

# あらすじとレビューを用いた

## 映画のキャッチコピー作成支援手法の提案

幅野裕貴 浦谷則好

東京工芸大学大学院工学研究科 電子情報工学専攻

### 1. はじめに

キャッチコピーは宣伝を目的とした「人を引き付ける言葉」であり、対象となるものの魅力や価値を短文でかつインパクトのある言葉で相手を惹きつけるものである。しかし、読み手の心を動かすようなキャッチコピーの作成は専門のコピーライターでなければ困難である。商品やサービスのキャッチコピー作成には「売れるキャッチコピーがスラスラ書ける本」[1]等が存在するが、小説や映画にはマニュアルとなる書籍はほとんどない。そこで本研究では映画に対するキャッチコピーの作成を支援する手法を提案する。

映画のあらすじとレビューを用いて、恋愛、アクション、ホラー映画の3つの分野に対してキャッチコピーを生成したのでその結果を報告する。

### 2. 関連研究

キャッチコピーの自動生成を行うシステムとしては、山根らのWeb情報を用いたキャッチコピー自動生成システムがある[2]。キャッチコピーにおいて述べる対象(例:景色)と、性質(例:美しい)を入力し、これをクエリとし、Webから文章の抽出を行う。そしてLDAを利用して潜在的トピックを抽出し、キャッチコピー生成にふさわしい単語を文章から獲得する。次にキャッチコピー候補生成にふさわしい分野の決定を行い、選択された分野におけるモデルとなるキャッチコピーを選択する。選択されたキャッチコピーの文法構造を利用し一部の単語をWeb上の情景描写していると考えられる単語と置き換え、キャッチコピー候補を生成している。

中野らは、コピーライターの役割を模倣した対話型の作成支援システムを提案している[3]。ユーザが対象(例:カルピス)を説明する文章を入力し、説明文に含まれる対象語句を取得する。次に対象語句と類似度が高い概念語を関連語句(例:味)として、自動的に選択する。ユーザも強調したい特徴を表す概念語を説明文中から選び、共通概念(例:甘酸っぱい)とする。そして対象語句と関連語句と類似度が低くかつ、共通概念と類似度が低い概念語を連想語句(例:初恋)として概念ベースから取り出す。最後に、関連語句と連想語句を「の」、「が」などの助詞で接続し文を生成する。

### 3. 提案手法

本提案手法におけるキャッチコピー作成の手順を図1に示す。まずユーザはキャッチコピーを作成したい映画名を入力する。システムは映画のあらすじをWeb上から取得し、あらすじの単語を検索キーワードとしてキャッチコピーDBからモデルとなるキャッチコピーを複数個選択する。次に映画の感想文(レビュー)をレビューDBから取得し、キャッチコピー作成時に使用する単語の候補を抽出する。この単語の意味素性と抽出されたモデルとなるキャッチコピーに含まれている単語の意味素性が同じものを探して単語を入れ替え、キャッチコピーの生成を行う。以下の節でそれぞれの処理を詳述する。

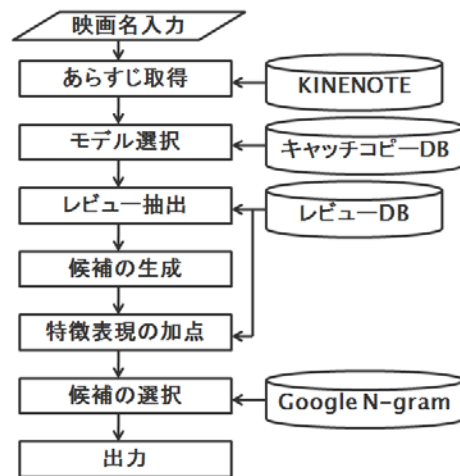


図1 作成手順

#### 3.1 あらすじの取得

図1のように入力された映画名のあらすじをキネマ旬報社が運営する KINENOTE[4]から獲得する。このあらすじを Mecab[5]で形態素解析し、品詞が名詞、動詞、形容詞の単語を抽出し、モデルキャッチコピーを選択する時の検索キーワードとする。

#### 3.2 モデル選択

前稿[6]では、レビューから取得したキーワードを用いてモデル選択を行っていたが、抽出してくるモデルの数が多くなりすぎることから、今回はあらすじから取得したキーワードを用いてキャッチコピ

一DBを検索することにした。検索対象となるモデルキャッチコピーは、キャッチコピーに含まれる自立語の数が11語以内のキャッチコピーとする。キャッチコピーはallcinema[7]から収集したものである。このDBにはキャッチコピーそのもの以外にキャッチコピーを構成している単語に対する意味素性が付加されている。この意味素性はキャッチコピー中の名詞、形容詞の単語に対しては日本語語彙大系[8]を用いて付加を行い、動詞の単語に対しては分類語彙表[9]を用いて付加した。

### 3.3 レビュー抽出

レビューをレビューDBから抽出する。これらのレビューはユーザがレビューを投稿することができるサイトYahoo映画[10]から取得したものである。DBからその映画に対するレビューを抽出した後、形態素解析器Mecabを用いて名詞、動詞、形容詞の単語を抽出する。3.2と同様、抽出した名詞、形容詞の単語に日本語語彙体系を用いて意味素性を付与し、動詞の単語に対しては分類語彙表を用いて意味素性の付加を行う。

### 3.4 候補の生成

3.3で獲得した意味素性と3.2で選択されたキャッチコピーモデルに含まれる単語の意味素性のマッチングを取る。マッチした意味素性が含まれている単語同士で置換を行いキャッチコピー候補を生成する。生成時に置換対象となる単語と3.3で抽出した単語が同じ単語であった場合、この単語も置き換え対象とする。なお、生成パターンとして置換だけではなく、キャッチコピーの置き換え候補となる単語をそのまま利用した生成も行う。

生成したキャッチコピーに対して、独創性の観点からスコアを以下の(1)式で算出する。Aはキャッチコピーに含まれる自立語の数である。Bは置き換え候補となる単語の数である。前稿ではBを2乗していたが、作品のイメージに沿わない長いキャッチコピーが多く含まれてしまうことが多いため、バランスを調整し、1.25乗に変更した。Cは置換単語の評価に関する項である。置き換え候補となる単語と同じ単語がレビューに含まれていて、この単語が置換された場合 $+\alpha$ とし、評価を上げる。置換する単語が別の単語の場合は $-\beta$ とする。キャッチコピーに含まれる単語をそのまま使用した場合は $-3\beta$ とする。実験では $\alpha, \beta$ とも0.5を用いた。以下の表1に「ハンガーゲーム」という作品で実験した結果を一部示す。

$$O = \frac{B^{1.25} + \sum C}{A} \quad (1)$$

表1 「ハンガーゲーム」の結果の一部

スコア(順位)	生成したキャッチコピー
3.0953 (1位)	生き残れるか!?それは、史上最も..
2.9827 (2位)	未来と戦え!生き残るために。
2.9320 (3位)	自分だけ、知らない

### 3.5 特徴表現の加算

$$\text{tfidf} = \text{tf} \times \log\left(\frac{D+1}{N}\right) \quad (2)$$

(1)式でスコア付けされたキャッチコピーを対象に、キャッチコピーに含まれる置き換え候補の単語毎のペアを抽出する。tfはここでは単語ではなくペアの数を示している。(2)式は単語ペアの妥当性を計るものである。生成されたキャッチコピー中から抽出した置き換え単語のペアが、レビュー内に含まれているか検索し、見つかった場合は(1)式で求めたOに対して(2)式で求めた値を加算し、マッチしない場合は-1とする。tfは入力された映画の全てのレビュー中から抽出した各単語のペアの出現頻度である。レビュー内からペアを抽出する範囲は、一文中に限定し、かつ一つの単語に対して抽出するペアの対象となる単語は自立語で5語以内の距離のものとする。DはキャッチコピーDB中の映画の各分野毎の作品の総数である。単語のペアのマッチングを取るときにカウント数が0になるのを防ぐため、Dには入力された映画の作品を含むものとする。Nは入力された映画のレビューから抽出した単語のペアを含んでいる作品の数である。これにより評価は(3)式に変更される。表1のそれぞれのスコアが、特徴表現の加算により変動した結果を表2に示す。

$$K = O + \sum \text{tfidf} \quad (3)$$

表2 変動したスコアの結果

スコア(順位)	生成したキャッチコピー
14.9858 (1位)	生き残れるか!?それは、史上最も..
4.3115 (165位)	未来と戦え!生き残るために。
3.3959 (221位)	自分だけ、知らない

### 3.6 候補の選択

文としての自然性はGoogle N-gram コーパス[11]を用いて次式で評価する。

$$P = \frac{\sum \log T}{n} \quad (4)$$

(4)式のnは生成されたキャッチコピーにおける単語の数である。TはGoogle N-gram コーパスに含まれる3グラムの出現頻度である。(4)式に重みwを乗じて(3)式と加算した(5)式を最終的な候補のスコアとする。実験ではw=0.5とした。以下の表3に表2から変動した結果を示す。

$$S = K + wP \quad (5)$$

表3 最終結果

スコア(順位)	生成したキャッチコピー
15.6593 (1位)	生き残れるか!?それは、史上最も..
7.0563 (31位)	自分だけ、知らない
3.2310 (375位)	未来と戦え!生き残るために。

#### 4. 実験

恋愛, アクション, ホラー映画のそれぞれ7作品, 計21作品に対して生成を行い, 提案手法の良否を検証した. 対象となるキャッチコピーは(5)式で求めたスコアの上位100位以内に含まれるキャッチコピーとする. その中から著者が映画のイメージに沿うと思われるキャッチコピーを選択し, 中野らの評価方法を参考に, 生成したキャッチコピーが「作品のイメージに沿ったキャッチコピーであるか」を,

1. とてもあてはまる(+2)
2. 少しあてはまる(+1)
3. なんともいえない(0)
4. あまりあてはまらない(-1)
5. 全くあてはまらない(-2)

の五段階評価で行い, 平均で0.5以上ならば良しとする. この評価を10人に対して行った. 恋愛, アクション, ホラー映画の評価結果をそれぞれ表4, 5, 6に示す. 評価は視聴した作品に限ったため人数にバラつきがある. なお, 表に記載した結果は1作品につき最低2人以上が評価したキャッチコピーである. また, 表7に恋愛映画, 表8にアクション映画, 表9にホラー映画の各作品に対して生成したキャッチコピーとオリジナルのキャッチコピーを対比した結果を示す.

表4 恋愛映画の評価結果

作品ID	生成したキャッチコピー	評価
L1	分からない目であなたを見つめていた	(1):1人 (2):1人
L2	そして, 私に残ったのは愛だけでした	(1):1人 (4):1人 (2):2人
L3	流れた時間が動いた時, そこが私の発火点	(2):3人 (3):3人
L5	深い深い海の底でも, 消せない愛。	(1):7人 (3):1人 (2):1人
L6	あなたに伝えたい, 本当の想い。	(1):2人 (2):1人
L7	恋をすれば男もキレイになる。	(1):2人 (3):1人 (2):5人 (5):1人

表5 アクション映画の評価結果

作品ID	生成したキャッチコピー	評価
A1	もはや, 敵なし。最強。	(1):2人
A2	結局兵器, それは武器	(1):2人 (4):2人 (2):2人
A3	恋の終わりは, 戦いの始まり。	(2):4人 (4):1人 (3):2人 (5):1人
A4	生き抜くことが, 彼のたったひとつの剣。	(2):2人 (4):3人 (3):1人
A6	生き残れるか!?それは, 史上最も過酷なサバイバルゲーム。	(1):3人

表6 ホラー映画の評価結果

作品ID	生成したキャッチコピー	評価
H1	さあ, ゲームをはじめよう	(1):4人
H2	その部屋に, 入ってはいけない	(2):2人
H4	誰も生きては戻れない	(1):2人 (2):2人
H6	この娘, どこかが変だ	(1):1人 (2):3人

評価を得られた作品のなかで「スパイダーマン(A3)」, 「るろうに剣心(A4)」を除いて, 平均値が0.5以上を越える結果となり, 作品のイメージに沿うキャッチコピーを生成できたことが示せた. 特に「タイタニック(L5)」, 「ハンガーゲーム(A6)」, 「ソウ(H1)」が高評価を得られた. 「タイタニック」と「ハンガーゲーム」のキャッチコピーは(1)式で求めたスコアが, (3)式によってスコアが大きく変動したため, トップの点数となったキャッチコピーである. 「ソウ」は特徴表現の加点ではスコアが僅かに変化し, 自然性の評価で大きく点数を上げていた.

一方, 特徴表現の加点の段階でスコアが大きく変動したキャッチコピーが, 自然性の評価により, ノイズとして扱われ, キャッチコピー候補から除外される作品もあった. 傾向として, 長い文のキャッチコピーや意味が伝わりにくいキャッチコピーが落とされる傾向にある. 今回, 生成したキャッチコピーも全体的に文字数が少ないキャッチコピーが多く, 評価の結果からも, 短い文の方が, 作品のイメージに沿ったキャッチコピーが生成できると考えられる.

今回, 映画のレビューから置き換え単語の抽出を行ったことにより, 置換できる単語の種類も増え, 作品を表現している単語が得られたことも, 作品の内容に沿ったキャッチコピーが生成できた要因だと考えられる.

モデルとなるキャッチコピーも様々な単語で構成されているので, 全ての単語と対応を取るためには置き換え単語の数も必要となってくる. そのため1作品で最低でも15000文字以上のレビューが得られることが望ましい. また, 生成されるキャッチコピーのバリエーションを増やすためにも, なるべく多くモデルとなるキャッチコピーを抽出できた方がよいので, 600文字以上で構成されたあらすじが必要であることが分かった.

問題点として, 自然性の評価でノイズとなるキャッチコピーを除外しても, スコアの上位の中に意味の伝わりにくい文章や作品のイメージに沿わないキャッチコピーが含まれていることがある. 原因として, 置き換え箇所が少ないキャッチコピーは, 特徴表現の加点を(3)式で行う際, 加点の対象となるペアが少ないことが考えられる. 意味の通らない文章であれば, 自然性の評価で除外されるはずだが, 「手を飛びたら,..」といった一般的には使われない文章が含まれていても, (5)式で除外されなかったり, 点数が更に加算されるケースもあった. これは, n-gramコーパスをWebから収集していることが原因だと考えられる.

#### 5. まとめ

本稿では映画のキャッチコピーの作成を支援する手法を提案し, 恋愛, アクション, ホラー映画のキャッチコピーの生成を行った. また, 生成したキャッチコピーに対しての評価を行い, 結果を検証した. 評価を得られた作品は, 2作品を除きイメージに沿ったものが生成できたと考えられる.

課題として, 作品のイメージに沿ったキャッチコピーが候補から除外される場合があるので, 作品の特徴を表現していると思われる単語のペアに対して評点を上げる方法の考案やn-gramに含まれる単語に対する評価が必要であると考えられる.

## 参考文献

- [1] 酒井とし夫: “売れるキャッチコピーがスラスラ書ける本”, 日本能率マネジメントセンター, 2007  
 [2] 山根宏彰, 荻原将文: Web 情報を用いたキャッチコピー自動生成システム. ファジィシステムシンポジウム講演論文集 28, pp. 1008-1013, 2012  
 [3] 中野俊亮, 鬼沢武久: ユーザー対話による意外性を持つキャッチフレーズ作成支援. 全国大会講演論文集, pp. 2-201-2-202. 情報処理学会, 2008  
 [4] KINENOTE - 映画ライフログサービス  
<http://www.kinenote.com/main/public/home/>  
 [5] Mecab  
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

- [6] 幅野裕貴, 浦谷則好: 映画のキャッチコピー作成支援手法. 言語処理学会第 19 回年次大会発表論文集, pp. 890-892, 2013  
 [7] 映画データベース - allcinema  
<http://www.allcinema.net/prog/index2.php>  
 [8] NTT コミュニケーション科学 基礎研究所: “日本語語彙大系”, 岩波書店, 1999  
 [9] 国立国語研究所: “分類語彙表-増補改訂版-”, 大日本図書, 2004  
 [10] Yahoo 映画: <http://movies.yahoo.co.jp/>  
 [11] Google 日本: 大規模日本語 n-gram データの公開  
<http://googlejapan.blogspot.com/2007/11/n-gram.html>

表 7 恋愛映画の比較表

作品名(ID)	生成したキャッチコピー	オリジナルのキャッチコピー
私の頭の中の消しゴム(L1)	分からない目であなを見つめていた	死より切ない別れがある。
ロミオ&ジュリエット(L2)	そして、私に残ったのは愛だけでした	最も偉大な永遠の愛の物語。
時をかける少女(L3) (アニメ版)	流れた時間が動いた時、そこが私の発火点	待つてられない未来がある
猟奇的な彼女(L4)	誰もが彼女に心取られる。	好きってイタイ!?
タイタニック(L5)	その時、運命に弄ばれた男と女。	運命の恋。誰もそれを裂くことはできない
君に届け(L6)	あなたに伝えたい、本当の想い。	本当に大切な想いは、ゆっくりと伝えればいい。
電車男(L7)	恋をすれば男もキレイになる。	A True Love Story 美女とヲタクの純情初恋物語

表 8 アクション映画の比較表

作品名(ID)	生成したキャッチコピー	オリジナルのキャッチコピー
エクスペンダブルズ(A1)	もはや、敵なし。最強。	最強無敵
アイアンマン(A2)	結局兵器、それは武器	装着せよ 強き自分
スパイダーマン(A3)	恋の終わりは、戦いの始まり。	Who is Spider-man?
るろうに剣心(A4)	生き抜くことが、彼のたったひとつの剣。	かならず、帰る。
ダークナイト ライジング(A5)	世界を変えるのは、破壊か、祈りか。	伝説が、壮絶に、終わる。
ハンガーゲーム(A6)	生き残れるか!? それは、史上最も過酷なサバイバルゲーム。	生存率 1/24 それは究極のサバイバル。戦うことでしか、生き残れないのか——?
特攻野郎 A チーム(A7) THE MOVIE	自分だけ、知らない	デカイことヤローゼ

表 9 ホラー映画の比較表

作品名(ID)	生成したキャッチコピー	オリジナルのキャッチコピー
ソウ(H1)	さあ、ゲームをはじめよう	Every piece has a puzzle.
クロユリ団地(H2)	その部屋に、入ってはいけない	誰から、死ぬ?
デッド・サイレンス(H3)	叫びだしたら、止まらない	叫んだら、死ぬ!?
サイレントヒル (H4)	誰も生きては戻れない	その街では、祈りさえも、悲鳴に変わる
13日の金曜日(H5)	音は、悪夢への招待状	なぜ全米で失神者が続出したのか!? 若者たちにとって楽しい夏休みキャンプのはずだった..
エクソシスト(H6)	この娘、どこかが変だ	この家の少女に想像を絶する何か起きている! すべてを託された男..
エルム街の悪夢(H7)	死にたくなければ、目をあけて眠れ。	目を覚ませナンシー! 眠ると奴がやってくる!