

## Twitter を用いたスポーツ試合中のイベント検出に関する検討

富田 大志<sup>†</sup> 道満 恵介<sup>††</sup> 井手 一郎<sup>†††</sup> 出口 大輔<sup>††††</sup> 村瀬 洋<sup>†††</sup>

<sup>†</sup> 名古屋大学 工学部 〒464-8603 愛知県名古屋市千種区不老町  
<sup>††</sup> 中京大学 情報理工学部 〒470-0393 愛知県豊田市海津町床立 101  
<sup>†††</sup> 名古屋大学 大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町  
<sup>††††</sup> 名古屋大学 情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町  
E-mail: <sup>†</sup>tomitat@murase.m.is.nagoya-u.ac.jp, <sup>††</sup>kdoman@sist.chukyo-u.ac.jp,  
<sup>†††</sup>{ide,murase}@is.nagoya-u.ac.jp, <sup>††††</sup>ddeguchi@nagoya-u.jp

**あらまし** 本稿では、Twitter に投稿されたツイートを基にスポーツの試合中におけるイベントを検出する手法に関する検討結果について報告する。近年、Twitter とスポーツの組み合わせに注目が集まっており、ツイートから得られる情報を用いたスポーツ映像の要約などが期待されている。スポーツ映像に対して試合内容を示すタグ付けなどを自動的に行うには、試合中に存在する得点や反則といったイベントの種類をそれぞれ判別し検出することが必要である。本研究では、種類に応じてイベントを検出することを目的とし、本報告では特にサッカーの試合中のゴールを検出する手法を検討する。従来手法では、各時刻に投稿されたツイート数のみを用いてイベントの検出を行っていた。一方検討手法では、各時刻に投稿されたツイートから、リツイートの数、興奮を表す文字列特徴も抽出・評価することで、様々なイベントの中でもゴールのみを検出する。評価実験の結果、従来手法に比べ、複数の特徴を導入した検討手法の方が高い検出精度が得られ、検討手法の有効性を確認した。

**キーワード** Twitter, イベント検出, スポーツ

## A Study on Event Detection in a Sports Game using Twitter

Taishi TOMITA<sup>†</sup>, Keisuke DOMAN<sup>††</sup>, Ichiro IDE<sup>†††</sup>,  
Daisuke DEGUCHI<sup>††††</sup>, and Hiroshi MURASE<sup>†††</sup>

<sup>†</sup> Faculty of Engineering, Nagoya University Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8603 Japan  
<sup>††</sup> School of Information Science & Technology, Chukyo University  
Tokodachi 101, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan  
<sup>†††</sup> Graduate School of Information Science, Nagoya University  
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan  
<sup>††††</sup> Information and Communications Headquarters, Nagoya University  
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan  
E-mail: <sup>†</sup>tomitat@murase.m.is.nagoya-u.ac.jp, <sup>††</sup>kdoman@sist.chukyo-u.ac.jp,  
<sup>†††</sup>{ide,murase}@is.nagoya-u.ac.jp, <sup>††††</sup>ddeguchi@nagoya-u.jp

**Abstract** We report the results of the study on a method for detecting events during a sports game based on tweets posted on Twitter. Recently, the combination of sports and Twitter is focused, and summarization of a sports video by information obtained from tweets is expected. To automatically annotate a sports video, it is necessary to determine the type of events present in the game, such as goal or foul. In this report, we address a method for detecting goals in a soccer game with the aim of detecting with their events. In the conventional method, only the number of tweets was used to detect events. In contrast, our method detects only goals among various events by using the number of retweets and features in the text that represents excitement in addition to the number of tweets. From the result of an evaluation experiment, the studied method achieved higher detection accuracy than the conventional method for the task. As a result, we confirmed the effectiveness of the studied method.

**Key words** Twitter, event detection, sports

## 1. はじめに

近年、マイクロブログは世界中に急速に普及し、個人だけでなく、企業も情報の発信や商品の宣伝に利用している。マイクロブログの一つである Twitter<sup>(注1)</sup>は、2006年に米国でサービスを開始して以来ユーザを増やし続けている。2012年3月時点のアクティブユーザ数は全世界で1億4000万人を超えており[1]、その内の70%以上は米国以外の国から利用している[2]。Twitterの特徴として、投稿できる文字数が最大140文字であること、画像や映像を直接投稿することができずテキストのみで他のユーザとやり取りをする等の手軽さがある。そのため、Twitterはコミュニケーションツールとしてだけでなく、速報性のある情報源として期待が高まっている[3][4][5]。また、最近ではTwitterとスポーツという組み合わせに注目が集まっている。2012年7月から8月にかけて開催されたロンドンオリンピックにおいては、大会期間中に投稿されたオリンピックに関するツイート（Twitterへの投稿）数は1億件以上になった[6]。オリンピックのような世界規模の大会でなくとも、アメリカンフットボールのプロリーグであるNFLやサッカーの欧州選手権であるUEFA EUROなど、ある程度の知名度をもつ大会であれば、オリンピックと同じようにその大会に関するツイートが多く投稿される[7]。実際にオリンピック期間中に投稿されたツイートの例を表1に示す。この例は、サッカーの試合中においてゴールが決まった瞬間のツイートである。得点に対する歓喜や失点に対する落胆といった、ゴールが決まったことに対するツイートが様々な国の言葉で数多く投稿されている。このことから、現在見ているスポーツの状況をTwitterへ投稿するユーザの多さが証明された。

このようなTwitterとスポーツの関係に着目し、小林らやLanaganらはツイートをを用いてスポーツ映像を要約する手法を提案した[8][9]。小林らは、ユーザのツイート数を基にして要約映像に含むべき試合映像の箇所を選択し、要約映像に含む得点シーンや、満塁などの得点が決まりそうなシーンなどのイベント検出が可能である手法を提案しており、野球の試合に適用、評価している。また、Lanaganらは、ツイート数とSBD (shot boundary detection) を用いて、得点シーンや反則のシーンなどのイベント検出が可能である手法を提案しており、サッカーとラグビーの試合に適用、評価している。しかし、これらの手法ではそのイベントがどのような種類であるか判別することはできず、イベントが発生したシーンへのタグ付けや、特定のイベントのみを検出することができない。一方Zhaoらは、イベントを検出し、その種類も理解する手法[10]を提案している。この手法は、ツイート数を基にイベントを検出し、そのツイートに含まれる語からイベントの種類を判別している。Zhaoらはこの手法をアメリカンフットボールの試合に適用、評価しており、タッチダウンやインタセプションなどといったイベントの種類判別が可能であることを示している。この手法はツイート中の語によってイベントを判別するため、対象とするスポー

ツのイベントを記述した辞書が必要となる。これら3つの手法は、Twitterから得られる情報としてツイート数のみを用いたものであり、イベント検出における再現率は高い。しかし、ツイート数が試合中のイベントだけでなく、試合終了時などの全体的なイベントの区切りにおいて増大することへの対処が不十分である。そのため、試合中のイベントではない場面を誤検出してしまふことが多い。

それに対し我々は、ツイート中の興奮を表す特徴や、他のユーザのツイートを引用したツイート（リツイート）の数などのツイートに含まれる特徴を考慮したイベントの検出手法を検討する。まずはじめに、イベントを個別かつ高精度に検出することを目指し、ゴール検出を行おうと考えた。イベントの中でもゴールを検出対象とした理由は、試合の勝敗を決定する上で最も重要なイベントであり、最低限ゴールさえ確認することができれば試合結果を把握することができると考えたからである。また、特にサッカーの試合を対象にすることにした。スポーツの中でもサッカーを対象とした理由は、イベントの存在する場面とそうでない場面において、視聴者の興奮の度合いの差が大きいと考えたからである。ゴールのような特定のイベントを個別に検出することが可能となることで、検出したイベントに対し、そのイベントに適切なツイートを抽出して解説文やコメントとして付与するといったことが実現できると考える。

検討する手法においては、従来手法で用いられていたツイート数に加え、新たに2つの特徴を導入する。興奮を表す特徴は、ゴールの瞬間に増大するため、ゴールを検出する精度をより高めることができる。また、リツイート数は全体的なイベントの区切りにおいて増大するという特徴であるため、これを考慮することで誤検出を除去することができる。これらの特徴を組み合わせるとゴールを検出するスコアを算出することで、試合終了などのサッカーの試合自体の区切りを検出せず、ゴールのみを検出することができ、より高精度なゴール検出が可能となると考える。

以降、2節ではTwitterについて述べ、3節では検討手法を詳述した後、有効性を確認するための実験について述べる。続く4節では、検討手法の有効性と改善の余地について考察する。最後に5節でまとめる。

## 2. Twitterについて

先述したとおり、Twitterはユーザが投稿できる文字数に140文字という制限がある非常に手軽なマイクロブログサービスである。Twitterへの投稿は「ツイート」と呼ばれ、基本的に「今何してる?」というTwitterの問いに答える形で行うが、他者との交流のために用意された機能も存在する。その中の一つである「リツイート」は、他者のツイートを引用して自分のツイートを投稿するという機能である。その例を表2に示す。このツイートは、XXXというユーザのツイートを引用して他のユーザが投稿したものである。リツイートを行う際は、引用するツイートの最初に“RT”とつけて投稿する。なお、特定のユーザを話題にしたツイートや、他のユーザに宛てたツイートを投稿する場合は、そのユーザ名の先頭に“@”をつける。

(注1) : Twitter, <http://twitter.com/>.

表 1 スポーツ試合中のツイート例 (ロンドンオリンピック 女子サッカー決勝 日本対米国)

日時	タイムゾーン	ツイート内容
2012/8/1/19:55:07	-06:00	GOOOOOOOOOOOOOOOOAAAAAAAAAAAAAAAAALLLLLLLLLLLLLLLLLLLLLL!!!
2012/8/1/19:55:07	-06:00	GOOOOOOAAALLLLLLLLLLLLLLLLLLLLLL!!!!!!!!!!!!!!!!!!!!
2012/8/1/19:55:08	-08:00	GOAL!!! :)
2012/8/1/19:55:08	-08:00	Go #TeamUSA!
2012/8/1/19:55:08	+09:00	また入れられた (;o;)
2012/8/1/19:55:09	+09:00	痛いなあ…落ち着こう!
2012/8/1/19:55:09	-05:00	GOAL!
2012/8/1/19:55:10	+02:00	RT @AP: U.S. leads Japan 1-0 at halftime in gold medal women's soccer match.
2012/8/1/19:55:11	+09:00	厳しいなあ…
2012/8/1/19:55:12	+09:00	うああああああああああ 切り替えの遅さが目立つなあ
2012/8/1/19:55:13	-05:00	GOOAAAAALLLLLLLLLLLLLL!!!
2012/8/1/19:55:14	+09:00	ぎゃああああああああああああああ
2012/8/1/19:55:14	-05:00	GOAL number 2 get at her....
2012/8/1/19:55:15	-05:00	GOAL #TEAMUSA !!!!!!!!!!!!!!!!!!!!!!! Up 2:0 in the 55th Minutes!!!
2012/8/1/19:55:15	+09:00	後半 9 分に 2 点目 (´ω´)
2012/8/1/19:55:16	+09:00	ぐあー… きっつー…
2012/8/1/19:55:16	-05:00	2-0!
2012/8/1/19:55:17	-06:00	Goal! Lloyd (54') give USA a 2-0 lead.

表 2 リツイート, ハッシュタグの例 (ユーザ名は全て XXX に置換)

日時	ツイート内容
2012/9/6/11:59:12	RT @XXX: 駒野のクロス最高!! #ntv #daihyo #ツイート観戦 #jfa
2012/9/6/12:07:24	RT @XXX: 清武の今日のパフォーマンスは素晴らしいね。#jfa #daihyo

他にも、ある特定の話題に関するツイートであることを明示するために、ツイートに「ハッシュタグ」というものを付記して投稿する使い方がある。表 2 に示したツイートの末尾に、「#」に続けて記されているのがハッシュタグである。このようにツイートすることで、特定の話題に関するツイートのみを検索することが可能となり、同じ事柄に興味を持ったユーザを見つけたり、他のユーザがその話題に対してどのようなツイートをしているのか調べるのに役立つ。

### 3. 検討手法

本稿では、サッカーの試合中におけるゴール検出手法を検討する。検討手法では、以下の 3 つの特徴を用いる。

- (1) 試合中のイベントを示す特徴:  $T$
- (2) ゴールを示す特徴:  $E$
- (3) イベントの区切りを示す特徴:  $R$

(1) は従来研究で用いられていた特徴であり、(2), (3) は新たに導入する特徴である。なお、これらの特徴を用いる理由については、後に予備実験で示す。

上記の 3 つの特徴を用い、ゴールスコアを次式で算出する。

$$S = T \times E \times (1 - R) \quad (1)$$

なお、スコアと各特徴量は最大値を 1, 最小値を 0 とし正規化する。また、これらの値は時刻  $t$  ごとに計算する。

このように検討手法では、 $T$  と  $E$  の積に  $R$  を 1 から引いた

値をかけることで、試合中のイベントを示す特徴  $T$  とゴールを示す特徴  $E$  が多く、イベントの区切りを示す特徴  $R$  が少ないほどゴール検出スコア  $S$  が大きくなり、ゴールと判定する。

以降、各特徴について詳述する。

#### 3.1 試合中のイベントを示す特徴

スポーツの試合中にゴール等の何らかのイベントが発生した場合、ユーザはそのことに対するツイートを投稿すると思われる。つまり、イベント発生時にツイートが増大するということがツイートの検出に用いることができるのではないかと考えた。この特徴は従来研究 [8] [9] [10] でも利用されており、ゴールやファウルなどのイベント検出が可能であると報告されている。しかし、この特徴だけではゴールとその他のイベントを区別できない。

#### 3.2 ゴールを示す特徴

ゴールの瞬間においては、試合を観戦しているユーザは興奮や落胆などの感情的な状態になり、その気持ちがツイートの文面に表れる傾向がある。よって我々は、ユーザが興奮状態であることを検出することができれば、ゴールの瞬間を検出できるのではないかと考えた。一方で、ツイートは口語体で投稿されることが多い。そのため、“!” は語気を荒げる記号として用いられると思われる。語気を荒げるということは興奮していることを示し、その文字数が多いほどより興奮しているといえる。よって我々は、“!” の文字数を興奮を表す特徴量として用いることができないかと考えた。

#### 3.3 イベントの区切りを示す特徴

過去研究 [11] において、イベントの区切りにおいてツイートの数が増大するということが示されている。よって、スポーツの試合の区切りやイベントの後においてもツイートの数が増大すると考えられ、ゴール検出に影響を及ぼすといえる。このため、試合やイベントの区切りを示す特徴が必要となる。我々はその

特徴量としてリツイート数に着目した。前述したとおり、リツイートは他者のツイートを引用してツイートを投稿する機能である。リツイートを行うには、気に入ったツイートを発見して先頭に“RT”と付記し、場合によっては自分のコメントもつけて投稿する必要がある。そのため、通常のツイートを投稿するよりも手間がかかる。このような作業を、まさにゴールが決まりそうな瞬間に行うとは考えにくい。他者のツイートを引用して議論を行いたいならば、イベントの終了時やその後に行うと思われる。よって、我々は試合やイベントの区切りとその他の場面を区別する特徴量として、リツイート数を用いることができないかと考えた。

## 4. 実験

4.1 では、前節で導入した特徴の正当性を予備実験によって確認する。4.2 では、検討手法の有効性を評価実験によって確認する。なお、ツイートの収集には TwitterStreamingAPI<sup>(注2)</sup> を用いた。

### 4.1 予備実験：各特徴の正当性の確認

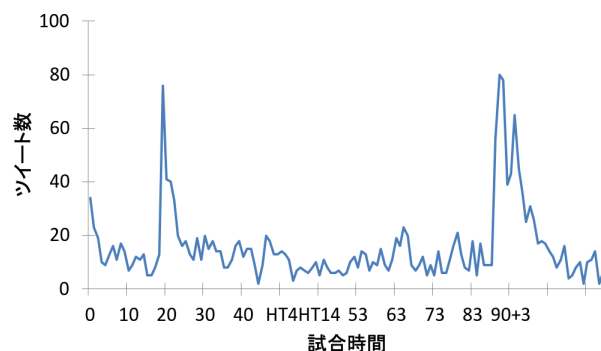
前節で導入した特徴の正当性を確認するため、調査実験を行った。サッカーの試合中のツイートを収集し、そのツイート中に含まれる“!”の文字数と、リツイート数を60秒あたりの値として算出し、その傾向を調査した。

対象とした試合は、サッカーの国際試合である UEFA EURO2012 のスペイン対フランスの試合で、ツイートはハッシュタグ“#EURO2012”、“#Euro2012”、“#euro2012”のいずれかを含むもののみを収集した。収集した期間は、2012年6月23日の試合開始時刻から試合終了までの112分間と試合終了後23分間を合わせた135分間である。

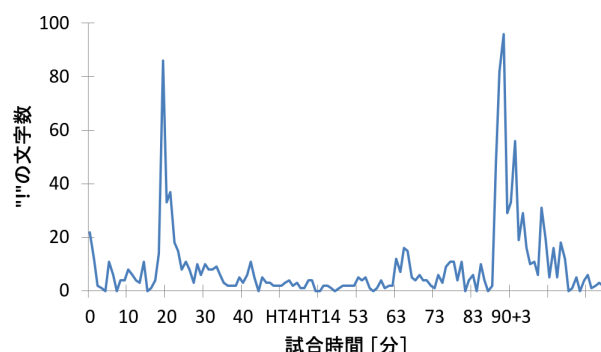
実験結果を図1に示す。この試合では、19分と90+1分のときにゴールが決まった。図1(a)はツイート数の推移を示している。同図においては、試合開始時、ゴール、試合終了時に特に高くなっている。図1(b)は“!”の文字数の推移を示している。同図においては、ツイート数の推移と同じく、試合開始時、ゴール、試合終了時に高くなっている。図1(c)はリツイート数の推移を示している。同図においては、試合開始時、ゴール直後、試合終了時に高くなっている。90+1分のゴールがペナルティキック(PK)によるものであること、終了間際のゴールであることなどから、その時点のリツイート数が増加しているが、イベントの区切りにおいて増加するというリツイート数の傾向は、その他の時点において確認することができた。この結果から、ツイート数によるイベント検出のスコアを“!”の文字数によるスコアでゴールの部分のみ増幅させ、リツイート数のスコアによってイベントの区切りを除去すれば、ゴールの瞬間のみを検出できると思われる。

### 4.2 実験：検討手法の有効性の確認

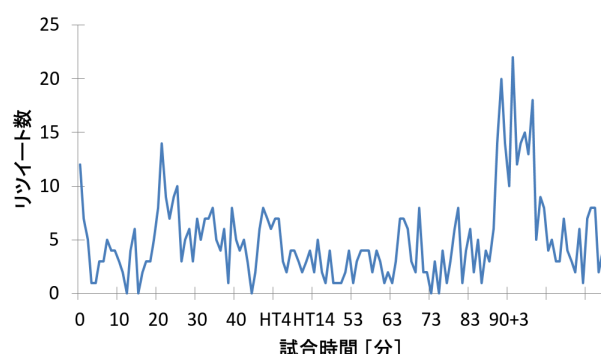
検討手法の有効性を確認するための評価実験について述べる。比較のため、以下の手法について検出精度の評価をした。



(a) ツイート数



(b) “!”の文字数



(c) リツイート数

図1 各特徴の調査結果 (HT はハーフタイムの意)

- 検討手法 (式 (1)) :  $S = T \times E \times (1 - R)$
- 従来手法 :  $S = T$
- 参考手法 1 :  $S = T \times E$
- 参考手法 2 :  $S = T \times (1 - R)$

従来手法は、単位時間当たりのツイート数  $T$  のみを用いる手法 [8] [9] [10] である。上記の合計 4 つの手法の検出精度を比較することで、検討手法において新たに導入した特徴の有効性を評価した。

#### a) 実験データ

評価実験には、表 3 に示す UEFA EURO2012 の試合中のツイートを使用する。ツイートは、ハッシュタグ“#EURO2012”、“#Euro2012”、“#euro2012”のいずれかを含むもののみを収集した。

(注2) : TwitterStreamingAPI, <https://dev.twitter.com/docs/streaming-apis/>.

表 3 実験データ

月日	対戦カード	時間	ツイート数
2012/6/21	準々決勝 チェコ対ポルトガル	135 分間	1299
2012/6/22	準々決勝 ドイツ対ギリシャ	135 分間	2566
2012/6/23	準々決勝 スペイン対フランス	135 分間	2092
2012/7/1	決勝 スペイン対イタリア	165 分間	410363

b) 評価方法

ゴールの判定に用いるスコア  $S$  は、0.0 から 1.0 までの値をとる。ゴールと判定するスコアの閾値を 1.0 から 0.0 まで 0.1 刻みで減少させていき、再現率と適合率を計算する。なお、ゴールと判定した時刻と、その 1 分前の合計 2 分の間に実際にゴールが存在した場合に検出成功とする。これは、ツイートの遅れと、分のまたぎ目でのゴールに考慮したものである。得られた再現率と適合率を用いて F 値を計算し、評価する。

c) 実験結果

表 3 に示した各試合に対して先述した評価手法を適用した場合の F 値を図 2 に示す。

4 試合中 3 試合において、いずれの閾値においても検討手法は従来手法の値を上回った。また、各比較手法の値に対しても全体的に上回っていることを確認した。このことから、興奮を示す特徴とイベントの区切りを示す特徴を新たに導入した検討手法の有効性を確認した。

## 5. 考 察

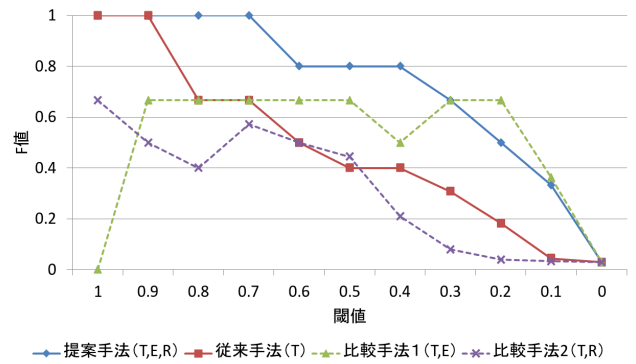
実験結果をうけ、検討したゴール検出手法の有効性と、改良の余地について考察する。

### 5.1 検討特徴の有効性

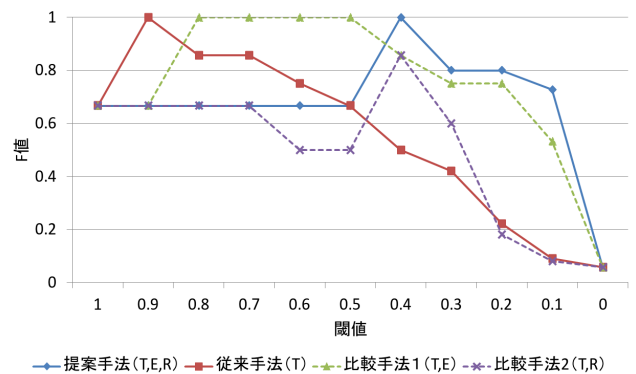
検討手法及び比較手法はほとんどの場合において従来手法の F 値を上回っている。図 3 にチェコ対ポルトガルの試合に対して、従来手法と検討手法を適用した場合のゴール検出スコアの遷移を示す。この試合においてゴールが決まったのは 79 分である。従来手法によって得られたスコアに比べ、検討手法によって得られたスコアはイベント以外の場面のスコアが減少し、前半終了時や試合終了時のスコアも減少していることが分かる。なお、この試合は 90+4 分であり、他の時点に比べその周辺のスコアが大きくなっているが、それについては後述する。ゴールの場面以外のスコアが減少したのは、“!” の文字数を利用し、ゴール時のスコアを高めたためであると考えられる。また、前半終了時や試合終了時のスコアが減少したのは、リツイート数を用いてスコアを減少させたためであると考えられる。このことから、本研究で新たに導入した特徴は高精度なゴール検出に効果があったといえ、検討手法の有効性が確認できた。検討手法の次に精度が良いと考えられるのは、 $T$  と  $E$  を用いた比較手法 1 であった。

### 5.2 検討特徴の改良の余地

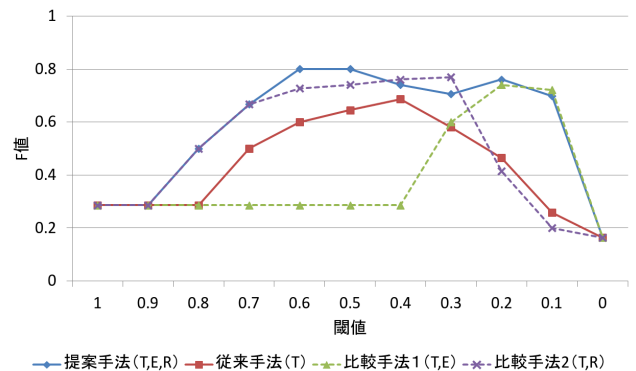
図 2 (b) に示したように、スペイン対フランスの試合に対して検討手法を適用した場合の F 値は、閾値が 0.5 になるまで従来手法と比較手法 1 の F 値を下回った。これは、この試合のゴールの中に PK によるものが含まれていたからだと考える。



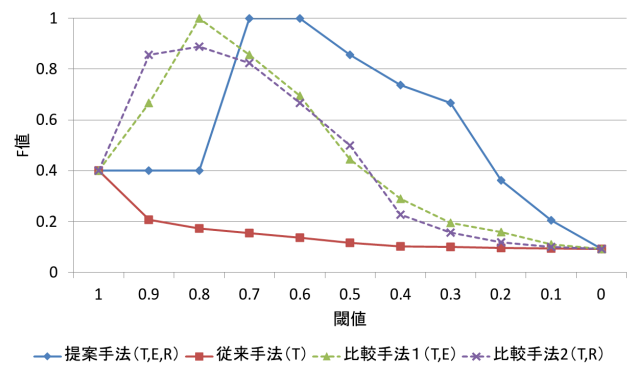
(a) 準々決勝 チェコ対ポルトガル



(b) 準々決勝 スペイン対フランス

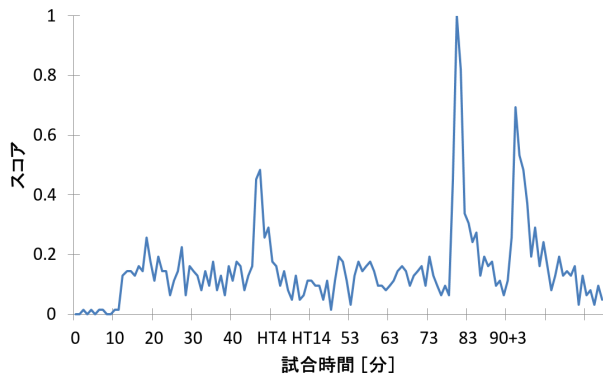


(c) 準々決勝 ドイツ対ギリシャ

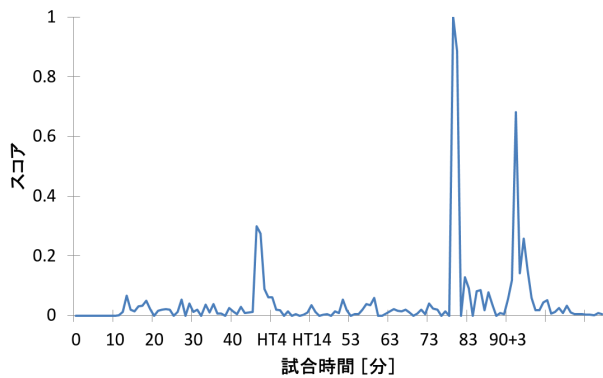


(d) 決勝 スペイン対イタリア

図 2 各手法を適用した場合の F 値



(a) 従来手法 (Tのみを利用)



(b) 検討手法 (T,E,Rを利用)

図3 チェコ対ポルトガルの試合中におけるゴール検出スコアの遷移 (HT はハーフタイムの意)

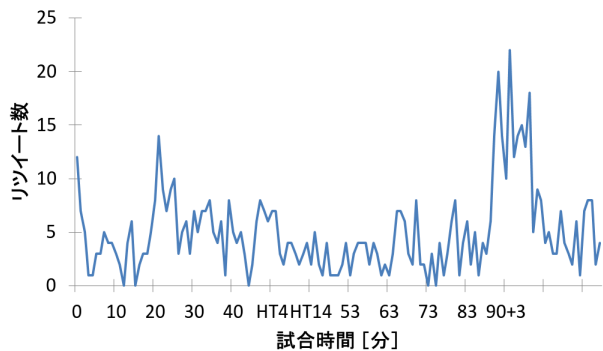


図4 スペイン対フランスの試合中におけるリツイート数の遷移 (HT はハーフタイムの意)

図4にスペイン対フランスの試合中におけるリツイート数の推移、図5にこの試合に対して検討手法を適用し算出したゴール検出スコアの遷移を示す。PKによってゴールが決まったのは90+1分の時点である。この時点でリツイート数が増大しており、ゴール検出スコアが減少していることが分かる。検討手法においてリツイート数をイベントの区切りを除去するために用いた理由の一つとして、流れの中で決まるゴールに関してはリツイートする余裕がないため、リツイート数が増大している箇所はゴールの場面ではないという考えがあった。しかし、PKは一旦試合が止まってから行われるため、観戦してい

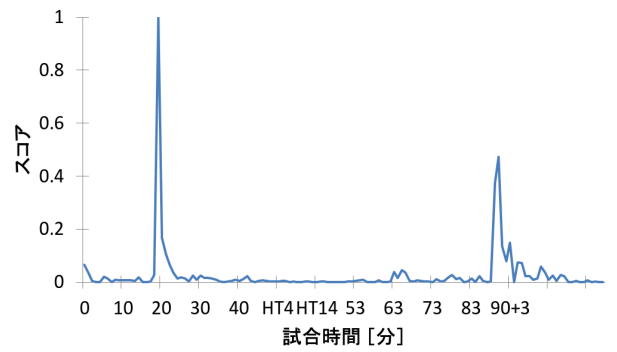


図5 スペイン対フランスの試合中におけるゴール検出スコアの遷移 (HT はハーフタイムの意)

るTwitterユーザにはリツイートをする余裕が生まれる。その結果、通常のゴールのときよりも早くリツイート数が増大し、ゴールと判定するスコアが小さくなってしまい、ゴールの検出精度が悪化するのだと思われる。この問題を解決するために、イベントの区切りにおいて増加せず、ゴールの瞬間において増加する特徴の更なる検討が必要である。

また、本手法は得点の決まりそうな場面を誤検出する傾向がある。図3(b)は、チェコ対ポルトガルの試合における、検討手法によって算出したゴール検出スコアの遷移である。この試合において、ゴールが決まったのは79分である。しかし、45分頃と90+3分頃にスコアが高くなっている。この時点の試合内容を確認すると、シュートがゴールの枠を直撃するというような得点になりかけた場面であった。検討手法では、“!”の文字数を興奮を表す特徴量として用いているため、興奮するような惜しい場面においても反応しているためであると考えられる。この問題を解決するために、今後はゴールのみに反応する特徴を探す必要がある。

本手法の傾向として、片方のチームを応援するツイートしか収集できない状態において、全てのゴールを検出できないということもあった。我々は、ロンドンオリンピック女子サッカー決勝の日本対アメリカの試合において、女子サッカー日本代表を示すハッシュタグ“#nadeshiko”を含むツイートを収集した。このハッシュタグを付記してツイートを投稿するユーザのほとんどは、日本代表を応援しているユーザである。そのツイートから、検討手法によってゴール検出スコアを算出した結果を図6に示す。この試合において、ゴールが決まったのは8分、54分、63分である。その内、日本代表のゴールは63分のものである。スコアを見ると、63分のゴールの場面のみ高くなっており、8分、54分の相手チームのゴールに反応していない。これは、興奮を表す特徴が相手チームのゴール時には増大しないためであると考えられる。この問題を解決するために、相手チームのゴール時に増大するであろう落胆を示す特徴か、両方のチームのゴール時に増大するような特徴を調査し、導入する必要がある。

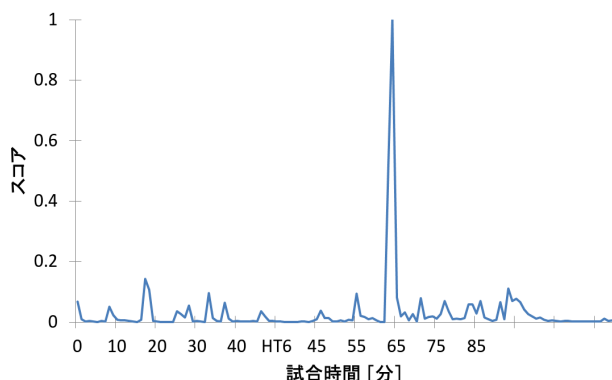


図 6 片方のチームを応援するツイートから検出手法によって算出したゴール検出スコアの遷移 (HT はハーフタイムの意)

## 6. ま と め

本稿では、Twitter に投稿されたツイートを基にスポーツの試合中におけるイベント検出を目的とし、ツイート中の特徴を用いた検出手法を検討した。検出手法では、従来手法で用いられていたツイート数のほかに、“!” の文字数、リツイート数を利用する。“!” の文字数は、興奮の度合いを示す特徴量として、ゴールの瞬間を検出するのに利用する。また、リツイート数は、リツイートは投稿するのに時間を要するという性質から、時間をかけて投稿する余裕の生まれる試合の区切りやイベントの区切りを検出・除去するための特徴量として利用する。評価実験の結果、検出手法は従来手法よりもゴール検出の性能に優れていることを確認した。今後は、ゴール検出スコアの計算方法の改善や、惜しいシーンとゴールの瞬間を区別、片方のチームを応援するツイートから両方のチームのゴールをそれぞれ検出するための新たな特徴の検証をし、より高精度なゴール検出手法を検討する。

## 文 献

- [1] Twitter Blog, <http://blog.twitter.com/2012/03/twitter-turns-six.html/>.
- [2] Semiocast, [http://semiocast.com/publications/2012\\_07\\_30\\_Twitter\\_reaches\\_half\\_a\\_billion\\_accounts\\_140m\\_in\\_the\\_US/](http://semiocast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US/).
- [3] Akshay Java, Tim Finin, Xiaodan Song, and Belle Tseng, Why we Twitter: Understanding microblogging usage and communities, In Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007, pp.56–65, 2007.
- [4] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, Earthquake shakes twitter users: Real-time event detection by social sensors, In Proceedings of the 19th International WWW Conference, pp.851–860, 2010.
- [5] Aron Culotta, Towards detecting influenza epidemics by analyzing twitter messages, In Proceedings of the First Workshop on Social Media Analytics, pp.115-122, 2010.
- [6] Twitter Blog, <http://blog.jp.twitter.com/2012/08/blog-post.13.html/>.
- [7] Twitter で 2011 年を振り返る, <http://yearinreview.twitter.com/ja/tps.html/>.
- [8] 小林 尊志, 野田 雅文, 出口 大輔, 高橋 友和, 井手 一郎, 村瀬 洋, Twitter の実況書き込みを利用したスポーツ映像の要約, 信学技報, MVE2010-162, 2011.
- [9] James Lanagan, and Alan F. Smeaton, Using Twitter to detect and tag important events in live sports, In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp.542–545, 2011.
- [10] Siqi Zhao, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan, Human as real-time sensors of social and physical events: A case study of twitter and sports games, Technical Report TR0620-2011, Rice University and Motorola Labs, 2011.
- [11] David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill, Tweet the debates: Understanding community annotation of uncollected sources, In Proceedings of the first SIGMM workshop on Social media, pp.3–10, 2009.