

クチコミによるネット炎上の定量化の試みとその検証

Verification of an Algorithm for Detecting the Negative Reactions Based on Viral Media

大曾根 匡[†] 福田 浩至[‡]
Tadashi OSONE Koji FUKUDA[‡]

[†] 専修大学 経営学部
[‡] 株式会社ループス・コミュニケーションズ
[†] School of Business Administration, Senshu University.
[‡] Looops Communications Inc.

要旨

Facebook やブログなどのソーシャルメディアの急速に普及により、誰でも気軽にネットに投稿できる環境が整っている。このため、企業の不祥事や個人の反感を招く言動は、瞬く間にネット上に拡散する。このような現象を「ネット炎上」と称し、「ネット炎上」に見舞われた企業・団体は、評判を貶められ、少なからず不利益を被ることになる。そこで、筆者らは、ネット上の該当するクチコミ数の時間推移データをもとに、炎上状態の検出と炎上規模の定量化を試みた。本稿では、「ネット炎上」を検出するアルゴリズムについて説明し、実データによる検証結果を報告する。

1. はじめに

Facebook やブログ、Twitter、LINE などに代表されるソーシャルメディアの急速に普及により、誰もがそれを媒介としてネット上に自らの考えを発信することが可能となった。例えば、全世界で 17 億人が利用する Facebook 上では、1 分間に 3 千万件を超えるメッセージと 270 万件の動画が投稿され続けているといわれている。このため、企業の不祥事や個人の反感を招く言動は、瞬く間にネット上に拡散する。このような現象を「ネット炎上」と称し、「ネット炎上」に見舞われた企業・団体は、評判を貶められ、少なからず不利益を被ることになる。そこで、筆者らは、ネット上の該当するクチコミ数の時間推移データをもとに、炎上状態の検出と炎上規模の定量化を試み、ネット炎上を検出するアルゴリズムを提案した。最初に、そのアルゴリズムについて説明し、次に、そのアルゴリズムを約 9 年間の 4 企業のブログデータに適用し、それによって得られた炎上期間や炎上規模に関する統計値を示す。最後に、アルゴリズムの妥当性を検証するために、炎上状態と判定された期間における原因事象を調査し、炎上規模と話題とを対応付けた結果を示す。

2. ネット炎上検出アルゴリズム

提案するネット炎上検出アルゴリズム [1] の概要について説明する。炎上日かどうか判定したい日を予測日とし、その予測日を含む観測期間内のクチコミ件数のうち、予測日を含むある一定期間を除いたクチコミ件数を観測値として使用する。そして、その観測値から回帰直線を導出し、予測日の実観測値が、回帰直線による推定値より大きく外れていると判断された場合、その予測日を話題拡散日とする。そして、話題拡散日におけるクチコミの内容をチェックし、ネガティブの内容である場合、その日を炎上日と認定する。ネット炎上日を検出するための具体的なアルゴリズムは以下のとおりである。

(1) 観測期間と除外期間の設定

予測日 x_0 が中心となるように観測期間と除外期間を設定する (図 1 参照)。除外期間を除いた観測期間内で、観測日 x_i に対するクチコミ件数 y_i を取得する。そのペアのデータを

$$(x_i, y_i) \quad (i = 1, 2, \dots, n)$$

とする。

(2) 回帰直線の導出回帰直線の導出

観測期間内の観測日とクチコミ件数のペアのデータ (x_i, y_i) ($i = 1, 2, \dots, n$) から回帰直線 $y = ax + b$ を求める。すなわち、

$$a = S_{xy}/S_{xx}$$

$$b = \bar{y} - a\bar{x}$$

であり、この式における

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

である。

(3) 予測日のクチコミ件数の推定量 \hat{y}_0 の平均と標準偏差の導出

予測日 x_0 に対するクチコミ件数 y_0 を推定するために、その推定量を表す確率変数 \hat{y}_0 を導入し、

$$\hat{y}_0 = Ax_0 + B$$

とする。ここで、 A と B は、それぞれ、回帰式の回帰係数 a と切片 b を表現する正規分布に従った確率変数であり、

$$A \sim N\left(a, \frac{V_e}{S_{xx}}\right)$$

$$B \sim N\left(b, \left\{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\}V_e\right)$$

である。また、 V_e は残差分散であり、

$$V_e = \frac{S_{yy} - aS_{xy}}{n - 2}$$

である。このとき、予測日 x_0 に対するクチコミ件数の推定量 \hat{y}_0 の分布は、下記の予測値平均 μ と予測値標準偏差 σ をもつ自由度 $n - 2$ の t 分布に従うことが知られている：

$$\mu = E[\hat{y}_0] = ax_0 + b$$

$$\sigma = \sqrt{V[\hat{y}_0]} = \sqrt{\left\{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\}V_e}$$

(4) 拡散指数 d の導入と話題拡散日の検出

データ数 n が十分大きいと仮定し、 t 分布を正規分布で近似する。そして、正規分布の 3σ の法則から、話題拡散日判定のための閾値を $\mu + 3\sigma$ とし、これを予測値上限と呼ぶことにする。そして、予測日 x_0 に対するクチコミ件数の観測値 y_0 が閾値である予測値上限 $\mu + 3\sigma$ を超えていたら話題拡散日と判定することにした。なぜなら、その閾値を超える確率は 0.003 以下だからである。その判断のために、予測日 x_0 の拡散指数 d を

$$d = \frac{y_0 - \mu}{\sigma}$$

と定義し、その拡散指数が3を超えていたら、その予測日 x_0 を話題拡散日とする。このようにして、話題拡散日を検出することにした。

(5) 話題拡散期間と話題拡散規模の定義

炎上は数日間続くことがある。また、最初の火種が小さくても、その小さな炎上が沈静化する前に新たな話題が発生すると大規模な炎上となったり、長期間の炎上となったりする。そこで、話題拡散日を検出した日の前後を観測し、拡散指数 d が1以上3未満の日を準話題拡散日と定義し、話題拡散日と準話題拡散日が連続する期間を話題拡散期間と定義する。なぜなら、クチコミ件数の観測値 y_0 が $\mu + \sigma$ を超える確率は0.16と比較的小さく、通常のバラツキの影響とは考えにくいからである。そして、その話題拡散期間における拡散指数の総和を話題拡散規模と定義した。

(6) 炎上日の判定

話題拡散期間における話題の内容を調査し、それがネガティブな話題の場合、その話題拡散期間に含まれる日をネット炎上日と判定する。

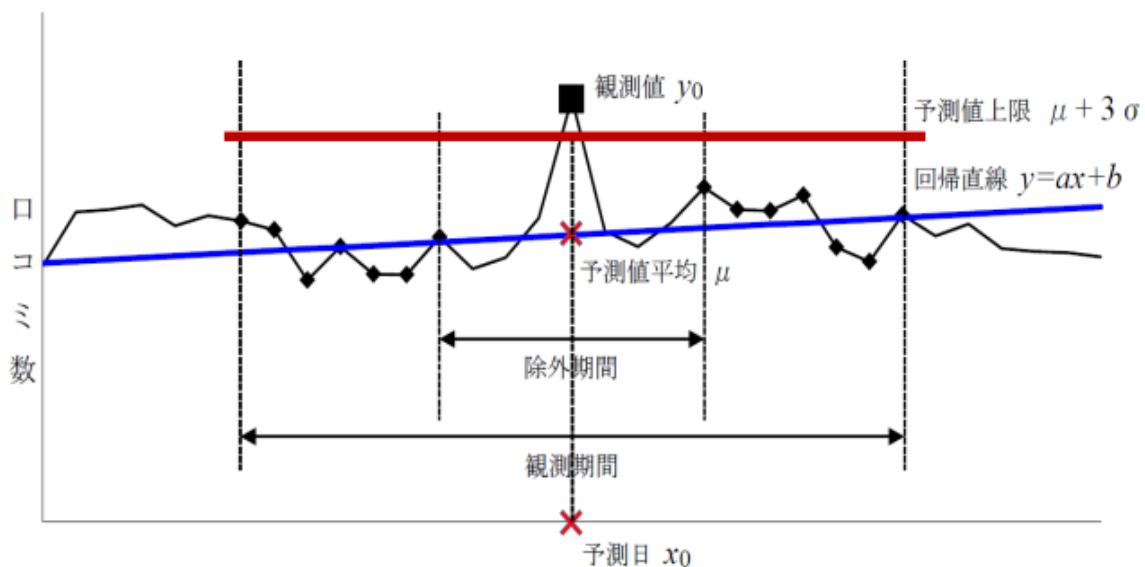


図1 炎上日検知アルゴリズムの図解

3. ネット炎上の定量化の事例

3.1. 観測期間と除外期間の設定

ネット炎上の定量化の事例として、すき家、吉野家、松屋、ワタミの飲食業界4企業におけるブログ上のクチコミ数を取り上げる。株式会社ホットリンクから提供していただいた2006年11月1日から2016年1月31日までの約9年間3379日のブログ上のクチコミ件数のデータに対しネット炎上検出アルゴリズムを適用し、ネット炎上の定量化を行った。その際、曜日によるクチコミ数の変動の影響を考慮し、除外期間を7日に設定した。そして、観測期間 T を21日間から105日間まで14日間づつ変化させ、話題拡散日数を測定した。そのグラフを図2に示す。これより、観測期間 T が49日間あたりで減少割合が小さくなり、変化が少なくなっていることがわかる。一方、話題拡散の平均間隔を4企業について調べてみると(表2参照)、最も短いワタミにおいて、それが約55日間であることがわかった。したがって、観測期間を55日間より長くすると同一の観測期間で2回以上の話題拡散が発生する確率が高くなり、それが回帰直線の生成に影響を与えるので、観測期間を49日に設定することにした。

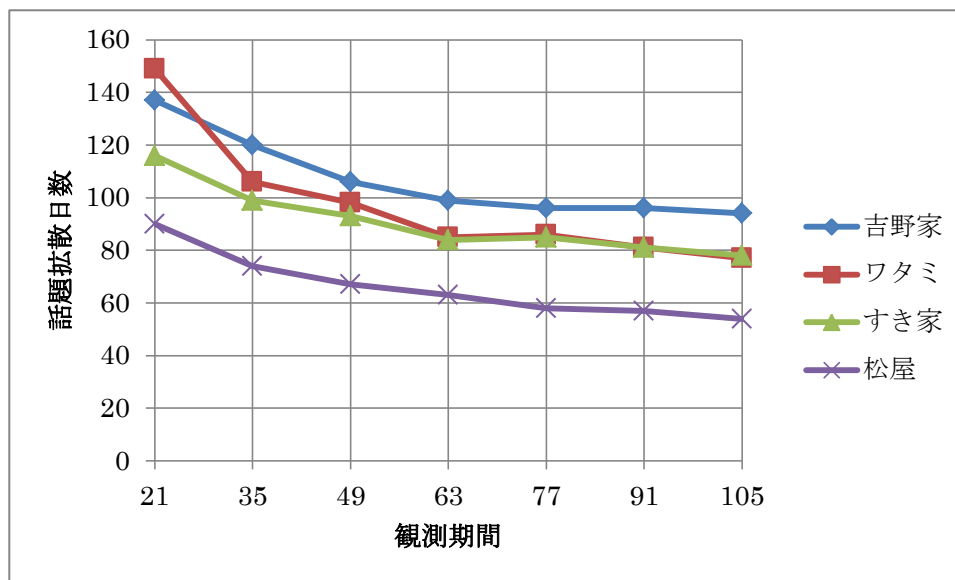


図2 観測期間と話題拡散日数との関係

3.2. 話題拡散の事例

ネット炎上の定量化の事例として、観測期間を49日間、除外期間を7日間としたときの、2015年12月の2週間のワタミと吉野家のクチコミ件数と拡散指数を表1に示す。ワタミで拡散指数が3を超え、話題拡散日と判断されたのは、2015年12月8日からの4日間である。さらに、12月12日も拡散指数が1以上であるので準話題拡散日とし、話題拡散期間に含める。したがって、話題拡散期間は5日間で、その期間の話題拡散規模は40.32となる。話題拡散の起点日となる12月8日は、ワタミの子会社で働いていた従業員が過労自殺したことに対し、遺族が損害賠償を求めて東京地裁に起こした訴訟の和解が成立した日であり、ワタミ側が1億3365万円を遺族に支払い、謝罪することになった。このように話題の内容がネガティブなので、この5日間は「炎上状態」と認定する。

一方、吉野家の話題拡散日は2015年12月10日の1日であり、それに引き続く2日間が準話題拡散日となった。したがって、話題拡散期間は3日間、話題拡散規模は12.96となる。この日は、「牛丼を3か月食べ続けても、健康リスクが増える兆しは見られなかった」という研究結果を吉野家が公表した日である。一見、ポジティブな話題にもみえるが、実験の人数が24名と少ないことや、吉野家自身による発表であったために、ネガティブな話題と捉え、この3日間も「炎上状態」と認定した。

表2に、2006年11月1日から2016年1月31日までの4企業のクチコミ件数と話題拡散件数と話題拡散の平均間隔と話題拡散日数を示す。クチコミ件数のボリュームが最大なのは吉野家の約110万件であり、最小のワタミの2倍以上であった。しかし、話題拡散件数では、クチコミ件数の最小のワタミが59件で最大であった。一方、話題拡散件数が少ないのが44件の松屋であった。このように、クチコミ件数のボリュームの大小に関係なく、提案したアルゴリズムは話題拡散日を検出していることがわかる。話題拡散の平均間隔については、ワタミが約55日間と最も短く、松屋が76日間と長いことがわかった。話題拡散日と認定された日数が最も多いのが106日の吉野家であり、最も少ないのが松屋の67日であった。

次に、4企業の話題拡散規模に関する基本統計値を表3に示す。これから、4企業のうち、話題拡散規模の平均が大きいのが15.51の吉野家である。一方、比較的小さいのが12.89と12.66のすき家と松屋であることがわかった。しかし、松屋は最大の話題拡散を起こしている。それは、2008/8/27の「並井を380円に30円値上げ発表」という話題であり、その話題拡散規模は107.20であった。

話題拡散期間についての基本統計値を表4に示す。話題拡散期間の平均が長いのはすき家と吉野家の約4日間であり、逆に短いのがワタミの3日間であった。最長の炎上期間は12日間であり、それは2014/3/21の「すき家の大量休店騒動」という話題であった。

表1 ワタミと吉野家のクチコミ件数と拡散指数

年月日	ワタミ		吉野家		年月日	ワタミ		吉野家	
	クチコミ件数	拡散指数	クチコミ件数	拡散指数		クチコミ件数	拡散指数	クチコミ件数	拡散指数
2015/12/01	49	0.008	102	-0.559	2015/12/08	93	4.062	84	-1.853
2015/12/02	36	-0.305	113	-0.157	2015/12/09	249	16.576	115	0.480
2015/12/03	43	-0.150	119	0.045	2015/12/10	172	10.951	239	10.060
2015/12/04	30	-0.461	100	-0.656	2015/12/11	105	6.358	122	1.124
2015/12/05	28	-0.485	104	-0.481	2015/12/12	69	2.370	131	1.773
2015/12/06	31	-0.435	120	0.196	2015/12/13	52	0.334	117	0.753
2015/12/07	35	-0.227	96	-1.081	2015/12/14	68	0.608	98	-0.495

表2 話題拡散の基本統計値

	すき家	吉野家	松屋	ワタミ
クチコミ件数	901,018	1,096,947	842,372	492,733
話題拡散件数	54	58	44	59
話題拡散の平均間隔	61.69	57.43	75.70	54.80
話題拡散日数	93	106	67	93

表3 話題拡散規模

	すき家	吉野家	松屋	ワタミ
平均	12.89	15.51	12.66	13.88
標準偏差	7.02	14.08	16.12	13.81
最大	33.47	79.20	107.20	68.44

表4 話題拡散期間

	すき家	吉野家	松屋	ワタミ
平均	4.07	4.00	3.73	3.07
標準偏差	2.23	2.00	2.20	2.04
最大	12	8	9	11

4. 話題拡散規模と事象との対応付け

話題拡散規模が20以上となった話題は40件あった。その話題の原因となったと思われる事象を調査し、話題拡散規模と対応付けを行い、拡散規模の数値の妥当性を検証した。紙面の関係で本稿ではそのうちの上位20件を表5に示す。この中には、ランキング2位の吉野家の「びっくりラーメンの買収」やランキング3位のワタミの「ゆずとのコラボレーション」など、ポジティブな話題も含まれている。そこで、ネガティブな話題と解釈できるものに「N」を、ポジティブな話題と解釈できるものには「P」を種別欄に記した。ランキング1位の松屋の「並井の30円値上げ発表」は、ネガティブな話題のように見えるが、松屋自身による発表であり、話題拡散により広報の目的が達成されたとも考えられるので、ネガティブな話題とは判断しなかった。

40件の中には対象とならない話題も混入していた。ランキング24位の話題は調べてみると百貨店の「銀座松屋」の話題であり、ランキング36位の話題はSEO対策によりクチコミ件数が急激に増加したことによる影響であることが判明した。このように、対象外の話題が混入することもあるので、話題の内容チェックは不可欠であることがわかった。また、ランキング6位の東日本大震災の話題の起点年月日が2011年3月10日となっている。これは、たまたま地震発生前日の拡散指数が1.8となっていて、この日が準話題拡散日と判定されてしまったことによる。

さて、40件の話題のうち、ネガティブな話題と認定した炎上話題は19件であった。その企業別内訳は、ワタミが12件、吉野家が4件、すき家が3件、松屋が0件であった。特に上位10件のうち、ワタミが6件、吉野家が4件と、この2企業がブログ上で大きなネット炎上を発生させていたことがわかった。このように、炎上規模の数値化はある程度妥当性があり、話題間や企業間のネット炎上の比較をするためのひとつの指標の役割を果たしていると考えられる。

表5 話題規模ランキング上位20件と原因事象との対応

順	発生年月日	企業	話題拡散規模	話題拡散期間	原因事象	種別
1	2008/08/27	松屋	107.20	5	「並井」380円に30円値上げ発表	
2	2007/08/30	吉野家	79.20	6	「びっくりラーメン」を買収	P
3	2012/01/11	ワタミ	68.44	2	「ゆず」とコラボレーション	P
4	2008/04/23	吉野家	63.50	4	吉野家向け米国産牛肉に特定危険部位	N
5	2012/02/22	ワタミ	62.68	5	ワタミ女性社員自殺の労災認定	N
6	2011/03/10	吉野家	55.75	4	地震による食材不足で24時間営業店が閉店	N
7	2007/06/11	ワタミ	45.87	5	コムスの介護施設買収に前向き	P
8	2007/12/03	吉野家	41.05	7	悪ふざけ動画を店員がアップ	N
9	2015/12/08	ワタミ	40.32	5	ワタミ女性社員の過労自殺裁判和解	N
10	2014/11/11	ワタミ	35.76	5	2015年3月末までに102店閉鎖と発表	N
11	2012/01/05	松屋	35.56	6	「豚めし」廃止発表	
12	2009/06/18	吉野家	34.97	5	ラーメン事業から撤退	N
13	2014/03/27	ワタミ	34.51	5	ワタミ裁判で社員が傍聴席を占拠	N
14	2009/04/15	すき家	33.47	5	残業代不払いで告発した店員を逆告訴	N
15	2012/10/08	ワタミ	27.79	3	ワタミ従業員の過労死記事掲載（毎日新聞）	N
16	2011/01/10	松屋	27.35	8	牛丼値下げ競争（240円）	
17	2008/08/28	すき家	27.05	7	「キン肉マン」とコラボ	P
18	2010/09/08	吉野家	26.78	7	低価格メニュー実施	
19	2009/12/07	すき家	26.38	3	牛丼最安値280円に値下げ	
20	2007/10/25	すき家	25.86	8	「メガ牛丼」発売	P

5. まとめ

本稿では、ネット炎上検出アルゴリズムにより、炎上規模の定量化を実現し、話題間や企業間のネット炎上の規模を比較することが可能となった。さらに、原因事象と比較すると、その炎上規模の数値がある程度妥当性のあることを確認した。また、商品の新発売などのポジティブな話題に関しては、その宣伝効果の指標としてもこの定量化手法が使えると考えられる。一方、その話題の内容の判断を目視によるチェックで行っており、その自動化が今後の課題である。さらに、関係のない企業の混入やSEO対策の影響による話題の排除も不十分である。実用化にあたっては、これらの課題を解決する必要がある。最後に、データの提供をいただいた株式会社ホットリンクに感謝したい。

参考文献

- [1] 福田浩至, 大曾根匡, “クチコミデータに基づく炎上日の解析と検証”, セキュリティ・マネジメント学会第30回全国大会予稿集, 2016.