Scalable population synthesis with Conditional Variational Autoencoder and combinatorial optimization

Daniel Horst[1,2], **Helen Ganal**[2], **Friedrich Krebs**[1,2]
Contact: Daniel Horst | +49 561 7294-263 | daniel.horst@iee.fraunhofer.de

[1] Universität Kassel
[2] Fraunhofer IEE

AgentHomeID

The purpose of **AgentHomeID** is to forecast energy-relevant investments in German residential buildings by their owners under assumed socio-economic boundary conditions until 2050. To do so, AgentHomeID is integrated into a detailed model of the German building stock.
The development of the building stock depends largely on the investment decisions made by the building owners when refurbishing the building envelope and installing new heating systems. To tackle this complexity, AgentHomeID considers both the heterogeneity in the building structure and the heterogeneous investment preferences of the building owners. One of the challenges is to provide a coherent mapping of the technical specifications of individual buildings and the respective investment preferences of its owner. In order to map developments locally and to analyze the influence of regulatory boundary conditions, detailed knowledge of the local building stock and the associated ownership structure is required. We present a combined approach of Synthetic Reconstruction (SR) and Combinatorial Optimization (CO) to generate a realistic local building stock and a linked agent population including socio-economic and -demographic profiles of resident owners.

Scalable Population Synthesis with Conditional Variational Autoencoder on Community Level

To map the local building stock including residents and owners, an approach based on a Variational Autoencoder framework with additional boundary conditions (CVAE) is used. In the unsupervised learning process of the CVAE, the joint probability distribution P(X) of the building stock is learned together with the socio-economic and demographic profiles of tenants and owners from the surveys of the Micro census Scientific Use Files[1] 2014 (SUF). Figure 1 shows a schematic representation of the CVAE structure with regional boundary conditions.

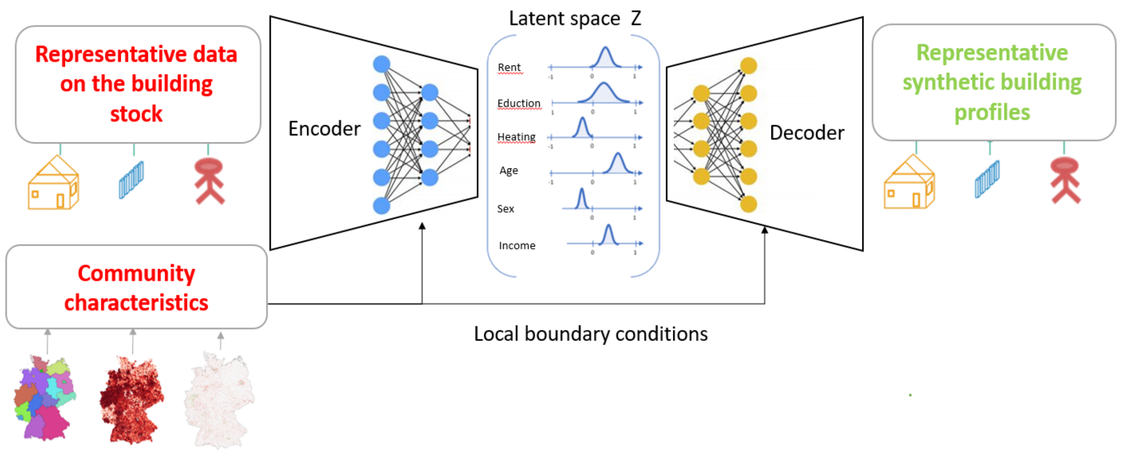

Figure 1: Schematic depiction of a CVAE. The CVAE learns to map ("decode") the distribution of the latent variable Z (for a multivariate Gaussian) into the data space to approximate P(X). This process is facilitated by learning the latent space representation ("encoding") of the data during the training.[2]

From the SUF, 25 categories with a total of 320 attributes are selected as representative data on the building stock. These describe the building stock, heating technology, and socioeconomic and demographic characteristics of tenants and owners. Regional characteristics such as the state and population of a municipality are included in the SUF, while the regional distribution of building ownership and structure is extracted specifically from the census 2011[3] at a municipal level.

After successful training, any number of profiles can be generated to sample the latent variables and map them to the observed space. This enables a community-level population to be modeled at a very high level of detail, including smaller zones, personal data and building information.

Generating Synthetic Population at Individual and Household Levels using Combinatorial Optimization

Second, the combinatorial optimization approach allows the profiles generated by the CVAE to be matched to available local entities, generating synthetic populations at the building level. The local units are from the 2011 Census survey[2] at a spatial resolution of 1 ha. The 2011 Census survey contains similar renter and owner categories as the 2014 SUF, allowing for easy combination of datasets.
A hill-climbing algorithm is used as the solution method. Meaning, the building and owner profiles corresponding to the number of buildings in a 2011 census cell are randomly selected from the CVAE profiles. Then, the deviation of the marginal distribution between the selected profiles and the spatial distribution of the census cell is determined. The total average error (TAE) is used as a measure of the deviation. A CVAE profile is now randomly swapped. If the new CVAE profile reduces the TAE, it is replaced. Otherwise, other possible replacements are randomly selected until no further improvements can be made to the selected combination. Figure 2 shows a schematic representation of the CO approach.

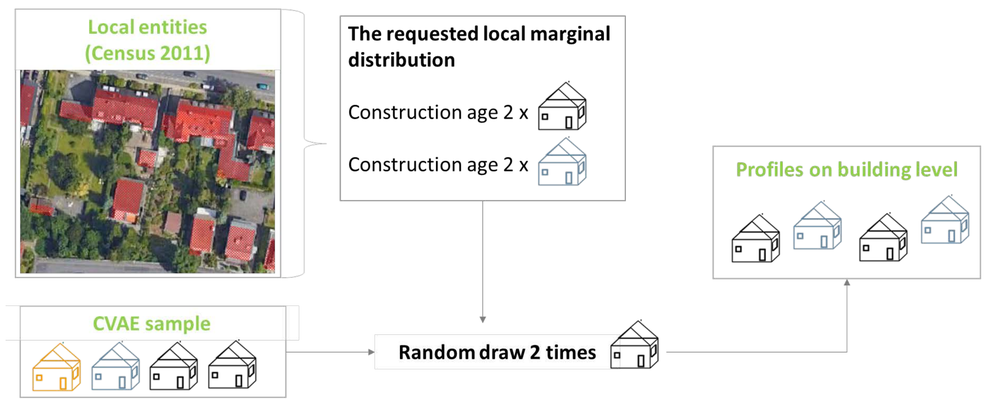

Figure 2: Schematic representation of the hill-climbing algorithm approach generating synthetic population at building level. The construction age is used as example.

The combined approach of CVAE and CO makes it possible to map regionally high-resolution and realistic building profiles that include socio-economic and demographic characteristics of residents and owners. Therefore, the presented approach can support agent-based modeling by enabling enriched synthetic populations with smaller zones and more detailed individual characteristics. In addition, the local information can be used for application in the field of regionalization of renewable energy producers and consumers.

Current and planned use cases of regional modeling of the building heating sector based on high-resolution data preparation:

- Influence of neighboring buildings and owners on each other.
- Expansion planning of the distribution grid considering the local development of the building sector.
- Planning of infrastructure development of district and local heating grids
- Feedback effects between decisions of building owners and expansion of infrastructure – providing support in the context of municipal heat planning.

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

1 Micro census | Research Data Center (2022). Available online at https://www.forschungsdatenzentrum.de/de/Households/microcensus, last updated on 07/25/2022, last checked on 07/25/2022.
2 Borysov, Stanislav S.; Rich, Jeppe; Pereira, Francisco C. (2018): Scalable Population Synthesis with Deep Generative Modeling (1). Available online at https://arxiv.org/pdf/1808.06910.
3 ZENSUS2011 (2022): ZENSUS2011 - population and housing census 2011. Available online at https://www.zensus2011.de/DE/Home/home_node.html, last updated on 07/25/2022, last checked on 07/25/2011