

社会シミュレーションのためのデータ合成

○喜多 一 (京都大学) 村田忠彦 (関西大学)

On Data Synthesis for Social Simulation

*Hajime Kita (Kyoto University) Tadahiko Murata (Kansai University)

概要— 合成人口データはリアルスケールの社会シミュレーションにおいて不可欠の技術として位置付けられる。これは従来、国勢調査など「個票」がデータとして存在するものについて匿名化などの方法で利用を検討してきたことと発想を異にし、集計量として公表されている統計に合致することを条件に仮想の人口を構成することで、個人情報を利用することなく、個人レベルのシミュレーションを可能にしたことにある。本稿では、合成人口の背景にある考え方を踏まえ、社会シミュレーションのためのデータ合成技術としての一般化を提唱する。

キーワード: 社会シミュレーション, 合成人口データ, データ合成

1 はじめに

合成人口データ／模擬個票^{1, 2)}はリアルスケールのエージェントベースの社会シミュレーション(以下RS-ABSS)において不可欠の技術として位置付けられる。従来、国勢調査など「個票」がデータとして存在するものについては匿名化などの方法で利用を検討してきた。合成人口データはこれと発想を異にし、公開されている統計に合致することを条件に仮想の人口を構成することにより、個人情報を利用することなく個人レベルのシミュレーションを可能にしている。本稿では、合成人口の背景にある考え方を踏まえ、RS-ABSSのためのデータ合成技術としての一般化を提唱する。

2 集計量からのデータの再構成

一般に集計量から、集計対象となった原データを再構成することは情報量が不足し、解は一意ではない。例えば N 件のデータが2つの属性 X, Y を持ち、 X と Y についての周辺分布しか分からず、 X, Y の同時分布が不明な場合を考える。この場合、他に論拠がなければ X, Y の相関がないと仮定して原データを復元させるという考え方がある。これは単に按分計算をしているだけであるが、未知の情報の不確かさとしてエントロピー最大³⁾を仮定した推計とも言える。周辺分布への適合を拘束条件としつつ、実際の統計調査などでのサンプルの重みを調整する方法として Iterative Proportional Fittingがあるが、これも同様の考え方に基づいている。

3 合成人口データの構成法

村田らの合成人口データ^{1, 2)}は国勢調査について公表されている統計量に基づき、地方自治体を単位に統計への合致を目的関数に模擬焼きなまし法(SA)を使って世帯構造を再構成し、さらに居住地、職種、所得などを他の統計などの情報源に基づいて割り当てている。世帯はいわば「夜間人口」であるが昼間人口を構成する就業先の割り当ても行っている。

個々の合成人口データは確定値であるが、人口データを拘束する統計量が少ないため、データは一意には定まらない。村田らはSAを複数回実行して、同一自治体に対し複数セットの合成人口データをサンプルし、サンプル群をアンサンプルとして提供している。

合成人口データの生成には個人情報を用いないが、これに加えアンサンプルとしてのデータ提供により同一地域に異なるセットのデータが割当てられており、個人

のプライバシー侵害が生じにくいという利点がある。また、実際の個票では利用に際して属性値を曖昧にする加工が行われるが、これ比べ統計で与えられる属性の精度を保存できるという利点もある。さらに異なる調査から得られる統計からのデータ合成の一元的な扱いも可能にしている。

4 ペルソナ法

サービスや商品の設計では、仮想の利用者を詳細・具体的に設定してユーザーエクスペリエンス(UX)などを検討するペルソナ法という手法がある。個人ベースの手法であるRS-ABSSではこのようなペルソナを大量に設定・利用しているとも解釈でき、「社会レベルのペルソナ法」を用いていることになる。データ合成の手法はこのための手法として位置付けることもできる。

5 ABSSのためのデータ合成

合成人口データでは「地方自治体の住民／世帯」をデータ合成の対象としているが、データ合成という観点からはその一例として捉えられる。同様の手法は対象とする集団について一定の統計量が利用可能ならば適用可能である。ただし、複数の統計間で接合可能な情報が使えない場合は2節で述べた按分計算に留まらざるを得ない。適用例として以下などが考えられるが、これらに仮想データを用いれば組織活動のシミュレーションによる活動案の策定などにも利用可能である。

- 企業活動における(実／潜在)顧客
- 大学等の教育機関における学生

6 おわりに

本稿はRS-ABSSにおいて重要な位置づけを持つ合成人口データの考え方を一般化した社会シミュレーションのためのデータ合成を提唱したものである。今後は具体的な課題を対象にデータ合成を検討して行きたい。

参考文献

- 1) 村田, 原田, 杉浦: リアルスケール社会シミュレーションの実現に向けて, シミュレーション, **36**-4, 58/62 (2017)
- 2) 村田, 原田: 仮想実社会データを用いたリアルスケール社会シミュレーションの実現, 計測と制御, **62**-1, 9/14 (2023)
- 3) 土屋和雄: 最大エントロピー法による複雑系の設計論, 日本ファジィ学会誌, **8**-5, 803/806 (1996)