

全国擬似人流によるモビリティ デジタルツインの構築

2024年10月8日

東京大学空間情報科学研究センター長
教授・関本義秀

擬似人流データの概要

- 2008年からCSISで「人の流れプロジェクト」を立ち上げ、約**25**都市圏の人の流れデータを提供し、約**300**件の共同研究を実施。
- キャリアの携帯人流もいいが、価格が高止まり気味。海外でも最近Syntheticな人流データ作成の動きが出始めている。
- パーソントリップ調査等がない地域でも、オープンな統計データや、共通の学習パラメータによるエージェントモデルに基づき、携帯電話データとの**0.81**高相関を持つ、国レベルでは世界初の**1.3億**人分の擬似人流を再現し、2022年4月からJoRAS提供開始。（構築・提供ではmdxもフル活用）

国勢調査（2015）

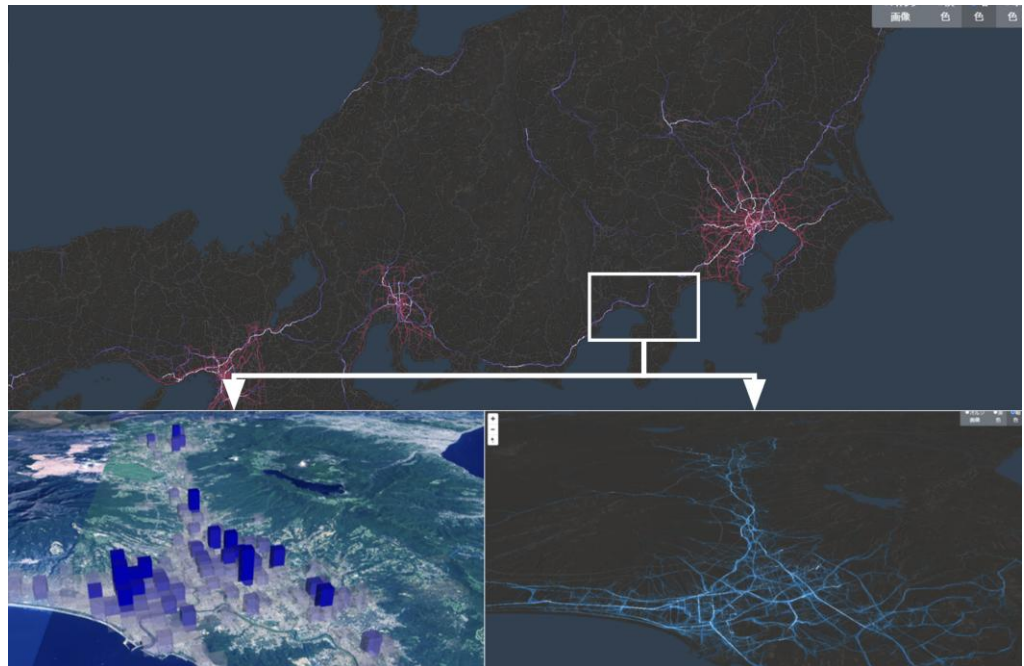
パーソントリップ調査

経済センサス

労働力調査

住宅・建物データ

道路ネットワーク



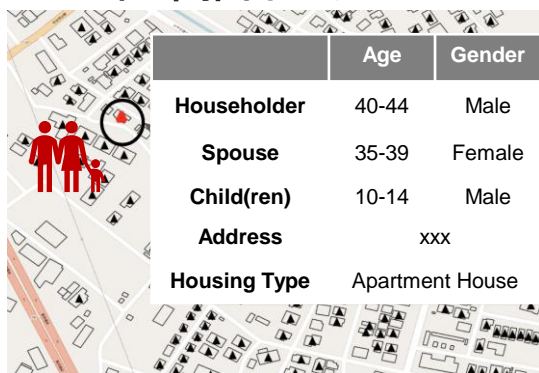
オープンな統計情報等を活用した日本全国の擬似人流再現

擬似人流データの生成方法

オープンな統計データや、共通の学習パラメータによるエージェントモデルに基づき、人々の典型的な日常行動に関する**時々刻々の活動内容、場所、交通手段、移動経路**を推定し、擬似的な表現できる合成データを作成する

① Household Estimation model

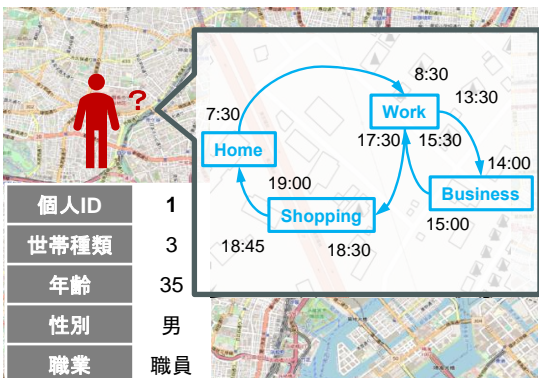
世帯推計モデル



What to do at What time of day ?

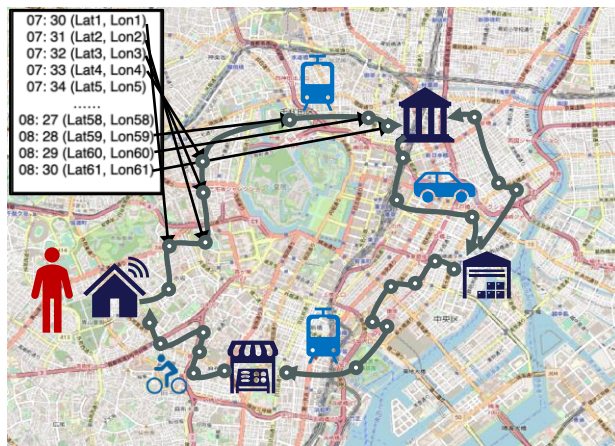
② Activity Generation model

活動生成モデル



⑤ Route Choice / Spatial-temporal Interpolation

経路選択・時空間内挿入処理

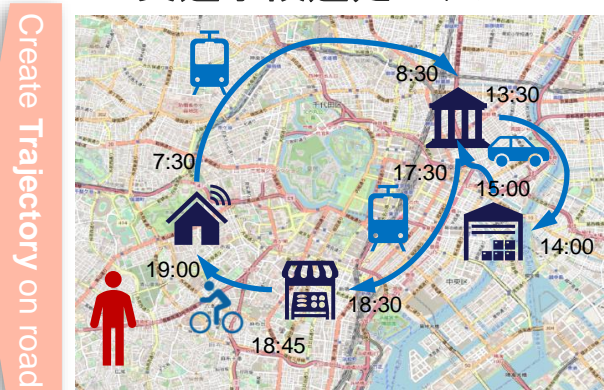


Pseudo People Flow Dataset
擬似人流データセット

From Where to Where ?

④ Transportation Mode Choice model

交通手段選定モデル

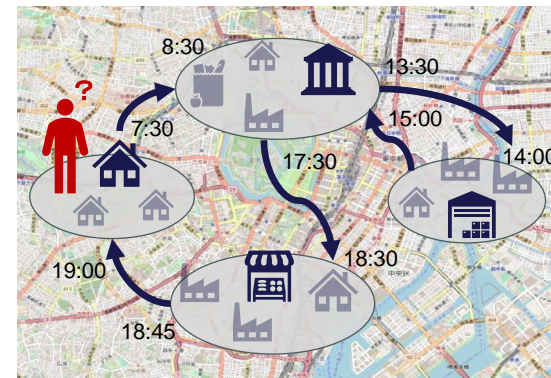


Create Trajectory on road

By what means of transportation ?

③ Activity Location Selection model

活動場所選定モデル

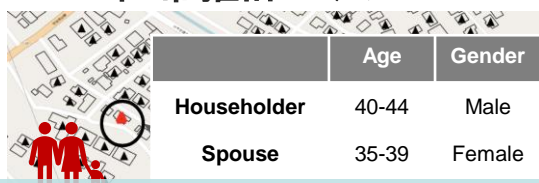


擬似人流データの生成方法

オープンな統計データや、共通の学習パラメータによるエージェントモデルに基づき、人々の典型的な日常行動に関する**時々刻々の活動内容、場所、交通手段、移動経路**を推定し、擬似的な表現できる合成データを作成する

① Household Estimation model

世帯推計モデル



	Age	Gender
Householder	40-44	Male
Spouse	35-39	Female
Child(ren)	10-14	Male
Address	xxx	
Housing Type	Apartment House	

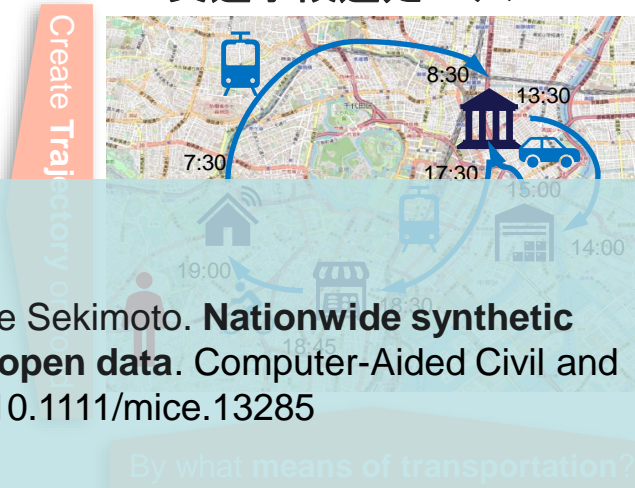
⑤ Route Choice / Spatial-temporal Interpolation

経路選択・時空間内挿入処理



④ Transportation Mode Choice model

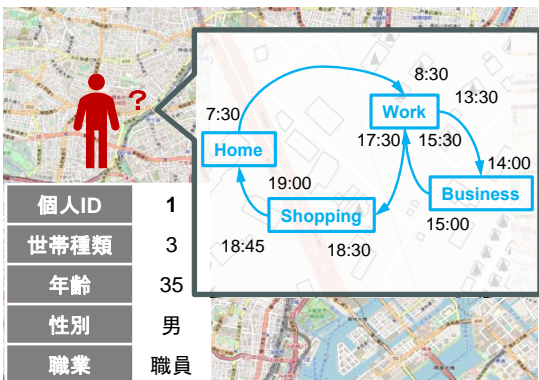
交通手段選定モデル



Takehiro Kashiwama, Yanbo Pang, Yuya Shibuya, Takahiro Yabe, Yoshihide Sekimoto. **Nationwide synthetic human mobility dataset construction from limited travel surveys and open data.** Computer-Aided Civil and Infrastructure Engineering. Available 10 June 2024 online. <https://doi.org/10.1111/mice.13285>

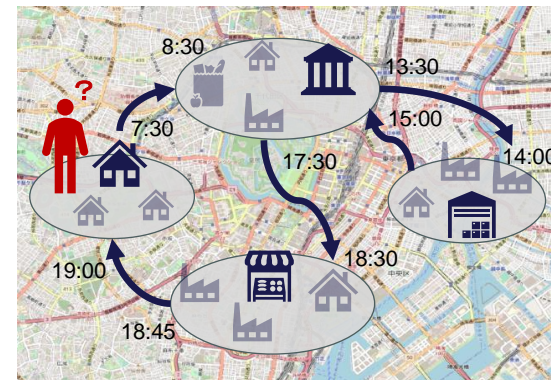
② Activity Generation model

活動生成モデル



③ Activity Location Selection model

活動場所選定モデル

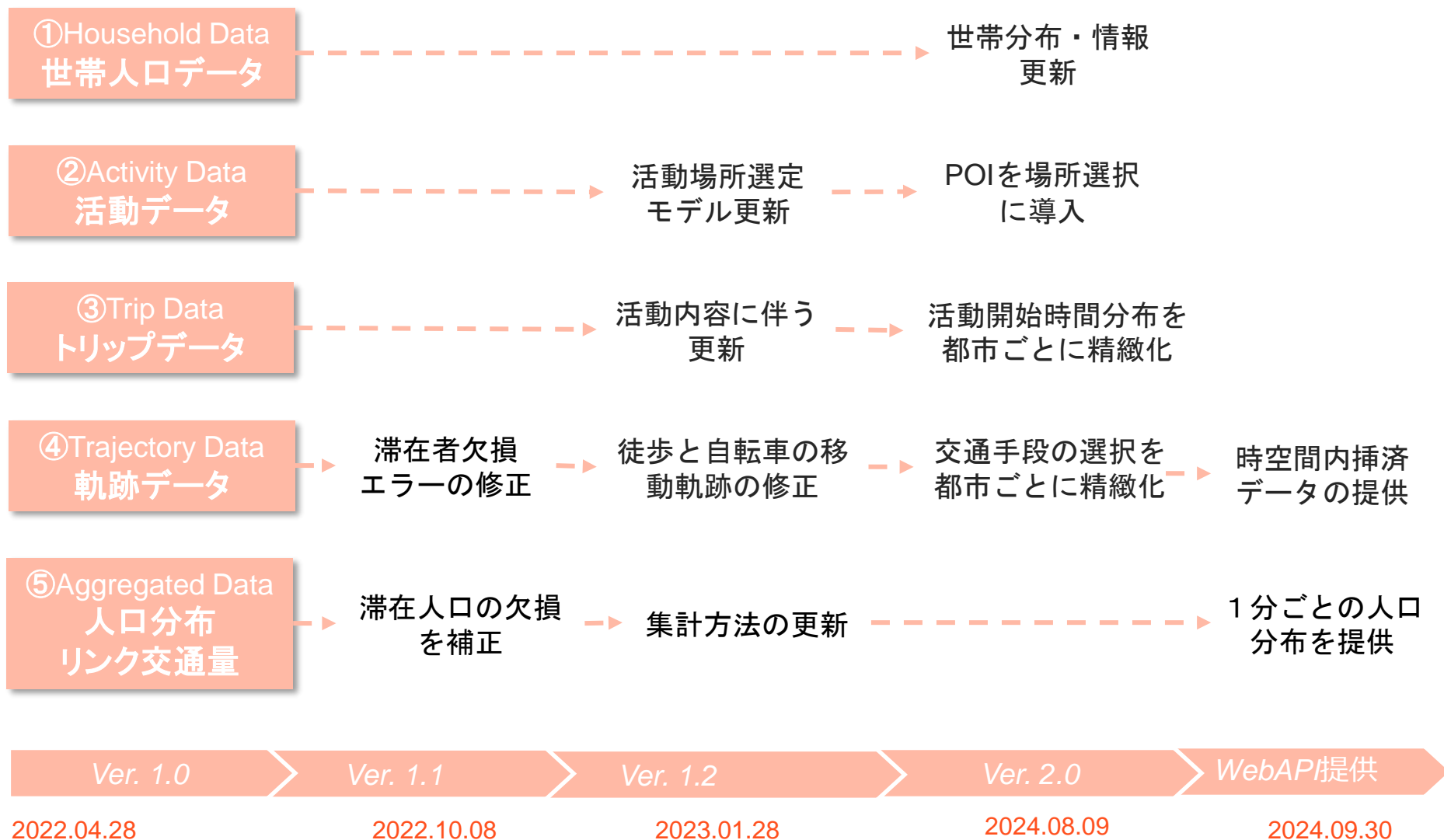


Pseudo People Flow Dataset
擬似人流データセット

From Where to Where ?

擬似人流データの変遷

2022年4月にバージョン1.0をリリースした後、バグ修正を行ったバージョン1.1を経て、現在はバージョン1.2に至る。さらに、本年度末にはバージョン2.0のアップデートを予定



擬似人流データの可視化結果

擬似人流の軌跡データは、GPSデータと同様に、より密な点の集合として表現されます。これにより、人々の移動パターンをより細かく、リアルタイムに追跡することが可能になります。（ブラウザーの性能制限のため、動画は約1%分の人口数の可視化結果）



擬似人流データ：リアルデータとの比較

擬似人流データとリアルデータの違いについて

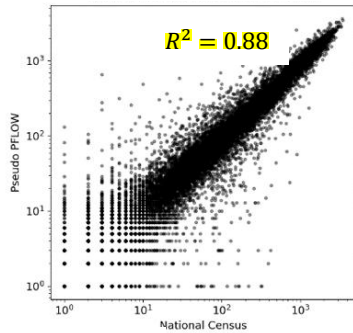
擬似人流データは高精度で補完が可能だが、推定値に基づくため、実データと比べて経路や属性に若干の差異が生じる

	擬似人流データ	リアルデータ(i.e., GPS)
時間解像度	任意の間隔で補完可能 (現在は1分ごとのバージョン)	数秒～数時間
空間解像度	道路網・建物単位で高精度	数メートル～
期間	典型的平日の1日(2020年)	長期間の取得が可能
サンプル率	全人口	1～5%
個人属性	年齢・性別・就業状態(推定値)	(一部のID) 年齢性別が取得可能
経路再現精度	公共交通機関の経路及び最短経路で推定される	一部の経路は道路網上で再現できないが、 迂回や個人の好みを反映可能
活動再現精度	日常的な活動が推定される	(一部のID) 全ての活動が取得可能

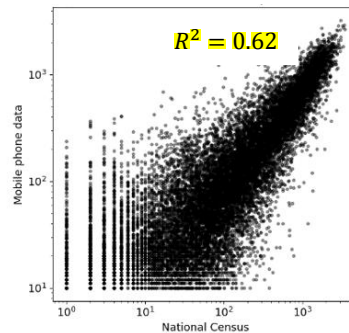
擬似人流データの品質について

擬似人流データは集計レベルで、国勢調査及び携帯電話データで検証した結果、時間帯による人口分布の精度が高いことが判明した

擬似人流 v.s. 国勢調査

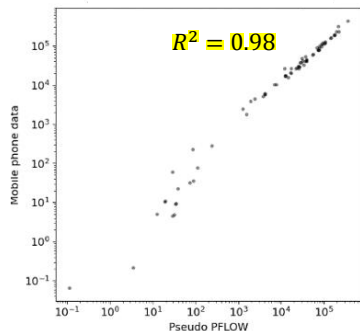


擬似人流 v.s. 携帯電話データ

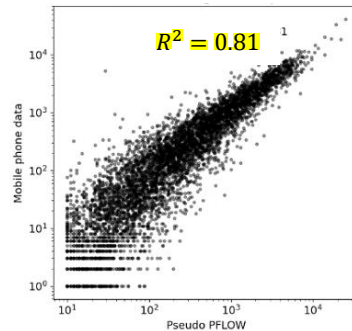


早朝時間帯において、携帯電話データよりも国勢調査データとの相関がより密接である

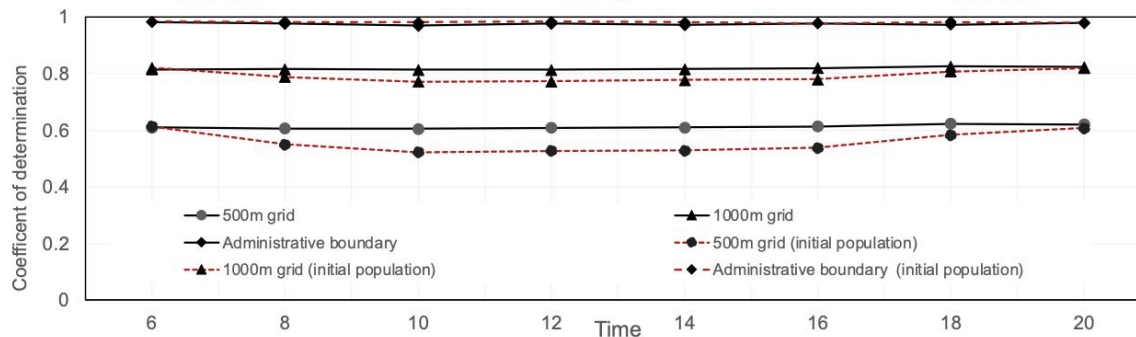
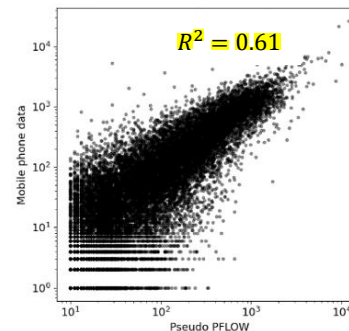
行政レベル(正午)



1km解像度(正午)



500m解像度(正午)



精度は空間解像度に大きく関わる。行政レベルで携帯電話データとの相関が0.98に達し、1000×1000mのメッシュでは精度がR²=0.81と高く、実データの有効な代替となります

擬似人流データの品質について

- 擬似人流データと令和3年度全国都市交通特性調査の比較により、外出率やトリップ数においては全体的に高い一致度が見られ、日常的な移動パターンの再現において擬似データの有効性が確認された。
- 一方で、交通手段別の構成比や移動目的においては、地域ごとに差異が見られ、一部の都市では公共交通機関の利用割合や自動車利用率の再現精度に改善の余地があることが分かりました

都道府県コード	都道府県	市町村コード	都市名	都市セグメント	都市セグメント	外出率	擬似人流		トリップ数		擬似人流		代表交通手段別構成比										
							ver1.2	ver2.0	グロス	ネット	ver1.2	ver2.0	鉄道	ver1.2	ver2.0	バス	ver1.2	ver2.0	自動車(運転)	自動車(同乗)	自動車(合計)	ver1.2(合計)	ver2.0(合計)
1	北海道	1100	札幌市	地方中核都市	中心都市	71.0	75.3	74.5	1.79	2.52	2.88	2.34	17.1	1.5	10.4	4.2	—	1.5	32.7	11.1	43.8	52.9	47.2
2	青森県	2202	弘前市	地方中核都市	中心都市	69.1	75.1	74.6	1.86	2.70	2.86	2.34	2.1	1.4	2.3	2.6	—	0.9	56.7	13.6	70.3	49.7	41.2
3	岩手県	3201	盛岡市	地方中核都市	中心都市	74.0	76.3	75.5	1.97	2.66	2.90	2.38	1.8	2.1	0.9	4.2	—	2.1	48.2	12.1	60.2	51.6	45.1
4	宮城県	4100	仙台市	地方中核都市	中心都市	72.2	75.5	74.9	1.87	2.60	2.88	2.35	11.0	1.9	12.6	4.6	—	0.4	38.5	11.6	50.2	54.9	36.7
5	秋田県	5207	湯沢市	地方中心都市	—	67.6	74.5	73.9	1.76	2.60	2.84	2.30	0.7	1.7	1.2	0.5	—	4.2	66.3	16.3	82.6	49.9	56.2
7	福島県	7203	郡山市	地方中核都市	中心都市	77.2	76.2	75.6	2.02	2.61	2.90	2.37	1.9	1.7	3.1	1.6	—	6.4	56.3	15.0	71.3	52.0	61.2
9	栃木県	9201	宇都宮市	地方中核都市	中心都市	78.1	76.4	62.8	2.15	2.75	2.91	2.67	2.2	2.1	1.8	2.5	—	5.8	52.2	14.7	66.9	53.7	53.1
11	埼玉県	11100	さいたま	三大都市圏	中心都市	74.0	76.2	63.1	1.94	2.63	2.91	2.32	26.5	3.7	31.3	1.5	—	3.4	17.8	8.0	25.8	57.1	25.4
12	千葉県	12100	千葉市	三大都市圏	中心都市	71.6	76.0	63.6	1.81	2.52	2.90	2.49	25.6	3.6	21.2	1.9	—	4.2	25.4	10.5	35.9	59.8	37.3
13	東京都	13100	東京都 23	三大都市圏	中心都市	67.8	76.4	64.8	1.76	2.60	2.91	2.56	35.6	3.0	30.1	3.0	—	4.5	6.3	5.7	12.0	59.8	21.1
14	神奈川県	14100	横浜市	三大都市圏	中心都市	73.0	76.6	64.2	1.89	2.58	2.92	2.43	31.2	3.5	18.5	4.9	—	11.4	14.7	7.2	21.9	56.5	15.3
14	神奈川県	14130	川崎市	三大都市圏	中心都市	69.2	76.6	64.2	1.71	2.48	2.92	2.43	34.4	3.5	26.7	4.2	—	8.9	11.6	5.7	17.3	56.5	20.3
15	新潟県	15222	上越市	地方中心都市	—	74.0	76.7	65.0	1.98	2.67	2.92	2.01	2.0	2.0	4.7	0.8	—	13.2	65.0	11.8	76.8	50.4	41.8
17	石川県	17201	金沢市	地方中核都市	中心都市	77.2	77.1	86.0	2.13	2.75	2.93	2.96	2.3	1.7	2.6	5.7	—	21.1	53.1	12.2	65.4	53.0	52.4
20	長野県	20209	伊那市	地方中心都市	—	78.7	77.0	85.9	2.13	2.71	2.90	2.94	1.9	2.0	4.1	0.3	—	10.7	66.4	15.2	81.6	53.2	48.4
22	静岡県	22100	静岡市	地方中核都市	中心都市	79.2	76.2	85.2	2.18	2.76	2.92	2.91	6.4	2.0	3.1	2.6	—	13.3	39.3	13.1	52.5	53.2	37.4
23	愛知県	23100	名古屋市	三大都市圏	中心都市	71.5	77.3	85.3	1.78	2.49	2.94	2.83	19.4	2.8	11.5	2.5	—	10.2	31.6	9.5	41.1	57.9	48.9
26	京都市	26100	京都市	三大都市圏	中心都市	75.9	75.8	62.1	2.00	2.64	2.89	1.91	20.4	2.5	12.3	5.1	—	9.6	14.0	7.9	21.8	56.2	38.4
27	大阪府	27100	大阪市	三大都市圏	中心都市	78.7	75.4	61.7	2.07	2.64	2.88	1.90	32.4	2.9	20.1	1.2	—	4.5	9.3	3.4	12.7	59.7	19.9
28	兵庫県	28100	神戸市	三大都市圏	中心都市	74.2	75.6	61.7	1.91	2.58	2.89	1.90	27.0	2.8	18.0	4.1	—	12.8	18.3	7.7	26.1	53.3	21.2
32	鳥取県	32201	松江市	地方中核都市	中心都市	73.8	76.2	62.7	1.89	2.57	2.90	1.91	1.4	1.3	3.1	1.4	—	7.6	55.8	13.0	68.8	49.0	41.6
34	広島県	34100	広島市	地方中核都市	中心都市	73.5	76.0	64.8	1.81	2.47	2.90	1.89	8.4	1.8	3.1	3.5	—	32.3	10.5	42.8	52.3	39.5	
35	山口県	35211	長門市	地方中心都市	—	68.9	74.9	59.8	1.82	2.65	2.86	1.76	1.3	1.9	3.1	0.3	—	58.5	17.4	75.9	49.9	39.6	
36	徳島県	36201	徳島市	地方中核都市	中心都市	76.3	75.1	61.7	2.07	2.71	2.86	1.86	1.1	1.9	3.1	1.0	—	4.2	50.7	16.9	67.6	53.3	41.3
38	愛媛県	38201	松山市	地方中核都市	中心都市	77.0	75.1	60.4	2.07	2.69	2.87	1.80	2.5	1.6	3.1	1.7	—	12.3	40.1	12.9	52.9	48.2	35.7
38	愛媛県	38202	今治市	地方中心都市	—	77.4	75.1	60.4	2.15	2.78	2.87	1.80	0.8	1.6	3.1	0.2	—	6.2	51.9	17.5	69.3	48.2	48.2
39	高知県	39201	高知市	地方中核都市	中心都市	76.5	74.9	61.3	2.08	2.72	2.86	1.81	0.3	1.4	3.1	1.5	—	4.0	46.0	13.6	59.6	50.4	38.6
40	福岡県	40100	北九州市	地方中核都市	中心都市	72.2	75.5	71.1	1.78	2.47	2.88	1.91	6.9	2.2	3.1	6.7	—	43.6	13.9	57.5	55.4	36.7	
40	福岡県	40130	福岡市	地方中核都市	中心都市	75.9	75.5	71.1	1.97	2.60	2.88	1.91	14.0	2.2	3.1	6.9	—	18.1	23.2	9.9	33.1	55.4	36.7
43	熊本県	43100	熊本市	地方中核都市	中心都市	77.0	76.1	63.7	2.09	2.72	2.90	1.95	1.2	1.6	3.1	3.7	—	12.3	48.7	12.5	61.2	49.4	42.5
43	熊本県	43203	人吉市	地方中心都市	—	73.9	76.1	63.7	2.06	2.79	2.90	1.95	0.0	1.6	3.1	0.3	—	9.0	60.9	15.7	76.7	49.4	42.5
46	鹿児島県	46201	鹿児島市	地方中核都市	中心都市	77.1	75.1	61.4	2.18	2.83	2.87	1.86	2.4	1.1	3.2	4.6	—	13.1	48.7	13.8	62.5	49.9	30.4

擬似人流データ：適用範囲と限界

擬似人流データの利用状況

交通予測，病気の拡散，都市計画，汚染などの社会問題広い研究領域で利用

1.3億人

典型的な1日の移動データ

1,724市区町村

日本全国網羅

22研究機関

幅広い研究課題に利用

代表者所属	代表者氏名	研究題目
東京大学大学院工学系研究科	薄井宏行	公共トイレの配置方針に関する自治体間の比較分析
京都大学経済研究所	森知也	都市集積と地域経済圏の同定
東京大学大学院工学系研究科	浅見泰司	都市の物的環境と歩行行動の関係
九州大学大学院 人間環境学研究院	蕭 耕偉郎	人流データ分析による都市環境の変遷に伴う人間行動変容の解明
法政大学 デザイン工学部	今井 龍一	GPSデータと都市開発をつなぐプロジェクト
東北大学 大学院情報科学研究科	井上 亮	移動履歴データに基づく地域分類手法の開発
長崎大学大学院 熱帯医学・グローバルヘルス研究科	遠藤 彰	Infectious disease modelling informed by spatial and social settings data
神戸大学大学院 人間発達環境学研究科	内山 愉太	居住地の環境特性と緑地への訪問パターンの統合的解析
東京大学生産技術研究所	櫻井雄大	歩行者のトリップ距離分布を考慮した歩行者ネットワークの評価
東京大学 情報基盤センター	川瀬 純也	配列アライメント手法による大規模移動データの類型化手法についての研究
東北大学 データ駆動科学研究センター	大山 智也	治安の評価と予測の適正化に関する研究
岡山理科大学	廣田 雅春	疑似人流データを用いた人々の移動パターンと地域の役割の分析
富山大学 都市・交通デザイン学科	土屋 泰樹	グルメレビューサイトの都市分析への応用の試み
東京大学 社会基盤学専攻	中尾俊介	松山における土地利用の変遷
Hong Kong University	卢佩莹	社会的隔離の緩和における公共交通機関の役割
神戸大学計算社会科学センター	上東 貴志	人流データを用いた少子化対策研究
慶應義塾大学	古谷知之	近隣地域への観光プロセス分析
...

擬似人流データの活用可能な分野と制約

適した研究・活動

1. 都市計画・交通シミュレーション

- 都市全体の人の流れや公共交通機関の利用状況を分析し、交通網の改善や新規路線計画のシミュレーションに活用可能。
- 大規模な都市デザインや、都市開発エリアでの交通需要予測に有効。

2. 防災・災害対策

- 災害時の避難シミュレーションや、人々の移動パターンを利用した防災計画の立
- 災害後の人流回復状況の予測やシミュレーションも可能。

3. 環境・エネルギー分析

- 人の移動パターンに基づいた環境影響（大気質、騒音、エネルギー消費量）評価が可能。自動車や公共交通の利用データを活用した二酸化炭素排出量の推定も

4. 商業エリアの活性化

- 人の流れを活用した商業施設の立地分析や、商圈のポテンシャル評価に適している。
- 観光地やショッピングエリアの人流れを可視化し、マーケティングや集客戦略の立案に貢献。

5. 教育への貢献

擬似人流データの活用可能な分野と制約

あまり向かない研究・使い方

1. 個人の詳細な行動解析

- 擬似人流データは集団ベースでのシミュレーションに優れているが、個々人の詳細な移動経路やライフスタイルに基づく精密な分析には不向き
- 特定の個人の移動履歴や行動パターンの再現には、リアルタイムのGPSデータの方が適している

2. リアルタイムの交通モニタリング

- 擬似人流データは過去のデータやモデルに基づいた予測であり、リアルタイムの交通状況を監視・反映する用途には向いていない
- 突発的なイベントや事故による交通変動を反映することは難しい

3. 細かい時間・空間解像度が求められる研究

- 建物内部や特定の細かいエリアにおける詳細な行動分析には向かない
- 分単位や秒単位の移動データを必要とする研究には制約がある

擬似人流データ：LLMの適用の可能性

擬似人流2024年度実施状況

7月から9月にかけて行われた擬似人流データの主な進捗を以下の3点にまとめています

- WebAPIの開発と検証：APIの機能検証を完了し、データ提供サービスの準備を進めています。これにより、WebAPI経由でのデータアクセスが可能となった
- 生成系AIの適用：Llama3をベースにファインチューニングを行い、擬似人流データの自動生成が可能になりました。小規模の実証実験において、リアルな擬似人流データの生成に成功し、実際のデータと高い相関を持つことが確認されている
- 擬似人流データと実データの比較分析：移動頻度かつ日々の変動を特定。現在、非定期的な行動の精度向上に向けたアルゴリズムの改善を進めている

	5	6	7	8	9	10	11	12	1	2	3
擬似人流（通常版）v2の改善と提供	API仕様検討と時間内挿版を準備		API開発と機能検証し、サービス提供開始			POI（商業・観光施設など）の属性を考慮し、目的地選択を精緻化			交通手段選択モデルの精緻化		
擬似人流への生成系アルゴリズムの適用による高度化	擬似人流仕様の生成系モデルを構築		生成系AIモデル（GPT等）を用いて擬似人流データの自動化生成を試作・検証			少量の実データで事前学習済みモデルをファインチューニングし、シナリオ特化したモデル・人流データを試作・検証					
擬似人流データと実データとの統合	動き統計データを用いて、擬似人流v2の地域ごとの移動量・人口分布を検証		ポイントデータを用いて擬似人流における非日常的行動の不足分を特定する			通院、日用品以外の買物、観光、社交、業務など行動モデルを構築し、擬似人流モデルに反映			動き統計データを用いて、擬似人流モデルの外出率・トリップ数などパラメータを補正		



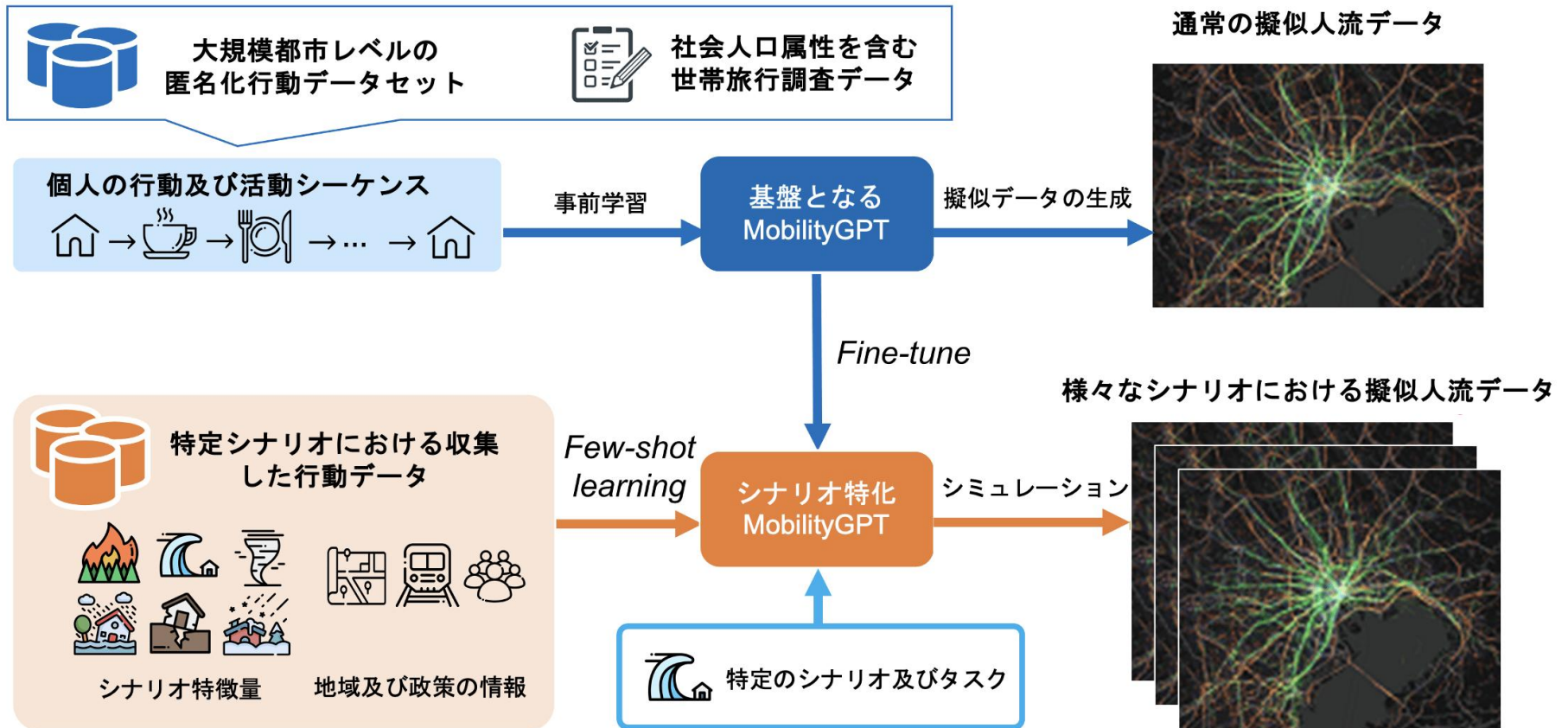
第1回目研究会



今段階の内容

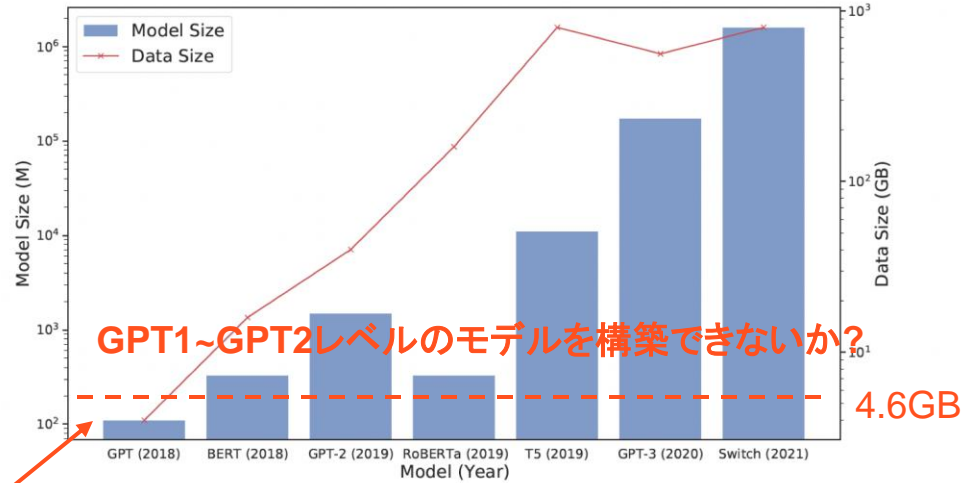
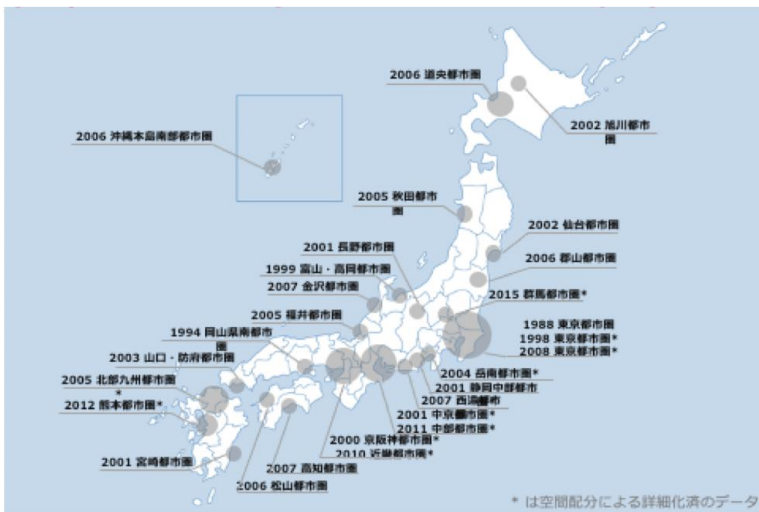
擬似人流×生成系AIについて

- 擬似人流はさまざまな地域、シナリオ（イベント、災害、政策シミュレーション）に対応する必要
 - 既存の手法では、事例ごとにデータ収集とモデル訓練を行う必要があり、莫大なコストが発生する
 - 通常の前学習データで事前学習後、特定タスク（例：イベント、災害、政策シミュレーション）のため、少量のデータ（GPS）で微調整を行い、学習コストを抑えるGenerative Pre-trained (GPT) アプローチを考えている



擬似人流x生成系AIについて

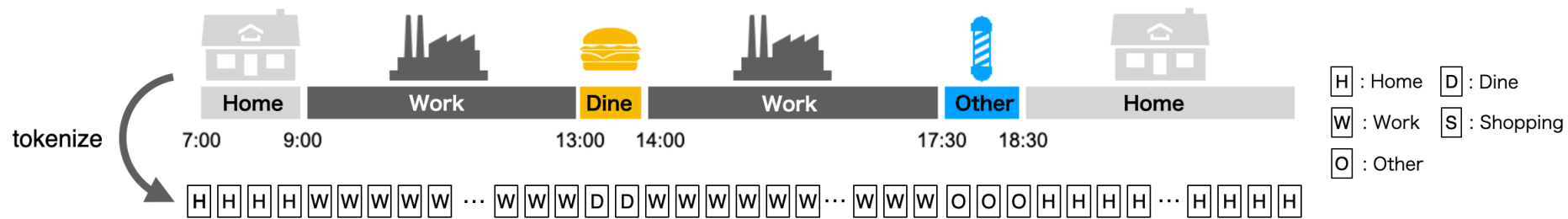
- 人流データをどのようにトークン化するべきか？
 - 自然言語では単語 (word) 単位でトークン化できるが、人流では？
 - 活動内容 (在宅、勤務、通学、買い物、外食、業務、通院) 単語とみなし→n_vocab=7, 1日を15分単位で区切りしたら→token長24h* 4 =96
- どのくらいデータが必要なのか？



CSIS所有する、全国**36**都市約**600**万人分のデータ、約**700GB**となり、前処理トークン化 (活動を単語化) 後、**4.6GB**の訓練データを得る

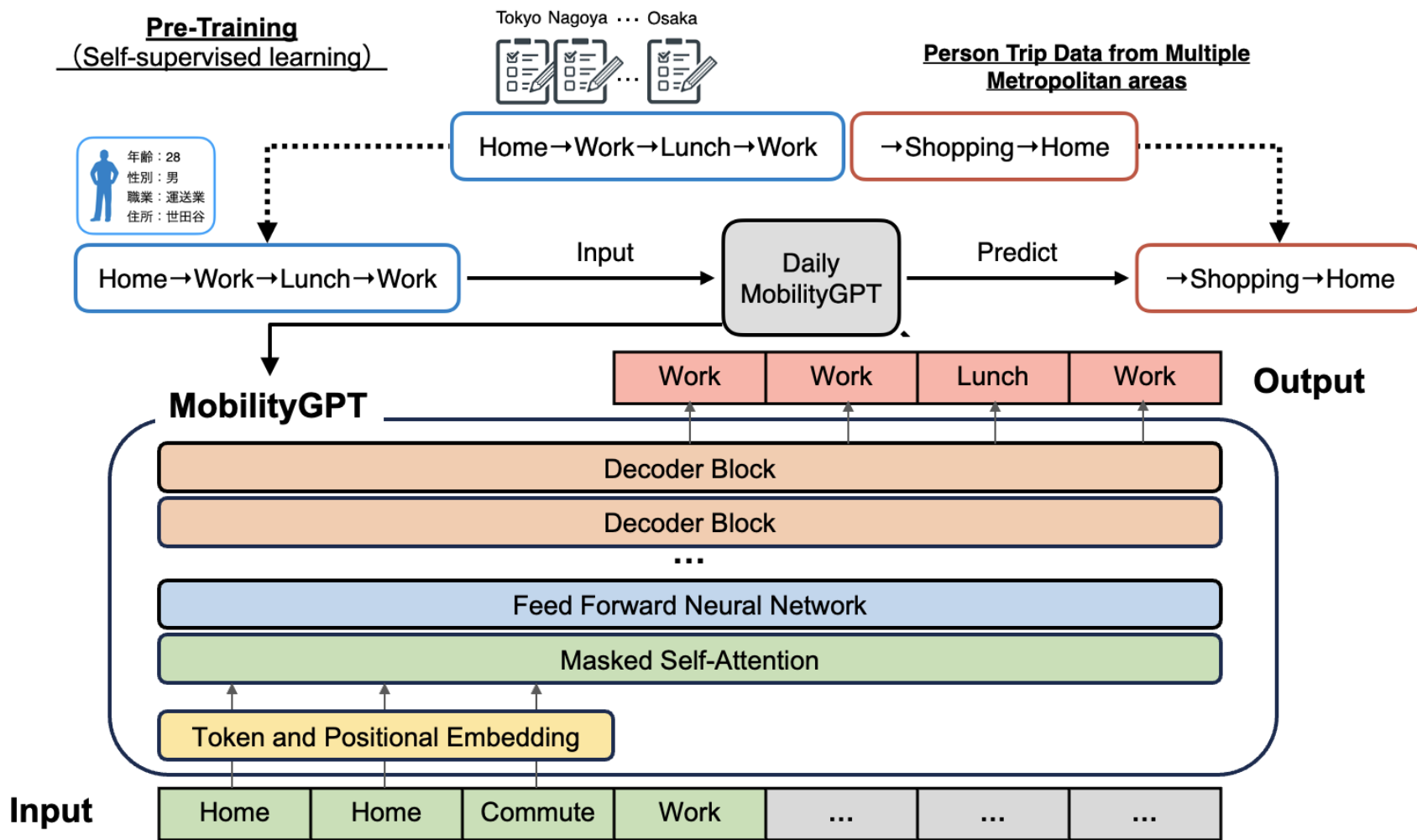
(b) The model size and data size applied by recent NLP PTMs. A base-10 log scale is used for the figure.

Han, Xu, et al. "Pre-trained models: Past, present and future." AI Open 2 (2021): 225-250.



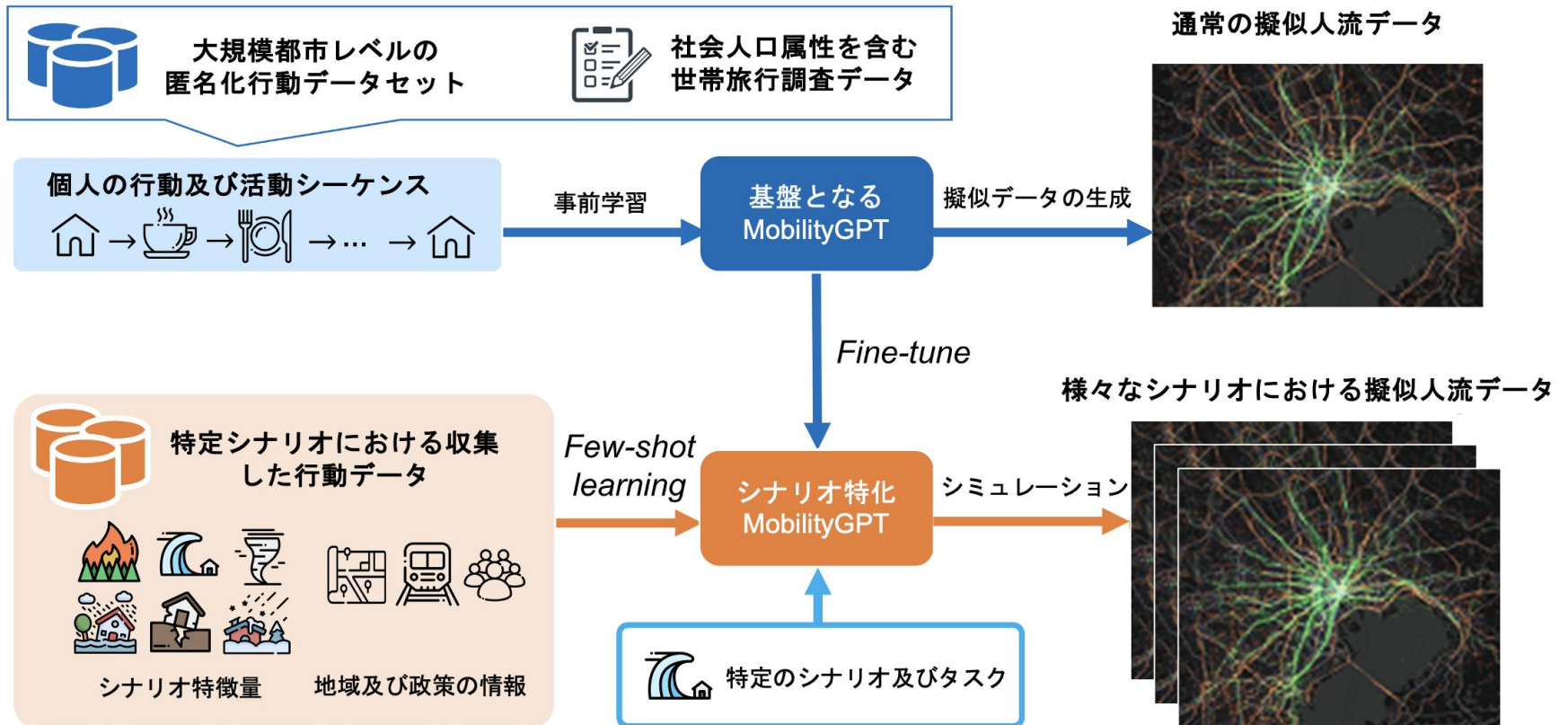
擬似人流x生成系AIについて

- 現段階で、パーソントリップ調査から日常的活動パターンを抽出し、それを用いてGPTモデルの開発に取り組んでいる



擬似人流x生成系AIについて

- a) トリップチェーンをGPTに倣い学習したアクティビティのビッグモデル構築による汎用化
- b) 通常のパーソントリップ調査(交通アンケート調査)では捕捉できないトリップの追加
- c) GPSなどの手段で収集した特定シナリオの行動データを用いて、基盤モデルをファインチューニングし、様々なシナリオにおける擬似人流データを生成



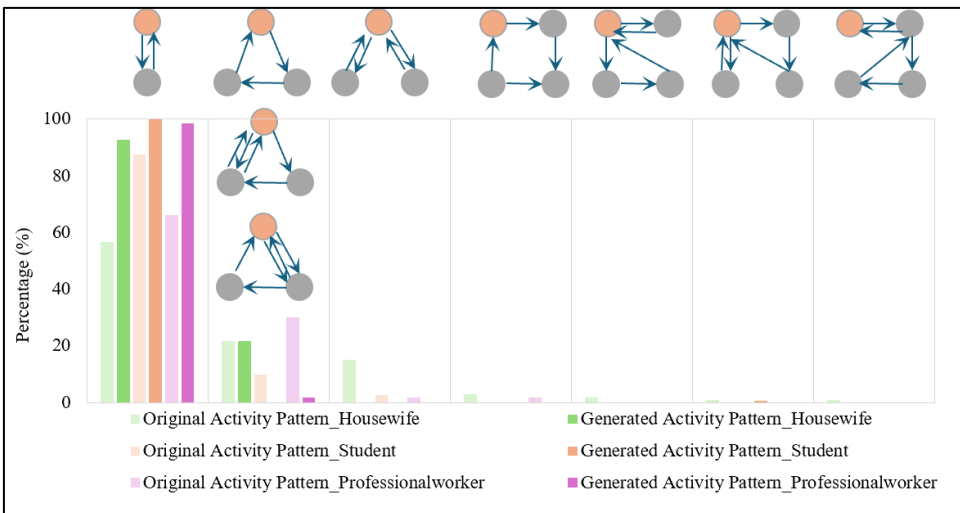
擬似人流x生成系AIについて

日常活動のパターン生成

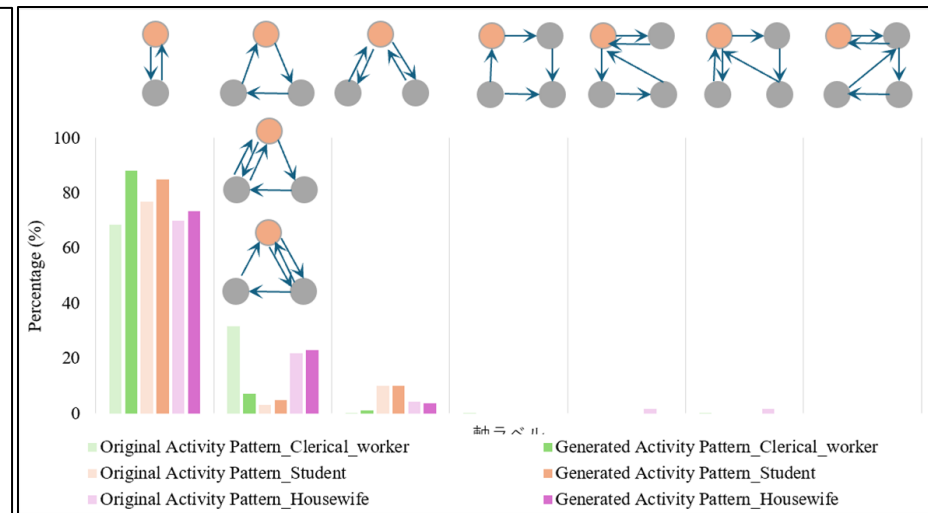
- LLMを用いて、大規模な擬似人流データから日常活動のパターンを生成可能。
- LLMは、活動の連続性や行動パターンの予測に優れ、個々の活動から都市全体の流れを予測する上で効果的

移動目的の推定

- LLMを使ったプロンプトエンジニアリングにより、移動目的の推定精度が向上。
- 少数のサンプルデータ (few-shot learning) でも、通勤・買い物・レジャーなどの目的を高精度で推定可能。



GPT-2による3つのグループでの生成活動パターンと元データの比較

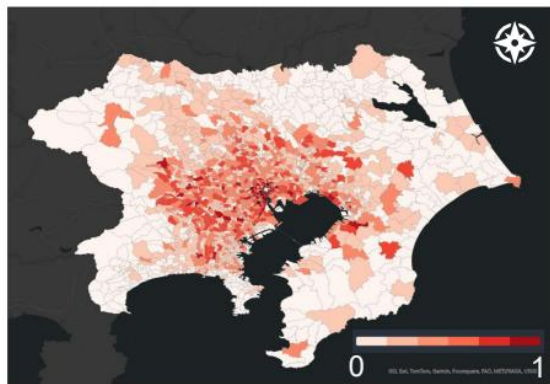


Llama-3による3つのグループでの生成活動パターンと元データの比較

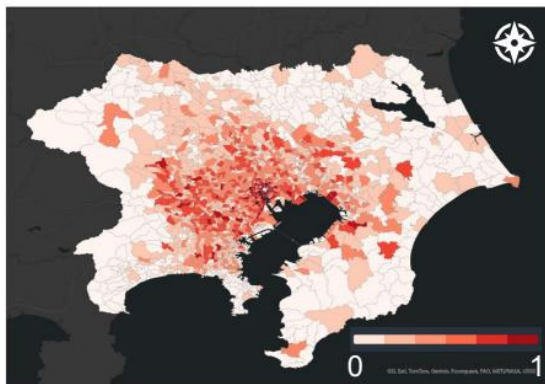
擬似人流×生成系AIについて

時空間データの生成

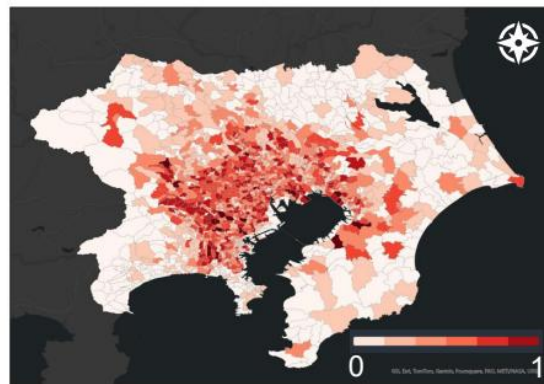
- 活動パターンの生成だけでなく、PTデータを利用して時空間データの生成も可能であることを確認した。
- 各時間帯における活動の分布を可視化し、実際のPTデータとの比較を行った
- 都市部を中心に、活動パターンが実データに近い形で生成されており、時空間的な変化を含んだデータ生成の有効性が確認された



9:00 – 14:00

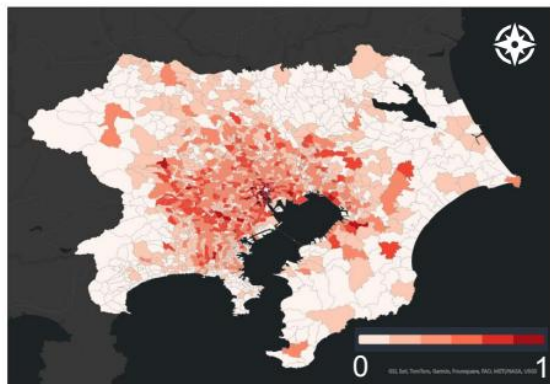


14:00 – 19:00

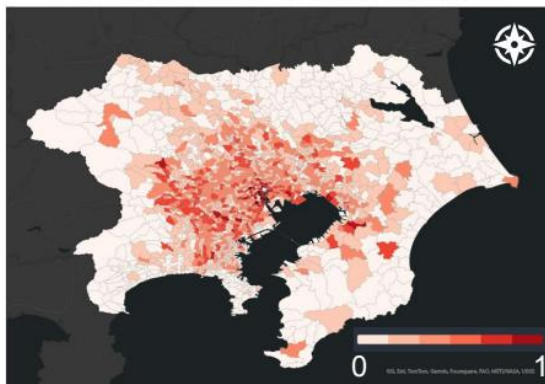


19:00 – 24:00

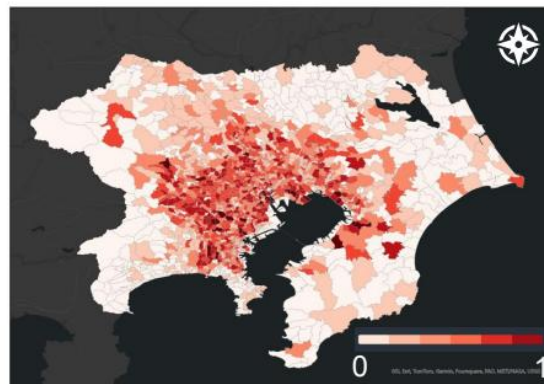
PT Ground Truth of Entire Activities



9:00 – 14:00



14:00 – 19:00



19:00 – 24:00

まとめ

- 擬似人流は全国の作成・提供を始めているが、様々なシミュレーションによる政策のbefore/after検討等、活用は幅広くポテンシャルは大きいと考えている。
- ただし、いくつかの精度ラベルを示していく必要がある。
- また、利活用環境の充実や再計算時間の短縮等も重要。
- また、GPS等、リアルデータとの融合（Few shot learning等）のような商業的価値向上にもポテンシャルが大きい。
- いずれにしても大規模計算となっているので、mdxのような研究クラウド環境が今後とも必須。
- GakuNinRDMは本格活用していないが、JoRASのような他のデータ提供システムなどとID連携ができるようになると使いやすくなる可能性も。