eVISON

eVision: Influenza Prediction

Department of Computer Science: Ron Huang
Department of Bioengineering: George Kouretas
Faculty Advisor: Prof. Navid Shaghaghi

IEEE GLOBAL HUMANITARIAN TECHNOLOGY
TENTH ANNIVERSARY

BioInnovation and Design Lab

EPIC
ETHICAL, PRAGMATIC, AND INTELLIGENT COMPUTING
LABORATORY

Background

Influenza is an umbrella term for a number of related viruses that are extremely infectious and widespread which cause respiratory complications. An estimated 3 to 5 million severe cases of influenza are reported annually, which lead to between 290 to 650 thousand respiratory deaths worldwide annually. Influenza vaccination is the primary strategy to prevent influenza and its complications, evidenced by vaccines preventing between 3.4 and 7.1 million flu cases during the 2018-2019 flu season.

Purpose

By creating a model capable of providing useful predictions of an influenza season over a month in advance, medical professionals and national health services could be given vital information to allow them to prepare for the eventual outbreak.

Program Design

eVision uses modern recursive neural network techniques in the form of a Long short-term memory (LSTM) neural network to surpass the Google Flu Trends project's linear model of predictions in accuracy and number of weeks ahead. The LSTM is fed data on the relative popularity of keywords being searched on Google using publicly available data from Google Trends which serve as the model's features.

Frugal Innovation Hub
SANTA CLARA UNIVERSITY
SCHOOL OF ENGINEERING

Santa Clara University
School of Engineering

Data Collection

The data that we use as features for our model come from three sources:

- FluView (CDC): The CDC provides weekly data on influenza-like illnesses reported in the USA going back to 1997. This data is used for our predictions
- FluNet (WHO). WHO provides data on influenza cases from 171 countries, areas, and territories. Only counts the specific cases analyzed by National Influenza Centers
- Google Trends: Extracted key search terms related to influenza as features. Data provides weekly trends from a 5-year window (2015-Present)

Implementation

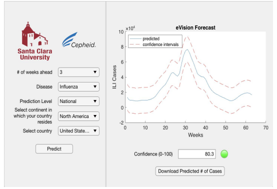
eVision is being implemented with two different programing languages:

- MATLAB: Run the LSTM neural network to yield prediction. Built the GUI of predictor using app designer.
- Python: Used to dynamically scrape web pages in order to extract data. Utilizes pytrends unofficial Google Trends API to acquire Google Trends, and selenium to scrape CDC and WHO Data

Graphical User Interface

Our GUI (see Fig. 1) provides the user simple controls that someone without any technical coding knowledge can use.

It allows the user to edit the parameters that are submitted to the model, which include the amount of weeks they would like to predict ahead and the region they would like to make a prediction for. There are also controls which allow the user to download images of the prediction graph, and a spreadsheet of the predicted number of cases


Santa Clara University
Cepheid.

# of weeks ahead: 3
Disease: Influenza
Prediction Level: National
Select continent in which your country resides: North America
Select country: United State...

Confidence (0-100): 80.3

Fig 1: The current design for our GUI

Results

We attempt to test our model with a variety of different parameters to see which yields the most accurate predictions. The primary way this has been achieved is through the granulation of our data. By this, we mean extracting data for more individualized regions that make up one region. For example, we are able to extract better results using the Google Trends data from all the sublevels (states) in the United States, rather than using the national level data from the United States.

Our best predictions have come when we granulate to metropolitan regions, which is the lowest level Google data can go. For the United States, we have managed to get our error rate as low as 8.57% (Fig. 2) using data from states (California, Texas, and New York) that feature national ports of entry.

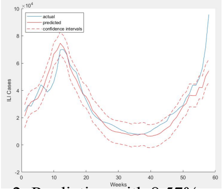


Fig 2: Prediction with 8.57% error

Future Work

There are a variety of features the eVision team have been working on implementing.

- Predictions for more than base influenza.
  - These include influenza strains A and B, respiratory syncytial virus (RSV), coronavirus (SARS-CoV-2), and more.
- Add more features.
  - We have been working on scraping Twitter and Facebook for even more data that may yield more robust predictions.
- Create our own domain and launch our application on a web server

Cepheid.

MATLAB

SANTA CLARA UNIVERSITY 1851