

Twitter における映画感想ツイートの自動抽出

瓜生奈月 三宅陽菜 (情報科学ゼミ)

1. はじめに

近年、Twitter、Facebook、LINE などのソーシャルネットワークサービス(SNS)は世界中に広まっており、日本でも多くの人々が利用している。Twitter は、リアルタイムでより多くの情報を共有できることから、LINE や Facebook などの他の SNS に比べ、感想や評判などを得る事に適したサービスであると言える。そのため、鑑賞する映画や食事するお店を選ぶときの参考に、Twitter を利用する人が多くなっている[1]。

映画を鑑賞する前に、その映画を既に観た人がどのような感想を持ったかを知るために Twitter を利用した場合には、いくつかの問題が存在する。一つは、Twitter の検索機能を使用して映画について検索をかけても、すべてのツイートを見るには時間がかかるため一部の Twitter しか参考にできない。もう一つは、「今日は〇〇を見にいった」「〇〇面白そう」など、感想が含まれていないツイートも表示されてしまうため、感想の含まれる有益な情報とそうでない情報とを自身で区別するという手間が必要となってくる。これらの問題を解決する方法として、Twitter を利用しない人がツイートをまとめたサイトや個人のブログなどを利用して感想や評判を調べる方法もあるが、少数の意見のまとめだったり、そのサイトを運営する個人的な意見だったり、偏った意見しか得ることができない。

そこで本研究では、映画を鑑賞した後につぶやかれたツイートの感想情報を自動的に収集し、ツイートに含まれる感想情報キーワードをもとにツイートを分類することで、Twitter 上にある映画の感想情報を集計する仕組みを提案する。本研究での感想情報キーワードとは、「面白い」「感動した」などの感想に関連する単語のことを呼ぶ。

2. 事前調査

感想情報を含むツイートを抽出する際、映画タイトルのみで検索した場合に、どのような問題があるのかを確認するため、「ラストレシピ」を例にして約 1 時間分のツイートを収集し調査した。調査結果から、映画の感想情報として有益でないツイートも一緒に収集されてしまうことが分かった。有益でないツイートの内容と、それを取り除くための改善方法を検討した。検討結果を下記に示す。

- ①bot 及びキャンペーンなどにより、同じ内容がリツイートされただけのツイート
⇒全く同じ内容のツイートを取り除く。
- ②舞台挨拶、トークショーなどのチケット取引のツイート
⇒「チケット」「舞台挨拶」等のワードを含む場合には、そのツイートを取り除く
- ③「面白そう」などの映画を鑑賞する前のツイート
⇒「●●そう」というキーワードが含まれるツイートを取り除く
- ④映画鑑賞後のツイートだが、感想情報が含まれていないツイート
⇒感想情報キーワードが含まれていないツイートを取り除く。

3. 感想情報キーワードの設定

本手法では、2017年に公開された映画をジャンルごとに2つずつ取り上げ、感想情報の含まれるツイートを収集し、感想情報キーワードとして適切なキーワードを収集した。

- ・アクション 「スパイダーマン」「亜人」
- ・サスペンス 「三度目の殺人」「22年目の告白」
- ・ミステリー 「ユリゴコロ」「ナミヤ雑貨店の奇蹟」
- ・恋愛 「ナラタージュ」「恋と嘘」
- ・アニメ 「メアリと魔女の花」「名探偵コナン-から紅の恋歌-」
- ・歴史 「関ヶ原」「忍びの国」
- ・コメディ 「斉木楠雄のΨ難」「帝一の國」

【収集の方法】

- ①Twitterの検索機能を用いて「映画のタイトル -そう since:公開日 until:公開日から4日後」というキーワードでツイートを収集する。【例】「スパイダーマン -そう since:2017-8-11 until:2017-8-15」
- ②映画タイトルのみで検索すると、「面白そう」などまだ映画を見ていない人のツイートまで表示されてしまう可能性があるため「そう」を省き収集することとする。
- ③検索結果から得られたツイートに1つずつ目を通し「面白かった」「感動した」などの感想ごとに手動で振り分ける。

[カウントするツイート例]

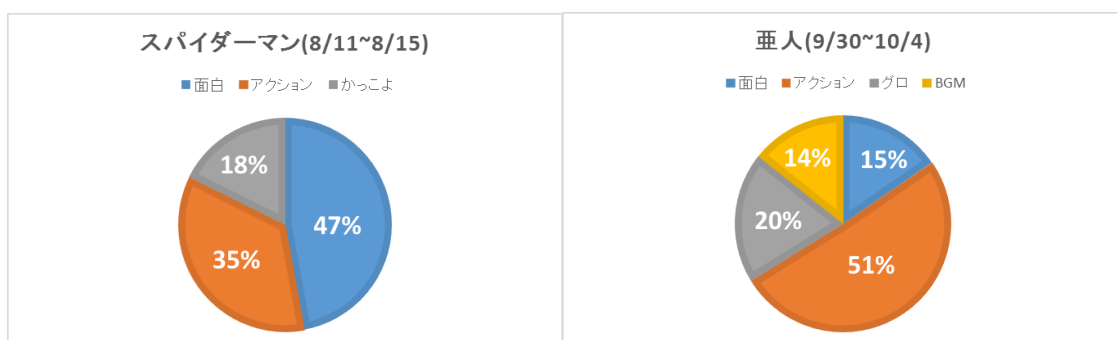
「スパイダーマン見てきた！面白かった！」
⇒「面白い」という感想情報が含まれている。

[カウントしないツイート例]

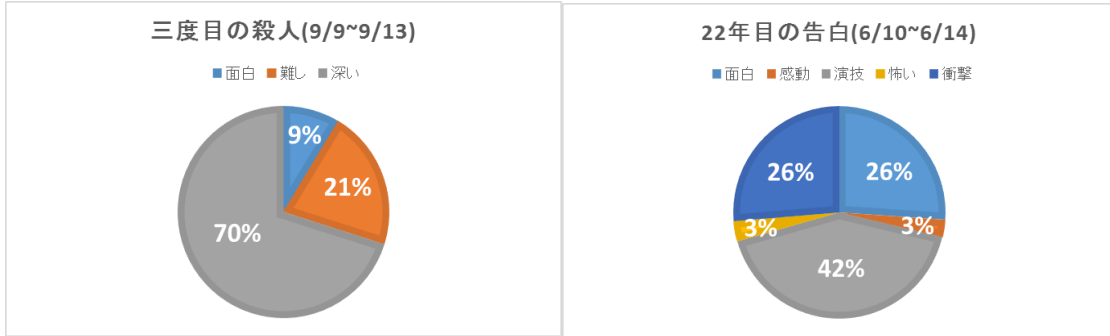
「スパイダーマン面白そう！」「スパイダーマン見てきた」
⇒「面白そう」という予想、「見てきた」という情報のみで感想情報が含まれていない。

【収集の結果】

- ・アクション



・サスペンス



・ミステリー



・恋愛



・アニメ



・歴史



・コメディ



これらの結果をもとにして多くの映画に共通して見られるキーワード、ジャンルごとによく見られるキーワードを表1に示す。なお、同じ意味・同じフレーズであっても言い回しが異なる為、1つの感想の種類につき複数の感想情報キーワードを採用した。例えば、「面白い」の感想情報キーワードとして、「面白」、「おもしろ」、「おもしろ」を採用した。

表1 感想情報のキーワード

感想情報の分類名	感想情報キーワード
面白い	「面白」、「おもしろ」、「おもしろ」
良かった	「良かった」、「よかった」
笑った	「笑った」、「笑え」、「笑い」
泣けた	「泣けた」、「泣いた」、「感動」
切なかった	「切な」
キュンキュンした	「キュン」、「きゅん」
怖かった	「怖」、「恐」
難しかった	「難し」
アクション	「アクション」
セリフ	「台詞」、「セリフ」
曲がいい	「主題歌」、「曲」
演技	「演技」

4. 自動収集の手法

以下の①～③の方法を用いて、自動収集を実施した。

- ① レンタルサーバから Twitter API を利用して、「映画のタイトル」でツイートを自動収集する。
- ② 広告ツイートを削除するため、収集したツイートのうち、まったく同じ内容のツイートを削除する。
- ③ 映画鑑賞後のツイートのみを抽出するため、ツイート内に「観た」というフレーズが含まれているツイートのみを抽出する。
- ④ 感想キーワード（「面白」、「泣いた」）を含むツイート数を感想キーワードごとにカウントする。

5. 結果と考察

4章で説明した自動収集方法をもとに映画の感想情報の自動抽出を実施した。抽出期間と抽出した映画を表2に示す。次に、抽出したツイートの総数を表3に示す。さらに、抽出した映画の感想情報を図1に示す。図1は、感想の種類各映画のツイート数を各映画のツイート総数で割った割合で示している。例えば、映画「泥棒役者」の「面白い」という感想が含まれたツイートの総数は、342ツイートであり、映画「泥棒役者」のツイート総数は2205ツイートであるので、映画「泥棒役者」の感想ツイートのうち「面白い」というツイートの割合は、約15.51% (342/2205) と計算した。

表2 抽出期間と抽出映画

抽出期間	2017年11月18日～2018年1月8日
抽出映画	「泥棒役者」「火花」「覆面系ノイズ」「鋼の錬金術師」「8年越しの花嫁」

表3 抽出したツイート数

映画名	ツイート数
泥棒役者	2,205
火花	1983
覆面系ノイズ	502
鋼の錬金術師	4,832
8年越しの花嫁	2,970

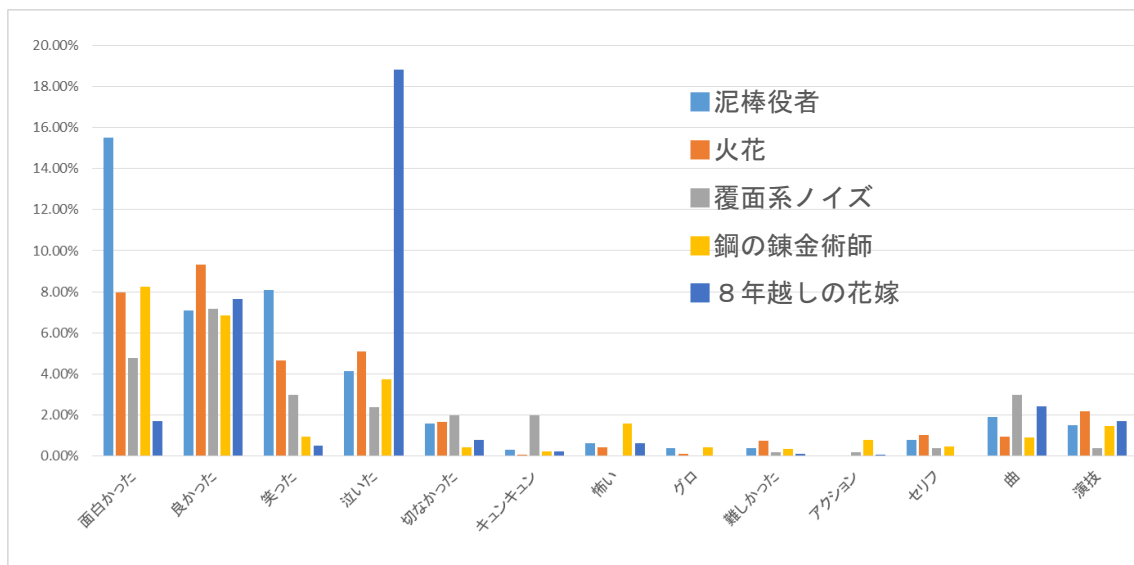


図1 収集結果

図1より、泥棒役者は「面白かった」「笑った」という感想が多いことから、コメディ要素の強い作品であることがわかる。それと同様に、火花は「難しかった」、覆面系ノイズは「曲」「キュンキュン」、鋼の錬金術師では「怖い」「アクション」、8年越しの花嫁は「泣いた」といった感想が多くみられ、映画によって感想情報が大きく異なっている。このことから、本手法で収集したツイートの感想情報の集計結果から、映画を観る前の人に、各映画がどんな映画であるかを推測できる情報を提供できていることを示している。ただし、今回収集結果で得られた映画の感想情報が信頼できるものであるか否かの評価を実施していないため、その信頼性については今後評価していく必要があると考えている。

6. おわりに

本研究では、Twitterの映画の感想ツイートを用いて鑑賞する上で参考となるサイトを立ち上げるため、Twitterのツイートから映画に関する感想情報を自動で抽出し、感想の種類ごとに分類する手法を提案した。今回、本研究の最終目的である「サイト作成」まで辿り着くことは出来なかったが、ツイート抽出を自動化し、映画を観た人のツイートの中に各感想の種類が含まれる割合の情報を提供するとともに

実現できた。しかし、自動抽出によって得られた感想情報が信頼できるものであるかの評価をしていないことと、その感想情報が映画を観る前の人の役に立つのかどうかの評価をしていないことが課題であり、今後それらを確認する方法を検討していきたい。

参考文献

- [1] 『Twitter を利用する&利用しない理由』, ITmedia ビジネスオンライン,
<http://bizmakoto.jp/makoto/articles/0912/28/news022.html>