5V-01

Twitter におけるリツイートに注目したスパムアカウントの検出

野村　信吾†　　中山　泰一‡

電気通信大学　情報理工学部†　　電気通信大学大学院　情報理工学研究科‡

1. 背景と目的

ソーシャルメディアの発展により容易に情報の取得と発信が可能となった．中でも Twitter はユーザ数を増やしており，2015 年 9 月での月間アクティブユーザ数は 3 億 2000 万に達している[1]．一方で，Twitter 内でのスパム行為が止むことはなく，2015 年ではアカウントを乗っ取りや不正なアプリ連携が話題となった[2]．この不正なアプリ連携はユーザの知らぬ間にスパムツイートをリツイートさせるもので，うっかりアプリ連携してしまったユーザがスパムの拡散に貢献してしまうものであった．

本研究ではリツイートに着目することでスパムツイートをするユーザとそれを拡散するユーザをまとめて検出するシステムの構築と検証を目的とする．

2. 関連研究

中村ら[3]は Twitter 社によって停止処分されたスパムユーザを学習データとした分類器を作成しスパムユーザフィルタの開発を行った．このフィルタは停止処分をされていないスパムユーザを 94.7%の割合で検出した．

若井ら[4]は Twitter 上のスパム行為となりすまし行為について様々な特徴から複数の項目を作成し，項目の組み合わせで検知を行う手法を提案した．スパム行為となりすまし行為のどちらも 90%以上の的中率を示した．

これらの研究では対象のユーザから最新のツイートを取得し判定に用いるが，判定材料となるスパムツイートが少ないと検出できない可能性があった．また，スパム間の繋がりは考慮されていない．スパムユーザには定期的にツイートを削除し，従来方法では判定が困難になると予想されるものがある．本研究ではこういったスパムユーザに対するツイート削除への対策と新たな判定基準としてリツイートに注目する．

Spam detection focused on Retweet in Twitter

† Shingo Nomura
Faculty of Informatics and Engineering,
The University of Electro-Communications, Japan
‡ Yasuichi Nakayama
Graduate School of Informatics and Engineering,
The University of Electro-Communications, Japan

3. 検出対象となるスパムユーザ

本研究ではリツイートを判定基準に用いるため，リツイートをされているスパムユーザが検出の対象である．

4. 提案手法

Twitter API を利用して検出を行う．実装には Twitter4J を用いた．以後，リツイートの元となるツイートを行うユーザを親ユーザ，親ユーザのツイートをリツイートするユーザを子ユーザと称する．提案手法を(1)～(3)に示す．

(1) 任意のユーザをストリーミングで監視する．任意のユーザが行った URL 付きツイートを記録する．

(2) 任意のユーザが行った URL 付きツイート毎にリツイートしたユーザを子ユーザとして記録する．また，任意のユーザが行った URL 付きリツイートの元となったツイートをしたユーザを親ユーザとして記録し，監視対象に加える．

(3) (1)で記録した任意のユーザのツイートのうち，リツイートされたものが 15 件になれば監視を終了し，任意のユーザが親ユーザであるかどうかを判定する．判定基準は 5.2 節で示す．

ストリーミングによるリアルタイムでの情報の取得により削除されたツイートも判定に用いることができる．

5. 判定基準の決定のための調査

5.1　調査内容

手作業で判別した 80 のスパムユーザ，100 の一般ユーザを 4 章の手法と同様に各ユーザを 2015 年 11 月 28 日から 2016 年 1 月 4 日の間にストリーミングで監視し，各ユーザのリツイートをされた URL を含むツイート 15 件とその各ツイートをリツイートした子ユーザ一覧をログとして取得した．判定基準の調査対象として 40 のスパムユーザ，50 の一般ユーザから親ユーザと子ユーザの傾向を見出し，判定基準を定める．そして，残りの 40 のスパムユーザと 50 の一般ユ

ーザの取得したログを用いてスパムユーザと一般ユーザの判別ができるかの評価実験を行う．

調査対象の親ユーザごとに 15 件のツイートの全子ユーザ数と，各子ユーザの 15 件中のリツイート数を集計した．親ユーザごとにリツイート数が 1～8，9～15 の子ユーザ数の合計を算出し，その割合を出した．

結果として，一般ユーザについて，50 の親ユーザ全てがリツイートされたツイート 15 件中 1～8 件のリツイートを行った子ユーザ数が 50%を超えた．逆にスパムユーザについては，40 の親ユーザ中 36 の親ユーザは 15 件中 9～15 件のリツイートを行った子ユーザ数が 50%を超えた．つまり，傾向として一般ユーザは同じ子ユーザが何度も親ユーザのツイートをリツイートすることが少なく，逆にスパムユーザは同じ子ユーザが何度も親のツイートをリツイートすることが多いことが分かる．

5.2 判定基準

5.1 節の結果から判定基準を以下のようにする．

(1)親ユーザのツイート 15 件中の 1～8 件をリツイートした子ユーザ数が全子ユーザ数の 50%以上であれば親ユーザを一般ユーザであるとする．また，その子ユーザを全て一般ユーザであるとする．

(2)親ユーザのツイート 15 件中の 9～15 件をリツイートした子ユーザ数が全子ユーザ数の 50%以上であれば親ユーザをスパムユーザであるとする．また，その子ユーザを全てスパムユーザであるとする．

6. 評価実験と考察

5.2 節の判定基準を実装した提案手法で調査に用いなかったスパムユーザ 40，一般ユーザ 50 のログに適応し正しく判定できるかの評価実験を行った．その結果を表 1 に示す．スパムユーザについては 40 ユーザ中 36 ユーザをスパムと判定し，一般ユーザについては 50 ユーザ中 49 ユーザを一般ユーザと判定した．本研究で提案したシステムの正答率は 94.4%であった．リツイートされているスパムのみの検出であるがリツイートによるスパム判定は可能であると考える．

誤検出してしまったユーザについて，まず一般ユーザであるのにスパムとされてしまった 1 件については特定のユーザのツイートを全てリツイートしているユーザが存在し，同時にこのユーザのみが子となっていたためツイート 15 件中 15 件リツイートしたユーザ 1 件で 100%となっていた．リツイートしたユーザ数の少なさが原因であると考えられる．

表 1. スパムユーザと一般ユーザの判定結果

		判定結果	
		スパムである	スパムでない
使用したユーザ	スパムユーザ(40)	36	4
	一般ユーザ(50)	1	49

スパムユーザであるのにスパムでないと判断された 8 件のユーザについては子ユーザごとに親ユーザの URL 付きツイートをリツイートするしないを分けているようで，子ユーザのリツイート数が少なくなっている．このようなふるまいの理由としては，親ユーザの URL 付きツイートをリツイートした人数が毎回ほぼ同じになるという不自然さを解消しアカウントの偽装の 1 つとしているからだと考えられる．

7. まとめと今後の課題

リツイートに着目したスパム検出手法とその基準を提案し検証した．本研究で提案する検出手法を使い，94.4%の精度でスパムユーザを検出することができた．

今後の課題として，誤検出も発生しているため，この誤検出を減らすためにリツイートしたユーザ数や度々現れる子ユーザ群を検出することを判定基準に組み込むことが挙げられる．

参考文献

[1] Twitter,Inc.について(オンライン),入手先<https://about.twitter.com/ja/company>(参考 2016-01-04).

[2] YOMIURI ONLINE ツイッタースパム撃退法,別ソフト利用の新手も(オンライン),入手先<http://www.yomiuri.co.jp/it/security/goshinjyutsu/20150306-OYT8T50177.html>(参考 2016-01-04).

[3] 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,情報科学技術フォーラム講演論文集,Vol.11,No.2,pp.99-100.(2012).

[4] 若井 一樹,佐々木 良一:Twitter のスパム検知機能となりすまし検知機能の開発と評価,情報処理学会論文誌,Vol.56,No.9,pp.1817-1825.(2015).