

Twitterにおける 映画感想ツイートの 自動抽出

情報科学ゼミ

65-9007 瓜生奈月 65-9053 三宅陽菜

Twitterとは

- ・2006年7月にオブビアウス社（現:Twitter社）が開始したウェブサービス。
- ・サービス名のTwitterとは英語で「さえずり・興奮」「無駄話」という意味。

参照 : <https://ja.wikipedia.org/wiki/Twitter>



● 検索

入力したワードを含むツイートを表示する機能

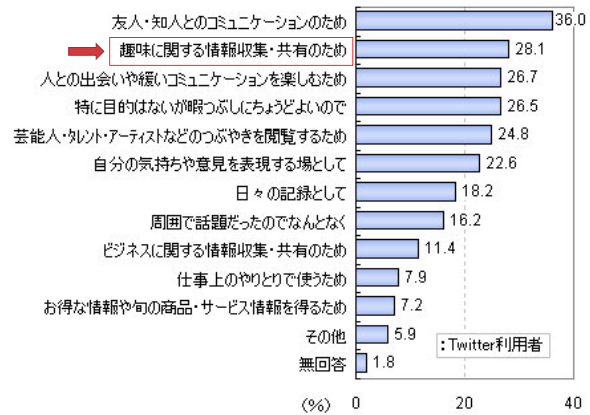
● tweet (ツイート)

Twitterでの投稿を指す。「鳥のさえずり」を意味する英語であり日本語では「つぶやき」と意識される。

お借りした画像<http://imgc.appbank.net/c/wpcontent/uploads/2017/04/333.jpg>

研究の動機

- ・学生が使う身近なウェブサービスの1つであるTwitterを使った研究を行いたいと考えた。
- ・「Twitterを利用する理由」の調査結果にもとづき、2番目の利用理由である「趣味に関する情報収集・共有のため」に着目した。趣味の中でも映画の情報をわかりやすく収集する方法を考えたい。



Itmediaビジネス ONLINE 「Twitterを利用する&利用しない理由」
<http://bizmakoto.jp/makoto/articles/0912/28/news022.html> より引用

研究の動機

- ・ 学生が使う身近なウェブサービスの1つであるTwitterを使った研究を行いたいと考えた。
- ・ 「Twitterを利用する理由」の調査結果にもとづき、2番目の利用理由である「趣味に関する情報収集・共有のため」に着目した。趣味の中でも映画の情報をわかりやすく収集する方法を考えたい。

研究の目的

「面白い」「感動した」などの感想を含むツイートを自動的に抽出・振り分けできれば、映画を観る上で有益な情報を簡単に得ることができる。それらのツイートの数や頻度から映画の評判を調査する。

Twitterを利用している人は勿論、利用していない人でも簡単にTwitter上の感想情報を得ることができれば良いと考えた。

本研究では
Twitterの映画の感想情報を用いた鑑賞する上で参考となるサイトを立ち上げる
 ことを目的とする。

自動抽出における問題点と改善方法

感想を含むツイートを自動的に抽出するにあたって、映画タイトルのみで検索すると有益でないツイートも一緒に抽出されてしまうという問題点がある。


- ① botや映画公式アカウント、コンビニなどとのコラボやキャンペーンのツイート

⇒ 全く同じツイートを省くよう設定する。



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
ワードを省くよう設定する

 フォローする

ラストレシビ公開記念トークショーの交換し
て下さる方を探しています。


【譲】 15:20 1枚 L~O列 同席
【求】 18:50 1枚

該当の方おられましたらDMにてご連絡くだ
さい。
よろしくお願いします


#ラストレシビ #ラストレシビ交換 #ラストレ
シビトークショー

③「面白そう」という予想や、
「見てきた」など感想の含まれていないツイート

⇒「そう」という予想のワードを省いたり
「見た」というワードも検索に入れる

 フォローする

みなさん、おやすみなさい!!明日はかなり遅め
のラストレシビを見られます👏とても面白そ
うだし、楽しみです🥳

 フォローする

ラストレシビ見てきた～～!
明日で終わりだって!!!!ギリ!!!!

これら3点を改善する必要があることがわかった。
そして、感想情報を分類するためのキーワードを
調査するため、次の研究を行った。

研究内容①

2017年に公開された映画を各ジャンルごとに2つずつ取り上げ
感感情報の含まれるツイートを収集した。


取り上げた映画は以下の通りである。

- ・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

1. Twitterで検索

- ・「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検索し、ツイートを収集する。

 スパイダーマン-そう since:2017-8-11 until:2017-8-15

2. 不必要なツイートの除去

- ・問題点で挙げたキャンペーン等のツイート、
チケット取引のツイートに関しては、目で見て確認し省く。

3. 検索結果の振り分け

- 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。

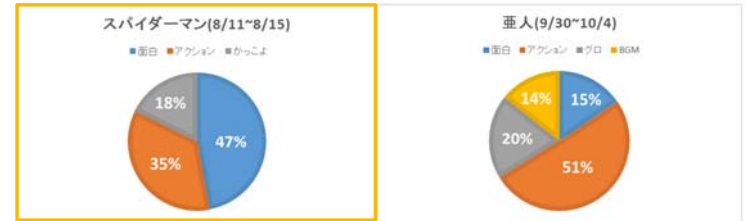
[カウントしないツイート例]

「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



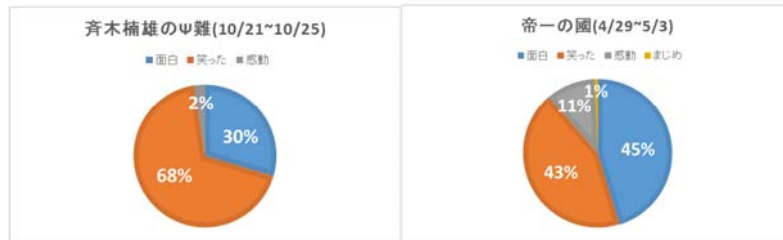
・アニメ



・歴史



・コメディ



先ほど取り上げた問題点に対する改善方法を手動で確認した結果、
ほぼ正確な結果を得ることができた。

研究内容②

手動で抽出した結果をもとにして
多くの映画に共通して見られるキーワード、
各ジャンルごとによく見られるキーワードを下記に取り上げた。

感知情報の分類名	感知情報キーワード
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」

収集の方法

1. Twitterから映画に関するツイートの自動収集

- ・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. 広告ツイート等の削除

- ・収集したツイートのうち、ツイート内容が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抽出

- ・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. 感知情報の抽出

- ・感想キーワード（「面白」、「泣いた」、「恐」）を含むツイート数を各感想キーワードごとにカウントする。

自動抽出の概要

- ・抽出期間：2017/11/18～2018/1/8

・抽出映画

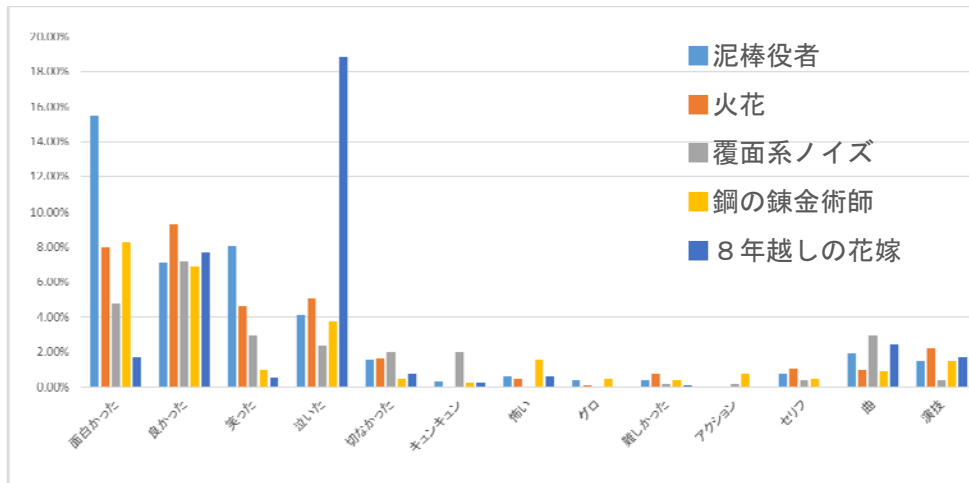
「泥棒役者」「火花」「覆面系ノイズ」
「鋼の錬金術師」「8年越しの花嫁」

・抽出したツイート数

（各映画名を含み、同一ツイート削除、「観た」含む）

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

🐦 収集結果 🐦



まとめ🐦

- ・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できなかったが、ツイート抽出を自動化し、感想情報の含まれたツイートのみをほぼ正確に得られた。
- ・今後の課題として、自動抽出の精度をさらに上げることと、実際にサイトを作成し公開できればと考えた。

ご清聴ありがとうございました

