Fairly Estimating Socioeconomic Status Under Costly Feature Acquisition

Anonymous authors
Paper under double-blind review

Abstract

Predictive models have become increasingly ubiquitous in our society. However, concern has been expressed on their ability to perpetuate inequality amongst subpopulations. Active feature-value acquisition has been suggested as a method of promoting both individual and group notions of fairness in a predictive model. In this work, we seek to use such active framework to create a predictive socioeconomic model. At the same time, satellite imagery has been utilized as a method of socioeconomic estimation. In this work, our goal is to integrate satellite imagery with an active framework to create a fair predictive socioeconomic model. This was tested on one real-world dataset. Results indicate an increase in accuracy resulting from the aggregation of the satellite imagery.

1 Introduction

Many organizations use data on economic livelihood to direct efforts to those in need of aid, including in sub-Saharan Africa (Varshney et al., 2015; Brass et al., 2018). However, obtaining such data at the household level is time-consuming, costly and logistically difficult. These costs have led to an increased reliance on predictive machine learning models for estimating socioeconomic status in practice. The use of machine learning raises concern for its ability to perpetuate inequity amongst subpopulations, resulting in a situation where one or more subpopulation is favored by the decision maker over the other (Noriega-Campero et al., 2020). This could potentially exclude certain subpopulations from receiving adequate financial assistance relative to their counterparts. Concern for bias in predictive models has resulted in a number of proposed definitions for fairness which can be classified into two broad categories: group fairness and individual fairness. *Group* or *statistical* definitions of fairness focus on balancing classification errors across subgroups to achieve equal error rates (*overall accuracy equality*), balanced false positive error rates (*predictive equality*), or equivalent accuracy for all subgroups (*conditional accuracy equality*). To the contrary, *individual* fairness requires that all individuals be treated similarly, utilizing a predefined distance function to measure similarity (Verma & Rubin, 2018). However, individual notions of fairness can be challenging to implement, and at times, even at odds with group fairness notions (Bakker et al., 2020).

At the same time, recent works have pushed to provide a framework to model a realistic situation, where information (features) can be acquired incrementally at incremental costs. Consider a doctor seeking to diagnose a patient. The doctor will begin by performing preliminary tests. Based on the results, the doctor would subsequently perform more tests to confirm the patient's condition. However, such tests can be costly, so the doctor must carefully select what information would most benefit his understanding of the patient's condition. Noriega-Campero et al. (2019) provide a similar setting, known as *active feature acquisition* (AFA), where a decision maker is not limited to the data provided, but rather can acquire features at a cost while also mitigating individual and group unfairness. However, inquiries must be limited because additional data is costly to acquire. Bakker et al. (2020) build on this framework by deriving a set of stopping criteria where an individual will be classified if sufficient features have been acquired to attain a certain level of confidence. This method also considers an individual's privacy by acquiring the smallest number of features. Both Noriega-Campero et al. (2019) and Bakker et al. (2020) only incorporate structured tabular features in their work.

In this work, we seek to build on the framework provided by Bakker et al. (2020) by incorporating satellite imagery as an additional non-tabular feature in a predictive socioeconomic model for distributing aid in sub-Saharan African countries such as Nigeria. High resolution satellite imagery has become widely available and relatively inexpensive, and past works have used it as a method of poverty detection (Varshney et al., 2015; Jean et al., 2016). From satellite imagery, key features have been correlated with economic development such as roads, large-scale buildings, and electrical poles can be used identify household socioeconomic status. Using deep learning techniques, such features can be extracted from images and used to estimate a household's socioeconomic status (Perez et al., 2017).

It is fairly well-understood that wealth estimation from satellite imagery and other household features is easier in rural areas of sub-Saharan Africa than urban areas because of greater visibility of features in lower density populations and greater correlation between household characteristics and wealth. Therefore, a machine learning-based method for targeting aid may be biased towards rural areas at the expense of urban areas. Acquiring household data through surveys is less costly in urban areas, however, because smaller distances must be traversed by survey takers. In this work, taking both cost and fairness as practical considerations, and building upon the work of Bakker et al. (2020), our primary contribution is to take preliminary steps towards developing a novel active feature acquisition model utilizing both satellite image and household survey features to predict wealth in sub-Saharan Africa. We demonstrate the promise of the model and approach on Demographic and Health Survey (DHS) data from Nigeria.

2 MATERIALS AND METHODS

2.1 DATASET

This work used the 2018 Nigerian DHS,[1] which contains data for 40,127 residences in Nigeria. Each residence pertains to one of five wealth index values: *poorest*, *poorer*, *middle*, *richer*, and *richest*, where poorest corresponds to 1 and richest corresponds to 5. This is the response variable to be predicted. In an aid allocation setting, one would only need to consider a binary response variable; therefore we binarized the wealth index as follows: *poorest* and *poorer* were categorized as “poor” or 0, while *middle*, *richer* and *richest* were categorized as “not poor” or 1. From the dataset, 27 household features were selected to create the model. Examples of such features include distance to the nearest source of water, material household's floor, and the number of rooms in a household. The Google Maps API was used to obtain 640 pixel by 640 pixel daytime satellite images using the latitude and longitude coordinates obtained from the DHS data.

2.2 ACTIVE FEATURE ACQUISITION

Following the framework provided by Bakker et al. (2020), let $(x^{(h)}, y^{(h)}) \sim P$ represent a household h where x is a d-dimensional set of features and $y^{(i)} \in \{0, 1\}$. We begin with an empty set denoted by $\mathcal{O}_0 := \emptyset$ at time t_0. Subsequently, a subset of features is selected from the set of unselected features such that $S_t^{(h)} \subseteq \{1, \ldots, d\} \setminus \mathcal{O}_{t-1}^{(h)}$. The feature will then be added to the set of acquired features $\mathcal{O}_t^{(h)} := S_t^{(h)} \bigcup \mathcal{O}_{t-1}^{(h)}$ which will be accessible to the classifier. The classifier is capable of handling partial features sets. Features will be acquired, until the stopping criterion is reached at time $T^{(h)}$. The set of features acquired can vary for different households.

2.3 CLASSIFIERS

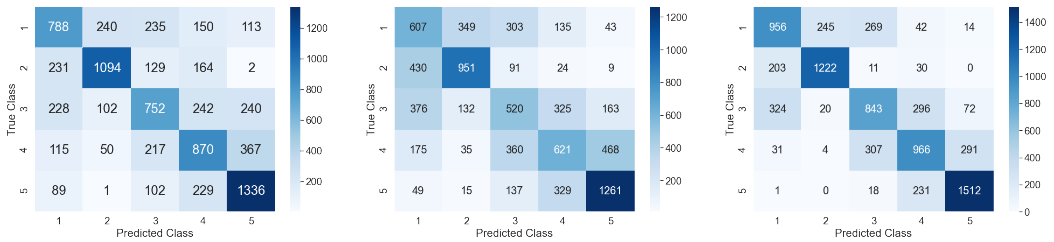
The household attributes from the DHS data was converted into feature vectors. A random forest classification was performed on this data. An 80/20 split was used to obtain the training and testing data. All random forests were created using 100 trees and a maximum of 150 leaf nodes.

Anaysis of the satellite imagery used a convolutional neural network (CNN) defined by a series of convolutional layers, where the output of one layer becomes the input of the next layer. Images were separated into RGB colored bands. The CNN consisted of a 2x2 kernel with a 1x1 stride and Rectified Linear Unit (ReLU) activation.

[1] https://dhsprogram.com/data/

Table 1: Accuracy results.

Feature Set	Classes	Overall	Urban	Rural
Household	5	0.60	0.51	0.66
CNN	5	0.50	0.49	0.52
Household + CNN	5	0.69	0.64	0.72
Household	2	0.83	0.75	0.89
CNN	2	0.82	0.79	0.85
Household + CNN	2	0.91	0.88	0.93



(a) (b) (c)

Figure 1: Confusion matrices for multiclass classification using (a) only household features, (b) only CNN features, and (c) both household and CNN features.

One feature in the d-dimensional dataset corresponds to a satellite image feature vector in the CNN features dataset $I^{(h)}$. At time t, as the satellite feature vector will be acquired from the CNN features dataset and aggregated to the partial dataset $\mathcal{O}_t^{(h)} := I_t^{(h)} \bigcup \mathcal{O}_{t-1}^{(h)}$.

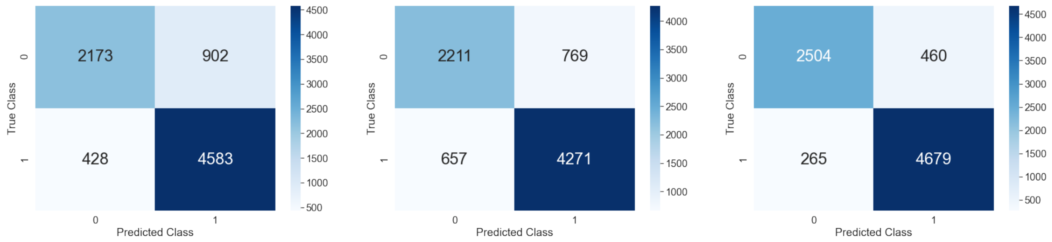
3 RESULTS

The first row of Table 1 provides the accuracy rates for the multiclass classification for the household features overall along with the accuracy rates for the urban and rural households. A disparity in accuracy is present between the urban and rural households. Figure 1(a) provides the confusion matrix of the classification.

The accuracy rates with the CNN features are included in the second row of Table 1. They were then combined with the household dataset, which resulted in an increase of accuracy in the third row of the table. The disparity between urban and rural accuracy was mitigated with the aggregation of the CNN features. Figure 1(b) shows the confusion matrix of the CNN features and Figure 1(c) shows the confusion matrix of the CNN and household features.

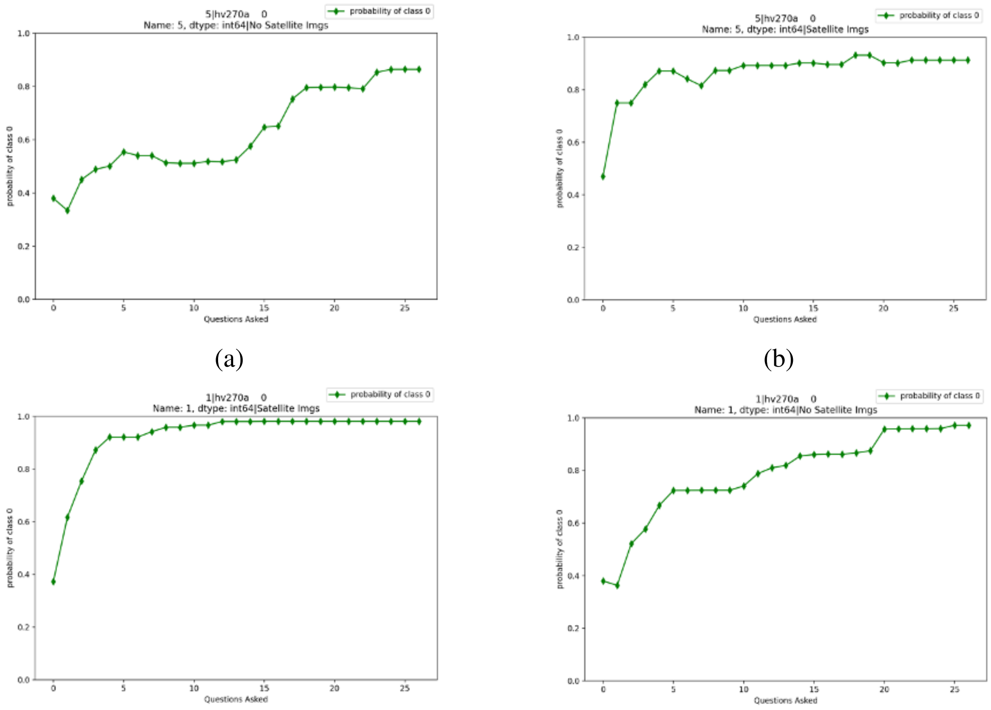
Another set of results were generated for a binary classification. The fourth row of Table 1 provides the overall accuracy for the base model. Similar to the multiclass model, a disparity between urban and rural households is prevalent. The confusion matrix of the classification is presented in Figure 2(a). The last two rows of Table 1 contain the accuracy for the CNN features, the CNN and household features, and the disaggregated urban and rural subpopulations. Figure 2(b) provides the confusion matrix for the CNN features and Figure 2(c) provides the confusion matrix for the CNN and household features.

Figures 3(a) and 3(b) provide an example of one of the households included in the survey. Subsequently, as more features are acquired (ordered optimally), the model is more certain of which class it pertains to. The graph presented in Figure 3(a) does not include satellite imagery, while the graph presented in Figure 3(b) does. Figures 3(c) and 3(d) present a different household, where the graph in Figure 3(c) does not include satellite imagery and the graph presented in Figure 3(d) does. The feature acquisition can be stopped at various points along the curve for different households to enable predictive equity, whose in-depth study will be our future work. Additional future work will attempt to include realistic costs for acquiring various types of features in rural and urban areas.



(a) (b) (c)

Figure 2: Confusion matrices for binary classification using (a) only household features, (b) only CNN features, and (c) both household and CNN features.



(a) (b)

(c) (d)

Figure 3: Probability of class 0 (poor) as a function of features acquired for (a, c) household features only, and (b, d) household + CNN features. The plots (a) and (b) correspond to one sample household and the plots (c) and (d) correspond to a different sample household.

4 Conclusion

This study attempted to keep feature acquisition costs down and mitigate disparity between urban and rural households in a socioeconomic predictive model using active feature acquisition. AFA allows for the creation of realistic predictive model applications, where a model is not limited to the data originally provided but rather can be acquired at a cost. In this specific model, one of the features acquired was low-cost satellite imagery. This model can assist in a more equitable distribution of funds in countries in sub-Saharan Africa and other regions of the Global South.

REFERENCES

Michiel Bakker, Humberto Riverón Valdés, Duy Patrick Tu, Krishna P. Gummadi, Kush R. Varshney, Adrian Weller, and Alex 'Sandy' Pentland. Fair enough: Improving fairness in budget-constrained decision making using confidence thresholds. In *Proceedings of the AAAI Workshop on Artificial Intelligence Safety*, pp. 41–53, February 2020.

Jennifer N. Brass, Wesley Longhofer, Rachel S. Robinson, and Allison Schnable. NGOs and international development: A review of thirty-five years of scholarship. *World Development*, 112: 136–149, December 2018.

Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B. Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, August 2016.

Alejandro Noriega-Campero, Michiel A. Bakker, Bernardo Garcia-Bulle, and Alex 'Sandy' Pentland. Active fairness in algorithmic decision making. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 77–83, February 2019.

Alejandro Noriega-Campero, Bernardo Garcia-Bulle, Luis Fernando Cantu, Michiel A. Bakker, Luis Tejerina, and Alex Pentland. Algorithmic targeting of social policies: Fairness, accuracy, and distributed governance. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, pp. 241–251, January 2020.

Anthony Perez, Christopher Yeh, George Azzari, Marshall Burke, David Lobell, and Stefano Ermon. Poverty prediction with public Landsat 7 satellite imagery and machine learning. arXiv:1711.03654, 2017.

Kush R. Varshney, George H. Chen, Brian Abelson, Kendall Nowocin, Vivek Sakhrani, Ling Xu, and Brian L. Spatocco. Targeting villages for rural development using satellite image analysis. *Big Data*, 3(1):41–53, March 2015.

Sahil Verma and Julia Rubin. Fairness definitions explained. In *Proceedings of the IEEE/ACM International Workshop on Software Fairness*, pp. 1–7, May 2018.