

道路事業に係わる行政相談資料及びSNSデータへのデータマイニング技術の適用性に関する考察

今井龍一・田嶋聡司・重高浩一

1. はじめに

私たちの社会には多種多様で膨大なデジタルデータが流通している。昨今は、“ビッグデータ”と称され、成長戦略などでも取り上げられている。その有効な活用例として、産業界では業務データやインターネット上のテキストデータを分析し、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が行われている。このような分析には、各データの項目間の相関やパターンを分析したり、文章を単語に分割して出現頻度や相関を分析したりして、これまで知り得なかった有益な知見を獲得するデータマイニング技術が用いられている。

道路事業も同様に、多種多様で膨大なデジタルデータが流通している。そのひとつとして、国民からの問合せ対応を行った内容を記録した行政相談資料がある。通常、行政相談資料は、問合せのあった事案毎に様式で整理されている。それら様式の全てを対象にデータマイニング技術を適用して分析することで、潜在する問題やニーズを発掘したり、道路管理者のノウハウ（経験知・暗黙知）を共有・継承できる形（形式知）に整理したりできる可能性がある。

一方、私たちの日常生活に目を向けると、Twitter¹⁾やfacebook²⁾に代表されるSNS（Social Networking Service）の浸透は目覚ましく、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声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している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、前述した国民の声を反映できると、道路事業評価の高度化の支援策となることが期待される。

本稿は、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データマイニング技術を適用して分析し、道路事業の計画や評価への適用可能性を検証した結果を報告する。

2. 行政相談資料の分析

2.1 分析対象

今回の分析には、平成23年度上期の千葉国道事務所の行政相談資料を用いた。資料の概要は表-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ている。

表-1 行政相談資料の概要

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、道の相談室等) 路線 相談区分(苦情、要望・意見等) 相談対象(自然環境、道路構造等)
テキスト	相談内容

2.2 分析方法及び結果

分析方法は、選択肢を持つ項目同士の組合せと、相談内容の文章を最小単位に分割した単語と項目との組合せの2通りとした（図-1）。

2.2.1 項目の組合せ分析

項目の組合せ分析では、選択肢を持つ項目同士の組合せに対して、統計的に見て有意に多いまたは少ない値を算出した。表-2は、「路線」と「相談区分」との組合せの分析結果例を示している。

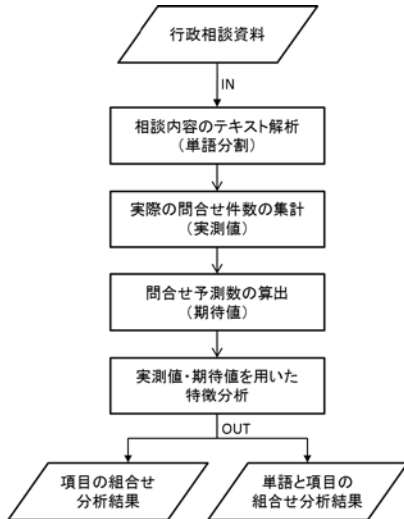


図-1 分析方法

表-2 路線と相談区分の組合せの分析結果例

相談区分	合計	間合せ	発見・通報	苦情	要望・意見
「路線」と「相談区分」の両項目が入力されている間合せ	1,176	85 (7.2%)	408 (34.7%)	390 (33.2%)	293 (29.4%)
127号線: 実測値	68	0	9	0	59
127号線: 期待値	68	4.91	23.59	22.55	16.94
127号線: 実測値÷期待値	-	0	0.38	0	3.48

表-2の各行の内容を次の(1)～(4)に示す。

- (1) 1行目：間合せ全件数1,205件に対する「路線」および「相談区分」の両項目が入力されていた合計数（1,176件）と相談区分毎の件数（割合）
- (2) 2行目：実際の127号線に関する間合せの件数（以下、「実測値」と言う。）
- (3) 3行目：「路線」の127号線に関する件数（68件）と、その件数に1行目の相談区分の割合を乗じた値（以下、「期待値」と言う。）
- (4) 4行目：実測値を期待値で除した値

表に示すとおり、「発見・通報」は期待値の0.38倍と小さく、「要望・意見」は期待値の3.48倍と大きくなっている。この分析結果を用いると、全体の統計量から予測（期待）される間合せ件数に対して、実際にどの程度の間合せが寄せられているかが把握できる。

2.2.2 単語と項目の組合せ分析

単語と項目の組合せ分析では、「相談内容」のテキストデータの文章をテキスト解析により単語へ分割し、その上で、各単語の出現数と、選択項目との組合せにて期待値と実測値を比較した。項

目の組合せにて得られた結果を表-3、単語と項目との組合せにて得られた結果を表-4に示す。この結果を用いて千葉県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。

表-3 項目の組合せとデータの偏り

項目の組合せ	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号は少ない 14号は道路構造に関する相談が多い
相談対象-性別	女性は道路構造に関する相談が少ない

表-4 単語と項目の組合せとデータの偏り

単語と項目の組合せ	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号は少ない
相談内容：死骸-路線	6号は死骸の相談が多い 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号は少ない
相談内容：夜間-性別	女性は夜間の相談が多い

3. 道路事業に係わるツイートの分析

3.1 分析対象

今回の分析では、圏央道開通の事例を対象とし、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日間に収集された「圏央道」の単語を含む約20,000ツイートをを用いた（表-5）。

表-5 分析対象のツイート

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」の単語を含むツイート
ツイート数	22,310 (ボット・スパム除去前) 19,410 (ボット・スパム除去後)
ユーザー数	11,364名 (ボット・スパム除去前) 11,357名 (ボット・スパム除去後)

ツイートの収集にはTwitter社から提供されているsearch APIを利用した。なお、分析には、ボットと呼ばれる機械による自動ツイートや作作的なスパム（拡散）は、プログラム処理にて除去したツイートを用いた。

3.2 分析方法及び結果

3.2.1 時系列分析

各ツイートは投稿時刻がわかるため、時系列に集計ができる。図-2は、圏央道に関するツイート数の日別の集計結果を示しており、開通の告知日や開通日にツイート数の急な盛り上がりが発生していることがわかる。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から、定性的な効果が把握できる。

3.2.2 評価表現分析

評価表現分析とは、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、文章の肯定・否定に関する意味的な内容を明らかにする方法である。

この方法を圏央道に関するツイートに適用し、肯定的な意見と否定的な意見との割合を調べたところ、肯定的な意見が約80%と否定的な意見を大きく上回っていることを確認した。具体的には図-2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わかる。

3.2.3 話題分析

話題分析とは、日毎に特徴的に出現し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方法である。図-2の2月26日と3月24日を見ると、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の原因を把握するために、話題分析を適用し、日毎に集計したツイートの中で特徴的に出現しているキーワードを抽出した。次の(1)・(2)の用語は、抽出結果を出現数の多い順に示している。

- (1) 2月26日：ゾウ、パオーン、行く、本物、高坂、もちろん、日曜日、PR、SA、…
- (2) 3月24日：歩く、見学、前、厚木、ウォーキング、ウォーク、神奈川、圏央道あつぎウォーク、…

抽出結果を元に、特徴的なキーワードを含むツイートの内容を確認したところ、2月26日は圏央道の開通PRのためのゾウのパフォーマンスのあるイベント（開催日は3月3日）情報が拡散していたこと、3月24日は圏央道を歩くイベントが開催されていたことが分かった。

著者らは、これらのイベントのことを把握していなかった。しかし、話題分析によって抽出した特異な事象からイベントの開催と併せて国民の声も把握することができた。

3.2.4 ユーザーの居住地域毎の特徴分析

Twitterにはユーザーがプロフィール情報として居住地を登録できる。今回のツイートを分析したところ、約4割のユーザーが都道府県単位の居住地を登録していた。この情報を利用してユーザーの居住地域毎の特徴を分析し、図-3の結果を得た。なお、都県でユーザー数が異なるため、図-3の縦軸の数値は、東京のユーザー数を基準に正規化して示している。

図-3を見ると、圏央道の開通対象である神奈川県と千葉県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

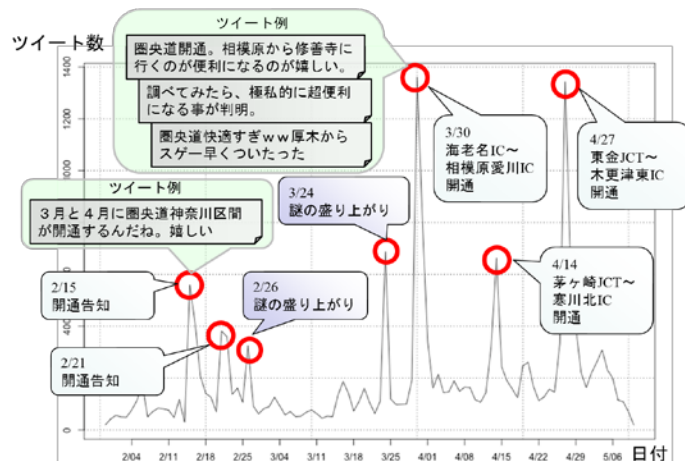


図-2 ツイート数の時系列推移

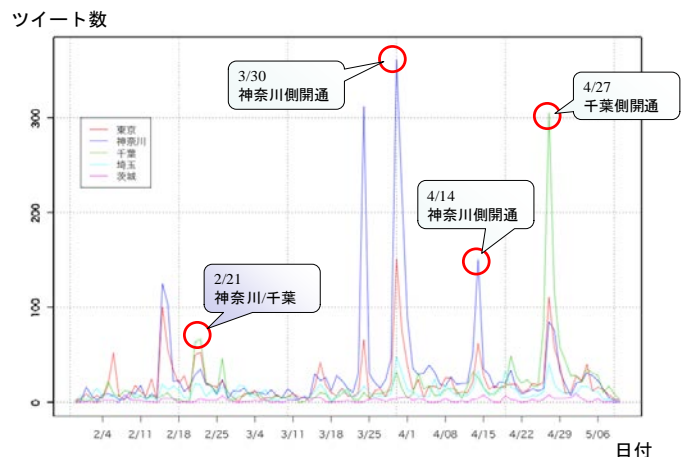


図-3 居住地域毎のツイート数の推移

いることがわかる。また、2月21日に着目すると、千葉県、東京都および神奈川県で盛り上がりが発生している。これに前述の話題分析を適用し、同日の話題を分析した結果、千葉県は道路の名称変更、東京都は国道16号と圏央道の接続、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。

4. おわりに

本稿は、行政相談資料および圏央道に関するツイートに対してデータ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性を考察した。

行政相談資料の分析では、問合せ内容の定量的な傾向の把握が可能となり、道路行政でも有効な方法であることを示した。

道路事業に係わるツイートの分析では、膨大なツイートから圏央道開通に係わるツイートを収集し、国民の関心の高さ、肯定的な評価や話題の地域差が抽出できることを示した。一方、課題としては、データマイニング一般の課題も含むが次の3点があげられる。

- (1) ツイートの信憑性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。
- (2) 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、否定表現や仮定表現などを区別する。
- (3) 今回は「圏央道」と道路名が明確であったため、関係するツイートを収集できたが、一般道を対象にしたツイートの収集は容易でない。道路や場所を特定するには、教師データ（道路名や通称名を列挙した地名辞典など）の作成などの措置を講ずる必要がある。

今回の結果に基づくと、行政相談資料やSNSデータへのデータマイニング技術を適用した分析は、道路行政サービスや道路事業評価の高度化に寄与することが期待される。なお、本分析結果の詳細は既往文献を参考されたい⁴⁾。

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など、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、行政相談資料を提供いただくとともに協力・助言をいただいた千葉国道事務所の方々、分析の協力をいただいた(株)富士通研究所の方々に感謝の意を表する。

参考文献

- 1) Twitter<<https://twitter.com/>>、(入手2014.5.7)
- 2) facebook : <<https://ja-jp.facebook.com/>>、(入手2014.5.7)
- 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) 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・講演集、Vol.48、土木学会、2013

今井龍一



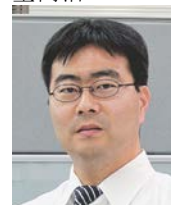
国土交通省国土技術政策総合研究所メンテナンス・基盤研究センターメンテナンス情報基盤研究室 研究官、博士(工学)
Dr. Ryuichi IMAI

田嶋聡司



国土交通省国土技術政策総合研究所メンテナンス・基盤研究センターメンテナンス情報基盤研究室 部外研究員
Satoshi TAJIMA

重高浩一



国土交通省国土技術政策総合研究所メンテナンス・基盤研究センターメンテナンス情報基盤研究室長
Koichi SHIGETAKA