

# スタイル制御を考慮した多様な広告文生成

川本 峻領<sup>†1†2</sup> 張 培楠<sup>†1</sup>

<sup>†1</sup>株式会社サイバーエージェント <sup>†2</sup>明治大学  
 {shunyo\_kawamoto, zhang\_peinan}@cyberagent.co.jp

## 1 はじめに

広告文は、ユーザの注意を惹きつけ、クリックやコンバージョンといった購買アクションに誘導する重要な役割を果たしている (図 1)。その一方で、日々増加する商品やサービスのすべてに対して、人手によって適切で魅力的な広告文を作成することは時間的・金銭的なコストが莫大なため現実的ではない。また、インターネット広告では、ユーザに何度も同じ内容の広告を提示するとユーザの反応が著しく低下することが知られている。これは「広告の疲弊」と呼ばれ、この問題を解決するためには同一の商品に対して複数の広告を製作する必要がある。そのため、多様な広告文を自動で大量に生成することは、広告制作者の負担を減らすことができる上、広告の疲弊を回避する手段にもつながる。

しかし自動生成タスクで頻繁に用いられる sequence-to-sequence (seq2seq) 手法は、単一入力に対応する最適な系列を出力するように学習するため、多様性と制御可能性を高めるのは簡単ではない。変分オートエンコーダ (Variational Auto-Encoder, VAE) [1] を用いることで、学習された潜在変数の分布から近い意味やスタイルの表現を多数サンプリングできるが、そのままでは生成する表現のスタイルを制御することができない。そこで我々は VAE をベースとした、品詞情報に着目した POS-VAE を提案する。また、既存の多様性の評価尺度では、各入力に対して多数の出力を行うモデルに対処できるものがなく、VAE のようなモデルを正しく評価できない。そのため、本研究では新たに多出力の場合を考慮した多様性の評価尺度を提案する。

## 2 関連研究

文生成における多様性の高低は、対話タスク [2, 3] に始まり、近年では広告文生成 [4, 5] でも注目されるようになってきている。

ニューラルネットワークで学習された対話モデルで



図 1: ショッピングサイトにおける広告文の例

は、応答文を生成する際に「そうですね」や「わかりません」といった当たり障りない発言を多用する解に陥ることがある。多くの場面で使用できるだけに多様性の低下に伴う人間味や自然さの喪失といった問題を引き起こす。この多様性を向上させるための一つのアプローチが変分オートエンコーダである。

**変分オートエンコーダ** 変分オートエンコーダは Kingma ら [1] によって提案された生成モデルの一種である。従来のオートエンコーダの中間表現に潜在変数  $z$  の正規分布を仮定し、そのパラメータである平均  $\mu$  と分散  $\sigma$  を推定する。その学習された潜在変数からサンプリングすることで、入力に近い表現を生成できる。また、変分オートエンコーダの入力時に制約となる要素を加えることで、出力をコントロールする CVAE [6, 7] も後続研究として考案されている。

**広告文生成** 広告文へのテキスト生成技術の適応は近年盛んに行われてきている。福田ら [4] は CVAE を用いて、キーワードを入力とした手法を提案した。出力文の制御という意味では本研究と同じだが、入力される単一の単語に依存するため、文構造単位でのスタイルの制御ができないという点で本研究と異なる。

Zhang ら [5] も広告文生成タスクにおいて CVAE を拡張したニューラルネットワークを使用している。「赤色」「メンズ」などといった商品特徴を持つ複数の単語を条件として入力することで、これらの条件に沿った広告文を生成する。しかし条件の単語を使用するよう

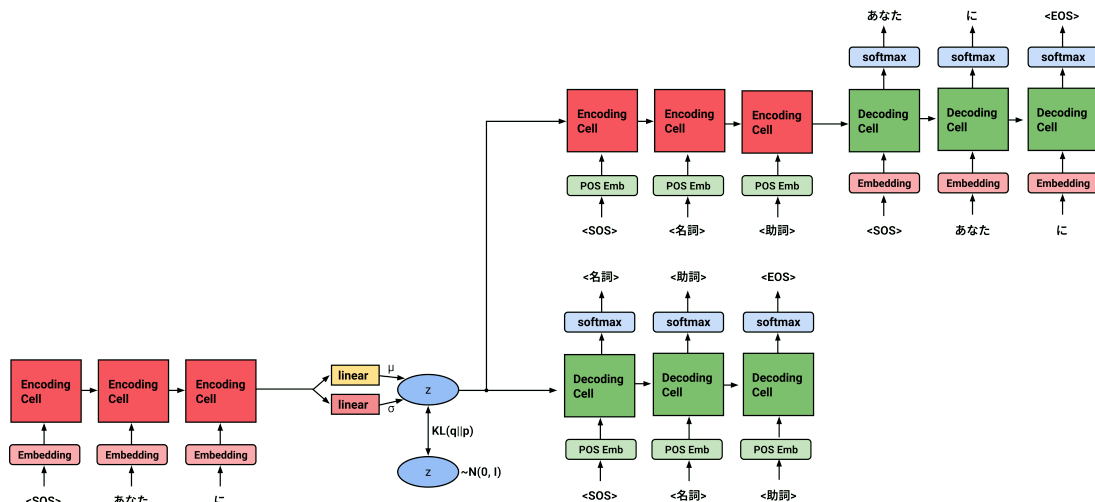


図 2: 提案モデル POS-VAE の全体像

に学習されるため、文構造をスタイルとして生成を行うことはできない。

### 3 提案手法

提案手法である POS-VAE の全体像を図 2 に示す。POS-VAE では VAE をベースとしており、品詞 (Part-Of-Speech, POS) 系列をスタイル情報とみなし、品詞系列を生成するネットワークと単語列を生成するネットワークを同時に学習する。品詞系列を生成するネットワークでは、VAE によってサンプリングされた潜在変数から品詞系列を多様に生成することを目指す。単語列を生成するネットワークでは、入力文をエンコードした中間表現と品詞系列の 2 つの系列から入力文を生成する。これにより、入力文の内容と品詞系列のスタイルに沿った文を生成するような学習が期待できる。

この手法の特徴は制約が緩いことである。単語のような強い制約を加えてしまうと、その単語をなんとかでも使うような学習が行われるため、不自然な文を生成してしまったり、実応用上において使いづらい点が生じてしまう。反対に全く制約を加えなければ、ユーザによるコントロールができないため、望ましい方向の生成がなされない可能性が高まる。したがってその中間である緩い制約が適しているのではないかと考えた。

$$\begin{aligned} \mathcal{L}(\theta; x) = & -\text{KL}(q_{\theta}(z|x)||p(z)) \\ & + \mathbb{E}_{q_{\theta}(z|x)}[\log p_{\theta}(x|z)] \\ \leq & \log p(x) \end{aligned} \quad (1)$$

VAE の学習では目的関数 (式 (1)) にある  $\log p(x)$  の変分下限を最大化する。変分下限では事後分布の近似

モデル  $q_{\theta}(z|x)$  を事前分布  $p(z) \sim N(O, I)$  の KL ダイバージェンスの最小化と、サンプリングした潜在変数  $z$  から入力  $x$  を再構築する期待値の最大化を行う。

$$\begin{aligned} \mathcal{L}'(\theta; x, x_{\text{pos}}) = & -\text{KL}(q_{\theta}(z|x)||p(z)) \\ & + \mathbb{E}_{q_{\theta}(z|x)}[\log p_{\theta}(x_{\text{pos}}|z)] \\ & + \mathbb{E}_{q_{\theta}(z|x)}[\log p_{\theta}(x|z, x_{\text{pos}})] \end{aligned} \quad (2)$$

提案モデルでは品詞系列の多様性のため、サンプリングした潜在変数  $z$  から品詞系列の生成を行う、品詞系列生成モデル  $\log p_{\theta}(x_{\text{pos}}|z)$  を導入する。また、文生成モデルは  $\log p_{\theta}(x|z, x_{\text{pos}})$  として品詞系列を考慮した生成を行う。提案モデルの目的関数を式 (2) に示す。

## 4 評価実験

### 4.1 実験設定

**実験データ** 実験では株式会社サイバーエージェントが保有する、美容・化粧品関連の広告データを使用した。前処理として URL 除去、ハッシュタグの除去、正規化、数字、低頻度語のマスキング、「(\*1)」といった注釈マークの除去を行ったのち、重複文を一つの文にまとめた。品詞は「名詞」や「形容詞」といった IPA 品詞体系における大分類<sup>\*1</sup>を用いた。総レコード数は 49,552 件であり、総単語数は 10,293、平均単語数は 15.9、総品詞数は 16 であった。実験では 44,596 件を学習用、2,477 件を開発用、2,479 件をテスト用に分割した。

**モデル** 品詞系列生成モデル、文生成ネットワークと

<sup>\*1</sup> 形態素解析には MeCab および mecab-ipadic-NEologd を使用した。

表 1: 自動評価の結果

	BLEU	dist-1	dist-2	local-dist-1	local-dist-2	local-dist-3
VAE	<b>0.14</b>	<b>0.053</b>	<b>0.18</b>	0.41	0.70	0.81
POS-VAE	0.12	0.038	0.15	<b>0.42</b>	<b>0.74</b>	<b>0.86</b>

表 2: 人手評価の結果

流暢性	妥当性
<b>2.96</b>	<b>2.30</b>
2.56	1.98

もに 256 次元 1 層の GRU を用いた。単語 Embedding 層は 300 次元、品詞 Embedding 層は 20 次元、潜在空間の次元数は 16 次元である。学習ではバッチサイズ 32 のミニバッチにて 10 epoch 学習させた結果を用いた。また潜在空間の学習を安定させるために KL cost annealing[1] と word dropout[1] を行った。

## 4.2 評価手法

生成された広告文に対して、複数の評価尺度での自動評価と人手評価を実施した。

**自動評価** BLEU は出力文と正解文の重なりを評価するため多様性を考慮できない。したがって多様性を測る指標として distinct [2, 3, 5] を用いて評価する。distinct を式 (3) に示す。

$$\text{distinct}_n(G) = \frac{\text{Unique}_n(G)}{\text{Count}_n(G)} \quad (3)$$

$G = \{g_1, g_2, \dots, g_M\}$  は  $M$  個の生成文集合を表し、 $\text{Unique}_n$  は  $n$ -gram の重複なしのカウント数、 $\text{Count}_n$  は  $n$ -gram の重複ありのカウント数を表す。

distinct はテストデータ全体に対してモデルが生成した文の多様性を示す。しかし提案手法が目標とする多様性とは、1 入力に対する多出力の多様性である。このような評価基準が存在しないため、我々は新たに local-distinct を提案する。local-distinct を式 (4) に示す。

$$\text{local-distinct}_n(G_{\text{multi}}) = \frac{\sum_{G \in G_{\text{multi}}} \text{distinct}_n(G)}{|G_{\text{multi}}|} \quad (4)$$

$G_{\text{multi}} = \{G_1, G_2, \dots, G_M\}$  は各テストデータに対して  $k$  回ずつ生成した生成文全体を表す。ここで  $G_i = \{g_{i1}, g_{i2}, \dots, g_{ik}\}$  は  $i$  番目のテストデータに対する  $k$  個の出力であり、 $g$  は生成文 1 サンプルを表す。実験では  $k = 10$  とした。

**人手評価** 人手評価では生成文の流暢性、妥当性をそれぞれ最低を示す 1 から最高を示す 4 の 4 段階で評価する。流暢性では「出力文自体が日本語的に流暢で

意味が通るかどうか」と聞かれ、生成文がどれだけ文法的・意味的なミスなく生成できているかを評価する。妥当性では「出力文が入力文に対して適切で嘘を言っていないか」と聞かれ、入力文の情報を生成文がどれほど保持できているかを評価する。評価にはテストデータから入力文を 30 件ランダムにサンプリングし、各入力文に対して手法ごとに 10 件の広告文を生成した。各生成文を作業員 4 名で評価スコアを付与した。

## 5 結果と考察

### 5.1 結果

自動評価の結果を表 1 に示す。dist- $n$  は式 (3) の  $n$ -gram の distinct を表し、local-dist- $n$  は式 (4) における  $n$ -gram の local-distinct を表す。BLEU, dist-1, dist-2 では VAE が優位である一方、local-distinct では POS-VAE が優位となった。特に local-distinct では  $n$  が大きくなるほど、POS-VAE が優位な傾向にあることが確認できる。

人手評価の結果を表 2 に示す。流暢性、妥当性ともに VAE が POS-VAE を上回る結果となった。

### 5.2 考察

表 1 より、提案手法が local-distinct で優位であることが確認できた。これは提案手法が同一の入力に対し品詞系列を確率的にサンプリングし、文生成時にスタイルとして制御することで多様な表現が得られたためであると考えられる。

表 3 に 1 入力に対するモデルの出力結果を示す。VAE では「シミは家で簡単に消せる」等のフレーズが一致しており、生成文全体では多様性に欠ける一方、POS-VAE では様々なフレーズが確認できる。

distinct では提案手法よりも VAE が優位であった。この結果は異なる入力間においては VAE がより多様に生成することを示している。これは POS-VAE では品詞系列を多様に生成している反面、文生成モデルが入力品詞系列に横断して適合しやすい単語列を生成している可能性を示唆している。distinct を保持しながら

表 3: 各モデルの生成結果

入力文	【レーザーじゃないシミ対策】 できる美魔女はみんなやってる簡単自宅ケア?
VAE	皮膚科医 (<num>) 「シミは家で消せるよ?」 簡単すぎる方法が大炎上中
	【手遅れのシミは簡単に消える】 皮膚科医 (<num>) 「シミは家で消せるよ」 簡単すぎる方法が大炎上中
	【レーザー治療はもう古い?】 皮膚科医の妻 (<num>) 「シミは家で簡単に消えるよ?」 効果が凄すぎて大炎上中
	【ファンデ不要】 スッピン美人急増中! シミの悩みにはこれ
POS-VAE	シミには並ばない! <person>も愛用したシミ消しの裏ワザが簡単すぎ
	<num>割の日本人女性が知らない? 美肌を維持した<num>歳主婦が簡単すぎ
	<num>代からのシミどこいった? <person>さんも愛用のジェルが話題
	シミしわに直撃! 1日<num>秒ケア付き<num>% off キャンペーン中

local-distinct を最大化することが今後の課題である。

また、POS-VAE がどれだけ品詞系列に沿って文を生成しているかを確認する。POS-VAE の生成文の品詞系列と入力品詞系列との完全一致率は約 20% であった。一方で VAE の生成文と入力文の品詞系列の完全一致率は 0% であった。また、品詞系列を指定して生成したサンプルを表 4 に示す。これらの結果から、POS-VAE は厳密に入力品詞系列に従うように学習しているわけではないが、文長や情報量など文の表現に貢献する部分を考慮して生成できていることがわかる。特に、品詞系列に記号を入力とした場合に正しく括弧が出力文に反映されている事例を鑑みると、この研究の目的に当たるスタイル制御が達成できていることが言える。

人手評価による流暢性、妥当性に関しては、VAE が提案手法を上回った。提案手法ではスタイルを制御できる反面、尤もらしいものではない単語を選択することになるため、所々不自然な表現になってしまうと考えられる。

## 6 まとめ

本研究では、品詞情報をスタイルとみなし、スタイル制御を考慮した多様な文生成手法として POS-VAE を提案した。また、多出力を考慮した多様性の評価指標も考案した。美容・化粧品関連の広告データで学習し、ベースラインの VAE との比較実験を行った。POS-VAE は自動評価で多様性が VAE よりも優位であった一方で、人手評価は VAE が POS-VAE よりも流暢で妥当な文を出力することが分かった。今後は多様性と制御性を維持した、より文法的に自然で内容的に正しい文を生成できるような手法を考えていきたい。

表 4: 品詞を指定した場合の POS-VAE の生成結果。品詞列において「名」は名詞、「副」は副詞、「記」は記号を指す

入力文	お肌が劇的に変わる♡プルプル日中美容クリームを使ってみた
品詞	生成文
名 名	角質層深部ケア
名 名 名	毛穴レス肌用美容液
副 名 名 名	まるでエステ帰りのすっきり肌
記 名 名 記 名	「これ 1 本」ファンデ不要の神コスメ
入力文品詞	お肌の状態を整え、透明感ある肌に導くクリームが凄かった

## 参考文献

- [1] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2014.
- [2] Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *ACL*, 2017.
- [3] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In *NAACL*, 2016.
- [4] 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, 2019.
- [5] Yuchi Zhang, Yongliang Wang, Liping Zhang, Zhiqiang Zhang, and Kun Gai. Improve diverse text generation by self labeling conditional variational auto encoder. *ICASSP*, 2019.
- [6] Diederik P. Kingma, Danilo J. Rezende, Shakir Mohamed, and Max Welling. Semi-supervised learning with deep generative models. In *NeurIPS*, 2014.
- [7] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *NeurIPS*, 2015.