

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～

今井 龍一¹・高橋 哲朗²・田嶋 聡司³・山影 譲²・重高 浩一¹

¹正会員 国土交通省 国土技術政策総合研究所 高度情報化研究センター 情報基盤研究室
(〒305-0804 茨城県つくば市旭一番地)

E-mail:imai-r92ta@nilim.go.jp, shigetaka-k258@nilim.go.jp

²非会員 株式会社富士通研究所 ソフトウェア技術研究所 インテリジェントシステム研究部
(〒211-8588 川崎市中原区上小田中4-1-1)

E-mail:takahashi.tet@jp.fujitsu.com, yamakage.yuzuru@jp.fujitsu.com

³非会員 国土交通省 国土技術政策総合研究所 高度情報化研究センター 情報基盤研究室
(〒305-0804 茨城県つくば市旭一番地)

E-mail:tajima-s924a@nilim.go.jp

産業界では、サービスや製品の評価、業務の問題点や顧客からの苦情の特性把握などの有益な情報を取得するためにテキストマイニング技術を活用している。同様に、道路事業に係わる国民からの意見に対して、テキストマイニング技術を適用することで、適切な事業執行を図るうえで重要となる潜在的な問題やニーズを発掘できる可能性がある。

本研究では、国道事務所への日々の問い合わせ内容を記録した行政相談資料および道路事業に係わるTwitterのつぶやき（ツイート）に対してテキストマイニング技術を適用し、道路事業の計画や評価への適用可能性を考察した。後者のツイートに対しては、道路事業に係わるツイートの収集可能性の検証とともに、圏央道開通に係わるツイートを収集しテキストマイニング技術を適用したケーススタディを実施した。

Key Words : text mining, public opinion, road maintenance, Twitter

1. はじめに

情報通信技術（ICT）の進展やコンピュータの処理能力の向上により、産業界では業務データやインターネット上のテキストにテキストマイニング技術を適用して分析し、自社提供のサービスや製品の評価、業務の問題点や顧客からの苦情の特性把握などを行っている。テキストマイニング技術とは、文章を単語に分割し出現頻度や相関関係を分析し、これまで知り得なかった有益な知見を獲得する方法である。

道路事業も同様に、電子化された多種多様で膨大な情報が流通している。そのなかでも国民からの道路事業に対する意見は、道路管理者にとって重要な情報である。時宜に即した道路行政を推進するには、あまねく国民からの意見を元に各事業を評価し、今後の計画を立案・実行していく必要がある。道路管理者が保有する国民意見

を収録した資料として、日々の問い合わせ対応を記録した行政相談資料、対話・説明会資料や定期的実施したアンケート調査結果などがある。このうち、行政相談資料は、問い合わせのあった個々の事案毎に様式で整理されている。その内容は、質問や要望、苦情が多くを占め、肯定的な意見は少ない特徴があるが、各様式を分析することで、道路管理者の暗黙知を形式知として整理したり、潜在する問題やニーズを発掘したりできる可能性がある。

一方、社会情勢に目を向けると、Twitter¹⁾やfacebook²⁾に代表されるSNS（Social Networking Service）の浸透は目覚ましく、蓄積されたデータは“ビッグデータ”と称され、現在、多様な分野で活用方法が研究されている。なかでも日本におけるTwitterのつぶやき（以下「ツイート」という。）数は1日5,000万件とも言われている。ツイートは、大規模な災害対応のみならず、電車遅延や交通渋滞などの日々の異常事象の情報共有でも活用されている。

このツイートの中には、道路事業に対する国民の意見が潜在している可能性が極めて高い。

道路整備による交通円滑性や定時性向上の効果は、整備地点での計測者による交通量調査、民間プローブデータを活用した旅行速度の分析やプローブカーによる実走行調査の結果を用いて把握するのが一般的である³⁾。これらの調査・分析結果から得られた定量的な交通実態に対し、前述した国民（道路利用者や周辺住民など）からの意見を反映できると、道路事業評価の高度化を図る支援策となることが期待される。

本研究の目的は、国道事務所への日々の問い合わせ内容を記録した行政相談資料および道路事業に係わるツイートに対してテキストマイニング技術を適用し、道路事業の計画や評価への適用可能性を検証することとした。第2章では、国道事務所で作成された行政相談資料に対してテキストマイニング技術を適用した分析結果を報告する。第3章では、膨大なツイートから道路事業に係わるツイートを抽出する方法および抽出結果を報告する。第4章では、圏央道開通に係わるツイートを抽出し、テキストマイニング技術を用いて分析した結果を報告する。第5章では、前章までの結果に基づく道路事業への適用可能性の考察と今後の課題を論ずる。

2. 行政相談資料に対するテキストマイニング技術の適用

(1) 分析対象データ

テキストマイニング技術を適用する行政相談資料は、平成23年度上期に千葉国道事務所にて問合せ処理された事実とした。詳細は表-1に示すとおりで、性別、受付機関、路線、相談区分や相談対象といった選択肢を持つ項目と、相談内容を書き起こしたテキストからなる。

なお、氏名、住所や電話番号などの個人を特定できる項目は予め削除された行政相談資料を使用した。

(2) 分析方法および結果

分析方法は、図-1に示すとおり、選択肢を持つ項目同士の組み合わせと、テキスト解析により分割した単語と項目との組み合わせの2通りとし、特異的に現われる組み合わせの抽出に着目して分析した。各分析方法を結果例とともに以下に示す。

1つ目の分析は、選択肢を持つ項目同士の組み合わせに対して、統計的に見て有意に多いまたは少ない値の抽出を試みる。例として、「受付機関」と「相談区分」との組み合わせの分析方法を説明する。なお、問い合わせ全件数1205件に対し、「受付期間」および「相談区分」の項目がともに該当する問い合わせの数は1,201件であ

るため、以下説明では全体の数を1,201件とする。まず、実際に問い合わせのあった数（実測値）を表-2に示す。

次に、全体の数に対する割合から期待される値（期待値）を表-3に示す。表-2の「道の相談室」には合計145件の問い合わせがあり、また、全体の1,201件の相談の中で「苦情」は397件であるため、「道の相談室」に寄せられることが期待される「苦情」の件数（期待値）は表-3の $145 \times (397/1,201) = 47.93$ となる。これに対して「道の相談室」に寄せられた実際の「苦情」の件数（実測値）は表-2の10件であるため、その比を求めると $10/47.93=0.21$ となる。この値は表-4に示すとおりであり、道の相談室に寄せられた「苦情」の件数は、全体の統計量から期待される件数の0.21倍であり、期待値よりも非常に少ないと解釈ができる。期待値と実測値との比が大きいかもしくは小さい組み合わせでは何らかの理由が背景にあることが考えられるため、そのような項目の組み合わせをすべて抽出した。表-4では、この比が2.0以上または0.5以下のセルに対してピンクと水色でそれぞれハイライトしている。

2つ目の分析は、テキストで記述された「相談内容」をテキスト解析により単語へ分割し、その上で、各単語が相談内容に現れる件数と、選択肢を持つ項目との組み

表-1 行政相談資料

期間	平成23年度上期
地域	千葉国道事務所
件数	1,205件
選択項目	性別 受付機関（千葉出張所、道の相談室、etc） 路線 相談区分（苦情、要望・意見、etc） 相談対象（自然環境、道路構造、etc）
テキスト	相談内容

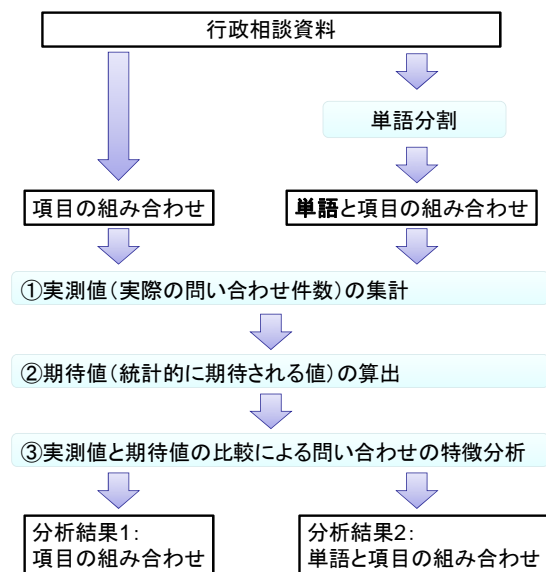


図-1 分析方法

合わせにて期待値と実測値との比を求める。そして、1つ目の分析と同様に、この比が大きいもしくは小さい組み合わせを抽出する。

項目の組み合わせによって得られた結果を表-5、単語と項目との組み合わせによって得られた結果を表-6に示す。この結果を用いて、千葉国道事務所の道路管理者にヒアリング調査したところ、経験的に感じていること（暗黙知として属人化されていた事象）が、定量的に形式知として示され、道路管理者間で共有できる情報としてまとめられているとの意見を得た。

3. 道路事業に係わるツイートの収集

本研究では、道路管理者の保有資料を学習データとして、1日5,000万件のツイートの中から、道路事業に係わる意見の抽出が可能かを検証した。

表-2 実測値（受付機関-相談区分）

相談区分 受付機関	問い合わせ	発見・通報	苦情	要望・意見	合計
千葉国道事務所	1	2	1	1	5
千葉出張所	31	84	195	40	350
酒々井出張所	8	10	9	83	110
木更津出張所	0	15	1	77	93
柏維持修繕出張所	16	136	80	29	261
船橋出張所	27	69	101	40	237
道の相談室	6	102	10	27	145
合計	89	418	397	297	1201

表-3 期待値（受付機関-相談区分）

相談区分 受付機関	問い合わせ	発見・通報	苦情	要望・意見	合計
千葉国道事務所	0.37	1.74	1.65	1.24	5
千葉出張所	25.94	121.82	116	86.55	350
酒々井出張所	8.15	38.28	36.4	27.2	110
木更津出張所	6.89	32.37	30.7	23	93
柏維持修繕出張所	19.34	90.84	86.3	64.54	261
船橋出張所	17.56	82.49	78.3	58.61	237
道の相談室	10.75	50.47	47.9	35.86	145
合計	89	418	397	297	1201

表-4 実測値と期待値の比較（受付機関-相談区分）

相談区分 受付機関	問い合わせ	発見・通報	苦情	要望・意見	合計
千葉国道事務所	2.7	1.15	0.61	0.81	5
千葉出張所	1.2	0.69	1.69	0.46	350
酒々井出張所	0.98	0.26	0.25	3.05	110
木更津出張所	0	0.46	0.03	3.35	93
柏維持修繕出張所	0.83	1.5	0.93	0.45	261
船橋出張所	1.54	0.84	1.29	0.68	237
道の相談室	0.56	2.02	0.21	0.75	145
合計	89	418	397	297	1201

(1) ツイート収集のAPI

ツイートを収集する方法は数種類あるが、主なものとしてTwitter社から提供されている2種類のAPIが無償で利用できる。それぞれのAPIは次に挙げる特徴を持つ。

- streaming API
 - ーリアルタイムデータ
 - ーランダムサンプリングによる1/100のデータ量
 - search API
 - ーキーワード指定による検索
 - ー過去1,500件・1週間の制限でほぼ全データを取得可
- streaming APIによるツイート収集は、1/100のサンプリングデータの取得制限がある。しかし、特定のキーワードを指定しなくてもツイートを収集できる利点がある。様々な話題のツイートを長期的に収集・蓄積できるので、

表-5 項目の組み合わせとデータの偏り

項目の組み合わせ	抽出された事象
受付機関-相談区分	木更津出張所は苦情が少ない 道の相談室は発見・通報が多く、苦情が少ない
区市町村名-相談対象	酒々井町、鋸南町は交通安全に関する相談が多い 南房総市、館山市は自然環境に関する相談が多い 八街市、習志野市は自然環境に関する相談が少ない
区市町村名-相談区分	富津市、木更津市には苦情が少ない
相談対象-路線	6号は清掃に関する相談が多い 51号、127号は清掃に関する相談が少ない 14号は道路構造に関する相談が多い 14号、127号は騒音・振動に関する相談が少ない
相談対象-性別	女性は道路構造に関する相談が少ない
路線-性別	126号は女性からの問合せが多い（照明に関するもの）
受付機関-性別	道の相談室には女性からの相談が少ない

表-6 項目の組み合わせとデータの偏り

単語と項目の組み合わせ	抽出された事象
相談内容：外灯-路線	51号、357号は外灯の相談が多い
相談内容：草刈-路線	6号、51号、409号は草刈の相談が多い 14号、126号、357号は草刈の相談が少ない
相談内容：死骸-路線	6号は死骸の相談が多い 51号、357号は死骸の相談が少ない
相談内容：見通し-路線	51号、127号は見通しの相談が多い 6号、16号、126号は見通しの相談が少ない
相談内容：夜間-性別	女性は夜間の相談が多い

収集開始日にまで逆のぼって多様な利用シーンに応じたマクロ的な分析ができる。

一方、search APIによるツイート収集は、ほぼ全ツイートを利用できる利点があるが、約1週間前までのツイートしか取得ができない。そのため、新しい分析対象に対して過去の時系列上での変化などを分析することは難しい。ただし、分析対象が決まれば、その分析対象を表すキーワードを設定し定期的かつ長期的にツイートの取得ができる。このため、分析対象が決まっている場合は、search APIによるツイート収集が有効である。

本研究では、道路整備に対する意見収集の用途を踏まえ、search APIによるツイート収集の方法を採用した。

(2) ツイート収集の方法

特定の事象に対するツイートを収集するために最も的確でシンプルな方法はキーワードを指定することである。上述した2つのAPIのどちらを用いても、分析対象を表すキーワードが既知の場合は、そのキーワードを含むツイートを収集できる。しかし、分析対象を表すキーワードが常に既知であるとは限らない。例えば第2章で用いた行政相談資料のような道路に対する問い合わせや要望・苦情に関するツイートを取得したい場合や、特定の地域の道路に関するツイートを取得したい場合、どのような

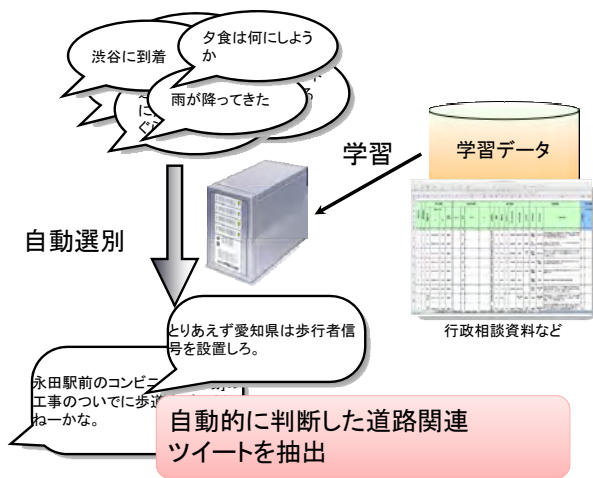


図2 文書分類によるツイート収集

順番	スコア	内容
0	4.36	新築マンションの工事のために、既存の道路に穴を開けた後また塞いだ後トラックが通るたびに段差でドン・ド
1	3.77	やはり横濱川崎はまともな道がほとんどないので人の住む所じゃない。道が狭くて歩道もなければ昔からの道は奥道でも
2	3.70	▼おもしろい写真などで切れて売れ下がっている電線には絶対に手を触れないでください。電線に絡木や看板、アンテナな
3	3.69	青森県も認める洪水による即電効果 →「青森県産エネルギーポータルサイト」ポータルサイト エネサポあおもり
4	3.25	永田駅前コンビニ、駐車場の工事のついでに歩道広げてくんなーかな。今のまま緑石置いたらかなり危ないだろ。ど
5	3.23	新青森街道でロードバイクと車の接触事故があったみたい。救急車きてたけど車道走っていき接触したのか。自分も車
6	3.11	江戸電の和田塚〜由比ヶ浜間で電車を撮影していた者と電車が接触する事故が発生したようだけど、和田塚〜由比ヶ浜間
7	3.07	もちろんならからですよ!! (笑) 当時付き合っていた彼と、環八の中央分岐帯に本当にカンナの花が咲いているか確
8	3.00	東北地方で震度5弱を最大とする地震が発生した模様です。この地震による津波の心配はありません。夜間です。食事の
9	2.99	朝職場の前の歩道のゴミを拾うのだが、毎日拾っても必ず清掃の回収車が落ちているので、これを拾った奴は一番長い間
10	2.96	2日前、店内の棚の下に小さな緑の完熟があった。片付けるのが面倒だったのでそのまま放置して帰った。今日出勤して
11	2.95	二段ベッド2台の四人部屋でカップルと一組の部屋だったんだけど三時くらいに自分の上で高の定まった。音、音浪の
12	2.93	川を渡る生き物の完熟をつくつとどじょうが飛び出し、岸壁を渡る完熟にはシャコがびしょりついてる
13	2.87	落ち葉のじょうたんは好きなんだけど、道路の側溝が詰まって大きな水溜まりが出来て、それを遠慮なくはねて行く車は
14	2.82	いつ事故が起きてもおかしなくらいに危ない通学路。あの交差点と歩道なしの道路は本当に危険。どうにか改善でき
15	2.79	【ゲーム】渋谷ヤマガダLAB1前、という車道に降りた場所にてipod nanoのBlueが落ちていた。落と

図3 文書分類によるツイート収集結果

キーワードを設定すればよいかは自明ではない。

この問題に対しては、テキストマイニングの次の2つの要素技術が適用可能である。

a) 文書分類

文書は単語の組み合わせから成り立っているため、どのような単語が含まれているかを元に、似た文書とそうでない文書との区別を機械的に行える。予め与えた文書(学習データ)を元に学習し、それらの文書と似ているかどうかを自動的に収集する方法を文書分類と呼ぶ。図-2は、例えば行政相談資料を学習データとして与え、これと似たツイートを収集する方法を示している。

図-3は、この方法による収集結果を示しており、予めキーワードを設定しなくても、分析対象としたいツイートが自動的に収集できる。行政相談資料などのように、すでに分析対象としたい内容の書かれた文書が利用可能な場合はこの方法が有効であるが、そうでない場合は、学習データを作成する必要があるため、その作成負荷が課題となる。

b) 特徴語抽出

ある文書を入力したときに、その文書に特徴的なキーワード(以下、「特徴語」という。)を抽出する方法がある。この方法は、一般的な文書における単語の出現数と、入力された文書における単語の出現数とを元に、各単語に特徴度を表すスコアを付与する。このスコアの高いキーワードを入力された文書の特徴語とする。例えば図-4のように千葉県道事務所のWebサイト⁴⁾にある「道路の現状」「記者発表」を入力した場合、「若松交差点」「東金JCT」「館山線」の単語は、一般の文書に出てくる確率よりも出現確率が高く、特徴度を表すスコアも高くなる。

本研究では、道路事業評価に有効な幅広い意見を収集する観点から、b)により抽出した特徴語をキーワードとして、道路事業評価に係わるツイート収集を実施した。なお、a)の文書分類は、現時点の学習データが行政相談資料のみである制約から、問合せ、要望や苦情などのツイートが多くなる収集結果となり、傾向が偏る懸念があったため今回は不採用とした。

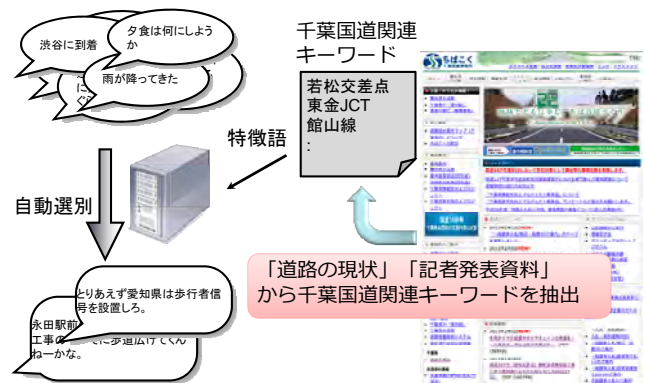


図4 特徴語抽出によるツイート収集

(3) ツイート収集の結果

search API を利用し、千葉国道事務所のWebサイトから特徴語抽出によって作成されたキーワードのリスト (図-5) を用いて収集したツイート数は、37,696 / 20日 (約1885件/日) となった。図-5における括弧内はキーワード毎のツイート数である。

収集されたツイートに対して、テキストマイニング技術の一つである評価表現分析を適用した。評価表現分析は、テキストに含まれる肯定的な表現 (例: 便利) と否定的な表現 (例: うるさい) とを抽出し、文書の肯定・否定に関する意味的な内容を明らかにする方法である。

この評価表現分析の結果の一例を図-6に示す。「東関東道」-「便利」「走りやすい」, 「国道16号」-「できる」など、千葉国道事務所の関連キーワードがポジティブな表現を含む文脈で用いられており、道路利用者から肯定的に評価されていることがわかる。

本分析結果から、膨大なツイートから道路関連キーワードを含んだツイートの収集は可能であり、キーワードに対する国民の見解も確認できることが明らかになった。これにより、ツイート収集が道路事業評価の有効な方策となり得る可能性が示されたといえる。

- 対象リスト**

 - 国土交通省 (2766)
 - 圏央道 (1615)
 - 水戸線 (319)
 - 6号線 (174)
 - 国道6号 (166)
 - 国道16号 (161)
 - 常磐自動車道 (124)
 - 加賀利 (109)
 - 東京湾アクアライン (95)
 - 東関東道 (48)
 - 東関東自動車道 (46)
 - 関東地方整備局 (40)
 - 首都圏中央連絡自動車道 (36)
 - 国道51号 (32)
 - 京葉道路 (28)
 - 国道298号 (18)
 - 房総スカイライン (12)
 - 51号線 (11)
 - 国道468号 (11)

図-5 抽出された千葉国道関連キーワード

東関東道 [pp:便利] target top	
1833	20130205 酒々井アウトレットモールのために東関東道のIC作るのか。成田空港から近いから便利だめ
1834	20130204 これは便利になるなあ。／東関東道：酒々井IC、4月10日開通 / 千葉 (毎日新聞) - Y1ニュース
東関東道 [p:走りやすい] target top	
1837	20130204 東関東道は走りやすかったですよ。
1838	20130203 サバゲー行くなび思うけど、東関東道ってめっちゃ走りやすい、右側車線はちゃんとベース早い車のための車線として使われてて余計なストレス溜まらなくて本当にやれない。滞り道に東名高速に乗った瞬間一番右側エコ走行してる人ばかりになって一気にストレスがマッハだけど。
国道16号 [p:できる] target top	
1580	20130208 国道16号は道路に並行して立派な自転車道があるわ。道路1車線分くらいの広さで、周りに自転車レンタルできる場所もたくさんあるから、自転車でも市内を回ってみるのも面白いわね。
1581	20130207 【事故<事故現場>】千葉県柏市02/07 08:22 国道16号外回り柏市大井付近。事故の為に車線規制がされて渋滞が2km位あります。通常の朝渋滞よりも通過時間がかかってます。迂回できる方は回避してください。
1582	20130217 国道16号が地元を走ってるから、寂しみのルート16の情報が鮮明に想像できる。

図-6 評価表現の抽出結果

4. 圏央道開通事例のツイート収集および分析

本章は、圏央道開通に係わるツイートを収集し、テキストマイニング技術を適用したケーススタディの結果を報告する。

(1) 分析対象

本分析では、第3章で示した方法で収集した「圏央道」の語を含む約20,000ツイートを対象とした。収集期間は表-7に示すとおり、圏央道の海老名IC～相模原愛川IC間、茅ヶ崎JCT～寒川北IC間および東金JCT～木更津東IC間の開通日を含む2013年1月30日～5月10日の101日間とした。なお、ツイート収集の際は、ボットと呼ばれる機械による自動ツイートや作為的な拡散はアルゴリズムにて除去した。

(2) 分析方法

今回のケーススタディでは、約20,000ツイートに対して、次の4種類の分析方法を適用した。

a) 時系列分析

ツイートが投稿された時刻の情報を利用してツイートを時系列上に集計し、急な盛り上がりの発生 (投稿数の多い時間帯) を検出する。

b) 評価表現分析

第3章でも適用したテキストマイニング技術の評価表現分析をツイートに適用し、肯定的な意見と否定的な意見とを抽出する。

c) 話題分析

テキストマイニング技術の話題分析をツイートに適用し、日毎に特徴的に現われているキーワードを抽出し、ツイート数の盛り上がりの原因を推測する。具体的には、すべての単語に対して日毎の出現頻度を集計し、その出現頻度を基にそれぞれの単語wの特徴度を現わすスコア $Score(w)$ を次の式により求める。

$$Score(w) = TF(w) \cdot \log \frac{N}{DF(w)}$$

表-7 圏央道開通事例の分析対象

収集期間:	2013年1月30日～5月10日 (101日間)
開通日:	海老名IC～相模原愛川IC間 (3月30日) 茅ヶ崎JCT～寒川北IC間 (4月14日) 東金JCT～木更津東IC間 (4月27日)
収集条件:	「圏央道」という語を含むツイート
ツイート数:	22,310 19,410 (ボット・スパム除去後)
ユーザー数:	11,364名 11,357名 (ボット・スパム除去後)

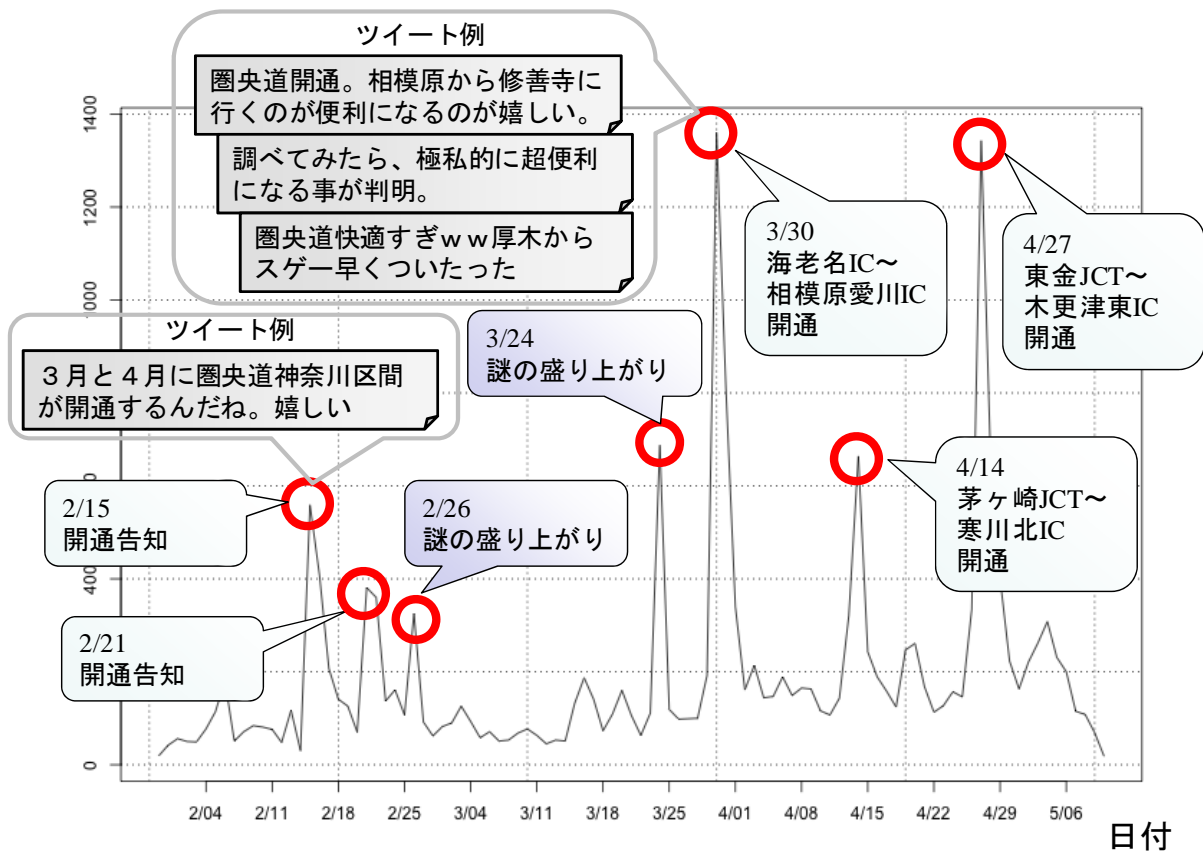


図-7 ツイート数の時系列推移

ここで $TF(w)$ は特定の日における単語 w の出現数 (Term Frequency) であり, DF は単語 w が含まれている文書数 (Document Frequency), M は全文書数である. このスコアが高いほど, その日の特徴を現わしており, 特徴的な単語のリストを確認することで, 一つひとつの記事を確認しなくてもその日に何が話題であったかを推測できる.

d) ユーザーの居住地域毎の特徴分析

ユーザーがプロフィール情報として登録している居住地域を利用して居住地域毎のツイートの特徴を分析する. 具体的には, a)で示した時系列分析結果を居住地域毎に集計後, c)で示したテキストマイニング技術の話題分析結果を用いて居住地域毎の盛り上がりの発生原因を推測する.

(3) 分析結果

a) 時系列分析

図-7は, 日別のツイート数を集計した結果を示しており, 開通情報の告知日や開通日にツイート数の急な盛り上がりがあることがわかる. 3月30日や4月27日には, 1日に約1,400件ものツイートが投稿されており, 国民の関心の高さが見て取れる. これらの盛り上がりと比較することにより, 国民の反応の相対的な把握ができる.

b) 評価表現分析

図-8に評価表現分析の時系列推移を集計した結果を示

す. 肯定的な意見 (便利, 快適, 嬉しいなど) が否定的な意見 (疲れる, 怖い, 悲しいなど) を大きく上まわっており, 肯定的な意見と否定的な意見との割合 (ポジティブ率) は, 概ね80%前後で推移していることがわかる (図-9). 具体的には図-7中のツイート例のような肯定的な意見が数多く見られ, 国民がそれぞれの主観的な言葉で圏央道の開通を肯定的にとらえていることが分かる.

c) 話題分析

図-7に示す2月26日と3月24日は, 開通日や告知日とは関係なく盛り上がりが発生している. これらの盛り上がりに対して, 話題分析を行った.

抽出された3月24日の特徴的なキーワードを表-8に示す. 表-8からは, 厚木でウォーキングのイベントがあったことが推測される.

より具体的に話題を把握するには, 元の文章 (ツイート) も確認すべきであるが, その際には上記の特徴的キーワードのリストと元の文章とを相互に見られるようなインターフェースがあると便利である. 一例として, ウェブブラウザを用いたインターフェースを図-10および図-11に示す. 2月26日は開通PRのためのゾウによるイベント (図-10), 3月24日は圏央道を歩くイベント (図-11) があったことが分かる.

この事例で示したように, 話題分析によって特異な事象の発見およびその内容の把握ができることがわかる.

ツイート数

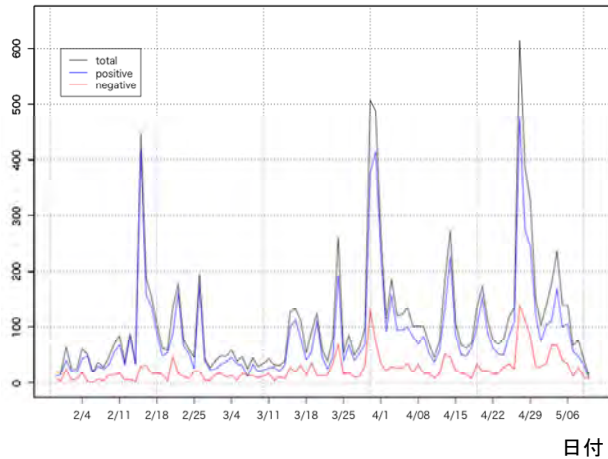


図-8 評価表現の時系列推移

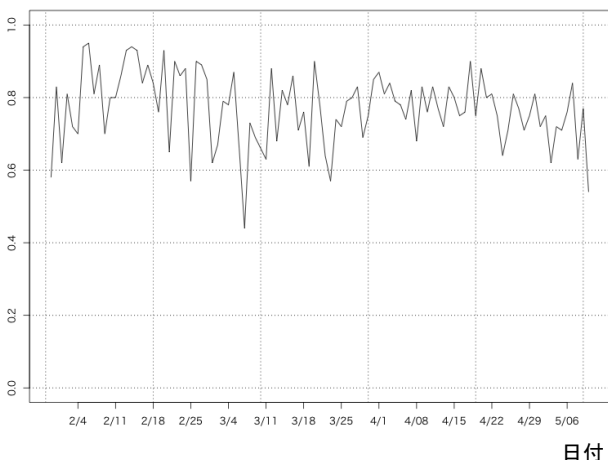


図-9 ポジティブ率の時系列推移

d) ユーザーの居住地域毎の特徴分析

Twitterではユーザーがプロフィール情報として居住地域を登録できる機能があり、約6割のユーザーが何らかの登録をしている。その登録内容は、「渋谷区」などの市区町村単位や「東京都」のような都道府県単位、そして「日本」のような国の単位にまで多岐に渡る。なかには「夢の中」のように地域を特定できないような登録もある。市区町村単位で登録されていれば、多くの場合は一意に都道府県単位への変換ができる。このため、本研究ではユーザーの居住地域の単位を都道府県とした。その結果、全ユーザーのうち都道府県の単位で居住地域が特定できるユーザーは、約4割となった。

都道府県単位で特定した居住地域を利用し、ユーザーの居住地域毎の特徴を分析し、図-12の結果を得た。なお、これらの数値は都県間を比較しやすいように、東京のユーザー数を基準として正規化している。図-12を見ると、圏央道の開通対象である神奈川と千葉の2県において、それぞれの地域での開通のタイミングに合わせてツイート数の盛り上がりが発生していることがわかる。

2月21日に着目すると、神奈川および千葉の2県で盛り

表-8 3月24日の特徴的キーワード

順位	キーワード	スコア
1	歩く	497.46
2	見学	491.53
3	前	470.54
4	厚木	418.60
5	ウォーキング	386.93
⋮	⋮	⋮

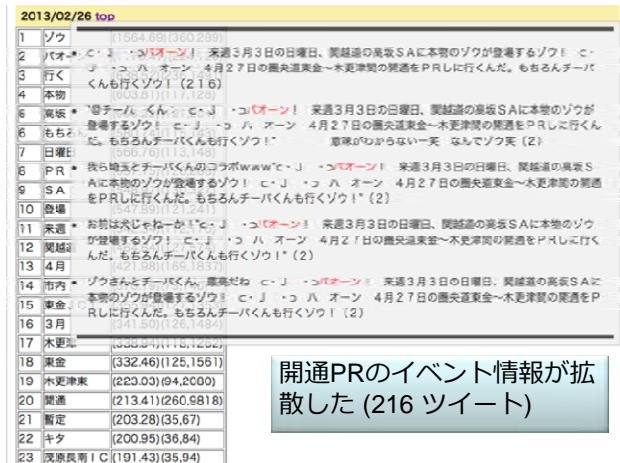


図-10 話題分析 (2月26日)

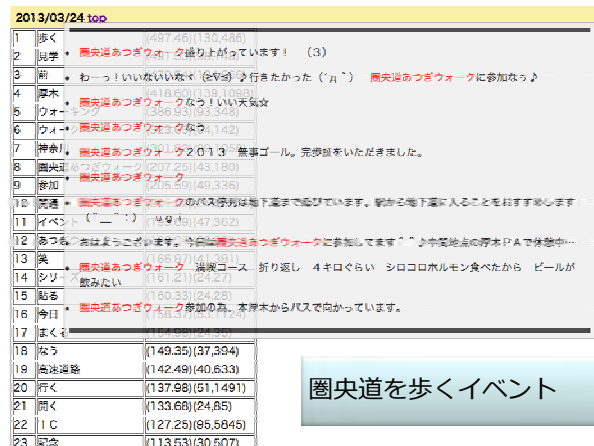


図-11 話題分析 (3月24日)

ツイート数

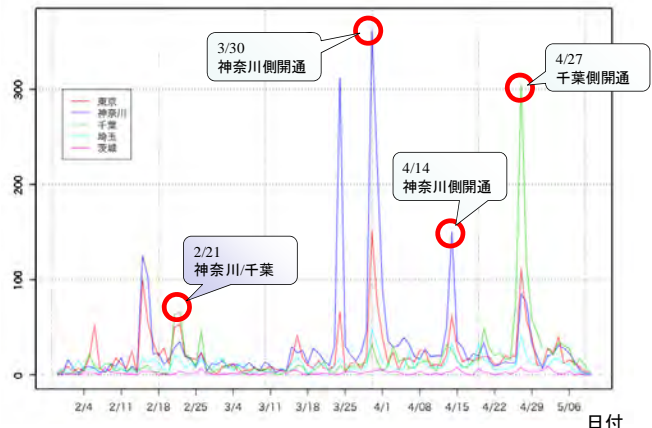


図-12 居住地域毎のツイート数の推移

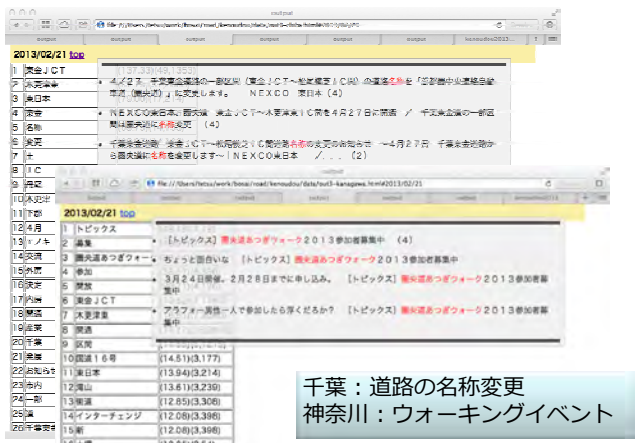


図-13 2月21日の話題の地域差

上がりが発生している。同日の話題を分析すると、千葉は道路の名称変更、神奈川はウォーキングイベントといった盛り上がりの話題の地域差が抽出された（図-13）。

5. おわりに

本研究は、国道事務所への日々の問い合わせ内容を記録した行政相談資料および道路事業に係わるツイートに対してテキストマイニング技術を適用し、道路事業の計画や評価への適用可能性を検証した。

第2章では、道路管理者が保有する行政相談資料に対してテキストマイニング技術を適用すると、問い合わせ内容の傾向の定量的な把握が可能となり、道路行政においても有効な方法であることが示された。第3章では、膨大なツイートから道路事業に係わるツイートの収集が可能であり、道路関連キーワードに対する国民の意見を取得できることを示した。第4章では、道路事業評価の

具体例として圏央道開通の事例を対象としたツイートを収集し、テキストマイニング技術を適用したケーススタディを実施した。その結果、圏央道に対する国民の関心の高さや肯定的に評価されていること、話題の地域差が抽出できることを示した。

本研究の成果から、テキストマイニング技術を用いた分析は、道路行政サービスや道路事業評価の効率化・高度化に寄与できる可能性が高いといえる。

今後の課題としては、テキストマイニング一般の課題ではあるが、次の2点が挙げられる。

- ツイートの信憑性の評価や作為的なツイート拡散を想定したクリーニング処理を確立する。
- 評価表現分析の際に詳細なテキスト解析を実施し、文脈における評価表現抽出の精度を高めるとともに、否定表現や仮定表現などを区別する。

今後は、交通量や旅行速度（民間プローブデータ）などの道路交通データを用いた客観的な分析結果と、本研究による主観的な分析結果とを組み合わせることで、道路事業評価の更なる高度化の可能性を模索する。

謝辞：本研究の遂行にあたり、千葉国道事務所から行政相談資料を提供いただくとともに、同事務所の各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) Twitter : <<https://twitter.com/>>, (入手 2013.8.1)
- 2) facebook : <<https://ja-jp.facebook.com/>>, (入手 2013.8.1)
- 3) 橋本浩良, 河野友彦, 門間俊幸, 上坂克巳 : 交通円滑化対策のためのプローブデータの分析方法に関する研究, 国土交通省国土技術研究会, 2010
- 4) 千葉国道事務所 : < <http://www.ktr.mlit.go.jp/chiba/>>, (入手 2013.7.23)

(2013.8.2 受付)

ANALYSIS OF PUBLIC OPINION FOR ROAD MAINTENANCE USING TEXT MINING

Ryuichi IMAI, Tetsuro TAKAHASHI, Satoshi TAJIMA, Yuzuru YAMAKAGE and Koichi SHIGETAKA