

改札機カウントデータと Wi-Fi データの統合分析について

藤井 秀夫^{†1} 溝口 雄斗^{†2} 盛江 佳史^{†3} 吉井 英樹^{†4} 曾根原 登^{†5}

概要: 個人情報やプライバシーの保護の観点から、事業者間でデータを共有し統合的に分析活用することはあまり行われていない。本稿では、鉄道事業者と通信事業者の保有するデータを統合的に分析することで生まれる新たな価値やサービスの可能性について検討を行った。さらに、Wi-Fi 検知数から実数を推定するモデルについても考察する。

キーワード: IoT、ビッグデータ、データ連携、官民データ活用

On big data analysis integrating ticket gate counting data with Wi-Fi access log data

HIDEO FUJII^{†1} YUTO MIZOGUCHI^{†2} YOSHIFUMI MORIE^{†3}
HIDEKI YOSHII^{†4} NOBORU SONEHARA^{†5}

Abstract: Due to personal information protection and privacy concerns, there is not much research to share data among different businesses, and analyze and utilize them in an integrated manner. In this paper, we examined the possibility of new values and services created by integrally analyzing the data possessed by railway carriers and telecommunications carriers. In addition, we also show a model that estimates real numbers from Wi - Fi detection numbers.

Keywords: IoT, BigData,

1. はじめに

IoT、ビッグデータが注目を集め、近頃は様々なデータの取得・蓄積が可能になってきた。しかし、個々のデータだけでは、その有効性や信頼性が不十分で上手く活用できない場合が多い。

また、プライバシー情報の保護や個人データの活用の観点から[1]、事業者間でデータを共有し統合的に分析し活用することはあまり行われておらず、各事業者が保有するデータを突合し、補完しあうことでデータの信頼性や有効性を増やす試みもあまり見られない。

本稿は、鉄道事業者と通信事業者がそれぞれ保有するデータを統合的に分析し突合することで、鉄道事業者のデータに属性情報を付与したり、通信事業者のデータから実数を予測するモデルを構築したりするなど、事業者間のデータ突合により得られる新たな価値やサービスの可能性について述べたものである。

2. 研究の目的

本研究は、鉄道事業者が自動改札機等から取得する乗降データ（以下、「改札機カウントデータ」と呼ぶ）と通信事業者がスマートフォンなどの端末と Wi-Fi アクセスポイントとの通信から取得するシステムログ（以下、「Wi-Fi データ」と呼ぶ）を突合し補完しあうことで、正確なお客様の駅等施設の利用状況を把握し、防災・減災、街づくり、観光地・商店街の活性化、利用者のサービス向上などにデー

タを活用することを目的に行ったものである。また、Wi-Fi データから実数を把握するモデルを構築することで、Wi-Fi アクセスポイントを設置することで人の滞留状況の把握や人流の検知精度の向上を目指したものである。

3. 本実証実験で用いるデータについて

3.1 鉄道事業者改札データ

① 改札機カウントデータ

自動改札機は、駅での乗降時に磁気乗車券や IC 乗車券などの券種、普通乗車券や定期券などの種類などを自動的に判別し、お客様から正しく運賃を収受するための機器である。そのため、有効期限の判別やどこで乗車し、どこで降車するかなど運賃計算に必要な情報を読み書きするが、その内容は乗車券の種類により異なるためデータフォーマットは統一されていない。また、保有者の情報は乗車券に保持していないため、性別や年齢などの属性把握不可能である。

鉄道事業者が保有するデータの扱いについては、2013 年 6 月に起きた Suica データ販売に係る議論との関係を整理する必要がある[2]。当時、東日本旅客鉄道株式会社は、「Suica 利用データから氏名、電話番号、物販情報等を除外し、生年月日を生年月に変換した上、さらに、SuicaID 番号を不可逆の別異の番号に変換したデータ」を株式会社日立製作所に提供することが明らかになり、多くの利用者から、「個人情報の保護、プライバシーの保護や消費者意識に対する配慮に欠けているのではないか」として批判や不安視

する声があがった。このことから、販売を見合わせた上で、Suica に関するデータの社外への提供についての有識者会議を立ち上げ、2015 年 10 月に議論の結果をとりまとめ公表している。

このまとめでは、「統計処理した分析結果を、公益目的をはじめ、利用者から理解が得られると考えられる目的に限って提供すること」や「利用者への配慮に十分な注意を用い、社会の期待に応えるようなビッグデータの利活用による成果を積み重ねることにより、今後、利用者の利便性向上はもとより、我が国の社会や経済の発展に寄与することが期待される」と結ばれている。

本稿では、自動改札機から取得できる様々なデータを、改札口単位で乗降別の通過人数を 1 時間単位に集約した「改札機カウントデータ」を用いている。改札機カウントデータは、磁気券については通勤定期、通学定期、定期券外毎に人数を集計しているが、ICカードは定期、定期券外の区別なくトータル人数しか把握できていない。また、保有者情報は、磁気券、ICカード共乗車券には保持していないため、性別や年齢などの属性把握は不可能である。

したがって、本研究で用いる改札機カウントデータは、統計データと考えられるが、先の報告を十分考慮すると共に 2017 年 5 月 30 日に施行された通称「改正個人情報保護法」を踏まえ更新された近畿日本鉄道（以下、「近鉄」と呼ぶ。）の個人情報保護方針に則り取り扱うこととした。また、データ活用のため特定の個人を識別できない態様にするため、統計値の粒度が 10 以上（例えば、改札機カウント数が 9 以下であれば切り下げて 0 と加工）になるように十分な匿名化を施した情報を分析に用いた。さらに、個人情報への考慮として、利用者が少ない時間帯のデータ（5 時台、24 時台）も省いている。

3.2 通信事業者 Wi-Fi データ

① Wi-Fi ログと Wi-Fi 位置情報

Wi-Fi 通信事業者は、サービス状態の把握、品質の計測、改善等を目的とし、様々なログをアクセスポイントから取得している。このうち、Wi-Fi 端末が Wi-Fi アクセスポイントと接続する際に交換されるプローブリクエスト/レスポンスやアソシエーションリクエスト/レスポンス等のログが位置情報として利活用される機会が、スマートフォンの普及と共に拡大している。

ログを位置情報として利用する方法には、端末側で位置情報化する場合と Wi-Fi 通信事業者側で行う場合の 2 通りがある。前者は、端末側で取得した Wi-Fi プローブやビーコン情報からアクセスポイント情報を抽出し、アクセスポイントと位置情報の関係を示したマッチングテーブルをルックアップすることで、位置情報にするものでほとんどのスマートフォンにはこの機能が搭載され、地下や建物内といった GPS では把握困難な位置情報の取得に使われている。

後者は、Wi-Fi 通信事業者が、自ら設置したアクセスポイントの場所と関連付けることで可能とする。前者に比べて位置情報の精度は一般的に良いと言われているが、一方、他事業者が設置したアクセスポイントからはログを収集することが出来ないため、網羅性では前者に劣る。

② 個人情報保護法と Wi-Fi 位置情報

2014 年 7 月、総務省が開催した緊急時等における位置情報の取扱いに関する検討会から、「位置情報プライバシーレポート～位置情報に関するプライバシーの適切な保護と社会的利活用の両立に向けて～」と題した報告書（以下、位置情報報告書）が公開された[3]。この位置情報報告書では、電気通信事業者が取り扱う位置情報として、基地局に係る位置情報、GPS 位置情報、Wi-Fi 位置情報の 3 つを挙げ、現状と課題を示した上で、保護と利活用のバランスに配慮した方向性が示されている。

位置情報報告書において、「アクセスポイント設置者たる電気通信事業者が取得できる Wi-Fi 位置情報は、大別して、①インターネット接続のための準備段階として行われる端末利用者とアクセスポイント設置者との間の通信に基づく位置情報と、②端末利用者がアクセスポイントから外部と通信を行うことで把握される位置情報に分けることができる。」とされ、前者は「今後個人情報保護法上保護される情報として取り扱われる可能性がある」情報、後者は「通信の秘密に該当する位置情報として取り扱うことが適当である」情報という見解が示された。

Wi-Fi 位置情報の取扱いについて、「通信目的で取得・利用する場合」、「通信以外の目的で利用・第三者提供する場合」、「十分な匿名化・低減データ化された場合」という 3 つに分けて整理し、最後のものについては、電気通信事業者は、利用者の同意なく利用・第三者提供することが可能と考えられるが、Wi-Fi 端末利用者に対し周知することと、位置情報の取扱いに係るオプトアウトの機能を設けることが好ましいとされた。

個人情報の保護に関する法律は、2015 年 9 月 9 日にその改正法（平成 2015 年 9 月 9 日法律第 65 号）が公布され、2017 年 5 月 30 日に全面施行された。これに先立ち、総務省から公開された電気通信事業における個人情報保護に関するガイドライン解説（2017 年 4 月 18 日版）では、位置情報の利用（第 35 条第 2 項関係）として、位置情報報告書の考え方を踏襲した内容が盛り込まれた。

③ Wi-Fi 位置情報と MAC アドレスランダム化

Wi-Fi 位置情報を利活用する場合、端末やアクセスポイントの MAC アドレスを識別子として用いることがほとんどである。この MAC アドレスは、ネットワークに接続された機器を（原則として）一意に識別するための 48 ビットの符号であり、当然、Wi-Fi 通信事業者等において契約者の

氏名等個人情報と容易に結びつく場合には、符号単体で個人識別性を有することになる。一方、単体では個人識別性がないとしても、ネットワークに接続されるほぼ全ての端末に付与され、ほとんどの個人がいくつかの端末を常に持ち歩くことを考えると、プライバシー保護の観点から個人情報保護法に抵触しない場合においても保護することがふさわしいという議論がなされて来た。

このような議論の延長線上に、MAC アドレス自体をランダム化し、接続する度に異なる MAC アドレスを利用するなど、MAC アドレスが付与された端末を利用するユーザのプライバシーを保護する動きが広がっている。

④ Wi-Fi 位置情報の利活用に関して

本稿で統計分析に利用した個人情報に該当する Wi-Fi 位置情報は、ソフトバンク株式会社が提供する公衆無線 LAN サービス「ソフトバンク Wi-Fi スポット」及び「Free Wi-Fi PASSPORT」の利用に際し、個人情報の取り扱いに関し同意を得られたユーザのうち、事後的にその利活用を拒否したユーザを除いたものの端末から得られた Wi-Fi プローブとアソシエーションログに基づいたものである。

また、その利用に際しては、ソフトバンク株式会社プライバシーポリシー「お客さま情報の利活用にあたってのプライバシー保護の取り組み」に従った取り扱いとし、ソフトバンク社内において、統計値の粒度が 10 以上（例えば、Wi-Fi 検知デバイス数が、9 以下であれば切り下げて 0 と加工）になるように十分な匿名化を施した情報のみを「Wi-Fi データ」として分析に利用した。

4. 改札機カウントデータと Wi-Fi データの突合の検討

4.1 近鉄奈良駅での実証実験



図 1 近鉄奈良駅 構内図と Wi-Fi 基地局設置場所

近鉄奈良駅は、①. 春日大社、東大寺、興福寺等世界遺産を抱える観光駅、②. 奈良県庁など行政中心であり同時に奈良女子大学等の学校も多数ある到着駅、③. 付近には多数の住宅地があり大阪方面への通勤・通学の出発駅、など異なる利用側面を持つ駅で、様々な利用者の動向分析が可

能と考えた。

実証実験は 2017 年 11 月～2018 年 3 月の間、図 1 の構内図に示す東西改札口に Wi-Fi 装置を設置して行った。なお、近鉄奈良駅は地下駅であり、実際に駅をご利用になるお客さま以外はほとんど Wi-Fi 装置に検出されない環境である。

4.2 近鉄奈良駅の改札機カウントデータの分析

① 改札機カウントデータからみる近鉄奈良駅の特徴

(a) 奈良駅の改札カウントデータ

改札カウントデータから、券種（IC や磁気など）に関係なく東西の改札利用者の合計数について入場と出場別に示したのが図 2（平日）、図 3（休日）である。1 日の乗降人員の平均は、平日が 55,000 人、休日が 50,000 人とあまり差はないが、時間別の利用状況は明らかに異なる。

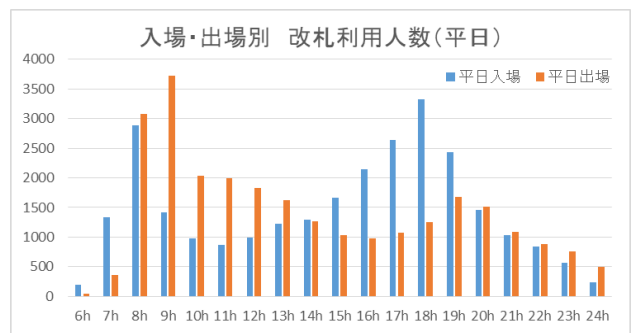


図 2 平日の改札利用人数

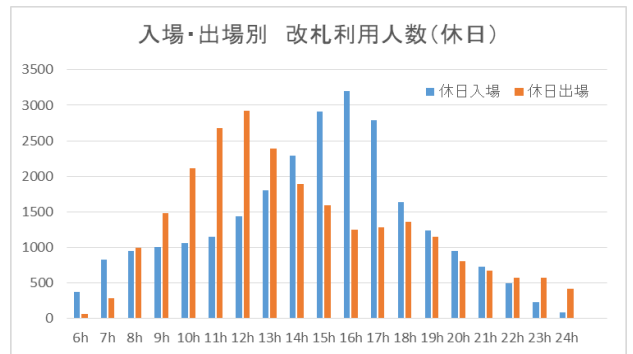


図 3 休日の改札利用人数

(b) 平日

まずは、平日の時間別の利用状況を把握するため、図 4 に磁気の定期と定期券外に分けた改札利用者の変化を示す。（以後図 7 まで同様）。これは、現在の改札機カウントデータでは IC カード利用者の定期、定期外券の区別ができないため、本節では磁気券のみで分析を行った。なお近鉄の磁気券と IC カードの比率は 4 : 6 程度である。

定期利用のピークは、6時から8時と17時から19時であり、通勤通学のお客さまが多いと推測される。一方、定期外利用のピークは9時から18時ごろであるが、図 5 に示すように、出場のピークは9時から11時、入場のピークは14時から18時であり、午前中に奈良を訪れたお客

様が昼過ぎから奈良を離れることが推測され、定期外は観光のお客様が多く含まれると考えられる。

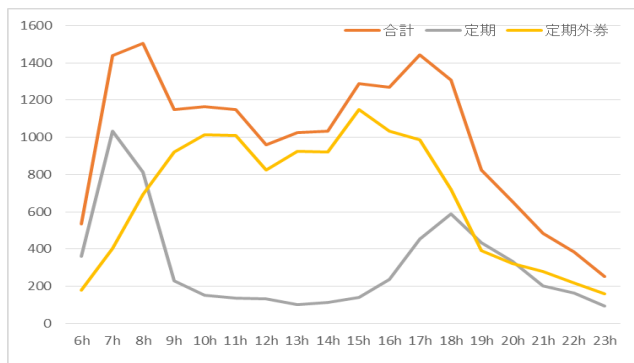


図4 券種別利用状況（平日、磁気券）

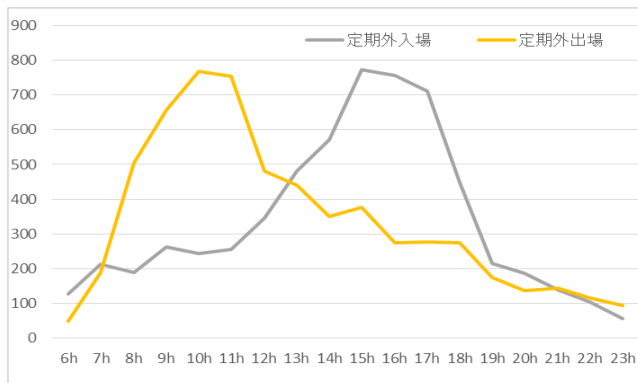


図5 定期券外 時間帯別入出場件数（平日、磁気券）

(c) 休日

利用者数の合計は平日と休日で大きな差はないものの、図6に示すように、休日は定期券外の利用がほとんどで、ピークは9時から18時ごろで平日と同じである。

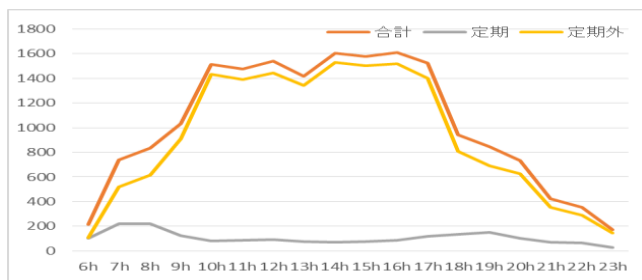


図6 券種別利用状況（休日、磁気券）

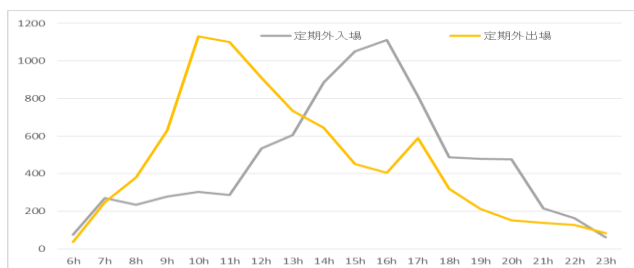


図7 定期外券 時間帯別入出場件数(休日、磁気券)

図7に休日の定期券外の入場、出場件者数を示すが、

平日と同様、午前中に奈良を訪れたお客様が昼過ぎから奈良を離れる傾向が見られるが、その件数は平日よりも多く、休日の近鉄奈良駅は、観光中心であることが推定できる。

以上の分析は、鉄道事業者のデータだけで、かつ利用状況については磁気券データのみで行ったものであるが、定性的に把握していた事象を定量的に表せたことに意義があると考えられる。今後は、乗降駅など様々な情報を保有するICカードも含めた分析を行って行きたい。

4.3 改札機カウントデータと Wi-Fi データの関係

図8は、1日あたりの改札利用者数（右目盛）と Wi-Fi アクセスポイントにより検知された端末数（左目盛）の関係を示したものである。

改札利用者数は、磁気券やICカードなど全ての券種の入出場を合計したもので、改札機を利用した「実数」と考えられる。実際、2017年12月9、10日の奈良マラソン開催日や2018年の年始及び1月27日の若草山焼きの日などには、改札利用件数の増加がみられお客様の動きを反映していると思われる。

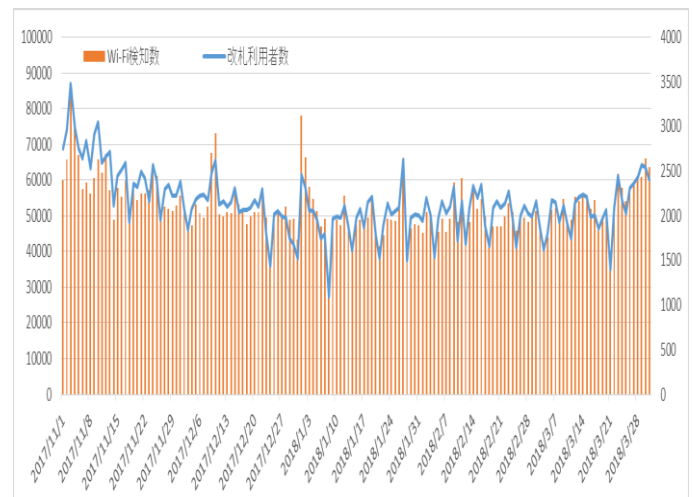


図8 改札利用者数と Wi-Fi 検知数（日別）

なお、スマートフォンなどの端末は、同一アクセスポイントと適時通信を行うため、Wi-Fi データには同一端末が複数回検出される可能性がある。そのため、Wi-Fi 検知数は、一定の期間内のユニークな端末数としたため、朝、近鉄奈良駅から入場し、夕方に出場した場合でも1日当たりのWi-Fi 検知数は1端末とカウントされる。一方、改札機カウントデータには出場と入場がそれぞれカウントされているため、図8では、実際のWi-Fi 検知数の2倍の値を採用している。

図8を見ると改札利用者数とWi-Fi 検知数の増減は近似しているため両者の相関分析を行った。その結果を図9に示すが、 $R^2=0.7179$ という高い相関が見られた。このことから、Wi-Fi 検知数は、全ての人員を把握しているとは限らないが、Wi-Fi 検知数からお客さまの実数を推測すること

や改札機を利用するお客様に Wi-Fi データが保有している属性の適用は可能と考えた。

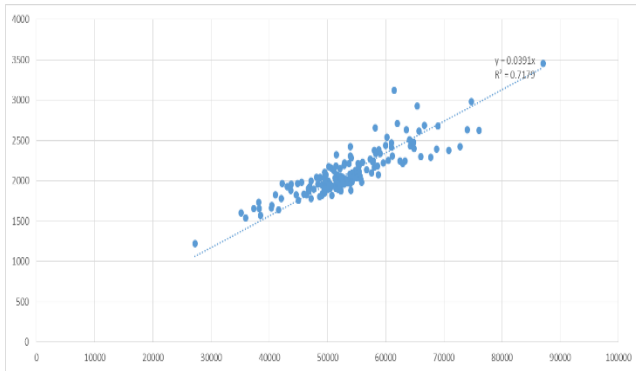


図 9 改札利用者数と Wi-Fi 検知数 (日別) 散布図

4.4 Wi-Fi データの属性からみる近鉄奈良駅の特徴

朝通勤時間帯は男性が多く、昼観光時間帯は女性が多いということは、交通事業者は感覚的に把握しているものの、その実態を定量的に把握することは難しく、アンケート等で実態把握を試みているのが実状である。

本節では、Wi-Fi データが保有する属性を用いて、近鉄奈良駅を利用する顧客属性の可視化を試みた。

① 平日

図 10 は Wi-Fi データが保有する時間帯男女比率を改札利用者数に適用したものである。この図から、通勤時間帯の 6,7 時台には男性が多く、周辺の観光施設等が開館する 10 時頃から女性の割合が増加し、昼間は女性のほうが多くなるという交通事業者の感覚と一致する。

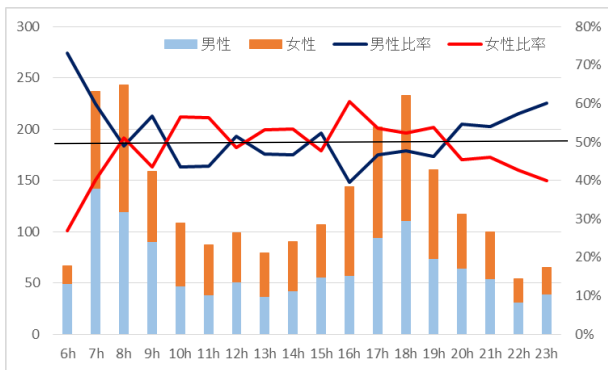


図 10 時間帯別 利用状況 性別比 (平日)

② 休日

平日と同様、図 11 に Wi-Fi データから得た男女比を休日の改札利用者数に適用したものを示す。1 日当たりの休日の利用率の男女比は男性 44% に対して女性 56% となる。時間帯別に見ても終日女性が上回るなど、奈良は女性の観光客の割合が多いことがわかる。

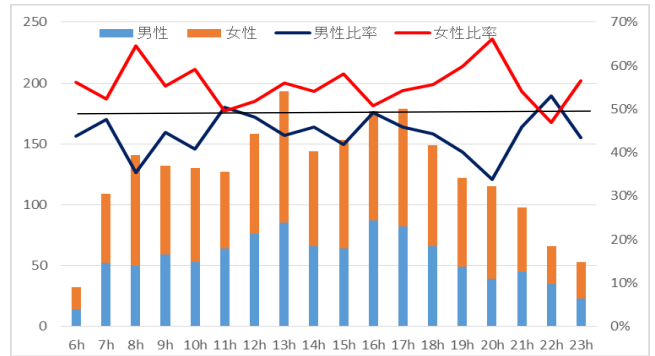


図 11 時間帯別 利用状況 性別比 (休日)

③ 年齢層別

図 12 に示すように、年齢層別に平日・休日毎で利用実態を比較してみると、若年層 (10 代、20 代)、プレシニア層 (50 代) の割合に変化は見られなかったが、ファミリー層 (30 代、40 代) は休日に多く、シニア層 (60 代、70 代) は平日に多いなどの傾向が見られた。

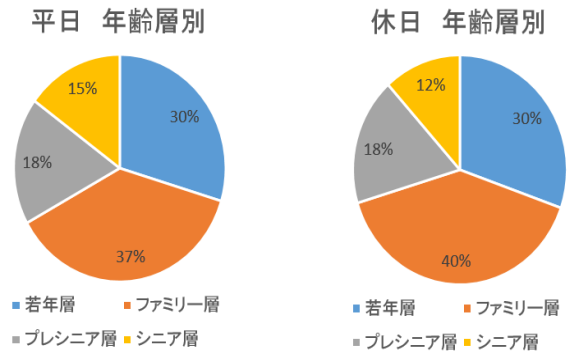


図 12 近鉄奈良駅 年齢層別 利用実態 平日・休日毎

④ 居住地別

図 13 に平日休日別に利用者の住所属性 (契約上の住所) の上位 5 県を示す。平日、休日とも奈良県が最も多いが、平日は東京都や神奈川県など遠方のお客様が多いことに驚かされる。また、休日には京都府、大阪府、兵庫県などのお客様が増加することから、近隣他府県からの流入が増えることがわかる。いずれにしても、平日休日ともに、近鉄奈良駅は地元のお客様と共に観光のお客様が相当数いることが推測できる。

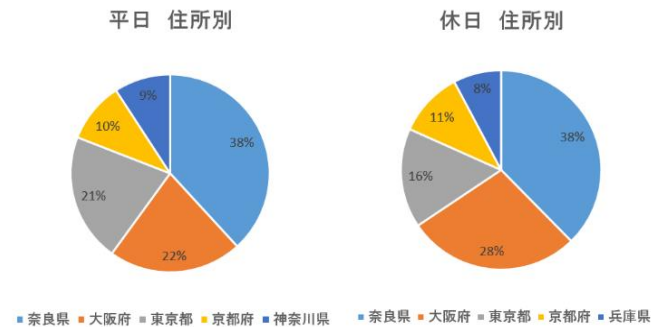


図 13 近鉄奈良駅 住所別 利用実態 平日・休日毎

以上、鉄道事業者データだけでもある程度利用状況の把握ができるが、通信事業者データと突合することでより今まで難しかったより詳細な利用状況の把握や見える化が可能になることを示せた。

5. Wi-Fi データから実数を推定するモデル

5.1 Wi-Fi 検知数を補正する欠損度

本章では、スマートフォンなどの端末が、Wi-Fi アクセスポイントでどの程度検知されるか？について分析を行う。

改札機カウントデータから得られる「改札利用者数」をその地点を通過した実際の人数、すなわち「実数」とし、アクセスポイントで検出された「Wi-Fi 検知数」との比較を行った。一般に、Wi-Fi 検知数から実数は

- ・各通信事業者のシェア率
- ・Wi-Fi 機能を on にしている確率

から計算により求めることができるが、なんらかの理由で Wi-Fi アクセスポイントが端末を捕捉できない度合として「欠損度」を考慮すると Wi-Fi 検知数と実数の関係は(1)式ようになる。なお、欠損度の逆数が「検知率」で、全ての端末が把握できる場合、欠損度 = 1 である。

$$\text{実数} = \frac{\text{欠損度 (A)}}{\text{キャリアシェア率} \times \text{Wi-Fi ON 率}} \times \text{Wi-Fi 検知数} \quad \dots (1) \text{式}$$

したがって、欠損度 (A) は(2)式で計算できる。

$$\text{欠損度(A)} = \frac{\text{実数} \times \text{キャリアシェア率} \times \text{Wi-Fi ON 率}}{\text{Wi-Fi 検知数}} \quad \dots (2) \text{式}$$

今回の実証実験で使用したソフトバンク社のシェアは、2018 年 3 月の電気通信事業者協会の事業者別契約者数から、「キャリアシェア率 = 22.3%」とした。一方、Wi-Fi ON 率は様々な統計があるが本稿では 70% とし(2)式を用いて欠損度の計算を試みた。

5.2 改札利用者数と欠損度の関係

① 平日休日別

(2)式に基づき実証実験期間中の Wi-Fi 検知数と改札利用者 (= 実数) から欠損度を日別に計算すると、平均 1.94 (最大値 2.34 最小値 1.54) となった。欠損度が 1.94 であることは、Wi-Fi データの検知率が 51.5% で、約半数程度の端末しか検出できていないことになる。

日別の欠損度を見るとある程度の周期性が見られたので、暦日の平日、休日に分けて欠損度を求めてみた。その結果、平日の欠損度の平均は 2.05 (最大 2.34 最小 1.74)、休日の欠損度の平均は 1.85 (最大 2.04 最小 1.54) となり、休日の方が平日より端末が検出されやすいことになる。

② 時間帯別

平日と休日に分類したうえで、時間毎に欠損度を算出した結果を図 14、15 に示す。平日の欠損度の平均は 2.99 (最大が 6 時台で 4.03、最小が 23 時台で 2.01)、休日の平均は 2.69 (最大が 7 時台で 3.09、最小が 23 時台で 1.99) となり、日平均と同様休日の欠損度の方が小さかった。

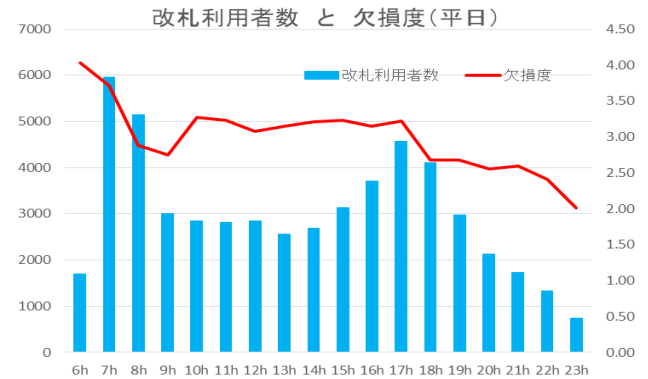


図 14 改札利用者数と欠損度の時間帯別推移 (平日)

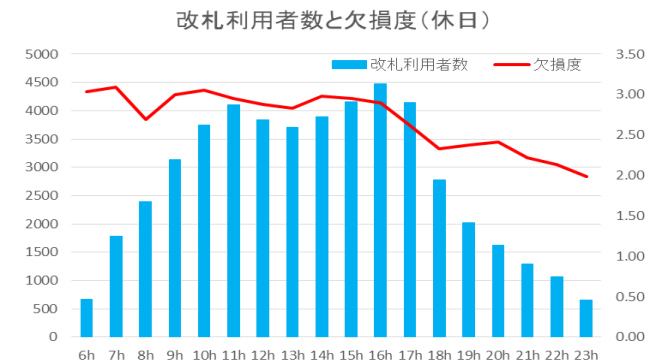


図 15 改札利用者数と欠損度の時間帯別推移 (休日)

5.3 Wi-Fi 検出件数から改札利用者数を求めるモデル

① 実数と欠損度の関係

図 14、15 をから改札利用者が最も多い平日朝の通勤時は欠損度が最も高く、18 時以降は徐々に低くなることなどから、駅混雑率が欠損度に影響を与えるのではないかと考え、平日、休日別に改札利用者数と欠損度の相関分析を行った。(図 16、17 参照)

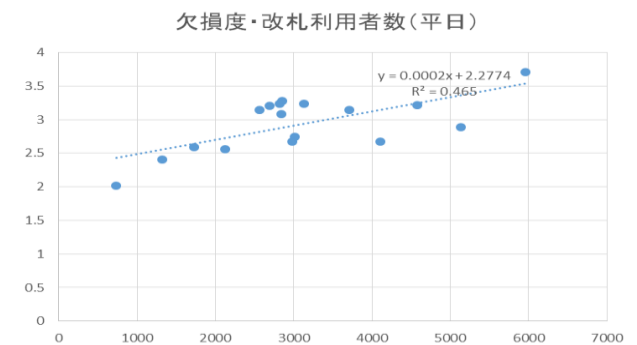


図 16 欠損度と改札利用者数 (平日)

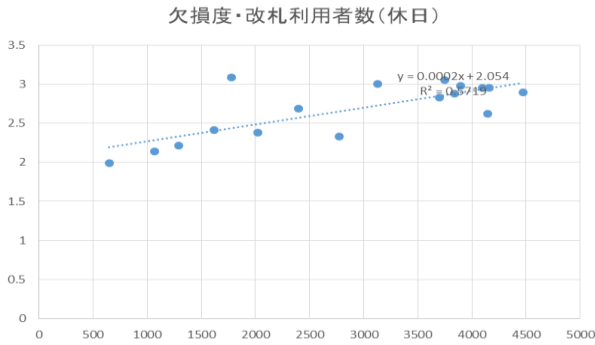


図 17 欠損度と改札利用者数 (休日)

その結果、平日で $R^2=0.46$ 、休日で $R^2=0.57$ の相関が見られ、改札利用者が増加すると欠損度も増加する傾向が見られた。なお、本分析では平日、休日とも、改札利用者数が少ないにもかかわらず欠損度の大きかった6時台は除いており、今後、6時台の欠損度についての分析は必要と考えている。

② 欠損度と Wi-Fi 検知数から実数を予測するモデル

図 1 6, 1 7 より実数 (改札利用者) と欠損度には

$$\text{平日欠損度} = 0.0002 \times \text{実数} + 2.2274 \quad \dots (3) \text{式}$$

$$\text{休日欠損度} = 0.0002 \times \text{実数} + 2.0540 \quad \dots (4) \text{式}$$

の関係がある。仮に、欠損度と実数である改札利用者の推計値の関係を(5)式で表し、(1)式の欠損度に(5)式を代入し整理することで、Wi-Fi 検知数から実数の推計値を計算する(6)式が導き出せる。

$$\text{欠損度}(A) = a \times \text{実数 (推計値)} + b \quad \dots (5) \text{式}$$

$$\text{推計値} = \frac{b \times \text{Wi-Fi 検知数}}{(\text{キャリアシェア率} \times \text{Wi-FiON 率} - a \times \text{Wi-Fi 検知数})} \quad \dots (6) \text{式}$$

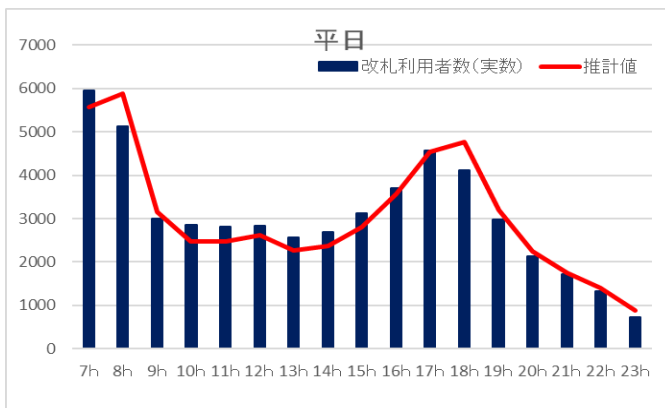


図 18 改札利用者数と推計値(平日)

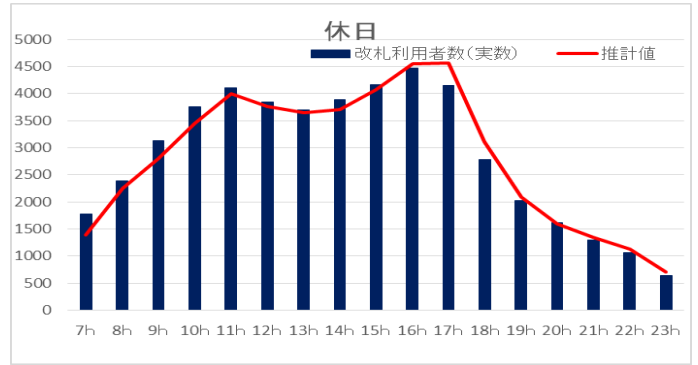


図 19 改札利用者数と推計値 (休日)

(6)式に平日は(3)式 ($a=0.0002, b=2,2274$)、休日は(4)式 ($a=0.0002, b=2,0540$) の値を用いて、それぞれ Wi-Fi 検知数から改札利用者の推計値を求めてみると、図 1 8, 1 9 に示すように、平日休日とも非常によく一致している (平日の相関係数 $r=0.97$ 、休日が $r=0.98$)。したがって、(6)式は、平日と休日別ではあるが、「奈良駅における Wi-Fi 検知数から実数を予測するモデル」と考えられる。

5.4 欠損度に影響を与えるその他の要素

前節では、改札利用者が増加すると欠損率が増加することを前提とした奈良駅モデルについて述べたが、ここでは、平日と休日の差異について検討を行うため、図 1 4, 1 5 を重ね合わせた図 2 0 を作成した。図 2 0 から、全体に欠損度は平日の方が高く、特に 1 0 時から 1 6 時は、改札利用者数の多い休日の方が平日の欠損度より低くなっている。

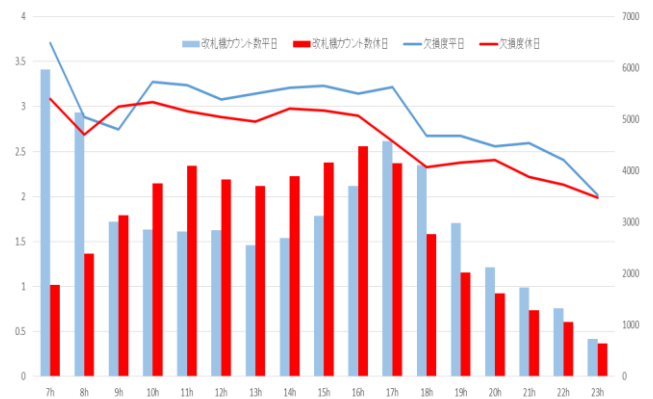


図 20 平日、休日の利用者数と欠損度

このことから、欠損度は単に改札利用者数だけではなく、別の要素が影響を与えていると考えられる。平日と休日の違いとして、駅を定常的に利用しているお客様と観光のように一時的に利用しているお客様の違いの可能性を考え、定期券利用者の割合を比較してみた (図 2 1 参照)。なお、定期券利用者の割合は、4 章と同様、磁気券のデータを用いている。この図を見ると、平日の方が全体的に定期券の利用割合が多く、定期客の行動特性 (歩行速度やルートな

ど)が欠損度に影響を与えている可能性が考えられる。すなわち、観光駅か通勤通学駅かといった駅の特徴により欠損度が変化する可能性が考えられる。

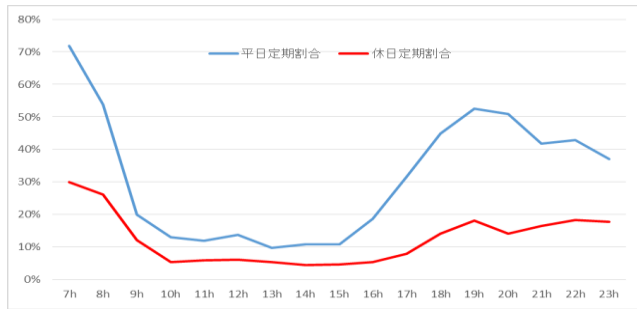


図 21 平日と休日の定期利用者割合

今後、近鉄奈良駅とは特徴の異なる駅、例えば、乗降客が少ない駅や非常に多い駅、あるいは、通勤のお客様が中心の駅や観光駅など、様々な駅で今回と同様の分析を行い、駅の特徴に応じた実数予想モデルの構築を目指したい。また、今回は携帯端末がスマートフォンかそれ以外のレガシー端末か否かの区別や年齢による Wi-Fi On 率の差異などは今後の課題であり、通信事業者などが保有する統計データなどを参考にモデルの更なる精度向上を図りたい。

6. まとめ

本稿では、異業種間データの統合的分析として、近鉄奈良駅の改札機カウントデータと Wi-Fi データの関係についての検討を行った。鉄道事業者が保有する改札機カウントデータから求められる「改札利用者数」と通信事業者が保有する Wi-Fi データから求められる「Wi-Fi 検知数」には強い相関が見られた。これは、Wi-Fi 検知数から実数予測が可能であることを示唆すると共に、改札利用者から Wi-Fi データから取得出来る性別や年代、居住地情報などを付加する妥当性を示すもので、今回、経験や勘から把握していた定性的な情報を具体的な数値で表すことができた。これは、異なる事業者間のデータを突合することで今まで難しかった情報の見える化が行えたと言える。

また、改札利用者数と Wi-Fi 検知数との比較から、奈良駅における平日、休日別に「実数予測モデル」を構築した。このモデルは、携帯端末が Wi-Fi 基地局により検出されない度合いとして「欠損度」という概念を導入したもので、欠損度は改札の利用者が増えるに従い上昇する（検知率が下がる）傾向が見られた。同時に欠損度はお客様の行動にも影響を受けると思われ、より精度の向上が必要であるが、今回のモデルは、Wi-Fi データを用いてリアルな時間空間での実数予測の可能性を示すものと言える。

さらに、今回は鉄道の駅における分析であるが、テーマパークなどのレジャー施設、百貨店などの商業施設でも同様の分析が可能と考える。すなわち、これらの施設が保

有するデータと通信事業者のデータを用いることで駅に限らずリアルな時間空間での人数予測や行動特性の把握なども可能と考える。

最後に、今回は実証実験のために設置した Wi-Fi 基地局を用いた分析であるが、キャリアなどが設置した Wi-Fi 基地局や携帯電話基地局などの情報と連携することで、駅だけでなく広範囲の導線分析も可能と考える。これらの分析により、防災・減災、街づくり、観光地・商店街の活性化、利用者のサービス向上などにも繋げて行きたい。

参考文献

- 1) 曾根原 登 (著), 宍戸 常寿 (著), 安岡 寛道 (編集), “ビッグデータ時代のライフログ —ICT 社会の“人の記憶””, 東洋経済新聞社, 2012/6/1
- 2) “Suica に関するデータの社外への提供について”
http://www.jreast.co.jp/information/aas/20151126_torimatome.pdf
- 3) “位置情報プライバシーレポート”.
http://www.soumu.go.jp/main_content/000434727.pdf