

VR 環境と生体情報を用いた映像コンテンツの評価及び分析

Evaluation and Analysis of Video Contents Using Physiological Information and VR Environment

浦地 勇人^{*1}
Hayato Uraji

松村 昂輝^{*1}
Koki Matsumura

フアン ロレンゾ ハガド^{*1}
Juan Lorenzo Hagad

福井 健一^{*2} 沼尾 正行^{*2}
Ken-ichi Fukui Masayuki Numao

^{*1} 大阪大学大学院情報科学研究科

Graduate School of Information Science and Technology,
Osaka University

^{*2} 大阪大学産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

Abstract: By evaluating the effect of the video content, the content creator can know how the viewer felt his work. In this evaluation process, traditionally self-report type questionnaire data has been used. However, since this method involves participant bias, experimenter bias, or human diversity, accurate evaluation is difficult. Also, in order to eliminate these as much as possible, it is necessary to obtain an appropriate subject, which is high cost. In order to deal with these problems, this study proposes a method to complement physiological information in addition to questionnaires when evaluating emotional response to video contents. Specifically, it is a combination of subjective self-report questionnaire and heart rate variability. It is the viewer of short television commercials and news programs that are watched via the VR headset platform. Analysis was done using support vector machine and random forest. As a result, effective models and analysis results were obtained.

1. はじめに

伝統的に、テレビコマーシャル(TVCM)やその他の消費者向けの映像コンテンツの評価はアンケートを用いて実施されてきた。しかしながら、近年の研究では映像コンテンツの評価に生体情報を利用する試みが広がっている。従来のアンケートを用いる手法では、複雑な感情や考えを言葉のみを用いて表現しなければならず、その回答は言葉による短いフィードバックに限られ、必ずしも本当の感情を取得できているとは限らない。言葉のみで正しい感情を必ずしも取得できない理由として、人間は社会的な動物であるので、実験者の期待に応じようとしてしまい、回答に影響を与えてしまうことがある。人間の意志決定は言葉で表現できるほど論理的ではない。それゆえ、人間の感情と考えを言葉のみで表現するのは難しく、その正確性は限られる。これらの理由で、生体反応に基づいて感情や考えを測定する取り組みが進んでいる [5]。しかしながら、生体情報(ECG)を用いて感情を予測する研究は多くなされているが、コンテンツそのものを評価するような高次の感情を予測する研究はほとんどない。また ECG は信頼性の高い生体情報であるが、その反応性はそう良くないと考えられており、日本のテレビコマーシャル(TVCM)にあたる 15 秒間といった動画にはほとんど用いられない。さらに、多くの研究では、実験室のような管理された空間で高価な器具を用いて実験を実施するため、アンケートよりも費用や労力を必要とし、実社会での企業等の利用は困難である。

本研究では、心電図(ECG)からニュース動画の視聴者が内容をわかりやすく感じているのかどうかのモデルと、TVCM の視聴者がその動画を好きかどうかを診断する 2 つのモデルを構築した。その結果 nn50 という特徴量に着目することで、ニュースのわかりやすさという高次の感情を予測できることが示された。それに加え、15 秒間という短い動画でも信頼性の高い生体情報である ECG から感情を予測できることも示された。実験には VR ヘッドセットと安価な器具を使用し、コンテンツへの没入感を高めることで、高い学習精度を実現するデータを取得した。

2. 関連研究

感情を取得する方法の一つとして心電図(ECG)を用いたものがある[2]。これらの感情は自律神経によって引き起こされる変化と強い相関があり、その変化を拡張することで、人間の感情にまで広げることができる[4]。そのため、脳波計(EEG)と皮膚コンダクタンス反応(GSR)も一般的に感情分析に用いられる。

Ramzan et al.[3]は心電図(ECG)と脳波(EEG)を用いて映像コンテンツを視聴した際の感情を予測した。彼らの研究では感情価と覚醒度を教師データとして機械学習モデルを構築している。多くの先行研究と同じように、感情を引き起こす刺激として映像コンテンツを用いており、使用される動画はある感情を引き起こすのに適したものとなっている。そのため、一般的な映像コンテンツより感情を引き出すことが容易であり、生体情報にも反映されやすい。TVCM と視聴者の感情との関連を分析したものに Granero et al.[1]の研究がある。この研究では ACE スコアを参考指標として使用している。ポジティブ、ネガティブおよびそのどちらでもないという3つの感情を割り当て、これを使用して視聴者がどのように感じたのかを識別した。彼らの研究では EEG、ECG そして GSR を用いて、感情を予測するのに最も良い特徴量とアルゴリズムを発見することを目的としている。その結果、GSR と ECG から最も良い特徴量が得られることがわかった。この研究ではアメリカの TVCM を用いているため、30 秒や1分間の動画によるデータとなっており、日本で一般的に使用される 15 秒間の TVCM の動画では、ECG と GSR の反応性の低さから正確な予測に有効な特徴量とならない可能性がある。またその動画の長さの違いから日本の TVCM では物語やメッセージ性を持たせられず、印象に残りやすいよう考えられた動画構成となっているのに対し、アメリカの TVCM では物語性を持たせるなど文化的な違いもある。

3. 目的

本研究では、アンケートデータと生体情報を用いて映像コンテンツの評価を実施した。そうすることで、映像コンテンツのより良い評価と調査の利便性向上を目指す。従来のアンケートデータのみを用いた手法では、上記の制約に加え、被験者の質

連絡先: 浦地勇人 大阪大学大学院情報科学研究科
(surname)@ai.sanken.osaka-u.ac.jp

を確保しなければならないという問題もある。質の高い被験者とは、実験に協力的で精力的に取り組み、実験者に極力配慮しない等の要素を持った人物である。評価実験のための被験者を集める際、こうした点に注意した上で被験者を集めるには多大なコストがかかってしまう。またデータを集めるためには、実験後等に回答を得る時間が必要となり、これも高いコストが必要となる。よって本研究では、質の高い被験者から得られたアンケートデータを用い、生体情報からモデルを構築し、映像コンテンツを評価する。そうすることで、評価の際にアンケートを実施することなく映像評価することができるようになり、被験者の質に関してもアンケートのみを用いるより注意が必要ではなくなる。それに加え、有用な特徴量を発見し、アンケートデータを用いずにモデルを構築する際の道標とすることが可能である。また生体情報による評価をするにあたり、コンテンツへの没入がより必要になる。生体情報には当然ながら評価対象のコンテンツ以外の刺激からの情報も含まれてしまい、コンテンツ以外の刺激を極力減らす必要がある。これらを踏まえ、本研究では、VR 環境下で実験を実施した。

4. 方法概要

本研究では、ニュース番組と 15 秒間の日本のテレビコマーシャルの2種類の映像コンテンツを分析した。

ニュース番組では、様々な種類のテロップを付与するかどうかで、内容のわかりやすさに影響があるのかを分析した。感情認識において最も信頼できる生体情報の一つであるので、ここでは ECG を使用する。同じニュース番組にテロップを付与した3種類の動画を比較した。1つ目はテロップをつけないもの(N)、2つ目は装飾のあるテロップ(A)、そして3つ目はシンプルなテロップ(B)のものである。全部で 24 名の被験者に対して実験を実施した。2分 15 秒間の動画を視聴したのち、複数の解答項目とともにその動画の内容がどのように理解しやすかったのかを取得した。これらの質問から、5段階のニュース動画のわかりやすさを取得した。機械学習モデルを構築し、これらのデータから学習とテストを実施し、分析に有用な特徴量が判明した。またそこからコンテンツのわかりやすさとの関連を分析した。

2つ目のデータセットはテレビコマーシャル(TVCM)から取得され、3 種類の 15 秒間の短い動画を比較した。その際、テレビコマーシャルを好き、嫌いまたそのどちらでもないのかを5段階で取得した。こちらのデータセットでも短い動画のデータにも ECG のデータが有用か調査するため ECG を使用した。

5. VR ON AIR TEST

本研究では、SOOTH 社によって開発された VR OAT(VR ON AIR TEST) [6]を用いてデータを取得した。このシステムでは、VR 環境下における、動画視聴者の動作を取得できる。タッチ式のデバイスを用いて、アンケートに回答し、複数の生体センサーがそのプラットフォームに内蔵されている。VR 環境はユーザーがコンテンツに没入できるといふ利点がある。さらに、既存のプラットフォームと比較して、使用しているツールが安価かつ着脱も容易である。このように、実社会で企業等が実施する映像コンテンツの評価により適していると期待される。

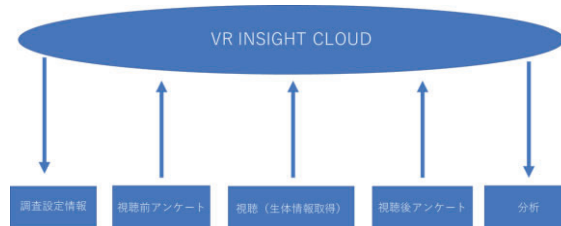


図1: VR ON AIR TEST

このように、タブレット端末から属性やアンケートな回答を取得でき、VR 上で映像コンテンツの視聴と EEG をはじめとする様々な生体情報を取得できる。そしてこれらのデータをクラウド上で管理することもできる。

6. 実験手続き

以下の項では2つの動画視聴の実験をどのように実施したかの詳細を述べる

6.1 ニュース動画

20代の男女24名の被験者が2分 15秒間のニュース番組を視聴した。アンケートは動画の視聴前と視聴後にそれぞれの動画に対して実施した。その間、ECG および EEG のデータを取得した。3種類のテロップを付与した動画を使用した。1つ目はテロップをつけないもの(N)、2つ目は装飾のあるテロップ(A)、そして3つ目はシンプルなテロップ(B)のものである。全ての被験者24名が動画 N を視聴し、その後24名を二つのグループに分割し、それぞれ A か B どちらかの動画を視聴した。全てのテロップは動画の内容について述べられたものである。

6.2 テレビコマーシャル

テレビコマーシャルの動画を20代から50代の男女 213名の被験者が視聴した。ニュース動画のようにセンサーを用いて生体情報を取得した。テレビコマーシャルは 15 秒間の動画 3 本で、同一商品について競合する 3 社によるものを用いた。アンケートも同様に、視聴前と視聴後に実施した。

7. 特徴量

7.1 ニュース動画

最初の実験のアンケートから、ニュース動画のわかりやすさを述べるため5段階の評価を得た。

- 1:非常にわかりにくい
- 2:わかりにくい
- 3:どちらでもない
- 4:わかりやすい
- 5:非常にわかりやすい

テロップなしの動画から5段階すべての評価の反応を得た。

しかしながら、テロップなしの動画からの評価はすべての評価が4以上であった。このことからテロップはニュース動画のわかりやすさを改善するのに効果的であったと考えられるが、追加の検証が必要である。それゆえ、アンケートの反応が生体反応と一致するかどうか分析した。段階1と2および段階 4 と 5 を区別することは困難であるので、1と2を「わかりにくい」、4 と 5 を「わかりやすい」に統合した。

7.2 テレビコマーシャル

TVCM から、11のシーンとそれぞれのシーンに番号を付与した。それぞれのシーンと時間は表 1 に示した。また動画の好き、嫌いに関しても同様に以下の 5 段階の評価を取得した。

- 1: 非常に嫌い
- 2: 嫌い
- 3: どちらでもない
- 4: 好き
- 5: 非常に好き

これらのアンケートに対してすべての動画において、嫌いと回答したものは非常に少なかった。そのため、「どちらでもない」及び好きの 2 段階に統合した。

7.3 ECG データ

心電図は、心臓内で発生する電気的な変化によって引き起こされる。心臓が収縮し、体表からの電位として検出されたとき、電気信号のピークは P、Q、R、S、T のように名付けられる。R は最高点を表し、R と R の間隔 (RRI) は常に一定ではなく変動する。この変動は感情との強い結びつきを持つ心理的影響も反映している。以下の図 2 にこれを示す。

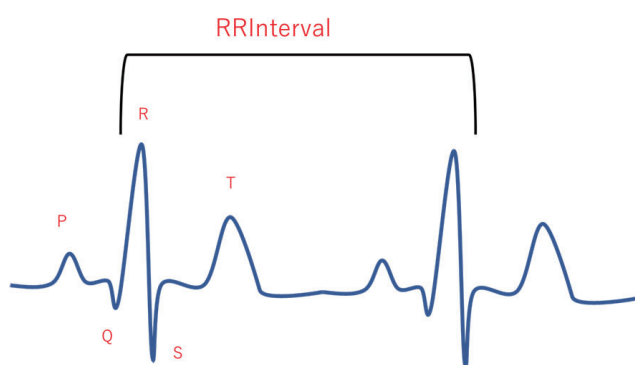


図2: 心電図と RRI

ここでは、Neurosky[7]を使用して ECG データを収集し、RRI から得られる以下の特徴量を抽出した。

- RMSSD - 連続する RR 間隔の差の二乗平均平方根
- SDNN - RR 間隔の標準偏差
- NN50 - 連続する隣接 RR 間隔の差が 50 ミリ秒を超えた数の割合
- PNN50 - 連続する隣接 RR 間隔の差が 50 ミリ秒を超えたものの割合
- HF - 高周波帯域のパワー (0.15~0.4 Hz)

8. モデル

アンケートからの教師データと生体情報から機械学習モデルを構築した。使用したアルゴリズムはサポートベクターマシン (SVM) とランダムフォレスト (RF) である。実装には python の機械学習ライブラリーの scikit-learn を使用した。

8.1 ニュース動画モデル

学習は ECG から得られる様々な特徴量を用いて実施した。機械学習のモデルを用いて、データの構造を分析するため ECG のデータは 10 秒間隔毎に分割し、それぞれから特徴量を

算出した。24 名の被験者のうち、欠損等のため信頼性の高いデータが得られたのは 15 名のみであった。それゆえ 15 名のデータを用い、4 分割の交差検証を実施した。教師データにはアンケートから得られたニュース動画のわかりやすさを使用し、SVM にてこれを予測した。その際、カーネル及びハイパーパラメータはグリッドサーチにて最適化を行った。

またそれに加え、RF を用いて特徴量の重要度も算出した。モデルのパラメータは同様にグリッドサーチにて最適化され、15 名すべてのデータを使用した。

8.2 テレビコマーシャルモデル

学習はニュース動画モデルと同様に ECG から得られる特徴量を用いて実施した。TVCM でも ECG のデータは前半 10 秒間と後半 5 秒間に分割され、値を算出した。欠損等のため信頼性の高いデータが得られたのは 20 名 (好き10名、どちらでもない10名) のみであった。それゆえ 20 名のデータを用い、4 分割の交差検証を実施した。教師データにはアンケートから得られた TVCM の好みを使用し、SVM にてモデルを構築した。TVCM のモデルでも、カーネル及びハイパーパラメータはグリッドサーチにて最適化を行った。

9. 結果

9.1 ニュース動画の結果

SVM によってニュース動画 (N) のわかりやすさを識別した結果を表 1 に示す。使用したカーネルは rbf カーネル、線形カーネル (linear)、多項式カーネル (poly) で表の model に示す。特徴量 (feature) は 7.3 のものを用いた。

表 1: 動画 N によるデータセットの予測結果

model	feature	training	test	recall	f-value
rbf	nn50	100	72.9	66.7	59.4
poly	rmssd	84.3	52.1	45.8	42.2
rbf	hf	46.6	45.8	33.3	20.8
linear	sdnn	97.7	33.3	33.3	23.6
rbf	pnn50	100	45.8	45.8	36.1

これらの結果から、最も有効な特徴量は nn50 であった。

次に、動画 N のデータから学習した 4 つの交差検証のモデルを使用して、テロップを付与した動画 (A,B) の値を予測した。表 1 から nn50 が最も高い精度を示したため、ここでは nn50 を使用する。注意すべき点として、テロップを付与した動画からは「わかりやすい」の単一の教師データしか得られておらず、トレーニングセットには使用できない。このデータセットは合計 20 (A が 11、B が 9) 個からなる。識別結果は以下の表 2 に示す。

表 2: テロップあり動画の予測結果

data	model	feature	test
A	rbf	nn50	100
B	rbf	nn50	100

予測結果はアンケートの回答と完全に一致した。このことから、テロップを付与することで被験者はニュース動画をわかりやすく感じるようになったということが生体情報からも検証された。

続いて動画 N を学習データ、アンケートを教師データとしてランダムフォレストにて学習器を構築した。この学習器から特徴量の重要度を得た。これを時間順にならべ、動画と比較した。その結果を以下の図 3 に示す。

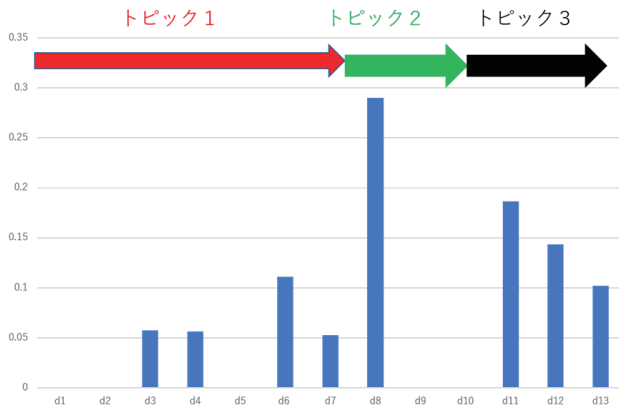


図3: 時間毎の特徴量の重要度とニュース動画の比較

横軸は動画の時間を表しており、縦軸は重要度の割合を表している。d1は動画開始時(0秒)からの10秒間を表しており、d2, d3...d13はそれぞれその後10秒間毎の値を表している。ここでは特徴量に nn50 を使用している。

この図から重要度の高い時間に注目し、動画と対応づけるとあるトピックの終盤に集中していることがわかる。このニュース動画は3つのトピックからなっており、重要度が高い時間帯はそれぞれトピックの終盤と一致している。このことから生体情報にもトピックの終盤にコンテンツのわかりやすさに対する反応が現れることがわかる。

9.2 テレビコマーシャルの結果

SVMによってテレビコマーシャルを好きか、そうでないかを識別した結果を表3に示す。

表3: テレビコマーシャルの予測結果

model	feature	training	test
linear	nn50	74.8	70.8
linear	sdnn	74.6	66.7

この結果から、日本のテレビコマーシャルのように短い動画であっても、ECGから感情を予測できることがわかる。また、こちらでも nn50 が最も有効な特徴量であった。

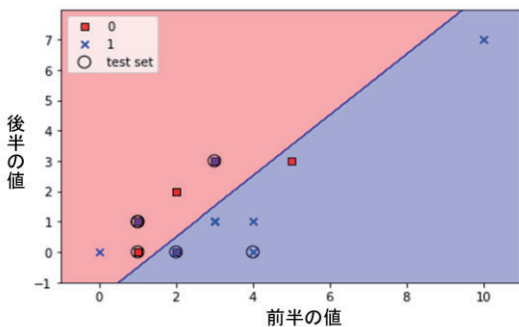


図4: TVCMモデルの nn50 決定境界

図4に nn50 を用いて得られた SVM の決定境界を示す。0が「どちらでもない」、1が「好き」を表す。このように、前半に比べ後半の nn50 値が減少すれば、好きと回答する傾向が見受けられた。

9.3 まとめ

ECG から得られる特徴量に基づき、ニュース動画のわかりやすさを識別する学習器を得た。テロップを使用した場合、全ての被験者がニュース動画を少なくともわかりやすかったと評価した。この結果が本当にそうであるのか、実験効果によるものかを検証するため、テロップ無しの動画から得られたデータセットでモデルを学習させた。その結果、検出されたニュースのわかりやすさとアンケート結果は完全に一致した。そのようにして、本研究ではアンケートの回答の信頼性をどのように検証するのかわを示した。またテロップの存在がニュース番組の理解において有効であることも判明した。さらに、ECG の特徴量の中でニュース番組のわかりやすさと最も強い相関を持つものが nn50 であることがわかった。また特徴量の重要度を算出し、その結果ニュース番組におけるそれぞれのトピックの終盤の重要度が高いことも判明した。

次に、同様に ECG を用いて TVCM から感情を推定するモデルを得た。この感情は TVCM が好きなのか、そうでないのかの2種類を使用した。その結果日本の TVCM のように短い動画でも ECG を用いて感情を推定できることが示された。

10. おわりに

本研究では、ECG から動画の視聴者の感情を診断するモデルを構築した。構築したモデルはニュース動画の内容をわかりやすく感じているのかを診断するモデルと、TVCM を好きかどうかのモデルである。その結果、ECG から取得できる nn50 という特徴量を用いることで動画の評価に関連する高次な感情を取得できることが示された。また TVCM のような短い動画でも nn50 を用いることで感情を予測できることが示された。これらは VR ヘッドセットと安価な器具を用いて実施しており、VR 環境によって使用者の没入感を高め、高精度の感情予測に使用可能なデータを取得できることも示された。

より信頼性の高い評価を得るため、これらの自動化された映像評価を従来のアンケート結果と組み合わせることが可能である。この自動化された映像評価の精度がさらに向上すれば、アンケートを使用しない評価も可能になると考えている。

参考文献

- [1] Granero, A. C.; Fuentes-Hurtado, F.; Ornedo, V. N.; Provinciale, J. G.; Ausin J. M. and Raya, M. A. Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents, Front. Comput. Neurosci., 15 July 2016
- [2] Shaffer, F. and Ginsberg, J. P. An Overview of Heart Rate Variability Metrics and Norms, Front. Public Health, 28 September 2017
- [3] Ramzan, N.; Palke, S.; Cuntz, T.; Gibson, R. and Amira, A. Emotion Recognition by Physiological Signals, 2016 society for Imaging Science and Technology
- [4] Quintana, D. S.; Guastella, A. J.; Outhred, T.; Hickie, I. B.; Kemp, A. H. Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition, International Journal of Psychophysiology 86 (2012) 168–172
- [5] 広志熊倉. ニューロマーケティングの現状、課題そして展望(特集ニューロマーケティング). オペレーションズ・リサーチ, Vol. 61, No. 7, pp. 421–428, jul2016
- [6] <https://www.vr-insight.com/en/oat.html>
- [7] <http://neurosky.com/>