

非階層型クラスタリング手法を用いた 観光地におけるプローブリクエストの分類

菅原 直樹¹・岡部 晃之²・菊池 輝³・泊 尚志⁴

¹学生会員 東北工業大学大学院 工学研究科 土木工学専攻 (〒982-8577 宮城県仙台市太白区八木山香澄町35-1)
E-mail:m204801@st.tohtech.ac.jp

²学生会員 東北工業大学 工学部都市マネジメント学科 (〒982-8577 宮城県仙台市太白区八木山香澄町35-1)
E-mail:s1614209@st.tohtech.ac.jp

³正会員 東北工業大学教授 工学部都市マネジメント学科 (〒982-8577 宮城県仙台市太白区八木山香澄町35-1)
E-mail:akikuchi@tohtech.ac.jp

⁴正会員 東北工業大学准教授 工学部都市マネジメント学科 (〒982-8577 宮城県仙台市太白区八木山香澄町35-1)
E-mail:tomari00@tohtech.ac.jp

近年、スマートフォン等のWi-Fi通信機器の普及に伴い、様々な場所にWi-Fiアクセススポットが設置されている。Wi-Fi通信機器はWi-Fi機能をONにしている場合、Wi-Fiアクセススポットに向けてプローブリクエスト(PR)と呼ばれる信号を発している。このPRを受信するセンサを用いた観光調査が近年盛んに行われている。既往研究では取得したPRの中から観光客のデータのみを抽出するため、データクリーニングが行われている。しかし、観光地によって異なる抽出条件を設定しなければならず、汎用的な手法ではない。そこで本研究では、データクリーニングに変わる手法として、クラスタリングによるPRデータの分類を実施する。クラスタリングによってどのようなカテゴリに意味づけされるのか確認する。

Key Words : *Wi-Fi packet sensing, MAC address, Tourists' movement, Clustering*

1. はじめに

近年、スマートフォン等のWi-Fi通信機器の普及に伴い、商業・公共施設や駅構内、観光地など様々な場所にWi-Fiアクセススポットが設置されている。Wi-Fi通信機器はWi-Fi機能をONにしている時、通信機器がアクセススポットと通信を行うためにプローブリクエスト(以下、PR)と呼ばれる管理パケットを発信している。PR内にはMACアドレス(以下、ID)と呼ばれる通信機器固有の識別子が含まれている。このIDを含むPRを取得するWi-Fiパケットセンサを用い、観光客の回遊行動を分析する研究が全国各地で盛んに行われている。いずれの既往研究^{1)~5)}も取得データから観光客のデータを抽出する際、データクリーニングにより滞在時間や滞在日数等でフィルタをかけ、観光客と思われるIDを抽出している。

(1) 既往研究

一井ら¹⁾はWi-Fiパケットセンサを長野県小布施町に2日間設置し、滞在人数や回遊経路を明らかにし、観光回遊行動の実態把握を行なっている。このエリアでは日帰りを訪れる観光客が多いことから、データクリーニングでは、複数日に渡って観測されたデ

ータを除外した。また、5分以上連続してデータが得られた場合にのみ滞在とみなし、それ以外のデータを除くことで通過交通のデータを分析対象から除外した。次に設置した27地点に関して時間帯別滞在者比率を説明変数としたクラスタリングを実施し、設置センサを4つのクラスタに分類した。最終的に小布施町において、観光回遊行動の実態が把握可能であると示されていた。

壇辻ら²⁾はWi-Fiパケットセンサを奈良県桜井市の長谷寺参道に2日間センサを設置し、回遊行動分析を通して特性を明らかにした。基礎分析として簡易的なアンケート調査から、自動車交通の現状や観光客の個人属性、旅行形態、観光行動について把握した。基礎分析の結果よりデータクリーニングでは2日間観測されたIDを除外した。その後、起終点の組み合わせと滞在時間によって鉄道駅・駐車場を利用した観光客を分類し、各グループの滞在時間特性を明らかにした。さらに来訪した時間帯の違いに着目して、データを細分化し分析した結果、滞在時間の分布に違いが存在していることを明らかにした。

堀籠ら³⁾は、宮城県女川町の狭域においてどの程度回遊行動を把握できるのか検証した。データクリーニングでは、「滞在時間10~480分・初観測時刻8

～23時台・滞在日数2日以内」の3つの条件を設けてデータクリーニングを行い、観光客以外のデータを除去した。集計・分析により、狭域において日別時間帯別の観光客の推移や滞在時間の算出は適用可能であることが示された。一方で観光客の域内移動に関して、センサ間のデータ取得範囲の重複など、課題が残る結果となった。

(2) 本研究の位置づけ

いずれの研究でも、Wi-Fiパケットセンサで取得した観測データで、データクリーニングから観光客のデータを抽出していたが、抽出条件は各研究で異なっていた。しかし、このプローブリクエストデータの量は膨大であり、クリーニングを施しても、その抽出条件が本当に正しく、観光客が含まれているデータであるのかは定かではない。

そこで本研究では、約1ヶ月間の長期において取得したデータを用いたクラスタリングを実施し、観光客を含むカテゴリの分類を行う。複数の説明変数によるクラスタリングによってカテゴリ別にどの程度まで特徴の抽出ができるのか、そして、どのようなカテゴリに分類できるのかを確認する。

2. 調査概要

(1) Wi-Fiパケットセンサ

本研究で用いたWi-Fiパケットセンサは、Raspberry Pi・無線LANアダプタ・micro SDカード等を用いた簡易に構築可能なデバイスである(図-1)。これを防水・防塵ケースに内含し、屋根程度の高さに設置した。

本センサはtshark⁶⁾と呼ばれるソフトウェアを用い、スマートフォン等のWi-Fi通信機器の発するPRを受信し、機器のIDと観測時刻を記録する。そのPRデータに各センサに対応した番号を付加した後、各センサのデータを1つに集約し表-1のように整理した。

(2) 調査概要

調査はJR女川駅前商店街を対象とし、センサを9台設置した。データ取得期間は2019年12月1日～2020年1月3日の34日間である。

また、センサの設置場所を選定するため、調査実施前に予備調査を行った。検証用のスマートフォン4台を用い、地域内の駐車場や施設の出入口・メイ



図-1 Wi-Fiパケットセンサ

表-1 PRデータの整理例

センサNo.	日付	時間	ベンダー	MACアドレス	電波強度	シーケンス番号
1	20191201	18:25:38	Apple	aa:bb:cc:01:23:45	-71	1012
5	20191201	18:28:54	Apple	aa:bb:cc:01:23:45	-73	1033
4	20191201	18:29:03	Apple	aa:bb:cc:01:23:45	-91	1056
11	20191201	18:29:21	Apple	aa:bb:cc:01:23:45	-88	1076
⋮	⋮	⋮	⋮	⋮	⋮	⋮

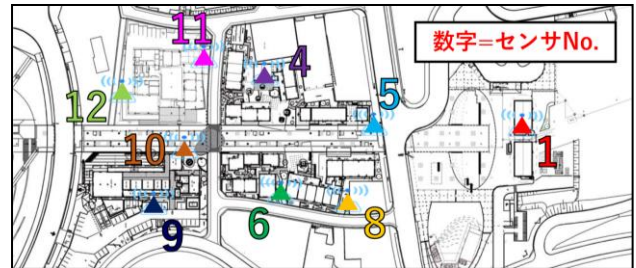


図-2 各センサの設置場所

ンストリート等の地点でPRデータを取得し、およそのセンサ範囲を推定した。予備調査の結果を踏まえて女川駅・シーパルピア女川・駐車場の全域を網羅するようセンサ位置を選定した。各センサの設置箇所を図-2に示す。

3. クラスタリングによるPRデータの分類

(1) 前処理

正確なクラスタリングを行うために、結果に大きく影響を及ぼす可能性があるPRデータを除外する必要がある。「パケット数が1つのみのID」は滞在時間の算出が不可能な為、除外した。また、「ベンダー”Google”のID」は全PR中の9割程度を占めており、その中でもパケット数1つまたは2つのIDが大半で、クラスタリング結果に影響を及ぼす可能性が高いため、除外した。また、一部の通信機器には個人情報保護を目的でIDがランダム化される機能が搭載されている。ランダム化されたIDでは同一IDの特定が不可能なため、同じく除外した。

(2) クラスタリング

除外後のID数は34日間合計で14410件であった。しかしこのIDは観光客だけではなく、従業員やOA機器、地域住民等も含まれている。そこで、これらのカテゴリを分類するためにk-means法を用いた。クラスタリングに用いた説明変数の一覧を表-2に示す。祝日は、対象地域のイベント日または年末年始の8日間、休日はイベント開催日を除く土日の7日間、平日はその他19日間とそれぞれ設定した。この説明変数の組み合わせを用い、12種のクラスタリングを行った。12種の各クラスタリングに使用した説明変数を表-3に示す。クラスタ数の選定には、各クラスタリングでエルボー法を行い、クラスタ数別のSSE値(クラスタ内誤差平方和)をプロットし、最適なクラスタ数を算出した。各エルボー法の結果、最適なクラスタ数は5～7つの間に留まったが、各結果を比

較するため、クラスタ数は一律で7つと設定した。紙面の都合上、No.12の結果(表-4)を例に特徴づけられた意味を読み取れるか検証する。表-4よりクラスタ1やクラスタ2、クラスタ3は観測日数が約1、2日のグループであることが分かる。クラスタ6やクラスタ7は観測日数が多く、滞在時間も長いことからOA機器や従業員の可能性が高い。クラスタ4やクラスタ5では際立った特徴は得られなかった。次に結果を裏付けるためにクラスタ毎に日別・時間帯別観測ID数を集計した。紙面の都合上、従業員らしい特徴を持つクラスタ6と最も観光客らしい特徴を持つクラスタ2について報告する。

a) クラスタ6の特徴

図-3はクラスタ6の日別ID数、図-4はクラスタ6の時間帯別の観測ID数・開始時間帯ID数・終了時間帯ID数をそれぞれ示す。各集計には、各IDの観測日別に始観測時間帯・最終観測時間帯・時間帯別に観測の有無(0:未観測, 1:観測)のデータベース(表-5)を作成した。図-3は日別に観測したID数を表す。図-4の開始・終了時間帯ID数は初観測・最終観測時刻を時間帯別に集計している。時間帯別観測ID数は表

表-2 クラスタリングに使用した説明変数一覧

説明変数名	説明
(ア) 観測日数(日)	期間内に観測された日数
(イ) 時間帯別滞在開始時間数(個)	1日の初観測時刻を時間帯別にカウントしたもの。定めた時間帯は8時台, 9-11時台, 12-13時台, 14-15時台, 16-22時台, その他の6つ
(ウ) 時間帯別滞在開始時間割合	(イ) 時間帯別滞在開始時間数(個)を(ア) 観測日数で除したものの
(エ) 平休祝日数(日)	観測された平日・休日・祝日のそれぞれの日数
(オ) 平休祝日割合	(エ) 平休祝日数(日)を(ア) 観測日数で除したものの
(カ) 平均滞在時間(分)	滞在時間の和を(ア) 観測日数(日)で除したものの
(キ) 平均滞在時間パラメータ	(カ)の値によって1~6のパラメータを付与 1: 10分未満 2: 10分以上60分未満 3: 60分以上90分未満 4: 90分以上180分未満 5: 180分以上360分未満 6: 360分以上

表-3 各クラスタリングに用いた説明変数

No.	説明変数	No.	説明変数
1	(イ), (カ)	7	(ア), (ウ), (カ)
2	(エ), (カ)	8	(ア), (オ), (カ)
3	(イ), (エ), (カ)	9	(ア), (ウ), (オ), (カ)
4	(イ), (キ)	10	(ア), (ウ), (キ)
5	(エ), (キ)	11	(ア), (オ), (キ)
6	(イ), (エ), (キ)	12	(ア), (ウ), (オ), (キ)

表-4 No. 12 のクラスタリング結果

クラスタ	ID数	平均滞在日数	平均滞在開始時間割合					平均平休祝日割合	平均滞在時間(分)			
			8時台	9-11時台	12-13時台	14-15時台	16-22時台			その他		
クラスタ1	7609	1.18	3.3%	27.1%	25.1%	16.5%	17.5%	2.2%	44.2%	27.0%	26.6%	20.46
クラスタ2	3547	1.26	3.7%	35.3%	27.3%	9.5%	10.3%	2.4%	47.6%	25.8%	23.8%	99.95
クラスタ3	1298	1.45	20.1%	33.6%	10.6%	4.0%	5.5%	8.6%	41.3%	25.4%	27.9%	314.07
クラスタ4	842	5.5	15.3%	31.3%	16.6%	10.7%	22.6%	3.5%	61.7%	17.2%	21.1%	164.45
クラスタ5	469	11.61	26.6%	27.6%	12.5%	9.2%	19.2%	4.8%	68.7%	13.1%	18.2%	252.02
クラスタ6	442	19.33	44.9%	25.8%	8.8%	5.6%	10.6%	4.3%	73.6%	10.5%	15.9%	383.77
クラスタ7	203	29.31	34.6%	33.1%	7.7%	5.0%	8.6%	11.0%	58.1%	19.8%	22.1%	510.78

表-5 集計用データベース例

ID	観測日	初観測時間帯	最終観測時間帯	各時間帯の観測の有無						
				3時台	4時台	5時台	~	21時台	22時台	
ab:cd:ef:12:34:56	12月1日	8	12	0	0	0	~	0	0	
ab:cd:ef:12:34:56	12月2日	3	4	1	1	0	~	0	0	
ab:cd:ef:12:34:56	12月3日	19	22	0	0	0	~	1	1	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

-5の各行の時間帯別観測の有無で観測した時間帯を集計している。

図-3では平日の観測ID数が多い一方、休日や28日以降の年末年始のID数が極端に少ないことが確認できる。図-4ではほとんどのIDが8時台に観測を開始している。また、多くのIDが8-17時台に渡って観測されているが、18時台以降から次第に減少している。これらの特徴からblueには範囲内で働く店舗従業員のスマートフォンやOA機器が多く該当している可能性が高い。

b) クラスタ2の特徴

図-5はクラスタ2の日別ID数、図-6はクラスタ2の時間帯別の観測ID数・開始時間帯ID数、終了時間帯ID数をそれぞれ示す。

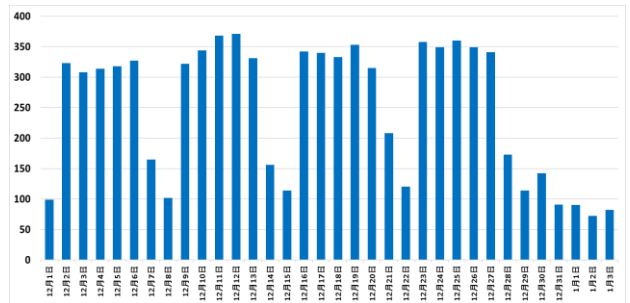


図-3 No. 12 クラスタ 6 の日別観測 ID 数

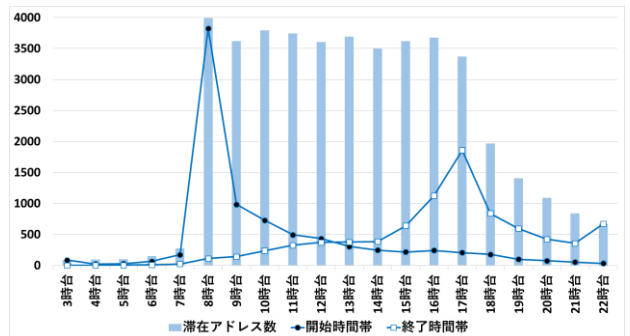


図-4 No. 12 クラスタ 6 の時間帯別 ID 数

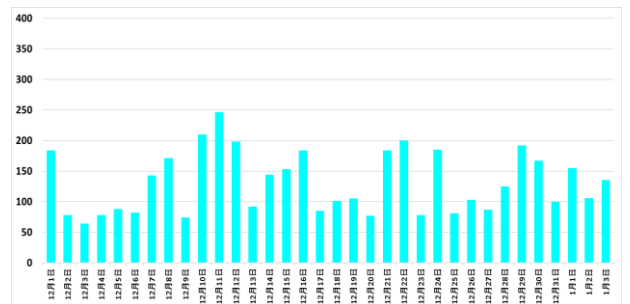


図-5 No. 12 クラスタ 2 の日別観測 ID 数

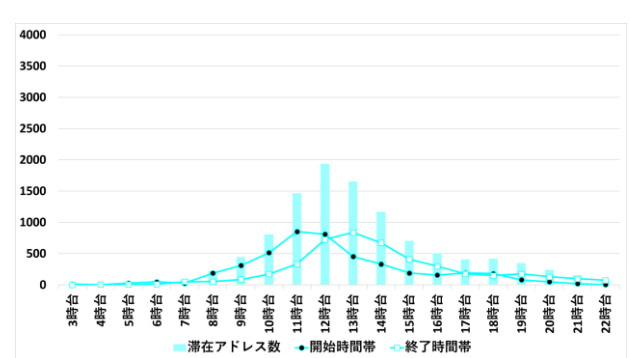


図-6 No. 12 クラスタ 2 の時間帯別 ID 数

帯ID数を示している。図-5、図-6の各集計はa)と同様の方法で集計した。図-5より、クラスタ2は基本的に休日やイベント日に観測されるIDが多い。また、図-6より11時から14時台の昼食時に観測されるIDが多いという特徴がある。これらの特徴より、クラスタ2は昼食目的の来訪者のIDである可能性が高いと考えられる。

また、これらの特徴は堀籠ら³⁾の研究で抽出した観光客のデータと概ね合致していたことより、観光客のIDが含まれていると考えられる。しかし現地の従業員によると観光客も多く来訪するが、地域住民や女川周辺の労働者も昼食目的で来訪するという。以上より、No-12クラスタ2は「昼食目的」の方が所持する通信端末が該当していると考えられる。しかし、その中でも観光客のIDのみならず、地域住民や労働者のIDも一定数存在している。

c) その他クラスタの特徴

blackは平均滞在日数・平均滞在時間がいずれも最も高く、日別観測ID数がほぼ一律であることから常時稼働しているOA機器が多いと断定できる。続いてクラスタ5の日別ID数は平日が多く、休日・年末年始が少ない。時間帯別ID数は8-17時台に多く観測され、18時台以降減少するなど、クラスタ6の特徴に酷似している。しかし、平均観測日数や平均滞在時間がクラスタ6と比較が少ないことから、アルバイトの従業員などのカテゴリが考えられる。クラスタ1は平均滞在時間が約20分と極端に短いクラスタであり、日別ID数では休日やイベント、年末年始の来訪が多いデータであった。滞在時間の長さより通過交通や卸業者、短時間滞在する地域住民、その他クリーニングしきれないデータなど色々なカテゴリが混在しているクラスタであると考えられる。クラスタ3は日別・時間帯別ID数から顕著な特徴は得られなかった。また、平均観測日数が少ない一方、平均滞在時間が314分と長い。以上より、観光客の中でも長時間滞在する方やイベント運営会社の従業員などに該当する可能性がある。最後にクラスタ4は日別・時間帯別ID数で際立った特徴は無く、観測日数は5日程度・滞在時間は2.5時間程度であった。観測日数と滞在時間より、週1回程度訪れる地域住民の可能性はある。

4. まとめ

(1) 本研究の成果

本研究では観測したPRデータを用いたクラスタリングを実施し、クラスタリング結果の意味づけを行った。「OA機器」や「従業員の所持するスマートフォン」等の滞在時間や滞在日数、滞在曜日等が固定化されやすいと考えられるカテゴリのIDは分類や意味づけが可能であった。また、昼食目的と考えられるクラスタでは、同じように分類や意味づけが可能であった。しかし、このクラスタ内には「観光客」「地域住民」「地域企業の従業員」などのカテ

ゴリが混在している。1ヶ月程度のデータでは「昼食目的」に混在している各カテゴリの特徴が酷似しており、これ以上の細分化は困難であった。

(2) 今後の展開

まず、データベースとしての活用を行う。現状の約1ヶ月分のデータでは「OA機器」や「従業員」等の特徴的なデータの分類は可能であった。その一方、「昼食目的」の中には「観光客」や「地域住民」「」複数のカテゴリが混在しており、このままではこれらのカテゴリを分類していくことはできない。そこで、今後も継続してデータを取得し、データベースとして運用する。そうすることで、例えば月に1度のペースで来訪していた地域住民なども複数日で観測されるようになるため、ここからさらに正確な観光客データの分類が可能になると考えられる。

また、クラスタリングと地点データの組み合わせによって詳細な分類が可能になると考えられる。今回はクラスタ2が滞在時間帯等の観点から昼食目的のクラスタと意味づけしたが、昼食以外の目的で来訪していたIDも存在している可能性がある。例えば、飲食店の範囲のみ限定的に取得するセンサを用い、クラスタリング結果と紐付けることで、詳細なデータ分類が可能になると思われる。

謝辞：本研究を進めるにあたり、多大なるご協力をいただいた、中央復建コンサルタンツ株式会社の末様と西村様、女川みらい創造株式会社の今野様、女川町商工会の茂木様、女川温泉ゆぼぼの吉田様、そしてセンサの設置にご協力いただいたシーパルピア女川の関係者の皆様に深謝の意を表します。

参考文献

- 1) 一井啓介, 寺部慎太郎, 柳沼秀樹, 康楠, 田中皓介: Wi-Fi パケットセンサを用いた散策型観光地における観光回遊行動の把握, 土木計画学研究発表会・講演集, Vol.57, 2018.
- 2) 壇辻貴生, 杉下佳辰, 福田大輔, 浅野光行: Wi-Fi パケットセンサを用いた観光客の滞在時間特性把握の可能性に関する研究-奈良県長谷寺における試み-, 公益社団法人日本都市計画学会都市計画論文集, Vol.52, No3, 2017.
- 3) 堀籠涼太, 高橋祥智, 中林果歩, 菊池輝, 泊尚志, 末祐介, 西村洋紀: 狭域における Wi-Fi パケットセンシング-女川シーパルピアを例に-, 土木計画学研究発表会・講演集, Vol.59, 2019.
- 4) 小林巴奈, 福田大輔, 中西航, 内田賢悦, 浅田拓海, 有村幹治, 菅芳樹: Wi-Fi パケットセンシングデータを用いた広域観光周遊行動のモデル分析, 土木計画学研究発表会・講演集, Vol.58, 2018.
- 5) 五百蔵夏穂, 田中謙大, 中西航, 神谷大介, 菅芳樹, 福田大輔: Wi-Fi パケットセンシングによる沖縄本島観光周遊パターンの特徴抽出, 土木計画学研究発表会・講演集, Vol.57, 2018.
- 6) tshark, <https://www.wireshark.org/>, 最終閲覧日: 2020/03/4.

(?受付)