Advances in Economics, Business and Management Research, volume 155
Proceedings of the 2020 2nd International Conference on Economic Management and Cultural Industry (ICEMCl 2020)

Nowcasting Influenza Using Google Flu Trend and Deep Learning Model

Ping Jiang[1, *]

[1]IB Course Centre of High School Affiliated to Shanghai Jiao Tong University, Shanghai, China
tyshe0411@hotmail.com

ABSTRACT
Understanding how Influenza-like Illness (ILI) spreads has a substantial societal impact on improving overall human well-being and effective government response. The accurate estimation of ILI occurrences in a timely manner helps to alert and make meaningful intervention decisions that could save lives. In this article, I leverage Google Flu Trends (GFT) data derived from Google general search queries in the United States during 2003-2015 to now cast the occurrence of ILI collected from CDC using various state-of-the-art deep learning models with different neural network architectures, such as a convolutional neural network (CNN), recurrent neural network (RNN), and sequence-to-sequence (Seq2Seq) models. The correlations between GFT and ILI data are first analyzed through the stationary test using classical time-series forecast models such as ARIMA and the recently developed Prophet model. Then, the deep learning models are used to predict the ILI occurrences at both the national and state levels. The results show that deep learning models are capable of now casting the ILI occurrence with good accuracy. Also, the model performance varies at the state level, the signaling occurrence of ILI may be impacted by idiosyncratic factors pertaining to the state and beyond what GFT may capture. Overall, this paper adds to the existing literature on using real-time alternative data sources such as general search queries to estimating ILI occurrence, which sheds light on achieving effective surveillance to achieve a better social good by preventing another pandemic.
Keywords: *Influenza-like illness, Google Flu Trend, Deep learning Model, Now casting*

1. INTRODUCTION

Illness like influenza (ILI) refers to a severe respiratory infection caused by human-to-human virus. The goal of ILI surveillance is to determine the timing, location, and magnitude of outbreaks by monitoring the frequency and progression of clinical case incidence [1]. The value of ILI surveillance is not only to identify hospitalized cases but also to track trends of the flu promptly. In addition, it can allow relevant government agencies to measure the overall burden of the community; improve understanding of influenza epidemiology and take precautionary measures to reduce such burden [2].

However, the ILI occurrence data from the Centers for Disease Control and Prevention (CDC) is the outcome, and medical institutions, government, and the public cannot take preventative actions before the outbreak of an influenza. To proactively resolve the uncertainty, researchers adopt now casting technique to predict ILI ahead of the healthcare-based reports [3] and even in near-real-time [4]. Therefore, data sources that can track the influenza activity raise high interests for prompt government response.

In November 2008, Google launched Google Flu Trends (GFT), an internet-based surveillance tool that uses anonymized and aggregated Google search data to estimate influenza activity in near-real-time [5,6]. For example, during the 2007–08 influenza season, Google Flu Trends estimates were highly correlated to CDC surveillance for ILI [7]. In 2009, H1N1 provides a chance for GFT to predict the flu outbreak and can be used to be validated by the Influenza-like Illness Surveillance Network (ILINet) [8], e.g., GFT models performed well during pH1N1(before the outbreak) and the summer mouths (Samantha 2014). Similarly, Orti et al. also used the historical GFT data from 2003-2007 and showed its usefulness in comparison to traditional surveillance system [8].

As GFT measures the changes in internet search behavior, researchers have been using GFT to now cast flu trend during the outbreak of the influenza in the past years and yet face some serious setbacks. In 2013, models based on GFT data were criticized for predicting more than double the proportion of doctor visits for ILI activities than the CDC reported numbers [9]. Some possible reasons include that the method built on the GFT data may alter human’s thoughts on engine search behavior, the health care system, and the physician testing practices, thus making Google overestimate the peak flu lever that year. As such, Butler reported that the big data approach may not as accurate as expected to substitute traditional surveillance networks [10].

Moreover, while Santillana et al. [11] and the following studies acknowledged the inaccuracies of GFT data, they also pointed out the methodological improvement may yield better influence detection. Given that now casting ILI occurrences using GFT data remains an underexplored issue, this work proposed to adopt three deep learning models including CNN, RNN and Sequence to Sequence, to estimate the retrospective, and prospective performance of GFT to capture the season-to-season epidemic outbreak. Results based on the recently developed deep learning models not only outperform the traditional models that raise questionable estimations, but also demonstrate the potentials for continuing to use alternative data sources such as GFT to complement the existing surveillance systems.

2. MATERIAL/DATA

2.1. Google Flu Trends (GFT)

As many as 72 percent of U.S. adults admit to checking for health information online in the past years [12]. About 90 million people search primarily for information about certain diseases such as cough or flu or treatments. More than three-quarters of online health information searchers start searching on Google, Bing, or Yahoo [5]. Hence, in 2008, Google uses aggregated search query data to predict flu trends in 25 countries, including the United States. The Internet-based influenza surveillance tool is called Google Flu Trends (GFT) [13]. First, from 2003 to 2008, Google record about 50 million common queries entered weekly within the United States. Every year, Google refresh with 45 of the most useful influenza-words. A linear regression model is used to compute the log-odds of Influenza-like illness (ILI) physician visit and the log-odds of ILI-related search query:

$$\log it(P) = \beta_0 + \beta_1 \times \log it(Q) + \varepsilon \qquad (1)$$

P is the percentage of illness people visit and Q is the ILI-related query fraction computed. β_0 is the intercept and β_1 is the coefficient, and ε is the error term.
In the work, I use the data from 9/28/2003 to 8/9/2015 in the United States, including 50 states, 10 HHS Regions, 97 cities, and the sum of the 50 states. The smallest value is from HHS Region 7 (IA, KS, MO, NE) on 7/12/2015 and the data is -78. The largest value is from Oklahoma on 1/13/2013 and the data is 24586. Then I focus on the data of the sum and some of the states. I use Google Flu Trends data describing the spread of flu in the United States during 2003-2015(Jeremy 2009).
Notably, GFT is a time-series data, which means a series of data points indexed in time order. Stock, weather, and GDP are examples of time-series data. Time-series data is always plotted via line charts, so auto-regression analysis is the most common method to predict the data, which means the latter data is highly related to the former data. Therefore, I can use classical time series models such as ARIMA and Prophet to test the GFT's stationary, which including seasonality trends. For example, the sales of ice creams will increase in summer and decrease in winter. GFT shows the same properties.

2.2. ARIMA

Autoregressive Integrated Moving Average Model, which is abbreviated for ARIMA, is a classical time-series model that stands as Auto-Regressive Integrated Moving Averages. The predictors depend on the parameters (p,d,q) of the ARIMA model. In the AR model, which is an auto-regression model, it contains:
y_{t-1} depends on y_t, and y_{t-2} depends on y_{t-1}. When y_{t-1} removes, y_t is independent on y_{t-2}. P is the number of AR (Auto-Regressive) terms. It represents the lags of the time series data itself used in the prediction model. In the MA model, which is a moving average model, it contains:

$$\begin{aligned} y_t &= \varphi_1 y_{t-1} + \varepsilon_t \\ y_{t-1} &= \varphi_1 y_{t-2} + \varepsilon_{t-1} \\ y_{t-2} &= \varphi_1 y_{t-3} + \varepsilon_{t-2} \end{aligned} \qquad (2)$$

y_t depends on et and et-1, and y_{t-1} depends on et-1 and et-2. Same as y_{t-2}. Q is the number of MA (Moving Average) terms. It represents the number of lags used in the prediction model. D refers to the number of no seasonal differences.
However, these are some limitations of ARIMA model. ARIMA requires the data to be stationary or via differencing. As a result, a new time-series model was designed, The Prophet Forecasting Model. Unlike with the ARIMA models, the measurements do not need to be regularly spaced, and we do not need to interpolate missing values e.g. from removing outliers.

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \\ y_{t-1} &= \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} \\ y_{t-2} &= \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3} \end{aligned} \qquad (3)$$

The formula is from https://blog.csdn.net/weixin _38502514/article/details 87986906, accessed on 22 April 2020.
The equation is $y(t) = g(t) + s(t) + h(t) + \epsilon$. $g(t)$ represents the non-periodic changes of trend function. $s(t)$ represents periodic changes. $h(t)$ represents the holidays which are irregular events. t represents special changes which are not contain in the model.
The model forms a cycle. First, using a flexible specification made by humans to model the time series. Second, using the model to predict baselines on historical data and making some adjustments to improve the model. Or if the problem is made by human's analysts, it will be targeted in a prioritized order. Finally, based on the results made some other adjustments to the model.

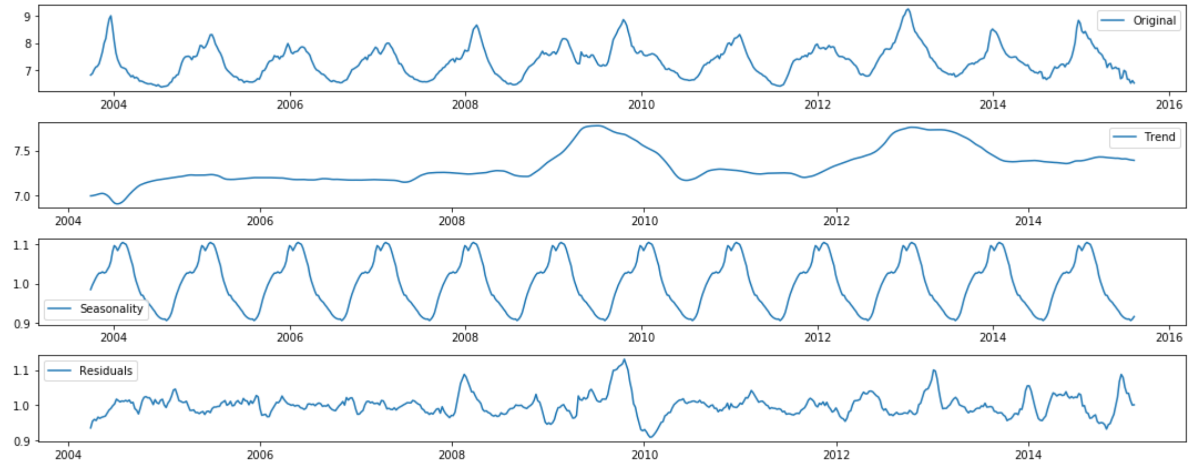

Figure 1 ARIMA model shows the stationary of GFT

2.3. PROPHET

From the decomposition of GFT data, both trend and seasonality are modeled separately. In the ARIMA, the data shows very strong seasonality and trend. In 2009 and 2013, there is a big outbreak, which makes sense. In 2009, the United States experiences H1N1 and in 2013, the pandemic is the H7N9. The whole trend is going upwards, which reflects the increasing population. Meanwhile, from the seasonality, we can conclude that influenza always occurs in the winter, and summer is a grace period. Indeed, because first of all, the difference between indoor and outdoor temperatures in winter makes people easy to catch colds and other flu. At the same time, there are fewer opportunities to open windows for ventilation, which increases the chance of infection. However, when the weather gradually warmed up, by April, the number of cases dropped very rapidly. The temperature at this time is not suitable for the survival of the virus, so the number of sick people has dropped.

I also use Prophet, a recently developed time series model to test the stationary of the data [14]. The results march the results I got by using ARIMA. Influenza always occurs in the winter from December to March, which is winter.

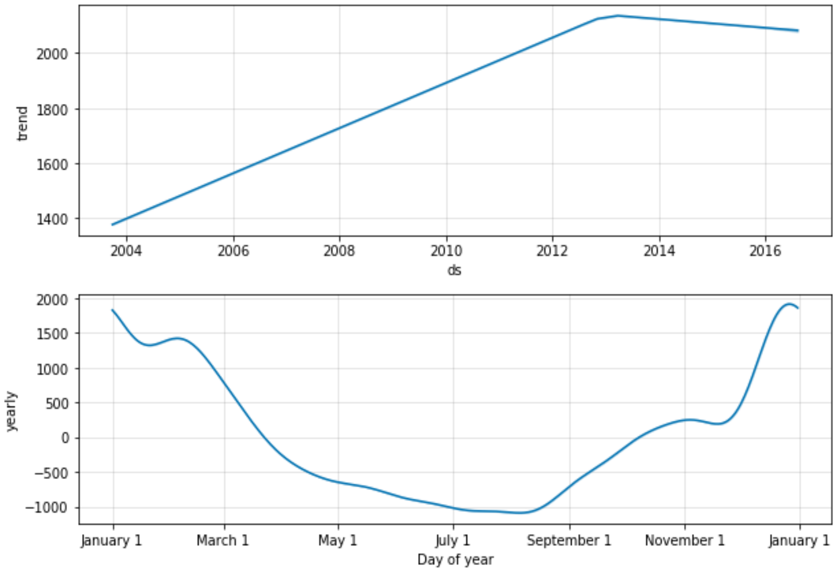

Figure 2 Prophet model shows the stationary of GFT

2.4. CDC

After testing the stability of GFT, I compare the data from GFT and CDC. Because seasonal influenza can cause serious consequences, the U.S. Centers for Disease Control and Prevention (CDC) monitors influenza-like diseases (ILI) in the United States and tracks and analyzes influenza activity to understand the incidence, prevalence, and incidence of international influenza throughout the

year ("10 flu myths" n.d.). The CDC monitors these numbers with data collected through multiple sources, including local and state health departments, 122 public health and vital statistics offices, nearly 3,000 outpatient health care facilities, more than 270 laboratories, and reports from the FluSurv-NET surveillance system.
a) Viral Surveillance — laboratory reports on the number of respiratory specimens taken that week and what percentage were, in fact, confirmed flu
b) Mortality — data on the proportion of pneumonia and influenza (P&I)-related deaths and reports of influenza-associated pediatric deaths
c) Hospitalizations — confirmed influenza-related hospitalizations
d) Outpatient Illness Surveillance — tracking the number of outpatient visits for ILI
e) Geographic Spread of Illness — the estimated level of flu activity by state, which could be widespread, regional, local, sporadic, or no activity. [13]
The data I used is from 10/4/2009 to 5/14/2017 and from 50 states.

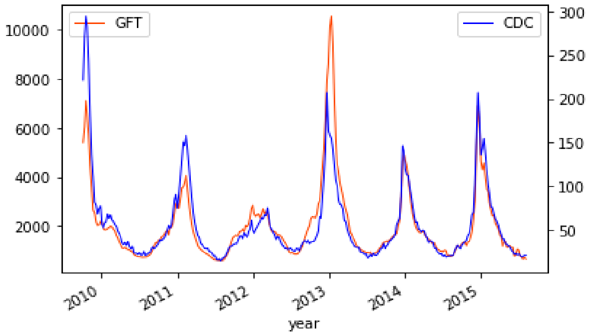

Figure 3 Compare the data from the CDC and GFT

At the beginning of each year (winter), the number of cases (CDC) will rise sharply. The red line represents GFT data, and the line represents CDC data, which is very close in terms of trends and shapes of the two lines, which means they are highly related. Both suddenly rose at the beginning of the year and fell in the middle of the year. However, there are situations in which extreme values do not match perfectly, such as in 2013.
The results can also be proved by Prophet. The overall trend is followed, but the still has a data leakage problem. To solve the problem, several deep learning models can be used.

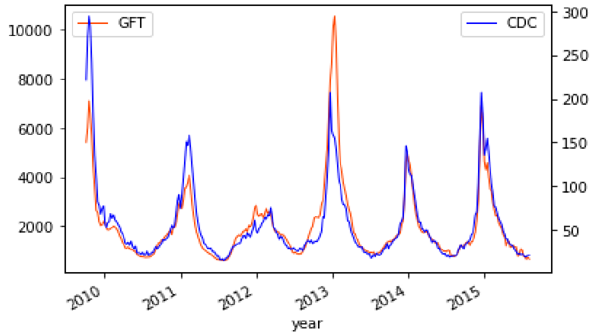

Figure 4 Prophet model prediction

3. METHODS

Both the time-series models Prophet and ARIMA prove that GFT data is time series data, so the data has certain characters, such as seasonality, repetition, and trend. Given that CDC and GFT are highly correlated, we can do some further and more accurate now cast by using deep learning models. In the deep learning models, I use three different models: Convolutional Neural Networks, Recurrent Neural Network, and seq2seq to estimate which one is the most suitable to predict the flu.

3.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) emerged from the study of the brain's visual cortex, and they have been used in image recognition since the 1980s. In the last few years, thanks to the increase in computational power, the amount of available training data, and the tricks for training deep nets, CNNs have managed to achieve superior performance on some complex visual tasks [14]. CNN consists of two parts: the feature learning part and the classification part. In feather learning parts, there are two layers: Convolution and Pooling. Convolution acts as a filter, which can filtrate important feathers of the data or pictures. Pooling's function is to reduce the dimensionality to in the network, and hence to also control overfitting. The classification part is the output part which can classify the sorts and assign a probability for the input image or data [16].

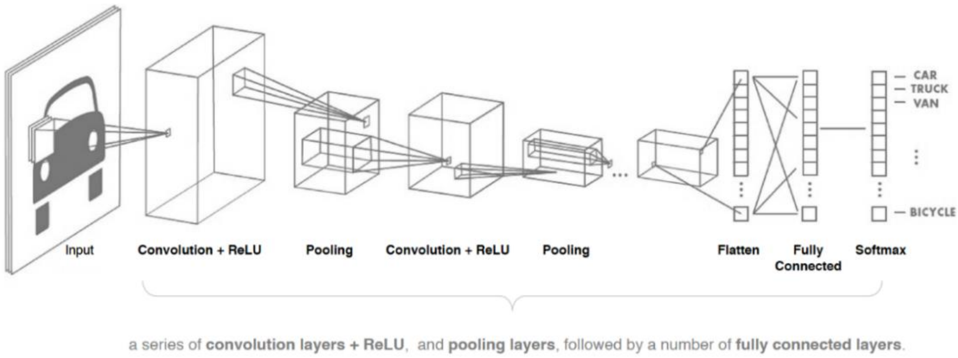

Figure 5 CNN Architecture ("Convolutional Neural Networks" n.d.)

3.2. Recurrent Neural Network

Another deep learning model is Recurrent Neural Network, which is called as RNN. Unlike CNN, RNN will add weights to different data. When analyzing a sequence of data, RNN preforms long short-term memory also called LSTM, which means the model will forget some previous data and remember the latest data. In LSTM, there are four gates: Forget gate, learn gate, update gate, and output gate. Forget gate is used to throw away some information from the cell state. Learn gate is used to decide what new information we're going to store in the cell state. Update gate is used to update the old cell state into the new cell state. Finally, output gate is used to output the data to the next cell. Comparing to the CNN model, RNN works better on time-series data. Since the data is highly related to the former ones and does not depend on very previous data [17].

3.3. Seq2seq

RNN also contains another model called seq2seq. RNN can be thought of as a natural extension of well-studied ARIMA models, but much more flexible and expressive. Since it is non-parametric, it can explore the "high-dimensional" time series setting, where a high quantity (100,000s+) of series must be forecast simultaneously. The model contains two parts: encoder and decoder. Train and validation both contain these two parts. In this article, I use a walk-forward split. The time frame for validation is shifted forward by one prediction interval relative to the time frame for training.

The work of the encoder is to transfer the data to an encoder state, and the work of the decoder is to code inside the loop gets a prediction from the previous step, and appends it to the input features for the current step [19].

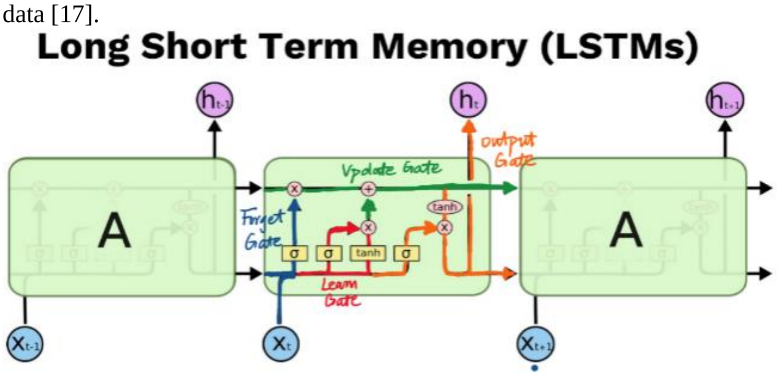

Figure 6 Principles of RNN [18]

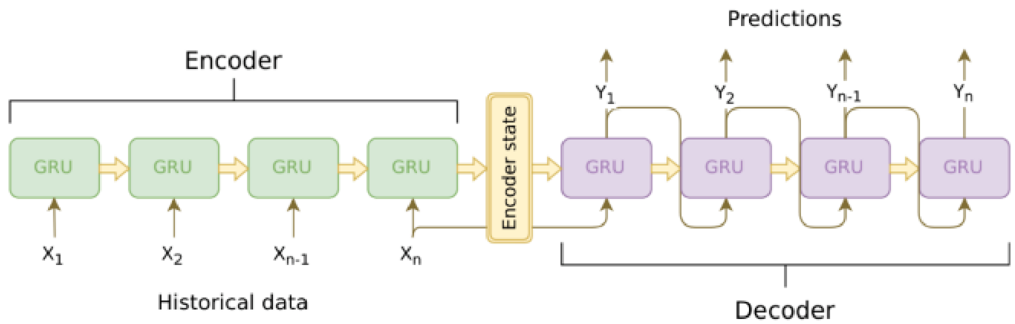

Figure 7 Principles of S2S [19]

4. RESULTS

4.1. CNN and RNN

I separate the dataset into train, validation, and test sets. I train the model on the train set. The validation set is used to evaluate the model after each training epoch and ensure that the model is not overfitting the training data. Then the data should be prepared by the following steps:

a) Filter the original dataset to include only that time period reserved for the training set

b) Scale the time series such that the values fall within the interval (0, 1)

c) Shift the values of the time series to create a database containing all the data for a single training example

d) Discard any samples with missing values

e) Transform this database into an array of shape (samples, features) for input into deep learning models

(All the above steps can be assisted by Python.)

After the model has finished training, we evaluate the model on the test set. Follow a similar process for the validation set and keep T hours from the training set in order to construct initial features. Ensure that the validation set and test set cover a later period in time from the training set, to ensure that the model does not gain from information from future time periods. Then implement the neural network with certain and suitable layers and neurons. Use Adam optimizer and mean squared error as the loss function, which can help me to determine the early stopping point.

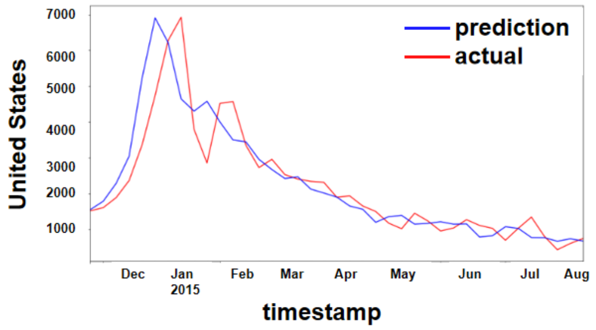

Figure 8 Result of CNN model

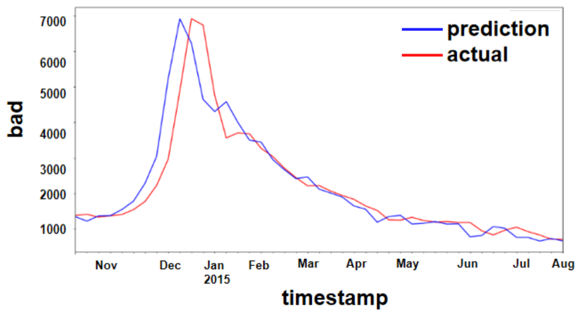

Figure 9 Result of RNN model

Two models both perform very well. Comparing the accuracy, the mean square error of RNN is 0.13 and that of CNN is 0.18, RNN performs 27.8% more accurately. Since the GFT is time-series data, our result validates the common findings where RNN typically performs better than CNN on time series data. Hence, the characteristic of the RNN helps the model performs better.

4.2. S2S

In S2S, the first step is to format the data for modeling: create 4 sub-segments of the data: Train encoding period, Train decoding period, Validation encoding period, Validation decoding period. Pull the time series into an array, save a date_to_index mapping as a utility for referencing into the array.

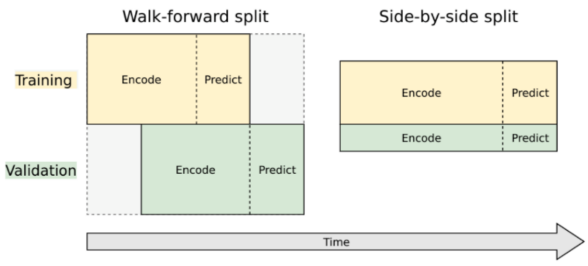

Figure 10 Data separation of S2S

a) Create a function to extract specified time intervals from all the series. Create functions to transform all the series. Smooth out the scale by taking log and de-meaning each series using the encoder series mean, then reshape to the (n_series, n_timesteps, n_features) format.
b) Build the Training Architecture model: during training, the true series values (lagged by one-time step) are fed as inputs to the decoder. At prediction time, the true values in this process will be replaced by predicted values for each previous time step. To get better results, the model can be adjusted to the hyperparameters, tweaked the learning rate and a number of epochs.
c) The third step is to build the inference model: use models to define an inference model that draws on the neural network to actually generate predictions.
(All the above steps can be assisted by Python.)
In a nutshell, this architecture starts by encoding the input series, then generates predictions one by one. The decoder gets fed initial state vectors from the encoder, but the state vectors are then iteratively updated as the decoder generates a prediction for each time step.
The results are quite good. The prediction (green line) is very close to the target series organized line.

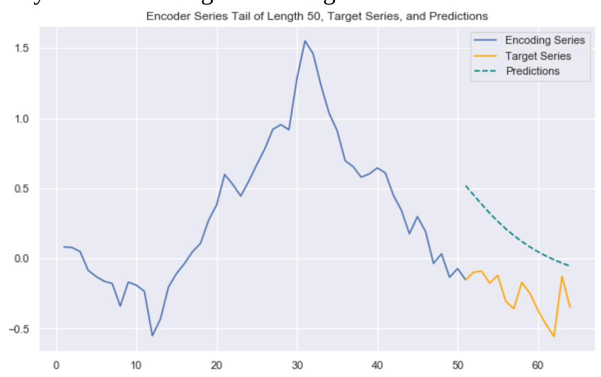

Figure 11 Results of S2S

4.3. Multivariate Time-Series Models

Multivariate also performs quite well. The green point is the predicted ILI occurrence using GFT and the blue one is the data from the CDC that records historical ILI occurrences. Multivariate time-series models even predict the small decline or increase very precise, such as 2015.

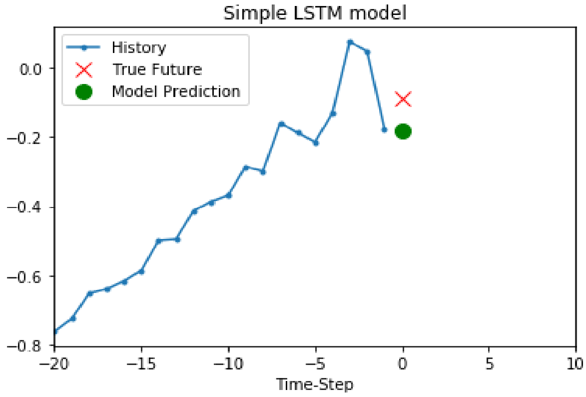

Figure 12 Result 1 of Multivariate

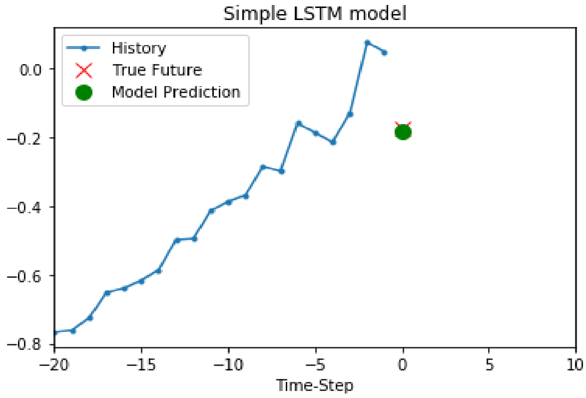

Figure 13 Result 2 of Multivariate

4.4. Further exploration

In the nutshell, the RNN model is the most suitable model to predict Influenza. Hence, I use RNN to predict the states in four different regions in America.

From the table, as for the settings, south and north performs better than east and west. South performs the best. Comparing with east and west, the east performs better. I conclude that the errors of prediction for states in the south and north are less than east and west, maybe because the flow of population is less, so the data from these settings are more credible. In the south, the weather is warmer, which means some viruses cannot survive in hot weather, so people will be infected are less. If the extreme number is not so big, the prediction will be better. Meanwhile, the Mean Square Error Value is varied from region to region, means that the for some regions the models can predict well, some cannot not, which is controversial to what Donald investigate: At least, in some regions the model performs well and the flu trend is predictable. However, the result remains more explorations in the future.

Table 1 Assessment of forecasts for the united states and its states

States	**Settings**	**Mean Square Error Value**	**Graph number**
Texas	South	0.09	14
New York	North East	0.11	15
Florida	Southeast	0.12	16
US		0.14	17
Washington	Northwest	0.16	18
Michigan	East	0.20	19
California	West	0.21	20
Kansas	Middle	0.25	21

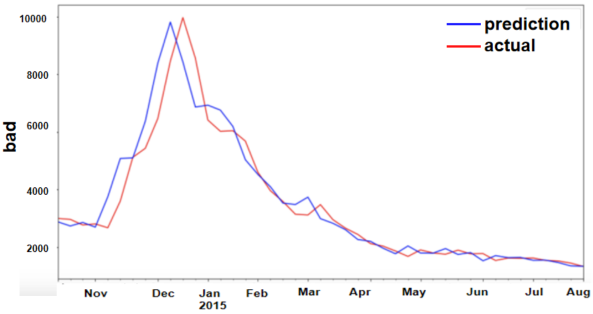

Figure 14 Texas

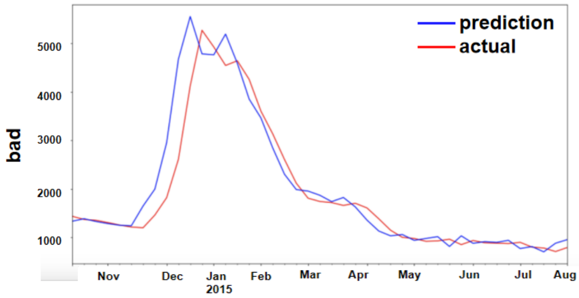

Figure 15 New York

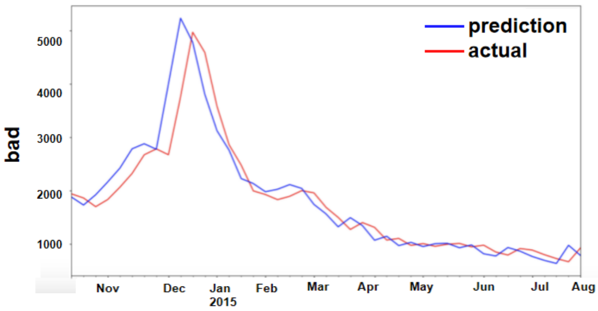

Figure 16 Florida

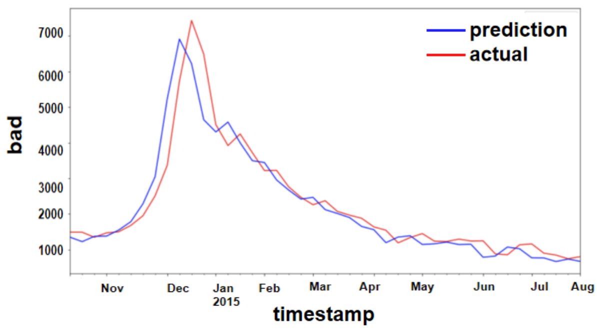

Figure 17 United State

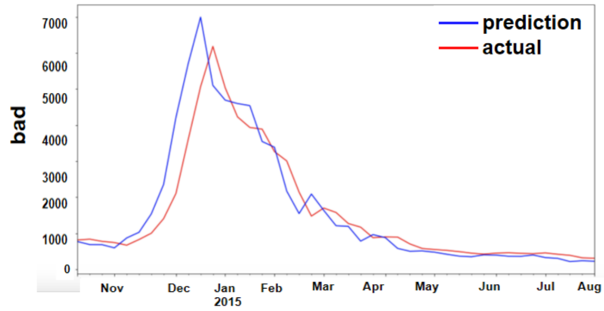

Figure 18 Washington

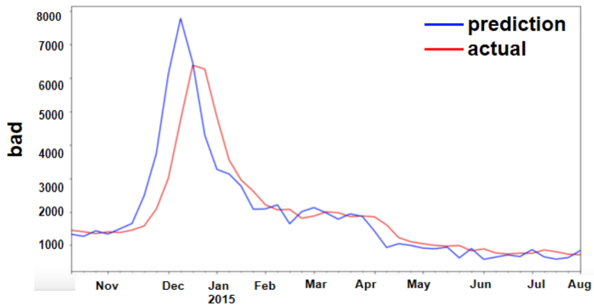

Figure 19 Michigan

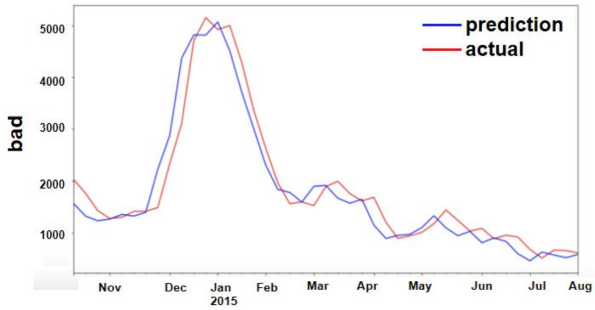

Figure 20 California

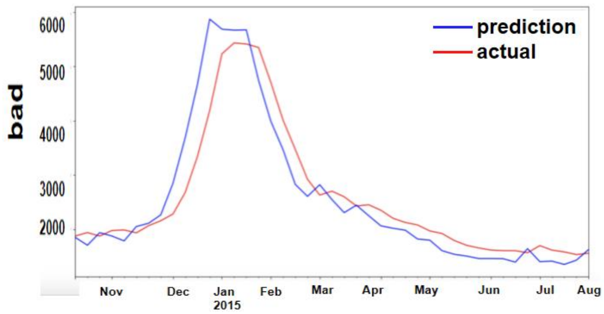

Figure 21 Kansas

5. CONCLUSION

I compare the data from GFT and CDC ILI and observe a strong correlation. However, given that CDC ILI is after-facts, scientists can only make conclusions based on past data which prohibit CDC to take a proactive role. On the contrary, GFT is nearly real-time and shows the people's attentional behaviors across different regions. If scientists can use GFT as a reliable source to estimate how many occurrences of ILI will be in the near future, governments and medical organizations can take precautions to alleviate emergencies. As a result, I use four models to estimate the retrospective and prospective performance of GFT to capture the season-to-season epidemic time. I found that the Deep Learning Model performs reasonably well. The results are consistent with the findings of Kandula and Shaman that the value of GFT data should be reappraised, if advanced models are introduced and now casted than classical epidemiological models used in prior studies [20]. Therefore, scientists can use the Deep Learning Model on GFT data to now cast the outbreak of influenza. It is worth noting that in this information age, people spend a significant amount of time online, and search engine is one of the most widely used sources to access to interesting content. With the population-level data from Google, the model can learn the real cases of the world with high accuracy.

In 2009, the influenza virus H1N1 spread all over the world. The outbreak provides a chance for the GFT model to predict the scale and testify the accuracy. The GFT model is highly correlated with ILINet data, which implies that it is likely to use GFT to now cast the H1N1 occurrences. However, the data in H1N1 is underestimated. Samantha Cook purposed several reasons why the result is underestimated. First, fewer people search the keywords, such as bronchitis and pneumonia, during pHIN1. Secondly, the eruption was breaking up in spring and summer. Thus, people may think the illnesses are not influenza, so they use different queries. Thirdly, the author uses the ILI to train the model, so the data may differ from ILINet which is different from ILI.

The models I use also have the same problem that the prediction is underestimated. According to Cook, I proposed the limitations of the study. First, the keywords may differ from people, because they have different symptoms, which means the GFT may miss some keywords that are also related to influenza. Second, people now are more knowledgeable. At first, they think they just have a cold or fever, which they have already grasped lots of solutions to overcome. Therefore, they will not use the search engine and the data from GFT will be less predictable of the reality. To remedy this issue, Lu et al. proposed to complement Google search frequency with the near-real time health record data and built ensemble machine learning models to now cast the ILI activity, which lays a future research direction [21].

At the same time, the result also raises some questions. First, now we know that the GFT is highly related to Influenza. Hence, how many people search the keywords such as cold, the government, or the board of health need to do some action? If the number is smaller, which means it is not serious influenza, the government will waste lots of money and energy. Or if the number is too big, maybe the situation has already reached an uncontrolled level. Second, some places, such as New York and California, have lots of temporary population, the data from these States cannot reflect the true situation. Thirdly the results that solely depend on the GFT data may be subject to the overestimation/underestimation of the actual ILI occurrences, which in turn result in a detrimental effect in the surveillance and responses to the ILI. As discussed in Olson et al. [22], classical time series models completely significantly underestimated the first wave of the 2009 influenza H1N1 pandemic, and overestimated the intensity of the H3N2 epidemic during the 2013 season. Nevertheless, by combing the data with deep learning models, the results with higher correlations and lower prediction errors. Meanwhile, our results should be used with caution, and require further analysis on how to combine its now cast prediction with other data sources.

Finally, the work of now cast ILI using GFT can also shed lights on monitoring and tracking the ongoing COVID-19 pandemic. Silverman et al., reported that if the surveillance system used to track the ILI had been also adopted to estimate the actual prevalence of COVID-19 cases, more than 80% of individuals with COVID-19 infections would have been detected just in March 2020 [23]. With this regard, the model used in this work may also be considered to now cast the COVID-19 occurrences with reasonable adaptation.

REFERENCES

[1] Helft M. Google Uses Searches to Track Flu's Spread. The New York Times; Sect A: 1. http://www.nytimes.com/2008/11/12/technology/internet/12flu.html?_r=1&, 2008. Accessed: 21-04-2020.

[2] R.E. Thomas. The case for ILI surveillance. Vaccine,33(48), 2015, pp. 6514-6514.

[3] Yang, S., Santillana, M., & Kou, S. C. Accurate estimation of influenza epidemics using google search data via argo. Proceedings of the National Academy of Sciences of the United States of America, 112(47), 2015, pp. 14473.

[4] Yang, S., Santillana, M., Brownstein, J.S., et al. Using electronic health records and Internet search information for accurate influenza forecasting. BMC Infect Disease 17, 332, 2017. doi： 10.1186/s12879-017-2424-7

[5] Ginsberg, J., Mohebbi, M., Patel, R. et al. Detecting influenza epidemics using search engine query data. Nature 457, 2009, pp. 1012–1014. doi: 10.1038/nature07634

[6] Cook S, Conrad C, Fowlkes AL, Mohebbi MH Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic. PLoS ONE 6(8), 2011, e23610.

[7] Gog, J. R., Ballesteros, Sébastien, Viboud, Cécile, Simonsen, L., Bjornstad, O. N., & Shaman, J., et al. Spatial transmission of 2009 pandemic influenza in the us. PLoS Computational Biology, 10(6), 2014, e1003635.

[8] Justin, R., Ortiz, Hong, Zhou, & David, et al. Monitoring influenza activity in the united states: a comparison of traditional surveillance systems with google flu trends. PloS ONE 6(4), 2011, e18687

[9] Lzaer, D., Kennedy, R., King, G., & A. Vespignani (2014). The Parable of Google Flu: Traps in Big Data Analysis, Science, 343, pp. 1203-1205.

[10] D. Butler, When Google Got Flu Wrong, Nature, 494, 2013, pp. 155-156.

[11] Santillana, M., Zhang, W., Althouse, B.M., & Ayers, J.W. What can Digital Disease Detection Learn from (an External Revision to) Google Flu Trends? American Journal of Preventive Medicine, 47(3), 2014, pp. 341-347.

[12] Fox, Susannah. "The social life of health information." Pew Research Center. http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/, 2014. Accessed:30-09-2020.

[13] Maria, T. How Google Flu Trends Works. How staff works. https://health.howstuffworks.com/medicine/modern-treatments/google-flu-trends.htm, 2014. Accessed: 13-04-2020.

[14] Taylor, S.J. & Letham, B. Forecasting at Scale, PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.3190v2. 2017.

[15] Douglass, M. J. J. Book review: hands-on machine learning with scikit-learn, keras, and tensorflow, 2nd edition by Aurélien Géron. Physical and Engineering in Medicine, 2017. 1-2.

[16] Krizhevsky, A., Sutskever, I., & Hinton, G. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25(2), 2012.

[17] Hochreiter, Sepp, Schmidhuber, & Jurgen. Long short-term memory. Neural Computation 9(8), 1997, pp. 1735-1780.

[18] "Understanding LSTM Networks" http://colah.github.io/posts/2015-08-Understanding-LSTMs, 2015. Accessed: 30-09-2020.

[19] Sutskever, I., Vinyals, O., & Le, Q. V. Sequence to sequence learning with neural networks. Advances in neural information processing systems, 2014.

[20] Kandula, S. & Shaman, J. Reappraising the utility of Google Flu Trends, PLoS Computational Biology, 15(8), 2019, e1007258.

[21] Lu, F. S., et al. Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. Nature Communications, 10(147), 2019, pp. 1-10.

[22] Olson, D. R., Konty, K. J., & Paladini, M., et al. Reassessing google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PloS Computational Biology, 9(10), 2013, e1003256.

[23] Silverman, J. D., Hupert, N., & Washburne, A. D. Using influenza surveillance networks to estimate state-specific prevalence of sars-cov-2 in the united states. Science translational medicine, 12(554), 2020, eabc1126.

[24] "Convolutional Neural Network 3 things you need to know" https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html, 2020. Accessed: 30-09-2020.

[25] Centers for Disease Control and Prevention. "Influenza (Flu)." Retrieved from: http://www.cdc.gov/flu/Accessed on 30, September 2020.

[26] Donald R., Konty, Kevin J., Paladini, & Marc, et al. Reassessing google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Computational Biology. 2013.

[27] Harvard Medical School - Harvard University. "10 flu myths." http://www.health.harvard.edu/flu-resource-center/10-flu-myths.htm, 2020. Accessed: 30-09-2020.

[28] Purpose and Methods. CDC. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/purpose-methods.html, 2020. Accessed: 23-04-2020.