

SNS 上でのサッカーの試合に対する直接的・間接的ネタバレの分析

白鳥裕士^{†1} 中村聡史^{†1, †2}

サッカーの試合を録画して視聴することを楽しみにしているユーザにとって、SNS 上で遭遇してしまうサッカーの試合結果を示唆するネタバレ情報は楽しみを減退させる忌むべきものである。本稿では、こうしたネタバレ情報を、単語から直接的に結果がわかる直接的ネタバレと、文脈から間接的に結果がわかる間接的ネタバレとに分類する。また、直接的ネタバレ、間接的ネタバレに関するデータセットを構築し、両者のネタバレの特徴を分析する。

Analysis of Direct and Indirect Spoilers of Football Match on SNS

YUJI SHIRATORI^{†1} SATOSHI NAKAMURA^{†1, †2}

Seeing the final score of a football match on a SNS often spoils the pleasure of a user who is waiting to watch a recording of this match on TV. In this paper, we classify spoiling information into direct spoilers and indirect spoilers and construct a dataset of spoilers about soccer matches on SNS. Then, we analyze the dataset and show characteristics of direct and indirect spoilers on SNS.

1. はじめに

スポーツは筋書きのないドラマであるため、勝つか負けるか分からないというハラハラ感を味わうためにリアルタイムでの視聴をしたいと考えている視聴者は少なくない。しかし、仕事や学業、その他の用事などでリアルタイムでの視聴ができないため、仕方なく録画予約をして時間ができたときに視聴しようとすることもある。そうした状況において、録画視聴を楽しみにしている視聴者が、視聴する以前にそのスポーツの試合結果を知ってしまうと、ハラハラ感や予想もしない展開に対する驚きが失われてしまう。こうしたスポーツの録画視聴を楽しみにしている視聴者にとって、試合結果はネタバレ情報と呼ばれており、多くの視聴者はこれを避けるため情報遮断を積極的に行っている。ここで、ウェブ上で視聴者がネタバレ情報に出会うのはニュースサイトやウェブログ、検索サイトなど様々であるが、特にネタバレとの遭遇機会が多いのが **Twitter** に代表される SNS である。

SNS 上にあるネタバレの遮断は、SNS サイトや SNS のためのアプリケーションを使用しないことで完全に断つことが可能である。しかし、SNS はコミュニケーションに利用されているものであり、完全に遮断することはコミュニケーションを遮断してしまうことにつながるため現実的ではない。

我々はこれまで、自身の都合によりリアルタイムでの視聴が不可能で、録画した試合を純粋に楽しむためにネタバレ

レを防止したい視聴者を対象に、スポーツの試合開始から視聴者の視聴開始までの間ネタバレ情報を遮断するための提示手法を提案[1]し、**Twitter** クライアントなどの形で実装[2]してきた。しかし、これまでの研究ではネタバレを画一的に扱っており、ネタバレがどのようなものなのかを深く分析出来ていなかった。

ここで、ネタバレについては文章内の単語などから簡単にネタバレと識別できる直接的なネタバレと、文脈からネタバレが暗に指し示される間接的なネタバレがあると考えられる。つまり、コンピュータで判断する際、直接的なネタバレは高精度に判定できるが、間接的なネタバレは判定が難しいと考えられる。

そこで本稿では、この直接的ネタバレと間接的ネタバレがどのように出現し、どのような特性を持っているのかを、サッカーの試合に関するネタバレデータセットを構築することで明らかにする。また、何人以上でデータセットを構築すべきかなどについても明らかにする。なお、SNS については特に **Twitter** を対象とし、さらにスポーツの試合についてはサッカーを対象とする。

2. 関連研究

ネタバレ防止に関連した研究として、情報推薦や情報フィルタリングがある。

ユーザの興味に関する情報をいかにして獲得するかという視点から情報推薦[3][4]に関する研究は以前から各分野において多くされてきた。このような情報推薦技術の場合、ユーザが興味のある情報を率先して提供してしまうため、ユーザの興味の対象の情報を遮断したいネタバレ防止

^{†1} 明治大学
Meiji University
^{†2} JST CREST

という観点からは、ネタバレ情報に触れる可能性を上昇させてしまうものであり、本稿で目指すネタバレ防止とは目的が異なる。

情報フィルタリングに関する研究において以前から多くされているのは、SPAM メールに代表されるコンテンツに対するフィルタリングである。しかし、これらのフィルタリング手法はほとんどのユーザーにとって興味のない有害なものを遮断するものであり、ネタバレのようなユーザーの興味の対象を動的に遮断するものとは異なる。

岩井ら[5]は機械学習アルゴリズムを用いてレビュー文のあらすじを分類・発見し、非表示にして提示することで、ストーリーに関する記述を読みたくないユーザーでもレビューを閲覧できる手法を提案している。あらすじが直接ネタバレとなるわけではないため、本稿で取り組んでいる内容とは異なる。また前田ら[6]は、ストーリーコンテンツに対するレビュー文をユーザーが参考にする際に、ネタバレを発見してしまうことを問題とし、ストーリーコンテンツに対する短文形式のネタバレデータセットを構築し、ネタバレに関する単語がストーリーコンテンツ内にどのように分布しているかを調査しており、コンテンツ自体からネタバレを判断する方法について検討している。我々は、ストーリーコンテンツではなく、決まったストーリー展開のないスポーツを対象コンテンツとしている。また、コンテンツではなくコンテンツに対するツイートからネタバレデータセットを構築し、ネタバレ判定を行うという点でアプローチが異なる。

また、Leavitt[7]らは、小説に対しネタバレ情報の提示の有無によって、ユーザーのコンテンツの楽しみ方にどのような差があるのかを実験により調査しており、その結果、ネタバレ情報はコンテンツの面白さを落とさないと主張している。この実験は、あらかじめあらすじを知っていることで、小説が読みやすくなるということを明らかにしているもので、我々が問題とするネタバレとは本質的に異なる。

一方、Golbeck[8]はテレビ番組に対して、放送時差による Twitter 上のネタバレコメントを判定する手法を提案している。Golbeck は対象コンテンツに対するあらゆるコメントを遮断することによって、再現率 100%を目指しているが、我々は SNS 上のコミュニケーションを可能な限り遮断せずに、試合の勝敗やスコアなどの致命的なネタバレ情報のみを遮断することを目的としている。

中村ら[1]は、スポーツの試合結果に関する記述を様々な提示手法により、ユーザーがネタバレ防止を可能とする方法を提案している。中村らはネタバレの判定にキーワードマッチングによる手法を用いているが、キーワードにマッチしないが文脈からネタバレを感じてしまうコメントも存在する。本稿では、こうしたネタバレも判定対象としている。

田中ら[9]はニコニコ動画などのコメント機能付動画共有サービスにおいてネタバレとなるコメントを検知する手

法を提案している。ニコニコ動画の性質上、そもそも視聴のタイミングはユーザーによって異なるものであり、SNS のようなリアルタイムに起きている事柄へのコメントではないため、対象が異なる。

3. 直接的ネタバレと間接的ネタバレ

本稿では、ネタバレを直接的ネタバレと間接的ネタバレに分類し、各々の特徴を分析した上でその両者を防止することを目的としている。ここで、直接的ネタバレと間接的ネタバレを下記のように定義する。

- **直接的ネタバレ:** 単語や記述パターンなどから試合の結果やスコアが直接伝わってしまうものを「直接的ネタバレ」と定義する。直接的ネタバレは「日本勝利!」「本田ゴール!!」「日本 2-1 アメリカ」などのようにパターンが明確であり、コンピュータで容易に判断できるものである。
- **間接的ネタバレ:** 「直接的ネタバレ」に対し、直接単語や記述パターンからは伝わらないが、文脈や時間的な要素などから試合の結果やスコアが読み取れてしまうものを「間接的ネタバレ」と定義する。間接的ネタバレは単語のみに注目すると多様な意味を表すことが多く、コンピュータで判断するのが難しい。例えば、試合終了後に「残念」や「悔しい」といった言葉があると、試合に敗北したことが分かるが、それ以外の時間帯であると分からない。また、文脈次第では試合についてのコメントかどうかさえも分からない場合もある。例として、「悔しい・・・」や「次こそは!」などが考えられる。

以上のように、サッカーの試合に対するネタバレと言っても直接的ネタバレと間接的ネタバレとは特性が大きく異なる。そこで本稿では、このサッカーの試合に対する Twitter 上での投稿を収集し、人手で分類を行ってデータセットを構築し、その特性を明らかにする。

4. データセット構築

ここでは、サッカーの試合に対する Twitter 上の投稿について、直接的ネタバレと間接的ネタバレの正解データを構築する。以降この Twitter 上の投稿をツイートと呼ぶ。

4.1 収集対象と方法

まず、サッカーの試合に対応するツイートを収集する。ここでは特にツイートが多く集まる日本代表の試合に注目し、下記の3試合を対象とした。

- FIFA ワールドカップブラジル 2014 グループリーグ初戦「日本 vs コートジボワール」で日本代表が敗北

- FIFA 女子ワールドカップカナダ 2015 決勝トーナメント準決勝「日本 vs イングランド」で日本代表が勝利
- 同大会決勝「日本 vs アメリカ」で日本代表が敗北

ここで、リアルタイムなコンテンツに対して Twitter 上でツイート（発信）する場合、ハッシュタグと呼ばれる検索・分類のためのテキストが付与される事が多い。例えば、サッカーの日本代表の試合においては、「#daihyo」や「#JPN」などのハッシュタグが用いられている。すべてのサッカーの試合に関するツイートに、サッカーに関するハッシュタグが付与されている場合は、そのハッシュタグを含むツイートをすべて遮断するだけでよい。しかし、実際にその試合に関連しているのにハッシュタグ無しでツイートされているものは多い。

あるサッカーの試合に対するツイートを全て集める場合、その時間にツイートされている全てのツイートを集め、そこからサッカーの試合に関するものを選別する必要がある。しかし、選別における精度の問題が生じるうえ、ツイートの内容を友人関係にしか開示していないユーザのツイートを集めることは出来ない。また、Twitter 社が提供している Stream API では全てのツイートを収集することは出来ない。

そこで本稿では、あるサッカーの試合に対して投稿されているハッシュタグ付きのツイートは、その試合に対して投稿されているすべてのツイートを代表していると仮定し、サッカーの日本代表の試合においてよく用いられる「#daihyo」や「#JPN」といったハッシュタグを含むツイートを Twitter Search API を用いて収集する。また収集においては、試合開始から、試合終了後から試合終了後 2 時間までの間ツイートを収集した。

4.2 データの整形

収集したデータの中には分類および分析において適切でないツイートも多い。そのため、下記の手順で不適切なツイートの除去およびツイートの整形を行う。

1. ワールドカップの試合などでは対戦相手国からのツイートも多数登場し、多言語となる。ここではデータセット構築者が日本人であることを考慮し、日本語以外のツイートを除去する。なお、日本語以外のツイートの除去については、ツイート取得時にあらかじめ言語コードを取得し、言語コードが「ja」かそうでないかによって判断する。
2. 先頭に「RT」を含むツイートは、Twitter のリツイート機能と呼ばれる他のユーザのツイートをそのままの形でツイートできる機能で、他のユーザのツイートを自分のツイートを見ている人に対して発信できるものである。これは元々のツイートと内容が重複するものであるため、正規表現により除去する。
3. 収集したツイートからハッシュタグの除去を行う。ハ

ッシュタグの除去については、「#」という文字から連続した空白・改行以外の文字までを正規表現により判定する。

4. 中身のないツイートを正規表現により削除する（ハッシュタグのみのツイートがあるため）。ここで先頭から最後までの間、空白および改行のみであった場合、そのツイートを除去した。
5. 無関係のスパムツイートを削除する。この作業については、1 つずつ目視で確認することによって除去する。

4.3 データの分類

収集したツイートが直接的ネタバレに該当するものか、それとも間接的ネタバレに該当するものなのか、それともネタバレではないのかということ进行分类するため、人手による分類を行った。

ここでは、先述の 3 試合のツイートからそれぞれ 1000 件ずつ無作為に抽出したツイートを対象とした。ここで 1000 件に限定した理由は、各データセットのデータ量を揃えるためである（もともとそれぞれの試合のツイート数は 1960 件、5554 件、11645 件）。

なお、分類作業を行ってもらうために、図 1 に示すウェブシステムを開発した。このシステムでは、ユーザは最初にアカウント名を入力してログインし、ページ上に提示されているツイートが直接的ネタバレなのか、間接的ネタバレなのか、非ネタバレなのかをラジオボタンにより入力するというものである（図 2）。ツイートは時間順ではなくランダムに提示される。またこのシステムでは、試合の状況が分かるように、対象とする試合の大会名や試合の展開などの詳細を載せた。



本サイトについて

ネタバレ进行分类するために用意したページです

じっくり読むというよりは全体を一見して判断してってください

今回対象とする試合は2014W杯「日本vsコートジボワール」です

(試合展開)

01:00:00頃-試合開始

01:16:00頃-日本先制(本田のゴール)

01:47:00頃-前半終了

図 1 開発したウェブシステム



図2 開発したウェブシステム (拡大)

データセット構築者には作成したウェブサイトアクセスしてもらい、ツイートの分類を行ってもらった。なお、ツイートに対する分類結果は1件毎にデータベースに記録されるため、途中で中断して再開することも可能となっている。データセット構築者は、Twitterを普段から用いている10代から20代の男子大学生8人であった。

5. 正解判定手法の検討

直接的ネタバレ・間接的ネタバレに対する評価は、人によってブレがあると考えられる。そこで、評価のブレをどのように判定したらよいかを考えるため、下記にあげる3つの判定手法を比較した。またここでは、ツイートに対する人数的な精度の変化も見ていく。

- **1票以上手法:** 誰か1人の評価者(データセット構築者)が直接的または間接的ネタバレと評価していたらそのネタバレとする手法
- **満票手法:** 全員が満場一致で同一のネタバレと分類している時にそのネタバレとする手法
- **1位手法:** 評価者によるネタバレの分類の内、最も票が集まっているネタバレとする手法

なお、ここでは各手法の8人で得られたデータを正解データとし、人数を7人、6人、5人と徐々に減らしていった時に、3試合で得られたデータの件数の平均がどう変化するかを図3に記載した。図の横軸は正解判定に反映させた人数で、縦軸はその人数での各手法・各ネタバレの平均ツイート件数である。また、人数および手法、ネタバレタイプ毎の精度を知るために適合率の平均を求めたものを図4に示す。図の横軸は正解判定に利用した人数で、縦軸はその人数での各手法各ネタバレの適合率の平均である。

ここで適合率とは、正解判定に利用した人数を k 人とするとき「 k 人で正解と判定されたもののうち、8人で正解と判定されたものの割合」のことである。なお、データはそれぞれの人数であり得るすべての組み合わせの平均を取ったものとした。

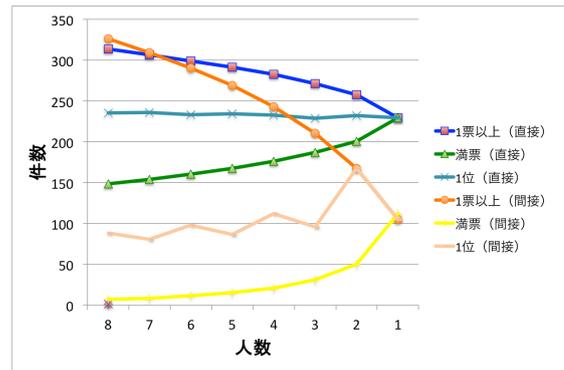


図3 人数による件数の変化 (3 試合平均)

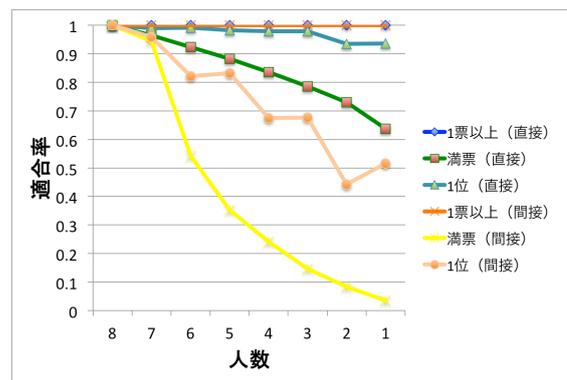


図4 人数による適合率の変化 (3 試合平均)

図3および4から、直接的ネタバレよりも間接的ネタバレの方が評価のブレが大きいことが分かる。これは、直接的ネタバレは文章内の表現自体がネタバレを表すことが多く判断に迷うことは少ないが、間接的ネタバレは人によって文章内の表現から伝わる内容が異なることを表している。

判定手法ごとに注目すると、1票以上手法は、その性質から他のどの手法よりも件数は上位で、人数を増やせば増やすほど差は広がっていることがわかる。この手法の場合、評価者(データセット構築者)すべてのあらゆる価値観や尺度でネタバレと判断されたものを全て網羅できるが、人数が増えれば増えるほどその件数は膨大になっていき、一般化が難しくなっていく。一方、満票手法は、その性質から人数の減少とともに件数は伸び、適合率は下がっていく。この手法の場合、評価者すべてで共通しているため、普遍的なネタバレを洗い出すことができると考えられるが、先述した間接的ネタバレの人によるブレが大きいという性質を考慮するとネタバレ防止においては適切ではない。1位手法はその性質から偶数人数と奇数人数の場合のブレが大きい。また間接的ネタバレにおいてブレが顕著であるが、これは人数を増やしていくことによって振幅は小さくなっていく。

以上のことから、今回はある一定以上の評価となっているこの1位手法で正解を利用して正解データを構築するとともに、直接的および間接的ネタバレの分析を行う。

なお、図4から人数的な精度の変化を考慮すると、直接的ネタバレを対象とするときには3人の評価者を集めるだけで良いが、間接的ネタバレを対象とするときには4人以上ではかなり結果が悪くなっているため、少なくとも5人以上の評価者が必要であることが分かる。

6. データセット分析

6.1 分類結果

1位手法で集計した結果、構築された3試合のデータセットは、平均すると直接的ネタバレが238.7件、間接的ネタバレが88.3件、非ネタバレが697.3件となった。「日本 vs コートジボワール」におけるツイートの分類結果の一部を表1に示す。

表1 コメントと分類の例

コメント	投稿日時	分類
本田～！！！！！！！！！！ GOOOOAL!!!!!!!!!!!!	2014-06-15 01:21:02	直接的
左で決めたんか本田めっちゃかっこいいやん	2014-06-15 01:50:26	直接的
日本負けたのか～さすがドログバだ!!心臓に悪いからやっぱ結果だけで良いな	2014-06-15 03:26:29	直接的
本田△!!!	2014-06-15 01:16:43	間接的
さすが本田△大舞台に強いな。	2014-06-15 01:19:31	間接的
ま、次とその次勝てばいいじゃん。	2014-06-15 02:59:57	間接的
うわーショック過ぎてうんこ漏れそう…	2014-06-15 03:00:15	間接的
後半41分：日本代表香川真司 out → 柿谷曜一朗 in	2014-06-15 02:43:43	非ネタバレ
なにしてんだよ～！！！！わちゃわちゃしすぎ！！！！	2014-06-15 02:23:29	非ネタバレ
さあキックオフ	2014-06-15 01:00:45	非ネタバレ

分類した結果、「直接的」のタグがつけられたものは、想定していた通り「○vs○」などのスコアの情報や、「ゴール」「負けた」など特定のパターン記述や単語が含まれているツイートであった。ただし、スコアの情報が書かれているにも関わらず、コメントが長くスコア部分が目に入らず「非ネタバレ」とされているものもあった。このような文章の見落としによりそのタグがつけられたと思われるものも数件存在したが、これはTwitterを実際に利用する場合もあり得るため、そのまま正解データとした。

一方、「間接的」のタグがつけられたものは、「悔しい」「ショック」などの感情的な単語を含むもの、「本田△」「本田～！！」などの名前が叫ばれているもの、「次」「切り替え」などの未来への変化を促す単語が含まれているものなどが存在した。ただし、2014-06-15 01:16:37（前半16分）に投稿された「本田△」が間接的ネタバレとされているのに対し、2014-06-15 02:13:52（後半11分）に投稿されてい

る「本田ああああ」は非ネタバレとされていた。これは、データセット構築の際に、試合の得点時間や得点者などの情報を、データセット構築者に知らせていたことが原因と考えられる。

6.2 試合経過に対するネタバレ数の変化

試合時間帯と直接的ネタバレ及び間接的ネタバレとの関係性を知るために、試合経過に伴うツイート数の変化（データセット内のツイートのみを使用）をグラフ化したものが図5～7である。図の横軸は試合開始からの経過時間（分）で、縦軸は1分間のツイート件数である。なお、図5が「日本 vs コートジボワール」、図6が「日本 vs イングランド」、図7が「日本 vs アメリカ」の試合のものである。

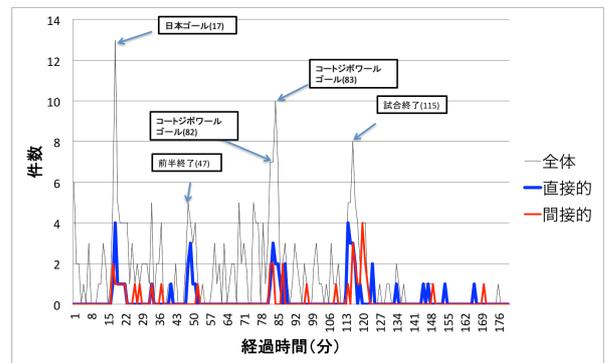


図5 1分毎のツイートの変化（「日本 vs コートジボワール」）

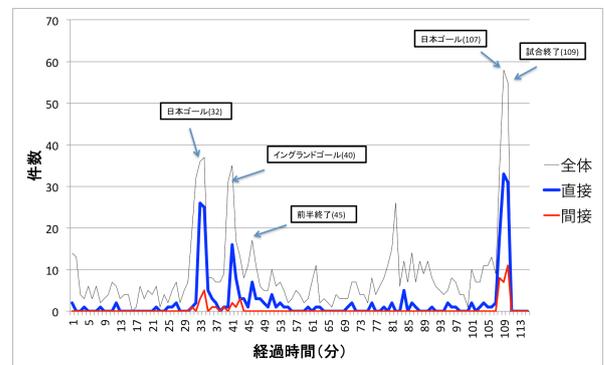


図6 1分毎のツイートの変化（「日本 vs イングランド」）

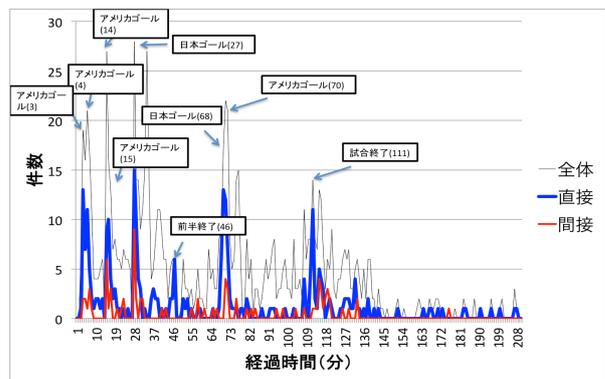


図7 1分毎のツイートの変化（「日本 vs アメリカ」）

まず図5~7から、全体のツイート数の上昇とともに直接的ネタバレと間接的ネタバレの数が上昇しており、ネタバレツイート数の上昇が全体のツイート数の上昇に影響していることがわかる。

試合展開に注目すると、直接的ネタバレと間接的ネタバレのどちらも得点が動いた時間帯には伸びているが、直接的ネタバレの方がより大きく伸びていることが分かる。また、試合終了の時間帯も得点が動いた時間帯と同じ動きをしているが、前半終了の時間帯には直接的ネタバレのみが伸びており、この時間帯では感想などの間接的ネタバレよりも、試合のスコアを記述した直接的ネタバレをツイートする傾向があることが分かる。さらに、この時間帯の傾向を見ると「日本 vs コートジボワール」の経過時間50分での1件を除き、間接的ネタバレはそれ以降点が動くまで一切登場しないということもわかる。これは、間接的ネタバレは試合の切れ目ではゴールやそれによる試合状況などを振り返るために登場し、それ以外の場所ではゴールの時間帯またはそれから数分の間で多く登場し、そのいずれもゴールやそれによる試合状況などに対して記述されたものが多いからであると考えられる。

正解の判定手法が変わると数はどう変化するかを調べるために、1票でも「間接的」に投票されていたら間接的ネタバレとみなしたときの時間的な流れを図8および9で示す。なお、「日本 vs アメリカ」では得点の動きが激しいためここでは省いた。図5, 6に比べ多少の差はあるものの、点が入るまでは間接的ネタバレと判定されていないことがわかる。一方、間接的ネタバレが前半終了から次のゴールにかけての時間帯において、少量ではあるが登場している。このことから、1人でもネタバレに遭遇しないようにするには、時間帯や試合展開(0対0以外の状況など)に応じてネタバレを厳しく判断しなければならないことがわかる。

試合開始から最初のゴールが入るまでの間、つまり得点の動きがゼロの区間でも直接的ネタバレが存在しているが、これはいずれもスコアの情報で、「0-0」という表現を含んでいるものであった。一方、間接的ネタバレについてもわずかながら登場しているが、こちらは、「ちょっとまだエンジンかかってないのかな。奪われたり、ミスする展開が続いている。」や、「日本おされてる」などのような試合展開の傾きに関するものであり、最初の点が入るまでは直接的ネタバレのみ注意すればよいことがわかる。

3試合目の「日本 vs アメリカ」では、試合終了後にも直接的ネタバレが多く登場しているが、これは他の2試合と比べ、大会の決勝戦であるため、優勝または準優勝という順位が決定する試合であり、試合終了後にその順位へのコメントが多くあったのが原因だと考えられる。順位のかかった試合では、試合終了後のこうしたツイートも防がなければならないと言える。

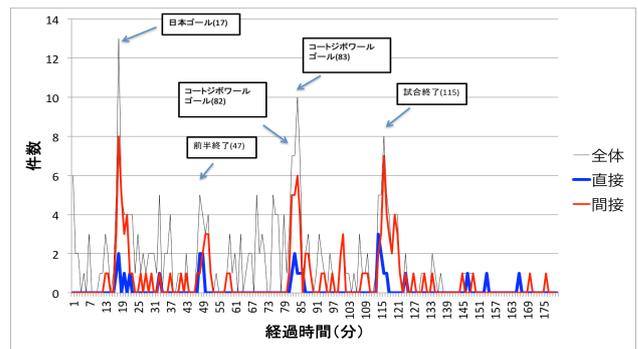


図8 1分毎のツイートの変化

(「日本 vs コートジボワール」・間接的ネタバレ:1票以上)

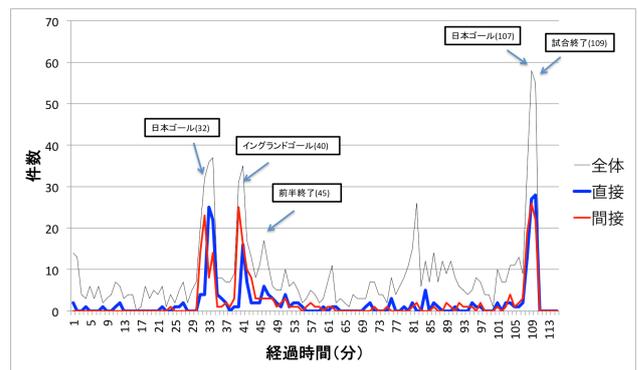


図9 1分毎のツイートの変化

(「日本 vs イングランド」・間接的ネタバレ:1票以上)

表2 バースト時間帯のネタバレの例

(括弧内の数字は試合開始からの経過時間(分))

直接的	間接的
日本ゴール (17)	
本田くんゴoooooooo!!	本田すごいよ、やっばり!
GOAL!!!! Japan!!!	本田△ ここで決めちゃうもんな〜 確実に持ってる!!!
ゴール!! 先制点!	突き刺さった(((o(*∇*)o)))
コートジボワール逆転 (83)	
2-1?!マジ?!	Just one more point!!!!
おおう。あっという間に逆転。	またさっきとおなじような形から、今度はジェルビーニョに叩き込まれて
試合終了 (115)	
敗戦。でもまだ終わってない!残り2戦ある!!!日本らしさが出ないまま試合終了。	本当にドログバにつける、ギリシャに勝ってコロンビアに引き分けないといけなくなった。やけ酒行っちゃーす
試合終了 日本1-2コートジボワール	気持ち切り替えて次にいこー!!
日本負けちゃった。。。	あああああー~~~~。残念だったなー。次は頑張ろう!!

「日本 vs コートジボワール」において、直接的ネタバレと間接的ネタバレがともに伸びている3箇所の時間帯に登場する、直接的ネタバレと間接的ネタバレの一部を表2に示す。表1の一部に記載したが、バースト時間帯以外の

コメントと見比べると、バースト時間帯以外の直接的ネタバレや間接的ネタバレは、それ以前にバーストした時間帯の出来事に対してコメントされているものが多いことがわかる。また点を決められてバーストした以降を見ると、その後に点が動くか、前半または後半が終了するまで間接的ネタバレは1件程度しか登場していない。これは、間接的ネタバレは感情を表現したものが多く、点を決められた時や直後にはショックや残念な気持ちからコメントをするが、それ以降は気持ちが萎えたために感情を表出するコメントが少なくなったのだと考えられる。

6.3 ネタバレごとの特徴語

直接的ネタバレと間接的ネタバレのツイートはどのような特徴を持つかを調べるため、直接的ネタバレと間接的ネタバレの単語の出現頻度を計算した。なお、単語には形態素解析エンジンである MeCab を用い、名詞と形容詞に絞って算出した。TF 値の上位 5 件ずつを表 3, 4 に示す。

表 3 名詞上位 5 件

直接的		間接的	
単語	TF 値	単語	TF 値
「日本 vs コートジボワール」			
日本	0.1054	本田	0.0567
コートジボワール	0.0687	点	0.0441
代表	0.0653	次	0.0420
本田	0.0435	日本	0.0210
点	0.0286	試合	
ゴール			
「日本 vs イングランド」			
イングランド	0.0706	宮間	0.1042
なでしこ	0.0696	大丈夫	0.0625
日本	0.0527	マジ	
先制	0.0414	なでしこ	
宮間	0.0376	コース	0.0417
		さん	
		イングランド	
		これ	
		日本	
		点	
「日本 vs アメリカ」			
点	0.0687	アメリカ	0.0511
アメリカ	0.0641	なでしこ	0.0483
なでしこ	0.0589	決勝	0.0227
分	0.0386	点	0.0199
優勝	0.0360	日本	

まず、名詞で直接的ネタバレにおいては、どの試合でも「ゴール」「先制」などの点の動きを表した単語や、「日本」や対戦している国名が多い。「ゴール」「先制」などはその単語自体が直接的ネタバレであるため明らかであるが、「日本」や対戦している国名についてはスコアの表記が理由と考えられる。また、間接的ネタバレでは点を決めた選手の名前が多かった。なお、「日本 vs コートジボワール」では「点」が直接的ネタバレと間接的ネタバレに共通して多く

出てきているが、直接的ネタバレでは「慎重になり過ぎてどうなるかと思ったけど先制点取った。」や、「日本1点リードのまま後半へ！」などの現在の状況や出来事を表す「点」であったが、間接的ネタバレは「ここ追加点欲しいね」や、「せめて同点に…」など次の展開への希望を表す「点」であった。

名詞は男子と女子、勝ちと負けにはあまり差はなく、トーナメントの上位に行くほど、「準決勝」や「決勝」「優勝」などそれにちなんだ単語が出ていた。

表 4 形容詞上位 5 件

直接的		間接的	
単語	TF 値	単語	TF 値
「日本 vs コートジボワール」			
ない	0.1373	欲しい	0.1569
良い	0.0980	悔しい	0.1373
早い	0.0784	ない	0.0980
いい		早い	0.0784
悪い	0.0588	素晴らしい	0.0588
甘い		良い	
「日本 vs イングランド」			
いい	0.2174	しんどい	0.125
うい	0.1084	ほしい	
ええ	0.0870	強い	
ない		素晴らしい	
強い	0.0435	凄い	
大きい		嬉しい	
ほしい		いい	
惜しい		良い	
「日本 vs アメリカ」			
いい	0.1165	悔しい	0.1961
早い	0.0874	強い	0.1765
すごい	0.0777	ない	0.0784
強い	0.0680	いい	0.0588
悔しい	0.0583	欲しい	
		素晴らしい	
		欲しい	

形容詞では、直接的ネタバレでは「良い」「いい」「悪い」など試合や選手に対する客観的な評価を表す単語が多く、間接的ネタバレでは「悔しい」「強い」「素晴らしい」など試合や選手を見た主観的な感動や感情を表す単語が多いことがわかった。

7. 考察

今回収集および分類したサッカーの 3 試合については、1 位手法で集計した結果に基づいて、直接的ネタバレと間接的ネタバレ、非ネタバレに分類することが適切であった。しかし、これは大多数の人がネタバレと感じたものを致命的なものとして捉えた場合に有効であり、時間的な試合経過における前半終了から次のゴールまでなど、ツイート数が比較的少ない時間帯については、1 票手法の導入も考えられる。実際、図 8, 9 では、図 5, 6 では出てこなかった間接

