

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

AKAKI, Fumihiko / 赤木, 文彦

(開始ページ / Start Page)

1

(終了ページ / End Page)

54

(発行年 / Year)

2016-03-24

(学位授与年月日 / Date of Granted)

2016-03-24

(学位名 / Degree Name)

修士(工学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

2015 年度 修士論文

論文題名：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

指導教員：藤井章博

大学院理工学研究科

応用情報工学専攻修士課程

学籍番号：14R4101

フリガナ：アカキフミヒコ

氏名：赤木文彦

目次

1. 序章.....	1
1.1. 概要.....	1
1.2. 目的.....	2
1.3. バーストの定義.....	3
1.4. 要約の定義.....	4
1.5. 論文の構成.....	4
2. 関連研究.....	5
2.1. Aggregation Pyramid.....	5
2.2. Tf-Idf 要約手法.....	8
2.3. 良い実況者に基づく要約手法.....	10
3. 要素技術.....	11
3.1. OAuth 認証.....	11
3.2. Twitter Streaming API.....	12
3.3. テキストマイニング.....	12
3.3.1. 形態素解析.....	13
3.3.2. 構文解析.....	13
3.3.3. 意味解析.....	14
3.3.4. 文脈解析.....	14
4. 実験データ.....	15
5. 提案手法.....	16
6. システム構成.....	19
6.1. バースト検出部.....	19
6.1.1. Twitter 上からのデータ取得.....	19
6.1.2. リアルタイムバースト検出.....	19
6.2. 要約生成部.....	20
6.2.1. 形態素解析器.....	20
6.2.2. 頻度集積.....	20
6.2.3. 要約生成.....	20
7. 評価指標.....	21
7.1. ROUGE.....	21
8. 結果.....	22
9. 考察.....	30

9.1. バースト検出に関する考察	30
9.2. 要約生成に関する考察	31
9.2.1. Tf-Idf 要約手法に関する考察	31
9.2.2. 良い実況者に基づく要約手法	31
10. 結論	31
参考文献	32
謝辞	32
付録	33

1. 序章

1.1. 概要

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することで情報の配信が可能になる. 本研究では, 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, ユーザが関心のある `tweet` を閲覧し, イベントストリーム内の全て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, 閲覧する場合にも問題が起こる. 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によってこの問題に対処する. 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[2]. 本研究では, イベントストリームに関してバースト検出とバースト検出内の解析の 2 つに分けることで, 時々刻々と変化するある特定の出来事に対する投稿から要約生成を行うことで実況状況を分かりやすく可視化し, その評価を行った.

1.2. 目的

本研究ではリアルタイム実況要約の可能性を検討するために過去のイベントについての Twitter 情報を蓄積し、そのデータを対象として分析を行う。すなわち、Twitter に対して投稿された FIFA ワールドカップの試合を要約することで、過去の tweet からユーザが試合の内容を理解することに対しての支援を目指す。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。

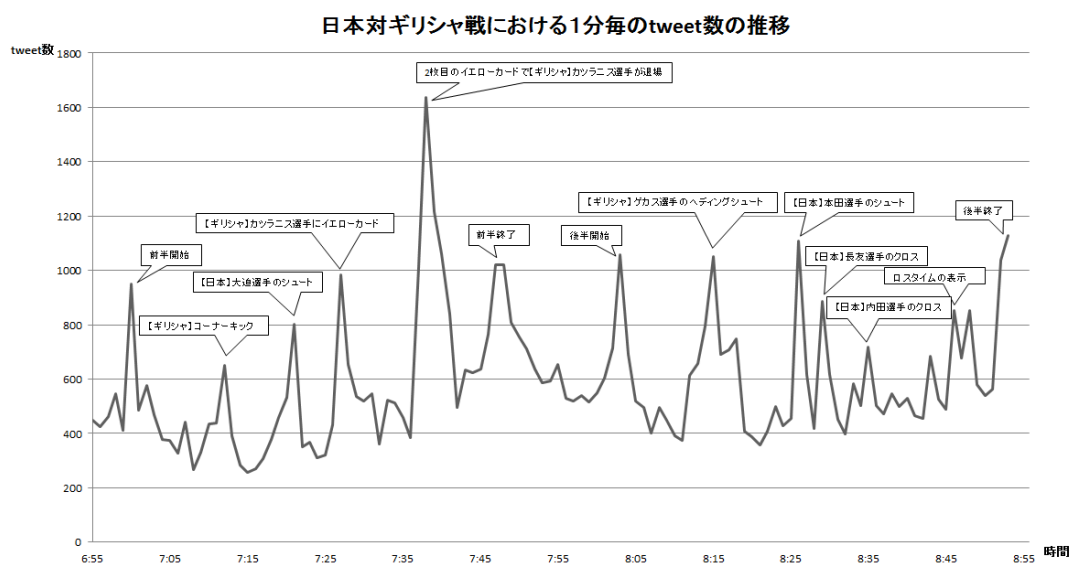


図 1 日本対ギリシャ戦における 1 分毎の tweet 数の推移

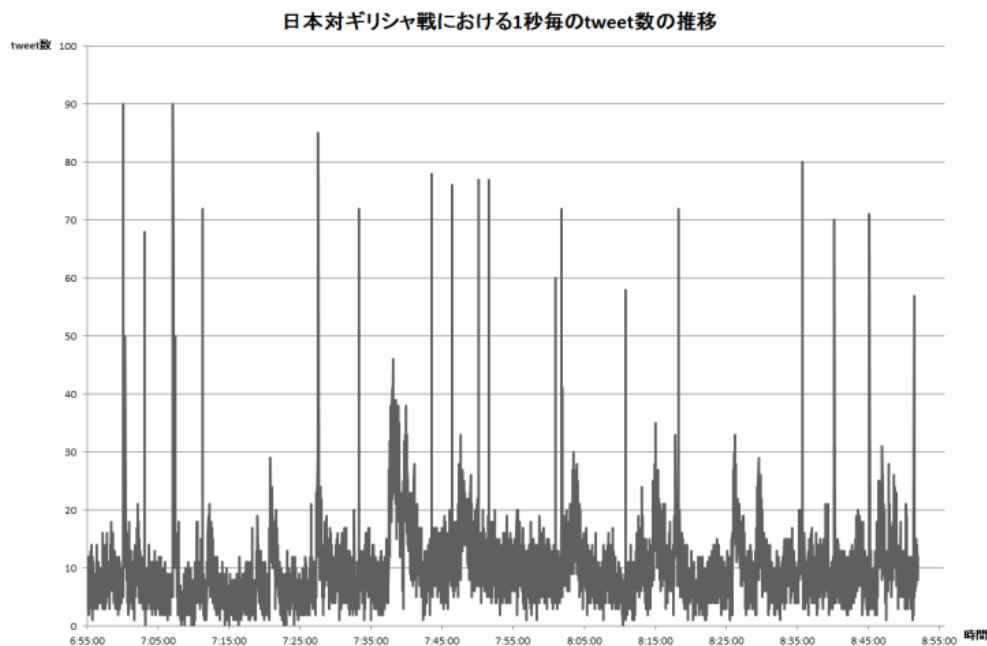


図2 日本対ギリシャ戦における1秒毎の tweet 数の推移

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るべきバースト箇所が含まれない問題も起こりうる。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。特に本研究では、2014年FIFAワールドカップの日本代表戦の試合に焦点をあてる。

1.3. バーストの定義

本研究ではバーストを、「ブログやマイクロブログに対する複数実況者の関心の高まりに起因する投稿数の急激な増加」とであると定義する。本研究では特に、投稿数ではなく、投稿される時間の投稿間隔に着目し、投稿時間の間隔が急激に狭まっている期間を発見することでバーストを検出する手法を用いる。

1.4. 要約の定義

本研究では要約を「バースト期間中の投稿内容に着目し，特に重要な状況描写が行われている投稿をバースト期間毎に 1 つずつ選択し，時系列順に並べたもの」とであると定義する．バースト内の投稿内容は，特定のイベントに対するそれぞれの実況者のリアルタイムな状況の描写であり，イベントを実際に観測していないユーザに対する重要な情報と成りうる．

1.5. 論文の構成

以下，第 2 章で関連研究について記述し，第 3 章では要素技術に関して，第 4 章で本研究において対象とする実験データに関して，第 5 章で提案手法について記述する．第 6 章でシステムの構成，第 7 章で提案手法の評価指標，第 8 章で提案システムによる結果，第 9 章で考察，第 10 章で結論と今後の課題について述べる．

2. 関連研究

2.1. Aggregation Pyramid

蝦名[3]らの提案したリアルタイムバースト検出手法である Aggregation Pyramid について述べる.

蝦名[3]らの提案する手法の利点は, バーストを単に解析するだけでなく, リアルタイムに解析するためにセルと呼ばれる概念を用いて, メモリを効果的な利用を考慮している点である.

坂本[4]らは時間によって話題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て用いている.

土屋[5]らは, 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用いている.

Aggregation Pyramid は, 投稿されるデータの時間間隔が急激に狭まっている期間を発見することで, リアルタイムバーストを検出手法である. Aggregation Pyramid はセルをデータとしたピラミッド構造である. セルのデータ構造のレベル 0 は N 個のセルを持つ. 最上位のセルは最下層のセルの情報を全て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 β ($0 < \beta < 1$) 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.

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

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.

各セルのデータの格納の流れ, Aggregation Pyramid について具体例を図 3, 図 4 に示す.

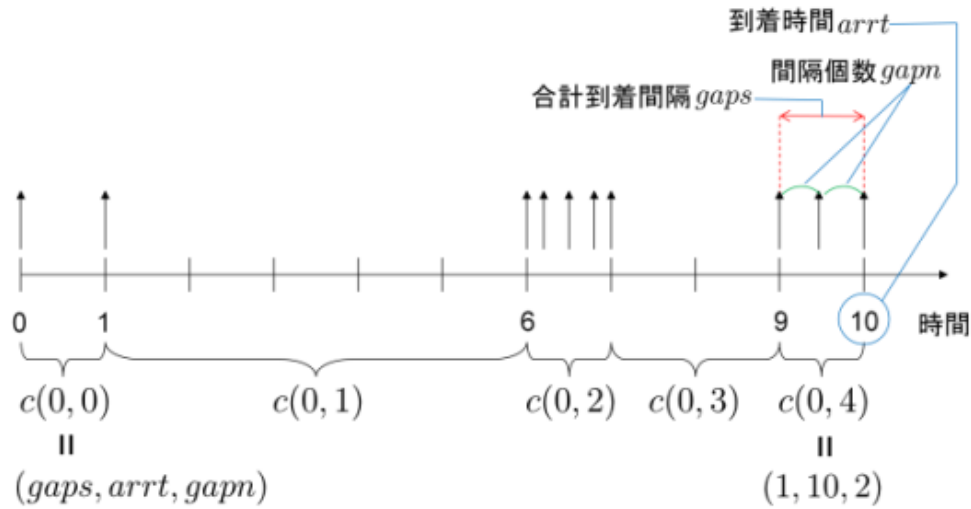


図3 各セルのデータ格納の流れ

図3は、1つのセルに格納されるデータを示している。

まず、1つのセルに格納される3つのパラメータはそれぞれ $gaps$, $arrt$, $gapn$ である。それぞれ合計到着間隔，到着時間，間隔個数を表している。合計到着間隔とは、1つのセル内に格納する全到着間隔の合計である。到着時間とは1つのセルに格納されるデータの最終的に到着した時間である。間隔個数とは、1つのセルに格納されるデータの到着間隔の数である。

ここでは、ドキュメントの更新回数を抑制するためのパラメータ $W_{min}=1$ として説明する。イベント発生時間列を $\{0,1,6,6,6,6,9,9,10\}$ ，発生間隔列 $\{1,5,0,0,0,0,3,0,1\}$ とする。

$C(h,t)$ を各セルのデータとし、 h はピラミッドのレベル、 t はセルの番号である。それぞれ更新回数を抑制する W_{min} と到着するドキュメントの到着間隔を比較する。 W_{min} を上回るような到着間隔の場合に W_{min} のサイズに圧縮する。したがって、レベル0に格納されるデータは $C(0,0) = (1,1,1)$ ， $C(0,1) = (5,6,1)$ ， $C(0,2) = (1,7,4)$ ， $C(0,3) = (2,9,1)$ ， $C(0,4) = (1,10,2)$ となる。

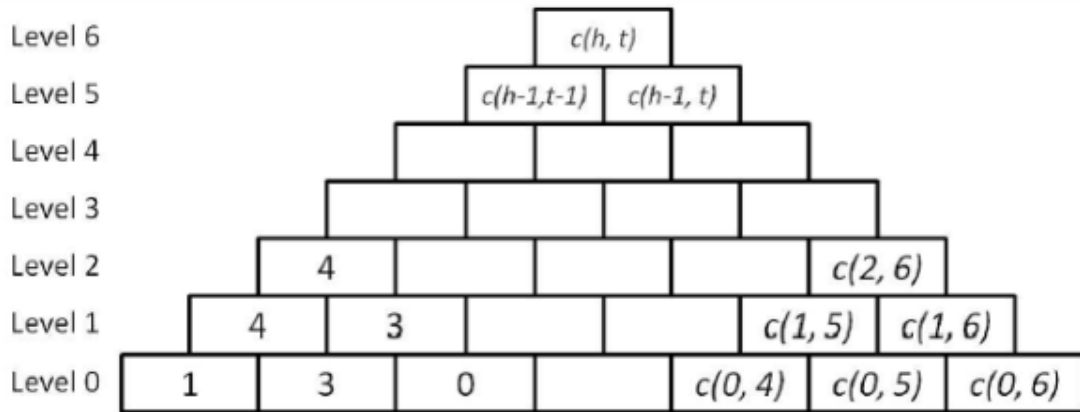


図 4 Aggregation Pyramid

図 4 は Aggregation Pyramid を示す。

リアルタイムバースト検出では、図 5 に示したように Level 0 の 1 つのセルのデータを決定する。Aggregation Pyramid(図 6 のように)では Level 0 として決定されたデータを基にして各セルに対して左から順に格納して計算する。Level 1 のセルは、Level 0 の隣接するセルの情報を利用して計算される。具体的には、 $C(1,0) = C(0,0) + C(0,1)$ 、 $C(1,1) = C(0,1) + C(0,2)$ 、 $C(1,2) = C(0,2) + C(0,3)$ 、 $C(1,3) = C(0,3) + C(0,4)$...のように計算する。

Level 2 は Level 1 の隣接するセルの情報を利用して保持する。具体的には、 $C(2,0) = C(1,0) + C(1,1)$ 、 $C(2,1) = C(1,1) + C(1,2)$ 、 $C(2,2) = C(1,2) + C(1,3)$ 、 $C(2,3) = C(1,3) + C(1,4)$...のように計算する。レベル 3 以降のセルも同様にして、求めたいセルの 1 つ下位の 2 つの隣接するセルの情報を利用して計算する。尚、計算の順序は逐次的に到着するデータを解析するために、具体的には $C(0,0)$ 、 $C(0,1)$ 、 $C(1,0)$ 、 $C(0,2)$ 、 $C(1,1)$ 、 $C(2,0)$ 、 $C(0,3)$ 、 $C(1,2)$ 、 $C(2,1)$ 、 $C(3,0)$...のように行う。

したがって、最大レベルを h とすると Level h のセルはピラミッドの大きさを決定する $N(\text{Level 0 のセルの数})$ のセルの情報を全て含む。実際にバーストの判定に使用するセルは最大の Level h のセル $C(h,t)$ である。セル数 N 毎に計算されるため、最大レベルのセルを計算終了後、ピラミッド内のセルに存在する各情報は、バースト判定に使用しない最大レベル以外のセルはメモリ効率の観点から消去される。

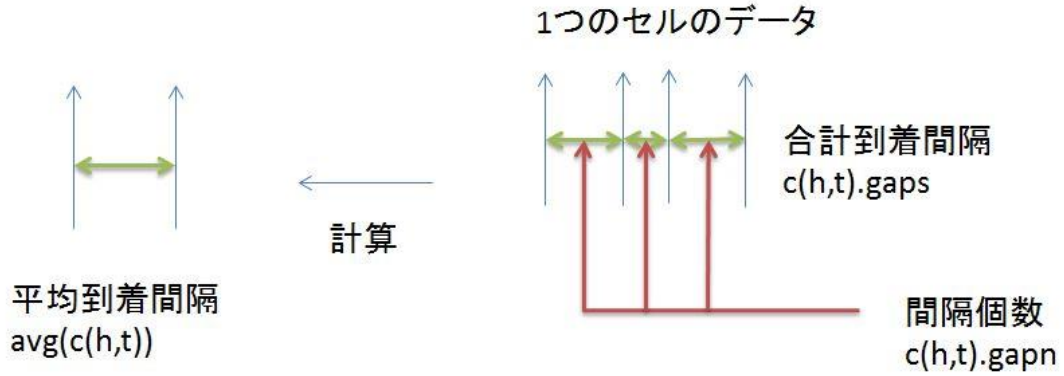


図5 平均到着間隔の計算イメージ

図5に平均到着間隔 $\text{avg}(c(h,t))$ の計算イメージを示す。

1つのセルのデータには、合計到着間隔(gaps)、到着時間(arrt)、間隔個数(gapn)の3つのデータが格納されている。この値で、合計到着間隔(gaps)、間隔個数(gapn)を利用して、各セルで平均して到着するデータの時間の間隔が存在するかということについて計算する。到着時間(arrt)は、バースト判定計算の際にいつバーストが発生しているのかということをも特定するために利用される。

最大の Level h のセルにおいて、この $\text{avg}(c(h,t))$ と次の $\text{avg}(c(N-1, t-1-h))$ を比較することで、バーストの判定を行う。 $\text{avg}(c(h,t)) \leq \beta \times \text{avg}(c(N-1, t-1-h))$ の条件を満たすとき、バーストが発生しているとみなす。 β ($0 < \beta < 1$) は任意の重みである。 $\text{avg}(c(h,t))$ と $\text{avg}(c(N-1, t-1-h))$ は最大の Level h のセルなので、投稿時間が直前よりも急激に短くなっている期間をバーストとして検出することが可能となる。

2.2. 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。

$\text{tf}(\text{term frequency})$ 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。

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

$\text{idf}(\text{inverse document frequency})$ とは、ある単語の文書頻度である $\text{df}(\text{document frequency})$ 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．

以下で Sharifi の提案した Tf-Idf 要約手法の流れを説明する．

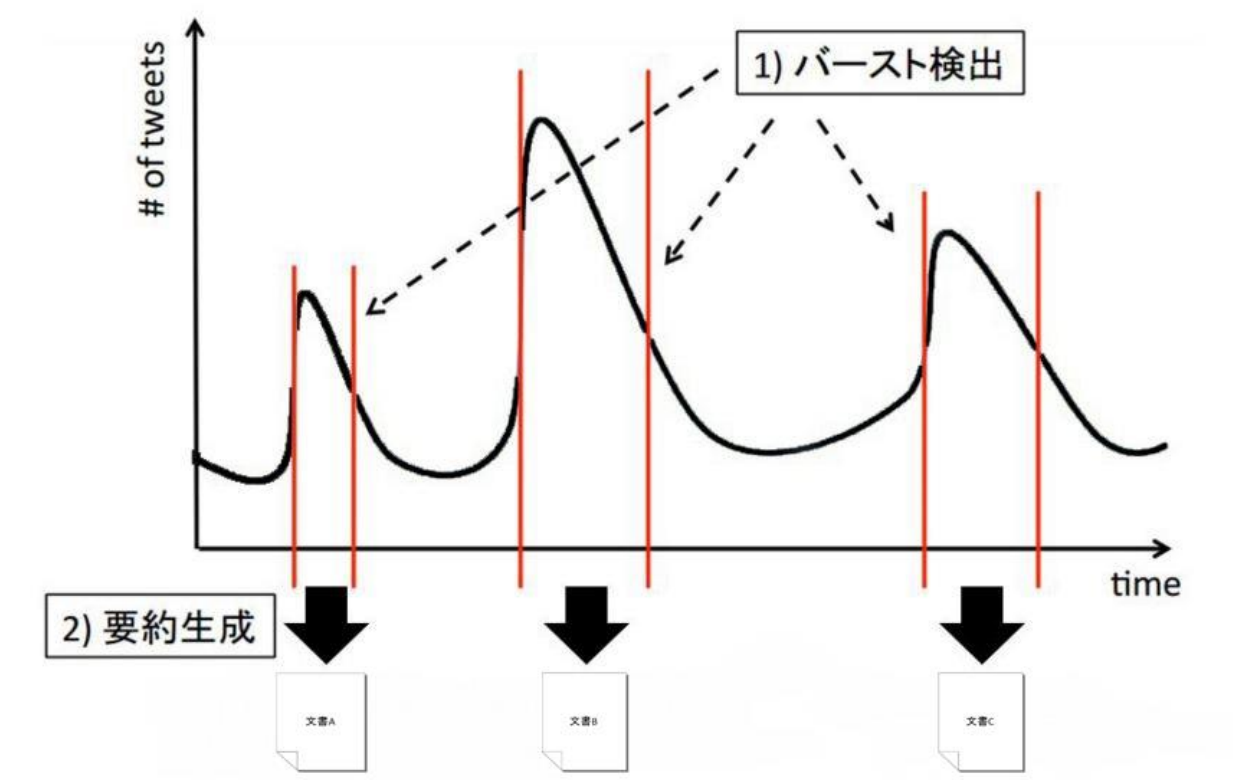


図 6 Sharifi の Tf-Idf 要約手法の文書イメージ

図 6 に Sharifi の提案する手法を示す．各文書は，特定の手法によってバースト検出された箇所を 1 文書とする．まず，tf をそれぞれ文書 A，文書 B，文書 C 内に対して用いることで，それぞれの文書内での頻度の高い単語を計算する．次に，それぞれの文書で固有の単語を特定するために，複数文書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，それぞれの文書中から 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．

2.3. 良い実況者に基づく要約手法

久保ら[6]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 つにわけ、速報性のある要約を提示するシステムを開発した。しかし、バーストを検出する際に 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。また、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。

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、久保[6]らの要約生成の手法を用いる。

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 (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。 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、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。

3. 要素技術

3.1. OAuth 認証

サーバーが Twitter に対して、データを取得するためには OAuth 認証という API 認可が必要となる。OAuth 認証は、サービス A の個人情報等の情報や獲得した情報を外部サービス B に受け渡すための認可の仕様である。尚、本研究ではサービス A が Twitter に相当し、サービス B が本システムに当たる。OAuth 認証の手順をシーケンス図で説明する。

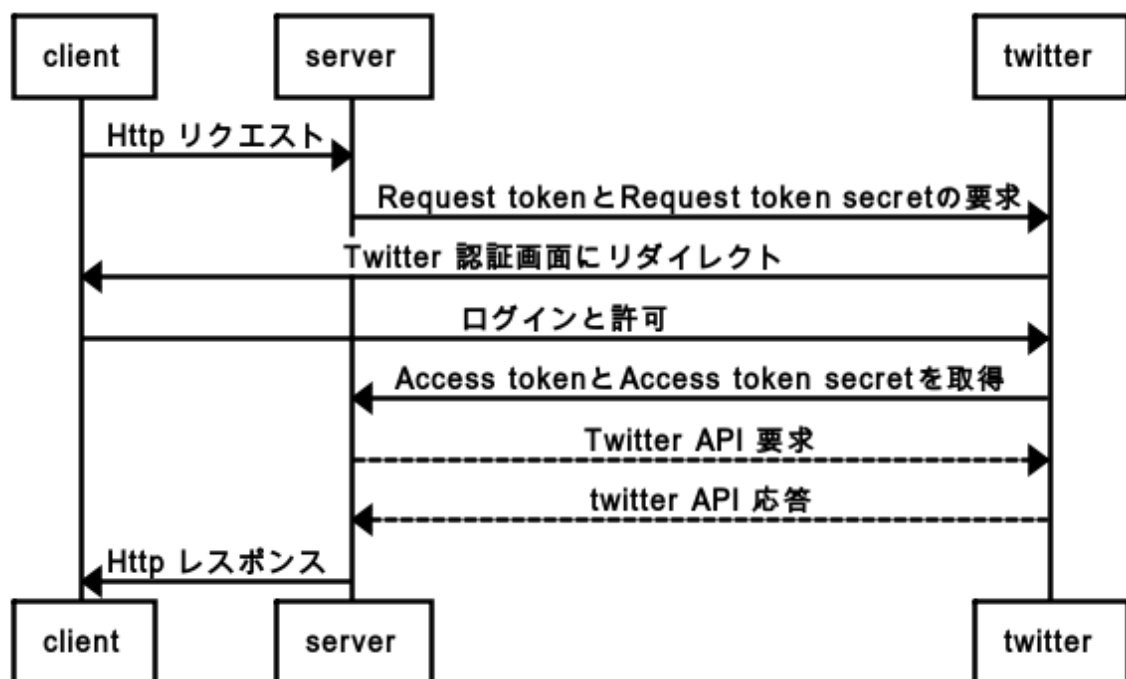


図 7 OAuth 認証のシーケンス図

OAuth 認証ではサーバーから twitter に対して, Request token と Request token secret を要求する。その後, twitter 認証が画面にリダイレクトされる。クライアントで twitter に対するログイン許可をする。ここで twitter を利用するユーザであるとユーザ名とパスワードを用いて判断された場合, Access token と Access token secret を取得することができる。ここで Request token, Request token secret, Access token, Access token secret の 4 つのキーを利用することで, Twitter 上の API 認可に必要な情報を獲得できる。これにより, Twitter 上の機能を Web API として利用できる。

3.2. Twitter Streaming API

Twitter[1]では、アプリケーションの開発者に対して Twitter 上のユーザ情報を利用することができるように API(Application Programming Interface)を提供している。Twitter Streaming API では、ロングポーリング技術を用いることで Twitter 上の特定の話題に関する内容を速報性のある形式で取得することができる。このときに、ハッシュタグと呼ばれる Twitter 上の複数投稿者が特定の話題に対して、投稿者以外が検索を容易にするための機能として利用される。このハッシュタグを利用することで、どの内容に関する話題であるかを特定し、取得することが可能となる。本研究では、ワールドカップ公式サイトで公開されている日本代表のハッシュタグや出場国のハッシュタグ、放送局で提示されたハッシュタグ等を用いた。Twitter Streaming API では、ハッシュタグを複数指定することができ、複数ハッシュタグの場合には、AND 検索でデータを取得するのではなく、OR 検索でデータを取得することができ、より大規模なデータの取得に利用することができる。本研究でも、複数のハッシュタグを指定することでデータを取得した。

3.3. テキストマイニング

通常の文章から成るデータを単語や文節で区切り、それらの出現頻度や共出現の相関、出現傾向、時系列などを解析することで有用な情報を取り出す、データ分析手法である。テキストマイニングで行う上で必要な自然言語処理での全体の流れを概略図 8 に示す。

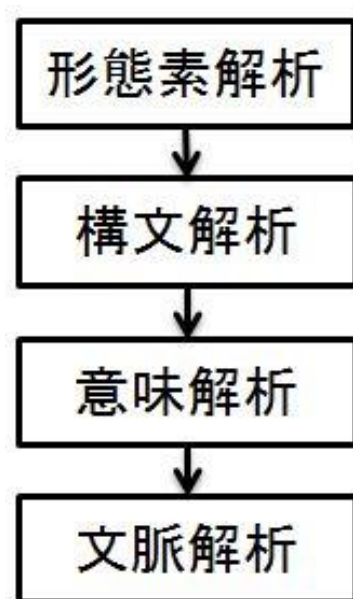


図 8 自然言語処理の概略

3.3.1. 形態素解析

文法的な情報の注記のない自然言語の文章から、対象の文法や、辞書と呼ばれる単語の品詞等の情報に基づき、形態素と呼ばれる言語で意味を持つ最小単位の列に分割し、それぞれの形態素の品詞等を判別する作業である。日本語として使用できるものには Mecab, Chasen, KAKASI, JUMAN 等が存在する。形態素解析処理の流れを図 9 に示す。まず、解析対象とする文字列を入力し、日本語分から書き辞書を参照、それぞれの形態素に分割し、解候補を作成、コストを計算し、計算の結果から分割する単位及び品詞の最適解を求めることで形態素解析を行う。

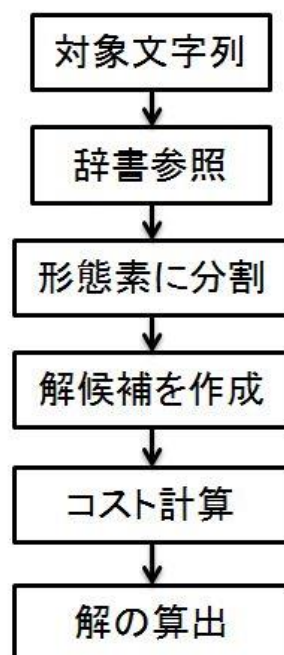


図 9 形態素解析処理の流れ

3.3.2. 構文解析

文章、マークアップ等の注記の入っていない文字列を、自然言語の場合形態素に切り分け、さらにその間の修飾や被修飾等の関係を構文論的な関係を図式化する等して明らかにする手続きである。自然言語の場合には、自然言語処理の要点の 1 つである。プログラミング言語等の形式言語の場合は、形式文法に従って構文木を決定する。文法規則によって、文章の構造を句や文節を単位として単語の係り受け解析の分析が可能となるが、本研究では利用しないものとする。

3.3.3. 意味解析

イベント固有の専門用語及び選手名を用いることで、自動的な要約文書を生成することが可能となる．本研究では，大規模なテキスト情報から要約文書を生成するために，久保[6]らの提案する手法を用いる．

3.3.4. 文脈解析

文脈解析は自然言語処理における一連の処理であるが，本研究では行わないものとする．

4. 実験データ

Twitter は投稿者が感じたことを **tweet** と呼ばれる 140 文字以内のテキストで投稿し、共有することができるマイクロブログと呼ばれるブログの一種である。Twitter に対して人々は毎日数百万の最新の動向を投稿する。これらのいくつかの投稿は、スポーツイベントや政策の議論、自然災害等のテレビ番組やラジオ放送のようなメディアを通して見ていたり、参加したりしている人々のイベントを含んでいる[2]。これらの投稿が、リアルタイムに実世界上の出来事に対する重要な状況を如実に描写している。そのため、本研究で対象とするイベントストリームのような移り変わりが頻繁で内容を容易に理解することができない問題を解決するための有意義な情報になる。

本研究の評価のために、2014 年 6 月 12 日から 7 月 13 日にかけてブラジルで開催された第 20 回 FIFA ワールドカップの決勝トーナメント 16 試合と日本代表戦 3 試合の **tweet** のデータセットを収集した。このデータセットは、Twitter Streaming API により収集したデータである。これらのうち、本研究では、実験対象を日本語による実況 **tweet** を多く含んだイベントストリームに絞りこむため、日本代表戦の 3 試合に減らす。

本研究では試合時間内の 169687 件のデータを対象にイベントストリームの解析を行った。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

5. 提案手法

本研究の提案手法では，バースト期間中の tweet を実況者の関心の高まりによって起こる重要なイベントであるとして，バーストが発生していない期間の tweet 集合は無視する．バースト期間中の tweet 集合に着目して解析することで，リアルタイムな状況要約を生成する．

提案手法は，大きく分けてバースト検出と要約生成の 2 つに分かれる．

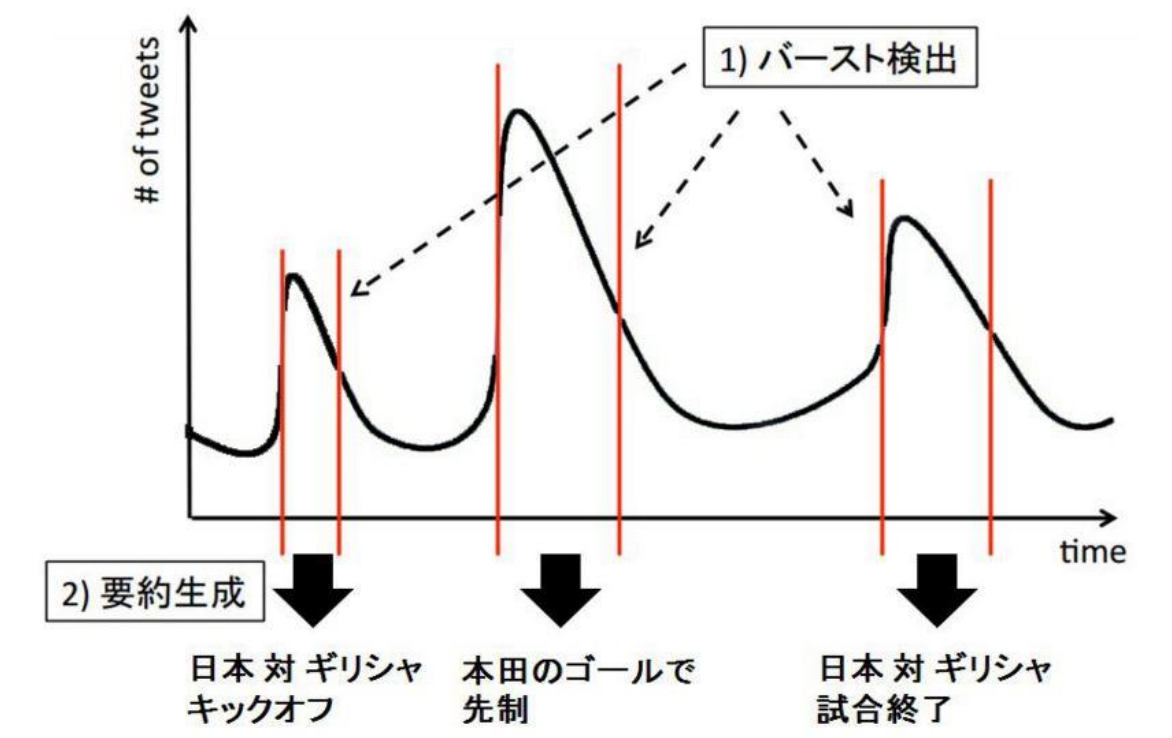


図 10 要約システムのイメージ

図 10 は本研究における要約システムのイメージである．

本研究では，特にバーストの発生に着目し，各バースト期間中の tweet 集合内の情報を話題の関心の高まりによる投稿数の増加とみなして要約生成を行う．

以下が 2 つのフェーズである．

- 1) バースト検出では，蝦名[3]らの提案したリアルタイムバースト検出手法である Aggregation Pyramid を用いる．
- 2) 要約生成は，久保[6]らの提案した良い実況者に基づく要約手法を用いる．

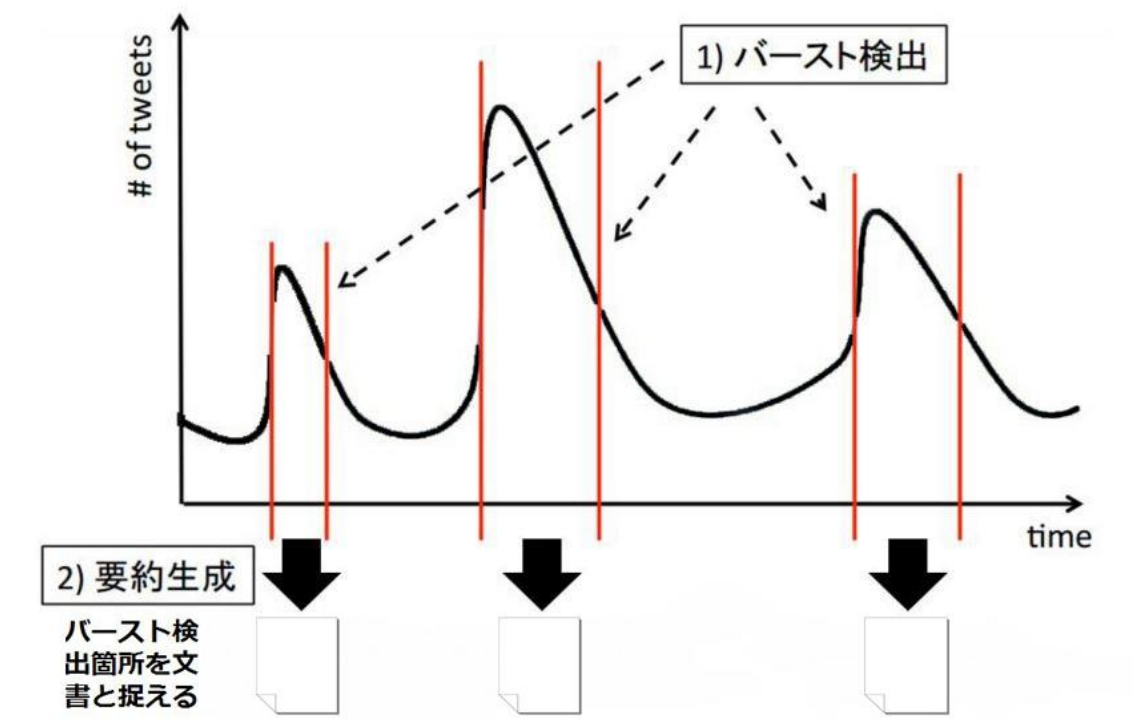


図 11 バースト検出からの要約生成の文書イメージ

本研究では，バースト検出手法によって検出された各期間をバーストであるとし，バースト期間中の tweet 集合を 1 文書とみなして要約生成を行う．

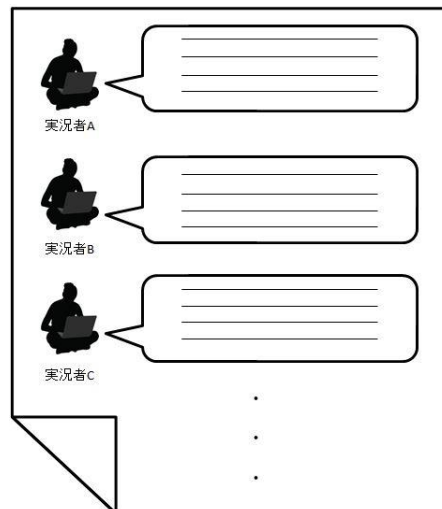


図 12 対象とする 1 文書

図 12 に本研究における 1 文書を示す．本研究では，バースト検出結果によって検出された期間を重要なイベントであると考え，重要なイベント中で複数の実況者の投稿を含んだ tweet 集合を 1 文書であるとする．重要なイベントでは，特定の話題に対する状況で関心が

高まっているような出来事(例えば、サッカーであればゴールシーンやキックオフ、ラグビーであれば、タッチダウンやファンブル等)を含んでいる。この 1 文書中で、久保[3]らの要約手法を用いることによってスコアリングを行う。このスコアが高いものを選択し、各文書中からバースト発生毎に選択する。このスコアリング結果が高いものを時系列順に並べたものを本研究では、要約であるとする。バースト検出からスコアリング処理による要約生成までの流れを図 13 に示す。まず、Twitter Streaming API で取得したイベントストリームを入力し、バースト検出、各文書の作成、各文書内の tweet に対してスコアリングを行い、各文書内で最高スコアの tweet を選択し、時系列順に並べ、要約として出力する。

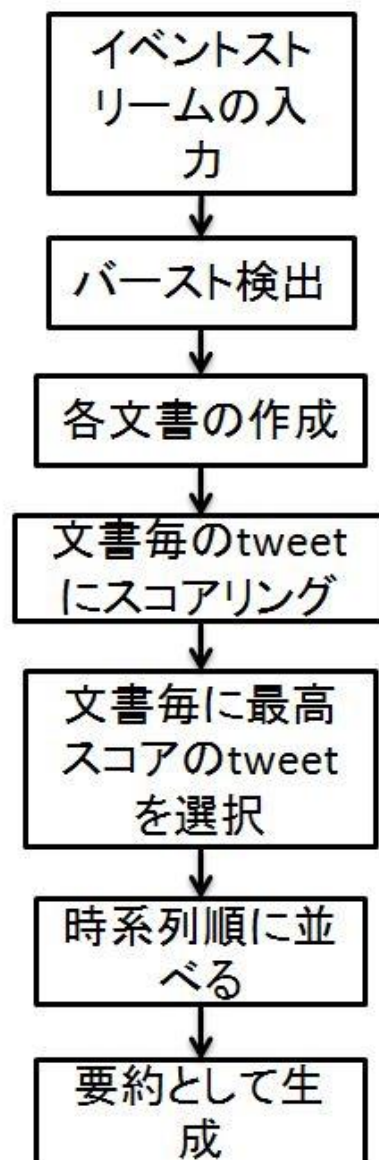


図 13 要約生成の流れ

6. システム構成

提案する要約システムを以下に示す．本研究のシステムは 2 つの部分から構成される．

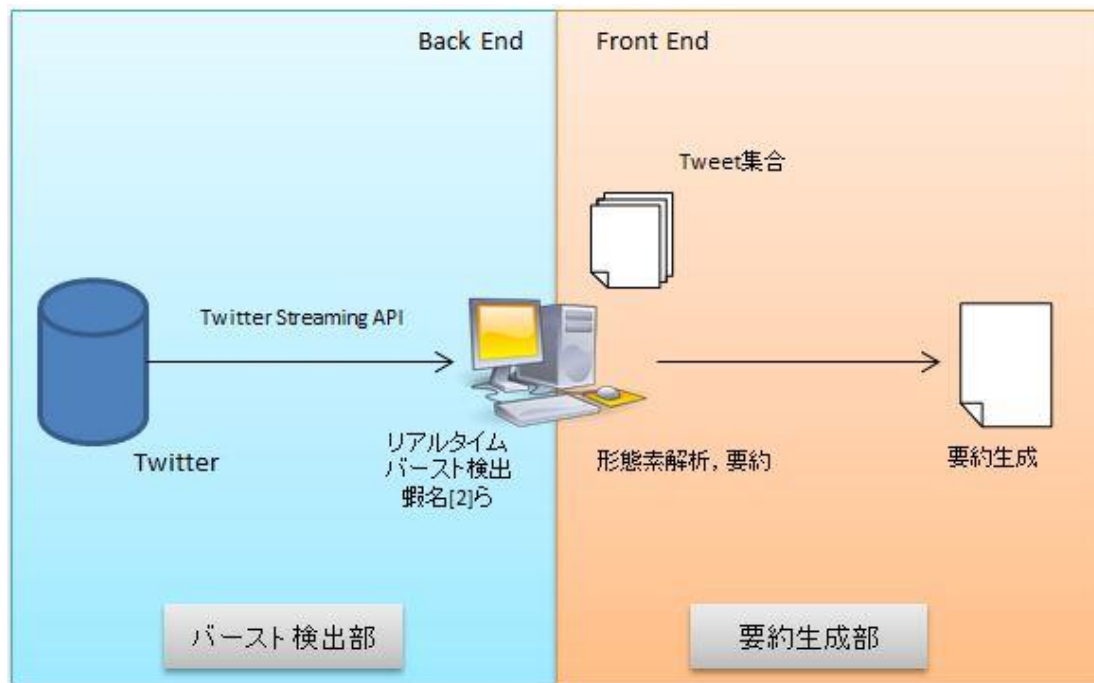


図 14 システム構成

6.1. バースト検出部

6.1.1. Twitter 上からのデータ取得

Twitter 上からデータを取得するために、Twitter が API(Application Programming Interface)として提供している Twitter Streaming API を使用することで、Twitter 上のデータを獲得する。尚、Twitter 上で実況者自身が特定のイベントに対するコメントとして付与した大文字、小文字の区別のないハッシュタグと呼ばれる検索ワードを指定することでイベントを特定した。Twitter API では、プログラミング言語 Java で簡易的に利用することができる Twitter4j を利用することで、ハッシュタグを含む tweet を取得する。

6.1.2. リアルタイムバースト検出

Aggregation Pyramid を利用することで、獲得した Twitter 上からのデータからイベントストリームに対してメモリの効率を考慮したリアルタイムなバースト検出が可能となる。

6.2. 要約生成部

形態素解析を用いることで、文章を単語毎に分割する．ここで、ストップワードと成りうる名詞・動詞・形容詞と比較して除外することで、利用すべき単語を選別する．その後、久保[6]らの手法を用いることで要約を出力する．

6.2.1. 形態素解析器

形態素解析器として、Java で利用することが可能となる Igo の辞書を利用し、MeCab 上で動作させた．解析結果に関しては、MeCab と互換である．

6.2.2. 頻度集積

形態素解析によって獲得した単語の出現回数を累積させることで出現頻度を分析する出現単語と出現回数のハッシュマップを作り、ストップワードを除去した全単語の頻度を累積させる．

6.2.3. 要約生成

良い実況者に基づいた要約手法[6]を用いることで、獲得した単語から「誰が何をしたのか」ということが分かるような良い実況者による実況を選択することができ、試合に出場する選手名及び専門用語をいずれも含むような tweet を客観的とみなすことで要約を生成することができる．ベースラインの手法として、Sharifi の提案する Tf-Idf 手法と比較することで要約の精度の向上などを観測する．その際、要約文書中の評価指標として時間差を 3 分とした修正版 ROUGE-1 を使用する．

7. 評価指標

7.1. ROUGE

文書要約評価手法には ROUGE を用いる．ROUGE-N はシステムにおいて要約文書を生じた際に，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．

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

$n\text{-gram}(\mathbf{C})$ は自動生成した要約文書の N グラム， $n\text{-gram}(\mathbf{R})$ は正解要約文書の N グラムを示す． $\text{Count}(e)$ は N グラムの出現頻度を数える関数であり， $\text{Count}_{\text{clip}}(e)$ は，自動生成した要約における出現頻度 $\text{Count}(e \in n\text{-gram}(\mathbf{C}))$ と正解要約における出現頻度 $\text{Count}(e \in n\text{-gram}(\mathbf{R}))$ の小さいほうを採用する．Lin[10]らは 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[6]らの評価指標同様に 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．正解要約には yahoo スポーツナビ[9]の提供する要約を使用した．

8. 結果

以下に Aggregation Pyramid による, 2014 年 6 月 20 日に行われた日本 vs ギリシャ戦のイベントストリームに対してバースト検出を行った結果を示す.

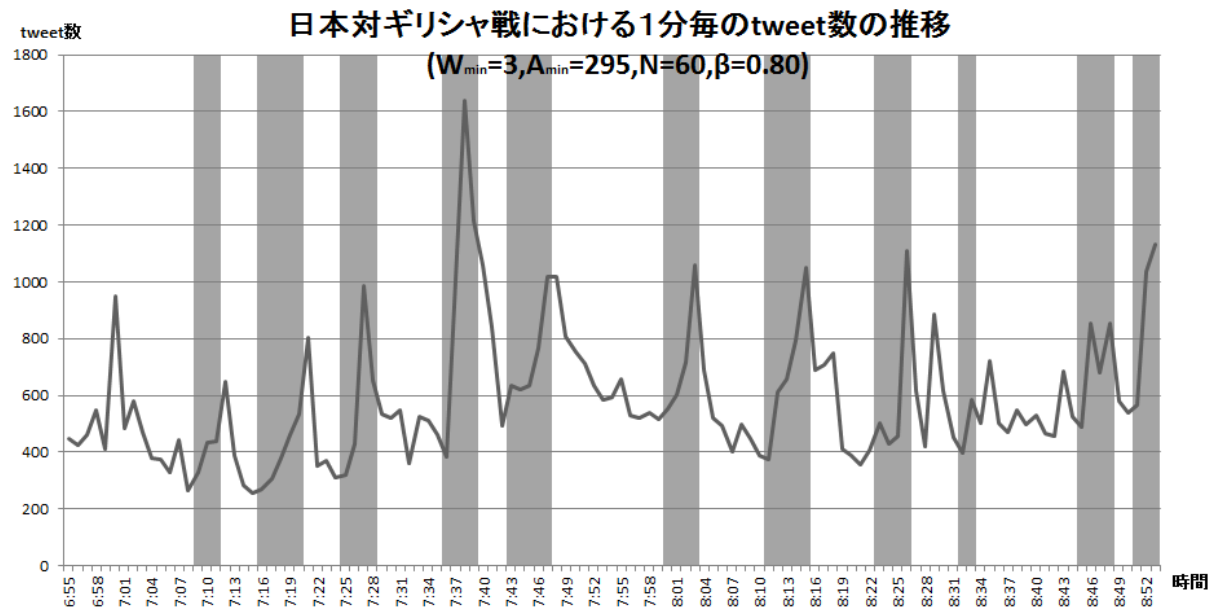


図 12 リアルタイムバースト検出結果

図 12 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[4], 土屋ら[5])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.

それぞれの図 12 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時間を以下の表 1 にまとめた.

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

以下に Tf-Idf による各バースト期間中の要約に利用される上位 10 件以内の単語を示す.

図 12 でバースト検出された各期間中で左から順にバーストの順序に対応する Tf-Idf の値を表 2, 表 3, 表 4, 表 14, 表 5, 表 6, 表 7, 表 8, 表 9, 表 10, 表 21 に示す.

表 2 1 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
感触	0.00269
スロットル	0.00185
ほんとう	0.00179
マキ	0.00154
すべて	0.00145
キプロス	0.00140
歴代	0.00126
日の丸	0.00109

表 3 2 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
オフサイド	0.00335
蛍光	0.00284
キャッチ	0.00211
オーバーヘッドクリア	0.00208
イラスト	0.00143

オーバー	0.00134
躍進	0.00131

表 4 3 番目のバースト内の Tf-Idf 値

ミューラー	0.00710
コツ	0.00710
エゴイスト	0.00710
ゲルト	0.00700
位置	0.00274
直後	0.00172
FK	0.00154
フリー	0.00136
キック	0.00130
右足	0.00128

表 5 4 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
退場	0.00830
カツラニス	0.00613
アクシデント	0.00538
交代	0.00524
右腕	0.00374
ゲカス	0.00362
イエロー	0.00277
レッド	0.00239
拳	0.00238
寒冷	0.00236

表 6 5 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
交代	0.00642
右腕	0.00524
カラグニス	0.00421
ゲカス	0.00345
拳	0.00331

寒冷	0.00331
殺人	0.00330
アディショナルタイム	0.00302
AT	0.00288
フォトファジデイス	0.00232

表 7 6 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
真っ先	0.00451
ハグ	0.00448
重点	0.00272
右腕	0.00265
闘志	0.00189
先ほど	0.00187
思い	0.00184
寒冷	0.00182
殺人	0.00182
交代	0.00181

表 8 7 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
保仁	0.00283
右腕	0.00189
交代	0.00186
OUT	0.00177
カッコ	0.00176
壁	0.00160
殺人	0.00133
寒冷	0.00133
拳	0.00132

表 9 8 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
交代	0.00989
右腕	0.00918
観客	0.00886
念	0.00882
マリア	0.00756
ジュリア	0.00756
日本語	0.00730
パンチング	0.00709
取材	0.00709

表 10 9 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
混戦	0.01250
パスワーク	0.00764
横	0.00457
斉藤	0.00431
ホイッスル	0.00233
陣	0.00183
神話	0.00140
フレー	0.00131
ギャグ	0.00131
バモ二	0.00131

表 11 10 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
トロシディス	0.00779
斉藤	0.00343
パワー	0.00335
交代	0.00324
ラスト	0.00286
豊田	0.00177
慎司	0.00174
きれい	0.00169

保仁	0.00167
FK	0.00164

表 12 11 番目のバースト内の Tf-Idf 値

Term	Tf-Idf 値
fk	0.00471
アディショナルタイム	0.00364
前線	0.00274
社員	0.00205
ラスト	0.00200
ロスタイム	0.00189
陣	0.00185
以前	0.00160
引き分け	0.00141
パワープレー	0.00133

以下の表 13 に Tf-Idf 要約手法による要約結果と表 14 に ROUGE-1 による評価結果を示す。

表 13 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 25 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で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 O(*・ω・*)O`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 14 ROUGE-1 による Tf-Idf 要約手法の評価結果

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 35 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 37 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。	0
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 25 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で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 O(*・ω・*)O`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 15 良い実況者の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 16 ROUGE-1 による良い実況者に基づく要約手法の評価結果

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

9. 考察

本研究で用いた手法は、まずバーストの判定を効果的に行うことに関する利点は、連続に発生する短文テキストである `tweet` に対して時系列上での解析とその内容に関する自然言語処理上の特徴がある。第一に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。

9.1. 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 12 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 12 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 12 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。

9.2. 要約生成に関する考察

9.2.1. Tf-Idf 要約手法に関する考察

表 14 Tf-Idf 要約手法による要約の評価結果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。

9.2.2. 良い実況者に基づく要約手法

良い実況者に基づく要約手法の評価に関しては、表 16 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、本研究で用いた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。

10. 結論

本研究では、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) <http://twitter.com>
- 2) 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.
- 3) 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
- 4) 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.
- 5) 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
- 6) 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.
- 7) <http://matome.naver.jp/odai/2140196031406402101>
- 8) <https://ja.wikipedia.org/wiki/%E3%82%B5%E3%83%83%E3%82%AB%E3%83%BC%E7%94%A8%E8%AA%9E%E4%B8%80%E8%A6%A7>
- 9) <http://sports.yahoo.co.jp/sports/soccer/japan/>
- 10) Lin, C.-Y:ROUGE:A Package for Automatic Evaluation of Summaries.Proc, The ACL-04 Workshop“Text Summarization Branches Out”,pp.74-81(2004).

謝辞

本研究を進めるにあたり, ご指導頂いた藤井章博准教授に感謝いたします. また, 共に切磋琢磨してきた藤井研究室の皆様に感謝いたします.

付録

実験データおよびプログラムを CD-ROM に付録として付与する。

以下で本研究の実験データに関して詳細を述べる。

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 Twitter 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。解析対象を日本語の tweet のみとした。そのため、日本語の投稿が多く含まれることが予想される日本代表戦の試合に解析対象データセットを限定した。

以下でハッシュタグに関する詳細について述べる。

各テレビ局で指定されているハッシュタグ一覧を以下に示す。

#nhk, #tvtokyo, #etv, #ntv, #tvasahi, #tbs, #fujitv, #tokyomx

本研究では、NHK, テレビ朝日, 日本テレビで試合の放送が行われたため、放送局のハッシュタグは #nhk, #ntv, #tvasahi, #tvasahi_soccer を使用した。

ブラジル開催 FIFA ワールドカップで使用された出場国 32 チームを予選リーグのグループ及びハッシュタグとして以下に示す。

グループ A	ブラジル #BRA	クロアチア #CRO	メキシコ #MEX	カメルーン #CMR
グループ B	スペイン #ESP	オランダ #NED	チリ #CHI	オーストラリア #AUA
グループ C	コロンビア #COL	ギリシャ #GRE	コートジボワール #CIV	日本 #JPN
グループ D	ウルグアイ #URU	コスタリカ #CTC	イングランド #ENG	イタリア #ITA
グループ E	スイス #AUI	エクアドル #ECU	フランス #FRA	ホンジュラス #HON
グループ F	アルゼンチン #ARG	ボスニアヘルツェゴビナ #BIH	イラン #IRN	ナイジェリア #NGR
グループ G	ドイツ #GER	ポルトガル #POR	ガーナ #GHA	アメリカ #USA
グループ H	ベルギー #BEL	アルジェリア #ALG	ロシア #RUS	韓国 #KOR

日本代表戦の公式サイトで公開されていた．#daihyo, #jfa を上記のハッシュタグに加え各試合に対して，指定したハッシュタグと日時を以下の表 17 に示す．

表 17 日本代表戦で用いたハッシュタグ，試合日時とハッシュタグ

試合日時	対戦国	ハッシュタグ
2014 年 6 月 15 日 AM10:00	日本 vs コートジボワール	#daihyo, #jfa, #nhk, #jpn, #civ
2014 年 6 月 20 日 AM7:00	日本 vs ギリシャ	#daihyo, #jfa, #ntv, #jpn, #gre
2014 年 6 月 25 日 AM5:00	日本 vs コロンビア	#daihyo, #jfa, #tvasahi, #tvasahi_soccer, #jpn, #col

尚，Twitter の仕様上ハッシュタグは，大文字，小文字の区別がなく，タグ付けされた実況 tweet からデータを検索することができるため，ここで記述したハッシュタグに関しては，大文字，小文字の区別は意味を持たない．

本研究の要約生成を行うときに，選手名及びイベント固有の専門用語を用いる．選手名は NAVER まとめ[7]から収集し，イベント固有の専門用語は，Wikipedia[8]と yahoo の提供するスポーツナビ[9]からサッカーの専門用語を収集した．以下に，日本代表戦の出場選手名を表 18，表 19，表 20 に，各試合において使用したイベント固有の専門用語を表 21 に示す．

表 18 日本 vs コートジボワール戦で使用した選手名辞書

日本選手名	コートジボワール選手
川島永嗣	バリー
権田修一	マンデ
西川周作	グボウオ
長友佑都	ヴィエラ
酒井宏樹	バンバ
酒井高德	ジャクパ
森重真人	ボカ
吉田麻也	アクパ・アクプロ
内田篤人	トゥ・レ
伊野波雅彦	オリエ
今野泰幸	ディアモンデ
山口蛍	セレイ・ディエ
長谷部誠	ティオテ
遠藤保仁	ヤヤ・トゥーレ
青山敏弘	ゾコラ
岡崎慎司	カルー
清武弘嗣	ドログバ
本田圭佑	ディディエ・ヤ・コナン
香川真司	ボリ
柿谷曜一朗	グラデル
大迫勇也	ボニー
大久保嘉人	シオ
齋藤学	ジェルヴィーニョ

表 19 日本 vs ギリシャ戦で使用した選手名辞書

日本選手名	ギリシャ選手
川島永嗣	カルネジス
権田修一	カピノ
西川周作	グリコス
長友佑都	ヴィントラ
酒井宏樹	モラス
酒井高德	パパスタソプーロス
森重真人	ツァヴェラス
吉田麻也	トロシディス
内田篤人	マノラス
伊野波雅彦	ホレバス
今野泰幸	サマリス
山口蛍	フェトファツィディ ス
長谷部誠	パナギオティス・タフ ツィディス
遠藤保仁	コネ
青山敏弘	ツィオリス
岡崎慎司	カラグニス
清武弘嗣	カツラニス
本田圭佑	クリストドゥロプー ロス
香川真司	ゲカス
柿谷曜一朗	サマラス
大迫勇也	ミトログル
大久保嘉人	サンピングディス
齋藤学	

表 20 日本 vs コロンビア戦で使用した選手名辞書

日本選手名	ギリシャ選手
川島永嗣	ヴァルガス
権田修一	モンドラゴン
西川周作	オスピナ
長友佑都	ヴァルデス
酒井宏樹	ジェペス
酒井高德	アリアス
森重真人	バランタ
吉田麻也	アルメロ
内田篤人	サパタ
伊野波雅彦	ロドリゲス
今野泰幸	キンテロ
山口蛍	サンチェス
長谷部誠	ファン・カミーロ・ス ニーガ
遠藤保仁	クアドラード
青山敏弘	グアリン
岡崎慎司	カルボネロ
清武弘嗣	メジャ
本田圭佑	アギラール
香川真司	ラモス
柿谷曜一朗	マルティネス
大迫勇也	イバルボ
大久保嘉人	グティエレス
齋藤学	バッカ

本研究では，イベント固有の専門用語は以下の辞書を用いる．収集したイベント固有の専門用語は 106 語となった．

表 21 イベント固有の専門専門用語辞書

イベント固有の専門用語
CK
アーリークロス
アウトサイドキック
アディショナルタイム
イエローカード
インサイド
オウンゴール
オーバーラップ
オープンスペース
オフサイド
カウンター
カット
キーパー
キープ
キッカー
キックオフ
キャッチ
くさび
グラウンダー
クリア
クリアボール
クリアミス
クロス
クロスバー
コース
コーナー
コーナーキック
ゴールマウス
ゴールライン
こぼれ球
コンビ

コンビネーション
サイド
サイドチェンジ
サイドバック
シュート
シュートミス
ショートコーナー
スペース
スライディング
スルー
スルーパス
スローイン
セービング
セーブ
センターバック
センタリング
ターン
ダイレクト
ダイレクトシュート
チャージ
ディフェンス
ディフェンスライン
トップ下
トラップ
ドリブル
ニアサイド
ネット
ハーフウェアライン
ハーフウェーライン
バイシクルシュート
バイタルエリア
パス
バックパス
ハンド
ビルドアップ

ファーサイド
ファーストシュート
ファインセーブ
ファウル
フォーメーション
フリー
プレス
プレッシャー
ブロック
ヘッド
ヘディング
ヘディングシュート
ペナルティアーキ
ペナルティーエリア
ペナルティエリア
ボール
ボールポゼッション
ポストプレー
ボランチ
ボレーシュート
マーク
マイナス
ミート
ミス
ミスパス
ミドルシュート
ラストパス
リスタート
リターン
リターンパス
レッドカード
ロビング
ロングシュート
ロングパス
ロングフィード

ロングボール
ワンツー
ワントップ

要約生成の正解要約は，人が手動で生成したヤフーの提供するスポーツナビを利用する．
表 22 に日本 vs コートジボワール戦で利用する正解要約，表 23 に日本 vs ギリシャ戦で利用する正解要約，表 24 に日本 vs コロンビア戦で利用する正解要約となる文書を示す．

表 22 日本 vs コートジボワールの正解要約

試合時間	テキスト
前半 0 分	コートジボワールボールでキックオフ、試合開始
前半 1 分	【日本】左サイドからの CK を獲得
前半 1 分	【日本】キッカーの香川はショートコーナーを選択。本田につなぐも、チャンスには至らず
前半 3 分	【コートジボワール】左サイドからの CK を獲得
前半 5 分	【コートジボワール】Y トゥレがオーリエにパスを出す。受けたオーリエは右サイドの敵陣深くから右足でクロスを供給するも、シュートには結び付かない
前半 6 分	【コートジボワール】右サイドからの CK を獲得
前半 7 分	【コートジボワール】キッカーのカルーが右足でクロスを入れる。だが、ニアサイドで DF にクリアされてしまう
前半 9 分	【コートジボワール】オーリエが右サイドの敵陣中央から右足でクロスを供給。だが、味方には合わず、ボールは逆サイドに流れてしまう
前半 11 分	【日本】山口が自陣からディフェンスラインに背後に浮き球のパスを出す。しかし、味方にはつながらず、バリーに処理されてしまう
前半 16 分	【日本】GOOOOOAL!!香川からパスを受けた長友が本田につなぐ。本田はペナルティエリア内で左足を振り抜くと、これがネットに突き刺さり、先制に成功する
前半 21 分	【日本】右サイドからの CK を獲得
前半 21 分	【日本】本田が右サイドの敵陣深くからシュート性のクロスを入れる。だが、バリーに処理されてしまう
前半 22 分	【コートジボワール】ペナルティエリア手前の中央で FK を獲得
前半 23 分	【日本】吉田にイエローカード
前半 24 分	【コートジボワール】キッカーの Y トゥレが右足で直接狙うも、シュートはクロスバーの上へ外れる
前半 27 分	【コートジボワール】Y トゥレがドリブルでペナルティエリア内に進入するも、得点には至らず
前半 27 分	【コートジボワール】カルーがペナルティエリア手前の中央から右足でシュートを放つ。しかし、枠をとらえられない
前半 30 分	【コートジボワール】ジェルビーニョが岡崎に倒され、ペナルティエリア手前からの FK を獲得

前半 31 分	【コートジボワール】キッカーのボカが左足で直接狙うも、わずかに枠をとらえられず
前半 36 分	【日本】左サイドからの CK を獲得
前半 36 分	【日本】キッカーの本田が左足でクロスを送るも、得点には至らない
前半 39 分	【コートジボワール】ディエがクロスを送ると、ペナルティエリア中央のボニーが右足で合わせる。だが、枠をとらえられない
前半 41 分	【コートジボワール】ボニーが左サイドの敵陣深くから右足でクロスを上げる。だが、至近距離で内田にブロックされてしまう
前半 43 分	【コートジボワール】ボカが左サイドの敵陣深くから左足でクロスを供給。カルーが頭で合わせるも、枠をとらえられず
前半 45 分	【コートジボワール】ジェルビーニョのドリブル突破は DF に阻まれるも、こぼれ球にボカが反応。ボカはペナルティエリア左脇からシュートを放つも、川島にキャッチされてしまう
前半 46 分	アディショナルタイムは 1 分の表示
前半 47 分	前半終了。1-0 と日本のリードで前半を折り返す
後半 0 分	日本ボールでキックオフ、後半開始
後半 2 分	【日本】香川が本田につなぎ、受けた本田はペナルティエリア内にスループスを出す。香川が走り込むも、得点には至らない
後半 3 分	【コートジボワール】オーリエがペナルティエリア右からシュート性のクロスを入れる。長友に当たって軌道が変わったボールにボニーが合わせるも、枠をとらえられない
後半 5 分	【日本】長谷部がペナルティエリア左から右足でシュートを放つ。だが、ゴールの左に外れてしまう
後半 7 分	【日本】内田が右サイドの敵陣中央からペナルティエリア内に浮き球を送る。しかし、バリーに直接キャッチされてしまう
後半 9 分	【日本】17 長谷部 OUT→7 遠藤 IN
後半 9 分	【日本】右サイドの敵陣中央で FK を獲得
後半 11 分	【日本】キッカーの本田が左足でクロスを送るも、DF にクリアされてしまう
後半 12 分	【日本】本田がペナルティエリア内にグラウンダーのスループスを出すも、バリーに処理されてしまう
後半 13 分	【コートジボワール】ゾコラにイエローカード
後半 15 分	【コートジボワール】ボカがカルーにつなぎ。受けたカルーはドリブルで持ち運び、敵陣中央から右足でシュートを放つ。だが、ゴールの左に外れてしまう
後半 17 分	【コートジボワール】20 ディエ OUT→11 ドログバ IN

後半 18 分	【コートジボワール】 ボニーが右サイドのジェルビーニョにパスを出す。ジェルビーニョはドリブルを仕掛けるも、フィニッシュには至らず
後半 19 分	【日本】 森重にイエローカード
後半 19 分	【コートジボワール】GOOOOOAL!!オーリエが右サイドの敵陣中央から右足で正確なクロスを供給。走り込んできたボニーが頭で合わせ、同点に追い付く
後半 21 分	【コートジボワール】GOOOOOAL!!オーリエが右サイドから右足でクロスを入れる。ジェルビーニョがヘディングシュートを放つと、これがゴールの右に決まり、逆転に成功
後半 22 分	【日本】 18 大迫 OUT→13 大久保 IN
後半 25 分	【日本】 本田が敵陣中央からディフェンスラインの裏にパスを出す。だが、反応した岡崎はボールを収められず
後半 30 分	【コートジボワール】 3 ボカ OUT→18 ジャクパ IN
後半 30 分	【日本】 内田が右サイドからクロスを上げる。ファーサイドの大久保が頭で折り返すも、ゾコラにクリアされてしまう
後半 32 分	【コートジボワール】 Y トゥレがペナルティエリア手前の中央からシュートを放つ。だが、枠をとらえられない
後半 33 分	【コートジボワール】 12 ボニーOUT→13 コナン IN
後半 35 分	【日本】 大久保がシュート
後半 35 分	【コートジボワール】 敵陣中央で FK を獲得
後半 37 分	【コートジボワール】 キッカーのドログバが右足でゴールの右を狙うも、川島のセーブに遭う
後半 40 分	【コートジボワール】 カルーのパスを受けたドログバが、ペナルティエリアでシュートを放つ。だが、森重に当たってゴールの右に外れてしまう
後半 41 分	【日本】 10 香川 OUT→11 柿谷 IN
後半 44 分	【コートジボワール】 右サイドからの CK を獲得
後半 45 分	【コートジボワール】 ショートコーナーを選択。ドログバが受けて時計の針を進める
後半 46 分	アディショナルタイムは 4 分の表示
後半 49 分	【コートジボワール】 ジェルビーニョが左サイドの敵陣深くからクロスを上げる。だが、DF に処理されてしまう
後半 51 分	試合終了。1-2 で日本が敗れた

表 23 日本 vs ギリシャの正解要約

試合時間	テキスト
前半 0 分	日本ボールでキックオフ、試合開始
前半 1 分	【日本】 吉田が自陣からグラウンダーの縦パスを入れる。しかし、味方には合 わず、カルネジスに処理されてしまう
前半 2 分	【日本】 山口がペナルティエリア手前の中央からシュートを放つ。だが、DF にブロックされてしまう
前半 4 分	【ギリシャ】 ミトログルがペナルティエリア内でシュートを放つ。しかし、得 点には至らない
前半 5 分	【ギリシャ】 左サイドからの CK を獲得
前半 6 分	【ギリシャ】 キッカーのホレバスが左足でクロスを供給するも、シュートには 結び付かず
前半 7 分	【日本】 本田がペナルティエリア内にパスを出す。だが、相手に阻まれてしま う
前半 11 分	【ギリシャ】 コネがセンターサークル内からドリブルで持ち上がる。コネはペ ナルティエリア手前の中央から右足を振り抜くも、シュートは川島に防がれて しまう
前半 12 分	【日本】 長谷部にイエローカード
前半 17 分	【日本】 本田がペナルティエリア内に浮き球を送る。しかし、大迫はオフサイ ドの判定
前半 19 分	【日本】 内田のパスを大久保がダイレクトで落とすと、大迫がペナルティエリ ア手前の中央から左足を振り抜く。だが、シュートはカルネジスに阻まれる
前半 21 分	【日本】 大迫がペナルティエリア手前の左から右足を振り抜く。しかし、シュ ートはゴールの右に外れてしまう
前半 23 分	【日本】 左サイドの敵陣中央からの FK を獲得
前半 25 分	【日本】 キッカーの本田が左足でクロスを入れる。しかし、精度を欠いてしま い、ボールはタッチラインを割ってしまう
前半 27 分	【ギリシャ】 カツラニスにイエローカード
前半 28 分	【日本】 大久保が倒され、ペナルティエリア手前の中央からの FK を獲得
前半 29 分	【日本】 キッカーの本田が左足で直接狙うも、カルネジスに防がれてしまう
前半 33 分	【日本】 長友が左サイドの敵陣中央からクロスを上げると、大久保が頭で合わ せる。しかし、ヘディングシュートは枠をとらえられず
前半 35 分	【ギリシャ】 9 ミトログル OUT→17 ゲカス IN
前半 36 分	【ギリシャ】 コネが左サイドの敵陣中央からペナルティエリア内にパスを出 す。しかし、味方には合わず、川島に処理されてしまう

前半 38 分	【ギリシャ】 カツラニスにイエローカード。2 枚目で退場となる
前半 39 分	【日本】 右サイドからの CK を獲得
前半 39 分	【日本】 キッカーの本田が左足でクロスを上げるも、DF にヘディングでクリアされてしまう
前半 40 分	【ギリシャ】 トロシディスがペナルティエリア内で右足を振り抜く。だが、川島の好セーブに遭う
前半 41 分	【ギリシャ】 18 フェトファツィディス OUT→10 カラグニス IN
前半 44 分	【ギリシャ】 敵陣浅い位置からの FK を獲得
前半 45 分	【ギリシャ】 キッカーのカラグニスが右足でペナルティエリア内に浮き球を入れるも、得点には至らない
前半 46 分	アディショナルタイムは 2 分の表示
前半 48 分	前半終了。0-0 と、スコアレスで前半を折り返す
後半 0 分	ギリシャボールでキックオフ、後半開始
後半 0 分	【日本】 17 長谷部 OUT→7 遠藤 IN
後半 4 分	【日本】 大久保がペナルティエリア手前の中央の大迫にパスを出す。しかし、大迫はボールを収められず
後半 5 分	【日本】 長友が倒され、敵陣中央の左からの FK を獲得
後半 6 分	【日本】 キッカーの遠藤が右足でペナルティエリア内に浮き球を送る。しかし、味方には合わず、ボールは直接ゴールラインを割ってしまう
後半 8 分	【日本】 内田が右サイドから右足でクロスを供給する。しかし、DF にクリアされてしまう
後半 9 分	【日本】 大久保がサマラスに倒され、敵陣中央からの FK を獲得
後半 10 分	【ギリシャ】 サマラスにイエローカード
後半 11 分	【日本】 キッカーの遠藤が右足で直接狙うも、シュートは壁に当たってしまう
後半 12 分	【日本】 18 大迫 OUT→10 香川 IN
後半 13 分	【ギリシャ】 左サイドからの CK を獲得
後半 14 分	【ギリシャ】 キッカーのカラグニスが右足でクロスを供給するも、得点には至らず
後半 17 分	【ギリシャ】 サマラスが大久保に倒され、敵陣中央の左からの FK を獲得
後半 17 分	【ギリシャ】 キッカーのホレバスが左足でペナルティエリア内に浮き球を送るも、シュートには結び付かない
後半 20 分	【日本】 自陣でボールを奪い、カウンターに転じる。香川がドリブルで持ち上がり、左サイドの長友につなぐ。受けた長友は左足でクロスを上げるも、シュートには結び付かない
後半 22 分	【日本】 キッカーの本田が左足で直接狙うも、得点には至らず

後半 23 分	【日本】 香川のパスに内田が反応。内田はペナルティエリア右から中央に折り返し、ファーサイドの大久保が左足で合わせる。だが、シュートは枠をとらえられず
後半 25 分	【日本】 本田がペナルティエリア手前からシュートを放つ。しかし、至近距離で DF にブロックされてしまう
後半 26 分	【日本】 長友が左サイドの敵陣深くから左足でクロスを上げると、ゴール前で混戦に。最後は内田がこぼれ球に反応し、シュートを放つも、わずかにゴールの右に外れてしまう
後半 30 分	【日本】 遠藤のパスを受けた長友が、左サイドの敵陣深くから右足でクロスを供給。だが、DF にヘディングでクリアされてしまう
後半 31 分	【日本】 大久保が倒され、右サイドの敵陣中央からの FK を獲得
後半 31 分	【日本】 キッカーの遠藤が右足でクロスを入れるも、得点には至らない
後半 32 分	【日本】 大久保が敵陣中央からミドルシュートを放つ。だが、カルネジスに阻まれてしまう
後半 36 分	【ギリシャ】 8 コネ OUT→14 サルピングディス IN
後半 38 分	【日本】 内田がペナルティエリア右から右足でグラウンダーのクロスを入れる。しかし、フィニッシュには至らない
後半 39 分	【日本】 本田が長友にパスを出す。受けた長友は左サイドから右足でクロスを上げるも、DF にクリアされてしまう
後半 40 分	【日本】 内田のクロスは DF にブロックされるも、右サイドからの CK を獲得
後半 40 分	【日本】 長友がペナルティエリア左から右足でシュート性のクロスを入れると、ゴール前の吉田が合わせる。しかし、得点には至らず
後半 43 分	【日本】 岡崎がトロシディスに倒され、ペナルティエリア手前からの FK を獲得
後半 44 分	【ギリシャ】 トロシディスにイエローカード
後半 45 分	【日本】 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
後半 46 分	アディショナルタイムは 4 分の表示
後半 46 分	【日本】 本田が自陣からドリブルで持ち上がり、ペナルティエリア内にスルーパスを出す。しかし、DF に対応されてしまう
後半 48 分	【日本】 遠藤がペナルティエリア内に浮き球を送る。しかし、競り合った吉田がファウルの判定を受けてしまう
後半 50 分	試合終了。スコアレスドローに終わった

表 24 日本 vs コロンビア戦

試合時間	
前半 0 分	日本ボールでキックオフ、試合開始
前半 3 分	【コロンビア】 キンテーロがドリブル突破を図るも、ペナルティエリア内で青山の対応に遭い、ボールを失ってしまう
前半 4 分	【コロンビア】 アリアスが右サイドの敵陣深くから右足でクロスを入れる。だが、今野にクリアされてしまう
前半 6 分	【日本】 長谷部がグアリンに倒され、ペナルティエリア手前の右からの FK を獲得
前半 7 分	【日本】 キッカーの本田が左足で直接狙うも、シュートは壁に当たり、オスピナに処理されてしまう
前半 10 分	【日本】 大久保がミドルシュートを放つも、枠をとらえられない
前半 12 分	【日本】 左サイドからの CK を獲得
前半 13 分	【日本】 キッカーの本田はショートコーナーを選択。ボールを受けた香川は左サイドの敵陣深くから右足でクロスを上げる。しかし、DF にヘディングでクリアされてしまう
前半 16 分	【コロンビア】 ラモスが今野に倒され、PK を獲得
前半 16 分	【日本】 今野にイエローカード
前半 17 分	【コロンビア】 GOOOOOAL!!キッカーはクアドラード。右足で放ったシュートがゴール中央に決まり、先制に成功する
前半 20 分	【日本】 右サイドからの CK を獲得
前半 20 分	【日本】 キッカーの本田が左足でクロスを供給するも、オスピナにパンチングで処理されてしまう
前半 21 分	【日本】 青山がペナルティエリア手前の右から浮き球を送る。しかし、オスピナに直接キャッチされてしまう
前半 22 分	【日本】 岡崎がバランタに倒され、ペナルティエリア手前の右からの FK を獲得
前半 23 分	【日本】 キッカーの本田が左足で直接狙うも、壁に阻まれる
前半 24 分	【コロンビア】 クアドラードが右サイドからディフェンスラインの背後にスルーパスを出す。マルティネスが反応するも、DF の対応に遭い、シュートまで持ち込めない
前半 26 分	【日本】 香川が DF をかわし、ペナルティエリア手前から右足でグラウンダーのシュートを放つ。だが、オスピナに防がれてしまう
前半 28 分	【日本】 大久保がペナルティエリア手前の右の内田につなぐ。内田は突破を図るも、アルメロの対応に遭い、ボールを失ってしまう

前半 30 分	【日本】 内田が右サイドの敵陣中央からペナルティエリア内にグラウンダーのパスを出す。岡崎がヒールで流すも、味方にはつながらず
前半 32 分	【日本】 ペナルティエリア手前の中央からの FK を獲得
前半 33 分	【日本】 キッカーの本田が左足で直接狙う。しかし、シュートはわずかにゴールの右に外れてしまう
前半 36 分	【日本】 内田が右サイドの敵陣深くから右足でクロスを上げる。ファーサイドの大久保が右足で合わせるも、シュートはゴールの左に外れてしまう
前半 38 分	【コロンビア】 こぼれ球に反応したキンテーロがペナルティエリア手前から左足でシュートを放つ。だが、シュートはゴールの右に外れてしまう
前半 39 分	【日本】 右サイドからの CK を獲得
前半 40 分	【日本】 キッカーの本田はショートコーナーを選択。香川が受けるも、チャンスには至らず
前半 42 分	【コロンビア】 ラモスがペナルティエリア右から右足でグラウンダーのクロスを入れる。中央のマルティネスがダイレクトで合わせるも、シュートはゴールの左に外れてしまう
前半 46 分	アディショナルタイムは 1 分の表示
前半 46 分	【日本】 GOOOOOAL!!本田が右サイドでボールを持ち、カットインからクロスを上げる。ゴール前で岡崎が頭で合わせると、これがネットを揺らし、同点に追い付く
前半 47 分	前半終了。1-1 と、同点で前半を折り返す
後半 0 分	コロンビアボールでキックオフ、後半開始
後半 0 分	【コロンビア】 20 キンテーロ OUT→10 ロドリゲス IN
後半 5 分	【コロンビア】 ロドリゲスがドリブルで中に持ち込み、ペナルティエリア手前からシュートを放つ。しかし、DF にブロックされてしまう
後半 7 分	【コロンビア】 ラモスが右サイドからパスを出す。マルティネスが抜け出すも、オフサイドの判定
後半 7 分	【コロンビア】 ロドリゲスが左サイドの敵陣深くから左足でクロスを上げる。しかし、シュートには結び付かない
後半 8 分	【コロンビア】 左サイドの敵陣中央からの FK を獲得
後半 9 分	【コロンビア】 キッカーのロドリゲスが左足でクロスを供給。ゴール前のカルボネロが合わせるも、シュートは枠をとらえられず
後半 10 分	【コロンビア】 GOOOOOAL!!アリアスが右サイドからドリブルで中に持ち込み、ロドリゲスにパスを出す。受けたロドリゲスはマルティネスにつなぐと、マルティネスはワントラップから左足でシュートを放つ。これが決まり、勝ち越しに成功する

後半 13 分	【日本】 青山が長谷部にパスを出し、受けた長谷部は香川につなぐ。香川はペナルティエリア内から左足でシュートを放つも、枠をとらえられない
後半 14 分	【日本】 長友が左サイドの敵陣深くから左足でクロスを供給。しかし、得点には至らず
後半 17 分	【日本】 14 青山 OUT→16 山口 IN
後半 18 分	【日本】 香川がグアリンに倒され、敵陣中央の左からの FK を獲得
後半 18 分	【コロンビア】 グアリンにイエローカード
後半 19 分	【日本】 キッカーの本田が左足で直接狙うも、シュートはオスピナに防がれてしまう
後半 20 分	【日本】 内田が右サイドの敵陣深くからクロスを供給。中央の大久保が右足で合わせるも、シュートは枠をとらえられない
後半 24 分	【日本】 9 岡崎 OUT→11 柿谷 IN
後半 24 分	【日本】 香川のシュートは DF にブロックされるも、右サイドからの CK を獲得
後半 25 分	【日本】 キッカーの本田が左足でクロスを供給するも、シュートには結び付かない
後半 27 分	【コロンビア】 マルティネスのシュートは DF に防がれるも、左サイドからの CK を獲得
後半 28 分	【コロンビア】 キッカーのグアリンが右足でクロスを上げるも、DF にヘディングでクリアされてしまう
後半 29 分	【日本】 今野がペナルティエリア手前からシュートを放つ。しかし、メヒアに当たり、オスピナにキャッチされてしまう
後半 31 分	【日本】 左サイドからの CK を獲得
後半 31 分	【日本】 キッカーの香川が右足でクロスを上げるも、ニアサイドで DF にヘディングでクリアされる
後半 32 分	【日本】 長友がペナルティエリア左から右足でシュートを放つ。しかし、オスピナにキャッチされてしまう
後半 35 分	【日本】 香川がペナルティエリア左でドリブルを仕掛けるも、DF の対応に遭い、シュートまで持ち込めず
後半 37 分	【コロンビア】 GOOOOOAL!!ロドリゲスがディフェンスラインの裏にスルーパスを出す。抜け出したマルティネスがペナルティエリア内で切り返し、左足でシュートを放つ。これがネットを揺らし、リードを広げる
後半 40 分	【コロンビア】 1 オスピナ OUT→22 モンドラゴン IN
後半 40 分	【日本】 10 香川 OUT→8 清武 IN
後半 43 分	【コロンビア】 アルメロがペナルティエリア左でドリブル突破を図る。しかし、

	相手の対応に遭い、ボールを失ってしまう
後半 45 分	【コロンビア】 GOOOOOAL!!ロドリゲスがペナルティエリア内で吉田をかわし、左足でループシュート。これが決まり、追加点を奪う
後半 46 分	アディショナルタイムは 3 分の表示
後半 49 分	試合終了。1-4 で日本が敗れた