ユーザの興味とコメント分布によるニコニコ動画の分析

小西 敦郎† 細部 博史‡

†法政大学 情報科学研究科 〒184-8584 東京都小金井市梶野町 3-7-2
‡法政大学 情報科学部 〒184-8584 東京都小金井市梶野町 3-7-2
E-mail: †atsuro.konishi.6b@stu.hosei.ac.jp, ‡hosobe@acm.org

あらまし ニコニコ動画とは，日本で有名な動画共有プラットフォームである．ニコニコ動画には約 1800 万件の動画が投稿されており，ユーザはキーワードによる検索や関連動画機能により動画を探索する．しかし，一般的に動画は短い文と少数のタグで構成されるため，それらからユーザが興味のある動画を見つけることは難しい．本研究では，ニコニコ動画の Time-Synchronized Comment とコメント分布の特徴を用いた動画の分析手法を提案する．本手法では動画をコメントしたユーザの集合とみなし，ユーザの共通の興味に基づいて動画をクラスタリングする．実験ではニコニコ動画に投稿された動画に対して提案手法を適用し，既存のテキストベースや画像ベースの手法との比較による定量的評価と，クラスタリング結果の主観的な評価を行った．定量的評価の結果，提案手法は同様のユーザベースの既存手法に比べて，テキストや画像ベースの既存手法やカテゴリーを正解のデータとした際の正規化相互情報量の値が高かった．主観評価の実験では，提案手法は既存の手法に比べて同等かそれ以上の結果を示した．

キーワード ニコニコ動画，ユーザベースクラスタリング，Time-Synchronized Comment，SNS，コメント分布

1. はじめに

近年，ユーザが動画を投稿するソーシャルメディア（YouTube, bilibili, ニコニコ動画等）の利用が盛んになっており，多数の動画が投稿されている．ニコニコ動画は日本で有名な動画共有プラットフォームの１つで,ニコニコ動画に投稿された動画はゲームやアニメ，音楽に関連するものを中心に約 1800 万件である．このような動画共有プラットフォームでの動画の探索にはキーワードによる検索と関連動画機能が用いられるが，これらでは不十分であると考える．キーワードによる検索では，動画のタイトルや説明文，タグにキーワードが含まれる動画を検索することができる．通常動画には少ない量の文字情報しか付与されず，またこれらは主に動画の投稿者が付与するものであるため，動画の内容に関係のない情報を加えることもできる．また，キーワードにより動画の内容に着目して検索することは，動画の内容を明確に示す単語が存在し，それを動画投稿者とユーザ両方が認識している必要がある．関連動画機能は選択された動画との関係性をもとに関連する動画を提示する．ニコニコ動画では 1 つの動画に対し 29 の動画が関連動画として提示される．関連動画機能のアルゴリズムは公開されていないが，同じカテゴリーの動画や文字情報が類似する動画が関連動画として提示される場合が多い．そのため，キーワード検索の結果と類似した動画や，関係のない動画が提示されることがある．

過去の研究では，主に動画の画像やメタデータをもとに分類や推薦を行っている．画像ベースの分類手法の問題点として，画像として類似した動画の内容が類似しているとは限らない点がある．例えば，画像の物体認識を用いた場合，人が画像の中心にいる動画がニュース番組か料理番組であるかの判別は難しい．画像の色を用いた分類では，芝生が緑色であるためサッカーの動画と動物の動画を似た動画とする可能性がある．分類に動画のメタデータを用いる場合，主にテキスト情報が用いられる．この場合，動画が持つテキスト情報が少ないため分類精度には疑問が残る．また，動画の内容に関係しないテキスト情報により誤った分類をする可能性がある．

本論文では，ニコニコ動画のコメント機能の特徴に着目し，ユーザの共通する興味をもとにした分析手法を提案する.提案手法は動画のコメント分布をもとに，ユーザが興味を持った動画を推定しクラスタリングを行う．コメント分布はニコニコ動画の機能である Time-Synchronized Comment (TSC)を用いて取得する．TSC はリアルタイムで動画の内容に関連付ける独特なユーザインタラクションレビューである．TSC の最大の特徴は，日付データと vpos データを持つことである．日付データはコメントが投稿された際のタイムスタンプであり，vpos はユーザがコメントをした際に見ていた動画上の時間的位置を表す．図 1 のように，あるユーザが動画を視聴すると，すでに投稿されたコメントの内容が vpos に対応する動画上の同じ画面に表示される．vpos により，ユーザは他のユーザが同じ画面を見て感じたことを知ることができ，また動画を盛り上げるためにコメントをすることができる．

図 1 TSC を表示したニコニコ動画の動画再生画

本研究では，TSC におけるコメント分布の特徴から動画に対するユーザの興味を推定する．動画におけるコメントの分布は一様でなく，動画ごとに異なる．一般的に動画の初めと終わりにコメントが集中し，残りの時間帯は動画の進行とともにコメントが少なくなる．例外として，動画の盛り上がりに対応する部分のコメント数は他の時間に比べ増加する．このことから，動画の盛り上がりの部分にコメントを投稿するユーザは動画の内容に強い興味を持っているという仮定をもとに，コメントしたユーザの動画に対する興味の程度を推定する．

実験において，提案手法をニコニコ動画に投稿された動画に対して適用し，定量的評価と定性的評価を行った．定量的評価では，正解となるラベルデータとして既存のテキストベースのクラスタリング手法（LDA [1], GSDMM [2]）や画像ベースのクラスタリング手法（IIC [3]）の結果を用い，正規化相互情報量（NMI [4]）を用いて評価を行った．この実験の目的は提案手法を既存のテキストベースや画像ベースの手法と比較し，妥当性を示すことにある．本研究で使用したデータセットにはラベルデータが付随していないためである．定性的評価では，1 つの動画が 2 つのクラスタのどちらに所属するかを回答するタスクを被験者が繰り返す主観実験を行い，結果をもとに精度を評価した．定量的評価の結果より，提案手法はユーザベースの既存手法に対して高い NMI の値を示した．定性的評価より，提案手法は既存手法と同等かそれ以上の精度を示した．

本論文は，著者らが国際会議で発表した論文 [5]をもとにしたものである．本論文ではさらに画像ベースの既存手法を用いた定量的，定性的評価実験とその結果についての報告を追加している．

2. 関連研究

2.1. 動画の分析と推薦

動画の分析や推薦を行う研究は数多くある．Zhou ら [6]は複数のユーザが共有するコミュニティへの動画推薦を提案した．この手法ではクリックされた動画を文書として扱っている．動画は時間的に連続したキーフレーム上に構築されたキューボイドシグネチャの組と，それにコメントしているユーザの集合として記述された社会的なつながりを表現した．彼らは動画間のシグネチャの社会的つながりを計算し，動画を推薦した．Jansen ら [7]は YouTube の大規模音声データにノンパラメトリックなクラスタリングアルゴリズムを適用し，音声イベントを検出した．彼らは DenStream と呼ばれるストリーミングクラスタリングアルゴリズムを用い，教師なし能動学習や弱い教師付き音声モデリング，教師なしの活動検出への期待を示した．Daniel [8]は YouTube のトランスクリプトドキュメントの集合からトピックを発見する手法を提案した．トランスクリプトドキュメントは動画の字幕を表すテキスト情報であり，トピックは LDA を用いて抽出された．Roy ら [9]は協調フィルタリングの問題はほとんど閲覧されていないアイテムを推薦することが難しい点にあると考えた．この問題点を軽減するため，ユーザとアイテム間の感情のつながりをもとにモデル化した潜在因子を学習する visual-CliMF を提案した．Zhao ら [10]は次の動画を推薦するための多目的ランキングシステムを提案した．このシステムはユーザからの 2 種類のフィードバックであるエンゲージメントと満足のふるまいをもとに学習する．前者はクリックや視聴に，後者は動画へのいいねなどに対応する．

2.2. TSC

TSC に関する研究は主に日本や中国の動画共有プラットフォームを対象に行われている．これは TSC を採用した有名な動画共有プラットフォームが日本のニコニコ動画や，中国の bilibili や AcFun であることによる．Ren ら [11]は bilibili に投稿された動画を，TSC を利用して分類している．彼らは多層ニューラルネットワークを組み合わせた深層ニューラルネットワークを用いている．この手法は各動画の TSC と各ユーザの TSC に着目し，動画の特徴とユーザの特徴，コメントの時間をもとに学習を行う．Tsukuda ら [12]は TSC の感情を動的に検出し，サポートベクター回帰で学習する SmartVideoRanking を提案した．このシステムは正規化されたコメントから 14 の特徴を計算し，ランク付けを行う．Yang ら [13]はレビューベースの推薦手法を提案した．彼らは TSC の，後にされるコメントが既にされているコメントに影響を受けるという文脈依存性（Herding Effect）の現象を活用して，画像とテキストを混合させたモデルを設計した．このモデルではユーザの特徴を，TSC のテキストとそれに対応する画像から取得する．Bai ら [14]は TSC（弾幕とも呼ばれる）のトピックを検出し，Yang ら [15]は時系列順の TSC から有効グラフを構築し，タグの抽出を行った．

3. TSC におけるコメント分布

この章では，実験で用いたデータセット内（表 1）の TSC におけるコメントの分布がどのようなものか説明する．一般的にコメント分布は一様でなく，動画ごとに特徴がある．我々はコメント分布により動画の内容に反応しているユーザの認識ができると仮定する．図 2 は表 1 に含まれる動画のコメント分布である．横軸は動画時間を 10 分割したものであり，縦軸は各動画での全体のコメント数に対する当該時間帯のコメント数の割合の平均である．図 2 より，コメントは動画の初めと終わりに集中しており，その他の時間では時間とともに減少していく．動画の初めと終わりにされているコメントの多くは挨拶のような，動画の内容に関係しないコメントである．この特徴は特にコメント数の少ない動画で顕著である．

表 1 データセット

	Dataset5	Dataset50	Dataset100
投稿日時	2018/9/1～30		
コメントしたユーザの最低数	5	50	100
動画数	45274	9470	4628
コメント数	6134029	4761992	3477685
コメントの投稿日時	2018/8/23～2020/1/25	2018/9/1～2020/1/25	
コメントしたユーザ数	1945830	1467446	1198719

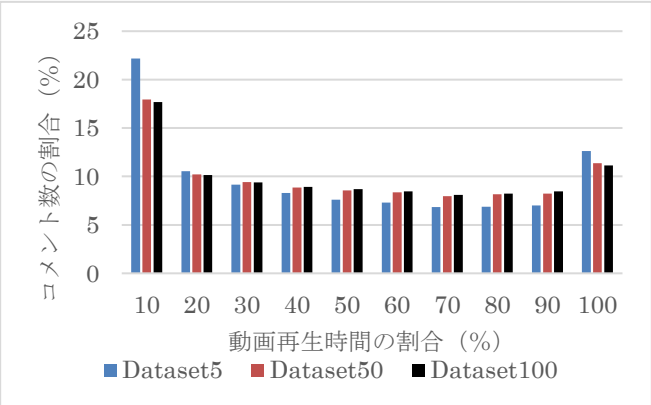

図 2 各データセットにおけるすべての動画の TSC のコメント分布

一方で個々の動画には，平均的なコメント分布とは異なる特徴的なコメント分布が存在する．図 3 の横軸は動画時間を 1 分区切りとしたものであり，縦軸はコメント数である．図 3 の動画は図 2 のように，多くのコメントが動画の初めに集中しており，これらの中には動画に独特な挨拶に対応するコメントが 36.7%を占めている．また，動画の終わり 20 秒に投稿されたコメントのうち 21.7%は動画の内容に直接言及するものではなかった．動画の初めにコメントが集中する一方で，動画の盛り上がりが動画の中盤にあるため，図 2 とは異なるコメント分布となっている．本研究では，このようなコメント分布の特徴の違いや共通点をもとに，動画の内容により強い興味を持つユーザを識別する．

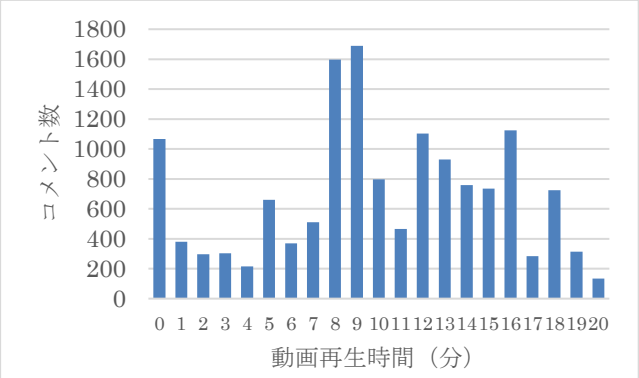

図 3 特定の動画での特徴的なコメント分布

4. 提案手法

本論文では，似たようなユーザがコメントした動画は共通する内容であるという仮定をもとに，コメント分布の特徴を用いた動画のクラスタリング手法を提案する．この手法は Uchida ら [16]の手法を拡張したものである．

本手法は以下の 3 つのステップからなる．

1. 全ての動画の組に対して，2 つの動画両方に対してコメントしたユーザの重複度を用いて動画間の類似度を計算する．
2. 計算した類似度をもとに，重み付き無向グラフを構築する．
3. Louvain 法 [17]を用いて重み付き無向グラフをクラスタリングする．

Louvain 法はクラスタ間のつながりの程度を表すモジュラリティを用いたクラスタリング手法で，分割したグラフのモジュラリティの最大化を目的とする．また，大規模なネットワークを高速で分割することができる．

ステップ 1 で，動画間の類似度を計算する必要がある．そのために，動画iに対するユーザkの重要度を表す値$r_{i,k}$を定義する．B_kはデータセットにおいてユーザkがしたすべてのコメントを表す多重集合である．

$$B_k = \bigcup_i \bigcup_n c_{i,k,n}$$

$c_{i,k,n}$は動画iに対するユーザkのn番目のコメントの vpos である．$r_{i,k}$は動画iに対してユーザkがしたすべてのコメントに重み関数$w(c_{i,k,n}, t)$を適用した値の和であり，B_kにより正規化される．

$$r_{i,k} = \frac{1}{|B_k|}\sum_n w(c_{i,k,n}, t)$$

$w(c_{i,k,n}, t)$は$c_{i,k,n}$の値とあらかじめ設定した時間幅tにより決定する重み関数である．$\boldsymbol{r}_i$は動画iに対応する行ベクトルであり，Uはデータセット内のユーザの総数である．

$$\boldsymbol{r}_i = (r_{i,0}, r_{i,1}, \ldots r_{i,U})$$

この行ベクトルを用いて動画iと動画jの類似度を，コサイン類似度をもとに計算する．

$$\mathrm{sim}(\boldsymbol{r}_i, \boldsymbol{r}_j) = \frac{|\boldsymbol{r}_i \cdot \boldsymbol{r}_j|}{|\boldsymbol{r}_i||\boldsymbol{r}_j|}$$

ここで，重み関数$w(c_{i,k,n}, t)$を以下の仮定のもとに定義する．動画におけるコメント分布は一様ではなく，動画の初めと終わりにコメントが集中する．それらの時間帯に投稿されたコメントの多くは挨拶など定型的なコメントである．そのようなコメントは動画の内容に言及するものでないため，その時間帯に投稿されたコメントは重要でない．また，動画の初めと終わり以外でコメント数が急激に増加している時間帯は動画の盛り上がりであると判断する．その時間帯にコメントしたユーザは動画の内容に対して強い興味を持っており，そのコメントは重要であると考える．

これらの仮定より，重み関数$w(c_{i,k,n}, t)$を以下のように定義する．$A_{c_{i,k,n},t}$は動画上の時間$c_{i,k,n}$を中心に幅tの間に投稿されたコメントを表す多重集合である．

$$A_{c_{i,k,n},t} = \left\{ c_{i,k',n'} \middle| c_{i,k,n} - \frac{t}{2} \le c_{i,k',n'} \le c_{i,k,n} + \frac{t}{2} \right\}$$

$w(c_{i,k,n}, t)$の値は$c_{i,k,n}$によって決定される．

$$w(c_{i,k,n}, t) = \begin{cases} 0.5 & \left(\begin{array}{l}\text{if } c_{i,k,n} \le \frac{3}{2}t \text{ or} \\ v_i - \frac{3}{2}t \le c_{i,k,n}\end{array}\right) \\ \dfrac{2|A_{c_{i,k,n},t}|}{|A_{c_{i,k,n}-t,t}| + |A_{c_{i,k,n}+t,t}|} & \text{(otherwise)} \end{cases}$$

v_iは動画iの再生時間であり，$w(c_{i,k,n}, t)$は$c_{i,k,n}$が動画の初めと終わりの$\frac{3}{2}t$以内であれば値を 0.5 とする．それ以外の時間であれば，動画上の時間$c_{i,k,n}$を中心に幅tの間に投稿されたコメント数と，その時間帯の外側のそれぞれ幅tの間に投稿されたコメント数の平均の比を値とする．本論文の実験では，時間幅tの値として 10 と 20 を用いた．

5. 実験

5.1. データセット

本研究では，国立情報学研究所の IDR データセット提供サービスにより株式会社ドワンゴから提供を受けた「ニコニコ動画コメント等データ」を利用した [18]．また，ニコニコ動画 API を用いて追加のコメントデータを取得した．「ニコニコ動画コメント等データ」は動画のメタデータ（タイトル，動画説明文，カテゴリー，タグ，投稿時間など）とコメントのメタデータ（コメント本文，投稿時間など）で構成されている．カテゴリーは 2018 年 11 月 8 日時点で音楽，アニメ，実況プレイ動画などの 37 種類であった．実験において，このカテゴリーデータを正解のラベルとして用いている．この提供されたデータには本手法で必要なユーザ ID や画像ベースの手法で用いるサムネイルは含まれていない．そのため，ニコニコ動画 API を用いてそれらのデータを収集した．実験で対象とした動画の条件は表 1 である．Dataset5 は 2018 年 9 月に投稿された動画の内，5 人以上のユーザがコメントをした 45274 件の動画を対象としている．Dataset50 と Dataset100 はそれぞれ，50 人以上，100 人以上のユーザがコメントした 9470 件と 4628 件の動画からなる．

5.2. 準備

提案手法の評価を行うため，NMI [4]による定量的評価と主観実験による定性的評価を行った，定量的評価は提案手法の妥当性を評価するために行う．データセットには正解となるラベルデータが存在しないため，3 つの既存手法の結果（LDA [1], GSDMM [2], IIC [3]）と動画のカテゴリーメタデータを正解のラベルとして用いた．主観実験は提案手法が，ユーザの知識や背景をもとに動画分類することの利点を評価するために行う．以下は実験に用いた既存手法（Uchida らの手法，LDA，GSDMM，IIC）の説明である．

4. Uchida らの手法は提案手法のもととなった手法である．本論文では，Uchida らの手法でのツイートにリツイートしたユーザが提案手法での動画にコメントしたユーザと対応している．Twitter ではユーザは同じツイートを 1 度しかリツイートできないため，$r_{i,k}$の値はユーザkがツイートiをリツイートしたかどうかにより 0 か 1 となり，本実験でも同様にコメントしたか否かで 0 か 1 の値をとる．行ベクトル$\boldsymbol{r}_i$は$r_{i,k}$により定義され，類似度は以下のようにシンプソン係数により計算される．

$$\mathrm{sim}(\boldsymbol{r}_i, \boldsymbol{r}_j) = \frac{\boldsymbol{r}_i \cdot \boldsymbol{r}_j}{\min(|\boldsymbol{r}_i|, |\boldsymbol{r}_j|)}$$

5. Latent Dirichlet Allocation (LDA)は 1 つの文書は複数のトピックからなると仮定した言語モデルであり，文書の集合に存在する隠れたトピックをもとに文書の分類を行う．LDA は確率分布（各トピックの単語の分布と，各文書におけるトピックの分布）を推定し，文書がどのトピックに所属する

かを計算する．本実験では，動画のタイトルのテキストを文書とした方法（LDA-T）と動画のタイトルと動画説明文のテキストを文書とした方法（LDA-TD）の 2 種類を用い，各文書における単語の値は tf-idf を用いて計算した．

6. A collapsed Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model (GSDMM)は短いテキストのクラスタリング手法である．Dirichlet Multinomial Mixture model [19]は生成プロセスに関する 2 つの仮定をもとにした文書の確率生成モデルである．GSDMM は初めに分割するクラスタ数を大きく設定すると，自動的に適切なクラスタ数を検出し分割することができる．その場合，多くのクラスタは要素数 0 となる．実験では，動画のタイトルを文書として扱った．
7. Invariant Information Clustering（IIC）は相互情報量を最大化することで学習を行う，深層学習を用いた画像クラスタリング手法である．IIC では元の画像と，その画像に加工を加えた画像をペアの入力とし，出力の相互情報量をもとに学習する．そのため，ラベルのないデータに対して学習を行うことができる．学習に画像を用いるため，データセットのサイズがテキストベースやユーザベースの手法に比べて大きくなる．そのため本実験では，Dataset100 の動画のサムネイルを入力として IIC を適用した．

LDA と GSDMM では各文書に含まれる単語が必要となる．そのため，日本語形態素解析器である MeCab を用いて文書の単語への分割を行った．LDA はあらかじめ分割するクラスタ数を決める必要がある．提案手法で用いた Louvain 法はクラスタ数をあらかじめ定めない手法であり，比較の際にはクラスタ数が同じであることが望ましい，しかし，各データセットにおいて LDA は動画をおよそ 100 クラスタに分割し，それ以上の値を設定しても分割するクラスタ数が頭打ちとなった．そのため，LDA での事前に定めたクラスタ数は 100 とした．また，LDA において 1 つの動画が複数のクラスタへ同じ確率で所属する場合があった．その際はその動画を要素数 1 の新しいクラスタとみなした．

5.3. NMI による定量的評価

提案手法の妥当性を評価するため，NMI による評価を行った．NMI は 2 つのクラスタ間の相互情報量を 0 から 1 の値に正規化したものであり，1 が完璧な関係を示す．NMI の値はクラスタの質を評価するために用いられ，クラスタ数の異なる 2 つの結果を評価することができる．2 つのクラスタリング結果λ_1とλ_2が与えられた場合，NMI の値は以下の式で表される．Nはデータセット内の要素数を，n_Xとn_Yはそれぞれλ_1のクラスタXとλ_2のクラスタYの要素数を表す．$n_{X,Y}$はλ_1のクラスタXとλ_2のクラスタY両方に属する要素数である．

$$\mathrm{NMI}(\lambda_1,\lambda_2)=\frac{\sum_{X\in\lambda_1}\sum_{Y\in\lambda_2} n_{X,Y}\log\frac{Nn_{X,Y}}{n_X n_Y}}{\sqrt{\left(\sum_{X\in\lambda_1} n_X\log\frac{n_X}{N}\right)\left(\sum_{Y\in\lambda_2} n_Y\log\frac{n_Y}{N}\right)}}$$

表 2，表 3，表 4 はそれぞれ Dataset5，Dataset50，Datast100 の NMI を示している．実験では正しいラベルとする既存手法として LDA, GSDMM, IIC を選択し，それらのクラスタリング結果とカテゴリーメタデータを正解のラベルとして扱った．

表 2　Dataset5 の NMI

		Uchida らの手法	提案手法 $t=10$	提案手法 $t=20$
	ラベル数	20779	21110	21120
LDA-T	100	0.520	0.554	0.555
LDA-TD	99	0.384	0.410	0.411
GSDMM	5089	0.788	0.831	0.832
Category	37	0.435	0.453	0.453

表 3　Dataset50 の NMI

		Uchida らの手法	提案手法 $t=10$	提案手法 $t=20$
	ラベル数	2259	2268	2265
LDA-T	99	0.460	0.560	0.561
LDA-TD	62	0.344	0.411	0.414
GSDMM	1925	0.697	0.802	0.802
Category	37	0.388	0.437	0.438

表 4　Dataset100 の NMI

		Uchida らの手法	提案手法 $t=10$	提案手法 $t=20$
	ラベル数	803	943	948
LDA-T	89	0.479	0.580	0.583
LDA-TD	55	0.356	0.432	0.432
GSDMM	414	0.659	0.773	0.775
IIC	191	0.322	0.422	0.427
Category	37	0.373	0.429	0.433

表 2，表 3，表 4 より，提案手法は Uchida らの手法に対してより高い NMI の値を示した．これにより，提案手法が Uchida らの手法よりも既存手法やカテゴリーメタデータをカバーする効果が高いことを示している．各結果において時間幅tの値を変えても，提案手法で得られた NMI の値に大きな差は見られなかった．これはデータが疎あるためと考える．より多くのユーザがコメントした動画を含むデータセットを作成することで，より明確な違いが出てくると考えられる．

表 2 において，提案手法と Uchida らの手法では，既存の手法に比べてクラスタ数が非常に大きくなっている．これはコメント数に対してコメントしたユーザが多いためデータが疎となり，要素数の少ない動画が

増えたことにあると考えられる．実際に Dataset5 では 17053 個の動画が要素数 1 のクラスタとなっている．これらの動画はデータセット内の他の動画と類似度が 0 であり，既存の手法と比較してクラスタ数に差が生じている．

5.4. 主観実験による定性的評価

Uchida らの実験 [16]をもとに．主観実験による定性的評価を行った．実験の概要は図 4 である．

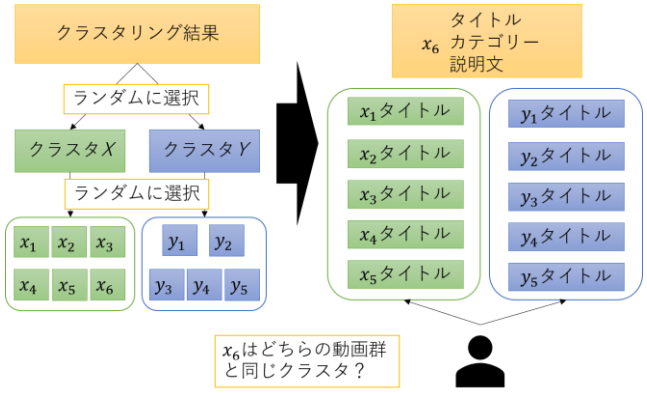

図 4　主観実験の概要

実験では，クラスタリング結果の適合率を以下のタスクにより評価した．はじめに，クラスタリング結果からクラスタXとクラスタYをランダムに選択する．その後，クラスタXより 6 つの動画x_1からx_6を，クラスタYより 5 つの動画y_1からy_5をそれぞれランダムに抽出する．x_6をターゲットの動画とし，被験者は「x_6がx_1からx_5の動画群と，y_1からy_5の動画群のどちらと同じクラスタに属するか」という質問に回答する．このタスクを繰り返すことで，被験者が正しいと思うクラスタリング結果を得ることができる．この回答をもとに，正解した割合を再現率として扱う．

この実験において，クラスタXとクラスタYはクラスタに属する動画の数とデータセットの動画数の比による確率でそれぞれ選択される．このタスクにおいて与えられる情報は，x_6に関してはタイトル，カテゴリー，動画説明文であり，x_1からx_5とy_1からy_5は動画のタイトルのみが与えられる．これらの情報は被験者が素早く回答できるために付与される．被験者として，著者の 1 人が Dataset5 と Dataset50 に関しては 6 つの手法（LDA-T，LDA-TD，GSDMM，Uchida らの手法，提案手法$t = 10$，提案手法$t = 20$）のそれぞれに対して 500 タスクを行い，Dataset100 に関しては IIC を加えた 7 つの手法に対して 500 タスクを行った．

図 5 は各データセットの結果である．Dataset5 では提案手法$t = 10$が最も良い結果であり，続いて GSDMM，提案手法$t = 20$という順であった．Dataset50 では GSDMM が最もよく，次いで提案手法$t = 10$，提案手法$t = 20$という順であった．Dataset100 では GSDMM，提案手法$t = 20$，提案手法$t = 10$という順であった．このタスクは動画のテキスト情報，主にタイトルをもとにして行われるため，テキストベースの手法は良い結果となる傾向にある．その条件において，提案手法がテキストベースの手法と同程度かそれ以上の結果となったことは提案手法の良さを表している．提案手法はクラスタリングにテキスト情報を用いていないが，ユーザの共通する興味によってテキスト情報の似た動画を同じクラスタとして分割することができている．この実験では，提案手法における時間幅tの値を変えたことによる結果の大きな違いは生じなかった．これも上述したデータが疎となることが関係すると考える．

5.5. 各クラスタの評価

提案手法のクラスタリング結果の詳細について説明する．表 5 は Dataset5 に対して提案手法$t = 10$を適用した結果の内，要素数が上位 10 クラスタの動画内容である．

表 5 のクラスタ No. 2 は，動画の内容に関してユーザの共通する興味をもとにしなければ分割できない，提案手法に特有の例である．このクラスタでは，半分以上の動画が新作のゲームやアニメに関するものである．ゲームに関する動画はゲームのイベントで公開されたプロモーションビデオや制作会社が直接公開した

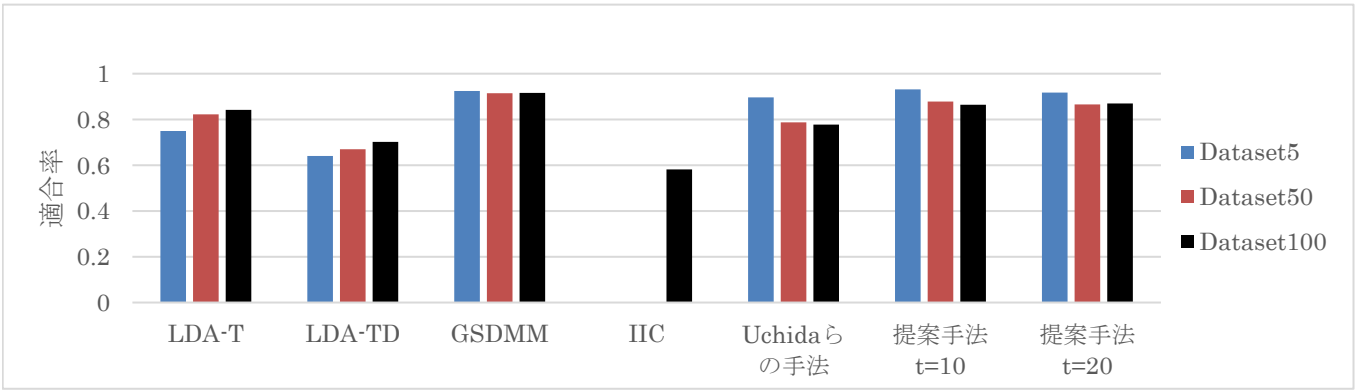

図 5　主観実験の結果

ものであり，その映像を視聴したユーザのリアクションが付随されているものもある．アニメに関する動画はすでに放送されているアニメのコマーシャルや，放送予定のアニメのビデオクリップなどである．また，同クラスタの Fate/Grand Order はゲームのタイトルであり，ゲームの新しいコンテンツについての動画である．これらの動画にはテキストや画像の類似はあまりないが，ユーザのゲームやアニメの新しいコンテンツに対する興味を反映し，1 つのクラスタとして分割することができている．

表 5　Dataset5 での提案手法$t = 10$の上位 10 クラスタの動画内容（手動で集計）

クラスタ No.	動画数	動画の内容（手動で集計）
1	590	Fortnite (43.4%) Splatoon (6.1%)
2	283	新作ゲーム (50.1%) Fate/Grand Order (9.5%) 新作アニメ (7.4%)
3	202	Azurlane (26.3%) VOICELOID 実況プレイ (10.9%)
4	195	声優(79.5%)
5	150	車/バイクの車載動画(87.3%)
6	148	KPOP (58.9%) 歌ってみた (7.4%)
7	145	仮面ライダー (19.3%) Dead by Daylight (14.5%)
8	143	ガンダム (70.6%)
9	138	Splatoon (34.1%) KPOP (15.9%)
10	126	うたってみた (53.1%) おそ松さん (12.7%)

クラスタ No. 6 と No. 9 には韓国のポピュラーミュージック（KPOP）に関する動画が含まれるが，クラスタの大部分は異なる．クラスタ No. 6 は歌ってみたや踊ってみたの動画が含まれ，クラスタ No. 9 にはゲームに関する動画が含まれている．加えて，クラスタ No. 9 には KPOP アーティストが出演するバラエティー番組が含まれており，これらは直接音楽に関係しない動画であった．これらより，クラスタ No. 6 の動画に対してコメントしたユーザは音楽に対して強い興味があり，クラスタ No. 9 は動画制作者や動画の企画に対して興味を持ったユーザがコメントしたと考える．

6. 議論

5.4 節の主観実験では，著者の内の 1 人のみが被験者として実験を行っている[1]．主観実験は複数の被験者により行われることが好ましい．しかし，被験者にはニコニコ動画に対する専門的な知識（アニメ，ゲーム，音楽など）が必要となるため，本論文では 1 人のみが実験を行った．主観実験の難しい例として図 6 がある．このタスクにおけるターゲットの動画（x_6に相当）は「Splatoon2」とタイトルにあり，カテゴリーは「実況プレイ動画」である．この例におけるテキストのみから判断すると，タイトルに実況プレイが含まれている右（青色）の動画群と類似しているように見える．しかし，ターゲットの動画は左（緑色）の動画群と同じクラスタである．この例に正しく回答するには，ターゲット動画のタイトルにある「Splatoon2」と左の動画群の「フォートナイト（Fortnite）」が TPS という同じジャンルのゲームであり主に若い年代に人気があることや，「スパッタリー」や「チャージャー」，「リッター」といった単語が「Splatoon2」に登場する固有名詞であるという知識が必要となる．そのため，主観実験はより多くのニコニコ動画に詳しい被験者による実験が必要である．

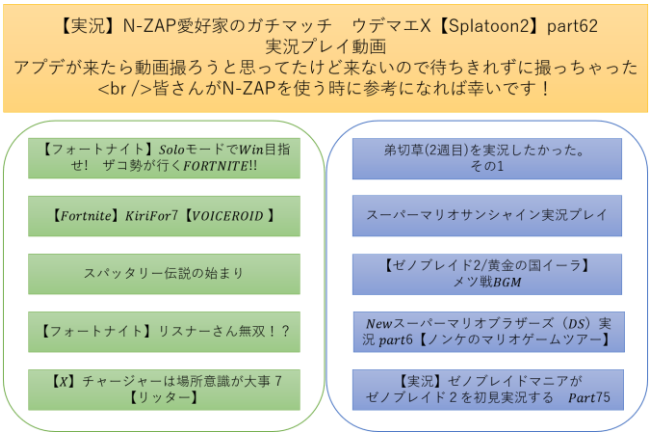

図 6　主観実験で判断が難しい例

提案手法のようにユーザのレビューを用いる手法には，レビューの少ないアイテムを取り扱うことが難しいという問題がある．実験において，Dataset5 の 37.7%，Dataset50 の 19.5%，Dataset100 の 13.9%の動画がデータセット内の他のすべての動画と類似度が 0 となった．これは，データセット内のコメントしたユーザの多くが 1 度しかコメントしていないことによると考える．この問題を解決するために，提案手法に他の情報を組み合わせる必要がある．現在のデータセットでは，タイトルなどのテキスト情報や画像などを使うことができる．しかし，これらの情報を組み合わせることで結果が既存手法に大きく近づいてしまう可能性があるため，本研究では用いなかった．

他のクラスタリング手法を提案手法に用いること

[1] 小さな規模の実験（提案手法$t = 10$，LDA-T，GSDMM のそれぞれに 100 タスク）を他の被験者 1 人が行い，著者の結果と似たような結果となった．

で，より正確な結果を得られる可能性がある．提案手法では，大きなネットワークを効率的に分割するために Louvain 法をクラスタリング手法として用いている．これは，一般的に動画にコメントしたユーザは数多く存在し，それに伴いネットワークが大きくなるためである．しかし近年，グラフのクラスタリングに対して多くのアプローチが提案されている．例えば，深層学習をもとにした DANMF [20]はオリジナルのネットワークと最終的なコミュニティとの階層的マッピングを学習することでネットワークのコミュニティを検出し，また効率よく学習することができる．行列分解をもとにした CNMMA [21]は，ネットワークのクラスタをノードの属性の共起から発見し，頂点数nのネットワークを計算量$O(n^2)$で計算することができる．これらの手法 Louvain 法の代わりに用いることで，より精度のよい結果を得ることができる可能性がある．

7. 結論

本論文では，TSC のコメント分布によるニコニコ動画に投稿された動画のクラスタリング手法を提案した．提案手法は動画をコメントしたユーザの集合として扱い，コメント分布の特徴によってユーザの重要度を計算した．定量的評価と定性的評価により，提案手法は既存のテキスト，画像ベースの手法や提案手法のもととなったユーザベースの手法に対して同等かそれ以上の結果を示した．今後の課題として，定性的評価に用いた主観実験の複数の被験者による実施や，コメントしたユーザ以外のデータを組み合わせることによるコメント数の少ない動画に生じる問題の解決，他のクラスタリング手法の適用による結果の評価が必要である．

参 考 文 献

[1] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *The Journal of Machine Learning Research,* vol. 3, pp. 993-1022, 2003.

[2] J. Yin and J. Wang, "A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering," *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 233-242, 2014.

[3] X. Ji, H. F. Joao and A. Vedaldi, "Invariant Information Clustering for Unsupervised Image Classification and Segmentation," *Proc. IEEE/CVF International Conference on Computer Vision (ICCV),* pp. 9864-9873, 2019.

[4] A. Strehl and J. Ghosh, "Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions," *Journal of Machine Learning Research 3,* pp. 583-617, 2002.

[5] A. Konishi and H. Hosobe, "Clustering Nico Nico Douga Videos by Using the Distribution of Time-Synchronized Comments," *Proc. IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT),* 2020.

[6] X. Zhou, L. Chen, Y. Zhang, D. Qin, L. Cao, G. Huang and C. Wang, "Enhancing Online Video Recommendation Using Social User Interactions," *The VLDB Journal,* vol. 26, pp. 637-656, 2017.

[7] A. Jansen, J. F. Gemmeke, D. P. W. Ellis, X. Liu, W. Lawrence and D. Freedman, "Large-Scale Audio Event Discovery in One Million YouTube Videos," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 786-790, 2017.

[8] C. Daniel, "Thematic Exploration of YouTube Data: A Methodology for Discovering Latent Topics," *Muma Business Review,* vol. 1, pp. 141-155, 2017.

[9] S. Roy and S. C. Guntuku, "Latent Factor Representations for Cold-Start Video Recommendation," *Proc. ACM Conference on Recommender Systems (RecSys),* pp. 99-106, 2016.

[10] Z. Zhao, L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy and X. Yi, "Recommending What Video to Watch Next: A Multitask Ranking System," *Proc. ACM Conference on Recommender Systems (RecSys),* pp. 43-51, 2019.

[11] H. Ren and D. Wang, "TRRS: Temporal Recurrent Recommender System Based on Time-Sync Comments," *Proc. International Conference on Machine Learning and Soft Computing (ICMLSC),* pp. 123-127, 2019.

[12] K. Tsukuda, H. Masahiro and M. Goto, "SmartVideoRanking: Video Search by Mining Emotions from Time-Synchronized Comments," *Proc. IEEE International Conference on Data Mining Workshops (ICDMW),* pp. 960-969, 2016.

[13] W. Yang, W. Gao, X. Zhou, W. Jia, S. Zhang and Y. Luo, "Herding Effect Based Attention for Personalized Time-Sync Video Recommendation," *Proc. IEEE International Conference on Multimedia and Expo (ICME),* pp. 454-459, 2019.

[14] Q. Bai, Q. Hu, G. Fang and L. He, "Topic Detection with Danmaku: A Time-Sync Joint NMF Approach," *Database and Expert Systems Applications,* vol. 11030, pp. 428-435, 2018.

[15] W. Yang, N. Ruan, W. Gao, K. Wang, W. Ran and W. Jia, "Crowdsourced Time-Sync Video Tagging Using Semantic Association Graph," *Proc. IEEE International Conference on Multimedia and Expo (ICME),* pp. 547-552, 2017.

[16] K. Uchida, F. Toriumi and T. Sakai, "Evaluation of Retweet Clustering Method: Classification Method Using Retweets on Twitter without Text Data," *Proc. IEEE/WIC/ACM International Conference on Web Intelligence (WI),* pp. 187-194, 2017.

[17] V. D. Blondel, J.-L. Guilaume, R. Lambiotte and E. Lefebvre, "Fast Unfolding of Communities in Large Networks," *Journal of Statistical Mechanics: Theory and Experiment,* vol. 2008, no. 10008, pp. 1-12, 2008.

[18] 株式会社ドワンゴ, "ニコニコ動画コメント等データ," 国立情報学研究所 情報学研究データリポジトリ, 2018. [Online]. Available: https://doi.org/10.32130/idr.3.1.

[19] K. Nigam, A. K. Mccallum, S. Thrun and T. Mitchell, "Text Classification from Labeled and Unlabeled Documents using EM," *Machine Learning,* vol. 39, pp. 103-134, 2000.

[20] F. Ye, C. Chen and Z. Zheng, "Deep Autoencoder-Like Nonnegative Matrix Factorization for Community Detection," *Proc. ACM International Conference on Information and Knowledge Management (CIKM),* pp. 1393-1402, 2018.

[21] T. He, K. C. C. Chan and L. Yang, "Clustering in Networks with Multi-Modality Attributes," *Proc. IEEE/WIC/ACM International Conference on Web Intelligence (WI),* pp. 401-406, 2018.