

人々の流動データや高精度画像利用による都市のマネジメント

関本義秀¹

概要: 本論文は 2008 年に立ち上げた「人の流れプロジェクト」の歩みや都市マネジメント技術を、「人の流れデータ提供」や携帯電話の GPS や CDR データを用いた研究等を中心にとまとめたものである。

Urban management using people flow and high-accuracy image

YOSHIHIDE SEKIMOTO¹

1. はじめに

最近のビッグデータブームの中で今や当たり前になりつつあるが、都市レベルの人の流動について始めたいと思ったのは 2000 年頃の博士課程の頃。しかし、博士論文が終る頃には建物の中のレベルの流動がちよっとできたくらいで[1]、自分の技術力のなさに情けなく思ったものである。

その後 2002 年から 5 年ほど在籍した国土交通省国土技術政策総合研究所で、もう少し何とかできないかと思っていた頃に東京都市圏のパーソントリップデータと出会った。マスターデータを借りることができ、雑なものであったがアニメーションができた時にはちょっとした興奮を覚えたものだった(本務は道路の基盤データ整備というものだったが、第二の本務とすべく科研費基盤 C「動くデジタル東京：人々の時空間行動把握のための情報統合インフラの構築に関する研究」が運よく獲得できたこともあった)。

技術的・研究的には GPS を配布することは頑張れば数十台～数百台規模であれば可能だがそれ以上の規模は難しいし、長期間のモニタリングをキープするととなると謝礼も必要となる。交通の分野ではこのような研究は見られ始めていたが[2]、交通計画・土木計画論や行動モデリング等、役所の長期政策のための数値的根拠作り、という側面が強く、可視化等による現状のモニタリングや医療・防災・マーケティング等、様々な分野の人々のためのデータ基盤という色彩はなく、そこに対して技術的・費用的・制度的に解決を目指していくことに大きな意味があると思った。

ということで財務省にもきちんと予算要求を行い、外部の有識者も集め、研究所の正式プロジェクトで 2006 年に「都市空間における動線解析プラットフォーム研究会」を立ち上げ、例えば 1 千万円の予算で 10 万人規模が何とか流動計測できないかなどと、携帯電話会社に見積もり交渉を一生懸命したことなどを思い出す。

その後、東京大学空間情報科学研究センター(CSIS)で

働くことができ、大学の自由な立場から再度、独立にプロジェクトを立ち上げることを考えた。CSIS は文部科学省が定める空間情報科学の拠点でもあり、全国の研究者に対するデータの無償提供による共同研究事業が盛んである。しかし GIS そのものは汎用化しつつあるツールなので、もっと時間軸も加わり時空間化していく方向で先端に行くのが将来像だろうと考え、この活動をきちんとプロジェクト化して、何とかその「都市空間における動線のデータ」(後に「人の流れデータ」と命名)を貸与できないか考えた。ただし単なる一若手の考えに過ぎないので、大仰な国のプロジェクトに仕立てて予算獲得というよりも継続的にサービスが続けられるように、CSIS 内部の若干の支援を頂きながら研究室の学生などと進めることにした(その後、運よく科研費若手 A「社会インフラとしての「人の流れ」データの効率的な再現技術の構築」を獲得でき大きく進めることができた)。当時、「流れ」という言葉にピンとこず、「人の動きプロジェクト」ではだめなのか、などの意見も学生からはあったが、一人のミクロな動きというよりは、都市の流動感を出すためだ、ということで「人の流れプロジェクト」にした記憶がある。

次章以降はそこから始まった人の流れプロジェクトの話であるが、様々な応用分野と情報処理分野の大きな接点の一つではないかと思われる。



図 1. 「人の流れプロジェクト」ウェブサイト (<http://pflow.csis.u-tokyo.ac.jp>)

¹ 東京大学生産技術研究所 人間・社会系部門 准教授
(工学系研究科社会基盤学専攻&空間情報科学研究センター 兼務)
Institute of Industrial Science, University of Tokyo

2. パーソントリップからの人の流れデータ

パーソントリップデータ（以下、PTデータ）は情報処理の分野ではあまりなじみがないかもしれないが、国や地方自治体が都市圏単位で行う交通やまちづくりの長期計画策定のために行う交通行動調査のデータで、1970年代くらいから5年や10年間隔で行われてきたものである。基本的にはA41枚程度のアンケート調査で、ある秋の平日の典型的な1日の行動を目的地群とそれぞれをつなぐトリップとさらに交通手段単位で分解したサブトリップの出発地・目的地（Origin-Destination: OD）を時刻と施設名・地名で記入するものである。アンケートはランダムサンプリングで都市圏住民全体の約1割程度の人に送付し、回収後、データの精査を経て3%前後のサンプルデータとなることが多い。

従って、これらトリップデータを人々の流動のデータ基盤、すなわち「人の流れデータ」として使いやすくするためには、時間的な間隔を一定にするとともに、現実の道路・鉄道のネットワークに沿った形でリアリティも持たせる必要がある。つまり、一定の時間間隔ごとにネットワークに沿った形でサブトリップのODを時空間内挿する必要があるが、何分くらいの間隔が適当だろうか？車や電車はかなり早い移動なので5分だとかかなり間が空いてしまうし、任意の時刻の検索もやりにくい、という理由で1分とした。

その処理プロセスを表したのが図1であるが、(a)でサブトリップの施設名・地名をジオコーディングして緯度経度に変換し、(b)で道路や鉄道のネットワークデータの位相構造を用いて最短経路を推定する。最短経路という仮定は必ずしも現実と一致するとは限らないが、サブトリップはそれなりに短い単位であるため、最短経路に近いはずという想定で行っている。最後に(c)で道路や鉄道のネットワークの幾何構造に沿った形で1分ごとの位置を内挿している。

また、属性についてはマスターデータが性別、年齢、職業分類等、多くのものを含んでいるが、マスターデータを再現不可能にすることが調査当局からの貸与条件であるため、人の流れデータの中では、多くの属性は削除し、年齢等も10歳刻み等に行っている（詳細はSekimotoら（2011）[3]を参照されたい）。

その後、トリップデータ中の公共施設名はピンポイントでわかるものの、地名はゾーンとして表現されるため、かなり広域であり、そのままジオコーディングしてしまうと一つのゾーンが一つのポイントに集約されてしまう点を解消するために、住宅地図のデータを参考にし、建物位置や種別、世帯人数等をベースに、ゾーンをジオコーディングする時の座標を確率的に決定している（詳細は秋山ら（2012）[4]）。

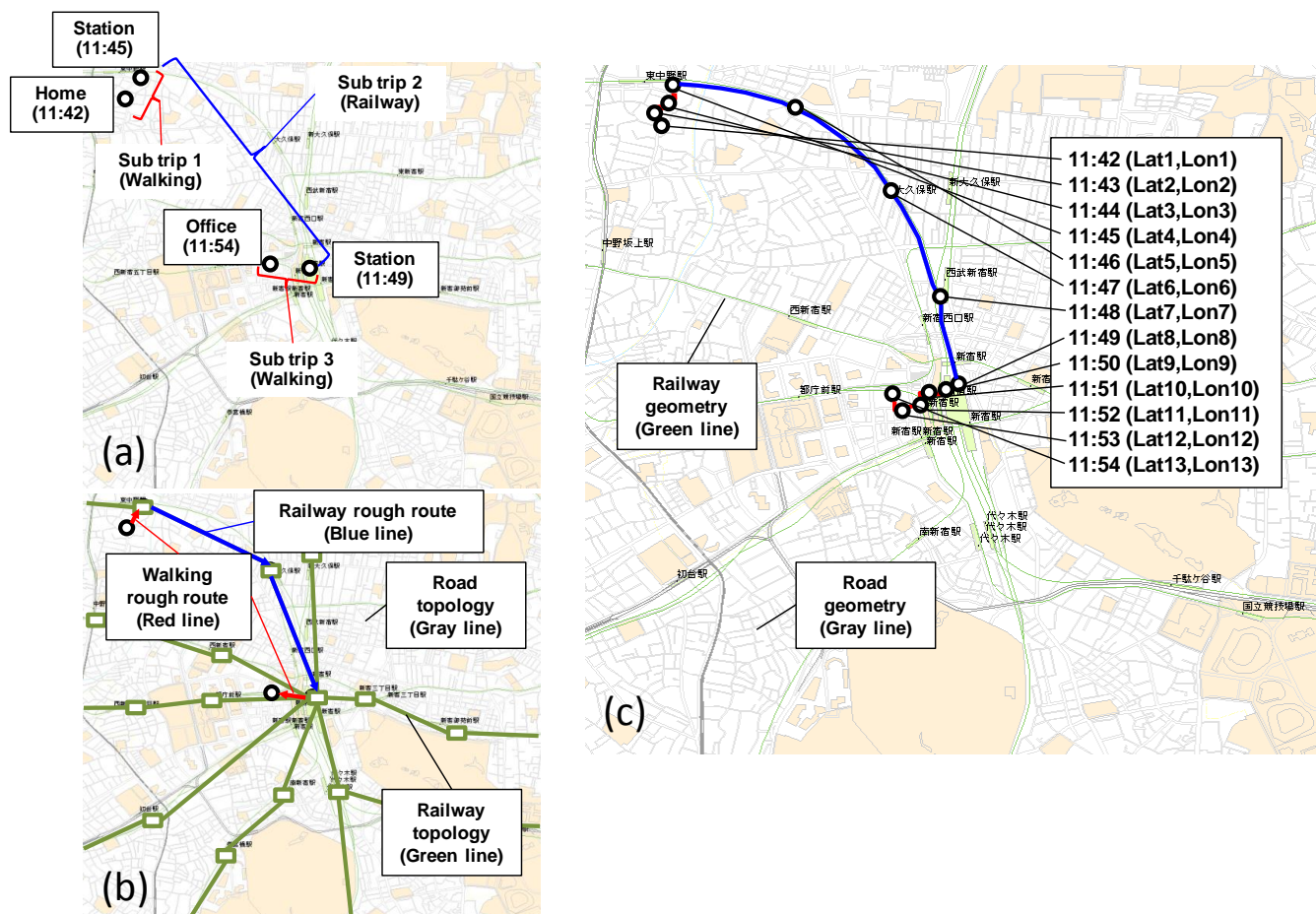


図1. パーソントリップデータの起終点情報からネットワークに沿った時空間内挿（Sekimotoら（2011）[3]から引用）

こうした形で 2015 年 5 月時点では 25 都市、約 440 万人分のデータを提供し、共同研究も 2009 年以降、200 件近くにも及んでいる（これら共同研究の概要はサイトに公開されている）。また、これらには JICA が行っている海外の PT データ 4 都市分も含んでおり、今後はこうした海外データのニーズもかなり増えてくるのではないと思われる。なお、これらの処理は極力、グローバルで適用可能なものを目指しているため、例えばネットワークデータは海外では OpenStreetMap データを使っている。ただし、住宅地図は日本だけなので、海外の PT データには適用できず、1km メッシュベースの土地利用データ等をベースに、粗くはあるが、ゾーン内人口を空間配分している（これらの詳細は Sekimoto ら (2013) [5]を参照のこと）。

一方で、これらを単なる PT データをもととした人の流れデータ提供だけではなく、ユーザーが自ら持ち込むタイプのデータの時空間内挿等、加工ツールとしてのプラットフォームという色彩も重要で、「動線解析プラットフォーム」という名前で人の流れプロジェクトを支えている。これらも当初から WebAPI ベースで多くの機能を提供しており、機能、処理速度の改良を重ね、WebAPI 仕様書が Ver. 2.70 となっている（WebAPI 仕様書は <http://pflow.csis.u-tokyo.ac.jp/wp-content/uploads/webapi.pdf> からダウンロード可能。また、これらの基本機能は関本ら (2011) [6]を参照されたい）。

3. 携帯電話データとの出会い

しかし、PT データはあくまで、アンケートベースであり、リアルタイムではないし、日々の変動が見える訳ではない。時流はやはり携帯電話のデータである。[3]で挙げた IEEE Pervasive Computing でも 2011 年当時、ビッグデータという言葉がはやる前であるが、“Optimistic Large-Scale Sensing”という名前で特集が組まれた。我々のグループのみ上記の PT データを使った流動再現で、それ以外の論文は全て欧米の携帯電話や SNS を使った都市分析だった記憶がある。この分野は Albert Barabasi と Marta Gonzalez のグループ（例えば[7]）や Carlo Ratti らの MIT Sensible City Laboratory（例えば[8]）などから始まった流れということができよう。とくに 2008 年の Nature に掲載された Marta Gonzalez の“Understanding individual human mobility patterns”では、Human mobility が交通という切り口だけではなく人の行動生態という切り口で扱われることを示したと言える。

そこで日本を振り返ると、研究でデータを利用できたり、製品が出てくるのはだいぶ後であるが、すでに 2009 年から NTT ドコモ社を中心に「モバイル空間統計による社会・産業の発展に関する研究会」が行われ、現在の「モバイル空間統計」のコアとなる議論が行われていたのは特筆すべきであろう。この当時の携帯電話を活用した人々の流動解析

のタイプや、法制度との関係は関本ら (2011) [9]を参照されたいが、もう一つ当時のホットピックはゼンリンデータコム社の「混雑度マップ」が公開されたことであり、これは無償であり、1 時間前のリアルタイムな分布状態が見られるという意味で、衝撃的であった。ほどなく担当の方が、人の流れプロジェクトと是非連携をしたいということで、研究ベースで利用することが可能になり、興奮した記憶がある。当時は東日本大震災の 2,3 か月後くらいの時期であり、何かわかりやすくその日を俯瞰して表すという意味で、柴崎教授のアイデアと上山特任研究員の卓越した視覚化の技術（なお、これは現在 Mobmap という Chrome 上で動く汎用ソフトウェアとして公開されている）により、2012 年に「東日本大震災時の首都圏における人々の流動に関するアニメーション」（例えば[10]）ができ、TEDxTokyo での柴崎教授の講演で使われたり、YouTube では 2013 年 1 月の 3 日間に約百万アクセスがあったりした。

しかし、一方で 2013 年夏に起こった Suica 炎上事件をきっかけに、個人情報保護法を遵守し、より安全に位置情報を扱うことの必要性を再認識することとなった。政府の IT 戦略本部で設置された「パーソナルデータ利活用検討会」では、こうしたデータを扱うものが匿名化することや個人情報保護委員会からの承認を受けることを必要とする個人情報保護法改正法案の動向をふまえ、各事業者も、安全確保のための規約作りやオプトアウト事項の整備等をより周到に行うこととなった。

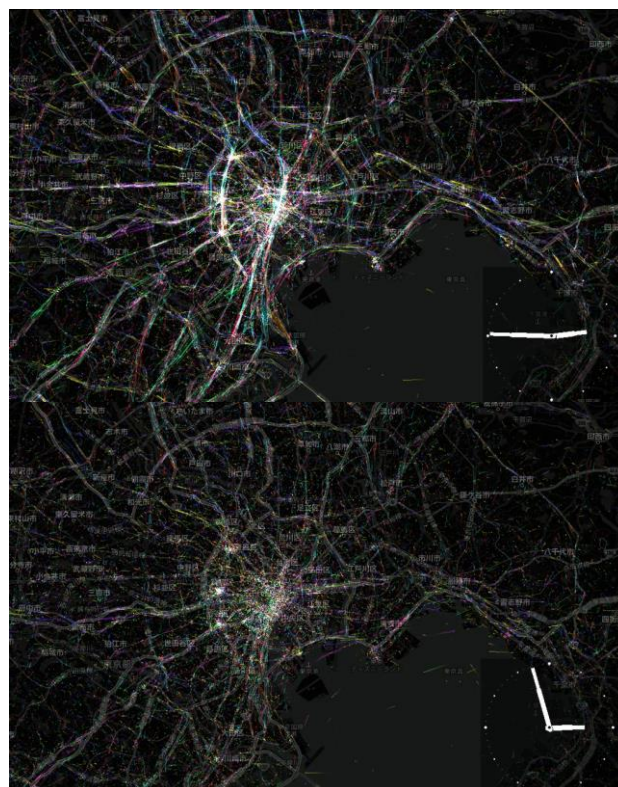


図 2. 東日本大震災前後の首都圏の人の流動状況の比較（関本 (2013) [10]から引用）

一方で、研究上の大量（150万人規模）・長期間（1年分等）のGPSデータは約92億行、600GByteにも及び、様々なデータ処理の効率化を再考するよい機会となった。こうしたデータ処理の原点は、Ashbrookら（2003）[11]の滞在点抽出やそれによるトリップセグメンテーション等が始まりと思われるが、規模が多いとインデキシングそのものが重要な問題である。全GPSのデータに対して、行政区画データ等と重ね合わせることで、メッシュコード、都道府県コード、市町村コードを付与した。これらのインデキシングは予め行うことができるので、Hadoop等の分散処理技術により効率的に行い、これらのインデキシングを用いて、当該地域のデータだけ予め処理対象として限定した上で、当該GPSデータを処理する事ができる。さらには、生成したトリップにはそれぞれ交通手段が推定できるとよい。ため、機械学習により、いくつか交通手段が既知のデータを教師データとして、トリップ全体の各交通手段を推定している（詳細はApichon（2013）[12]を参照されたい）。

しかし、国際的に見ると、大規模のGPSデータ収集はプライバシーの問題や携帯端末のバッテリーの問題もあり、それほど事例は多くなく、CDR（Call Detail Record）と呼ばれる、携帯端末の基地局利用状況を記録したデータを使った分析の方が多いように思われる。そうしたものは、位置表現は基地局ベースとなるため、位置精度は粗いものの、ある程度の傾向は見えてくる。こうした研究は、Kurokawaら（2013）[13]、Wangら（2012）[14]、関本ら（2015）[15]などがある。また、図3は、バングラディッシュとの共同研究におけるダッカ市内のある時刻の分布状況を可視化したものである[16]。

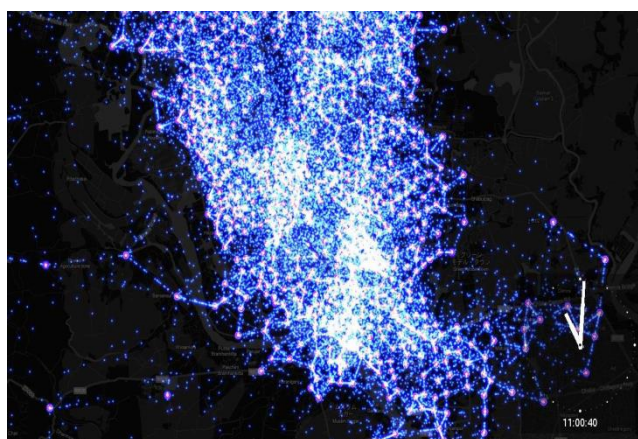


図3. 携帯電話基地局の位置情報を用いたダッカの人々の分布（関本（2014）[16]からの引用）

4. シミュレーションとの融合

ここまで来て、再度気づくことは、2章のようなPTベースのデータは公共色が強く、無償で入手ができる可能性があり、ODの実サンプルデータは用いているが一種のシミュレーションデータである。

一方で、3章で述べたものは携帯端末の契約に基づく民間データであり、かなり高価であり、また、個人個人のデータを大規模で入手できることは個人情報の観点からなかなかない。そこで、リアルタイム観測ベースの携帯端末をメッシュ単位で集計した時間帯別のメッシュ人口データと、2章のようなPTデータに基づく典型的な都市圏の流動データを組み合わせ、リアルタイムな推定流動を求めるアプローチも行っている。これらは、モデルとデータの融合という意味で、データ同化などと呼ぶこともあり、観測データが取得できない将来の予測をモデルから行えるメリットもある（例えば都市圏レベルの流動という意味では中村ら（2013）[17]がある）。

図4は実際に、東日本大震災時の首都圏全体の流動をPTデータによる平常時の分布状況を初期状態として、15時以降に1時間に1回メッシュ人口データ（この場合1kmメッシュで実験を行っている）として得られたGPSベースの混雑統計データを観測データとして用いて、データ同化を行い、推定を行った事例である。(a)の方は、2時間に1回しか観測データで同化ができなかった場合に間をシミュレーションで補った精度（上が相関値、下がRMSE/平均人口）、(b)の方は、4時間に1回しか同化ができなかった場合にシミュレーションで補った精度である。

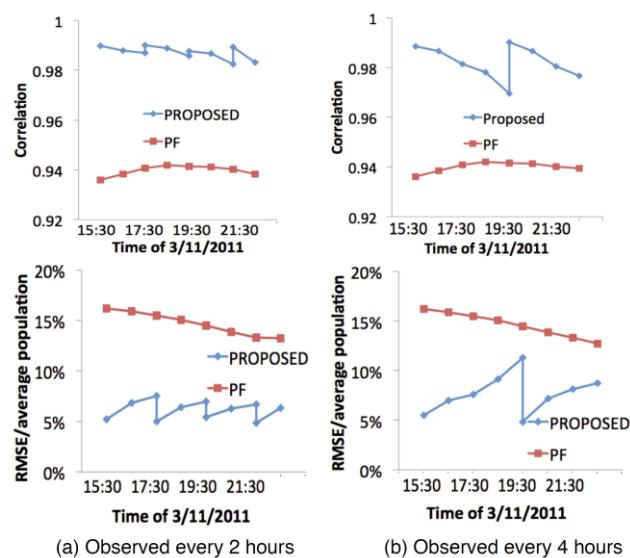


図4. 2時間ごと、4時間ごとにデータ同化を行った場合の再現精度の回復状況の比較（PFとあるのは、パーティクルフィルターを用いた場合。Proposedとあるのは、同化の際になるべく観測値に近くなる形で値自体を補正する手法を用いた場合）（須藤特任研究員提供）

5. 国際的な動向

最後に、最近の国際的な動向について述べたい。携帯電話を利用したHuman mobilityの分析は急速に増えつつある。しかし携帯電話ログの活用については個人情報保護の観点

からもバランスの取れた利用が望まれ、個別のビジネス等の利益のために用いることへの合意形成が簡単には形成しにくいこともある。そこで、まずは途上国支援のような公益のために使うことに絞り、アイデアを募ることとしたものが、D4D (Data for Development) であり、世界最大手の携帯通信キャリアの一つであるフランスのオレンジ社である。これは、2013年から始まり一回目はコートジボアールのデータ、2014年はセネガルのデータとなっていて多くの参加者があった。2013年のコートジボアールのデータについては、新井ら[18]によると、応募が80件あり、社会・経済開発が28%、交通・都市計画が35%、保健・感染症対策が19%、残りが残りはデータの可視化とマイニングに関する提案であった。

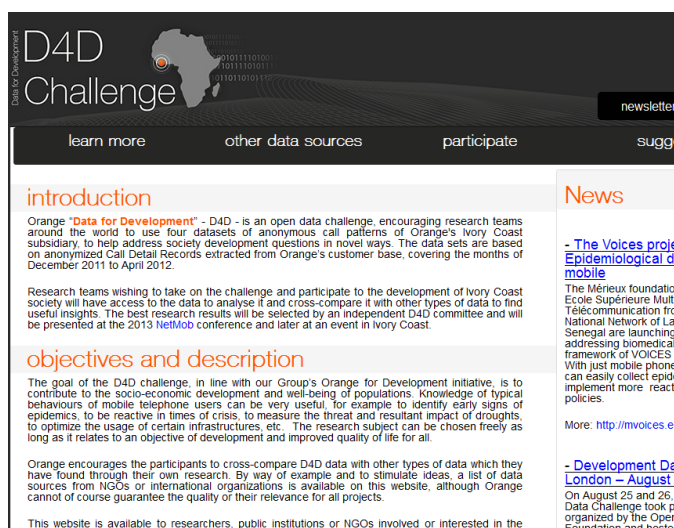


図5. 携帯電話会社オレンジ社による D4D チャレンジのウェブサイト (<http://www.d4d.orange.com>)

とくに最近では、世界的にもエボラ出血熱による感染症の拡大などもあり、感染症での適用等も増えている。感染症の研究はもちろん古くから存在しているが、感染源がどの辺で起こり、どの辺まで広がっているかの全体像をつかむのはなかなか簡単ではなかった。しかし近年の携帯電話の普及でこれらが少しずつ可能になりつつあり、ここでは、Caroline Buckee らハーバード公衆衛生大学院 (Harvard School of Public Health) における携帯電話のログ情報を用いた感染症伝播分析の研究[19]を紹介したい。

彼女らのグループはマラリアのホットスポットと見られていたケニアのビクトリア湖で詳細な感染状況を分析するために、その地域で働く人々、とくに茶の農園で働く多くの移民がどこから来て、どこに行くのかを携帯電話のログ情報 (CDR) を用いた分析した。図6はその結果であり、ビクトリア湖を含む感染源 (Source) とそれに対する感染先 (Sink) がどのように分布しているかを地図で示したものである。それと同時に彼女らのグループはケニアの CDR

データ 1500 万人分を用いて、一般のアンケート調査との違いなども調査を行い、CDRの方が実態に近い面も多いことを2014年にScienceで発表している[19]。

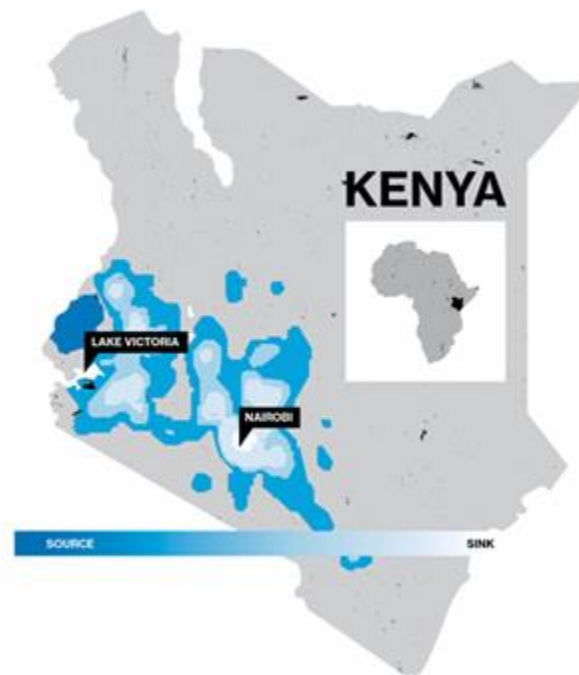


図6. 携帯電話利用データを用いたケニアにおける感染症伝播の分析 ([19]からの引用)

6. おわりに

以上のように、著者らの研究紹介が多くなってしまったが、人の流れプロジェクトの歩みを中心に人々の流動解析技術に関して述べた。我々自身も人の流動と言っても、都市レベルの流動だけでなく、まちやターミナル駅周辺レベルの流動もあるし、商業施設やキャンパスレベルのものもある。そうした様々なスケールでの流動推定的手法、適切なモデリングが重要である。また、公共データ、民間データが相乱れる中で、共有できるデータ、パラメータを考えることも重要である。また、学生向けへのエントリーという意味では昨年度から人の流れデータの PostGIS 上での簡単な処理や Mobmap を使った可視化等のチュートリアルなども進めている。一方で、最近、欧米の研究者を中心に、Flow minder という法人もできつつあり、日本は携帯データ利用については後発ではあるが、プレゼンスを示していくことも今後重要であると思われる。

謝辞

なお、これらについては、人の流れプロジェクト事務局のメンバーである、金杉特任研究員、瀬戸特任助教、檜山学術研究支援職員、須藤特任研究員、福島特任研究員と過去在籍した薄井特任准教授 (名古屋大学)、南特任助教 (総合地球環境学研究所) の協力なくしてはありえず、ここに

感謝したい。また、柴崎教授からは常に大所高所の視点から適切なアドバイスを頂いた。最後に、プロジェクトとして継続的に支援を頂いている東京大学空間情報科学研究センターの皆様にも感謝を申し上げる次第である。

[19] MIT Technology Review, 10 Breakthrough Technologies 2013, <http://www.technologyreview.com/featuredstory/513721/big-data-from-cheap-phones/> (参照 2015-5-15)。

参考文献

- [1] 関本義秀, 柴崎亮介: 多様な観測データや知識を用いた地物の時空間変化の再構成手法, GIS-理論と応用, Vol.11(2), pp.123-132, 2003.
- [2] 例えば, 朝倉康夫, 羽藤英二, 大藤武彦, 田名部淳: PHSによる位置情報を用いた交通行動調査手法, 土木学会論文集, No. 653, pp. 95-104, 2000.
- [3] Yoshihide Sekimoto, Ryosuke Shibasaki, Hiroshi Kanasugi and Tomotaka Usui, Yasunobu Shimazaki: PFLOW: Reconstruction of people flow recycling large-scale social survey data, IEEE Pervasive Computing, Vol.10, No.4, pp.27-35, Oct.-Dec. 2011.
- [4] 秋山祐樹, 金杉洋, 関本義秀, 柴崎亮介: 住宅地図データを用いた人々の時刻別空間位置の精細化, 第32回交通工学研究発表会論文集, pp.317-323, 2012.
- [5] Yoshihide Sekimoto, Atsuto Watanabe, Toshikazu Nakamura, Hiroshi Kanasugi, Tomotaka Usui: Combination of spatio-temporal correction methods using traffic survey data for reconstruction of people flow, Pervasive and Mobile Computing Journal, Elsevier, Vol.9, pp. 629-642, 2013.
- [6] 関本義秀, 薄井智貴, 金杉洋, 増田祐介: 都市空間における効率的な動線解析の共通基盤のあり方について, 土木学会論文集 F3 (土木情報学), Vol.67, No.2 (特集号), pp.170-180, 2011.
- [7] Gonzalez, M., Hidalgo, C., & Barabasi, A.: Understanding individual human mobility patterns, Nature, 453, 779-782, 2008.
- [8] Ratti, C., Pulselli, R. M., Williams, S., & Frenchman, D.: Mobile landscapes: using location data from cell phones for urban analysis, Environment and Planning B: Planning and Design, 33(5), 727-748, 2006.
- [9] 関本義秀, Horanont, T., 柴崎亮介: 解説: 携帯電話を活用した人々の流動解析技術の潮流, 情報処理, Vol.52, No.12, pp.1522-1530, 2011.12.
- [10] 関本義秀: 人の流動と時空間データセット最前線, オペレーションズ・リサーチ誌, Vol.58, No.1, pp.24-29, 2013.1
- [11] Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users, Personal and Ubiquitous Computing 7 (5), 275-286, 2003.
- [12] Apichon, W.: Study on human activity analysis with large scale gps data of mobile phone using cloud computing platform, In partial fulfillment of the requirements for the degree of Doctor of Engineering, Department of Civil Engineering The University of Tokyo, 2013.
- [13] Kurokawa, M., Watanabe, T., Muramatsu, S., Kanasugi, H., Sekimoto, Y., and Shibasaki, R.: "Extracting People's Stays from Cellular Network Data", NetMob2013, 2013.05
- [14] Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., Gonzalez, M. C.: Understanding road usage patterns in urban areas, Nature Scientific Reports, Vol.2, No.1001, 2012.
- [15] 関本義秀, 樫山武浩, 長谷川瑤子, 金杉洋: スパースな携帯電話通話履歴を用いたリンク交通量の推定〜ダッカの事例, 交通工学論文集, Vol.1, No.4, pp.A_1-A_8, 2015.4.
- [16] 関本義秀: まちづくりにおける新しいデータの活用事例, 雑誌「都市と交通」, Vol.98, pp. 17-19, 2014.
- [17] 中村敏和, 関本義秀, 薄井智貴, 柴崎亮介: パーティクルフィルターを用いた都市圏レベルの人の流れの推定手法の構築, 土木学会論文集 D3 (土木計画学), Vol.69, No.3, pp.227-236, 2013.
- [18] 新井亜弓, 関本義秀: 解説: 大規模人口流動データの利活用について, 写真測量とリモートセンシング, Vol. 52, No. 6, pp. 327-331, 2013.12.