

ダッカ首都圏における匿名化された携帯電話データ分析による ユーザーの世帯属性推計に関する研究

新井亜弓・ウィタヤンクーン・アピチョン・邵肖偉・柴崎亮介

Understanding Household Attributes of Mobile Phone Users from Anonymized Call Detail Records: A Case Study of Greater Dhaka in Bangladesh

Ayumi ARAI, Apichon WITAYANGKURN, Xiaowei SHAO, and Ryosuke SHIBASAKI

Abstract: This study provides a clue to understand household attributes of mobile phone users analyzing calling behavior from anonymized Call Detail Records (CDR). Statistical analysis results on calling locations indicate there are distinctive differences in females' trends on non-primary activity day according to the presence of kids within their households. Our findings can contribute to understand whereabouts of kids who are considered to be vulnerable to disasters but difficult to be located from CDR.

Keywords: 携帯電話データ (Call Detail Records, CDR), 世帯属性 (household attributes), 端末使用パターン (calling behavior)

1. はじめに

携帯電話やタブレット等のモバイル端末の普及により, GPS データや CDR (Call Detail Records, 携帯電話データ) 等の大規模人口の流動を捉えたデータが蓄積されるにつれ, 人間の行動パターンや移動軌跡に関する研究は近年急速に発展している。特に大規模人口の流動については, 行動パターン抽出やデータマイニングに関する手法開発が進んでいる。González *et al.* (2008) は, 人間は社会的規範に則った生活を送ることから, 結果として限られた特定の場所を繰り返し行き来することを明らかにしている。それらの場所を時間軸と組合せて解析することにより, 自宅や職場, 学校等に滞在する時間分布を明らかにし, 軌跡から活動パターンを推計することも行われている (Isaacman *et al.*, 2011)。

これらの研究に用いられるデータは, 大規模人口の流動を捉える上で有用である一方で, 端末利用者しか含まれないことにより生じる偏りや, 匿

名化されているためデータが軌跡の集積を表しているに過ぎないという点も指摘されている (Arai & Shibasaki, 2014)。近年, 匿名化されたユーザーの性別等の属性を, 端末使用パターン分析により推計しようとする試みは, 徐々に始められている (Brdar, *et al.* 2012; Nadeem, 2012)。本研究では, CDR が携帯電話使用者 (以下, 端末使用者) しか含まないという点を利用し, 端末使用者の端末使用傾向から, 端末使用者が属する世帯属性推計に必要な特徴量の抽出を試みる。特徴量は, バングラデシュのダッカ首都圏にて実施した端末使用傾向と属性に関するインタビュー調査データと CDR を用いて考察, 抽出する。

2. データ

2.1 携帯電話データ (CDR)

本研究では, バングラデシュにおける主要携帯電話事業者一社の携帯端末利用者の CDR を用いる。CDR は, 2013 年 11 月～12 月の通話発信記

録の時刻、通信した基地局の位置情報、通信時間で構成されており、これに世帯構成に関する情報を追加収集した。

2.2 インタビュー調査データ

2013年11～12月にかけて、前出の携帯電話事業者の端末利用者を少なくとも一名含む810世帯を対象とした調査を行った。標本抽出には、土地利用と所得レベルの二段階層化抽出法を用い、世帯構成、構成員の性別、年齢、職業等の属性、携帯電話の保有状況、調査日に最も近い一日分の端末使用履歴に関するデータ収集を行った。

3. 研究方法

3.1 インタビュー調査標本の記述統計分析

調査回答者を端末使用者と非使用者に分類し、両者の属性について考察を行う。本調査の標本は所得レベルによる層化抽出法を用いて標本抽出を行ったため、分析も所得レベル別に行う。

3.2 行動パターンに影響する要因の特定

次に、既往文献から、個人の行動パターンと相関があるとされる世帯属性を抽出する。既出の通り、性別等の個人属性推計に関する試みは進みつつあることから、ここでは男女別に世帯属性に関連すると考えられる特徴量について考察する。

3.3 端末使用傾向分析による特徴量抽出

3.2の既往研究から導いた要因が、実際に端末使用傾向として抽出可能な特徴量であるかについて考察する。本稿では、インタビュー調査とCDR両データから端末使用傾向を分析し、キーボードとなる特徴量の抽出を試みる。

4. 分析結果

4.1 端末使用者と非使用者の属性

表1に所得レベル別の(A)世帯人数、(B)端末利用者数（通信会社不問）、(C)対象端末使用者数をまとめた。なお括弧内は世帯人数に占める割合を示している。(B)通信会社不問の場合、所得レベルが高いほど端末使用者数は大きいが、調査対象端

末に限ると所得レベルによる差は殆ど無く、何れも端末使用者1人に対して、世帯あたり約2～3人の同端末不使用者が存在することになり、端末使用者一名に対し、その自宅付近にCDRに含まれない人口が平均2.6人存在することが推測される。なお、本調査は対象端末使用者を一人以上含む世帯のみをインタビュー調査しているため、(B)の値は実際よりも過小評価、(C)の値は過大評価される傾向にあると考える必要がある。

表1 所得レベル別世帯当たり平均人数分布(括弧内は世帯人数(A)に対する割合)

	(A)世帯人数	(B)端末使用者数 (会社不問)	(C)対象端末 使用者数
高所得	4.00	2.64 (66%)	1.16 (29%)
低所得	4.09	2.33 (57%)	1.15 (28%)
中所得	4.14	1.90 (46%)	1.14 (28%)
スラム	4.07	1.73 (43%)	1.07 (26%)

図1は(a)全標本と(b)対象端末使用者における男女分布を表している。(a)における男性の割合は平均51%で、ダッカの男性の割合54% (Bangladesh Bureau of Statistics, 2012) と変わらない。一方、(b)は男性の割合が比較的大きいことが分かる。

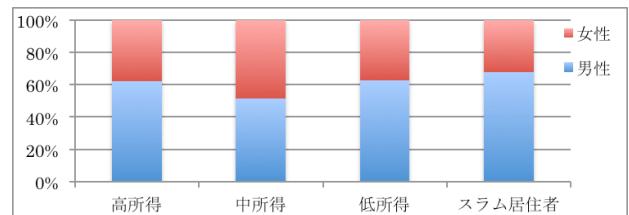
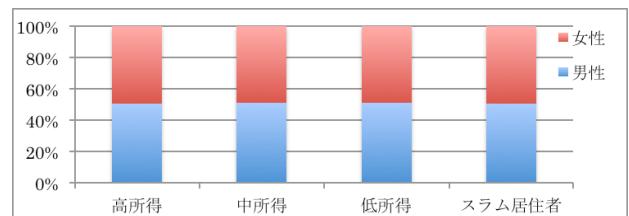


図1(a) インタビュー調査全標本(b)同標本内の対象端末使用者における所得層別男女分布 (a)上(b)下

図2は各所得レベルの対象端末使用者を100人ずつ抽出し、所得レベルごとに図2(a)男性と図2(b)女性に分け、その世帯内続柄の人数内訳を表したものである。どの所得レベルでも男性は世帯主、女性は配偶者が大半を占めている。これは、当該端末使用者の多くが既婚であり、家庭内意思決定に関わる立場にある者であることを示している。

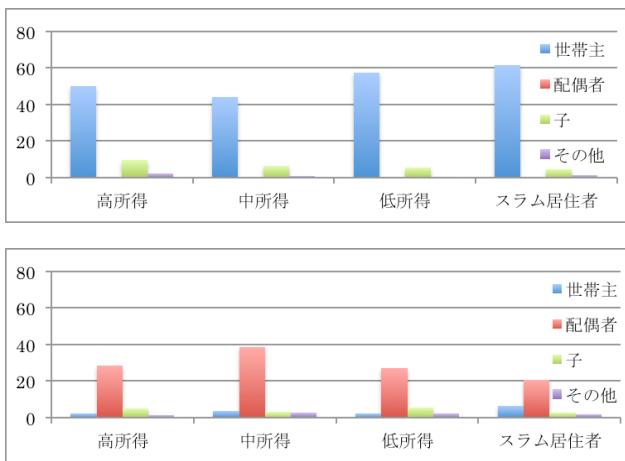


図2 インタビュー調査の対象端末使用者100人に占める(a)男性と(b)女性の世帯内続柄の所得層別分布人数 (a)上(b)下

図3は図2で抽出した各100人に対して世帯内に存在する当該端末非使用者の内訳を表している。所得レベル別の合計人数と男女内訳(括弧内は男:女の人数比)は高所得244人(112:132), 中所得256人(130:126), 低所得256(115:141), スラム居住者168人(124:154)となっている。所得レベルによる大きな違いはなく、男女とも子供の割合が圧倒的に多い。これは、CDRデータ分析による、子供の分布の把握の難しさを示している。逆に、100人分のCDRデータを分析すれば、その自宅周辺エリアにはCDRデータには反映されない130~160人程度の子供の存在が想定できる。なお子供の約半数は0~9歳の年齢グループが占めている。

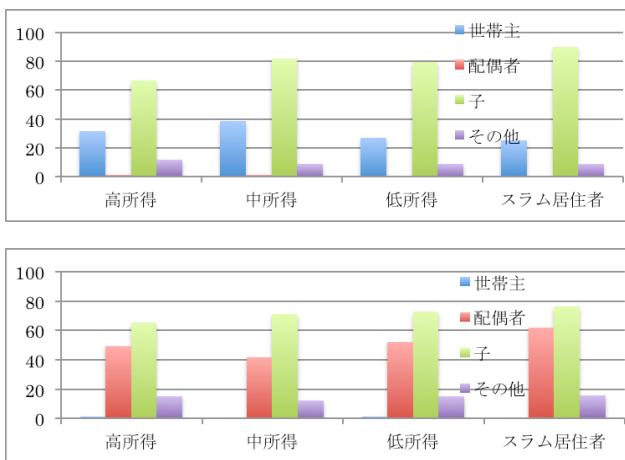


図3 インタビュー調査の対象端末使用者100人に対する当該端末不使用者の占める(a)男性(b)女性別世帯内続柄の所得層別分布人数 (a)上(b)下

4.2 行動パターンに影響を与える世帯要因

交通や都市計画分野では、都市OD調査(Origin-Destination調査)に見られるような質問票ベースの調査により、人間の移動や活動パターンに関する研究が長きに渡って行われてきた。その中で、女性の活動パターンは世帯内の子供の有無に影響を受けることが議論されている(Kitamura, 1988)。そこで本研究では、世帯における12歳以下の子供の有無と女性の端末使用パターンを比較し、女性の端末使用パターンから、女性が属する世帯における12歳以下の子供の有無を抽出することができるような特徴量を抽出する。

4.3 端末使用傾向分析

図4(a), (b)はインタビュー調査データを用いて、対象端末使用女性の端末使用傾向を比較したものである。女性を所属する世帯における12歳以下の子供の有(赤)無(青)によって分類し、さらに端末使用履歴が調査対象者の主要活動日(例えば、働いている者の場合仕事をする日)である場合は(a)、非主要活動日(ルーティン以外の活動を主に行う日)である場合は(b)に振り分けている。図の縦軸は、各端末使用場所からの発信が総発信数に占める割合を示す。端末使用場所は、自宅と自宅外主要活動場所の二種類を用いており、自宅外主要活動場所は、職場や学校、マーケット等、自宅以外で頻繁に訪れる場所を表している。図4(a), (b)を比較すると、12歳以下の子供がいる世帯の女性はそうでない世帯の女性に比べて、非

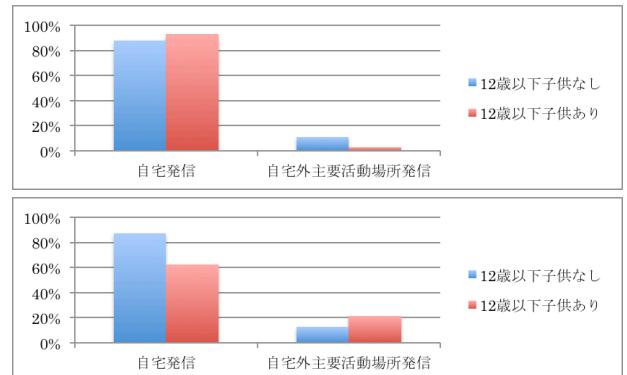


図4 インタビュー調査データによる対象端末使用女性の(a)主要活動日と(b)非主要活動日別、自宅発信・職場発信が総発信数に占める割合 (a)上(b)下

主要活動日に自宅外主要活動場所から発信する割合が高い傾向があることが分かる。そこで、同様の比較を CDR で行った結果を図 5 に示す。図 4 ほど傾向は顕著ではないが、非主要活動日の端末使用傾向は図 5 も同様の傾向を示していることが分かる。以上から、個人ベースで蓄積される CDR データから端末使用者の所属する世帯属性を推計するためには、女性の非主要活動日における、自宅外主要活動場所からの発信割合が重要な特徴量の一つであることを導くことができた。

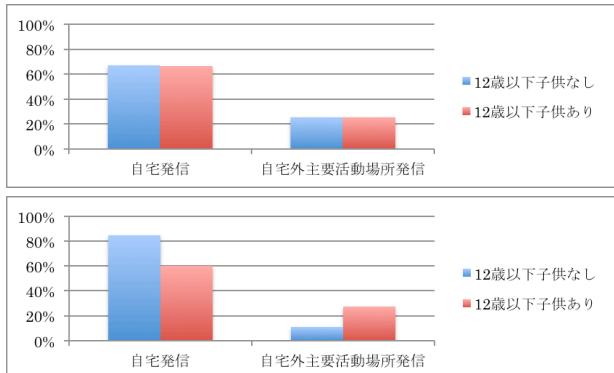


図 5 CDR による対象端末使用女性の(a)主要活動日と(b)非主要活動日別、自宅発信・職場発信が総発信数に占める割合 (a)上(b)下

5. おわりに

携帯電話データ (CDR) は大規模人口の流動動向を把握できることから交通・災害対策等に活用されることが期待されているデータである。一方で、CDR は端末使用者の動向のみを反映することから、カバーすることができる人口層には偏りがあることが指摘されている。本研究では、CDR から抽出することができる端末使用傾向から、端末使用者の所属する世帯の 12 歳以下の子供の有無を判定するために有用な特徴量を抽出した。本研究成果は、CDR を分析することにより CDR に含まれない幼い子供のいる世帯分布状況を把握する一助となることから、災害弱者の一部である子供の分布把握に貢献することが期待される。

本研究では、女性の端末使用傾向から世帯属性推計のための特徴量を抽出したが、男性の端末使用傾向と世帯属性についても今後検討を行い、抽

出した特徴量を実際に CDR に適用して世帯属性推計を行うことを目指している。

謝辞

本論文執筆にあたりご協力頂いた携帯電話事業者、CDR 提供者の皆様、調査参加者の方々に、深く感謝致します。本研究の一部は、文部科学省「グリーン・ネットワーク・オブ・エクセレンス (GREENE) 環境情報分野」の支援のもとに実施されました。

参考文献

- Arai, A., and R. Shibasaki. 2014. Estimation of Human mobility patterns and attributes analyzing anonymized mobile phone CDR. In *Proceedings of the 2nd AGILE PhD School 2013*. Frauenwörth, Lake Chiemsee, Germany, September 30 - October 2, 2013. ISSN1613-0073, online <http://ceur-ws.org/Vol-1136/paper2.pdf>.
- Bangladesh Bureau of Statistics. 2012. Community report: Dhaka Zila June 2012.
- Brdar, S., D. Culibrk, and V. Crnojevic. 2012. "Demographic attributes prediction on the real-world mobile data. In *Proceeding of Mobile Data Challenge by Nokia Workshop in Conjunction with International Conference on Pervasive Computing*, Newcastle, UK.
- González, M. C., Hidalgo, C. A., and Barabási, A. L. 2008. Understanding individual human mobility patterns. *Nature*, 453, 7196, 779-782.
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M. J., Rowland, J., and Varshavsky, A. 2011. Identifying important places in people's lives from cellular network data. In *Pervasive Computing*, 133-151. Springer Berlin Heidelberg.
- Kitamura, R. 1988. An evaluation of activity-based travel analysis. *Transportation*, 15, 9-34.
- Nadeem, S. 2012. Demographic prediction of mobile user from phone usage. *Age*, 1, 16-21.