GPS に基づくプライバシー保護を考慮した非集計データの都市計画分野における活用可能性

森山　拓洋[1]・江端　杏奈[2]・鈴木　宙顕[3]・樋田　英能[4]・陣内　寛大[5]

[1]正会員　株式会社 GEOTRA（〒100-0004 東京都千代田区大手町 1-2-1）
E-mail: moriyama@geotra.jp

[2]正会員　株式会社 GEOTRA（〒100-0004 東京都千代田区大手町 1-2-1）
E-mail: ebata@geotra.jp

[3]正会員　株式会社 GEOTRA（〒100-0004 東京都千代田区大手町 1-2-1）
E-mail: suzuki@geotra.jp

[4]正会員　株式会社 GEOTRA（〒100-0004 東京都千代田区大手町 1-2-1）
E-mail: hida@geotra.jp

[5]正会員　株式会社 GEOTRA（〒100-0004 東京都千代田区大手町 1-2-1）
E-mail: jinnai@geotra.jp

都市計画において人流データは重要な基礎データの一つであり，パーソントリップ調査や交通量調査などが古くから活用されてきた．特に携帯電話 GPS 位置情報から得られる人流ビッグデータは詳細な移動履歴を保持しており，交通需要を評価・予測するにあたって有用である．一方でこれまで，プライバシー保護の観点から個人が特定されないように集計加工されたデータのみが提供されてきた．しかし，集計済みデータは扱いにくく粒度も低いという問題がある．株式会社 GEOTRA は機械学習を活用し，データプライバシーを保護しつつ非集計で一人ひとりの行動を表現した人流データを生成する技術を開発した．本稿では，非集計データの概要を記載し，データの作成方法と既存調査データとの比較検証結果を記載する．また，比較検証を踏まえて，データの活用可能性を提案する．

Key Words: *city planning, GPS Big Data, machine learning, disaggregated human mobility data*

1. はじめに

都市の現状把握とシミュレーションにおいて人流データは重要な基礎データの一つであり，パーソントリップ調査や交通量調査などが古くから活用されてきた．交通需要を評価，予測するため携帯電話 GPS 位置情報から得られる人流データは移動経路を詳細に示すことができる．

GPS データの活用に関する既往研究として都市交通分野[1)2)]，観光[3)]，防災[4)]がある．また，携帯電話の基地局データ，GPS データの信頼性検証について既往研究[5)6)7)]がある．既往研究では，プライバシー保護の観点から，従来は個人が特定されないように集計加工されたデータのみが利用されてきた．プライバシーの保護は重要だが，集計加工されたデータは扱いにくく，情報の粒度も低いという問題がある．したがって，今後人流ビッグデータをさらに高次元で活用するためには，プライバシーを保護しながらデータの粒度を下げない技術の開発が重要である．三井物産株式会社と KDDI 株式会社は株式会社 GEOTRA を立ち上げ，機械学習を活用してプライバシーを保護しつつ，非集計で一人ひとりの行動を表現した人流データ，GEOTRA Activity Data（以下 GAD とする．）を生成する技術を開発した．本稿においては，非集計データの概要を記載するとともに，データの作成方法と既存調査データと比較検証をおこなった結果を記載する．また，比較検証を踏まえて，データの活用可能性や留意点を整理する．

2. GEOTRA Activity Data の概要

(1) データの概要

KDDI が保有する携帯電話から得られる au GPS ビッグデータを利用する．au GPS ビッグデータは，KDDI が au ユーザーの許諾を元に収集した GPS 位置情報であり，数分に一度のサンプリングによる高精度・高粒度な人流データである．指定された出発地から到着地までの推定移動人口を抽出しており，最小 125m メッシュ間の移動実態を表している [8]．尚，携帯キャリアから取得可能なデータは，GPS データと基地局データに分かれるが，GPS データは，基地局データに比べてメッシュサイズが細かく，データ取得頻度が高いため，詳細な移動経路が取得可能である．基地局データの特徴については図 -1，GPS データの特徴については図- 2 に示す．尚，au GPS ビッグデータは，上述の通り，au ユーザーの許諾 [9]を元に収集した GPS 位置情報が元となるが，公的統計情報に基づき，人口と同数に拡大推計されたデータとして，GEOTRA に提供されている．また，性別，年代，居住地等の属性情報が付与されており，多角的かつ網羅的な分析が可能である．さらに，利用許諾を元にデータを収集した上で個人を特定出来ない様に秘匿化処理及び統計処理を実施しており，個人情報を保護し分析を行うことが可能である．

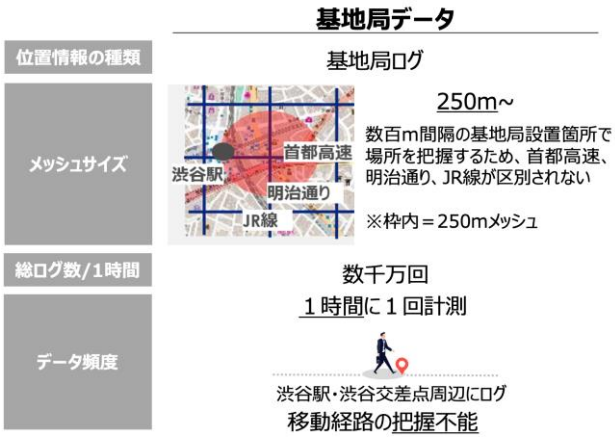


図-1 基地局データの特徴



図-2 GPS データの特徴

(2) データの加工技術

au GPS ビッグデータ（集計データ）を元に，GEOTRA が保有するデータ加工技術を用いて，データの加工を行う．具体的には，Simulated Annealing 手法を始めとした複数の加工技術等により，トリップデータ（非集計データ）を作成する．また，同データを元に，交通シミュレーション技術及び経路分析技術を用いて，地図データや道路ネットワークデータ等を統合することで，経路情報を補完し，トリップに加え，経路分析が可能なデータを作成する．

GAD の作成プロセスについて図-3 に示す．

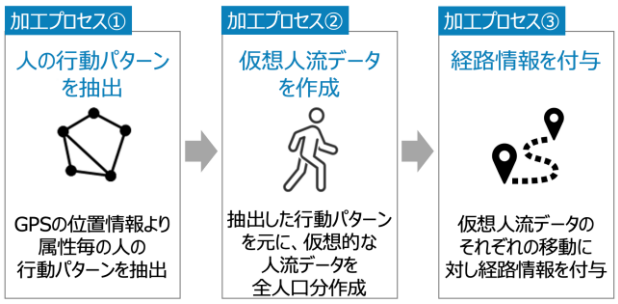


図-3 GAD 作成のプロセス

＜加工プロセス①＞

データ加工については，まず au GPS ビッグデータより，属性毎の人の行動パターンを確率モデルとして抽出する．具体的には，性別，年代，居住地，勤務地などの属性をもとに，au GPS ビッグデータから属性毎の移動履歴を抽出の上，これらの移動履歴を元に，属性毎の行動パターンを地点間の遷移確率として作成する．また，au GPS ビックデータは 15 分以上同じ場所にとどまると滞在と判定し，滞在と滞在の間の移動をトリップと定義している．

＜加工プロセス②＞
抽出した行動パターンを元に，全人口分の仮想的な人流データを作成する．これらは，対象都市の全人口分の仮想的な人口を国勢調査の性別、年齢、居住地（市区町村単位）の集計と合致するように作成したものであり，これらのデータと抽出した行動パターン（遷移確率）に基づいて各人口一人一人の行動データを作成する．その後，パターン抽出元の元データとの統計的な差異が一定範囲以下に収まるまで最適化計算を繰り返すことにより人流データを作成する．

＜加工プロセス③＞
最後に，作成した人流データのそれぞれの移動に対して，MATsim[10]を用いて経路情報を付与する．抽出した行動パターンに基づいて作成された人流データはあくまで地点間の移動データ（A 地点→B 地点へ移動等）であるため，交通シミュレーション手法を用いて各移動に対して経路情報を付与することにより最終的な非集計データを作成する．
このように生成したデータを用いることで，一人一人のトリップデータを用いた都市全域の OD 分析（Origin-Destination）や経路分析，またそれらに性別・年代等の属性情報や移動目的（通勤，帰宅，買い物・・等），移動手段（車・鉄道・徒歩・・等）に関する情報を組みあわせた多角的な交通流分析が可能となる．尚，上記により生成される GAD は，データ取得期間における平休日毎の平均的な移動を表現したデータとして出力される．このときのデータフォーマットのイメージを図-4 に示す．

ID	性別	年代	出発時間	到着時刻	移動目的	移動手段	始点(経度)	始点(緯度)	…
034	**男性**	**30代**	**7:12**	**8:00**	**通勤**	**車**	**139.11**	**36.44**	**…**
111	女性	40代	7:14	8:58	通勤	鉄道	139.11	36.44	…
006	女性	60代	7:31	7:54	買い物	徒歩	139.11	36.44	…
239	男性	20代	7:33	8:33	通学	鉄道	139.11	36.44	…
099	男性	50代	8:00	8:45	出勤	鉄道	139.11	36.44	…
542	女性	20代	8:10	8:30	食事	徒歩	139.11	36.44	…
090	男性	30代	8:16	8:40	通院	車	139.11	36.44	…
034	**男性**	**30代**	**8:00**	**8:25**	**食事**	**徒歩**	**139.29**	**34.32**	**…**
…	…	…	…	…	…	…	…	…	…

図-4　GAD データフォーマット

(3)　データの仕様

前述の方法により作成される GAD の仕様を表 -1 に示す．GAD の移動手段については端末の移動速度情報と道路・鉄道路線等外部データより取得した，交通ネットワーク情報より判定する．また，移動目的については滞在先の位置情報と同位置にある施設情報を組み合わせ，その滞在時間から判定する．

表-1　GAD の仕様

データ項目	値・属性に関する説明
性別	男性/女性
年代	20 代/30 代/40 代/50 代/60 代以上（※10 代以下はデータ無し）
居住地	居住地の緯度経度
勤務地	勤務地の緯度経度
移動目的	帰宅/通勤/食事/買い物/アウトドア/食品・日用品購入/通院/エンターテインメント/その他
移動手段	車/鉄道/徒歩/自転車・その他
移動時間	単位：分（移動に要した時間を算出）
移動距離	単位：km（移動に要した距離を算出）
滞在時間	単位：分（滞在先での滞在時間を算出）
出発時刻	0:00-24:00
到着時刻	0:00-24:00
出発地	出発地点の緯度経度
到着地	到着地点の緯度経度
移動経路	OD 間の経路情報（利用した道路，路線等）に関する情報

また，OD データはエリアを指定して作成する必要がある．エリアは「基準エリア」と「対象エリア」の 2 エリアを指定する．基準エリアにおいて，OD データでは，基準エリア内に一度でも足を踏み入れた人の移動を捕捉し，基準エリア内に一度も足を踏み入れていない人の移動は捕捉されない．対象エリアにおいて，OD データでは基準エリア内にいた人の対象エリア内の移動を捕捉する．基準エリアにいた人の対象エリア外の移動は捕捉されない．対象エリアは通常，基準エリアと同じエリア或いは基準エリアを包含するより広範なエリアとして設定される．エリアについて図示したものを図-5 に示す．

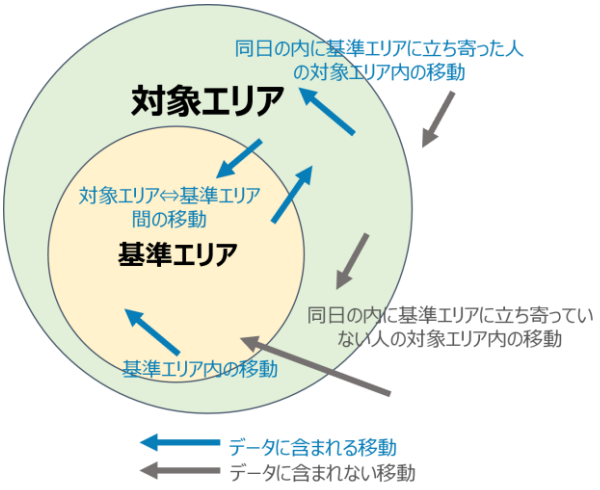


図-5　基準エリアと対象エリアの説明

また，パーソントリップ調査（以下，PT 調査）と GAD の仕様に関する比較を表-2 に示す．

表-2　PT 調査と GAD の比較

項目	PT 調査	GAD
調査対象	都市圏内居住者（標本として数%を抽出）	(株)au の携帯電話所有者（許諾を得た場合）
調査日	特定の 1 日	集計対象期間の毎日
調査頻度	概ね 10 年に 1 度	毎日
調査地域	都市圏単位	日本全国
属性	性別・年代・居住地	性別・年代（20 代以上）※・居住地・勤務地
時間解像度	分単位	分単位
空間解像度	中ゾーン（最小は夜間人口約 1 万 5 千人を目安とした小ゾーン）	100m~（来訪判定は 10m~）
移動目的	自宅ー勤務/自宅ー通学/自宅ー業務/自宅ー私事/帰宅/勤務・業務/その他私事/不明	帰宅/通勤/食事/買い物/アウトドア/食品・日用品購入/通院/エンターテインメント/その他
移動手段	鉄道/路線バス/乗用車など 14 区分	車/鉄道/徒歩/自転車・その他

※プライバシー保護の観点から 10 代以下の属性は除く

3.　データの有用性評価

(1)　対象地域の選定

PT 調査を実施した都市圏として「東京都市圏」を選定した．具体的な対象地域としては，中央区，千代田区，港区を基準エリアとし，さらに東京都，神奈川県，埼玉県，千葉県を対象エリアとして取り上げる．対象エリアの交通量を図-6 に示す．また，OD として対象エリアから基準エリアの来訪者のイメージを図-7 に示す．全ての属性を一度に可視化することは図示を複雑化させるため，具体的な一例として，図-7 では，20 代の男性が千代田区を訪れるケースに焦点を当てた可視化を行っている．

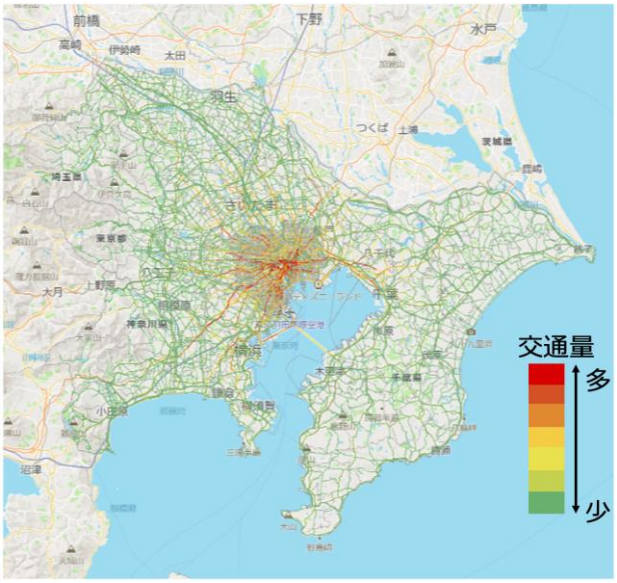


図-6　対象地域の交通量

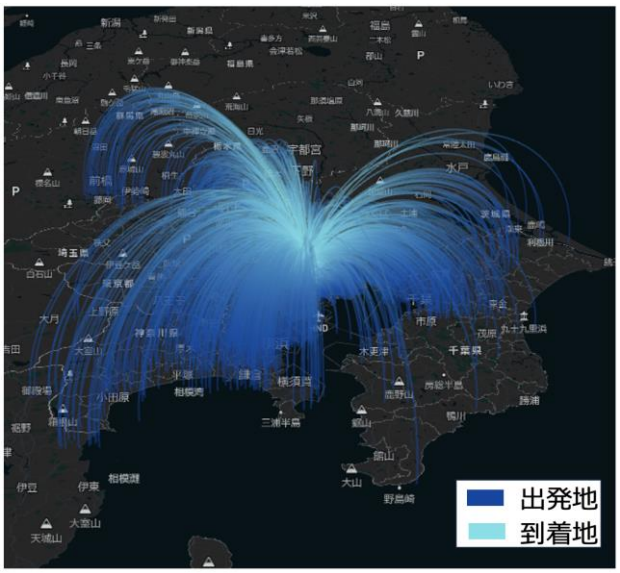


図-7　対象地域の千代田区への来訪者
（20 代の男性のみを可視化）

(2)　対象日の選定

対象とする日のデータ選定に際しては，PT 調査との比較を考慮し，できる限り時期が接近している日を選択した．具体的には，GAD としては，2019 年 10 月から 11 月にかけての 2 か月間の平日のデータを平均化し，その 1 日分を採用した．一方，PT 調査のデータとしては，2018 年秋の平日における 1 日分のデータを使用している．

(3)　検証方法

移動パターンの集計に際して，以下の基準を設けて分析を行った．まず，出発地あるいは到着地のいずれかが基準エリア内に位置する移動量（トリップ数）を対象として集計を行った．さらに，出発地と到着地の両方が対

象エリアを越えることのない移動も同様に集計した．PT 調査とのデータの比較については，単一の指標に頼るのではなく，データの類似度を多角的に評価するため，複数の統計手法を用いて精度検証を施した．精度検証に当たり，用いた統計手法は表 -3 の通り．また，これらの手法を用いて，精度比較を分析した．

スピアマンの順位相関係数は，2 つのデータの大小関係の一致度を-1 から 1 の範囲で示すものである．1 に近いほど，データ間の一致度が高いことを示す．ピアソンの相関係数は，量的変数間の線形関係を評価し，データが正規分布を有していると仮定するが，スピアマンの順位相関係数は非正規分布データにも対応しており，交通量のような必ずしも正規分布にならない統計量の評価に適しているため，本稿において採用する．

ウィルコクソンの符号順位和検定は，2 つのデータの中央値の差の一致度である p 値を 0 から 1 の範囲で示す．p 値が 0.01 を上回る場合，比較対象の 2 つのデータの間に差異は無いと判断される．t 検定はデータが正規分布を有しているという前提の下，サンプル間の平均値の差を検定するが，ウィルコクソンの符号順位和検定は，スピアマンの順位相関係数同様，データが正規分布に従っているかどうかの前提を必要としない．よって本稿ではウィルコクソンの符号順位和検定を用いるこの検定はデータ量の評価が可能であるが，データの大小関係の傾向については，相関係数を用いて評価する必要がある．

表 -3　精度検証に当たり用いた統計手法

スピアマンの順位相関係数	2 つのデータの大小関係の一致度合いを検証
ウィルコクソンの符号順位和検定	2 つのデータの中央値の差の一致度合いを検証
平均相対誤差	2 つのデータ間の相対誤差の平均を評価

(4)　検証結果

GAD および PT 調査データの両方において，通勤時間帯としての 6 時から 8 時，及び 16 時から 20 時にピークを観察することができる．さらに，PT 調査データは GAD に対して，トリップ数が顕著に多いことが確認される．要因として，GAD には 20 歳未満の通勤・帰宅の移動が含まれていないことが挙げられる．各区の GAD と PT 調査のトリップ数の比較は千代田区：図 -8，中央区：図-9，港区：図-10 の通りである．

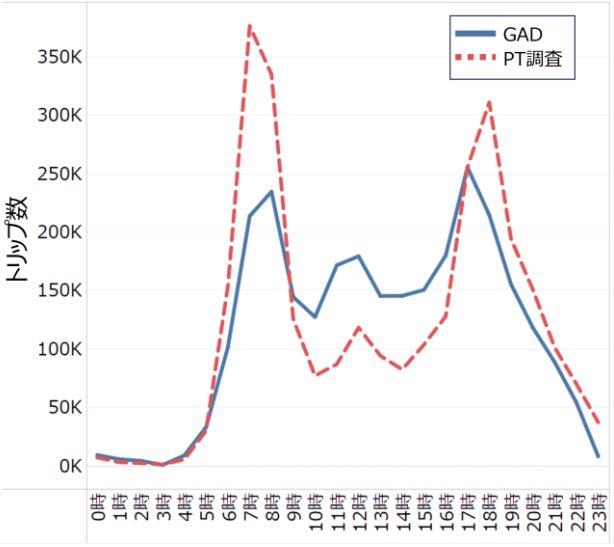


図 -8　GAD と PT 調査のトリップ数の比較（千代田区）

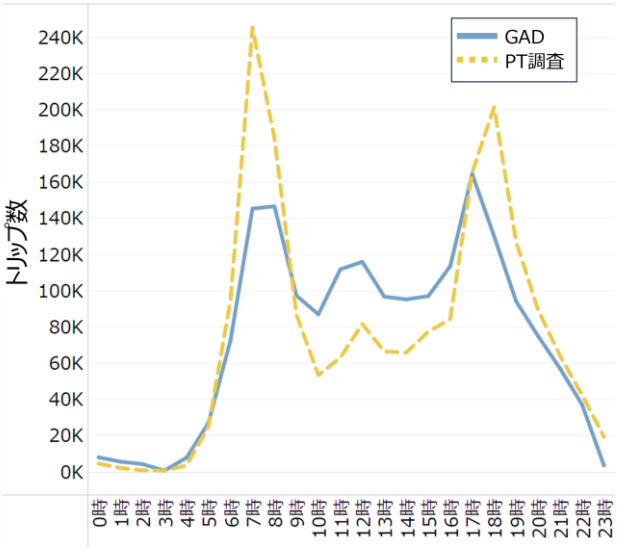


図 -9　GAD と PT 調査のトリップ数の比較（中央区）

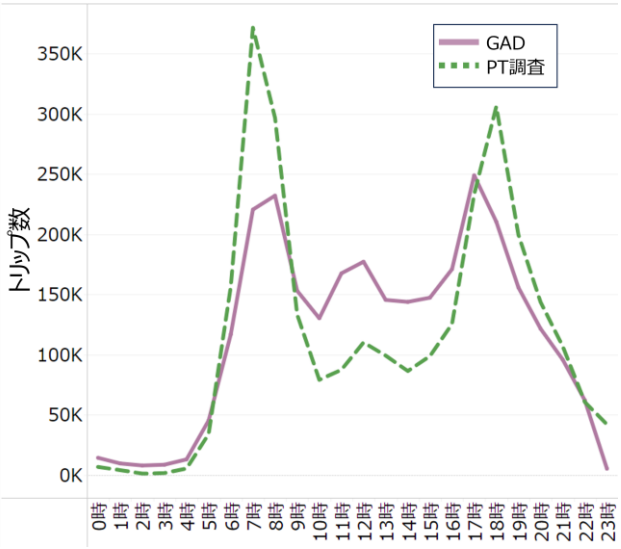


図 -10　GAD と PT 調査のトリップ数の比較（港区）

精度検証結果を表- 4 に示す．スピアマンの順位相関係数は 0.95 と評価され，一定の信頼性を有していることが確認された．さらに，千代田区，中央区，および港区と対象エリア間のトリップ数に関する散布図とその相関関係を記載する．（千代田区：図 -11，中央区：図 -12，港区：図- 13）各区における相関係数も 0.94 以上となっており，一貫した高い一致度が示されている．これらの図では共通して GAD のトリップ数が PT 調査データよりも数倍大きい外れ値が存在するが，これは注目している基準エリア内での移動を指している．GAD では GPS データをもとにデータを生成しているためデータ取得頻度が高い．それに対して PT 調査データは，出発地から到着地までの移動ごとのカウント（目的トリップ）または交通手段ごとのカウント（手段トリップ）が反映されており [11]，同一区内の近距離の移動が記録されにくいことが原因として挙げられる．ウィルコクソンの符号順位和検定の結果，帰無仮説が成立し，絶対量の差は大きいことがわかる．原因として大きく二つが推察される．1 つ目は PT 調査と集計時期による差が考えられる．2 つ目は調査方法による差が考えられる．具体的には GAD は 15 分以上同じ場所にとどまると滞在と判定していることでアンケート調査ベースの PT 調査と比較し，滞在判定が異なってくる可能性がある．以上により，昼間の GAD が PT 調査データのトリップ数を上回る結果となっている．

表-4　精度検証結果

スピアマンの順位相関係数	0.95
ウィルコクソンの符号順位和検定	p 値= 1.0×10^{-80}
平均相対誤差	0.43

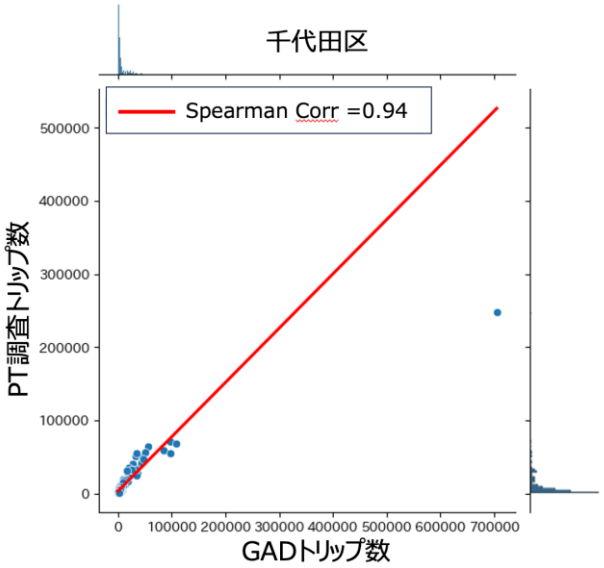


図-11　PT 調査 GAD 相関関係（千代田区）

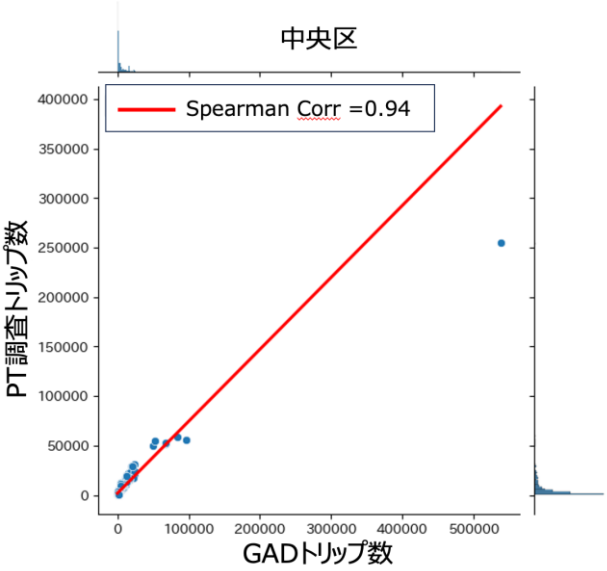


図-12　PT 調査 GAD 相関関係（中央区）

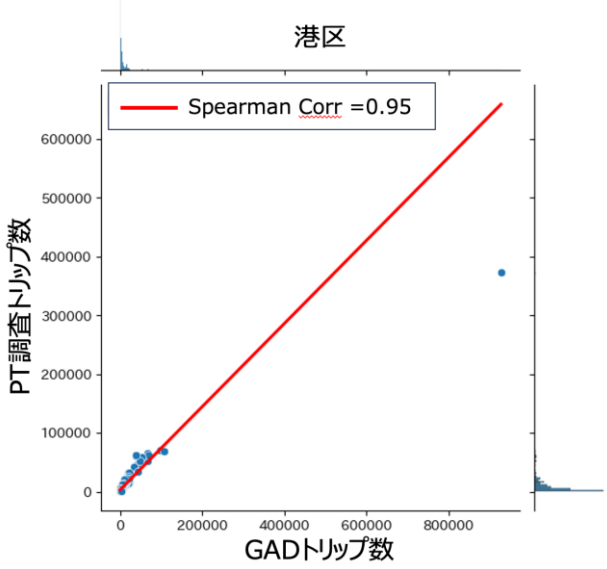


図-13　PT 調査 GAD 相関関係（港区）

また，時間帯によるスピアマンの順位相関係数の詳細は図 -14 の通り．6 時から 22 時までの間は相関係数が 0.60 以上となっており，これにより GAD と PT 調査との間には十分な相関が認められる．しかしながら，23 時から翌日の 5 時までの深夜から早朝の時間帯では，日中と比較して相関係数が低下していることが確認される．この現象の背後には，夜間のトリップ数の減少が影響していると推察される．



図 -14　時間帯別のスピアマンの順位相関係数

4.　おわりに

(1)　結論

本稿では，非集計データの概要を記載し，データの作成方法と既存調査データとの比較検証結果を記載した．日中のデータ精度は一定担保できた．

(2)　今後の課題

本稿では，データの有用性評価に関して，人流量が多い東京都市圏を対象に行った．今後は，より小規模な地域での精度検証についても実施し，より汎用的なデータ構築を進める．また，今回は移動量（トリップ数）に関する比較検証を行ったが，自動車の交通量とトラフィックカウンターデータとの比較検証も実施したい．

(3)　データの活用可能性

今後のデータ活用の可能性として以下三つのキーポイントを挙げる．

新規交通網における検討でのデータ活用：
多くの地方部では，交通需要を正確に評価・予測するための基礎データが不足している．この問題を解決するためには，携帯電話 GPS 位置情報やその他のビッグデータを活用して現況の交通量を再現することが有効である．これにより，新たな交通網の必要性や最適なルートを正確に判断するための基盤となる．

観光促進への活用：
観光業界において，データ活用の重要性は日増しに高まっている．特に，各トリップを具体的に追跡できるデータは，移動経路の解析や居住地，勤務地，また公共施設への入場データとの比較研究において，顕著な価値を提供する．これにより，観光客の移動や滞在の実態，すなわち回遊パターンを精緻に把握することが可能となり，その結果，ターゲットとする観光客層に対するより効果的な施策の策定やアプローチが可能となる．

大規模開発における人流データ活用：
大規模な都市開発や再開発において，人流の変化やその属性の変動は非常に重要な指標となる．特に，新型コロナウイルスの影響や都市の景観変化に伴って，人々の行動パターンや滞在時間，出入りの頻度などが変わる可能性がある．これらのデータを詳細に追跡・分析することで開発計画の最適化やリスク管理が進められる．
以上のように，データの活用は多岐にわたる分野でその価値を発揮している．今後も継続的なデータ収集と分析技術の向上により，更なる地域振興や都市計画の質的向上に寄与したい．

謝辞：本研究を進めるにあたって，一般財団法人計量計画研究所の石井主任研究員に多大なご協力を賜りました．この場を借りて，心より感謝いたします．

参考文献

1) 柴崎真理子，薄葉康生，高山知拡，CHEN Quanjun，高安洋：都市交通分野等に利活用できるリアルタイムモバイルデータの開発，土木計画学研究・講演集 Vol.67，2023.
2) 今井 龍一，藤岡啓太郎，新階 寛恭，池田 大造，永田 智大，矢部 努，重高 浩一，橋本 浩良，柴崎 亮介，関本 義秀：携帯電話網の運用データを用いた人口流動統計の都市交通分野への適用に関する研究，第 52 回土木計画学研究発表会・講演集，Vol.52，2015
3) 生形嘉良，関本義秀，Horanout Teerayut：大規模・長期間 GPS データによる観光統計調査の活用可能性~石川県を事例に~，土木計画学研究・論文集 Vol.53, 2016
4) 山口裕通，奥村誠，金田穂高， 土生恭祐：携帯電話 GPS 情報から分かる熊本地震による行動パターンの被災・回復過程，土木計画学研究・論文集 Vol.73, 2017.
5) 石井良治，末成浩嗣，越智健吾，関信郎，大塚賢太，酒井幸輝，會田優，南川敦宣：携帯電話 GPS ビッグデータの都市交通分野における活用に向けた信頼性検証，土木計画学研究・講演集，Vol.58，2018.
6) 生形嘉良，関本義秀，Horanout Teerayut：携帯電話 GPS ビッグデータの都市交通分野における活用に向けた信頼性検証，土木計画学研究・論文集 Vol.30, 2016
7) 新階 寛恭，今井 龍一，池田 大造，永田 智大，森尾 淳，矢部 努，重高 浩一，橋本 浩良，柴崎 亮介， 関本 義秀：携帯電話網運用データに基づく人口流動統計とパーソントリップ調査手法との比較による活用可能性に関する研究，第 53 回土木計画学研発表会・講演集，Vol.53，2016
8) KDDI Location Data | KDDI 株式会社，（https://k-locationdata.kddi.com/，last access 2023/10/13）
9) 位置情報の活用について | 公開情報 | KDDI 株式会社，（https://www.kddi.com/corporate/kddi/public/bigdata/，last access 2023/10/13）
10) Multi-Agent Transport Simulation，（https://www.matsim.org/，last access 2023/10/13）
11) パーソントリップ調査について |都市交通調査・都市計画調査 | 国土交通省，（https://www.mlit.go.jp/toshi/tosiko/toshi_tosiko_tk_000031.html，last access 2023/10/13）

UTILIZATION OF GPS DISAGGREGATED DATA IN THE FIELD OF URBAN PLANNING CONSIDERING PRIVACY PROTECTION

Takuhiro MORIYAMA, Anna EBATA, Hiroaki Suzuki, Hidetaka HIDA and Nobuhiro JINNAI

Person-trip surveys and traffic volume surveys have long served as foundational data in urban planning. In particular, the use of big data derived from mobile phone GPS location tracking offers significant advantages in evaluating and forecasting traffic demand due to its ability to capture detailed movement histories. However, to ensure privacy, only aggregated data – which is challenging to handle and lacks granularity – has been provided. GEOTRA Co., Ltd. has pioneered a technology that leverages machine learning to generate human mobility data. This data maintains the richness of individual behaviors in a disaggregated manner without compromising privacy.

This paper presents an overview of this disaggregated data, the methodology behind its creation, and a comparison with traditional survey data. Based on our findings, we also propose the possibility of utilizing this data.