

# 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<sup>†</sup> 中村 聡史<sup>‡</sup> 土屋 駿貴<sup>†</sup> 大野 直紀<sup>†</sup>

<sup>†</sup> <sup>‡</sup> 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: <sup>†</sup> {ev30658, ev30616, ev30508}@meiji.ac.jp <sup>‡</sup> 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。検索ユーザを支援するためサムネイル動画や要約動画を生成する試みはあるが、これまでに実現されている手法は検索意図に応じたものではない。そこで本稿では、音楽動画検索クエリ中の印象語を検索ユーザの意図として考え、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用いて自動生成する手法を提案する。また、自動生成されたサムネイル動画を用いた評価実験によって、印象によってサムネイル動画を生成するのに適している手法は異なっていることを明らかにした。

**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したり、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印象語が含まれる場合、その印象語をユーザの意図であると考え、印象語に応じたサムネイル動画を生成してユーザに提示する手法を提案する。ここでは、動画のどのシーンがその印象語に適しているかどうかを、動画に対して投稿されているソーシャルコメントから判断する。そして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、動画を最初から最後ま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、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」という印象には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、音楽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することが出来ない。

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、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。

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サムネイル動画に着目した研究でないうえ、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。

ソーシャルコメントと印象の関係性を明らかにする研究もいくつかなされている。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）に対する印象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。

### 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を実現するためには、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、本稿ではユーザの検索意図に沿ったサムネイル動画自動生成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を用いて、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、その動画内で視聴者の受ける印象がどのように変化していったのかという情報が必要となる。ここで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いえないが、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 秒分の印象推定の可能性を検討する。

なお、本来は考えるすべての印象に対応するような分析を行う必要があるが、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

本稿で用いる印象評価データセットは、音楽動画のサビ部分(RefrainD[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。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。

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、陽気な、心地よい、楽しい気持ちにさせる
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、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

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いた、再生数が多い上位500個を抽出したのとなっている。

本稿では、この印象評価値の3人分の平均を計算し、それぞれの印象軸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段階評価をされていた。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

各印象でどのようなコメントが付与されているかを明らかにするために、[7]において構築した印象評価データセットに該当する500個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。また、明確な印象の差による分析を行うため、8つの印象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した(それぞれPositive集合、Negative集合とする)。

例としてC1という印象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つの印象軸に対して行う。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、その動画に該当する単語がコメントの一部として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で、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。

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

C6(かわいい)に「かわいい」という単語が頻出しているのは問題ないが、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。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、動画に投稿されている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。そこで、「かっこいい」というコメントがどのように利用されているか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。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。

ここで、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。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る。このことから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。

C5(滑稽な)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, 伸びる, 大好き, 苦手, ベース
C5(滑稽)	面白い, 意味, 性, かawaii, 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、表2と異なり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、印象ごとに使用されている単語の違いが出てきていることが分かる。例えば、C3では「調教」や「声」といったVOCALOIDのボーカルに関する単語と、「綺麗」や「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」や「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」や「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」や「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(ここではある動画に対して投稿されたコメント集合に含まれる単語の比率)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」の単語はそもそもTF値が高いことが分かる。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推定する方法が考えられる。

#### 4. 検索意図に応じたサムネイル動画生成

本稿では、ユーザの検索意図は検索クエリに埋め込まれた印象語に特にあらわれると考え（固有名詞やジャンルなどの名詞は最低限の絞り込みのための語として考える）、サムネイル動画自動生成においてはその印象語に対応した該当部分を切り出して利用する。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、サムネイル動画の生成においては、先述の形態素解析を用いた手法に加え、ニコニコ動画上のコメントが形態素解析に適していない可能性を考慮してトリグラムを用いた手法を用意した。

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より表れている動画のみに対してサムネイル動画を生成している。

一方、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、コメントされた時間で並べ、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。

#### 5. 評価実験

提案する印象に基づくサムネイル動画自動生成手法の有用性を検証するため、ユーザベースでの評価実験を行った。

##### 5.1. 実験方法

実験ではまず、形態素解析およびトリグラムによる2手法を用い、8つの印象値が高いものについて、3.2節の印象評価データセットで用いた500個の動画に対してサムネイル動画を生成した。この生成されたサムネイル動画について、8つの印象、2つの生成手法のそれぞれに対して5個ずつランダムに提示した。

評価者には、提示されるサムネイル動画に対し、8つの印象軸すべてについて-2~+2の5段階で評価してもらった。すべての印象について評価してもらった理由は、その提示されている印象に影響を受けないようにするためである。なお、評価者は20代の男子大学生4名である。

##### 5.2. 実験結果

評価者による印象評価値が、目的とする印象をどの程度推定できていたのかの平均をまとめたものが表5である。

形態素解析を用いた手法ではC6が高い値を示しており形態素解析が効果的に働いていることがわかるが、ほかの軸に関してはあまり高い値が得られなかった。一方、Valence, Arousalについては評価値が0を下回っておりうまくいっていないことがわかる。

トリグラムを用いた手法では、C1, C2, C4, C5, Valenceが高い値を示した一方で、C3, C6, Arousalは評価値が0を下回っていた。

次に形態素解析を用いた手法とトリグラムを用いた手法を比べると、C3, Arousal以外の軸で形態素解析とトリグラムで評価値に大きな差があることが分かる。その中でもC6やValenceのようにどちらかの手法では0を超える評価、もう片方の手法では0を下回る評

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	0.10	0.05	0.05	0.25	0	0.80	-0.35	-0.15	0.09
トリグラム	0.60	0.50	-0.25	0.80	0.55	-0.70	0.40	-0.10	0.23

価となった軸があり，これらの軸については 0 を超える手法が特にサムネイル動画の抽出に適した手法であると考えられる。

### 5.3. 考察

形態素解析を用いた手法において，C1，C2，C6，Arousal の 4 軸で「かわいい」という単語が頻出していたにも関わらず C6 でのみ値が高くなっていた。この理由としては「かわいい」という単語がそもそも C6 の印象軸を表す語として用いられているために C1，C2，Arousal の 3 軸よりも印象がコメントに出てきやすいからであると考えられる。

形態素解析による手法で Valence，Arousal の評価値が 0 を下回っていた理由としては，頻出語が印象に与える影響が少ないということが考えられる。Valence については，「PV」「理解」といったようにどの印象にも影響を与えることがなさそうな単語が多く出現しており，Arousal については，「萌え」「元気」といったような「激しい」とは少し違う印象を与えるような語が多く出現していた。これらの語が多い箇所がサムネイル動画抽出の箇所選ばれてしまったために評価値が低くなってしまったのではないかと考えられる。

トリグラムを用いた手法においては，C1，C2，C4，C5，Valence が高評価となっていた。この理由としては，C1，C2 の軸では，頻出する単語が 1 文字や 2 文字のものが多く，同じ「かわいい」が頻出していた C6 や Arousal よりも頻出する単語が表れやすかったことが考えられる。C4，Valence に関しては「かっこいい」という単語の一部が多く出現しており，また C4 に関しては「ロック」という音楽のジャンルに関する語が，Valence に関しては「www」という語が多く出現しており，この部分が抽出され高い評価が出たのではないかと考えられる。C5 に関しては「www」という語がほかに比べ圧倒的に多く抽出されており，「滑稽」という印象とあっていただけにうまく抽出できたと考えられる。一方，C3，C6，Arousal が低評価となっていたが，この理由としては，トリグラムにおいてこのような 3 文字の塊がうまく抽出されておらず，印象にあった箇所がうまく抽出されなかったのではないかと考えられる。C6 に関しては，「かわいい」という単語の表記の方法として平仮名やカタカナ，漢字などといった違いや「か

わゆい」といったような表記が混ざってしまい抽出の箇所がぶれてしまったということが考えられる。Arousal に関しては C6 の軸と同じ理由に加えて，頻出していた単語がトリグラムで抽出した際にほとんど現れなくなっていたことが原因と考えられる。

形態素解析を用いた手法とトリグラムを用いた手法の比較から，それぞれの軸でサムネイル動画の生成に適している手法が違っているため，ユーザの求める印象に応じて切り替えることが重要であると考えられる。

今回の実験で抽出したサムネイル動画のうち，同一の動画内で複数の印象軸でもサムネイル動画がうまく生成できた例が多数あった。ここでは，C1，C2，C6，Arousal からなる「かわいい」が特徴的なグループ，C3，C4，Valence からなる「かっこいい」が特徴的なグループ，C5 からなる「www」が頻出するグループのうち，異なるグループに属する軸に対して抽出する場合にうまく抽出できていることが多かった（例えば C1 と C5 についてはうまく抽出できることが多く，C1 と C6 についてはうまく抽出できることが少ない）。この理由としては，「かわいい」「かっこいい」「www」の 3 つの単語は印象の方向性が全く違うものであるということが考えられる。同一の動画内で印象が大きく変化するものでは「かわいい」とコメントされる箇所，「かっこいい」とコメントされる箇所，「www」とコメントされる箇所がはっきりと分かれていることが多いため，印象に適した箇所がそれぞれの軸で生成できたのではないかと考えられる。

一方，著者の主観では動画内から他にサムネイル動画として適している箇所があるにもかかわらず，適していない箇所を抽出してしまっている例もあり，これは C5（滑稽）の軸での抽出が大半であった。この理由としては，「www」という語が関係のないコメントの後ろにとりあえず付与するという感覚で用いられることが原因として考えられる。特に動画の序盤には「うぼつ」といったような，どの動画にも同じように用いられる決まり文句のようなコメントがあり，そのようなコメントに「www」というコメントが用いられてしまうことで本来抽出すべき箇所が抽出できなくなっていると考えられる。また，C5 の印象にあっていない箇所に付与されるコメントは「www」という語が単体で用い

