

# SNS における影響力と SNS 外部における影響力の関係

## Relation between the influence on SNS and the influence outside of SNS

12N8100003B 情報工学専攻 梅田 博之 UMEDA Hiroyuki

### 1. はじめに

人の影響力というものは多種多様な場面において効果を及ぼす。例えば職場の人間関係、選挙の結果、広告の効果、近所の人々とのつきあいなどが考えられる [8]。さらに近年では SNS という新しい人との交流媒体が出現した。この新しい媒体の出現により、人同士の交流の形態が変化し、人の影響力も形を変えつつある。実際に SNS が商品の売上に影響したり [9]、消費者の行動に影響したりということが報告されていて [10]、その影響は多大なものであることがわかる。

従って SNS において効率的に宣伝等を行いたいと考えるのは自然な考えであり、その要望に従って影響力最大化問題というものがここ 10 年ほど盛んに研究が行われ、性能の良いアルゴリズムも提案されてきている [6]。

しかし、ここで一つ問題がある。それはある人からある人へ影響が伝搬する確率を所与のものであると仮定している論文が、大半を占めているということである [6]。この問題を解決するために、影響力の経験的研究が行われてきている。それらの研究のほとんどが SNS に対する研究である。これは人の交友関係がある程度可視化できるためである。

例えば [1] では、ツイッターにおける影響力をフォロワー数、コメント数、リツイート数という 3 つの尺度で考えている。

SNS での影響力測定で問題になるのは、観測不可能な要因である。たとえば、[2] は観測不可能な要因は多大な影響を及ぼしているという結論を下している。観測不可能な要因の一つとして考えられるのは、外部メディアの影響である。具体的には、有名人は SNS においてほとんど活動をしていなくても、多大な影響を SNS に対し行使できる [3]。

そこで本研究では外部的な影響の例として TV を取り上げ、SNS の例としてツイッターを取り上げる。そしてツイッターにおける影響力と TV における影響力には関係があるのか？という疑問に答える。驚くべきことに二つの媒体における影響

力はほとんど関係がなく、独立な変数として振る舞うことが分かった。

### 2. データ概要

本研究で用いるデータは株式会社オモシロから提供して頂いたデータである。このデータには 3 つのテーブルがある。以下の表はその 3 つのテーブルのテーブル名とそれぞれが持つカラムを表したものである。

表 1: データテーブル一覧

Tv person	Tv daily perform	Tv tweet speed
氏名	出演番組の年月日	ツイート速度
ふりがな	1 日の出演時間	取得時刻
性別	1 日の出演回数	
職業・肩書き		
職業属性		

まず Tv person テーブルのカラムについて述べる。このテーブルは NHK 総合、NHK 教育、TBS TV、TV 朝日、フジ TV、日本 TV、TV 東京に 2008 年 1 月から 2013 年 11 月 30 日までに出演した 37614 人のデータである。性別は 0 か 1 の変数で与えられ、職業・肩書きは文字列データで与えられる。職業属性とは、俳優、芸人、歌手・音楽関係、アイドル、スポーツ選手、アナウンサー、製作者、声優、タレント、ファッションモデルの 10 種類があり、各人物に与えられる。また複数の属性を一人の人物が同時に兼ねることも許されている (たとえば、歌手・音楽関係とアイドルなど)。

次に Tv daily perform テーブルについて述べる。このテーブルは、Tv person テーブルと同条件の 37614 人に対するデータで、その人達の TV 出演頻度に関するテーブルである。データの収集は番組表を用いて行っている。1 日の出演時間、回数とは、出演番組の年月日に記載された日付に出演した時間、回数である。また番組内の一部の時間しか出演していなくても、番組放送時間全体に渡って出演したと考えて集計を行っている。出演時間、出演回数は時間帯毎にも集計を行っている。具体的にはゴールデンタイム (18:50~21:50)、プライ

ムタイム（18:50～22:50）、深夜（0:00～3:50）時間帯における出演時間、出演回数のデータもテーブルに含まれているが、本論文の分析では用いていないため、省略する。

最後の Tv tweet speed テーブルは、データを取得した時刻、ツイート速度が与えられている。データの取得方法はまず、人物の氏名でツイッターを検索する。そして最新 30 件のツイートを取得し、30 件のツイートが投稿されるのにかかった時間を算出する。それを 1 時間当たりのツイート速度に直した値がツイート速度である。取得期間は 2013 年 6 月から 2013 年 11 月 30 日までである。また全ての人物に対し取得が成功したわけではなく、成功した人数は 23815 人である。またその 23815 人も定期的にツイートを取得しているわけではなく、人物毎に取得頻度はばらつきがある。

以上 3 つが本研究で用いたデータである。

### 3. 分析概要

まず人物の集合を  $N = \{1, 2, \dots, n\}$  とする。また時刻  $T = (\{1, 2, \dots, t\}, \leq)$  は順序集合である。また、人物  $i$  の時刻  $j$  におけるツイート速度を  $tw_{ij}$ 、出演時間を  $ap_{ij}$  とする。 $tw_{ij}$  はテーブル Tv tweet speed から、 $ap_{ij}$  はテーブル Tv daily perform から得ることができる。生データのままであるデータで、分析に適さないため、今回は 1 週間毎の値に修正して使用している。

$tw_{ij}$  と  $ap_{ij}$  を一組として考え、 $(tw_{ij}, ap_{ij})$  とかくこともある。

まず、 $tw_{ij}, ap_{ij}$  のヒストグラムを作成する。そのためにビンで区切る必要がある。ここでは平方根選択を行う。 $max_{tw} = \max_{i,j}(tw_{ij})$ ,  $max_{ap} = \max_{i,j}(ap_{ij})$  とし、 $max_x = \lceil \sqrt{max_{tw}} \rceil$ ,  $max_y = \lceil \sqrt{max_{ap}} \rceil$  とすれば、ビンの集合は

$$X \times Y, X = \{1, 2, \dots, max_x\}, Y = \{1, 2, \dots, max_y\} \quad (1)$$

となる。ここでビン  $(x, y) \in X \times Y$  に関する定義する。まず、ビン  $(x, y)$  に  $(tw_{ij}, ap_{ij})$  が含まれることを  $(tw_{ij}, ap_{ij}) \in (x, y)$  とかき、

$$max_{tw} * \frac{x-1}{max_x} \leq tw_{ij} \leq max_{tw} * \frac{x}{max_x} \quad (2)$$

$$max_{ap} * \frac{y-1}{max_y} \leq ap_{ij} \leq max_{ap} * \frac{y}{max_y} \quad (3)$$

が同時に成立するとき、 $(tw_{ij}, ap_{ij}) \in (x, y)$  である。またビン  $(x, y)$  に含まれる  $(tw_{ij}, ap_{ij})$  の数を

度数といい、 $Deg(x, y)$  で表す。さらに、 $Tw(x, y)$  をビン  $(x, y)$  に含まれる全ての  $tw_{ij} \in (tw_{ij}, ap_{ij})$  の平均値とする。また同様に  $Ap(x, y)$  をビン  $(x, y)$  に含まれる全ての  $ap_{ij} \in (tw_{ij}, ap_{ij})$  の平均値とする。すなわち、 $Tw(x, y), Ap(x, y)$  は集合  $X \times Y$  に関する関数である。

最後に、 $Tw(x, y), Ap(x, y)$  に関するヒストグラムを以下に示す

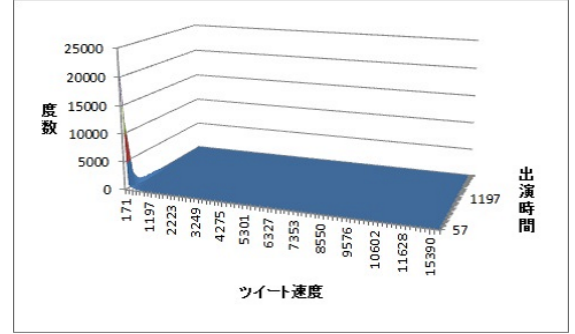


図 1: ヒストグラム

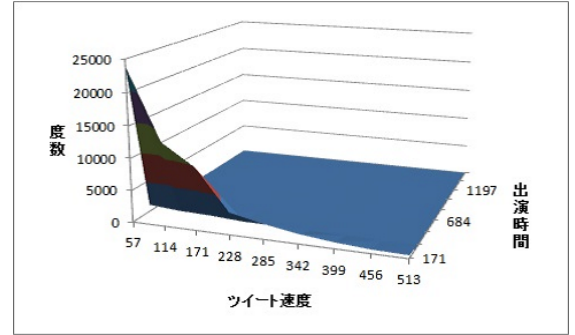


図 2: 拡大図

図からべき乗則に従うことが推察できる。従って次のような分析を行う。

#### 3.1. 重回帰分析

$(Tw(x, y), Ap(x, y), Deg(x, y)), \forall (x, y) \in X \times Y$  を重回帰分析する。ここでは全ての変数に対数をとったもの、すなわち

$(\log_{10} Tw(x, y), \log_{10} Ap(x, y), \log_{10} Deg(x, y))$  に対し重回帰分析を行う。結果は次のようになった。

表 2: 相関行列

	出演時間	ツイート速度	度数
出演時間	1.000	-0.343	-0.278
ツイート速度	-0.343	1.000	-0.614
度数	-0.278	-0.614	1.000

まず相関行列を見ると、この結果を見てみると、ツイート速度と度数に最も強い負の相関が存在す

ることがわかる。また、出演時間と度数にも弱い負の相関がある。興味深いのはツイート速度と出演時間には弱い負の相関がみられることである。

表 3: R 値

R	修正 R
0.804	0.803

表 4: 変数の有意性

変 数	出演時間	ツイート速度	定数項
偏 F 値	382.348	806.470	1024.845
t 値	-19.553	-28.398	32.013
P 値	0.000	0.000	0.000

表 5: 回帰式の有意性

F 値	P 値
457.923	0.000

上の 3 つの表はそれぞれ R 値と変数の有意性に関する表である。まず修正 R 値が 0.8037 であることから、当てはまりが良いことが伺える。また、偏 F 値、t 値ともに良好であり、有意水準 5% で帰無仮説を棄却できている。P 値が非常に低いことからそのことを伺うことができる。さらに、回帰式の有意性の表を見ると、F 値は良好でそれによる P 値はほぼ 0 であることから、回帰式は有意であるといえる。

表 6: 回帰係数

変 数	出演時間	ツイート速度	定数項
偏回帰係数	-0.906	-0.889	5.721
標準誤差	0.046	0.031	0.178
下限値	-0.997	-0.950	5.370
上限値	-0.815	-0.827	6.072
標準偏回帰	-0.553	-0.803	
単相関	-0.278	-0.614	
偏相関	-0.658	-0.785	

表 7: 変数の共線性

	出演時間	ツイート速度
トレランス	0.882	0.882
VIF	1.132	1.132

最後の表は重回帰分析の偏回帰係数に関するもの、共線性に関するものの結果である。偏回帰係数を見てみると、出演時間は-0.9063、ツイート速度は-0.8891 となっている。また偏回帰係数の 95%信

頼区間を見ると、出演時間の方は下限が-0.9974、上限が-0.8152 となっている。またツイート速度の方は下限が-0.9506、上限が-0.8276 となっており、いずれも負の値である。標準偏回帰係数を見ると、出演時間の方は-0.5533、ツイート速度の方は-0.8036 であり、ツイート速度の方が度数への影響が高い。また共線性の値も良好でツイート速度と出演時間の間に共線性がないことがわかる。つまり、相関がないといえる。

以上の事実から以下の回帰式に従うと結論付けることができる。

$$\log_{10} Deg(x, y) = -0.889 \log_{10} Tw(x, y) - 0.906 \log_{10} Ap(x, y) + 5.721 \quad (4)$$

### 3.2. グレンジャー因果性検定

3.1 では時刻の影響は消えていた。そのため時刻による影響を考慮した分析を行う。

$tw_{ij}, ap_{ij}$  に対し、全ての人物の平均値を代表値とする。具体的には、 $Atw_j = ave(tw_{ij}, \forall i)$ ,  $Aap_j = ave(ap_{ij}, \forall i)$  と定義した  $Atw_j, Aap_j$  に対してグレンジャー因果性検定を行う。

この検定を行うには VAR モデルで記述する必要があるが、そもそもこのモデルで記述するためには定常性が必要である。

そのため、定常性の検定を行うが、 $Atw_j, Aap_j$  は定常と判定されなかったため、各データ  $Atw_j, Aap_j$  の差分系列、 $diffAtw_j = Atw_{j+1} - Atw_j$ ,  $diffAap_j = Aap_{j+1} - Aap_j$  に対してそれぞれ PP 検定を行った。その結果は以下ようになった。

表 8: PP 検定

	統計量	P 値
$diffAap$	-4.8165	0.01
$diffAtw$	-6.8214	0.01

P 値が 0.01 なので有意水準 5% ならば、定常性があるといえる。

次に AIC によって VAR モデルの次数を決定する。AIC の値は以下ようになった。

表 9: AIC

次数	1	2	3	4	5	6
AIC(n)	8.344	7.995	8.010	8.195	8.152	8.132

この結果から、次数は AIC が極小な 2 を採用する。

そして最後にグレンジャー因果性検定を行い、時差相関の検定を行った。以下が結果である。

表 10: グレンジャー因果

因果の方向	統計量	P 値
$diffAap \rightarrow diffAtw$	0.031	0.969
$diffAap \leftarrow diffAtw$	2.643	0.0856

上の結果を見ると、帰無仮説は有意水準 5% で棄却されることがわかる。このことからグレンジャー因果性が存在しないと結論づけることができる。

#### 4. おわりに

本論文では TV における出演時間と、ツイッターにおけるツイート速度の関係をマクロ的に解析した。

そこで分かったことは、TV の出演時間とツイート速度、言い換えると TV における影響力とツイッターにおける影響力は関係がないということである。

まず、重回帰分析でわかったことは、TV の出演時間とツイート速度は度数に対して独立にべき乗則を示すということである。

また、グレンジャー因果性検定により、ツイート速度と TV 出演時間は因果がないという結果が出たため、時差を考慮しても因果がないことが分かった。従って、この二つの変数は驚くべきことに、ほとんど関係をもたない分る。

#### 5. 謝辞

本研究で使用したデータを提供して頂いた株式会社オモシロ様に感謝申し上げます。

#### 参考文献

- [1] Cha, Meeyoung, et al. “Measuring user influence in Twitter: The million follower fallacy.” ICWSM 10 (2010): 10-17.
- [2] Fang, Xiao, et al. “Predicting adoption probabilities in social networks.” Information Systems Research 24.1 (2013): 128-145.
- [3] Myers, Seth A., Chenguang Zhu, and Jure Leskovec. “Information diffusion and external influence in networks.” Proceedings of

the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012.

- [4] Kempe, David, Jon Kleinberg, and Eva Tardos. “Maximizing the spread of influence through a social network.” Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2003.
- [5] J Scott, PJ Carrington and eds. “The SAGE handbook of social network analysis.” SAGE publications, 2011.
- [6] Wang, Chi, Wei Chen, and Yajun Wang. “Scalable influence maximization for independent cascade model in large-scale social networks.” Data Mining and Knowledge Discovery 25.3 (2012): 545-576.
- [7] Weng, Jianshu, et al. “Twiterrank: finding topic-sensitive influential Twitterers.” Proceedings of the third ACM International Conference on Web Search and Data Mining. ACM, 2010.
- [8] Easley, David, and Jon Kleinberg. “Networks, Crowds, and Markets.” Cambridge University Press, 2010.
- [9] 奥田輔, 安田孝美, 水野政司. “ソーシャルメディアにおける情報の伝播がもたらす商品の売れ行きに対する影響について: ツイッターのつぶやき数の推移と iTunes AppStore のモバイルアプリの売れ行きの相関性.” 電子情報通信学会技術研究報告. SS, ソフトウェアサイエンス 111.268 (2011): 1-6.
- [10] 清水麻衣. “CGM が消費者の購買意思決定プロセスに及ぼす影響: 消費者発信情報と企業発信情報の比較.” 商学論集, Vol. 81, No. 3, (2013) 93-121.
- [11] 勝又他, “タレント属性の TV 広告効果測定ータレントイメージと製品カテゴリーとの適合関係分析ー” オペレーションズ・リサーチ, Vol. 59, No. 1, (2014) 42-51.
- [12] 沖本竜義. “経済・ファイナンスデータの計量時系列分析.” 朝倉書店, (2010).