T-Cube Web Interface as a tool for detecting disease outbreaks in real-time: A pilot in India and Sri Lanka

IEEE RIVF International Conference on Computing and Communication Technology
Research, Innovation, and Vision for the Future

November 02, 2010
Vietnam National University, Hanoi, Vietnam

Nuwan Waidyanatha
LIRNEasia
Email: nuwan@lirneasia.net
http://www.lirneasia.net/profiles/nuwan-waidyanatha
Mobile: +8613888446352 (cn) +94773710394 (lk)

This work was carried out with the aid of a grant from the International Development Research Centre, Canada.
Problem the RTBP is trying to solve in developing countries

Reduce 15 - 30 day delays to Minutes

Introduce Comprehensive active surveillance

- **Black arrows**: current manual paper/postal system for health data collection and reporting
- **Red lines**: RTBP mobile phone communication system for health data collection and reporting
Data collection, Event detection, and Situational-wareness/Alerting

1. Record health data
2. Digitize indicators
3. Transmit records to database
4. Analyze data
5. Detect adverse events
6. Generate/issue alert
7. Transmit alert to mobile device
8. Receive health alerts

Public Health Indicators

m-HealthSurvey

Sahana Alerting Module Interface

GSM network

Health Officials

T-Cube Web Interface

Epidemiologists

Skip the paper
Research Question: “Can software programs that analyze health statistics and mobile phone applications that send and receive the health information potentially be effective in the early detection and mitigation of disease outbreaks?”

**Data Collection**

*mHealthSurvey* is a data entry software that works on any standard java-enabled mobile phone. A typical record contains the patient visitation date, location, gender, age, disease, symptoms, and signs. Data is transmitted over GPRS cellular networks.

**Event Detection**

*T-Cube Web Interface* (TCWI) is an Internet browser based tool to visualize and manipulate large spatio-temporal data sets. Epidemiologists can pin down a potential outbreak of, for instance, a gastrointestinal disease among children in the Sevanipatti PHC health division.

**Alerting**

*Sahana Alerting Module* (SAM) allows for the generic dissemination of localized and standardized interoperable messages. Selected groups of recipients would receive the single-entry of the message via SMS, Email, and Web.
T-Cube Web Interface (TCWI) by Auton Lab

- AD Tree data structure
- Trained Bayesian Networks
- Fast response to queries
- Statistical estimations techniques
- Data visualization over temporal and spatial dimensions
- Automated alerts
Pre-Screening using Massive Temporal Scan

Minimum Support: 0

Start Date: 09/22/2009
End Date: 08/31/2010

Scan Option: UpperTail
Baseline: All Data
Reference Type: All Past
Reference Window: 20

Use Adaptive Baseline
Use Drill-Down

Run Screening

Screening Results

The following screenings are available:

- Escalating Fever Diseases
- Escalating Non-communicable Diseases
- Escalating Notifiable Diseases

This screening shows data about diseases classified as high-impact and notifiable. It will show instances of these diseases that have been statistically unusual recently (21 queries).

<table>
<thead>
<tr>
<th>Id</th>
<th>Query</th>
<th>Date</th>
<th>PValue</th>
<th>Window</th>
<th>Count</th>
<th>Expected Count</th>
<th>Ranked PValue</th>
<th>Alt Date</th>
<th>Alt PValue</th>
<th>Alt Window</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>disease(Chicken_Pox),age_grp=(15-20),ds_priority=(Notifiable)</td>
<td>09/31/2010</td>
<td>6.79F-03</td>
<td>21</td>
<td>4</td>
<td>0.06</td>
<td>2.859E-03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>disease(Chicken_Pox),age_grp=(15-20),location=(Sandalankanwa),ds_priority=(Notifiable)</td>
<td>09/31/2010</td>
<td>3.09E-04</td>
<td>21</td>
<td>3</td>
<td>0</td>
<td>2.899E-03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>age_grp=(15-20),location=(Sandalankanwa),ds_priority=(Notifiable)</td>
<td>09/31/2010</td>
<td>2.78E-03</td>
<td>21</td>
<td>4</td>
<td>0.14</td>
<td>2.899E-03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>location=(Sandalankanwa),ds_priority=(Notifiable)</td>
<td>09/31/2010</td>
<td>4.69E-03</td>
<td>21</td>
<td>4</td>
<td>0.51</td>
<td>2.899E-03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>disease(Chicken_Pox),location=(Sandalankanwa),ds_priority=(Notifiable)</td>
<td>09/31/2010</td>
<td>5.29E-03</td>
<td>21</td>
<td>9</td>
<td>0.22</td>
<td>2.899E-03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>age_grp=(15-20),ds_priority=(Notifiable)</td>
<td>09/31/2010</td>
<td>1.00E-02</td>
<td>21</td>
<td>9</td>
<td>3.26</td>
<td>2.899E-03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

On 08/31/2010, we note that the counts in last 21 days are significantly higher than expectation. The chance of seeing anything more extreme is about one in 26,000 cases.

Dump Query       Use For Drill-Down       Save To List       Name: Escalating Notifiable Disease       Save Results to File       Print List

Rate usefulness of the result: 5
User Name: Comments:

Submit Feedback
T-Cube Web Interface – Spatio – Temporal Presentation
Overview of the T-Cube data structure and computations

**Complaint:**
Contingency Tables can reach enormous sizes (number of cells) if the underlying data is highly dimensional and involved variables can assume many different values.

**Raw data**

<table>
<thead>
<tr>
<th>SriLankan?</th>
<th>ArrackDrinker?</th>
<th>CricketFan?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

- N=7 Records
- M=3 Attributes

**Replace contingency table with a tree**
It represents the same counts of co-occurrences

- **Data**
  - A1=1
  - A2=2
  - A1=2
  - A2=3
  - A1=3
  - A2=4

- **Vary A1**
  - C=0
  - C=1
  - C=2
  - C=3
  - C=4

- **Vary A2**
  - A1=1
  - A2=1
  - A1=2
  - A2=2
  - A1=3
  - A2=3

**“Vary” nodes contain queries in which specific attributes are instantiated**

**“Count” nodes store queries and the corresponding counts of the records of data matching them**

---

**All Historical Data**

**Today’s Environment**

**What should be happening today?**

**Today’s Cases**

**What’s strange about today, considering its environment? And how significant is this?**
Research design

- 24 Health Sub Center Village Nurses
- 4 Public Health Center Sector Health Nurses, Health Inspectors, and Data Entry Operators
- 1 Integrated Disease Surveillance Program Unit of the Deputy Director of Health Services
- 12 District/Base Hospitals and Clinics
- 15 Sarvodaya Suwadana Center Assistants
- 4 Medical Officer of Health divisions & 1 Regional Epidemiology Unit
- Kurunegala District, Wayamba Province, Sri Lanka
Evaluation of TCWI

- **Replication study** :: Sri Lankan Weekly Epidemiological Return (WER) reports published at www.epid.gov.lk notifiable disease counts tabulated by District was semi synthesized by distributing the weekly counts as daily counts taking day-of-week effect, gender distribution, and age representations.

- **Study the reliability and effectiveness** :: significant events detected by T-Cube is compared with the ground truth and also weighed on the response actions or inaction

- **Competency exercise** :: injected fake data over a period of 5 days and the subjects, unaware of the prefabricated events, were asked to detect most significant events

- **T-Cube Acceptance** :: a questionnaire was designed based on the Technology Assessment Methodology (TAM) and was subject to TCWI users as well as health official associated with T-Cube who make decisions on whether or not to take action

- **Cost analysis** :: compare the economic efficiencies and cost effectiveness between present detection/analyses system and T-Cube
Replication study using synthesized WER data


- Synthesized the data to match that similar to the RTBP dimensions by distributing the district weekly aggregates

- day-of-week visitation densities (M - F)

- female to male ration

- age-groups (0-5, 6-14, 15-20, 21-45, 46-65, above 65)
Replication study using Sri Lanka WER data 2007 - 2009

Food poison spike as detected by spatial scan around Feb 15, 2007 in Nuwara Eliya, which was reported as outbreak by health department.

In addition TCWI detected spikes in Kandy and Vauvniya areas.

Spatial scan is run by 7 days windows size.
Another Food poison spike as detected by spatial scan around June 17, 2009 in Nuwara_Eliya, the same location.

Spatial scan is run by 7 days windows size.
Dengue Fever
Seasonal and spatial pattern

May 1, 2007

Aug 30, 2007

May 21, 2008

April 15, 2009

May 28, 2009

www.lirneasia.net
Progression of Dengue Fever outbreak in April - June 2009

First day an elevated global score noted, lead by region Kandy

Situation in Kandy intensified, together other regions

Southern Regions began to see increased cases

Southern region continue to see progression, while other region subsides
Most frequently occurring wide spreading infectious disease outbreaks

These findings are from TCWI's spatial scan algorithms

Common Cold, Sivaganga District – India, 18 outbreak episodes to date with over 23,188 cases.

Worm Infestation, Sivaganga District – India, 13 outbreak episodes to date with over 1,236 cases.

Dysentery, Sivaganga District – India, 5 outbreak episodes to date with over 1,541 cases.

Common cold is the most popular but gastrointestinal infectious are, relatively, the most visible

Cough, Kurnegala District – Sri Lanka, 11 outbreak episodes to date with over X cases.

Respiratory Tract Infection, Kurnegala District – Sri Lanka, 09 outbreak episodes to date with over X cases.

Tonsilitis, Kurnegala District – Sri Lanka, 07 outbreak episodes to date with over X cases.

Respiratory infectious diseases, a correlated with environmental factors, are the most common
Trends in selected noncommunicable disease
These findings are from TCWI's statistical estimation and pivot table analysis methods

Hypertension (High Blood Pressure) has a linearly increasing trend over the one year period in both countries with Females and Males over 45 years of age showing to be the most vulnerable. The trend in India shows an unusual increase between March and May 2010; while the reported cases are consistent throughout the year in Sri Lanka.

Diabetes-Mellitus has a linearly increasing trend over the one year period in both countries with Indians over 40 years of age and Sri Lankan over 45 years of age to be the most vulnerable groups.

Given that the Male to Female ratios, approximately, in Tamil Nadu, India and Kurunegala, Sri Lanka are both 1 : 1; statistics to date show females to be more susceptible to the above mentioned life style diseases.
Trends in selected noncommunicable disease
These findings are from TCWI's statistical estimation and pivot table analysis methods

Arthritis and Rheumatoid-Arthritis has a linearly stagnant trend over the one year period in both countries with Males over 45 years of age and Females over 35 years of age to be the most susceptible in India; similarly Males over 45 and Females over 31 years of age to be the most vulnerable groups.

Asthma has a linearly decreasing trend over the one year period in both countries; the trend shows the counts to increase during the rainy season, India: Sept'09-Jan’10 and Sri Lanka: Nov ‘09-Jan ‘10. In India, only males over 45 years of age are affected but females in all age groups are affected. Both Male and Female over 31 years of age are in Sri Lanka are equally vulnerable.

Given that the Male to Female ratios, approximately, in Tamil Nadu, India and Kurunegala, Sri Lanka are both 1 : 1; statistics to date show females to be more susceptible to the above mentioned life style diseases.
TCWI Competency Assessments with Injected Synthetic data

Used “Epigrass” to generate synthetic data with a SEIR model

Susceptible ➞ Exposed ➞ Infected ➞ Recovered

With a Network Flow

Injected 3 sets of data
1) Notifiable disease :: Dysentery
2) Other-Communicable disease :: ADD
3) Syndrome :: Fever, Pain, RTI
TCWI Actual Usage by Health Departments

3 of 14 potential users spend less than 30 minutes each time once a week on detection analysis; remaining 9 did claim to be too busy to use TCWI.

75% of the 9 Sri Lankan users spend more than 30 minutes each time every day of the week on detection analysis.
TCWI Preferred functions

India health officials' primary preferences are screening for fever, other-communicable diseases, and using the pivot table.

Sri Lanka health officials' primary preferences are screening the notifiable, fever, and other-communicable diseases.
Quality of the digitized data

Data quality = Signal to Noise Ratio (SNR); i.e. number records with errors/records submitted

The 23% noisy data in India subsided to less than 4% after informing the consequences of false detections (SNR for sub intervals: 0.18, 0.40, 0.31, 0.04, 0.07)

Assistants in Sri Lanka with no formal health training and no affiliation to the hospitals/clinics had no incentive to correct the 45% errors (SNR for sub intervals: 0.58, 0.30, 0.53, 0.57, 0.17)

1 Low quantities of data received from Health Sub Centers
2 Volume of records were better after including Primary Health Centers
3 Holiday effect: no records received
4 Learning curve getting medical officers to adopt to the new procedures of writing the diagnosis
5 Release of mHealthSurvey v1.3 with better predictive text

www.lirneasia.net
# Digitizing problems that affect the categorical data

<table>
<thead>
<tr>
<th>PROBLEM</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of synonyms</td>
<td>goal fever = jail fever = typhus fever&lt;br&gt;dementia = memory loss&lt;br&gt;enteric fever = typhoid fever; encephalitis = meningitis</td>
</tr>
<tr>
<td>Inserting symbols and extra spacing between words</td>
<td>body ache = body-ache, body pain = body pain</td>
</tr>
<tr>
<td>Changing the order of words</td>
<td>muscle weakness = weakness in muscle&lt;br&gt;stomach pain = pain in the stomach</td>
</tr>
<tr>
<td>Inclusion and exclusion of adjectives</td>
<td>‘severe’ memory loss vs memory loss</td>
</tr>
<tr>
<td>Using local language when terms are unknown</td>
<td>leg vettuthal [Tamil] = broken leg [English]</td>
</tr>
<tr>
<td>Using preposition and conjunctions between terms</td>
<td>nasal stuffiness or sneezing over bleeding with abdominal pain</td>
</tr>
<tr>
<td>Long sentences</td>
<td>not able to identify color white and shining patches without any sense</td>
</tr>
<tr>
<td>Mistaking treatment for diagnosis</td>
<td>oral pills, remove catheter, vaccination</td>
</tr>
<tr>
<td>Prepopulated instructions in text boxes propagating to database”</td>
<td>please specify details&lt;br&gt;specify symptoms</td>
</tr>
<tr>
<td>UK vs USA spelling</td>
<td>diarrhoea = diarrhea&lt;br&gt;vomitting = vommitting</td>
</tr>
<tr>
<td>Test results as symptoms or signs</td>
<td>BP 140/90, BP 120/100</td>
</tr>
<tr>
<td>Singular vs plural</td>
<td>fit / fits, cut / cuts</td>
</tr>
<tr>
<td>Inconstancies in the verb tense</td>
<td>faint, fainted, fainting</td>
</tr>
</tbody>
</table>

**SNOMED-CT**

**LOINC**
Observations of the data digitizing uncertainties

No observations for India in this quadrant → Data submitted by health workers in India is consistent

Diseases with higher counts but occurring only in a single location; hence suspected of possible mis-coding by Sri Lankan assistants

Fever greater than 7 days concentrated in February and March of 2010, mainly from a single location, during the non rainy season

The likelihood of a measles outbreak emerging only in a single location without spreading to other areas, given that it is a viral disease, is highly unlikely.

The assistant entering the data had submitted data for “Toxide vaccine” as Tetanus.

These diseases occurred only once in one location

Uniformity of geographic distribution of disease cases
( low: concentrated in a few locations, high: spread over)
Timeliness of data submission

Timeliness = submitting the patient’s record the same day as the patient visitation

Finding time to complete the records without disrupting current work flow was a significant barrier for real-time data submission (sub interval delay rates 0.28, 0.09, 0.21, 0.38, 0.44, 0.48, 0.68).

Data entry assistants have no other role besides digitizing records but see delays proportional to the patient visitation counts (sub interval delay rates: 0.10, 0.27, 0.25, 0.36, 0.53, 0.21).

¹ Users with dysfunctional phones where sharing and were sending data on the weekends or when friends phone was available for borrowing.
Distribution of expenses calculated in terms ToC

Comparison of expenses in relation to the data collection, event detection, and alerting components

India and Sri Lanka invest very little or no resources on event detection and alerting

RTBP can reduce direct expenses, increase timeliness, and introduce detection and alerting components

India and Sri Lanka can reduce overall expenses by 50% and 30%, respectively, with ICT

Comparison of expenses in relation to the health facility, health department, and health workers

Bulk of the expenses are in health departments invested for data collection and consolidation, which can be reduced by RTBP with the introduction of mHealth at the point of care

Invest more in alerting to empower health workers with information on the state of affairs of the health in their regions

[ Existing (IN) = present system in India (Integrated Disease Surveillance Program); Existing (LK) = present system in Sri Lanka (Disease Surveillance and Notification Program); RTBP (IN), RTBP (LK) = Real-Time Biosurveillance Program in India and Sri Lanka, respectively]
Incremental Cost Effectiveness Ratios (ICER)

Going from Existing system to RTBP

**Collection** – mHealthSurvey can reduce the expense and enhance system to collecting all data opposed to reporting a small subset of diseases once only

**Detection** – TCWI can introduce syndromic and disease surveillance opposed to no rapid detection analyses done at present

**Alerting** – not an existing practice but health workers will be better informed of the public health status in the geographic area for better response and mitigation
“We can use this rich and comprehensive dataset and analysis tools for our annual planning, now our planning relies on professional perception and not necessary data.”
- Deputy Director Planning, Kurunegala District, Sri Lanka, Consulted (06.10.09)

“Epidemiologists want TCWI to facilitate the old ways of monitoring outbreaks based on thresholds opposed to statistical significance. For example, a single case of Malaria is regarded as an outbreak in India, which requires response actions.”
- Deputy Director of Health Services, Sivaganga District, India, Consulted (19.12.09).

“It is important to monitor escalating fever cases, notifiable disease cases, and common clusters of symptoms.”
- Regional Epidemiologist, Kurunegala District, Sri Lanka, consulted (19.12.09).

“Medical Officers, Nurses, Health Educators, etc, who are interested in learning of outbreaks see the benefit and are happy with TCWI detection analysis methods but the staff at the Integrated Disease Surveillance Program are not ready to accept change and want to stick to the traditional system unless state or national level Authorities mandate it.”
- Senior Project Officer, RTBI, India, consulted (19.08.10).
“Pharmacists’ perceptions are such that a separate computer should be given for detection analysis and they do not want to share their computers, which are used for medicine and birth information.”
- Senior Project Officer, RTBI, India, Consulted (08.07.2010).

“RTBP’s real-time biosurveillance capabilities will enhance the present day passive or non-active passive surveillance to an active surveillance system.”
- Wayamba Provincial Director of Health Services, consulted (07.07.10).

“All cases can be viewed in TCWI in real-time for detecting outbreaks swiftly, which otherwise would take several days before the hospitals/clinics send the notification paper forms, by which time the patient may be dead or discharged.”