

“良い実況者”に着目したTwitterからのスポーツ速報生成

久保 光証[†] 笹野 遼平[‡] 高村 大也[‡] 奥村 学[‡]

[†]東京工業大学 総合理工学研究科, [‡]東京工業大学 精密工学研究所

[†]kubo@lr.pi.titech.ac.jp, [‡]{sasano, takamura, oku}@pi.titech.ac.jp

1 はじめに

Twitter では tweet と呼ばれる 140 字以内の短いテキスト情報をベースに、様々な情報がやり取りされている。テレビで中継されるイベントの実況や感想もその内の一つであり、ニールセンの調査¹によれば、アメリカにおける Twitter のアクティブユーザの約三分の一がテレビ番組に関する tweet をしている。このようなイベントに関する tweet は、ハッシュタグと呼ばれる検索を容易にするためのキーワードとともに投稿されることも多い。ユーザはハッシュタグを検索キーワードとして検索し、そのイベントに関連して投稿されている tweet をスマートフォンなどで得ることができる。しかしながら、瞬時に流れていく膨大な量の tweet から、特定のイベントの状況を過去の経緯を含め正確に把握するのは容易ではない。

一方、膨大な量のイベント関連 tweet には重要な出来事(サブイベント)に関する情報が含まれており、それらを機械的に要約することで、分かりやすいイベント要約を作成できると考えられる。そこで本研究ではイベント関連 tweet からその要約を速報という形で生成する手法を提案する。本稿では時間とともに到着するイベントの実況や感想をイベントストリームと呼ぶ。イベントストリームでは、イベント中で重要な出来事が起きると、多数のユーザが tweet するため急激に tweet 数が増加する、という特徴がある。このような現象を本稿ではバーストと呼ぶ。本研究ではバースト時の tweet を迅速に要約することによって、イベントで起こっている重要なことを、速報としてリアルタイムで分かりやすくユーザに提示することを目的とする。

イベントストリームにおける tweet の例を図 1 に示す。イベントストリームに現れるユーザの中には、ユーザ B のように、試合状況を迅速に、かつ、冷静に tweet するユーザもいる。このようなユーザを本稿では“良い実況者”と呼ぶ。つまり、ユーザ A のように、

サブイベント1		サブイベント2	
id:1	ユーザA キター!!!!	ユーザC 長友にイエロー	id:4
id:2	ユーザB 松井のパスから本田のシュートで先制!	ユーザA まじかよ、この時間にもらうのは痛いな...	id:5
id:3	ユーザC ナイスシュート!	ユーザB 遅延行為で長友にイエローカード	id:6

図 1: イベントストリームにおける tweet の例

個人の単なる感想が tweet の中心であるユーザは“良い実況者”ではない。また、ユーザ B とユーザ C を比較して分かるように、“良い実況者”であるといってもその度合いには大小がある(ユーザ B > ユーザ C)。

提案手法ではユーザ B のような“良い実況者”に着目することにより、そうでない場合と比べて、イベントにおける重要な場면을説明するために必要な情報を、多く被覆している tweet を選択し、速報という形で出力する。

2 関連研究

Sharifi ら [4] はイベントストリームの要約を、与えられたクエリを含む頻出単語列パターンを抽出することで生成した。Sharifi らの手法は、流行しているキーワードなど特定のフレーズを含む膨大な量の tweet の中から、その中心となっている単語列を見つけることを目的としており、時間によってトピックが移り変わることを捉えようとしたものではなく、問題設定や目的が本研究と異なる。

Chakrabarti ら [1] はサブイベントの移り変わりを隠れマルコフモデルで表現した。Chakrabarti らの手法ではサブイベントの時間幅を固定する必要があるため、アメリカンフットボールの試合のように、サブイベントの区切りがはっきりしているイベントに対しては有効であるが、本稿で扱うサッカーの試合のような、サブイベントの区切りがはっきりしていないイベントに対しては適切ではない。

高村ら [5] はイベントストリームから代表的な tweet

¹<http://blog.nielsen.com/nielsenwire/social/2012/>

を選択することによりそのイベントの要約とした。具体的には、イベントストリームから代表的な tweet を選択するという問題を施設配置問題の一種に落としこみ、線形分割条件を加えることによって、問題を高速に解く手法を提案した。だが高村らの手法はイベントが終了してからバッチ処理を行うことを前提としており、イベントストリームをリアルタイムで処理することはできない。

そして、Nichols ら [3] は本稿の問題設定と類似したイベントストリームの要約手法を提案した。Nichols らの手法ではまず、イベントストリームから単位時間あたりの tweet 数の増減を利用してサブイベントの区間を特定する。その後、各サブイベント中の tweet 集合に対して、Sharifi らの手法をベースにフレーズグラフを構築し、各 tweet に単語の頻度でスコアを付ける。そしてスコアの高いものから順に、トピック単語が重複しないように、設定した個数の tweet を選択し要約とする。しかしながら、Nichols らの手法は要約生成の際に、その tweet 集合に含まれる単語の頻度に大きく依存するという特性がある。図1で見たように、サブイベント開始直後には id:1, id:3 のような tweet が多数を占め、id:2 のような tweet は少数である。そのため Nichols らの手法では、id:3, id:4 のようなあまり説明的でない tweet が選ばれる傾向にあり、状況を客観的に描写する速報として望ましい id:2, id:6 のような tweet が選ばれにくい。そこで本研究では、説明性を担保する基準を導入し、この問題を解決する。

3 提案手法

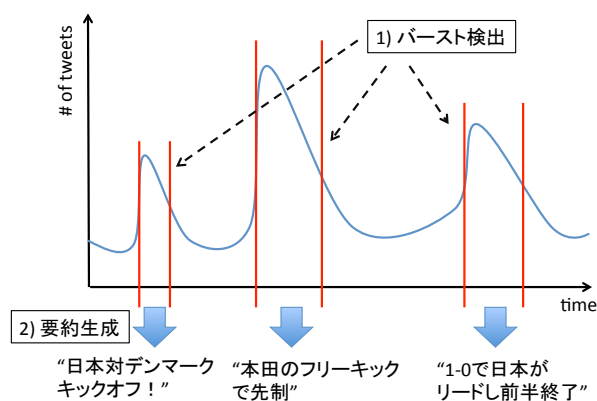


図 2: 提案手法の概要

図2に示すように、提案手法は大きく分けて、1) イベントストリームにおけるバースト検出と、2) 各セグメントに対する要約生成から成る。まず、入力データであるイベントストリーム中の tweet を、バースト状態にある tweet 集合ごとにセグメントに分ける。次に

セグメントのサイズがあらかじめ定めた閾値以上を超えたときイベント中にサブイベントが発生したものと考え、サブイベントの発生から短い時間が経過した時点で要約を生成し、速報として出力する。その際に、サブイベントの開始直後に説明的な tweet をしている“良い実況者”に着目することで、サブイベント開始直後であっても、説明的で冷静な tweet を選択することができる。

3.1 バースト状態の検出

イベントストリームの特徴として、サブイベントが発生した瞬間に急激に時間あたりの tweet 数が増える、というバースト現象が挙げられる。本稿では 10 秒あたりの tweet 数を時間あたりの tweet 数とし、その tweet 数が、非バースト状態のときの中央値の 2 倍よりも大きくなっている状態をバースト状態とした。イベントストリームにおいて、バースト時に得られる tweet は、その大半が同一のサブイベントについて言及している。そこでバースト状態が連続する部分が、1 つのセグメントとなるようにセグメンテーションを行った。

3.2 速報としての要約生成

バースト状態として検出された各セグメントに対して、セグメントの最初の tweet が到着した時刻をサブイベントの開始時刻と見なし、サブイベントの開始から一定時間 T が経過した時点でそのセグメントの要約を生成し、速報として出力する。

要約の生成時にはユーザ及び tweet の説明性を考慮する。多くの場合、イベントにおいて、サブイベントは動作の主体である名前と、その専門用語の組み合わせで説明できる。例えばサッカーであれば「(選手名)のシュート」、「(選手名)にイエローカード」、野球であれば「(選手名)がホームラン」などである。つまり、選手名と専門用語の共起回数が多ければ、サブイベントをより詳細に説明していると考えられるため、本研究ではこれを説明性の指標の 1 つとする (tweet スコア)。

また、そのような説明的な tweet を、サブイベントの開始から迅速に投稿しているユーザのスコアが高くなるようなスコア (ユーザスコア) をユーザごとに割り当て、これを考慮することによって最終的な速報を出力する。

3.2.1 説明性を考慮した tweet スコアの設定

速報として tweet 集合の要約を出力する際には、その tweet 集合において、誰が何をしたのかを具体的に

描写している, “良い実況 tweet” を選択することが望ましい. そのような tweet を選択するため, バースト検出された各セグメント中の tweet tw に対して, 以下の (1) 式で tweet スコア S_{TW} を定める. tweet スコアが高いほど状況を説明的に描写しているといえる:

$$S_{TW}(g_n, tw) = sim_{cos}(g_n, tw) \cdot num_{co}(tw). \quad (1)$$

ここで $sim_{cos}(g_n, tw)$ は n 番目のセグメント g_n に含まれる tweet 集合と tweet tw との余弦類似度であり, 与えられた tweet がどの程度セグメント全体の情報を表しているかの指標となる. $num_{co}(tw)$ はイベント(試合)に出場している選手名 (p) とイベント固有の専門用語 (d) のうち, tw において少ない方の出現回数である. 図 1 のサブイベント 1 の例では, num_{co} の値は, $id:1$ が 0, $id:2$ は $p = \{ \text{松井, 本田} \}, d = \{ \text{パス, シュート} \}$ なので 2, $id:3$ は $p = \{ \}, d = \{ \text{シュート} \}$ なので 0 となる. sim_{cos} と num_{co} を掛け合わせた数値が大きければ良い実況 tweet であるとみなす.

3.2.2 速報として最適な tweet の選択

前節で述べた tweet スコアをユーザごとに, またサブイベントごとにその平均値を計算する. このスコアが高いユーザは, サブイベントにおいて良い実況 tweet をしている, “良い実況者” であると考えられる.

n 番目のサブイベントが終了した時点でのユーザ u のユーザスコア S_{U_n} は, (2) 式のようになる:

$$S_{U_n}(u) = \frac{1}{c_n(u)} \sum_{k=1}^n S_{TW}(g_k, tw_{u_k}). \quad (2)$$

ここで $c_n(u)$ はユーザ u が n 番目までのサブイベントのうち, 少なくとも 1 回の tweet をその中で投稿しているサブイベントの数である. また, tw_{u_k} は k 番目のセグメントにおいて, ユーザ u の tweet のうち, S_{TW} の値が最も大きい tweet である.

最後に, (1) 式と (2) 式を合わせて考え, n 番目のサブイベントの要約として (3) 式のレポートスコア S_{R_n} の値が最も大きい tweet を選択して出力する:

$$S_{R_n}(u, g_n, tw) = \lambda \cdot S_{U_{n-1}}(u) + (1 - \lambda) \cdot S_{TW}(g_n, tw). \quad (3)$$

ただし, λ は $S_{U_{n-1}}$ への重みであり, $0 \leq \lambda \leq 1$ を満たす.

4 実験

4.1 データセット

サッカーの試合がテレビ中継されている時間帯のデータを 6 試合分用意した. 2010 年 FIFA ワールド

表 1: 各試合の統計量

対戦国	試合日	tweet 数	ユニークユーザ数
カメルーン	2010 年 6 月 14 日	61,666	9,809
オランダ	2010 年 6 月 19 日	56,976	8,201
デンマーク	2010 年 6 月 24 日	93,336	10,771
オマーン	2012 年 6 月 3 日	33,113	5,365
ヨルダン	2012 年 6 月 8 日	32,091	4,463
オーストラリア	2012 年 6 月 12 日	49,866	5,943

カップグループステージの日本代表戦 3 試合 (対カメルーン, 対オランダ, 対デンマーク) と, 2012 年に行われた 2014 年 FIFA ワールドカップアジア最終予選の日本代表戦 3 試合 (対オマーン, 対ヨルダン, 対オーストラリア) である. これらすべての試合の tweet は Twitter の Streaming API の status/filter² を用い, 関連するハッシュタグ (#soccer, #jfa, #daihyo など) を含むものだけを集めた. これらの試合の統計量について表 1 にまとめた.

また出場する選手名やチーム名は試合が行われる前に分かるためそれらは既知であるとし, そのリストを作成した. サッカー用語については過去に行われたサッカーの試合の速報テキストから複数回出現しているものを抽出し, 人手で整備したものを用いた. そして評価で使用する正解要約にはヤフーが提供するスポーツナビ³によるテキスト速報を用いた.

4.2 スポーツ速報のための評価尺度

イベントストリームの要約のための自動評価尺度を導入する. まず正解要約として Web 上で公開されている新聞社やスポーツニュースサイトのテキスト速報を用いる. 文書要約の評価指標としてよく用いられる ROUGE[2] は, 正解要約に含まれる単語のうち, 生成された要約に含まれる単語の割合を表す. しかしイベントストリームの要約にはそれぞれイベント開始からの時刻が紐付いており, たとえ正解要約中の単語が生成要約中に出現していても, それらが時間的に離れた位置にあるならば正解と見なすべきではない. そこで自動評価指標として高村ら [5] と同様に修正版 ROUGE を利用する. これは通常の ROUGE とは異なり, 正解要約中の単語は, 決められた時間差内の生成要約にその単語が出現した場合のみ被覆されていると見なされる. 本稿での実験では速報という観点からこの差を 3 分に定めた. また ROUGE-1 の計算には, 名詞, 動詞, 形容詞のみを用いた.

²<https://dev.twitter.com/docs/streaming-apis/parameters>

³<http://sportsnavi.yahoo.co.jp/>

表 2: ROUGE-1 スコア, $T = 30$ 秒

対戦国	Nichols ら	ユーザスコアなし	提案手法
カメルーン	0.0690	0.0979	0.110 ($\lambda = 0.3$)
オランダ	0.102	0.140	0.137 ($\lambda = 0.3$)
デンマーク	0.0837	0.117	0.132 ($\lambda = 0.3$)
オマーン	0.0876	0.121	0.127 ($\lambda = 0.3$)
ヨルダン	0.107	0.131	0.130 ($\lambda = 0.3$)
オーストラリア	0.117	0.118	0.155 ($\lambda = 0.3$)

表 3: ROUGE-1 スコア, $T = 60$ 秒

対戦国	Nichols ら	ユーザスコアなし	提案手法
カメルーン	0.0765	0.107	0.103 ($\lambda = 0.2$)
オランダ	0.106	0.157	0.163 ($\lambda = 0.3$)
デンマーク	0.0971	0.127	0.125 ($\lambda = 0.3$)
オマーン	0.0958	0.130	0.150 ($\lambda = 0.3$)
ヨルダン	0.137	0.149	0.159 ($\lambda = 0.3$)
オーストラリア	0.142	0.128	0.180 ($\lambda = 0.3$)

4.3 実験設定

サブイベントの開始から速報としての要約生成を行うまでの時間 T には 30, 60 秒の 2 つの値を用いた。また、ある 1 試合の λ は、他の 5 試合のデータを用いて決めた。

また、本稿で比較するベースライン手法として Nichols ら [3] の手法を用いた。ただしイベントストリームのセグメンテーションは同様の手法で行い、ベースライン手法で生成された要約を提案手法の要約結果と比較した。

4.4 実験結果

実験の結果を表 2 と表 3 に、出力例を表 4 に示す。表中のユーザスコアなしというのは提案手法で $\lambda = 0$ とした場合である。 λ の値は、 $T = 30$ 秒の場合、すべての試合において $\lambda = 0.3$ が使われ、 $T = 60$ 秒の場合はカメルーン戦を除き、すべての試合において $\lambda = 0.3$ であった。いずれの場合でも提案手法が Nichols らの手法を上回った。ただし Nichols らの手法では選手名やサッカー用語辞書を使用していないため、提案手法に有利な条件であると考えられる。

サブイベントの開始直後は感情的な tweet が大半を占め、客観的に状況を描写している tweet は多くない。Nichols らの手法はそのクラスターで使われている単語の頻度に大きく影響を受けるため、本稿のように迅速に速報を出力することにはあまり向いていないことが分かる。一方提案手法では選手名とサッカー用語が共起している頻度を考慮しているため、速報に必要な情報を含む、より客観的な tweet を抽出できている。さ

表 4: 対オマーン, $T = 30$ 秒

時刻	手法	速報内容
31 分	Nichols ら	工(` π `)工 これが本田と岡崎の違いか ... 今のは決めないと ...
	提案手法	本田から長友のクロス、岡崎のヘッドはゴール右 !!
44 分	Nichols ら	今日は岡崎の日じゃないな w しかし本田は長友をよく見てる
	提案手法	本田→長友のクロス。最後は岡崎が合わせるもミートせず ...
106 分	Nichols ら	すげえ、今キーパーどうやって止めた?
	提案手法	酒井のスルーパスから本田～ さらに清武一好セーブに阻まれたああ

らにはユーザスコアを考慮することにより、すべての試合で、そうでない場合より修正版 ROUGE-1 で上回るか、ほぼ同等であることが確認できた。

わずかながらユーザスコアを考慮していない場合が上回っている試合においても、それぞれに最適な $\lambda (\neq 0)$ を設定することによって、ROUGE-1 が最も高い値になることが確認できた。 λ の決め方についての詳細な調査は今後の課題である。

5 おわりに

本稿では試合状況を迅速に伝える、“良い実況者”に着目したスポーツ速報生成手法を提案した。単なるイベント要約ではなく、速報という形で状況を伝える新たな要約システムを提案した。

今後の課題として、異なる種類のスポーツへの適用、さらには選手名やイベント専門用語を既知としないような手法の開発が挙げられる。また同一人物を指す、複数の呼称（選手の愛称など）の名寄せなども精度向上に役立つであろう。

参考文献

- [1] Chakrabarti Deepayan and Punera Kunal. Event summarization using tweets. In *Proceedings of the Fifth International AAI Conference on Weblogs and Social Media*, pp. 66–73, 2011.
- [2] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In Stan Szpakowicz Marie-Francine Moens, editor, *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pp. 74–81, 2004.
- [3] Jeffrey Nichols, Jalal Mahmud, and Clemens Drews. Summarizing sporting events using twitter. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, IUI '12*, pp. 189–198, 2012.
- [4] Beaux Sharifi, Mark-Anthony Hutton, and Jugal Kalita. Summarizing microblogs automatically. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10*, pp. 685–688, 2010.
- [5] Hiroya Takamura, Hikaru Yokono, and Manabu Okumura. Summarizing a document stream. In *Proceedings of the 33rd European conference on Advances in information retrieval, ECIR'11*, pp. 177–188, 2011.