

On the Origins of Memes by Means of Fringe Web Communities

Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro,
Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil.

情報工学科 4 年 寺岡研究室
61619027 安森 涼

2019 年 7 月 27 日

1 背景

Webの台頭により考えや文化、画像、映像などが今までにない速度で拡散されるようになった。その中でも政治思想や心情の拡散にミーム (meme) というインターネットで流行する画像や言い回しが利用されている。その中でも人種差別的、政治的意味をもつ攻撃的なミームの特定の需要が増加している。そこで、SNS ごとにミームの分布、拡散の傾向の解釈が必要となっている。本論文では画像ミームのみを対象に分析している。

2 関連研究

ミームの分析として様々な研究が行われている。例えば、複数の Twitter のハッシュタグからのミームの人気度調査 [1], Facebook 間の伝搬の過程で新しいミームが生まれることの証明 [2], 4chan, Reddit で口汚い言葉を用いた投稿の検知 [3] などが行われている。しかしながらこれらの事前研究では 2 点実施されていないことが 2 点存在する。まず, 1 点目として, 検証対象外の SNS データセットへの適用が不可能である点である。具体的に言うと, Twitter に対して調査を行った手法は, Facebook などの他の SNS では利用ができないという意味である。つまり, 複数 SNS のミームを意味づけすることが必要となっている。これをハッシュ化した画像のクラスタリングを実施して解決へと導いている。2 点目がミームの発生元となる SNS が調査されていないという点である。具体的に言うと, Twitter で拡散されているミームがどの SNS で初めて観測されたかなどの調査はしていないと言う点である。これはつまり, 複数 SNS 間のミームの伝搬を検知する必要があると言うことで, 本論文では伝搬性を持つ事象を表現できるモデルを作成することで解決に導いている。

3 データセット

利用するデータセットとして2つの種類のデータセットが存在する。1つ目が調査対象の SNS で、2つ目がミームのまとめサイトである。前者から分析対象のミームが存在する SNS、後者からミームにメタデータを結びつけるために利用するデータを取得する。まず調査対象の SNS について、2 項目の分類が存在する。それぞれ、Mainstream コミュニティ (以後 Mainstream と記述)、Fringe コミュニティ (以後 Fringe と記述) である。それぞれ、悪意のないミームが拡散されやすい SNS、悪意を含むミームが拡散されやすいと考えられる SNS を示している。ここで言う悪意のあるという言葉はヘイトや人種差別のことを指す。Mainstream として Twitter と Reddit という SNS を利用する。Twitter は 140 字以内の短い記事を投稿し合うサイト、Reddit はニュース記事、画像やテキストの投稿サイトを示す。Fringe としては The_Donald (T.D)、/pol/, Gab という SNS を利用し、それぞれ、Reddit のトランプについてのチャンネル、匿名の掲示板 4chan の政治チャンネル、ほぼ規制がない、言論の自由を尊重するサイトを示す。ミームのまとめサイトとしては Know Your Meme (KYM) と言うものを利用する。これはミームに対するメタデータ (由来、キーワードタグ、説明、例、イメージギャラリー) を供給するものである。

Fringe				Mainstream			
Spot			Cost	Spot			Cost
Entry	Points (C)	Entry	Points (C)	Entry	Points(C)	Entry	Points (C)
1	100	1	100	1	100	1	100
2	90	2	90	2	90	2	90
3	80	3	80	3	80	3	80
4	70	4	70	4	70	4	70
5	60	5	60	5	60	5	60
6	50	6	50	6	50	6	50
7	40	7	40	7	40	7	40
8	30	8	30	8	30	8	30
9	20	9	20	9	20	9	20
10	10	10	10	10	10	10	10
11	0	11	0	11	0	11	0
12	0	12	0	12	0	12	0
13	0	13	0	13	0	13	0
14	0	14	0	14	0	14	0
15	0	15	0	15	0	15	0
16	0	16	0	16	0	16	0
17	0	17	0	17	0	17	0
18	0	18	0	18	0	18	0
19	0	19	0	19	0	19	0
20	0	20	0	20	0	20	0
21	0	21	0	21	0	21	0
22	0	22	0	22	0	22	0
23	0	23	0	23	0	23	0
24	0	24	0	24	0	24	0
25	0	25	0	25	0	25	0
26	0	26	0	26	0	26	0
27	0	27	0	27	0	27	0
28	0	28	0	28	0	28	0
29	0	29	0	29	0	29	0
30	0	30	0	30	0	30	0
31	0	31	0	31	0	31	0
32	0	32	0	32	0	32	0
33	0	33	0	33	0	33	0
34	0	34	0	34	0	34	0
35	0	35	0	35	0	35	0
36	0	36	0	36	0	36	0
37	0	37	0	37	0	37	0
38	0	38	0	38	0	38	0
39	0	39	0	39	0	39	0
40	0	40	0	40	0	40	0
41	0	41	0	41	0	41	0
42	0	42	0	42	0	42	0
43	0	43	0	43	0	43	0
44	0	44	0	44	0	44	0
45	0	45	0	45	0	45	0
46	0	46	0	46	0	46	0
47	0	47	0	47	0	47	0
48	0	48	0	48	0	48	0
49	0	49	0	49	0	49	0
50	0	50	0	50	0	50	0
51	0	51	0	51	0	51	0
52	0	52	0	52	0	52	0
53	0	53	0	53	0	53	0
54	0	54	0	54	0	54	0
55	0	55	0	55	0	55	0
56	0	56	0	56	0	56	0
57	0	57	0	57	0	57	0
58	0	58	0	58	0	58	0
59	0	59	0				

図1 SNSごとに投稿されるミームの割合

4 提案手法 1 | 複数 SNS のミームの意味付け, 評価

ある要素の出力が次の要素の入力となるような処理を繰り返すことで複数 SNS のミームに意味付けを実施する。まずミームの一部を似た画像に近い値を保持するという特徴をもつ 64bit ハッシュ値を用いて、ハッシュ化をする。続いて、ハミング距離準拠のアルゴリズムを利用してクラスタリングを行い、クラスタごとに Medoid というクラスタ内の自身以外の画像との距離の総和が最小となる、クラスタを代表する画像を決定する。続いて、ハッシュ化した KYM データを利用し、クラスタとメタデータを結びつけ、その結果を他のデータセットへ適用する。この処理の入力が画像のため任意の SNS に対応可能であり、前述の検証対象外の SNS データセットへの適用が不可能という問題を解決している。

4.1 SNS ごとに投稿されるミームの割合

前述の処理を行った後に出力された結果を元に分析を行い、出力された結果が図 1 である。この図において、左 2 つ (/pol/, Gab) が Fringe を、右 2 つ (Twitter, Reddit) が Mainstream を示している。これを見ると人種差別的ミーム (図中赤) が Fringe で上位であるのに対して Mainstream では悪意のないミームが上位を占めている。ここで述べた人種差別的というのは 'racist', 'antisemitism' (反ユダヤ主義) などのタグを持つものである。一方で政治的なミームは Mainstream, Fringe 問わずどこにでも存在していることがわかる。

4.2 ミームの投稿数の推移

人種差別的ミームおよび政治的ミームの投稿数の推移はそれぞれ図 2, 図 3 で表される。共に 2017 年に Gab の投稿が増加していることが観察できる。これは Gab が 2016 年後期にサービスを開始し、2017 年に徐々に普及してきたことが原因として考えられる。続いて、人種差別的ミームについて観察すると、Mainstream ではほとんど投稿がないのに対し、/pol/(Fringe) では継続的に投稿があることがわかる。一方で政治的ミームに関しては Mainstream, Fringe に関係なく全てのコミュニティで継続的な投稿があることがわかる。これは 4.1 節での評価の裏付けとなっている。また、政治的ミームについては第 2 回 US 大統領選挙討論会や 2016 US 大統領選挙において投稿数が伸びていることから政治的ミームの投稿数推移は実世界の出来事と密接に関係していることがわかる。

5 提案手法 2 | 複数 SNS 間のミームの伝搬を検知

複数 SNS 間のミームの伝搬の検知にはミームの投稿が起因する SNS を調査できるモデルを作成し利用することで解決に導いてい

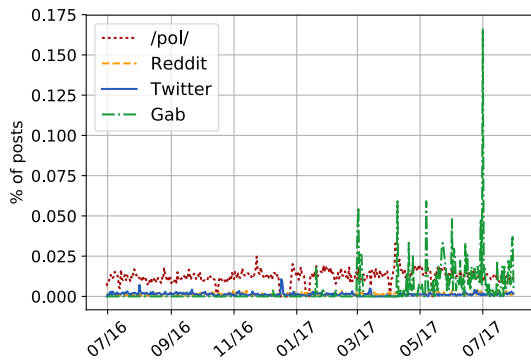


図2 人種差別的ミームの投稿数の推移

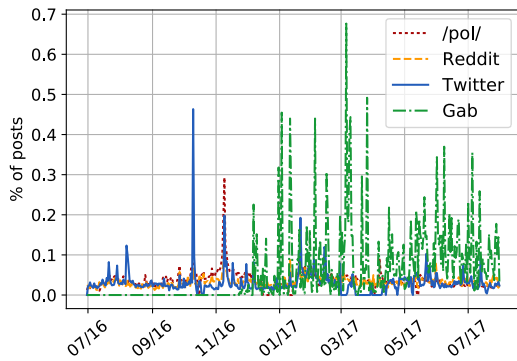


図3 政治的ミームの投稿数の推移

る。本項では説明のため SNS B, C, A に順にミーム 1, 2, 3 の投稿があった場合にミーム 2 の発信元 SNS を特定することを考える。図 4 を元に説明する。まず各投稿発生時にポール 1, 2, 3 を立て、各投稿発生直後全ての SNS に波を描画する。そして各ポール時点での発生確率を「投稿の発生要因」として抽出し、「投稿の発生要因」の 2 について観察する。観察を行うと「投稿の発生要因」の 2 の紅色で表されている部分はポール 1 によって発生した波に起因する部分で、緑色で表されている部分は SNS C に起因していることがわかる。紅色で表される部分は SNS B に 100% 起因している部分であるから、紅色を橙色に置換した結果を投稿が起因する SNS として表示する。これによりミーム 2 の発信元 SNS が B, C と判明する。同様の手順でミーム 3 についても投稿が起因する SNS を求めることが可能である。

5.1 拡散された人種差別的ミームの割合

前述の手順で処理を行った後に出力された結果を元に拡散された人種差別的ミームを求めたものが図 5 である。図の縦はミームの拡散元を、横はミームの拡散先を、図は拡散されたミームの割合を示している。図の R と NR は人種差別的、非人種差別的なミームを示している。この図を観察すると、/pol/ が拡散元の時他の SNS への影響力が高いことがわかる。つまり他のコミュニティに広まったミームに /pol/ 発祥のものが多くことがわかる。

5.2 人種差別的ミームが拡散される確率

図 6 は 5.1 の図について送信元の SNS で正規化したものである。この図より T.D が送信元の時に拡散される確率が高いことから、T.D にミームを投稿すると他の SNS に広がる可能性が高いことがわかる。一方で、影響力が最大であった /pol/ は拡散される確率という面では最悪で、投稿のうち一流のものしか残らないことがわかる。政治的ミームについても本論文で言及があったが少しの変化しかなかったため割愛する。

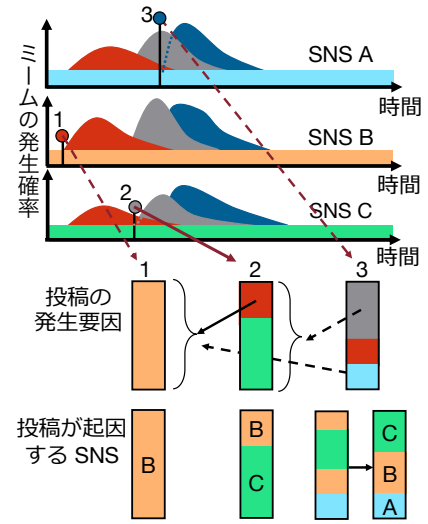


図4 政治的ミームの投稿数の推移

	R : 6.36% NR : 3.86%	R : 4.31% NR : 3.12%	R : 18.83% NR : 13.08%	R : 15.04% NR : 16.35%
R : 0.35% NR : 1.25%		R : 2.48% NR : 4.79%	R : 1.29% NR : 9.01%	R : 9.52% NR : 8.89%
R : 0.20% NR : 0.97%	R : 2.22% NR : 3.49%		R : 1.26% NR : 9.21%	R : 2.10% NR : 5.08%
R : 0.05% NR : 0.09%	R : 0.54% NR : 0.15%	R : 0.54% NR : 0.15%		R : 0.22% NR : 0.56%
R : 0.06% NR : 0.73%	R : 1.77% NR : 2.11%	R : 1.77% NR : 2.11%	R : 2.54% NR : 9.30%	

図5 拡散された人種差別的ミームの割合

	R : 0.4% NR : 1.5%	R : 0.3% NR : 1.8%	R : 0.2% NR : 0.4%	R : 0.2% NR : 0.9%	R : 1.1% NR : 4.5%
R : 5.1% NR : 3.3%		R : 2.9% NR : 7.1%	R : 0.2% NR : 0.7%	R : 1.4% NR : 1.3%	R : 9.5% NR : 12.4%
R : 2.4% NR : 1.7%	R : 1.9% NR : 2.3%		R : 0.1% NR : 0.5%	R : 0.3% NR : 0.5%	R : 4.7% NR : 5.0%
R : 5.3% NR : 3.0%	R : 4.0% NR : 1.9%	R : 0.5% NR : 3.1%		R : 0.2% NR : 1.0%	R : 10.0% NR : 9.1%
R : 6.3% NR : 13.6%	R : 12.2% NR : 15.0%	R : 2.5% NR : 12.6%	R : 2.3% NR : 5.1%		R : 23.3% NR : 46.2%

図6 人種差別的ミームが拡散される確率

6 結論

政治思想や信条の拡散にミームが利用され、その中でも人種差別的、政治的な意味を含む攻撃的なミームの特定の需要が増えている。そこで、SNS ごとにミームの分布や拡散の傾向の解釈が必要であると考えられている。本論文では複数 SNS のミームの意味付けおよび複数 SNS 間のミームの伝搬を検知している。結果として人種差別的ミームが Fringe に多く投稿されている点、政治的ミームが普遍的に存在している点、実世界の出来事は政治的ミームの投稿数に影響する点、/pol/ の他の SNS への影響力が最大である点、T.D にミームを投稿すると他の SNS に広がる可能性が高い点などが判明した。

7 参考文献

- [1] L. Weng, F. Menczer, and Y.-Y. Ahn. Predicting Successful Memes Using Network and Community Structure. In ICWSM, 2014.
- [2] L. A. Adamic, T. M. Lento, E. Adar, and P. C. Ng. Information Evolution in Social Networks. In WSDM, 2016.
- [3] E. Chandrasekharan, et al., The bag of communities: Identifying abusive behavior online with preexisting Internet data. In CHI, 2017.