Extracting Information on Poverty and Wellbeing from Aerial and Satellite Images using Deep Learning

Name and Contact of the Supervisors

Prof. Dr. Jan van den Brakel, Dept. of Quantitative Economics, Maastricht University & Methodology Department, Statistics Netherlands, Heerlen

ja.vandenbrakel@cbs.nl

Dr. Rui Jorge Almeida, Dept. of Quantitative Economics & Data Analytics and Digitalisation, Maastricht University

rj.almeida@maastrichtuniversity.nl

Description of the project

The aim of this project is to improve deep learning algorithms to extract statistical information on poverty and wellbeing from satellite and high resolution aerial images and develop advanced prediction methods for poverty indicators on low regional levels. Regional statistical information on poverty is traditionally compiled from survey samples using Small Area Estimation (SAE) methods, see e.g. Pratesi (2016). SAE refers to a class of model-based inference methods (multilevel models) that use sample data to make detailed regional predictions at a level where only a few of even no sample data in the separate areas are available, see Rao and Molina (2015). The success of these methods depends on the availability of related covariates, usually derived from registers or censuses.

Parallel to the development of the SAE literature, several authors have proposed methods for combining survey data with sensor data such as satellite and aerial images or mobile phone data with the purpose of making detailed regional predictions for poverty and wellbeing. Noor et al. (2008) used nighttime light intensity as a measure for household income in Africa. Yeh et al. (2020) uses deep learning models on satellite images to predict survey-based estimates of asset wealth across African villages. Engstrom et al. (2017) uses daytime satellite images to predict wellbeing on a fine regional level in Sri Lanka. Blumenstock et al. (2015) used mobile phone data to predict poverty and wellbeing on low regional detail in Rwanda. These methods are predominantly applied in developing countries and combat areas were no high standard official statistical systems are in place. Most of these methods use survey sample data to train a machine learning algorithm to predict poverty or wellbeing using (remote) sensor data. In a next step the algorithm is applied to the sensor data only to predict the target indicators on low regional levels. The major drawback of this approach is the risk of overreliance on the algorithm used to produce predictions.

The purpose of this project is to develop novel deep learning algorithms to distract statistical information from satellite and high resolution aerial images. We will develop efficient sampling strategies in combination with active learning to annotate images using sampling information on

poverty and wellbeing. We will use the poverty predictions as covariates in the afore mentioned SAE models to avoid the risk of overreliance on an algorithmic model. This project will contribute to the existing methods to produce low regional estimates for poverty in developing countries but it will also improves the statistical quality of poverty estimates in developed countries because more timely related covariates to build SAE models becomes available from these new data sources.

Good to know

This project is made in close collaboration with Statistics Netherlands. This collaboration provides us with access to income data, derived from different registers, for all households residing in the Netherlands on an annual frequency. This allows to perform simulations with all kind of sampling strategies and active learning methods without spending costs and time to collect survey data on income to annotate images.

Features of a good candidate

This is a project for a very good, student with strong analytical skills with the ambition to conduct this research successfully within a period of four years resulting in a PhD. dissertation. The student should be proactive and cooperative. A strong background in econometrics, statistics or computer science is desired. Interest in programming is highly recommended, particularly parallel, distributed and GPU programming. Support is available to develop programming skills. Finally, the student must have a good command of the English language.

Signature for approval Head of Department Quantitative Economics

Dries Vermeulen

Maastricht 12-10-2022

References

Blumenstock, J., Cadamuro, G. and On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science, 350, pp. 1073-1076.

Pratesi, M. ed., 2016. Analysis of poverty data by small area estimation. John Wiley & Sons

Rao, J.N. and Molina, I., 2015. Small area estimation. John Wiley & Sons.

Engstrom, R., Hersh, J.S. and Newhouse, D.L., 2017. Poverty from space: using high-resolution satellite imagery for estimating economic well-being. World Bank Policy Research Working Paper, (8284).

Noor, A.M., Alegana, V.A., Gething, P.W., Tatem, A.J. and Snow, R.W., 2008. Using remotely sensed night-time light as a proxy for poverty in Africa. Population Health Metrics, 6(1), pp.1-13.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S. and Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nature communications, 11(1), pp.1-11.