視聴者の時刻同期コメントを用いた動画の特徴シーンの推定

末永　智彦†　　早川　智一††　　疋田　輝雄††

† 明治大学大学院理工学研究科基礎理工学専攻　〒 214–8571 神奈川県川崎市多摩区東三田 1-1-1
†† 明治大学理工学部　〒 214–8571 神奈川県川崎市多摩区東三田 1-1-1
E-mail: †{suenaga,t_haya,hikita}@cs.meiji.ac.jp

あらまし　本論文では，動画共有サービスに投稿された動画内容を効率的に把握するため，動画内の特徴的な盛り上がりを見せるシーンを発見し，それがどのようなシーンなのかを表す特徴語を推定する手法を提案する．実際に動画を見た視聴者の投稿したコメントは動画内容を反映していると仮定し，シーンの発見と特徴語の推定に用いる．例えばサッカーの試合の動画では，スーパープレーやゴールがあったシーンに「すげええ」や「うおおおお」といったコメントが集中する．提案手法では，ニコニコ動画に投稿された動画を対象に，機械学習で作成したネガポジ分類器によりコメントの感情を推定し，各感情のコメントが集中しているシーンを特徴的な盛り上がりのシーンとし，そのシーン内の特徴的なコメントを特徴語とした．提案手法の推定結果と実際に動画を見た評価者の回答との比較評価を行った．

キーワード　動画共有サービス，自然言語処理，動画要約

1. はじめに

近年，動画共有サービスの普及と発展に伴い，ユーザが動画共有サービスの利用に使う時間が大きくなっている．また，動画を作成し Web へアップロードすることが容易となり，より多くの人が動画共有サービスに動画を投稿できるようになったことで，Web 上の動画の数も増加している．Youtube(注1)に 1 分間で投稿される動画の総再生時間数は，2017 年時点で 400 時間 [1] とされ，comScore(注2)によると，日本人 1 ユーザあたりの 1ヶ月間の動画共有サービス利用分数は，2015 年 11 月時点で 2747 分を記録した [2].

これを受けて，ユーザが効率的に動画共有サービスを利用し，目的の動画をより短時間で視聴できる手法が必要とされていると我々は考える．

本研究では，ユーザの動画視聴時間を削減することによる効率化を目的とし，そのために動画の特徴シーンを用いる．ここで，特徴シーンとは，動画内で特徴的な盛り上がりを見せる重要なシーンを意味する．目的の動画における特徴シーンの再生時間帯と，それがどういうシーンなのか表す言葉をユーザに提示することで，ユーザが動画の内容を効率的に把握できるようになり，視聴時間の削減につながると我々は考える．

しかし，特徴シーンの推定には，動画の特徴シーンがどこにあるか（動画がどこで盛り上がるか）は実際にユーザが動画を見てみなければわからない——という課題がある．

この課題を解決するため，動画に付随する情報を用いて自動的に動画の特徴シーンの再生時刻とそのシーンを表す特徴語とを推定する手法を提案する．動画のどの再生時刻が特徴的で，そこはどういうシーンなのかをユーザが動画を視聴する前に提示することで，ユーザは視聴するシーンを取捨選択でき，視聴時間の削減に繋がると我々は考える．

本研究では特に，動画に付随する情報として時刻同期コメント [3] を利用する．ここで，時刻同期コメントとは，動画の再生時刻に同期して投稿されたコメントのことを指す．時刻同期コメントを投稿できる点に着目し，日本の動画共有サービスであるニコニコ動画 (注3)の動画を対象として，提案手法のプロトタイプの実装および実験と評価を行う．

時刻同期コメントは，「作品全体に対する通常のコメントに比べ，時刻同期コメントはその内容が質的に異なり，その瞬間ごとの感情をより多く表現する傾向がある」 [3] と指摘されており，これを用いることで動画の特定のシーンの盛り上がりやその内容が推定できると考える．

特に，本研究ではコメントのネガポジ感情に注目する．例えばサッカーの試合を映した動画では，スーパープレーがあったシーンやゴールシーンがあった再生時間帯に対して，「すげええ」や「うおおおお」といったコメントが集中して投稿される．この場合，それらのシーンは多くのユーザが興奮した，つまりポジティブに盛り上がったシーンと言える．コメントのネガポジ感情推定は，サポートベクターマシン（SVM）を用いた教師付き学習によるコメントの感情分類により行った．

また，あるシーンの中で特徴的なコメントはそのシーンの内容を表していると考えられるため，それをそのシーンの特徴コメントとして利用する．例えば「すごいゴール」といったコメントが，動画内のあるシーンに集中して現れるなら，「すごいゴール」がそのシーンがどのような内容なのかを表していると言える．

我々の提案手法の特徴は次の 3 点である：

（1） SVM を用いた教師付き学習により，ニコニコ動画のコメントをネガティブ・ポジティブ・ニュートラルの 3 感情に分類した．

(注1)：https://youtube.com
(注2)：https://www.comscore.com
(注3)：http://www.nicovideo.jp/

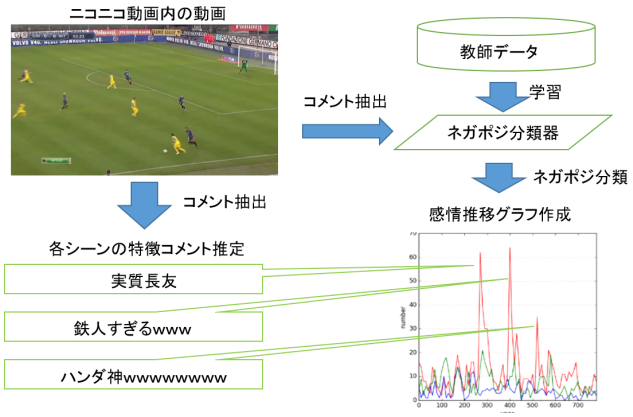

図 1　提案手法の概要図

（2） 動画の特徴シーン推定を時刻同期コメントの時間的特性と（1）で分類した感情とを組み合わせて行った．

（3） 動画のあるシーン内のコメントから，そのシーンの内容を表す特徴コメントを推定した．

提案手法の概要図を図 1 に示す．提案手法のプロトタイプの実行結果をユーザが動画を視聴する前から得ることで，どのシーンを視聴すれば良いかがわかり，視聴時間の削減に繋がると考える．

今回作成した提案手法のプロトタイプは，次の 2 点をユーザに提供する：

（1） コメントの感情や集中度合いから動画の感情ごとの盛り上がり推移をグラフ化して特徴シーンを推定し，；

（2） 特徴シーン中のコメントから，そのシーンの内容を表す特徴コメントを推定する．

提案手法のプロトタイプを用いた評価実験では，特徴シーン推定の正答率は 84 %だった．評価実験は，ニコニコ動画に投稿されている動画に対しての実行結果と，実際に動画を視聴した人の回答とを比較して行った．

本稿の構成は以下の通りである．2 章では関連研究について述べる．3 章では本研究で対象とするサービスについて述べる．4 章では提案手法について具体的に述べる．5 章では評価実験について述べる．6 章で本稿をまとめ，今後の展望について述べる．

2. 関 連 研 究

本章では，動画に対するユーザのコメントを用いた研究を本研究に近いものとして挙げる．動画を対象とした研究は，動画の検索（2.1 節）と動画分析（2.2 節）とに分けて考えられるので，各節でそれぞれについて述べる．

2.1 動 画 検 索

我々の研究は動画要約や内容の分析を行うものであるため，動画検索を行う研究とは異なるが，時刻同期コメントを用いた研究は，動画検索の分野でも多く行われている．

佃ら [4] [5] は，ニコニコ動画の時刻同期コメントを用いることで，通常の動画のタイトルやタグに基づいたキーワード検索ではない，特定の人物の動画内での活躍度合いに注目した動画検索手法 [4] と，ユーザが自分好みの動画のランキングを生成できるシステム [5] といった，よりユーザの意図を反映した検索が行える手法を提案した．コメントを視聴者の反応として扱う点は我々と同じだが，これらの研究では動画全体の内容や特定の人物の活躍を知るためにコメントを用いており，我々は動画の特定のシーンの内容や盛り上がりの推移を知るために用いているという違いがある．

磯貝ら [6] や齋藤ら [7] は，視聴者からのコメントを元に動画の短時間なダイジェスト動画を作成し，それをユーザに見せることでその動画を視聴するかどうかの意思決定を助ける手法を提案した．これらの研究は実質的に動画要約を行っており，重要なシーンの推定に視聴者からの時刻同期コメントを用いるという我々と同じアプローチを取っている．しかし，盛り上がり度合いとして用いているのは単純なコメント数や，インターネット上で「笑い」を意味するスラングである「w」のついたコメント数であり，我々はコメントをネガティブ・ポジティブ・ニュートラルといった感情に分類して扱うため違いがある．

2.2 動 画 分 析

時刻同期コメントを用いた動画要約の研究は，我々の研究に最も近いものとして挙げられる．山内ら [8] は，視聴者の観点の入れ替わりに基づいた特徴的シーン抽出を行った．彼らの研究では，動画内で視聴者のポジティブ・ネガティブ感情が入れ替わるタイミングにはそのきっかけとなる出来事があると考え，そこを特徴的なシーンだと仮定している．視聴者の感情推定には時刻同期コメントを用いており，その点は我々と同じである．だが，コメントの感情推定にはコメント内の単語と単語の感情極性の評価表現辞書とを用いており，SVM を用いて感情推定を行っている我々とは手法が異なる．我々も以前から本研究と同様の目的の研究 [9] を行ってきた．これも SVM を用いたコメントの感情分類を行っているが，本研究では分類時のコメントに対する処理に変更を加えていること，新たにあるシーンの特徴コメントの推定を行っていることに違いがある．

動画に付随するコメントだけでなく，音声や画像といった動画自体の情報から動画要約を行う研究も多く行われている．中村ら [10] は，音楽を扱う動画に対して，音響特徴量とコメントの感情とを組み合わせたサビ部分検出を行うことでサムネイル画像を生成する手法を提案した．また松原ら [11] は，画像特徴量の変化に基づいて動画を複数の区間に分割し，それぞれの区間中で最もコメントが盛り上がっているシーンを用いた動画の支持的ようやくサムネイルを生成する手法を提案した．これらの研究でも我々と同様にコメントをいくつかのクラスに分類して用いているが，辞書やあらかじめ決めた単語とのパターンマッチングで分類をしているため，SVM を用いている我々とは手法が異なる．

3. 対象とするサービス

3.1 ニコニコ動画

ニコニコ動画は，2018 年 1 月時点で投稿動画数 1,500 万件を

図 2　特徴的なシーンの例

超える，ドワンゴ [注4] の運営する動画共有サービスである．ニコニコ動画は動画に対するコメント機能が他の動画共有サービスと異なっており，動画の再生時刻に同期したコメントを投稿できる．Youtube などの一般的な動画共有サービスでは，ユーザのコメントは動画の枠外にリストで並べられ，ユーザ自身がコメント内に動画の再生時刻を記入するなどしなければ，それらは動画全体へのコメントとして見られる．これに対しニコニコ動画では，ユーザの全てのコメントが動画の特定の再生時刻に対して投稿される．ユーザが動画を視聴すると，あるコメントが投稿された再生時刻になると画面上にそのコメントが流れる．

このため，動画内の特徴的シーンにはコメントが集中する（図 2[注5] は，野球の試合で決勝点が入った瞬間のシーン）．多くの人がコメントを投稿したくなるような出来事があったシーンには，その出来事に関するコメントが大量に投稿される．

本研究では，この点に着目し，時刻同期コメントを動画の特定のシーンに対する視聴者の反応として扱う．具体的には，コメントの感情をそのシーンを見ている視聴者の感情とみなし，動画の感情推移を推定する．また，あるシーンに特徴的なコメントはそのシーンの内容を表していると仮定し，特徴コメントとする．

3.2　ニコニコデータセット

本研究では，動画に投稿されたコメントのデータセットとして，国立情報学研究所により提供されているニコニコデータセット [注6] を用いる．ニコニコデータセットは，ニコニコ動画のサービス開始当初から 2016 年 8 月 31 日までに投稿された約 1,400 万件のメタデータと，それらに対する合計約 35 億件のコメントデータを含むデータセットである．動画のメタデータについては，動画に付いたタグと動画のコメント数とを使用する．コメントデータについては，時刻同期コメントとして扱うため，コメント本文とコメントが投稿された動画上の再生時刻情報とを使用する．

本研究では，データセットの中でもコメント数が 1000 件以上ある動画を対象とする．コメントの集中度合いにから特徴シーンを推定するには，ある程度以上多くのコメントが投稿されている必要があると考えたためである．

(注4)：http://dwango.co.jp/
(注5)：http://www.nicovideo.jp/watch/sm6535703
(注6)：https://www.nii.ac.jp/dsc/idr/nico/nico.html

4.　提 案 手 法

提案手法では，動画内の時刻同期コメントを感情分類することによって得られた動画の感情推移をグラフ化したものと，各シーンにおける特徴コメントとをユーザに提示する．感情推移グラフにおいて山となっているシーンが動画の特徴シーンと言える．また，ここで特徴コメントとは，あるシーンの内容を最もよく表しているコメントのことを指すものとする．コメントの感情分類には SVM を用いる．コメントの素性をベクトル表現として SVM へ入力するために，形態素解析を行い，コメントを Bag-of-Words モデルに変換する．

提案手法の具体的な流れは，以下の通りである：

（1）　コメントに正規化処理を施す（4.1 節）．
（2）　コメントを Bag-of-Words モデルに変換する（4.2 節）．
（3）　SVM によるコメントの感情分類を行い，それぞれの感情のコメント数推移から特徴シーンを推定し，グラフ化する（4.3 節）．
（4）　特徴シーン中のコメントから，シーンの内容をよく表す特徴コメントを推定する（4.4 節）．

4.1　コメントの処理

コメントに対する前処理として，正規化を行う．提案手法ではコメントを形態素解析して扱うが，ニコニコ動画に投稿されるコメントは基本的に短文であり単語の表記揺れも多く，そのまま扱うことが難しいためである．このような表記揺れが起きている同じ意味の単語をそのまま別々の単語として扱うと，特徴シーンの推定や特徴コメントの推定において各単語の数がまばらになり，結果が不正確になる可能性がある．

提案手法では，佃ら [5] の手法に倣い，以下の手順でコメントを正規化する．

（1）　スペース及び「!」と「?」以外の記号文字（「+」「#」など）を除去する．
（2）　半角の片仮名と英数字，記号とを，全角に変換する．
（3）　平仮名及び片仮名の小書き文字（「ぁ」「ァ」「っ」「ッ」など）を大文字に変換する．
（4）　英語の小文字を大文字に変換する．

以上（1）から（4）までの処理により，例えば「すげぇえええええ!!!」というコメントは「すげえええええええ !!! 」という文字列に正規化される．佃ら [5] の手法は Brody ら [15] の手法に倣い，同じ文字が 2 回以上繰り返して記述されている場合に繰り返し回数を 1 回にするという処理を行っていたが，我々はこれを行わない．4.2 節で述べるコメントの Bag-of-Words モデル化に関係するためである．

4.2　コメントの Bag-of-Words モデル化

コメントを感情分類のための SVM への入力や，特徴コメント推定に用いる TF-IDF 値の計算をするため，コメントを Bag-of-Words モデルに変換する．Bag-of-Words モデルとは，文章における各単語の出現回数のみを考え，単語の出現順など

は考慮しない単純なモデルである．このモデルをコメントのベクトル表現とし，SVM への入力とする．

Bag-of-Words モデルへの変換のため，コメントの形態素解析を行う．形態素解析には，日本語形態素解析エンジンである MeCab(注7)を用い，システム辞書として mecab-ipadic-NEologd [12] [13] [14] を用いた．さらに,「w」,「８」,「!?」と,「ああ」のように母音を繰り返している文字列を感動詞として辞書に追加した．ニコニコ動画には多くのスラングが存在するが，特に感情を表していながら多くの場面で記述される「w」（笑いを意味する）,「８」（拍手を意味する）,「!?」（驚きを意味する）も扱う必要があると考えた．母音の繰り返しについては，ニコニコ動画のコメントには,「すげえええええ」や「おおおおおおおお」と言ったコメントのように，強い感情を表すために母音を繰り返して記述される傾向があるためである．Brody ら [15] は Twitter(注8)上のコメントには例えば「cool」が「cooooollll!!!」になるように，単語の意味を強調するために同じ文字を繰り返して記述される傾向があることを報告している．

分解されたコメントの各品詞のうち，名詞と，形容詞，動詞，感動詞とを Bag-of-Words モデルに含める．

ここで，コメント中に同じ単語が複数回出現しても，ひとつのコメントにおけるその単語の出現回数は 1 とする．つまり，変換した結果のベクトルは，全体の語彙中の単語がそれぞれコメント中に出現しているかどうかを 0 か 1 で表す．

4.3 特徴シーンの推定

SVM による線形分類器で，コメントの感情をネガティブ，ポジティブ，ニュートラルの 3 つの感情に分類する．SVM モデルの作成には，scikit-learn(注9)を標準的なパラメータで用いた．学習用の教師データとして，実際のコメントにそれぞれの感情のラベルをつけたものを人手で作成した．今回作成した提案手法のプロトタイプでは，教師データには 50 件の動画から無作為に 100 件ずつ抽出した 5000 件のコメントを用いた．

感情別にコメント数の時間推移を測り，コメント数のもっとも多いシーン 3 つを特徴シーンとする．複数の感情のコメント数が多いシーンについては，重複はせずにコメント数が多い感情の特徴シーンとする．コメント数は 10 秒単位で測る．また，動画の冒頭と終わりには投稿者に対しての挨拶コメントが多数投稿されることが多く，それらのコメントは動画の盛り上がりとは関係のないものであるため，動画の最初の 10 秒と最後の 10 秒は除外する．

また，各感情ごとに対象の動画におけるコメント数の時間推移をグラフ化する．グラフにおいて，各感情のコメント数が突出している山の部分がその動画の特徴シーンを表している．

4.4 特徴コメントの推定

特徴コメントの推定には，コメント中の単語の TF-IDF 値を求めて使うというアプローチを取る．動画全体を 10 秒毎に区切り，それぞれのシーン中にあるコメント群をひとつの文書集合，動画全体のコメント群を全文書集合として単語の TF-IDF 値を計算する．例えば，動画内の特定のシーンにのみ「かっこいい」という単語が多く登場していた場合，そのシーン中の「かっこいい」という単語の TF-IDF 値は高くなる．

コメント中に含まれる単語の TF-IDF 値を合計をそのコメントのスコアとし，スコアがもっとも高いコメントをそのシーンの特徴コメントとする．この時，コメント本文があまり長い場合はそのシーンと無関係な内容について記述されていることが多いと考え，本文が 30 文字以下のコメントのみを対象とする．

(注7)：http://taku910.github.io/mecab/

(注8)：https://twitter.com

(注9)：http://scikit-learn.org

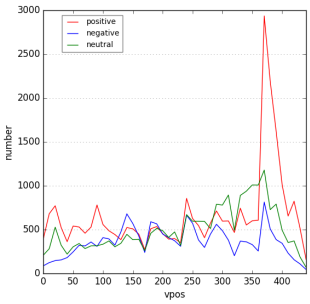

図 3　S1 の盛り上がり推移（提案手法）

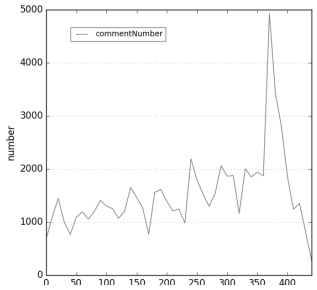

図 4　S1 の盛り上がり推移（コメント数のみ）

5. 実　　験

5.1 実験方法

ニコニコ動画に投稿されている動画に対して提案手法のプロトタイプから得られた結果と，実際にその動画を視聴した 5 人の評価者 (注10)からの回答とを比較して評価実験を行った．評価者はニコニコ動画のコメント表示機能をオフにした状態で動画を視聴した．特徴シーンの推定についてと，特徴コメントの推定について，同じ動画を対象としてそれぞれ評価を行った．

実験では 5 件の動画を対象とする．これらの動画はそれぞれ，ニコニコ動画内の「スポーツ」カテゴリの動画が 2 件,「ゲーム」カテゴリの動画が 2 件,「ニコニコ技術部」カテゴリの動画が 1 件である．

また，コメントの感情を考慮せずに単純にコメント数のみで動画の盛り上がりを推定する手法でも同様の実験を行い，提案手法と比較する．

5.2 特徴シーンの推定

実行結果の例として,「スポーツ」カテゴリの動画 2 件（それぞれ S1，S2 とする）について提案手法とコメント数のみの手法とで作成したグラフ，提案手法で推定した特徴シーンを図 3，4，5，6 と表 1 に示す．図 3，4 が，S1 に対してそれぞれ提案手法とコメント数のみの手法とで作成した盛り上がり推移のグラフで，図 5，6 が，同様に S2 に対してのものである．各グラフは，縦軸が 10 秒あたりのコメント数，横軸が動画の再生時刻を表しており，提案手法で作成したグラフについては，赤線がポジティブなコメント，青線がネガティブなコメント，緑線がニュートラルなコメントの数を意味している．

例として挙げている動画の内容は，S1 が野球の国際大会の

(注10)：評価者の中に著者は含まれていない．

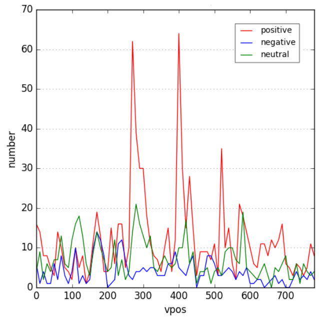

図 5　S2 の盛り上がり推移（提案手法）

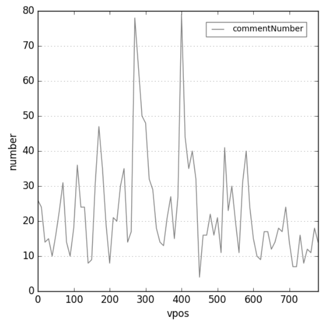

図 6　S2 の盛り上がり推移（コメント数のみ）

表 1　推定された特徴シーン

S1		S2	
提案手法	コメント数のみ	提案手法	コメント数のみ
6 分 20 秒	6 分 20 秒	6 分 40 秒	6 分 40 秒
5 分 10 秒	4 分 50 秒	4 分 50 秒	4 分 50 秒
4 分 00 秒	4 分 00 秒	8 分 40 秒	2 分 50 秒

表 2　S1 の特徴コメント

再生時刻	特徴コメント
6 分 20 秒	うおおおおおおおおおおおおおおおおお
5 分 10 秒	この場面で今の球見逃せるのはすげーわ
4 分 00 秒	とりあえず野球見ようぜ野球

表 3　S2 の特徴コメント

再生時刻	特徴コメント
6 分 40 秒	鉄人すぎるｗｗｗ
4 分 50 秒	実質長友
8 分 40 秒	半田神ｗｗｗｗｗｗｗｗ

決勝戦で決勝点が入った打席の一部始終，S2 が日本人選手も出場している海外サッカーの試合のハイライト映像である．

評価者は対象とする 5 つの動画を視聴して，もっとも重要だと思ったシーンを動画内の再生時刻（10 秒単位）で回答した．それらの回答が，提案手法のプロトタイプが推定した 3 つの特徴シーンに含まれていれば正解とし，正解率を出した．

5.3　特徴コメントの推定

5.2 節で提案手法により推定された S1 と S2 との特徴シーンの特徴コメントの推定結果をそれぞれ表 2，表 3 に示す．

全ての動画に対して表 2，3 のように推定されたコメントについて，「そのシーンの内容を表していると思うか」という質問を評価者に行い，評価者は「思う」,「思わない」の 2 通りで回答した．

6.　評　　価

6.1　特徴シーンの推定

提案手法で作成されたグラフとコメント数のみの手法で作成されたグラフと（図 3 と図 4，図 5 と図 6）を定性的に見比べると，見た目に大きな差は見られなかった．これは，提案手法の盛り上がり度合が，感情別とはいえ，コメント数のみの手法と同様にコメント数を元に測られているためだと考えられる．

表 4　評価者の回答

評価者	動画 S1	動画 S2
A	6 分 20 秒	9 分 30 秒
B	6 分 20 秒	4 分 30 秒
C	6 分 20 秒	4 分 50 秒
D	6 分 20 秒	4 分 30 秒
E	6 分 10 秒	6 分 30 秒

表 1 より，結果に多少の差は見られた．

実験で示した例と同じ動画である，S1 と S2 とに対する評価者らの回答を例として表 4 に示す．

5 件の動画に対しての推定結果と評価者の回答とを比較すると，正答率は提案手法とコメント数のみの手法とでそれぞれ 64 %，56 %だった．これは高い数値とは言えない．

しかし，表 1 と表 4 とを見ると，推定結果と回答とずれが 10 秒しかないものがいくつかあり，実際に動画を見るとこれらは同じシーンを指していると考えられた．表 4 を見ると，動画 S1 に対して A から D までは全員 6 分 20 秒が特徴シーンだと回答して E だけは 6 分 10 秒と回答しており，これも同様に同じシーンを指していた．

このように，時間の区切りや個人の感覚によって，同じシーンを指していてもずれが起こりうるので，推定結果の前後 10 秒の回答を許容して正答率を測ったところ，それぞれ 84 %，76 %に上昇した．提案手法の方が，8 %高い正答率を得た．

動画ごとの正答率に注目すると，動画 S1 の 100 %に対して，動画 S2 は 40 %と低い．図 3，図 5 を見ればわかるように，動画 S1 はグラフの山が動画全体で 1 箇所に集中しているのに対し，動画 S2 は山が数箇所にある．また，動画 S2 の 4 分 50 秒部分や 8 分 50 秒部分を見ると，このコメントの盛り上がりは確かに動画の内容に関係のあるものだと考えられたが，普通に映像を視聴しているだけではそのシーンに起こっている出来事に気が付きにくいものだった．その出来事についてのコメントが表示されることによって気がついた他の視聴者がまたコメントを投稿していき盛り上がったシーンで，評価者はコメントを非表示にして動画を視聴したためにそのシーンの特徴に気が付かなかったものだと考えられる．このように，動画を視聴する際にコメントを表示しているかしていないかによって，動画への印象が異なることが評価者の回答に影響していると考えられる．

6.2　特徴コメントの推定

推定された特徴コメントを「そのシーンの内容を表していると思う」と評価者が回答した割合は 70.6 %だった．

回答の中で特に「思う」という回答が少なかったのは，動画 S1 の 4 分 00 秒時点の特徴コメントである「とりあえず野球見ようぜ野球」（表 2）であり，これは評価者 A から E までの全員が「思わない」と回答した．このシーンを見ると，コメント同士で言い争いが行われており，動画内容に関係があるコメントとは言えなかった．

逆に，動画 S2 の 4 分 50 秒時点と 8 分 40 秒時点と（表 3）についてはどちらも 5 人全員が「思う」と回答した．これらは，

そのシーン中に活躍している人名を含むコメントが推定されたためだと考えられる．

7. おわりに

今回我々は，動画の時刻同期コメントとその感情とを用いて動画の特徴シーンを推定し，またそのシーンの特徴コメントを推定する手法を提案した．コメントの感情抽出には SVM による線形分類器を用いた．また，5 件の動画と 5 人の評価者とを用いて，提案手法とコメント数のみから特徴シーンを推定する手法との比較評価を行った．特徴シーン推定については正答率が 84 %と，コメント数のみの手法と比較して高い結果を得た．しかし，盛り上がり推移の推定から作成されるグラフの形については大きな差は見られなかった．これに大きく差を出すためには，シーンごとの盛り上がり度合を，各感情のコメント数だけでなくコメントの感情の強さを考慮するなどと言った方法が考えられる．特徴コメント推定については，正しく推定できていると評価者が回答した割合は 70.6 %だった．

提案手法により，動画の盛り上がる部分やその内容を事前に把握することが可能で，それによりユーザの視聴時間の削減につながると我々は考える．

今後の課題としては，以下の 2 点が挙げられる：

（1） 特徴シーン推定と特徴コメント推定の精度に関する評価実験をより大規模に行う．

（2） 提案システムで得られる情報を実際にユーザに活用してもらい，視聴時間の削減という本来の目的を達成できるかどうかを評価する．

また，テレビ番組に対する Twitter の実況ツイートからテレビ番組のダイジェストを自動作成する研究 [16] [17] も行われているため，本研究も動画共有サービスだけでなく，ニコニコ生放送や Youtube Live，Twitch(注11)などといったリアルタイムのコメント付き映像配信サービスに対しても同様に適用できる可能性がある．今後はインターネット映像配信のアーカイブも対象として実験を行っていく予定である．

文　献

[1] Google blog, “Improving our brand safety controls” https://blog.google/topics/google-europe/improving-our-brand-safety-controls/ (2018 年 2 月 13 日参照)

[2] comScore, “日本の動画サイト利用動向 - comScore, Inc” https://www.comscore.com/jpn/Insights/Press-Releases/2015/1/comScore-Releases-Japan-Desktop-Online-Video-Rankings (2018 年 2 月 13 日参照)

[3] 後藤真孝, “CGM の現在と未来: 初音ミク，ニコニコ動画，ピアプロの切り拓いた世界：1. 初音ミク，ニコニコ動画，ピアプロが切り拓いた CGM 現象” 情報処理学会論文誌, Vol. 53, No. 5, pp. 466–471, 2012.

[4] 佃洸摂, 中村聡史, 田中克己, “視聴者のコメントに基づく動画検索及び推薦システムの提案” WISS2011, 2011.

[5] 佃洸摂, 中村聡史, 田中克己, “SmartVideoRanking : 視聴者の時刻同期コメントに基づく動画ランキングシステム” 第 8 回 Web とデータベースに関するフォーラム論文集, pp. 56–63, 2015.

[6] 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ダイジェスト動画作成アルゴリズムの検討” 情報処理学会第 73 回全国大会, Vol.2011, No. 1, pp. 347–348, 2011.

[7] 齋藤義仰, 磯貝佳輝, 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” マルチメディア，分散協調とモバイルシンポジウム 2011 論文集, Vol. 2011, pp. 1251–1256, 2011.

[8] 山内嶺, 北山大輔, “ダイジェスト映像自動生成のための観点の入れ替わりに基づいた特徴的シーン抽出” 第 6 回データ工学と情報マネジメントに関するフォーラム, F4-2, 2014.

[9] 末永智彦, 早川智一, 疋田輝雄, “時刻同期コメントを用いた動画の特徴シーンの推定” 情報処理学会第 79 回全国大会, Vol. 2017, No. 1, 2017.

[10] 中村聡史, 山本岳洋, 後藤真孝, 濱崎雅弘, “視聴者反応と音響特徴量に基づくサムネイル動画の生成手法” 情報処理学会論文誌データベース, Vol. 6, No. 3, pp. 148–158, 2013.

[11] 松原宏和, 新妻弘孝, 太田学, “画像特徴量とコメントを用いたニコニコ動画の支持的要約サムネイルの生成手法” 研究報告データベースシステム, Vol. 160, No. 18, pp. 1–9, 2014.

[12] 佐藤敏紀, 橋本泰一, 奥村学, “単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討” 言語処理学会第 23 回年次大会 (NLP2017), NLP2017-B6-1, 2017

[13] 佐藤敏紀, 橋本泰一, 奥村学, “単語分かち書き用辞書生成システム NEologd の運用 — 文書分類を例にして —” 自然言語処理研究会研究報告, NL-229-15, 2016

[14] Toshinori, Sato, “Neologism dictionary based on the language resources on the Web for Mecab” https://github.com/neologd/mecab-ipadic-neologd（2018 年 2 月 13 日参照）

[15] S. Brody and N. Diakopouls “Cooooooooooooooollllllllllll!!!!!!!!!!!!!! Using Word Lengthening to Detect Sentiment in Microblogs” EMNLP, pp. 562–570, 2011

[16] 羽山徹彩, “Twitter データを用いたテレビ番組の自動ダイジェスト生成に対するユーザ分類の適用” 研究報告セキュリティと心理学とトラスト, Vol. 9, No. 14, pp. 1–7, 2014.

[17] 高津戸章人, “Twitter におけるテレビ番組の盛り上がり検出” 人間工学, Vol. 50, Vol. Supplement, pp. 222–223, 2014.

(注11) : https://www.twitch.tv/