

Study of Poverty Prediction using Remote Sensing Data

¹Rekha S. Thorat, ²Dr. Praveen Shetiye, ³Dr. Avinash K. Gulve

¹Student, ²Associate Professor, ³Associate Professor
Computer Science and Engineering,
Government college of Engineering, Aurangabad, India

Abstract: In this paper, a study on poverty prediction using remote sensing data is done. Remote sensing method of prediction poverty is an efficient method in terms of time consumption, cost, and effort required than the household survey method. For this approach various machine learning methods like classification, regression, clustering and dimension reduction are used to train the model. In the training phase, household survey data is used. This GDP, school enrollment, CO2 emissions, poverty headcount ratio, life expectancy at birth, GNI per capita and census data is made freely available for research purpose by World Bank Group in the form of statistics, and DHS (demographic health survey) data is available on DHS program's site. Satellite daytime and nighttime data can be taken from public and private domains of the satellite. After data collection and execution of model with machine learning methods, various results are computed with different and maximum accuracy.

Index Terms: Remote sensing data, Machine Learning, DHS data.

I. INTRODUCTION

From so many years poverty is serious problem till now. Billions of people are suffered from it for food, shelter, health etc. Income gap between developed and undeveloped regions always remains large. Government conducts various programs, creates various policies to recover from poverty and fill the gap between various regions according to development. But those facilities and policies of government did not rich to too many peoples due to various reasons like corruption, accurate information of poverty.

Various methods used by government agency to reduce poverty and to find undeveloped regions.

Earlier gross domestic product (GDP) was used to track socioeconomic activities. But it had many limitations as it does not differentiate costs from benefit and not distinguishes sustainable practice from unsustainable [1].

Later traditional household survey method comes, which collects data by reaching each door. This method is effective till now. Only the limitation of this is it is expensive as more personnel is required, more efforts required, and it is more time consuming process.

The solution to these 4 shortcomings is that we can combine remote sensing data with strong machine learning algorithms to predict poverty, reduction in poverty and development analysis.

Aim and Objectives

The aim of this study is to find another way of mapping poverty using satellite images and machine learning. This technique can predict the poverty from only satellite images using machine learning model, which is trained using census data from household survey. And this early prediction without household data makes profitable to government to make policy for poverty area. Also the artificial intelligence is evolving rapidly day by day, so it will give more accuracy for this research area.

The objective of this study to fulfill following 2 steps:

- Find correlation between household survey and satellite image features, and accordingly train the machine learning model.
- Apply trained model on input satellite image, and map the poverty area on it with optimum accuracy [2].

Scope

- There is rapid progress in use of remote sensing data in various fields for various applications, as it easily available with minimum cost, less efforts and time.
- Advancement in artificial intelligence and revolving strong algorithms.
- Increasing demand of policymaker to get frequent estimation of areas on the basis of economics [3].

These things make scope for this study in future.

II. DATA COLLECTION

Public domain

United states make satellite imagery data freely available to scientific use [4].

Landsat

Optical Landsat imagery provides 30 m resolution .It is oldest program to observe the earth. It started with Landsat5, now Landsat7 and Landsat8 are currently working and Landsat9 is in progress.

Modis

Since 2000 it captures daily satellite images of earth and uses 36 spectral bands.

Sentinel

ESA is on the way to develop the sentinel constellation of satellite.

It has planned for 7 missions, already launched sentinel 1, sentinel 2 and sentinel 3.

Aster

Terra a satellite launched in 1999 in December 18, at Vandenberg Air Force Base, California in USA. And Aster is sensor which captures 15 meter, 14 band multispectral resolution images.

Meteostat

It is high impact whether satellite. The first Meteostat satellite was launched in 1977 on 23 November and operated until 1979.

Private domain

Private companies also launched and maintained their satellites. But there images are not freely available, to access them user has to be paid. Some of these domains are GeoEye, Spot Image, DigitalGlobe, BlackBridge, ImageSat International etc.

Other Data

With satellite data census data can be combined to train the machine learning model. In supervised learning approach household survey data is combined with remote sensing data, and correlation between them is calculated. Once model is trained only remote sensing data can be used for prediction.

Census Data

Census is a scientific method that every country uses it to collect population data. Census data can show the condition of state, nation or world. For resource management, policy making and for other programs, population data plays great role. There are many websites, web-applications which provides census data, one of these site is CensusReporter.org. This is largest and best free source of census data in U.S. under of Census Bureau’s American Fact Finder [5].

World Bank Data

World Bank data is provided by World Bank’s group, and data contains the economic statistics of nation. World Bank provides data in the form of statistics, graphs, analysis etc. and data about GDP, population, school enrollment, CO2 emissions, poverty headcount ratio, life expectancy at birth, GNI per capita and statistical capacity score [world bank].

DHS Data

DHS (demographic health survey) is a program which collects data about population’s health and facilities provided for their health in developing countries. Data can be taken from dhsprogram.com site of DHS program [6].

III. METHOD FOR SATELLITE IMAGE PROCESSING

There are various methods in machine learning which can applied on satellite imagery and census data. Machine learning methods are useful to find statistical relationship between different quantities and to find statistical formulation for prediction [7].

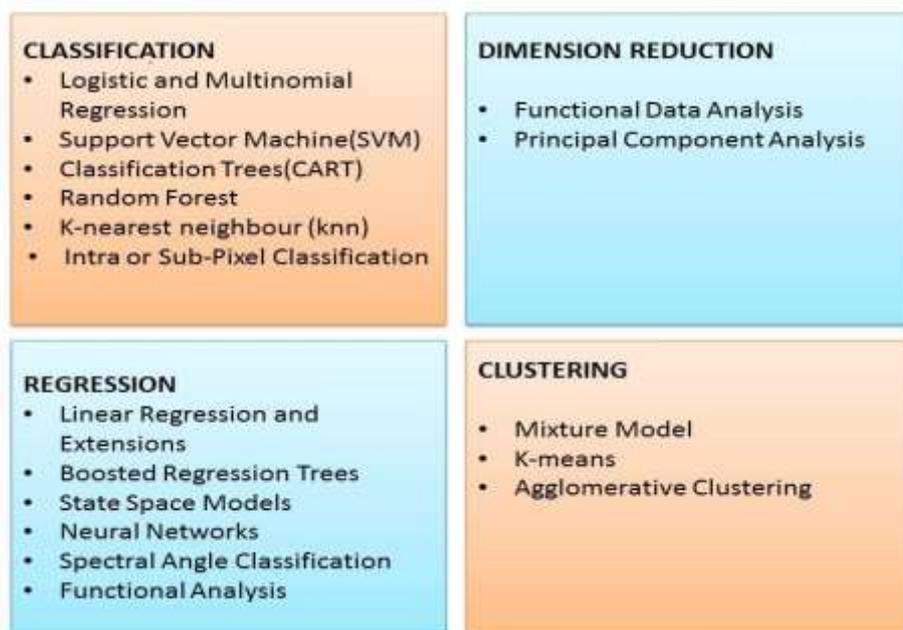


Fig 1: Machine Learning Methods

Classification

Classification is based on predicting certain output based on given input. Categorize specified data into a given number of classes this is main aim of classification. Classification is a systematic approach to build classification modules from input data. Classification may use mathematical concepts such as decision tree, linear programming, Statistics. There different algorithm are available for performing classifications. Such as support vector algorithm, K-nearest neighbor, Bayesian algorithm. Classification real time application like fraud detection, weather predation .Classification process involves two phases, Learning phase and Classification phase. In learning phase training data analysis and in classification phase apply classification algorithm to produce the unknown

results. After training the model i.e.-neural network, can be used to classify microscopic images such as cancerous images to macroscopic images like remote sensing data [8].

Clustering

Clustering is process of partitioning a set of data into meaning full set of sub class called as clusters. Clustering is set of grouping objects in such way as object in same group are group together. i.e.-when doing clustering or cluster analysis firstly partitioning the data into groups based on similarity find between them and analysis of different attributes in data. In clustering algorithm directly apply on row data to find clusters of that data. There are different applications of clustering such as market research, search engines, Pattern recognition. There are some advantages of clustering technique such as Scalability. It can work with different kinds of attributes, Interpretability. Similarly for the poverty predation clustering play a very important roll to cluster poverty area on basis of the censors data and attributes map on it.

Dimensional Reduction

In Classification or clustering they basically work on different factors on which classification or clustering is done. These factors are basically different features. More the number of features harder to work on training set in classification or clustering. In that case dimensionality reduction plays and very important roll. In Dimensionality redaction process readied the numbers of variables or features under consideration, by obtaining a set of principal variables. It can be divided into feature selection and feature extraction. In Feature selection process finding the subset of original set of variables which can use to module the problem. In feature extraction process convert the data into high dimensional space to low dimensional space. There are different modules of feature selection like Principal Component Analysis, Linear Discriminant Analysis. There are some advantages of dimensionality reduction such it helps in data compression, and hence reduced storage space, it reduces computation time, it also helps remove redundant features, if any. In poverty prediction process PCA uses to reduce features dimensionality.

Regression

Regression in machine learning is supervised leaning process, where relation between variables is calculated. There are different kind of module have different kind of relationship for dependent and independent kind of variables. When variable X is not directly related to variable Y then it is non-linear regression. In this method relation between those variables find by mathematical equation.

IV. RELATED WORK

The Machine learning on satellite images is become in demand topic today [9]. Early work done on ‘poverty prediction using remote sensing’ in 2014 is by data scientists from DataKind working concurrently with the NGO GiveDirectly. Their aim was to find poorest village in western Kenya, to direct transfer money to the poor people of village. They use template matching method for identifying roofs whether thatched roof or metal roof from satellite images and used hand-designed conventional computer vision feature to classify thatched roofs and metal roofs. After classifying calculated metal roof ratio. Result of village with less ratio will be poor village of western Kenya [10].

In 2007 by Christopher D. Elvidge and team presented a concept of whether satellite data can predict poverty. For this research disaggregated population data is taken from Landscan. Origin of LandScan is the U.S. Department of Oak, Energy Ridge National Laboratory. This LandScan data include input from three satellite data sources that is NASA Modis landcover data, Shuttle Radar Topography Mission’s topographic data and human settlement outline from high resolution imagery. After this DMSP’s (Defense Meteorological Satellite Program of U.S.) night time satellite data is used with different radiances and tried to find any correlation between population count verses light intensity. After execution on three areas, Chicago, Bangkok and Lagos it is observed that there is dimmer or no light in poverty area. This observation then opens the door for more research on this study area [11].

After this work many researchers did work on this study area. And most recent work done using additional information of mobile data is done by Blumenstock et al. They used additional data with remote sensing and household survey data is mobile phone call record data. Large mobile provider from Rwanda gives Blumenstock a call record dataset to calculate asset-based wealth index from whole country [12].

In 2015 Bailang Yu, is a member of IEEE worked on poverty evaluation using remote sensing data. They used nighttime light composite data of china from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day–Night Band (DNB) .which is taken from the Suomi National Polar-orbiting Partnership (NPP) Satellite. In their study ten socioeconomic variables involved with integrated poverty index and then calculated the correlation with Average Light Index (ALI). ALI is computed from night time satellite image and correlation between IPI and ALI with coefficient of determination (R²) of 0.8554, is evaluated using linear regression method [13].

Later the students of Indian Institute of Technology Ropar, focused on this research area for estimation of poverty in India. For this work they used fully convolution deep neural network for concurrently predicting the material of roof, from where people get drinking water and lightning condition of area. Pre-trained model of deep neural network is used to extract features from satellite image. Census data of year 2011 is used and correlated with extracted features of image. And satellite image is requested from Google Static Map API, of size 1920 × 1920, at zoom level 16. After execution with multitasking deep neural network optimum accuracy is obtained by team [14].

V. CONCLUSION

This paper has reviewed that there is efficient way to predict poverty using remote sensing data, without going through the earlier time consuming, costly procedure. Also this is an era of machine learning, and intelligent algorithms are developing day by day. Additionally household survey data, census data, DHS data and World Bank data is also freely available for research purpose which

is used to train the model by various machine learning methods like classification, regression, clustering and dimension reduction. From the previous work it can be seen that there exists correlation between remote sensing data and poverty, also the prediction of poverty using reviewed methods has given better result with maximum accuracy.

REFERENCES

- [1] Bailang Yu et al., "Poverty Evaluation Using NPP-VIIRS Nighttime Light Composite Data at the County Level in China", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 3, pp. 1217-1229, 2015.
- [2] Noorita Sahrman et al., "Urban Poverty Area Identification Using High Resolution Satellite Imagery: A Preliminary Correlation Study", 2013 IEEE International Conference on Control System, Computing and Engineering, pp. 430-434, 29 Nov. - 1 Dec. 2013, Penang, Malaysia.
- [3] The World Bank Data of all country. Available online: <https://data.worldbank.org/country/>.
- [4] VIIRS Day/Night Band Nighttime Lights from satellite of NASA. Available online: https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html
- [5] P.C. Sutton, D. Roberts, C. Elvidge and K. Baugh, "Census from heaven: an estimate of the global population using nighttime satellite imagery," International Journal of Remote Sensing, vol. 22, pp. 3061-3076, 2001
- [6] ICF. The DHS Program. Available online: <https://dhsprogram.com/data/>
- [7] Jacinta Holloway and Kerrie Mengersen, "Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review", Remote Sens. 2018, doi:10.3390/rs10091365.
- [8] P. C. Shetiye, et al., "Detection of Breast Cancer Using Electrical Impedance and RBF Neural Network", International Journal of Information and Electronics Engineering, Vol. 5, No. 5, September 2015
- [9] 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>.
- [10] Abelson, B., Varshney, K. R., and Sun, J. "Targeting direct cash transfers to the extremely poor", 2014 Proceedings of the 20th ACM SIGKDD
- [11] Christopher D. Elvidge, et al. "Can Poverty Rates Be Estimated Using Satellite Data?", 2007 Urban Remote Sensing Joint Event, 2007.
- [12] Blumenstock, J., Cadmuro, G., et al., "Predicting poverty and wealth from mobile phone metadata". Science. 350, 6264 (Nov. 2015), 1073- 1076. DOI= 10.1126/science.aac4420.
- [13] Xizhi Zhao, et al., "Estimation of Poverty Using Random Forest Regression with Multi-Source Data: A Case Study in Bangladesh", Remote Sens. 2019, doi: 10.3390/rs11040375
- [14] Shailesh M. Pandey, Tushar Agarwal and Narayanan C. Krishnan, "Multi-Task Deep Learning for Predicting Poverty from Satellite Images", The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18), pp.7793-7798.