A Collaborative, Weighted Density Ensemble Approach to Influenza Forecasting in the U.S.

Logan Brooks, Spencer Fox, Craig McGowan, Sasikiran Kandula,
Dave Osthus, Evan Ray, Nicholas G Reich, Roni Rosenfeld, Jeffrey Shaman,
Abhinav Tushar, Teresa Yamana [authorship list to be finalized]

April 23, 2018

Contents

1 Introduction

Outline

- Ensembles are good - weather and inf. disease.
- Simple ensembles ok, evidence weighted ensembles better
- CDC forecasting flu for a while - history of challenge means number of teams with models to potentially combine
- Formation of FluSight Network
- Overview of goals, methods, etc.

Ensemble models, or models that fuse together predictions from multiple different models, have long been seen as a valuable method for improving predictions over any single model. This "wisdom of the crowd" approach (where the "crowd" can be thought of as a throng of models) has both theoretical and practical advantages. First, it allows for an ensemble forecast to incorporate signals from different data sources and models that may highlight different features of a system. Second, combining signals from models with different biases may allow those biases to offset and result in an ensemble that is more accurate than the individual component models. Weather and climate models have utilized ensemble systems for these very purposes, and recent work has extended ensemble forecasting to forecasts of infectious diseases, including influenza , dengue fever, and Ebola hemorrhagic fever

Since the 2013/2014 influenza season, the Centers for Disease Control and Prevention (CDC) has run an annual prospective influenza forecasting competition, known as the FluSight challenge, in collaboration with outside researchers. Participating teams submit probabalistic forecasts for a variety of influenza targets weekly from early November through mid May. During the 2015/2016 and 2016/2017 challenges, analysts at the CDC built a simple ensemble model by taking the arithmetic mean of submitted models. This model was one of the top performing models each season (cite McGowan et al when accepted).

Given the success of a simple ensemble that incorporated no information about the relative performace of component models, an ensemble taking component model performance into account has the potential for further improvments. In March 2017 the FluSight Network, a collaborative group of influenza forecasters who have worked with the CDC in the past, was established to facilitate the pooling of resources to develop an ensemble that could incorporate past performance of models. This group worked throughout 2017 to create a set of guidelines and an experimental design that would enable submission of a publicly available, multi-team, real-time submission of an ensemble model with validated and performance-based weights for each model.

This paper describes the development of

2 Methods

Outline

- ILI definition
- Forecast structure
- Forecast evaluation
- Component models
- Ensemble construction

2.1 Influenza Data

Forecasting targets for the CDC FluSight challenge are based on the US Outpatient Influenza-like Illness Surveillance Network (ILINet). ILINet is a syndromic surveillance system that measures the weekly percentage of outpatient visits due to influenza-like illness (ILI) from a network of more the 2,800 providers, and publishes a weighted estimate of ILI (wILI) based on state populations. Estimates of wILI are reported weekly by the CDC's Influenza Division for the United States as a whole as well as for each of the 10 Health and Human Services (HHS) regions. Reporting of 'current' wILI is typically delayed by approximately two weeks as data are collected and processed, and each weekly publication can also include revisions of prior reported values if new data become available.

2.2 Forecast Targets and Structure

3 Results

Outline

- Description of models included
- Ensemble comparisons using CV data by log score
- Ensemble comparisons using CV data by model rank
- Prospective comparisons of 2017-2018 by log score and rank

4 Discussion

Outline

- Mostly TBD depending on results
- Comparisons to unweighted average
- Comparisons to component models
- Strengths and weaknesses of ensemble approach
- Practicality and real-world impacts

References