ソーシャルコメントからの音楽動画印象推定手法の提案

土屋 駿貴[†1]　大野 直紀[†1]　中村 聡史[†1]　山本 岳洋[†2]

†1 明治大学総合数理学部 〒164-8525 東京都中野区 4-21-1
†2 京都大学大学院情報学研究科 〒606-8501 京都府京都市左京区吉田本町 4-5-6
E-mail: †1 {ev30616 , ev30508}@meiji.ac.jp, satoshi@snakamura.org
†2 tyamamoto@dl.kuis.kyoto-u.ac.jp

あらまし　我々は音楽動画の印象に基づく検索や推薦の実現，またそのためのデータセットの容易な構築・拡張を可能とすることを目的としている．そのために，これまでの研究で構築してきた音楽のみ，映像のみ，音楽と映像のセットという 3 つのメディアタイプに対する印象評価データセットを用いて，ソーシャルコメントからの各メディアのコンテンツに対する印象を自動推定する手法を提案する．また，その精度について実験を行い，用いるソーシャルコメント中の単語を品詞で使い分けることで，メディアや印象によって推定精度に差が出ることを明らかにした．

キーワード　ソーシャルコメント，音楽動画，印象推定

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，音楽動画を検索する方法は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一般的である．これらの検索方法では，目的とする音楽動画に直接辿り着くことが可能であるが，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る．

一方，目的とする音楽動画が明確でなく，また目的とするキーワードが曖昧な状態で音楽動画の検索を行いたいというニーズは存在している．例えば,「面白い音楽動画を見て楽しい気分になりたい」や「切ない気分だから悲しい音楽動画を視聴して泣きたい」などといった，ユーザが音楽動画を見て感じると期待される印象かを用いた検索・推薦が挙げられる．こうした印象を用いた検索や，視聴中の音楽動画の印象に類似した印象での音楽動画の推薦が可能となれば，ユーザは今までにない新しい視点からの検索や推薦が可能となるうえ，新たなる音楽動画との出会いも生まれると期待される．

印象に基づく検索や推薦を可能とするには，そのそれぞれの音楽動画にユーザがどのように印象を抱くのかということを，人手で付与，または機械的に自動推定する必要がある．印象の付与を人手で行う場合，その印象付与にかかる時間的なコストは膨大であり，多くの人手を集めることも容易ではない．また，機械的に自動推定することも容易では無いため，正解を含む大きなデータセットを用意して手法を検討していくことになるが，そうしたデータセットを用意するのもまた困難である．そのため，この印象評価データセットを手軽に構築・拡張することが可能となれば印象検索や印象推定の研究に寄与し，手法実現に近づくと考えられる．

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，こうしたコメントがある一定以上集まっている音楽動画であれば，視聴中の音楽動画に対する主観的な印象の推定が可能になると期待される．また，コメントデータはニコニコ動画上において膨大な数存在しているため，事前にある程度のデータセットを構築しておき，そのデータセット内で再現率ではなく適合率を重視した判定手法を実現することにより，最初に作ったデータセットを拡張できると期待される．

なお，コメントを用いた音楽動画の印象推定についてはすでに多くの研究がなされているが，その多くは音楽動画全体（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]．これまでの我々の研究[6]において，サビ部分と音楽動画全体とでは視聴者の受ける印象が大きく異なることを明らかにしてきた．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．また，音楽動画に対するコメントは，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．

こうした問題に対処するため，我々はこれまでの研究[6]で，500 曲の音楽動画のサビ部分についてメディアタイプ（音楽，映像，音楽と映像の組み合わせ）・印象タイプの組み合わせに対する印象評価データセットを構築してきた.またそのそれぞれ音楽動画について，

ソーシャルコメントから単語ベクトルを生成する手法を提案する．生成の際には，全ての品詞を用いる手法と形容詞のみを用いる手法を提案し，コメントからどの程度印象推定可能なのかを実験的に明らかにしてきた．しかし，この 2 つの手法ではメディア・印象タイプによっては精度がかなり低いものとなっており，十分ではなかった．

そこで本稿では，ソーシャルコメントから音楽動画に対する単語ベクトルを生成する際に，以前の実験で用いなかった品詞を用い，様々な手法を検討することで，より多くのメディア・印象タイプの印象推定精度向上を図る．

以下，2 章では関連する研究についてまとめ，3 章で印象評価データセットの詳しい説明を行う．4 章では評価実験について記し，5 章では実験の結果から考察し，6 章でまとめを行う．

2. 関連研究

音楽情報処理の分野では印象に基づく検索や推薦を実現するために，印象の推定や印象に基づく多くの研究が多数なされている．

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

本稿では，[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 件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．

表 1　8 つの印象軸[6][12]

印象クラス名	印象を表す形容詞
C1（堂々）	堂々とした，どっしりとした 心躍る，賑やかな
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，感情的な 興奮させる，感情あらわな
C5（滑稽）	滑稽な，ユーモラスな，面白げな 奇抜な，気まぐれ，いたずらっぽい
C6（かわいい）	可愛らしい，愛くるしい，かわいい 愛おしい
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 穏やかな，消極的な，弱気な

表 1 はデータセットで用いられている 8 つの印象軸をまとめたものである．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

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-3 することによって-2～+2 に変換した．

4. 評価実験

人手で構築された音楽動画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．

評価実験の方法は，印象評価データセットにおいて印象評価値が一定以上あり，人によってブレが少ない印象を機械的に推定可能かどうかについて，SVM を用いて検証する．具体的には，あるメディア・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生成方法について述べ，データの量に関する基礎検討を行う．また，その単語ベクトルを生成する方法によって，各メディア・印象タイプでどの程度推定可能なのかを示す．その結果により，各メディア・印象タイプにおける印象推定の適切な単語ベクトル生成方法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全てのコメントを，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，終了時間に基づき，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に対する単語ベクトルを生成する。まず，抽出した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，各単語の出現頻度を数えたものを音楽動画に対する単語ベクトルとする．

ここで，これまでの研究[13]では全ての品詞を利用する all 手法と形容詞だけを利用する adj 手法の 2 つの手法で実験を行った．その結果，adj 手法を用いると C6（かわいい）と Arousal の精度が高く，ソーシャルコメントからの印象推定の可能性があることが明らかとなっていた．しかし，他の印象に関してはあまり高い精度を得ることができなかった．その理由として C6 と Arousal の印象を文字として表現する際に形容詞を多く用いるからであると考えられる．

そこで本稿では，C6 と Arousal 以外の印象についても高い推定精度を得るために他の手法を用意する．具体的には，以前の研究で用いていた形容詞に加え，動画によって特徴が出ると考えられる名詞と動詞，「もっと」や「とても」など受ける印象の程度を表すと考えられる副詞の 4 つの品詞を用いる．また，それぞれの品詞を 2 つずつ組み合わせた手法も用意する．以上 10 個の手法と，これまでの研究で用意した All 手法，今回使用する 4 つの品詞を用いた All2 手法の合計 12 個の手法を用意した．それらのすべての手法名と使用する品詞については下の表 2 にまとめた．

表 2　単語ベクトル生成手法

手法名	用いる品詞
All 手法	全ての品詞
All2 手法	名詞，動詞，形容詞，副詞
Noun 手法	名詞
Verb 手法	動詞
Adj 手法	形容詞
Adv 手法	副詞
Noun-Verb 手法	名詞，動詞
Noun-Adj 手法	名詞，形容詞
Noun-Adv 手法	名詞，副詞
Verb-Adj 手法	動詞，形容詞
Verb-Adv 手法	動詞，副詞
Adj-Adv 手法	形容詞，副詞

4.2 手法による精度比較

先述の通り，本稿では各メディア・印象タイプにお

いて評価値が高いものと低いものに分け，その評価値が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，評価値が 1 以上（高評価群）と-1 以下（低評価群）の動画集合を構築し，それらの音楽動画集合から各単語ベクトルを使用して機械学習を行う．具体的には，高評価群を正例，低評価群を負例としてそれぞれを 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，正例の適合率を計算する．また，我々のこれまでの研究[13]において，メディアタイプ・印象タイプの組み合わせによっては高評価群と低評価群の音楽動画数に大きな偏りがあり，不均衡データ問題が起きていた．これに対処するため，アンダーサンプリングを行うことで高評価群と低評価群の音楽動画数を同一にして行っていた．従って，本稿でもその問題に対処するためにアンダーサンプリングを行い実験した．

以下の表 3～14 は用意した全ての手法で単語ベクトルを生成し，実験を行った時の，各メディアタイプ・印象タイプの SVM による正例に関する適合率の平均を示したものである．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．また，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 を意味する．

表 3　All 手法の適合率

	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 4　All2 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.645	0.814	0.705	0.765	0.728	0.792	0.694	0.822	0.745
Audio	0.738	0.658	0.566	0.750	0.725	0.787	0.736	0.778	0.717
Visual	0.880	0.786	0.390	0.725	0.564	0.776	0.814	0.870	0.725
平均	0.754	0.753	0.554	0.747	0.672	0.785	0.748	0.823	0.730

表 5　Noun 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.575	0.720	0.644	0.653	0.704	0.680	0.646	0.652	0.659
Audio	0.698	0.606	0.528	0.621	0.721	0.661	0.708	0.650	0.649
Visual	0.700	0.640	0.608	0.600	0.620	0.688	0.552	0.641	0.631
平均	0.658	0.655	0.593	0.625	0.682	0.676	0.635	0.648	0.647

表 6　Verb 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.667	0.627	0.440	0.544	0.642	0.714	0.575	0.574	0.597
Audio	0.615	0.622	0.133	0.658	0.587	0.500	0.600	0.551	0.533
Visual	0.588	0.549	0.606	0.517	0.584	0.573	0.508	0.654	0.572
平均	0.623	0.599	0.393	0.573	0.604	0.596	0.561	0.593	0.568

表 7　Adj 手法の適合率

	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	0.669
Visual	0.714	0.736	0.733	0.759	0.536	0.829	0.603	0.850	0.720
平均	0.705	0.747	0.679	0.725	0.595	0.814	0.653	0.808	0.716

表 8　Adv 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.618	0.586	0.522	0.576	0.520	0.481	0.556	0.603	0.557
Audio	0.679	0.600	0.580	0.537	0.545	0.481	0.642	0.538	0.575
Visual	0.879	0.759	0.211	0.632	0.519	0.451	0.777	0.805	0.629
平均	0.725	0.648	0.438	0.582	0.528	0.471	0.658	0.649	0.587

表 9　Noun-Verb 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.687	0.699	0.648	0.620	0.681	0.714	0.661	0.636	0.668
Audio	0.683	0.580	0.489	0.642	0.689	0.672	0.729	0.658	0.642
Visual	0.881	0.760	0.308	0.614	0.595	0.639	0.805	0.859	0.682
平均	0.750	0.680	0.482	0.625	0.655	0.675	0.732	0.718	0.665

表 10　Noun-Adj 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.662	0.854	0.690	0.780	0.750	0.778	0.694	0.800	0.751
Audio	0.754	0.644	0.612	0.750	0.707	0.772	0.740	0.806	0.723
Visual	0.888	0.792	0.409	0.706	0.657	0.768	0.821	0.874	0.739
平均	0.768	0.763	0.570	0.745	0.705	0.773	0.752	0.827	0.738

表 11　Noun-Adv 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.592	0.714	0.644	0.654	0.722	0.673	0.656	0.649	0.663
Audio	0.672	0.589	0.538	0.621	0.711	0.661	0.694	0.632	0.639
Visual	0.879	0.763	0.372	0.636	0.622	0.683	0.805	0.852	0.701
平均	0.714	0.689	0.518	0.637	0.685	0.672	0.718	0.711	0.668

表 12　Verb-Adj 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.781	0.811	0.711	0.684	0.667	0.856	0.652	0.784	0.743
Audio	0.692	0.627	0.520	0.714	0.682	0.740	0.673	0.707	0.669
Visual	0.921	0.734	0.400	0.734	0.511	0.764	0.779	0.871	0.714
平均	0.798	0.724	0.544	0.711	0.62	0.787	0.701	0.787	0.709

表 13　Verb-Adv 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.667	0.568	0.535	0.531	0.657	0.630	0.600	0.660	0.606
Audio	0.677	0.560	0.458	0.566	0.587	0.513	0.589	0.581	0.566
Visual	0.882	0.729	0.250	0.622	0.488	0.529	0.724	0.814	0.629
平均	0.742	0.619	0.414	0.573	0.577	0.557	0.638	0.685	0.601

表 14　Adj-Adv 手法の適合率

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	0.700	0.837	0.679	0.690	0.681	0.848	0.695	0.844	0.746
Audio	0.733	0.646	0.581	0.634	0.683	0.743	0.667	0.718	0.675
Visual	0.911	0.765	0.477	0.653	0.622	0.757	0.840	0.884	0.738
平均	0.781	0.749	0.579	0.659	0.662	0.783	0.734	0.815	0.720

まず All 手法と All2 手法を比較する．各メディア・印象タイプごとにみると All2 手法の方が 0.8 を越す高い値が多くなっており，全体の平均値も All 手法に比べ高くなっていることがわかる．しかし，C3（切ない）の印象に関してはすべてのメディアタイプにおいて値が低くなっている．

次に，1 つのみの品詞を使用している手法を比較する．Noun 手法，Verb 手法，Adv 手法では，0.8 を越す高い値が Adv 手法に少し見られるのみで，Adj 手法と比べると推定精度はあまり良くない結果となっていることがわかる．特に Adv 手法では 0.6 を下回る低い値が多く見られる．また，Adj 手法に関しては，特に C6（かわいい）と Arousal の精度が高いことが表から改めてわかる．

2 つの品詞を用いた手法を比較すると，品詞を組み合わせることで高い値が少なくとも 2 つ以上は見られることがわかる．特に形容詞を含んでいる手法では高い値が多いことがわかる．さらに，Audio に関しては 0.8 を越す高い値がこれまで見られなかったが Noun-Adj 手法の Audio-Arousal において唯一 0.8 を越す結果となった．一方，C3 の印象に関してはどの手法を用いても高い値が見られず，0.6 を下回る低い値が多いことがわかる．

各メディア・印象タイプごとについて見ていくと，Visual-C1（堂々）や Movie-C2（元気が出る），Visual-Arousal はどの手法を用いても比較的高い値が見られる．このように，高い値が見られるメディア・印象タイプにある程度の偏りがあることもわかる．

5. 考察

ソーシャルコメントからの推定精度は各手法によって，各メディア・印象タイプによって差がでることがわかった．

All 手法と All2 手法を比べてみると，All2 手法の方が高い値が多く見られる．また，全体の平均値も All2 手法の方が高くなっている．これは，All 手法では鉤括弧や顔文字などで使用される記号なども含め全ての品詞を用いているため，印象を表現していると考えにくい品詞も用いているためであると考えられる．しかし，All 手法では他の手法と比べて唯一 C3（切ない）で 0.6 を下回る低い値が見られないため，C3 の印象は All 手法のみで用いている印象を表しにくい品詞を用いることで，精度の向上が図れるのではないかと考えられる．

1 つの品詞のみを用いた手法を見ると，Adj 手法では高い値が現れているが，他の 3 つの手法では高い値が見られず，全体の平均値もとても低くなっていることがわかる．このことから，名詞，動詞，副詞は印象を表す際にはあまり用いられない，もしくは，印象によって使われる単語に特徴がないと考えられる．そのため，ユーザは形容詞を用いて印象を表現することが最も多く，また使われる単語に特徴があるのではないかと考えられる．

次に，2 つの品詞を組み合わせた手法を見ると，1 つの品詞のみを用いた手法とは違い，ある特定のメディア・印象タイプで 0.8 を越す値が出ていることから，そのメディア・印象タイプを表す際の品詞に特徴があると考えられる．特に，Noun-Verb 手法と Noun-Adv 手法の Visual-Valence の結果に関しては高い値になっているが Noun 手法，Verb 手法，Adv 手法のいずれにおいても Visual-Valence で高い値は見られず，組み合わせることで高い値になっていることがわかる．このように，単語ベクトル生成の際に使用する品詞の組み合わせによって結果に大きな差が出ることから，今回使

用しなかった品詞も他の品詞と組み合わせることによって推定精度が高くなるのではないかと考えられる．しかし，C3 の印象に関しては，どの組み合わせにおいても値が低くなっている．今回の実験で用いた品詞では特徴が出にくく，先述したように記号等の品詞で特徴が出ることも考えられるが，名詞，動詞，形容詞，副詞は文章を構成する際の重要な品詞であり，それらを用いて値が低くなるということは，C3 の印象はソーシャルコメントから推定することは困難であることも考えられる．

印象タイプごとに見ると，Visual-C1（堂々）や Movie-C2（元気が出る），Visual-Arousal のようにどの手法においても比較的高い値が出ているメディア・印象タイプがある．これは，このメディア・印象タイプがソーシャルコメントから推定しやすいということも考えられるが，そもそも今回用いたソーシャルコメントに特徴があるということも考えられるため，今後は使用しているソーシャルコメントの量や使用されている単語に関しても考察していく必要があると考えられる．

表 15　各メディア・印象タイプで最も高い値を出した手法

	C1	C2	C3	C4	C5	C6	V	A
Movie	Verb-Adj	Adj	All	Noun-Adj	Noun-Adj	Verb-Adj	All	Adj-Adv
Audio	Noun-Adj	All	Noun-Adj	Noun-Adj	All2	All2	Noun-Adj	Noun-Adj
Visual	Verb-Adj	Noun-Adj	All	Adj	Noun-Adj	Adj	Adj-Adv	Adj-Adv

表 16　各メディア・印象タイプで最も高い値

	C1	C2	C3	C4	C5	C6	V	A	平均
Movie	**0.781**	**0.869**	**0.713**	**0.780**	**0.750**	**0.856**	**0.783**	**0.844**	**0.797**
Audio	**0.754**	**0.671**	**0.612**	**0.750**	**0.725**	**0.787**	**0.740**	**0.806**	**0.731**
Visual	**0.921**	**0.792**	**0.752**	**0.759**	**0.657**	**0.829**	**0.840**	**0.884**	**0.804**
平均	**0.819**	**0.777**	**0.692**	**0.763**	**0.711**	**0.824**	**0.788**	**0.845**	**0.777**

表 15，16 は，各メディア・印象タイプで最も高い値を出した手法とその値をまとめたものである．表 15 を見ると，全てのメディア・印象タイプで形容詞を含む手法が最も高い値を出していることがわかる．このことから，印象を表す際には形容詞が用いられ，その使われる形容詞に特徴が出やすいことが考えられる．また，ソーシャルコメントからの音楽動画印象推定の際には形容詞が重要な品詞であることも考えられる．次に，その値について見ると，（3 つのメディアタイプ）×（8 つの印象タイプ）で 24 パターンあるうちの 20 のメディア印象タイプで 0.75 を越す値が出ていることがわかる．今回使用しているデータセットの評価値は 3 人の評価の平均値であり，ブレがあるため 0.75 を越す分類精度は比較的有効であると考えられる．特に，0.8 を越す値に関しては，8 割の精度で分類可能であるため信用できる値であると考えられる．

メディアタイプに関して見ると，Audio の平均と Visual の平均に大きな差が出ている事がわかる．この事からコメントは映像に対して付与される傾向があり，映像の印象に関してはコメントからの推定が有効であることが考えられる．しかし，今回の実験では VOCALOID 楽曲のみを扱っているため，初音ミクなどのキャラクターが映像中に登場すると音楽に関係なくキャラクターに関してのコメントが付与される傾向があると考えられる．そのため Visual の精度が高くなっているということも考えられる．

これらの結果から実際に音楽動画の印象を推定する際には，各メディア・印象タイプごとに適した手法を用いることでソーシャルコメントからの音楽動画の印象推定が可能になるのではないかと考えられる．しかし，本研究では推定精度を分類精度によって評価している．そのため，実際にデータセットの拡張を実現するためには，具体的な評価値を推定する手法が必要となるため，今後はその手法についても検討していく．

6. まとめ

本稿では 500 曲，3 メディアタイプ，8 軸の印象評価軸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，音楽動画に対する単語ベクトルを作成し，各メディア・印象タイプごとに SVM を用いて印象推定し，それについて考察した．特に，単語ベクトル生成の際，4 つの品詞やその組み合わせによる手法で生成し，それぞれの手法の比較と各メディア・印象タイプの有効な手法について考察した．その結果，各手法において推定精度に差がでることがわかった．また，全てのメディア・印象タイプにおいて形容詞を含む手法が最も高い値を出すということがわかった．

今後の課題として，データセット拡張のために，具体的な評価値を推定する手法を検討すること，また今回使用した印象評価データセットは，評価者が 3 人と少なく評価値にそもそもブレがあると考えられるため，評価者全員が一定の評価をつけた音楽動画に対しての推定精度について考慮していくことが必要であると考えられる．さらに，今回の研究ではコメントの量が与える影響については調べきれていない．そこで，今後はこうしたコメントの量の影響とそれによりニコニコ動画上でどの程度の音楽動画をデータセット化できるのかについても検討する．

謝辞

本研究の一部は, JST CREST, 明治大学重点研究 A, 重点研究 B の支援を受けたものである.

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ト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).

[13] 土屋駿貴, 中村聡史, 山本岳洋: ソーシャルコメントからの音楽動画印象推定に関する考察, 研究報告会グループウェアとネットワークサービス（GN), 2015-GN96, vol 3, pp.1-6(2015-09-25).