

ユーザレビューを用いた TVCM の特徴語抽出に基づく時系列分析

澤田 悠治[†] 北山 大輔^{††} 角谷 和俊[†]

[†] 関西学院大学総合政策学部メディア情報学科 〒669-1337 兵庫県三田市学園 2-1

^{††} 工学院大学情報学部システム数理学科 〒163-8677 東京都新宿区西新宿 1 丁目 24-2

E-mail: [†]{fps41167,sumiya}@kwansei.ac.jp, ^{††}kitayama@cc.kogakuin.ac.jp

あらまし TVCM は商品・サービスの特徴をアピールすることで、ユーザの購買に影響を与えられる媒体とされる。また、ユーザレビューにはその TVCM を見て購入したユーザによる投稿も含まれており、TVCM 配信後で特に投稿されている可能性がある。そこで本研究では、TVCM でアピールされた特徴とユーザレビューで評価される特徴の推移に何らかの関連性があるとして、TVCM の字幕・セリフから抽出した特徴語の TVCM とユーザレビューでの時系列な出現パターンを比較することで、TVCM でアピールされた特徴に対するユーザの反応の分析を行った。その結果、TVCM でアピールされる特徴がユーザレビューで反応を得るまでに時間差が生じ、その特徴の内容によって生じる時間差の大きさが異なることが分かった。

キーワード テキスト処理, TVCM, 特徴語, 不満情報

1 はじめに

近年、同一のカテゴリの中でも多様な特徴を持った商品・サービスが流通されている。その中で、「消臭が出来る洗剤が欲しい」、「眠くなりにくい薬が良い」などのユーザの様々な動機が広告媒体などに誘引されることによって、ユーザに適した商品・サービスが選択されていると考えられる。また、EC サイトや SNS などのユーザレビューでは、ユーザが興味を持って購入した商品の感想や評価が投稿されており、投稿したユーザの中にはこれらの広告媒体を見て購入したユーザも存在する。これらの宣伝によって商品・サービスを購入したユーザは、アピールされた特徴に対する期待が大きく、ユーザレビューでも TVCM でアピールされた特徴について評価されていると考えられる。

また、TVCM は時系列に応じてアピールされる内容や特徴が変化する場合がある。図 1 の過去に配信された商品 A の TVCM では、2007 年から 2009 年に配信された TVCM で「急須」、「にごり」といった見た目に関する特徴がアピールされているのに対し、2011 年以降では「急須」、「旨み」といった味に関する特徴に変化している。このような変化に応じて、ユーザレビューでも TVCM でアピールされた特徴について評価されると考えられることから、TVCM でアピールされた特徴とユーザレビューで評価された特徴の変遷には何らかの関連性があると考えられる。そこで、本研究では各 TVCM でアピールされた特徴を示す特徴語を字幕・セリフから抽出し、その特徴に対する言及の推移を TVCM とユーザレビューから比較する事で、TVCM がユーザの投稿に与える影響を分析する。

本稿の構成について以下に述べる。2 章では本研究に関連する TVCM の影響に関する研究、及びキーワードの抽出とテキストの時系列な投稿パターンを用いた研究について紹介する。3 章では、本研究で行った特徴語の抽出と時系列な出現パターンの算出手法について説明する。4 章では実際に商品・サービ



図 1 TVCM の変遷とアピールされる特徴

スの実際の TVCM から抽出した特徴語を用いて反応の検証を行った結果について述べる。

2 関連研究

本研究は、TVCM でアピールされた特徴について評価した投稿から、TVCM が与えるユーザへの影響を分析することを目的としている。本章は、本研究に関連する TVCM の影響と効果に関する研究、テキスト情報から有益な情報を抽出する手段とされるキーワードの抽出手法及びテキストの時系列な投稿パターンを用いた研究について紹介する。

2.1 TVCM の影響・効果

TVCM が視聴者に与える効果の分析において、顧客が TVCM のメッセージをどのように処理しているかを認識する事は重要とされており、TVCM に含まれるコンテンツ情報や表現要素と視聴者の認知の間にある関連性について検討されている。牧野は、視聴者が TVCM に親近感を覚える内容は、商品への関心と購入意図を高められると示唆している [1]。越川は TVCM に含まれるメッセージのうち、七五調のメッセージは音や感情を処理する右脳と言語を処理する左脳を両方刺激することから、記憶を活性化しやすい音声的な特徴を持つと主張している [2]。浅川は TVCM が視聴者に与える印象を 5 つの因子とその特性値によって求めることで、TVCM を視聴印象のタイプ毎に分類した [3]。河原は TVCM に含まれる表現要素で、最も消費者の反応に影響を与える要素と各表現要素が消費者に影響を与え

る要因について Random Forest 法を用いて分析した [4]。また、小迫らは TVCM に結びつく TV 番組の種類や内容によって宣伝の効果は異なるとして、TVCM と TV 番組のテキスト情報、及び出演タレント・キャラクターなどのコンテンツ情報から TVCM と TV 番組の類似度を算出する手法を提案した [5]。

以上の研究では 1 つの TVCM 自体の内容、表現に注目した研究内容が主であり、TVCM でアピールされた特徴がユーザに与える影響を過去の TVCM の配信の時系列から注目した研究は少ない。本研究はその点を踏まえ、TVCM でアピールされた特徴に対するユーザの反応をユーザレビューの投稿の時系列から分析する手法の提案を目指す。

2.2 特徴語・評価属性の抽出

ユーザレビュー等のテキスト情報から有益な情報を取得する手法として、投稿から評価される属性や商品を特徴づける単語を抽出する手法があげられる。属性の抽出について、波多野らは類似する文は同じアスペクトを持つとして、評価文に含まれる内容語の出現頻度からクラスタリングを行い、これらのクラスタの中心に位置する評価文にアスペクトを付与する事で、機械的に全ての評価文のアスペクトを推定する手法を提案した [7]。中野らや村らは各文節に含まれる係り受け関係や単語の分散表現等の素性を基にベクトルを生成し、Support Vector Machine を用いて属性を抽出している [8] [9]。村松らはユーザレビューの時系列における出現パターンに着目し、投稿レビュー数及びトピック数の増加、トピックの推移から商品の属性を予測する手法を提案した [10]。

2.3 時系列な投稿の関連性

Twitter や Instagram 等のマイクロブログをソーシャルセンサーとして、投稿内容の時系列からイベントとの関連性を分析、検出する研究は多数行われている。若宮ら [11] は各地域におけるインフルエンザに関する Tweet 数のピークとインフルエンザ感染者のピークに関連があることを示し、磯ら [12] は両者のピークの間隔から現時点での感染者数と将来的な感染者数を Tweet 数から予測する手法を提案した。また、 Deng 熱、胃腸炎などその他感染症に対するアプローチに加え [15] [16]、株価収益率や地震に対する反応などのイベントとの関連についても検討されている [13] [14]。本研究では、TVCM とユーザレビューの言及の推移から関連性を分析する点でこれらの研究と類似し、TVCM による影響を TVCM とユーザレビューの出現パターンを分析することで、TVCM の広告としての効果を評価することを目指す。

3 提案手法

本研究では、TVCM でアピールされた特徴に対するユーザの反応を分析する為、TVCM の字幕・セリフに含まれる単語から TVCM とユーザレビューの出現パターンを示す算出する。本章では、TVCM でアピールされる特徴を示す特徴語の抽出方法及び TVCM とユーザレビューで各特徴語の時系列データを作成する手法について説明する。

3.1 特徴語の抽出

TVCM の字幕・セリフで頻出される単語を、TVCM でアピールされた特徴を示す特徴語として 2 つの期間を基準に頻出する単語を抽出する。

3.1.1 全期間における頻出単語

対象商品の TVCM が配信された全ての期間で頻出する単語を、全ての TVCM を通じてアピールされた特徴を示す特徴語として用いる。これらの特徴語は各 TVCM の字幕・セリフを文書として閾値以上の DF 値を示した単語を抽出した。

3.1.2 該当期間における頻出単語

各 TVCM で頻出する単語を TVCM が放送された期間で特に強調されている特徴を示す特徴語として用いる。これらの特徴語については全期間と同様に各 TVCM を文書として各 TVCM で閾値以上の TF 値を示した単語を抽出した。

3.2 時系列データの作成

3.1 より抽出した特徴語を用いて、TVCM とユーザレビューでの出現パターンを求める。特徴語は TVCM を通じて対象商品をアピールした特徴を示すことから、ユーザレビューでもその特徴語を用いて投稿されると考えられる。そこで、特徴語の TVCM での出現頻度とユーザレビューでの出現投稿数を元に特徴量を算出し、TVCM での主張と TVCM に対する反応のパターンを示す時系列データとして作成する。

まず TVCM での時系列データについて、各月で配信された全ての TVCM の字幕・セリフに含まれる特徴語の TF 値を特徴量として月毎に算出する。なお、各 TVCM の配信期間が不明である為、TVCM が配信されてから次の TVCM が配信されるまでの期間を配信期間とし、複数の TVCM が同じ期間で配信開始された場合は同時配信と仮定した。

ユーザレビューでの時系列データについては、各月で投稿されたレビューで特徴語を含む投稿の DF 値を特徴量として算出する。また、ユーザレビューで特徴語を含む投稿には以下の 2 つの尺度による評価が存在する。

- 眠くなりにくいと CM であったが、やっぱり眠い。
- 眠くなる。頭が痛い。とにかく辛い

上部の投稿は、TVCM で使用されたフレーズや“CM”や“コマーシャル”などの TVCM に関する用語を用いているなどの TVCM を視聴した上で商品进行评估していることが明白に分かる投稿である。しかし、下部の投稿では TVCM でアピールされている特徴について評価されているものの、その TVCM を視聴したかについては文章からは不明である。本研究ではこれら 2 つの投稿を同質な評価とし、特徴語を含む全ての投稿を TVCM の主張に対する反応として特徴量を算出した。

4 評価実験

本稿では、実際に各商品・サービスの TVCM から提案手法を元に抽出した特徴語を用いて、TVCM とユーザレビューの時系列データから比較を行った。

4.1 使用データセット

4.1.1 字幕データ

対象商品・サービスの TVCM から特徴語を抽出する為、Youtube 等の動画配信サイトで投稿されている CM 動画、公式サイトで配信された TVCM プレス情報等から過去の TVCM の字幕・セリフ、配信開始日等の CM 情報を収集し、字幕データとして作成した。実際に作成した字幕データの一部を表 1 に示す。

表 1 字幕データの作成例

cm	year	month	caption	text
1	2007	10	字幕	急須のお茶ににごりがあるのに
1	2007	10	セリフ	急須のお茶ににごりがあるのに
1	2007	10	字幕	ペットボトルはなぜ透明だったのか
1	2007	10	セリフ	ペットボトルはなぜ透明だったのか
1	2007	10	字幕	にごりを除くと、お茶はやせる。
1	2007	10	セリフ	にごりを除くと、お茶はやせる。
2	2007	10	セリフ	三度振んのどすえ
2	2007	10	セリフ	ひい、ふう
2	2007	10	セリフ	こうしとくれやす
2	2007	10	セリフ	ああ、おおきに

“cm” は全 TVCM の配信順を示し，“year”，“month” は TVCM の配信が開始された時期を表す。“caption” は各 TVCM で出現した字幕・セリフの種類，“text” は実際に出現した字幕・セリフを記述したテキストを示す。重複して出現するフレーズは重要度が高い可能性がある為、同時に類似する内容が字幕・セリフで出現した場合も、双方のテキストを使用した。また、ユーザーレビューの投稿期間に対応する為、対象商品の歴代の TVCM のうち 2015 年 1 月から 2017 年 3 月にかけて配信された TVCM のみを使用し，“No ポイ捨て，Yes リサイクル”などの商品と直接関係しない情報を含む文字サイズの小さい字幕についてはノイズとして除去した。

4.1.2 ユーザーレビュー

ユーザーレビューのデータセットに関しては、NII データリポジトリで株式会社 Insight Tech が提供している不満買取センターの投稿（以下、不満データ）を使用した。不満データは 2015 年 3 月から 2017 年 3 月の期間で 106,173 人のユーザーによって投稿された 5248,820 件の投稿が提供されており、全ての投稿がユーザーの家庭的な不満及び商品・サービスに関するネガティブな評判情報である特徴を持つ。本研究では不満データの“company_name”，“product_name”のカテゴリで対象商品・サービス名を含む投稿を使用し，“text”及び“created_at”から時系列データを作成した。

4.1.3 時系列データ

洗剤 A，医薬品 B，携帯 C の 3 つの商品・サービスを対象とした字幕データから抽出された特徴語の一部を表 2 に示す。

洗剤 A は 11 件，医薬品 B は 9 件，携帯 C は 32 件の TVCM を使用し，特徴語は NEologd¹ を辞書として Mecab² から抽出

表 2 特徴語（一部）

	特徴語
洗剤 A	ジェルボール，洗濯，抗菌，ウイルス，漂白剤
医薬品 B	くしゃみ，鼻炎，効果，花粉，説明，鼻水，アレルギー，薬，眠い
携帯 C	スマホ，1980 円，コミコミ，学生さん，ワンキユッパ，缶，ニヤン

した名詞，形容詞，副詞を対象とし，全期間で 0.45，該当期間で 0.2 以上の出現頻度を示した単語を抽出した。次に，洗剤 A，医薬品 B，携帯 C で抽出されたそれぞれ 2 つの特徴語から作成した TVCM と不満データの言及の推移を示すグラフを図 2,3,4 に示す。

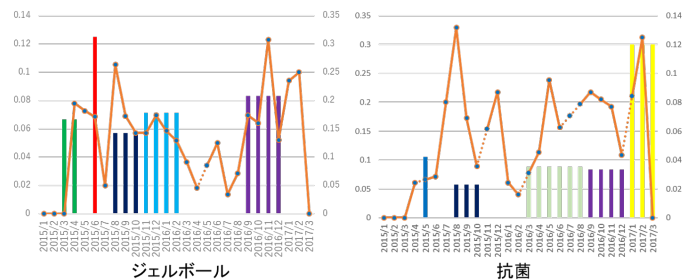


図 2 洗剤 A の時系列データ

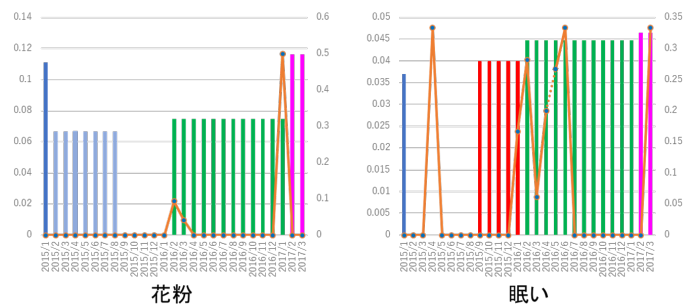


図 3 医薬品 B の時系列データ

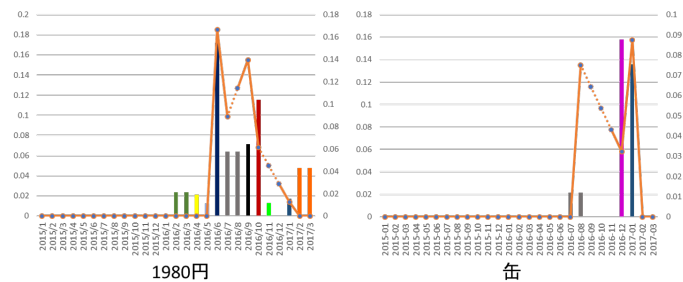


図 4 携帯 C の時系列データ

洗剤 A は 669 件，医薬品 B は 143 件，携帯 C は 893 件の投稿を使用し，ヒストグラムは各特徴語の TVCM の言及の推移，折れ線はユーザーレビューの言及の推移を示す。左の縦軸は TVCM での特徴量，右の縦軸は不満データでの特徴量を示す。なお，各特徴語の表記揺れに対応する為，予め Word2Vec で分散化した不満データから，cos 類似度が 0.9 以上の単語を特徴語と同義の単語として含め，時系列データを作成した。

以上の図から，特徴語について TVCM でアピールされてからユーザーレビューの特徴量が増加するまでに時間差が生じる

1 : <https://github.com/neologd>

2 : <http://taku910.github.io/mecab>

TVCM が存在する事が分かる。実際に図 2 の“抗菌”では全ての TVCM が配信された月に特徴量が増加しているのに対し、“ジェルボール”では 2015 年 11 月に配信されてから 1ヵ月後にユーザレビューでの特徴量が増加されている。また、医薬品 B の“眠い”については 2015 年 9 月に配信されてから 4ヵ月後の 2016 年 1 月に特徴量が増加されている。また、これらの結果から、TVCM が配信されてから反応を得るまでの時間差の違いは、商品と特徴語の関係性によって変化する事が考えられる。その点を踏まえ、抽出した全ての特徴語で TVCM とユーザレビューでの時間差をまとめた結果、洗剤 A の“抗菌”以外の全ての特徴語で時間差が生じる TVCM が存在し、“洗濯”、“漂白剤”などの時間差の大きい TVCM が存在する特徴語は、既に商品の特徴として知られた特徴である可能性がある。

4.2 今後の課題

本研究では、TVCM でアピールされる特徴を字幕・セリフで頻出される単語から抽出したが、抽出された単語からでは TVCM でアピールされた特徴について正しく言及されていない問題がある。実際に、医薬品 B の“花粉”は花粉症に効く特徴を示す特徴語となる。しかし、TVCM では“花粉”という単語を用いて表現されているのに対して、ユーザレビューでは“花粉症”と言及されているなどの特徴について言及される内容の違いによって特徴量が異なる場合があり、TVCM の視聴と無関係な投稿も含まれている可能性がある。また、TVCM でのアピールとして字幕・セリフ以外に映像や音声などの演出要素も挙げられることから、TVCM を視聴したユーザによる言及のみをユーザレビューから抽出し、TVCM の字幕・セリフ以外の演出要素を加味した特徴量から反応を分析する必要がある。また、今回ユーザレビューとして使用した不満データは基本的にネガティブな評価を集積したものである為、アピールされた特徴がポジティブな評価を得たことで時間差が生じた可能性も考えられることから、今後 Twitter などのその他のユーザレビューを用いた検証が必要となる。

5 まとめ

本稿では、TVCM でアピールされた特徴とユーザレビューの投稿内容の変化に関連性があるとして、TVCM の字幕・セリフに含まれる特徴語から時系列な出現パターンを TVCM とユーザレビューから比較する事で、TVCM がユーザの評価に与える影響について分析を行った。そして、実際に抽出された特徴語を用いて検証を行った結果、ほぼ全ての特徴語で TVCM でのアピールからユーザレビューでの反応までに時間差が生じ、その特徴語と商品との関係性によって生じる時間差が異なることが分かった。今後は、Twitter 等の肯定的な情報を含んだ投稿も含んだユーザレビューを用いた分析を検討している。

謝辞

本研究では、株式会社 Insight Tech が国立情報学研究所の協力により研究目的で提供している「不満調査データセット」を

利用した。

文 献

- [1] 牧野幸志, “広告効果に及ぼすコンテンツ情報の影響に関する研究 (2)” 経営情報研究, 第 16 巻, 第 1 号, pp. 1-11, 2008.
- [2] 越川靖子, “広告における記憶研究に関する一考察” 経営情報研究, 第 16 巻, 第 1 号, pp. 1-11, 2008.
- [3] 浅川雅美 “テレビ CM の「視聴印象」の多次元的特性の分析” 行動計量学, Vol.36, No.1, pp. 47-61, 2009.
- [4] 河原達也, “TVCM 表現要素の消費者反応に対する効果” 行動計量学, Vol. 43, No. 1, pp. 85-105, 2016.
- [5] 小迫大, 宮林卓郎, 坂本真樹, “コンテンツ情報に着目した TV 番組と TVCM の類似性算出に関する研究” 広告科学 54(0), pp. 33-49, 2012.
- [6] 鈴木元也, 生田目崇, “消費者行動に着目したテレビ CM 効果の考察” 経営情報学会全国研究発表大会要旨集 2012, 144, 2012.
- [7] 波多野匡, 嶋田和孝, 遠藤勉, “クラスタリングを利用した評価文のアスペクト推定” 情報科学技術フォーラム講演論文集, Vol. 9, No. 2, pp. 103134, 2010.
- [8] 中野裕介, 湯本高行, 新居学, 上浦尚武, “機械学習による商品レビューの属性-意見ペアの抽出” 研究報告データベースシステム 2015-DBS-162 14 1-8, 2015-11-19.
- [9] 村本直樹, 大島裕明, 湯本高行, “係り受けと分散表現を考慮したレビューからの属性と意見の抽出” 信学技報 DE2018-3, 2018-06, pp.9-14.
- [10] 村松直哉, 佐藤哲司, 伏見卓恭, “レビュー観点の推移パターンに基づく商品属性の抽出手法”, 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM), Mar 2017.
- [11] Shoko Wakamiya, Yukiko Kawai and Eiji Aramaki, “After the Boom No One Tweets: Microblog-based Influenza Detection Incorporating Indirect Information” In Proc. of the 6th International Conference on Emerging Databases (EDB 2016), pp. 98-106, 2016.
- [12] Hayate Iso, Shoko Wakamiya and Eiji Aramaki, “Forecasting Word Model: Twitter-based Influenza Surveillance and Prediction” In Proc. of the International Conference on Computational Linguistics (COLING), pp. 76-86, 2016.
- [13] Gabriele Ranco, Darko Aleksovski, Guido Caldarelli, Miha Grar and Igor Mozeti, “The effects of twitter sentiment on stock price returns” PLoS ONE, 10(9), 121.
- [14] Takeshi Sakaki, Makoto Okazaki and Yutaka Matsuo. “Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors” Proceedings of the 19th International Conference on World Wide Web(WWW), pp. 851-860, 2010.
- [15] Janana Gomide, Adriano Veloso, Wagner Meira Jr, Virglio Almeida, Fabrcio Benevenuto, Fernanda Ferraz and Mauro Teixeira, “Dengue surveillance based on a computational model of spatio-temporal locality of Twitter” Proceedings of the 3rd International Web Science Conference on - Web-Sci '11, pp.1-8, 2011.
- [16] Ryo Takeuchi, Hayate ISO, Kaoru Ito, Shoko Wakamiya, Eiji Aramaki, “Multi Liner Regression of Symptom-related Tweets for Infectious Gastroenteritis Scale Estimation” In Proc. of the Workshop on Digital Disease Detection using Social Media (DDDSM) of International Joint Conference on Natural Language Processing (IJCNLP), 2017.
- [17] Kensuke Mitsuzawa, Maito Tauchi, Mathieu Domoulin, Masanori Nakashima and Tomoya Mizumoto, “FKC Corpus: a Japanese Corpus from New Opinion Survey Service”, In proceedings of the Novel Incentives for Collecting Data and Annotation from People: types, implementation, tasking requirements, workflow and results, pp.11-18, Portoro, Slovenia, May 2016.