Finding Poverty in Satellite Images

Neal Jean
Stanford University
Department of Electrical Engineering
nealjean@stanford.edu

Rachel Luo
Stanford University
Department of Electrical Engineering
rsluo@stanford.edu

ABSTRACT

The lack of reliable poverty data in developing countries poses a major challenge for making informed policy decisions and allocating resources effectively in those areas of the world. Unfortunately, it can be prohibitively expensive to frequently conduct comprehensive surveys that track measures of economic progress. A cheap and scalable method of producing poverty maps would greatly facilitate economic progress in these developing countries. In this paper, we present a method for predicting socioeconomic indicators directly from satellite images. We take the output of a convolutional neural network (CNN), a 4,096-element feature vector, and use these image features along with known survey data from certain parts of Uganda and Tanzania to perform linear regression on continuous wealth measures. We then use our models to predict consumption-based wealth measures and asset-based wealth measures for both Uganda and Tanzania. We find that our machine-learning based predictions approach survey accuracy at a fraction of the cost.

Keywords

Remote sensing; poverty mapping; computer vision; convolutional neural networks; cross validation

1. INTRODUCTION

The rise of Big Data in recent years has precipitated an explosion of new applications; from robotics to computer vision, from advertising to movie recommendation algorithms, big data and machine learning are now almost ubiquitous. However, one area that could benefit from further attention is the allocation of resources for third-world countries. One challenge that has been impeding progress in the area of developmental economics is the difficulty of obtaining reliable data in these areas of the world. We hope to improve this state of affairs by accurately predicting poverty levels through the proxies of consumption-based wealth measures and asset-based wealth measures without the use of expensive surveys.

This project builds on work done by the Ermon group in the Computer Science department at Stanford. Previously, Xie *et al.* [1] trained a convolutional neural network (CNN) to predict nighttime light intensities from daytime satellite imagery. This CNN output a 4,096-dimensional feature vector.

However, due to the scarcity of poverty data, we cannot train the CNN directly on poverty. Instead, we take a transfer learning approach with the assumption that the features useful for predicting nighttime light intensities are also useful for predicting poverty levels, and use linear regression to predict poverty measures from satellite images with the feature vector output by the CNN. Thus, our inputs are daytime satellite images and our outputs are continuous wealth measure predictions. Figure 1 below displays the pipeline for our project.

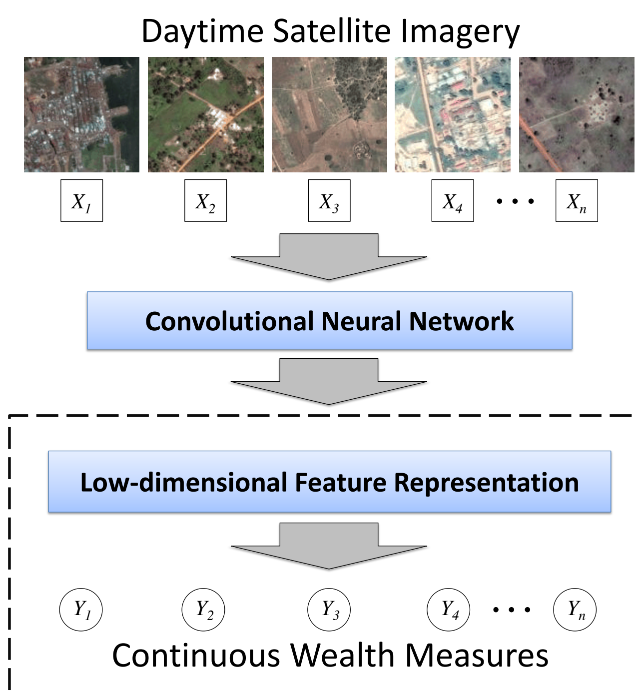

Fig 1. Project pipeline. Focus of our project in dotted box.

2. RELATED WORK

The field of computational social science or computational sustainability is a relatively young one, so opportunities for new applications of machine learning abound. Early work on identifying poverty using remote sensing was presented in 2014 by data scientists from DataKind working in conjunction with the NGO GiveDirectly. Their goal was to identify the poorest villages in western Kenya in order to deliver direct cash transfers to the very poor. In their approach, they used template matching to find roofs in satellite images, then hand-designed conventional computer vision features such as color histograms that would allow them to classify roofs as either metal or thatched [2, 3]. They then used the ratio of metal roofs in each village as a proxy for the amount of poverty present.

A more recent approach to poverty prediction was published in *Science* just a few weeks ago. Blumenstock *et al.* used mobile phone call record data from a large mobile provider in Rwanda to predict asset-based wealth measures nationwide [4]. In addition to using survey data, they conducted a small-scale phone survey to collect personalized wealth data specific to their call record dataset. For their poverty measure, they used a wealth index based on the first principal component of the data collected on survey respondent asset ownership, a method summarized by Filmer and Pritchett in [5].

This project is a continuation of work in Prof. Stefano Ermon's group in the Stanford AI Lab. As described earlier, Xie *et al.* [1] trained a CNN to predict nighttime light intensities from daytime satellite imagery, and then used the learned filters to predict poverty. However, they worked exclusively on the problem of predicting binary poverty labels (classifying each location as either "in-poverty" or "poverty-free"), while we attempt to predict continuous measures of socioeconomic status in this project. In going from a binary classification problem to a continuous wealth regression problem, we hope to provide more useful and fine-grained poverty data for the World Bank and other entities on the ground in Africa.

3. DATASET AND FEATURES

We use satellite images from Google Earth [6] covering all of Uganda and Tanzania to predict poverty in both countries. Our Uganda set includes 641 household clusters, where each cluster contains around 10 households surveyed. Since up to 5 km of jitter is added to each reported cluster location to preserve the anonymity of the survey respondents, we tile each location with 100 images of 1 km^2 each to cover a 10 km by 10 km area. Our Tanzania set includes 400 locations, where each location is subdivided as in the Uganda set. All images are 400 pixels × 400 pixels—at Google Maps zoom level 16, this corresponds to a resolution of approximately 3 meters per pixel. Some examples of our satellite images are shown in the first row of Figure 1.

Our project uses two separately obtained continuous wealth measures: consumption expenditure-based and asset-based. For consumption-based wealth measures, we procure training and test examples from the Living Standards Measurement Study (LSMS). The LSMS is administered by the World Bank and collects consumption expenditure data at the household level [7]. These consumption expenditures represent the amount a household spends in a given period of time, and are generally considered a better measure of living standards than traditional metrics such as income.

For asset-based wealth measures, we employ training and test examples from the Demographic and Health Surveys (DHS), an independent organization that collects health and household asset data in developing countries [8]. The survey includes questions such as whether the household owns a refrigerator, whether it owns a car, and whether it has access to electricity, and these asset-based questions are then used to create a continuous wealth index that we try to predict. The wealth index is calculated by creating a matrix from the answers to the survey questions, taking the first principal component of this matrix, and using a linear combination of its elements [5].

4. METHODS

We use linear regression to predict these wealth measures in Uganda and Tanzania. For both the consumption expenditures and the asset-based wealth measures, we try ordinary least squares (OLS) regression, lasso (L1) regression, and ridge (L2) regression.

OLS regression seeks to minimize the distance between the predicted wealth values and the true wealth values. We use $x^{(i)}$ to denote the input variables (the image features in this case), and $y^{(i)}$ to denote the output variables (the continuous wealth measures). Let h represent our hypothesis, and θ_i represent the unknown parameters mapping from x to y. Also, let m be the number of training examples and n be the number of features (4,096 here). Then $h = \theta^T x$, and OLS chooses θ to minimize the cost function

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

L1 regression is similar to OLS, but it includes a regularization term to prevent overfitting the training set. Thus, it chooses θ and c to minimize

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + c\sum_{i=1}^{n}|\theta_i|$$

This method forces many of the θ_i parameters to zero, effectively finding a sparse feature representation.

L2 regression is also similar to L1 regression, but it uses a different regularization term. The cost function that it minimizes is

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + c\sum_{i=1}^{n}|\theta_i|^2$$

and this regularization encourages the θ_i parameters to be very small.

To train our model and evaluate its performance, we use a 10-fold nested cross-validation scheme for the L1 and L2 regression methods. In the outer loop of this procedure, we divide our training data into 10 equally sized, disjoint sets. We then take 9 of these sets, combine them, and again subdivide this combined set into 10 equally sized, disjoint subsets. In the inner loop of the procedure, we train our data on the subsets using standard 10-fold cross-validation and select a model. We then go back to the outer loop and test it on the 10^{th} (previously unseen) set to get an estimated generalization error. We repeat this process 10 times, leaving out each set once, and average the estimated generalization error from each run.

The prediction metrics we use are the average r^2 values over the 10 folds, with r^2 calculated as

$$r^2 = \frac{\sum_{i=1}^{m}(y - \hat{y})^2}{\sum_{i=1}^{m}(y - \bar{y})^2} = \frac{SSE}{TSS}$$

where SSE is the sum of squared prediction errors and TSS is the total sum of squares, a measure of how much the data varies around its mean. The higher the r^2, the better the model predicts the continuous wealth measure.

5. RESULTS AND DISCUSSION

We run OLS, L1, and L2 regression on the Uganda and Tanzania data separately to train models for each country, and we also use the Uganda model to predict wealth measures in Tanzania to see how well that model generalizes.

5.1 Predicting Consumption

In predicting consumption expenditures, we obtain the results shown in Table 1 below for the Tanzania data and similar results for the Uganda data. By comparing the satellite test r^2 columns, we can see that ridge regression performs the best of the three, approaching survey test accuracy. The choice of regularization hyperparameter is extremely important—since we are using high-dimensional feature vectors and training on relatively small datasets, we must avoid overfitting.

We then use the Uganda model to predict consumption expenditures in Tanzania and plot the average test r^2 for both countries vs. the average Uganda training r^2 to obtain Figure 2. This graph shows where the model begins overfitting for the Uganda data (around r^2 = 0.45), as well as the optimal training r^2 for generalization to Tanzania (around r^2 = 0.3). Based on these results, we believe that computer vision based methods for poverty prediction should generalize well across national borders as long as the visual features of both countries are similar (Uganda and Tanzania are neighboring countries in sub-Saharan Africa, with comparable geographies and socioeconomic conditions).

	OLS		Lasso (L1)		Ridge (L2)	
Features	Survey	Satellite	Survey	Satellite	Survey	Satellite
Training r^2	0.407	1.000	0.403	0.383	0.396	0.461
Test r^2	0.327	-1.007	0.339	0.187	0.331	0.296

Table 1. Linear regression models predicting consumption expenditures in Tanzania.

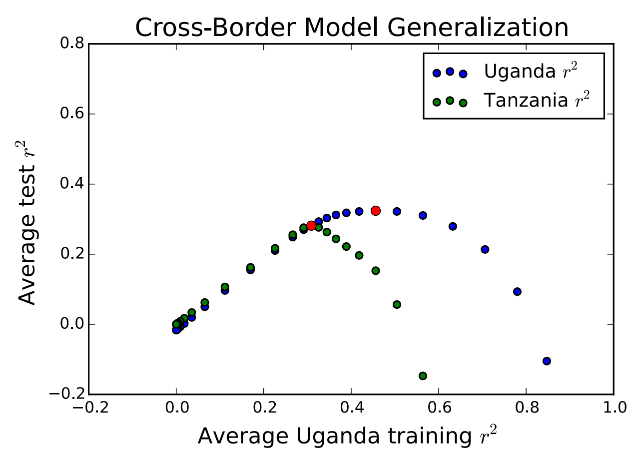

Fig. 2. Average test r^2 vs. average Uganda training r^2. Optimal Uganda training r^2 for both Uganda and Tanzania models shown in orange.

5.2 Urban/rural divide in predicting consumption

We then examine whether our model's performance shows significant discrepancies between urban and rural areas. Our first approach is to perform ridge regression for all of the combined data, and then to look at the urban and rural households separately. With this approach, we obtain the results shown in the "combined" row of Table 2 and in Figure 3. This figure and table come from the Uganda data set, but both countries yield similar results. Clearly, our model does a better job discriminating between households in rural areas than in urban areas. In fact, in the urban household group, we would attain better results simply by predicting the mean urban expenditure rather than using our model.

	Urban		Rural		Overall	
r^2	Training	Test	Training	Test	Training	Test
Combined	0.043	-0.227	0.324	0.145	0.442	0.303
Separate	0.439	0.303	0.294	0.212	-	-

Table 2. Combined vs. separate urban and rural regression models.

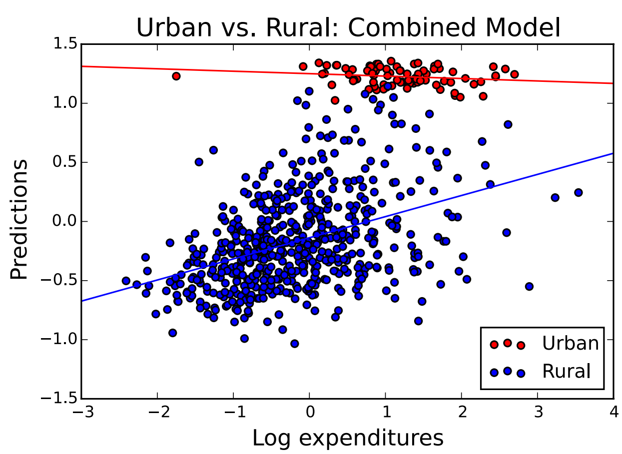

Fig. 3. Uganda combined model with all data points. Urban points are shown in red and rural points are shown in blue. The urban and rural lines look very dissimilar.

Our second approach involves pre-processing the data to separate out the urban and rural points, and then applying ridge regression to these two sets separately to build two distinct models. We obtain the results shown in Figure 4 and in the "separate" row of Table 2 (again from the Uganda data set). This approach significantly improves the r^2 values of both the urban and rural sets.

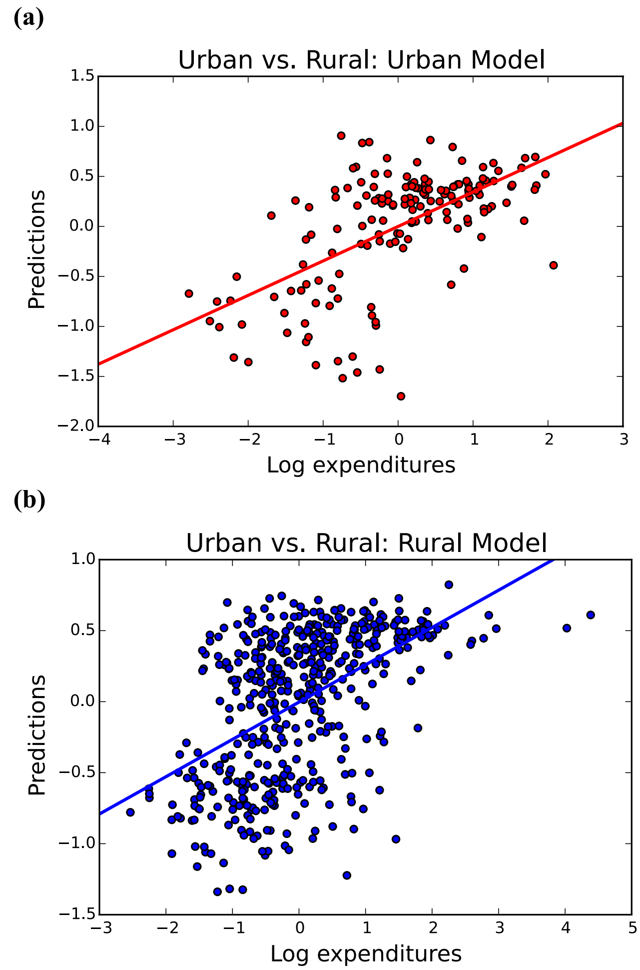

Fig. 4. (a) Uganda urban points with best-fit line; **(b)** Uganda rural points with best-fit line.

5.3 Predicting Assets

Finally, we run OLS, L1, and L2 regression on the asset-based wealth measures, with the results shown in Table 3 for Uganda, (Tanzania looks reasonably similar). Figure 5 plots our predicted wealth index against the true wealth index; the goal is to obtain a linear relationship between the two with a slope of 1, such that our predicted values map to the true values. Looking at the plot, we do in fact find a linear trend with $r^2 = 0.653$; this improves upon the results published in *Science* in [4], which found a trend with $r^2 = 0.46$ using mobile phone record data. Our hypothesis is that asset-based wealth measures are easier to predict using satellite images than consumption-based measures—assets can actually be seen in some cases, while consumption cannot.

	OLS		Lasso (L1)		Ridge (L2)	
r^2	Training	Test	Training	Test	Training	Test
Our model	0.978	0.000	0.676	0.634	0.795	0.653

Table 3. Linear regression models predicting asset-based wealth in Uganda

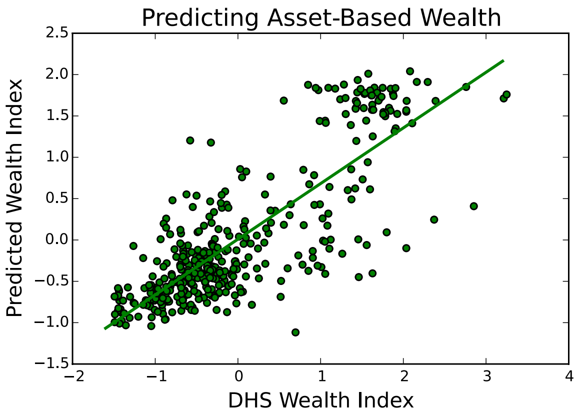

Fig. 5. Our predicted wealth index values vs. the true wealth index values in Uganda.

6. CONCLUSION

Our predictions for consumption expenditures approach survey accuracy, while our predictions for asset-based wealth measures improve upon recent state-of-the-art results. Additionally, our poverty estimation techniques based on remote sensing and machine learning can be applied on a global scale at a fraction of the cost of traditional survey-based methods. Cheaply and accurately predicting poverty in developing countries can help governments and non-profits better allocate their resources and create effective policies. In the future, we hope to extend this work by applying these techniques to other countries and creating poverty maps with greater geographic coverage. In addition, we hypothesize that poverty dynamics include predictable temporal and spatial relationships. We hope to obtain satellite images from different points in time and then create a probabilistic graphical model that can take into account poverty relationships between points that are neighbors in both time and space.

7. REFERENCES

[1] Xie, M., Jean, N., and Ermon, S. 2015. Transfer learning from deep features for remote sensing and poverty mapping. *CoRR*. http://arxiv.org/abs/1510.00098.

[2] Abelson, B., Varshney, K. R., and Sun, J. 2014. Targeting direct cash transfers to the extremely poor. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.* 1563-1572. DOI= 10.1145/2623330.2623335.

[3] Varshney, K. R., Chen, G. H., Abelson, B., et al. 2015. Targeting villages for rural development using satellite image analysis. *Mary Ann Liebert, Inc.* 3, 1 (March 2015), 41-53. DOI= 10.1089/big.2014.0061.

[4] Blumenstock, J., Cadmuro, G., and On, R. 2015. Predicting poverty and wealth from mobile phone metadata. *Science*. 350, 6264 (Nov. 2015), 1073-1076. DOI= 10.1126/science.aac4420.

[5] Filmer, D. and Pritchett, L. H. 2001. Estimating wealth effects without expenditure data – or tears: an application to educational enrollments in states of India. *Demography*. 38, 1 (Feb. 2001), 115-132. DOI= 10.1353/dem.2001.0003.

[6] Google Static Maps API. Tanzania, Uganda. November 2015.

[7] World Bank. 2015. Living standards measurement study (various) [datasets].

[8] ICF International. 2015. Demographic and health surveys (various) [datasets].