

# ブログにおける話題関心の減退曲線推定の試み

大浜毅美

株式会社日経リサーチ

## 1. はじめに

「人の噂も七十五日」とは古来噂話への関心が長く続かないことの喩えだが、現代社会においてはある話題が2ヶ月半も取り上げられ続けることの方がむしろ珍しく、より短期間で急速に、その関心は減退していく。特にマスメディアが発信源となるような社会的な話題ではこの傾向が顕著であり、ほとんどの記事(=ニュース)は発生時にその関心(=情報量)がピークとなるが、以後新事実が発見されない限り、急速にメディアのその記事に対する関心は減退することとなる。

これに対して消費者生成メディア(CGM)、例えばブログの場合、例えば「株」や「子育て」などの特定のテーマだけを取り上げるサイトや、それらのテーマの情報集約自体が目的であるサイト(いわゆる「まとめサイト」)が高い割合で存在する(総務省, 2008)[1]ため、マスメディアやリアルコミュニティに比べればある話題に対する持続性は比較的高いことが予想される。福原ら(2005)[2]によるブログ記事の時系列推移の波形分類においても、周期型、突発型、関心持続型などの5つが見られるとしており、話題の内容によってブログ執筆者の反応が異なることが取り上げられている。

とはいえ、CGMにおいても社会的話題に関する一次提供元は新聞やテレビなどのマスメディアであることが多く、その影響は強く受けざるを得ない。このため、これらの話題の場合、全体的に見れば関心の推移は他のコミュニケーション手段と同様に急速な関心の増加と速やかな減退、というパターンをとることが多いと考えられる。

一方、そもそもこうした社会的話題に対する関心は実際どの程度持続するか、どのように減退してゆくのかといった測定は、記事数の推移調査やアンケート・観察調査などといった手法である程度は行われていたものの、いずれも量(記事数や被験者数)と質(時系列の密度や客観的な数値の取得など)の両立は難しく、関心

の減退を定量的に測定することは困難であった。

しかし、ブログについては日本語で記述されたサイトに限ったとしても総数1500万サイト以上の記事が公開されており(総務省, 2008)、いわゆるスパムブログを除いたとしても十分に統計的に安定した測定が可能である。また、ブログの特徴の1つであるRSSフィードにより、一般のWebサイトに比べて公開日や執筆者の情報が整理されており、そのキーワードが示す「話題」に対する反応(=記事数)を定量的に、時間単位の時系列で測定することが可能である。

本研究では、こうしたブログ記事数の推移を元に、「話題」について反応の時系列推移を観測し、人の噂の減退状況の検討を行うことを目的とする。

## 2. 方法

社会的話題(時事・経済・芸能などの話題)に対するブログ記事の反応を調べるため、以下の方法で実験を行った。

### 1) 対象話題の選択

ブログ検索エンジン「Ask.jp」で公開されている「話題のニュースランキング」を用い、2008年5月1日から2008年5月15日までの1位~5位のニュースから、重複記事を除き計69本を検索対象の「話題」とした。このランキングは、インターネット上で公開されたニュース記事のURLがAsk.jpの保有する記事DBにどの程度含まれているかによって順位付けられるため、例えばニュースサイトのクリック数に基づくランキングに比べ、ブログ記事数との関連が高いという特徴を持つ。

### 2) ブログ記事の検索

それぞれの話題についてその内容を特徴づける検索文字列を決定し、日経リサーチのブログ分析サービス「blogViz センサー(<http://viz.nikkei-r.co.jp/>)」を用いてその語が含まれるブログ記事

の URL を取得した。なお、blogViz センサーの機能により、内容がスパムブログと判断される記事はこの段階で除外されている。取得した URL 情報を元に各記事の RSS 情報および記事の HTML の解析を行い、その記事の公開日を推定し、日毎に記事数をカウントした。

3) 話題数によるピーク(関心最大日)の検出  
各話題について、1 日あたりの記事数が最大となる日をピークとし、この日から 15 日を分析対象期間とした。ただし、いったん記事数が減退した後に再度記事数が増加したものの、つまり複数のピークが見られるものは目視により最後の山がピークとなるよう調整した。また、ピーク日の記事数が 50 未満のもの(36 記事)は誤差が大きいため今回の曲線推定の対象外とした。この結果 33 の話題が今回の分析対象となった。

#### 4) べき曲線推定

ピーク日からの日別記事数の時系列推移を、従属変数を記事数、独立変数を日次の時間軸として曲線推定した。「話題への関心は急速に減退する」という仮説のもと、いくつかのモデルの中からべき乗モデルを採用してあてはまりを検討した。なお、SPSS では変数に 0 が含まれていると対数の算出ができないため、記事数に便宜的に 1 を加えた数値を解析対象とした。

### 3. 結果

表 1 に対象となった話題のピーク日記事数、べき乗モデル推定結果の決定係数(R2)、F 値、係数を、図 1 に話題の中で特にモデルの当てはまりの良かった「母の日」を例にとり時系列グラフを図示する。なお表中の係数はべき乗モデル

$$E(Y_t) = a \cdot t^{-b}$$

の推定値(b0,b1)を示す。この結果、33 の話題中 25 以上について、決定係数 0.7 以上の高い当てはまりを示すことが確認された。また、これらの多くが指数-1 以下の値を示し、それぞれ話題に対する関心が急激に減退し、収束しているといえる。

また、これらの傾向は執筆された記事の数、すなわち話題自体の注目度や、話題のジャンル・重要さといった内容との関連は特に見られなかった。

表 1 分析対象キーワードとモデル要約

Keyword	Peak	R2	F	0	1
母の日	29459	0.976	518.054	31732.463	-1.600
小栗旬 山田優	485	0.946	228.847	577.346	-1.399
嵐 24時間	1015	0.937	192.823	706.684	-1.553
船場吉兆	529	0.901	117.854	580.343	-1.260
魔王 嵐	245	0.892	107.493	176.844	-0.997
伊達公子	373	0.885	100.164	448.451	-1.000
井ノ原朝香 拳式	64	0.883	98.446	57.835	-1.627
暫定税率	3270	0.883	98.239	3646.398	-1.405
中丸 環境問題	194	0.872	88.560	172.622	-1.768
アグリ	1100	0.870	87.251	1464.585	-1.422
野茂 楽天	185	0.868	85.560	129.848	-1.938
歴史的 和解 美味しんぼ	62	0.860	80.005	101.740	-1.689
四川 地震	1288	0.850	73.388	1343.734	-0.501
サポーター 浦和 ガンバ	149	0.848	72.336	158.601	-1.789
氷室 セルフカバー	57	0.836	66.416	28.484	-1.412
地震 カエル	75	0.830	63.348	57.222	-0.935
中居 倅田	100	0.825	61.278	72.443	-1.315
口移し 事故	79	0.786	47.869	39.819	-1.428
古畑任三郎	551	0.780	46.110	350.866	-1.396
マイクロソフト ヤフー 買取	92	0.764	42.082	149.842	-1.313
真夏日	484	0.757	40.521	698.999	-1.030
マクドナルド 値上げ	127	0.756	40.297	98.844	-1.002
琢磨 ホンダ	213	0.732	35.544	241.133	-1.359
被災者 支援物資 奪い合い	60	0.730	35.202	76.336	-1.001
110番 コキブリ	93	0.713	32.333	96.741	-1.426
ハーゲンダッツ 超高級,1000円	85	0.685	28.319	63.393	-1.075
ガソリン 価格	1060	0.657	24.899	633.927	-0.694
ラルク バリ	85	0.620	21.200	62.861	-0.944
水谷豊 ライブ	69	0.535	14.970	29.460	-0.826
海外旅行 離れ	50	0.460	11.080	20.482	-0.661
緊急速報	70	0.460	11.058	25.502	-0.797
タスボ	158	0.457	10.941	126.721	-0.227

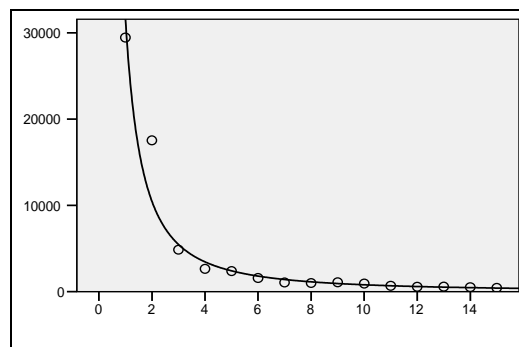


図 1 キーワード「母の日」5/11からの推移

#### 4. まとめ

本研究の結果により、ブログにおいて社会的話題のほとんどが急速に減退し、その曲線は指数関数的に収束していくことが明らかになった。またこの傾向は話題の注目度やテーマによらず、広い内容で見られることが示された。しかし、話題によってはこのようなモデルが当てはまらず、断続的に話題が続くもの、急に盛り返すものなども存在している。これらの要因の研究については、今後の課題としたい。

#### 5. 文献

- [1] 総務省情報通信政策研究所 ブログの実態に関する調査研究. 2008
- [2] 福原知宏, 村山敏泰, 中川裕志, 西田豊明 ウェブログ記事を用いた関心解析システム, 人工知能学会全国大会 2005