Impact of Human Mobility on Spread of Dengue in Sri Lanka

Conference Paper · April 2017

5 authors, including:

Lasantha Fernando
LIRNEasia

Sriganesh Lokanathan
LIRNEasia

Hasitha A Tissera
Epidemiology Unit & Dengue Control Unit

Impact of Human Mobility on Spread of Dengue in Sri Lanka

Lasantha Fernando
LIRNE*asia*, Colombo
Sri Lanka
Email: lasantha@lirneasia.net

Amal Shehan Perera
Dept. of CSE, University of Moratuwa
Sri Lanka
Email: shehan@cse.mrt.ac.lk

Sriganesh Lokanathan
LIRNE*asia*, Colombo
Sri Lanka
Email: sriganesh@lirneasia.net

Azhar Ghouse
Epidemiology Unit, Ministry of Health
Sri Lanka
Email: docazharghouse@gmail.com

Hasitha Tissera
Epidemiology Unit, Ministry of Health
Sri Lanka
Email: dr_korelege@yahoo.co.uk

Rohan Samarajiva
LIRNE*asia*, Colombo
Sri Lanka
Email: rohan@lirneasia.net

***Abstract*—Human mobility plays a significant role in spatio-temporal propagation of infectious diseases. But how much of an impact does human mobility have on propagating a dengue outbreak in a dengue endemic country such as Sri Lanka? We show that a proxy value for human mobility, derived from mobile network big data, has a significant correlation with dengue incidence using past case data. Furthermore, we discuss the applicability of this proxy measure of mobility to increase accuracy of several prediction models based on machine learning techniques. These improved models can be used to do spatio-temporal forecasting of dengue outbreaks which can help medical officers and related governmental or non-governmental agencies execute preemptive measures before the outbreak occurs in actuality.**

I. Introduction

The significance of human mobility in propagation of infectious diseases has been established in multiple works done previously [1, 2]. Until recently, the datasets and methodologies available for estimating human mobility have been limited. However, there has been a significant increase in research done to estimate human mobility using call detail records (CDR) with applications in multiple, wide ranging domains [3–5].

The main focus of our work is in measuring the impact of human mobility in propagating a vector borne tropical disease such as dengue in a dengue endemic country. A recent study from Pakistan [6] developed a model that validates the impact of human mobility in introducing dengue to naive regions as well as regions where dengue incidence is low. However in regions where dengue is already prevalent throughout the year, human mobility would not effect disease propagation in the same manner. The forecasting models developed for regions where the human population is immunologically naive to dengue would not be applicable to a country like Sri Lanka, where dengue is endemic to most regions of the country.

In the next section, we describe the proxy measure for human mobility developed by us, which we use to determine the correlation between that value and dengue incidence. We compare the correlation of this value with other values such as temperature and rainfall (factors previously identified in the literature as contributing factors to disease propagation [7]) to show that there is a significant correlation between human mobility and dengue incidence. In section III, we compare the accuracy of the proxy measure by incorporating the mobility value to multiple prediction models to determine whether mobility plays a significant role in spreading dengue in a dengue endemic country such as Sri Lanka.

II. Mobility Model

Most of the mobility models developed in related work have been based on the number of trips between different regions[6, 8]. Additionally, considering only overnight stays of subscribers in a region other than his or her home, an approach taken for a study on malaria in Namibia [9], is not suitable for modeling dengue propagation. The reason is that the primary vector of dengue, aedes aegypti, is known to be most active during daytime [10]. Our focus was on developing a mobility measure that can be applied to a geographic region so that it can be incorporated directly into machine learning models. In the context of disease forecasting, machine learning methods can be described as working broadly on the principle of identifying a set of attributes that can describe dengue incidence characteristics for a specific spatial region, and training a model to learn the weightage of those attributes that contributes toward disease incidence. For such a forecasting model, we would need a mobility measure that corresponds to each spatial and temporal unit of consideration.

In our project, the smallest spatial unit for which data is available is a Medical Officer of Health (MOH) division, which is an administrative area defined by Sri Lanka's Ministry of Health. Number of dengue cases reported from each MOH division in a given week for the year of 2013 was used to represent dengue incidence. Pseudonymized call detail records from multiple mobile operators in Sri Lanka for the year of 2013 was used to derive the mobility measure. The mobility model derived using the available mobile phone data is described below.

A. Probabilistic Mobility Model

The probabilistic model was developed by building upon available literature on the subject where it is assumed that the number of calls taken or received by a subscriber in a particular region is proportional to the amount of time spent in that region [11]. Our model builds on the same assumption and obtains a normalized mobility value for each MOH division. The mobility value for a subscriber is calculated by the number of calls made by a subscriber who is not a resident of that particular MOH in a given week. This value is normalized by dividing from the total number of calls taken by that subscriber in that week. This derived value per subscriber is aggregated to get a mobility value for an MOH division.

If we consider M as a set of all MOH divisions, and S as a set of all subscribers, our model can be defined as follows:

$CDR(m_i, s_j, w_k)$ = No. of CDR in MOH division m_i, for subscriber s_j during week w_k where $\forall m_i \in M, \forall s_j \in S$

Mobility of subscriber $\mathbf{s_j}$ at MOH $\mathbf{m_i}$ can be defined as

$$mob(m_i, s_j) = \frac{CDR(m_i, s_j, w_k)}{\sum_i^M CDR(m_i, s_j, w_k)} \tag{1}$$

where $\forall m_i \in \{M - Home(s_j)\}, \forall s_j \in S$

Mobility for MOH $\mathbf{m_i}$ can be defined as

$$mob(m_i) = \frac{\sum_j^N mob(s_j)}{N} \tag{2}$$

where N - No. of subscribers travelled to $\mathbf{m_i}$ for that week

III. Results & Prediction Models

The distance correlation was used as a measure to determine the dependence between dengue incidence and mobility. The correlation graph for mobility and other input features is depicted in Fig. 1. This graph shows that mobility has a significant correlation when compared to the other input parameters considered in our predictive models such as temperature, rainfall and the no. of dengue cases in previous weeks.

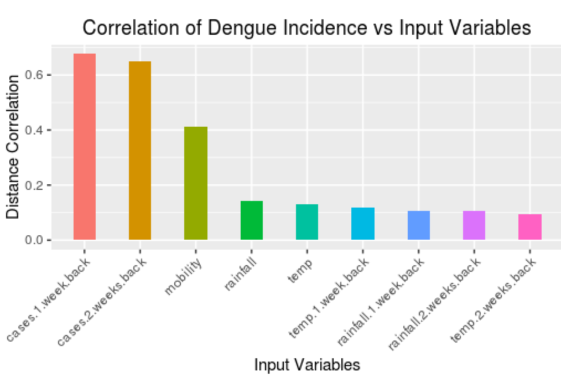

Fig. 1. Correlation of input parameters with dengue incidence

TABLE I
RMSE AND r^2 VALUES OF DIFFERENT MACHINE LEARNING METHODS

Model	Mobility Used	RMSE	r^2
Neural Networks	No	10.923	0.194
Neural Networks	Yes	9.867	0.342
XGBoost [12]	No	10.023	0.321
XGBoost	Yes	9.548	0.384
SVR	No	7.445	0.139
SVR	Yes	7.297	0.173

Several different machine learning methods were tried out to come up with a dengue outbreak forecasting model with and without the derived mobility measure. The results from our own preliminary analysis is also given in Table I. The RMSE and the r^2 values for each method improved with the introduction of the mobility measure. Amongst the methods that were tried, Support Vector Regression (SVR) [13] provided the greatest accuracy, reflecting recent literature [14] where the performance of SVR has been shown to be better than Neural Networks. Therefore, SVR was selected as the method to carry out further work on our predictive model. The model was trained for 5 MOH divisions that were identified beforehand and the resultant model was used to predict for the sixth MOH division. After tuning the SVR model, we were able to obtain an RMSE value of 6.463 without mobility and 6.236 with the introduction of mobility. Similarly, the r^2 values were 0.351 and 0.396 for SVR models without and with mobility respectively. The related prediction graphs for Moratuwa MOH division are given in Fig.2 & 3.

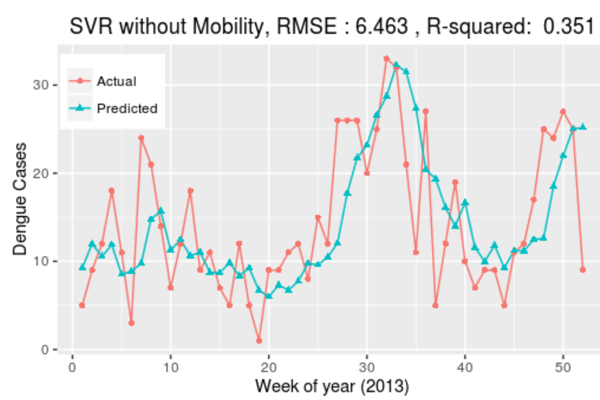

Fig. 2. Predicted vs actual dengue incidence for Moratuwa MOH (without mobility)

IV. Ongoing and Future Work

We are currently working on developing a trip based mobility model where the derived value can be aggregated to an MOH division. This model can be compared with the mobility model described above to determine which approach yields the greatest accuracy and fit. It can also be used as another validation of the observed improvement in predictive accuracy due to the introduction of mobility. Since the mobility measure

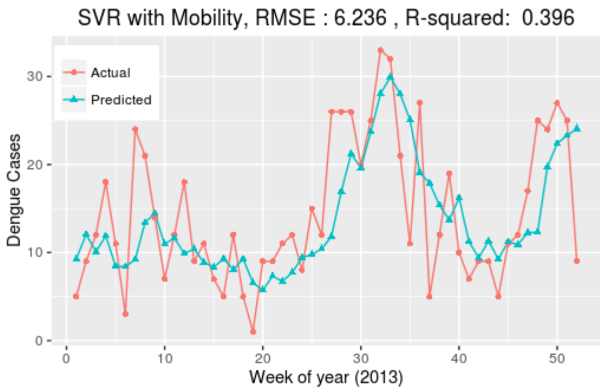

Fig. 3. Predicted vs actual dengue incidence for Moratuwa MOH (with mobility)

is derived independently of the modeling of the disease itself, this measure can be used in forecasting other infectious disease outbreaks as well. Validating the feasibility of using this measure in forecasting other infectious diseases can be done as a part of future work.

V. Conclusion

We have proposed a new methodology to calculate human mobility that provides a mobility measure for each administrative region. The advantage of this measure is it can be directly applied to any machine learning model without having to calculate a separate measure for different machine learning techniques. The results show that this measure is significantly correlated with dengue incidence. Additionally, we were able to obtain an improvement in prediction accuracy of all the considered models by introducing the mobility measure. Based on these results, we can conclude that mobility does have a significant impact in dengue propagation even in a dengue endemic country such as Sri Lanka.

Acknowledgment

The authors would like to thank the Epidemiology Unit, Ministry of Health, Sri Lanka for providing dengue incidence data as well as expert knowledge on the disease dynamics and the epidemiology of dengue. This research was funded through a grant from the International Development Research Centre (IDRC) of Canada and a grant from the Senate Research Committee (SRC) of University of Moratuwa.

References

[1] Michele Tizzoni et al. “On the Use of Human Mobility Proxies for Modeling Epidemics”. In: *PLOS Computational Biology* 10.7 (July 2014), pp. 1–15.

[2] Dirk Brockmann. “Human Mobility and Spatial Disease Dynamics”. In: *Reviews of Nonlinear Dynamics and Complexity* 2 (2010), pp. 1–24.

[3] Enrique Frias-Martinez, Graham Williamson, and Vanessa Frias-Martinez. “An Agent-Based Model of Epidemic Spread using Human Mobility and Social Network Information”. In: *3rd International Conference on Social Computing (SocialCom’11)* (2011), pp. 49–56.

[4] Sibren Isaacman et al. “Identifying important places in people’s lives from cellular network data”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6696 LNCS.June (2011), pp. 133–151. ISSN: 03029743.

[5] R Samarajiva et al. “Big data to improve urban planning”. In: *Economic and Political Weekly* 50.22 (2015), pp. 42–48. ISSN: 00129976 (ISSN).

[6] Amy Wesolowski et al. “Impact of human mobility on the emergence of dengue epidemics in Pakistan”. In: *Proceedings of the National Academy of Sciences* 112.38 (2015), pp. 11887–11892.

[7] Suchithra Naish et al. “Climate change and dengue: a critical and systematic review of quantitative modelling approaches.” In: *BMC infectious diseases* 14.1 (2014), p. 167. ISSN: 1471-2334.

[8] Amy Wesolowski et al. “Quantifying the impact of human mobility on malaria.” In: *Science (New York, N.Y.)* 338.6104 (Oct. 2012), pp. 267–70. ISSN: 1095-9203.

[9] Nick W Ruktanonchai et al. “Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data.” In: *PLoS computational biology* 12.4 (2016), e1004846. ISSN: 1553-7358.

[10] Duane J Gubler and Gary G Clark. “Dengue/dengue hemorrhagic fever: the emergence of a global health problem.” In: *Emerging infectious diseases* 1.2 (1995), p. 55.

[11] Flavio Finger et al. “Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks”. In: *Proceedings of the National Academy of Sciences* 113.23 (2016), p. 201522305. ISSN: 0027-8424.

[12] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2.

[13] Debasish Basak, Srimanta Pal, and Dipak Chandra Patranabis. “Support vector regression”. In: *Neural Information Processing-Letters and Reviews* 11.10 (2007), pp. 203–224.

[14] Yuhanis Yusof and Zuriani Mustaffa. “Dengue Outbreak Prediction : A Least Squares Support Vector Machines Approach”. In: *International Journal of Computer Theory and Engineering* 3.4 (2011), pp. 489–493. ISSN: 17938201.