

一般画像認識手法を応用したディスプレイ広告のユーザ属性推定とクリック率予測

山元 浩平[†] 茂木 哲矢^{††} 田頭 幸浩^{††} 小林 隼人^{††} 小野 真吾^{††}
中山 英樹[†]

[†] 東京大学大学院情報理工学系研究科

〒 1138656 東京都文京区本郷 731

^{††} ヤフー株式会社

〒 1076211 東京都港区赤坂 971 ミッドタウン・タワー

E-mail: [†]yamamoto@nlab.ci.i.u-tokyo.ac.jp, ^{††}{tmotegi,yutagami,hakobaya,shiono}@yahoo-corp.jp,

^{†††}nakayama@ci.i.u-tokyo.ac.jp

あらまし クリック課金型ディスプレイ広告のクリック率 (click-through rate; CTR) 予測において、レイテンシとコールドスタート問題—履歴の乏しい広告や属性が未知のユーザに対する予測の困難さ—は解決すべき主要な問題である。本稿では、計算コストを抑えつつコールドスタート問題に対応可能なクリック率予測モデルとして、一般画像認識手法を応用した内容ベースクリック率予測手法を提案する。具体的には (1) 広告画像から畳み込みニューラルネットワーク (convolutional neural networks; CNN) 等により抽出した画像特徴量を説明変数とし、ユーザの属性を目的変数として識別器を構築することで、広告画像から未知のユーザの属性を推定する (2) 推定したユーザ属性を広告主 ID や広告 ID 等の ID 情報と組み合わせたものを予測モデルの素性として用いてクリック率予測を行う。本手法では画像特徴量を用いて学習を行うため、履歴の乏しい広告や属性が未知のユーザに対しても画像の類似性から予測・推定が可能である。最終的に、提案手法を広告のアクセスログのデータセットに適用し、その有効性を検証する。キーワード 広告クリック率予測, 一般画像認識, 畳み込みニューラルネットワーク, 画像特徴量, ユーザ属性推定

Predicting User Demographics and Click Through Rate of Display Ads Applying a Generic Image Recognition Methodology

Kohei YAMAMOTO[†], Tetsuya MOTEGI^{††}, Yukihiro TAGAMI^{††}, Hayato KOBAYASHI^{††}, Shingo ONO^{††}, and Hideki NAKAYAMA[†]

[†] Graduate School of Information Science and Technology, The University of Tokyo

731 Hongo, Bunkyo-ku, Tokyo 1138656, Japan

^{††} Yahoo Japan Corporation

Midtown Tower 971 Akasaka, Minato-ku, Tokyo 107-6211, Japan

E-mail: [†]yamamoto@nlab.ci.i.u-tokyo.ac.jp, ^{††}{tmotegi,yutagami,hakobaya,shiono}@yahoo-corp.jp,

^{†††}nakayama@ci.i.u-tokyo.ac.jp

Abstract In the field of click-through rate (CTR) prediction of pay per click display ads, the latency and the cold-start problem, which is a difficulty in CTR prediction regarding users and ads with little historical data, are fundamental problems. In this work, we propose a content-based CTR prediction method applying a generic image recognition methodology. Our proposed method is completely content-based using image similarity; therefore, we expect it to enable a CTR prediction even for users and items with little historical data. Finally, we evaluate our proposed method with datasets of display ads and confirmed its effectiveness.

Key words Ad CTR Prediction, Generic Image Recognition, Convolutional Neural Networks, Image Features, User Demographics Estimation

1. はじめに

オンライン広告に関する技術はアドテクノロジーと呼ばれ、産業面だけでなく学術面においても、機械学習や推薦システム等の情報技術の新たな適用分野として、近年大きな注目を集めている [1][2][3]。オンライン広告の例としては、検索エンジンでユーザーが検索したキーワードに関連した広告を表示する検索連動型広告や、ニュースやブログ記事のページに関連した広告を表示するコンテキスト広告などが挙げられる。

本稿ではオンライン広告のうち、クリック課金型ディスプレイ広告を扱う。クリック課金型広告とは、広告が配信されているページを閲覧しているユーザーが広告をクリックする度に、広告主に課金が行われる広告配信方式のことである。またディスプレイ広告とは、ページの一部に埋め込まれて表示される画像や Flash、動画などによる広告のことを指す。その一部は、バナー広告と呼ばれることもある。オンライン広告には、コンテンツを享受するユーザーや広告掲載を依頼する広告主の他に、広告掲載ページを提供するパブリッシャーと呼ばれる事業者が存在する。オンライン広告の処理の流れとユーザー、パブリッシャー、広告主の関係を図 1 に示す。パブリッシャーは各広告主の広告を掲載することで収益を得るため、ユーザーやページによってそれぞれ収益を最大化させるような広告を選択して掲載する必要がある。

これらのクリック課金型ディスプレイ広告では、ユーザーが広告をクリックした場合のみ広告主に課金が行われるため、ある広告を配信した場合のパブリッシャーの期待収益は、広告主が設定した入札額とその広告のクリック率によって決まる。ここでクリック率とは、表示された広告がクリックされる確率のことを指す。したがって、パブリッシャーが収益を最大化させるためには、クリック率を正確に予測することが非常に重要である。一般的に広告のクリック率は、広告やユーザーの情報を予測モデルの素性として用いることが多い [4][5]。したがって、新規の広告や属性が未知のユーザーに対してクリック率の予測を行うことは困難である。このような問題をコールドスタート問題 [6] と呼ぶ。一般的に、広告はいれ替わりのサイクルが短いことが多いため、新規の広告は常に大量に存在する。また、広告が掲載されるパブリッシャーのページを、ユーザーが必ずしもログインして利用するとは限らないため、属性が未知のユーザーも多く存在する。したがって、広告のクリック率を正確に予測するうえで、コールドスタート問題は本質的な問題の一つである。

また、実際の広告配信システムでは、ユーザーがページを訪れてから数十ミリ秒で広告を表示することが求められるため、クリック率予測の計算コストも考慮しなければならない。したがって、計算コストを抑えつつコールドスタート問題に対応したクリック率予測モデルを開発することは、非常に重要な研究課題である。

2. 関連研究

ディスプレイ広告のクリック率予測に用いる広告の情報としては、広告主や広告キャンペーン、広告画像などにそれぞれ付

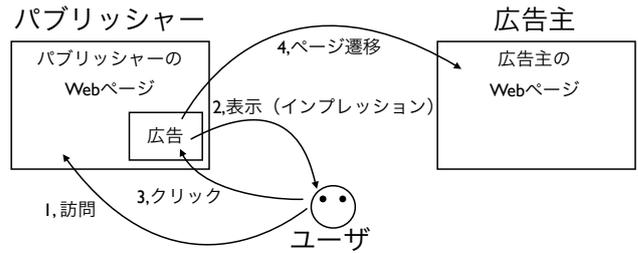


図 1 オンライン広告の処理の流れとユーザー、パブリッシャー、広告主の関係。

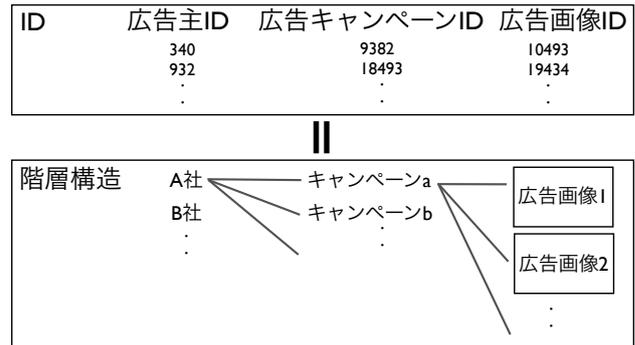


図 2 ID の階層構造

与されている ID がある。これらの ID は階層構造になっており、広告主、広告キャンペーン、広告画像の順に粒度が小さくなっていく (図 2)。また ID は、Weinberger ら [7] の feature hashing という手法により、バイナリにハッシングして用いられることが多い。ID はクリック率予測に有効であるうえ計算コストも小さいため、多くの研究で予測モデルの素性として用いられている [5][8][9]。しかし、ID は広告の類似性の表現力が乏しいため、協調フィルタリング [10][11] を用いた推薦システムと同様に、システム内に新規の広告が多い場合、クリック率を予測することが難しい。たとえば、一番粒度の大きい広告主 ID が新規の広告である場合、類似性をはかる手がかりがなく予測を行うことができない。

推薦システムにおいては、このようなコールドスタート問題に対応する手段として、アイテムの内容の類似性を推薦に反映させる内容ベースのアプローチが検討されている [12]。ディスプレイ広告のクリック率予測においても、広告から得られるマルチメディア特徴量を素性として用いる内容ベースアプローチで、コールドスタート問題に対応する手法が提案されている [5][13]。

Cheng ら [5] は、ディスプレイ広告のマルチメディア特徴量を用いて、新規広告に対してクリック率を予測する手法を提案した。彼らはまず、ディスプレイ広告の様々なマルチメディア特徴量と広告のクリック率の関連性を比較調査し、マルチメディア特徴量の中にはクリック率と相関のある特徴量が多く存在する可能性があることを示した。また最終的にこの研究では、広告のクリック率と相関のあったマルチメディア特徴量を、予測モデルの素性として用いることで、コールドスタート時の

クリック率予測の精度を向上させている。この研究はディスプレイ広告のクリック率予測の分野において、本格的にマルチメディア特徴量とクリック率との関連性を調査した初めての研究である。

しかし一方で、比較に用いられたマルチメディア特徴量は低次元のプリミティブなものが多く、一般画像認識で用いられるような、画像のセマンティックな内容を表現可能な特徴量是用いられていない。それは、一般画像認識で用いられるような特徴量は、連続値の密な高次元ベクトルであることが多いため、計算コストが大きいことが原因であると考えられる。そこで本稿では、一般画像認識で用いられている画像特徴量を用いて、計算コストを抑えつつコールドスタート問題に対応可能なクリック率予測モデルを提案し、その有効性を検証する。

3. 提案手法

広告のクリック率予測におけるコールドスタート問題とは、履歴の乏しい広告とユーザ属性が未知のユーザに対する予測の困難さであることは先に述べた。本研究では、画像特徴量を用いて未知のユーザの属性を推定（画像特徴量をユーザ属性にマッピング）し、推定したユーザ属性を広告主 ID、広告 ID 等の階層構造を持った ID と統合して用いることで、計算コストを抑えつつコールドスタート時のクリック率予測の精度を向上させる手法を提案する。提案手法のパイプラインを図 3 に示す。本手法は以下の 2 ステップから構成される。

Step 1. ユーザ属性推定

まず、 j 番目の広告画像 x_j から Gist [14], Fisher vector [15], CNN [16] などを用いて画像特徴量を抽出し、画像一枚ごとに特徴ベクトル $f(x_j)$ を作成する。次に、特徴ベクトル $f(x_j)$ を説明変数とし、ユーザ属性 c_j を目的変数として識別器を構築する。たとえばユーザの性別を推定したいときは、ユーザ属性 $c_j \in \{+1, -1\}$ は、男性であるか非男性（女性）であるかを表す変数になり、 $c_j = +1$ のとき男性、 $c_j = -1$ のとき女性を表す。 d 次元の特徴ベクトルをもつ n 枚の画像の学習データセット $D = \{f(x_j), c_j\}_{j=1}^n$ があるととき、以下の式で表されるロジスティック回帰モデルを用いた。

$$p(c | x) = \frac{1}{1 + \exp(\sum_{i=1}^d w_i f_i(x))} \quad (1)$$

なお、 $f_i(x)$ は、 d 次元の特徴ベクトルの i 番目の特徴を、 $w_i \in \mathbb{R}$ はその特徴に対応する重みを表している。過学習を避けるため、正則化項としてユークリッドノルム (L2 ノルム) を加えている。また、 $p(c = 1 | x)$ は、画像 x が与えられたとき、その画像をクリックしているのが男性である確率を表している。この確率が 0.5 以上のとき男性、0.5 未満のとき女性と識別することでユーザ属性を推定する。

Step 2. クリック率予測

Step 1 で推定されたユーザ属性と ID の特徴ベクトルをかけたものを説明変数とする。たとえば、性別の特徴ベクトルが $[0, 1]$, ID の特徴ベクトルが $[0, 0, 1, 0]$ のとき、性別と ID をかけあわせた特徴ベクトルは、 $[0, 0, 0, 0, 0, 1, 0]$ となる。ま

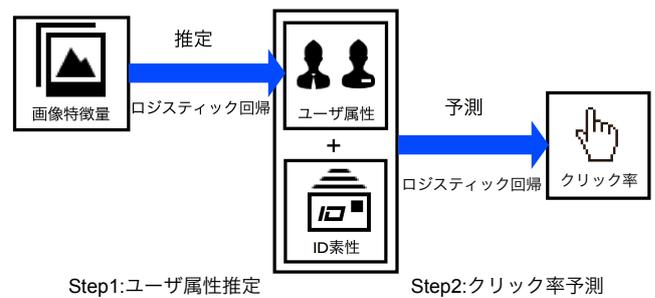


図 3 提案手法のパイプライン。

た目的変数は、クリックされていれば +1、表示（インプレッション）のみでクリックされていなければ -1 としてモデルを構築する。識別には目的変数の事後確率を用いるため、Step 1 と同様にロジスティック回帰モデルを適用し、目的変数の事後確率よりクリック率を予測する。ここでは、ロジスティック回帰によって求めた事後確率が予測クリック率になる。クリック率予測モデルの素性に直接密な画像特徴量を用いるのではなく、あらかじめ画像特徴量からユーザ属性にマッピングを行ない、ユーザ属性と ID をかけあわせたスパースな素性を予測モデルの素性として用いることで、計算コストをおさえることが可能である。

4. 評価実験

4.1 実験内容

ディスプレイ広告の広告画像とアクセスログのデータセットを用いて提案手法の評価実験を行った。本評価実験ではまず予備実験として、広告のクリック率予測タスクに有効な画像特徴量の比較検討を行なった。次に、コールドスタート時における提案手法の有効性を検証するために、訓練データとテストデータ内の広告画像（画像 ID）の重複率を 0% から 10% に徐々に変化させていった時のユーザ属性推定の精度変化を実験 1 で、クリック率予測の精度変化を実験 2 で比較した。ベースラインとして、ID を用いたモデルと比較した。

予備実験 画像特徴量の比較

予備実験では、ディスプレイ広告のクリック率予測タスクにおいて有効な画像特徴量を検討した。Fisher vector や CNN などを用いて広告画像から画像特徴量を抽出し、各種画像特徴量をモデルの説明変数として、広告ごとの実際の過去のクリック率の値にリッジ回帰させた。各特徴量によるモデルの性能比較は、予測クリック率と実際の過去のクリック率との Mean Absolute Error (MAE) で行った。

実験 1 ユーザ属性の推定

実験 1 では、コールドスタートな状況におけるユーザ属性推定の評価実験を行った。具体的には、まずデータセットに前処理を行い、各性別の教師ラベルがついた広告のクリックログを得た。これについては、次項のデータセットを参照されたい。次に、そのデータセット内の広告画像から画像特徴量を抽出し、その画像特徴量を説明変数、各性別を目的変数としてロジス

表 1 データセットのアクセスログ数と各素性のユニーク数

	アクセスログ数	広告主 ID	広告 ID	広告画像 ID
訓練データ	23,906,738	955	8,506	6,996
検証データ	3,236,631	1,005	10,237	8,262
テストデータ	3,253,943	1,009	10,325	8,360

ティック回帰で広告画像からクリックしたユーザの性別を推定した。画像特徴量として、予備実験の結果より CNN を用いて抽出した特徴量を用いた。提案手法の性能を評価するため、ID を用いたモデル (Baseline) と画像特徴量を用いたモデルをそれぞれ AUC で比較した。

実験 2 クリック率予測

実験 2 では、コールドスタートな状況におけるクリック率予測の評価実験を行った。具体的には、ユーザ属性と ID をかけあわせたものを説明変数とし、クリックを +1、インプレッションのみを -1 として、ロジスティック回帰モデルを適用することでクリック率を予測した。提案手法の性能を評価するため、ID のみを説明変数としたモデル、CNN を用いて抽出した画像特徴量のみを説明変数としたモデル、ID にユーザ属性をかけたモデルを AUC でそれぞれ比較した。

4.2 データセット

データセットとして、マルチメディア分野の国際会議である ICME2014 の Tencent Multimedia Ads pCTR Challenge [17] で提供されたデータセットを用いた。このデータセットは、Tencent という中国で検索エンジンを運営している会社のディスプレイ広告の広告画像とアクセスログのデータセットである。ここでいうアクセスログとは、ユーザに対するディスプレイ広告の表示 (インプレッション) と、ユーザがディスプレイ広告に対して行ったクリックのログデータである。このデータセットには以下のようなものが含まれる。

- ID 素性 (広告主 ID, 広告 ID, 広告画像 ID)
- ユーザ属性 (性別, 年齢)
- 広告画像

各広告画像のサイズは 160×210 pixels である。また、データセットのアクセスログ数と各 ID 素性のユニーク数をまとめたものを表 1 に示す。なお、表中の広告画像 ID の数が広告画像の枚数にあたる。

実験 1 のユーザ属性推定では、ユーザ属性の推定精度を適切に評価するうえで必要なデータセットを整備するために、データセットに対して前処理を行った。本実験で用いたデータセットは広告のインプレッションとクリックのログで構成されている。これをクリックされたものだけに絞れば、ユーザによる広告のクリックという行動は、動画視聴サイトにおける動画の視聴のように、ユーザの嗜好を反映した行動であるとみなすことができる。そこでまずデータセットの中から、クリックログだけを抽出し、ユーザ属性 (ここでは性別) ごとにクリックされている広告画像のデータセットを得た。次に、データセットのバイアスをなくす処理を行った。これはアクセスログに、配信の際のターゲティングなどによるバイアスが入っている可能性

があるためである。たとえば、50 代男性のみにしか配信されていない広告や、20 代女性のみしか配信されていない広告などターゲティングがかかった配信がアクセスログの中に存在している可能性がある。そこで本実験では、ユーザ属性推定の精度を適切に評価するために、この配信時のターゲティングによるバイアスを除いた。具体的には、各広告画像に対し性別ごとに 10 倍以上インプレッション数の差がある場合、ターゲティングバイアスがかかっていると判断し、そのようなデータを除いた。

4.3 画像特徴量と ID

大域特徴量である Gist と、局所特徴量の SIFT [18], C-SIFT [19], Opponent SIFT [20], RGB-SIFT [20] をそれぞれ Fisher vector に変換し、画像一枚の大域的な特徴ベクトルとしたものと、CNN (Caffe の ImageNet pre-trained model [23]) の最終層の出力を特徴量としたものを比較した。各画像特徴量の具体的な実装を以下にまとめる。

Gist

Gist は画像のシーン情報を記述する際によく利用される大域画像特徴量の一つである。本実験では、画像を 4×4 の小領域に区切り、各領域に対し 20 方向のフィルタバンクの反応を計算した。これを R, G, B の各色に対して行うことで、画像 1 枚に対して、 $4 \times 4 \times 20 \times 3 = 960$ 次元の特徴ベクトルを得た。計算コストの問題上、最終的に PCA で 256 次元に圧縮した。

Fisher vector (SIFT, C-SIFT, Opponent SIFT, RGB-SIFT)

Fisher vector は局所記述子の分布を表現する Bag-of-visual-words [21] ベースの state-of-the-art の手法である。特徴点ごとに、SIFT, C-SIFT は 128 次元, Opponent SIFT, RGB-SIFT は 384 次元の特徴ベクトルが得られる。本実験では、5 pixels ごとにグリッド状に特徴点を取り、それぞれ半径 8 pixels で局所記述子を抽出していった。次にそれらに対して PCA を適用し 16 次元に圧縮した後、混合数 16 の Gaussian mixture model (GMM) を用いて Fisher vector に変換した。また、画像内の大まかな位置情報を考慮するために、画像の上中下と全体の 4 領域で Fisher vector を計算し、画像 1 枚に対して、 $2 \times 16 \times 16 \times 4 = 2,048$ 次元の特徴ベクトルを得た。Gist 同様、最終的に PCA で 256 次元に圧縮した。

CNN (Caffe ImageNet pre-trained model)

CNN はニューラルネットワークの一種で、現在一般画像認識でもっとも良い精度を達成している。本実験では、Jia ら [22] を中心に開発が進められている deep learning のオープンソース・ソフトウェアである Caffe [23] を用いた。Caffe には、Krizhevsky ら [24] が、一般画像認識の精度を競うコンペティションである ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 [25] 用に、一般画像認識用のデータセットの ImageNet で学習した CNN のモデルが用意されている。本実験ではこの pre-trained モデルを用いている。このモデルに画像データセットを入力し、Donahue ら [26] の手法に基づき、ネットワークの中間層の出力を特徴量として抽出し、

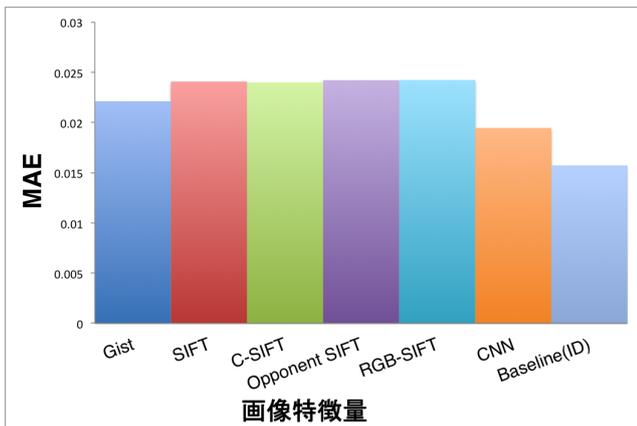


図 4 各画像特徴量によるモデルの MAE 比較。

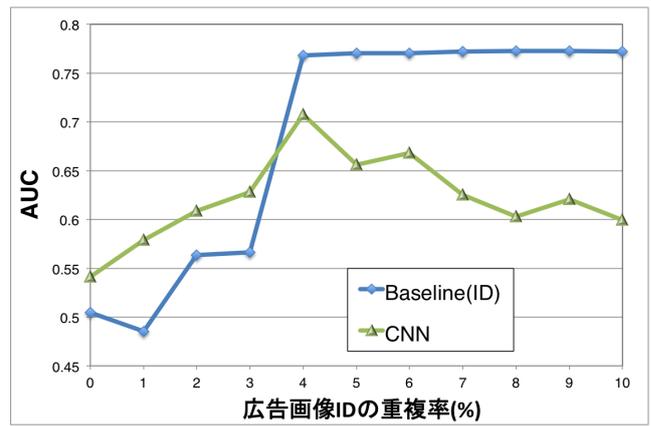


図 5 ユーザ属性推定の AUC 比較

画像一枚あたり 4,096 次元の特徴ベクトルを得た。他の画像特徴同様、最終的に PCA で 256 次元に圧縮した。

Baseline (ID)

ベースラインの特徴量として、ID (広告主 ID, 広告 ID, 広告画像 ID) を用いた。ID はそれぞれ Weinberger ら [7] の feature hashing に基づき、20-bit のバイナリに変換した。

4.4 実験結果

予備実験 画像特徴量の比較

リッジ回帰によって得られた各画像特徴量のモデルの性能を比較するため、各モデルの予測クリック率と実際の過去のクリック率の MAE を比較した結果を図 4 に示す。参考のために、ベースラインとして ID を用いたモデルも評価を行っている。

画像特徴量のなかでは CNN を用いたものが最も良かった。理由としては、このタスクにおいてはそこまで画像内の局所的な形上の情報は必要なく、むしろ画像内に写っている物体をある程度特徴として表現できることが必要だったことが考えられる。また、Fisher vector によるモデルの性能が最も悪かった原因としては、計算コストの問題上、全ての画像特徴量を 256 次元に圧縮して用いたため、Fisher vector のように次元数の大きい特徴ベクトルで画像を豊かに表現する特徴量は不利であったことが考えられる。

実験 1 ユーザ属性の推定

アクセスログの中からクリックログだけを抽出し、各クリックのユーザの性別を教師ラベルとした。CNN を用いたモデルと ID を用いたモデルの性能を AUC で比較した結果を図 5 に示す。図 5 よりコールドスタート時には、画像特徴量の方が性別をより正確に推定できていることが分かる。

また図 6 は、各性別クラスに属する事後確率が高かったトップ 16 の広告画像を示している。図 6 より、各性別の嗜好がよく反映されていることが分かる。例えば、女性側であれば女性ファッションの画像が多く、男性側であればゲームや紳士ファッションの画像が多い。このように広告の男性度・女性度が分かることで、未知のユーザであってもクリックした広告から、画像の類似性を用いて推定されたユーザ属性にマッピングを行うことが可能である。



図 6 性別ごとに事後確率が高かったトップ 16 の広告画像

実験 2 クリック率予測

ID によるモデル、CNN によるモデル、ID にユーザ属性をかけたモデルの性能を AUC で比較した結果を図 7 に示す。図 7 の User_add は、未知のユーザを画像特徴量によって完全に推定できたと仮定したときの AUC である。もし、画像特徴量から完全に推定ができた場合、ID とかけあわせることで、コールドスタート時の予測精度を大きく向上させると期待できる。本稿執筆時点で、実際に画像特徴量から推定したユーザ属性と ID をかけあわせた場合の性能評価は行っておらず、今後の課題としたい。提案手法のクリック率予測モデルでは、素性として直接画像特徴量を用いるのではなく、推定したユーザ属性と ID をかけあわせたスパースな素性を用いることで計算コストを抑えることが可能である。

5. まとめと今後の課題

本研究では、計算コストを抑えつつコールドスタート問題に対応可能なクリック率予測手法として、広告画像の画像特徴量をあらかじめユーザ属性にマッピングし、推定されたユーザ属性と ID をかけあわせたものを予測モデルの素性として用いることでクリック率を予測する手法を提案した。また評価実験により、コールドスタート時における画像特徴量からのユーザ属性推定手法の有効性を確認した。提案手法は、広告画像の画像特徴量を用いた内容ベースのアプローチなので、コールドス

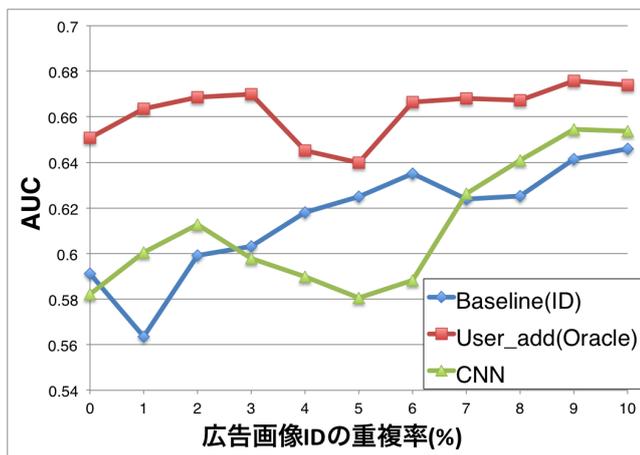


図 7 クリック率予測の AUC 比較

タート時にも対応性があり、クリック率予測モデルの素性には、推定されたユーザ属性と ID をかけあわせたスパースな素性を用いるため、計算コストを抑えることができる。今後の課題としては、より最適な画像特徴量のマッピング法や、他のモダリティの特徴量との統合を検討したい。

文 献

[1] A. Ghosh, B. I. P. Rubinstein, S. Vassilvitskii and M. Zinkevich. Adaptive bidding for display advertising. In Proc. WWW, pp. 251-260, 2009.

[2] D. G. Goldstein, R. P. McAfee and S. Suri. The effects of exposure time on memory of display advertisements. In Proc. EC, pp. 49-58, 2011.

[3] S. Balseiro, J. Feldman, V. Mirrokni and S. Muthukrishnan. Yield optimization of display advertising with ad exchange. In Proc. EC, pp. 27-28, 2011.

[4] H. Cheng, R. Rosales and E. Manavoglu. Postclick conversion modeling and analysis for non-guaranteed delivery display advertising. In Proc. WSDM, pp. 293-302, 2012.

[5] H. Cheng, R. V. Zwol, J. Azimi, E. Manavoglu, R. Zhang, Y. Zhou and V. Navalpakkam. Multimedia features for click prediction of new ads in display advertising. In Proc. KDD, pp. 777-785, 2012.

[6] D. Maltz and K. Ehrlich. Pointing the way: active collaborative filtering. In Proc. SIGCHI, pp. 202-209, 1995.

[7] K. Weinberger, A. Dasgupta, J. Langford, A. Smola and J. Attenberg. Feature hashing for large scale multitask learning. In Proc. ICML, pp. 1113-1120 2009.

[8] O. Chapelle, E. Manavoglu and R. Rosales. Simple and scalable response prediction for display advertising. In ACM TIST, 2013.

[9] L. Yan, W.-J. Li, G.-R. Xue and D. Han. Coupled group lasso for web-scale ct. prediction in display advertising. In Proc. ICML, 2014.

[10] D. Goldberg, D. Nichols, B. M. Oki and D. Terry. Using collaborative filtering to weave an information tapestry. In CACM, Vol. 35, No. 12, 1992

[11] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In Proc. CSCW, pp. 175-186, 1994.

[12] N. J. Belkin and W. B. Croft. Information filtering and information retrieval: two sides of the same coin? In CACM, Vol. 35, No. 12, pp. 29-38, 1992.

[13] J. Azimi, R. Zhang, Y. Zhou, V. Navalpakkam, J. Mao and X. Fern. The impact of visual appearance on user response in online display advertising. In Proc. WWW Companion, pp.457-458, 2012.

[14] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. In IJCV, Vol. 42, No. 3, pp. 145-175, 2001.

[15] F. Perronnin, J. Sanchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In Proc. ECCV, pp. 143-156, 2010.

[16] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner. Gradient-based learning applied to document recognition. In Proc. IEEE, Vol 86, No. 11, pp. 2278-2324, 1998.

[17] ICME2014. Tencent Multimedia Ads pCTR Challenge. <http://www.icme2014.org/tencent-multimedia-ads-pctr-challenge/>.

[18] D. G. Lowe. Object recognition from local scale-invariant features. In Proc. ICCV, pp. 1150-1157,1999.

[19] G. J. Burghouts and J. M. Geusebroek. Performance evaluation of local color invariants. In CVIU, Vol. 113, No. 1, pp. 48-62, 2009.

[20] K. E. A. van de Sande, T. Gevers and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. In IEEE TPAMI, Vol. 32, No.9, pp. 1582-1596, 2010.

[21] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray. Visual categorization with bags of keypoints. In Proc. ECCV Workshop on Statistical Learning in Computer Vision, pp. 1-22, 2004.

[22] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell. Caffe: convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093, 2014.

[23] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. <http://caffe.berkeleyvision.org/>.

[24] A. Krizhevsky, I. Sutskever and G. Hinton. Imagenet classification with deep convolutional neural networks. In Proc. NIPS, pp. 1097-1105, 2012.

[25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei. Imagenet large scale visual recognition challenge. arXiv preprint arXiv:1409.0575, 2014

[26] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell. Decaf: a deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531, 2013