

科学研究費助成事業 研究成果報告書

令和元年6月15日現在

機関番号：20105

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K00713

研究課題名（和文）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

研究課題名（英文）A Study on Extraction of Potential Knowledge Oriented toward "Local Community Revitalization Design" by Text Mining

研究代表者

城間 祥之（SHIROMA, Yoshiyuki）

札幌市立大学・デザイン学部・教授

研究者番号：90113571

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本研究では、札幌市南区在住65歳以上の高齢者（892人）の健康に関するニーズ調査の集計データ（自由記述文書）をマイニング分析し、属性（性別や年齢区分など）ごとの特徴（潜在知識）を抽出した。また、学生や行政担当者らが自らテキストマイニング分析を容易に行なえるよう、マイニング分析技術や分析結果の解釈方法などを例示した解説書を作成するとともに大学院生を対象とした講習会を実施し、その有効性を確認した。さらに、研究の概要を紹介するリーフレットを作成し、地域創生デザインに係わる関係者に配布するとともに、ウェブサイトにてそのpdf fileを掲載し、広く社会に公開している。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、文書ビッグデータ分析手法として注目を集めているテキストマイニング手法を地域住民の意識調査文書データ（テキスト）の分析に適用することを他者に先駆けていち早く着想・適用し、分析結果から住民の潜在的要望や活動欲求などが抽出できることを明らかにしたことである。同時に、研究過程で得られた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その有効性を確認したことは社会的意義が深いものと考えている。

研究成果の概要（英文）：In this research, tabulated data (free description document) of needs survey on health of 892 elderly people aged 65 years or older living in Minami-ku, Sapporo city was analyzed by text mining, and characteristics of each attribute of sex, age group, etc. was extracted. In addition, in order to facilitate the text mining analysis by students and administrative personnel, we prepared an explanatory manual on mining analysis technique and how to interpret analysis results, and held a seminar for graduate students, confirmed its effectiveness. Furthermore, we have created a leaflet to introduce the outline of the research, distributed it to the people involved in local community revitalization design, and loaded the pdf file on the website to make it widely available to society.

研究分野：デザイン学

キーワード：テキストマイニング ビッグデータ分析 地域創生デザイン 潜在知識抽出 文章解析 質的データ分析 住民意識調査 地方創生

様式 C-19、F-19-1、Z-19、CK-19（共通）

1. 研究開始当初の背景

本研究者は地域再生・活性化の問題にいち早く着目し、「タイム・スペースシェアリング型地域連携による地域創成デザイン研究（日本学術振興会科学研究費助成事業 基盤研究(A)、平成25年度～平成27年度）」を行ってきた[1][2][3][4]。これは、大都市（札幌市）と周辺の過疎市町村の住民がお互いの時間や空間を共有することにより、両地域住民の相互交流を通じた人的ネットワークを構築し、大都市への人口の一極集中と周辺地域の衰退を防止するというコンセプトの社会実験研究であった。その研究の中で収集した住民の「短期居住体験」実験報告書のテキストデータが膨大な量になり、分析者がテキストを読んで主観的に分類・分析する従来型的手法では対応できなくなっていた。そこで、テキストマイニング手法を適用し、テキストデータを計量的に分析し、客観的に解釈することを試みたところ、「ポジティブ・ネガティブの感情表出や話題の中心、気になること」などが客観的な数量データやことばネットワーク図などで可視化できたことから、それを解釈することで新たな知識を抽出できるのではないかと考えたことが背景の一つである。

また、研究開始当初はもちろん、令和元年（2019年）現在においても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[5][6]、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考えたことが背景の二つ目である。

2. 研究の目的

本研究の目的は、地域住民の意識調査文書データをテキストマイニング分析し、得られるキーワードの出現頻度やことばの共起関係、係り受け関係などから、潜在的要望や活動欲求などを抽出する知見を得ることである。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、その有効性を確認することである。

3. 研究の方法

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』[7]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」[8]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ことにした。

この健康ニーズ調査は平成24年度に実施されたが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。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が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保健福祉部保健福祉課地域福祉係の担当者が行った。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。

本研究では、Text Mining Studio 6.1.[9]を用いて、892人の回答文書データを以下の手順で分析した[10]。

- (1) 回答文書データファイル(csv file)をText Mining Studioへ読み込み、「分ち書き」を実行した。文章は形態素（言語で意味を持つ最小単位）に区切ら、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. 研究成果

テキストマイニング分析によって得られた基本情報と品詞出現回数（表1参照）および属性別集計（表2参照）からテキストデータ全体の量的特徴として以下のことが明らかになった。

- (1) 総行数は892行であった。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。
- (2) 平均行長（文字数）は57.5であった。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。
- (3) 総文章数は1641であった。総文章数を総行数で除すると約1.84となり、上記同様、1行あたりの文章数1.84が算出される。
- (4) 平均文章長（文字数）は31.3であった。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。
- (5) 延べ単語数は10887であった。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。
- (6) 単語数の内訳を品詞別出現回数で見ると、形容詞と副詞を合わせた出現回数は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る。
- (7) 単語種別数は4296であった。単語種別数を総行数で除すると約4.82となり、回答者1人あたり他者と異なる単語が4.82個含まれていることを意味する。他者と異なる単語数4.82は1行あたりの単語数12.2の約39.5%に相当し、回答者それぞれが簡潔な文章（回答）のなかにも独自の表現を使っていることが窺える。
- (8) 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。
- (9) 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人いた。

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%)

表2 属性別集計と欠損（属性記入漏れ）

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

気になることば“欲しい”の係り受け表現や肯定的・否定的な係り受け表現の抽出、好評語・不評語の抽出などから、以下のことが明らかになった。

- (10) “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）は性別2（女性）からが多く、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（図1参照）。

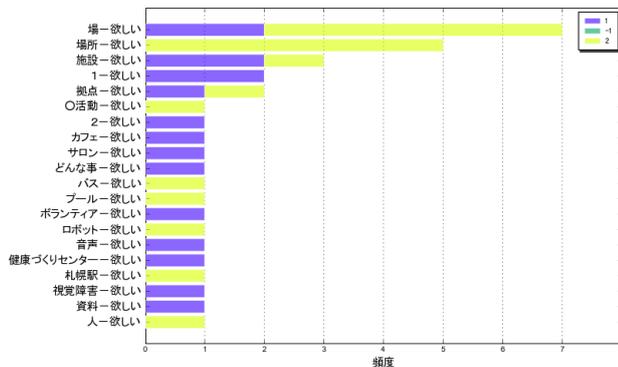


図1 “欲しい”の係り受け表現（男女別集計）

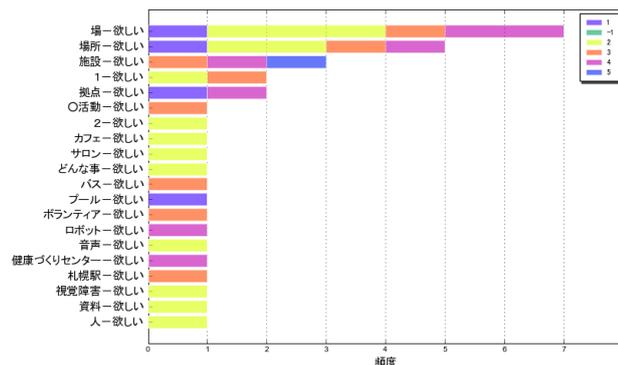


図2 “欲しい”の係り受け表現（年齢別集計）

(11)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(図2参照)。

(12)“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、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・場所についての肯定的な意見(要望)が多く出たが、これらの意見(要望)は性別2(女性)からが多かった(図3参照)。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。(図4参照)。

(13)肯定的な意見(要望)は全体的に男女では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)は圧倒的に性別1(男性)が多く、“場所”、“利用”、“活動”に関する肯定的な意見(要望)は性別2(女性)に多かった。女性は社交的な場や活動に積極的であることが好評語・不評語の抽出からも明らかとなった。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)は性別2(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ながらものごとを考えていると推測される(図5、図6参照)。

(14)ことばを介した男女の特徴比較から、性別1(男性)は性別2(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、女性はどのような活動をしたいのか、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(図7参照)。

(15)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(図8参照)。

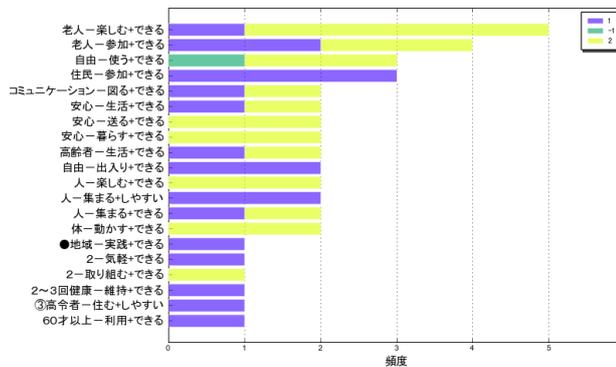


図3 肯定的な係り受け表現(男女別集計)

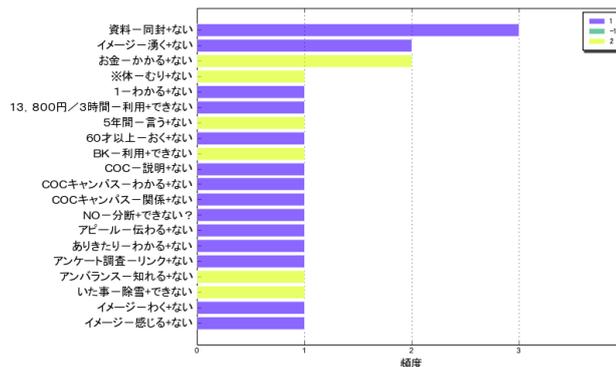


図4 否定的な係り受け表現(男女別集計)

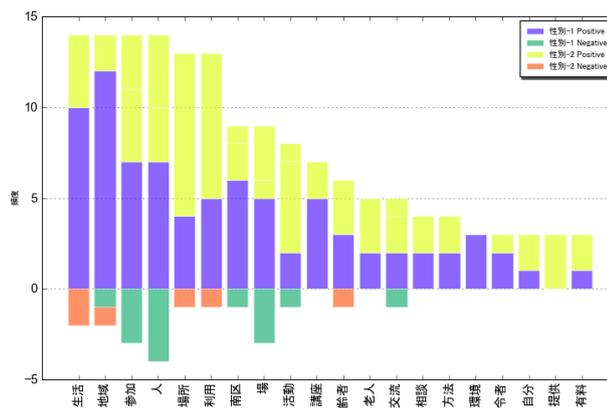


図5 好評語ランキング(男女別集計)

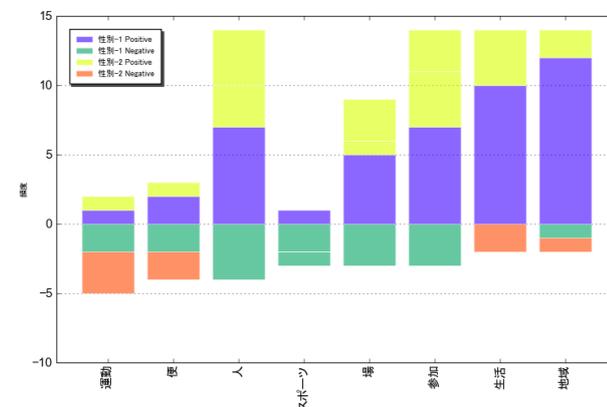


図6 不評語ランキング(男女別集計)

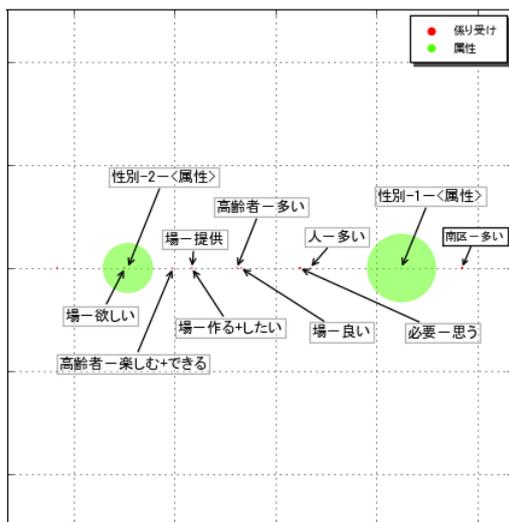


図7 ことばを介した男女の特徴比較

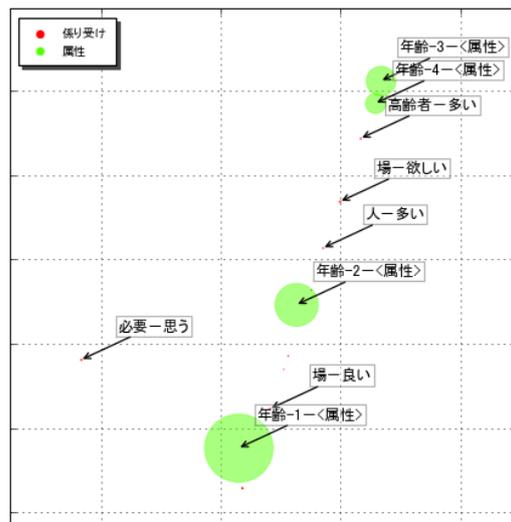


図8 ことばを介した年齢区分の特徴比較

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大容量文書データのテキストマイニング分析から、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と判断している。また、その過程で得られた新しい知見や分析技術を伝授すべく解説書を作成したことから、今後、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、分析技術の伝授を図りたい。

<参考文献>

- ① 柿山浩一郎, 城間祥之, 中原宏, 原俊彦, 石井雅博, 蓮見孝, Art & Design を活かした地域創生手法の検討を目的とした住民調査, 第 61 回日本デザイン学会春季研究発表大会, 2014.
- ② 酒井正幸, 蓮見孝, 城間祥之, 上遠野敏, 中原宏, 原俊彦, Art & Design を活かした地域創生手法に関する研究-1, 第 1 報 - 地域創生デザイン学の確立のための TSS 研究 -, 第 62 回日本デザイン学会春季研究発表大会, 2015.
- ③ 柿山浩一郎, 片山めぐみ, 石井雅博, 山田良, 齊藤雅也, 上田裕文, Art & Design を活かした地域創生手法に関する研究-2, 第 2 報 - 地域を対象とした研究における仕掛けづくり, 第 62 回日本デザイン学会春季研究発表大会, 2015.
- ④ 城間祥之, 蓮見孝, 酒井正幸, 柿山浩一郎, 中原宏, 原俊彦, 石井雅博, Art & Design を活かした地域創生手法に関する研究 - 短期居住体験者の日報のテキストマイニング分析 -, 第 17 回日本感性工学会大会, 2015.
- ⑤ 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
(2018 年 10 月 1 日参照)
- ⑥ 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
(2018 年 10 月 1 日参照)
- ⑦ 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.
- ⑧ 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.
- ⑨ (株) NTT データ数理システム, Text Mining Studio Version 6.1 Manual, 2018. 1.
- ⑩ 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.

5. 主な発表論文等

〔雑誌論文〕（計 0 件）

〔学会発表〕（計 1 件）

城間祥之、柿山浩一郎：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 —、感性フォーラム札幌 2019、2019 年 2 月、1-6.

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

○取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

(1) ホームページ等

「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」
成果報告サイト

<http://kakiyama.info/research/kaken/2018/tm>

(2) 発表者名：城間祥之、柿山浩一郎 発表表題：大容量文書データのテキストマイニング分析 — 札幌市南区在住 65 歳以上高齢者の意識調査データを対象として —
学会等名：2018 SCU 産学官金研究交流会 発表年：2018 年 11 月

6. 研究組織

(1) 研究分担者

研究分担者氏名：柿山 浩一郎

ローマ字氏名：KAKIYAMA Koichiro

所属研究機関名：札幌市立大学

部局名：デザイン学部

職名：准教授

研究者番号（8 桁）：30410517

(2) 研究協力者

研究協力者氏名：

ローマ字氏名：

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。