

# オンライン広告における CTR 予測モデルの素性評価

田頭 幸浩<sup>†</sup> 山本 浩司<sup>†</sup> 小野 真吾<sup>†</sup> 塚本 浩司<sup>†</sup> 田島 玲<sup>†</sup>

<sup>†</sup> ヤフー株式会社 〒107-6211 東京都港区赤坂 9-7-1 ミッドタウン・タワー

E-mail: <sup>†</sup>{yutagami,koyamamo,shiono,kotsukam,atajima}@yahoo-corp.jp

**あらまし** オンライン広告はインターネットの経済を支える大きな柱の一つであり、ビジネスと学術の両方から大きな注目を浴びている。本稿ではオンライン広告のうち、クリック課金型のテキスト広告に注目し、そのクリック率 (click-through rate; CTR) を予測するモデル (CTR 予測モデル) の素性について評価を行う。CTR 予測モデルの精度向上は適切な広告選択に不可欠であり、収益と直結した重要な部分を占めるため、その素性は現在までに多岐に渡って提案されてきた。本稿では実際の配信システムログを用いてこれらの素性を評価し、その有用性を検証する。

**キーワード** オンライン広告, CTR 予測, 機械学習, 統計モデル

## 1. 導 入

オンライン広告はインターネットの経済を支える大きな柱の一つである。そのため、この分野はビジネス的に、また学術的にも大きな注目を浴びている [6], [22]。オンライン広告の例としては、検索サイトにおける検索連動型広告、ポータルサイトにおけるディスプレイ広告、ニュースやブログ記事のページにおけるコンテキスト広告が挙げられる。

本稿ではオンライン広告のうち、クリック課金型のテキスト広告を扱う。クリック課金型広告とは、広告が配信されたページを閲覧しているユーザーが、広告をクリックして広告主の設定したページに移動した場合に、広告主があらかじめ入札していた金額に基づき課金される仕組みである。クリック課金以外の課金モデルとしては、広告が表示された回数に応じて課金されるインプレッション課金や、商品の購入手続きや資料請求などの広告主があらかじめ定めた目標にユーザーが至った時に課金されるコンバージョン課金などがある。また、テキスト広告とは、広告自体がテキストのみからなる広告である。

先に述べたようにクリック課金ではユーザーが広告をクリックした場合のみ広告主はその代金を支払う。そのため、ある広告を配信した場合の期待収益は、広告主が設定した入札額と、クリック率 (click-through rate; CTR) によって決まる。あるページに対し、複数の広告候補から配信する広告を適切に選択して収益を最大化する問題を考えた時、前者の入札額は広告主が設定しているため既知であるが、後者の CTR は未知であるため予測を行う必要がある。もしも CTR の予測が真の値とかけ離れたものであれば、配信候補の選択を適切に行うことができず、ユーザーがページを訪れた時に得られる収益も最大とは程遠いものになるであろう。そのため、クリック課金モデルの広告配信において CTR 予測は収益と直結した非常に重要な部分を占めるのである。

現在の広告配信システムでは、広告配信ログをもとに機械学習で CTR 予測モデルを構築している。本稿では実際の広告配信ログを用いて CTR 予測モデルの素性を評価し、その有用性

を検証する。

本稿の構成は以下のとおりである。2 章では先行研究でどのような予測モデルと素性 (特徴量) が提案されてきたかについて述べる。続く 3 章では広告配信システムの全体概要について述べる。4 章では実際の広告配信ログを用いた評価を行う。これらを受け 5 章では本稿を結び、将来の展望について述べる。

## 2. 予測モデルと素性

### 2.1 予測モデル

CTR 予測や問題設定が良く似たコンバージョン予測のモデルとしては一般的に線形モデルが用いられ、その中でも CTR が確率であることから、確率値を出力とするロジスティック回帰モデルが用いられることが多い [7], [8], [18]。他のより複雑なモデル、例えばカーネルを用いた SVM や Random forest [3], Gradient boosted decision trees [10], [12] を用いることも考えられるが、後述するように素性 (特徴量) の種類の多さと実際のシステムログから得られる学習データの膨大さ、配信時の計算コストが比較的少ないことから、ロジスティック回帰モデルがこの問題には適しているといえる。ロジスティック回帰モデル以外の線形モデルとしては、プロビット回帰が使われる例もある [13]。

### 2.2 素 性

予測モデルの入力として用いられる素性はさまざまなものが提案されている。大まかに、以下のような素性を挙げることができる。

- 配信対象のページと広告の類似度
- ユーザーの属性と広告の関連性
- 広告そのものの情報
- 過去の配信実績情報

まず直観的なものとして、広告を配信するページと広告の語句的な類似度がある [21]。これは、ページに関連した広告を配信した方がユーザーはクリックしやすいであろうという考えに基づいている。しかし語句的な類似度だけではミスマッチが容易に起こるため、ページとカテゴリをそれぞれ意味カテゴリや

クラスタに割り振り、それらの類似度を用いることも行われている [5], [17].

ページと広告の語句や意味的な類似度と同様に、ユーザーの持つ属性と広告の関連性も重要な素性である [7]. ユーザーの属性としては年齢や性別、住んでいる地域がある. 例えば多くの 20 代の女性はオートバイの広告よりも化粧品の広告に興味を持ちクリックしやすいと推測される. また、ユーザーが住んでいる場所に関連した広告の方が全く関係がない地域のものよりもクリックされやすいであろう. これらは人間の持つ直観や知識によるものでしかないが、モデルの素性としてユーザーの属性と広告の関連性を用いることにより、実際のクリックログから、関連性のうち重要なものを見出すことが可能になると考えられる.

年齢、性別、住んでいる地域などは静的な情報だが、動的な情報としてユーザーの行動履歴を挙げることができる [2]. 例えば直前に「ハワイ 旅行」というクエリで検索を行っていたユーザーは、現時点でお菓子のレシピページを見ていたとしても、ハワイ旅行に興味を持っており、これに関連した広告はクリックされやすいと考えられる. ページの場合と同じように、ユーザーの行動履歴から推定される興味は語句のレベルだけではなくカテゴリやクラスタレベルの情報も用いられる [2].

ユーザーの行動履歴から現時点の興味を推定する場合に重要な要素になるのはイベントの種類とそれがどれほど前に行われたかである. ここでいうイベントとは、先の例のような検索行動に加えて、ページの閲覧や広告の閲覧、クリック、コンバージョンなどを指す. 直観的に受動的な広告の閲覧に比べて、ユーザーが能動的に行う検索行動や広告のクリックなどの方が、より興味を推定するための強い手がかりになるであろう. また、同じ検索行動であっても 1 週間前と 1 時間前では現時点の興味への寄与度は異なるであろうし、検索クエリの内容によっても同様である. ユーザーの行動履歴を全て保持しておくことはストレージ量の問題があるため効率的ではない. そのため適切な素性選択を行い、CTR 予測に有用な情報のみを残すことが重要である [2].

ユーザーの行動履歴から年齢や性別を推定し、それらの推定された情報を素性として用いることもある. これは間接的に興味推定を行っており、一見非効率に見えるが、年齢や性別が既知であるユーザーの割合が少ない場合において、予測モデルの精度を向上させることが報告されている [19]. 一つの広告に対する一人のユーザーの広告閲覧回数が増加するにしたがって、そのユーザーと広告の組み合わせにおける CTR は低下するという報告がされており、これも CTR 予測モデルの素性として用いられる [7].

CTR 予測の素性としては、広告そのものの情報も用いられる. 例えば広告のテキストに広くユーザーの目を引く単語が含まれている場合などが分かりやすいであろう [14]. また、その広告の掲載を依頼した広告主や、キャンペーン、広告グループと呼ばれる、その広告に関する階層情報も使用される. これらの情報が CTR 予測に影響を及ぼす原因としては、広告主が意図している広告効果がオンラインもしくはその他の媒体を通し

てユーザーに伝わり、広範囲の人が興味を持っていたり、より単純に広告テキストが洗練されていることが考えられる.

ここまで列挙してきた素性は、ページ、ユーザー、広告それぞれの情報そのものと、二者間の類似度であった. これに加えて、更に重要な素性として実績 CTR が用いられる [15]. 実績 CTR とは、過去にその広告がクリックされた回数を配信された回数で割った値である. この実績 CTR はさまざまな粒度で集計して用いられる. 例えば、その広告の単位や、広告主単位、ユーザーの性別や年代単位、またはそれらをかけ合わせた単位などである. 粒度を細かくすればするほど、現在広告配信を行おうとしている状態に近づくが、過去に行われた配信回数とクリック回数が共に少なくなるため、信頼性が下がってしまう. そのため、どの粒度の実績 CTR を用いるかを選択したり、分母と分子に定数を足してスムージングを行う [20] などの工夫が重要である.

CTR に影響を与えるその他の要素としては、広告が表示されるページ中のポジションがある [9]. 誤クリックの場合を除き、広告がクリックされるためには、ユーザーが広告を認識することが必要である. 広告の表示位置は、ユーザーが広告を認識するかどうかに影響を与え、CTR にも影響を与える. これは検索結果の表示でもみられる現象である [9]. 先に述べた実績 CTR の集計においても、このポジションによる CTR への影響を考慮し、各表示位置の平均 CTR を用いて標準化を行った click over expected click(COEC) [23] が用いられることがある. COEC は各ポジションごとのクリック数  $c_{pos}$  とインプレッション数  $i_{pos}$ , CTR の平均  $CTR_{pos}$  をもとに以下のように表現される.

$$COEC = \frac{\sum_{pos=1}^P c_{pos}}{\sum_{pos=1}^P i_{pos} \cdot CTR_{pos}}$$

4 章では、以上の素性を用いて CTR 予測モデルを構築し、評価を行った結果を示す.

### 3. 広告配信システムの全体概要

2 章で述べたように CTR 予測モデルはさまざまな素性を用いて CTR の予測を行う. 配信対象のページやユーザーが異なれば予測 CTR も異なるため、配信時にリアルタイムで予測を行う必要があるが、配信候補の広告の数は膨大であり、全ての候補に対して CTR 予測を行うことは、実システムで許容される計算コストの制約から難しい. そのため、予測モデルで CTR の予測を行う前に配信候補を簡単な計算によって絞り込み、実際に予測を行う広告の数を減らす必要がある.

広告の絞り込みには情報検索の top- $k$  retrieval [4], [11] という手法が用いられる. top- $k$  retrieval ではクエリに対してスコアが高い  $k$  件のドキュメントをシステムから取得する. 一般的にクエリとドキュメント間の語句やフレーズ、メタデータなどの重なりに応じたスコアが用いられ、特にクエリとドキュメントを表す二つのベクトルの内積によってスコアが表現されることが多い. コサイン類似度や BM25 [16] もクエリベクトルとドキュメントベクトルをあらかじめスカラー倍することによって

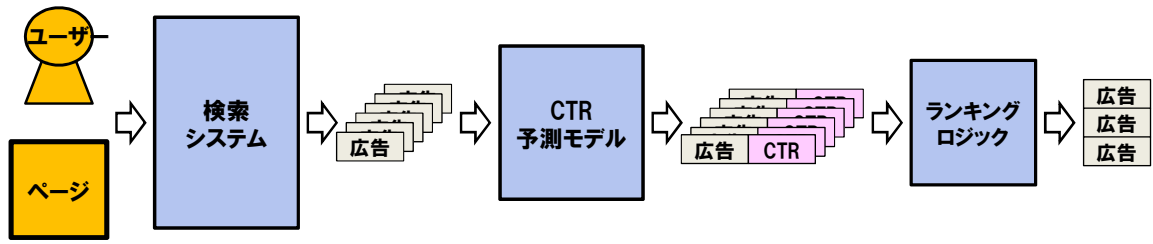


図 1 広告配信システムの概要. ユーザーとページから素性を抽出してクエリとし, 検索システムを用いて候補となる広告を取得する. その後候補の広告に対して CTR を予測し, ランキングロジックにて最終的に配信する広告と順序を決定する.

内積計算として扱うことが可能である.  $\text{top-}k$  retrieval では計算途中で既に上位  $k$  件に入らないことが分かっているドキュメントに対する評価を行わないことで効率的に上位  $k$  件のドキュメントを取得する [4], [11].

広告の絞り込みの場合, ドキュメントとして広告情報をシステムに格納し, クエリとして広告配信ページやユーザーの情報をを用いて検索を行う. 検索システムを用いて広告の絞り込みを行った後に CTR 予測モデルを用いてスコアの計算を行う様子を図 1 に示す.

モデルでの予測の前に広告を絞り込む場合, 最終的に目的とする値が大きくなるであろう候補を高速に取得できることが求められる. 高速に候補を取得する点は  $\text{top-}k$  retrieval の文脈で盛んに研究が行われている [4], [11]. また,  $\text{top-}k$  retrieval のスコアとして, 最終的に目的とする値を近似するように, 標準的な検索システムで扱うことができる二つのベクトルの内積で表現される関数形のモデルを, 機械学習で構築する研究も行われている [1].

## 4. 実験

この章では 2.2 節で述べた CTR 予測モデルの素性の評価を行う. まず 4.1 節で評価に用いるデータと素性セットについて述べ, 続いて 4.2 節で評価値の比較を行う.

### 4.1 実験設定

素性の評価には、『Yahoo!ディスプレイアドネットワーク』の, ある 7 つのウェブサイトでの実際の配信システムログを 28 日分用いた. この広告配信ログをウェブサイトごとに分割し, 更に前半の 14 日分を学習データ, 続く 7 日分をバリデーションデータ, 後半 7 日分をテストデータとして扱った. データの各サンプルは配信された広告一つに対応しており, クリックされたか否かがラベル付けされている. それぞれのデータは同じ規則でフィルタリングとサンプリングを行い, 最終的に得られた学習データとテストデータそれぞれの広告数やユーザー数などの統計情報は表 1 にまとめている.

素性は種類に応じてグループ分けを行った. それぞれのグループの詳細は表 2 にまとめている.

ページと広告の類似度として, 語句ベクトルのコサイン類似度と, その二つで共起する語句を用いた. ページの語句は tf-idf

表 2 素性グループ

素性グループ	素性詳細
ポジション	ページ内での広告表示位置
ページと広告の類似度	語句の類似度, 共起する語句
ユーザーと広告の関連度	カテゴリの類似度, 共起するカテゴリ
広告の情報	語句, カテゴリ, 広告 ID, 広告グループ ID, キャンペーン ID, 広告主 ID
過去の配信実績情報	広告の実績 CTR

をベースとして, HTML のタグや出現位置を考慮して重みの調整を行いベクトル化した. 広告も同様に, 語句の出現位置がタイトルか説明文かによって重みを調整した.

ユーザーと広告の関連度としては, 双方にカテゴリを付与し, カテゴリベクトルのコサイン類似度と, 両方で共起するカテゴリを用いた. ユーザーのカテゴリは, そのユーザーの興味を表現するカテゴリで, 行動履歴を用いた興味推定のシステムで得られたカテゴリを用いた. 広告のカテゴリは, 先に述べた語句のベクトルをもとに分類器でカテゴリを付与した. ユーザーと広告のカテゴリは同じ体系を用いた.

広告の情報としては, 語句とカテゴリの他に, 広告に結びついた情報を用いた. この情報は階層構造になっており, 広告, 広告グループ, キャンペーン, その広告を出稿した広告主の順で階層が高くなる. 広告の情報として, これらの広告, 広告グループ, キャンペーン, 広告主に付与された ID を用いた.

過去の配信実績情報として, 広告の実績 CTR を用いた. 本実験では, 広告情報の各粒度の実績値をポジションを考慮して集計し, 各粒度とその配信数を考慮して重みづけた値を用いた.

7 つのウェブサイトそれぞれに対して, 上記の各素性セットごとに, 以下の式で表現されるロジスティック回帰モデルを用いて CTR 予測モデルを構築した.

$$p(c | p, u, a) = \frac{1}{1 + \exp(-c \cdot \mathbf{w}^T \mathbf{f}(p, u, a))}$$

なお,  $c \in \{+1, -1\}$  はクリックされたか否かを表す変数であり,  $c = +1$  の時にクリックされたことを,  $c = -1$  の時にクリックされなかったことを表す. また,  $p(c = +1 | p, u, a)$  はページ  $p$ , ユーザー  $u$ , 広告  $a$  が与えられた時にクリックされる確率を表す.  $\mathbf{f}(p, u, a)$  はその 3 つから抽出された素性ベク

表 1 各データの統計情報

ウェブサイト	データ種類	サンプル数	広告 ID 数	広告グループ ID 数	キャンペーン ID 数	広告主 ID 数
A	学習	543,006	33,689	28,271	15,423	7,723
	バリデーション	329,070	19,556	16,602	10,140	5,292
	テスト	242,057	17,899	15,116	9,209	4,915
B	学習	1,875,952	55,634	46,184	20,524	9,755
	バリデーション	1,089,336	33,851	28,410	14,197	7,168
	テスト	1,231,439	35,546	29,719	14,503	7,213
C	学習	7,760,783	135,962	107,063	34,799	14,689
	バリデーション	4,848,506	86,815	68,604	26,077	11,558
	テスト	4,476,524	85,821	67,171	25,526	11,451
D	学習	1,277,352	62,769	51,234	21,036	10,342
	バリデーション	766,223	38,651	33,112	14,387	7,358
	テスト	672,299	36,915	31,511	13,631	7,087
E	学習	515,457	33,340	27,807	14,732	7,525
	バリデーション	205,215	18,115	15,344	9,261	4,846
	テスト	168,364	17,217	14,497	8,844	4,733
F	学習	231,330	21,891	18,740	8,000	4,208
	バリデーション	131,890	13,312	11,581	5,147	2,821
	テスト	126,267	14,260	12,443	5,324	2,952
G	学習	1,803,434	122,090	99,385	33,622	14,200
	バリデーション	887,859	84,250	68,861	26,437	11,711
	テスト	846,326	84,595	69,221	26,064	11,663

トルを、 $\mathbf{w}$  はその素性に対応する重みベクトルを表現している。  
過学習を避けるため正則化項として  $\mathbf{w}^T \mathbf{w} / 2$  を加え、以下の最適化問題を解き  $\hat{\mathbf{w}}$  を得た。

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \log(1 + \exp(-c_i \cdot \mathbf{w}^T \mathbf{f}_i(p, u, a)))$$

ここで  $C > 0$  は正則化パラメータである。 $C$  を変化させて学習データで学習を行い、バリデーションデータでの評価値が高い  $\hat{\mathbf{w}}$  を用いてテストデータの評価を行った。評価値としては area under the ROC curve(AUC) [16] を用いた。

#### 4.2 評価

A から G の 7 サイトそれぞれのデータを用いて、各素性グループごとに CTR 予測モデルを構築し、AUC で評価を行った。なお、広告が表示されたポジションはとても強力な要素であるため、全ての場合において素性として用いた。結果が表 3 である。各ウェブサイトでは一番高い AUC を太字で記している。7 つのウェブサイトのうち 6 つのサイトで、全ての素性を用いた場合に一番高い AUC を示し、続いて広告情報の素性を用いた場合の AUC が高かった。A のサイトでは、この二つの順番が僅差で逆転した。ページ、ユーザーと広告の類似度を比較すると、ウェブサイトによって素性の影響度が異なり、B, D, F, G ではページと広告の類似度を、A, C, E ではユーザーと広告の類似度を追加した方が他方よりも高い AUC を示した。過去の配信実績情報も各ウェブサイトにおいて安定して AUC の向上に寄与していることが確認された。

続いて、広告の情報の素性グループの中でもどの種類の素性が有用であるかを検証した。結果が図 4 である。表 3 と同じく、各ウェブサイトでは一番高い AUC を太字で記している。7 つの

ウェブサイト全てにおいて、広告の語句を素性として用いた時に AUC が高くなった。広告の各 ID を素性とした場合、粒度を広告主 ID から広告 ID へと細かくするにつれ AUC も減少する傾向が見られた。広告のカテゴリ情報は他の素性との比較すると、各ウェブサイトでは寄与度は変化した。安定して AUC を向上させていることが分かった。

#### 5. まとめと今後の課題

本稿ではオンライン広告のうち、クリック課金型のテキスト広告に注目し、その CTR 予測モデルの素性について、『Yahoo! ディスプレイアドネットワーク』の配信システムログを用いて評価を行った。

今後の課題としては、さらなる素性の追加や、より複雑なモデルを用いた場合の評価に加え、マルチメディア広告への拡張を挙げることができる。

本稿ではテキスト広告の CTR 予測モデルに着目した。しかし同じクリック課金の広告でも画像や動画などからなるマルチメディア広告もあり、近年では画像から素性を抽出して CTR 予測モデルに用いる研究がされている [8]。この問題への取り組みも今後の大きな課題といえる。

#### 文 献

- [1] Deepak Agarwal and Maxim Gurevich. Fast top-k retrieval for model based recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pp. 483–492, New York, NY, USA, 2012. ACM.
- [2] Mohamed Aly, Andrew Hatch, Vanja Josifovski, and Vijay K. Narayanan. Web-scale user modeling for targeting.

表3 各素性グループを素性とした場合の AUC

素性グループ	A	B	C	D	E	F	G
ポジションのみ	0.8925	0.8250	0.7563	0.7938	0.7980	0.7342	0.7250
ページと広告の類似度	0.8937	0.8597	0.7594	0.8120	0.8047	0.7930	0.7368
ユーザーと広告の関連度	0.8940	0.8324	0.7649	0.8057	0.8200	0.7614	0.7286
広告の情報	<b>0.8954</b>	0.8775	0.7727	0.8304	0.8306	0.8220	0.7424
過去の配信実績情報	0.8938	0.8597	0.7648	0.8160	0.8144	0.7899	0.7339
全て	0.8953	<b>0.8781</b>	<b>0.7731</b>	<b>0.8322</b>	<b>0.8313</b>	<b>0.8234</b>	<b>0.7469</b>

表4 広告の情報を素性とした場合の AUC

素性詳細	A	B	C	D	E	F	G
語句	<b>0.8955</b>	<b>0.8761</b>	<b>0.7707</b>	<b>0.8282</b>	<b>0.8279</b>	<b>0.8208</b>	<b>0.7419</b>
カテゴリ	0.8950	0.8523	0.7676	0.8176	0.8247	0.8090	0.7330
広告 ID	0.8947	0.8617	0.7687	0.8125	0.8142	0.7822	0.7334
広告グループ ID	0.8948	0.8617	0.7691	0.8135	0.8157	0.7888	0.7342
キャンペーン ID	0.8956	0.8717	0.7696	0.8240	0.8219	0.8071	0.7367
広告主 ID	0.8947	0.8712	0.7685	0.8235	0.8232	0.8121	0.7381

In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pp. 3–12, New York, NY, USA, 2012. ACM.

- [3] Leo Breiman. Random forests. *Mach. Learn.*, Vol. 45, No. 1, pp. 5–32, October 2001.
- [4] Andrei Z. Broder, David Carmel, Michael Herscovici, Aya Soffer, and Jason Zien. Efficient query evaluation using a two-level retrieval process. In *Proceedings of the twelfth international conference on Information and knowledge management*, CIKM '03, pp. 426–434, New York, NY, USA, 2003. ACM.
- [5] Andrei Broder, Marcus Fontoura, Vanja Josifovski, and Lance Riedel. A semantic approach to contextual advertising. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pp. 559–566, New York, NY, USA, 2007. ACM.
- [6] Andrei Broder and Vanja Josifovski. Introduction to computational advertising. <http://www.stanford.edu/class/msande239/>. Accessed: 14/12/2012.
- [7] Haibin Cheng and Erick Cantú-Paz. Personalized click prediction in sponsored search. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pp. 351–360, New York, NY, USA, 2010. ACM.
- [8] Haibin Cheng, Roelof van Zwol, Javad Azimi, Eren Manavoglu, Ruofei Zhang, Yang Zhou, and Vidhya Navalpakkam. Multimedia features for click prediction of new ads in display advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pp. 777–785, New York, NY, USA, 2012. ACM.
- [9] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, pp. 87–94, New York, NY, USA, 2008. ACM.
- [10] Kushal S. Dave and Vasudeva Varma. 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*, SIGIR '10, pp. 897–898, New York, NY, USA, 2010. ACM.
- [11] Marcus Fontoura, Vanja Josifovski, Jinhui Liu, Srihari Venkatesan, Xiangfei Zhu, and Jason Zien. Evaluation strategies for top-k queries over memory-resident inverted indexes. In *Proceedings of the 37th International Conference on Very Large Data Bases*, Vol. 4, pp. 1213–1224, 2011.
- [12] Jerome H. Friedman. Stochastic gradient boosting. *Comput. Stat. Data Anal.*, Vol. 38, No. 4, pp. 367–378, February 2002.
- [13] Thore Graepel, Joaquin Quinonero Candela, Thomas Borchert, and Ralf Herbrich. 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*, pp. 13–20, 2010.
- [14] Andrew Hatch, Abraham Bagherjeiran, and Adwait Ratnaparkhi. Clickable terms for contextual advertising. In *ADKDD*, 2010.
- [15] Kuang-chih Lee, Burkay Orten, Ali Dasdan, and Wentong Li. Estimating conversion rate in display advertising from past performance data. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pp. 768–776, New York, NY, USA, 2012. ACM.
- [16] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [17] Adwait Ratnaparkhi. 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.
- [18] Rómer Rosales, Haibin Cheng, and Eren Manavoglu. Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pp. 293–302, New York, NY, USA, 2012. ACM.
- [19] Lucas Silva, Aaron Davis, and Henrique Ribeiro. A feature engineering approach for click-through rate prediction: Kdd cup track 2. In *KDD Workshop*, 2012.
- [20] Kuan-Wei Wu, Chun-Sung Ferng, Chia-Hua Ho, An-Chun Liang, Chun-Heng Huang, Wei-Yuan Shen, Jyun-Yu Jiang, Ming-Hao Yang, Ting-Wei Lin, Ching-Pei Lee, Perng-Hwa Kung, Chin-En Wang, Ting-Wei Ku, Chun-Yen Ho, Yi-Shu Tai, I-Kuei Chen, Wei-Lun Huang, Che-Ping Chou, Tse-Ju Lin, Han-Jay Yang, Yen-Kai Wang, Cheng-Te Li, Shou-De Lin, and Hsuan-Tien Lin. A two-stage ensemble of diverse models for advertisement ranking in kdd cup 2012. In *KDD Workshop*, 2012.
- [21] Wen-tau Yih and Ning Jiang. Similarity models for ad rel-

evance measures. In *MLOAD - NIPS 2010 Workshop on online advertising*, pp. 32–38, 2010.

- [22] Shuai Yuan, Ahmad Zainal Abidin, Marc Sloan, and Jun Wang. Internet advertising: An interplay among advertisers, online publishers, ad exchanges and web users. *CoRR*, 2012.
- [23] Wei Vivian Zhang and Rosie Jones. Comparing click logs and editorial labels for training query rewriting. In *A workshop at the 16th International World Wide Web Conference*, 2007.