

ネット上の検索行動の数理モデル

○石井晃 芦田昇 少林俊道 岡野のぞみ (鳥取大学), 川畑泰子 (群馬大学)

Mathematical model for search actions on the Internet

*A Ishii, N Ashida (Tottori University), Y Kawahata (Gunma University)

概要— インターネット上の人々の関心を表す検索行動を, 社会物理学の立場で理論として数理モデル化した. この理論を用いて Google Trend から検索行動を解析すると, マスメディア広告などの影響の他, ブログや Twitter の影響も定量的に評価することができる.

キーワード: Google Trend, 数理モデル, ヒット現象の数理モデル

1 序論

インターネット上の検索行動は, 往々にして興味ある事柄, 特に新たに目に入った新しい事柄で興味を持つ事を検索すると考えら得るので, ネット上の検索行動はその人の興味関心の移り変わりを表すと考えられる. Google Trend によるサイトでの検索のデータは, 世の中の人々の多くが何に興味を持って検索するかを表すデータになっていると考えられる. ブログや Twitter などソーシャルメディアの書込も社会の人々の興味関心を表す指標になるが, 世の中には興味関心があっても, それをすぐにソーシャルメディアには書かない人も少なくない. 大きな社会的事件があったとして, それについてソーシャルメディアに書く人がいる一方で, 関心を持ちつつも何も書かない人が多い. そうした書かない人々も, 関心がある話題についてインターネット上で検索するであろう. 従って, 人々の検索行動を解析することは, ソーシャルメディアへの書込とはまた違った意味で, 世の中の人々の興味関心を解析することになるはずである.

本研究ではそうした人々の検索行動をヒット現象の数理モデル^{1, 2)}にならう形で社会物理学の考え方で数理モデルにまとめ, いくつかの実例で計算して, Google Trend で実測される人々の検索数の時系列的な変化を我々の数理モデルがいかに記述するかを調べる. 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が構築されているが³⁾, 本研究で提案するモデルはより具体的なマスメディア露出やブログ, Twitter の影響を含めての人々の検索行動の動力学を記述するモデルである.. 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.

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 も影響を与えよう. 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 Google Trends 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, Google Trends 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・関心を定量化して方程式にしている. 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 つがあるとする. 会話や LINE などによる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」とよび, 街中の噂やソーシャルメディア上でのやり取りなどで影響を受けることを「間接コミュニケーション」と呼ぶことにする. 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, 以下の形で表せると仮定できる^{1, 2)}.

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

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, このままの形では分析を行うのは難しい. そこで, 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。

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

導出の詳細は^{1), ?)}を参照。

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。つまり実際にはこの $\Delta I(t)$ を計算している。

これまでのヒット現象の数理モデルでも話題の新鮮さの減衰は考えられていた⁴⁾。ここではさらに広告宣伝項の影響の減衰が指数関数的に減少すると推測し、その項を追加しよう。分析日を $t = t_0$ として、

式3の $A(t)$ を $\sum_{i=0}^N \sum_{\xi} c_{\xi_i} A_{\xi_i}(t) \exp(t_i - t_0)(\alpha_{\xi})$ (α_{ξ} は減衰のパラメータ) と書き換えると、

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

となる。N 日前の $t = t_N$ と表せる。

2.2 検索行動の数理モデル

本研究ではヒット現象の数理モデルを検索行動について立て、さらにブログや Twitter, さらには Wikipedia での当該の話題の検索数も考えて次のように数理モデルをたてる。

$$\begin{aligned} \frac{dI(t)}{dt} = & DI(t) + PI(t)^2 \\ & + \sum_{i=0}^N c_{tvi} A_{tvi}(t) \exp(t_i - t_0) \alpha_{tv} \\ & + \sum_{i=0}^N c_{newsi} A_{newsi}(t) \exp(t_i - t_0) \alpha_{news} \\ & + \sum_{i=0}^N c_{blogi} A_{blogi}(t) \exp(t_i - t_0) \alpha_{blog} \\ & + \sum_{i=0}^N c_{twitteri} A_{twitteri}(t) \exp(t_i - t_0) \alpha_{twitter} \\ & + \sum_{i=0}^N c_{wikipediai} A_{wikipediai}(t) \exp(t_i - t_0) \alpha_{wikipedia} \end{aligned} \quad (7)$$

このモデルでは人々の興味関心の指標である $I(t)$ を GoogleTrends 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。

本研究ではこのヒット現象の数理モデルを元に分析を進める。

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^{5, 6)}。本研究では以下の式を用いる。

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

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と言える。

3 解析結果

本研究で提案した理論を実際の事例に応用してみる。本研究は以下の流れで分析を行っている。

1. データの取得

Blog, Twitter, 5ch(2ch), ネットニュース書き込み件数 (クチコミ@係長から取得 ホットリンク社より提供)

Tv 露出秒数 (クチコミ@係長から取得 エムデータ社より提供)

関心度 (GoogleTrends から取得)

Wikipedia による検索数は Pageviews Analysis を用いて取得した。

2. 日毎の関心度件数の推移を数理モデルで再現 (フィッティング)

3. 得られたパラメータの関係性を探る

分析区間を区切り、区間ごとにフィッティングを行いそれぞれでパラメータを算出する。本研究では2016年に大ヒットした映画の「君の名は。」と、花見、紅葉や水着などの季節性のある話題に焦点を当てて分析する。映画「君の名は。」の分析区間の区切り方を以下、表1に示す。分析区間は公開前1か月から公開終了日までとしている。

(公開日: 2016/8/26)

Table 1: 「君の名は。」フィッティングの区間の区切り方

period1	2016/7/25~8/25
period2	8/26~9/25
period3	9/26~11/9
period4	11/10~2017/1/10
period5	1/11~4/14

また、季節性の話題の解析では、分析期間は1年間で、それを1ヶ月ごとに数理モデルで解析して、毎月の各パラメータの値の時間的推移を見る。

3.1 「君の名は.」計算結果

図1に分析対象「君の名は.」のフィッティンググラフを示す。分析期間は period1 から period5 までの全体的ものである。

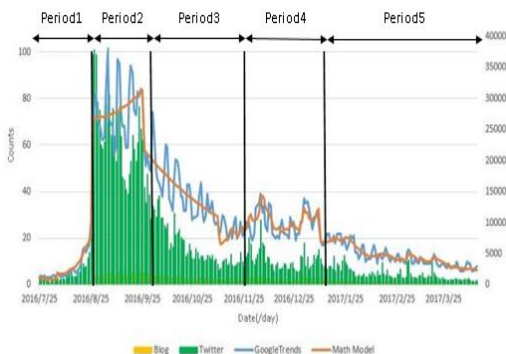


Fig. 1: 君の名は. フィッティング結果. : 青線が関心度, 赤線が検索行動の数理モデルによる再現, 緑の impulse が Twitter の件数, 橙の impulse が Blog の件数である. 横軸は1日おきの日付, 縦軸は数値であり, 左側を関心度のスケール, 右側を Blog, Twitter の書き込み件数のスケールで表している.

次にパラメータを比較していく。以下に period1 から5までのそれぞれ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, Tv の影響 $C_{adv,t}$ とネットニュースの影響 $C_{adv,n}$ の比較グラフを図4, 図5, Blog の影響 $C_{adv,Blog}$ と Twitter の影響 $C_{adv,twitter}$ の比較グラフを図6, 図7に示す。

さらに, Twitter の書き込み件数を人々の興味・関心の指標とした従来型数理モデルでフィッティングしたパラメータ D, P の比較グラフを図8, 図9に示す。

検索行動の数理モデルでの D が period4 以降減少しており, ブログ投稿量の時系列的推移を計算で再現するヒット現象の数理モデルの D は公開後, 一貫して高い値をとっている。ブログや Twitter などは話題を書きたい人々, 顕在層の関心を示していて, Google Trend はソーシャルメディアに書き込むことはしないが関心を持っているという潜在層の関心を表していると考えられることことから, 顕在層の人々間での話題にはなっている一方で, period4 以降の潜在層の人々が顕在層に移ったのではないかと考察できる。これは period4 以降での Twitter の影響が増大していることから示唆できる。また, 検索行動の P が公開後高い値をとっていることからちょっとした会話にでていると言えるだろう。

次に, 公開前の Tv, ネットニュースの影響が強まっており, 検索行動に結びつく人々の興味関心は Tv, ネットニュースから影響を受けていることがわかる。この時点で Blog や Twitter の書込は多くないので, 潜在層が Tv やネットニュースから情報を得ることから関心を得ていると言えるだろう。Blog, Twitter の影響は徐々に高まっていて, 特に Twitter の影響は徐々に上昇しており, 先程の D の考察から潜在層を引き込むことになったのが長期的に人気となった一因であるだろうと考察できる。

また, 今回の検索行動の数理モデルによる直接・間接

コミュニケーションの値, 図2・図3と従来から使われているヒット現象の数理モデルによって Blog 投稿数を計算した時の直接・間接コミュニケーションの値, 図8・図9を比較すると, 明らかに両者は異なっており, 今回の数理モデルで Blog や Twitter が関心に影響を与えるとしたメカニズムは少なくない効果であると言える。

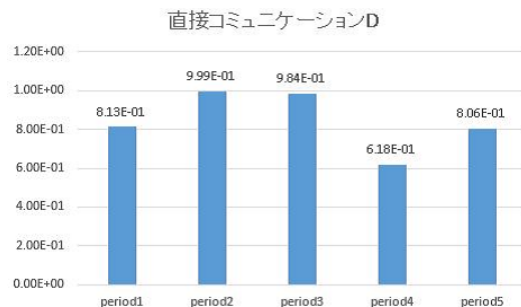


Fig. 2: 「君の名は.」における検索行動のモデルの直接コミュニケーション D 比較. 横軸はそれぞれの分析区間, 縦軸はパラメータの数値. 分析区間に関しては表1参照.

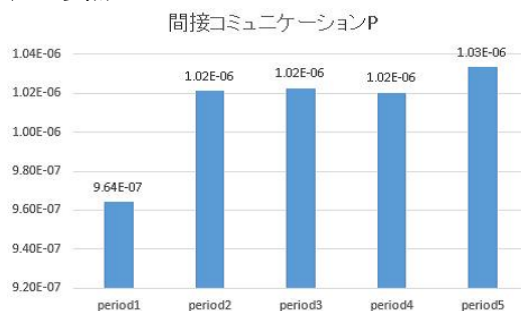


Fig. 3: 「君の名は.」における検索行動のモデルの間接コミュニケーション P 比較. 横軸, 縦軸, 分析区間は図2と同じ.

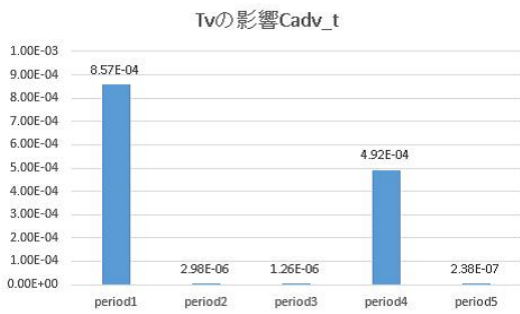


Fig. 4: 「君の名は.」における検索行動のモデルのテレビの影響 Cadv_t 比較. 横軸, 縦軸, 分析区間は図 2 と同じ.

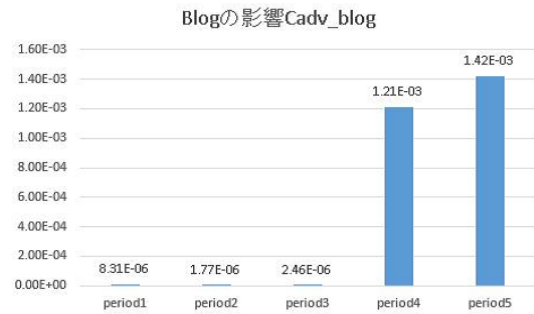


Fig. 6: 「君の名は.」における検索行動のモデルの Blog の影響 Cadv_blog 比較. 横軸, 縦軸, 分析区間は図 2 と同じ.

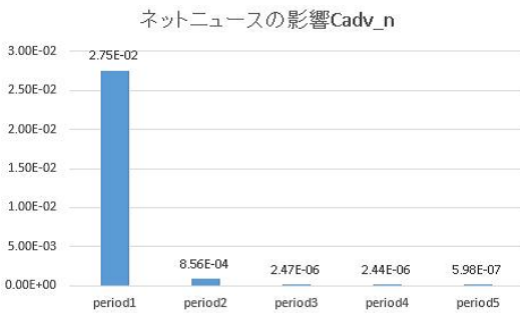


Fig. 5: 「君の名は.」における検索行動のモデルのネットニュースの影響 Cadv_n 比較. 横軸, 縦軸, 分析区間は図 2 と同じ.

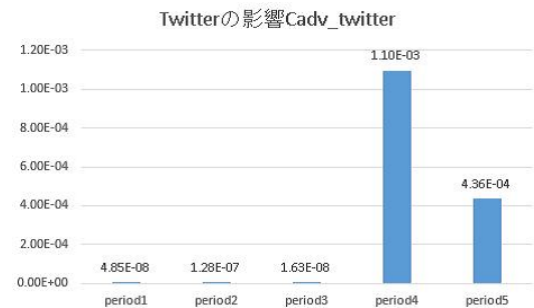


Fig. 7: 「君の名は.」における検索行動のモデルの Twitter の影響 Cadv_twitter 比較. 横軸, 縦軸, 分析区間は図 2 と同じ.

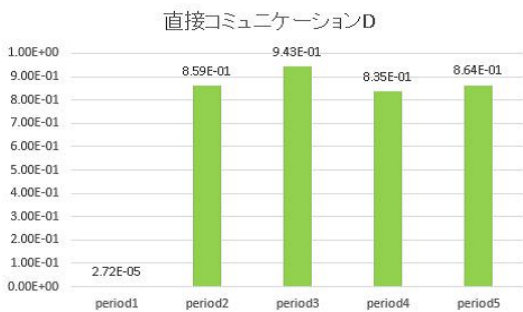


Fig. 8: 「君の名は。」におけるヒット現象の数理モデルの直接コミュニケーションDの比較。横軸，縦軸，分析区間は図2と同じ。

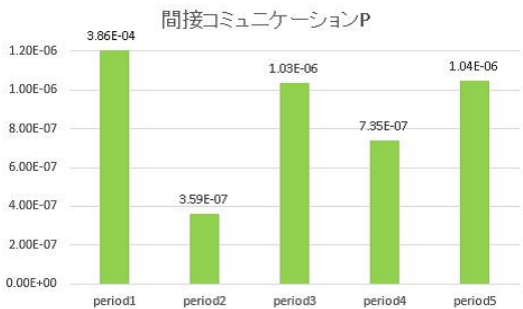


Fig. 9: 「君の名は。」におけるヒット現象の数理モデルの間接コミュニケーションPの比較。横軸，縦軸，分析区間は図2と同じ。

3.2 花見と紅葉の解析結果

季節性のある話題としては，冬にやるスキー，夏にやる水泳やそこで着る水着，さらにお花見や紅葉，クリスマス，バレンタインデーなどが考えられる。まず，ここでは春の花見と秋の紅葉をとりあげる。いずれも話題となる時期は春と秋に限定されているが，調べる区間は2017年1年間で，毎日のデータを取得して1ヶ月ごとに数理モデルで解析した結果である。

以下直接コミュニケーションDを図2，間接コミュニケーションPを図3，Tvの影響Cadv_tを図12，ネットニュースの影響Cadv_nを図13，Blogの影響Cadv_Blogを図14，Twitterの影響Cadv_twitterを図15，Wikipediaの影響Cadv_wikipediaを図??に示す。

計算結果でまず目立つのは，花見の直接コミュニケーションが事前で高いことである。これは特に桜の花の満開時期は予測が必要な上に満開時期も短いので，場所ごとにいつ花見に行くかで花見前にいろいろと意見の交換が必要であることを意味しているであろう。紅葉は比較の見頃の時期が長いので，花見ほどの事前の検索行動の高さは見られない。数理モデルでの計算結果から，花見も紅葉も検索のピーク時点以降でテレビやニュースの影響を受けている。これは，各地の見所や時期あるいは具体的な桜の花や紅葉の画像などがテレビやニュースで話題となることで検索意欲を高めていると考えることができる。また，検索数のピークの

前後で花見も紅葉もBlogの値は下がり，Twitterの値は上がっていることがわかる。これはBlogの方がピンポイントに場所が特定できる上，写真付きでの投稿がある場合はリアルに今が見どころかどうかわかるのでそれを参考に検索するからだと考える。一方で検索のピークを過ぎると，検索する立場からするとまだ見れるところはないかと探すが普通ではないだろうか。そのため，Blogよりも拡散性のあるTwitterの影響を受けているのではないかと考えられる。

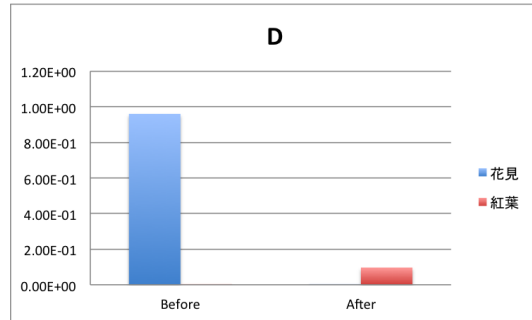


Fig. 10: 花見と紅葉における検索行動の数理モデルの直接コミュニケーションD比較。横軸は日付け，縦軸はパラメータの数値。

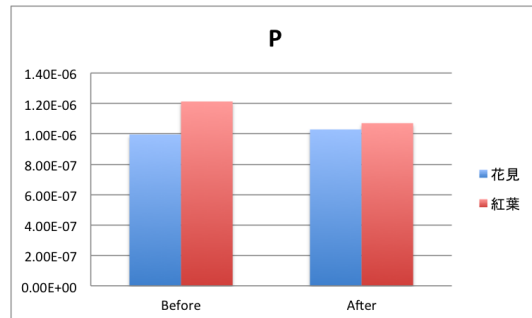


Fig. 11: 花見と紅葉における検索行動の数理モデルの間接コミュニケーションD比較。横軸は日付け，縦軸はパラメータの数値。

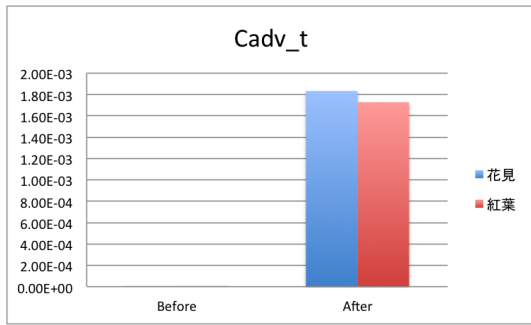


Fig. 12: 花見と紅葉における検索行動の数理モデルのテレビの影響 Cadv_t 比較. 横軸, 縦軸, 分析区間は図 10 と同じ.

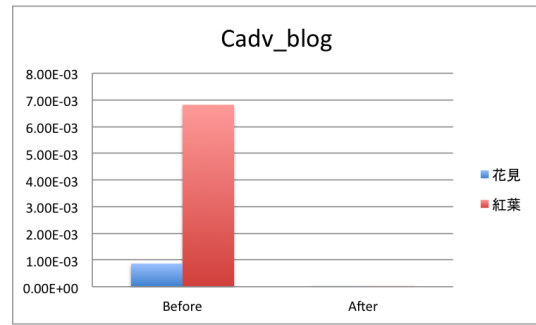


Fig. 14: 「花見と紅葉における検索行動の数理モデルの Blog の影響 Cadv_blog 比較. 横軸, 縦軸, 分析区間は図 10 と同じ.

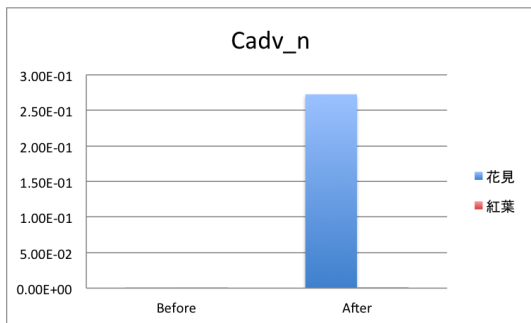


Fig. 13: 「花見と紅葉における検索行動の数理モデルのネットニュースの影響 Cadv_n 比較. 横軸, 縦軸, 分析区間は図 10 と同じ.

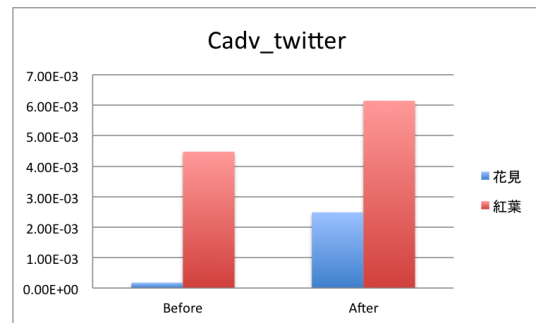


Fig. 15: 花見と紅葉における検索行動の数理モデルの Twitter の影響 Cadv_twitter 比較. 横軸, 縦軸, 分析区間は図 10 と同じ.

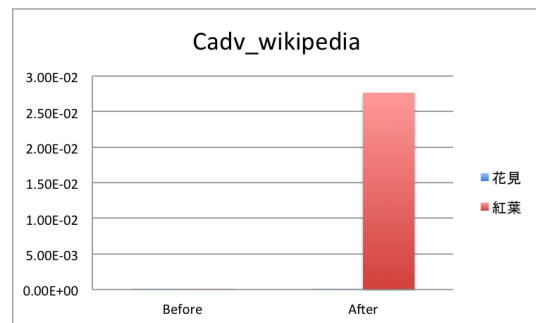


Fig. 16: 「花見と紅葉における検索行動の数理モデルの Wikipedia の影響 Cadv_wikipedia 比較. 横軸, 縦軸, 分析区間は図 10 と同じ.

3.3 水着の解析結果

最後に水着についての解析を示す. 一般に水着というと男女それぞれの水着や競泳用水着, スクール水着など多様であるが, ここでは「bikini」をキーワードとして調べた例なので, 女性のビキニの水着に限定されていて, かつ, 「ビキニ」ではなく「bikini」なので Google

Trend では世界中のビキニ水着に関する検索が集まる。ここでも日本だけでなくアメリカなど北半球の主要な国々を含めたデータである。調べる区間は2017年1年間で、毎日のデータを取得して1ヶ月ごとに数理モデルで解析した結果である。

水着についての直接コミュニケーションDを図17, 間接コミュニケーションPを図18, Tvの影響Cadv.tを図19, ネットニュースの影響Cadv.nを図20, Blogの影響Cadv.Blogを図21, Twitterの影響Cadv.twitterを図22, Wikipediaの影響Cadv.wikipediaを図??に示す。

数理モデルでの計算結果から、水着のシーズンである夏に入る前の5月, 6月にTwitter, Blog値が高くなっている。その年のトレンドなどがSNSによって広まることで検索意欲を高めていると考えられる。これから、ビキニ水着の検索はビーチやプールでビキニの水着の女性を見ることでの関心の高まりではなく、事前に水着を買う女性の購買意欲に表しているのではないかと考察する。女性は毎年水着を買う傾向があるので、今年の夏の水着は何を買おうかなという関心の高まりが検索行動に出ていると考えられる。ネット上の情報の影響Cadv.nはシーズンの7月-9月までの値が高く、ネットで話題となったりニュースになった水着を調べるという検索行動になっているのではないかと考察できる。

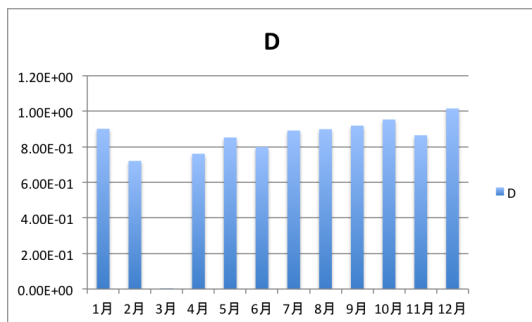


Fig. 17: ビキニ水着についての検索行動の数理モデルの直接コミュニケーションD比較。横軸は日付け, 縦軸はパラメータの数値。

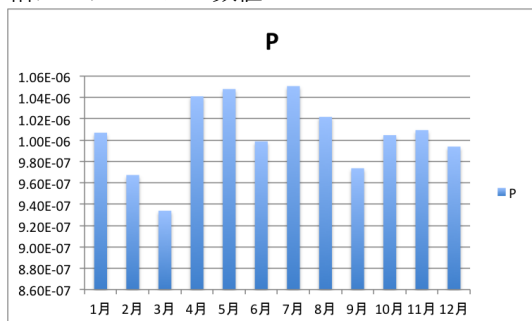


Fig. 18: ビキニ水着についての検索行動の数理モデルの間接コミュニケーションD比較。横軸は日付け, 縦軸はパラメータの数値。

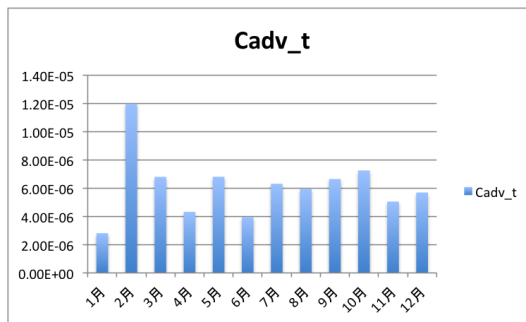


Fig. 19: ビキニ水着についての検索行動の数理モデルのテレビの影響Cadv.t比較。横軸, 縦軸, 分析区間は図17と同じ。

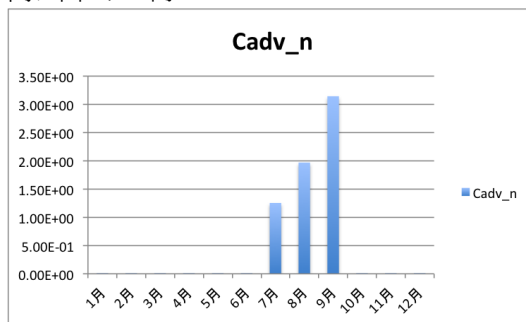


Fig. 20: 「ビキニ水着についての検索行動の数理モデルのネットニュースの影響Cadv.n比較。横軸, 縦軸, 分析区間は図10と同じ。

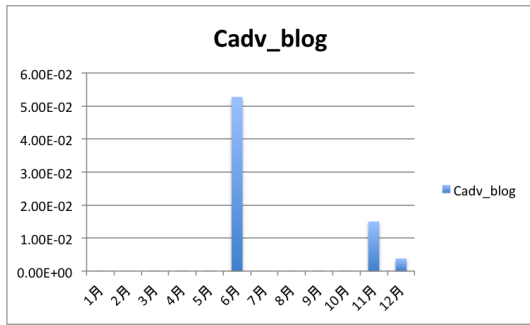


Fig. 21: 「ビキニ水着についての検索行動の数理モデルの Blog の影響 Cadv_blog 比較. 横軸, 縦軸, 分析区間は図??と同じ.

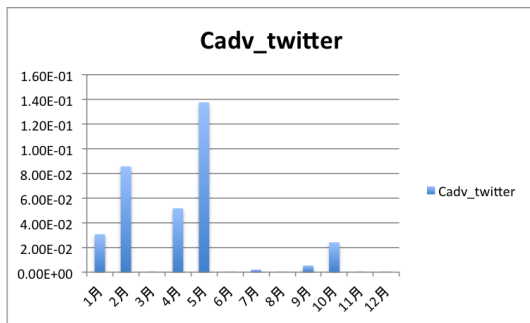


Fig. 22: ビキニ水着についての検索行動の数理モデルの Twitter の影響 Cadv_twitter 比較. 横軸, 縦軸, 分析区間は図 17 と同じ.

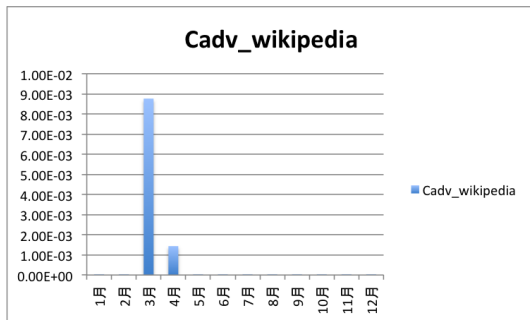


Fig. 23: ビキニ水着についての検索行動の数理モデルの Wikipedia の影響 Cadv_wikipedia 比較. 横軸, 縦軸, 分析区間は図 17 と同じ.

4 結論

本研究では検索行動を表現する数理モデルを提案した. 映画「君の名は。」「花見と紅葉」「水着」について本研究の数理モデルによる計算は実測をよく説明する.

興味深い点として, ブログと Twitter を比べると, 検索行動にはブログの影響の方が係数で 2 桁程度大き

いということがわかった. 即ち, 人々の興味関心には, Twitter よりもブログの方がよく効くことを示している. このことは, AKB 選抜総選挙の予測計算ではブログの解析の方が良いという石井・太田の報告⁷⁾とも一致している. まとめて, 本研究は検索行動を記述する新しい数理モデルを提案し, 映画や季節性のある話題の実例で Google Trend を検索のデータとして用いる場合, 非常にうまく検索数の時間的変化を記述できることがわかった. こうした人々の検索行動を数理モデルとして社会物理学で記述できることは, 検索という人々の興味関心のバロメータを記述できることになる. これはマーケティングなどに広く応用できる数理モデルであると思われる.

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) 063018 (22pp)
- 2) Ishii A, ”Analysis and Predictions of Social Phenomena via social media using Social Physics method”, A.Ishii, INFORMATION 20 (2017) 7073/7086
- 3) 大知正直・長濱憲・榎剛史・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』広報研究 11 (2016) 35/50
- 4) 石井晃, 吉田就彦, 新垣久史, 山崎富美『ヒット現象の数理モデルとマーケティング・サイエンス』鳥取大学工学部研究報告 37 (2007) 107/113
- 5) Pendry JB, 1980, ” Reliability factors for LEED calculations”, J. Phys.C13 937/944 (1980)
- 6) 漆谷たみこ『GRP を用いたヒット現象の数理モデルのパラメータ考察』(鳥取大学応用数理工学科 卒業論文) (2010)
- 7) 石井晃, 太田奨, 「ヒット現象の数理 映画から AKB 総選挙まで」 応用数理, 25(2) (2015) 50/58