

# Twitterにおけるアニメのネタバレツイート判定手法の提案

## A Proposal of Method to Detect Animation Spoilers on Twitter

田島 一樹<sup>†</sup> 中村 聡史<sup>‡</sup>

<sup>†</sup> 明治大学大学総合数理学部 〒164-8525 東京都中野区 4-21-1

E-mail: <sup>†</sup> kazu.t.1205@gmail.com, <sup>‡</sup> satoshi@snakamura.org

**あらまし** アニメなどテレビ番組を視聴しながら SNS 上で感想などの情報発信することは一般的であり、リアルタイムな感想共有は視聴体験を高めることにつながっている。一方で、こうしたネタバレ情報は、未視聴のユーザーにとって本来作品を通して体験するはずだった興奮や感動的な体験を無くしてしまう忌むべきものである。我々はこれまでの研究で、放送時間差などの都合でネタバレに遭遇してしまう可能性を明らかにし、ネタバレデータセットを構築および分類してきた。本研究では複数のアニメコンテンツに対する Twitter 上のネタバレデータセットを構築し、ネタバレ判定手法を提案するとともに、実験を行うことでそのネタバレの推定可能性を検証する。

**Abstract** People generally transmit their impression about Animation on social network service in viewing it. Sharing impression enhance users' enjoyment. However, like these information on social network services often spoil users' enjoyment. In previous our study, we clarify the possibility of encounter spoiler by broadcast time difference and we built the dataset of story spoilers and classified spoilers into several categories. In this paper, we built the dataset of some animation spoilers on Twitter and we clarify the possibility of accuracy to detect spoilers with proposing method to detect animation spoiler.

**キーワード** ネットバレ防止, ストーリーコンテンツ, 機械学習, Twitter

### 1.はじめに

TwitterやFacebookに代表されるソーシャルネットワークサービス (SNS) は、友人や知人などとの交流や情報収集の場として必要不可欠な存在となりつつある。こうしたSNSでは互いの近況を報告するだけでなく、思ったことや感じたことをリアルタイムで発信していき、他者と共有するということが日常的に行われている。ここで、ドラマやアニメで思ったことや感じたことをTwitterで他者と共有するため発信することも多く、番組の放送に応じてTwitter上が盛り上がることも多い。こうした番組に連動したツイート(140字以内のTwitter上での投稿)は、その番組を視聴している人にとっては楽しいものである。一方、その番組を視聴するのを楽しみにしているが、何らかの事情でリアルタイムに視聴できない人にとっては、そうしたツイートはネタバレにつながるため悩ましいものである。

ここでネタバレとは、物語を視聴することを通して本来得られたはずの興奮を奪い去ってしまうものである。株式会社社会情報サービスが運営しているサイト[1]で行われたコンテンツにおけるネタバレに関するアンケートの結果、受けた人の内の約6割が映画などのネタバレに対して不満をもっているものとしている。こうした問題はSNSサイトにアクセスすること、アプリケーションを使わないことで回避することができる。しかし、SNSを遮断することは友人とのコミュニケーションを遮断するとともに、ニュースなどの情報への接触機会を減らしてしまうため、その対策方法としては現実的ではない。また、ネタバレされないためには

リアルタイムで視聴したら良いと考えられるが、仕事や学校の都合上、リアルタイムで視聴できないケースは多々ある。さらに、地域による放送時間の違いも問題の一つである。SNSが普及するまでであれば、こうした地域が離れた視聴者同士は、その物理的な距離によりコミュニケーションを取る機会は限られており問題となりにくかったが、SNSなどでこうしたユーザー同士がつながっているために問題となっている。

我々はこれまでの研究でドラマやアニメの放送時間が地域によって異なるという点に注目し、その時間差によってネタバレに遭遇してしまう可能性があるユーザーの規模について調査し、視聴者の約7割がネタバレに遭遇してしまう可能性があることを明らかにしてきた。また、ストーリーコンテンツにおける人々が共通して致命的なネタバレと考える出来事についてアンケートを利用して調査した結果、アニメにおいてはネタバレとして正体、生死、人物特徴、勝敗に関する出来事に関するものが多いことがわかった。そこで、カテゴリ毎のネタバレデータセットを構築し、SVMによる分類を行った結果、勝敗に関するネタバレの判定はしやすく、正体に関するネタバレの判定はしにくいことが明らかになった。

しかし、これまでの我々の研究[2]ではネタバレデータセットを構築する際、著者がデータに対してラベリングを行っており、一般的にネタバレと考えられているツイートを判定するためには不適切なデータセットである可能性があり、著者以外の者にラベリングを行ってもらう必要があると考えられる。また、我々は放

送中のアニメ番組に連動したツイートに含まれるネタバレでなく、ユーザの知人や友人が発言するジャンルが一様でないツイートに含まれるネタバレを判定対象としているため、それをどの程度判定できるのか明らかにする必要があると考えられる。

そこで本研究では、複数のアニメコンテンツに対するネタバレに関するデータセットを構築し、ネタバレ判定手法を提案するとともに評価実験によって推定可能性を明らかにする。具体的には、選定したアニメコンテンツを同一作品、ミステリー系、バトル系の3つのジャンルに分け、アニメ番組に連動したツイートと一定期間内に投稿された全てのツイートから無作為に選定したものから単語ベクトルを作成し、SVMを用いてジャンルごとのネタバレツイートの判定精度や判定しにくいネタバレツイートの特徴などを明らかにする。

## 2. 関連研究

ネタバレを防止することを目的とした研究はこれまでにもいくつかなされてきている。

我々は過去の研究[3]において、諸事情によってリアルタイムでスポーツの試合を見ることができないユーザがウェブページを閲覧している際、試合の内容に関するネタバレに遭遇してしまうことを問題とし、そうしたユーザのためにウェブページにおけるテキスト情報の曖昧化処理によってネタバレを防止する手法を提案している。しかし、この研究では、ネタバレの判定は事前に用意したネタバレ用正規表現辞書とのマッチングを前提としており、正規表現辞書をメンテナンスする手間などがあった。そこで、白鳥ら[4]はそういったサッカーのネタバレを大まかに直接的ネタバレと間接的ネタバレに分類し、正規表現のみでは判定できない間接的に試合結果が分かってしまうネタバレの判定可能性を示した。本研究は、こうしたネタバレ判定のための分類器を機械的に構築することを目的としている。

Golbeckら[5]は時差により放送時間が異なることによってTwitterでネタバレされてしまうアメリカの事例を紹介しており、世界的にもネタバレは問題となっていることが分かる。また、この研究ではドラマやスポーツに関するワードが登録されているブラックリストを生成することによってテキストのネタバレを検知し、そのツイートをブロックするためのミュートボタンを実装している。Golbeckらは全てのネタバレを検知し、再現率100%を目指しているという点で、我々と同じであるが、本研究ではストーリーコンテンツにおけるネタバレを分類および判定するという点で異なる。

Twitterを対象としてクライアントを実装することによりネタバレ防止を行う手法として[3]や[5]がある。

こうした研究ではハッシュタグ付きツイートから時間的にバーストする単語を抽出し、その単語を含むツイートを非表示にすることでネタバレを防止している。しかし、バーストする単語が必ずネタバレというわけではない。本研究ではストーリーコンテンツに対する複数ジャンルのネタバレデータセットを構築し、ストーリーコンテンツにおける本質的なネタバレの防止を目指しているという点で異なる。

池田ら[6]はAmazon.comや価格.comのようなオンラインショッピングサイトにおいて、ユーザが商品購入の意思決定する際に参考にするレビュー文に小説、映画、ゲームなどのストーリーコンテンツに関するあらすじが含まれることを問題に挙げており、人名辞書と意見辞書を用いてあらすじを表す文と意見文をそれぞれ判定し、あらすじ部分のみを隠して表示するシステムの提案と実装を行っている。判定する文章がストーリーコンテンツに関するものという点では同じだが、オンラインショッピングサイトとTwitterでは、感嘆文や登場人物の発言の有無など投稿される文章の傾向が大きく異なると考えられる。

前田ら[7]は、ストーリーコンテンツに対するレビュー文をユーザが参考にする際にネタバレを発見してしまうことを問題とし、また、ストーリーコンテンツに対する短文形式のネタバレデータセットを構築し、ネタバレに関する単語がストーリーコンテンツ内にどのように分布しているかを調査しており、コンテンツ自体からネタバレを判断する放送について検討している。我々は、コンテンツではなくコンテンツに対するツイートからネタバレ分類器を構築し、ネタバレ判定を行うという点でアプローチが異なる。

田中ら[8]はニコニコ動画においてネタバレとなるコメントが動画視聴中に流れてきてしまうことを問題に挙げており、ルールベース手法と機械学習によりネタバレコメントを検知する手法を提案している。またここでは、ネタバレとなる重要な単語を捉える際には単語バーストを利用している。田中らは動画共有サイト上の蓄積された動画コンテンツに対するネタバレを防ぐことを目指しているが、本研究はリアルタイムのイベントであるアニメなどに対するネタバレを判定するものである。

Leavittら[9]はストーリーコンテンツにおいて作品を楽しみにしている最中にネタバレを知ってしまうために否定的な感情が生まれると考え、作品を知る前からその作品の知識を得ることによって途中でネタバレをされても楽しみを損なわずに作品を楽しめるかどうかの実験を行っているが、本研究では人々が既にストーリーについての知識を持っていることを前提としているものである。

### 3.判定手法

ここでは、アニメ番組に連動した Twitter 上での投稿（ツイート）から SVM のための単語ベクトルを生成するにあたり、形態素解析と係り受け解析による単語ベクトル生成手法を提案する。

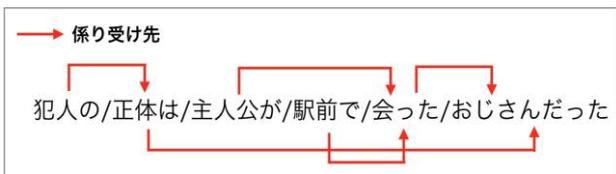
#### 3.1 単語ベクトル生成手法

アニメのネタバレツイートには日常的には使用しない特徴的な語が含まれると考えられる。そこで、我々はネタバレ特有語を検出することでネタバレのツイートとネタバレでないツイートを区別する手法を提案する。具体的には、形態素解析エンジンの Mecab を使用することで単語に分割し、得られた単語の中からネタバレが含まれる文章に特有であると考えられる名詞、動詞、形容詞、連体詞、副詞の 5 つの品詞を利用する。また、それぞれ得られた単語については原形を使用し学習する。

しかし、形態素解析では単語の情報が得られるが、アニメのネタバレツイートを判定する上で重要であると考えられる「誰がどうしたのか」といった文節同士の修飾関係を考慮することができない。そこで文節同士の修飾関係を考慮することが可能である係り受け解析を使用したネタバレ判定手法を提案する。係り受け解析には日本語の係り受け解析器の CaboCha を使用する。この手法ではツイートを係り受け解析にかけて文節ごとに分割し、得られた文節により単語ベクトルを生成する。具体的には、文節単体と文節同士の関係性を考慮したものを利用する。この手法による単語ベクトル生成手順を図 1 に示す。

#### ツイート例

犯人の正体は主人公が駅前で会ったおじさんだった



#### 生成した単語ベクトル

- ・文節のみ  
犯人の、正体は、主人公が、駅前で、会った、おじさんだった
- ・文節同士の関係性を考慮  
犯人の正体は、主人公が会った、駅前で会った、会ったおじさんだった

図 1 CaboCha を用いた単語ベクトル生成手順

これら方法で単語ベクトルを生成したものをベースライン手法とする。

#### 3.2 正規化、人物名一般化処理

精度の向上のため、ツイートを単語または文節に分割した時点でテキスト処理を行った上で単語ベクトルを生成する手法を 3 つ提案する。

Brody ら[10]は Twitter に代表される SNS 上で単語の一部を連続させるなどして変化させることによって投稿者の強い感情を表す語を検出するという手法を提案している。そこで、アニメ視聴者が衝撃的な展開に感情を左右されて用いると考えられる「wwwwww」「勝ったあああああ」のような連続した記号と単語の末尾の母音部分を正規表現で「w」や「勝ったあ」の形に変換する。これを正規化手法とする。これにより、投稿にユーザの感情が含まれているかどうかを区別することができ、ネタバレ判定精度を向上させることができる。と期待される。

また、ストーリーコンテンツにおける主人公やライバル、犯人や被害者などの登場人物名は、作品と話数ごとに大きく異なる。過去のアニメ作品におけるネタバレから、新しいアニメ作品のネタバレを推定するには、この人物名を一般化することが重要になる。そこで、物語の進行を左右するような影響力のある人物名を「主要人物」、物語への影響力の少ない人物名を「モブ」と置き換えを行う。「主要人物」と「モブ」の区別には番組情報が掲載されているサイト[11]と Wikipedia を利用して行った。具体的には、番組情報が掲載されているサイトにおいて出演者として掲載されている人物名を「主要人物」と定義し、Wikipedia に掲載されている全ての人物名から先述の「主要人物」を除いた人物名を「モブ」とした。これにより、例えば「主要人物が死んだ」と「モブが死んだ」とものを区別することができ、精度の向上が期待される。

最後に、先述した正規化手法と人物名一般化手法の二つで行った処理を同時に行う正規化・人物名一般化組み合わせ手法を提案する。これによりネタバレ判定精度のさらなる向上が期待できる。

### 4 ネットバレーデータセット構築

ツイートデータを収集・整形し、複数のアニメ作品に対するネタバレデータセットを構築する。

#### 4.1 データ収集

ここでは、アニメに連動した投稿されたツイートとリアルタイムに投稿されるツイートの収集方法を示す。

##### 4.1.1 番組連動ツイートの収集

今回データセットを構築するにあたって 9 個のアニメ作品から 12 話分を選定した。選定したアニメの放送開始と同時に連動して投稿されるツイートを収集する。ここで、アニメやドラマなどの視聴者のすべてのツイートを収集するためには、その時間帯のすべてのツイートを収集し、選別する必要がある。精度問題が生じる。また、フォローされている人のみにツイートを公開されている場合に、そうしたツイートを収集することは出来ない。

ここでアニメなどの作品を視聴しながらリアルタイムで投稿する際、ハッシュタグと呼ばれる検索およびタグ付けを可能とするキーワードをツイートに付与することが多い。一方で、ハッシュタグを付与せずに作品の内容を投稿しているユーザも多く、扱いが難しいものである。ここでは、アニメ番組に対するハッシュタグが付与されたツイートは、そのコンテンツに対するハッシュタグが付与されていないツイートも代表していると考え、学習および分類に使用する。

なお、ここで選定した番組について、そのアニメ番組などに関してTwitter上で実況に用いられるハッシュタグ(#シャロット, #Sakurakosan, #tokyomxなど)を設定し、Twitter Search APIを利用して関連するツイートを収集した。

#### 4.1.2 一般ツイートの収集

日本語で投稿された全ツイートの中から無作為に一部を取得するためにTwitterのStreamingAPIを利用し、日本語で投稿された全ツイートの中から無作為に5000件収集した。このとき、データにアニメのネタバレが極力含まれないように収集する時間帯を考慮し、東京の地域ではアニメ番組が放送されていなかった2015年1月9日16時から1時間ツイートを収集した。

#### 4.1.3 データ整形

収集した投稿の中には分類を行うデータとして不適切なものが含まれており、下記に示すパターンマッチによるテキスト処理を行った。

(1)ボット(bot)と呼ばれる自動発言システムによる番組に連動した投稿の多くは、放送開始・終了等を知らせる広告であり、番組の内容について言及するものではなく不要である。そこで、「【自動】」または「【定期】」を含む投稿をボットの発言として除去した。

(2)スパムツイートは番組に無関係のため不要である。そこで、「http」を含む投稿をスパムツイートとして除去した。

(3)リツイート(RT)と呼ばれる他人の発言を引用できる機能による投稿はテキストデータが重複するため不要である。そこで、「RT」を含む投稿をリツイートとして除去した。

(4)番組連動ツイートに含まれるネタバレとは無関係の文字列であるハッシュタグを「#」から改行までの文字列として除去した。また、一般ツイートを収集する場合、ネタバレが含まれる可能性があるアニメに関するツイートを避ける必要がある。そこで「#」を含む投稿をアニメに関するツイートとして除去した。

#### 4.2 番組連動ツイート評価システム

機械学習のため、収集したツイートがネタバレかどうかということを手による分類により行った。ここでは、選定したアニメ番組に連動したツイートからそ

れぞれ2000または3000件ずつ無作為に抽出したツイートを対象とし、1話につき3人の評価者に分類を行ってもらった。

なお、分類作業を行ってもらうために、図1に示すウェブシステムを開発した。このシステムでは、ユーザは最初にアカウント名を入力してログインし、ページ上に提示されているツイートに対してネタバレと感じるものを複数選択するというものである。ツイートは4.1.3項と同じテキスト処理を行ったものであり、時間順ではなくランダムに提示される。

図2 開発したウェブシステム

データセット構築者には作成したウェブサイトアクセスしてもらい、直前の話の内容を確認した後、ツイートの分類を行ってもらった。なお、ツイートに対する分類結果は100件毎にデータベースに記録されるため、途中で中断して再開することも可能となっている。データセット構築者はTwitterを普段から用いており、かつ分類するアニメ作品の選定した話数まで視聴済みの20代の大学生の男性16名と女性3名である。

#### 4.3 一般ツイート・番組連動ツイートで構成されたデータセット

一般的に、Twitterにおいてユーザの友人や知人の投稿内容は投稿ごとに言及するジャンルが異なることが多く、その中にネタバレが含まれてしまっているという状況を想定したデータセットを構築する必要がある。そこで、ネタバレ分類システムにおいて評価者3人の

うち2人以上がネタバレと判定したツイートをネタバレツイート、4.1.2項で収集したツイートをネタバレでないツイートとして使用し、アニメ作品を同一作品、ミステリー系、バトル系の3つのジャンルに分けてデータセットを構築した。また、ネタバレツイートと非ネタバレツイートの学習量に偏りを無くすため、アンダーサンプリングを行った。ここでは、ネタバレツイートと同数の非ネタバレツイートを無作為に選定した。なお、このデータセットの情報を表1に示す。

表1 構築したデータセット

ジャンル	作品名	ネタバレツイート数
同一作品	Charlotte(シャーロット) 第4話	232
	Charlotte(シャーロット) 第7話	91
	Charlotte(シャーロット) 第9話	520
	Charlotte(シャーロット) 第13話	359
ミステリー系	六花の勇者 第12話	283
	すべてがFになる 第10話	133
	櫻子さんの足元には死体が埋まっている 第11話	37
	終物語 第5話	259
バトル系	Fate/stay night[Unlimited Blade Works] 第24話	381
	遊戯王 ARC-V 第82話	480
	ワンパンマン 第11話	140
	黒子のバスケ 第75話	338

## 5. 評価実験

ここでは一般ツイートに含まれるアニメの放送に連動して投稿されたネタバレツイートの判定精度を算出する。

### 5.1 評価尺度

本研究ではクラスは2つあり、このクラスはネタバレとなるクラス(正例)と非ネタバレ(負例)である。この2クラスにおける評価尺度として、適合率(Precision)と再現率(Recall)を用いる。ここで、クラス $C_i$ に対して、適合率と再現率は以下のように算出される。

$$Precision(C_i) = \frac{\text{正しく } C_i \text{ に分類されたツイート数}}{C_i \text{ に分類されたツイート数}}$$

$$Recall(C_i) = \frac{\text{正しく } C_i \text{ に分類されたツイート数}}{C_i \text{ に属するツイート数}}$$

番組を楽しみに待っている視聴者にとってネタバレは可能な限り回避したいものであるためネタバレを

含むツイートをできるだけ網羅することが重要である(正例をできるだけ網羅することが重要)。そこで本研究では、番組が放送されるまでの期間中のみある程度ネタバレでないツイートを遮断しても仕方ないものとし、正例(ネタバレ)分類に対する再現率に特に注目して評価を行う。

## 5.2 結果と考察

同一作品では過去の話数分のツイートを学習データとして利用し、新しい話数をテストデータとして適合率、再現率を算出する。ミステリー系、バトル系では3話分を学習データとして利用し、残り1話分の作品をテストデータとして精度を算出する。これを作品全てに対して算出し、適合率と再現率の平均を計算する。なお、機械学習におけるネタバレの判定精度については、3章で提案したベースライン手法、正規化手法、人物名一般化手法、正規化・人物名一般化組み合わせ手法でそれぞれ算出する。

判定精度を図3~10に示す。なお、図3~6の横軸は判定した話数を表しており、図7~10は利用したツールを表している。また、図3, 5, 7, 9の縦軸は適合率、図4, 6, 8, 10の縦軸は再現率を表している。

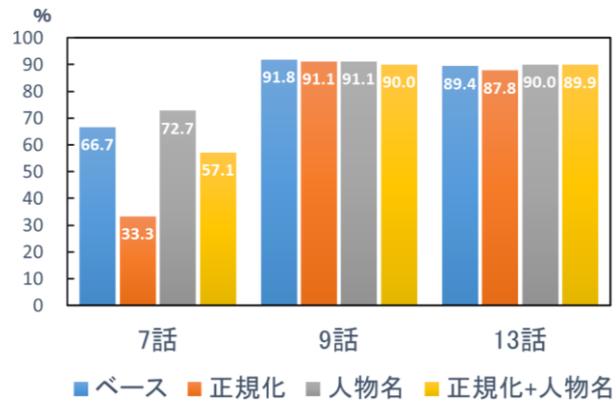


図3 同一作品の話数ごとの適合率 (形態素解析)

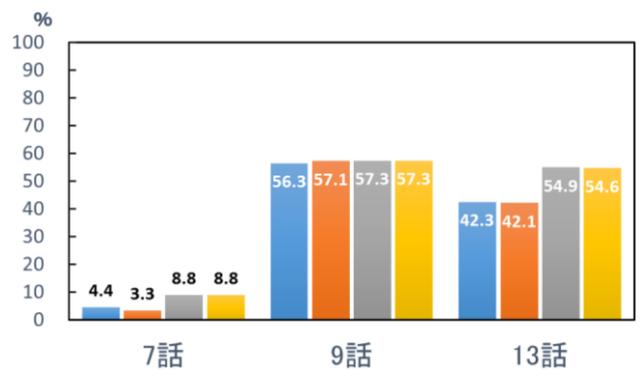


図4 同一作品の話数ごとの再現率 (形態素解析)

図3, 4より、同一作品を判定する上で形態素解析

を用いたとき、7話のネタバレ判定がほぼできず、それ以降の話数では再現率、適合率ともに上昇しているが、13話の判定では9話と比べて再現率が下がっていることが分かる。また、正規化手法では7、13話の両方の精度が下がっており、人物名一般化手法では再現率が全て改善されている。

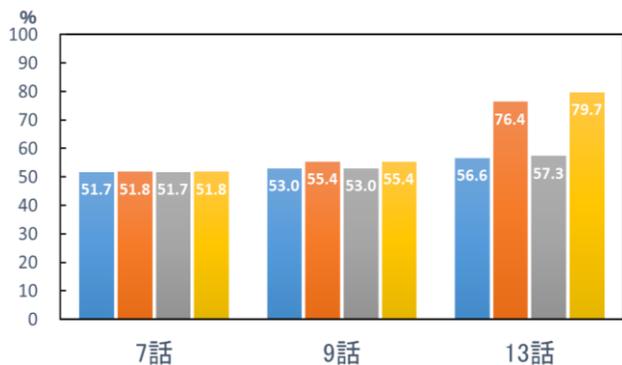


図5 同一作品の話数ごとの適合率（係り受け解析）

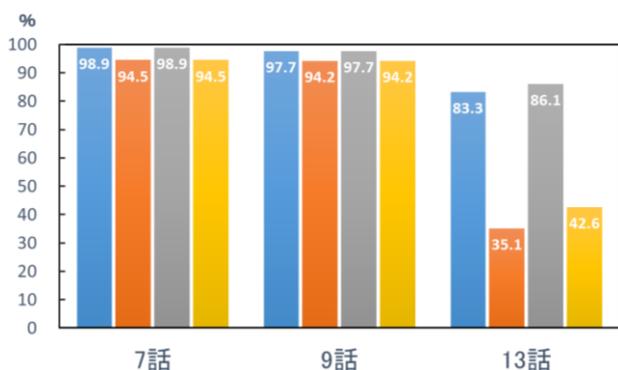


図6 同一作品の話数ごとの再現率（係り受け解析）

図5、6では係り受け解析を用いた結果、形態素解析と比べて全体的に再現率がかなり高くなることが分かる。特に7話から再現率が高く、先の話数に進むにつれて少しずつ再現率が下がっていることが分かる。正規化手法では全ての再現率が下がっており、13話の判定では特に下がった。

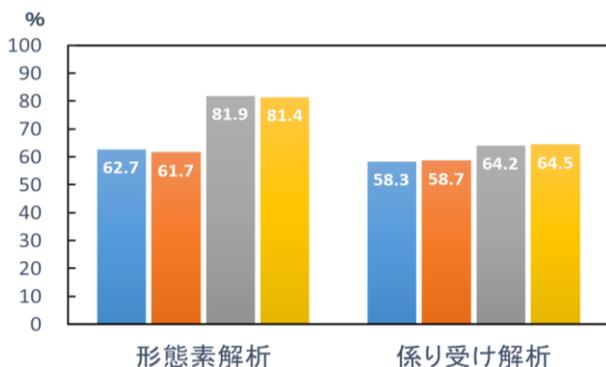


図7 ミステリー系における手法ごとの適合率

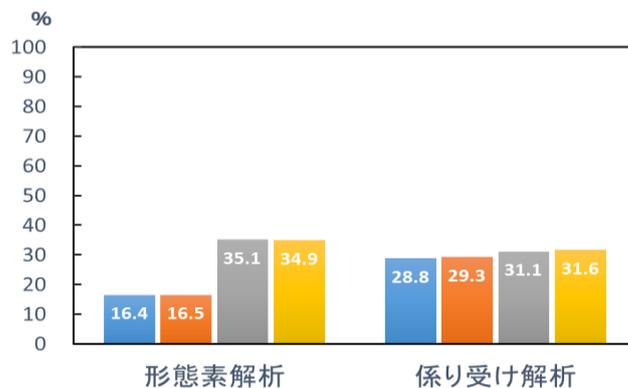


図8 ミステリー系における手法ごとの再現率

図7、8では、全体的にかなり再現率が低かったが、形態素解析を用いた場合、人物名一般化手法によって適合率と再現率がかなり改善され、係り受け解析よりも判定精度が高くなっていることが分かる。特に形態素解析では適合率が約20%、再現率が約19%改善した。

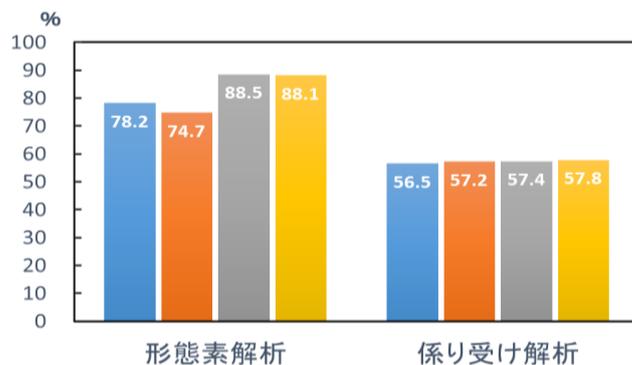


図9 バトル系における手法ごとの適合率

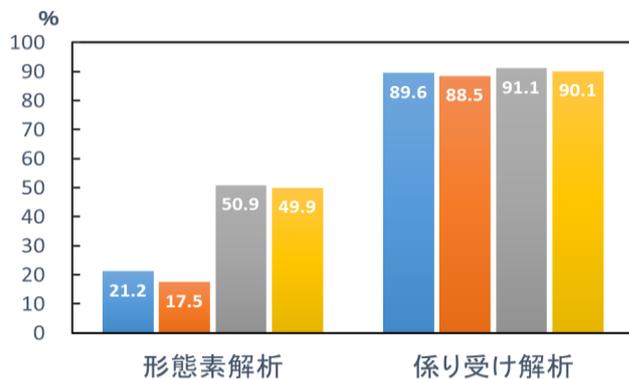


図10 バトル系における手法ごとの再現率

図9、10では、係り受け解析を用いたときに再現率がかなり高くなることが分かる。また正規化手法では精度の改善はせず、人物名一般化手法では形態素解析

と係り受け解析の両方で精度の改善が見られた。特に形態素解析では再現率が約 30%改善した。

ここで全体的な結果として、まず、形態素解析と係り受け解析を用いた判定結果を比較したところ、基本的には係り受け解析をしたときにアニメのネタバレ判定をしやすいことが明らかになった。しかし、ミステリー系においては形態素解析をしたときの精度が比較的やや高かった。これはミステリー系のアニメにおいては文節ごとの関係性よりも単語の情報が重要であったためと考えられる。そこで、アニメジャンルによって用いる単語ベクトル生成手法を変える必要があると考えられる。

次に、ジャンルごとの結果をまとめると同一作品とバトル系のネタバレ判定はしやすく、ミステリー系のネタバレ判定は困難であることが明らかになった。その理由として、同一作品ではその作品に特有な語を学習可能であり、また、バトル系では「勝った」、「負けた」といった勝敗に関する語や「死んだ」、「生きていた」といった人物の生死に関する語が多用され、それらが判定に有効な形態素となったためであると考えられる。一方、ミステリー系では作品ごとに特有な語が多く出現し、判定に有効な形態素が少なかったためであると考えられる。また、係り受け解析を用いるとき後半の話数のネタバレ判定精度が下がった理由として、アニメでは作品全体の前半部分と後半部分の展開が異なることが多く、有効な形態素が前半部分と後半部分で異なったためと考えられる。そこで、同一作品の最新話のネタバレを判定するときは過去に投稿されたその作品全ての番組連動ツイートを利用してデータセット構築するのではなく、最新話から数話前までの番組連動ツイートのみを利用する必要があると考えられる。

最後に手法ごとの結果をまとめると、正規化手法ではほぼ精度の改善は見込めないが、人物名一般化手法では精度が改善することが明らかになった。また、正規化手法が有効でなかったため正規化・人物名一般化手法も人物名一般化手法と比べて有効でなかった。正規化手法については正例、負例のどちらにおいても感情を表現する連続語がほぼ等しく出現したため精度が向上せず、人物名一般化手法についてはネタバレツイートのみに一般化された登場人物名が多く含まれていたため精度が向上したと考えられる。

ここで、実際に機械学習を通して出力されたデータを確認した上でネタバレと判定することが困難であったツイートの特徴を下記に示す。

- (1)作品に特有な語・専門用語が含まれている。
- (2)登場人物名があだ名や名称に置き換えられた単語が含まれている。
- (3)登場人物と「!」などの記号・絵文字の組み合わせ

のみで記述されている。

以上が Twitter におけるアニメのネタバレを判定する上での課題であると考えられる。

## 6.まとめ

本研究では複数のアニメ作品の番組に連動したツイートと一般ツイートを収集し、システムを使用して人手でツイートがネタバレかどうか分類することでネタバレデータセットを構築し、それぞれの分類器を作成した。また、ネタバレを SVM で学習した結果、Twitter におけるアニメのネタバレを判定する上で係り受け解析による単語ベクトル生成は基本的に必須であること、手法では人物名一般化手法で判定精度が向上すること、また、同一作品とバトル系のネタバレ判定はしやすく、一方でミステリー系のネタバレ判定は困難であることが明らかになった。今後の展開としては扱うストーリーコンテンツの量を増やすこと、作品ごとに特有な語をパターンマッチにより判定すること、人物名一般化手法において登場人物の性別を区別することによってネタバレ判定精度の向上を目指す。

## 謝辞

本研究の一部は、JST CREST, 明治大学重点研究 A, 重点研究 B の支援を受けたものである。

## 参 考 文 献

- [1] アンケート 100 人に聞きました!, <http://www.enquete.ne.jp/hundred/>
- [2] 田島一樹, 中村聡史:ストーリーコンテンツに対するネタバレの基礎調査とその判定手法の検討, 研究報告グループウェアとネットワークサービス (GN), 2015-GN-96, Vol.7, pp.1-6(2015).
- [3] 中村聡史, 小松孝徳: スポーツの勝敗にまつわるネタバレ防止手法: 情報曖昧化の可能性, 情報処理学会論文誌 54(4), pp. 1402-1412 (2013).
- [4] 白鳥 裕士, 中村 聡史: SNS 上でのサッカーの試合に対する直接的・間接的ネタバレの分析, 研究報告グループウェアとネットワークサービス (GN), 2015-GN-96, vol 8, pp.1-8 (2015-09-25).
- [5] Jennifer Golbeck: The Twitter Mute Button: A Web Filtering Challenge, Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI 2012), pp. 2755-2758 (2012).
- [6] 中村聡史, 川連一将: スポーツのネタバレを防止する Twitter クライアントの開発と諸検討, 第 4 回 ARG Web インテリジェンスとインタラクション研究会 (2014).
- [7] 池田郁, 土方嘉徳, 西田正吾: レビュー文からのあらすじ除去と人名特定に関する基礎検討, 自動制御連合講演会講演論文集, 52(0), pp.239-239 (2009).
- [8] 前田恭佑, 土方嘉徳, 中村聡史, ストーリー文書内のネタバレの記述に関する基礎的調査, 第 6 回 ARG Web インテリジェンスとインタラクション研究会, 2015.
- [9] 田中駿, 廣田壮一郎, 高村大也: コメント機能付動画共有サービスにおけるネタバレ検知, 第 29

