

Twitter を用いたテレビ番組の分析と評価 Analysis and Evaluation of Television Programs by Using Twitter

海老原 将也

Masaya Ebihara

法政大学情報科学部デジタルメディア学科

E-mail: masaya.ebihara.7w@stu.hosei.ac.jp

Abstract

In recent years, it is becoming more difficult to evaluate and compare TV programs along with the diversification of channels. In particular, it is difficult to evaluate animation programs because there are 50 or more programs simultaneously and because the typical period of their broadcast is 3 months. Also, since many programs are broadcasted late at night, audience rating is hard to use for their evaluation. In this paper, we organize information on TV programs obtained from Twitter, and analyze it from a long-term perspective. Specifically, we analyze it to evaluate TV programs by using on-the-spot tweets of real-time viewers and comments of those who watch the recorded programs. We calculate emotion values by using information from Twitter, and investigate changes of viewers' impressions about programs. Also, we examine the changes of numbers of Twitter users who watch programs. As a result, we found that it was difficult to calculate emotion values from short comments of Twitter, but that programs had tendencies for changes in viewers and tweets. Also, our results were different from existing evaluation measures such as audience rating. Although we used the Streaming API and hashtags to obtain tweets, it turned out that they were not sufficient because of their properties.

1. はじめに

近年、チャンネルの多様化に伴い、放送されるテレビ番組の評価、比較を行うことが容易ではなくなっている。特にアニメ番組に関しては同一期間で約 50 番組と膨大ながら、放送期間が通常 3 か月単位であり、評価しにくい。また多くの番組が深夜帯に放送されているため、視聴率が評価材料になりにくい。こうした中、マイクロブログで番組の感想や実況を行う人が増加しており、統計的な数値だけでなく番組の情報を文章で得られやすくなった。特に近年も利用者の拡大が続いている Twitter ではリアルタイム性や利用の手軽さから様々な番組に関するツイートが見られる。

そこで本研究では Twitter から得られるテレビ番組に関する情報を整理し、長期的な視点から分析と評価を行う。具体的には、リアルタイムに番組を見る人たちの Twitter における実況的なツイートや、録画を見た人たちの感想を用いて、番組の評価のための分析を行う。分析方法と

しては、Twitter から得られる情報をもとに感情値を算出し、番組に対する印象の変化が見られるかを調べる。また、番組を視聴する Twitter ユーザの数の変化等も調べる。結果としては、Twitter のような短文から感情値を読み取ることが難しいものの、番組ごとに視聴者やツイート数の変化の傾向があることが分かった。また、視聴率等の既存の評価方法とは大きく異なる結果も得られた。本研究ではツイートの取得に Streaming API とハッシュタグを利用したが、その特性のために十分な結果を得ることができないことが分かった。また分析の精度をより高めることによって、番組の傾向予測といった分野への応用も考えられる。

2. 関連研究

Twitter の分析についての先行研究としては古川ら [1] のものがある。この研究では社会的事象の抽出や活用方法について、特に犯罪情報を対象にして調査した。その結果、投稿者自身の目撃、被害に関する投稿、公共機関の発表を引用した投稿、ニュース記事の引用の 3 種類があることを発見した。また犯罪種別傾向の違いから Twitter 独自の犯罪情報が存在することを突き止めた。

また志岐の研究 [2] では視聴者が番組のツイートで注目する点や Twitter を用いて番組を話題にする動機を調査した。その結果、驚き、応援、感動、感謝といった点を多くの人が注目していることや、動機として Twitter 上で行われる共有は自分の意見の表出や、他者の意見の確認を行っていたことが分かった。

3. 準備

3.1. Twitter

Twitter は Twitter 社が運営する情報ネットワークである。マイクロブログの 1 種であり、利用者は全世界で 3 億人を超えており、国内の利用者も 4000 万人と広く利用されている。Twitter を利用したメッセージの発信はツイートと呼ばれ、全世界で毎日 5 億件ものツイートが行われている。ツイートには他人のツイート本文をそのままツイートするリツイート機能や特定のユーザに向けてツイートするプライベイトといったものがある。また特定の内容に関するツイートを検索できるようにするため、「#」を用いてタグ付けするハッシュタグ機能などがある。

Twitter をプログラムで利用するための関数や機能の集合である Twitter API に、Rest API と Streaming API がある。Rest API は各ツイートに対してリクエストを行い、デー

タを利用する。毎回リクエストを行うため、一定時間ごとに利用できる回数が設定されているが、複雑な検索や過去にさかのぼってのツイートの利用ができる。一方、**Streaming API** では設定した情報をもとにリアルタイムでツイートを利用する。そのため、一度リクエストを行えば停止するまで、ツイート利用でき、回数の制限はない。しかし、リアルタイムでしか利用できないため、ネットワークが安定している環境が必要となる。またリアルタイムで無作為に取得できるツイートは全体の約 1%程度であり、すべての **Twitter** の情報を利用できるわけではない。

3.2. 感情辞書

感情辞書 [3] [4] は日本語の用言を主観的、客観的の 2 種類に分類し、さらにそれぞれをポジティブ、ネガティブに分類している。また名詞に対しては主観的、客観的に加えて、出来事や行為といった分類がされている。収録表現数は用言が約 5000 件、名詞表現が約 8500 件となっている。

辞書作成の際に単語感情極性対応表 [5] を用いて分類している。これは語彙ネットワークを利用して計算された感情極性値と単語の対応表である。感情極性値は -1 から 1 までの実数値であり、-1 に近づくほどネガティブ評価、1 に近づくほどポジティブ評価となっている。一方、感情辞書では感情値をポジティブ、ネガティブの二値属性として扱っている。

4. 分析手法

本研究では **Twitter** から得られるテレビ番組に関する情報を整理し、長期的な視点から分析と評価を行う。番組の評価には一回の放送ではなく、放送期間全体からデータを読み取り、分析する必要がある。そこで、放送期間全体のツイートから感情値やユーザ数の変化などを求め、番組の傾向を分析し評価に役立てる。また事前に行われた期待度のアンケートや視聴率といった **Twitter** 以外の媒体との比較も行う。

4.1. データの収集

本研究では **Java** の **Twitter API** を利用してデータを収集した。最初に、**Rest API** か **Streaming API** のどちらを利用すべきか決めるための実験を行った。実験では 2 つの **API** をそれぞれハッシュタグによるキーワード指定と未指定による収集を行った。これは過去のツイートデータを得られる **Rest API** では **Streaming API** と比べ、データの取りこぼしが少ないのではないかと考えたからである。またキーワード指定による収集を比較しているのは、無作為に行うデータ収集の方がハッシュタグを含まない対象ツイートを取得することができるかといったことを実験するためである。その結果、**Rest API** では細かくデータを取得できる一方、リクエスト制限による問題が発生した。またキーワードの未指定によるデータ収集では目的の内容を含んだデータが得られないことが分かった。そのため、リクエストに制限がない **Streaming API** を利用した。また確実に番組に関する内容を得られるように公式アカウントが推奨するハッシュタグを利用している。

4.2. テキスト処理

テキストを分析するために日本語形態素解析エンジン **kuromoji** を利用する。そこから **URL** などの不要な情報を排除し、ツイート毎にハッシュタグによる番組の分類を付加する。

4.3. 感情分析

各ツイートに感情値を設定して分析に利用するために、3.2 節で述べた 2 つの辞書を用いる。単語をポジティブまたはネガティブな主観的評価、ポジティブまたはネガティブな客観的評価、ポジティブまたはネガティブなそれ以外の評価の計 6 種類に分類する。それぞれの単語はポジティブなら 1、ネガティブなら -1 の感情値が与えられる。それらを主観的評価、客観的評価、それ以外の評価についてツイート内の感情値を持った単語の平均をそのツイートの感情値としている。

4.4. ツイートの分別

大量のツイートを得られるが、感情値を持たないツイートやリツイートによる重複した内容のツイートがあるため、そういったツイートを分別する必要がある。本研究ではスパムやボットのような性質を持つツイートやリツイートを分析対象から排除している。

4.5. 感情値の変化

番組内容に応じて時系列的に変化するツイート内容を分析する。例えばクライマックスに近づくにつれ、視聴者の期待が高まり、感情値がポジティブな値を示すといった内容である。これらから番組全体の印象がつかめるかどうかを調べる。

4.6. **Twitter** での話題性の変化

番組ごとの話題性について分析する。各番組についてのツイートをしているユーザ数の変化や感情値を持ったツイートの増減から分析する。

4.7. 既存の評価基準との比較

既存の評価基準と **Twitter** での評価との比較を行う。その 1 つに視聴率があるが、本研究で対象とする番組では公表されない場合がある。そのほかにも放送開始前のアンケート調査の結果と 4.5、4.6 節の結果を比較する。

5. 実験

実験期間は 2016 年 10 月 1 日から 2016 年 12 月 31 日であり、対象としたアニメ番組は期間中に新規に放送が開始された 58 番組とした。収集したツイートデータはテキストに加えて、ユーザ情報、投稿時間を含んでいる。収集する際に使用したハッシュタグは各番組の公式アカウントが推奨する中で番組 1 つにつき 1 種類とした。

取得したデータの中にもスパムやボットの影響を受けたツイートが存在していたため、それらを排除する必要があった。スパムやボットの多くは本来無関係なタグを同時に使用することで検索に引っかかりやすいようにしている。そこで対象番組のハッシュタグを同一ツイート内に複数持つ場合、分析対象から除外した。

5.1. 番組の長さによる違い

取得したデータを番組ごとに分類した際、それぞれのツイート数が大きく異なっていた。5分枠、15分枠の番組は放送日であっても、1000件以下の場合が多かった。一方、30分枠の番組の放送日では多くの場合、2000件以上のツイートを取得できた。

5.2. 感情値の変化

番組の内容に応じた変化を期待したが、大きな変化の傾向は見られなかった。原因として、ほとんどのツイートは短文であることに加え、インターネット上では独特な言葉が使われることが多く、感情値を計算できなかったことが考えられる。

全体的にはポジティブな傾向があり、主観的な感情値の変化が大きい番組が多かった。感情値を伴うツイートは各番組の放送時間外に高い割合で存在した。これは実況的なツイートが放送中に行われており、キャラクターの様子や映像の状況に関する内容のみを言及しており、感情値を含む単語が少ないからだと考えられる。

5.3. Twitterでの話題性の変化

ツイート数やユーザ数の変化にいくつかの傾向が見られた。多くの番組で見られる傾向として、初回放送日に多くのユーザやツイートがあり、その後、減少傾向になり、最終回の放送日で再び増加するパターンがあった。これは視聴を継続しつつも、内容をリアルタイムに言及するほどではないと感じたユーザが節目となる最初と最後の放送でツイートをした結果だと考えられる。図1は「#夏目友人帳」(夏目友人帳 伍)を含んだツイート数の変化を示したものである。初回放送日である10月25日からツイート数が減少しているが最終放送日である12月21日には初回放送日と同数近くまで増加している。

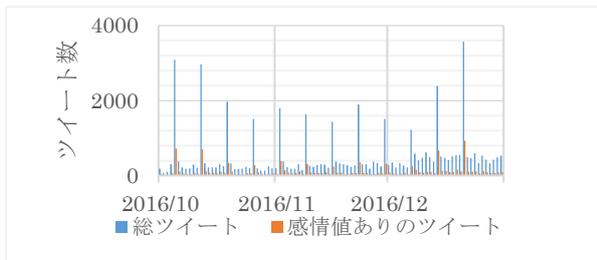


図1:「#夏目友人帳」のツイート数の変化

一方で中盤から終盤にかけて極端に変化した番組が存在した。この番組は放送期間中の季節に合ったテーマで構成されており、ある放送回で有名人を起用することを公式アカウントが発表している。その直後の放送からユーザ、ツイート数ともに大きな伸びを見せており、最終回まで増加傾向が続いた。このことから番組内容だけでなく、制作者側の広報活動がTwitter上の番組全体の話題に影響する可能性があることが示唆される。図2は「#yurionice」(ユーリ!!! on ICE)を含んだツイートをしたユーザ数の変化を示したものである。感情ユーザは感情値を持ったツイートをしたユーザであり、総ユーザは感情値を持たないユーザも含めたものである。この番組は

フィギュアスケートを題材としており、12月6日にプロフィギュアスケーターである織田信成の出演が告知された。

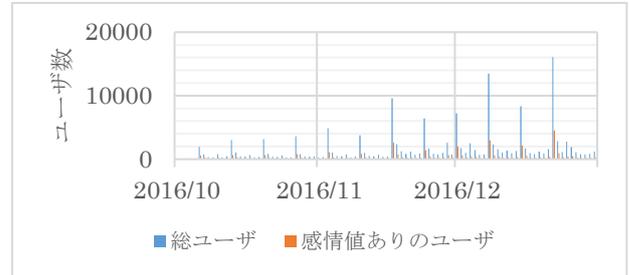


図2:「#yurionice」のユーザ数の変化

多くの番組に共通していたこととして、総ユーザ数の変動がある場合でも、感情値を持つツイートをするユーザの数は総ユーザの増減率にあまり影響を受けなかった。

5.4. 既存の評価との比較

本研究の分析結果を、インターネット上での前評判である、2016年9月14日から9月21日まで行われたgooランキング[6]の結果と比較した。上位10作品のうち本研究で対象とした番組は9番組あった。そのうち5番組が続編となっており、アニメオリジナル作品は1作品のみだった。インターネット上での前評判のランキングには、原作の知名度や、以前の放送の感想が大きく影響しているといえる。しかし、前評判のランキングと本研究のTwitterの分析結果には大きな違いがあった。ランキング1位である「ブブキ・ブランキ 星の巨人」はTwitter上ではあまり言及している人はいなかった。ツイート数の伸びはないものの、ユーザ数はあまり変化していなかった。このアニメは続編であり、一定の視聴者の確保はできているため、ランキングの上位になったのだと考えられる。図3、図4は「#ブブキ」を含んだツイート数とユーザ数の変化である。あまりツイート数は多くないものの、比較的安定したユーザ数の変化である。



図3:「#ブブキ」のツイート数の変化



図4:「#ブブキ」のユーザ数の変化

さらに本研究の結果と視聴率をもとにしたデータを比較した。比較したデータはビデオリサーチが公開している関東地方の視聴率をもとに制作した。そのデータでは今回の分析対象の1つである「タイムボカン 24」についてはほぼ毎週上位10位以内に入っていた。これは放送時間が週末の夕方であり、家族で見ている人が多かったからだと考えられる。しかし、Twitterではあまり話題に上らず、ツイート数やユーザ数は減少傾向にあった。これはTwitterの利用者とこの番組の視聴者層の違いによるものだと考えられる。図5はツイート数の変化を示したものであり、上位に入ったにもかかわらず初回放送以降に減少傾向にあるだけだった。



図5: 「#タイムボカン 24」のツイート数の変化

6. 議論

本研究では Streaming API とハッシュタグを用いて実験を行ったが、番組ごとに大きな結果の差が生まれた。それぞれの仕様がその原因の1つだと考えられる。Twitterから情報を集める際、本来はハッシュタグによるフィルタリングなどではなく、無作為に広範囲にわたって集めなければ、様々な種類のツイートを取得できない。しかし、3.1節で述べたように Streaming API は全体の約1%しかツイートを取得できない。そのため、無作為に広く集めようとする対象の情報を含むツイートを取得できない可能性がある。そのため、本研究ではハッシュタグを指定することで不要な情報を除外することを試みた。

一方、ハッシュタグは自由に設定できるものであるため、非常に大量に存在している。例えば漢数字と算用数字、全角文字と半角文字といった違いだけでも異なるハッシュタグとして認識される。本研究では、対象となる番組の公式アカウントが推奨するハッシュタグのアカウント1つを使用してツイートデータを集めたため、この点が問題となった。「3月のライオン」などでは漢数字表記や全角数字のものも多く用いられており、取得結果が予想以上に少なかった。

またこうした文字の違いだけでなく、Twitter上で番組のハッシュタグが新たに発生した場合に、公式アカウントの推奨タグと大きく異なることもある。例えば「オカルティック・ナイン」では公式アカウントが「#オカン」をハッシュタグとしている中で Twitter 上では「#occultic_nine」というタグが生まれていた。このタグを原作者や公式アカウントが放送途中から使い始めたため、取得できたデータにずれが生じてしまった。

本研究では、比較的にスパムアカウントなどによる影響は少なかったが、さらにデータの取得範囲を広げる場

合には考慮する必要もある。その方法の1つに事前に調査によるボットアカウントの排除といったものがある。

感情値について本研究では一般的な感情語に関する辞書を用いて算出したため、あまり良い結果とならなかった。Twitterの特性に合わせた感情辞書を作成できれば、より精度を上げることはできると考えられる。

7. おわりに

本研究では Twitter から長期的に得た情報を用いてテレビ番組に関する分析を行い、評価に役立てることができるか調べた。既存の情報と異なった傾向や番組形態による違いなどが確認できた。このような点から Twitter の特徴であるリアルタイムな情報やデータ量は様々な分析に使うことが可能だと考える。

しかし、ツイートデータの収集方法については課題点が多く、対象のデータの設定方法やハッシュタグの自由度といった問題がある。また、テキストの分析には、通常の言語処理だけでなく Twitter のようなマイクロブログ特有の言葉に配慮した改善も必要である。

文 献

- [1] 古川忠延, 阿部修也, 安藤剛寿, 岩倉友哉, 志賀聡子, 高橋哲郎, 井形伸之, "Twitter からの犯罪情報抽出の可能性調査," 情報処理学会研究報告: デジタルドキュメント (DD), vol. 2011-DD-82, no. 3, pp. 1-6, 2011.
- [2] 志岐裕子, "テレビ番組を話題とした Twitter 上のコミュニケーションに関する検討," 慶応義塾大学メディア・コミュニケーション研究所紀要, vol. 65, pp. 135-148, 2015.
- [3] 東山昌彦, 乾健太郎, 松本裕治, "述語の選択選好性に着目した名詞評価極性の獲得," 言語処理学会第14回年次大会論文集, pp. 584-587, 2008.
- [4] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "意見抽出のための評価表現の収集," 自然言語処理, vol. 12, no. 3, pp. 203-222, 2005.
- [5] 高村大也, 乾孝司, 奥村学, "スピンモデルによる単語の感情極性抽出," 情報処理学会論文誌, vol. 47, no. 2, pp. 627-637, 2006.
- [6] goo ランキング編集部, "これは外せない! 10月スタートの「深夜アニメ」ランキング ※30分枠のアニメのみ," NTT レゾナンス, 2016. [Online]. Available: https://ranking.goo.ne.jp/ranking/category/026/vote_798/.