

Wi-Fiプローブパケットと目視データを用いた 公園内の訪問者流動推定手法と効果的なセンサ配置の検討

Effective Arrangement of Wi-Fi Sensors and Method for Population and Flow Estimation in a Park

玉田 拓[†] 豊木 博泰[‡] 渡辺 喜道[‡]
Taku Tamada Hiroyasu Toyoki Yoshimichi Watanabe

1. はじめに

スマートフォン等の携帯端末の普及に伴い、携帯端末に備わっている Wi-Fi 機能を利用して、端末から発信されるプローブパケットを取得し、分析することで人の流動を解析する研究は多く行われている([1]-[3]など)。一方で、NTT docomo の「モバイル空間統計」のような大手携帯電話通信各社によるネットワークを利用した大規模な人流解析が行われている。最近の COVID-19 禍への社会的対応に貢献するため、基地局のエリア付近の携帯端末情報を取得し、基地局付近の人の量のデータとして 1 時間ごとに提供する無料サービスが行われ広く知られるようになった。これらのサービスは、相対的な人の流れや人の量の変動を把握することができ、観光客の流動や街づくり、防災計画にも役立つものとして非常に有用性があるが、予算面での難しさや、基地局のないエリアや地方の観光地などの細かいエリアにおける人流解析は難しいという点があるため、安価で任意的な地点の観測を行うことの出来る Wi-Fi パケットセンシングシステムは地方観光業者や中小企業などから多く期待が寄せられている。

Wi-Fi プrobeパケットデータの観測では人流や歩行者数を相対的に把握することは可能だが、実数把握は困難である。また、市街地などでのセンシングにおいては、アクセスポイントや車などからのプローブパケットによる雑音が多く、歩行者や訪問者だけのデータとするためにはデータクリーニングなどが必要である。また近年、プローブパケットにはランダム MAC アドレスを利用する端末が増加しており、地点間移動量の観測が困難になってきた。本研究グループによる甲府市内約 20カ所の継続観測では、地点によってランダムアドレスの割合が異なり、時間変動も大きい。市街地では Wi-Fi アクセスポイントが多く、それらを利用中の端末はプローブにもグローバルアドレスを使用することがあるために変動が大きくなると考えられる。本グループが甲府市中心市街地に設置したセンサから得られたアドレス比と本研究の対象である県立公園でのデータを図 1 に示す。公園では、グローバルアドレスの比率がほぼ 20%で時間変動がほとんどない。アクセスポイントや住

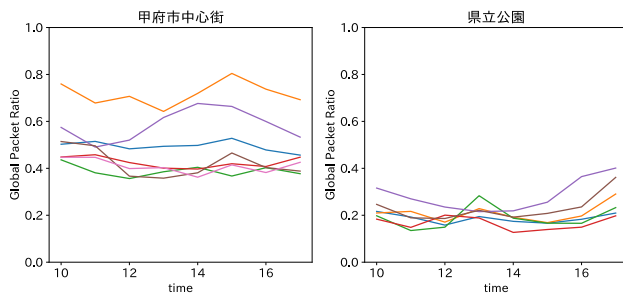


図 1 グローバル/ランダムアドレスの割合

[†] 山梨大学大学院工学専攻電気電子工学コース

[‡] 山梨大学大学院総合研究部工学域,
University of Yamanashi

居・オフィスの固定端末がほとんどないエリアは、訪問者実数計測を行うための好環境であるといえる。本研究はそのような地域における訪問者実数推定の方法を検討する。

管理者からの了解が得られた県立自然公園を研究対象とし、昨年 7 月に 9 個のセンサを設置して常時観測を開始し現在も継続中である。得られたデータは逐次サーバに送信し、日ごとの集計を行って即日 Web 公開するシステムを構築した。そのサービスは関係者から好評を得ている。本研究では、さらに精度の高いデータ提供システム構築を目指し、現地の従業員が毎日行なっている目視による利用者数調査のデータ（以下、目視データ）とセンサで得られた Wi-Fi パケットデータ（以下、Wi-Fi データ）に関する機械学習分析を行う。また、この結果からセンサの最適配置についても考察する。

2. Wi-Fi パケットセンサの概要

Wi-Fi パケットセンサは、安価であり、複製が容易である点から、Raspberry Pi 3 model B(以下、RasPi)を用いた。Wi-Fi パケットを取得するために、RasPi に内蔵されたオンボード Wi-Fi チップを利用し、サーバとのデータ通信を行うために 3G ドングルを USB 端子に接続する。OS は Raspbian GNU/Linux9.4(stretch)に研究室独自の設定プログラムを組み込んだものを使用し、プローブパケットの取得には tcpdump を用いた。図 2 にシステム構成を示す。

RasPi には、取得したプローブパケットを 1 日分蓄積し、前日のデータを深夜にサーバへ scp によって送信し、送信に成功した際に RasPi 内のデータを消去するシステムを構築する。また、サーバにポートフォワーディングを含む ssh 接続を行うことでセンサとサーバ間にセッションを張り、サーバからセンサへの ssh 接続を可能にした。それぞれのセンサにつけた名称からポート番号を一意的に得るルールを設定し、台数が増えても管理を容易にした。以上の設定を用いて、センサの常時監視、プログラムのアップデート等の遠隔操作を行っている。

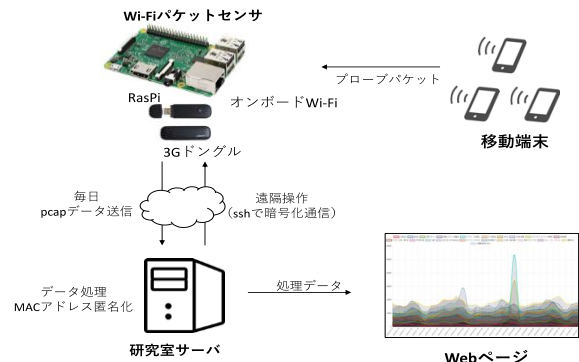


図 2 システム構成 概略図

センサや tcpdump プロセスが動作しているかを cron により定期的に確認し、動作していなければ再起動するようにすることで、24 時間常時計測を可能にしている。同様に、3G 回線の接続についても定期的に確認し、不具合がある場合は、再接続するようにしている。現在、他地域を含め 40 台以上のセンサを設置しているが問題なく管理ができている。

3. サーバ上でのデータ処理

サーバではセンサから送られてきたプローブパケットの情報をバイナリ形式で保存したもの（以下、pcap データ）を定期的に処理し、データ分析や Web ページでデータの可視化を行いやすい形式にしている。データ処理の流れを図 3 に示す。具体的には、pcap データから日毎のデータ、アドレスごとの滞在時間データ、2 地点間の移動データの順に抽出するプログラムを日ごとに定期実行する。

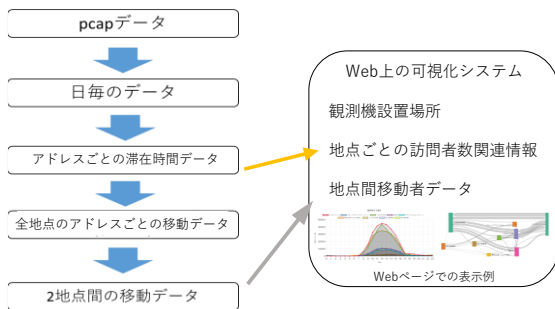


図 3 サーバ上のデータ処理の流れ

3.1 pcap データの処理

pcap データからデータの抽出は、パケット分析アプリケーション Wireshark の tshark で行う。この際、プライバシー保護の観点から MAC アドレスの匿名化を行う。匿名化の例を図 4 に示す。MAC アドレスを SHA256 によってハッシュ化し、ハッシュ値の上位 24byte を切り捨てたものを匿名アドレスとして抽出する。また、匿名化の際に MAC アドレスの種類の情報が失われてしまうため、MAC アドレスの 7bit 目からグローバルアドレスかランダムアドレスかを判別し、アドレスフラグとして抽出する。

これらの処理を通じて、pcap データからプローブパケットの受信時刻、匿名アドレス、受信信号強度、アドレスフラグを抽出し、日毎のデータとして csv 形式のファイルに変換する。

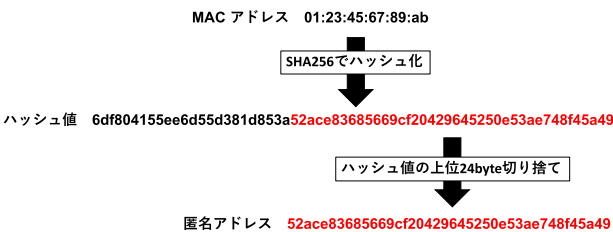


図 4 MAC アドレスの匿名化

3.2 端末の滞在時間データの抽出

日毎のデータから匿名アドレスごとの滞在時間の抽出を行う。滞在時間抽出方法の例を図 5 に示す。匿名アドレス A を初めて取得した時刻を滞在時間開始時刻として、同じ

匿名アドレスが最後に取得した時刻を滞在時間終了時刻として滞在時間データを抽出する。また匿名アドレス B のように同じ匿名アドレスが 10分以上間隔を空けて取得された場合、別の端末として抽出する。

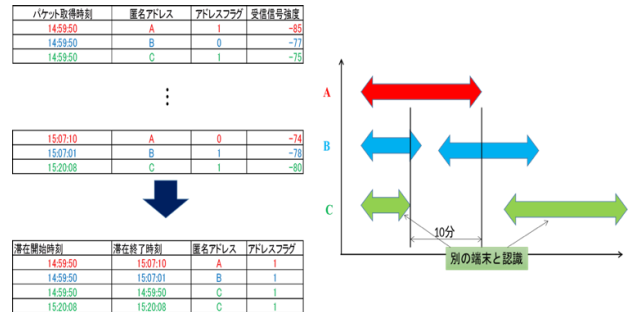


図 5 端末の滞在時間データ抽出手法

3.3 端末の地点間移動データの抽出

先述の滞在時間データにセンサ ID を付け足し、全地点のアドレスごとの滞在時間データをまとめ、そこから地点間移動データを抽出する。その手法のスキームを図 6 に示す。同一の匿名アドレスが 2 地点で観測されているとき、取得時刻の早い地点の滞在時間終了時刻を出発時刻とし、その後取得した地点データの滞在時間開始時刻を到着時刻として、端末の地点間移動のデータを抽出する。



図 6 端末の地点間移動データ抽出手法

4. 公園内の訪問者実数推定

4.1 センサ設置と観測

県立公園の広さは東西約 800 m、南北約 300m であり、公園北部にホテルが建っている以外に人家はなく周辺は果樹園である。中心部に透明板で覆われたドーム状の施設があり、屋外には人工池を備えた広場や遊歩道がある。センサは電源が取れ、公園全体のデータが取得できる配置で設置した。設置箇所を図 7 に示す。上述の処理を行った集計デ



図 7 センサ設置箇所

ータを用いて、日毎のアドレス数の地点別表示や地点間移動を Web ブラウザ上で閲覧できるシステムを作成した。図 8 にその表示例を示す。毎週末にアドレス数が増え、また、8 月の夏休みなどの時期においても、他の月に比べアドレス数が増えていることなど、季節や日毎の傾向が確認できた。同様に、イベントなどがある日にもアドレス数が増えており、これらのデータは現地の従業員の直観と一致している。また、地点間移動図からは、中心部の施設を行き来する数が多いことが確認できる。地点間移動が観測できるのはほぼグローバルアドレスに限られるが、図 1 に示したように、このエリアではグローバルアドレス比率がほぼ一定であるため、移動量の相対値としては正しい傾向を示していると考えられる。

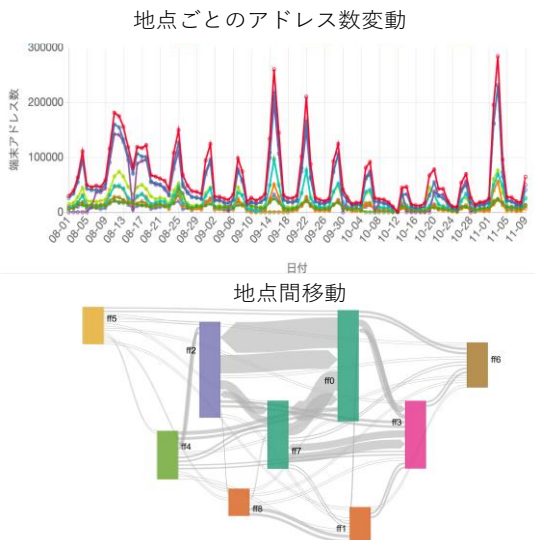


図 8 実際の Web ページ表示例

4.2 センサの検知範囲測定

目視データと Wi-Fi データとの比較を行う上で、センサの検知範囲を把握しておく必要がある。本研究では、グローバルアドレスを含むプローブパケットを常時発信させるプログラムと GPS ドングルにより位置情報を記録する機能を持たせた RasPi を作成し、これを実験地内で持ち歩いて、各センサの検知範囲を測定した。測定の結果、9つのセンサの検知範囲を重ねると公園内全てをカバーしていることがわかった。また、センサ個々の検知範囲は、配置環境によって異なること、最大で約 200m まで検知されること、センサ同士で検知範囲の重なり（以下、オーバーラップ）が起きていることが確認できた。図 9 は、1つの目視地点に対して、センサの検知範囲がオーバーラップしている例である。センサが検知した範囲は、受信強度を用いてヒートマップ



図 9 検知範囲のオーバーラップ

で示してある。このように、単一の端末からのパケットが 2 つ以上のセンサで同時観測されるケースがかなりの割合で含まれる。市街地の回遊行動を同様なシステムで分析する研究では[3]、信号強度から端末の位置を推定する方法を採用しているが、本研究対象地域では推定したい目視地点とセンサの位置が面的に入り組んでいるため、その方法を採用せず、各センサのカウント数を説明変数として、目視地点の滞在者数を推定することとした。

4.3 推定手法の検討

現地の従業員が毎日 11 時と 15 時の 2 回、14 地点で行っている目視データと同じ時間帯にセンサそれぞれで得られたアドレス数（以下、「アドレス数」）を比較することにより推定を行う。以下の分析では 2019 年 7 月から 12 月までのデータを用いた。

目視データとその目視地点に近接するセンサで得られた Wi-Fi データとの比較から単回帰による推定を行った。中心部の施設屋内広場などでは、単一のセンサデータとの相関は良いが、屋内のセンサであっても建物外のパケットも拾っているため、他のセンサデータと組み合わせることによって、よい精度を高められる可能性もある。近隣にセンサがない公園周縁部の訪問者推定には複数のセンサデータを組み合わせることが必須である。

以下では、すべてのセンサデータを説明変数とした、重回帰、サポートベクトルマシン（以下、SVR）、ニューラルネットワーク（以下、NN）の 3 つの推定手法を用いて推定を行い、精度比較を行う。本報告では、重回帰と NN の 2 つについて述べる。推定には、python の統計パッケージ scikit-learn を用いた。

重回帰による推定

すべてのセンサの「アドレス数」を説明変数とし、目視データを目的変数として重回帰による推定を行う。推定値と目視数の比較を図 10 に示す。全体の相関係数は 0.88 となり、地点ごとに相関係数を算出すると、0.39 から 0.90 となり、少しばらつきが見られた。

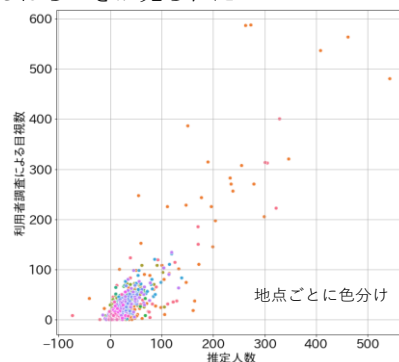


図 10 重回帰による推定人数と目視数

ニューラルネットワークによる推定

NN は、機械学習を用いた推定手法であり、複数の説明変数の互いの影響度を算出し、何度も学習することで推定を行う。影響度を算出するためには、様々なパラメータをユーザーが設定する必要があり、最適なパラメータを選択することで、推定精度を向上させることができる。本研究

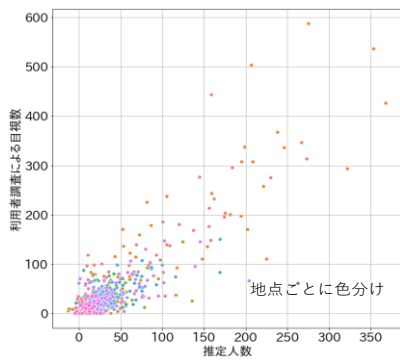


図 11 NN による推定人数と目視数

では、隠れ層を 1~9 層、ニューロン数を 1~200 個と設定し、全ての組み合わせの中から最適パラメータをグリッドサーチにより決定した。結果として、最適パラメータは隠れ層が 8 層、ニューロン数は各層に 100 個を採用した。また、重回帰と同様に、各センサが計測した「アドレス数」それぞれを説明変数とし、各地点の目視データを説明変数として推定を行う。

推定値と目視数の比較を図 11 に示す。全体の相関係数は 0.87 となり、地点ごとに相関係数を算出すると 0.10 から 0.88 となり、かなりばらつきが見られた。

交差検定による推定精度の比較

半年間のデータから 7:3 の割合でトレーニングデータとテストデータをサンプリングし、トレーニングデータで推定モデルを構築し、テストデータで推定精度を検証した。推定値と目視データの相関係数の平均値と平均値の標準偏差を算出したところ、標準偏差はどの地点においても小さく、サンプルによる推定精度の変化が少ないのでそれを用いずに、相関係数の平均値だけを推定精度の評価に用いる。単回帰、重回帰、NN の交差検定による地点ごとの相関係数の平均値を図 12 に示す。

複数のセンサの計測範囲がオーバーラップしている地点では、単回帰よりも重回帰か NN の少なくともどちらかの相関が良くなる。図 13 にセンサのオーバーラップの数を示す。例外は、M、N である。これらの地点は遊歩道上であり、屋内施設や広場に比べて滞在時間が短い。頻繁にアドレスを変更するランダムアドレス端末が多いため、「アドレス数」は端末数だけではなく滞在時間に依存する。加えて、目視が瞬間的な滞在者数であるのに対し、Wi-Fi 「アドレ

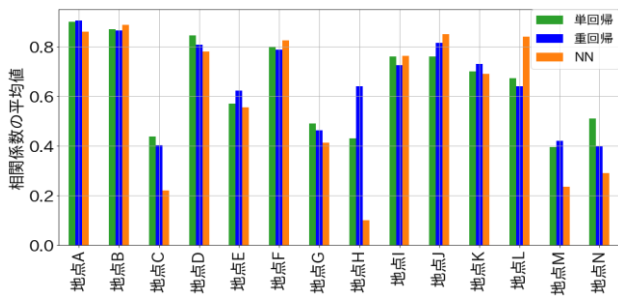


図 12 交差検定による相関係数の平均値

地点	A	B	C	D	E	F	G	H	I	J	K	L	M	N
オーバーラップ数	6	6	2	1	2	1	4	3	4	5	5	5	5	5

図 13 センサのオーバーラップ数

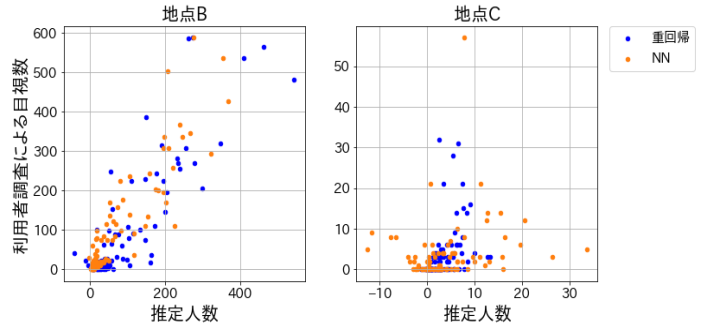


図 14 相関係数の良い地点 (左) と悪い地点 (右)

ス数」は時間の取得したアドレスの合計値であることの違いが推定を困難にしている。

図 14 に相関係数が良い地点と悪い地点を示す。C は遊歩道上にあり目視数が少ない。一般に目視数が少ない地点の精度は悪い。目視数の多い点では、左図に示されているように、人数が少ないケースは NN の方が重回帰よりもよく推定できているが、人数の多い場合、推定数が目視数に比べて小さいほうへ大きくずれる。B は広場で、イベントなど催し物が行われるときには訪問者が急増する。このような場合には、Wi-Fi パケットの衝突により、強い信号のみが捕捉されるため実効的な検出範囲が狭くなってしまうことから、推定数が下振れすると考えられる。検出範囲の端末数依存性を考慮した分析方法の改善が今後の課題である。また、検出範囲に入っていないセンサを説明変数から除くことも精度の向上につながると考えられる。

なお、SVR による推定との比較を行ったところ、NN や重回帰の方がよい精度が得られた。これは、外れ値が一方に偏っていることが原因と考えられる。

5. おわりに

本研究では、9 個のセンサを用いて 14 カ所の地点の目視データと比較することにより目視地点ごとの滞在者数推定を行なった。センサデータは相対的な人の多さを表す指標として役立つが、実数推定は、訪問者数が多い中心部近くの地点に限り有効であることがわかった。単純に目視地点の推定を行うのではなく一定のエリア内の人数を推定することのほうが望ましい。センサの密度、位置関係と推定エリアの大きさと適切に設定する方法を今後検討していきたい。

参考文献

- [1] A. Kurkcu, K. Ozbay, “Estimating Pedestrian Densities, Wait Time, and Flows with Wi-Fi and Bluetooth Sensors”, Transportation Res. Record No.2644, 72-82 (2017).
- [2] H. Hong, G.D. De Silva, M.C. Chan, “CrowdProbe: Non-invasive Crowd Monitoring with WiFi Probe”, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 2, No. 3, Article 115 (2018).
- [3] Khaimook Sippakorn, 葉 健人, 八木 優治, 青木 保親, 猪井 博登, 土井 健司, “限定設置条件下での Wi-Fi パケットセンサに基づく来街者の回遊行動の把握”, 第 59 回土木計画学研究発表会・講演集, 54 (2019).