

Adaptively stacked ensembles for influenza forecasting with incomplete data

Thomas C. McAndrew, Nicholas G. Reich

University of Massachusetts at Amherst, Dept. of Biostatistics and Epidemiology

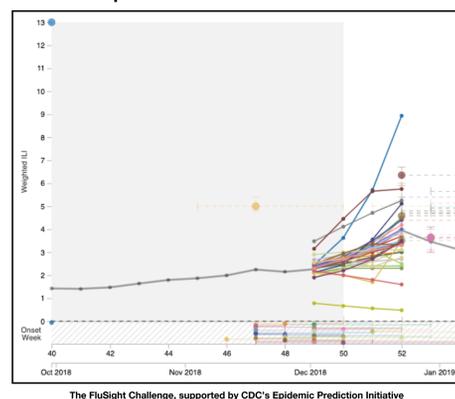
Introduction

Seasonal Influenza infects an average 30 million people in the United States every year¹, overburdening hospitals during weeks of peak incidence. Named by the CDC as an important tool to fight the damaging effects of these epidemics, accurate forecasts of influenza and influenza like illness (ILI) forewarn public health officials about when, and where, seasonal influenza outbreaks will hit hardest. Extending an existing ensemble implementation², we developed a new method for combining component models that relies on recently observed, in-season data to adaptively estimate a convex combination of models.

To protect the adaptive ensemble framework from relying too heavily on recent, revision-prone ILI data, we developed a Bayesian model that uses a time-dependent prior to regularize ensemble weights.

Data

Every week throughout the season, for 10 different reporting regions and a national average, the CDC publishes surveillance data on influenza-like illness (ILI). ILI is defined as the percentage of patients presenting with a fever~(greater than 100F) plus cough or sore throat with no known cause other than influenza.

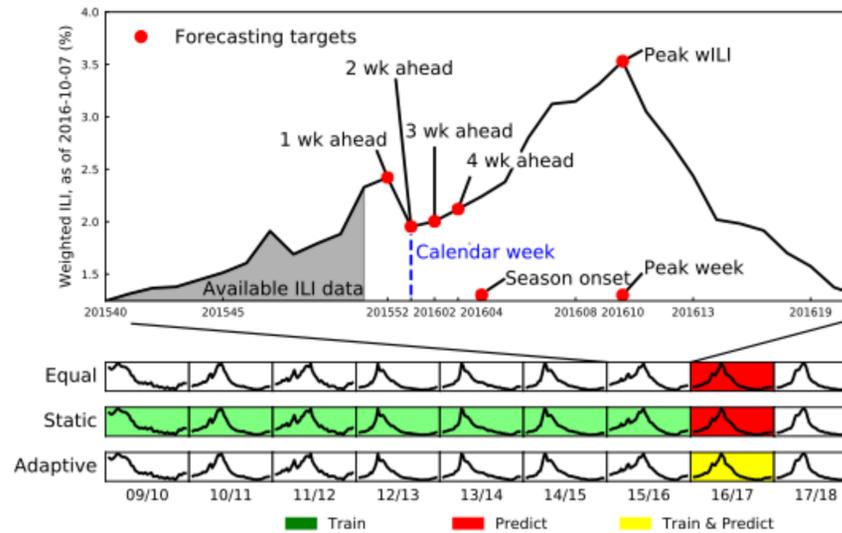


21 Component model forecasts

Real-time forecasts from 2010/2011 to present

The FluSight Network (FSN) is a collaborative group of influenza forecasters, using historical performance of models to build ensemble forecasts.

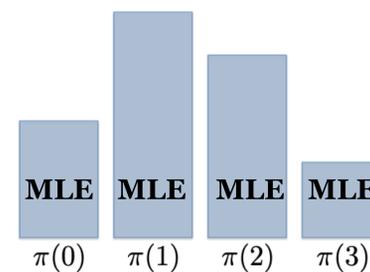
Methods



Assume ILI data is generated by a mixture of component models

$$p(\text{ILI}_t | \pi) = \sum_{m=1}^{\# \text{models}} \pi(m) f(m, \text{ILI}_t)$$

find MLE



Prior is time-dependent

$$p[\pi(t)] = \text{Dir}[\pi | \alpha(t)]$$

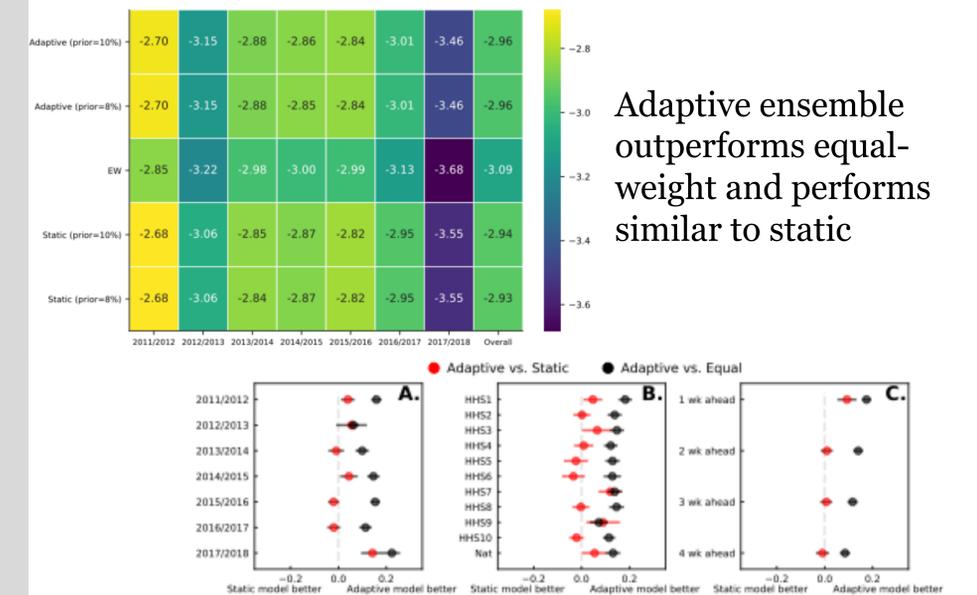
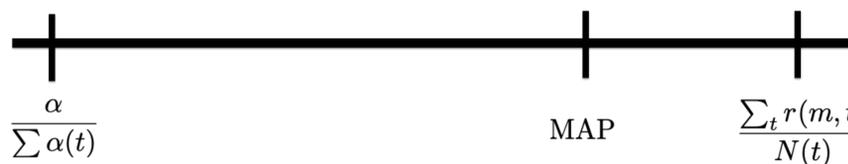
$$\alpha(t) = \frac{\rho N(t)}{M}$$

percent training examples / models

Maximum a posteriori (MAP) is a convex combination btw prior and MLE

$$\text{MAP} [\pi(m)] = \left[\frac{\alpha_m(t)}{\sum_m \alpha_m(t)} \right] \left(\frac{\rho}{1 + \rho} \right) + \left(\frac{1}{1 + \rho} \right) \left[\frac{\sum_{t=1}^T r(m, t)}{N(t)} \right]$$

$r(m, t)$ = responsibility ρ = Percent towards equal weighting



Adaptive ensemble outperforms equal-weight and performs similar to static

Discussion

Our adaptive ensemble can forecast from revision-prone, noisy ILI data by relying on a prior. We show this adaptive algorithm outperforms an equal-weight ensemble and shows similar, or better performance against a static ensemble.

Able to generate quick forecasts from sparse or noisy data, an adaptive ensemble could serve as a valuable tool to public health officials needing to make informed decisions under uncertainty

References

- Thompson, M. G., et al. "Estimates of deaths associated with seasonal influenza-United States, 1976-2007." *Morbidity and Mortality Weekly Report* 59.33 (2010): 1057-1062.
- Reich, Nicholas G., et al. "A Collaborative Multi-Model Ensemble for Real-Time Influenza Season Forecasting in the US." *bioRxiv* (2019): 566604.

This work was funded by the National Institute of General Medical Sciences (NIGMS) Grant R35GM119582, and the Defense Advanced Research Projects Agency. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the NIH or the NIGMS. The funders had no role in study design, data collection and analysis, decision to present, or preparation of the presentation.



Adaptively stacked ensembles for influenza forecasting with incomplete data

Thomas C. McAndrew, Nicholas G. Reich

University of Massachusetts at Amherst, Dept. of Biostatistics and Epidemiology

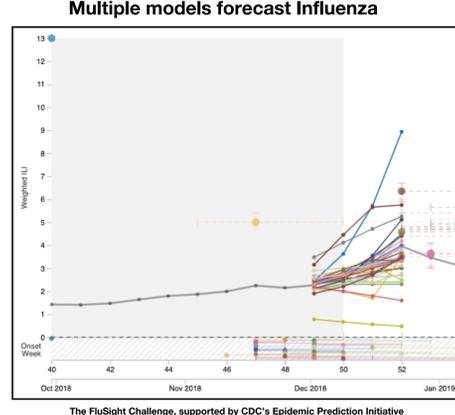
Introduction

Seasonal Influenza infects an average 30 million people in the United States every year¹, overburdening hospitals during weeks of peak incidence. Named by the CDC as an important tool to fight the damaging effects of these epidemics, accurate forecasts of influenza and influenza like illness (ILI) forewarn public health officials about when, and where, seasonal influenza outbreaks will hit hardest. Extending an existing ensemble implementation², we developed a new method for combining component models that relies on recently observed, in-season data to adaptively estimate a convex combination of models.

To protect the adaptive ensemble framework from relying too heavily on recent, revision-prone ILI data, we developed a Bayesian model that uses a time-dependent prior to regularize ensemble weights.

Data

Every week throughout the season, for 10 different reporting regions and a national average, the CDC publishes surveillance data on influenza-like illness (ILI). ILI is defined as the percentage of patients presenting with a fever~(greater than 100F) plus cough or sore throat with no known cause other than influenza.

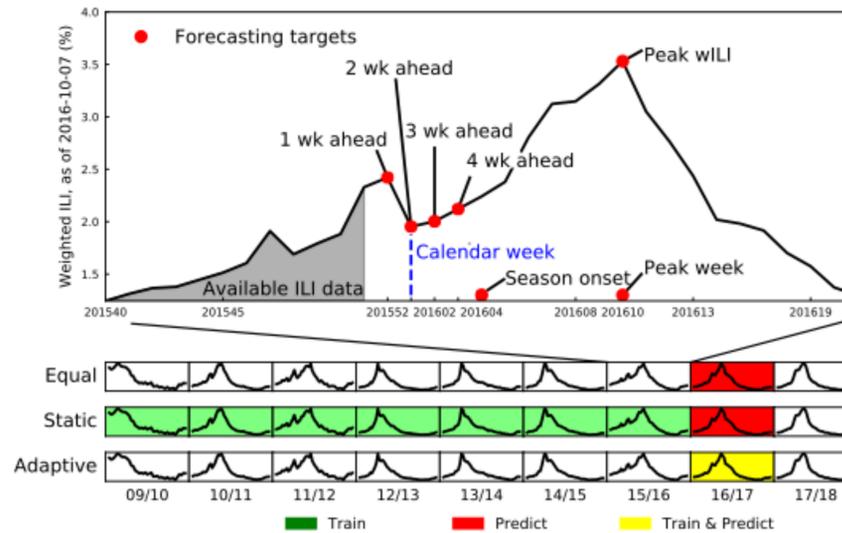


21 Component model forecasts

Real-time forecasts from 2010/2011 to present

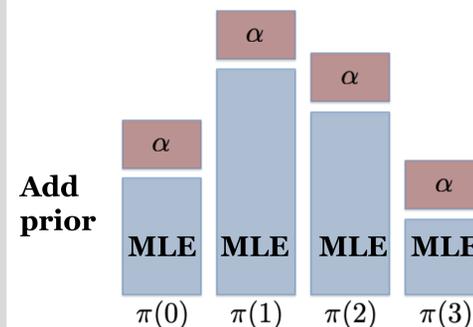
The FluSight Network (FSN) is a collaborative group of influenza forecasters, using historical performance of models to build ensemble forecasts.

Methods



Assume ILI data is generated by a mixture of component models

$$p(\text{ILI}_t | \pi) = \sum_{m=1}^{\#\text{models}} \pi(m) f(m, \text{ILI}_t)$$



Prior is time-dependent

$$p[\pi(t)] = \text{Dir}[\pi | \alpha(t)]$$

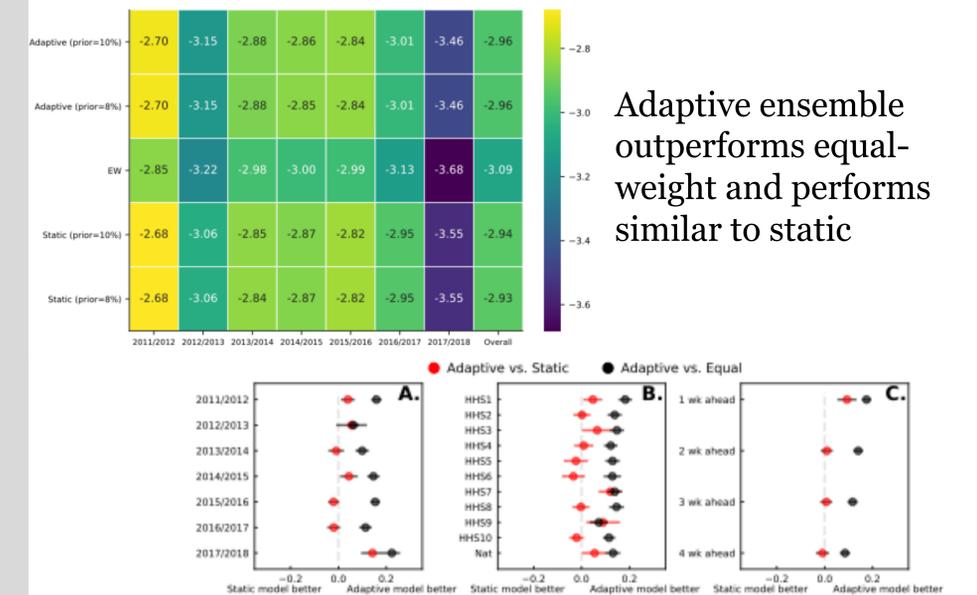
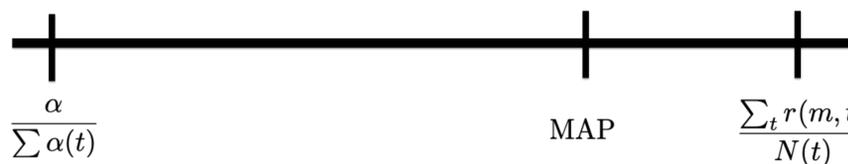
$$\alpha(t) = \frac{\rho N(t)}{M}$$

percent training examples
models

Maximum a posteriori (MAP) is a convex combination btw prior and MLE

$$\text{MAP} [\pi(m)] = \left[\frac{\alpha_m(t)}{\sum_m \alpha_m(t)} \right] \left(\frac{\rho}{1 + \rho} \right) + \left(\frac{1}{1 + \rho} \right) \left[\frac{\sum_{t=1}^T r(m, t)}{N(t)} \right]$$

$r(m, t)$ = responsibility ρ = Percent towards equal weighting



Adaptive ensemble outperforms equal-weight and performs similar to static

Discussion

Our adaptive ensemble can forecast from revision-prone, noisy ILI data by relying on a prior. We show this adaptive algorithm outperforms an equal-weight ensemble and shows similar, or better performance against a static ensemble.

Able to generate quick forecasts from sparse or noisy data, an adaptive ensemble could serve as a valuable tool to public health officials needing to make informed decisions under uncertainty

References

- Thompson, M. G., et al. "Estimates of deaths associated with seasonal influenza-United States, 1976-2007." *Morbidity and Mortality Weekly Report* 59.33 (2010): 1057-1062.
- Reich, Nicholas G., et al. "A Collaborative Multi-Model Ensemble for Real-Time Influenza Season Forecasting in the US." *bioRxiv* (2019): 566604.

This work was funded by the National Institute of General Medical Sciences (NIGMS) Grant R35GM119582, and the Defense Advanced Research Projects Agency. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the NIH or the NIGMS. The funders had no role in study design, data collection and analysis, decision to present, or preparation of the presentation.



Adaptively stacked ensembles for influenza forecasting with incomplete data

Thomas C. McAndrew, Nicholas G. Reich

University of Massachusetts at Amherst, Dept. of Biostatistics and Epidemiology

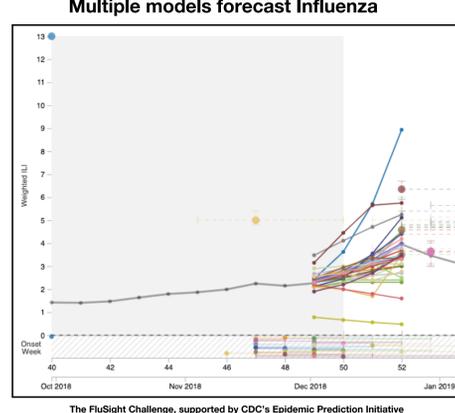
Introduction

Seasonal Influenza infects an average 30 million people in the United States every year¹, overburdening hospitals during weeks of peak incidence. Named by the CDC as an important tool to fight the damaging effects of these epidemics, accurate forecasts of influenza and influenza like illness (ILI) forewarn public health officials about when, and where, seasonal influenza outbreaks will hit hardest. Extending an existing ensemble implementation², we developed a new method for combining component models that relies on recently observed, in-season data to adaptively estimate a convex combination of models.

To protect the adaptive ensemble framework from relying too heavily on recent, revision-prone ILI data, we developed a Bayesian model that uses a time-dependent prior to regularize ensemble weights.

Data

Every week throughout the season, for 10 different reporting regions and a national average, the CDC publishes surveillance data on influenza-like illness (ILI). ILI is defined as the percentage of patients presenting with a fever~(greater than 100F) plus cough or sore throat with no known cause other than influenza.

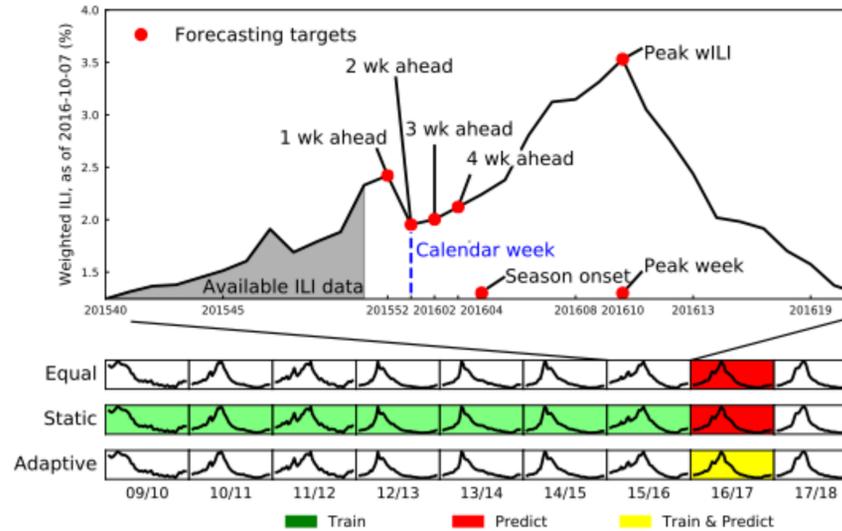


21 Component model forecasts

Real-time forecasts from 2010/2011 to present

The FluSight Network (FSN) is a collaborative group of influenza forecasters, using historical performance of models to build ensemble forecasts.

Methods



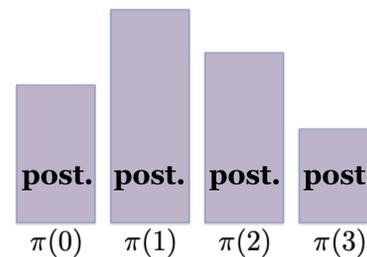
Assume ILI data is generated by a mixture of component models

$$p(\text{ILI}_t | \pi) = \sum_{m=1}^{\# \text{models}} \pi(m) f(m, \text{ILI}_t)$$

Prior is time-dependent

$$p[\pi(t)] = \text{Dir}[\pi | \alpha(t)]$$

norm



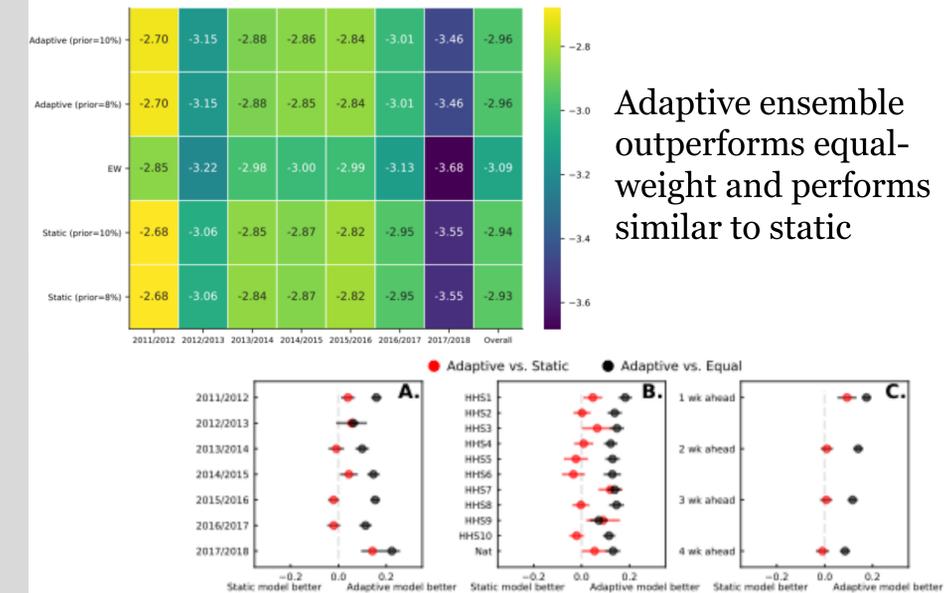
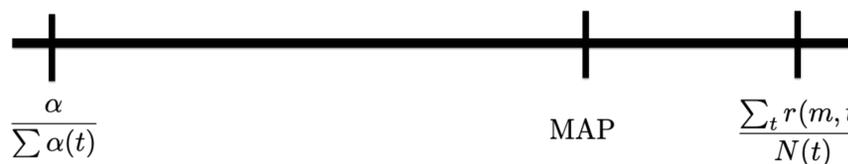
$$\alpha(t) = \frac{\rho N(t)}{M}$$

percent training examples
models

Maximum a posteriori (MAP) is a convex combination btw prior and MLE

$$\text{MAP} [\pi(m)] = \left[\frac{\alpha_m(t)}{\sum_m \alpha_m(t)} \right] \left(\frac{\rho}{1 + \rho} \right) + \left(\frac{1}{1 + \rho} \right) \left[\frac{\sum_{t=1}^T r(m, t)}{N(t)} \right]$$

$r(m, t)$ = responsibility ρ = Percent towards equal weighting



Adaptive ensemble outperforms equal-weight and performs similar to static

Discussion

Our adaptive ensemble can forecast from revision-prone, noisy ILI data by relying on a prior. We show this adaptive algorithm outperforms an equal-weight ensemble and shows similar, or better performance against a static ensemble.

Able to generate quick forecasts from sparse or noisy data, an adaptive ensemble could serve as a valuable tool to public health officials needing to make informed decisions under uncertainty

References

- Thompson, M. G., et al. "Estimates of deaths associated with seasonal influenza-United States, 1976-2007." *Morbidity and Mortality Weekly Report* 59.33 (2010): 1057-1062.
- Reich, Nicholas G., et al. "A Collaborative Multi-Model Ensemble for Real-Time Influenza Season Forecasting in the US." *bioRxiv* (2019): 566604.

This work was funded by the National Institute of General Medical Sciences (NIGMS) Grant R35GM119582, and the Defense Advanced Research Projects Agency. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the NIH or the NIGMS. The funders had no role in study design, data collection and analysis, decision to present, or preparation of the presentation.

