

複数の指定語句を必ず含むリスティング広告の 広告文自動生成

黒木 開¹ 石塚 湖太^{1,2} 川上 孝介¹ 中田 和秀²

¹negocia 株式会社 ²東京工業大学

kai_kurogi@negocia.jp

概要

Google などの検索結果に表示されるリスティング広告の制作現場では、ブランド名や検索キーワードなどの効果が高かった指定語句を広告文に含めることで、広告効果を高めたいニーズがある。しかし、従来の広告文自動生成モデルでは、指定語句を必ず含む保証ができない課題や、複数の指定語句に対応できない課題があった。我々は機械翻訳タスクで提案された指定語句を含む翻訳文生成技術を応用して、複数の指定語句を必ず含む広告文自動生成技術、及び、実運用に向けた高速化手法を提案した。結果として、実用時間内に複数の指定語句を含む広告文が生成可能となった。

1 はじめに

リスティング広告とは、ユーザーが検索エンジンで検索すると、その検索キーワードに関連した広告文が表示される広告である。リスティング広告の広告文は、商品の特徴やアピールポイントなどの訴求点を含み、かつ、検索キーワードと連動した顧客にクリックされやすい魅力的な広告文であることが求められる。そのため、制作現場では広告文を魅力的に作り込む必要があるが、広告文は多様なキーワードに合わせて大量に作成する必要があり、制作現場の限られた人材で、それを行うことは非常に難しい。さらに、リスティング広告の広告取引量は急速なスピードで増加していることもあり、広告文の作成を支援する技術が求められている。

この課題解決に向け、自然言語処理技術を活用した広告文自動生成技術が数多く提案されている。大曾根ら [大曾根 21] は広告文から抽出したキーワードとカテゴリ情報を入力として、GPT-2 を用いた生成技術を提案している。Ishizuka ら [Ishizuka 21] はランディングページ (LP) から抽出した HTML のタ

グ情報と検索キーワードを入力に、BERT2BERT を用いた広告文生成モデルにより、人と同等品質な広告文が生成されることを示している。

ところで、これら広告文自動生成タスクには、ブランド名や検索キーワードなどの効果が高かった指定語句を広告文に含めることで、広告効果を高めたいニーズがある。黒木ら [黒木 21] は、Decode 時に指定されたキーワードの挿入確率を上げることで、キーワードが含まれた広告文を生成する手法を提案している。山田らは [Yamada 21]、キーワードを中心に双方向に文章を生成することで、100%指定語句を含む文章生成手法を提案している。

しかし、前者は指定語句が必ず含まれる保証がなく、後者は指定語句を複数含めることができない問題があった。そこで、本研究では、広告文自動生成タスクにおいて、指定語句が複数指定されていた場合であっても、全ての指定語句を必ず含む広告文を生成する手法である Phrase-based Grid Beam Search(Phrase-based GBS) を提案する。

加えて、指定語句の評価を先頭トークンのみに限定することで、Phrase-based GBS における探索時の評価速度に改善を加えた Phrase-based Approximate Grid Beam Search(Phrase-based AGBS) も併せて提案する。Phrase-based AGBS を用いることで、生成品質をほとんど落とすことなく、生成速度を向上させることに成功した。

なお、本研究は Chousa ら [Chousa 21] が提案した翻訳文生成に対する指定語句を含ませる手法を参考としている。Chousa ら [Chousa 21] は Grid Beam Search(GBS) に LeCA モデル [Chen 20] と呼ばれる手法を組み合わせることで、翻訳精度を落とすことなく指定語句を必ず含む翻訳文生成手法を提案している。我々は、Chousa ら [Chousa 21] が提案したアルゴリズムを広告文生成タスクへ適用し、かつ、実運用に向けた高速化手法を開発した。

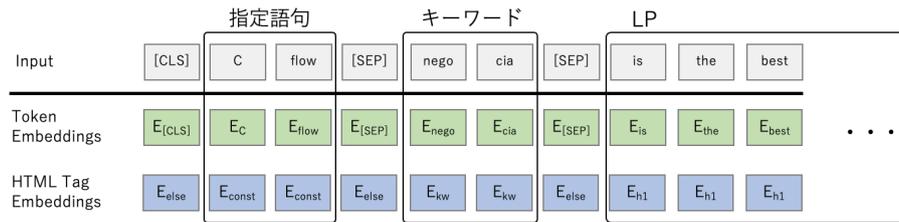


図 1 本モデルの埋め込み層の図。指定語句、キーワード、LP のタグ情報の順に文字列を連結した上で、それぞれに対応するタグを割り当てている。Encoder への入力は 2 つの埋め込みの和である。

2 提案手法

2.1 モデル構造と入力形式

本研究における広告文自動生成モデルは、様々な自然言語処理タスクを "Text to Text" で解くモデルである T5 [Raffel 20] をベースとして用いた。また、より高品質な広告文を生成するために、Ishizuka らの手法 [Ishizuka 21] と LeCA モデルを組み合わせた入力方式を採用する。自動生成モデルの概要を図 1 に示す。埋め込み層に HTML のタグ情報による Tag Embeddings を追加した他、入力は指定語句、キーワード、および LP の HTML からタグ抽出して得られたタグ情報を連結した文字列を用いる。

入力文字列先頭の指定語句は LeCA モデルによるものである。LeCA モデルは入力に指定語句を含めることで、モデルが指定語句を考慮しながら生成を行えるようにする手法であるが、広告文生成の学習時には指定語句の情報がない。そこで本論文では、広告文からランダムに抽出した複数の単語を指定語句として学習時の入力に使用した。これにより、モデルは指定語句を考慮しながら生成する方法を学ぶため、推論時に実際に広告文内に含めたいフレーズを指定語句としたときに、指定語句を含んだ自然な広告文を生成することが可能となる。

2.2 生成手法

リスティング広告の制作現場では、指定語句を広告文に含めたいニーズの他に、サービス名やテンプレート文など、任意のフレーズを一言一句間違えずに生成したいニーズがある。そこで特定の表現をフレーズ単位で扱いつつ、複数の指定語句を必ず含む広告文を生成するために、Phrase-based Decoding と Grid Beam Search (GBS) を組み合わせた Phrase-based GBS を提案する。また、Phrase-based GBS の評価回数を減らすことで高速化させた Phrase-based AGBS を併せて提案する。

2.2.1 Phrase-based Decoding

任意の表現をフレーズ単位で扱うために、Zhang ら [Zhang 17] が提案した Phrase-based Decoding を採用する。Phrase-based Decoding は、生成時に各トークンを生成するのではなく、事前に登録された、複数トークンからなるフレーズ単位で文章を生成する手法である。フレーズ単位で生成することで、事前に登録したブランド名やテンプレート文などのフレーズをそのまま出力することが可能となる。

2.2.2 Phrase-based Grid Beam Search

Phrase-based GBS は、Phrase-based Decoding と GBS を組み合わせた手法である。その疑似コードはアルゴリズム 1 の通りである。GBS は、**Start** (指定語句の最初のトークンを生成する)、**Continue** (生成中の指定語句の次のトークンを生成する)、**Generate** (制約なく次のトークンを生成する) の 3 つの状態を考慮しながら文章生成を行う。それに対して Phrase-based GBS では、**Start** と **Continue** を一体化させることで、指定語句を生成する際にフレーズ単位で生成する。このとき、 k をビームの幅、 C を指定語句の数とすると、タイムステップあたりの計算量オーダーは $O(kC^2)$ となる。

2.2.3 Phrase-based Approximate Grid Beam Search

前述の Phrase-based GBS を用いることで、複数の指定語句を必ず含む広告文を生成できるようになったが、Phrase-based GBS は、各タイムステップごとに全ての指定語句の生成確率を評価しているため、指定語句の数の二乗に比例して計算時間が増える。そこで、指定語句の生成時に、各指定語句の 2 トークン目以降の生成確率を計算せず、その評価を先頭トークンの生成確率のみで行うことにより評価回数を減らす Phrase-based AGBS 法を提案する。本手法により、全ての指定語句の評価を 1 回で行えるため、計算量オーダーを $O(kC)$ まで減らすことができる。

表 1 実験結果の表

アルゴリズム	指定語句数	BLEU	ROUGE-1	ROUGE-2	生成時間 (s)
Phrase-based GBS	2	0.63	0.74	0.68	4.23
Phrase-based GBS	5	0.69	0.84	0.78	13.64
Phrase-based AGBS	2	0.59	0.72	0.65	3.40
Phrase-based AGBS	5	0.67	0.82	0.76	9.57

アルゴリズム 1 Phrase-based GBS の擬似コード

```

1: output ← dict()
2: output[0] ← [BOS]*k
3: constraints ← [None + 全指定語句]
4: for t = 0 to 広告文の最大長 do
5:   numC_list ← sort_constraints(output[t])
6:   for n in numC_list do
7:     for ph in constraints do
8:       if ph is None then
9:         scores ← decoder(output[t])
10:      else
11:        scores ← phrase生成(output[t], ph)
12:      end if
13:      len ← 生成したトークン数
14:      beam ← KBEST(scores)
15:      output[t+len] ← output[t+len] + beam
16:    end for
17:  end for
18: end for

```

3 実験

3.1 実験設定

本実験では、Phrase-based GBS と Phrase-based AGBS のそれぞれで実際に広告文を生成し、広告文の生成品質および、生成速度を比較する。具体的なデータセットは検索キーワードと LP と広告文の関係を学習するために、独自にスクレイピングすることで得られたデータを用いた。各データは検索キーワード、リスティング広告で表示された広告文、LP の HTML から構成されており、学習データ 741,506 件とテスト用データ 82,320 件に分割して使用した。

3.2 評価方法

生成品質は文章の流暢性や類似度を評価する手法である BLEU[Papineni 02] と ROUGE[Lin 04]

で評価する。ROUGE は 1-gram(ROUGE-1) および 2-gram(ROUGE-2) の 2 通りで評価した。ただし今回用いたテストデータはスクレイピングによって得られたデータで指定語句の情報がない。そこで本実験では、広告文からランダムに抜き出した複数のフレーズを指定語句に指定して広告文を生成した。指定語句の数は $C = 2$ or 5 とし、生成時のビーム幅は $k = 8$ とした。

4 結果と考察

4.1 結果

自動評価の結果を表 1 に、指定語句数ごとの生成速度の比較を図 2 に示す。同じ指定語句数で比較した場合、BLUE、ROUGE-1、ROUGE-2 の全ての生成品質指標で、Phrase-based GBS が優位であった。逆に、生成速度は Phrase-based AGBS が常に速かった。

また、表 2、表 3 に実際の生成例を示す。表 2 は Phrase-based GBS による生成文、表 3 は Phrase-based AGBS による生成文である。それぞれの生成文はリスティング広告のフォーマットに則り「タイトル 1 - タイトル 2 - 説明文 1」の 3 要素から構成されている。

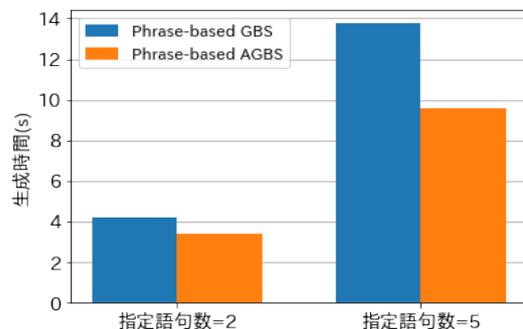


図 2 指定語句数と生成時間のグラフ。指定語句が増えれば増えるほど、2つのアルゴリズムの生成速度差は開く。

4.2 考察

表 1 の結果から、Phrase-based AGBS は生成品質を 3% ~5% 下げる代わりに、生成速度を 24% ~43% 上げ

表 2 Phrase-based GBS による生成例

LP の URL	指定語句	生成文
https://commerce-flow.com	-	Amazon 広告運用ツール - キャンペーン構成を最適な形に - Commerce Flow は国産で唯一のスポンサー広告運用ツールです。
https://commerce-flow.com	AI	AI 広告運用ツール Commerce Flow - キャンペーン構成を最適な形に - 入札額や日予算の最適化により、お客様のビジネスを強力にサポートいたします。
https://commerce-flow.com	効果的, ターゲティング, AI	【公式】Commerce Flow - キャンペーン構成を最適な形に - AI が広告運用を効率化し、効果的に商品ターゲティングを実現!日本最大級の広告ツールです。
https://commerce-flow.com	最適な広告運用	【公式】Commerce Flow - 国産で唯一の広告運用ツール - 入札額・日予算を最適化するために、Commerce Flow が最適な広告運用方法をご提案します。

表 3 Phrase-based AGBS による生成例

LP の URL	指定語句	生成文
https://commerce-flow.com	-	Commerce Flow - キャンペーン構成を最適な形に - 国産で唯一のスポンサー広告運用ツール。入札額・日予算最適化機能付き!
https://commerce-flow.com	AI	AI で商品を世界に発信 - 広告運用ツール/Commerce Flow - 入札額や日予算の最適化により、商品戦略を最大化させます。
https://commerce-flow.com	効果的, ターゲティング, AI	Commerce Flow で広告運用 - キャンペーン構成を最適な形に - AI によるターゲティング効果最大化。キーワードレコメンド機能や入札額最適化機能も効果的。
https://commerce-flow.com	最適な広告運用	国産で唯一の広告運用ツール - キャンペーン構成を最適な形に - Commerce Flow は、スポンサー広告運用を最適な広告運用ツールです。

るアルゴリズムであると言える。品質をほとんど劣化させないまま生成速度を上げることができるため、その実用性は高い。また、図 2 から分かるように、指定語句の数が多いほど Phrase-based AGBS の有効性は高くなると言える。

次に、実際の生成例に注目する。表 2、表 3 はいずれも Commerce Flow と呼ばれる Amazon の広告運用ツールの LP に対して生成された広告文である。表 2 の phrased-based GBS を用いた場合には、いずれの場合にも自然な広告文が生成された。指定語句が複数設定された場合や、複合語が設定された場合にも適切に生成できていることが確認できる。

次に表 3 の Phrase-Based AGBS を用いた場合を確認する。こちらも全体的に自然な広告文を生成できることが確認できた。しかし指定語句に複合語、つまり「最適な広告運用」を指定した場合には、「広告運用」というフレーズが多様され、生成品質の明らかな低下が確認された。

この原因は Phrase-based AGBS の近似の仕様によるものと推測できる。Phrase-based AGBS は指定語句を先頭トークンで近似するため、「最適な広告運用」は「最適」に近似される。つまり後半の「広告運用」部分が見落とされることになるため、同フレーズの多用に繋がったと考えられる。Phrase-based AGBS を利用する場合には指定語句に複合語を設定しないなど、やや工夫が必要であると言える。

5 まとめ

本研究では、複数の指定語句を必ず生成文内に含める Phrase-based GBS を提案した。加えて、Phrase-based GBS を高速化する手法である Phrase-based AGBS を提案した。実験の結果、Phrase-based GBS を用いることで指定語句が全て含まれた高品質な広告文を生成出来た他、Phrase-based AGBS によって生成品質をほとんど落とすことなく高速化出来たことが確認できた。今後は、生成品質を維持しながら、より多様に、より高速に広告文を生成するような手法を考えていきたい。

参考文献

- [大曾根 21] 大曾根 宏幸, 張 培楠. GPT-2 の転移学習によるキーワードを考慮した広告文生成, 2021 年度人工知能学会全国大会 (第 35 回), 2021.
- [Ishizuka 21] Ishizuka, K., Kurogi, K., Kawakami, K., Iwai, D., Nakata, K. Generating Search Text Ads from Keywords and Landing Pages via BERT2BERT, 2021 年度人工知能学会全国大会 (第 35 回), 2021.
- [Zhang 17] Zhang, J., Utiyama, M., Sumita, E., Neubig, G., Nakamura, S. Improving Neural Machine Translation through Phrase-based Forced Decoding, In Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017), pages 11–20, November 28-30, 2017.
- [黒木 21] 黒木 開, 岩井大志, 川上孝介, 石塚湖太, 中田和秀. キーワードを考慮した BERT2BERT による広告文生成, 2021 年度人工知能学会全国大会 (第 35 回), 2021.
- [Yamada 21] 山田 康輔, 人見 雄太, 田森 秀明, 岡崎 直観,

乾 健太郎: 指定語句を確実に含む見出し生成, 言語処理学会第 27 回年次大会, 2021.

- [Hokamp 17] Hokamp, C., Liu, Q. Lexically constrained decoding for sequence generation using grid beam search. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1535–1546, Vancouver, Canada. Association for Computational Linguistics., 2017.
- [Chen 20] Chen, G., Chen, Y., Wang Y., Victor O.K.Li. Lexical-constraint-aware neural machine translation via data augmentation. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization. Main track.(IJCAI-20), pages 3587–3593, 2020.
- [Chousa 21] Chousa, K., Morishita, M. Input Augmentation Improves Constrained Beam Search for Neural Machine Translation: NTT at WAT 2021, The 8th Workshop on Asian Translation(WAT), 2021.
- [Raffel 20] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou Y., Li, W., J. Liu. P. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67., 2020.
- [Papineni 02] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pages 311–318., 2002.
- [Lin 04] Lin, C.-Y. Rouge: A package for automatic evaluation of summaries. In Text summarization branches out, pages 74–81., 2004.