

スポーツにおけるネタバレの特性調査と判定手法の検討

白鳥 裕士^{1,a)} 牧 良樹¹ 中村 聡史¹ 小松 孝徳¹

受付日 2017年6月11日, 採録日 2017年12月8日

概要: スポーツの録画視聴を楽しみにしている人にとって、意図せず遭遇してしまうネタバレ情報は問題であるとされている。本研究では、ネタバレが本当に問題であるのか検証を行うため、サッカーを対象としてネタバレが視聴者の観戦態度にどのような影響を及ぼすのか調査を行った。その結果、視聴者の緊張感に大きく影響することや、ネタバレの影響が大きい試合が存在することから、ネタバレが視聴者にとって問題であることを明らかにした。また、ネタバレ防止のための足がかりとして、サッカーを対象としたSNS 上でのネタバレに関するデータセットを構築することにより、ネタバレの文章特性の調査を行った。その結果、試合展開とネタバレ内容の連動性を確認した。さらに、構築したデータセットを用いてネタバレ判定実験を行い、3つの判定手法を比較した結果、SVM + 試合展開手法が優れていることが明らかになった。

キーワード: ネットバレ防止, 機械学習, スポーツ, Twitter

Analysis and Detection of Sports Spoilers on SNS

YUJI SHIRATORI^{1,a)} YOSHIKI MAKI¹ SATOSHI NAKAMURA¹ TAKANORI KOMATSU¹

Received: June 11, 2017, Accepted: December 8, 2017

Abstract: It is said that spoilers on the SNS services such as the results of sports matches spoil viewer's enjoyment when watching recorded matches. In this paper, we verified whether spoilers are a problem. It was revealed that spoilers are a problem in terms of lack of tension for viewers. Therefore, we investigated the characteristics of spoilers. We clarified the relationship between the conditions of sports matches and spoilers. In addition, we compared three methods for detecting spoilers and show the usefulness of a method that uses SVM with Condition of Matches.

Keywords: diminish spoilers, machine learning, sports, twitter

1. はじめに

スポーツは筋書きのないドラマであるため、どちらが勝つか分からないという緊張感や予想もしない試合展開に対する驚きを味わうことができる。そのため、リアルタイムで観戦したいと考えている視聴者は少なくない。しかし、仕事や学業などの時間の関係で、スポーツの試合をリアルタイムで視聴観戦することが困難な場合がある。そこで、あらかじめ録画予約をしておき、時間に余裕があるときに改めて視聴するということが珍しくない。ここで、録画視

聴を楽しみにしている人が、視聴前にそのスポーツの試合結果を知ってしまうと、緊張感や驚きが失われてしまう可能性がある。こうした緊張感や驚きを大事にしている視聴者にとって、試合のスコアや勝敗といった「ネタバレ情報」は避けたいものであるため、視聴するまでの間、情報遮断を積極的に行っている。しかしTwitterのようなSNSや、Googleのような検索サービスは気軽にアクセスすることが可能であるため、何気なく使用するユーザが多く、その際に「本田ごおおおおおおお！」のようなネタバレ情報を目にしてしまうことも少なくない。

こうしたネタバレ情報を遮断するための研究は現在さかんに行われている。たとえば、中村らは、ネタバレ情報を動的にフィルタリングすることを目的とし、ユーザの意図

¹ 明治大学
Meiji University, Nakano, Tokyo 164-8525, Japan
^{a)} swany181@gmail.com

に基づいてウェブページに含まれるユーザの興味のある情報をフィルタリングする手法や、ユーザとネタバレとの関わり方に注目し、情報の提示を工夫することでネタバレを回避する手法について提案している [1], [2]。しかし、これまで行われてきた研究では、スポーツのネタバレは視聴者にとって忌むべきものであるということを前提として防止手法の提案をしており、本当にネタバレが視聴者の視聴行為に影響を及ぼすのか、また、具体的にどのような要因が試合をつまらなくさせてしまうのかについては明らかにされていなかった。さらに、防止手法についても、これまでの研究ではユーザとネタバレとの関わり方に関するインタラクション手法に注目しており、ネタバレはどのような文章特性を持っているのか、また、高精度に判定するにはどうしたらよいかといった点については取り組まれていなかった。

そこで本研究では、スポーツのネタバレは本当にネタバレ後に視聴する楽しみを奪うのか、また、奪うとすればどのような楽しみを奪うのかを実験により明らかにする。さらに、SNS 上でのスポーツの試合についての投稿に関するネタバレデータセットを構築することで、スポーツのネタバレの文章特性について分析し、それらを高精度に判定するための手法についての検討も行う。

2. 関連研究

2.1 ネットバレルの影響調査に関する研究

ネタバレの影響調査については、Leavitt らが小説に着目し、ネタバレ情報の提示の有無によって、ユーザ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いる [3]。実験の結果、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が、小説を読むという行為とスポーツを視聴するという行為は本質的に異なるものであり、また Leavitt らの研究は、どちらかといえばあらすじ提示によって、内容や人物関係の理解を支援し、結果的に小説が読みやすくなるということを示唆する研究である。これについては、Rosenbaum らが、小説を読み慣れていない人はネタバレをされた方がストーリーを面白いと感じ、読み慣れている人はネタバレをされない方がストーリーを面白いと感じるということを実験により明らかにしている [4]。

このようにネタバレの影響を調査した研究はいくつか存在するが、スポーツにおいてネタバレがどのように影響するのかは調査されていない。そのため、スポーツにおいてネタバレ問題が存在するかどうか調査することは重要な課題である。

2.2 ネットバレル防止に関する研究

ネタバレに類する情報を遮断する研究としては、インターネット上のレビュー文に対する研究が広く行われてい

る。Ikeda らは、ストーリーコンテンツに対するレビュー文にあらすじが含まれてしまうことを問題としており、機械学習の手法を用いてあらすじを除去している [5]。Pang らは、レビュー文に対して、どの文章にあらすじではない意見が書かれているのかを SVM (Support Vector Machine) などを用いて特定している [6]。岩井ら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部分を分類・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視聴者でもレビューを閲覧できる手法を提案している [7]。コンテンツの時間的な流れに注目した前田らの研究では、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レビュー文に含まれるあらすじを判定する方法について検討している [8]。

これらのレビュー文に対する研究では、あらすじ部分をすべてネタバレとして判定しているが、スポーツではあらすじ (試合内容) が直接ネタバレとなるわけではないため、本研究で問題としているネタバレとは少し性質が異なる。

Twitter のような SNS でのネタバレを問題視した研究として、田島らはテレビアニメのようなストーリーコンテンツにおいて、放送時間差によって、SNS 上でネタバレをされてしまうことを問題としており、致命的なネタバレとなる「生死」「勝敗」などのトピックに対し、人物名一般化などの事前処理を施してから機械学習を行うことでネタバレを判定する可能性について明らかにしている [9]。同じく SNS 上でのネタバレを問題視した研究として Jeon らは、Twitter 上でのコメントに対して、「固有表現」や「頻繁に使用される動詞」「時制」などに注目した機械学習を用いてネタバレ検出をする手法を提案している [10]。同研究では、テレビ番組に関するコメントを用いて実験を行うことで、これまでのキーワードマッチングや LDA (Latent Dirichlet Allocation) を用いた手法に比べて高い適合率でネタバレを検知することを可能とし、有用性を示している。こうした研究に対し、スポーツのネタバレは勝敗や試合結果に関するものが多く、これらの研究で扱われているネタバレとは内容が異なる。

中村らはリアルタイムでスポーツの試合を視聴できないため、ネタバレを遮断したいと考える視聴者を対象に、スポーツの試合開始から視聴者の視聴開始までの間、ウェブページにおいてネタバレ情報を遮断する手法を提案している [2]。同研究では、ネタバレ情報を遮断する手法として、テキスト情報の曖昧化処理による表現手法を 4 つ提案している。さらに、Twitter を対象として、時間的にバーストする単語を抽出し、クライアントの形でネタバレを防止するシステムを実装している [11]。こうしたスポーツに対するネタバレ防止について、本研究でネタバレの影響を調査することで研究の基盤を構築する。また、スポーツのネタ

バレの文章特性についても分析し、高精度にネタバレ判定をするための手法についても検討を行う。

2.3 スポーツの観戦支援に関する研究

スポーツ観戦に関する研究としては、試合のダイジェスト生成が多く行われている。インターネットの普及により、コンテンツが多様化してきている中で、コンテンツの内容を要約して短時間で把握できるダイジェストは有用であり、自動的にコンテンツの内容を抽出しダイジェストを生成することができれば、より多くのコンテンツを消費することができる。しかしこれまでの研究では、ダイジェスト映像に得点シーンが含まれてネタバレしてしまうものばかりであり、ダイジェストのみで視聴が完結するようなものが多く、ダイジェスト映像を、試合を視聴するかどうかの指標とする人にとって、試合を実際に視聴したいと思えるようなものではなかった。ネタバレをしないようなダイジェストを生成することが可能であれば、緊張感を奪うことなく、録画した試合の面白さや価値を知ることができるのではないかと考えられるため、我々はダイジェスト作成技術に注目している。

ダイジェストとして抽出する部分の決定に関して、岡本らは、教師付き学習プログラムを利用して、その試合の持つストーリーをダイジェストで表現する手法を提案し、野球を対象としてシステムを実装している [12]。さらに、そのシステムによって生成されたダイジェストが、人手によって作成されているテレビのダイジェスト放送と同等のものになっていたことを、実験により明らかにしている。また、黒田らは、一般に放送される野球映像において、試合の展開を分かりやすく伝えるための手段としてテロップが用いられていることを利用して、野球映像に表示されているテロップからシーンの識別を行っている [13]。さらに、テロップに表示されている情報を解析し、適切に再配置することで、余分なシーンを排除し、時間を短縮したダイジェストを生成している。同様に田中らも、テロップ情報を用いて野球中継番組のダイジェストを自動生成している [14]。同研究では、イベントの発生確率からエンタロピーに基づくシーンの重要度を算出し、重要なシーンのみをダイジェストに含む手法を提案している。野球に限らず様々なスポーツに適用できるような手法として、Tjondronegoroらは、ホイッスルの音、観客の盛り上がり、テキスト情報の有無から重要なシーンを検出できる要約手法を提案している [15]。また、Zhaoらは、たいていのスポーツではリプレイシーンの前後にロゴショットが挿まれることを利用して、リプレイシーンを検出する手法を提案している [16]。

また、個人の趣味や嗜好を重視したダイジェスト生成に関して、橋本らは、テレビ受信端末で視聴者にあらかじめ好きな選手やチームを選択させておいて、映像から視聴者ごとの重要なシーンを動的に検索できるダイジェストを作

成している [17]。さらに橋本らは、ギリギリで入らなかったシュートのような不成功プレーの重要度を判定し、ダイジェストに反映できるような、ターニングポイント解析という概念を提案している [18]。ほかにも、視聴者視点を反映したものとして、Miyamoriらは、番組実況チャットに書き込まれるコメントを利用して、番組の盛り上がり場面や視聴者が注目した場面を抽出することで、シーンのランキング表示やダイジェスト生成などに視聴者の趣味や嗜好を適用できることを確認している [19]。

これらのダイジェストに関連した観戦支援研究は、いずれも重要なシーンをどのように抽出するかを検討したものである。しかし、どの試合を視聴するかの検討材料として用いる場合、これらの抽出シーンは重要なシーンであるため、視聴者はネタバレと感じてしまう可能性がある。そこで、本研究でスポーツにおけるネタバレの特性を調査することにより、こうした観戦支援にネタバレ防止を適用できる可能性がある。

3. ネタバレ影響調査実験

3.1 実験概要

本章では、実験協力者にサッカーの試合映像を視聴してもらい、ネタバレをされた人とされていない人でその試合の面白さの感じ方に変化はあるのかを実験により調査した。

観戦対象の試合は表 1 に示す 2 試合とした。協力者はサッカーの試合観戦に興味のある 19~22 歳の日本人の大学生で、前者の試合は 8 人（男性 8 人、女性 0 人）、後者の試合は 6 人（男性 6 人、女性 0 人）に協力を依頼し、ネタバレ組と非ネタバレ組に分けてそれぞれ観戦態度に変化が生じるのかを観察した（うちオーバーラップ 3 人）。視聴させるスポーツをサッカーにした理由は、全世界のファンの数が第 1 位であり、興味のある協力者を招集しやすいと判断したためである [20]。また、試合観戦に興味のある協力者に依頼したのは、興味度合いが低い視聴者は試合を視聴しようとする機会が少なく、ネタバレの影響について議論すべき対象には含まれないと判断したためである。なお、協力を依頼するにあたっては、倫理面を考慮して、協力者にはネタバレをする可能性があることについて事前に承諾を得た。

ここで、ネタバレ組に試合を視聴させる際に、リアルタ

表 1 観戦対象

Table 1 Matches for the experimental test.

試合名	スコア	開催日
国際親善試合 「日本代表 vs オマーン代表」	日本 - オマーン 4 - 0	2016/11/11
FIFAW 杯 2018 アジア最終予選 「日本代表 vs サウジアラビア代表」	日本 - サウジ 2 - 1	2016/11/15



図 1 結果表の例

Fig. 1 An example of match's results.

イムの試合をネタバレすることは不可能であるため、リアルタイムで試合が行われている間は、事前に承諾を得たうえで、スマートフォンなどのインターネットにつながる機器を回収し、試合の情報を見られない状態で待機してもらった。その後、リアルタイムの試合が終了したタイミングでネタバレ組にはネタバレをして、非ネタバレ組にはネタバレをせずに視聴してもらった。また、ネタバレ情報は、図 1 に示した結果表に記載されるような、最終スコア、前後半スコア、得点者、得点時間のみとした。

試合を視聴してもらった後、アンケートに回答してもらった。ここでは、プレーへの興奮やゴールへの喜びなど、面白さを多面的に評価するために、下記の項目について 5 段階のリッカート尺度で回答してもらった。

- プレーに興奮したか (興奮度)
- 試合の緊張感を感じたか (緊張感)
- 「悔しい」「嬉しい」など試合展開に一喜一憂したか (一喜一憂度合い)

スポーツ観戦において、多面的に面白さを評価するための確立した指標を示した先行研究は存在しない。そこで押見らの研究 [21] のサッカー観戦における感動因子分析の、8 つの因子 (「共鳴・一体感場面」「スタジアムライブ観戦場面」「ドラマ的展開場面」「卓越したプレー場面」「劣勢からの活躍場面」「懸命な姿場面」「ヒューマニティ場面」「付加的要素場面」) からまず実験環境の制約から使用できない 2 因子を排除 (= スタジアムライブ観戦場面、付加的要素場面) し、残った 6 つを以下の 3 つに集約して評価指標として選定した。

- 興奮度: 卓越したプレー場面、劣勢からの活躍場面
- 緊張感: 懸命な姿場面、ヒューマニティ場面
- 一喜一憂: 共鳴・一体感場面、ドラマ的展開場面

なお、この 3 つへの集約においては、著者間で合議を行った。

前者の「日本代表 vs オマーン代表」の試合では、協力者 8 人のうち 4 人にはネタバレした後に試合を視聴してもらい、残り 4 人にはネタバレせずに試合を視聴してもらった。後者の「日本代表 vs サウジアラビア代表」の試合では、協力者 6 人のうち 3 人にはネタバレした後に試合を視聴してもらい、残り 3 人にはネタバレせずに試合を視聴してもらった。

実験環境については、他人の様子をうかがうなど、多少



図 2 実験の様子

Fig. 2 The condition of this experiment test.

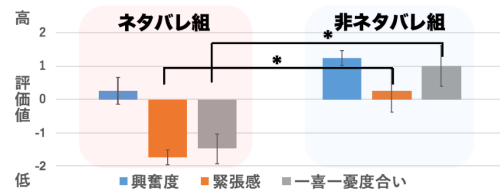


図 3 アンケート結果 (平均) 「日本代表 vs オマーン代表」

Fig. 3 Average evaluation for each question for "Japan vs. Oman".

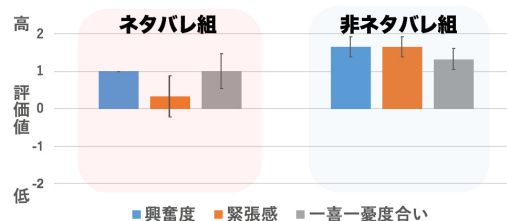


図 4 アンケート結果 (平均) 「日本代表 vs サウジアラビア代表」

Fig. 4 Average evaluation for each question for "Japan vs. Saudi Arabia".

のコミュニケーションが発生してしまう可能性があるため、ネタバレ組と非ネタバレ組を別の部屋にしたうえで、同じ組の中でもダンボールで空間を隔てた。さらに、純粋に試合の面白さのみを評価するために、コミュニケーションが発生しないよう協力者 1 人 1 人にヘッドフォン (PHILIPS 社製密閉型 SHL3260BK) を装着してもらった。実験の様子を図 2 に示す。飲食については、試合の面白さへの評価に影響しないと判断したため承認した。また、試合映像は録画したものをテレビ画面 (東芝社製 55Z700X) に映して視聴してもらった。

3.2 実験結果

それぞれの試合について、実験結果をネタバレ組、非ネタバレ組ごとに平均した結果を図 3 と図 4 に示す。図 3 に示すように「日本代表 vs オマーン代表」の試合では、緊張感と一喜一憂度合いで、非ネタバレ組に比べてネタバレ組の評価値がさわめて低かった。このことより、ネタバレが視聴観戦の楽しみを奪っていたことが示唆された。それに対し、図 4 に示す「日本代表 vs サウジアラビア代表」の試合では、非ネタバレ組に比べてネタバレ組の評価値が低

いものの、ネタバレ組の評価値もプラス評価となっていた。

ここで、興奮度、緊張感、一喜一憂度合い、それぞれの評価値に対するネタバレの効果を検討するために、対応のないt検定を行った(独立変数:ネタバレ/非ネタバレ, 従属変数:興奮度/緊張感/一喜一憂度合い)。その結果, 図3に示すように「日本代表 vs オマーン代表」の試合で, 緊張感, 一喜一憂度合いの評価値について, 非ネタバレ組よりもネタバレ組の方が有意に低いことが観察された(緊張感: $t(6) = 2.20$, $p = .04$, 一喜一憂度合い: $t(6) = 2.79$, $p = .02$)。

3.3 評価項目に注目した考察

ネタバレ組と非ネタバレ組で, 2試合ともに興奮度には有意差が観察されなかった。これは, 今回ネタバレするにあたって文字と口頭で試合の情報を伝えたが, 文字や口頭によるネタバレをしても, 実際に映像で見るプレー自体の面白さは損なわれなかったためであると考えられる。

一方, 緊張感は「日本代表 vs オマーン代表」の試合において, ネタバレすることで有意に減少した。これは, ネタバレによって試合の結果を知ってしまうと, 試合の行方にハラハラしなくなってしまうためであると考えられる。実際に, ネタバレ組の協力者から, 「チャンスの場面にはまったくドキドキしなかった」「勝ってるのが分かってるからドキドキはしなかった」といったフィードバックが得られた。

一喜一憂度合いについても緊張感と同じく, 「日本代表 vs オマーン代表」の試合で有意に減少した。これは, 得点時間もネタバレ情報として提示したため, ネタバレ組は, 試合のピンチやチャンスといえる場面で, 得点が入るかどうかを事前に悟ってしまい, 喜びや悔しさが抑制されてしまったためであると考えられる。実際に協力者から, 日本代表の試合について, 「得点場面やチャンスの場面であまり感情が動かなかった」というフィードバックが得られた。一方で, 「日本代表 vs サウジアラビア代表」の試合については, 「ピンチの場面では結末が分かっているにもかかわらず一喜一憂した」というフィードバックも得られ, 試合の重要度(大きな大会であるか否か, 勝利すれば予選突破できる試合であるか否かなど)や場面の種類によってはネタバレされても楽しめるということが示唆された。

こうした結果から, 人によってはネタバレをされてもプレー自体の面白さを楽しめたり, 場面次第で一喜一憂できたりするものの, ネタバレの影響は存在しており, 特に緊張感や一喜一憂度合いには大きく影響することが示された。

3.4 試合の種類に注目した考察

試合の種類ごとに, ネタバレによる影響の違いを考察する。まず「日本代表 vs オマーン代表」の試合では, ネタバレ組のすべての評価値が低くなっていた。さらに, 緊張感と一喜一憂度合いの2つの評価値で有意差が観察された。

この試合は日本代表が大差で勝つが, 得点場面以外は比較的落ち着いた試合展開であったため, 得点に関するネタバレをされてしまうと面白さを見い出せる場面が少なかったことが, ネタバレの影響が大きかった要因だと考えられる。

一方, 「日本代表 vs サウジアラビア代表」の試合では, すべての評価について, ネタバレ組の評価値が非ネタバレ組の評価値よりも低くなっていたものの, その下がり幅は「日本代表 vs オマーン代表」の試合よりも小さくなっている。これは, 試合内容が, ピンチやチャンスの場面が多い試合であったため, 試合のスコアが動いた場面以外でも, 協力者にとって印象に残った場面が多かったからだと推測される。また, この試合は日本が勝てばグループ1位になる可能性があるが, 負ければ予選通過圏外で折り返すことになってしまう重要な試合であった。そのため, ネタバレをすれば大きく面白さが減退するものだと推測していたが, 実際はネタバレの影響は小さいものとなった。これは, 試合のコンテキストが重要な試合であると, 選手に気迫があり, 試合内容がピンチやチャンスの場面が多いような面白い試合になりやすい傾向があるため, 視聴者はたとえネタバレをされても十分に試合を楽しめるということを示唆している。しかし, ネタバレの影響は小さいものの確かに存在し, 実際に協力者からも「ネタバレをされなければもっと楽しく観戦できたと思う」というフィードバックを得ているため, こうした試合についてもネタバレは考慮すべき問題だと考えられる。

こうした結果から, ネタバレの影響の大きさは試合によって差があり, 特に試合内容が落ち着いた試合であるとネタバレが大きく影響することが示された。

3.5 他の種類のコンテンツへの適用可能性

今回の実験は, 視聴者の興味度合いが高そうな世界的なレベルの試合を対象として行った。国内リーグや高校サッカーといった他のレベルの試合については実験を実施していないが, 地元のチームを応援している視聴者や, 高校サッカーが好きという視聴者も, それぞれの試合に対して興味度合いが高いと推測される。そのため, 視聴者が試合の視聴に臨む姿勢には変化は起こらず, 同様の結果になると考えられる。また, 他のレベルの試合では試合の質が大きく異なる可能性があるが, 試合の質によるネタバレの影響度合いの変化についてはすでに3.4節で考察している。これらのことは, コミックやアニメ, ドラマなどの作品中のサッカーの試合や, テレビゲームの実況動画におけるサッカーの試合, さらに, 野球やバスケットボール, 陸上競技といった他のスポーツにおいても, 視聴者の試合に対する興味度合いが高ければ同様のことがいえると考えられる。

また, 今回観戦対象とした試合は, 偶然にも2試合とも日本代表が勝利した試合であったが, 応援しているチームが敗北した試合は視聴する意欲が失われると考えられるた

め、日本代表が敗北した試合ではネタバレのマイナスの影響がさらに大きくなる可能性がある。このような試合の結果によるネタバレの影響度合いの違いについても今後調査していきたい。

4. ネタバレデータセット構築

前章より、ネタバレが視聴者にマイナスの影響を及ぼす傾向が見られたことから、ネタバレ防止の必要性が明らかになった。本章では、ネタバレを防止していくための足がかりとして、どういった情報がネタバレとなるのか、ネタバレの特性を分析する。ここで、視聴者がネタバレに遭遇する媒体としては、ニュースサイトやウェブログ、検索サイトなど様々なものが考えられるが、特にネタバレとの遭遇機会が多いのが Twitter のような SNS である。Twitter では、アクセスするだけで友人の現在の状況を知り、気軽にコミュニケーションをとることができるため、何気なくアクセスするユーザが多く、その際にネタバレ情報も目にしてしまう可能性が高い。そこで、本章ではスポーツの試合に対する Twitter 上の投稿を収集し、ネタバレデータセットを構築することで、ネタバレの特性を分析していく。以降この Twitter 上の投稿をツイートと呼ぶ。

データセットの構築にあたって、まずスポーツの試合に対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ようなサッカー日本代表の試合に注目した。収集した試合の情報を表 2 に示す。

表 2 収集した試合

Table 2 Matches for generating dataset.

試合名	スコア	開催日
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs イングランド代表」	日本 - イングランド 2 - 1	2015/07/01
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs アメリカ代表」	日本 - アメリカ 2 - 5	2015/07/05
EAFF 東アジアカップ 2015 「日本代表 vs 韓国代表」	日本 - 韓国 1 - 1	2015/08/05
EAFF 女子東アジアカップ 2015 「日本代表 vs 中国代表」	日本 - 中国 2 - 0	2015/08/08
EAFF 東アジアカップ 2015 「日本代表 vs 中国代表」	日本 - 中国 1 - 1	2015/08/09
FIFA ワールドカップロシア 2018 「日本代表 vs カンボジア代表」	日本 - カンボジア 3 - 0	2015/09/03
FIFA ワールドカップロシア 2018 「日本代表 vs アフガニスタン代表」	日本 - アフガニスタン 6 - 0	2015/09/08
国際親善試合 「日本代表 vs イラン代表」	日本 - イラン 1 - 1	2015/10/13
FIFA ワールドカップロシア 2018 「日本代表 vs シンガポール代表」	日本 - シンガポール 3 - 0	2015/11/12

ここで、リアルタイムなコンテンツに対してツイートをする場合、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ことがある。たとえば、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。ネタバレ対象のスポーツの試合に関するツイートに、対象スポーツ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なしでツイートされているものは多く、それらを遮断するためにはツイートの内容を分析する必要がある。

しかし、あるスポーツの試合に対するツイートをすべて集める場合、その時間にツイートされているすべてのツイートを集め、そこからスポーツ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、ツイートの内容を限定公開にしている視聴者のツイートは集めることができない。また、Twitter 社が提供している Streaming APIs ではすべてのツイートを収集することはできない。

そこで、本研究ではハッシュタグ付きのツイートとハッシュタグなしのツイートは、ハッシュタグの有無に違いはあるものの、そのツイート内容に大きな差は存在しないと考え、収集効率を優先し、ハッシュタグ付きのツイートを収集することとした。ここでは、サッカーの日本代表の試合で一般的に用いられる、「#daihyo」や「#JPN」のようなハッシュタグを試合開始前に選定しておき、そのハッシュタグを含むツイートを Twitter 社の提供している The Search API を用いて収集した。なお、収集においては、試合開始から試合終了後 2 時間までの間ツイートを収集した。

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かった。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った。

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投稿されるため、収集されたツイートは多言語となる。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した。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した。
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視聴者のツイートをそのままの形でツイートできる機能で、他の視聴者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した。
3. 収集したツイートからハッシュタグを除去した。これについては、「#」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した。また、中

身のないツイートについても正規表現により削除した(ハッシュタグのみのツイートがあるため)。ここではツイートの先頭から文末までの間、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。

4. スポーツのハッシュタグに対して URL を投稿しているツイートに試合とは無関係のスパムツイートが多かったため、「http://t.co/」もしくは「https://t.co/」を含むツイートを正規表現により判定し除去した。

上記の処理を行った後、ツイートをネタバレと非ネタバレに分類してもらうために、図 5 に示すウェブシステムを開発した。このシステムは、ページ上に提示されているツイートがネタバレだと感じたらクリックすることでネタバレ、チェックされていない場合は非ネタバレに分類される。また、ツイートからネタバレを感じるタイミングとして、独立したツイート以外にも、「香川打て!」と「ナイス!!!」が連続して投稿されていたときや、「ゴ」「ー」「ル」が連続して投稿されていたとき、急にその試合に対するツイート数が増えていたときなども考えられるが、分類時の提示方法や判断基準が複雑かつ困難なることを考慮し、本研究では独立したツイートをネタバレ判断基準とした。なお、ツイートすべてを分類すると膨大な時間がかかってしまうため、提示するツイートは 1 試合につ

き任意の 1,000 件とした。ここで、1 試合ずつ提示してしまうと、試合の詳細情報を知っている人が分類をする場合に、1 つのツイートから試合の前後関係や内容が鮮明に伝わってしまう恐れがある。たとえば、香川選手が得点を決めた試合を分類してもらおうと仮定して、「2018W 杯予選日本 vs シンガポール戦開始!」のような、対象としている試合が把握できてしまうツイートをその試合の結果を知っている人が見してしまうと、「かがわー」といった、試合の詳細を知らない人はネタバレとはいい難いツイートまでネタバレと判断してしまう可能性がある。これは、実際に録画視聴を楽しみにしている視聴者を想定した場合、視聴者は試合の事前情報を知ることはあっても、試合内容までは知ることはないはずであるので好ましくない。そこで、9 試合を 3 試合ずつまとめて、ランダムに提示することとした。つまり、合計 3,000 件のツイートが 3 グループできることになる。

また、視聴者は試合開始からのある程度の経過時間については、視聴していなくても把握することができる。経過時間が影響してしまう例として、「守備固めてもしょうがない、攻めていこう」というツイートがあったとすると、これが試合開始時点であれば、単なる意気込みととらえる視聴者が多いと推測されるが、試合後半であれば、試合に負けているという状況を考慮したうえでのコメントととらえる視聴者が多いと推測され、ネタバレの判断に必要であると考えた。そこで、ツイートは時間順ではなくランダムに提示され、各ツイートの下には、そのツイートがされた試合開始からのおおよその経過時間を表示した。たとえば、ツイートの下に「60」と表示されていたら試合開始 51 分から 70 分までのツイートとなっている。また、おおよその時間にした理由は、実際に録画視聴をしている視聴者を想定した場合に、視聴者は細かな時間までは知りえないという点を考慮したためである。

ウェブシステムにアクセスすると、1 ページごとに 50 件ずつツイートが表示される。また、1 グループあたり 60 ページの分量となっている。図 5(b) のようにクリックすると背景色が赤に変わり、選択済みであることを示す。

データセット構築者は、1 グループにつき 5 人で、サッカーの試合観戦に興味があり、Twitter をふだんから利用している 19 歳から 22 歳の大学生とした。

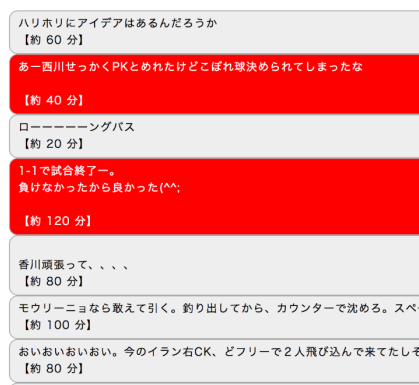
5. ネタバレデータセットの分析

本章では、前章で構築したデータセットの内容について分析する。

データセット構築結果および、分類されたツイートのネタバレ率をそれぞれ表 3、表 4 に示す。5 人全員がネタバレと判断したツイートは、「完勝だな!これで首位」「試合終了。イラン 1-1 日本」といった、最終的な試合結果を述べているものが多かった。これは、試合の最終



(a) 全体



(b) 拡大

図 5 開発したウェブシステム

Fig. 5 Screenshot of the evaluation system.

表 3 データセット構築結果
Table 3 Examples of the dataset.

ツイート	経過時間	分類
おおお！ 香川がゴール！！！！	20	ネタバレ
既に2失点(‘D’)	0	ネタバレ
さあキックオフ	0	非ネタバレ
ぬーんやはり パスミスは駄目だわ	20	非ネタバレ

表 4 ネットバレットツイート一致率
Table 4 Concordance rate of judging spoilers.

ネタバレ 一致人数 (人)	ツイート数 (件)	一致率 (%)	ネタバレ内 一致率 (%)
5	351	3.90	14.03
4	680	7.56	27.18
3	620	6.89	24.78
2	217	2.41	8.67
1	634	7.04	25.34
0	6498	72.20	-

的な結果は誰もがネタバレと感じているためだと考えられる。3人および4人がネタバレと判断したツイートは、「西川よく止めた！先制点取られたのは悔しいけど、めっちゃかっこよかった！」「原口の仕掛けから香川の先制ゴール！1vs0」といった、ゴールなどの試合が動いたことを述べているものが多かった。これは、試合の最終的な結果のみを重要と考え、それ以外の重要な場面をネタバレと考へない視聴者も一定数存在することを表している。また、1人および2人のみネタバレと判断したツイートは、「吉田、あまりに不用意な反則。しかも選手は決められるもん決めてつけてこぼれ球に反応できない失態。ひどすぎる。」といった、間接的に得点が決められたことを述べているものや、「あのウサミンのシュートはセンスを感じる。」といった、重要度がそこまで高くはない場面について述べたものが多かった。これは、ツイートを熟読する度合いや対象スポーツへの精通度合い、また、ネタバレへの敏感度合いが視聴者によって異なるためだと考えられる。なお、0人がネタバレと判断したツイート（誰もネタバレと判断しなかったツイート）は、「ハリルホジッチ笑顔」といった試合内容とは関係のないものや、「岡崎頑張れー！」といった単純な応援メッセージ、「本田からだはってるwけがしないだね。」といった重要度が低い場面について述べたものなど様々なツイートが存在した。

上記の理由から今回は、半数以上、つまり3人以上がネタバレと判定したものをネタバレツイートとして分析していく。つまり、ここでは分類したツイート9,000件のうち、

表 5 表5 試合展開別ネタバレ頻出単語
Table 5 Terms used frequently in spoiler tweets.

勝ち		負け		同点	
単語	TF-IDF	単語	TF-IDF	単語	TF-IDF
[選手]	0.742	[選手]	0.531	[選手]	0.627
[チーム]	0.422	[チーム]	0.475	[チーム]	0.552
ゴール	0.261	先制	0.238	ゴール	0.226
[数字]点目	0.173	失点	0.238	同点	0.201
[数字]点	0.131	止める	0.224	試合	0.151
[数字]	0.117	後半	0.112	[数字]-[数字]	0.136
勝つ	0.106	取る	0.112	終了	0.125
試合	0.099	すぎる	0.112	代表	0.110
点	0.095	られる	0.112	前半	0.105
先制	0.083	最後	0.098	引き分け	0.100

ネタバレツイートは1,651件ということになる。

ネタバレツイートはどの試合も共通して、「○-○」「勝つ」「○点」など、特定のパターン記述や単語が含まれているものが多かった。また、勝っているときは「キター」「嬉しい」、負けているときは「悔しい」「最悪だ」など、試合展開によって内容が異なっていた。

そこで、日本代表が勝っている時間帯、負けている時間帯、同点である時間帯別に、非ネタバレツイートと比較してネタバレツイートに出現する頻度 (TF-IDF [22]) が高かった単語上位10件を表5に示す。なお、単語分割にはMeCabを用い、1文字の連続した名詞は連結（「憲」「剛」を「憲剛」,[2]「点」「目」を「2点目」などとして処理するため）し、単語として意味を持たない助詞、助動詞や、名詞以外の基本形が辞書に定義されていない1文字の単語（「。」「▽」などは単語として意味を持たないため）、その他単一では意味を持たないと判断した「する」「てる」「の」「ん」「いる」「れる」「なる」「さん」は除去した。また、「押し込んだああああ」「キターー」といった繰り返し表現がノイズとなるため、「あ」「い」「う」「え」「お」「あ」「い」「う」「え」「お」「～」「一」「-」「ー」「-」「!」「!」「w」が2回以上繰り返されていた場合、その部分を除去した。さらに、岩井ら[23]の固有名詞などの一般化手法を適用し、数字は[数字]として機械的に一般化し、選手名は[選手]、チーム名は[チーム]、監督名は[監督]としてパターンマッチにより一般化した。

表5より時間帯別で見ると、負けている時間帯の頻出単語は「[選手]」「[チーム]」「先制」という単語以外、勝っている時間帯や同点である時間帯と異なっていた。また、勝っている時間帯と同点の時間帯についても、「[選手]」「[チーム]」「ゴール」「試合」という単語以外は異なっていた。勝っている時間帯では「勝つ」、同点の時間帯では

「同点」のような、試合展開を直接表すような単語が頻出していった。さらに、勝っている時間帯では「[[数字] 点目]」「[[数字] 点]」「点」のような得点に関する単語、同点である時間帯では「[[数字] - [数字]]」のようなスコア情報に関する単語、負けている時間帯では「失点」のような失点に関する単語が頻出しており、時間帯でネタバレ内容が異なっていることが示唆された。

6. ネタバレ判定実験

6.1 実験手順

本章では、構築したデータセットを用いて、ネタバレを高精度に判定するための手法についての検討を行った。前章より、スポーツのネタバレは単語に大きく特徴が表れていたため、本研究では単語をベースとした、下記のパターンマッチ手法、SVM 手法、SVM + 試合展開手法という 3 つの手法を比較した。

- **パターンマッチ手法：** ネタバレとして出現頻度の高い単語をキーワードとし、キーワードにマッチする単語を含むツイートをネタバレと判定した。ここでは、前章で時間帯別に分析したときのルール（1 文字の連続した名詞は連結など）で単語分割し、TF-IDF 値が 0.100 以上の単語をキーワードとした。ここで、0.100 を閾値に設定したのは、0.000 から 0.300 まで 0.050 ずつ閾値を変化させて分析した結果、0.100 で最も F 値が高かったためである。
- **SVM 手法：** 判定対象となる試合以外の試合のツイートを用いて SVM のモデルを作成し、そのモデルを用いて判定対象となる試合のツイート（1,000 件のテストデータ）を判定した。モデルの作成にあたっては、ネタバレツイートよりもネタバレでないツイートの数が多くなってしまったため、アンダーサンプリングを行ってデータ量を調整した（訓練データ数は判定対象となる試合の種類によって異なるため、表 6 に示す）。また、SVM のための特徴ベクトルは、各ツイートの BoW (Bag-of-Words) [24] で生成した。なお、単語分割は前章で時間帯別に分析したときのルールで行った。そのうえで、グリッドサーチを行い、学習率 0.01 の線形カーネルを学習モデル作成時のパラメータとして設定した。また、それぞれの次元（単語）のスケールを同じに揃えるため、標準化処理も行った。
- **SVM + 試合展開手法：** 前章より、試合展開によって大きく出現単語が異なっていたため、SVM のモデルの作成において試合展開を考慮し、判定対象となる試合以外の試合のツイートのうち、日本代表が勝っている時間帯のツイートでは勝ちモデル、負けている時間帯では負けモデル、同点の時間帯では同点モデルと 3 つの学習モデルを作成した（それぞれのモデルの訓練データ数は判定対象となる試合の種類によって異なるため、表 7 に示す）。そして、判定対象となる試合のツイートを判定する際に、そのツイート時点で日本代表が勝っていれば勝ちモデルを、負けていれば負け

表 6 SVM 手法の訓練データ数

Table 6 The number of training data for the SVM-method.

試合名	訓練データ数
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs イングランド代表」	3248
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs アメリカ代表」	3240
EAFF 東アジアカップ 2015 「日本代表 vs 韓国代表」	3704
EAFF 女子東アジアカップ 2015 「日本代表 vs 中国代表」	3544
EAFF 東アジアカップ 2015 「日本代表 vs 中国代表」	3710
FIFA ワールドカップロシア 2018 「日本代表 vs カンボジア代表」	3424
FIFA ワールドカップロシア 2018 「日本代表 vs アフガニスタン代表」	3350
国際親善試合 「日本代表 vs イラン代表」	3624
FIFA ワールドカップロシア 2018 「日本代表 vs シンガポール代表」	3596

表 7 SVM + 試合展開手法の訓練データ数

Table 7 The number of training data for the method of SVM with condition of match.

試合名	勝ちモデル	同点モデル	負けモデル
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs イングランド代表」	1704	674	870
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs アメリカ代表」	2106	954	180
EAFF 東アジアカップ 2015 「日本代表 vs 韓国代表」	2106	764	834
EAFF 女子東アジアカップ 2015 「日本代表 vs 中国代表」	1794	880	870
EAFF 東アジアカップ 2015 「日本代表 vs 中国代表」	2106	770	834
FIFA ワールドカップロシア 2018 「日本代表 vs カンボジア代表」	1624	930	870
FIFA ワールドカップロシア 2018 「日本代表 vs アフガニスタン代表」	1528	952	870
国際親善試合 「日本代表 vs イラン代表」	2106	756	762
FIFA ワールドカップロシア 2018 「日本代表 vs シンガポール代表」	1774	952	870

るため、表 7 に示す）。そして、判定対象となる試合のツイートを判定する際に、そのツイート時点で日本代表が勝っていれば勝ちモデルを、負けていれば負け

表 8 SVM + 試合展開手法のテストデータ数

Table 8 The number of test data for the method of SVM with condition of match.

試合名	勝ちモデル	同点モデル	負けモデル
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs イングランド代表」	328	672	0
FIFA 女子ワールドカップカナダ 2015 「日本代表 vs アメリカ代表」	0	12	988
EAFF 東アジアカップ 2015 「日本代表 vs 韓国代表」	0	897	103
EAFF 女子東アジアカップ 2015 「日本代表 vs 中国代表」	345	655	0
EAFF 東アジアカップ 2015 「日本代表 vs 中国代表」	0	838	162
FIFA ワールドカップロシア 2018 「日本代表 vs カンボジア代表」	842	158	0
FIFA ワールドカップロシア 2018 「日本代表 vs アフガニスタン代表」	932	68	0
国際親善試合 「日本代表 vs イラン代表」	0	797	203
FIFA ワールドカップロシア 2018 「日本代表 vs シンガポール代表」	855	145	0

モデルを、同点であれば同点モデルを適用して判定した（それぞれのモデルのテストデータ数は判定対象となる試合の種類によって異なるため、表 8 に示す）。単語分割やアンダーサンプリング、SVM のパラメータ（学習率）やカーネルについては SVM 手法と同様に行った。また、本手法では時間帯ごとに判定するため、テストデータが極端に少ない試合やまったく存在しない試合があった。テストデータにおけるネタバレツイート数が 20 件以下の場合、例外的なツイートが大きく結果に影響してしまうと考えられるため、結果から除外した（0 件の場合も適合率などの結果が算出不可能であるため除外した）。なお、それぞれのツイート時点での試合の状況を判断する時間が必要であるため、試合中のツイートの表示に多少の遅延を持たせる必要があることや、国内リーグなど応援するチーム（または選手）がどちらかに決定しにくい場合、勝ちモデルと負けモデルを同時に適用させる必要がある時間帯が存在することを前提としている。

以上の 3 つの手法について、適合率、再現率、F 値で比較した。また、3 つの手法すべてにおいて、実験は 9 試合分（データセットの試合数）試行し、9 試合の平均を結果として算出した。ただし、それぞれの手法において、適合率は「ネタバレと判定したツイートのうち正しくネタバレと判定できたツイートの割合」、再現率は「実際にネタバレであるツイートのうち正しくネタバレと判定できたツイー

表 9 手法ごとのネタバレ判定結果

Table 9 Accuracy in detecting spoilers for each method.

	適合率	再現率	F 値
パターンマッチ	0.270	0.668	0.372
SVM	0.617	0.601	0.598
SVM+試合展開	0.698	0.565	0.611

トの割合」を表し、F 値は式 (1) で表されるものとする

$$F \text{ 値} = \frac{2 \cdot (\text{適合率}) \cdot (\text{再現率})}{(\text{適合率}) + (\text{再現率})} \quad (1)$$

6.2 実験結果

それぞれの試合について、パターンマッチ手法、SVM 手法、SVM + 試合展開手法で判定した結果の適合率、再現率、F 値の平均を表 9 に示す。

表 9 に示すように、F 値は SVM + 試合展開手法が他の手法よりも高い結果となった。また、適合率は SVM + 試合展開手法が最も高いが、再現率は SVM 手法が最も高い結果となった。

6.3 考察

SVM + 試合展開手法の F 値が最も高く、他の手法よりも優れているという結果となった。特に、適合率が他の手法よりも優れていた。パターンマッチ手法よりも SVM 手法や SVM + 試合展開手法の適合率が優れていたのは、パターンマッチ手法では「芝が走らないから香川のパスミスが怖い」のようなツイートで、「香川」という選手名のみでネタバレと判断されてしまうケースが多かったが、SVM 手法では特定の選手名のみではネタバレと判断されず、「ゴール」などといった単語が同時に存在することでネタバレと判断されるためであると考えられる。SVM 手法よりも SVM + 試合展開手法の適合率が優れていたのは、SVM 手法で間違っして学習されてしまっていたツイートが、時間帯別によって学習されなくなったからだと考えられる。実際に、勝っている時間帯の「久しぶりにスッキリした」「久しぶりの快勝劇を見た。」といったツイートが正しく判定可能となっていたが、これは SVM 手法では、試合開始時点の同点である時間帯における「トップ下香川久しぶりかな?」「ものすごい久しぶりに代表戦観てる~!!」といったツイートが非ネタバレとして学習されてしまうが、SVM + 試合展開手法では学習されなかったためだと考えられる。

それに対して、再現率は SVM 手法の方が SVM + 試合展開手法よりも優れていた。これは、SVM + 試合展開手法では時間帯によって学習するデータも 3 つに分割するため、単純に SVM 手法よりも学習データが少なくなったことが原因だと考えられる。実際に、SVM 手法では正しくネタバレと判断できていた「森重押し込んだっ」や「押し

表 10 SVM+試合展開手法におけるモデルごとのネタバレ判定結果
 Table 10 Accuracy of detecting spoilers with SVM with condition of match for each model.

	適合率	再現率	F 値
勝ちモデル	0.716	0.646	0.664
同点モデル	0.656	0.528	0.585
負けモデル	0.773	0.315	0.447

込んだあああああ」といったツイートが SVM + 試合展開手法では非ネタバレと判断されていたが、テストデータでは勝っている時間帯で「押し込む」という単語が出現しているにもかかわらず、学習データでは同点の時間帯のみ出現するため、SVM + 試合展開手法ではネタバレとして学習されなかったことが原因だと考えられる。そのため、SVM + 試合展開手法については、学習データの試合数を増やすことで再現率が改善され、さらに F 値が向上する可能性がある。また、再現率についてはパターンマッチ手法が最も優れていたが、TF-IDF の閾値を 0.100 以上にする、試合間で共通のネタバレとなる単語が多く存在するためだと考えられる。しかし、その分だけ適合率が低い結果となっている。

結果として 3 つの手法いずれも F 値がそこまで高くならなかった。その原因としては、「本田のゴールが見たい」「勝つぞ」といったツイートがネタバレと判定されてしまっていることが考えられる。このような希望や意気込みなどの未来へ向けた発言は非ネタバレとして処理する必要があるが、日本語は未来時制がなく文法から判断することは不可能であるため、形態素解析とは異なる手法を用いる必要がある。この未来へ向けた発言の判定手法の検討については今後の課題としたい。また、SVM を用いた 2 手法において訓練データ数が少なかったことも F 値がそこまで高くならなかった原因として考えられる。特に、表 7 に示したように、SVM + 試合展開手法の負けモデル作成時の訓練データ数が少なかった。実際に、表 10 に示す SVM + 試合展開手法におけるモデルごとのネタバレ判定結果を見ると、勝ちモデルの F 値 0.664 に対して、負けモデルの F 値は 0.447 となっており、訓練データ数が影響した可能性を示唆している。

7. おわりに

本研究では、スポーツのネタバレが視聴者の観戦態度に及ぼす影響を調査し、ネタバレが視聴者の緊張感や一喜一憂度合いに影響すること、得点シーン以外のピンチやチャンスが少ない落ち着いた試合ではネタバレの影響が大きいことを確認した。また、スポーツの試合に対するツイートのネタバレデータセットを構築することでネタバレの特性を調査した。データセットを分析した結果、ネタバレの内容は試合展開によって異なることを明らかにした。さらに、

ネタバレの判定精度について、パターンマッチ手法、SVM 手法、SVM + 試合展開手法で比較した結果、SVM + 試合展開手法が他の手法よりも F 値が優れていることを示した。

今後は、学習データを増やしたりデータの事前処理をさらに工夫したりすることで、ネタバレの判定精度を向上させていき、実際にネタバレ防止をクライアントなどの形でシステム化することを考えている。また、本研究では独立したツイートをネタバレ判断基準としたが、前後のツイートやツイート数といった要素も視野に入れたネタバレ特性調査やネタバレ判定手法を検討していきたい。さらに、他のスポーツジャンルへの適用実験も行っていく予定である。

関連研究の 2.3 節でも触れたとおり、スポーツの試合のダイジェスト生成においてはネタバレの問題が大きい。そこで今後の研究では、本研究で取り組んでいるツイートのネタバレ判定技術を応用し、どのシーンが最もネタバレ度が高いかなどを考慮して、ネタバレをせずにそのスポーツの試合を見たいと感じさせるようなスポーツの試合のダイジェスト生成技術を実現していく予定である。

謝辞 本研究の一部は JST CREST, JST ACCEL (グラント番号 JPMJAC1602) の支援を受けたものである。

参考文献

- [1] Nakamura, S. and Tanaka, K.: Temporal Filtering System for Reducing the Risk of Spoiling a User's Enjoyment, *Proc. IUI 2007*, pp.345–348 (2007).
- [2] 中村聡史, 小松孝徳: スポーツの勝敗にまつわるネタバレ防止手法の検討, 情報学会論文誌, Vol.54, No.4, pp.1402–1412 (2013).
- [3] Leavitt, J.D. and Christenfeld, N.J.S.: Story Spoilers Don't Spoil Stories, *Psychological Science* (2011).
- [4] Rosenbaum, J.E. and Johnson, B.K.: Who's afraid of spoilers? Need for cognition, need for affect, and narrative selection and enjoyment, *Psychology of Popular Media Culture*, Vol.5, No.3, pp.273–289 (2016).
- [5] Ikeda, K., Hijikata, Y. and Nishida, S.: Proposal of Deleting Plots from the Reviews to the Items with Stories, *Proc. SNSMW'10*, CDROM, Vol.6193, pp.346–352 (2010).
- [6] Pang, B. and Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *Proc. ACL'04*, pp.271–278 (2004).
- [7] 岩井秀成, 池田 郁, 土方嘉徳, 西田正吾: レビュー文を対象としたあらすじ分類手法の提案, 電子情報通信学会論文誌 D, Vol.J96-D, No.5, pp.1222–1234 (2013).
- [8] 前田恭佑, 土方嘉徳, 中村聡史: ストーリー文書内のネタバレの記述に関する調査とレビュー文書でのネタバレ検出の試み, 第 8 回 Web とデータベースに関するフォーラム論文集, Vol.2015, pp.32–39 (2015).
- [9] 田島一樹, 中村聡史: ストーリーコンテンツに対するネタバレの基礎調査とその判定手法の検討, 研究報告グループウェアとネットワークサービス (GN), Vol.2015-GN-96, No.7, pp.1–6 (2015).
- [10] Jeon, S., Kim, S. and Yu, H.: Don't Be Spoiled by Your Friends: Spoiler Detection in TV Program Tweets, *Proc. 7th International Conference on Weblogs and Social Media (ICWSM)* (2013).
- [11] 中村聡史, 川連一将: スポーツのネタバレを防止する

- Twitter クライアントの開発と諸検討, 第4回 ARG Web インテリジェンスとインタラクション研究会 (2014).
- [12] 岡本道也, 鎌原淳三, 植田和憲, 下條真司, 宮原秀夫: シナリオテンプレートによるスポーツダイジェスト自動生成機構, 電子情報通信学会論文誌, Vol.85, No.8, pp.1269–1276 (2002).
- [13] 黒田智也, 椋木雅之, 浅田尚紀: テロップを利用した野球映像のダイジェスト生成, 電子情報通信学会技術研究報告, Vol.107, No.129, pp.73–78 (2007).
- [14] 田中孝志, 小方善貴, 片岡充照, 黒木修隆, 沼 昌宏: 野球中継番組におけるテロップ情報を用いたダイジェスト映像自動生成, 映像情報メディア学会技術報告, Vol.33, No.54, pp.25–28 (2009).
- [15] Tjondronegoro, D., Chen, Y.P.P. and Pham, B.: Integrating highlights for more complete sports video summarization, *IEEE Trans. Multimedia*, Vol.11, No.4, pp.22–37 (2004).
- [16] Zhao, F., Dong, Y., Wei, Z. and Wang, H.: Matching logos for slow motion replay detection in broadcast sports video, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.1409–1412 (2012).
- [17] 橋本隆子, 白田由香利, 真野博子, 飯沢篤志: TV 受信端末におけるダイジェスト作成及び視聴システムの試作, 電子情報通信学会技術研究報告, Vol.99, No.202, pp.7–12 (1999).
- [18] 橋本隆子, 白田由香利, 飯沢篤志, 北川博之: ターニングポイントの解析に基づくダイジェスト作成方式, 情報処理学会論文誌 (データベース), Vol.43, No.5, pp.1–11 (2002).
- [19] Miyamori, H., Nakamura, S. and Tanaka, K.: Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web, *Proc. ACM Multimedia' 2005*, pp.853–861 (2005).
- [20] Top 10 Most Watched Sports In The World, available from (<http://top-10-list.org/2010/10/04/10-most-watched-world-sports/>) (accessed 2017-01-27).
- [21] 押見大地, 原田宗彦: スポーツ観戦における感動場面尺度, Vol.2, No.2, pp.163–178 (2010).
- [22] Baeza-Yates, R.A. and Ribeiro-Neto, B.A.: *Modern information retrieval: the concepts and technology behind Search, 2nd Edition*, Addison-Wesley Professional, p.944 (2011).
- [23] 岩井秀成, 池田 郁, 土方嘉徳, 西田正吾: レビュー文を対象としたあらすじ分類手法の提案, 電子情報通信学会論文誌 D, 情報・システム, Vol.96, No.5, pp.1222–1234 (2013).
- [24] Manning, C.D. and Schtze, H.: *Foundations of Statistical Natural Language Processing*, The MIT Press, p.728 (1999).



白鳥 裕士 (学生会員)

1994 年生. 2017 年明治大学総合数理学部先端メディアサイエンス学科卒業. 現在, 同大学大学院先端数理科学研究科博士前期課程在学中. スポーツのネタバレ防止やグループ型音楽鑑賞等の研究活動に従事. 学士 (理学).



牧 良樹 (正会員)

士 (理学).

2017 年明治大学総合数理学部先端メディアサイエンス学科卒業. 現在, 同大学大学院先端数理科学研究科博士前期課程在学中. コミックにおけるネタバレの影響, 能動的音楽鑑賞支援, 没入型動画鑑賞等の研究活動に従事. 学



中村 聡史 (正会員)

1976 年生. 2004 年大阪大学大学院工学研究科博士後期課程修了. 同年独立行政法人情報通信研究機構専攻研究員. 2006 年京都大学大学院情報科学研究科特任助手, 2009 年同特定准教授, 2013 年明治大学総合数理学部先端メディアサイエンス学科准教授, 現在に至る. サーチとインタラクションや, ネットバレ防止技術, 平均手書き文字等の研究活動に従事. ヒューマンインタフェース学会等の会員. 博士 (工学).



小松 孝徳 (正会員)

1997 年芝浦工業大学工学部卒業. 2003 年東京大学大学院総合文化研究科博士課程修了. 博士 (学術). 2003 年公立ほこだて未来大学システム情報科学部助手. 2007 年信州大学ファイバーナノテク国際若手研究者育成拠点助教, 2012 年信州大学繊維学部准教授, 2013 年明治大学総合数理学部先端メディアサイエンス学科准教授, 現在に至る. 人間の「いい加減な能力」という視点から, 人間と人工物とのインタラクションを考察する研究活動に従事. 日本認知科学会, 人工知能学会, ACM 各会員.