

Twitter データを用いたテレビ番組のイベント検出に関する研究

深沢 知明[†] 高島 真之介[†] 羽山 徹彩[†]

金沢工業大学

1. はじめに

テレビ番組は多種多様なテーマに対し、映像や音声を駆使して製作され、その長年に渡り蓄積され続けてきた番組動画データは膨大な知識資源である。それにも関わらず、その利活用技術はほとんど整備されてこなかった。近年、各番組動画に対し、手作業でメタタグを付与したり、音声・画像特徴を利用し予め定義された重要シーンに対し特定する方法が開発されたりしてきた[1][2]。しかしながら、逐次蓄積される番組動画の利活用性を高めるには番組内の重要なシーンに対し、多様なメタタグを自動的に付与する必要があるが、現状の技術ではほぼ不可能といえる。

近年、テレビを視聴しながら Twitter を使用するユーザが膨大となった。我々はそのような Tweet データを収集し解析することで、重要なシーンとしての番組内のイベント時間帯とその内容が把握できると考えた。そこで本研究では Twitter データを用いて、テレビ番組のイベントを高精度に検出する方法の開発を目的とし、実施した。

2. Twitter データを用いた番組内のイベント検出

2.1 設定

本研究のイベント検出では番組時間に発信された Tweet の頻度時系列データに対し、Kleinberg のバースト検出法[3]を適用する。

Tweet 頻度時系列データには番組に関する Twitter ユーザとその Follower を対象に対象番組時間内に発信された Tweet を収集し、番組放映時間に対して Tweet 頻度を時系列化して用いる。番組に関する Tweet を大量に収集するために、番組を視聴している可能性のあるユーザが発信した Tweet を収集している。

Kleinberg のバースト検出法は時系列データの異常を検出するために、最も利用されているアルゴリズムのひとつである。このアルゴリズムではイベントが急激に発生するバースト期間と非バースト期間、およびバースト度合いを割り当てることができる。その手順を簡単に説明すると、イベントが発生する時間にバースト・非バーストを判断す

るためにイベント発生時間間隔を用いた期待値を設定し、各イベント発生時間に対しその発生間隔に応じて状態遷移確率を割り当てる。また不自然な状態遷移を防ぐために、状態遷移のコストを付与する。最後に、時系列データのバースト・非バーストの全ての状態遷移の組み合わせから、Viterbi アルゴリズムにより最小コストの状態遷移を決定する。その結果、バースト期間が番組時間帯のなかで番組時間内に重要なイベントが発生した時間帯として検出する。

2.2 番組のイベント検出に対する問題点

番組時間内の Tweet 頻度時系列データに対し、バースト検出法をそのまま適用すると、以下のような問題が発生する。

- (1) 検出精度の問題：全てのイベントの検出ができない場合がある
- (2) 正確性の問題：検出されたイベントのバースト期間が実際よりも長い場合がある

このような問題に対し、以下の原因が考えられる。

- 番組を視聴しながら発信された Tweet であっても、コミュニケーションや他の Tweet へのコメントである場合には発信時間と実際のイベント発生時間とのズレが生じる。
- 番組の Tweet 収集数が少ない場合には番組に関係ない Tweet の影響を受け易い。

3. 改善手法

本研究の番組内のイベント検出法の手順では、収集した Tweet データを頻度時系列化し、バースト検出法を適用する。本改善手法では、Twitter データを頻度時系列化する過程で、以下の 3 つを適用する。

(1) リツイート・リプライツイートの除去

番組のイベント発生後に、それを話題として他のユーザに対しコミュニケーションしたり伝達したりするようなツイートを阻害するで、検出精度が高められる。

(2) 頻出語を発信するユーザのツイートの選出

番組時間帯に頻出語（名詞）を含んだツイートをするユーザだけを選出し、そのユーザが発信するツイートだけでバースト検出する。それによって、番組を視聴してツイートを発信しているといった信頼性の高いツイートだけをイベント検出に利用することができる。

Study on event detection of TV programs using Twitter data

Fukasawa Tomoaki · Kanazawa Institute of Technology
Shinnosuke Takasima · Kanazawa Institute of Technology
Tessai Hayama · Kanazawa Institute of Technology

(3) 時間系列におけるツイート頻度の偏差値化

各番組放映時間のツイート頻度に対して、番組全体のツイート頻度数での位置付けた値に変換することで、ツイートの反応が鈍い番組に対しても、一定数のイベントを検出することができる。

4. 評価実験

4.1 評価方法

本実験では本改善手法の有効性を示すために、Twitter データを用いたテレビ番組の主要なイベント検出精度によって従来手法と比較した。従来手法には Kleinberg のバースト検出を Twitter データに対しそのまま適用した方法を用いた

実験データには 2014 年に開催されたサッカーブラジルワールドカップの 5 試合に対し、サッカー関係の Twitter ユーザとその Follower が番組放映時間中に発信した Tweet を収集し、使用した。主要なイベント検出結果の正解データには日刊スポーツの Web ニュースの記事をもとに、ゴールシーン、シュートシーン、反則シーン、および選手交代シーンに対してのイベント発生時間を用いた。検出精度については正解データの各イベントに対し、算出した。

4.2 評価結果

番組内の主要なイベントの検出精度の結果とツイ

ート時系列データの例について、表 1 および図 1 に示す。

本改善手法のそれぞれが、従来手法よりも検出精度が高かった。そのため、本改善手法はテレビ番組の主要なイベント検出に有効であるといえる。しかしながら、シュートシーンに対してはいずれの方法も比較的検出できていない。このことは Twitter ユーザがそれほどは反応していないことが原因である。今後は微細な Tweet 反応に対しても、新たな改善手法を開発していきたい。

5. まとめ

本研究では Twitter データを用いて、テレビ番組のイベントを高精度に検出する方法の開発を実施し、検出精度の改善を行った。

参考文献

- [1] 山本大樹, 清水大輔, 渡邊 睦: サッカー映像のシーン自動解析の研究, 電子情報通信学会技術報告書, vol.104, no. 573, PRMU2004-176, pp. 73- 78 (2005)
- [2] Jinjun Wang, Changsheng Xu, Engsiong Chng and Qi Tian: Sport Highlight Detection from Keyword Sequences using HMM, In Proc. IEEE ICME, pp. 27-30, 2004.
- [3] J. Kleinberg. Bursty and hierarchical structure in streams. In Proc. 8th SIGKDD, pp. 91- 101, 2002.

表 1. 番組内の主要なイベントの検出精度の結果

番組 (試合)	ゴールシーン				シュートシーン				反則シーン				交代シーン			
	Pr	Re	Se	St	Pr	Re	Se	St	Pr	Re	Se	St	Pr	Re	Se	St
JPNvsCIV	3/3	3/3	3/3	3/3	0/23	1/23	0/23	6/23	1/4	1/4	1/4	3/4	0/6	1/6	1/6	4/6
JPNvsCOL	5/5	5/5	4/5	4/5	2/31	3/31	2/31	7/31	1/2	1/4	1/4	4/4	2/6	3/6	2/6	2/6
ESPvsNED	6/6	6/6	6/6	6/6	4/16	5/16	3/16	8/16	1/4	1/4	1/4	3/4	0/6	1/6	3/6	4/6
GERvsBRA	8/8	8/8	8/8	8/8	8/24	8/24	8/24	2/24	0/1	0/1	0/1	1/1	1/6	1/6	1/6	6/6

Pr.:従来手法, Re.:不要ツイートの除去, Se.:ユーザ選定, St.: 偏差値化

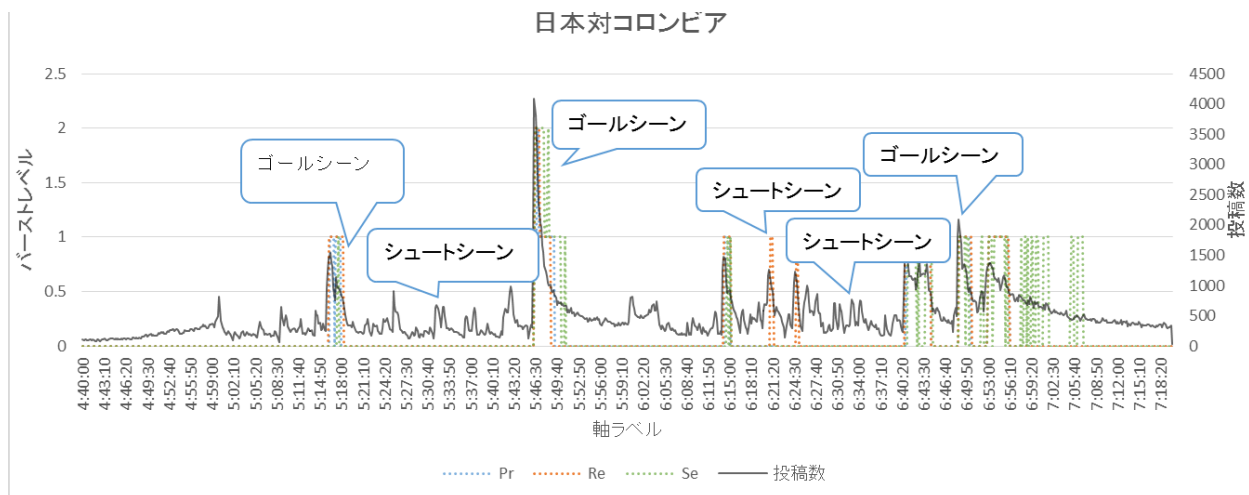


図 1. テレビ番組のツイート時系列データとバースト検出結果の例