

ソーシャルメディア上の季節性のある書込の解析

少林俊道¹ 岡野のぞみ¹ 川畑泰子² 石井晃¹

¹鳥取大学持続性社会創生科学研究科

²群馬大学社会情報学部

要旨：スキーや水着など、季節性のある話題について Twitter、Blog の書込と Google Trends、Wikipedia による検索数の測定を用い、それぞれの関連や時間差などを解析する。日本に限らず世界各国での書込の季節性も分析する。

1. はじめに

近年のスマートフォンやタブレット端末の普及により、私たちはSNSをより手軽に利用することが可能になった。そのため、TwitterやBlogなどで情報を発信でき、リアルタイムで情報を共有、拡散できるようになった。また、調べたいことやわからないこと、困ったことがあっても Googleをはじめとする検索エンジンですぐに検索ができ、情報を得ることが可能になった。

私たちの住む日本は、四季がありそれぞれの季節において行楽やレジャーなど様々な楽しみ方がある。日本のみならず海外からも日本の四季への関心は高く、長野県の白馬村では主に欧米圏の人々をターゲットにしたインバウンド着地型ツアーが人気である[1]。日本のスキー、スノーボード人口は減少傾向であるが、こうしたインバウンドの取り組みから改めて日本の良さ、四季の素晴らしさを感じることができるといえる。また、Google Trendsを用いて季節性のあるキーワードを入れて検索すると季節性が綺麗に見て取れる（図1-2）。

本研究では、スキーや水着などの季節性のある話題について、Twitter、Blogの書込とGoogle Trends、Wikipediaによる検索数の測定を用いた解析を行う。

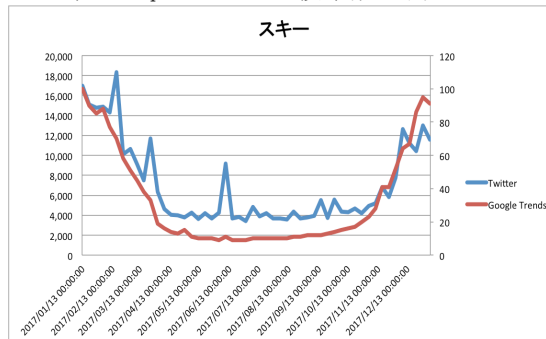


図1 Google TrendsとTwitterのスキーについての書き込み

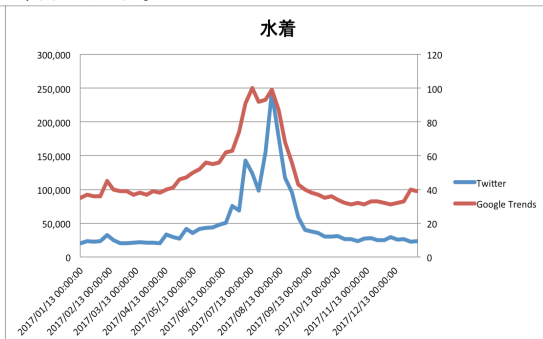


図2 Google TrendsとTwitterの水着についての書き込み

2. 方法

2.1 分析対象

季節性のある書き込み。以下にその例を示す。

- ・ スキー、スノーボード
- ・ 水着 (Bikini)
- ・ 花見、紅葉
- ・ クリスマス

2.2 データ取得

2.2.1 クチコミ@係長

Twitter、Blog の書き込み件数のデータはホットリンク社提供のクチコミ@係長を用いた。なお、取得できる Twitter の書き込み件数は 1/10 のサンプリングである。

2.2.2 Google Trends

クチコミ@係長では精度の高い検索数が取れないため、Google Trends でデータを取得した。Google Trends は 0 から 100 までの値を取っており、キーワードの検索数を全検索数で割り、その最大値が 100 になるように規格化されている。

2.2.3 Pageviews Analysis

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

2.3 分析方法

2.3.1 取得データに基づく分析

Twitter、Blog、Google Trends、Wikipedia の取得データを時系列ごとに関連するワードも一部含めて分析する。

2.3.2 数理モデルを用いた分析

今回は Google Trends を用いたヒット現象の数理モデルを用いて分析を行った。数理モデルの説明は 3 で行う。

3. 数理モデル

3.1 従来型のヒット現象の数理モデル

従来のヒット現象の数理モデルを示す[2]。

$$\frac{dI(t)}{dt} = \sum_{\xi} c_{\xi} A_{\xi}(t) + DI(t) + PI(t)^2 \quad (1)$$

$I(t)$ は、ある物事に対する興味・関心を表す量である。左辺は興味・関心の時間変化を表す。右辺の第 1 項はテレビやネットニュースによる影響を表し、第 2 項は会話による影響（直接コミュニケーション）を表し、第 3 項は噂による影響（間接コミュニケーション）を表す。

本論文では、ヒット現象の数理モデルの c, D, P の係数のことをパラメータとよんでいる。

3.2 Google Trends を用いたヒット現象の数理モデル

本研究に用いる Google Trends を解析するヒット現象の数理モデルを示す[3]。

$$\begin{aligned} \frac{dI(t)}{dt} = & DI(t) + PI(t)^2 + \sum_{i=0}^N c_{tv} A_{tv}(t) + \sum_{i=0}^N c_{news} A_{news}(t) + \sum_{i=0}^N c_{blog} A_{blog}(t) \\ & + \sum_{i=0}^N c_{twitter} A_{twitter}(t) + \sum_{i=0}^N c_{wikipedia} A_{wikipedia}(t) \quad (2) \end{aligned}$$

主な変更点として、従来のヒット現象の数理モデルでは興味・関心の指標として Twitter、Blog の書き込み数を用いていたが、本研究では Google Trends の関心度（検索数を規格化したもの）をターゲットとする。[3] では「Wikipedia」の影響は考えられていないが今回は「Wikipedia」の影響も加えることにする。「テレビ」、「ネットニュース」、「Blog」、「Twitter」「Wikipedia」の 5 つの影響を考えるため、それぞれの広告宣伝効果の和で表す。 A_{tv} はテレビの露出秒数、 C_{tv} はテレビの影響の強さを表す定数、 A_{news} はネットニュースの件数、 C_{news} はネットニュースの影響の強さを表す定数、 A_{blog} は Blog の書き込み件数、 C_{blog} は Blog の影響の強さを表す定数、 $A_{twitter}$ は Twitter のツイート件数、 $C_{twitter}$ は Twitter の影響の強さを表す定数、 $A_{wikipedia}$ は Wikipedia の検索件数、 $C_{wikipedia}$ は wikipedia の影響の強さを表す定数である。Google Trends の関心度についてのデータは検索数、You tube 検索数のいずれかである。

4. 計算結果

4.1 スキー・スノーボード

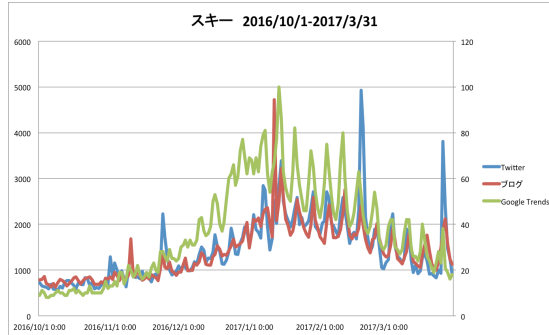


図3 スキーの書き込み件数と検索件数

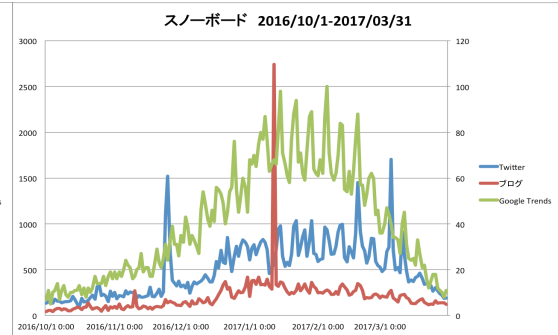


図4 スノーボードの書き込み件数と検索件数

図3、図4は2016年の10月1日から2017年の3月31日までのスキーとスノーボードの書き込み件数と検索件数の1日ごとの推移を示している。書き込み件数はスノーボードよりもスキーの方が圧倒的に多い。また、SNSへの書き込みが増えてくるよりも前から検索数が伸びてきていることがわかる。

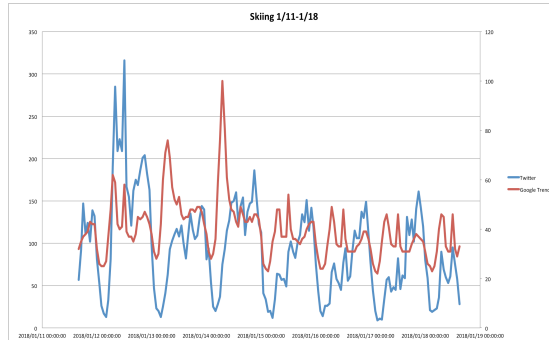


図5 1週間のスキーの書き込み件数と検索件数

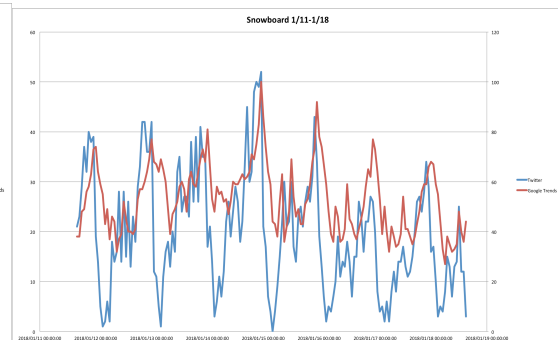


図6 1週間のスノーボードの書き込み件数と検索件数

図5、図6は2018年の1月11日から1月18日までの1週間のスキーとスノーボードの書き込み件数と検索件数の1時間ごとの推移を表している。図3のスキーは、それぞれの日ごとの検索件数のピークが書き込み件数のピークとずれている。一方で図6のスノーボードは検索件数と書き込み件数のピークがほぼ一致している。

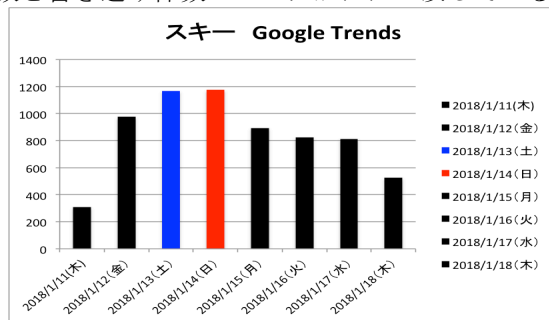


図7 スキーの検索件数

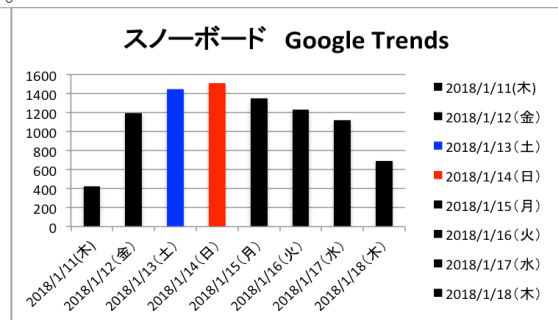


図8 スノーボードの検索件数

図7、図8は2018年1月11日から1月18日までの1週間のスキーとスノーボードの検索件数の1時間ごとのデータを足し合わせて1日単位に集計したものである。グラフから見て分かる通り土曜や日曜といった週末に検索件数が増えていることがわかる。

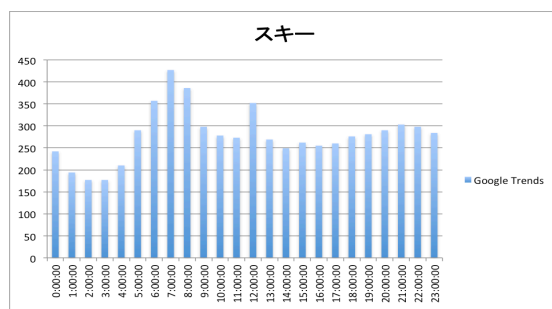


図9 1時間単位のスキーの検索件数

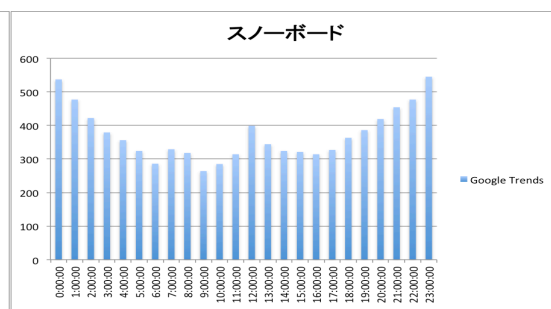


図10 1時間単位のスノーボード検索件数

図9、図10は2018年1月11日から1月18日までの1週間のスキーとスノーボードの検索件数の各日にちの同時刻の検索件数を足し合わせて1時間単位で集計したものである。スキーは午前7時付近にピークがあることがわかる。スノーボードは深夜0時付近にピークがあることがわかる。

以下はGoogle Trendsを用いたヒット現象の数理モデルの計算結果である。表1はキーワードにスキーを、表2はキーワードにSkiingを用いて計算して得られた値である。図11-17は各パラメータの値を示したグラフである。

表1 スキーのパラメータの値

スキー	10月	11月	12月	1月	2月	3月
D	9.95E-01	8.95E-01	3.44E-05	2.48E-02	2.66E-06	7.88E-01
P	9.88E-07	1.03E-06	1.09E-06	9.74E-07	7.29E-07	9.98E-07
Cadv_t	1.45E-04	5.32E-04	2.87E-08	5.51E-08	1.35E-05	1.69E-08
Cadv_n	5.27E-04	7.37E-05	6.29E-06	2.57E-01	5.48E-06	7.07E-06
Cadv_blog	3.08E-06	1.84E-05	2.01E-02	1.78E-05	2.63E-02	6.88E-07
Cadv_twitter	6.46E-05	1.88E-03	1.52E-02	2.37E-02	7.45E-04	5.17E-04
Cadv_wikipedia	1.66E-06	4.43E-06	2.13E-07	2.66E-03	2.20E-07	6.41E-03
R_factor	0.007	0.006	0.013	0.013	0.019	0.028

表2 Skiingのパラメータの値

Skiing	10月	11月	12月	1月	2月	3月
D	6.28E-01	8.29E-01	1.02E+00	9.89E-01	2.39E-04	7.89E-01
P	9.61E-07	9.84E-07	9.82E-07	9.92E-07	9.36E-07	9.48E-07
Cadv_t	3.73E-04	1.59E-03	1.19E-06	2.10E-07	5.41E-04	2.75E-08
Cadv_n	3.68E-05	2.29E-01	8.76E-04	1.29E-04	1.46E-04	1.01E-05
Cadv_blog	1.25E-02	6.82E-06	1.74E-04	2.31E-06	2.40E-02	8.31E-07
Cadv_twitter	6.63E-05	1.67E-05	2.08E-04	6.09E-06	2.15E-03	1.65E-06
Cadv_wikipedia	2.20E-06	7.89E-06	7.19E-06	3.42E-06	8.48E-03	9.15E-03
R_factor	0.005	0.007	0.007	0.008	0.003	0.012

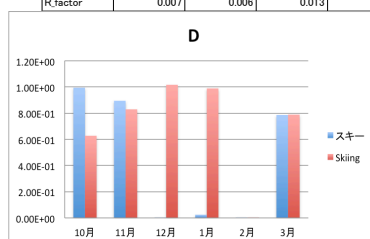


図11 Dの値

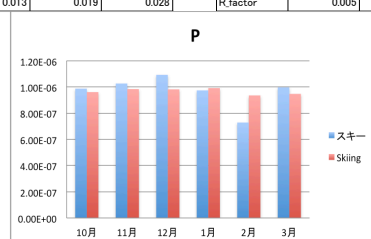


図12 Pの値

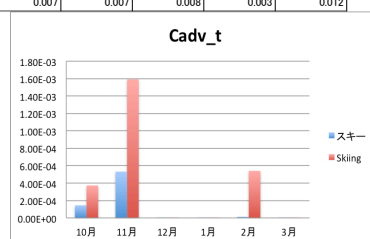


図13 Cadv_tの値

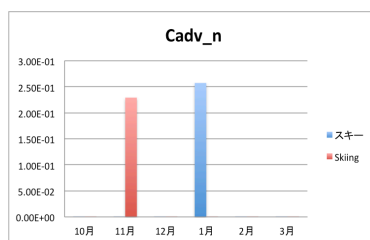


図14 Cadv_nの値

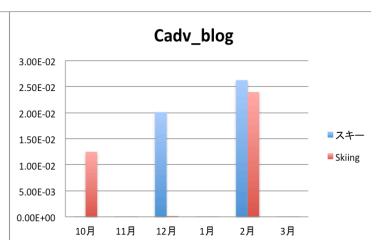


図15 Cadv_blogの値

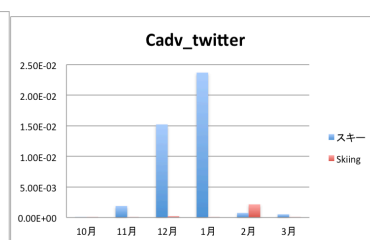


図16 Cadv_twitterの値

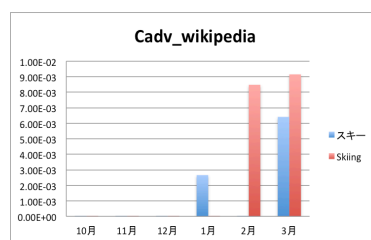


図17 Cadv_wikipediaの値

以下はGoogle Trendsを用いたヒット現象の数理モデルの計算結果である。表3はキーワード

にスノーボードを、表 4 はキーワードに Snowboard を用いて計算して得られた値である。図 18-24 は各パラメータの値を示したグラフである。

表 3 スノーボードのパラメータの値

スノーボード	10月	11月	12月	1月	2月	3月
D	8.71E-01	9.88E-01	1.02E+00	4.32E-01	8.87E-05	7.79E-01
P	1.00E-06	9.89E-07	1.06E-06	1.04E-06	8.52E-07	1.02E-06
Cadv_t	8.77E-06	6.70E-06	8.62E-06	1.75E-05	1.48E-06	2.10E-04
Cadv_n	4.86E-04	1.56E-01	1.89E-03	1.88E+00	6.25E-05	1.13E-04
Cadv_blog	1.38E-05	1.36E-05	6.90E-06	2.82E-03	7.92E-02	5.95E-06
Cadv_twitter	9.99E-03	2.60E-03	4.55E-06	4.87E-02	5.11E-02	1.00E-02
Cadv_wikipedia	5.25E-04	1.27E-05	5.23E-05	1.82E-05	1.71E-02	2.45E-04
R_factor	0.016	0.011	0.014	0.005	0.01	0.02

表 4 Snowboardのパラメータの値

Snowboard	10月	11月	12月	1月	2月	3月
D	9.05E-01	9.98E-01	1.01E+00	6.11E-01	9.91E-01	6.04E-01
P	9.66E-07	1.04E-06	9.68E-07	9.62E-07	1.04E-06	9.61E-07
Cadv_t	1.37E-05	2.56E-05	7.96E-06	2.29E-05	1.78E-06	7.17E-04
Cadv_n	2.11E-04	3.99E-01	7.16E-04	2.49E+00	1.19E-04	2.70E-04
Cadv_blog	1.42E-05	1.09E-05	5.60E-06	4.87E-03	8.38E-06	1.03E-05
Cadv_twitter	3.73E-05	8.50E-04	3.87E-06	2.86E-02	4.84E-06	4.37E-05
Cadv_wikipedia	1.81E-02	1.96E-05	1.46E-05	1.51E-05	3.54E-06	4.06E-02
R_factor	0.009	0.005	0.007	0.007	0.008	0.006

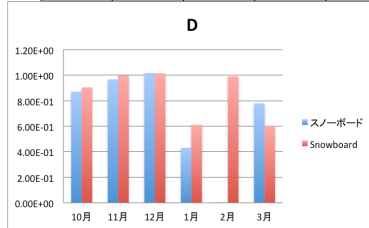


図 18 D の値

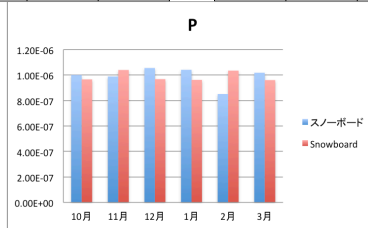


図 19 P の値

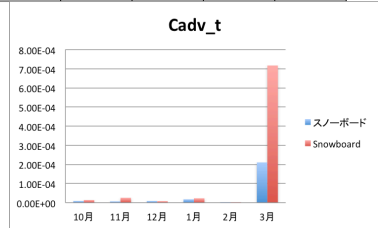


図 20 Cadv_t の値

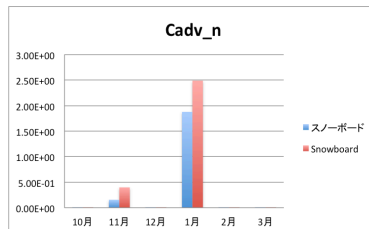


図 21 Cadv_n の値

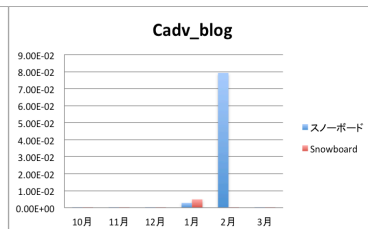


図 22 Cadv_blog の値

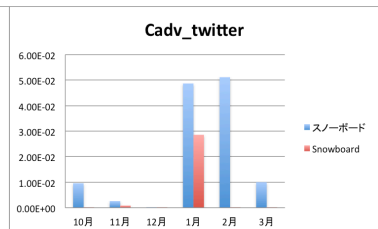


図 23 Cadv_twitter の値

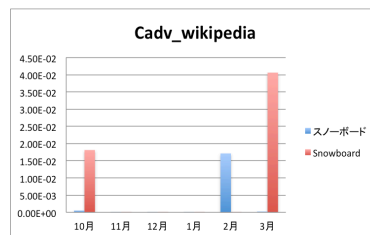


図 24 Cadv_wikipedia の値

4.2 水着(Bikini)

以下は Google Trends を用いたヒット現象の数理モデルの計算結果である。表 5 はキーワードに Bikini 用いて計算して得られた値である。図 25-31 は各パラメータの値を示したグラフである。

表 5 Bikiniのパラメータの値

Bikini	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
D	9.02E-01	7.21E-01	4.32E-05	7.61E-01	8.53E-01	8.00E-01	8.92E-01	9.00E-01	9.19E-01	9.54E-01	8.65E-01	1.02E+00
P	1.01E-06	9.67E-07	9.34E-07	1.04E-06	1.05E-06	9.99E-07	1.05E-06	1.02E-06	9.74E-07	1.00E-06	1.01E-06	9.94E-07
Cadv_t	2.81E-06	1.20E-05	6.80E-06	4.33E-06	6.81E-06	3.95E-06	6.31E-06	5.95E-06	6.65E-06	7.26E-06	5.05E-06	5.70E-06
Cadv_n	1.61E-04	4.83E-04	3.14E-04	6.90E-04	1.77E-03	1.99E-04	1.25E+00	1.97E+00	3.14E+00	8.10E-04	6.16E-04	2.52E-04
Cadv_blog	5.38E-06	1.15E-05	2.35E-06	1.40E-05	1.35E-05	5.28E-02	3.15E-05	3.03E-06	1.55E-05	1.54E-05	1.50E-02	3.83E-03
Cadv_twitter	3.08E-02	8.56E-02	9.58E-07	5.17E-02	1.38E-01	4.97E-06	2.03E-03	2.46E-05	5.39E-03	2.41E-02	6.57E-05	3.01E-06
Cadv_wikipedia	9.83E-08	2.77E-06	8.76E-03	1.43E-03	2.56E-05	1.65E-05	2.15E-07	7.09E-06	7.81E-06	9.38E-06	1.54E-06	7.57E-08
R_factor	0	0	0	0	0	0	0	0	0	0	0	0.001

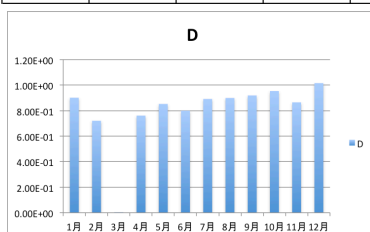


図 25 D の値

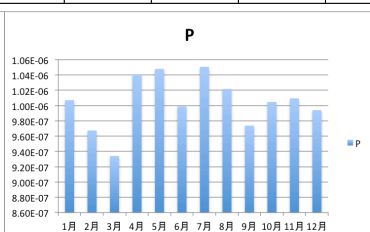


図 26 P の値

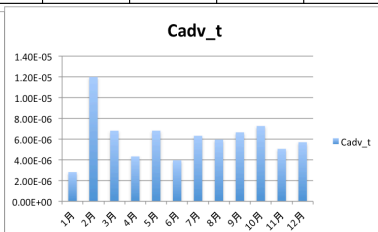


図 27 Cadv_t の値

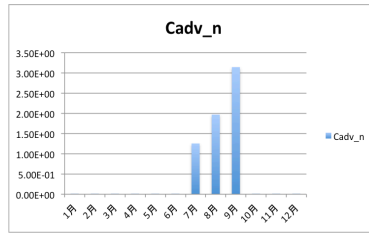


図 28 Cadv_n の値

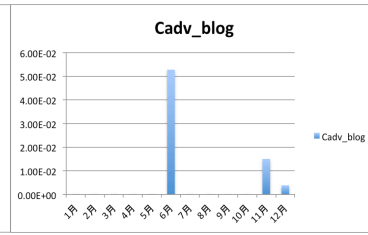


図 29 Cadv_blog の値

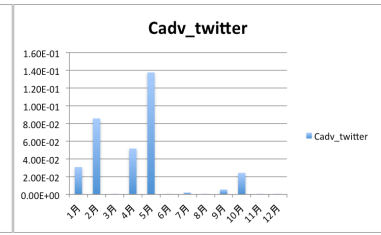


図 30 Cadv_twitter の値

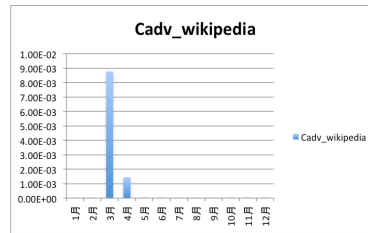


図 31 Cadv_wikipedia の値

4.3 花見・紅葉

以下は Google Trends を用いたヒット現象の数理モデルの計算結果である。表 6 はキーワードに花見を、表 7 はキーワードに紅葉を用いて計算して得られた値である。図 32-38 は各パラメータの値を示したグラフである。

表 6 花見のパラメータの値

花見	Before	After
D	9.61E-01	8.23E-05
P	9.96E-07	1.03E-06
Cadv_t	2.47E-06	1.83E-03
Cadv_n	4.22E-05	2.73E-01
Cadv_blog	8.68E-04	2.37E-06
Cadv_twitter	1.78E-04	2.48E-03
Cadv_wikipedia	2.03E-06	1.34E-06
R_factor	0.006	0.018

表 7 紅葉のパラメータの値

紅葉	Before	After
D	2.18E-05	9.67E-02
P	1.21E-06	1.07E-06
Cadv_t	2.13E-07	1.73E-03
Cadv_n	9.52E-06	1.22E-03
Cadv_blog	6.82E-03	5.09E-07
Cadv_twitter	4.47E-03	6.15E-03
Cadv_wikipedia	1.56E-07	2.77E-02
R_factor	0.014	0.019

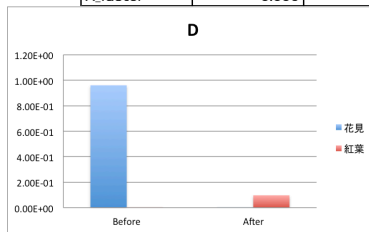


図 32 D の値

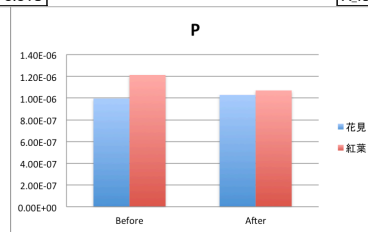


図 33 P の値

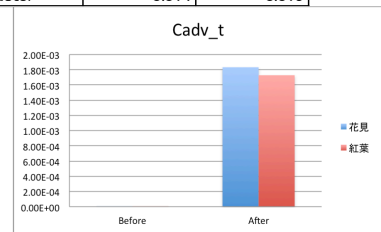


図 34 Cadv_t の値

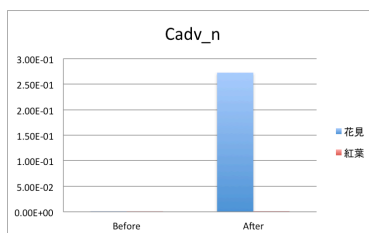


図 35 Cadv_n の値

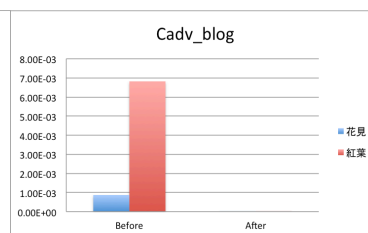


図 36 Cadv_blog の値

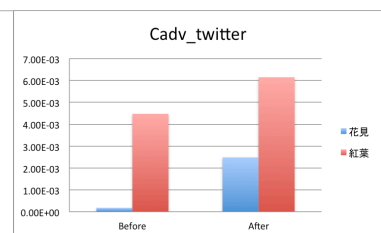


図 37 Cadv_twitter の値

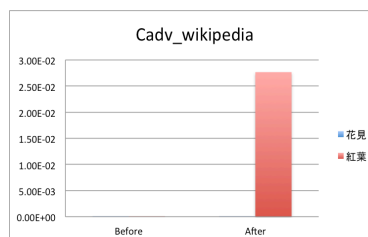


図 38 Cadv_wikipedia の値

4.4 クリスマス

以下は Google Trends を用いたヒット現象の数理モデルの計算結果である。表 8 はキーワードにクリスマス用いて計算して得られた値である。図 39-45 は各パラメータの値を示したグラフである。

表 8 クリスマスのパラメータの値

クリスマス	Before	After
D	5.63E-06	8.84E-06
P	5.43E-07	9.65E-07
Cadv_t	2.74E-08	2.57E-04
Cadv_n	1.50E-02	7.63E-07
Cadv_blog	4.50E-09	2.59E-04
Cadv_twitter	1.76E-04	2.75E-09
Cadv_wikipedia	1.03E-09	5.79E-09
R_factor	0.056	0.009

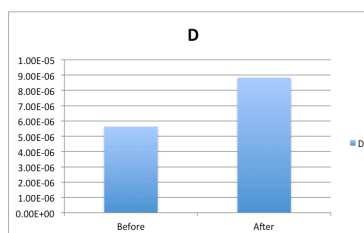


図 39 D の値

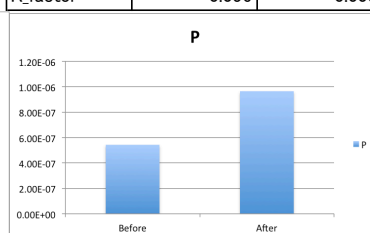


図 40 P の値

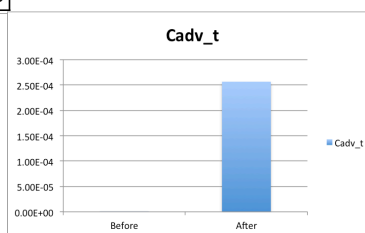


図 41 Cadv_t の値

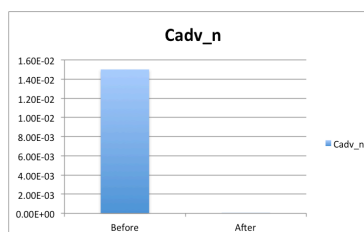


図 42 Cadv_n の値

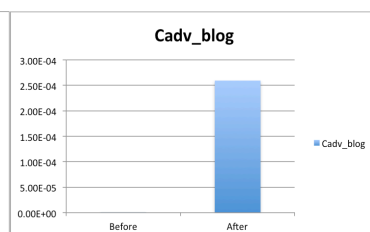


図 43 Cadv_blog の値

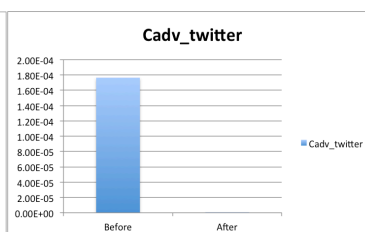


図 44 Cadv_twitter の値

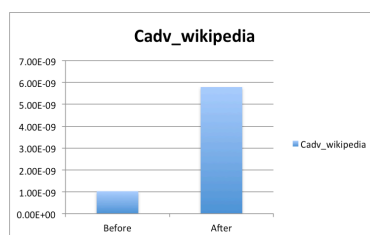


図 45 Cadv_wikipedia の値

5. 考察

5.1 スキー・スノーボード

図 7、図 8 より、スキーもスノーボードも土日を中心に検索件数が増えていることから、平日よりも休日にスキー、スノーボードを楽しむ人が多いことがわかる。また、図 9、図 10 より、スキーとスノーボードでは検索件数のピーク時点が違うことがわかった。スキーが朝の 7 時付近にピークがある要因としては、ゲレンデに行く前にゲレンデの天気やコンディションを調べていることや、車で移動する際は場所をあらかじめ調べているということがあげられる。また、スノーボードが深夜 0 時付近にピークがある要因としては、スキーよりもスノーボードの方が若い人で人気であり、若い人を中心として夜行バスでゲレンデに行くということや、You Tubeなどでスノーボードの動画を閲覧していることがあげられる。またこの結果から、主にゲレンデのオープン時間は朝の 8 時から 9 時までのところがほとんどであるから、車で日帰りできるゲレンデはおおよそ片道 1 時間から 2 時間であるという予想もできる。

数理モデルでの計算結果から、スキーもスノーボードもカタカナ検索の場合シーズンの12月、1月、2月に多いことがわかり日本では検索行動に SNS が影響を及ぼしていることが考えられる。東京は別として、地方では地元のグレンデまで片道1、2時間は珍しくないであろう。

5.2 水着 (Bikini)

数理モデルでの計算結果から、シーズンに入る前の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。これは、水着の検索はビーチやプールで水着を見るのではなく、事前に水着を買う人の購買意欲につながっているのではないかと考察する。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つながると考察できる。

5.3 花見・紅葉

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地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が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かわかるからだと考える。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の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と考えられる。

5.4 クリスマス

数理モデルの計算結果から、クリスマスの前後では、クリスマスの前の方がニュースと Twitter が検索に影響を及ぼし、クリスマス後はテレビとブログが検索に影響を及ぼしている。

クリスマスの前は、ニュースや Twitter で様々な情報が飛び交っていることが検索に影響を及ぼしていると考えることができ、クリスマス後は実際にクリスマスがどうだったのかという報告が検索に影響を及ぼしているのではないかと考察できる。

6. まとめ

従来のヒット現象の数理モデルではなく、Google Trends を用いたヒット現象の数理モデルを使うことで、人々の検索行動に対する各メディアの影響が見て取れた。季節性のある書き込みでも Google Trends を用いたヒット現象の数理モデルが適用できたと言える。また、検索と書き込みに時間的なズレが生じやすく、その点が季節性のある書き込みの特徴である。しかし、季節性に関する書き込みは、自然現象や天候に左右されることが多いが、本研究では、そうした自然現象や天候を考慮していない。今後この研究を深めていく上では自然現象や天候を考慮した上での分析が必要となると思われる。

参考文献

- [1] インバウンド着地型観光の手引き <<http://www.mlit.go.jp/common/001091713.pdf>> (2018.02.27 アクセス)
- [2] A.Ishii, H.Arakaki, N.Matsuda, S.Umemura, T.Urushidani, N.Yamagata and N.Yoshida. "The 'hit' phenomenon: a mathematical model of human dynamics interactions as a stochastic process". New Journal of Physics 14(2012)
- [3] 芦田昇,石井晃,川畑泰子 "Google Trends を用いた検索の数理モデル".第二回計算社会科学ワークショップ (2018)