

アンケートの単語の出現傾向による  
酒田市の生活満足度分析

広瀬研究室 3年 C1172313 佐藤利紀

## 概要

近年、都市部への人口の集中、少子高齢化などの影響により地方では過疎化が進んでおり、地方自治体は対策を求められている。そこで現状の把握や課題の認識、地方の魅力再発見などを行う手段の一つとしてアンケートがあげられる。アンケートの回答形式の中には選択式と自由記述があり、定性的でかつ未知の情報を得るためには自由記述が適している。しかし自由記述の回答の集計は人手で行われることが多く、時間がかかってしまう。そこで本研究ではテキストマイニングを利用し、アンケートの自由記述の回答内容を概観する。具体的には本学生を対象に酒田市での生活に関するアンケートをとり、その内容をテキストマイニングする。その過程で記述者の酒田市における生活満足度の高低で回答内容を分類することで傾向を比較する。回答内容の傾向を調べ考察することで、酒田市のブランディングや生活環境改善を試みた。その結果、酒田市の街のイメージについての質問では生活満足度の高低どちらのグループにおいても自然に関する単語が頻出した。同様に酒田市の自然環境についての質問では風に関する単語、酒田市における不満についての質問では交通機関に関する単語が頻出し、これら3つの項目における酒田市の特徴を示した。(520文字)

# 目次

第1章	はじめに	5
第2章	テキストマイニングと先行研究	7
2.1	テキストマイニングの概要	7
2.1.1	形態素解析の概要	7
2.1.2	テキストマイニングのツール	7
2.2	自由記述におけるテキストマイニングの先行研究	7
第3章	アンケートの分析	9
3.1	アンケートの概要	9
3.2	分析方法	9
3.3	アンケートの集計結果	10
3.4	自由記述の形態素解析	10
3.5	街のイメージについて	10
3.6	自然環境について	11
3.7	不満について	11
第4章	今後の展望	13



## 第1章 はじめに

近年、晩婚化や団塊の世代の高齢化などによって少子高齢化が進んでいる。これに加え都市部への人口の集中が地方の過疎化を加速させている。それが著しい地域の一つに山形県があげられる。山形県の人口増減率は2017年から2018年にかけて-1.04%を記録しており、これは全国で下から6番目の値となっている[1]。中でも過疎化が顕著にあらわれているのが酒田市だ。酒田市の人口は年々減少傾向にあり、なおかつ老年人口の割合は増加している[2]。そのため、酒田市は典型的な少子高齢化社会といえる(図1.1 図1.2)。

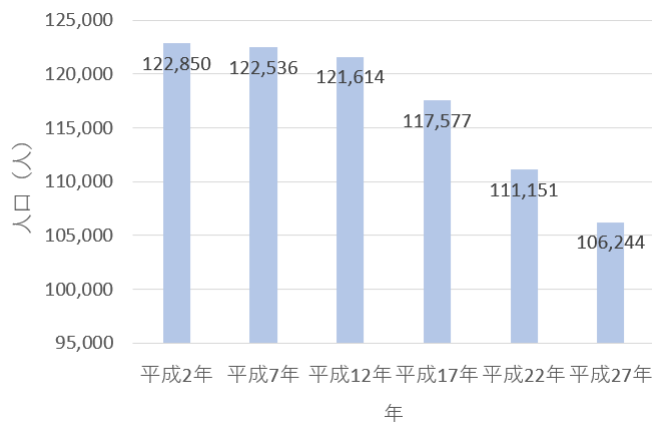


図 1.1: 酒田市の人口の推移 (総人口)

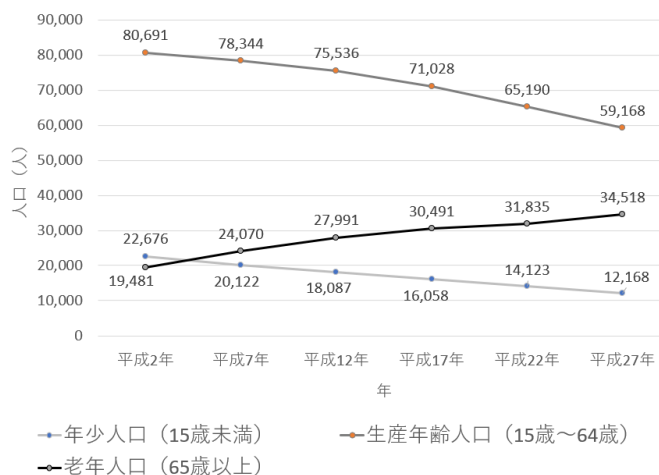


図 1.2: 酒田市の人口の推移 (年齢人口別)

そこで酒田市は多様なライフスタイルの受け入れや市街地の生活支援などを方針とした施策を講じている [3]。しかし、この施策は高齢者が安心して暮らせる居住環境の整備や生活を支える機能の強化など高齢者向けのものが多く散見される。平成 28 年に行われた酒田市公共施設等に関する市民アンケートでは、アンケートの回収数 868 のうち 18 歳から 29 歳までの回答者のサンプルデータは 67 しかなく、若者の意見が正当に反映されていない可能性がある [4] (表 1.1)。そこで本研究では、酒田市在住の東北公益文科大学の学生 (以下本学生) を対象に酒田市での生活に関するアンケートをとる。そして生活満足度の数値と自由記述の内容をテキストマイニングを用い分析する。若者の視点による酒田市のイメージや不便な点を明らかにし、酒田市のブランディングや生活環境改善へとつなげることを試みる。

表 1.1: 平成 28 年酒田市公共施設等に関する市民アンケートの年齢別回収数

年齢	回答数
18-29 歳	67
30-39 歳	102
40-49 歳	163
50-59 歳	206
60-69 歳	294
70 歳以上	33
無回答	3
合計	868

## 第2章 テキストマイニングと先行研究

本章ではテキストマイニングの概要および先行研究について述べる。

### 2.1 テキストマイニングの概要

テキストマイニングとは形式化されていない大量のテキストデータを分析して有用な情報を抽出することである [5][6]。これによりテキストデータ処理の効率化や新たな知見を得る試みなどがなされている。しかし、日本語の多くは分かち書きされていないため、何らかの方法で形式化し分ける必要がある。そこで本研究ではその手段の一つである形態素解析を利用し、テキストデータの分析を行う。

#### 2.1.1 形態素解析の概要

形態素解析とは対象の品詞情報や文法などをもとに、文章を単語の最小単位である形態素に分解することである [5][6]。日本語などの分かち書きがされない言語をテキストマイニングする際に利用され、形態素や品詞情報、出現頻度などの情報を取得することができる。

#### 2.1.2 テキストマイニングのツール

テキストマイニングを行うツールは JUMAN<sup>1</sup>や cabocha<sup>2</sup>など多様にある。本研究では R<sup>3</sup>を通して形態素解析ツール MeCab<sup>4</sup>を操作する RMeCab というパッケージを使い、アンケートの自由記述の形態素解析を行う。

### 2.2 自由記述におけるテキストマイニングの先行研究

アンケートの自由記述の分析にテキストマイニングを用いた研究として小林らの研究があげられる [7]。これはアンケートの自由記述の内容から、生活環境の評価や地域の課題を見つけることを目的としている。具体的には記述内容を一文ごとに「プラス評価」、「マイナス評価」などに分類し、相関分析を行っている。しかし、「プラス評価」、「マイナス評価」の分類が研究者の主観的なものとなっている。そこで本研究では生活満足度の高いグループ、低いグループそれぞれの記述内容における頻出の単語を主要な単語とし、客観性の高い研究を目指す。

---

<sup>1</sup><http://nlp.ist.i.kyoto-u.ac.jp/>

<sup>2</sup><http://taku910.github.io/cabocha/>

<sup>3</sup><https://www.r-project.org/>

<sup>4</sup><http://code.google.com/p/mecab/>





## 第3章 アンケートの分析

本章ではアンケートの概要および分析方法について述べる。

### 3.1 アンケートの概要

アンケートは今年度開講していた「プロジェクト型応用演習(学生を対象にアンケートを実施しよう)」のデータを使う。分析で使用する質問項目は以下のとおりである。

- 現在どこに住んでいますか。(1. 酒田市内、2. 酒田市外・山形県内、3. 山形県外)
- あなたは酒田市での生活に満足していますか。(1. まったく満足していない、2. 満足していない、3. 満足している、4. とても満足している)
- あなたは酒田市をどのような街だと思えますか。自由に記入してください。
- あなたは酒田市の気候や自然環境についてどのようなことを思えますか。自由に記入してください。
- あなたは酒田市で何か不便を感じることはありますか。自由に記入してください。

「現在どこに住んでいますか。」という質問は酒田市で生活している人のみを対象にするための項目、「あなたは酒田市での生活に満足していますか。」という質問は生活満足度の高低を分類する項目である。それ以外は回答が自由記述による質問であり、回答内容のばらつきを抑えるためそれぞれ「街のイメージ」、「自然環境」、「不満」を問う3つの項目を設けた。これにより回答内容の偏りを作り、類似した意見を集めることで単語の出現傾向による差異を見出しやすくした。

### 3.2 分析方法

まず酒田市に在住しているか否かでフィルタリングする。酒田市に在住している人の中で生活満足度の高低を分類する。これは生活満足度の高低における傾向やその差異を比較するためである。前述の「あなたは酒田市での生活に満足していますか。」という質問に対し「1. まったく満足していない」、「2. 満足していない」と答えた人を生活満足度の低いグループ、「3. 満足している」、「4. とても満足している」と答えた人を生活満足度の高いグループとみなす。そして生活満足度の高低それぞれのグループ群で自由記述の内容を形態素解析する。そして単語の出現頻度を求め、頻出の単語から生活満足度の高低における傾向や特徴を探る。

### 3.3 アンケートの集計結果

アンケートの集計結果について述べる。集計の結果、アンケートの回収数は165、有効回答数は135となった。そのうち生活満足度の高いグループは70、生活満足度の低いグループは65である。これらの回答内容のデータをもとに分析を行う。

### 3.4 自由記述の形態素解析

本節では RMeCab を使い自由記述の回答内容を生活満足度の高低および質問ごとに形態素解析した結果について述べる。解析結果において、助詞や助動詞など単語で意味をなさない品詞の単語を除外することで、出現頻度による傾向を見出す。

### 3.5 街のイメージについて

まず「街のイメージ」を表す「あなたは酒田市をどのような街だと思いますか。」という質問の回答内容について述べる。生活満足度の高低ごとに記述内容を形態素解析し、単語の品詞情報と出現頻度を分析する。その結果をそれぞれ表3.1、表3.2に示す。

表 3.1: 酒田市のイメージに関するアンケート回答結果の頻出上位 10 単語 (生活満足度の高いグループ)

単語	品詞	品詞小分類	出現頻度
街	名詞	一般	16
自然	名詞	形容動詞語幹	10
強い	形容詞	自立	7
人	名詞	一般	7
豊か	名詞	形容動詞語幹	7
多い	形容詞	自立	6
町	名詞	一般	6
まち	名詞	一般	5
いい	形容詞	自立	4
おいしい	形容詞	自立	4

表 3.2: 酒田市のイメージに関するアンケート回答結果の頻出上位 10 単語 (生活満足度の低いグループ)

単語	品詞	品詞小分類	出現頻度
ない	形容詞	自立	11
強い	形容詞	自立	8
街	名詞	一般	8
風	名詞	一般	7
自然	名詞	形容動詞語幹	6
多い	形容詞	自立	5
車	名詞	一般	5
町	名詞	一般	5
田舎	名詞	一般	5
不便	名詞	形容動詞語幹	5

表 3.1 においては「自然」、「豊か」などの単語、表 3.2 においては「風」、「田舎」などの単語が頻出していることから、酒田市は自然に関するイメージが強いことを示唆してい

る。そのため、自然に着目したブランディングの可能性が考えられる。また表 3.1 において「おいしい」などの単語が出現していることから、食文化のブランディングへの活用可能性も考えられる。

### 3.6 自然環境について

次に「自然環境」を表す「あなたは酒田市の気候や自然環境についてどのようなことを思いますか。」という質問の回答内容について述べる。前述と同様に形態素解析した結果をそれぞれ表 3.3、表 3.4 に示す。

表 3.3: 酒田市の自然環境に関するアンケート回答結果の頻出上位 10 単語 (生活満足度の高いグループ)

単語	品詞	品詞小分類	出現頻度
風	名詞	一般	50
強い	形容詞	自立	45
寒い	形容詞	自立	7
冬	名詞	一般	6
つよい	形容詞	自立	4
夏	名詞	一般	4
雨	名詞	一般	3
雪	名詞	一般	3
つらい	形容詞	自立	2
激しい	形容詞	自立	2

表 3.4: 酒田市の自然環境に関するアンケート回答結果の頻出上位 10 単語 (生活満足度の低いグループ)

単語	品詞	品詞小分類	出現頻度
風	名詞	一般	37
強い	形容詞	自立	31
つよい	形容詞	自立	5
天気	名詞	一般	5
多い	形容詞	自立	4
海	名詞	一般	4
山	名詞	一般	4
さむい	形容詞	自立	3
ない	形容詞	自立	3
雷	名詞	一般	3

表 3.3、表 3.4 のどちらにおいても「風」、「強い」という単語が頻出している。同様の単語が表 3.2 にも頻出していることから、酒田市は強風が特徴的な気候であり、なおかつそれが低い生活満足度の一要因であると考えられる。

### 3.7 不満について

次に「不満」を表す「あなたは酒田市で何か不便を感じることはありますか。」という質問の回答内容について述べる。こちらも前述と同様に形態素解析した結果をそれぞれ表 3.5、表 3.6 に示す。

表 3.5: 酒田市の不満に関するアンケート回答結果の頻出上位 10 単語 (生活満足度の高いグループ)

単語	品詞	品詞小分類	出現頻度
ない	形容詞	自立	19
交通	名詞	一般	14
少ない	形容詞	自立	6
車	名詞	一般	6
店	名詞	一般	6
便	名詞	一般	6
多い	形容詞	自立	4
機関	名詞	一般	4
悪い	形容詞	自立	3
バス	名詞	一般	3

表 3.6: 酒田市の不満に関するアンケート回答結果の頻出上位 10 単語 (生活満足度の低いグループ)

単語	品詞	品詞小分類	出現頻度
少ない	形容詞	自立	16
ない	形容詞	自立	15
交通	名詞	一般	12
店	名詞	一般	10
遠い	形容詞	自立	7
バス	名詞	一般	5
不便	名詞	形容動詞語幹	5
機関	名詞	一般	4
車	名詞	一般	4
公共	名詞	一般	3

表 3.5、表 3.6 のどちらにおいても「交通」、「車」、「バス」などの単語が頻出している。これは酒田市の交通面に関する不満を抱いている人が多いことを示唆している。しかし、酒田市立地適正化計画によると酒田市は歩いても暮らせるまちとしての基盤が整っていると述べられている [3]。この見解は今回のアンケート結果とは異なり、本学生との見識の相違がある。そのため学生の視点による意見が反映されていない可能性があると考えられる。

## 第4章 今後の展望

本研究ではアンケートの自由記述の内容を生活満足度の高低ごとに形態素解析し、その傾向から地域のブランディングの可能性を示した。今後の課題としては3つのことがあげられる。1つはアンケートの改善だ。本研究ではアンケートの回収数165に対して有効回答数135となった。有効回答の割合を向上させるために回答者の心理的に作用させるバイアスなどを考慮したアンケートの構成や文章作りが必要であるとする。2つ目は分析結果の統計処理だ。本研究では出現頻度などの数値を用いることで研究の客観性を高めた。しかし考察は筆者の主観にとどまってしまった。そのため多変量解析や検定などの統計的手法を利用した分析結果の妥当性の検証が必要である。3つ目はほかのテキストマイニング技術の複合だ。本研究では自由記述のテキストマイニングに形態素解析を利用したが、単語のみでは正確な考察は難しいと考えられる。たとえば自由記述から「自然」という単語を抽出したとする。この1単語だけでは「自然」に対してどのような印象や感情を抱いているかが判断できない。そのため単語同士の関係性をみる係り受け解析や共起などのテキストマイニング技術を組み込み、分析の正確性を高める必要がある。



## 参考文献

- [1] 総務省 . 人口推計 (2018 年 (平成 30 年)10 月 1 日現在). 総務省統計局 . <http://www.stat.go.jp/data/jinsui/2018np/index.html>
- [2] 酒田市 . 統計資料 . 酒田市公式ウェブサイト . <http://www.city.sakata.lg.jp/shisei/toukei/toukeishiryou/index.html>
- [3] 酒田市 . 立地適正化計画 . 酒田市公式ウェブサイト . [https://www.city.sakata.lg.jp/jyutaku/toshikeikaku/tokeimaster\\_tekisei/tekiseikakeikaku/ritteki0329.html](https://www.city.sakata.lg.jp/jyutaku/toshikeikaku/tokeimaster_tekisei/tekiseikakeikaku/ritteki0329.html)
- [4] 酒田市 . 公共施設等に関する市民アンケート結果 . 酒田市公式ウェブサイト . <http://www.city.sakata.lg.jp/shisei/kaikaku/shisetsutekiseika/koukyousisetu.html>
- [5] 石田基広,R によるテキストマイニング入門 第 2 版, 森北出版,2017
- [6] 小林雄一郎,R によるやさしいテキストマイニング, オーム社,2017
- [7] 小林祐二, 寺田充伸, 佐藤誠治. テキストマイニングを活用したアンケートにおける自由回答の分析と生活環境評価. 日本建築学会計画系論文集 第 77 巻 第 671 号,pp.85-93,2012
- [8] 西尾和恭, 岡野拓巳, 芝千尋, 戸次陸, 須子統太. アンケートデータにおける選択バイアス補正法に関する一考察. 第 81 回全国大会講演論文集 2019(1),pp.687-688,2019