

インターネット・コミュニティ・データを使った テレビCMの商品イメージ形成効果測定

Analysing the Image Building Effects of TV Advertisements Using Internet Community Data

上原 宏
Hiroshi Uehara

筑波大学 大学院 ビジネス科学研究科
Graduate School of Business Science, University of Tsukuba
uehara@gssm.otsuka.tsukuba.ac.jp

佐藤 忠彦
Tadahiko Sato

(同 上)
sato@gssm.otsuka.tsukuba.ac.jp

吉田 健一
Kenichi Yoshida

(同 上)
yoshida@gssm.otsuka.tsukuba.ac.jp

keywords: web mining, on-line community, web marketing, data analysis

Summary

This paper proposes a method to measure the effects of TV advertisements on the Internet bulletin boards. It aims to clarify how the viewers' interests on TV advertisements are reflected on their images on the promoted products. Two kinds of time series data are generated based on the proposed method. First one represents the time series fluctuation of the interests on the TV advertisements. Another one represents the time series fluctuation of the images on the products. By analysing the correlations between these two time series data, we try to clarify the implicit relationship between the viewer's interests on the TV advertisement and their images on the promoted products. By applying the proposed method to an Internet bulletin board that deals with certain cosmetic brand, we show that the images on the products vary depending on the difference of the interests on each TV advertisement.

1. はじめに

インターネット掲示板, Blog, SNS などのオンラインコミュニティには, 社会的な話題全般にわたってユーザーの意見が日々大量に書き込まれている。製品・サービス(以下, プロダクト)のテレビCMも, オンラインコミュニティが注目する話題のひとつであり, CM 放映開始直後から, 宣伝されたプロダクトに対する様々な意見が書き込まれる。こうした現象は, ブランド認知からプロダクト購買に至る消費者行動に変化をもたらしていると言われている [山崎 07]。

従来は, より多くのユーザーの興味を喚起できることがテレビCMの最も重要な効果とされ, 当該CMを視聴したユーザー数と, 当該CMの1人あたりの平均視聴回数の積として計算されるGRP(Gross Rating Point: 延べ到達率)など, ユーザーへの到達度で広告効果を測定することが一般的であった。しかし, オンラインコミュニティが普及した現状では, 従来のように到達率だけでは広告効果を十分に捉えきれない。マスメディア広告によるブランド認知からプロダクト購買に至る過程に, オン

ラインコミュニティでの意見や経験の共有, 関連情報の検索といった消費者行動が介在する。その過程で広告が伝えたメッセージがどのように浸透したかを捉えることが, これからの広告効果を捉える上で不可欠であるとの認識が実務家の間で広がっている [秋山 07, 加藤 07, 横山 07]。

このような広告効果を表す新たな指標はまだ提案されていないが, オンラインコミュニティで書込まれるプロダクトに対する意見や利用経験の活性度合いで捉えるという考え方が示されている [秋山 07]。本稿では, このような考え方に即し, テレビCMがインターネット掲示板でのプロダクトイメージ形成にどのような効果を及ぼしたかを捉える広告効果測定手法を提案する。

インターネット掲示板の書込みは概して断片的であり, テレビCMから影響を受けているのかどうかを文脈から読みとれるものはごく希である。いくつかのblog検索サイト*¹では, 話題となっている語彙でキーワード検索

*1 Yahoo! ブログ検索 : <http://blog-search.yahoo.co.jp/> ,
Technorati Japan : <http://www.technorati.jp/> , Ameba 口コミ評判検索 : <http://buzz.ameba.jp/>

すると、その評判やイメージを意味する語彙を表示する機能を提供している。しかし、それらの語彙のうち、テレビ CM への関心と相関するものがどれなのかは判別できず、取得した語彙が広告のプロダクトイメージ形成効果を反映しているのかどうかを明らかにすることはできない。

提案手法は、テレビ CM への関心が持続した期間を特定し、その期間におけるユーザーのプロダクトに対するイメージと CM への関心との相関を調べ、相関が認められたプロダクトイメージを抽出してその特徴を調べる。すなわちテレビ CM によるイメージ形成効果を、CM への関心が集まっている期間でのプロダクトイメージへの意見の活性度合いで捉えることが本稿での目標である。

なお、オンラインコミュニティの書き込みは、何を話題としているかを判別するための手がかりが文中から欠落していることがしばしば見受けられる。プロダクトイメージを抽出するにあたっては、各書き込みが何を話題としているかを明確化する必要があるため、タイトル等の補完情報を手がかりに話題を特定できるような文書構造を持つ書き込みが本提案での解析対象となる。

以下では、2 章で関連研究を紹介し、3 章で提案手法を述べ、ついで 4 章にて提案手法による実験結果を報告する。5 章では 4 章の実験結果を使ったテレビ CM によるプロダクトイメージ形成効果の分析事例を示す事により提案手法と実験結果の妥当性を示す。最後に本稿の結論を述べる。

2. 関連研究

従来テレビ CM の効果測定に関する研究は、アンケートやインタビューによって、形容詞の選択肢からテレビ CM を表象しているものを視聴者に選択させるという手法を用いてきた [Aaker 81, Wells 71, Schlinger 79]。しかしながら、これらの手法は膨大な人手を要する点や、サンプリングやアンケート回収に時間を要する点に課題を残している。

本研究では上記課題を克服するために、テレビ CM のオンエアにともないオンラインコミュニティで形成される関心の度合いとその時間的な変動を分析する事で、テレビ CM の効果測定を試みる。

オンラインコミュニティで形成される関心の時系列な分析を試みた研究としては、テレビ番組オンエア中に視聴者間で取り交わされるチャットから関心の高い区間を認識して番組のダイジェストを生成する研究 [Uehara 05, 宮森 05]、blog に出現する語彙の出現頻度の時系列推移を可視化しグラフの形状から社会的関心の分類を行う研究 [Gruhl 04]、連接する形態素の出現頻度の時系列推移の同期性を評価して関心と呼んでいる複合語を抽出する研究 [村上 06]、同様な手法に意外性という評価尺度を加味して掲示板の書き込みから話題を抽出しようとする研究 [石

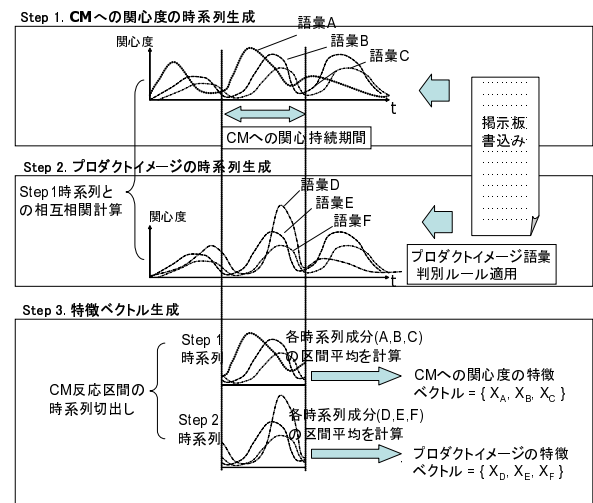


図 1 提案手法概要

井 05]、書き込み数の時間的な密集度から、掲示板での盛り箇所を特定する研究 [井上 04, Haga 02] 等がある。

しかし、後述するように、これらの手法を単純に適用しても、短い間隔でオンエアされるシリーズ CM において前後でオンエアされた CM の影響を分離して解析する事ができず、テレビ CM の効果測定に応用するには改良を要する。さらに、上記の手法は、個々の語彙・語句をベースに関心を捉えるか、もしくは関心の対象を特定せずに盛り箇所を抽出しようとしたもので、語彙間の関連性を構造化して関心を捉える分析には至っていない。

本研究では、CM 自体への関心を表明する語彙と、プロダクトイメージの関心語彙との時系列な相関から CM への関心と関連するプロダクトイメージの抽出を試みる。なお、プロダクトイメージ語彙を抽出するには、書き込みからプロダクトイメージへの意見を表明している文を判別することが必要である。評判解析、あるいは意見マイニングと呼ばれる諸研究では、一般に候補となる語彙を登録した辞書を構築し、それを参照して対象文書を解析し意見を抽出する。文書解析にあたっては、タグ付コーパスや機械学習を用いてより高精度化を試みる研究も見られる [矢野 04, 鍛冶 07, 小林 06, 飯田 05]。本研究も、辞書の構築については、これらの研究の考え方を踏襲する。しかし、意見抽出の精度を追及することが主眼ではないため、文書解析についてはより簡便な方法を用いることとした。

3. 提案手法

図 1 に前述の関連研究の問題点をふまえて、テレビ CM の効果測定手法の流れを示す。提案手法は Step1: 初めにインターネット掲示板の書き込みからテレビ CM に対する関心を表明する語彙群を抽出し、

時系列データ生成アルゴリズムを適用して関心度の時系列データを得る。次にこの時系列データの自己相関を分析して CM への関心が持続した期間を推定する。

Step2: Step1 と同様に、プロダクトのイメージを表明する語彙群の時系列データを生成する。イメージ語彙は、同一句内でプロダクトの属性を表す語彙と共起することを抽出条件とする。

Step3: 最後に Step1, Step2 それぞれの時系列データについて、CM への関心が持続した期間に対応する部分時系列を切り出し、この期間における関心度の平均値を各語彙それぞれについて求め、それらを成分とするベクトルを生成する。

Step3 で生成するベクトルは、掲示板が CM に反応した期間において、どのようなプロダクトイメージが形成されたのかを示す特徴ベクトルである。

以下、提案手法の詳細を述べる。まず、時系列遷移データを生成する手法を説明した後、それをを用いたプロダクトイメージ形成効果測定の手順を述べる。

3.1 関心度の時系列遷移データ生成手法

提案手法では、3.2 節で説明するプロダクトイメージ形成効果測定のプロセス (具体的には 3.2 節 Step 1 と 2) で、[Uehara 05] の手法をもとに、インターネット掲示板の書き込みからテレビ CM に対する関心度の時系列遷移データを生成する。[Uehara 05] では、書き込みの時間的な密度が非常に高い期間に限定して同手法を適用しているが、本研究では、書き込みの粗密に大きな変動があるような長期の観測期間に適用する。

以下では [Uehara 05] の手法の特徴を述べ、本研究で対象とする書き込みに適用する際の課題を指摘した後、修正点を記し、予備実験によって修正点の妥当性を確認する。

§ 1 基本アルゴリズム

著者らが [Uehara 05] で提案したアルゴリズムは、「短い間隔で同じ語彙が出現する程その語彙に関する関心度が高い」として関心度の時系列遷移データを生成する。即ち、このアルゴリズムは、以下の手順で語彙の時系列な偏り度合いを関心度として計算する。

- (1) 各書き込みにはタイムスタンプが付されているものとする。観測期間において各書き込みは、タイムスタンプ順に並んでいるものと仮定し、それらのうちで語彙 w_m が出現する各書き込みにシーケンス番号 $m_j (j = 1, \dots, j_{max})$ をつける。ここで、 j_{max} は観測区間で w_m が出現する最後の書き込みの番号を示し、 m_k から $m_{k+1} (1 \leq k \leq j_{max} - 1)$ までの間にある書き込み数をインターバルと呼び I_k で表す (I_k は、最小値を 1 とする正の整数)。

インターバルの例を図 2 (a) に示す。図においての各長方形は、時系列に並べた書き込みを表わし、長方形内に m_k の表記があるものは語彙 w_m が出現する

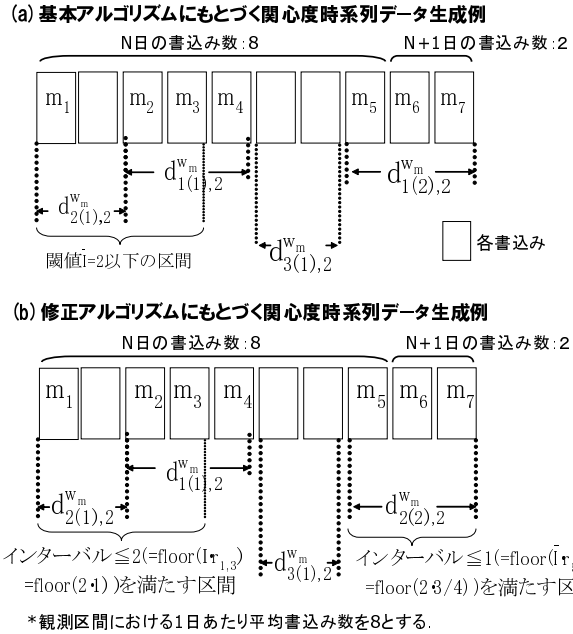


図 2 関心度時系列データ生成の例

書き込み、表記がないものは w_m が出現しない書き込みを意味する。図で語彙 w_m は、 m_1 で最初に出現し、 m_2 で再現するまでに 2 つの書き込みが挟まる (m_1 の書き込みを含んで数える) ので、インターバル $I_1 = 2$ である。同様に、 m_2 で出現してから m_3 で再現するまでに 1 つの書き込みが挟まるので、インターバル $I_2 = 1$ である。

- (2) インターバルの閾値 \bar{I} (\bar{I} は正の整数) を 1 から順次増やしていき、各 \bar{I} 毎に下記条件にあてはまる区間をインターバル I_k の昇順に全観測区間について調べあげ、該当する区間を全て抽出する。

条件 1 隣り合うインターバル $I_k, I_{k+1}, \dots, I_{k+T}$ が全て \bar{I} 以下であるような全ての連続した区間 (T は、 w_m がインターバル閾値 \bar{I} で連続して出現する繰り返し回数)

条件 2 条件 1 を満たして、かつ、 $J < \bar{I}$ であるような $d_{J(l),T}^{w_m}$ を含まない区間

ここで、各 \bar{I} のもとで抽出された区間にシーケンス番号 l をつけたものを $d_{\bar{I}(l),T}^{w_m}$ で表す。また、上述した設定の下で区間 $d_{\bar{I}(l),T}^{w_m}$ における語彙 w_m の関心度 $A_{\bar{I}(l),T}^{w_m}$ を下記のように定義する。

$$A_{\bar{I}(l),T}^{w_m} = \bar{I}_{max} - \bar{I}$$

$\bar{I}_{max} : \bar{I} \times T$ が観測区間の総書き込み数
 以下であるような \bar{I} の最大値

なお、 $d_{\bar{I}(l),T}^{w_m}$ は、書き込みの数で表される区間のため、最後にこれらを時間単位での区間にマッピングしなおす。例えば日次でのマッピングを行う場合、区間 $d_{\bar{I}(l),T}^{w_m}$ に含まれる書き込みのタイムスタンプを調べ、日付が更新され

る前後で $d_{\bar{I}(l),T}^{w_m}$ を分割する．この処理を全ての $d_{\bar{I}(l),T}^{w_m}$ に対して行い，タイムスタンプが同一日付の $d_{\bar{I}(l),T}^{w_m}$ をまとめる．このようにすると，同一日において関心度の異なる複数の $d_{\bar{I}(l),T}^{w_m}$ が存在することがあるが，そのような場合，当該日の関心度は，当該日の $d_{\bar{I}(l),T}^{w_m}$ に含まれる書込み数による加重平均とする．

図 2(a) は $T = 2, \bar{I}_{max} = 5$ とした時のこのアルゴリズムによる抽出例である．図では最初の 8 つの書込みのタイムスタンプは N 日，残りの 2 つの書込みのタイムスタンプは N+1 日と想定している．つまり，N 日は書込みが密，N+1 日は疎な区間と想定する．

まず，インターバル閾値を初期値 $\bar{I} = 1$ に設定し，この間隔で 2 回語彙 w_m が連続して出現する区間を左から順に探索すると，最初に $d_{1(1),2}^{w_m}$ が得られる．同じ条件で探索を続けると，更に $d_{1(2),2}^{w_m}$ が得られる．次に $\bar{I} = 2$ に設定して同様に探索すると m_1 から m_3 までの区間が得られるが， $J < 2$ であるような区間を除外するので，結果として $d_{2(1),2}^{w_m}$ が得られる．以降， \bar{I} を順次増やして，同様の手順で区間 $d_{\bar{I}(l),2}^{w_m}$ を全て求める．この後，各 $d_{\bar{I}(l),T}^{w_m}$ での関心度を求め，更に N 日，N+1 日にマッピングして，関心度を計算すると以下のとおりとなる．

N 日の関心度: 3.3

$$((5-2)*2 + (5-1)*3 + (5-3)*2 + (5-1))/8$$

N+1 日の関心度: 4.0 (= (5-1)*2/2)

ここで書込みが疎な区間 N+1 日のほうが N 日に比べて関心度が高いと計算されている点には注意を要する．極端な例を想定すると，1 日あたりの書込み数が 1 件でも，語彙が隣り合う書込みに連続して出現するような場合，上述のように既存研究の手法を単純にあてはめたのでは，関心度が非常に高く計算されてしまう．テレビ CM への関心は，多くのユーザーの同調を伴うものであり，こうした疎な状態での語彙の偏りを，テレビ CM への関心と捉えるのは適切ではない．

§2 修正アルゴリズム

前述のように既存研究の手法を単純にあてはめた関心度の計算方法には問題があった．そこで本研究では下記のように区間抽出条件を改良した修正アルゴリズムを開発した．

条件 1 隣り合うインターバル $I_k, I_{k+1}, \dots, I_{k+T}$ を，全て $\text{floor}(\bar{I} \cdot r_{k,k+T})$ が閾値以下であるような全ての連続した区間とした．ここで， $\text{floor}(A)$ は A を超えない最大の整数を意味するガウス記号である．また， $r_{k,k+T}$ は， $m_k, m_{k+1}, \dots, m_{k+T}$ の書込み日付において書込みの総数が少なかった場合，つまり書込みが疎な状態でインターバルを大きくする (結果として時間間隔を同程度にする) ために導入した補正項で，下式により計算する．

$$r_{k,k+T} = \frac{m_k, \dots, m_{k+T} \text{ と同じ日付けの全書込み数}}{\text{観測期間における 1 日あたり平均書込み数} * m_k, \dots, m_{k+T} \text{ の区間日数}}$$

また，例えば，500 回に 1 回語彙が出現することと 499 回に 1 回出現することの関心度の差異よりも，2 回に 1 回出現することと毎回出現することの関心度の差異のほうが大きい．このような違いを考慮して，インターバル閾値 \bar{I} が小さくなるにつれて関心度の増大幅がより大きくなるよう下記のように関心度の計算式を修正した．

$$A_{\bar{I}(l),T}^{w_m} = \frac{1}{\bar{I}}$$

図 2(b) は，修正アルゴリズムにもとづく抽出例である．観測期間における 1 日あたり平均書込み数を 8 とする．図 2(a) と同様に初期値 $\bar{I} = 1$ にて探索を開始すると， m_2, m_3, m_4 は，インターバル 1 で出現し，かつ $r_{2,4} = 1$ なので， $d_{1(1),2}^{w_m}$ が得られる．同じ条件で探索を続けるとインターバル 1 で出現する m_5, m_6, m_7 に到達するが， $r_{5,7} = \frac{5}{8} (= \frac{2+8}{8*2})$ なので，これらのインターバルは，閾値 $\text{floor}(\bar{I} \cdot r_{5,7}) = \text{floor}(1 \times \frac{5}{8}) = 0$ を満たさず，抽出されない．次に $\bar{I} = 2$ に設定して同様に探索すると， $d_{2(1),2}^{w_m}$ が得られる．同じ条件で探索を続けると，インターバル 1 で出現する m_5, m_6, m_7 に到達するが，今度は，閾値 $\text{floor}(\bar{I} \cdot r_{5,7}) = \text{floor}(2 \times \frac{5}{8}) = 1$ を満たすので， $d_{2(2),2}^{w_m}$ が得られる．以降， \bar{I} を順次増やして，同様の手順で区間 $d_{\bar{I}(l),2}^{w_m}$ を全て求める．

以下，関心度の計算と N, N+1 へのマッピングを行うと，各日の関心度は以下のとおりとなり，図 2(a) とは逆に密な区間 N のほうが N+1 日に比べて関心度が高いと計算される．

N 日の関心度: 0.6 (= ((1/2*2) + (1/1*3) + (1/3*2) + (1/2))/8)

N+1 日の関心度: 0.5 (= ((1/2*2)/2))

これは，図 2(a) でインターバル閾値 $\bar{I} = 1$ に相当すると認識された， m_5, m_6, m_7 に対応する区間が，図 2(b) では，N+1 日の疎な状態を反映して，インターバル閾値 $\bar{I} = 2$ に相当する区間と認識された結果である．

§3 予備実験

観測期間において書込みの粗密に大きなばらつきがあるようなデータを用いて，上記修正点の妥当性を確認した．対象データには，2ちゃんねる*2より，ある化粧品ブランドの CM をテーマとする掲示板の 2005.8.21. ~ 2006.8.31. に該当する 1629 個の書込みを用いた．この期間における 1 日あたりの平均書込み数は 4，同最大書込み数は 98，同最小書込み数は 0，分散は 132 であった．図 3 (a) は，基本アルゴリズムから得た時系列遷移データである．y 軸は，3・1 §1 節で定義した関心度 $A_{\bar{I}(l),T}^{w_m}$ を表す．各々の折線は，この CM に登場した各タレントを意味する語彙を表す．

関心度の時間遷移は，振幅 150-200 の間で比較的小さな振動を示している．図 3(b) は図 3(a) からこの範囲を拡大したグラフである．CM 放映時期と同期して関心度が高くなっているものの，例えば，2006 年 4 月など CM

*2 2ちゃんねる: <http://www.2ch.net>

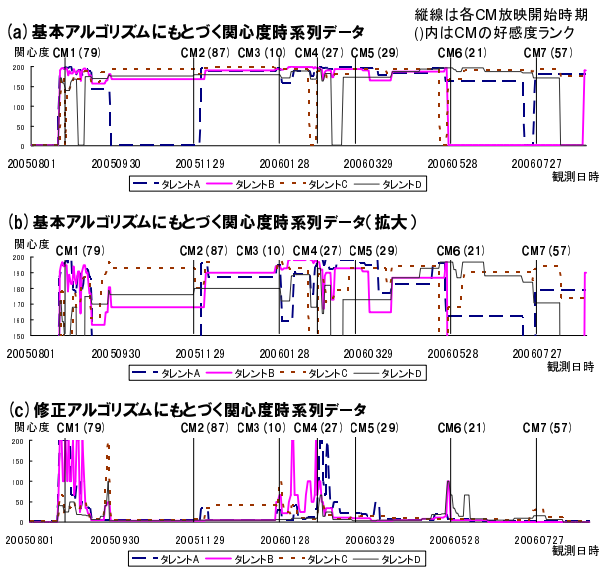


図 3 関心度の時系列データ (予備実験結果)

が放映されておらず、書き込みが非常に疎な(4月の平均書き込み数は1.3)期間でも関心度は高く表示される。

図3(c)は、修正アルゴリズムから得た時系列遷移データである、関心度は、3.1 §2節のとおり $A_{I(t), T}^{w_m} = \frac{1}{T}$ であり、最大値は1であるが、図3(a),(b)と比較の尺度をそろえるために200倍して表示した。関心度の遷移は、振幅0-200の間で大きな振動を示しており、CM放映開始直後で、関心度が非常に高くなっている。また、CMが放映されない期間で図3(a)で高い関心度が得られた箇所は、図3(c)では書き込みが疎な状態であるために低い関心度に抑えられている。

これらの結果を専門家によるCM区間の好感度調査結果([CM INDEX], 図3(a),(b),(c)上部に[CM INDEX]に記載されたCM区間の好感度順位示す)と比較すると、修正アルゴリズムによるグラフの関心度は、好感度順位の高いCM区間に高い関心度を与えるという意味で整合していた。この予備実験の結果から、修正アルゴリズムにもとづく関心度の測定は妥当と考え、本研究での時系列遷移データ生成アルゴリズムとして採用した。なお、上記CMの好感度調査は、毎月2回、専門家により首都圏3000名のモニターユーザーを対象に実施され、各ユーザーが関心を抱いたCMを集計して頻度順にランキングしたものである。

3.2 テレビCMのプロダクトイメージ形成効果の測定手法

§1 Step 1: CMへの関心度の抽出

CMは、タレント、キャッチコピー、場面設定などを通じてプロダクトのイメージを伝えようとする。そこで、これらを表象する語彙(例えば、タレント名、セリフ、ロケ地名など)を予め辞書登録し、同辞書を参照しながら

3.1節で説明した時系列遷移データ生成アルゴリズムを実行する。こうして得られる各語彙の時系列は、CMが伝えたイメージの何にどの程度関心が集まったかを表す。図3(c)で見たとおり、アルゴリズムがCMへの関心を適切に捉えているならば、関心度の盛上りは、CMのオンエア開始時期が起点となり、次第に減衰するようなカーブを描く。従って、掲示板がCMに反応した区間は、CMオンエア開始時期を始点、カーブの減衰が終了する地点を終点とする区間となる。

しかし、カーブが長いテールを描くケースや、テール上で余震を生じるケース、もしくは短い期間に複数のCMがオンエアされて反応が重なり合うようなケース(例えば、図3(c)のCM3, CM4, CM5)では、カーブの終了地点を時系列遷移データ生成アルゴリズムの出力から直感的に判断しにくいことがある。そこで本研究では、終了地点までの区間を切り出す際の恣意性を排除するため、自己相関関数を用いて終了地点を客観的に算出する事を試みた。具体的には、まず前述のアルゴリズムで抽出した時系列データに対して、同一時系列内における任意の時刻 t におけるデータ y_t と、 t から時間差 k に位置するデータ y_{t-k} との自己相関 R_k を下式のとおり求める。

$$R_k = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t)Var(y_{t-k})}}$$

Cov : 相関係数

Var : 分散

時系列データがCMからのインパクトのみを反映し、CMからのインパクトがない区間でランダムならば、自己相関関数の値は $k=0$ を起点に徐々に減衰するカーブを描くはずである。後述の実験結果より、この解析は妥当である事が確認されたので、この自己相関関数値に閾値を定め、減衰によって閾値を下回る時点を $k=n$ とした時の、 $k=0$ から $k=n$ の期間をCMへの関心が持続した期間とした。

以上 Step 1 の処理により、インターネット掲示板の書き込みから、テレビCMに対する関心を表明する語彙群(関心度の時系列データ)と、CMへの関心が持続した期間の推定を得る。

§2 Step2: プロダクトイメージの時系列データ生成

Step 2 では、掲示板参加者がプロダクトに対して抱いたイメージを表象する語彙(以下、プロダクトイメージ語彙)を文献[飯田05]で提案されている方法にならい辞書を構築し、別途計算する適合率を加味して Step 1 と同じアルゴリズムでプロダクトイメージの時系列データを生成する。

まず次の手順で辞書の構築、および適合率の計算を行う。

- 書き込みから人手でプロダクトイメージへの意見を抽出し、それらに出現する語彙を候補語彙として辞書登録する。人手でプロダクトイメージの意見かどうかを選別するにあたり、以下のクライテリアを用いた。

- プロダクト自体, もしくはその属性に対する評価を表明している文 (但し, 単に, 良い/悪いというようにイメージ評価をしていないものは除外する)
 - プロダクトの効果を表明している文
- 抽出した意見文からプロダクトイメージを意味する形容詞・形容動詞語幹をプロダクトイメージ語彙辞書に登録する.
- 同様にして, 意見文からプロダクト (およびその属性) を意味する語彙, たとえば商品名・型番・色などをシンボルワード辞書に登録する.
 - 次に掲示板の書込みを句読点で区切った文字列に分解した後, シンボルワード辞書に登録された語彙とプロダクトイメージ語彙辞書に登録された語彙が, 間に句読点を挟まない文字列の中で共起している文字列を抽出する.
 - 抽出した文字列のなかで, プロダクトイメージ語彙 $image_j$ がシンボルワードを正しく修飾している文字列を手でカウントし, 以下の定義に従い各シンボルワード $symbol_i$ の適合率とする.

$$\text{適合率} = \frac{\text{image}_j \text{ が } symbol_i \text{ を正しく修飾している文字列の数}}{\text{symbol}_i \text{ とプロダクトイメージ辞書上の語彙の共起総数}}$$

プロダクトイメージ語彙の時系列データを生成する際には, 上記の適合率を利用する. すなわち, 各掲示板書込みについて, 間に句読点を間に含まない文字列でプロダクトイメージ語彙とシンボルワードが共起している文字列を検索し, そのような文字列があった場合にプロダクトイメージ語彙が適合率回出現したとして処理を行う.

時系列データ生成アルゴリズムでは, 時系列データを形成する語彙を抽出する際に, 一定回数 (T , 後述の実験では 6) 以上, 且つ, 一定インターバル以内で連続的に語彙が出現することを条件としている. 出現回数に適合率を用いることは, 適合率が低いほど T を満たす条件が厳しくなることを意味し, 適合率の悪い出現パターンにペナルティを与える.

以上により生成する時系列データは, CM への関心とは無関係なものも含め, 全てのプロダクトイメージを含んでいる. 以下の解析では CM への関心がプロダクトイメージに与える影響を明らかにしたいので, 次に, プロダクトイメージの時系列と CM への関心語彙の時系列との相互相関を計算し, 相関が認められたプロダクトイメージの時系列のみを抽出する. 具体的には下式にもとづき, プロダクトイメージの時系列データ $y_t(i)$ と, 時間差 k に位置する CM への関心度の時系列データ $y_{t-k}(j)$ との相互相関を求め, Step1 で求めた CM 反応区間を時間差の最大値として, この区間内での相互相関が予め定めた閾値を上回ったプロダクトイメージを, CM への関心と何らかの関連性があるものとして抽出する.

$$R_k(i, j) = \frac{Cov(y_t(i), y_{t-k}(j))}{\sqrt{Cov(y_t(i), y_t(i))Cov(y_t(j), y_t(j))}}$$

§ 3 Step3: 特徴ベクトルの生成

Step3 では CM への関心を表す語彙の時系列から, CM 放映開始時期を始点とする CM 反応区間に対応する部分時系列を抽出する. この区間における各語彙の部分時系列の関心度平均値を求め, これらの値を成分とする特徴ベクトルを構成する (図 1 Step3 x_A, x_B, x_C に相当). このベクトルは, CM への関心の特徴を表す. 同様に, プロダクトイメージの時系列から CM 反応区間に対応する部分時系列を取り出し, 各語彙の関心度平均値を成分とする特徴ベクトルを構成する (図 1 Step3 x_D, x_E, x_F に相当). Step2 で CM への関心を表す語彙と相互相関を示すプロダクトイメージを抽出しているので, このベクトルが示すプロダクトイメージは, CM への関心との関連性において形成されたイメージとしての特徴を持つ. どのような特徴を反映しているかについては, 一般的なデータ解析によって明らかにすることができる. この点は 5 章で確認する.

4. 実 験

提案手法にもとづき, CM のイメージ効果を測定した. 実験では, 化粧品専門掲示板^{*3}への書き込みの中から, ある国産化粧品ブランドに対する意見に関するものをデータとした. このブランドは, 幅広い年齢層, ライフスタイルをもつユーザーをターゲットにしており, 世代, タイプの異なる 4 人のレギュラータレントによるシリーズ CM をオンエアしている. CM の各回においては, 主演タレントやロケ地を変えるなど, 少しづつ変化を持たせており, 掲示板にもそれぞれ異なる影響を与えることが予想される. このような相違が, 掲示板における化粧品へのイメージにどのような変化を与えるかを実験によって確認する.

4.1 実験データ

表 1 に, 2005 年 8 月から 2006 年 8 月にかけてオンエアされた実験対象の化粧品ブランドに対するシリーズ CM の概要を示す. このブランドは, 口紅, アイシャドウ, などを含むメイク系プロダクトの総合ブランドである. CM は, ドラマ仕立てのシーンでタレントにキャッチコピー的なセリフを語らせるような作りになっており, 主演タレント, シーン, キャッチコピーなどを各 CM 毎に替えている.

この化粧品に係わる書込みを化粧品専門掲示板から取得し, 実験データとして用いた. この掲示板は, 購入後の利用感, 試供品の利用感, 店頭販売員の説明からの印象, マスメディア広告からの印象など, ユーザーが書き込んだ様々な意見の集合体である. その書込みは, 市場に流通するほとんど全ての化粧品を網羅しており, 総書込み数は 460 万と言われている. 各書込みの見出し部に

*3 @cosme : <http://www.cosme.net>

表 1 国産化粧品ブランドのシリーズ CM (2005.8 - 2006.8)

	主演 タレント	出演 タレント	放映 開始時期	好感度ランク [CM INDEX]
CM1		A,B,C,D	2005/8/22	79
CM2	A	A,B,D	2005/11/20	87
CM3	B,C	A,B,C,D	2006/1/21	10
CM4		A,B,C,D	2006/2/24	27
CM5	A,C	A,C	2006/3/21	29
CM6	B,D	B,D	2006/5/21	21
CM7	C	C	2006/7/20	57
CM8	B,D	B,D	2006/8/21	43

は、ブランド名、プロダクト名、書き込み時刻の表示がある。ユーザーは、これらをキーにして必要な書込みを検索抽出することができる。上記国産化粧品ブランド名および上記期間をキーとして抽出した全ての書込み約1万コメントを実験データとして用いた。

4.2 CM への関心度の抽出

各 CM に登場したシーン、タレントを表象する語、およびキャッチコピー、セリフなど CM への関心を表す語彙を辞書登録し、時系列生成アルゴリズムを適用した。図 4 に結果を示す。時系列が生成された語彙は、タレントを表象する語彙のみとなった。CM 好感度調査 [CM INDEX] によれば、出演者に関する因子が CM の好感度要因の上位を占めており、実験結果は同調査と符合する。これらのグラフは、各 CM のオンエア開始時期と同期して、CM 関連語彙への関心度が高まっていることを示しており、またその後次第に関心度が減衰することを示している。例えば CM6 は、アイシャドウをメインにタレント D が主演した CM であるが、グラフの対応する期間を見ると、タレント D への関心度が高くなっている。また CM7 は、口紅をメインにタレント C が主演した CM であるが、この期間のグラフは、タレント C への高い関心度を示している。このように、生成した時系列は、各 CM 毎に異なる関心の特徴を反映している。また、CM 放映期間と関心度の盛り上がり方を見ると、CM1, CM3, CM6, CM7, CM8 への関心度は高く、それに対して CM2, CM4, CM5 への関心度は低かったこともわかる。この結果も上記の CM 好感度調査結果と整合する。

これらの時系列データからタレント毎に自己相関を計算した結果を図 5 に示す。95 パーセント信頼区間を割り込む時間差は、グラフ間でほぼ似通った値を示しており、平均値は 17 となった。これは CM への掲示板の平均反応区間が 17 日であることを示唆している。そこで、後述する時系列データからの区間切り出しの基準となる CM への反応区間を 17 日とした。なお、95 パーセント信頼区間を割り込んだ後、40 - 60 日付近でわずかな周期性が見られるが、これらは一連のシリーズ CM が 1ヶ月から 2ヶ月程度の間隔でオンエアされたことによるものと

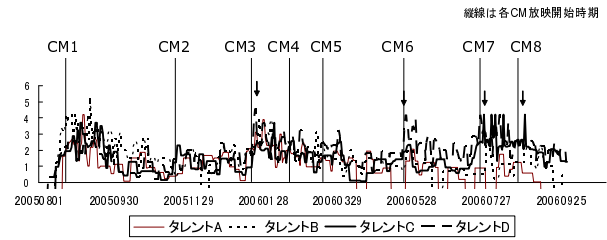


図 4 化粧品シリーズ CM への関心度時系列グラフ

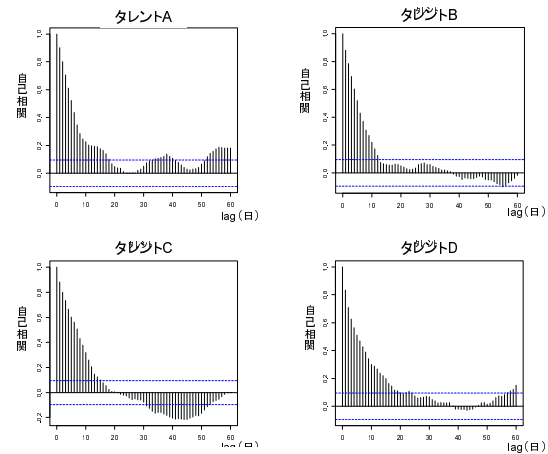


図 5 自己相関グラフ (化粧品 CM への関心度)

思われる。

4.3 プロダクトイメージの時系列データ生成

3.1 §2 節 Step2 で述べた手順に従い、@cosme の書込み 10,576 個をもとに、人手で化粧品に対するプロダクトイメージの意見を抽出し、シンボルワード辞書、プロダクトイメージ語彙辞書を登録し、シンボルワード毎に適合率を計算した。獲得したシンボルワード語彙数は 43、プロダクトイメージ語彙数は 128 であった。表 2 に、シンボルワードの一部と適合率を示す。なお、同表の 1 行目の品番は、掲示板では番号で出現している。この適合率を出現頻度として時系列生成アルゴリズムを適用した結果を図 6 に示す。CM への反応区間では、そうでない時期と比較して、より多くのイメージ語彙が高い関心度を示していることがわかる。また、図 4 と図 6 ので示した位置を比較すると、例外もあるが、CM への関心の時系列は CM オンエア開始から遅延なくピークを示しているのに対して、プロダクトイメージの時系列のピークは遅れて現れる傾向が見られる。

この時系列と、CM への関心度の時系列との相互相関を計算し、相関が認められなかった語彙を除外した。図 7 に相互相関が認められた例 (a) と、認められなかった例 (b) を示す。図 7(a) は相関が認められた例である。時間差 0 付近をピークに山形のグラフを描いている。図 7(b)

表 2 化粧品プロダクトイメージのシンボルワードと適合率

シンボルワード	品番	色合い	パール	印象	目頭	睫毛
適合率	85%	91%	92%	97%	84%	90%
シンボルワード	雰囲気	仕上げ	効果	見た目	顔色	感蝕
適合率	100%	96%	87%	95%	77%	90%

表 3 特徴ベクトル (化粧品 CM・プロダクトイメージ)

(a) 化粧品CMへの関心度の特徴ベクトル

タレント	高関心度CM					低関心度CM		
	CM1	CM3	CM6	CM7	CM8	CM2	CM4	CM5
タレントA	17	16	4	1	2	3	5	4
タレントB	38	21	5	4	5	5	7	5
タレントC	14	11	5	21	13	5	5	6
タレントD	12	22	21	24	11	5	9	8

(b) 化粧品プロダクトイメージの特徴ベクトル

イメージ言葉	高関心度CMにおけるプロダクトイメージ					低関心度CMにおけるプロダクトイメージ		
	CM1	CM3	CM6	CM7	CM8	CM2	CM4	CM5
ナチュラル	24	22	15	13	25	27	19	21
はなやか	4	5	4	6	3	0	4	4
上品	8	7	6	7	4	8	8	4
強い	3	3	0	2	3	0	2	2
可愛い	12	17	13	9	8	14	17	17
色っぽい	3	1	0	4	3	5	2	0
深い	1	0	0	6	4	0	2	0
素敵	3	1	2	0	3	0	0	0
大人な	5	3	6	6	4	0	4	2
濃い	4	4	11	13	6	11	6	11
薄い	3	3	9	4	3	3	8	4
明るい	3	3	2	4	4	3	2	4
落ち着いた	3	0	0	0	4	0	2	2
綺麗	22	33	32	28	25	30	25	30

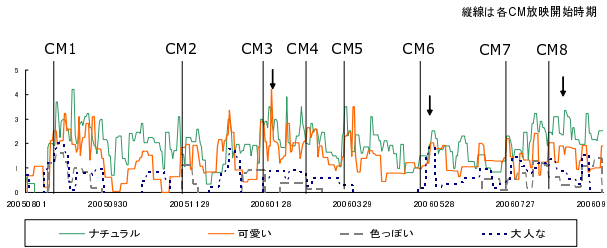


図 6 化粧品プロダクトイメージの時系列グラフ

は、相関が認められなかった例である。時間差 0 付近にピークはなく、そこからはずれた位置で不規則に振動している。相互相関が認められなかったイメージ語彙は、“繊細”の 1 語であった。

4.4 特徴ベクトルの生成

CM への関心度時系列について、8 つの CM それぞれに対応する反応区間 (図 4 の縦線で示した地点を始点として、17 日間の区間) における時系列成分の関心度の平均値を求め、これらを要素とする特徴ベクトルを生成した。表 3 (a) に、CM への関心度の特徴ベクトルを示す。図 4 でも確認したとおり、CM1, CM3, CM6, CM7, CM8 の特徴ベクトルの成分値は、CM2, CM4, CM5 に比較して大きな値を示している。

同様に、プロダクトイメージの時系列 (但し、相互相関が認められた語彙に限る) から生成した特徴ベクトルを表 3 (b) に示す。なお、特徴を単純化するため、

各 CM の特徴ベクトルの成分構成比率が非常に小さく (2%以下), CM 間での差異が微小な語彙は除外した。表 3 (b) 中、低関心度 CM におけるプロダクトイメージにおいて高い値を示しているものがあるが、これは、CM への関心度は低かったものの、その期間で特定のプロダクトイメージを強く誘発していることを示す。

5. 実験結果の分析

マーケティングの実務では、アンケート調査で表 3 のような結果を取得した後、データの相関関係から因果関係を推定するような作業を通してマーケティングに有用な知見を得ようとする事も多い。例えばテレビ CM の実務では、誰を起用タレントに選ぶかが広告効果を左右する重要なファクターである。CM 放映の結果、起用タレントのイメージがプロダクトイメージと整合したのかを知ることは、起用タレントの適切性を評価するうえで重要視されている [秋山 07]。

本章ではそのような評価の 1 例として、表 3 の結果をベースに、コレスポンス分析を用いてタレントイメージとプロダクトイメージとの関係を分析する。コレスポンス分析は多次元ベクトルを二次元平面に射影し、複数のベクトル間の類似性、差異を平面上の距離として可視化する手法である。特徴を単純化して捉えられるため実務で一般的に採用されている [小塩 04]。

以下では提案手法による解析結果 (表 3 の結果) が実務にも利用可能である事を示し前節の結果がマーケティングの分野において重要な意味を持つ事を説明する。なお本章の分析中、相関関係から因果関係を推定する過程は理論的な裏付けに乏しいが、実務利用上は有用な知見と考

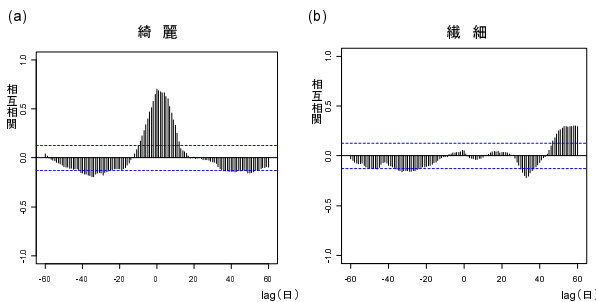


図 7 化粧品シリーズ CM の関心度とプロダクトイメージの相互相関グラフ

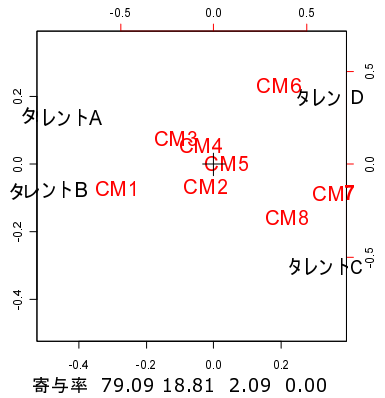


図 8 コレスポネンスプロット (CM への関心)

え記載した。

5.1 特徴ベクトルの可視化

特徴ベクトルの解釈を容易にするため、表 3 (a),(b) それぞれをコレスポネンス分析を用いて 2 次元平面に可視化した (図 8, 図 9)。CM への関心度の特徴ベクトルにコレスポネンス分析を適用した結果 (図 8) を見ると、CM1, CM3, CM6, CM7, CM8 は、いずれかのタレント名のプロット位置付近に現れ、各 CM の演出上の特徴を反映している。例えば、CM6 は、タレント C, タレント D が共演した CM である (表 1 参照) が、これを反映してタレント D 付近にプロットされている。CM7 も同様に主演したタレント C 付近にプロットされている。一方、ベクトル成分値が小さい CM2, CM4, CM5 は、全タレントの中間地点に現れ個々 CM の演出による差異がほとんど見られない。なお、CM3 も中間位置に現れているが、特徴ベクトル成分を見ると、タレント B, タレント D の要素値が大きく、それらの中間にプロットされたものと見られる。CM3 は、タレント B, タレント C が主演した CM である。

次に、プロダクトイメージの特徴ベクトルにコレスポネンス分析を適用した結果を見る。図 9(a) は、関心度の高かった CM のプロット、(b) は、関心度の低かった CM のプロットである。図 9(a) を見ると、“可愛い” という語彙は表 3 (a) のとおり、CM1, CM3, CM6 で高い成分値を示しており、それらのちょうど中間地点にプロットされている。語彙“ナチュラル”は、CM1, CM3, CM8 で高い成分値を示しており、それらの中間地点である CM1 の付近にプロットされている。同様に、“はなやか”は、CM3, CM7 で高い成分値を示し、それらの中間にプロットされている。CM3, CM6, CM8 の主演には、いずれもタレント B が含まれ、なおかつこれらの CM 区間ではタレント B への関心が集中している。同様に、CM3, CM7 の主演には、いずれもタレント C が含まれ、かつ関心が集中している。各 CM 区間でのプロダ

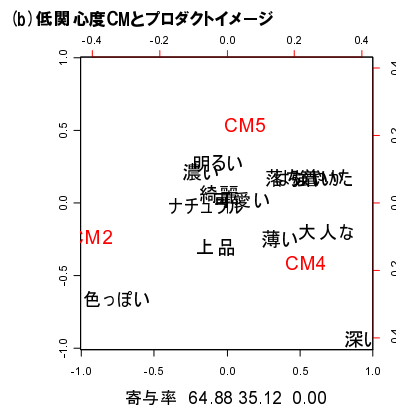
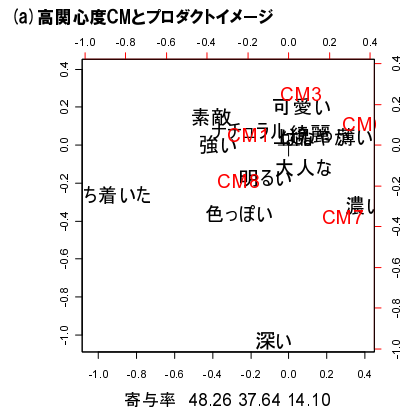


図 9 コレスポネンスプロット (プロダクトイメージ)

クトイメージの違いは、関心と呼んだ主演タレントへのイメージの違いを反映している可能性がある。

5.2 タレントイメージとプロダクトイメージの関連性

各 CM 区間でのプロダクトイメージと、タレント自身のもつイメージとの類似性を評価した。タレントのイメージを表す語彙は、化粧品の掲示板でも出現するが、出現頻度が低く、なおかつ CM のメイキング等他の要素から影響を受けている可能性もあるので、別途、各タレントのファンが書き込む掲示板*4から取得した。この掲示板から生成したタレントイメージ語彙を成分とする特徴ベクトル、およびそれにコレスポネンス分析を行った結果を図 10 に示す。

表 3 (b) と図 10 から各 CM 区間でのプロダクトイメージと、主演タレントのイメージの類似性を分析した結果を表 4 に示す。表中 a は、高い関心度を示した CM 区間の分析結果、b は相対的に関心度が低かった CM 区間の分析結果である。まず、図 10 の特徴ベクトルにおいて、各語彙の成分値を 4 人のタレント間で比較し、成分値が最も高いタレントをその語彙が特徴付けているものとして抽出した。例えば、語彙“自然”の図 10 での成分値は、タレント A, タレント B, タレント C, タレント D そ

*4 2ちゃんねる：http://www.2ch.net

(a) 特徴ベクトル(タレントイメージ)

イメージ語彙	タレントA	タレントB	タレントC	タレントD
かっこいい	0	2	0	0
くっきり	6	0	0	0
ぱっちり	0	0	0	2
はなやか	0	0	2	1
ぶりっこ	0	1	0	1
可愛い	14	23	17	70
自然	4	14	4	1
柔らかい	0	2	0	0
上品	0	0	4	1
色っぽく	3	8	0	0
清楚な	0	0	2	0
素敵	1	2	1	0
大人な	0	1	2	3
知的	0	0	3	0
透明な	0	2	0	0
美しい	49	13	43	15
目立つ	1	4	1	1
綺麗	21	26	22	6

(b) コレスポネンスプロット(タレントイメージ)

表 4 タレントイメージとプロダクトイメージの関連性

	主演	高関心度 タレント	タレントを特徴付けるイメージ語彙											タレントイメージにないプロダクトイメージの特徴語彙			
			タレントA			タレントB			タレントC			タレントD					
			くっきり	美しい	自然	綺麗	可愛い	目立つ	素敵	上品	知的	大人な	華やか		美しい	綺麗	可愛い
a. 高関心度 CM	CM1	タレントB								◎◎		○				○	強い 落ち着いた
	CM3	タレントB タレントC	タレントD タレントB			◎	◎						○	◎	○	強い	
	CM6	タレントD タレントB	タレントD			○	○			○		◎		○	◎	薄い	
	CM7	タレントC	タレントD タレントC		○		○				○	◎◎			◎	深い 濃い 明るい	
	CM8	タレントB タレントD	タレントC タレントD		○					◎						強い 深い 濃い 明るい 落ち着いた	
	b. 低関心度 CM	CM2	タレントA	なし		◎◎	◎				◎		◎				濃い
		CM4		タレントD						◎		◎			◎		強い 薄い
		CM5	タレントC タレントA	タレントD					◎				◎		◎		強い 濃い 明るい

図 10 タレントイメージの特徴ベクトルとコレスポネンスプロット

それぞれ、4, 14, 4, 1 なのでタレント B を特徴付ける語彙として抽出している。表 4 中の“タレントを特徴付けるイメージ語彙”は、この要領で抽出した語彙群である。次に、表 3 (b) の特徴ベクトルにおいて、各語彙の成分値を 8 つの CM 区間で比較し、成分値が最も高い CM 区間、もしくは次に高い CM 区間をその語彙が特徴付けているものとし、タレントを特徴付けるイメージ語彙と共通している場合に、該当箇所に もしくは でマーキングした。例えば、語彙“綺麗”は、CM3 における成分値が 33 で最も高く、CM6 における成分値が 32 で次に高いので、CM3, CM6 を特徴付ける語彙と見なしている。この語彙は、タレント B, タレント C を特徴付ける語彙でもあるので、表中 CM3, CM6 の行と、タレント B, タレント C の語彙“綺麗”の列とが交差する箇所にマーキングしている。は、CM 期間での成分値が最も高い語彙、は、その次に成分値が高い語彙である。更に、各 CM 区間 (各行) でマーキング箇所が相対的に多いタレントがその CM で主演したタレントと一致する場合は、該当箇所を網掛け表示している。例えば、CM3 の行を見ると、タレント B, タレント C へのマーキング数

が他のタレントよりも大きい。タレント B, タレント C は CM3 の主演タレントなので、該当箇所を網掛け表示している。つまり、網掛け表示は主演タレントと、その CM 期間におけるプロダクトイメージとのイメージの一致性を表す。

網掛け表示箇所を各行についてみると、高い関心を呼んだ CM の期間 (表中 a) では、明らかに CM 区間毎のプロダクトイメージの特徴は、主演したタレントイメージを反映していることがわかる。また、主演したタレントは、図 8 で見たとおり、CM への関心の中心でもある。一方、関心度が低かった CM の期間 (表中 b) では、網掛け表示箇所は見られず、プロダクトイメージの特徴と主演タレントのイメージとの共通性は認められない。また、図 8 で見たとおり、CM への関心の高かったタレントと主演タレントとの間にも共通性がない。

以上の分析から、この化粧品 CM は、以下のような特性・効果があったことがわかる。

- (1) CM のインパクトが持続した期間は、平均 17 日であった。
- (2) CM によって高い関心度を得たものとそうでないものがある。
- (3) 4.3 に示したように CM のオンエアを契機としたタレントに対する関心度の変化はプロダクトイメージの変化よりも早く現れる。その点もデータの相関関係を合せて考慮すると、高い関心度を得た CM では、主演タレントに関心が集まり、そのイメージがプロダクトイメージに影響を及ぼした事が推察できる。
- (4) 関心度が低かった CM では、プロダクトのイメージと CM への関心との間での関連性は認められず、CM によるイメージ形成効果は認められなかった。

なお、プロダクトイメージ語彙には、タレントイメージ語彙には出現しない語彙もある。表 4 の最右列にこれらの語彙を記載した。図 9 (a) において、 $y=0.0$ を境界にこれらの語彙を見ると、上側には、“薄い”が、下側

には、“濃い”，“深い”がプロットされている．上側にプロットされるCM1，CM3，CM6は，春夏物の化粧品のCM，下側にプロットされるCM7，CM8は，秋冬物化粧品のCMである．このようにプロダクトイメージ語彙のなかには，主演タレントのイメージの他に，季節的イメージを反映していると思われる語彙も見られる．同じタレントを起用しても，CMのメイキングによってタレントのイメージは変化する．上記の語彙は，季節性に応じたCMのメイキングによってイメージが変化したことを反映したものであると思われるが，詳細な分析は今後の研究に譲りたい．

6. 結 論

本稿では，テレビCMによりプロダクトイメージがオンラインコミュニティ上でどのように形成されるかを捉える手法を提案した．本提案によれば，テレビCMによるプロダクトイメージの形成効果に関する以下のようなデータを取得することができる．

- (1)CMへの関心が持続した期間
- (2)CMが呼び起こした関心と関心度
- (3)CMへの関心と時系列な相関をもつプロダクトイメージの上記期間における特性

高関心度CMでは，主演したタレントのイメージがプロダクトイメージと類似するなど，CMによるプロダクトイメージ形成効果を裏付ける特徴が抽出された．一方，低関心度CMでは，主演したタレントのイメージとプロダクトイメージとの類似性は認められずプロダクトイメージ形成効果は認められなかった．このように本手法を用いることで，掲示板上の書き込みの文脈からは読み取れない，CMへの関心とプロダクトイメージとの関連性を捉えられることを確認した．

本稿では，テレビCMを時間的事件とするインターネットコミュニティでのイメージ形成効果を分析した．提案手法は，テレビCMに特化したものではなく，広くニュースやキャンペーンなど様々な時間的事件を契機とするインターネットコミュニティ上での意見形成の解析に応用できる可能性がある．一方「外部イベントの発生(CMの放送)だけでなく，消失(CMへの関心の消失)後のプロダクトイメージにどう影響を残すか．それがブランドイメージの形成にどうつながっていくか？」等本研究でカバーできない重要な課題も多い．今後，こうした様々な時間的事件の分析に応用範囲を広げ，提案手法の有効性を確認したい．

謝 辞

マーケティング研究の視点から貴重なアドバイスを下さった筑波大学 西尾教授に感謝致します．

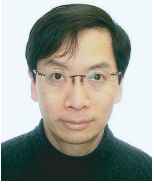
◇ 参 考 文 献 ◇

- [Aaker 81] Aaker, D. and Bruzzone, E. : Viewer perceptions of prime-time television advertising, *Journal of Advertising Research*, Vol.21, pp.15-23 (1981).
- [秋山 07] 秋山隆平：情報大爆発，宣伝会議 (2007) .
- [CM INDEX] “月刊CM INDEX,” 2005年8月号-2006年9月号
- [Gruhl 04] Gruhl, D. , Guha, R. , Liben-Nowell, D. , and Tomkins, A. : Information Diffusion Through Blogspace, *Proceedings of WWW Conference*, pp.491-501 (2004).
- [Haga 02] Haga, H. : Combining video and bulletin board systems in distance education systems, *Elsevier Science, Internet for Higher Education*, Vol.5, pp.119-129 (2002).
- [飯田 05] 飯田龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一:意見抽出を目的とした機械学習による属性-評価値対同定, 情報処理学会研究報告「自然言語処理」, 2005-NL-165(4), pp.21-28 (2005) .
- [井上 04] 井上智雄, 小林哲郎, 池田謙一, 重野寛, 岡田謙一:ウェブ掲示板を対象としたネットワークコミュニティ分析支援システム:CMINER, 情報処理学会論文誌, Vol.45, No.1, pp.131-141 (2004) .
- [石井 05] 石井恵, 井沢味奈子, 片岡良治:ネット上の会話からの話題即時抽出技術の評価について, 情報処理学会研究報告「情報学基礎」, 2005-FI-80(6), pp.33-40 (2005) .
- [鍛冶 07] 鍛冶伸裕, 喜連川優:自動構築した評価文コーパスからの評価表現の獲得, 電子情報通信学会, 第18回データ工学ワークショップ, D2-5 (2007) .
- [加藤 07] 加藤智明, 中谷有紀:CGMマーケティング, 毎日コミュニケーションズ (2007) .
- [小林 06] 小林のぞみ, 乾健太郎, 松本裕治:意見情報の抽出/構造化のタスク仕様に関する考察, 情報処理学会研究報告「自然言語処理」, 2006-NL-171(18), pp.111-118 (2006) .
- [小塩 04] 小塩 真司:SPSSとAmosによる心理・調査データ解析 因子分析・共分散構造分析まで, 東京図書 (2004) .
- [宮森 05] 宮森恒, 中村聡史, 田中克巳:番組実況チャットを利用したテレビ番組のメタデータ自動抽出方式, 情報処理学会論文誌, Vol.46, No.SIG 18, pp.59-71 (2005) .
- [村上 06] 村上明子, 渡辺日出雄:時系列情報を利用した複合語キーワードの抽出, 情報処理学会研究報告「自然言語処理」, 2006-NL-172(1), pp.1-8 (2006) .
- [Schlinger 79] Schlinger, M. : A profile of responses to commercials, *Journal of Advertising Research*, Vol.19, pp.37-46(1979).
- [Uehara 05] Uehara, H. and Yoshida, K. : Anotating TV Drama based on Viewer Dialogue -Analysis of Viewers' Attention Generated on an Internet Bulletin Board-, *SAINT2005*, pp.334-340 (2005)
- [Wells 71] Wells, D. , Leavitt, C. and McConville, M. : A reaction profile for TV commercials, *Journal of Advertising Research*, Vol.11, pp.11-15 (1971).
- [山崎 07] 山崎秀夫, 兼元謙任:ネット広告がテレビCMを超える日, 毎日コミュニケーションズ (2007) .
- [矢野 04] 矢野宏実, 目良和也, 相沢輝昭:嗜好を考慮した評判情報検索手法, 情報処理学会研究報告「自然言語処理」, 2004-NL-164(28), pp.165-170 (2004) .
- [横山 07] 横山隆治, 関良樹, 松矢順一, 太駄健司, 三輪宗久:次世代広告コミュニケーション, 翔泳社 (2007) .

〔担当委員：松村 真宏〕

2007年9月13日 受理

— 著 者 紹 介 —



上原 宏(正会員)

筑波大学大学院ビジネス科学研究科博士後期過程在籍。インターネットコミュニティにもとづく広告効果の解析，メタデータの自動生成技術の研究に従事（株）NTT ドコモ 法人営業本部担当課長。情報処理学会，各会員。



佐藤 忠彦(正会員)

2004 年総合研究大学院大学数物科学研究科統計科学専攻，博士（学術）。現在，筑波大学大学院ビジネス科学研究科・准教授。計量的マーケティング研究に従事。日本統計学会，応用統計学会，日本マーケティングサイエンス学会，各会員。



吉田 健一(正会員)

1980 年東京工業大学理学部情報科学科卒業，同年，日立製作所入社。1992 年 9 月 博士（工学）。2002 年より筑波大学大学院ビジネス科学研究科教授。インターネット上の各種データを，機械学習の手法を使って解析する研究に従事。情報処理学会，各会員。