A Bayesian Hierarchical Model for Estimating Influenza Epidemic Severity

Nicholas L. Michaud* and Jarad Niemi

Iowa State University Statistics Department, Ames, IA, USA

Objective

To use multiple data sources of influenza epidemic severity to inform a model which can estimate and forecast severity for the current influenza epidemic season by accounting for the bias from each source.

Introduction

Timely monitoring and prediction of the trajectory of seasonal influenza epidemics allows hospitals and medical centers to prepare for, and provide better service to, patients with influenza. The CDC's ILINet system collects data on influenza-like illnesses from over 3,300 health care providers, and uses this data to produce accurate indicators of current influenza epidemic severity. However, ILINet indicators are typically reported at a lag of 1-2 weeks. Another source of severity data, Google Flu Trends, is calculated by aggregating Google searches for certain influenza related terms. Google Flu Trends data is provided in near-real time, but is a less direct measurement of severity than ILINet indicators, and is likely to suffer from bias. We create a hierarchical model to estimate epidemic severity for the 2014 - 2015 epidemic season which incorporates current and historical data from both ILINet and Google Flu Trends, allowing our model to benefit both from the recency of Google Flu Trends data and the accuracy of ILINet data.

Methods

To forecast for the 2014 - 2015 influenza epidemic season, we provide our model with both ILINet and GFT data from previous seasons, starting with the 2004 - 2005 epidemic season, and going through the 2013 - 2104 epidemic season. Our model has a hierarchical structure, which allows ILINet and GFT data from previous seasons to inform epidemic severity prediction in the current season.

ILINet data is modeled as being an unbiased but noisy estimate of the true, unknown influenza severity. GFT severity measurements, on the other hand, are influenced by external factors such as media coverage. These factors could consistently bias GFT severity estimates to over or under-estimate the true epidemic severity depending on the intensity of media influenza coverage in a season. To account for this potential bias in GFT data, we include an autoregressive error term, which allows over or under-predictions made by GFT data in one week to carry over into the next. Estimation is performed using the Bayesian statistical software Stan (http://mc-stan.org/).

Results

We examine the increase in forecast accuracy that GFT data provides by comparing the forecasting ability of our model using both GFT and ILINet data to that of a model given only ILINet data. The two models are evaluated for their ability to predict the week of maximum epidemic severity using only data from early points in the influenza season. Examining 95 % credible intervals for the peak severity week, we find that the model which uses only ILINet data consistently predicts that the peak severity week will occur later than it was observed to occur. The model using ILINet data and GFT data, on the other hand, is able to accurately estimate a range for the peak severity week several weeks before it was observed to occur.

We also evaluated each model's ability to forecast epidemic severity one week into the future. Figure 1 shows a comparison of the one week ahead forecasting abilities of these models.

Conclusions

Combining up-to-date Google Flu Trends data with accurate ILINet data improves epidemic severity forecasting ability significantly. Additional data sources, such as data from Twitter or Wikipedia, could likely further benefit forecasting.

One week ahead ILINet predictions
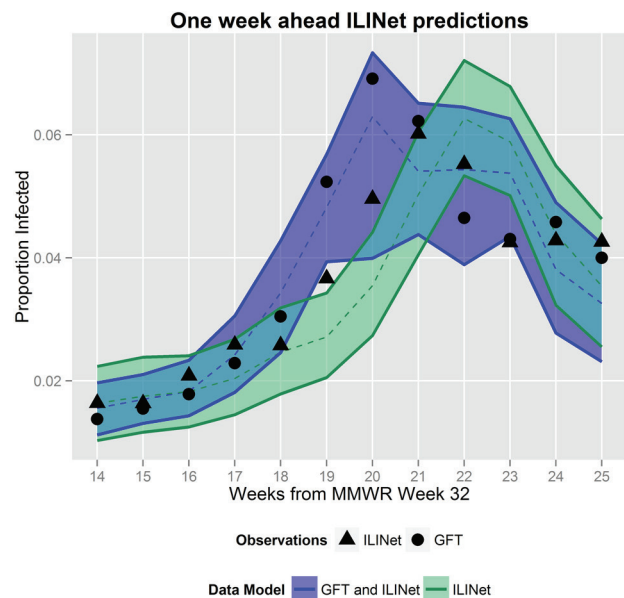
The green band shows a 95% credible interval for a forecast of ILINet data in week t, using the ILINet model given data up to week t-1. The blue band shows a 95% credible interval for a forecast of ILINet data in week t, using the Combined model given data up to week t-1.

Keywords

Influenza; Google; Model; Bayesian; Hierarchical

References

1. Chretien JP, George D, Shaman J, Chitale RA, McKenzie, FE. Influenza forecasting in human populations: A scoping review. PLoS ONE. 2014 Apr;9(4):e94130
2. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009 Feb 19;457(7232):1012-4.

*Nicholas L. Michaud

E-mail: michaud@iastate.edu