

大規模公園環境における WiFi パケットセンサデータの利活用に関する分析と課題

村井 大地¹ 浦野 健太¹ 青木 俊介² 米澤 拓郎¹ 河口 信夫^{1,2}

概要: 地下鉄やショッピングモールなどの公共エリアにおいて, 利用者の行動パターンの把握は, 混雑度の緩和やマーケティングの提案において重要である. 一方, Wi-Fi パケットセンサを用いた人流分析が登場してきた. Wi-Fi パケットセンサ 1 つで混雑度推定や人流推定を行うことができる. しかし, プライバシーの観点から, MAC アドレスをランダム化する機能を搭載したスマートフォンが増えており, 人流分析や混雑度推定において影響されることが考えられる. 我々は 59 ヘクタールの広大な敷地を有する愛知県名古屋市の東山動植物園に, 35 箇所の Wi-Fi パケットセンサを取り付け, 継続的なデータの収集を行っている. 本研究ではランダムアドレスのデータ, ユニークアドレスのデータそれぞれにおいて, 東山動植物園という実環境において有用性を検証する.

Analysis and Issues on Utilization of WiFi Packet Sensor Data in Large-Scale Park Environment

DAICHI MURAI¹ KENTA URANO¹ SHUNSUKE AOKI² TAKURO YONEZAWA¹
NOBUO KAWAGUCHI^{1,2}

1. はじめに

近年, COVID19 の影響により, リモートワークや長期休暇における外出の自粛など, 感染のリスクを避けるようになっている. そのような人々の行動の変化に伴い, ショッピングモールやレジャー施設, 駅などの公共エリアにおいて, 人々の行動の把握はより重要になっている. なかでも, 混雑度の推定や人流推定は, 密集を避けるような対策や新たなマーケティングの提案に繋がる.

人々の分布状況や行動を把握するため, これまでに様々な手法が提案されてきている. Global Positioning System (GPS) による手法では, 携帯電話から得られる位置情報から, 人々の行動パターンを分析し, 属性の推定や混雑推定などに利用されている [1][2]. しかし, 屋内施設での位置推定は困難であり, 特に商業施設や地下鉄などでの利用は難しい. 固定カメラによる手法では, 画像処理技術を用い

ることで, 人の移動軌跡から, どのような行動を起こしたかがわかる [3][4]. しかし, カメラ一台で捉えることができる範囲は限定的であり, 広範囲を対象とする場合, 複数台のカメラを設置しなければならず, 導入コストが高くなってしまふ. また障害物による死角が生じてしまうため, 設置場所や角度などの調節が難しいといった課題が生じる.

一方, Wi-Fi 機能を持つスマートフォンの普及に伴い, スマートフォンから発せられる信号を「人」とみなし, 人流を計測する Wi-Fi パケットセンサを用いた手法が登場してきた [5][6][7]. スマートフォンなどの端末は Wi-Fi ルータとの接続を行うために, 数秒から数分程度の間隔でプローブ要求をブロードキャストしている. Wi-Fi パケットセンサを任意の環境に設置することで, プローブ要求を収集し, そのフレームに含まれる MAC アドレスをキーとすることで混雑度推定と端末ごとの人流解析の両方を行うことができる. そのため, 店内や地下鉄といった屋内施設やレジャー施設や都市といった屋外施設など, 周りの環境に依存せず利用できる. しかし, 近年ではプライバシーの観点から, MAC アドレスをランダム化する機能を搭載

¹ 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

² 名古屋大学社会創造機構
Institutes of Innovation for Future Society, Nagoya University

したスマートフォンが登場してきている [11]. このランダム化機能により, 人流分析や混雑度推定をする際に, 大きな影響があると考えられる. 本研究では, 愛知県名古屋市が保有する東山動植物を対象に, 分析を行う. 東山動植物園における入園者の行動を分析する際, 追跡ができるユニークアドレスを用いて入園者の推定を行う必要がある. しかし, ランダム化機能を搭載した端末の増加により, ユニークアドレスのデータ数が減少すると考えられる. そこで, 実際の入園者数データと2日分の入退場者数のデータを用いて, ランダムアドレスのデータとユニークのデータそれぞれの有効性を検証した. またランダム化されていないアドレス(以降ユニークアドレスと呼ぶ)の利活用に関する分析を行った. その結果, ユニークのアドレスでは実データと比べて, 少ないデータ数となった.

2. 関連研究

Wi-Fi パケットセンサを用いた研究は多く行われている. Cristian らは, オランダの都市, アッセンにおいて, TT フェスティバル期間での人々の移動軌跡データを推定した. 用いたデータは追跡が可能なユニークアドレスのみを用いている [9]. Hande hon らは, 美術館内における来場者の館内回遊ルート推定を行った [10]. この研究ではまず, ユニークデータの移動遷移をもとに, マルコフモデルを作成し, その遷移確率をもとにランダムアドレスに適用することで, ランダム化機能を持ち合わせている端末の追跡を行った. Fang-Jing Wu らは, ウェリントン駅(屋内)とショッピングモール(屋外)における人数推定を行った [8]. カメラデータによる人数と Wi-Fi データにおけるデータの乖離に対し, 動的線形回帰を行った. 使用したデータはランダムアドレスを含む全てのアドレスデータを使用している. このように, 様々な環境で Wi-Fi パケットセンサを用いた検証が行われており, 環境により扱うデータが異なる. そこで我々は屋内や屋外に設置したセンサに対し, ランダムアドレスとユニークアドレスを実際の入園者データと入退場データをもとに比較し, 有用性を検証する.

3. センサ概要

3.1 Wi-Fi パケットセンサ

本研究では, 株式会社システム総合研究所の Wi-Fi パケットセンサ(型番: DPS-ODJP)を用いる. 同センサはスマートフォンなどの Wi-Fi 機器から送信されるプローブ要求を匿名性を高めた上で収集し, サーバへアップロードする仕組みを有する. プローブ要求には機器固有の MAC アドレスが含まれているが, 同センサは匿名性を確保するために MAC アドレスにハッシュ関数による変換処理を行っている. この処理を施したアドレスを AMAC(Anonymous MAC) アドレスと呼び, 匿名化の処理をした上で表 1 に示すデータを保存する. ハッシュ関数は一定期間でシードを

表 1: センサデータ

データ概要	備考
TIMESTAMP	プローブ要求の取得時間
AMAC	ハッシュ化した MAC アドレス
OUI	製品ベンダーコード
SC	シーケンス番号
RSSI	受信強度

変更するため, 長期に渡り継続した同一ユーザ認識は不可能であるが, AMAC アドレスをキーとすることで, 少なくとも1日単位で個々のユーザ端末を識別できるため, 人の流動などの把握することが可能となっている.

3.2 対象環境

本研究では, 東山動植物園を対象とする. 東山動植物園は名古屋市が所有・管理し, 59 ヘクタールの敷地面積を有し年間 250 万人近くが来園する日本最大級の動植物園である. 園内は, 動物園本園, 北側動物園, こども動物園, 植物園, 遊園地, スカイトワーなどのゾーンに区切られ, スカイトワー以外は一旦入場料を支払えば, 回遊が可能である. また, 園内にはお土産屋やレストラン施設などが複数地点に存在する. 動物園本園, 北側動物園, こども動物園, 植物園は一般道路によって区切られている構造になっており, それぞれ高架橋により一般道路を横切ることができる. 同園は, その敷地面積の広さから6箇所の出入り口ゲートが存在し, 入園者はどのゲートからも出入りが可能となっている. また, 同園に訪れるには自家用車の他, 西側の正門に近い地下鉄東山公園駅と, 北東の星ヶ丘門に近い地下鉄星ヶ丘駅を利用できる. 複数の有力な交通手段, 出入り口が存在する一方, その広さから入園者の園内での行動は把握しづらく, 行動パターンを抽出し, 園内運営の効率化が求められている.

我々は, 東山動植物園に, 2019 年 11 月初旬に Wi-Fi パケットセンサを 35 箇所設置した(図 1). データ収集は 2019 年 11 月に開始し, 2021 年 g 月現在で, 一年半分のデータが溜まっている. また Wi-Fi パケットセンサを設置するにあたり, センサ自体にデータ収集中のステッカー(図 2)を貼り付けるとともに, 収集データに関する説明やプライバシーポリシー, オプトアウトのための問い合わせ先などが記載されたホームページへのリンクを掲載している. なお, 設置にあたっては名古屋大学未来社会創造機構の倫理委員会の審査を得た.

4. MAC アドレス

Wi-Fi 機能を有するスマートフォンはプローブ要求を, 数秒から数分程度の間隔でブロードキャストしている. この送信間隔は, 端末に依存している. 図 3 は, ある 1 日に収集したデータから各アドレスの観測間を表したものである.



図 1: 東山動植物園におけるセンサ設置場所

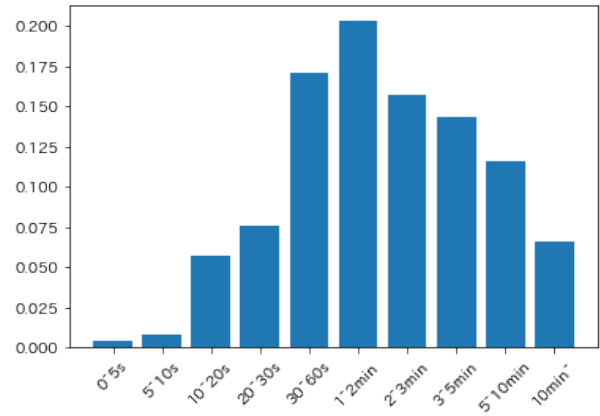


図 3: プローブ要求の送信間隔



図2: 観測機の設置箇所周辺に設置するステッカー

図 2: センサに貼付されたステッカー

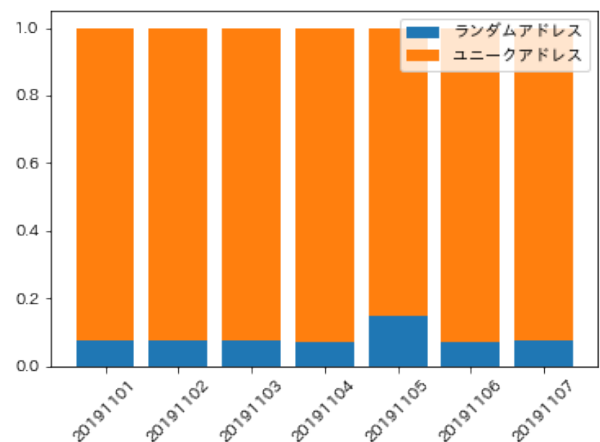


図 4: ランダムアドレスとユニークアドレスの比率

多くの端末はプローブ要求を送信してから、次の送信を行うまで1分以上かかることがわかる。

4.1 ユニークアドレス

スマートフォンなどの機器には固有の MAC アドレスが存在する。ランダム化機能がない端末では、プローブ要求を送信する際、MAC アドレスに対し定期的に変換処理を行わないため、AMAC アドレスと TIMESTAMP をもとに、端末を保持しているユーザがどのように移動しているかがわかる。

4.2 ランダムアドレス

機器固有の MAC アドレスでの追跡によるプライバシーの侵害を避けるために、Apple では iOS8 以降に、Google の Android 端末ではバージョン 5.0 から、MAC アドレスのランダム化機能を搭載し始めた [12]。このランダム化機能により、定期的に MAC アドレスをランダム化するため、異なるタイミングで得られたレコード間での移動履歴の照合が困難になる。

4.3 各アドレスの割合

図 4 は、2019 年 11 月 1 日から 7 日までの一週間におけるランダムアドレスのデータ数とユニークアドレスのデータ数の割合である。一週間を通して、ランダムアドレスのデータは全体の 9 割近くを占めており、トラッキングが可能なユニークアドレスのデータは 1 割前後しかない。

5. 分析

Wi-Fi パケットセンサを用いて、東山動植物園における入場者の行動を分析する際、入園者のアドレスを抽出しなければならない。しかし、入園者の推定を行うにはある程度追跡が可能であるユニークなアドレスを用いる必要があるが、前章で述べたように、ユニークなアドレスのデータは Wi-Fi パケットセンサのデータの 1 割しか満たない。そのため、ユニークのアドレスのみを用いる場合、データ量として不十分な可能性がある。そこで、本章では実際の入園者数のデータと入退場データを用いて、ランダムアドレス、ユニークアドレス、そしてユニークアドレスをもとに推定した入園者のアドレスが分析において有用性があるかどうかを検証していく。

5.1 入園者アドレスの抽出

プローブ要求の送信範囲は、端末を中心に放射線状に数百 m まで届く。園内には一般道路が存在しており、また近隣に住宅や商業施設もあることから、入場者以外からのプローブ要求も多く受信してしまう。入園者の回遊パターンや混雑度を推定するには園外からのデータはノイズになるため、入園者を判別するフィルターが必要になる。本研究では、ユニークアドレスから、滞在時間が 30 分以上、かつ複数のセンサで観測されたアドレスを入園者と判断した。

5.2 入園者データ

今回用いる実際の入園者数のデータの期間は、2019 年 11 月 1 日から 2020 年 12 月 31 日である。園内に設置してある全てのセンサで観測したアドレスから、ランダムアドレス数とユニークアドレス数を算出した。また評価には、時系列の相関を用いた。

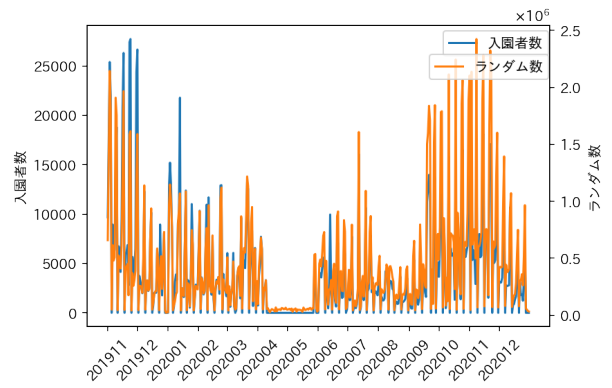
5.2.1 結果

図 5 に各アドレスのデータと入園者数のデータを示す。表 2 に、各月の金土日の 3 日間のデータを示す。図 5a のように、ランダムアドレスでは、入園者数と比べ 100 倍近い値が見られた。それに対し、図 5b の推定した入園者数では入園者数の 10 分の 1 であった。それぞれのデータと入園者数の時系列データの相関は、ランダムアドレスと推定入園者数で 0.9 以上と高い相関が見られた。ユニークのデータでは 0.6 とその他のアドレスと比べて低い相関になった。

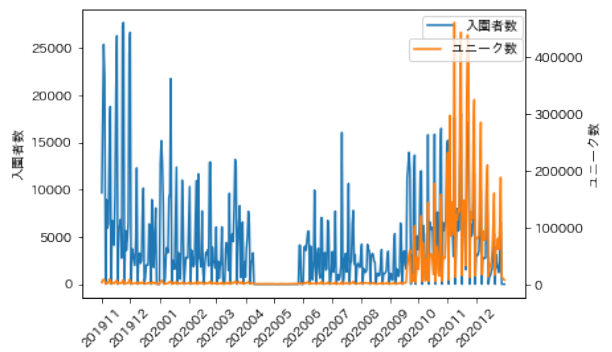
5.2.2 考察

ランダムアドレスと推定した入園者数ともに、実際の入園者数データとの相関が 0.9 以上と高く、入園者数の傾向を捉えることができている。ランダムアドレスの場合、ランダム化機能の影響から 1 台の端末が複数台としてカウントされてしまう。そのため入園者数と比べて、過剰に判定してしまう。一方で、推定した入園者数は実際の入園者数の 10 分の 1 の人数しか抽出できていない。例えば、表 2 の 2020 年 6 月 19 日のように、入園者数が 634 人に対し、56 人のアドレスしか抽出できていない。東山動植物園のようなレジャー施設では、イベントなどを、その日の入園者の行動の傾向を調べるのに 56 人では把握することが難しいと考えられる。以上のことから、入園者数といった長い年月で比べる場合はユニークのアドレスだけでも有効であるが、イベントを行った日など、特定の日にける入園者の動向を分析するにはユニークのアドレスのみでは読み取れない恐れがある。そのため、データ数を増やすためにランダムアドレスを用いた手法を検討する必要がある。

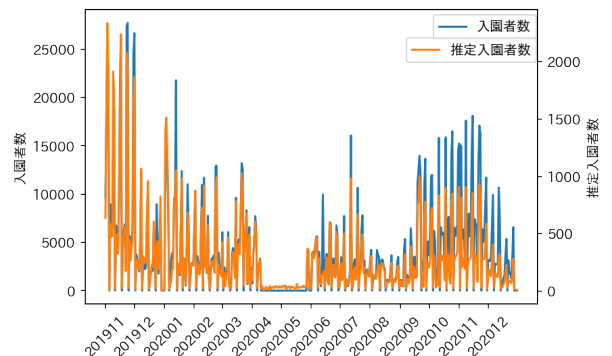
ユニークアドレスでは、2020 年 9 月以降、アドレス数が増大している。図 6 では、アドレスが増加する以前と以降において、各アドレスが何回観測されたかを示している。図 6b では、1 回しか観測されていないアドレスの数が異常に増えている。考えられる原因として、東山動植物園



(a) 入園者数とランダムアドレス数



(b) 入園者数とユニークアドレス数



(c) 入園者数と推定入園者数

図 5: 各アドレスと入園者数データとの比較

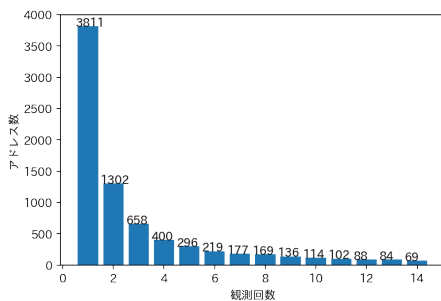
の交通量の増加と OS のアップデートによるランダム化の変更である。交通量の増加に関しては、COVID19 の影響で外出が減少していたため、交通量が増えたことは考えにくい。OS のアップデートによる変更は、ユニークのアドレスもランダム化処理がされている可能性が考えられる。

5.3 入退場カウントデータ

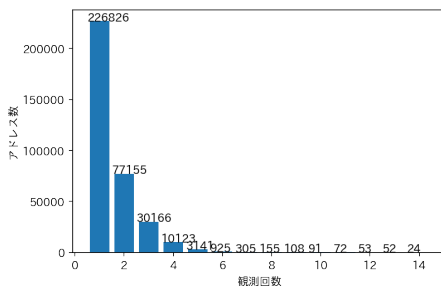
我々は 2021 年 3 月 26 日、27 日の 2 日間、正門と星ヶ丘門の 2 箇所のゲートで 30 分おきの入退場者をカウントした。(図 7) 正門は東山動植物園におけるメインゲートであり、時間帯によっては 600 人以上の人がゲートを通過する。また、ゲートは屋外にあり、正門付近のセンサも屋外に設

表 2: 週末における実際の入園者数と各アドレス観測数, 推定入園者数の比較

	20191101	20191102	20191103	20191206	20191207	20191208	20200110	20200111	20200112
日計入園者数	9678	17365	25392	1980	4845	12304	3324	9134	9622
ランダムアドレス	658141	423416	2144357	195764	504823	1139187	202468	803239	921579
ユニークアドレス	4252	8729	9686	2561	3544	4635	2502	4928	4650
推定入園者	638	1619	2338	186	479	1066	195	805	840
	20200207	20200208	20200209	20200306	20200307	20200308	20200403	20200404	20200405
日計入園者数	2113	6714	10955	2441	6061	626	5002	7701	6464
ランダムアドレス	213879	599582	764024	242441	470286	81454	541022	680182	575886
ユニークアドレス	2379	3998	4447	3031	4386	1667	4195	5151	3710
推定入園者	163	520	726	173	468	49	458	605	484
	20200619	20200620	20200621	20200703	20200704	20200705	20200807	20200808	20200809
日計入園者数	634	7038	7031	1274	2257	7726	1450	4200	3723
ランダムアドレス	108057	867980	910382	153501	285718	794589	221011	321323	444634
ユニークアドレス	2049	5133	5053	2057	2959	4690	3432	3428	3154
推定入園者数	56	601	583	94	174	503	106	161	225
	20200904	20200905	20200906	20201002	20201003	20201004	20201106	20201107	20201108
日計入園者数	432	4214	5348	6338	11853	11990	5775	2921	17597
ランダムアドレス	91908	519915	652687	772423	1783115	1790081	754567	498349	2423566
ユニークアドレス	2377	4090	2965	53641	108936	121439	145855	88918	460899
推定入園者数	47	223	322	288	686	726	250	192	906
	20201204	20201205	20201206						
日計入園者数	3550	6819	9926						
ランダムアドレス	422487	981793	1393407						
ユニークアドレス	87106	200197	285532						
推定入園者数	127	305	408						



(a) 2019/11/16 における入園者数と推定入園者数



(b) 2020/11/21 における入園者数と推定入園者数

図 6: ユニークアドレスのカウント回数の分布

置したある。星ヶ丘門は東山動植物園の植物園エリアにあるゲートであり、ゲートを通過する人は多くて 200 人であ

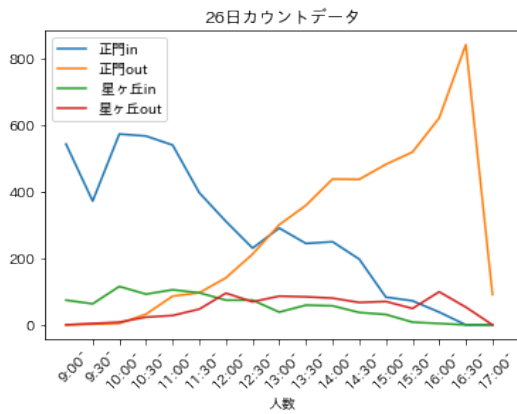
る。ゲートは屋内にあり、星ヶ丘門に設置したセンサも屋内にある。今回は、ゲートに設置したセンサから得られたランダムアドレスのデータと、推定した入園者データのうち、正門と星ヶ丘門で入退場したアドレスを用いる。ただし、ランダムアドレスはデータのスケールを合わせるために、受信強度 (RSSI) が-65 以上のデータを用いた。

5.3.1 結果

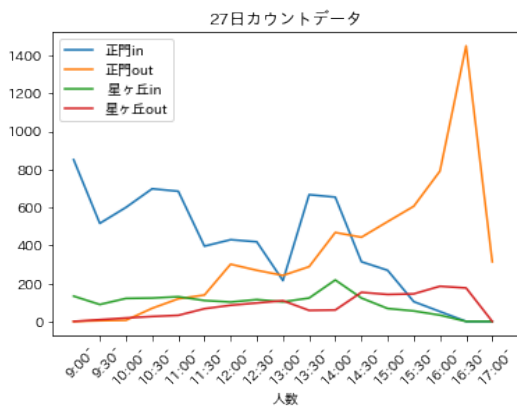
図 8 は、時間帯ごとの入退場カウントデータとランダムアドレスの観測数を表している。正門では入場者の傾向が取れている、しかし、閉園時間における退場者数はうまく捉えることができていないことがわかる。図 9 は、時間帯ごとの推定した入退場者数を表している。入園者数と退場者数の大まかな傾向は捉えられているが、実際の入退場者数よりも 10 分の 1 のデータ数であった。

5.3.2 考察

図 8a より、ランダムアドレスの場合、退場者の傾向が取れていない原因として、来場者の移動速度とプローブ要求の送信間隔にあると考えられる。東山動植物園では、入園する際に、職員にチケットを見せる必要がある。また、パンフレットやマップを見るために立ち止まる人も多い。そのため、ゲート付近にいる時間が長い。対して、退園時では、入園時のように立ち止まる人が少なく、一直線に帰る



(a) 26日のデータ



(b) 27日のデータ

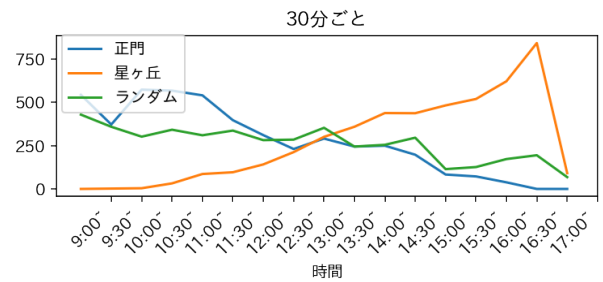
図 7: 30分おきの入退場カウント

人が大多数である。プローブ要求は送信間隔が数分以上ある端末が多いため、センサ付近にいる時間が短いほど観測される可能性が少なくなる。

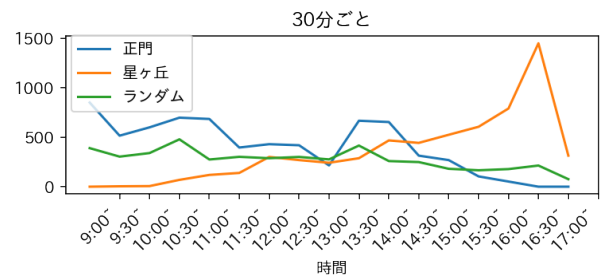
推定した入退場者数では、しかし、実際の入退場者数と比べ、データ数が少ない。特に星ヶ丘門では時間帯によっては0人と推定されてしまう。以上から、正門のような利用者が多い箇所では、ユニークなアドレスでも一定のデータ数は得ることができる。しかし、利用者の移動速度によっては得られるデータ数が異なるため、利用者の特徴を含めていく必要がある。また、星ヶ丘門のように、利用する人数が少ない場合、推定したデータだけでは、データが極端に少なくなってしまう可能性がある。そのため、ランダムアドレスのデータを考慮する必要がある。

6. まとめ

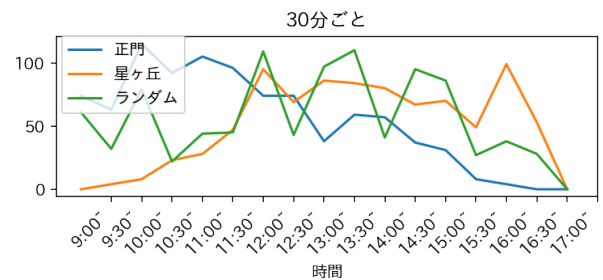
本項では東山動植物園における Wi-Fi パケットセンサデータと実データを比較を行った。ランダムアドレス機能により今後ランダムアドレスが増えていくと考えられる。大都市や観光地帯における人数の分布ではユニークアドレスのみでも十分なデータが得られるが、今回対象としたレジャー施設のように利用者を判別が必要な場合やレストラン、売店などの局所的な分析においては、データ数が不足してし



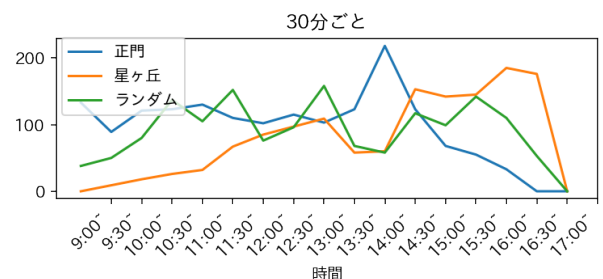
(a) 26日正門



(b) 27日正門



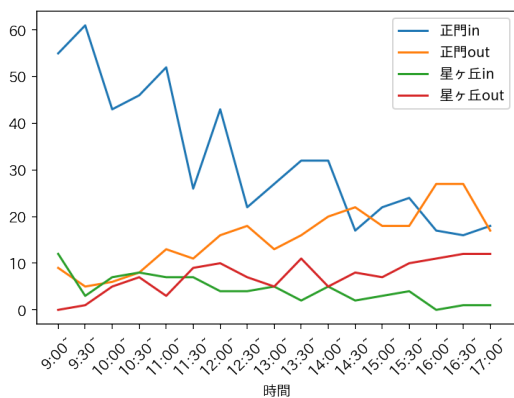
(c) 26日星ヶ丘門



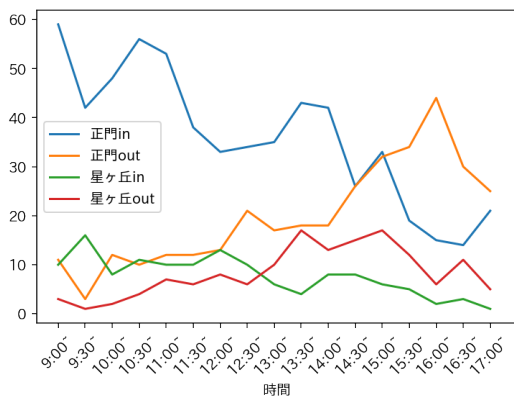
(d) 27日星ヶ丘門

図 8: 入退場カウントとランダムアドレスの比較

まう可能性があり、それを防ぐにはランダムアドレスを用いた検討を行う必要がある。このように、センサの周りの環境により扱うデータを取捨選択をしなければならない。また、対象となる利用者の行動も考慮する必要がある。プローブ要求の送信間隔から、センサの受信範囲内においても観測されない可能性がある。そのため、ある時間における人数推定を行う際の過少評価や、対象とするアドレスの追跡を行う際、間欠な移動軌跡データが得られる。そのため、移動速度や属性などの特徴をもとにデータを分析していく必要がある。



(a) 26 日における推定した入退場者データ



(b) 26 日における推定した入退場者データ

図 9: 30 分ごとの推定した入退場者数

らの行動分析のための同一端末推定手法” 情報処理学会第 80 回全国大会, 6S-04

- [7] A.B.M. Musa, Jakob Eriksson, ”Tracking Unmodified Smartphones Using Wi-Fi Monitors”, *SenSys'12: Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems* November 2012 Pages 281–294
- [8] Fang-Jing Wu, Gurkan Solmaz, ”CrowdEstimator: Approximating Crowd Sizes with Multi-modal Data for Internet-of-Things Services”, *MobiSys'18*, 2018
- [9] Cristian Chilipirea, Andreea-Cristina Petre, Ciprian Dobre, ”Presumably simple: monitoring crowds using WiFi”, 2016 17th IEEE
- [10] Hande Hong, Girisha Durrel De Silva, Mun Choon Chan, ”CrowdProbe: Non-invasive Crowd Monitoring with WiFi Probe”
- [11] K. Skinner and J. Novak. 2015. Privacy and your app.
- [12] Dan Goodin. 2017. Shielding MAC addresses from stalkers is hard and Android fails miserably at it. <https://arstechnica.com/information-technology/2017/03/shielding-mac-addresses-from-stalkers-is-hard-android-is-failing-miserably/>. [Online].
- [15]

謝辞

本研究の一部は、総務省委託研究 SCOPE (0159-0110)、NICT 委託研究 (21406)、JST CREST (JPMJCR1882) の支援を受けています

参考文献

- [1] K. Nishida, H. Toda, T. Kurashima, and Y. Suhara, ”Probabilistic Identification of Visited Point-of-Interest for Personalized Automatic Check-in,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 631–642. ACM, 2014.
- [2] Junchuan Fan, Cheng Fu, Kathleen Stewart, Lei Zhang, ”Using big GPS trajectory data analytics for vehicle miles traveled estimation”, 2021 Elsevier B.V.
- [3] 宮城 優里, 大西 正輝, 渡辺知恵美, 伊藤 貴之, ”文字列かによる人流データの圧縮と可視化”, *DEIM Forum 2015* F8-5
- [4] Kang Hao Cheong, Sandra Poeschmann, Joel Weijia Lai, Jin Ming Koh, U.Rajendra Acharya, Simon Ching Man Yu, Kenneth Jian Wei Tang, ”Practical Automated Video Analytics for Crowd Monitoring and Counting”, *IEEE Access* 7:183252-183261
- [5] Kai Li, Chau Yuen, Salil S. Kanhere, Kun Hu, Wei Zhang, Fan Jiang, Xiang Liu, ”An Experimental Study for Tracking Crowd in Smart Cities”, *Journal Paper CISTER-TR-181201*
- [6] 中田 恵史, 岡本 東, 堀川 三好, ”Wi-Fi プローブ要求か