

携帯電話通信履歴による位置情報を用いた 次滞在地推定とその評価

石塚宏紀¹ 黒川 茂莉¹ 渡邊 孝文¹ 村松 茂樹¹ 小野 智弘¹

概要：昨今、スマートフォンに搭載された各種センサを用いた位置履歴情報の収集が容易になったため、ユーザの行動予測に関する研究が注目を集めている。しかしながら、GPS センサ等による定常的な位置履歴情報取得は、スマートフォンの電池を著しく浪費するため、行動予測の基礎データである位置情報履歴の大規模な収集は困難であり、小規模のユーザに対する行動予測の研究しか行われていない。一方で、移動体網を経由して取得される携帯電話通信時に得られる位置情報履歴の利活用は、全国を網羅的且つリアルタイムに人々の移動が把握可能であるため災害時における首都圏の避難対策や都市交通施策等さまざまな分野で期待されている。そこで、本研究では、携帯電話の通信履歴による位置情報履歴を用いて大規模な行動予測を実施した一例を示し、その可能性について考察する。本論では、一都六県在住の18～60歳のスマートフォン利用者に個別許諾を得た約2000名分の1ヶ月間(2013年2月及び2014年3月において各個別の約1000名分)における位置情報履歴に基づいて各ユーザの次滞在地推定手法とその評価について述べる。

A Next Place Estimation Prediction using Individual Mobile Communication Logs

HIROKI ISHIZUKA¹ MORI KUROKAWA¹ TAKAFUMI WATANABE¹ SHIGEKI MURAMATSU¹
CHIIHIRO ONO¹

1. まえがき

昨今、スマートフォンに搭載された各種センサを用いた位置履歴情報の収集が容易になったため、ユーザの行動予測に関する研究 [1] が注目を集めている。しかしながら、GPS センサ等による定常的な位置履歴情報取得は、スマートフォンの電池を著しく浪費するため、行動予測の基礎データである位置情報履歴の大規模な収集は困難であり、小規模のユーザに対する行動予測の研究しか行われていない。一方で、移動体網を経由して取得される携帯電話通信時に得られる位置情報履歴の利活用 [2][3] が注目を集めている。大規模なユーザの位置情報履歴を活用することによって、各ユーザの滞在地・滞在時間を分析し、特定エリアにおける時間帯ごとの人口分布等の時空間統計情報が作

成可能となった。こうした時空間統計情報は、移動体通信事業者の通信エリア設計最適化から災害時における首都圏の避難対策や都市交通施策等への応用が期待されている。防災分野における利活用の例として、NTTドコモは、工学院大学と共同研究 [4] を行い、首都直下地震が発生した場合の帰宅困難者数の推計と各市区町村の特徴や課題抽出、対策の検討をあわせて行うことで、東京都における帰宅困難者数の推計に活用する取組を行っている。

また、行政における都市交通施策のあり方を検討する上で、都市特性とその交通特性との関係を把握し、都市特性に応じた交通施策を準備するため、国土交通省を始め各自治体ではパーソントリップ調査 (PT 調査) を定期的に実施している。PT 調査では、交通行動の出発地、到着地、目的、行動時間帯など1日の詳細な交通データの大規模な収集を目的としている。しかし、大都市圏や全国レベルのPT 調査では、1回の調査に3～4年を要し、実施費用として数10億円程度の予算が必要となり、リアルタイムな交

¹ 株式会社 KDDI 研究所
KDDI R&D Laboratories, Inc., 2-1-15 Ohara, Fujimino,
Saitama, 356-8502 Japan

通状況の把握や常時観測による費用の平準化が喫緊の課題となっている。そこで、国内において網羅的に収集可能な携帯電話通信時に得られる位置情報履歴の統計的な分析がPT調査の課題解決に期待されている。網羅的な行動分析において、人々の移動を予測し、個々人の移動をモデル化することで利活用が促進されると考えられる。しかしながら、既存の携帯電話通信時に得られる位置情報履歴による行動分析では、各ユーザの滞在地・滞在時間の推定技術に留まり、ユーザの次滞在地推定を含めた移動のモデル化技術の研究はいまだ不十分である。本稿では、携帯電話の通信履歴による位置情報履歴を用いて大規模な行動予測を実施した一例を示し、その可能性について考察し、一都六県在住の18~60歳のスマートフォン利用者に個別許諾を頂いた約2000名分の1ヶ月間(2013年2月及び2014年3月)において各約1000名分)における位置情報履歴に基づいて各ユーザの次滞在地推定を実施し、その評価について述べる。

2. 関連研究

本章では、位置情報履歴から滞在地・滞在時間を検出する既存研究を述べ、それらの情報を基に次滞在地推定を行う研究を述べた後、本研究の位置づけについて述べる。

2.1 滞在地・滞在時間検出

位置情報履歴からの滞在地・滞在時間検出を行う手法は、位置情報を基準としたクラスタリング手法 [6][7][8][9] と時間情報と位置情報の両方を基準としたクラスタリング手法 [10][11][12][13] がある。このうち、[10][11] は位置をIDとして表現し、[6][7][8][9][12][13] は位置を2次元空間上の点として表現する。また、位置情報履歴の情報源として、[6][7][8][9] はGPSによる定期的な位置情報履歴を用い、[12] はWiFiによる比較的密な位置情報履歴を用い、[10][11][13] は通信時の情報のみを利用する疎な通信履歴を用いている。滞在地・滞在時間検出は、提案手法の滞在地推定においてデータ前処理として実行され、滞在地推定の結果に大きな影響を与える。本論における滞在地・滞在時間検出では、位置情報履歴の情報源として通信履歴を想定し、位置と時間の両方を考慮したクラスタリング手法である [13] を利用する。

2.2 次滞在地予測

山田等 [1] は、滞在地を訪問地と経由地として抽出し、移動経路を推定することで、次滞在地を推定している。Ying等 [15] は、GIS情報と滞在地における位置情報の両方を活用して次滞在地推定し、推定手法として、類似行動ユーザのクラスタリングを行うことで予測を行っている。また、Scellato等 [14] は、NextPlaceという滞在時間に着目した非線形時系列分析によって次滞在地推定を行い、次滞在地

とその場所での滞在時間を予測可能としている。[16] は、自動車移動に関するGPS位置情報履歴を活用し、自動車利用者の日常的な走行経路や行動から走行目的地を推定している。Pianese[17]等は、Foursquareなどのソーシャルメディアによるチェックイン情報を利用した行動分析を行い、次滞在地推定を試みている。いずれも個々人の次滞在地推定を行っているが、推定に利用している位置情報履歴情報は、GPSなどによる高頻度・高精度な位置情報もしくは、より高次のコンテキスト情報を活用している。本論では、携帯電話通信履歴による位置情報のような時空間的に疎な位置情報履歴を活用して、個々人の次滞在地推定を一般的な手法で実施し、その精度評価から活用の可能性について議論する。

3. 行動予測手法

3.1 滞在地/滞在時間抽出

本論において、次滞在地推定においてもまず位置情報履歴から滞在地・滞在時間を検出することを前提としている。本論で用いる滞在地・滞在時間検出手法は、筆者ら [13] が考案した、時間的に近接するデータ区間内の位置情報の分布に着目し、単峰である場合に滞在地と判定する手法を利用する。本手法では、一定の時間幅(T)のタイムウィンドウに、一定数(N_{min})以上、位置が観測された場合に、位置の集合に対してMeanShift法 [18] を適用し、滞在地・滞在時間を判定する。図1に示すようにMeanShift法は、連続変数に関する分布のモードを検出する手法であり、各位置に対して、 Th_1 の円周内に収まる位置の集合に対して窓関数を繰り返し適用し、収束した点が Th_2 の円周内に収まる場合、分布のモードが1つである、すなわち、分布が単峰であるとみなすことができる。この条件を満たす場合に、滞在地候補点と判定する。窓関数は矩形カーネルを利用する。滞在地候補点に再度、MeanShift法を適用し、検出されたモードを最終的な滞在地とする。

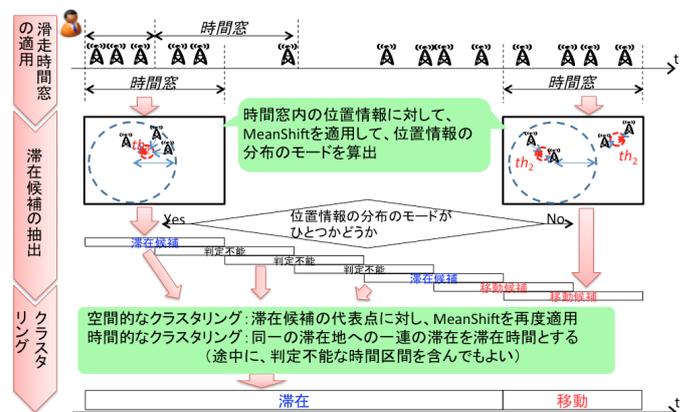


図1 MeanShift法を用いた滞在地・滞在時間検出手法

3.2 次滞在地推定

本論における次滞在地推定は、滞在地・滞在時間抽出において明らかとなった各ユーザの滞在地を1日単位で分割し、1ヶ月間における各日の滞在地遷移をカウントし、時空間的な遷移確率を算出する。算出された遷移確率に基づいて k 次マルコフモデルにて次滞在地推定を行う。本論の滞在地推定手法で抽出した2週間分のあるユーザにおける滞在地遷移の様子を??に示す。各ユーザの滞在地履歴情報 $H = [l_1, l_2, \dots, l_m]$ に基づいた k 次マルコフモデルにおいてあるユーザの一連の行動は、 $c = H(m - k + 1, m)$ と定義できる。現在の滞在地 l_m から次の滞在地 l_{m+1} への遷移確率は、 $P(l_{m+1} = l | c, H)$ にて与えられる。1ヶ月間の位置情報履歴の内、前半2週間のデータをモデル生成の学習データとして利用し、残り後半2週間分のデータを用いて生成したモデルの精度評価を実施するものとする。尚、平日と休日では、行動が大きく変化するため、本モデルでは、平日のみの位置情報履歴のみを利用した。

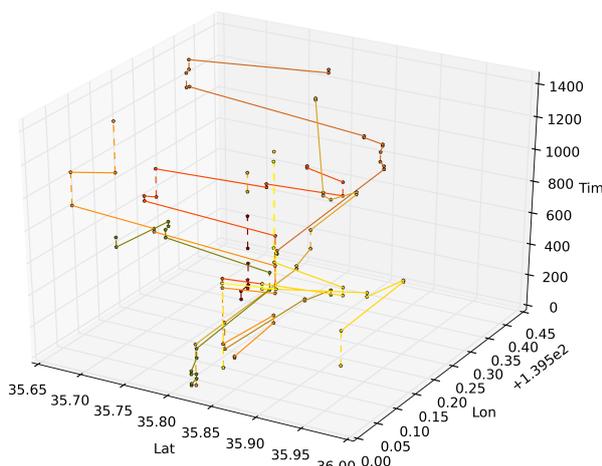


図2 あるユーザにおける2週間分の滞在地間遷移の様子

4. 利用データセットとその特徴

本論にて利用する携帯電話通信履歴^{*1}による位置情報は、移動通信インフラの運用・保守のために移動体網を経由して取得される通信に関するイベントログであり、ログ取得のタイミングは、主に通話、SMS、データの送受信時である。取得データとしては、端末またはユーザの識別子、通話/通信の開始終了時刻、通信時に利用した基地局の情報となる。図3に携帯電話通信履歴による位置情報取得の流れを示す。

データ収集の被験者として一都六県在住の18歳(高校生は除く)~60歳の男女にて、スマートフォンをご利用の方

^{*1} 本実験における携帯電話通信履歴は、総務省の研究会において、事業者による自主的なガイドライン等を策定のための指針となる利用者情報の活用の際の「配慮原則」に則って移動や行動の研究分析を行うことを明記した事前同意を取得した上で利用している。



ユーザ識別子	通信開始時	基地局緯度	基地局経度	通信終了時
AAAAAA	2013/11/11 11:11:11	35.9999	139.9999	2013/11/11 11:11:21
AAAAAA	2013/11/11 12:12:12	34.9999	138.9999	2013/11/11 12:12:22
AAAAAA	2013/11/11 13:13:13	35.1111	139.1111	2013/11/11 13:13:23

図3 携帯電話通信履歴による位置情報取得の流れ

募集を行い、事前に個別同意を得たうえで通信履歴を2013年2月1日から2月28日までの約1ヶ月間に1250人から取得し、2014年3月1日から3月31日までの約1ヶ月間に1000人から取得し、合計で約2000名分の1ヶ月間における移動履歴を基に、前半2週間のデータをモデル生成の学習データとして利用し、残り後半2週間分のデータを用いて生成したモデルの精度評価を行った。図4に実験参加者の性別・年齢属性分布を示す。尚、2013年におけるデータは、3G通信のみに対応した端末を対象として実施し、2014年は3G/LTE両方に対応した端末を対象に実施した。通信方式による出力のタイミング上、2014年におけるデータの方が、データ数が多い状況である。

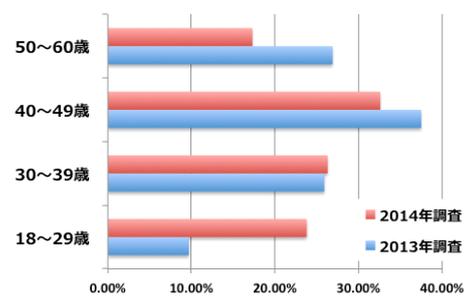
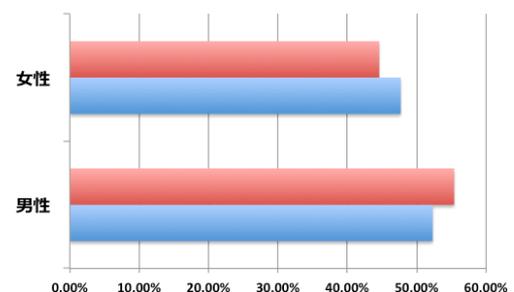


図4 利用データの性別・年齢属性分布

5. 評価実験

5.1 滞在地/滞在時間抽出

滞在地・滞在時間の検出手法におけるパラメータは、 $T = 5$ 分、 $N_{min} = 4$ 、 $Th1 = 4000m$ 、 $Th2 = 1000m$ として抽

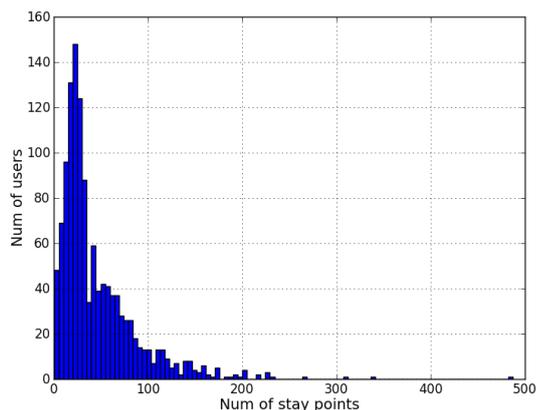


図 5 2013 年 2 月における調査データの滞在地抽出結果

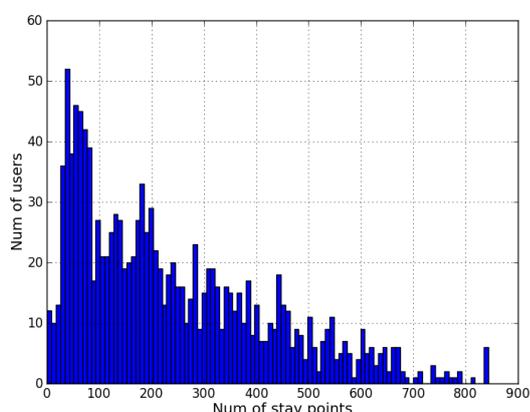


図 6 2014 年 3 月における調査データの滞在地抽出結果

出した。本パラメータ設定は、[13]での評価実験により通信履歴に対して適した設定であると検証された結果を利用している。2013 年におけるデータでは、平均滞在地抽出数が、46 地点であり、図 5 のような分布であった。また、2014 年におけるデータでは、平均滞在地抽出数が、240 地点であり、図 6 のような分布であった。2013 年におけるデータは、3G 通信時のみにおける通信履歴を利用していることからデータ量も 2014 年と比べて少なく、その点において、滞在地抽出量も減ったものと考えられる。一方で、2014 年におけるデータは、LTE 通信も含めた通信履歴となり、且つスマートフォンの急激な普及やアプリの多様性により 2013 年におけるデータと比較してユーザ毎の履歴数が多く、その分多地点での滞在地が抽出されたものと考えられる。

5.2 次滞在地推定

滞在地推定では、2 年間のデータに対して、平日における次滞在地を単純マルコフモデルを用いて予測し、1 ヶ月間の位置情報履歴の内、前半 2 週間のデータをモデル生成の学習データとして利用し、残り後半 2 週間分のデータを

用いて生成したモデルの精度評価を行った。尚、正解データとしての滞在地についても滞在地推定結果を利用するものとし、予測正解判定は、次滞在地と判定された滞在地と正解である滞在地の空間的距離が、1000m 以内、時間的距離が、3 時間以内となったものを正解として判定した。これらの指標は、滞在地判定のパラメータ及び、滞在時間より算出したものである。予測結果として、2013 年のデータに対する予測精度の平均値は、55%であり、2014 年度のデータに対する予測精度は、51.9%であった。2013 年、2014 年のデータに対する単純マルコフモデルでの精度検証結果を、それぞれ図 7、図 8 に示す。また、同様に 2 次マルコフにおける予測精度も検証し、2013 年のデータで、56.6%、2014 年のデータで、53%と多少ではあるが精度向上が見られる結果となった。

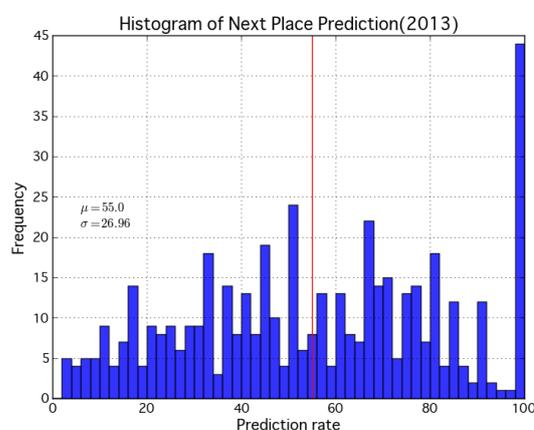


図 7 2013 年 2 月における調査データの次滞在地推定結果

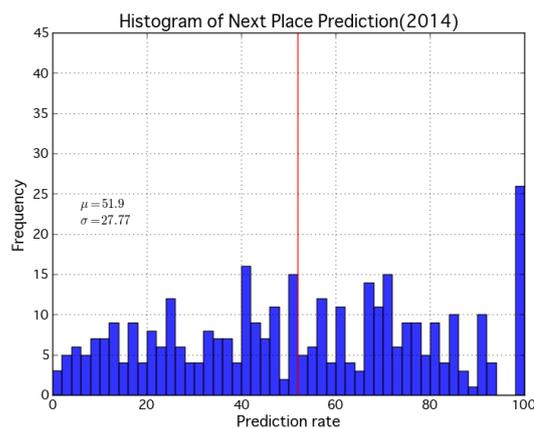


図 8 2014 年 3 月における調査データの次滞在地推定結果

6. まとめ

本論では、携帯電話通信履歴による位置情報を用いた次滞在地推定の評価を行い、2013 年及び 2014 年の 2 回に分けて各々 1 ヶ月間におよび取得された約 2000 名の通信履

歴から滞在地・滞在時間抽出を行い、マルコフモデルでの次滞在地推定を行った。評価実験では、予測精度の平均値が、50%程度であり、且つ、正解として判定された時空間範囲も1km及び3時間の空間と、かなり大きなものであった。また、精度の極端に高いユーザについて再度調査したところ、抽出された滞在地が極端に少ないため、予測が容易だったと推測される。今後は、モデル化手法を改変し、滞在地への意味付けや通信履歴ならではのコンテキストを利用することで、精度向上を試みる。

参考文献

- [1] 山田直治, 磯田佳徳, 南正輝, 森川博之, 屋外行動支援のためのGPS搭載携帯電話を用いた移動経路の逐次的精練手法, 情報処理学会論文誌 Vol.52 No.6, pp1951 - 1967(2011).
- [2] R. Becker, R. Caceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, C. Volinsky, Human mobility characterization from cellular network data, Communications of the ACM, vol.56 no.1, January 2013
- [3] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. Nature, 453, June 2008.
- [4] 村上正浩, 岡島一郎, 鈴木俊博, 山下仁, “7413 モバイル空間統計を活用した滞留者・帰宅困難者数の推定と具体的対策の検討(避難行動, 都市計画),” 学術講演梗概集, F-1, 都市計画, 建築経済・住宅問題 2011, pp.893-894, 2011.
- [5] パーソントリップ調査データ概要, 東京都圏交通計画協議会, <http://www.tokyo-pt.jp/data/01.html>
- [6] D. Ashbrook, and T. Starner, 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.
- [7] M. Ester, H. P. Kriegel, Jörg Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining, pp.226-231, 1996.
- [8] P. Nurmi, and S. Bhattacharya, “Identifying meaningful places: The non-parametric way,” Proc. of the 6th International Conference on Pervasive Computing, pp.111-127, 2008.
- [9] S. Isaacman, R. Becker, R. Caceres, S.G. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky, “Identifying Important Places in People’s Lives from Cellular Network Data,” Proc. of the 9th International Conference on Pervasive Computing, pp.133-151, 2011.
- [10] M.A. Bayir, M. Demirbas, and N. Eagle, “Mobility profiler: A framework for discovering mobility profiles of cell phone users,” Proc. of the International Conference on Pervasive and Mobile Computing, vol.6, no.4, pp.435-454, 2010.
- [11] 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,” 本会論文誌 J95-D(4), pp.722-733, 2012.
- [12] J. Liu, O. Wolfson, H. Yin, “Extracting Semantic Location from Outdoor Positioning Systems,” Proc. of the 7th International Conference on Mobile Data Management (MDM 2006), pp.73, 2006.
- [13] M. Kurokawa, T. Watanabe, S. Muramatsu, H. Kanasugi, Y. Sekimoto, and R. Shibasaki, “Extracting People’s Stays from Cellular Network Data,” Third International Conference on the Analysis of Mobile Phone Datasets(NetMob 2013), 2013.
- [14] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, “Nextplace: a spatio-temporal prediction framework for pervasive systems,” in Pervasive computing, 2011, pp. 152-169.
- [15] J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng, “Semantic trajectory mining for location prediction,” in Proc. of ACM SIGSPATIAL GIS ’11, New York, NY, USA, Nov. 2011, p. 34
- [16] J. Krumm and E. Horvitz, “Predestination: Inferring destinations from partial trajectories,” In Eighth International Conference on Ubiquitous Computing (UbiComp 2006), Jan 2006.
- [17] F. Pianese, X. An, F. Kawsar, and H. Ishizuka. Discovering and predicting user routines by differential analysis of social network traces. In WoWMoM, 2013.
- [18] Y. Cheng, “Mean Shift, Mode Seeking, and Clustering,” IEEE Trans. Pattern Anal. Mach. Intell, Vol. 17, No. 8, pp. 790-799, 1995.