



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume3, Issue2)

Epidemic Disease Detection and Forecasting: A Survey

Siddharth Satish

ISE, R.V College of Engineering
sidd.satish95@gmail.com

Smitha G. R

ISE, R.V College of Engineering
smithagr@rvce.edu.in

Abstract: The objective of this paper is to present a precise methodology to project and forecast the spread behaviors of epidemic diseases well before they occur. With numerous cases of widespread outbreaks of epidemics being reported in densely populated areas, these methodologies can help restrict the outbreak only to a small confined area. This would ensure that a better coping mechanism is provided to study the spread of infectious diseases and adequate control mechanisms are provided to reduce casualties in the form of human death. Through the course of this paper, we hope to develop a well-defined prediction methodology that can predict the likeliness of an individual being affected by a particular epidemic through assessing of early symptoms.

Keywords: Epidemic Diseases, Disease Forecast, Pathogen Detection, Biosurveillance, Epidemic Breakout Detection.

I. INTRODUCTION

With the recent upturn in tourism and global nature of modern life, diverse societies are evolving to form new communities of people in different parts of the world. With this form of vibrancy, people are placing an increased importance on health and the threat of diseases. The health officials are doing everything they can to obtain timely information on epidemic outbreaks. They are putting several mechanisms in place to identify the first signs of an epidemic and to contain them as soon as possible. The conventional methods that are being utilized to obtain data from practitioners and aggregate them is a slow process. Officials and governments are looking in tandem to find more efficient means to predict the outbreak of an epidemic and to control its impact.

The timeline of epidemic infections and outbreaks is usually divided into several stages. The first stage is called as the Susceptible state. A transition happens when a person who is infected and a person who is in the Susceptible state come in contact with one another. When a host organism becomes infected, the communicable agent begins to reproduce inside of the host, to enable the infected host to spread the epidemic onto another Susceptible organism. This time between which a host is infected and is capable of spreading the infection onto another host is called as the Latency state.

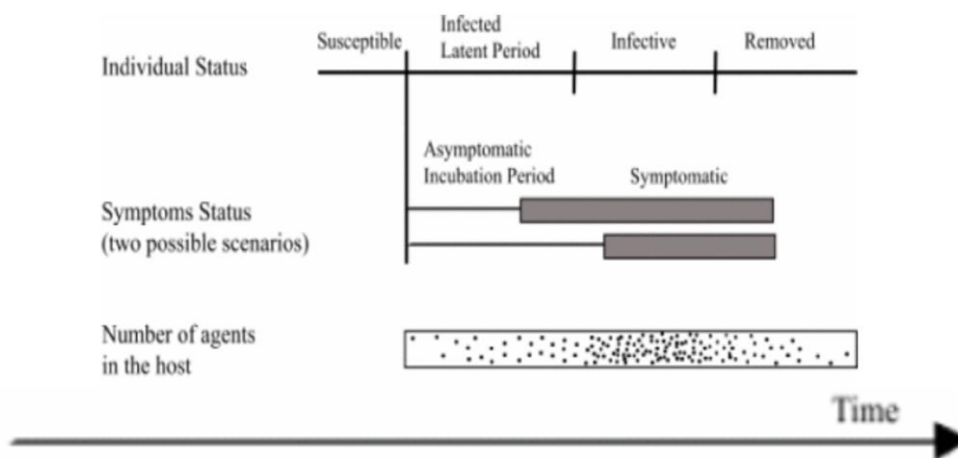


Fig 1: Epidemic Infection Timeline

The period before which the infected host starts to show early signs of the epidemic is called as the Incubation state. It is possible for a host to be infectious and spread the epidemic onto other organisms, but show no signs of the epidemic itself. This may be as a result of an immunity that the individual may possess.

Specific symptoms can then be tracked based on the type and degree of the outbreak. Particular conditions can also be tracked by examining permutations of these symptoms. We attempt to predict the spread pattern of the epidemic based on the number of cases that are initially reported in a sporadic manner and also by utilizing social media symptom data.

II. METHODOLOGY

Several models are used to represent the symptom data that is obtained from various sources. Some of the most used data models are represented below

A. Risk Assessment Model

It correlates specifics of a particular place – such as humidity, rainfall, weather which are the risk factors to calculate the risk of the epidemic.

B. Epidemic Prediction Model

It assigns a probability to each epidemic based on specific data inputs related to the epidemic and other variables. It provides a prediction of whether an epidemic is likely to occur under the current circumstances.

C. Spatial Model

It predicts the outbreak of the epidemic in that particular geographical location based on the relationship between the outbreak and the geographic factors. They keep changing as time progresses as the geographic factors of a place might vary considerably over time.

D. Dynamic Model

It illustrates how a particular epidemic is likely to surge through a given population. It includes finite parameters such as restriction of movements, the presence of an immunity. It can also be used to find out when the pathogenicity of an infectious disease may change.

E. Event Detection Model

It tries to identify the outbreak of an epidemic by analysis of real-time clinical data and social media symptomatic data. Through this, it tries to identify likely causes of a certain epidemic or first signs of a particular infectious disease.

All the data that is collected at the initial stage of an epidemic or in situations where an outbreak of an epidemic is expected is aggregated and is subjected to the above data modeling process. The modeled data is then used to classify the epidemic on various aspects and identify sections of populations that are most likely to be affected by such an event of an outbreak.

The Bayesian Model was also incorporated in order to further model and analyse the data. Bayesian Model is a type of Spatial Model that is mainly utilised to predict the portions of the population that are likely to be affected next. This model has seen large amounts of success after being applied to other infectious diseases such as Pneumonia and Influenza Virus. The Bayesian Model was able to make accurate predictions related to the percentage of people likely to be affected in the next window frame.

III. RESULTS

Once the data modeling of the data was complete, it was revealed that close to 84% of all transmission of infectious diseases causing epidemics is either through direct contact or through vector-borne means. Water and other means of spread of diseases accounted for the rest.

The Bayesian Model was used to make weekly and monthly predictions on several infectious diseases which have the potential to finally lead to an epidemic. The communicable diseases that were chosen for this experiment were Influenza Virus, RS- Virus, Rhinovirus, and Pneumonia.

TABLE I
ROOT MEAN SQUARED PREDICTIONS

Disease	Weekly	Monthly
Influenza A	0.0721	0.1713
RS-virus	0.1299	0.2437
Rhinovirus	0.1498	0.2545
Norovirus	0.1243	0.1863
Mycoplasma pneumoniae	0.1424	0.1735

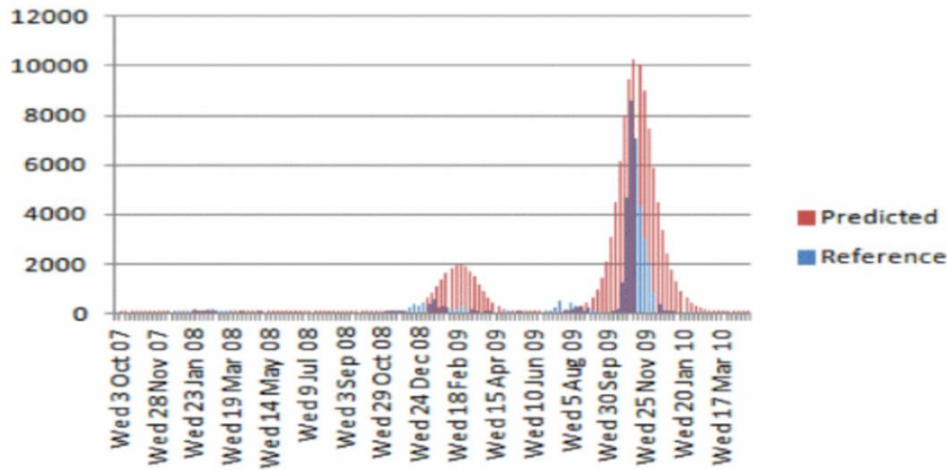


Fig 2: Prediction of Influenza Virus using Bayesian Model

The results showed that the Bayesian Model was closer in its weekly predictions but wavered further away during the monthly predictions.

CONCLUSION

We have discussed the various data modeling techniques to classify large-scale data that is obtained from various sources. Each of the data is modeled and is assigned a particular level of likeliness based on the spatial features and other parameters. We have also provided a well-defined and well-trusted means for prediction of the outbreak of the epidemic. The Bayesian Prediction Model takes into account likelihood and other important factors as it tries to assess the weekly outbreak percentages of the epidemic.

The main contributions of this paper are three-fold. Firstly, we have developed a neoteric technique to collect large amounts of symptomatic data from various sources and model it easily. Secondly, we have developed a generalized epidemic prediction approach based on the current activity of the epidemic. Finally, we attempt to forecast future activity and spread patterns of the outbreak based on patterns viewed from the Bayesian Model.

ACKNOWLEDGEMENTS

This research was carried out in collaboration with the Information Science and Engineering Department of R.V. College of Engineering.

REFERENCES

- [1] Daphne Lopez, M. Gunasekaran, B. SenthilMurugan, *Spatial big data analytics of influenza epidemic in Vellore, India*, IEEE International Conference on Big Data, 10.1109/BigData.2014.7004422, 2014.
- [2] NadraGuizani, ArifGhafoor., *Modeling and evaluation of disease spread behaviors*, Wireless Communications, and Mobile Computing Conference, 10.1109/IWCMC.2014.6906491, 2014.
- [3] Konstantinos P. Exarchos et al, "Prediction of coronary atherosclerosis progression using dynamic Bayesian networks", 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2013.
- [4] Courtney D. Corley et al, "Disease Prediction Models and Operational Readiness", in *PMC3960139*, 2014.
- [5] Nicholas Then, et al, "DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response", in *plos - /journal.pone.0155417*, 2016.
- [6] João Andrade, Artur Arsenio, *Epidemic Spreading Over Social Networks Using Large-scale Biosensors: A Survey*, Volume 5, Pages 922-931, Procedia Technology, 2012.
- [7] Lauren N. Carroll, et al, "Visualization and analytics tools for infectious disease epidemiology: A systematic review", Volume 51, Pages 287-298, Journal of Biomedical Informatics, 2014