

Extraction of Disease Occurrence Patterns Using MiSTIC: Salmonellosis in Florida

Vipul Raheja* and K. S. Rajan

International Institute of Information Technology Hyderabad (IIIT-H), Hyderabad, India

Objective

This work leverages spatio-temporal data mining (ST-DM), the MiSTIC (Mining Spatio-Temporally Invariant Cores)[1,6] method for infectious disease surveillance, by identifying

- Extent of spatial spread of disease core regions across populations-scale of disease prevalence
- Possible causes of the observed patterns-for better prediction, detection & management of infectious disease & its outbreaks

Introduction

Infectious diseases, though initially tend to be limited geographically to a reservoir; a subsequent spatial variation in disease prevalence (including spread & intensity) arises from the underlying differences in physical-biological conditions that support pathogen, its vectors & reservoirs. Different factors like spatial proximity, physical & social connectivity, & local environmental conditions which add to its susceptibility influence the occurrence[2].

In Disease management, analysis of historical data over various aspects of geography, epidemiology, social structures & network dynamics need to be accounted for. Large amounts of data raise issues of data processing, storage, pattern identification, etc. In addition, identifying the source of disease occurrence & its pattern can be of immense value.

ST-DM of disease data can be an effective tool for endemic preparedness[3], as it extracts implicit knowledge, spatial & temporal relationships, or other patterns inherent in such databases.

Here, Core Region is defined as a set of spatial entities(eg.counties) aggregated over time, which occur frequently at places having high values in a defined region (considering areas of influence around them)[1].

Methods

Here, MiSTIC algorithm detects spatio-temporally invariant cores with respect to disease occurrence. It involves both a spatial analysis step to detect focal points & a spatio-temporal analysis over the time period of study to identify core regions, which are then classified as – CHD, CLD & CND. They refer to Cores with High, Low and No (mostly random) dominating points respectively based on frequency of occurrences of disease. The predominantly occurring focal points capture the localized behavior of the disease whereas the neighborhood constraints capture the nature (dynamic or non-dynamic) of the event.

Results

County-level annual data of Salmonellosis incidence from Florida Department of Health [3] covering a period of 50 years (1961-2010) is used.

Two types of cores were identified based on type of neighborhood - Contiguous (CC) & within a defined Radius (CR).

Table1 shows the analysis of counties according to valid frequency criteria for both CC & CR ($t=2$) & their sub-classification.

Salmonellosis etiology shows that it is caused by tainted food, hygiene, local environment etc. which are largely sanitation-related [4]. Taking the level of urbanization [5] as a proxy for sanitation, it can be seen from Fig.1, 12 of 19 cores occur in rural counties.

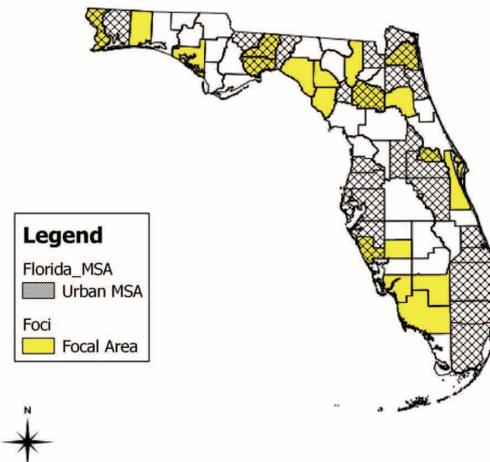
Conclusions

It is observed that CC is better indicator of cores than CR, implying that Salmonellosis manifests itself in a highly localized manner. Thus, use of MiSTIC is promising & provides a way for identifying disease “hot-spots”. It also provides valuable insight into the understanding of disease prevalence in different regions based on their history over space and time.

Classification of Core Polygons

	CC	CR
CHD	5	1
CLD	11	13
CND	3	5

0 2e+06 degrees



Map showing overlay of Metropolitan Areas and Cores

Keywords

Disease Cores; Salmonellosis; Spatio-Temporal Data Mining; Patterns

References

- K. S. Rajan, Vipul Raheja: Spatio-Temporal Mining of Core Regions - Study of rainfall patterns in Monsoonal India. 11th IEEE International Conference on Data Mining Workshops (ICDMW) 2011, pp.30-37
- Chris Bailey-Kellogg et al.: Spatial Data Mining to Support Pandemic Preparedness. ACM SIGKDD Explorations Newsletter. 2006; 8(1):80-82
- <http://www.floridacharts.com> [20/5/2012]
- <http://www.cdc.gov/healthypets/diseases/salmonellosis.htm> [2/7/2012]
- <http://www.census.gov> [25/5/2012]
- K. S. Rajan; MiSTIC; MS Thesis. IIIT Hyderabad

*Vipul Raheja

E-mail: vipul.raheja@research.iiit.ac.in

