Healthcare Mapping in Nigeria

Yuan Gao, Ruohan Zhan
Stanford University

ICME Xplore IDM

Abstract

Geographic modeling on healthcare outcomes can facilitate policy making, which can contribute to more systematic and targeted care to be provided in developing country. However, high-resolution mapping of healthcare outcomes are not standard and currently available data are expensive and have limited resolutions in time and space. Recent research (Blumenstock et al. 2015, Jean et al. 2016) has shown the potential of machine learning as an alternate mapping approach which makes use of passively-collected high-resolution data such as satellite and cellphone call data records to infer a spatial distribution of socioeconomic variables. This project extends previous work to obtain healthcare map in Nigeria, which has the potential to identify gaps in care and thus enable improved access to care and quality of the care available.

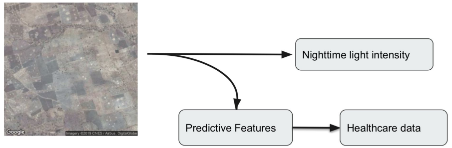
Data

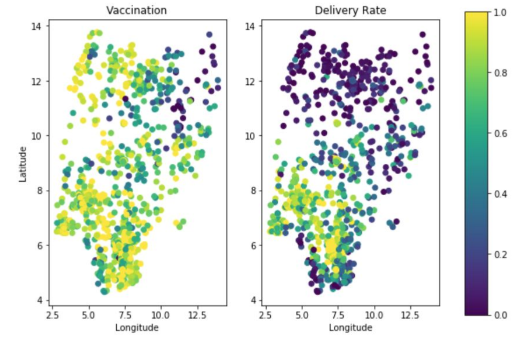
The healthcare datasets are from Demographic and Health Survey, while the satellite images are downloaded from Google Static Map. Specifically, we focus on survey done in Nigeria in 2013 which includes 904 clusters. Datasets needed to be plugged into our entire pipeline are listed as follows:

- 25 Satellite images for each cluster from Google Static Map (Cluster GPS is shifted up to 5km for privacy protection);
- Healthcare data
  - Maternal and neonatal child health(Birth recode: place of delivery)
  - Vaccination

Transfer Learning Model

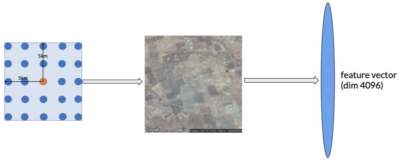
- Low level feature learning: pre-trained model on ImageNet to obtain low-level features such as edges and corners
- High level feature extraction: train the model to map satellite data to nightlight intensities to acquire high-level socioeconomic features embedded in the satellite data
- Supervised training: use the extracted features from the last layer of model acquired from the previous step to do a supervised training with healthcare data





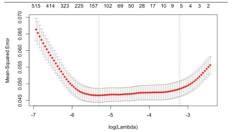
Experiments

We extracted features for each cluster's satellite image data. Each feature has a dimension of 4096



Then fitted vaccination data with lasso regression

- MSE
  - Training: 0.03187561
  - Testing: 0.04728266
- 138 Nonzero coefficients



Next Steps

- Systematically evaluate results
- Explore other machine learning methods to make prediction from extracted features to labels.
- Explore other unsupervised learning methods to extract features from satellite images based on geographical similarities [(Jean et al. 2018)
- Explore the impacts of changing geographies and solve the mistach between satellite images and healthcare data
- Adjust models to make more accurate prediction in target districts

Contact {gaoy, rhzhan}@stanford.edu

References

1. Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science , 353 (6301), 790-794.
2. Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. Science , 350 (6264), 1073-1076.
3. Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2016, March). Transfer learning from deep features for remote sensing and poverty mapping. In Thirtieth AAAI Conference on Artificial Intelligence.
4. Perez, A., Yeh, C., Azzari, G., Burke, M., Lobell, D., & Ermon, S. (2017). Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning. arXiv preprint arXiv:1711.03654.
5.Jean, Neal, et al. "Tile2Vec: Unsupervised representation learning for spatially distributed data." CoRR (2018).