

タグマイニングによるテレビ視聴率への影響要因分析

浦田 紀果¹ 大槻 明¹

概要: 近年、テレビを見ながら SNS に感想を投稿する視聴者が増加している。こうした傾向から、Twitter に投稿されたツイートの内容とテレビ番組の視聴率に焦点を当てた研究もいくつか行われているが、ハッシュタグに注目した研究はほとんど見られない。ゆえに本研究では、タグマイニングのアプローチを提案し、テレビドラマの視聴率に影響を与えていた要因について分析した。本研究におけるタグマイニングのアプローチについて次の通り述べる。まず、ツイートからタグセットを抽出し、抽出したタグセットを対象にアソシエーション分析を行うことで、Confidence（信頼度）の観点から共起頻度の高いタグセットを抽出する。そして、この Confidence スコアと視聴率との相関係数を求めることで、テレビ視聴率に影響を与えていた要因について分析する。

キーワード: Twitter, TV 視聴率, ハッシュタグ, タグマイニング, アソシエーション分析

1. はじめに

近年、スマートフォンの普及に後押しされ、Twitter や Facebook, Instagram などの SNS を利用する人が増加している。また、それによって、テレビ番組を見ながら SNS に感想を投稿するソーシャルビューイングという視聴スタイルも定着してきた。スカパーJSAT 株式会社[1]は、SNS 利用者の約3割、10代・20代では4割以上に、ソーシャルビューイングの経験があるという調査結果を得ている。2014年には、株式会社ビデオリサーチ社[2]が、ツイッターTVエコーという、ツイートが表示された総量の「インプレッション数」と、ツイートが表示された人数の「インプレッションユーザー数」を指標として、Twitter 上でのテレビ番組への反応を測るサービスの提供を開始している。このように、Twitter とテレビ番組の関係は以前から注視されてきた。

しかし先行・関連研究では、テレビ番組に関する書き込みの抽出や内容の解析する手法については明らかにしているが、SNS 上での発言量には発信者数の偏りが見られるため、正確な評価が難しいことへの指摘もされている。

ゆえに本研究では、タグマイニングのアプローチから分析を行うことで、視聴率に影響を及ぼしている要因について明らかにするアプローチを提案する。本研究におけるタグマイニングとは次に述べる通りである。まず、ツイートからタグセットを抽出し、抽出したタグセットを対象にアソシエーション分析を行うことで、Confidence（信頼度）の観点から共起頻度の高いタグセットを抽出する。そして、この

Confidence スコアと視聴率の相関係数を求めることで、テレビ視聴率に影響を与えていた要因について分析する。評価実験の結果、俳優というファクターがドラマ視聴率に影響を与えていたことを明らかにし、本研究が提案するタグマイニングのアプローチの有効性を確認することができた。

2. 先行・関連研究

インターネットとテレビ番組の関係についての研究は、これまでもいくつか提案されてきた。

小林ら[3]は、テレビ番組の視聴などのイベントに際し、この内容を実況しているユーザーの書き込みを検出する手法を提案している。具体的には、「なう」や「はじまた」といったイベント内容に依存しない表現を抽出し、イベント開始直後に書き込みをしたユーザーにイベントを観覧しているかいないかのラベル付けを行った。次にラベルの付いたユーザーがイベント開始直後以外に書き込んだ内容について解析することで、イベント固有の情報を抽出する検出器を構築し、イベントの評判解析に用いることができる実況書き込みの検出を行った。

齋藤ら[4]は、大きな話題となった朝ドラの「あまちゃん」が視聴者にどう見られたかを明らかにするため、①視聴率分析、②世論調査、③インターネット調査、④ソーシャル・リスニング調査という4つの調査を実施した。④ソーシャル・リスニング調査では、投稿されたツイートの内容について、定量・定性分析のアプローチからその傾向を探っている。この分析の結果、SNS での発信の推移が視聴率と関連しつつ

¹ 日本大学 経済学部
Nihon University College of Economics

も独自の動きを見せていることを発見し、SNSが視聴率とは異なる評価の指標となり得る可能性を述べている。

テレビ番組の視聴率に影響を及ぼす要因については、福島ら[5]が放送前の情報のみを用いてテレビドラマの視聴率を予測する研究を行っている。放送局や時間帯、出演する役者やスタッフといった情報から、サポートベクター回帰とランダムフォレストの2つのモデルを適用し、テレビドラマの初回視聴率の予測を行った。この結果、予測した視聴率と実際の視聴率の間に強い相関関係を確認し、また予測モデルから視聴率に影響を与える要因についての解明も行っている。

以上に外観した先行・関連研究では、ツイート数や投稿された内容を中心として分析が行われているが、ツイートに含まれるハッシュタグに着目した研究は行われていない。ゆえに本研究では、ハッシュタグの共起頻度をアソシエーション分析によって解析することで、ツイートの投稿者が強調したい内容や、ツイートの特徴を抽出できるのではないかと考えた。この一連のアプローチを、本研究ではタグマイニングと呼称する。

3. 提案コンセプト

3.1 はじめに

はじめに、本研究で提案するタグマイニングのアプローチについて述べる。タグマイニングは、TV視聴率に影響を及ぼす要因について特定するためのアプローチであり、まずは分析対象のコンテンツ（本研究ではドラマを3つ「シャーロック」、「同期のサクラ」、「モトカレマニア」）を選定し、これらのドラマに関わるツイートを、これらのドラマの放送前から最終回までの間、放送時間中につぶやかれたツイートを取得する（3.4節）。同時に、全放送分の視聴率データも取得する。

次に、取得したツイート単位でタグセットの共起頻度リストを抽出し、アソシエーション分析を行うことで、タグセット単位でConfidence（信頼度）のスコアを求める（3.5節）。

そして最後に、全放送分の視聴率とタグセットのConfidenceスコアの相関分析を行うことで、視聴率へ継続的に影響を与えていた要因を抽出する。

3.2 ハッシュタグ

ハッシュタグとは、キーワードやフレーズの前にハッシュタグ記号（#）を付けたもののことであり、Twitterにおいては、興味・関心のあるトピックへのフォローを容易にするためのラベルとして使用されるものである¹。これを含んだツイートの投稿が行われることで、ユーザーはトピックの検索がしやすくなったり、同じ興味・関心を持つユーザーとトピックの共有が容易になったりして、SNSの機能性を向上

させることができる。ハッシュタグはTwitterに限らず、Facebook、Instagramなど他のSNSでも活用されている。

本研究では、このハッシュタグに注目し、タグマイニングのアプローチから分析することで、視聴率に影響を及ぼす要因（特徴）を抽出することを目指す。

3.3 視聴率データ

視聴率データは、株式会社ビデオリサーチ社²が関東地区で調査したものを取得した。今回分析対象に選定した3つのドラマ（「シャーロック」、「同期のサクラ」、「モトカレマニア」）の全放送回の視聴率の推移を図1～3に示す。

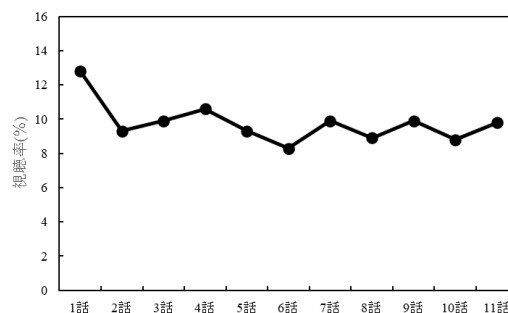


図1 「シャーロック」視聴率の推移

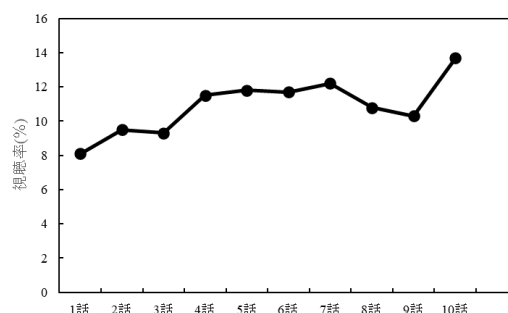


図2 「同期のサクラ」視聴率の推移

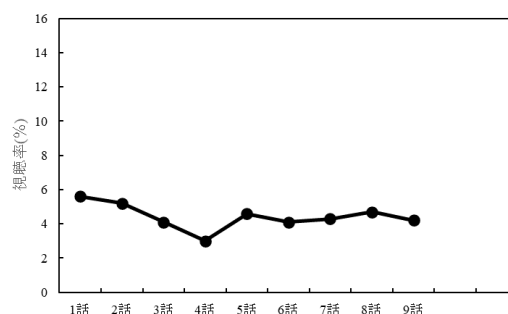


図3 「モトカレマニア」視聴率の推移

3.4 ツイートの収集

Otsukiら[6]が開発したTwitter APIを用いた検索スクリプトを用いてツイートを収集した。

¹ <https://help.twitter.com/ja/using-twitter/how-to-use-hashtags>

² <https://www.videor.co.jp/tvrating>

まず、収集対象としては、①2019年10月に放送が開始する、②午後9時～午後10時を放送開始時刻とする、③放送回数10回前後のテレビドラマという条件を基準に、2019年10月1日時点で公式アカウントのフォロワー数が多かったものから、「シャーロック」（フジテレビ）、「同期のサクラ」（日本テレビ）、「モトカレマニア」（フジテレビ）の3タイトルを選出した。

次に、対象となるドラマの名称を使用したハッシュタグを含むツイートを、第1話放送前から放送終了まで週次で取得した。放送開始前は任意の時間帯に2～3時間、放送中はドラマの放送終了時間から1～2時間後から過去にさかのぼってツイートを取得した。そして、2019年10月6日から2019年12月17日までの期間、ツイートの収集を行った。そして、収集したツイートについて、重複ツイートの削除といったスクリーニングを施した。この結果、取得した総ツイート数は1,178,457件となった。具体的に、ツイート取得に用いたキーワード（ハッシュタグ）と実際に取得できたツイート数を表1にまとめる。

表1 番組名とハッシュタグ、ツイート取得数一覧

番組名	ハッシュタグ	ツイート取得数
シャーロック	#シャーロック	623,763 件
同期のサクラ	#同期のサクラ	316,860 件
モトカレマニア	#モトカレマニア	237,834 件

また、取得したツイートを放送回数ごとにまとめたグラフを図4～6に示す。

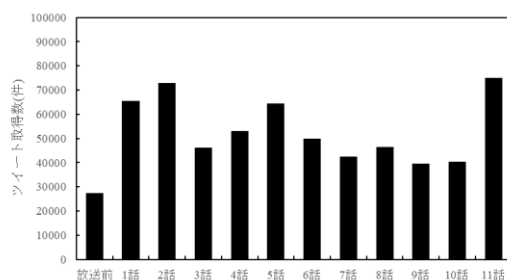


図4 「シャーロック」放送回ごとのツイート取得数推移

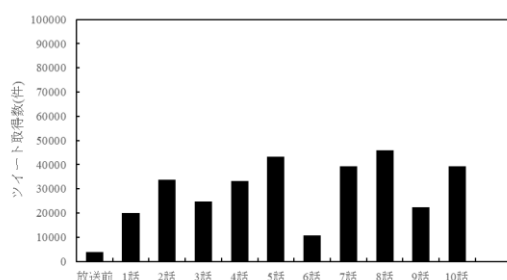


図5 「同期のサクラ」放送回ごとのツイート取得数推移

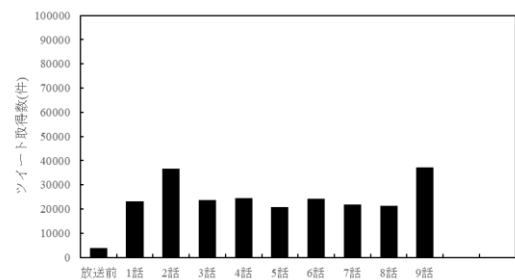


図6 「モトカレマニア」放送回ごとのツイート取得数推移

3.5 アソシエーション分析

各テレビドラマの放送回数ごとに、取得したツイートをテキストマイニングして、共起タグセット頻度リストを作成する。共起タグセット頻度リストの作成例を表2に示す。

表2 共起タグセット頻度リストの作成例

共起タグ1 (番組名タグ)	共起タグ2 (番組名と共起関係 があるタグ)	共起頻度
#シャーロック	#ディーンフジオカ	9
#同期のサクラ	#高畑充希	99
#モトカレマニア	#新木優子	999

次に、アソシエーション分析を行い、各タグセットの信頼度（Confidence）のスコアを放送回数ごとに求めた。このスコアとドラマの視聴率データの相関関係を分析することで、視聴率に影響を及ぼしていた要因を明らかにした。

4. 分析結果

相関係数 R が「 $R > 0.30$ 」もしくは「 $-0.30 > R$ 」の場合に相関関係があると解釈して分析したところ、視聴率に影響を与えた要因をいくつか発見することができた。この結果を表3～5及び図7～12に示す。なお、表3～5は相関関係があるタグセットのみを表示している。

ただし、今回のツイート取得では、継続的に同じタグの共起頻度を抽出することができなかったため、抽出できた放送回のみで分析を行っている。タグ共起頻度のデータを取得できた回数は、表3のタグセット①では全11話中9回、②では11回。表4のタグセット③では全10話中7回、④では6回。表5のタグセット⑤では、全9話中8回、⑥では5回であった。

表3 「シャーロック」相関分析の結果

タグセット	相関係数	結果解釈
①#シャーロック,#ディーン フジオカ,#佐々木蔵之介⇒# 岩田剛典	-0.54	負の相関関係あり
②#シャーロック,#佐々木蔵 之介⇒#岩田剛典	0.47	正の相関関係あり

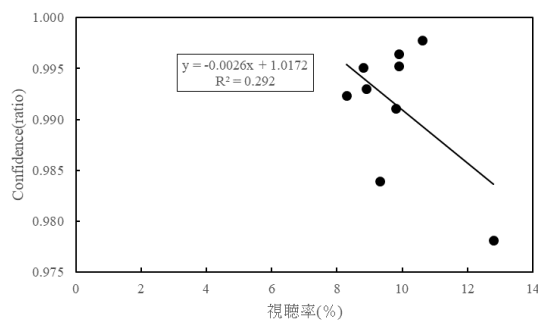


図 7 「シャーロック (表 3)」タグセット①と信頼度の関係

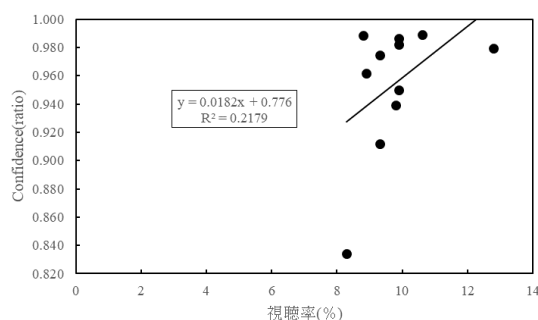


図 8 「シャーロック (表 3)」タグセット②と信頼度の関係

タグセット	相関係数	結果解釈
③#高畑充希,#同期の サクラ⇒#新田真剣佑	-0.37	負の相関関係あり
④#橋本愛,#同期のサ クラ⇒#高畑充希	-0.62	負の相関関係あり

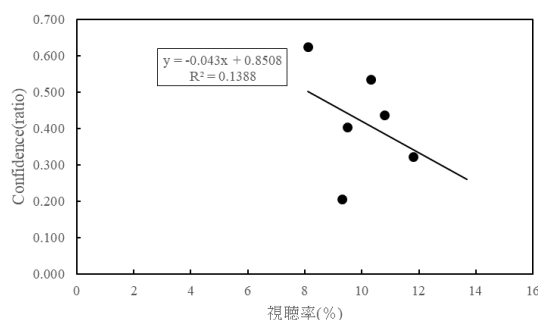


図 9 「同期のサクラ (表 4)」タグセット③と信頼度の関係

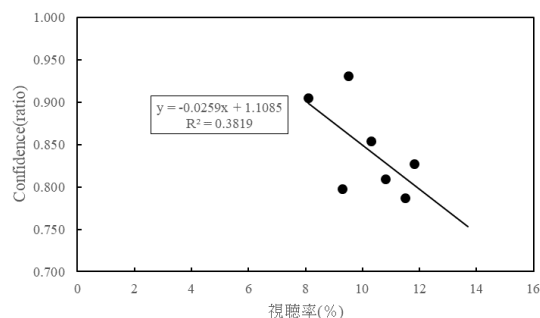


図 10 「同期のサクラ (表 4)」タグセット④と信頼度の関係

表 5 「モトカレマニア」相関分析の結果

タグセット	相関係数	結果解釈
⑤#モトカレマニア,# 高良健吾⇒#新木優子	0.63	正の相関関係あり
⑥#モトカレマニア,# 高良健吾⇒#浜野謙太	0.97	正の相関関係あり

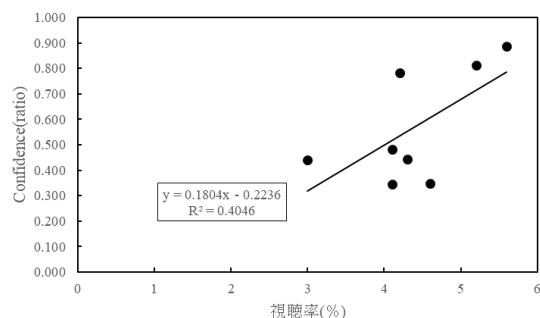


図 11 「モトカレマニア (表 5)」タグセット⑤と信頼度の関係

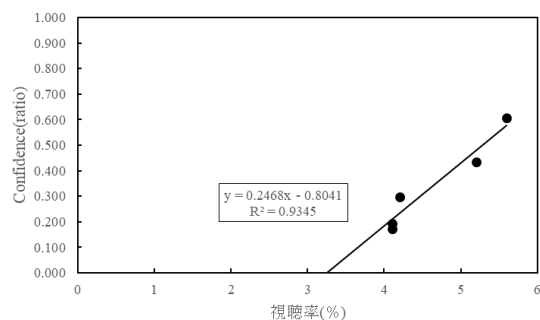


図 12 「モトカレマニア (表 5)」のタグセット⑥と信頼度の関係

5. 分析結果の考察

本章では、前章の分析結果を元に考察を行う。まず、「シャーロック」であるが、表3及び図7, 8の結果から、「#シャーロック, #佐々木蔵之介⇒#岩田剛典」のタグセットの場合は、視聴率との相関が約47%であったが、「#シャーロック, #ディーンフジオカ, #佐々木蔵之介⇒#岩田剛典」のタグセットにおける視聴率との相関係数は-54%であった。つまり、#佐々木蔵之介と#岩田剛典のタグセットは視聴率とポジティブな関係であったが、そこに#ディーンフジオカが入ると視聴率とネガティブな関係になることが明らかとなった。

次に、「同期のサクラ」であるが、「#高畑充希, #同期のサクラ⇒#新田真剣佑」のタグセット（相関係数=約-37%）も、「#橋本愛, #同期のサクラ⇒#高畑充希」のタグセット（相関係数=約-62%）も、負の相関関係が認められたため、#高畑充希と#新田真剣佑、#橋本愛のタグセットは視聴率とネガティブな関係にあることが明らかとなった。

最後に、「モトカレマニア」であるが、「#モトカレマニア, #高良健吾⇒#新木優子」の場合（相関係数=約63%）も、「#モトカレマニア, #高良健吾⇒#浜野謙太」の場合（相関係数=約97%）も、視聴率と正の相関関係が認められたため、#高良健吾と#新木優子、#浜野謙太のタグセットは視聴率とポジティブな関係であることが明らかとなった。

6. まとめと今後の課題

本研究では、TV視聴率に影響を及ぼす要因を特定するためのアプローチとしてタグマイニングを提案した。そして、評価実験の結果、今回選定したドラマにおいて、視聴率に影響を及ぼしている要因が俳優であることを明らかにした。

しかし、今回の分析では、視聴率に負の影響を与えている要因として抽出できたものは俳優だけであった。例えばシャーロックであれば、俳優以外にも原作のアーサー・コナン・ドイルの「シャーロック・ホームズ」シリーズが好きでこのドラマを視聴した人がいるかもしれないが、このような俳優以外の要因を抽出することができなかった。この原因として考えられることは、フリーのTwitterAPIでは1%のサンプルしか取得できないため、満足なサンプルが抽出できていない、もしくは、偏ったサンプルが抽出された可能性が考えられる。

しかし、視聴率に影響を与えていた要因を抽出すること自体はできたため、本研究が提案するタグマイニングという手法の有効性は確認することができたと考えられる。

他方、本研究で用いた相関関係だけでは、因果関係までは明らかにすることはできない。つまり、第4章、5章の結果からは逆の因果も考えられる。例えば、視聴率が悪いがゆえに俳優タグセットが頻出したとも考えられるということである。ゆえに、今後の課題として、タグセットと視聴率の因

果関係についても研究していきたい。

参考文献

- [1] スカパーJSAT 株式会社, “テレビ視聴に関する調査2018” (2018)
https://oshirase.skyperfectv.co.jp/general/release_20180920.pdf, (参照 2019-06-16)
- [2] 株式会社ビデオリサーチ社, “ビデオリサーチ「Twitter TV 指標(サービス名:Twitter TV エコー)」の提供を開始ー1日のインプレッションユーザー数は約700万ー”
<https://www.video.co.jp/press/images/b8fbfb99d10690683a04f239bbc14263.pdf>, (参照 2019-06-16)
- [3] 小林尊志, 野田雅文, 出口大輔, 高橋友和, 井手一郎, 村瀬洋: Twitterにおける実況書き込み検出手法の検討, 電子情報通信学会「信学技報」, MVE2010-45, pp.129-130, 20
- [4] 齋藤建作, 二瓶互, 関口聡: 朝ドラ『あまちゃん』はどう見られたか ~4つの調査を通して探る視聴のひろがり と視聴熱~, NHK文化放送研究所「放送研究と調査」, pp.12-41, 2014.
- [5] 福島悠介, 山崎俊彦, 相澤清晴: 放送前の情報のみを用いたテレビドラマの視聴率予測, 映像情報メディア学会誌, Vol.70, No.11, pp.J255-J261, 2016
- [6] Akira Otsuki, Masayoshi Kawamura: The Study about the Analysis of Responsiveness Pair Clustering of Social Network Bipartite Graph, Advanced Computing: An International Journal (ACIJ), AIRCC, Volume 4, Number 6, pp.1-14, 2013.