印象に関する検索意図を考慮したサムネイル動画自動生成手法の提案

前島 紘希† 中村 聡史‡ 土屋 駿貴† 大野 直紀†

†明治大学総合数理学部 〒164-8525 東京都中野区中野 4-21-1

E-mail: †{remi.nokotti , bad.ukr.mbr.pr , kas.naoki.0212}@gmail.com, ‡satoshi@snakamura.org

あらまし 動画検索においては様々な方法が考えられるが，その一つが印象に基づく検索である．ここで，動画検索結果のスニペットとして提示されるのはサムネイル画像とタイトル，簡単な説明文程度であり，その動画が自身の意図に沿っているかどうかを判断するには不十分である．サムネイル動画や要約動画を生成する試みはあるが，それは検索意図に応じたものではない．そこで本稿では，音楽動画検索の支援のため，検索クエリ中の印象語に応じたサムネイル動画を，ソーシャルコメントを利用して自動生成する手法を提案する．また，その手法の有効性について簡易的な実験を通じて考察する．

キーワード サムネイル，動画，ソーシャルコメント

1. はじめに

YouTube やニコニコ動画など動画共有サイトが人気を博すようになり，一般ユーザによって創り出される動画の数が飛躍的に増加している．ユーザがこうした動画共有サイト上で視聴する動画を探す場合，動画ランキングから目的のものを探すか，目的の動画に該当しそうなキーワードを入力することで検索したりすることが一般的である．

この動画共有サイト上での動画の検索では，「初音ミク」や「VOCALOID」，「JAZZ」などの固有名詞を用いるだけでなく，「かわいい」「壮大な」「泣ける」などの印象に基づく検索が行われることも少なくない．実際に，ニコニコ動画などでは「泣ける動画」や「涙腺崩壊」のように動画に印象タグをつける試みが多々行われており，機能している．しかし，山本の調査[10]にあるように，そのタグは十分に付与されているわけではない．また，印象語を付与した検索を行った場合に出てくる検索結果の数は多く（例えば 2016 年 2 月時点で，ニコニコ動画で「初音ミク きれい」で検索するとおよそ 1000 件の動画が検索の候補としてでてくる），どれが求めている動画なのかを選別することは容易ではない．特に動画の検索においては，ユーザは視聴対象とする動画をタイトルや動画の説明文などのテキスト情報やサムネイル画像を見て選ぶことが一般的であり，テキスト情報や画像といった動きのない情報からその動画がどのようなものであるかを判断することは困難である．

そうした問題を解決するため，中村らは音楽動画のサビの部分とニコニコ動画において動画につけられたコメントの量やコメントに込められた感情に注目してサムネイル動画を抽出するといった手法を提案してきた[4]．しかし，この手法ではコメントに含まれている感情の種類をひとくくりにしてしまっている．そのため，特定の印象語を用いた検索の際に利用することができないといった問題があった．

そこで本稿では，ユーザの検索クエリに含まれる印象語に応じたサムネイル動画を生成し，ユーザに提示する手法を提案する．ここでは，動画に対して投稿されているソーシャルコメントから動画の各シーンにおける印象を推定し，その印象度合いに応じてシーンの切り出しを行う．

我々の手法によりユーザは検索結果が自分の求めていた動画か，それとも自分が探していたものと違うものなのかという判断がしやすくなり，動画をフルで視聴するか否かを判断する指標になるのではないかと期待される．

2. 関連研究

ダイジェスト動画の生成に着目した研究は多数行われている．石黒らの研究[1]では，ドラマやアニメの次回予告のセリフを利用したダイジェスト動画生成を行っている．しかし，この手法は，次回予告が必須となっているため，動画共有サイト上の動画へ利用することはできない．

小川らの研究[2]では，ニコニコ動画におけるソーシャルコメントが多い箇所を判定し，その箇所のみを再生するといったダイジェストの疑似生成を行っている．しかし，単純にコメントが多いからといってその部分が重要なシーンであるとは限らず，またユーザの検索意図に応じたダイジェストの生成はできない．なお，ソーシャルコメントが少ない動画の際にダイジェスト再生がされないことが問題点として挙げられているが，これについては我々の研究も同様の問題を抱えていると言える．

磯貝らの研究[3]では笑いの意味を含んだコメントである「w」を利用しておもしろい動画を探している人向けのダイジェスト動画の作成アルゴリズムの手法

を提案している．しかし，この研究では「おもしろい」という印象についてにしか焦点を当てておらず，多種多様な印象語に対応できていない．また，「w」が使われているとしても必ずしも面白いとは限らない．

中村らの研究[4]では，楽曲動画のサビ検出技術，感情コメントの数を利用した視聴者の盛り上がり検出技術またはその両方を用いて 15 秒のサムネイル動画の生成を行っている．しかし，この提案手法では感情コメントの種類には着目しておらず，動画検索の際には視聴者の検索意図を考慮できていない．

土屋らの研究[6]では下記で説明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音楽動画のメディアタイプ（映像，音楽，映像と音楽），印象タイプを推定する実験を行っている．本稿では，動画に対する印象によるソーシャルコメント内に出現する単語の違いを調べ，それらを利用し印象ごとのサムネイル動画の自動生成手法について検討していくものである．

ウェブ上の動画を対象としているわけではないが，Miyamori らの研究[11]では，テレビ番組を見ながらチャットをしている人の情報を利用してテレビ番組のシーンのインデックス化およびビューを生成する手法を提案している．この研究はチャットというテキスト情報を扱っており，我々のソーシャルコメントを用いている点と類似しているが，シーン検索かサムネイル動画の生成かという用途が大きく異なっている．

サムネイル動画の生成に着目した研究ではないが，高見らの研究[5]ではウェブ検索の検索意図に応じて検索結果のスニペットを再構築する手法を提案している．この研究は我々のクエリに応じて動画サムネイルを生成するという点と類似しているが，対象とするメディアが異なるためアプローチも大きく異なっている．

3. 印象評価データセットに基づく分析

3.1. シーン毎の印象推定の方針

本稿で目的としている動画の印象に基づくサムネイル動画自動生成においては，動画のシーン毎の印象を推定する必要がある．動画のシーン毎の印象推定においては，音響的な特徴量を使うことや，映像の視覚的な特徴量を使うことも考えられるが，本稿では研究の第一段階として，動画の再生時間に対するソーシャルコメントからの，動画のシーン毎の印象推定を行う．

動画のシーン毎の印象推定を行うには，その動画内での視聴者がどのように印象変化させていったのかという情報が必要となる．ソーシャルコメントが存在する動画に対する印象評価データセットとしては，我々がこれまでに構築した動画全体に対する印象評価データセット[9]と，動画のサビ部分（30 秒）の音楽のみ，映像のみ，音楽と映像の組み合わせに対する印象評価データセット[7]が存在する．シーン毎の印象推定においては，両データセットともに最適とは言いがたいが，動画の 30 秒に対する印象推定が可能になれば，シーン毎の印象推定もある程度可能になると考えられる．そこで，本稿ではまず我々がこれまで構築した 30 秒の動画に対する[7]において構築した印象評価データセットを利用し，その 30 秒分の印象推定可能性を検討する．

なお，本来はすべての印象に対応するような分析を行うべきであるが，本稿では研究の第一段階ということで，後述する印象評価データセットで用いられている 8 つの印象軸についてのみ取り組む．

3.2. 印象評価データセット

この印象評価データセットは，音楽動画のサビ部分(RefraiD[8]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，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．

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，[7]および[9]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．本稿では，この印象評価値の 3 人分の平均を計算し，それぞれの印象タイプに対する評価値とする．

表 1　8 つの印象軸

C1(堂々)	堂々とした，どっしりとした，心踊る，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，気が滅入る，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-3 することによって-2〜+2 に変換した．

3.3. 分析方法

[5]において構築した印象評価データセットに該当する 500 曲の音楽動画のサビ部分（30 秒間）に対して付与されたコメントを収集した．ここで収集されたコメントの総数は 4,780,872 件であった．さらに，8 つの評価軸に対してそれぞれ評価値が 1 以上の動画集合と -1 以下の動画集合を作成する（それぞれ Positive 集合，Negative 集合とする）．

例として C1 という評価軸について考えると C1 の評価値が 1 以上の動画集合と評価値が-1 以下の動画集合を作成する．この操作を 8 つの評価軸に対して行う．その後，Positive 集合と Negative 集合のそれぞれについてすべての単語の DF 値（ここで DF 値とは，動画を 1 つのドキュメントとして捉え，その動画にコメントが投稿されているかで算出）を計算する．なお，単語については形態素解析を行うため，Mecab を用いた．最後に Positive 集合と Negative 集合の DF 値の差をとり，その差の大きさによってどのような単語が出現しやすいかの分析を行った．その際，DF 値の差を「0.2 以上」と「0.1 以上 0.2 未満」の 2 種類に分けて出現する単語の種類の確認を行った．

3.4. 分析結果

分析結果は表 2，表 3 の通りである．表 2 は，Positive 集合と Negative 集合における DF 値の差が「0.2 以上」のものをピックアップしたもの，表 3 は DF 値の差が「0.1 以上 0.2 未満」のものをピックアップしたものである．

表 2　DF 値の差が「0.2 以上」の
8 つの印象軸の特徴的な単語

C1（堂々）	かわいい，www
C2（元気が出る）	かわいい，www
C3（切ない）	かっこいい，綺麗
C4（激しい）	かっこいい，声
C5（滑稽）	www，中毒
C6（かわいい）	かわいい，萌え
Valence	かっこいい，サビ
Arousal	かわいい

表 2 の結果より，C1，C2，C6，Arousal の 4 軸では「かわいい」という単語が，C3，C4，Valence の 3 軸では「かっこいい」という単語が，C5 では「www」という単語が多く表れていることが分かる．

C6 はかわいいという分類であるため「かわいい」という単語が頻出語として出てくるのは問題ないが，C1（堂々），C2（元気が出る），Arousal（積極的）などにおいても「かわいい」という単語が出てきてしまっている．ここで，C1 や C2，Arousal の評価が C6 と一致するのであれば問題ないが，そうでないため「かわいい」という単語がどのような利用のされ方をしているのかが問題となる．また，「かわいい」という単語は，C6 の評価値が-1 以下のものにもそれなりの頻度で登場していた．そこで実際に，コメントを調査してみたところ，動画のお約束としての「かわいい」，単純に初音ミクを利用しているから「かわいい」などのように，「かわいい」という言葉が音楽動画自体の印象とは別のものとして利用されていた．そのため今回のような結果になったと考えられる．つまり，単純に「かわいい」という言葉を利用するだけでは，本来の「かわいい」シーンを抽出することは困難であると考えられる．

一方，C3，C4，Valence の 3 軸において「かっこいい」という単語が頻出している．C4（激しい）においては「かっこいい」という単語が頻出語として出てくるのは問題ないが，C3(切ない) や Valence（楽しい）などにおいても「かっこいい」という単語が出てきてしまっている．ここで，C3 や Valence の評価が C4 と一致するのであれば問題ないが，そうではないため，「かっこいい」という単語がどのような利用のされ方をしているのかが問題となる．そこで実際に，コメントを調査してみたところ，鏡音レンなどのキャラクターの見た目に対しての「かっこいい」，楽曲のストーリー中のキャラクターの性格に対しての「かっこいい」などのように，「かっこいい」という言葉が音楽動画自体の印象とは別のものとして利用されていた．そのために今回のような結果になったと考えられる．つまり，「かわいい」同様，単純に「かっこいい」という言葉を利用するだけでは，本来の「かっこいい」シーンを抽出することは困難であると考えられる．さらに，C3 に現れた「かっこいい」という単語の DF 値の差が C4 や Valence に現れた「かっこいい」という単語の DF 値の差よりも値が低くなっていることも分かった．これは，C3 の「かっこいい」は C4 や Valence の「かっこいい」よりも印象として弱いものとなっているのではないかと考えられる．実際に C3, C4, Valence の Positive 集合の DF 値を確認したところ，C3 の「かっこいい」の DF 値は C4 や Valence における「かっこいい」の DF 値より 0.1 以上少ないことが分かった．さらに，C3 以外の印象では「かわいい」，「かっこいい」，「www」といった単語の別の表現も DF 値の差が「0.2 以上」の特

徴的な単語として多く出現しているのに対して, C3 は「かっこいい」以外の表現については, DF 値の差が 0.2 以上になっているものがなかった．これから表記の違いによって DF 値の差が低くなっているのではないことが分かり，このことからも C3 の「かっこいい」という印象が C4 や Valence の「かっこいい」よりも印象として弱くなっていることが考えられる．

また，C5 のみ「www」という単語が特に頻出しており，この印象軸がほかの印象軸とは特にかけ離れた固有の特徴を持っていることが分かった. C5 の印象軸のみが固有の特徴を持った理由としては,「面白い」という意味の表現のネットスラングは「www」という表現以外ほぼ使われていないためであると考えられる．

表 3　DF 値の差が「0.1 以上 0.2 未満」の 8 つの印象軸の特徴的な単語

C1(堂々)	最高，イラスト，今，アニメ，萌え，明日，嫁，love，元気，愛，歳，調教，胸，頭，投稿，神，幸せ，楽しい，大好き，リン，天使，絶対，友達，流れ，誕生，ミリオン，爽やか，再生，本家，結婚，カラオケ，一番，夏，行く，高い，おめでとう，レン
C2(元気が出る)	萌え，爽やか，歌，絵，アニメ，普通，嫁，love，サビ，胸，投稿，元気，天使，ミク，誕生，弾幕，明日，幸せ，夏，画質，青春，人気，恋，今日，歳
C3(切ない)	鳥肌，調教，声，イケレン，人間
C4(激しい)	リン，ギター，希望，鳥肌，カラオケ，サビ，最高，イラスト，評価，伸びろ，大好き，苦手，ベース
C5(滑稽)	面白い，意味，性，かわいい，嫌，楽しい，センス，シュール，おめでとう，人，癖，怖い，動く，ひどい，市場，カオス，不思議，頭，子
C6(かわいい)	俺，歌，www，アニメ，弾幕，爽やか，ミク，嫁，画質，絵，普通，職人，目，泣ける，誕生，恋，色
Valence	www，ギター，リン，PV，中毒，絵，神，服，理解，流石，配信，動画，相変わらず
Arousal	www，萌え，元気，爽やか，アニメ，普通，ミク

表 3 は，DF 値の差が「0.1 以上 0.2 未満」の特徴的な単語についてまとめたものである．

「かわいい」という単語が頻出語として出てきた C1, C2，C6，Arousal の間には「爽やか」「ミク」「嫁」などのように重複している部分が多く，あまり違いがないことが分かる．これは上述したように「かわいい」には様々な使われ方があるが，これらの「かわいい」に含まれているイメージがほぼ同じになっているということが原因であると考えられる．

一方,「かっこいい」という単語が頻出語として出てきた C3，C4，Valence の間には違いが出てきていることが分かる．例えば，C3 では「調教」,「声」といった VOCALOID のボーカルに関する単語と「綺麗」,「イケレン」といった映像に関する単語が出てきている．また，C4 では,「ギター」,「ベース」といった楽器の名前が多くあることからバンド調の音楽動画が多いことが考えられる．Valence では「www」「中毒」といった C5 に多く出現していたコメントが出現しているが, これは C3，C4 にない特徴であり,「かっこいい」という印象に加えて「面白い」という印象も一緒に持っている音楽動画が多いことが考えられる．

「www」というコメントが頻出していた C5 については「シュール」,「カオス」,「不思議」といったようなほかの印象にほとんど出現していない単語が数多く使用されている．このことからも C5 の印象軸が固有の特徴を持っていることが分かる．

また文書内で出現する頻度を表す TF 値による閾値を利用した分析も行った. TF 値が 0.01 以上かつ DF 値が 0.1 以上となった単語をピックアップしたものを表 4 にまとめる．

表 4　TF 値 0.01 以上かつ DF 値 0.1 以上の 8 つの印象軸の特徴的な単語

C1(堂々)	かわいい，絵，www
C2(元気が出る)	かわいい，絵，歌，www
C3(切ない)	かっこいい，声，調教，神，綺麗，すごい，曲，素敵，絵，歌，レン，鳥肌
C4(激しい)	かっこいい，最高，声，やばい，リン，サビ，大好き，鳥肌
C5(滑稽)	www，かわいい，中毒
C6(かわいい)	かわいい，ミク，絵，歌
Valence	かっこいい，www，サビ，最高，リン，ルカ，絵，やばい
Arousal	かわいい，www

表 2 と表 4 を比較すると表 2 にある単語はどれも表 4 に出てきており，DF 値「0.2 以上」の単語は文書内での出現率も高いことが分かった．また，表 3 と表 4 を比較してみると,「かわいい」が頻出語として出てきた C1，C2，C6，Arousal についてはどれも出現する単語の種類が激減していることが分かった．また,「かっこいい」が頻出語として出現している C3, C4，Valence ではそれぞれに特徴が出ていた．C3 は「神」「曲」「素敵」のように表 3 で表れていた単語ではない単語もいくつか出ていた．C4 に関しては，表

3 では「ギター」「ベース」といった楽器に関する単語が特徴的な単語となっていたが，TF 値を利用した表 4 では出現しなくなっている．Valence では表 3 と表 4 で出現する単語が大きく変わっている．「www」が頻出語として出現している C5 については表 3 で出現していた単語は「かわいい」しか残っておらず，繰り返し用いられる単語はほとんど存在しないということが分かった．

以上のことより，特に頻出する語句である程度の印象カテゴリを推定し，それ以外に出てくる語でその中でもどのような印象に分類されるのかということを推定する方法が考えられる．

4. サムネイル動画

4.1. クエリに応じたサムネイル動画生成

本稿で提案するサムネイル動画自動生成手法は，音楽動画検索においてユーザが入力したクエリの中に印象語が含まれている場合に，その印象語を該当部分切り出しに利用する．ここで印象語については，表 1 に示す印象軸で説明されている語句のみとした．なお，サムネイル動画の長さについては，一般的なテレビ CM と同じ 15 秒に設定した．

システムは，まずユーザの入力したクエリに応じて動画集合を限定する．次に，印象語がクエリに含まれている場合に，表 1 を元にした辞書を用いてユーザがどの印象軸を求めているかを判定する．また，その印象語に最も適切である 15 秒を抽出し，ユーザにサムネイル動画として検索結果とともに提示するものとなっている．

なお，毎回上記の計算を行うのは無駄であるため，事前に用意した 8 つの印象軸についてそれぞれの動画で適している 15 秒を計算し，データベースに格納しておき，そのデータベースから呼び出すことでサムネイル動画をユーザに返す．

今回，私たちは形態素解析を用いた手法とトリグラムを用いた手法の 2 種類の手法でサムネイル動画を生成した．

形態素解析を用いた手法では，まず表 2，表 3 にある単語が入っているコメントを抽出し，それらをコメントが行われた時間軸で並べる．その後，一番コメントの量が多かった，連続した 15 秒を抽出する．さらに抽出した 15 秒内のすべてのコメントの中での表 2，表 3 にある単語が含まれるコメントの割合を計算し，その値が 0.25 という閾値を超えていた場合のみサムネイル動画として抽出するといった手法を用いている．これにより 8 つの印象軸のそれぞれに対して印象が高い動画のみに対してサムネイル動画を生成している．

トリグラムを用いた手法では，まず，すべてのコメントを 3 文字ずつに区切る．その中で印象軸ごとに出現頻度が高い 3 文字の塊を収集する．その後，形態素解析を用いた手法と同様に，印象軸ごとに収集された 3 文字の塊を含むコメントを抽出し，コメントがされた時間軸で並べ，一番コメントが多かった 15 秒を抽出する．さらに抽出したすべてのコメントの中で収集された 3 文字の塊が含まれるコメントの割合を計算し，0.2 という閾値を超えていた場合のみサムネイル動画として抽出するといった手法を用いている．

4.2. 生成されたサムネイル動画の特性

提案する印象に基づくサムネイル動画自動生成手法の特性を検証するため，各印象について生成されたサムネイル動画を視聴し，評価を行った．ただし，ここでは，被験者ベースの実験ではなく，著者が判断してどうだったかということを分析する事前調査にとどめた．

最初に形態素解析を用いた手法，トリグラムを用いた手法のそれぞれの手法により生成されたサムネイル動画の特性について考察し，その後，形態素解析を用いた手法，トリグラムを用いた手法の両方の手法について著者の評価を-2~+2 の 5 段階評価で評価し，どちらの手法が良いサムネイル動画が生成できていたかを判断する．

まず，形態素解析を用いた手法についてだが，「かわいい」が頻出語であった C1，C2，C6，Arousal は同一の動画でサムネイル動画が生成されることが多く，同じ場所を抽出してしまう動画が多く存在した．これを解消するには印象評価データセットの楽曲数を増やし，特徴的な単語のさらに深い分析をする必要がある．「かっこいい」が頻出語であった C3，C4，Valence については同一の動画でサムネイル動画が生成されることは多くなく，分析であった通り，同じ「かっこいい」という単語でも違った使い方がされていることが分かった．サムネイル動画の内容としては，C1，C2，C3，C4，C6 として抽出されたサムネイル動画は表 1 にある印象を受ける動画が多く出現していた．しかし，C5，Valence，Arousal として抽出されたサムネイル動画は表 1 にある印象とは違ったものが少なからず出現していた．それぞれの理由について考察していくと C5 については「www」というコメントは「面白い」とあまり感じない場合でも「とりあえずコメントの最後に入れる」といった形で使用されることがあることが原因だと考えられる．この問題を解消するには「www」という文字以外が入っていないコメントのみを利用するといった方法が考えられるため，今後の研究の際には

表 5　形態素解析，トリグラムのそれぞれの手法で生成されたサムネイル動画の評価

	C1 堂々	C2 元気が出る	C3 切ない	C4 激しい	C5 滑稽	C6 かわいい	Valence 楽しい	Arousal 積極的	平均
形態素解析	0.20	0.37	0.35	0.30	-0.17	1.04	0.36	0.32	0.35
トリグラム	0.20	0.17	-0.08	0.45	-0.16	0.56	0.12	-0.08	0.15
平均	0.20	0.27	0.14	0.38	-0.16	0.80	0.24	0.12	0.25

この方法も試していく必要がある. Valence, Arousal につい
ては C1~C6 と評価の方法が違ったことが原因と考えられる．印象評価データセットの項でも触れたが，C1~C6 はその印象に当てはまっているかどうかで 1~5 の 5 段階評価を行っているが，Valence，Arousal については「明るい気持ちになる」と「暗い気持ちになる」，「激しい」と「穏やか」のような真逆の印象に対してどちらの印象に近いかを-2~+2 の 5 段階で評価している．Valence，Arousal の評価方法だと「激しい」と「穏やか」で比べると「激しい」に近いといった考え方で評価が高くなってしまうことが考えられる．評価軸の設定を変えることで表 1 のような印象に近づく可能性があるためデータセットを再構築していく必要があると考えられる.

次にトリグラムを用いた手法についてだが，こちらの手法でも「かわいい」が頻出語であった C1, C2, C6, Arousal は同一の動画でサムネイル動画が生成されることが多かった. サムネイル動画の内容は C1, C2, C4, C6 は形態素解析を用いた手法と同じように表 1 にある印象を受けるサムネイル動画が多く生成されていた. C5, Arousal も形態素解析を用いた手法と同じで表 1 の印象と違ったサムネイル動画が存在していた．形態素解析を用いた手法と大きく違っていた印象軸が C3, Valence の 2 軸であった．C3 については形態素解析を用いた手法では表 1 の印象にあったサムネイル動画が生成されていたが，トリグラムを用いた手法ではこの印象から少し外れた印象のサムネイル動画が生成されていた．この理由としては，形態素解析を用いた分析を行った際には, 声に関する単語が多く表れていたが, トリグラムの場合，声に関する 3 文字の塊が出現していなかったことが挙げられる. C3 において出現頻度の高かった 3 文字の塊としては「きれい」「イケレ」などといった分析において C3 の印象軸で表れていた単語の一部も出ていたが「ギター」といった C3 の印象軸では表れていなかった単語の一部が多く出現していたためだと考えられる．Valence については C3 とは逆に形態素解析を用いた手法では表 1 の印象から外れた動画が生成されていたが，トリグラムを用いた手法では印象に近い動画が多く生成された. この理由としては，形態素解析では「かっこいい」という単語が多く使われていたが，これは Valence の印象である「明るい気持ちになる」「楽しい」というものと違いがあったと考えられる．一方，トリグラムを用いた手法では,「カッコ」「www」といった Valence で頻出していた単語の一部の塊のほかに,「可愛い」「センス」といったような「かわいい」「www」が頻出語として出現していた印象軸の単語の一部のような塊が出現していた．これにより「明るい気持ちになる」「楽しい」といった印象に近づいたのではないかと考えられる.

続いて，著者が形態素解析を用いた手法，トリグラムを用いた手法について-2~+2 の 5 段階で評価したものを表 5 に示す. この表から分かるように形態素解析，トリグラムの両方の手法において C6 の評価が高くなった．また，形態素解析，トリグラムのそれぞれの評価値の平均を算出した結果，形態素解析を用いた手法のほうがサムネイル動画に対する評価が高くなった．しかし, C4 についてのみ形態素解析を用いた手法よりもトリグラムを用いた手法のほうが評価値は高くなった.

5. まとめ

本稿では，500 曲，8 軸の印象評価からなる印象評価データセットを用い，それぞれの印象に対して頻出する単語の分析を行った. その結果, C1, C2, C6, Arousal からなる「かわいい」が頻出するグループ，C3，C4, Valence からなる「かっこいい」が頻出するグループ, C5 からなる「www」が頻出するグループの 3 つのグループに大きく分けられることが分かった.さらに，C1, C2，C6，Arousal の 4 軸については出現する単語にはあまり違いが見られなかったが，C3，C4，Valence の 3 軸については違いが見られた．このことから特に頻出する単語を用いてある程度の印象推定を行い，それ以外に出てくる語でさらに印象を絞り込んでいくという方法が有効ではないかと考えられる．また，形態素解析を用いた手法とトリグラムを用いた手法の 2 種類のサムネイル動画の生成手法を提案し，事前調査を実施した.

今後の課題は，まず今回の評価は著者の主観によって行われたものであり，サムネイル動画が多くのユーザの検索の助けになるかどうかの判断には不十分である．そこで多数の被験者に実際にサムネイル動画を視聴してもらい，有用性を検証する評価実験を行う必要がある．

また，今回の研究では，単語間の共起関係に関しては十分に調査・分析できていない．頻出語との組み合わせによって印象の絞り込みが行えると考えられるので今後調査する必要がある．また，今回の研究ではサビのみに限定していたが，音楽動画によってはサビ以外の部分に印象語が多く出現する動画も存在すると考えられる．また，サビとその他の部分で印象が異なる音楽動画も存在すると考えられる．したがって，今回の結果と違う結果が出てくることも考えられるので調査の必要がある．

今回,サムネイル動画の長さを 15 秒と設定したが，この 15 秒という時間をさらに短くすることで内容がより伝わりやすくなる動画の存在も考えられる．そのような動画の調査と，サムネイル動画の長さを固定の長さではなく可変の長さにする手法の考案についても今後行っていく必要がある．

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] 石黒信啓，白井治彦，黒岩丈介，小高知宏，小倉久和:文章要約の手法を用いたダイジェスト動画の製作手法，情報処理学会創立 50 周年記念(第 72 回)全国大会(5W-1),pp.491-492(2010)

[2] 小川一昭，服部哲，速水治夫:視聴者からのコメント情報を用いたダイジェスト動画疑似生成方法の提案，情報処理学会研究報告(2009-GN-71)，pp.146-150(2009)

[3] 磯貝佳輝，齋藤義仰，村山優子:視聴者コメントを用いた動画検索支援のための紹介動画作成手法の提案，情報処理学会論文誌　コンシューマ・デバイス＆システム,Vol.2 No.1 pp74-81(2012)

[4] 中村聡史，山本岳洋，後藤真孝，濱崎雅弘:視聴者反応と音楽的特徴量を用いたサムネイル動画の自動生成，WebDB Forum 2012

[5] 高見真也，田中克己:ウェブ検索結果に応じたスニペット生成，情報処理学会論文誌，Vol.49 No.4 pp1648-1656(2008)

[6] 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検討，情報処理学会論文誌

[7] 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).

[8] 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)

[9] 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).

[10] 山本岳洋，中村聡史:視聴者の時刻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，Vol6, No3, pp61-72(2013)

[11] Miyamori，H.，Nakamura，S.，and Tanaka，K:Generation of Views of TV Content Using TV Viewers' Perspectives Expressed in Live Chats on the Web，In Proc. of ACM Multimedia2005，pp853-861(2005)