Responsible use of big data for development

Sriganesh Lokanathan

Global Affairs, Ottawa

11th March 2016

LIRNEasia

Pro-poor. Pro-market.

IDRC CRDI

Canada

This work was carried out with the aid of a grant from the International Development Research Centre, Canada and the Department for International Development UK..

UKaid

from the British people

Our mission

Catalyzing policy change through research to improve people's lives in the emerging Asia Pacific by facilitating their use of hard and soft infrastructures through the use of knowledge, information and technology.

LIRNEasia
Pro-poor. Pro-market.

Where we work

LIRNEasia
Pro-poor. Pro-market.

What is big data?

Huge in volume

Generated at high speeds

Many different forms of data

Exhaustive

High-resolution

Relational

Big data: a simpler definition?

- An all-encompassing term for any collection of data sets so large and/or complex that it becomes difficult to process using traditional data processing applications.
- Examples:
  - 100 million Call Detail Records per day generated by Sri Lanka companies
  - 45 Terabytes of data from Hubble Telescope

Why big data? Why now?

- Vast drops in the cost of storing and retrieving information has lead to increased 'datafication'
- Exponential growth in computing power and memory
- Major improvements in techniques for performing machine learning and reasoning
- Democratization of analyses: open-source tools for processing (MapReduce/ Hadoop), analyses (R), as well as visualization

If we want comprehensive coverage of the population, what are the sources of big data in developing economies?

- Administrative data
  - E.g., digitized medical records, insurance records, tax records
- Commercial transactions (transaction-generated data)
  - E.g., Stock exchange data, bank transactions, credit card records, supermarket transactions connected by loyalty card number
- Sensors and tracking devices
  - E.g., road and traffic sensors, climate sensors, equipment & infrastructure sensors, mobile phones communicating with base stations, satellite/ GPS devices
- Online activities/ social media
  - E.g., online search activity, online page views, blogs/ FB/ twitter posts

If we want comprehensive coverage of the population, what are the sources of big data in developing economies?

- Administrative data
  - E.g., digitized medical records, insurance records, tax records
- Commercial transactions (transaction-generated data)
  - E.g., Stock exchange data, bank transactions, credit card records, supermarket transactions connected by loyalty card number
- Sensors and tracking devices
  - E.g., road and traffic sensors, climate sensors, equipment & infrastructure sensors, mobile phones communicating with base stations, satellite/ GPS devices
- Online activities/ social media
  - E.g., online search activity, online page views, blogs/ FB/ twitter posts

How does LIRNEasia leverage big data for public purposes?

Mobile network big data + other data ➔ rich, timely insights that serve private **as well** as public purposes

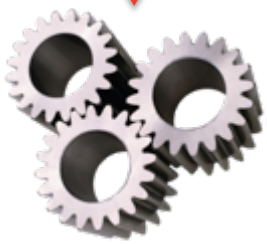
Mobile network big data (CDRs, Internet access usage, airtime recharge records)

Other Data Sources

1. Data from Dept. of Census & Statistics
2. Transportation data
3. Health data
4. Financial data
5. Etc.

Construct Behavioral Variables

1. Mobility variables
2. Social variables
3. Consumption variables



Dual purpose insights

Private purposes

1. Mobility & location based services
2. Financial services
3. Richer customer profiles
4. Targeted marketing
5. New VAS

Public purposes

1. Transportation & Urban planning
2. Crises response + DRR
3. Health services
4. Poverty mapping
5. Financial inclusion

What are some of the insights that we have developed?

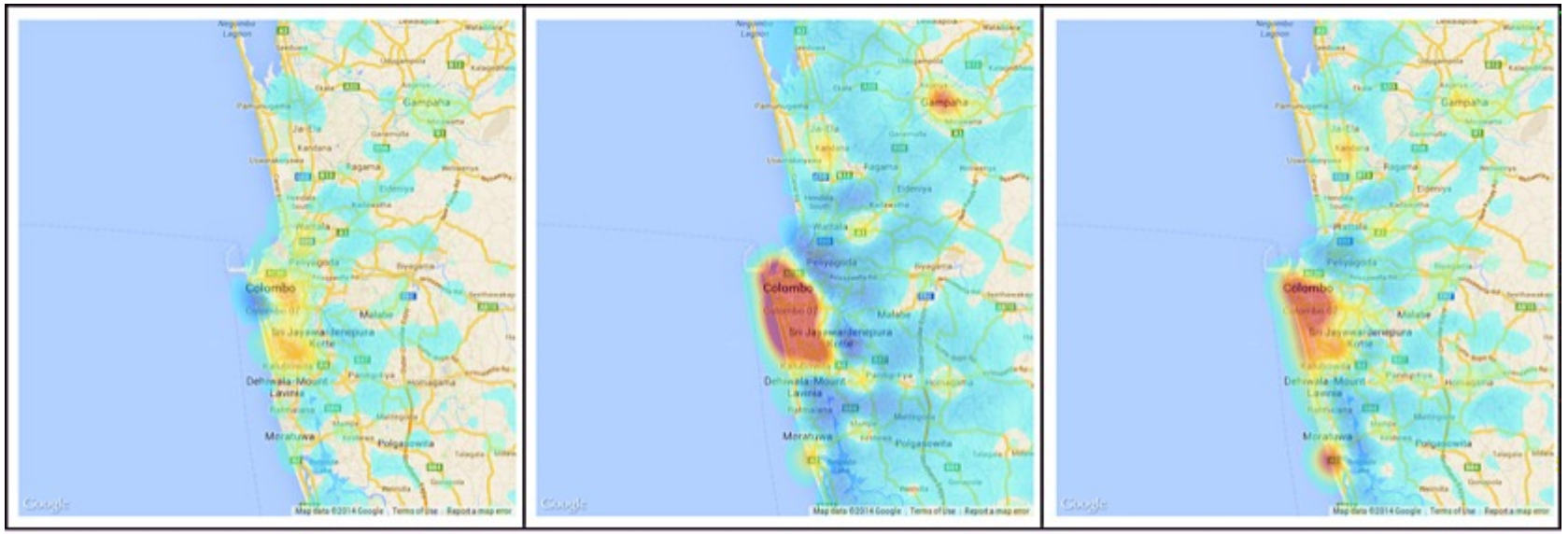
Population density changes in Colombo region: weekday/ weekend

Pictures depict the change in population density at a particular time relative to midnight

Weekday

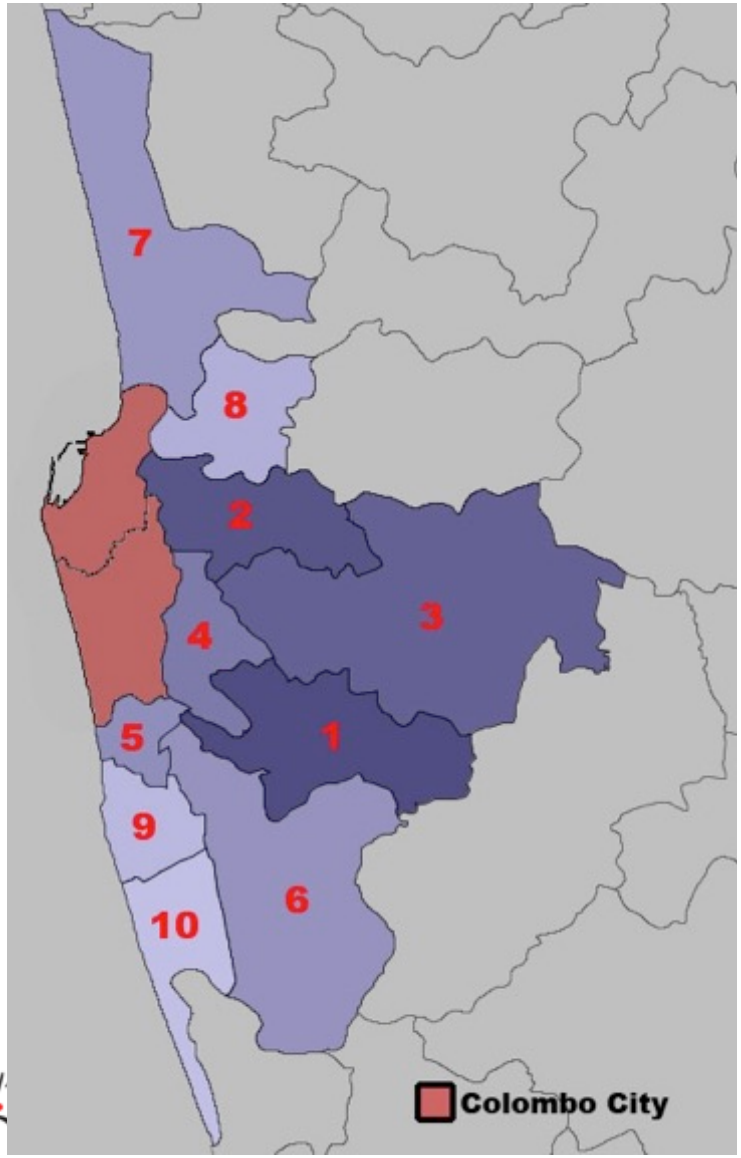
Time 06:30 Time 12:30 Time 18:30

Sunday



Decrease in Density Increase in Density

46.9% of **Colombo City's** daytime population comes from surrounding areas, but some surprises . . .

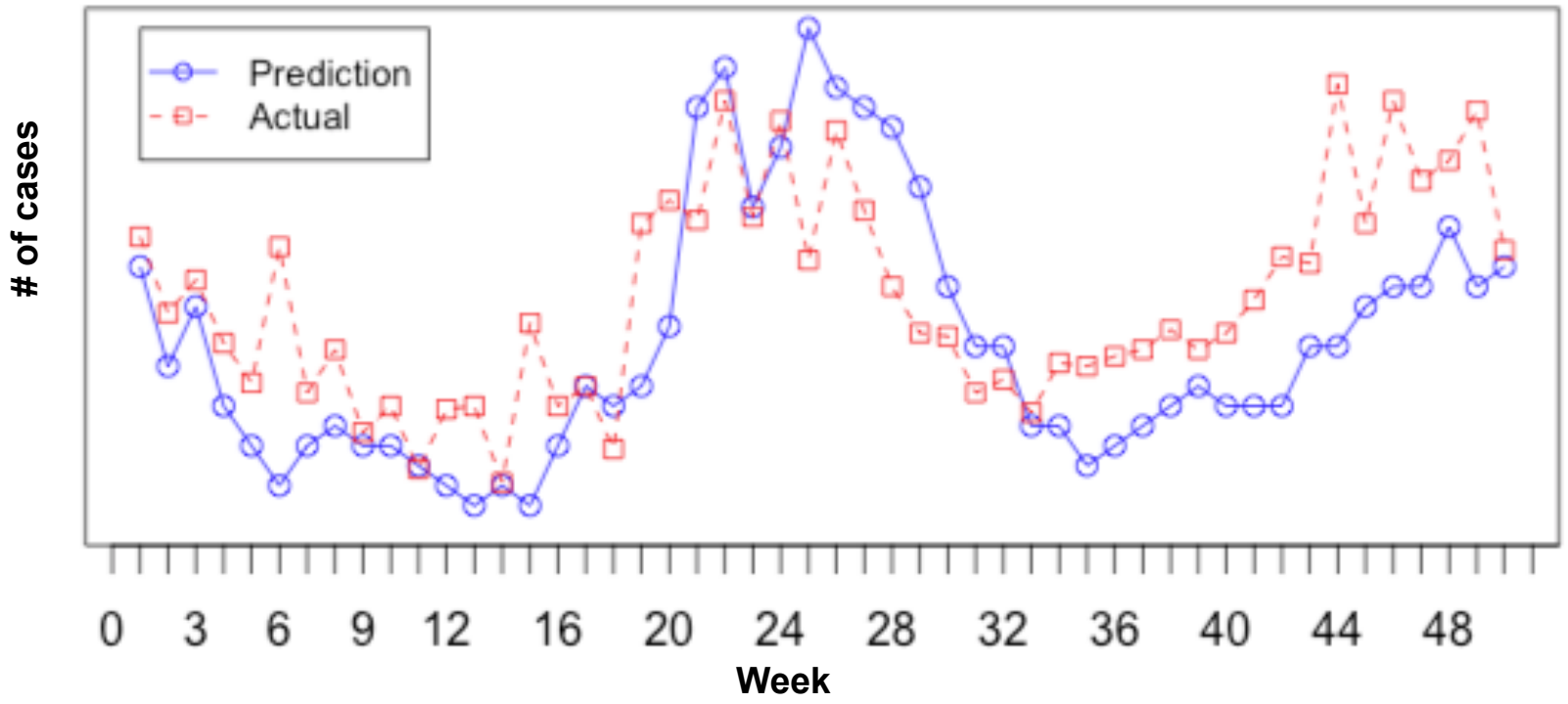


Colombo city is made up of Colombo and Thimbirigasyaya DSDs

Home DSD	%age of Colombo's daytime population
Colombo city	53.1
1. Maharagama	3.7
2. Kolonnawa	3.5
3. Kaduwela	3.3
4. Sri Jayawardanapura Kotte	2.9
5. Dehiwala	2.6
6. Kesbewa	2.5
7. Wattala	2.5
8. Kelaniya	2.1
9. Ratmalana	2.0
10. Moratuwa	1.8

Prediction of dengue outbreaks

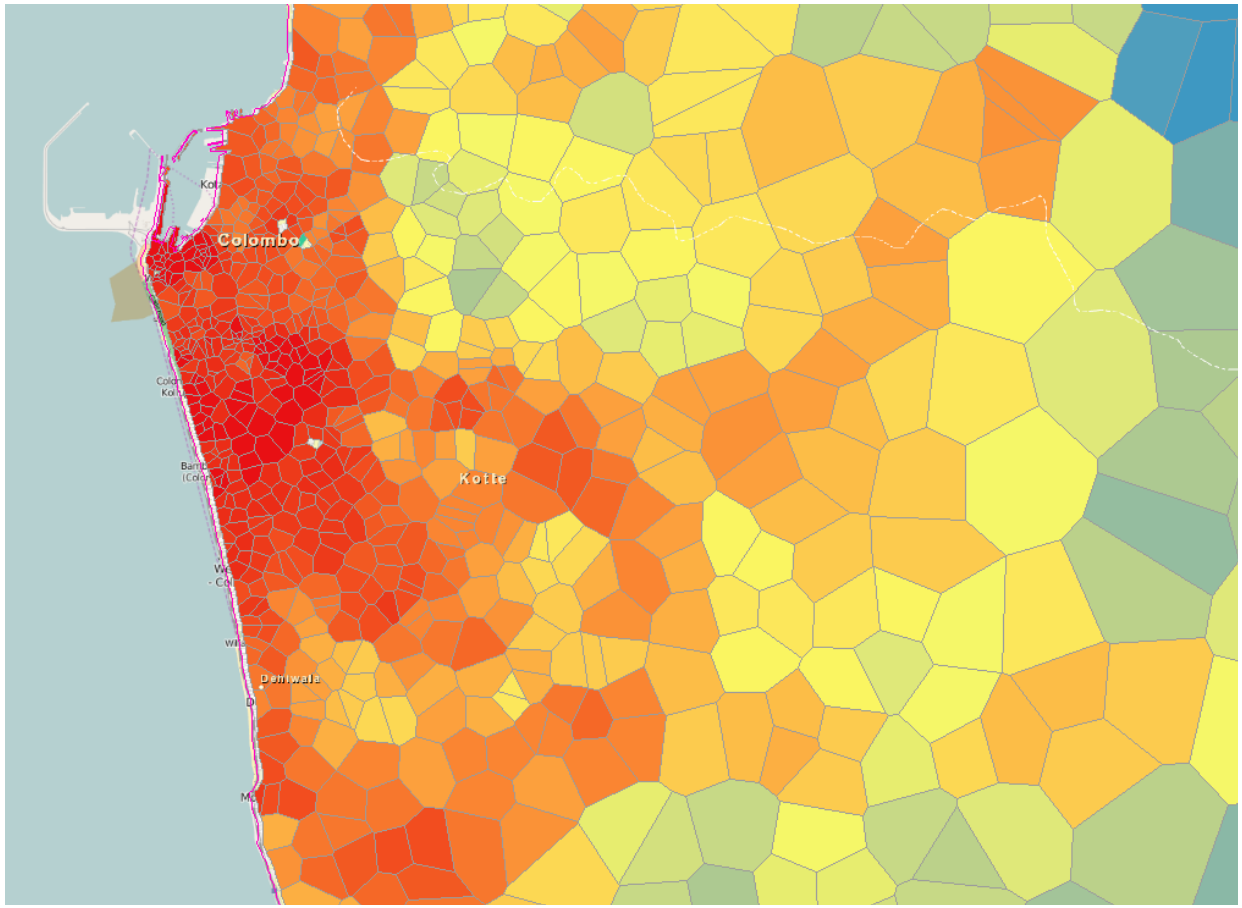
Comparing predicted & actual dengue outbreaks for Colombo city in 2014



- Such analyses will be very important should any cases of zika be detected in Sri Lanka

New proxy measures of economic activity

Estimated log(wage)



High Estimated Wage

Low Estimated Wage

Advantages:

- High temporal frequency and spatial resolution
- Can capture informal economic activity
- Can be utilized to map poverty

How are our big data insights being used?

- Big data insights used in formulating the Western Region Megapolis Plan (WRMP)
  - LIRNE*asia* insights used to support case for further development of specific peripheral regions in the province
  - Arguments for the need to alleviate daily congestion in Colombo city again use LIRNE*asia* insights
- Collaborations with the Urban Development Authority (UDA) and local universities to improve land use mapping

Advantages of using mobile network big data in developmental policy

- High frequency
- High resolution
- Near real time
- (Almost) universal coverage

CHALLENGES

Challenge: accessing government data

- Data from government
  - Requires greater 'datafication' especially in developing economies
  - Being spearheaded through the open data movement

Challenge: accessing private sector data

- Not amenable to open data initiatives
- There can be competitive implications to data release, even if pseudonymized

Potential opportunities for accessing data:

- The value of 'mashing' data (both private as well as government data) is greater than the sum of the parts
- The state of the art is still developing, enabling unique collaborations e.g. UN Global Pulse innovation labs, Orange Data for Development challenges, UP Singapore, LIRNE*asia*
- There is a potential role for intermediaries that interface between private business interests and public uses e.g. ongoing work at LIRNE*asia*

Challenge: skills

- Statistics can work differently in a big data world
- Its not just about data science skills: domain knowledge equally important

Opportunity:

- New collaborations
- Inter-disciplinary teams will be essential if we are to leverage big data for public purposes

Challenge: privacy

- The higher the resolution of your insights, the greater the privacy implications
- Mixing non-personal data with other sources can reveal personal attributes
- Informed consent is meaningless in a big data world
- People often don't know what their generalizable privacy needs are or how their preferences might evolve

We should also be asking the following:

- Should we concentrate on defining and enforcing privacy rights or reducing harms?
- When it comes to developing economies, should we worry about inclusion or exclusion in big data?
- What about competition?

KEY TAKEAWAYS FROM OUR EXPERIENCES

Key takeaways

- Knowing local context matters
- First target low hanging fruits/ quick wins
- Breaking data silos is possible by showing benefits early
- Extensive policy enlightenment with all levels of the symbolic environment of policy domains leads to greater acceptance

Thank you.

More info at:

http://lirneasia.net/projects/bd4d/