

# テレビコマーシャルの好感度予測

伊藤 鈴奈<sup>†</sup> 長尾 智晴<sup>††</sup>

<sup>†</sup> 横浜国立大学 理工学部

<sup>††</sup> 横浜国立大学 大学院環境情報研究院

## 1. はじめに

より良い広告制作のためには、その広告の質を適切に評価する必要がある。そのためにテレビコマーシャル(以下、CMと呼ぶ)では、全体の好感度、および、詳細な観点別の点数評価をユーザアンケートで収集しているが、この方法ではコストが非常に大きく評価に時間がかかる。本稿では機械学習法を用い、過去のアンケート結果を学習することで、CM 動画から要因別の好感度を予測する手法を提案する。

## 2. 提案手法

図1に本手法の概要を示す。CM 動画から複数の特徴量を抽出し、それらを入力としたネットワークモデルにて要因別好感度を求める。入力する特徴量は、CM 動画の評価に有効であることが期待される物体、動き、表情、音、および色情報を特徴量として用いる。過学習の抑制を目的として、一部の特徴量において時間軸方向に基本統計量を取る。

### 2.1 特徴抽出

**物体特徴:** 物体特徴の抽出には、物体認識器として学習した VGG16[1]を使用する。このネットワークに動画の各フレーム画像を入力し、畳み込み層後の特徴量を抽出する。各動画の長さを15等分し、epochごとにそれぞれの区間からランダムに1個の特徴量をサンプリングすることで、15個の特徴量を使用する。

**動き特徴:** 連続する2フレーム間の密なオプティカルフローを求める。フローごとに動きベクトルの大きさの最大・最小・平均・中央値を算出する。

**表情特徴:** 各フレーム画像に対して、顔検出とそこで検出された顔に対する表情認識を行い、顔の数と happiness スコアの総和を算出する。時間軸方向の基本統計量は顔が検出されたフレームのみ、全フレームをそれぞれ対象として求める。また、動画における顔が検出されるフレームの割合も算出する。

**音特徴:** 人間の知覚特性を考慮した特徴量である、メル周波数ケプストラム係数を動画音声から算出する。

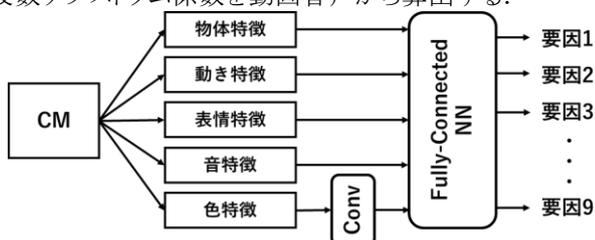


図1. 提案手法の概要

表1. 予測結果

モデル	MAE								
	要因1	要因2	要因3	要因4	要因5	要因6	要因7	要因8	要因9
ベース	0.406	0.417	0.283	0.344	0.349	0.320	0.361	0.296	0.341
提案	0.341	0.322	0.279	0.325	0.313	0.277	0.306	0.280	0.269

**色特徴:** 各フレーム画像に対して、RGB 各8階調で表せる512色の色割合を算出する。動き特徴と同様、15個の色割合特徴量を epoch ごとにサンプリングして使用する。

物体、動き、表情、音特徴に対しては、算出した各値における、時間軸方向の最大・最小・平均値を NN の入力特徴量とする。

### 2.2 好感度の予測

2.1 で動画から抽出した各特徴量を入力とする、1層の畳み込み層と3層の全結合層で構成された NN を用い、要因別好感度の予測を行う。損失関数として全ての予測要因の平均二乗誤差を用い、RMSprop によりパラメータの最適化を行う。

## 3. 実験設定

本実験では、15秒または30秒のCM動画を対象とする。モデル学習に2323本、検証で581本を用いた。要因は順に{出演者・キャラクター、ユーモラス、宣伝文句、音楽・サウンド、商品にひかれた、心がなごむ、ストーリー展開、映像・画像、かわいらしい}の9種類あり、予測精度は平均絶対誤差(MAE)で示す。ベースラインとして、各要因において予測値を固定値としたときの最小 MAE を示す。

## 4. 実験結果

表1に結果を示す。提案する特徴量を用いた好感度の予測は、ベースラインよりも高精度であることが分かるが、実応用の水準には達していない。

ベースラインからの向上率を要因別でみると、要因2{ユーモラス}に対する予測精度が最も高く、使用した特徴量との相関が強かったと示唆される。

## 5. まとめ

本稿では、CMの要因別好感度を予測するモデルを提案した。今後は精度の向上と、CM動画において好感度に影響を与える要素の解析を行い、新たな知見の獲得を目指す。

## 参考文献

[1] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.