

指導教員（主査）：山本祐輔 講師

副査：福田直樹 准教授

2019 年度 静岡大学情報学部 卒業論文

ゲーム実況動画のハイライトシーン 自動検出

静岡大学 情報学部 行動情報学科 所属

学籍番号 70612025

佐藤 颯哉

2020 年 2 月 10 日

概要

本稿では、YouTube に投稿されているゲーム実況の生配信アーカイブ動画についている時刻同期コメントを用いて、該当のシーンが面白いシーンかどうかを推定する分類器の作成方法を提案する。コメントに含まれる名詞の出現頻度とコメントの文末表現に着目し「w」、「笑」、「草」が含まれるコメントの割合をシーンごとについてそれぞれベクトル化し特徴量とする。実際に面白いと判断されたシーンのコメントデータを正解データ、面白くないと判断されたシーンのコメントデータを不正解データとしてラベル付けを行う。また、それらのデータセットを用いてサポートベクターマシンによる機械学習を行い、分類器を構築する。分類器の評価実験として、名詞の出現頻度を考慮した特徴量とした分類器、笑いの感情を示す文末表現がコメントに含まれる割合を特徴量とした分類器、両方を特徴量とした分類器に分け、適合率で評価を行い比較した。その結果、両方を特徴量とした分類器で最も高い適合率を記録した。そのため、名詞の出現頻度と笑いの感情を示す文末表現がコメントに含まれる割合を考慮した特徴量の分類器が面白いシーンを検出する手法として比較3手法の中で最も優れていることが明らかになった。

目次

第 1 章	はじめに	5
第 2 章	関連研究	8
2.1	コメント分析以外によるシーン推定と検出	8
2.2	コメント分析によるシーン推定と検出	8
2.3	本稿との比較	9
第 3 章	利用技術	12
3.1	サポートベクターマシン	12
3.2	クロスバリデーション	14
第 4 章	提案内容	17
4.1	コメントデータの取得	17
4.2	機械学習に用いるための教師データ	18
4.3	特徴ベクトルの作成	20
4.4	ラベル付けと分類器の作成	21
第 5 章	分類器の評価実験	23
5.1	適合率	23
5.2	3 手法の分類器	24
第 6 章	結果	25
第 7 章	考察	26
7.1	提案手法	26
7.2	データ数	26
7.3	提案手法の改善点	28

第 8 章	まとめ	29
参考文献		30

目次

1.1	動画とシーンの関係	6
3.1	線形分離可能なデータモデル	13
3.2	5 分割交差検証の例	15
3.3	層化 3 分割交差検証すべき例	16
3.4	図 3.3 において層化 3 分割交差検証の例	16
4.1	生配信アーカイブ動画とその再生時間についてコメント欄の様子 引用元 https://www.youtube.com/watch?v=_gfopK4jJFs に加筆 . . .	19
7.1	配信開始直後に書き込まれるゲーム内容に関わらないコメントの例 引用元 https://www.youtube.com/watch?v=C6bSNdN2wyc に加筆 . .	27

表目次

6.1	分類器の評価結果	25
-----	--------------------	----

第 1 章

はじめに

近年，スマートフォン端末やタブレット端末の普及により，ウェブの利用が多くの人の日常の一部となっている．ウェブ上には様々なサービスが展開しているが，その中のコンテンツの 1 つに動画配信サービスが存在する．YouTube^{*1}やニコニコ動画^{*2}をはじめとしたユーザが動画を投稿するオンデマンドの動画配信サービスや，TVer^{*3}や FOD^{*4}といったテレビ局が放送した番組を配信するサービス，NETFLIX^{*5}や Hulu^{*6}といった有料会員制のドラマやアニメを配信しているサービスが存在する．総務省情報通信政策研究所が 2019 年に発表した「平成 30 年度情報通信メディアの利用時間と情報行動に関する調査」^{*7}によると YouTube やニコニコ動画などのオンデマンドの動画配信サービスの利用率は 10 代で 87.9%，20 代で 90.9% を記録している．また平日 1 日の利用時間においても 10 代では平均 60.1 分，20 代では 39.8 分，休日 1 日では 10 代 95.4 分，20 代 70.2 分であり 1 日における時間比率も大きいと言える．配信サービスの内容も様々であるが本稿では YouTube に投稿されているゲーム実況動画に着目した．ゲーム配信のカテゴリは生配信の中でも人気のコンテンツであり配信者数，視聴者数ともに多い．近年では企業からのサポートを受けてゲーム実況の生配信活動するストリーマー，インターネット生配信活動を主に行って収益を得るライバーと呼ばれる職業も台頭してきている．動画を配信するプラットフォームの増加や e-sports 業界発展の影響を受け，ゲーム実況動画の視聴を

^{*1} <https://www.youtube.com/>

^{*2} <https://www.nicovideo.jp/>

^{*3} <https://tver.jp/>

^{*4} <https://fod.fujitv.co.jp/s/>

^{*5} <https://www.netflix.com/jp/>

^{*6} <https://www.hulu.jp/>

^{*7} https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000082.html

趣味としている人も増えている。上述の背景から、ゲーム実況動画の視聴者は多く配信者も多いものの、1日に動画を視聴できる時間は限られていることが分かった。そのため、視聴者は効率的に面白いシーンを探す手法を求めているのではないかと考えた。

本稿においての言葉の定義を行う。動画とは1本の動画全体を指し、シーンとは動画に含まれるその動画の1部を指す。(図 1.1)

ハイライトシーンとは動画の面白いシーンであると考え、そのシーンを検出することを試みる。

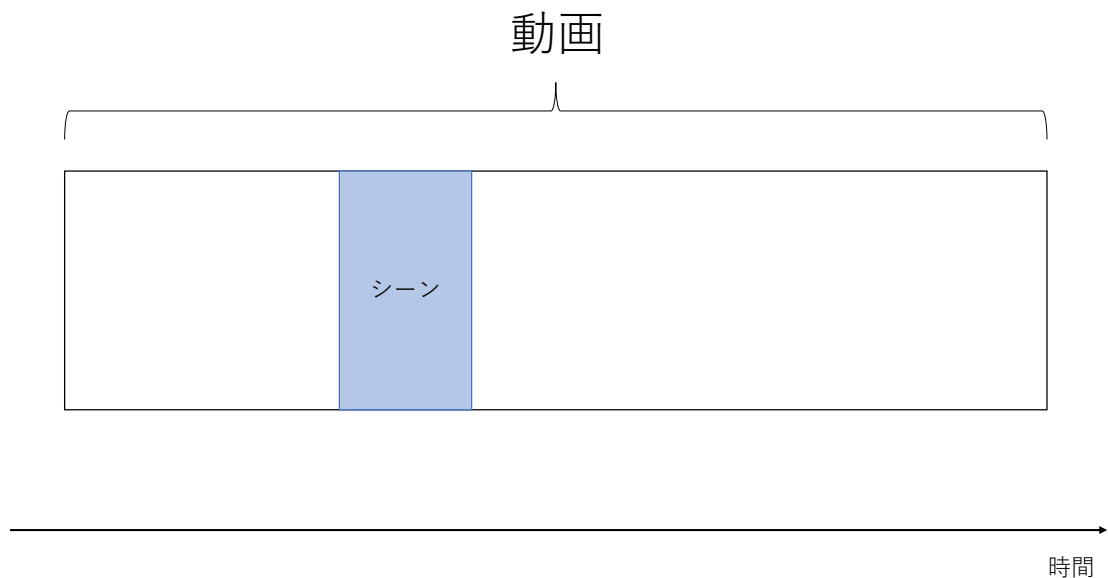


図 1.1 動画とシーンの関係

本稿ではゲーム実況動画の中でも、YouTube で行われたゲーム実況の生配信アーカイブ動画を対象とする。その理由として主に3つの理由が挙げられる。1つ目の理由としてコメントの特性の差が挙げられる。YouTube において、生配信アーカイブ動画には時刻同期コメント [1] がついている。時刻同期コメントとは動画の再生時間に同期したコメントのことである。視聴者が生配信で、あるシーンを見たときそれに対する反応がコメントとして書き込まれ、そのコメント情報が生配信アーカイブ動画にも残る。これを時刻同期コメントと呼ぶ。コメントの内容とシーンが紐づいているため、コメントからシーンを推定するために有用であると考えた。2つ目の理由として生配信アーカイブ動画はシーンを検出する意義が高いことが挙げられる。通常の動画は投稿者により編集されたものであ

り、時間も比較的短い。編集者により編集されたものという点で、ある程度面白いシーンがピックアップされている場合が多いと考えた。この場合、改めて面白いシーンを検出することの必要性が感じられない。しかし、生配信アーカイブ動画は長時間に及ぶものが多く、冗長な場面も多い。編集された動画より面白いシーンを検出する意義があるのではないかと考えた。3つ目の理由として生配信の視聴者数の違いが挙げられる。2019年現在において、他のプラットフォームで行われているゲーム実況の生配信より、YouTubeで行われているゲーム実況の生配信の方が他の配信プラットフォームに比べて圧倒的に視聴者数多く、その差はコメント数の差にも現れると考えた。生配信の視聴者数の判断については、リアルタイムで配信しているチャンネルにどのくらい視聴者がいるのかを確認することができるサービスである「配信者勢いランキング」*8を利用して行った。以上の3つの理由から対象をYouTubeで行われているゲーム実況の生配信アーカイブ動画に決定した。

本稿では、効率的に面白いシーンを探すためにコメントデータから面白いシーンであるかどうかを推定する手法を提案する。本手法で作成した分類器を用いて、面白いシーンが高い精度で検出できるようになれば長時間に及ぶ動画の効率的な視聴が可能になることが期待される。

以下、2章では関連研究について、3章では利用技術について、4章では提案手法について、5章では評価実験について、6章では評価実験の結果について、7章では結果に対する考察について、8章では本稿のまとめを行う。

*8 <http://ikioi-ranking.com>

第 2 章

関連研究

コメントからシーンを推定する研究はニコニコ動画についての研究が多く，様々な視点から研究が行われている．2 章では 2.1 節でコメント分析以外のシーン推定の関連研究について述べる．2.2 節でコメント分析によるシーン推定の関連研究について述べ，2.3 節で関連研究と本稿の比較を行う．

2.1 コメント分析以外によるシーン推定と検出

宮崎らの研究ではサッカーの動画を動画内の画像の特徴からハイライト生成する方法を提案している [2]．画面にボールとゴールポストが映りこんだ時が特徴的なシーンである可能性が高いと考えた．サッカーコート芝生の色である緑色を検出した上でゴールポストとボールが検出された時，それらの位置関係を考慮しシーンを検出している．吉高らの研究では BGM などの音声情報，撮影方法や編集での演出技法から受ける視聴者の印象を用いて映画のダイジェストを生成することを試みた [3]．同じ区間内により多くの演出技法が使われているほうが視聴者に与える影響は大きく，BGM が与える印象も演出技法と同じくらいの重要度であることが述べられていた．

2.2 コメント分析によるシーン推定と検出

末永らは動画に書き込まれた「すげえええ」や「うおおおお」といった同じ文字の繰り返し表現を用いて感情を大きく揺さぶられている様子が表れているコメントに着目した研究を行った [4]．コメントの内容から感情の極性を判断し，極性の値が高いものを特徴シーンとした．佃らは時刻同期コメント自体をその動画のキーワードとして検索システム

に組み込んだ [5]. 動画に書き込まれたコメントが、検索クエリに対してどのくらい適合しているのかを計算する必要があるため、コメントの文字数や頻度、検索クエリとコメントの内容一致率や形容詞の有無など様々な特徴量を用いて有用度を推定するアルゴリズムを考案した. そのアルゴリズムで計算された有用度の高いコメントから動画が与える印象ごとにカテゴリ分けをして、検索クエリから動画をランキング表示できるシステムを提案した. また、佃らの他の研究では「泣ける」や「笑える」などの感情を表すコメントに着目してコメント分析を行った [6]. 感情とコメント内容の紐付いた辞書を作成し、辞書に当てはまるコメントの分析や分布からシーンの感情属性を推定した. また、佃らの別の研究では、動画内に登場する人物に着目してコメント分析とシーンの推定を行っている [7]. 登場人物の名前がコメントに多く含まれる区間は、その人物が活躍しているシーンであるとした. 登場人物名には動画についたタグを用いて、動画内に書き込まれたコメントと一致するものをシーン推定の対象とした. 尾崎らは動画についたコメントを 8 つの感性に分類し、動画を見ている視聴者がどのような感情遷移を行うかシーンごとに可視化を行った [8]. それに伴い動画が与える感情を用いた検索を可能にしようと試みた. 山内らはコメントから視聴者の観点の入れ替わりに着目してシーンを検出し、それを用いてダイジェストを生成することを試みた [9]. 視聴者の観点は同一視聴者によるコメントのネガティブ、ポジティブ度で判定を行った. コメントを書き込んだ視聴者を特定して、コメントを追って視聴者の極性ごとにクラスタリングを行った. このクラスタ間の極性の転換点に特徴的なシーンがあるのではないかと仮定している. この極性の転換点を特徴的なシーンとして検出した. 齋藤らはコメントの量に着目し動画の中の特徴的なシーンを検出することを行った [10]. コメント量のみでシーン抽出を試みると目的に合わない不要なシーンが存在するため、シーンのカット数やコメント数から傾向を読み取り、取り除く方法を考察している. また、別の齋藤らの研究では動画についたコメントを用いて特徴的なシーンを推定し、広告の挿入タイミングを決定するアルゴリズムを提案している [11]. この提案は特徴シーンを推定することができれば、動画の視聴を妨げず広告の挿入が可能であるという仮説に基づく. 特徴シーンの推定には単位時間当たりのコメント量と映像の切り替わり、すなわち動画の場面の転換点を考慮して推定していた.

2.3 本稿との比較

本稿ではシーン検出のためにコメント分析を行うが、それ以外のアプローチも考えられる. 2.1 節では画像処理によるシーン検出方法と、BGM などの音声情報や演出技法の特徴によるシーン検出方法を提案手法とする関連研究を挙げた. これらの手法はゲーム実況

動画のシーンを検出するための手法として適していないと考えた。画像処理による検出については、特徴的なシーンを検出するために、ゲーム実況動画に共通する特徴を表すものを考えなくてはならない。ゲーム実況というジャンルは、同じジャンル内でも扱うゲームタイトルによって動画の内容に大きな差がある。面白いシーンとなりうるシーンに現れる共通の特徴を表すものを動画内で設定するのは非常に困難であると考えた。BGMなどの音声情報や演出技法の特徴によるシーン検出に関してもゲームタイトルに幅があるという同様の理由で難しいと考えられる。ゲーム実況のジャンルの中でゲームタイトルやゲームジャンルを絞って考えることで、画像処理やゲーム内の BGM と SE、配信者の声の大きさ、リアクションの大きさなどから面白いシーンを推定できると考えられる。しかし、それらの特徴量とした場合の分類器は、特定の動画にしか適用できない。それに対して、動画に書き込まれたコメントは視聴者が感じた面白さや面白さにつながる感情が現れている可能性が高い。面白いシーンを推定するために、どのゲームタイトルにおいてもコメント情報は汎用的な利用が可能である。この点がコメント分析によるシーン推定の優れている点であると考えた。

2.2 節ではコメント量や感情に関するコメント、ある動画においての特殊なコメントが多いシーンにその動画の特徴シーンが現れるという関連研究を挙げた。この中でも末永らの研究 [4] が本稿と最も類似していると考えられる。末永らの研究における特徴的なシーン推定の提案手法は、シーンに書き込まれたコメントの極性を用いてシーンの推定である。事前に集めた動画からコメントを取得し、ポジティブ、ニュートラル、ネガティブの 3 種類にラベル付けを行う。このラベル付けされたデータから教師あり学習を行いモデルを作成する。このモデルを用いて対象の動画につくコメントを感情ごとにスコア付けを行い、そのスコアが高いシーンを特徴シーンとして検出している。そして検出されたシーンのコメント分析を行い、シーンを表す特徴語を推定するために TF-IDF を用いている。本稿においてシーン推定の提案手法はシーンに書き込まれたコメント内の笑いの感情を示す文末表現と名詞の TF-IDF 値を考慮して推定される。ポジティブ、ネガティブではなく、明確な感情である笑いに限定することでより面白いシーンが検出できるのではないかと考えた。また TF-IDF 値も特徴量に用いることで他の動画にはない、より特徴的なシーンの検出ができるのではないかと考えた。この 2 つを考慮して特徴量を決定し、分類器を用いてシーン検出を行っていることが先行研究と本稿の差である。動画に書き込まれる時刻同期コメントについても「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にある」と述べられている [1]。これらの先行研究から得た知見を用いて、時刻同期コメントから笑いを示す文末につく感情表現の割合と名詞の出現頻度に着目してベクトル化し、それを特徴量とす

る．それらの特徴量を用いた分類器でコメントデータから面白いシーンを推定することを試みる．本稿で用いた特徴量についての詳細は 4 章で述べる．

第 3 章

利用技術

本研究では分類器構築のためにサポートベクターマシンを利用した。また分類器の適合率を検証するためにクロスバリデーションを行った。

3.1 サポートベクターマシン

サポートベクターマシン (SVM) [12][13] は、分類問題や回帰問題の教師あり学習に用いられる機械学習のアルゴリズムで、本稿では教師あり学習の二値分類問題の手法として用いた。多次元空間にラベルがつけられたデータが存在し、決定境界によってデータが分離できる場合を線形分離可能と言う。この適切な分離を行うことができる決定境界を求める問題である。図 3.1 の場合、A, B のラベルがつけられたデータが 2 次元空間に存在するため、決定境界は緑の直線となる。

SVM と別のアルゴリズムとの大きな特徴の違いはマージン最大化である。マージンとはデータが完全分かれているとき、決定境界と決定境界に最も近い訓練データの距離のことを指す。決定境界と訓練データ間のマージンを大きくすることで、より妥当な決定境界を決定するという特徴である。これを行うことにより誤判定を減らすことができる。また、マージンの内側にデータが入り込むことを許容しないことをハードマージンと呼ぶ。データによって、線形分離可能でない場合も存在する。この場合マージンの内側にデータが入ることを許容して学習を行うことがある。このことをソフトマージンと呼び、線形分離可能でないデータについても学習することができる。どれくらいデータが内側に入るかを許容するかを決めることができる。データが増えると直線で分離することができなくなる場合がある。その場合 SVM においてはカーネル法を用いる。カーネル法を用いることでデータを高次元に変換し線形分離可能でないベクトルも、線形分離可能に変形すること

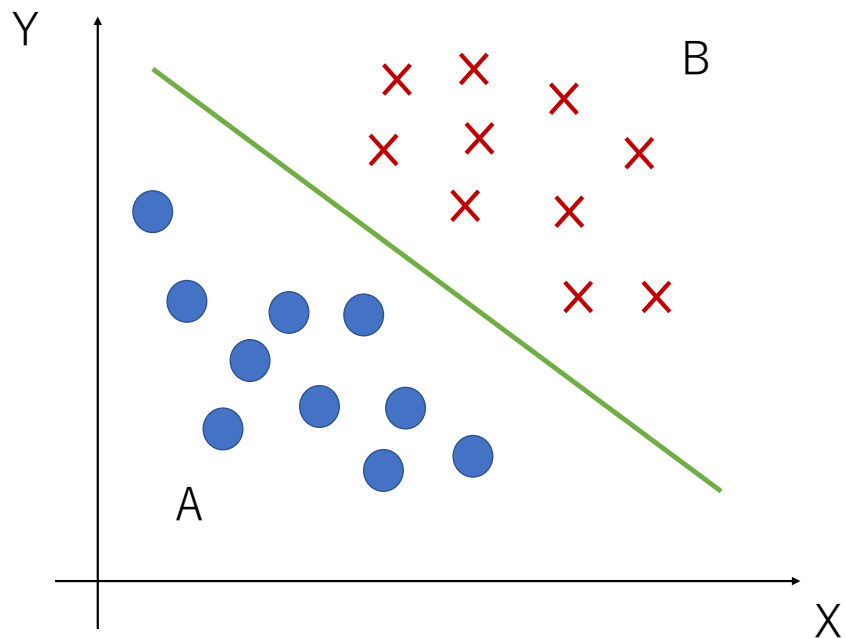


図 3.1 線形分離可能なデータモデル

ができる．これを非線形 SVM と呼ぶ．主なカーネル関数には線形カーネル，シグモイドカーネル，多項カーネル，RBF カーネルなどが存在する．本稿では二値分類問題を扱うため，分類器構築のためにサポートベクターマシンを選んだ．またカーネル法を用い，関数には分類問題において最も一般的に使われている RBF カーネルを採用した．ハイパーパラメータにはコストパラメータを 1000，gamma 値を 0.1 に設定した．

3.2 クロスバリデーション

クロスバリデーション（交差検証）[13] とは訓練データとテストデータを k 分割し、分類器の性能 k 回測定した結果の平均値を計算することによって分類器の妥当性を検証する方法である。訓練データとテストデータの分割時、 k 回のうち 1 回はすべてのデータが訓練データとテストデータとして利用されるため、この検証で得た値はデータの偏りによって偶然算出されたものではないといえる。図 3.2 ではデータを 5 回分割し、5 回検証を行っている。

本稿では、クロスバリデーションの種類として stratified k -fold cross validation（層化 k 分割交差検証）を用いた。この手法はデータを分けるとき訓練データとテストデータ内で正解データと不正解データの比率をもとのデータセット内の比率を同じにしたのちクロスバリデーションを行う検証方法である。この方法は分類問題に多く使われる。例えば図 3.3 のように先頭からラベル A,B,C が 1:1:1 の割合でつけられたデータセットが存在し、そのデータセットを前から 3 分割してクロスバリデーションを行うとする。赤の破線でデータが区切られることになるため、1 回目はラベル A のデータがすべてテストデータになり、ラベル B,C が訓練データとなる。2 回目はラベル B のデータが全てテストデータ、ラベル A,C のデータが訓練データ、3 回目はラベル C のデータが全てテストデータ、ラベル A,B のデータが訓練データとなってしまう。しかし、層化 3 分割交差検証を行うことにより図 3.4 のようになる。1 回目の検証を例に挙げるとラベル A がついたデータ内で 1:2 にテストデータと訓練データに分かれる。ラベル B,C のデータにおいても同様に分割され検証が行われる。また正解データと不正解データの偏りが大きい場合にも有効である。層化 k 分割交差検証を行うことにより、データの偏りを少なくすることができる。

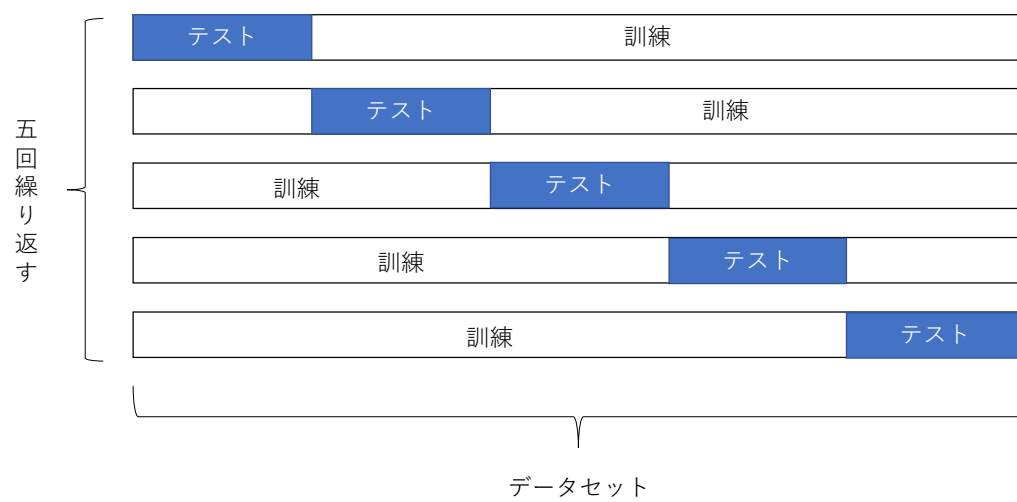


図 3.2 5 分割交差検証の例

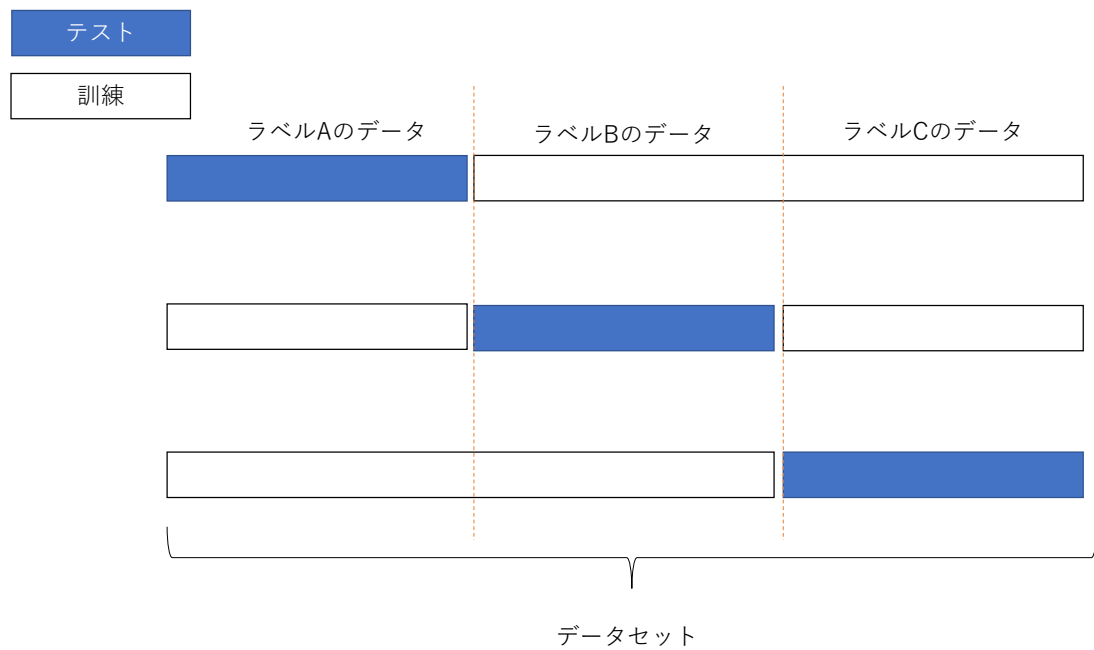


図 3.3 層化 3 分割交差検証するべき例

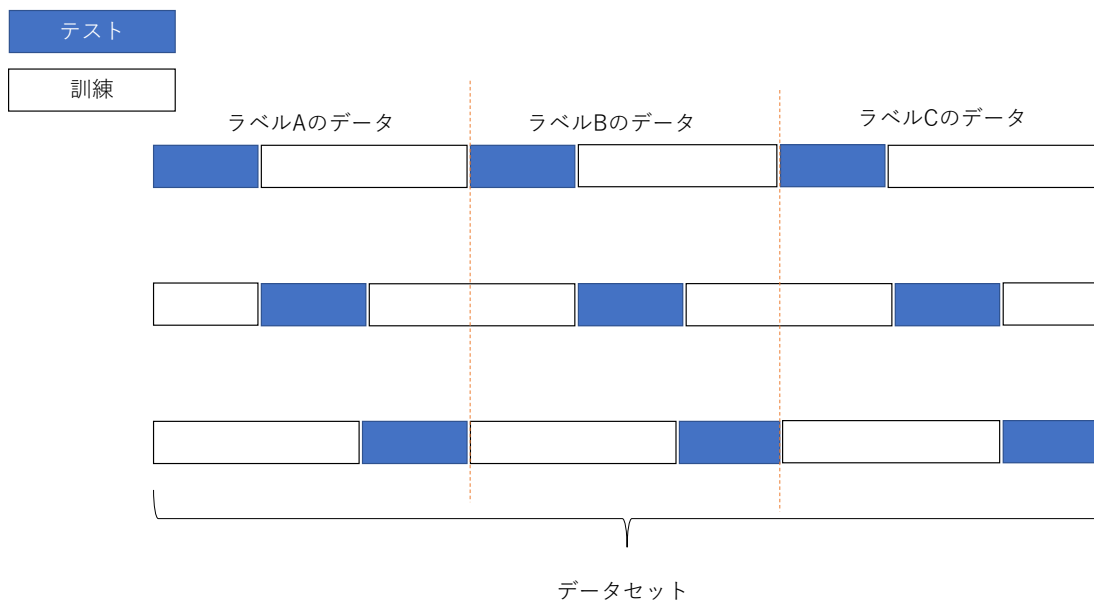


図 3.4 図 3.3 において層化 3 分割交差検証の例

第 4 章

提案内容

本章ではコメントデータから面白いシーンであるかどうか推定する分類器の構築方法について述べる．主なステップについては以下の通りである．

- YouTube で配信されたゲーム実況の生放送動画のアーカイブからコメントデータを取得する（4.1 節）．
- クラウドソーシングサービスを利用し，教師データを作成する（4.2 節）．
- シーンごとのコメントデータに含まれる名詞の出現頻度と，コメントの文末表現着目し「w」，「笑」，「草」が含まれるコメントの割合をシーンごと特徴量ベクトル化する（4.3 節）．
- 面白いと評価されたシーンのコメントデータを正解データ，面白くないと評価されたシーンのコメントデータを不正解データとしラベル付けをする．その後 SVM による機械学習を行い，分類器を構築する（4.4 節）．

4.1 コメントデータの取得

4.1.1 YouTube と生配信のアーカイブ動画の特性

動画配信プラットフォームの最大手である YouTube では日々様々な動画が投稿されたり，配信が行われたりしている．2005 年 12 月 15 日に正式サービスを開始した．2019 年現在では世界で 20 億人以上のユーザがおり，毎日のべ 10 億時間視聴されている．*⁹今回コメントデータを使用するにあたってゲーム実況の生放送アーカイブ動画を使用する．基

*⁹ <https://www.youtube.com/intl/ja/about/press/>

本的に YouTube では投稿された動画に対して視聴者は 1 つのコメントを残す。しかし、生配信についてはその時起きた事象に対してコメントを書き込むことができる。その生配信のアーカイブ動画にはその時書き込まれたコメントが時刻同期コメントとして通常のコメントとは別に残る。図 4.1 は生配信のアーカイブ動画とその再生時間についている時刻同期コメントである。本稿でコメントデータを集める際、考慮したことは以下の項目である。

1. 2019 年 12 月 7 日～12 月 13 日の間に行われた生配信のアーカイブ動画
2. 1 の配信を行った配信者の過去 3 ヶ月以内に行われたゲーム実況の生配信アーカイブ動画で再生数の最も多い動画

上述の 1 か 2 の条件を満たしている生配信アーカイブ動画の時刻同期コメントを取得した。配信者の選定は、恣意的な選定にならないために「配信者勢いランキング」*⁸を用いプラットフォームを YouTube に絞って閲覧した。閲覧した時間帯で上位 3 人の配信者の生配信アーカイブ動画を選んだ。

4.1.2 コメントデータ

YouTube Data API*¹⁰を用いてゲーム実況の生配信アーカイブ動画から時刻同期コメントと、そのコメントが動画のどの再生時間に書き込まれたものなのか示すタイムスタンプをセットで取得する。これらをコメントデータとして 1 動画ごとに保存する。コメントデータの取得には Python を用いた。

4.2 機械学習に用いるための教師データ

4.2.1 データの準備

分類器を構築する際に教師データを用意する必要がある。正解データを面白いシーンについたコメントデータ、不正解データには面白くないシーンについたコメントデータを利用する。そのためにコメントデータを集めた動画の中から面白いシーンと面白くないシーンを判定し抽出する必要がある。コメントデータを集めた 100 本の動画を用いて、コメントの多いシーンを上から 5 つ抜き出す。コメントの数え方としては動画を開始から

*⁸ <http://ikioi-ranking.com/v/youtube>

*¹⁰ <https://developers.google.com/youtube/v3/docs>

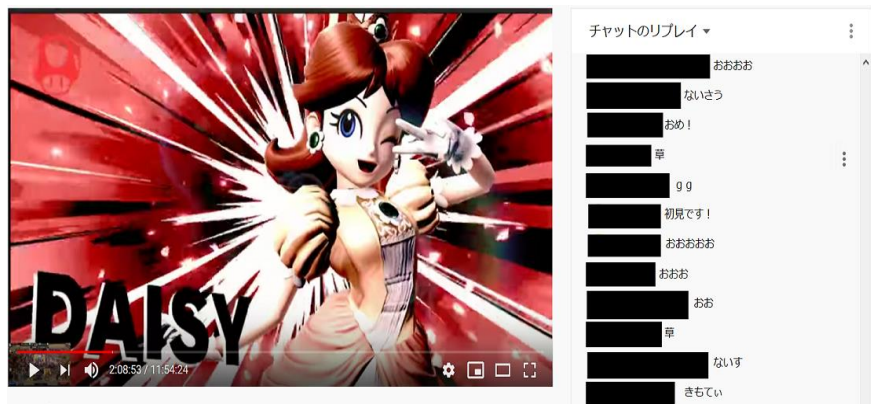


図 4.1 生配信アーカイブ動画とその再生時間についてのコメント欄の様子
引用元 https://www.youtube.com/watch?v=_gfopK4jJFs に加筆

10 秒ごとに区切り，その間のコメント数を数える．利用したコメントは「w」，「8」のみで構成されたものを除き，日本語で書き込まれたものを利用した．本稿で用いる時刻同期コメントは，生配信上で書き込まれたコメントであるため，何か事象が起きた後にそれに対するリアクションとしてコメントが書き込まれることが多い．その特性を加味し，コメントがされた対象の事象を面白さの判定を行うシーンに含めるためにコメント数が多く選ばれた 10 秒の区間とその前の 10 秒を 1 シーンとした．1 シーン 20 秒という条件を保持するため，冒頭の 10 秒が選ばれた場合のみ冒頭の 10 秒とそのあとの 10 秒で 1 シーンとした．コメントデータを集めた動画にこの処理を行い 1 動画ごと 20 秒のシーンを 5 つ抽出した．それらが面白いかわからないかで評価を行う．

4.2.2 クラウドソーシング

シーンの評価には客観的かつ，恣意的でない視点が必要であるため，クラウドソーシングサービスを利用し評価を行う．本稿ではランサーズ^{*11}を用い評価を行うワーカーを雇った．このタスクの参加報酬として 1 件当たり 22 円支払った．

^{*11} <https://www.lancers.jp>

4.2.3 タスク

ランサーズで行った評価タスクについては以下の通りである。「20 秒の動画を 5 本視聴してそのシーンが面白いかわからないかを判断してください」という説明文をワーカーに与えた。1 シーンごとに面白い、面白くないの 2 値で、1 動画当たり 5 人のワーカーに評価を行ってもらった。面白いシーン、面白くないシーンの基準として半数以上のワーカーが面白いと評価したものを面白いシーン、面白くないが半数以上のものを面白くないシーンとした。コメントデータの不備や回答の不備を除いた結果、91 本の動画で合計 455 シーンのデータが取得できた。その結果、455 シーンのうち面白いシーンは 66 シーン、面白くないシーンは 389 シーンに分けられた。

4.3 特徴ベクトルの作成

取得したコメントデータから 2 つの観点を用いてシーンを特徴ベクトル化した。1 つ目はコメント内においての名詞の出現頻度を考慮してシーンを TF-IDF[13] で特徴ベクトル化した。TF-IDF とは TF (Term Frequency) で表される 1 つの文書 d におけるある単語 X の出現頻度と IDF (Inverse Document Frequency) で表されるある単語 X が出現する文書頻度の逆数の積であり、以下の数式を用いて表される。

$$tf(X_i, d_j) = \frac{f(X_i, d_i)}{\sum_{X_k \in d_j} f(X_k, d_j)}$$

$$idf(X_i) = \log \frac{N}{df(X_i)} + 1$$

$$TF-IDF = tf(X_i, d_j) \cdot idf(X_i)$$

$f(X_i, d_i)$ = ある単語 X の 1 つの文書 d_i の中での出現回数

$\sum_{X_k \in d_j} f(X_k, d_j)$ = 1 つの文書 d_j の中でのすべての単語の出現回数を足した数

$N = \text{全文書数}$

$df(X_i) = \text{ある単語 } X \text{ が出現する文書数}$

本稿では 1 シーン中のコメント群を 1 つの文書、455 シーンについてコメント群の集合を全文書として、コメント内の名詞で TF-IDF 値を計算する。例えば 455 シーンのうち、1 つのシーン中にのみ「運」というコメントが多く書き込まれていた場合、そのシーン中の「運」という単語の TF-IDF 値は高くなる。固有名詞やそのシーンの内容を表す名詞がコメントされた場合、TF-IDF 値は高くなりやすいと考えられるため特徴的なシーンの抽出が可能になる。この特徴的なシーンは面白いシーンと関連があるのではないかと考えた。2 つ目はシーンに付いているコメントの文末表現に、「w」、「笑」、「草」が含まれるコメントの割合を特徴量としてシーンごとにベクトル化した。文末表現に「w」が含まれるコメントの割合、文末表現に「笑」が含まれるコメントの割合、文末表現に「草」が含まれるコメントの割合、文末表現に「w」、「笑」、「草」のいずれかが含まれるコメントの割合を特徴量として 4 次元の特徴ベクトルを作成した。これは面白いシーンには上記のような笑いの感情を示す文字が記号的な意味で用いられることが多いのではないかという仮説に基づいている。本稿で笑いを表す文末表現として「w」、「笑」、「草」を用いているがこのうち「w」と「草」はネット上で笑いを表現するためのネットスラングである。「w」は「warai」の頭文字に由来し、「草」は「w」が並んだ際、草原のように見えることが転じて笑いを表現するネットスラングとして定着した。人が感じる面白さには様々な種類があり、面白さには感情が付随する。笑いの感情というものは面白さを判断する上で非常に明確な判断基準であり、多くの人の面白さ基準となっている。それ故に面白いシーンを推定するために重要な特徴量なのではないかと考えた。

4.4 ラベル付けと分類器の作成

91 動画 455 シーンのうち、面白いと判断された 66 シーンを 1、面白くないと判断された 389 シーンを 0 としてラベル付けを行った。これらのステップを踏まえて分類器を構築し、手法の比較を行う。二値分類問題であるため分類器の構築には SVM を利用する。カーネルには RBF カーネルを採用した。SVM のほかに分類器を構築する代表的な機械学習の手法としてランダムフォレスト [14] が挙げられる。ランダムフォレストの特徴として SVM に比べて調整するパラメータが少なく、訓練データの正規化が必要ないこと、学習にかかる計算量が少ないことなどが挙げられる。しかしランダムフォレストは主にデータセットが十分に揃っているときに用いられる手法であるため、データセットが少ない本

稿では SVM を採用した．面白いと判断されたシーンのデータを正解データ，面白くないと判断されたシーンのデータを不正解データとし，教師あり学習を行って分類器を作成した．

第 5 章

分類器の評価実験

5.1 適合率

評価指標には適合率，再現率，F 値を用いる．本研究における適合率 (precision)，再現率 (recall)，F 値 (F-value) の定義を以下に示す．

$$precision = \frac{\text{正解と分類できた正解データ数}}{\text{正解と分類したデータ数}}$$

$$recall = \frac{\text{正解と分類できた正解データ数}}{\text{実際に正解であるデータ数}}$$

$$F\text{-value} = \frac{2recall \cdot precision}{recall + precision}$$

本稿では評価指標のうち，適合率で比較を行う．適合率を主に用いる理由としては，入力されたコメントデータから面白いシーンであるかどうかの判定を行うことが目的であるため，分類器が面白いと推定して実際にそのシーンが面白いシーンであることが最も重要である．再現率は適合率とトレードオフの関係にあり，適合率と再現率の調和平均を取ったものが F 値である．

5.2 3 手法の分類器

構築した分類器についての評価実験を行った。以下の 3 つの手法について性能を比較した。

tf-idf: コメントに含まれる名詞の出現頻度のみを特徴量とした分類器

sub: コメントの文末表現着目し「w」、「笑」、「草」が含まれるコメントの割合のみを特徴量とした分類器

tf-idf+sub: 提案手法のコメントに含まれる名詞の出現頻度とコメントの文末表現に着目し「w」、「笑」、「草」が含まれるコメントの割合をベクトル化し特徴量とした分類器

3 手法のとも、学習には SVM を用いた。以下は評価実験の手順である。

1. 面白いシーンのコメントデータを正解データ、面白くないシーンのコメントデータを不正解データとする。
2. 層化 2 分割交差検証を行うため、データを分割する際、訓練データとテストデータ内で正解データと不正解データの比率が同じになるように分割する。
3. 訓練データに対して層化 2 分割交差検証を行い分類器を構築する。
4. 分類器にテストデータを入力し適合率、再現率、F 値を計算する。

本稿では正解データが 66 個、不正解データが 389 個と正解データが少ないため、偏りを無くすために層化 k 分割交差検証を用いた。

第 6 章

結果

本章では 5 章で述べた 3 手法の適合率 (precision), 再現率 (recall), F 値 (F-value) での評価結果を表 6.1 に示す.

本稿での評価指標での比較は前章でも述べた通り, 適合率によって行う. 名詞の出現頻度のみを特徴量とした場合, 適合率は 0.426, 再現率は 0.486, F 値は 0.454 となった. コメント文末に「w」, 「草」, 「笑」を含む割合のみを特徴量とした場合, 適合率は 0.588, 再現率は 0.547, F 値は 0.553 となり, 名詞の出現頻度のみを特徴量とした場合より適合率が高いことが明らかになった. 最後に名詞の出現頻度とコメント文末に「w」, 「草」, 「笑」を含む割合の両方を特徴量とした場合, 適合率は 0.653, 再現率は 0.529, F 値は 0.524 となり, 本稿において提案特徴量を用いた分類器が最も高い適合率を記録した.

表 6.1 分類器の評価結果

	precision	recall	F-value
手法			
tf-idf	0.426	0.486	0.454
sub	0.588	0.547	0.553
tf-idf+sub	0.653	0.529	0.524

第7章

考察

本章では6章で得た結果を踏まえて、提案手法についての考察とどのようなコメントが面白さを判断する上で重要であるかの考察を行う。

7.1 提案手法

6章で示した結果より、tf-idfにおいて、適合率は0.426となった。この結果から、コメント中に含まれる名詞の頻度を考慮した特徴量1つでは面白いシーンを正確に検出できないと考えられる。subに着目すると、適合率は0.588であった。tf-idfよりsubのほうが面白いシーンの推定には重要な特徴量であると同時に、笑いの感情は面白さの判断に重要な要素であると考えられる。tf-idf+subが比較手法の中で一番高い適合率を示したことについて理解できることは、tf-idfのみを用いた場合より、tf-idfとsubの結合ベクトルとして特徴量とすると面白いシーンを検出するための特徴量となりうるということである。tf-idfとsubについての考察を深めると、あるシーンで起きた特徴的な事象に対して多数の視聴者が笑いの感情を抱き、その内容についてのコメントを行う。そのコメント内に笑いを表現する文末表現が含まれる割合が高い時、そのシーンは面白いシーンである可能性が高いと考察できる。

7.2 データ数

今回は機械学習に十分なデータ数を準備することができなかった。この原因には2つの問題がある。1つ目は準備した動画数の問題である。抽出するシーンを5シーンから増やすことでデータの総数を増やすことを検討したが、動画時間や総コメント数にばらつき

があるため、1 シーン当たりのコメント数の平均に差が出てしまうため適切ではないと考えた。2 つ目の問題はタスク設計である。面白いという判断は人によって基準がさまざまであり、思うように面白いシーンの取得ができなかった。また 20 秒のシーンのみ視聴しての判断であったため、話のオチだけを視聴し面白さを判断させるようなシーンがあった。事前の流れを理解していなかったり、シーン外で事前に張られた伏線の回収のシーンであったりする場合、面白いシーンとして判定されにくいと考えた。こう言ったシーンのデータの取得方法、分類方法は今後の課題である。また、動画冒頭の生配信自体は開始しているがサムネイルが表示され、配信者が登場していないシーンにつく定型コメントのようなものが評価の対象になっていたものがあった。図 7.1 においては、ゲーム配信は始まっていないが多くの視聴者が「待機」というコメントを書き込んでいる。こういったシーンのコメントは、生配信の内容に関わらない上に配信者独自のコメントが多く書き込まれる例があるため、TF-IDF 値を重要視するのであれば除外してタスクを設計するべきだと考えた。

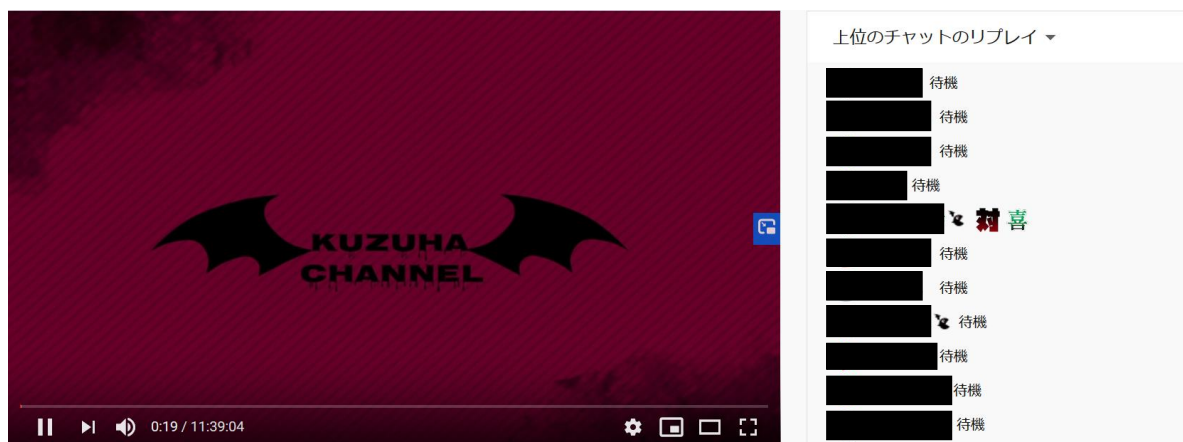


図 7.1 配信開始直後に書き込まれるゲーム内容に関わらないコメントの例
引用元 <https://www.youtube.com/watch?v=C6bSNdN2wyc> に加筆

7.3 提案手法の改善点

本稿においては，分類器の作成にとどまってしまった．実際に面白いシーンをハイライトシーンとして検出するためには，動画 1 本のコメントデータから面白いシーンを検出する機能の実装が必要である．適合率も決して高い数値ではないため，分類器自体の性能向上についても考える必要がある．また，本稿では SVM を用いてカーネル関数に RBF カーネルを採用したがハイパーパラメータの値を最適化できていない．グリッドサーチやランダムサーチを用いて最適化することによって性能の向上ができる可能性がある．本稿では，笑いの感情を示す文末表現を含むコメントの割合に着目した．これとは別の特徴量として，別の感情を示すコメントの文末表現を含む割合を特徴量に加えることで分類器の性能が向上するのではないかと考えた．動画コメントにおいて感情を表すコメントの文末表現として考えられるものを例に挙げると，驚きを示す「！」や，混乱を示す「？」，気まずさを示す「…」などが存在する．これらの感情表現を含むコメントの割合を特徴量として分類器を作成し，適合率の比較を行うことで，どの感情が面白さに与える影響が大きいのかを考えることができる．どの感情が面白さに大きな影響を及ぼしているかを理解することで，重み付けを行い，特徴量を決定することでさらなる適合率の向上が見込めるのではないかと考えた．亀井らの研究によると，動画の内容とそこに書き込まれるコメントによってほかの視聴者との共有体験が行われ，相乗的に視聴者の感情が高揚するということが述べられている [15]．本稿ではシーンの面白さ評価のタスクの際，動画のみを視聴して評価を行った．そのシーンに書き込まれたコメントを表示しながら評価を行うことによって，感情の影響を面白さの判定に，より大きく反映させられるのではないかと考えた．分類器の性能向上についてはどこまで適合率を向上させればいいのかという明確な基準を設けることがまだできていない．今回の検証において適合率が 0.653 でありこれはまだ十分ではないと感じた．この基準を明確にするためにも未知の動画からシーン検出を行い，客観的な評価を得て適切な適合率を検討する必要があると考えた．

第 8 章

まとめ

本稿では YouTube に投稿されているゲーム実況の生配信アーカイブ動画に書き込まれた時刻同期コメントを用いて、面白いシーンかどうかを推定する分類器の作成方法を提案した。取得したコメントデータから、コメント内においての名詞出現頻度を考慮することと、シーンに書き込まれているコメントの文末表現に「w」、「笑」、「草」が含まれるコメントの割合の 2 つの観点を用いて、シーンの特徴ベクトル化した。それらを用いて SVM による機械学習を行い面白いシーンであるかどうかの分類器を作成し、層化 2 分割交差検証で評価を行った。その結果、両方の特徴量を用いた分類器が最も高い適合率を記録した。今後の課題については、データセットの総数を増やし、学習の量を増やすことや、笑い以外の感情を示すコメントの文末表現に着目して分類器の性能向上を図ることが挙げられる。ハイライトシーンの検出の観点から見ると、未知の動画のコメントデータから面白いシーンを検出し、そのシーンが面白いものなのかの客観的な評価を行うことが必要である。

参考文献

- [1] 後藤真孝. CGM の現在と未来：初音ミク，ニコニコ動画，ピアプロが切り拓いた世界. 情報処理学会論文誌, Vol. 53, No. 5, pp. 466–471, 2012.
- [2] 宮崎英吾, 大関和夫. サッカーにおける画像処理によるハイライト生成. 映像情報メディア学会技術報告, Vol. 35, No. 8, 2011.
- [3] 吉高淳夫, 田中壮詩, 平嶋宗. 映画等を対象としたダイジェスト生成のための映像特徴に関する検討. 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 (HCI) , Vol. 68, pp. 79–86, 2007.
- [4] 末永智彦, 早川智一, 疋田輝雄. 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. 第 10 回データ工学と情報マネジメントに関するフォーラム, pp. E3–4, 2018.
- [5] 佃洸撰, 中村聡史, 田中克己. 視聴者のコメントに基づく動画検索および推薦システムの提案. 第 18 回インタラクティブシステムとソフトウェアに関するワークショップ, 2011.
- [6] 佃洸撰, 濱崎雅弘, 後藤真孝. 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 7 回データ工学と情報マネジメントに関するフォーラム, 2015.
- [7] 佃洸撰, 中村聡史, 山本岳洋, 田中克己. 映像に付与されたコメントを用いた登場人物が注目されるシーンの推定. 情報処理学会論文誌, Vol. 52, No. 12, pp. 3471–3482, 2011.
- [8] 林貴宏, 尾内理紀夫. 動画とライブ配信のコメントデータを対象としたリアルタイム時系列感性変化量・可視化・検索システム. 人工知能学会論文誌, pp. C4–1, 2019.
- [9] 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, pp. F4–2, 2014.

- [10] 齊藤義仰, 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. 情報処理学会論文誌コンシューマ・デバイス&システム (CDS), Vol. 2, No. 1, pp. 74–81, 2012.
- [11] 齊藤義仰, 村山優子. 視聴者コメントを用いた広告動画挿入タイミング決定アルゴリズムの提案と評価. 情報処理学会論文誌, Vol. 52, No. 2, pp. 520–528, 2011.
- [12] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory (COLT 1992)*, pp. 144–152, 1992.
- [13] 秋庭伸也, 杉山阿聖, 寺田学. 機械学習図鑑 見て試してわかる機械学習アルゴリズムの仕組み. 株式会社 翔泳社, 2019.
- [14] Leo Breiman. Random forests. *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.
- [15] 亀井且有, 豊田晃史, 串田淳一. 擬似同期を用いた動画共有によるビデオ視聴者の感情高揚. 知能と情報, Vol. 24, No. 5, pp. 944–953, 2012.