) ランダム配置手法  
広告を掲載する際の WEB ページ上の位置をランダムに決定する手法
- (iv) 広告・WEB ページ別配置手法  
広告と WEB ページのペアで最大 CTR となる位置を推定する方法(ナイーブ手法)

なお、提案手法は、比較手法(iv)にユーザの広告クリック情報を加えた手法である。以下、4.2 では、提案手法について述べる。

### 4.2 提案手法

4.1 に示した概要の具体的な説明を行う。以下に示す過程を経て、広告の WEB ページ上の配置位置の推定を行う。

- 4.1(1)に示したように、クリックした広告からユーザ分類を行う。具体的には各ユーザについて、クリックした広告が持つ特徴の集合を作り、LDA[9]を用いることで、ユーザ毎にトピックが付与される。付与されたトピックのうち、最も尤もらしいトピックをグループとしてユーザに付与する。広告の特徴としてキャンペーン(宣伝内容)、WEB ページ、時間、広告サイズ、広告の位置を用いた。この LDA を用いて推定する過程の概念図を図 8 に示した。以上から、広告 CTR に影響を及ぼす特徴を用いることで、効果的なユーザ分類ができると考えた。
- 4.1(2)に示したように、広告クリックにおいて類似する性向を持つユーザ集合から LDA を用いて広告を分類する。具体的には各広告について、クリックしたユーザの特徴で集合を作り、LDA を用いて、特徴の集合から広告にグループを付与する。ユーザの特徴として、ユーザがクリックしたキャンペーン・WEB ページ・時間・サイズ・広告の位置を用いた。以上の特徴それぞれについてグループを付与した。図 9 に概要を示した。
- 4.1(3)に示したように、広告の分類結果を用いて、各広告を WEB ページ毎に最大 CTR となる位置を求める。具

体的には、一定期間において、WEB ページ毎に同一グループの広告を集約し、位置ごとの CTR を算出し、最大 CTR となる位置を求める。求められた WEB ページ・同一グループ毎の最大 CTR となる位置を、当該グループに属する全ての広告に適用する。以上から各広告に対する広告掲載 WEB ページ毎の最適な位置を決定する。

本来 LDA は、トピックモデルにおけるアルゴリズムである。文書を構成する単語集合から、各文書におけるトピックを推定するものである。今回は単語ではなく、広告のキャンペーン ID や位置などの特徴を用いた。なお、トピック数の推定は HDP[10]を用いた。

5. 実験と評価

前項で提案した手法に対する評価実験を行う。5.1 では実験データ概要、5.2 では評価方法を示す。5.3 ではベースライン手法に対する評価、5.4 では提案手法に対する評価を行う。

5.1 実験データ

実験では用いる特徴として、広告の位置・キャンペーン ID・バナーID・サイト ID・ユーザ ID を用いる。ただし、ユーザ ID はクッキー等により追跡可能であったユーザを示しており、必ずしも 1 ユーザが 1 ユーザ ID になるとは限らない。実験期間としては、2016 年 11 月 01 日から 2016 年 11 月 21 日までの 21 日間をモデル作成の期間とする。そして、2016 年 11 月 22 日から 2016 年 11 月 28 日までの 7 日間をテスト期間として、作成したモデルを用いて広告の配置位置を推定する。また、学習・テスト期間の各期間における用いる特徴の統計情報は表 2 に示す。なお、モデル作成期間とテスト期間で出現する広告は異なる。よって、実験時にはモデル作成期間とテスト期間の両期間に出現した広告のみを評価対象とする。

5.2 評価方法

評価実験では、各広告におけるユーザの多クラス分類結果を特徴として用いることで、各広告の WEB ページへの配置位置推定への有効性を確認し、各広告における最高 CTR となる位置(4 値)を推定し評価する。評価方法についての条件を以下に記す。

- 位置を示す 4 値はヘッダー・記事下・サイドバー・フッターである。また 4 値が示す実際の WEB ページ上における位置は、WEB ページごとに詳細な広告の位置は異なる。
- 特定位置のみに広告を掲出可能な WEB ページを除外するため、全位置に広告を掲出可能な WEB ページを対象とした。さらに、クリック数が上位 10 位以内となる WEB サイトに属するページを対象とした。対象となる WEB ページの情報を表 3 に示す。
- ベースラインとなる手法として、(i)WEB ページ別広告配置手法、(ii)広告別広告配置手法、(iii)ランダム配置手法、(iv)広告・WEB ページ別配置手法を用いる。テスト期間のデータに基づき、各手法により推定された位置に広告を掲載した場合に得られる CTR により評価した。位置の推定方法は、WEB ページ毎(i)、広告毎(ii)、または WEB ページ・広告毎(iii)(iv)に集約し、位置ごとの CTR

を算出し、最大 CTR となる位置を求める。求めた位置を、(i)では WEB ページ毎に掲載される全ての広告に適用し、(ii)では広告毎に掲載する全ての WEB ページに適用する。最後に(iii)と(iv)では、広告と WEB ページ毎に適用する。以上から各手法を用いて、各広告に対する広告掲載 WEB ページ毎の最適な位置を決定し、評価する。

なお、1つの WEB ページ上に存在する各広告枠は同一のゾーンカテゴリとは限らない。ゆえに、広告枠に割り当てられたゾーンカテゴリ番号(最大 2 種類)を表 3 の最終列に示す。なお、ゾーンカテゴリが 1 種類しか存在しない場合は、2 種類目のゾーンカテゴリを“—”と表している。

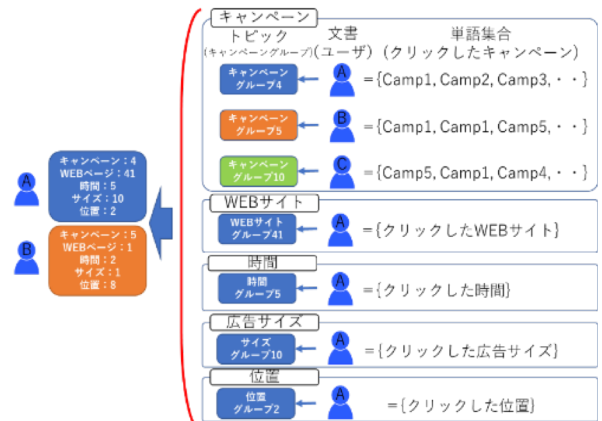


図 8 ユーザ分類の方法

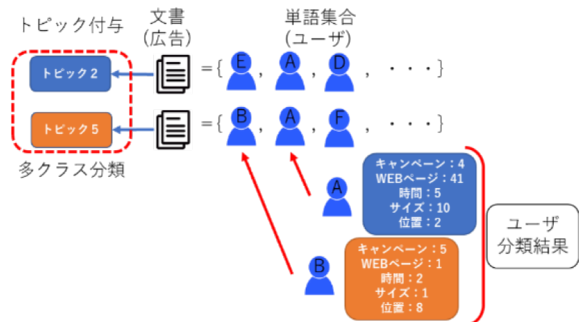


図 9 広告分類の方法

表 2 実験データ

期間	インプレッション [回]	クリック [回]	ユーザ [人]	キャンペーン [個]	バナー [個]	サイト [個]
モデル作成	1,077,727,483	1,574,635	34,141,120	445	9,440	4,013
テスト	37,303,721	95,969	15,113,842	313	5,284	3,449

表 3 対象となる WEB サイトにおける統計情報

サイト ID	クリック数 [回]	掲載されたユニーク広告数 [個]	位置別クリック数 [回]				広告枠に当てられたゾーンカテゴリ	
			ヘッダー	記事下	サイドバー	フッター	1	2
A	221,278	1,098	25,020	182,828	13,358	72	1	—
B	122,020	1,004	2,895	59,571	57,860	1,694	9	3
C	90,193	1,012	21,437	46	68,705	5	22	—
D	11,628	688	37	257	7,049	4,285	3	—
E	8,831	1,154	4,522	168	308	3,833	22	3
F	8,285	885	4,573	2,239	113	1,360	22	—
G	7,626	738	4	7,408	49	165	1	—
H	3,952	773	19	3,147	373	413	1	—
I	2,831	594	121	2,621	26	63	3	—
J	2,615	582	46	1,988	40	541	1	—

### 5.3 ベースライン手法の実験結果

4つのベースライン手法における結果を表4, 表5, 表6, 表7に示した. 表4におけるWEBページにおける掲出可能な位置数が4ヶ所の場合, 表5, 表6, 表7における広告掲出位置数4ヶ所の場合の4つに対して, CTRを比較する. その結果, 広告情報に比べ被広告掲載WEBページ情報を用いることで位置推定の精度が向上していることがわかる. また, 広告やWEBページの情報を使わず, ランダムによる配置位置決定の精度は, WEBページ単体に劣ることが分かった. ランダムによる配置位置決定が性能的に劣る理由として図1に示したようなWEBページにおける配置位置別のCTRは均一ではなく, ばらつきがあるためだと考えられる.

### 5.4 提案手法の実験結果と評価

#### 5.4.1 ユーザ分類

ユーザのクリック情報からユーザ分類を行った. テスト期間においてクリック行動を行なったユーザは717,352人であった. LDAにおけるトピック数の決定にはHDPを用いて決定し, HDPにより推定されたトピック数を表8に示す. 表8より, キャンペーンとサイトはID数が多いため, トピック数が多い. 一方, 広告サイズは12種類, 位置は4種類と種類数が少ないため, トピック数は少ない. 時間は1週間で28分割しているが, 得られたトピック数が非常に多くなっており, 時間においてユーザのクリック行動パターンが大きく変化する多様性を持っていることが分かる.

#### 5.4.2 広告の多クラス分類

広告をクリックしたユーザ集合を用いた, 広告の多クラス分類をした. 対象となる広告数は4,088個であった. ユーザの特徴として, 前項で用いたユーザ分類を用いる. なお, 5つの分類結果を組み合わせることで1つのクラスとした. HDPを用いたLDAにおけるトピック数の推定結果を表9に示す.

#### 5.4.3 広告配置位置の推定

5.4.2における広告の多クラス分類の結果を用いて, WEBページにおける広告配置位置の推定を行った. 表3に示したWEBサイトを対象に, 各広告を配置する際に最大CTRとなる位置を各広告の掲出位置数別に推定を行った. 表10には掲出位置数2~4ヶ所の広告を提案手法を用いて配置位置を推定した際の結果を示す. 表7と表10から, 提案手法において広告のWEBページにおける配置位置の推定に際し, ユーザ分類を用いた広告分類を行うことは広告・WEBページ別配置手法と比較し, ほぼ同等の結果が得られた. 図10は比較手法と提案手法における位置推定の結果をまとめた. 図10より, 広告またはWEBページ単体の情報に比べ, 広告情報と被広告掲載ページ情報を組み合わせることで位置推定の精度が向上する.

### 6. おわりに

本研究ではオンライン広告を配信する一般的な広告配信サービスを対象として, CTR向上を目的とした被掲載WEBページにおける最適な広告配置位置の推定を行う手法を提案した. 一般的な広告配信サービスが用いることができないユーザの詳細情報を用いないことを前提とした. 広告CTRが最大となる広告の配置を行い, CTRの向上を

表4 WEBページ別広告配置手法の推定結果

	広告の掲出位置数(広告枠数)		
	2ヶ所	3ヶ所	4ヶ所
インプレッション数	62,732,115	103,341,363	76,757,936
クリック数	136,180	133,393	67,802
CTR [%]	0.31	0.4	0.42

表5 広告別広告配置手法の推定結果

	WEBページ上に掲出可能な位置数		
	2ヶ所	3ヶ所	4ヶ所
インプレッション数	10,526,973	15,954,429	262,098,499
クリック数	6,346	9,463	340,189
CTR [%]	0.06	0.06	0.13

表6 ランダム配置手法による結果

	広告の掲出位置数(広告枠数)		
	2ヶ所	3ヶ所	4ヶ所
インプレッション数	62,579	2,163,227	21,406,032
クリック数	32	1,729	27,827
CTR [%]	0.05	0.08	0.13

表7 広告・WEBページ別配置手法

	広告の掲出位置数(広告枠数)		
	2ヶ所	3ヶ所	4ヶ所
インプレッション数	62,579	2,163,227	21,406,032
クリック数	98	1,958	98,646
CTR [%]	0.16	0.09	0.46

表8 ユーザ分類時のトピック数の推定結果

ユーザ分類に用いる特徴	キャンペーン	サイト	時間	広告サイズ	位置
推定されたトピック数[個]	169	379	167	13	12

表9 広告分類時のトピック数の推定結果

広告分類に用いる特徴	キャンペーン	サイト	時間	広告サイズ	位置
推定されたトピック数[個]	95	56	56	14	23

表10 提案手法による位置推定結果

	広告の掲出位置数(広告枠数)		
	2ヶ所	3ヶ所	4ヶ所
インプレッション数	62,579	2,163,227	21,406,032
クリック数	98	1,958	98,700
CTR [%]	0.16	0.09	0.46

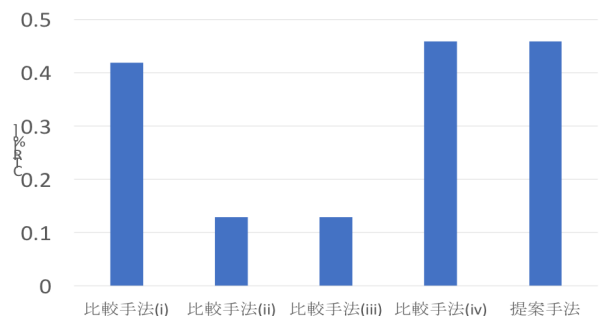


図10 各手法における位置推定の結果  
(広告の掲載位置数4ヶ所)

本研究の目的とした。提案手法では、各ユーザのクリックした広告内容、広告の被掲載 WEB ページ及びクリックした時間と位置に基づいて、ユーザ分類を行う。次に各広告におけるクリックしたユーザの集合を元に、広告を分類する。広告の分類結果を用いることにより、複数の広告枠が定められた WEB ページにおいて、広告配置する際に、最大 CTR となる広告枠の位置を推定した。広告・WEB ページ別配置手法(ナイーブ手法)と比較し、さらにユーザ情報を用いた分類では CTR が 0.46%と同等の実験結果を得られた。本研究を通して、広告配信における様々な特徴と CTR との相関関係を確認した。

本研究を通して、今後の研究として 2 点挙げられる。(i) 広告クリック情報だけでなく、精度向上のために広告インプレッション情報を用いた手法の提案。(ii)位置を示す特徴として、WEB ページを解析し絶対的な広告枠の位置を算出し、手法に適用すること。が挙げられる。以上をはじめとして多くの追加研究課題が考えられる。

#### 参考文献

- [1] Y. Chen and T. W. Yan, "Position-normalized click prediction in search advertising," *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD*, pp. 795–803, 2012.
- [2] S. Schroedl, A. Kesari, and L. Neumeier, "Personalized Ad Placement in WEB Search," *Proceedings of the 4th Annual International Workshop on Data Mining and Audience Intelligence for Online Advertising, ADKDD*, 2010.
- [3] Y. Yuan, D. Zeng, H. Zhao, and L. Li, "Analyzing positioning strategies in sponsored search auctions under CTR-based quality scoring," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 4, pp. 688–701, 2015.
- [4] I. Koutsopoulos, "Optimal Advertisement Allocation in Online Social Media Feeds," *Proceedings of the 8th ACM International Workshop on Hot Topics in Planet-scale mObile Computing and Online Social neTworking*, pp. 43–48, 2016.
- [5] H. Cheng, E. Manavoglu, Y. Cui, R. Zhang, and J. Mao, "Dynamic ad layout revenue optimization for display advertising," *In Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy, ADKDD*, pp. 1–9, 2012.
- [6] J. H. Friedman, "GREEDY FUNCTION APPROXIMATION:A GRADIENT BOOSTING MACHINE," *Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [7] A. P. Dempster, N. M. Laird, D. B. Rubin, P. Dempster, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, pp. 1–38, 1977.
- [8] IAB, "Networks & Exchanges Quality Assurance Guidelines v1.5," 2011. [Online]. Available: <http://www.iab.net/media/file/IAB-NE-QA-Guidelines-v1.5-November-2011-FINAL.pdf>.
- [9] David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, pp. 993–1022, 2003.
- [10] C. Wang, J. Paisley, and D. M. Blei, "Online Variational Inference for the Hierarchical Dirichlet Process," *Icais*, vol. 15, pp. 752–760, 2011.