

ハッシュタグの種類とツイートダイナミクスとの関係性

武市融紀[†]笹原和俊[†]鈴木麗壘[†]有田隆也[†][†]名古屋大学 大学院情報科学研究科

1 はじめに

TwitterなどのSNSを用いて、ソーシャルメディア時代の人間行動に関する知見を得ようとする試みが盛んである。ソーシャルデータは口語やノイズが多いため、キーワードや投稿の数のみを調べる既存の分析手法だけで有益な知見を得ることは難しい。しかし、実世界の出来事に即時的に反応し、大規模でフラットな社会的相互作用が発生するソーシャルメディアの特徴を捉えることで、興味深い人間行動の性質が明らかになる場合がある。例えば、笹原[2013]はTwitterの投稿数の概日周期に着目して、人々のマクロレベルでの集会的な注意を定量化することに成功した。またMichael[2012]は、政治に関するハッシュタグによりツイートを収集し、フォローネットワークやリツイートネットワークが左寄り・右寄りの人々を中心として二極化している様相を示した。本研究では、ソーシャルセンサーとしてのTwitterユーザーの性質を明らかにするために、特定の実世界イベントと関連したユーザーの投稿行動や相互作用の分析を行う。主要なスポーツイベントにおいては高強度のツイートとリツイートが生じることが知られているため(笹原[2013])、今回は日本のプロ野球を題材として、ハッシュタグの種類とツイートダイナミクスとの関係性について調べた。

2 プロ野球に関するツイートの収集

本研究ではプロ野球の2013年の日本シリーズを分析対象とした。日本シリーズに関するツイートを収集するために、TwitterのSearch API¹を用い、プロ野球に関するハッシュタグを検索語に指定し、ツイートを収集した。使用するハッシュタグは、ハッシュタグクラウド²の情報から各球団を表すもので最も利用されているものを選定した。今回の解析には、日本シリーズで対戦した読売ジャイアンツ(#giants)と楽天イーグルス(#rakuteneagles)のハッシュタグを用いた。

3 ツイートとリツイートのダイナミクス

日本シリーズ第6戦はジャイアンツがイーグルスを4対2で下し勝利した。例として、試合開始から終了

までのツイート数とリツイート数の時間変化を図1に示す。なお、リツイート数とは公式リツイートの数であり、ツイート数とは全投稿数のうち公式リツイートを除いた数である。

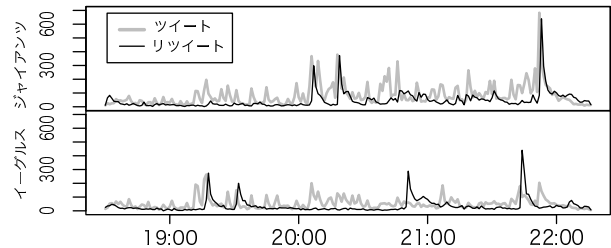


図1: 日本シリーズ第6戦におけるツイート数とリツイート数の時間変化。横軸は時刻。

イーグルスが先制しさらに2点目を追加した19時台では、イーグルス側のツイートとリツイートが連動した2回のバーストが見られる。この間、ジャイアンツ側はツイートのバーストは見られるがリツイート数は増加していない。一方、ジャイアンツが同点になり、さらに決勝点を獲得した20時台では、ジャイアンツ側にツイートとリツイートの2回のバーストが見られるが、イーグルス側には同様の現象は見られない。このように、自チームが得点した場合など、ポジティブなイベントの発生に際しては、ツイート数に連動してリツイート数もやや遅れて増加する傾向があるのに対して、相手チームが得点した場合などは、ツイート数は増加するが、それと連動するリツイート数の増加は生じない傾向があった。これは、リツイートにはポジティブな情報を選択的に拡散する傾向があることが原因だと考えられる。ツイートとリツイートの連動に関するこのような傾向は他の試合でも確認できた。このことから、勝ったチームの方が負けたチームよりもツイートとリツイートの連動性が高いことが予想される。このことを確かめるために、日本シリーズの全7戦について、試合開始直後から終了後1時間までのツイート時系列とリツイート時系列の相互相関を算出してその最大値を求めた。ウィンドウは10秒とし、ラグはツイート数に対してリツイート数を負の方向に移動させる、0分から5分の間で相関を最大にするものを選定した。図2に示したように、すべての試合で、勝利したチームの相互相関の値は敗北したチームより大きいことがわかった。勝利したチームは、ツイートとリツイートの行動

The Relationship Between Types of Hashtags and Tweet Dynamics
[†] Yuki Takeichi Kazutoshi Sasahara Reiji Suzuki Takaya Arita
[†] Graduate School of Information Science, Nagoya University

¹<https://api.twitter.com/1.1/search/tweets.json>

²<http://hashtagcloud.net>

が連動する傾向があり、リツイートのピークはツイートのピークに対して平均 110 秒遅れて発生している。2 チームの相関の差は試合の戦況を表していると考えることが出来る。第 1 戦は一方的な試合展開になっていたこと、第 5 戦は切迫した試合展開が続いていたことが相関の差から分かる。以上の分析から、ツイートとリツイートの連動性の大きさがチームの戦況や勝敗と深く関連する、という興味深い事実が得られた。

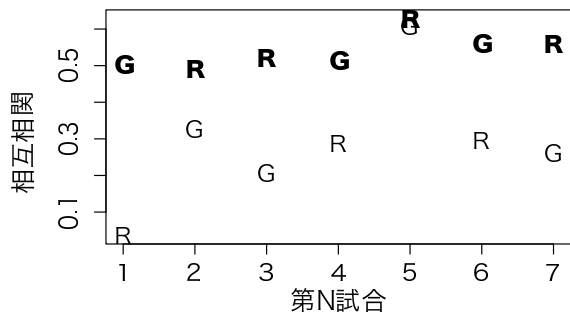


図 2: 日本シリーズにおけるツイートとリツイートの相互相関 (G=ジャイアンツ, R=イーグルス, 太字=勝ちチーム, 細字=負けチーム)

4 リツイートと社会的相互作用

野球観戦におけるユーザー間の社会的相互作用を調べるために、リツイートネットワークを分析した。ノードはユーザーで、エッジは公式リツイートを行ったユーザーから、元となったツイートをしたユーザーに張られる。例えば、ユーザー A のツイートを B がリツイートしてそれを見た C がリツイートした場合、エッジは B から A へ、また C から A へと張られる。

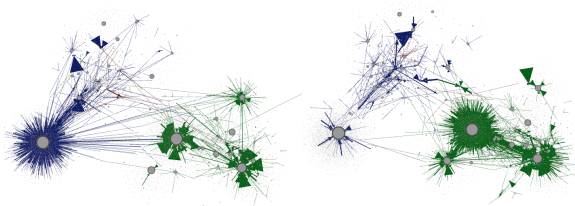


図 3: リツイートネットワーク (パネモデルでの作図。左=19 時台, 右=20 時台。ノードの大きさは試合開始から終了後 1 時間までの $\log(k)$ に対応している。リンクの色はリンクの色はリツイートに含まれるハッシュタグを表している。緑=#giants, 青=#rakuteneagles, 赤=両方を含む。)

図 3 は、前節と同じ日本シリーズ第 6 戦のデータで 19 時間帯と 20 時台のリツイートネットワークである。リツイートネットワークは 2 つのチーム間で二極化しており、イーグルス側にはリツイートが集まる中心的な役割を果たすノードが 1 つ存在しているが、ジャイアンツ側には複数存在している。また、わずかではあるが

両チームのハッシュタグを含むリツイートをするユーザーもいた。19 時台と 20 時台では、大規模なリツイートが生じるに寄与したノードが異なっていることは注目に値する。

図 4 は、図 3 の 2 つのネットワークの累積入次数分布である。両チームとも次数分布はスケールフリー性を示し、リツイートがパーストした時間帯の分布は、傾きの絶対値が小さい。以上の結果から、ジャイアンツファンはジャイアンツファン同士と、イーグルスファンはイーグルスファン同士と専らツイートのやり取りをし、両チーム間をリツイートが行き交うことは稀であることが分かった。また、次数分布はスケールフリーであることから、野球チームのオフィシャルアカウントやそれに類する振舞いをするアカウントが支配的にリツイートされることが示唆された。

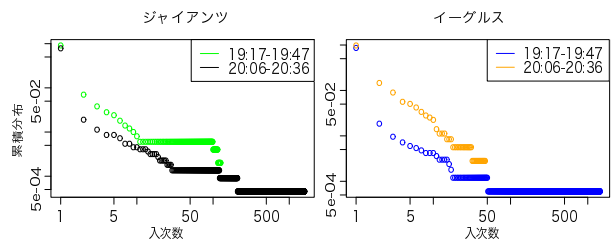


図 4: リツイートネットワークの累積入次数分布

5 おわりに

本稿では、ツイートとリツイートの相関やリツイートネットワークによる社会的相互作用の定量化が、ソーシャルセンサーとしての Twitter ユーザーの性質を調べる上で有効であることを示した。日本シリーズのデータにおいて、ツイートとリツイートの連動性は試合の戦況や勝ち負けと相関すること、リツイートネットワークは二極化したスケールフリー構造をしていることを明らかにした。これらの特異な性質は、Twitter ユーザーが野球というイベントをどのように知覚・反応したのかを反映している。今後は、日本シリーズ以外の試合でも同様の特徴が見られるのかを確認し、さらに、他のスポーツイベントのハッシュタグを用いて比較を行う。

6 参考文献

[1] Sasahara K, Hirata Y, Toyoda M, Kitsuregawa M, Aihara K (2013) Quantifying Collective Attention from Tweet Stream. PLoS ONE 8(4): e61823.
 [2] Michael D C, Bruno G, Alessandro F, Filippo M (2012) Partisan Asymmetries in Online Political Activity. EPJ Data Science 1:6.