学位論文 博士(工学)

キャッチコピーの自動生成に関する研究

2014 年度

慶應義塾大学大学院 理工学研究科

山根 宏彰

目次

論文要旨	á	1
第1章	序論	2
1.1	はじめに	2
1.2	人工知能	3
	1.2.1 人工知能の歴史	3
	1.2.2 人工知能と創造性	6
1.3	ことば工学	6
	1.3.1 ことば工学の認知・心理学的背景	7
	1.3.2 ことば工学の研究例	8
	1.3.3 期待される応用分野	9
1.4	自然言語処理	9
	1.4.1 コーパス 1	10
	1.4.2 統計的アプローチ	11
	1.4.3 テキストマイニング 1	12
	1.4.4 文生成	12
	1.4.5 ことば工学と自然言語処理	12
1.5	キャッチコピー 1	13
	1.5.1 キャッチコピーとその目的	13
	1.5.2 キャッチコピーの変遷と社会的背景	14
	1.5.3 キャッチコピーに関するシステム 1	15
1.6	本研究の位置づけ	17
1.7	本論文の構成 1	19
第2章		20
2.1		20
	2.1.1 キャッチコピーコーパスと比較対象とするコーパス 2	20
	2.1.2 コーパスの前処理と解析方法 2	21
2.2	他コーパスとの比較による特徴分析 2	22

	2.2.1	全体の品詞使用割合の分析	22
	2.2.2	品詞列に着目した分析	25
2.3	キャツ	チコピーコーパスにおける特徴	29
	2.3.1	キャッチコピーコーパスにおける単語	29
	2.3.2	キャッチコピーコーパスにおける係り受け関係	30
	2.3.3	キャッチコピーコーパスにおける分野ごとの特徴	30
2.4	統計的]手法を用いたキャッチコピー自動生成システム	31
	2.4.1	分野特有の関連語取得 (I.)	34
	2.4.2	キャッチコピー候補生成 (II.)	37
	2.4.3	キャッチコピー候補選択 (III.)	37
2.5	実験		40
	2.5.1	実験概要....................................	40
	2.5.2	実験条件	40
	2.5.3	生成例	41
	2.5.4	実験結果....................................	41
2.6	考察		43
	2.6.1	それぞれのテーマ	43
	2.6.2	全テーマ	44
2.7	まとめ		44
第3章	Web	上の文情報を用いたキャッチコピー自動生成システム	46
3.1			46
3.2			
3.3		この文情報を用いたキャッチコピー自動生成システム	49
0.0	3.3.1	- 提案システムの流れ	
	3.3.2	知識獲得	50
	3.3.3	候補生成	52
	3.3.4	·····································	55
	3.3.5	文の適切度による選択	56
3.4	実験		57
	3.4.1	実験条件....................................	57
	3.4.2	生成例	59
	3.4.3	実験結果・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	60
3.5	考察		61
3.6			63

第4章	SNS 上の嗜好情報を用いたキャッチコピー自動生成システム	65
4.1	背景	65
4.2	Facebook ファンページ	67
4.3	SNS 上の嗜好情報を用いたキャッチコピー自動生成システム	67
	4.3.1 知識取得	68
	4.3.2 候補生成	68
	4.3.3 キャッチコピー候補生成	70
	4.3.4 候補選択	70
4.4	評価実験	74
	4.4.1 実験概要	74
	4.4.2 実験条件	75
	4.4.3 実験結果	77
4.5	まとめ	80
第5章	結論	81
5.1	まとめ	81
5.2	今後の課題	83
	5.2.1 評価に関して	83
	5.2.2 関連する発展的なテーマ	83
謝辞		85
参考文献	就 数	87
付録		103
付録Ai	章 統計的手法を用いたキャッチコピー自動生成システム	103
A.1	パラメータ	103
	A.1.1 分野特有の関連語取得	103
	A.1.2 キャッチコピー候補生成	105
	A.1.3 キャッチコピー候補選択	105
A.2	評価実験	106
	A.2.1 評価実験に用いたキャッチコピー	106
	A.2.2 失敗例	109
付録Bi	章 Web 上の文情報を用いたキャッチコピー自動生成システム	111
	マイニングされた文の例	111

B.2	取得された単語群	112
В.3	生成例	113
付録Ci	章 SNS上の嗜好情報を用いたキャッチコピー自動生成システム	116
C.1	嗜好情報取得の詳細	116
	C.1.1 全ファンページの傾向	116
	C.1.2 嗜好情報のまとめ	117
C.2	生成例	118
付録Di	章 出力に対する専門家の意見	121

論文要旨

キャッチコピーは短い文字数で注目を引き,端的かつ効果的に対象の特徴を述べ,人々の心を捉え,引きつける.その応用先には,雑誌やカタログ,CM などに代表される広告利用,一般の人々でも関わる機会が多い出し物やイベント等の呼び込み広告がある.このようなキャッチコピーを分析し,自動生成を行うことは言語学のみならず,広告や産業界からの需要にも合致すると考えられる.

本論文では,キャッチコピーの自動生成に関する研究について述べた.本論文の目的は,自然言語処理における統計的手法をキャッチコピーの自動生成に応用することで,ユーザに対して興味を引く文を実現することである.そのためのアプローチとして,(1) 最大級のキャッチコピーコーパス,(2) 大規模 N-gram コーパス,(3) 膨大な嗜好情報を利用し,この課題の解決を目指した.

第1章では,キャッチコピーに至る人工知能,自然言語処理の変遷,およびその周辺に関して関連する研究の歴史的背景について述べた.

第2章では,独自に構築したキャッチコピーコーパスに対して自然言語処理のツールを用いて統計的分析を行い,文構造及び出現単語の傾向を調べた.分析によって,例えばキャッチコピー特有の表現が存在することやテーマに固有な単語が頻出することが明らかになった.また,これらの知見を基に,その特徴を模倣するキャッチコピー自動生成システムを提案した.

第3章においては,統計的手法のみでは解決が困難なキャッチコピーの間接的な表現に着目した.キャッチコピーには「これは丈夫です」よりも「象が踏んでも大丈夫」のように,間接的な表現の方が印象に残ることが多い.この方向でのアプローチを行うため,Webからの知識取得による常識獲得,およびキャッチコピーコーパスの更なる拡張をすることにより,間接的なキャッチコピーを生成するシステムの提案を行った.

第4章では、第3章の手法のシステムをより実世界的な Social Networking Service (SNS) サイトに適用することにより、投稿に相応しいキャッチコピーの生成を行った、SNS 上の嗜好情報に着目し、bag-of-words モデルを用いた線形回帰分析を適用することによって、より好まれる候補の選択を行った.これによって、システムによる生成のキャッチコピーはプロ作成には及ばないものの、人手選択に近いレベルでユーザの興味を引くことが示された.

第5章では,各章で得られた内容をまとめ,本研究の成果を要約した.また,将来の展望に対しても言及した.

第1章

序論

1.1 はじめに

本論文は,キャッチコピーについて分析を行い,その自動生成を行う生成システムについての研究である.キャッチコピーの研究は分野横断的である.本章では,キャッチコピーに至る人工知能,自然言語処理の変遷,およびその周辺に関連する研究の歴史的背景について述べる.

コンピュータの登場以降,人工知能の研究は急速な発展を続けてきた.人工知能の 定義は「理解し,推論し,行動することを可能にする計算の研究」であり,その工学 的目標は「知識表現を行い,知識を利用し,知識を組み立て,現実世界の問題を解く こと」である[1].

社会における生産性向上のためのアプリケーションが数多くある中で,人間しか持たないとされてきた創造性に関する研究も存在する.

例えば、ことば工学は、人間の感性、知性・遊び心などを扱う分野である、社会の加速的な情報化が引き起こすビッグデータの興隆により、人間の知性、意志および情動に関する情報が蓄積されている。同時に、計算機の処理速度の飛躍的向上によって、これらの情報を処理し、実世界に応用する流れは一層強くなっている。

認知科学等人間に関する学問も急速に発展している現代において,生産性を向上させる知的なアプリケーション以外にも,人間の情動・感性,意志に訴えかける人工知能の実現が期待される.

本章は以下のように構成される.まず,1.2では人工知能の歴史を概説した後,最近の研究例を挙げ,人工知能と創造性の関係性について述べる.1.3では,人工知能における創造性を言語に適用したことば工学について説明する.1.4では,ことばを扱うために必要な要素技術について概説する.1.5では,本論文においてターゲットとするキャッチコピー研究,および関連研究を含んだその歴史について俯瞰する.

1.2 人工知能

人が行う知的作業は,推論,記憶,認識,理解,学習,創造といった現実世界に適応するための能力を指す[2].人工知能は,人間の知的な働きと同じ動きを,機械によって実現することを目的とする科学技術,あるいは,そのような科学技術によって実現される具体的なシステムである[3].人工知能は「知能のある機械」である強いAIと,知能があるようにも見える機械,つまり,人間の知的な活動の一部と同じようなことをする弱いAIに区分される[4].

1.2.1 人工知能の歴史

人工知能の歴史を文献 [5] を参考に,以下のように年代別に分けて概説を述べる.

1950年~60年代

1956 年ダートマス会議において「Artificial Intelligence (人工知能)」が John Mc-Carthyによって提唱された.人間そのものの知能や知性を実現する,強い AI が目標とされていた.当初は知的とみなされる振る舞いをするプログラムが多く登場した.1966年に Joseph Weizenbaumによって,おしゃべり対話システムの ELIZA が開発されている [6]. MIT において Fortran や LISP の開発が行われ [7],問題解決を目標とする野心にあふれた General problem solver (GPS) [8] を始め,コンピューターに記号積分を行わせることに成功したり,チェスや数学の定理証明なども盛んに研究され,すぐにでも人間並みの知能が実現されるかのような楽観的雰囲気であったと言われている.

1970年代

1969年に,有限の情報処理能力しか持たない人工知能には,対象世界に存在する全ての問題に対処できないという「フレーム問題」が指摘される [2].定理証明やゲームのような閉じた Toy Problem と呼ばれる小規模問題から脱却できず,実世界の開いた複雑な問題を AI に解決して欲しいという産業界からの期待には応えられなかった.そのため,AI は Toy Problem を扱うだけと批判され,この時代,AI の研究開発は停滞していった [5].

1980年代

汎用の問題解決プログラムで高性能システムを開発するのは困難であり,専門知識 および専門的ノウハウとヒューリスティックを持つ者が専門家である考えから [9] ,エ キスパートシステムと呼ばれる知識を持つ人工知能が登場した.これは「知識の形で 表された専門的技術に推論のメカニズムを加える事によって働くコンピュータシステ ム」と定義されている [10] .

代表的な例は,1970年代中盤にスタンフォード大学で開発された感染症の診断を行う MYCIN であり,500 余りのルールから得られた診断結果は若い医者よりも優れたものであった [7].この時代のシステムの運用形態は,相談システムが中心であった [11].エキスパートシステムで 70 年代の終盤から 80 年代にかけて人工知能は第二次ブームを迎えたと考えられている.

エキスパートシステムを始めとして,人工知能に関する産業に一般企業が参入し始めた時代である.例えば,デュポン社では 1988 年までに 100 のエキスパートシステムが運用されており,500 が開発途中だったと言われている [7] . また,エキスパートシステムの課題であった一般ユーザに対して,より親和的に扱えるように行った研究も存在している [12] .

また,この時代には日本の第5世代コンピュータプロジェクト,米国の MCC (Microelectronics and Computer Technology Corp.) コンソーシアム,欧州のESPRIT (欧州情報技術研究開発戦略計画)のような国家プロジェクトが発足している[5].

ソフトコンピューティングに関しても,実用化がされてきた時代であり,例えば1980年後半には,ファジィ制御を取り入れた家電製品が作られ,次いでニューラルネットワークが産業界でも取り入れられ,炊飯器・エアコン等を始めとして,様々なものに「ニューロファジィ」を冠した家電が登場した.

1990年代

人間のもつ知識の奥深さや幅広さをコンピュータ内で表現することの困難さ,質の良い知識を獲得することに大きなコストがかかること,獲得された知識を維持することも大変であることが次第に認識され,世界で5,000程度開発されていたエキスパートシステムの2/3以上は動かなくなり,産業からのAI研究開発への期待は薄れていった[5]. 一方で,90年代以降,大規模データマイニング分析法としての機械学習とデータマイニング,モデリング基礎としてのオントロジーと Semantic Web 等,AI の基礎技術は着実に進展した[5]. 一方で,1997年には IBM のディープブルーが当時のチェスプレイヤーチャンピオンに勝利を収めている[13].

2000年代

2000 年代前半には, Web 上のデータを処理対象とし,自然言語処理技術,人工知能技術と融合した Web インテリジェンスの研究が行われるようになった [2].この時代に行われた,テスト分類や推薦システムなど,特定タスクに適した機械学習,データマイニングが進展し,現在のビッグデータブームの技術基盤を与えている [5].

2010年~

2010年以降, AIの第三次ブームを迎えていると考えられている.

人工知能に関して様々なブレイクスルーが起きている.2010 年に,あから 2010 が 清水女流王将を破った [14] のを皮切りに,将棋においてもコンピュータプログラムが 有段のプロ棋士に勝つなどしている [15]. ウクライナ在住 13 歳の少年という設定の下,チューリングテストをクリアした人工知能もある [16].

現在の人工知能の要素技術を用いることで,センター試験及び二次試験の試験問題に回答し,東京大学に合格を目指す「ロボットは東大に入れるか」[17]のプロジェクトも盛り上がりを見せている.

教師なしで抽出前の生データに近い信号から,概念そのものの学習が可能な深層学習 [18] も脚光を浴びている.

初期の AI が科学者達の知的好奇心に支えられていたのに対し,現在の AI は厳しい 競争を勝ち抜いてきた強豪企業が,あくまでも自らの事業に役立てる目的で研究開発 をしている [19].

最近のホットな研究例としては,Googleを始めとして他に多くの自動車会社が,しのぎを削って開発を進める自動運転の技術が挙げられる [20,21].会話可能な Apple の Siri を筆頭として,Google Now,そして新たにリリースされた Microsoft の Cortana も次の会話型インタフェースとして,注力がなされている [22].

機械学習,データマイニングとオントロジーおよび Semantic Web の研究開発は,コンピュータの高性能化,ネットワークの広帯域化というハードウェア基盤の進展と相まって,大きく進展するとともに,自然言語,画像,音声理解のような応用 AI 技術の進展も加速させ,これらの要素技術を統合したグランドチャレンジ的なタスク指向型,生活支援型 AI サービスが登場する時代を迎えたといえる [5].

以上の研究からわかるように「機械の知能化」と「人間知識の拡大」に対して注力されてきた.一方で、創造性に関しても人工知能の立場から扱う動きがあり、身近な知識の斬新な組み合わせを考慮できる点において、新しいアイデアを創出可能であるという指摘がある[23].次項では、人工知能と創造性について触れる.

1.2.2 人工知能と創造性

創造性を要する分野においては,アイデア,デザインに関する研究が盛んに行われてきた「発想支援システム」は「発散的思考,収束的思考,アイデア結晶化」までの人間の創造的問題解決プロセスを支援するコンピュータシステムのことである[24].このような分野では,例えば,対話型遺伝的進化計算を用いて工業製品デザインのアイデアを喚起させることを意図したシステム[25]や評価エージェントを用いてインテリアレイアウト支援を行うシステムがある[26].

芸術に関する分野,例えば,絵画においては自動化された画家システムの提案が行われている[27].また,生け花の分野でもシステムによる作成支援の提案がある[28].

音楽の分野でも,人工知能による創造性の研究が盛んである [29] . 例えば,作曲の分野でも個々の作曲者の作風の特徴量を適用することにより,バッハ,ショパン,ラフマニノフ,モーツアルト,ストラヴィンスキーの作品であると認識される曲を作ることに成功している [30] . また,より完成度を高めるために奏者自身のテンポ・強弱等の演奏効果に着目した研究がある [31] . 対話型遺伝的プログラミングを用いて,クラシック音楽の作成支援を行うシステム [32] や,バロック期に用いられていた明示的な和音の指定がない通奏低音に対して,人間のような即興的伴奏を付与可能なシステム [33] の提案が行われている .

このように創造性を生み出す人工知能には様々なアプローチがある.次節では,この人工知能の創造性を言語に適用したことば工学について説明する.

1.3 ことば工学

遊び心を含めた [34]「イメージ,感性や感情・情動」などを中心に据えた言葉をコンピュータを用いて扱う研究が盛んに行われている.その中にことば工学 [35] がある.ことば工学は,ことばの世界に閉じない世界を積極的に扱いながら,ことばの問題を論じることであり,このような領域は人間の知的活動の多くを占めているものと考えられる.

図 1.1 にことば工学に関する概観を示す. 図からわかるようにことば工学は,言語学を始めとして脳科学に至るまで,様々な分野にまたがっている. 以下に,ことば工学の認知・心理学的背景,研究例,期待される応用分野について述べる.

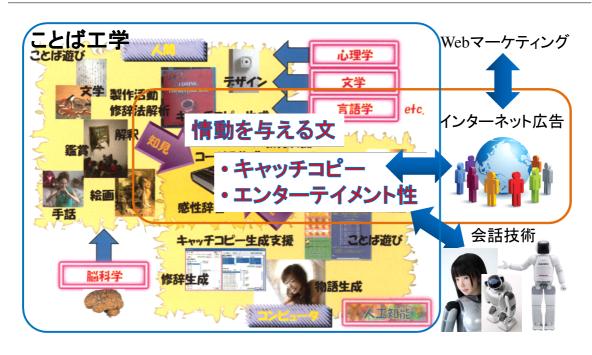


図 1.1 ことば工学 [35-37] と対象テーマの概観 (ロボット等 [38-41])

1.3.1 ことば工学の認知・心理学的背景

比較的古くから主に言語学や心理学において,人間の言葉の感性について研究が行われてきた [42]. その中で,特に言語の感情的な側面からのアプローチもある [43]. 言語を刺激とする情緒反応の研究として,機器を用いて計測したものや,アンケート形式で連想された単語等を記録した研究も多い [44,45].

例えば,文献 [44] は,秋田が 40 年あまりの歳月をかけてアンケート収集した,様々な刺激語に対するイメージ等が含まれており,一例として 100 人を超える被験者を対象とした 1,150 語の形容詞刺激語に対する連想反応語が収録されている.

他にも,NTT コミュニケーション科学基礎研究所が公開している親密度および心像性に関する言語資源が存在している [46,47].ここで,親密度とは該当単語がどれだけ馴染みがあるかの指標であり,心像性とはある単語におけるイメージの容易さを表す.例えば「リンゴ」などの具体的な名詞は想像しやすいのに対し「一般」などの抽象的な名詞は想像しにくい.このような各々の単語における親密度,心像性の度合いを収録している.

一方では,fMRI (functional Magnetic Resonance Imaging: 機能的磁気共鳴画像)や NIRS (Near-infrared Spectroscopy: 近赤外分光法)を用いて脳活動を計測する研究が活発化している.例えば,前述の心像性を脳活動から分析を行っている研究 [48] が存在する.また,動植物や道具といった単語属性の差異によって側頭葉における脳のマッピング位置に相違が見られることが明らかになっている [49].単語間の連想関係や連

想概念距離が脳活動に影響を与え,より人間らしいオントロジーモデルの構築につながる可能性も示唆されている[50].

最近では,工学分野の自然言語処理において用いられている LDA (Latent Dirichlet Allocation) [51] と組み合わせて,脳の活性化部位から文章生成を試みる融合的研究も行われるようになってきている [52] .

1.3.2 ことば工学の研究例

書き言葉では,書籍等に見られる情報伝達に重きを置いた比較的長い文章が用いられる.これに関しては,物語に関する研究が存在しており,小説家,SF作家であった星新一の小説の自動生成を行うことを目標とするプロジェクトが発足している[53].ここでは,短い小説であるショートショートの物語構造の計量的解析[54]等を基に,現在「星新一らしさ」に関する考察[55]等,研究が盛り上がりを見せている.

一方で感性的な要素を多く持つキャッチコピー,標語,俳句,会話文,遊び言葉に分類される地口(駄洒落)等の比較的短い文が存在する.この際,修辞のようなテクニックを用いることがあり,様々なアプローチの研究例がある[56].ここでは,キャッチコピー以外の例として,オノマトペ,オクシモロン,地口について示す.

オノマトペは擬音語,擬態語のことであり,感情や感性を表現するものとして広く用いられている.認知科学の分野において,このオノマトペとテクスチャとの関連について分析を行っているものがある[57].工学の分野でも,状態を表現することが容易であることから,様々なアプリケーションが生まれており,例えば音象徴から印象を推定を行ったり[58],ロボットの動きに適用する研究例[59]がある.

また,オノマトペは幼児期の言語発達において重要であるという指摘[60]がある.さらに,単語のみや対象の音のみに比べ,脳の活動範囲も広い[61]という知見も明らかになっている.これを基に,簡素なオンライン上のキャッチコピーにオノマトペを付与し,魅力を向上させることを目的としたシステムもある[62].

オクシモロン (Oxymoron) は「常識的には両立しないふたつの対義語または対義表現を無理に結合することばづかい」と定義され、具体的な例としては「小さな巨人」「近くて遠い国」などが挙げられる[63]. オクシモロンには様々な種類がある[64]. このオクシモロンに関して、相互情報量を基に生成するアプローチがある[65].

言葉を扱ったユーモアに関しては次のような研究が行われてきた.駄洒落[66-68]やことば遊び[69,70]などである.これらの駄洒落研究においては,音韻の類似度や概念ベースを利用して計算機で生成を行っている.ことば遊びでは,なぞなぞがフレームワークとして利用されており,同音異義語および因果関係の成立のしやすさなどが生成の際の指標として用いられている.海外でも,英語の駄洒落を意味する"pun"[71-73]

を始めとする Computational Humor の研究が盛んである [74]. これらに加え,すかしというフレームワークおよびファジィルールを用いて,笑いを誘うような面白いことわざの自動生成を試みる研究も存在している [75].

また、枠に収まらない言語の処理として、不自然言語処理の研究についても盛り上がりを見せている。海外においては、AND (Analytics for Noisy Unstractured Text Data) やSANCL (Syntactic Analysis of Non-Canonical Language) などのワークショップも主要国際会議と併設して開催されており、ウェブや電子メール上の構造化されていない、非正規的な言語の処理に対して、注目が集まっている[76].

国内においても不自然言語処理の分野に関心が向けられており,顔文字[77]等の研究も行われている.

1.3.3 期待される応用分野

社会の需要として,ニューロマーケティング等,認知心理学・脳科学等の知見を用いてビジネスに応用しようという動きが加速している[78].従って,これら心理学等で得られた知見やデータを情報処理技術の分野に適用していく意義は今まで以上に大きくなっていると考えられる.

ことば工学の成果によって生成されるこれらの文は,短いながらも多様な感情を表現し,コミュニケーションを円滑にする.ユーザに情動を与える点で,ことば工学はインターネット広告やロボットの会話等の社会に還元可能であると考えられる.例えば,キャッチコピーは対象となる商品のメリットを伝え,心を動かすことによって,人々に行動を促す特徴を持つ.このような文の特徴を明らかにし自動生成を試みることは,人間の知性を計算機で実現するにとどまらず,Webマーケティング,人間を楽しませ円滑なコミュニケーションを行うロボット等のインターフェイスの基礎技術にも応用できる可能性がある点で,意義深いと考えられる.

1.4 自然言語処理

ここでは,ことば工学を扱う上で必要な自然言語処理の要素技術について述べる. 自然言語処理は,自然言語で書かれた文書の集まりから情報を抽出し,構造化し,再 構成する[79]ような処理のことをいう.自然言語処理においては,膨大な情報の中か ら必要なものを検索したり,日々生産される情報を自動分類したり,要約や可視化に よって分析を行う「情報加工技術」が重要である.

自然言語処理研究の究極の目標は,言語を理解する機械をつくること,すなわち自然言語理解である.

自然言語処理は,規則に基づく全盛時代を経て,1990年初頭のパラダイムシフトを受けて,コーパスおよび統計的手法に基づく研究が主流になっている.その理由として,以下の要因が挙げられている[80].

- 曖昧性解消の問題: 論理的制約に基づく文法記述と構文解析では本質的な曖昧性に対処することが難しく,曖昧性爆発を起こすか,より厳しい制約を記述する必要性があったこと。
- 頑健な処理に関する問題: 制約条件が厳しすぎる文法記述は厳格すぎて,わずかな文法誤りも許容できず解析結果が得られないことがある.多少の文法誤りにも破綻しない頑健な言語解析を行うことに困難があったこと.
- 規則の拡張の問題: 自然言語処理システムの規模が大きくなるにつれ,文法規則の規模が増大し,人手でのメンテナンスが文法開発以上の問題になっていたこと.
- 例外への対応: 言語が美しい規則性を持つ反面,言語現象には様々な例外があり, 単純な文法規則の拡張では,全ての言語現象をカバーする文法を構築するのが難 しかったこと.
- 電子化データの増大: 電子化された大規模テキストデータやタグ付けコーパスの 出現,および World Wide Webの普及により電子化データの蓄積が急速に加速 されたこと. 同時にこのような実データの解析に対する要望が高まったこと.
- 計算能力の進歩: コンピュータの処理能力や記憶容量が飛躍的に増大し, それまでディスクに置く以外扱えなかったような大規模データをすべて主記憶に読み込んで処理することが問題なく行えるようになってきたこと.

自然言語処理において課題とされているテーマには,形態素解析,構文解析等の基礎的技術,テキストマイニング等の応用的技術がある.また,理解のために必要なコーパス構築も重要である.

以下に,コーパスおよび統計的手法による代表例を示す.

1.4.1 コーパス

自然言語処理において,用いるコーパスは大変重要である.コーパスには様々な種類があるが,構造(主に木構造またはネットワーク構造)を持つものとそうでないものに大別される.

構造を持つもの

構造を持つコーパスとしては, WordNet [81,82] や CYC [83], SCONE [84] などが挙げられる. 特に WordNet に関しては,様々な辞書との統合が図られ,例えば SentiWordNet

[85] のような感情,情動情報が登録されたコーパスも登場している.

日本語においては , 日本語 WordNet [86] を始めとして , EDR [81,87] , 日本語語彙大 系 [88,89] , 連想概念辞書 [90] が存在する . より意味論に踏み込んだ Fillmore のアイデアに基づいた日本語フレームネット [91] も提案されている .

一般に,これらのコーパスは明示的に知識が表現されているため,意味理解へと応用が期待できる.一方で,これらは人手で構築するためコストがかかるのに加え,新しい概念に対応できない.

構造を持たないもの

ここでは,ノードを持たない構造である主に均衡コーパス等の例を挙げる.

英語圏でのBrown Corpus [92] の登場等,大規模なコーパスが構築されてきた.

日本でも,ビッグデータの興隆もあり [93],2,550 億単語から構築された Web 日本語 Nグラム [94,95] を始めとして,書籍・雑誌・新聞など 11 種類の媒体におけるデータから構成されている BCCWJ [96,97] などの大規模言語資源が利用できるようになっている.N-gram のみにとどまらず,日本語に特有な格に着目し,用言とそれに関係する名詞を用言の各用法ごとに自動整理したコーパスも構築されている [98].

上の構造を持つものとのハイブリッドという視点から、Wikipedia 等の百科事典から自動的に構造を構築する手法がある。例えば、DBpedia [99] は膨大な量のオントロジーを持つに至っている。更にこれに対し、同義語・クラス階層についてより詳細に記述されている Wikipedia オントロジーが利用可能になっている [100].

以上のように,様々なコーパスが存在している.これらのコーパスを用いる手法が,次に紹介する統計的アプローチである.

1.4.2 統計的アプローチ

上で述べたように,統計的手法は現在の自然言語処理において主流の方法である.Zipf-Mandelbrot の法則 [101,102] と始めとして,統計的手法は確率的言語モデルによって構築され,N-gram モデルがよく用いられる.言語のモデル化において,文字列や単語列などの実際のコーパスでの頻度を利用すると,有効な結果が得られることが知られている [103].言語モデルのうち,特に重要で代表的であるものは,N-gram モデル,隠れマルコフモデル,確率文脈自由文法である [104].

それぞれの要素となる N-gram に対して,Jaccard,Dice,Simpson [105],相互情報量 (PMI), χ^2 値などが用いられて処理が行われる.その中でも,Web を大規模コーパスと見立てた上で,Web 上の検索エンジンによるクエリに対する結果の頻度を利用し

て,関連度を推定する方法が用いられている [106] . また $\mathrm{TF} ext{-}\mathrm{IDF}$ 等の手法も利用されている .

12

また,統計的手法をテキストマイニングに応用した手法も存在する.

1.4.3 テキストマイニング

テキストマイニングは,自然言語処理やデータマイニングなど多様な技術を組み合わせた複合技術である.対象となる文書データを扱うための自然言語処理に加え,膨大なデータから有用なパターンを発見するためのデータマイニングの技術や,テキスト分類などの機械学習のアルゴリズムが必要となる[107].Google のような大規模なシステム上で行われているものや[108],テンプレートを用いて口コミの情報検索を行ったり[109],単語の印象を推定する研究もある[110].フレーズに対応する表現を用いることで,知識獲得を行っている.例えば,文献[110]では,11個のフレーズを用いることで名詞と動詞に関する組み合わせに対して特徴を自動的に獲得し,例えば「花が咲く」時の印象を「綺麗」であると推定している.

文献 [109] は主に三つ組のような,テンプレートに当てはめる手法を用いているが, 高野ら [111] のように可変長の知識表現を行っているもの,因果関係に着目し知識抽出 を行うもの [112] も存在している.

1.4.4 文生成

これまで触れたように,N-gram から文を構築するものが文生成であり,マルコフ連鎖等が用いられている.例えば,応用先には文章要約のための文生成がある.Bankoらの,1文よりも短いヘッドラインの生成を目標とした統計モデル [113] や Web ページのヘッドライン作成のための統計的要約 [114] などがある.また「いつも正確な機械操作」を「いつも正確に機械を操作する」に変換するように,一文に対して言い換えの生成を行う手法の提案も行われている [115] .

1.4.5 ことば工学と自然言語処理

認知言語学の中に比喩がある [116,117].ことば工学と関連して,比喩の分析が自然言語処理においても行われている.ここでも,人間の知能と深く関わる比喩は言語処理の中でも,統計的手法に基づく頻度によって意味の尤もらしさの推定を行っている [118].

以上で見たように,統計的手法によってキャッチコピーにアプローチすることを試 みる.

名称
キャッチコピー,標語,スローガン,惹句,キャッチフレーズ
tags, tag lines, taglines, theme lines
end lines, endlines, straplines
claims

表 1.1 各国のキャッチコピーの名称

1.5 キャッチコピー

国名

オランダ, イタリア

(日本以外)一般に

日本

アメリカ イギリス ドイツ

ベルギー

フランス

1.5.1 キャッチコピーとその目的

baselines

signatures

slogans

pay-offs, payoffs

言葉を用いたコミュニケーション , 特に書き言葉は人間社会において重要な役割を果たしている . 広告活動においてキャッチコピーは大きな影響を私達に与える . 一般の書き言葉と異なり , キャッチコピーは短い文字数で注目を引き , 端的かつ効果的に対象の特徴を述べるものである . その目的は , 受け手の注意を引いて , その関心と興味をかき立て , さらに送り手の期待するような欲望を起こさせて , 最後にはその欲望に沿って , 受け手にある種の行動を行わせることである [119] . 営利目的のキャッチコピーの場合は 「, 利益を極大化するために , 購買者を説得する意図を持った情報のマス・コミュニケーション 」[120] という定義がある .

キャッチコピーには,主にブランド構築を意図する広告キャッチコピーと,商品の販売促進を促す商品キャッチコピーが存在する.どちらの種類のキャッチコピーも,ブランドあるいは商品となる対象に対して好意的なイメージを持たせることを目的としている.従って,キャッチコピーには目的とする対象に関して,情報を伝え消費者の心を捉える必要性があると考えられる.

キャッチコピーには様々な名称がある.各国の違いについて表 1.1 に示す [121].日本語において,キャッチコピーはキャッチフレーズとほぼ同義であり,標語として用いられることもある.なお,本論文のキャッチコピーの英語訳はアメリカで使われている "tag line" を利用することとする.

1.5.2 キャッチコピーの変遷と社会的背景

時代によってキャッチコピーの中身は変化してきており,単に商品の登場を告げることを第一にした「知らせればいい」時代から,マーケティング理論に基づいた生活提案型の広告が展開された時代,感性や気分を重視した,いわゆる「ムード広告」「フィーリング広告」の時代を経て,現在では広告のエンタテインメント化は一層進みつつあると言われている[122].現代では生き残りをかけて,企業が価格競争よりもブランド構築を重要視するようになってきているとも言われている[123].

また,広告文の側面も持つキャッチコピーは,経済学や言語学からも盛んに研究が行われている. Kohli ら [124] は,効果的なキャッチコピーを作成するために必要な要素の考察を行っている. Lowrey ら [125] は,ブランド名における語音象徴に関する研究をしている. 新井 [126] は,「言葉はなぜ通じるのか」という「関係性の理論」からキャッチコピーの特性分析を行っている.日本社会における独特の文化を背景とする,「場」に焦点を当てた考察 [127] も存在する.

越川 [128] はブランドネームについて語感に関する分析を行っている.また,脳科学の見地からのブランドの考察も試みられている [129,130].

キャッチコピーの応用先には,雑誌,カタログや CM などに代表される広告利用,一般の人々でも関わる機会が多い出し物やイベント等の呼び込み広告がある.前者はプロのコピーライターがキャッチコピーを作成するが,後者のような例も多く存在し,プロでない人がキャッチコピーを作成するケースが少なくない.

また,近年ではオンライン上での広告の需要も増大している.例えば,AmazonやeBayを始めとして,世界規模でオンラインショッピングが普及している.日本でも価格.comや楽天におけるショッピングが多くの人々に利用されているが,EC(electronic commerce: オンラインショッピング)化比率は3.1%と主要国と比べても低く,2014年10兆円である市場規模が2020年には25兆円に増加するという予測もある[131].2012年の調査によれば,日本におけるインターネット広告費は8,680億円であり,直近10年間に渡って増加しており,広告全体の市場において14.7%を占めるまでになっている[132].このことからWebにおけるキャッチコピーを代表とする広告媒体の重要性が今まで以上に増大していることが示唆される.

さらに,これはパーソナライズ化の流れを作っている.インターネットの最大の特徴は,広告媒体であることに加え,インターネット内でユーザと情報発信者相互にコミュニケーションを取りつつ,ユーザに具体的なアクションを起こすように誘導できる点にある[133].代表的な例が,利用ユーザによる口コミである.

Web で情報収集を行い,商品を購入する消費者数が増大している.従来,広告業界では,Attention,Interest,Desire,Memory,Action(AIDMA)のプロセスに分けて,

それぞれの認知要素 , 知識要素 , 情緒・感情要素 , 意図要素 , 行動要素に消費者の態度を構成するモデル化が行われていた . それに対して , 従来の AIDMA プロセスではなく , AISAS (Attention, Interest, Search, Action, Share) が提唱されており [134] , 検索とシェアを考慮したマーケティングが必要となっている .

このように,個人がよりインタラクティブに社会に影響を与えることが可能となっているため,一対多に代表されるマス広告よりも,より特定層や個人に対して,訴えかける広告の需要が増加していると考えられる.具体的には,Amazon に代表される協調フィルタリング [135] 等の技術を使うことが,今後のトレンドとして考えられる.以上のような背景によって,よりパーソナライズされたオンデマンドなキャッチコピーを作る需要は増しているといえる.

1.5.3 キャッチコピーに関するシステム

このように,人間の心を捉えるキャッチコピーを自動生成するシステムへの需要もあると考えられる.以下,海外(英語圏)と国内のキャッチコピーに関するシステムについて述べる.

海外のキャッチコピーに関するシステム

英語圏では,キーワードをテンプレート形式に当てはめ,キーワードを含むキャッチコピー生成システム [136,137] が存在する.また,特にオンライン広告に関する技術に関して数多くの特許が申請されている.1990 年代後半から 2000 年代中頃にかけてのインターネットの興隆を受けて,E メールにおける広告の需要が増大した.これによって,受信後クライアント側のローカルマシンでも広告表示可能である特許 [138] やユーザに応じてテキストにメタデータを付与する [139] 等数々の広告のアイデアが具現化された.このような背景から,各ユーザの電子メッセージに対してキャッチコピーを選択する技術 [140] やメールのやり取りに従って,作成済みの広告の中身が遷移するようなアイデア [141] も採択されている.

国内のキャッチコピーに関するシステム

日本においても,インターネットの普及とともにオンライン上での広告という形でキャッチコピー関連の研究が行われている.オンラインショッピングの需要の高まりを受け,アクセスログ解析を用いたサイト作成支援方法についての特許も存在している [142].以上のような背景からキャッチコピーに関するシステムが提案された.特に,

酒井ら[143]は,予め用意したテンプレートに対して,ユーザの履歴からキーワードを取得し,ユーザに対してキャッチコピーの生成をするシステムの提案を行っている.

この流れとは別に,キャッチコピーの生成支援を行うシステムが提案されている.松平ら [144,145] は,質より量,批判厳禁を旨とする発散的発想支援的な立場から,遺伝的プログラミングを用いたキャッチコピー生成システムを提案した.西原ら [146-148] は,研究発表に限定し,タイトル毎に興味を引きやすいかどうかをランク付けするアルゴリズムを提案した.中野ら [149,150] はブログ記事を用いて目を引くキャッチコピーを作るための基礎的研究を行った.森本ら [151] によって,商品の属性と属性値を利用したキャッチコピーの自動生成も提案されている.幅野ら [152,153] によって映画のキャッチコピー生成を目指したものが存在する.

以下の表 1.2 に,国内外のキャッチコピー研究を時系列順にまとめる.

表 1.2 キャッチコピーに関するシステムの変遷

年	出来事
2004 & 05	遺伝的プログラミングを用いたキャッチコピー生成支援システム [144, 145]
2007	聴講者 (多く) の興味をひく研究発表タイトルの作成支援 $[146,147]$
	目を引くキャッチフレーズ生成に関する基礎的研究[149]
2008	Title-composing support system for reaching new audiences [148]
	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 [150]
2009	Method and apparatus for adding advertising tag lines to electronic messages [140]
	キャッチコピー生成装置およびキャッチコピー生成プログラム [143]
	Systems, methods, network elements and applications for modifying messages [141]
2010	商品の属性と属性値を利用したキャッチコピーの自動生成 $[151]$
2013	映画のキャッチコピー作成支援手法 [152]
2014	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[153]

海外のもので確認されているものは,予め作成されたキャッチコピーが含まれる広告を呈示するもののみであり,文生成の技術を用いているものは見受けられない.

国内においてもこれらの手法が提案されているが,実際に評価実験まで述べているものは,2004年と2005年の松平ら[144,145]および西原ら[146-148],幅野ら[153]のもののみである.最終的に実世界で利用可能なアプリケーションを目指し,キャッチコピーの自動生成システムの構築し,評価実験を行うことで生成されたキャッチコピーの質の検証を行う.次節に本研究の位置づけを示す.

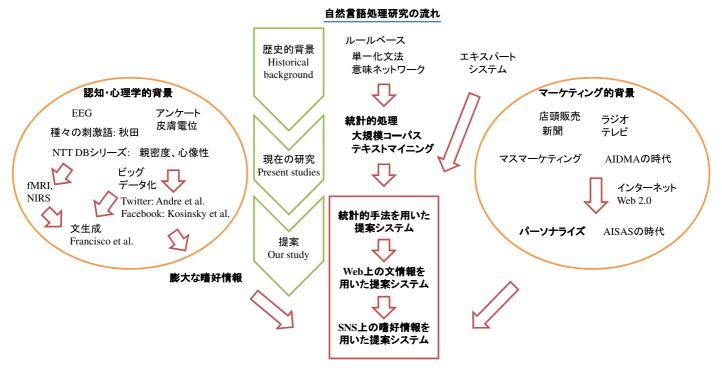


図 1.2 本研究の位置づけ

1.6 本研究の位置づけ

これまでに述べた本研究の位置づけを図 1.2 にまとめた . 本論文において , キャッチコピーに着目しその自動生成を行う研究について述べる .

本研究は,認知・心理学およびマーケティングの背景を考慮して,統計的自然言語処理のアプローチを行っている.まず,ルールベースから始まった自然言語処理では,ルールベースでの柔軟性への閉塞感がある一方で,大規模なデータが入手可能であることから,現在においても統計的手法によるアプローチが主流である.本研究の対象であるキャッチコピーに関しても最大級のコーパスの構築が可能であったことから,この流れに則り,統計的手法を用いたキャッチコピー自動生成システムの提案を行う.

この統計的手法では生成が難しい,間接的な表現を含むキャッチコピーについても解決を試みる.Webの発展に伴い,コーパスに含まれない知識の取得が可能になっている.これを用いて伝え方に着目したキャッチコピー自動生成システムの提案を行う.

また,マーケティングにおいて,ラジオ・テレビ等のマスマーケティング(AIDMA)の時代から,インターネット,Social Networking Service (SNS)を始めとするWeb2.0の時代に入り,パーソナライズ等が必要になってきている.この需要によって,キャッチコピーにもよりユーザの嗜好を反映できることが求められている.

一方で、認知・心理における研究はアンケートや皮膚電位、脳波によるもの等から

表 1.3 提案システムの技術と狙い

名前	入力	生成方法	作成可能な表現	実世界向きか
オンライン [136]	キーワード	作成済みの	キーワード含む	クリエイター向け
システム [137]		ものを選択		
海外特許	メッセージ	作成済みの	作成はしない	ログデータ
[140, 141]	(メール等)	ものを選択		を考慮
松平ら	単語	動的	EDR コーパス	作成支援
[144, 145]		(対話型 GP)	に依存	クリエイター向け
西原ら	研究発表タイトル	作成済みの	作成はしない	作成支援
[146-148]		ものを選択		クリエイター向け
中野ら	ユーザ作成による	動的	与える知識量	作成支援
[149, 150]	対象の説明文		に依存?	クリエイター向け
酒井ら	ユーザ履歴	動的	テンプレートに	個々のユーザの
[143]			制約	自然言語入力要
森本ら	商品名	動的	より普遍的な	作成支援
[151]	(またはカテゴリ)		テンプレート	クリエイター向け
幅野ら	映画名	動的	あらすじと	作成支援
[152, 153]			レビューに依存	クリエイター向け
統計的手法	キーワード	定量を鑑み	キーワード含む	意図して
(第2章の提案)	(テーマ , キーワード)	動的に生成		いない
Web 上の文情報	キーワード	動的	キーワードに	意図して
(第3章の提案)	(対象,性質)		対して間接的に	いない
SNS 上の嗜好情報	ファンページ上の	動的	投稿内容に	ユーザの嗜好
(第4章の提案)	投稿		依存	情報を考慮

始まり、近年のfMRIやNIRSなど空間、時間分解能が高いツールを用いるものに発展している。アンケートによって得られた単語特性に関する研究も存在しており、NTTが提供している親密度、心像性に関するデータベース、秋田による種々の刺激語に関するコーパス等が存在している。しかしながら、キャッチコピーに応用するには、データ収集のコストがかかる点で困難である。一方で、SNSサイトの興隆から人々の嗜好情報に関してデータ取得が可能になってきており、特にそこでの嗜好情報は個人の属性を強力に反映するという報告がある[154]。そこで、今回のアプローチとして代表的SNSから膨大な嗜好情報の取得を行い、キャッチコピーに応用する。これによって、よりユーザに好まれるキャッチコピーの選択が期待出来る。

本論文における,既存手法と比較した技術的な特徴を表 1.3 に示す.

表に示すように,それぞれのシステムの特徴を,入力,生成方法,作成可能な表現, 実世界向きかの点で評価付けを行う.第2章で提案を行う統計的手法を用いたキャッチ コピー自動生成システムでは、分析を行った上で定量的にキャッチコピーらしさを考慮して生成する点を特徴としている.第3章で提案するWeb上の文情報を用いたキャッチコピー自動生成システムでは、Webを大規模なコーパスと見立てて知識取得を行い、間接的なキャッチコピーの生成を目指している.第4章で提案するSNS上の嗜好情報を用いたキャッチコピー自動生成システムでは、投稿に対するSNS上のユーザの嗜好情報を利用することで、実世界向きのアプリケーションを目指している.

1.7 本論文の構成

本論文は全5章から構成される.

第2章では,独自に構築を行ったキャッチコピーコーパスに対して自然言語処理の ツールを用いて統計的分析を行う.また,ここで明らかになった知見を基に,その特 徴を模倣する統計的手法を用いたキャッチコピー自動生成システムを提案する.

第3章においては,統計的手法のみでは解決が困難なキャッチコピーの間接的な表現に着目する.キャッチコピーには,対象について直接的な表現を含まないが,間接的な表現をすることによって印象に残るものが多い.この方向でのアプローチを行うため,Webからの知識取得による知識獲得,およびキャッチコピーコーパスの更なる拡張をすることにより,間接的なキャッチコピーを生成する手法の提案を行う.

第4章では,第3章の手法のシステムを,より実世界にSNSサイトに適用することによって,投稿に相応しいキャッチコピーの生成を目指す.生成された候補に対して嗜好情報を用いることによって,さらにユーザの興味を引くキャッチコピーの選択を目指している.

第5章では,本論文で得られた成果についてまとめる.

第2章

統計的手法を用いたキャッチコピー自動生成シス テム

本章では,キャッチコピー特性の統計的側面を考慮し,生成に応用する.前半(2.1 - 2.3)では統計的処理によりキャッチコピーの特性を分析する.他コーパスとの比較, およびキャッチコピーコーパスにおける単語,係り受け,キャッチコピーが属する分野について分析を行うことでキャッチコピーの性質を明らかにする.

後半 (2.4-2.7) ではその知見を基に,入力されたキーワードを含む,統計的手法を用いたキャッチコピー自動生成システムの提案と評価を行う.

2.1 キャッチコピー分析の概要

前半部では,比較的最近の営利目的のキャッチコピーが大量に収録されている文献 [155](以下キャッチコピーコーパス)を分析の対象とする.これらのキャッチコピーは,すべてが確実にプロ作成のものと断定は出来ないが,少なくともカタログやチラシに掲載されたキャッチコピーである点で,十分に練られ作成されていると考えられる.

さらに,他の種類の文章が含まれているコーパスとの比較および分析を行うことによりキャッチコピーの特徴を明らかにする.

2.1.1 キャッチコピーコーパスと比較対象とするコーパス

比較対象とするコーパスは,文法的に正しく適切な説明文が用いられているインターネット百科事典 Wikipedia コーパス [156],様々な言い回しや話題を扱っているブログコーパスである KNB コーパス [157],文法的には必ずしも厳密ではないが,自由度の高い会話文コーパス [158] である.表 2.1 にそれぞれの詳細を示す.

表からわかるように,それぞれ数千の文からなるコーパスを用いて分析実験を行う. キャッチコピーコーパス [155] は,食品・美容・生活用品を始めとする 12 の分野を包含 しており,全体で 6,466 個のキャッチコピーが掲載されている.百科事典コーパス [156]

コーパスの種類				
	内訳: 食品 (495個), 美容 (451個)			
	生活用品 (1,133 個), ファッション (1,066 個)			
	デジタル関連製品 (671 個)			
キャッチコピー	キャッチコピー 家庭用電化製品 (649 個)			
コーパス [155]	・パス [155] 交通・レジャー (450 個), サービス (253 個)			
	ハウジング (429 個), 自動車 (385 個)			
	メディア (341 個), その他 (143 個)			
百科事典コーパス [156]				
ブログコーパス [157]				
	会話文コーパス [158]	5,384		

表 2.1 コーパスの種類と含まれる文の数

については、情報量が多く処理に時間がかかるため、全体から約1万文をランダムに 抽出し用いる.

2.1.2 コーパスの前処理と解析方法

前半では,キャッチコピーの言語工学的観点からの分析を行う.具体的にはまず2.1.2で,用いるコーパスの前処理と解析方法について述べた後,2.2で品詞,単語特徴,係り受け関係について他種類のコーパスとの比較を行い,キャッチコピー固有の特徴について明らかにする.さらに2.3で,キャッチコピーコーパスそのものに含まれる単語,係り受け関係,分野ごとの名詞に着目した分析・考察を行う.

図 2.1 に , コーパスの前処理と解析の流れを示す . 前処理では , 各コーパスに含まれている分析に不必要な成分の削除を行う . 例えば , 括弧 "()"内に入っている文や , 不必要な記号などがあげられる . さらに形態素解析器 MeCab [159] , 構文解析器 CaboCha [160]を用いた解析を行う際に , エラーを発生させる成分の削除・変更も行う . 例えば , MeCabでは半角記号に対して , 全て名詞判定を行ってしまう . そこで , 全てのテキストデータに関して半角記号を全角に変更する . ただし年号のように数字が連続する場合には半角に戻し , 数字列を名詞として扱わせる , などの例外的な処理も行っている . 前処理の後 , 一文ずつ形態素解析を行い , 各コーパスにおける品詞を取得する . CaboChaの場合には , 係り受け関係において係る側と係られる側の形態素列を取得する .

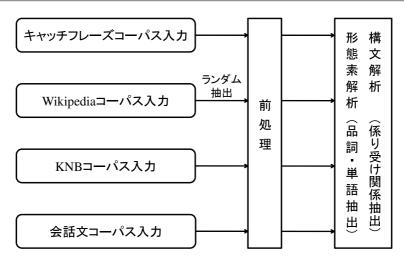


図 2.1 解析の流れ

2.2 他コーパスとの比較による特徴分析

ここでは,キャッチコピーコーパスと他コーパスについて,品詞の使用割合および 品詞列に着目した分析を行う.

2.2.1 全体の品詞使用割合の分析

図 2.2 に,解析によって得られたキャッチコピーコーパス全体における品詞の使用割合を示す.全体の品詞使用割合の傾向としては名詞の割合が一番多く,助詞・記号・動詞と続いていることがわかる.キャッチコピーにおいて,名詞が重要であるといえる.

次に,図 2.3 に百科事典 (Wikipedia) コーパス [156] における品詞の使用割合を示す.図 2.2 のキャッチコピーコーパスの場合と比較すると,割合が違うものの,品詞割合の順位がかなりの部分で一致していることがわかる.一方,百科事典コーパスにおいては特に名詞割合が多い.理由として,Wikipedia は名詞として扱われる年号や専門用語が多く,名詞の連語が多かったからと推測できる.また,Wikipedia にはほとんど形容詞が見られなかった.

図 2.4 には,ブログコーパス [157] の品詞割合を示す.ブログコーパスは,図 2.3 の百科事典コーパスと比較すると,より図 2.2 のキャッチコピーコーパスの割合に近いことがわかる.理由としては,ブログでは様々な言い回しが用いられているため,品詞の使用割合がよりキャッチコピーに類似したためと考えられる.

図 2.5 に会話文コーパス [158] の品詞割合を示す.図から会話文コーパスは,他のコーパスの品詞使用割合と大きく異なっていることがわかる.これは会話文特有に,話の途中で終わってしまったり,主語となる名詞の省略などが起こったりするケースが

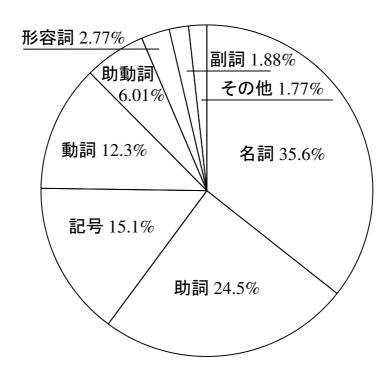


図 2.2 キャッチコピーコーパスにおける品詞の使用割合

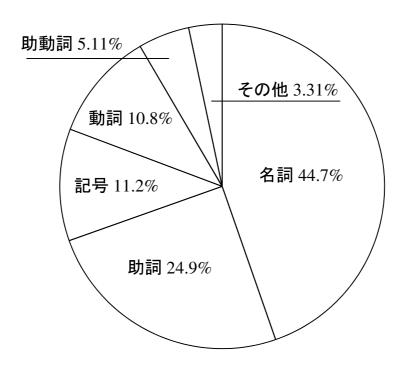


図 2.3 百科事典コーパスにおける品詞の使用割合

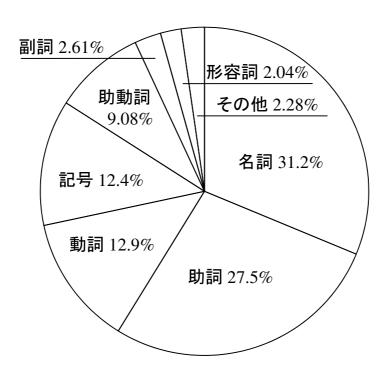


図 2.4 ブログコーパスにおける品詞の使用割合

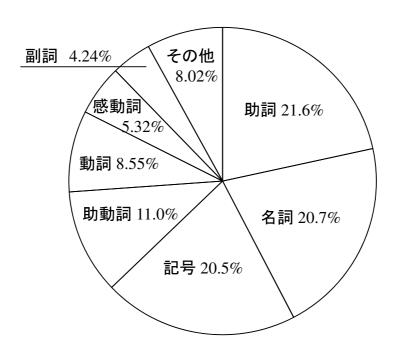


図 2.5 会話文コーパスにおける品詞の使用割合

名詞,	助詞,	名詞,	名詞,	助詞,	•••
接頭詞,	名詞,	名詞,	記号,		
名詞,	名詞,	名詞,	名詞,	名詞,	
名詞,	名詞,	助詞,	名詞,	名詞,	
名詞,	記号,	名詞,	名詞,	名詞,	• • • •
助詞,	動詞,	動詞,	記号,		
名詞,	助詞,	名詞,	名詞,	名詞,	
名詞,	名詞,	記号,	名詞,	名詞,	• • • •
:	:	:	•	:	÷
1gram	2gram	3gram	4gram	5gram	

図 2.6 順方向の品詞 N-gram 解析方法

頻繁に見られ,名詞の使用割合が下がったためと考えられる. また,感動詞の割合が 多いことも特徴として挙げられる.

以上から,会話文コーパスを除く,三種類のコーパスにおいては名詞が占める割合 が最も多く,キャッチコピーコーパスに含まれる名詞の割合は百科事典コーパスとブログコーパスの中間程度であることがわかった.

2.2.2 品詞列に着目した分析

キャッチコピーの文法構造に着目した分析を行うために単語の連なりを扱う N-gram モデル [104] を用いた.まず,形態素解析で得られた品詞列により,文頭から 1-7gram の頻度を計算する.図 2.6 に N-gram に対し頻度を求める方法を示す.各キャッチコピー,および文に対して品詞解析を行い,N-gram 目までの品詞が一致した場合加算して,N-gram の頻度を求める.

一般に N-gram においては,Nの数が増大するほど,一致する確率は下がっていく. 具体的には,順方向品詞 N-gram に関して,上位 10 件の品詞列の全体に占める割合は,N=1 のとき 99.8%(6,453 個),N=2 のとき 82.9%(5,363 個),N=3 のとき 63.8%(4123 個),N=4 のとき 40.7%(2,630 個)であり,N=7 のときは 7.14%(462 個)である.以上のことを考慮し,順方向品詞 N-gram については N=3 のときを対象とする.後述する逆方向品詞 N-gram については,キャッチコピーにおいて頻出することが予想される「体言止め」かつ「記号終わり」など,終端の特徴を調べるために N=2 のときを対象とする.

表 2.2 にキャッチコピーコーパスにおいて , 高頻度であった順方向の品詞 $3 \operatorname{gram}$ 上位 10 件の全体に占める割合を示す . また , 該当する品詞 $N \operatorname{-gram}$ が他コーパスにおいて占める割合を右側に併載する . 表よりキャッチコピーは名詞から始まるものが多い

品詞 N-gram	キャッチコピー(%)	百科事典(%)	ブログ(%)	会話文(%)
名詞,助詞,名詞	21.6	17.9	18.0	2.32
名詞,助詞,動詞	12.1	2.12	6.43	0.947
名詞,名詞,助詞	8.71	18.8	8.61	2.02
名詞,助詞,記号	5.00	3.04	2.21	1.50
名詞,助動詞,名詞	3.94	0.508	1.04	0.223
形容詞,名詞,助詞	2.74	0.254	0.584	0.0557
名詞,助詞,助詞	2.63	3.12	2.95	0.409
名詞,名詞,名詞	2.47	15.5	3.73	0.799
名詞,助詞,形容詞	2.37	0.333	1.37	0.297
連体詞,名詞,助詞	2.24	2.91	3.15	0.910

表 2.2 各コーパスの順方向品詞 3gram の頻度割合

ことがわかる. なお,名詞で始まるものは全体の約7割(4,515/6,466)を占めている. 他コーパスとの比較では,キャッチコピーコーパスでよく出現した「名詞,助詞,名詞」のパターンが,百科事典・ブログコーパスにおいても高い割合で見られることがわかる.会話文コーパスについては,特に類似するパターンは見られない.

終端からの解析も同様に行う.逆方向の解析方法を図 2.7 に示す.逆方向であるため,表中の左側が終端となっている. 逆向きに一致する品詞 N-gram を,順方向と同様に加算し頻度を算出する.表 2.3 にキャッチコピーコーパスにおいて,高頻度であった逆方向品詞 2-gram 上位 10 件の全体に占める割合と該当する品詞 N-gram が他コーパスにおいて占める割合を示す.解析が逆方向であるため,表中の左側が終端となっている. 表より,キャッチコピーコーパスにおいて記号で終わっているものが全体のほとんどを占めていることがわかる (5,107/6,466).その中でも,名詞が記号の前に来るものが多く,名詞で終わるキャッチコピーが合計で 2,383 個と全体の 4 割近く占めていることがわかる.他コーパスとの比較では,一番多かった「記号,名詞」のパターンに関して,百科事典コーパスに類似していることがわかる.ここで,キャッチコピー特有の品詞列の定量的な分析を行う.具体的には,キャッチコピーコーパスと他コーパスとの違いを明確にするために次式のような頻度比を用いる.

$$S_f = \frac{F_c/N_c}{F_a/N_a} \tag{2.1}$$

 F_c はキャッチコピーコーパスにおける品詞 N-gram の頻度, F_a は他のコーパスにお

•••	名詞,	助詞,	名詞,	名詞,	記号
		接頭詞,	名詞,	名詞,	記号
• • •	名詞,	名詞,	助詞,	名詞,	記号
• • •	助詞,	名詞,	名詞,	名詞,	記号
• • •	記号,	名詞,	名詞,	名詞,	記号
		助詞,	動詞,	動詞,	記号
•••	名詞,	名詞,	名詞,	名詞,	記号
•••	名詞,	記号,	名詞,	名詞,	記号
:	:	•	:	:	:
	5gram	4gram	3gram	2gram	1 gram

図 2.7 逆方向の品詞 N-gram 解析方法

表 2.3 各コーパスの逆方向品詞 2gram の頻度割合

品詞 N-gram	キャッチコピー (%)	百科事典(%)	ブログ (%)	会話文(%)
記号,名詞	29.1	20.7	10.2	6.65
記号,助動詞	15.7	39.2	41.5	20.2
記号,動詞	15.3	34.7	14.1	2.95
記号,助詞	13.0	0.631	12.4	46.7
名詞,名詞	4.58	0.0175	1.83	0.204
名詞,助詞	3.19	0.00877	0.686	0.204
記号,形容詞	2.71	2.07	4.09	0.501
動詞,助詞	2.66	-	0.229	0.149
記号,記号	2.20	1.39	5.74	0.316
名詞,動詞	1.62	0.00877	0.279	0.0929

キャッチコピー VS. 百科事典		キャッチコピー VS. ブログ		キャッチコピー VS. 会話文	
品詞 N-gram	頻度比 S_f	品詞 N-gram	頻度比 S_f	品詞 N-gram	頻度比 S_f
動詞,助詞,動詞	39.4	動詞,助詞,記号	14.0	形容詞,名詞,助詞	49.1
動詞,助動詞,助詞	26.5	名詞,記号,動詞	11.0	動詞,名詞,助詞	33.9
形容詞,助詞,記号	24.7	名詞,動詞,記号	10.4	名詞,助動詞,名詞	17.7
動詞,助詞,助詞	22.0	動詞,形容詞,名詞	7.92	副詞,助詞,名詞	15.4
動詞,記号,動詞	17.6	動詞,名詞,助詞	6.75	名詞,動詞,名詞	15.0

表 2.4 順方向品詞 3gram 頻度比

ける品詞 N-gram の頻度である.さらに正規化のために,キャッチコピーコーパスにおける文の数 N_c ,他コーパスにおける文の数 N_a を用いている.

式 (2.1) によって , 順方向において各コーパスと比較を行った際の頻度比 S_f を表 2.4にまとめる.まず,表の左側の百科事典との比較では,名詞が支配的なのではなく,動 詞・形容詞から始まるところがキャッチコピーの特徴であることがわかる.表中央の ブログコーパスとの比較では「動詞」助詞」記号」の品詞列の組み合わせが最も頻度 比において高かった.このような例としては「贈っ $(\leftarrow$ 動詞)て $(\leftarrow$ 助詞), $(\leftarrow$ 記号)も らって,大人気」というキャッチコピーが挙げられる.2番目,3番目についても文頭 に近い場所で記号が現れている.従って,キャッチコピーはブログと異なり,端的で歯 切れのよい構造を持つ傾向が強いことがわかる、表の右側の会話文コーパスとの比較 では左側,中央と比べると,右側は全般的に上位の値が高いことがわかる.特に「形 容詞、名詞、助詞」の構造を頻繁に用いられることが会話文との大きな違いであると 言える.これは、会話文では対象となる名詞は文脈上共有できているが、キャッチコ ピーは対象の名詞についてどのようなものか知らせる必要がある、という違いがある からと考えられる.また,キャッチコピーの文頭から $3 \operatorname{gram}$ 分の品詞列が全体として どのような傾向を持っているのか,表2.2及び頻度スコア式を基に計算を行った.表 2.2 において全体の5 割以上(3,318/6,466) を占める上位5 位までの品詞列は,百科事 典コーパスの42.4% (4,839/11,406), ブログコーパスの36.3% (1,429/3,947), 会話文 コーパスの 7.02% (378/5,348) で用いられている. 従って, 文の前半において, キャッ チコピーの品詞列は百科事典及びブログのものに近いことがわかる.

次に,逆方向の頻度比の分析を行う.表 2.5 に逆方向において各コーパスと比較を行った際のスコア S_f を示す.まず,表の左側の百科事典との比較では,名詞終わりの体言止めが多いことがわかる.この例としては「ひと口で幸せな気分になれる,グラスの中の (\leftarrow 助詞) パラダイス (\leftarrow 名詞)」のようなキャッチコピーが挙げられる.百科事典は基本的に説明文であり「である」のような助動詞で終わることが多く,体言止めが多く見られるキャッチコピーコーパスとの大きな違いとなったと考えられる.表

キャッチコピー VS. 百科事典		キャッチコピー VS. ブログ		キャッチコピー VS. 会話文	
品詞 N-gram	頻度比 S_f	品詞 N-gram	頻度比 S_f	品詞 N-gram	頻度比 S_f
名詞,助詞	363	名詞,形容詞	23.7	動詞,名詞	24.2
名詞,名詞	261	動詞,助詞	11.6	名詞,名詞	22.4
名詞,動詞	185	名詞,副詞	6.09	動詞,助詞	17.9
名詞,助動詞	106	名詞,動詞	5.81	名詞,動詞	17.5
記号,副詞	28.2	助詞,名詞	5.48	名詞,形容詞	16.2

表 2.5 逆方向品詞 3gram 頻度比

中央のブログコーパスと比較では、体言止めが多いことの他に、助詞終わりも特徴として考えられることが分かる。これはキャッチコピーが「ようこそ日本の、ファーストクラスへ」などのようにイメージ形成のために「へ」や「に」などを多用しているからと考えられる。表の右側の会話文コーパスとの比較では「動詞、名詞」(語順は「名詞、動詞」)のパターンが多く見られる。この例として「本気モードをクールに刺激する」がある。日本語における会話において「、みたいな」などの曖昧な表現が多い一方で、キャッチコピーにおいては終端で歯切れのよい表現が多いためと考えられる。以上より、キャッチコピーにおける前半の品詞列は「名詞、助詞、名詞」などのパターンが多く、百科事典やブログの構造に似ていると言える。さらに、後半の品詞列は名詞で終わることが多く、体言止めが多用されている傾向があることがわかる。

2.3 キャッチコピーコーパスにおける特徴

ここでは,キャッチコピーコーパスそのものに含まれる単語,係り受け関係,分野 ごとの名詞について分析を行い,その特徴を明らかにする.

2.3.1 キャッチコピーコーパスにおける単語

キャッチコピーコーパスにおいて多く用いられる名詞について検討を行う.表 2.6 に 6,466 個のキャッチコピーから抽出された,高頻度で現れた名詞・固有名詞を示す.表より,名詞については「人」「あなた」「私」「心」などの人を意識したメッセージ性が強い単語が高頻度で現れていることがわかる.一方,固有名詞では地名が多く見られた.2.2.2 の品詞 N-gram の分析の逆方向解析において,体言止めがキャッチコピーの一つの大きな特徴であることが分かった.表 2.7 において具体的な例,文末に名詞が来る体言止めのケースを示す.表から「スタイル」「暮らし」など,実際に行動を起こした後の未来をイメージさせるような名詞が多用されていることがわかる.

名詞	頻度	固有名詞	頻度
人	217	日本	49
肌	197	東京	16
こと	191	イタリア	11
あなた	183	ニッポン	10
夏	173	パリ	9
私	151	ドコモ	7
心	144	銀座	7
自分	122	九州	6
日	119	ハワイ	6
世界	108	フランス	6

表 2.6 高頻度で現れた名詞・固有名詞 (上位 10件)

2.3.2 キャッチコピーコーパスにおける係り受け関係

前後からの N-gram を用いた品詞解析や含まれる単語解析の他に,係る単語,係られる単語を特定することも,キャッチコピーの特性を知る上で重要である.ここでは,構文解析器 CaboCha [160] を用いてキャッチコピーコーパスにおける係り受け関係を分析する.表 2.8 に係る・係られる側で高頻度で現れた単語を示す.表より,係る側として「この」,「その」,「もっと」のような強調する役割を果たすと考えられる指示語・副詞「あなたの」のような関係性を強調するもの「新しい」、「いい」,「美しい」などのポジティブなもの「夏」や「大人の」などの特定のイメージを喚起する単語が高頻度で現れていることがわかる.係られる側では「ある」、「なる」などの係る側を現実にする単語が多く現れていることがわかる.

2.3.3 キャッチコピーコーパスにおける分野ごとの特徴

ここでは,キャッチコピーコーパスにおける分野ごとの特徴の違いを検討する.具体的には,分野ごとに使用される名詞の頻度を取得する.表 2.9,表 2.10 に食品分野において高頻度で出現した名詞を示す.

まず食品分野では人々の食生活において,表現に欠かせない「味」や「人」などが頻繁に利用されていることがわかる.美容分野に関しては,表 2.6 に示したキャッチコピーコーパス全体においても,非常に高頻度で「肌」が現れていることがわかる. 具体的に「肌」と表現することで対象を明確にして「美」などの目標を提示することによって,購買意欲をかき立てていることが推察される.生活用品分野においては「肌」

名詞	頻度
スタイル	30
暮らし	22
時間	21
夏	19
空間	19
デザイン	16
キッチン	16
幸せ	15
贅沢	14
旅	13

表 2.7 高頻度で現れた体言止めにおける単語 (上位 10件)

のような,非常に高頻度で現れる名詞は存在しなかった.一方で「自分」「人」「私」のような単語が頻度に関して上位になっている.これは人の内面を捉えることと共に,高頻度だった「部屋」などの単語を取り入れることで,消費者に対して気づきを与えているためと考えられる.ファッション分野では「夏」「日」といった状況に関する単語「私」のような内面に関する単語「服」「スタイル」といったイメージの湧く単語がよく現れた.与えられた状況に関して,対象となる消費者の心を捉えて,行動に移すことを意図して作られていることが考えられる.デジタル関連製品分野においては,ケータイやカメラなどの具体的なモノと「あなた」や「世界」といった外界との繋がりを強調するような名詞が選ばれていることがわかる.最後に,ハウジング分野においては対象を指すときに必要な「空間」という単語,住まう上で欠かせない「キッチン」といった名詞がよく現れていることがわかる.

以上のことから,分野ごとに高頻度で現れる名詞が違うことがわかり,特に目的が明白である美容などの分野に関しては,特定の名詞がよく現れる傾向が見出された.

2.4 統計的手法を用いたキャッチコピー自動生成システム

以上得られた知見を基に,ここでは統計的手法を用いたキャッチコピー自動生成システムを提案する.2.5では評価実験を述べる.2.6では実験結果について考察し,2.7をまとめとする

図 2.8 に提案システム全体の流れを示す.まず始めに,キャッチコピーのテーマ (例: 食品) およびそのキーワード (例: ジュース,かき氷) が入力される.次に,キーワー

係る側	頻度	係られる側	頻度
この	143	ある	222
その	108	なる	156
もっと	105	1111	102
あなたの	94	ない	78
ある	92	ために	72
新しい	74	こと	62
L1L1	74	する	46
夏の	61	夏の	44
美しい	58	あなたの	40
大人の	47	あります	34

表 2.8 係る・係られる側で高頻度で現れた単語 (上位 10件)

表 2.9 各分野における高頻度で現れた名詞 1

食品分		美容分野 総数: 451 個		生活月	月品分野 総数: 1,133 個
単語	頻度	単語	頻度	単語	頻度
味	33	肌	173	自分	39
人	30	感	31	こと	31
心	23	美	27	人	31
夏	19	夏	23	私	30
幸せ	14	白	20	部屋	29

ドを基に,キャッチコピーコーパス[155]におけるテーマに特有の関連語(例:果汁,風物詩,夏,名物等)が選択される.3番目に,候補はテーマにどれくらい近いかの基準,キャッチコピーの文法的基準,相互情報量の基準によって選択される.最後に,それらのスコアが高いものが最終的に出力される.

ここで用いられるキャッチコピーコーパスは,食品や美容,生活用品を始めとする 12 の分野を包含しており,全体で 6,466 個のキャッチコピーが掲載されている.この キャッチコピーコーパスを分析して明らかになったテーマ特有の単語およびキャッチコピー特有の文構造を以降で用いる.

	役 2.10 日月到120月 3日頃反 55177 日間 2				
食品分野	総数: 495 個	美容分野	総数: 451個	生活用品分	分野 総数 : 1,133 個
単語	頻度	単語	頻度	単語	頻度
夏	63	ケータイ	33	空間	36
日	36	写真	31	キッチン	33
私	35	あなた	26	暮らし	25
服	33	世界	25	窓	25
スタイル	33	カメラ	24	家	21

表 2.10 各分野における高頻度で現れた名詞 2

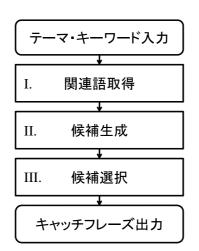


図 2.8 システム全体の流れ

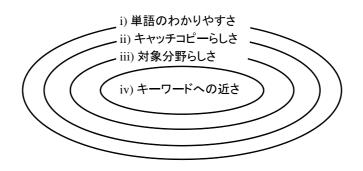


図 2.9 キャッチコピーにおける単語の包含関係

2.4.1 分野特有の関連語取得 (I.)

テーマ特有でキーワードを考慮したキャッチコピーを生成するために,テーマおよびキーワードが入力される.ここでは,キャッチコピーにおける単語の包含関係を考慮する.図 2.9 に,キャッチコピーにおける単語の包含関係を示す.図は,i)単語のわかりやすさ,ii)キャッチコピーらしさ,iii)対象分野らしさ,iv)キーワードへの近さが示されており,キーワードへの近さを考慮するため,i)単語のわかりやすさ,ii)キャッチコピーらしさ,iii)対象分野らしさの条件を満たす単語取得を行う.

理由として,キャッチコピーは消費者の購買欲を喚起することを目的としている.従って,目的とする対象の商品の説明をする前に,キャッチコピーは食品や生活用品など特定の分野に属しているかどうかを知らせる必要性があると考えられる.キャッチコピーの根幹となる単語はこれらの基準を満たすことが良いと考えられる.

i) わかりやすさ

わかりやすさは、キャッチコピーそのものを理解してもらうために必要な条件である。キャッチコピーが一般的に知られていない単語を含んでいる場合、理解を妨げる可能性があるため、除去する必要性がある。Web 日本語 Nグラムコーパスにおける 1gram は、形態素列における一要素のみ含まれており、日本語で用いられているほとんどの単語をカバーしている。また、Web 上における頻度が併記されており、それぞれ単語の使用頻度が示されている。ここで、分かりやすく親しみやすい単語は、一般に多く Web 上に出現し、使用頻度が高いと考えられる。従って、Web 日本語 Nグラムコーパスの 1gram において頻度が閾値以下のものは、候補として除外する。

ii) キャッチコピーらしさ

通常の文章で用いられる単語とキャッチコピーにおいて使用される単語は異なることが考えられ、Web上における単語の頻度とキャッチコピーにおける頻度の違いを利用し、よりキャッチコピーに高頻出な単語を選択することを考える.つ

まり,ある単語 W_k のキャッチコピーコーパスにおける頻度 $F_c(W_k)$ と Web 日本語 N グラムコーパスにおける頻度 $F_g(W_k)$ の違いに着目し,以下のような式を用いて,キャッチコピーらしい単語を求めるスコア S_c を計算する.

$$S_c = \frac{\left(\frac{F_c(W_k)}{\frac{1}{n}\sum_{k=1}^n F_c(W_k)}\right)^{c1}}{\left(\frac{F_g(W_k)}{\frac{1}{n}\sum_{k=1}^n F_g(W_k)}\right)^{c2}}$$
(2.2)

ここで, $\frac{1}{n}\sum_{k=1}^nF_c(W_k)$ はキャッチコピーコーパスにおける正規化のための平均値, $\frac{1}{n}\sum_{k=1}^nF_g(W_k)$ は正規化のための平均値,c1,c2 は頻度の違いを明確にするための累乗のパラメータである.

iii) 対象分野らしさ

同様に、キャッチコピーコーパスにおいても、各分野ごとに中心となる単語が異なることが考えられる。利用するキャッチコピーコーパス [155] は複数のテーマのキャッチコピーを独立に収録しており、それぞれの分野のキャッチコピーにおける単語の使われ方は、テーマごとの単語頻度で表現されていると考えられる。従って、ii)と同様に、キャッチコピーコーパス全体とキャッチコピーコーパスにおける各分野ごとの頻度の違いを利用し、よりキャッチコピーの各分野に高頻出な単語選択を行う。ある単語 W_k のキャッチコピーコーパスの各分野における頻度 $F_c(W_k)$ とキャッチコピーコーパス全体における頻度 $F_c(W_k)$ の違いに着目し、キャッチコピーコーパスにおける各分野に相応しいスコア S_t を計算する。

$$S_{t} = \frac{\left(\frac{F_{t}(W_{k})}{\frac{1}{n}\sum_{k=1}^{n}F_{t}(W_{k})}\right)^{c3}}{\left(\frac{F_{c}(W_{k})}{\frac{1}{n}\sum_{k=1}^{n}F_{c}(W_{k})}\right)^{c4}}$$
(2.3)

ここで, $\frac{1}{n}\sum_{k=1}^n F_t(W_k)$ はキャッチコピーコーパスにおける正規化のための平均値, $\frac{1}{n}\sum_{k=1}^n F_c(W_k)$ は正規化のための平均値,c3,c4 は頻度の違いを明確にするための累乗のパラメータである.

これらのスコアに関して S_f と S_t の線形和を計算し,一定値以上のものをキャッチコピーの各テーマに相応しい単語として次のステップ iv) に渡す.

iv) キーワードへの近さ

入力されたキーワードに対して,より関連性のある単語を抽出すると良いと考えられる.そのための理論背景,および具体的方法を述べる.

関連語取得の理論的背景 [106]

関連語を抽出する際に、従来は相互情報量が用いられてきた.しかしながら、

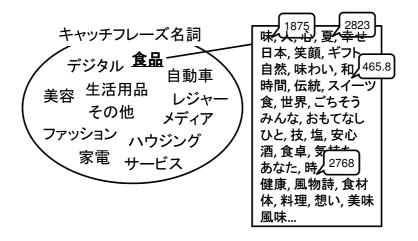


図 2.10 関連語取得の流れ

相互情報量を利用した場合,頻度に偏りがあるときは良好な結果が得られないことが分かっている.従って,提案システムにおいては, χ^2 値を用いることで関連語の抽出を行う. χ^2 値では,語群を構成する単語の出現頻度を正規化要素とし,値の正規化を行った上で共起の偏りを算出するので,出現確率のばらつきによる影響を抑えることができる.このため,値のばらつきが大きい Web 日本語 N グラムコーパス [94] の共起頻度を用いて関連語を算出する場合, χ^2 値を計算指標として用いることがより適切であると考えられる.語 w_i と語 w_j の実際の共起頻度を $n(w_i,w_j)$,語 w_i と語群 G の語との共起頻度の和を $S_{wi}=\sum_k n(w_i,w_k)$,全ての共起頻度の和を $S_G=\sum_{w_i\in G} S_{wi}$ とするとき,語 w_i と語 w_j に関する χ^2 は次式で表される.

$$\chi^{2}(w_{i}, w_{j}) = \frac{n(w_{i}, w_{j}) - E(w_{i}, w_{j})}{E(w_{i}, w_{j})}$$
(2.4)

$$E(w_i, w_j) = S_{wi} \times \frac{S_{wj}}{S_G} \tag{2.5}$$

この関連度計算のアルゴリズムを用い,各テーマにおいてキーワードに関して関連語を取得する.

関連語取得の具体的方法

図 2.10 に関連語取得の流れを示す.図に示すように,選択されたテーマに関して,テーマ特有の名詞が存在している.それらの名詞に関して,式 (2.4) および式 (2.5) を適用することにより,それぞれのテーマとキーワードに対して関連する単語の取得を行う.

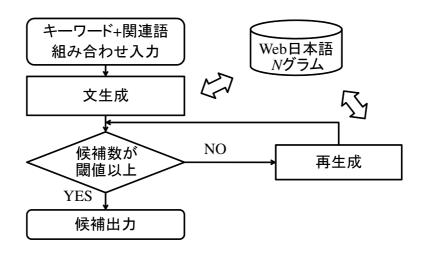


図 2.11 候補生成の流れ

2.4.2 キャッチコピー候補生成 (II.)

図 2.11 にキャッチコピーの候補生成の流れを示す.図に示すように,キーワードと関連語がまず最初に入力される.それらの複数の単語を含む文を 2,550 億単語のデータ量を持つ Web 日本語 N グラムコーパスから抽出する.なお,高速なアクセスが必要となるため検索ツールである ssgnc [161] を利用している.抽出後,候補数が閾値以上であるならば候補を出力する.もし閾値以下であるならば以前に生成した文の一部および関連語を含む文を再抽出し,閾値を超えるまで再生成を行う.

2.4.3 キャッチコピー候補選択 (III.)

キャッチコピーの選択を行う際,以下の3つを指標を用いる.

- a) χ^2 値 (関連度) によるスコア
- b) 文構造スコア
- c) 斬新な名詞の組み合わせスコア

a) χ^2 値 (関連度) によるスコア

キャッチコピー候補において,関連語が含まれていた場合にその値をその関連度を 加算することで,それぞれの候補の関連度スコアを算出する.

b)-i キャッチコピーコーパスにおける文構造スコア

キャッチコピーにおいて,より出現頻度が大きいほど,キャッチコピーに相応しい文構造とみなすことが出来る.従って,ある品詞列Lがキャッチコピーコーパスにおいて現れる確率 $P_c(L)$ は,F(L)を頻度,Nをキャッチコピーコーパスの文の数としたとき,

$$P_c(L) = \frac{F(L)}{N} \tag{2.6}$$

で表される.

この確率に関して,情報量を用いてスコア計算を行う.情報量の定義式に対して,式 (2.6) を代入し,N-gram 分 $(N_{length}=7)$ の合計スコア $S_{\rm I}$ を

$$S_{\rm I} = -\sum_{k=1}^{N_{length}} \log \frac{F(L)}{N} \tag{2.7}$$

のように算出する.

この処理をキャッチコピー候補を品詞解析した後,前方と後方の両方からスコア算出を行う.

b)-ii 複数コーパスを用いた文構造スコア

他コーパスと比較したときに,キャッチコピーコーパス特有に含まれている品詞列を優先的に抽出することも,高品質なキャッチコピー選択において重要と考えられる.よって,ある品詞列Lが存在しているときに,それがキャッチコピーコーパスにどれくらいの確率で含まれているかを計算する必要がある.しかしながら,複数コーパスにおいて,必ずしも品詞列の数が同じとは限らない.従って,これを考慮するため各コーパスの品詞列の数を正規化した確率 $P_v(L)$ を考える.それぞれのコーパスをC,特に対象となるコーパス(ここではキャッチコピーコーパス)を C_k として,

$$P_v(L) = \frac{P(L|C_k)}{\sum_C P(L|C)}$$
(2.8)

と表す.式 (2.8) に関して,それぞれキャッチコピーコーパス,百科事典コーパス [156], ブログコーパス [157], 会話文コーパス [158] を適用し,

$$\frac{P(L|C_k)}{\sum_{C} P(L|C)} \approx \frac{F_{C_k}(L)/N_{C_k}}{\sum_{C_x \in corpora} F_{C_x}(L)/N_{C_x}}$$
(2.9)

とする.従って,求める複数コーパスを用いた文構造スコア $S_{\rm II}$ は,情報量に関して N- ${
m gram}$ 分 $(N_{length}=7)$ を取得して,

$$S_{\rm II} = -\sum_{k=1}^{N_{length}} \log \frac{F_{C_k}(L)/N_{C_k}}{\sum_{C_x \in corpora} F_{C_x}(L)/N_{C_x}}$$
(2.10)

となる. なお, いずれのコーパスにも文構造が含まれていない場合は, 1 件存在する ものとみなす.

キャッチコピーコーパス単体の時と同様に,前方と後方の両方からスコア算出を行う.

c) 斬新な名詞の組み合わせスコア [146-148]

斬新な名詞の組み合わせスコアにおいては,相互情報量 (PMI) を用いてキャッチコピーらしさを考慮する.西原ら [146-148] は相互情報量を研究タイトルの評価指標に用いている (相互情報量の定義が分母分子が反転しているが,この式を本研究の相互情報量として計算を行う).提案システムにおいても,相互情報量を斬新な名詞の組み合わせスコアとして取り入れる.相互情報量 (PMI) は,H() を検索エンジンにおけるヒット件数の頻度, w_a と w_b をそれぞれの名詞,ALL を検索エンジン上の全てのページ数とした時に,

$$PMI = \frac{H(w_a) \times H(w_b)}{H(w_a \wedge w_b) \times H(ALL)}$$
(2.11)

で表される . ここで , H(ALL) は 10^{10} として計算を行う . 提案システムでは , ssgnc [161] を用いて Web 日本語 Nグラムコーパスから抽出された N-gram 頻度を取得する . そのため , それぞれの単語の頻度 ($H(w_a)$, $H(w_b)$ or $H(w_a \wedge w_b)$) は ,

$$F_w = \sum_{N=1}^{7} F_{Nw_x} \tag{2.12}$$

で計算される.ここで, F_{Nw_x} はコーパスにおける, $word_x$ または $word_y$ の $word_z$ の組み合わせの頻度である.

斬新な名詞の組み合わせスコア S_n は,Wを候補における単語の集合, n_i を単語数とした時,組み合わせを考慮して

$$S_n = \frac{1}{n_i^2 - n_i} \sum_{w_a, w_b \in W; w_a \neq w_b} PMI(w_a, w_b)$$
 (2.13)

で表される.

提案システムは,1. 文構造によるスコアリングの後 (閾値以下は除去), $2. \chi^2$ 値によるスコアリングでソートを行っている.その後最終的に,3. 単語の組み合わせの相互情報量が一定範囲内のものを出力する.ここで3. に関して,優先順位は設けていないため, χ^2 値によるスコアリングでソートされた順に出力が得られる.

2.5 実験

ここでは,生成されたキャッチコピーに関して主観評価実験を行い,質を検証する. 具体的には,提案システム生成のものとプロ作成のもの,Web日本語 N グラムコーパスよりキーワード指定して得られたベースラインのものに関して主観評価実験を行う.

2.5.1 実験概要

それぞれのテーマに関して,キーワードを入力する「システム生成のキャッチコピー (提案システム)」と「キャッチコピーコーパスからランダムに選択したもの (プロ)」,「Web 日本語 N グラムコーパスに対して二つのキーワード単語を入力し,得られた複数の文からランダムに抽出したもの (ベースライン)」について比較を行う.

実験項目

- 1. 適切な文か (3 段階評価)
 - 3. 文として成り立っている 2. 意味はわかる 1. 意味不明
- 2. テーマに関して適しているか (3 段階評価)
 - 3. そのテーマで通用する 2. 拡大解釈すれば通用しそう 1. 難しい
- 3. キーワードに対して適当な文か (3 段階評価)
 - 3. そのテーマで通用する 2. 拡大解釈すれば通用しそう 1. 難しい
- 4. キャッチコピーとしての質(総合評価)
 - 5. 良い 4. やや良い 3. どちらでもない 2. やや悪い 1. 悪い

2.5.2 実験条件

評価対象:

3 テーマ 2 キーワードから生成されたキャッチコピー 60 個 , キャッチコピーコーパスから取得したプロ作成のものと Web 日本語 N グラムコーパスから取得した合計 120 個をランダムに混合し , 比較する . 被験者は 20 代の大学生および大学院 13 名である .

	文の適切さ	テーマ	キーワード	総合評価
提案システム	2.62	2.68	2.42	2.85
ベースライン	1.72	2.49	2.56	2.01
プロ	2.91	2.55	1.52	3.73

表 2.11 食品分野の主観評価実験結果

2.5.3 生成例

提案システムが生成したキャッチコピーを以下に示す.

テーマ: 食品

キーワード: ジュース, かき氷

夏のかき氷は懐かしい味。

ジュースとヨーグルトの優しいハーモニー

テーマ: 交通・レジャー

キーワード: 温泉, リラックス

温泉で心安らぐ森の休日

リゾートの風を感じてリラックス

テーマ: メディア

キーワード:映画,感動

涙と笑いの青春映画。

最高の感動がここにある

2.5.4 実験結果

提案システム生成のもの,ベースラインのもの,プロ作成のものを混合し,4つの項目 (1. 適切な文か 2. テーマとして適しているか 3. キーワードに対して適当な文か 4. 総合評価) について主観評価実験を行い,その結果を示す.

それぞれのテーマ

表 2.11 に食品分野に関する評価値を示す.表より提案システムは文法的に適切な文を生成している傾向がベースラインよりも強いことがわかる.テーマにおいては,プロのものよりもわかりやすいという結果となっている.キーワードについては,キーワードをそのまま含む文を抽出したベースラインに肉薄している.総合評価については,ベースラインを上回る結果となっている.

	文の適切さ	テーマ	キーワード	総合評価
提案システム	2.46	2.55	2.60	2.86
ベースライン	2.03	2.58	2.71	2.61
プロ	2.88	2.66	1.98	3.77

表 2.12 交通・レジャー分野の主観評価実験結果

表 2.13 メディア分野の主観評価実験結果

	文の適切さ	テーマ	キーワード	総合評価
提案システム	2.56	2.53	2.53	3.11
ベースライン	1.92	2.39	2.55	2.19
プロ	2.94	2.48	2.23	4.13

表 2.12 に交通・レジャー分野に関する評価値を示す.表より提案システムは文法的に適切な文を生成している傾向がベースラインよりもやや強いことがわかる.テーマにおいては,ベースラインのものよりもややわかりにくいという結果となっている.キーワードについては,キーワードをそのまま含む文を抽出したベースラインに接近している.総合評価については,ベースラインをやや上回る結果となっている.

表 2.13 にメディア分野に関する評価値を示す.表より提案システムは文法的に適切な文を生成している傾向がベースラインよりも強いことがわかる.テーマにおいては,プロのものよりもわかりやすいという結果となっている.キーワードについては,キーワードをそのまま含む文を抽出したベースラインに肉薄している.総合評価については,ベースラインを大幅に上回る結果となっている.

全テーマ

表 2.14 に食品,交通・レジャー,メディアに関する評価値の平均を示す.

表より,システム生成のキャッチコピーはベースラインよりも文として成立し,テーマに関しても有意水準5%でベースラインを上回っている.キーワードについては,プ

表 2.14 全テー	マの王観計	<u>"</u> 伽美	疑結 5	果
\ _ \ - \ - \-				-

	文の適切さ	テーマ	キーワード	総合評価
提案システム	2.55	2.59	2.51	2.94
ベースライン	1.89	2.49	2.61	2.27
プロ	2.91	2.56	1.91	3.88

	文の適切さ	テーマ	キーワード	総合評価
提案システム	0.159	0.0630	0.0645	0.423
ベースライン	0.355	0.0650	0.0460	0.475
プロ	0.0137	0.247	0.199	0.219

表 2.15 全テーマの主観評価実験結果の分散

表 2.16 評価値の割合

	評価値 4 (%)	評価値 5 (%)
提案システム	22.3	7.05
ベースライン	12.4	1.67
プロ	36.7	31.3

口のものよりも分かりやすく、総合評価においてはベースラインよりも有意水準 5%で ベースラインを上回っていることが示された.

表 2.15 に食品,交通・レジャー,メディアに関する評価値の平均の分散を示す.

文の適切さに対しては、プロ作成のものが一番分散が小さく安定した結果が得られ ている、システム生成のものも、ベースラインよりも評価値が安定していることがわ かる.テーマやキーワードに関しては、システム生成やベースラインの方がプロ作成 のものよりも分散が小さいことがわかる.総合評価については,システム生成の方が ベースラインよりも分散が小さいものの、プロ作成のものよりも大きく様々な質のも のが生成されていることがわかる.

表 2.16 にシステム生成,ベースライン,プロ作成のものに関する評価値を示す.表 に示すように,評価値4の「やや良い」に関してシステム生成のものはプロ作成のも のとベースラインの中間程度となっている.評価値5の「良い」に関しては,ベース ラインに対して大幅に上回っている.

2.6 考察

2.6.1 それぞれのテーマ

システム生成のキャッチコピーの総合評価において、メディア分野の時に一番評価 値が高かった.生成例としても「涙と笑いの青春映画。」「最高の感動がここにある」 などわかりやすく、人の心に訴えかけやすいと思われるキャッチコピーが出力されて いる.一般にメディアでの映画関係の宣伝が多く,一般人がブログ等に書く機会が多

かったため,キャッチコピーとしても理解可能で高品質なN-gramの文字列が比較的多かったからと考えられる.

2.6.2 全テーマ

適切な文かという指標に関して,生成システムはプロよりも劣るものの,安定してベースラインを上回っている.これは,各テーマごとの文に関しても,文構造スコアが有効に機能し,文法が正しいものが選ばれたからと考えられる.

テーマに関して適しているかについては,システム生成,ベースライン,プロ作成のものについて値が近い場合が多かった.これはプロについてはテーマごとにまとめられているので,テーマらしく,ベースラインについては,キーワード生成なので必然的にテーマに近くなる一方で,システム生成に関しては各テーマについて取得した χ^2 値による関連語がテーマらしさの評価の時に有効に働いているからと考えられる.キーワードに関しては,提案システム及びベースラインともにプロ作成のものよりも良好な結果が得られた.提案システムがベースラインに肉薄した理由として χ^2 値による関連語取得がキーワードに対して適切という指標に対して有効に働いたと考えられる.

総合評価に関しては,システム生成が有意にベースラインを上回っている.これは,「適切な文」という指標をクリアした上で,キーワードや関連語を含む文を出力できているからであると考えられる.評価値4の割合については,システム生成がプロに対して6割ほどであり,大きな差はないと思われる.しかしながら,評価値5に関しては,システム生成の占める割合の4倍以上がプロ作成のものであり,特に心に響くような高品質なキャッチコピーの生成を行うには更なる改善が必要なことが示唆される.

また,キーワードを含むのみの文であるベースラインに関しては,提案システムと同様の言語資源を用いているが,質の点で課題がある.例えば,キャッチコピーに極めて近いもの(キーワード:温泉、リラックス 「週末リラックス出来る至宝の温泉宿」)も出現しているが,全体としてクオリティが低いと考えられるため,比較対象としての更なる検討を行う必要性があると考えられる.

2.7 まとめ

ここでは,定量的分析に基づいたキャッチコピーの自動生成システムを提案した. キャッチコピーコーパスを品詞 N-gram および単語使用の観点から分析を行う.分析により明らかになった特徴に基づいて,キャッチコピーの自動生成において,大規模な 言語資源である Web 日本語 N グラムコーパスを利用してキャッチコピーを生成する.ユーザのキーワードに基づいて,各テーマごとに特有の名詞を χ^2 値を用いた関連度の指標に基づいて重み付けを行い,それを含む候補を大量に取得する.候補に関して, χ^2 値を用いた関連度のスコア,キャッチコピーの文法構造に基づくスコアおよび単語の組み合わせの斬新さに基づくスコアを用いてソートを行い,キャッチコピーを取得する.評価実験により,提案システムはテーマとキーワードに対して,適切なキャッチコピーを生成可能であることが示唆された.

第3章

Web 上の文情報を用いたキャッチコピー自動生成 システム

本章では、Web 上の文情報を用いたキャッチコピー自動生成システムの提案を行う. キャッチコピーには、直接的に対象の特徴を表現せず、間接的な描写で魅力を伝えているものが数多く存在する.このようなキャッチコピーを生成するため、以下の二つのアプローチを行う.

- (1) Web を用いた文に関する知識獲得
- (2) 様々な表現を目的とした,膨大な量のキャッチコピーを含むコーパスの利用

それぞれの意図として, Web を用いることにより, 人手作成のコーパスには含まれていない知識取得を行うこと, また新たに拡充した2万個を超える最大級のキャッチコピーコーパスを利用することで, より洗練された伝え方をする様々なキャッチコピーの表現を取得することを目指す.

主観評価実験によって,システム生成のキャッチコピーの質の検証を行った結果,生 成されたキャッチコピーの一部はプロ作成のものに匹敵する評価が確認されている.

3.1 背景

自然言語生成は自然言語処理においても,重要なトピックである.Web上には膨大な量の情報が蓄積されていることから,Web情報を用いたアプリケーションの需要は大きいと考えられる.また,人間に関する研究も発達してきている.脳神経科学から,行動心理学に至るまで,多くの興味深い知見が明らかになっている[162,163].工学的見地に戻ると,自然言語処理によって人間の感性を捉えることを目的とする研究も存在する[164].感情を扱う自然言語処理の研究[165,166]も活発である.広告で用いられるキャッチコピーは短い文にも関わらず,短時間で人の目を惹きつける.従って,キャッチコピー研究は自然言語処理と情緒両方に関連すると考えられる.

広告宣伝において,キャッチコピーに関して様々な分析が行われている [167,168].また,キーワードをテンプレート形式に当てはめ,キーワードを含むキャッチコピー生成システム [136,137] も存在する.文生成に関して,Banko [113] らが統計的手法を用いて文生成を行っている.

国内においても、キャッチコピーに関するシステムの提案が行われている・松平ら [144,145] は、質より量、批判厳禁を旨とする発散的発想支援の立場から、遺伝的プログラミングを用いたキャッチコピー生成システムを提案した.西原ら [146,147] は、研究発表に限定し、タイトル毎に興味を引きやすいかどうかをランク付けするアルゴリズムを提案した.中野ら [149,150] はブログ記事を用いて目を引くキャッチコピーを作るための基礎的研究を行った.これらの手法・システムの成果に統計的な視点を加えるため、前章の統計的手法によるキャッチコピー自動生成システムが提案された.このシステムにおいては、品詞列および使用単語の側面を考慮してキャッチコピー生成が行われている.しかしながら、このシステムはキーワードを含んだキャッチコピー生成が行われている.しかしながら、このシステムはキーワードを含んだキャッチコピーは成が困難であった.一方で、間接的な事実しか伝えないのにも関わらず、情動を伝えるのに効果的なキャッチコピーが存在する.このことから「伝えること」よりも「どう言うのか」に焦点を当てるべきであるといえる.

自然言語生成に関しては,伊藤ら [169] が格フレームを用いて一般的な用途の文の自動生成を行っている.間接的なキャッチコピーを生成するためには,単語の他に用いる文の構造が重要となると考えられる.第2章の分析で得られた知見から,キャッチコピーの特徴は通常の文の構造と異なっていることがわかっているため,多くのキャッチコピーを含むコーパスの利用が効果的であると考えられる.

従って,複数のキャッチコピー大百科の統合を行い,キャッチコピーの構造を取得し 利用する.

さらに、Web 情報を用いて知識源とする様々な手法も提案されている [170, 171].膨大な量の情報が Web 上に存在しており、ビッグデータの応用にも期待が高まっている [172]. Blog や口コミサイト等、オンライン上には多岐にわたる文が存在しており、新しい単語や常識等、辞書には含まれていない知識を含んでいる.一方で、このような情報はノイズを多く含んでいる.Bollegela [171] らは検索エンジンから取得可能なスニペットを用いて、単語間類似度の計算を行うアルゴリズムの提案を行った.日本語に関しても、杉木ら [109] がテンプレートを利用して、掲示板の宿情報に関する口コミを効率的に探索するための手法を提案している.

これらの背景から、Web 上の情報を用いて表現のスパース性を克服するシステムを 提案する.また、実在する間接的なキャッチコピーを模倣することにより、提案シス テムは入力に相応しいキャッチコピーの生成を行う.生成されたキャッチコピーの質を 検証するため,評価実験では,被験者によるキャッチコピーの評価を行う.

以降,まず3.2 にて,目的とするキャッチコピーについて説明し,3.3 において提案システムについて詳細を述べる.3.4 で実験について説明し,3.5 で結果について考察を行い,3.6 をまとめとする.

3.2 生成目標とするキャッチコピー

キャッチコピーには,主にブランド構築を意図する広告キャッチコピーと,商品の販売促進を促す商品キャッチコピーが存在する.どちらの種類のキャッチコピーも,ブランドあるいは商品となる対象に対して好意的なイメージを持たせることを目的としている.従って,キャッチコピーには目的とする対象に関して,情報を伝え消費者の心を捉える必要性があると考えられる.

キャッチコピー生成に関連するシステムとして,発想支援に基づくもの等が存在する [144,145]. 生成システムとしては,前章の統計的手法を用いたキャッチコピー自動生成システムを含め,キーワードを含む文を生成するものにとどまっているものが多かった.しかしながら,直接的に対象に言及するよりも,間接的に対象の魅力を述べるほうが効果的な可能性がある.例えば,丈夫である筆箱をアピールする時に「頑丈な筆箱です」という表現をするよりは「象が踏んでも壊れない」という表現にした方が,受け手となる消費者にインパクトを与えられると考えられている [173].他の例としては「いい景色が見えます」よりは「宝石のような灯火が散りばめられた夜景が見えます」の方がより効果的と考えられる.

このような表現を生成する際には,対象の知識を獲得し,こなれた日本語表現に変化させる必要があると考えられる.ここでは,2つのアプローチを利用することにより,この問題の解決を目指す.

まず第一に,多くのキャッチコピーを用いた文構造の取得である.キャッチコピー大百科 [174] を始めとする大百科から,食品・美容・生活用品等の複数のテーマにわけられた 20,000 個を超えるキャッチコピーを取得し,キャッチコピーコーパスとして用いる.様々な言い回しを用い,読み手を惹きつけるキャッチコピーの文構造を得ることで,こなれた日本語表現を目指す.図 3.1 に具体的に用いたキャッチコピーコーパスを示す.それぞれの大百科 [155,174-176] は約 6,000 個のキャッチコピーを含んでおり,食品・美容・生活用品等のテーマが存在する.合計 24,472 個 44 テーマのキャッチコピーが候補となる.

もう一つは,トピックモデルの利用である.Latent Dirichlet Allocation (LDA) はテ



図 3.1 新たに構築したキャッチコピーコーパスの基となった 4 つのキャッチコピー大百科 [155, 174-176] の表紙

キストにおける,潜在的トピックの推定等に用いられている[51].本章では,得られた潜在的トピックを間接的表現に必要な知識源と仮定することにより,キャッチコピー生成を行う.これを実現するために,LDAを利用し,Web文書の潜在的トピックのクラスタリングを行うことにより,キャッチコピー生成にふさわしい単語の獲得を目指す.

3.3 Web上の文情報を用いたキャッチコピー自動生成システム

Web 上の文情報を用いたキャッチコピー自動生成システムを提案する. Web から取得した文書に対し, LDA を適用することによりトピックを獲得し, これを知識ベースとして用いることでキャッチコピーを生成する. より洗練されたキャッチコピーを生成するため, キャッチコピーの構造をキャッチコピーコーパスから取得する.

これは、Web上の知識を用いることにより、キャッチコピーらしい間接的な言い回しを取得することを主眼としている。つまり、Web情報を用いることで「景色が美しい」や「ケーキがおいしい」などの「対象、性質」を含む述語項構造を持つ入力対に対して、膨大な量の知識を得ることによって、より効果的なキャッチコピーの生成を目指す。

多くのキャッチコピーを含むコーパスを利用することに関しては,適切なキャッチコピーの選択が考慮される.取得された知識はモデルキャッチコピーと統合され,キャッチコピー候補の生成が行われる.(単語の一部のみを入れ替えるアプローチでは,文構造自体はオリジナルなものと類似すると考えられる.しかしながら,キャッチコピーは単語の特徴が大きな性質を占めていると考えられる.例えば,みんなを幸せにする会社」はある会社を宣伝するキャッチコピーになるが,みんなを幸せにする苺」では,美味な苺が乗っている食品を宣伝していることになる.従って,異なる単語の変換を行う点においては,類似しすぎたものになることは多くないと考えられる.)

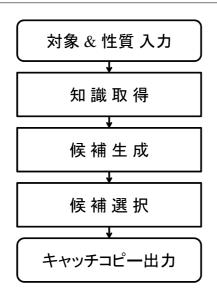


図 3.2 システム全体の流れ

これらのアプローチにより,効果的なキャッチコピーの生成を目指す.

3.3.1 提案システムの流れ

図 3.2 に提案システムの流れを示す.

まず始めに,キャッチコピーにおいて述べる対象 (例: 景色) とその性質 (例: 美しい) が入力される.その後,前処理において Web を用いて対象および性質に関する情報の取得を行い,トピックおよび単語の知識を構築する.次に,得られた知識およびキャッチコピーコーパス [155,174-176] の分析によって得られた特徴を用いてキャッチコピーの候補が生成される.これらの候補は,文の妥当性,斬新さの観点によって選択され,上位のものが出力される.

3.3.2 知識獲得

ここで,提案システムは対象と性質についての情報取得を行う.図 3.3 に知識獲得の流れを示す.まず「景色」「美しい」のような述語項構造である対が入力される.これらの対はクエリ化の後に Web に投げられ,文章取得を行う.その際,完全一致検索を行うことにより,少なくともクエリが含まれている文が一文以上存在する文書が取得されることになる.

ここで,述語項構造からクエリ化の作成方法について述べる.クエリは因果関係に 合致するように生成される.つまり,ある理由が示されている時には,その結果が存

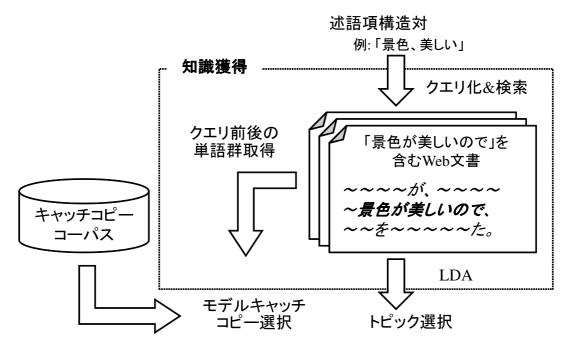


図 3.3 知識獲得の流れ

	对家	助詞	性質	助訶
クエリ1 クエリ2 クエリ3 クエリ4 クエリ5 クエリ6	景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景景	ががががががががが	美しくっ 美しい 美しかった 美しい 美しいった	ててののかかか

図 3.4 「景色」「美しい」から生成されたクエリ

在する」ことを前提にクエリ化を行う.具体的には,対象と性質に対して助詞が付加される.

図 3.4 に「景色」「美しい」から生成されたクエリの例を示す.

このような処理を行うために IPAdic [177] が用いられる.これらのクエリは,検索エンジンに投げられる.

表 3.1 に例で用いたクエリを含む文の例を示す.

図に示すように「対象」「性質」からなる因果関係を示す 6 種類のクエリが生成される、得られた Web 文書に関して , 1. クラスタリング , 2. 単語取得が行われる .

表 3.1 取得された文の例 (入力対: 「景色」、「美しい」 \rightarrow 景色が美しくて)

部屋から眺める秋の景色が美しくて一句読めそうだぜ. 移動中,助手席から常に地中海が見えていて顔をあげて 窓の外を見るたびに景色が美しくてハッとする.

雲一つない青空と海の景色が美しくて何度も車を停めて 西湖は中国の伝統的な庭園,絵,文化などを移し,

一年中の景色が美しくて,絶対見逃せない観光地です.

寝室のガラス戸に写り込んだ景色が美しくて・・一枚

大浴場の窓から見える富士山の景色が美しくて癒されます

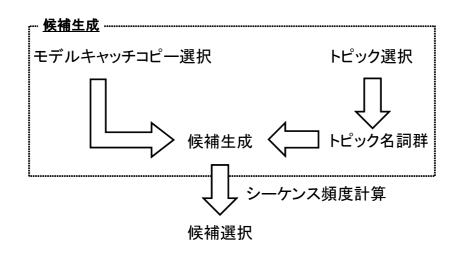


図 3.5 候補生成の流れ

3.3.3候補生成

候補生成における流れを図3.5に示す.

図に見られるように候補生成においては、モデルキャッチコピーとトピック選択か ら得られた知識の統合が行われる.

モデルキャッチコピー選択

モデルキャッチコピーには次のステップからなる.

- I. 適切なテーマ選択
- II. テーマにおけるキャッチコピー選択

I. 適切なテーマ選択

3.3.2 において,入力対に関連した文の取得が行われている.この情報を用いて,提案システムは生成にふさわしいテーマを選択する.抽出した Web 情報は入力対に関して潜在的な意味を含む場合があるため,Web 上のより多くの単語を持つものがより最適であると考えられる.

具体的には,形態素解析器 MeCab [159] を用いて,提案システムは文書からの名詞,動詞,そして形容詞の取得を行う.同様に,名詞,動詞,形容詞に関してもキャッチコピーコーパスから取得を行う.

その後,テーマは以下の式によって計算される.一般に,より対象についての情報を持っているほど,そのテーマは対象に相応しいと考えられる.従って,Tをキャッチコピーコーパスにおけるテーマ,W(T)をテーマにおける単語群, W_{WEB} を Web から取得した名詞・動詞・形容詞からなる単語群, $W_s(T)$ を W(T)と W_{WEB} の共通集合, W_x を $W_s(T)$ に含まれる単語, $F(W_x)$ を単語 W_x の頻度, $F(W_y)$ を単語 W_y の頻度とした時,それぞれのテーマのスコア S(T) は,

$$S(T) = \frac{\sum_{W_x \in W_s(T) = \{W(T) \cap W_{WEB}\}} F(W_x)}{\sum_{W_y \in W(T)} F(W_y)}$$
(3.1)

で与えられる.よって,提案システムが選択するテーマ T_f は,

$$T_f = \arg\max_{T} \left\{ \frac{\sum_{W_x \in W_s(T) = \{W(T) \cap W_{WEB}\}} F(W_x)}{\sum_{W_y \in W(T)} F(W_y)} \right\}$$
(3.2)

となる. 次にテーマの決定後,モデルキャッチコピーの選択のため,テーマ T_f におけるキャッチコピーが提案システムに入力される.

II. テーマにおけるキャッチコピー選択

テーマ選択の後, 文構造及び単語の連なりを利用するため, モデルとなるキャッ

チコピーが選択される.テーマと同様に,Web上の関連する単語がより多く含まれるキャッチコピーが最適であると考える.

具体的には,L をキャッチコピー,W(L) をキャッチコピーにおける単語群, $W_s(L)$ をキャッチコピーと Web から取得された単語群の共通集合, W_z を $W_s(L)$ における単語,F(W(L)) を W_z の頻度,N(L) を単語数とした時に,モデルキャッチコピーの相応しさ S(L) は,

$$S(L) = \frac{\sum_{W_z \in W_s(L) = \{W(L) \cap W_{WEB}\}} F(W_z)}{N(L)}$$
(3.3)

で表される.キャッチコピーの中でスコアが高かったものがモデルキャッチコピーとして選択され,次のフェーズで用いられる.

トピック名詞選択

与えられた入力対に関連するキャッチコピーを生成するために,提案システムは入力対とそれぞれのトピック名詞に対して類似度の計算を行う.

それぞれの Web 文書に対して,形態素解析器 MeCab により名詞が抽出され,クラスタリングが行われる.クラスタリングには自動的にトピック単語群の取得が可能な LDA [51] を用い,トピック数=10 として計算を行っている.取得された名詞に関して,相互情報量を用いて入力対との関連度が計算される.

具体的には,名詞群取得の後,提案システムは次の式によってトピック名詞群を決定する. W_i を入力対における名詞, W_a をそれぞれのトピック名詞群における名詞, $F(W_i,W_a)$ を二つの単語の共起とした時に,それぞれの名詞群のスコアS(N) は,

$$S(N_G) = \sum_{W_i \in input} \sum_{W_a \in N_G} \frac{F(W_i, W_a)}{F(W_i)F(W_a)}$$
(3.4)

として与えられる.

生成

候補生成における流れを図3.6にまとめる.

まず,図の下部に示すように,モデルキャッチコピーの構造が抽出される.その後,入力対に関連する単語群を用いてキャッチコピー候補が生成される.

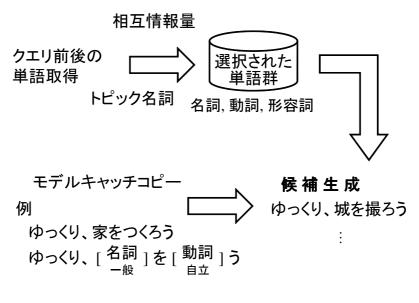


図 3.6 キャッチコピー候補生成の詳細

I. モデルキャッチコピーの構造抽出

提案システムはモデルキャッチコピーの構造を模倣し,モデルの一部を変更を することにより候補生成を行う.具体的には,以下のルールを用いる.

- 名詞・動詞及び形容詞を置換の対象とする.(MeCab が名詞として認識する ため形容動詞に関しても同様に置換の対象とする.)
- 助詞,助動詞,記号等,その他の品詞に対しては変更を行わない。

II. 生成

名詞,動詞,形容詞それぞれに関してモデルキャッチコピーと同様の活用形を用いるために IPAdic [177] を利用する.図 3.7 に入力が 対象「景色」,特徴「美しい」の場合の生成例を示す.モデルキャッチコピー「カラダの中にある」に対して「メリーランドの中に居る」等のキャッチコピー候補が生成されていることがわかる.

3.3.4 候補選択

生成の後,キャッチコピー候補の選択が行われる.まず最初に,候補は記号等が含まれているかの基準により分割される.次に,提案システムによって文としての適切さの基準により候補選択が行われる.

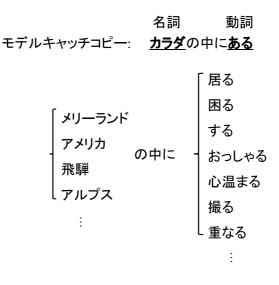


図 3.7 候補生成の例

キャッチコピー候補分割

文の整合性を推定する際,より細かな単位に分割するとより効率的であると考えられる.理由としては,(1) それぞれの単位ごとの意味的一貫性の推定が容易,(2) 文全体を考慮しないので,候補数の爆発的増加を抑えられる,からである.ここでは,分割に関連する記号が存在した時に分割することにより,意味的,文法的である文単位への分割を行う.キャッチコピーコーパスに対する予備実験の結果によって,キャッチコピーは名詞節を連結する際に記号「、」および助詞「は」「に」「を」を用いる場合が多いことがわかっている.ここでは,以下のように分割の優先順位を設けることで,分割を行う.

- 1. 「、」「は」「が」
- 2. **「を」「に」「で」**
- 3. 「から」「より」

3.3.5 文の適切度による選択

ここでは,提案システムによって意味の持たない候補の削除を行う.モデルキャッチコピーが決定されると,その構造を利用して品詞が合致するものに対して単語の入れ替えが行われ,候補が生成される.従って,文法的に正しい場合でも無意味の候補が存在する.ここで,2,550 億単語から構築された [95] Web 日本語 N グラムコーパス [94] に含まれる文の連なりを意味的に一貫性のあるものとして扱う.妥当性のない不自然な候補はこのコーパスにアクセスし,存在確認を行うことにより削除を行う.

さらに、西原らによって提案された相互情報量 (PMI) [146-148] と類似した手法によって選択を行う、最低限の文としての妥当性を考慮した後は、斬新な文の取得を行うため、フレーズに含まれる単語の頻度がそれらの単語の連なりであるフレーズの頻度よりも相対的に大きい時に、差別化を図るように式の設定を行う、ここでは、この処理は短い文に対して行うため、提案システムは名詞、動詞、形容詞に関して同列に扱う、これらの単語列のスコアが加算することにより、候補のフレーズのスコア算出が行われる、具体的には、P をそれぞれのフレーズ, P_E をフレーズより細分化された単位 $(N\text{-}\mathrm{gram})$, $W(P_E)$ をフレーズに対する細分化された単位における単語群, $W_p(P_E)$ を $W(P_E)$ と名詞・動詞・形容詞群との共通集合、 $F(W_k)$ を Web 日本語 N グラムにおける細分化された単位の頻度とした時,候補におけるフレーズのスコア S(P) は,

$$S(P) = \sum_{P_E \in P} \frac{\prod_{W_k \in W_p(P_E) = \{W(P_E) \cap \{W_n, W_v, W_a\}\}} F(W_k)}{F(P_E)}$$
(3.5)

で与えられる.この処理は,候補となるフレーズが多い時に行われ,出力が得られる. スコアの高いものをより斬新な候補として選択する.

提案システムにおいては,数が少ない時トピック名詞との共起頻度が高い単語順に生成が行われるため,上位には共起頻度の高い単語が上位に集まる.一方で,キャッチコピー候補数が多い場合は,Web日本語 Nグラムを用いて相互情報量のアイデアを基に斬新な候補の選択が行われる.従って,この場合はスコアリングが行われるため,斬新なスコアが高い候補が先に出力される.

3.4 実験

ここでは,伝え方に着目したキャッチコピー自動生成システムによって生成された キャッチコピーの質の検証を行う.

3.4.1 実験条件

「対象」、「性質」からなる述語項構造の対が入力される.実験における入力は、

- I. "景色, 美しい"
- Ⅱ. "ネタ,面白い"
- III. "ケーキ,美味しい"



図 3.8 主観評価実験用インタフェース

である.生成時には組み合わせ爆発が起こり,時に数が数百万を上回ってしまうため,処理時間が現実的ではない.従って, 式 (3.5) において,提案システムは候補フレーズの数の上限を 1.000 個として処理を行う.

生成されたキャッチコピーの質を検証するため,実験を行う.図3.8 に評価実験用のインタフェースを示す.

評価するキャッチコピー

生成されたキャッチコピーの質を評価するため,チューリングテスト的な実験を行う.図に示した主観評価プログラムは,それぞれの入力対に対してシステム生成のキャッチコピー 100 個と,プロ作成のキャッチコピー 10 個をランダムに混合し,提示する.被験者はこれら合計 330 個のキャッチコピーに対して,どれがシステム生成のキャッチコピーに該当しているのか知らされずに評価を行う.なお,被験者は 20 代大学生,大学院生および社会人 11 名である.

ここで,評価実験における指標を示す.

実験項目

- 1. 理解できるか (5 段階評価)
 - 5. 理解できる 4. やや理解できる 3. どちらでもない 2. あまり理解できない 1. 理解できない
- 2. 間接的な表現か (5段階評価)
 - 5.間接的である 4.やや間接的である 3.どちらでもない 2.あまり間接的でない

生成例			
ゆっくり、楽園をつくろう	ゆっくり、写真を撮ろう		
ゆっくり、美術館をつくろう	ゆっくり、写真を知ろう		
ゆっくり、シーンをつくろう	ゆっくり、夢をつくろう		
ゆっくり、プールをつくろう	ゆっくり、城を撮ろう		
ゆっくり、水着を撮ろう	ゆっくり、道を帰ろう		
ゆっくり、あたりを走ろう	ゆっくり、霜を撮ろう		
ゆっくり、遺産をつくろう	ゆっくり、街を飾ろう		
ゆっくり、ルートを知ろう	ゆっくり、心を知ろう		

表 3.2 「ゆっくり、家をつくろう」から生成されたキャッチコピー例

1.間接的でない

- 3. キャッチコピーとしての質 (5 段階評価)
 - 5. 良い 4. やや良い 3. どちらでもない 2. やや悪い 1. 悪い

また,被験者が評価に要する時間の計測も行った.

3.4.2 生成例

生成された例を以下に示す.

I. に関しては式(3.2)によって「建築・不動産」が選択された. Web から取得された単語例としては「楽園」「珍しい」「遺産」「山頂」「飾る」等が挙げられる.

表 3.2 にモデルキャッチコピー「ゆっくり家をつくろう」から生成されたキャッチコピーを示す.

II. に関しては , 式 (3.2) によって「メディア」が選択された . Web から取得された 単語例として , 武士」「はまる」「楽しむ」「大阪」がある .

表 3.3 にモデルキャッチコピー「いつもそこに人間」から生成されたキャッチコピー を示す.

III. については,式(3.2)により「流通」がテーマとして選択された. Web から取得された単語例として「カフェ」「人気」「上品」「豆乳」「季節」などが挙げられる.

表 3.4 にモデルキャッチコピー「みんなを幸せにする会社」から生成されたキャッチコピーを示す.

	<u> </u>				
生成例					
いつもそこに武士。	いつもそこにオチ。				
いつもそこに実況。	いつもそこに辛口。				
いつもそこに曲名。	いつもそこにボロボロ。				
いつもそこにアトリエ。	いつもそこに星空。				
いつもそこに書房。	いつもそこに寸前。				
いつもそこに本編。	いつもそこにキレ。				
いつもそこに落語。	いつもそこに動機。				

表 3.3 「いつもそこに人間。」から生成されたキャッチコピー例

表 3.4 「みんなを幸せにする会社」から生成されたキャッチコピー例

生成例				
みんなを爽やかにする小倉	みんなを上品にするカラフル			
みんなを爽やかにする常連	みんなを爽やかにする相方			
みんなを爽やかにする固め	みんなを爽やかにするかぼちゃ			
みんなを上品にする苺	みんなを上品にする百貨店			
みんなを上品にするキリ	みんなを上品にするフォーク			
みんなを爽やかにする陸	みんなを静かにするおばさん			
みんなをきれいにする理	みんなをきれいにする喉			
みんなをきれいにするクリ	みんなをきれいにする豆乳			

3.4.3 実験結果

表 3.5 に入力述語項が「景色」「美しい」の時のシステム生成のキャッチコピーとモデルキャッチコピーとの比較を示す.

表 3.6 に述語項が「ネタ」「面白い」の時のシステム生成のキャッチコピーとモデルキャッチコピーとの比較を示す.

表 3.7 に述語項が「ケーキ」「おいしい」の時のシステム生成のキャッチコピーとモデルキャッチコピーとの比較を示す.

これらの表より,評価に要する時間以外の側面に関しては,システム生成のキャッチコピーに対してモデルキャッチコピーの方が良好な結果を示した.

表 3.8 に「景色」「美しい」の場合における,システム生成のキャッチコピーの全体に占める割合を示す.

表 3.9 に「話」「面白い」の場合における,システム生成のキャッチコピーの全体に 占める割合を示す.

		理解	間接的	全体	時間
		評価 (1-5)	評価 (1-5)	評価 (1-5)	(秒)
提案	平均	3.10	3.17	2.82	11.3
システム	σ	0.727	0.471	0.610	2.90
モデル	平均	4.22	3.66	3.94	10.9
キャッチコピー	σ	0.390	0.677	0.524	1.58

表 3.5 入力対が「景色」「美しい」の時の比較結果

表 3.6 入力対が「ネタ」「面白い」の時の比較結果

		理解	間接的	全体	時間
		評価 (1-5)	評価 (1-5)	評価 (1-5)	(秒)
提案	平均	2.65	2.91	2.52	10.8
システム	σ	0.701	0.395	0.535	1.96
モデル	平均	3.88	3.41	3.61	10.3
キャッチコピー	σ	0.419	0.595	0.491	1.71

表 3.10 に「ケーキ」「おいしい」の場合における,システム生成のキャッチコピー の全体に占める割合を示す.

これらの表より,提案システム生成のキャッチコピーに関しても上位に見られるも のも存在していることがわかる.

3.5 考察

Web 情報を用いて,入力対に関する知識抽出を行うことにより,様々な単語の取得 が可能にであることがわかる.また,提案システムは入力対に関する隠れた情報を取 得することが可能であることが示唆されている.

		理解	間接的	全体	時間
		評価 (1-5)	評価 (1-5)	評価 (1-5)	(秒)
提案	平均	2.58	2.95	2.53	8.51
システム	σ	0.731	0.455	0.623	2.06
モデル	平均	3.62	3.77	3.78	8.36
キャッチコピー	σ	0.604	0.498	0.481	1.34

表 3.8 「景色」「美しい」の場合における,システム生成のキャッチコピーの全体に占める割合

*は同じス	コアを同順位と	・してカウン	トレている
		ひしかう	こしん しゅしんご

が原位としてパックトしている				
	Top5	Top10	Top30	All
理解	2/5	5/10	22/30	(100/110)
間接的	2/5	6/10	24/30	(100/110)
全体の質	2/8(*)	6/13(*)	22/30	(100/110)
所要時間	5/5	10/10	28/30	(100/110)

表 3.9 「話」「面白い」の場合における,システム生成のキャッチコピーの全体に占める割合*は同じスコアを同順位としてカウントしている

	Top5	Top10	Top30	All
理解	3/5	5/11(*)	21/30	(100/110)
間接的	2/5	13/16(*)	28/34(*)	(100/110)
全体の質	3/6(*)	5/10	20/30	(100/110)
所要時間	5/5	10/10	28/30	(100/110)

例えば、「景色」、「美しい」の述語項の組が提案システムに入力される.ここで、景色が美しい場合の人々の行動の帰結を考慮する.この場合、人々は写真を撮る等、何らかのアクションを起こしたい衝動に駆られると考えられる.実際、出力されたキャッチコピーには「写真を撮るのに相応しい」情景を描写しているものも存在する.具体的な例として、「ゆっくり城を撮ろう」における城は美しい場合が多く、人々が写真を撮るうとする可能性もある.

入力対が「ケーキ」「おいしい」の時に関しては、提案システムは「みんなを上品にするカラフル」等のキャッチコピーらしい出力を得ることが出来ている.

さらに、評価実験によって、システム生成による一部のキャッチコピーに関しては、

表 3.10 「ケーキ」,「おいしい」 の場合における,システム生成のキャッチコピーの全体に占める割合

*は同じスコアを同順位としてカウントしている

	Top5	Top10	Top30	All
理解	3/5	7/11(*)	24/32(*)	(100/110)
間接的	2/6(*)	5/10	22/31(*)	(100/110)
全体の質	3/5	7/11(*)	24/33(*)	(100/110)
所要時間	5/5	10/10	26/30	(100/110)

可順位としてカワントしている						
	Top5	Top10	Top30	All		
理解	3/5	7/10	16/30	(300/330)		
間接的	1/5	4/10	19/31(*)	(300/330)		
全体の質	2/5	4/11(*)	18/32(*)	(300/330)		
所要時間	5/5	10/10	28/30	(300/330)		

表 3.11 全ての入力における,システム生成のキャッチコピーの全体に占める割合*は同じスコアを同順位としてカウントしている

プロ作成によるキャッチコピーに匹敵する質であることが確認されている.表 3.11 における,高評価であったキャッチコピーには理解レベルにおいて上位であった「みんなを上品にする苺」,間接的と評価された「神秘が降る、海のテレビ」,全体の質において上位であった「みんなを幸せにするかぼちゃ」が挙げられる.これらの出力結果により,モデルキャッチコピーに対して Web から取得した単語を挿入する手法の有効性が示唆される.

より短い抽象的なモデルキャッチコピーの方が,入れ替え対象となる許容範囲が広く,理解できる出力が多いと考えられる.例えば,モデルキャッチコピー「ゆっくり、家をつくろう」に対する生成例「ゆっくり、楽園をつくろう」「ゆっくり、街を飾ろう」などからもわかるように,場所に関するポジティブな単語の取得が行われている時に,理解が容易である出力が得やすいと推察される.一方で,モデルキャッチコピー「神々が選んだ島ギリシャ」に対する生成例「街々が並んだ近畿夏」「道々が並んだ東京中国」はそもそもの理解が困難であり,良いものが取得されにくくなっていると考えられる.理由として,名詞を連続させつつも,自然で意味が通る候補の生成が難しいからであることが考えられる.

このように,多くの理解できない出力も存在する.特に,キャッチコピーが長くなると,高品質のキャッチコピーを得ることが困難になる.従って,より洗練された候補選択が今後の課題である.

3.6 まとめ

Web 上の文情報を用いたキャッチコピー自動生成システムの提案を行った.

キャッチコピーは短い文にも関わらず,人の目を惹きつける.今回は,直接的表現に対して間接的表現を用いたキャッチコピーに着目した.間接的なキャッチコピーを生成するため,以下の二つのアプローチを行った.

(1) Web を用いた「対象,性質」からなる入力対に関する知識獲得

(2) 様々な表現を目的とした,膨大な量のキャッチコピーを含むコーパスの利用

まず,提案システムに「対象,性質」からなる述語項が入力される.次に,入力対についてクエリ化を行い,Web からの知識抽出によってテーマ選択が行われる.次に,テーマに相応しい単語が取得される.また,Web からの知識に基づきモデルキャッチコピーの選択が行われる.モデルキャッチコピーに対して,名詞・動詞・形容詞の挿入をすることにより,キャッチコピー候補の生成が行われる.これらのキャッチコピーはWeb 日本語 N グラムコーパスを用いて文としての適切度を基に選択が行われる.

主観評価実験によって,システム生成のキャッチコピーの質の検証を行った.実験結果により,生成されたキャッチコピーの一部はプロ作成のものに匹敵する評価が得られた.

第4章

SNS上の嗜好情報を用いたキャッチコピー自動生成システム

本章では, SNS 上の嗜好情報を用いたキャッチコピー自動生成システムを提案する. 第3章での取り組みを実世界における SNS サイトに適用することにより, 投稿に相応しいキャッチコピーの生成を目指す. 具体的には, SNS サイトからの情報抽出によりユーザの単語に対する嗜好情報の取得を行う.

一般にコストがかかるため、心理実験等でも膨大な量の嗜好情報を得ることは困難であった.今回のアプローチでは特に課題であったユーザの嗜好情報の取得に関して、SNS サイトの投稿における「いいね!」に着目し bag-of-words に対して線形回帰モデルを適用することで、解決を図っている.

このような応用を行うことにより,低コストでユーザが好むキャッチコピーを選択でき,社会的需要と考えられるパーソナライズへの対応が期待できる.

評価実験により,提案システムは人手選択に近いレベルでよりユーザの興味を引く キャッチコピーの選択が可能となった.

4.1 背景

ブランド構築や商品の販売促進のためにキャッチコピーはマーケティングにおいて重要である.近年では,インターネットにホームページを置き,インターネットにおける宣伝にも力を入れている企業も多い.また,膨大な数の人がソーシャル・ネットワーキング・サービス(SNS)を利用するようになり[178],SNS,Twitter等のソーシャルメディアにおいても広告活動を行うことがスタンダードになりつつある[179].例えば,Facebook上にファンページ等が開設され,プロモーションが行われている[180].従って,以前に比べ広告主は,ユーザに対してより親密なアプローチを取ることが容易になったと考えられる.キャッチコピーは通常,コピーライターが作成することが多い.しかしながら,様々な嗜好を持つ個々のユーザに対してそれらを考慮し,キャッチコピーを作成するのは現実的ではない.

さらに、物質的側面では社会が成熟し、需要に対して商品の供給が十分に行われることが多くなった.従って、単にモノやサービスを提供するだけでなく、消費者の感性と行動に着目したビジネスが求められている[181].それゆえ、近年のキャッチコピーは機能性に加えて、環境に対する優しさや社会性、感性等を内容に盛り込み、他の商品との差別化を図る必要性が出てきていると考えられる.ユーザの個性に関する研究には以下のようなものがある.例えば、Kosinski ら [154] は Facebook を対象とした研究を行い、ユーザの「いいね!」とその属性の間に強い相関があることを見出した.また、文の嗜好性に関する研究も行われている.Andréら [182] は Twitter を分析し、情報共有や思いついたアイデアの方が、自分本位志向や現在の状態を知らせるものより価値を置かれることを発見した.国内の類似研究としては、沢井ら [183] による文書に対する興味の強さの推定の研究がある.このようにその時代・時期のトレンドやユーザによって嗜好性が異なるため、動的に新しい文章を生み出すことが求められていると考えられる.

キャッチコピーに関する分析及びシステムには以下のものがある.松平ら [144,145] は,質より量,批判厳禁を旨とする発散的発想支援の立場から,遺伝的プログラミングを用いたキャッチフレーズ生成システムを提案した.西原ら [146-148] は,研究発表に限定し,タイトル毎に興味を引きやすいかどうかをランク付けするアルゴリズムを提案した.中野ら [149,150] はブログ記事を用いて,目を引くキャッチフレーズを作るための基礎的研究を行った.これらの成果に加え,第2章ではキャッチコピーの特性の分析を行い,その定量的分析結果に基づいたキャッチコピー生成システム,第3章では間接的なキャッチコピーの生成を主眼とするキャッチコピー生成システムの提案を行った.幅野ら [153] によって映画のキャッチコピー生成を目指したものが存在する.

このようにキャッチコピー生成を実社会のドメインに応用していくことは重要であると考えられる.特に,情報化が進み,消費者に対してインタラクティブにアプローチ出来るようになった現代社会において,ユーザに好まれる広告を提供する重要性は増大していくと考えられる.

キャッチコピーの生成を行うシステムは,松平ら [144,145] 及び第2章および第3章のキーワードを入力するもの(例:「夏」「かき氷」),中野ら [149,150] 及び幅野ら [153] のドメイン名を入力するもの(例:映画名)に大別できる.第3章の提案システムを実世界への応用するため,後者のようにドメイン名を入力として扱う.また,キャッチコピーシステムにおいては,システム中で生成された候補の選択が非常に重要な問題である.松平ら [144,145] 及び中野ら [149,150] のシステムは作成者の評価による選択,西原ら [146-148] の手法は相互情報量を用いた面白さの推定,幅野ら [153] は Web コーパスの頻度を用いた候補選択を行っている.松平ら及び中野らのシステムは,人手に

よる評価が必要な点においてコストがかかり,西原らの手法は予め存在するタイトルにおける単語の組み合わせの面白さを推定するため,自然な文の選択は扱っていない.

以上入力及び選択手法二点の問題を解決するため,(1) SNS における投稿の情報を利用することにより,キャッチコピーの生成を行う.同時に,(2) ユーザに好まれるキャッチコピーを考慮に入れることにより,現実的に利用可能なシステムを目指す.具体的には,Facebookのファンページへの投稿を利用し,キャッチコピー候補の生成を行い,投稿において好まれる単語を推定することで,ユーザに好まれると推測されるキャッチコピーを出力する.

以降,本章では4.2で Facebook におけるファンページについて説明する.4.3でユーザの嗜好を反映させたキャッチコピー生成システムを提案する.4.4では,評価実験について述べ,最後に4.5をまとめとする.

4.2 Facebook ファンページ

Facebook にはユーザのプロフィール,フィード,メッセージ等があり,本研究で対象とするファンページは,会社などが自社のブランドイメージの向上を意図して利用している。このファンページには「いいね!」「シェア」「コメント」等の属性があり,ここでは特に,投稿 (post) の本文及び「いいね!」に着目する.投稿はファンページにおける記事の最小単位である「いいね!」は Facebook ユーザー人が各記事に対して一回のみ肯定的評価を下せるもので,一般にこれを多く有している投稿は高評価で魅力的であると考えられている.これらの特徴から,時間の変化により移り変わる好みであるトレンドについても「いいね!」の数を基に取得できると考えられる.

4.3 SNS上の嗜好情報を用いたキャッチコピー自動生成システム

図 4.1 にシステム全体の流れを示す.まず,システムにファンページにおける投稿が入力される.次に投稿の知識獲得が行われる.その後,キャッチコピーコーパスに含まれるキャッチコピーをモデルとして,キャッチコピー候補が生成される.候補選択において,文の自然さ,意味的関連性,嗜好度合いが考慮され,キャッチコピーが出力される.

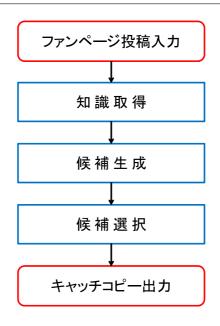


図 4.1 提案システムの流れ

4.3.1 知識取得

ここでは, Facebook におけるファンページの投稿の知識を取得する. 具体的には, 日本語構文・格解析システム KNP [184, 185] によって, 文中に含まれる単語のカテゴリの取得を行う.

4.3.2 候補生成

候補生成においては、モデルとなるキャッチコピー選択、及び知識取得で得られた 単語のモデルキャッチコピーへの適用を行う、第3章の提案システムのように、構造 の模倣を行うことにより、キャッチコピーの生成を行う、すなわち、モデルキャッチコ ピーを決定した後、その構造を利用し、一部の品詞を新たな単語に置き換えることに より、その投稿に相応しいキャッチコピーの生成を目指す、

モデルキャッチコピー選択

より文脈にふさわしいキャッチコピーを選択する際に、そのキャッチコピーが具備する特徴が重要になると考えられる.提案システムで用いるキャッチコピーコーパスには、ファッションや生活用品等のテーマが存在し、キャッチコピーはそれぞれのテーマに属している.まずこれらのテーマを設定し、その後でキャッチコピー選択を行うことによって、より投稿の文脈にふさわしいキャッチコピーの選択を試みる.そのため1)テーマ設定、2)キャッチコピー選択の二つのフェーズを順に説明する.なお、モデ

ルキャッチコピー選択における選択手法は前章の伝え方に着目したキャッチコピーにおける選択手法,式 (3.1-3.3) と同一のものである.

1) ファンページにおける投稿の情報がキャッチコピーコーパスにおけるテーマ内のキャッチコピーにより多く含まれるほど,生成に相応しいテーマであると考えられる.つまり,投稿における単語とキャッチコピーコーパスにおけるテーマ内の単語において,よりテーマ内の単語が多く含まれるほどスコアを与えるように式を組み立てる.具体的には,日本語形態素解析器 $\operatorname{MeCab}[159,186]$ を用いて,名詞および動詞,形容詞の頻度の取得を行う.同様に,キャッチコピーコーパスからも名詞および動詞,形容詞の単語頻度を取得する.この処理の後,テーマは以下のようにして決定される.T をテーマ,W(T) をそのテーマに含まれる単語, W_{POST} を投稿における名詞および動詞,形容詞に関する単語群, W_x を W(T) と W_{POST} の共通集合, W_y を W(T) における単語,F() を単語頻度として,テーマスコア S(T) を,

$$S(T) = \frac{\sum_{W_x \in W_s(T) = \{W(T) \cap W_{post}\}} F(W_x)}{\sum_{W_y \in W(T)} F(W_y)}$$
(4.1)

で表す.

従って,提案システムの選択するキャッチコピーコーパスにおけるテーマ T_f は,

$$T_{f} = \arg\max_{T} \left\{ \frac{\sum_{W_{x} \in W_{s}(T) = \{W(T) \cap W_{POST}\}} F(W_{x})}{\sum_{W_{y} \in W(T)} F(W_{y})} \right\}$$
(4.2)

で決定される.

2) テーマ決定の後,そのテーマ T_f に含まれるキャッチコピーに対して選択が行われる.テーマにおける選択と同様に,より投稿の情報を多く含むキャッチコピーが選択される.具体的には,L をキャッチコピー,W(L) をキャッチコピーにおける単語, W_z を W(L) と W_{POST} の共通集合,N(L) をキャッチコピー中に含まれる単語数とした時,キャッチコピーの投稿への相応しさ S(L) は,

$$S(L) = \frac{\sum_{W_z \in W_s(L) = \{W(L) \cap W_{POST}\}} F(W_z)}{N(L)}$$

$$(4.3)$$

で与えられる.このスコアが高いものがモデルキャッチコピーとして選択される.

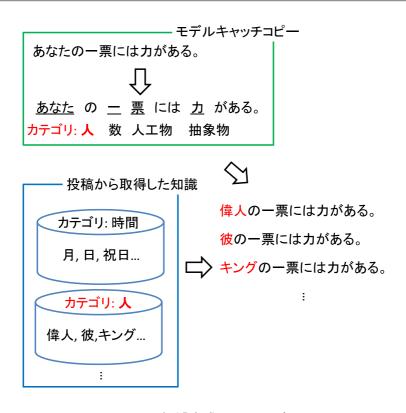


図 4.2 候補生成のイメージ図

4.3.3 キャッチコピー候補生成

モデルキャッチコピーに対して,知識取得でカテゴリが合致する単語の探索を行い,見つかった場合にキャッチコピー候補の生成を行う.図4.2に候補生成の例を示す.図からわかるように,モデルキャッチコピー「あなたの一票には力がある」に対して,カテゴリ分けが行われ,知識取得において合致した「人」に関して,キャッチコピーの候補生成が行われている.意味の相違を最小限にするため,品詞が名詞であり,かつそのカテゴリが一致したもののみの変更を行う.

4.3.4 候補選択

システムによる生成のキャッチコピーは文法的には正しいが,文の意味・自然さの 点で不完全なものも少なくない.ここでは,自然さと投稿との意味的関連性及び嗜好 度合いの三つの観点からキャッチコピーの選択を行う.

文の自然さ

ここでは,提案システムは不自然な候補の削除を行う.語用論的な自然度の尺度は,統計的な手法により定量化できる.具体的には Web 上の 2,550 億単語から構築された大規模なコーパスである Web 日本語 N グラム [94,95] を用いて自然さのレベルの計算を行う.まず,コーパスにおいて候補における単語シーケンスの存在の有無の確認を行うことで,文として成立しない候補の排除を行う.コーパスにおいて,規定の gram数 (=3-gram) 以上存在しているものに関して,文の自然さに関するスコアリングを行う.つまり,ある複数の単語がシーケンス上に同時に現れるときに,相対的にそれぞれの単語頻度に対して大きい場合に関して,より大きなスコアを与えるように式を立てる.具体的には,相互情報量のアイデアを基に以下の計算を行う.キャッチコピーをL,その最小の構成要素(単語シーケンス)を L_E ,単語群を $W(L_E)$,F() を Web 日本語 N グラムにおける頻度とした時,そのキャッチコピーのスコア $S_P(L)$ は,

$$S_p(L) = \sum_{L_E \in L} \left(\sum_{W_k \in W_p(L_E) = \{W(L_E) \cap \{W_n, W_v, W_a\}\}} \log(F(W_k)) - \log(F(L_E)) \right)$$
(4.4)

として計算される.この場合は,スコアが低いものが自然な文となる.なお,通常は頻度の積を取ることで計算を行うが,コーパスにおける頻度のスコアが大きくなる場合には総積の値が大きくなり過ぎる場合があること,また頻度の差異に対してロバストにするため,式 (4.4) においては対数を取っている.

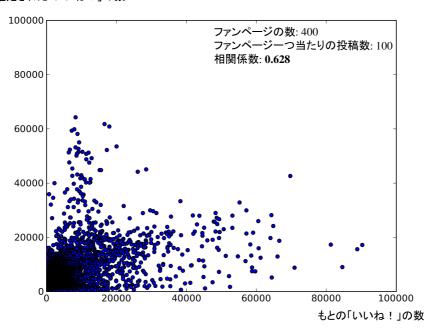
意味的関連性

式 (4.3) で取得されたキャッチコピーのスコアリング情報を利用する. つまり, モデルキャッチコピー中に投稿に関連する単語がより多く含まれているものが, 意味的関連性が大きいものと見なしている.

嗜好度合い

Facebook へ投稿された文章に対して, bag-of-words モデルを用いて線形回帰を行うと, 単語のみでも嗜好度合いを推測できることが著者らによって報告されている[187]. 図 4.3 に x 軸に実際に「いいね!」が押された数と, y 軸に線形回帰に基づいて予測した投稿の「いいね!」の相関図を示す.

図から,ファンページにおける投稿に対して実際に「いいね!」が押された数と,推定された「いいね!」の数の間には相関関係があることが確認できる.従って,線形



推定された「いいね!」の数

図 4.3 「いいね!」の相関図

回帰による重み付けの単語を考慮することによって,より嗜好度合いの高いキャッチコピーの取得が出来ると推測される.以下にその全体の概要を述べる.

まず、Facebook からのファンページがシステムに入力される.ファンページにおける URL 等のコンテンツに関係ないものは前処理において削除される.その後、単語の重要度取得において、メッセージは形態素に分けられ、嗜好度合いに応じて重み付けされる.最終的に、これらの単語の重みを用いてファンページにおけるそれぞれの投稿の嗜好度合いが再現される.

次に,図 4.4 に詳細な処理の手法について示す.ファンページでは,それぞれの投稿に対して固有の「いいね!」等の嗜好情報が付加されている.ここではシンプルなモデルとして「いいね!」の数を投稿における文の単語数で割ることにより,正規化を行う.

あるファンページの投稿 P における単語 w のスコア P(w) は, L_p を「いいね!」の数, N_p を投稿中の単語数として,

$$P(w) = \sum_{p \in Posts} \frac{L_p}{N_p} \tag{4.5}$$

として求める.

次に,嗜好のスコアは投稿全体に含まれる単語数により正規化される.従って,各単語のスコアS(w) は,all を全投稿,N(w) を all 中の単語数として,

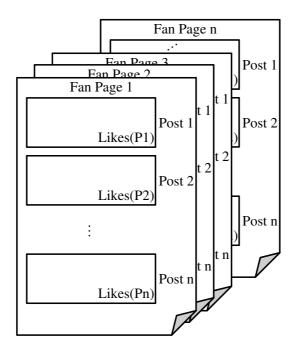


図 4.4 嗜好度合いの推定のイメージ図

$$S(w) = \frac{\sum_{p \in all} P(w)}{N(w)} \tag{4.6}$$

で表される.ここまでの処理においては,単語の重要度は均等に扱ってきた.しかしながら,単語によって重要度に関してばらつきがあることが推測される.ここでは一投稿を一文書と見なし,式 (4.5) において IDF を用いることにより,単語の重要度の推定を行う. IDF は,D をファンページの全てのファンページの投稿数,分母は文書中において単語 t が現れる数として,

$$IDF(t,D) = \log\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right)$$

$$(4.7)$$

となる.従って,式(4.5)は、

$$PI(w) = \sum_{p \in Posts} \frac{L_p}{\frac{IDF_w}{\sum_p IDF_w}}$$

$$\tag{4.8}$$

のように変形される.ここまでの処理により,それぞれのファンページにおける単語の 嗜好度合いに関して,ある程度の推定ができていると考えられる.しかしながら,こ れを文章の嗜好度合いに適用すると,人気のファンページであるため単語が特異的に 高いスコアになってしまうような場合,文章の嗜好度合いの推定に困難が生じてしまうと考えられる.従って,さらにファンページにおける投稿全体の「いいね!」の合計で正規化を行い,ファンページ間の嗜好度合いの差異の解消を行う.

$$P''(w) = P'(w) \frac{L_p}{\sum_{p \in Posts} L_p}$$
(4.9)

以上の処理を行った後, キャッチコピーLの嗜好度合いE(L)は,

$$E(L) = \sum_{w \in W_L} P''(w) \tag{4.10}$$

として計算される.提案システムでは,このスコアが高いものがユーザに相応しいキャッチコピーとして選択される.

以上,それぞれ式(4.4),式(4.3),式(4.8) に対応する「文の自然さ」「意味的関連性」「嗜好度合い」三つの観点に基づき,キャッチコピーLの最終スコア $S_F(L)$ は,

$$S_F(L) = \frac{1}{S_p(L)} \times S(L) \times E(L)$$
(4.11)

として計算される.このように提案システムにおいては,自然さ,意味的関連性,嗜好度の総積のスコアが高い候補が出力される.

4.4 評価実験

4.4.1 実験概要

提案システムによって生成されたキャッチコピーの質の評価を行う.その際,様々な基準から生成および選択されたキャッチコピー同士を比較する.具体的には,

- キャッチコピーコーパスに含まれているもの(=モデルキャッチコピー)
- 生成されたものを人手で選択したもの
- 生成されたものを自然度及び意味的関連性を考慮した後で取得したもの(嗜好度 合い考慮しない)
- ◆ 生成されたものを自然度,意味的関連性及び嗜好度合いを考慮した後で取得したもの(嗜好度合い考慮)すなわち式(4.11)
- 生成されたものから自然度のみを考慮した後でランダムに取得したものの 5 つの基準を考慮して比較を行う.

また,キャッチコピーを判断する指標として,

- 1. 投稿の内容が反映されているか
 - 1. 反映されている
 - 2. やや反映されている
 - 3. どちらでもない
 - 4. あまり反映されていない
 - 5. 反映されていない
- 2. 興味を持ったか
 - 1 興味を持った
 - 2. やや興味を持った
 - 3. どちらでもない
 - 4. あまり興味を持たなかった
 - 5. 興味を持たなかった
- 3. キャッチコピーとしての質
 - 1. 良い
 - 2. やや良い
 - 3. どちらでもない
 - 4. やや悪い
 - 5. 悪い

の三つの観点を基に主観評価を収集し,さらに同時に評価に要する時間計測も行い, キャッチコピーの質の検証を行った.

4.4.2 実験条件

Facebook 企業/ブランド公式ページ [188] に掲載されている 500 のファンページ上の企業 (2013 年 6 月 14-18 日に取得) で取得可能であった 486 企業の直近 100 投稿を利用する.ここで用いたキャッチコピーコーパスとして,久野が作成した 4 冊のキャッチコピー大百科 [155, 174-176] を用いた.それぞれのコーパスは 1980 年から 2008 年までの特定期間にわたって新聞・雑誌広告,カタログ・チラシ等に掲載された約 6,000 個のキャッチコピーを含んでいる.例えば,大百科 [174] は 1980 年から 1995 年までの雑誌広告における「食品」「化粧品」「ファッション」「旅行・サービス」「通信・精密機器」等のテーマのキャッチコピーを収録している.これら 4 つのコーパスにおけるテーマを統合した結果,44 のテーマにおける合計 24,472 個がモデルキャッチコピーの選択



図 4.5 評価実験のインタフェース

候補となっている.他のキャッチコピーコーパス[189]の利用も検討したが,テーマ数,キャッチコピーそのものの数を考慮した結果,キャッチコピー大百科の利用が最適であると判断した.

また,嗜好度合い推定の際には,MeCab による形態素解析の後,名詞および動詞, 形容詞のみ用いた.更に,ストップワードとして沢井ら [183] が用いた「する」「ある」 「よる」「いる」「なる」「いう」「みる」「できる」を利用し,これらの単語を除いて,嗜 好度合いの算出を行った.

図 4.5 にユーザインタフェースを示す.被験者は画面上部に提示される投稿内容を確認した後,その下に表示されているキャッチコピーについて実験概要で述べた三つの観点に関して評価を行う.キャッチコピーは「アメリカン・エキスプレス」「富士通FMV」「セブンネットショッピング」3 社のファンページにおける投稿から生成される.評価実験に用いられたキャッチコピーは合計約 100 個であり,それらがランダムに提示される.具体的には「嗜好あり・嗜好なし」に関しては,それぞれのファンページにおける投稿から生成されスコアが最も高かったもの(上位一件ないしは二件)が提示される.なお,キャッチコピー実験の被験者は 20 代大学生および大学院生 9 名である.

4.4.3 実験結果

表 4.1 に提案システムによる生成例を示す.

「コーパス内」はモデルキャッチコピーに「人手選択」はシステム生成の候補を人間が選択したものに「嗜好なし」は自然さと意味的関連性の観点から選択したものに「嗜好あり」は自然さ、意味的関連性及び嗜好度合い全てを用いて選択したものに「ランダム」はシステム生成の候補をランダムに提示したものに、それぞれ対応している.この中で「嗜好あり」が最終のスコアリングまで適用したシステム生成のキャッチコピーである.例えば「贈るギフトは、夏のにおいがしたから。それだけ。」は、セブン&アイグループのセブンネットショッピングのギフトに関する投稿から生成されている.まず、モデルキャッチコピーとして「贈る理由は、夏のにおいがしたから。それだけ。」が選択された後、生成された.

表 4.2 に評価実験の結果を示す.表 4.1 と同様「コーパス内」はモデルキャッチコピーに「人手選択」はシステム生成の候補を人間が選択したものに「嗜好なし」は自然さと意味的関連性の観点から選択したものに「嗜好あり」は自然さ、意味的関連性及び嗜好度合い全てを用いて選択したものに「ランダム」はシステム生成の候補をランダムに提示したものに、それぞれ対応している.

表からわかるように「投稿の内容が反映されているか」に関しては,人手選択によるものが高く,次に嗜好度合いを考慮したキャッチコピーのスコアが高かった.一対の標本による平均の検定では,人手選択が嗜好度合いを考慮したキャッチコピーを有意水準5%で上回らなかった (p=0.142>0.05) が,嗜好度合いを考慮したキャッチコピーは考慮していないものに対して反映されていることが有意水準5%(p=0.000663<0.05)で認められた.また,人手選択と嗜好ありのものは相関係数0.615 と嗜好あり・なしの相関係数0.438 よりも大きく,類似性が見られた.嗜好度合いを考慮しなかったキャッチコピーは,ランダムのものに比べ有意水準5%(p=0.0492<0.05) でより投稿に対して反映されていると評価を得た.一方で,キャッチコピーコーパスから取得したものは反映度が他と比較すると低いことがわかる.標準偏差に関しても,1.23 と他の生成されたキャッチコピーに対してばらつきが大きいことが確認された.

「興味を持ったか」については,キャッチコピーコーパスのものが高く,次に来る人手選択に準じる形で,嗜好度合いを考慮したものが高評価であった.一対の標本による平均の検定では,キャッチコピーコーパスにおけるモデルキャッチコピーが他のものに比べ有意に高く $(p=2.58\times 10^{-8})$,人手選択及び嗜好度合いを考慮したものに関しては,有意差は認められなかった (p=0.346>0.05).一方で,嗜好度合いを考慮したものは考慮していないものに対して有意差 5%(p=0.000182<0.05) で高評価であった.また,人手選択との相関は,嗜好度合いなしの相関係数 0.378 に対して,嗜好度

表 4.1 提案システムによるキャッチコピーの生成例 アメックスはアメリカン・エキスプレス , FMV は富士通 FMV , 7 ネットはセブンネットショッピングを示している

対象	ファンページ	 ンページ キャッチコピー例	
コーパス内 (モデルキャッチコピー)	アメックス	達成感の先にあるもの。	
		それは新たな挑戦。	
	FMV	「かけがえのない命」	
		が生まれた日。	
	7ネット	贈る理由は、夏のにおいが	
		したから。それだけ。	
人手選択	アメックス	参加ライブの先にあるもの。	
		それは新たな挑戦。	
	FMV	「かけがえのないイベント」	
		が生まれた日。	
	7ネット	夏のスタートを	
		知っている一着。	
嗜好なし	アメックス	私、若い頃は紹介に向いて	
		ないと思っていたんですよ。	
	FMV	世界の他を	
		変える時が来た。	
	7ネット	母の夏は母と私の、	
		特集の夏。	
嗜好あり	アメックス	卒業ライブの先にあるもの。	
		それは新たな卒業。	
	FMV	世界の文字を	
		変える時が来た。	
	7ネット	贈るギフトは、夏のにおいが	
		したから。それだけ。	
ランダム	アメックス	構成で、	
		応援しよう。	
	FMV	特徴の数だけ	
		紹介販売を業務します。	
	7ネット	母の日は母と婦人の、	
		販売の日。	

		反映	興味	質	時間
		(1-5)	(1-5)	(1-5)	(s)
コーパス内	平均	3.09	4.01	3.98	15.6
	σ	1.23	0.96	1.00	13.2
人手選択	平均	3.61	3.55	3.47	15.0
	σ	1.14	1.12	1.15	13.6
嗜好なし	平均	3.25	3.15	2.95	18.1
	σ	1.06	1.00	1.09	16.8
嗜好あり	平均	3.52	3.51	3.26	17.3
	σ	1.14	1.09	1.16	22.1
ランダム	平均	3.13	3.22	3.06	20.0
	σ	1.10	1.12	1.19	31.8

表 4.2 キャッチコピーの評価実験結果 (σ は標準偏差)

合いありの時に 0.469 と , 相関係数の向上が見られた . 総じて , 嗜好度合いの導入が , キャッチコピー候補の選択に有効であるといえる .

「キャッチコピーとしての質」に関しては,モデルキャッチコピーが最も高く,人手選択,嗜好ありのキャッチコピーの順に高評価を得た.モデルキャッチコピーは人手選択に,人手選択は嗜好ありに対して,それぞれ有意差 $(p=2.09\times10^{-8}<0.05)$ p=0.00827<0.05 が認められている.嗜好度合いを考慮したものは考慮していないものに対して有意差 5%(p=0.00171<0.05) で上回っていることが確認された.また,人手選択との相関は,嗜好度合いなしの相関係数 0.394 に対して,嗜好度合いありの時には 0.544 と,相関係数の向上が見られ,人手選択に接近する結果となった「興味を持ったか」と同様に,嗜好度合いはより高品質なキャッチコピーの選択に貢献しているといえる.

また,時間に関して,人手選択のもの及びコーパス内のものがより短い時間で評価を得られていた一方で,ランダムのものはより多くの時間がかかることが見出された. 具体的には,一対の標本による平均の検定では,人手選択に対しては有意水準 5%(p=0.0134<0.05) で,モデルキャッチコピーに比べて有意水準約 $5\%(p=0.0568\approx0.05)$ でランダムに選択されたキャッチコピーは評価に時間を要した.更に,標準偏差も 31.8 秒と群を抜いて大きいことが確認された.このことから,評価が容易なわかりやすいキャッチコピーを生成および選択することが,更なる質の向上につながる可能性が示唆される.

4.5 まとめ

本章では、SNS上における嗜好情報を用いたキャッチコピー自動生成システムを提案した。本システムの最大の特徴は SNS ファンページ上の投稿を知識源として、キャッチコピーの生成、選択に利用することである。提案システムは、企業等が宣伝を行っているそれぞれの投稿の記事内容を反映させたキャッチコピーの自動生成を目的としている。まず、SNS上の投稿がシステムに入力される。投稿における内容のカテゴリ知識及び 24,472 個の膨大な数のキャッチコピーをモデルキャッチコピーとして利用することにより、キャッチコピー候補の生成を行う。候補選択においては、大規模コーパスを用いた文の自然さ、キャッチコピーコーパスを用いた意味的関連性、SNS から取得した情報を用い嗜好度合いを考慮した。特に、嗜好度合いに関して、SNS ファンページ上の情報に対して、bag-of-words に基づいた線形回帰モデルを用いることでユーザに好まれる単語の推定を行った。これらの単語に基づくスコアリングを行うことにより、トレンドを加味し、キャッチコピーを自動的に選択することが可能となった。

評価実験により,ランダムに生成された候補に比べ,文の自然さ及び意味的関連性を考慮したものがより投稿記事の内容を反映していることが示された.また,投稿に対する嗜好度合いを単語レベルで考慮することにより,人手選択に近いレベルでよりユーザの興味を引くキャッチコピーの選択が可能となった.

第5章

結論

5.1 まとめ

本論文では、キャッチコピーの分析、およびその自動生成について述べた.3通りの自動生成システムを提案し、それぞれ統計的手法を核としたキャッチコピーの模倣、キャッチコピーの間接的な表現に着目した手法、これを実世界のSNSサイトに適用したアプローチを行った.

人々の心を動かす文には様々なものがあるが,キャッチコピーは短い文でありながら,人目を引きつけるものが多い.このキャッチコピーを工学分野から扱ったものは数が少なく,まず,キャッチコピーの定量的側面の分析が不可欠であった.

第2章では,この点を克服するため,独自に構築を行ったキャッチコピーコーパスに対して自然言語処理のツールを用いて統計的分析を行った.また,ここで明らかになった知見を基に,その特徴を模倣するキャッチコピー自動生成システムの提案を行った.

第3章においては,統計的手法のみでは解決が困難なキャッチコピーの間接的な表現に着目した.キャッチコピーには「これは丈夫です」よりも「象が踏んでも大丈夫」のように,間接的な表現の方が印象に残ることが多い.この方向でのアプローチを行うため,Webからの知識取得,およびキャッチコピーコーパスの更なる拡張をすることにより,間接的なキャッチコピーの生成を行った.

第4章では、第3章のシステムをより実世界に適用する試みである.具体的には、Social Networking Service (SNS) サイトに適用することにより、投稿に相応しいキャッチコピーの生成を目指している.具体的には、SNS サイトにおけるファンページの投稿に付随する嗜好情報に対して、bag-of-words モデルを用いた線形回帰分析を適用することで、より好まれる候補の選択を行った.これによって、提案システムは、人手選択に近いレベルでよりユーザの興味を引くキャッチコピーの選択が可能となっている.上記のように、第4章のシステムは第2章、第3章のものに比べ、より実用性が増しユーザフレンドリーになっていることがわかる.

結論 82

以上3つの提案システムの特徴を以下にまとめる.

統計的手法を用いたキャッチコピー自動生成システム

- テーマ,キーワードを入力とする,キーワードを含んだキャッチコピーらしい キャッチコピーの生成が可能
- キーワードの情報伝達という点で有用
- プロ作成のものには質では比肩しにくい

有効なシーン: 含めたいキーワードが既知である時

Web 上の文情報を用いたキャッチコピー自動生成システム

- 間接的なキャッチコピー生成が可能
- 一部はプロ作成のものにも匹敵することを確認
- 意外性を持つキャッチコピーが生成できる可能性
- 解釈困難な出力が多い

有効なシーン: 含めたいキーワードがわからない場合

SNS 上の嗜好情報を用いたキャッチコピー自動生成システム

- 投稿内容を反映したキャッチコピーの生成を確認
- 興味の点では人手に匹敵するほどの選択が可能

有効なシーン: ファンページを基にキャッチコピー生成を行いたい時

これら3つの提案システムは、それぞれ統計に根ざしたアプローチ、ウェブにおける知識獲得手法、SNSにおける嗜好情報取得を行っている。まず、統計的にキャッチコピーらしさを追求することで、キーワードを含んだキャッチコピーらしいキャッチコピーの生成が可能であり、また知識獲得によってより斬新で広範な表現が期待でき、動的に SNS 上のトレンドを反映した嗜好情報を用いてより好まれる候補の選択が可能となっている。

以上の提案システムは,統計的手法に基づくため,データ量が増加するほど性能が 向上すると考えられる.ビッグデータの流れにより今後もデータ量の増加が見込まれ ることから,その意義は大きくなっていくものと考えられる. 結論 83

5.2 今後の課題

5.2.1 評価に関して

本研究では評価実験において、素人による評価を行ったが、専門家による評価は大変有効であると考えられる.特に、専門家の場合、素人には知覚出来ないメタ認知を持ち、素人の主観評価によらない評価が可能であると考えられる.一方で、全てを専門家に依頼した場合に、そのコストが膨大になる懸念がある.その点では、心理学実験のような主観評価によらない評価実験を一般人に行うことも考えられる.例えば、インターネット広告の効果を測定する際には、クリックスルーレートを用いることが多く、また、最近ではアイトラッカーによる視線追跡の研究も行われている.このような指標、機器を用いれば、ターゲット層、世代別により踏み込んだキャッチコピーの興味度評価が行えると考えられる.

5.2.2 関連する発展的なテーマ

今後の発展的な研究テーマとして「言い易さ」「韻を踏む」等の音韻及び形式的要素の考慮,生成モデル等を用いた文生成の洗練化,そして深層的な意味分析をキャッチコピーにおいて行うことが考えられる.特に最後の意味分析に対するアプローチには,キャッチコピーにおけるより詳細なメタデータの付与が挙げられる.現状のキャッチコピーコーパスの問題点として,

- 1. 計算機で処理するための噛み砕いた意味が付与されていないこと
- 2. より細かなカテゴリ分類が行われていないこと
- 3. 受け手のより詳細な嗜好情報

が挙げられる。1. については「ものが頑丈である」ことを伝えるときに「象が乗っても大丈夫」と表現するキャッチコピーが存在する。しかしながら、単語群「象」「乗る」「大丈夫」からは「ものが頑丈である」ことを計算機では直接処理できない。この場合には「象が乗っても大丈夫」に対して「もの」「頑丈」のような夕づ付けが別途必要であると考えられる。2. については、食品だけでなく食品のどのカテゴリに属するのかを詳細に夕づ付けすることによって、より高精度なキャッチコピーの生成ができると期待できる。3. については、感情を含めた嗜好情報を取得すると良いと考えられる「いいね!」に対応するポジティブさの情報だけでなく、喜怒哀楽等の細かな情動情報を得ることにより、生成したキャッチコピーが受け手にどのような影響を与えるのかより詳細な予測が可能となる。

さらに、エキスパートによる知見を取り入れる方向性もある.エキスパートがどのような尺度で単語の組み合わせの取捨選択を行っているかの知見を取り入れ (例えば、単語間の意味的類似度や音韻等の特徴化を行う、あるいは word2vec [190] のような大規模なテキストコーパスから単語のベクトル化を行うツールを用い、可能な限りの感性、イメージ情報の取得を行う)、キャッチコピーの良し悪しの評価を教師データとして与えることにより、キャッチコピー生成への応用の可能性があると考えられる.

また,広告は画像が付与されているものも多い.従って,与えられたテーマに関して,相応しいキャッチコピーを生成するとともに,画像を選択あるいは生成することも今後の更なる展望として挙げられる.

謝辞

本研究を進めるにあたり、様々なアドバイスを下さった萩原将文先生に心から感謝いたします。学部一年生の時に個別にして頂いた研究室説明から、学部の三年生の研究室配属を経て、現在に至る長きにわたって、萩原先生には研究の他、大学生活全般に関して親身になって相談に乗って下さいました。また、本研究に必要不可欠なキャッチコピーのコーパスを入手、加工する際にもご尽力頂き、研究に対しても深いご理解を頂き、自分自身大変快適で有意義な研究生活を送ることが出来ました。重ねて御礼申し上げます。

大変お忙しい中,副査をお引き受け下さり,有益で貴重なコメントを下さった斎藤 博昭先生,杉本麻樹先生,山口高平先生には深く感謝申し上げます.

また,学会や講演会等で様々な形で研究のアイデアを下さった方々に感謝いたします.特に,慶應義塾大学三田キャンパスで行われていた「脳の講習会」で毎回お世話になった元京都大学霊長類研究所所長の小嶋祥三先生,ディスカッションをして下さった梅田聡先生,柴田みどり博士,高山緑先生に感謝いたします.

キャッチコピーの研究を支援して下さった電通の吉田秀雄記念財団および,経済的に支えて下さった日揮実吉奨学金,慶應義塾先端科学技術研究センター,日本学生支援機構,そして特別研究員として採用して頂いた日本学術振興会には感謝申し上げます.間接的なキャッチコピーのアイデアをご教授下さり,出力されたキャッチコピーに対しても貴重なコメント,ご意見を下さったコピーライターとしてもご活躍されている松尾順氏,言語資源やツールを開発して下さった作成者,開発者の方々に深謝いたします.

多大なご助言,ご協力を下さった萩原研究室の皆様に深く感謝致します.研究室で過ごした時間は,かけがえのないものとして記憶に残り続けると思います.

まず学部時代からお世話になった,先輩の畑さん,家田さん,橋本さん,高野さん, 幸塚さん,大浦さんに感謝いたします.また,特に堀田創さんと清水浩平さんには研 究や,研究の組み立て方や博士課程に関する有益なアドバイスを頂きました.ここに 御礼申し上げます.

有意義で愉快な時間を共に過ごさせて頂いた,同期の長谷川宏聡くん,相良司くん, 小倉卓也くん,新井寛くん,松尾健司くん,井上友美さんに感謝いたします.特に同 謝辞 86

期の皆さんとのディスカッションは実に実り多いものでした.研究内容のみならず人間的にも得るものが多かったです.本当にありがとうございます.

I greatly appreciate nice advices from Dr. Ji-Soo Keum and Dr. Kevin Binkley. I am also grateful to Michael Waechter, Arnaud Rachez, Lubin Maxime, Wolfer Geoffrey, Mattia Desana, Zelda Mariet, and An Nguyen for funny discussions (valuable advices of course!) and enjoyable time.

後輩の皆さんにも心から感謝いたします.特に,吉田くん,伊藤さん,大竹くん,本間くん,菅生くん,江口くん,沖くん,長谷くん,そして現 M1の山本くん,佛木さん,和泉くん,関くん,小川さん,自ら論文チェックを申し出てくれた松井くんにはとても感謝しております.皆さんのご協力のお陰で,このように今謝辞を書くことが出来ます.ありがとうございます.

紙面の都合上ここに書ききれませんが,私をいろいろな形で応援して下さった皆様 に感謝いたします.

最後に,経済的にも精神的にも様々な形で支えて下さった家族,弟,母,そしてこの工学博士への導きを与えてくれた天国にいる父に厚く感謝いたします.

- [1] P. H. Winton, "Artificial Intelligence Third Edition," Addison-Wesley Publishing Company, 1992.
- [2] 小林 一郎, "人工知能の基礎," サイエンス社, 2008.
- [3] 人工知能学会, "人工知能ハンドブック," オーム社, 1990.
- [4] 人工知能学会,"人工知能研究," http://www.ai-gakkai.or.jp/whatsai/AIresearch.html.
- [5] 独立行政法人科学技術振興機構開発研究戦略センター, "研究開発の俯瞰報告書電子情報通信分野 (2013 年) 知能システムの基礎," http://www.jst.go.jp/crds/pdf/2012/FR/CRDS-FY2012-FR-05.pdf.
- [6] 人工知能学会,"人工知能学事典," 共立出版, 2006.
- [7] C. T. Leondes, "Expert Systems: The technology of knowledge management and decision making for the 21st century, Vol.1," Academic Press, A Harcourt Science and Technology Company, 2002.
- [8] A. Newell and H. Simon, "Human problem solving," In *Prentice-Hall*. Englewood Cliffs, 1972.
- [9] E. A. Feigenbaum, B. G. Buchana, and J. Lederberg, "On generality and problem solving: a case study using the Dendral program," Machine Intelligence, Vol.6 Edinburgh University Press, 1971.
- [10] A. グッダル、戸内 順一、"エキスパートシステムとは何か、" 啓学出版、1987.
- [11] 安信 千津子, 重見 一秀, "エキスパートシステムの諸事例-V:金融ディーリング 支援エキスパートシステム,"情報処理, Vol. 33, No. 9, pp. 1087-1091, 1992.

[12] 山口 高平, 溝口 理一郎, 田岡 直樹, 小高 浩, 野村 康雄, 角所 収, "深い知識に基づく一般ユーザー向けの説明機能," 人工知能学会誌, Vol. J70-D, No. 11, pp. 2083-2088, 1987.

- [13] IBM100, "Icons of progress deep blue," http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/.
- [14] 日経トレンディネット, "女流王将が苦杯を喫した将棋コンピューター「あから 2010」とはナニモノだ!?," 日経 BP 社, 2010.
- [15] 新 清士, "九段破れたり AI とプロ棋士のタッグ戦、思わぬ番狂わせ," 日本経済 新聞, 2013.
- [16] IBM100, "'Eugene' becomes first computer to pass the Turing Test," The Washington Post, 2014.
- [17] 新井 紀子, 松崎 拓也, "ロボットは東大に入れるか? : 国立情報学研究所「人工頭脳」プロジェクト (<特集>ロボットは東大に入れるか?)," 人工知能学会誌, Vol. 27, No. 5, pp. 463-469, 2012.
- [18] Q. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean, and A. Ng, "Building high-level features using large scale unsupervised learning," In International Conference in Machine Learning, 2012.
- [19] 小林 雅一, "クラウドから AI へ アップル、グーグル、フェイスブックの次なる 主戦場 (朝日新書) [新書],"朝日新聞出版, 2013.
- [20] "車の自動運転なぜ進む?情報集積、新たな価値生む," 5月 13日付.日本経済新聞社, 2014.
- [21] 青木 啓二, "自動運転技術の開発動向と実用化に向けた課題," ISITS 第 14 回カーエレクトロニクス研究会「自動運転技術の開発動向と実用化に向けた課題」. 公益財団法人九州先端科学技術研究所, 2013.
- [22] J. Dolcourt, "MS の音声アシスタント「Cortana」の実力は 「Siri」「Google Now」と比較," CNET News, 2014.
- [23] M. A. Boden, "Creativity and artificial intelligence," *Artificial Intelligence*, Vol. 103, No. 1-2, pp. 347–356, 1998.

[24] 國藤 進, "発想支援システムの研究開発動向とその課題 (<特集>「発想支援システム」)," 人工知能学会誌, Vol. 8, No. 5, pp. 552-559, 1993.

- [25] 小谷 淳司, 萩原 将文, "構造的表現を用いた進化的形状デザイン支援システム," 電気学会論文誌 C, 電子・情報・システム部門誌, Vol. 121, No. 5, pp. 933-939, 2001.
- [26] 伴場 裕介, 小谷 淳司, 萩原 将文, "評価エージェントを用いた対話型進化計算法によるインテリアレイアウト支援システム (知的創造作業支援)," 情報処理学会論文誌, Vol. 46, No. 11, pp. 2804-2813, 2005.
- [27] S. Colton, "The painting fool: Stories from building an automated painter," In J. McCormack and M. dnverno, editors, Computers and Creativity, pp. 3–38. Springer Berlin Heidelberg, 2012.
- [28] 是永 基樹, 萩原 将文, "対話型進化計算法によるフラワーレイアウト支援システム," 感性工学研究論文集, Vol. 4, No. 2, pp. 81-88, 2004.
- [29] R. L. de Mantaras and J. L. Arcos, "AI and music from composition to expressive performance," AI Mag., Vol. 23, No. 3, pp. 43–57, 2002.
- [30] T. Hajdu, J. Hardesty, J. Lubin, and A. Milburn, "Method and apparatus for composing original musical works," EP Patent App. EP19,990,905,967, 2000.
- [31] A. Kirke and E. R. Miranda, "A survey of computer systems for expressive music performance," *ACM Comput. Surv.*, Vol. 42, No. 1, pp. 1–41, 2009.
- [32] 安藤 大地, P. Dahlstedt, M. Nordahl, 伊庭 斉志, "対話型 GP を用いたクラシック音楽のための作曲支援システム," 芸術科学会論文誌, Vol. 4, No. 2, pp. 77-86, 2005.
- [33] M. Niitsuma and H. Saito, "Automatic thorough-bassrealization in baroque music," In *International Computer Music Conference*, pp. 425–428. ICMC, 2007.
- [34] J. ホイジンガ著, 高橋 英夫, "ホモ・ルーデンス 人類文化と遊戯," 中央公論社, 2002.
- [35] 松澤 和光, 堀 浩一, 金杉 友子, 阿部 明典, "ことば工学入門," 人工知能学会誌, Vol. 15, No. 3, pp. 446-455, 2000.

[36] 阿部 明典, "ことば工学研究会 (SIG-LSE)(<特集>研究会総覧)," 人工知能学会誌, Vol. 25, No. 4, pp. 547-548, 2010.

- [37] 阿部 明典, "「ことば工学」特集にあたって,"知能と情報: 日本知能情報ファジィ学会誌, Vol. 23, No. 5, pp. 664-665, 2011.
- [38] 森山 和道, "産総研の女性型ロボ「HRP-4C」開発者座談会 (その 1)," Robot Watch, 2009.
- [39] 森山 和道, "富士ソフト、人を覚えて会話でき「成長」するロボット発売," International Business Times, 2010.
- [40] Honda Robotics, "ASIMO について," http://www.honda.co.jp/ASIMO/.
- [41] 株式会社 HTP , "SEO(検索エンジン最適化) とは," http://www.htp.vc/seo/.
- [42] 杉村 健, 栗山 広治, "刺激の具体性と心像性," 奈良教育大学紀要 (人文・社会科学), Vol. 21, No. 1, pp. 223-236, 1972.
- [43] 五島 史子, 太田 信夫, "漢字二字熟語における感情価の調査," 筑波大学心理学研究, Vol. 23, pp. 45-52, 2001.
- [44] 秋田 清, "種々の刺激語資料の尺度付け," あいり出版, 2012.
- [45] 水島 りか, 柳谷 啓子, 清河 幸子, 川上 正浩, "連想語頻度表 3 モーラの漢字・ひらがな・カタカナ表記語 ," ナカニシヤ出版, 2011.
- [46] 天野 成昭, 近藤 公久, "単語親密度によって分類した日本語単語およびモーラの 統計的分布," 電子情報通信学会技術研究報告. SP, 音声, Vol. 96, No. 160, pp. 9-15, 1996.
- [47] 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, "NTT データベースシリーズ『日本語の語彙特性』第8 巻単語心像性 (1)," 三省堂, 2005.
- [48] D. Sabsevitz, D. Medler, M. Seidenberg, and J. Binder, "Modulation of the semantic system by word imageability," *NeuroImage*, Vol. 27, No. 1, pp. 188 200, 2005.
- [49] 大槻 美佳, "言語機能の局在地図," 高次脳機能研究: 日本高次脳機能障害学会 誌, Vol. 27, No. 3, pp. 231-243, 2007.

[50] T. Nao, O. Jun, and I. Shun, "Evaluating semantic relations and distances in the associative concept dictionary using nirs-imaging," In *In proceeding of: Proceedings of the International Conference on Language Resources and Evaluation*, pp. 3857–3860. LREC, 2010.

- [51] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," J. Mach. Learn. Res., Vol. 3, pp. 993–1022, 2003.
- [52] F. Pereira, G. Detre, and M. Botvinick, "Generating text from functional brain images," Frontiers in Human Neuroscience, Vol. 5, No. 72, 2011.
- [53] The Hoshi Library, "きまぐれ人工知能プロジェクト 作家ですのよ," http://www.hoshishinichi.com/project/1.html.
- [54] 村井 源, 松本 斉子, 佐藤 知恵, 徃住 彰文, "物語構造の計量分析に向けて," 情報 知識学会誌, Vol. 21, No. 1, pp. 6-17, 2011.
- [55] 松原 仁, "ショートショートの星新一らしさに関する考察," 人工知能学会 第 28 回年次大会 発表論文集, pp. 1C2-OS-14a-1pp.1-2. 人工知能学会, 2014.
- [56] 阿部 明典, "特集「修辞の認知科学」に対して," 認知科学, Vol. 14, No. 4, pp. 562-565, 2007.
- [57] 佐々木 康成, 鈴木 孝典, 坂東 敏博, "陰影のあるテクスチャ画像のオノマトペ を用いた質感評価 音韻と画像特徴の関係の統計的分析 ," 認知科学, Vol. 18, No. 3, pp. 477-490, 2011.
- [58] 土屋 誠司, 鈴木 基之, 任 福継, 渡部 広一, "モーラ系列と音象徴ベクトルによるオノマトペの印象推定法," 自然言語処理, Vol. 19, No. 5, pp. 367-379, 2012.
- [59] 小松 孝徳, 秋山 広美, "ユーザの直感的表現を支援するオノマトペ表現システム (エージェントデザイン,<特集>人とエージェントのインタラクション論文)," 電子情報通信学会論文誌. A, 基礎・境界, Vol. 92, No. 11, pp. 752-763, 2009.
- [60] C. Kit, "How does lexical acquisition begin? A cognitive perspective," *Cognitive Science*, Vol. 1, No. 1, pp. 1–50, 2003.
- [61] T. Hashimoto, N. Usui, M. Taira, I. Nose, T. Haji, and S. Kojima, "The neural mechanism associated with the processing of onomatopoeic sounds," *NeuroImage*, Vol. 31, No. 4, pp. 1762–1770, 2006.

[62] H. Yamane and M. Hagiwara, "Web Catchphrase Improve System Employing Onomatopoeia and Large-Scale N-gram Corpus," International Journal of Fuzzy Logic and Intelligent Systems, Vol. 12, No. 1, pp. 94–100, 2012.

- [63] 伊藤 薫, "オクシモロンにおける反義のプロトタイプ性について," 第 18 回年次 大会 発表論文集, pp. 38-41. 言語処理学会, 2012.
- [64] J. Gibbs, Raymond W. and L. R. Kearney, "When parting is such sweet sorrow: The comprehension and appreciation of oxymora," *Journal of Psycholinguistic Research*, Vol. 23, No. 1, pp. 75–89, 1994.
- [65] H. Yamane and M. Hagiwara, "Oxymoron Generation Using an Association Word Corpus and a Large-scale N-gram Corpus," In 14th International Symposium on Advanced Intelligent Systems. ISIS, 2013.
- [66] キム ビンステッド, 滝澤 修, "日本語駄洒落なぞなぞ生成システム "BOKE"," 人工知能学会誌, Vol. 13, No. 6, pp. 920-927, 1998.
- [67] 横川 壽彦, "調音類似に基づく駄洒落候補の自動生成," 第 10 回インテリジェント・システム・シンポジウム, pp. 253-256, 2000.
- [68] 田辺 公一朗, "駄洒落のコンピュータによる処理: 駄洒落生成システムの基本設計," 産能大学紀要, Vol. 26, No. 1, pp. 65-74, 2005.
- [69] 濱田 真樹, 鬼沢 武久, "ファジィ測度・積分モデルで決定した面白さに基づいたなぞなぞの生成," 日本感性工学会論文集, Vol. 7, No. 2, pp. 267-274, 2007.
- [70] 濱田 真樹, 鬼沢 武久, "同音異義語の意味の多様性を構造にもつなぞなぞの生成," 知能と情報:日本知能情報ファジィ学会誌, Vol. 20, No. 5, pp. 696-708, 2008.
- [71] G. Ritchie, "Computational mechanisms for pun generation," in Proceedings of the 10th European Natural Language Generation Workshop, pp. 125–132, 2005.
- [72] A. Waller, R. Black, D. A. O'Mara, H. Pain, G. Ritchie, and R. Manurung, "Evaluating the standup pun generating software with children with cerebral palsy," *ACM Trans. Access. Comput.*, Vol. 1, No. 3, pp. 1–27, 2009.
- [73] H. B. Anthony and O. Ethel, "Automatically extracting word relationships as templates for pun generation," In *CALC '09: Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, pp. 24–31, Morristown, NJ, USA, 2009. Association for Computational Linguistics.

[74] B. Kim, B. Benjamin, C. Seana, N. Anton, S. Oliviero, S. Carlo, R. Graeme, M. Ruli, P. Helen, W. Annalu, and O. Dave, "Computational humor," *IEEE Intelligent Systems*, Vol. 21, No. 2, pp. 59–69, 2006.

- [75] 山根 宏彰, 萩原 将文, "笑いを生むことわざすかしの自動生成システム," 知能と情報:日本知能情報ファジィ学会誌, Vol. 24, No. 2, pp. 671-679, 2012.
- [76] 萩原 正人、"特集号「不自然言語処理 枠に収まらない言語の処理」、" 自然言語 処理、Vol. 19, No. 5, pp. 365-366, 2012.
- [77] 江村 優花, 関 洋平, "マイクロブログにおける感情・コミュニケーション・動作タイプの推定に基づく顔文字の推薦," 自然言語処理, Vol. 19, No. 5, pp. 401-418, 2012.
- [78] 横断型基幹科学技術研究団体連合,"分野横断型科学技術アカデミック・ロードマップ報告書,"経済産業省,2009.
- [79] 乾 健太郎, 浅原 正幸, "自然言語処理の再挑戦:統計的言語処理を超えて (<特集 >テキストの可視化と要約)," 知能と情報:日本知能情報ファジィ学会誌, Vol. 18, No. 5, pp. 669-681, 2006.
- [80] 松本 裕治, "自然言語処理から見た第五世代コンピュータと未来 (特集第五世代コンピュータと人工知能の未来)," 人工知能: 人工知能学会誌, Vol. 29, No. 2, pp. 127-133, 2014.
- [81] D. Lenat, G. Miller, and T. Yokoi, "CYC, WordNet, and EDR: Critiques and responses," *Commun. ACM*, Vol. 38, No. 11, pp. 45–48, 1995.
- [82] G. A. Miller, "WordNet: A lexical database for English," Commun. ACM, Vol. 38, No. 11, pp. 39–41, 1995.
- [83] D. B. Lenat, "CYC: A large-scale investment in knowledge infrastructure," Commun. ACM, Vol. 38, No. 11, pp. 33–38, 1995.
- [84] S. E. Fahlman, "Marker-passing inference in the scone knowledge-base system," In Proceedings of the First International Conference on Knowledge Science, Engineering and Management, KSEM'06, pp. 114–126, Berlin, Heidelberg, 2006. Springer-Verlag.

[85] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," In *In Proceedings of the 5th Conference on Language Resources and Evaluation*, pp. 417–422, 2006.

- [86] F. Bond, H. Isahara, S. Fujita, K. Uchimoto, T. Kuribayashi, and K. Kanzaki, "Enhancing the Japanese Wordnet," In *Proceedings of the 7th Workshop on Asian Language Resources, Association for Computational Linguistics*, pp. 1–8, 2009.
- [87] 荻野 孝野, "EDR 電子化辞書について: 言語資源の観点で," 情報処理学会研究報告,情報メディア, Vol. 98, No. 106, pp. 31-38, 1998.
- [88] 岩波書店, "日本語語彙大系 CD-ROM 版," http://www.kecl.ntt.co.jp/mtg/resources/GoiTaikei/.
- [89] 白井 諭, 大山 芳史, 池原 悟, 宮崎 正弘, 横尾 昭男, "日本語語彙大系について," 情報処理学会研究報告.IM, Vol. 98, No. 106, pp. 47-52, 1998.
- [90] 岡本 潤, 石崎 俊, "概念間距離の定式化と既存電子化辞書との比較," 自然言語処理, Vol. 8, No. 4, pp. 37-54, 2001.
- [91] 小原 京子, 大堀 壽夫, 鈴木 亮子, 藤井 聖子, 斎藤 博昭, 石崎 俊, "日本語フレームネット: 意味タグ付きコーパスの試み," 言語処理学会第 20 回年次大会, pp. S6-1. 言語処理学会, 2005.
- [92] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini, "Building a large annotated corpus of English: The penn treebank," *Comput. Linguist.*, Vol. 19, No. 2, pp. 313–330, 1993.
- [93] ビクター・マイヤー=ショーンベルガー,ケネス・クキエ,"ビッグデータの正体情報の産業革命が世界のすべてを変える,"講談社,2013.
- [94] グーグル株式会社, "Web 日本語 N グラム第 1 版," 特定非営利活動法人 言語資源協会, 2007.
- [95] 工藤 拓, 賀沢 秀人, "大規模日本語 n-gram データの公開," http://googlejapan.blogspot.com/2007/11/n-gram.html, 2007.
- [96] 前川 喜久雄, "代表性を有する大規模日本語書き言葉コーパスの構築," 人工知能 学会誌, Vol. 24, No. 5, pp. 616-622, 2009.

[97] 小原 京子, 加藤 淳也, 斎藤 博昭, "日本語フレームネットにおける BCCWJ への 意味アノテーション," 特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ (研究成果報告会) 予稿集, pp. 513-518, 2011.

- [98] 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.
- [99] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia," *Semantic Web Journal*, 2014.
- [100] 玉川 奨, 森田 武史, 山口 高平, "日本語 Wikipedia からプロパティを備えたオントロジーの構築," 人工知能学会論文誌, Vol. 26, No. 4, pp. 504-517, 2011.
- [101] L. A. Adamic and B. A. Huberman, "Zipf's law and the internet," *Glottometrics*, Vol. 3, pp. 143–150, 2002.
- [102] B. Mandelbrot, "An informational theory of the statistical structure of language," Communication Theory, No. 36, pp. 486–502, 1953.
- [103] 言語処理学会,"言語処理学事典,"共立出版, 2011.
- [104] 北 研二, "確率的言語モデル," 東京大学出版会, 2001.
- [105] E. Rasmussen, "Clustering algorithms. in W.B. Frakes and R. Baeza-yates (eds.)," *Information retrieval : Data structures and algorithms*, pp. 419–442, 1992.
- [106] 榊 剛史, 松尾 豊, 内山 幸樹, 石塚 満, "Web 上の情報を用いた関連語のシソーラス構築について," 自然言語処理, Vol. 14, No. 2, pp. 3-31, 2007.
- [107] 那須川 哲哉, 河野 浩之, 有村 博紀, "テキストマイニング基盤技術 (<特集>「テキストマイニング」)," 人工知能学会誌, Vol. 16, No. 2, pp. 201-211, 2001.
- [108] S. Brin, "Extracting patterns and relations from the World Wide Web," In Selected Papers from the International Workshop on The World Wide Web and Databases, WebDB '98, pp. 172–183, 1999.
- [109] 杉木 健二, 松原 茂樹, "消費者の意見に基づく商品検索," 情報処理学会論文誌, Vol. 49, No. 7, pp. 2598-2603, 2008.

[110] 清水 浩平, 萩原 将文, "名詞と動詞の組み合わせに対する印象推定法," 日本感性 工学会論文誌, Vol. 10, No. 4, pp. 505-514, 2011.

- [111] 高野 敦子, 池奥 渉太, 北村 泰彦, "因果関係に着目した口コミ Web サイトからの評価表現抽出," 人工知能学会論文誌, Vol. 24, No. 3, pp. 322-332, 2009.
- [112] T. Masaaki, T. Kentaro, and D. S. Stijn, "Toward finding semantic relations not written in a single sentence: An inference method using auto-discovered rules," In Proceedings of the 5th International Joint Conference on Natural Language Processing, 2011.
- [113] M. Banko, V. O. Mittal, and M. J. Witbrock, "Headline generation based on statistical translation," In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, ACL '00, pp. 318–325, 2000.
- [114] 廣嶋 伸章, 長谷川 隆明, 奥 雅博, "Web ページのヘッドライン生成のための統計 的要約," 自然言語処理, Vol. 12, No. 6, pp. 113-128, 2005.
- [115] 藤田 篤, 降幡 建太郎, 乾 健太郎, 松本 裕治, "語彙概念構造に基づく言い換え生成 : 機能動詞構文の言い換えを例題に(自然言語)," 情報処理学会論文誌, Vol. 47, No. 6, pp. 1963–1975, 2006.
- [116] 樅山 洋輔, "日本語表現で学ぶ 入門からの認知言語学,"研究社, 2009.
- [117] レイモンド・W・ギブス, "比喩と認知 心とことばの認知科学,"研究社, 2008.
- [118] L. Gandy, N. Allan, M. Atallah, O. Frieder, N. Howard, S. Kanareykin, M. Koppel, M. Last, Y. Neuman, and S. Argamon, "Automatic identification of conceptual metaphors with limited knowledge.," In M. des Jardins and M. L. Littman, editors, AAAI. AAAI Press, 2013.
- [119] 鵜月 洋, "広告文の歴史 キャッチフレーズの 100 年," 日本経済新聞社, 1965.
- [120] 北村 日出夫, 山路 龍天, 田吹 日出硯, "広告キャッチフレーズ," 有斐閣, 1981.
- [121] T. R. V. Foster, "The art & science of the advertising slogan," ADSlogans Unlimited, 2001.
- [122] 深川 英雄、"キャッチフレーズの戦後史、"岩波新書、1991.
- [123] 阿久津 聡, 石田 茂, "ブランド戦略シナリオ," ダイアモンド社, 2002.

[124] C. Kohli, L. Leuthesser, and R. Suri, "Got slogan? guidelines for creating effective slogans," *Business Horizons*, Vol. 50, No. 5, pp. 415–422, 2007.

- [125] T. M. Lowrey and L. J. Shrum, "Phonetic symbolism and brand name preference," *Journal of Consumer Research*, Vol. 34, No. 3, pp. 406–414, 2007.
- [126] 新井 恭子, "関連性理論における「広告のことば」の分析," 東洋大学経営学部 経営論集 (68), pp. 79-91. 東洋大学経営学部, 2006.
- [127] P. Wetzel, "広告にみる日本の言語世界," 名古屋大学大学院国際言語文化研究科 国際シンポジウム「異文化としての日本」記念論文集, pp. 73-82. 名古屋大学大 学院国際言語文化研究科, 2009.
- [128] 越川 靖子, "ブランド・ネームにおける語感の影響に関する一考察 音象徴に弄ばれる私達 ," 商学研究論集, pp. 47-65. 明治大学, 2009.
- [129] 山田 理英, "脳科学から広告・ブランド論を考察する," 評言社, 2007.
- [130] 田邊 学司, "なぜ脳は「なんとなく」で買ってしまうのか?," ダイアモンド社, 2013.
- [131] GMO ペイメントゲートウェイ株式会社 , "EC 市場の進化・拡大を支え、高成長を継続する," http://corp.gmo-pg.com/assets/files/pdf/140206_gmo_pg_kessan.pdf.
- [132] 電通, "2012年日本の広告費 媒体別広告費," http://www.dentsu.co.jp/books/ad_cost/2012/media.html.
- [133] 宮下到, "ネット広告の効果を100%上げ続けるために," 株式会社宣伝会議, 2008.
- [134] 近藤 史人, "AISAS マーケティング・プロセスのモデル化," JSD 学会誌『システムダイナミックス』, 2009.
- [135] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. in Artif. Intell.*, Vol. 2009, pp. 4:2–4:2, January 2009.
- [136] Sloganizer.net, "Instant slogans with our slogan generator," http://www.sloganizer.net/en/, available in June, 2014.
- [137] THE-PCMAN-WEBSITE, "Free slogan generator," http://www.thepcmanwebsite.com/media/free_slogan_generator/index.php, available in June, 2014.

[138] D. Shaw, C. Ardai, B. Marsh, M. Moraes, D. Rudolph, and J. Auliffe, "Electronic mail system for displaying advertisement at local computer received from remote system while the local computer is off-line the remote system," US Patent 5,809,242, 1998.

- [139] H. Jonsson, "User-based semantic metadata for text messages," US Patent App. 13/130,850, 2011.
- [140] J. Bosarge and R. Little, "Method and apparatus for adding advertising tag lines to electronic messages," US Patent 7,599,852, 2009.
- [141] J. Aaltonen and S. Saru, "Systems, methods, network elements and applications for modifying messages," EP Patent App. EP20,080,159,355, 2009.
- [142] 菅 隆彦, 門間 洋二郎, "オンラインショッピングサイト作成支援方法," 公開特許 公報 (A) 特許公開 2005-202783. 日本国特許庁, 2005.
- [143] 酒井 宏明, 麻岡 正洋, 仲尾 由雄, 丸橋 弘治, 山川 宏, "キャッチコピー生成装置 およびキャッチコピー生成プログラム," 公開特許公報 (A) 特許公開 2009-116548. 日本国特許庁, 2009.
- [144] 松平 智史, 萩原 将文, "電子化辞書と遺伝的プログラミングを用いたキャッチコピー作成支援システム,"電気学会論文誌 C (電子・情報・システム部門誌), Vol. 124, No. 1, pp. 164–169, 2004.
- [145] 松平 智史, 萩原 将文, "対話型遺伝的プログラミングと電子化辞書を用いたキャッチコピー作成支援システム," 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 125, No. 4, pp. 616-622, 2005.
- [146] 西原 陽子, 砂山 渡, 谷内田 正彦, "聴講者の興味をひく研究発表タイトルの作成 支援," 言語処理学会年次大会発表論文集, pp. 448-451. 言語処理学会, 2007.
- [147] 西原 陽子, 砂山 渡, 谷内田 正彦, "多くの興味をひく研究発表タイトルの作成支援," 第 21 回人工知能学会全国大会, pp. 2H4-5. 人工知能学会, 2007.
- [148] Y. Nishihara and W. Sunayama, "Title-composing support system for reaching new audiences," In *Data Mining Workshops*, 2008. ICDMW '08. IEEE International Conference on, pp. 816–822, 2008.

[149] 中野 俊亮, 鬼沢 武久, "目を引くキャッチフレーズ生成に関する基礎的研究," 日本感性工学会大会予稿集, p. B71. 日本感性工学会, 2007.

- [150] 中野 俊亮, 鬼沢 武久, "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," 全国大会講演論文集, pp. 201-202. 情報処理学会, 2008.
- [151] 森本 直樹, 玉川 洋輔, 牧野 正治, 韓 東力, "商品の属性と属性値を利用したキャッチコピーの自動作成," 第 16 回年次大会 発表論文集, pp. 74-77. 言語処理学会, 2010.
- [152] 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. P6-19. 言語処理学会, 2013.
- [153] 幅野 裕貴, 浦谷 則好, "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," 言語処理学会 第 20 回年次大会 発表論文集, pp. P7-16. 言語 処理学会, 2014.
- [154] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences*, 2013.
- [155] 久野 寧子, "カタログ・チラシ キャッチコピー大百科," ピエ・ブックス, 2008.
- [156] "Wikipedia 日本語アーカイプ," http://dumps.wikimedia.org/jawiki/, available in June, 2014.
- [157] 京都大学大学院情報学研究科黒橋研究室, "KNB コーパス (Kyoto-University and NTT Blog コーパス)," http://nlp.kuee.kyoto-u.ac.jp/kuntt/, 2009.
- [158] 北九州市立大学国際環境工学部情報メディア工学科上村研究室 , "インタビュー形式による日本語会話データベース," http://www.env.kitakyu-u.ac.jp/corpus/texts/index.html, 1996.
- [159] K. Taku, Y. Kaoru, and M. Yuji, "Applying conditional random fields to Japanese morphological analysis," In *In Proc. of EMNLP*, pp. 230–237, 2004.
- [160] K. Taku and M. Yuji, "Japanese dependency analysis using cascaded chunking," In in Proc. of Natural Language Learning, pp. 63–69, 2002.
- [161] 矢田 晋, "Search System for Giga-scale N-gram Corpus," http://code.google.com/p/ssgnc/.

- [162] 安西 祐一郎, "心と脳 認知科学入門," 岩波新書, 2011.
- [163] 小泉 英明, "脳の科学史 フロイトから脳地図、MRI へ," 角川 SSC 新書, 2011.
- [164] K. Shimizu and M. Hagiwara, "A new electronic dictionary with meaning description of case frame," *JACIII*, Vol. 9, No. 3, pp. 304–313, 2005.
- [165] Z. Kozareva, B. Navarro, S. Vazquez, and A. Montoyo, "UA-ZBSA: A headline emotion classification through web information," In *Proceedings of the Fourth* International Workshop on Semantic Evaluations (SemEval-2007), Association for Computational Linguistics, pp. 334–337, Prague, Czech Republic, 2007.
- [166] A. Abbasi, H. Chen, S. Thoms, and T. Fu, "Affect analysis of web forums and blogs using correlation ensembles," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 20, pp. 1168–1180, 2008.
- [167] C. Kohli, L. Leuthesser, and R. Suri, "Got slogan? Guidelines for creating effective slogans," *Business Horizons*, Vol. 50, No. 5, pp. 415–422, 2007.
- [168] 北村 日出夫, 山路 龍天, 田吹 日出硯, "広告キャッチフレーズ," 有斐閣, 1981.
- [169] N. Ito and M. Hagiwara, "Natural language generation using automatically constructed lexical resources," In *Neural Networks (IJCNN)*, The 2011 International Joint Conference on, pp. 980–987, 2011.
- [170] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Extracting large-scale knowledge bases from the web," In *Proceedings of the 25th VLDB Conference*, pp. 639–650, 1999.
- [171] D. Bollegala, Y. Matsuo, and M. Ishizuka, "A web search engine-based approach to measure semantic similarity between words," *IEEE Transactions on Knowl*edge and Data Engineering, Vol. 23, pp. 977–990, 2011.
- [172] OPEN METI プロジェクト, "技術戦略マップ, ソフト 2 サービス工学分野," 経済産業省, 2010.
- [173] NTT コムウェア, "ニッポン・ロングセラー考 Vol.73 アーム筆入「象が踏んでも壊れない!」印象的なテレビ CM で大ヒット," http://www.nttcom.co.jp/comzine/no073/long _seller/index.html.
- [174] 久野 寧子, "雑誌広告 キャッチコピー大百科 業種別," ピエ・ブックス, 1998.

- [175] 久野 寧子、"新聞広告キャッチコピー大百科 2、" ピエ・ブックス、2005.
- [176] 久野 寧子, "新聞広告キャッチコピー大百科 3," パイインターナショナル, 2011.
- [177] SourceForge.jp, "IPAdic legacy," http://sourceforge.jp/projects/ipadic/releases/, 2007.
- [178] K. Hampton, L. S. Goulet, L. Rainie, and K. Purcell, "Social networking sites and our lives," Technical report, Pew Research Center's Internet & American Life Project, 2011.
- [179] 野口 智雄, "「顧客から顧客へ」 ファンを拡大する SNS 活用法," プレジデント PRESIDENT Inc, http://www.president.co.jp/pre/backnumber/2011/20111003/20597/20603/, 2011.
- [180] facebook Inc., "facebook," https://www.facebook.com/, available in June 2014.
- [181] 小阪 裕司, "感性マーケティング"でビジネスを変えよう!," PHP Biz Online 衆 知, http://shuchi.php.co.jp/article/67, 2010.
- [182] P. André, M. Bernstein, and K. Luther, "Who gives a tweet?: evaluating microblog content value," In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pp. 471–474, New York, NY, USA, 2012.
- [183] 沢井 康孝, 山本 和英, "文書に対する大衆の興味の強さの推定," 自然言語処理, Vol. 15, No. 2, pp. 101-136, 2008.
- [184] 河原 大輔, 黒橋 禎夫, "自動構築した大規模格フレームに基づく構文・格解析の 統合的確率モデル," 自然言語処理, Vol.14, No.4, pp.67-81, 2007.
- [185] 黒橋・河原研究室, "日本語構文・格解析システム KNP," http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP, available.
- [186] T. Kudo, "MeCab: Yet another part-of-speech and morphological analyzer," http://mecab.sourceforge.net/, available.
- [187] H. Yamane and M. Hagiwara, "Preference estimation for SNS fan pages based on statistical information," In Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on, Vol. 1, pp. 397–401, 2013.

[188] facenavi, "Facebook 企業/ブランド公式ページ ベスト 500," http://facebook.boo.jp/featured-company.

- [189] 安藤 隆, 岡本 欣也, 仲畑 貴志, 前田 知巳, 小野田 隆雄, 佐々木 宏, 山本 高史, 児島 令子, 一倉 宏, 澤本 嘉光, "日本のコピーベスト 500," 宣伝会議, 2011.
- [190] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," In *Advances in Neural Information Processing Systems* 26, pp. 3111–3119. Curran Associates, Inc., 2013.

付録 A

統計的手法を用いたキャッチコピー自動生成シ ステム

A.1 パラメータ

本章では,第2章の提案システムにおけるパラメータの詳細について述べる.

A.1.1 分野特有の関連語取得

関連語取得におけるパラメータの値の詳細について説明する.

わかりやすさ

わかりやすさはキャッチコピーにおいて重要な条件である.第2章の提案システムにおいて,Web 日本語 Nグラム [94] における $1 \operatorname{gram}$ の閾値以下のものは除外しているが,その根拠を示す.図 A.1 に実際の単語の順位と頻度の関係を示す.なお横軸に順位を,縦軸に頻度の対数をプロットしている.図からわかるように,普段よく用いられそうな単語は非常に多く現れ,そうでない単語が実際にはかなりの割合を占めていることがわかる.頻度が一定値以下になると,一般では知らないような単語の割合が増える.今回は経験的に頻度が 300 未満を足切している.なお,頻度が 300 程度のものとして (ショコラ・シェーヌ: 300),(レナート・ニルソン: 200) が挙げられる.この経験則に従って,閾値が 300 以上のものを,次のキャッチコピーらしさおよび対象分野らしさのステップで用いている.

キャッチコピーらしさ・対象分野らしさ

取得したキャッチコピーらしい単語を示す S_c とその対象分野らしい単語を示す S_t のスコアに関して、経験的に $S_c+2\times S_t$ の値が大きい順に関連語取得のステップに渡し

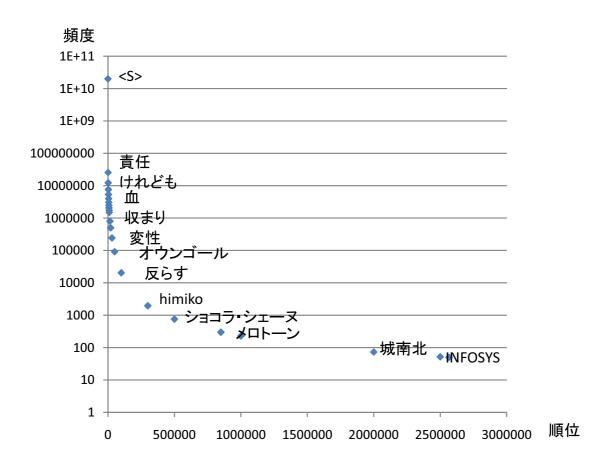


図 A.1 単語の順位と頻度の関係

ている.なお,共起頻度の計算コストの観点から上位 400 個を選択している.(全分野の単語を計算する場合, χ^2 値を求める計算時間が数ヶ月と予想されたため)

A.1.2 キャッチコピー候補生成

生成

キーワードと関連語をリクエストする際に ,ssgnc [161] において語順はUNORDERED , 生成は 7gram でそれぞれにつき上限は 1,000 個としている .

再生成

候補数が閾値以下の時に,再生成を行う.回は 10,000 個以下のときに再生成を行っている.また,再生成の時には,最初の生成の段階で得られた候補の語尾 2 gram と関連語を,ssgnc における ORDERED を用いて最大 50 個取得している.10,000 個を超えると,再生成は完了する.

A.1.3 キャッチコピー候補選択

キャッチコピーの候補を3段階で選択している.

第1候補選択 文構造スコア利用

最初に,キャッチコピーにおける文構造スコアと複数コーパスを用いた文構造スコアの線形和をとり,合計の情報量が100.0以下のものを次のステップに渡す.なお,キャッチコピーにおける文構造スコア(順方向と逆方向)を S_{If} , S_{Ib} ,複数コーパスを用いた文構造スコア(順方向と逆方向)を S_{IIf} , S_{IIb} とし,合計文構造スコア S_{Gtotal} は

$$S_{Gtotal} = 1.0 \times S_{If} + 1.0 \times S_{Ib} + 0.3 \times S_{IIf} + 0.3 \times S_{IIb}$$
 (A.1)

として計算を行っている.

第2候補選択 関連度スコアおよび文構造スコア利用

次に, χ^2 (関連度) と文構造スコアの線形和をとり,上位 $1{,}000$ 個について第 3 候補選択に回している.なお,第 2 候補選択の式は, χ^2 値の関連度を S_{χ^2} として,

$$S_{both} = S_{Gtotal} + \log_2 S_{v^2} \tag{A.2}$$

計算を行う.

第3候補選択 斬新な名詞の組み合わせスコア利用

相互情報量 (西原らの定義するもの [146-148] . 分母分子が反転) が $0.05 \le \alpha \le 1.0$ のものにから,ランダムに抽出し最終的なシステム生成のキャッチコピーを取得している. $0.05 \le \alpha \le 1.0$ のものを選んだ理由として,相互情報量が大きく奇抜すぎるものは理解が出来ずキャッチコピーとして相応しくないと考えられるからである.同様に,相互情報量が小さいものはありきたりな文になり,これもキャッチコピーに相応しくないと判断した.

A.2 評価実験

本章では,評価実験の詳細について述べる.

A.2.1 評価実験に用いたキャッチコピー

食品分野のキャッチコピー

以下に食品のテーマで、評価実験に用いたキャッチコピーの一部を示す、

システム生成

- 1. 日本初の味付きかき氷を
- 2. フルーツジュースをお腹一杯摂る
- 3. ワインがジュースのような味
- 4. 季節の果実でジュースにこだわる
- 5. 香りのジュースの出来上がりです。
- 6. タクダーネ社の大人気健康ジュース果汁
- 7. かき氷とフランクフルトも夏ですね
- 8. 夏の期間だけかき氷屋さん
- 9. 菓子やジュースが買える山形物産
- 10. ジュースの味が忘れられない

ベースライン

- 1. 生ジュースを使い、かき氷の
- 2. フレッシュジュースとかき氷のお店
- 3. 洲崎にあるフレッシュジュースとかき氷
- 4. ジュースにしたり、かき氷に
- 5. かき氷、沖縄ぜんざい、ジュース、

- 6. ジュースはもちろんのこと、かき氷
- 7. かき氷売ってたり、ジュースが
- 8. マンゴージュース&フルーツシャーベットフルーツかき氷
- 9. ジュース&フルーツシャーベットフルーツかき&(
- 10. ジュースが楽しめます かき氷も

プロ

- 1. 舌で知る江戸の味。
- 2. あのひとにぴったりの美味しさを贈りたい。
- 3. 刻み栗入りの栗饅頭をこし餡で包み、栗を模して焼き上げた栗饅頭でございます。
- 4. イヴのお相手はキラキラ宝石のようなケーキたち。
- 5. くいだおれの街が自信を持ってオススメするなにわの味を召し上がれ!
- 6. 夏もちっ。夏ふわっ。夏ヨーグルト。
- 7. お得意様だけの"おもたせ"の味。
- 8. ごはんがおいしくなるスーパー
- 9. 春って、ぷるぷる。
- 10. すこやか乳酸菌

交通・レジャーのキャッチコピー

以下に交通・レジャーのテーマで,評価実験に用いたキャッチコピーの一部を示す.

システム生成

- 1. 温泉は世界有数のラジウム温泉
- 2. ホテルの滞在に温泉の安らぎ
- 3. 湯を日常に、温泉付
- 4. 自噴の日の出温泉は24時間
- 5. 温泉街の夜景を眺め露天
- 6. 黒湯の天然温泉は心
- 7. 森の中の温泉リゾート。
- 8. 湖畔の温泉宿は全て自噴
- 9. あなたの心も体もリラックス
- 10. 極上の温泉体験が毎日できる

ベースライン

- 1. 本物の別府温泉でリラックス】
- 2. 層雲峡温泉の湯でリラックスも

- 3. 温泉休暇 リラックス『楽山館
- 4. 夜は温泉施設でリラックスを
- 5. 温泉の恵みでリラックス (要
- 6. 温泉でリラックス感をどうぞ。
- 7. 週末リラックスできる至宝の温泉宿
- 8. 天然温泉大浴場とリラックスラウンジ
- 9. 温泉岩風呂でリラックスできます
- 10. 天然温泉でリラックス 旅の

プロ

- 1. その孤島は、瑞々しい自然の楽園だった
- 2. 「世のもり洗う四万温泉」と、上毛カルタに詠まれた名湯
- 3. どうしてだろう。はじめてなのにすべてが懐かしい。ここはずっと思い描いてい た夢のかたち。
- 4. 大自然が創り出す芸術にふれる旅。
- 5. 湯煙おおう、世界有数の温泉の町へ。
- 6. 素足が心地いい、パウダーサンドの浜辺
- 7. ちひろが願ったこと
- 8. しれとこ 地球上で選ばれしその存在
- 9. 無色無臭のさらりとした湯に、すっぽり包まれる。
- 10. "何もしない" という贅沢な休日... 夢をかなえる旅はいつもルック JTB から

メディアのキャッチコピー

以下にメディアのテーマで、評価実験に用いたキャッチコピーの一部を示す、

システム生成

- 1. 親子の愛を描いた感動
- 2. 心を幻想と感動の世界
- 3. 世紀を舞台にした映画
- 4. 誰でも心に残る名作映画
- 5. 感動があなたの心にあふれる
- 6. 大人のピアノ永遠の映画音楽
- 7. 心の感動を写真で表現
- 8. 感動のストーリーが砂時計を通して心
- 9. 奇跡と感動の冒険物語。
- 10. 愛と涙の感動ストーリー。

ベースライン

- 1. あの大好きな映画の感動シーン
- 2. サントラが映画の感動を何
- 3. 映画を見ていたく感動。
- 4. たまにディズニーの映画に感動
- 5. 、恋愛、感動映画あります
- 6. 感動、涙のおすすめ映画
- 7. 映画を見て感動したり
- 8. 映画に感動した身として
- 9. 久しぶりに感動した映画でし
- 10. 、感動した映画、本

プロ

- 1. 人はみんな幸せになれるようにできているんですよ
- 2. 大人は誰も愛してくれない。
- 3. 海の向こうは不可思議の国だった。
- 4. 結婚とは、夫婦とは?苦しみは分かち合い、歓びは一人占め これがホントの '夫婦のススメ'
- 5. スマイルな朝を、あげる。
- 6. この夏、あなたの時計を巻き戻してみませんか?
- 7. 「言葉」は、いつでも誰にでも、力をくれる。この世界を生き抜く力を。
- 8. 喫茶店始めたぐらいじゃ、人生変わらないと思ってた。
- 9. 美人じゃなくてもこんなに元気、こんなに幸せ。アンは女の子の永遠の憧れです。
- 10. その浮世絵は、日本人の心の風景になった。

A.2.2 失敗例

キャッチコピーとは呼べない失敗例を示す.

- ジュースやガムを買う時。
- かき氷とフランクフルトも夏ですね
- かき氷やアイスに極上の黒
- お茶やジュースを避け、水
- 味のかき氷が好きですか

失敗例として,文構造は頻出するものだが,普通の文章になっているものが挙げられる.また,形態素解析器 MeCab が「~味」の~に該当すべきものをカットしてしまっているため,不自然な文章が選択されることも散見された.これらを改善するために,形態素解析器の辞書の拡充や,システムにおいて意味も考慮に入れる必要性があることが示唆される.

付録 B

Web上の文情報を用いたキャッチコピー自動生成システム

ここでは,第3章の提案システムにおける詳細について述べる.

B.1 マイニングされた文の例

述語項構造「景色、美しい」から取得された例を以下の引用に示す.

「景色が美しくって」を含む Web 文書

私はあの橋からの景色が美しくって、気持ちよくって、楽しかったです 電車の窓から見える、特に日が暮れ始めたころの景色が美しくって。駅前のアウトレットモールも夕暮れ時は景色が美しくって沈丁花の花の甘い香りが、ふんわり漂ってきます桜の花も一気に開花し、景色が美しくって、通勤が楽しい会社のお花見はあるけど、わんこちゃんたち連れてお弁当持ってお花見に行きたいなぁ~春はわくわく、気分が明るくなります Android 携帯からの投稿加賀は何処も景色が美しくって、バスに揺られてるだけで幸せになります。 政次郎尾根の展望ない登りや大倉尾根を下る事を考えれば、景色が美しくってとても楽しい。山、そして谷の景色が美しくって!景色が美しくって、気が散るかな~~~って、ちょっと思いましたが、とてもまじめに練習を重ねていってくれます。 ウブドってのどかで景色が美しくって、人は優しくて…とイメージしていましたが、junjun さんのお写真を見て、もう行きたくなっちゃいました。保津川は一度行ったことがありますが、本当に景色が美しくって、感動した覚えがあります。下界の景色が美しくって … ヒトって本当にちっぽけなんですよねぇ~そちらは景色が美しくってよろしいなぁ? Commentedbyvon-braunat2009-02-2400:37 最近忙しくご無沙汰しています。オアシス 21 の最上階から見る夕暮れの栄の景色が美しくって…寒さの中ぼーっと見入ってしまいました。Commentedby miriy unat 2007-05-2800:14 景色が美しくって、歴史が深く、おいしそうな香りまで想像してしまいました。2 日共天候に恵まれて、緑がきれいで、風がそよそよしていて、景色が美しくって、電車も座れてご飯もおいしくって最高でした!

「景色が美しくて」を含む Web 文書

ゆらりと暮らす日々**:朝の景色が美しくて朝の景色が美しくて今日はあんまり空など景色が美しくて、う れしくて思いついたままにしゃべりながらを日記にしてみました。大学の池の景色が美しくて飄々としていて **陶酔させられて、たまに薄い霜はぼんやりして、更に夢の世界と同じに、たくさんの青年の結婚の契りを結ぶ地** で、いくらかのロマンチックな風情を増やします。朝ランしながら景色が美しくて落淚。ステージから見た景色 が美しくてみんなの笑顔がもう忘れらんない。2匹目のドジョウを狙っているわけではありませんが余りにも景 色が美しくて思わず降りてカメラを取り出してしまいます毎日毎日ただ景色が美しくて国家ソフトウェア産業 基地インキュベーターセンターは景色が美しくて、空気の清新な紫金山の東側 徐庄に位置しており、南京市 の 2006 年の十大近代的なサービス業の重点プロジェクトの 1 つである。植物園は景色が美しくて、空気がきれ いと思います。植物園は景色が美しくて、空気がきれいだと思います。植物園は景色が美しくて、空気がきれい だと思いました。すごくノスタルジーな気持ちになりますね昔この映画を見てすごく感動しました切なくて優 しくて尾道の景色が美しくて涙がとまりませんでしたありがとうございました? 本物の心地良さについて考え ていくと、見える景色が美しくて、見えない景色も整っているってこと、環境がすっきりしていると自分の思考 もスッキリ! 車窓の景色が美しくて、太陽神ラーに感謝 降り注ぐ景色が美しくて、ちらちらと雪が舞う中で撮 りました。今朝は多摩川沿いを走ってみたんだけど、景色が美しくて感動した。ガイドブックに赤土の島で、景 色が美しくて・・・・と書いてあり、そんなに移動中、助手席から常に地中海が見えていて顔をあげて窓の外を見る

度に景色が美しくてハッとする。岡山市と広島市の間にある尾道は、坂と山と海の景色が美しくて大好きな街です。何より山道の景色が美しくてドライブも全然飽きません。それにしても、窓から見える景色が美しくて、まさに癒しの景色です。大町はご飯が美味くて、景色が美しくてとっても楽しいところだから、僕の Tweet を見て旅行の参考にしてね! 遊園地から見える日本海の景色が美しくてしばしうっとり。でもやっぱりいくら景色が美しくても、人が多すぎるのは疲れますね。天気が良いと、カメラ越しに見る景色が美しくて、何度もシャッターをきってしまいます。特に景色が美しくて、おいしいものがある場所を探しています。お部屋へ戻ると既に16:00.寝室のガラス戸に写り込んだ景色が美しくて・・・一枚。島原周辺・・・とにかく景色が美しくて。

B.2 取得された単語群

ケーキ、おいしいから取得された Web 文書に対して, LDA を用いてクラスタリングを行った際の単語群を示す.

- トピック 1 こと コメント トラック リンク 日 グッジョブ 日時 方 ワタ シ 沖縄 次 中 イベント 今 表示 開催 何 仕事 今回 みたい 分 サイト 予定 あたし ここ 平成 紹介スレ 会場 ページ メール 名 先 豌 気 設定 介護 的 商品 必要 産業 無事 萓 手 話定例 地域 電話 アップ
- トピック 2 月 日 こと 私 旅行 日記 最新 名前 写真 更新 投稿 読者 名無し 中 結婚 世界 次ここ 場 屋 予約 返信 確認 日々 トラック アメリカ 最近 数 分 何 彼 会社 テーママイ 火 土 あと 型 市 金 本 笑 札幌 エリア やつ 禁止 手作り 友達 イベント素敵
- トピック 3 ケーキ 日 レシピ 料理 質問 人 円 者 ブログ 誕生 私 表示 知恵 分 スイーツ 簡単 お菓子 グルメ ホテル 和泉 あなた 型 お気に入り シフォン ロールケーキ 数 クック 選択 趣味 おすすめ すべて ココナッツ まとめ 日本橋 お出かけ 美容 恋愛 ぉ 編 悩み 全体 品 材料 プライバシー 人気 順 件 ニックネーム 解決 飲食
- トピック 4 東京 区 月 名 屋 県 日 人 駅 最近 味 肉 コメント 大分 紹介 魚 円 気 レストラン グルメ 寿司 みたい 写真 チョコレート トラック 女子 公園 ところ 広島 者 金沢 酒 バック 携帯 水 雰囲気 近く 餃子 駅前 人形 気持ち 菓子 羊毛 今 軒 系 ー 何 焼肉 玉
- トピック 5 月 様 ケーキ 詳細 コメント 人 カード 時 今日 青 回答 今津 場 気 方 家 子供 今回 予約 カフェ 前 たち ロールケーキ コーヒー とき 散歩 ロール 大変 写真 お茶 お菓子 ー 気持ち 映画 最新 埼玉 風 ドラマ 購入 友達 作品 円 鎌倉 ファン 水カテゴリ メール 店員 記事
- トピック 6 口コミ ケーキ 点数 利用 駅 おすすめ ランキング 規約 位 金額 カフェ 用途 マイレビュアー 連絡 県 評価 愛知 レストラン グルメ 写真 件 訪問 代 レビュアー

日 栄 ランチ ユーザー 的 周辺 度 男性 洋菓子 地図 席 こと リスト 円 雰囲気 ジャンル 編集 中 特集 レビュー 西葛西 削除 人 可 可能 喫煙

- トピック 7 ケーキ クリーム 中 店 生地 感じ 感 チョコレート 今回 投稿 リンク 用 コメント 濃厚 バイキング バック お気に入り デザート 私 焼き ソース 気 上 的 方 値段 ショートケーキ 入り バター みたい チェック ロ ー ロール 度 抹茶 メニュー 満 足 新宿 時間 サラダ ショコラ 階 皿 店員 風味 紅茶 注文 インド ショーケース
- トピック 9 月 花 教室 中学 猫 君 運動 美 身 健康 動物 選手 すき グッジョブ 展 豆腐 学生 綾 冷蔵庫 月間 トップページ オ プロ 士 回 比嘉 幼稚園 ベーカリー 杯 トマト 社長 アニメ 大城 バスケ コメント チャレンジ 保育 力 金城 フェア 雑記 サンタ 酵素 身体 会議 位 収納 宮城 企業 中部
- トピック 10 ゅ ょ 迪 こ 遙 ぉ 阪 蠎 律 代 峨 蜿 倥 医 險 ゆ 莉 抵 偵 翫 雋 譬 豌励 騾 邯 壹 邏 ヶ ヵ 豐 蜉 辟 撰 晁 蟄 怦 隧 蜀咏 窶 舌 蠢 昴 呈 菴懊 霑 ヰ 閏 鬟 筵 リ月 人 何 誰 ケーキ 時間 ブログ 円 時 ところ 車 店 今日 感じ クリスマス 家 気 お願い 自分 目 最近 どこ 後 買い物 街 記事 説明 方面 近く 道 今回 ここ ポチッ 言葉 ページ ちゃ 弁当 事 間 用 バス ダイエット 品 的 たま 風 旅行 歩 場

B.3 生成例

キーワード「景色」、「美しい」

- きれいなあなたが居るところ
- きれいなあなたが止まるところ
- ぴったりなあなたが止まるところ
- 確かなあなたが立ち止まるところ
- 幸せなあなたが伝わるところ
- 確かなあなたがくるところ
- おだやかなあなたが撮るところ
- すてきなあなたが走るところ
- 静かなあなたが立ち止まるところ
- 忍はうつくしい。
- 美女は空しい。
- 水族館はたまらない。
- 重点は無い。
- 気候は空しい。
- ユーロは深い。
- ぶどうは短い。

- 忍は涼しい。
- 山頂は美味しい。
- ひとときはつまらない。
- 以前は心のお客。
- さっきは画像の秋。
- 今日は駅の日。
- 以前は思い出の自分。
- ほかは情報の皆様。
- 日曜日は心のお客。
- すべては情報の費用。
- 今度は画像の秋。
- その後は車の自分。
- よい花にできることは面白い。
- 楽しい割に見えることは楽しい。
- 遠いあたりに感じることは遅い。
- 美しい割にいることはよい。
- 珍しい文化に向けることはうれしい。
- 高いエリアに居ることは珍しい。

キーワード「ネタ」、「面白い」

- あなたは三回やめますか!?
- あなたは千回みますか!?
- あなたは三回誉めますか!?
- あなたは千回売れますか!?
- 実況を読もう。登録にしれよう。
- 爆弾を読もう。質問に間違えよう。
- 夜空をはさもう。期待に笑えよう。
- オチを読もう。設定にできよう。
- 動機を読もう。話に盛り上げよう。
- 申レをはさもう。リンクにくれよう。
- ボロボロをはさもう。メールに許せよう。
- オチを楽しもう。更新に盛り上げよう。
- いまへ。一歩ずつ、前へ。
- 昨日へ。一歩ずつ、昨日へ。
- 毎日へ。一歩ずつ、今日へ。
- 全てへ。一歩ずつ、全てへ。
- 前へ。一歩ずつ、時間へ。
- 昨日へ。一歩ずつ、昼へ。
- 今へ。一歩ずつ、今へ。
- 当日へ。一歩ずつ、昨日へ。
- ついにフリーとオードリーテーマ園にいってしまいました!
- ついに武士と明國園にいってしまいました!
- ついに土管と吉野家リメイク園に行ってしまいました!

- ついに客と大阪性格園にいってしまいました!
- ついに作家とじゅん単発園に行ってしまいました!
- ついに国道と日パンクブーブー園に行ってしまいました!
- 知りたいことが増えてる子は忙しい!
- 分かりたいことが読めてる気は多い!
- 関わりたいことがくれてるゲームは高い!
- ハマりたいことが出てる掲示板は詳しい!
- わかりたいことが消えてる印象は上手い!
- 知りたいことが見てる気持ちは楽しい!

キーワード「ケーキ」「おいしい」

- 季節に確認を、プリン久々に二円を。
- 口に仕事を、女子風呂に一円を。
- ライトに紹介を、始まり右に何円を。
- ライトに紹介を、レッスン帰りに何円を。
- パンに追加を、イマイチ自分に三円を。
- 音楽にコメントを、地上人に一円を。
- ライトに紹介を、感じ隣に三円を。
- 季節に確認を、イマイチ自分に三円を。
- 気にお願いを、ケーキ生地に二円を。
- フォークがダメ。
- 小倉が安直。
- おばさんが安直。
- キリが確か。
- カラフルがシンプル。
- クリが普通。
- 小倉が苦手。
- カラフルが確か。
- かぼちゃが正直。
- みんなを上品にする苺
- みんなを爽やかにするカラフル
- みんなを爽やかにする陸
- みんなを幸せにするかぼちゃ
- みんなを静かにする豆乳
- みんなを安直にする喉
- みんなを上品にするおばさん
- みんなを大好きにする珈琲
- みんなをヘルシーにする相方
- みんなをメジャーにするクリ

付録 C

SNS上の嗜好情報を用いたキャッチコピー自動 生成システム

ここでは,第4章の提案システムにおける詳細について述べる.

C.1 嗜好情報取得の詳細

嗜好情報を取得した際の詳細についてここで述べる.

C.1.1 全ファンページの傾向

形態素解析器 MeCab [159, 186] を用いて、元のものと推定される「いいね!」の数を 比較する.

実験条件

- Facebook 企業/ブランド公式ページ [188] に掲載されている 100 のファンページ 上の会社 (2013 年 5 月 11 日のもの) を使用する
- 名詞・動詞・形容詞のみ用い、日本語ストップワードとして沢井ら [183] の「する」「ある」「よる」「いる」「なる」「いう」「みる」「できる」を利用する
- 単語情報のみ用いる
- 式 (4.5) と (4.7) による想起の違いを比較する

実験条件

全ファンページの結果を図 C.1 に示す. また, 式 (4.5) の使用時よりも , 第 4 章の IDF によって正規化した式 (4.7) を用いることにより, 相関係数 0.336 から 0.490 と向上が見られた.

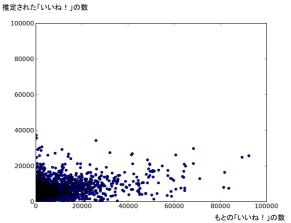
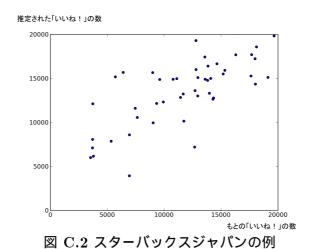


図 C.1 日本語での全ファンページの結果



実験条件

- 形態素解析器 MeCab を用いる
- 同じ 100 の会社のファンページが入力される
- 日本語ストップワードを除いた名詞・動詞・形容詞
- 単語情報のみを利用し、式 (4.7) を用いる

実験結果

図 C.2 にスターバックスジャパンの例を示す. 相関係数は 0.693 であった.

C.1.2 嗜好情報のまとめ

Facebook における企業のファンページへのユーザの嗜好について分析を行った. 嗜好の指標となる「いいね!」の数を再現するために、全体のファンページとそれぞれの

ファンページについて bag-of-words モデルを用い、単語に対して嗜好の度合いを振り分け、投稿に対する「いいね!」の再構築を行った.

まず、全ファンページについて単語と嗜好に関して相関があるか分析を行った. 単語に対し IDF による重み付けを行うことにより、再現された「いいね!」の数は実際の投稿の「いいね!」により近い結果となった. 具体的には、日本語については 0.34 から 0.49 へと向上が見られた. このことにより投稿における単語と嗜好の間には、少なからず相関があると考えられる. また、IDF による重み付けは有効であることから、より低頻度の単語の方が「いいね!」に対する寄与が大きいと考えられる.

それぞれのファンページに対しても分析を行った。それぞれのファンページの相関係数の平均は 0.69 であった。

C.2 生成例

評価実験に用いたキャッチコピーについて例示する.プロ作成のもの,システムが 生成したものを人手で選択したもの,システム生成のものをランダムに選択したもの, 嗜好情報以外の情報を用いて選択したもの,嗜好情報を含めた全ての情報によって選 択したもの,について示す.

コーパス内(プロ作成)

- 1. 母の日は母と私の、きずなの日。
- 2. 夏のマナーを知っている一着。
- 3. 贈る理由は、夏のにおいがしたから。それだけ。
- 4. ゴルフは、ボールを選べるスポーツです。
- 5. 世界のサッカーを変える時が来た。
- 6. 「かけがえのない命」が生まれた日。
- 7. この華やかな夜の裏側で、いまも「人身取引」が行われている。
- 8. 私、若い頃は結婚に向いてないと思っていたんですよ。
- 9. 達成感の先にあるもの。それは新たな挑戦。
- 10. がんばりたい、現役合格したい、という人たちのためにここはある。

人手選択

- 1. 母の日は母と婦人の、スタートの日。
- 2. 夏のスタートを知っている一着。
- 3. 夏の気持ちを知っている一着。
- 4. 世界のスタイルを変える時が来た。
- 5. この華やかな夜の裏側で、先週も「イベント開催」が行われている。

- 6. あなたの一票には力がある。
- 7. がんばりたい、卒業ライブしたい、という人たちのためにここはある。
- 8. マラソンスタートで、完走しよう。
- 9. 「かけがえのないイベント」が生まれた日。
- 10. 母の日は母と私の、ギフトの日。

システム生成 (ランダム)

- 1. 母の夏は母と私の、対象の日。
- 2. 日のスタートを知っている一着。
- 3. 日の実施を知っている一着。
- 4. キーボードやキーボードよりも、つののうちのスタイルと他を買ってもらいたい
- 5. 使用は、キーボードを選べるスタイルです。
- 6. この華やかな夜の裏側で、先週も「到着予定」が行われている。
- 7. 特徴の数だけ紹介販売を業務します
- 8. ジュニアからの暴力に苦しんでいるのはキングだけではありません。
- 9. 「ハッキリ言ってお母さんとわたしの"紹介"は違うのよ」と思いながら、お父さんみたいな男性を探してるわたしです。
- 10. 構成で、応援しよう。

システム生成(嗜好情報なし)

- 1. 紳士の日に、紳士の服。
- 2. 夏の注文を知っている一着。
- 3. 他は、画面を選べるスタイルです。
- 4. 世界の他を変える時が来た。
- 5. 私、若い頃は紹介に向いてないと思っていたんですよ。
- 6. 応援応援でスタートしよう。
- 7. 日のセットを知っている一着。
- 8. 贈る対象は、日のにおいがしたから。それだけ。
- 9. 母の夏は母と婦人の、販売の夏。
- 10. 特徴の数だけ紹介販売を業務します

システム生成 (嗜好情報あり)

- 1. 卒業ライブの先にあるもの。それは新たな卒業。
- 2. 夏のセットを知っている一着。
- 3. 母の夏は母と私の、感謝の夏。
- 4. 私、若い頃は式に向いてないと思っていたんですよ。
- 5. 夏の注文を知っている一着。

- 6. 文字は、キーボードを選べるスタイルです。
- 7. 世界の文字を変える時が来た。
- 8. この華やかな夜の裏側で、昼も「労働参加」が行われている。
- 9. 「かけがえのない渋滞」が生まれた日。
- 10. 「ハッキリ言ってお母さんとわたしの"平均"は違うのよ」と思いながら、お父さんみたいな男性を探してるわたしです。

付録 D

出力に対する専門家の意見

それぞれのシステムの出力に関して,専門家の松尾順氏に意見を伺った.3つのシステムの中でも,特に第3章のシステムの「爆弾を読もう。質問に間違えよう。」「さっきは画像の秋。」のような出力がちぐはぐな組み合わせで面白いのではないかとのことであった.

また,それぞれのシステムは以下のようにエッセンス型,サプライズ型,共感型に分けることが可能ではないかとのことであった.

エッセンス型 (第2章: 統計的手法を用いたキャッチコピー自動生成システム)

トピックのエッセンスをアピールすることがこのタイプのキャッチコピーの特徴である、最近流行りのマンションポエムがこの代表的な例と考えられる、

例えば「信長公が望んだ高台にそびえ立つマンション」が挙げられる.これは,気 取ることによって対象の良さを伝えている.

統計的なキャッチコピーシステムはこのような出力を得るのに適しているのではないか。他には「遮るものがない見渡す限り360°の視界から見える、清盛公の大地」等が挙げられる.

最近,出版社が要約化の技術を進めているが,このような出力は究極の要約と考えられる。

サプライズ型 (第3章: Web 上の文情報を用いたキャッチコピー自動生成システム)

アイキャッチ効果を狙うことが出来る意外な組み合わせのものがこのタイプのキャッチコピーの特徴である.具体的には,有名なキャッチコピーである「おいしい生活」のような斬新な組み合わせを考慮することである.

例えば,エッセンス型のキャッチコピーの例として「温泉は世界有数のラジウム温泉」があるが,これを変換して「赤ちゃんもラジウム好き」にするとサプライズ型となる.意外性を基につかみを狙う手法と考えられる.ちぐはぐな組み合わせを利用するオクシモロンの要素を持つ.

共感型 (第4章: SNS 上の嗜好情報を用いたキャッチコピー自動生成システム)

受け手の隠れた思いを代弁してくれるものが,このタイプのキャッチコピーの特徴と考えられる「温泉は世界有数のラジウム温泉」に対して「ラジウムでほっこりしますよ」が共感型に対応する.受け手の潜在ニーズを代弁しているため,提示された際に「確かに!」と思うインサイト型であるとも考えられる.

また,サプライズ型と共感型は親和性が高く,例えば「プール冷えてます」は両者のハイブリッドであると考えられる.一方で,エッセンス型は伝達を主な目的としているため,前出の両者との領域と重複しない.

以上より「本論文における提案システムは,それぞれ先に挙げた性質を持っているため,その時々の状況によって使い分けると良いのではないか」とのコメントを頂いた.