

時刻同期コメントを用いた動画の登場人物の 活躍度推定とその検索への応用

佃 洸 撰^{†1} 中村 聡 史^{†1} 田中 克 己^{†1}

現在の動画検索では、人物名をキーワードとして検索を行うと、キーワードがタイトルやタグなどに含まれる動画のみが検索の対象となる。そのため、ある人物が活躍している動画などの検索が行えないという問題がある。そこで本研究では、動画に投稿された時刻同期コメントを用いて、動画の登場人物がどの程度視聴者の注目を集めるような活躍をしているかを推定する手法および、それを応用した動画検索システムを提案する。

Estimating the Activity of Characters in a Video Clip based on Comments and a Search Application

KOSETSU TSUKUDA,^{†1} SATOSHI NAKAMURA^{†1}
and KATSUMI TANAKA^{†1}

When we search for video clips, we often use a character name as a query. In this case, however, only video clips that include the query in their titles or tags are shown as a search result. Therefore, we cannot search video clips in which a character is active. In this study, we propose a method for estimating the activity of a character in a video clip using comments posted to the video clip. Moreover, we implement a search system based on the method.

^{†1} 京都大学大学院情報学研究科
Department of Social Informatics, Graduate School of Informatics, Kyoto University

1. はじめに

近年、YouTube^{*1}やニコニコ動画^{*2}などの動画共有サービスが広く利用されている。YouTubeには2010年1月の時点で10億本以上、ニコニコ動画には2011年6月の時点で610万本以上の動画が投稿されており、非常に多様な動画の視聴が可能になっている。その一方で、動画の数が多くなりすぎているため、視聴したいと考える動画を検索するのは困難である。また、ユーザには限られた時間しかないため、発見した全ての動画を視聴することは難しい。そのため、ユーザには限られた時間の中で自分の視聴したい動画を効率良く検索したり視聴したりすることを望んでいる。具体的には以下のような例が挙げられる。

- 動画を最初から最後まで全て視聴する時間はないので、ある人物が注目を集めているシーンを集めたダイジェスト動画を視聴したい
 - キーワードで動画を検索したときに、検索結果に含まれる動画を全て視聴する時間はないので、ある人物がより動画全体に渡って視聴者の注目を集めている動画や、動画の中で雰囲気が大きく変わるきっかけとなる活躍をしている動画を優先的に視聴したい
 - ユーザが気に入ったある動画の関連動画として、タイトルや説明文などのテキストが類似している動画ではなく登場人物の活躍パターンが類似している動画を推薦してほしい
- 動画のダイジェスト生成や動画の盛り上がり度を計算するため、映像を分析してシーンの切り換わりを検出したり⁴⁾、動画中の音声进行分析することでシーンの盛り上がりを検出する⁶⁾研究は行われているが、視聴者の情報を利用していないため、コンテンツ作成者の意図を反映しているにすぎず、動画の作成者が注目させようと意図している人物と、視聴者が注目している人物は必ずしも一致するとは限らない。一方、動画に対して付与されたコメントを視聴者反応とみなし、視聴者の番組に対する感想を反映した情報の測定を行う研究はなされている²⁾。しかし、登場人物に対して付与されたコメントに特化していないため、上記のような要求を満たしていない。また、膨大な動画に対する適用事例がない。

ある人物がある動画内のどのシーンで注目されているかは、動画のタイトルやタグ、説明文といったテキスト情報からは判定不可能である。また、画像解析によってその人物が画面に映っているかどうかを判定することはある程度可能であるが、注目の度合いは判定できない。そこで本研究では、動画の中で各登場人物が視聴者の注目を集める言動をとっている

*1 <http://www.youtube.com/>

*2 <http://www.nicovideo.jp/>

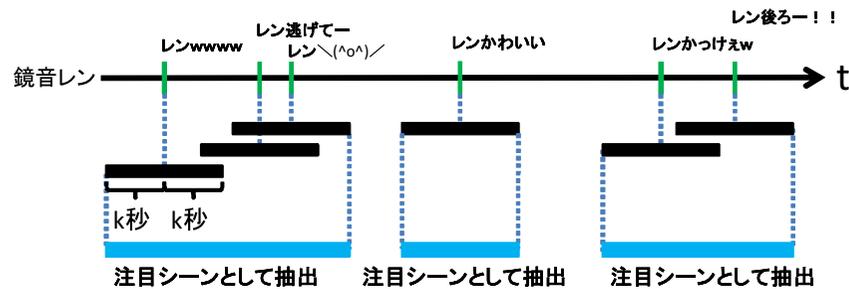


図 2 名前を含むコメントを用いた注目シーンの抽出

2.2 注目シーンの抽出

コメントの中にある人物名が含まれていれば、そのコメントが投稿された時には、その人物が視聴者の注目を集める言動をとっている可能性が高いと考えられる。ある動画 v における、全てのコメントの投稿時間の集合を $T = \{t_1, t_2, \dots, t_n\}$ とし、 T の内、人物 x の名前を含む全コメントの投稿時間の集合を T_x とする。このとき、各 $t \in T_x$ の前後 k 秒間は人物 x が注目を集める言動をとっていると仮定する。つまり、時間 t_i 以降に投稿されたコメントで x の名前を含むものの内、最も近い投稿時間を t_{i+1} とすると、 $t_i + k \leq t_{i+1} - k$ を満たしていれば、人物 x は $t_i - k$ 秒から $t_{i+1} + k$ 秒にわたって視聴者の注目を集めていると言える。このように、 $2k$ 秒以内に連続して同一人物の名前を含むコメントが投稿されている期間をその人物の注目シーンとして定義する。例えばある動画において、「鏡音レン」という人物の名前を含むコメントが動画の再生時間軸上に図 2 のようにマッピングされたらとすると、図中の青い線分で示された時間帯が「鏡音レン」の注目シーンとして抽出される。

2.3 注目の大きさの数値化

ある人物が視聴者の注目を集める言動をとっているときに、視聴者は必ずしもその人物の名前を含むコメントを投稿するとは限らない。本研究では、名前を含むコメントが投稿された時間と近い時間に投稿されたコメントは、その人物の言動に対して投稿されたコメントであると仮定し、その数が多いほど、視聴者のその人物に対する注目の大きさは大きいと考える。ただし、動画に投稿されるコメントは、全てが動画に登場する人物に対するものであるとは限らない。コメントの中には、動画の内容を説明しているコメントや視聴者同士で会話のやりとりをするように投稿されるコメント、動画の内容とは全く関係の無いコ

メントなどがある。ここで、動画の登場人物の言動に対して投稿されたコメントには、何らかの感情を表す語が含まれていると仮定する。例えば、ある人物が視聴者を笑わせる言動をとっているシーンでは、その人物の言動に対して喜びの感情を含むコメントが投稿され、ある人物が視聴者の批判を買う言動をとっているシーンでは、その人物の言動に対して否定的な感情を含むコメントが投稿されると考えられる。そこで本研究では中村ら⁹⁾ が作成した喜び、悲しみ、驚き、肯定、否定のそれぞれに対応する語の集合から成る辞書を用いて、辞書に登録されている語を含むコメントのみを解析の対象とする。

動画に投稿されたコメントの内、人物名を含まず、感情を表す語を含むコメントの投稿時間の集合を T_{imp} とし、注目シーン r の開始時間を r_{start} 、終了時間を r_{end} と表すと、人物 x のある活躍シーン r における注目の大きさ $Attention(r)$ は次式により表される。

$$Attention(r) = |\{t | r_{start} \leq t \leq r_{end}, t \in T_x \cup T_{imp}\}|$$

つまり、注目シーン r の間に投稿されたコメントのうち、人物 x の名前を含むコメントと感情を表す語を含むコメントの数の和により表す。ただし、名前を含むコメントの中にはその人物が注目を集める言動をとっていないときに投稿されるものもあるため、 $Attention(r)$ の値が閾値 α 以下のシーンはノイズとして除去する。

3. 実験

本節では、登場人物が視聴者の注目を集める言動をとっているシーンではその人物に対するコメントが投稿されるかということと、注目シーンを推定するために感情を含むコメントのみを対象とすることの有用性を調べるために、被験者を用いた評価実験を行った。今回評価の対象とした動画は、「アイドルマスター」、「MMD(MikuMikuDance)」、「政治」のカテゴリに属する動画である。アイドルマスターは、アイドルを育成するゲームシリーズの名称であり、ニコニコ動画ではゲームの登場人物を用いて作成したアニメや、ゲーム中の動画に別の音楽を重ねた動画などが投稿されている。MMDは無償で利用できる3DPV作成ツールであり、キャラクターのモデルを読み込むことで様々なキャラクターを用いて動画を容易に作成することを可能としている。MMDのカテゴリにはこのツールを用いて作成された動画が含まれる。政治のカテゴリには、政治家を登場人物として作成したアニメや、国会での討論等の動画が含まれる。各カテゴリに属する動画集合は、タグに「アイドルマスター」、「MMD」、「政治」のそれぞれの単語を含む動画とした。これらのカテゴリを選択した理由は、以下の3つである。

- 1つの動画に多くの人物が登場する

- 様々な人物の組み合わせから成る動画が存在する
- 特定の人物が目立つ動画が極端に多いということはなく、動画によって登場人物の役割が様々である

3.1 データ収集

実験を行うにあたり、我々はニコニコ動画に投稿されている動画の中でコメントが 1000 件以上投稿されている約 17 万件の動画について、それぞれのタイトル、説明文、タグ、コメント、再生数、コメント数、お気に入り数を収集し、データベースに格納した。なお、コメントはクロール時の最新の 1000 件を収集した。

3.2 名前を含むコメントの割合

まず、提案手法の適用可能性を示すために、全てのコメントに対する名前を含むコメントの割合を調べた。そのために、「アイドルマスター」、「MMD」、「政治」の 3 つのカテゴリについて、各カテゴリに含まれる全ての動画に投稿された全てのコメントからランダムに 3000 件ずつ抽出した。その後、抽出された全てのコメントを著者が目視で確認し、名前を含むコメントの割合を求めた。その結果、名前を含むコメントの割合は「アイドルマスター」カテゴリで 9.6%、「MMD」カテゴリで 10.1%、「政治」カテゴリで 7.9%であった。

3.3 正解セットの作成

次に、被験者実験により、動画の各登場人物の注目シーンの正解セットを作成した。そのために、「アイドルマスター」、「MMD」、「政治」の各カテゴリから、動画をランダムに 12 本ずつ選択した。ただし、今回は再生時間が 2 分 30 秒以上、6 分以内の動画を評価の対象とした。被験者数は 5 名で、内訳は 20 代の男性 4 名と女性 1 名である。そして、1 つの動画につき 3 名の被験者が評価を行った。評価方法は、動画ごとに、動画に登場する人物の一覧と動画を 15 秒ごとに区切った表を用意し、各人物が各区分においてどの程度注目を集める言動を取っているかを 3 段階で評価してもらった。3 段階の内訳は、0 が注目を集める言動を全くとっていない、1 がやや注目を集める言動をとっている、2 がかなり注目を集める言動をとっている、とした。なお、本評価を行う際に用いた人物名辞書は 44 件からなる。被験者は動画を視聴する際、コメントを表示しないようにし、動画のみを閲覧して評価を行った。評価の結果の例を表 1 に示す。

その後、各動画ごとに 3 名の被験者の評価結果を集約し、正解セットを作成した。正解セットの作成方法は次の通りである。ある動画 v で人物 x の各時間区分において 3 名の被験者のうち 2 名以上が 1 または 2 をつけている区間を全て求め、その区間を動画 v において人物 x がやや注目を集める言動をとっている、またはかなり注目を集める言動をとって

表 1 1 人の被験者の 1 つの動画に対する評価結果の例

	0:00-0:15	0:15-0:30	0:30-0:45	...	4:45-5:00	5:00-5:09
初音ミク	0	1	0	...	0	0
弱音ハク	1	2	2	...	1	0
巡音ルカ	0	0	0	...	2	1
KAITO	0	0	2	...	0	2

いるシーンとする。これを全ての動画の全ての登場人物について求めたものを正解セット 1 とする。

さらに、ある動画 v で人物 x の各時間区分において 3 名の被験者のうち 2 名以上が 2 をつけている区間を全て求め、その区間を動画 v において人物 x がかなり注目を集める言動をとっているシーンとする。これを全ての動画の全ての登場人物について求めたものを正解セット 2 とする。

3.4 注目シーンの抽出精度の検証

動画の登場人物が視聴者の注目を集める言動をとっているときはその人物に対するコメントが投稿されるという仮説を検証するために、3.2 節のパラメータ k および 3.3 節の閾値 α の値を変化させたときの適合率、再現率、 F 値を求めた。適合率、再現率、 F 値は各動画の各登場人物ごとに次のようにして求められる。

まず、正解セット 1 において 2 人以上が 1 または 2 をつけた区間を 1 とし、それ以外の区間を 0 とすると、動画 v の登場人物 x の注目シーンは 0 と 1 の配列として表される。これを $list_1$ とする。3 章で述べた手法により、 $k = k_i$ 、 $\alpha = \alpha_j$ を決めると $Attention(r)$ の値が求められる。ただし、シーン r が複数の区間にまたがっている場合は、各区分での注目の大きさを求め、その値が α_j 以上の区間を 1、 α_j 未満の区間を 0 とする。例えば、人物 x の注目シーンが 20 秒から 41 秒までであった場合、20 秒から 30 秒までと 30 秒から 41 秒までで分割し、20 秒から 30 秒までの間の注目の大きさが α_j 以上であれば 15 秒から 30 秒は x の注目シーンとなり、30 秒から 41 秒までの間の注目の大きさが α_j 未満であれば 30 秒から 45 秒は x の注目シーンとはならない。これにより提案手法により求められる注目シーンも 0 と 1 の配列として表される。これを $list_2$ とする。このとき、動画 v の登場人物 x の適合率 $P(v, x)$ 、再現率 $R(v, x)$ 、 F 値 $F(v, x)$ はそれぞれ次の式により求められる。

$$P(v, x) = \frac{\text{common}(list_1, list_2)}{\text{notzero}(list_2)}$$

表 2 正解セット 1 における各カテゴリの F 値の最大値とそのときの k および α の値

	MMD	アイドルマスター	政治	全カテゴリ
F 値	0.691	0.695	0.819	0.679
k (秒)	2.5	1.0	0.50	2.0
α (個)	2	4	6	12

表 3 正解セット 2 における各カテゴリの F 値の最大値とそのときの k および α の値

	MMD	アイドルマスター	政治	全カテゴリ
F 値	0.715	0.670	0.762	0.655
k (秒)	6.5	1.5	5.5	0.50
α (個)	53	13	83	8

$$R(v, x) = \frac{\text{common}(\text{list}_1, \text{list}_2)}{\text{notzero}(\text{list}_1)}$$

$$F(v, x) = \frac{2 \cdot P(v, x) \cdot R(v, x)}{P(v, x) + R(v, x)}$$

ここで、 $\text{notzero}(\text{list}_i)$ は list_i の要素の 1 の個数を表し、 $\text{common}(\text{list}_1, \text{list}_2)$ は list_1 と list_2 の同じ時間区間の値が共に 1 となっている区間の数を表す。これを全ての動画の全ての登場人物についてもとめ、その平均値を k 、 α におけるスコアとする。本実験では、 k の値は 0.5 から 10 まで 0.5 ずつ変化させ、 α の値は 1 から 100 まで 1 ずつ変化させた。 α の値が大きくなると、注目シーンとして抽出されることが少なくなるため、適合率は上昇したが再現率が低下した。一方で k の値が大きくなると、様々な区間が注目シーンとして抽出されるため、再現率は上昇したが適合率は低下した。 F 値の最大値は $k = 2.0$ 、 $\alpha = 12$ のときの 0.679 であった。また、このときの適合率は 0.801、再現率は 0.676 であった。さらに、各カテゴリの F 値の最大値とそのときの k と α の値を表 2 に示す。

次に、正解セット 2 において 2 人以上が 2 をつけた区間を 1、それ以外の区間を 0 として、先ほどと同じ範囲で k と α を変化させ、 F 値の最大値を求めたところ、 $k = 0.50$ 、 $\alpha = 8$ のときに最大値 0.655 となった。このときの適合率は 0.678、再現率は 0.748 であった。各カテゴリの F 値の最大値とそのときの k と α の値を表 3 に示す。

3.5 感情を含む周辺コメントを使用する有用性の検証

3 章では、名前を含むコメントの周辺コメントのうち、感情を含むコメントのみを使用する手法の提案を行った。そこで本節では、次に挙げる 2 つの手法との精度の比較を行い、提案手法の有用性を検証する。

- 比較手法 1：3.3 節において、注目の大きさを求める際に動画に投稿された全てのコメントを対象とする。そして、各区間の注目の大きさが閾値 β 以下であればノイズとして注目シーンから除く。
- 比較手法 2：動画を 15 秒ごとに区切り、人物 x の注目シーンを、各区間に含まれる x の名前を含むコメントの数のみから求める。この場合も、ある区間の名前を含むコメントの数が閾値 γ 以下であればその区間をノイズとして注目シーンから除く。

テストセット 1 およびテストセット 2 に対して提案手法と 2 つの比較手法を適用したときの各カテゴリの F 値の最大値を表 4 および表 5 に示す。なお、 β と γ の値はいずれも 1 から 100 まで 1 ずつ変化させ、最適な値を求めた。この結果から、いずれのテストセットにおいても全カテゴリの F 値の最大値は提案手法が両比較手法を上回っていた。

4. 考 察

4.3 節の結果より、正解セット 1 における全カテゴリの F 値の最大値は 0.679 であり、比較的高い精度で注目シーンの特定が行えていると言える。被験者がスコアとして 1 をつけたシーンとしては、画面に少数の人物がある程度の大きさで映っていたり、会話をしたりしているものが多かった。よって、動画に投稿されたコメントを解析することで、動画のタイトルやタグ、説明文には書かれていないが動画に登場している人物の判定などを行うことができると考えられる。カテゴリ別に見ると、特に政治のカテゴリにおいては F 値の最大値が 0.819 と高い値になった。4.2 節で名前を含むコメントの割合を調べた際に、政治カテゴリは他の 2 つのカテゴリに比べてその値が低かったにも関わらず、最適な k の値は 0.50 と小さい。このことから、政治カテゴリの動画では、ある人物が視聴者の注目を集める言動をとったときに集中的にその人物の名前を含むコメントが投稿されていると言える。また、MMD やアイドルマスターのカテゴリに含まれる動画では、政治カテゴリの動画に比べて、短時間に次々と画面に映る人物が変わる動画が多く、名前を含むコメントが分散しやすいということも、 F 値が低かった原因として考えられる。

正解セット 2 の全カテゴリの F 値の最大値は 0.655 であり、こちらも比較的高い精度で注目シーンの特定を行っていた。被験者がスコアとして 2 をつけたシーンとしては、視聴者を笑わせるような言動や感動させる言動をとっているシーンが多かった。表 2 と比べて、 k の 1 秒あたりの α の値が大きいことから、視聴者の注目をより集めるシーンでは 1 秒あたりのコメント数が多いことがわかる。全カテゴリの最適な k の値が 0.50 であるのに対して、MMD と政治のカテゴリではそれぞれ最適な k の値が 6.5 と 5.5 というように大きく異

表 4 テストセット 1 における各手法の F 値の最大値の比較

	MMD	アイドルマスター	政治	全カテゴリ
提案手法	0.679	0.695	0.819	0.679
比較手法 1	0.671	0.699	0.801	0.671
比較手法 2	0.664	0.677	0.794	0.664

表 5 テストセット 2 における各手法の F 値の最大値の比較

	MMD	アイドルマスター	政治	全カテゴリ
提案手法	0.715	0.670	0.762	0.655
比較手法 1	0.671	0.661	0.675	0.633
比較手法 2	0.582	0.643	0.667	0.612

なる値となったが、今回の実験からは原因はわからなかった。

次に 4.4 節の結果から、テストセット 1 ではいずれのカテゴリでも手法間で大きな差は見られなかった。一方でテストセット 2 では特に MMD と政治のカテゴリにおいて提案手法と比較手法の間で差が見られ、特に視聴者の注目を集めるシーンの抽出においては感情を含む周辺コメントを考慮することが有用であることが示せた。

5. 登場人物の活躍パターンを指定した動画検索

実験で得られた知見をもとに、動画検索システムを実装した。本システムでは、ユーザは登場人物とその活躍パターン、そして検索結果の絞り込みのためのキーワードを指定した動画の検索を行うことができる。指定可能な活躍パターンとして、我々は以下の 4 つの活躍パターンをあらかじめ用意した。

- ある人物が動画全体に渡って活躍している「全体型」
- ある人物が動画の後半から活躍している「後半型」
- ある人物が動画の前半にのみ活躍している「前半型」
- ある人物が動画の一部でのみ活躍している「ピンポイント型」

これら 4 つの活躍パターンに加えて、動画の再生時間を 10 分割し、各区間の活躍度を 3 段階で指定可能な機能も実装した。本章では、各活躍パターンに対する各動画のスコアの計算方法および、それらにもとづく検索例について述べる。

5.1 活躍パターンによる動画のランキング

システムが検索結果をユーザに提示するまでの流れは以下の通りである。

- (1) ユーザは検索を行いたい人物 1 名の名前 c を入力し、その人物の活躍パターンを 1 つ

指定する。また、キーワードにより検索結果を絞り込みたい場合はキーワード k を入力する。

- (2) システムはユーザが入力した人物の活躍シーンが 1 つ以上存在する動画集合 V_c を求める。
- (3) ユーザがキーワードを入力していた場合、システムはタイトル、タグ、説明文のいずれかにそのキーワードを含む動画集合 V_k を求め、 $V_c \leftarrow V_c \cup V_k$ とする。
- (4) 4 つの活躍パターンに応じて、各動画 $v \in V_c$ のスコアを以下のように求める。

(a) 全体型

動画 v における人物 c の注目シーンの集合を R とする。動画 v のスコア S_v を次式により求める。

$$S(v) = \sum_{r \in R} Attention(r)$$

(b) 後半型

動画 v の前半部分における人物 c の注目シーンの集合を F 、後半部分における人物 c の注目シーンの集合を L とする。動画 v のスコアを次式により求める。

$$S(v) = \sum_{l \in L} Attention(l) - \alpha \sum_{f \in F} Attention(f)$$

(c) 前半型

動画 v のスコアを次式により求める。

$$S(v) = \sum_{f \in F} Attention(f) - \alpha \sum_{l \in L} Attention(l)$$

(d) ピンポイント型

ピンポイント型では、人物 c に対するコメントが動画の一部分にしており、かつそのコメント数および周辺コメント数が多いほどその動画のスコアは高くなる。そのため、まず次式により人物 c の名前を含むコメントが付与された時刻の偏りの大きさを求める。

$$C(v) = 1 - \sqrt{\frac{1}{|T_x|} \sum_{t \in T_x} \left(\frac{t}{t_v} - \bar{t} \right)^2}$$

ここで、 $\bar{t} = \frac{1}{|T_x|} \sum_{t \in T_x} t$ 、 t_v は動画 v の再生時間である。そして動画 v のスコアを次式により求める。



図3 検索結果の例

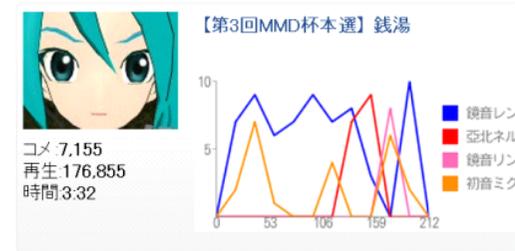


図4 各登場人物の活躍パターンを表すグラフ

内で最も重要な箇所の判別や映像の要約を試みている．また，中村ら⁹⁾ はニコニコ動画の動画に対して投稿されたコメントから喜び，悲しみなどの印象情報を抽出し，インデックスを作成することで，印象に基づく動画検索及びランキングを可能としている．いずれの研究においても，動画に登場する人物には着目していない点で本研究とは異なる．

Diakopoulos ら³⁾ は Shamma ら¹⁰⁾ の行った研究を基に，テレビで放送された大統領討論会に対する Twitter^{*1} 上のコメントを用いて，動画に登場する人物に対してポジティブな感情を持ったユーザとネガティブな感情を持ったユーザのどちらが多いかを分析している．しかし，この研究では動画内で誰がどの時間にどのような内容を話しているかといった情報をあらかじめ用意した上でコメントを分析しているため，こうした発信情報が無い動画に対しては手法を適用できない．我々が対象とする動画共有サイトに日々アップロードされる膨大な動画はそうした情報をもたないため，Diakopoulos らの手法は利用できないが，我々の手法は視聴者の反応を利用するものであるため，一般のユーザが作成した動画に対してもその動画に登場する各人物に対する視聴者の反応の大きさを求めることが可能である．

テレビ番組と同期してコメントの書き込みが行われる Web 上の実況チャットを利用した研究では，上原ら¹²⁾ が実況チャットのコメントを分析し，番組内である人物やテーマが話題となっているシーンを求めている．また，宮森らは実況チャット上に現れる特徴的な表現を処理することで，番組の盛り上がり場面や視聴者の嗜好・趣味に沿ったリアクションなど，視聴者視点に関連するメタデータを抽出したり⁷⁾，それに基づいて視聴者の盛り上がりや自分と類似した嗜好をもつ他人が興味を示す部分など，様々なビューを作成したり⁸⁾ している．これらの研究では，ある人物がより注目されている動画を見たいというユーザの欲求を

$$S(v) = C(v)^{\beta} Attention(c)^{\gamma}$$

(5) V_c の動画を S_v の値が高い順にランキングし，結果をユーザに提示する．

5.2 検索例

登場人物を「鏡音レン」，活躍パターンを「全体型」，検索結果の絞り込みのキーワードを「MMD 杯本選」としたときの検索結果を図3に示す．検索結果の各動画には，動画のサムネイルやタイトル，再生時間等に加えてグラフが表示される(図4)．このグラフの横軸は動画の再生時間，縦軸はコメントの数であり，動画の中で名前が言及されている各登場人物について，動画の再生時間を10分割したときの各区間に付与された各登場人物の名前を含むコメントの数を表している．

6. 関連研究

動画の再生時間に応じてユーザがコメントを投稿できるサービスを対象とした研究は近年盛んに行われている．

青木ら¹⁾ はニコニコ動画の動画に対して投稿されるコメントの出現頻度を用いて，動画

*1 <http://twitter.com/>

満たすことはできないが、本研究では動画間での人物の注目の大きさを考慮し比較することで、そのような意図を持ったユーザに動画をランキングして提示することが可能である。

また、独自のアノテーションシステムを開発し、それを利用した研究も行われている。山本ら¹³⁾は任意のビデオシーンに対するコメントの投稿や、ブログへの引用ができるオンラインビデオアノテーションシステム Synvie を開発し、増田ら⁵⁾がそのシステムを用いて、動画のシーンに対して投稿されたコメントや引用元のブログ記事の文章からタグを自動生成し、動画のシーン検索を可能としている。しかし、彼らのシステムを利用することで、ある人物が登場するシーンを検索することは可能になるが、人物に対する注目の大きさに基づいて検索することはできない。

7. ま と め

本稿では、動画に投稿された時刻同期コメントを用いて動画に登場する人物が視聴者の注目を集める言動をとっているシーンを推定し、それぞれのシーンにおいてその人物に対する注目の大きさを推定する手法の提案を行った。今回提案した手法では、人物の名前を含むコメントの分布から各人物が視聴者の注目を集めているシーンを、人物の名前を含むコメントの数とその周辺の感情を含むコメントの数からその人物に対する注目の大きさを求めた。また、提案手法を適用した動画の検索システムを実装した。

実験では、被験者が注目を集める言動をとっていると判定した人物について、そのシーンの特定の精度を検証した。その結果、いずれのカテゴリにおいても比較的高い精度で各登場人物が注目されているシーンを特定することができていた。さらに、被験者が特に注意をひく言動をとっていると判定したシーンの特定については感情を含むコメントを用いることが有効に働くことがわかった。一方で、登場人物が注目されているシーンを特定するために最適なパラメータの値はカテゴリにより大きく変わるため、今後はコメントの傾向とパラメータの値の関係をより詳細に考察することで、適切な値を推定する手法を考える必要がある。さらに、現在は名前を含まないコメントが誰に向けられたものであるかを十分に考慮できていないが、例えばある人物と共に出現しやすい語を求めることで、名前が明記されていなくても誰に対するコメントであるかを判定できるようにする予定である。

参 考 文 献

1) 青木秀憲, 宮下芳明: ニコニコ動画における映像要約とサビ検出の試み, 情処研報 2008-HCI-128/2008-MUS-75, Vol.2008, No.50, pp.37-42 (2008).

- 2) 大黒泰平, 加藤友規, 土居清之, 亀山涉: 番組に対する視聴者入力情報からの時系列キーワード抽出の改善に関する検討, 情報科学技術フォーラム一般講演論文集, pp.81-82 (2004).
- 3) Diakopoulos, N.A. and Shamma, D.A.: Characterizing debate performance via aggregated twitter sentiment, Proceedings of the 28th international conference on Human factors in computing systems, pp.1195-1198 (2010).
- 4) 橋本隆子, 白田由香利, 真野博子, 飯沢篤志: TV 受信端末におけるダイジェスト視聴システム, 情報処理学会論文誌: データベース, Vol.41, No.SIG3(TOD6), pp.71-84 (2000).
- 5) Masuda, T., Yamamoto, D., Ohira, S. and Nagao K.: Video Scene Retrieval Using Online Video Annotation, Lecture Notes on Artificial Intelligence, LNAI 4914: JSAI 2007 (K. Satoh, et al. Ed.) Springer-Verlag, pp.54-62 (2008).
- 6) Miyamori, H.: Automatic annotation of tennis action for content-based retrieval by integrated audio and visual information. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2728, Springer 318-327 (2003).
- 7) 宮森 恒, 中村聡史, 田中克己: 番組実況チャットを利用したテレビ番組のメタデータ自動抽出方式, 情報処理学会論文誌 データベース Vol.46, No.SIG 18 (TOD 28), pp.59-71 (2005).
- 8) Miyamori, H., Nakamura, S. and Tanaka, K.: Generation of Views of TV Content Using TV Viewers' Perspectives Expressed in Live Chats on the Web, Proceedings of the 13th Annual ACM International Conference on Multimedia (ACM Multimedia2005), pp.853-861 (2005).
- 9) Nakamura, S. and Tanaka, K.: Video Search by Impression Extracted from Social Annotation, Proceedings of the 10th international conference on Web Information Systems Engineering (WISE2009), LNCS 5802, pp. 401-414 (2009).
- 10) Shamma, D.A., Kennedy, L. and Churchill, E.: Tweet the debates, ACM Multimedia Workshop on Social Media (2009).
- 11) Tahara, Y., Tago, A., Nakagawa, H. and Ohsuga, A.: NicoScene: Video Scene Search by Keywords Based on Social Annotation, 2010 International Conference on Active Media Technology, pp.461-474 (2010).
- 12) Uehara, H. and Yoshida, K.: Annotating TV drama based on viewer dialogue-analysis of viewers' attention generated on an Internet bulletin board, The 2005 Symposium on Applications and the Internet, 2005. Proceedings, pp.334-340 (2005).
- 13) Yamamoto, D. and Nagao, K.: iVAS: Web-based Video Annotation System and its Applications, Proceedings of the 3rd International Semantic Web Conference (ISWC2004), demo 29 (2004).