

2-3

短期居住体験者の日報のテキストマイニング分析

Author

城間祥之 地域創生先行事例分析班 / 暮らし満足度定量化手法開発班

Keywords

短期居住者

日報

テキストマイニング

■「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」の一環として、大都市（札幌市）と周辺地域（壮瞥町）の住民を対象に「短期居住体験」実験が行われました。札幌市民は壮瞥町で、逆に壮瞥町民は札幌市で短期居住を体験するという実証実験です。居住体験者の日々の出来事や感想などが書かれた日報（テキスト）を通読して分類・分析すると分析者の主観的な解釈を伴うという欠点があります。そこで本研究では、日報の分析にテキストマイニングを適用して計量的に分析し、客観的に解釈することを試みました。その結果、札幌市民と壮瞥町民それぞれの思いが対比化・視覚化され、客観的な分析から興味深い新たな知見が得られました。

1. 研究の背景と目的

日本創成会議・人口減少問題検討分科会の報告書「ストップ少子化・地方元気戦略」によると、“2010年から2040年までの間に「20歳～39歳の女性人口」が5割以上減少する自治体数は、現在の推計に比べ大幅に増加し、896自治体、全体の49.8%にもものぼる結果となった”とのことです¹⁾。さらに、896自治体のうち、2040年時点で人口が1万人を切る市区町村は523自治体で、全体の29.1%にのぼり、これらは、このままでは消滅可能性が高いと断言されています。2014年5月、“消滅可能性都市”と定義された896自治体の一覧とともに当該報告書が公開されるや否や、日本全国に衝撃が走り、改めて、地域の再生・活性化が喫緊の国民的課題として注目されています。報告書では、「若者に魅力のある地域拠点都市」を中核としたコンパクトな拠点形成の一環として、空き家を活用した「二地域居住」の提言、同様に、「中高年の地方移住の支援」の一環として、「都市高齢者の地方への住み替え支援」などが提言されています。

報告書の提言に先鞭をつけること1年、本研究者らは「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」を行ってきました²⁾³⁾⁴⁾。これは、大都市と周辺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理念の社会デザイン研究です。本研究は、その一環として行われた「短期居住体験」実験参加者の日報分析を試みるものです。

本研究の目的は、短期居住体験者の日報（文章）をテキストマイニングにより計量的に分析し、文脈を客観的に解釈し、体験者の思い、興味・関心事などを明らかにすることです。最終的には、大都市とその周辺町村の住

民相互の「二地域居住」や「短期居住」を推進するための新たな知見を得ることが目的です。

2. 研究方法

2-1. 「短期居住体験」実験について

「短期居住体験」実験を行うため、まず、札幌市に1か所、壮瞥町に2か所の住居を用意しました。壮瞥町は洞爺湖のほとりに位置し、有珠山と昭和新山を有し、観光と農業を主要産業とする町です。壮瞥町と札幌市との自動車での移動時間は2時間弱、距離は約107kmです。実験に際し、札幌市立大学倫理委員会の承認を得た上で、①実験参加者の募集、②説明会の開催、③参加申込、④書類選考などを行いました。居住体験者には、⑤同意書の提出、⑥開始時インタビュー、⑦日常生活ガイダンス、⑧指定イベントへの参加、⑨日報・週報の提出、⑩居住体験終了報告書の提出、⑪終了時インタビューなどを課すこととしました⁴⁾。以上の実験計画に基づき、「短期居住体験」実験を行いました。その内訳は、札幌市での短期居住は2組(3名)、壮瞥町での短期居住は3組(4名)でした。壮瞥町での短期居住1組(2名)だけは冬季間、残りの4組は夏から秋にかけて実施されました。

2-2. テキストマイニングの概要

テキストマイニングとは、鉱山で大量の土砂の中から極めて少量の金塊を採掘するように、大量の文章（自由記述文）データをマイニング(採掘)することで数量化し、可視化する分析手法です。従来のテキストを通読し、分類・分析する主観的分析手法からは想像が及びにくい潜在的知識や法則性を発見することが期待される分析手法です。本研究では、短期居住体験者の日報（文章）の分析をText Mining Studio 5.0.2⁵⁾を用いて行います。

2-3. テキストマイニング手順

日報のテキストマイニングは以下の手順で行われます。

(1) 日報のテキストデータ(csv file)は、Text Mining Studioへ読み込まれるとともに「分かち書き」されて形態素(言語で意味を持つ最小単位)に区切られ、分析のためのオリジナルテキストが生成されます。その際、形態素ごとの品詞が確定されます。

(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)、総文数、平均文長(文字数)、延べ単語数、単語種別数)を算出します。品詞出現回数なども算出し、日報をテキストの数量データとして把握します。また、形容詞や副詞の出現回数から感性的表現の多少を量的に推測します。

(3) 単語頻度解析を用いて、単語の出現頻度を分析します。同様に、係り受け頻度解析を用いて、主語と述語のような係り受け表現から、居住体験者がポジティブ(またはネガティブ)に表現した事柄を把握します。

(4) 評判分析を用いて、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)を量的に把握します。

(5) ことばネットワーク分析を用いて、品詞間の係り受け関係を矢印付き直線で構成されるネットワーク図として、視覚的に把握します。この図では、矢印の根元の単語は条件部を表し、矢印の先は結論部を表すことで、係り受け関係をたどることができます。また、係り受け関係を構成する単語間の組み合わせ数の多い・少ないが密(または疎)なネットワーク図として描画されるので、この機能を用いて、日報のテキスト情報量を視覚的に把握します。さらに、クラスターの数を変更することで、単語間の関わり合いの強いクラスターを見つけ出すことができるので、この機能を用いて、居住体験者の話題の中心は何か? ということや話題の数などを把握します。

(6) 注目分析では、「気になることば」を設定すると、そのことばと一緒に出現することばのネットワーク図が作成されるので、居住体験者が「気になることば」を中心にどのような話を展開しているのかを把握します。

3. 結果

札幌市で短期居住(2014年9月16日~11月2日)した壮瞥町の男性(49歳)を「壮瞥町男性A」、同様に、壮瞥

町で短期居住(2014年10月20日~11月30日)した札幌市の男性(46歳)を「札幌市男性B」と記すこととしますが、両者の日報のテキストマニング分析比較から以下のことが明らかになりました。

(1) 日報の基本情報と品詞出現回数に関して、札幌市男性Bの平均行数(文字数)は壮瞥町男性Aの4倍と多く、札幌市男性Bは日々の活動を大量に報告していることが判ります。一方、感性(感情)表現と関係が深いと言われる形容詞、副詞の日報全体に占める割合は壮瞥町男性Aの方が約2%高いですが、両者の感性(感情)表出割合はほぼ同様と見なせます(表1、表2)。

表1 壮瞥町男性Aの日報の基本情報と品詞出現回数

項目	値	品詞	回数 (%)
総行数	1	名詞	597 (61.87%)
平均行数(文字数)	2554	動詞	168 (17.41%)
総文数	174	形容詞	81 (8.39%)
平均文長(文字数)	14.7	副詞	79 (8.19%)
延べ単語数	965	連体詞	15 (1.55%)
単語種別数	579	接続詞	7 (0.73%)
		記号	17 (1.76%)

表2 札幌市男性Bの日報の基本情報と品詞出現回数

項目	値	品詞	回数 (%)
総行数	1	名詞	4743 (61.6%)
平均行数(文字数)	10216	動詞	1458 (18.94%)
総文数	1311	形容詞	513 (6.66%)
平均文長(文字数)	15.6	副詞	612 (7.95%)
延べ単語数	7700	連体詞	135 (1.75%)
単語種別数	2876	接続詞	83 (1.09%)
		記号	120 (1.56%)

(2) 係り受け頻度解析に関して、壮瞥町男性Aの方が肯定的な意見が多く出ており、都会の利便性の良い機能と密接に関係しています。札幌市男性Bの肯定意見は、「パノラマ、山系、魚影、洞爺湖」などの単語と係り受け関係にあり、壮瞥町の自然環境が肯定的意見に結びついています(図1、図2)。

(3) 評判抽出に関して、壮瞥町男性Aの好評語は、「店、天気、ご飯、味、料理」などの単語を含み、その出現数も多い。これは図1に示す肯定的な意見とも合致しています。その一方、不評語として、「信号、交通量、町中、環境」などの単語が並んでおり、大都市の喧騒が気になったものと推測されます。札幌市男性Bの好評語は、「場所、風景、景観、天気」などで、図2に示す肯定的な意見とも合致しています。その一方、不評語は「道、天気、居

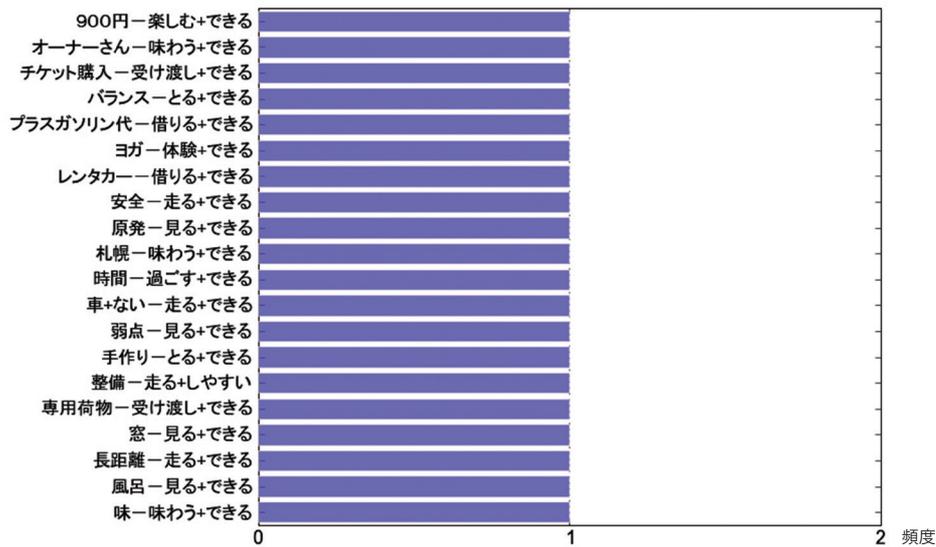


図1 壮瞥町男性Aの係り受け頻度解析(肯定的な意見)

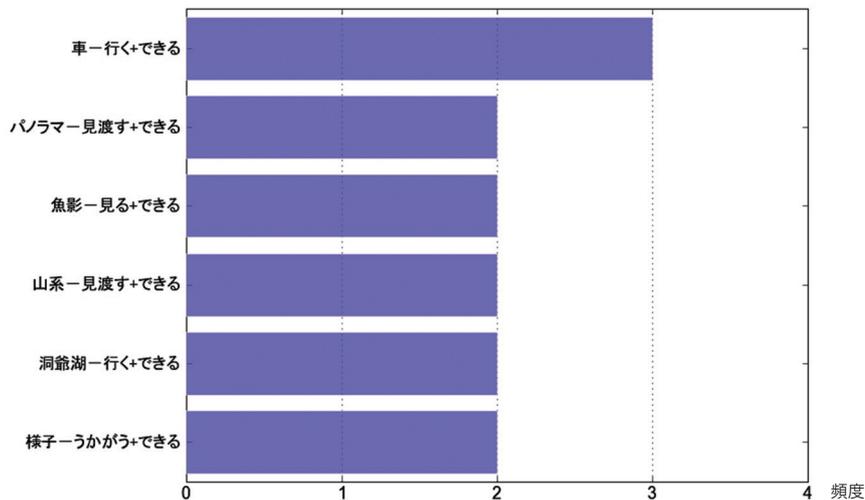


図2 札幌市男性Bの係り受け頻度解析(肯定的な意見)

心地」などで、田舎の道路状況や「短期居住体験」実験のために用意された住居の住み心地が影響しているものと推測されます。「天気」は好評語と不評語の両方に現れており、天気の良し悪しが短期居住者の活動にプラス・マイナスに作用しているものと推測されます(図3、図4)。

(4) 注目語情報のネットワークに関して、壮瞥町男性Aの注目語(都会)は、「シビア、冷たい、イメージ、スピード」などの単語と連結しており、大都市のネガティブなイメージのことばネットワークを構成しています。また、注目語(田舎)は、「若年層、家族連れ、スピード、アマゾン」などの単語と連結しており、一見、田舎とは結び難いイメージのことばネットワークを構成しています。詳細に調べてみると、これらの単語は都会の実態を表現したのですが、田舎との違いを比較する文章表現に使

われたため、注目語(田舎)に連結していることが判明しました(図5)。札幌市男性Bの注目語(景色)は、「洞爺湖全般、気に入る、360度、ベストセット、味わい深い」などの単語と連結しており、壮瞥町のポジティブなイメージのことばネットワークを構成しています。また、注目語(生活)は、「20時、先、見越す、ネット環境、左右、アート感、もの作り感、共感」などの単語と連結しており、脈絡がないように見えます。しかし詳細に調べてみると、「20時、先、見越す」は、壮瞥町のお店が早く閉店することに関する感想のキーワードでした。同様に、「ネット環境、左右」は、ネット環境によって札幌市男性B自身の日々の生活が左右されていることを省察したもの、「アート感、もの作り感、共感」は、若者イベントへの感想のキーワードでした(図6)。

(5) ことばネットワークに関して、壮瞥町男性Aの日報

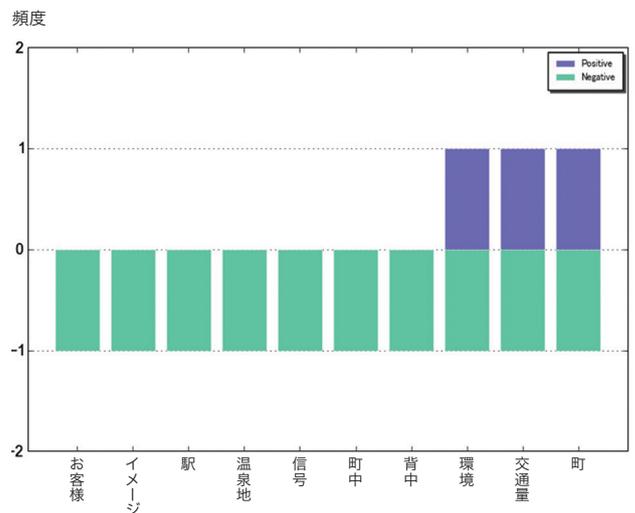
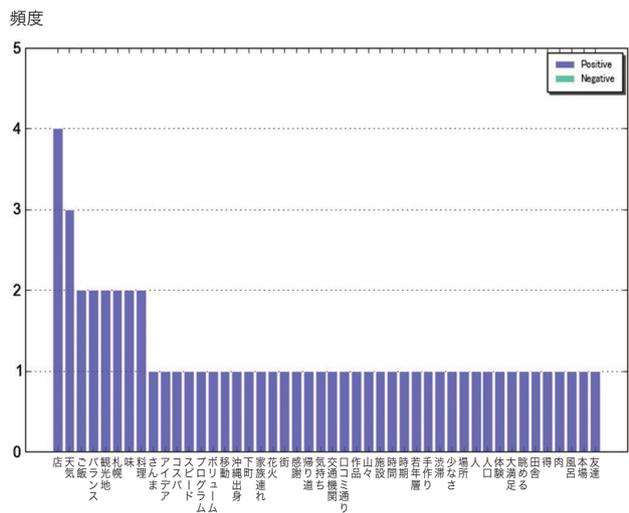


図3 壮瞥町男性Aの評判抽出：好評語（左）と不評語（右）

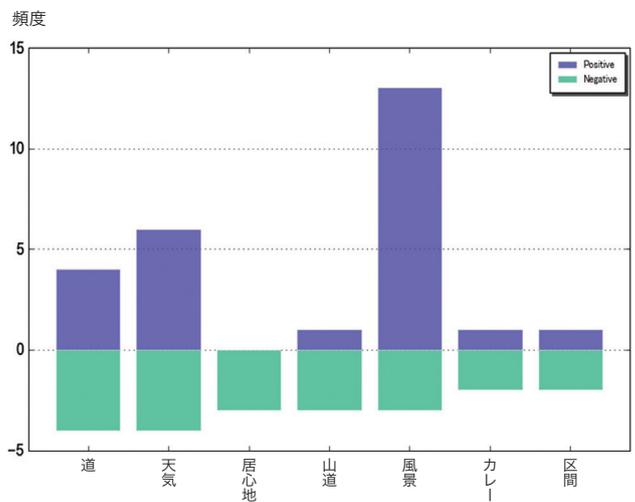
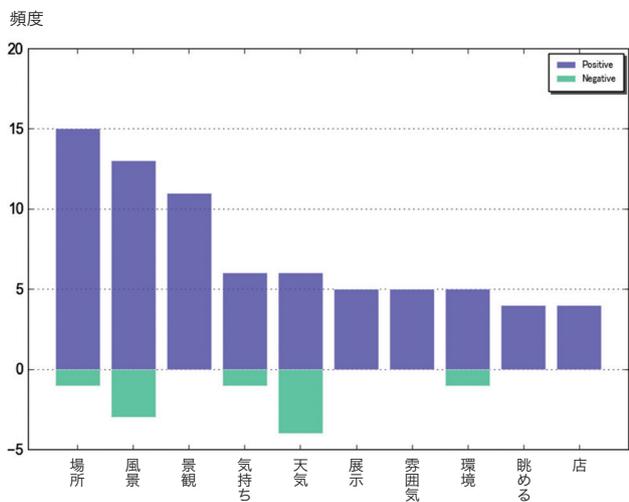


図4 札幌市男性Bの評判抽出：好評語（左）と不評語（右）

の話題を3クラスターに絞り、そのうちの2つのクラスターにおける“ことばとことばの共起関係”をネットワーク図で示したものが、図7の左図です。ここでは、「一苦勞、片付く、帰る、芸森スタジオ、友人達、ありがたみ、覗く、カフェ、出店」が一つの話題を構成しています。また、「マクロビ、絶品、人口、増やす、年、1回、1回＋ない」がもう一方の話題を構成しています。図7の右図は、係り受け関係のことばネットワークです。分析の際のパラメータとして、話題一般（名詞－形容詞・形容動詞・形容詞）を指定したため、膨大な量の係り受け関係が表示されています（図7）。札幌市男性Bの日報の話題を3クラスターに絞り、そのうちの2つのクラスターにおける“ことばとことばの共起関係”をネットワーク図で示したものが、図8の左図です。ここでは、「場所、農家、中学生、75点、キャンプ場、気、環境、人、興味深い」が一つの話題を構成しています。また、「CAFE、

興味、リング、作業、楽しい、参加、ストレス」がもう一方の話題を構成しています。図8の右図は、係り受け関係のことばネットワークです。分析の際のパラメータとして、行動（名詞－動詞・サ変接続名詞）を指定したため、札幌市男性Bの日報の文章量が壮瞥町男性Aの4倍あるにも関わらず、すき間の多いネットワーク図となっています（図8）。

なお、図7、図8の共起関係ネットワーク図において、単語の傍の丸い点の大きさはその単語の出現割合の大きさを表しますが、それぞれの話題を構成する単語の出現回数は均等でした。

4. 考察

短期居住体験者は居住先の特徴を短期間に明確に捉えており、それが肯定的／否定的意見に結びついているこ

