

# 野球における実況ツイートを用いた 試合の山場の見逃し防止手法の検討

木頃伸哉<sup>†1</sup> 白鳥裕士<sup>†2</sup> 中村聡史<sup>†1</sup>

**概要**：プロ野球は1試合の平均時間が3時間以上あり、サッカーなどの他のスポーツと比べて試合時間が長い。そのため、テレビやインターネットの生放送での観戦では、入浴や食事、仕事と重なって視聴を断念してしまったり、一方的な試合展開による視聴意欲の低下により視聴を途中で断念してしまったりする視聴者も存在する。視聴を途中で断念した後に、に試合展開が大きく変わり面白い場面を見逃すこともあり、視聴者が後悔してしまうことも珍しくない。そこで、我々は野球の試合において山場山場となりそうなシーンになったときに視聴者を、視聴誘導するシステムの構築を試みる。本研究ではその足がかりとして、野球における山場となりそうな場面を推定する手法を検討する。本研究では、野球の試合に並行して進行する実況ツイートの中に、期待にまつわる特徴語が現れることに着目した手法を検討した。また、その手法の可能性を検証するため、山場になりそうな場面に関する正解データセットを構築し、複数の手法を検討することにより精度の検証を行った。その結果、全体的な精度は低いものの、ツイートをそのまま扱うのではなく、どちらのチームのファンなのかという点に注目して判定すると精度が上がるのが分かった。

**キーワード**：Twitter, スポーツ, 野球, 見逃し防止, 期待

## 1. はじめに

スポーツは、ときに誰にも予想のつかない展開を見せることがあり、そのような場面に遭遇することがスポーツ観戦の楽しみの一つである。Shiratoriら[1]の研究でも明らかのように、どちらが勝つかわからないといったハラハラ感をもって視聴したいユーザは多くいる。人気のあるスポーツの多くは視聴するのに時間がかかるものも多いが、例えばサッカーやラグビーでは試合時間が2時間以内であるのに対し、日本のプロ野球では2018年の平均試合時間が3時間13分<sup>\*1</sup>と特に長いため、テレビやインターネットの観戦では、入浴や食事、仕事など様々な理由で視聴を中断してしまう視聴者も存在する。また、思い入れがそこまで強くないファンの場合、特にプロ野球などでは明日も試合があるからと、試合を視聴し忘れることも珍しくない。さらに、試合序盤に大差がつくなど、一方的な展開の試合であると、途中で視聴を断念してしまう視聴者もいる。ここのでも、途中で視聴を中断した試合や、そこまで理由がなく視聴しなかった試合で、応援しているチームが逆転劇を演じて勝利したとしたら、リアルタイムで視聴しなかったことを後悔することになるだろう。

こうした後悔をなくするための手段として、視聴を断念した後も、スマートフォンやPC等でSNS上の実況コメントや、ウェブ上の速報を随時チェックし、チャンスの場面で試合に復帰することが考えられるが、スマートフォンやPCから目を離さないようにする必要があるので、視聴者の負担が大きい。また、どうしても得点した場合や、逆転した

場合など、一番いいシーンを逃してしまう可能性がある。つまり、ネタバレを防止するために視聴していないのではなく、単純に忙しいから・序盤で大差がついてしまって少し興味がないから・やや思い入れが少ないからなどの理由で試合を視聴していないユーザに対し、後で「あのシーンをリアルタイムで見たかった」と後悔させてしまうようなことを防ぐため、そういったシーンの直前に視聴誘導する仕組みが重要であるといえる。以上を踏まえ、我々の大きな目標は、「この試合は視聴しなくてもいいかな」と考えて視聴をしていない、または視聴をやめてしまったユーザに対し、「この試合ちょっと面白い感じになりそうなので見てはどうか？」という通知を行い、視聴誘導するアプリケーションを実現することである。

こうした仕組みを実現するにあたり、まず我々は野球に注目し、中でも得点が入ったといった盛り上がった場面ではなく、「これから盛り上がりそうだ」「これから期待できそうだ」という場面をリアルタイムに推定する手法を検討する。ここで、野球は他のスポーツに比べ進行が遅いため、選手の成績や走者の情報などメタデータを用いたものも考えられるが、そもそもそのチームの状態（順位や残り試合数など）や各選手のその日の成績や通算対戦成績、得点圏での成績など考慮する要素が膨大になってしまうためメタデータだけから正確に推定するのは難しいと考えられる。また、その視聴者が楽観的なのか悲観的なのか、どのチームのファンなのか、誰のファンなのかなどによっても、チャンスと思えるシーンは異なってくるであろう。

そこで本研究では、プロ野球の試合を見ながらリアルタイムにツイートし、他者と情報の共有を図っている実況ツイートに注目し、そのツイートからこれから何かが起こるかも期待できるシーン（他の視聴者が試合で何かが起こると期待しているシーン）を推定する手法を検討する。プ

<sup>†1</sup> 明治大学

Meiji University

<sup>†2</sup> 明治大学大学院

Graduation School, Meiji University

\*1 <http://npb.jp/statistics/2018/time.html>

口野球の実況ツイートでは主に、状況を淡々と述べるツイートやチームや選手を応援するツイートがある。後者の応援ツイートの中には、期待や喜び、落胆など多くの感情が含まれていると考えられる。

ここで、ホームランで得点が入った場合や、抑えの切り札がしっかりと抑え込んだようなシーンにおいては特に感情的なコメントが出現すると考えられるが、これから何かが起こるかもと期待しているシーンにおいては単純な喜びや悲しみといった感情は利用できないであろう。その一方で、これから何かが起こるという期待や、願望にまつわる語が多数現れると考えられる。そこで、こうした期待や願望に関する語の頻度を用いてこれから山場を迎えるシーンを推定する。また、単純にすべてのツイートをを用いるのではなく、そのチームを応援している視聴者のツイートを特に使用することにより、どこまで山場を見逃さないための手法について検討を行い、提案手法の有効性を検証する。さらに、どういった状況において有効で、どういった状況において有効でないのかといった点についても検証を行う。

## 2. 関連研究

Twitter を利用して現実世界で何が起きているのかを検知する研究は多数行われている。

Sakaki ら[2]は、Twitter 上でのユーザの投稿が社会的なセンサーとして使えることに注目し、地震に関するツイートから、その震源地を予測する研究に取り組んでおり、高い精度で判定できることを示している。Sakaki らの研究に限らず多くの研究において、Twitter 上の投稿を一種のソーシャルセンサーとしてとらえて現実世界のイベントなどを推定する研究は行われており、その価値は高いといえる。

Twitter から予測をする研究についても多数取り組まれている。Bollen ら[3]は Twitter 上の株価に関するツイートから、その株価の変動について予測できるかについて検討を行っている。具体的には、ツイートからネガポジや6つの感情因子を抽出し、利用することで、1日当たりの株価の変動を高い精度で推定できることを明らかにしている。また、Tumasjan ら[4]は Twitter 上における政党名に言及するツイートから各政党の得票数を予測する研究をしており、得票数を予測できるという結果を示している。さらに那須野ら[5]は、Twitter における候補者の情報拡散度合いによって当選者予測をする研究をしているうえ、Hotlink[6]は AKB 総選挙の予測を、ソーシャルデータをもとに行うなどしている。他にも、Asur ら[8]は、Twitter 上での映画に関する投稿から、映画の興行収入の予測を行っており、好意的な意見かどうかといった情報を付与することにより、高い精度で予測できることを明らかにしている。また、保住ら[9]は、Twitter 上での投稿に加え、検索エンジンに入力されているクエリの傾向と、Wikipedia の分析を加えることによ

て、消費トレンドを予測する手法を提案している。以上のように、Twitter 上の様々な投稿を通じて未来予測をする研究は多数取り組まれており、我々の研究もこうしたものと類似したものであるといえる。なお、多くの研究では推定できるという結論が出ているが、その推定はできないという反例も示されている[12]。

一方 Kubo ら[7]は、Twitter 上に存在している良い実況者に着目し、そのツイート内容からスポーツ速報を生成する研究を行っており、その有用性について明らかにしている。我々の研究もスポーツに注目しており、その試合の重要なシーンを明らかにすることを目的としているが、Kubo らが何かが起こるシーンを抽出するのに対し、我々はユーザが何かが起こると期待しているシーンを抽出しようとしているという点で異なる。ただ、Kubo らの手法はユーザにチャンスの情報を提示する際に、有効に使えらると思える。

## 3. 山場の見逃し防止手法の検討

本研究で目的とするシステムは、「ユーザが興味のある試合について、今まさに盛り上がっている山場なので通知する」のではなく、「ユーザが興味のある試合について、もしかしたら今から盛り上がって山場がくるかもしれないので、その山場を見逃さないために通知する」というものである。ここで、「スポーツは筋書きのないドラマである」という言葉がある通り、得点のチャンスであってもその選手がホームランを打つのか、ヒットを打つのか、はたまた三振をするのかはわからない。また、ここを抑えれば勝利という場面においても、ピッチャーを抑えることができるのか、それとも打たれてしまうのかといった未来予測をすることはできない。つまり、我々が目指すものは、この後どうなるかはわからないが、「これから盛り上がりそう」「何かが起こりそう」というシーンを推定し、ユーザに提示することである。

こうした推定においては、ランナーがどこにいるのか、バッターとピッチャーの対戦成績はどうなっているのか、点差は何点なのかといった、試合の状況を利用することがまず考えられる。試合にまつわる様々な情報は、逆転するかどうかといった期待値を計算するうえである程度参考になるものであるが、試合の流れや雰囲気判断することは難しい。また、観客の声やアナウンサーの声[11]を使用する方法もあるが、こちらで取得可能なのは今まさに盛り上がっている場面の推定であるうえ、観客の声をどの程度の音量で再生するのかに依存したり、アナウンサー1人の主観に依存したりしてしまう。

試合状況や、コンテンツ自体を使わない方法として、Miyamori ら[10]が取り組んでいるように、スポーツの試合と並行して進行する実況チャットの投稿数が通常より多い、または感情にまつわる語が多数登場しているということか

ら推定する方法が考えられる。しかし、Miyamori らの手法も今まさに盛り上がっているということを検知するものである。今回我々が取り組む手法も、試合と並行して進行する Twitter 上への投稿からシーンを推定するものであるが、盛り上がり（山場）を事前に推定するものであるため、単純に喜びや悲しみといった感情語の出現頻度などを用いることはできない。

ここで、これからバッターがヒットを打って逆転するのか、凡退して得点チャンスを逃すのか、ピッチャーがピンチを抑えるのか、打たれて逆転されるのかといった、こうしたまさに盛り上がるであろうという状況においては、視聴者は選手に対して、「打ってくれ！」や「打ってくれるはず！」、「抑えてくれ！」「彼なら抑えるはず！」などのように、何らかの「期待」や「願望」をもって視聴を続けていると考えられる。そこで本研究では、ツイートに現れる他者への期待や願望にまつわる「期待語」に注目し、その期待語の出現により推定する手法について検討を行う。具体的には、選手に対する「頼む」や「お願い」といった語にまず着目し、その語とよく一緒に使われている語を、推定することによって期待語として抽出する。そのうえで、この期待語の出現頻度が、全ツイートの中である閾値を越えたときに、今まさに何が起こりそうだと期待が高まっているシーンとして推定する。

一方、これから山場を迎えようとしているシーンにおいて、攻撃側のファンと、守備側のファンでは、その期待や願望といった振る舞いが、変わってくると考えられる。そこで本研究では、試合に並行して進行する Twitter 上での投稿を、どちらのファンのものであるかを分離し、その分離したツイートからそれぞれ期待度合を推定し、その平均を求めることによって「これから山場を迎えようとしているシーン」として判定する手法を提案する。

#### 4. 山場を見逃さないためのデータセット構築

実験にあたり、データセットを構築する。ここではまず複数のプロ野球の試合を選定して、ハッシュタグを用いてツイートを収集する。また、各試合のチャンスである時間の正解データの作成を人力で行う。ここでユーザが求めているのは、山場を見逃さないためのものであるが、先述の通り、未来予測をすることはできないため、「この後何が起こるかも」と視聴者が期待できるかどうかという観点で正解データセットを構築し、その後何かが起こるかどうかは考慮せず、視聴者のツイートからその期待しているシーンを推定精度について検証する。

##### 4.1 データセット

表 1 に示す通り、2018 年に行われたセ・リーグ公式戦 8 試合と 10 月 18 日に行われたクライマックスシリーズ・ファイナルステージ第 2 戦のセ・リーグとパ・リーグの各試

合の合計 10 試合分のツイートを収集した。収集には、各球団が公式に利用していたり、ファンがよく利用したりしているハッシュタグを用いた。本研究で提案する手法は、どちらのファンであるのかに注目するものであるが、本研究では簡単のためこのハッシュタグを使うことにより、どちらのファンのツイートであるかを分離した。

ここで、山場については攻撃側、守備側両方の立場がありうるが、今回は単純のため攻撃側のみを対象としてデータセットを構築する。また、各試合のこれから山場を迎えそうな時間（以降、攻撃側のみを対象とするためチャンスと表記する）の正解データは、著者が日本野球機構のウェブサイトに掲載されている試合経過と収集した実況ツイートを見て作成した。また、チャンスである時間の基準は様々なものが考えられるが、今回はデータセット構築の都合上、得点圏と呼ばれる状況になった場合とした。得点圏とは、走者が二塁または三塁に走者がいる状態のことである。ただし、点差が大きく同点や逆転の可能性がないとき、また打者が投手の場合などは、ファンの期待があまり高くないと考え、正解としなかった。

表 1 収集対象とした試合

日付	対戦カード
2018 年 7 月 8 日	巨人 6 - 4 広島
2018 年 7 月 20 日	広島 11 - 9 巨人
2018 年 7 月 22 日	広島 8 - 6 巨人
2018 年 7 月 26 日	阪神 1 - 3 広島
2018 年 8 月 5 日	DeNA 5 - 6 広島
2018 年 8 月 15 日	ヤクルト 6 - 5 巨人
2018 年 8 月 19 日	巨人 6 - 5 中日
2018 年 8 月 23 日	広島 9 - 8 ヤクルト
2018 年 10 月 18 日	広島 4 - 1 巨人
2018 年 10 月 18 日	西武 5 - 6 ソフトバンク

##### 4.2 期待語の選定

チャンスとなってファンが選手に期待をかける場面のツイートでは「頼む」「お願い」「決める」といった単語が多く使われているため、まずその 3 語とそれらのひらがなの表現を期待語として選定し、分析を行う。また、我々はチャンスの時にそれらとともに使われる語があるのではないかと考えた。そこで、word2vec を用いて類似語を抽出した。「頼む」「たのむ」「お願い」「おねがい」「決める」「きめる」の 6 つの単語からそれぞれ抽出され、類似度が 0.5 以上のものを対象とする。word2vec による学習は名詞、動詞、形容詞、副詞、感動詞の 5 つの品詞を利用し、MeCab を用いて名詞以外の基本形を取得した。これにより抽出された語を含めた場合の分析も行う。重複があったため、追加された単語の数は 42 であった。抽出された語は直接期待

を表すものでないものも含まれるため、「頼む」などの 6 語が含む場合は重みを 1 とし、追加された語のみを含む場合の重みは小さくして割合を出すこととする。

#### 4.3 比較手法

本研究では、ツイートをそれぞれのファンのものに分離して比較する。ここで比較するのは、下記の 3 手法である。

**All 手法:** どちらのファンであるかを区別せず、単純に単位時間あたりの期待にまつわる語の出現頻度からチャンスシーンを推定する手法。これを比較のためのベースライン手法とする。

**Team 手法:** それぞれのチームのファンによる単位時間あたりの期待にまつわる語の出現頻度から、それぞれのチームのチャンスシーンを推定する手法。

**Both 手法:** 各チームのファンによる単位時間あたりの期待にまつわる語の出現頻度をそれぞれ計算し、その平均値を用いてチャンスシーンを推定する手法。

上記の 3 手法を用いて、チャンスシーンの推定可能性について検討を行う。なお、出現頻度の閾値については、それぞれ値を変更した実験を実施して適合率と再現率からなるグラフを作成し、そのグラフから手法の検討を行う。

### 5. 実験結果

#### 5.1 精度の判定方法

チャンス判定の閾値は、期待語の出現頻度が 2% から 30% まで 2% ごとに 15 段階設定してそれぞれ実験を行った。判定精度の評価指標は、適合率 (Precision)、再現率 (Recall) を使用する。また、語拡張を行わない場合と、語拡張を行う場合についても比較を行う。

ここで、適合率は、チャンスとしてシステムが推定したシーン内、実際にチャンスだったものはどれだけあるのかで求める。一方再現率は、数分間連続したチャンスシーンを 1 つの塊とみなし、その塊をいくつ判定することができたのかということで求める。

図 1 に 2018 年 7 月 8 日の巨人対広島の実況ツイートを分毎の期待語の割合のグラフにした結果を示す。図 1 の緑の線が巨人側の、青が広島側のツイートを用了場合、赤は両チームのツイートを用了場合の期待語の割合の推移である。紫線は閾値を示しており、この図では 12.5 となっている。背景が黄色の部分巨人のチャンス、青の部分広島チャンスを示している。この部分で割合が閾値を超えたときに正しい判定となる。この試合では、巨人側のみのツイートを用了場合の適合率が 2/13、再現率が 2/4 であり、広島側のみの場合の適合率が 4/17、再現率が 2/3 である。両チームのツイートを用了場合、適合率が 5/13、再現率が 3/7 となる。

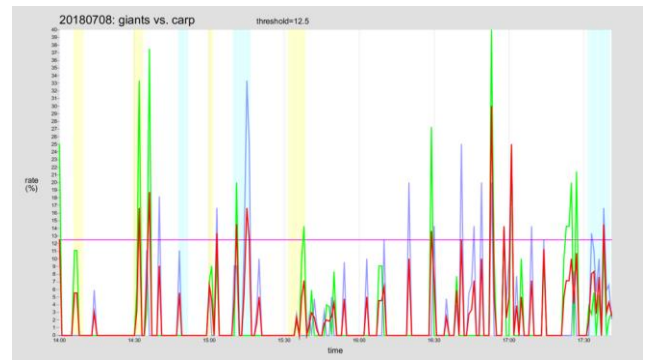


図 1 時間ごとの推定値

#### 5.2 語拡張なしでの精度

「頼む」、「お願い」、「決める」の 3 語とそのひらがな表現を含むツイートの頻度の割合を用いる分析を行った。このときの適合率と再現率の散布図を図 2 に示す。ここで図の横軸は再現率、縦軸は適合率を意味している。また先述の通り、All 手法は全ツイートを単純に用いるもの、Both 手法はチームごとに計算して平均化するもの、Team 手法はそれぞれで推定する手法である。

閾値を低くすると、再現率が上がるが、適合率が下がっていく。一方、閾値を高くすると再現率が下がるという傾向がある。結果より、適合率と再現率両方を考慮した場合、Both 手法が最も良い結果であることがわかる。

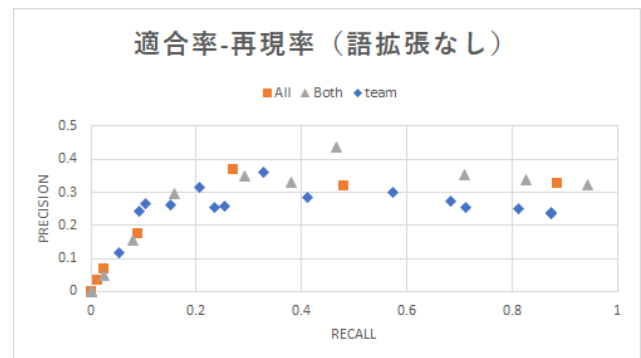


図 2 語拡張なしのときの手法間の精度比較

#### 5.3 word2vec で抽出された語を含めた分析

word2vec を用いて抽出された 42 語を期待語に含めた場合の結果を図 3、4 に示す。図 3、4 ともに横軸に再現率、縦軸に適合率をとっている。また、図 3 は語拡張された語への重みを 0.1 としたもの、図 4 は重みを 0.2 にしたものを示している。

どちらの結果においても、適合率・再現率の両者を高くしているのは Both 手法であるといえる。

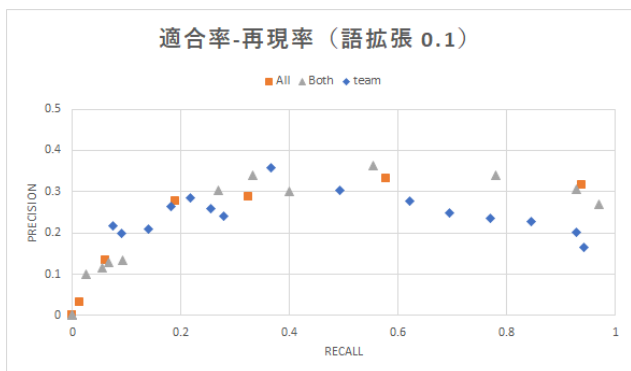


図 3 重み 0.1 で語拡張したときの適合率・再現率

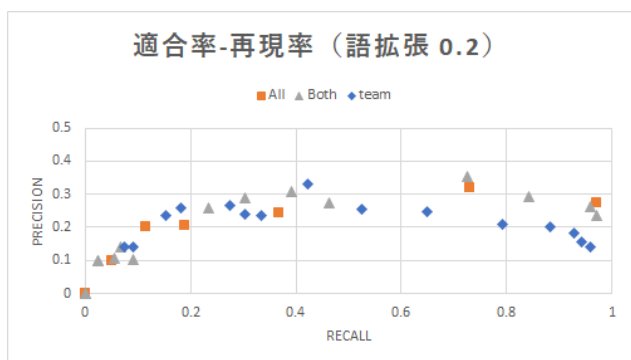


図 4 重み 0.2 で語拡張したときの適合率・再現率

次に、精度が高かった Both 手法について、語拡張を行っていないもの、語拡張を 0.1 の重み付けで行っているもの、0.2 の重み付けで行っているものを比較しているのが図 5 である。この結果より、語拡張を 0.1 の重み付けで行っている場合に高い精度になっていることがわかる。

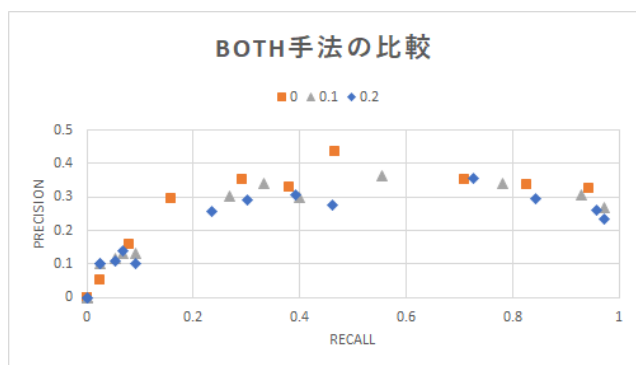


図 5 Both 手法内での語拡張なし、0.1、0.2 の重み比較

## 6. 考察

実験より、提案している Both 手法が最も良い結果であったものの、適合率、再現率ともに十分ではなく、特に適合率は 0.5 を超えることがなかった。そのため、All 手法や Team 手法に比べ、Both 手法が有効であることは分かった

ものの、そもそも提案手法が十分に有効であるとはいえない結果であった。今回の結果で適合率が低い原因の一つとして、攻守交代時などのツイート数が特に少ない時間帯であるために、図 6 のように期待語を含むツイートの割合が極端に高くなり、誤判定しているものいくつか見られた。さらに、投手交代時に中継ぎや抑えの切り札といわれているような選手が出場した場合に誤判定しているものもいくつか見られた。今回の実験では攻撃側にとっての山場を迎えるシーンについてデータセットを構築したが、こうした中継ぎや抑えの切り札登場は、守備側のファンにとってはこれから山場を迎えるシーンであるため、ある意味正解であるとも考えられる。そこで今後は、攻撃側だけでなく、守備側も考慮したデータセットを構築し、精度の判定を行っていく予定である。

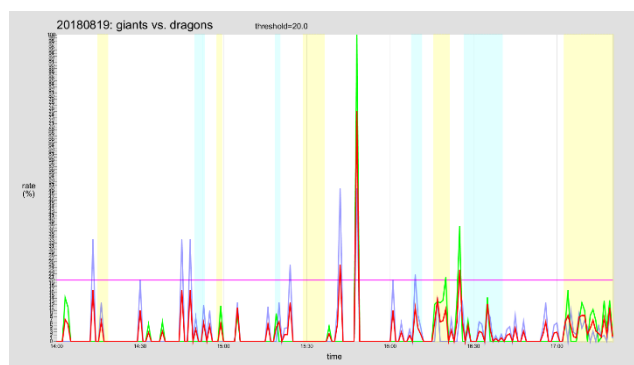


図 6 誤判定してしまった試合の推定値

一方、本研究ではチャンスの時間の正解データの作成を著者の主観で行ったため、適切でない正解データとなっている可能性が考えられる。最初にも述べた通り、楽観的な人、悲観的な人によってチャンスの評価も変わるであろう。また、攻撃側だけを今回対象としていたが、投手交代で守護神が登場する場面など、今回正解としなかった守備側の場面も視聴したいと思うユーザもいると考えられる。そこで今後は試合の山場である場面の基準の調整や複数人で正解データを作成する予定である。

word2vec を用いて抽出した語を含めた場合の精度について、使わない場合に比べ多少向上したものの十分でなかった。その原因として、類似語を抽出するときの学習データが不十分であったことが挙げられる。今回は収集したツイートを全て無差別に学習させたが、チームによって選手名や応援歌などの特徴語があるため、ツイートをチーム別に学習させ、チームによって判定に用いる語を変えることで精度が向上するのではないかと考えられる。

また、全体を通して Team 手法の精度は高くなかったが、ここでホームチームなのかビジターチームなのかによって精度に大きな差があった。図 7 は、Team 手法について、ホームとビジターに分けて適合率と再現率を求め、その精度



をプロットしたものである。なおここでは、ホームとビジターのみならず、勝敗についても考慮して精度を出している。ここで、図の横軸は再現率、縦軸は適合率を意味している。

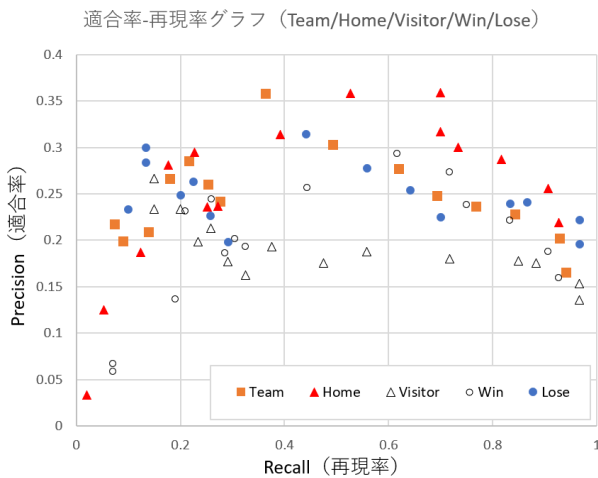


図7 ホーム・ビジター・勝敗による精度の違い

この結果より、ホームチームはビジターチームの2倍以上高い精度で判定することができていたが、ビジターチームの精度が足を引っ張る状態となっていたことがわかる。つまり、ホームとビジターを考慮することにより、精度を向上させることができると期待される。このホームとビジターによる違いが影響している点は興味深い。一方、図7では勝敗の影響についてもプロットしているが、この結果からはその差を見出すことはできない。このことより、最終的な結果ではなく、試合展開を考慮する必要があるのではと考えられる。例えば、チャンスシーン（これから山場になるであると考えられるシーンの直前）では、そのタイミングで負けているチームの方がより期待や願望に関するツイートをしており、その影響が出た可能性がある。つまり、現在負けている状態での期待や願望なのか、それとも勝っている状態での期待や願望なのかでは、そこに表出するものも違うと考えられる。そこで今後は、こうした点を考慮して勝ちチームならではの、負けチームならではの語を推定し、そこから判定する予定である。実際、Shiratoriら[1]は試合展開を考慮してネタバレツイートを判定しており、その手法をそのまま応用可能であると考えられる。今後は、そうした点を考慮した手法を検討していく予定である。

## 7. まとめと今後の課題

本研究では、プロ野球の試合においてこれから山場になると予想されるシーンを推定し、ユーザに視聴誘導することを目的とし、プロ野球の試合に連動したTwitter上での投稿から推定する手法について検討を行った。本研究では特

に、これから山場を迎えるシーンにおいてツイートの中に現れる期待や願望といった語が出現することに注目し、またツイートをファン毎に分類してチャンスシーンを推定する手法を提案した。また、10試合分のデータセットを構築し、その試合についてどの程度推定可能かについて、閾値を変更しつつ推定することによって、チームごとに推定する手法の可能性を明らかにした。しかし、適合率・再現率ともに十分な精度であるとは言えなかった。

今回は単純なパターンマッチでの推定を行っているが、今後SVMなどの手法によりどの程度推定可能かについても検証を行っていく予定である。また、勝っているのか負けているのかといった試合展開を考慮した手法を考慮することや、試合の詳細なデータと組み合わせた手法についても検討を行っていく予定である。さらに、アナウンサーの声の大きさといった情報も組み合わせて使用可能であると考えられる。

一方、今回実験に利用したデータセットは十分なものであるとは言えなかった。そこで今後はチャンスに関するデータセットをどのように構築すべきなのかといった点について深く研究するとともに、再現性の高いデータセットを構築するとともに、再度検証を行っていく予定である。

## 参考文献

- [1] Yuji Shiratori, Maki Yoshiki, Satoshi Nakamura and Takanori Komatsu: Detection of Football Spoilers on Twitter, The 10th International Conference on Collaboration Technologies (CollabTech 2018), pp.129-141.
- [2] Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo: Earthquake shakes Twitter users: real-time event detection by social sensors, Proceedings of the 19th international conference on World wide web (WWW 2010), pp. 851-860.
- [3] Johan Bollen, Huina Mao, Xiaojun Zeng: Twitter mood predicts the stock market, Journal of Computational Science, 2 (1), 2011, pp. 1-8.
- [4] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpe: Predicting elections with Twitter: What 140 characters reveal about political sentiment. Proceedings of 4th International AAI conference on Weblogs and Social Media (ICWSM 2010), pp. 178-185.
- [5] 那須野薫, 松尾豊: Twitterにおける候補者の情報拡散に着目した国政選挙当選者予測, 第28回人工知能学会全国大会論文集 (JSAI 2014), 1K3-1, pp.1-4.
- [6] Hotlink: 2014年AKB総選挙 ソーシャルデータ予測の舞台裏, <https://www.hottolink.co.jp/customer/hottosearch/20140605-2>
- [7] Mitsumasa Kubo, Ryohei Sasano, Hiroya Takamura, Manabu Okumura: Generating Live Sports Updates from Twitter by Finding Good Reporters, 2013 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technologies, pp.527-534.
- [8] Sitaram Asur, Bernardo A. Huberman: Predicting the Future with Social Media, 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 429-499.
- [9] 保住純, 飯塚修平, 中山浩太郎, 高須正和, 嶋田絵理子, 須賀千鶴, 西山圭太, 松尾豊: Webマイニングを用いたコンテンツ消費トレンド予測システム, 人工知能学会論文誌 29 巻 (2014) 5号, pp.449-459.

- [10] Hisashi Miyamori, Satoshi Nakamura and Katsumi Tanaka: Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web, Proceedings of the 13th annual ACM international conference on Multimedia, pp. 853-861.
- [11] Dian Tjondronegoro, Yi-Ping Phoebe Chen, Binh Pham: Sports video summarization using highlights and play-breaks, Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval (MIR 2003), pp. 201-208.
- [12] Andreas Jungherr, Pascal Jürgens, Harald Schoen: Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment", Social Science Computer Review, 30(2): 229-23, 2012.