

Twitterデータを用いた 地域拠点ワードの共起パターン分析

山本 岳¹・浅田 拓海²・鈴木 貴文³・有村 幹治⁴

¹HRS株式会社 環境防災部 (〒047-0005 小樽市勝納町8-39)

E-mail:g.yamamoto@h-r-s.co.jp

²正会員 室蘭工業大学 大学院工学研究科 暮らし環境系領域 (〒050-8585 室蘭市水元町27-1)

E-mail:asada@mmm.muroran-it.ac.jp

³HRS株式会社 環境防災部 (〒047-0005 小樽市勝納町8-39)

E-mail:ta.suzuki@h-r-s.co.jp

⁴正会員 室蘭工業大学 大学院工学研究科 暮らし環境系領域 (〒050-8585 室蘭市水元町27-1)

E-mail:arimura@mmm.muroran-it.ac.jp

本研究では、室蘭市を対象とし、Twitterコメントデータから抽出した地域拠点ワードをユーザーレベルで関連付けし、人々の関心先として共有される拠点について分析した。具体的には、ユーザーが対象期間中に発生させた複数の拠点ワードを関連付け、それを全ユーザーで束ねることで拠点ワード間の関連度合いをネットワークとして表した。その共起パターンを地図上に可視化し、拠点の位置関係とともに、それらの関心先としての関連性について明らかにした。さらに、各拠点の中心性指標を算出した結果、主要駅や観光名所はネットワーク内の中心的な存在として評価される一方、単独発生数や次数中心性が大きいものの、所属するネットワークが小さい、すなわち関心を共有するユーザーコミュニティが小さい拠点が幾つか抽出された。

Key Words : *Twitter, Town Center, GIS, Network Data, Sharing Interests, Centrality Measures*

1. はじめに

現在、我が国では、観光立国の実現に向けた施策が積極的に推進されている。最近では、成長著しいアジア諸国をはじめとする観光需要の取り込みにより、訪日外国人観光客数が著しく増加している状況にある。北海道においても、食、景観、温泉、アウトドアなど様々な地域の魅力を求めて、国内外からの観光客数がシーズンを通して増加している。さらには、2016年4月、北海道新幹線が開通し、首都圏、東北エリアと函館間の観光交流に期待が高まっている状況である。このような観光客の流れやニーズがめまぐるしく変化する今日において、さらなる新規客、リピーターを確保するためには、観光コンテンツの充実とともに、それらを的確に把握し、情報公開することが極めて重要である。

観光情報については、最近では、スマートフォンの普及などから、WEB サービスやアプリなどによる動的なデジタルマップを利用し、旅行者が自ら目的地や周遊ルートを選択するケースが増えている。中

でも、ユーザーの「ナマの声」として、SNS によるリアルタイムな情報を観光旅行の参考にするケースも少なくはない。このようなメディアの多様化にせよ、重要となるのは、このような情報提供を効果的なものとするためには、まず第一に、地域内の観光資源の特長を広域的かつ網羅的に把握することが挙げられる。

近年、Twitter や Facebook などの SNS データの活用が様々な分野で着目されており、観光、防災¹⁾、交通システム²⁾などにおいても利活用が検討されつつある。土木計画分野における SNS データ活用に関する研究は、現在、初動期の段階あり、その多くは、コメント発生の地理的分布からユーザーの移動状況や関心拠点の把握である。このようなネットインフラの充実化、SNS の利用率の高まりから、今後は、SNS のような個人の投稿データにも目を向けた「ネットコンシャス」なまちづくり³⁾も重要となろう。最近では、地域の拠点に関するキーワードの発生頻度から、人々の関心の間、すなわち潜在的な交流ポイントの把握を試みた事例も報告されている⁴⁾。このように、

地域の交流拠点に関するコメントの発生量やその位置についての分析は報告されているが、人々の関心先として「共有」される各拠点の関連性について明らかにした事例はほとんどない。

そこで、本研究では、室蘭市を対象とし、Twitter の投稿データを用いて、ユーザー毎に複数発生される地域拠点に関するキーワード（以下、地域拠点ワード）を関連付けし、その共起パターンから人々の関心先として共有される交流拠点の空間的な分布やつながりを可視化することで、対象地域の拠点特性を明らかにする。

2. データの概要

(1) 地域拠点ワードの設定

本研究では、室蘭市を分析の対象として、地域に関する Twitter 投稿コメントを専用 API により取得した。地域拠点ワードについては、地域の各拠点がユーザーの関心先としてどのように関連しあっているのかを分析するため、表-1 に示すような規模レベルで選定した。レベル 1 は、室蘭市内の地区名 57 語、レベル 2 は室蘭市内の大型施設名 33 語、レベル 3 は室蘭市内の小型施設名 96 語であり、この他に、北海道道内の各市町村名を加え、計 2360 語を設定した。これらのワードは、室蘭市 HP の観光情報や各種 WEB サイトを参照して設定した。

(2) Twitter データの取得

上記の地域拠点ワードは、室蘭市以外の地域でも該当するものも含まれている。したがって、これらのワードで Twitter コメントを検索しても関係のないコメントも同時に抽出されてしまう。そこで、室蘭市固有のキーワードとして、「室蘭」「輪西」「だんぱら」「測量山」「地球岬」「絵鞆」「白鳥大橋」「母恋」「マスイチ」「イタンキ」「スワンフェスタ」「ペトル」の 12 語を設定し、これらのいずれかが含まれるコメントを抽出した²⁾。抽出対象期間は、2015 年夏季と秋季（7月から11月の5ヶ月間）とし、Twitter API を利用して自動的に取得した。なお、上記の 12 語で抽出されるコメントには、地域拠点ワード 186 語のいずれかも含まれる可能性が高いことを事前に確認している。

(3) 地域拠点ワードの基礎集計

取得したコメントデータから検出した各レベルの地域拠点ワード発生数を図-1 に示す。ワード発生数は 8 月が最も多く、11 月にかけて徐々に減少する。内

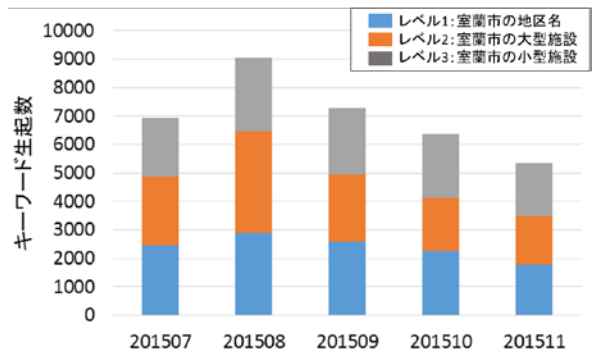


図-1 月別・レベル別の地域拠点ワードの発生回数

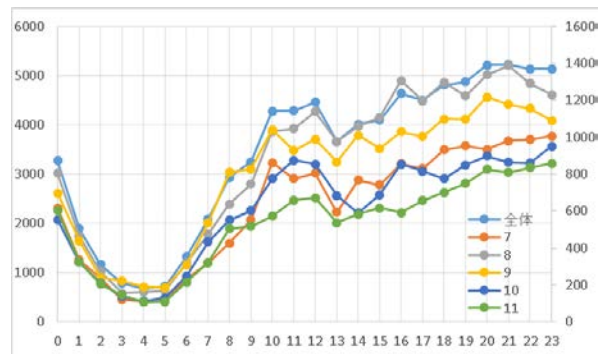


図-2 時間帯別・月別の地域拠点ワードの発生回数

表-1 全期間における地域拠点ワードの発生順位

順位	L1:地区名	L2:大型施設	L3:小型施設
1	東室蘭 3,696	地球岬 2,709	室蘭駅 2,384
2	母恋 2,052	室蘭港 2,707	東室蘭駅 1,839
3	絵鞆 1,257	白鳥大橋 1,899	水族館 626
4	東町 619	イタンキ 1,482	道の駅みたら 414
5	本輪西 280	測量山 1,083	母恋駅 409
6	栄町 272	トツカリシヨ 402	日本製鋼所 359
7	白鳥台 271	室蘭岳 269	栄高 355
8	日の出 264	大黒島 226	新日本製鐵 314
9	祝津町 241	金屏風 198	マスイチ 248
10	海岸町 240	鷺別岳 196	室蘭工業大 241

訳としては、レベル 1 が室蘭市の地区名、レベル 2 が市内の大型施設、レベル 3 が市内の小型施設となっている。中でも、レベル 2 の大型施設名は 8 月にその割合が急増していることから、イベントなどが開催された施設名が一時的に増加したものと考えられる。また、時間帯別に示したものを図-2 に示す。朝方 6 時から発生数が急激に増加し、正午をピークとして、若干減少するが、夕方、深夜にかけて再度増加に転じることがわかる。特に、7、8 月は、夜間での発生数増加が著しい。室蘭市では、測量山、白鳥大橋、室蘭港および周辺の工場群から形成される「工場夜景」の鑑賞ツアーの popularity が高まっており、特に夏季においては、地元のお祭りなどのような短期的

なイベントの他にも、上記のようなナイトツアーへの関心も SNS コメントに現れていると思われる。

次に、レベル別の地域拠点ワード発生順位を表-1 に示す。レベル 1 (地区名) に関しては、商業施設や病院などの建物が付近にある地区が発展している場所が多くなっていると考えられる。レベル 2 (大型施設) では、地球岬や白鳥大橋など有名な観光施設関係のワードが多く見受けられる。これは、市内の人だけではなく、他所からくる人も関連キーワードをツイートしていると考えられるからである。レベル 3 (小型施設) では、市内では需要が比較的に高い駅のキーワードが多くなっている。このことから、レベル毎に共通して考えられることは、人が多く集まる場所から多くの発信がされていることがわかる。

なお、抽出された Twitter データのユーザーの居住地情報は不明であるため、観光客であるのか室蘭市居住者であるのかは判断できない。したがって、本研究では、そのような区分を考慮せずに地域拠点への関心度やそれらのつながりを分析することとなる。観光という視点からの分析においては、そのような点は極めて重要な要素であることから、今後、コメントの履歴などからユーザーの居住地が対象地域の中なのか外なのかを判定するモデルを検討する予定である。

3. 地域拠点ワードの共起パターン分析

(1) 分析方法

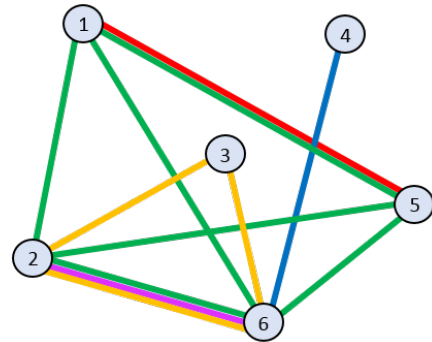
本研究では、対象地域内の各拠点への関心度に加え、関心先の関連性を把握するために以下に示すような方法で地域拠点ワードの共起パターン⁷⁸⁾分析を行った。まず、図-3a に示すように、ユーザー毎に全期間におけるコメントから地域拠点ワードを抽出した。なお、ここでは、地域拠点ワードの発生回数は考慮せず、単純に、各ユーザーの「関心先」をリストアップしただけである。次に、ユーザー毎の関心先を結ぶと、図-3b のようになる。これをまとめると図-3c のような行列データが得られる。

図-3b に示したように、例えば、地域拠点 2 と 6 は、3 名のユーザーが関心先としてワードを共に発生 (共起) させている。したがって、地域拠点 2 と 6 は、「セット」として多くの人から関心を得ていると判断できる。また、地域拠点 6 においては、これと接続 (関連) する地域拠点が 5 つと多く、中心的な存在となっている。一方、地域拠点 4 は、共に関心先となる地域拠点が 6 のみであり、他の地域拠点から孤立する形となっている。以上から、共起数が多い拠点の組

a) 対象期間におけるユーザー毎の発生ワード

ワード ↓ユーザー	1	2	3	4	5	6	総数
A	○				○		2
B				○		○	2
C	○	○			○	○	4
D		○				○	2
E		○	○			○	3

b) ワードの共起パターン※位置に意味は無い



c) ワードの共起数・単独発生数・接続数

ワード	1	2	3	4	5	6	接続数
1	—	1	0	0	2	1	3
2	1	—	1	0	1	3	4
3	0	1	—	0	0	1	2
4	0	0	0	—	0	1	1
5	2	1	0	0	—	1	3
6	1	3	1	1	1	—	5
単独発生数	2	3	1	1	2	4	

図-3 地域拠点ワードの共起パターンのイメージ

合せは、それらの間で交流 (人の往来) が生じた、もしくは生じる可能性が高いと考えることができる。

(2) 対象エリアにおける共起パターンの可視化

室蘭市市内の地域拠点ワードと室蘭市以外の道内市町村名を用いて、上記の方法で共起パターンを解析し、その結果を地理空間上に可視化したものを図-4 に示す。札幌市などの市レベルの地域と共起する傾向が確認でき、その中でも近隣都市である苫小牧市、函館市などは室蘭市とともに関心先としてワードを発生しているユーザーが多い。あくまでも WEB 空間での関心のつながりではあるが、一つの観光圏、交流圏が形成されているものと言える。

次に、室蘭市内における地域拠点ワード 186 語の場合の地理空間的共起パターンを図-5 に示す。単独発

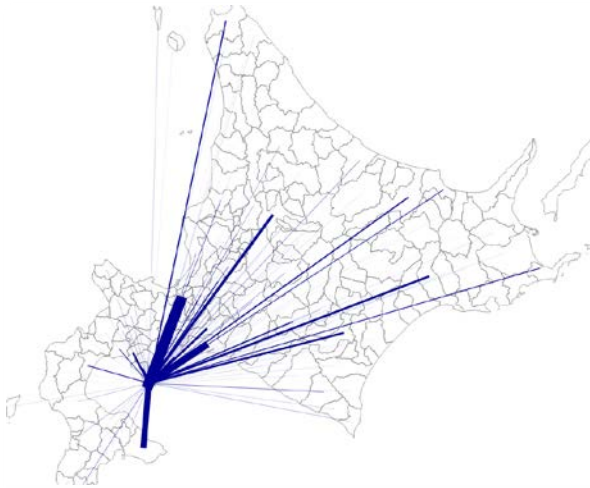


図-4 室蘭市内各地域拠点と他の市町村との共起パターン

生数は、東室蘭駅と室蘭駅の市内主要駅や、母恋、絵鞆地区、また、白鳥大橋、地球岬、イタンキ（浜）などの観光スポットが大きいことが分かる。これらの市内南側に位置する拠点は、互いの共起数が大きいことから、各拠点への関心度が大きく、これらを結ぶ大きな交流圏が生じていると考えられる。一方、中央部の住宅地や東室蘭地域の商業地区では、広範囲に渡って地域拠点ワードが発生しているが、他点との共起数は少ないことから、関心のつながりは小さいと思われる。

(3) 地域拠点ワードの中心性評価

次に、共起パターンの行列データ (図-3c) を用いて、各地域拠点の中心性について評価した。中心性指標⁹⁾とは、通信網、交通網、人間関係、食物連鎖などのネットワークにおけるノードやリンクの重要性を表す幾つかの指標であり、最近では、これらを用いてSNSにおけるユーザーのネットワーク分析などを行った事例が着目されている。本研究では、提案されている幾つかの指標のうち、次数中心性と媒介中心性を採用した。次数中心性は、接続する他点の数 (図-3c における接続数) を示し、居所的な中心性を表す。媒介中心性は、他点から接続する点群への影響度であり、全体における中心性を表す。

地域拠点ワード 186 語の次数中心性と媒介中心性を算出し、それぞれ標準化した。両指標の関係を図-6 に示す。「白鳥大橋」は、両指標ともに大きいことから、局所的にも全体としても中心的な交流拠点として評価される。また、「室蘭駅」「東室蘭駅」「地球岬」「絵鞆」なども、両指標が比較的に大きい。これらと次数中心性は同程度であるが、「室蘭港」「イタンキ」「測量山」などは媒介中心性が小

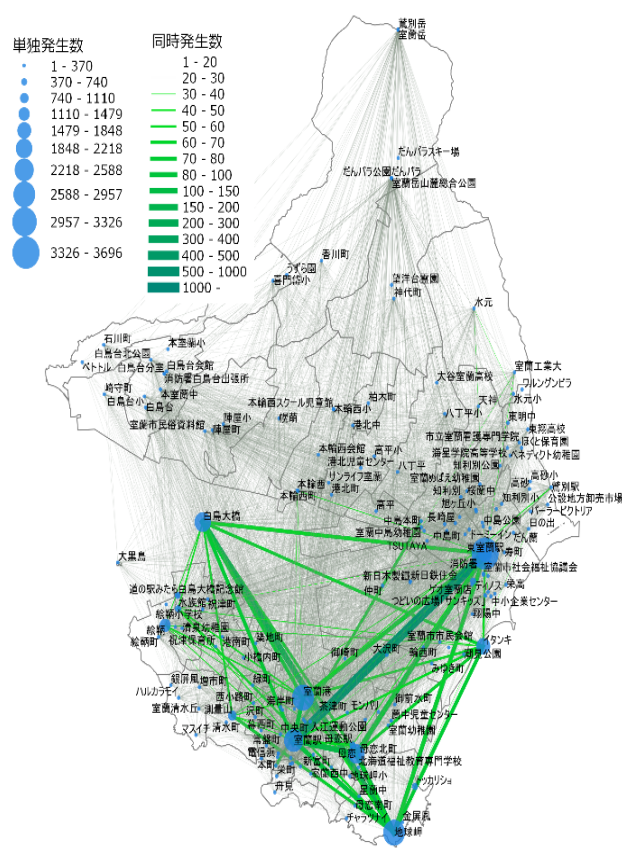


図-5 室蘭市における地域拠点ワードの共起パターン

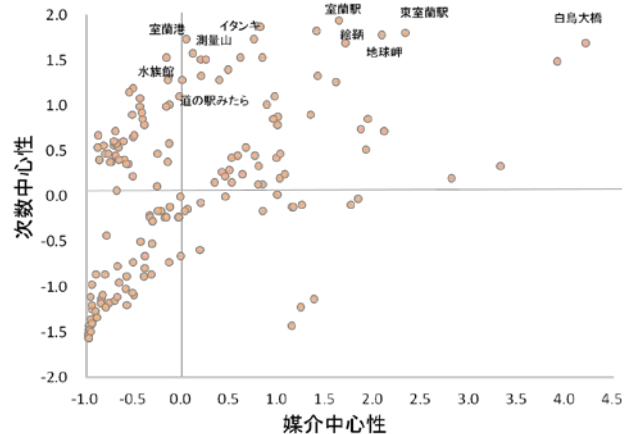


図-6 媒介中心性と次数中心性の関係

さい。次数中心性が同程度でも、媒介中心性が大きい場合は、当該ノードが大きなネットワークに属していることを意味する。したがって、言い換えると、「室蘭港」「イタンキ」「測量山」などは、「室蘭駅」「東室蘭駅」「地球岬」「絵鞆」などよりも、小さな共起パターンのネットワークに存在することから、関心空間上のネットワーク内での露出が少ない、いわゆる「ローカル」な拠点である可能性が高い。

以上のように、中心性指標によって各地域拠点の特性を評価することができるが、これだけでは説明できない部分も大きい。今後は、他の指標も導入しつつ、ネットワークを分解して、より詳細な特性を明らかにしたい。

4. 結論

本研究では、室蘭市を対象とし、Twitter の投稿コメントデータから抽出した地域拠点ワードをユーザーレベルで関連付けし、その共起構造について分析した。その結果、まず、単独での地域拠点ワードの発生数を求め、その順位から、地域内における関心度の高い拠点を把握した。次に、地域拠点ワードの共起パターンを地理空間上で示し、対象地域における各拠点のロケーションと拠点間の関連度を同時に把握した。さらに、そのデータから中心性指標を算出し、各拠点の特性について明らかにした。

以上のような結果が得られたが、課題は多い。まず、186 語の地域拠点ワードを設定したが、対象地域内の拠点をすべてカバーしているとは言いがたい。また、ユーザーの居住地が不明であるため、地域の居住者、観光客の区別がつかない。今後は、形態素解析などの言語処理や機械学習などの技術を用いて、コメントの抽出や居住地を判定するモデルを構築する予定である。

参考文献

- 1) 藤生慎, 高田和幸, 富田智之, 伊東大輔: ゲリラ豪雨時の社会現象に関する一考察—twitter を用いた検討—, 土木計画学研究・講演集, Vol.49, 論文番号 190, 2014.
- 2) 榎本甫, 桑野将司, 小池淳司: 災害時のソーシャルメディアと行動の関連性分析, 土木学会論文集 D3 (土木計画学) Vol.70, No.1, pp.102-112, 2014.
- 3) 岡野大輔, 久保田尚, 上野俊司, 花村嗣信: Twitter を活用した新しいバス情報システムの普及可能性に関する研究, 土木計画学研究・講演集, Vol.44, 論文番号 163, 2011.
- 4) 矢野晋哉, 伊藤秀昭, 安田幸司: 道路開通に関するツイッター情報の分析事例, 土木計画学研究・講演集, Vol.48, 論文番号 203, 2013.
- 5) 富永透見, 肥後洋平, 谷口守: ネットコンシャスな拠点計画に関する一考察-チェックインスポットに着目して-, 土木計画学研究・講演集, Vol.50, 論文番号 37, 2014.
- 6) 谷口守, 星野奈月, 富永透見: 自治体の観光資源に対する認識と SNS 上の「話題」とのギャップ分析, 土木計画学研究・講演集, Vol.50, 論文番号 35, 2014.
- 7) 藤井雄太郎, 安藤哲志, 伊藤孝行: 2 単語間の共起情報と距離情報を考慮した有害文章判別手法の提案: 情報処理学会会研究報告, Vol.2010-ICS-159, No.1, 2010.
- 8) 石田和成: 地域特有の単語共起にもとづく位置推定と地域トピックの考察: 研究報告情報基礎とアクセス技術 (IFAT), Vol.2015-IFAT-117, No.2, pp.1-6, 2015.
- 9) 篠田孝祐, 松尾豊, 中島秀之: 複数の中心性尺度によるネットワーク生成モデル, 知能と情報 (日本知能情報フュージ学会誌), Vol.20, No.3, pp.410-422, 2008.