

SNSにおけるヒット現象の数理モデル

グループ演習7班

北島慧 内田大樹 清水純平

アドバイザー教員 高安亮紀

1 研究背景・目的

近年インターネットは急速に普及しており、総務省の調査によると、インターネットを利用している個人の割合は83.5%であり依然として上昇している [1]。利用手段においても多様性が増してきている。検索エンジンを利用したホームページの閲覧や、電子メールなどの利用に加えて、Facebook, Twitter, ブログなどのソーシャルネットワークサービス（以下 SNS）といわれる、簡易に情報を受信・発信できるメディアの活用が普及している。20 歳代の3分の2以上が SNS・動画投稿を利用しており、SNS の普及率は、各年代で差は見られるものの人々の生活に浸透しているといえる (図 1)。

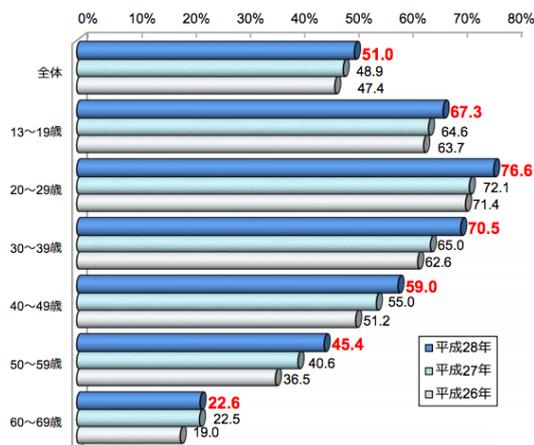


図 1: 年代別 SNS 利用者の割合 ([1] より引用)

さらには個人だけでなく、企業においてもイン

ターネット利用は進んでおり、25%以上の企業がインターネットを広告媒体として活用している (図 2)。このように、社会に広く周知されている SNS において、企業・人々からインターネットに書き込まれた内容は社会の興味・関心を表しているといえる。また、デジタル情報であることから検索・集計することで定量的に扱うこともできる。

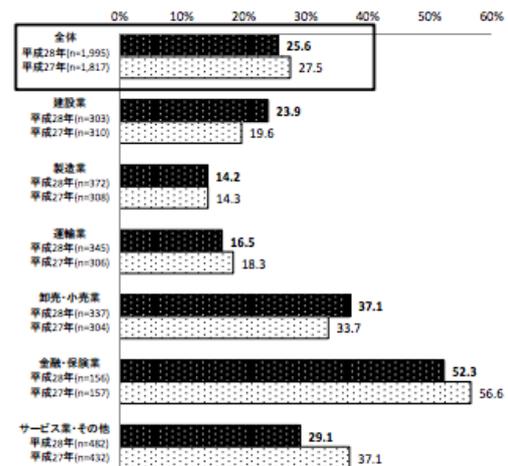


図 2: 産業別インターネット広告の実施状況 ([1] より引用)

石井らは、人々の興味・関心が定量的に評価できると仮定し、ヒット現象の数理モデルを提唱した [2]。ヒット現象とは通常、商品の販売数等が爆発的に増えることや、テレビの視聴率がある数字を越えていくものである。このヒット現象をテレビメディア等への露出と人々の間での直接・間接コ

コミュニケーションから予想できるとしている。石井らはこのモデルを用いて映画やドラマ、コンサートなどが SNS でどれだけ反響を呼ぶかどうかの予測を行い、要因の解析とヒット現象が予測可能であることを示した。

石井らの先行研究では、広告投入量としてテレビの露出量などをもとに、その評判である SNS への投稿数を予測するものであったが、本研究においては入力を Twitter での宣伝がリツイートされた数として、視聴率と Twitter への投稿数を予測する。社会に広く周知されている SNS において、企業からの宣伝ツイートに十分な宣伝効果が認められるのであれば、これにより視聴率と Twitter への投稿数が変化することが予測できる。企業による SNS 上での宣伝からテレビメディア等への影響力を計ることで、SNS 上での宣伝効果の検証を行いその有効性を計る。なお、視聴率データなどの入手しやすさから、本研究では対象を連続テレビドラマに設定した。

2 ヒット現象の数理モデル

2.1 モデルの概要

ヒット現象の数理モデルの内容を石井らの先行研究 [2] をもとに説明する。このモデルでは人々に意欲を駆り立てる要因を以下の 3 つに分類している。

- 宣伝広告の影響
- 直接コミュニケーション（他人からの薦めによる影響）
- 間接コミュニケーション（噂話やネット検索で見かけること等からの影響）

これらについて、意欲の時間的な変化を追う微分方程式を立てることで数理モデル化する。広告投入量を $A(t)$ 、広告の強さを C とする。人間の意欲が定量化できるものであると仮定し、社会全体

の意欲を $I(t)$ とする。直接コミュニケーションの強さを D 、間接コミュニケーションの強さを P とする。社会の中の 1 人 1 人の意欲の方程式は以下のようなになる。

$$\frac{dI_i(t)}{dt} = CA(t) + \sum_{j \neq i}^N D_{ij} I_j(t) + \sum_j \sum_k P_{ijk} I_j(t) I_k(t)$$

これを平均場近似で簡単化し、 N 人全ての消費者が全く同じ動きをするように社会を単純化すると、方程式は以下のようなになる。

$$\frac{dI(t)}{dt} = CA(t) + DI(t) + PI^2(t)$$

右辺第 1 項は単純なマーケティング理論に対応する形となる。右辺第 2 項は直接コミュニケーションが他者の意欲に比例することを表し、右辺第 3 項は間接コミュニケーションが 3 体相互作用することを表す。

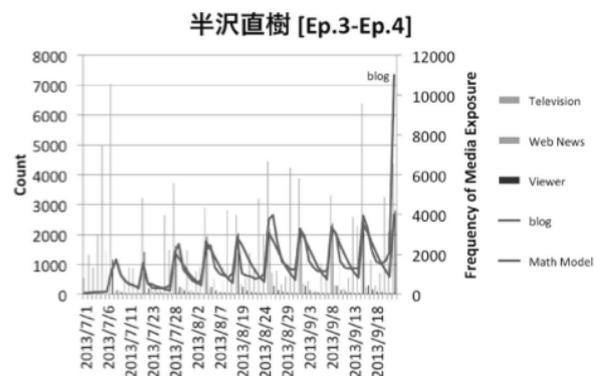


図 3: ヒット現象の数理モデルでドラマ「半沢直樹」を解析した結果 [2]

2.2 計算方法

ドラマのヒット予測を行う場合の計算方法を説明する。まず広告投入量 $A(t)$ として、日ごとのテレビ露出時間のデータを入力する。その結果が、日ごとの SNS（ブログや Twitter）投稿数と一致

するまで、 C, D, P の値を乱数で探索する。この C, D, P の探索は 4 話までのデータをもとに行い、探索結果をもとに 5 話以降の SNS 投稿数を予測する。

図 3 は 2013 年夏放送の「半沢直樹」での例で、予測結果と実測値がよく合っていることが分かる。

3 研究内容

3.1 本研究の位置づけ

本研究においても計算方法は先行研究と同様であるが、入力及び結果予測に用いるデータが異なる。広告投入量 $A(t)$ として、「ドラマの Twitter 公式アカウントが投稿したツイートがリツイートされた数」のデータを用いる。ドラマの公式アカウントが行った宣伝が Twitter 上でどれだけの人に目に触れられたのかを表している。結果の予測には「ドラマのタイトルを含んだツイートが投稿された数」のデータと、視聴率のデータを用いる。ドラマがどれだけ反響を呼ぶのかを Twitter・テレビの 2 つの媒体でそれぞれ予測し比較する。 C, D, P の各係数が持つ意味は次のようになっている。

- C : 宣伝ツイートの強さ
- D : 宣伝ツイートがリツイートで回ってきたのを見たときの強さ
- P : リツイート以外で宣伝ツイートを目にするときの強さ

3.2 対象

2017 年 7~9 月にかけて放送されたドラマ 6 作品を対象とした。以下に一覧を示す。

- 愛してたって、秘密はある (日本テレビ, 日 22:30~)
- ごめん、愛してる (TBS, 日 21:00~)
- 過保護のカホコ (日本テレビ, 水 22:00~)
- コード・ブルー (フジテレビ, 月 21:00~)
- 僕たちがやりました (フジテレビ, 火 21:00~)
- 黒革の手帖 (テレビ朝日, 木 21:00~)

3.3 データの収集

計算で用いるデータの収集は Ruby 言語の Twitter API を扱うライブラリを使用して行った。あるアカウントに対するリツイートと、あるキーワードに関するツイートを、それぞれ過去に遡りながら取得するようなスクリプトを作成し、CSV ファイルとしてまとめた。Twitter API の仕様上約 10 日以上前のデータは取得できないため、数日に 1 度の間隔でこのスクリプトを実行した。最終的にこの CSV ファイルを作品ごとにまとめて日ごとのリツイート数・ツイート数の表を作成した。対象のドラマ作品はそれぞれ Twitter 公式アカウントを開設しており、そのアカウントから投稿されるツイートを宣伝ツイートとみなしてそれらがリツイートされた数を集計した。また、ドラマのタイトルまたはその略称¹がハッシュタグ²として含まれているツイートを視聴者の反響とみなして、その数を集計した。視聴率のデータは「Audience Rating TV」³という、ビデオリサーチ社が提供する関東地区の連続ドラマ視聴率を一般人がまとめたウェブサイトから取得した。

3.4 ヒット予測の計算

MATLAB を使用して数理モデルのパラメータ決定とモデルの解の挙動のシミュレーションを行った。数理モデルのパラメータは方程式の初期解 $I(t=0)$ 及び C, D, P であり、これらをドラマの放送開始から 4 週目までの実データを教師データとしたフィッティングを行い、決定した。フィッティングにはモンテカルロ法を用い、4 週目までのデータとの 2 乗誤差の総和を評価関数として与えた。フィッティングによって得られたパラメータを用いて 4 週目以降のシミュレーションを行った。

¹ドラマタイトルの略称は、公式 Twitter アカウントによって提供されたもののみを用いる。

²ツイート文の中に付加することで検索性を高めたり話題性を容易に確認できるようにする機能のこと。

³<https://artv.info/>

4 結果

4.1 視聴率の予測

一日あたりの宣伝ツイートのリツイート数から各日のツイート数を予測した結果を以下に示す。図中の横軸はドラマ1話放送日7日前からの日数、縦軸は視聴率である。赤色で示される点は実際の視聴率であり、緑色の線が数理モデルの出力結果である。

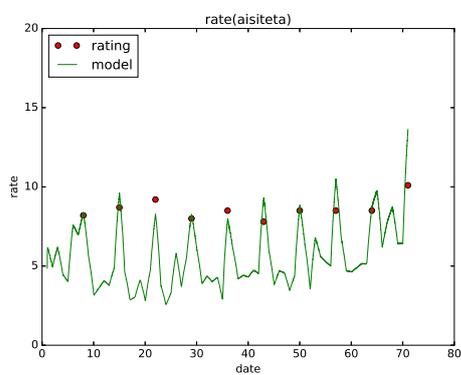


図 4: 「愛してたって、秘密はある」の視聴率予測結果

4週目以降も実際の視聴率よりも高く見積もっているものの、全体としてはある程度正確な予測ができています。

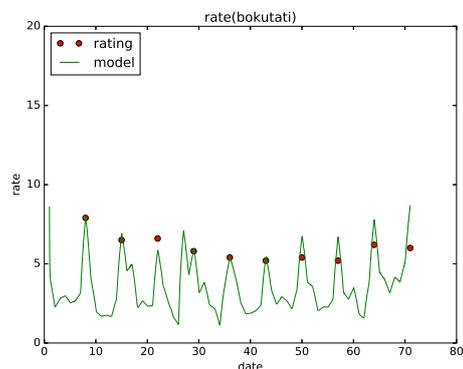


図 5: 「僕達がやりました」の視聴率予測結果

5週目、6週目は正確な予測となった。後半に宣伝ツイートのリツイート数が伸びたためか、実際の数値より高く見積もっている。

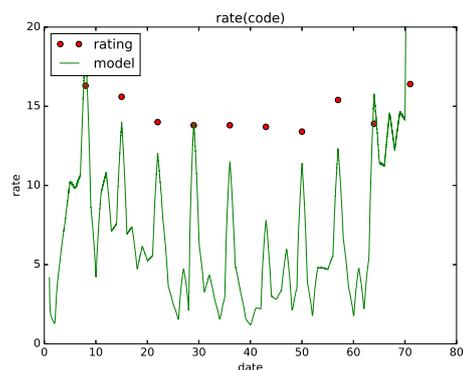


図 6: 「コード・ブルー」の視聴率予測結果

図 4, 図 5 とは異なり、5週目以降の視聴率が低く見積もられている。また、最終回の視聴率は実際の数字より非常に大きく異なっている。最終話の放映日に公式 twitter アカウント上で映画化の告知がなされ、その告知が広く拡散されたためこのような結果が得られたと考えられる。

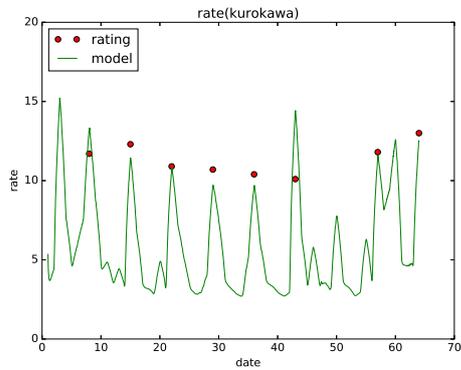


図 7: 「黒革の手帖」の視聴率予測結果

正確な予測ができています。宣伝ツイートのリツイート数が比較的少なかったため、最終話の視聴率まで正確な予測となった。

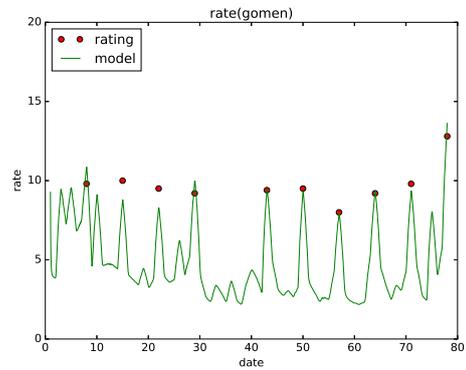


図 9: 「ごめん、愛してる」の視聴率予測結果

全体的に非常に正確な予測ができています。最終話の視聴率も正確に予想できている。

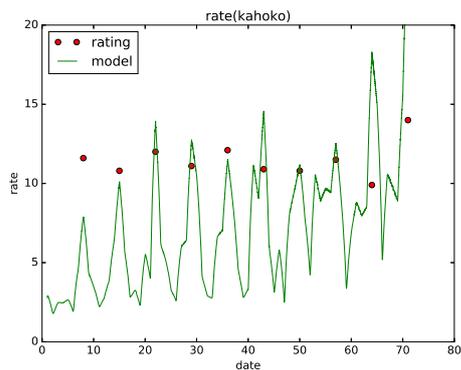


図 8: 「過保護のカホコ」の視聴率予測結果

数理モデルの出力結果の変動が激しく、9話、10話に関しては正確な予測結果は得られなかった。しかし、8話までは実際の値に近い結果が得られた。

4.2 ツイート数の予測

一日あたりの宣伝ツイートのリツイート数から各日のツイート数を予測した結果を図 10～図 12 に示す。横軸はドラマ 1 話放送日 7 日前からの日数、縦軸はツイート数である。赤色の棒グラフが実際のツイート数、緑色の折れ線グラフが数理モデルによる予測を示す。

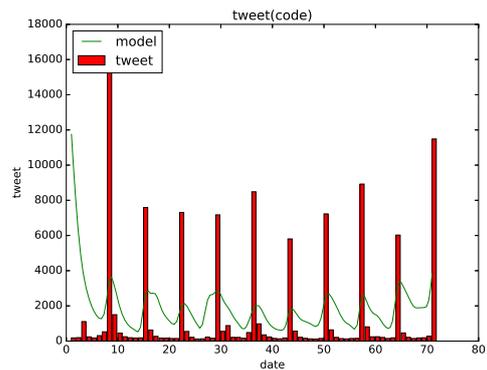


図 10: 「僕達がやりました」のツイート数予測結果

ドラマの放映日にツイート数が大きく伸びるという傾向を反映した結果となった。

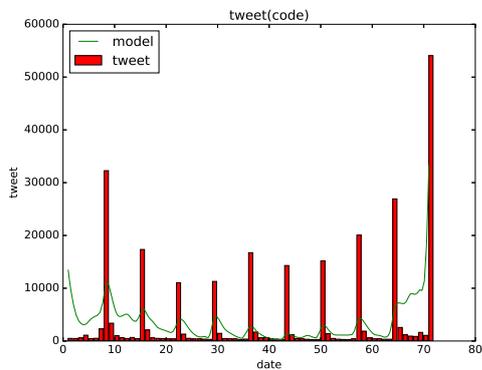


図 11: 「コード・ブルー」のツイート数予測結果
 ドラマの序盤と終盤の反響が大きかったためか、ドラマ中盤のツイート数が非常に小さく予測された。

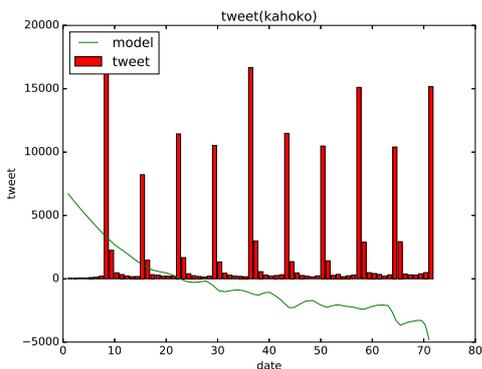


図 12: 「過保護のカホコ」のツイート数予測結果
 パラメータのフィッティングがうまくいかなかったため、不自然な挙動を示している。7日目までのツイート数と、8日目のツイート数に大きく差があるためだと考えられる。

実際のツイート数に対し、モデルの出力は非常に小さくなっている。これは、ドラマ放送日にのみ急激にツイート数が伸びるが、モデルが連続モデルであるため、あまり急激な変化に対応できないことが現れている。

5 まとめ

視聴率の予測に関しては、ドラマによって、高く見積もる・低く見積もるなどの違いはあるものの、ある程度正確に予測できたといえる。最初の話題性が高い場合、その話題性を維持するのが難しく、モデルは視聴率を低く見積もってしまう。逆に、話題性が低く宣伝ツイートのリツイート数が少ないドラマであれば、より正確な予測が可能である。一方、宣伝ツイートのリツイート数から、ツイート数の予測は難しいことが分かった。これは視聴率の変動と比較して、ツイート数の日毎の変動は非常に大きいことが関係している。連続モデルでは急激な変化を再現することが難しいため、日毎の変動が大きいツイート数の再現が難しい。また、本研究での宣伝量は、ドラマ作品の公式アカウントが出演者のアカウントのツイートをリツイートする、という部分は考慮していない。したがって、実際の Twitter での宣伝とは厳密に異なっているためヒット予測にずれが生じたのではないかと考えられる。

参考文献

- [1] 総務省. 平成 28 年通信利用動向調査の結果. pp.847-853, 2017/6/8.
- [2] 石井 晃, 川畑 泰子. ヒット現象の数理モデル. 人工知能 = journal of the Japanese Society for Artificial Intelligence 30(1), pp.97-103, 2015-01-01.