

Wi-Fi パケットセンシングによる 広域観光圏における 時空間周遊行動パターン分析

遠藤 幹大¹・高橋 央亘²・浅田 拓海³・有村 幹治⁴

¹学生会員 室蘭工業大学 大学院工学研究科 環境創生工学系専攻 (〒050-8585 室蘭市水元町27-1)
E-mail:18041012@mmm.muroran-it.ac.jp

²学生会員 室蘭工業大学 大学院工学研究科 環境創生工学系専攻 (〒050-8585 室蘭市水元町27-1)
E-mail:17041037@mmm.muroran-it.ac.jp

³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

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「旭川富良野広域圏 Wi-Fi 調査」によって収集されたデータを用いて周遊パターンの分析を行った。マーケティング分野などで使われる「系列パターンマイニング」を用い、各スポット間の訪問順列を考慮した分析を行った。それにより検出された訪問順列パターンと三つの指標から考察と観光施策の提案などを行い、大規模周遊行動パターン分析における系列パターンマイニングの有用性を示した。

Key Words : *Travel pattern, Wi-Fi, Big Data, Sequence pattern mining*

1. はじめに

近年の北海道ではインバウンド観光客を含め、多くの観光客が一年を通して訪来し賑わいをみせている北海道は広域分散型地域構造であり、都市や観光スポットが点在しており、時空間的距離が長いと言える。また北海道観光の主な移動手段は自動車のため、観光シーズンでは渋滞が発生しやすいと言った問題もある。

それらに伴い北海道の各観光スポット間では周遊観光客に対する広域的な観光地域間連携が求められるが、これらの検討に当たっては、観光客の周遊行動の時空間的な観測と分析が基礎情報として必要となる。

観光客の周遊行動調査としては、従来アンケート調査が一般的である。しかし、結果が回答者の主観と記憶に依存するため正確性に欠ける場合があり費用や実施回数による制限もあるためサンプル数が少なくなる傾向にある。また昨今では各種のプローブ調査も採用されている。しかしながら、これらも調査コストや回答結果のバイアス、観測デバイスの普及率の問題を含み、特に広域

的、長期的な対象では適用が難しい。

本研究では Wi-Fi パケットセンシングを用いて行われた旭川富良野広域 WiFi 調査で得られた結果を元にデータのクレンジングを行い、さらに基礎集計を行って各スポットの来客数推移などの基礎的な情報を得ることができた。さらにそのデータを用いて時系列データの解析に用いられる手法の一つである系列パターンマイニングを用いて得られた 3 つの指標からスポット間の相関関係の把握を行なった。

2. 関連研究

望月ら¹⁾は防災計画、災害対応へ利用するために Wi-Fi パケットセンサを用いた匿名人流解析システムのシステム構築についての研究を行っている。これは MAC アドレスを収集し、そのデータを分析用サーバにアップロードすることでリアルタイム観測を行おうとするものである。なお、取得する MAC アドレスは各端末に固有に割

り振られているが、ハッシュ関数により変換が行われ収集したデータの匿名性を確保している。使用したセンサーは Raspberry PI をベースに制作されているため、単価が安価であり、複数個所に設置することが可能である。Wi-Fi 通信機能を持つ通信機器が定期的に発信する Probe Request を受信する為、基本的に観測対象者のハードウェア及びソフトウェアに手を加える必要がないのが特徴である。

田中ら²⁾は Wi-Fi パケットセンサーを用いた調査と実際の施設の集客数、さらにアンケート調査との比較から Wi-Fi パケットセンサーを用いた調査の有意性を示した。パーソントリップ調査の考え方である「誰が、いつ、OD、目的、交通手段、移動経路」の観点から調査がなされている。更に美ら海水族館入館者数とこの調査方法から得られた入館者数を比較し、比較的安定していることを示した。本調査法およびア OD と来場者数を比較し、アンケート調査では得られない OD の導出が可能であることを示し、アンケート調査の問題点である、回答者の記憶の曖昧さや訪問者が相対的に少ない観光地などの確認ができ、Wi-Fi パケット調査の意義が確認できた。

福田ら³⁾は広域圏における観光周遊行動の調査のために Wi-Fi 機器からのパケット送信の情報に基づくパッシブ型広域観光周遊行動調査を行った。得られたデータに対し、4つの項目によるクレンジング手法を提案し観光客に絞った基礎集計を行った。

3. データ概要

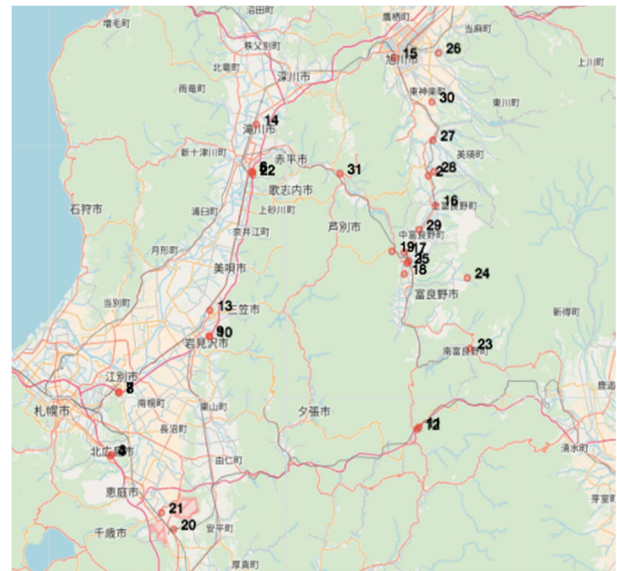
本研究では、福田ら³⁾により実施された旭川富良野広域圏 Wi-Fi 調査にて収集されたデータを使用する。この調査では、地域未来研究所(株)の作製した Wi-Fi パケットセンサーを、対象エリアの観光スポット、高速道路 SA・PA、公共施設等に設置(図-1)した。調査期間は、2017年6月19日(水)～7月23日(日)の33日間とし、夏季における富良野観光シーズンの周遊パターンについて分析する。

(1) データのクレンジング

福田ら³⁾の研究を参考に、以下の手順でデータクレンジングを行った。

a) 複数箇所設置のセンサーのデータ統合

使用した Wi-Fi パケットセンサーの受信範囲はおおよそ 400m であるが、周辺の建物、人混み、天候などによってユーザーからの管理パケットの受信が困難となる場合が生じる。そのため、大型施設となるスポットには複数箇所センサーを設置した。分析においてはこの複数箇所の受信データをそのスポットのデータとして結合した。



No	name	No	name
1	富良野 ナチュラルックスホテル	16	フラワーランドかみふらの
2	ラベンダーオーナー園	17	カンパーナ六花亭
3	輪厚 PA 上り	18	富良野チーズ工房
4	輪厚 PA 下り	19	ラベンダーの森 ハイランドふらの
5	砂川 SA 下り	20	新千歳空港トヨタレンタカー
6	砂川 SA 上り	21	道の駅サーモンパーク千歳
7	野幌 PA 下り	22	砂川ハイウェイオアシス
8	野幌 PA 上り	23	道の駅南ふらの
9	岩見沢 SA 下り	24	五郎の石の家
10	岩見沢 SA 上り	25	フラノマルシェ
11	占冠 PA 下り	26	旭山動物園
12	占冠 PA 上り	27	道の駅びえい丘のくら
13	道の駅三笠	28	四季彩の丘
14	道の駅たきかわ	29	中富良野町営ラベンダー園
15	道の駅あさひかわ	30	旭川空港トヨタレンタカー
		31	道の駅スタープラザ芦別

図-1 Wi-Fiパケットセンサーの設置箇所

b) スポット到着時・出発時のデータ抽出

元データでは、数秒間隔で連続的に受信記録が残っているが、スポット内での初回受信(到着)時と最終受信(出発)時の記録以外は不要であるため除外した。

c) 周遊期間 10 日以上ユーザー除外

上記のスポット毎の到着、出発データを MAC (ユーザー) 毎に時系列で並べ、最初のスポットの到着時と最後のスポットの出発時の差分から、当該ユーザーの周遊時間を求めた。10 日以上滞るは、地元住民や業務目的のユーザーである可能性が高い。観光客のみの周遊行動を対象とするため、10 日以上ユーザーは除外した。

(2) 基礎集計

クレンジング後のデータから、スポット毎に、日別および時間帯別の訪問客数（MAC アドレス合計値）を集計した。一例として、「ラベンダーの森ハイランド富良野」の結果を図-2、図-3 に示す。日別訪問客数の推移を見ると、休日をピークに変動していることがわかる。また、ラベンダー見頃の夏に向けて、増加していることがわかる。さらに詳細な時間帯別の集計では、昼 12~13 時と夕方 18 時頃にピークがあることが確認できる。また、日によって、傾向が変わっており、この理由としてはイベントや天候などの影響が考えられるが、これについては今後の課題とする。

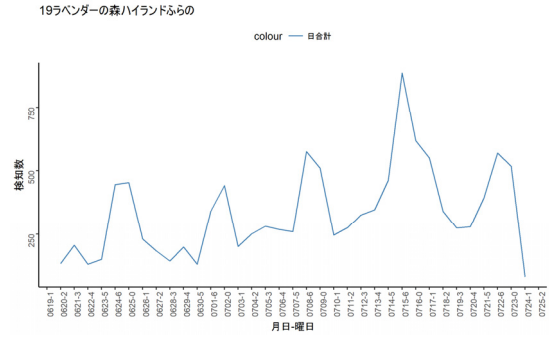


図-2 日別訪問客数
(ラベンダーの森ハイランドふらの)

4. 系列パターンマイニングによる周遊行動分析

(1) 系列パターンマイニングの概要

本研究では、観光客の周遊パターン、すなわち、どのような順番でスポットを巡っているのか、また、どのような巡回パターンが代表的なのかについて検討する。このような順番を考慮したパターンの抽出手法として、系列パターンマイニングがある。系列パターンマイニングは、マーケティング分野や SNS、テキストマイニングなどのビッグデータ分析で使われる相関ルール分析の一種であり、「商品」がどのような組み合わせで、どのような順番で「購入」されるのかを定量的に分析するために用いられている。本研究のような大規模な観光周遊調査から得られるデータへの適用は稀であるが、「商品」を「訪問地」、「購入」を「滞在（立ち寄り）」と置き換えることで同様の分析が可能と考えられる。系列パターンマイニングでは、1 ユーザーが各スポットを周遊した時、それらスポットの順列をトランザクションと呼ぶ。さらに、トランザクションは条件部 X（途中観測地順列）と帰結部 Y（最終観測地）に区分される（図-4）。これら X と Y により、以下の3つの指標が算出される。

・支持度 (Support)

以下の式(1)で表されるように、全トランザクションにおける対象トランザクションの発生割合である。すなわち、この値が高いと、X のパターンで周遊するユーザーが多いと判断できる。

$$\text{Support}(X,Y) = \frac{X,Y \text{ を満たすトランザクション数}}{\text{全トランザクション数}} \quad (1)$$

・確信度 (Confidence)

式(2)で表される。ある途中観測地順列 X を考えたとき、その最終観測地 Y は複数存在し、各 Y の発生確率がそ

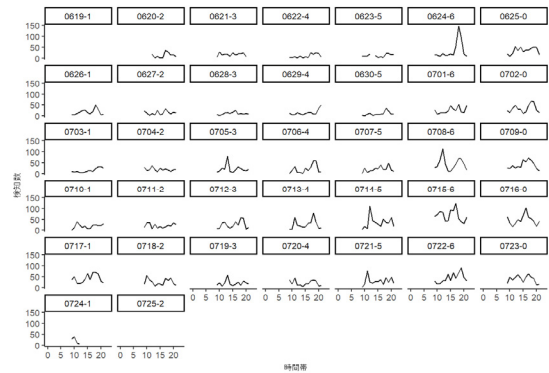


図-3 時間帯訪問客数
(ラベンダーの森ハイランドふらの)

mac	トランザクション (全観測地順列)	
	途中観測地順列：条件部	最終観測地：帰結部
1	D→C→B→	A
2	A→D→C→	B
3	E→	D
4	D→C→B→	A
5	D→C→B→	E
6	D→B→	A
7	A→D→	C
8	C→D→A→B→E	F

計算例

$$\begin{aligned} \text{Support}(D \rightarrow C \rightarrow B, A) &= 2/8 = 0.25 \\ \text{Confidence}(D \rightarrow C \rightarrow B, A) &= 2/3 = 0.67 \\ \text{Lift}(D \rightarrow C \rightarrow B, A) &= \frac{0.67}{3/8} = \frac{0.67}{0.36} = 1.78 \\ \text{Lift}(D \rightarrow C \rightarrow B, E) &= \frac{0.67}{1/8} = \frac{0.67}{0.125} = 5.33 \end{aligned}$$

図-4 系列パターンマイニングの概要

それぞれの確信度となる。

$$\text{Confidence}(X,Y) = \frac{X \text{ と } Y \text{ を共に満たすトランザクション数}}{X \text{ を満たすトランザクション数}} \quad (2)$$

・リフト値 (Lift)

式(3)で表されるように、確信度 (X, Y) を、Y を含むトランザクション発生確率で除した値であり、X と Y の相関度を表す。最終観測地 Y への訪問確率が低いほどリフトが大きくなることから、途中観測地順列 X の増加による Y への訪問数増加の効果を表すと言える。

$$\text{Lift}(X, Y) = \frac{\text{Confidence}(X, Y)}{\text{Support}(Y)} \quad (3)$$

(2) 対象エリアの周遊パターン全体像

算出した上記 3 つの指標の関係性から、対象エリア・期間における周遊パターンの傾向を把握する。なお、系列パターンマイニングでは、意味のあるトランザクションを抽出するために、支持度、確信度、リフトにおいて閾値を設定し、データのスクリーニングを行う場合が多い。本研究における周遊パターン (トランザクション) 数は極めて多いことから、支持度は小さい傾向にある。そのため、支持度の閾値 0.001 を設定した。

支持度と確信度の関係を図-5 に示す。支持度、確信度がともに小さい範囲に集中しており、両指標がともに大きいケースはほとんどない。支持度が大きい図中 A では、これらのトランザクションを代表的な訪問パターンと捉えることができるが、確信度が小さいことから最終的な訪問地へのアクセスは定着していないと言える。一方、確信度が大きい図中 B では、訪問パターンの発生は少ないものの、途中までの訪問から、その後高い確率でアクセスさせる訪問地 (最終観測値 Y) が抽出されている。

次に、支持度とリフト値の関係を図-6 に示す。図-5 と同様に、両指標が共に大きいトランザクションはほとんど無い。支持度が大きい図中 C では、その訪問パターンの発生は比較的多いものの、最終的な訪問地とそこまでの訪問地の相関性は小さい。リフト値が大きい図中 D では、途中までの訪問パターンと最終的な訪問地に強い相関があり、現状では、最終的な訪問地へのアクセスが少ないことから、今後の促進施策の対象となり得る潜在性の高いスポットや訪問パターンが抽出されている。

最後に、確信度とリフト値の関係を図-7 に示す。リフト値は、確信度を最終目的地 Y の発生確率で除した値であることから、おおよそ相関関係にある。確信度が高い図中 E は、途中の訪問地から最終的な訪問地への典型的なアクセスパターンと捉えることができるが、前者と後者の相関性は低い。一方、確信度は低いが、リフト値が大きい図中 F は、途中訪問地から最終訪問地へのアクセス発生は少ないが、両者の相関は強い。したがって、途中訪問地の促進によるその後の訪問地へのアクセス向上効果が大きいことから、観光周遊施策として重要なスポ

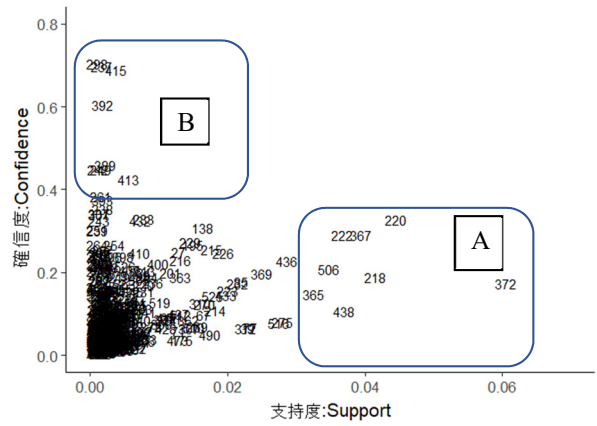


図-5 支持度と確信度の関係

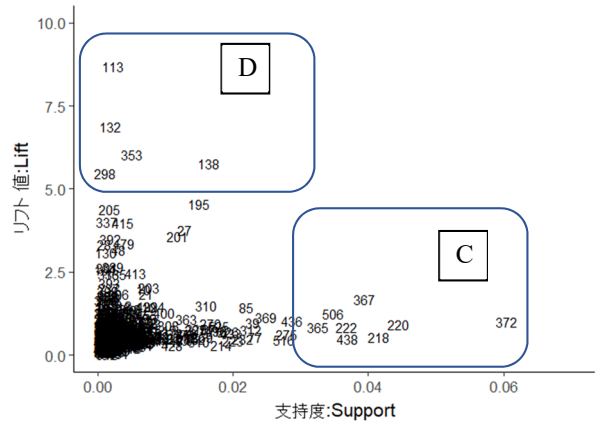
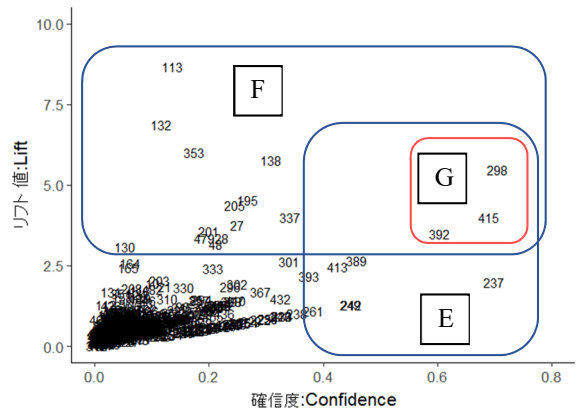


図-6 支持度とリフト値



の関係

ットおよび周遊パターンであると言える。両者の値が比較的高い図中 G では、観光施策の実施などによる途中観測値順列の訪問客数増加の影響が効率的かつ効果的に最終観測値の訪問客数に影響を与えると考えられる。

(3) 対象エリアにおける代表的な周遊パターン

以上の 3 つの指標の関係から、対象エリアにおける周遊パターンの全体像を把握する事ができた。次に、各指標が大きいトランザクションを抽出し、対象エリアにおける広域観光の代表的な周遊パターンおよび訪問客数増加に向けた方策について考察を行う。考察のために、

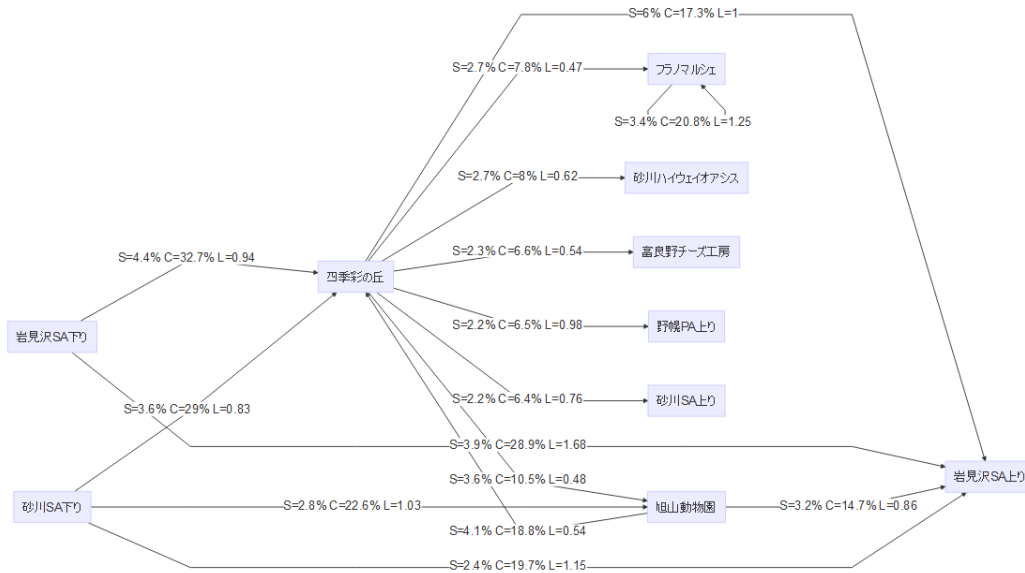


図-8 支持度上位 15 のトランザクション

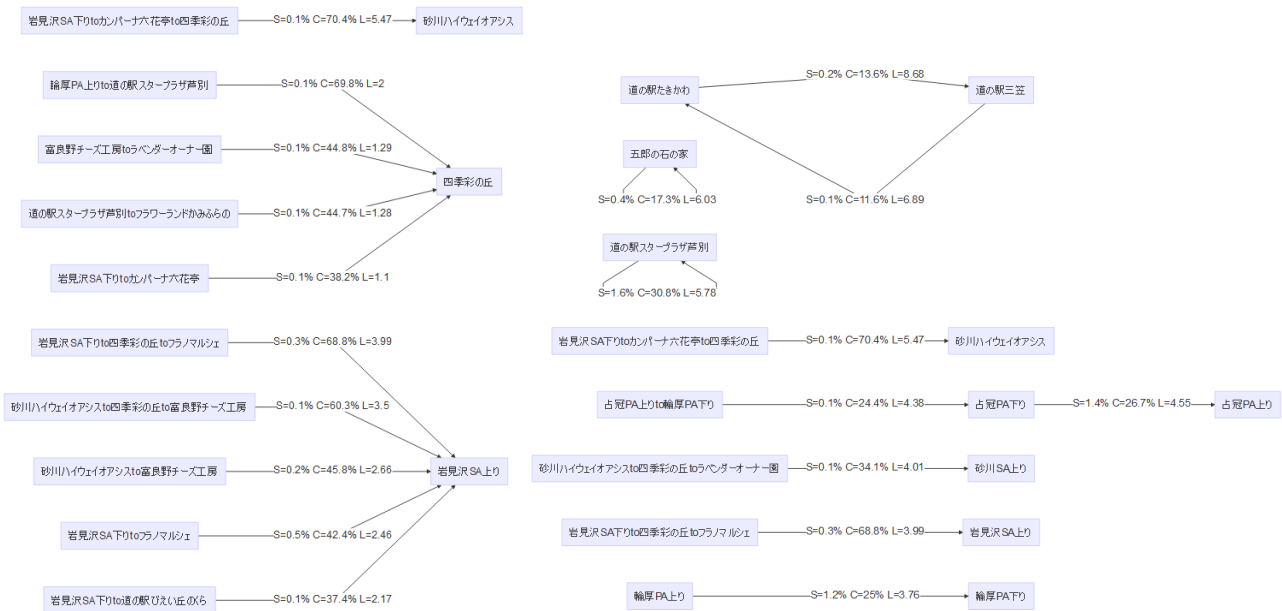


図-9 確信度上位 10 のトランザクション

図-10 リフト値上位 15 のトランザクション

支持度，確信度，リフト値において，それぞれ値が大きい順に並べ，上位のトランザクションを抽出し，ネットワーク図として可視化した（図-8，図-9，図-10）．これらの図には，X（途中観測地順列）から Y（最終観測地）への方向を矢印線で示し，また，その線上に支持度（S），確信度（C），リフト値（L）を記載している．

支持度上位 15 のトランザクションを図-8 に示す．四季彩の丘や旭山動物園が全体のハブ的なスポットになっていることがわかる．また，例えば，岩見沢 SA 下りや砂川 SA 下りを経由して，四季彩の丘に行くパターンが発生率が 4%程度あり，四季彩の丘の後にフノマルシェ等観光スポットでの周遊で終わるケース，また，SA・PA 上りを最終地とするパターンも見られる．これは，主に札幌方面から富良野エリアに観光で訪れるパタ

ーンと考えられる．これらは対象ユーザー全体における代表的，典型的なパターンであると言えるが，確信度やリフト値は低い．したがって，最終地へのパターン化は定着しておらず，また，現状においてすでに多くの訪問客が訪れていると考えられることから，今後，促進すべき周遊パターンであるとは言えない．

次に，確信度上位 15 位のトランザクションを図-7 に示す．最も確信度が大きいケースとして，「岩見沢 SA 下り→カンパーナ六花亭→四季彩の丘」の後に 7 割のユーザーが砂川ハイウェイオアシスに訪問しているパターンが見られた．この四季彩の丘や岩見沢 SA 上りは，他の行程からの確信度も高く，これらに向かう周遊コースはパターン化されていると言える．四季彩の丘は，無料で利用できることから，観光客が集中する場合が多い．

これらの途中経由地の情報を見て、四季彩の丘の訪問者数を予測しつつ、途中訪問地での滞在時間をコントロールすることができれば、より満足度の高い周遊を提供できると思われる。また、岩見沢 SA 下りについては、富良野エリアでの観光の後、札幌への帰路の際に立ち寄る典型パターンと言える。観光への関心が高い人々が立ち寄るこの SA において、観光案内等によるリピーター確保や他の地域への関心拡大など、効果的な観光施策を展開できると考えられる。

最後に、リフト値上位 15 位のトランザクションを図 8 に示す。道の駅たきかわと道の駅三笠の双方向のリフト値が極めて高い。確信度が 10% と低く、パターンとして定着されているとは言いがたいが、これは、昨今、この地域で需要が高まっているワインツーリズムにおける周遊パターンを表しているものと推測される。現状では両者ともに訪問者は多くはないが、連携することで、互いに集客数が大きく増加する見込みがあると言える。五郎の石の家や道の駅スタープラザ芦別は、それらの中で行程が巡回している。後者に関しては、札幌側と富良野側の経由地であることから、日をまたいだ往復時にそれぞれ立ち寄っているものと思われる。後者は、周遊期間中に観光地として複数回訪問されている可能性が高い。これらについては、元のデータから確認する必要がある。その他、最終目的地として SA・PA が多いことから、富良野圏での周遊増加が、高速道路の交通量増大、SA・PA 利用数に大きく影響を及ぼしているものと推測される。特に、「岩見沢 SA 下り→カンパーナ六花亭→四季彩の丘→砂川ハイウェイオアシス」や「岩見沢 SA 下り→四季彩の丘→フラノマルシェ→岩見沢 SA 上り」は、確信度が 60% 以上となり、最終地の訪問者増加を予測することができることから、将来における富良野エリアの観光客増加の予想の下、SA における適切な集客容量などに関する検討が可能となる。

5. まとめ

本研究では、Wi-Fi パケットセンシング調査により得られたデータに系列パターンマイニングを適用して、富良野エリアにおける広域観光周遊パターンの分析を行った。その結果、支持度、確信度、リフト値の 3 つの指標の関係性から対象エリアにおける周遊パターンの全体像を把握できることを示した。また、各指標の上位のトランザクションをネットワーク図として表現することで、途中訪問地と最終訪問地の関係を容易に把握できることを示し、それらの詳細から周遊観光施策に関する提案を行った。今後は、短期間毎に算出した上記の指標による訪問客数のリアルタイム推定方法などを検討し、Wi-Fi パケットセンシングの活用の幅を広げていきたい。

謝辞：本研究は国交省・道路政策の質の向上に資する技術「ETC2. 0 プローブ情報等を活用したデータ駆動型交通需要・空間マネジメントに関する研究開発」からの支援を受けて行われた。

参考文献

- 1) 望月祐洋, 上善恒雄, 西田順二, 中野秀男, 西尾信彦: Wi-Fi パケットセンサを利用した匿名人流解析システムの構築, 情報処理学会研究報告, Vol.2014-MBL-70 No.45, 2014.
- 2) 田中謙大, 神谷大介: 周遊観光行動の調査方法に関する基礎的考察, 第 72 回土木学会年次学術講演会講演概要集, 2017.
- 3) 福田大輔, 小林巴奈, 五百蔵夏穂, 田中謙大, 中西航, 浅田拓海, 有村幹治, 内田賢悦, 菅芳樹: Wi-Fi パケットセンシングによる北海道広域観光周遊行動調査, 観光情報学会第 16 回研究発表講演論文集, pp. 58-61, 2017.

(2018.4.27 受付)

AN ANALYSIS OF SPATIAL-TEMPORAL PATTERN IN WIDE TOURISM AREAS BY USING WI-FI PACKET SENSING

Mikihiro ENDO, Hironobu TAKAHASHI, Takumi ASADA
and Mikiharu ARIMURA

In recent years, the number of tourists in Hokkaido has increased. The attraction spots (e.g., tourist spot, hotel, SP/AP) and trips information can be collected for traveler prediction. The distance and travel time affected to traveling pattern in Hokkaido area, because of the most attraction spots are far away and take time. In this research, cruise pattern was analyzed by using Wi-Fi survey data, which collected in "Asahikawa and Furano wide area". The first, conducted a basic analysis and examined of each spots. Next, the three indicators such as support, confidence and lift parameters calculated by "sequence pattern mining analysis", which generally used in the marketing purposes. As results, usefulness of sequence pattern mining model provided the large scale of migration behavior patterns.