Improving Disease Outbreak Forecasting Models for efficient targeting of Public Health Resources

Dengue is a vector-borne infectious disease that is estimated to have 390 million infections annually, out of which 96 million manifest. WHO estimates half the global population to be risk from this neglected tropical disease. Accurate disease outbreak forecasting is a essential to reduce mortality as well as morbidity from infectious diseases such as dengue. It enables limited public health resources to be utilized for maximum impact by executing disease prevention and control mechanisms on targeted high-risk regions.

Arboviral infectious diseases such as dengue are primarily propagated across distant regions via the movement of humans since the spatial range of the mosquito vector is limited. Therefore identifying aggregate human movement patterns is critical for fighting vector-borne diseases just as it is for contagious diseases. New sources of big data such as Mobile Network Big Data (MNBD) is able to shed light on human mobility patterns at spatial and temporal granularities that had hitherto not been possible with census or survey based data. Combining such rich spatio-temporal information extracted from MNBD data sources with other epidemiological data as input features for machine learning methods, we can develop more accurate forecasting models that can predict infectious disease outbreaks at a high spatial and temporal granularity. Based on data from Sri Lanka, our research shows that MNBD can be used to derive human mobility models and obtain insight into how human movement patterns influence propagation of vector-borne infectious diseases such as dengue. It also shows that human mobility plays a significant role even in regions where dengue is endemic.

MAIN FINDINGS & RECOMMENDATIONS

1. **Even in regions where the disease is endemic, human mobility is a critical factor for dengue propagation:** We observed that human mobility is highly correlated to dengue incidence in our work. We were also able to obtain improved predictions when human mobility was integrated into predictive models even in regions where dengue is endemic.
2. **Disease outbreak forecasting models can be used to formulate epidemic disease policy and allocate public health sector resources efficiently**: Obtained epidemic models can be used to identify high-risk regions and prepare policy accordingly. Forecasting regions with high dengue incidence 2 weeks of ahead of time will help allocate limited public health sector resources for maximum impact.
3. **Models developed for dengue can easily be extended to predict propagation of other infectious diseases**: Resources need not be allocated to conduct additional studies if a similar infectious disease such as zika gets introduced to the country. The already developed models can be applied to predict the spread of such an emerging diseases and identify regions that need to be targeted to prevent spreading of the disease further.

THE RESEARCH

LITERATURE/PRIOR WORK

Prediction of dengue outbreaks has been the focus of multiple studies globally including Sri Lanka. Recent research studies exist that also make use of Mobile Network Big Data (MNBD) as a means of deriving large scale human mobility patterns for disease outbreak prediction and epidemic modeling (Wesolowski et al., 2015).

Many studies have also made use of different machine learning techniques to predict dengue incidence (Wu, Lee, Fu, & Hung, 2008; Herath, Perera, & Wijekoon, 2014). However, we were able to find comparatively a fewer number of studies that made a comparison between different methodologies. A Malaysian study (Yusof & Mustaffa, 2011) compared two techniques, namely Least Squares - Support Vector Machines (SVM) vs. Neural Networks. Another Malaysian study (Wu, Lee, Fu, & Hung, 2008) made use of wavelet decomposition, SVM and Genetic Algorithms (GA) to detect climatic factors that contribute towards dengue incidence. We make use of a similar technique to improve the accuracy of our predictive models.

THE DATA

Our work uses Call Detail Record (CDR) data (which is MNBD) spanning more than 1 year for nearly 10 million SIMs from multiple mobile operators in Sri Lanka to derive aggregate human mobility patterns. The data is completely pseudonym-ized by the operator.

Weekly reported dengue cases for each Medical Officer of Health (MOH) division, which is the smallest health administrative unit for Sri Lanka, as well as rainfall and temperature data are made use of in our research. We also utilise the mean Normalized Difference Vegetation Index (NDVI) values, which were derived using Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data. All input data were projected to its corresponding MOH division with a temporal scale of 1 week.

METHODOLOGY

We initially developed multiple human mobility models that provide a mobility value for each MOH division by processing the CDR dataset. Then we developed predictive models that integrate different data sources using multiple machine learning techniques.

Human Mobility Models: Separate mobility models were developed based on the methodology used to obtain a mobility value for an MOH division.

a. ***Probabilistic Mobility Model***: Calculates the proportion of calls initiated/received by a subscriber within his/her home MOH division vs. calls initiated or received in an outside MOH division and aggregates this value of all subscribers residing with an MOH division to obtain a mobility value
b. ***Trip-based mobility model***: Based on the number of trips taken by a mobile subscriber to a particular MOH division
c. ***Risk-based mobility model***: A risk value is assigned based on whether a subscriber is at home, work or other place and obtains a risk value for an MOH division

Predictive Models: All input data sources are aggregated to an MOH division and used to train different machine learning models (see Table 1) to obtain forecasts for the year 2014. A genetic algorithm based approach was used to select input features that contributed most to the accuracy of the model.

RESULTS

By extracting spatio-temporal information related to human mobility from the CDR data set, we were able to map MOH divisions with high risk of dengue due to human mobility. The risk map in Fig.1 is shown as an example. Integrating the values obtained from these mobility models into different machine learning techniques showed improvement in predictive performance when mobility was introduced as an input feature. The results are summarized in Table 1.

Model	R^2		RMSE	
	- Mobility	+ Mobility	- Mobility	+ Mobility
Random Forests	0.628	0.639	6.907	6.812
Neural Networks	0.063	0.335	10.966	9.239
XGBoost	0.63	0.64	6.892	6.794
SVR	0.68	0.704	6.408	6.17

Table 1: Summary of machine learning models with/without mobility as an input feature

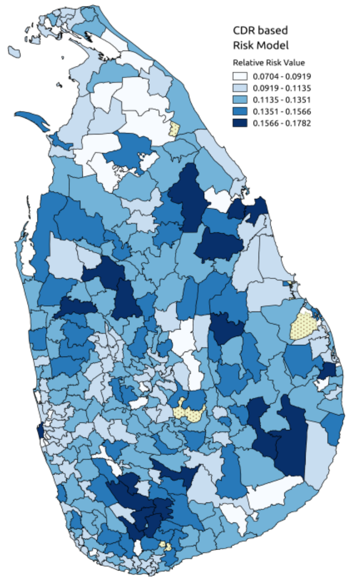

Fig. 1: Risk map of Sri Lanka based on risk-based mobility model for week 32 of year 2013

ACKNOWLEDGMENTS

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.

SOURCES

Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rasheed, S. B., … Buckee, C. O. (2015). Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences, 112*(38), 11887–11892.

Wu, Y., Lee, G., Fu, X. J., & Hung, T. (2008). Detect climatic factors contributing to dengue outbreak based on wavelet, support vector machines and genetic algorithm. *World Congress on Engineering 2008 Vols III, 1*, 303–307.

Yusof, Y., & Mustaffa, Z. (2011). Dengue Outbreak Prediction: A Least Squares Support Vector Machines Approach. *International Journal of Computer Theory and Engineering, 3* (4), 489–493.

AUTHOR/S

Lasantha Fernando| LIRNE*asia*| 12, Balcombe Place, Colombo 00800, Sri Lanka | lasantha@lirneasia.net | www.lirneasia.net

Sriganesh Lokanathan | LIRNEasia | 12, Balcombe Place, Colombo 00800, Sri Lanka | sriganesh@lirneasia.net | www.lirneasia.net

Amal Shehan Perera | Dept. of Computer Science & Engineering, University of Moratuwa |Moratuwa, Sri Lanka | shehan@uom.lk | www.cse.mrt.ac.lk

Azhar Ghouse | Epidemiology Unit, Ministry of Health | 231 De Saram Pl, Colombo 00800, Sri Lanka | docazhar@gmail.com | www.epid.gov.lk

Hasitha Tissera | Epidemiology Unit, Ministry of Health | 231 De Saram Pl, Colombo 00800, Sri Lanka / *dr_korelege@gmail.com* | www.epid.gov.lk