

キーワード条件つき変分 Autoencoder による広告文生成

Keyword Conditional Variational Autoencoder for advertising headline generation

福田 宏幸^{*1}
Hiroyuki Fukuda

^{*1}株式会社 電通
DENTSU INC.

Recently, fast and massive advertising headline production is highly demanded by growing number of digital ad. Many types of headline generation systems have been developed. But, most of these systems generate headlines systematically by rules and lack generation variety. On the hands, the systems which generate headlines almost randomly satisfy such variety but these headlines are not relevant to ad objective. Until now, it is still difficult to satisfy both variety and relevancy. To this end, we propose Keyword Conditional Variational Autoencoder for advertising headline generation. We regulate generation process by relevant keyword while keeping variety by randomly selected input hidden variables. In our experiment, It can generate variety of headlines and obtain headlines which include a relevant keyword.

1. 背景

近年、広告制作の実務においては、デジタル広告のシェア拡大に伴い、早く、大量に、広告文(いわゆるキャッチコピー)を制作することが求められている。デジタル広告においては、掲載単価が安いことに加え、入稿の手続きも WEB 上で簡単に出来ることから、次々と新しい広告を投入することが推奨されている。また、このように大量の広告を制作して、運用していくことで、広告の鮮度を保つことができ、同時に、どういう広告文が消費者の心を掴むのかという知見も得ることができる。結果として、制作の現場への作業負荷が増大しており、AI を活用した負荷の軽減が求められている。

これまでも、ルールベースを中心として様々な広告文生成の手法が開発されてきた[Yamane 14 等]が、生成手法が単純である程、確実な生成が期待できる一方で、バリエーションに限られるという欠点があった。また、人間がルールを設定している為、人間の発想の範疇を超えた、セレンディピティのようなアイデアが出る可能性も少ないと考えられる。

そこで、本研究では、近年研究が盛んになってきている深層生成モデルの一種である変分 Autoencoder (VAE)を用い、バリエーションに富んだ創発的な広告文の生成を目指す。一方で、広告には商品を売るという目的があり、広告文と商品との強い関連性が求められる為、ランダムな文生成ではなく、キーワードによって条件づけられた広告文生成を提案する。

2. 関連研究

近年、深層生成モデルの研究は、深層学習の研究の中でも特に活発に研究がなされているトピックの1つである。特に変分 Autoencoder (VAE) [Kingma 13]や敵対的生成ネットワーク (GAN) [Goodfellow 14]が発表されて以来、関連論文の数が急増している。

これら生成モデルの研究は、画像の生成を対象としたものが中心で、それらが実用的な精度になりつつある一方で、短文に対する生成モデルの研究は、まだ発展途上と考えられている。手法としては、画像同様に VAE を用いたもの[Bowman 15]と、GAN を用いたものがある。GAN においては、画像に比べ、自

然言語の離散的な性質から、Discriminator の勾配が逆伝播しないという課題があり、強化学習の Policy Gradient の手法を用いた SeqGAN [Yu 17]や、誤差関数に MMD による誤差を加えた TextGAN [Zhang 17]など、様々な工夫を凝らした研究がある。また、垂種としてテンプレート文を元に編集ベクトルによって文を生成するモデルも存在する[Guu 17]。

3. 提案手法

広告文と商品との関連性を高める為、訴求内容(価格や機能やイメージ)や訴求ターゲットで条件づける方法が考えられるが、本論文ではキーワード(単語)によって生成を条件づける手法を提案する。広告文のノウハウ本には、「キャッチコピーは、キーワード探し」というような記述もあり、商品特徴や伝えたいメッセージを短文に表現するために、キーとなる単語の選択が重要と考えられている。

具体的には、まずキーワードが選択され、そのキーワードを条件として広告文が生成されたと仮定し、条件つき変分 Autoencoder モデルを適応する。今回の実験では、簡便性の為、キーワードは名詞のみとした。

Keyword:ビール



広告文:暑い夏に、うまいビール。

図1 提案手法におけるキーワードによる条件づけ

3.1 データ

独自に収集した過去の新聞・雑誌の広告文のうち、15 単語以内の広告文およそ 30 万文を使用した。前処理として、MeCab および mecab-ipadic-NEologd を用いて形態素解析をし、id 化を行った。語彙数は 1 万語で足りし、残りは未知語として処理した。データセットを 8:2 に分割し、8 割の訓練データで学習した後、2 割の validation データの精度(誤差)によって、ハイパーパラメータチューニングを行った。

3.2 モデル

VAE のモデルは、[Bowman 15]をベースにし、Encoder および Decoder の双方にキーワードを条件として入力した。Encoder

と Decoder は共に 1 層 LSTM を使い、隠れ層の次元は 200 に設定した。入力文および入力キーワードの単語は共に 200 次元のベクトルの埋め込み表現として入力した。

上記によるネットワークの概略図 2 に示す。推論時には、入力キーワードとランダムにサンプリングされた z を Decoder に入力することで、新たな広告文を生成した。また、ビーム探索も用いている。最適化には Adam を使い validation の精度によって、推論で使用する Epoch を決めた。

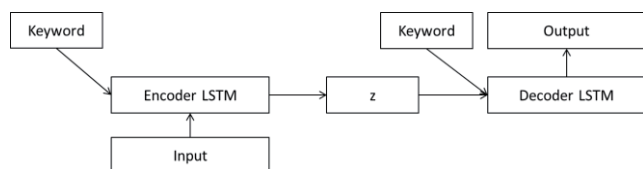


図2 提案手法のネットワーク概略図

4. 実験

4.1 広告文生成例

前述のモデルを学習することにより、実際に得られた広告文の例を図 3 に示す。広告出稿が比較的多い業種である、トイレタリー・飲料・自動車の広告作業を想定し、「シャンプー」「ビール」「燃費」の 3 つをキーワードとして入力した。今回の実験では、各キーワードに対して 10000 文を出力した。変分 Autoencoder は、正しい文を潜在空間上に確率的にマッピングする為、比較的正しい文を生成しやすいと言われている。但し、実際には広告文になっていないものや、そもそも文として成立していないものも多く、人間が目視で選別している。一方で、「低燃費旅、妻対策」など、人間がなかなか思いつかないような言葉の組合せも出現しており、バリエーションの豊富さにも伺える。また、これらの広告文の類似表現が、元のデータに無いことも確認している。

Keyword	生成広告文
シャンプー	「時短シャンプー」そして、 シャンプーリゾート。 春が気持ちいいシャンプーです。 がんばるシャンプーが、ここに。 シャンプー0 円ライフ
ビール	ビールは、美味しく、楽しめる! No.1 ビールの贈り物なら 新ブランドビールで生まれ変わりつづける ビールが時代の街づくりを 春のクラシックビールはもっと楽しんで!
燃費	低燃費コンパクト 「燃費が自由自在」をサポート! 低燃費のヒミツ。 低燃費旅、妻対策 低燃費がヒロインなのには!?

図3 生成された広告文の例

4.2 キーワード出現率

本研究のもう1つの目的は、商品との関連性の高い広告文の生成であり、その為に、キーワードによる条件付けモデルを提案

した。ここでは、その結果の検証として、生成された広告文中のキーワードの出現率を計算した。結果、キーワードを入力していても、必ずしも生成文中にキーワードが出現する訳ではないが、元データにおける出現率に比べて約 50~200 倍程度の大幅な出現率の向上が確認された。

キーワード	生成文での出現率	元データでの出現率
シャンプー	8.96%	0.04%
ビール	2.88%	0.06%
燃費	3.64%	0.03%

図4 提案手法によるキーワードの出現率の比較

5. 結論および考察

本論文では、キーワードを入力条件とした変分 Autoencoder による広告文生成を試みた。結果、元データでの出現率に比べて高い確率で、目的のキーワードが入った広告文を生成することが可能となった。また、得られる広告文にはバリエーションがあり、時には人間が思いつかないようなセレンディピティ的な広告文も生成された。

一方で、必ずしもキーワードが含まれない場合があったり、文として成立していない出力も多く存在した。通常、深層生成モデルの研究では 100 万文~1000 万文以上のデータセットが用いられるが、今回は広告文というデータの特殊性から、利用可能なデータが少なく、精度に影響したと考えられる。また、通常の条件つき変分 Autoencoder は、ユニークなクラスラベルを元に条件付けを行うが、今回の場合は1つの広告文に複数のラベルがついている状態である。結果、より複雑な構造の学習を強いることになり、今回の LSTM による学習器がその構造を十分に捉えられなかったものと思われる。

今後は、文として成立している広告文の精度を上げていくと共に、広告においては、様々なタイプの文生成が要求される為、訴求内容や訴求ターゲット等の他の条件においても効率的な広告文の生成ができるように研究を進めていきたい。

参考文献

- [Yamane 14] 山根 宏彰, 萩原 将文: ウェブにおけるユーザの嗜好を反映したキャッチコピー自動生成システム, 日本感性工学会論文誌, 2014.
- [Kingma 13] D. P. Kingma and M. Welling: Auto-encoding variational Bayes., arXiv, 2013.
- [Goodfellow 15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio.: Generative adversarial nets. In Advances in Neural Information Processing Systems, 2014.
- [Yu 17] Yu, Lantao, Zhang, Weinan, Wang, Jun, and Yu, Yong.: Seqgan: sequence generative adversarial nets with policy gradient., In AAAI, 2017.
- [Zhang 17] Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, Lawrence Carin: Adversarial Feature Matching for Text Generation, arXiv, 2017
- [Bowman 13] Bowman, Samuel R., Vilnis, Luke, Vinyals, Oriol, Dai, Andrew M., Jozefowicz, Rafal, and Bengio, Samy: Generating sentences from a continuous space. CoRR., 2015.
- [Guu 17] Kelvin Guu, Tatsunori B. Hashimoto, Yonatan Oren, Percy Liang: Generating Sentences by Editing Prototypes., arXiv, 2017.