Vol.2014-IS-130 No.1
2014/12/8

Twitterにおける言及関係によるクラスタリングを利用したスパムアカウント判定手法の検討

菊池望† 吉村博幸†
Nozomi Kikuchi Hiroyuki Yoshimura

近年，Twitter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いる．しかしその中には有害なサイトへの誘導等，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，一般ユーザーを装う手口が巧妙化している．そこで本稿では，Twitterにおいてアカウント同士の言及関係のネットワークに着目し，各アカウントのクラスタ係数と，言及数を指標にした繋がりの強弱を利用してスパムアカウントの判定を行った．

Extraction of the SPAM accounts by use of clustering of the mention-related network in Twitter

NOZOMI KIKUCHI† HIROYUKI YOSHIMURA†

In recent years, users of Twitter have increased and everyone owns his account freely. However, the SPAM account which diffuses the information becoming disadvantageous for users, such as guidance to a harmful site, has also existed, and a means pretending to be a general user has become skillful. So, in this paper, the SPAM account was judged using the cluster coefficient and the number of references of each account by paying attention to the mention-related network of account in Twitter.

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キングサービス（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，匿名での利用が可能である上，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多く存在する.スパムの存在により，必要な情報を探す際にスパム投稿に埋もれ見つけ出しづらくなるといった問題点もある．

Twitter を管理する Twitter 社においてもスパム報告を受けてアカウントの凍結・削除を行う等スパムアカウントへの対応は行われているが，スパムアカウントの数は日々増加しており，対応が追い付いていないのが実情である．スパムによる被害を防ぐためにはTwitter社における対応を待つだけでなく，利用者においてもスパムアカウントに対処する必要がある．そのため，まずは利用者がスパムアカウントを発見・判定できることが重要である．

そこで，一般の利用者が得ることができる情報を用いて，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とする．本論文では，その足掛かりとして，各アカウントにおける言及関係を繋がりとみなし，スパムアカウントの特徴を捉え判定基準の検討を行った．

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．[4][5]の報告によると，アフィリエイトへのリンクが多く張られているものはスパムブログである確率が高くなり，また，アフィリエイトサイトへの誘導のためアクセス数稼ぎを行うスパムブログも存在する．近年では情報発信手段の変化により，Twitter においてもブログと同様の手口のスパムが増加している．

筆者の前回の研究[6]では，言及関係を数え上げ，スパムアカウント同士で言及関係があることを示した．今回はより定量的な指標で示す方法を提案する．

3. 提案手法の概要

3.1 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，リツイート（以下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に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や携帯ゲーム等のダウンロードページへの誘導が含まれる．また，2014年に入ってからは，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[7]．

一方，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，閲覧者が多くなることが必要になる．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．

† 千葉大学大学院工学研究科

図 1　スパムアカウントによる投稿の一例

図 1 にスパムアカウントの投稿の一部を示した．スパムアカウントの多くはこのような形で閲覧者を増やすための投稿や，誘導目的の投稿を行っている．

3.2 言及数について

Twitter では，他のユーザーのアカウント名を@を含める形式で記述することにより，他者への言及（mention）を行うことができる．

方法として代表的なものを以下に挙げる．

リプライ：他者の投稿に対する返信
非公式 RT：他者の投稿を引用して自分の言葉を付加
公式 RT：他者の言葉をそのまま引用

閲覧する際のクライアントにより表示のされかたは異なるが，投稿内の文字情報ではリプライは@ユーザー名が投稿の先頭にくる形で記述され，RT の場合は RT@ユーザー名の形で記述される．そのため，言及先を取得するために@のついた投稿に着目する．

フォロー関係によるクラスタ係数の算出は行われてきた[8]が，フォロー関係については 2014 年現在 API の仕様上一定数を超えたフォロー関係の取得が困難であり，スパムアカウントは個人のアカウントと比べフォロー数・フォロワー数が大きくなることから傾向が掴みにくい．また，急激にフォローを増やした場合 Twitter 社によって対象アカウントの凍結が行われるようになったことから，フォローを増やさないスパムアカウントが出現してきた．そのため，フォロー関係によるクラスタリングとは別の観点からのアプローチを加えることが重要であると考える．

3.3 分析手順，データについて

分析データの取得方法について以下説明する．対象とするアカウントを一つ決め，起点とする．今回起点とするアカウントとして，筆者の個人アカウントと，スパムと確認したアカウントの二つのアカウントを取り上げ，比較を行う．以下，個人アカウント起点で取得したデータを Ⅰ 群，スパムアカウント起点で取得したデータを Ⅱ 群とする．TwitterAPI により起点とするアカウントの最新ツイート 200 件を取り出す．その中から，@のついた投稿を抜きだし，@付きで言及されているアカウントのリスト（言及先アカウントリスト）を作成する．その後，言及先アカウントリストにあるアカウントに対しても同様の手順で最新 200 件のツイートを取得し，言及先をリスト化する．これを行うと，理論上最大 200×200+1（起点アカウント）=40,001 のアカウントが出現する．これから重複を取り除いたアカウント数が総数となる．実際には，最新 200 件のツイートすべてが言及ありのツイートになることは稀であること，加えてすべての言及先が違うことはほとんどないため，総数はこれよりも少なくなる．

今回は傾向を見るため，個人アカウントとスパムアカウントそれぞれの最新 200 件のツイート中において言及数が高かった上位 5 アカウント，さらに言及先アカウントにおいても各上位 5 アカウントに限定してアカウントの取得・言及関係の分析を行った．この場合，Ⅰ 群 Ⅱ 群それぞれ 5×5+1 で最大 26 件のアカウントが出現する．そこから重複を取り除いたものが，今回取り扱うアカウント総数である．

次に，（アカウント総数）×（アカウント総数）の接続行列を作成する．言及関係は向きと重みを持つが，接続行列作成においては考えない．アカウント 1 とアカウント 2 についての関係を考えるとき，1 から 2 または 2 から 1 へ言及が行われていれば行列の値を 1 とし，どちらからも言及が行われていなければ 0 とする．言及関係を可視化し，重みを除く手順をグラフで示したものが図 2 である．

言及数については数値データとして保持しておく．言及数データは言及数のばらつきを見るときに用いる（後述）

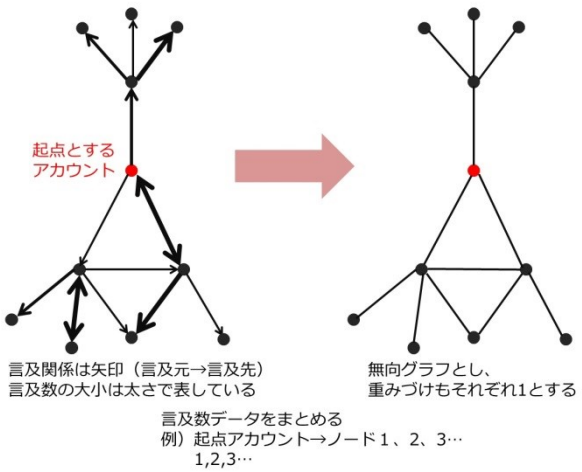

図 2　言及関係のグラフ化手順

出来上がった接続行列データを用い，クラスタ係数を算出する．今回の計算には R を用いた．

3.4 クラスタ係数

クラスタとは，3 つのノード（点）がそれぞれ繋がって三角形を作っている状態を指す．今回の分析では，ノードは各アカウントに相当し，繋がりは言及関係に相当する．クラスタ係数とは，ネットワークの中でクラスタを見出す確率である．[9]

ネットワーク上にあるノード i におけるクラスタ係数を C_i とすると，クラスタ係数は以下のように表される．

$$C_i = \frac{2E_i}{k_i(k_i-1)} \quad \cdots(1)$$

ここで，k_i はノード i と繋がっているノードの数で，E_i はノード i から繋がっている 2 つのノードが繋がっている数である．

また，ネットワーク全体のクラスタ係数は以下のようになる．

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \quad \cdots(2)$$

これは，ネットワーク内それぞれのノードにおけるクラスタ係数の平均をとったものである．

3.5 使用データの可視化

使用したデータ（Ⅰ群，Ⅱ群）について，グラフで表したものが次に示す図 3，図 4 である．

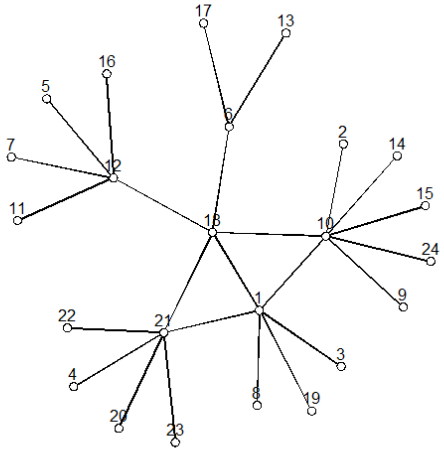

図 3　Ⅰ群（個人アカウント起点）

可視化を行うと，Ⅰ群とⅡ群の間では形状に違いが見られる．実際のスパム判定においては，グラフ描写を行わなくても判定ができるように，クラスタ係数の算出によって定量的な判定を行う．

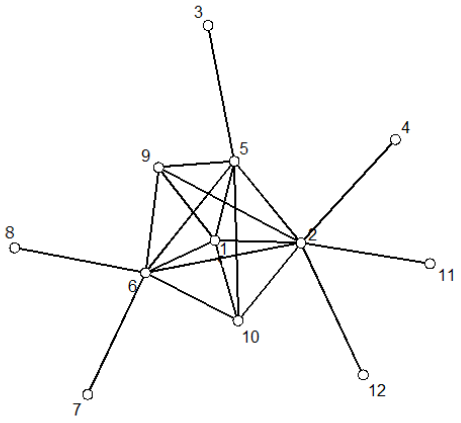

図 4　Ⅱ群（スパムアカウント起点）

3.6 標準偏差の比較

Ⅰ群，Ⅱ群それぞれの言及先リストの抽出にあたって，単純に言及数の比較を行ったところ，大きな差異は見られなかった．しかし，言及数のばらつきに違いが見られたため，言及数の平均・標準偏差について分析を行い傾向について比較を行った．

言及数分析の対象としたのは，クラスタ係数の算出の際使用したアカウントと同一のもので，起点とするアカウントとその言及先アカウントについて言及先とその回数を記録し，言及回数において平均・標準偏差を算出した．

4. 分析結果，考察

4.1 分析結果

算出した結果について以下に示す．

表 1　Ⅰ群・Ⅱ群におけるクラスタ係数

	アカウント総数	クラスタ係数
Ⅰ群	24	0.0187
Ⅱ群	12	0.354

表 2　言及数の平均・標準偏差

	平均	標準偏差
Ⅰ群	3.37	2.26
Ⅱ群	1.47	1.14

表 3　言及数の平均・標準偏差
（言及数 0 を含まない場合）

	平均	標準偏差
Ⅰ群	3.74	2.06
Ⅱ群	2.20	0.523

表 1 より，クラスタ係数はⅠ群において 0.0187、Ⅱ群においては 0.354 と一桁以上の差が表れた．言及数の標準偏差についてはⅠ群が 2.26，2 群が 1.14，言及数 0 を除いた場合の結果は，Ⅰ群が 2.06，Ⅱ群が 0.523 となった．

4.2 考察

表 1 より，クラスタ係数については I 群より II 群の方が高い値を示した．これより，II 群のアカウントは特定の範囲内で言及関係が築かれていると言える．特定のアカウント同士での言及関係が行われるのは，言及を行うことにより言及元の投稿の評価を高めることができるためである．また，言及を行うことにより一般のユーザーに見せかけることができるといった狙いがあると考えられる．

続いて，表 2 について考察を行う．最新 200 件の投稿を対象にしているため，言及数の上限は I 群，II 群ともに 200 である．表 2 は言及数 0 のアカウントについても 0 として算出の対象にしたものであり，表 3 は言及数 0 のアカウントは算出の対象としていないものである．数値のばらつき具合を表す標準偏差の値を比較すると，II 群と比べ I 群が大きい値を示し，ばらつきが大きいことがわかった．表 3 の言及数 0 の数値を取り除き，算出を行った結果では差がより顕著に表れる．

言及数は，アカウントによってばらつきを持つ．機械的に処理されているアカウントと比べ，人間の手によって管理がなされているアカウントでは，言及先によって言及数のばらつきが大きくなると考えられる．これは，言及の主たる目的として会話が挙げられること，Twitter において対話が行われると複数回にわたって言及付きの投稿がなされることによる．また，RT で他者の投稿を引用する場合は，引用元のアカウントは対話しあう関係内にとどまらず，言及関係もその時限りであることも珍しくない．そのため，言及数の大小が生じると推測される．一方，機械的に作成されたスパムアカウントにおいては，どのアカウントに対しても同じように言及を行うため，言及数のばらつきが小さくなると考えられる．

5. まとめ

今回の分析により，スパムアカウントは一般のアカウントと比べ，クラスタ係数は高く，言及数の標準偏差は小さい傾向にあることがわかった．そのため，スパムアカウント判定基準として言及関係によるクラスタ係数の大小と言及数のばらつきに着目することが有効であると考えられる．なお，今回は傾向を見るため言及数上位 5 アカウントずつに限定して分析を行ったが，今後は言及されているすべてのアカウントに対象を広げて分析を行っていきたい．

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) 菊池望，吉村博幸，“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”（第 13 回情報科学技術フォーラム，2014）
7) “「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）
8) 晒谷亮輔，“Twitter 上の人間関係ネットワークの抽出とその分析”（千葉大学都市環境システム学科平成 23 年度卒業論文）
9) Duncan J. Watts，Steven H. Strogatz，″Collective dynamics of 'small-world' networks″（Nature 393,pp440-442,4 June 1998）