International Cartographic Association
Association Cartographique Internationale

Vector-Based Super-Resolution Mapping with the Population Dynamics Foundation Model

Hongyu Zhang [a,*]

[a] *Earth, Geographic, and Climate Sciences, University of Massachusetts Amherst – honzhang@umass.edu*

* Corresponding author

Keywords: super-resolution mapping, foundation model, young adult debt, United States

Abstract:

Image super-resolution is a well-established research problem in computer vision, aimed at improving the quality of low-resolution images. The advancement of machine learning algorithms has introduced innovative methods for image super-resolution. This topic is particularly significant in remote sensing, where access to high-resolution images and advanced sensor technologies is often limited (Wang et al., 2022). Most existing research on super-resolution focuses on raster images, processing pixel-based data. Can machine learning techniques be applied to vector data super-resolution, such as administrative boundaries?

Foundation models, pre-trained on large-scale datasets, have been widely applied to natural language and computer vision tasks. Recent applications in geospatial artificial intelligence (GeoAI), as discussed by Mai et al. (2024), have demonstrated the potential of using foundation models for geospatial analytics. One notable advantage of using foundation models is their general-purpose nature: applying them to specific tasks often requires minimal downstream training and fine-tuning.

Google has released the Population Dynamics Foundation Model (PDFM) embeddings, which are generated by a Graph Neural Network (GNN) model (Agarwal et al., 2024). The model is trained on datasets including search trends, geospatial data, busyness metrics, and climate conditions. The embeddings comprise 330 features and are aggregated at the county and ZIP code levels, covering the contiguous United States. One application of the PDFM embeddings is super-resolution on vector data.

To illustrate, we use young adult debt data in America from the Urban Institute (Martinchek et al., 2024). This debt data is sourced from a major credit bureau that was last updated in August 2023. For young adults (consumers aged 18 to 24), the finest resolution available is at the state level due to sample size limitations. However, at the state level, spatial patterns of debt are often difficult to discern, which complicates the identification of regions with high levels of young adult debt and hinders the implementation of targeted support for those most in need. PDFM embeddings are applied to increase the resolution of the data. Specifically, the model is trained using state-level embeddings and debt data, and it predicts debt data at the county level using county-level embeddings.

The steps of debt data super-resolution are described as follows: first, I calculated the state-level PDFM embeddings by averaging each of the 330 features at the ZIP code level using Python. Next, I loaded the PDFM embeddings at the state and county levels into a Python data frame. Then, I joined the young adult debt data at the state level to the state-level PDFM embeddings. Columns containing NULL values were dropped. I used the ridge regression model from the scikit-learn package to train the model and make predictions. Ridge regression was selected because it is well-suited for estimating coefficients involving highly correlated independent variables (Hilt and Seegrist, 1977), as is the case with the PDFM embeddings. The predicted outcomes were saved in CSV format and visualized in ArcGIS Pro alongside the original state-level data (Figure 1).

The results in Figure 1 show that the predicted data (c and d) exhibit a pattern similar to the training data (a and b). Regarding the share of people with a credit bureau record who have any debt in collections, at the state level, a north-south divide was evident, with higher debt share percentages in the South and economically underdeveloped regions (e.g., Wyoming, Arkansas, Mississippi, and Louisiana). At the county level, the trend of high debt shares in the South extended into neighboring states, such as Texas and Georgia. In terms of the median amount of all debt in collections among those with any debt in collections, at the state level, Nevada had the highest median. However, at the county level, the highest medians were observed in California, Arizona, Wyoming, and Texas in addition to Nevada. This demonstrates the PDFM embeddings' capabilities in economic indicator modeling. By combining the two metrics, counties with the highest share

Abstracts of the International Cartographic Association, 10, 321, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-321-2025 © Author(s) 2025. CC BY 4.0 License.

of debt and median debt can be identified, which enables recommendations for localized economic policies to assist young adults in need.

In conclusion, this study demonstrates that machine learning techniques, specifically foundation models, can be effectively applied to vector data super-resolution. The case study extends super-resolution mapping from pixel-based data to administrative boundaries, highlighting its potential for vector-based applications. Future work may build on this foundation by integrating additional data sources and advanced GeoAI methods for improved super-resolution mapping.

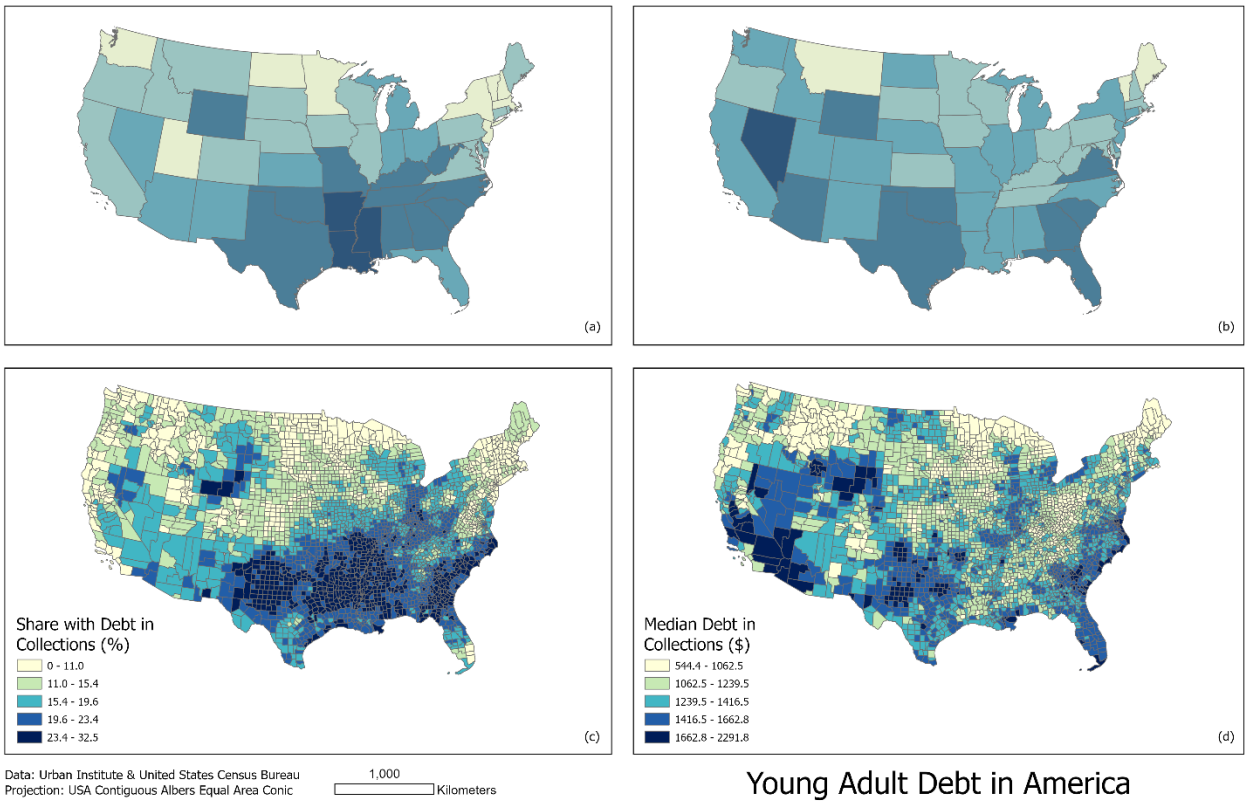

Figure 1. Young adult debt in America[1]. (a) State-level share of individuals with a credit bureau record who have any debt in collections. (b) State-level median amount of debt in collections among individuals with any such debt. (c) County-level super-resolution map of the share of individuals with a credit bureau record who have any debt in collections. (d) County-level super-resolution map of the median amount of debt in collections among individuals with any such debt.

References

Agarwal, M., Sun, M., Kamath, C., Muslim, A., Sarker, P., Paul, J., Yee, H., Sieniek, M., Jablonski, K., Mayer, Y., Fork, D., de Guia, S., McPike, J., Boulanger, A., Shekel, T., Schottlander, D., Xiao, Y., Manukonda, M.C., Liu, Y., Bulut, N., Abu-el-haija, S., Eigenwillig, A., Kothari, P., Perozzi, B., Bharel, M., Nguyen, V., Barrington, L., Efron, N., Matias, Y., Corrado, G., Eswaran, K., Prabhakara, S., Shetty, S. and Prasad, G., 2024. General Geospatial Inference with a Population Dynamics Foundation Model. *arXiv preprint arXiv:2411.07207*.

Hilt, D.E. and Seegrist, D.W., 1977. *Ridge, a computer program for calculating ridge regression estimates*. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R. and Lao, N., 2024. On the Opportunities and Challenges of Foundation Models for GeoAI (Vision Paper). *ACM Transactions on Spatial Algorithms and Systems*, vol. 10, no. 2, pp. 11:1–11:46.

Martinchek, K., Santillo, M., Braga, B. and McKernan, S.-M., 2024. *Debt in America: Updated September 18, 2024*. Available at: https://datacatalog.urban.org/dataset/debt-america-2024

Wang, P., Bayram, B. and Sertel, E., 2022. A comprehensive review on deep learning based remote sensing image super-resolution methods. *Earth-Science Reviews*, *232*, p.104110

[1] Legend classifications are consistent within each column.

https://doi.org/10.5194/ica-abs-10-321-2025 © Author(s) 2025. CC BY 4.0 License.