

# テキストマイニングによる「地域創生デザイン」志向潜在知識抽出に関する研究

— 札幌市南区の65歳以上住民を対象とした意識調査データを分析例として —

## A Study on Extraction of Potential Knowledge Oriented toward “Local Community Revitalization Design” by Text Mining

- An Analysis of Consciousness Survey Data Targeted Residents Aged 65 and Older in Sapporo City Minami Ward -

(キーワード：テキストマイニング，潜在知識抽出，地域創生デザイン，ビッグデータ分析)

(KEYWORDS: text mining, potential knowledge extraction, local community revitalization, big data analysis)

○城間祥之，柿山浩一郎（札幌市立大学デザイン学部）

### 1. 研究の背景と目的

テキストマイニングは、統計学、パターン認識、人工知能などのデータ解析技法を大量の文書（テキスト）データに網羅的に適用することで潜在知識を抽出する手法であり、発見的な知識獲得が期待される分析手法である。しかし、文書データをテキストマイニングソフトに入力さえすれば、自動的に潜在知識が出力されるものではない。多種多様な条件（パラメータ）を試行錯誤的に設定して分析を繰り返し、ことばの数量データやネットワーク図などで可視化し、そこから潜在知識を読み取る（抽出する）ものである。このため、問題毎にパラメータの設定変更と分析を繰り返す必要があり、分析者（データアナリスト）にはマイニング（採掘）熟練技術、すなわち、(1)大量の文書データをふるいにかける感度調整技術、(2)ふるいの中から金塊（重要なキーワードや関係性）を見つけ出す能力、(3)分析結果から重要と思われる潜在知識を読み取る（抽出する）能力など、が要求される。

平成31年（2019年）現在、大容量文書データを分析できるデータアナリストは全国的に不足しており<sup>1), 2)</sup>、ましてや地域創生を担うステークホルダー（普通地方公共団体、地域住民、地域の企画運営会社・調査会社など）にそのような人材はほとんどいないのが実態である。仮に、地域創生関係者、特に市町村の行政担当者にテキストマイニング分析手法を伝授することができれば、行政が大量に保有する住民の要望・意見データなどを担当者自らがマイニング分析することで、住民の希望や生きがい、地域愛に繋がる潜在的要望や活動欲求などを掘り起こすことが可能となり、地域創生につながる施策の手がかりを得ることが継続的に行えるようになると思われる。これが本研究を始めた動機である。

そこで本研究では、テキストマイニング分析により、住民の意識調査文書データから潜在的要望や活動欲求などを抽出する知見を得ることを目的とし、同時に、研究過程

で得られる新しい知見やマイニング分析技術を地域創生関係者へ伝授すべく解説書にまとめることを目標として、大容量文書データのテキストマイニング分析を試みる。すなわち、ここでは平成25年度 文部科学省「地（知）の拠点整備事業」に採択された札幌市立大学『ウェルネス x 協奏型地域社会の担い手育成「学び舎」事業』<sup>3)</sup>で実施された、「札幌市南区在住65歳以上高齢者の健康に関するニーズ調査」<sup>4)</sup>の集計データ（自由記述文書）をマイニング分析し、潜在的要望や活動欲求などの抽出を試みる。

### 2. 分析対象データの概要

前述のように「札幌市南区在住65歳以上高齢者の健康に関するニーズ調査」は、札幌市立大学『ウェルネス x 協奏型地域社会の担い手育成「学び舎」事業』の一環として平成24年度に実施された。当時の札幌市南区の人口は146341人で、そのうち65歳以上高齢者は37610人（約25.7%）であった。健康ニーズ調査は、郵送法による無記名自記式質問紙による調査形式で行うこととし、10%（約3700人）の回収を想定した。一般に高齢者からの回収率は高いことを考慮して、回収率を40%と推測し、調査対象人数を約9000人と算定した。選定方法は住民基本台帳からの無作為抽出とし、その作業は札幌市南区の担当者が行った。調査項目は、(1)世帯構成、(2)健康状態や健康への関心に関する項目、(3)外出や取り組んでいる活動に関する項目、(4)日常生活や困難・不安に関する項目、(5)札幌市と札幌市立大学が協働で取り組む事業に関する考えの5項目であった。そのうち5番目の項目については質問への回答を自由記述形式で収集しており、ここでは自由記述回答文のうち、以下に示す問21への回答文書データを分析対象に選定した。

問21 「旧真駒内緑小学校」の跡地を利用し、札幌市と札幌市立大学が協働してCOCキャンパス（同封資料参照）を作る予定です。南区の住民として、そこで行って欲しいことや期待することを何でもお書きください。

札幌市南区在住 65 歳以上高齢者、約 9000 人を対象に実施された健康ニーズ調査への回答者数は 2998 人 (33.3%)、問 21 への有効回答者数は 892 人 (回答者の 29.8%) であった。

### 3. テキストマイニング分析の概要

本研究では、Text Mining Studio 6.1.<sup>05)</sup>を用いて、前述の 892 人の回答文書データを以下の手順で分析した<sup>6)</sup>。

- (1) 回答文書データファイル(csv file)を Text Mining Studio へ読み込み、「分かち書き」を実行した。文章は形態素 (言語で意味を持つ最小単位) に区切り、形態素ごとの品詞が確定された分析のためのオリジナルテキストに変換された。
- (2) テキスト情報ツールを用いて、基本情報 (総行数、平均行長 (文字数)、総文章数、平均文章長 (文字数)、延べ単語数、単語種別数) や品詞出現回数など、テキストの数量データを算出した。
- (3) ビジュアル集計ツールを用いて、属性 (性別、年齢、町内会など) ごとの回答者数を算出した。
- (4) 単語頻度解析ツールを用いて単語の出現頻度を分析し、同様に、係り受け頻度解析ツールを用いて主語と述語のような係り受け表現から、回答者が記述したポジティブ (またはネガティブ) な表現事例を導出した。
- (5) 評判抽出ツールを用いて、ポジティブ (またはネガティブ) な表現に用いられた単語 (名詞や自立動詞) の頻度を導出した。
- (6) ことばネットワーク分析ツールを用いて、矢印付き有向直線で構成される係り受け関係ネットワーク図を出力した。
- (7) 対応バブル分析ツールを用いて、ことばを介した属性同士の関係図を出力した。

## 4. 分析結果

### 4.1 テキスト情報

表 1 基本情報と品詞出現回数

項目	値	品詞	回数 (%)
総行数	892	名詞	7512 (69.00%)
平均行長 (文字数)	57.5	代名詞	169 (1.55%)
総文章数	1641	動詞	1610 (14.7%)
平均文章長 (文字数)	31.3	形容詞	499 (4.58%)
延べ単語数	10887	形容動詞	333 (3.06%)
単語種別数	4296	連体詞	105 (0.96%)
		副詞	408 (3.75%)

基本情報と品詞出現回数 (表 1 参照) から以下のことが読み取れる。

- ・総行数は 892 行である。これは 892 人の回答者それぞれ

の意見 (テキスト) が回答文書データファイル(csv file) の各 1 行に記述されていることを意味する。

- ・平均行長 (文字数) は 57.5 である。これは 1 行あたりの文字数が 57.5 文字であることを意味する。平均行長を平均文章長 (文字数) で除すると約 1.84 となり、1 行あたりに 1.84 個の文章が記述され、その文章は 57.5 文字で構成されていることを表す。これは回答者が比較的簡潔に意見を記述していることを意味する。
- ・総文章数は 1641 である。総文章数を総行数で除すると約 1.84 となり、1 行あたりに 1.84 個の文章があることを意味する。
- ・平均文章長 (文字数) は 31.3 である。これは一つの文章が 31.3 文字で構成されていることを意味し、簡潔な文章 (回答) だと判断される。
- ・延べ単語数は 10887 である。これを総行数で除すると約 12.2 となり、1 行あたり 12.2 個の単語が含まれていることを意味する。
- ・単語数の内訳を品詞別出現回数で見ると、名詞が 7512 (69.00%)、代名詞が 169 (1.55%)、動詞が 1610 (14.7%)、形容詞が 499 (4.58%)、形容動詞が 333 (3.06%)、連体詞が 105 (0.96%)、副詞が 408 (3.75%) である。また形容詞と副詞を合わせた出現回数は 907 (8.33%) であり、回答文章の中に感性を表す単語は少なく、感情的な回答は少なかったと判断される。
- ・単語種別数は 4296 である。

### 4.2 属性情報

表 2 属性別集計と欠損 (属性記入漏れ) 数の関係

年齢	男性	女性	性別未記入	合計
1: 65歳~69歳	178	173	1	352
2: 70歳~74歳	138	84	0	222
3: 75歳~79歳	100	53	0	153
4: 80歳~84歳	65	40	0	105
5: 85歳以上	34	25	0	59
年齢未記入	1	0	0	1
合計	516	375	1	892

属性別集計と欠損 (属性記入漏れ) 数の関係 (表 2 参照)

から以下のことが読み取れる。

- ・男女別の回答者数は、男性が 516 人 (57.9%)、女性が 375 人 (42.1%) であり、すべての年齢区分で女性より男性の回答者が多い。
- ・年齢ごとの回答者数は、区分 1 (65 歳~69 歳) が 352 人 (39.5%) で最も多く、区分 2 (70 歳~74 歳) は、222 人 (24.9%)、区分 3 (75 歳~79 歳) は 153 人 (17.2%)、区分 4 (80 歳~84 歳) は 105 人 (11.8%)、区分 5 (85 歳以上) は 59 人 (6.6%) であり、年齢が上がるにしたがって男女ともに回答者数は漸減している。

・区分 1 (65 歳～69 歳) の回答者で性別の記入漏れが 1 人、同様に、男性で年齢区分の記入漏れが 1 人いる。

#### 4.3 気になることば“欲しい”に関する係り受け表現の抽出

係り受け表現とは、「何がどうした」や「どんな何か」のように、文章中で意味のつながりのある単語と単語の組み合わせのことである。前述の質問文は、問 21「・・・COC キャンパス・・・で行って欲しいこと・・・」と問うているので、回答文における気になることばは、“欲しい”に絞り込んで係り受け表現を抽出した。

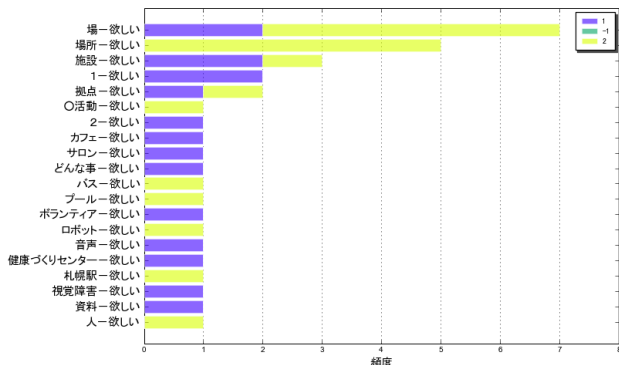


図 1 気になることば“欲しい”の係り受け表現 (男女別集計)

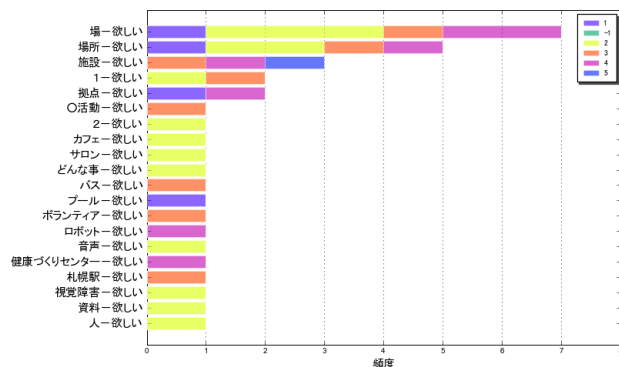


図 2 気になることば“欲しい”の係り受け表現 (年齢別集計)

気になることば“欲しい”の係り受け表現 (図 1、図 2 参照) から以下のことが読み取れる。

- ・頻度が高い係り受け表現は、“場—欲しい”、“場所—欲しい”、“施設—欲しい”であり、いずれも高齢者が活動する“場”や“場所”に関する意見 (要望) である。また、男女の比較では、これらの意見 (要望) は性別 2 (女性) から多く、特に、“場所—欲しい”は女性だけから出ており、女性の方が地域での交流や活動の場を積極的に求めていることがわかる。
- ・年齢別の比較では、“場—欲しい”や“場所—欲しい”の意見 (要望) は区分 2 (70 歳～74 歳) が最も多く、区分 4 (80 歳～84 歳)、区分 3 (75 歳～79 歳) と続いてお

り、区分 4 (80 歳～84 歳) の高齢者の積極性が特筆される。

#### 4.4 肯定的・否定的な係り受け表現の抽出

肯定的・否定的な係り受け表現 (図 3、図 4 参照) から以下のことが読み取れる。

- ・肯定的な係り受け表現は、“老人—楽しむ+できる”、“老人—参加+できる”、“自由—使う+できる”、“住民—参加+できる”などであり、高齢者が楽しく参加できる活動や自由に使うことができる場・場所について肯定的に意見 (要望) を述べている。男女別では性別 2 (女性) からの肯定的な意見 (要望) が多いが、“住民—参加+できる”は性別 1 (男性) からだけの意見 (要望) である。

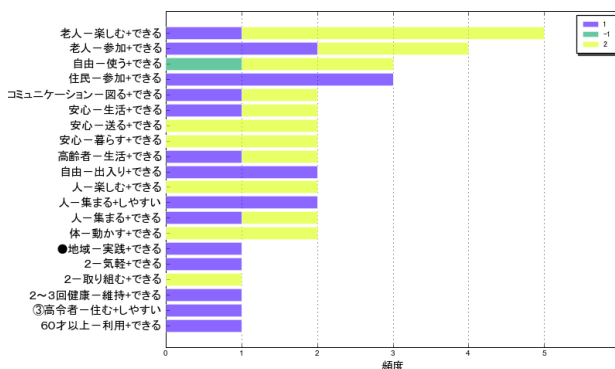


図 3 肯定的な係り受け表現 (男女別集計)

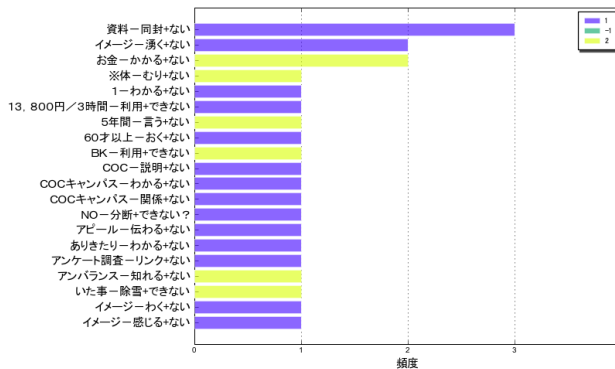


図 4 否定的な係り受け表現 (男女別集計)

- ・否定的な係り受け表現は、“資料—同封+ない”、“イメージ—湧く+ない”、“お金—かかる+ない”などである。このうち、“資料—同封+ない”、“イメージ—湧く+ない”は郵送されてきた封筒に札幌市立大学『ウェルネス x 協奏型地域社会の担い手育成「学び舎」事業』や COC キャンパスについての説明資料がない (実際には説明書類は同封されているが見つけれなかったものと推測される) ことに対する否定的な意見 (要望) である。その他の否定的な係り受け表現もほとんど同様である。一方、“お金—かかる+ない”は、お金がかからないことを表現した意見 (要望) であり、否定的な係り受け表現に分類され

ているが、否定的な意見（要望）ではなく、むしろ肯定的な意見と捉えるべきである。

#### 4.5 好評語・不評語の抽出

好評語・不評語ランキング（図5、図6参照）から以下のことが読み取れる。

- 抽出されたことば（単語）に関する肯定的な意見（要望）は、全体的に男女でほぼ同等であるが、“地域”や“生活”に関する肯定的な意見（要望）は男性が圧倒的に多い。一方、“場所”、“利用”、“活動”に関する肯定的な意見（要望）は性別2（女性）に多く、“提供”に関する肯定的な意見（要望）は女性だけである。女性は社交的な場や活動に積極的であると判断される。
- 抽出されたことば（単語）に関する否定的な意見（要望）は、全体的に男性が多い。特に、“人”、“参加”、“場”、“南区”、“活動”、“交流”に関する否定的な意見（要望）は性別1（男性）だけであり、男性は社交的な場や活動に消極的であると判断される。一方、“生活”と“高齢者”に関する否定的な意見（要望）は女性だけである。

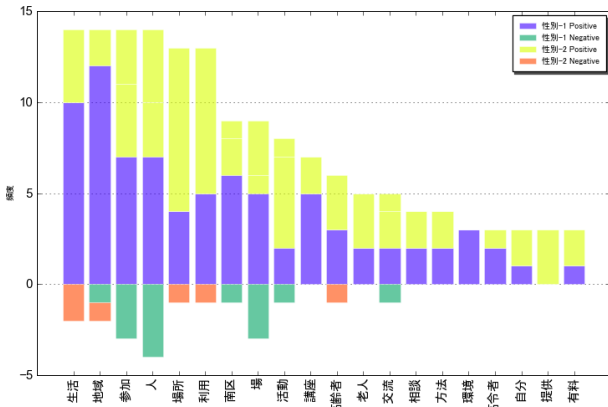


図5 好評語ランキング（男女別集計）

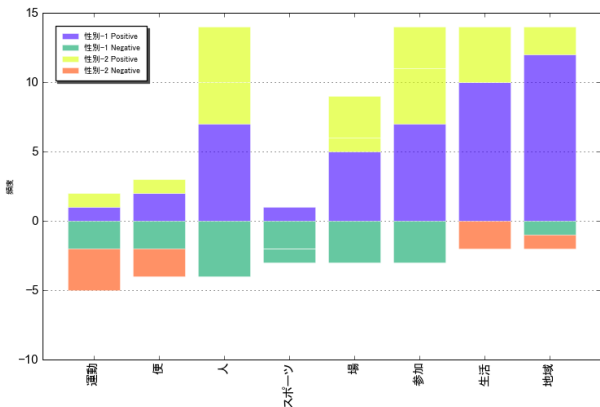


図6 不評語ランキング（男女別集計）

#### 4.6 属性と係り受け表現の関係

ことばを介した男女の特徴比較（図7参照）から以下のことが読み取れる。

- 性別1（男性）と性別2（女性）の距離が比較的離れている。これは男性と女性の意見（要望）に隔たりがあることを意味している。
- 性別1（男性）の丸印は性別2（女性）の丸印と比較して大きく、男性の意見（要望）が女性よりも多いことを意味している。しかし、男性は“人—多い”、“南区—多い”、“必要—思う”という係り受け表現と関連が深く、人・モノ・ことの多少や必要性和関連する表現をしているが、何をしたいのか、何をしてほしいのか、意見（要望）としては不明確である。一方、女性は“場—欲しい”、“高齢者—楽しむ+できる”、“場—提供”、“場—作る+したい”という係り受け表現と関連が深く、どのような活動をしたのか、どのような場がほしいのか、明確に意見（要望）を述べている。

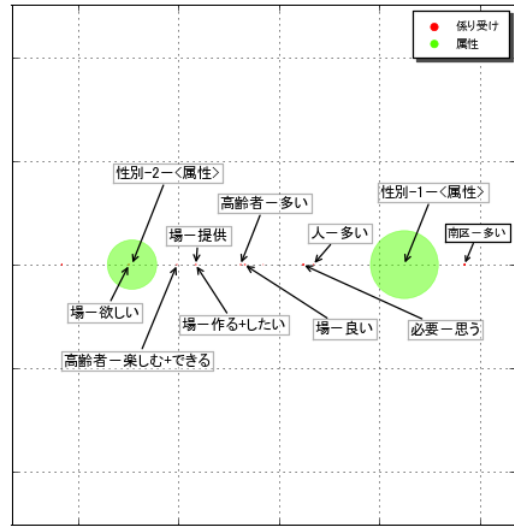


図7 ことばを介した男女の特徴比較

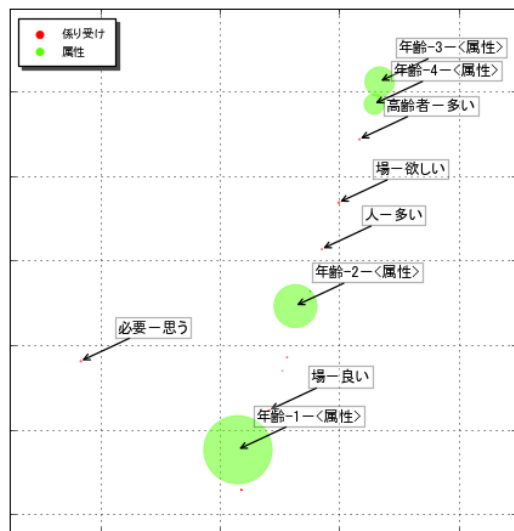


図8 ことばを介した年齢区分の特徴比較（1）

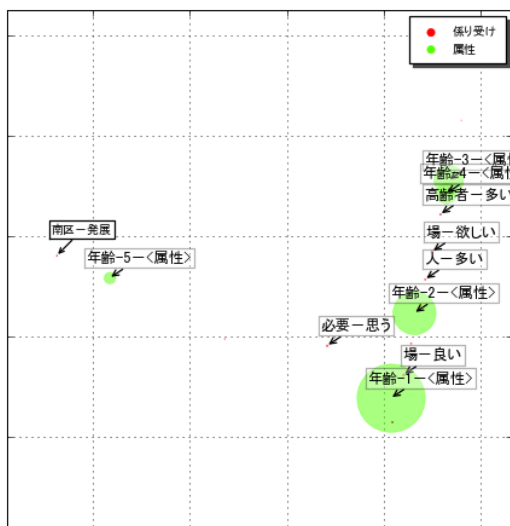


図9 ことばを介した年齢区分の特徴比較（2）

ことばを介した年齢区分の特徴比較（図8、図9参照）から以下のことが読み取れる。

- ・区分1（65歳～69歳）の丸印が一番大きく、区分2（70歳～74歳）から区分3（75歳～79歳）・区分4（80歳～84歳）・区分5（85歳以上）の順に丸印は小さくなっている。これは区分1（65歳～69歳）の意見（要望）が一番多く、年齢が上がるにしたがって意見（要望）が少なくなっていることを意味している。
- ・区分1（65歳～69歳）と区分2（70歳～74歳）の丸印は比較的距離が近い。同様に、区分3（75歳～79歳）と区分4（80歳～84歳）の丸印はかなり距離が近いが、区分1・区分2と区分3・区分4の丸印の距離は比較的離れている。また、区分5（85歳以上）の丸印は、他の区分のいずれとも距離がかなり離れている。距離の近い／遠いは意見（要望）の近い／遠いを意味している。
- ・区分1（65歳～69歳）の意見（要望）は、“場—良い”という係り受け表現と関連が深い。同様に、区分2（70歳～74歳）の意見（要望）は、“人—多い”、“場—欲しい”という係り受け表現と、区分3（75歳～79歳）・区分4（80歳～84歳）の意見（要望）は、“高齢者—多い”という係り受け表現と、区分5（85歳以上）の意見（要望）は、“南区—発展”という係り受け表現と関連が深い。
- ・“場—良い”と“場—欲しい”という似通った係り受け表現から見ても区分1（65歳～69歳）と区分2（70歳～74歳）の意見（要望）は比較的近い。同様に、“高齢者—多い”というほぼ共通の係り受け表現を持つことから区分3（75歳～79歳）・区分4（80歳～84歳）の意見（要望）はかなり近い。“南区—発展”という係り受け表現は、どの係り受け表現とも距離が大きく離れており、これら

の位置関係からも区分5（85歳以上）の意見（要望）は他の年齢区分のいずれの意見（要望）ともかなり離れていることが分かる。

## 5. 考察

本研究では、「札幌市南区在住65歳以上高齢者の健康に関するニーズ調査」における892人分の回答文書データ（総単語数10887）をテキストマイニング分析した結果、以下のことが明らかになった。

- (1) 回答者一人当たり平均57.5文字で回答しており、比較的簡潔に意見（要望）を記述していた。
- (2) 形容詞と副詞を合わせた出現回数が907（8.33%）と少なく、感性的・感情的な回答は少なかったと判断された。
- (3) 回答者はすべての年齢区分で女性より男性の方が多かった。これは世帯を代表して男性が多く回答したことによるものと推測される。また、年齢が上がるにしたがって男女ともに回答者数は漸減していた。
- (4) “場—欲しい”、“場所—欲しい”、“施設—欲しい”など、高齢者が集い活動する“場”や“場所”に関する意見（要望）が多く出たが、これらの意見（要望）は女性から多く、女性の方が地域での交流や活動の場を積極的に求めていることが分かった。
- (5) 年齢別の比較では、“場—欲しい”や“場所—欲しい”の意見（要望）は区分2（70歳～74歳）が最も多いが、区分4（80歳～84歳）、区分3（75歳～79歳）と続いており、交流や活動の場を求める健康で積極的な高齢者が多いと推測される。
- (6) 高齢者が楽しく参加できる活動や自由に使える場・場所についての肯定的な意見（要望）は男性よりも女性が多かったことが肯定的・否定的な係り受け表現の抽出から明らかとなった。活動や場・場所に関する否定的な意見（要望）はなかったが、郵送されてきた封筒に説明資料が入ってなかったとの否定的な意見（要望）があった。
- (7) 肯定的な意見（要望）は全体的に男女でほぼ同等であったが、“地域”や“生活”に関する肯定的な意見（要望）は圧倒的に男性が多く、“場所”、“利用”、“活動”に関する肯定的な意見（要望）は女性に多かったことが肯定的・否定的なことば（単語）の抽出から明らかとなり、女性は社交的な場や活動に積極的であることが分かった。否定的な意見（要望）は、全体的に男性に多かった。特に、“人”、“参加”、“場”、“南区”、“活動”、“交流”に関する否定的な意見（要望）は男性だけであり、男性は社交的な場や活動に消極的であるこ



とが分かった。一方、”生活“と”高齢者“に関する否定的な意見（要望）は女性だけであり、女性は生活に対する一抹の不安を脳裏に浮かべてものごとを考えていると推測される。

- (8) ことばを介した男女の特徴比較から、男性は女性よりも多く意見（要望）を述べているが、何をしたいのか、何をしてほしいのか、不明確であった。一方、女性はどうのような活動をしたのか、どのような場がほしいのか、明確に意見（要望）を述べていた。
- (9) ことばを介した年齢区分の特徴比較から、区分1（65歳～69歳）と区分2（70歳～74歳）の意見（要望）は比較的近く、同様に、区分3（75歳～79歳）と区分4（80歳～84歳）の意見（要望）はかなり近いが、区分1・区分2と区分3・区分4の意見（要望）は比較的離れていることが明らかとなった。また、区分5（85歳以上）の意見（要望）は他の年齢区分のいずれの意見（要望）ともかなり離れていることが明らかとなった。

## 6. 結言

札幌市南区在住 65 歳以上高齢者の健康ニーズ調査で収集した自由記述文書データのテキストマイニング分析を通して、大量の文書データをふるいにかける感度調整技術、ふるいの中から金塊（重要なキーワードや関係性）を見つけ出す能力、分析結果から重要と思われる潜在知識を抽出する能力などが養われ、高齢者の潜在的要望や活動欲求などを抽出することができたことから研究目的は達成されたと判断している。また、得られた新しい知見や分析技術を解説書にまとめる準備ができたことから、今後、解説書を作成し、地域創生関係者を対象としたテキストマイニング講習会を実施し、地域創生関係者への分析技術の伝授を図りたい。

## 参考文献

- 1) 総務省：第4節 本格的なデータ活用社会の到来、平成26年版 情報通信白書、2016。  
[www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/html/nc134020.html](http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/html/nc134020.html)（2018年10月1日参照）
- 2) 経済産業省：IT人材の最新動向と将来推計に関する調査結果～報告書概要版～、経済産業省 商務情報政策局 情報処理振興課、2016.6。  
[http://www.meti.go.jp/policy/it\\_policy/jinzai/27FY/ITjinzai\\_report\\_summary.pdf](http://www.meti.go.jp/policy/it_policy/jinzai/27FY/ITjinzai_report_summary.pdf)（2018年10月1日参照）
- 3) 札幌市立大学：平成25年～平成29年度 文部科学省「地（知）の拠点整備事業（大学COC事業）」ウェルネスx協奏型地域社会の担い手育成「学び舎」事業 平成25年度成

果報告書、2014.3.

- 4) ウェルネスx協奏型地域社会の担い手育成「学び舎」事業 研究企画推進チーム：南区にお住まいの65歳以上の方の健康に関するニーズ調査 平成25-26年度報告書、文部科学省 平成25年度採択「地（知）の拠点整備事業」ウェルネスx協奏型地域社会の担い手育成「学び舎」事業、2015.3.
- 5) (株)NTT データ数理システム：Text Mining Studio Version 6.1 Manual、2018.1.
- 6) 城間祥之：2-3 短期居住体験者のテキストマイニング分析、「地域創生をデザインする 都会と地方の魅力相乗モデル開発」タイム・スペースシェアリング型地域連携による地域創生デザイン研究報告書、札幌市立大学 地域創生デザイン研究会、2016.3.

## 謝 辞

本研究では、平成25年度文部科学省「地（知）の拠点整備事業」に採択された札幌市立大学『ウェルネスx協奏型地域社会の担い手育成「学び舎」事業』において実施された、「札幌市南区在住65歳以上高齢者の健康に関するニーズ調査」の集計データ（自由記述文書）をテキストマイニング分析した。本研究への協力を快諾してくれたCOC部門長（事業担当者）の中原宏教授、および集計データを無償提供してくれた公立大学法人札幌市立大学に深く感謝申し上げる。

なお、本研究は日本学術振興会科学研究費助成事業（学術研究助成基金助成金）（基盤研究(C)（一般））「テキストマイニング分析手法による『地域創生デザイン』志向潜在知識抽出に関する研究」（課題番号16K00713）の助成を受けて実施したものである。