CH-003

# 場面のラベル付けによるＣM好感度の分析

# Analysis of TV commercial favorability by scene labeling

福寄 幹太† 松井 くにお† 辰元 晃‡
Kanta Fukuyori Kunio Matsui Koh Tatsumoto

## 1. はじめに

テレビを通じて企業やその商品を宣伝するテレビコマーシャル（以下，CM）は，1953 年から始まった．CMは，今日までに数多くの企業と商品の宣伝を行ってきており，今後も発展していくことが期待できる．そんな日本のCMの総広告費は，年々増加の傾向にあり，2019 年で 6 兆 9381 億円にまで上った[1]．このように，CMには多額の費用が投じられていると分かる．しかし，どのようなCMが視聴者からの好感を得られるかというのは，ビッグデータ社会の今日でも未だに解明されていない．

CM好感度調査を 30 年以上行っている株式会社東京企画CM総合研究所では，毎秒ごとのCM好感度の分析に成功した[2]．しかし，秒単位では各秒数での映像がどのようなものか分からないという問題がある．そこで，よりデータに意味を持たせるために場面ごとに評価する必要があると考えた．そのため，本研究ではCMを場面ごとに分割し，どのような場面が高い好感度を得られるのかというCM制作に有益な情報を分析する．

## 2. 研究概要

### 2.1 研究の流れ

研究の流れ図を図 1 に示す．

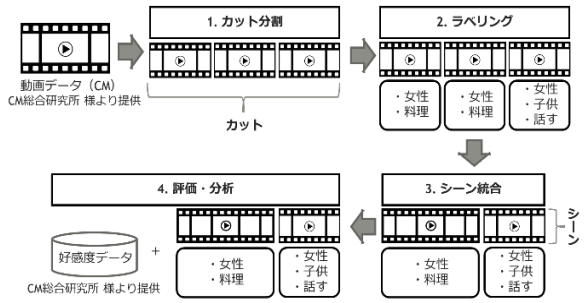


**図 1　研究の流れ図**

流れとしては，まず，動画データであるCMを映像の基本単位であるカット（カメラが切り替わるまでの映像のこと）ごとに分割する．この分割した各カットに対して，物体認識で「何が映っているか」，動作認識で「何をしているか」のラベルを付与する．続いて，類似しているラベルかつ隣接しているカットをシーン（カットの集合で表現の最小単位）として統合する．このラベルの付いたシーンを 1 つの「場面」として扱う．作成したこの場面データ（場面ごとの時間データ＋ラベルデータ）をCMの好感度データとマッチングさせて評価・分析を行う．なお，評価・分析はシーンごとに行う．

### 2.2 利用データ

本研究で利用するCMの動画データ，好感度データは，共同研究先の株式会社東京企画 CM総合研究所より提供していただいている．動画データは，15 秒尺の 1000 種類のCMであり，第 3 章で作成する分析データはこの 1000CMのデータで作成する．また，好感度データは，CM総合研究所の「Mnavi」というモデルによって算出された毎秒ごとのスコアデータを使用している．Mnavi とは，fMRI による画像特徴量と脳活動（血流）のマッチングから，CM視聴時の人間の脳活動を実測したデータをもとに作られた仮想脳モデルである[2]．今回，この Mnavi によって，動画データと同じ 1000CMで算出した毎秒ごとのスコアデータを使用する．毎秒ごとのスコアデータには，CM好感度，試用意向，好感要因というデータが存在する．まず，CM好感度とは，アンケートからCMの好感度を数値化したものである．次に，試用意向とは，消費者の「購入したい」や「試してみたい」という度合いを数値化したものである．さらに，好感要因とは，CMのどのような要因が好感を得たのかを数値化したものである．この好感要因は 15 項目存在し，それぞれの要因を数値化している．15 項目の好感要因とその略称（以後，略称で明記）を表 1 に示す．

表 1　15 項目の好感要因と略称

| 好感要因 | 略称 |
|---|---|
| 出演者・キャラクター | 出演者 |
| ユーモラスな所が | ユーモラス |
| セクシーだから | セクシー |
| 宣伝文句が印象的 | 宣伝文句 |
| 音楽・サウンドが印象的 | 音楽・サウンド |
| 商品にひかれた | 商品にひかれた |
| 説得力に共感した | 説得力に共感 |
| ダサいけど憎めない | ダサイけど憎めない |
| 時代の先端を感じた | 時代の先端 |
| 心がなごむ | 心がなごむ |
| ストーリー展開がおもしろい | ストーリー展開 |
| 企業姿勢にウソがない | 企業姿勢 |
| 映像・画像がよい | 映像・画像 |
| 周囲の評判もよい | 周囲の評判 |
| かわいらしい | かわいらしい |

第 4 章で行う分析では，このCM好感度と試用意向，15 項目の好感要因の計 17 項目のデータとマッチングさせて行う．

† 金沢工業大学　Kanazawa Institute of Technology
‡ 株式会社 東京企画　Tokyo Kikaku

## 3. 分析データの作成

本章では，カット分割，ラベル付け，シーンの統合により分析用の場面データを作成する．

### 3.1 カット分割

#### 3.1.1 カット点検出の手法

カット分割する上で必要になる技術に，カット点検出がある．カット点検出とは，動画データから編集点であるカット点を自動で検出して，映像として扱いやすい単位に分割する技術である．方法としては，前後のフレームで画素単位での色情報の差異を取り，大きく変化した部分をカット点として検出する．今回，カット点は自動で検出するため，判定のためのデータが必要になる．そこで，前フレームとの変化の度合いを示すデータ（以下，変化割合）を使い，閾値を超えた場合にカット点とするようにした．変化割合の算出には，隣接するフレームの画像同士で差分画像を作成し（式(1)），その差分画像の平均二乗誤差（式(2)）を変化割合としている．ここで，$I(x,\ y)$とは$(x,\ y)$のピクセル位置の画素データ（RGB）であり，$I_b(x,\ y)$は 1 つ前のフレーム画像，$I_d(x,\ y)$は差分画像の画素データである．

$$I_d(\mathrm{x},\mathrm{y}) = I(x,y) - I_b(x,y) \tag{1}$$

$$\mathrm{MSE} = \frac{1}{n}\sum I_d(\mathrm{x},\mathrm{y})^2 \tag{2}$$

算出した変化割合の例を図 2 と図 3 に示す．

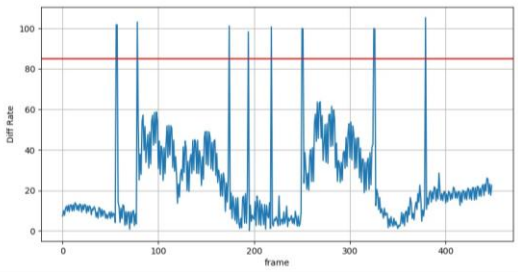


図 2　検出し易い変化割合

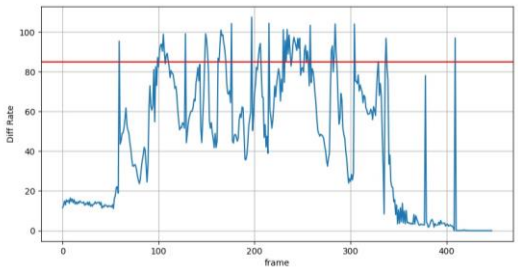


図 3　検出しづらい変化割合

図 2 の大きく突出した部分がカット点に該当する．このようにカット点は変化割合の値が一時的に大きくなる特徴があり，この値が閾値を超えた場合にカット点と判定する．しかし，実際には図 3 のような閾値を連続で何度も超えるような変化割合のＣＭが多く，カット点が検出しづらいため，工夫したカット点検出をする必要があった．

そこで，変化割合が閾値を超えたフレームを暫定的なカット点として，その中から不正解の疑いがあるフレームをカット点から削除していく．このような段階的なカット点の修正を行うことで，閾値を何度も超えるＣＭの対策を行った．これらを踏まえたカット点検出のフローチャートを図 4 に示す．

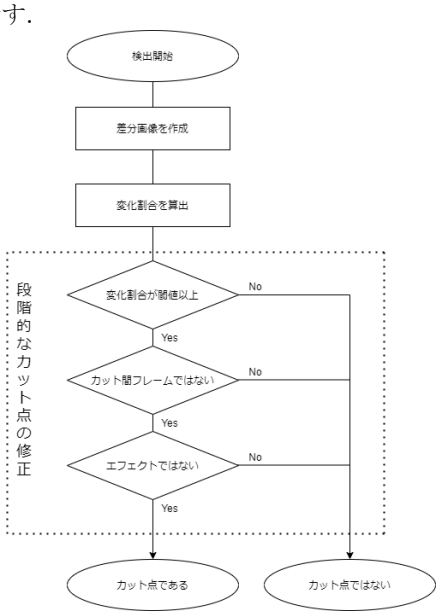


図 4　カット点検出のフローチャート

図 4 のカット間フレームとエフェクトの説明と判定方法について説明する．

カット間フレームとは，カットとカットの間に現れる不要なフレームである．これは，カット点と誤検出することがあり対策する必要があった．対策として，カット間フレームかどうかを判定して削除することにした．判定方法は，該当フレームと前後のフレームそれぞれで 2 つの差分画像を作成し，その作成した差分画像同士の差分画像を更に作成する．作成した差分画像の平均二乗誤差が閾値以下の時にカット間フレームとして判定した．これは，2 つの差分画像は非常に酷似した画像が生成されるため，平均二乗誤差は小さくなるためこのような方法を取った．

続いて，エフェクトの判定方法について説明する．ここでいうエフェクトとは，瞬間的な映像効果のことである．具体的には，人工的な光であるフラッシュや瞬間的に物体がカメラの前を通るなどの数フレームだけ何らかの変化の後に数フレーム前の映像に戻ることである．このエフェクトもカット点と誤検出してしまうためエフェクトを検出・削除の対策を講じる．エフェクトの判定方法は，カット点と次のカット点を輝度ヒストグラムとヒストグラムインタセクションを使用して，画像同士の類似度を算出し，類似度が高い時には該当フレームをカット点から削除する．これは，隣接するカット点は基本的に非常に異なる画像になる性質を利用している．

#### 3.1.2 カット点検出の評価

図 4 のフローチャートの処理を行った後，カット点検出の精度を評価した．このカット点検出の精度が高ければ高いほど，次の工程である各場面へのラベル付けが過不足無く行える．評価方法は，検出したフレームがカット点かどうかを示した正解データを 120ＣＭ分作成し，適合率，再現率，Ｆ値，正答率の 4 項目を算出して評価を行った．ここでの正答率とは，完璧に分割出来たＣＭ数の割合を指す．評価結果の推移と関連研究との比較を表 2 に示す．

表 2　評価結果の推移と関連研究との比較

| | 閾値以上 | カット間フレーム削除後 | エフェクト削除後 | 関連研究 |
|---|---|---|---|---|
| 適合率 | 62.19% | 70.37% | 87.39% | 72.00% |
| 再現率 | 91.02% | 90.72% | 86.99% | 89.00% |
| F 値 | 73.90% | 79.26% | 87.19% | 80.00% |
| 正答率 | 24.17% | 40.00% | 38.33% | |

間違っているカット点だけを削除して，適合率を大幅に上げつつ，正しいカット点はなるべく削除しないようして，再現率は下がらないように行ってきた．そのため，総合的な評価ができるＦ値が，関連研究[3]と比べて 7% 以上も高くなり，本研究のカット分割の精度は関連研究を上回った．

### 3.2 ラベル付け

本研究のラベル付けでは，画像・動画像に映っている物体を幅広いカテゴリーで制約なく認識する一般物体認識を行う．一般物体認識は，個人単位での一からの学習が困難となるため事前学習されたモデルを使用する．分割した各カットに対して物体認識で名詞ラベルを，動作認識で動詞ラベルを付与して，最後にラベルを翻訳・整形・結合する．

#### 3.2.1 物体認識によるラベル付け

物体認識には，Google が作成した事前学習モデルである Faster-RCNN[4]を使用した．まず，RCNN（Regional CNN）とは，最初に物体が含まれている可能性のある領域を提案し，その後，提案した領域内の物体を認識するという 2 段階構造のオブジェクト検出のアルゴリズムである．この RCNN を高速化に発展させたものが Faster-RCNN である．Faster-RCNN の特徴は，ネットワーク内に RPN（Region Proposal Network）という領域提案のネットワークを追加して直接行っているため，領域提案から画像認識（クラス分類）までを CNN 化することにより，高速化を実現していることである[5]．次に，データセットは，Open Image v4 を使用しており，最大 600 種類のラベルを認識できる．これにより，汎用性の高いラベルの付与が可能となる．

分割した各カットからブレの少なかった，カットの最後のフレーム画像を 1 枚取り出し，上記のモデルを使用して物体認識による名詞のラベル付けを行った．その際，ラベルのスコアが 0.25 以上の時のみ付与を行っている．

#### 3.2.2 動作認識によるラベル付け

動作認識とは，人間が行う「食べる」や「踊る」などの動作を認識する技術で，近年更なる活発化を見せている．この動作認識を用いて，各カットの動画データを読み込ませて動詞のラベル付けを行った．使用したのは，画像認識において高い予測性能を持つ ResNet(Residual Network)の畳み込み層を 3D 領域に拡張した 3D-ResNet をモデルとし，400 種類の人間の動作を認識できる kinetics をデータセットとしている事前学習されたモデル[6]を使用して行った．3D-ResNet は，入力した動画に対して，2D の空間情報と 1D の時間情報をまとめて 3D の畳み込みを行うことで，時空間情報を考慮した動画の動作認識が可能となる[7]．入力の際は，動画を 16 フレームの長さに分割して入力し，その入力ごとにクラス分類を行う．なお，16 フレーム未満の短いカットには何のラベルも付与せず，32 フレーム以上のカットで異なるラベルが付与された場合，その異なるラベル全てを付与する．また，物体認識と同様に認識スコアが閾値以上（7.0 以上）のラベルを付与している．

#### 3.2.3 ラベルの翻訳と整形，結合

物体認識と動作認識によるラベル付けの結果は，英語のラベルが付与されているが，結果の見やすさなどの観点から，日本語のラベルに翻訳する．その際，翻訳には Google 翻訳を使用している．ただ，「stretching leg」の翻訳結果は「ストレッチ脚」のような意味の通らないラベルのまま出力されてしまう．これを避けるために，自動翻訳ではカバーできない部分を「足のストレッチ」のように手動による意訳をすることで補っている．また，「顔」や「人」，「服」といったどのカットにも付くような冗長なラベルは，今後の分析の時にノイズとなるため削除する．

このように，物体認識と動作認識のラベルを翻訳・整形した後に，同じカットのラベルは 1 つに結合してラベル付けの工程を終える．

### 3.3 シーンの統合

隣接するカットかつ類似するラベル同士は統合してシーンとし，統合しなかったカットは単一でシーンとする．統合するか否かは，隣接するカットのラベルに対して，TF-IDF とコサイン類似度から算出したラベル同士の類似度で判定し，0.93 以上の場合に統合した．ラベルの統合では，共通のラベル + お互いに無いラベルで結合し，カットの時間データは，前カットの開始時間から後ろカットの終了時間とする．

以上の方法で作成した分析データは，全部で 8194 シーンとなった．

## 4. CM 好感度との分析と評価

第 3 章で「シーンごとの時間データ＋ラベルデータ」の場面データの作成を行った．本章では，このデータに「毎秒ごとのＣM好感度データ」を時間軸でマッチングさせ，好感度が高い時，低い時のシーンのラベルを分析する．

マッチングには，シーン範囲に近い秒数の好感度の平均値をそのシーンの好感度として付与する．続いて，各シーンの好感度を降順にソートし，上位と下位の各 1000 シーンに付与されているラベル件数 TOP10 を調査した．その結果を表 3 に示す．

表 3 各区分 1000 シーンに付与されたラベル結果

| 順位 | 上位 | | | 下位 | | |
|---|---|---|---|---|---|---|
| | ラベル | 件数 | 正規化 | ラベル | 件数 | 正規化 |
| 1 | 女性 | 415 | 0.13 | 女性 | 393 | 0.13 |
| 2 | 男性 | 318 | 0.13 | ポスター | 294 | 0.18 |
| 3 | 女の子 | 145 | 0.14 | 男性 | 219 | 0.09 |
| 4 | ポスター | 134 | 0.08 | 女の子 | 137 | 0.14 |
| 5 | おもちゃ | 112 | 0.35 | クリームを塗る | 65 | 0.45 |
| 6 | 食べ物 | 83 | 0.19 | ドレス | 62 | 0.18 |
| 7 | ドリンク | 49 | 0.15 | ボトル | 61 | 0.16 |
| 8 | お菓子 | 46 | 0.19 | 携帯電話 | 60 | 0.31 |
| 9 | ボトル | 45 | 0.12 | 手 | 40 | 0.14 |
| 10 | 靴 | 44 | 0.14 | スーツ | 37 | 0.12 |

単純な件数だけだと、よく付与されるラベルが高い順位にくるため，各ラベル件数を全シーンに付与されたラベル

件数で除算して正規化した値も確認した．その結果，上位の「おもちゃ」が 0.35，下位の「クリームを塗る」が 0.45 と特に数値が高くなった．その区分（上位もしくは下位）にだけ付与された特有のラベルやその区分で多く付与されたラベルは重要度が高いと判断し，そのラベルが付与されることで好感度は増減しているのではないかと考察する．これが正しいと仮定すると，「おもちゃ」というラベルが付与されると好感度が高くなり，「クリームを塗る」のラベルが付与されると好感度が低くなるということになる．

同様な方法で他の好感要因の結果についても見ていく．なお，ここでは筆者が特に興味深いと感じた「試用意向」と「商品にひかれた」，「出演者」の好感要因について触れたいと思う．

まず，「試用意向」での結果を表 4 に示す．

**表 4　「試用意向」での結果**

| 順位 | 上位 | | | 下位 | | |
|---|---|---|---|---|---|---|
| | ラベル | 件数 | 正規化 | ラベル | 件数 | 正規化 |
| 1 | 女性 | 363 | 0.12 | 女性 | 396 | 0.13 |
| 2 | 男性 | 281 | 0.12 | ポスター | 329 | 0.20 |
| 3 | 食べ物 | 142 | 0.33 | 男性 | 247 | 0.10 |
| 4 | 女の子 | 130 | 0.13 | 女の子 | 143 | 0.14 |
| 5 | ポスター | 123 | 0.08 | クリームを塗る | 59 | 0.41 |
| 6 | おもちゃ | 107 | 0.33 | ドレス | 56 | 0.16 |
| 7 | お菓子 | 69 | 0.29 | ボトル | 50 | 0.13 |
| 8 | ドリンク | 58 | 0.18 | 携帯電話 | 50 | 0.26 |
| 9 | 調理する | 53 | 0.35 | スーツ | 47 | 0.15 |
| 10 | 女性 | 363 | 0.12 | 木 | 41 | 0.10 |

先ほどのＣM好感度と比べて，「食べ物」，「調理する」，「食べる」などの食事関連のラベルと，「おもちゃ」という商品のラベルが多く付与されていた．これは，試用意向という評価指標の特性が大きく関わっているためである．試用意向は，消費者の「購入したい」という度合いを数値化したものであるため，視聴者が食事シーンや調理シーンなどを見て，自分もこれを食べたいと感じた結果だと思われる．また，この試用意向の結果は，「商品にひかれた」という好感要因の結果と非常に似たものであった．このように食事している場面を挿入することで，「試用意向」や「商品にひかれた」の数値は大きく増加すると考察する．

続いて，「出演者」の好感要因での結果を表 5 に示す．

**表 5　「出演者」の好感要因での結果**

| 順位 | 上位 | | | 下位 | | |
|---|---|---|---|---|---|---|
| | ラベル | 件数 | 正規化 | ラベル | 件数 | 正規化 |
| 1 | 女性 | 363 | 0.12 | 女性 | 396 | 0.13 |
| 2 | 男性 | 281 | 0.12 | ポスター | 329 | 0.20 |
| 3 | 食べ物 | 142 | 0.33 | 男性 | 247 | 0.10 |
| 4 | 女の子 | 130 | 0.13 | 女の子 | 143 | 0.14 |
| 5 | ポスター | 123 | 0.08 | クリームを塗る | 59 | 0.41 |
| 6 | おもちゃ | 107 | 0.33 | ドレス | 56 | 0.16 |
| 7 | お菓子 | 69 | 0.29 | ボトル | 50 | 0.13 |
| 8 | ドリンク | 58 | 0.18 | 携帯電話 | 50 | 0.26 |
| 9 | 調理する | 53 | 0.35 | スーツ | 47 | 0.15 |
| 10 | 女性 | 363 | 0.12 | 木 | 41 | 0.10 |

この好感要因では，「女性」「男性」，「女の子」といった人物のラベルが他の好感要因より多く付与されていた．また，正規化した値では，「踊る」という動作ラベルが 0.23 と高くなっていた．このことから，「出演者」の好感要因は，人物のラベルが非常に重要であること，その人物による「踊る」という動作ラベルも数値を上げる大きな要因になっていることが明らかになった．

## 5. おわりに

本研究では，カット分割，ラベル付け，シーンの統合，マッチング・分析の 4 つの工程を経て，場面ごとのＣM好感度の分析を行った．結果として，「おもちゃが映っている場面」は好感度が高くなり，「クリームを塗っている場面」は好感度が低くなる傾向にあることが分かった．

今後の課題は，まず，カット点検出の更なる精度向上が挙げられる．関連研究と比べて高い精度にはなったが，短時間で様々な表現技法を使用しているＣMでは，やはり難易度が高く，完璧なカット点の検出はまだまだ出来ていない．これについては，徐々に次のカットへ遷移するフェードの検出や物体の動きをベクトルで表現するオプティカルフローによる動きの激しいＣMへの対策が求められる．

次に，誤認識の多かったラベル付けの精度向上も今後の課題であると考えている．このラベル付けの精度が低いと，場面ごとの分析・評価の時に正しい結果が得られないためである．今回の既存モデルを利用するラベル付けでは限界であると分かったため，手動によるアノテーションで補うことで，これを改善したいと考えている．具体的には，場面の動画像とラベルの結果を人間が確認し，ラベルの成否を確認・訂正する．これにより，間違っているラベルの訂正やＣM独自の新たなラベル（商品別等）の付与，訂正結果の追加学習ができるようになる．このように，ＣMに特化させたラベル付けを行うことにより，より詳細な場面の分析が可能となるだろう．

### 謝辞

最後に，データの提供，研究への助言等の多大なる支援を頂きました株式会社東京企画 ＣM総合研究所の皆様に感謝いたします．

### 参考文献

[1] dentsu ： 『2019 年 日本の広告』，〈https://www.dentsu.co.jp/news/release/2020/0311-010027.html〉，（参照 2020-11-16）
[2] ＣM総合研究所 ： 『CM 好感度 AI による 動画評価システム【Mnavi】』，〈https://www.cmdb.jp/service/consulting/m-navi/〉，（参照日 2020-12-25）
[3] 佐藤駿介，青野雅樹：『色特徴量と動き特徴量を用いたショット分割手法』，情報処理学会 第 75 回全国大会講演論文集 5U-7
[4] TensorFlowHub ： 『FasterRCNN Openimages v4』，〈https://tfhub.dev/google/faster_rcnn/openimages_v4/inception_resnet_v2/1〉，（参照日 2020-12-12）
[5] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun :『Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.』，Advances in Neural Information Processing Systems . Vol. 28，2015.
[6] Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh :『Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?』，arXiv preprint，2017
[7] Wei Xu, Ming Yang, Kai Yu, “3D Convolutional Neural Networks for Human Action Recognition”，IEEE Transactions on Pattern Analysis and Machine Intelligence ( Volume: 35, Issue: 1, Jan. 2013)