

jh190056-MDH

リアルスケール社会シミュレーションのための人口合成とその応用

村田忠彦（関西大学）

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構築を目指す。RSSS とは、実際の規模の世帯構成をもつ人口を用いた社会シミュレーションである。実規模の人口を対象としたシミュレーションを実施することにより、シミュレーション結果のより現実的な解釈を可能とすることが目的である。本年度は、大阪大学の計算機を用いて、現実と同じ規模の世帯構成をもつ人口を合成し、北海道大学で合成人口データベースの構築に取り組んだ。現在、国レベルの合成人口は、米国と英国で公開されているのみであり、アジア初の RSSS 基盤として、日本の合成人口を公開する意義は大きく、実規模の社会シミュレーションの研究コミュニティの活性化が期待できる。

1. 共同研究に関する情報

(1) 共同研究を実施した拠点名

大阪大学サイバーメディアセンター
北海道大学情報基盤センター

(2) 共同研究分野

■ 超大規模数値計算系応用分野

■ 超大規模データ処理系応用分野

(3) 参加研究者の役割分担

村田忠彦（関西大学）	研究統括
原田拓弥（青山学院大学）	人口合成
李 皓（静岡大学）	人口合成
伊達 進（大阪大学）	プログラム高速化
市川 学（芝浦工業大学）	DB インタフェース
後藤裕介（岩手県立大学）	DB インタフェース
棟朝雅晴（北海道大学）	データベース構築
杉木章義（北海道大学）	データベース構築

2. 研究の目的と意義

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構築を目指す。RSSS とは、実際の

規模の世帯構成をもつ人口を用いた社会シミュレーションである。合成人口を用いることにより、実規模の人口を対象としたシミュレーションが可能となる。RSSS により、従来の仮想空間における社会シミュレーションと異なり、シミュレーション結果のより現実的な解釈が容易となる。

RSSS においても、個々の市民の意思決定モデルを構築することが重要となるが、様々な研究者が仮想空間において構築し、現実に観察されている現象と同様の現象を、構築したモデルを用いたシミュレーションで観察することにより、モデルの妥当性を確認してきた。とはいえ、そこで同定されるパラメータの値はあくまでも仮想空間における値であり、現実社会ではどのようなパラメータ値になるかが検討できなかった。そこで、本研究課題では、仮想空間におけるシミュレーションを、現実社会に適用するためにまず必要となる対象地域の世帯構成を統計情報から構築する人口合成手法により、日本全国の市区町村別の人口を合成する研究課題を構想した。

研究代表者の村田は、合成人口の研究に 2013 年から取り組んでおり、2017 年に執筆

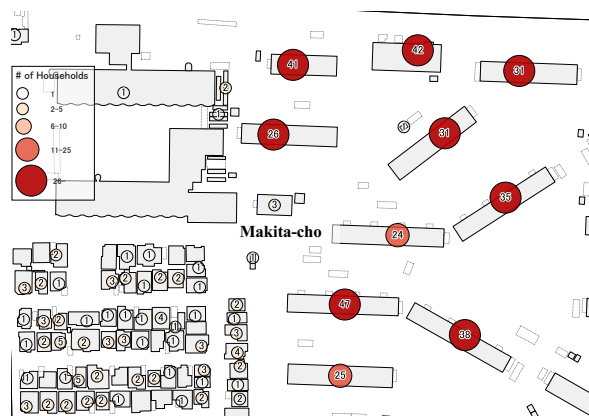


図 1：合成世帯の地図へのマッピング結果。
円内の数字は対応する建物に居住する世帯数。
(地図は国土地理院の基盤地図情報により作成)

した合成した人口を地図上の建物矩形に割り当てる論文[1, 5]で 2019 年度計測自動制御学会論文賞を受賞した。本研究課題では、その技術を利用し、日本全国の合成人口世帯を図 1 のように地図上の建物への割り当てを行なった。

現在、国レベルの合成人口は、米国と英国で公開されているのみであり、アジア初の RSSS 基盤として、日本の合成人口を公開する意義は大きく、実規模の社会シミュレーションの研究コミュニティの活性化が期待できる。なお、米国や英国の合成人口は、各地域における世帯情報として提供されているのみであり、本研究課題で実施しているような建物矩形への世帯の割り当てを行なっているわけではない。また、合成人口の合成においては、統計的にデータに整合する様々な世帯の合成が考えられることから、合成データは一意に定まるものではないが、本研究課題の拠点のみが、希望する研究者に複数セットの合成人口を提供している。このように、本研究課題の合成人口は、世界的に見ても RSSS を実施するための貴重なデータであり、現実的な社会シミュレーションに取り組む研究者に必要なデータを供給することにつながる。

3. 当拠点公募型研究として実施した意義

本年度は、大阪大学の計算機を用いて、現実と同じ規模の世帯構成をもつ人口を合成し、北海道大学が提供するアカデミッククラウドを用いて、合成人口データベースの構築に取り組んだ。合成人口データの研究者向けの提供を行うために、北海道大学のハイパフォーマンスインタークラウドを用いて公開できる体制を整えている。

本研究課題は、RSSS のための日本全国の合成人口データベースを構築するため、人口の合成とデータベースを連動させる課題である。実社会を対象とした社会シミュレーションを行おうとする研究者が、自分のモデルを研究対象の自治体に適用するだけであれば、個々の研究者が自ら人口を合成すればよい。しかし、その場合、研究者によって、合成する際に考慮する統計にばらつきがでるため、他の研究者による追試が容易に行える環境となるわけではない。個々の研究者が個別に人口合成を行う現状は、(1) RSSS に関する研究の精度が高まらないばかりか、(2) 研究者コミュニティを充実させることにもつながらない。そこで本研究課題での取り組みを通して、研究代表者が開発した人口合成手法を用いて、日本全国の人口合成を行い、希望する研究者が活用できるようにデータを提供する環境を整備している。

また、現在、2000 年、2005 年、2010 年、2015 年の国勢調査に基づく人口合成を行なっているが、国勢調査は 5 年に 1 度予定されており、経年的にデータが拡充していくことが見込まれる。本拠点公募型研究として採択されたことにより、HPCI 共用ストレージに応募することができ、2020 年 2 月 1 日よりストレージの利用が可能となった[21]。このようなシステムを活用することにより、過去の年度に遡って、シミュレーションを実施できる環境が整っていることから、RSSS における今後の研究の活性化が期待できる。



※非就業者の場合 所得が0円で他の就業属性が全てN/A

図2：合成人口における世帯情報の例

4. 前年度までに得られた研究成果の概要

継続課題ではない。

5. 今年度の研究成果の詳細

本年度は、研究者向けのデータ提供システムを構築するにあたり、まず合成人口を提供するためのルールを定めた[2-4, 7]。図2に本研究課題により合成される世帯情報の例を示す。これは、研究代表者の所属する関西大学高槻キャンパス内に世帯があると仮定した場合の世帯の例である。76歳の女親と38歳の娘の2名の世帯で、娘は、卸売業・小売業の業種に勤めており、その所得は月額300,000円となっている。これらは、公開されている統計から計算される世帯の情報であり、実在する世帯の個人情報は一切用いていないため、特定の世帯のプライバシーを侵害することはない。また、本研究課題では、必ず複数のセットの人口世帯を提供することにより、着目する地域や建物にひもづく世帯の特徴が一意に提示されないようにしている。

一方で、合成人口データの複数セットを提供することにより、計算により求められた一

表1：セキュリティレベル表

セキュリティレベル	粒度	所得	対象※
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	○	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

※研究内容によっては、より高いレベルのデータ提供も考慮する。

定の幅で、年齢・業種・所得の特徴をもつ合成世帯がデータベース閲覧者の着目する地域や建物に存在する可能性を示すことになるともいえる。したがって、商業利用を考えた場合、広告配布計画や出店計画、商圈分析に容易に活用できるデータとなる。合成人口データはすべて公開されている統計情報から合成されているため、商業利用を計画する事業者にとっても、合成可能なデータであるが、本研究課題が提供するデータにより、対

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

Download Reset

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3：合成人口データダウンロード用インターフェース

象地域を特定する労力を大幅に軽減できることになる。

本研究課題の成果物の利用に関して、個人情報保護に関する法律（平成十五年法律第五十七号）と照らし合わせても、上述の商業利用は、公開されている統計情報から個人を特定しない範囲での世帯情報の合成となっていることから、認容できる範囲であると考えられる。しかし、昨今の個人情報保護に対する国民の意識の高まりに配慮し、本研究課題で合成される合成人口データの使用目的を、公的目的に限り、無償で提供することとした。すなわち、学術機関における教育研究目的や自治体や公益法人などにおける政策・事業立案の目的に限って合成人口データを提供することとした。

なお、学術機関においては、学生の利用も

考えられるが、提供対象者に応じて、表1のようなセキュリティレベルを設けて、提供する合成人口データの内容や粒度を定めた。セキュリティレベルが高くなるほど、粒度が細かく、所得データのついた合成人口データを提供できることとした。

上述のようなセキュリティポリシーのもと、申請内容を許可された対象者にのみ、利用を許可するデータベースを構築した。データベースから、図3のようなインターフェースにより、年度、都道府県、市町村、町丁目を指定すると、該当の地図が表示され、地域を確認して、データのダウンロードができるようになっている。

本年度は、研究代表者が JHPCN 課題研究者との共同研究を推進するため、共同利用・共同研究拠点 関西大学ソシオネットワーク

ク研究機構の公募研究課題に応募し、研究課題「合成人口データの配布方法と活用に関する研究」[19]により、本研究課題を推進するための予算を獲得した。2019年12月26日、27日に北海道大学情報基盤センターにて研究会を開催し、人口合成に使用する統計と、インタフェースに関する確認を行なった。

2019年11月に千葉大学で実施された計測自動制御学会システム・情報部門学術講演会（SSI2019）では、原田・村田は、人口合成手法をより正確におこなうための手法[9]を提案し、SSI 2019 優秀発表賞を受賞した。また、同学術講演会において、李は静岡県浜松市を対象として、本研究課題で提供している合成人口の動態推定を行う手法[8]を、原田は合成人口の世帯を割り付ける対象となる建物の用途を、深層学習手法を用いて同定する手法[10]を提案した。

また、社会シミュレーションに関する研究発表として、計測自動制御学会第22回社会システム部会研究会（2020年3月15-17日）にて研究発表を行った[11-16]。中でも、岩手県大船渡市、陸前高田市、住田町、釜石市、大槌町の合成人口を用いて行われた災害復興における経済支援政策としてのキャッシュネットワークの有効性に関する後藤の研究[12]は、計測自動制御学会第22回社会システム部会研究会優秀賞を受賞している。

国際的な活動として、研究代表者の村田は、2019年10月に、University College Dublin（アイルランド）の計算社会学研究科のPh.Dセミナーにおいて、研究科の研究者と大学院生を対象に、合成人口を用いた社会シミュレーションのレクチャーを行った[20]。また、2019年11月にDenverで開催されたSupercomputing Conference（SC19）の関西大学展示ブースにおいて、合成人口の提供ルールの研究[4]に加えて、合成人口の各世帯の所得の合成手法[5]と世帯の建物への割付手法[6]に関するポスター展示を行い、国内

外の研究者に合成人口データのデータベース構築に関する情報提供を行った。

さらに、本研究課題の推進を広告するため、大阪大学サイバーメディアセンターはHigh Performance Scientific Computing NewsのVol.6として、「リアルスケール社会シミュレーション」というタイトルのビデオ[18]を作成し、2019年7月1日付で、大阪大学サイバーメディアセンターのホームページで公開している。さらに北海道大学情報基盤センターは、刊行している北海道大学情報基盤センター大型計算機システムニュース Vol.50で、「リアルスケール人工社会で展開するデータサイエンスとシミュレーションの世界」と題するインタビュー記事[22]を作成し、合成人口データの紹介を行なっている。

6. 今年度の進捗状況と今後の展望

本年度は、大阪大学の計算機を用いて、表2のような2000年から2015年の4カ年の国勢調査を用いた人口合成を行う予定であった。町丁目レベルの100セットの人口合成が全ての年度に対して達成できている。建築物カバー率とは、図1のように合成人口の各世帯を地図上の建築物に割り当てることのできた町丁目の割合である。研究代表者らが用いた手法[1]では、国土地理院が提供する2015年の基盤地図情報を用いたが、2014年7月31日以前の地図については、文献[1]の手法を適用可能な地図が全ての地域に対して得られず、2010年の建築物の割り当ては一部に留まった。なお、2015年のカバー率が100%になっていないのは、建物が含まれていない町丁目が存在していたためである。

表2に示したように1カ年1試行のデー

表2：合成人口の合成状況

年度	2000	2005	2010	2015
市区町村	完了	完了	完了	完了
町丁目	完了	完了	完了	完了
建築物カバー率	地図 X	地図 X	42.6%	99.7%
ファイルサイズ	2.5GB	5.4GB	7.4GB	10.4GB

タは平均で約 5GB~10GB のデータ量となっており、データベース用でのファイルサイズは約 3 倍の 15~25GB (2015 年度については、建築物データを含めて 25GB 程度) になった。4 年分、100 試行では、約 8~10TB のデータサイズとなっている。

また、合成手法を改良する場合、他の研究者による追試環境を維持するため、以前のバージョンの人口を残しながら、新たな合成手法を用いたデータを公開する必要がある。合成手法を改良するとともに、1 年分 100 試行あたり 2.5TB のデータサイズが必要となる。すでに、2018 年に開発した改良手法を用いた人口合成が 2020 年度の JHPCN 研究課題として採択[24]されており、データの増加のために備えておく必要がある。さらに、2020 年の国勢調査により新たな統計が公開された場合にも、データの増加が見込まれる。合成手法の改良、統計データの公開のたびに、データ量が增大することが明らかである。そこで、HPCI 共用ストレージを申請し、採択された[21]。

上述のように、本研究課題の研究者の様々な研究活動や、大阪大学の研究紹介ビデオ、北海道大学のインタビュー記事の効果、データベースの利用者向けのセキュリティポリシーの明確化、本研究課題によるデータベースの整備を行うことにより、本研究課題の共同研究者に加えて、北海道大学、芝浦工業大学、筑波大学、東京工業大学、聖路加国際病院の研究者からの問い合わせを受けており、合成人口データの提供を開始している。

また、社会シミュレーション以外のデータ分析等に合成人口を利用する研究プロジェクトを推進するため、研究代表者は、2020 年度関西大学研究拠点形成支援経費 (2020 年 4 月~2022 年 3 月) に申請し、採択されている[23]。本研究課題代表者である村田を研究代表者とし、中元康裕 (経済学・関西大学)、中井豊 (社会学・芝浦工業大学)、松下光範 (情報学・関西大学)、佐々木美絵 (医療・兵庫県立大学)

と共に合成人口を利用した研究を推進する。さらに松本渉 (人口統計学・関西大学)、原田拓弥 (人口合成・青山学院大学) と共に、合成人口の信頼性を高める研究を並行して推進する。すでにこれらの研究者と各研究に必要な合成人口データの調整を行っている。

さらに、共同研究者の原田は 2019 年 8 月から、研究代表者の村田は 2020 年 4 月から合成人口に関係した科学研究費[17, 25]を獲得している。また、研究代表者の村田と原田は、年金制度、雇用制度に関する研究プロジェクトに参画し、合成人口データを活用したシミュレーションを行う予定である[26]。

加えて、共同研究者の市川は、内閣府が推進する戦略的イノベーション創造プログラム (SIP2) のプロジェクトに参画しており、災害復旧に関係するプロジェクトにおける合成人口データの利活用を計画している。

このように、本研究課題の成果である合成人口データを活用した新規研究開発プロジェクトが動き出しており、それぞれのプロジェクトの今後の研究成果が期待される。

7. 研究業績一覧 (発表予定も含む)

(1) 学術論文 (査読あり)

[1] (2019 年度計測自動制御学会論文賞)
Takuya Harada, Tadahiko Murata,
“Projecting Households of Synthetic Population on Buildings Using Fundamental Geospatial Data,” *SICE Journal of Control, Measurement, and System Integration*, Vol. 10, No. 6, pp. 505–512, 2017.

(2) 国際会議プロシーディングス (査読あり)

[2] Tadahiko Murata, Takuya Harada,
“Synthetic and Distribution Method of Japanese Synthesized Population for Real-Scale Social Simulations,” *Proc. of the 33rd Annual Conference*

of the Japanese Society for Artificial Intelligence 2019, 3 pages, 2019.

[3] Tadahiko Murata, Takuya Harada, Manabu Ichikawa, Yusuke Goto, Lee Hao, Susumu Date, Masaharu Munetomo, Akiyoshi Sugiki, “Distribution of Synthetic Populations of Japan for Social Scientists and Social Simulation Researchers,” Proc. of International Conference on Machine Learning and Cybernetics: ICMLC 2019 (Kobe, Japan, 2019/7/7-10), 4 pages, 2019.

(3) 国際会議発表 (査読なし)

[4] (ポスター展示) Tadahiko Murata, Manabu Ichikawa, Yusuke Goto, Lee Hao, Takuya Harada, Susumu Date, Masaharu Munetomo, Akiyoshi Sugiki, “Synthetic Population for Real-Scale Social Simulation,” IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis: SC 19 (Booth 1976, Denver, USA, 2019/11/18-21).

[5] (ポスター展示) Takuya Harada, Tadahiko Murata, “Projecting Synthetic Household on Buildings in Maps for Real-Scale Social Simulations,” IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis: SC 19 (Booth 1976, Denver, USA, 2019/11/18-21).

[6] (ポスター展示) Sho Sugiura, Tadahiko Murata, Takuya Harada, “Income Estimation for Workers in Synthetic Population for Real-Scale Social Simulations,” IEEE/ACM International Conference for High

Performance Computing, Networking, Storage and Analysis: SC 19 (Booth 1976, Denver, USA, 2019/11/18-21).

(4) 国内会議発表 (査読なし)

[7] 村田忠彦, 原田拓弥, 日本の全人口合成データ配布システム, 第 35 回ファジィシステムシンポジウム (大阪, 2019/8/29-31), 4 pages, 2019.

[8] 岩田恵人, 李皓, 社会動態推定のための初期仮想個票データベースの構築, 計測自動制御学会システム・情報部門学術講演会 2019 (千葉, 2019/11/23-25), pp. 28-29.

[9] (2019 年度計測自動制御学会システム・情報部門学術講演会優秀発表賞) 原田拓弥, 村田忠彦, 出生コーホートによる出生数を考慮した人口個票の合成, 計測自動制御学会システム・情報部門学術講演会 2019 (千葉, 2019/11/23-25), pp. 554-559.

[10] 田口栞, 原田拓弥, 大内紀知, 現実に近い社会シミュレーション実施のための pix2pix を用いた建造物の用途の判別の検討, 計測自動制御学会システム・情報部門学術講演会 2019 (千葉, 2019/11/23-25), pp. 566-567.

[11] 原田拓弥, 高橋真吾, 村田忠彦, 仮想都市の統計情報を用いて合成した合成人口の評価, 計測自動制御学会第 22 回社会システム部会研究会資料 (沖縄, 2020/3/15-17), pp. 15-21.

[12] (計測自動制御学会第 22 回社会システム部会研究会優秀賞) 阿部蕉太, 後藤裕介, 震災復興過程における地域特性を踏まえた有効な CFW のシミュレーション分析, 計測自動制御学会第 22 回社会システム部会研究会資料 (沖縄, 2020/3/15-17), pp. 22-29.

[13] 岩田恵人, 輿野木龍, 李皓, 掛川市の年少人口実績データを用いたマイ

- クロシミュレーションモデルの検証,
計測自動制御学会第 22 回社会システム部会研究会資料 (沖縄, 2020/3/15-17), pp. 79-84.
- [14] 塚窪孝裕, 市川学, 大澤絵里, 越智真奈美, 峰友紗, 0~3 歳児を持つ親のニーズの可視化に関する研究, 計測自動制御学会第 22 回社会システム部会研究会資料 (沖縄, 2020/3/15-17), pp. 85-97.
- [15] 漆原史也, 吉田壮一郎, 岩崎翔吾, 柳沢隆己, 小森健一郎, 市川学, 中井豊, 災害シミュレータによるインフラ被害の可視化とその活用, 計測自動制御学会第 22 回社会システム部会研究会資料 (沖縄, 2020/3/15-17), pp. 98-105.
- [16] 田口栞, 原田拓弥, 大内紀知, 現実に近い社会シミュレーション実施のための pix2pix を用いた建造物の用途の判別, 計測自動制御学会第 22 回社会システム部会研究会資料 (沖縄, 2020/3/15-17), pp. 124-128.
- (5) その他 (特許, プレスリリース, 著書等)
- [17] (科学研究費) 原田拓弥, 全国規模エージェントシミュレーションの為に街区毎の特徴を保持する個人属性の合成, 科研研究活動スタート支援 (2019/8/30~2021/3/31) .
- [18] (ビデオ) 村田忠彦, 原田拓弥, リアルスケール社会シミュレーション, High Performance Scientific Computing News (大阪大学サイバーメディアセンターホームページ内), 6, 2019/7/1.
- [19] (共同利用・共同研究拠点研究費) 村田忠彦, 市川学, 後藤裕介, 李皓, 原田拓弥, 伊達進, 棟朝雅晴, 杉本章義, 合成人口データの配布方法と活用に関する研究, 2019 年度関西大学ソシオネットワーク戦略研究機構公募研究課題 (2019/7/15~2022/3/31) .
- [20] (招待講演) Tadahiko Murata, “Toward the realization of real-scale social simulations,” University College Dublin Ph. D Seminar, (Dublin, Ireland, 2019/10/4)
- [21] (HPCI) 村田忠彦, 市川学, 後藤裕介, 李皓, 原田拓弥, 伊達進, 棟朝雅晴, 杉本章義, 埜敏博, 合成人口のためのデータストレージ, 令和元年度 HPCI 共用ストレージ利用研究課題 (2020/2/1~2021/1/31) .
- [22] (インタビュー) 村田忠彦, インタビュー: リアルスケール人工社会で展開するデータサイエンスとシミュレーションの世界, iic-HPC (北海道大学情報基盤センター大型計算機システムニュース), Vol. 50, pp. 2-7, 2020/3.
- [23] (関西大学研究費) 村田忠彦, 佐々木美絵, 中井豊, 中元康裕, 原田拓弥, 松下光範, 松本渉, 合成人口の利活用に関する研究, 関西大学研究拠点形成支援経費 (2020/4/1~2022/3/31) .
- [24] (JHPCN) 村田忠彦, 市川学, 後藤裕介, 李皓, 原田拓弥, 伊達進, 棟朝雅晴, 杉本章義, 埜敏博, 社会の分析とシミュレーションのための合成人口データ提供システム, 2020 年度学際大規模情報基盤共同利用・共同研究拠点公募型共同研究課題.
- [25] (科学研究費) 村田忠彦, 合成人口データを用いた心停止発生シミュレーション, 科研基盤研究 (C) (2020/4/1~2023/3/31) .
- [26] (科学研究費) 稲垣誠一, 小塩隆士, 高山憲之, 山田篤裕, 村田忠彦, 原田拓弥, 老後生活に関するダイナミックマイクロシミュレーション, 科研基盤研究 (B) (2020/4/1~2023/3/31) .