

携帯電話通信履歴に適した移動滞在状態推定手法の提案

小林 直^{1,a)} 石塚 宏紀¹ 南川 敦宣¹ 村松 茂樹¹ 小野 智弘¹

受付日 2016年9月9日, 採録日 2017年1月7日

概要: 近年, 高速で大容量なモバイル通信の普及にともない, 携帯端末を所持するユーザの行動分析に注目が集まっている. ユーザの行動を把握できると, 災害時の避難対策や交通施策への活用だけでなく, 行動を考慮した緻密な通信エリアの設計も可能になる. また, ユーザの行動を把握するために携帯端末の位置情報を用いた移動滞在状態推定技術の研究も多々検討されており, 活用する位置情報としては, アプリケーションのインストールや端末の電力消費等の負担を課することなく端末の位置を把握でき, 網羅的に位置情報の収集が可能な基地局情報に基づく位置情報履歴 (CDRs) が注目されている. しかしながら CDRs は, 基地局と端末間の無線通信情報に基づいており, 通信時やサービス運用上の処理を契機とした不規則な位置情報であるため, 時空間的に密であり均一間隔に取得された GPS 等を前提とした既存の移動滞在状態推定手法の適用は大きな課題を生じる. そこで本論文では, CDRs を用いて時空間の変化に適応した移動滞在状態推定手法を提案する. 本論文の貢献は, 可変時間窓を用いた取得間隔の疎密吸収と 2 段階距離閾値を用いた測位誤差吸収による CDRs 利用時の移動滞在状態推定精度の向上にある. 提案手法の有意性評価においては, 事前許諾を得た参加者 582 名の 1 カ月分の CDRs を用い, 同時取得した GPS から推定した移動滞在状態を正解として精度検証を行った. 評価結果から, 提案手法は既存手法より高精度に状態推定が可能であることが明らかとなった.

キーワード: 携帯電話通信, 携帯電話通信履歴, 移動滞在状態推定, 位置情報

An Estimation of Human Moving Patterns for Call Detail Records

NAO KOBAYASHI^{1,a)} HIROKI ISHIZUKA¹ ATSUNORI MINAMIKAWA¹ SHIGEKI MURAMATSU¹
CHIHIRO ONO¹

Received: September 9, 2016, Accepted: January 7, 2017

Abstract: Researches on user behavior analysis have attracted huge attention recently due to the popularity of smart mobile devices and the availability of high-speed mobile communication services. Proper user behavior modeling not only provides insight into the design of emergency and traffic planning in catastrophic situations, but also gives hints on sophisticated wireless area design. To this end, researches on the topic of understanding human mobility patterns by estimating user mobility status via users' spatial-temporal data such as GPS have been developed. To alleviate the problem of aggressive battery consumption due to GPS sensing, approaches utilizing location data embedded in Call Detail Records (CDRs) from base stations become intriguing and promising. Compared with high-resolution GPS data, the location data in CDRs is less accurate. Besides, the sampling of locations in CDRs is spatial-temporally irregular due to the irregular communication patterns and infrastructure maintenances. For these reasons, tuning existing methods to adapt for CDRs is difficult and may be impractical. In this paper, we propose a method to estimate end-user mobility status that tolerates the spatial-temporal irregularity in CDRs' location data. Concretely, we employ an adaptable sliding window to group records to address the irregularity problem and a 2-level hierarchical distance estimation scheme to reduce positioning error in CDRs. To evaluate our proposed method, we collected both GPS and CDR data of 582 end-users with user consents for a time span of 30 days (wherein the GPS data is regarded as ground-truth), and we tested our method on this dataset. The results prove the feasibility of our method and demonstrate higher accuracy compared with existing methods.

Keywords: mobile communications, Call Detail Records, human moving patterns, locations

1. まえがき

スマートデバイスの普及による高速で大容量なモバイル通信時代において、携帯端末を所持するユーザの行動分析に注目が集まっている。ユーザの行動を把握できると、災害時の避難対策や交通施策への活用だけでなく、行動を考慮した緻密な通信エリアの設計も可能になる。

行動分析技術の分野において、行動の最小単位である移動滞在状態推定技術の研究がさかんに行われている [1], [2], [3], [4], [5], [6], [7], [8], [9]。高精度な移動滞在状態推定を実現することで、主要滞在地の推定 [10], [11], [12] や滞在目的の推定 [13], [14], [15] 等の応用的な行動分析が可能となる。移動滞在状態推定における既存研究では、スマートデバイスから直接的に取得された GPS や Wi-Fi 等の時空間的に高解像度であり、均一間隔で取得された位置情報の利用を前提としたアルゴリズム設計と評価を行っている。

一方、近年では基地局情報に基づく位置情報履歴 (CDRs: Call Detail Records) が注目されている。CDRs は、ユーザの通信や通信ネットワークの運用上の処理を契機に接続した基地局情報を用いて、通信事業者側の制御サーバで間接的に測位された位置情報である。そのため、ユーザに対して、位置情報収集用のアプリケーションのインストールや同アプリケーションによる端末の電力消費等の負担を課することなく測位可能であり、網羅的な位置情報の収集が実現できる。本論文では、CDRs からユーザの行動を把握することを目的とした移動滞在状態推定技術について取り扱う。

CDRs は通信に付随して生成される位置情報であるため、GPS や Wi-Fi に基づく位置情報と比較して、以下の課題が存在する。

課題 1. 不規則な取得契機

CDRs はユーザの移動行動、通信行動を契機に生成されるため時空間的に不規則である。

課題 2. GPS に比べると低精度

CDRs の位置情報は通信を行った基地局の位置であり、GPS に比べると位置精度ははるかに低く、かつ、同じ場所に滞在していてもエリアをカバーする基地局が複数あれば位置情報がばらつく可能性がある。

このため、CDRs に適した移動滞在状態推定はユーザの移動行動、通信行動、通信環境の 3 点の多様性に適したアルゴリズムが必要となる。我々は、上述した 2 つの課題が与える影響を軽減する方式として、可変時間窓および 2 段階距離閾値を用いた移動滞在状態推定手法を提案する。提案手法は (1) 可変時間窓によって CDRs の取得間隔のばら

つきの影響を受けず、移動滞在状態推定に (2) 2 段階距離閾値を使うことによって測位誤差の影響を軽減することができ、既存技術に比べ高精度に移動滞在状態を推定できる。

本論文では提案手法の有効性を確認するために、事前許諾を得た参加者 582 名の 2014 年 3 月における 1 カ月分の CDRs と GPS 情報をもとに評価した。評価は GPS から推定した移動滞在状態を正解として、提案手法と既存の移動滞在状態推定手法の精度を比較した。その結果、移動滞在状態推定の結果は提案手法が最も高精度であり、滞在状態推定の結果は既存手法と提案手法ともに移動滞在状態推定結果よりも高精度であった。さらに、移動滞在状態推定についてはユーザの移動行動、通信行動、通信環境の変化への耐性を確認するため、実験データを各種特徴においてクラスタリングした評価データセットを多数作成し、各々を評価した。その結果、1,708 の評価条件のうち、84% の評価条件で提案手法が他手法に比べ有意であり、ユーザの移動行動、通信行動、通信環境のあらゆる変化に対して柔軟に適応可能な手法であることを明らかにした。以下に、本論文の貢献についてまとめる。

貢献 1. CDRs の特徴を分析

携帯電話サービスの実契約者を参加者とした長期間におよぶ CDRs と GPS の同時取得による確かな評価データセット作成と、ユーザの移動行動、通信行動、通信環境の関係性についての分析

貢献 2. CDRs に適した移動滞在状態推定手法の提案

可変時間窓および 2 段階距離閾値を用いた移動滞在状態推定手法を導入した手法の提案と評価

本論文の構成は以下である。2 章で既存の行動分析技術について述べ、3 章で提案手法について述べる。4 章では評価概要を示し、5 章で評価を行い、最後に結論を述べる。

2. 行動分析技術

本章では行動分析技術の活用例と行動分析技術の基盤である移動滞在状態推定技術について述べる。

2.1 行動分析技術の活用

行動分析技術においてはユーザの行動だけでなく、行動から推定できるユーザの状況、未来予測等様々な検討がなされている。たとえば、滞在頻度からユーザの自宅や職場といった主要な滞在場所の推定 [10], [11], [12]、滞在場所に付与されたラベリング情報から滞在目的の推定 [13], [14], [15]、機械学習を用いた行先予測 [16], [17], [18], [19]、移動情報から学生や社会人といった属性の推定 [20] がある。さらに、特定の場所で生じるユーザ群の移動量を推定 [21], [22] する研究も行われている。これらの技術をより高精度に実施するためには基盤となるユーザの行動を移動と滞中に正確に分離することが重要である。

¹ 株式会社 KDDI 総合研究所
KDDI Research, Inc., Fujimino, Saitama 356-8502, Japan

a) no-kobayashi@kddi-research.jp

2.2 既存の移動滞在状態推定技術

既存の移動滞在状態推定技術では位置情報のみから移動滞在状態を推定する技術 [1], [2], [3], [4], 位置情報と時間情報を併用して推定する技術 [5], [6], [7], 加速度等のセンサ情報を利用する技術 [8], [9] がある. 本論文で扱う CDRs には基地局の位置情報および基地局と端末との通信時刻が記録されているので, 位置情報と時間情報を用いる技術が適用できる.

位置情報のみを用いる移動滞在状態推定技術は定期的取得される位置情報に対して適用できる. Kang ら [4] が提案する手法は位置情報のみを用いる手法であり, 蓄積した位置情報に対し, 移動滞在状態および滞在位置を推定する. 本手法は移動滞在状態推定の対象とする位置情報と, 直前に生じた位置情報との距離が閾値 (D) 内である場合, 同一のクラスタとする. また, 同一のクラスタとなった場合, クラスタ内の全位置情報の重心を求め, 次に生じた位置情報との距離を比較する. 距離が D の範囲内であれば, 次に生じた位置情報も同一のクラスタとし, 新たな重心を求める. ここで, 位置情報が連続して D の範囲外である場合, 最初に D の範囲外となった位置情報の直前までが同一のクラスタとなる. ただし, 連続して D の範囲外であってもあらかじめ設定された数を超える前に, クラスタに含まれる位置情報が生じた場合, D の範囲外である位置情報は外れ値として扱われ, クラスタの重心に置き換えられる. また, 連続したクラスタが生じた場合, クラスタの重心どうしの距離が $D/3$ 以内であれば同一のクラスタとし, 各クラスタの中心の重心が滞在位置となる. そして, クラスタの開始時刻と終了時刻から導出される滞在時間がそれぞれあらかじめ設定された閾値を超えていた場合, クラスタは滞在状態と推定される.

位置情報と時間情報を用いる移動滞在状態推定技術は不定期に取得される位置情報に対して適用できる. Liu ら [6] が提案する手法は位置情報と時間情報を併用する手法であり, 蓄積した位置情報と時間情報に対し, 移動滞在状態および滞在位置を推定する. 本手法は移動滞在状態推定の対象となる位置情報が発生した時刻からあらかじめ設定された時間 ($mintime$) を遡り, $mintime$ に含まれる全位置情報との距離を算出し, 各位置情報との距離が閾値 (D) の範囲内であれば同一の滞在位置と推定する. また, 同一の滞在位置と推定された場合, 各位置情報間の距離を総当りで求め, 合計距離が小さくなる位置情報を滞在位置としている. なお, 以降に発生した位置情報は滞在位置との距離が $D/2$ 以内であれば同一の滞在位置とし, 条件を満たさない場合は直前の位置情報までが滞在状態と推定され, 新たな位置情報は開始点として, 次の位置情報を処理する.

2.3 既存の移動滞在状態推定技術の課題

Kang らの手法は新たに生じた位置情報が連続してクラ

スタの重心から D 以上離れたときに移動状態と推定するため, 時間的に疎な移動区間において, 実際移動状態であっても外れ値として推定されてしまう. さらに, 推定対象ごとに固定時間窓をスライドさせるため, 滞在位置の重心が変化する. 本影響により, 低速移動が検知し難いといった問題がある. 一方, Liu らの手法は移動状態と推定されている間は Kang らの手法と同様に, 固定の時間幅を用いるが, ひとたび滞在状態と推定されると基準点を固定するので, Kang らの手法の欠点である低速移動時の検知性能を改善している. しかしながら Kang らの手法と同様, 推定に用いる D が外れ値の影響を受けやすく, D が大きいと移動を検知し難いといったトレードオフの問題が残っている. 本論文ではこれら既存の課題を軽減させることを目的とした移動滞在状態推定手法を提案する.

3. 提案手法

本章では提案手法について具体的に記載する. 図 1 は滞在場所 A から滞在場所 B へ移動したユーザの CDRs の位置情報の軌跡と発生時刻の関係性を示している. 前述のとおり, CDRs は時空間的に粗密があるため, Kang らの手法, Liu らの手法を適用すると 2.3 節で記述した課題の影響が顕著に現れる. この粗密の影響を軽減するため, 提案手法では以下の手順で処理を実行する.

- (1) 可変時間窓による移動滞在状態推定対象区間の選定
- (2) 移動滞在状態推定処理
- (3) 2段階距離閾値による移動開始, 終了時刻の補正

3.1 可変時間窓を用いた移動滞在状態推定対象区間の選定

既存の移動滞在状態推定手法は一定時間間隔で位置情報が生成されることを前提にしており, 特に Kang らの手法は特定区間の時間窓の幅を固定している. しかし, CDRs のように生成間隔に粗密がある場合, 移動状態時に生成さ

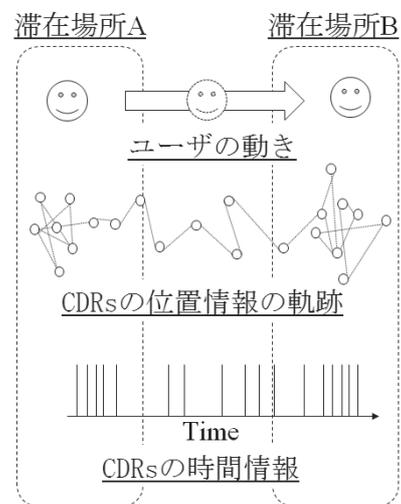


図 1 ユーザの動きと CDRs の関係性

Fig. 1 Association of user movement and CDRs.

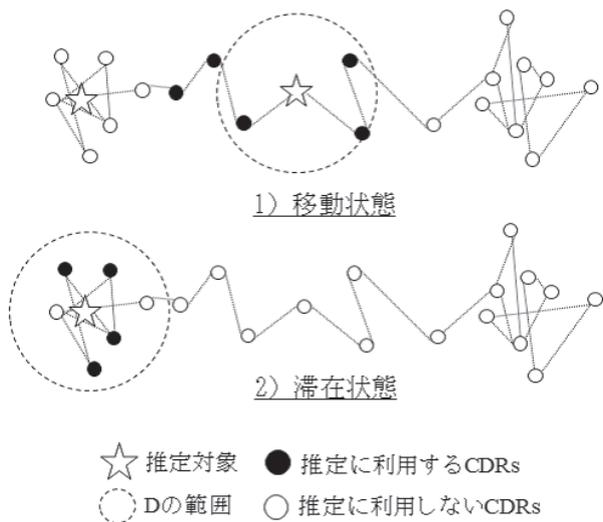


図 2 移動状態推定と滞在状態推定

Fig. 2 Estimation of human movement patterns.

れるレコード数が少ないと短時間で往復のような移動をした際に、外れ値として検知され、除外される可能性が高くなる。そこで、提案手法は Liu らの手法と同様に滞在時のみ可変の時間窓幅を用いる。推定対象の CDRs を n , n の生成時刻を $T(n)$, 移動時の固定時間窓幅を t とすると, Liu らの手法の対象区間は $[T(n) - t, T(n) + t]$ となる。提案手法では以下の条件を満たす場合に時間窓の開始時刻を固定し, 時間窓幅を可変とする。

(1) $n - 1$ での推定結果が ‘滞在状態’

(2) $T(n) - T(n - 1) < t$

のときに, 滞在開始時刻は $T(n - m)$ (ただし $m > 1$) まで遡り, 時間窓の開始時刻は $T(n - m) - t$, 時間窓の区間は $[T(n - m) - t, T(n) + t]$ となる。

3.2 移動滞在状態推定処理

時間窓内に存在する CDRs のうち, 推定対象の CDRs との距離がすべて閾値以内であれば滞在状態と推定し, 滞在状態以外を移動状態と推定する。図 2 は選定された CDRs に基づく移動推定と滞在推定の概念図であり, 図中の星が推定対象を示している。また, 推定対象を囲む破線が推定に用いる距離の閾値 (D) の範囲であり, 黒丸が推定区間内の CDRs を示している。図中の 1) のように破線外に CDRs が 1 つでもあれば移動状態, 2) のように CDRs がすべて破線内であれば滞在状態と推定される。

3.3 2段階距離閾値による移動開始, 終了時刻の補正

本節で提案する手法は推定に2段階の距離 ($D, D2: D > D2$) を用いて移動滞在状態推定結果を補正する。図 3 中の 3), 4) は $D, D2$ に基づく移動滞在状態推定の概念図であり, D は $D2$ に比べ, 移動状態と推定された位置情報が少なく, 滞在状態と推定された位置情報が多いため

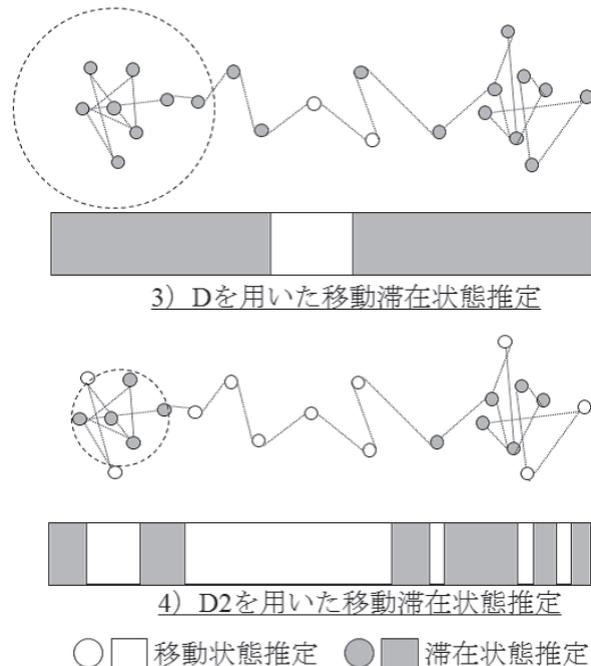


図 3 D と $D2$ に基づく移動滞在状態推定結果

Fig. 3 Human movement patterns estimation based on D and $D2$.

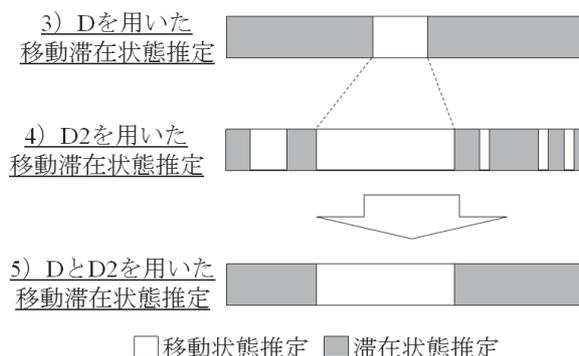


図 4 2段階距離閾値を用いた移動滞在状態補正

Fig. 4 2-level hierarchical distance estimation scheme.

る。提案手法は距離が長い D の推定結果で移動状態と推定された CDRs が $D2$ でも移動状態と推定されていれば, 該当の CDRs が含まれる $D2$ の移動区間の開始時刻と終了時刻を正とする。図 4 は本手法の概念図であり, 本手法によって滞在中の測位誤差の影響を軽減しつつ, 移動状態と推定されると移動区間の開始, 移動時刻を精度良く検知することが可能になる。このように2段階距離閾値を用いることで, 2.3 節で記述した Kang らの手法と Liu らの手法の課題であったトレードオフ問題を軽減できる。

4. 評価概要

本論文では提案手法と既存手法を用いて, ユーザの移動行動, 通信特性, 通信環境等から作成した複数のデータセットを評価する。評価対象となるデータセットは事前許諾を

表 1 移動時間と移動距離におけるトリップ数の分布

Table 1 Statistics of trip number in moving time and moving distance.

	0 [s]	900 [s]	1800 [s]	2700 [s]	3600 [s]	4500 [s]	5400 [s]	6300 [s]	7200 [s]	8100 [s]	9000 [s]-	合計
0 [km]	2621	3304	1393	664	372	167	93	39	23	9	12	8697
2 [km]	817	1592	745	376	201	134	57	37	18	14	20	4011
4 [km]	146	931	558	254	181	111	56	30	16	12	28	2323
6 [km]	51	449	479	264	155	88	61	30	21	12	26	1636
8 [km]	12	201	327	242	132	85	61	48	22	19	39	1188
10 [km]	10	156	302	225	129	81	48	32	32	12	33	1060
12 [km]	10	68	211	194	129	82	62	39	18	9	31	853
14 [km]	9	37	159	162	124	74	36	37	27	10	29	704
16 [km]	6	17	99	131	119	83	58	20	20	14	36	603
18 [km]	7	15	91	96	112	67	47	34	24	18	34	545
20- [km]	48	65	156	417	624	693	587	494	351	284	1113	4832
合計	3737	6835	4520	3025	2278	1665	1166	840	572	413	1401	26452

得た日本在住の 20~50 代の男女 582 名の 2014 年 3 月の 1 カ月分の CDRs と GPS を用いる。ここで、参加者の携帯電話の利用方法に関しては制約を持たせていないが、推定結果と正解状態の比較のため、参加者の端末には 5 分間隔で GPS による測位結果を自動記録するアプリケーションをインストールしている。また、今回使用した GPS は緯度経度に加え、衛星数も記録しており、衛星数から求めた測位信頼度が低い GPS は除外している。上記から得られた GPS と CDRs に対して移動滞在状態を推定し、GPS が存在する時刻の前後 30 秒に存在する最も近い時刻の CDRs と比較を行う。ここで、GPS の移動滞在状態は位置情報を定期的に得られることから時間情報と位置情報を利用する Liu らの手法を適用して推定し、正解の状態とした。さらに、本論文の評価はユーザによる行動の違いを考慮するため、トリップ単位で評価を行う。トリップは正解の状態が 5 分以上の異なる滞在状態に 1 分以上の移動状態が挟まれていれば 1 回のトリップとして定義した。なお、本定義による評価対象トリップ数は 26,452 トリップであった。

4.1 データセットの構築

本論文ではユーザの移動行動、通信行動、通信環境の変化に対する提案手法の耐性を確認するため、トリップを各種特徴においてクラスタリングしたデータセットを複数作成する。

- ユーザの移動行動
 - トリップの移動時間、移動距離をもとにデータセットを作成する。
- 通信行動
 - 移動時に行われた通信数、移動時の通信時間間隔と距離間隔の標準偏差をもとにデータセットを作成する。
- 通信環境
 - 移動時に端末が接続した基地局の周囲に存在する他の基地局数をもとにデータセットを作成する。

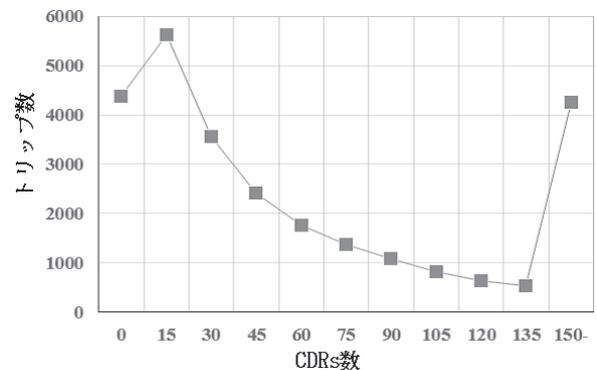


図 5 移動時の CDRs 数とトリップ数の分布

Fig. 5 Distribution of CDRs number and trip number during movement.

また、各データセットを掛け合わせた「ユーザの移動行動と通信行動」、「ユーザの移動行動と通信環境」、「通信行動と通信環境」についても評価を行う。

4.2 データセットの詳細

最初にユーザの移動行動として、表 1 は移動時間と移動距離に基づきトリップを分類した結果である。本表からトリップの 46% は移動時間 2700 秒未満かつ移動距離は 6 km 未満の短い移動であることが分かる。次にユーザの通信行動として、図 5 は移動時の CDRs 数、図 6 は移動時の通信時間間隔における標準偏差、図 7 は移動時の距離間隔における標準偏差に基づきトリップを分類した結果を示している。本結果からも通信数および標準偏差はユーザまたは行動によって一定ではないことが分かる。最後に通信環境として、図 8 はトリップごとに接続した通信局の周囲 1 km に存在した基地局数の中央値に基づきトリップを分類した結果を示しており、本結果からもトリップによって周囲の基地局数が異なることが分かる。したがって、CDRs で得られる位置情報はユーザの移動行動、通信行動、通信環境によって各特徴が変動するので、これらの変化に対応

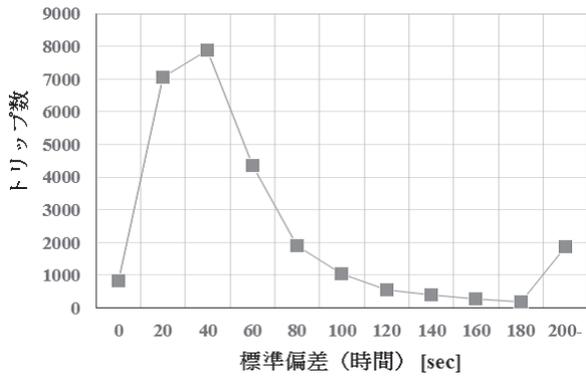


図 6 標準偏差 (時間) とトリップ数の分布

Fig. 6 Distribution of standard deviation (time) and trip number.

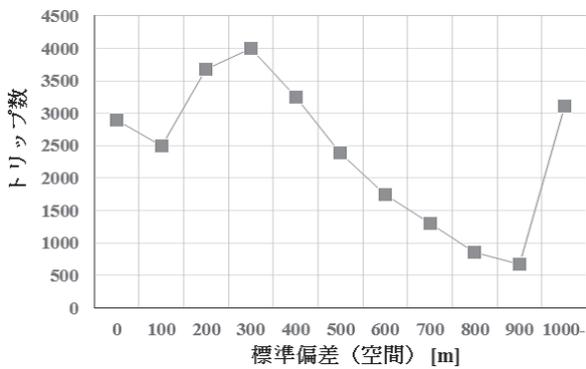


図 7 標準偏差 (空間) とトリップ数の分布

Fig. 7 Distribution of standard deviation (space) and trip number.

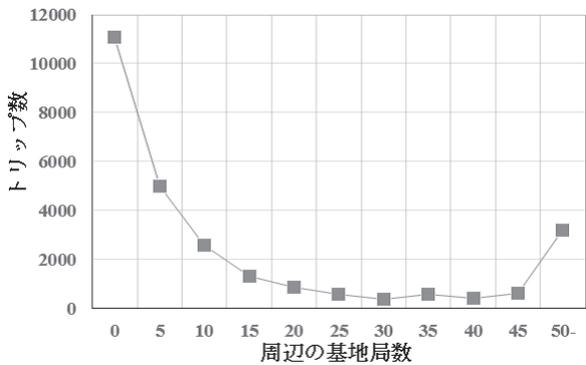


図 8 周辺の基地局数とトリップ数の分布

Fig. 8 Distribution of the number of neighbor base stations and trip number.

できる移動滞在状態推定手法が望ましい。

4.3 各手法における最適な距離評価概要

既存手法である Kang らの手法, Liu らの手法と提案手法は共通して, 移動滞在状態推定の閾値に距離 (D) を用いるため, D によって移動滞在状態が変化する可能性がある。そこで, 各手法における D の変化にともなう移動滞在状態の影響を確認し, 移動状態と滞在状態の両方に最適な

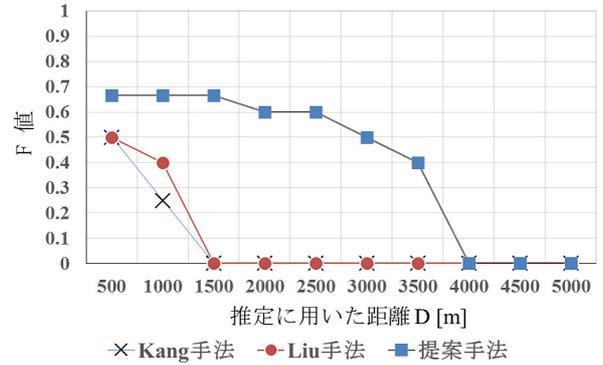


図 9 移動状態推定精度

Fig. 9 F-measure of moving state estimation.

D を求める。なお, D は 500 m, 1,000 m, ..., 5,000 m と 500 m 間隔で変化させ, 各 D における移動滞在状態を推定し, 評価する。 D 以外の設定値として, Kang らの手法ではクラスタの終了を示す閾値を 3, 滞在の定義はトリップの最小滞在時間である 5 分以上と設定した。また, Liu らの手法の *mintime* もトリップの最小滞在時間から 5 分と設定した。提案手法についても最小の滞在時間となる固定の時間窓は 5 分とし, 2 段階距離閾値に用いる距離 D_2 は最小の D である 500 m より短い 400 m とした。なお, 評価は CDRs を用いて推定した移動滞在状態が GPS と同一の状態であれば正解状態とし, 各手法から推定した状態と GPS から推定した状態の適合率と再現率から算出できる F 値を求める。そして, 全トリップに対して算出した F 値の中央値を用いて各データセットを評価する。

Precision

$$= \frac{\text{CDRs と GPS の状態が一致した数}}{\text{GPS から推定した移動あるいは滞在状態推定数}}$$

Recall

$$= \frac{\text{CDRs と GPS の状態が一致した数}}{\text{CDRs から推定した移動あるいは滞在状態推定数}}$$

$$\text{F-Value} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. 評価結果

本章では最初に各手法における最適な距離を求める。次に複数のデータセットに対し, 最適な距離を用いた Kang らの手法, Liu らの手法, 提案手法による移動滞在状態の推定結果と正解の移動滞在状態を比較し, 評価する。

5.1 各手法における最適な距離評価結果

図 9 は各 D における移動状態の評価結果を示しており, 全手法共通して D の増加にともない, F 値が減少する傾向があることを確認した。これは D の増加によって, D 以下の距離の移動状態を推定できなくなったことが考えられる。また, 各手法における最も高い F 値と, 推定に利用した D は Kang らの手法および Liu らの手法は $D = 500$ のときに最も高い 0.5 となり, 提案手法は $D = 500 \sim 1500$ の

表 2 ユーザの移動行動に基づくデータセット

Table 2 Dataset constructed from user's movement behavior.

移動距離 [km]	0	2	4	6	8	10	12	14	16	18	20-
トリップ数	8697	4011	2323	1636	1188	1060	853	704	603	545	4832
Kang らの手法	0	0.4	0.5	0.571	0.6	0.667	0.667	0.667	0.667	0.667	0.744
Liu らの手法	0	0.462	0.572	0.667	0.667	0.715	0.75	0.727	0.75	0.75	0.8
提案手法	0	0.667	0.727	0.8	0.8	0.833	0.857	0.833	0.857	0.857	0.889
移動時間 [sec]	0	900	1800	2700	3600	4500	5400	6300	7200	8100	9000-
トリップ数	3737	6835	4520	3025	2278	1665	1166	840	572	413	1401
Kang らの手法	0	0	0.445	0.5	0.556	0.625	0.632	0.667	0.667	0.667	0.667
Liu らの手法	0	0.333	0.5	0.6	0.667	0.667	0.727	0.75	0.75	0.764	0.778
提案手法	0	0.5	0.667	0.769	0.8	0.824	0.857	0.861	0.875	0.885	0.889

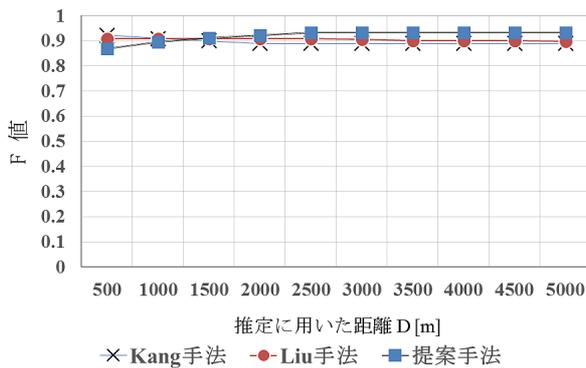


図 10 滞在状態推定精度
Fig. 10 F-measure of staying state estimation.

ときに最も高い 0.667 となった。
次に図 10 は D における滞在状態の評価結果を示している。本結果から滞在状態推定に関しては全手法、全 D において 0.85 以上の F 値であることが分かり、移動状態に比べて推定が容易であることを確認した。また、各手法における最も高い F 値と、推定に利用した D は Liu らの手法が $D = 500 \sim 2500$ のときに最も高い 0.909 となり、Kang らの手法は $D = 500$ のときに最も高い 0.923 となった。提案手法に関しては $D = 2500 \sim 5000$ のときに最も高い 0.933 となった。本論文では移動状態と滞在状態の両方を推定することを想定しているため、Liu らの手法は $D = 500$ 、Kang らの手法は $D = 500$ 、提案手法は $D = 1500$ を最適な D として設定した。

5.2 各データセットの評価結果

前節で求めた各手法の最適な D を用いてユーザの移動行動、通信行動、通信環境に基づく複数のデータセットを評価する。表 2 はユーザの移動行動として、移動距離を 2km 間隔、移動時間を 900 秒間隔でそれぞれ 11 段階の評価条件にトリップを分類し、評価条件ごとに F 値の中央値を求めた結果である。本結果から全手法共通して、移動距離および移動時間の増加にともない、F 値が向上していることが分かる。また、提案手法と各手法を比較すると提案手法が移動距離 2 km 以上、移動時間 900 秒以上において、

最も高い F 値であることを明らかにした。ただし、2 km 未満の移動距離、900 秒未満の移動時間に関しては半数以上のトリップで移動状態を推定することができなかった。この原因として、移動距離が特定の基地局のカバーエリア内に収まり、他の基地局にハンドオーバーしなかったことが考えられる。

表 3 は通信行動として、移動時の CDRs 数を 20 レコード間隔、通信時間間隔の標準偏差を 20 秒間隔、通信距離間隔の標準偏差を 100 m 間隔でそれぞれ 11 段階の評価条件にトリップを分類し、評価条件ごとに F 値の中央値を求めた結果である。本結果から、時空間の標準偏差は全手法影響がないこと、移動時の通信数は多い方が F 値が高いことを確認した。また、通信行動に関しても提案手法が多くの評価条件で Kang らの手法、Liu らの手法に比べ高い F 値であることを明らかにした。ただし、通信数が 15 回未満、通信時間間隔の標準偏差が 20 秒未満、通信距離間隔の標準偏差が 200 m 未満の評価条件に関してはトリップの半数以上が推定できないことを確認した。これらもユーザの移動行動と同様に移動時間や移動距離が非常に短いトリップが原因と考えられる。

表 4 は移動中に接続した各基地局の周囲 1 km に存在する基地局数の中央値を 5 局間隔で 11 段階の評価条件に分類し、評価条件ごとに F 値の中央値を求めた結果である。本結果から基地局数によって Kang らの手法は F 値が 0~0.5、Liu らの手法は F 値が 0.4~0.667 に変動しているのに対し、提案手法は 0.667~0.778 と他手法に比べ変動が小さく、基地局の粗密の影響が小さいことが分かる。また、F 値に関しても提案手法が全評価条件で最も高い結果であることを確認した。

5.3 複数のデータセットに基づく評価

本節では「ユーザの移動行動と通信行動」、「ユーザの移動行動と通信環境」、「通信行動と通信環境」を組み合わせたデータセットの比較、評価を行う。なお、各データセットは 11 段階の評価条件に分類し、組み合わせたデータセットは 1 組あたり 121 の評価条件で評価を行う。また、各

表 3 通信行動に基づくデータセット

Table 3 Dataset constructed from communication behavior.

CDR 数	0	15	30	45	60	75	90	105	120	135	150-
トリップ数	4374	5626	3564	2416	1759	1373	1081	824	640	536	4259
Kang らの手法	0	0	0.5	0.5	0.571	0.615	0.666	0.6	0.667	0.645	0.6
Liu らの手法	0	0.325	0.5	0.667	0.667	0.715	0.727	0.714	0.727	0.727	0.667
提案手法	0	0.5	0.667	0.75	0.8	0.833	0.857	0.818	0.857	0.857	0.842
標準偏差 (時間) [sec]	0	20	40	60	80	100	120	140	160	180	200-
トリップ数	821	7063	7893	4348	1910	1035	566	410	288	184	1865
Kang らの手法	0	0.5	0.5	0.4	0.429	0.445	0.5	0.444	0.4	0.5	0.5
Liu らの手法	0	0.571	0.571	0.5	0.5	0.5	0.589	0.571	0.558	0.534	0.615
提案手法	0.161	0.667	0.667	0.667	0.667	0.667	0.75	0.667	0.667	0.667	0.769
標準偏差 (空間) [m]	0	100	200	300	400	500	600	700	800	900	1000-
トリップ数	2892	2495	3679	3993	3242	2397	1748	1302	856	669	3110
Kang らの手法	0	0	0.333	0.5	0.6	0.667	0.625	0.667	0.667	0.667	0.667
Liu らの手法	0	0	0.445	0.666	0.667	0.667	0.667	0.667	0.75	0.715	0.727
提案手法	0	0	0.667	0.715	0.8	0.8	0.8	0.824	0.857	0.828	0.842

表 4 通信環境に基づくデータセット

Table 4 Dataset constructed from communication environment.

接続局周辺の基地局数	0	5	10	15	20	25	30	35	40	45	50-
トリップ数	11081	4998	2574	1293	837	551	373	554	388	622	3181
Kang らの手法	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.333	0.5	0.4	0
Liu らの手法	0.5	0.667	0.615	0.667	0.631	0.545	0.632	0.4	0.667	0.5	0.4
提案手法	0.667	0.714	0.667	0.714	0.667	0.667	0.714	0.667	0.778	0.667	0.667

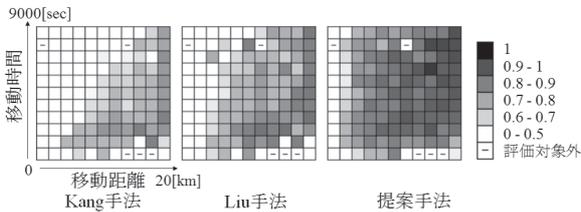


図 11 移動距離と移動時間に基づくデータセット

Fig. 11 Dataset constructed from moving distance and moving time.

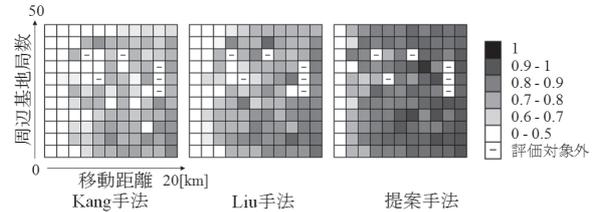


図 12 移動距離と周辺基地局数に基づくデータセット

Fig. 12 Dataset constructed from moving distance and number of neighbor base stations.

データセットで最も F 値が高かった評価条件の数と、最も F 値が高かった手法と次に高かった手法の F 値の差が 0.15 以上あった評価条件の数を求める。ただし、トリップ数が 10 以下である評価条件は稀なトリップ、あるいは、異常なデータの影響が考えられるため評価対象外とした。図 11 はデータセットの組合せとして、移動距離と移動時間における各手法の評価結果の分布図を示している。各分布図のセルは評価条件に対応し、左から右にかけて移動距離が長く、下から上にかけて移動時間が長く、セルの色は推定結果の精度を白黒の濃淡で示している。ここで、セルの色は黒に近いほど F 値が高く、色が白に近いほど F 値が低いことを示しており、トリップ数が 10 以下の評価対象外に関してはセル内に“-”を表示している。本図から提案手法は他手法に比べ、広範囲かつ高精度に移動状態を推定できていることが分かる。具体的な評価結果としては 121 の評

価条件の中で評価対象となった評価条件は 116 あり、Liu らの手法は 1、提案手法は 99 の評価条件で最も高い F 値となり、16 の評価条件では有意差が確認できなかった。また、評価対象全体の F 値の中央値は Kang らの手法が 0.52、Liu らの手法が 0.67 であったのに対し、提案手法は最も高い 0.8 となった。同様に図 12 は移動距離と周辺の基地局数の評価結果を示しており、評価対象全体の F 値の中央値は Kang らの手法が 0.6、Liu らの手法が 0.67 に対し、提案手法は最も高い 0.8 となった。また、図 13 は移動距離と標準偏差 (時間) の評価結果を示しており、評価対象全体の F 値の中央値は Kang らの手法が 0.5、Liu らの手法が 0.67 に対し、本評価でも提案手法は最も高い 0.76 であった。一方で、図 11、図 12、図 13 に共通して、移動距離が 2km 以内の短い移動に関してはトリップの半数以上が推定できていないことを確認した。これらの原因も移動距

表 5 複数のデータセットの評価結果

Table 5 Evaluation results of two types of datasets.

データセット 1	データセット 2	評価条件の数	Kang らの手法	Liu らの手法	提案手法	有意差無	評価対象外
移動距離	移動時間	116	0	1	99	16	5
移動距離	CDR 数	118	2	1	99	16	3
移動距離	接続局周辺の基地局数	115	0	1	100	14	6
移動距離	標準偏差 (時間)	107	0	0	94	13	14
移動距離	標準偏差 (空間)	111	0	1	102	8	10
移動時間	CDR 数	103	1	2	95	5	18
移動時間	接続局周辺の基地局数	120	0	1	103	16	1
移動時間	標準偏差 (時間)	111	1	0	99	11	10
移動時間	標準偏差 (空間)	120	0	0	103	17	1
標準偏差 (時間)	接続局周辺の基地局数	100	0	5	88	7	21
標準偏差 (空間)	接続局周辺の基地局数	121	0	2	99	20	0
標準偏差 (時間)	標準偏差 (空間)	116	0	3	89	24	5
標準偏差 (時間)	CDR 数	108	0	1	91	16	13
標準偏差 (空間)	CDR 数	121	0	0	103	18	0
接続局周辺の基地局数	CDR 数	121	2	3	96	20	0
合計		1708	6	21	1460	221	107

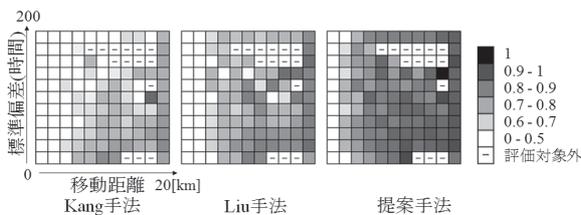


図 13 移動距離と標準偏差 (時間) に基づくデータセット

Fig. 13 Dataset constructed from moving distance and distribution of standard deviation (time).

離が特定の基地局のカバーエリア内に収まり、他の基地局にハンドオーバーしなかったことが考えられる。

最後に表 5 は複数のデータセットの全組合せの評価結果であり、各手法において最も高い F 値となった評価条件の数、有意差がなかった評価条件の数、評価対象外であった評価条件の数を示している。また、合計は評価対象の評価条件の数、Kang らの手法、Liu らの手法、提案手法、有意差なしの合計の評価条件の数を示している。本結果から全組合せにおいて、評価を行った 1,708 の評価条件のうち、提案手法は 1,460 (85%) の評価条件で Kang らの手法と Liu らの手法より高精度に移動状態を推定できることを確認した。また、提案手法が最も高精度であった 1,460 の評価条件のうち、921 の評価条件で次に精度が高かった手法との F 値の差が 0.15 以上あることを確認した。

6. おわりに

本論文では位置精度が GPS より悪く、取得頻度が不規則な CDRs の特徴に適した移動滞在状態推定手法を提案、評価した。事前承諾を得た 582 名の 2014 年 3 月の 1 カ月分の CDRs と GPS から生成した 26,452 のトリップに対し

て、既存技術である Kang らの手法、Liu らの手法と提案手法の移動滞在状態推定の精度を評価した。評価結果から滞在状態の推定では全手法 0.85 以上の推定精度であったが、移動状態は Liu らの手法と Kang らの手法が 0.5 であったのに対し、提案手法は 0.667 となった。また、移動時間に関しては、ユーザの移動行動、通信行動、通信環境の各種特徴で分類した評価データセットをもとに評価した。結論として、提案手法は 1,708 の評価条件のうち、1,460 (85%) の評価条件で既存手法に比べ、高い推定精度であった。ただし、移動時間が短いトリップや移動距離が少ないトリップに対しては既存手法と変わらない結果であることを確認した。本原因として、移動距離が特定の基地局のカバーエリア内に収まり、他の基地局にハンドオーバーしなかったこと、位置情報である CDRs 数が少なかったことが考えられる。これらの問題に対し、過去の自身の CDRs あるいは行動が類似している他人の CDRs を利用した補正等を検討しながら高度化を進める。

謝辞 本技術は、総務省の『G 空間プラットフォームにおけるリアルタイム情報の利活用技術に関する研究開発』による委託を受けて実施した研究開発による成果です。

参考文献

- [1] Ashbrook, D. and Starner, T.: Using GPS to learn significant locations and predict movement across multiple users, *Journal of Personal and Ubiquitous Computing*, Vol.7, No.5, pp.275-286 (2003).
- [2] Ester, M., Kriegel, H.P., Sander, J. and Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise, *Proc. 2nd International Conference on Knowledge Discovery and Data Mining*, pp.226-231 (1996).
- [3] Nurmi, P. and Bhattacharya, S.: Identifying meaning-

ful places: The non-parametric way, *Proc. 6th International Conference on Pervasive Computing*, pp.111-127 (2008).

[4] Kang, J.H., Welbourne, W., Stewart, B. and Borriello, G.: Extracting places from traces of locations, *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol.9, No.3 (2005).

[5] Bayir, M.A., Demirbas, M. and Eagle, N.: Mobility profiler: A framework for discovering mobility profiles of cell phone users, *Proc. International Conference on Pervasive and Mobile Computing*, Vol.6, No.4, pp.435-454 (2010).

[6] Liu, J., Wolfson, O. and Yin, H.: Extracting Semantic Location from Outdoor Positioning Systems, *Proc. 7th International Conference on Mobile Data Management (MDM 2006)*, p.73 (2006).

[7] Kurokawa, M., Watanabe, T., Muramatsu, S., Kanasugi, H., Sekimoto, Y. and Shibasaki, R.: Extracting People's Stays from Cellular Network Data, *3rd International Conference on the Analysis of Mobile Phone Datasets (NetMob)* (2013).

[8] Huang, C., Ying, J. and Tseng, V.: Mining User's Behaviors and Environments for Semantic Place Prediction, *The Mobile Data Challenge 2012 (by Nokia) Workshop*, June 18-19 (2012).

[9] Zhu, Y., Zhong, E., Lu, Z. and Yang, Q.: Feature Engineering for Place Category Classification, *The Mobile Data Challenge 2012 (by Nokia) Workshop*, June 18-19 (2012).

[10] Isaacman, S., Becker, R., Cáceres, R., Kobourov, S.G., Martonosi, M., Rowland, J. and Varshavsky, A.: Identifying Important Places in People's Lives from Cellular Network Data, *Proc. 9th International Conference on Pervasive Computing*, pp.133-151 (2011).

[11] Umair, M., Kim, W.S., Choi, B.C. and Jung, S.Y.: Discovering personal places from location traces, *Proc. ICACT '14*, IEEE, pp.709-713 (2014).

[12] Yang, P., et al.: Identifying Significant Places Using Multi-day Call Detail Records, *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, IEEE (2014).

[13] 黒川茂莉, 横山浩之, 吉井和佳, 麻生英樹: 携帯電話通信時に得られる疎な位置情報履歴を用いた有意位置検出, *電子情報通信学会論文誌*, Vol.J95-D(4), pp.722-733 (2012).

[14] Zhu, Y., Sun, Y. and Wang, Y.: Predicting Semantic Place and Next Place via Mobile Data, *The Mobile Data Challenge 2012 (by Nokia) Workshop*, June 18-19 (2012).

[15] Zhu, Y., Zhong, E., Lu, Z. and Yang, Q.: Feature Engineering for Place Category Classification, *The Mobile Data Challenge 2012 (by Nokia) Workshop*, June 18-19 (2012).

[16] Fan, Z., Xuan, S., Shibasaki, R. and Adachi, R.: CityMomentum: An Online Approach for Crowd Behavior Prediction at a Citywide Level, *Proc. Ubicomp* (2015).

[17] Manoranjan, D., et al.: Next place prediction by understanding mobility patterns, *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, IEEE (2015).

[18] Zhang, Y.: User mobility from the view of cellular data networks, *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, IEEE (2014).

[19] Donna, D., Aniello, S., Cantelmo, G. and Viti, F.: A Markov chain dynamic model for trip generation and dis-

tribution based on CDR, *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, IEEE (2015).

[20] Becker, R.A., et al.: Clustering anonymized mobile call detail records to find usage groups, *PURBA*, San Francisco, CA, USA (2011).

[21] Shimosaka, M., Maeda, K., Tsukiji, T. and Tsubouchi, K.: Forecasting Urban Dynamics with Mobility Logs by Bilinear Poisson Regression, *Proc. Ubicomp* (2015).

[22] Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A. and Willinger, W.: Human mobility modeling at metropolitan scales, *Proc. 10th international conference on mobile systems, applications, and services*, ACM (2012).

[23] Felde, I., et al.: Characterization of urban traffic by using mobile phone traces, *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE (2014).



小林 直

2011年早稲田大学大学院国際情報通信研究科情報通信システム学専攻博士課程修了。同年 KDDI 株式会社入社。現在、株式会社 KDDI 総合研究所データマイニング応用グループ研究主査。位置推定, 行動認識, 行動モデリング, データマイニング等の研究開発に従事。電子情報通信学会会員。博士 (国際情報通信学)。



石塚 宏紀 (正会員)

2012年東京大学大学院情報理工学系研究科電子情報学専攻博士課程修了。同年 KDDI 株式会社入社。現在、株式会社 KDDI 総合研究所データマイニング応用グループ研究主査。通信ログや位置情報のデータマイニング技術, 行動分析の応用技術等の研究開発に従事。博士 (情報理工学)。



南川 敦宣

2005年京都大学大学院情報学研究科通信情報システム専攻修士課程修了。同年 KDDI 株式会社入社。現在、株式会社 KDDI 総合研究所データマイニング応用グループ研究マネージャー。行動認識, ユビキタス, ITS, 放送通信融合等の研究開発に従事。電子情報通信学会 2009 年度学術奨励賞受賞。電子情報通信学会会員。



村松茂樹 (正会員)

1999年東京大学大学院工学系研究科電子情報工学専攻修士課程修了。同年KDD株式会社(現, KDDI株式会社)入社。現在, 株式会社KDDI総合研究所フューチャーデザイン2部門3グループアナリスト。位置推定, 行動認識, データマイニング, データベースシステム等の研究開発に従事。本会2011年度論文賞受賞。



小野 智弘 (正会員)

1992年慶應義塾大学理工学部電気工学科卒業。1994年同大学大学院理工学研究科修士課程計算機科学専攻修了。同年KDD株式会社(現, KDDI株式会社)入社。1999年9月~2000年9月スタンフォード大学電気工学科客員研究員。2004年4月~2008年3月独立行政法人産業技術総合研究所情報技術研究部門協力研究員(併任)。現在, 株式会社KDDI総合研究所グリーン・クラウド部門長兼データマイニング応用グループグループリーダー。この間, 行動分析, 利用者嗜好抽出, ソーシャルメディア解析, データベース等の研究に従事。電子情報通信学会(シニア)会員。博士(工学)。

(担当編集委員 鷹野 孝典)