FOSSR (FOSTERING OPEN SCIENCE IN SOCIAL SCIENCE RESEARCH) Integrating statistical data into social simulation



Conferenza Nazionale di Statistica

ROCCO PAOLILLO – CNR-IRPPS rocco.paolillo@cnr.it X: PaolilloRoc | MARIO PAOLUCCI – CNR-IRPPS mario.paolucci@cnr.it X: @mario_paolucci

Abstract

Agent-based models are computer simulations for studying and experimenting on the dynamics of collective phenomena that emerge from the interaction of individual social actors and institutions. Within the FOSSR project for an Italian scientific open cloud, we aim to provide an automated service to initialize agent-based models representative of the Italian population. Our goal is to provide algorithms for detailed reconstruction of synthetic populations from data of the FOSSR research

meta-infrastructure.

Key Objectives

- Social simulation studies the emergence of collective phenomena through the interaction of virtual agents that replicate human behaviours, cognitions and goals and that can adapt to their environment
- Our goal is to develop a Software as a Service (SaaS) in the FOSSR open cloud for the initialization of simulations representative of the information stored in the cloud



Agent-based Modelling

Method to build artificial societies and study the emergence of collective phenomena from the **interaction** of agents



Synthetic Populations

 Artificial societies where virtual agents replicate the socio-demographic attributes of the target population

| | | | # | # | #

• Synthetic reconstruction: computation of joint distributions of the target population integrating different

 Validation of IPF algorithm with ISTAT 2022 data on gender and age distribution in Italy



 Synthetic populations can enhance the specifity of behavior based on the segment of the target population they represent

Conclusions

 The Iterative Proportional Fitting is the archetypal technique for extraction of synthetic populations from aggregated and independent sources of information, as it can be in the FOSSR open cloud

FÔSSR sources of information **Iterative Proportional Fitting** Fostering Open Science in Social Science Research no compute weight for each compute weight for update fitted benchmark crossed cell for cell for dimension 2 met? marginals dimension 1 (row) (column) GENDER yes → stop male female 31 mln fitted marginal observed ≤ 50 Equivalentt to ≤ 50 marginals raking in statistics 22 mln fitted Computing **joint** ¥ 51 ≤ • • • marginal observed **51 ≤ 80** marginals probablities from 80 known marginals 4 mln fitted observed marginal ≥ **81** in the population marginals ≥ 81 • Estimating fitted fitted At step 0, each marginal marginal weights for each cell has value 1 female male cross-category 30 mln 28 mln Target: Total Absolute Error ≈ 0 observed marginal value weight = fitted marginal value

Impact

- Synthetic populations enhance the potential of agentbased modelling for policy testing and social research where intersectionality of demographic attributes need to be considered
 As part of the FOSSR open cloud services, the extraction of synthetic populations will be easier to integrate with other innovative methods for data collection and knowledge extraction for the social sciences
- We will provide algorithms to its extension for different type of data and requirements of users

Contact Information:

For more details, please visit www.fossr.eu or contact fossr.dissemination@cnr.it

