Vol.2015-GN-96 No.3
2015/10/2

ソーシャルコメントからの音楽動画印象推定に関する検討

土屋駿貴[†1, †2]　中村聡史[†1,†2]　山本岳洋[†3,†2]

本稿では印象に基づく音楽動画検索実現に向け，音楽動画の再生時間に対するソーシャルコメントを用いて音楽動画を印象分類する手法を検討する．ここでは，500 曲の音楽連動動画のサビ部分について音楽のみ，映像のみ，音楽と映像の組み合わせに対して付与された 8 軸の印象評価を利用し，そのそれぞれのメディアタイプ，印象タイプにおけるコメントを利用した印象推定手法を検討し，その精度について考察を行うとともに，コメントからの印象推定可能性について検討を行う．

A Study on Estimating Impression of Music Video Clip by Using Social Comments

SHUNKI TSUCHIYA[†1, †2]　SATOSHI NAKAMURA[†1,†2]
TAKEHIRO YAMAMOTO[†3,†2]

In this paper, we consider a method to estimate impression of a music video clip by using social comments in order to realize the impression-based music video search. Here, we use the dataset consists of 500 music video clips with evaluation score in each media type and in each impression type. Then, we evaluate the precision of estimating the impression of a music video clip in each media type and impression type by using social comments and discuss the possibility of estimation.

1. はじめに

YouTube やニコニコ動画，piapro に代表される CGM サイトの広がりや，VOCALOID に代表される DTM ソフトウェアの普及により，Web 上の音楽動画の数が飛躍的に増加している．ここで，音楽動画が増加している一方で，ユーザが求める音楽動画を検索する方法は多様であるとは言いがたく，アーティスト名や楽曲名，ユーザが付与したタグなどの情報から検索する方法が主となっている．これらの検索方法では，事前にアーティスト名や楽曲名などの音楽動画に関する情報を知っている必要がありユーザの求める楽曲を探しだすことは困難であるといえる．

こうした問題を解決するために音楽情報検索の分野では，音楽動画からユーザが受ける印象に基づく検索についての研究が進められている．印象とは，「元気が出る」「かわいい」「激しい」などのユーザの主観的な感情のことであり，印象からの検索が可能となれば，ユーザは今までにない新しい視点からの検索が可能となる．しかし，印象に基づく検索を行うためにはそれぞれの音楽動画に対してあらかじめユーザが主観的な印象を評価しておくことが必要となる．先述したように音楽動画は増加を続けているため，すべての音楽動画に対して印象評価を行うことは不可能である．そのために，機械的に音楽動画に対する印象を推定する手法が求められているが，音楽動画から印象を推定することは容易ではない．

ここでニコニコ動画では，動画を視聴中のユーザが動画の任意の時間に対して自由にコメントすることができる．このコメントは動画を視聴したユーザが感じた印象をリアルタイムに文字にして表現していると考えることができる．つまり，こうしたコメントを使うことによって，視聴中の音楽動画に対する主観的な印象推定が可能になると期待される．なお，コメントを用いた音楽動画の印象推定についてはいくつかの研究がなされているが，そのいずれも音楽動画全体（音楽動画の最初から最後まで）を対象としたものである[4][5]．ここで，我々のこれまでの研究[6]により，音楽動画のサビ部分と，音楽動画全体とでは印象評価が大きく異なる事がわかっている．つまり，音楽動画の部分に注目して印象評価を行い，その印象評価の時間的評価によって音楽動画全体を評価する必要があると言える．一方，ニコニコ動画のコメントに代表されるソーシャルコメントはその音楽動画の再生時間に対するコメントであるものの，そのコメントが音楽に対するものか，映像に対するものか，双方に対するものかも不明である．

そこで我々は，ニコニコ動画のソーシャルコメントに注目し，音楽動画に対するコメントがどのメディアタイプ（音楽，映像，音楽と映像）でどういった印象の時に印象を推定可能なのかを検討する．ここではまず，我々が構築した印象評価データセット[6]を用いる．このデータセットは，

†1 明治大学
Meiji University
†2 JST　CREST
JST CREST
†3 京都大学
Kyoto University

音楽動画のサビ部分のみを対象として，音楽のみ，映像のみ，音楽と映像の組み合わせという 3 つのメディアタイプへの，8 つの印象軸について主観評価を行ったものである．このデータセットにおける音楽動画の該当部分に対するソーシャルコメントを収集し，そのコメントからの音楽動画の各メディアタイプ・印象タイプの推定精度を実験により示す．

2. 関連研究

音楽情報処理の分野では印象に基づく検索を実現するために，印象の推定や印象に基づく検索に関する研究が多数行われている．

2.1 楽曲の印象モデル

楽曲の印象の表現方法については多数のアプローチが提案されている．まず，楽曲の印象のクラスタリングに関しては，Hevner の研究[1]がある．この研究では楽曲に対する印象を，8 グループの印象群としてクラスタリングしている．また楽曲の印象推定に用いられるモデルとして，Russell が提案した Valence-Arousal 空間がある[2]．Valence は快-不快を表す次元，Arousal は覚醒-鎮静を表す次元であり，この 2 つの次元で印象を表現するという考え方である．

2.2 楽曲の印象推定

楽曲の印象推定に関する研究は近年様々に取り組まれている．その多くは楽曲の音響信号に基づく特徴量を利用した手法[8]であるが，他に楽曲の歌詞に基づく特徴量を利用する手法も提案されている[3][7]．また，音楽動画にユーザによって付与されたタグによる印象推定も行われている．本稿で行うコメントを利用した音楽動画の印象推定は新たな印象推定手法の 1 つになると考えられる．

2.3 メディア間の印象の差異

音楽動画に関して各メディア間から受ける印象の違いについては様々な研究がある．佐藤らの研究[9]では，視覚刺激が印象評価に大きく影響するとの結果が示されている．つまり，映像により受ける印象が強いことがわかっている．また，長谷川らの研究[10]では，ユーザの好みのジャンルが静止画と音楽の印象の類似に影響を与えることが明らかとなっている．しかし，こうした研究では大規模なデータセットを用いているわけではない．本稿では大規模なデータセットを用いることにより，このメディア間の印象の差異を明らかにしつつ，ソーシャルコメントからの推定可能性について検証するものである．

3. 印象評価データセット

本稿で検証するソーシャルコメントによる印象評価の推定可能性について，[6]において構築した印象評価データセットを利用する．この印象評価データセットは，音楽動画のサビ部分（RefraiD[11]によって推定されたサビ開始の 5 秒前から 30 秒間）のみを対象として，音楽のみ，映像のみ，音楽と映像の組み合わせのそれぞれのメディアタイプについて，8 軸の印象評価を 3 人以上が行ったものである．なお，評価対象となっている音楽動画は，動画共有サイトであるニコニコ動画上に投稿された音楽動画のうち，2012 年 8 月時点で「VOCALOID」というタグが付与されており，再生数が多い上位 500 個を抽出したものとなっている．

なお，印象評価については，音楽情報検索ワークショップである MIREX で用いられている 5 つの印象クラスと，Russell らの Valence-Arousal 空間という 7 つの軸に，[12]の研究で用いている「かわいい」という軸が追加されている．

データセットで用いられている 8 つの印象軸を表 1 に記す．表中の「印象クラス名」は，[6]および[12]において便宜上付与されている印象を表すラベル名である．なお，「印象を表す形容詞」は，データセット構築において評価者から評価値を収集する際に，その印象クラスを表現するために用いられたものである．本稿では，この印象評価値の 3 人分の平均を計算し，それぞれのメディアタイプ・印象タイプに対する評価値とする．

表 1　8 つの印象軸 [6][12]

C1（堂々）	堂々とした，どっしりとした 心躍る，にぎやかな
C2（元気が出る）	元気が出る，楽しい気持ちにさせる 陽気な，心地よい
C3（切ない）	切ない，悲痛な，ほろ苦い 気がめいる，哀愁の
C4（激しい）	アグレッシブな，激しい，興奮させる 感情的な，感情あらわな
C5（滑稽）	滑稽な，ユーモラスな，おもしろげな 奇抜な，気まぐれ，いたずらっぽい
C6（かわいい）	可愛らしい，愛くるしげ，愛おしい かわいい
Valence	明るい気持ちになる，楽しい 暗い気持ちになる，悲しい
Arousal	激しい，積極的な，強気な 穏やか，消極的な，弱気な

なお，印象評価データセットでは，C1 から C6 については 1（全くそう思わない）～5（とてもそう思う），Valence に対しては-2（暗い気持ちになる，悲しい）～+2（明るい気持ちになる，楽しい），Arousal に対しては-2（穏やか，消極的な，弱気な）～+2（激しい，積極的な，強気な）の各 5 段階評価を行われている．そこで，C1 から C6 に対する評価については，Valence-Arousal と比較しやすくするため，1～5 の評価値を単純に-2 することによって-2～+2 に変

換した.

4. 評価実験

人手で構築されたある音楽動画のあるメディアタイプに対する印象評価を，ソーシャルコメントから機械的にどの程度推定可能かを検討するため，印象評価データセットを用いた評価実験を行う．印象の推定においては，重回帰分析などを行うことによって印象評価式を求め，その印象評価式の性能を測る方法と，印象評価値によって動画集合を複数のクラスによって分類し，SVM（Support Vector Machine)などによって分類器を構築してその分類性能を測る方法とが考えられる．

今回用いる印象評価データセットでは，評価者 3 名であり評価に多少のぶれがあることから，印象評価値自体を推定することは適切で無いと考えられる．そこである一定以上の印象評価値をもち，人によってぶれが少ない印象を，機械的に推定可能かどうかを SVM により検証する．具体的には，あるメディアタイプ・印象タイプにおける印象評価値に基づき，高評価群・低評価群という 2 つの音楽動画集合を構築する．また，各集合を学習データとテストデータに分け，SVM で学習およびテストし，交差検定を行うことによってソーシャルコメントからの高評価群の分類性能を評価する．

ここではまず，ソーシャルコメントの収集および SVM のための単語ベクトルの生成方法について述べ，データの量に対する基礎検討を行う．また，高評価群，低評価群の閾値と,単語ベクトルを生成する方法を切り替えたときに,各メディアタイプ・各印象タイプをどの程度推定可能なのかを示す．またこの結果により，適切な閾値や単語ベクトルの生成方法，メディアタイプおよび印象タイプにおける印象推定可能性について議論を行う．

4.1 音楽動画に対する単語ベクトル生成

ソーシャルコメントから音楽動画の各メディアタイプの各印象における推定精度を考察するため，印象評価データセットに該当する音楽動画のコメントを収集する．ここでは,該当する音楽動画に対するすべてのコメントを,2015 年 7 月 23 日にかけてニコニコ動画 API を利用して収集し，860,455 個のコメントを集めた．その後，印象評価データセットで用いられた音楽動画の開始時間，終了時間にもとづき，各動画のサビ区間内に投稿されたコメントを抽出する．これにより，132,036 個のコメント（1 動画あたり平均 264.1 個）が抽出された.

次に，ソーシャルコメントからの音楽動画の単語ベクトルを生成するために，各動画のサビ区間のコメントを，MeCab を用いて形態素解析することによって単語に分割し，各音楽動画における単語の出現頻度を数える．ここでは，印象は形容詞に影響されると考え，すべての品詞を利用する all 手法，コメント中の形容詞だけを用いて単語ベクトルを作成する adj 手法を用意した.

4.2 印象分類の基礎検討

先述の通り，本稿ではその評価値が高いものと低いものに分け，その高い印象に分類される音楽動画をどの程度判定できるかを評価指標とする．ここではまず機械学習における基礎検討を行うため，メディアタイプごとに各印象の評価値が 1 以上（高評価群）と-1 以下（低評価群）の場合と，評価値が 0.5 以上（高評価群）と-0.5 以下（低評価群）の場合の 2 通りについて動画集合を構築した.

それぞれのメディアタイプ・印象タイプにおける音楽動画数は表 2～5 の通りである．なおこれ以降の表において，Movie は音楽動画を，Audio は音楽のみを，Visual は映像のみを意味する．また，V は Valence を A は Arousal を意味している.

表 2　評価値 1 以上の音楽動画数

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	76	105	87	54	83	104	101	150	95.0
Audio	133	127	46	69	49	73	124	178	99.9
Visual	21	50	142	49	81	78	57	111	73.6
平均	76.7	94.0	91.7	57.3	71.0	85.0	94.0	146.3	89.5

表 3　評価値-1 以下の音楽動画数

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	105	169	191	209	178	215	62	94	152.9
Audio	65	92	232	195	180	209	61	43	134.6
Visual	252	272	165	247	207	234	96	155	203.5
平均	140.7	177.7	196.0	217.0	188.3	219.3	73.0	97.3	163.7

表 4　評価値 0.5 以上の音楽動画数

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	135	142	140	90	119	138	157	200	140.1
Audio	205	187	91	107	91	115	189	238	152.8
Visual	44	83	192	65	114	112	100	144	106.8
平均	128.0	137.0	141.0	87.3	108.0	121.7	148.7	194.0	133.2

表 5　評価値-0.5 以下の音楽動画数

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	157	217	224	267	235	263	107	139	201.1
Audio	98	146	277	253	255	270	107	89	186.9
Visual	312	319	190	308	276	267	163	213	256.0
平均	189.0	227.3	230.3	276.0	255.3	266.7	125.7	147.0	214.7

これらの音楽動画集合から各特徴ベクトルを使用して機械学習を行うが，Visual-C1 や，Audio-C1 のようにメディアタイプ・印象タイプの組み合わせによってデータの偏りがあり，不均衡データ問題となると考えられる．そこで

まず，データの不均衡を考慮せず，高評価群を正例，低評価群を負例としてそれぞれ 5 分割し，その内の 4 つを SVM の訓練データ，1 つをテストデータとして交差検定（5-fold クロスバリデーション）を行い，正例，負例それぞれの適合率を計算する．

ここで印象評価値が 1 以上を正例，-1 以下を負例としたものの判定に関する適合率の平均を求めたものが表 6 と表 7 である．また，表 8 および 9 は，そのそれぞれのメディアタイプ・印象タイプにおける動画数から計算されるランダムサンプリングした時に正例および負例となる確率である．なお，単語ベクトルの生成においては all 手法を用いた．

表 6　all 手法の適合率（正例：1 以上）

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.645	0.769	0.56	0.568	0.553	0.734	0.794	0.837	0.682
Audio	0.790	0.741	0.283	0.600	0.425	0.569	0.821	0.886	0.639
Visual	0.263	0.471	0.669	0.360	0.594	0.667	0.529	0.713	0.533
平均	0.566	0.660	0.504	0.509	0.524	0.657	0.715	0.812	0.618

表 7　all 手法の適合率（負例：-1 以下）

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.653	0.777	0.547	0.463	0.519	0.690	0.85	0.865	0.670
Audio	0.742	0.664	0.289	0.435	0.354	0.465	0.787	0.886	0.578
Visual	0.238	0.48	0.688	0.367	0.481	0.623	0.482	0.759	0.515
平均	0.545	0.640	0.508	0.422	0.451	0.593	0.706	0.837	0.588

表 8　ランダムサンプリングした時に正例となる確率（1 以上）

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.42	0.383	0.313	0.205	0.318	0.326	0.62	0.615	0.4
Audio	0.672	0.58	0.165	0.261	0.214	0.259	0.67	0.805	0.453
Visual	0.077	0.155	0.463	0.166	0.281	0.25	0.373	0.417	0.273
平均	0.39	0.373	0.314	0.211	0.271	0.278	0.554	0.612	0.375

表 9　ランダムサンプリングした時に負例となる確率（-1 以下）

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.58	0.617	0.687	0.795	0.682	0.674	0.38	0.385	0.6
Audio	0.328	0.42	0.835	0.739	0.786	0.741	0.33	0.195	0.547
Visual	0.923	0.845	0.537	0.834	0.719	0.75	0.627	0.583	0.727
平均	0.61	0.627	0.686	0.789	0.729	0.722	0.446	0.388	0.625

この結果より，正例についてはランダムサンプリングした時に正例となる確率を上回ることが多いものの，負例については全体的にランダムサンプリングした時に負例となる確率を下回っていることが分かる．また，メディアタイプ・印象タイプによって精度（適合率）が大きく異なっており，単純に動画数が多いものが高く評価されているにすぎないことがわかる．そのため，どのメディアタイプ・印象タイプがどの程度の精度で推定可能かについては検討が行えない．

このことから，アンダーサンプリングを行うことによりメディアタイプ・印象タイプごとの高評価群，低評価群の動画数を同一にして学習を行わないと，メディアタイプ・印象タイプによる差を比較できないことが分かる．このことより，これ以降ではアンダーサンプリングを行い，実験および評価を行う．

4.3 メディアタイプ・印象タイプによる精度比較

表 10〜13 は，all 手法，adj 手法それぞれで単語ベクトルを生成し，印象評価値を「1 以上を高評価群，-1 以下を低評価群とした場合」「0.5 以上を高評価群，-0.5 以下を低評価群とした場合」のそれぞれの動画集合から実験を行ったときの，各印象タイプ・各メディアタイプの SVM による分類の適合率を示したものである．なお，ここでも動画集合を 5 つに分け，4 つを訓練データ，1 つをテストデータとし，交差検定（5-fold クロスバリデーション）により適合率平均を計算した．

ここで「1 以上／-1 以下」と，「0.5 以上／-0.5 以下」条件を用意しているのは，閾値をどの程度とるのが適切かを調べるためである．各表において値が 0.8 以上のものをオレンジ色で，0.6 以下のものを青色で示している．

表 10　all 手法（1 以上）の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.720	0.830	0.713	0.765	0.718	0.758	0.783	0.777	0.758
Audio	0.742	0.671	0.612	0.661	0.600	0.712	0.704	0.744	0.681
Visual	0.611	0.680	0.752	0.714	0.603	0.797	0.660	0.743	0.695
平均	0.691	0.727	0.692	0.713	0.640	0.756	0.712	0.755	0.711

表 11　adj 手法（1 以上）の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.733	0.869	0.710	0.750	0.667	0.838	0.650	0.842	0.757
Audio	0.667	0.635	0.595	0.667	0.581	0.775	0.706	0.733	0.670
Visual	0.714	0.736	0.733	0.759	0.536	0.829	0.603	0.850	0.720
平均	0.705	0.745	0.679	0.725	0.595	0.814	0.653	0.809	0.716

表 12　all 手法（0.5 以上）の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.648	0.754	0.669	0.795	0.676	0.805	0.796	0.741	0.735
Audio	0.646	0.638	0.621	0.727	0.598	0.645	0.708	0.738	0.665
Visual	0.571	0.671	0.689	0.725	0.657	0.779	0.684	0.763	0.692
平均	0.622	0.688	0.660	0.749	0.644	0.743	0.729	0.747	0.698

表 13　adj 手法（0.5 以上）の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.650	0.786	0.717	0.714	0.663	0.889	0.670	0.782	0.734
Audio	0.568	0.634	0.542	0.706	0.644	0.706	0.683	0.604	0.636
Visual	0.614	0.634	0.656	0.738	0.594	0.767	0.652	0.778	0.679
平均	0.610	0.684	0.638	0.720	0.633	0.787	0.668	0.721	0.683

それぞれの表を比較すると，全体の適合率の平均はおよそ 7 割前後であることがわかる．また，各手法における評価値が「1 以上／-1 以下」,「0.5 以上／-0.5 以下」の全体の適合率の平均は，どちらも「1 以上／-1 以下」の方が高い値を示しているが，適合率の差は 2～3%程度であることも分かる．

Movie，Audio，Visual を比較すると，平均化した場合，いずれの手法および条件においても Movie が最も適合率が高く，次に Visual で，Audio が最も悪い結果となっていることが分かる．また，印象タイプや評価値の閾値，手法よらず Movie の適合率が他のメディアタイプと比べ最下位になっていないことがわかる．さらに，どのメディアタイプにおいても C5 についての結果は悪いものとなっている．

一方，all 手法と adj 手法を比較すると， all 手法の方が全体的に良い結果であるが，C3，C5，Valence については all 手法が，C6，Arousal については adj 手法が良い結果となっていることが分かる．また，C2，C3，C6，Arousal では Audio が，Valence では Visual が最も悪い結果となっていることがわかる．さらに，C3，C5，Valence を見ると all 手法が adj 手法よりも精度が良いことがわかる．

各印象タイプについて見ていくと，C2 ではメディアごとの精度が Movie，Audio，Visual の順に良いことがわかる．また，C3 では Visual の結果が良い，C6 では Audio の結果が悪いことがわかる．さらに，C6 に関しては，唯一 adj 手法の方が良い精度を示した．これは，かわいさにおいて形容詞が効果的に働いたからであると考えられる．一方，Valence と Arousal に関しては，どちらも Movie 結果が良く，Visual の結果が悪いことがわかる．

5. 考察

各印象・メディアタイプにおいてソーシャルコメントからの推定精度には差があることがわかった．

表 10～13 の比較で「1 以上／-1 以下」「0.5 以上／-0.5 以下」を比較した時に，「1 以上／-1 以下」条件の方が精度は高かったが，2%程度であり大きな差ではなかった．一方，表 2～5 において，「1 以上／-1 以下」条件の動画数に比べ「0.5 以上／-0.5 以下」条件の動画数は 1.37 倍となっている．そのため，「0.5 以上／-0.5 以下」条件を用いることで，評価のブレが減るのではと期待される．

印象タイプごとに見ると，C6（かわいい）に対する印象推定精度は他と比べて高いことがわかった．特に adj 手法を用いると，all 手法より推定精度が高くなっていることから，C6 に対するコメントは形容詞に特徴があると考えられる．また，Movie と Visual の値が Audio よりも高く出ていることから C6 に対するコメントは音楽に対するものでなく，音楽動画もしくは映像に対するものであると考えられる．特に，表 11 において Movie-C6 の精度が高いため，ソーシャルコメントによる印象推定が C6 の印象タイプに対して非常に有効であると考えられる．

一方，C1（堂々），C3（元気が出る），C5（滑稽）の推定精度は他の印象タイプと比べて低いことがわかった．C1 に関しては，使用する学習データの数が少ないことによって精度が低くなっているとも考えられるが，これら 3 つの印象タイプに関しては，評価値「1 以上／-1 以下」の C1 を除き，adj 手法により精度が下がっていることから，形容詞を使用することはこれらの印象タイプを推定する際には適切ではないことがわかる．今後は，用いなかった別の品詞を使用して音楽動画に対する単語ベクトルを生成することで結果を検証する予定である．

表 14　各メディアタイプで各印象タイプのみが 1 以上それ以外は 1 より小さい動画の数

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	3	2	51	4	83	10	31	23	18.4
Audio	9	8	31	6	13	8	19	30	15.5
Visual	1	0	108	8	31	17	10	22	24.6
平均	4.3	3.3	63.3	6	22.3	11.7	20	25	19.5

表 14 は，各メディアタイプにおいて，各印象タイプの評価値のみが 1 以上で，それ以外の印象タイプの評価値は 1 以下の動画の数である．この表から，C3（切ない）の評価値のみが 1 以上を示している動画の数が他の印象タイプと比べても非常に多いことがわかる．このことから，C3 は他の印象とは独立していることが顕著であると言える．また表 2 より，C3 のコメント数は Visual に対するものが最も多く，Audio に対するものが最も少ない．推定精度に関しても Visual が高く，Audio は低くなっていることから C3 は他の印象タイプとは独立していて，かつ映像から受ける影響が大きいことがわかる．

「1 以上／-1 以下」において，adj 手法を用いると，all 手法との精度が印象タイプごとに違うことがわかった．C6（かわいい）ではすべてのメディアタイプにおいて精度が上がっていることから adj 手法が有効であるといえる．一方，C3（切ない）と C5（滑稽）についてはすべてのメディアタイプで精度が下がっている．この違いについては，印象タイプを直接表す形容詞をコメントとして用いる頻度によるものであると考えられる．この違いが何によるもの

なのか，今後の研究により明らかにする予定である．

メディアごとの精度に関しては，Movie の推定精度が最も高い．このことから，平均化するとソーシャルコメントの多くは音楽と映像がミックスされた音楽動画自体に投稿される傾向があると考えられる．

今回用いた動画はいずれもコメント数が多いものであり，また他者からの評価も高いものである．しかし一般的にニコニコ動画上の音楽動画に対して投稿されるコメント数は今回用いた音楽動画のように多いものではない．そこで今後は今回のデータセットで用いたコメント数を減らし，コメント数の変化によってどの程度精度が変化するかなどについても検証予定である．

6. まとめ

本稿では，500 曲，3 メディアタイプ，8 軸の印象評価からなる印象評価データセットを用い，ソーシャルコメントからメディアタイプ・印象タイプごとの印象推定の可能性について実験を行った．ここでは特に，音楽動画の単語ベクトルを作成し，各印象・メディアタイプごとに SVM を用いて印象推定し，それについて考察した．その結果，コメントによる推定精度が各印象・メディアタイプごとに差があることを明らかにした．コメントからの印象推定精度は C6（かわいい）と Arousal に関して特に高く，この 2 つの印象タイプについてはコメントからの推定が有効であると考えられる．一方で，C1（堂々），C3（切ない），C5（滑稽）では精度がそれほどよいものでないため，コメントからの推定を有効なものにするためには更なる検討が必要である．メディアタイプに関しては，音楽動画に対する精度が他のメディアタイプと比べ高いため，音楽動画に対するコメントからの印象推定が有効であると言える．

また，2 つの手法で単語ベクトルを生成し，それぞれについて実験を行いその精度について考察を行った．形容詞を用いることで，印象評価値が「1 以上／-1 以下」条件において精度がよくなった．特に，C6 では，すべてのメディアタイプにおいて精度が上昇している．一方，C3 と C5 のすべてのメディアタイプで精度が下がっていた．この結果により，用いる品詞によって各印象タイプの推定精度に違いが出ることが考えられる．

今回の実験で, 全体としての推定精度は 7 割程度であり，それほど高い値であるとは言えない結果であった. しかし，印象・メディアタイプごと見ると，精度が 8 割を越すものもあることから，単語ベクトルの取り方や評価値の閾値の設定を変えることで精度は上げることができると考えられる．また，今回の実験で用いた動画数が少ないものもあったため，より大規模なデータセットを用いることによってより正確な精度を出すことができると考えられる．こうした点は今後の課題である．

参考文献

1) Hevner, K.: Experimental studies of the elements of expression in music, The American Journal of Psychology, Vol.48, No.2, pp.246-268 (1936)
2) Russell, James A.: A Circumplex Model of Affect, Journal of Personality and Social Psychology, 39(6), pp.1161-1178 (1980).
3) 舟澤慎太郎，北市健太郎，甲藤二郎: 楽曲推薦システムのための楽曲波形と歌詞情報を考慮した類似楽曲検索に関する一検討, 情報処理学会研究報告オーディオビジュアル複合情報処理, pp.1-5 (2013)
4) 中村聡史，田中克己: 印象に基づく動画検索，情報処理学会研究報告ヒューマンコンピュータインタラクション (2009-HCI-131)，pp.77-84 (2009).
5) 山本岳洋，中村聡史: 視聴者の同期コメンロを用いた楽曲動画の印象分類，情報処理学会論文誌，Vol.6，No.3，pp.66-72(2013)
6) 大野直紀，中村聡史，山本岳洋，後藤真孝: 音楽動画への印象評価データセット構築とその特性の調査，情報処理学会研究報告，Vol.2015-MUS-108, No.7, pp.1-9 (2015).
7) 西川直毅，糸山克寿，藤原弘将，後藤真孝，尾形哲也，奥乃博: 歌詞と音響特徴量を用いた楽曲印象軌跡推定法の設計と評価, 情報処理学会研究報告，Vol.2011-MUS-91，No.7, pp. 1-8 (2011).
8) 絵本詩織，糸山克寿，奥乃博: 音響特徴量を用いた楽曲印象分布の推定，情報処理学会 76 回全国大会，pp.391-392 (2014).
9) 佐藤淳也，佐川雄二，杉江昇: 音と映像の組み合わせによる主観的印象の変化，映像情報メディア学会誌，Vol.55，No7, pp.1053-1057 (2001).
10) 長谷川優，武田昌一: 好みの音楽ジャンルに着目した静止画と音楽の組み合わせに関する考察: −個人の属性に着目した静止画と音楽に対する印象度の相互比較−，日本感性工学会論文誌, Vol.11，No.3，pp.435-442 (2012).
11) 後藤真孝: SmartMusicKIOSK: サビ出し機能付き音楽視聴機，情報処理学会論文誌，Vol.44，No.11，pp.2737-2747 (2003)
12) 山本岳洋，中村聡史: 楽曲動画印象データセットの作成とその分析，ARG 第 2 回 Web インテリジェンスとインタラクション研究会 (2013).