

# ソーシャルメディアを用いた臨場感のある スポーツハイライト自動生成

11471011 伊藤 大貴 (灘本研究室)

あらまし：本研究では、スポーツ観戦においてリアルタイムで試合を観戦できなかった人に、リアルタイムで観戦したような盛り上がりや臨場感を感じるスポーツハイライト自動生成手法の提案を行う。具体的には、複数のソーシャルメディアを用いて臨場感、盛り上がりが伝わる情報を抽出し、統合する手法の提案を行う。

## 1. はじめに

近年、Twitter や Instagram といったソーシャルメディアが普及している。スポーツ観戦においても、各シーンにおける観戦者の感情を他の観戦している人々と共有するためにソーシャルメディアの活用が増加している。リアルタイムでスポーツを見ている人たちはこのようなソーシャルメディアを用いることで、現地に行かなくても人々と情報や感情を共有することで観戦時の盛り上がり等を共有することができる。それに対し、リアルタイムでスポーツを見れなかった人たちは、録画したものを見たり、試合結果だけを確認することとなる。よって、その場で他の人々と情報や感情、盛り上がりを共有することができない為、リアルタイム観戦のように臨場感のあるスポーツ観戦を楽しめないのが現状である。実際にプロ野球を例に、試合のリアルタイム Tweet とハイライト映像のみを見た人の Tweet を比較してみると、ハイライト映像のみを見た人の Tweet は細かい試合情報がわからない為、盛り上がりには欠けている。そこでニコニコ動画やスポーツ用の掲示板を見ることで、リアルタイムの試合情報や観戦者の感情や盛り上がりを知るという方法がある。しかしながら、ニコニコ動画には様々なコメントがあり、情報がまとめられていない為、試合情報がわかりにくいという問題がある。また、スポーツ用の掲示板ではコメントが多すぎて盛り上がった場面がわかりにくいという問題がある。

そこで、本研究ではハイライト映像のみを見た人もリアルタイムで観戦したような盛り上がりを感じることを目的とし、複数のソーシャルメディアを活用し臨場感、盛り上がりが伝わるスポーツハイライトを自動生成する。本研究では、プロ野球のハイライト映像を研究の対象とする。

## 2. 提案手法

以下に提案システムの生成手順を示す。

- ① あるプロ野球の試合におけるコメントを取得する。コメントは Twitter, Yahoo の運営する掲示板である textream, ニコニコ動

画の 3 つのソーシャルメディアから抽出する。

- ② Twitter から試合内容がわかる一般 Tweet を抽出する。
- ③ textream から選手名、野球用語が含まれるコメントを抽出する。
- ④ ニコニコ動画から盛り上がりが伝わるコメントを抽出する。
- ⑤ ②③④からわかりやすく盛り上がりも伝わるコメントを抽出し、ハイライト映像を見ているユーザに提示する。

### 2.1. Tweet の抽出

本研究で作成したプロ野球球団辞書 144 語をクエリとして、Twitter から対象のプロ野球の試合が行われている時間のプロ野球に関する Tweet を抽出する。次に、抽出した Tweet を Repeated Bisection 法<sup>[1]</sup>を用いてクラスタリングし、選手名や野球用語が含まれる Tweet とそれ以外の Tweet に分類する。ここで、分類したプロ野球の Tweet と textream のプロ野球のコメントそれぞれの TF 値を比較すると、Tweet には一般的に用いる単語が多いことが判明した。そこで、分類したプロ野球の Tweet の中で出現頻度の高い 10 単語をクエリとし、再度 Tweet を抽出する。ここで抽出した Tweet を一般的野球 Tweet と呼び、提示する。表 1 に提示する一般的野球 Tweet の例を示す。

表 1：一般的野球 Tweet の例

広島やっぱ強いなあ〜
今日の阪神は良かった
坂本ここで打って！

### 2.2. textream コメントの抽出

対象のプロ野球の試合のリアルタイムコメントを手動で作成した野球用語辞書 288 語をクエリとして textream から抽出する。ここで言う野球用語辞書とは、プロ野球球団辞書に野球の専門用語を追加したものである。次に、リアルタイムのコメントの感情を共有する為に、抽出したコメントの感情値を算出する。感情値の算出には小林らが構築している日本語評価極性辞書<sup>[2]</sup>

を用いる。日本語評価極性辞書の極性は p(ポジティブ), n(ネガティブ), e(ニュートラル)の3軸から成る。抽出したコメントと日本語評価極性辞書を照合し、以下の式を用いて感情値を求めらる。

$$s_i = \frac{|p| - |n|}{|p| + |n| + |e|}$$

$s_i$ はコメント  $i$  の感情値を示し、 $|p|$ ,  $|n|$ ,  $|e|$ はコメント  $i$  に含まれる各属性を持つ単語の個数を示す。算出した感情値から、感情がポジティブなコメントを応援コメント、感情がネガティブなコメントを批判コメントとし、これらのコメントを提示する。表 2 に提示する応援コメントと批判コメントの例を示す。

表 2：提示するコメントの例

応援コメント	坂本ナイスヒット！
批判コメント	藤浪コントロール悪すぎ

### 2.3. ニコニコ動画コメントの抽出

対象のニコニコ動画のプロ野球ハイライト映像のコメントから、正規表現を用いて叫喚フレーズを抽出する。叫喚フレーズの抽出には、浅井ら[4]の提案している手法を用いる。尚、浅井らの叫喚フレーズの定義に以下の条件を追加する。

追加条件：「w」「ー」「8」が2回以上繰り返して付加されている。

抽出した叫喚フレーズを盛り上がったコメントとして提示する。表 3 に提示する叫喚フレーズの例を示す。

表 3：叫喚フレーズの例

丸タイムリー—————！！！！！！
きたああああ

### 2.4. コメントの統合

抽出した Tweet と textream コメントを時間軸でソートし、1分ごとの出現頻度上位 2 単語をクエリとしてハイライトのシーンごとに分類したニコニコ動画のコメントに挿入し、統合して提示する。提示画面を図 1 に示す。

## 3. 評価実験

### 3.1. 実験内容

提案手法の有用性を示すために、評価実験を行った。被験者は 8 名である。被験者にある日のプロ野球のハイライトとして、図 1 に示す提案システムとニコニコ動画を見てもらい、被験者の野球関心度、提案手法の評価、提案手法とニコニコ動画の比較評価の 3 つの視点で評価した。設問には「言える」「どちらかといえば言える」「どちらとも言えない」「どちらかといえば言えない」「言えない」の 5 段階で評価を行った。

### 3.2. 結果と考察

提案手法の評価において、試合の盛り上がり



図 1：提案システム

験者がニコニコ動画は 12.5%だったのに対し、提案システムは 25%だった。これにより、提案システムの方が比較的盛り上がり

が伝わることがわかった。また、リアルタイムで観戦しているような臨場感が伝わるかの問いには、ニコニコ動画は 62.5%が「言えない」もしくは「どちらかといえば言えない」という回答に対し、提案システムは 62.5%が「言える」もしくは「どちらかといえば言える」と回答した。これにより、提案システムの方が臨場感が伝わり

と言え

る。さらに、試合の内容が理解できるかの問いに対して「言える」と答えた被験者がニコニコ動画はいなかったのに対し、提案システムは 37.5%いた。よって、提案システムの方が試合の内容が理解できると言える。以上の結果より、提案システムの方が試合内容が理解しやすく、臨場感、盛り上がり

が伝わる

ことがわかった。また、野球関心度が低い被験者は、提案システムで特定のシーンでの盛り上がり

が伝わるかの問いに対し、40%が「どちらかといえば言えない」もしくは「言えない」と回答した。よって、野球関心度が低いと盛り上がり

が伝わり

にくいこともわかった。

### 4. まとめと今後の課題

本研究では、複数のソーシャルメディアを活用し臨場感、盛り上がり

が伝わる

スポーツハイライト自動生成手法の提案を行った。さらに提案手法の評価実験も行った。評価実験により、ニコニコ動画を見るよりも盛り上がり

や臨場感、試合内容が伝わることを確認した。今後の課題は、コメントのネタバレ防止やシーン判定の改善を行うことである。さらに、対象のスポーツの知識が少ないと試合内容や盛り上がり

が伝わり

にくい傾向がある為、ハイライト映像に対象のスポーツの補足情報を提示することも今後の課題である。

#### 参考文献

[1] Y. Zhao and G. Karypis, "Comparison of Agglomerative and Partitional Document Clustering Algorithms", University of Minnesota, pp.2-14, 2002.

[2] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "意見抽出のための評価表現の収集", 自然言語処理, Vol.12, No.3, pp.203-222, 2005.

[3] 浅井洋樹, 秋岡明香, 山名早人, "きたああああああああああ！！！！！！11:マイクブログを用いた教師なし叫喚フレーズ抽出", DEIM Forum, A4-4, 2013.