DRAFT

ASSESSMENT OF A WISDOM OF CROWDS APPROACH TO FORMING ENSEMBLE MODELS FOR FORECASTING OF INFLUENZA LIKE ILLNESS

Jeff Morgan

Statement of the Problem and Background: Accurately determining the epidemic trajectory of infectious disease outbreaks is difficult, preventing decision-makers and public health officials from developing optimal mitigation strategies. Multiple models predict the characteristics of the outbreak, but they yield different results. Decision-makers may incorporate their own judgment and experience with the range of available models, but the forecasts may be more accurate if it appropriately incorporated the judgment and experience of many people.

Seasonal influenza is an important public health concern and offers an annual opportunity to develop and evaluate infectious disease models. The global health burden includes millions becoming ill and thousands dying on an annual basis. Starting with the 2013-2014 flu season, the Influenza Division of the Centers for Disease Control and Prevention (CDC) has held a "Predict the Influenza Season Challenge" to encourage the scientific community to make advances in the field of influenza forecasting. An observation from earlier challenges was that a simple average of the submitted forecasts seemed to outperform nearly all the individual models. During the 2017-2018 season, I explored a crowdsourcing, "Wisdom of Crowds" approach to capture the input from a handful of individuals (average of 3.5 participants per week) and found that this small crowd outperformed the simple average.

Purpose. The purpose of this study is to collect data which may suggest whether the accuracy of influenza forecasting can be improved by creating ensemble models with a weighting strategy based on votes from willing participants. While this study is focused on influenza, it has potential applications to synthesizing multiple models for any other infectious disease such as Ebola, Zika, Chikungunya, and measles.

Methodology. On a weekly basis throughout the challenge, multiple modeling teams will provide discrete probability distributions for seven targets at the National and regional level and for six targets at the state level. The targets are based on CDC's use of weighted Influenza Like Illness (wILI) percentage (wILI %) as a metric for influenza activity. The CDC derives wILI% as the percentage of people attending participating health facilities with symptoms that suggest an influenza like illness, weighted based on state population. The seven targets are season onset week, season peak week, season influenza peak wILI %, one week ahead wILI%, two week ahead wILI%, three weak ahead wILI%, and four week ahead wILI%. State forecasts do not include season onset as a target.

CDC releases wILI% data on the FluSight website on Fridays, enabling modeling teams to submit forecasts on Mondays. CDC upload these probability distributions to GitHub on Tuesday morning or early afternoon. I will create visual displays of these models and place them on a website, and also in a pdf document, later that day to collect votes from willing participants by Wednesday evening. Participants will vote, either online or via email, on the probability distributions they perceive to be most accurate at the national level and for California, Minnesota, Region 2, Region 3, and Region 6. These votes will be used to assign weights for the remaining areas. I will also provide influenza data for the current season, the epidemiological curves for influenza since the 2004-2005 season. I will use the percentage of votes for each model as the weights for combining the models.

Predictions are scored against values that the CDC reports at the end of the influenza season. For targets based on weeks (i.e., season onset and season peak week), the probability distribution bins are weeks. For these targets, scoring is based on the probability assigned to the bin containing the correct week as well as one bin (representing one week) before and after the correct value. For targets based on wILI%, points are awarded for the probability assigned to the bin containing the correct wILI% as well as five bins (representing 0.5%) below and five bins above. All bins within the scoring range yields equal scores. This scoring methodology will be used to compare the results of the crowd forecast with the results of individual models and their average. If the crowd statistically (p=0.05) outperforms the average model, it will suggest that this approach to combining individual models is preferable to taking a simple average.

When people register for the challenge, they will be asked to answer questions regarding their job responsibilities, education, experience with infectious disease models, and desire to be eligible for prizes based on random drawings and performance. They will also be asked if they would be interested in being a co-author in a future article describing this study. These answers will be used to address the extent that the crowd satisfies the conditions of diversity, independence, and decentralization as explained by James Surowicecki in his popular book, The Wisdom of Crowds. Depending upon participation levels, comparisons may be made between the performance of individual groups. Throughout the season, crowd members will be given updates of their tentative scores as well as the scores of the models. During holiday weeks (e.g., Christmas Day, New Years Day), CDC will adjust due dates to accommodate the obligations of participants.

A “Wisdom of Crowds” approach to influenza forecasting has been demonstrated, but those approaches have asked participants to sketch epidemic trajectories or rank possible trajectories from most likely to least likely. It is important to determine methods for combining multiple models to enable better public health strategies aimed at mitigating the infectious disease risk. This “Wisdom of Crowds” approach may be equally beneficial for Ebola, measles, cholera, or other diseases.

Selected Bibliography

"Building a collaborative ensemble to forecast influenza" http://reichlab.io/2017/11/28/flusight-ensemble.html

Eldon, Y. Li, Tung, Chen-Yuan, Chang, Shu-Hsun. "The wisdom of crowds in action: Forecasting epidemic diseases with a web-based prediction market system". International Journal of Medical Informatics. Volume 92, August 2016, Pages 35-43. https://doi.org/10.1016/j.ijmedinf.2016.04.014

Spice, Byron. "Carnegie Mellon's Flu Forecasts Prove Most Accurate". 6 September 2017. https://www.cs.cmu.edu/news/carnegie-mellons-flu-forecasts-prove-most-accurate

Seasonal Influenza, More Information. https://www.cdc.gov/flu/about/season/flu-season.htm

CDC Overview of Influenza Surveillance in the United States: https://www.cdc.gov/flu/weekly/overview.htm

"FluSight Forecasts" https://github.com/cdcepi/FluSight-forecasts

"Forecasting Influenza in the U.S. Is a Group Effort" Biocompare posted December 28, 2017. https://www.biocompare.com/Life-Science-News/345776-Forecasting-Influenza-in-U-S-Is-a-Group-Effort/

Spice, Byron. "Carnegie Mellon's Flu Forecasts Prove Most Accurate". 6 September 2017. https://www.cs.cmu.edu/news/carnegie-mellons-flu-forecasts-prove-most-accurate

Suroweicki, James. The Wisdom of Crowds. Doubleday. New York: 2004.