# Detecting Influenza Outbreaks in United States by Analyzing Climatic Heat Maps Using Convolutional Neural Network

Juhyeon Kim

*Abstract*—**Detecting influenza infection pattern is critical as influenza is threatening global human health conditions for hundreds of years. There had been many different types of researches to estimate the incidences of influenza infection. However, detecting influenza infection pattern requires data which is accumulated for long-term. Commonly it is not easy to gain data and manage it constantly. Nevertheless, climate data which is highly correlated with influenza infection is usually well accumulated constantly as most governments and weather centers are very attentive to climate. There are several kinds of climatic data such as numerically recorded data or data which is indicated with heat map. In data conducted as heat map case, it is very effective to apply them to convolutional neural network model which is new emerging machine learning technique. In this research, we tried to estimate the influenza infection incidences using climate heat map with CNN and compared its prediction accuracy with prediction accuracy of support vector regression model using numerical climate data.**

## I. Introduction

IT is important to response rapidly to a health epidemic so we can reduce the loss of life. In the past, existing methods mostly rely on expensive surveys of hospitals. However, these hospital surveys are usually not the newest data but has lags about 2 to 3 weeks. Recently, researchers suggest many different methods that does not requires hospital survey data including machine learning models to detect the pattern of influenza infection. Instead of using hospital survey data, researchers tend to use data that can be obtained easily such as climatic data, social network service data, and so on recently.

Autoregression model was used to forecast for 1 to 10 week ahead of the incidences of influenza-like illnesses in [1]. In [1], the incidences of influenza-like illnesses in France are used and authors proved that appropriate for forecasting influenza-like illnesses incidences up to 10 weeks in advance during the epidemic and pre-epidemic periods. Some researchers used Twitter data to prove that it can improve seasonal influenza prediction. In [2] authors claim that Twitter data is highly correlated with the influenza-like illness rates across different regions within USA and can be used to effectively improve the accuracy of prediction. Research showed best fit on age groups of 5-24 and 25-49 years because people in these groups are the most active user age groups on Twitter. Authors of [3] collected 3.6 million flu-related tweets from 0.9 million Twitter users starting from 2008 to 2010. In [3], a probabilistic graphical Bayesian approach based on Markov Network model was used and real-time surveillance system is suggested. Not only Twitter data but al so some other web data such as Google Flu Trends was used in other researches. The combination of three different data which are Google Flu Trends, meteorological data, and temporal information was used in [4] to predict one week influenza cases. Research showed that adding Google Flu Trends data enhanced the forecasting accuracy dramatically.

While various methods for predicting the number of influenza infected people are being found, there were no attempt of using image data with convolutional neural network yet. Recently, convolutional neural network is considered in diverse fields as it shows outstanding prediction performance when there are enough amounts of data. The success of convolutional neural network is attributed their ability to learn mid-level images. For this reason, we considered using convolutional neural network for estimating the number of influenza infected people in this research using image data.

The remainder of the paper is organized as follows. The next section introduces convolutional neural network, which is the algorithm that we are challenging in this research and support vector regression, which is the algorithm that is used for prediction performance comparison. The section on experiments provides the experimental results for a comparison of convolutional neural network and support vector machine. We used the temperature heat map of United States for 104 weeks which is produced by National Oceanic and Atmospheric Administration (NOAA) as independent variables for convolutional neural network and the minimum weekly temperature of United States accumulated from 424 different stations which is produced by United States Historical Climatology Network as independent variables for support vector regression. Moreover, the number of influenza infection counts in United States produced by FluNet is used as predictor. The final section presents our conclusions.

## II. Methods

### A. Convolutional Neural Network

Convolutional neural network incorporates constraints and achieve some degree of shift and deformation invariance using three ideas: local receptive fields, shared weights, and

This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea Government (MSIP) (No. CRC-16-01-KRICT)

Juhyeon Kim is with the Center for Convergent Research of Emerging Virus Infection, Korea Research Institute of Chemical Technology, 141, Gajeong-ro, Yuseong-gu, Daejeon 34114, Korea (e-mail: juhyeon@krict.re.kr).

spatial subsampling. The use of shared weights also reduces the number of parameters in the system aiding generalization. A typical convolutional neural network is shown in figure 1[5]. The network consists of a set of layers each of which contains one or more planes. Approximately centered and normalized images enter at the input layer. Each unit in a plane receives input from a small neighborhood in the planes of the previous layer. The idea of connecting units to local receptive fields dates back to the 1960's with the perceptron and Hubel and Wiesel's [6] discovery of locally sensitive orientation selective neurons in the cat's visual system [7]. The weights forming the receptive field for a plane are forced to be equal at all points in the plane. Each plane can be considered as a feature map which has a fixed feature detector that is convolved with a local window is scanned over the planes in the previous layer. Multiple planes are usually used in each layer so that multiple features can be detected. These layers are called convolutional layers. Once a feature has been detected, its exact location is less important. The network is trained with the usual backpropagation gradient-descent procedure. A connection strategy can be used to reduce the number of weights in the network.

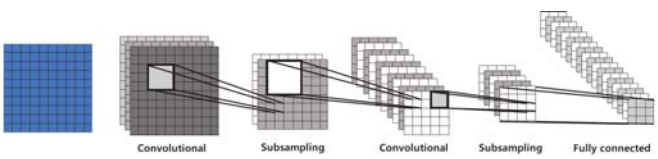


Figure 1. A typical convolutional neural network

### B. *Support Vector Regression*

The support vector regression is recently introduced machine learning theory based on support vector machines. First of all, to solve a nonlinear problem in SVR, let $x_i$ be mapped into a feature space by a nonlinear function $\phi(x)$; the decision function becomes:

$$f(x) = w \cdot \phi(x) + b \tag{1}$$

where, $w$ is weight vector, $b$ is a constant, and $w \cdot \phi(x)$ describes the dot production in the feature space. Similarly, the nonlinear regression problem can be expressed as the following optimization problem (2).

$$\begin{aligned} &\min_{w,b,\xi,\xi^*} && \frac{1}{2}w^2 + C\sum_{i=1}^{l} \xi_i + \xi_i^* \\ &\text{subject to} && y_i\text{-}(w \cdot \phi(x_i) + b) \le \varepsilon + \xi_i \\ & && (w \cdot \phi(x_i) + b)\text{-}y_i \le \varepsilon + \xi_i \\ & && \xi_i, \xi_i^* \ge 0, i = 1, 2, \dots, l \end{aligned} \tag{2}$$

Then, the dual form of the nonlinear SVR can be expressed as (3).

$$\begin{aligned} &\min_{\underline{\alpha}_i, \overline{a}_i} && \frac{1}{2}\sum_{i,j=1}^{l} (\underline{\alpha}_i\text{-}\overline{a}_i)(\underline{\alpha}_j\text{-}\overline{a}_j)\{(x_i) \cdot \phi(x_j)\} \\ & && +\varepsilon\sum_{i=1}^{l}(\underline{\alpha}_i + \overline{a}_i)\text{-}\sum_{i=1}^{l} y_i(\underline{\alpha}_i\text{-}\overline{a}_i) \\ &\text{subject to} && \sum_{i=1}^{l}(\underline{\alpha}_i\text{-}\overline{a}_i) = 0 \\ & && 0 \le \underline{\alpha}_i \le \mathrm{C}, i = 1, 2, \dots, l \\ & && 0 \le \overline{a}_i \le \mathrm{C}, i = 1, 2, \dots, l \end{aligned} \tag{3}$$

While it is very complex to perform the computation of $\phi(x_i) \cdot \phi(x_j)$ which is in the feature space, it is not need to be use nonlinear function $\phi(x_i)$ in SVM. The computation in input space can be performed using a "kernel" function $K(x_i, x_j) = \{\phi(x_i) \cdot \phi(x_j)\}$ to yield the inner products in feature space, circumventing the problems intrinsic in evaluating the feature space. Functions that meet Mercer's condition can be proven to correspond to dot products in a feature space [10]. Therefore, any functions that satisfy Mercer's theorem can be used as kernel. Some commonly used kernels in SVM are as follows:

Linear kernel

$$K(x_i, x_j) = x_i \cdot x_j \tag{4}$$

Polynomial kernel

$$K(x_i, x_j) = [\gamma(x_i \cdot x_j) + c]^d \tag{5}$$

Sigmoid kernel

$$K(x_i, x_j) = tanh[\gamma(x_i \cdot x_j) + c] \tag{6}$$

Radial basis function kernel

$$K(x_i, x_j) = \exp\,(\text{-}\gamma|x_i\text{-}x_j|^2) \tag{7}$$

Finally, the kernel function allows the decision function of nonlinear SVR to be expressed as (8).

$$f(x)\sum_{i=1}^{l}(\text{-}\underline{\alpha}_k + \overline{a}_k)K(x_i, x_k) + b \tag{8}$$

The parameter $c$ controls the smoothness or flatness of the approximation function. A greater $c$ value, corresponding to a greater penalty of errors, indicates that the objective is only to minimize the empirical risk, which makes the learning machine more complex. A smaller $c$ value may cause the errors too be excessively tolerated, yielding a learning machine with poor approximation. If the data are noisy, then smaller $c$ values, which penalize the errors less, may be preferred. The parameter $\varepsilon$ also affects the smoothness or complexity of the approximation function. More importantly, $\varepsilon$ dominates the number of support vectors, since it governs the accuracy of the approximation function. Smaller values of $\varepsilon$ may lead to more support vectors and result in a complex learning machine. Greater $\varepsilon$ values may cause the $\varepsilon$-insensitive tube to encompass too many data that are unseen by the learning machine, so some important information hidden in these data may be lost, resulting in "flattening" the regression function. The SVM model used a Polynomial kernel function that has been known as an excellent performance model relatively.

## III. EXPERIMENTS

The number of influenza infection counts in United States dataset from FluNet was used for the experiment, which is a global web-based tool for influenza virological surveillance first launched in 1997. FluNet claims to have one of the most comprehensive collections of virological data. They service data at country level and are publically available which are updated weekly. Furthermore, data are presented in various formats including tables, mas and graphs. We used the temperature heat map of United States for 410 weeks which is produced by National Oceanic and Atmospheric Administration (NOAA) as independent variables for convolutional neural network. The samples of heat maps are shown in figure 2. As a feature for support vector regression model, daily temperature observations from the National Climatic Data Center's (NCDC) Global Historical Climatology Network (GHCN) daily database is used. Each temperature data points are collected from 424 different stations through all over the United States. As well as the counts of influenza infection, the average temperature of 104 weeks in United States is used in this study. 32 Weeks (about 30% of whole data set), which are the rest of data were reserved for out-of-sample evaluation and comparison of performances among the two models.

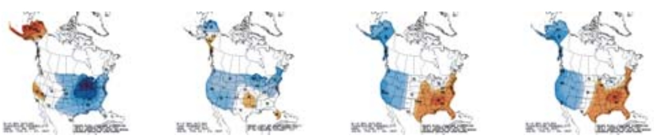

Figure 2. Samples of heat maps

We used the root mean square error (RMSE) between predicted variance of infection number and real data as performance measure. RMSE is a measure of the difference between locations that have been interpolated or digitized. It is derived by squaring the differences between known and unknown points, adding those together, dividing that by the number of test points, and then taking the square root of that result. RMSE can be explained as (9).

$$RMSE = \sqrt{\frac{e_1^2 + e_2^2 + e_3^2 + \cdots + e_n^2}{n}} \tag{9}$$

The convolutional neural network model and support vector regression model were used to perform forecasting for the number of influenza incidences. Through whole experiments, data were standardized and results are also shown as standardized values.

## IV. RESULTS

Table 1 shows a comparison of the results with convolutional neural network and support vector regression in terms of root mean square error. The best performance among the two models is marked in bold face. In terms of root mean square error, support vector regression produced root mean square error of 0.5854 which was dramatically better than that of convolutional neural network. Figure 3 summarizes the performance of the two models and original values using graphs from $1^{st}$ week to $32^{nd}$ week.

Table 1. The prediction results of CNN and SVR

| Date | Original Values | CNN | SVR |
|---|---|---|---|
| 2009.5.21 | 0.0075 | 0.6432 | 0.1619 |
| 2009.5.28 | 0.0020 | 0.3459 | 0.0668 |
| 2009.6.4 | 0.0020 | 0.3249 | 0.0335 |
| 2009.6.11 | 0.0020 | 0.2970 | 0.0265 |
| 2009.6.18 | 0.0035 | 0.0564 | 0.0068 |
| 2009.6.25 | 0.0035 | 0.3827 | 0.0042 |
| 2009.7.2 | 0.0015 | 0.3467 | 0.0080 |
| 2009.7.9 | 0.0000 | 0.5243 | 0.0033 |
| 2009.7.16 | 0.0035 | 0.6884 | 0.0004 |
| 2009.7.23 | 0.0040 | 0.1810 | 0.0000 |
| 2009.7.30 | 0.0100 | 0.4734 | 0.0005 |
| 2009.8.6 | 0.0075 | 0.3682 | 0.0035 |
| 2009.8.13 | 0.0140 | 0.3689 | 0.0475 |
| 2009.8.20 | 0.0085 | 1.0000 | 0.0325 |
| 2009.8.27 | 0.0206 | 0.6494 | 0.0193 |
| 2009.9.3 | 0.0150 | 0.4954 | 0.0197 |
| 2009.9.10 | 0.0211 | 0.2511 | 0.0238 |
| 2009.9.17 | 0.0301 | 0.2722 | 0.0190 |
| 2009.9.24 | 0.0316 | 0.0000 | 0.0504 |
| 2009.10.1 | 0.0291 | 0.3107 | 0.0731 |
| 2009.10.8 | 0.0366 | 0.5660 | 0.0568 |
| 2009.10.15 | 0.0622 | 0.6227 | 0.0272 |
| 2009.10.22 | 0.0687 | 0.8364 | 0.0400 |
| 2009.10.29 | 0.0822 | 0.3281 | 0.0400 |
| 2009.11.5 | 0.1219 | 0.0405 | 0.0440 |
| 2009.11.12 | 0.1745 | 0.2995 | 0.0483 |
| 2009.11.19 | 0.2442 | 0.3622 | 0.0128 |
| 2009.11.26 | 0.2929 | 0.5243 | 0.0169 |
| 2009.12.3 | 0.3084 | 0.4257 | 0.0745 |
| 2009.12.10 | 0.4208 | 0.5585 | 0.1861 |
| 2009.12.17 | 0.7192 | 0.3069 | 0.9264 |
| 2009.12.24 | 1.0000 | 0.6014 | 1.0000 |
| **RMSE** | | 2.3920 | **0.5854** |

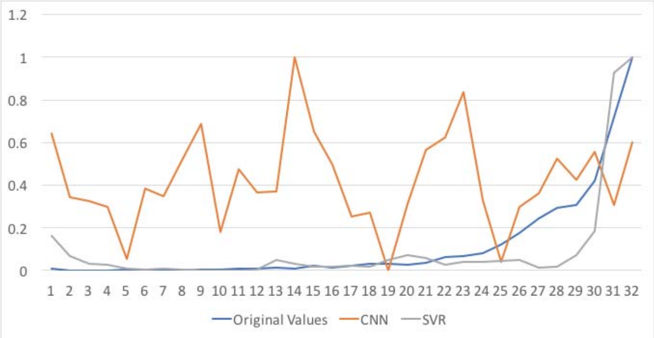


Figure 3. The prediction results of CNN and SVR

## V. Conclusion

Convolutional neural network is one of the most emerging machine learning technique these days. However, convolutional neural network requires voluminous data to get acceptable prediction result. Even though, researchers are attracted to using convolutional neural network model, the performance of convolutional neural network is dramatically low as shown in figure 3. Thus, as there are no enough climatic and incidence of influenza data it is not suitable to use convolutional neural network on forecasting the number of influenza infection yet. On the other hand, support vector regression model showed satisfactory performance compared to convolutional neural network model even though there were not enough data to use. Even though it shows little gap between original values, it follows the trend of original data.

In forecasting the number of influenza infected people problem, there are several obstacles when we try to use machine learning models. However, support vector regression showed that it follows the trend of original data. Nevertheless, in this research, only ground version of support vector regression is used, more sophisticated methods will be required to get better precision performance in the future. Furthermore, to make the best use of convolutional neural network model, we need to find the solution of gathering sufficient amount of data or amplifying gathered data.

## Acknowledgment

This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIP) (No. CRC-16-01-KRICT).

## References

[1] C Viboud, PY Boelle, F Carrat, et al. Prediction of the spread of influenza epidemics by the method of analogues. Am J Epidemiol. 158, 10 (Nov. 2003), 996-1006

[2] H Achrekar, A Gandhe, R Lazarus, SH Yu, B Liu. Twitter improves seasonal influenza prediction. In International Conference on Health Informatics (HEALTHINF'12). Vilamoura, Portugal: Nature Publishing Group, based in London, UK; 2012:61-70

[3] J. Li and C. Cardie. Early stage influenza detection from twitter. arXiv prepreint arXiv:1309.7340, 2013.

[4] Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, et al. (2013) Influenza forecasting with Google Flu Trends. Plos One 8: e56176

[5] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network, 1989. Neural Information Processing Systems (NIPS).

[6] D. H. Hubel, T. N. Wiesel, Receptive fields binocular interaction and functional architecture in the cat's visual cortex, J. Physiology (London), vol. 160, pp. 106-154, 1962.

[7] M. A. Arbib, Convolutional networks for images speech and time series, The Handbook of Brain Theory and Neural Networks, MIT Press, PP. 255-258, 1995.

[8] U. Thissen, R. Van Brakel, A.P. De Weijer, W.J. Melssen, L.M.C. Buydens, Using support vector machines for time series prediction, Chemom. Intell. Lab. Syst. 1-2 (2003) 35-49.