

コンテキスト広告における クリックログを用いた効率的な広告引き当て手法

Efficient Retrieval for Contextual Advertising Utilizing Past Click Logs

田頭 幸浩
Yukihiro Tagami

ヤフー株式会社
Yahoo Japan Corporation
yutagami@yahoo-corp.jp

堀田 徹
Toru Hotta

(同 上)
thotta@yahoo-corp.jp

田中 祐介
Yusuke Tanaka

(同 上)
yuustana@yahoo-corp.jp

小野 真吾
Shingo Ono

(同 上)
shiono@yahoo-corp.jp

塚本 浩司
Koji Tsukamoto

(同 上)
kotsukam@yahoo-corp.jp

田島 玲
Akira Tajima

(同 上)
atajima@yahoo-corp.jp

keywords: contextual advertising, learning-to-rank, click feedback, modeling.

Summary

Contextual advertising is a form of textual advertising usually displayed on third party Web pages. One of the main problems with contextual advertising is determining how to select ads that are relevant to the page content and/or the user information in order to achieve both effective advertising and a positive user experience. In this study, we propose a translation method that learns the mapping of the contextual information to the textual features of ads by using past click data. The contextual information includes the user's demographic information and behavioral information as well as page content information. The proposed method is able to retrieve more preferable ads while maintaining the sparsity of the inverted index and the performance of the ad retrieval system. In addition, it is easy to implement and there is no need to modify an existing ad retrieval system. Extensive evaluations showed the effectiveness of our approach.

1. はじめに

オンライン広告は、昨今のインターネットのエコシステムを支える重要な役割を担っており、日本国内の市場に限定しても年間数兆円規模まで成長している。オンライン広告の例として、検索連動型広告やコンテキスト広告、ディスプレイ広告などがあげられる [Yuan 12]。本稿では、その中でもコンテキスト広告に着目する。コンテキスト広告は、主にニュースサイトやブログなどに表示される、短いテキストメッセージで構成される形態の広告である。広告を出稿する広告主は、広告効果を高めるためにも、その広告に関連するユーザーに対して広告を表示したいと考え、自らのウェブサイトには広告を表示するパブリッシャーは、広告によるユーザー体験の損失を最小限に抑えるように気を遣う。これらの2つの目的を満たすため、アドネットワークは、広告を表示するウェブ

ページやユーザーに関連した広告を選択しようとする。本稿では、広告を選択する際の指標として、CTR (Click-Through Rate; クリック率) に着目する。CTR はユーザー満足度やパブリッシャーの収益、広告効果に密接に関わる指標である。

典型的なアプローチでは、広告とウェブページの関連性をそれらのテキスト内の単語の共起度合いによって測り、ウェブページにとって関連性の高い広告を選択する。より具体的には、広告およびウェブページに対して tf-idf によって重みづけられた単語ベクトルを算出し、ベクトル間の類似度が高い広告を関連性が高いとみなし、スコア上位の広告をユーザーに対して提示する。そのため、このタスクは、情報検索分野でよく研究されている、高次元かつスパースな空間上での類似検索として捉えられる。このタスクに対して、高次元空間上のスパース性を考慮し、転置インデックスを用いた手法を用いることで、

広告配信に求められるシビアな応答性能を満たすような高速な類似検索が可能になる。しかしながら、このように単語ベクトルの空間上で広告とウェブページの関連性を測るアプローチは、広告とウェブページ間で単語の共起度合いが高い時は効果的である一方、広告で用いられる単語とウェブページで用いられる語彙が異なる場合には、うまくいかない。

上記の問題を解決するため、先行研究ではマッチング関数の中に、階層化されたカテゴリ構造の導入 [Broder 07] や、隠れクラスを用いた確率モデル [Ratnaparkhi 08] を用いるなどの提案が行われてきた。しかしながら、求められる応答性能の中でこれらのアプローチを適用するためには、上述した転置インデックスを用いた既存の広告検索システムを拡張したり、カテゴリやクラスを扱えるように新たな検索インデックスを作成する必要がある。加えて、広告配信システムの運用のためには、カテゴリ数や階層構造の見直しやクラスタの定期的な作り直しが不可欠であり、システムの規模が大きくなるにつれ容易なことではなくなる。

上記の問題を解決するため、我々は、既存の広告検索システムを変更することなく、転置インデックスを用いて2つの単語ベクトル間のマッチング関数を計算するアプローチを開発した。言い換えると、このアプローチでは、ウェブページやユーザーの情報を含むコンテキスト情報を、広告のテキスト空間に変換するものである。コンテキスト情報の特徴ベクトルは、変換行列により、広告検索システムの入力単語ベクトルに変換される。このプロセスを図1に示す。

一般的に、コンテキスト情報の特徴量空間と広告のテキスト空間はともに巨大であるため、この変換行列も巨大なものになってしまう。行列の低ランク近似を用い、過去のクリックデータからこの変換行列を学習することは可能かもしれない。しかしながら、学習を効率的にできたとしても、変換後のベクトルをできるだけスパースにする必要がある。なぜなら、転置インデックスと枝刈りアルゴリズムにより構成される広告検索システムは、高次元かつスパースな空間に対して性能を発揮されるように設計されているため、入力ベクトル内の非ゼロ要素の数が増えるにつれて応答性能が悪化してしまうからである [Ding 11, Fontoura 11, Strohmman 07]。そのため、我々は過去の CTR を元にした基準で、コンテキスト情報に関連した広告の単語特徴量を選択し、その後変換行列を学習する。言い換えると、まず行列の要素の部分集合を選び、その要素のみを過去のクリックログを用いて学習することで、広告検索システムの応答性能を悪化させることなく、CTR がより高い広告を効率的に検索する。

本稿の構成は以下の通りである。2 章ではコンテキスト広告のシステムについての概観を示す。3 章ではコンテキスト情報を広告の単語ベクトル空間に変換する提案手法について述べる。4 章では実験設定と結果の詳細に

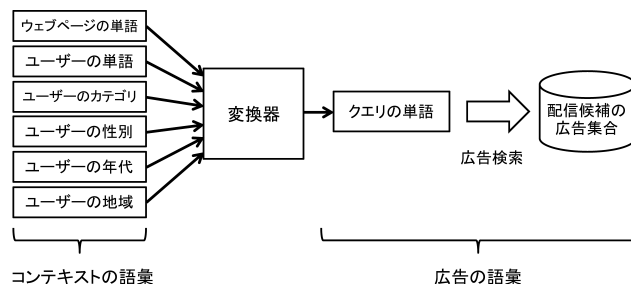


図1 提案手法の概要図。コンテキスト情報の特徴量を、広告検索システムの入力単語ベクトルに変換する。

について述べる。最後に、5 章で我々の知見をまとめる。

2. 概要および関連研究

この章では、コンテキスト広告の概要と関連研究について述べる。

2.1 コンテキスト広告の概要

コンテキスト広告にはパブリッシャー、広告主、アドネットワーク、ユーザーという4種類のひとびとが関わっている。パブリッシャーはウェブサイトの運営者であり、ウェブページの一部を広告のために確保している。広告主はアドネットワークに対して、広告を出稿する。パブリッシャーはウェブページにユーザーが訪れるたびに、アドネットワークに対して広告をリクエストする。広告リクエストを受けたアドネットワークは、配信候補の広告の中から適切な広告を選択し、パブリッシャーに返却する。パブリッシャーは、ウェブページ上の確保していた領域に返却された広告を表示し、ユーザーはそれを目にすることになる。アドネットワークによる広告の返却が遅れると、ウェブページの一部が表示されずにユーザー体験を損なうため、広告配信はシビアな応答性能（数十ミリ秒から数百ミリ秒程度）が求められる。コンテキスト広告でよく見られる、クリック報酬型（pay-per-click）のモデルでは、ユーザーが広告をクリックもしくはタップして広告主のウェブサイトを訪れた時のみ、広告主はパブリッシャーとアドネットワークに料金を支払う。そのため、各広告が表示された時に期待される収益額は、入札額と CTR の関数で表現される。入札額とは、広告がクリックされるたびに支払ってもよいと広告主が考える額であり、これは広告を出稿する際に設定する。そのため、アドネットワークの広告配信システム上はこの額は既知である。一方、各広告の CTR はウェブページやユーザーなどのさまざまな要素（コンテキスト）によって大きく変化しうる。それゆえ、限られた時間内で、いかに CTR を正しく予測し、CTR が高い広告をデータベースから効率的に検索するかは、コンテキスト広告の大きな問題の一つである。

近年では、広告主に代わり広告効果を最適化する demand-side platform (DSP) や、パブリッシャーに代わり収益の最大化を行う supply-side platform (SSP) もよく見られるようになった。これらはともに、複数のアドネットワークとやりとりすることによって、広告効果や収益を向上させることを目的としている。また、広告入稿時に入札額を決定するのではなく、それぞれの広告リクエストごとに入札額をユーザーなどの情報に基づいてリアルタイムで決定する、real time bidding (RTB) と呼ばれる仕組みも広く浸透している。これらを含む場合においても、本稿で示す提案手法を適用可能である。

2.2 関連研究

一般的な広告を表示するより、ウェブページの内容にあった広告を表示することで、より良いユーザー体験を提供することができると考えられるため、ページに広告をマッチさせる手法を提案した先行研究はこれまでいくつかなされてきた。これにより、ユーザーの現時点での興味に適合した広告を提示できるので、ユーザーが広告をクリックする確率が向上すると考えられる。これらの研究では、ウェブページに広告をマッチさせる問題を、ベクトル空間上での類似検索として扱っていた。ある広告とウェブページ間の関連性は、それらのコンテンツに含まれる単語の重なりを元に、tf-idf スコアによって表現される。Chakrabarti らや、Karimzadehgan らは、ウェブページと広告に含まれる単語の重みを、ウェブページの HTML タグや広告のセクションに基づき、学習する手法を提案した [Chakrabarti 08, Karimzadehgan 11]。これらの手法では、高次元空間上でのスパース性を考慮し、転置インデックスを用いることで、広告集合からスコア上位 K 件の広告を高速に取得できる。

ウェブページと広告を同じ空間にマッピングする際に、それぞれで用いられる語彙が異なるため、関連性を適切に測ることが難しいという問題がある。この問題を解決するために、様々なアプローチが提案されてきた。Broder らはウェブページと広告間のマッチング関数に、6000 ノードからなる階層的な意味カテゴリを導入した [Broder 07]。また、Ratnaparkhi は、ページの単語と広告の単語間の意味的な関連性を、隠れクラスを用いた確率モデルで表現した [Ratnaparkhi 08]。

Joshi らはコンテンツ連動型広告において、ユーザーのデモグラフィック情報やウェブ上での行動情報を用いてマッチングの改善を行う手法を示した [Joshi 11]。この手法では、ユーザーに紐づく非テキスト情報を、広告のテキスト情報にマッピングしている。

異なる研究の方向性として、広告の CTR を予測することが挙げられる。これらの研究では、コンテキスト広告だけでなく、同様にクリック報酬型の課金モデルを採用している検索連動型広告も対象としている。広告の CTR 予測は、一般的に過去のクリックデータから学習した統計モデ

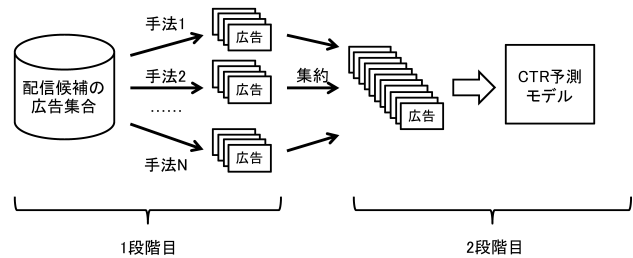


図2 広告配信システムにおける2段階のアプローチ。1段階目では複数の方法により広告が検索される。2段階目では、1段階目で得られた広告に対して CTR の予測を行う。

ルを用いて行われる。これらのモデルの例として、ロジスティック回帰 [Chapelle 14, McMahan 13] やプロビット回帰 [Graepel 10]、ブースティング木 [Dave 10, Trofimov 12]、field-aware factorization machines [Juan 16]、それらの組み合わせ [He 14, Ling 17] が挙げられる。

2.3 広告配信システム

2.2 節で述べたように、機械学習によって得られた CTR 予測モデルはより精緻な予測を行うことができるが、ウェブページの表示を高速に行うという目的のもとでは、数十万から数百万の、候補集合内のすべての広告に対して予測を行うには時間がかかりすぎてしまう [Agarwal 12]。そのため、我々の広告配信システムでは、先行研究 [Agarwal 12, Chakrabarti 08] と同様に、2段階のアプローチを採用している。まず、1段階目では、転置インデックスを用いて、候補集合の中から K 件の広告を取得する。2段階目では、その K 件の広告に対して CTR 予測モデルを適用し、 k 件 ($k \ll K$) の広告を選択する。1段階目においては、複数の手法により、並列かつ独立に広告が検索される。2段階目では、1段階目で得られた広告に対して CTR 予測を行う。その際、広告は予測 CTR と入札額に応じてランキングされ、表示される。このプロセスを図2に示す。本稿では、1段階目を「広告検索システム」と呼び、この部分に焦点を合わせる。我々は、WAND アルゴリズム [Broder 03, Ding 11, Fontoura 11] を採用した、最適化された検索エンジンを用いて広告の検索を行っている。

本稿では、先述したベクトル空間での類似検索で用いられる既存のインデックスを流用して、新規の検索インデックスを構築することなく、効率的に CTR が高い広告を検索する手法を提案する。単純に、コンテキスト情報と広告のマッチ度合いを測る目的に対しては、近年注目を浴びている、ニューラルネットワークなどにより得られた分散表現を用いる手法 [Covington 16, Huang 13] も有力な選択肢として考えられる。しかし、分散表現は一般的に、低次元かつデンスなベクトルであるため、転置インデックスと WAND アルゴリズムによる検索システムでは効率的な検索ができない。このような低次元かつ

デンスなベクトルを効率的に検索するためには、他の検索手法 [Liu 05] を用いる必要があり、これらの手法を大規模な実サービスに運用可能な水準で実装するのはコストが高い。そのため、本稿では、既存の検索システムで高速な検索が可能な、高次元かつスパースなベクトルを生成する手法に絞り、提案および比較を行う。

3. 引き当て手法

この章では、まずコンテキスト情報と広告間のマッチング関数を定義し、コンテキスト情報を広告のテキスト空間にマッピングする我々のアプローチについて述べる。なお、以下では、コンテキスト情報を、広告検索の入力になるという観点から、“クエリ”と呼び換えて説明する。

3.1 マッチング関数

クエリの特徴量ベクトルを $\mathbf{q} = (q_1, \dots, q_{D_q})^T$ 、広告の特徴量ベクトルを $\mathbf{a} = (a_1, \dots, a_{D_a})^T$ と定義する。クエリの特徴量ベクトル $\mathbf{q}$ は、広告リクエストに紐づくコンテキスト情報から構成され、ウェブページやユーザーの情報を含むものである。加えて、変換行列 $\mathbf{W} = [w_{ij}]_{D_q \times D_a}$ を用いて、 $\mathbf{q}$ と $\mathbf{a}$ に対するマッチング関数の一般的な形を以下のように定義する。

$$\text{mscore}(\mathbf{q}, \mathbf{a}) = \mathbf{q}^T \mathbf{W} \mathbf{a} = \sum_{i=1}^{D_q} \sum_{j=1}^{D_a} w_{ij} q_i a_j \quad (1)$$

ここで、 D_q と D_a はそれぞれ、クエリ特徴量と広告特徴量の次元数を表す。

式 (1) は、マッチング関数の一般形を表現している。もし、 D_a と D_q が同じ数であり、行列 $\mathbf{W}$ が単位行列 $\mathbf{I}$ の場合、マッチングスコア $\text{mscore}(\mathbf{q}, \mathbf{a})$ は単純な内積になる。もちろん、 $\mathbf{q}$ と $\mathbf{a}$ を共に正規化することで、コサイン類似度を計算することも可能である。加えて、一般的な場合では、クエリ特徴量ベクトル $\mathbf{q}$ に変換行列 $\mathbf{W}$ をかけ、 $\mathbf{q}' = \mathbf{W}^T \mathbf{q}$ とし、この $\mathbf{q}'$ を従来の転置インデックスを用いた広告検索システムの入力ベクトルとすることで、 $\text{mscore}(\mathbf{q}, \mathbf{a}) = \mathbf{q}^T \mathbf{W} \mathbf{a} = \mathbf{q}'^T \mathbf{a}$ を計算する。

3.2 変換行列の学習

1 章で述べたように、本稿では、転置インデックスを用いた広告検索を効率的に行うため、変換後の単語ベクトルをスパースにすることを目的とする。この目的を達成するためには、 $\mathbf{W}^T \mathbf{q}$ に対する L1 正則化項を目的関数に追加し、alternating direction method of multipliers (ADMM) により最適化を行うことが考えられる [Sprechmann 13]。しかし、このような正則化項を含む目的関数の最適化は比較的複雑なものになる。そのため本稿では、変換後の単語ベクトルがスパースになることを目指して、よりシンプルかつ直接的に、変換行列 $\mathbf{W}$ のスパース性を調整

する手法を提案する。この手法では、L1 正則化を用いる代わりに、まず行列の中から要素の部分集合を選択し、その後、それぞれの w_{ij} について学習を行う。

学習データ内に現れる、 q_i と a_j のペアに対して、以下のスコア m_{ij} を計算する。

$$m_{ij} = \frac{\text{CTR}(q_i, a_j)}{\max(\text{CTR}(q_i), \text{CTR}(a_j))}$$

ここで、 $\text{CTR}(q_i)$ はクエリ特徴量ベクトルが q_i を含む (q_i の値が非ゼロの) 場合の CTR を表し、 $\text{CTR}(a_j)$ は a_j を含む広告の CTR を表す。同様に、 $\text{CTR}(q_i, a_j)$ はクエリ特徴量ベクトルに q_i を含んでいる時の、 a_j を含む広告の CTR を表現している。 m_{ij} が大きい場合は、クエリ特徴量ベクトルが q_i を含む時に、 a_j を含む広告がよりクリックされやすいことを意味している。この m_{ij} の値が大きなペアの集合を閾値 T を用いて以下のように選択する。

$$P = \{(i, j) \mid m_{ij} > T\},$$

閾値 T を大きくするに従って、 P 内のペアの数は減少する。この P を用いて、マッチング関数 (式 (1)) を以下のように置き換える。

$$\text{mscore}(\mathbf{q}, \mathbf{a}) = \sum_{(i,j) \in P} w_{ij} q_i a_j$$

もし、相互情報量や L1 正則化を用いた従来の特徴量選択を行った場合、正の値をとる w_{ij} と同じく、負の値をとる w_{ij} が選択されることが予想される。しかし、2.3 節で述べたように、我々の目的は、 $\text{mscore}(\mathbf{q}, \mathbf{a})$ が大きな上位 k 件の広告を、広告集合から取得することである。この時、負の値をとる w_{ij} を用いても、 $\text{mscore}(\mathbf{q}, \mathbf{a})$ が高い広告を取得することはできない。そのため、正の値をとる w_{ij} が選択されることを期待し、上記のアプローチを採用した。

上の w_{ij} を過去のクリックデータを用いて学習する。2 つの特徴量ベクトル $\mathbf{q}$ と $\mathbf{a}$ を用いて、CTR に関連したスコアを、以下のような線形の式で定義する。

$$\begin{aligned} \text{score}(\mathbf{q}, \mathbf{a}) &= \text{mscore}(\mathbf{q}, \mathbf{a}) + \text{bscore}(\mathbf{q}, \mathbf{a}) \\ &= \sum_{(i,j) \in P} w_{ij} q_i a_j + \mathbf{w}_{\text{basic}}^T \mathbf{x}_{\text{basic}} \\ &= \mathbf{w}_{\text{match}}^T \mathbf{x}_{\text{match}} + \mathbf{w}_{\text{basic}}^T \mathbf{x}_{\text{basic}} \\ &= \mathbf{w}^T \mathbf{x} \end{aligned} \quad (2)$$

ここで、 $\text{bscore}(\mathbf{q}, \mathbf{a}) = \mathbf{w}_{\text{basic}}^T \mathbf{x}_{\text{basic}}$ はベーシックスコアである。 $\mathbf{x}_{\text{basic}}$ は広告自身のクリックされやすさなどの特徴量を含むベクトルであり、 $\mathbf{w}_{\text{basic}}$ はそれに対応する重みベクトルである。 $(i, j) \in P$ に対し、 $\mathbf{w}_{\text{match}}$ と $\mathbf{x}_{\text{match}}$ をそれぞれ、 w_{ij} と $q_i a_j$ をその要素として構成されるベクトルとして定義する。 $\mathbf{w}$ は $\mathbf{w}_{\text{match}}$ と $\mathbf{w}_{\text{basic}}$ を連結したベクトル、 $\mathbf{x}$ は $\mathbf{x}_{\text{match}}$ と $\mathbf{x}_{\text{basic}}$ を連結したベクトルで

ある。CTR に比例したスコアは、クエリと広告間のマッチングスコアだけでなく、性別や年代などによるユーザーのクリックしやすさの傾向や、広告自身のクリックされやすさ、ウェブページ上の表示位置などにも依存しているため、 $\text{mscore}(\mathbf{q}, \mathbf{a})$ だけでなく $\text{bscore}(\mathbf{q}, \mathbf{a})$ も考慮する必要がある。

2.3 節で述べたように、本稿では、2 段階からなる広告配信システムの 1 段階目を改善することに注力する。2 段階目での上位 k 件の結果を良くするためには、1 段階目で取得する上位 K 件を改善する必要がある。そこで、 w_{ij} として、単純に上記の m_{ij} を用いるのではなく、クリックログから学習を行う。

次に、重みの学習において、クリックログをどのように用いるかについて述べる。コンテキスト広告の設定では典型的に、ウェブサイトのオーナーであるパブリッシャーは、1 つのページに 1 つ以上の広告を表示するため、同時に複数の広告をアドネットワークにリクエストする。言い換えると、ログ中の 1 つの広告リクエスト r は、 $N^{(r)}$ 個の広告インプレッション（広告表示および、その回数）を含む。

$$(\mathbf{q}^{(r)}, \mathbf{a}_1^{(r)}, y_1^{(r)}), (\mathbf{q}^{(r)}, \mathbf{a}_2^{(r)}, y_2^{(r)}), \dots, (\mathbf{q}^{(r)}, \mathbf{a}_{N^{(r)}}^{(r)}, y_{N^{(r)}}^{(r)})$$

各広告のインプレッション情報は、3 つ組 $(\mathbf{q}^{(r)}, \mathbf{a}_i^{(r)}, y_i^{(r)})$ からなる。ここで、 $\mathbf{q}^{(r)}$ はその広告リクエスト r に対応するクエリ特徴量ベクトルを表し、 $\mathbf{a}_i^{(r)}$ は広告リクエスト r 中の i 番目のインプレッションに対応する広告の特徴量ベクトルを表す。出力変数 $y_i^{(r)}$ は、ユーザーがその広告をクリックすれば 1 を、そうでなければ 0 をとるものとする。

本稿では先行研究 [Tagami 13] と同様に、一部の広告リクエストに着目し、ランキング学習の手法を用いて重みの学習を行う。

データ R 中には、以下のように 2 種類の広告リクエストが存在する。

$$R^+ = \{r \mid \exists i (y_i^{(r)} = 1)\}$$

$$R^- = \{r \mid \forall i (y_i^{(r)} = 0)\}$$

R^+ は、クリックされた広告インプレッションを 1 つ以上含む広告リクエストの集合を表す。そこで、この R^+ を**クリックドリクエスト**と呼ぶ。一方、 R^- はクリックが 1 つもなかった広告リクエストの集合であり、**ノンクリックドリクエスト**と呼ぶ。もちろん、 $R^+ \cup R^- = R$ かつ $R^+ \cap R^- = \emptyset$ という関係が成り立つ。

ここでは、広告リクエスト内の 1 つの広告インプレッションを、情報検索の文脈における 1 つの文書とみなし、ランク学習の手法を用いて重みを学習する。各クリックドリクエストから、以下のように、広告のペアを作成する。

$$\{(\mathbf{a}_i^{(r)}, \mathbf{a}_j^{(r)}) \mid \forall i, j (y_i^{(r)} = 1 \wedge y_j^{(r)} = 0)\}$$

上記の広告ペア $(\mathbf{a}_i^{(r)}, \mathbf{a}_j^{(r)})$ に対して、クエリ特徴量ベクトル $\mathbf{q}^{(r)}$ に対する広告特徴量ベクトル $\mathbf{a}_i^{(r)}$ への CTR に比例したスコアは、 $\mathbf{a}_j^{(r)}$ よりも高くなることが期待される。このスコアの大小関係を、 $\text{score}(\mathbf{q}, \mathbf{a})$ を用いて表現し、式 (2) を用いて変形すると以下ようになる。

$$\begin{aligned} \text{score}(\mathbf{q}^{(r)}, \mathbf{a}_i^{(r)}) &> \text{score}(\mathbf{q}^{(r)}, \mathbf{a}_j^{(r)}) \\ \Leftrightarrow \mathbf{w}^T \mathbf{x}_i^{(r)} &> \mathbf{w}^T \mathbf{x}_j^{(r)} \\ \Leftrightarrow \mathbf{w}^T (\mathbf{x}_i^{(r)} - \mathbf{x}_j^{(r)}) &> 0 \end{aligned} \quad (3)$$

二乗ヒンジロスを用いて、RankSVM [Joachims 02] のようにペアワイズの損失関数 $L(\mathbf{w})$ を定義する。

$$L(\mathbf{w}) = \sum_{r \in R^+} \sum_{i: y_i^{(r)}=1} \sum_{j: y_j^{(r)}=0} \max(0, 1 - \mathbf{w}^T (\mathbf{x}_i^{(r)} - \mathbf{x}_j^{(r)}))^2$$

正則化項を足し、以下の最小化問題を解くことで、重みベクトル $\hat{\mathbf{w}}$ を求める。

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \cdot L(\mathbf{w})$$

ここで、 $C > 0$ は損失パラメーターである。学習した重みベクトル $\hat{\mathbf{w}} = [\hat{\mathbf{w}}_{\text{match}}^T, \hat{\mathbf{w}}_{\text{basic}}^T]^T$ 中の、 $\hat{\mathbf{w}}_{\text{match}}$ の各要素の値から、変換行列 $\mathbf{W}$ を復元する。ただし、 $(i, j) \notin P$ に対しては $w_{ij} = 0$ とする。

3.3 広告コーパスからの引き当てと応答性能

学習した変換行列 $\mathbf{W}$ を用いて、それぞれの広告リクエストのクエリ情報に対して、広告検索システムの入力ベクトルを $\mathbf{q}_{\text{input}} = \mathbf{W}^T \mathbf{q}$ と変換して作成する。提案手法では、この変換のみが必要となる。それゆえ、既存の転置インデックスを変更したり、新しいインデックスを構築する必要はなく、高速な検索が可能であり、実装も容易である。

この変換後の入力単語ベクトルは、変換前のクエリ特徴量ベクトルの非ゼロ要素の数に応じて、いくつかの非ゼロ要素を含む。しかしながら、この入力ベクトルの非ゼロ要素の数が増えるに従い、広告検索システムの応答性能も低下する [Fontoura 11]。そのため、ハイパーパラメーター M を用いて、以下のように非ゼロ要素の数を制限する必要がある。

$$\|\mathbf{q}_{\text{input}}\|_0 \leq M$$

$\|\mathbf{q}_{\text{input}}\|_0$ は入力ベクトル $\mathbf{q}_{\text{input}}$ の L_0 ノルムを表す。つまり、非ゼロ要素の数である。ここでは単純に、大きな値を持つ、上位 M 個の要素を選択する。

上記の変換行列 $\mathbf{W}$ を用いたベクトルの変換にかかる計算コストは、それを入力ベクトルとして行う検索自体の計算コストに比べて、非常に小さい。そのため、検索システムの応答性能は、主に入力ベクトル中の非ゼロ要

素の数 M に依存する。なお、 M を変化させた時の応答時間の増加率は、線形より大きい (superlinear) ことが知られている [Ding 11, Fontoura 11, Strohmaier 07]。3.2 節で導入した、行列のスパース性を調整する閾値 T は、応答性能にはそれほど影響しない。しかし、小さな M の値を採用する場合には、それに応じて T の値を大きくすることがよいと考えられる。なぜなら、 T の値を大きくすることで行列 $\mathbf{W}$ をスパースにすることができ、それによって $\mathbf{W}^T \mathbf{q}$ の非ゼロ要素を減少させ、 M で制限する際の影響を少なくできるからである。実システムへの導入時には、4.2 節にて後述するように、まず求められる応答性能を満たすように M を決定し、その後 M に対して最適な T を選択することで、ハイパーパラメータの値を決定する。

4. 実 験

この章では、オフラインとオンラインでの評価について述べる。オフライン評価では過去のクリックログを用いて提案手法の評価を行い、オンライン評価では実際の広告配信システムに提案手法を実装し、ユーザーの反応による評価を行った。ビジネス上の秘匿事項のため、結果を示す時は、絶対的な値ではなく相対的な値のみを報告する。

4.1 オフライン評価

本稿の目的は、既存の広告配信システムに大きな変更を加えることなく、広告の CTR を向上させる手法を開発することである。その検証のためには、実際の配信システムに提案手法を適用し、オンライン環境で A/B テストを行い、既存の配信ロジックとの比較を行えばよい。しかしながら、オンライン環境での評価を行う際には、実システムへの実装のコストやテスト実行によるユーザー満足度や広告売り上げの低下リスクがあるため、まずは比較的低コストで実行可能な過去のクリックログを用いたオフライン評価を行い、オンライン環境での既存手法に対する改善度合いや、ハイパーパラメータの変化の影響を見積もることとする。なお、提案手法の応答性能については、4.2 節にて後述するオンライン環境でのテスト時に担保するため、このオフライン評価の対象としない。

以下では、まずデータセットや特徴量、モデル、評価指標などの実験設定について述べる。その後、既存手法と提案手法の比較や、ハイパーパラメータ T や M を変化させた時のモデルのパフォーマンスを示す。

§1 データセットおよび特徴量

オフライン評価には、あるアドネットワークの 8 週間のログから抽出したデータを用いた。前 6 週間を学習データ、7 週目をバリデーションデータ、そして最終週をテストデータとして扱った。

前述したように、データ中の各サンプルは、 $(\mathbf{q}^{(r)}, \mathbf{a}_i^{(r)}, y_i^{(r)})$

の組からなる、広告インプレッション 1 つに対応する。ユーザーが広告をクリックすると、出力変数 $y_i^{(r)}$ は 1 となり、そうでなければ 0 となる。クエリ特徴量ベクトル $\mathbf{q}^{(r)}$ はウェブページとユーザーの情報を含む。ウェブページ情報として、ウェブページから抽出した単語を用いた。単語は、ウェブページ上の位置や HTML タグに基づいてスコア付けされ、そのスコアによっていくつかの単語が選択される。ユーザーの情報としては、性別、年代、地域に加え、ユーザーの興味を表す単語とカテゴリを用いる。ユーザーの興味を表す単語やカテゴリは、Aly らの研究 [Aly 12] で行われたように、ページ閲覧や検索クエリ、広告クリックなどのユーザーの行動履歴から抽出される。このカテゴリは、Broder らの研究 [Broder 07] で用いられたような、階層化されたもので、その数は約 900 である。広告の特徴量 $\mathbf{a}_i^{(r)}$ としては、広告のタイトルや詳細部 (ディスクリプション) の単語を tf-idf で重みづけたテキスト情報を用いる。これらの特徴量について、表 1 にまとめた。

3.2 節で述べた $\mathbf{x}_{\text{basic}}$ として、ウェブページ上での表示位置や、その広告や広告主の過去の実績 CTR などを用いた。先に述べたように、ユーザーの属性によっても CTR は変化するため、それらも $\mathbf{x}_{\text{basic}}$ に含むことが自然である。しかしながら、学習方法としてペアワイズの損失関数を採用したため、広告とは独立なクエリ情報量 $\mathbf{q}$ を $\mathbf{x}_{\text{basic}}$ に含めても、式 (3) でベクトルの差を取った際に打ち消しあってしまう。言い換えると、ある 1 つの広告リクエストに含まれる広告間の CTR の大小関係のみを考慮して学習を行っているので、リクエスト中の全ての広告に対して同じ量だけ CTR を変化させる、クエリに関する情報を $\mathbf{x}_{\text{basic}}$ に含める必要は結果としてなくなった。そのため、 $\mathbf{x}_{\text{basic}}$ には広告に関連する情報のみを含めた。これらの特徴量についても同じく表 1 にまとめている。

ウェブサイトによる効果の違いを見るため、ニュース、ブログ、Q&A、ファイナンス、スポーツ、天気、旅行を含む、8 つの多彩なウェブサイトを選択した。ウェブサイト内のページやそこに訪れるユーザーの傾向も異なるため、ウェブサイトごとにそれぞれ評価するモデルを構築した。各ウェブサイトのデータの統計量について表 2 にまとめた。なお、クリックドリクエストあたりのクリックされた広告数は、おおよそ 1 である。

§2 既存手法と評価指標

提案手法を、広告配信システムで用いられていた 3 つの既存手法 (既存手法 1, 2, 3) と比較した。これらの手法では、4.1 節 §1 と表 1 内に記した特徴量と同一の情報を用いている。既存手法 1 は、広告検索のためにウェブページ中の単語情報のみを用いた手法であり、Karimzadehgan らの手法 [Karimzadehgan 11] と同様の方法で最適化を行い、単語の重み付けを行っている。既存手法 2 は、ユーザーの行動履歴 (ウェブページ閲覧や検索キーワード、広

表 1 特徴量の概要

特徴量のタイプ	取得元	詳細
クエリ特徴量 $\mathbf{q}$	ウェブページ	ウェブページから抽出した単語
	ユーザー	行動履歴から抽出した単語, 行動履歴に基づくカテゴリ, 性別, 年代, 地域
広告特徴量 $\mathbf{a}$	広告	Tf-idf で重みづけられた単語
基礎特徴量 $\mathbf{x}_{\text{basic}}$	過去のクリックログ	広告および広告主単位の過去の CTR, ウェブページ上での表示位置

表 2 評価で用いるウェブサイトごとのデータの統計量

種別	A	B	C	D	E	F	G	H
学習	711,539	2,676,577	1,648,118	919,870	905,842	153,849	297,814	4,644,350
$ R^+ $ バリデーション	142,649	429,760	137,646	92,547	217,165	23,836	53,306	780,037
テスト	119,464	356,578	134,168	81,077	169,627	17,879	42,098	498,407
平均インプレッション数	7.00	4.99	4.42	4.10	4.33	5.00	8.12	4.46

告クリックなど) から単語情報を獲得し, それを広告検索のクエリ単語とした手法である. 既存手法 3 は, ユーザーの行動履歴からカテゴリ情報を獲得して用いた手法である. 既存手法 2 と 3 は, Aly らの手法 [Aly 12] を元に, 単語情報およびカテゴリ情報をユーザーに付与した.

2.3 節で述べたように, 本稿では実システムへの適用という観点から, 転置インデックスと WAND アルゴリズムを用いた既存の検索システムに適した, 高次元かつスパースなベクトルを作成する手法に絞って比較を行う. 上記の 3 つの既存手法は, クエリと広告を高次元かつスパースなベクトルとして表現することで, この制限を満たしている. 一方, ニューラルネットワークの学習などにより得られたベクトルは一般的に低次元かつデンスになり, これを用いた手法は, 既存の検索システムでは求められる応答性能を満たさないため, 今回の比較対象から外した.

上記の 3 つの既存手法のスコアとして, 式 (2) 中の $\text{score}(\mathbf{q}, \mathbf{a})$ の代わりに以下の $\text{score}_{\text{existing}}(\mathbf{q}, \mathbf{a})$ を用いた.

$$\text{score}_{\text{existing}}(\mathbf{q}, \mathbf{a}) = w \cdot \text{escore}(\mathbf{q}, \mathbf{a}) + \text{bscore}(\mathbf{q}, \mathbf{a})$$

ここで, $\text{escore}(\mathbf{q}, \mathbf{a})$ は既存手法によって計算されたマッチングスコアである. それぞれの手法は, 異なる $\text{escore}(\mathbf{q}, \mathbf{a})$ を持つ.

3.3 節で述べたように, 広告検索システムの応答速度に対する制限を満たすため, クエリ単語の数を制限する必要がある. 実験では, M の値を変化させて評価を行った. その際の予測時のスコア関数は以下ようになる.

$$t\text{-score}(\mathbf{q}, \mathbf{a}) = t\text{-mscore}(\mathbf{q}, \mathbf{a}) + \text{bscore}(\mathbf{q}, \mathbf{a}) \quad (4)$$

ただし, $t\text{-mscore}(\mathbf{q}, \mathbf{a})$ は切り捨てられたクエリベクトル $\mathbf{q}_{\text{input}}$ を用いたマッチング関数を表す. M を変化させてクエリ単語ベクトル $\mathbf{q}_{\text{input}}$ を切り詰めるのは, 学習後に行う. ここで注意すべきなのは, M を制限した時の評価は, 実際のオンラインでの設定をうまく反映できてい

ないかもしれないということである. オフライン評価においては $t\text{-score}(\mathbf{q}, \mathbf{a})$ を用いるが, 実際のオンラインの設定では $t\text{-mscore}(\mathbf{q}, \mathbf{a})$ を用いて広告の検索を行う. つまり, $\text{bscore}(\mathbf{q}, \mathbf{a})$ の影響により, 同一の広告リクエストに対する, $t\text{-mscore}(\mathbf{q}, \mathbf{a})$ と $t\text{-score}(\mathbf{q}, \mathbf{a})$ を用いた広告の順序は異なる場合がある. これは過去のクリックログを用いることによる制限である.

モデルの性能は, Mean Average Precision (MAP) [Manning 08] を用いて評価した. MAP の算出には, まず各クリックドリクエスト中に含まれる広告を, 予測スコア $\text{score}(\mathbf{q}^{(r)}, \mathbf{a}_j^{(r)})$ で並び替えて平均 precision を計算し, その後, 全クリックドリクエストに対して平均を取った. ビジネス上の制約のため, 本稿では, 学習およびバリデーション時に $\text{bscore}(\mathbf{q}, \mathbf{a})$ のみを用いたベーシックなモデルによる手法の評価値 $\text{MAP}_{\text{basic}}$ を用いて正規化した値のみ報告する. 評価指標の値は, 以下のように表される.

$$\Delta\text{MAP} = \left(\frac{\text{MAP}}{\text{MAP}_{\text{basic}}} - 1 \right) \times 100$$

ただし, $\mathbf{x}_{\text{basic}}$ に含まれる広告の表示ポジションは非常にクリックを予測するためには有用な特徴量であるため, $\text{MAP}_{\text{basic}}$ はそれなりに高い値となる. 実際の広告検索は数十万から数百万の配信候補全てを対象とするのに対し, この評価では過去の配信ロジックによって選択された少数の広告のみをスコア計算の対象としている点に注意されたい. これも過去のクリックログを用いて評価を行うことによる制限である.

§3 実験結果

まずはじめに, 閾値パラメーター T を変化させた時の, 提案手法について評価を行った. 3.2 節で述べたように, T を大きくするに従って, $\mathbf{W}$ 中の非ゼロ要素の数は減少する. 逆に, T を小さくすることにより, モデルの性能は向上すると期待される. この実験設定における評価では, M についての制限は設けない. 実験結果を表 3 に示す. **太字**の値は, 手法の中で一番良い結果を表している.

提案手法を用いることで、既存手法の場合よりも評価値が向上していることが分かる。予想された通り、MAP の値は T を減少させることにより改善している。学習データが多い、ウェブサイト B と H では、他のウェブサイトよりも ΔMAP が大きく、 T を変化させた時の影響も比較的小さい。ウェブサイト A においては、既存手法 1 の MAP は、 $T = 0.20$ や 0.15 の時よりも高くなっている。しかし、 T を減少させるにつれ、評価値の向上度合いは増している。この結果から、大量のデータで学習したモデルの性能の向上は、大きくかつ頑健であることが分かる。既存手法どうしを比較すると、それぞれのウェブサイトで得手不得手があることが分かる。ウェブサイト A, E, F では、ウェブページの情報をを用いた既存手法 1 がそのほかの 2 つよりも優れている。ウェブサイト B, C, D, G, H では、ユーザーの情報をを用いた既存手法 2 がより良い結果を示した。このことから、ウェブサイトによって、広告検索を行うのに有効な特徴量が異なることが分かる。提案手法は広告検索に対してウェブページとユーザー情報の両方を用いるため、これらの既存手法と比較して、より良い結果を示したのだと考えられる。

次に、 M の値を変化させた時の、それぞれの T に対応するモデルのパフォーマンスについて調査する。4.1 節 §2 で述べたように、 M を変化させることにより、評価時において、クエリ単語ベクトル $\mathbf{q}_{\text{input}}$ の要素を切り捨てる。例えば、 $M = 20$ の場合は、広告検索システムへの入力として、スコア上位の 20 単語のみを用いるということである。実験結果を図 3 に示す。こちらも予想通り、 M を減少させるに従い、MAP の値は低下している。また、ウェブサイト F における $M = 10$ の時のように、 M を減少させると、 T として小さい値を選択した方が高い評価値を示す場合があることも確認できた。この結果は、3.3 節で述べたように、 M に応じて T の値を調整する必要性を示唆している。ウェブサイト H における $T = 0.05$ かつ $M = 50$ の場合のように、 ΔMAP が負の値であることにより、提案手法がベーシックなモデルよりも悪化していると、もしかしたら考えるかもしれない。しかしながら、先に 4.1 節 §2 で述べたように、 $t\text{-score}(\mathbf{q}, \mathbf{a})$ と $t\text{-mscore}(\mathbf{q}, \mathbf{a})$ の違いのため、このオフライン評価は実際のオンラインでの性能をうまく反映できていないと考えられる。 M を制限した時の既存手法と提案手法の比較は、次の節でのオンライン評価に持ち越すことにする。

4.2 オンライン環境での評価

前節でのオフライン評価で改善の見込みが確認できたので、本節ではオンライン環境で提案手法の性能を測ることにする。

オンライン環境での提案手法の性能を測るため、提案手法を実際の広告配信システムに適用した。この広告配信システムは、2.3 節と図 2 に示したような、2 段階のアプローチを採用している。1 段階目の検索手法の 1 つとして提案

手法を加え、A/B テストによりオンラインでの性能を比較した。ハイパーパラメータとしては、($T = 0.20, M = 20$) を用いた。3.3 節で述べたように、システムの応答性能により M の値を決定し、その後、オフライン評価の結果と学習時間を考慮して T の値を決定した。システムの応答性能は、事前のシステムテストにより見積もり、A/B テストの対象である 2 つの配信ロジックの応答性能がほぼ同一（数十ミリ秒）になるようにした。同様に、公正な比較のため、1 段階目で取得する広告の総数も同じ数に揃えた。これは、取得する広告の数を単純に増加させるだけで、CTR が向上する可能性があるためである。2 段階目で用いる CTR 予測モデルは、どちらの場合においても、同じモデルを用いた。この予測モデルは、過去のクリックデータを用いて学習された統計モデルである。それぞれのウェブサイトで、2013 年 11 月の、ある一週間でオンラインのテストを行った。

オンラインテストの評価値として、CTR, CPC (Cost Per Click), RPR (Revenue Per Request) の 3 つを用いた。これらの指標は以下のように定義される。

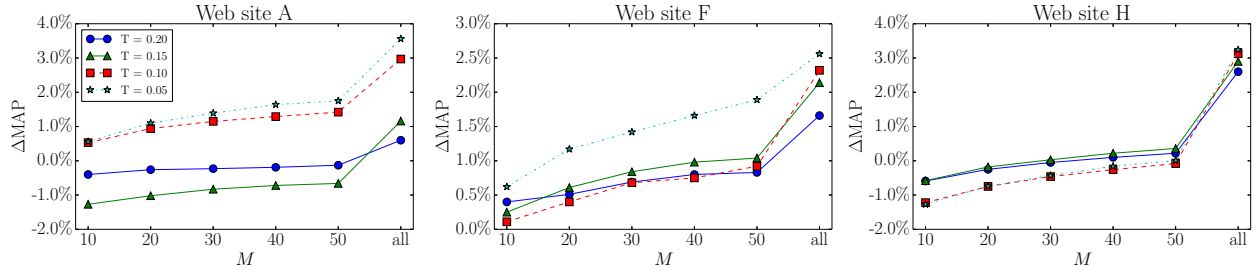
$$\text{CTR} = \frac{\#\text{clicks}}{|R|}, \text{CPC} = \frac{\text{revenue}}{\#\text{clicks}}, \text{RPR} = \frac{\text{revenue}}{|R|}$$

$|R|$ は広告リクエストの数を表す。revenue は広告主が支払った総額（広告収益）である。

表 4 に実験結果をまとめた。これらのパーセントの値も、相対的な向上度合いを示している。ウェブサイト A 以外の全てのサイトで、CTR が向上した。オフライン評価の結果である表 3 で示したように、ウェブサイト A では $T = 0.20$ の時に既存手法 1 と 2 よりも提案手法は悪い結果であったが、ここでは全てのウェブサイトで同じハイパーパラメータ ($T = 0.20, M = 20$) を用いて評価を行った。そのため、この結果は妥当なものである。ハイパーパラメータ M はシステムの応答性能によって決定するためウェブサイトごとに変化させることが難しいが、 T はそれぞれのサイトに合わせて調整した方がよいと言える。学習データが多いウェブサイト B と H では、オフライン評価の傾向と同じく、他のウェブサイトよりも CTR の向上比率は大きかった。ウェブサイト A, B, D, E, H における CTR の結果は、5% の水準 ($p\text{-value} < 0.05$) で統計的に有意であった。RPR も、ウェブサイト A 以外のすべてのサイトにおいて向上した。一方、ウェブサイト B, E, G, H では CPC が減少した。この A/B テストにおいては、収益を考慮して広告のランキングおよび表示を行っている。提案手法で取得された広告は、CTR が高く入札額が比較的低いものであったため、このような結果を示したのだと考えられる。1 章で述べたように、本稿では CTR を向上させることに注目している。その結果として、広告収入も向上させることができたと言える。

表3 実験結果. 値は ΔMAP . 太字の要素は手法間で一番良い結果を示している.

手法	ウェブサイト							
	A	B	C	D	E	F	G	H
既存手法 1	+1.90%	+0.01%	-0.04%	+0.01%	+0.87%	+1.22%	+0.31%	+0.00%
既存手法 2	+0.85%	+2.71%	+0.53%	+0.24%	+0.85%	+0.59%	+0.98%	+1.00%
既存手法 3	+0.04%	+0.37%	+0.04%	+0.03%	+0.02%	+0.38%	+0.01%	+0.11%
$T = 0.20$	+0.60%	+6.18%	+1.42%	+0.50%	+1.70%	+1.66%	+1.73%	+2.60%
$T = 0.15$	+1.16%	+6.60%	+1.50%	+0.60%	+1.79%	+2.14%	+1.86%	+2.90%
$T = 0.10$	+2.97%	+6.83%	+1.55%	+0.64%	+1.84%	+2.32%	+2.00%	+3.13%
$T = 0.05$	+3.56%	+6.93%	+1.54%	+0.68%	+1.81%	+2.56%	+2.12%	+3.24%

図3 M を変化させた時の実験結果. 紙面の制限のため, 一部のウェブサイトにおける実験結果のみを抜粋した.

5. ま と め

コンテキスト広告は, 一般的なウェブページに表示されるテキスト形式の広告である. 効果的な広告効果とより良いユーザー体験のため, ウェブページの内容やユーザー情報に関連した広告を, 膨大な候補の中からどのように選択するかというのは, コンテキスト広告の大きな問題の一つである.

本稿では, 転置インデックスと WAND アルゴリズムを用いた既存の検索システムの優れた応答性能を活用する形で, コンテキスト情報から広告のテキスト特徴量へのマッピングを学習する変換手法を提案した. コンテキスト情報には, ウェブページの情報や, ユーザーの属性や行動履歴を元にした情報が含まれる. この提案手法を適用するにあたっては, 学習した変換行列により, コンテキスト情報を広告検索システムの入力単語ベクトルに変換するだけでよい. そのため, 既存の転置インデックスを修正したり, 新規にインデックスを追加する必要もなく, 実装が容易である.

あるアドネットワークの実際のデータを用いたオフライン評価で, 提案手法は既存の手法と比較して良い結果を示した. さらに, 実際の広告配信サービスに提案手法を適用し, 既存の配信システムに対する向上を確認した.

◇ 参 考 文 献 ◇

- [Agarwal 12] Agarwal, D. and Gurevich, M.: Fast Top-k Retrieval for Model Based Recommendation, in *Proceedings of the 5th ACM International Conference on Web Search and Data Mining* (2012)
- [Aly 12] Aly, M., Hatch, A., Josifovski, V., and Narayanan, V. K.: Web-scale User Modeling for Targeting, in *Proceedings of the 21st International Conference on World Wide Web* (2012)
- [Broder 03] Broder, A. Z., Carmel, D., Herscovici, M., Soffer, A., and Zien, J.: Efficient Query Evaluation Using a Two-level Retrieval Process, in *Proceedings of the 12th International Conference on Information and Knowledge Management* (2003)
- [Broder 07] Broder, A., Fontoura, M., Josifovski, V., and Riedel, L.: A Semantic Approach to Contextual Advertising, in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (2007)
- [Chakrabarti 08] Chakrabarti, D., Agarwal, D., and Josifovski, V.: Contextual Advertising by Combining Relevance with Click Feedback, in *Proceedings of the 17th International Conference on World Wide Web* (2008)
- [Chapelle 14] Chapelle, O., Manavoglu, E., and Rosales, R.: Simple and Scalable Response Prediction for Display Advertising, *ACM Trans. Intell. Syst. Technol.*, Vol. 5, No. 4, pp. 61:1–61:34 (2014)
- [Covington 16] Covington, P., Adams, J., and Sargin, E.: Deep Neural Networks for YouTube Recommendations, in *Proceedings of the 10th ACM Conference on Recommender Systems* (2016)
- [Dave 10] Dave, K. S. and Varma, V.: Learning the Click-through Rate for Rare/New Ads from Similar Ads, in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2010)
- [Ding 11] Ding, S. and Suel, T.: Faster Top-k Document Retrieval Using Block-max Indexes, in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2011)
- [Fontoura 11] Fontoura, M., Josifovski, V., Liu, J., Venkatesan, S., Zhu, X., and Zien, J.: Evaluation Strategies for Top-k Queries over Memory-Resident Inverted Indexes, in *Proceedings of the 37th International Conference on Very Large Data Bases* (2011)
- [Graepel 10] Graepel, T., Candela, J. Q., Borchert, T., and Herbrich, R.: Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine, in *Proceedings of the 27th International Conference on Machine Learning* (2010)
- [He 14] He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., Shi, Y., Atallah, A., Herbrich, R., Bowers, S., and Candela, J. Q.: Practical Lessons from Predicting Clicks on Ads at Facebook, in *Proceedings of the 8th International Workshop on Data Mining for Online Advertising* (2014)
- [Huang 13] Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., and Heck, L.: Learning Deep Structured Semantic Models for Web Search Using Clickthrough Data, in *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management* (2013)
- [Joachims 02] Joachims, T.: Optimizing Search Engines Using Click-

表 4 オンライン環境での A/B テストの結果. 評価指標は CTR (click-through rate), CPC (cost per click), RPR (revenue per request). 値は相対的な向上割合で示している. CTR の結果に対しては有意差検定を行った.
*: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001

評価指標	ウェブサイト							
	A	B	C	D	E	F	G	H
CTR	-3.67% **	+4.60% ***	+0.48%	+2.82% *	+2.47% **	+1.42%	+3.27%	+4.02% ***
CPC	+3.63%	-2.00%	+1.62%	+1.31%	-1.01%	+7.51%	-2.42%	-2.94%
RPR	-0.18%	+2.51%	+2.10%	+4.17%	+1.44%	+9.04%	+0.77%	+0.97%

through Data, in *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2002)

[Joshi 11] Joshi, A., Bagherjeiran, A., and Ratnaparkhi, A.: User Demographic and Behavioral Targeting for Content Match Advertising, in *Proceedings of the 5th International Workshop on Data Mining and Audience Intelligence for Advertising* (2011)

[Juan 16] Juan, Y., Zhuang, Y., Chin, W.-S., and Lin, C.-J.: Field-aware Factorization Machines for CTR Prediction, in *Proceedings of the 10th ACM Conference on Recommender Systems* (2016)

[Karimzadehgan 11] Karimzadehgan, M., Li, W., Zhang, R., and Mao, J.: A Stochastic Learning-to-rank Algorithm and Its Application to Contextual Advertising, in *Proceedings of the 20th International Conference on World Wide Web* (2011)

[Ling 17] Ling, X., Deng, W., Gu, C., Zhou, H., Li, C., and Sun, F.: Model Ensemble for Click Prediction in Bing Search Ads, in *Proceedings of the 26th International Conference on World Wide Web Companion* (2017)

[Liu 05] Liu, T., Moore, A. W., Yang, K., and Gray, A. G.: An Investigation of Practical Approximate Nearest Neighbor Algorithms, in *Advances in Neural Information Processing Systems 17* (2005)

[Manning 08] Manning, C. D., Raghavan, P., and Schütze, H.: *Introduction to Information Retrieval*, Cambridge University Press (2008)

[McMahan 13] McMahan, H. B., Holt, G., Sculley, D., Young, M., Ebner, D., Grady, J., Nie, L., Phillips, T., Davydov, E., Golovin, D., Chikkerur, S., Liu, D., Wattenberg, M., Hrafnkelsson, A. M., Boulos, T., and Kubica, J.: Ad Click Prediction: A View from the Trenches, in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013)

[Ratnaparkhi 08] Ratnaparkhi, A.: A Hidden Class Page-ad Probability Model for Contextual Advertising, in *Workshop on Targeting and Ranking for Online Advertising at the 17th International World Wide Web Conference* (2008)

[Sprechmann 13] Sprechmann, P., Litman, R., Ben Yakar, T., Bronstein, A. M., and Sapiro, G.: Supervised Sparse Analysis and Synthesis Operators, in *Advances in Neural Information Processing Systems 26* (2013)

[Strohman 07] Strohman, T. and Croft, W. B.: Efficient Document Retrieval in Main Memory, in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (2007)

[Tagami 13] Tagami, Y., Ono, S., Yamamoto, K., Tsukamoto, K., and Tajima, A.: CTR Prediction for Contextual Advertising: Learning-to-rank Approach, in *Proceedings of the 7th International Workshop on Data Mining for Online Advertising* (2013)

[Trofimov 12] Trofimov, I., Kornetova, A., and Topinskiy, V.: Using Boosted Trees for Click-through Rate Prediction for Sponsored Search, in *Proceedings of the 6th International Workshop on Data Mining for Online Advertising and Internet Economy* (2012)

[Yuan 12] Yuan, S., Abidin, A. Z., Sloan, M., and Wang, J.: Internet Advertising: An Interplay among Advertisers, Online Publishers, Ad Exchanges and Web Users, *arXiv preprint arXiv:1206.1754* (2012)

[担当委員: 風間 一洋]

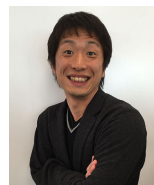
2017 年 5 月 2 日 受理

著者紹介



田頭 幸浩(正会員)

2010 年東京工業大学大学院情報理工学研究科計算工学専攻修士課程修了。同年、ヤフー株式会社入社。ディスプレイ広告のクリック率予測モデルの構築など、機械学習技術の実サービス適用に携わる。現在、京都大学大学院情報科学研究科知能情報学専攻博士課程に在籍中。



堀田 徹

2009 年早稲田大学大学院基幹理工学研究科情報理工学専攻修士課程修了。同年、ヤフー株式会社入社。決済金融サービスのシステム運用を経て、現在はディスプレイ広告における配信ロジックの研究開発に取り組む。



田中 祐介

2009 年中央大学大学院情報理工学研究科情報工学専攻修士課程修了。同年、ヤフー株式会社入社。レコメンドシステム・ターゲティング広告のシステム・ロジック開発を経て、広告プロダクト・営業の分析業務に従事。



小野 真吾

2009 年東京大学大学院情報理工学系研究科数理情報学専攻博士課程修了。同年、ヤフー株式会社入社。ウェブ検索サービスの開発を経て、オンライン広告や E コマースサイトへの機械学習技術の適用に取り組む。博士(情報理工学)。



塚本 浩司

1998 年東京大学大学院工学系研究科化学系生命系専攻修士課程修了。株式会社富士通研究所、スタンフォード大学言語情報研究センターを経て、2009 年ヤフー株式会社入社。Yahoo! JAPAN 研究所所属後、2015 年より D&S 統括本部サイエンス本部長。



田島 玲

1992 年東京大学大学院工学系研究科航空工学専攻修士課程修了。日本アイ・ビー・エム株式会社東京基礎研究所、A.T. カーニーを経て、2012 年よりヤフー株式会社 Yahoo! JAPAN 研究所所長。2000 年に東京大学大学院理学系研究科情報科学専攻博士課程修了。博士(理学)。