

# 鉄道における Bluetooth 受信特性を用いた 乗車車両および混雑の推定手法

前川 勇樹<sup>1</sup> 内山 彰<sup>1</sup> 山口 弘純<sup>1</sup> 東野 輝夫<sup>1</sup>

概要：本稿では、快適な鉄道利用を実現するため、鉄道旅客の乗車車両および各車両の混雑状況を推定する手法を提案する。提案手法は、乗客の持つ携帯端末が受信した近隣端末の Bluetooth シグナルをサーバに集約し、各端末間で観測された RSSI から、それらが同じ車両に存在する確率（同一車両確率）および端末間の混雑確率を算出する。次に、得られた同一車両確率および一部の端末の（信頼度の高い）乗車車両情報を用いて全端末の乗車車両を推定し、その結果と端末間の混雑確率を用いて車両毎の混雑を推定する。この際、電車内の乗客移動は一般にあまり見られないことを利用し、同一車両確率を継続的に更新することで、乗降車が発生しても高精度かつ迅速な推定を実現する。大阪都市部における 4 路線において 259 分間に渡り収集したデータを用いて提案手法の評価を行った結果、16 名の各車両位置を精度 83%、車両毎の混雑の有無を F 値 0.75 で推定できることがわかった。

キーワード：位置推定, Bluetooth, ユーザ参加型センシング, 鉄道

## 1. はじめに

我々の都市生活において快適な移動を実現するためには、電車や地下鉄のような公共交通機関での移動をよりスムーズで快適にすることが重要である。東京などの大都市では、電車や地下鉄が都市交通システムとして十分に発達する一方で、鉄道駅自体が巨大化し、複雑な構造になっている。一つのエリアに複数の階段やエレベータなどの出口が混在し、それぞれが異なる方面や別路線の乗換口へと繋がるため、慣れない旅客は電車降車後にどの方向に向かうべきかも把握しにくい。さらに、身体的障害を抱える旅客、大きな荷物やベビーカーを運ぶため移動に負担が伴う旅客らは、空いている路線や車両など、より負担が少なく安全な移動を望んでいる。このような旅客のニーズに応じた快適な鉄道利用をサポートするためには、プラットフォーム/電車内に存在する旅客の位置特定や電車内の混雑状況把握が有用であると考えられる。

しかし、現状ではそれぞれに課題が存在する。旅客の車両特定については、バスや電車自体の位置が追跡可能である GPS に基づく位置推定手法 [1] を応用可能とも考えられるが、乗車車両推定やプラットフォーム上の移動端末の位置推定に必要な粒度の位置情報が得られない。混雑状況の

把握については、ジョルダンライブなどで混雑情報を提供しているが、これらは全てユーザの入力に頼っている。そのため、混雑情報を入力する手間がかかる上、どの車両に関する混雑状況の入力であるかがわからない場合もある。

本稿では、これらの課題を解決し快適な鉄道利用を実現するため、旅客の乗車車両および車両毎の混雑を推定する手法を提案する。提案手法は、乗客の持つ携帯端末が受信した近隣端末の Bluetooth シグナルをサーバに集約し、各端末間で観測された RSSI から、それらが同じ車両に存在する確率（同一車両確率）および端末間の混雑確率を算出する。次に、得られた同一車両確率および一部の端末の信頼度の高い乗車車両情報を用いて全端末の乗車車両を推定し、その結果と端末間の混雑確率を用いて車両毎の混雑を推定する。この際、電車内の乗客移動はあまり見られないことを利用し、同一車両確率を継続的に更新することで、乗降車が発生しても高精度かつ迅速な推定を実現する。

大阪都市部の 4 路線で 259 分間収集したデータを用いて提案手法を評価した結果、16 名の各車両位置を 83% の精度で推定し、車両毎の混雑の有無を F 値 0.75 で推定できることがわかった。

## 2. 関連研究

世界各国で公共交通機関向けの位置情報サービスが整備され始めており、旅客により快適な交通機関の利用を促す

<sup>1</sup> 大阪大学大学院情報科学研究科  
Graduate School of Information Science and Technology, Osaka University, Japan

ための移動支援に関する研究が盛んとなっている。たとえば、ジョルダンライブや NAVITIME[2] では、混雑や運行情報といったコンテキスト把握を目的とし、ユーザからの状況報告を共有するためのプラットフォームを提供している。一方、携帯端末を保持するユーザの参加型センシングにより、バスや電車の位置に加えて到着時刻推定を行うシステム [1], [3] など、携帯端末の位置推定技術を活用した事例も報告されている。中には、本稿で目指す乗車車両推定に応用可能な方法もあると考えられるが、電車内での運用はコストや精度などに関する課題がある。たとえば、文献 [4] にあるように、GPS 信号による端末位置の検出も可能と考えられるが、GPS では乗車車両の特定に十分な精度を得られないことに加え、群衆の存在は検出できない。歩行者や群衆のトラッキングにはカメラなどの動画像を用いることが多い(文献 [5])。しかし、駅などに設置されたカメラの映像は、プライバシーの問題から人身の安全やセキュリティの目的以外で利用することは困難である。また、車両内においてはコストの問題もあり、導入されていないことが多い。一方、低コストに実装可能な携帯端末を用いた群衆推定や人数計測技術も提案されている。屋内や狭領域における混雑推定では、端末内部のセンサや端末間での相互通信を用いる場合が多く、端末間で音による通信を行うことで環境内の端末数を推定する手法 [6] などが検討されている。さらに、文献 [7] では、近隣の Bluetooth 端末の検出状況を集約し、近隣端末数などの特徴量から群衆密度を推定する手法を提案している。しかし、電車内の混雑を対象とする場合には、車両毎に混雑が異なるため、混雑状況だけでなくその車両も把握する必要がある。

本稿では、鉄道における快適な移動をサポートするため、旅客の乗車車両および車両毎の混雑という2つのコンテキストを低コストで高精度に推定する手法を提案する。旅客のリアルタイムな位置情報および車両の混雑状況を携帯端末の Bluetooth 情報のみからインフラに頼らず推定する試みは、我々の知る限りこれまでに提案されていない。提案手法では、ベイズ推定に基づく逐次的な確率更新によって、それらを同時に求めるアルゴリズムを考案している。このように相関する2つのパラメータを時系列的に同時更新していく技術は SLAM[8] においても利用されているが、我々は電車を対象とし、(i) Bluetooth RSSI はユーザの車両位置関係や車両混雑の影響を受ける、(ii) 乗客の車両移動はあまり見られないという電車環境の特徴を考慮したシステムの設計を行っている。

### 3. 事前実験による分析

電車車両内における Bluetooth 信号の RSSI 特性を調査するため、事前実験を行った。実験では実験参加者 17 名(以下、ユーザ)が1台ずつ携帯端末 Nexus S (Android 4.1) を首から下げ、計 259 分間に渡る Bluetooth サンプル

表 1: 各路線における列車仕様

路線	駅数	編成	車両長	車幅	車高	座席定員
M	23	10 両	18.74m	2.89m	3.74m	44 人/両
OL	19	8 両	19.50m	2.80m	4.14m	54 人/両
H	11	7 両	18.90m	2.80m	4.09m	54 人/両
OM	5	4 両	14.60m	2.98m	5.20m	40 人/両

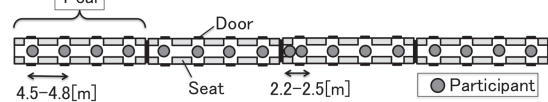


図 1: ユーザの配置

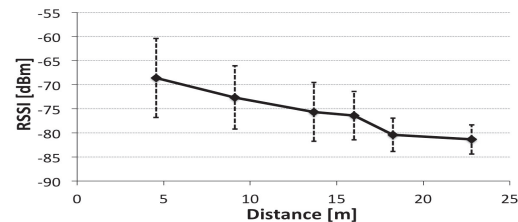


図 2: 距離 vs. 全サンプル平均 RSSI (誤差棒は標準偏差)

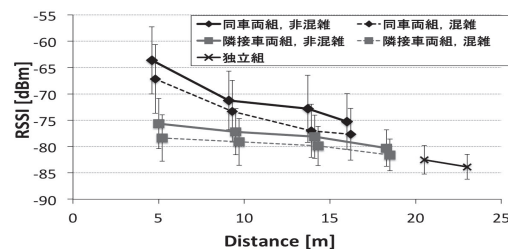


図 3: 距離 vs. 分類別平均 RSSI (誤差棒は標準偏差)

およそ 5 万個を収集した。ユーザの車内配置は図 1 に従う。各端末では Bluetooth の端末検出機構で用いられる近隣端末問い合わせ (inquiry) および問い合わせ待受 (inquiry scan, 約 12 秒間) を交互に行うアプリケーションを動作させた。また、実験は大阪市営地下鉄御堂筋線 (M), 大阪環状線 (OL), 阪急千里線 (H) の計 3 つの路線で行った。各路線の詳細を表 1 に示す。

まず、電車内の距離に対する全サンプル平均 RSSI の測定結果を図 2 に示す。電車内においても自由空間同様に距離による信号減衰が見てとれるが、端末間距離が近い場合は分散が非常に大きい。一方で、端末間距離が車両長(約 18m)を超える場合は、車両間扉による影響を受けるため平均 RSSI や分散が似通っている。そのため、RSSI に基づく正確な距離推定を行うことは困難であるといえる。

そこで、RSSI の減衰や分散に影響を与える要因として考えられるユーザの相対的な車両位置関係や混雑状況により図 2 のデータを分類し、RSSI により距離を推定するのではなく、それらの区別が可能かを調べた。以下では、ある 2 人のユーザが同じ車両に乗車している場合を同車両組、隣り合う車両に乗車している場合を隣接車両組、それ以外の場合を独立組とする、3 つの位置関係を定義する。また、表 1 中の座席定員に基づき、対象車両に 60 人以上の乗客が存在する場合に混雑、それ以外是非混雑と定義する。

これらの分類に基づく平均 RSSI とその分散を図 3 に示

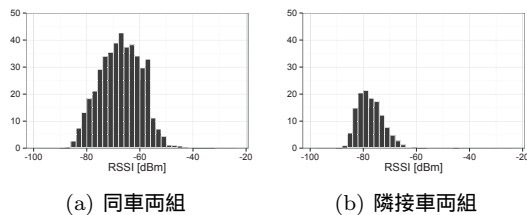


図 4: RSSI の分布状況

す．図 3 における隣接車両組の車両混雑は，少なくとも一方の車両が混雑している場合に混雑とし，それ以外は非混雑としている．図 3 において，独立組は inquiry がほとんど受信されず，全サンプル中約 0.002% のサンプル数しか存在しなかった．実際に観測された値も非常に小さく，inquiry 受信の閾値（約 -90dBm）に近い値であることがわかる．これは複数の車両間距離が減衰に大きく影響するためであると考えられる．このことから，ユーザ間で inquiry が観測される場合はほぼ同車両組か隣接車両組とみなすことができるといえる．また，RSSI が強い場合（-75dBm 以上）は，高い確率で（相対距離が短い）同車両組であるといえるものの，RSSI がそれ以下の場合は同車両組か隣接車両組かの単純な判定は難しいこともわかる．さらに，図 3 では同じ位置関係にあっても，混雑するほど人体の遮蔽による影響が顕著になるため，混雑時と非混雑時では平均 RSSI に多少の差があることがわかる．しかし，ユーザ組の位置関係に見られるほどの差ではない．以上の結果から，距離が近い場合の同車両組，および独立組の判定は比較的高い信頼度で可能であり，隣接車両組と同車両組の分別は単純でないことがわかった．したがって，強い RSSI が得られる場合はその結果を利用してユーザ組の分類を行い，そうでない場合は信頼度の高い情報に基づいて信頼度をあげたり，観測を重ねて信頼度をあげる工夫が必要である．また，混雑の判定は単独では難しいため，ユーザ組の判定後に行って精度を上げるべきであると考えられる．

最後に，RSSI で同車両組と隣接車両組の判定がどの程度可能かを調べるため，混雑や非混雑，ユーザ間距離を無視した場合に同車両組と隣接車両組で観測される平均 RSSI の分布を調べた．図 4 に，様々な状況下で得られたサンプルを混合した際の各組に関する RSSI の分布状況を示す．図 4 では，ノイズの影響を軽減するために 60 秒間に観測された平均 RSSI を示している．図 4 の同車両組および隣接車両組のヒストグラムには違いが見られるため重ならない部分に関しては判別可能だが，重なる部分に関しては曖昧性が認められる．そのため，あるユーザ組で観測される RSSI からその位置関係を正確に求めるためには図 4 に示されるような統計的な違いを反映する必要がある．

以上の知見を基に，提案手法ではユーザ組の相対的な車両位置関係の確からしさを積み重ねることで統計的に正しい位置関係が得られるようにする．さらに，RSSI のみでは限界があるため，電車内の乗客移動はあまり見られない

という特徴を利用して過去の確からしさを引き継ぎ，得られた各ユーザ組の確からしさに基づくトポロジ情報によって位置関係を補完する方法を採用する．

#### 4. 乗車車両および混雑の同時推定手法

提案手法は，Bluetooth 通信可能な携帯端末と集約した Bluetooth 端末間の RSSI を解析するサーバで構成され，電車が各駅間の経路（区間）を走行中に，ユーザによる観測（観測フェーズ）とサーバによる推定（推定フェーズ）の順に実行される．観測フェーズでは，観測期間  $T$  の間，各ユーザが Bluetooth の端末検出機構を用いて，観測される周辺端末のユーザ ID と RSSI を取得した後，携帯電話網を介して観測した全ユーザ ID とその平均 RSSI をサーバに集約する．続く推定フェーズでは，3 章での知見に基づき，平均 RSSI を用いて各ユーザ組の相対的な乗車位置関係を算出する．ここでは，独立組であるユーザ間はほとんど観測が得られないことから観測が得られたユーザ組が同車両組または隣接車両組のいずれかを求める．次に，信頼度の高い乗車車両情報が得られる一部ユーザを仮定した上で，それらの絶対位置情報および算出した相対位置関係からユーザの乗車車両を推定する．さらに，乗車車両推定において同一車両に乗車していると推定されたユーザ間の平均 RSSI を用いて各車両の混雑状況を推定する．

以降では，上述のような信頼度の高い乗車車両情報が得られる少数のユーザをランドマークと呼ぶ．ランドマークに選出されるユーザには様々な場合が考えられる．たとえば，電車車両内に WiFi 基地局が将来的に配置されることがあれば，その付近の乗客は WiFi の電波強度に基づき自身の乗車車両が高い確率で推定でき，他の乗客に対するランドマークとなり得るが，本稿では，簡単のためランドマークは自身の乗車位置を把握しており，手動でシステムに乗車車両番号を通知するものとする．提案手法は少数のランドマークを想定しているが，位置基準情報が十分でない場合のみ必要となり，毎区間ランドマークが必要というわけではない．特に，サーバに十分な情報が集約されていない初期区間ではランドマークが必要だが，初期の区間以降では，乗降車などによるユーザの入れ替わりが発生しても新規ユーザに既存ユーザの推定結果を伝搬させることで，各ユーザの位置情報を算出可能である．

以降では，単一電車におけるユーザの乗車車両と車両別の混雑状況の推定に着目する．この理由は，文献 [1], [3] のように時刻表などの運行情報と GPS や WiFi などの粗い位置情報を利用することで，ユーザが乗車している電車は容易に特定可能なためである．たとえば，粗い位置情報の時系列から電車の移動方向を推定することで，周辺を運行している電車のうち，最も移動方向に近い電車にユーザが乗車していると予測できる．したがって以降では，車両列  $V = \{v_1, v_2, \dots, v_{|V|}\}$  ( $v_i$  と  $v_{i+1}$  は隣接車両) で構成され

る単一電車およびそれに乗車するユーザのみを対象に提案手法の説明を行う。

#### 4.1 観測フェーズ

相互に観測を行ったユーザ  $i, j$  がどちらもシステム参加ユーザである場合、観測期間  $T$  後のそれぞれの平均 RSSI  $r_{ij}$  および  $r_{ji}$  を算出し、サーバに送信する。サーバ上では、得られた両平均 RSSI からユーザ  $i, j$  間の平均 RSSI を決定する。ユーザ間距離が長い場合に強い信号が観測されることは稀であるのに対し、短い場合では人体の遮蔽など様々な要因が影響し、弱い信号が観測されることも多い。つまり、電波強度が弱い信号よりも強い信号が信頼できると考えられる。よって、ユーザ  $i, j$  間の平均 RSSI を  $r_{ij} = r_{ji} = \max(r_{ij}, r_{ji})$  と定義する。これにより、たとえば  $i, j$  の組に対し、 $j$  がシステムの参加ユーザでない乗客が保持するデバイスであった場合も、 $i$  が参加ユーザであれば、 $i, j$  をユーザ組として扱える利点もある ( $r_{ji}$  は報告されないが、上式を用いて  $r_{ij}$  から推測できる)。提案手法では、非参加端末  $j$  を参加ユーザと同様に扱うことで、できる限り多くの情報を収集し、推定精度を向上させる。

#### 4.2 推定フェーズ

推定フェーズでは、ユーザから収集した観測データを基にその乗車車両および車両毎の混雑状況を推定する。ここで、区間  $s$  においてユーザ  $i$  が車両  $v$  に乗車している確からしさを乗車車両確率  $l_i^s(v)$  とする。観測フェーズ後、サーバでは事前学習によって構築した RSSI に関する尤度関数を用いて、観測値  $r_{ij}$  からユーザ  $i, j$  が同車両組である確率 (同一車両確率) を算出する。次に、算出した各組の同一車両確率とランドマーク位置情報を用いて、各ユーザ  $i$  の乗車車両確率  $l_i^s$  を算出することでその乗車車両を推定する。

また、乗車中のユーザによる車両移動は頻繁には起こらないという仮定に基づき、区間  $s$  における同一車両確率および乗車車両確率は、前区間  $s-1$  の結果を考慮したベイズ推定による逐次更新を行う。さらに、各ユーザの非観測情報を利用して、更新結果の補正を行う。非観測情報とは、あるユーザ間で RSSI が全く観測されなかったことを指す。ユーザ  $i, j$  が互いに観測されなかった場合、3章の分析結果から  $i, j$  は同一車両に乗車していないとみなせるため、これを利用することで推定精度の向上を図る。

最後に、各車両が混雑している確率 (混雑確率) を各ユーザ組の平均 RSSI から算出する。混雑確率は、同一車両確率と同様に RSSI に関する尤度関数を用いて算出される。各車両  $v$  の混雑状況は、車両  $v$  に乗車していると推定された全ユーザ組の混雑確率を乗車車両確率による重み付き多数決した結果から判定する。

##### 4.2.1 同一車両確率の算出

区間  $s$  においてユーザ  $i, j$  が同一車両に乗車している事象を  $S_{ij}^s$  とすると、 $i, j$  の同一車両確率は平均 RSSI  $r_{ij}^s$  が

#### Algorithm 1 区間 $s$ における乗車車両確率の更新

```

Require: ユーザ集合  $N^s$ , ランドマーク集合  $L^s (\subseteq N^s)$ , 電車  $V$ 
Ensure:  $\{l_i^s(v) | \forall i \in N^s, \forall v \in V\}$ 
for  $i \in N^s, v \in V$  do
    乗車車両確率  $l_i^s(v)$  を初期化 (式 3 または式 4)
end for
for  $i, j \in N^s$  do
    同一車両確率  $P(S_{ij}^s | r_{ij}^s)$  を算出 (式 1)
end for
for  $t = 1$  to  $R$  do
    for  $i \in N^s - L^s, v \in V$  do
        乗車車両確率  $l_i^s(v)$  を更新 (式 5)
    end for
    for  $i \in N^s - L^s, v \in V$  do
         $l_i^s(v) \leftarrow l_i^s(v) / \sum_{w \in V} l_i^s(w) / *$  正規化  $*/$ 
    end for
end for

```

観測された下で仮説  $S_{ij}^s$  が正しい確率である。すなわち、同一車両確率は事後確率  $P(S_{ij}^s | r_{ij}^s)$  に等しく、ベイズの定理にしたがって次のように表される。

$$P(S_{ij}^s | r_{ij}^s) = \frac{P(r_{ij}^s | S_{ij}^s) P(S_{ij}^s)}{P(r_{ij}^s | S_{ij}^s) P(S_{ij}^s) + P(r_{ij}^s | \bar{S}_{ij}^s) P(\bar{S}_{ij}^s)} \quad (1)$$

尤度関数  $P(r_{ij}^s | S_{ij}^s)$  および  $P(r_{ij}^s | \bar{S}_{ij}^s)$  は事前学習によって定義する。この尤度関数の詳細については 4.3 節で説明する。 $P(S_{ij}^s)$  は事象  $S_{ij}^s$  の事前確率であり、ユーザ  $i, j$  が前区間でも同じ電車に乗車していた場合、乗車中の車両移動は起こらないという仮定に基づき次式で与える。

$$P(S_{ij}^s) = P(S_{ij}^{s-1} | r_{ij}^{s-1}) \quad (2)$$

なお、 $i, j$  の少なくともいずれか一方が区間  $s$  からの新規参加ユーザである場合には、事前知識 (前区間の推定結果) がないので、 $i, j$  がそれぞれがランダムに車両を選んだ場合に同一車両となる確率  $1/|V|$  を事前確率  $P(S_{ij}^s)$  に設定する。

##### 4.2.2 乗車車両確率の更新手順

区間  $s$  における乗車車両確率の更新手順をアルゴリズム 1 に示す。まず初期化のため、前区間  $s-1$  に引き続きユーザ  $i$  が電車に乗車していた場合、 $l_i^s(v)$  に  $l_i^{s-1}(v)$  を設定し、 $i$  が区間  $s$  からの新規参加ユーザならば、 $l_i^s(v) = 1/|V|$  とする (式 3)。なお、 $i$  が車両  $v$  に存在するランドマークならば、 $l_i^s(v) = 1$  とし、その他の確率は 0 とする (式 4)。

$$i \in N^s - L^s \Rightarrow l_i^s(v) = \begin{cases} l_i^{s-1}(v) & (i \in N^{s-1}) \\ 1/|V| & (otherwise) \end{cases} \quad (3)$$

$$i \in L^s \Rightarrow l_i^s(v) = \begin{cases} 1 & (i \text{ が車両 } v \text{ に存在}) \\ 0 & (otherwise) \end{cases} \quad (4)$$

次に、ランドマークのように、より信頼できるユーザの乗車車両確率を位置基準点として固定し、4.2.1 節で算出した同一車両確率に基づいて、位置基準点の乗車車両確率を伝搬させることで他ユーザの乗車車両確率を更新する。たとえば、ユーザ  $i$  の車両  $v$  に関する乗車車両確率  $l_i(v)$  は、



$i$  との間で観測を持つユーザ  $j$  の保持する乗車車両確率  $l_j$  に基づき更新する．この際、 $i, j$  が同車両組である確率が高いほど、 $l_i(v)$  は車両  $v$  に関する  $j$  の乗車車両確率  $l_j(v)$  に近づけることが望ましい．一方、 $i, j$  が隣接車両組である確率が高いほど、 $l_i(v)$  は車両  $v$  の隣接車両  $v_k$  に関する  $j$  の乗車車両確率  $l_j(v_k)$  をより強く反映させることが望ましい．そこで、 $i, j$  が同一車両組である確率  $P(S_{ij}|r_{ij})$  および隣接車両組である確率  $1 - P(S_{ij}|r_{ij})$  をそれぞれ重みとした  $j$  の乗車車両確率  $l_j(v), l_j(v_k)$  の和によって  $l_i(v)$  を更新する．提案手法では、位置基準とするランドマークとの間で観測をもつユーザから始めて  $R$  回この確率更新を行うことで位置基準情報を全ユーザに伝搬、反映する．

この方針に基づき、現在の更新回数が  $t$  である場合のユーザ  $i$  の車両  $v$  に関する乗車車両確率を  $l_i^{s,t}(v)$  とすると、 $t+1$  回目の更新  $l_i^{s,t+1}(v)$  を次のように計算する．

$$l_i^{s,t+1}(v) = \sum_{j \in N_i} \left[ l_j^{s,t}(v) P(S_{ij}^s | r_{ij}^s) + \frac{1 - P(S_{ij}^s | r_{ij}^s)}{|V_k|} \sum_{v_k \in V^k} l_j^{s,t}(v_k) \right] \cdot \text{negative}(i, v, s, t) \quad (5)$$

ここで、 $N_i$  は  $i$  との間で RSSI が観測されたユーザの集合、 $V_k \subseteq V$  は車両  $v \in V$  の隣接車両集合である．また、関数  $\text{negative}$  は更新結果の非観測情報による補正を目的とし、以下の式で定義される．

$$\text{negative}(i, v, s, t) = \frac{\sum_{j \in N - N_i} (1 - l_j^{s,t}(v))}{\sum_{j \in N - N_i, w \in V} (1 - l_j^{s,t}(w))} \quad (6)$$

関数  $\text{negative}$  を導入した理由を述べる．たとえば、 $i$  が車両  $v$  に乗車している場合を考える． $i, j$  の同一車両確率が低い場合、 $j$  は隣接車両  $v-1$  または  $v+1$  のどちらか一方に乗り込んでいると考えられる．しかし、 $j$  がどちらに乗り込んでいるか判断がつかないため、式 5 では車両  $v-1$  および  $v+1$  に関する  $i$  の乗車車両確率が均等に上がってしまう．この確率がさらに伝搬されると各ユーザで算出される乗車車両確率はランドマークの存在する車両を中心として対象な位置関係にある 2 つの車両についてほぼ等しくなり、実際の存在状況を正しく反映できない．そこで、提案手法ではこのような対称性による不確実性をなるべく軽減するため、関数  $\text{negative}$  を用いて  $i$  と観測を持たないユーザ集合  $N - N_i$  が乗り込んでいる確率が高い車両に関する  $i$  の乗車車両確率を小さくするよう重みづける．

以上のようにして、1 回の更新が完了する．ユーザの乗車車両確率は各更新の終了時に正規化を行い次回更新時の計算に利用する．提案手法では、ランドマークが存在した区間の結果を事前確率として引き継ぎながら現区間の乗車車両確率を計算するため、各区間に必ずランドマークが必要というわけではない．ランドマークが存在しない区間では閾値  $l_{TH}$  を超える乗車車両確率を持つユーザを擬似ランドマークとみなし、その他ユーザの乗車車両確率を計算

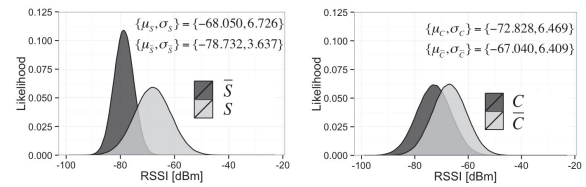


図 5: 路線 M に関する尤度関数

する．このとき、擬似ランドマークの乗車車両確率を基準に他のユーザの乗車車両確率を決定するため、擬似ランドマークに選出されたユーザの乗車車両確率は更新しない．

#### 4.2.3 混雑推定

乗車車両推定の更新結果およびユーザ間で観測された平均 RSSI に基づき各車両の混雑状況を推定する．区間  $s$  において、ユーザ  $i, j$  の乗車車両が混雑している事象および混雑していない事象をそれぞれ  $C_{ij}^s, \bar{C}_{ij}^s$  とすると、その事象の確率は平均 RSSI  $r_{ij}^s$  が観測された下で仮説  $C_{ij}^s, \bar{C}_{ij}^s$  が正しい確率である．すなわち、混雑/非混雑確率はベイズの定理に従う事後確率として次のように定義できる．

$$P(C_{ij}^s | r_{ij}^s) \propto P(r_{ij}^s | C_{ij}^s) P(C_{ij}^s) \quad (7)$$

$$P(\bar{C}_{ij}^s | r_{ij}^s) \propto P(r_{ij}^s | \bar{C}_{ij}^s) P(\bar{C}_{ij}^s). \quad (8)$$

なお、混雑状況の定義はアプリケーションなどの要求や環境に依存するが、本稿では“車両内の座席に座れそうかどうか”を基準とする．

次に、車両  $v$  の乗車車両確率が最も高いユーザの集合  $N(v)$  による一種の「多数決」により車両  $v$  の混雑状況を推定する．この際、 $i, j$  が高確率で同一車両  $v$  に乗り込んでいる場合に、 $i, j$  間の観測から得られる混雑確率が大きく反映されるようにしている．これにより、推定する車両  $v$  の混雑確率が車両  $v$  における観測から生じている確からしさを考慮する．区間  $s$  における車両  $v$  の混雑確率対混雑確率比（混雑尤度比） $m^s(v)$  は、次のように定義される．

$$m^s(v) = \frac{\sum_{i, j \in N(v)} l_i^s(v) l_j^s(v) P(r_{ij}^s | C_{ij}^s) P(C_{ij}^s)}{\sum_{i, j \in N(v)} l_i^s(v) l_j^s(v) P(r_{ij}^s | \bar{C}_{ij}^s) P(\bar{C}_{ij}^s)}$$

一般的に混雑状況は区間毎に大きく変動するため、前区間の確率を引き継ぐのは妥当ではない．しかし、朝夕の通勤ラッシュなど時間帯や路線によって一定傾向があるとも考えられるため、そのような事前知識がある場合は事前確率  $P(C_{ij}^s), P(\bar{C}_{ij}^s)$  に反映できる．そのような事前知識が得られず混雑傾向が不明の場合は、各事前確率に  $1/2$  を設定する．混雑状況は尤度比と混雑/非混雑閾値  $m_{TH}^C, m_{TH}^{\bar{C}}$  に基づき決定するものとする．すなわち、車両  $v$  に関して  $m^s(v) > m_{TH}^C$  が成り立つ場合は混雑とし、同様に  $m^s(v) < m_{TH}^{\bar{C}}$  が成り立つ場合は非混雑と判定する．

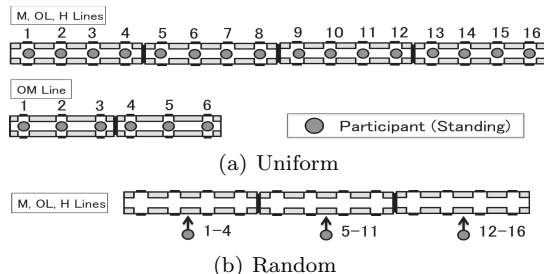
#### 4.3 尤度関数の定義

提案手法で用いる尤度関数  $P(r|S), P(r|\bar{S}), P(r|C), P(r|\bar{C})$  は、各路線で事前に収集した学習データから構築する．これは各路線における車両の構造や材質などから生



(a) 非混雑 (0-60 人/両) (b) 混雑 (61-140 人/両)

図 6: 混雑/非混雑時における車両内状況



(a) Uniform

(b) Random

図 7: 評価実験におけるユーザの配置図

じる Bluetooth の電波特性の違いを尤度関数に反映することにより、精度を向上させるためである。以下では路線 M における尤度関数の構築方法について述べる。

事前学習として、16 台の携帯端末を図 7(a) に示すように網羅的に配置し、各事象  $S, \bar{S}, C$ 、および  $\bar{C}$  が成立する下で観測される Bluetooth サンプルを 1 時間収集した。各事象において観測された RSSI 分布をガウス分布でフィッティングした確率モデルを尤度関数とする。フィッティングパラメータの推定には最尤法を用いる。平均  $\mu$ 、標準偏差  $\sigma$  で表される正規分布の確率密度関数を  $f(x, \mu, \sigma)$  とし、事象  $E$  に関するガウス分布の推定パラメータである平均および標準偏差をそれぞれ  $\mu_E, \sigma_E$  と表すと、事象  $E$  に関する尤度関数を次のように定義できる。

$$P(r|E) = f(r, \mu_E, \sigma_E)$$

以上に基づき構築した各事象の尤度関数およびその推定パラメータの値を図 5 に示す。

## 5. 性能評価

提案手法の評価には路線 M, OL, H に加えて大阪モノレール線 (OM) の計 4 路線において収集した実データを用いる。これらの路線の詳細は表 1 に示す。路線 M, OL, H においては、ユーザ 16 名に携帯端末 Nexus S を 1 台ずつ首から下げてもらい、12 秒毎に Bluetooth 端末検出機構の結果を記録した。ユーザ配置は図 7 に示す Uniform および Random の二種類である。Uniform (図 7(a)) では被験者に乗車中の位置を指定するが、Random (図 7(b)) では乗車車両のみ指定した。一方、路線 OM については、車両間の扉が無い場合の影響を調べるために、ユーザ 6 名で Uniform シナリオ (図 7(a)) によるデータ収集を行った。また、評価における混雑状況の定義として表 1 の座席定員に基づき、対象車両内の乗客が 60 人以上である場合を混雑事象  $C$  とする。混雑/非混雑時における実際の車両内

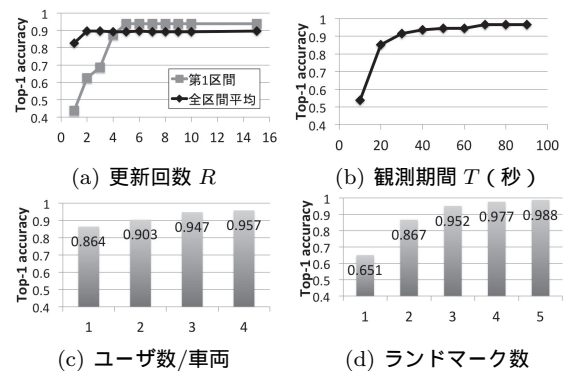


図 8: 各パラメータ変化時における平均乗車車両推定精度の様子を図 6 に示す。パラメータの規定値は、 $R = 10$  回、

$T = 60$  秒、 $m_{TH}^C = m_{TH}^{\bar{C}} = 0.0$  とし、路線 M の学習データを用いて構築された尤度関数を用いる。ランドマークは路線 M, OL, H ではユーザ番号 5 および 12、路線 OM ではユーザ番号 6 とした。特に指定がない限り、以降ではこれらの設定を利用する。

また、乗車車両確率の正しさを評価するため、Top-k 精度と呼ぶ指標を導入する。推定の結果が得られたユーザ  $x$  の乗車車両確率を高い順に並べたときに、 $x$  の正解車両番号が上から第  $k$  位までに含まれていることを “ $k$ -correct” と定義する。Top-k 精度は次の式により算出される。

$$\frac{k\text{-correct であるユーザ数}}{\text{全ユーザ数}}$$

特に指定がない限り、以降の評価では Top-1 精度 (完全に正解したユーザの割合) を用いる。

### 5.1 パラメータや環境の影響

パラメータや環境要因が乗車車両の推定精度に与える影響を調べるため、路線 M における Uniform シナリオのもとで全ユーザの平均乗車車両推定精度を評価した。まず、更新回数  $R$  に関する評価結果を図 8(a) に示す。図 8(a) 中の直線はそれぞれ評価実験における最初の区間のみの推定精度および全区間の平均推定精度を表している。最初の区間では、推定精度が収束するためにおよそ 5 回の更新が必要であることがわかる。これはランドマークが保持する情報が 1 回更新では十分に伝搬していないためである。一方、全区間の平均推定精度は  $R = 2$  のときにほぼ収束している。これは前区間の結果を引き継ぎながら乗車車両確率の更新を行うことから、十分にランドマーク情報が伝搬した後の区間では、少ない更新回数でも高い精度が得られるためである。しかし、実際にはユーザの乗降車が頻繁に起こり、多くのユーザが前区間の情報を保持していない状況も考えられるため、推定精度向上のためには更新ラウンドを区間毎に数回実行する必要がある。

次に、観測期間、参加ユーザ数、ランドマーク数に対する全区間の平均推定精度を図 8(b)-図 8(d) にそれぞれ示す。結果から、それぞれのパラメータが大きいほど平均精度が良いことがわかるが、これは観測期間が長いほど多く



表 2: 各尤度関数による平均乗車車両推定精度

用いる尤度関数	M	OL	H	OM
路線 M	0.957	0.804	0.975	0.791
各路線	-	0.814	0.990	0.875

の RSSI サンプルが得られ、ユーザ数やランドマーク数が多いほど相対位置情報および絶対位置情報が多く得られるので、RSSI の平均化やトポロジ情報による相互補完によってノイズの影響が抑えられるためと考えられる。また、ランドマーク数 1 つの場合と比べ、2 つ以上の場合に推定精度が大きく改善されている。これはランドマークが 1 つしか存在しない場合の乗車車両確率更新において、対称性による不確実性の影響が生じるためと考えられる。

次に、車両構造の違いが推定精度に与える影響を調べるため、路線毎に尤度関数を構築し、比較を行った。路線 M で収集した学習データに基づき構築した尤度関数を利用した場合と、各路線で収集した学習データに基づき構築した尤度関数を利用した場合における平均推定精度を表 2 に示す。全体的に各路線の学習データに基づく尤度関数を用いた場合の推定精度が良いが、路線 OM を除く 3 つの路線では路線 M の学習データに基づく尤度関数を用いても、ほぼ同じ推定精度が得られている。これは各路線の車両が車両間扉を持つなど構造に多くの共通点があるためである。一方、路線 OM は他路線の車両構造とは異なり、車両間扉がない。そのため、路線 M の学習データに基づく尤度関数を用いるよりも路線 OM の学習データに基づく尤度関数を用いる方がより精度が良くなっている。これは、車両間扉がない場合、扉による RSSI の減衰が無く、車両間扉がある路線と尤度関数が大きく異なるためである。以上より、各路線に応じた尤度関数を構築することで高精度な推定が可能と考えられるが、共通の特徴を持つ車両構造であれば同じ尤度関数が利用できる。

### 5.2 ランドマークの影響

提案手法では、ランドマークを基準として各ユーザの乗車車両を推定するため、ランドマークが複数存在する場合に、そのランドマーク組の配置関係が推定精度に影響を与えられられる。そこで、ランドマーク数を 2 とし、その配置関係を同車両組、隣接車両組、2 車両離れた独立組、3 車両離れた独立組とした場合に全ユーザから選出されるランドマーク組の全ての組み合わせについて平均推定精度を算出すると、同車両組の場合から順に 0.736, 0.929, 0.928, 0.875 となった。同一車両内にランドマーク組が存在する場合は、図 8(d) のランドマーク数 1 の場合と同様、対称性による不確実性の影響が生じるため、ランドマーク組が異なる車両に存在する場合と比べ精度が低い。一方、ランドマーク組が 2 両離れて存在する場合、および隣接車両組の場合と比べてランドマーク組が 3 両離れて存在する場合の精度は低下している。これは、ランドマーク

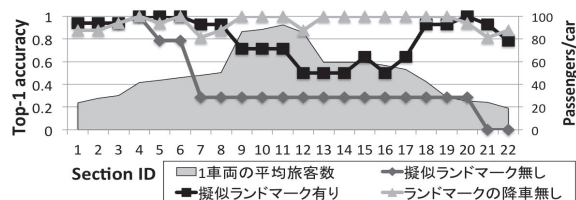
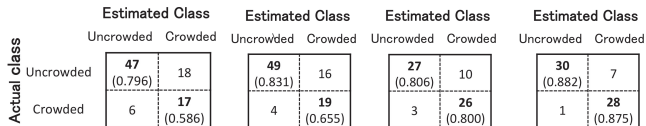


図 9: 擬似 LM の影響 (区間 4 で全 LM 降車)



(a) Uni, 推定 (b) Uni, 理想 (c) Ran, 推定 (d) Ran, 理想

図 10: 混雑推定における混同行列 (括弧内は F 値)

周辺のユーザと比べて遠方のユーザに伝搬されるランドマークの乗車車両確率がより曖昧になるためである。

次に、擬似ランドマークが提案手法の推定精度に与える影響を示す。図 9 は、全ランドマークが降車した際に、いずれかの車両の乗車車両確率が  $l_{TH}$  を超えるユーザを擬似ランドマークとした場合の推定精度を示している。 $l_{TH}$  は経験的に 0.4 とした。擬似ランドマーク無しの場合、各ユーザの乗車車両確率是对等な関係で等しく伝搬され、最終的に全車両の乗車車両確率が同様の値をとるため、推定精度が大幅に低下する。一方、擬似ランドマーク有りの場合は中間区間を除き精度低下が抑制されていることがわかる。中間区間は主要駅が多い混雑した区間であり、人体による遮蔽の影響によって RSSI が弱くなるため、同一車両確率に基づくランドマーク情報伝搬がうまくいかず精度低下を招いていると考えられる。この傾向はランドマークが降車しない場合でも一部見られるが、擬似ランドマーク有りの場合はランドマークが降車しない場合と比べ、精度が全体的に低い。これは、擬似ランドマークの保有する乗車車両確率がランドマークと比べて曖昧なためである。ただし、混雑区間では車両内のユーザ数が増え、新規ランドマークの乗車も十分に期待できるため、実際にはこのような精度低下は軽減されられる。

### 5.3 混雑推定の精度

路線 M における Uniform (Uni) および Random (Ran) シナリオに対する混雑推定の精度評価を行った。図 10 に “区間数 × 車両数” を総数とした推定対象に対する結果を示す。図 10 では車両推定精度の影響を除外した混雑推定精度を調べるため、正しい乗車車両がわかると仮定した場合 (理想値) の混雑推定精度も示している。理想の場合は各ユーザの乗車車両確率について正解車両を 1, 他車両を 0 とした。また、推定対象のうち正しく混雑または非混雑と推定されたものの割合を精度、推定結果のうち実際に混雑または非混雑であったものの割合を再現率と定義し、精度と再現率の調和平均によって定義される値 (F 値) を指標として用いる。

図 10 から、同配置での混雑判定において推定結果を用

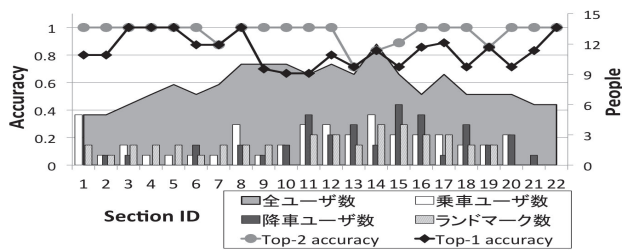


図 11: ユーザ乗降車が与える位置推定精度への影響

いた場合は理想値と比べて少し F 値が低いものの、どちらも同程度の精度が得られている。一方で、Uniform シナリオにおける混雑事象の推定精度は低い。これは Uniform シナリオの下でのユーザ間距離が長いためと考えられる。ユーザ間距離が長いと観測される RSSI が弱くなるため、空いていても空いていると確信するに足る強い観測を得られないことが多く、その結果混雑していると推定する傾向がある。実際、よりユーザが密な位置関係にある Random シナリオでは、混雑事象の推定精度が F 値 0.8 を超えている。現実的なケースでは、混雑に伴ってユーザ数が増加するとともにその位置関係はランダムに分散すると考えられる。したがって、Random シナリオにおける結果から、実世界において提案手法の混雑推定は平均して F 値 0.8 の精度を達成できると考えられる。また、精度および再現率はトレードオフの関係にあることから、提案手法では混雑尤度比に対する閾値  $m_{TH}^C$ ,  $m_{TH}^{\bar{C}}$  を導入し、これらを調整することで要求される精度および再現率が実現できる。

#### 5.4 乗降車の影響

最後に、実環境に近いシナリオを用いて、提案手法の実用性について示す。Random シナリオで収集したデータを基に各駅でユーザの乗降車が発生する擬似的なシナリオを作成し、評価実験を行った。このシナリオでは、主要駅に近づくにつれてユーザ数および乗客が増え、また主要駅ではユーザおよび乗客の乗降車が頻繁に起こる。ランドマークは、ユーザからランダムに 4 つ選出した。

本シナリオにおける各区分でのユーザ数やランドマーク数、ユーザ乗降車数、乗車車両推定精度を図 11 に示す。この結果から、提案手法の精度は全区間平均で 0.83、Top-2 精度では 0.96 であり、ユーザの乗車車両を高精度に推定できていることがわかった。また、ランドマークが存在しない区間でも、擬似ランドマークにより比較的高い精度を維持できていることがわかる。一方、混雑推定精度については混雑/非混雑の判定精度がそれぞれ F 値 0.79, 0.71 であり、平均 F 値は 0.75 であった。以上の結果から、現実的な環境においてもユーザの乗車車両および車両混雑の推定は提案手法によって十分に実現可能と考えられる。

#### 6. まとめ

本稿では、鉄道旅客の保持する携帯端末間で観測された Bluetooth の RSSI およびベイズ推定に基づく継続的な更新アルゴリズムにより、旅客の乗車車両確率および混雑確



図 12: ナビゲーションアプリのプロトタイプ

率を算出することで、各旅客の乗車車両と各車両の混雑状況を高精度に推定する方法の提案を行った。評価では、大阪都市部の鉄道において収集したデータを用い、ユーザ 16 名の乗車車両を 83% の精度で推定し、車両毎の混雑の有無を F 値 0.75 で推定できることがわかった。

現在、提案手法で得られる情報を活用したナビゲーションアプリのプロトタイプ (図 12) を構築している。図 12 のように、提案手法で得られたユーザの乗車車両情報および各車両の混雑状況を用いることで、乗車ユーザに対して降車後のホームにおける最適出口案内や乗車する旅客に対して車両レベルの混雑情報提示が可能になるなど、鉄道移動における高度なナビゲーションへの応用が期待できる。

謝辞 本研究の一部は、KDDI 財団ならび文部科学省国家課題対応型研究開発推進事業-次世代 IT 基盤構築のための研究開発 - 「社会システム・サービスの最適化のための IT 統合システムの構築」(2012 年度 ~2016 年度) の助成を受けたものです。

#### 参考文献

- [1] Biagioni, J., Gerlich, T., Merrifield, T. and Eriksson, J.: EasyTracker: automatic transit tracking, mapping, and arrival time prediction using smartphones, *Proc. of SenSys'11*, pp. 68–81 (2011).
- [2] Arikawa, M., Konomi, S. and Ohnishi, K.: Navitime: supporting pedestrian navigation in the real world, *IEEE Pervasive Computing*, Vol. 6, No. 3, pp. 21–29 (2007).
- [3] Zhou, P., Zheng, Y. and Li, M.: How long to wait?: predicting bus arrival time with mobile phone based participatory sensing, *Proc. of MobiSys'12*, pp. 379–392 (2012).
- [4] Chintalapudi, K., Padmanabha Iyer, A. and Padmanabhan, V. N.: Indoor localization without the pain, *Proc. of MobiCom'10*, pp. 173–184 (2010).
- [5] Teixeira, T., Jung, D. and Savvides, A.: Tasking networked CCTV cameras and mobile phones to identify and localize multiple people, *Proc. of UbiComp'10*, pp. 213–222 (2010).
- [6] Kannan, P. G., Venkatagiri, S. P., Chan, M. C., Ananda, A. L. and Peh, L.-S.: Low cost crowd counting using audio tones, *Proc. of SenSys'12*, pp. 155–168 (2012).
- [7] Weppner, J. and Lukowicz, P.: Bluetooth based collaborative crowd density estimation with mobile phones, *Proc. of PerCom'13*, pp. 192–199 (2013).
- [8] Durrant-Whyte, H. and Bailey, T.: Simultaneous localization and mapping: part I, *Robotics Automation Magazine, IEEE*, Vol. 13, No. 2, pp. 99–110 (2006).