

# 広告クリエイティブ評価のための深層確率埋め込みの学習

## Learning Deep Probabilistic Embedding for Ad Creative Evaluation

北田 俊輔<sup>\*1</sup> 彌富 仁<sup>\*1</sup> 関 喜史<sup>\*2</sup>  
Shunsuke Kitada Hitoshi Iyatomi Yoshifumi Seki

<sup>\*1</sup>法政大学 大学院 理工学研究科 応用情報工学専攻  
Major in Applied Informatics, Graduate School of Science and Engineering, Hosei University

<sup>\*2</sup>株式会社 Gunosy  
Gunosy Inc.

The performance and quality of ad creatives in online advertising can generally be evaluated by the rate at which the creatives are clicked on by users, and action rate such as the purchase of products are taken. On the other hand, such user behavior is highly noisy and uncertain, and it often depends on the potential attractiveness of the ad campaign and the target product of the ad. In this paper, we propose a new method for learning probabilistic ad creative embeddings to evaluate the ad creatives, which represent each ad creatives in the campaign as a Gaussian distribution in the latent space. Our probabilistic embedding can properly capture features of the creative from uncertain user behaviors and multiple ad creatives associated with the campaign. We evaluated our proposal using the record of real-world 200,000 ad creatives provided by Gunosy Inc. We confirmed that our probabilistic embeddings can accurately capture the serving performance of the ad creatives regardless of the text encoder (e.g., LSTM, BERT). Furthermore, we observed that our proposal achieves prediction with small uncertainty by using BERT, which is the recent state-of-the-art model.

### 1. はじめに

広告クリエイティブ (以下、クリエイティブ) は購買対象となる顧客に対して効果的に商品の情報を届ける重要な役割を担っている。クリエイティブの良し悪しは、表示されたクリエイティブをユーザーがクリックする割合 (click through rate; CTR) や、クリック後に商品を購入する等の行動が起きる割合 (conversion rate; CVR) で評価するのが一般的である。一方で、こうしたユーザー行動は非常にノイズが多く、不確実性が高い。先行研究における広告の CTR 予測 [Chen 16] や CVR 予測 [Lu 17] では、学習される広告表現は点推定が主であり (図 1a を参照)、こうした不確実性に脆弱であることが考えられる。不確実性を考慮するために、確率的な埋め込みを導入する試みが推薦システム [Jiang 19] や、広告の CTR 予測モデル [Liu 18] で検討されている。我々は今回、確率的な埋め込みを広告クリエイティブの性能予測に導入することを試みる。<sup>\*1</sup>

本研究では、クリエイティブの評価において発生する 2 つの不確実性 (1) ユーザー行動による不確実性、(2) 1 つのキャンペーンに紐づく複数のクリエイティブの不確実性に対して取り組むために、確率的広告クリエイティブ埋め込み (*Probabilistic Ad Creative Embedding*; PACE) を提案する。埋め込み表現において不確実性を考慮する研究として、確率埋め込みを元にした顔画像埋め込み表現の学習 [Shi 19] がある。本研究ではこの研究を元に図 1b のような入力される広告クリエイティブの埋め込み表現を潜在空間上において点ではなく確率分布として得られるように予測モデルを訓練する。この分布の平均は最も可能性の高い潜在表現であると解釈でき、分布の分散は予測結果の不確実性を表す。我々の提案手法は確率埋め込みを用いる既存の CTR 予測モデル [Liu 18] や推薦システム [Jiang 19]

連絡先: 北田 俊輔, 法政大学 大学院 理工学研究科 応用情報工学専攻, 東京都小金井市梶野町 3-7-2 S603, [shunsuke.kitada.8y@stu.hosei.ac.jp](mailto:shunsuke.kitada.8y@stu.hosei.ac.jp)

<sup>\*1</sup> 広告の CTR 予測は配信されている広告のユーザーに対する予測 CTR を推定するのに対し、広告クリエイティブの性能予測は配信前の広告の性能を予測するものであり、タスクとして大きく異なる。

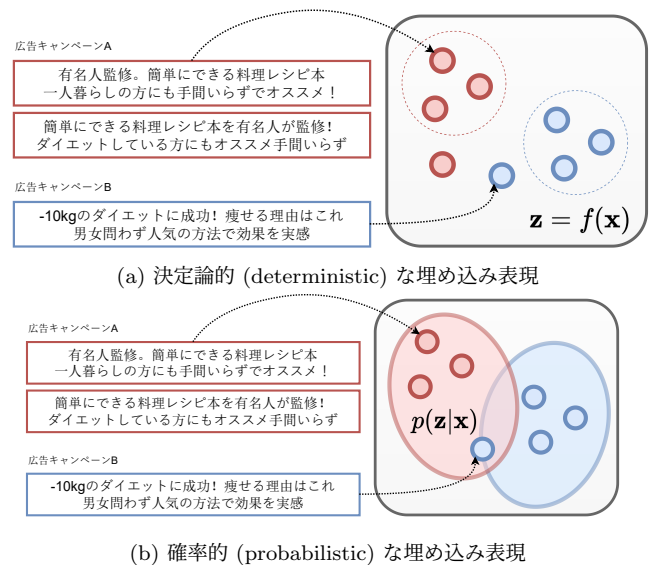


図 1: 広告クリエイティブにおける、決定論的 (deterministic) な埋め込み表現と確率的 (probabilistic) な埋め込み表現の違い。決定論的な埋め込み表現は潜在空間中で各広告クリエイティブを点として表現する。一方で確率的な埋め込み表現は潜在空間中で広告クリエイティブの特徴を分布として表現する。

と比べて確率埋め込みに対するサンプリングを必要としない、非常にシンプルかつ効率の良い埋め込みを実現できる。PACE によって推定された不確実性を利用して、入力される広告クリエイティブの多様性の検討や、高い配信効果を見込める広告クリエイティブの作成支援の可能性が期待できる。本研究の貢献は次のとおりである: (1) クリエイティブの埋め込み表現を点推定するのとは比べて、確率分布を推定することで不確実性に頑健なクリエイティブの評価が可能になった (2) 提案法においてより高性能な特徴抽出モデルを使うことでクリエイティブの評価の不確実性が小さくなることを確認した。

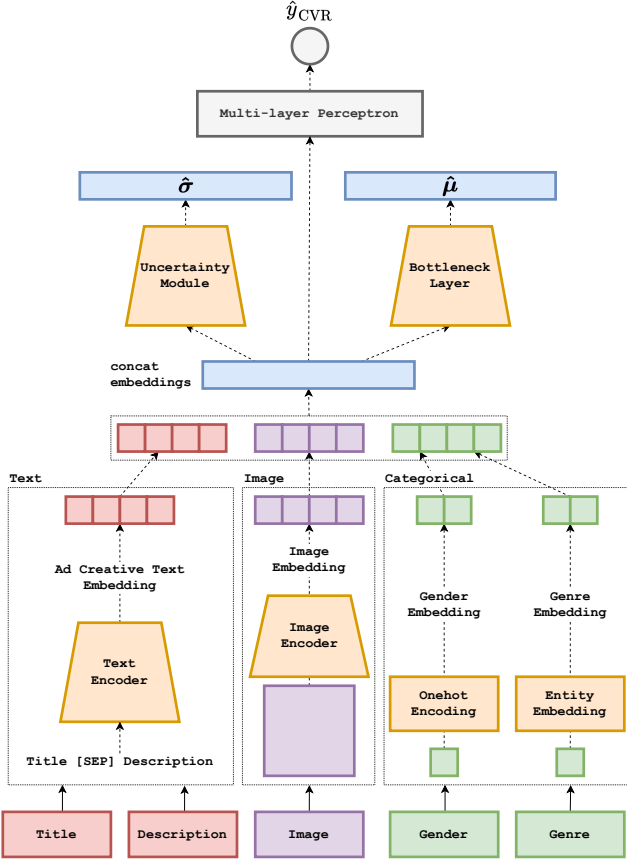


図 2: 広告確率埋め込み (probabilistic ad creative embedding; PACE) を含むネットワーク。不確実性機構 (uncertainty module) と ボトルネック層 (bottleneck layer) により確率的埋め込み表現を学習する。

## 2. 確率的広告クリエイティブ埋め込み表現

### 2.1 決定論的な埋め込み表現の限界

本節では、決定論的埋め込みの問題点を経験的および理論的の両面から説明する。まず  $\mathcal{X}$  を入力空間、 $\mathcal{Z}$  を  $D$  次元の潜在特徴空間とする。理想的な潜在特徴空間  $\mathcal{Z}$  は、入力に対して特有で顕著な特徴のみを符号化し、特有さは無関係な特徴から切り離されていなければならない。このように、特有さは各広告クリエイティブを最もよく表す固有の埋め込み  $z \in \mathcal{Z}$  を表すべきであり、各広告クリエイティブ  $\mathbf{x} \in \mathcal{X}$  は  $p(\mathbf{x}|z)$  からサンプリングされた観測である。広告クリエイティブの埋め込みにおける学習過程は、このような潜在特徴空間  $\mathcal{Z}$  の探索と逆写像  $p(z|\mathbf{x})$  の学習が共存する過程であるとみなすことができる。確率的な埋め込みについては、逆写像はディラックのデルタ関数  $p(z|\mathbf{x}) = \delta(z - f(\mathbf{x}))$  であり、ここで  $f$  は埋め込み関数である。どのような空間  $\mathcal{Z}$  でも、 $\mathbf{x}$  のノイズの可能性を考えると、どれだけ多くの訓練データがあっても正確な  $z$  を推測することは非現実的である。

### 2.2 確率的広告クリエイティブ埋め込み表現の概要

本研究では、図 2 に示すネットワークを元にして、入力である広告クリエイティブ  $\mathbf{x}$  から配信情報を用いない形で確率的な広告埋め込み表現を学習する。具体的には、ネットワークの入力としてテキスト特徴 (タイトル、説明文)、画像特徴 (バナー画像)、カテゴリーカル特徴 (配信対象の性別、広告クリエイティブのジャンル) を使用する。テキスト特徴は BERT [Devlin 19]、

画像特徴は ResNet34 [He 16] を使用した。これらの特徴量を結合したベクトルは、埋め込みを学習するボトルネック層 (bottleneck layer) と不確実性機構 (uncertainty module) へと入力され、埋め込みの中心を表すベクトルと不確実性を表すベクトルを推定する。

データの不確実性から発生する前述の問題に取り組むために、上記の入力とネットワークを元に広告埋め込み表現に対して不確実性を符号化する。具体的には、潜在空間中で点推定を与えるネットワークを構築するのではなく、提案手法を含むネットワークは本来広告クリエイティブに備わっている特性を表すように潜在空間中において分布  $p(z|\mathbf{x})$  を推定する。特に、提案手法は多変量正規分布を用いる:

$$p(z|\mathbf{x}_i) = \mathcal{N}(z; \boldsymbol{\mu}_i, \boldsymbol{\sigma}_i^2 \mathbf{I}) \quad (1)$$

ここで  $\boldsymbol{\mu}_i$  と  $\boldsymbol{\sigma}_i$  は共に  $i$  番目の入力  $\mathbf{x}_i$  に対してネットワークによって予測された  $D$  次元のベクトルである。なお、我々は学習する埋め込み表現の複雑さを減らすために対角共分散行列を考慮した。我々の確率的広告埋め込み表現は次の性質を持つことを期待できる: (1) 中心を表すパラメータ  $\boldsymbol{\mu}$  は最も広告クリエイティブを表現する特徴を入力からエンコードしている (2) 不確実性を表すパラメータ  $\boldsymbol{\sigma}$  は各特徴次元においてネットワークが出力する予測の信頼度を符号化している。上記に加えて、我々は単一モデルのみで分布を予測することを期待できる。

### 2.3 相互尤度スコアによる確率的埋め込み表現の学習

入力として同一キャンペーン内の 2 つの広告クリエイティブ ( $\mathbf{x}_i, \mathbf{x}_j$ ) を与えられると、それらが同一キャンペーンに属する (同じ潜在表現を共有する) 可能性を以下のように直接推定することができる:

$$p(\mathbf{z}_i = \mathbf{z}_j) = \int p(\mathbf{z}_i|\mathbf{x}_i)p(\mathbf{z}_j|\mathbf{x}_j)\delta(\mathbf{z}_i - \mathbf{z}_j)d\mathbf{z}_i d\mathbf{z}_j. \quad (2)$$

実際には代わりに以下で与えられるような対数尤度を利用する:

$$\begin{aligned} s(\mathbf{x}_i, \mathbf{x}_j) &= \log p(\mathbf{z}_i = \mathbf{z}_j) \\ &= -\frac{1}{2} \sum_{l=1}^D \left( \frac{(\mu_i^{(l)} - \mu_j^{(l)})^2}{\sigma_i^{2(l)} + \sigma_j^{2(l)}} + \log(\sigma_i^{2(l)} + \sigma_j^{2(l)}) \right) \\ &\quad - \text{const} \end{aligned} \quad (3)$$

ここで  $\text{const} = \frac{D}{2} \log 2\pi$  であり、 $\mu_i^{(l)}$  は  $\boldsymbol{\mu}_i$  における  $l$  次元目を指し、 $\sigma_i^{(l)}$  も同様である。

この対称的な尺度は、ある入力の潜在表現の尤度の期待値として見る事ができる。つまり、他の条件付きの入力の潜在表現の尤度の期待値としてみなすことができる:

$$\begin{aligned} s(\mathbf{x}_i, \mathbf{x}_j) &= \log \int p(\mathbf{z}|\mathbf{x}_i)p(\mathbf{z}|\mathbf{x}_j)d\mathbf{z} \\ &= \log \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z}|\mathbf{x}_i)} [p(\mathbf{z}|\mathbf{x}_j)] \\ &= \log \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z}|\mathbf{x}_j)} [p(\mathbf{z}|\mathbf{x}_i)]. \end{aligned} \quad (4)$$

[Shi 19] では、これを相互尤度スコア (mutual likelihood score; MLS) と呼んでいる。全ての不確実性が同じであると仮定した場合、2 乗ユークリッド距離は MLS の特殊なケースと等価であることが示されている [Shi 19]。

決定論的埋め込み  $f$  は適切に最適化されていれば、PACE は次のような特性を満たすことができる: (1) 埋め込み空間は

表 1: 実世界の広告クリエイティブデータセットを用いた決定論的埋め込みと我々の確率埋め込みを用いたネットワークの予測性能。

Dataset date	Text Encoder	Embedding type	Logloss ↓	NDCG ↑
2019-01-01 ~ 2020-01-01	LSTM	Deterministic	0.044 ± 0.004	79.60 ± 4.04
		<b>Probabilistic</b>	0.044 ± 0.003	<b>80.66 ± 3.11</b>
	BERT	Deterministic	0.043 ± 0.004	85.67 ± 3.06
		<b>Probabilistic</b>	0.041 ± 0.003	<b>87.27 ± 1.31</b>
2020-01-01 ~ 2021-01-01	LSTM	Deterministic	0.061 ± 0.005	76.02 ± 3.78
		<b>Probabilistic</b>	0.072 ± 0.012	<b>79.30 ± 2.80</b>
	BERT	Deterministic	0.060 ± 0.005	78.88 ± 3.61
		<b>Probabilistic</b>	0.065 ± 0.005	<b>83.04 ± 1.35</b>

入力特有で顕著な潜在空間である。(2)  $f(\mathbf{x})$  は与えられた最も可能性の高い特徴を示す。そのため、本研究では、段階的な訓練戦略を検討する。まず事前に訓練された埋め込みモデル  $f$  が与えて、そのパラメータを固定し、 $\mu(\mathbf{x}) = f(\mathbf{x})$  を得た後、不確実性機構で  $\mu(\mathbf{x})$  を推定できるようにパラメータを最適化する。不確実性機構が埋め込みネットワークの同じデータセットで訓練される場合、この段階的な訓練戦略により、end to end な訓練戦略よりも PACE と元の埋め込み  $f(\mathbf{x})$  をより公平に比較することが可能である。

不確実性機構は、ボトルネック層と同じ入力を共有する 2 つの全結合層を持つネットワークである。最適化基準は全ての  $(\mathbf{x}_i, \mathbf{x}_j)$  の相互尤度スコアを最大化することである。よって、我々は以下の損失関数を最小化する：

$$\mathcal{L} = \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} -s(\mathbf{x}_i, \mathbf{x}_j), \quad (5)$$

ここで、 $\mathcal{P}$  は対象のキャンペーン内におけるクリエイティブの組み合わせの集合であり、 $s(\dots|\dots)$  は式 (3) で定義される。実際には、損失関数は各ミニバッチ内で最適化される。直感的には、この損失関数は、 $p(\mathbf{z}|\mathbf{x})$  を最大化する代わりであると解釈できる。すべての可能な組み合わせの潜在分布に大きなオーバーラップがある場合、潜在表現  $\mathbf{z}$  は、対応する  $\mathbf{x}$  に対して大きな尤度  $p(\mathbf{z}|\mathbf{x})$  を示す必要がある。

### 3. 評価実験

#### 3.1 実験設定

提案手法を評価するために、評価用データセットとして株式会社 Gunosy から提供された 2019 年から 2021 年の実世界における広告クリエイティブデータセット約 200,000 件を使用した。このデータセットをそれぞれ 2019 年から 2020 年、2020 年から 2021 年の 1 年ずつのデータセットに分割した。その後、それぞれのデータセットに対して広告キャンペーンに基づいた 5-fold cross validation を実施し、予測性能を比較した。

評価指標として、一般的な logloss の他、本研究では予測値の順位を評価する normalized discounted cumulative gain (NDCG) を使用した。著者らは順位を評価可能な NDCG が配信効果が高い広告 (本研究における CVR の高さ) を正しく高いと予測できるかを評価できると主張しており [Kitada 19]、本研究でも同様の動機で利用した。

提案手法のネットワークでは事前学習モデルを 2 つ使用した。1 つ目は BERT encoder における東北大乾研究室によって

表 2: 最先端手法と提案手法の比較

Model	NDCG
DeepFM [Guo 17]	67.52±2.82
Wide & Deep [Cheng 16]	67.44±2.66
<b>PACE (Proposed)</b>	<b>83.04±1.35</b>

公開されている日本語事前学習済み BERT\*<sup>2</sup> である。2 つ目は Image encoder における ImageNet 事前学習済み ResNet34 である。それぞれの encoder は訓練時には全て重みを固定した。

提案する確率的埋め込みを学習する PACE の有効性を確認するため、ネットワーク構造が同じで決定論的埋め込みを学習するネットワークを用いた。また、ネットワーク中の text encoder を word embedding + long short-term memory (LSTM) [Hochreiter 97] としたものを比較した。

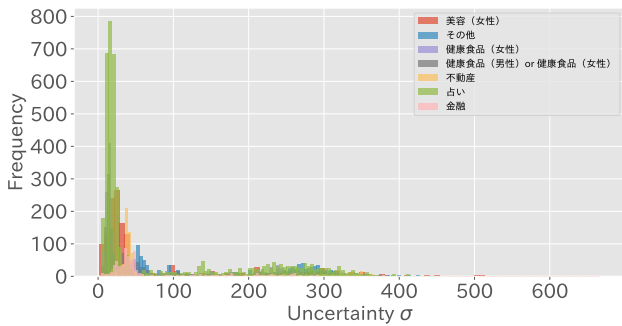
#### 3.2 実験結果

表 1 に実験結果を示す。決定論的な埋め込みに比べて、提案手法である確率的な埋め込みを利用することで text encoder の種類に限らず NDCG が良い結果になった。logloss の観点においても提案手法の確率的埋め込みは決定論的な埋め込みと同程度の値になっていることが確認できた。異なる期間の評価用データセットを使用したときにおいても、大きな予測性能の劣化は見られなかった。

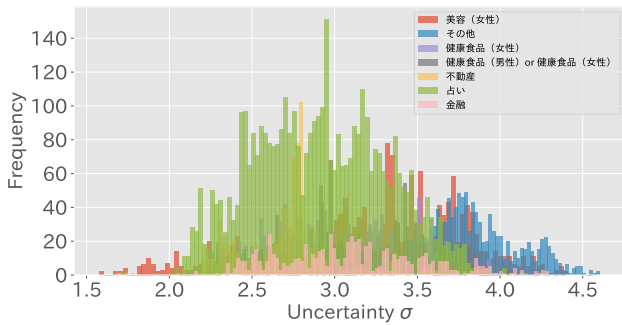
図 3 にクリエイティブ数上位 7 ジャンルにおける、学習済み PACE によって推定された不確実性の分布を示す。このときの不確実性  $\sigma$  は特徴量のすべての次元に渡る調和平均である。それぞれの図の横軸に着目すると、text encoder に LSTM を用いたときに比べて BERT を用いると小さい不確実性を示した。図 3a では、大部分が 100 を下回る不確実性を示した一方で、300 程度の不確実性を示す広告クリエイティブが存在した。図 3b では、美容やその他のジャンルにおける広告クリエイティブが大きい不確実性を示した。不動産や占いといったジャンルの広告クリエイティブは幅広い不確実性を示した。

図 4 に不確実性  $\sigma$  と実際の CVR の関係を示す。text encoder において BERT は LSTM よりも小さい不確実性を示した。また BERT を text encoder として用いた場合、実際に高い CVR を記録している広告クリエイティブは不確実性の高いものから生まれている可能性を示唆している。

\*2 <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>. 2021/02/26



(a) LSTM



(b) BERT

図 3: 推定された不確かさの分布。ここで、 $\sigma$  は特徴量のすべての次元にわたる調和平均を示す。横軸に着目すると、text encoder に LSTM を用いたときと比べて BERT を用いると小さい不確かさを示している。

## 4. 議論

### 4.1 決定論的埋め込み vs. 確率的埋め込み

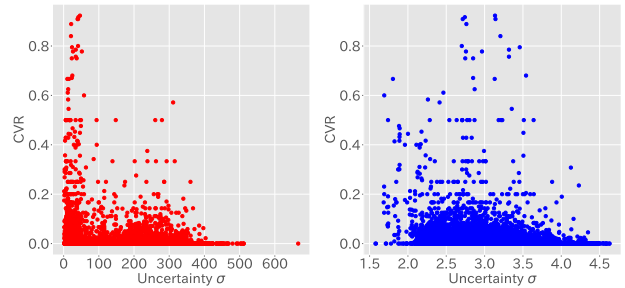
広告クリエイティブに対する click や conversion といった行動は不確か性が高く、また広告クリエイティブ自身も不確か性が高い（広告対象の製品に対して同様の傾向を持つクリエイティブが多数作られる）ことが知られている。提案手法である確率的な埋め込みは決定論的な埋め込みと比較して高い予測性能を示したが、これは広告クリエイティブの有する不確か性を適切に捉えていることを示唆している。

### 4.2 LSTM vs. BERT

我々が使用したネットワークはテキストを特徴量へと変換する text encoder を備えている。本研究ではこの text encoder として LSTM と BERT を考慮した。評価実験では、特に配信効果が高い広告を効果が高いと予測できたかを測る NDCG において LSTM よりも BERT を利用したほうが高かった。これは BERT が CVR 予測の文脈で効果の高いクリエイティブを適切に評価できていることを示唆している。提案手法である PACE は埋め込み  $\mu$  と同時に不確か性  $\sigma$  を出力する。このとき text encoder として LSTM と BERT では、特に BERT において小さい不確か性を示すことが確認した。

## 5. 結論と今後の展望

本研究では確率的な広告クリエイティブ埋め込み表現 (PACE) を提案した。PACE はクリエイティブが有する不確か性を適切に捉え、配信効果が高いクリエイティブを配信履歴情報なしに高い精度で予測することを可能とした。また PACE を学習



(a) LSTM

(b) BERT

図 4: 不確か性  $\sigma$  と実際の CVR の関係

する際に予測性能の高い BERT を用いることで、不確か性の小さい予測を得ること、すなわち信頼度の高い予測を得ることが可能となる。こうした信頼度の高い予測の提示はより精度の高いクリエイティブの評価につながると考えられる。

PACE が示す不確か性はクリエイティブの作成支援に繋がると考えている。クリエイティブは広告キャンペーンに対して複数種類のクリエイティブを作成するが、これらは文字を入れ替えたりといった小さい変更に留まることが多い。広告キャンペーンに対するクリエイティブの多様性を確保するために、PACE の不確か性が大きくなるようなクリエイティブを作成することでより高い配信効果を期待することができる。不確か性が大きいことは必ずしも良い結果を期待することができないため、今後は PACE による不確か性の提示と配信効果の高いクリエイティブの作成支援の文脈で検討を続けていく。

## 参考文献

- [Chen 16] Chen, J., Sun, B., Li, H., Lu, H., and Hua, X.-S.: Deep ctr prediction in display advertising, in *Proc. of ACM MM*, pp. 811–820 (2016)
- [Cheng 16] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., et al.: Wide & deep learning for recommender systems, in *Proc. of DLRSW*, pp. 7–10 (2016)
- [Devlin 19] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in *Proc. of NAACL*, pp. 4171–4186 (2019)
- [Guo 17] Guo, H., Tang, R., Ye, Y., Li, Z., and He, X.: DeepFM: a factorization-machine based neural network for CTR prediction, in *Proc. of IJCAI*, pp. 1725–1731 (2017)
- [He 16] He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in *Proc. of CVPR*, pp. 770–778 (2016)
- [Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997)
- [Jiang 19] Jiang, J., Yang, D., Xiao, Y., and Shen, C.: Convolutional Gaussian Embeddings for Personalized Recommendation with Uncertainty, in *Proc. of IJCAI*, pp. 2642–2648 (2019)

- 
- [Kitada 19] Kitada, S., Iyatomi, H., and Seki, Y.: Conversion Prediction Using Multi-task Conditional Attention Networks to Support the Creation of Effective Ad Creatives, in *Proc. of KDD*, pp. 2069–2077 (2019)
- [Liu 18] Liu, W., Tang, R., Li, J., Yu, J., Guo, H., He, X., and Zhang, S.: Field-aware probabilistic embedding neural network for ctr prediction, in *Proc. of RecSys*, pp. 412–416 (2018)
- [Lu 17] Lu, Q., Pan, S., Wang, L., Pan, J., Wan, F., and Yang, H.: A practical framework of conversion rate prediction for online display advertising, in *Proc. of AD-KDD*, pp. 1–9 (2017)
- [Shi 19] Shi, Y. and Jain, A. K.: Probabilistic face embeddings, in *Proc. of ICCV*, pp. 6902–6911 (2019)