

(21) リアルタイムに交通現象を獲得するための キーワードの選定方法の検討

中村 健二¹・田中 成典²・藤本 雄紀³・外山 諒⁴

¹正会員 大阪経済大学准教授 情報社会学部 (〒533-8533 大阪市東淀川区大隅2-2-8)

E-mail: k-nakamu@osaka-ue.ac.jp

²正会員 関西大学教授 総合情報学部 (〒569-1095 大阪府高槻市霊仙寺町2-1-1)

E-mail: tanaka@res.kutc.kansai-u.ac.jp

³非会員 関西大学大学院 総合情報学研究科 (〒569-1095 大阪府高槻市霊仙寺町2-1-1) /
国土交通省国土技術政策総合研究所 社会資本マネジメント研究センター 社会資本情報基盤研究室

E-mail: k398881@kansai-u.ac.jp

⁴非会員 関西大学大学院 総合情報学研究科 (〒569-1095 大阪府高槻市霊仙寺町2-1-1)

E-mail: k943166@kansai-u.ac.jp

我が国の道路網では事故などの交通現象が日々発生している。これらの監視には、監視カメラやトラフィックカウンターなどが用いられているが、定点監視であるため監視範囲には限界がある。そこで、SNSの一種であるマイクロブログを解析し、交通現象に係わる情報取得を補完する方法が提案されている。しかし、あらかじめ設定したキーワードで探索するため、取得できるツイート数に多寡が発生する。

そのため、本研究では、交通現象発生時に関連するキーワードがバーストする特性を用いて、交通現象の獲得に有用なキーワードを選定する手法を提案し、その有効性を評価する。

Key Words: burst, real-time, microblog, traffic behavior, information extraction

1. はじめに

我が国に張り巡らされた道路網は、国内の人流、物流を支える要であり、人々の生活の基礎となっている。このような状況下において、各道路の状況把握は、道路を適切に管理する上で重要な役割を担っている。そのため、道路上の各所において、監視カメラやトラフィックカウンターが設置されており、事故などの事象（以下、交通現象）の発生を常時把握する仕組みが実現されている。しかし、これらの監視手法は定点監視であり、監視範囲も制限されていることから、交通現象を把握する他の方策と併用することが望まれている。

交通現象の把握に係わる新たな取り組みの一つとして、近年、広く普及した SNS (Social Networking Service) の一種であるマイクロブログを解析する手法^{1,2)}が提案されている。マイクロブログの Twitter では、日々膨大なつぶやき（以下、ツイート）がインターネット上に投稿されている。これらのツイートの中には、交通現象や道路利用に係わる内容も多く含まれているが、その情報を獲

得するためには、膨大なツイートを解析する必要がある。既存手法^{3,4)}では、有用な情報のみをツイート群から選択的に取得するため、交通現象に係わるキーワードをあらかじめ設定している。しかし、キーワードを設定した場合、次に示す2つの課題が発生する。

一つ目は、出現数の少ない未登録のキーワードへ対応できない課題である。交通現象は、「ぶつかりたてホヤホヤ」などの多様な語句で表現される。そのため、あらかじめキーワードを設定した場合、交通現象を漏れなく抽出することが難しい。

二つ目は、選定したキーワードが恒常的にツイートに含まれる課題である。例えば、「事故」や「渋滞」などのキーワードは、交通現象に係わるツイートにも含まれるが、ニュースなどでも使用され、取得対象となるツイート数が膨大になり、ノイズも多く含まれる。

そこで、本研究では、道路に係わるツイートから交通現象に関するリアルタイム性の高いツイートを獲得するためのキーワードの選定方法を提案する。

2. 研究の概要

(1) 課題への対応方策

本研究では、交通現象の発生に際し、関連する内容を含むツイート数が増加する特性があることに着目した。そこで、「出現数の少ない未登録のキーワードへ対応できない課題」と「選定したキーワードが恒常的にツイートに含まれる課題」に対して、普段の出現数とは異なる傾向を示すバーストの概念を適用することで解決を図る。

バースト検出手法^{5,6)}では、出現数が平常時と比較して急激に変化するキーワードを発見できる。そのため、出現数が恒常的に多いもしくは少ないキーワードであった場合も、交通現象発生時に出現数が増加する傾向であれば、適切に検出できる。そのため、2つの課題を解消可能であると考えられる。

(2) 研究内容

本研究では、交通現象を獲得するために有効なキーワード選定手法として、バースト検出手法を用いることを提案する。バースト検出手法には、蓄積された時系列データからバースト期間を検出する手法⁵⁾やリアルタイムに時系列データを解析してバーストの有無を検出する手法⁶⁾（以下、リアルタイムバースト手法）がある。本研究では、交通現象をリアルタイムに把握することから、キーワードの選定にリアルタイムバースト手法を用いる。

(3) 処理の流れ

提案手法の手順を図-1に示す。本手法は、交通現象発生後のツイート抽出機能および交通現象の獲得に有効なキーワード抽出機能から構成される。

交通現象発生後のツイート抽出機能では、道路に関する内容を含むツイート群から、交通現象に係わるツイートを取得し、そこに含まれるキーワードを蓄積する。まず、道路に関する内容を含むツイート群から、交通現象の発生後一定時間以内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形態素解析し、交通現象に係わる語句としてデータベースに登録する。

交通現象の獲得に有効なキーワード抽出機能では、交通現象に係わる語句から、交通現象発生後にバーストするキーワードを抽出する。バースト度合い算出処理では、各キーワードが含まれるツイートとその投稿時間を用いて、リアルタイムバースト検出を行い、バースト度合いを算出する。次に、バースト判定処理では、算出したバースト度合いが閾値を超えているかを判定する。そして、交通現象の獲得に有効なキーワード選定処理では、交通現象の発生時刻から一定時間以内にバーストが発生した回数の多いキーワードを選定する。ここで、バースト回数が一定数以上のものは、ノイズとして除去する。

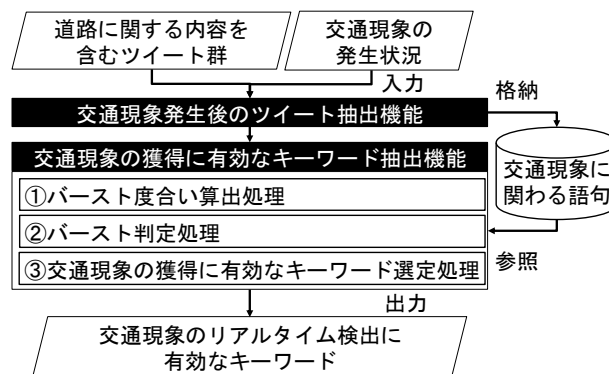


図-1 提案手法の手順

本処理の詳細は既存研究⁶⁾を参照されたい。

3. 抽出キーワードの評価実験

(1) 実験計画

本研究では、既存手法における2つの課題を解消できているかを評価するため、「実験1：選定キーワードの比較実験」と「実験2：バースト発生時間と交通現象発生時間の比較実験」を行う。一つ目の実験では、既存手法と提案手法で選定されたキーワードを比較し、提案手法で新たに選定されたキーワードの件数を評価する。二つ目の実験では、提案手法で得られたキーワードと交通現象との関係を分析し、選定したキーワードを用いて交通現象を高精度に獲得可能であることを検証する。

(2) 実験データの収集

a) 交通現象の収集

交通現象発生情報は、公益財団法人日本道路交通情報センター（以下、JARTIC）の公開する規制情報を利用する。JARTICのWebサイトでは、路線名称、方向、規制区間、原因、規制内容などの情報が公開されており、それらの情報を5分間隔で収集する。今回は試験的に2015年1月24日から2015年12月3日の間に公開された近畿圏の規制情報を対象とした。5分間隔で情報を取得するため、5分以上継続公開されている交通現象は重複して取得される。そのため、取得した内容が同一の場合は一つの交通現象として統合して登録する。その結果、1,662,888件の交通現象を収集した。

b) ツイートの収集

道路に係わる内容を含むツイート群は、Twitter社提供のAPIを用いて収集する。まず、収集の際に必要な検索キーワード集（以下、地名辞典）を作成する。一般に交通現象に関する情報を含むツイートには、路線名、サービスエリア名やインターチェンジ名などの単語が含まれていることが多い。また、これらの単語には略称や通称などの類義語もよく利用されているため、正式

表-1 各手法で選定できたキーワード（各上位 30 件）

順位	既存手法 A	既存手法 B	提案手法 C	提案手法 D	順位	既存手法 A	既存手法 B	提案手法 C	提案手法 D
1	高速	阪神	燃え	ボンネット	16	号	カード	キャラ	ラジオ
2	阪神	高速	不正	ほとんど	17	現在	渋滞	ニュース	クルーズ
3	線	線	やっぱり	俊三	18	交通	現在	開始	超え
4	し	時	スタート	エンゼル	19	環状	道路	もの	淡路島
5	時	神戸	伊勢	リボン	20	車	系統	下さい	向かつ
6	情報	情報	オンライン	かかつ	21	湾岸	日	なんば	スカイライン
7	道	湾岸	ネットワーク	一橋大	22	あり	重大	交通	ウイング
8	大阪	分	コレクション	クレジット	23	通行止め	堺	時間	もらえ
9	神戸	し	公園	出せ	24	てる	月	言っ	アクセス
10	分	都市	豊中	地図	25	カード	ところ	更新	天保山
11	日	大阪	名古屋	摩耶	26	さ	環状	明日	ブルー
12	道路	道	トレーラー	パーキング	27	れ	只今	車両	須磨
13	渋滞	首都	マスコット	もらい	28	する	号	阪和線	西大阪線
14	月	通行止め	山口	おばさん	29	首都	あり	画像	周回
15	都市	交通	わから	ビジネス	30	市	等	阪和	全く

名称だけでなく、それらを含めて 971 件を地名辞典へ登録した。次に、Twitter 社提供の Search API に地名辞典のキーワードを指定し、2014 年 8 月 22 日から 2016 年 6 月 09 日の期間を対象にツイートを集めた。その結果、3,793,301 件のツイートを収集した。

(3) 実験条件

本実験では、交通現象の中でも特に関連する内容を含むツイート数が増加する事故を対象とする。そして、ツイートは事故発生から 2 時間以内に投稿されたものを対象とした。

また、リアルタイムバースト手法のパラメータを $N = 50$, $\beta = 0.4$, $W_{min} = 1$ 日, $C_{min} = 10$, $A_{min} = 10$, $W_{max} = 1$ 日に設定し、実験を行った。パラメータの詳細は既存研究⁹⁾を参照されたい。

(4) 実験 1: 選定キーワードの比較実験

a) 実験内容

本実験では、既存手法と提案手法で選定されたキーワードを比較し、提案手法で新たに選定されたキーワードの件数に基づき評価する。本実験では、キーワード選定手法として、次に示す 4 つの手法を用いる。

- 出現数が高い単語を用いる手法（既存手法 A）
- TF-IDF 値が高い単語を用いる手法（既存手法 B）
- 交通現象の発生時刻から一定時間以内にバーストが発生した回数が多い単語を用いる手法（提案手法 C）
- 提案手法の中でも誤判定率の低い単語を用いる手法（提案手法 D）

b) 結果と考察

各手法にて選定されたキーワード上位 30 件を表-1 に示す。表-1 において、手法 B から D のキーワードで既

存手法 A でも選定されているものは黒塗りにしている。表-1 の結果を確認すると、既存手法 B の場合は一致率 80%であるのに対して、提案手法 C, D はそれぞれ 3%, 0%となった。このことから、既存手法の課題である「出現数の少ない未登録のキーワードへ対応できない課題」について解消できていることがわかる。

(5) 実験 2: バースト発生時間と交通現象発生時間の比較実験

a) 実験内容

本実験では、実験 1 の 4 つの手法で選定したキーワードが、実際の交通現象の獲得に有効であるかを評価する。まず、関連する内容を含むツイート数が増加した事故を抽出する。次に、既存手法 A, B および提案手法 C, D のキーワードの上位 5 件のいずれかが含まれる事故を抽出する。そして、その事故の発生した日の該当キーワードが含まれるツイート件数を 1 時間ごとに集計し、その結果をグラフとして可視化する。可視化した事故発生当日のツイート数の推移から交通現象をリアルタイムに獲得するためのキーワードとして有効であるか検討する。

実験対象として、事故に係わるツイートが多くみられた 2015 年 6 月 24 日の 19 時 40 分頃に阪神高速道路 13 号東大阪線で発生した車両火災事故⁷⁾を設定した。

b) 結果と考察

本実験の結果を図-2 に示す。図-2 のグラフは、事故発生当日の該当キーワードが含まれるツイート件数の推移を示している。

実験結果（図-2）を確認すると、既存手法で選定したキーワードを含むツイートは恒常的に投稿されていることがわかる。また、提案手法 D で選定したキーワードを含む投稿はなされていなかった。それに対し、提案手法 C のキーワードは、事故発生時刻に 1 件のツイートが

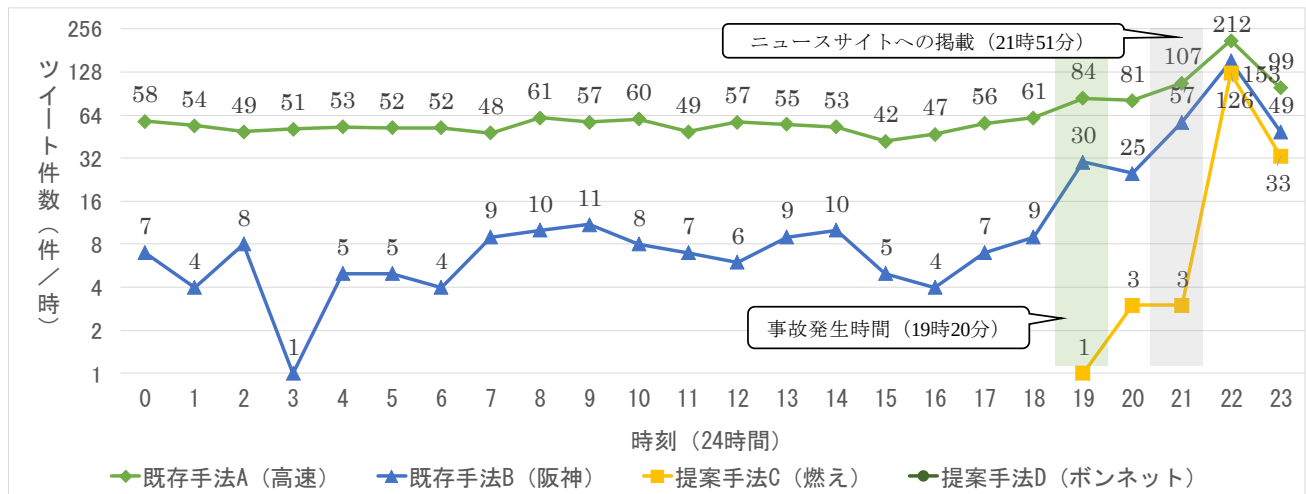


図-2 各手法で選定したキーワードを含むツイート件数と事故との関係の例

あり、その後、ツイート数が増加する傾向となっている。また、提案手法Cで選定したツイート166件のうち、実際の事故に関するツイートは164件であり、高精度に事故に係わる内容を取得できていることが分かった。このことから、「選定したキーワードが恒常的にツイートに含まれる課題」を解消し、交通現象の獲得に有効なキーワードを抽出できていることが明らかとなった。

4. おわりに

本研究では、リアルタイムに交通現象を獲得するためのキーワード選定を目的として、バースト検出手法を用いた交通現象に関連性の高いキーワードの抽出手法を提案した。評価実験を通して、交通現象に係わるキーワードをあらかじめ設定した際の課題である「出現数の少ない未登録のキーワードへ対応できない課題」と「選定したキーワードが恒常的にツイートに含まれる課題」を解消できていることを確認した。このことから、提案手法にて選定したキーワードは、リアルタイムな交通現象の獲得に有用であることがわかった。

本研究を通して、次に示す2つの課題が明らかとなった。一つ目は、Twitterに投稿するキーワードは、投稿者に依存する点である。今回の検証では、投稿者を考慮せずにすべてのツイートを解析したため、不要な語句も散見された。この課題に対しては、投稿者を考慮して解析することで対応する予定である。二つ目は、提案手法により選定したキーワードがリアルタイム監視に有用であるかを定量的に評価する必要がある点である。本研究では、交通現象の抽出に有用であるかどうかを評価したが、交通現象の発生直後に即座に反応したかどうかを評価できていない。この課題へは、長期間、実際にリアルタイムに投稿をモニタリングすることで対応する。

今後は、これらの課題へ対応するとともに、本提案手

法の土木分野以外での活用方法を検討する予定である。

謝辞：本研究はJSPS科研費JP15K00163の助成を受けたものです。

参考文献

- 1) 今井龍一、高橋哲朗、田嶋聡司、山影譲、重高浩一：道路事業に係わる行政相談資料及びTwitterのつぶやきに対するテキストマイニング技術の適用～道路事業評価の高度化支援に向けた一考察～、土木計画学研究発表会・講演集、土木学会、Vol.48、2013.
- 2) 今井龍一、中村健二、田中成典、藤本雄紀：マイクロブログを用いた道路事業に係わる意見の抽出と位置情報推定に関する研究、土木学会論文集F3、土木学会、Vol.71、No.2、pp.188-195、2015.
- 3) 伊達友裕、井上博之：マイクロブログを利用したローカルコンテンツの選択のためのキーワード抽出機構、マルチメディア、分散協調とモバイルシンポジウム2014論文集、情報処理学会、pp.1667-1673、2014.
- 4) 邱起仁、樫山淳雄：ニュース記事に関連するTwitterの投稿の収集手法の提案、研究報告データベースシステム、情報処理学会、Vol.22、pp.1-6、2013.
- 5) Kleinberg, J.: Bursty and Hierarchical Structure in Streams, Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.91-101、2002.
- 6) 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、情報処理学会、Vol.5、No.3、pp.86-96、2012.
- 7) 産経WEST：阪神高速で車炎上 車の調子悪くなり「停車したら燃えだした」 東大阪線が一部通行止め、入手先<<http://www.sankei.com/west/news/150624/wst1506240096-n1.html>>、(入手2016.6.25)。