

ツイートのカテゴリ化によるアニメーション作品のヒット要因に関する研究

Research on hit factors of animation work by categorizing tweets

田爪 聡^{*1} 大知 正直^{*1} 榊 剛史^{*2} 坂田 一郎^{*2} 森 純一郎^{*2}
 So Tazume Masanao Ochi Takeshi Sakaki Ichiro Sakata Junichiro Mori

^{*1} 東京大学工学部システム創成学科 Faculty of Engineering, The University of Tokyo
^{*2} 東京大学 The University of Tokyo

In this research, we propose a method to quantify the work attribute of each animation work by estimating the time change of tweet number for each topic from tweet data including title of animation work.

Specifically, we set tweet data as one sentence for each hour and apply topic model analysis to those sentences. For each topic outputted in this method, each topic is classified into a more abstract topic with reference to the word generation probability of the topic. In this research, we call that abstract topic 'a topic category'. Based on the topic generation probability in each sentence, time transition of the number of tweets of each topic category is estimated and the relationship between the sales of the work and the tweet data is examined by designing the feature quantity for each work.

1. 内容概論

本研究ではアニメーション作品のタイトルを含むツイートデータからトピックごとにツイート数の時間変化を推定することで、各アニメーション作品の作品属性を定量化する手法を提案する。

具体的にはツイートデータを時間ごとに一つの文章とし、これらの文章に対してトピックモデル解析を施す。こうして出力された各トピックに対し、そのトピックの単語生成確率を参考にして、各トピックをより抽象的なトピックに分類する。本研究ではその抽象的なトピックをトピックカテゴリーと呼ぶ。各文章におけるトピック生成確率を元に各トピックカテゴリーのツイート数の時間推移を推定し、作品ごとに特徴量を設計することで、作品の売り上げとツイートデータとの関係を調べる。

2. 目的と貢献

2.1 目的

近年、ツイッターのデータを用いた市場分析が広く行われている。中でもツイートの数を用いて映画の興行収入を予測する研究や、他のウェブサービスからのデータと組み合わせることで、漫画の売り上げ予測の精度を上げる研究などが行われている[1][2]。しかしデータのテキストの意味を考慮していない単純な言及の量で売り上げの予測を行う場合、予測を行う対象によっては売り上げに関係のない要素がノイズとして働き、予測の精度を著しく下げることが先行研究でも示されている[3]。

そこで本研究の目的は、全ツイートを要素ごとに分類することで、それらとアニメーションの売り上げの関係を分析することである。例えばアニメーション作品に関するツイートにおいて「アニメが面白かった」等のアニメの感想に関するツイートは映像ソフトの売り上げに大きく相関していると考えられるが、「原作者が炎上した」という話題は売り上げと相関が小さいと予想できる。そこで本研究ではそれぞれの話題に関するツイートを各トピックカテゴリーに分類し、各カテゴリーと売り上げとの関係を調べる。

2016年は多くのアニメーション作品がヒットした。しかしそのヒ

ット要因の分析は主観的なものが多く、客観的な分析が行われていないのが現状である。制作者にとってヒットした作品のヒット要因の分析は次のヒット作品の制作の手助けになる。そこで本研究では各要素に対して特徴量を設計し、その特徴量を用いることでヒット要因の分析手法を提案する。

2.2 貢献

本研究の貢献は以下の通りである。

- ツイートのテキストの内容を考慮した特徴量の設計に成功し、その特徴量を用いることでアニメーション作品の売り上げ予測の精度を向上させる手法の可能性を示した。
- 各トピックから得られた特徴量を用いて、作品属性の抽出に成功した。また作品属性とアンケート調査の結果を比較することで、被験者の好みの傾向を確認することに成功した。
- 各トピックのツイートの量を時系列に可視化することで、本手法がヒット作品のヒット要因の分析に適用できることを示した。

3. 実験

3.1 解析で用いたデータの概要

本研究では株式会社ホットリンクが収集したツイートデータを用いて、解析を行った。今回扱ったデータは全日本語ツイートの10パーセントサンプリングに対し、アニメーションのタイトル名を含むツイートを抽出したものである。本研究では全部で12本のテレビアニメと映画「君の名は。」を扱った。

3.2 提案手法

ここでは本研究の提案手法についてまとめる。本研究ではツイッターデータを1時間ごとに1文章として Latent Dirichlet Allocation(以下では LDA とする)を適用した。LDA を行うことで、各文章におけるトピックの生成確率、および各トピックにおける語彙生成確率が出力される。本研究では、各文章におけるトピックの生成確率とその文章に対応する時間帯のツイート数との積を計算することで、その時間帯における各トピックのツイート数を推定し、分

析に用いた。

次に各トピックをより抽象的なトピックであるトピックカテゴリーへ分類した。本研究ではトピックカテゴリーはアニメ関連のツイート話題として考えられる、「曲 (music)」、「原作 (author)」、「ネット動画 (netvideo)」、「グッズ (goods)」、「放送後の感想 (on air)」、「twitter 固有、ハッシュタグや提携アプリ (twitter)」、「声優 (actor)」、「映像ソフト (disk)」、「イベント (event)」、「キャラクター (character)」、「宣伝やキャンペーン (CM)」、「まとめサイト等話題やニュース (news)」、「二次創作 (fan art)」、「他作品 (other)」を用いた。これらのトピックカテゴリーは経験的に考えられたものであり、再度評価する必要がある。

3.3 テレビアニメに対する提案手法の適用

ここではテレビアニメに対して提案手法を適用する。テレビアニメのタイトルを含むツイートに LDA を適用し、各トピックをその単語生成確率から手でトピックカテゴリーに分類する。そうすることで各トピックカテゴリーの時間ごとのツイートの数を推定することができる。本研究ではこの推定されたツイート数から特徴量を設計する。特徴量はトピックカテゴリーごとに、ツイートを収集した全期間の一時間ごとの平均ツイート数、一時間当たりの最大ツイート数、期間全体の総ツイート数の3つと、全期間をアニメ放送開始前とアニメ放送開始後の2期間に分けた上で、それぞれに平均値、最大値、総数の3種類を設計した。つまり一つのトピックカテゴリーにつき9つの特徴量が設計される。またトピックカテゴリーは14種類あるので、一つのアニメにつき特徴量が126個設計されたことになる。

まずそれぞれの特徴量を説明変数、映像ソフトの売り上げを目的変数とした単回帰分析を行った。この時評価の指標として、R2乗値を用いた。提案手法を適用していない状態のツイート数で単回帰分析を行った結果、一時間当たりのツイート数を説明変数として用いた時に R2乗値が0.110となり最も良いスコアとなった。次に提案手法をツイートデータに適用し、得られた各トピックカテゴリーから設計された特徴量を用いた。この時「映像ソフト」のトピックカテゴリーから設計された放送開始後の一時間当たりの平均ツイート数が最も良い結果となり、その R2乗値は0.823であった。その結果を以下のグラフで表す。左側が適用前のツイート数、右側が「映像ソフト」のトピックカテゴリーから得られた特徴量で回帰した結果である。縦軸は映像ソフトの売り上げ枚数を示している。

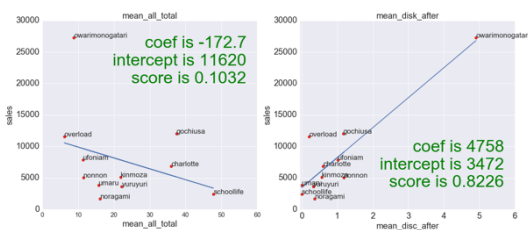


図1. 単回帰分析の結果の比較

図からもわかるように、トピックカテゴリーから得られた特徴量を用いた方がよりアニメーション作品の売り上げを説明できていることがわかる。

次に各トピックカテゴリーに属するツイート数が全体のツイート数に占める割合アニメ放送開始前と、アニメ放送開始後に分け

て算出した。この時距離の計算には JS ダイバージェンスを用いた。それらの値に関して全アニメの平均距離を算出した結果、全作品間の距離の平均は 0.150 となった。またアンケート調査を20人に対して行い、今回扱った作品の中で面白かった作品を答えてもらい、その面白かった作品同士の平均距離を算出した。その結果3作品以下の作品を評価している人の平均距離は 0.127 であり、少ない作品を評価している人は、距離の近い作品を好むということが示された。これはアニメの推薦システムに用いることが期待できる。

また面白いと答えた作品の数が増えるほど、平均距離が伸びた。これは多くの作品が好きなおの好みは幾つかのジャンルにまたがっていることを示しており、これは現実の社会から予測される結果と重なる。

3.4 「君の名は。」に対する提案手法の適用

本手法を「君の名は。」に適用した結果、音楽のトピックカテゴリーとニュースのトピックカテゴリーに特徴的な波形が見られた。以下にそのグラフを示す。

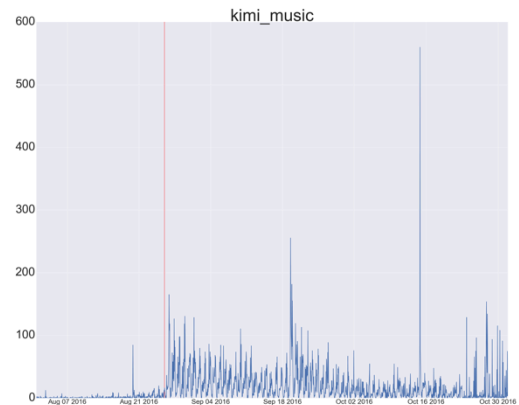


図2. 音楽に関するツイート数の推移

音楽のトピックカテゴリーは映画公開後すぐにそのツイート数を伸ばしている。その後減衰することなく盛んにつぶやかれている。今回の映画では映画を観終わった後に音楽に関するつぶやきが増えたことで、アニメに対して抵抗のある層にも映画館に足を運びやすい状況ができたのではないかと考えられる。それが今回のヒットの第一の要因だと考えられる。

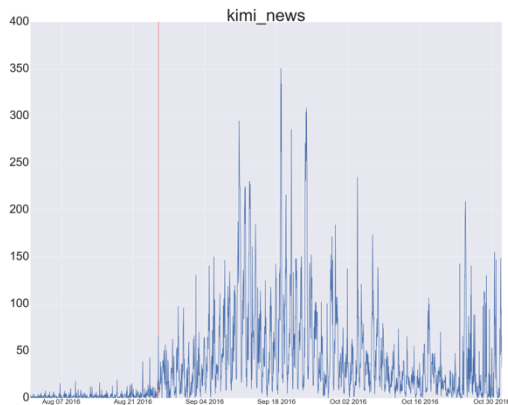


図3. ニュースに関するツイート数の推移

9月の中旬頃は聖地巡礼のトピックでツイート数が伸びていた。これは最近の若者が SNS を通じて聖地巡礼を写真で公開することが今回の映画のヒットを加速させた要因の一つとなっていることを示している。また10月の後半においてツイート数が増加している。この時期は「君の名は。」の興行収入がこれまでの作品の記録を更新していくことが頻繁にニュースに取り上げられていた時期と重なっている。

以上の結果から、「君の名は。」のヒットの流れをまとめると、初期は監督やミュージシャンのコアなファンが映画を見て、その感想を SNS を通じて発信した。その時映画の感想だけでなく劇中歌の感想も同時に流れたことで、アニメに抵抗を持っている層が気兼ねなく作品を見ることができるようになった。次に作品を見た人が増え始めると、流行に敏感な層が聖地巡礼を行い、それが社会現象として流行し始めた。そして、ここまで伸び続けていた興行収入が歴代の記録を塗り替え初めたことで、流行に疎い客層も動員することに成功した。これが今回の分析で得たトピックの移り変わりから考えられるヒット要因である。

4. 結論

4.1 結論

本研究では、アニメーション作品に関するツイートを収集し、単純なツイート数に加え、各トピックの時間毎のツイート数を推定することでアニメーション作品のヒット要因分析を行った。まず、各アニメーション作品のタイトルを含む時間ごとのツイートデータに対し LDA 解析を行い、出力された全トピックをトピックカテゴリーに分類することで、各トピックカテゴリーの時間ごとのツイート数の推移を推定した。また推定された時間毎のツイート数の推移を用いて特徴量を設計した。それらの特徴量を用いることで各作品の作品属性を抽出することが可能となった。

また各トピックカテゴリーに属するツイート数の推移の推定量を用いることで、アニメーション作品「君の名は。」のヒット要因の分析を行った。

4.2 今後の展望

本研究では扱ったアニメーション作品のサンプル数が 12 本であり、ヒット要因の分析には不十分であった。しかし作品数を増やすことで、要因分析の精度向上が期待される。

サンプル数不足の一因として、各トピックを経験的に定めたトピックカテゴリーに手動で分類したことが挙げられる。また、トピックカテゴリーへの分類に外部知識を用いたクラスタリング手法

を用いることで、より精緻なトピックカテゴリーが得られると考えられる。

クラスタリング手法をカテゴリー分類に適用することで、十分なサンプル数確保と精緻なカテゴリー分類が可能になり、要因分析の精度が向上すると考えられる。

参考文献

- [1] Hyunyoung Choi and Hal Varian. Predicting the present with google trends. *Economic Record*, Vol. 88, No. s1, pp. 2–9, 2012.
- [2] ohan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, Vol. 2, No. 1, pp. 1–8, 2011.
- [3] Sharad Goel, Jake M Hofman, Sebastien Lahaie, David M Pennock, and Duncan J Watts. Predicting consumer behavior with web search. *Proceedings of the National academy of sciences*, Vol. 107, No. 41, pp. 17486–17490, 2010.