Agenda

- Motivation
- Problem description
- Methods for signal generation
- Challenges for signal generation from social media
- Conclusions and future work
Motivation: The Threat

• Infectious diseases lead to public health threats

• Contributing factors:
  ▾ global warming, globalization, air traffic
  ...mutation, spread, ...

• Technologies to early detection are required

 ➩ Early Detection by Disease Surveillance
 ➩ Broad range of information sources
How fast is communication?

# Generating signals: Event ➔ Indicator ➔ Signal

<table>
<thead>
<tr>
<th>Event</th>
<th>Indicator</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measles outbreak</td>
<td>Instances where the term measles or related symptoms are mentioned</td>
<td>Number of mentions exceeds specific threshold</td>
</tr>
</tbody>
</table>
Methods for Signal Generation

• Problem:
  ■ existing algorithms developed for processing structured, verified data
  ■ search for aberrations of the observed values from an expected level

• Research Question:
  ■ What needs to be considered when applying them to social media data?
Methods for Signal Generation

- Simple Thresholds
  - calculate threshold from data from a previous period
- Regression Analysis (Farrington)
  - Builds a linear model
- Quality Control Measure (CUSUM)
  - Focus on several consecutive periods and sum up the aberrations in one particular direction
M-Eco: Generated Signals

Browse Signal Definitions

Predefined Signals

<table>
<thead>
<tr>
<th>Disease</th>
<th>Location</th>
<th>Date range</th>
</tr>
</thead>
<tbody>
<tr>
<td>bird flu</td>
<td>China</td>
<td>11/05/2009-11/05/2009</td>
</tr>
<tr>
<td>bird flu</td>
<td>China</td>
<td>25/04/2009-25/04/2009</td>
</tr>
<tr>
<td>bird flu</td>
<td>China</td>
<td>24/04/2009-24/04/2009</td>
</tr>
</tbody>
</table>

Tags: patients, flu, tamiflu, hong, kong

Unsupervised Signals

Recommended Signals

<table>
<thead>
<tr>
<th>Disease</th>
<th>Location</th>
<th>Date range</th>
</tr>
</thead>
<tbody>
<tr>
<td>h1n1, swine flu</td>
<td></td>
<td>25/04/2009-01/05/2009</td>
</tr>
</tbody>
</table>

Tags: 150mg/day, 75mg, accordingly, added, adjustments, adult, allowed, and, announced, antalya, avian, bird, case, city, clinical, doctors, dosage, dose, doses, doubling, flu, h1n1, h5n1, higher, hong, hospital, hui, ill, in, kong, long, march, model, pandemic, patients, pneumonia, prince, rate, severely, standard, swine, tamiflu, the, treatment, viral, virus, wales, who, world

Documents

Source: Blog  Open in new window: 🔗

... The standard adult course of Oseltamivir Tamiflu for seasonal flu has long been 2 - 75mg capsules per day, for 5 days. ... Or a total of 10 pills. ...
Challenges for Generation from Social Media

• Problems occur already for indicator-based data:
  ▪ unequal distribution of the population
  ▪ different diagnostic and reporting behavior

• Data is very noisy

• Interpretation is not always straightforward

• How to aggregate data from different sources?

• Which symptoms make a health event?

• Incompleteness of data
Conclusions and Future Work

• Testing biosurveillance algorithms on social media data
• Signal generation from multiple sources
• Studying open questions
• Creation of evaluation data set for comparing approaches
References


Challenges for Signal Generation from Medical Social Media

Dr. Kerstin Denecke
Forschungszentrum L3S
denecke@L3S.de

http://www.meco-project.eu

Thank you for your attention!
The Medical Ecosystem – Personalized Event-based Surveillance

Funded by the European Union
January 2010 – July 2012
Project coordination: L3S Research Center

Can the Health of a Society’s Individuals Be Isolated from Today’s Web?