

基地局通信履歴とGPSトレースを用いた都市レベルの旅行推定手法の提案

山田 遊馬^{*1} 内山 彰^{*1} 廣森 聡仁^{*1} 山口 弘純^{*1} 東野 輝夫^{*1}

^{*1} 大阪大学 大学院情報科学研究科

{ yuma-ymd, uchiyama, hiromori, h-yamagu, higashino }@ist.osaka-u.ac.jp

1 はじめに

首都圏などの広範囲にわたる人々の動きを把握するため、これまでにGPSにより得られる位置情報を用いた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。

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。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て通信した基地局 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。しかし、通信履歴では通信時に接続した基地局 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。また、通信する基地局は電波受信強度を基に決定されるが、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、必ずしも最寄りの基地局と通信するわけではない。これらの課題を解消するために、本研究では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、学習結果と推定対象ユーザ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、移動手段および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。本稿では様々な移動手段の中でも最も利用者の多い電車旅客に着目し、推定対象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。

2 想定環境

2.1 通信履歴の特性

通信履歴はユーザが通信し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。

基地局と通信した際にユーザ端末 ID、通信時刻、基地局 ID (緯度、経路) が基地局側で通信履歴として記録され

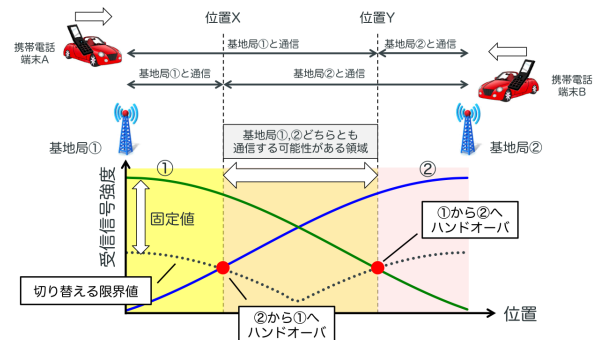


図 1: ハンドオーバーの一例

る。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。

2.2 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、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 [2] 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。

3 提案手法

図 2 の矢印のように、同じ路線に乗車している電車ユーザであっても、乗車位置などによって基地局遷移パターンが異なる場合がある。また、ユーザによって通信頻度が大きく異なるため、必ずしも隣接した基地局にハンドオーバーするとは限らない。ゆえに、基地局遷移パターンの組み合わせ数が膨大になってしまい、大量の学習データを必要とする。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、これら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 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

City-Scale Trip Estimation Using LTE Control Signal Records and GPS Traces

Yuma Yamada^{*1} Akira Uchiyama^{*1} Akihito Hiromori^{*1} Hirozumi Yamaguchi^{*1} Teruo Higashino^{*1}

^{*1} Graduate School of Information Science and Technology, Osaka Univ., Osaka Japan

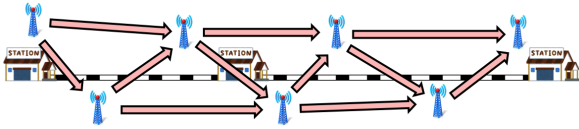


図 2: 駅間を移動するユーザの基地局遷移パターンの例

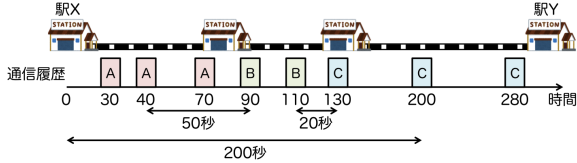


図 3: GPS ユーザから得られる教師データの例

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。そして、教師データから得られた学習結果を用いて、トリップの尤度を算出し、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。

3.1 基地局遷移パターンの学習

異なる2つの基地局において発生し得る通信間隔の学習結果を学習テーブル Tbl とする。学習テーブルはトリップ属性ごとに定義され、出発駅 i から到着駅 j へのトリップの学習行列を $Tbl^{(i \rightarrow j)}$ とする。また、学習テーブル $Tbl^{(i \rightarrow j)}$ の各要素 $Tbl^{(i \rightarrow j)}(m, n)$ 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、要素は通信間隔の最小値と最大値からなる区間である。

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、到着駅をそれぞれ $st_{orig}(T_r)$, $st_{dest}(T_r)$ 、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 (1) のように定義できる。

$$\mathcal{TD}^{(i \rightarrow j)} = \{C_u^{(t_{orig}(T_r), t_{dest}(T_r))} \mid st_{orig}(T_r) = i \wedge st_{dest}(T_r) = j, T_r \in \mathcal{T}_r^{(u)}, u \in \mathcal{TU}\} \quad (1)$$

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 (2) 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。

$$C_u^{(t_1, t_2)} = \{c \mid t_1 \leq t(c) \leq t_2, c \in C_u\} \quad (2)$$

そして、駅 i から駅 j へ移動するトリップの学習テーブル $Tbl^{(i \rightarrow j)}$ 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 $BS(c_1)$ と通信してから基地局 $BS(c_2)$ と通信するまでに要する時間を学習し、学習区間 $Tbl^{(i \rightarrow j)}(BS(c_1), BS(c_2))$ を更新する。ゆえに、学習区間 $Tbl^{(i \rightarrow j)}(m, n)$ 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。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 $Tbl^{(i \rightarrow j)}(O, m)$, $Tbl^{(i \rightarrow j)}(m, D)$ で表す。

3.2 電車トリップ候補の抽出

3.1 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。学習データの要素 $Tbl^{(i \rightarrow j)}(m, n)$ 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。学習データですべての通信間隔を網羅できているわけではないので、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、一定の誤差を許容するよ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 $l_{u,t}^{(i \rightarrow j)}$ を式 (3) のように定義する。 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、 $p(t, m)$ 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) で定義する。 $p(t, m)$ は通信間隔 t が学習区間の中央値であれば1、学習区間の両端であれば $\exp(-\frac{1}{2})$ となり、学習区間から遠く離れるほど小さい値になるように設定している。

$$l_{u,t}^{(i \rightarrow j)} = \prod_{c_1, c_2 \in C_u^{(t, t+\Delta T)}} p(t_{c_1, c_2}, Tbl^{(i \rightarrow j)}(BS(c_1), BS(c_2))) \quad (3)$$

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

尤度 $l_{u,t}^{(i \rightarrow j)}$ が閾値以上であれば、出発時刻を t とする駅 i から駅 j へのトリップであると推定する。ここで、検証区間 $[t, t + \Delta T]$ は駅 i から駅 j へのトリップであるか否かを推定する区間、 ΔT は学習で得られた駅 i から駅 j へのトリップに要する時間を表している。

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設定されている。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。

4 まとめと今後の課題

本稿では携帯電話の基地局通信履歴を用いて、電車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学習させることで、電波伝搬を考慮した推定を行った。

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] 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).