

大規模言語モデルによる脚本データの解析: プロダクト・プレイスメント挿入箇所の探索と評価

山木良輔¹ 檜木悠士² 長沼大樹^{3,4}

¹立命館大学 ²早稲田大学 ³モントリオール大学 ⁴Mila

yamaki.ryosuke@em.ci.ritsumeai.ac.jp yuji.1277@akane.waseda.jp

naganuma.hiroki@mila.quebec

概要

本研究では、プロダクト・プレイスメントの自動化を実現するための手法を提案する。映像作品の脚本を大規模言語モデルによって解析することによって、脚本中において顕在性が高い箇所及び商品が登場する箇所をプロダクト・プレイスメントを挿入可能なシーンとして抽出する。映画脚本における実験の結果より、我々の提案手法が脚本中における重要シーン及びプロダクト・プレイスメント挿入可能シーンを抽出可能であることが示された。

1 はじめに

プロダクト・プレイスメント (PPL: Product Placement) とは、ドラマ・映画等の劇中において、飲食物・役者の小道具・背景として実在する企業ロゴや商品を露出する広告手法である。通常の広告に比べ、映像視聴者が無理に宣伝されているストレスを感じにくく宣伝効果が高いとされているほか [1], 「ブランド認知施策」として欧米や韓国の映像コンテンツで注目を集めており、例えば、米国の 2019 年の PPL 市場は 1 兆 2500 億円と報告されている [2]。また、PPL は CM と違い映像コンテンツが視聴可能な期間中は広告効果を発揮し続けるため、今後も更なる市場価値向上が見込まれている [3]。

PPL を映画やドラマの作品に挿入する際には、担当者が脚本を人力で解析して、PPL を打診する必要がある。この脚本の人力による解析においては、PPL を可能なアイテムの抽出だけではなく、そのアイテムがどれだけ注目されるかを担当者の経験値をもとに評価し、PPL の打診を行う。このプロセスでは、担当者が脚本を受け取ってから、広告主に打診するまでに数ヶ月かかることがある。

本研究では、上記の課題を昨今の大規模言語モデルを用いることで解決可能であるかを検証する。具

体的には、事前学習済みの GPT-2 [4] によって計算されるエントロピーの情報を活用することで、脚本中において顕在性が高く、注目が集まりやすい箇所を評価する指標を複数提案し、また、GPT-3 [5, 6] が持つゼロショット学習能力を活用することで PPL を挿入可能な箇所の抽出を行った。実在の映画脚本を用いた実験により、我々の提案手法を定性的に評価し、有効性を示す結果が得られた。

2 提案手法

本研究では GPT-2 [4] 及び GPT-3 [5, 6] を活用することで脚本中の顕在性・重要性の高い箇所及び PPL 挿入が可能な箇所の抽出を行う。

2.1 GPT-2 を用いた顕在性評価

まず、GPT-2 を用いた手法に関して説明する。ここでは GPT-2 によって計算されるサブワードの出現確率の情報を活用した 2 つの手法を提案する。

2.1.1 サブワードの出現確率に基づくエントロピー

本手法では事前学習済みの GPT-2 を用いることで、脚本中の部分サブワード列に関してサブワードあたりの平均エントロピーを計算する。そして、エントロピーが高いサブワード列を脚本中において顕在性が高い箇所として評価する。これは、エントロピーが高くなるようなサブワード列、すなわち脚本中での表現は周りの文脈から予測することが比較的難しく、他の部分に比べて顕在性が高いという筆者らの仮説に基づく。

まず、脚本中において i 番目に登場するサブワード w_i が出現する確率は文脈情報 c_i に条件付けられた確率 $P(w_i|c_i)$ で表され、この確率は広範なデータセットによって事前学習された大規模言語モデルである GPT-2 によって計算する。そして、 N 個のサブワードを含む部分サブワード列

$S_i = w_{i:i+N} = [w_i, \dots, w_{i+N-1}]$ に関して、サブワードあたりの平均エントロピー $H(S_i)$ は次式で計算される。

$$H(S_i) = -\frac{1}{N} \sum_{j=0}^{N-1} \log P(w_{i+j}|c_{i+j}) \quad (1)$$

ここで、文脈情報 c_{i+j} はハイパーパラメータ M を用いて次の通りに定義する。

$$c_{i+j} = w_{i-M:i+j} = [w_{i-M}, \dots, w_{i+j-1}] \quad (2)$$

以上の定式化に従って、GPT-2 によって計算されるサブワードあたりの平均エントロピー $H(S_i)$ が高い箇所を、脚本中において顕在性が高い箇所として評価する。なお、ハイパーパラメータである N, M は全ての実験において、 $N = 30, M = 100$ に設定した。

2.1.2 エントロピーの比率

前節で説明した手法では、特定の表現に関するサブワードあたりの平均エントロピー $H(S_i)$ のみを顕在性の評価指標として用いる。これに対して、本手法では、特定の表現に関する顕在性の評価指標として、その表現に関するエントロピー $H(S_i)$ とその次に続く表現のエントロピー $H(S_{i+N})$ の比率の情報をを用いる。これは、局面の切り替わりを示す表現においては、その表現自身のエントロピーは高くなることが予想されるが、それに続く表現のエントロピーは低くなるという筆者らの仮説に基づく。

i 番目の表現 S_i とそれに続く表現 S_{i+N} に関するエントロピー $H(S_i), H(S_{i+N})$ の比率を次のように定義する。

$$R(S_i) = \frac{H(S_i)}{H(S_{i+N})} \quad (3)$$

この比率 $R(S_i)$ を脚本中での表現 S_i に関する顕在性の評価指標とし、この値が大きいほど顕在性が高い箇所として評価する。

2.2 GPT-3 を用いた PPL 可能箇所の抽出

次に、GPT-3 [5, 6] を用いた手法に関して説明する。GPT-3 は約 1750 億個のパラメータから構成される大規模言語モデルである。GPT-3 は広範かつ膨大なデータセットから学習された結果として高い汎化性能を獲得しており、文書生成、翻訳、要約などの広範な自然言語処理タスクを指示プロンプトによるゼロショット学習で実現可能なことが知られている。そこで、本研究では GPT-3 のゼロショット学習能力を活用した PPL 挿入可能箇所の抽出手法を 2 つ提案する。

2.2.1 GPT-3 のゼロショット学習

本手法では、GPT-3 に対して、特定カテゴリの商品が登場する文を抽出することを指示するプロンプトと脚本の一部分を与えることで、脚本中において PPL が挿入可能な箇所を抽出する。例えば、飲料の PPL を挿入可能な箇所を検索する場合には、GPT-3 に対して “Select all of the following sentences that contain beverages” というプロンプトと脚本の一部分を入力することで、脚本中において飲料に関する表現が登場する文の抽出を行う。ただし、前章で説明した GPT-2 を用いた顕在性の評価手法とは異なり、本手法は必ずしも脚本中における表現の顕在性の高さを考慮することはできない。

2.2.2 SimCSE によるテンプレートマッチング

GPT-3 と SimCSE [7] を用いた商品登場箇所の抽出方法を検討する。SimCSE は対照学習を用いた文ベクトル獲得手法であり、文の等価性を評価する Semantic Textual Similarity タスクにおいて高い性能を示した。SimCSE では、同じ文で異なる Dropout から得られる 2 つの文ベクトルを正例、異なる 2 文から得られる 2 つの文ベクトルを負例として対照学習を行うものである。我々は SimCSE を 2 文の類似度を測る手法として用いることで、脚本中において商品が登場する箇所の抽出を行う。

商品登場箇所の抽出方法は次の 2 段階に分かれる。(I) 商品登場箇所のテンプレート生成と (II) テンプレートに類似した箇所の探索である。まず (I) では GPT-3 を用いて、特定カテゴリの商品が登場するような自然な文 (テンプレート) を複数生成する。例えば、飲料が登場する文を生成する場合には、“Generate 50 sentences in which a beverage appears, as it would be in a movie scenario.” のようなプロンプトを GPT-3 に与えることで、“She nervously sipped her coffee as she waited for her date to arrive.” といったテンプレートを複数生成する。そして (II) では、(I) で生成したテンプレートと SimCSE による類似度が高い箇所をシナリオ中から探索する。これによって商品が登場すると考えられる箇所をシナリオ中から抽出できる。また、本手法も前節で説明した GPT-3 によるゼロショット学習と同様に、必ずしも脚本中における表現の顕在性の高さを考慮することはできない。

3 実験

前章で説明した各手法による映画脚本の解析を行う。なお、実応用の場面では、様々なカテゴリの商品が PPL の対象となりうるが、本研究では評価の簡便性のために、対象となる商品カテゴリを飲料に限定する。

3.1 実験データ

著名なハリウッド映画の脚本 3 本を [IMSDb](https://www.imsdb.com/)¹⁾ から取得し実験に用いた。脚本の選定にあたっては、アクション・コメディ・ロマンスの 3 つの異なるジャンルから 1 作品ずつを選定した。以下が実験に用いた作品の一覧である。括弧は抽出候補となる表現の総数(シーン数に比例)を示す。

- アクション：Top Gun (899)
- コメディ：The Hangover (326)
- ロマンス：La La Land (833)

3.2 事前学習済みモデル

GPT-2 の事前学習済みモデルには Hugging Face [8] で公開されている gpt2-xl を用いた。また、GPT-3 には OpenAI API によってアクセスし、事前学習済みモデルには text-davinci-003 を使用した。さらに、SimCSE の事前学習済みモデルには sup-simcse-roberta-large を使用した。

3.3 評価方法

各提案手法において脚本から重要箇所または PPL 挿入可能箇所として抽出された表現を定性的に評価する。それらの表現が登場する映画中でのシーンが映画全体の中でどのように重要な局面か、また、そのシーンにおいて PPL を挿入することが可能か否かに関して分析を行い、手法間での比較を行った。サブワードあたりの平均エントロピー及びその比率を評価する手法においては、計算された評価指標が上位 10 % 以内の表現を抽出し、定性評価を行った。

4 結果

実験に用いた 3 作品の脚本に関して、4 つの提案手法がそれぞれ抽出したシーンの一部を抜粋したものを表 1, 2, 3 に示す。各表において、Entropy, Ratio, GPT-3, SimCSE はそれぞれ 2.1.1 節, 2.1.2 節, 2.2.1 節, 2.2.2 節で説明した各手法を表しており、

1) フェアユース規定に則り、著作物を研究目的で使用した。
<https://imsdb.com/>

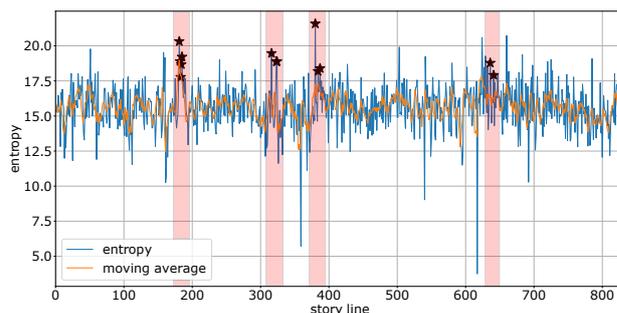


図 1 La La Land におけるエントロピーの推移

Entropy 及び Ratio の列に記載した括弧付き数字は各手法において抽出された表現の総数を表している。

また、La La Land の脚本におけるサブワードあたりの平均エントロピーの推移を図 1 に示す。なお、グラフ中で★が付されている部分は、表 3 に記載した各シーンに含まれる表現が脚本中において出現した箇所を示している。

5 考察

5.1 解析手法間の比較

ここでは、実験結果より確認された各解析手法の特徴に関する考察を述べる。まず、サブワードあたりの平均エントロピーを評価する手法 (Entropy) は脚本中における多くの重要シーンを抽出することに成功している。例えば、表 2 の 3 つ目のシーンのように、物語の重要な転換点となるシーンを抽出することができる。また、飲料に関する表現を抽出することを主たる方策とはしていないものの、抽出されたシーンの中には実際に飲料が登場したり、飲料が登場させることが可能なものが含まれており、PPL 挿入の観点からも妥当な結果が得られていることがわかる。

一方で、エントロピーの比率を評価する手法 (Ratio) は多くの場合において適切なシーンを抽出できないことがわかった。本手法では、急激なエントロピー変化を捉えることはできるものの、図 1 及び表 3 に示した 1 つ目のシーンのように、あるシーンに含まれる表現のエントロピーが継続的に高い、すなわち顕在性が高い状態が続いているシーンに関しては、隣接する表現間のエントロピーの比率が小さくなるため、このようなシーンを抽出することはできない。

次に、GPT-3 のゼロショット学習を用いた手法では、飲料が登場するシーンを高精度で抽出可能であることがわかった。一方で、本手法では表現の顕在性を考慮しないため、物語全体の流れにおいて重要

表 1 Top Gun から抽出されたシーン

Scene	Entropy	Ratio	GPT-3	SimCSE
1 バーでマーヴェリックとチャーリーが初めて会うシーン	○ (10)	×	○	○
2 訓練後、マーヴェリックとヴァイパーがバーで話すシーン (実際の映画中ではロッカールームでのシーンに変更)	○ (1)	×	×	○

表 2 The Hangover から抽出されたシーン

Scene	Entropy	Ratio	GPT-3	SimCSE
1 物語の冒頭においてパーティー後の荒れたスイートルームのシーン (実際の映画中では MONSTER ENERGY の空き缶が登場)	○ (1)	×	○	○
2 ストリップクラブでのシーン (実際の映画中ではエンドロールに登場)	○ (5)	○ (1)	○	○
3 物語中のメインの課題である人質の受け渡しを砂漠で行うシーン	○ (4)	×	×	○

表 3 La La Land から抽出されたシーン

Scene	Entropy	Ratio	GPT-3	SimCSE
1 主人公の男女がパーティーで偶然再開し初めて会話するシーン (実際の映画中では Mountain Dew が登場)	○ (5)	×	○	○
2 主人公の男女が映画館で初めてキスをしようとするシーン (シーン上、飲料を含めることも可能)	○ (3)	×	×	×
3 男性主人公がその後の成功するバンドのリーダーの男に ジャズバーで会い加入を誘われるシーン (実際の映画中では緑の瓶ビールが登場)	○ (3)	×	×	×
4 男性主人公が実家に帰っていた女性主人公を車で迎えに行き 二人の新たな挑戦が始まるシーン (実際の映画中では女性主人公がコーヒーを持ってくる)	○ (2)	○ (2)	○	○

ではないシーンも抽出される傾向がある。また、飲料が明示的に含まれていない表現であっても、周囲の文脈を考慮した上でその表現を飲料に関するものとして抽出可能であることがわかった。

最後に SimCSE を用いたテンプレートマッチングによる手法は、GPT-3 による手法に比べて飲料に関する表現をより網羅的に抽出可能であることがわかった。すなわち、飲料に関する表現の抽出に関して、再現率は高いが適合率が低くなる傾向がある。これは、実験時に設定した表現間の類似度に関する閾値に依存するため、実応用を考えた場合にはこれらのハイパーパラメータを慎重に設定する必要があると考えられる。また、SimCSE が参照するテンプレート表現に関しても、より表現の多様性を大きくするなど、さらなる改善が必要であると考えられる。

5.2 映画ジャンル間の比較

ここでは、映画のジャンルごとに抽出されたシーンの比較及び考察に関して述べる。まず、Top Gun のようなアクション映画では、戦闘シーンなどの日常生活とは乖離の大きいシーンが脚本中に多く含まれているため、他ジャンルの映画に比べて PPL を挿入可能なシーンが相対的に少ない。そのため、顕在性に基づいて抽出されたシーンの中で PPL が挿入可

能なものは限られていた。

一方で、The Hangover のようなコメディ映画及び La La Land のようなロマンス映画では日常的描写が脚本中に多く含まれているため、比較的多くの PPL 挿入可能箇所が抽出可能である。ただし、The Hangover の場合には、表 2 の 2 つ目のシーンのように、実際の商品露出による広告効果の観点から、商品にネガティブな印象を与えかねないシーンが抽出されてしまう。このため、実応用を考えた場合にはシーンの顕在性評価と感情分析を組み合わせた形で用いることが望ましいと考えられる。

6 おわりに

本研究では PPL の自動化という目的に対して、大規模言語モデルを用いて映像作品の脚本から顕在性の高い箇所及び商品が登場する箇所を抽出する手法を複数提案し、実験及び評価を行った。実験結果より、提案手法は映画脚本から重要シーン及び PPL 挿入可能シーンを抽出可能であることが示された。

本研究で提案した各手法にはそれぞれの長所及び短所が存在することから、実応用に向けて、これらの手法を統合することでさらなる改善を行っていくことが今後の課題として挙げられる。

謝辞

本研究は、2022 年度 NEDO Technology Commercialization Program の支援を受けたものである。

参考文献

- [1] Cristel Antonia Russell. Expanding the agenda of research on product placement: A commercial intertext. **Journal of Advertising**, Vol. 48, No. 1, pp. 38–48, 2019.
- [2] Hanna Górska-Warsewicz and Olena Kulykovets. Product placement as an effective tool for the brand positioning. **Brand Management: Advancing Insights on**, Vol. 21, , 2017.
- [3] A Kjaernested and TM Nielsen. Product placement in tv shows—assessing product placement effectiveness and the influence on brand equity. **MSc Marketing communications Management MSc International Marketing and Management. Department of Marketing, CBS**, 2012.
- [4] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. **OpenAI blog**, Vol. 1, No. 8, p. 9, 2019.
- [5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. **Advances in neural information processing systems**, Vol. 33, pp. 1877–1901, 2020.
- [6] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. **arXiv preprint arXiv:2203.02155**, 2022.
- [7] Tianyu Gao, Xingcheng Yao, and Danqi Chen. Simcse: Simple contrastive learning of sentence embeddings. **arXiv preprint arXiv:2104.08821**, 2021.
- [8] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface’s transformers: State-of-the-art natural language processing. **arXiv preprint arXiv:1910.03771**, 2019.