

社会的な出来事がツイッターへの投稿に与える影響を分析するソフトウェア開発

呉 思奇†

華山 宣胤‡

尚美学園大学芸術情報学部†

尚美学園大学芸術情報学部‡

1. はじめに

日本国内で最も一般的な SNS であるツイッターのユーザーは、ツイートと呼ばれる 280 文字以内の文章を何時でも投稿することが出来る。このため、ある期間に投稿されたツイートの内容の全体は人々（ユーザー）のその期間の関心事を表していると考えられる。

「人々が何に関心を持っているか」だけではなく「ある物事（キーワード）に対する人々の関心がどの出来事に影響されているか」を把握することもできる。例えば、台風による「豪雨災害」の直後に「防災グッズ」というキーワードを含むツイートやリツイートが増えれば「豪雨災害によって人々の防災グッズへの関心が高められた」と考えられる。

そこで本研究では、あるキーワードを含むツイートとリツイートの投稿数を集計し、そこからキーワードに対する人々の関心がどの出来事にどの程度影響されたかを把握するためのソフトウェアを開発し、実データを収集・集計した結果を示す。

2. Tweet データの集計方法

ある出来事や事件が起きると、その後（直後）には、その出来事や事件に関係するメッセージ（ツイート）のツイッターへの投稿が増えると考えられる。しかし、リツイート（共感を持ったツイートのコピーの投稿）は、元のツイートをユーザーが読んだ後に投稿されるため、元のツイートの投稿より遅れて投稿されるため、その出来事や事件が起きてから時間が経ってから投稿数が増えると考えられる。このためリツイート数の増減は、元のツイートが投稿された時点に起きた出来事や事件とリツイートが投稿された時点に起きた出来事や事件の両方に影響されると考えることが出来るであろう。

このような考え方を説明／把握するために、本研究では、ツイートを「投稿された時点に生まれて、リツイートにより延命される生命体」

と考え、レクシス図[1]のアイデアを応用してツイートとリツイートを図1のように表す（把握すること）を提案する。図の中では、ツイートを○で、そのリツイートを●で表し、斜め45度の直線で結ばれた元のツイートとリツイートを表している。

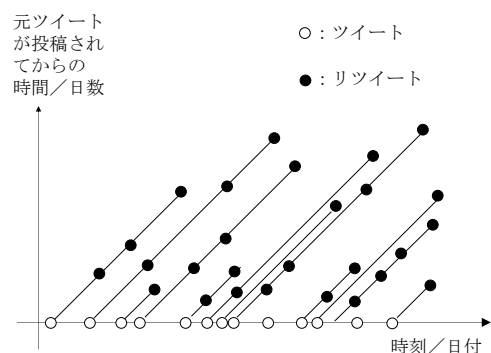


図1：投稿時刻・元ツイートが投稿されてから時間を座標軸としたツイートとリツイートのプロットイメージ図

実際のツイート・データは Python により構築したソフトウェアを用いてツイッターAPI を介して収集し、適当な時間区分（図2の平行四辺形の区分）で集計する。そして、集計されたデータについて「投稿時刻別ツイート数」を対応する時間帯に起きた出来事や事件の「即時効果」, 「元ツイートが投稿された時刻別のリツイート数」を「漸次効果」, 「元ツイートが投稿されてからの経過時間別リツイート数」を「経時効果」を表すと考える。

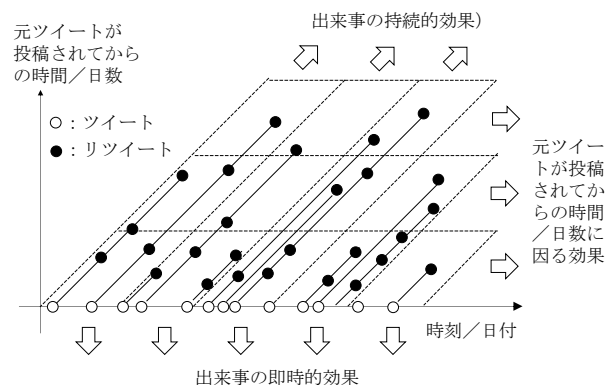


図2：出来事や事件の「即時効果」「漸次効果」「経時効果」のイメージ図

A tool for estimating effects of socially big incidents on numbers of submissions of message to Twitter

†Wu Siqi · Shobi University Faculty of Art and Information

‡Nobutane Hanayama · Shobi University Faculty of Art and Information

3. “今年の漢字”を含むツイートの集計結果

表1に、2020年12月25日6時0分～15時35分に“今年の漢字”というキーワードで検索・収集したツイートとリツイート数を、図2に示した方法で集計した結果を示す。

表1：“今年の漢字”で検索・収集し集計したツイートとリツイート数

投稿時刻 別ツイ ート数	リツイート数									
	0~9	1~9	2~9	3~9	4~9	5~9	6~9	7~9	8~9	
	時間 未 満	時間 未 満	時間 未 満	時間 未 満	時間 未 満	時間 未 満	時間 未 満	時間 未 満	時間 未 満	
6時	38	4	3	3	3	2	2	2	1	1
7時	76	14	9	7	7	7	3	3	2	0
8時	66	3	1	0	0	0	0	0	0	
9時	66	3	2	2	0	0	0	0		
10時	72	3	1	1	1	0	0			
11時	72	79	33	12	8	1				
12時	149	9	2	2	0					
13時	103	1	0	0						
14時	78	5	0							
15時	52	4								
元ツイートが投稿されてからの経過時間別リツイート数	125	51	27	19	10	5	5	3	1	

図3に“即時効果”を示す。図から、12時台にピークがあることが分かる。これは、この日の午前中に岸防衛大臣が「今年の漢字」として「異」を選んだことがお昼のニュースで報道されたことが影響していると考えられる。

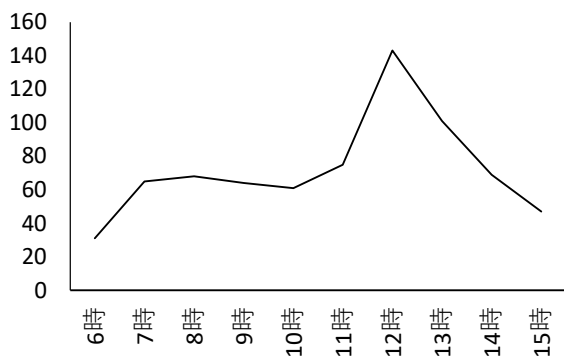


図3：“今年の漢字”をキーワードとして検索されたツイートに関する“即時効果”

図4に“漸次効果”を示す。図から、11時台にピークがあることが分かる。これは、昼休み中にツイッターがチェックされリツイートされるため、元のツイートは昼休みの直前に投稿されたものが多いためであると考えられる。

図5“経時効果”をグラフで表したものである。図から、元ツイートが投稿された1時間後から、減少が緩やかになることが分かる。

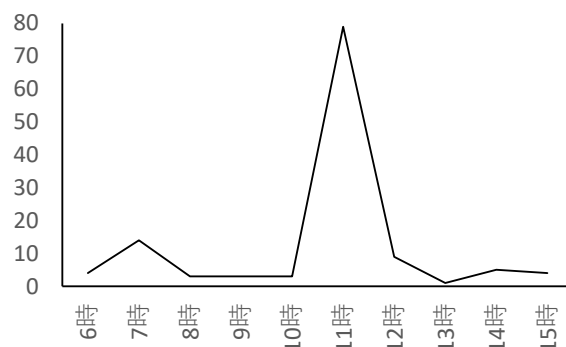


図4：“今年の漢字”をキーワードとして検索されたツイートに関する“漸次効果”

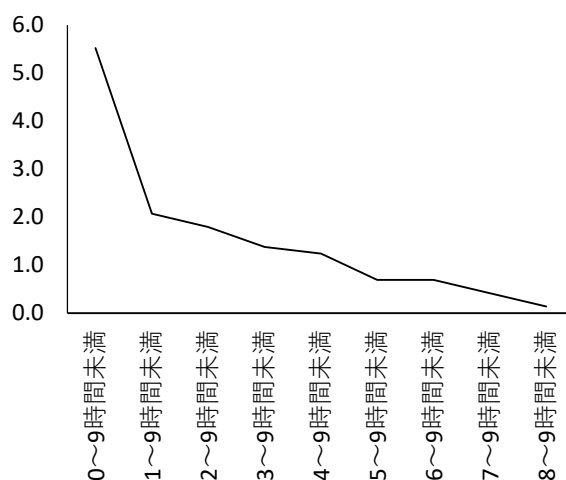


図5：“今年の漢字”をキーワードとして検索されたツイートに関する経時効果

4. 議論

本報告ではツイート数とリツイートの集計方法の提案を行った。今後は、統計モデルを用いて集計したデータを分析する手順について検討を行うことを課題である。

リツイート行動に関する研究[2]からは「過去のリツイートの量が現在のリツイート行動の影響する」という結論が得られている。このようなリツイートに関する研究結果を本研究に应用することにより、いわゆる「炎上」や「バズり」といった現象と社会的な出来事の関係の分析に発展させることが出来る可能性がある。

<参考文献>

[1] Keiding, N. (1990). 'Statistical inference in the Lexis diagram'. Philosophical Transactions of the Royal Society of London Series A 332, 487-509.
 [2] 塩田 茂雄, 中島 圭佑. Twitter データに見られる特徴と人間のリツイート行動. 人工知能学会全国大会論文集, 2019, JSAI2019 巻, 第33回全国大会(2019).