Le nuove sfide dell’epidemiologia digitale: implicazioni e potenzialità in sanità pubblica

Caterina Rizzo

Istituto Superiore di Sanità
Digital epidemiology in communicable diseases surveillance

• Early detection of health threats
• Surveillance of CD through social media data (e.g. Twitter). However, still limited use in Italy
• Evaluation of population compliance with national recommendations (e.g. vaccine hesitancy)
“passive” techniques: monitoring search engine queries
correlazione
query su influenza – dati CDC

J. Ginsberg et al., Nature 457, 1012 (2009)
Explore flu trends around the world

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. Learn more »

Download world flu activity data - Animated flu trends for Google Earth - Compare flu trends across regions in Public Data Explorer
National flu outbreak widens
FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.

- Red: Google Flu Trends
- Blue: CDC data
- Green: Flu Near You

Google’s algorithms overestimated peak flu levels this year.
Google Flutrends ha **sovrastimato**
il trend influenzale reale.
“passive” techniques: monitoring user-generated content
Using Twitter to track the flu

Researchers find a better way to screen the tweets

Phil Sneideman / ☎ January 24, 2013
Posted in Health, Science+Technology
Tagged computer science, flu, influenza, tweets, twitter, mark dredze

Sifting through social media messages has become a popular way to track when and where flu cases occur, but a key hurdle hampers the process: how to identify flu-infection tweets. Some tweets are posted by people who have been sick with the virus, while others come from folks who are merely talking about the illness. If you are tracking actual flu cases, such conversations about the flu in general can skew the results.

To address this problem, Johns Hopkins computer scientists and researchers in the School of Medicine have developed a new tweet-screening method that not only delivers real-time data on flu cases, but also filters out online chatter that is not linked to actual flu infections. Comparing their method, which is based on analysis of 5,000 publicly available tweets per minute, to other Twitter-based tracking tools, the Johns Hopkins researchers say their real-time results track more closely with government disease data that takes much longer to compile.

"When you look at Twitter posts, you can see people talking about being afraid of catching the flu or asking friends if they should get a flu shot or mentioning a public figure who seems to be ill," said Mark Dredze, an assistant research professor in the Department of Computer Science who uses tweets to monitor public health.

These U.S. maps indicate the Twitter systems rate of influenza in each state in the first week of January (higher flu rates are marked with darker red). They show that the country is awash in a high flu rate in 2013 (the bottom map), yet was relatively unscathed during the same week in 2012 (the top map).
My doctor is so nice. She always makes me feel like everything is going to be ok. By the way I got the flu
Think I got the flu. Hate this. These body aches are no joke
anche “flu” OR “influenza” su Twitter
sovrastimano il trend influenzale
Influenza-like illness (ILI)

– Sudden onset of symptoms

AND

– At least one of the following four systemic symptoms:
  – Fever or feverishness
  – Malaise
  – Headache
  – Myalgia

AND

– At least one of the following three respiratory symptoms:
  – Cough
  – Sore throat
  – Shortness of breath
((fever) OR (feverishness) OR (malaise) OR (headache) OR (myalgia)) AND ((cough) OR (pharyngitis) OR (dyspnea))
Think I might be getting sick. Body starting to ache and cough like crazy. Ain't nobody got time for that. Need to get my flu shot.
<table>
<thead>
<tr>
<th>TT</th>
<th>NT</th>
<th>pattern</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>dyspnea</td>
<td>difficult</td>
<td>is the medical term for</td>
<td>dyspnea &lt;is the medical term for&gt; difficult breathing &lt;or&gt; shortness of breath</td>
</tr>
<tr>
<td></td>
<td>breathing</td>
<td>is the technical word for</td>
<td>dyspnea &lt;is the technical word for&gt; difficult breathing</td>
</tr>
<tr>
<td>shortness of</td>
<td>or</td>
<td>is another name for</td>
<td>dyspnea &lt;is another name for&gt; shortness of breath or breathlessness, and is often a symptom of cardiac problems</td>
</tr>
<tr>
<td>breath</td>
<td></td>
<td></td>
<td>dyspnea (&lt; shortness of breath ) can be caused by many conditions, including asthma and cardiomyopathy</td>
</tr>
<tr>
<td>troubled</td>
<td>, a medical word</td>
<td></td>
<td>dyspnea &lt;, a medical word for&gt; troubled breathing is a common symptom produced due to lung and heart problems</td>
</tr>
<tr>
<td>breathing</td>
<td>for</td>
<td></td>
<td></td>
</tr>
<tr>
<td>breathlessness</td>
<td>or</td>
<td></td>
<td>dyspnea &lt;is another name for&gt; shortness of breath &lt;or&gt; breathlessness, and is often a symptom of cardiac problems</td>
</tr>
</tbody>
</table>
correlazione dati CDC – twitter ILL

0.97

Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control

Marcel Salathé*, Shashank Khandelwal

Center for Infectious Disease Dynamics, Department of Biology, Penn State University, University Park, Pennsylvania, United States of America

Abstract

There is great interest in the dynamics of health behaviors in social networks and how they affect collective public health outcomes, but measuring population health behaviors over time and space requires substantial resources. Here, we use publicly available data from 101,853 users of online social media collected over a time period of almost six months to measure the spatio-temporal sentiment towards a new vaccine. We validated our approach by identifying a strong correlation between sentiments expressed online and CDC-estimated vaccination rates by region. Analysis of the network of opinionated users showed that information flows more often between users who share the same sentiments and less often between users who do not share the same sentiments than expected by chance alone. We also found that most communities are dominated by either positive or negative sentiments towards the novel vaccine. Simulations of infectious disease transmission show that if clusters of negative vaccine sentiments lead to clusters of unprotected individuals, the likelihood of disease outbreaks is greatly increased. Online social media provide unprecedented access to data allowing for inexpensive and efficient tools to identify target areas for intervention efforts and to evaluate their effectiveness.


Editor: Lauren Ancel Meyers, University of Texas at Austin, United States of America

Received May 10, 2011; Accepted July 30, 2011; Published October 13, 2011

Copyright: © 2011 Salathé, Khandelwal. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MS acknowledges funding from Society in Science: the Branco Weiss fellowship. http://www.society-in-science.org/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: salathe@psu.edu
“active” techniques: crowdsourced disease tracking
Surveillance Sans Frontières: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project

John S. Brownstein*, Clark C. Freifeld, Ben Y. Reis, Kenneth D. Mandl

The Opportunity
As developed nations continue to strengthen their electronic disease surveillance capacities [1], the parts of the world that are most vulnerable to emerging disease threats still lack essential public health information infrastructure [2,3]. The existing network of traditional surveillance efforts managed by health ministries, public health institutes, multinational agencies, and laboratory and institutional networks has wide gaps in geographic coverage and often suffers from poor and sometimes suppressed information flow across national borders [4]. At the same time, an enormous amount of valuable information about infectious diseases is found in Web-accessible information sources such as discussion forums, mailing lists, government Web sites, and news outlets [5,6,7]. These resources can support situational awareness by providing current, highly local information about outbreaks, even from areas relatively invisible to traditional global public health efforts.

Summary Points

- Valuable information about infectious diseases is found in Web-accessible information sources such as discussion forums, mailing lists, government Web sites, and news outlets.
- Web-based electronic information sources can play an important role in early event detection and support situational awareness by providing current, highly local information about outbreaks, even from areas relatively invisible to traditional global public health efforts.
- While these sources are potentially useful, information overload and difficulties in distinguishing “signal from noise” pose substantial barriers to fully utilizing this information.
- HealthMap is a freely accessible, automated real-time system that monitors, organizes, integrates, filters, visualizes, and disseminates online information about emerging diseases.
- The goal of HealthMap is to deliver real-time intelligence on a broad range of diseases, including those for which surveillance infrastructure remains weak.

department visits with acute respiratory illness [9,10]. This was followed by media reports of a respiratory disease among health care workers in February 2003, all captured by the Public Health Agency of Canada’s Global Public Health Intelligence Network (GPHIN) [10,11,12]. In parallel, online discussions on the ProMED-mail system referred to an outbreak in Guangzhou, well before official government reports were issued [13].

These Web-based data sources not only facilitate early outbreak detection, but also support increasing public awareness of disease outbreaks prior to their formal recognition. Through low-cost and real-time Internet data-mining, combined with openly available and

Funding: This work was supported by grants R21LM009263-01 and R01LM007970-01 from the National Library of Medicine, the National Institutes of Health, the Canadian Institutes of Health Research, and a research grant from Google.org. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
http://healthmap.org
The future hopefully.....
Enhancing Infectious Disease and Severe Adverse Events Following Immunizations Surveillance with Novel Data Streams

Examples include:
- Web crawl data (https://commoncrawl.org/)
- Data from social networks like Twitter (Streaming of tweets from Public API) and Facebook (public pages)
- Data from internet-based biosurveillance systems (e.g., HealthMap, US; ProMed, US; MedISys, European Commission)
- Data from health forums
- Search engines query keywords (from Google AdWords)
- Participatory electronic surveillance systems (Influenza, etc.)
- Real-time health data from GPs, pediatricians, Emergency room access, hospitalizations, drug and vaccine sales
- Proxies for collective attention to specific topics, such as Wikipedia page view data (http://blog.wikimedia.org/2015/11/13/pageview-data-easily-accessible/)

Screening, mapping and review of Novel Digital Sources (NDS) of potential use in the surveillance and monitoring of infectious diseases and public health threats

Digital Data Streams

Development and use of a dedicated IT Platform for data integration and the continuous monitoring of data flows and the early detection of signals, using technologies of machine learning. Technologies used will ensure data protection issues are respected.

Analysis of complementarity between NDS and data traditionally used for surveillance

Traditional Data used for surveillance/monitoring of Infectious diseases/Public Health Threats

Epidemi Intelligence integrating:
- Indicator based surveillance
- Event based surveillance

Model of collaboration and networking between data scientists and public health professionals that support the use of NDS in public health policies

Case studies in the field comparing information from NDS analysis with ground truth as documented by clinical and public health reference data
Grazie per l’attenzione

caterina.rizzo@iss.it