

# 比喩生成を応用したキャッチコピー作成支援システムの提案 Support System for Catchphrase Generation Applying Metaphor Generation

呉 航平<sup>†</sup>, 寺井 あすか<sup>†</sup>

Kohei Kure, Asuka Terai

<sup>†</sup> 公立ほこだて未来大学

Future University Hakodate

g2121019@fun.ac.jp

## 概要

キャッチコピーにおいてよく使われている修辞技法として比喩があることが指摘されている。本研究ではキャッチコピーにおける比喩に着目し、キャッチコピーにおいて対象となる語を被喩辞、キャッチコピーにおいて強調したい特徴を特徴語とみなすことで、比喩生成を応用したキャッチコピー作成支援システムを提案した。また、先行システムと提案システムの比較を通じ、キャッチコピー生成における比喩生成機構の有効性を示した。

**キーワード：**キャッチコピー生成, 比喩生成

## 1. はじめに

近年、SNSの普及やネット広告等の増加により、世の中の情報量が急激に増えている。その中で、一言で商品の特長や伝えたいことを表現するキャッチコピーの需要は高まっており、書籍の表紙やCM等の日常のさまざまな場面で使用されている。キャッチコピーの作成は、強調したい特徴を定めたうえで様々なアイデアを出し、それを文章にまとめ上げる活動であり、創造的思考を要する。

創造的思考の基礎要素として概念結合があり、2つの概念の融合による創造的思考を含む言語表現として比喩が存在する。比喩生成の計算論的アプローチを試みた研究としては北田ら [1] や中条ら [2] のものがあり、これらの研究では被喩辞（喩えられる語）と特徴語（喩えられる語が持つ性質や特徴）から喩辞（喩える語）の推定を行う。

また、キャッチコピーにおいてよく使われている修辞技法として比喩があることが指摘されており [3]、比喩表現と意味の重ね合わせを用いたキャッチコピー生成システムの構築が行われている [4]。しかし、このシステムはキャッチコピーテンプレートにおける動詞を比喩表現に変更するものであり、キャッチコピーにおいて強調したい特徴について考慮されていない。

そこで、本研究ではキャッチコピーにおける比喩に着目する。キャッチコピーにおいて対象となる語を被喩辞、キャッチコピーにおいて強調したい特徴を特徴語とみなし、比喩生成を応用したキャッチコピー作成支援システムを提案する。

本研究では、はじめに、連想に基づくキャッチコピー生成の先行システム [5] の連想部を比喩生成モデルに置き換えることで、比喩生成に基づくキャッチコピー作成支援システムとしてキャッチコピーで使用する語句を提案可能なシステムを構築する。次に、提案システムと先行システムとの比較を通じて、キャッチコピー生成に対する比喩生成の影響を検討する。

比喩生成に基づく提案システムと連想に基づく先行システムの最も明確な違いとして比喩生成に基づくシステムでは喩辞の評価 [6] を行う。

## 2. 比喩生成に基づく提案システム

比喩生成に基づく提案システム（以下「提案システム」という。）の概要を図1に示す。ここで、関連語はキャッチコピーにおける生成対象に最も関連が高く、キャッチコピーの主語、特徴語は関連語が持つ特徴、喩辞は対象語を喩える語である。

システムへの入力には映画のあらすじであり、出力は関連語とランキングされた喩辞のリストである。システムの出力は「（関連語）は（喩辞）だ」「（喩辞）の（関連語）」「（喩辞）のような（関連語）」といったキャッチコピーとして使用することを想定している。

関連語と喩辞はそれぞれ名詞1語とし、特徴語は動詞・形容詞・形状詞に加えて、以下の普通名詞に限定する。

- サ変可能, 形状詞可能, サ変形状詞可能

システムの中身は大きく分けて、前処理、キーワード抽出、喩辞生成の3つのステップがある。まず前処理では、ユーザが入力した映画のあらすじを取得し、あらすじの内容語集合を取得する。次にキーワード抽

出では、関連語候補と特徴語候補 5 語をユーザに提示する。ユーザはそこからそれぞれ一語を選択し、それらを関連語、特徴語とする。最後に喩辞生成では、抽出した関連語と特徴語から喩辞候補を生成し、喩辞の評価に基づくランキングを行う。

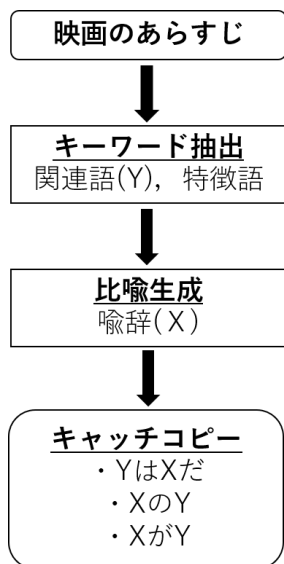


図 1 提案システムの概要

## 2.1 前処理

入力された映画のあらすじをキネマ旬報社が運営する KINENOTE[7] から取得した。このあらすじを形態素解析することで意味のある品詞のみを取り出し、あらすじの内容語集合とする。

この時、あらすじには丸括弧の中に俳優名が示されている場合があるため、その場合は丸括弧ごと俳優名を除去する。また、310 件の日本語ストップワードが記載されている Slothlib[8] を用いて、ストップワードを除去する。

## 2.2 キーワード抽出

以下のスコア  $r$  を利用し、内容語集合から関連語候補と特徴語候補の抽出を行う。

### あらすじとの類似度と TF-IDF 値の乗算値 (スコア $r$ )

内容語集合内の各単語に対して、あらすじの文章ベクトルとのコサイン類似度と TF-IDF 値を掛け合わせた値を算出する。あらすじの文章ベクトルは、文章内で使用されている内容語集合の単語ベクトルの平均と

する。また、TF-IDF の算出における対象文章は、対象作品と公開年が前後 5 年以内の 1000 作品とし、ランダムに選択する。

### 2.2.1 関連語の抽出

関連語候補の抽出手順は以下のとおりである。

1. スコア  $r$  が高い 5 語を関連語候補としてユーザに提示する。
2. ユーザが 1 つ選び、それを関連語とする。

### 2.2.2 特徴語の抽出

特徴語候補の抽出手順は以下のとおりである。

1. 関連語を含む文を係り受け解析し、関連語にかかっている語と関連語がかかっている語を  $x$  語抽出する。
2. スコア  $r$  が高い  $(5 - x)$  語を抽出する。
3. 1・2 で抽出した特徴語候補 5 語をユーザに提示する。
4. ユーザが特に強調したい特徴を 1 つ選び、それを特徴語とする。

## 2.3 喩辞生成

比喩生成モデル [2][6] に基づき、喩辞候補の生成と評価の 2 ステップにより喩辞を生成する。

### 2.3.1 喩辞候補の生成

まず、以下の 3 種類のフィルタリングを行い、残ったものを喩辞候補とする。

- (i) 品詞細分類  
IPA 辞書における一般名詞のうち、「サ変接続・形容動詞語幹・副詞可能・被自立」を喩辞候補から除外する。
- (ii) 関連語のカテゴリ  
分類語彙表 [9] を使用して、関連語と同じカテゴリに属する語を喩辞候補から除外する。
- (iii) 意味空間  
意味ベクトルを持たない語を喩辞候補から除外する。

次に、以下の 4 つのスコア ( $v_1, v_2, v_3, v_4$ ) の線形和を喩辞スコア ( $v$ ) とし、喩辞スコアが高い上位 30 語を喩辞候補とする。各スコアに対する重み

を  $w_1, w_2, w_3, w_4$  とした時, 喩辞スコアの計算式は式 1 で表される.

$$v = v_1 w_1 + v_2 w_2 + v_3 w_3 - v_4 w_4 \quad (1)$$

### 関連語・特徴語の合成ベクトルと喩辞候補ベクトルのコサイン類似度 (スコア $v_1$ )

合成ベクトルの算出には, predication algorithm[10] を使用する. 関連語・特徴語の合成ベクトルを取得する手順は以下の通りである.

1.  $j_1$  個の概念を含む特徴語の近傍を特徴語ベクトルとの類似度に基づき推定する.
2. 特徴語の近傍に含まれ, 関連語ベクトルと類似度が高い語を  $k$  個抽出する.
3. 関連語・特徴語・ $k$  個の語のセントロイドを計算し, 合成ベクトルとする.

### 喩辞候補の具象度 (スコア $v_2$ )

単語心像性の辞書 [11] を利用して, 抽象度の低い語に対して高いスコアを与える. 心像性とは, 単語の想起のしやすさを示す指標であり, 心像性を 0 から 1 の範囲に正規化したものを具象度として定義する. 具象度は式 2 で表される.

$$\text{具象度} = \begin{cases} (\text{心像性} - 1)/6 & (\text{心像性} \leq 4.5) \\ (4.5 - 1)/6 & (\text{上記以外}) \end{cases} \quad (2)$$

### 文脈ベクトルと喩辞候補ベクトルのコサイン類似度 (スコア $v_3$ )

文脈ベクトルは, 関連語の直前  $m$  個と直後  $n$  個の内容語の重心ベクトルとする. ただし, 内容語を取得する際に関連語や特徴語と一致する語は例外としてスキップする.

### 関連語ベクトルと喩辞候補ベクトルのコサイン類似度 (スコア $v_4$ )

このスコアを喩辞スコアから減算することで, 関連語と似たような意味を持つ喩辞が生成されることを防ぐ.

#### 2.3.2 喩辞候補の評価

喩辞候補 30 語に対して, 比喩表現として入力された特徴をどの程度表現できているかどうかを以下のスコア  $v_5$  を用いて,  $v_5$  の値が高い単語が喩辞としてより適切であると評価する.

### 対象語・特徴語の合成ベクトルと対象語・喩辞の合成ベクトルのコサイン類似度 (スコア $v_5$ )

それぞれの合成ベクトルの算出には, predication algorithm[10] を使用する. 関連語・特徴語の合成ベクトルを取得する手順は, スコア  $v_1$  と同様である. 関連語・喩辞の合成ベクトルを取得する手順は以下の通りである.

1.  $j_2$  個の概念を含む喩辞の近傍を喩辞ベクトルとの類似度に基づき推定する.
2. 喩辞の近傍に含まれ, 関連語ベクトルと類似度が高い語を  $k$  個抽出する.
3. 関連語・喩辞・ $k$  個の語のセントロイドを計算し, 合成ベクトルとする.

### 3. 連想に基づくシステム

先行システム [5] への入力の対象を説明する文章であり, キャッチコピーの作成過程では, 対象語・関連語・共通概念 (特徴語)・連想語の 4 つの語が関与する. 先行システムの語の抽出構造を図 2 に示す.

まず, 説明文を近似した「対象語」, 対象語と類似度が高い「関連語」を抽出し, 強調したい特徴を表す「共通概念」を選択する. 次に, 対象語・関連語との類似度が低く, 共通概念との類似度が高い「連想語」を抽出する. 最後に, 適切な副詞や助詞を補うことで, 関連語と連想語からなるキャッチコピーを生成する.

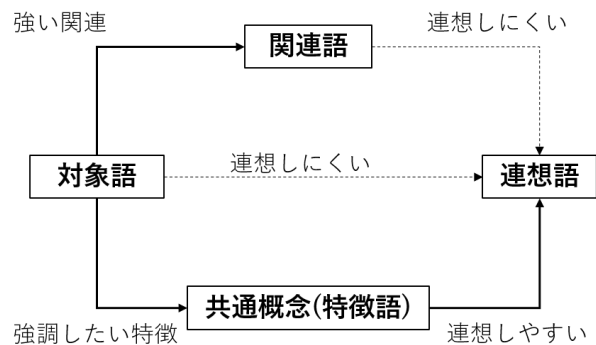


図 2 先行システム [5] での語句の抽出構造

先行システムに基づき, 連想に基づくシステムを設計した. 手順は以下の通りである. まず, 提案システムと同様に関連語, 特徴語の抽出を行う. 次に, 以下の 4 つのスコア ( $a_1, a_2, a_3, a_4$ ) の線形和を連想語スコア ( $a$ ) とし, 連想語をランキングする.

- (i) 特徴語ベクトルと連想語候補ベクトルのコサイン類似度 (スコア  $a_1$ )
- (ii) 連想語候補の具象度 (スコア  $a_2$ )

- (iii) 対象語ベクトルと連想語候補ベクトルのコサイン類似度 (スコア  $a_3$ )
- (iv) 関連語ベクトルと連想語候補ベクトルのコサイン類似度 (スコア  $a_4$ )

それぞれのスコアに対する重みは提案システムに基づき設定する。連想語スコア ( $a$ ) の計算式は式 3 で表される。

$$a = a_1w_1 + a_2w_2 - (a_3 + a_4)w_4/2 \quad (3)$$

#### 4. システムの評価

各システムの入力に対して、正解リストを作成し、正解リストに基づく評価値を定義することで、適切なパラメータの推定を行った。

また、決定したパラメータにおける各システムの評価値を比較することで、システムの評価を実施した。使用した映画は合計 6 作品である。最終的に決定したパラメータにおけるシミュレーション結果を表 1 に示す。表 1 は「四月は君の嘘」という作品を対象として、関連語として「ピアノ」、特徴語として「自由」を選択した際の各システムにおける出力である。

表 1 各システムの入力上位 10 件

	提案システム	連想に基づくシステム
1	マイペース	自分
2	遊び場	個人
3	余暇	決まり
4	一人っ子	言論
5	一人	代わり
6	自然体	好み
7	人生	都合
8	天真爛漫	良心
9	僕	スタイル
10	庭	空間

#### 4.1 正解リストの作成

各システムのパラメータの全ての組み合わせにおける出力上位 30 語に対して、適切性を 5 段階で評価した。評価者は 3 名であり、2 名以上の評価者が 4 以上の評価を与えたものを正解とする。これを 6 作品すべてで行い、各作品に対する正解リストを作成した。各作品の正解リストの語数を表 2 に示す。

表 2 各作品の正解リストの語数

	語数
作品 1	45
作品 2	32
作品 3	34
作品 4	38
作品 5	14
作品 6	28

#### 4.2 評価値の算出

システムの入力上位 30 語に対して、正解リストに含まれる語の順位の逆数の和を計算し、そのパラメータにおける評価値とする。

#### 4.3 パラメータ推定

映画 6 作品の入力を用いて比喩生成に基づくシステムの適切なパラメータを推定した。パラメータの全ての組み合わせに対して評価値を計算し、6 作品の評価値の平均が最も高いものを適切なパラメータとした。提案システムのパラメータと値の選択肢を表 3 に示す。最終的に適切なパラメータとして決定したものを太字で示している。

表 3 提案システムのパラメータと値の選択肢

パラメータ	値の選択肢
j1	10, <b>50</b> , 100
j2	100, 300, <b>500</b>
k	3, 5, <b>10</b>
m	<b>3</b>
n	<b>3</b>
w1	<b>1</b>
w2	<b>1</b>
w3	<b>0.4</b>
w4	<b>0.1</b> , 0.2, 0.4

#### 4.4 結果

決定したパラメータにおける各システムの評価値を比較した。各作品における各システムの評価値を表 4 に示す。作品 5 を除く 5 作品の評価値が提案システムの方が高い結果となった。

表4 各システムの評価値

	提案システム	連想に基づくシステム
作品1	3.10	2.04
作品2	2.82	2.45
作品3	2.92	2.51
作品4	3.25	3.20
作品5	1.03	1.52
作品6	3.16	2.54
平均値	2.71	2.38

## 5. まとめ

本研究では、キャッチコピーにおける比喩に着目し、比喩生成を応用したキャッチコピー作成支援システムを提案した。また、連想に基づくシステムとの比較を通じ提案システムの妥当性を示すことで、キャッチコピー生成における比喩生成機構の有効性を示した。

しかし、本研究のシステムの出力は、対象語と特徴語、喩辞といったキャッチコピーのキーワードのみである。今後は、そこから適切な助詞を選択してキャッチコピーを生成するシステムの構築を考えている。キャッチコピーのパターンとしては、本稿で示した「(対象語)は(喩辞)だ」、「(喩辞)の(対象語)」、「(喩辞)のような(対象語)」の他にも検討が必要である。

## 文献

- [1] 北田純弥, 萩原将文, (2001) “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.
- [2] 中條寛也, 松吉俊, 内海彰, (2017) “意味空間に基づく文脈情報をを用いた比喩生成” 研究報告音声言語情報処理, Vol. 2017, No. 14, pp. 1-10.
- [3] 丹羽彩奈, 岡崎直観, 西口佳佑, 亀山千尋, 毛利真崇, (2019) “キャッチコピーの自動生成に向けた分析” 言語処理学会第25回年次大会発表論文集, pp.558-561.
- [4] 梅村奏子, 狩野芳伸, (2021) “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” SIG-SLUD, Vol.5, No. 03, pp. 24-30.
- [5] 寺井あすか, 中川正宣, (2011) “言語統計解析に基づく評価メカニズムを含む比喩生成モデルの構築” 日本認知科学会第28回大会発表論文集, pp. 388-389.
- [6] 中野俊亮, 鬼沢武久, (2008) “ユーザ対話による意外性を持つキャッチフレーズ作成支援” 情報処理学会第70回全国大会講演論文集, pp. 201-202.
- [7] KINENOTE  
<http://www.kinenote.com/main/public/home/>
- [8] 大島裕明, 中村聡史, 田中克己, (2007) “SlothLib:Webサーチ研究のためのプログラミングライブラリ” 日本データベース学会, Vol. 6, No. 1, pp. 113-116.
- [9] 分類語彙表  
[https://pj.ninjal.ac.jp/corpus\\_center/goihyo.html/](https://pj.ninjal.ac.jp/corpus_center/goihyo.html/)

[10] Kintsch, W, (2000) “Metaphor comprehension: A computational theor” Psychonomic Bulletin & Review, Vol. 7, No.2, pp. 257- 266.

[11] 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳格, 田中正之, 天野成昭, 近藤公久, (1999) “NTT データベースシリーズ日本語の語彙特性第3期 第8巻”, 三省堂.