

道路除排雪に対する住民の要望・苦情の内容分析と事前予測

○石川大貴¹・小野寺誠²・上村靖司³・杉原幸信³

(1:長岡技術科学大学大学院 工学研究科 2:長岡技術科学大学 工学部 3:長岡技術科学大学 機械系)

1. はじめに

日本有数の経済都市でありながら、多量の雪が降る札幌市は冬季の経済と市民の生活を支えるために雪対策に大きな予算を費やして、その額は年間 200 億円にも及ぶ¹⁾。それにもかかわらず、市民からの除雪に対する要望・苦情は毎年 1~2 万件、雪の多い年にはそれ以上の件数の要望・苦情が寄せられる。本研究ではこれらの寄せられた要望・苦情を分析し、札幌市民が直面している除排雪の困り事への理解を深める。さらに、どのような気象条件でどの程度の要望・苦情が寄せられるのかを事前に予測する手法の開発を目指す。

2. 方法

2.1 内容分析

札幌市に寄せられた要望・苦情を分析する方法として、テキストマイニングを用いた。テキストマイニングとは自然言語解析の一手法で、文章を単語や文節に分割し、それらの出現頻度や相関関係を分析することで有益な情報を抽出することができる。本研究では、テキストマイニングソフトウェアである KH-coder を用い、単語出現頻度表と共起ネットワーク、から平成 26 年度から平成 28 年度の間に札幌市に寄せられた 52,622 件の要望・苦情を分析した^{2,3)}。

2.2 予測手法

要望・苦情件数を事前に予測する方法として、機械学習による予測を試みる。機械学習は、コンピュータ内でデータを反復的に学習することによってデータの中に潜む特徴や規則性を見つけ出す手法である。本研究で行う機械学習を用いた予測手順を図 1 に示す。はじめに、学習に用いるデータを収集する。今回は平成 26 年度から平成 28 年度の札幌市の気象データ⁴⁾と要望・苦情件数を用いる。気象データと要望・苦情件数の相関係数を求め、要望・苦情件数の増減傾向を分析する。相関が強い気象データを用いて、教師あり学習の一つである重回帰分析を行うことで、要望・苦情件数の予測モデルを構築する。構築したモデルから要望・苦情件数を出力し、実際の件数と比較を行い、精度を検証する。

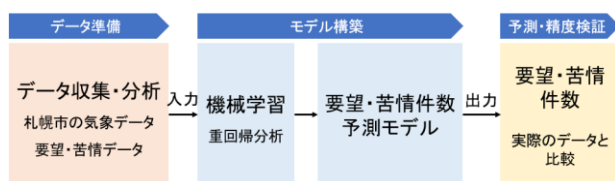


図1 機械学習の流れ。

3. 内容分析結果

3.1 抽出語

3 年度分の要望・苦情データにおける単語出現頻度表(上位 10 単語)を表 1 に示す。それぞれの年の出現回数が多い語はほぼ共通していることから、毎年同じような要望・苦情が寄せられていると考えられる。詳細な単語を見ると、3 年間共通して「家」「前」「間口」「置く」「車」「道路」「ザクザク」といった語が多く出現していることが分かる。このことから、自動車や道路の状況に関する要望・苦情、家の前の除雪に関する要望・苦情が多く寄せられていると推測できる。

3.2 共起ネットワーク

テキストデータ内においてある語と他の語が一緒に出現することを共起といい、共起する語を線で結んだものを共起ネットワークという。共起ネットワークの読み方として、円の大きさは語の出現頻度、円を結ぶ線の太さは関連の強さを示す。似た出現パターンの語は同じ色の円で色分けされる。

札幌市全体における平成 26 年度から平成 28 年度の 3 年度分の要望・苦情の共起ネットワークを図 2 に示す。図 2 より、札幌市には自宅前の除雪に関すること、道路状況に関すること、現地確認の要望・苦情、砂箱に関する要望・苦情、交差点や道角の除雪に関する要望・苦情が寄せられていると考えられる。

3.3 月別の要望・苦情分析

図 3 に外部変数を月とした共起ネットワーク分析の結果を示す。図を見ると、11 月、12 月、1 月、2 月、3 月というように月に応じてまとまりができていくことがわかる。それぞれの月を詳細に見ていくと、11 月の周囲には「砂箱」、「砂」、「補充」が集まっており、砂箱関連の要望・苦情が多い。12 月には「間口」、「駐車」、「均等」、「砂」、「砂箱」、「補充」、「処理」が結びついており、家および車庫の間口の除雪や砂箱関連の要望・苦情が多く寄せられたと考えられる。1 月は交差点や路面の状態、現地確認の要望・苦情が多く寄せられている。そして、2 月や 3 月になると路面の整正作業に関する要望・苦情が増加することがわかる。これらの結果から、月ごとに要望・苦情内容の傾向は変化することがわかった。また、月によって降雪量や平均気温などの気象条件は大きく変わるため、要望・苦情内容と気象条件には大きな関係があると考えられる。

表1 単語出現頻度表.

平成26年度		平成27年度		平成28年度	
抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
雪	7871	雪	5290	欲しい	7677
欲しい	7592	欲しい	3814	雪	7349
除雪	5116	除雪	3619	除雪	5380
置く	4088	置く	3105	前	3989
前	3576	前	2245	置く	3939
ザクザク	3555	間口	1289	車	3599
車	3198	家	1193	道路	2964
道路	2764	車	1094	ザクザク	2871
間口	2273	雪山	1088	排雪	2183
家	1902	歩道	1068	歩道	2104

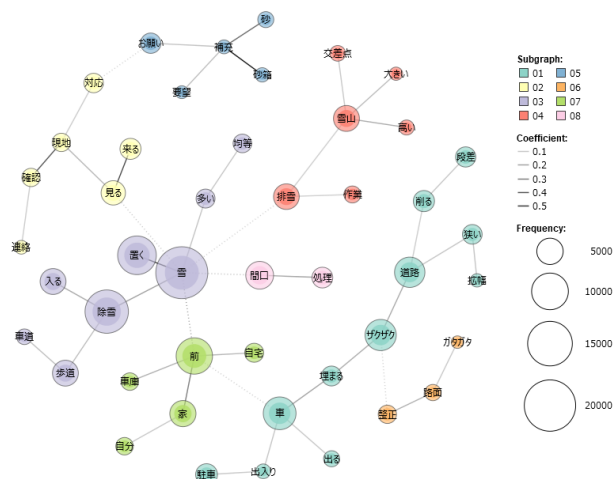


図2 札幌市の要望・苦情の共起ネットワーク.

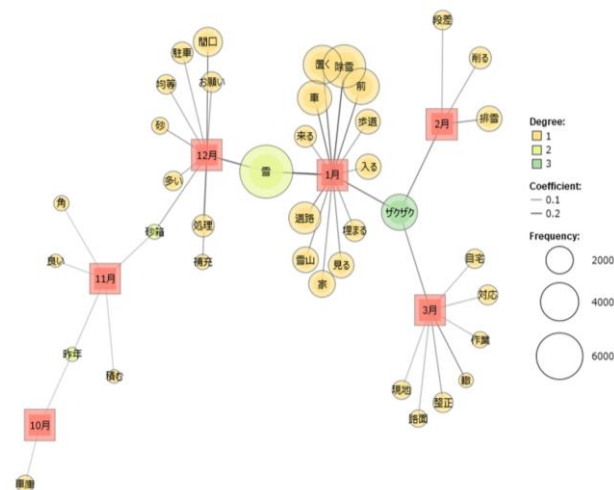


図3 外部変数を月とした共起ネットワーク分析.

4. 要望・苦情件数の予測結果

4.1 特徴量と相関係数

要望・苦情内容と気象条件の関係を踏まえて、札幌市の気象データから要望・苦情件数の予測を行う。予測を行うための特徴量として、札幌市の降雪量、前日降雪量、積雪深、平均気温、最高気温、最低気温、前日平均気温の計7つを使用する。目的変数は、平成26年度から平成28年度の1

日の間に寄せられた要望・苦情件数とする。特徴量と目的変数の相関係数を求めた結果を表2に示す。相関係数の値が大きいほど特徴量と目的変数の間に強い相関があり、予測に有効な特徴量であると判断することができる。表2より、前日降雪量の相関係数が0.5281と最も高く、次いで積雪深の値が高い結果となり、前日降雪量と積雪深の2つの特徴量が要望・苦情件数に大きく影響していることがわかった。いずれの相関係数の値も決して高いとは言えないが、本研究ではこれらの特徴量を使用して予測モデルを作成する。

表2 特徴量と要望・苦情件数との相関係数

No.	特徴量	相関係数
1	降雪量	0.2350
2	前日降雪量	0.4494
3	積雪深	0.5281
4	平均気温	0.3134
5	最高気温	0.2985
6	最低気温	0.3233
7	前日平均気温	0.2999

4.2 要望・苦情件数予測

重回帰分析による予測モデルの作成には、Pythonの機械学習ライブラリscikit-learn⁹⁾を使用する。3.3節に示した7つの特徴量のうち、使用する特徴量を変えて4つの予測モデルを作成した。それぞれの予測モデルに対する決定係数を求めた結果を表3に示す。決定係数は、0から1の間の値をとり、値が大きくなるほど予測値と実測値の誤差が少ない回帰式であることを意味する。表3より、積雪深と前日降雪量の2つの特徴量を用いた予測モデルの決定係数が0.3861と最も高くなったが、他の特徴量を用いた予測モデルとほとんど差はない。この理由の1つとして、使用した特徴量の相関係数がもともと低く、予測に有効な特徴量ではなかったことが考えられる。

また、平成26年度における実際の要望・苦情件数と、特徴量に積雪深と前日降雪量を用いた予測結果との比較を図4に示す。予測の結果、実際の要望・苦情件数の定性的な傾向を再現できていることがわかった。一方で、定量的な予測には改善の余地がある。今回は平成26年度から平成28年度に寄せられた全ての要望・苦情を用いて予測を行った。そのため、適用範囲が広く、1つ1つの要望・苦情に対する強い傾向がつかみにくくなってしまったことから精度も悪くなってしまったと考えられる。要望・苦情の内容と気象条件の関係から、期間を定めて重回帰分析を行うことで有効な特徴量も変化し、精度も向上することが期待できる。

表3 予測結果の決定係数

予測に使用した特徴量	決定係数
積雪深, 前日降雪量	0.3861
積雪深, 前日降雪量, 最低気温	0.3763
積雪深, 前日降雪量, 最低気温, 平均気温	0.3766
全ての特徴量	0.3741

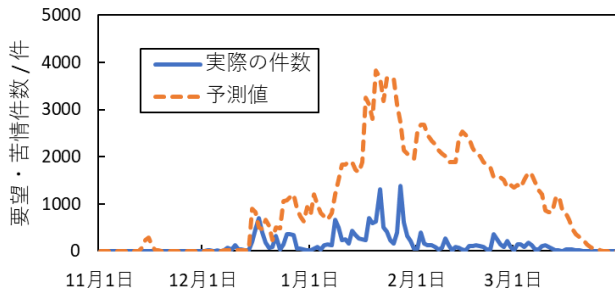


図4 要望・苦情件数の予測結果(平成26年度).

4.3 家の前の除雪に関する要望・苦情件数の予測

予測精度を向上させるために、機械学習の適用範囲を狭めて予測モデルの構築を行う。3.3節において、12月では家の前の除雪に関する要望・苦情が多く寄せられていることがわかった。そこで、使用する特徴量を札幌市の平成26年度から平成28年度の12月の気象データとし、目的変数を同年の12月に寄せられた家の前の除雪に関する要望・苦情件数に限定して予測モデルの構築を行った。

はじめに、特徴量と家の前の除雪に関する要望・苦情件数との相関係数を求める。4.1節で用いた特徴量に加え、平年比積雪深を用いた。平年値は1991-2020年の30年間の観測値の平均をもとに算出しており、平年比積雪深は1日の間の観測値と平年値との比をいう。特徴量と家の前の除雪に関する要望・苦情件数との相関係数を求めた結果、表4に示すように3つの特徴量が家の前の除雪に関する要望・苦情件数と強い相関をもつことが確認できた。使用するデータを12月に限定することで、相関係数の値が大きくなり、予測に有効な特徴量を準備することができた。

表4に示す特徴量を使用して予測モデルを構築し、予測を行った。その結果、予測モデルに対する決定係数は0.5952となり、精度の向上が確認できた。また、実際の家の前の除雪に関する要望・苦情件数と、予測結果との比較を図5に示す。図4と比較すると、定量的な予測に改善が見られ、定性的な傾向の再現度も高くなっていることがわかる。これらの結果から、適用範囲を狭めて予測を行うことで特徴量と要望・苦情件数の相関が強くなり、予測精度も向上することができた。

表4 特徴量と家前の除雪に関する要望・苦情件数との相関係数

No.	特徴量	相関係数
1	前日降雪量	0.7400
2	平年比積雪深	0.6483
3	積雪深	0.5595

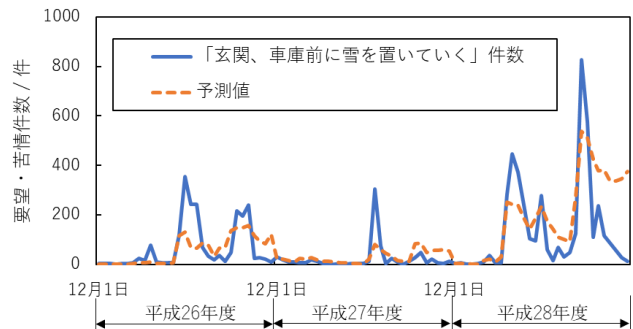


図5 家の前の除雪に関する要望・苦情件数の予測結果.

5. 結言

本稿ではテキストマイニングにおいて用いられる単語出現頻度表と共起ネットワーク図を用いて、平成26年度から平成28年度に札幌市に寄せられた要望・苦情の分析を行った。分析結果より、札幌市に多く寄せられる要望・苦情の内容を把握し、要望・苦情の大まかな分類を行うことができた。要望・苦情内容と気象条件には大きな関係があり、分析をするうえで非常に重要な関係であると考えられる。

また、札幌市の気象データの特徴量として、教師あり学習の一つである重回帰分析を用いて要望・苦情件数の予測を行った。しかし、予測精度が低かったため、機械学習の適用範囲を定めて予測モデルの構築を行った。その結果、予測精度は向上し、要望・苦情件数の傾向を再現することができた。今後は更なる精度向上に向けて特徴量や分析方法を模索し、有効な予測モデルを検討する。

文献

- 1) 札幌市 HP (2023 年 4 月 10 日閲覧): 雪対策費実績,
<https://www.city.sapporo.jp/kensetsu/yuki/library/budget.html>
- 2) 杉原幸信ら, 2022: 札幌市の道路の除排雪に対する住民の要望・苦情のテキストマイニング分析, 日本雪工学会論文集, 83-95.
- 3) 石川大貴ら, 2022: 道路除排雪に対する住民の要望・苦情のテキストマイニング分析, 日本雪氷学会北信越支部総会および研究発表会・製品発表検討会予稿集, 14p.
- 4) 気象庁ホームページ (2023 年 4 月 10 日閲覧): 過去の気象データ検索,
https://www.data.jma.go.jp/obd/stats/etm/index.php?prec_no=14&block_no=47412&year=2000&month=2&day=25&view=
- 5) scikit-learn (2023 年 4 月 10 日閲覧)
<https://scikit-learn.org/stable/>