

東京都市圏における長期の GPS データを用いた移動経路の推定に関する研究

大野夏海・関本義秀・中村敏和・Horanont Teerayut・柴崎亮介

Estimation of Routes Using Long-Term GPS Data in Tokyo

Natsumi ONO, Yoshihide SEKIMOTO, Toshikazu NAKAMURA, Teerayut HORANONT, and Ryosuke SHIBASAKI

Abstract: Information about the daily travel patterns of people is becoming increasingly important, and in recent years, may be obtained through Global Positioning System (GPS) logs instead of traditional travel diary surveys. Such logs automatically collect time and location data but little else; thus, this study uses GPS data from Tokyo, Japan, to infer specific trip routes. While previous similar studies have conducted experiments with GPS loggers, our data has been collected from ordinary mobile phone users, and therefore is inconsistent. In this paper, we identify significant locations and trips by using continuous data taken for individuals over the course of a year.

Keywords: GPS (GPS), 携帯電話 (mobile phone), 人の流れ (people flow), 長期データ (long-term data)

1. はじめに

近年、防災やマーケティングなどにおいて、人々の流動を把握する重要性は高まっている。特に昨年の東日本大震災では、東京都市圏でも帰宅困難者で人々があふれる場所も続出し、日々の流動特性を把握する必要性が顕著となっている。

また一方で、携帯電話の普及により、大規模（数千人～数十万人）かつ長期（数か月～数年）の GPS データが蓄積されてきている。データの時間や空間の解像度はそれほど高くないものの、例えば株式会社ゼンリンデータコム「混雑度マップ」(<http://lab.its-mo.com/densitymap/>) など、リアルタ

イムにおける人口分布を推定するためにも用いられている。

しかし、GPS データそのままでは点列に過ぎず（図 1）、一人一人の具体的な移動経路や利用した交通手段等、意味を持つレベルまで把握することが重要である。

最終的には、こうした GPS データから個人の詳細な時空間経路や交通手段等の推定を行うことを目的とするものの、本研究では、その前段階としてトリップパターンの把握を行った。



図-1 携帯電話による 1 日分の GPS データ例

大野夏海 〒153-8505 東京都目黒区駒場 4-6-1

東京大学生産技術研究所 Cw-503 号室

Phone: 03-5452-6417

E-mail: nono@csis.u-tokyo.ac.jp

2. 手法

2.1 全体の流れ

本研究では、大規模・長期の GPS データとして、数十万人レベル、1 年程度のボリュームを想定し、また GPS の取得間隔として数分～数十分、空間的には数 m～数百 m の誤差が存在するレベルを想定している。

大規模・長期の GPS データから個々のトリップを抽出するため、まず各ログが移動中か否か判別する。既存研究では、各ログにおける速度による判定 (de Jong et al. (2003)) や、ログのクラスターによる滞在場所の抽出 (Ashbrook et al. (2003)) が行われた。しかし本研究で用いるデータは取得の時間間隔が大きい上に不規則であるため、正確な速度は求められず、また、ログ数は滞在時間の指標とならない。

従って、連続するログの位置関係と経過時間を両方考慮して滞在場所 (「ステイポイント」と呼ぶ) を求める。長期データを重ねる際、同一と考えられるステイポイントを特定する必要があるため、位置の近いものをクラスターする (「ステイエリア」と呼ぶ)。

次に、同じステイエリアを発着地点にもつトリップを抽出し、移動した可能性が最も高い経路を推定する。

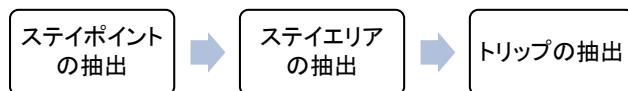


図-2 全体の流れ

2.2 ステイポイント及びステイエリアの抽出

まず各個人の長期データについて、150 メートル以内の範囲に 20 分以上滞在したログを、まとめてステイポイントとして抽出する。

次に、求めたステイポイントを、Ashbrook のアルゴリズムでクラスターする (Ashbrook et al. (2003))。

アルゴリズムの原理は K 平均法と同じだが、パラメーターがクラスターの数でなく半径となる。データの点を中心点として円を描き、その半径内に存在する点を求め、これらの点の平均値を計算する。この平均値を新たな中心点として、この操作を繰り返す。中心点が動かなくなると、その円を新しいクラスターとし、含まれる点をデータセットから取り除く。データが全て取り除かれるまで、このプロセスを繰り返す。

本研究ではクラスターの半径を 500m に設定し、各ステイポイントの中心点を用いて、ステイエリアを作る。結果のイメージを図 3 に示す。

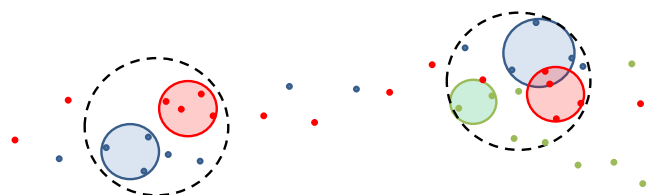


図-3 GPS ログ (点)・ステイポイント (実線)
・ステイエリア (点線) の関係

2.3 トリップの抽出

ステイエリア間のデータをトリップとして抽出する。OD 表を作成し、各トリップのサンプル数を確認する。今回は往路と復路を合わせて 10 サンプル以上抽出できたトリップを有効とした。

3. 実験

3.1 データの紹介

本研究では株式会社ゼンリンデータコム of 混雑統計データを用いる。2010 年 8 月～2011 年 7 月の期間で、移動時のみ、最短 5 分の間隔で取得された GPS ログとなる。最も多くのログが東京都内で観測されたデータは約 25 万人分あるが、そのうち 5000 以上のログが存在し、その 95%以上が一都三県 (東京都・埼玉県・千葉県・神奈川県) に位置する、9 つの ID (A～I) をランダムに選択した。

表-1 データ及び抽出結果について

個人 ID	A	B	C	D	E	F	G	H	I	平均値
全ログ数	8031	15998	12102	10410	6764	12133	25400	25703	12327	14318.7
ステイポイント数	731	890	516	782	646	969	679	1060	864	793
ステイエリア数	52	69	65	55	34	33	53	229	53	71.44
1 サンプル以上のトリップ数	2	53	9	13	9	15	22	43	59	25.0
10 サンプル以上のトリップ数	0	11	2	4	4	4	9	7	8	5.4

3.2 ステイポイント及びステイエリアの抽出

選択した ID について、ステイポイント及びステイエリアを抽出した。図 4 は ID「I」の全ログ及びステイポイントを可視化したものである。また図 5 は図 4 の白枠内の拡大図であり、ステイポイントのクラスターとステイエリアの関係を示す。9 つの ID の平均値で考えると、ステイポイントが年間 793 か所（一日当たり約 2.17 か所）抽出されたのに対し、ステイエリアを約 71 か所に絞り込むことができた。

3.3 トリップの抽出

作成したステイエリアを用いて OD 表を作成し、各トリップのサンプル数を確認した。表 1 では、1 サンプル以上抽出できたトリップ数及び 10 サンプル以上抽出できたトリップ数を示す。また、表 2 は ID「I」の OD 表を、10 ステイポイント以上のステイエリアのみを用いて作成したものである。赤色の数字は、往路と復路を合わせて 10 サンプル以上抽出できたトリップを示す。

表-2 ID「I」の OD 表

	1	3	4	8	16	18	19	20	22	29	30	31	35	38	44	47	48
1	147	9	14	5	8	9	14	3	12	10	8	5	5	4	3	2	5
3	7	9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	9	2	4	0	0	1	0	0	0	0	0	0	1	0	0	1	0
8	10	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
16	8	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0
18	9	0	2	1	0	12	0	0	0	0	0	0	0	0	0	0	0
19	12	0	0	0	0	0	20	0	0	0	1	0	0	0	0	0	0
20	2	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
22	10	0	1	0	1	0	1	0	14	0	0	0	0	0	0	0	0
29	7	0	0	1	1	0	0	0	0	13	11	0	0	0	0	0	0
30	5	0	1	2	0	0	0	0	0	8	14	0	0	0	0	0	0
31	4	0	0	1	0	0	0	0	0	0	0	9	0	0	0	0	0
35	2	0	1	0	1	0	0	0	0	0	0	0	10	0	0	0	0
38	3	0	1	1	0	0	0	0	0	0	0	0	0	5	0	0	0
44	2	0	1	0	0	0	0	0	0	0	0	0	0	0	8	0	0
47	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
48	7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	7

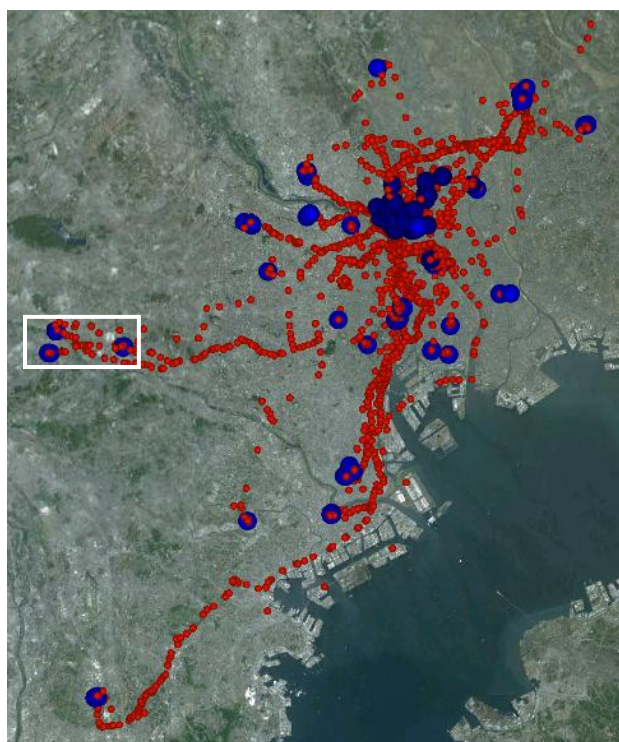


図-4 ID「I」の全ログ (赤) 及びステイポイント (青)

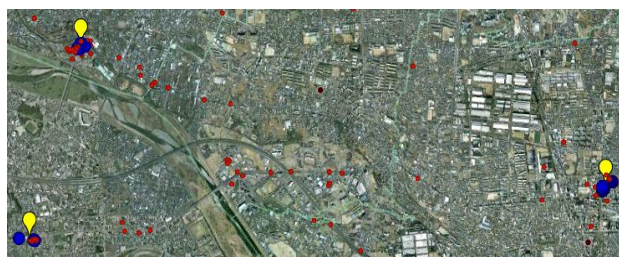


図-5 ID「I」のステイポイント (青) 及びステイエリア (黄) (図-4 の白枠内を拡大したもの)

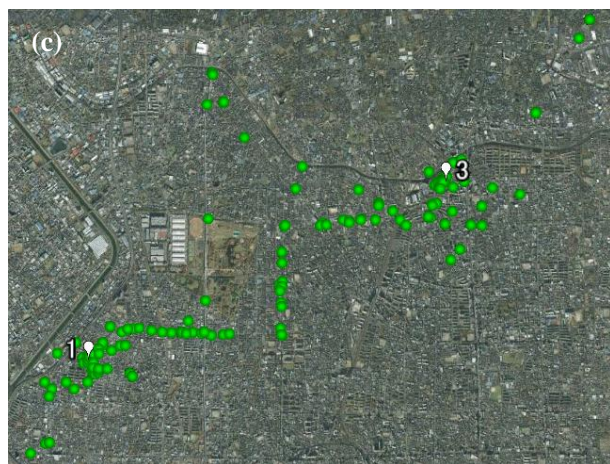
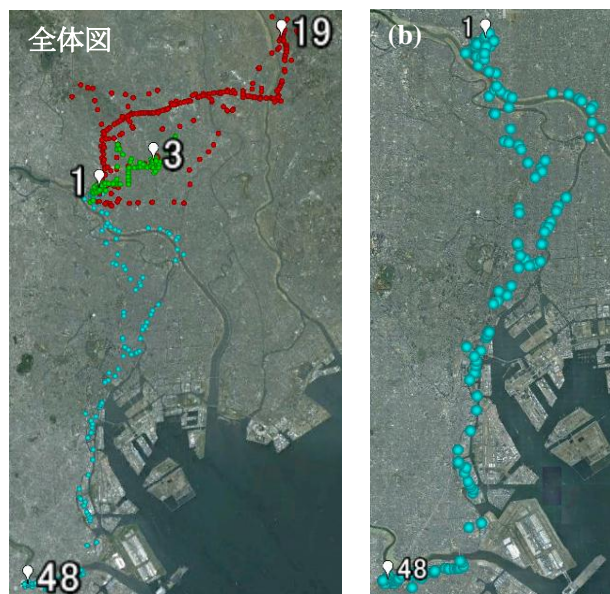


図-6 ID「I」のトリップの例

ID「I」から抽出したトリップの例を図6に示す。特に図6(a)(1と19を結ぶトリップ)の範囲は図1と同じであるため、一日分のログと比較し、長期データを用いると移動経路が分かりやすいことを確認できる。また、図6の(b)(1と48を結ぶもの)と(c)(1と3を結ぶもの)はトリップの距離が大きく異なるものの、いずれも抽出結果より経路を推定できると言える。最後に、同一のトリップにおいて複数の経路が抽出されることも目立った。

4. おわりに

本研究では、大規模・長期のGPSデータを用いてトリップパターンの把握を行った。今後の研究では、抽出したトリップのデータをクリーニングした上で、実際の交通ネットワークとマッチングを行う。具体的には、東京都市圏の鉄道ネットワークや道路ネットワークと重ね合わせ、速度等を参考にしながら、利用した交通手段を判定する。また最終的には、詳細の移動経路や交通手段等が確認済みのデータを用いて、本研究の手法を検証する。

謝辞

株式会社ゼンリンデータコムの皆様には分析用データを提供して頂いた。ここに深謝の意を表します。

参考文献

- Ashbrook, D. and Startner, T., 2003. *Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users*. *Personal and Ubiquitous Computing*, v.7 n.5, p.275-286.
- de Jong, R. and Mensonides, W., 2003. *Wearable GPS device as a data collection method for travel research*. Working Paper, ITS-WP-03-02, Institute of Transport Studies, University of Sydney