

jh240068

合成人口プロジェクト：合成人口データへの従業地属性の追加

村田忠彦（大阪大学）

概要

本研究課題では、リアルスケール社会シミュレーション(RSSS: Real Scale Social Simulation)の基盤を構築するため、国勢調査をもとに日本全体の市区町村ごとの世帯構成員を含む人口を合成するとともに、合成された人口を RSSS で活用するためのデータベースの作成に取り組んでいる。RSSS とは、実際の規模の世帯構成をもつ人口データを用いた社会シミュレーションである。今年度の研究課題の目的は、すでに合成されている世帯の就業者の居住地における統計を用いて従業地を割り当てるアルゴリズムを日本全国規模に展開し、国勢調査に整合する日本全国の従業者への従業地割当てを実現することである。今年度開発した手法により、就業者の常住地自治体と従業者の従業地自治体双方の統計との整合性を高めることが可能となった。

1. 共同研究に関する情報

(1) 共同利用・共同研究を実施している拠点名

北海道大学 情報基盤センター

東京大学 情報基盤センター

大阪大学 D3 センター

(2) 課題分野

データ科学・データ利活用課題分野

(3) 参加研究者一覧と役割分担

総括・人口合成アルゴリズム

村田忠彦, 原田拓弥, 李 皓, 堀上駿太

人口合成プログラム配備

伊達 進

mdx 環境整備・合成人口 DB 構築・運用

棟朝雅晴, 杉木章義, 塙 敏博

合成人口 DB インタフェース

市川 学, 後藤裕介

2. 研究の目的と意義

本研究課題では、実規模の社会を対象としたリアルスケール社会シミュレーション(RSSS: Real Scale Social Simulation)を実現

するため世帯単位に世帯構成員の属性を合成した人口データの合成を行っている。実際の地域を対象としたリアルスケールな社会シミュレーションを実行するためには、対象となる地域の人口データが必要不可欠となる。しかし、世帯構成員の年齢・性別・従業する産業分類・企業規模・所得などは個人情報であり、研究者が利用することは困難である。そこで、本研究プロジェクトでは、公開されている都道府県、市区町村、町丁目単位の人口統計を用いて、対象とするコミュニティの仮想的な世帯個票データを合成している。公開されている統計に合致する合成人口データは、多様な組合せが考えられるが、多様な組合せの人口データを合成することにより、同じ住居に異なる世帯構成が割り当てられる。同一住居に割り当てられたどの世帯が正しい世帯であるか（もしくは、いずれの世帯も正しくないか）は、その住居の実際の居住者以外にはわからないため、個人情報を秘匿することが可能となる。

2019 年度から JHPCN において、合成人口プロジェクトを開始し、すでに 2000 年、2005

年、2010 年、2015 年の国勢調査に基づく人口合成を完了し、リアルスケール社会シミュレーションの実施を計画する研究者や公的機関に対して、自治体単位でのデータの提供を開始している。

2019 年末に勃発したコロナ禍において、世界各国で感染シミュレーションを行うため、対象とする国や地域の人口合成が行われている。その中で、国単位の合成人口データを研究者向けに利用可能にしているのは、アメリカ、イギリス、ベルギー、日本の 4 カ国のみであり、地域や国のシミュレーションを実施する上で、世界的に重要な技術になりつつある。国が主導して収集している統計として、日本では、人口 20 万人以上の自治体では 1 歳階級の人口統計が、20 万人未満の自治体では 5 歳階級の人口統計が公開されている。さらに 20 万人以上、市部、郡部別に 1 歳階級の夫婦年齢差が、都道府県単位で 1 歳階級の父子年齢差、母子年齢差が公開されている。さらに、町丁目別の 5 歳階級の人口統計、世帯人員別世帯数、家族類型別世帯数、家族類型別人口が公開されている。これらの人口統計を都道府県単位で一括して重ね合わせるにより、都道府県、市区町村、町丁目の特徴を保持しながら仮想的な世帯構成を合成している。日本ではこれらの統計を e-Stat を用いて公開しており、市区町村単位の統計や 5 歳階級別の統計のみから合成している人口データよりも高品質な人口データの合成が可能となっている。

今年度は、合成人口データの従業者に対して従業地を割り当てるアルゴリズムを日本全国規模に展開し、国勢調査に整合する日本全国の従業者への従業地割当てを実現する。

3. 当拠点の公募型共同研究として実施した意義

本プロジェクトは、2017 年と 2018 年に大阪大学サーバーメディアセンター公募型利用制度「若手・女性研究者支援萌芽枠」で開

始し、2019 年度から JHPCN で合成人口に関する共同研究を実施している。研究代表者の村田は、2020 年 4 月から 22 年 3 月に、当時の所属である関西大学の研究拠点形成支援事業により、経済学・社会学・情報学・医療の分野への合成人口データの利活用の研究を推進した。さらに、2020 年 7 月から 2023 年 10 月まで、内閣官房 COVID-19 AI・シミュレーションプロジェクトに合成人口データを提供した。さらに 2020 年 11 月から JST 未来社会創造事業探索研究で合成人口データを用いた研究を行った。2021 年 12 月には進化計算学会において社会シミュレーションの最適化コンペティションを実施し、査読論文として成果を報告した[1-5]。JST 未来社会創造事業は、2023 年 4 月から本格研究に採択され、自治体と共同して取り組むプロジェクトを展開している。

本プロジェクトにより合成された人口データは、これまで本プロジェクトの共同研究者が所属する岩手県立大学、芝浦工業大学、静岡大学、大阪大学以外に、北海道大学、筑波大学、東京大学、東京工業大学、京都大学、奈良先端科学技術大学院大学、神戸大学、富山県立大学、滋賀県立大学、東京理科大学、早稲田大学、慶應大学、青山学院大学、国際医療福祉大学、創価大学、関西大学、国立保健医療科学院、国立情報学研究所、防災科学技術研究所、聖路加国際病院から問合せをうけ、合成人口データの配布を行なってきた。今年度はさらに、東北大学、東京女子大学、中央大学、大阪工業大学、大阪医科薬科大学の研究者にも提供した。それらの研究機関では、科学研究費基盤研究 A、同基盤研究 B、同基盤研究 C、JST 未来社会創造事業探索研究、同本格研究などの研究プロジェクトが展開されており、様々な規模の研究を支援できる体制が整っている。

本年も本プロジェクトが提供した合成人口データを用いた多数の研究が実施され、計



都道府県 ID:27 大阪		市区町村 ID:27207 高槻		町丁目 ID:27207209002 霊仙寺町 2 丁目	
緯度 34.876799			経度 135.575662		
世帯 ID:5		世帯類型 ID:4 女親と子供世帯		世帯人員数 2	
個人 ID:7	年齢 38	性別 ID:1 女性	役割 ID:21 子供 (女性)		
所得 (月額) 300,000 円			産業分類 ID:90 I 卸売業・小売業		
雇用形態 ID:10 一般労働者			企業規模 ID:1000 1000 人以上		
個人 ID:8	年齢 76	性別 ID:1 女性	役割 ID:11 妻・女親		
所得 (月額) 0 円			産業分類 ID:N/A N/A		
雇用形態 ID:N/A 非就業者			企業規模 ID:N/A N/A		

出典：国土交通省
基盤地図情報

図 1：合成人口データの世帯情報（大阪府高槻市霊山寺町のサンプル）

表 1：データ保護レベル

データ 保護レベル	粒度	就業状況と所得※1	対象※2
1	都道府県	×	学部生
2	市区町村	×	学部生
3	都道府県	○	院生
4	市区町村※3	○	院生
5	町丁目	×	院生
6	位置情報	×	研究者
7	町丁目	○	研究者
8	位置情報	○	研究者

※1 研究内容により就業状況のみのデータ提供の場合がある。

※2 研究内容により、高いレベルのデータ提供も考慮する。

※3 人口 1,000 人未満の市区町村を除く。

立案のための意思決定支援）、**優秀賞**（柴山りく，後藤裕介，販売記録データと社会調査データの融合による顧客像に基づく商圈分析方法の開発）、**学生賞**（宮武俊作，村田忠彦，世帯の連続性を考慮した合成人口補間手法）を受賞した論文は，いずれも合成人口データを用いた研究であり，リアルスケール社会シミュレーションを実施する研究者コミュニティにおいて，合成人口データを用いることが標準となってきた。

Whole Japan Synthesized Population by HPCI-JHPCN

Source:
Dataset: 2015

Target area:
Prefecture: 東京都 City: 千代田区 Town: All

Primary data:
☐ Building ☐ Household ☐ Person ☒ Worker

Format:
☒ JSON ☐ CSV

Data count (generated): 25,861

Downloaded file includes 10 sets of Synthesized Population of the specified area. Please implement your data analysis or simulation at least ten times using these data sets.



図 2：合成人口ダウンロード用インターフェース

測自動制御学会社会システム部会で発表された 42 件の論文のうち，**最優秀賞**（神部尚樹，貝原俊也，國領大介，マルチスケール社会シミュレーションを用いた健幸都市政策

4. 前年度までに得られた研究成果の概要

本プロジェクトは 2019 年度から継続して採択されている。2019 年度は，大阪大学の計算機を用いて，図 1 のような日本全国の個票データを国勢調査に基づいて合成した。図 1 は大阪府高槻市にある建築物に居住する世帯の一例である。本プロジェクトでは，表 1 のデータ保護レベルにしたがって，図 1 に色分けされた属性を含めたデータの提供を行っている。さらに，北海道大学のインタラクティブシステムを用いて，図 2 のように地図上で自治体を指定して世帯や就業者の情報をダウンロードできるシステムを構築した。北海道大学で継続的にデータベースを維持すると共に 2020 年 2 月から HPCI 共用ストレージにてバックアップを行っている。

2020 年度は以下の研究により，合成人口データの精度を向上した。

表 2：割当てる属性と使用する統計表の出典

割当てる属性	使用する統計名	公開されている対象	表題/境界名称	表
従業地（市区町村）	平成27年国勢調査 / 従業地・通学地による人口・就業状態等集計（人口、就業者の産業（大分類）・職業（大分類）など）	21大都市（東京都都区部と政令指定都市）とその区、県庁所在市、人口20万以上の市	常住地による従業市区町村、産業（大分類）別15歳以上就業者数	A
		全市区町村	常住地又は従業地（9区分）による雇用者（3区分）、産業（大分類）、男女別15歳以上就業者数	B
			常住地による従業・通学市区町村、男女別15歳以上就業者数及び15歳以上通学者数（15歳未満通学者を含む通学者一特掲）	C
従業地（小地域）	平成26年経済センサス-基礎調査 / 町丁・大字別集計	全市区町村	経営組織（2区分）、産業（中分類）・従業者規模（6区分）別全事業所数及び男女別従業者数-市区町村、町丁・大字	D
従業地（小地域の図形中心点座標）			平成26年経済センサス-基礎調査 町丁・大字別境界データ	—

表 3：従業地における従業者の統計表

使用する統計名	公開されている対象	表題/境界名称	統計表
平成 27 年国勢調査/ 従業地・通学地による人口・就業状態等集計	21 大都市（東京都特別区部および政令指定都市）とその各区、県庁所在市、人口 20 万以上の市	従業地による常住市区町村、産業（大分類）別 15 歳以上就業者数	E
	全市区町村	常住地又は従業地（9 区分）による雇用者（3 区分）、産業（大分類）、男女別 15 歳以上就業者数	B

- ・合成世帯の建物へのマッピング
- ・施設世帯を含めた人口合成
- ・人口動態に関する研究
- ・シミュレーションプラットフォーム開発
- ・救急医療の環境整備に関する研究
- ・新型コロナウイルス感染症に関する研究
- ・地震時の避難行動に関する研究
- ・ベーシックインカムに関する研究

2021 年度は、次の 2 つの課題に取り組んだ。

- ・就業者の従業地の割当てに関する研究
- ・建造物の利用用途推定に関する研究

従業地割当てアルゴリズムの研究は 2021 年度計測自動制御学会第 27 回社会システム部会優秀賞を受賞し、2022 年度には IEEE Transactions on Computational Social Systems に採択され、AAAS Eureka Alert!や、芝浦工業大学、関西大学からプレスリリースされた。まず、国勢調査から得られる常住地による従業市区町村別就業者数の分布比率に基づいて各従業地（市区町村）に割当てる人数を決定する。表 2 の表 A より、就業者の常住地（市区）別、産業分類別、従業地（市区町村）別の構成比が得られ、これに比例した人数の合成人口データを無作為に抽出することで従業地（市区町村）を割当てる。これが公開されていない市町村においては、表

B および表 C を用いて段階的に割当てる手法を提案した。次に、表 D を用いて、割当てた従業地（市区町村）において、経済センサス-基礎調査を用いてさらに詳細な小地域への割当てを行った。

建造物の利用用途推定については、空中写真から住宅の利用用途を判別するアルゴリズムを用いて、合成人口データの適切な割り当てを試みた。首都圏と中部圏のデータを用いて U-Net を訓練し、近畿圏のデータで評価したところ、近畿圏の建物（面積 25 m² 以上）の内、82.8 %の用途を正しく判別できた。

2022 年度には、mdx 上で合成人口 DB インタフェースを構築するため、608vCPU で仮想ディスク 100GB の環境を確保し、北海道大学で運用している合成人口データをコンテナ化することにより、mdx の Kubernetes 上に移動させ、保護レベル別合成人口データベースを構築した。

2023 年度には、2020 年度に実施された国勢調査に基づく合成人口データの合成に取り組んだ。2020 年調査結果のうち、人口合成に必要な統計表でもっとも遅いものは 2023 年 3 月に公開される。2020 年度から一部の統計の公開フォーマットが変更されたため、2023 年度での合成には至らなかったが、

JHPCN 事務局の支援により、2024 年 3 月から総務省の連携先の検討を行い、東京大学空間情報科学研究センター（CSIS）の仲介で、統計センター情報システム部の担当者との意見交換を開始することができている。まだ、2020 年度の人口合成には至っていないが、公的統計の公表方法に関する継続性の観点での議論を進めている。

上記のような経緯で、2019 年度より合成人口プロジェクトを推進している。2024 年度は、2021 年度に実施した常住地単位の従業地属性合成手法を、表 3 のような従業地ごとの統計にあわせて調整することにより、常住地と従業地の双方の統計に合致する従業地割り当て手法の開発に取り組んだ。

5. 今年度の研究成果の詳細

2024 年度は、大阪府高槻市を対象にして、市外から高槻市に流入する従業者の従業地割当てを、従業地側である高槻市の従業地統計を用いて実施した。これは、論文（Murata et al., Workplace Assignment to Workers in Synthetic Populations in Japan, IEEE Trans. Computational Social Systems, 10 (4), 1914/1923）で提案した常住地に在住する従業者に対して市内外の従業地を割当てするアルゴリズムを、従業地自治体に流入する従業者が常住地とする多数の自治体に対して実施した上で、従業地側の統計に合致するようにアルゴリズムを構成するものである。

具体的には、従業地を高槻市として実施し、表 3 のような高槻市を従業地とする統計を用いて調整した。国勢調査「従業地による常住市区町村、産業（大分類）別 15 歳以上就業者数」により、高槻市を従業地とする従業者は 109,557 人存在し、それらの従業者の常住地自治体数は、高槻市を含めて、417 自治体である。なお、高槻市を常住地とする従業者数は全従業者の 59.8%（65,506 人）だった。

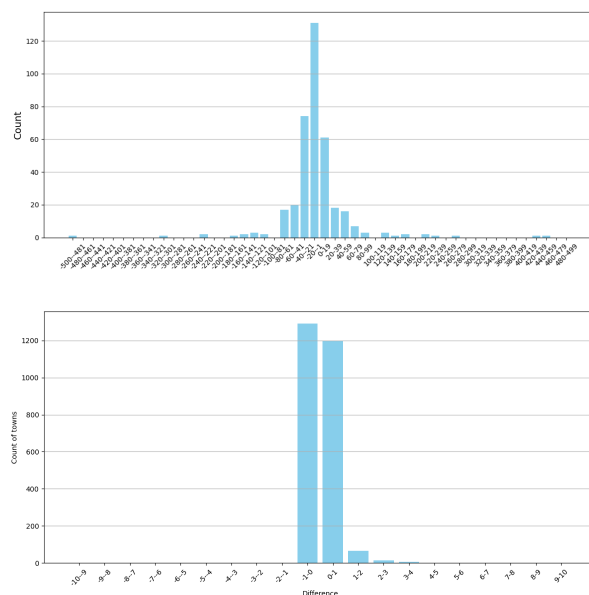


図 3：従業者数の一致度別小地域数

90%の従業者が常住する高槻市以外の自治体数を計上したところ、417 自治体のうち、70 自治体に常住していることがわかった。これらの 70 自治体に対して、前述の従業地割当てアルゴリズムを適用して、就業者への従業地割当てを行い、高槻市における小地域単位の従業者数と、従業地割当てアルゴリズムにより割り当てた従業者数の差を求めた。結果を図 1 上に示す。横軸は従業者数の一致度（統計での従業者数とのズレ幅）で、縦軸は一致度幅の小地域数である。図 1 上では、20 人単位の小地域数を示している。真ん中の最も高い小地域数のあたりがズレなしの小地域数となる。

従業地割当てアルゴリズムでは、小地域（町丁目）単位の従業地の割り当てを確率的に行っているため、一部の小地域に関しては、過少または過大な従業者が確率的に割り当てられる場合があることがわかった。そこで、過多に割り当てられた小地域からランダムに従業者を選び、確率的に再割り当てを行う処理を施したところ、図 3 下のように誤差の少ない小地域にすることが可能となった（図 3 下は 1 人単位の小地域数となっている）。今回の試行では従業者数が 10%ほど少ないた

め、統計値よりもやや少ない従業者数が割り当てられることになっており、修正アルゴリズムにより、適切に従業者の割り当てが行えることがわかった。

図 4 に高槻市内に割り当てられた産業分類ごとの就業者を示す。従来手法ではある地域に存在しない産業分類に従事する従業者の割り当てが確率的に行われていたが、そのような割り当てを回避するアルゴリズムを導入したことにより、地域全体における産業分類ごとの従業者数もやや改善できたことがわかる。

各自治体の従業者の従業地割当ては、それぞれ独立に行えるため、分散アルゴリズムとなるが、各自治体の従業者数により計算量が異なる。そこで、 n ノード使用可能である場合に、各自治体の従業者数を $p_1, p_2, \dots, p_{1718}$ とすると、各ノードに割り当てるべき従業者数は、 $(p_1 + p_2 + \dots + p_{1718})/n$ 人となる。各自治体を従業者数が多い順にソートし、もっとも従業者数の多い自治体の従業者数を $p_{[1]}$ 、最も少ない従業者数を $p_{[1718]}$ とし、 $p_{[1]}$ の自治体から順に $(p_1 + p_2 + \dots + p_{1718})/n$ 人を超えない範囲で各ノードに割り当てることで、ノードごとの計算量を平滑化するスケジューリングプログラムを現在構築中である。

6. 進捗状況の自己評価と今後の展望

上記のように、従業地単位の統計を用いた従業地割当てアルゴリズムを完成させることができた。上記の試行では、従業数の 90% にあたる自治体からの流入従業者の従業地割当てを行っているが、これを日本国内の全 1718 自治体に対して実施することで、自治体ごとに 100% の流入元自治体での従業地割り当てが可能になる。

上記の従業地割当てプログラムを実施後、今年度提案したアルゴリズムにより、従業地の調整を行うことにより、従来手法よりも統計に合致した従業地属性合成が可能となる。

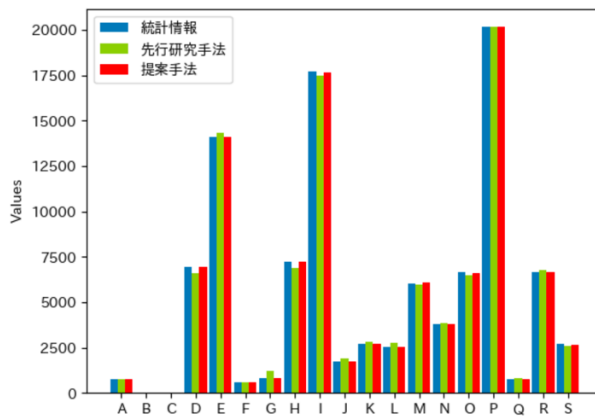


図 4：産業分類ごとの高槻市内の従業者数

今年度、可能であれば、全国の割り当てまで完了させたかったが、従業地割当ての分散プログラムの構築と、各自治体内での従業地の調整アルゴリズムの構築に手間取ってしまった。今年度達成したかった本質的なアルゴリズムの開発を行うことができたため、70%程度の進捗であるといえる。