

インターネット上の検索行動の数理モデル

Mathematical model for search action on the Internet

石井 晃^{*1} 芦田 昇^{*1} 川畑泰子^{*2}
Akira Ishii Noboru Ashida Yasuko Kawahata

^{*1}鳥取大学 ^{*2}群馬大学
Tottori University Gunma University

Mathematical model to analyze search action on the Internet is presented on the sense of sociophysics. As an extension of the mathematical model for hit phenomena, in the present model, we assume that the search action of ordinary people in society is affected not only by mass media campaign, direct communications and indirect communications, but also by Twitter and blog posting.

1. はじめに

人々のインターネット上の検索行動は人々がその時その時に何に興味あるかをよく示していると考えられる。Blog, Twitter, Facebook などへの書込も人々の興味関心を反映させているが、書込まではせずに検索で調べるだけという人も少なくなく、そうした人々の行動は検索行動の方で反映される。Blog, Twitter などへ書き込む人々を自らの意見を顕在化させる「顕在層」とすれば、そうした書込をせずに検索をする人々は潜在化している「潜在層」である。本研究はそうした潜在層の興味関心を調べる、人々の検索行動について動力的な数理モデルを提案する。

Blog, Twitter, Facebook などのソーシャルメディアの普及が進み、インターネット上には様々な話題について書き込まれている。

そうした書込の解析にはヒット現象の数理モデルが有用である [1, 2]。過去の研究として、映画の興行収入の予測や AKB 選抜総選挙の順位予想などエンターテインメントの分野への応用に成功している [3]。従来は人々の興味関心の指標として Blog, Twitter を用いる数理モデルに対し、別の指標でも応用できないかと考え、本研究では検索数を人々の興味・関心の指標として計算する数理モデルを提案する。先行研究でも検索数を人々の興味関心の指標とし、SIR モデルを用いて情報の拡散過程を説明するモデルを構築していたが、本研究で提案するモデルは検索行動の動力学を記述するモデルである [4]。それにより、今までからは読み取れなかった潜在層の人々の動きや、及ぼされる外的影響から新たな知見が得られると考えられる。

本研究で、人々の検索行動は Google Trend を用いて調べた。さらに特定の話題ごとの Blog や Twitter の書込はホットリンク社のクチコミ@係長で調べ、その話題のテレビやネットへの露出はエムデータ社調べを用いた。

2. 数理モデル

本研究はヒット現象の数理モデルにならう形で、社会の人々による検索行動の動力学を記述する新しい数理モデルを構築することを目的とする。その数理モデルでは人々の検索行動に、Blog や Twitter も影響を与えたと考え、その影響とマスメディアの影響の比較も行う。本研究で特に注目していることは以下である。

1. 計算が実測の GoogleTrends を説明できるか、その精度を確認、条件を変えて比較する。
2. 従来のヒット現象の数理モデルと各パラメータを比較、考察する。
3. 新たに外力項に追加した Blog, Twitter のパラメータがどのように影響しているのかをみる。

なお、GoogleTrends で用いている検索数のデータは 1 から 100 に規格化されたものであり、以下、ここでは関心度と呼ぶことにする。(Google はこれを人気度と呼んでいる)

2.1 ヒット現象の数理モデル

ヒット現象の数理モデルでは、社会における人々の 1 人 1 人が抱く興味・関心を定量化する。ある人「i さん」が抱く興味・関心を $I_i(t)$ と定義し、興味・関心 ($I_i(t)$) を掻き立てる要員として、

1. メディアによる影響
2. 会話による影響
3. 噂による影響

の 3 つがあると考え、面と向かった会話やインターネットを用いた直接的なやりとりによる影響のことを「直接コミュニケーション」、街中の噂やソーシャルメディア上でのやり取りなどに影響を受けることを「間接コミュニケーション」と呼ぶ。それらについて興味・関心の時間的変化を追う微分方程式を立て、数理モデル化すると、以下の形になる [1]。

$$\frac{dI_i(t)}{dt} = \sum_{\xi} c_{\xi} A_{\xi}(t) + \sum_{j \neq i}^N D_{ij} I_j(t) + \sum_j \sum_k P_{ijk} I_j(t) I_k(t) \quad (1)$$

ここで D_{ij} は直接コミュニケーションの強さを表す係数、 P_{ijk} は間接コミュニケーションの強さを表す係数である。広告宣伝の影響は外力と考える。ある話題に関する日毎のテレビでの露出秒数や、ネットニュースの件数を A_{ξ} とし、その係数を c_{ξ} とする。 ξ はメディアの種類を表す添え字である。

式 (1) は個々の視点に基づいた式であるが、簡単化のために平均場近似を行う。社会全体の構成員の数を N 人とし、社会

全体で平均化された人々の意欲・関心を $I(t)$ として、以下で定義する。

$$I(t) = \frac{1}{N} \sum_i I_i(t) \quad (2)$$

この平均場近似を用いると、 $I(t)$ に従う方程式は次のようになる [1]。

$$\frac{dI(t)}{dt} = \sum_{\xi} c_{\xi} A_{\xi}(t) + DI(t) + PI(t)^2 \quad (3)$$

実際に計算する際には左辺の微分の箇所を、

$$\frac{\Delta I(t)}{\Delta t} = \sum_{\xi} c_{\xi} A_{\xi}(t) + DI(t) + PI(t)^2 \quad (4)$$

と表す。本研究では 1 日 1 日の動向をこのヒット現象の数理解モデルで探るので、 $\Delta t = 1[\text{日}]$ として、

$$\Delta I(t) = \sum_{\xi} c_{\xi} A_{\xi}(t) + DI(t) + PI(t)^2 \quad (5)$$

となる。

マスメディアの影響など広く人々に影響する外部からの影響 $A_{\xi}(t)$ の時間的減衰が指数関数的に減少するとする。分析日を $t = t_0$ とし、式 3 の $A(t)$ を $\sum_{i=0}^N \sum_{\xi} c_{\xi i} A_{\xi i}(t) \exp \alpha_{\xi}(t_i - t_0)$ (α_{ξ} は減衰のパラメータ) と書き換えると、

$$\frac{dI(t)}{dt} = \sum_{i=0}^N \sum_{\xi} c_{\xi i} A_{\xi i}(t) \exp \alpha_{\xi}(t_i - t_0) + DI(t) + PI(t)^2 \quad (6)$$

となる。N 日前の $t = t_N$ と表せる。

2.2 検索行動の数理解モデル

本研究ではヒット現象の数理解モデルの考え方で検索行動の意欲について数理解モデルを立て、さらにブログや Twitter、さらには Wikipedia での当該話題の検索数も考えて次のように数理解モデルをたてる。

$$\begin{aligned} \frac{dI(t)}{dt} = & DI(t) + PI(t)^2 \\ & + \sum_{i=0}^N c_{tv_i} A_{tv_i}(t) \exp(t_i - t_0) \alpha_{tv} \\ & + \sum_{i=0}^N c_{news_i} A_{news_i}(t) \exp(t_i - t_0) \alpha_{news} \\ & + \sum_{i=0}^N c_{blog_i} A_{blog_i}(t) \exp(t_i - t_0) \alpha_{blog} \\ & + \sum_{i=0}^N c_{twitter_i} A_{twitter_i}(t) \exp(t_i - t_0) \alpha_{twitter} \\ & + \sum_{i=0}^N c_{wikipedia_i} A_{wikipedia_i}(t) \exp(t_i - t_0) \alpha_{wikipedia} \end{aligned} \quad (7)$$

このモデルでは人々の興味関心の指標である $I(t)$ を Google-Trends の関心度とし、従来のヒット現象のモデルの広告宣伝

の項には Tv 、ネットニュースの影響しか加味していなかったが、第 5 項の Blog の影響、第 6 項の Twitter の影響、第 7 項に Wikipedia 検索の影響を追加した。実際の計算では話題ごとに Wikipedia 検索は入れる場合と入れない場合がある。

本研究ではこの数理解モデルで分析を進める。

2.3 精度の計算方法

また、係数をフィッティングの際に精度を表す指標として、2 本の曲線の一致度を測る指標である R_factor を用いる [?, ?]。R_factor は物理学の電子線回折で実験と理論計算の曲線の類似度を測る指標として 30 年以上用いられている確立された技術である。R_factor は以下の式である。

$$R = \frac{\sum_i (f(i) - g(i))^2}{\sum_i ((f(i))^2 + (g(i))^2)} \quad (8)$$

ここで、例えば本研究では $f(i)$ はターゲット (映画など) に対する関心度の件数、 $g(i)$ は検索行動の数理解モデルによるシミュレーション結果の値を用いる。

R は、 $0 \leq R \leq 1$ の範囲で、値が小さいほどフィッティングの精度が良い。

3. 計算結果

3.1 「君の名は。」計算結果

図 1 に 2017 年に公開された映画「君の名は。」の計算結果を示す。分析期間は公開前から公開後 2ヶ月の全体のものである。C,D,P などのパラメータは Google Trend の測定データに合うように、つまり乱数によって変更したパラメータを用いた計算毎に R-factor を計算し、10 万回以上試行して最も R-factor が低い計算結果を採用したものである。図 1 でわかるように最終的な計算結果は Google Trend の測定データをきれいに説明する。

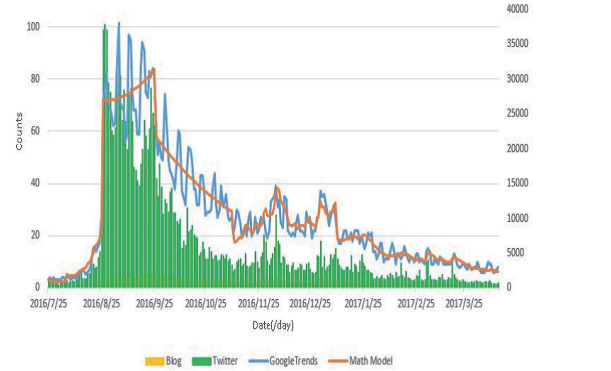


図 1: 君の名は。計算結果。: 青線が関心度、赤線が数理解モデルによる再現、緑の impulse が Twitter の件数、橙の impulse が Blog の件数である。横軸は 1 日おきの日付、縦軸は数値であり、左側に関心度のスケール、右側を Blog, Twitter の書き込み件数のスケールで表している。

次に 2016 年に話題になったピコ太郎による PPAP (Pen Pineapple Apple Pen) の動画についての計算を示す。この動画が大人気になったのは 2016 年 9 月であり、ここでは盛り上がりの初期からほぼ終息している 2016 年 9 月 1 日から 2017 年 4 月 30 日を分析期間とした。

次にパラメータを比較していく．以下に毎月の直接コミュニケーション D 、間接コミュニケーション P の比較グラフを図 2、図 3 に、TV の影響 $Cadv_t$ とネットニュースの影響 $Cadv_n$ の比較グラフを図 4、図 5 に、Blog の影響 $Cadv_{Blog}$ と Twitter の影響 $Cadv_{twitter}$ の比較グラフを図 6、図 7 に示す．

4. 考察

従来のヒット現象の数理モデルでの D, P を顕在層の人々のみのパラメータ、本研究でのモデルの D, P を潜在層の人々のパラメータが大量に含まれていると仮定して考察を進める．それは Blog や Twitter が自分の考えを外に向けて発信していく人であるのに対し、GoogleTrends による検索行動には発信をしない潜在層の人々の関心も多く含まれているからである．

ピコ太郎のヒットについてのヒット現象の数理モデルによる分析は先行研究 [5] があり、この解析によるとピコ太郎のヒットを 2016 年 9 月 - 10 月で分析すると、ブレイク時に間接コミュニケーションの強さ P は大きな値になり、一方で直接コミュニケーションの強さ D はブレイク後に減っている．

今回の数理モデルで間接コミュニケーションの強さ P は分析区間全部にあたって一定の強さを示している一方、直接コミュニケーションの強さ D はブレイク時と年末年始前後、そして翌年春に大きくなっている．これはブレイク時に特に強い支持があったと言える．なお、年末年始に D が大きいのは紅白歌合戦にピコ太郎が出たことによると思われる．TV やネットニュースの影響はブレイク時に限られ、逆に Blog や Twitter の影響はブレイク後に高まっていることがわかる．

他に季節性のある話題についてこの検索行動の数理モデルを用いた研究もあり [6]、それも顕在層と潜在層の関心の動きの違いなどが示されているので、本論文で提示した検索行動の数理モデルは有用であると思われる．

5. 結論

本研究では人々の検索行動のダイナミクスを数理モデルとして提出した．従来の Blog や Twitter への書込行動を計算するヒット現象の数理モデルと比べ、ピコ太郎の動画の計算で異なる傾向が得られ、書き込む顕在層と検索だけをする潜在層で関心を持つタイミングが異なることがわかった．検索行動の解析や予測にこの数理モデルを用いるのはマーケティング等に有効であり、またバースト現象の解析にも有効であろうと思われる．

参考文献

- [1] Akira Ishii, Hisashi Arakaki, Naoya Matsuda, Sanae Umemura, Tamiko Urushidani, Naoya Yamagata and Narihiko Yoshida. "The 'hit' phenomenon: a mathematical model of human dynamics interactions as a stochastic process". New Journal of Physics 14(2012)
- [2] Ishii A, "Analysis and Predictions of Social Phenomena via social media using Social Physics method", INFORMATION-An International Interdisciplinary Journal 20(10) 7073-7086 (2017)
- [3] Ishii A, Ota S, Koguchi H and Uchiyama K, the proceedings of the 2013 International Conference on Biometrics and Kansei Engineering(ICBAKE2013) 143-147 DOI 10.1109/978-0-7695-5019-0/13 (2013)

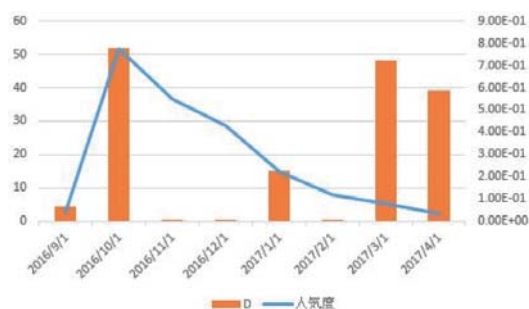


図 2: ピコ太郎における本研究の数理モデルの直接コミュニケーション D 比較．横軸はそれぞれの分析区間、縦軸はパラメータの数値．分析区間は 2016 年 9 月 1 日から 2017 年 4 月 30 日．

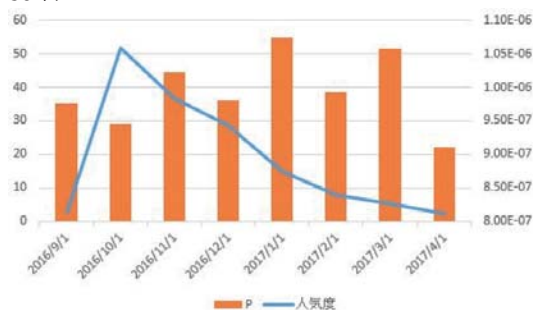


図 3: ピコ太郎における本研究のモデルの間接コミュニケーション P 比較．横軸、縦軸、分析区間は分析区間は 2016 年 9 月 1 日から 2017 年 4 月 30 日．

- [4] 大知正直, 長濱憲, 榊剛史, 森純一郎, 坂田一郎『口コミ指数による事例類型化に基づく複数メディアのヒット前の露出を先行指標とした情報拡散過程の分析』(2016)
- [5] Ishii A, Fukui K, Oda T, Miki A, Fujiwara S, Okano N and Kawahata Y, "Analysis of social epidemic phenomena on Social Media using social physics approach" Adv. Sci. Lett. in press
- [6] 少林俊道, 岡野のぞみ, 川畑泰子, 石井晃『ソーシャルメディア上の季節性のある書込の解析』第 2 回計算社会科学ワークショップ (2018)

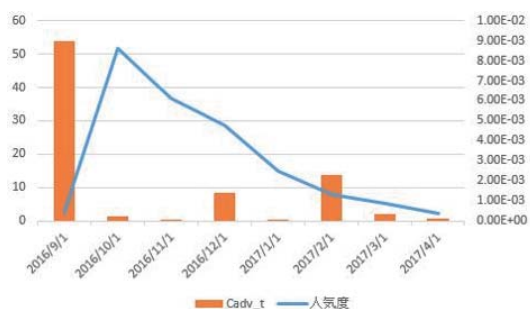


図 4: ピコ太郎における本研究のモデルのテレビの影響 Cadv_t 比較. 横軸, 縦軸, 分析区間は分析区間は 2016 年 9 月 1 日から 2017 年 4 月 30 日.



図 5: ピコ太郎における本研究のモデルのネットニュースの影響 Cadv_n 比較. 横軸, 縦軸, 分析区間は分析区間は 2016 年 9 月 1 日から 2017 年 4 月 30 日.

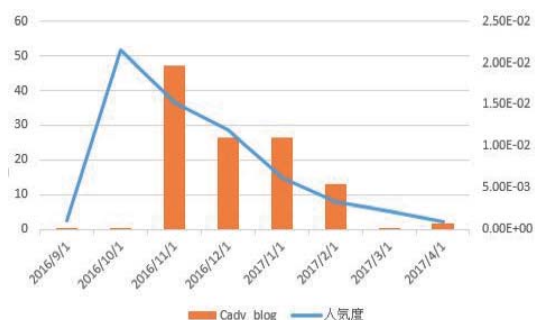


図 6: ピコ太郎における本研究のモデルの Blog の影響 Cadv_blog 比較. 横軸, 縦軸, 分析区間は分析区間は 2016 年 9 月 1 日から 2017 年 4 月 30 日.

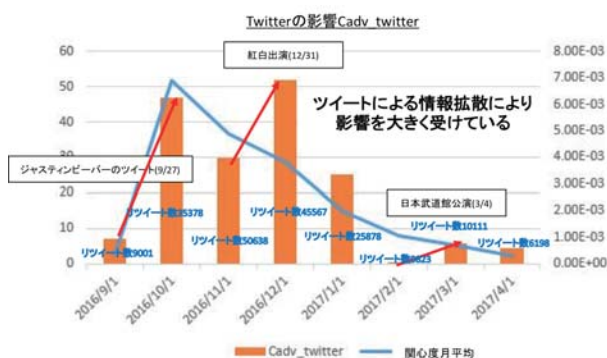


図 7: ピコ太郎における本研究のモデルの Twitter の影響 Cadv_twitter 比較. 横軸, 縦軸, 分析区間は分析区間は 2016 年 9 月 1 日から 2017 年 4 月 30 日.