

(83) 交通現象に係わるマイクロブログの投稿記事の信頼性分析

藤本 雄紀¹・今井 龍一²・中村 健二³・田中 成典⁴・
有馬 伸広⁵・荒川 貴之⁶

¹学生会員 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

²正会員 東京都市大学准教授 工学部 (〒158-8557 東京都世田谷区玉堤1-28-1)
E-mail: imair@tcu.ac.jp

³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

⁴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

⁵正会員 阪神高速技研株式会社 システム事業本部 (〒550-0011 大阪市西区阿波座1-3-15)
E-mail: nobuhiro-arima@hanshin-tech.co.jp

⁶非会員 阪神高速技研株式会社 システム事業本部 (〒550-0011 大阪市西区阿波座1-3-15)
E-mail: takayuki-arakawa@hanshin-tech.co.jp

道路交通分析において、道路上に設置された様々な機器のデータを用いた定量的な分析に加えて、マイクロブログから道路利用者の意見を抽出する定性的な分析が行えるようになってきた。しかし、マイクロブログは誰もが自由に投稿できるため、各投稿の信頼性が不明である。

本研究では、信頼性の観点からTwitterのツイート内容を分析し、道路管理で有用なツイートを抽出するための信頼性分析手法を考案し、その有用性を検証した。

Key Words : *microblog, traffic phenomena, reliability, text mining*

1. はじめに

道路交通分析では、主に事故や渋滞などの交通現象の発生要因の特定、道路整備後の効果測定といった分析が行われている¹²⁾。これらの分析では、道路上に設置された監視カメラの映像や走行車両に取り付けられた機器から得られるプローブデータなどの定量的なデータを用いられることが多い。一方、SNS (Social Networking Service) の一つであるマイクロブログのTwitterに注目すると、膨大な量のつぶやき (以下「ツイート」という。) がインターネット上に日々投稿され続けている。これらのツイートの中には、道路の利用時の所感や交通現象に係わる内容が写真や動画などと共に投稿されている。この特性に着目し、ツイートの情報を解析して道路交通分析への活用を検討する既存研究がある³⁴⁾。その一例として、ツ

イートを解析して道路利用者の移動目的を抽出する研究³⁾や、鉄道の運転状況をツイートから検知する研究⁴⁾がある。これらの研究成果を活用することで、道路利用に係わる所感などの主観的な情報を取得できる。しかしながら、マイクロブログは誰でも自由に投稿できるため、必ずしも正確な情報のみが投稿される保証はない。さらに、交通現象が発生すると、一時的に大量のツイートが投稿されるため、道路管理者はツイートを一様に確認し、正確な情報を判別する必要がある。これは、道路管理へのマイクロブログの活用の弊害となり、リアルタイムに情報を獲得できない要因の一つとなる。

そこで本研究の目的を道路利用者のツイートの信頼性を分析することとした。この結果を活用すると、交通現象が発生した際の情報の取捨選択ができるなど、ツイート分析の高度化が期待できる。本研究では、投稿者の日

表-1 投稿者のタイムラインの集計結果

投稿者	阪神高速道路に関するツイート数 (件)	居住地	近畿圏の位置情報 の付与数 (件)
投稿者 1	1	不明	9
投稿者 2	4	不明	4
投稿者 3	3	兵庫県	1
投稿者 4	1	大阪府	1
投稿者 5	1	不明	0

々の投稿したツイートを分析した結果から得られる特定道路の話題への反応度合いと投稿者の位置情報とを用いて投稿の信頼性を評価する手法を提案する。

なお、本研究におけるツイートの取扱いは、Twitter社のプライバシーポリシー⁵⁾に準じているため、個人情報やプライバシーには抵触しない。

2. 阪神高速道路に係わるツイートの収集

本研究では、交通現象に係わるツイートを収集し、その内容の信頼性を分析する。また、ツイート内容だけでなく投稿者自身の属性も分析し、交通現象に係わるツイートの傾向から、信頼性の手掛かりとなるデータを明らかにする。

(1) 収集手順

本研究では、2015年9月から近畿圏の一般道路名、高速道路名及びその路線名、SA・PA名などを含むツイートをTwitter社提供のSearchAPI⁶⁾により継続して収集している³⁾。今回の分析では「阪神高速」またはその略語である「阪高」の語句を含むツイートを用いる。ツイートの収集期間は2016年6月13日から6月19日の7日間とした。収集手順は、まず阪神高速道路上で発生した交通現象に対するツイートから、事象発生時の状況などを目視にて確認する。次に、外部ニュースサイトなどからその事象に関する内容を参照し、ツイートの内容と比較する。最後に、これらのツイートの投稿者の情報と、過去1か月分のツイートのログ（以下「タイムライン」）とを確認し投稿者の傾向を分析する。

(2) ツイートの収集結果

ツイートの収集結果を図-1に示す。投稿数は2017年6月18日が最も多いが、阪神高速道路との関連性の低い投稿が取得されていた。17日の阪神淡路大震災に関するツイートに次いで多い2017年6月16日に着目すると、阪神高速道路池田線にて発生したトラックと軽乗用車の衝突

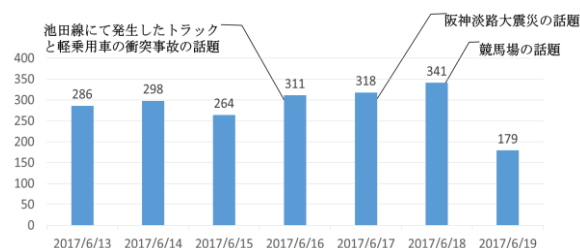


図-1 2017/6/13～2017/6/19のツイートの投稿数と話題

事故⁷⁾に係わるツイートが多く見られた。この事象に係わるツイートの投稿者は49名でツイート数は52件（リツイートは39件）であった。これらのツイートの投稿者のうち、無作為に5名を選出してタイムラインを集計した結果を表-1に示す。表中の居住地は投稿者自身の情報から把握できた地名である。また、近畿圏の位置情報の付与数はGeoタグまたは文章中の地名が近畿圏内のツイートを集計した。この結果から、事象に反応する投稿者には様々な傾向があることがわかった。

(3) ツイートの分析結果

収集したツイートを分析した結果、次の2点の知見が得られた。1点目は交通現象に係わるツイートには重要性の高いものと低いものとが含まれていることである。今回の事象に係わるツイートの内容を大別すると、事象の状況を報告するツイートと、事象の発生を拡散するためのニュース記事や他人のツイートを再度投稿したツイートの2通りが見られた。このうち、前者は事象の発生直後から投稿される傾向がある。また、状況を把握できる写真などが付与されて投稿されている。そのため、これらのツイートは、道路交通分析において有用な情報が含まれているといえる。一方で後者は、事象の発生直後から1～2日程度投稿され続けるが、ニュースの内容をそのまま引用しているため、既知の情報しか含まれていない。しかし、これらのツイートのうち、事故発生直後に投稿されたツイートは情報を拡散させる役割を果たしているため、重要性はあると言える。そのため、事象に係わるツイートを一様に確認するのではなく、重要性の高いツイートのみを抽出できる必要がある。

2点目は投稿者の属性が異なることである。表-1の結果を確認すると、投稿者によって、阪神高速道路に関するツイート数や、近畿圏の位置情報に関するツイート数が異なることがわかる。また、1点目で挙げた事象の状況を報告するツイートの投稿者は、特定道路のツイートや位置情報付きツイートの数が比較的多い傾向がみられた。居住地は、投稿者自身の情報として登録しているユーザは少ないが、投稿の傾向から都道府県の粒度で居住地を推定できる可能性がある。

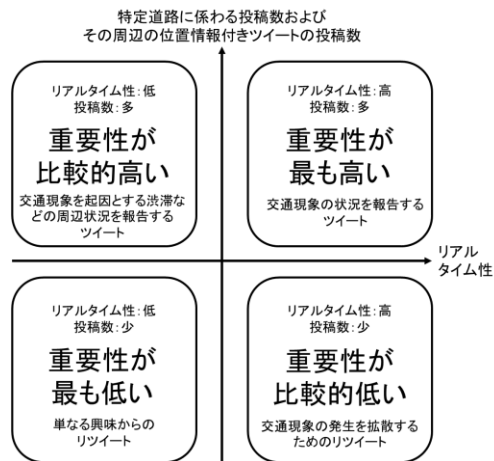


図-2 特定道路に係わる投稿数とリアルタイム性の関係

以上の2点から、事象に関するツイートの傾向には図-2に示すような傾向があることがわかった。右上の投稿者は、交通現象の状況を投稿する可能性が最も高く、重要性が高いといえる。左上の投稿者は、交通現象を起因とする渋滞などの情報を提供する場合があり、重要性は比較的高い。右下の投稿者は、右上の投稿者が投稿したツイートを拡散目的でリツイートなどを行うため、情報の新規性はないが交通現象の検知などに活用できる。どちらの尺度も低い左下の投稿者は、単に興味などから交通現象に関するツイートをリツイートするため、重要性が最も低い。これらの関係性から、最も重要性の高い右上の投稿者を抽出する手法を次章にて詳説する。

3. ツイートの信頼性分析手法の考案

(1) 分析手法の概要

本研究では、事象に関するツイートの投稿者のタイムラインを時系列に分析することで、ツイートの信頼性を分析する手法（以下「本手法」という。）を考案した。図-3に示すとおり、本手法は投稿者単位でツイートを収集して信頼性を分析する。図-3の網掛け箇所は既存研究³⁾と同様の処理を指す。本手法は、道路利用者のツイート抽出機能および信頼性分析機能から構成される。

(2) 道路利用者のツイート抽出機能

本機能では、道路利用者が投稿したと考えられるツイートを抽出する。まず、既存研究³⁾と同様に事前に作成した地名辞典を利用して対象道路のツイートを収集する。次に、道路交通情報から得た事象に係わる単語を指定して交通現象に係わるツイートを抽出する。そして、そのツイートの各投稿者の過去のタイムラインを収集して道

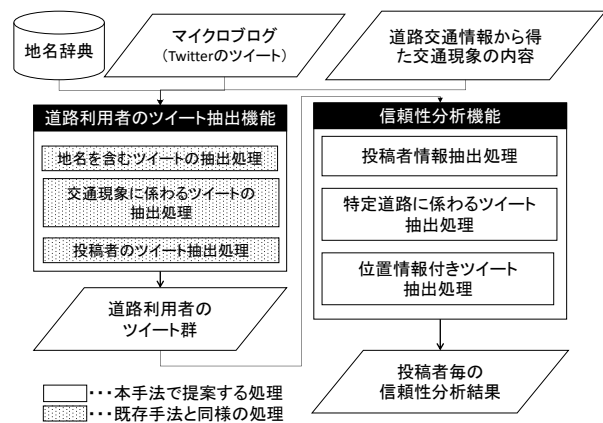


図-3 本手法の手順

路利用者のツイート群として出力する。

(3) 信頼性分析機能

本機能では、前述の機能にて収集した道路利用者のツイートを分析して信頼性を分析する。投稿者情報抽出処理では、投稿者の居住地をAPIを用いて取得する。本研究では、兵庫、大阪、京都、和歌山および奈良の各府県居住者を阪神高速道路周辺とした。また、居住地は投稿者自身が自由に記述できるため、存在しない地名である場合は居住地を不明として設定する。特定道路に係わるツイート抽出処理では、投稿者のタイムラインに対し、特定道路の単語が含まれているツイートを抽出する。位置情報ツイート抽出処理では、Geoタグ付きツイートや形態素解析を経て「地域」として判定された単語を含むツイートから、阪神高速道路周辺の位置情報付きツイートを抽出する。以上の処理に合致するものが多い投稿者を信頼性の高い投稿者としてラベル付けする。

4. 本手法の有用性の検証

(1) 実験内容

本章では、道路利用者のツイートの信頼性の分析可能性を本手法を適用して検証する。検証は、阪神高速道路で発生した交通現象に対し、各投稿者の信頼性の分析結果を集計する。そして、信頼性が高い投稿者のツイートを確認し、重要性の高低を評価する。検証対象は、Twitter上でも話題となった2017年6月8日に発生した阪神高速道路上で豚19匹が脱走した交通現象⁸⁾とする。検証データは、第2章で収集したツイートのうち阪神高速道路に関するツイートが多い投稿者50名と、新規に検証対象の交通現象に反応した投稿者56名を追加した計106名および、2017年6月に投稿された対象者のタイムライン43,204件とした。

表-2 信頼性分析の実験結果

投稿者	阪神高速道路に関するツイート数 (件)	居住地	近畿圏の位置情報の付与数 (件)	ツイート内容	投稿時間
投稿者 1	57	奈良県	156	RT 阪神高速に豚が放たれた様子です	2017/6/8 11:00
投稿者 2	5	不明	170	RT 事故のトラックから豚 1 9 頭逃走大阪の阪神高速	2017/6/8 20:11
投稿者 3	81	大阪府	39	RT 阪神高速池田線、豚がいますよ	2017/6/19 20:14
投稿者 4	3	不明	59	RT 【阪神高速 豚 19 頭が逃げ出す】 大阪府の阪神高速で「豚が逃げている」と料金所職員から連絡。	2017/6/8 16:40
投稿者 5	3	不明	32	RT 阪神高速に豚が放たれた様子です	2017/6/8 10:34
投稿者 6	4	不明	6	RT 『阪神高速ブタがたくさんあるいてる』が投資家の間で話題に。	2017/6/8 11:04
投稿者 7	3	兵庫県	5	阪神高速に豚おんねんて笑見に行こ	2017/6/8 12:28
投稿者 8	5	不明	0	RT 危ないけど超可愛い！？阪神高速で脱走したブタにネットがざわついた	2017/6/8 14:00
投稿者 9	2	兵庫県	2	阪神高速池田線、豚がいますよ	2017/6/8 10:27
投稿者 10	2	大阪府	0	RT 阪神高速に豚が放たれた様子です	2017/6/9 16:15

(2) 検証結果

本検証で分析した信頼性の高い投稿者10名を集計した結果を表-2に示す。表-2を確認すると、多くのツイートがリツイートである。そのうち、投稿者7や投稿者9のように交通現象発生 of 早い段階で写真付きで状況を報告するツイートを投稿している場合もあり、重要性の高いツイートの投稿者の抽出ができていたことがわかった。

(3) 考察

本手法を用いて交通事象のツイートの信頼性を確認することで、道路管理者にとって有用な情報を抽出できる可能性があることがわかった。しかし、本手法から得られた最も信頼性の高い投稿者のツイートを検証すると、リツイートで重要性の低いツイートも含まれていることがあった。これは、自ら発信をせずリツイートのみを目的としている投稿者であったためである。また、本手法は過去のタイムラインに基づいて分析するため、これまでほとんど投稿していなかった投稿者が重要性の高いツイートを投稿した場合には抽出ができない。そのため、今後は、タイムラインだけでなく、投稿者の投稿形態や嗜好性なども考慮して信頼性を分析する必要がある。

5. おわりに

本研究では、まず、交通現象に係わるツイートを収集し、その内容の信頼性を分析した。その分析結果に基づいて、ツイートの過去のタイムラインや属性に着目して信頼性を分析する手法を考案した。本手法を適用した検証結果から、ツイートを分析すると道路管理者にとって

有用なツイートを抽出できる可能性を確認した。

今後は、さらに詳細な投稿者の属性を分析した上で信頼性を分析する手法を研究し、分析精度の向上を図る。

謝辞：本研究の遂行にあたり、阪神高速道路（株）の岡山真人氏から貴重なご意見を賜った。ここに記して感謝の意を表する。

参考文献

- 1) 金進英, 宇野巧, 岩里泰幸, 大藤武彦: 量的・質的データを融合した事故要因分析による新たな交通事故分類と安全対策の方向性, 交通工学論文集, 交通工学研究会, Vol.3, No.4, pp.A_92-A_101, 2017.
- 2) 吉村仁志, 福山祥代, 山田康右, 加納英明, 鹿野島秀行, 牧野浩志: ETC2.0 対応カーナビを用いた車線利用適正化情報提供サービスの導入効果評価, 交通工学研究発表会論文集, 交通工学研究会, Vol.1, No.26, pp.333-337, 2016.
- 3) 藤本雄紀, 今井龍一, 中村健二, 田中成典, 有馬伸広, 荒川 貴之: 道路利用者のマイクロブログ分析の可能性評価, 土木情報学シンポジウム論文集, 土木学会, Vol.41, No.1, pp.79-82, 2016.
- 4) 堀江岳, 関谷聡大, 金子雄一郎: 公式 Twitter に基づく都市鉄道の運転見合わせの発生状況と情報提供の実態分析, Vol.2, No.71, pp.II_28-II_33, 2015.
- 5) Twitter, Inc.: Twitter Privacy Policy, <https://twitter.com/privacy?lang=en>, (入手 2017.6.28).
- 6) Twitter, Inc.: Twitter Developers, <https://dev.twitter.com/>, (入手 2017.6.28).
- 7) 産経 WEST: 衝突事故で 1 人心肺停止 大阪・豊中の阪神高速, <http://www.sankei.com/west/news/170616/wst1706160047-n1.html>, (入手 2017.6.28).
- 8) 毎日新聞: 接触事故で豚 1 9 匹逃走 通行止めに, <https://mainichi.jp/articles/20170608/k00/00e/040/281000c>, (入手 2017.6.28).