

バースト検出と良い実況者に基づくリアルタイム状況要約手法

赤木, 文彦 / AKAGI, FUMIHIKO

(出版者 / Publisher)

法政大学大学院理工学・工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学・工学研究科編 / 法政大学大学院紀要. 理工学・工学研究科編

(巻 / Volume)

57

(開始ページ / Start Page)

1

(終了ページ / End Page)

8

(発行年 / Year)

2016-03-24

(URL)

<https://doi.org/10.15002/00013087>

バースト検出と良い実況者に基づくリアルタイム状況要約手法

REAL-TIME SUMMERISING BASED ON BURST DETECTION AND AUTHENTIC REPORTERS

赤木 文彦

FUMIHIKO AKAKI

指導教員 藤井 章博

法政大学大学院理工学研究科応用情報工学専攻修士課程

Recently Microblogs such as Twitter become popular. The application treats short messages in real time. Users can indicate whether the messages are categorized by attaching so called ‘hash-tag’ according to their interests. When a certain social event such as sport game is happening, the sequence of message with a group of ‘hash-tag’ describes the discourse of the event. Once a special incident such as shooting in a soccer game occurs, sequence of tweets shows burst in time-line. In this research, in order to understand the situation in large social event, both burst phenomenon of messages and words appears in messages are utilized. Algorithm named ‘aggregation pyramid’ is used for detecting group of messages in an event and analysis based on ‘tf-idf’ for extracting words to describe the situation when a burst is happening. To evaluate proposed automatic extraction of specific incident in a soccer game, human summary and the results of machine extraction is compared. The results shows good matches and availability of proposed method.

1. はじめに

近年, Twitter[1]などのマイクロブログが多くの人々に利用され, 情報の配信や共有が行われている. Twitter を利用するユーザは 140 字以内の tweet と呼ばれるテキストをやりとりすることで情報の配信が可能となる. Twitter の利用方法の中でハッシュタグに注目する. これは, テレビ番組に対する感想や内容を投稿するために用いられる場合に, 検索を容易にするためのキーワードとして付与される. 付与されたハッシュタグによって, ユーザは関心のあるイベントを特定して tweet を閲覧するこ

とができる. このような, ある特定のイベントに関して投稿された tweet は” イベントストリーム” と呼ばれ, イベント実況状況を如実に表現している. しかし, イベントストリームは移り変わりが頻繁であり, ユーザが関心ある tweet を閲覧し, イベントストリーム内の内容を理解することは容易ではない. そこで, イベントストリームの解釈を適切に行えばリアルタイムに状況の把握が行える. また, テレビ番組をリアルタイムに視聴していないユーザが過去の tweet をハッシュタグから特定し, 閲覧した際にも同様の問題が起こる. そこで本研究では

イベントストリーム中の特定の状態を抽出することで状況を適切に把握することを目的とする。イベントストリームでは、視聴者の関心の高まりにより tweet 数が急激に上昇する現象がある。この現象は”バースト”と呼ばれている。バーストの生起要因は様々考えられるが、特にニュースやスポーツ番組、ラジオ放送といった速報性の高いメディアとの関連が強い[6]。

本研究では、イベントストリームに関してバーストをリアルタイムに解析し速報性のある要約を生成することを目的とし、過去の tweet からユーザが視聴番組の内容を理解することに対する支援を目指す。ここで、情報の配信が最も多く行われた日本対ギリシャ戦の推移を以下に示す。図1は1分毎の tweet 数の推移、図2は1秒毎の tweet 数の推移である。

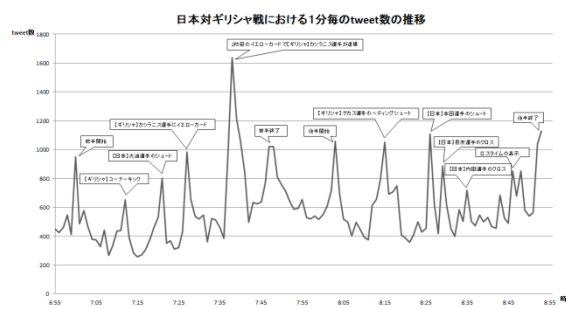


図1 日本対ギリシャ戦(1分毎の tweet 数の推移)

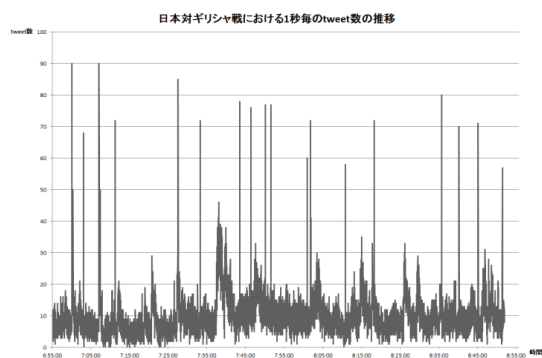


図2 日本対ギリシャ戦に(1秒毎の tweet 数の推移)

本研究では、イベントの内容を要約するためにイベントストリーム内のバーストを利用する。イベントを決定するためにバースト検出を適用し、

バーストが発生している期間をイベントであるとする。この時、各イベント期間内における tweet を解析することで要約を生成し、時間順に並べたものを要約であるとする。

図1, 図2, より、同様のイベントストリームにおける1分毎の tweet 数の推移と、1秒毎の tweet 数の推移を特定の時間幅に分割する。この時、tweet 数の上昇する箇所を比較するバースト検出では、イベントストリームの分割する時間によって、要約生成の際に検出すべきバースト箇所が含まれない問題も起こる。そこで、本研究では経験的な値を用いて、投稿される tweet の間隔に着目することで分割する時間に影響されにくいバースト検出手法を利用した要約システムを提案する。

2. 関連研究

久保ら[3]はイベント内で要約を生成する際の提案手法をバースト検出とバースト内の Tweet 集合からの要約生成の2つに分け、速報性のある要約を提示するシステムを開発した。彼らは、バーストを検出する際に10秒あたりの tweet を時間あたりの tweet 数とし、その tweet 数が非バースト状態の中央値の2倍になっているときをバーストであるとしている。しかし、分割したセグメント内にバーストが含まれているかどうかは、セグメントに分割した際の時間に影響されてしまう可能性が存在する。

坂本ら[5]は時間によってトピックの移り変わりを捉える目的で要約を生成し、Twitter 上からキーワードを抽出する際に、バーストの断続性に着目したリアルタイムバースト検出を行っている。

土屋ら[7]は、マイクロブログから鉄道運行トラブルの発生時間および復旧時間の検出に本稿で用いる蝦名らのリアルタイムバースト検出を用いている。

本研究では、これらの従来手法をふまえて、新たに効果的な方法の提案を行う。

3. 提案手法

3.1. リアルタイムバースト検出

図3に要約を行うためのシステム構成を示す. 本研究では, 蝦名[2]の提案したリアルタイムバースト検出手法を用いる. 蝦名らの手法は, Zhang ら[8]の手法を改良したものであり, 一定時間毎に tweet 数の上昇を観測し, バーストを解析するのではなく, tweet 毎にバーストの解析を行うことができ, 短時間に大量の tweet が行われた際に, 高速に処理を行うことができるため, リアルタイムバースト検出に適したアルゴリズムである. 蝦名らの提案したリアルタイムバースト検出手法では, Aggregation Pyramid と呼ばれるセルをデータとしたピラミッド構造を用いる.

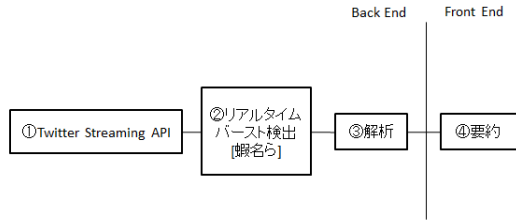


図3 システムの構成

図3で示した要約システムの構成では, Twitter 上から取得したデータに対してバースト検出を行い, その後バーストと判定された期間に対して解析を加えることで要約を生成し, 結果を出力する.

図4に Aggregation Pyramid を示す. セルのデータ構造のレベル0はN個のセルを持つ. 上層のセルは下層のセルの情報を統合したデータを持つ. 合計到着間隔 (gaps), 到着時間 (arrt), 間隔個数 (gapn) をセルの各データとする. 各セルを同条件で比較するために1つのセル内の到着間隔の1つあたりの平均値を求める平均到着関数を以下のように定義し, $avg(c(h, t))$ と $avg(c(N-1, t-1-h))$ を比較する.

$$avg(c(h, t)) = \frac{c(h, t).gaps}{c(h, t).gapn}$$

バーストを判定するパラメータ $\beta (0 < \beta < 1)$ を用い

て以下の条件を満たすとき, バーストが発生していると定義する.

$$avg(c(h, t)) \leq \beta \times avg(c(N-1, t-1-h))$$

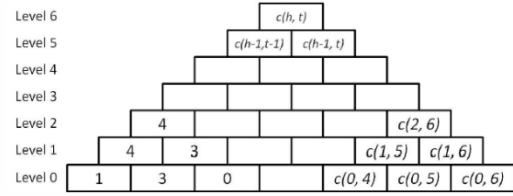


図4 Aggregation Pyramid

ピラミッド構造を決定するN, バースト判定係数 β , セルの最小ウィンドウサイズ W_{min} , バースト判定を行う最低記事数 A_{min} は, 経験的に $N = 60, \beta = 0.80, W_{min} = 3000$ (ミリ秒), A_{min} はバースト判定時に(判定を行うセルのレベル) $\times 5$ とした.

3.2. バースト内 tweet からの要約生成

3.2.1. Tf-Idf 要約手法

Sharifi の提案した Tf-Idf 要約手法について述べる. まず Tf-Idf について述べる.

tf (term frequency) とは, ある単語が文書 d 中に現れる回数を, 文書 d 中の全ての重複を許す単語数で割った値を意味する.

$$tf(t, d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}}$$

idf (inverse document frequency) とは, ある単語の文書頻度である df (document frequency) の逆数である. ある単語が出現する文書数である df の逆数ではあるが, 実際に使用する際には対数を用いる.

$$idf(t) = \log \frac{N}{df(t)} + 1$$

N は全文書数とする. 対数を取り, 文書 d で重複して出現する単語の影響を小さくする. 最後に, tf と idf の積を計算することによって文書中においてどれだけ特徴的な単語であるかということの1つの指標とする.

Sharifi の Tf-Idf 要約手法では, バースト検出期間を tweet 集合とする文書として Tf-Idf を用いる.

それぞれのバースト内での頻度の高い単語を計算する．次に、それぞれのバーストで固有の単語を特定するために、複数バーストに横断的に出現する単語を重要であるとみなさないために、idf を計算する．最後に、tf と idf の積を計算し、それぞれのバースト中から Tf-Idf の値の合計数をスコアリングする文章の長さに影響されないように、総単語数で割ったものが最も高いものを要約として出力する．

3.2.2. 良い実況者にに基づく要約手法

本研究では、バーストとして検出された tweet に対して、冷静な tweet をするユーザを”良い実況者”であるとし、解析を行った久保ら[2]の手法を用いる．久保らの手法は誰が何をしたのかを具体的に描写している”良い実況 tweet”を選択するために、バーストとして検出された各セグメント中の tweet tw に対して、以下の式の tweet スコア S_{tw} のように定める．

$$S_{tw}(g_n, tw) = Sim_{cos}(g_n, tw) \cdot num_{co}(tw)$$
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になる． $num_{co}(tw)$ はイベント(試合)に出場している選手名とイベント固有の専門用語のうち、tw において少ない方の出現回数である． 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．

$$S_{U_n}(u) = \frac{1}{C_n(u)} \sum_{k=1}^n S_{tw}(g_k, tw_{uk})$$

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 を行っているサブイベントの数である．また、 tw_{uk} 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} の最も大きい tweet である．

最後に 2 つの式を合わせて考え、n 番目のサブイ

ベントの要約として、最終式のレポートスコア S_{Rn} の値は最も大きい tweet を選択して出力する．

$$S_{Rn}(u, g_n, tw) = \lambda \cdot S_{U_{n-1}}(u) + (1 - \lambda) \cdot S_{Tw}(g_n, tw)$$

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. 評価実験

提案手法の有効性の検証のために評価実験を行う

4.1. 実験データ

本研究では、Twitter Streaming API を用いて #soccer, #daihyo, #jfa などのハッシュタグを指定して収集した 2014 年 6 月 12 日から 2014 年 7 月 13 日にかけて行われたブラジル開催 FIFA ワールドカップ決勝トーナメント 16 試合及び日本代表戦 3 試合から tweet のデータセットを収集した．本研究では試合時間内の 169687 件のデータを対象に解析を行った．

尚、データは 2014 年 6 月 15 日、日本 vs コートジボワール戦:tweet 数 4342 件、ユニークユーザ数:2564、2014 年 6 月 20 日、日本 vs ギリシャ戦:tweet 数 97185 件、ユニークユーザ数:19700、2014 年 6 月 25 日、日本 vs コロンビア戦:tweet 数 68160 件、ユニークユーザ数:9190 となった．

4.2. 評価指標

文書要約の評価手法には ROUGE[4]を用いる．ROUGE-N はシステムにおいて要約文書を生成した際に、正解要約と自動生成した要約との間の N グラムの一致する割合を計算するものである．

$$ROUGE(C, R) = \frac{\sum_{e \in n\text{-gram}(C)} Count_{clip}(e)}{\sum_{e \in n\text{-gram}(R)} Count(e)}$$

$n\text{-gram}(C)$ は自動生成した要約文書の N グラム、 $n\text{-gram}(R)$ は正解要約文書の N グラムを示す．

$Count(e)$ は、N グラムの出現頻度を数える関数であり、 $Count_{clip}(e)$ は自動生成した要約における出現頻度 $Count(e \in n\text{-gram}(C))$ と正解要約における出現頻度 $Count(e \in n\text{-gram}(R))$ の小さいほうを採用する．

Lin ら[4]は N を 1 から 4 まで変化させ、検証を行った．その結果 N が 1, 2 で最も高い相関があるこ

とを示した。久保ら[3]の評価指標同様に Twitter から速報を生成した時間と tweet が紐づいていることから高村らの提案する修正版 ROUGE を使用し、その際 N を 1 とした修正版 ROUGE-1 を用い、時間差を 3 分として名詞、形容詞、動詞のみを用いた。

5. 結果

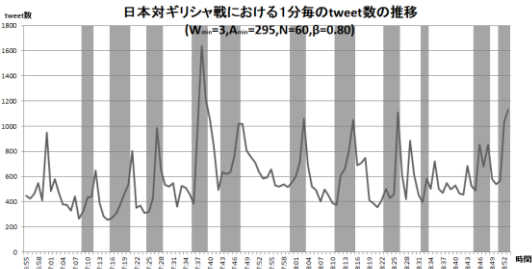


図 5 リアルタイムバースト検出

図 5 に日本 vs ギリシャ戦のリアルタイムバースト検出の結果を示す。

バースト判定に使用する各パラメータは、経験的な値(坂本ら[5], 土屋ら[7])を採用し、ピラミッド構造を決定する N, バースト判定係数 β , セルの最小ウィンドウサイズ $W_{\min}$, バースト判定を行う最低記事数 $A_{\min}$ は、経験的に $N = 60$, $\beta = 0.80$, $W_{\min} = 3000$ (ミリ秒), $A_{\min}$ はバースト判定時に(判定を行うセルのレベル) $\times 5$ とした。

表 1 バースト発生期間と tweet の件数

バースト	バースト発生期間	Tweet 件数
1	AM7:08~AM7:11	773 件
2	AM7:16~AM7:20	1415 件
3	AM7:24~AM7:28	2391 件
4	AM7:36~AM7:39	3032 件
5	AM7:43~AM7:48	3677 件
6	AM7:59~AM8:04	2459 件
7	AM8:10~AM8:15	1880 件
8	AM8:22~AM8:26	1408 件
9	AM8:31~AM8:33	405 件
10	AM8:44~AM8:48	2039 件
11	AM8:50~AM8:53	1639 件

表 1 にそれぞれの図 5 に示した検出されたバーストの左からの順序, バースト内の tweet 集合の件数とバースト発生期間をまとめた。

表 2 Tf-Idf 要約手法による要約文書

バースト	テキスト
1	コートジボワール戦より前向けて、縦パスも入りますね。いいんじゃないの！？ 前試合より明らかに前を向けてる感触あるよ！
2	大迫オフサイド
3	ギリシャが前からガンガンは来ないので、CB の位置でかなり組み立てられている。そこから縦パスで大迫が落とし、本田が前へ運ぼうとしたところでファウル。
4	カツラニスがイエロー 2 枚目で退場！！これでギリシャが 10 人になります。日本も吉田選手、長谷部選手がイエロー 1 枚ずつなので気をつけないと！
5	前半 3 5 分にギリシャのミトログル選手がゲカス選手と交代。アクシデントの模様です。そして 3 7 分、カツラニス選手に 2 枚目のイエロー、退場です！
6	スタメンから外れてしまった香川の元にピッチに入ると真っ先に向かった本田。“大丈夫だ、絶対勝とう” そんな思いがこもってる 2 人のハグ。
7	後半から：日本代表長谷部誠 out→遠藤保仁 in
8	運命の残り 45 分!!GO!!!!
9	後半 2 5 分過ぎ、ゴール前の混戦で内田が詰めたが外れる。これまた惜しい！ワールドカップギャグの中でも最

	流行のギャグいいすファール取れないのかよ・・・
10	岡崎慎司を倒したトロシディスにイエローカードそしてゴール前でのフリーキック GRE 決めろ 0(*・ω・*)0ゝまたフリーキック www これ決めないと無理ぽ。頼む～
11	アディショナルタイムは4分

表 2 は、図 5 に示した検出されたバーストの左からの順序と Tf-Idf 要約手法によって出力した要約文書の対応を示したものである。

表 3 ROUGE-1 による Tf-Idf 要約の評価

テキスト	ROUGE-1
コートジボワール戦より前向けて、縦パスも入りますね。いいんじゃないの！？ 前試合より明らかに前を向けてる感触あるよ！	0
大迫オフサイド	1.00
ギリシャが前からガンガンは来ないので、CB の位置でかなり組み立てられてる。そこから縦パスで大迫が落とし、本田が前へ運ぼうとしたところでファウル。	0.0667
カツラニスがイエロー2 枚目で退場！！ これでギリシャが 10 人になります。日本も吉田選手、長谷部選手がイエロー1 枚ずつなので気をつけないと！	0.429
前半 3 5 分にギリシャのミトログル選手がゲカス選手と交代。アクシデントの模様です。そして 3 7 分、カツラニス選手に 2 枚目のイエロー、退場です！	0.0769
スタメンから外れてしまった香川の元にピッチに入ると真っ先に向かっ	0

た本田。“大丈夫だ、絶対勝とう” そんな思いがこもってる 2 人のハグ。	
後半から：日本代表長谷部誠 out→遠藤保仁 in	0.889
運命の残り 45 分!!GO!!!!	0
後半 2 5 分過ぎ、ゴール前の混戦で内田が詰めたが外れる。これまた惜しい！ワールドカップギャグの中でも最流行のギャグいいすファール取れないのかよ・・・	0.0769
岡崎慎司を倒したトロシディスにイエローカードそしてゴール前でのフリーキック GRE 決めろ 0(*・ω・*)0ゝまたフリーキック www これ決めないと無理ぽ。頼む～	0.500
アディショナルタイムは4分	1.00

表 3 は ROUGE-1 によって、Tf-Idf 要約手法を利用したシステムが生成した要約を評価した結果を示している。

表 4 提案手法による要約文書

バースト	テキスト
1	長谷部にイエロー
2	大迫勇也のシュートは GK がキャッチ！
3	ギリシャのカツラニス選手にイエローカード
4	ギリシャ代表 カツラニスが 2 枚目のイエローカードで退場！！！！
5	川島永嗣ナイスセーブ！！！！
6	長谷部誠に代わって遠藤保仁ですかね。キャプテンマークを遠藤保仁が巻いてますね
7	サマラスにイエロー
8	香川が内田のオーバーラップに合わせてピタリとパスを通し、折り返し

	に大久保が詰めるもゴールならず (涙)…。その調子で攻めろ～～～
9	大久保ファウルもらうのうまい！
10	遠藤保仁の直接フリーキックはGKが セーブ！！！！
11	吉田どんだけファール取られるんだ よ！！

表4は、図5に示した検出されたバーストの左からの順序と提案システムが出力した要約文書の対応を示したものである。

表5 ROUGE-1 による提案手法の評価

テキスト	ROUGE-1
長谷部にイエロー	1.00
大迫勇也のシュートは GK がキャッチ！	0.400
ギリシャのカツラニス選手にイエローカード	0.750
ギリシャ代表 カツラニスが2枚目のイエローカードで退場！！！！	0.833
川島永嗣ナイスセーブ！！！！	0.500
長谷部誠に代わって遠藤保仁ですかね。キャプテンマークを遠藤保仁が巻いてますね	0.300
サマラスにイエロー	1.00
香川が内田のオーバーラップに合わせてピタリとパスを通し、折り返しに大久保が詰めるもゴールならず(涙)…。その調子で攻めろ～～～	0.461
大久保ファウルもらうのうまい！	0.250
遠藤保仁の直接フリーキックは GK がセーブ！！！！	0.142
吉田どんだけファール取られるんだよ！！	0.667

表5は ROUGE-1 によって、提案手法を利用したシステムが生成した要約を評価した結果を示して

いる。

6. 考察

本研究で用いた手法では、第一にバーストの判定を効果的に行うことに特徴がある。ドキュメントの到着に対して、不要な更新回数を抑制する目的でセルに圧縮しているため、イベントが発生していない無駄な計算を削除し、リアルタイムに解析することが可能である。第二に Tf-Idf 法を導入して、バースト内の個々の Tweet 集合を単位として特徴語を抽出した。これによってバースト間の比較を行うことで個々のバースト間の特徴を際立たせた。そして、バースト内の個々の Tweet 集合を単位とした特徴語によって、バースト内の個々の tweet に対して特徴的な単語を含む tweet を選択することで要約を生成した。第三に良い実況者に基づいた要約手法を用いることで要約を生成した。第四に Tf-Idf 手法で生成した要約と良い実況者に基づいた要約手法で生成した要約をそれぞれ ROUGE-1 によって評価した。

6.1. バースト検出に関する考察

バーストの判定条件 $avg(c(h, t)) \leq \beta \times avg(c(N-1, t-1-h))$ と図5の日本 vs ギリシャ戦のバースト検出の結果から投稿間隔が急激に狭まっている期間の左半分を正確に検出することができていることが確認できた。

その一方で、図5の日本 vs ギリシャ戦において試合開始前後の6時58分から7時2分にかけて投稿数が増加しており、前半戦開始時間と一致していることから、関係性があると考えられるが、この期間を本研究で利用した手法の経験的なパラメータ (A_{min} , N , W_{min} , β) として採用した値では検出できなかった。その1つの原因として、Aggregation Pyramid のピラミッド構造を決定するための値 N がデータの観測する期間となるが、この値を $N=60$ としたことで、ピラミッド内に前半開始時間に起因する投稿期間以外のデータが格納さ

れてしまったことが考えられる。また、同様の理由で図5において 左からバースト検出期間の8番目と9番目の間に投稿数が増加している期間をバーストとして検出していると考えられる。このように、連続して出現するような投稿数の急増を本研究の手法で捉えることができるかということに対する考察は私の知る限り、研究が行われていないため考慮した実験が今後必要になると考える。

6.2. 要約生成に関する考察

6.2.1. Tf-Idf 要約手法に関する考察

表3からバースト固有の特徴語のスコアが高いものを含むが、サッカーの試合の状況に対する実況に対してではなく、「選手に対するコメントや試合に対する感想、選手に対する状態など」を示すものも特徴語としてしまったため、評価のスコアが低くなったと考えられる。

6.2.2. 良い実況者に基づく要約手法

表5から特にROUGE-1のスコアが0.250, 0.1242と低くなったものに着目すると、出力要約が「大久保ファウルもらうのうまい!」, 「遠藤保仁の直接フリーキックはGKがセーブ!!!!」, 正解要約で「大久保が倒され、右サイドの敵陣中央からのFKを獲得」, 「キッカーの遠藤が右足でゴール右を狙うも、シュートはカルネジスに防がれてしまう」のように同一の事象に対する実況ではあるが別の表現がなされているものとなった。そのため、本研究の提案手法ではイベントストリーム中のイベントに対する状況を表現した実況要約tweetを出力可能である。出場選手とイベント固有の専門用語を解析の事前データとして必要とするが、選手の愛称などを考慮することで更なる精度向上が期待できる。

6.3. システム全体に関する考察

Twitterと呼ばれるマイクロブログ上の特定の話題に関する内容の突発的なバーストをリアルタイムバースト検出手法によって解析した。しかし、

適応限界であるバースト規模の見積もりのためには、実験データの容量と提案手法によって実際にシステムで使用されるメモリの関係を考慮する必要がある。また、リアルタイムバースト検出では経験的な4つのパラメータ(A_{min} , N , W_{min} , β)が存在している。この経験的な値の調査方法として、バースト検出精度の評価指標として再現率(Recall)R, 適合率(Precision)P, F値(F-measure)Fのそれぞれ3つの意味(網羅性, 正確性, 調和平均)のグラフの可視化を行うことが1つの経験的な値の決定の指標にもなると考えられ、今後の課題であると言える。

参考文献

- [1]<https://twitter.com>
- [2] 蝦名亮平, 中村健二, 小柳滋:リアルタイムバースト検出手法の提案, 日本データベース学会論文誌, Vol. 9, No. 2, November 2010.
- [3] 久保光証, 笹野遼平, 高村大地, 奥村学:”良い実況者”に着目したTwitterからのスポーツ速報生成, 言語処理学会, 第19回年次大会, 発表論文集.
- [4] Lin, C.-Y.:ROUGE:A Package for Automatic Evaluation of Summaries.Proc, The ACL-04 Workshop “Text Summarization Branches Out”, pp.74-81(2004).
- [5] 坂本翼, 廣田雅春, 横山昌平, 福田直樹, 石川博:Twitterストリームのバーストの断続性に着目したキーワード抽出.
- [6] 水沼友宏, 池内淳, 山本修平, 山口裕太郎, 佐藤哲司, 島田諭: Twitterにおけるバーストの生起要因と類型化に関する分析, 情報社会学会誌, Vol. 7, No. 2, 原著論文.
- [7] 土屋圭, 豊田正史, 喜連川優. マイクロブログを用いた鉄道の運行トラブル発生期間および付帯情報の抽出, DEIM Forum 2014 B3-2
- [8] Xin Zhang and Dennis Shasha. Better Burst Detection. ICDE, pages 146-149 2006.