

# プロ野球ツイートの バースト現象の分析

矢富 匡祐

## 研究背景

TwitterやFacebookなどのSNSの利用ユーザーがどのような行動をとっているのかに、着目した研究が多く行われるようになってきている。

## 参考文献

[1]水沼 友宏(2013)

Twitter におけるバーストの生起要因と類型化に関する分析

情報社会学会誌 Vol.7 , No.2, pp41-50, 2013-03

[2]武市 融紀(2014)

ハッシュタグの種類とツイートダイナミクスの関係性

第76回全国大会講演論文集.

2014-03-11 , pp145-146

## 関連研究

### [1] Twitter におけるバーストの生起要因と類型化に関する分析

- バースト時と非バースト時の特徴の違いを明らかにした.
- イベントごとにバースト時の特徴があることを明らかにした.

### [2] ハッシュタグの種類とツイートダイナミクスの関係性

- リツイート(RT)にはポジティブな情報を選択して拡散する傾向がある.
- ツイートとリツイートの連動性がの大きさがチームの戦況や勝敗と深く関連することを明らかにしている.

# 研究内容

プロ野球の試合に関するツイートを収集して

- ① バースト時と非バースト時の比較
- ② 試合状況ごとのバーストの特徴
- ③ バーストの類型化(クラスタリング)
- ④ ツイートとリツイートの連動性

を分析することによって

ユーザーの行動を明らかにしていく。

# ツイート収集方法

TwitterのAPIを用い、

プロ野球の2015年のセ・リーグの

- シーズン中の試合**27試合**

**(8/8,9,12,18,19,20)(9/5,6,8,9,10)**

- クライマックスシリーズ**3試合**

**(10/10,11,12)**

- 日本シリーズ**5試合**

**(10/24,25,27,28,29)**

# 野球の補足説明

セリーグ代表



上位3チームで  
代表を争う



セ・リーグ  
6チーム

パリーグ代表



上位3チームで  
代表を争う



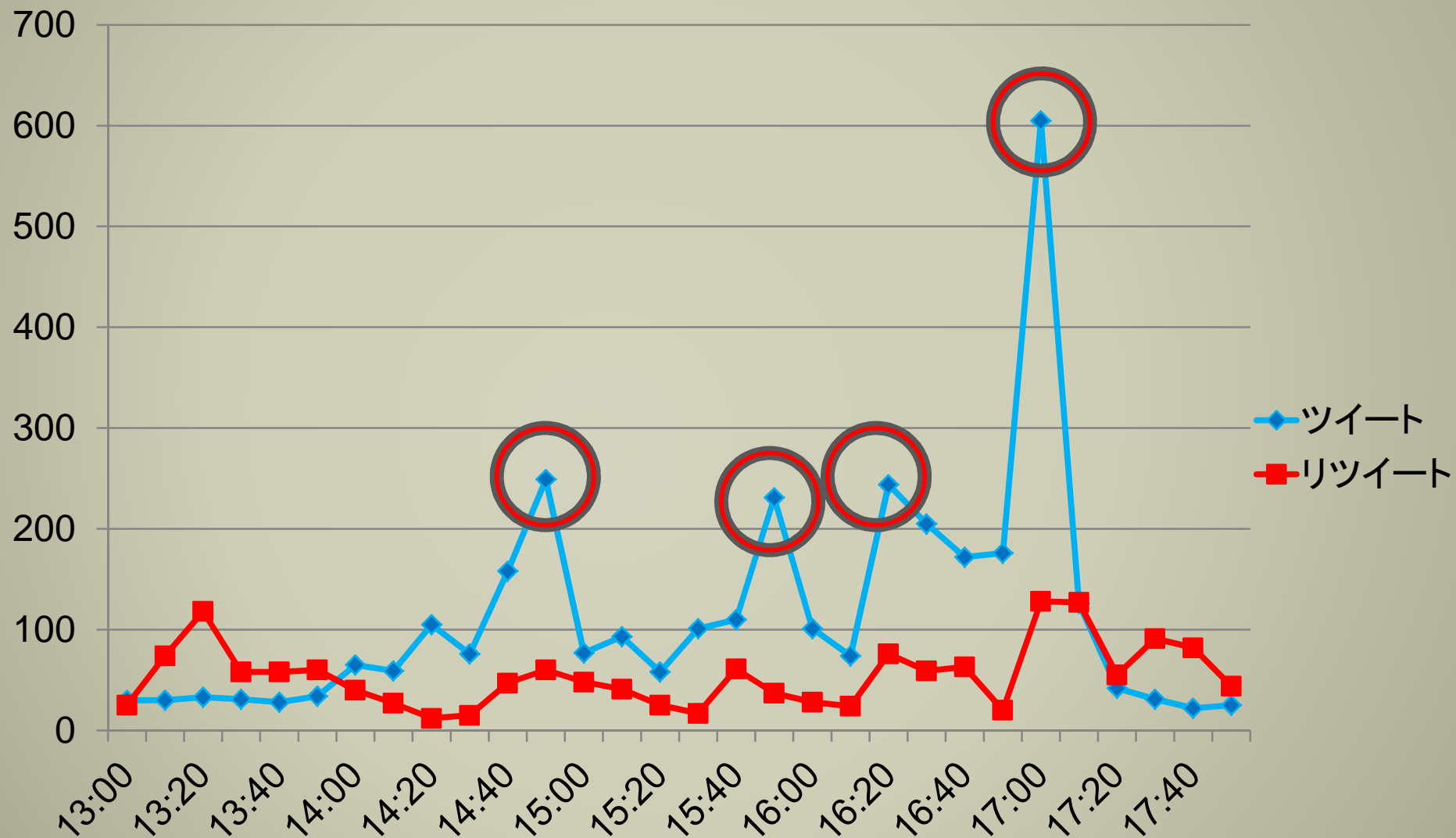
パ・リーグ  
6チーム

# 収集データの基本統計

ユニークユーザー数	118,739
総ツイート数	1,425,772
総リツイート数	616,680
総リプライ数	633,585
URL を含むツイート数	412,769
画像を含むツイート数	314,475



# バーストの例



# バーストとの定義

ただし・・・

収集したツイートについて

本文を全て確認してバーストを見つけるのは  
不可能である。

バースト検出の閾値を設定した。

# バースト閾値の式

$$\underline{\text{ツイートの閾値 } Nnt(t) = N(m) + 1.86\sigma(m)}$$

$$\underline{\text{リツイートの閾値 } Nnt(t) = N(m) + 1.5\sigma(m)}$$

ある試合に( $m$ )における閾値を $Nnt(t)$  とする。

$N(m)$ はある試合( $m$ ) における

10 分当たりの平均ツイートまたはリツイート数。

$\sigma(m)$  はある試合( $m$ ) における標準偏差である。

# 分析1

## バースト時と非バースト時の比較

バースト時と非バースト時を比較するために

(1)リツイート(*RT*)が含まれる比率

(2)リプライ(@)が含まれる比率

(3)ツイート内に含まれる画像や動画の比率

(4)*URL*の比率

(5)ツイートの平均文字数

算出し, バースト時の特徴を検出する.

# チームごとの集計

ヤクルト	非バースト時	バースト時
RTの比率(%)	46.23	41.25 <sub>4</sub>
@の比率(%)	0.29	0.13
URLの比率(%)	62.48	5.83
画像の比率(%)	26.66	20.69
平均文字数(%)	62.48	45.18

# チームごとの集計

巨人	非バースト時	バースト時
RTの比率(%)	<b>29.77</b>	<b>22.49</b>
@の比率(%)	<b>0.1</b>	<b>0.14</b>
URLの比率(%)	<b>33.76</b>	<b>3.87</b>
画像の比率(%)	<b>17.43</b>	<b>9.04</b>
平均文字数(%)	<b>56.13</b>	<b>32.52</b>

# チームごとの集計

阪神	非バースト時	バースト時
RTの比率(%)	<b>56.17</b>	<b>53.99</b>
@の比率(%)	<b>0.25</b>	<b>0.16</b>
URLの比率(%)	<b>54.11</b>	<b>43.05</b>
画像の比率(%)	<b>17.12</b>	<b>7.95</b>
平均文字数(%)	<b>68.04</b>	<b>60.94</b>

# チームごとの集計

広島	非バースト時	バースト時
RTの比率(%)	41.79	34.29
@の比率(%)	0.52 <sub>4</sub>	1.45
URLの比率(%)	40.17	10.03
画像の比率(%)	34.39	23.99
平均文字数(%)	61.91	40.70



# チームごとの集計

中日	バースト時	非バースト時
RTの比率(%)	<b>39.93</b>	<b>31.28</b>
@の比率(%)	<b>0.91</b>	<b>1.31</b>
URLの比率(%)	<b>32.47</b>	<b>5.37</b>
画像の比率(%)	<b>24.26</b>	<b>17.61</b>
平均文字数(%)	<b>59.73</b>	<b>39.61</b>

# チームごとの集計

横浜	非バースト時	バースト時
RTの比率(%)	<b>43.08</b>	<b>35.36</b>
@の比率(%)	<b>0.5</b>	<b>0.33</b>
URLの比率(%)	<b>41.35</b>	<b>11.72</b>
画像の比率(%)	<b>28.5</b>	<b>19.53</b>
平均文字数(%)	<b>58.61</b>	<b>41.11</b>

# 集計結果

	シーズン		CS		日本シリーズ	
	非バースト時	バースト時	非バースト時	バースト時	非バースト時	バースト時
RTの比率 (%)	42.83	36.44	54.04	54.89	58.25	51.83
@の比率 (%)	0.43	0.59	7.56	3.75	0.26	0.23
URLの比率 (%)	39.06	13.31	22.52	8.53	13.07	6.93
画像の比率 (%)	24.72	16.47	23.12	14.51	21.75	14.53
平均文字数 (%)	61.15	43.34	57.34	57.98	56.00	48.21

# バースト時の特徴

収集結果から . . .

- バースト時は文字数が**減少傾向**がある。
- バースト時はRTの割合は**減少傾向**がある。
- バースト時はリプライの割合は**減少傾向**がある。
- バースト時はURLの割合は**減少傾向**がある。
- バースト時は画像などの割合は**減少傾向**がある。
- シーズンの試合と比較して、  
CSと日本シリーズのRTの割合は高くなる。

# 分析2

## 試合状況ごとのバーストの特徴

これらのバーストは

試合の状況によって特徴が変わるのでは??

ということで試合の状況ごとに分類をして  
分析をする

# 試合の分類

試合の状況を今回は

(1)得点が入った場合 (139回)

(2)失点した場合 (83回)

(3)チャンスで無得点の場合 (52回)

(4)ピンチで無失点の場合 (35回)

バーストは全部で309回あり

括弧内の数字は状況別で出てきたバーストの回数。

# 分類結果

リツイート(RT)にはポジティブな情報を選択して拡散する傾向がある[2]

ポジティブな  
内容

	RT(%)	@(%)	URL(%)	画像(%)	文字数(%)
得点した時	31.21	0.60	11.66	11.28	38.81
失点した時	25.77	0.39	12.47	8.41	37.53
チャンスで無得点	20.56	0.65	10.69	7.07	35.59
ピンチで無失点	17.51	0.52	5.98	8.08	31.86

# 分析3 クラスタリング

	RT(%)	@(%)	URL(%)	画像(%)	平均文字数(%)
第一 クラス	65.52	0.10	55.19	5.31	66.70
第二 クラス	32.24	0.23	5.47	8.30	38.00
第三 クラス	12.78	0.17	5.09	6.60	29.3
第四 クラス	49.60	0.25	22.17	7.90	54.25
第五 クラス	18.03	2.96	6.96	13.70	31.94
第六 クラス	31.74	0.36	11.07	25.52	38.36



# クラスタリングの中身(チーム別)

	東京	巨人	阪神	広島	中日	横浜
第一 クラスタ	0	0	22	0	0	0
第二 クラスタ	15	17	10	3	4	3
第三 クラスタ	27	41	11	6	26	21
第四 クラスタ	2	2	16	0	0	0
第五 クラスタ	0	0	0	25	11	1
第六 クラスタ	9	2	3	9	0	3

## 分析4

### ツイートとリツイートの相関関係

[2]では、  
試合で勝ったチームの方が  
ツイートとリツイートの連動性が  
高いと証明した。

しかし・・・  
試合データは日本シリーズしかないため  
シーズンの試合などに当てはまるかわからない

# ツイートとリツイートの相関関係

引き分け1試合を除く, 35試合で調べた結果  
35試合中28試合で論文と同じ結果が得られた.

しかし、残りの7試合では  
負けているチームの方が高かった.

# まとめ

- 野球におけるTwitterユーザーの特徴が明らかになった。
- 試合状況にバーストの違いが明らかになった。
- チームの特徴が表れた。
- ツイートとリツイートの相関があることが明らかになった(例外あり)

# 今後の課題

- **収集データを増やす。**  
今回の実験ではデータが少ないため。  
単純に増やす必要がある。
- **試合状況以外でバーストの影響を分析。**  
連勝や連敗, その時の順位などが  
どのように影響してくるか。