

広告効果を報酬とした強化学習に基づく広告文の自動生成

上垣外 英剛*^{1,a)} 張 培楠*^{2,b)} 高村 大也^{1,3,c)} 奥村 学^{1,d)}

概要: デジタル広告の増加とターゲティング精度の向上に伴い、広告制作物（クリエイティブ）の需要はますます高まってきている。しかしクリエイティブが人手によって制作されることが主である現状では、この需要を十分に満たすことができていない。また、クリエイティブを自動的に生成する場合においても、既存の自然言語生成手法では、テンプレートに単語や句を埋め込む手法や、最尤推定法に基づいて系列変換モデル (Seq2Seq) を学習するものがほとんどであるが、これらの手法では、広告の効果を直接考慮した上でクリエイティブを生成することが困難である。本論文ではこの問題を解決するために、事前に推定された広告効果を強化学習の報酬として用いることにより、実際の広告配信での効果の向上を直接の目的とした広告文の生成を学習可能な手法を提案する。さらに、従来手法と提案手法を自動評価および人手評価により比較することで、各指標においての有効性を検証した。

キーワード: 自然言語処理, テキスト生成, Seq2Seq, 強化学習

Effect-Rewarded Reinforcement Learning for Ad Text Generation

HIDETAKA KAMIGAITO*^{1,a)} PEINAN ZHANG*^{2,b)} HIROYA TAKAMURA^{1,3,c)} MANABU OKUMURA^{1,d)}

Abstract: With the increase of digital advertising and the improvement of targeting accuracy, the demand for advertisements (hereinafter referred to as creatives) is increasing. However, because of its large quantity, human hands are not sufficient to satisfy this demand. Also, even if generating creatives automatically, commonly used natural language generation approaches such as template-based and sequence-to-sequence (Seq2Seq)-based approaches do not explicitly consider the effects of generated creatives. To solve this problem, we propose Seq2Seq-based methods that can take into account the effects of creatives as a reward of the reinforcement learning approach. In this paper, we investigated the effectiveness of automatically generated creatives by comparing the conventional and our proposed methods on various domains of advertisements.

Keywords: Natural Language Processing, Text Generation, Seq2Seq, Reinforcement Learning

1. はじめに

Webの普及に伴いデジタル広告の数は増加し、コンバージョン*¹をより増やすための広告のパーソナライズ、つま

りターゲティングも自動化され、デジタル広告は高速化・高精度化されてきている。そのような状況において、クリエイティブは未だに人手で、過去の経験や知識を活かしながら制作されているため、その需要の増加速度に追いつけないケースも存在する。このような課題から、クリエイティブを製作者の考え方のパターンや、あるいは過去に作ったものから学習することで自動的に生成する試みは少なくない [1], [2]。しかし、これらの手法はテンプレートに対する slot filling に基づく手法や、最尤推定により学習された Seq2Seq を用いる手法であり、多様性を増すのが困難な上に広告表現の質や配信による効果を考慮していない。

*Both authors are equally contributed.

¹ 東京工業大学 科学技術創成研究院 未来産業技術研究所

² 株式会社サイバーエージェント アドテック本部 AI Lab

³ 産業技術総合研究所

a) kamigaito@lr.pi.titech.ac.jp

b) zhang_peinan@cyberagent.co.jp

c) takamura@pi.titech.ac.jp

d) oku@pi.titech.ac.jp

*1 商品の購入や申込といったサービスが目標としているアクションを実行すること

入力	配信ページ本文: 人気の保険をランキング形式でご紹介! ■■■が運営する保険の国内最大級の比較サイト「■■■ ■」で人気のあった保険を、保険種類ごとに資料請求順・ 申込順など様々な形式で比較できます。 配信キーワード: [保険, 比較] ユーザの検索クエリ: [保険, の, 比較]
広告	タイトル: 保険はどれが一番お得なの? 最新ランキング 内容: 国内最大級の保険比較サイト<■■■> ランキン グ上位からオススメします。

表 1 今回生成対象とする広告文と入力対の例

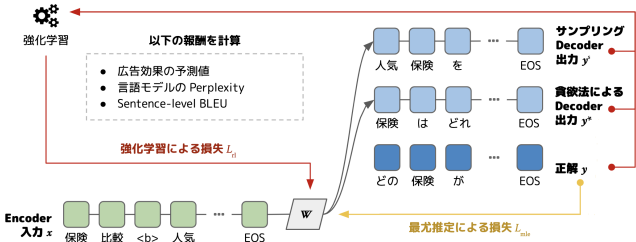


図 1 提案手法の概要

我々はこの問題を解決するために、強化学習手法を取り入れることで多様性に富んだ、報酬として広告効果を学習時に考慮可能な広告文の生成手法を提案する。

2. 生成対象とする広告文

広告文は、配信ページ、配信キーワード、ユーザの検索クエリに応じて提示される。実際の入力と広告文のペアを表 1 に示す。入力の中で、ユーザの検索クエリは事前に配信者が決定できないため、機械学習手法に基づくシステムの入力としては利用できる場面が限られる。この前提に基づき、入力系列として下記の二種類の形式を使用した。

- 配信ページ本文 配信キーワード 検索キーワード
- 配信ページ本文 配信キーワード

なお、は区切りを表す記号である。また、Javascript 等の影響により、配信ページが直接クロールできない場合には、URL を単一のトークンとして代わりに使用した。そして、出力系列として以下の形式を使用した。

- 広告タイトル 広告内容

次節では、これらの形式に基づく広告文の生成方法についての説明を行う。

3. 提案手法

提案手法は図 1 に示すように、Seq2Seq の重みパラメータを self-critical policy gradient (SCST)[3] に従い更新する [4]。これにより、最尤推定では考慮することが難しい広告効果を、あらかじめ学習された予測器を通じ、報酬として Seq2Seq の学習に反映することができる。次節以降では各機構の詳細についての説明を行う。

3.1 広告文生成に用いる Seq2Seq

Seq2Seq では、入力系列 $\mathbf{x} = x_1 \cdots x_n$ に対し、出力確率が最大となる広告文 $\mathbf{y} = y_1 \cdots y_m$ を次式に従い出力する：

$$\mathbf{y} = \arg \max_{\mathbf{y}} \prod_{t=1}^m P(y_t | \mathbf{x}, y_{t-1} \cdots y_1). \quad (1)$$

広告文中の単語 y_t を出力する確率は、単語埋め込み $Emb(\cdot)$ 、両方向 LSTM から成るエンコーダ $Enc(\cdot)$ 、右方向 LSTM から成るデコーダ $Dec(\cdot)$ 、デコーダの各時刻での隠れ状態 s_t 、エンコーダへの注意 $AttEnc(\cdot)$ の出力結果 c_t^{enc} 、デコーダへの注意 $AttDec(\cdot)$ の出力結果 c_t^{dec} を用いて、

$$\begin{aligned} P(y_t | y_{t-1} \cdots y_1, \mathbf{x}) &= g(s_t, c_t^{enc}, c_t^{dec}), \\ c_t^{dec} &= AttDec(s_1 \cdots s_t), \\ c_t^{enc} &= AttEnc(h_1 \cdots h_n, s_t), \\ s_t &= Dec(s_{t-1}, Emb(y_{t-1})), \\ h_i &= Enc(Emb(x_i), h_{i-1}, h_{i+1}), \end{aligned} \quad (2)$$

のように計算される。エンコーダへの注意は過去の注意を考慮した上で次式により計算される：

$$\begin{aligned} c_t^{enc} &= \frac{\exp(e_t)}{\sum_{k=1}^t \exp(e_k)}, \\ e_t &= \sum_{j=1}^{T_x} \alpha_j^t h_j, \\ \alpha_j^t &= \frac{\exp(s_t W_{att}^{dec} h_j)}{\sum_{k=1}^n \exp(s_t W_{att}^{dec} h_k)}. \end{aligned} \quad (3)$$

デコーダへの注意についても同様に、過去の注意を考慮した上で次式により計算される：

$$\begin{aligned} c_t^{dec} &= \frac{\exp(d_t)}{\sum_{k=1}^t \exp(d_k)}, \\ d_t &= \sum_{j=1}^t \beta_j^t s_j, \\ \beta_j^t &= \frac{\exp(s_t W_{att}^{enc} s_j)}{\sum_{k=1}^n \exp(s_t W_{att}^{enc} s_k)}. \end{aligned} \quad (4)$$

出力層ではコピー機構に基づいて入力中の単語を直接出力するために、コピーする確率 $P(u_t = 1)$ を用いて、 $P(y_t | u_t = 1)$ と $P(y_t | u_t = 0)$ の重み付き和

$$\begin{aligned} &g(s_t, c_t^{enc}, c_t^{dec}) \\ &= P(u_t = 1)P(y_t | u_t = 1) + P(u_t = 0)P(y_t | u_t = 0), \end{aligned} \quad (5)$$

を最終的な単語の出力確率として使用する。それぞれの確率は次式によって計算される：

$$\begin{aligned} P(u_t = 1) &= \sigma(W_u[s_t, c_t^{enc}, c_t^{dec}] + b_u), \\ P(u_t = 0) &= 1 - P(u_t = 1), \\ P(y_t | u_t = 1) &= \text{softmax}(W_{out}[s_t, c_t^{enc}, c_t^{dec}] + b_{out}), \\ P(y_t = x_j | u_t = 0) &= \alpha_j^t. \end{aligned} \quad (6)$$

出力層と入力層で同じ単語ベクトルを使用するために、出力層の単語の重み行列は入力側の単語埋め込み W_{emb} と変換用の重み行列 W_{proj} を用いて、 $W_{out} = \tanh(W_{emb}W_{proj})$ のように計算される。

3.2 目的関数

提案手法の目的関数は、最尤推定に基づく損失 L_{mle} と、強化学習に基づく損失 L_{rl} のハイパーパラメータ γ による重み付き和 $L_{mixed} = \gamma L_{rl} + (1 - \gamma)L_{mle}$ で表される。 L_{mle} は訓練データ中に存在する出力広告文の系列 \mathbf{y}^* を用いて次のように計算される：

$$L_{mle} = - \sum_{t=1}^m \log P(y_t^* | y_{t-1}^* \cdots y_1^*, \mathbf{x}). \quad (7)$$

L_{rl} は、訓練時の出力確率に基づいて $P(y_t^s | y_{t-1}^s \cdots y_1^s, \mathbf{x})$ のようにサンプリングされた系列 \mathbf{y}^s と貪欲法により得られた系列 $\hat{\mathbf{y}}$ を用いて次のように計算される：

$$L_{rl} = (r(\hat{\mathbf{y}}) - r(\mathbf{y}^s)) \sum_{t=1}^m \log P(y_t^s | y_{t-1}^s \cdots y_1^s, \mathbf{x}). \quad (8)$$

$r(\cdot)$ は与えられた系列に対する報酬を返す関数であり、次の3.3節でより詳細に述べる。

3.3 報酬

提案手法の報酬 $r(\cdot)$ は以下のものを組み合わせて用いた：

- 広告効果 生成する広告文の広告効果を向上するためのスコア。広告配信時に配信元から与えられる広告効果を回帰モデルにより予測して使用している。広告効果の予測についての詳細は次節にて解説する。
- 言語モデル perplexity 広告文の文法性を担保するための報酬。生成された広告文に対して文字を入出力とする LSTM 言語モデル*2によって perplexity を計算し、負の値としたものを報酬として用いている。
- Sentence-BLEU 生成する広告文が訓練データと類似するようにするための報酬。Sentence-BLEU の結果を100倍して使用している。

3.3.1 広告効果の予測

過去に配信した広告文と、その際に得られた配信効果を示すラベルからなるデータセットを用意した。このデータセットを用いて、広告文から配信効果を予測するモデルを構築した。配信効果ラベルは最低を示す1から最高を示す10までの10段階からなっている。効果予測では、入力である広告文を FastText[5] により文の分散表現にエンコードしたのち、Gradient Boosting Regression Tree (GBRT) で回帰モデルを構築して配信効果ラベルを予測した。配信効果の予測値は強化学習の報酬として用いられるため、予

*2 2層、100次元の隠れ層に対し、学習率20のSGDで日本語Wikipediaを事前学習した後に、Seq2Seqの学習データに含まれる広告文により fine-tuning されたものを使用した。

	保険 1	保険 2	金融
訓練データ	33,633	13,766	94,475
開発データ	1,000	1,000	1,000
テストデータ	1,000 (30)	1,000 (30)	1,000 (30)

表 2 各データセットに含まれる文対の数。() 内の数字は人手評価に使用した対の数を表す。

測値を推定する速度が重要である。RNN などのより複雑なモデルを採用することも可能だが、精度はわずかに上昇する代わりに推定が遅くなり強化学習の学習速度の低下に直結する。よって、精度と速度のバランスが最適である本手法により広告効果を予測し、強化学習の報酬とした。

4. 実験と考察

4.1 実験設定

表 2 に示すように実験には保険・金融ドメインのクライアント 3 社の配信データを用いた（以降、保険 1、保険 2、金融と記載）。入出力が一致する対が二回以上存在した場合には、データセット中で単一の対となるように、二回目以降に出現した同一の対をデータセットから除外した。単語分割には MeCab を使用した。学習は文献 [4] と同じパラメータの元で 20 回試行し、最も出力単語の予測正解率が高いモデルを採用した。勾配爆発により学習が停止する場合には γ を半減させることで対応し、広告文の生成には貪欲法を使用した。

4.2 評価手法

生成された広告文に対し、自動評価と人手評価を実施した。自動評価では BLEU [6] と 3.3.1 節で作成した広告効果の予測モデルで予測される値 (Ad Effect Score; AES) を用いるが、これらの指標では本研究の目標である、より広告効果があり、かつ多様性に富んだ広告文の評価には不十分である。そのため、あわせて人手評価も実施した。

人手評価では Pairwise Ranking Agreement による手法 [7], [8] を用いてモデルを評価した。具体的には、入力となる配信ページ本文・配信キーワード・検索キーワードを評価者に見せ、広告文候補を良い順に並べ替える。広告文候補はベースライン手法と二つの提案手法、そして正解の 4 つであり、これらをランダムな順序で表示して評価者に順序を入力させる。そして得られた順序結果からペアを生成する。例えば広告文候補 (A, B, C, D) の評価結果が $(4, 2, 1, 3)$ の場合、次のようにペアが生成される。(1) $A < B, A < C, A < D$, (2) $B < C, B > D$, (3) $C > D$ 。この生成されたすべてのペアに対して TrueSkill [9] を実行し、得られた μ 値をモデルの良さを示す指標とする。

また、評価者間の一致度をクリップンドルフの α 値 [10]、コーエンの κ 値 [11]、スピアマンの順位相関係数 r_s [12] そしてケンドールの順位相関係数 τ [13] で測定した。

データセット	自動評価 (BLEU)			自動評価 (AES)			人手評価			平均順位	
	保険 1	保険 2	金融	保険 1	保険 2	金融	全て	保険 1	保険 2	金融	全て
Gold	–	–	–	7.35	7.08	6.79	25.30	25.66	22.95	26.58	2.00
ベースライン手法	34.28	32.04	31.88	7.41	7.23	6.79	23.93	24.99	23.29	24.74	2.06
LM+QS	32.89	20.25	29.04	7.35	7.08	6.74	25.16	24.93	26.16	25.37	2.27
LM+QS+BLEU	33.67	23.91	28.63	7.40	7.21	6.77	25.13	25.00	24.43	26.47	2.17

表 3 各手法の自動評価および人手評価の結果. +LM は言語モデル, +QS は広告効果, +BLEU は Sentence-BLEU の報酬をそれぞれ学習に使用したことを示している.

α	κ	r_s	τ
0.208	0.101	0.233	0.202

表 4 評価者間の一致度

4.3 実験結果と考察

実験の結果を表 3 に示す. 自動評価では, BLEU と AES 両方において全てベースライン手法が他を上回った. 一方, 人手評価では 7 人の評価者に答えてもらったところ, 提案手法である LM+QS と LM+QS+BLEU がベースライン手法を上回った.

自動評価と人手評価に結果の乖離が生じた原因としては, 入力に対して広告が一意には定まらないためであると考えられる. BLEU, AES は共に訓練データに近い出力に高いスコアを与えるため, 今回の評価では最尤推定により訓練データに近い広告文を出力する ベースライン手法が自動評価において高い評価を得たと考えられる. 人手評価では報酬を用いた手法が高い評価を得ていることから, 広告効果の予測値を報酬とすることで, 出力される広告文の質を向上可能なことが分かる. しかし, 上記の結果からは Sentence-BLEU が広告文の生成において有用な報酬ではないことも分かる.

一方で, 人手評価でも平均順位においては提案手法が低くなる傾向がある. 目視の結果, 強化学習手法の出力は文法的な誤りを生じる回数がベースライン手法よりも多く, この影響により順位が下がったと考えられる. この問題を解決するためにはより効果的な言語モデルが求められる. また, 表 4 に示すように広告の評価は評価者に強く依存するため, 実運用上はどのようなユーザに対して生成された広告が有効であるかの調査が必要であると考えられる.

5. 結論

本論文では, 広告効果の予測値を報酬とする強化学習が可能な Seq2Seq による Web ページの広告文の自動生成手法を提案した. 提案手法は人手評価において, 報酬を用いない Seq2Seq よりも高い TrueSkill スコアを達成した. 今後は言語モデルの報酬を改良することで, 広告文の文法性を向上させることを課題としたい.

参考文献

- [1] 山根宏彰, 萩原将文: Web 情報を用いたキャッチコピー自動生成システム, 日本知能情報ファジィ学会 (2012).
- [2] 福田宏幸: キーワード条件つき変分 Autoencoder による広告文生成, 人工知能学会全国大会 (2019).
- [3] Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J. and Goel, V.: Self-critical sequence training for image captioning, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017).
- [4] Paulus, R., Xiong, C. and Socher, R.: A Deep Reinforced Model for Abstractive Summarization, *International Conference on Learning Representations* (2018).
- [5] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T.: Enriching Word Vectors with Subword Information, *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135–146 (2017).
- [6] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation, *Proceedings of the 40th annual meeting on association for computational linguistics* (2002).
- [7] Bamman, D. and Smith, N. A.: Open Extraction of Fine-Grained Political Statements, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (2015).
- [8] Bojar, O., Chatterjee, R., Federmann, C. et al.: Findings of the 2016 Conference on Machine Translation, *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers* (2016).
- [9] Herbrich, R., Minka, T. and Graepel, T.: TrueSkill (TM): A Bayesian Skill Rating System, *Advances in Neural Information Processing Systems 20* (2007).
- [10] Krippendorff, K.: Estimating the Reliability, Systematic Error and Random Error of Interval Data, *Educational and Psychological Measurement*, Vol. 30, No. 1, pp. 61–70 (1970).
- [11] Cohen, J.: A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46 (1960).
- [12] Spearman, C.: The proof and measurement of association between two things, *American journal of Psychology*, Vol. 15, No. 1, pp. 72–101 (1904).
- [13] Kendall, M. G.: A new measure of rank correlation, *Biometrika*, Vol. 30, No. 1/2, pp. 81–93 (1938).