

携帯電話の基地局通信履歴と 地理情報を用いたパーソントリップ推定法の提案

山田 遊馬^{1,a)} 内山 彰^{1,b)} 廣森 聡仁^{1,c)} 山口 弘純^{1,d)} 東野 輝夫^{1,e)}

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、電車および自動車で移動する旅客とその移動経路の推定を行う。通信履歴から得られる位置情報は基地局のセルサイズに依存しており、この位置情報のみからユーザの移動経路を得ることはできない。そこで提案手法では、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。まず地理情報および推定移動速度を利用して電車旅客と自動車旅客を抽出する。その後、電車、自動車旅客それぞれに対して尤度に基づく乗車電車の推定および自動車の移動経路推定を行う。基地局との距離に依存した通信モデルに基づくシミュレーションを行った結果、電車旅客抽出について通信頻度が1時間に9回を超える場合に再現率は88%、精度は89%となり、乗車した電車を85%で推定できることが分かった。また、自動車旅客抽出については再現率は75%、精度は70%となり、経路推定では平均一致度0.644を達成できた。

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、コストをかけずに詳細なPT調査を行うことが可能となっている。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。このため、これまでにGPS（Global Positioning System）や携帯電話の通話履歴など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。しかし、GPSを用いたものは、ユーザが位置情報収集アプリケーションをインストールしなければならないため、年齢などのユーザ属性に偏りが生じ、真の移動パターンを把握できない可能

性がある [1]。

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では、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。これにより、複雑多岐にわたる人々の動きを移動手段の特性に応じて詳細に分類・解析できる可能性がある。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、WiFiやGPSと比較して位置誤差が大きい。従って、これらの特性に合わせた解析方法を設計する必要がある。

本研究では様々な移動手段の中でも特に利用者が多い電車および自動車に着目する。まず、ユーザが滞在しているときの通信履歴の安定性に着目し、滞在地点と期間を抽出する。その後、滞在地点間の移動について、通信履歴から電車・自動車に乗車しているユーザ（電車旅客・自動車旅客）ならびに電車旅客が乗車した電車の推定および自動車旅客の移動経路を推定する。提案手法では、電車旅客について、駅付近の基地局と通信したユーザはその駅にいと仮定し、一定時間内に隣の駅へと移動したユーザを電車旅客とみなす。電車推定では、時刻表により推定される各電車の位置履歴とユーザの通信履歴から得られる移動履歴とのマッチングにより、乗車した電車を推定する。また、自動車旅客について、自動車の平均速度の特徴に着目し、平

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

a) yuma-ymd@ist.osaka-u.ac.jp

b) uchiyama@ist.osaka-u.ac.jp

c) hiromori@ist.osaka-u.ac.jp

d) h-yamagu@ist.osaka-u.ac.jp

e) higashino@ist.osaka-u.ac.jp

均移動速度が一定値より速いユーザを自動車旅客とみなす。自動車旅客の経路推定では、通信履歴から様々な移動経路について尤度を算出することにより自動車旅客の経路を推定する。

提案手法の有効性を評価するために、シミュレータ Scenargie[2] を用いて通勤・通学者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、基地局との距離に応じた確率的な通信モデルによる性能評価を行った。その結果、電車旅客推定は通信頻度が1時間あたり9回を超える人に対して88%の再現率および89%の精度、電車推定については約85%の精度を達成した。また、自動車旅客推定は再現率75%、精度70%となり、自動車経路推定は平均一致度0.644を達成した。

2. 関連研究

2.1 パーソントリップ推定法

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[3], [4], [5]。文献[6]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。提案されている手法では、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[7]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[8]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、タクシーのトレース情報を利用した取り組みもある。文献[9]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や携帯電話の通話履歴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], [10], [11]。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[11]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。文献[10]は、人の動きが社会的な関係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。

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[12]。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[13]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[14]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。

2.2 本研究の位置づけ

本研究では、携帯端末の基地局通信履歴を用いて、人の動きを把握するアルゴリズムを提案している。従来の位置情報をアプリケーションにより収集する方法やPT調査では、ユーザの網羅率が低く、ユーザ層(年齢や性別など)に偏りがあるため正確なユーザの動きが把握できないという問題がある。これに対し、提案手法は日本国民の94.8%[15]が所持している携帯電話の基地局との通信履歴を用いることで、前述の問題点を解消することを狙う。通信履歴では各ユーザの推定位置の精度が基地局のセルサイズに依存するため、位置誤差が約数百mから十数kmにも及ぶことが課題となる。このため、提案手法では駅、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。

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シグナリング情報の履歴である。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

表 1: 通信履歴の例

端末 ID	タイムスタンプ	基地局 ID	緯度, 経度
123456	2014-04-01 07:00:10	AABBAA	(135.012, 34.123)
090909	2014-04-01 07:00:14	DDEEEE	(135.678, 34.888)
123456	2014-04-01 07:01:45	CCCCC	(136.123, 34.666)
90abcd	2014-04-01 07:02:22	AABBAA	(135.012, 34.123)

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。

実際の通信履歴を用いた既存研究 [16], [17] 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 に通信履歴の例を示す。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。これは GPS の誤差である 10m と比較すると非常に大きいといえる [18]。

3.2 通信基地局選択モデル

接続対象となる基地局は電波強度に応じて決まるが、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。これらの影響は非常に複雑なため、厳密なモデル化は困難である。そこで本研究では基地局からの距離に応じた確率的なモデルに基づき、接続する基地局が決定されるものとする。自由空間における受信電波強度は送信機と受信機の距離の 2 乗に反比例することが知られている [19]。このため、基地局と携帯端末の距離の 2 乗に応じて確率的に接続対象となる基地局を決定するものとした。すなわち、各基地局 BS_1, BS_2, \dots, BS_N とユーザの距離を d_1, d_2, \dots, d_N とすれば、ユーザが通信を行うときに基地局 BS_i が選択される確率 P_i は以下で定義される。

$$P_i = \frac{1/d_i^2}{\sum_{j=1}^N 1/d_j^2}. \quad (1)$$

4. 提案手法

4.1 概要

図 1 に示すように、ユーザの移動は異なる地点における滞在の間に発生する。本研究ではまず通信履歴から滞在地点 (SP) の抽出とその滞在期間の推定を行う。こうして抽出された滞在地点 SP_x から SP_{x+1} への移動 M_x について、移動手段 (電車, 自動車, その他) を推定する。最後に、推定された移動手段が電車の場合は乗車車両の推定, 自動車の場合は移動経路の推定を行う。以降の節では、詳細な推定アルゴリズムについて述べる。

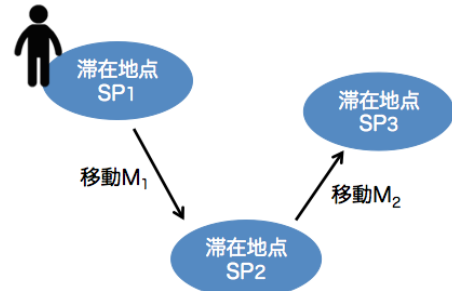


図 1: 滞在地点と移動の関係の例

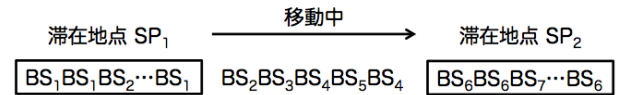


図 2: 通信履歴と滞在地点

4.2 滞在地点の抽出

図 2 に示すように、ユーザが滞在中の場合、位置が変わらないため通信する基地局はほぼ同じであると考えられる。よって、同一基地局と通信した時間が閾値以上の場合、その期間は当該基地局周辺に滞在していたとみなせる。しかしながら、3.1 節で述べたように同じ場所でも異なる基地局に接続される場合がある。したがって、提案手法では同一の基地局と一定期間以上、かつ一定割合以上の通信を行っていれば、当該基地局周辺に滞在していたとする。

まず、ユーザ i の携帯電話の通信履歴の系列 P_i を式 (2) のように定義する。ここで、 $p_i(x)$ はユーザ i が x 番目に通信した基地局を表している。

$$P_i = \{p_i(1), p_i(2), \dots, p_i(m)\} \quad (2)$$

滞在と見なす期間に対する閾値を $T_{stayTime}$ とする。また、ユーザ i が基地局 $p_i(x)$ と通信した時刻を $\text{Time}(p_i(x))$ とする。 P_i のうちの一部の通信履歴 $\{p_i(x), \dots, p_i(y)\} (1 \leq x < y \leq m)$ に対して、以下の式 (3) を満たすものを滞在と判断する。

$$\begin{aligned} & \text{Time}(p_i(y)) - \text{Time}(p_i(x)) > T_{stayTime} \\ & \wedge \frac{\sum_{k=x}^y \text{same}(p_i(x), p_i(k))}{y - x + 1} > T_{stay} \end{aligned} \quad (3)$$

ユーザの滞在判定の条件式を式 (3) に示す。 $\text{same}(p_i(x), p_i(k))$ は $p_i(x) = p_i(k)$ であれば 1, そうでなければ 0 を返す関数であり、 T_{stay} は滞在とみなすための同一基地局との通信回数割合の閾値である。ユーザ i の全通信履歴に対して、全ての x, y の組合せに対して、式 (3) の条件を満たすかどうか判定する。なお、滞在と判定された異なる x, y について期間が重複する場合、期間最長となる x, y を滞在期間として採用する。

4.3 移動手段判別

滞在地点の抽出によって得られたそれぞれの移動 M_x に

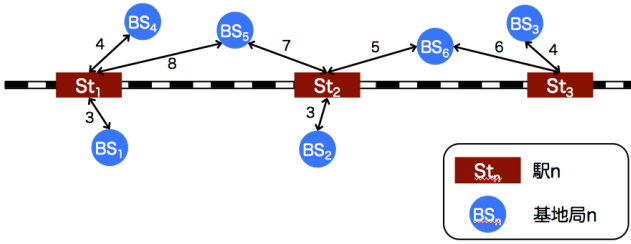


図 3: 駅と基地局の関係図の例

ついて、移動手段の推定を行う。電車による移動は線路に沿って移動し、道路網と比べて空間的特徴が顕著であると考えられる。このため、まず移動 M_x が電車によるものか否かを判断する。もし、電車によるものであると判断できなければ移動 M_x が自動車によるものか否かを判断する。

4.3.1 電車旅客の推定

ユーザが駅に滞在していることを検知するため、各駅に対応する基地局を決定する。駅に最も距離が近い基地局のみを対応する基地局にすると、正しくユーザの滞在を検知できない場合がある。これは 3.1 節で述べたように駅に滞在していても最も距離の近い基地局と通信しているとは限らないからである。そこで、駅と基地局を 1 対 1 で対応させるのではなく、1 対多で対応させる。

各駅に対応する基地局は、以下のように決定する。任意の路線に属する駅を St_1, St_2, \dots, St_N と定義する。ここで、 St_i の次に到着する駅を St_{i+1} とする。まず、駅 St_i に滞在時に通信する可能性のある基地局の集合を S_i とする。駅 St_i に最もユークリッド距離が近い基地局 BS_j を集合 S_i に追加する。そこから、駅 St_i から基地局 BS_k までの距離 $d(St_i, BS_k)$ が $d(St_i, BS_k) \leq 2 \times d(St_i, BS_j)$ を満たす基地局 BS_k を順次、集合 S_i に追加することで、駅 St_i に対応する基地局の集合 S_i を決定する。つまり、駅 St_i との距離が、駅 St_i と最も近い基地局との距離の 2 倍以内である基地局は駅 St_i に滞在しているとみなす基地局とする。ここで、1 つの基地局に対して、複数の駅が該当する場合は基地局との距離が最も近い駅の集合に追加する。これを論理式で表すと式 (4) のようになる。

$$S_i = \{BS_j | d(St_i, BS_j) \leq 2 \times d_{min}(St_i)\} \quad (4)$$

ここで、 St_n は駅 n 、 BS_n は基地局 n 、 $d(Obj_i, Obj_j)$ は Obj_i と Obj_j のユークリッド距離、 $d_{min}(St_n)$ は駅 n に最も距離が近い基地局との距離である。また、 Obj_i は駅または基地局を表す。例としてこの手法を図 3 に示す地形に適用すると、 $S_1 = \{BS_1, BS_4\}$ 、 $S_2 = \{BS_2, BS_6\}$ 、 $S_3 = \{BS_3\}$ となる。

4.1 節で述べた空間的特徴より電車旅客 i は式 (5) を満たす。式 (5) は、ユーザの位置が時系列的に駅 St_{j-1} 、駅 St_j 、駅 St_{j+1} へと順次移動していることを示している。

$$\begin{aligned} \exists a \exists b \exists c ((p_i(a) \in S_{j-1}) \wedge (p_i(b) \in S_j) \\ \wedge (p_i(c) \in S_{j+1}) \wedge (a < b < c)) \end{aligned} \quad (5)$$

また、時間的特徴より電車旅客 i は式 (6) を満たす。ここで $t_i(x)$ はユーザ i が x 番目に基地局と通信した時刻を表している。式 (6) は、駅 St_{j-1} から駅 St_{j+1} への所要時間が $T_{threshold}$ 以内であることを示している。この $T_{threshold}$ は電車の時刻表に基づき決定する閾値である。

$$\begin{aligned} \exists a \exists b ((p_i(a) \in S_{j-1}) \wedge (p_i(b) \in S_{j+1}) \\ \wedge (t_i(b) - t_i(a) \leq T_{threshold})) \end{aligned} \quad (6)$$

よって、式 (5)、(6) の 2 つの条件を満たすユーザ i を電車旅客と判別する。

4.3.2 自動車旅客の推定

徒歩と自動車の大きな違いは平均速度であり、自動車の方が圧倒的に平均速度が速い。その特性に注目し、移動 M_x が自動車によるものか否かを推定する。交差点の位置および交差点間の接続関係は既知であると仮定する。また、移動 M_x の始点 SP_x と終点 SP_{x+1} に最も近い交差点をそれぞれ移動開始地点 IS_x および移動終了地点 IS_{x+1} とする。文献 [20] より、自動車旅客の平均速度は 30 ~ 40[km/h]、高速道路上では 80[km/h] なので、滞在地点間の平均移動速度が 30[km/h] 以上であれば、そのユーザは自動車旅客であるとする。平均移動速度は交差点 IS_x から IS_{x+1} までの道路網の最短距離を移動時間で除して算出する。

4.4 電車推定

電車旅客の推定後、さらに乗車している電車の推定を行う。ユーザの通信履歴ごとに複数の電車に対して尤度を配分し、尤度の最も高かった電車に乗車していると推定する。ユーザ i と i が乗車している電車の移動軌跡は時間的に同一である。よって、ある時刻におけるユーザ i の位置からの距離が近い電車に高い尤度が与えられるように尤度の配分方法を定義する。

電車 i に割り振られる尤度 l_i を式 (7) のように、全電車に割り当てる重みの総和に対する割合で決定する。ここで、 K は対象となる電車の総数、 w_i は電車 i に割り振る尤度の重みをそれぞれ表している。

$$l_i = \frac{w_i}{\sum_{j=1}^K w_j} \quad (7)$$

そして、重み w_i の決定方法として、次の 4 通りを挙げる。

- (1) $w_i = 1/d_i$
- (2) $w_i = 1/d_i^2$
- (3) $w_i = 1/d_i^3$
- (4) $w_i = \begin{cases} 1 & (d_i = \min(d_1, d_2, \dots, d_N)) \\ 0 & (\text{otherwise}) \end{cases}$

ここで、 d_i はユーザと電車 i の距離を示しており、(1) は

重みを距離の逆数, (2) は重みを距離の逆数の 2 乗, (3) は重みを距離の逆数の 3 乗としている. (1)~(3) の中では, (3) が最も距離に近い電車に割り振る尤度が大きい重み決定方法である. また, (4) は最も近い電車に尤度 1.0 全てを割り振る方法である. 5 章の性能評価ではこれら 4 通りの重み決定方法を適用し, それぞれの重み決定方法に対して評価を行う.

4.5 自動車経路推定

自動車の移動特性として, 道路に沿って移動することが挙げられる. そこで, 一般にユーザは目的地までの最短経路を選択するものと仮定し, 滞在地点 SP_x, SP_{x+1} をそれぞれ始点, 終点とする経路のうち, 短い経路から順に N 個の経路 $Path_i$ を移動経路の候補とする. それぞれの経路候補に対して, ユーザの通信履歴から尤度を算出することで移動経路の推定を行う.

経路は交差点毎に区切られた道路 R の系列により構成される. 各道路では通信を行う場合に基地局の距離に応じて 3.2 節で述べたモデルに従って確率的に通信する基地局を決定する. 道路 R と基地局の距離は, 道路 R の平均的な位置, すなわち中心位置と基地局との距離によって定義する.

移動経路 $Path_i$ の尤度は, 基地局 BS との通信履歴が得られた時に $Path_i$ を構成する道路 R を通過していた事後確率 $P_{road}(R|BS)$ に基づき定義する. ベイズの定理により $P_{road}(R|BS)$ は式 (8) で表せる.

$$P_{road}(R|BS) = \frac{P_{conn}(BS|R)P_{car}(R)}{\sum_{i \in ROADS} P_{conn}(BS|i)P_{car}(i)} \quad (8)$$

道路 R を通過する事前確率を $P_{car}(R)$, 道路 R を通過している時に基地局 BS と通信する確率を $P_{conn}(BS|R)$ とする. $P_{conn}(BS|R)$ は 3.2 節で述べた通信基地局選択モデルに従って定義される. $P_{car}(R)$ は交通調査などの事前知識がある場合は, それに基づき決定する. そうでない場合は, 一様分布に従うものとする.

$Path_i$ の尤度 L_i は以下のように算出する. 経路 $Path_i$ を構成する道路を始点から終点まで順に R_1, R_2, \dots, R_M , 通信履歴により得られた基地局の系列を $p(1), p(2), \dots, p(N)$ とする. 通信履歴が得られるタイミングは一定間隔ではないため, $Path_i$ 上の全ての道路で通信が行われるわけではない. 逆に同一の道路で複数回の通信が行われる場合もある. したがって, 提案手法では道路の順序関係と通信履歴の順序関係を保つような組み合わせ a について, 尤度を算出する. 通信履歴 $p(x)$ が道路 R_y で発生したことを表す関数を $f(p(x)) = y$ とすると, この順序関係は以下の論理式 (9) で表される.

$$\forall i \forall j (i < j \Rightarrow f(p(i)) \leq f(p(j))) \quad (9)$$

式 (9) を満たす道路と通信履歴の組み合わせ a における

対数尤度 $L(a)$ を式 (10) の通り定義する.

$$L(a) = \log_{10} \prod_{x \in Hist} P_{road}(R(x)|p(x)) + V \log_{10} \frac{\text{mindist}}{d(Path_i)} \quad (10)$$

V は通信履歴の数, mindist は IS_x から IS_{x+1} までの最短経路の距離, $d(Path_i)$ は $Path_i$ の距離を表している. SP_x から SP_{x+1} までの最短距離に近いほど自然な経路と考えられるため, 式 (10) の第 2 項で経路の距離に対するペナルティを付与している. なお, 1 つの経路 $Path_i$ に対して順序関係を満たす全組み合わせの計算量は動的計画法を用いることで $O(MN)$ となり, 十分実用可能である. 通信間隔によっては複数の経路の尤度が同一になる場合があるため, $Path_i$ について算出された全ての組み合わせ a の対数尤度 $L(a)$ のうち上位 U 個の平均値を $Path_i$ の尤度 L_i とする. 以上のようにして N 個の経路候補全てに対して尤度を計算し, 最も尤度が高い経路を推定結果とする.

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 4 に示す $6,000\text{m} \times 6,000\text{m}$ の領域であり, 100 個の基地局を 666m 間隔の格子状に配置した. このような環境で 12 時間を 9,000 秒間に圧縮しシミュレーションを行った. 評価には 5 回シミュレーションを行った平均値を用いる. 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 節の基地局選択モデルに従う.

図 4 にシミュレーションに用いた領域を示す. 各ユーザには就労者 1, 就労者 2, 学生の行動パターンのうち 1 つを割り当てる. 就労者 1 は図 4 の左上または右下に位置する駅のどちらか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オフィス (図 4 の桃色) へ向かう. オフィス到着後, シミュレーション開始から 3,000 秒までオフィスに滞在した後に元の位置に戻る. 就労者 2 は住宅地 (図 4 の薄緑) 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オフィスへ向かう. そして, 就労者 1 と同様にオフィスで 3,000 秒まで滞在した後に店舗 (図 4 の橙色) へ向かい店舗にて 1,000 秒間過ごした後, 住宅地の元の地点に戻る. 学生は住宅地にランダムに配置され, シミュレーション開始と同時に最寄りの学校 (図 4 の青色) へ向かう. そして, 学校で一定時間過ごした後ランダムに選択した公園 (図 4 の緑色) へ向かい, その後住宅地の元の地点に戻る. 就労者 1, 就労者 2 はそれぞれ 750 人, 学生は 500 人とした. 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.

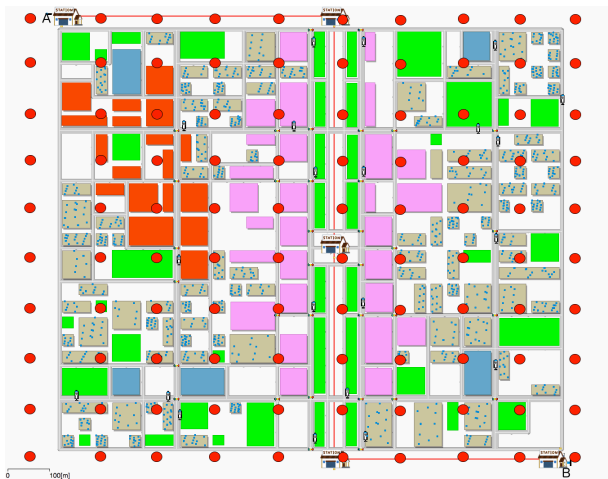


図 4: シミュレーション環境

通信間隔 T については、式 (11) に示すような指数分布とする。 α は通信間隔の期待値、 x ($0 < x < 1$) は一様乱数値を表している。今回は $\alpha = 30$ に設定し、擬似的に通信履歴を生成し、評価に利用する。

$$T = -\alpha \ln x \quad (11)$$

5.2 電車旅客の推定精度

電車旅客と判別したユーザのうち正しかった割合を精度、全電車旅客のうち判別できた割合を再現率として性能評価を行った。ユーザの通信頻度に応じて性能が変化すると考えられるので、通信頻度ごとに精度および再現率を算出した。電車旅客の判別の精度および再現率を図 5 に示す。図 5 の横軸の値 x は最低通信頻度を表しており、通信頻度が x 回以上のユーザを対象とした精度および再現率をそれぞれ示している。通信頻度に制限を持たせない場合の再現率は約 45% となり、電車旅客のうち半分以上は検出できていなかった。これは提案手法では少なくとも 3 回の通信履歴が必要であり、通信履歴が 1 回および 2 回の電車旅客は必然的に検出できなくなる点が原因であると考えられる。最低通信頻度が大きくなるにつれて再現率が上昇し、通信頻度が上位半数（通信回数が 1 時間に 9 回以上）のユーザの再現率は 88% であった。以上の結果より、通信頻度が多い電車旅客の推定精度が非常に高いことを確認した。

5.3 乗車電車の推定精度

乗車電車の推定アルゴリズムの性能評価を行うため、4.4 節で示した 4 通りの尤度に対する重み付け方法について性能評価を行った。評価結果を図 6 の箱ひげ図で示す。(1) が重みを距離の逆数、(2) が重みを距離の逆数の 2 乗、(3) が重みを距離の逆数の 3 乗、(4) が最も距離が近い電車に全尤度を割り振った場合である。図 6 より、距離の逆数に応じて尤度を配分した (1) の精度が最も低くなり、(2)~(4)

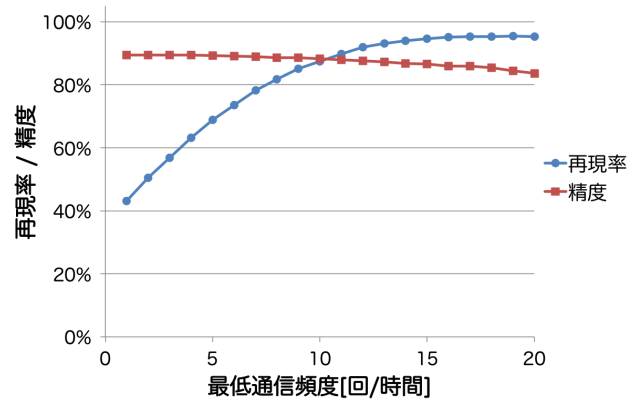


図 5: 電車旅客推定の精度と再現率

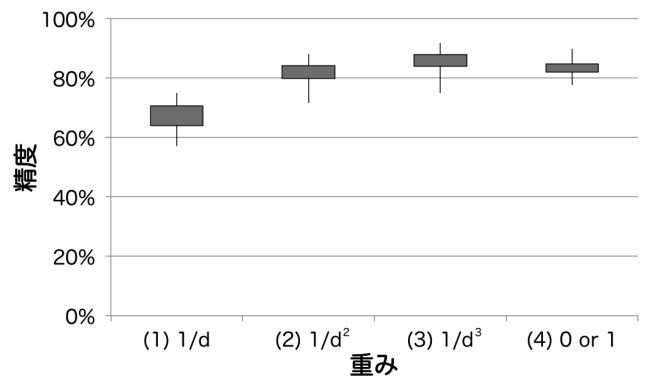


図 6: 乗車電車推定の精度

の配分法の精度はほぼ同じであった。中でも (3) の精度が約 85% と最も高かった。

この結果から、基地局付近で行われた通信を重点的に利用することが精度の向上に寄与することが分かる。これは、ユーザが通信した基地局と複数の電車との距離の差が大きい場合は、最も距離が近い電車に乗っている可能性が高くなる一方で、基地局と複数の電車との距離の差がほとんどない場合は、どの電車に乗車しているか判断しにくいためである。

5.4 自動車旅客の推定精度

総移動距離毎の自動車旅客推定の精度および再現率を図 7 に示す。道路 R を通過する事前確率 $P_{car}(R)$ は一様分布とした。推定精度および再現率の平均はそれぞれ 70%、75% であり、F 値は 0.851 となった。誤検出する旅客の特徴として、移動距離が著しく短い特徴があった。基地局通信の位置誤差は 100m 以上にも及ぶため、距離が短い移動では総移動距離に対する誤差の割合が大きくなる。そのため、平均移動速度が正確に推定できず誤検出につながったと考えられる。

5.5 自動車旅客の経路推定精度

正しく推定できた自動車旅客に対して、4.5 節で定義し

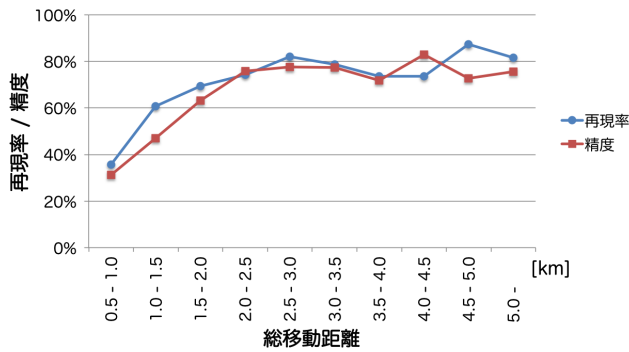


図 7: 自動車旅客の推定精度と再現率

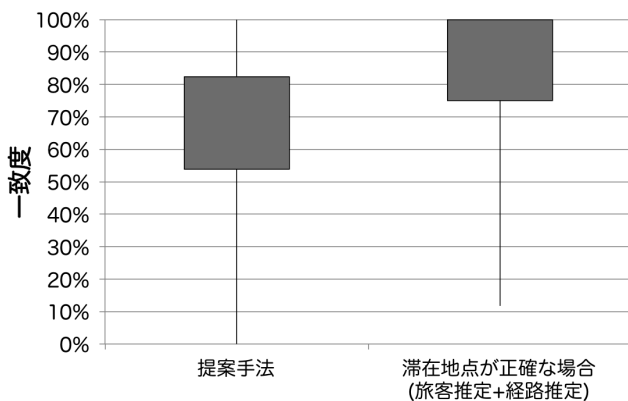


図 8: 自動車経路推定の一致度

たパラメータを $U = 5, N = 50$ に設定し、経路推定を行いその一致度を評価した。文字列の類似性を示す指標として使われる編集距離の一つであるレーベンシュタイン距離 [21] $LevDist(IS_t, IS_e)$ を用いて、経路の一致度 AL を式 (12) のように定義する。

$$AL = 1 - \frac{LevDist(IS_t, IS_e)}{\max(P, Q)} \quad (12)$$

IS_t を実際に自動車旅客が経由した交差点の系列 $IS_t = \{IS_{t1}, IS_{t2}, \dots, IS_{tP}\}$, IS_e を推定された経路が経由する交差点の系列 $IS_e = \{IS_{e1}, IS_{e2}, \dots, IS_{eQ}\}$ とする。

評価結果を図 8(左) に箱ひげ図で示す。全体の一致度の平均値は 0.644, 中央値は 0.688 となりおおむね正しい経路が推定ができていたことが分かる。しかし、約 1/4 の自動車旅客の一致度が 0.5 以下となった。一致度が著しく低いユーザは移動を抽出する元となる滞在地点の推定精度が悪く、経路の端点が真値と大きく異なっていることが原因として挙げられる。

そこで、滞在地点が正確に推定できたと仮定して経路推定を行った評価結果を図 8(右) に示す。図 8(右) より、滞在地点が正確に推定できていた場合の一致度の平均値は 0.838, 中央値は 0.894 となり、経路推定には滞在地点の抽出が重要であることがこの評価結果から伺える。

6. おわりに

本研究では網羅性の高い携帯電話と基地局との通信履歴を活用し、電車および自動車で移動するユーザ (電車旅客・自動車旅客) の抽出、および乗車している電車、移動経路を推定する手法を提案した。基地局通信の位置情報は基地局のセルサイズに依存しており位置誤差が比較的大きいが、提案手法では路線や道路網から得られる地理情報を併用することでこの課題を解消した。シミュレーションによる評価の結果、電車旅客推定は通信頻度が 1 時間あたり 9 回を超える人に対して 88% の再現率および 89% の精度を達成した。乗車している電車の推定については約 85% の精度が得られた。また、自動車旅客の推定は再現率 75%, 精度 70% となり、移動経路の推定では平均 0.644 の一致度を達成できた。

今後の課題として、滞在地点の抽出精度を改善する手法について検討を進めたいと考えている。また、提案手法を拡張し通信頻度が少ないユーザでもより正確に判別できるように実用性を高めていきたい。

参考文献

- [1] Teerayut Horanont, Apichon Witayangkurn, Yoshihide Sekimoto, and Ryosuke Shibasaki. Large-Scale Auto-GPS analysis for discerning behavior change during crisis. *IEEE Intelligent Systems*, Vol. 28, No. 4, pp. 26–34, 2013.
- [2] Space-Time Engineering, LLC. Scenargie. <http://www.spacetime-eng.com/>.
- [3] 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.
- [4] 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.
- [5] Tomer Toledo and Tanya Kolehckina. 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.
- [6] Corrado de Fabritiis, Roberto Ragona, and Gaetano Valenti. Traffic estimation and prediction based on real time floating car data. In *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems (ITSC 2008)*, pp. 197–203, 2008.
- [7] Bruce R. Hellinga. *Estimating dynamic origin-destination demands from link and probe counts*. PhD thesis, Queen's University, 1994.
- [8] Yoshihide Sekimoto, Atsuto Watanabe, Toshikazu Nakamura, Hiroshi Kanasugi, and Tomotaka Usui. 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.
- [9] Raghu Ganti, Mudhakar Srivatsa, Anand Ranganathan, and Jiawei Han. Inferring human mobility patterns from taxicab location traces. In *Proceedings of the 2013 ACM*

- International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 459–468, 2013.
- [10] Daqiang Zhang, Min Chen, Mohsen Guizani, and Haoyi Xiong. Mobility prediction in telecom cloud using mobile calls. *IEEE Wireless Communications*, Vol. 21, No. 1, pp. 26–32, 2014.
- [11] Francesco Calabrese, Francisco C Pereira, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. The geography of taste: analyzing cell-phone mobility and social events. In *Proceedings of IEEE Pervasive Computing*, pp. 22–37, 2010.
- [12] 寺田雅之, 永田智大, 小林基成. モバイル空間統計における人口推計技術 (社会・産業の発展を支える「モバイル空間統計」: モバイルネットワークの統計情報に基づく人口推計技術とその活用), 第 20 巻. NTT DoCoMo テクニカル・ジャーナル, 2012.
- [13] 風間一洋, 鳥海不二夫, 榊剛史, 篠田孝祐, 栗原聡, 野田五十樹. 東日本大震災時の twitter データを用いた単語間の関係の時系列変化の分析. 第 26 回人工知能学会全国大会予稿集, 2012.
- [14] Clodoveu A. Davis Jr, Gisele L. Pappa, Diogo R. R. de Oliveira, and Filipe de L. Arcanjo. Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, Vol. 15, No. 6, pp. 735–751, 2011.
- [15] 総務省. 平成 26 年度版 情報通信白書. 2013.
- [16] Murat Ali Bayir, Murat Demirbas, and Nathan Eagle. Mobility profiler: A framework for discovering mobility profiles of cell phone users. *Pervasive and Mobile Computing*, Vol. 6, No. 4, pp. 435 – 454, 2010.
- [17] Hiroshi. Kanasugi, Yoshihide. Sekimoto, Mori. Kurokawa, Takafumi. Watanabe, Shigeki. Muramatsu, and Ryosuke. Shibasaki. Spatiotemporal route estimation consistent with human mobility using cellular network data. In *Proceedings of 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 267–272, 2013.
- [18] Michael G. Wing, Aaron Eklund, and Loren D. Kellogg. Consumer-grade global positioning system (GPS) accuracy and reliability. *Journal of Forestry*, Vol. 103, No. 4, pp. 169–173, 2005.
- [19] Goldsmith Andrea. ワイヤレス通信工学. 丸善株式会社, 2007.
- [20] 国土交通省道路局. 平成 22 年度道路交通センサ調査結果 (集計結果整理表、箇所別基本表、時間帯別交通量表) . <http://www.mlit.go.jp/road/census/h22-1/index.html>.
- [21] Vladimir I. Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, Vol. 10, p. 707, February 1966.