Poverty Prediction by Selected Remote Sensing CNN Features

Stanford University

Zhihan Jiang, Yicheng Li, Zhaozhuo Xu

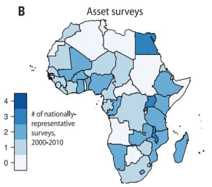
CS 229 Machine Learning

Motivation

Accurate poverty measurements of certain area critically shape the decisions of local governments about how to allocate scarce resources, and to track the progress toward improving human livelihoods.

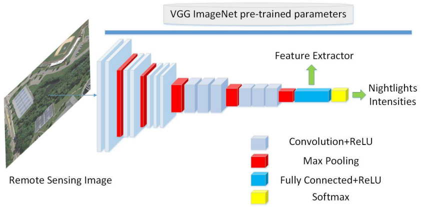
Despite the lack of the ground truth asset survey data, we propose a poverty predictor[1] based on the satellite images[2], according to the following steps:

- Train CNN as a feature extractor rather than a classifier.
- Reduce overfitting by wisely choosing CNN features.
- Improve prediction performance by using stronger regression models.



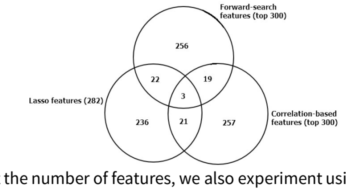
CNN Based Feature Extraction

We use a fully convolutional model converted from the VGG F model pre-trained on ImageNet. We train this VGG model by using remote sensing images with corresponding night-lights intensities as labels. The input to our VGG model is 400x400 pixel and the 4096 fully connected layer features produced by our model are selected as our CNN features.

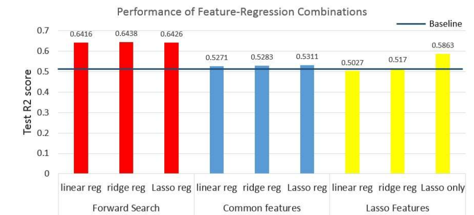


Experiment

Applying feature selection methods discussed in the previous session yields the Venn diagram shown below. We then use these features to train all the regression models, and compare their training/test R^2 scores, which provides a measure of how well observed outcomes are replicated by the model based on the proportion of total variation of outcomes explained by the model. (Refer to reports for more details)



To limit the number of features, we also experiment using only the intersected features in the diagram above.



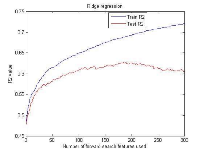
Model Specification

Coarse Feature Selection

- Correlation-based feature selection computes each feature's correlation with the label (asset index). Then we use the features with the highest correlation in our regression.
- Principal components analysis (PCA) is often used in unsupervised learning to find a low-dimensional coordinate system that preserves the separation of data points. Here, we try to apply PCA to the 4096 CNN features and see if the principle components would give a good regression performance.

Fine Feature Selection

- Lasso regression automatically selects some features to use, while reducing the coefficients of others to zero.
- Forward-search feature selection starts from an empty feature set, each time adding one single feature that raises the test performance the most. We use ridge regression as the evaluation metric.



Feature Selection

- Correlation
- PCA
- VAE
- Lasso Regression
- Forward Search

Regression Model

- Linear Regression
- Ridge Regression
- Lasso Regression
- XGBoost

Linear Regression

Linear regression a naive model for CNN feature regression, which models the error as normal distributions. We consider linear regression as our baseline for all features and regression models.

$$\text{minimize} \ \ (X\theta - y)^T (X\theta - y)$$

Ridge Regression

Ridge regression is a technique for analyzing multiple regression data that suffers from multicollinearity, which might result in high variances. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. It can also help mitigate overfitting to some extend.

$$\text{minimize} \ \ (X\theta - y)^T (X\theta - y) \ \ \text{s.t.} \sum_{i=1}^{n} \theta_i^2 \leq t$$

XGBoost Regression

XGBoost is an open-source software library which provides the gradient boosting framework for machine learning. XGBoost is widely used in Kaggle competitions and other machine learning objectives due to its scalable, portable and distributed features.

XGBoost

Lasso Regression

Lasso (Least Absolute Shrinkage Selector Operator) regression is similar to ridge regression, but instead of using L2 regularization, Lasso uses L1 regularization, which automatically discards features that are not helpful, and uses only a small portion of the features. For this reason, Lasso is even less susceptible to overfitting than ridge regression.

$$\text{minimize} \ \ (X\theta - y)^T (X\theta - y) \ \ \text{s.t.} \sum_{i=1}^{n} |\theta_i| \leq t$$

Conclusion

By researching into various feature selection methods and regression models, we manage to find sets of CNN features that are able to yield a satisfactory asset prediction. To be specific:

- Forward search provides the most robust set of features for all regression models,
- Correlation feature selection also provides a good generalization to the model,
- Lasso regression provides features that are only helpful to itself,
- PCA and VAE are not valid selection methods for this project.

For the regression models, the R^2 score relies heavily on the features fed into the model. We also find that:

- Lasso regression is the best regressor to reduce overfitting, ridge regression also reduce overfitting to some extent,
- Linear regression is most vulnerable to overfitting,
- XGBoost relies heavily on parameter tunings to yield good result.

[1] Combining satellite imagery and machine learning to predict poverty, N Jean, M Burke, M Xie, WM Davis, et. al - Science, 2016

[2] Transfer learning from deep features for remote sensing and poverty mapping M Xie, N Jean, M Burke, D Lobell, S Ermon - arXiv preprint arXiv:1510.00098, 2015