OJPHI

ISDS 2016 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Using a Bayesian Method to Assess Google, Twitter, and Wikipedia for ILI Surveillance

Danielle Sharpe*[2, 1], Richard Hopkins[2], Robert L. Cook[2] and Catherine W. Striley[2]

[1]Emory University, Atlanta, GA, USA; [2]University of Florida, Gainesville, FL, USA

Objective

To comparatively analyze Google, Twitter, and Wikipedia by evaluating how well change points detected in each web-based source correspond to change points detected in CDC ILI data.

Introduction

Traditional influenza surveillance relies on reports of influenza-like illness (ILI) by healthcare providers, capturing individuals who seek medical care and missing those who may search, post, and tweet about their illnesses instead. Existing research has shown some promise of using data from Google, Twitter, and Wikipedia for influenza surveillance, but with conflicting findings, studies have only evaluated these web-based sources individually or dually without comparing all three of them[1-5]. A comparative analysis of all three web-based sources is needed to know which of the web-based sources performs best in order to be considered to complement traditional methods.

Methods

We collected publicly available, de-identified data from the CDC ILINet system, Google Flu Trends, HealthTweets.org, and Wikipedia for the 2012-2015 influenza seasons. Bayesian change point analysis was the method used to detect change points, or seasonal changes, in each of the web-data sources for comparison to change points in CDC ILI data. All analyses was conducted using the R package 'bcp' v4.0.0 in RStudio v0.99.484. Sensitivity and positive predictive values (PPV) were then calculated.

Results

During the 2012-2015 influenza seasons, a high sensitivity of 92% was found for Google, while the PPV for Google was 85%. A low sensitivity of 50% was found for Twitter; a low PPV of 43% was found for Twitter also. Wikipedia had the lowest sensitivity of 33% and lowest PPV of 40%.

Conclusions

Google had the best combination of sensitivity and PPV in detecting change points that corresponded with change points found in CDC data. Overall, change points in Google, Twitter, and Wikipedia data occasionally aligned well with change points captured in CDC ILI data, yet these sources did not detect all changes in CDC data, which could indicate limitations of the web-based data or signify that the Bayesian method is not adequately sensitive. These three web-based sources need to be further studied and compared using other statistical methods before being incorporated as surveillance data to complement traditional systems.

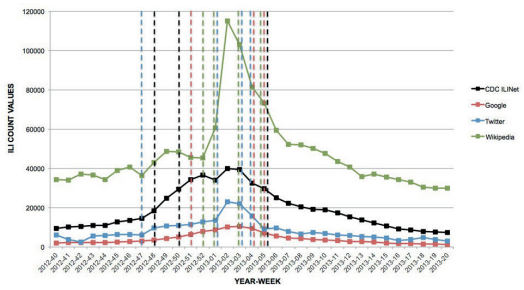

Figure 1. Detection of change points, 2012-2013 influenza season

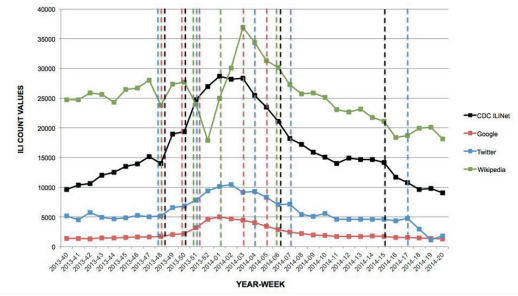

Figure 2. Detection of change points, 2013-2014 influenza season

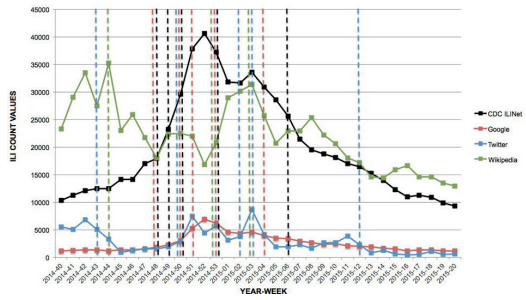

Figure 3. Detection of change points, 2014-2015 influenza season

Keywords

Google; Twitter; Wikipedia; Bayesian change point analysis; influenza surveillance

References

1. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using Twitter. Proceedings of the 2011 Conference on Empirical Natural Language Processing Conference; 2011:1568-1576.
2. Broniatowski DA, Dredze M, Paul MJ, Dugas A. Using social media to perform local influenza surveillance in an inner-city hospital: a retrospective observational study. Eysenbach G, ed. JMIR Public Health and Surveillance. 2015;1(1):e5.
3. McIver D, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol. 2014;10(4):e1003581.
4. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, Brownstein JS. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res. 2014;16(10):e236.
5. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Currents. 2014;6: ecurrents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117.

***Danielle Sharpe**
E-mail: danielle.sharpe@emory.edu