Describing Patterns of Socio-Economic Development at Fine Spatial and Temporal Resolutions

Chahat Bansal
Ph.D. Scholar at School of Information Technology
Indian Institute of Technology-Delhi, India
chahat.bansal@cse.iitd.ac.in

ABSTRACT

Our vision is to describe, at fine spatial and temporal scales, the nuances of socio-economic development taking place as measured through a mix of data sources including censuses and surveys, satellite data, agricultural commodity prices, and qualitative data from mass media and participatory media networks. Towards this goal, we intend to build a system that automatically pools data together from diverse data sources, detects noteworthy facts and trends from the data, describes them automatically in natural language, and surveys a large community of users to probe the observations in more detail.

KEYWORDS

machine learning, census, satellite imagery, socio-economic development, mass-media, social-media

ACM Reference Format:
Chahat Bansal. 2020. Describing Patterns of Socio-Economic Development at Fine Spatial and Temporal Resolutions. In *7th ACM IKDD CoDS and 25th COMAD (CoDS COMAD 2020), January 5–7, 2020, Hyderabad, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3371158.3371214

1 RESEARCH METHODOLOGY

Our goal is to inform people accurately about the state of economic growth and social development happening in different parts of the world, to generate new evidence to settle some long-standing debates about the type of development outcomes produced by different kinds of economic policies [7]. The system we propose has multiple parts: It will automatically pool data from diverse data sources, detect noteworthy facts and trends from the data, describe them automatically in natural language, and survey a large community of users to probe the observations in more detail. Figure 1 presents the bird-eye view of the proposed idea.

We will build a system initially for the Indian context to facilitate a collection of variables obtained from different data sources like NSSO surveys, population and economic census, nightlights and daytime satellite imagery, social media, mass media, and other participatory media streams available at an ongoing basis. Out of

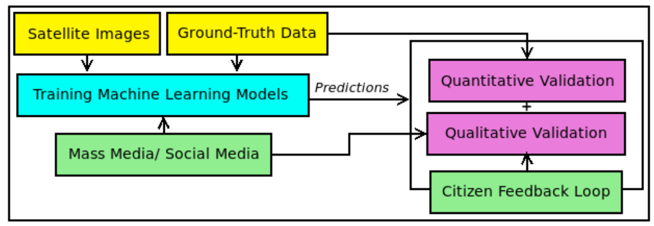

Figure 1: Overall proposal summary

these data sources, the census and survey data have the least temporal granularity but contain the ground truth of socio-economic indicators on a cross-sectional basis. Our strategy will be to build machine learning/deep learning-based models that can use more frequently available satellite data to serve as a proxy to the census and survey-based socio-economic indicators. We have already built models at the village level, which can use data from the Landsat and Sentinel satellite systems to yield a few socio-economic indicators at accuracy of almost 80%. We will attempt to improve this with more data from social/mass media, along with continued analysis to understand better the spatial and temporal granularities at which these mappings work well. We have also collected news articles published since 2011 by seven leading English national newspapers to analyze this data to understand if certain events reported in the news could have led to different economic and development growth noticed from the socio-economic indicators yielded by the satellite imagery-based models above. Most studies on event detection from time-series data have operated either on only socio-economic data or on only social/mass media data, but not together [1–5]. Models we will build to operate jointly on the two datasets may help even to generate explanations about the nature of the events spotted in the data. Finally, a citizen feedback loop will be developed to further investigate such interesting observations by directly contacting people to answer customized surveys.

2 CONCLUSIONS

We leverage the advances made in big-data analytics and apply them to the field of socio-economic development with a goal to create a more informed citizenry that can interpret the changes of different policy approaches. We are aware of some pitfalls in such research. Social media and mass media is susceptible to bias and misuse, which can lead to incorrect indicators. Similarly, machine learning models built on satellite and census data can reflect the errors and biases that might be present in the census data [6]. While we will attempt to develop methods that can work on noisy data, our overarching strategy to address such concerns will however be to involve the citizens in the system itself so that this system can be community driven just like Wikipedia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CoDS COMAD 2020, January 5–7, 2020, Hyderabad, India
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7738-6/20/01…$15.00
https://doi.org/10.1145/3371158.3371214

REFERENCES

[1] Scott R Baker, Nicholas Bloom, and Steven J Davis. 2016. Measuring economic policy uncertainty. *The quarterly journal of economics* 131, 4 (2016), 1593–1636.

[2] Sunandan Chakraborty, Ashwin Venkataraman, Srikanth Jagabathula, and Lakshminarayanan Subramanian. 2016. Predicting socio-economic indicators using news events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1455–1464.

[3] J Vernon Henderson, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American economic review* 102, 2 (2012), 994–1028.

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

[5] Alejandro Llorente Pinto, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. 2015. Social Media Fingerprints of Unemployment. (2015).

[6] Cathy O'neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

[7] Amitendu Palit. 2013. Indian economy, Amartya Sen and Jagdish Bhagwati. *ISAS Brief* 290 (2013), 26.