Map-Reduce based Modeling and Dynamics of Infectious Disease

Chinmayee Mohapatra, Leena Das, Siddharth Swarup Rautray, Manjusha pandey
School of Computer Engineering, KIIT University, Bhubaneswar, Odisha , India
chinmayeemohapatra19@gmail.com, ldasfcs@kiit.ac.in, siddharthfcs@kiit.ac.in, manjushafcs@kiit.ac.in

***Abstract*— The rapid increase in population creates an issue in handling and analyzing the population data for the traditional data base management system. So Big data came into figure to solve the issue. Big data is more efficient in comparison to the traditional data base system due to some of its basic features like Velocity, Veracity, Volume , Verity and Value. Day by day the disease are growing and becoming harmful to the society irrespective of treatments that are available. Infectious disease is caused by infectious agents including Viruses, Prions, Bacteria, Nematodes etc. Population dynamics is a branch of life science which includes the study of population size and age composition of dynamic system and the biological and environmental process managing them.This proposed paper consider an infectious disease i.e Dengue Fever and divides the population dynamic into three parts those are High Vulnerable, Mid vulnerable, Low vulnerable to Dengue. Then suggest the preventive measure like Forced preventive for high Vulnerable, Efficient preventive measure for Mid vulnerable and Delayed preventive measure for Low vulnerable areas by utilizing the benefits of big data.**

Keywords—Infectious Disease; Population Dynamics; Hadoop; Map-Reduce;

I. INTRODUCTION

Data is a raw and unorganized form if information. Now a days data are coming in a very large and very fast manner. Which becomes difficult for the normal traditional database so big data came into figure.The main characteristics of big data are volume, velocity, veracity, verity, Value. Volume indicates the amount of data that may in Gbs or Tbs or more than that. Velocity is considered as how fast the data are created and collected. Verity means how many types of data are present that may be structured, semi-structured or unstructured. Veracity means the truthfulness of data. Value means, how much the data is valuable.

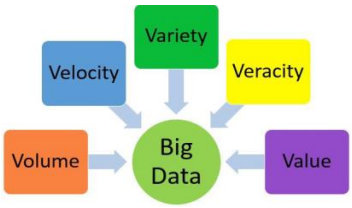


Fig-1. 5V's of Big Data

Infectious disease are caused by pathogenic micro-organisms like bacteria,viruses, parasites, fungi [1]. Infection spreads directly or indirectly from one person to another. Among many infectious diseases, dengue fever is one , which can convert a life to death. Dengue fever is a mosquito borne infection that causes a severe flue-like illness. So it is also called as break-bone fever. Initially Dengue is transmitted to humans by the Aedes aegypti and Aedes albopictus female mosquito[1]. Once the virus enters into human body then, it multiply and increase inside the human body.

Population dynamics is the study of any type of change in size or structure of population depending on the time. The main factors that affects population dynamics are rates of reproduction, death and migration[2]. As the world is changing and also the technology changing , so for more efficient and accurate result Big data is used to store the data and the process the data. Population also varies according to geographic region like population is less in forest and hills in comparison to other areas. Cause of infectious disease varies with population depending on the atmosphere.

The main aim of this paper is to diagnosis the cause of dengue and suggestion of preventive measures to the population affected by dengue. Here, population is considered as three different three parts like Highly vulnerable, mid vulnerable and low vulnerable to dengue. And then suggest some prevention procedure like forced prevention, efficient prevention and delayed prevention respectively.

II. OVERVIEW INFECTIOUS DISEASE(DENGUE)

Dengue fever is commonly found in urban parts of subtropical and tropical areas. The main cause of dengue fever is mosquito bite. It's seriousness varies from mild to severe. The severe form of dengue includes dengue shock syndrome(DSS) and Dengue hemorrhagic fever(DHF). DHF occurs due to the virus like DV-1, DV-2, DV-3 and DV-4[3]. DV is a positive-stranded encapsulated RNA virus[4]. The Fig.1 shows the life cycle of dengue virus , how they become adult from a egg and increase their family.

Fig 2. Life cycle of Dengue Virus

© IEEE 2017. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

After 4-7days the symptoms of mosquito bite come into figure. Dengue can be transmitted from one infected person to another uninfected person through mosquito. DSS is a type of dengue which is the worst form that leads to death. The Fig.2 shows how the infection spreads from one person to another through the mosquito. The symptoms of dengue fever are[5]

- High fever
- headache
- Pain behind the eyes
- Aching muscles and joints
- Vomiting
- Bleeding from your mouth/gums, Nosebleeds
- Internal bleeding
- Small blood spots under your skin
- Heavy bleeding
- Blood vessels leaking fluid

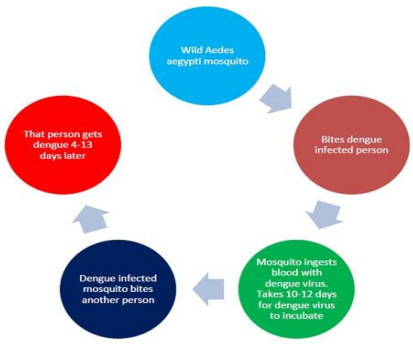


Fig 3. Dengue infection life cycle

The fig-3 shows how a person gets infected by the dengue virus. Initially the mosquito intakes blood then the virus grow inside the mosquito. Now the when it bites a human it ingest the infection inside the human body. Now the virus grows inside the body of human. In the next step the human was declared as infected person. Now again the mosquito bites the infected person then goes to bite a healthy person and transfers the virus inside the healthy person. Now the healthy person found with dengue infection. In this way the infection spreads in the population through virus.

III. Population Dynamics

Population dynamics is the study of life science which deals with the study of change in population with time[6]. It also affected by the atmosphere and disease. If in a particular location a harmful disease is seen then automatically the population of that area gets affected. Like Dengue is a very harmful disease which has a bad effect on the population. The Dengue fever mostly seen in the forestry areas where a large no of tree present. And also the dengue mosquito are seen in the most areas where the water is stored without proper care.

IV. Related Work on Big Data in Disease Diagnosis

A overview about Big data analysis for cardiovascular disease detection system using map reduce technique, its causes and the basic challenges of big-data in Cardiovascular. Also identifies the big-data capabilities to support health-care[7]. Similarly another paper Proposed a mechanism for A review to predictive methodology to diagnose chronic kidney disease[8]. There are many papers where big data is used to help people for diagnosis of disease and also provides some preventives and also recommend drugs for the patients. One of the model considers the population and imagines the possible future scenarios and determine corresponding trajectories of infected population in different regions. Then, the information is used to find an optimal distribution of bed across countries/regions [9]. As now a days no one have the interest to read books or some materials, So some video games are invented for the better under standing on the cause of dengue and also how to prevent the dengue from spreading[10]. The human is designed as a robot and the dengue mosquito is designed bit larger size then the human[11].

V. Proposed Prototype

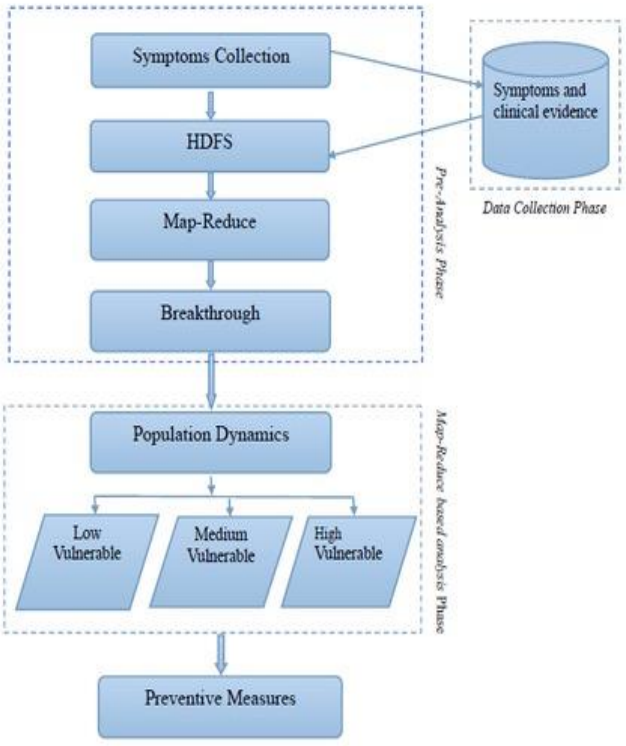


Fig 4. Proposed Mechanism

Here in fig-4 a mechanism is proposed for the better prevention of dengue. In the initial stage we collect various symptoms from various resource. The information can be collected from hospitals, from people as feedback , From the social media or from search engines i.e. from which place more search according to dengue fever then people in that area are more effected. In the next step the data that are collected stored inside the HDFS(Hadoop Distributed File System). As

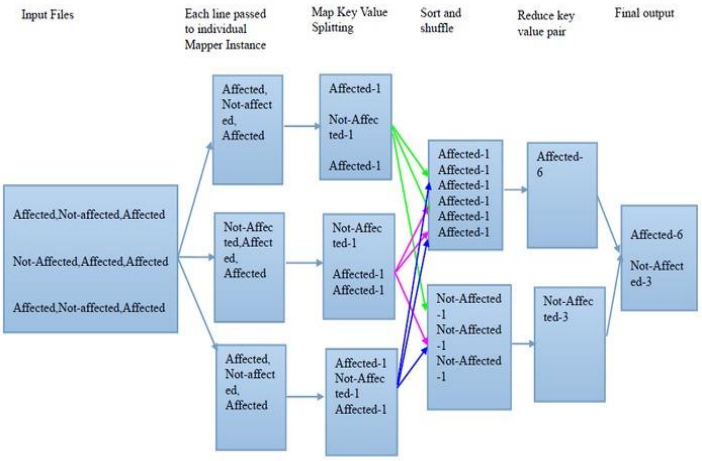
here the data stored into small small files[12]. Now Map-Reduce technique is used to manipulate/analyze the data. Below we have taken an example of map-reduce frame work.

Data can be provide to the system using tools for cleansing, abstraction and logical storage for successful mining in the further steps. It's become a difficult task for database to manipulate such a large volume of data which is coming with high velocity . So Hadoop come into the figure, for efficient manipulation of data-set. Hadoop has two main components are HDFS and Map-Reduce. This paper use Map-Reduce methodology to save data in a globally classified manner according to the population dynamics.

The proposed implementation done using some of the current existing tools and techniques of big data analysis, some of them like Hadoop, Map-Reduce. As the infection is increasing day by day and more cases of dengue are coming into figure. So the data set is also increasing rapidly so it becomes difficult to handle the data for the traditional data base. Hadoop consist of two main components that are storage and processing. Storage part deals with HDFS, Hadoop Distributed File System. HDFS leverages large block size and moves computation where data is stored. A file can be replicated into a number of times. HDFS breaks the larger files into smaller blocks and stores them in a cluster. The processing part is done by Map-Reduce frame work. Map-Reduce programming helps to process massive amount of data in parallel in a efficient manner.

Map-Reduce works in two phases mapping phase and Reducing phase. By using Map-Reduce programming here the count of people is done, who are affected by the dengue fever according to the population dynamics. In Map-Reduce the data set is divided into small chunks then those chunks are mapped by the mapper function in a parallel manner. The output of the mappers are automatically shuffled and stored by the frame work. The out put is sorted on the basis of key element i.e dengue. Here the data set is divided into three parts That are Highly vulnerable, Mid vulnerable and Low vulnerable to dengue fever. And provides the preventive measures to them respectively. Forced preventive measure for the area or population which is highly affected by dengue or which is highly vulnerable to dengue fever. Efficient preventive measure for the medium vulnerable area and delayed preventive measure for low vulnerable area/population to Dengue.

Here in Fig.5 example of Map-Reduce is given for better understanding of the working process of Map-Reduce. In this one file is taken into consideration, which belongs to a particular area. There are two results affected and not affected, affected means one person is affected by dengue and not-affected means that person is not affected by dengue. And according to the result it can be decided that the area is highly vulnerable or mid vulnerable or low vulnerable.



VI. Conclusion

Health care is a very vast platform for various research works. The application of big data in health care improves the process of disease diagnosis and also provides more efficient preventives for the disease. This proposed system can be used by the health wale-fare organization for saving lives from dengue fever on the basis of common signs and symptoms. And based on the result marks the areas into highly vulnerable, Medium Vulnerable and low Vulnerable to dengue. Due to the utilization of big data it become possible to predict the dengue fever within a sort period of time,which helps to take steps to save life from the spreading of dengue virus. The same mechanism can be used for the diagnosis of other disease which will help to protect the human-life and also the animals life also.

References

[1] D. Saikia and J. C. Dutta, "Early diagnosis of dengue disease using fuzzy inference system," 2016 International Conference on Microelectronics, Computing and Communications (MicroCom), Durgapur, 2016, pp. 1-6.

[2] A. Zvoleff and J. Ahumada, "Understanding the link between population dynamics and biodiversity conservation through remote sensing and gridded population data integration," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, 2015, pp. 2560-2563.

[3] Gupta, Nivedita, et al. "Dengue in India." The Indian journal of medical research 136.3 (2012): 373.

[4] D. Guhar, G. Mustafa, S. F. Rehmani and R. Bilal, "Immunodiagnostic of dengue fever: Primary and secondary infections," 2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, 2016, pp. 75-76.

[5] W. T. Sesulihatien, S. Sasaki and Y. Kiyoki, "Ecological context-dependent analysis and prediction using MMM: A case of dengue fever disease," Electronics Symposium (IES), 2015 International, Surabaya, 2015, pp. 227-232.

[6] J. Barreiro-Gomez, N. Quijano and C. Ocampo-Martinez, "Distributed resource management by using population dynamics: Wastewater treatment application," Automatic Control (CCAC), 2015 IEEE 2nd Colombian Conference on, Manizales, 2015, pp. 1-6.

[7] G. Vaishali and V. Kalaivani, "Big data analysis for heart disease detection system using map reduce technique," 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), Kovilpatti, India, 2016, pp. 1-6.

[8] A. Batra, U. Batra and V. Singh, "A review to predictive methodology to diagnose chronic kidney disease," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2016, pp. 2760-2763.

[9] Evans, Robin J. and Musa A. Mammadov. “Predicting and controlling the dynamics of infectious diseases.” CDC (2015).

[10] T. M. Porcino, E. Strauss and E. G. Clua, "Hugo against dengue: a serious game to educate people about dengue fever prevention," 2014 IEEE 3nd International Conference on Serious Games and Applications For or Health (SeGAH), Rio de Janeiro, 2014, pp. 1-5.

[11] Mohan D Kashinkunti, Shiddappa, Dhananjaya M."A Study of Clinical Profile of Dengue Fever in a Tertiary Care Teaching Hospital " Scholars Journal of Applied Medical Sciences (SJAMS) ,2013; PP.280-282.

[12] What is big data?, in: Big Data Now: 2012 Edition, 1st Edition, O'Reilly Media, Inc, 1005 Gravenstein Highway North, Sebastopol, CA 95472, 2012,pp. 3.