ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

A Proposal for Epidemic Prediction using Deep Learning

[1]Sagar Palao, [2]Abhishek Shahasane, [3]Siddhesh Dighe, [4]Harsh Gupta, [5]Anjali Yeole
[1][2][3][4] B.E., Computer Engineering, VESIT, Mumbai, India
[5] Professor, Computer Department, VESIT, Mumbai, India

***Abstract* - HealthCare has been a surging need in India. The inception of Smart India, Smart Villages has further emphasized the need of a Smart & Healthy India. However little reflection of such digitally empowered country has been seen in the field of healthcare. This has further increased the importance of the definition of ubiquity of information and computing technology in healthcare that demand constant surveillance and vigilance of healthcare to predict epidemic outbreak and biological attack. The paper's primary focus is to analyze and determine the spread of diseases and epidemic in cities/villages. And using this analysis to predict where the next outbreak of epidemic will be. This prediction helps the health authorities to take necessary action in terms of assuring that sufficient resources are available to suffice the need and if possible stop the occurrence of such epidemic by taking necessary actions. To achieve this, we use deep neural network as the heart of our prediction. It receives its training from the past experiences of data which we have collected from hospitals and our spread network. Using this training with our dynamic data it makes predictions as well as adaptively learns from the real time data.**

***Index Terms:*— epidemic prediction, deep learning, recurrent neural network, machine learning, disease spread network**

I. INTRODUCTION

HealthCare has been a surging need in India. The inception of Smart India, Smart Villages has further emphasized the need of a Smart & Healthy India. However little reflection of such digitally empowered country has been seen in the field of healthcare. This has importuned the definition of ubiquity of information and computing technology in healthcare that demand constant surveillance and vigilance of healthcare to predict epidemic outbreak and biological attack.

Our primary motivation is to analyze and determine the spread of diseases and epidemic in cities/villages. And using this analysis to predict where the next outbreak of epidemic will be. This prediction helps the health authorities to take necessary action in terms of assuring that sufficient resources are available to suffice the need and if possible stop the occurrence of such epidemic by taking necessary actions.

Our secondary motivation is to determine biological threats in cities/villages, primarily by computational intelligence.

The objective outcome is to determine with what probability a disease will outbreak in a particular region and what is the possible number of individuals who will be prone to be affected by it. This prediction will be backed with accuracy level which will act as an alarm for the health authorities who are monitoring the health status of the city or village.

II. METHODOLOGY

The heart of the project is the model which predicts the possibility of epidemic and disease in a region. The implementation of the model is done using two technologies which has a pervasive impact in computation. The technology and their role in our project is as given,

A. Deep Learning RNN

Deep learning (also known as deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning based on a set of algorithms that attempt to model high level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations. We are using Recurrent Deep Neural Networks for our implementation. RNNs have become the method of choice for processing information which has a sequential context.

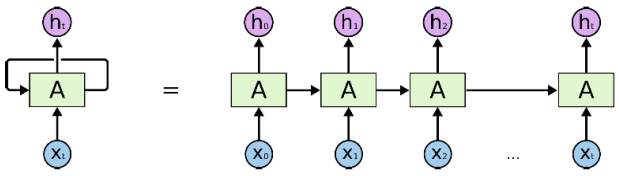

Fig. 1. Deep RNN

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

B. Hard Computing

We used hard computing to use our spread network to blend with our Deep Recurrent Neural Network. The Deep Neural Network primarily learns form past experiences which we have gained from the information of hospitals. In order to blend our information of spread networks – sewage network, air flow network, human contact network along with the environmental condition during that phase of time in our model hard computing plays a vital role.

III. REQUIREMENT

A. Epidemic Outbreak Prediction

The key functional requirement of SpateAugur is it should predict epidemic/disease outbreak in different region of city. The prediction should come as:

1) Which region of a city is to be infected by epidemic and which epidemic?
2) What is the probability that the disease will outbreak in the region?
3) What are the expected number of people to be infected by the disease in that region?

The epidemic is prioritized based on its impact in the city and the health officials are notified and so they can maintain constant surveillance.

B. Visualization

The system gives the health official the visualization of the spread of disease. The visualization has the following requirements.

1) Timeline: The system displays the timeline which defines the flow of spread of disease from region to region in the city. This visualization helps the official to determine the nature of spread of each disease.
2) Disease Heat Map: It defines heat map of how the infectious the disease is in a region in form of a heat map.
3) Prediction Heat Map: It defines how latent is it for a disease to outbreak in a region in form of a heat map.

C. Fetch Dynamic Data

The system is a self-learning and ever learning model. To learn it needs dynamic and periodic data. It fetches this data from various sources. Data is the pivotal mean for the model to learn. It fetches data as,

1) Plugin: A plugin is a part of SpateAugur software which collects data from the health center, hospitals and dispensaries. The plugin can read data from multiple form of databases including support for Oracle, MS-Access, MS SQL Server.
2) API: A set of APIs are used to collect data such as Sewage Network (offered by MCGM), Weather network (offered by CPCB), Transport frequency (offered by TISS) and Air Flow network (offered by CPCB).

D. Public Notifications

When a disease is predicted to be outbreak in a region, SpateAugur notifies the users in the region to take precautionary measures. For example, If it is predicted Cholera to spread in a regions. The users will be notified as: 'Please drink safe water, boiled water or purified water. Don't eat street food for a while. The health officials has predicted an outbreak of cholera in your area.'

This notification will be send to the users by mean of SMS/Push Notifications, so that all grade of public in the region are notified and are ready to fight back the epidemic.

IV. PROPOSED DESIGN

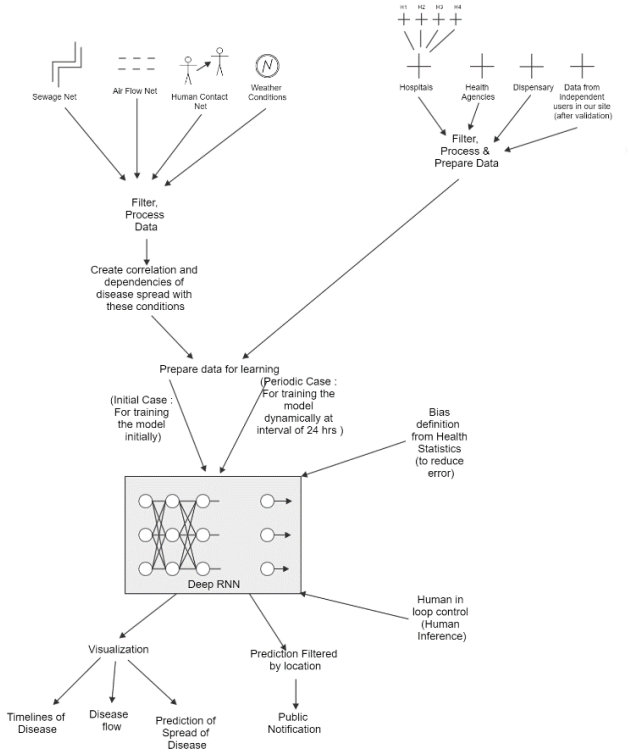

Fig. 2. System/Conceptual Diagram

The diagram gives a generic flow of the entire software in its most abstract view. We can perceive the diagram in 3 stages. The stages are:

A. Data Collection Unit

It is responsible to collect data in real time. Primarily two set of data is collected. It includes:

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

1) Data from Health Agencies, Hospitals and dispensaries of the patient: Data key ingredients are:

- Region of the patient
- Detected diseases
- Timestamp when patient enrolled and when patient was discharged.

2) Data about the Disease Spread Network

- Sewage Network form MCGM
- Air Network form CPCB
- Human Contact Network from TISS
- Weather Condition from CPCB
- The collected data is processed to make it suitable for stage B.

B. Deep RNN (Recurrent Neural Network)

This is the model which learns about the historic evidence of disease spreads and from current situation to make prediction of spread of disease.

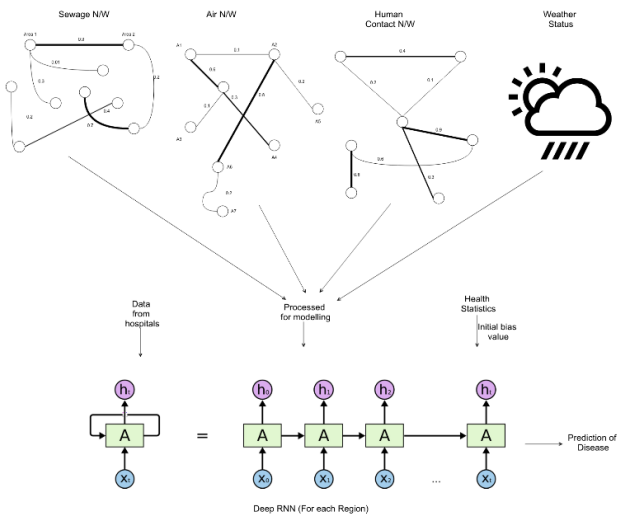

Fig. 3. Deep RNN Abstract Design

To train and make the Deep RNN periodically learn we need to provide it 3 inputs:

1) Data from hospitals: which is the input to the Deep RNN.

2) Three networks: Sewage Network, Air Network and the Human Contact Network are weighted as per their impact. For e.g. Sewage network weights is in terms of the number of drainage, open/close drainage and WQI of the sewage constitutes to its weight.

3) Health statistics: from NGOs constitute for the bias weight in the RNN.

The RNN is constructed as a deep network which trains itself periodically as well as from past history records.

C. Visualization & Public Notification

By means of varying visualization we make the health control board understand the situation of diseases and its spread probability to different regions. The public in the region where outbreak can occur are notified by Push Notifications/SMS.

V. INFORMATION GATHERING

To train and test our deep recurrent neural network we need data, in fact a lot of data. The data was collected from the following sources.

A. Data from Hospitals and Health Clinic:

We targeted hospitals present all over Mumbai region. The hospitals are governmental hospitals. The sole purpose of selection of such hospitals was to get hold of the diseases which are spread among the common people who usually admit to such hospitals. These hospitals maintain information about each patient which is enrolled in their hospital in digital format and regularly provide this information to NGOs and health institute for research operations. The hospitals has been targeted such that they cover entire region of Mumbai. This helps us in actually predicting the spread of disease and finding the next outbreak, We got hold of this information directly from the hospitals.

The information we collected from this hospitals primarily include:

- The name of the hospital
- The area of the patients address
- The disease which was detected to the patient
- The date- time when patient was enrolled
- The date-time when patient was discharged

B. Spread Network

We needed three set of key information to define our spread network. Most of the disease spread through the medium of air, human contact, and sewage/drainage water. The information was collected from the following sources:

1) CPCB – Central Pollution Control Board:

The Central Pollution Control Board (CPCB), statutory organization, was constituted in September, 1974 under the Water (Prevention and Control of Pollution) Act, 1974. Further, CPCB was entrusted with the powers and functions under the Air (Prevention and Control of Pollution) Act, 1981.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

Air Quality Monitoring is an important part of the air quality management. The National Air Monitoring Program (NAMP) has been established with objectives to determine the present air quality status and trends and to control and regulate pollution from industries and other source to meet the air quality standards. It also provides background air quality data needed for industrial siting and towns planning.

Information regarding AQI (Air Quality Index) and air flow direction was collected from here for Mumbai for the entire year of 2015. It provide open source easily accessible data.

2) MCGM – Municipal Corporation of Greater Mumbai

The Mumbai Metropolis has a historic tradition of strong civic activism dedicated to the cause of a better life for all its citizens. And it's the Municipal Corporation of Greater Mumbai (MCGM), the primary agency responsible for urban governance in Greater Mumbai.

Information regarding sewages/drains in Mumbai was collected from here. For each sewage/drain the category of nalla (major/minor), the count of nalla between regions and the WQI (Water Quality Index) of the major nallas was acquired from here. It provide open source easily accessible data. The data is updated twice a year.

3) The Tata Institute of Social Sciences (TISS), Mumbai

The Tata Institute of Social Sciences (TISS) was established in 1936. As a mean of the project Engendering Mumbai's Suburban Railway System statistics related to the frequency of commutation between different stations was obtained. This helped us to estimate the mass movement of people from one place to another place thus carrying with them diseases which spread through human contact.

The same project also estimates the movement of people from one place to another through medium of road. The estimation is on basis of the traffic and BEST buses which commute between these regions.

These information is combined by us to get our needed information of flow of human form one region to another in Mumbai

Information regarding human movement from one region to another and the probability and frequency of travel was obtained from a recent project carried out by TISS. It shows recent statistics of the year 2015.

V. USE CASE

A. Health Control Board

They are capable of Adding a new Disease and its Spread route in the system. The need to login/register in the system to be a part of the system. The key role is Visualization. The board can visualize the spread of disease and see the prediction made by the system.

B. Hospitals

It provides data of the patients in a periodic basis which is directly fetched from its databases.

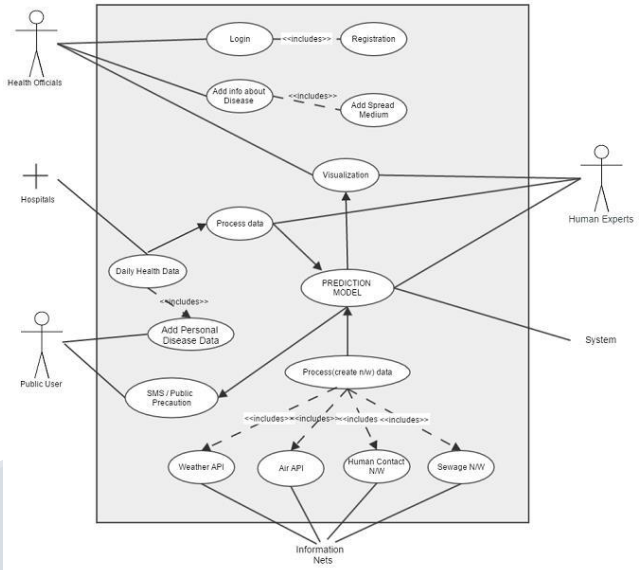

Fig. 4. Use Case for SpateAugur

C. API's

The play a role in providing real time data such as Weather data, Sewage data, Air data and the estimated human contact data (based on statistics of travelling).

D. System

It is responsible for maintaining a prediction model and making predictions. It takes data from Hospitals and APIs for the same.

E. Public Users

They get notification of the epidemic which are about to cause in their region, along with precautionary measures which they must take.

CONCLUSION

The paper is aimed at predicting the epidemic in local areas. For this, deep neural network is used to train the system with the past data and make a prediction for the future. The data for training is obtained from the local hospitals and government databases. The system after training makes predictions which is sent to the local residents of the areas using SMS notifications and let the health authorities maintain and manage the predictions using visualization. The graphs and data of the prediction can also be viewed on a website. Thus, it aims to create mitigation plan which can be used by government agencies to help them carry out the actions to tackle disease

outbreak in a better and efficient manner. This is a proposal paper which we propose to implement.

Acknowledgement

We would like to acknowledge CPCB – Central Pollution Control Board, MCGM – Municipal Corporation of Greater Mumbai, The Tata Institute of Social Sciences (TISS), Mumbai and all hospitals who provided the essential data and data set for training the RNN model.

REFERENCES

[1] R. Chakoumakos, "Predicting Outbreak Severity through Machine Learning on Disease Outbreak Reports", Stanford Web, 2010.
[2] Y. Zhang, W. Cheung and J. Liu, "A Unified Framework for Epidemic Prediction based on Poisson Regression", IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 11, pp. 2878-2892, 2015.
[3] V. Yu and L. Madoff, "ProMED-mail: An Early Warning System for Emerging Diseases", Clinical Infectious Diseases, vol. 39, no. 2, pp. 227-232, 2004.
[4] Online detection and quantification of epidemics. BioMed Central Ltd, 2007.
[5] "VoroGraph: Visualization Tools for Epidemic Analysis - IBM", Researcher.watson.ibm.com, 2016. [Online]. Available: http://researcher.watson.ibm.com/researcher/view_group.php?id=5992. [Accessed: 28- Oct- 2016].
[6] "Epi Info™ | CDC", Cdc.gov, 2016. [Online]. Available: http://www.cdc.gov/epiinfo/index.html. [Accessed: 28- Oct- 2016].
[7] "Programs - EcoHealth Alliance", EcoHealth Alliance, 2016. [Online]. Available: http://www.ecohealthalliance.org/programs. [Accessed: 28- Oct- 2016].
[8] "Espacenet - Home page", Worldwide.espacenet.com, 2016. [Online]. Available: https://worldwide.espacenet.com/. [Accessed: 28- Oct- 2016].
[9] "patft", Patft.uspto.gov, 2016. [Online]. Available: http://patft.uspto.gov/netahtml/PTO/index.html. [Accessed: 28- Oct- 2016].