

# 行政相談資料を用いたテキストマイニングによる国民意見の分析

国土交通省 国土技術政策総合研究所 ○田嶋聡司  
 同 今井龍一  
 同 重高浩一  
 株式会社富士通研究所 ソフトウェア技術研究所 高橋哲朗  
 同 山影 譲

## 1. はじめに

情報通信技術（ICT）の進展により、日常業務の大半の文書が電子化されている昨今、産業界ではテキストマイニング技術を用いて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などを分析し、改善を図っている。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することで、これまで知り得なかった有益な情報を取得する手法である。

道路行政においても様々な文書が電子化されており、そのひとつとして、国民からの日々の問い合わせ対応を行った内容を電子化した行政相談資料がある。通常、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。それら様式のすべてを対象にテキストマイニング技術を適用して各事案を多角的かつ複合的に分析することで、道路管理者の経験知（暗黙知）を形式知として整理（見える化）したり、潜在する気付きやニーズを発掘したりできる可能性がある。

本稿は、千葉国道事務所にて作成された行政相談資料にテキストマイニング技術を適用し、問い合わせ内容の傾向や行政へのニーズなど、道路行政サービスの検討に資する有益な知見の取得可能性を考察する。

## 2. テキストマイニング技術を適用する行政相談資料

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い合わせ処理された1,205件の事案とした。行政相談資料にて整理されている項目は、性別、受付機関（出張所、相談室など）、路線、相談区分（問い合わせ、発見・通報、苦情など）、相談対象（工事情報、清掃、道路構造など）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、本研究では、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データを使用した。

## 3. 分析方法および結果

本章では、2つの分析方法を結果例とともに報告する。

1つ目の分析では、選択肢を持つ項目の対において、統計的に見て多い／少ない値の発見を試みた。例として、表-1では「受付機関」と「相談区分」という組み合わせにおいて、期待値と実測値の比を記している。期待値とは全体の数に対する割合から期待される値であり、実測値とは実際の問い合わせのあった数である。例えば表-1の「道の相談室」には合計145件の問い合わせがあり、また、

表-1 データの偏りの調査（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
千葉出張所	1.20	0.69	1.69	0.46	350
千葉国道事務所	2.70	1.15	0.61	0.81	5
木更津出張所	0.00	0.46	0.03	3.35	93
柏	0.83	1.50	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
酒々井出張所	0.98	0.26	0.25	3.05	110
合計	89	418	397	297	1201

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」に寄せられるであろう「苦情」の件数（期待値）は、 $145 \times (397/1,201) = 47.93$ である。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10件であるため、その比を求めると $10/47.93 = 0.21$ となり、この値を表に記している。この値から、道の相談室に寄せられた苦情は全体の統計量から期待される件数の0.21倍であり、少なかったと解釈ができる。期待値と実測値の比が大きい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-1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つ目の分析では、まずテキストで記述された「相談内容」に対してテキスト解析を適用して単語へ分割した。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合わせにおいて、期待値と実測値との比を求めた。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した。

項目の組み合わせによって得られた結果を表-2、単語と項目の組み合わせによって得られた結果を表-3に示す。2つの分析によって抽出された事象の中には、道路管理者が経験的に感じていたことが形式知として整理（見える化）されている事象に加え、これまで認識していなかった事象も含まれていたため、気づきの抽出やニーズの発掘にも活用できることが確認できた。

### 3. おわりに

本研究では、行政相談資料にテキストマイニング技術を適用することで、問い合わせ内容の傾向や行政へのニーズなどが抽出できることを確認した。換言すると、行政相談資料へのテキストマイニング技術の適用は、道路行政サービスの効率化や高度化に寄与することを示唆する結果が得られた。

今後は、道路管理者が簡便にテキストマイニング技術を用いた行政相談資料の分析ができるツールの開発に取り組んでいきたい。

**謝辞:** 本研究の遂行にあたり、千葉国道事務所からは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。

表-2 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をあまり受けない 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	16号は清掃に関する相談が少ない 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
相談対象 - 性別	女性は道路構造についてはあまり相談しない
路線 - 性別	126号は、女性からの問合せが多い (特に照明に関するもの)
受付機関 - 性別	道の相談室には、女性からの相談が少ない
相談区分 - 性別	女性は発見・通報をしない 男性よりも女性の方が苦情を言う

表-3 単語と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号で、外灯に関する相談が多い
相談内容:「草刈」 - 路線	6号、51号、409号で、草刈の依頼が多い 14号、126号、357号で、草刈の依頼が少ない
相談内容:「死骸」 - 路線	6号で死骸に関する相談が多い 357号で死骸に関する相談が少ない
相談内容:「見通し」 - 路線	51号、127号で見通しに関する相談が多い 6号、16号、126号では少ない
相談内容:「夜間」 - 性別	女性は「夜間」に関する相談が多い