

Geo Visualization and monitoring of diseases in India using GIS

Lakhya Hari Chandana P, Waquar Haidry, Sivagami G

School Of Computer Science and Engineering Srm University, Chennai, Tamil Nadu, India.

Abstract

The geographical information system (GIS) is an assigned framework used for collecting and analysing geographical data. This can be used to map and represent disease specific information and its spread in ranges where the infection reconnaissance is finished. Epidemiology information has become part of health data repository in health care management system. Epidemiology along with GIS helps increase the efficiency of health care services. Information visualization method can potentially be employed to assist the work of epidemiologists and other health care professional. Mapping the spread of communicable diseases in areas where the task of disease surveillance encompasses temporal elements such as changes in climate, land use and population movements. Communicable disease such as Malaria, HIV, Hepatitis, and Measles and tuberculosis are charted out on the map using colour codes.

The proposed system can be used as a tool for pattern recognition with data set measuring epidemics.

Keywords: GIS, geographical information system, epidemiology

1. Introduction

In recent times due to the advent of internet, information on any topic is easily available; but accurate and specific details of the type of disease, it's prevalence in an area, it's geographical distribution and analysis of its spread is not. Epidemiologists have identified certain co-relation between a disease and its distribution spatially by analysing the adjacent climatic condition and environment, both natural and manmade.

This can be successfully applied to identify pattern of disease outbreak and occurrence by surveillance at community national as well as international levels. The various attributes of the disease such as incidence, persistence, distribution, spread and mortality rates or its long term implication can be taken into consideration. The age group affected, the specific population which is plagued by a form of health condition can be analysed and this helps immensely in the field of health care on a large scale. Any health condition that affects a human being can be traced and correlated with various adjoining attributes that play a role in causing that disease.

This information is recognised and used by many organizations in the field of medical case studies and for improving the health condition on a particular location. Various analysis tools such as bar chart, pie chart, are used. For mapping specially, Google map is used for geo-visualization using colour code such as blue to red indicating least to most affected locations. The study of diseases and their numbers on the states of India is performed in this system. This enables a clear perception of the actual numbers in concise and coherent way. Users can have quick access to the geographical data using specialized information layers. This method makes it accessible to the general public instead of skilled domain experts. Keyhole mark-up language is used by Google map to facilitate this.

When large volumes data is to be mapped for epidemiological information exploration, it becomes a challenging task to correctly analyse and decipher each large data value with its corresponding attribute and weights of each attributes spatially. Spatial cluster analysis is commonly used in disease surveillance and a number of clustering is available. The most

classical spatial clustering algorithm is SOM, it achieves mapping of higher and lower dimensional map space. It provides a compact representation of the data distribution and has wide application for higher dimensional data.

Web-Epi: it is a web GIS service that locates disease alerts and presents locations geographically for the health epidemiology unit. The aim of web-Epi is to explore patterns of different variables of a disease. When considering large multidimensional datasets, it becomes difficult to extract patterns and discover new knowledge; hence a novel health GIS application for the need of analysis data is required by health researchers.

Previously, SOM has been used for pattern recognition and classification, it is more effective. K-means based visualization algorithm was also developed to achieve clustering of variables epidemiology-wise.

In the proposed system TOP-K algorithm has been used to order dataset values at a faster rate than the previous algorithms could. A sort will take $O(n \log n)$ time, which when leading with lots of items, will take several minutes using the heap sort technique for finding the top k item can be done in $O(n \log k)$ time, which is much faster.

Strategically it is done by going through the entire list once and keeping a list of top k elements that are found so far, efficiency of this search increases when the smallest element is in this top k. It is easily maintainable to find different between a linear search and a heap search. To find the top k elements, let us analyse the time complexity of the following code:

```
#!/usr/bin/python
import heapq
import random
import time

def create Array ()
    Array = range (10 * 1000 * 1000)
    Random. Shuffle (array)
    Return array
```

```

def linear Search (big Array, k)
    Return sorted (big Array, reverse=True) [k]
def heap Search (big Array, k)
    Heap = []
    # Note: below is for illustration. It can be replaced by
    # heapq.nlargest ( big Array, k )
    For item in big Array:
    # if we have not yet found k items, or the current item
    is larger than
    # the smallest item on the heap,
    If len (heap) < k or item > heap[0]:
    # if the heap is full, remove the smallest element on
    the heap.
    If len (heap) == k: heapq.heappop( heap )
    # add the current element as the new smallest.

heapq.heappush (heap, item)
Return heap
Start = time. Time ()
Big Array = create Array ()
Print "Creating array took %g s" % (time.time () -
start)
Start = time. time ()
Print linear Search (big Array, 10)
Print "Linear search took %g s" % (time. time () -
start)
Start = time. time ()
Print heap Search (big Array, 10)
Print "Heap search took %g s" % (time. time () - start)

```

For the linear search for a sample input table 7.1514 sec. and the heap search for the same table 2.6637 sec. and it is four times faster than linear search.

2. Statement

The underlying idea is to enhance the usability of the datasets that are derived from the epidemiological surveillance, to analyse, cluster and find the pattern among the numbers.

Previously this information was restricted to only a few, but it can be made accessible to the general public.

3. AIM and objective of the study

The study focuses on in-depth revision and need for accurate, up-to-date and exhaustive surveillance and data collection at the ground level. The major task is to take into consideration each and every factor that plays a part in the equation of human and animal health.

The objective is to join in the pieces of puzzle, which is the root cause of each epidemic. It has been noted that each disease requires specific factors to propagate and remain persistent at a location, and the environment at the location has a major role to play. The food, air, soil quality and the staple diet being consumed, all play a role in one's state of health.

4. Scope of the Study

The vastness of the proposed system can only be understood with the amount of research and information collection, each one requires. The scope of the study next to unlimited, from the use in medical research, studies, to governmental organizations, epidemiologists, doctors and even to the general public. Each sector using the details and finding of the study uniquely in benefiting themselves and the elevating the health conditions present at the present scenario.

5. Application Components

- List of five diseases:
 1. HIV (2009 - 2015)
 2. Malariya (2011 - 2015)
 3. Tuberculosis (2012 - 2015)
 4. Measels (2013 - 2015)
 5. Hepatitis (2012 - 2014)
- Analysis
 1. Bar chart
 2. Pie chat
 3. GIS map method
 4. Detailed disease information

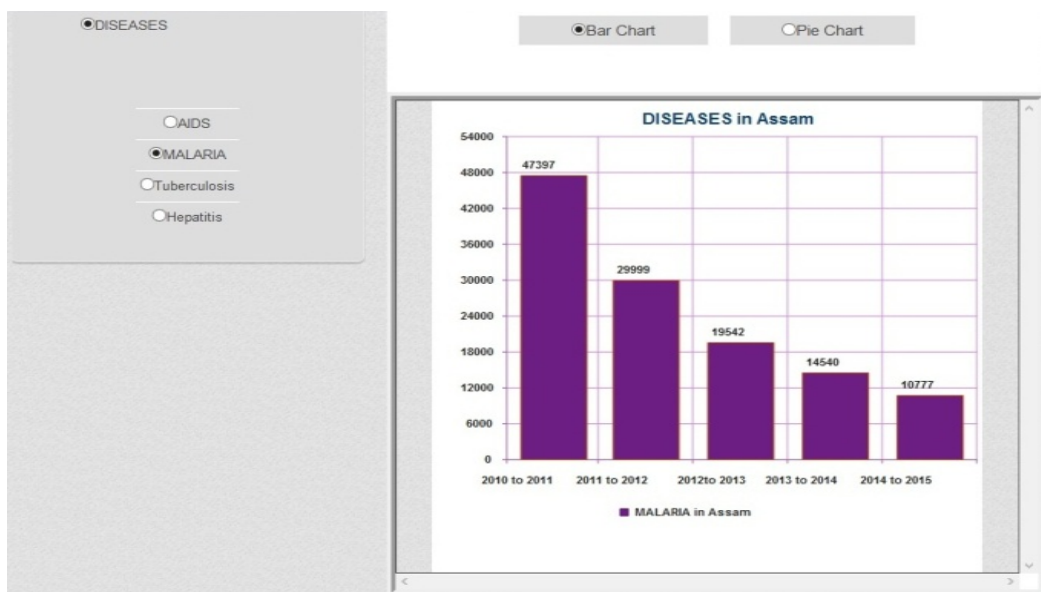


Fig 1: Bar chart

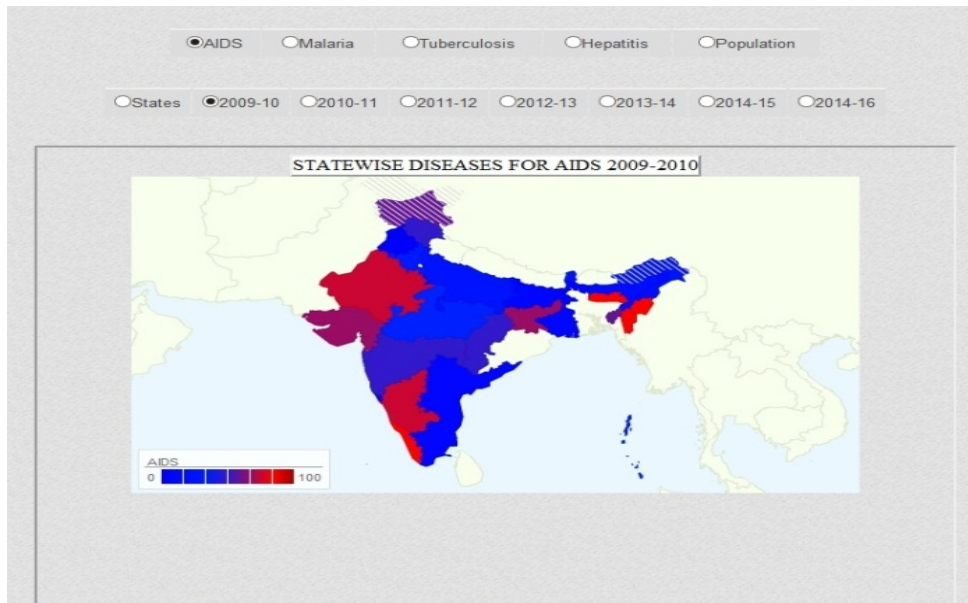


Fig 2: GIS map method

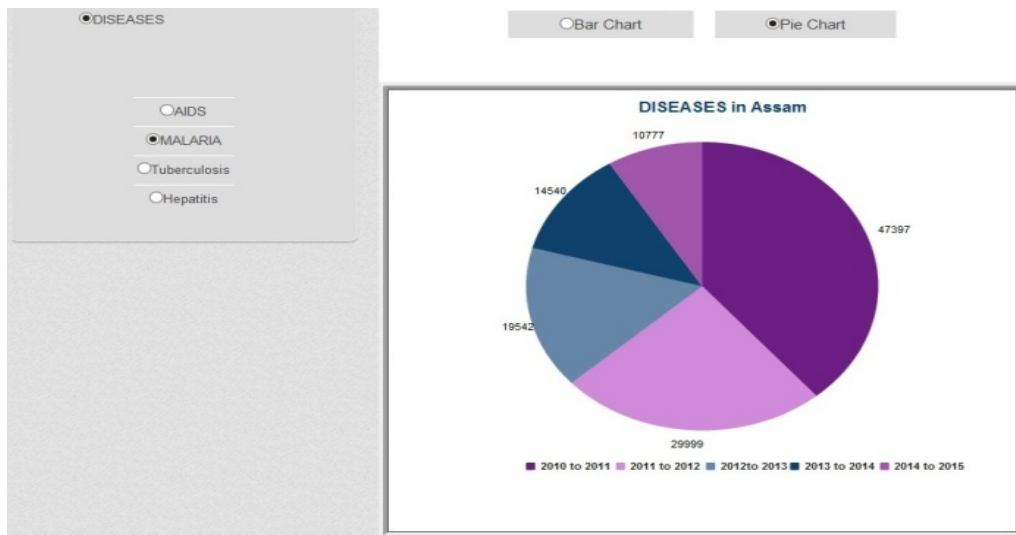


Fig 3: Pie chat

SEARCH KEYWORD HERE....

AIDS

About AIDS.txt Summary:

- The Symptoms of HIV/AIDS.
- HIV infection comes in three stages.
- The first stage is called acute infection or seroconversion, and it typically happens within two to six weeks after exposure or becoming infected.
- This is when the body's immune system puts up a fight against HIV.
- The symptoms of acute infection look similar to those of other viral illnesses and are often compared to those of the flu.

Fig 4: Detailed disease information

6. Materials and Method

The system is developed using software tool such as visual studio for the user interface using *c#* language and java script html and SQL server is used for the data base management on the backend.

• Methods

1. The back end module, deals with the management of datasets in the form of table using SQL commands.
2. The front end design is created using HTML and the functionality is enhance using *c#* language.
3. Date is mined from the table that are input to the backend to display the bar chart and pie chart of corresponding datasets.
4. Google map is used for GIS aspect of the project i.e. visual perception of data.
5. Use of TOP-K sorting algorithm for enhancing efficiency while working with the large datasets of higher values.

7. Result

1. An efficient way to perceive numerical data for a better analysis of health care and wellness statistics in Indian states.
2. Access to the corresponding disease detailed is achieved by even those who are not experts in the field.

8. Discussion Of Result

The results of the analysis can further be enhanced by the use of in-depth and exhaustive surveillance methods, which also includes the district-wise and also city-wise statistics. Patient details such as residence, age and also the month in which the disease affected the patient. This will further lead to a clarified approach towards disease monitoring and controlling and equips the government for making informed decisions in health care.

9. Conclusion

Modern tools have brought in a major development in which data can be perceived in multiple files and formats for a better understanding of the user regarding the given information.

From converting numerical values, to visualizing in geo-spatial mapping and also in various other methods help to reach at a thorough analysis of a disease much faster.

The implication of such a system can be harnessed by improving surveillance methods and bringing health care system from the traditional pen-paper reports to an electronic/software medium. Maintaining electronic records of health data helps from the gross root such as hospital and cities to international level such as WHO and other relevant institutions.

10. References

1. Self-Organizing Map Methodology, Google Maps Services for Geographical Epidemiology Mapping" Jingyuan Zhang, Hao Shi and Yanchun Zhang.