

# 笑顔に着目した動画の評価取得手法の検討

田村 俊\* 大久保 雅史\*

## Study of Evaluation Method for Video Contents Focusing on Smile

Shun Tamura\* and Masashi Okubo\*

**Abstract** - Recently, video sharing services and video on demand services like YouTube, Netflix and Hulu have come to prominence worldwide. These services try to get the user's evaluation of videos for the better recommendation. For example, some of them prepare the good/bad buttons and star buttons in their services to get user's evaluation. We have some doubts an effect of that method. The reason why these method needs user's conscious operation. In this study, we focus on facial expression especially smile while watching videos. In this paper, we have developed the system that can analyze the user's facial expression, and we have investigated the relation between evaluation of videos and facial expression by using the proposed system. As a result of experiment, we found strong correlation between the ranking by the user's scoring and the ranking based on the amount of smile while watching videos in many participants. This result suggest possibility of recommendation method based on facial expression.

**Keywords:** facial expression, recommendation method, video on demand, usability

### 1. 緒論

近年、YouTube などの動画投稿サイトや Netflix, Hulu といったビデオ・オン・デマンド・サービスが世界的に人気を博している。これらのサービスではユーザが膨大な動画の中から観たい動画を探す手段として、キーワードによる検索が行われる。検索は主に動画のタイトルや動画説明の文章、動画に付与されたタグなどのテキスト情報に基づいて行われるが、これらを投稿者が自由に設定できる投稿サイト等では、動画の内容とキーワードが一致しておらず、ユーザが観たい動画を探すことを阻害している。また、検索によってフィルタリングされた結果も膨大な数の動画を含み、ユーザが興味を示す可能性のあるコンテンツにたどり着けないことがある。

これらの課題に対する解決策として、ユーザの興味に即した動画を推薦する方法が重要な役割を担っている。アメリカの大手ビデオ・オン・デマンド・サービス Netflix は、同サービスにて配信している動画の再生時間の 8 割が推薦によってもたらされていると発表している<sup>[1]</sup>。

推薦の手法としては、ユーザに動画を評価するためのボタンを押させるものや、動画タイトルやウェブサイト上でのユーザの行動履歴などの付加情報から動画への興味を推定するものなどが挙げられる。しかし、これらのユーザの能動的操作に依存する方法は多くのユーザにとって煩雑であり、能動的に評価を付けるユーザは少ないと考えられる。また動画の付加情報を用いる方法はユーザが動画内のどの部分に興味を示したか考慮していない。

そこで本研究では動画視聴時のユーザの表情に注目する。表情は人の感情表現の中で重要な役割を占めている

とされる<sup>[2]</sup>。そのため、動画視聴時の表情は動画への興味の度合いや集中度に関係していると考えられ、動画を評価する指標となる可能性がある。また、動画中の同じシーンで同様の表情を示すユーザ同士は動画への評価や嗜好が類似すると考えられる。

### 2. 目的

#### 2.1 関連研究

動画視聴時のユーザの反応から興味を推定する手法として、菅沼らは眼球情報（注視点、サッケード、瞳孔径）や脳波（背景脳波）を用いるものを提案している<sup>[3]</sup>。

また、夏目らは動画視聴時の視線の動きに注目し、ユーザ間の視線移動軌跡の類似性に基づいて動画を推薦する手法を提案している<sup>[4]</sup>。

しかし、いずれの手法も脳波計測器や眼球運動計測器といった特別な機器を必要とする。またユーザはそれらの機器を身に着ける必要があり、少なからず負担を強いられ、より自然な状態での動画視聴が可能なシステムが求められる。

#### 2.2 表情を用いた動画推薦システム

本研究では一般的な動画視聴シーンへの応用を想定しており、特別な機器を利用するのは好ましくない。これに対し、我々は一般的なノート PC に付属しているウェブカメラを利用する。ユーザの顔をウェブカメラで撮影し、表情から動画への評価を推定し、図 1 に示すユーザの嗜好に合わせた動画を提案する推薦システムの構築を目指す。

本稿では、まず動画視聴時の表情を取得するシステムを構築している。さらにこのシステムを用いて評価実験を行い、動画視聴時の表情を用いた動画評価手法の検討を行っている。

\*: 同志社大学大学院 理工学研究科

\*: Graduate School of Science and Engineering, Doshsha University

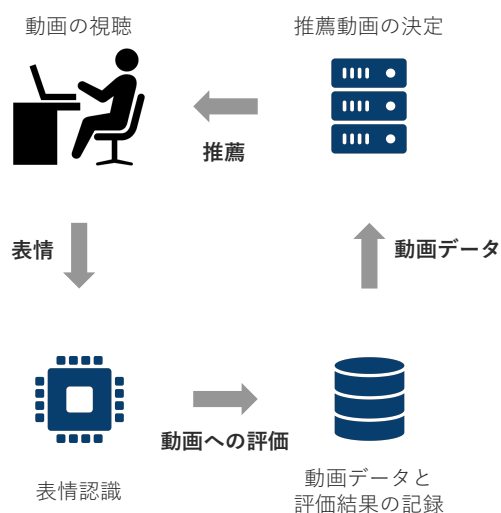


図1 提案システムの概要  
Fig.1 Outline of proposed system

### 3. 実験1：動画視聴中の表情と評価の関係性

#### 3.1 実験目的

動画視聴中のユーザの表情と動画への評価との関係性を検証するため、同志社大学の8名に対して実験を行う。

#### 3.2 実験システム

ユーザに負担をかけずに動画視聴時の表情を取得することを目的とし、ウェブカメラによる顔画像の撮影、および表情認識の結果を記録するシステムを開発した。システムの概要を図2で示す。

システムでは実験者が用意した動画を視聴できる。ユーザが動画の再生ボタンを押すと同時にウェブカメラの撮影が始まる。ウェブカメラは動画が終了するまで3秒毎に撮影を繰り返す。撮影した画像は動画の再生が終了した後、後述する外部のWeb APIへ送信する。APIから表情認識の結果を受信後、PC内部に記録する。

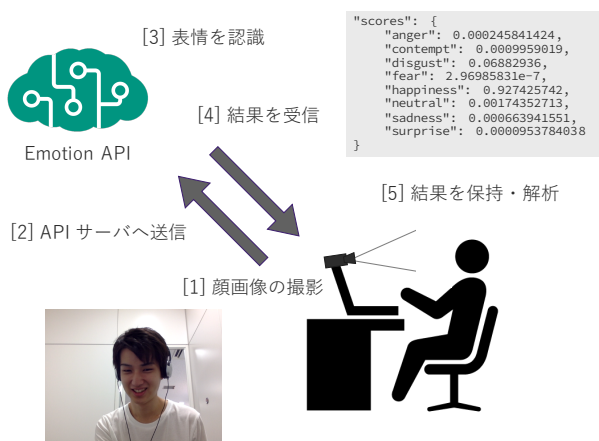


図2 実験1システムの概要<sup>1</sup>  
Fig.2 Outline of 1st experimental system

### 3.3 Emotion API

表情認識には Microsoft 社が公開している Emotion API を利用する<sup>[5]</sup>。Emotion API はリクエストに含まれる画像の中から人の顔貌を検出し、表情の認識と感情の推定を行う。推定される感情には anger (怒り), contempt (軽蔑), disgust (嫌悪), fear (恐怖), happiness (幸せ), sadness (悲しみ), surprise (驚き), neutral (いずれでもない) の8つがある。本実験ではユーザに動画を視聴させた後、取得した顔画像を Emotion API へ送信し、分析結果を得る。

### 3.4 実験方法

実験協力者に個室に入らせ、表1に示す内容の異なる4本の動画を、実験システムを用いて視聴させる。各実験協力者には動画の視聴を終了する度に動画への興味度などを聞くアンケートに回答させる。

表1 実験1に用いた動画

Table.1 Video contents used in 1st experiment

動画	内容
A	自然風景
B	フラッシュモブ
C	映画の別れのシーン集
D	マジック

各動画の視聴を終える毎に動画を観て湧いた感情や動画への興味度をアンケートで回答させる。5つの質問項目を表2に示す。

表2 動画視聴後アンケートの質問項目

Table.2 Questionnaires after watching each video

質問番号	内容	回答方式
Q1	この動画を既に見たことがあったか	ある or ない
Q2	この動画を観てどんな感情が湧いたか	怒り、軽蔑、嫌悪、喜び、悲しみ、驚き、その他から複数選択
Q3	この動画は興味があるものだったか	5段階評価
Q4	Q3に関して、それは何故か	記述式
Q5	同様の動画をもう一度観たいか	はい or いいえ

### 3.5 実験結果

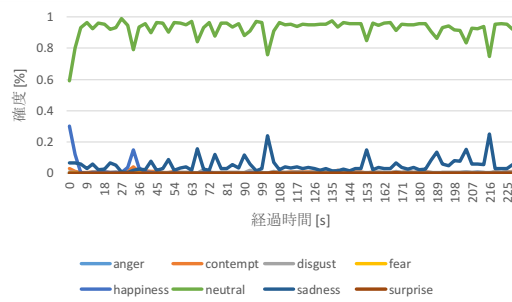
例として実験協力者1のアンケートの回答を表3に、表情の遷移を図3に示す。

表3 実験協力者1のアンケートの回答 (抜粋)

Table.3 Answers of participant 1 (extracted)

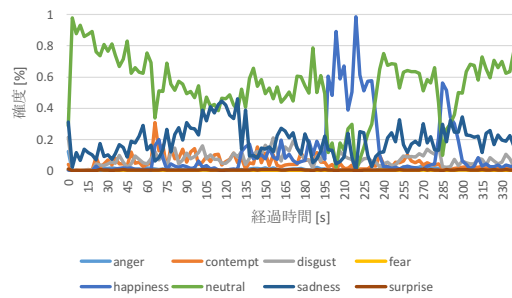
質問番号	動画A	動画B	動画C	動画D
Q1	ない	ない	ない	ない
Q2	退屈	幸せ	悲しみ	幸せ、驚き
Q3	1	4	2	5
Q4	いいえ	はい	いいえ	はい

1: 写真は実験協力者の承諾を得て掲載



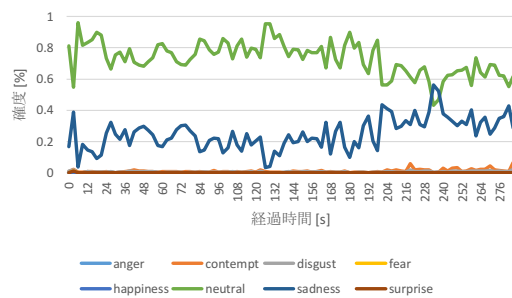
(a) 動画 A

(a) Video A



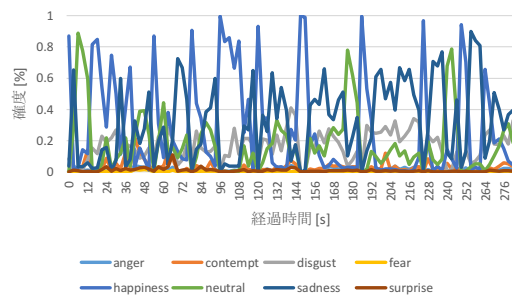
(b) 動画 B

(b) Video B



(c) 動画 C

(c) Video C



(d) 動画 D

(d) Video D

図3 実験協力者1の表情の遷移

Fig.3 Transition of estimated facial expression of participant 1

表3と図3にて結果を示した実験協力者1においては、動画への興味度と表情に関係性が見られた。1は動画Aに対して退屈な感情を抱き、興味度は1と回答している。またシステムを通して認識された動画視聴中の表情はneutralを維持している。動画Bには幸せの感情を持ち、

興味度は4と高い。表情の遷移では200秒付近、225秒付近にてhappinessの確度が高まっている。動画Cには悲しみを感じ、興味度は2と評価している。表情では一定の確度で悲しみが見られる。動画Dに対しては幸せと驚きを感じ、最も高い興味度5を付けており、表情の遷移はhappinessとsadnessが交互に現れる形となっている。全体としては興味度の高かった順に表情変化が多くみられる結果となった。

以上より、表情が大きく変化する動画に対しては表情の表出から動画への評価を推定し得る可能性が示された。また、実験協力者1は幸せを感じる動画を好み、悲しみをを感じる動画を好まないという推測が可能であり、好みを表情から推定し、ユーザが好む動画を推薦し得る可能性が示唆される。

他の4名については、4つの動画を通して表情と動画への評価の関係性が現れる実験協力者はなかったものの、一部の動画について実験協力者1が示したような関係性が見られた。

しかしながら、他の3名の実験協力者に関してはこのような表情の変化と動画への評価の関係性を読み取ることができなかった。例として実験協力者2の動画Bにおける表情の遷移を図4に示す。実験協力者2は他の動画A, C, Dにおいても動画視聴中に全く表情が変化せず、同様の結果となった。

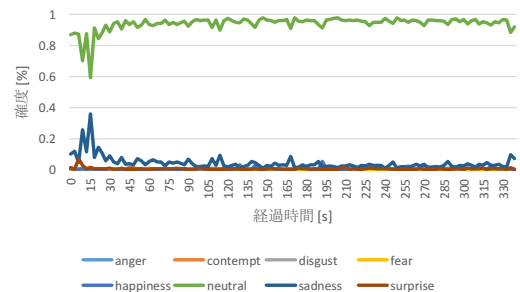


図4 実験協力者2の表情の遷移(動画B)

Fig.4 Transition of estimated facial expression of participant 2 (Video B)

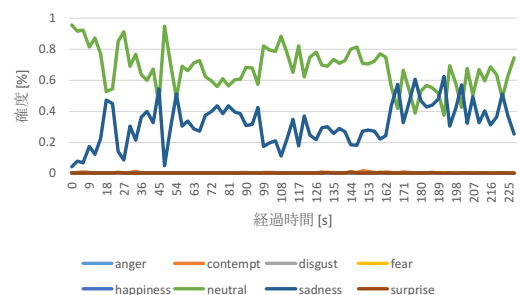


図5 実験協力者3の表情の遷移(動画A)

Fig.5 Transition of estimated facial expression of participant 3(Video A)

また、実験協力者 3 は Emotion API によって自然な表情と笑顔が **sadness** と判定され、4 つの動画全てで図 5 のような **sadness** の割合が多い結果となった。

加えて、実験協力者 4 は Emotion API によって表情を認識することができなかった。

### 3.5.1 ユーザ間の表情表出の類似性

また、実験協力者同士の表情表出の類似性について述べる。図 6 に実験協力者 1 と 5 の動画 B における表情遷移を示す

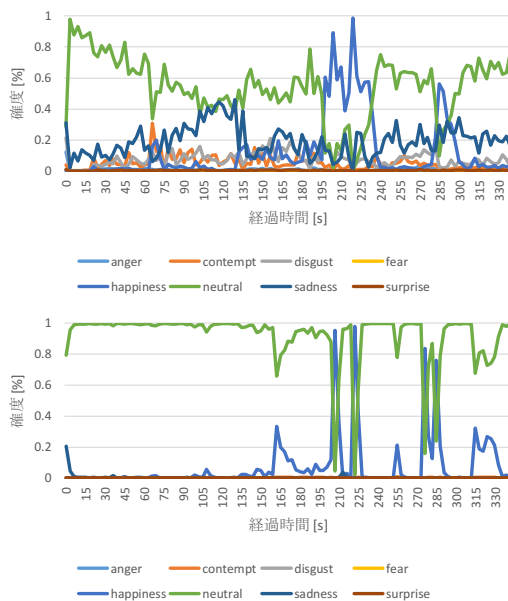


図 6 実験協力者 1 (上) と 5 (下) の表情表出の比較 (動画 B)

Fig.6 Transitions of estimated facial expression of participant 1 (upper) and 5 (lower) (Video B)

図 6 の 210 秒付近と 225 秒付近、285 秒付近の 3 箇所で実験協力者 1、5 ともに **happiness** の確度が高まっている。また、動画視聴後のアンケートではどちらも興味度を 4 としている。動画中の同様の箇所で同じ表情を示し、また同様に高評価を付けている点から、実験協力者 1 と 5 は動画の嗜好も類似している可能性が示唆される。

### 3.6 考察

実験結果を踏まえ、動画コンテンツの推薦システムへの応用を考察する。例えば表情が表出している時間を積分し、その多寡を動画への評価とするなど、動画への評価を表情から算出することが可能となれば、既存の推薦システムでよく用いられる協調フィルタリングに利用できる。また、ユーザの表情が変化した動画シーンの解析が可能であれば、同様のシーンを含む動画を推薦するなど、コンテンツの内容に沿った推薦が可能となる。

## 4. 実験 2：動画視聴中の笑顔と評価の関係性

### 4.1 目的

実験 1 により動画視聴中は特に笑顔が表出・認識しやすいこと、また動画への評価と表情の変化量に関係がある可能性が示された。そこで対象とする表情を笑顔に限定し、コメディ動画を視聴するユーザの笑顔と動画への評価を検証するため、同志社大学の実験協力者 8 名に対して実験を行う。

### 4.2 実験システム

実験 1 と同様のシステムで実験を行ったが、推薦システムへの応用を想定し、よりリアルタイムな分析を目指すべく、表情認識に Affectiva 社が開発する感情認識 SDK である **Affdex** を利用する。**Affdex** は一般的なウェブカメラ等で撮影された顔画像から特徴点を抽出・分析し、表情 (アクションユニット) の分類を行い、この表情の組み合わせから感情を推定する<sup>[6]</sup>。図 7 に実験 2 システムの概要を示す。



図 7 実験 2 システムの概要

Fig.7 Outline of 2nd experimental system

**Affdex** は外部の WebAPI にリクエストを送ることなく顔画像の分析が可能である。動画視聴と同時に **Affdex** による顔画像の分析を開始、終了時に表情分析結果を保存するシステムを開発し、実験を行う。

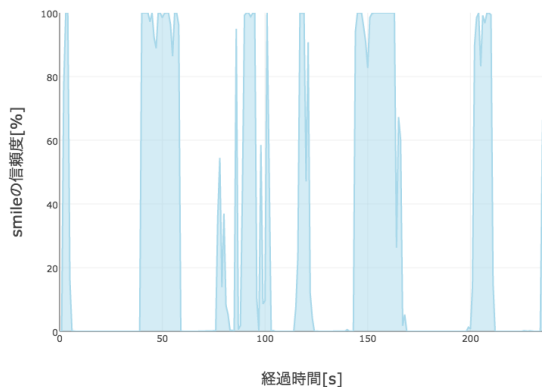
### 4.3 実験方法

実験協力者には用意した 1 本 4 分程度の漫才ビデオを、前述の実験 2 システムを通して 9 本視聴させる。動画の視聴を終了する度に 100 点満点で採点させ、特に面白かったシーンを自由記述、複数可で回答させる。動画は実験協力者毎にランダムな順番を指定するが、全被験者で共通の採点基準を作るためにいずれの実験協力者も動画 A から視聴させる。

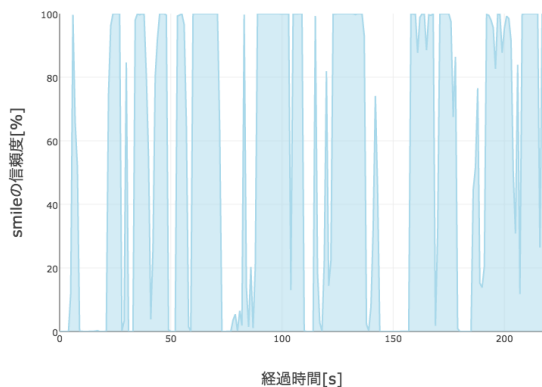
## 4.4 結果

### 4.4.1 評価方法

まず実験 2 システムを通して得られた表情分析結果から, smile 項目の信頼度を示す時系列データを抽出する. システムを用いて動画を視聴している間, ウェブカメラのフレームレートが一定でないため, 時系列データには実験協力者毎, 視聴動画毎にデータ数のばらつきが見られる. そのため, 時系列データを 1 次スプライン補間し, 得られる補間関数から 1 秒毎にサンプリングし標本を得る. こうして得られる smile 項目の信頼度の時間遷移を図 8 に示す.



(a) 動画 A  
(a) Video A



(b) 動画 D  
(b) Video D

図 8 実験協力者 3 の smile の遷移

Fig.8 Transitions of estimated smile expression of participant 3

次に, 得られた標本の平均を笑顔の量と定義し, その大小を比較して各実験協力者の動画への評価順位を算出する. 各実験協力者の各動画に対する笑顔の量を表 4 に, 笑顔の量から決定される評価順位を表 5 に示す. なお, 実験協力者 5 は動画 D を視聴した際, システムを通して分析結果を得ることができなかったため, 以下 NaN(Not a Number)とし, 実験協力者 5 については動画 D を評価対象から除外している.

表 4 各実験協力者の各動画に対する笑顔の量

Table.4 Average of smile each participant to each video

実験協力者	A	B	C	D	E	F	G	H	I
1	40.31	66.25	50.83	69.21	53.95	54.64	57.02	0.10	52.89
2	0.07	15.81	2.59	18.75	0.79	13.28	0.00	3.56	0.03
3	29.31	30.57	27.12	56.54	42.85	40.53	22.43	16.39	8.93
4	16.73	39.52	20.08	17.88	32.97	23.64	21.75	19.94	12.45
5	0.001	6.273	0.780	NaN	0.409	4.033	3.634	0.031	9.756
6	19.00	36.82	5.29	47.65	7.46	36.20	13.77	9.74	25.40
7	56.98	55.33	44.14	68.01	37.39	35.89	14.77	19.91	21.45
8	84.88	91.29	73.71	89.67	60.89	90.51	89.25	69.52	86.12

※アルファベットは動画を表す

表 5 各実験協力者の各動画に対する  
Smile の平均による評価順位

Fig.5 Evaluated ranking for each video based on  
amount of each participant's smile

実験協力者	A	B	C	D	E	F	G	H	I
1	8	2	7	1	5	4	3	9	6
2	7	2	5	1	6	3	9	4	8
3	5	4	6	1	2	3	7	8	9
4	8	1	5	7	2	3	4	6	9
5	8	2	5	NaN	6	3	4	7	1
6	5	2	9	1	8	3	6	7	4
7	2	3	4	1	5	6	9	8	7
8	6	1	7	3	9	2	4	8	5

※アルファベットは動画を表す

またアンケートにて採点させた結果についても実験協力者毎に各動画に対する採点結果の大小を比較し, 表 6 に示す評価順位を得る.

表 6 各実験協力者の各動画に対する  
採点による評価順位

Table.6 Evaluated ranking for each video  
based on participant smile

実験協力者	A	B	C	D	E	F	G	H	I
1	8	1	3	2	7	5	4	9	5
2	7	4	4	2	1	2	7	7	6
3	6	4	8	1	3	2	9	6	5
4	8	2	8	3	1	5	3	5	5
5	8	2	6	NaN	6	2	4	5	1
6	6	3	8	1	9	2	5	7	4
7	3	1	7	2	4	4	9	8	6
8	7	3	9	1	8	3	6	3	1

※アルファベットは動画を表す

このようにして得られた 2 つの評価順位データから, 動画視聴中の表情から推定した評価と実験協力者の採点評価との関係性を確認するため, スピアマンの順位相関係数を利用する. スピアマンの順位相関係数は順位に変



換した 2 変数から求められる相関の指標であり，次式で定義される．算出される相関係数は  $-1 \leq r_s \leq 1$  の値をとり，順位が完全に一致する場合は 1 を，全て逆順となる場合は  $-1$  をとる．

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (1)$$

$n$  : 動画の数,  $d_i$  : 2 つの順位データの差

#### 4.4.2 評価結果

前述したスピアマンの順位相関係数を実験協力者毎に求めた結果を表 7 に示す．実験協力者 8 名中 5 名で相関係数が一般に強いと言われる 0.7 以上となり，強い正の相関が見られた．特に実験協力者 6 は 0.97 と極めて強い正の相関を示している．他の 3 名についても  $0.4 \leq r_s \leq 0.7$  となっており，関係性があることが示された．

表 7 各実験協力者のスピアマンの順位相関係数  
Table.7 Spearman's rank correlation coefficient

実験協力者	順位相関係数
1	0.770
2	0.487
3	0.745
4	0.530
5	0.916
6	0.976
7	0.826
8	0.545

#### 4.5 考察

実験結果から，笑顔を誘発させる内容の動画を対象とした場合に，動画視聴中の笑顔の量を指標として，動画コンテンツを評価し得る可能性が示唆された．この評価は既存の推薦アルゴリズムである協調フィルタリングへの応用が可能であると考えられる．つまりユーザの能動的操作を必要とせずに，動画の嗜好が似ているユーザを発見したり，あるユーザに推薦すべき動画を発見したり，似ているアイテムを発見したりすることが可能となる．

### 5. 結論

本研究では動画視聴時のユーザの表情に注目し，動画視聴時の表情を取得するシステムの開発を行った．さらにこのシステムを用いた評価実験を行い，動画視聴時の表情，特に笑顔と動画への評価の関係について検討した．

実験 1 の結果，表情をよく表出し，かつシステムが認識可能なユーザにおいては動画への評価の順番と動画視

聴中の表情の変化量に関係性が見られた．また，動画を観て湧いた感情と表情を認識して推定される感情が概ね一致していた．

次に実験 2 の結果，特に笑顔を誘発する動画においては，動画視聴中の笑顔の量と動画への評価に強い相関がある可能性が示された．

これらの結果から，表情を動画コンテンツの評価指標として用いることの出来る可能性が示唆される．従来の推薦システムに用いられてきた協調フィルタリング等のアルゴリズムにおいて，ユーザの能動的な評価に換えて本稿で示した手法が有用であると考えられる．今後は表情認識を通して得られる結果を用いた推薦システムを実際に構築し，推薦を行う．

また，本稿では対象とする表情を笑顔のみに限定したが，悲劇やホラーなどさまざまな表情・感情を誘発する動画コンテンツが数多く存在する．これらに対応する指標を検討することも考えられる．

#### 参考文献

- [1] Gomez-Urbe, A. C., Hunt, N.: The Netflix Recommender System: Algorithms, Business Value, and Innovation; Journal ACM Transactions on Management Information Systems (TMIS), Vol. 6, No. 4, Article 13 (2015)
- [2] Ekman, P, Friesen, W. V.: Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions, New Jersey: Prentice-Hall (1975)
- [3] 菅沼, 川村, 亀山 : 生体情報による動画視聴時の興味度推定に関する基礎的検討; 2011 映像情報メディア学会冬季大会講演予稿集, 7-4 (2011)
- [4] 夏目, 入部 桂田, 新田: 動画視聴時の注目点を考慮した動画推薦方法の提案; 情報処理学会全国大会講演論文集, Vol. 70, pp. 857-858 (2008)
- [5] Microsoft Cognitive Services - Emotion API , <https://www.microsoft.com/cognitive-services/en-us/emotion-api>, (2016/10/27 アクセス)
- [6] Senechal, T., McDuff, D., Kaliouby, R.: Facial Action Unit Detection Using Active Learning and an Efficient Non-linear Kernel Approximation; IEEE International Conference on Computer Vision Workshop, pp.10-18 (2015)