

イベントの時空間推定に基づく ブログとマイクロブログの統合閲覧システムの提案

野呂 勇太[†] 三浦 惇貴^{††} 廣田 雅春^{†††} 野澤 浩樹^{††††} 横山 昌平^{†††††}

[†] 静岡大学大学院情報学研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

^{††} 静岡大学情報学部 〒432-8011 静岡県浜松市中区城北 3-5-1

^{†††} 静岡大学創造科学技術大学院/日本学術振興会特別研究員 DC 〒432-8011 静岡県浜松市中区城北 3-5-1

^{††††} 株式会社シーポイント 〒432-8002 静岡県浜松市中区富塚町 1933-1 佐鳴湖パークタウンサウス 2F

^{†††††} 静岡大学大学院情報学研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: [†]gs13026@s.inf.shizuoka.ac.jp, ^{††}cs10094@s.inf.shizuoka.ac.jp, ^{†††}dgs11538@s.inf.shizuoka.ac.jp, ^{††††}hiro@c-point.co.jp, ^{†††††}yokoyama@inf.shizuoka.ac.jp

あらまし 本研究では、ブログとマイクロブログの記事を統合することにより、自然言語処理を用いずにイベント内の盛り上りを要約する手法を実現する。マイクロブログの記事は、イベントに参加している人々の行動や感想が逐一に投稿されるため、投稿数のバーストからイベント内での盛り上りを検知することができる。しかしながら、1つの記事の文字数が少なく、また、連続する記事間の関係も明示的ではないため、バーストの内容を把握することが困難である。それに対し、ブログの記事は整形された文章であるため、イベントの内容を把握しやすいが、イベント全体を単位とした記事が多く、イベント内のバーストの検知はできない。そこで、本論文では、マイクロブログからバーストを検知し、ブログ記事が内包する写真の撮影時刻からバースト毎に記事を分割し、要約として提示する。

キーワード 記事統合, イベント発見, ブログ, マイクロブログ, Hamazo, Twitter

1. はじめに

近年、ブログやマイクロブログなどの Consumer Generated Media (CGM) の普及に伴い、多くの人々の意見や感想を Web 上から取得することが容易になった。このような人々の意見や感想は、マーケティングや社会分析の観点から重要であり、それらを情報源として実世界のイベントに関する情報の抽出や分析を行う研究が増加している [1] [2] [3] [4] [5]。

近年は、代表的なマイクロブログである Twitter^(注1)を対象とした研究が増加しており、例えば、Twitter に投稿されるツイートからイベントの発見を行う研究が行われている。そのような研究として、Sakaki ら [1] と Li ら [2] の研究がある。Sakaki らは、実世界のイベントのうち、地震や台風のような災害に着目し、イベントの発見を行う手法を提案している。Sakaki らの手法は、対象とする特定のイベントの発生に関係するかどうかを SVM を用いて分類し、イベントの発生に関係すると分類されたツイートが、短期間に多数投稿されている場合にイベントとして検出する。一方、Li らは、イベントを発見するためにツイートセグメントという 1-gram 以上の単語のまとまりを定義し、同様のツイートセグメントが出現するツイートの急激な増加（バースト）を検出することでイベントの発見を行う。このようなマイクロブログに対する研究において、ブログに対する既存研究を適用することは困難であると考えられる。その原因として、ブログとマイクロブログの特徴が異なるということが

あげられる。

マイクロブログの 1 つである Twitter には、以下に述べる特徴がある。Twitter に投稿されるツイートと呼ばれる記事は、140 字以内という制限がある。このような文字数の制限により、ツイートには、省略表現などの特有の表現が用いられることがある。さらに、連続するツイート間の関係が明示的ではない場合が多いため、投稿者の正確な意図を把握することが困難である。また、マイクロブログの特徴として、従来のブログと比較してリアルタイム性の高い内容の記事が多く投稿されることや記事の更新頻度が高いことがあげられる。これについて、Java ら [3] は、マイクロブログが容易に記事の投稿が可能であることや記事の内容を考える負担が軽減したことが要因の 1 つであると述べている。このような特徴により、投稿数のバーストからイベント内の盛り上りを検知することが容易であると考えられる。

一方、ブログは、ユーザが内容を慎重に考えて記事を書く傾向があるため、マイクロブログと比較して文字数が多く、整形された文章という特徴がある。そのため、マイクロブログの記事を閲覧する場合と比較して、ユーザは少数の記事を閲覧することで、ブログ記事からイベントの内容を把握することが容易であると考えられる。また、記事において言及されているイベントの発生期間と記事の投稿時刻には、差があるという特徴があげられる。あるイベントに参加したブログのユーザは、そのイベントの終了後にイベントに関する記事を投稿する傾向が多いため、ブログ記事の投稿時刻からイベントの発生期間を推定することは困難である。

(注1) : <https://twitter.com/>

そこで、本研究は、実世界で起きたイベントに基づいてブログとマイクロブログを統合することで、2つのメディアの特徴を併せ持つデータを生成する。マイクロブログの記事は、イベントに参加している人々の行動や感想が逐一に投稿されるため、投稿数のバーストからイベント内での盛り上がりを検知することが可能である。そのため、マイクロブログに投稿された記事の投稿時刻に基づいて投稿数の急激な増加を検出することで、イベントの発生期間を推定する。記事の統合は、同一の実世界のイベントについて言及している記事について行う。その際、ブログ記事が投稿された時刻ではなく、ブログ記事に書かれているイベントが発生した時間を推測することで、その時間情報に基づいてブログとマイクロブログの記事を統合可能であると考えられる。統合の結果として、同一の実世界のイベントに関するブログとマイクロブログの各記事は、関連づけられて、実世界のイベントに関する記事群となる。

2つの異なる種類のメディアを統合した結果は、イベントの要約生成、評判分析、実世界の動向予測などの研究に応用可能であると考えられる。本研究では、2つの異なる種類のメディアを統合した結果を用いた研究の1つとして、イベントの要約を生成し、ユーザに提示するシステムを提案する。イベントの要約を提示することで、イベントに関する情報を得るために大量のブログ記事やツイートを閲覧するというユーザの負担の軽減を目指す。また、本研究では、イベント内の盛り上がり毎に要約を提示することで、時間とともに変化するイベントの様子の把握を容易にすることを目指す。

本論文の構成を以下に示す。第2節では、関連研究について述べる。第3節では、ブログとマイクロブログの記事を統合し、要約を提示する手法と、それを用いたシステムについて述べる。第4節では、提案システムの動作例と考察について述べる。最後に、第5節では、本論文のまとめと今後の課題について述べる。

2. 関連研究

2.1 2つの異なる種類のメディアの統合を行う研究

2つの異なる種類のメディアの統合を行う研究として、Tsagkiasら[6]とBronら[7]の研究がある。TsagkiasらとBronらの研究は、ソーシャルメディアの記事のほとんどは、ニュースに取り上げられたイベントに関係しているという仮定に基づいて、オンラインニュースの記事とソーシャルメディアの記事を統合する手法を提案している。これらの手法では、オンラインニュースとソーシャルメディアの記事の投稿時刻の差が閾値よりも小さい記事を統合している。しかし、オンラインニュースやブログ記事は、イベント中ではなく、イベント終了時以降に記事が投稿されることが多い。また、イベントの種類や規模などによって、適切な閾値は異なると考えられる。そのため、本研究では、イベントの発生期間を推定し、ブログ記事に含まれる写真の撮影時刻に基づいて統合を行うことで、写真を含むブログ記事については、適切な統合を実現する手法を提案した。

オンラインニュースの記事は、記事の大部分を文章が占め、

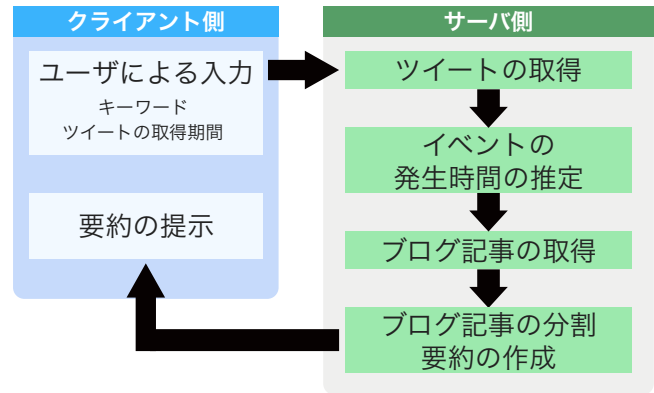


図1 システム構成図

イベントの発生期間に撮影されたイベントに関する写真を内包していることは少ない。そのため、TsagkiasらやBronらの手法では、記事中の文章のみに着目して記事の統合を行っている。そのような場合、記事中にイベントの発生期間に関する記述が存在しない限り、同様の期間に開催された同様のイベントに関する記事を統合することは困難である。時間情報を用いずに統合した場合、あるイベントに関する記事とその1年前に開催された同様のイベントに関する記事が統合されてしまうことが考えられる。定期的に開催されるイベントの場合、発生期間が異なるとイベントの内容が異なると考えられるため、発生期間の異なるイベントに関する記事同士を統合することは適切でないと考えられる。

2.2 要約に関する研究

要約に関する研究として、坂本ら[4]とShenら[5]の研究がある。坂本らは、Twitterからリアルタイムにイベントを検出し、随時イベントの要約を生成する手法を提案している。坂本らの手法では、Twitterから検出したイベントに対して、代表となる1つのツイートを選出し、そのイベントの要約としている。しかし、1つのイベントに対して、代表となる1つのツイートのみを選択し、要約としているため、イベントの内容を十分に表す要約の作成が困難であるという課題があげられる。

一方、Shenらは、イベントの参加者に着目して、Twitterから発見したイベントの要約を行っている。Shenらの手法では、スポーツの試合に出場している選手や企業などのカンファレンスにおける講演者の名前が出現するツイートから、イベントの要約を表す代表的なツイートを抽出し、要約を行っている。しかし、参加者の名前を含んでいるツイートを発見することで要約を行っているため、イベントの進行に関わる人物の名前がツイートに出現しないようなイベント(e.g. エア・フェスタ)に対して、要約を作成することはできない。

本研究では、イベントの要約としてブログ記事の文章を用いることで、代表的なツイートを提示することと比較してイベントの概要が容易に把握できる要約の作成を目指す。

3. 提案システム

本節では、提案するイベントに関する記事の統合閲覧システムについて説明する。本システムは、実世界のイベントに基づ

いてブログとマイクロブログの記事を統合し、その結果を用いて、イベントに関する要約と写真を提示する。本システムは、次に述べる仮説に基づいてイベントの要約を提示する。ブログに投稿される記事は、イベント全体を単位とした記事が多く、イベントの概要を把握することが可能であると考えられる。また、ブログ記事に含まれる写真の周辺には、その写真について説明している文章が書かれていると考えられる。また、あるイベントに関するブログ記事は、そのイベントの終了後に投稿される事が多い。そのため、ブログ記事が言及しているイベントの発生期間を、記事の投稿時刻を用いて推測することは適切ではない。そこで、イベントに関する記事が内包する写真は、そのイベントが発生している時間に撮影されていると考えられるため、バーストしている時間中に撮影された写真を含み、ユーザが入力したキーワードを含むブログ記事を統合する記事として取得する。そして、写真の撮影時刻を用いて、バーストを検出した時間毎にブログ記事を分割し、それをイベント内の盛り上がり要約とする。

本システムは、図1のシステム構成図において示した手順で記事の統合とイベント内の盛り上がり要約の提示を行う。最初に、ユーザにより入力されたキーワードと記事の取得期間を用いて、キーワードを含み、取得期間中に投稿されたツイートを取得する。次に、取得したツイートをを用いてイベントの発生期間の推定を行う。そして、推定したイベントの発生期間に基づいて、ブログ記事が内包する写真の撮影時刻を用いて、統合するブログ記事を取得する。そして、イベントの発生期間に基づいて取得したブログ記事とバーストを検出した時間中に投稿されたツイートを統合する。次に、バーストしている時間中に撮影された写真を内包するブログ記事を写真の撮影時刻を用いて分割し、分割した文章をバーストの盛り上がり要約とする。最後に、記事の統合とイベント内の盛り上がり要約の結果として、バーストを検出した時間、統合したブログ記事とツイート、バーストしている時間中に撮影された写真、その時間中のイベントの要約を表示する。

3.1 イベント発生期間の推定

本システムで行うイベント発生期間の推定では、マイクロブログに投稿された記事の投稿時刻に基づいてバースト検出を行う。バーストの検出には、蝦名らが提案した手法[8]を用いる。本システムでは、蝦名らの手法を用いて、バースト検出を行った結果として、バーストしている時間をイベント発生期間として扱う。

蝦名らの手法では、データの到着間隔が急激に短くなっている期間をバーストと定義しており、到着間隔の変化率が大きいほど、バーストの度合いは大きくなる。バーストの度合いは、aggregation pyramid というデータ構造を用いて、あるセルとそのセルの直前の最上位のレベルのセルの到着間隔の変化率を算出することで求める。最小ウィンドウサイズとして期間(e.g. 1時間, 1日)を設定し、その期間中に発生したイベント(本研究の場合、ツイートの投稿)の平均到着間隔を各セル毎に算出し、バーストの度合いの算出に利用する。また、一定時間毎にバーストを検出するのではなく、イベントの発生毎(本研究

表1 システムの実行時に用いたデータセット

	はまぞう		Twitter	
	期間	記事数	期間	記事数
データセット1	2013/10/13~10/27	32,414	2013/10/12~10/21	12,083
データセット2	2014/1/13~2/6	51,467	2014/1/12~2/11	940,882

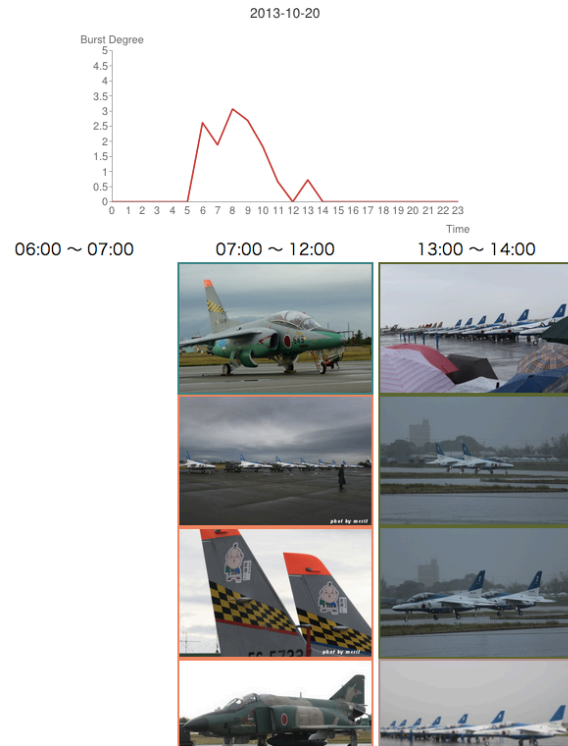


図2 システムの動作例(検索語:“エアフェスタ” or “エア・フェスタ”)

の場合、ツイート投稿毎)にバーストを解析する。本システムでは、最小ウィンドウサイズを1時間に設定し、出力として得られた1時間毎のバーストの度合いを用いて、イベントの発生期間を推定する。

3.2 統合するブログ記事の取得

推定したイベントの発生期間とユーザが入力したキーワードを用いて、ブログ記事の取得し、ツイートと統合を行う。その際、前述したとおり、ブログ記事が内包する写真の撮影時刻を用いる。バーストしている時間中に撮影された写真を含み、ユーザが入力されたキーワードを含むブログ記事を統合する記事として取得する。ここで取得したブログ記事を用いて、イベントの要約を作成する。

3.3 イベントの要約提示のためのブログ記事の分割

本研究では、イベントを通して1つの記事が書かれることが多いブログ記事を用いて要約を生成する。ブログ記事において、バースト期間中に撮影された写真の前後に書かれている段落中の文章は、そのイベントについて書かれている可能性が高いと考えられる。そのため、バースト期間中に撮影された写真の前後の段落を抽出し、それをイベントの要約とする。ここで、本研究における段落とは、ブログ記事を空行で区切った際の文章の集合である。最終的に、抽出した段落から重複する文章を削除し、ユーザに要約として提示する。

表 2 検索語“エアフェスタ” or “エア・フェスタ”を与えた際の要約結果

バーストしている時間	要約
06:00～07:00	なし 昨夜から降り続いた雨も朝には上がり 天候が持ちこたえてくれたと思ったのもつかの間で雨のエアフェスタ 2013 になり 501SQ のフライトまで実施されましたがその後はキャンセルになってしまいました 雨にも関わらず多くの人で会場は賑わっていましたブルーの列線。 空が暗い… 家康君塗装機 当然ご本人も登場 (笑) RF-4E CH-47J 海自の SH-60J U-4 C-1 昨年は大行列が出来て断念しましたが、今日は入れました (iv UH-60J U-125A F-15J(1 衛校) F-2A(1 衛校) 警空 E-2C & AWACS あああーーーーーっ! 30 周年アニバーサリーの記念マークがっ! キティちゃんがラブリー o(≧▽≦)o T-400 T-33 昨年はギンギラギンでしたが今年はそれらしく復元されます d(´-`)o AWACSJr. のタクシードアウト (笑) 個人的には E-7 67 ジュニアがお気に入りですただのパロディーバイクと思いきや、円盤も回るし、翼の両端も赤と緑のランプが点滅して意外と本格的です 昨日はネイル検定一級の試験官として名古屋でお仕事でしたのでサロンはお休みさせていただきました! 昨日は朝からエア・フェスタ浜松に来ています。 サロンは明日 10 時から営業いたします。 よろしく願い申し上げます☆以前は書かれてなかった富士山裾野の文字。 轟音で起きました (^^;) エアフェスタのならし飛行かな、一機ずつ旋回してるみたい新型の航空消防車 A-NB-2 昨日参加し損ねた先着無料体験の木工工作を娘と一緒にやってきました ん工作ではないか… 背もたれに絵を描いただけですからー 大きな音が気になるようです横を向いていても耳だけ動かして音を聞いてます 2ヶ500 円のジャンボ焼き帆立を食べました テントがあっても横殴りな雨が降りかかって グルメでは…まず、外の会場でやってた 東北女川の方々がやってた無料秋刀魚と…そんでもって次にシーポイントかね田さんが 絶賛してたスバイス 6 のチキンカマージを うん確かに見た目は大したことないけど…(笑) 今回は当店スタッフ中村の一押しフレイムのご紹介です。 もちろん愛用してますよ。 FaceFonts FF-121 Col 01 牡丹 です。 フロントメタル マゼンタ フロント樹脂 ライトグレー テンプル アイボリー の 3 カラー使用。 コンビネーションフレームでおしゃれですよ。 鼻パッドも滑りにくく、優しい肌触りの全面シリコンパッドを使用しています。 パッドアームにもβチタンを採用して、鼻への負担を軽減しています。 テンプルには軽くて錆にくいチタニウムを使用して、しっかり感も演出しています。 超軽量フレームで掛け心地抜群なんです。 しかもフロントはβチタンを使用しており、とってもしなやか。 このモデルは 4 カラー展開になっております。 ピアン正明堂さんの珍品トマト大箱を あっ、アイドルなライブも観ましたノリノリだぜベイバー (笑) 予定では 13:00 からブルーインパルスの展示飛行の予定です。 パイロットの紹介からです。 雨の中とはいえ人垣がすごくて見えないです。 ことまちの手前にあった久米吉さんで こんにく料理無料バイキングやってたので 当然立ち寄りましたよ。ヘルシーまいうー しかし、このあとはパレードだけで展示飛行は中止でした。 来年までお預けです。すみませんでした。 飛んでるところを見なかったなあ。。。 今日、航空自衛隊浜松基地にてエア・フェスタ in 浜松 2013 が開催され式典に参加してきました。 式典は大勢の人で賑いましたが、ブルーインパルの展示飛行は荒天のため残念ながら中止となり地上滑走のみ行われました。 浜松地区での大きなイベントで県外からも大勢の人が楽しみにして来るだけに残念ですね。 私も雲ひとつない晴天の中で展示飛行を見たのは 2 回ぐらいでしょうか? 当日、晴天の確立は低いですね。
07:00～12:00	
13:00～14:00	

4. システムの動作例

本節では、提案システムによる、イベントの抽出に基づいたブログ記事とマイクロブログの記事の統合と、要約の提示例について述べる。記事の統合は、静岡県浜松市の地域ブログポータルサイト「はまぞう^(注2)」の記事と Twitter のツイートを用いて行った。本論文では、表 1 に示す 2 つのデータセットを用いる。これら 2 つのデータセットに対する提案システムの動作例について以下に述べる。

データセット 1 に対して、期間として 2013 年 10 月 18 日から 2013 年 10 月 21 日を与え、キーワードとして“エアフェスタ”と“エア・フェスタ”を与えて OR 検索を行い、提案システムの記事統合を行った結果の一部を図 2 に示す。キーワードとして与えた“エア・フェスタ”は、航空自衛隊による航空機の展示飛行が行われるイベントのことであり、静岡県の航空自衛隊浜松基地で開催される“エア・フェスタ浜松 2013”は、2013 年 10 月 20 日の 09:00 から 15:00 の間に行われた。図 2 は、バーストを検出した 10 月 20 日の結果となっており、図中のグラフ

は、横軸が 00:00 から 23:00 までの時間、縦軸がバーストの度合いを表している。また、グラフの下には、バーストの盛り上がり毎の時間とその時間中に撮影された写真を表示している。

20 日に検出したバーストには、06:00 から 07:00、07:00 から 12:00、13:00 から 14:00 の 3 つの盛り上がりが存在した。10 月 20 日に発見した 3 つの盛り上がり毎の要約結果を表 2 に示す。06:00 から 07:00 の間に撮影された写真や要約が存在しないという結果から、その時間はエア・フェスタが行われている時間とは異なると推測可能である。06:00 から 07:00 の間にバーストを検出した原因は、エア・フェスタに注目しているユーザによる“これからエア・フェスタに行ってきます”というような内容の投稿が増加したためであると考えられる。

13:00 から 14:00 の間に撮影された写真は、ブルーインパルスという展示飛行を披露する専門のチームが使用する航空機を撮影した写真が多い。また、要約を提示した結果には、“予定では 13:00 からブルーインパルスの展示飛行の予定です。”や“式典は大勢の人で賑いましたが、ブルーインパルの展示飛行は荒天のため残念ながら中止となり地上滑走のみ行われました。”という文章があり、13 時からブルーインパルスによる演出が始

(注2) : <http://www.hamazo.tv/>

表 3 検索語“豆まき”を与えた際の要約結果

パーストしている時間	要約
10:00～11:00	鬼をやっつけ、一段落した後につどいの始まりです。今日は月曜日ですが、イレギュラーな開催ですが、「すこやか」さんも遊びに来てくれました。 つどいには、各クラスが作った、鬼のお面を被って入場です。とてもかわいかったですよ つどいの中では魔法使いさんが節分について分かりやすく教えてくれました。そしてお待ちかねの豆まきです 一応十分な数は用意したのですが、子ども達も一斉にお菓子を捨てていました。こういう時って性格出ますよね すこやかさんも豆まきに参加。「キョトン」としている子もいましたが、楽しんでくれたようです 不安な表情をしながらも体操を可愛くやっていました。足かawaiiでしょ 節分のお話を聞いて鬼のパンツの体操を元気にやりました。さあ!自分の中にいる鬼を退治するぞ!! おにはそと～!!と豆を投げるみんな…泣きそうになりながらも必死に頑張っていました。みんなとっても強かったね 鬼さんと体操…鬼さんもみんなと仲良しになりましたおまけじゅんばん…じゅんばんね 鬼さんをやっつけた後は福の神からみんなの健康と幸せを願ってお菓子まきとともたくさんお菓子をもらえて大満足そうな子どもたちでした。 そしてたんぼぼさんに新しいお友達が来ましたよまだ小さなお友達です
12:00～13:00	晩飯は、長女療宅で恵方巻 右上：ローリングペンケース 840 円 左上：折尺 294 円 右下：ルーラーカリキュレーター 735 円 左下：消しゴム 84 円 地元の「ゆるきやら」連も勢揃い。私は初めて「出世大名家康くん」を見て大喜び。 3 体いたけど、一番人気だったと思います。うなぎの「膳」もちゃんと戻ってきました。 「福はうち」だけに「福 (お客様) はうち (山基整堂)」に来てくれますように 【今日の昼食メニュー】 ・ご飯 ・お味噌汁 ・アジの青海苔フライ ・菜の花のオカカ和え ・甘辛牛蒡炊きのゴマ和え ・香の物 ・杏仁豆腐 法多山の節分について来ました 人生お初 家康くんも登場。もちろんどんまにやかさんとフィッビーも登場 大変賑やかな豆まきとなりました。福は内、福は内 豆まきの後は、恵方巻きをガブッと、、、願いを思い浮かべながら…頂きます。縁起物はやりましょう!!!
15:00～17:00	昨日は旦那さんが帰ってきてから豆まきしたよお厄年も終わったし大殺界も抜けたしこれからはきつといーこといっぱいだなあ 昨日は節分、中泉八幡宮の節分豆まきのお手伝いに行ってきました。私も一緒に年男の厄払いをいっていただけてきました。これで今年一年はいい年を越せるでしょう。 そして今日は立春! 暦のうえでは春到来というわけです。 なんだか庭の木々や草花が一気に芽吹いて花も咲き出したような気がします。 今日からカフェのスイーツも春を感じさせるケーキが登場してきますよ。お楽しみに! 今日は節分。豆まきをします～ 恵方巻きを頂きます。。 今年の恵方東北東を向いて… 一年間の無病息災・商売繁盛・世界の平和を…願って… 2 月の潮見表無料の可愛いイラストを貼りました。やっぱり、私が書くよりも 華があります! …う～ん。毎月変わり映えないような気が、…。 節分と言えば『恵方巻き』。母と一緒に作りましたちゃびは大人しく見学たくさん出来ました 海鮮巻と田舎巻? (かんぴょう・干し椎茸・牛蒡・人参連根・厚焼き卵・胡瓜) 納豆巻き (長男用) 茶碗蒸しと一緒に。 次は『いちご大福』作り 長女と次女が帰宅し一緒に…。 おかげで早く出来ました作りながら、幾つかお腹の中に 子供たちがいつのまにか、玄関に節分の折り紙を折って飾ってありました。上手にできました。 洋服部屋作ったよ!とのことで、新しい作業部屋で楽しい時間でした。いつも美味しいコーヒーありがとうございます!!
18:00～24:00	恵方巻きをお腹一杯いただきました。さあ一今から豆まきです 福は内!! 福は内!! もちろん例にもれず遣伝子組み換えてない国産豆を買い、ちゃんと豆まきしました。 今年のお姿はやや控えめですね。 今日は節分でしたね。日中は春?を思わせる陽気でした。 しかし、週後半には氷点下の予報も出て居ますので油断なさらぬ様にね。 昨日の節分には、当町内でも豆まきが行われました。 豆まきのあとは「恵方巻き」長いまま食べるのは納得いかず、プチツとちぎりながら食べていました! あの大きな恵方巻き。次男には足りなかったようです 僕は贅沢に中トロ入り～ イェイ (こーゆうとこでお金惜しみたくない。) 東北東向いての恵方巻き妻と娘は安いサラダ卷ですが… 鬼はー外福はー内 さてさて、昨日は節分。恒例の花嶋家の豆まきにたくさんのお友達が来てくれましたあ～ 年々子供たちが大きくなっていくので大人も入りきらない状態です ということで行政書士鈴木家の鬼です青い目がコワイでしょ (笑) さんざん「福はうち」を叫んだあと。。。しっかり年の数。。いやいや年の数以上に (笑) お豆を食べました～!! すると奥さんが息子のお豆を横取りしようと試みると。。。「豆を取るな」(笑) こんばんは一オザワです。節分に豆まきは昔からですが、恵方巻は、ココ最近で定着したような気がします その年に良いとされる方向を向いて食べるのが恵方巻で今年は、東北東と言われても… 東西南北が全く分からない私です。。 今日のイチオシ☆☆☆ ナットクラッカー。その名の通り、ナットを砕くための道具です 人気のこちら、フリーズ&リリーズで冷やしても、バーナーなどで熱しても、、、 サビ付いてしまったり、固着してしまって外れないナットの最終手段砕くが素早くできます この丸の中にナットをセットしたら、お尻部分を締めていくだけ これで刃がナットを砕いてくれます また、砕いたナットの飛び散りを最小限にしてくれるので安全です 大きさは、9～12mm 用、1 2～1 6 mm 用とあります もう、どうしようもない!!って状態に陥った時の最終手段にどうぞもうひとつ プラリペア こちらは、破損してしまったプラスチックの応急処置のための道具です 同じ用に形を作って直すんです破損してしまったけど、取り換え用の部品が無い時に役立ちますよ 使い方も簡単です どちらも、できれば順番があって欲しくは、ないけれど、いざという時にあると便利な 2 点ですよ 工具店でどうぞ工具は、通販をしておりませんので店頭にておもとめください。 facebook もやってます 「いいね」お願いします アートツール HP よりミニカーの通販も開始しました↓↓ http://www.arttool.jp/

まったが、悪天候により飛行はしなかったということが分かる。

一方、07:00 から 12:00 の間に撮影された写真は、ブルーインパルスが使用する航空機以外の様々な航空機を撮影した写真が多い。要約の提示結果には、RF-4E、CH-47J、SH-60J、U-4 や C-1 のような航空機の名称が含まれていることから様々な航空機の展示が行われていたことが分かる。また、“昨夜から降り続いた雨も朝には上がり”という文章から、イベントの序盤には雨が降っていなかったということが分かる。

また、記事の統合の結果として、エア・フェスタに参加する

予定であったが、悪天候のため予定を変更して別のイベントに参加したという内容のブログ記事が存在した。そのような記事を統合してしまった影響により、要約を作成した結果として、エア・フェスタに関係がないと考えられる文章が含まれてしまった。

次に、データセット 2 に対して、期間として 2014 年 2 月 2 日から 2014 日 2 月 4 日を与え、キーワードとして“豆まき”を与え、提案システムの記事統合を行った結果の一部を図 3 に示す。また、2 月 3 日に検出したイベントの要約結果を表 3 に

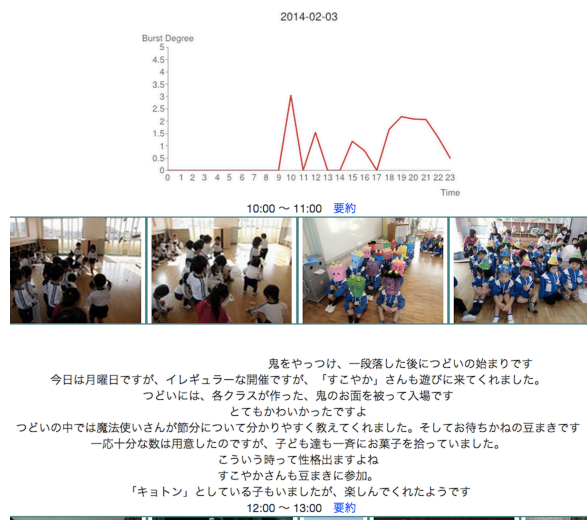


図 3 システムの動作例（検索語：“豆まき”）

示す。図 3 中のグラフから、2 月 3 日に検出したバーストには、10:00 から 11:00、12:00 から 13:00、15:00 から 17:00、18:00 から 24:00 の 4 つの盛り上がりが存在したことが分かる。

10:00 から 11:00 の要約を提示した結果から、子供たちが豆まきをしている様子が分かる。それ以降の時間は、正午や夕方のような食事を取る時間であることから、要約を提示した結果として、豆まきだけでなく恵方巻きを作成して食べたという内容の文章も存在する。また、12:00 から 13:00 の要約を提示した結果には、“法多山の節分について来ました”という文章が含まれていることから、法多山で節分のイベントが行われていたことが分かる。要約において言及されている法多山は、静岡県袋井市にある法多山尊永寺という寺院のことを示しており、初詣や節分のような季節の催しの際、多くの人が訪れる場所である。このように、イベントの様子を示す文章が存在する一方、12:00 から 13:00 の要約に含まれる“ルーラーカリキュレーター 735 円”や“消しゴム 84 円”、18:00 から 24:00 の要約に含まれる“今日のイチオシ☆☆☆ ナットクラッカー。その名の通り、ナットを砕くための道具です”のようなイベントに関係がないと考えられる文章が含まれてしまった。

データセット 1 やデータセット 2 に対する要約を提示した結果として、イベントに関係がないと考えられる文章が含まれている。その原因の 1 つとして、ブログとマイクロブログの統合を行う際、ユーザが与えたキーワードのみを文章中に含んでいるかということを考慮してブログ記事を取得しているためであると考えられる。そのため、統合するブログ記事を取得する際、イベントに関する複数のキーワードが含まれているかどうかを考慮し、イベントに参加していないと考えられる記事は統合しないようにする必要がある。イベントに関する複数のキーワードを決定する手段として、TF-IDF などの特徴量を用いて、バーストしている時間中のツイートから特徴的な単語を抽出し、それをキーワードとして扱うことが考えられる。

5. おわりに

本研究は、ブログとマイクロブログを統合することで、2 つ

のメディアの特徴を生かし、イベントの発見とそのイベント内の盛り上がり毎の要約をユーザに提示するシステムを提案した。提案システムの動作結果から、写真の撮影時刻を利用する事で、マイクロブログからバースト検出を行うだけでは困難な実際のイベントの発生期間を推定できることを示し。また、バーストの盛り上がり毎に要約や写真を提示することで、イベント内の盛り上がりの様子を知ることができた。

今後の課題として、記事を統合する際、イベントに関係しない内容を多く含む記事 (e.g. “悪天候のため、別のイベントに参加した”) を除去する手法の検討があげられる。解決方法として、マイクロブログからイベントに関するキーワードを抽出し、そのキーワードがブログ記事に含まれているかどうかを考慮することや統合する記事が同様のトピックに属しているかを考慮することが考えられる。

文 献

- [1] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web, WWW '10*, pages 851–860, New York, NY, USA, 2010. ACM.
- [2] Chenliang Li, Aixin Sun, and Anwitaman Datta. Twevent: segment-based event detection from tweets. In *Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM '12*, pages 155–164, New York, NY, USA, 2012. ACM.
- [3] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: Understanding microblogging usage and communities. *Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, August 2007.
- [4] 坂本 翼, 横山 昌平, 福田 直樹, 石川 博. マイクロブログを対象としたリアルタイムな要約生成システムの試作. In *第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM 2011)*, 2011.
- [5] Chao Shen, Fei Liu, Fuliang Weng, and Tao Li. A participant-based approach for event summarization using twitter streams. In *Proceedings of NAACL-HLT*, pages 1152–1162, 2013.
- [6] Manos Tsagkias, Maarten de Rijke, and Wouter Weerkamp. Linking online news and social media. In *Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11*, pages 565–574, New York, NY, USA, 2011. ACM.
- [7] Marc Bron, Bouke Huurnink, and Maarten de Rijke. Linking archives using document enrichment and term selection. In *Proceedings of the 15th international conference on Theory and practice of digital libraries: research and advanced technology for digital libraries, TPD L'11*, pages 360–371, Berlin, Heidelberg, 2011. Springer-Verlag.
- [8] 蝦名 亮平, 中村 健二, 小柳 滋. リアルタイムバースト解析手法の提案. *情報処理学会論文誌データベース (TOD)*, 5(3):86–96, sep 2012.