

Twitterにおけるリンク構造を利用したスパムアカウント抽出手法の検討 Extraction of Spam Accounts by Use of the Link Structure in Twitter

菊池望十 吉村博幸
Nozomi Kikuchi Hiroyuki Yoshimura

あらまし 近年、スマートフォン等の普及に伴い、ソーシャルネットワークサービスの利用者は増え、誰もが気軽にアカウントを所有するようになってきている。しかしその中には、個人情報取得や、有害なサイトへの誘導、デマ等、利用者にとって不利益となる情報の配信を狙って情報の拡散を行う、スパムアカウントも存在している。また、一般ユーザーを装い、情報を拡散させるための手口も巧妙化している。そこで本稿では、Twitterに着目してスパムアカウントの特徴を考察し、フォロー・フォロワー情報等の定量的なデータと投稿内容の文字データを用いて、スパムアカウントの抽出が可能であるか検証および考察を行った。

1. はじめに

現在、スマートフォン等の普及により、ソーシャルネットワークサービス(以下 SNS)も手軽に利用されるようになってきている。SNSをはじめとするソーシャルメディアは、若年層の利用も多い。中でも Twitter[1]は、他の SNS と比べて登録時に電話番号等の個人確認がないため、利用が手軽な反面、スパムアカウントを作りやすい。また拡散力があるため、不快な投稿を目にする機会も存在する。そこで、スパムアカウントを効率的に発見・抽出できる方法を見つけ出すことを目標に、本論文では、スパムアカウントとのつながりが大きいアカウントと、通常アカウントとのつながりが大きいアカウントについて、スパム含有率を調査分析した。

2. 関連研究

Twitterに関する研究は数多く行われている。特に、Twitterにおけるスパムアカウントを分別する研究として、分類器を用いて機械学習を行う方法[2]、投稿時のクライアント名や自己紹介文といったユーザー情報とフォローとフォロワー数の情報により判定する方法[3]等がある。

また、同じくソーシャルメディアとして挙げられるブログにおけるスパムの特徴を挙げたもの[4]、アフィリエイトとスパムブログについての関係性が述べられているもの[5]等がある。

3. スパムアカウントについて

今回着目する「スパムアカウント」の定義を、1. アフィリエイトや有害なサイトへの誘導を目的としたアカウント、および 2. フォロワー稼ぎや、RT 稼ぎを主目的としているアカウントとする。

1 については、情報商材や個人情報を得ることを目的としたサイトへの誘導、アフィリエイトを目的としたスパムブログへの誘導が含まれる。2014 年に入ってから、Twitter の機能を逆手に取ったアプリ連携のサイトへの誘導もみられる[6]

2 については、有害サイトや誘導目的のアカウントが効果を発揮するためには、閲覧者が多くなる必要がある。そのため、フォロワーを増やしたり、RT によりシェア数を増やすことで公式アカウントや著名人のアカウントのような挙動に見せかけることを目的としている。1 の要素も含み、有害サイトへの誘導を行っているアカウントも存在する。

4. 解析・考察

解析対象として、スパムアカウントと思われるアカウントをフォローしたダミーアカウントと、スパムアカウントをフォローしていない筆者の個人アカウントを準備した。前者については、「稼ぐ」・「無料」・「相互フォロー」等、スパムでよく使用される語句で検索した中で、スパムだと目視で確認したものや、ネット上のコミュニティにおいてスパムだとされているものをフォローして準備した。フォロー情報はそれぞれ、50 アカウント・46 アカウント分である。

さて、Whotwi[7]というサービスにより、各アカウントの投稿内での言及数が多いアカウントを「仲良し」としてランキング表示することができる。これを利用して、まず、ダミーアカウントでフォローしたアカウントと関わりの大きい、「仲良し」とされたアカウントの中からそれぞれ上位 10 アカウントについて情報を収集した。次に、関わりの大きいとされたアカウントの中で 2 回以上現れた頻度の高いアカウントについて、リンク付ツイート数および RT 数を調べた。個人アカウントについても同様の手順でデータを収集した。対象となったアカウントは、スパムと思われるアカウントから辿ったものと筆者の個人アカウントから辿ったもの、それぞれ 24 アカウントと 34 アカウントである。以後これを I 群、II 群とする。さらに、投稿内容で判断し目視で確認した。リンク付ツイート数または RT 数のどちらかが 8 割以上含まれているアカウントには、結果欄に○印が付してある。なお、今回は日本語のツイートのみを対象とした。

表 1、表 2 に結果を示す。表 1 において、最下段に示された色つきのアカウント X は、Twitblock[8]というスパム検出サービスによりスパムだと判定されたアカウントである。A から W については、スパムアカウントとして検出されなかった。

表 2 で欄が—となっているものは、プロテクトされているアカウントで、筆者からは閲覧権限がないため、投稿については確認できなかった。しかし、プロテクトされているユーザーは情報の拡散ができないため、スパムではないと考えられる。

また、I 群において、一番頻出であったアカウントは、スパムアカウントではなく、管理ツールクライアントのアカウントであった。この管理ツールとは、フォロー・フォロワーの管理や新規フォローの開拓、他のアカウントにフォローされたときに自動的にフォローし返す機能などが付いたものであり、いかに人の手を使わずにフォ

† 千葉大学大学院工学研究科

ロワーを増やすかという観点からスパムアカウントでの使用が多かったと考えられる。

表 1: I 群における解析結果

番号	回数	リンク付ツイート数	RT数	結果	目視
A	7	1	0		×
B	4	6	10	○	○
C	3	2	10	○	○
D	3	10	8	○	○
E	3	1	10	○	○
F	3	2	2		×
G	2	8	6	○	○
H	2	2	10	○	○
I	2	2	10	○	○
J	2	10	0	○	○
K	2	10	10	○	○
L	2	2	10	○	○
M	2	2	10	○	○
N	2	1	10	○	○
O	2	0	10	○	○
P	2	2	9	○	○
Q	2	2	9	○	○
R	2	2	9	○	○
S	2	2	9	○	○
T	2	2	9	○	○
U	2	1	8	○	○
V	2	1	0		○
W	2	2	8	○	○
X	1	0	10	○	○

表 2: II 群における解析結果

番号	回数	リンク付ツイート数	RT数	結果	目視
1	6	2	2		×
2	6	3	0		×
3	6	2	2		×
4	4	1	0		×
5	3	3	1		×
6	3	1	2		×
7	3	—	—		×
8	3	2	3		×
9	3	3	1		×
10	3	2	0		×
11	3	1	0		×
12	3	0	0		×
13	3	2	3		×
14	2	1	0		×
15	2	0	0		×
16	2	2	6		×
17	2	9	0	○	×
18	2	7	4		×
19	2	0	0		×
20	2	1	0		×
21	2	2	0		×
22	2	3	4		×
23	2	1	3		×
24	2	1	2		×
25	2	—	—		×
26	2	0	0		×
27	2	—	—		×
28	2	7	1		×
29	2	1	2		×
30	2	9	0	○	×
31	2	2	0		×
32	2	2	0		×
33	2	1	0		×
34	2	5	4		×

表 3: スпам含有率 (%)

	I 群	II 群
リンク付ツイート・RT 数から算出	87.5	5.9
目視で確認	91.7	0

表 3 は, I 群, II 群についてスパムアカウントの含有率を示したものである。リンク付ツイートと RT 数によって算出されたものと, 目視で確認したものの, どちらも I 群のほうが高い値を示している。これは, 通常アカウントと比較して, スпамアカウントはスパムアカウント同士でのつながりが強いことを表している。

5. おわりに

本論文では, スпамアカウントとのつながりが大きいアカウントと, 通常アカウントとのつながりが大きいアカウントについて, スпамの含有率を調査分析した。その結果, リンク付ツイート数と RT 数より算出した場合はそれぞれ 87.5%, 5.9%, 目視で確認した場合はそれぞれ 91.7%, 0%の割合でスパムが含まれていた。これより, 通常アカウントより, スпамアカウントとのつながりが大きいアカウントの方がスパム含有率は極めて高いことが示された。スパムアカウント同士のつながりを用いて, 従来のスパムアカウント判定方法と組み合わせることにより, さらに効果的にスパムを発見できるようになると考えられる。

また, 今回はリンク先 URL 等の情報を加味していなかったが, これを加えることによりスパム判定の信頼性は増すと考えられる。より大量のデータを用いたリンク解析や精度を高めるための手法については, 今後の検討課題とする。

参考文献

- [1] Twitter <https://twitter.com/>
- [2] 中村悠一, 山田剛一, 絹川博之, “Twitter におけるスパムユーザフィルタの開発とその評価” (第 11 回情報科学技術フォーラム, 2012)
- [3] 若井一樹, 岡田泰輔, 鎌田祐輔, 佐々木良一, “Twitter の表示系を発展させスパム発見機能を強化したアプリケーション LookUpper の開発と評価” (マルチメディア, 分散, 協調とモバイルシンポジウム, 2013)
- [4] 寒河江昭博, 勝野裕文, “日本語ブログ空間におけるスパムブログ発見手法の提案” (情報処理学会第 71 回全国大会, pp. 1-635, 1-636)
- [5] 原正憲, 長谷巧, 山本匠, 山田明, 西垣正勝, “スパムブログとアフィリエイトの関連性に関する一考察” (情報処理学会論文誌, Vol. 50 No. 12 3206-3210, 2009)
- [6] “「ドラえもん打ち切り」など, Twitter で広がる悪質なデマツイートに注意” マイナビニュース (2014/02/03)
- [7] whotwi - グラフィカル Twitter 分析 <http://whotwi.com/>
- [8] twitblock <http://twitblock.org/>
<http://news.mynavi.jp/news/2014/02/03/084/>