Trabalho apresentado no XXXIX CNMAC, Uberlândia - MG, 2019.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Deep Learning-Based Dengue Cases Forecasting with Synthetic Data

Juan V. Bogado [1]
Diego Stalder [2]
Santiago Gómez [1]
Christian Schaerer [1]
[1] Polytechnic School, National University of Asunción, Paraguay
[2] Paraguayan Space Agency, Paraguay

1 Introduction

Dengue fever is a viral disease transmitted to humans mainly by the *Aedes aegypti* mosquito acting as a vector, with greater incidence in urban areas. Worldwide, an estimation indicates that in tropical and sub-tropical countries, approximately 975 million people living in urban areas are at risk of infection. It is estimated that more than 50 million infections occur each year, including 500,000 hospitalizations for dengue hemorrhagic fever [2]. Over the last decade, there has been a dramatic increase in dengue infections in South American countries such as Colombia, Ecuador, Paraguay[2], Peru, Venezuela and Brazil. It is also known that dengue has an endemic characteristic, and this is why it is considered a public health problem in tropical and sub-tropical regions.

Causal and statistical models have been developed to predict the incidence of the disease as a function of time, considering the heterogeneity between different sub-region locations and populations. Several models capture dengue dynamics and allow to study various scenarios and to evaluate the effects of various intervention types[3] [1]. The main limitation when fitting those models is the scant surveillance data. However, their predictions can be used as synthetic data to train intelligent computational models, such as deep neural networks.

2 Methodology

Recently, Wang et al [4], proposed an epidemic forecasting framework that integrates the strengths of artificial neural networks, eg. Long Short-Term Memory (LSTM), and causal methods for influenza-like illness (ILI) with promising results. This work applies the same theory-guided data science paradigm, using synthetic data generated by causal models in conjunction with surveillance data to improve the dengue outbreaks forecast. A similar starting point is considered, synthetic data is generated by a dengue multi-agent model, which considers several factors, as the population dynamics, weather conditions and geographic heterogeneities [3].

3 Preliminary results

The deep neural networks considered for this work are composed by sets of LSTM cells. Those cells are capable of learning long-term dependencies, characteristics which makes it

---

[1] juan.vicente.bm@gmail.com, dstalder@aep.gov.py, sgomepz@gmail.com, cschaer@pol.una.py

[2] Paraguay reports more than 100.000 confirmed infections between December 2012 and July 2013

[3] A nowcasting system for the surveillance of arboviruses has been developed in Brazil: https://info.dengue.mat.br

the best option for time series forecasting.

To generate synthetic data, we consider a Dengue multi-agent model (eg. SIR-SI[4]) with their parameters estimated based on surveillance data [3]. Then the deep neural network (LSTM) weights are trained based on the synthetic data. The time dependent dataset is divided into 80% for network training and 20% for the test. Finally we perform a fine tune of the neural network architecture to adjust the parameters of the network, eg. number of LSTM cells and the windows size which is the number of previous records in the time series to be considered for the next prediction. The metric used to evaluate the forecasting performance was: root mean squared error (RMSE) considering real data[5], the loss function returned the values of 0.005 and 0.003 for training and validation respectively. Figure 1, present a real dengue time series where only 25 LSTM neural network layers yields quite good predictions.

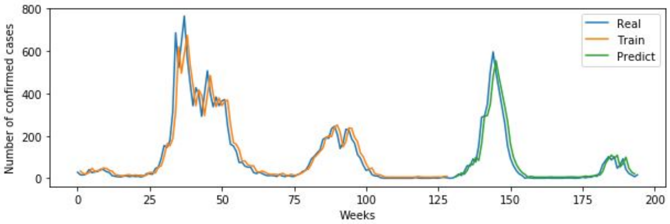

Figure 1: Preliminary LSTM results Dengue cases in Caaguazú City, 2013-2015

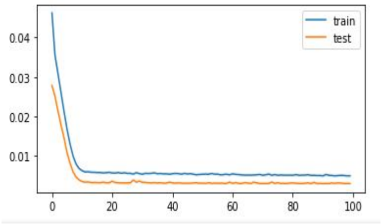

Figure 2: Performance evaluation of the network with the loss function RMSE

4 Conclusion

This preliminary result motivates the introduction of additional parameters such as demographic, geographic and environmental variables to characterize the behavior of the epidemic across regions.

Acknowledgement. This work is funded by Paraguay's CONACyT research grant PINV15-706 "Construcción de un modelo de incidencia de dengue para comunidades de Paraguay (Comidenco)".

References

[1] M. Andraud, N. Hens, C. Antonie and P. Beutels. Dynamic epidemiological models for dengue transmission: a systematic review of structural approaches, *PLoS One*, 7:e49085, 2012. DOI: 10.1371/journal.pone.0049085.

[2] M. Guzmán, S. Halstead, H. Artsob et al. Dengue: a continuing global threat, *PLoS One*, 8:S7, 2010. DOI: 10.1038/nrmicro2460.

[3] T. Lima, R. Lana, L. Senna and C. Davis. Dengueme: A tool for the modeling and simulation of dengue spatiotemporal dynamics, *International journal of environmental research and public health*, 13:920, 2016. DOI: 10.3390/ijerph13090920.

[4] L. Wang, J. Chen and M. Marathe. DEFSI: Deep Learning Based Epidemic Forecasting with Synthetic Information, *Proceedings of the 30th innovative Applications of Artificial Intelligence(IAAI)*, 2019.

---

[4] SIR-SI: Susceptible-Infective-Recovered for human populations; Susceptible-Infective for vector populations

[5] Epidemiological Bulletin, Health Surveillance Directorate. Paraguay