

# 熱狂現象の計算社会科学： 日本プロ野球のソーシャルメディア分析

杉森真樹\*

水野誠†

笹原和俊‡

## 概要

熱狂は、人間が有する重要な性質であり、個人の行動の原動力となるとともに、人々を結びつけ体験の共有を促すなどの人間特有の集合現象を引き起こす。熱狂の本質を理解するために、日本で最も人気のあるスポーツの一つである日本プロ野球（NPB）に関するソーシャルメディア（Twitter）のデータ分析を行った。ツイートとリツイートの投稿数の急激な増加の共起を測定することによって、熱狂を検出、定量化するとともに、熱狂と試合結果や記録的な出来事との間に関連性があることを確認した。さらに、投稿から形成される集合注意の意味を分析し、球団ごとの意味的性質を比較することで、球団によって単語の関連性のパターンが異なっており、球団の性質などを反映していることを明らかにした。また、情報の拡散過程が球団の性質などによって異なることを確認し、熱狂の拡散過程も同様に異なる可能性があることを明らかにした。

## 1 導入

熱狂とは、ある出来事に対して興奮を覚えたり、その出来事に熱中したりする状態のことである。熱狂は、個人行動の原動力となるのみならず、時に人々を盲目的にしたり、人々を結びつけ体験の共有を促すなど、人間特有の集合現象を引き起こす。このような熱狂現象はスポーツやエンターテインメントの根本を支えるものである。従来の熱狂現象は、その

原因となっている出来事に直接またはテレビなどの映像を通して間接的に接している人々の集団内で発生する現象であった。このような従来の熱狂現象を検出、定量化することは困難であったが、ウェブのソーシャル化がこの状況を一変させた。

ソーシャルメディアの登場によって、オンライン上での人々の行動痕跡はデジタルデータとして記録されるようになり、これら进行分析することで、熱狂現象の検出・定量化を試みるのが可能になった。さらに、ソーシャルメディアの普及により、情報が時間や空間を超えて共有・拡散されるようになり、熱狂はより多くの人々が関与する複雑な集合現象になった。言い換えると、デジタル時代において、ソーシャルメディアは熱狂を定量化するためのツールであるとともに、熱狂を形づくるものの一部であると言える。したがって、集合現象としての熱狂を理解するためには、オンライン上での人々の行動をいかに定量化するかが大きな鍵となる。このような場合に有効なのが、オンラインデータから実社会における人間行動や社会現象を定量的に理解する計算社会科学 (Computational Social Science) のアプローチである [4]。

本研究では計算社会科学の手法を用いて、なぜ私達は興奮し、その興奮（感情）がなぜ、どのように共有・拡散されるのかという熱狂現象の創発過程を探索する。そのための格好の素材として、日本で最も人気のあるスポーツの一つであり、多くの熱狂的なファンが存在する日本プロ野球に関するソーシャルメディアのデータを対象として、熱狂現象についての分析を行う [11]。熱狂現象には様々な側面があるが、本研究では、特に熱狂の時系列的性質と意味

\* 名古屋大学大学院情報学研究科, sugimori@nagoya-u.jp

† 明治大学商学部, makoto@meiji.ac.jp

‡ 名古屋大学大学院情報学研究科, JSTさきがけ  
sasahara@nagoya-u.jp

的性質、情報の拡散過程の3つに焦点を当て、その他の側面については別の機会に議論する。

## 2 データセット

本研究では、大規模なデータの収集が可能なソーシャルメディアであるTwitterのデータを利用する。2018年度のプロ野球が開幕する前日である3/29から、日本シリーズが終了した翌日である11/4までの約7ヶ月間に渡り、データの収集を行なった。データの収集にはTwitter Search APIを利用し、表1に示すプロ野球球団に関連する特定のハッシュタグを含む投稿を収集した。利用するハッシュタグは、各球団について言及する際に頻繁に利用され、かつ、別用途に利用される可能性が低いものを選択した（例えば、「#楽天」はオンラインショッピングなどにも利用される）。なお、収集したデータの中には日本語以外の言語（主に英語）が確認されたため、それらは分析の対象から除外し、Twitterのシステムによって日本語と判定された投稿のみを抽出し、後述の分析に利用するものとする。

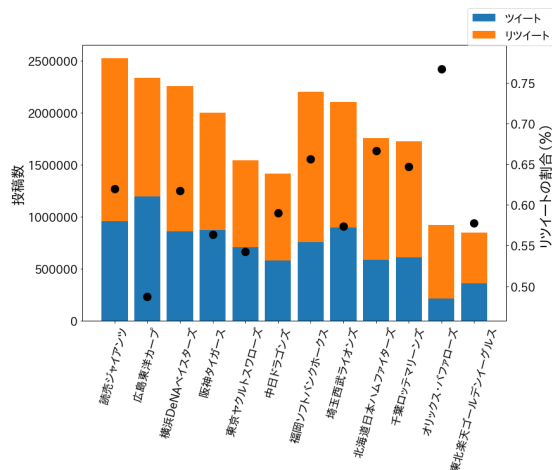


図1：各球団ごとの投稿数

図1は、収集したデータをツイート（通常の投稿）とリツイート（他のアカウントの投稿のコピー）に分類し、球団ごとに積み上げグラフとして示した図である。図中の黒点は全投稿に占めるリツイートの割合を示している。この図から、データ数には球団ごとにばらつきがあるものの、全投稿におけるリッ

ツイートの割合が多いことが確認できる。例えば、広島東洋カープを除く全ての球団はリツイートが全投稿の50%以上を占めており、特にオリックス・バファローズは約75%とリツイートの割合が投稿の大多数を占めている。リツイートの割合が多いということは、同一のツイートが多数のファンに拡散され、ある種のエコーチェンバー的状况になっていることを示唆している。その意味において、広島東洋カープに関するツイートの投稿パターンは例外的である。

## 3 熱狂現象の検出と定量化

### 3.1 関連研究と問題点

関連研究として、ソーシャルメディアの情報から実社会の出来事の検知を試みる「ソーシャルセンサー」の研究があげられる [7,9]。ソーシャルセンサーの研究の中には、スポーツを対象として、投稿数の急激な増加（バースト）を検知する研究が多数存在する [3,10]。単なる現実社会の出来事であれば、ソーシャルメディア上での特定の単語を含む投稿のバーストに注目することで検出が可能であると考えられる。

しかし、熱狂現象はスポーツの試合中に起きる出来事への単なる反応ではなく、その結果として発生する、現実社会の出来事に対する人間の感情に基づく現象である。そのため、単に投稿のバーストに注目するだけでは、そのイベントが熱狂現象と結びつくポジティブな感情に結びつくものであるか、落胆やヤジなどのネガティブな感情に結びつくものであるのか判別することは困難である。また、データから熱狂をどのように定義するかも自明でなく、先行研究もほとんど存在していない。

### 3.2 ツイートとリツイートの共起的バースト

そこで本研究では、熱狂現象を検出するために、ツイートとリツイートの投稿数の急激な増加の共起（共起的バースト）に注目する。この問題について、Takeichiらは、ある集団にとってポジティブな出来事であれば、その出来事に関する情報を拡散しようとするため、ネガティブな出来事よりもツイートのバーストの直後にリツイートのバーストが発生する

表1：収集に利用したハッシュタグ

チーム	ハッシュタグ
読売ジャイアンツ	#giants, #kyojin, #巨人, #読売ジャイアンツ, #ジャイアンツ, #読売巨人軍, #g党
阪神タイガース	#hanshin, #hanshintigers, #阪神タイガース, #タイガース, #tigers
広島東洋カープ	#carp, #カープ, #広島カープ, #広島東洋カープ
中日ドラゴンズ	#dragons, #ドラゴンズ, #中日ドラゴンズ
横浜DeNAベイスターズ	#baystars, #ベイスターズ, #横浜ベイスターズ, #横浜denaベイスターズ
東京ヤクルトスワローズ	#swallows, #yakultswallows, #ヤクルト, #ヤクルトスワローズ, #スワローズ, #東京ヤクルトスワローズ
福岡ソフトバンクホークス	#sbhawks, #ホークス, #ソフトバンクホークス, #福岡ソフトバンクホークス
オリックス・バファローズ	#orix_buffaloes, #orix, #オリックス, #バファローズ, #オリックスバファローズ, #orixbuffaloes, #bs2018
北海道日本ハムファイターズ	#lovefighters, #fighters, #日ハム, #日本ハム, #ファイターズ, #日本ハムファイターズ, #北海道日本ハムファイターズ
千葉ロッテマリーンズ	#chibalotte, #marines, #千葉ロッテ, #千葉ロッテマリーンズ
埼玉西武ライオンズ	#seibulions, #西武ライオンズ, #ライオンズ, #埼玉西武ライオンズ
東北楽天ゴールデンイーグルス	#rakuteneagles, #eaglenow, #楽天イーグルス, #イーグルス, #東北楽天ゴールデンイーグルス

可能性が高くなると仮定し、共起的バーストの分析を行っている [8]。その結果、試合の盛り上がり具合や勝敗と共起的バーストの相関係数が関係していることを確認した。本研究でも、この仮定を採用し、ツイートとリツイートの共起的バーストに注目することで、熱狂現象の検出と定量化を試みる。

図2は、2018年度の日本シリーズにおけるツイートとリツイートの投稿数のダイナミクスを示した図である。リツイートがバーストする直前に、ツイートがバーストしているケースが複数存在することが確認できる。また、第3戦（10/30）8回表の満塁ホームランや最終戦（11/3）のソフトバンク優勝などの際には大きなピークが発生しており、実際に起きた熱狂現象を反映していると考えられる。

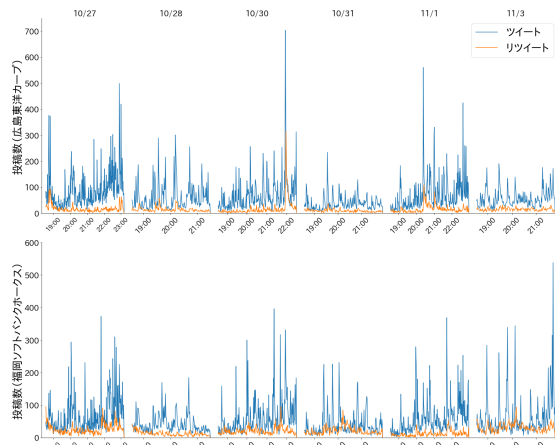


図2：日本シリーズの試合中の投稿数

### 3.3 共起的バーストネス

本研究では、このツイートとリツイートのバーストの共起の度合いを相互相関関数（cross-correlation function）によって測定する [1]。相互相関関数の定義式を以下に示す。

$$r_{xy}(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \left( \frac{x_i - \bar{x}}{\sigma_x} \right) \left( \frac{y_{i+\tau} - \bar{y}}{\sigma_y} \right)$$

ツイートとリツイートの共起的バーストに注目する場合、 $x$ は時間毎のツイート数、 $y$ は時間毎のリツイート数となる。ツイート数とリツイート数は10秒ごとに集計するものとする。相互相関関数の計算は各チーム、各試合に対して行い、試合中の投稿のみを利用する。 $\bar{x}$ ,  $\bar{y}$ は平均値、 $\sigma_x$ ,  $\sigma_y$ は標準偏差であり、 $\tau$ は時間間隔（0秒から300秒）、 $N$ は集計回数である。集計回数は、試合時間（秒）/10（秒）に相当する。本研究では、上記の定義式における最大値を「共起的バーストネス」と定義する。共起的バーストネスは、ツイートとリツイートの投稿数の増加（バースト）のパターンに最も相関があるときに1、最も逆相関がある時に-1、全く相関がなければ0となる。

次に、プロ野球の試合において最も重要な要素の一つである試合結果と熱狂現象の関連性をみるために、2018年度の全試合の共起的バーストネスを計算し、この値と試合結果（勝敗、点差）の関係を調査した。

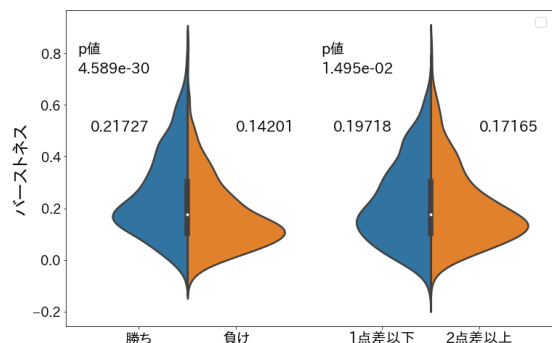


図3：共起的バーストネスの分布（図中の値：中央値）

図3は、試合結果ごとの共起的バーストネスの分布を示したものである。ここから、試合に勝った場合は共起的バーストネスが高くなる傾向が確認できる。点差については、勝敗ほど顕著ではないものの、1点差以下の試合と2点差以上の試合では前者の方が共起的バーストネスが高くなる傾向が見える。マン・ホイットニーのU検定でこれらの違いの確認を行った結果、勝敗については有意水準1%、点差については有意水準5%の統計的な有意差があることがわかった。このことから、試合に勝利した場合と、1点差以内で試合が拮抗している場合により大きな熱狂現象が発生することが明らかになった。

チームごとに試合中の共起的バーストネスの値が大きい日の投稿内容に注目すると、リーグ優勝やスター選手の引退、2000本安打達成などの試合内容よりも大きな注目を集める、記録的な出来事に関する単語が頻出していることが確認できた。これらの結果から、ツイートとリツイートの共起的バーストに注目することで、熱狂現象の検出と定量化が可能であることが確認できた。

## 4 熱狂の意味的性質の可視化

### 4.1 連想ネットワーク

次に、熱狂現象の意味的性質について、集団の特性によってどのような違いが生じるかを探るために、投稿内容（テキスト）に注目して、各チームのファンが有する集合注意（単語同士の関連性）について分析し、比較を行う。

各チームのファンが有する意味内容を比較するために「連想ネットワーク」を利用する[2]。連想ネッ

トワークは、連想ゲームのように意味的な関連性の高い単語を結びつけることで構築されるネットワークである。連想ネットワークでは、Mikolovらが提案した「word2vec」と呼ばれる技術を用いて、単語をベクトルに変換し、その類似度に注目する[5, 6]。word2vecにより生成されたベクトルは文脈を考慮した単語の意味を反映しているため、単語の意味の演算や関連性の高い単語の算出が可能なベクトルである。具体的には、シードとなる単語を決定し、そのシード単語との類似度が閾値 $s$ を上回る、関連性が高い単語の集合から上位 $N$ 個を抜き出す。そして、これらの単語に対して同様の処理を行い、抜き出した単語同士を結びつけることで構築されるネットワークである。なお、シード単語が複数ある場合は全てのシード単語に対して同様の処理を行う。

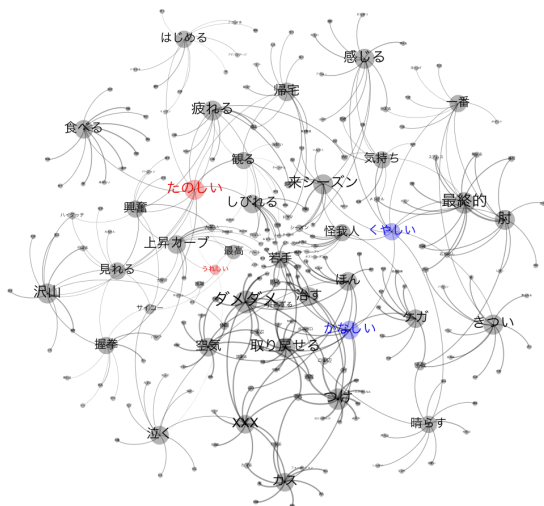
### 4.2 連想ネットワークの比較

本研究では、2018年度の日本シリーズに出場した広島東洋カープと福岡ソフトバンクホークスを対象として、日本シリーズの試合が行われた6日間の投稿のテキストから、それぞれword2vecのモデルを構築し、連想ネットワークを構築した。ちなみに、2018年度の日本シリーズは福岡ソフトバンクホークスが優勝している。シード単語には、熱狂ととりわけ関連性がある感情に関係する単語の中から、ポジティブまたはネガティブな単語である「うれしい」、「たのしい」、「かなしい」、「くやしい」を採用した。なお、シード単語については表記揺れを無くすために、漢字をひらがなに変換した上で処理を行なっている（例：楽しい  $\Rightarrow$  たのしい）。また、本研究におけるword2vecの次元数は200、連想ネットワークのパラメーターは、 $s = 0.6$ 、 $N = 10$ とする。

図4は、構築した連想ネットワークを示した図である。なお、一部の人名についてはプライバシー保護のため伏字（XXX）として記載する。

日本シリーズを制した福岡ソフトバンクホークスの連想ネットワークでは、「たのしい」が「観る」、「興奮」などといった単語と繋がっていることが確認できる。しかし、「うれしい」と関連する単語が少ないことから、日本シリーズを制した事実が他

一方で、日本シリーズに敗れた広島東洋カープの連想ネットワークでは、「うれしい」が「喜ぶ」、「泣く」などの単語とつながっている。また、「たのしい」は「緊張感」や「ドキドキ」、「面白い」などの単語と繋がっていることが確認できる。ネガティブな単語については、「かなしい」が「涙」、「つらい（辛い）」などの単語とつながっていることに加え、「くやしい」が「負ける」、「完敗」などの単語とつながっており、日本シリーズに敗れた結果が反映されていると考えられる。さらに、「うれしい」という単語が「たのしい」という単語と「幸せ」という単語を介して繋がっており、「かなしい」という単語と「泣く」や「泣ける」といった単語を介して繋がっていることから、正負に関わらず様々な感情同士が結びついていることが確認できる。



A complex network graph visualization showing relationships between various Japanese words. The nodes are represented by circles of varying sizes, and the edges are lines connecting them. The graph is highly interconnected, with many nodes having multiple connections. The words are in Japanese, and some are highlighted in red or blue. The overall structure is a dense web of relationships.

Key words visible in the graph include:

- 堅実 (Kenjitsu)
- しんどう (Shindou)
- AWAKE
- 考へ頭 (Kaohegata)
- ドキドキ (Dokidoki)
- 面白い (Omowai)
- 寝歌 (Nemika)
- 負ける (Makaru)
- たのしい (Tanoshii) - highlighted in red
- 見える (Mieru)
- くせいい (Kusei) - highlighted in blue
- 盛り上がる (Moridaru)
- 幸せ (Shiawase)
- 通り越す (Toorisosu)
- しょう (Shou)
- 感じ (Kanjie)
- 皆さん (Minna-san)
- うまい (Umai) - highlighted in red
- 伝わる (Dawaru)
- かっこいい (Kakkoii)
- 喜ぶ (Yoroku)
- 羨む (Enmu)
- 羨望 (Enbou)
- 泣ける (Nakeru)
- かなしい (Kanashii) - highlighted in blue
- 笑笑 (Shoushou)
- ぼんま (Bonma)
- 思い出す (Omoidasu)
- つらい (Tsurai)
- 久しぶり (Hisashiburi)
- 頂 (Itadaki)

これらの比較から、福岡ソフトバンクホークスは、日本シリーズを制したにも関わらず、ポジティブな感情と結びつく単語が少なく、ネガティブな感情は試合の結果や内容よりも、怪我に関する情報などに関連性があることが確認できる。一方で、広島東洋カープは、ポジティブかネガティブかによらず、感情同士に関連性があることが確認できる。このように球団によって、感情に対する集合注意（単語同士の関連性）が異なることが確認できたことから、球団の特性によって熱狂現象の意味内容が異なる可能性があることが確認できた。

熱狂の創発過程を探索する上で、熱狂がなぜ、どのように拡散されるのかを知ることは非常に重要である。本研究では、熱狂現象が拡散する原因と過程を探るための第一歩として、ソーシャルメディア (Twitter) における情報の拡散過程に注目する。

5



築されるネットワークであり、リツイートネットワークを分析することで、ソーシャルメディア上の情報の拡散過程について理解を深めることができる。本研究では、2018年度の日本シリーズにおけるリツイートネットワークを分析、比較することで、ソーシャルメディアにおける情報の拡散過程について調査を行った。なお、今回注目するのは情報の拡散過程であり、その中に含まれる熱狂に限定をした分析、議論については別の機会の行うものとする。

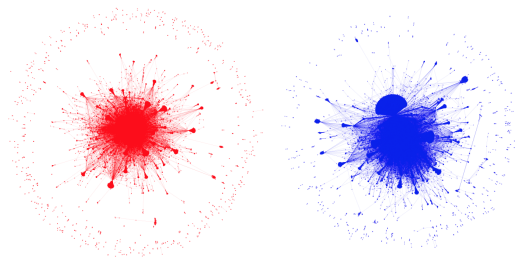


図5：リツイートネットワーク  
(赤：広島東洋カープ、青：福岡ソフトバンクホークス)

図5は、日本シリーズの試合が行われていた6日間におけるリツイートネットワークを示したものである。次数が大きく、固有ベクトル中心性やページランクなどが大きいユーザーの多くはニュースアカウント、公式アカウントであったが、一部、一般のファンのアカウントの存在も確認された。これらのユーザーに注目すると、福岡ソフトバンクホークスのネットワークには、二番目に次数が大きいユーザーの約4倍という、非常に高い次数を有するユーザーが存在し、これが球団の公式アカウントであることを確認した。加えて、球団のマスコットキャラクターのアカウントの次数が大きいことを確認した。これらの事実から、球団の公式情報が多く拡散されていたことが確認できる。一方で、広島東洋カープのネットワークでは、とりわけ高い次数を有するユーザーは存在せず、球団の公式アカウントの情報ではなく、主にニュースアカウントや一般のファンが発信した情報が拡散されていたことが確認できる。このことから、球団の特性やソーシャルメディアでの広報活動の取り組み、試合（日本シリーズ）の結果などによって情報が拡散する過程やその中で重要な

役割を担うユーザーが異なる可能性があることを確認した。

熱狂はこれらの情報の中に含まれるものであるため、熱狂が拡散する過程についても球団の特性や試合の結果などによって異なる可能性が考えられる。今後は、他の球団を含めた上で、熱狂自体の拡散過程に注目した分析やネットワークの特徴量の時系列分析、熱狂の拡散において重要な役割を担っているユーザーの特定、分析などを行うことで、熱狂現象の拡散過程についてさらなる研究を行う。

## 6 結論と考察

本研究では、計算社会科学の手法を用いて、ソーシャルメディア上における日本プロ野球に関するデータを対象として、熱狂現象の時系列的性質を定量化するとともに、意味的性質、拡散過程について分析を行った。

ある球団にとってポジティブな内容の投稿はツイートの投稿数が急激に増加するだけでなく、その直後にリツイートの投稿数も急激に増加するという仮定のもと、この二種類の投稿数のダイナミクスとその相互相関関係に注目をした分析を行った。その結果、試合の結果（勝利や点差が1点以下）によって、ツイートとリツイートの投稿数の増加の共起性が高まることを、共起的バーストネスの計測から確認した。また、共起的バーストネスの値が大きい日の投稿内容から記録的な出来事を検出した。

熱狂現象は人々の感情と深く関わる集団現象であるため、球団ごとに感情に対する連想ネットワーク（単語の関連性）の分析を行なった。その結果、連想ネットワークの構造が球団により異なり、球団の性質などを反映していることが確認できた。

また、熱狂の創発過程における拡散の過程やその原因を探るための第一歩として、情報そのものが拡散する過程について分析を行なった。結果として、情報の拡散過程が球団の性質などによって異なることを確認し、熱狂の拡散過程も同様に異なる可能性があることを明らかにした。

今後は、各集団の特性の分析や、より詳細なテキスト分析や感情分析、複雑ネットワークに基づく手

法などを用いて分析を行うとともに、本研究では扱わなかった熱狂の異なる側面について研究を行う。さらに、日本プロ野球以外を対象として同様の分析を行うことで、多様な熱狂現象に関する研究を行う。

## 謝辞

本研究はJSPS科研費(JP15H03446, JP17H06383JST), JSTさきがけ(JPMJPR16D6), JST CREST(JPMJCR17A4)の助成を受けたものです。

## 参考文献

- [1] Reinsel G. Ljung G. Box G., Jenkins G. *Time Series Analysis: Forecasting and Control*. Wiley, 2015.
- [2] Sasahara K. Visualizing collective attention using association networks. *New Generation Computing*, 34, 2016.
- [3] Smeaton A. F. Lanagan J. Using twitter to detect and tag important events in live sports. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 2011.
- [4] Adamic L. Aral S. Barabási A. L. Brewer D. Christakis N. Contractor N. Fowler J. Gutmann M. Jebara T. King G. Macy M. Roy D. Van Alstyne M. Lazer D., Pentland A. Computational social science. *Science*, 323, 2009.
- [5] Corrado G. Dean J. Mikolov T., Chen K. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, 2013.
- [6] Corrado G. Dean J. Mikolov T., Chen K. Efficient estimation of word representations in vector space. In *Proceedings of International Conference on Learning Representations*, 2013.
- [7] Matsuo Y. Sakaki T., Okazaki M. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on WWW*, 2010.
- [8] Suzuki R. Arita T. Takeichi Y., Sasahara K. Concurrent bursty behavior of social sensors in sporting events. *PLoS ONE*, 2015.
- [9] Leonardi E. Weng J., Yao Y. Event detection in twitter. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 2011.
- [10] Wickramasuriya J. Vasudevan V. Zhao S., Zhong L. Human as real-time sensors of social and physical events: A case study of twitter and sports games. In *Proceedings of CoRR*, 2011.
- [11] 水野誠, 三浦 麻子, 稲水伸行. プロ野球「熱狂」の経営科学:ファン心理とスポーツビジネス. 東京大学出版会, 2016.