

視聴者はドラマの「どこ」に「どう」反応しているのか？ —実況ツイートを利用したドラマのシーン特徴の抽出—

牛島 実桜[†] 南 大智^{††} 牛尼 剛聡^{†††}

[†]九州大学芸術工学部 〒815-8540 福岡県福岡市南区塩原 4-9-1

^{††}九州大学大学院芸術工学府 〒815-8540 福岡県福岡市南区塩原 4-9-1

^{†††}九州大学大学院芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: [†]1DS13175E@s.kyushu-u.ac.jp ^{††}minami@kyudai.jp ^{†††}ushiana@design.kyushu-u.ac.jp

あらまし TVドラマには「恋愛もの」、「刑事もの」等、様々なジャンルが存在する。しかし、最近のTVドラマは単一のジャンルに分類できず、複数のジャンルの特徴が混在するものが少なからず存在する。そこで、本研究では、TVドラマ放送中のTwitterの実況ツイートから、TVドラマの進行に沿ったユーザの反応を分析し、TVドラマのシーンの特徴を明らかにする。そして、シーンの特徴に基づいて、TVドラマ全体に対して、ジャンルとは異なるドラマの「展開パターン」に基づいて、クラスタリングする。具体的には、TVドラマ中の窓区間中に投稿された実況ツイートを一つの文書としてLDAを用いてトピックを抽出し、それぞれのシーンの特徴をトピック分布としてモデル化する。

キーワード SNS, Twitter, ソーシャルビューイング, TV, ドラマ, シーン特徴, LDA

1. はじめに

近年、SNS(ソーシャルネットワークサービス)が広く普及し、老若男女を問わず、多くの人がSNSを日常的に利用するようになった。中でも、Twitterは代表的なSNSの1つであり、コミュニケーションや情報収集など、様々な目的のために世界中で利用されている。

近年、SNSの利用方法の一つとして、「ソーシャルビューイング」が注目されている。ソーシャルビューイングは、SNSへの投稿を行いながらTV番組を視聴する視聴形態である。ソーシャルビューイングでは、SNSを通して他のユーザとリアルタイムで感想や意見の共有を行うことができる。ソーシャルビューイングの普及に伴い、SNSとの連携機能を搭載したテレビ[1]や、ソーシャルビューイングを支援するサービスやアプリとして、JoinTV[2]やemocon[3]などが開発されている。

現在、ソーシャルビューイングのためのSNSとして、Twitterが一般的に用いられている。また、ソーシャルビューイングの対象になるテレビ番組の種類としては、テレビドラマやテレビアニメなどがあげられる。ソーシャルビューイングのために投稿されたツイートは、対象とするドラマのタイトルやその略称を含むハッシュタグにより、リアルタイムに取得できる。

近年、TVドラマはジャンルとしての分類が複雑化しており、単一のジャンル名を厳密に付与することが困難になっている。例えば、恋愛ドラマが社会派ドラマの要素を含んでいたり、刑事ドラマと宣伝されていたものが、実は事件を解決することよりもコメディ感

を重視しているものであったりと、ジャンルと実際のドラマの内容との差異が存在する場合も多い。こうした中で、従来の単純なジャンル分類によらず、ドラマの特徴を適切に抽出することが重要になっている。ドラマの特徴が抽出できれば、検索、推薦の高度化等、様々な応用が期待できる。

本論文では、Twitterを利用したTVドラマのソーシャルビューイングに注目し、ドラマ放送中に投稿された実況ツイートをを用いて、ドラマの時系列でのシーンの特徴の抽出を行う。これにより、ジャンルとは異なるドラマの「展開パターン」でTVドラマの特徴付けを行う。

2. 関連研究

これまでにも、Twitterの実況ツイートを利用し、テレビ番組との関連性を抽出する研究が行われている。

若井ら[4]は、感情語辞書を用い、実況ツイートからテレビで放送された映画の時系列毎の視聴者の感情値を、Twitterの特有表現を考慮して算出する方法を提案している。また、山内ら[5]は感情極性辞書を利用し、テレビ番組に関連したツイートを対象に、実況ツイートがつぶやかれた際のシーンの視聴者の感情を推定し、感情をグラフとして視聴者に提示するシステムの提案を行っている。本研究では、感情に着目するのではなく、視聴者が示す反応に焦点をあて、反応のパターンを見出す。

感情以外でテレビ番組のシーンに特徴付けを行う研究として、中澤ら[6]の研究がある。この研究では、実況ツイートを解析することで、テレビ番組の重要シ

ーンを検出を行い、またそのシーンの内容を表すラベルを付与している。また、中原ら[7]はテレビ番組の実況ツイートを解析することにより、ツイートの投稿数が急激に増えたシーンと、台詞と投稿内容が一致したシーンにおいて、解析者が興味を持つと考えられるツイートを要約して示す方法を提案している。しかし、これらの研究は、時系列でテレビ番組の内容を追うのではなく、突出してツイート投稿数が多いシーンに注目する点で本研究とは異なる。

大田垣ら[8]は、Twitterでのソーシャルビューイングにおいて、リアルタイムにタイムラインの盛り上がりを検出し、類似したツイートを集約して、ユーザがタイムラインの特徴を捉えやすいよう提示する手法を提案している。本研究は、視聴者の反応をもとに作成したトピックモデルの分布の推移によりTVドラマの特徴を得るという点で異なる。

上記を踏まえ、本研究では視聴者の実況ツイートでの反応をもとに、時系列でのトピックの分布の推移によってTVドラマのシーンの特徴を明らかにすることを目的とする。

3. アプローチ

本研究では、TVドラマ放送中の視聴者の反応をTwitterの実況ツイートから分析することで、ジャンルとは異なるTVドラマの「展開パターン」により、ドラマのシーンの特徴を見出すことを目的とする。

本研究では、ドラマの特徴をモデル化するために、ユーザ（視聴者）の反応パターンを考える。つまり、TVドラマには、ユーザに対して特定の印象を与えるいくつかのパターンが存在し、それらの組み合わせによりTVドラマは構成されていると考える。

このように考えるとき、TVドラマに対するユーザの反応パターンを規定することが重要である。これまでに、ユーザの反応を「喜・怒・哀・楽」に代表されるセンチメントとして分類することが考えられている。関連研究でも示したように、このアプローチに基づいた多くの研究が行われている。

しかし、TVドラマに対するユーザの印象は、それらのセンチメントの組み合わせとして表現することは困難であると考えられる。その理由として、TVドラマは、日常生活の中での息抜きとしての役割が期待されているものが少なからず存在していることが挙げられる。それらは必ずしも、「喜・怒・哀・楽」に代表されるセンチメントの要素の分布に還元するのが適切でない場合があるのではないかと考える。

上記の問題を解決するために、本研究では、複数のドラマに対するユーザの実況ツイートを分析することにより、様々なドラマの様々なシーンに共通して表れ

るユーザの典型的なパターンを、ボトムアップ的に抽出する。そして、ボトムアップ的に抽出した反応パターンの分布を利用して、ドラマのシーンの特徴をモデル化する。

本研究の概要を図1に示す。また、本研究の手順を以下に示す。

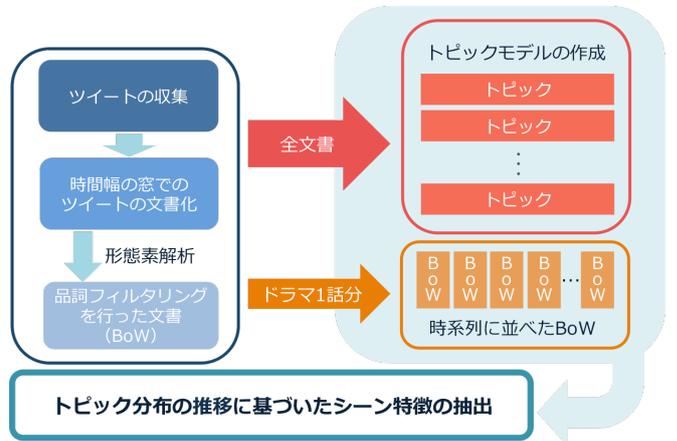


図 1 本研究のイメージ図

1. 対象とするTVドラマに対する実況ツイートを収集する。具体的には、対象とするドラマのタイトルをハッシュタグとして含むツイート、もしくはドラマのタイトルの略称をハッシュタグとして有するツイートを収集する。収集の対象となるのは、ドラマの放送時間中に投稿されたツイートである。収集にはTwitter APIを利用する。
2. TVドラマに対して時間的なセグメンテーションを行う。具体的には、窓関数を利用してドラマ中の特定の時間区間を規定し、規定された時間区間に投稿されたツイートを抽出する。それぞれの時間区間中に投稿されたツイートを連結して一つの文書として扱う。
3. それぞれの文書に対して形態素解析を行う。
4. 形態素解析の結果に基づいて文書をBoW(Bag of Words)として表現する。ここで、すべての単語を使用するわけではなく、使用する品詞のフィルタリングを行う。
5. Latent Dirichlet Allocation (LDA) [10]を用いて文書からトピックモデルを作成し、それぞれの時間区間に対するトピック分布を求める。それぞれの時間区間のトピック分布の推移をドラマのシーン特徴として捉える。
6. シーン特徴に対応する時間区間のツイートをかけ合わせるにより、シーンごとの特徴ベクトルを導く。また、特徴ベクトルの集合から、ドラマ毎の特徴を表すベクトルを算出する。

4. 提案手法

4.1. ドラマのモデル化

本研究では、TV ドラマ放送中の実況ツイートを用いてシーンの特徴抽出を行うため、放送されているシーンと同時刻に投稿されたツイートの対応付けを行う必要がある。そのため、TV ドラマのハッシュタグによって集められた実況ツイートを投稿時刻に基づいて時系列化する。次に、この時系列に対して、あらかじめ設定しておいた幅の時間窓を適用して、時系列をセグメントに分解する。このセグメントを、ドラマを構成する原始的な単位としてモデル化を行う。このセグメントをシーンセグメントと呼ぶ。

排他的なセグメント化を行うと、一つの意味のあるまとまりが分断されてしまい、ドラマの特徴が適切に抽出できない場合がある。この問題に対応するために、窓幅の半分が次の窓と重複するようにしてセグメント化を行う。

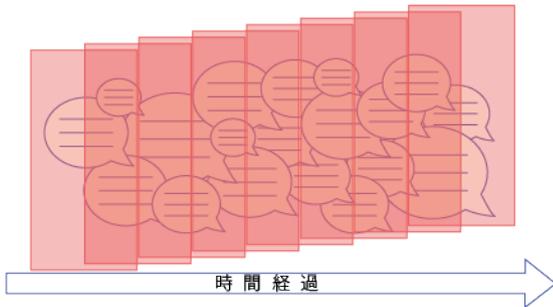


図 2 時間幅を利用したドラマのセグメント化

4.2. シーンセグメントに対する反応のモデル化

シーンセグメントに対応する反応をモデル化する。モデル化のためには Twitter に投稿された実況ツイートを利用する。それぞれのシーンセグメントが放送された時間区間上に設定した時間幅の窓で区切って作られた文書に含まれるツイートを、MeCab[9]による形態素解析を行い、ツイートに含まれる単語を抽出する。それぞれのシーンセグメントに対応して、その特徴を BoW としてモデル化する。

本研究では、ユーザの反応を抽出するための BoW として、全ての品詞を利用するわけではない。本研究では、個々のドラマの題材に依存せず、ユーザに対して与えた印象を表す単語を抽出したい。そのため、ドラマのタイトルや登場人物といった固有名詞だけでなく、ユーザの反応とは考えにくい単語を極力省きたい。

そこで、BoW を構成する単語を、品詞に基づいてフィルタリングする。形態素解析により分類した品詞の中から、シーンの特徴を抽出するために適切であると思われる、視聴者の反応が現れやすい品詞を選択して使用する。品詞の選択には、MeCab での形態素解析時

に割り振られる、品詞 ID を用いる。品詞 ID には、名詞や形容詞といった品詞名だけでなく、名詞の中でも一般名詞なのか、固有名詞なのかといった、より細かい分類が含まれている。

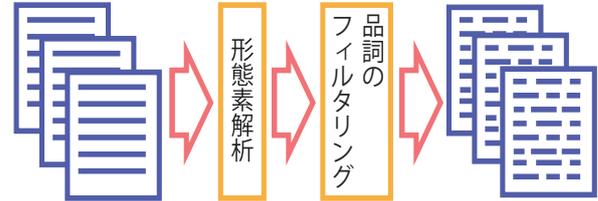


図 3 BoW を用いた反応のモデル化

4.3. シーンセグメントに対する反応のトピックモデルの構築

本研究では、ドラマのシーン特徴を抽出する際、前述の手法によって得られる BoW を利用して、シーンセグメントに対する反応のトピックモデルを構築する。本研究ではトピックモデルとして LDA を使用する。LDA を用いることで、文書中に一緒に出現する可能性のある単語の集合をトピックとして得ることができる。

トピックモデルをジャンルとは異なるシーン特徴を見出す指標とし、文書ごとにそれぞれのトピック分布を算出する。トピック分布とは、それぞれのトピックが出現する確率である。シーンセグメントのトピック分布を利用してシーン特徴を決定する。

4.3.1 コーパス

本研究では、対象とするドラマ集合の実況ツイートをコーパスとして、LDA を利用しトピックモデルを構築する。具体的には、ドラマの実況ツイートを特定の時間幅で区切り、文書化したものを形態素解析することによって得た BoW を、コーパスとして使用する。

4.3.2 シーンセグメントを特徴づけるトピックの抽出

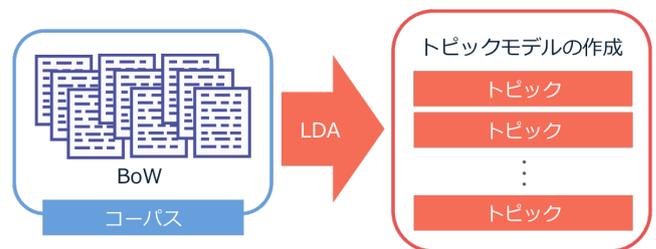


図 4 コーパスを用いたトピックの作成

コーパスを利用して、トピックを抽出する。抽出するトピックの個数はパラメータとして明示的に指定する。ここで抽出したトピックは、様々な種類に含まれる様々なシーンにおいて一般的に存在するトピックで

あり、ドラマに対するユーザ反応トピックと呼ぶ。

4.3.3 ユーザ反応トピックを利用したシーンセグメントの特徴付け

抽出したユーザ反応トピックを利用して、シーンセグメントの特徴付けを行う。具体的には、上記の手法で抽出したユーザ反応トピックを利用し、それぞれのシーンセグメントに対するユーザ反応トピックの出現確率に、そのシーンセグメントに対応する窓区間に含まれるツイート数をかけ合わせ、得られる数値の分布を求める。この分布をベクトルとしたものをシーンセグメントに対する特徴ベクトルとする。

LDA によって得られるユーザ反応トピックの分布からは、対象とする文書中でそれぞれのトピックが出現する確率を得ることができる。しかし、その時点でのドラマの盛り上がりがどの程度かを知ることはできない。そのため、本研究ではシーンセグメントに対するユーザ反応トピックの分布に対し、対応する窓区間に含まれるツイート数をかけ合わせる。これにより、実況ツイートの数の推移から得ることのできるドラマの盛り上がりの様子を含んだ、シーンセグメントの特徴ベクトルを作成する。

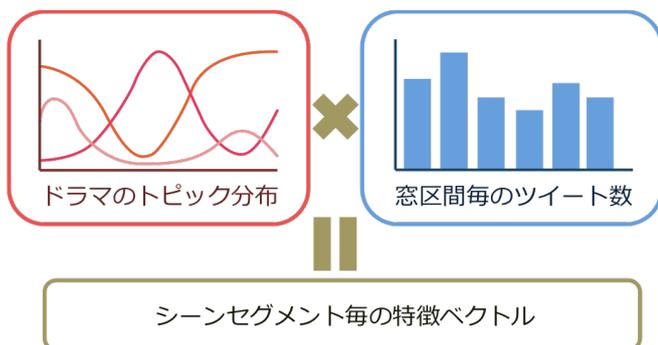


図 5 特徴ベクトルの算出

4.4. 各話の特徴を用いたドラマの特徴の算出

本研究では、特徴ベクトルをドラマごとに全話分集め、各ドラマの特徴を抽出する。具体的には、先述の手法で求めたドラマ 1 話毎の特徴ベクトルを、それぞれのトピックにおいて全て足し合わせる。得られた 30 トピック分の特徴ベクトルの和について、それらを足し合わせる際、和が 1 となるようにそれぞれのトピックにおいて正規化を行う。これにより、1 話の中で全トピックを相対的に見た場合に、ツイート数の盛り上がりを殺すことなく、特徴ベクトルを均一な数値にそろえることが可能となる。これらをドラマごとに全話分集めたものを、それぞれのドラマのドラマベクトルとし、ドラマの特徴を捉える指標とする。

5. 予備実験

提案手法の有効性の検証に先立ち、検証条件を定めるための 3 つの実験を行った。

5.1. シーンセグメントのための窓幅の決定

シーンセグメントを構成するための窓の時間幅を変更した場合の効果について検討する。時間幅を 1 分にして文書化しトピックモデルを得たものと、同様に窓の時間幅を 2 分にしたものの 2 種類の比較を行う。比較に使用した TV ドラマは表 1 の通りである。

この実験では表 1 に示したドラマと同時期である 2016 年 10～12 月に放送されていた 14 つのドラマの 1～8 話を用い、トピックを作成した。

表 1 実験の比較に使用した TV ドラマ

タイトル	放送日時	総ツイート数
地味にスゴイ！校閲ガール・河野悦子	2016年10月5日22時	3124
逃げるは恥だが役に立つ	2016年10月11日22時	20467

窓の時間幅が 1 分のとき、「地味にスゴイ！校閲ガール・河野悦子」はトピック分布にまとまりがなかったが、「逃げるは恥だが役に立つ」では全体的にトピック分布に波があることを確認した。対して、窓の時間幅が 2 分のときには、どちらのドラマでもトピック分布の波が確認され、時間幅が 1 分のときに比べるとシーンの特徴がトピック分布に表れていた。

この結果により、本研究では窓幅を 2 分に設定する。

5.2. 使用する品詞の決定

2016 年 10～12 月に放送されていた 14 つのドラマの実況ツイートを用い、形態素解析の後に BoW として表現する際、Mecab の品詞 ID により選択する品詞を変更して、LDA でトピックモデルを得る実験を行った。実験は、選択する品詞を表 2 のように変更して行った。

表 2 実験に使用した品詞

パターン	選択した品詞
①	動詞、自立 名詞、一般・サ変接続・形容動詞語幹 副詞、一般 形容詞、自立
②	動詞、自立 副詞、一般 形容詞、自立
③	名詞、一般・サ変接続・形容動詞語幹 副詞、一般 形容詞、自立
④	副詞、一般 形容詞、自立

品詞を選択する際、ある単語が出現する文書数を示す値である df 値が 25 を下回る単語と 7500 を上回る単語は除外している。ただし、例外として「かわいい」

という単語は df 値が 7500 を上回っているが、視聴者の反応を捉えるためには必要な単語であると判断し、あえて加えた。

結果として、①、②、③のパターンでは、それぞれトピックに含まれる単語がいずれかのドラマに依存している状態となっており、品詞の選択が適切であるとは言えない。

④は、トピックに含まれる単語の大多数が形容詞だが、本研究で必要としている視聴者の反応が反映されていると考えられる単語が多く含まれていた。また、特定のドラマによる影響が小さく、視聴者の反応を捉えることが容易なトピックが得られた。

以上より、本研究には④の副詞と形容詞を使用するパターンが最も適していると考えられる。

5.3. トピックの数の決定

得られるトピック数を変化させてトピックモデルを取得し、トピックの内容を比較することで、ユーザの反応を捉えるために適切なトピックの数を検討する。

トピック数は 10、20、30、40、50 の 5 種類に設定し、トピックモデルを得た。このとき、使用する文書は全て同じとした。品詞は形容詞とオノマトペを中心に抜粋した副詞を使用した。

トピック数 10、20 ではトピック数が少ないため、1 つのトピック内に複数もしくは数種類の視聴者の反応とみられる言葉が含まれている状態であった。

対してトピック数 40、50 では似た単語が含まれ、類似するトピックが多数存在しており、トピックにまとまりがなくなっていた。

この実験結果より、最もトピックがユーザの反応を分類できていると判断できた、トピック数 30 を利用する。

6. 検証

提案手法の有効性を評価するため、上記の提案手法に基づいて検証を行った。具体的には、以下の 3 種類の内容について検証を行った。

- ① シーンセグメントに対する特徴ベクトルがドラマのシーンの特徴を表しているか
- ② 連続ドラマの各話の特徴がドラマベクトルに表れているか
- ③ ジャンルと異なるドラマの類似度をはかれるか

6.1. 検証 1

6.1.1. 検証方法

提案手法に沿ってドラマ 1 話分のトピック分布を求め、算出した特徴ベクトルが実際のドラマに対する視聴者の反応を適切に表しているかを確認した。具体的

には、ドラマ内のどのようなシーンで何が起きているかを 30 秒毎に書き出したもの(シーンリストとする)と、提案手法によって得られた特徴ベクトルとを、対応する時間においてトピック通りにユーザの反応を表すことができるかを比較する。

前述の予備実験の結果により、窓の時間幅・使用する品詞・トピックの数を以下のように設定する。

- ① 窓の時間幅：2分
- ② 使用する品詞：形容詞、副詞
- ③ トピックの数：30

検証対象とした TV ドラマを表 3 に示す。

シーンリストと特徴ベクトルの比較の際、表 3 で記述した 3 つのドラマについて、得られた 30 個の特徴ベクトルの中から、特徴ベクトルの数値が高いものを 5 つ選出し、グラフ化したものを使用した。

この検証では、表 4 のドラマの全話を使用しトピックモデルを得た。表 4 で示したジャンルは、各ドラマの Web ページで紹介されていたものを使用している。

表 3 検証で比較に使用した TV ドラマ

タイトル	放送日時	総ツイート数
逃げるは恥だが役に立つ	2016年10月11日22時	20467
IQ246～華麗なる事件簿～	2016年12月4日21時	9670
THE LAST COP/ラストコップ	2016年12月10日21時	26673

表 4 2016年10～12月に放送されたドラマ一覧

ドラマ名	時間帯	ジャンル
カインとアベル	月曜21時	ラブストーリー
メディカルチーム レディ・ダ・ヴィンチの診断	火曜21時	医療ミステリー
逃げるは恥だが役に立つ	火曜22時	社会派ラブコメディ
相棒 Season15	水曜21時	刑事
地味にスゴイ! 校閲ガール・河野悦子	水曜22時	仕事
ドクターX～外科医・大門未知子～	木曜21時	医療
Chef～ミツ星の給食～	木曜22時	エンターテイメント
砂の塔～知りすぎた隣人	金曜22時	サスペンス
家政夫のミタゾノ	金曜23時台	ヒューマンドラマ
勇者ヨシヒコと導かれし七人	土曜0時台	冒険
THE LAST COP/ラストコップ	土曜21時	刑事アクションコメディ
IQ246～華麗なる事件簿～	日曜21時	ミステリー
キャリア～掟破りの警察署長～	日曜21時	刑事
レンタル救世主	日曜22時台	コメディ

6.1.2. 検証結果・考察

「逃げるは恥だが役に立つ」では、「かわいい」が含まれたトピックが軒並み高い数値を示していた。中でも、トピック中の「かわいい」の確率が 0.351 と非常に高いトピック 3 は、1 話を通して波を示しているだけでなく、開始 49～54 分や、開始 61～63 分の小さな「かわいい」が現れるシーン、そして開始 66 分頃に流

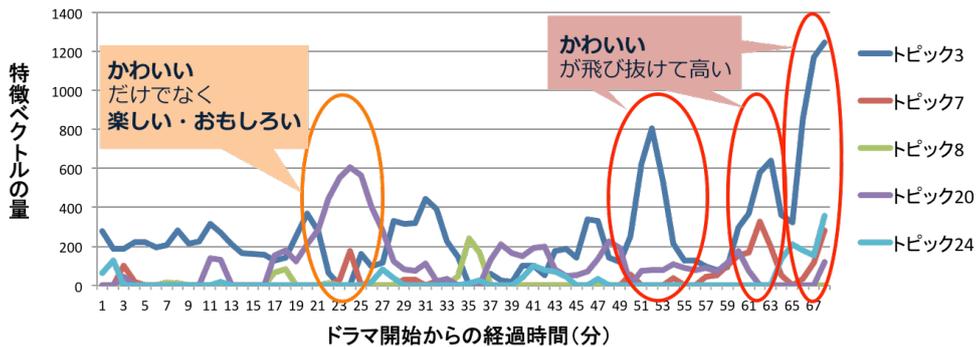


図 6 「逃げるは恥だが役に立つ」の1話中のトピックの推移

表 5 「逃げるは恥だが役に立つ」の上位5トピック

トピック	トピック内容(上位3単語)
3	0.351×かわいい、0.030×なんて、0.021×うい
7	0.090×楽しい、0.049×上手い、0.047×うまい
8	0.078×かわいい、0.036×早い、0.029×赤い
20	0.102×かわいい、0.053×楽しい、0.031×おもしろい
24	0.121×おもしろい、0.060×かわいい、0.050×おもしろい

れたエンディングの「恋ダンス」において数値が非常に高くなっている。また、情熱大陸のパロディが取り入れられた開始 21～24 分と、男性の仕事場でハプニングが起きた開始 35～38 分では、「かわいい」だけでなく「楽しい」「おもしろい」が含まれるトピック 20 が高くなっていた。これは、トピックモデルによりドラマの特徴を捉えたからこそ表れたものである。全体を通して、「逃げるは恥だが役に立つ」では「かわいい」という感想が非常に高かったが、シーンによって「かわいい」にも少しずつ違いがあることが確認できた。

「IQ246～華麗なる事件簿～」では、「かっこいい」が主となるトピックが前半から中盤にかけて高い数値を示し、後半に向かうにつれ、様々なトピックの反応が得られた。この回では、主人公が開始 3 分で逮捕されるが、それに対して「驚き」「どうして」といった反応よりも「かっこいい」を含むトピックの反応が大きい点が予想と異なった。このような「予想とは異なるがシーンとの結びつきが観測される」点が、特徴ベクトルを用いることによってドラマを特徴付ける際に、ジャンルとは異なる特徴を得られる点であると考えられる。

比較を行った 4 つのドラマの中で最も総ツイート数の多い「THE LAST COP/ラストコップ」では、シーンの盛り上がりが一と目でわかる特徴ベクトルの分布を得ることができた。

4 つのドラマのシーンリストとの比較を通して、ドラマのシーンと対応付いた特徴ベクトルを得ることができていると確認できた。また、特徴ベクトルを得る

ことによって予想とは異なる視聴者の反応を知ることができることもわかった。ジャンルとは異なるドラマの特徴を見出すにあたって、このような反応が得られることは重要であると考えられる。

6.2. 検証 2

6.2.1. 検証方法

提案手法に沿って求めた各話の特徴を表すドラマベクトルがドラマの内容を反映しているかどうか、および 1 クールのドラマ全体の特徴を表しているかどうかに関する検証を行った。

具体的には、表 4 に示すドラマそれぞれに対してドラマベクトルを求め、グラフ化する。グラフには、特徴ベクトルの和の値が 0.1 を超えているトピックを使用した。グラフと公式ホームページ上に公開されているあらすじを比較することにより、各話での特徴がドラマベクトル上に出現しているか、また全話を通して見た際にドラマの特徴が表れているかを確認する。

6.2.2. 検証方法

ここでは、4 つのドラマに対する結果と考察を示す。

「逃げるは恥だが役に立つ」は、「かわいい」を含むトピックが多く反応していることがわかり、その中でも「かわいい」が高確率となっているトピックが全話を通じて高い数値を維持している。しかし、第 8 話では、「切ない」「辛い」「悲しい」「もどかしい」といった反応を含むトピック 14 が、「かわいい」の確率が高いトピック 3 を上回っている。主人公が男性と同居している家から出て行き、男性が思い悩むシーンが多く描かれているためであると考えられる。このように、印象的なエピソードを含む放送回では、「かわいい」とは異なるトピックの数値が高くなっていることが確認できた。

同様に、「砂の塔～知りすぎた隣人」では、「怖い」「こわい」が大半を占めるトピックが第 6 話まで大きく推移していた。しかし、第 7 話でトピック構成が大きく変わっている。この原因として、前半ではタワー

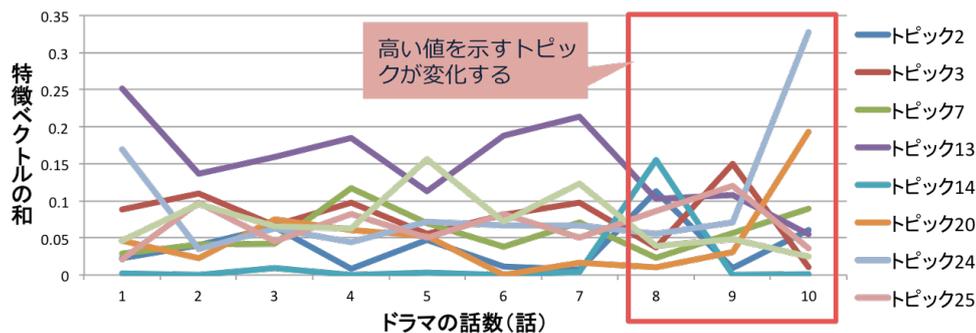


図 7 「THE LAST COP/ラストコップ」の各話の特徴

表 6 「THE LAST COP/ラストコップ」の抜粋トピック

トピック	トピック内容(上位3単語)
2	0.142×多い, 0.055×黄色い, 0.048×痛い
3	0.351×かわいい, 0.030×なんて, 0.021×うい
7	0.090×楽しい, 0.049×上手い, 0.047×うまい
13	0.194×カッコいい, 0.066×カッコよい, 0.033×カッコイイ
14	0.095×切ない, 0.055×早い, 0.051×辛い
20	0.102×かわいい, 0.053×楽しい, 0.031×おもしろい
24	0.121×おもしろい, 0.060×かわいい, 0.050×おもしろい
25	0.095×なんて, 0.048×ええ, 0.044×まさか
28	0.048×さすが, 0.043×悪い, 0.029×なんて

マンションでのママ友同士の陰湿ないじめやヒエラルキーに焦点があてられていたのに対し、後半に進むにつれて初期に親切だった隣人が実は悪人かもしれないという点に焦点があてられ、注目の集まるポイントが異なるからであると考えられる。それによって視聴者の示す反応が変化していったのであろう。

「THE LAST COP/ラストコップ」でも、上記の2ドラマと同じく、1つのトピックが前半で大きな数値を示している。しかし、「砂の塔～知りすぎた隣人」と同様に、後半にあたる第8話でトピック構成に変化が表れている。また、「逃げるは恥だが役に立つ」も、第8話でトピック分布に変化が見られる。このことから、ドラマでは終盤に差し掛かる辺りで、視聴者が示す反応に変化が出るような仕掛け・演出が施されていることが考えられる。つまり、終盤の変化は、ドラマを制作する上での特徴であることが推測される。

「勇者ヨシヒコと導かれし七人」では、突発的な特徴ベクトルの盛り上がりが多かった。上記の3つのドラマは、全体的にトピックに緩やかな波があった。しかし、「勇者ヨシヒコと導かれし七人」では、放送回毎に大きな反応を示しているトピックが異なる。これは、1話完結型のドラマである点が要因としてあげられる。1話完結型のドラマでは、放送回により雰囲気や印象、エピソードが大きく異なることにより、視聴者の反応

が固定されず、毎回異なるトピックが高い数値を示すことが考えられる。放送回により様々なトピックが反応するため、他のドラマに比べて特徴ベクトルの和の値が0.1を超えたトピックも必然的に多くなる。

この特徴は、何年も続くシリーズものである「相棒 Season15」や「ドクターX～外科医・大門未知子～」でも見られた。毎回変化があるため、視聴者が飽きずに視聴することができ、長く愛される、「長寿ドラマの特徴」であるとも考えられる。

各ドラマ各話から得られたドラマベクトルは、実際に放送されたドラマの内容と結び付きがあった。多くは1つのメインとなるトピックが存在していた。また、ドラマ後半や、1話完結型のドラマにおけるトピック構成に複数のドラマで同様の変化が見られた。これは、ドラマを制作する上での特徴が表れているものであると考えられる。

以上より、ドラマベクトルの動きからも、ドラマの特徴を得ることが可能であると考えられる。

6.3. 検証 3

6.3.1. 検証方法

提案手法に沿って求めたドラマベクトルを用い、コサイン類似度によってドラマ同士の類似度を測った。使用したドラマは表4で示した14つのドラマである。算出した類似度を各ドラマの公式 Web ページに記載されているジャンルでの類似と比較することで、ジャンルとは異なる指標での類似度が導き出しているか、確認を行った。

6.3.2. 検証結果・考察

表7は、表4で示した14つのドラマの中から、類似度の高い組み合わせ順に10個示したものである。

最も類似度が高かったのは、「キャリア～掟破りの警察署長～」と「レンタル救世主」である。ジャンルはそれぞれ「刑事」と「コメディ」の分類になっており、ジャンルから類似を想定することは難しいといえる。しかし、ドラマベクトルにおいて、どちらのドラマも「カッコいい」を主とするトピック13の値が高いものであった。また、どちらも主人公を男性俳優が務め、

毎回問題を解決していく形のドラマであり、ジャンルとは異なる類似性を検出できていると考えられる。

2番目に類似度が高いペアは、「THE LAST COP/ラストコップ」「レンタル救世主」であった。この2つのドラマはどちらもジャンルに「コメディ」が含まれており、一般的な印象に近いと考えられる。

3番目に類似度が高いペアは「砂の塔～知りすぎた隣人」と「家政夫のミタゾノ」である。この2つはそれぞれ「サスペンス」と「ヒューマンドラマ」となっており、ジャンルから類似は予測できない。しかし、どちらも人間関係や人間の醜さに焦点をあてたサスペンス系のドラマとなっており、類似しているといえる。

このように、ここで示した10個のペアの内、7つのペアは単純なジャンルとしては類似性が予測できず、提案手法により、ジャンル以外の指標での類似性が判定できる可能性が示された。

表 7 ドラマ同士の類似度

順位	ドラマ1	ドラマ2	コサイン類似度
1	キャリア～突破りの警察署長～(刑事)	レンタル救世主(コメディ)	0.891
2	THE LAST COP/ラストコップ(刑事アクションコメディ)	レンタル救世主(コメディ)	0.873
3	砂の塔～知りすぎた隣人(サスペンス)	家政夫のミタゾノ(ヒューマンドラマ)	0.872
4	THE LAST COP/ラストコップ(刑事アクションコメディ)	キャリア～突破りの警察署長～(刑事)	0.8643
5	ドクターX～外科医・大門未知子～(医療)	キャリア～突破りの警察署長～(刑事)	0.8637
6	メディカルチーム レディ・ダ・ヴィンチの診断(医療ミステリー)	相棒 season15(刑事)	0.838
7	メディカルチーム レディ・ダ・ヴィンチの診断(医療ミステリー)	家政夫のミタゾノ(ヒューマンドラマ)	0.814
8	メディカルチーム レディ・ダ・ヴィンチの診断(医療ミステリー)	ドクターX～外科医・大門未知子～(医療)	0.808
9	ドクターX～外科医・大門未知子～(医療)	勇者ヨシヒコと導かれし七人(冒険)	0.808
10	相棒 season15(刑事)	ドクターX～外科医・大門未知子～(医療)	0.793

7. まとめ

本研究では、TV ドラマに対する実況ツイートを利用して視聴者の反応を抽出し、ドラマ1話にどのような特徴があるかについて、時間区間を定め、トピック分布を求めることによって分析した。また、その分析から得られたドラマ1話毎の特徴を用いることによってドラマ全体の特徴を見出し、他のドラマとの類似度の算出を行った。

結果として、ドラマ1話の時間区間における特徴づけは、実際のシーンとの結びつきがはっきりとしている箇所が多く、特徴を捉えることができていた。また、これを用いて求めたドラマ全体の特徴についても、各話とトピックとの関連が見受けられた。しかし、コサイン類似度により求めたドラマ同士の類似度について

は、類似度の高い順に並べた場合、出現するドラマに偏りがあつた。何らかの問題により偏りが発生している可能性があるため、今後、その要因を排除し、類似度算出の精度を向上させていく必要があると考える。

また、本研究において、ドラマベクトルを用いてドラマ全体の特徴づけを行った際、部分的にはあるが、ドラマの展開のパターンに規則性が見られた。このことから、今後は、視聴者の反応から抽出したドラマの特徴に基づき、より詳細なドラマの「展開パターン」を抽出し、展開パターンに応じたドラマの検索や推薦を行うことを検討している。

謝辞

本研究は(公財)電気通信普及財団の助成およびJSPS 科研費 JP16K12534 の助成をうけたものです。

参考文献

- [1] SONY “Twitter 連携 | ネットサービスを楽しむ | 液晶テレビ BRAVIA ブラビア | ソニー”, <http://www.sony.jp/bravia/technology/internet/twitter.html> (2017-1-14)
- [2] 安藤聖泰 “金曜ロードSHOW!などで実施したソーシャルテレビ視聴サービス“JoinTV(ジョインティービー)””, 映像情報メディア学会誌, 68, 3, pp.J111-J116, 2014
- [3] VOYAGE GROUP “グリーと VOYAGE GROUP、ソーシャルビューイングアプリ「emocon」をリリース”, <https://voyagegroup.com/news/press/469-3/> (2017-1-14)
- [4] 若井祐樹, 山本湧輝, 熊本忠彦, 灘本明代 “映画の実況ツイートにおける時系列毎の感情抽出手法の提案”, 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(DEIM Forum 2014), E6-1, 2014
- [5] 山内崇資, 林佑樹, 中野有紀子 “日本語解析による Twitter の感情分析とシーンインデキシングへの応用”, 第75回全国大会講演論文集, 2013, 1, pp.315-316, 2013
- [6] 中澤昌美, 帆足啓一郎, 小野智弘 “Twitter によるテレビ番組重要シーン検出及びラベル付与手法”, 第73回全国大会講演論文集, 2011, 1, pp.517-519, 2011
- [7] 中原孝信, 前川浩基, 羽室行信 “テレビ番組視聴時における Twitter 投稿からのトピック検知”, オペレーションズ・リサーチ: 経営の科学, 58, 8, pp.442-448, 2013
- [8] 大田垣翔, 牛尼剛聡, 角谷和俊 “ソーシャルビューイングにおける盛り上がりの効果的な提示のためのツイート集約化手法”, 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(DEIM Forum 2015), G4-4, 2015
- [9] <http://mecab.sourceforge.net/>
- [10] David M Blei, Andrew Y Ng, and Michael I Jordan. “Latent dirichlet allocation”, the Journal of machine Learning research, 3, pp.993-1022, 2003