
市民の声のテキストマイニング分析

－ 自治体世論調査の自由記述データを事例にして －

公共コミュニケーション学会

第2回事例交流・研究発表大会

自治体広報広聴研究所

行政広報アドバイザー

金井 茂樹

目次

1. 研究の背景と目的

2. 市民の声の分析の現状

- (1)実務における市民の声の収集の枠組み
- (2)報告書における市民の声の分析事例
- (3)テキストマイニングによる市民の声の分析事例

3. 分析

- (1)対象データ:自治体世論調査で得られた自由記述データ
- (2)コレスポンドンス分析による課題抽出
- (3)テキストマイニング分析

4. 結論

1. 研究の背景と目的

■ 自治体の広聴活動は半世紀以上の歴史

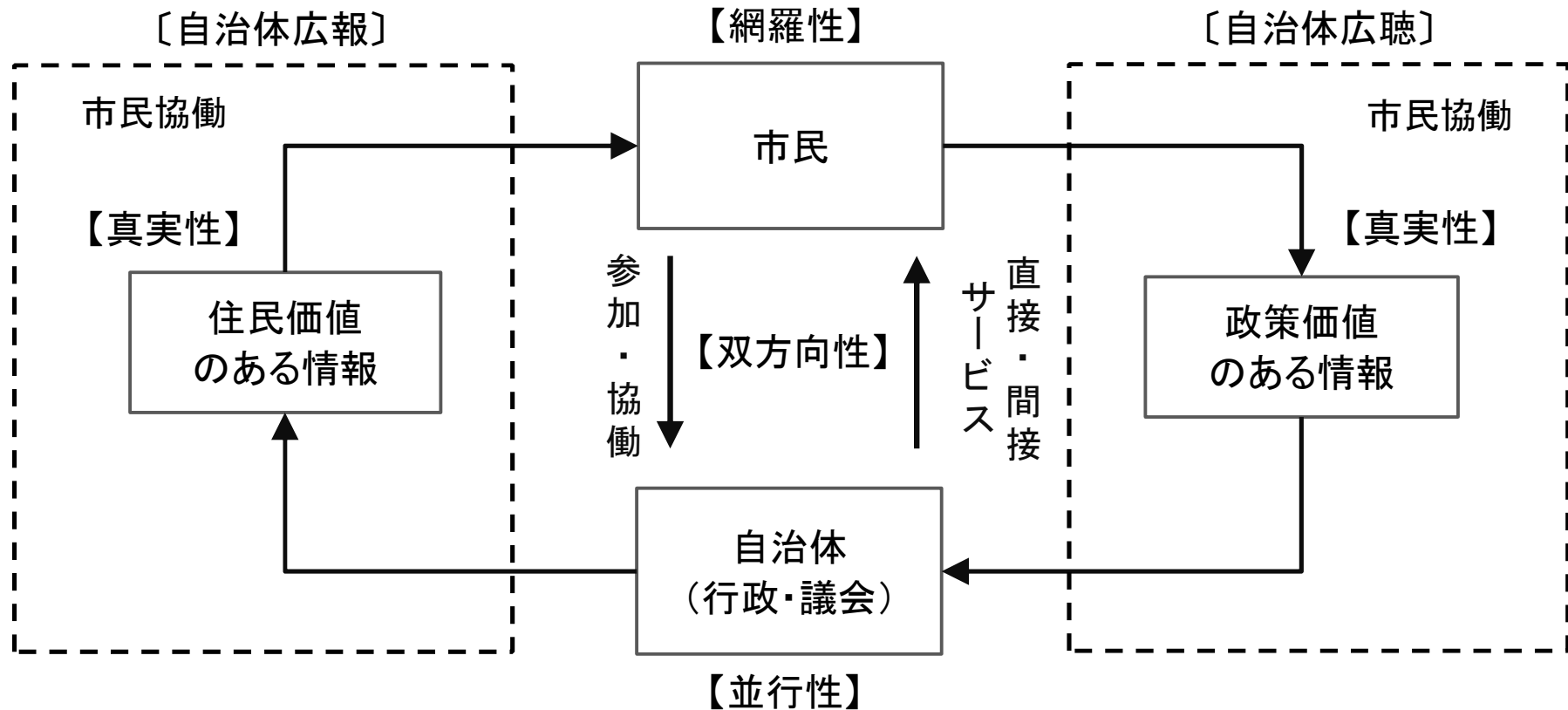
- 世論調査・対話集会・個別広聴
- 自治体の広聴活動は行政施策に特段の影響をもたらすことなく、首長や行政の「聴く姿勢」をアピールすることにとどまり、世論調査や住民意識調査も政策形成や行政評価に活用されていないとの批判もなされてきた(上野,2000)
- 市民の声を無視した政策形成はありえない
- 迅速な対応／市民の声の個別性や代表性の問題／恣意的判断

■ 研究の目的

- テキストマイニング分析を活用して政策形成に有用な情報の手がかりを得ること

2. 市民の声の分析の現状

■ 行政広報の基礎フレーム



広聴活動

⇒ 収集した情報から政策価値のある情報を創出すること

2. 市民の声の分析の現状

■ 実務上の分類

収集態度 情報の 特性等	受動的活動 住民が主体		能動的活動 自治体が主体
実務上分類	個別広聴	集団広聴	調査広聴
手 法	面談、手紙・はがき、 電話、メールなど	対話集会、懇談会、 モニター制など	世論調査・意識調査 ウェブ調査など
情報の性質	【集まるデータ】 個別的な意見・要望・苦情など		【集めるデータ】 提示した問題群に対する意 見構造分布
データ形式	主にテキストデータ		主に数値データ 一部テキストデータ
傾 向	情報の量的増加		情報の質的低下
問 題 点	意見の個別性・代表性		調査票の制約

2. 市民の声の分析の現状

■ 報告書における市民の声の分析事例

- 東京23区足立区『区民の声報告書』『足立区政に関する世論調査報告書』
 - ⇒性質別(苦情・不満・要望・意見・質問・感謝)による分類分析
 - ⇒政策別(約70項目)による分類分析

■ テキストマイニングによる市民の声の分析事例

- 近藤(2003)、仙台都市総合研究機構(2003)
 - ⇒集まる市民の声を分析対象
 - ⇒回答者の性別や年齢といった属性情報には欠落が多い
 - ⇒議論のための素材として有用な情報
- 金井(2015)
 - ⇒自治体世論調査で集めた情報の分析
 - ⇒自由記述データについてテキストマイニング(クラスタリング)分析によりデータの構造を明らかにした

3. 分析

■ 分析対象データ

調査名	A自治体世論調査
測定方法	郵送配付／訪問回収・郵送回収
抽出方法	無作為抽出法
標本数	2,000件
回収数（率）	1,512件（75.6%）

○現況評価

個別評価：①買い物、②交通、③道路・歩道、④公園・水辺、⑤緑化、⑥景観、⑦バリアフリー、⑧ごみ、⑨リサイクル、⑩治安・安全性、⑪災害への備え、⑫近所づきあい、⑬文化施設、⑭スポーツ施設、⑮熟年者の生活環境、⑯子育て環境、⑰健康相談・健康診査、⑱病院・診療所の利便性・整備状況

総合評価：A自治体の現況全体について

評価尺度：「満足」「やや満足」「ふつう」「やや不満」「不満」「わからない」

3. 分析

○個別評価項目

項目 \ 評価	満足	やや満足	ふつう	やや不満	不満	わからない
ア. 買い物の便	1	2	3	4	5	6
イ. 交通の便	1	2	3	4	5	6
ウ. 道路・歩道の整備	1	2	3	4	5	6
エ. 公園・水辺の整備	1	2	3	4	5	6
オ. 緑化の推進	1	2	3	4	5	6
カ. 街の景観	1	2	3	4	5	6
キ. 街のバリアフリー	1	2	3	4	5	6
ク. ごみの収集	1	2	3	4	5	6
ケ. リサイクルの推進	1	2	3	4	5	6
コ. 地域の治安・安全性	1	2	3	4	5	6
サ. 災害への備え	1	2	3	4	5	6
シ. 近所づきあい	1	2	3	4	5	6
ス. 文化施設の充実度	1	2	3	4	5	6
セ. スポーツ施設の充実度	1	2	3	4	5	6
ソ. 熟年者の生活環境	1	2	3	4	5	6
タ. 子育て環境	1	2	3	4	5	6
チ. 健康相談・健康診査	1	2	3	4	5	6
ツ. 病院・診療所の充実度	1	2	3	4	5	6

3. 分析

○自由記述回答の項目

あなたが、A自治体で暮らしているなかで、A自治体の「良さ」または「改善してほしい」と感じるのは、どのようなことですか。子育て支援、熟年者施策、生活環境、産業振興、災害対策、まちづくりなど、どのような分野でもかまいません。できるだけ具体的にお書きください。

「良さ」

「改善してほしい点」

3. 分析－分析手順

○数値データの分析(現況評価)

個別評価18項目から得られた数値データに対してコレスポネンス分析(対応分析)を行い、政策課題を抽出する。

○テキストデータの分析(「改善」に関する自由記述)

抽出された政策課題に対する評価が高い市民と低い市民の「改善」に関する特徴的な語句を明らかにして、政策形成に有用な情報の手がかりを得る。

政策課題に対する評価	「良さ」 自由記述データ	「改善」 自由記述データ
評価が高い市民	—	分析
評価が低い市民	—	分析

3. 分析－政策課題の抽出

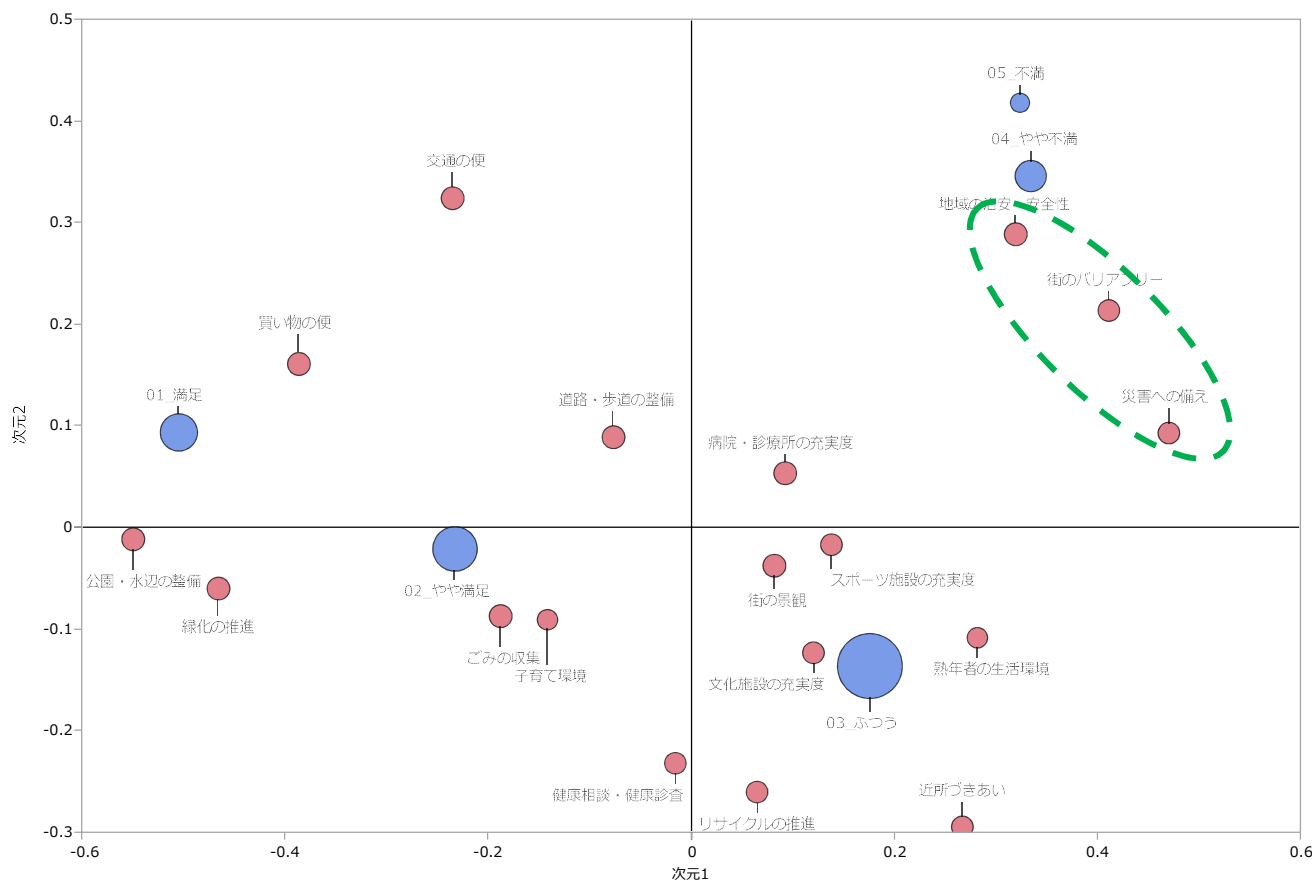
A自治体の現況に対する18項目の評価(18×5のクロス表)の対応分析

	01_満足	02_やや満足	03_ふつう	04_やや不満	05_不満	06_わからない
買い物の便	455	369	480	145	51	4
交通の便	391	355	439	213	96	6
道路・歩道の整備	269	365	615	175	67	13
公園・水辺の整備	456	471	469	56	19	33
緑化の推進	397	479	518	51	23	34
街の景観	201	297	758	159	61	25
街のバリアフリー	91	190	703	308	84	124
ごみの収集	319	355	683	102	29	19
リサイクルの推進	189	250	827	80	24	133
地域の治安・安全性	125	250	659	291	132	48
災害への備え	61	148	751	231	96	217
近所づきあい	104	213	934	103	35	114
文化施設の充実度	137	304	733	140	37	148
スポーツ施設の充実度	138	299	678	166	57	162
熟年者の生活環境	85	200	700	132	56	327
子育て環境	191	375	540	85	34	265
健康相談・健康診査	183	313	728	75	21	183
病院・診療所の充実度	177	331	681	177	83	59

3. 分析－コレスポンス分析

■ 政策課題の抽出

現況18項目の評価(18×5のクロス表)に対するコレスポンス分析



- ＜政策課題＞
- ・治安・安全性
 - ・バリアフリー
 - ・災害対策

固有値の数は、 $K = \min\{18\text{項目}, 6\text{階級}\} - 1 = 5$ (個)まで得られる。累積寄与率から、はじめの2成分で全情報(総変動)の約95%を占める。

3. 分析－自由記述データ

■ 回答者の特徴

全回答者1,512名のうち自由記述回答数

- ・「良さ」 = 670人(全体の44.3%)
- ・「改善」 = 735人(全体の48.6%)
- ・性別の記入率 男性47.6% 女性52.4%
- ・年齢別の記入率 40歳代(53.4%) 30歳代(49.8%) 60歳代(49.12%)
50歳代(49.1%)

■ 「改善」についての自由記述データの特徴

出現頻度が「1」以上つまり全構成要素数が23,992(語)

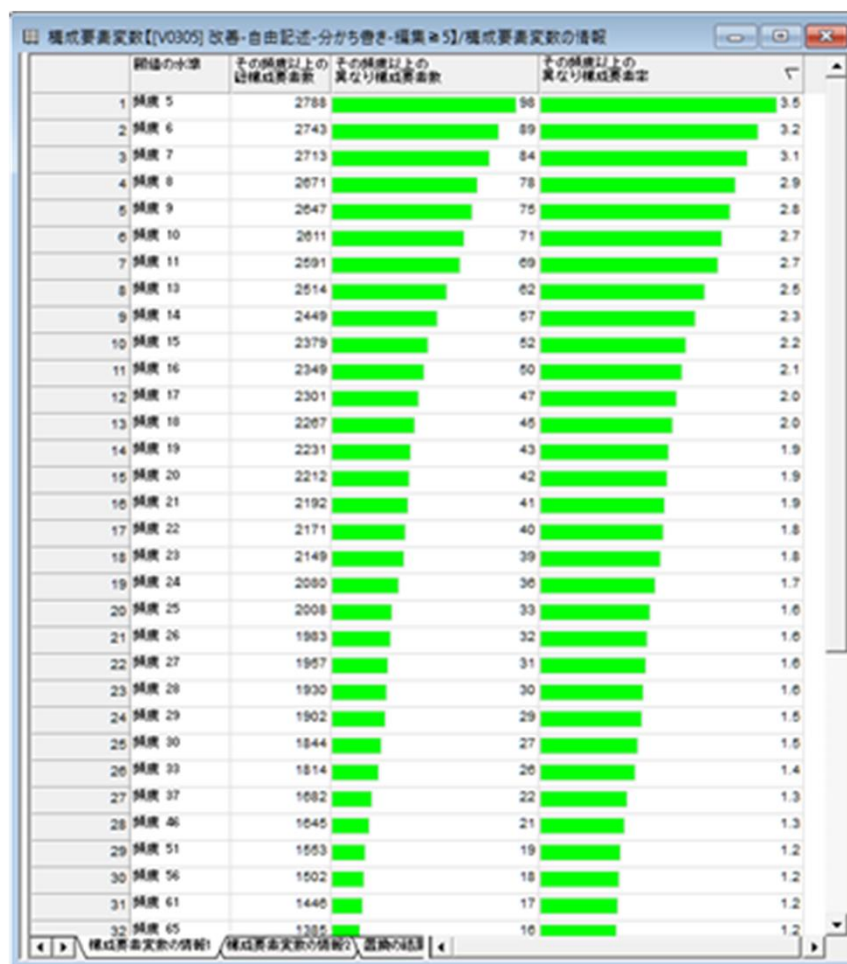
異なり構成要素数が3,875(語)

異なり構成要素率が、 $3,875 \div 23,992 = 16.2(\%)$

ここでの異なり構成要素率は、回答者が記述した語句の重複の程度を示す1つの指標と考えられ、この値が大きいほど同じ語句が繰り返し使われたことになる。

3. 分析－対象とする構成要素（語句）

ここでは、閾値を5と指定した。つまり出現頻度が4以下の構成要素は除外。
頻度5以上の全構成要素数(3,875語)に対して異なり構成要素数は98語



<特徴的な語句>

「子ども」107回

「自転車」93回

「災害対策」88回

「バス」75回

「交通ルール」73回

「リサイクル」「呼び込み」「防犯カメラ」5回

「学力」「景観」6回

頻度分布(5語以上)

3. 分析－構成要素×質的変数(1)

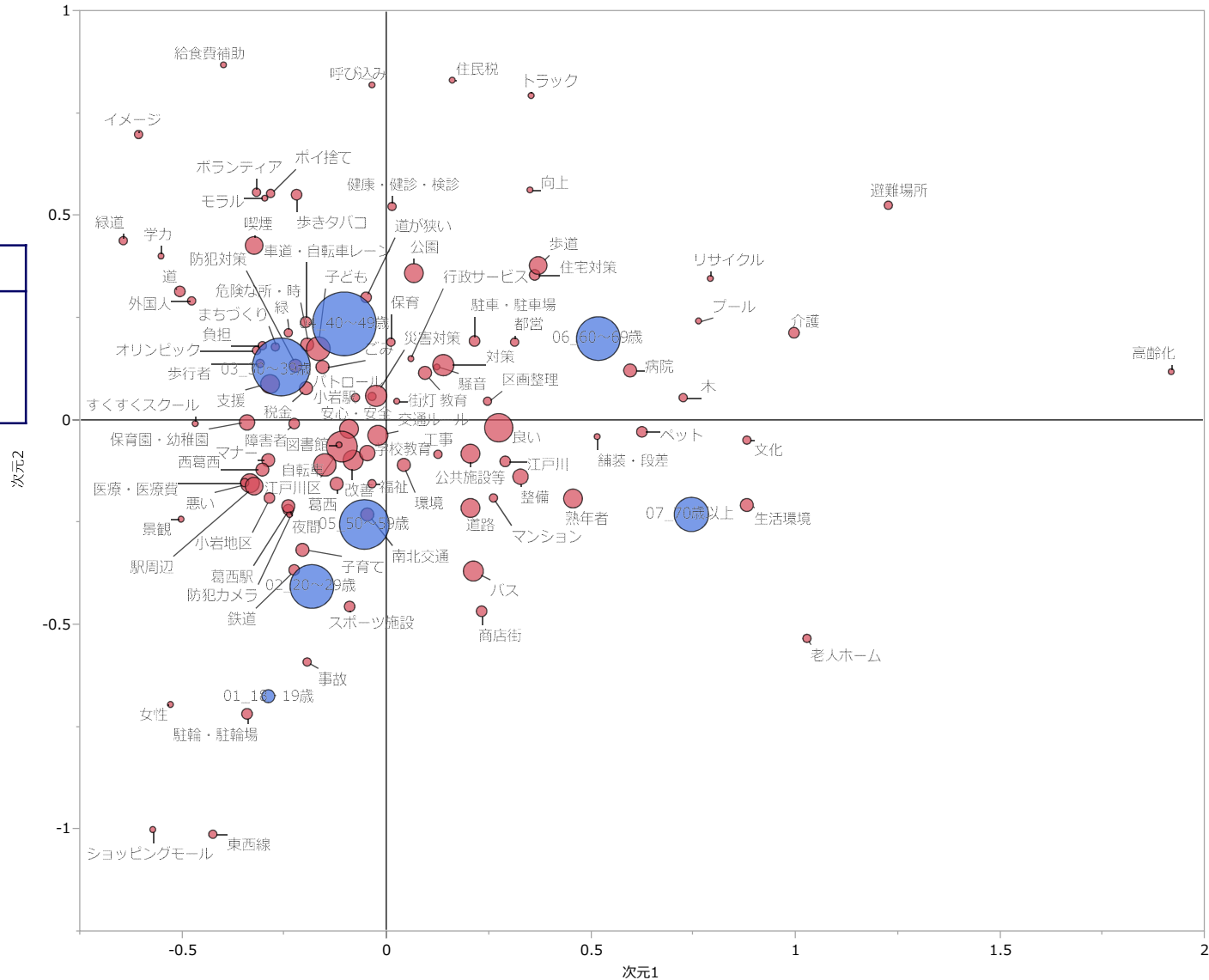
構成要素×質的変数＝98語の構成要素×年齢区分(7階級)の2元データ表(98×7)にテキストマイニング分析を適用した。

<質的変数>

18・19歳 20代 30代 40代 50代
60代 70以上

	← 年齢階級(7階級) →
↑ 構成要素数 (98語) ↓	98×7

固有値の数は、 $K = \min\{98\text{語}, 7\text{階級}\} - 1 = 6$ (個)まで得られる。累積寄与率から、はじめの2成分で全情報(総変動)の約50.8%を占める。



3. 分析－年代別の特徴的な語句

■ 年代別の特徴的な語句

18・19歳	20歳代	30歳代	40歳代	50歳代	60歳代	70歳以上
悪い	子育て	ごみ	給食費補助	学校教育	避難場所	生活環境
A自治体	駐輪・駐輪場	保育園・幼稚園	公園	整備	歩道	老人ホーム
バス	スポーツ施設	喫煙	防犯対策	自転車	良い	熟年者
改善	ショッピングモール	ポイ捨て	モラル	防犯カメラ	プール	バス
子育て	鉄道	緑道	健康・健診・検診	税金	高齢化	介護
行政サービス	環境	イメージ	道	商店街	住宅対策	病院
		支援	子ども		都営	ペット
		危険な所・時			介護	商店街

3. 分析—構成要素×質的変数(2)

■ テキストマイニング分析

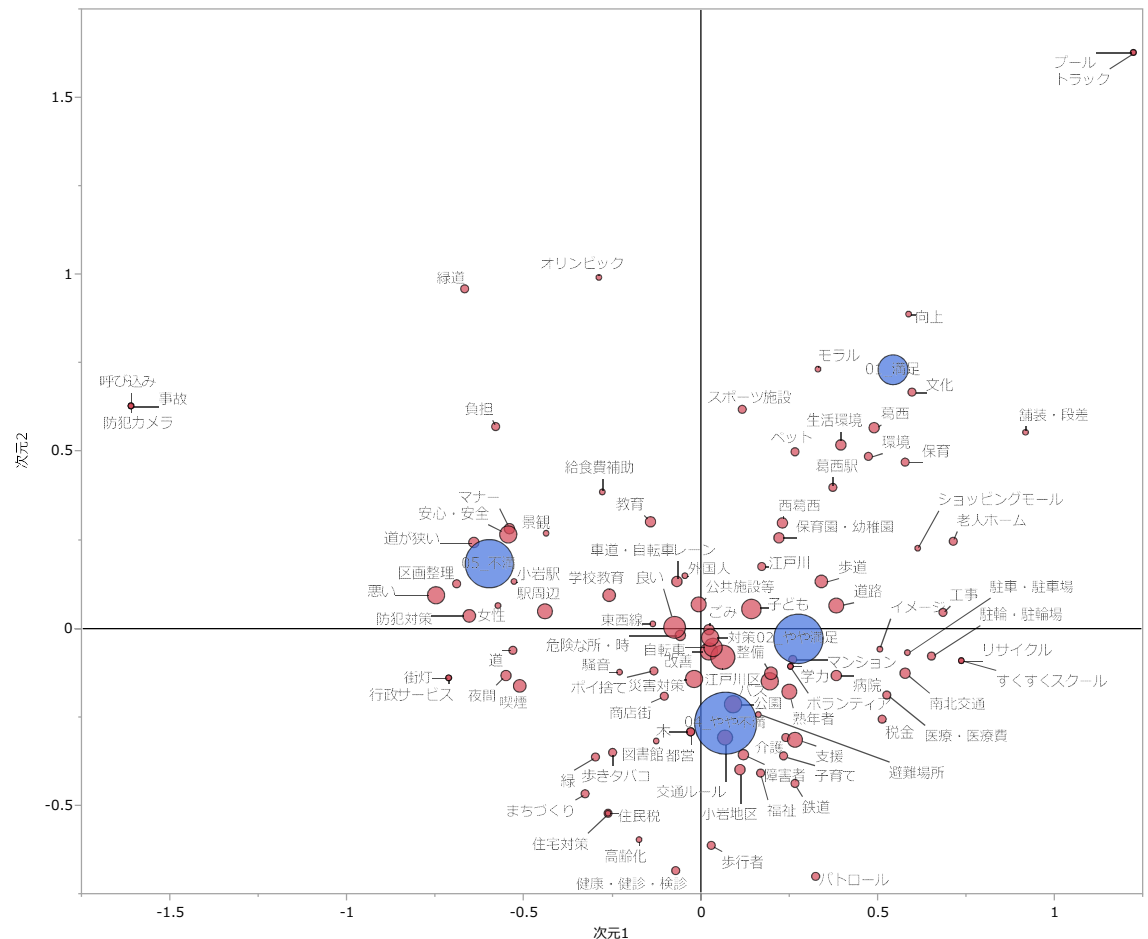
政策課題に対する評価別の語句の特徴を明らかにするために、構成要素×質的変数=98語の構成要素×政策課題に対する評価(5階級)の2元データ表(98×5)にテキストマイニング分析を行った。

<質的変数>

満足 やや満足 ふつう やや不満 不満

	←地域の治安・安全性への評価(5階級)→
↑ 構成要素数 (98語) ↓	98×5

固有値の数は、 $K = \min\{98\text{語}, 5\text{階級}\} - 1 = 4$ (個)まで得られる。累積寄与率から、はじめの2成分で全情報(総変動)の約63.3%を占める



3. 分析－治安・安全性評価別の特徴的な語句

■ 治安・安全性評価別の特徴的な語句

満足	やや満足	ふつう	やや不満	不満
プール	学力	環境	高齢化	安心・安全
文化	駐輪・駐輪場	子育て	まちづくり	緑道
生活環境	保育	保育園・幼稚園	パトロール	防犯対策
トラック	景観	税金	支援	呼び込み
向上	向上	歩道	介護	マナー
	ボランティア	駐車・駐車場	歩行者	区画整理
		歩きタバコ	緑	道が狭い
				夜間
				景観
				駅周辺

⇒これらの語句を手かがりに自由記述の原文の内容を検討する

⇒735人の自由記述データの効率的な分析が行える

3. 分析—地域の治安・安全性に関する要望（仮説）

- 治安・安全性に関する自由記述回答からみる改善要望（仮説）
 - 夜間の駅周辺で飲食店への呼び込みなどが多く、安全への不安がある
 - その対策として駅周辺のパトロールや防犯カメラの設置を求められている
 - また、喫煙に関して、歩きタバコ、ポイ捨てなどのマナーが守られていないことや子どもへの悪影響への懸念である。
 - 評価項目のなかで評価の高かった公園・水辺の整備および緑化の推進に関連する緑道に関する改善要望がある。これは、数多くの緑道が整備されたがゆえに、その管理の不十分さを指摘するものである。

4. 結論

- 選択肢型質問と市民の声データの構成要素(語句)とを合わせたテキストマイニング分析によって、課題に対する具体的な市民の声を抽出
- 地域の課題発見・問題解決のためにはテキストデータの分析は不可欠
- 従前の分類分析とは異なる視点からの新たな情報を得ることができるのではないか
- 量的データと質的データが混在するミドルデータやビッグデータの解析に活用可能性をもつのではないか
- 自治体にとってもテキストマイニング分析は実務レベルで活用可能な効率的かつ客観的な分析手法として期待できるのではないか

5. 参考文献ほか

- 上野征洋(2000)「自治体広報広聴の現状と課題-変化から新たな時代の展望へ-」『判例地方自治』(202),ぎょうせい.
- 近藤田津(2003)「市民の声を市政に活かすためのIT活用の研究」『行政&ADP』7月号,行政情報システム研究所.
- 仙台都市総合研究機構(2003)『「市民の声」の活用法に関する調査研究』2003SURF研究報告.
- 金井茂樹(2015)「テキストマイニングによる「市民の声」の分析:自治体世論調査における自由記述データを事例にして」『自治体学』自治体学会.

統計ソフトウェア

SAS社JMP(ジャンプ) / 富士通エフ・アイ・ピー・システムズ株式会社WordMiner(ワードマイナー)