

Geo-based Social Media Analytics and SMART Dashboard for Tracking Influenza Outbreaks

Ming-Hsiang Tsou^{1,3,4}, Michael Peddecord*^{1,2,3}, Jeffrey Johnson^{5,6} and Chin-Te Jung^{1,3,4}

¹Graduate School Public Health, San Diego State University, San Diego, CA, USA; ²Quality Health Analytics, San Diego, CA, USA; ³Center for Human Dynamics in the Mobile Age, San Diego, CA, USA; ⁴Geography Department, San Diego, CA, USA; ⁵Epidemiology & Immunization Branch, San Diego, CA, USA; ⁶San Diego County Health and Human Services Agency, San Diego, CA, USA

Objective

We developed geo-targeted social media application program interfaces (APIs) for Twitter and a web-based social media analytics and research testbed (SMART) dashboard to analyze “flu” related tweets. During the 2013-14 flu season, for 10 cities with active surveillance for influenza (ILI), we correlated weekly tweeting rates and visual patterns of flu tweeting rates. To facilitate widespread use and testing of this system, we developed an interactive web-based dashboard “SMART” that allows practitioners to monitor and visualize daily changes of flu trends and related flu news.

Introduction

Active surveillance for influenza is a useful but costly endeavor. In recent years *infoveillance* tools have been developed to track and analyze data available on the Internet and social media (Eysenbach 2011). While *infoveillance* tools have been developed, few tools focus on geo-targeted data collection at a local level combined with Geographic Information Systems (GIS) capability.

Methods

Tweets are collected continuously with the keyword “flu” using customized geo-targeted Twitter Search APIs from 31 U.S. cities based on a 17 mile radius from Week 39, 2013 to Week 10, 2014. Our model is based on improvement of previous methods (Nagel et al., 2013). We compared the temporal patterns of weekly flu tweeting rates in each city against the CDC regional ILI data using 2010 census data and GIS methods (Tsou et al. 2013). The correlation (r) between flu tweeting rates and the regional CDC ILI records were calculated and trends were represented graphically. Our online SMART dashboard was built by using Python®, JavaScript®, and the Node.js® platform (Figure 1).

Results

We used regional CDC ILI data to compare each city’s flu tweeting rate within each region. Correlation (r -value) between weekly aggregated flu tweeting rates and disease occurrence (CDC regional ILI rates) were very high in many cities (Examples: Atlanta vs. Region-4 ILI $r=0.79$, Nashville vs. Region-4 ILI $r=0.93$, Chicago vs. Region-5 ILI $r=0.75$, Detroit vs. Region-5 ILI $r=0.84$, Los Angeles vs. Region-9 ILI $r=0.88$, Seattle vs. Region-10 ILI $r=0.78$). We also compare weekly confirmed influenza cases from the San Diego County public health laboratory with weekly flu tweeting rates in San Diego. The r of 0.93 indicates strong correlation of temporal patterns between confirmed influenza cases and flu tweeting rates (Figure 2).

Conclusions

Geo-targeted social media analytics with GIS methods are feasible to track and monitor flu outbreaks at the local level. A key advantage of our tool is the ability to track flu outbreaks automatically every day. Capabilities of the SMART dashboard have great potential to assist local public health agencies by providing real-time information to investigate and make appropriate responses to large-scale disease outbreaks.

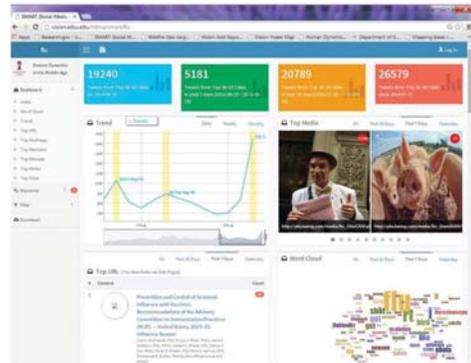


Figure 1. SMART Dashboard — Realtime Monitoring of “Flu”

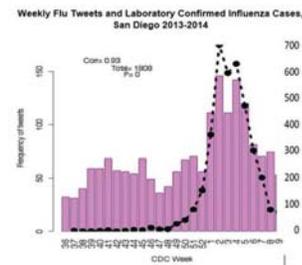


Figure 2. Comparing weekly laboratory confirmed influenza cases (black line) and weekly flu tweeting rates (pink bars). Note: Total laboratory confirmed cases= 1808

Keywords

Influenza; social media; surveillance; *infoveillance*; GIS

Acknowledgments

This study is partially supported by the National Science Foundation Grant # 1028177, “Mapping Cyberspace to Realspace”. We especially thank Anoshe Aslam for her data analysis.

References

- Tsou, MH et al (2013) Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 US Presidential Election. *Cartography and Geo Info Sci*, 40(4).
- Nagel, AC et al. (2013) The complex relationship of realspace events and messages in cyberspace: case study of influenza and pertussis using tweets. *J Med Internet Res*, 15(10).
- Eysenbach, G. (2011). Infodemiology and *infoveillance* tracking online health information and cyberbehavior for public health. *Am J Prev Med*, 40(5). 154-8.

*Michael Peddecord

E-mail: mpeddeco@mail.sdsu.edu

