Temasek Defence Systems Institute

The Use of Twitter to Predict the Level of Influenza Activity in the United States

Ng Kok Wah
Assoc. Prof. Samuel E. Buttrey

Objectives

- Provide first responders of an influenza outbreak with situation awareness.
- Develop prediction models that uses Twitter messages to predict influenza-related statistics that indicates the level of influenza activity.

Approach

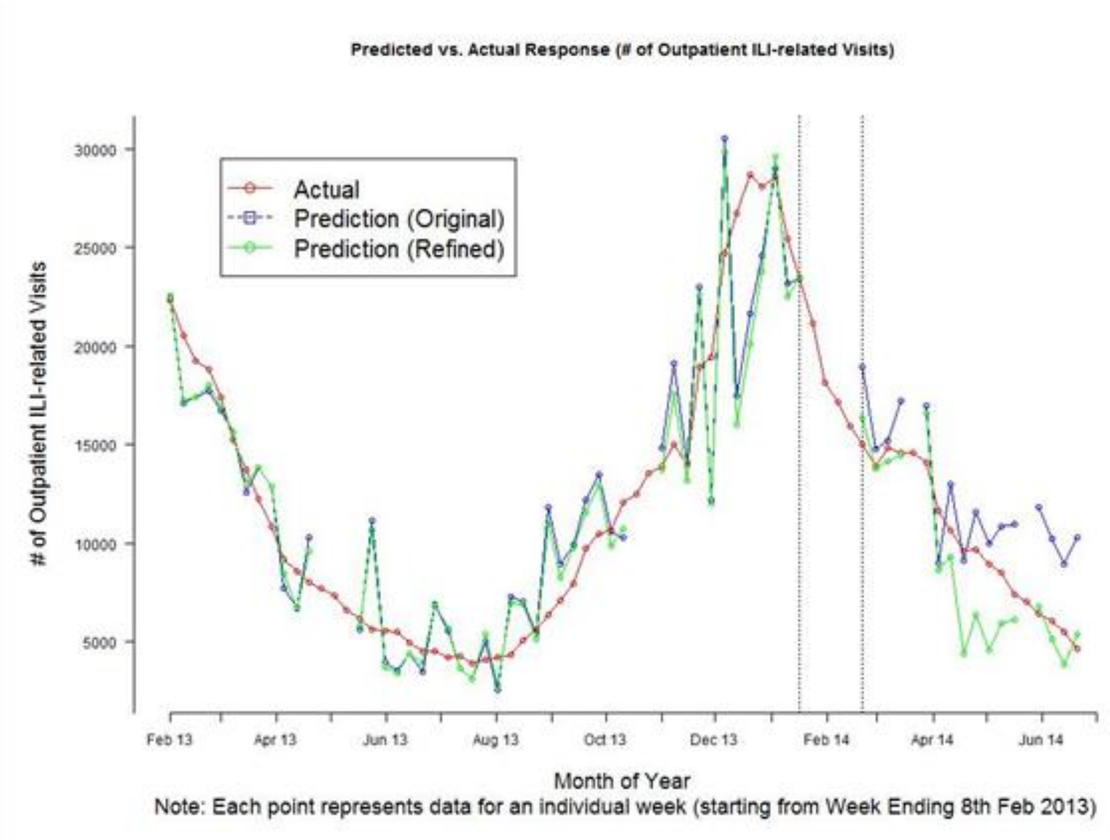
- Explores the method of aggregating frequencies of categories of hand chosen terms as predictor variables.
- Use CDC's ILI and Virologic surveillance network data that tracks the number of Influenza-like Illnesses Outpatient visits and number of respiratory specimens collected and tested positive for influenza type A and B as response variables.
- Generate predictions by using the regression models constructed for each response variables
- Determine the Pearson's correlation coefficient that describes the correlation between the generated predictions and actual CDC surveillance data.

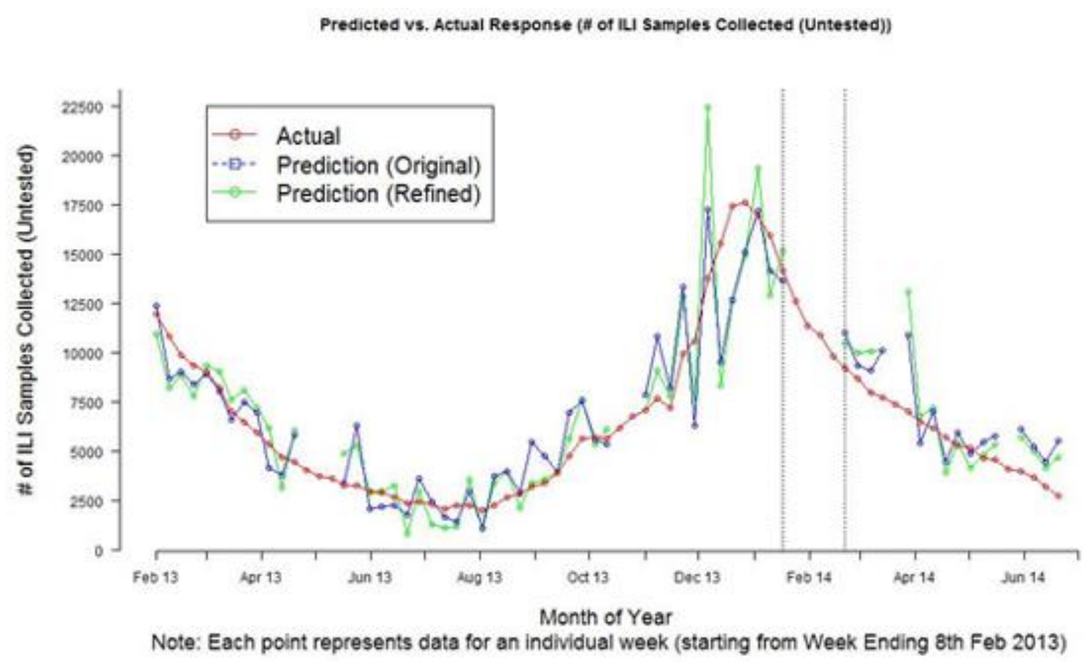
Results

- Results are promising for the models constructed for the national level (entire U.S.); the models are well fit (adjusted R2>0.6) and their predictions are highly correlated with CDC's surveillance ILI and Virologic data.
- Pearson's Correlation Coefficient between the test set predictions and actual CDC ILI surveillance data: 0.900 (95% CI: 0.732, 0.965)
- Pearson's Correlation Coefficient between the test set predictions and actual CDC Virologic surveillance data (Number of respiratory specimens collected): 0.833 (95% CI: 0.574, 0.940).
- The observed high Pearson's correlation coefficient suggests the presence of correlation between Twitter messages and CDC surveillance data.
- Low adjusted R2 (<0.6) are observed for the majority of the regional and state level models.

Benefits

- First responders are able to respond promptly and accurately to influenza outbreaks.
- Additional lead time to enhance logistics operations and preparations





Future Works

- Evaluate the proposed approach in the future using new data
- Apply the proposed approach to predict the level of influenza activity in other countries
- Refine keyword selection method
- Use of a Twitter Geo-location prediction tool to determine location of tweet sender

5 Indicative Predictor Variables						5 Supportive Predictor Variables				
Group	Flu Activities	Flu Terms	Flu Symptoms	Medicines	Flu Complications	Rest Activities	Verbs	Adjectives	Pronouns	Emoticons
Examples	Doctor	Flu	Chesty	Medicine	Pneumonia	Medical Certificate	Diagnose	Bedridden	I	:'-(
	Clinic	Influenza	Fever	Tylenol	Bronchitis	Need Some Rest	Got	Unwell	You	:'(
	Hospital	H1N1	Sore Throat	Vicks	Sinus Infection	Day Off	Down	Weak	He	>:[

NUS National University of Singapore