

## メタファーの自動生成による意味的な 重ね合わせのあるキャッチコピー生成器の構築

常勤研究者の部



代表研究者 狩野 芳伸  
静岡大学情報学部  
准教授

共同研究者 梅村 奏子  
静岡大学大学院  
修士課程2年

### はじめに

近年、AIによってキャッチコピーを自動生成するシステムを見かけることが多くなった。商用利用されているものから一種のジョークツールとして公開されているものまで幅広く存在している。しかし、これらのシステムにおいて良いキャッチコピーとは何かについてはあまり追及されていない。多くのシステムが人間のキャッチコピー作成の支援システムの側面が強く、最終的に良いものを人が選択するという運用がなされているためである。

そこで、まず良いキャッチコピーとは何かに関する考察を行う。呉本ら [1] の研究によると、キャッチコピーが世間で話題になり、消費者にその文言を記憶してもらうことが良いキャッチコピーにおいて重要なことである。更に、修辞技法を用いた広告の方が消費者の記憶に残りやすい [2] ことや、キャッチコピーのコンテストで良い成績を収めるような、プロのコピーライターが「良い」と判断するキャッチコピーには意味の重ね合わせが使われている

[1] ことが分かっている。以上を総合すると、修辞技法を用いることで消費者の記憶に残りやすくし、更に意味の重ね合わせを含んだキャッチコピーが良いキャッチコピーということになる。そこで、良いキャッチコピーを生成するために、修辞技法の1つである比喻表現に着目する。比喻表現は1つの事柄を別の事柄で例えることにより成立する表現であることから、修辞技法としての役割にとどまらず意味の重ね合わせが生まれることも期待できる。また、キャッチコピー生成に関する研究の多くは商用利用を目的としているため、単なる広報手段の1つとしてキャッチコピーを捉えているケースが散見されるが、キャッチコピーは発信者（多くの場合は営利企業）と受け手（多くの場合は消費者）との間に成立するコミュニケーションの1種である。その証拠として、日常会話においても頻繁に起こり得るメッセージの解釈の不安定性がキャッチコピーにおいても存在する[1] ことが挙げられる。解釈の不安定性とは、コミュニケーションの特徴の1つであり、発信者がある意図を込めて発信した言語表現に対して、受け手が発信者の意図通りの解釈を行わなかったり、受け手によって解釈が異なったりする現象の事である。このような、コミュニケーションとしての特徴を踏まえてキャッチコピーについて考察すると、ある特定の意図を込めたキャッチコピーについて全ての受け手に対してどのような状況においても一意に解釈してもらうことは、解釈の不安定性が発生するために不可能であることが分かる。しかし、本研究のような修辞技法を用いた生成手法では、最初から発信者側の意図を限定せず解釈の幅を広く持たせた状態で受け手に発信するため、意図が全く伝わらないという状況が起きにくく、解釈の不安定性を生成の段階で持たせることにより幅広い範囲にメッセージを届けることができると期待される。なお本研究では、ウェブ広告によくみられるような特定の商品やその属性そのものを直接的に表現する類のコピーではなく、ブランド広告のようなより抽象的な表現を必要とするキャッチコピーにフォーカスを置く。

以上より、本研究では、キャッチコピーをコミュニケーションの手段として捉え、比喻表現と意味の重ね合わせを用いた良いキャッチコピーの生成を目指す。

## 関連研究

### 比喻について

比喻表現生成に関する計算論的アプローチを試みた研究は多くない[3]。そ

して、そのほとんどが名詞メタファーや直喩を対象にしたものであり、知識源や手法に違いはあるものの、被喩辞と特徴語を入力として適切な喩辞を選択することで喩を生成するという方針は一致している。北原ら [4] の行った辞書等の言語資源に基づく概念間の類似度計算を用いたシステムをはじめ、近年では中條ら [5] の文脈情報を用いたシステムや、佐藤らの word2vec とソーラスを用いたシステム [6] 及び単語の分散表現と意味カテゴリを用いたシステム [7] などが提案されている。これらの研究は、喩生成支援システムという側面が強く、システムの生成結果を人間が選択するもしくは参考にすることを前提としている研究が多い。本研究ではキャッチコピー生成システムの一部として喩生成を行う関係上、人手での選択を行うことは難しい。

喩表現には、直喩のような名詞や形容詞を用いるものだけではなく、動詞を用いるものも存在する。その代表例が暗喩（メタファー）である。Lakoff ら [8] は、概念間の写像関係を表す概念メタファーをもとにメタファーが生成されるという概念メタファー理論を提唱した。概念メタファー理論の詳細については図 1 に示す。



図 1 概念メタファー

鍋島 [9] はこの理論の動詞メタファーへの適応可能性について「水は感情」という概念メタファーを例に証明した。また、大石 [10] による研究では、名詞と動詞の共起情報から機械的に概念メタファーを同定可能であることを示唆した。更に、我々の先行研究 [11] では、以上の研究を踏まえた概念メタファー理論に基づいた喩表現の生成に成功している。

以上より本研究の目的である意味の重ね合わせを達成するためには、直喩のような名詞や形容詞によるメタファーではなく、動詞による概念間の関係を考慮したメタファーが最適であると考え、概念メタファー理論に基づく喩生成をベースラインとする。

## キャッチコピーについて

キャッチコピー生成に関する研究は、大きく分けて 2 つの手法に分けられる。

1つ目はテンプレートを用いる手法である。この手法では、もととなるキャッチコピーの一部（多くの場合は名詞部分）を変換して新たなキャッチコピーを生成する。山根らは Web 情報を元にした知識ベースとトピックモデルによるテーマ選択を組み合わせた生成システム [12] や SNS 上の知識を元に bag-of-words に基づく線形回帰モデルを用いてユーザに好まれる単語を推定し生成を行うシステム [13] を提案した。岩間ら [14] の研究ではプロのコピーライターが作成した独自のキャッチコピーテンプレートを用いて、単語のベクトル化や格フレームの分析に基づいた生成を行うシステムを提案した。これらのシステムでは、人間では思いつかない斬新な表現の創出に成功しており、人手でのキャッチコピー選択の必要性は存在するものの、人間がキャッチコピーを考える際の補助として用いるには十分な性能を発揮することが分かった。

もう1つは機械学習を用いて生成を行う手法である。この手法は前者とは異なりキャッチコピー全体を生成する。福田 [15] は、キーワードを入力とした変分 Autoencoder による生成を行った。張ら [16] はサンプリング手法の一種である MH アルゴリズムを用いて単語や語句を得る手法である GCMH を使ってキーワードからキャッチコピーを生成するシステムを提案した。脇本ら [17] は seq2seq をベースに Web ページから広告対象に最適なキーワードを抽出して利用するシステムを提案した。これらの研究では、学習データとして用いたキャッチコピーには存在しない表現など斬新でパリエーション豊富なコピーが生成できることを示した。一方で、日本語の文や広告文として成立しないものも多く生成されている。これは、人間の発想では出てこないような文が生成されるという強みがある一方、理解可能な生成結果が限られるというマイナス面もあるという機械学習による文生成一般に言える内容がキャッチコピー生成においても当てはまることを示唆している。また、教師付き機械学習においては学習データの量が生成結果に影響を及ぼすため、利用可能な大規模データのないキャッチコピー分野においては特に精度が低くなりやすいのではないかと考えられる。

以上の内容を踏まえ、機械学習における十分な学習データ確保の困難さや、ルールベースで行う比喩生成との組み合わせやすさを考慮して本研究ではテンプレートを用いた生成手法を使用することにする。

## 提案手法

### 抽象度推定

名詞単語がどの程度抽象的であるかに関する指標として抽象度を策定する。抽象度は1を最も具体的、5を最も抽象的として連続値で算出する。抽象度の算出方法については、日本語抽象度辞書 [18] に準拠するが、固有名詞や顔文字、本研究では使用しない動詞や形容詞などが多く含まれているため、これらを取り除いたものと新たに人手による抽象度付与を行った名詞とを合わせて学習データとする。学習データに出現する単語を word2Vec でベクトル化したものについて多層パーセプトロンで予測モデルを構築する。以上の手法で構築した抽象度推定器を用いて抽象度の推定を行う。

### 多義語クラスター

動詞メタファー内の動詞がその名詞領域において複数の語義で使用される場合、その比喩表現は多義的に解釈できると仮定し、多義名詞のクラスタリングを行う。多義名詞の各語義においてどのような動詞表現が使用され得るかをまとめた辞書やコーパス等は見つけることができなかつたため、新たに作成することにした。小説投稿サイト小説家になろう<sup>1</sup>で純文学ジャンルに分類される小説 1059 作品について、kuromoji [19] を用いて形態素解析を行い、名詞と判定された単語かつ小学館の中日・日中統合辞典 [20] の見出し語の語義数が2つ以上のものを多義語と定義した。クラスタリング手法は、主に速度の観点から K-Means 法を用いる。手順は以下の通りである。

1. 多義名詞の抽出
2. 1. で抽出した各名詞について出現するごとに前後 100 単語を取って共起する語をカウントする
3. 名詞のクラスタリング
  - ・ k-means 法でクラスタリング
  - Ex.) 名詞「猫」  
その出現回数分の「猫」について2を実施し、それら全ての「猫」をクラスタリングする
  - ・ クラスタ数については語義数 +3 とした

<sup>1</sup> <https://syosetu.com/>

## 比喩生成

先行研究と同様に、比喩表現の対象となる名詞1単語の入力から名詞と動詞からなる比喩表現の生成を行う。ただし、本研究では名詞-動詞間の共起情報に加え、目標領域となる名詞と起点領域となる名詞の間の抽象度の差から概念メタファーの同定を行う。また、本研究の目的の1つである多様な解釈が可能な意味の重ね合わせのある表現の生成を行うため、多義名詞と、その名詞領域の複数語義で使用される動詞からなる比喩表現の生成を行うことを目指す。生成の手順は以下の通り。

1. 入力名詞と共起する（格関係を持つ）助詞-動詞の組を取得する
  - ・共起の取得には京都大学格フレーム [21] を用いる
2. 1 と共起する名詞を取得する
3. 2 について抽象度を求め、入力名詞との抽象度の差が絶対値 0.5 以上となる名詞を起点領域候補とする
4. 3 及びキャッチコピーテンプレートで用いられる助詞と共起する動詞を取得する
  - ・1 で得た動詞と一致するものは除外する
5. 4 の動詞の中から入力名詞の複数語義で使用されるものを選択する
  - ・入力名詞が多義語クラスターに含まれない場合や、複数語義で使用される動詞が存在しない場合は、4 の中からランダムで選択する
6. 入力名詞、テンプレートで用いられる助詞、5 の動詞を合わせて比喩表現とする

## キャッチコピー生成

本研究では、テンプレートを用いた生成手法を応用し入れ替えの際に比喩を用いることにする。手順は以下の通り。

1. 生成に用いるテンプレートをランダムに選択する
2. テンプレートから入れ替えの対象となる格関係を持つ名詞と動詞の組及びその格助詞を探す
  - ・格関係解析には日本語構文・格・照応解析システム KNP4.19[22] を用いる
3. 入力名詞、2 で見つけた格助詞を用いて比喩表現を生成する
  - ・入力単語が複数ある場合は、ランダムに使用する単語を選択する

- ・ 入れ替えを行う部分以外との整合性を保ち、キャッチコピー全体の文法性を担保するためにテンプレート内で使用されている格助詞を用いて比喩生成を行う
4. 生成した比喩表現の動詞を2の動詞と同様の活用に変換する
    - ・ 3と同様の観点から動詞の活用形及び、動詞の後ろについていた助動詞や接続助詞についてもテンプレートと同じものを用いる
  5. 生成した比喩表現について名詞と動詞をそれぞれ1で見つけた名詞、動詞と入れ替える

## 実験と考察

本研究では、生成したキャッチコピーについて主観評価実験を行った。実験では、文生成において重要な観点である日本語として意味が通るものが生成されているか、キャッチコピーとして重要な受け手に関心を持たせることができるか、更に比喩表現を用いたことによる意味の重なりを感じられるかといった点を評価するために以下の6項目の評価軸を用意した。

- 1 日本語として成立しているか
- 2 色々な意味で解釈できるか
- 3 意図を理解できる気がするか
- 4 興味深い表現だと感じるか
- 5 おもしろいか
- 6 コピーの対象について知りたくなるか

各項目について1(全くそう思わない)から5(とてもそう思う)までの5段階評価で、評価者3名によって実験を行った。また、ベースラインとして抽象度推定器及び多義語クラスターの導入されていない状態のシステムで生成したキャッチコピーを用いた。提案手法、ベースラインともに600作品のキャッチコピーを生成し、そこからランダムで100作品を選んで評価実験用データとした。

評価軸ごとの評価者の2名以上一致率を以下表1にまとめる。なお、各評価軸については上記の項目番号を用いて表した。

表1 キャッチコピー評価実験での評価者2名以上一致率

評価軸	1	2	3	4	5	6	合計
ベースライン	0.71	0.53	0.74	0.62	0.66	0.65	0.65
提案手法	0.73	0.51	0.67	0.63	0.50	0.57	0.60
全体	0.72	0.52	0.71	0.63	0.58	0.61	0.63

2名以上一致率は概ね高い値を示しているが、評価軸2の「色々な意味で解釈できるか」に関してはベースライン、提案手法共に低くなっている。6つの評価軸の中で1番、評価者ごとの感性に左右されやすい項目であることが分かる。評価軸3の「意図を理解できる気がするか」が高い一致率を示しているにもかかわらず2の一致率が低いということは、評価者によって汲み取った意図が異なるということを示している。また、ベースラインと比較して提案手法の一致率が下回る傾向にあることから、比喩表現によって評価者による解釈の異なりが生まれたのではないかと推察する。

評価軸ごとの平均値を表2にまとめる。

表2 キャッチコピー評価実験結果（平均）

	1	2	3	4	5	6	合計
ベースライン	3.08	2.91	2.33	2.59	2.56	2.73	16.19
提案手法	3.14	2.85	2.38	2.70	2.58	2.77	16.41

評価軸2を除いて、提案手法がベースラインを上回っているが、評価者によるばらつきが大きかった。

以下、全評価軸の平均値の合計が上位20位に入ったキャッチコピーについて考察する。ベースラインでは「よし、加速で死のう!」、「素敵を教える電気」、提案手法では「読みます、あなたの熱意を。」、「どうせ払うなら、自分の社会に誕生する方で。」などの動詞メタファーを含むキャッチコピーの生成に成功している。これによって抽象度を用いない共起情報のみによる概念メタファー推定でも比喩表現が生成可能であることを示唆している。また、比喩生成器の出力結果にかかわらず元のキャッチコピーテンプレートに含まれていた単語と組み合わせさせた結果として良いコピーが生まれる場合があることから、テンプレート内における文脈について考慮しながら生成を行うことで、更なる精度の向上を図ることができると考えられる。一方、生成に失敗した例では、テンプレートを解析する際の形態素解析や活用形変換の精度に問題があることが分か



る。これらについては、個別事例にたいしてルールを作って対応するか、解析器の側に与える辞書を工夫することで解決を試みるかいずれかの手段を講じる必要がある。

## 課題と展望

多義語の問題は、テキストの字面のみでは判断が難しいため、自然言語処理分野一般にも未解決の困難な問題である。本研究で試みた多義語クラスターでも、十分な精度を得ることができなかった。本研究で十分に検証することのできなかった多義語分類結果の妥当性に関して、検証を行い精度の向上を図ることが必要である。加えて、分類の際のクラスター数の決定方法についてもより幅広く検討する必要がある。比喻生成器については、抽象度推定器を組み込むことによる生成結果の向上を確認することができなかったことが大きな課題である。実際に概念メタファーを形成する上でそれぞれの概念間にどの程度の抽象度差が存在するべきかについて検証し、最適化を図ることが課題解決の糸口となるだろう。キャッチコピー生成器全体の課題としては、多義語クラスター及び抽象度推定器を導入した比喻生成器を用いた場合に明確な数値的精度の向上が見られなかったことが挙げられる。また、前章で述べたように一部で日本語として成立しないキャッチコピーが生成されることもあることから、テンプレートの解析や動詞の活用変換のさらなる精度向上も課題として挙げられる。更に、生成に用いる元のキャッチコピーテンプレートの文脈情報を加味して生成を行うことで更なる性能の向上が期待できると考えられる。

今後の展望としては、まずは上述の課題の解決を図り、比喻表現、キャッチコピーともに生成の質を上げることを目標としたい。また本研究では、ブランド広告によくみられる抽象的な表現の多いキャッチコピーを対象としたが、背景情報として商品等の属性を与え適切にテキスト表現にくみこむことで、ウェブ広告によくみられる特定の商品そのものの購買を促すタイプのキャッチコピーにも拡張を試みたい。

## 参考文献

- [1] 呉本彰子, 赤岡仁之. 広告コピーの解釈多元性に関する研究 -- 美容関連製品のキャッチコピーをケースにして. 武庫川女子大学紀要 人文・社会科学編. vol. 55. pp. 99-105, (2007).
- [2] Tom, Gail, Eves, Anmarie. The Use of Rhetorical Devices in Advertising. Journal of Advertising Research. vol. 39, no. 4. p. 39, (1999).
- [3] 内海彰. 比喩理解への計算論的アプローチ. 認知科学. vol. 20, no. 2. pp. 249-266, (2013).
- [4] 北田純弥, 萩原将文. 電子辞書を用いた比喩による文章作成支援システム. 情報処理学会論文誌. vol. 42, no. 5. pp. 1232-1241, (2001)..
- [5] 中條寛也, 松吉俊, 内海彰. 意味空間に基づく文脈情報を用いた比喩生成. 情報処理学会音声言語情報処理研究会 (IPSJ-SLP) 研究報告, vol 14, pp. 1-10 (2017).
- [6] 佐藤遼河, 杉本徹. コーパスとシソーラスを用いた比喩生成. 2018年度人工知能学会全国大会論文集. (2018).
- [7] 佐藤遼河, 浅見一樹, 杉本徹. “単語分散表現と喩辞の意味カテゴリを用いた比喩の作成支援”. 情報処理学会第80回全国大会講演論文集. 2018, pp. 565-566.
- [8] Lawler, John M., Lakoff, George, Johnson, Mark. Metaphors We Live by. Language. vol. 59, no. 1. p. 201, (1983).
- [9] 鍋島弘治朗. 日本語のメタファー. くろしお出版, 2011.
- [10] 大石亨. “共起情報を用いた概念メタファーの発見”. 言語処理学会年次大会発表論文集. 2005, pp. 392-395.
- [11] 梅村奏子, 狩野芳伸. “類似単語を元にしたメタファー生成の検討”. メタファー研究会. 関西大学千里山キャンパス, 2018.
- [12] Yamane, Hiroaki, Hagiwara, Masafumi. “Tag Line Generating System Using Information on the Web”. Proceedings of the Fuzzy System Symposium. 2012, pp. 1008-1013.
- [13] Yamane, Hiroaki, Hagiwara, Masafumi. Advertising Slogan Generation System Reflecting Preferences of Users on the Web.

- Transactions of Japan Society of Kansei Engineering. vol. 13, no. 4. p. 493-500, (2014).
- [14] Iwama, Kango, Kano, Yoshinobu. “Japanese Advertising Slogan Generator using Case Frame and Word Vector”. The 11th International Natural Language Generation Conference (INLG 2018). Tillburg, Netherlands, 2018, pp. 197-198.
- [15] 福田宏幸. キーワード条件つき変分 Autoencoder による広告文生成. 2019 年度人工知能学会全国大会論文集. (2019).
- [16] 張浩達, 上垣外英剛, 高村大也, 奥村学. 複数の言語モデルを考慮したキーワードからの広告文生成. 2019 年度人工知能学会全国大会論文集. (2020).
- [17] 脇本宏平, 川本峻頌, 張培楠. インターネット広告におけるキーワードに基づく広告文の自動生成. 2020 年度人工知能学会全国大会論文集. (2020).
- [18] 奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室. “日本語抽象度辞書「AWD-J: Abstractness of Word Database for Japanese common words」概要”. <http://sociocom.jp/~data/2019-AWD-J/>.
- [19] “kuromoji: japanese morphological analyzer”. <http://www.atilika.org/>.
- [20] 小学館. 中日・日中統合辞典. 小学館, 1998.
- [21] 河原大輔, 黒橋禎夫. 格フレーム辞書の漸次的自動構築. 自然言語処理, vol. 12, no. 2. pp. 109-131, (2005). <https://ci.nii.ac.jp/naid/10015447374/>.
- [22] “Japanese Dependency and Case Structure Analyzer KNP”.