

Daniel B. Neill

Assistant Professor of Information Systems

Heinz School of Public Policy and Management

August 31st, 2007 3:00pm

Guillermo Marinero

Research Topic

Detecting Epidemic breakouts

Research Problem

How to speed up the detection of epidemic outbreak using stochastic methods?

Problem Statement

Given a set of dimensions construct a model capable of detecting if an epidemic outbreak is on the verge of reaching its tipping point.

Dimensions are characteristics that can be used as indicatives of an epidemic, such as amount of cold medicines sold, number of doctor visits, symptoms reported by doctors or any other dimension considered of interest by an epidemiologist.

Tipping point is considered the time when an disease can be considered an epidemic.

Problem Description

The best technique to stop an epidemic is to stop it before it reaches a “tipping point”. Detecting a potential outbreak of a disease in its early stages is not trivial. The data to be analyzed is of tremendous proportions (dimensions and quantity) and by the time it gets processed by the current technique, the epidemic might be all too obvious and causing havoc among the population. A fast and quick response before the disease reaches critical mass is the best approach but currently not the most feasible. Applying stochastic methods is sometimes more viable than using traditional deterministic methodologist. The professor is exploring how machine learning technique can speed up the process without introducing an unreliable process.

Computer Science Perspective

In this problem computer science is known as Machine Learning. By using stochastic methods the problems become manageable to the point that an analysis that took about a week to complete it is now conducted in 20 minutes. Mostly, the techniques used are clustering and learning algorithms.

Disciplines actively involved

Epidemiology

Description of Disciplines Involved

Several epidemiologist are involved helping define the problem. Defining the problem involves delimiting a natural breakdown of unit areas and the attributes to be observed and collected in order to detect anomalies.

Stakeholders

Epidemiologist: The most obvious stakeholder as they are the ones that have the need for information

Pharmaceutical Companies: Their sales data of drugs is a good metric to use when looking for outbreaks, unfortunately HIPA and industrial rivals need for information might curb how willing a company might of disclosing data.

GIS Companies: The break up of maps into manageable chunks of data is possible by GIS technologies. GIS companies might start a race to capture that niche market.

Health Policy Makers: The result of detecting outbreak might shed some light on faulty health policies, so policy makes might be a little reluctant to massive the use of this system.

FEMA and similar organizations: When Katrina struck New Orleans two years ago, FEMA was slow on reacting, ever since the agency has sought ways of addressing emerging crisis. This tool could enable a quicker reaction in cases like controlling West Nile virus or a Flu outbreak.

References

General Information about the researcher

<http://www.cs.cmu.edu/~neill/>

A KDD on clustering

<http://www.cs.cmu.edu/~neill/presentations/Neill-KDD.ppt>

Time space clusters

<http://www.cs.cmu.edu/~neill/papers/sss-kdd05-abstract.html>

By: Guillermo Marinero

Updated: Guillermo Marinero, September 27, 2007