

デジタル広告におけるクリエイティブの自動生成手法の検討

兵頭 沖 栗飯原 俊介
株式会社オプト

{f.hyodo, partner_s.aihara}@opt.ne.jp

1 はじめに

デジタル広告においてユーザーに提示される文や画像などの広告クリエイティブはユーザーと商品の接触地点であり、商品情報をユーザーに届ける役割を果たす [1]。広告配信における広告効果の説明変数はクリエイティブ、ユーザー、配信日時、配信量などがあるが、他が同条件でもクリエイティブの違いにより広告効果が大きく異なることは広告代理業を通して経験的に知られている。

一方で、クリエイティブによる広告効果はユーザーの性質などの測定が困難な変数の影響が無視できず、その推定は難しい研究課題である。また、同一のクリエイティブを配信し続けると広告効果が減少していくことが指摘されている [2]。

この背景を踏まえた広告効果を高めるアプローチとして、配信先の媒体で適用されるバンディットアルゴリズムによりクリエイティブの効果を検証し、効果が低いクリエイティブを差し替えることで広告効果を維持する方法が考えられる。このアプローチを有効にするには、クリエイティブを作り続けることが必須の条件となり、自動でクリエイティブを生成する機構が望まれる。

このような背景から、本研究では広告クリエイティブのうち、商品訴求を行う文章である広告文の自動生成タスクに焦点を当て、課題の整理と関連する要素技術の比較検討を行った。

2 デジタル広告におけるクリエイティブ作成の課題

2.1 広告代理店における広告運用の課題

広告代理店は広告主が広告を出すプロセスのサポートを行う事業者である。その収益を確保する重要な要因にユーザーに表示される広告クリエイティブがある。

広告の目的に沿った成果を効率的に出せるクリエイティブの探索を行うため、入稿クリエイティブの監視

を行い、効果の低いクリエイティブの停止・差し替えを行うことが広告代理店の業務の一つである。

自動生成によりクリエイティブを継続的に作ることができれば差し替えにより広告効果の高いクリエイティブを効率的に探索できるため、同一商品に対して新規のクリエイティブを自動生成するシステムの開発は広告代理店にとって重要な課題である。

2.2 課題解決のアプローチ

2.1 節で述べた課題を解決するための、広告文に焦点を当てたクリエイティブの自動生成手法の条件を以下に列挙する。また、各条件を満たすためのアプローチを述べる。

1. 破綻していない文章を出力できる
2. 広告対象の商材情報を反映した文章を出力できる
3. 広告効果が高い文章を出力できる
4. 多様な文章を出力できる

条件 1, 2 については、デジタル広告の遷移先ページなどの商材情報が含まれる情報源から抽出した文章集合を入力として、文章要約を行うことで広告文を作成するアプローチが考えられる。

条件 3 は、CTR などの広告効果を目的変数として、強化学習のアプローチを用いて言語モデルとしての損失の最小化と広告効果の最大化を行うマルチタスクラーニングの手法が考えられる。

条件 4 を満たすためには、単一の入力から複数の異なる文を出力する必要がある。確率的な文生成モデルを学習し、分布からサンプリングすることで複数の文を生成するアプローチや、beam-search による探索の際に N-best 全てを出力するものが考えられる。

本論文ではこれらの条件のうち 1, 2, 4 に焦点を当てる。広告文の生成モデルとして文章要約モデルを用い、beam-search による N-best をモデルの出力とする方法を取る。

3 検証する文生成手法

要約タスクに標準的に使われるモデルとして Seq2Seq[3] に要約文と入力文の単語間関係の強弱を表す機構である attention を付与した Seq2Seq with attention がある。また、その発展形としてさまざまな要素技術が提案されている。これらの要素技術のうち、2.2 節で述べた条件の一部を満たす手法について以下で述べる。

条件 1 を満たすため、文の長期的な依存関係を捉えられるモデルである Transformer[4] を活用する。また Transformer のモデルをベースにした事前学習済みの言語モデルである BERT[5] を活用することで自然な文章を出力する効果を狙う。

条件 2 を満たすために copying mechanism[6] を活用する。copying mechanism は Out of Vocabulary (OOV) の問題を解決するために考案されたものだが、入力から必要な情報を抜き出す機構として考えると商品情報を生成文に反映させる効果が期待できる。

以下で各文生成手法について述べる。

3.1 Seq2Seq with attention

エンコーダー・デコーダーそれぞれにリカレントニューラルネットワークを用いたモデルである。エンコーダーにより入力単語列のベクトル表現 $\{h_s\}$ を計算し、デコーダーの計算をエンコーダーの出力 $\{h_s\}$ で条件付けて行う。またデコーダーにより時刻 $t-1$ に生成した単語 y_{t-1} の表現ベクトル h_{t-1} から入力単語列への attention a_{t-1} を求める。出力単語 y_t の分布を a_{t-1} により $\{h_s\}$ の加重和を計算することで、1 単語ずつ出力単語列を得る。

3.2 Transformer

Transformer は入力単語列を並列に入力し、入力単語列のベクトル表現を attention によりエンコード、デコードするモデルである。エンコーダーとデコーダーはそれぞれ、self-attention を含むエンコーダーレイヤと、self-attention と source-target attention を含むデコーダーレイヤを N 層スタックすることで構成される。

self-attention と source-target attention は複数ヘッドによる attention により実現され、入力を query, key, value の 3 種類のベクトルに変換し、query と key の類似度を元に決定した重みを用いて value の加重和を計算する。query を Q , key を K , value を

V , key の次元を d_k とすると、attention は

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

と表せる。ただし、query と key の類似度計算には $\sqrt{d_k}$ で内積の結果をスケールリングする scaled-dot product を用いている。

self-attention では自身の出力から query, key, value を計算することで同一文内の単語間の関係の強弱を計算する。また source-target attention はデコーダーの隠れ層の出力から query を、エンコーダーの出力から key と value を計算することで入力文と出力文の単語間の関係の強弱を計算する。

3.3 BERT

Transformer のモデルをベースにした汎用なタスクに使えるモデルとして BERT がある。事前学習の手法やモデルの詳細についてはここでは割愛するが、今回は単純にエンコーダー部分に BERT を用いたモデルを比較に用いる。

3.4 copying mechanism

copying mechanism は入力文の部分系列を陽にコピーする機構であり、Seq2Seq のモデルに付加することで OOV の問題に対処する。

学習データに存在する語彙 \mathcal{V} に UNK という OOV を表す特殊な単語を追加した集合を考える。加えて、入力単語列 X 内のユニークな単語の集合 \mathcal{X} を考える。 X に対する語彙は $\mathcal{V} \cup \text{UNK} \cup \mathcal{X}$ となる。

モデルの語彙 \mathcal{V} から生成する generate mode (g) と、入力単語列からコピーする copy mode (c) それぞれについて、 y_t を生成する同時確率を計算したのちに合算する。

generate mode と copy mode それぞれにおいて、単語 y_t を生成する確率は以下ようになる。

$$p(y_t, g|\cdot) = \begin{cases} \frac{1}{Z} e^{\psi_g(y_t)}, & y_t \in \mathcal{V} \\ 0, & y_t \in \mathcal{X} \cap \bar{\mathcal{V}} \\ \frac{1}{Z} e^{\psi_g(\text{UNK})}, & y_t \notin \mathcal{V} \cup \mathcal{X} \end{cases} \quad (2)$$
$$p(y_t, c|\cdot) = \begin{cases} \frac{1}{Z} \sum_{j: x_j = y_t} e^{\psi_c(x_j)}, & y_t \in \mathcal{X} \\ 0, & \text{otherwise} \end{cases}$$

ここで、 $\psi_g(\cdot)$ と $\psi_c(\cdot)$ はそれぞれ generate mode と copy mode のスコア関数、 Z は正規化項で $Z = \sum_{v \in \mathcal{V} \cup \{\text{UNK}\}} e^{\psi_g(v)} + \sum_{x \in \mathcal{X}} e^{\psi_c(x)}$ である。

モデルの語彙に存在しない単語 $w \in \bar{\mathcal{V}}$ の出力確率は generate mode のみの場合 0 だが、copying mechanism

により $w \in \mathcal{X}$ であれば入力単語列からコピーすることで生成が行えるため、OOV に対処ができる。

4 実験

3章で述べた要素技術を組み合わせて比較を行う。入稿された広告文と遷移先 URL の組を学習データの元にする。遷移先 URL が示すページであるランディングページから商品情報に関する文章を抽出して入力文を作成し、入力文と広告文の組を学習データとして用いる。広告文生成モデルとして文章要約モデルを用い、学習時は広告文と MLE による予測値から損失を計算し、推論時は1つの入力文から beam-search により計算された尤度が高い文字列を上から4件出力する。

ベースラインの Seq2Seq モデルではエンコーダーに BiLSTM、デコーダーに LSTM を用いた。また Transformer と、それぞれのエンコーダー部分を BERT に変更した以下のモデルを用意した。

- BiLSTM&LSTM
- Transformer
- BERT&LSTM
- BERT&Trans-dec

さらに、各モデルに copying mechanism を付加したモデルを用意して実験を行った。ただし、Trans-dec は [4] のアーキテクチャのデコーダー部分を表し、”&”で区切られた左側はエンコーダー、右側はデコーダーを表す。また広告文生成モデル X に copying mechanism を付加した機構を X+COPY と表す。

実験で使用したすべてのモデルにはデコーダーからエンコーダーへの attention を組み込んだ。Transformer のエンコーダー・デコーダーと Trans-dec のレイヤはそれぞれ3層にし、ヘッドの数は4とした。分かち書きには sentencepiece[7] を用い、sentencepiece と BERT の学習済みモデルは [8] のモデルを使用した。

4.1 データ

データは株式会社オプト内にある化粧品に関する広告配信実績を使用した。クリエイティブの種類を問わずデータを収集し、遷移先ページから本文抽出した入力文と広告文のペアとしての重複を除去することで3860件のデータを獲得し、学習用・検証用・評価用に8:1:1の割合で分割した。入力文・広告文のユニークな件数はそれぞれ3248件、493件となった。

4.2 自動評価

配信済み広告文との類似性を見るために BLEU を、出力の多様性を見るために DIST-n[9] を計算した。

評価結果を表1の BLEU, DIST-1, DIST-2 に示す。copying mechanism を付加したモデルは付加していないモデルと比較して、Transformer を除き BLEU, DIST-1, DIST-2 のスコアが向上した。また、BERT&Trans-dec+COPY は BLEU において Transformer のスコアを上回る結果となった。

4.3 人手評価

株式会社オプトに所属する広告事業に関わる社員27名に対してアンケートを実施した。

- 文として不自然でない
- 広告の対象商品に関する内容である

の2点を満たす広告文を多く出力していると回答者が感じる順番に、同数の場合は出力の質の高さでモデルを順位付けた。評価用データからランダムにサンプリングした20件をアンケート用のデータとして用い、各モデルの出力4件からランダムに2件ずつ抽出した。また実験対象のモデルを

- BiLSTM&LSTM+COPY
- Transformer
- BERT&LSTM+COPY
- BERT&Trans-dec+COPY

の4つに限定した。限定した理由は BiLSTM&LSTM と BERT&LSTM の出力が意味のとれるものが少なく、Transformer+COPY と BERT&Trans-dec の出力が入力文を考慮したものが少なかったためである。

アンケートの評価結果を表1の最右列に示す。BERT&Trans-dec+COPY の平均順位が最も高く、次点の Transformer と比較して0.18高かった。

4.4 考察

入力文と各モデルの出力例を表2に示す。Transformer の出力は自然で多様だが、商品に関連が薄い文も多かった。copying mechanism を付与したモデルの出力は多様性は Transformer に劣るが商品との関連が強いものが多かった。これは Transformer のスコアが DIST-1 では1位だが人手評価では2位だったことに関係する可能性がある。人手評価ではアンケートの対象者が商品と関係が薄い出力に対して低い順位を付けた結果、Transformer の順位が下がったと考えられる。

文生成モデル	BLEU	DIST-1	DIST-2	人手評価順位の平均
BiLSTM&LSTM	0.329	0.009	0.018	-
BiLSTM&LSTM+COPY	0.595	0.025	0.071	2.74
Transformer	0.653	0.037	0.088	2.46
Transformer+COPY	0.284	0.013	0.021	-
BERT&LSTM	0.111	0.002	0.005	-
BERT&LSTM+COPY	0.653	0.033	0.092	2.52
BERT&Trans-dec	0.204	0.009	0.023	-
BERT&Trans-dec+COPY	0.656	0.031	0.079	2.28

表 1: エンコーダー, デコーダー, copying mechanism を組み合わせたモデルの自動および人手の評価結果.

入力文	出力文
BiLSTM&LSTM+COPY	... ヘアカラー・化粧品・コスメ シャンプーで落ちるマスカラタイプ... 髪によってさまざまなセクシーさりヘアカラーブランド.
Transformer	髪を美しく保湿までスキンケアお試しセットが〇〇限定〇〇円
BERT&LSTM+COPY	髪をもっと, マイペースに.
BERT&Trans-dec+COPY	髪を美しく染めるヘアカラーブランド.

表 2: 入力文と広告文生成モデルの出力例. 入力文は一部のみ, モデルは人手評価の実験対象のみで各モデルの出力は1つのみ表示している.「〇〇」は商品情報のためマスクをかけた部分である.

自動評価の実験において Transformer+COPY, BERT&LSTM, BERT&Trans-dec のスコアが低かったのは実装上の問題である可能性があるため, 今後調査を行う必要がある.

5 おわりに

本研究では広告クリエイティブの自動生成システムに求められる条件を考察し, 条件の一部を満たす広告文生成モデルを作成し, 比較検討を行った. 実験を通して, Transformer と copying mechanism が広告文生成タスクにおいて有用であることを確認した.

実験では事前学習済みモデルを組み込んでも性能は大幅には向上しなかった. このことから, 言語モデルだけでなく文生成を直接事前学習の際に行う学習済みモデルが必要だと思われる.

本研究では多様性を考慮した問題設定にしつつも, 陽に多様性を扱うモデルを取り扱っていない. 文生成の多様性に焦点を当てた研究は [10] や [11] などがある. 今後はこれらの機構を広告文生成のタスクに合わせて改良したモデルを検討していく必要がある.

参考文献

- [1] Olivier Chapelle, Eren Manavoglu, and Rómer Rosales. Simple and scalable response prediction for display advertising. *ACM TIST 2014*, Vol. 5, No. 4, pp. 61:1–61:34, 2014.
- [2] マーケティング・サイエンス入門: 市場対応の科学的マネジメント, pp. 174:16–175:22. 有斐閣, 2011.
- [3] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *NIPS 2014*, pp. 3104–3112, 2014.
- [4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS 2017*, pp. 5998–6008, 2017.
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT 2019*, pp. 4171–4186, 2019.
- [6] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O. K. Li. Incorporating copying mechanism in sequence-to-sequence learning. In *ACL 2016*, 2016.
- [7] Taku Kudo and John Richardson. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *EMNLP 2018*, pp. 66–71, 2018.
- [8] Yohei Kikuta. Bert pretrained model trained on japanese wikipedia articles. <https://github.com/yoheikikuta/bert-japanese>, 2019.
- [9] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In *NAACL-HLT 2016*, pp. 110–119, 2016.
- [10] Jiachen Du, Wenjie Li, Yulan He, Ruifeng Xu, Lidong Bing, and Xuan Wang. Variational autoregressive decoder for neural response generation. In *EMNLP 2018*, pp. 3154–3163, 2018.
- [11] Ashwin K. Vijayakumar, Michael Cogswell, Ramprasaath R. Selvaraju, Qing Sun, Stefan Lee, David J. Crandall, and Dhruv Batra. Diverse beam search for improved description of complex scenes. In *AAAI 2018*, pp. 7371–7379, 2018.