

jh200022-DAH

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

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。2020 年度は、都道府県の統計に過度に調整されていた世帯構成を調整した最新の人口合成手法を用いて、現実と同じ規模の世帯構成をもつ人口を合成し、合成人口データベースの構築に取り組んだ。本年度は、内閣官房が公募する COVID-19 AI・シミュレーションプロジェクトにもデータの提供を行い、実スケールのシミュレーションによる感染対策・経済対策に取り組むことができた。

1. 共同研究に関する情報

(1) 共同研究を実施した拠点名

北海道大学 東京大学 大阪大学

(2) 共同研究分野

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

総括

村田忠彦

人口合成アルゴリズム

原田拓弥, 李 皓

人口合成プログラム配備

伊達 進

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・応用

市川 学, 後藤裕介

用するために、高度な解釈を行ってきた。そこで、本研究プロジェクトでは、現実と同じスケールの世帯構成をもつ人口を合成し、その人口に対するシミュレーション結果から、より現実に応用しやすい解釈を可能とするリアルスケール社会シミュレーション基盤を構築する。現在、国レベルの合成人口データの公開は、アメリカおよびイギリスで行われているのみであり、アジア初の RSSS 基盤として、日本の合成人口データベース構築には大きな意義がある。

2020 年度は、文献[C]の最新の人口合成手法を用いた人口合成を実施し、研究者向けに提供する。文献[C]の手法は、従来手法(文献[A])において、全ての市区町村に対して、その所属する都道府県単位の統計に過度に調整されていた世帯構成を、都道府県全体で調整できるように改良した手法である。

図 1 に従来手法[A]と文献[C]の違いを示す。従来手法では、全ての市区町村が、所属する都道府県の統計に同時に適合するように人口合成が行われていた。その際、図 1 (a)のように都道府県単位の統計を各市区町村の単位に調整するため、都道府県単位の統計で少数の世帯や個人の合成が行われなくなる可能性があった。文献[C]の手法では、(b)のように同じ都道府県に所属する市区町村の人口合成を行いつつ、全ての市区町村の合成人口の全体として、都道府県の統計に適合するように合成が行われている。そのため、各市区町村の特徴を保持しながら、市区町村の人口全体として都道府県の統計に合致するように合成を行うことが可能となる。

2. 研究の目的と意義

2019 年度から JHPCN 公募研究として開始している合成人口プロジェクトでは、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を目指している。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。これまで、多くの社会シミュレーションでは、時にはトイプロBLEMと呼ばれる小規模なシミュレーションから得られた含意を、現実の社会に適

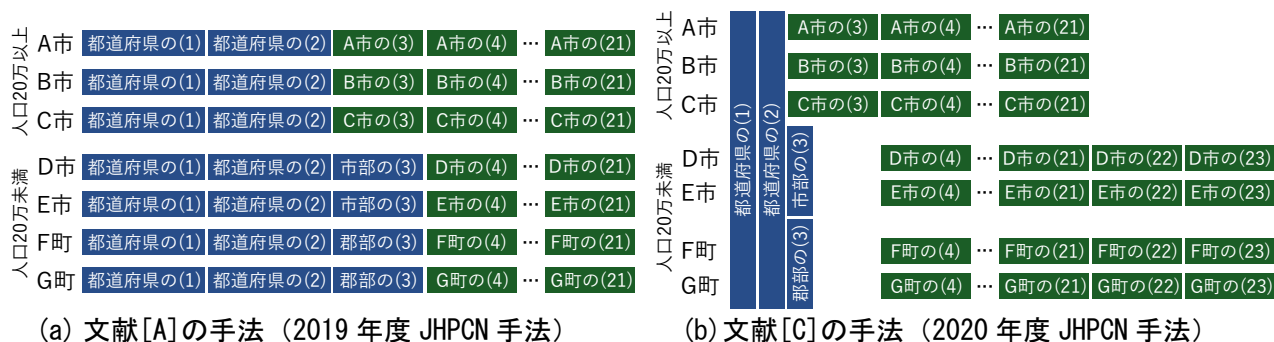


図 1：合成人口における統計の適用方法

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。

3. 当拠点公募型研究として実施した意義

本年度は、2019 年度に引き続き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、北海道大学の計算機を用いて、合成人口データベースを構築する。研究者向けに合成人口データを提供するために、表 1 のようなデータ保護レベルを設け、データ提供を行なっている[業績 1]。2020 年度は新型コロナウイルス感染症への対策を検討するため、2020 年 7 月に内閣官房 COVID-19 AI・シミュレーションプロジェクトが公募され、本公募型研究代表者は感染症対策に活用できるデータとして合成人口データを申請した。その結果、9 月から開始した 2 つの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、長野県の観光地における感染症対策に関するシミュレーション、急激な感染者数の増加を避けるための Stay with Community の効果を確認する首都圏と横浜市を対象としたシミュレーションに活用された[業績 10]。本プロジェクトで、日本の全自治体を対象とした合成人口データを整備することができたため、迅速なデータ提供につなげることができた。

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。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。

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。

- [A] Tadahiko Murata, Takuya Harada, Daiki Masui, Comparing Transition Procedures in Modified Simulated-Annealing-Based Synthetic Reconstruction Method Without Samples, *SICE Journal of Control, Measurement, and System Integration*, Vol. 10, No. 6, pp. 513-519, June 2017.
- [B] 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, June 2017.
- [C] Tadahiko Murata, Takuya Harada, Synthetic Method for Population of A Prefecture Using Statistics of Local Governments, *Proc. of IEEE International Conference on System, Man & Cybernetics* (Miyazaki, Japan, Oct. 7-10, 2018), pp. 1171-1176, October, 2018.

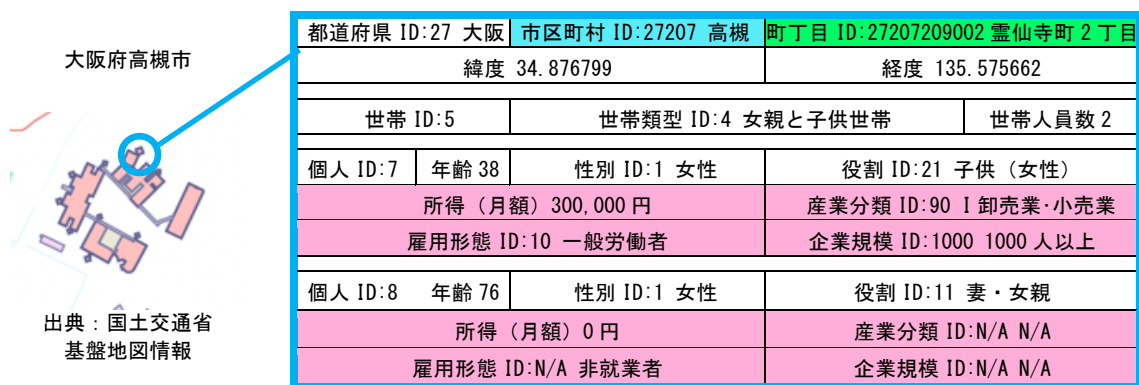


図 2：合成人口データの世帯情報（大阪府高槻市霊山寺町のサンプル）

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

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

4. 前年度までに得られた研究成果の概要

2019 年度は、大阪大学の計算機を用いて、図 2 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 2 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。現在、表 1 のデータ保護レベルにしたがって、図 2 中の色分けされた属性を含めるかどうかを決定して、データの提供を行っている。

また、北海道大学のインタークラウドシス

テムを用いて、図 3 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。

2019 年度は、表 2 のように 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、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。研究代表者らの手法では、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、研究代表者ら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

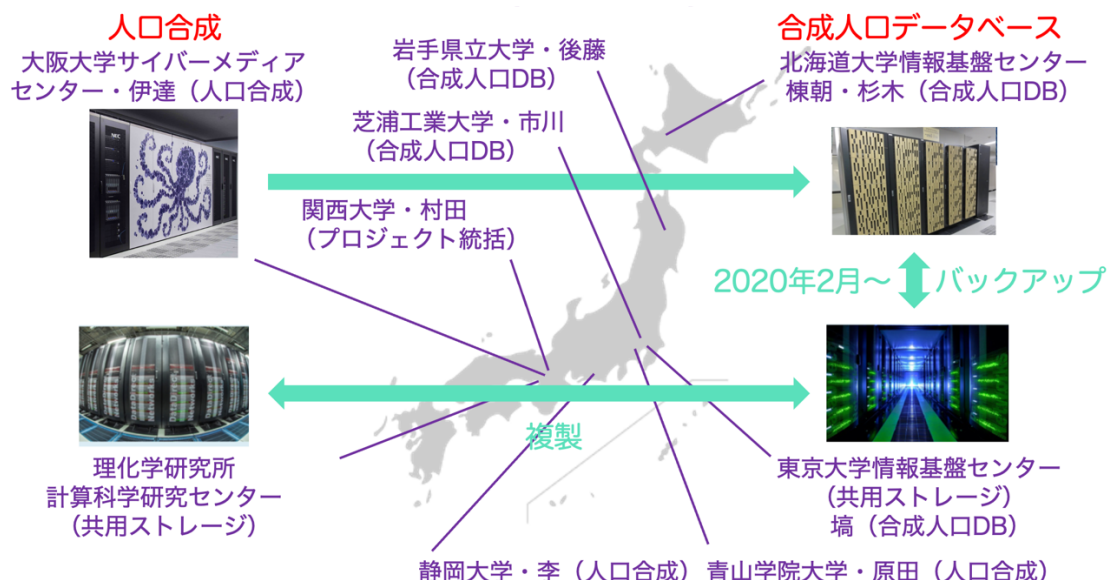


図4：HPCI-JHPCN 合成人口プロジェクト

まれている町丁目が 0.3%存在していたためである。

5. 今年度の研究成果の詳細

2020 年度は、2019 年度と同様に人口合成を大阪大学の計算機を用いて行い、合成人口データベースを北海道大学のインタークラウドシステムを用いて構築した。人口合成には、文献[C]の手法を用いて、2010 年と 2015 年の市区町村レベルの人口を合成し、合成世帯の町丁目および建物へ割付けている。建築物カバー率は、国土地理院の提供する基盤地図情報の制約のため、表 2 と同様である。また、2020 年 2 月以降 HPCI 共用ストレージを用いて、北海道大学情報基盤センターの合成人口データベースのバックアップを、東京大学情報基盤センターと理化学研究所計算科学研究センターのストレージを用いて行っている（図 4）。

さらに、2019 年度から開始している合成人口データの研究者向け公開を、表 1 のデータ保護レベルにしたがって開始し、本プロジェクトのメンバーである岩手県立大学、芝浦工業大学、青山学院大学、静岡大学以外に、北海道大学、芝浦工業大学、筑波大学、東京大学、東京工業大学、国際医療福祉大学、神戸大学、国立保健医療科学院、聖路加国際病院、国立情報学研究所から問合せをうけ、合成人口データの配布を行なっている。

2020 年度において、本プロジェクトの共同研究者が取り組んだ合成人口データを用いた研究は以下のように大別される。

1) 合成人口データの配布方法に関する研究

- データ保護レベル別提供方法[業績 1, 8, 9, 22]

表 1 に示したデータ保護レベルに関する

検討を行った。また、個票データの特定化をさけるために行われている k-匿名化と、人口合成手法による複数の個票データを合成する手法の違いに関する検討を行った。合成人口データの提供方法に関する発表で、2020 年度日本知能情報ファジィ学会ファジィシステムシンポジウムにおいて、ポスター・デモセッション最優秀賞を受賞した[業績 8]。

2) 合成人口データの新手法の開発

- 合成世帯の建物へのマッピング[業績 6, 14, 21]

合成された世帯を地図上の建造物に割り当てる際、その建造物が立地する土地利用用途を検討することにより、一層適切な割当を行うことが可能となる。土地利用用途を推定する手法の検討を行った。

- 所得の割当て[業績 7]

合成世帯における就業者の所得の割当手法に関する検討を行った。

- 施設世帯を含めた人口合成[業績 11]

施設に居住する世帯を推定し、人口の合成を行った。

- 人口動態に関する研究[業績 18, 20]

結婚、出生、転居などのライフイベントを合成世帯に対して適用することにより、自治体における人口動態の推定を行った。

3) シミュレータの開発

- 社会シミュレーションプラットフォームの開発[業績 16]

都道府県間の移動を伴うシミュレーションを行うための計算機環境に関する研究を行った。

4) アプリケーションの開発

- ・救急医療の環境整備に関する研究[業績 2, 13, 19]

突発的心停止に備えるため AED（自動体外式除細動器）の設置場所に関する検討を行った。

- ・新型コロナウイルス感染症に関する研究[業績 3, 4, 10, 12]

新型コロナウイルス感染症の感染拡大を防止するための方策や、観光における密をさけるための方策に関するアプリケーションを開発した。北海道全体を対象としたシミュレーションにより、第 24 回計測自動制御学会社会システム部会研究会において、優秀賞を受賞した[業績 12]。

- ・地震時の避難行動に関する研究[業績 15]

熊本地震の際の熊本市における避難行動を事例としてモデル化し、避難所への移動状況や避難所ごとの年齢分布などを分析するためのアプリケーションを開発した。

- ・ベーシックインカムに関する研究[業績 17]

浜松市を対象としてベーシックインカムを導入することにより、所得が増加する世帯と減少する世帯の割合を示し、ベーシックインカム制度の検討を行った。

このように、本プロジェクトでは、1) 合成人口データの利活用を研究者コミュニティに呼びかけつつ、2) 合成人口データの精度向上と属性の追加、3) 合成人口データを用いた大規模シミュレーション環境の構築、4) 合成人口データを用いたアプリケーションの開発を行うことにより、RSSS（リアルスケール社会シミュレーション）に関する技術の改善をはかるとともに、実務に役立てることができる環境の構築に取り組んでいる。

6. 今年度の進捗状況と今後の展望

今年度は、国内学会での招待講演[業績 24-26]、国際学術団体でのチュートリアル[業績 27]、放送大学での本プロジェクトの紹介[業績 28]、SC20 での研究展示[業績 5-7]など、様々な機会に合成人口データを紹介し、その活用方法を示すことができた。

合成人口データの利活用に関する案内を積極的に行うことにより、2021 年 3 月に開催された計測自動制御学会社会システム部会研究会では、全 52 件の研究発表中、25%を超える 14 件の発表が合成人口データを用いた研究となり、本プロジェクトが推進するリアルスケール社会シミュレーション（RSSS）の研究手法の新たな潮流が始まっていることがわかる。

また、合成人口データを活用するための研究費も獲得している[業績 29-34]。経済学や社会学の分野に合成人口データを応用するため

の研究プロジェクト[業績 29]や、内閣官房による新型コロナウイルス対策のための医療研究プロジェクト[業績 33]、本プロジェクトの共同研究者が参画して、JST 未来社会創造事業の探索研究課題として採択されている[業績 34]。また、本プロジェクトの若手研究者が単独で獲得した所属機関における研究費[業績 32]もあり、本プロジェクトで構成された合成人口データを基盤として、さまざまな研究が展開されている。

7. 研究業績一覧（発表予定も含む。投稿中・投稿予定は含まない）

(1) 学術論文（査読あり）

(2) 国際会議プロシーディングス（査読あり）

- [1] Tadahiko Murata, Susumu Date, Yusuke Goto, Toshihiro Hanawa, Takuya Harada, Manabu Ichikawa, Hao Lee, Masaharu Munetomo, Akiyoshi Sugiki, Distribution System for Japanese Synthetic Population Data with Protection Level, Proc. of International Conference on Machine Learning and Cybernetics: ICMLC 2020 (Adelaide, Australia, December 4-6, 2020), 5 pages
- [2] Tadahiko Murata, Atsuki Fukushima, Takuya Harada, Mie Sasaki, Social Awareness from Analysis of Available Time for Automated External Defibrillators in a City, Proc. of IEEE International Conference on Cybernetics (Online, June 8-10, 2021), 5 pages (Accepted).
- [3] Tadahiko Murata, Kanta Yamashita, Estimating the Effect of COVID-19 Contact Tracing Application Using Agent-Based Simulation, Proc. of IEEE International Conference on Cybernetics (Online, June 8-10, 2021), 6 pages (Accepted).
- [4] Tadahiko Murata, Kohei Totsuka, Agent-Based Simulation for Avoiding the Congestions of Tourists, Proc. of IEEE International Conference on Cybernetics (Online, June 8-10, 2021), 5 pages (Accepted).

(3) 国際会議発表（査読なし）

- [5] (ポスター展示) Tadahiko Murata, Susumu Date, Yusuke Goto, Toshihiro Hanawa, Takuya Harada, Manabu Ichikawa, Hao Lee, Masaharu Munetomo, Akiyoshi Sugiki, Synthetic Population for Real-Scale Social Simulations, Exhibition at IEEE/ACM International Conference for High Performance Computing, Networking,

Storage and Analysis: SC 20 (ONLINE, USA, Nov. 17-19, 2020)

- [6] (ポスター展示) Takuya Harada, Tadahiko Murata, Projecting Synthetic Household on Buildings in Maps for Real-Scale Social Simulations, Exhibition at IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis: SC 20 (ONLINE, USA, Nov. 17-19, 2020)
- [7] (ポスター展示) Sho Sugiura, Tadahiko Murata, Takuya Harada, Income Estimation for Workers in Synthetic Population for Real-Scale Social Simulations, Exhibition at IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis: SC 20 (ONLINE, USA, Nov. 17-19, 2020)

(4) 国内会議発表 (査読なし)

- [8] **2020 年度日本知能情報ファジィ学会ファジィシステムシンポジウム ポスター・デモセッション最優秀賞受賞**, 村田忠彦, 市川学, 後藤裕介, 杉木章義, 伊達進, 塙敏博, 原田拓弥, 棟朝雅晴, 李 皓, 日本人口の保護レベル別合成データ配布システム, 第 36 回ファジィシステムシンポジウム (ビデオ会議, 久留米, 福岡, 9 月 7-9 日, 2020), pp. 269-272
- [9] 村田忠彦, e-Stat で公開された統計情報に基づく合成人口データを用いたシミュレーションの展開, 官民オープンデータ利活用の動向及び人材育成の取組 (オンライン, 11 月 17 日)
- [10] 大澤幸生, 村田忠彦, 坪倉正治, 倉橋節也, 早矢仕晃章, 前川知英, 長谷大輔, ステイ・ウィズ・コミュニティの実践方法〜コロナ禍での安全生活と価値創造〜, 電子情報通信学会人工知能と知識処理研究会 (AI) (ハイブリッド開催, 人吉, 熊本, 2 月 12 日, 2021), 信学技報 AI2020-24, pp. 12-17
- [11] 原田拓弥, 村田忠彦, 国勢調査結果を用いた全ての一般世帯と施設などの世帯を含む全世界帯の合成, 計測自動制御学会第 24 回社会システム部会研究会資料 (オンライン, 3 月 6-7 日, 2021), pp. 17-23
- [12] **第 24 回計測自動制御学会社会システム部会研究会優秀賞** 嶋直紀, 松井陽太郎, 市川学, 国内における COVID-19 の流行シミュレーションの構築, 計測自動制御学会第 24 回社会システム部会研究会資料, pp. 31-39 (2021)
- [13] 江尻雄一, 原田拓弥, 大内紀知, 村田忠彦, 佐々木美絵, 住宅内心停止への AED 利活用促進に向けた社会シミュレーション分析, 計測自動制御学会第 24 回社会システム部会研究

会資料 (オンライン, 3 月 6-7 日, 2021), p. 80-86

- [14] 福崎竜希, 山崎幹矢, 原田拓弥, 大内紀知, Pix2Pix を用いた建物の用途判別における土地利用ラベルの簡素化 -合成人口の改良に向けて, 計測自動制御学会第 24 回社会システム部会研究会資料, pp. 126-132 (2021)
- [15] 渡邊康平, 小田中悠, 中井豊, 市川学, エージェントベースモデルによる避難行動シミュレータの構築, 計測自動制御学会第 24 回社会システム部会研究会資料, pp. 131-137 (2021)
- [16] 松井陽太郎, 嶋直紀, 渡邊康平, 市川学, SOARS Toolkit を用いた社会シミュレーションプラットフォームの構築, 計測自動制御学会第 24 回社会システム部会研究会資料, pp. 133-138 (2021)
- [17] 根岸美知, 李皓, マイクロシミュレーションを用いた日本におけるベーシックインカム導入の一考察, 計測自動制御学会第 24 回社会システム部会研究会資料, pp. 154-161 (2021)
- [18] 太田奎祐, 李皓, 雇用増加に伴う人口増減予測モデルの構築, 計測自動制御学会第 24 回社会システム部会研究会資料, p. 172 (2021)
- [19] 福島惇輝, 村田忠彦, 原田拓弥, 佐々木美絵, 自動体外式除細動器 (AED) の利用可能時間に関する分析, 計測自動制御学会第 24 回社会システム部会研究会資料 (オンライン, 3 月 6-7 日, 2021), pp. 181
- [20] 岩田恵人, 李皓, 仮想個票データに基づく浜松市における将来の人口動態を推定するマイクロシミュレーション, 計測自動制御学会第 24 回社会システム部会研究会資料, p. 189 (2021)
- [21] 山崎幹矢, 福崎竜希, 原田拓弥, 大内紀知, pix2pix を用いた建造物の用途判別精度向上のための複数点による判定, 計測自動制御学会第 24 回社会システム部会研究会資料, p. 195 (2021)
- [22] 村田忠彦, リアルスケール人工社会 データサイエンスとシミュレーションによる未来デザイン基盤, 社会・経済システム (社会・経済システム学会誌), 第 40 号, pp. 1-50 (2020 年 7 月 11 日招待講演 講演録)

(5) 公開したライブラリなど

- [23] 村田忠彦, 合成人口データリクエストフォーム, <http://www.res.kutc.kansai-u.ac.jp/~murata/rsss-distribution/>

(6) その他 (特許, プレスリリース, 著書等)

- [24] (招待講演) 村田忠彦, リアルスケール人工社会 データサイエンスとシミュレーション

ンによる未来デザイン基盤，社会・経済システム学会研究会「新しい資本主義と社会・経済システムのデザイン」（ビデオ会議システム，ホスト京都大学），2020 年 7 月 11 日

- [25]（招待講演）村田忠彦，合成人口を用いたリアルスケール社会シミュレーション，計測自動制御学会第 3 回スマートワールド実現のための新たなシステムズアプローチの実展開を目指す調査研究会（ビデオ会議システム，ホスト同志社大学），2020 年 8 月 1 日
- [26]（招待講演）村田忠彦，合成人口を用いたリアルスケール人工社会，日本知能情報ファジィ学会ファジィ学問塾（ビデオ会議システム），2020 年 9 月 7 日
- [27]（チュートリアル）Tadahiko Murata, Data Distribution for Japanese Synthesized Population and Real-Scale Social Simulations, Social Simulation Week 2020 (Video Conference, Hosted by Milan, Italy, September 14-18, 2020), September 15, 2020
- [28]（収録講義）村田忠彦，合成人口を用いたリアルスケール社会シミュレーション，放送大学キャリアアップ講座「データサイエンス基礎から応用」第 7 回「AI によるデータサイエンスとシミュレーション」（第 1 部）（放送大学にて収録，2022 年まで再放送）
- [29]（研究費）村田忠彦，関西大学研究拠点形成支援経費（2020 年，2021 年度），合成人口データの利活用に関する研究，採択.
- [30]（研究費）村田忠彦（研究代表者），科学研究費基金 基盤研究（C）一般（2020-22 年度），合成人口データを用いた心停止発生シミュレーション，採択.
- [31]（研究費）村田忠彦（研究分担者），科学研究費基金 基盤研究（B）一般（2020-22 年度），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，採択.
- [32]（研究費）原田拓弥，青山学院大学アーリーイーグル研究支援制度（2020 年），国勢調査に基づく人工社会への属性追加のための属性依存関係抽出手法の開発，採択.
- [33]（研究費）村田忠彦（研究分担者），内閣官房受託研究（2021 年 1 月まで），「スマートライフ実現のための AI 等を活用したシミュレーション調査研究業務」，採択.
- [34]（研究費）村田忠彦，後藤裕介，市川学，原田拓弥（研究分担者），「社会政策立案に向けたマルチスケール ABSS 手法」（研究代表者・貝原俊也），JST 未来社会創造事業「超スマート社会の実現」領域「異分野共創型の AI・シミュレーション技術を駆使した健全な社会の構築」，採択.