

基地局遷移パターンの学習に基づく 通信履歴からの電車旅客推定法の提案

山田 遊馬¹ 内山 彰¹ 廣森 聡仁¹ 山口 弘純¹ 東野 輝夫¹

概要: 本研究では、携帯電話での通信時に基地局で記録される制御情報の履歴 (通信履歴) を用いて、電車で移動するユーザの推定ならびにその移動経路推定を行う。接続した基地局 ID から分かる大まかな位置情報のみを基にトリップを推定することは困難なため、駅、路線、道路の位置や接続関係といった地理情報を用いる。また、都市部における電波伝搬は複雑なためモデル化が困難であり、接続する基地局の選択基準は単純に距離で決められない。このような複雑な電波伝搬状況を考慮するため、本研究では電車旅客を対象に協力ユーザから得られる GPS トレースと通信履歴を利用して、基地局通信履歴とハンドオーバーの特徴を基地局遷移パターンとして予め学習する。電車旅客の推定では、学習した遷移パターンとの照合を行うことで、尤もらしいトリップを推定する。受信信号強度に基づく基地局選択モデルによるシミュレーションを行った結果、平均通信間隔が 40 秒のデータセットに対して、電車トリップの推定において再現率 90.2%、精度 78.5%、F 値 83.9%を達成できた。

Proposal of Train Travel Estimation Using Control Signal Records Based on Base Station Transition Pattern Learning

YUMA YAMADA¹ AKIRA UCHIYAMA¹ AKIHITO HIROMORI¹ HIROZUMI YAMAGUCHI¹
TERUO HIGASHINO¹

1. はじめに

首都圏などの広範囲にわたる人々の動きの把握はマーケティング、都市開発、防災・減災など多方面で活用されている。人々の動きを把握するため、約 10 年に 1 度の頻度で主要都市において、人々の動きを調べ交通機関の実態を把握するパーソントリップ調査 (PT 調査) が行われている。しかし、PT 調査は実施のための労力が大きく、時間帯や曜日、季節に応じて時々刻々と変化する人々の動きを把握することは難しい。近年では交通系 IC カードの利用履歴を用いることで、詳細な PT 調査を行うことも可能となっている。しかし、人は鉄道だけでなく、徒歩、自転車、自動車など様々な手段により移動するため、都市全体の人々の動きを把握するには不十分である。そのため、これまでに GPS (Global Positioning System) により得られる位置情

報から、人々の動きを把握する手法が提案されている [1]。しかし、位置情報を取得するためには位置情報収集のためのアプリケーションが携帯電話に導入されている必要があり、対応機種や年齢層といったユーザの網羅性の観点から課題が残されている。また、これまでに携帯電話の通話履歴 (Call Detail Record; CDR) を用いた研究がされてきたが、通話はそれほど頻繁に行われなため、位置情報を得られる頻度が低いという問題点がある [2]。

そこで、本研究では携帯電話が基地局と通信を行ったときの制御情報の履歴 (通信履歴) を利用する。基地局との通信履歴は、対象キャリアの携帯電話を保持している全ユーザに対し通信した基地局 ID と時刻が得られるため、網羅率が高いという利点がある。通信履歴はアプリケーションの利用状況によって記録頻度が変化する。また、通信履歴では通信時に接続した基地局の ID しか分からないため、位置の精度は基地局のセルサイズに依存し、Wi-Fi や GPS と比較して位置誤差が大きい。通信履歴を用いた移動経路の推定を行う手法も存在するが、日常習慣の移動を対象に

¹ 大阪大学大学院 情報科学研究科
Graduate School of Information Science and Technology, Osaka University

しており、普段と異なる突発的な移動パターンの検出が難しい [3]。このため、我々はこれまでに通信履歴を用いて、駅、路線、道路の地理的な特徴と、移動速度や時刻表から推測される電車の位置といった時間的な特徴を利用し、高精度な推定を実現させる手法を考案している [4]。しかし、文献 [4] の手法では現実的な電波伝搬が考慮されておらず、建物などの遮蔽物が乱立する都市部では反射や回折などの影響が避けられないため、実環境での適用には依然として課題が残されていた。

本研究では、建物などの遮蔽物による電波伝搬を考慮するために、トリップ属性ごとに基地局の通信履歴とハンドオーバーの特徴を予め学習しておき、その学習した情報と推定対象の基地局通信履歴を照合することでトリップ推定を行う。ここでトリップ属性とは、移動手段だけではなく、どの地点からどの地点へ移動したのかといった詳細なトリップの特徴のことを言う。基地局の通信履歴とハンドオーバーの特徴を学習するにあたって、事前に一部の協力ユーザから GPS トレースが得られることを想定する。そして、GPS トレースから移動手段が正確に把握できると仮定し、GPS 提供ユーザの基地局通信履歴を教師データとする。この教師データは実際のハンドオーバーによって生成された基地局通信履歴から生成されているため、建物などの遮蔽物を考慮した電波伝搬状況を反映している。この教師データを用いて、電波伝搬を考慮したパーソントリップ推定法を提案する。電車旅客は乗降車する駅が同じであれば、その経路はほぼ同じであるが、自動車旅客は同一地点間の移動においても、経路が無数に考えられるため、自動車旅客の推定をするためには膨大な教師データが必要であると考えられる。そのため、本研究では電車旅客のみに着目し、推定対象のユーザが電車を利用したか否かに加えて、電車を乗降車した駅まで推定することを目的とする。

提案手法の有効性を評価するために、ネットワークシミュレータ Scenargie[5] を用いて、携帯電話ユーザが様々な交通手段を利用して目的地へ向かうシナリオを再現し、受信信号強度を基にしたハンドオーバーモデルによる性能評価を行った。その結果、平均通信間隔が 40 秒のデータセットに対して、再現率 90.2%、精度 78.5%、F 値 83.9% を達成した。

2. 関連研究

車両に搭載されている GPS や速度計などのセンサに加えて、無線通信機器を搭載し、位置情報や速度情報などの様々な情報を収集するプローブカーを活用して交通量を調査する取り組みもなされている [6], [7], [8]。文献 [9] は、600,000 台以上の車両から収集された 3 分間隔のプローブカーデータから、ローマの環状高速道路における交通状況の推定と予測を行う手法を提案している。同手法では、ニューラルネットワークを利用したパターンマッチングに

基づき、30 分後の平均速度を 3.5~9.5[km/h] の誤差で予測できている。文献 [10] は、各道路リンクにおけるプローブカーの通行車両数から交通量を推定する手法を提案している。この手法においては、全車両のうち 4 割程度の車両からプローブカーデータを収集することができれば、20% 程度の誤差で交通量を推定できることが示されている。

人の動きを把握する手法としては、PT 調査データを利用する手法が挙げられ、様々な取り組みがなされている。文献 [11] は、PT 調査データに時空間補正方法を組み合わせた手法を提案している。PT 調査はアンケートベースの統計であるため、出発時刻が 30 分区分切りとなることが多く、不自然なトリップデータが得られるという問題がある。この手法ではカーネル密度推定を用いて出発時刻を平滑化することにより、より自然なトリップデータを生成し、最短経路に基づき 1 分毎の位置を予測している。

また、タクシートのトレース情報を利用した取り組みもある。文献 [12] は、タクシー移動の遠回り度合いを表すストレッチ係数を新たに定義し、人がどこでタクシーに乗車し、どこで降車したかを解析する手法を提案している。この手法では、隠れマルコフモデルに基づくアルゴリズムを適用することで 90~94% の精度と再現率を達成している。

一方、GPS などのビッグデータを利用した取り組みもなされている [1], [13]。文献 [1] は、92 億レコードを超える GPS データを解析し、東日本大震災後の避難者の滞在位置を推定する手法を提案しており、東日本大震災後、1 ヶ月間の避難者の動きを誤差 5% 以下の精度で推定している。文献 [13] は、コンサートや野球などの人気のあるイベントと人の動きの関係性を解析する手法を提案している。このような解析結果により、イベントのジャンルと人の居住地域の相関性を見出し、交通輸送の管理などへの応用が可能である。

モバイル空間統計では、携帯電話から定期的に送信される位置情報から 250m メッシュレベルでの時刻ごとの人口推定を行っている [14]。Twitter メッセージを解析することでユーザの位置やイベントを推定する取り組みも行われている。文献 [15] は、Twitter 解析から多数の人々が訪れるイベントの解析を行っている。文献 [16] は、ユーザ同士の関係性を利用して、つぶやきを解析し、位置推定を行っている。

また、携帯電話の通話履歴を利用した取り組みもある [2], [17]。文献 [2] は、人の動きが社会的な関係性に密接に相関していることに着目し、携帯電話の通話履歴 (CDR) を用いてユーザの位置を推定する手法を提案している。さらに、クラウドベースの移動予測システム NextMe を導入することで、人の動きの予測に成功している。文献 [17] は、携帯電話の通話履歴を用いて、ユーザが乗車している電車を推定する手法を提案している。基地局のセルサイズを仮定することで、20 分間隔で発着する路線において乗車した

表 1: 通信履歴テーブルの例

端末 ID	タイムスタンプ	基地局 ID
123456	2016-04-01 07:00:10	AABBAA
090909	2016-04-01 07:00:14	DDEEEE
123456	2016-04-01 07:01:45	CCCCC
90abcd	2016-04-01 07:02:22	AABBAA

表 2: 基地局テーブルの例

基地局 ID	緯度, 経度
AABBAA	(135.012, 34.123)
DDEEEE	(135.678, 34.888)
CCCCC	(136.123, 34.666)

電車を高精度に推定している。

これらの手法に対し提案手法は携帯電話の運用上、インフラ側で記録される制御信号を用いて、人々の移動手段および経路を推定するという点が異なる。また、同様の制御信号を用いた手法として、日常習慣を対象としたトリップ推定手法が提案されている [3].

3. 想定環境

3.1 通信履歴の特性

通信履歴はユーザが通信を行った際に携帯電話網の機器間で発生する制御情報の履歴である。通信履歴が記録されるタイミングには (1) アプリケーションによる通信を行った場合、(2) 異なる位置登録エリアに移動した場合、(3) 通信履歴が無い状態で長時間が経過した場合、の 3 種類が存在する。位置登録エリアとは、LTE ネットワークにおいて携帯電話が存在する領域を管理する単位領域であり、通常は複数の基地局をまとめた領域として定義される。また、スマートフォン端末には Wi-Fi の電波状況が悪い時に自動的に LTE 回線で通信する機能があり、Wi-Fi のアクセスポイントが多い都市圏でも一定の LTE 通信が期待できるものとする。

実際の通信履歴を用いた既存研究 [3], [18] では、ユーザ端末 ID、通信時刻、基地局 ID の 3 つが基地局側で記録されることが報告されている。したがって、本研究でも同じデータが記録されるものとする。また、基地局 ID ごとの位置情報 (緯度, 経度) は既知なので、基地局 ID と紐付けることで、ユーザがどの位置の基地局と通信したかを特定できる。表 1, 2 に通信履歴と基地局のデータセットの例を示す。通信履歴は携帯電話網の制御情報の履歴であるため、位置情報としては通信した基地局しか分からない。このため、位置の精度は基地局のセルサイズに依存し、その誤差は 100m 以上に及ぶ場合もある [19]。これは誤差が 10m 程度である GPS と比較すると非常に大きいといえる [20].

3.2 通信基地局選択モデル

接続対象となる基地局は受信信号強度に応じて決まるが、

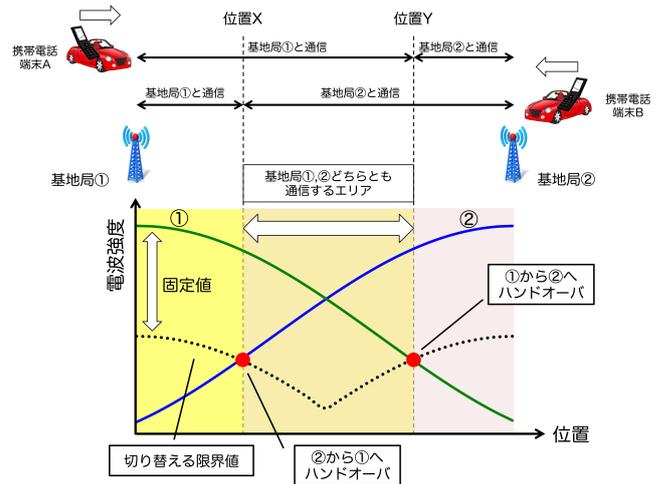


図 1: ハンドオーバーの一例

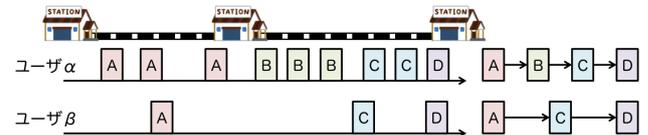


図 2: 通信頻度が異なるユーザの通信履歴の例

ハンドオーバーや電波の揺らぎによって、必ずしも最も近い基地局に接続されるとは限らない。文献 [21] では、ハンドオーバー時の基地局切り替えの詳細が述べられている。ハンドオーバー時には、まず複数の無線基地局から届く電波のうち最も強い受信信号強度を基準値とする。そして、その基準値から各携帯電話事業者が定める固定値を引いたレベルを限界値とし、現在通信中の基地局の受信信号強度がこの限界値を下回れば、その基地局との通信は切断し、その時点の最も受信信号強度が強い基地局にハンドオーバーする。ゆえに、図 1 に示すように、同一地点でも通信する基地局が異なる場合があり得る。また、図 1 では基地局からの距離に応じて電波強度が変化しているが、実際にはビルなどの影響によって遠方でも強い電波強度になるといった複雑な電波伝搬状況となる。ユーザのトリップ推定にあたっては、これらを考慮することが必要不可欠である。

4. 提案手法

4.1 概要

図 2 のように、ユーザによって通信頻度が大きく異なる。そのため、単純に通信履歴上で隣り合う基地局の遷移パターンからトリップを推定すると、考慮すべき遷移パターンの組み合わせ数が膨大になるため推定に適さない。そこで、本研究では通信間隔に基づいた推定を行うことで、ユーザごとに異なる通信頻度の問題を解消している。ここで通信間隔とは、ユーザの通信履歴の異なる 2 つの通信において、一方の通信を行ってから他方の通信を行うまでの時間を言う。

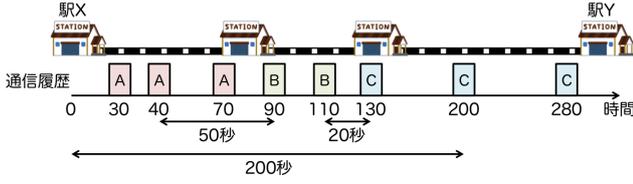


図 3: GPS ユーザから得られる教師データの例

図 3 に示すような駅 X から駅 Y へ移動したユーザの通信履歴が教師データとして与えられた場合を考える。この教師データより、駅 X から駅 Y へ移動するにあたって、移動開始してから 200 秒後に基地局 C と通信することや、基地局 A と通信してから 50 秒後に基地局 B と通信するといった知見が得られる。このような異なる 2 つの基地局の通信間隔を学習し、トリップ推定に利用する。教師データから得られた学習結果を用いて、トリップの尤度を決定し、トリップ候補の抽出を行う。最後に、得られたトリップ候補それぞれの尤度に従って適切なトリップを選択することで、電車旅客の推定および移動経路の推定を行う。以降の節では、詳細な学習アルゴリズムおよび推定アルゴリズムについて述べる。

4.2 基地局遷移パターンの学習

4.1 節で述べたように、教師データから異なる 2 つの基地局において発生し得る通信間隔を学習した結果を学習行列 Mat とする。学習行列はトリップ属性ごとに定義され、駅 i から駅 j へのトリップの学習行列を $Mat^{(i \rightarrow j)}$ とする。また、学習行列 $Mat^{(i \rightarrow j)}$ の各要素 $Mat^{(i \rightarrow j)}(m, n)$ には基地局 m の後に基地局 n と通信する場合の通信間隔の学習結果が格納されており、要素は式 (1) で表される通信間隔の最小値と最大値からなる区間である。

$$Mat^{(i \rightarrow j)}(m, n) = \left[Mat_{\min}^{(i \rightarrow j)}(m, n), Mat_{\max}^{(i \rightarrow j)}(m, n) \right] \quad (1)$$

学習に用いる教師ユーザ群を \mathcal{TU} 、ユーザ u のトリップの集合を $\mathcal{T}_r^{(u)}$ 、トリップ T_r の出発駅、終着駅をそれぞれ $st_{orig}(T_r), st_{dest}(T_r)$ 、出発時刻、到着時刻をそれぞれ $t_{orig}(T_r), t_{dest}(T_r)$ と定義する。すると、駅 i から駅 j へのトリップの教師データ $\mathcal{TD}^{(i \rightarrow j)}$ は式 (2) のように定義できる。

$$\mathcal{TD}^{(i \rightarrow j)} = \{C_u^{(t_{orig}(T_r), t_{dest}(T_r))} \mid st_{orig}(T_r) = i \wedge st_{dest}(T_r) = j, T_r \in \mathcal{T}_r^{(u)}, u \in \mathcal{TU}\} \quad (2)$$

ここで C_u はユーザ u の通信履歴の系列、 $C_u^{(t_1, t_2)}$ は時刻 t_1 から t_2 の間に発生した通信履歴の系列を表しており式 (3) のように定義される。また、 $t(c)$ は通信 c が発生した時刻を表している。

$$C_u^{(t_1, t_2)} = \{c \mid t_1 \leq t(c) \leq t_2, c \in C_u\} \quad (3)$$

そして、駅 i から駅 j へ移動するトリップの学習行列

Algorithm 1 Learning Trips

Input: $\mathcal{TD}^{(i \rightarrow j)}$
Output: $Mat^{(i \rightarrow j)}$

```

/* initialization */
while  $Mat_{\min}^{(i \rightarrow j)}(m, n) \in Mat_{\min}^{(i \rightarrow j)}$  do
   $Mat_{\min}^{(i \rightarrow j)}(m, n) \leftarrow +\infty$ 
end while
while  $Mat_{\max}^{(i \rightarrow j)}(m, n) \in Mat_{\max}^{(i \rightarrow j)}$  do
   $Mat_{\max}^{(i \rightarrow j)}(m, n) \leftarrow -\infty$ 
end while
/* learning */
while  $C \in \mathcal{TD}^{(i \rightarrow j)}$  do
  while  $c_1 \in C$  do
     $m \leftarrow BS(c_1)$ 
    while  $c_2 \in C$  do
      if  $BS(c_1) \neq BS(c_2) \wedge t(c_1, c_2) > 0$  then
         $n \leftarrow BS(c_2)$ 
         $Mat_{\min}^{(i \rightarrow j)}(m, n) \leftarrow \min(t(c_1, c_2), Mat_{\min}^{(i \rightarrow j)}(m, n))$ 
         $Mat_{\max}^{(i \rightarrow j)}(m, n) \leftarrow \max(t(c_1, c_2), Mat_{\max}^{(i \rightarrow j)}(m, n))$ 
      end if
    end while
  end while
   $Mat_{\min}^{(i \rightarrow j)}(O, m) \leftarrow \min(t(c_1) - t_{orig}, Mat_{\min}^{(i \rightarrow j)}(O, m))$ 
   $Mat_{\max}^{(i \rightarrow j)}(O, m) \leftarrow \max(t(c_1) - t_{orig}, Mat_{\max}^{(i \rightarrow j)}(O, m))$ 
   $Mat_{\min}^{(i \rightarrow j)}(m, D) \leftarrow \min(t_{dest} - t(c_1), Mat_{\min}^{(i \rightarrow j)}(m, D))$ 
   $Mat_{\max}^{(i \rightarrow j)}(m, D) \leftarrow \max(t_{dest} - t(c_1), Mat_{\max}^{(i \rightarrow j)}(m, D))$ 
end while
end while

```

$Mat^{(i \rightarrow j)}$ を教師データ $\mathcal{TD}^{(i \rightarrow j)}$ から学習する。与えられた教師データ $\mathcal{TD}^{(i \rightarrow j)}$ に属する通信履歴 C で発生した異なる 2 つの通信 c_1, c_2 ($t(c_1) < t(c_2)$) の組から基地局 $BS(c_1)$ と通信をしてから基地局 $BS(c_2)$ を通信するまでに要する時間を学習し、学習区間 $Mat^{(i \rightarrow j)}(BS(c_1), BS(c_2))$ を更新する。ゆえに、学習区間 $Mat^{(i \rightarrow j)}(m, n)$ には、教師データから得られた基地局 m と通信をしてから、基地局 n と通信するまでに要する最短時間と最長時間が記録されることとなる。詳細な学習アルゴリズムを Algorithm 1 に示す。ここで、 $t(c_1, c_2)$ は通信 c_1 が発生してから通信 c_2 が発生するまでの時間を表している。また、移動開始から任意の基地局 m との通信が発生するまでの時間や、任意の基地局 m との通信が発生してから移動終了までの時間も学習しており、その学習行列はそれぞれ $Mat^{(i \rightarrow j)}(O, m)$ 、 $Mat^{(i \rightarrow j)}(m, D)$ で表す。

4.3 電車トリップ候補の抽出

4.2 節で作成した学習行列を基にユーザ u が時刻 t を出発時刻として、駅 i から駅 j へのトリップを行ったと仮定したときの尤度 $l_{u,t}^{(i \rightarrow j)}$ を算出する。学習行列の要素 $Mat^{(i \rightarrow j)}(m, n)$ には基地局 m から基地局 n への通信間隔の区間が記録されている。推定対象ユーザ u の通信履歴において基地局 m と基地局 n の通信間隔は、必ずしも学習区間が実際の基地局の通信間隔と一致するわけではない。このため、学習した区間に対して、一定の幅を許容するよ

Algorithm 2 Calculate Likelihood

Input: $C_u^{(t,t+\Delta T)}$, $Mat^{(i \rightarrow j)}$
Output: $l_{u,t}^{(i \rightarrow j)}$

```

/* initialization */
 $l_{u,t}^{(i \rightarrow j)} \leftarrow 1.0$ 
/* calculate likelihood */
while  $c_1 \in C_u^{(t,t+\Delta T)}$  do
   $m \leftarrow BS(c_1)$ 
  while  $c_2 \in C_u^{(t,t+\Delta T)}$  do
    if  $BS(c_1) \neq BS(c_2) \wedge t(c_1, c_2) > 0$  then
       $n \leftarrow BS(c_2)$ 
       $l_{u,t}^{(i \rightarrow j)} \leftarrow l_{u,t}^{(i \rightarrow j)} \times p(t(c_1, c_2), Mat^{(i \rightarrow j)}(m, n))$ 
    end if
  end while
   $l_{u,t}^{(i \rightarrow j)} \leftarrow l_{u,t}^{(i \rightarrow j)} \times p(t(c_1) - t, Mat^{(i \rightarrow j)}(O, m))$ 
   $l_{u,t}^{(i \rightarrow j)} \leftarrow l_{u,t}^{(i \rightarrow j)} \times p((t + \Delta T) - t(c_1), Mat^{(i \rightarrow j)}(m, D))$ 
end while

```

うに尤度を定義する必要がある。

このことを踏まえて、尤度 $l_{u,t}^{(i \rightarrow j)}$ を Algorithm 2 のようにして算出する。Algorithm 2 中に登場する関数 p は、通信間隔 t と学習区間 m から尤度を算出する関数であり、式 (4) で定義される。関数 p は図 4 のように、通信間隔 t が学習区間の中央値であれば 1 を、学習区間の両端であれば $\exp(-\frac{1}{2})$ を返し、学習区間から遠く離れるほど小さい値が返るようにした。

それぞれの通信組において、関数 p から通信間隔の尤度を算出し、その総乗が閾値以上であれば、出発時刻を t とする駅 i から駅 j へのトリップとして、ユーザ u のトリップ候補 $\mathcal{TC}^{(u)}$ に追加する。しかし、Algorithm 2 では、検証区間 $[t, t + \Delta T]$ の間に 1 度も通信が行われなかった場合や、同一の基地局のみとしか通信を行わなかった場合において、尤度 $l_{u,t}^{(i \rightarrow j)}$ が 1 となってしまう。これはトリップ候補の抽出において適切でないと考えられるため、基地局通信が行われなかった場合や同一の基地局のみとしか通信をしなかった場合は、トリップ候補に追加しないこととする。ここで、検証区間 $[t, t + \Delta T]$ はトリップであるか否かを推定する区間、 ΔT はトリップごとに異なる学習で得られたトリップに要する時間を表している。

$$p(t, m) = \exp\left(-\frac{(t - \frac{m_{\min} + m_{\max}}{2})^2}{2 \cdot (\frac{m_{\max} - m_{\min}}{2})^2}\right) \quad (4)$$

前述のトリップと判定するための尤度閾値は $\{\exp(-\frac{1}{2})\}^N$ で定義する。ここで、 N は推定対象ユーザの検証区間 $[t, t + \Delta T]$ 中に発生した通信の組み合わせ数を表す。この閾値は検証区間 $[t, t + \Delta T]$ の全ての通信の組において、通信間隔が学習区間内であれば必ず閾値以上になるように定義されている。また、通信間隔が学習区間外である通信の組があった場合でも、他の通信の組が学習区間の中央値に近ければ尤度が閾値以上となるため、トリップの検出漏れを防ぐことができる。

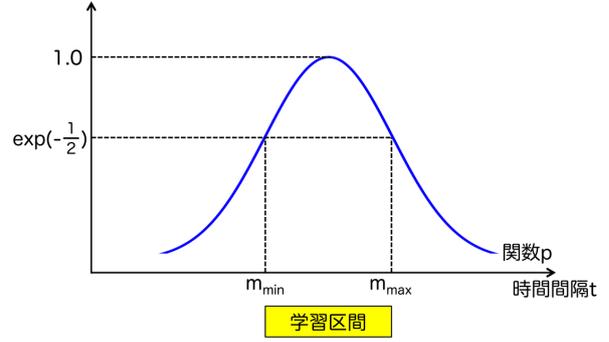


図 4: 学習区間と尤度関数 p の関係

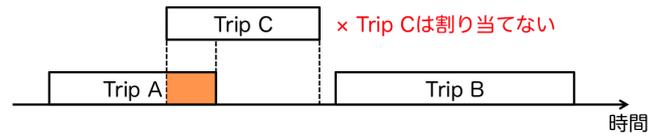


図 5: トリップ候補から 1 日のトリップ推定を行う例

4.4 トリップ推定

この節では 4.3 節で検出したトリップ候補 $\mathcal{TC}^{(u)}$ から、ユーザの全トリップを推定する手法を述べる。ユーザの基地局遷移パターンによっては、同一時間に複数のトリップ候補が混在することが考えられる。しかし、ユーザは同一時間に複数のトリップを行うとは考えられないため、トリップ候補の尤度に応じたトリップ推定を行う。トリップ候補のうち、尤度が最も高いトリップからユーザ u のトリップとして順に割り当てていく。そして、図 5 のように、割り当て区間に既に他のトリップが存在していれば、割り当てを行わずにそのトリップ候補を除外する。このように、尤度の高いトリップ候補から順に割り当て、ユーザの全トリップを推定する。

5. 性能評価

5.1 評価環境

ネットワークシミュレータ Scenargie を用いて、ユーザの移動と携帯電話による通信を再現し、提案手法の性能評価を行った。

シミュレーション領域は図 6 に示す $4\text{km} \times 4\text{km}$ の領域であり、29 個の基地局 (図 6 の青色) をランダムに配置した。このような環境で 2 時間のシミュレーションを行った。ユーザの通信頻度は指数分布に従うものとし、通信時に接続する基地局は 3.2 節で述べた文献 [21] の受信信号強度に基づいたハンドオーバーモデルに従って決定される。なお、ビルの高さは 20m、基地局の高さは 22m、電波を受信する携帯電話の高さは 1.5m とした。

図 6 にシミュレーションに用いた領域を示す。各ユーザはビル内 (図 6 の黄色) にランダムに配置され、シミュレーション開始と同時に、ランダムに選択された開始地点とは異なるビルへ向かう。ユーザのうち 50% は自動車を所持し

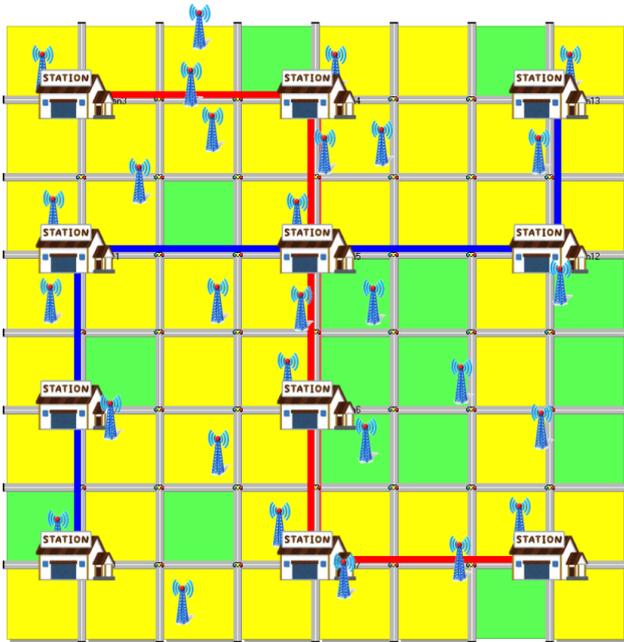


図 6: シミュレーション環境

ており、自動車を所持しているユーザは目的地へ自動車に向かう。残りのユーザは電車または徒歩で目的地まで移動する。また、駅は全部で11駅あり、赤線と青線の2つの路線が交差するような構造となっている。駅間距離は1.0～1.5kmであり、電車が駅間を移動する時間は80～110秒に設定し平均速度が約45[km/h]となるようにした。さらに、各路線上下線ともに電車が6分間隔で発着するような時刻表を設定し、都市部の過密なダイヤグラムを再現した。図6の緑色は公園であり、遮蔽物がないエリアである。ユーザ数は1,000人として、ユーザの移動経路はScenargieのMulti Agent モジュールを使って再現した。

通信が発生してから次の通信が発生するまでの通信頻度 T は、式(5)に示すような指数分布とする。 α は通信頻度の期待値、 $x(0 < x < 1)$ は一様乱数値を表している。性能評価では通信頻度の期待値 α を変化させ、通信頻度に応じた精度の差を検証した。

$$T = -\alpha \ln x \quad (5)$$

5.2 通信頻度による電車トリップの推定精度

通信頻度による推定精度の違いを検証するために、通信頻度 α を40秒、60秒、120秒の3通りについて性能評価を行った。5.1節で述べたシミュレーションを5回行い、そのうちの4回分を教師データ、1回分をテストデータによる5分割交差検証を行った。また、距離に応じた基地局選択モデルに基づく既存手法[4]を用いて比較を行った。

電車旅客のうち乗降車した駅まで推定ができた場合を正解、電車旅客ではないが電車旅客であると推定した場合を誤検出として、混同行列に評価結果を示す。まず、

表 3: 提案手法の推定結果

(a) $\alpha = 40$

		予測	
		Positive	Negative
正解	Y	655	71
	N	179	4095

(b) $\alpha = 60$

		予測	
		Positive	Negative
正解	Y	597	129
	N	184	4090

(c) $\alpha = 120$

		予測	
		Positive	Negative
正解	Y	465	261
	N	160	4114

$\alpha = 40, 60, 120$ 秒としたときの提案手法の評価結果を表3の混同行列および図7に示す。どの結果においても90%以上の正解率を達成し、 $\alpha = 40$ 秒の正解率は95.0%にもなった。 $\alpha = 40$ 秒においては再現率90.2%、精度78.5%となり、F値は83.9%を達成し、非常に高精度に推定できていることが分かる。また、図7より、 α が大きくなるにつれて再現率が大きく低下しており、得られる通信履歴数が少なくなるとトリップの検出漏れが多くなることがこの結果からいえる。

同様に既存手法[4]の評価結果を表4に示す。 $\alpha = 40$ 秒の場合では再現率55.5%、精度86.4%となり、F値は67.6%となった。この結果からわかるように既存手法では検出漏れが非常に多くなっていることが分かる。既存手法では通過する駅付近の基地局と必ず通信しておく必要があるため、検出漏れが多くなったと考えられる。提案手法では、教師データで得られた学習区間に幅をもたせることで電車旅客の検出漏れを防いでいるため、高い再現率を達成できているといえる。

5.3 教師データ量による電車トリップの推定精度

教師データ量による推定精度の差を検証するために、教師データとして用いるシミュレーションのデータセット数を1～4セットに変化させ推定精度および再現率を評価した。なお、通信頻度は $\alpha = 60$ 秒とし、データ1セットあたりの各トリップ属性の教師データ数は約5件である。データセット数ごとの精度、再現率、F値の関係を図8に示す。図8より、データセット数が少なくなると精度が上昇する反面、再現率が低下していることが分かる。これは、教師データ数が多くなるにつれて通信間隔の学習区間が広

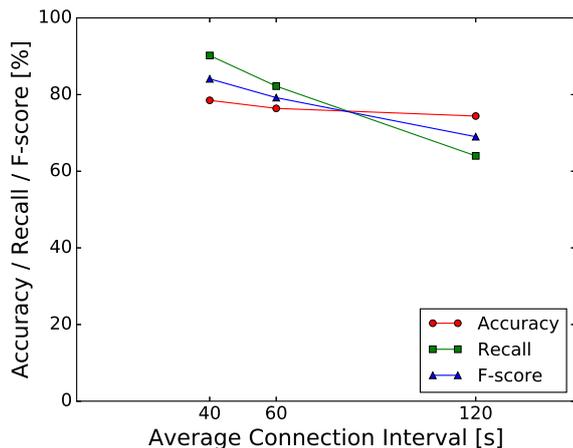


図 7: 通信頻度と推定精度の関係

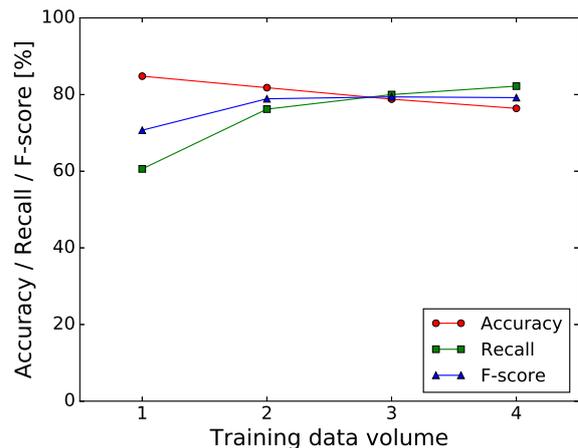


図 8: 教師データ量と推定精度の関係

表 4: 既存手法 [4] の推定結果

(a) $\alpha = 40$

		予測	
		Positive	Negative
正解	Y	403	323
	N	63	4211

(b) $\alpha = 60$

		予測	
		Positive	Negative
正解	Y	366	360
	N	51	4233

(c) $\alpha = 120$

		予測	
		Positive	Negative
正解	Y	232	494
	N	31	4243

くなり、電車旅客の検出漏れが防げているからであるといえる。しかし、通信間隔の学習区間が広がることで誤検出も多くなり、精度が低下している。だが、教師データ量の増加とともに F 値も増加傾向にあるため、精度、再現率の両方を同一のバランスで評価した場合、教師データ量が多い方が良い推定ができています。

データセット数が 1 セットの場合でも F 値が 70.7% を達成していることから、ごくわずかな GPS トレースを用いて学習を行ったとしても、一定水準の電車旅客推定が可能であるといえる。

6. まとめ

本研究では、網羅性の高い携帯電話の基地局通信履歴を用いて、電車で移動するユーザの推定ならびにその移動経路の推定手法を提案した。複雑な電波伝搬状況を考慮する

ために、協力ユーザから得られる GPS トレースと通信履歴を利用し、トリップ属性ごとに基地局通信履歴とハンドオーバーの特徴を予め学習させることで、電波伝搬を考慮した推定を行った。

シミュレーションによる評価の結果、ビルが乱立し電波状況の複雑な都市部においても、ユーザの平均通信間隔が 40 秒の場合で F 値 84.1% を達成した。また、教師データ量による性能評価を行った結果、各トリップ属性の教師データ数が約 5 件といった少ない教師データからでも、平均通信間隔が 60 秒の場合で F 値 71.7% を達成し、一定水準の推定が可能であることが示された。

今後の課題として、自動車旅客の推定手法の検討が挙げられる。自動車によるトリップは出発地点と到着地点が同一であっても、その移動経路は無数に考えられるため、本手法を適用するには膨大な教師データが必要であると考えられる。学習に必要な教師データ数を削減するため、頻りに利用されるような主要道路や最短経路といった自然な行動に関する事前知識の併用を検討している。

参考文献

- [1] Horanont, T., Witayangkurn, A., Sekimoto, Y. and Shibasaki, R.: Large-Scale Auto-GPS Analysis for Discerning Behavior Change during Crisis, *IEEE Intelligent Systems*, Vol. 28, No. 4, pp. 26–34 (2013).
- [2] Zhang, D., Chen, M., Guizani, M. and Xiong, H.: Mobility prediction in telecom cloud using mobile calls, *IEEE Wireless Communications*, Vol. 21, No. 1, pp. 26–32 (2014).
- [3] Kanasugi, H., Sekimoto, Y., Kurokawa, M., Watanabe, T., Muramatsu, S. and Shibasaki, R.: Spatiotemporal route estimation consistent with human mobility using cellular network data, *Proceedings of 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 267–272 (2013).
- [4] 山田遊馬, 内山彰, 廣森聡仁, 山口弘純, 東野輝夫: 基地局通信履歴と地理情報を用いたパーソントリップ推定法の提案, マルチメディア通信と分散ワークショップ

- 2015 論文集, pp. 128–135 (2015).
- [5] Space-Time Engineering, LLC: Scenargie, <http://www.spacetime-eng.com/> (accessed February 10, 2015).
- [6] 堀場庸介, 松本幸正, 松井 寛, 高橋政稔: プローブデータに基づく推定経路交通量への観測誤差の影響分析と推定経路交通量の更新手法, 土木計画学研究・論文集, Vol. 22, No. 3, pp. 495–505 (2005).
- [7] 三輪富生, 山本俊行, 竹下知範, 森川高行: プローブカーの速度情報を用いた動的 OD 交通量の推定可能性に関する研究, 土木学会論文集 D, Vol. 64, No. 2, pp. 252–265 (2008).
- [8] Toledo, T. and Kolehkina, T.: Estimation of dynamic origin - destination matrices using linear assignment matrix approximations, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 2, pp. 618–626 (2013).
- [9] de Fabritiis, C., Ragona, R. and Valenti, G.: Traffic estimation and prediction based on real time floating car data, *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems (ITSC 2008)*, pp. 197–203 (2008).
- [10] Hellinga, B. R.: Estimating dynamic origin-destination demands from link and probe counts, PhD Thesis, Queen's University (1994).
- [11] Sekimoto, Y., Watanabe, A., Nakamura, T., Kanasugi, H. and Usui, T.: Combination of spatio-temporal correction methods using traffic survey data for reconstruction of people flow, *Pervasive and Mobile Computing*, Vol. 9, No. 5, pp. 629–642 (2013).
- [12] Ganti, R., Srivatsa, M., Ranganathan, A. and Han, J.: Inferring human mobility patterns from taxicab location traces, *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 459–468 (2013).
- [13] Calabrese, F., Pereira, F. C., Di Lorenzo, G., Liu, L. and Ratti, C.: The geography of taste: analyzing cell-phone mobility and social events, *Proceedings of IEEE Pervasive Computing*, pp. 22–37 (2010).
- [14] Terada, M. and Nagata, Tomohiro, K. M.: Population Estimation Technology for Mobile Spatial Statistics, *NTT DOCOMO Technical Journal*, Vol. 20, No. 3, pp. 10–15 (2012).
- [15] Yano, Y., Hashiyama, T., Ichino, J. and Tano, S.: Behavior extraction from tweets using character N-gram models, *Proceedings of 2014 IEEE international conference on Fuzzy Systems*, pp. 1273–1280 (2014).
- [16] Davis Jr, C. A., Pappa, G. L., de Oliveira, D. R. R. and de L. Arcanjo, F.: Inferring the location of twitter messages based on user relationships, *Transactions in GIS*, Vol. 15, No. 6, pp. 735–751 (2011).
- [17] Kanno, T., Kanasugi, H., Sekimoto, Y. and Shibasaki, R.: Real-time passenger location estimation using CDRs and train objects generated from crowdsourced timetables, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pp. 1197–1205 (2015).
- [18] Bayir, M. A., Demirbas, M. and Eagle, N.: Mobility profiler: A framework for discovering mobility profiles of cell phone users, *Pervasive and Mobile Computing*, Vol. 6, No. 4, pp. 435 – 454 (2010).
- [19] 蓑輪 正, 三浦 周: 周波数共用方式の課題と有効性について, 情報通信研究機構研究報告, Vol.61, No.1, p. 10 (2015).
- [20] Wing, M. G., Eklund, A. and Kellogg, L. D.: Consumer-grade Global Positioning System (GPS) Accuracy and Reliability, *Journal of Forestry*, Vol. 103, No. 4, pp. 169–173 (2005).
- [21] 中嶋信生, 有田武美, 樋口健一: 携帯電話はなぜつながるのか 第2版, 日経 BP 社 (2012).