

クチコミ拡散の定量化手法の開発とメディア別の傾向分析

福田 浩至[†] (株式会社ループス・コミュニケーションズ)大曾根 匡[‡] (専修大学 経営学部)

はじめに

ソーシャルメディアの普及とともに、ネット上のクチコミが企業などの組織やブランドに与える影響力が高まってきた。事業活動の過程で生まれる新商品や不祥事など様々な話題について、生活者は多様な観点で関心を持ち、今この瞬間にも語られている。その反響は、クチコミが増えている状況を感覚的にとらえて評価される傾向がある。筆者らはクチコミの拡散現象の定量化を試み、その拡散の発生タイミングや規模を数値化することを可能とした。ネット上で拡散した事象について、その規模を比較できるため、クチコミデータを企業の評判把握やプロモーションの目標設定等にも活用可能となる。さらに、本稿ではソーシャルメディアの種類による傾向の違いについて解説する。

1 クチコミ拡散状況の定量化

提案するクチコミ拡散状況の定量化アルゴリズム [1] について説明する。話題拡散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。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) ($i = 1, 2, \dots, n$)

とする。

(2) 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、

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

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

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

である。

(3) 予測日のクチコミ件数 \hat{y}_0 の標準偏差の導出

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、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。ここで、 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 \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知

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とす

Development of a Method for Detecting the Reactions Based on Viral Media and Analysis of the Characteristics of Each Media

[†] Koji FUKUDA Loops Communications

[‡] Tadashi OSONE Senshu University

る予測値上限を $\mu + 3\sigma$ とし、予測日 x_0 に対するクチコミ件数の観測値 y_0 が閾値とする予測値上限 $\mu + 3\sigma$ を超えていたら外れ値と判定することにした。それを判断するために、予測日 x_0 の拡散指数 d を

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

と定義し、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、話題拡散日を検知することにした。

(5) 話題拡散期間と話題拡散規模の定義

拡散は数日間続くことがある。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。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。

2 メディア別のクチコミ拡散特性

2011/1 から 2015/12 までの Twitter とブログを対象として、「すき家」「吉野家」「松屋」「ワタミ」というキーワードを含むクチコミの拡散状況の定量化を試みた。表 1 は Twitter の拡散回数の経年推移を、表 2 にはその規模を示す。同様に表 3 にはブログの拡散回数の経年推移を、表 4 にはその規模を示す。

表 1 年別 Twitter 拡散回数

年	すき家	吉野家	松屋	ワタミ
2011	5	9	7	7
2012	7	5	11	10
2013	8	8	7	9
2014	11	7	12	7
2015	10	9	7	9
合計	41	38	44	42

表 2 Twitter 拡散規模の統計値

	すき家	吉野家	松屋	ワタミ
累計拡散規模	437. 48	529. 92	479. 22	717. 23
拡散回数	41	38	44	42
平均拡散規模	10. 67	13. 95	10. 89	17. 08
最大拡散規模	38. 56	43. 34	52. 43	139. 56

表 3 年別ブログ拡散回数

年	すき家	吉野家	松屋	ワタミ
2011	5	6	4	6
2012	4	6	5	9
2013	7	6	4	6
2014	6	5	3	7
2015	5	6	3	7
合計	27	29	19	35

表 4 ブログ拡散規模の統計値

	すき家	吉野家	松屋	ワタミ
累計拡散規模	320. 75	371. 58	200. 80	574. 62
拡散回数	27	29	19	35
平均拡散規模	11. 88	12. 81	10. 57	16. 42
最大拡散規模	23. 45	55. 75	35. 56	68. 44

Twitter はブログに比較して拡散回数が多い傾向がある。また、拡散規模は平均すると顕著な差異はみられないが、最大拡散規模は Twitter のほうがブログに比較して大きい傾向がみられる。即ち、Twitter のほうがネット上の拡散力が強いことを示している。また、比較対象キーワードでは、「ワタミ」の累計拡散規模、平均拡散規模、最大拡散規模すべてにおいて Twitter、ブログともに、他キーワードと比較して、最大値となっている。2012 年の社員自殺の労災認定に端を発し、何か話題が出る度に、企業に対する批判意見が拡散する状態が定着している状況が伺える。

3 まとめ

本稿では、クチコミ拡散状況の定量化を実現し、メディア、キーワードごとのクチコミの拡散特性を比較することが可能になった。クチコミの拡散傾向は、その原因が、好意的な意見か否定的な意見か、あるいは過去からの評判の蓄積によって異なると考えられる。今後、これらをより詳細に分析し、企業活動により有益な結果を提示してゆきたい。

最後に、クチコミ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.