

# 美容品レビューのクリック予測に向けた マルチモーダルデータの利用

平尾 礼央<sup>1,2,a)</sup> 小町 守<sup>1</sup> 岡 照晃<sup>1</sup>

**概要:** 商品レビューサイトのような CGM (Consumer Generated Media) サービスの維持には、ユーザの継続的獲得・確保が求められ、ユーザの欲するコンテンツを常に把握し提供し続けていくことが重要である。類似の分野として、ログデータに基づくオンライン広告のクリック率 (Click-Through Rate; CTR) 予測の研究がある。これに対し CGM ではサービス内に掲載されているテキストの情報や、表示されている画像の情報もユーザ行動に強く影響すると考えられる。そこで本稿では、美容品レビューサービス LIPS を対象に、ログデータだけでなくサービス内の画像データやテキストデータも取り入れた CTR 予測モデルを構築し、それぞれの情報が CTR に与える影響について分析を行った。

## Using Multimodal Data to Predict Clicks in Beauty Product Reviews

**Abstract:** In order to maintain Consumer Generated Media (CGM) services such as product review sites, it is important to understand user's interests and provide the content that users want in order to continuously acquire and keep users. In the field of online advertisement, there are many researches on predicting Click-Through Rate (CTR) using metadata based on system logs. However, in the field of CGM, unlike advertisement, we cannot ignore the influence of the attractiveness of the text and images displayed in the service on user behavior. In this paper, we developed a CTR prediction model for a beauty product review service, LIPS, that incorporates not only metadata but also images and texts in the service, and analyzed the effectiveness of each piece of information.

### 1. はじめに

CGM (Consumer Generated Media) サービスとは、掲示板やレビューサイトのような、ユーザがコンテンツを作成していくメディアのことである。CGM サービスでは、ユーザの継続的獲得・確保に向け、ユーザが求めるコンテンツを常に把握し、提供し続けていくことが重要であり、ユーザに求めるコンテンツを提供できているかを測る指標としてコンテンツのクリック率 (Click Through Rate; CTR) が挙げられる。CTR は  $\frac{\text{クリック数}}{\text{閲覧数}}$  で表すことができ、CTR が高いほど表示しているコンテンツに興味を持って貰える可能性が高く、閲覧候補を選択するシステムとして良いものであると言える。CTR は主に推薦システムや検索システム、オンライン広告配信システムで使用さ



図 1 LIPS アプリ内の投稿一覧画面

Fig. 1 Post list screen in the LIPS application.

<sup>1</sup> 東京都立大学  
Tokyo Metropolitan University

<sup>2</sup> 株式会社 AppBrew  
AppBrew, Inc.

a) hirao-reo@ed.tmu.ac.jp

表 1 CTR 予測に使用した素性一覧  
Table 1 List of features used for CTR prediction.

素性の種類	素性に紐づく情報	素性名
数値	投稿	投稿されてからの日数, いいね数, コメント数, クリップ数
	投稿者	投稿者の獲得いいね数, 投稿数, フォロワー数, フォロイー数
	閲覧者	閲覧者の獲得いいね数, 投稿数, フォロワー数, フォロイー数
カテゴリ	投稿	投稿ジャンル, プロモーションステータス
	投稿者	投稿者の年代, 肌質, 収益化ステータス, アカウント種別
	閲覧者	閲覧者の年代, 肌質, 収益化ステータス, アカウント種別, 使用デバイス

れる指標であり, サービスのクオリティや収益に大きな影響を与えるためビジネス上の重要な指標である. ユーザ数が多い大規模なサービスは, CTR の予測精度の少しの改善でさえ全体の収益を大きく増加させる可能性があるため, 近年は学術領域に限らず, ビジネス領域でも盛んに研究が行なわれている [1], [2]. 本研究では自社\*1の CGM サービス LIPS (図 1) を対象にデータの収集, CTR 予測の実験を行なった.

CTR 予測に用いるデータは一般的に WEB サービスやアプリケーションで発生するログデータが多いが, ユーザ, コンテンツが大規模であることによりデータスパースなことが多い. スパースなデータを効率的に処理するモデルの研究が進み [3], [4], [5], [6], 近年は精度の高い CTR 予測を行なうことが可能となった. これらのモデルはユーザ, コンテンツのカテゴリデータや, 「いいね数」のような数値データのようなシステムのログデータを使用しているが, CGM サービスを対象とした場合, このようなログデータだけでは不十分である. CGM サービスは様々なユーザがコンテンツを作成するため, サムネイルや文章でユーザごとの個性が出やすく, 実際にユーザに閲覧され, 大きな影響を与えているにも関わらず, これまでのモデルだと十分に活用されていない. 画像データやテキストデータを使用した CTR 予測が行なわれない理由として, 研究利用可能なデータに画像とテキストが含まれていないことが挙げられる. 本研究では, 実際に LIPS サービス内に掲載されているテキストや表示されている画像を含め CTR の予測を行なった.

本稿では, 画像とテキストを利用した CTR 予測に焦点を当て, CTR 予測で重要度の高いデータを調査し, ユーザが関心を持つ確率が高いコンテンツ (投稿) にどのような傾向があるかを分析, システムに組み込むことでユーザ体験を改善することを目指す.

## 2. 関連研究

近年の CTR 予測及び CVR (Conversion Rate) 予測は主にオンライン広告の領域で利益を最大化するために行な

われており, 学術分野だけでなくビジネス分野でも盛んに研究されている [1], [2]. オンライン広告では, どの広告がクリックやコンバージョン (e.g. 購買行動) に繋がるかが重要であり, 広告を出す前に事前に CTR を予測することができればよりユーザの嗜好に合った広告をシステムが選択して提示でき, 費用対効果を最大化することができる. また, 作成した CTR 予測モデルが重要視しているデータが何であるかを分析することで, どのようなクリエイティブが利益を出しやすいかを予想することができるため, 広告クリエイティブの補助の役割として CTR, CVR 予測を用いる研究 [7] も行なわれている.

CTR 予測はロジスティック回帰等の線形モデルを使用した研究 [8] から始まり, その後スパースなデータに対しても相互作用を考慮できる FM (Factorization Machines) [3] が提案された. LR と FM は現在もベースライン手法として比較されている. 特に FM は CTR 予測だけでなく, ランキングや推薦問題など様々な問題に適用できるモデルとして一般化されている.

近年は Deep Learning を活用した予測モデル [4] の研究が盛んで, Deep Learning と FM を組み合わせたモデルや更に発展させたモデルが提案されている. 画像やテキスト分野の Deep Learning モデルの進化に伴い, CTR 予測の分野でもそれらのデータを使う手法が提案されている. 画像を CTR 予測に活用する手法として, 広告画像を CNN で埋め込みベクトルとして扱い, CTR 予測を行なう手法 [9] がある. テキストの場合は広告クリエイティブの補助的な役割として CVR を予測し, テキストのどの部分がどの属性のユーザに対して重要度が高いかを Conditional Attention によって推定する手法 [7] が提案されている. これらの研究は画像, テキストを活用することで性能向上しているが, 画像とテキストの両方を使用し, それぞれの相互作用を考慮した手法の検証は行なわれていない. 本研究では, 画像とテキストの両方を使用して CTR を予測し, CTR 予測における各情報の重要度の分析を行なう.

## 3. マルチモーダル CTR 予測モデル

図 2 に本研究で使用したモデルのアーキテクチャを示

\*1 株式会社 AppBrew

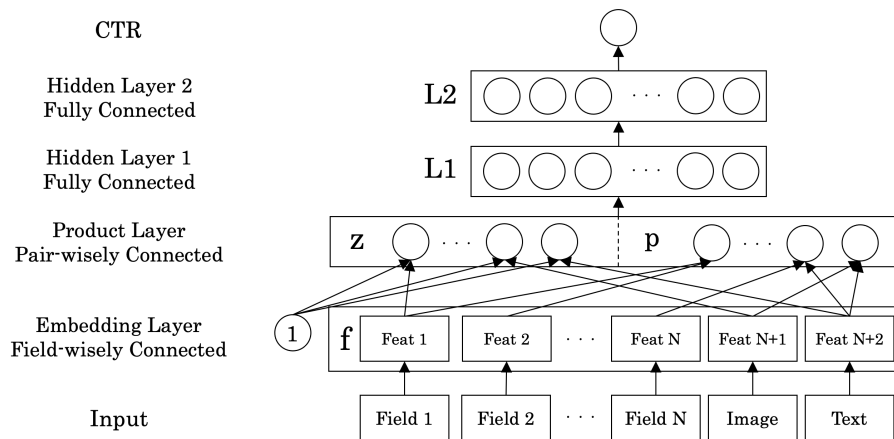


図 2 本研究で用いたマルチモーダルデータを利用した CTR 予測モデルアーキテクチャの概要  
Fig. 2 Overview of the multimodal data-based CTR prediction model architecture used in this study.

す。このモデルは先行研究の PNN (Product-based Neural Networks) [10] を参考に、シンプルな形で画像とテキストをモデルに組み込めるようにした。図 2 では、カテゴリと数値データからなる素性、画像、テキストが入力として与えられ、 $f$  層で埋め込みベクトルに変換される。この時、LIPS のアプリ画面では図 1 のように投稿のサムネイル画像と本文の冒頭約 30 文字が表示されるため、サムネイル画像を埋め込みベクトル化する Image Encoder と本文冒頭を埋め込みベクトル化する Text Encoder を使用し、ベクトル化を行なった。次に各埋め込みベクトルはバイアスのみ加算される  $z$  層と、 $f$  層の各 2 つの埋め込みベクトルの組み合わせから求められる内積及び外積からベクトルを生成する  $p$  層に入力される。 $z$  層は各データ単体の属性を表現し、 $p$  層は各データが及ぼす相互作用を表現する層になっている。 $z$  層と  $p$  層で計算されたベクトルは、最後に 2 つの全結合層を経由して CTR の予測値として出力される。

## 4. 実験

### 4.1 データセット

本研究では、株式会社 AppBrew の協力により、同社が運営する美容品レビューサービス LIPS<sup>\*2</sup> のデータを使用した。LIPS はアプリの累積ダウンロード数 700 万、月間ユーザー数約 1,000 万のサービスで、購入した美容品のレビューをユーザーが作成するメディアである。主なユーザーは 10 代後半から 20 代の女性で、毎日約 3,000 件のレビューが作成されている。

実験では、LIPS の iOS/Android アプリにおいて 2021 年 6 月 26 日から 2021 年 7 月 2 日の期間で閲覧、タップされたログデータを対象として評価を行なった。このデータのうち、データの偏りをなくするため表示回数が 100 回を超える投稿は使用するデータを 100 件以内に制限し、画像と本

文がない投稿は除外し、データの合計が 30 万件程度になるようにランダムに収集した。学習に 234,304 件、検証に 29,280 件、評価に 29,312 件のログデータを使用し、タップは全体の 10.7% 程度であった。ログデータに付随するメタデータを表 1 に記す。

表 1 上段は数値データの一覧であり、上から投稿、投稿者、閲覧者に紐づく情報である。下段はカテゴリデータの一覧であり、こちらも上から投稿、投稿者、閲覧者に紐づく情報が記載されている。投稿ジャンルは、コスメ、スキンケア、ヘアメイク、ネイル、ダイエット、恋愛、その他の 7 項目の中から選択され付与されている。全投稿の 8 割はコスメのカテゴリで登録されており、2 番目にスキンケア、3 番目にヘアメイクの投稿が多い。プロモーションステータスは投稿がプロモーション (宣伝) として表示されているかを表している。次にユーザーに関する素性だが、年代は 10 代以上を 10 代前半、10 代後半、20 代前半のように 50 代後半まで 5 歳単位で分けて登録し、60 代以上はまとめて 60 代以上として登録している。肌質は普通肌、脂性肌、乾燥肌、混合肌、敏感肌、アトピー肌の 6 項目の中から選択され付与されている。収益化ステータスとはユーザーがアカウントを収益化しているかを表し、アカウント種別は普通のアカウント、ブランド公式アカウント、インフルエンサーのどれに分類されるかを表している。最後の使用デバイスは iOS か Android のいずれかとなっており、収集データの 8 割以上は iOS となっている。

### 4.2 比較手法

LIPS のデータセットを使用した既存手法の性能評価とマルチモーダルデータを使用する事による予測精度の向上を調査するため、以下のモデルと比較実験を行なった。

- **Logistic Regression (LR).** ロジスティック回帰 [8] は CTR 予測において最もシンプルなベースラインモ

\*2 <https://lipscosme.com>

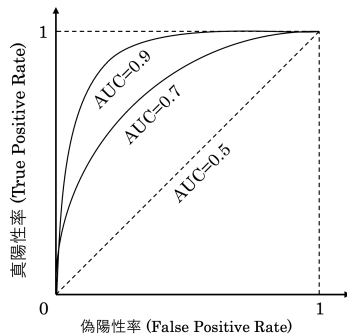


図 3 AUC スコアの例

Fig. 3 Sample image of AUC score.

デルであり, Google によって提案された FTRL [11] によりオンライン学習が可能となっている.

- **Factorization Machines (FM)**. FM [3] は LR で捉えることができなかった特徴の相互作用を密なベクトルに埋め込み, 特徴の相互作用に対応する埋め込みベクトルも内積として表現するモデルで, スパースなベクトルを扱う様々な問題に利用されている.
- **Wide&Deep**. Wide&Deep [4] は Google によって提案されたモデルで, コンテンツ間の共起を取り込む Wide 部とコンテンツの属性を取り込む Deep 部を組み合わせ, 双方を考慮した予測を行なうことができるモデルである.
- **DeepFM**. DeepFM [5] は Wide&Deep の Wide 部分に FM を取り入れ相互作用を明示的に取り込むようにしている.
- **xDeepFM**. xDeepFM [6] は Compressed Interaction Network という高次元の相互作用を表現するネットワークを使用し, FM よりも高次の相互作用を表現することができる.
- **Adaptive Factorization Network (AFN)**. AFN [12] は Logarithmic Transformation Layer を導入し, 高次のノイズを取り除くことを目指したモデルである.
- **Product-based Neural Networks (PNN)**. PNN [10] は特徴ベクトル同士の内積及び外積を Deep Neural Network に入力することで相互作用を表現したモデルである.

#### 4.3 評価方法

CTR 予測の精度を測る評価指標として, Area Under the Curve (AUC) [13] と LogLoss [14] を使用した. これらの評価指標は CTR 予測モデルの性能評価でよく使用される指標であり, 公開データセットの SoTA の比較は主に AUC を用いて行なわれる. 図 3 に AUC の計算に利用する ROC 曲線及び対応する AUC のスコアを示す. AUC は ROC 曲線の下側の面積のことで, ランダムに選択した場合 0.5 と

表 2 LIPS データセットを用いた CTR 予測の実験結果

Table 2 Experimental results of CTR prediction using the LIPS dataset.

Model	AUC	LogLoss
LR	0.5423	0.3588
FM	0.5997	0.3144
Wide&Deep	0.5958	0.3159
DeepFM	0.6004	0.3120
xDeepFM	0.6065	0.3108
AFN	0.5916	0.3166
PNN	0.6121	0.3086
PNN +Image	0.6134	0.3084
PNN +Text	<b>0.6150</b>	<b>0.3075</b>
PNN +Image +Text	0.6126	0.3080

なり, 性能が向上するほど 1 に近づいていく. ROC 曲線は縦軸に真陽性率, 横軸に偽陽性率を取る曲線であり, 陽性と陰性の比率に偏りがある場合の評価においてデータの偏りを考慮した評価を行なうことができる. Logloss の計算式を式 (1) に示す.

$$\text{LogLoss} = -\frac{1}{N} \sum_{i=1}^N (y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i)) \quad (1)$$

式 (1) において,  $N$  はデータ数,  $y_i$  は  $i$  番目の正解ラベル,  $\hat{y}_i$  は  $i$  番目の予測確率を表している. LogLoss は Binary Cross Entropy (BCE) と同義であり, BCE は本研究の学習で損失関数として使用した. AUC は大きい方が優れ, LogLoss は小さい方が優れている.

#### 4.4 実験設定

マルチモーダルデータを使用するにあたり, Text Encoder として日本語学習済みの DistilBERT [15] を使用した<sup>\*3</sup>. このモデルは日本語データで学習済みの BERT-base を教師モデルとして, 蒸留を行なっている. 指定した長さのベクトルを得るため, テキストをトークン化する際に先頭に付与される [CLS] トークンに対応する出力ベクトルに対して MLP (Multi Layer Perceptron) を使用し, 出力する次元数を変更した. Image Encoder は ImageNet で学習済みの ResNet18 を使用し, 最終層の出力次元数を変更した. これらの Encoder で生成する埋め込みベクトルと画像とテキスト以外の素性の埋め込みベクトルの次元数は 16 で統一した.

全てのモデルの学習において, 損失関数として LogLoss を使用し, パラメータの最適化には Adam [16] を使用した. 学習率は 0.001 から始め, 学習が進むにつれて徐々に小さくなるように設定した<sup>\*4</sup>. また, 過学習されないように early stopping を導入し, 3 エポック以上 validation の

<sup>\*3</sup> <https://github.com/BandaiNamcoResearchInc/DistilBERT-base-jp>

<sup>\*4</sup> 5 エポックで 1/10 になるように設定

表 3 各素性の入力をマスクした場合の AUC と LogLoss の変化

Table 3 Differences in AUC and LogLoss when the input of each feature is masked.

素性名	AUC Diff	LogLoss Diff
投稿のクリップ数	-0.0145	0.0056
投稿からの日数	-0.0113	0.0013
デバイスの種類 (iOS/Android)	-0.0096	0.0012
閲覧ユーザのフォロワー数	-0.0075	0.0010
閲覧ユーザの獲得いいね数	-0.0071	0.0011
投稿のコメント数	0.0007	0.0001
投稿者の年代	0.0007	-0.0001
投稿者の収益化ステータス	0.0010	0.0140
投稿者の投稿数	0.0012	-0.0003
投稿者の獲得いいね数	0.0016	-0.0005

AUC が改善されない場合、最も AUC が高いパラメータが保存されるようにした。各モデルの性能を評価するにあたり、異なるシード値で 5 回実験を繰り返し、その平均 AUC 及び LogLoss で評価を行なった。

#### 4.5 実験結果

LIPS データセットを用いた CTR 予測の AUC 及び LogLoss のスコアを表 2 に示す。表の下 3 つの提案手法であるマルチモーダルデータを利用した CTR 予測モデルは、いずれも他の比較手法と比べて AUC, LogLoss の双方で優れていることが確認できた。

今回使用したデータだと LR の AUC スコアが極端に悪く、他のモデルは 0.59 から 0.62 の間に収まった。先行研究で提案されていた手法の中では PNN が最も AUC, LogLoss のスコアが最も優れていた。Wide&Deep や AFN が他の手法と比べてスコアが悪いことから、本実験で使用したデータには比較的シンプルな構造を持つモデルが有利であると考えられる。

PNN に画像、テキストをそれぞれ単体で追加したモデルは、画像単体よりテキスト単体で追加したモデルの方が優れていた。画像とテキストの両方を追加したモデルは双方を追加していない PNN よりは優れているが、画像、テキストを単体で追加したモデルには劣る結果になった。画像とテキストを CTR 予測に組み込むと精度向上を見込めるが、二つを同時に取り扱う場合はどちらかが予測の妨げにならないか注意が必要であると言える。

### 5. 考察

#### 5.1 CTR 予測における素性について

マルチモーダルデータが AUC 及び LogLoss のスコアに与える影響は表 2 で確認することができた。画像とテキスト以外の素性の影響を調査するため、画像とテキストを使用しない PNN において、推論時に各素性の埋め込みベク

表 4 Encoder の学習を行わない場合の AUC と LogLoss の変化

Table 4 Difference in AUC and LogLoss without encoder training.

Model	AUC Diff	LogLoss Diff
PNN +Image	-0.00018	-0.00027
PNN +Text	0.00066	0.00008
PNN +Image +Text	0.00145	0.00004

トル (図 2 中 f の部分) を 0 ベクトルでマスクし、使用しないことでそれぞれの素性が AUC と LogLoss にどの程度影響を与えているかを測定した。AUC において変化が大きかった上位 5 つ、下位 5 つの素性を表 3 に示す。AUC は大きい程予測の精度が高いため、その素性を除いた時の AUC の下り幅が大きい素性は重要度が高い素性と言える。それぞれの素性について考察を行なう。

最も重要度が高い素性は投稿のクリップ数で、クリップとは選択した投稿を後から確認するための機能である。比較的いいねが気軽に付けられるのに対して、クリップは気になる投稿、再度確認したい投稿を保存しておく用途で使われる。投稿の質とクリップ数に相関があるのがわかる。投稿からの日数は、2 番目に重要度が高い素性であり、美容品は流行りや季節性があるため、投稿の鮮度が重要であるためであると考えられる。3 番目に重要度が高い素性であるデバイスの種類については、既知の情報として iOS ユーザの方が Android ユーザに比べて 30% 程 CTR が高い事が挙げられる。これは iOS/Android 間のユーザー層の違いや機能の差によるものであると考えられる。4,5 番目の閲覧ユーザのフォロワー、いいね数は相互フォローの知り合い間でいいねされたりいいねし返すという流れで、知り合いが多いユーザーほどタップが増える傾向にあるためであると考えられる。表にはないが、6~9 番目まで閲覧ユーザの素性が固まっており、CTR 予測において閲覧ユーザの情報は重要度が高い傾向がある事がわかる。

次に重要度が低い素性に注目すると、下から 1 から 4 番目の素性は投稿者に関する素性だった。投稿者の獲得いいね数や投稿数は投稿者の知名度との相関が強く、クリック率に及ぼす影響が大きいと考えていたが、本実験で使用した他の素性と比べるとあまり重要ではないことが確認できた。下から 5 番目の投稿のコメント数は、投稿一覧の画面においてコメント数が確認できるが CTR に与える影響は小さいと考えられる。

現在の投稿一覧画面ではクリップ数が確認できないが、新たにクリップ数の項目を表示することで効率良くユーザーを興味ある投稿に導くことができると考えられる。

#### 5.2 Encoder の学習について

画像とテキストデータの双方を使用したモデルの改良方法として、学習時のパラメータを減らすため Image Encoder



図 4 入力画像において注目されている部分の可視化

Fig. 4 Visualization of the parts of the input image that are of interest.

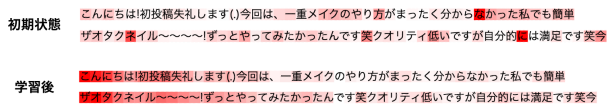


図 5 入力テキストにおいて注目されている部分の可視化

Fig. 5 Visualization of the parts of the input text that are of interest.

と Text Encoder のパラメータを事前学習モデルのパラメータで固定，埋め込みベクトル以降のパラメータを学習するように設定し，検証を行なった。

表 4 に Encoder のパラメータを固定した場合に固定しない場合と比較して AUC と LogLoss にどの程度変化があったかの数値を示す。画像データのみを追加した PNN モデルは Image Encoder のパラメータを固定しても AUC は改善せず，テキストデータのみを追加したモデルは Text Encoder のパラメータを固定することで AUC が僅かに改善した。画像，テキストの双方を追加した PNN モデルは双方の Encoder のパラメータを固定することで AUC が改善した。この結果から表 2 において画像，テキスト双方を使用した PNN モデルの AUC スコアがよくなかったのは，学習パラメータの大幅な増加により学習が上手くできなかったためであると考えられる。

今後の改善点として，公開されている学習パラメータを用いた Encoder では，LIPS のデータセットとドメインが違いすぎる可能性が高いため，適したドメインになるように再学習を行ない，そのパラメータで固定した Encoder を用いてマルチモーダルデータを利用した CTR 予測を行なうことが挙げられる。

### 5.3 画像とテキストの注目されている部分について

画像では，学習したパラメータの重みから画像の CTR

予測値を高めている（注目されている）部分の可視化を行なった。図 4 に入力画像において注目されている部分を可視化した例を示す。図中の上 2 つの画像は，画像内のテキスト部分が最も注目されており，下 2 つの画像は瞳の部分が最も注目されている。これらの結果は図で示した画像以外のデータでも同様の傾向が見られた。この可視化の結果から，引きが強いテキストや実際にメイクを行なった部分（特に瞳）がクリックに寄与している可能性が高いことがわかる。

テキストでは，DistilBERT の最終層における [CLS] トークンに対する各トークンの attention を出力することで，テキスト内のどの部分が埋め込みベクトル生成における重要度の高さを調査した。図 5 に入力テキストにおいて注目されている部分を可視化した例を示す。初期の状態ではテキスト内の最初から最後まで満遍なく注目されているのに対して，学習後を見ると最初の部分に集中している。テキストの最初の部分が注目される要因として，最初の方に投稿の内容が記されていることが多いことや，投稿一覧画面では 1 行目以降はクリックして確認することが多いため，2 行目以降の影響は少ないことが考えられる。

## 6. おわりに

本稿では CTR 予測における画像とテキストの必要性に注目し，美容品レビューアプリ LIPS のシステムログを用いて素性の評価を行なった。画像とテキストを CTR 予測に組み込むのは精度向上を見込める一方で，画像テキスト双方を使う場合は必ずしも精度が向上するとは限らないことがわかった。また，画像，テキスト以外の素性に関する調査も実施し，CTR に影響を与える可能性が高い素性を特定することができた。今後の課題として，画像とテキストの両方を使用して精度向上できるより複雑なモデルを用いた予測や，CTR 予測を活用したシステム改善の導入が挙げられる。

## 参考文献

- [1] Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. Introduction to computational advertising. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL 2008), June 15-20, 2008, Columbus, Ohio, USA, Tutorial Abstracts*, p. 1. The Association for Computer Linguistics, 2008.
- [2] Shuai Yuan, Ahmad Zainal Abidin, Marc Sloan, and Jun Wang. Internet advertising: An interplay among advertisers, online publishers, ad exchanges and web users. *CoRR*, Vol. abs/1206.1754, , 2012.
- [3] Steffen Rendle. Factorization machines. In Geoffrey I. Webb, Bing Liu, Chengqi Zhang, Dimitrios Gunopulos, and Xindong Wu, editors, *Proceedings of The 10th IEEE International Conference on Data Mining (ICDM 2010), Sydney, Australia, 14-17 December 2010*, pp. 995–1000. IEEE Computer Society, 2010.
- [4] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal

- Shaked, Tushar Chandra, Hrishi Aradhya, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. Wide & deep learning for recommender systems. In Alexandros Karatzoglou, Balázs Hidasi, Domonkos Tikk, Oren Sar Shalom, Haggai Roitman, Bracha Shapira, and Lior Rokach, editors, *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016*, pp. 7–10. ACM, 2016.
- [5] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: A factorization-machine based neural network for CTR prediction. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pp. 1725–1731. ijcai.org, 2017.
- [6] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. xDeepFM: Combining explicit and implicit feature interactions for recommender systems. In Yike Guo and Faisal Farooq, editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pp. 1754–1763. ACM, 2018.
- [7] Shunsuke Kitada, Hitoshi Iyatomi, and Yoshifumi Seki. Conversion prediction using multi-task conditional attention networks to support the creation of effective ad creatives. In Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis, editors, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pp. 2069–2077. ACM, 2019.
- [8] Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through rate for new ads. In Carey L. Williamson, Mary Ellen Zurko, Peter F. Patel-Schneider, and Prashant J. Shenoy, editors, *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pp. 521–530. ACM, 2007.
- [9] Junxuan Chen, Baigui Sun, Hao Li, Hongtao Lu, and Xian-Sheng Hua. Deep CTR prediction in display advertising. In Alan Hanjalic, Cees Snoek, Marcel Worring, Dick C. A. Bulterman, Benoit Huet, Aisling Kelliher, Yiannis Kompatsiaris, and Jin Li, editors, *Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016, Amsterdam, The Netherlands, October 15-19, 2016*, pp. 811–820. ACM, 2016.
- [10] Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. Product-based neural networks for user response prediction. In Francesco Bonchi, Josep Domingo-Ferrer, Ricardo Baeza-Yates, Zhi-Hua Zhou, and Xindong Wu, editors, *Proceedings of IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*, pp. 1149–1154. IEEE Computer Society, 2016.
- [11] H. Brendan McMahan, Gary Holt, David Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad click prediction: a view from the trenches. In Inderjit S. Dhillon, Yehuda Koren, Rayid Ghani, Ted E. Senterator, Paul Bradley, Rajesh Parekh, Jingrui He, Robert L. Grossman, and Ramasamy Uthurusamy, editors, *Proceedings of The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013*, pp. 1222–1230. ACM, 2013.
- [12] Weiyu Cheng, Yanyan Shen, and Linpeng Huang. Adaptive factorization network: Learning adaptive-order feature interactions. In *Proceedings of The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 3609–3616. AAAI Press, 2020.
- [13] Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.*, Vol. 30, No. 7, pp. 1145–1159, 1997.
- [14] Christopher M. Bishop. *Pattern recognition and machine learning, 5th Edition*. Information science and statistics. Springer, 2007.
- [15] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, Vol. abs/1910.01108, , 2019.
- [16] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *Conference Track Proceedings of 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015*, 2015.