Attention-Based Recurrent Multi-Channel Neural Network for Influenza Epidemic Prediction

Bofeng Fu[1]
Tianjin University
College of Intelligence and Computing
Tianjin, China
jxmg686000@163.com

Yaodong Yang[1]
Tianjin University
College of Intelligence and Computing
Tianjin, China
yydapple@gmail.com

Yu Ma
Guangzhou Center of Disease Control and Prevention
Guangdong, China
freedomsky_220@163.com

Jianye Hao
Tianjin University
College of Intelligence and Computing
Tianjin, China
haojianye@gmail.com

Siqi Chen
Tianjin University
College of Intelligence and Computing
Tianjin, China
siqi.chen09@gmail.com

Shuang Liu
Tianjin University
College of Intelligence and Computing
Tianjin, China
shuang.liu@tju.edu

Tiegang Li
Guangzhou Center of Disease Control and Prevention
Guangdong, China
155179402@qq.com

Zhenyu Liao*
Pony Testing International Group
Tianjin FoodSafety Inspection Technology Institute
Tianjin, China
liaozy08@163.com

Xianglei Zhu*
CATARC
Automotive Data Center
Tianjin, China
zhuxianglei@catarc.ac.cn

***Abstract*—Influenza is a contagious respiratory disease that can lead to serious illness. Due to its serious threat to public health, accurate real-time prediction of influenza outbreaks has a great value. In this paper, a novel deep neural network architecture is employed to provide a real-time ILI% in Guangzhou, China. Because of the long-term structure property and the diversity of influenza epidemic data, long short-term memory (LSTM) network can yield accurate prediction accuracy. We design a Multi-channel LSTM network to extract fused descriptor from multiple types of input. We further improve prediction accuracy by adding attention mechanism. This structure allows us to handle the relationship between multiple inputs more appropriately. The proposed model can make full use of information in the dataset, solving the actual problem of influenza epidemic prediction in Guangzhou with pertinence. The performance evaluates by comparing with different architectures and other state-of-art methods. The experiments show that our model has the most competitive result, and can provide the effective real-time prediction.**

***Index Terms*—Influenza Epidemic Prediction, Attention Mechanism, Multi-channel LSTM Neural Network**

I. Introduction

Influenza is an acute respiratory infection caused by influenza virus. Influenza can cause seasonal epidemics and global pandemics, which has a major effect on human health and socio-economics [1] [2]. Therefore, accurate real-time monitoring and prediction of influenza outbreaks have great values to the public health officials [3].

Influenza-like-illness (ILI) is an acute respiratory infection measurement defined by the World Health Organization (WHO). Recently, increasing researchers focus on accurate real-time monitoring, early detection, and prediction of influenza outbreaks. By using information from online search or social networks like Twitter, Google Correlate, influenza outbreaks prediction has become one of the most active research areas [4] [5] [6]. These approaches are usually based on some commonly used linear model, such as least absolute shrinkage and selection operator (LASSO) or penalized regression [4] [6] [7]. Some researchers also use deep learning methods to solve the influenza epidemic prediction problem [8] [9]. However, these models tend not to provide sufficiently accurate one-week ahead prediction of ILI%.

In this paper, we focus on using the deep neural network to solve the prediction problem. In recent years, deep learning models have achieved outstanding results in a variety of applications from computer vision, speech recognition to climate prediction [10] [11] [12]. We use Long-short term memory (LSTM) neural network [13] as the basic method for prediction due to the time-sequence property of influenza data. By designing a multi-channel structure, we can better extract the sequential information from data. We further improve our model by adding attention mechanism. This structure allows us to handle the relationship of input data between different districts more appropriately. We named our model as Att-MCLSTM, which stands for Attention-based multi-channel LSTM.

Our main contributions in this paper can summarize as

[1]Equal contribution
*Corresponding author

978-1-5386-5488-0/18/$31.00

follow: (1) We study on Guangzhou influenza surveillance dataset, which is authentic and reliable. It has multiple features and obvious characteristics. (2) We design an Attention-based multi-channel LSTM architecture that combines multiple well-performing methods. This architecture fully considers the prediction problem itself and the data characteristics. Our method can be viewed as a feasible alternative to predict influenza outbreaks in other regions. Experiment results also prove the effectiveness of our model. To the best of our knowledge, this is the first study that applies LSTM architecture to the field of influenza epidemic prediction.

The rest of the paper organize as follow. The proposed model describes in detail in Section II. Section III evaluates the performance of the proposed model by comparing with different neural network architectures and other state-of-art methods. Section IV gives the conclusion and the prospect for future work.

II. Methodology

The accuracy of the prediction can improve by combining several different methods [14]. In this paper, we designed an LSTM architecture to solve the influenza prediction problem in Guangzhou. To explain our model more clearly, we first describe our dataset. The following subsections give a further elaboration of the dataset, the general idea of our model, the details of LSTM, Attention mechanism, and Attention-based multi-channel LSTM.

A. Dataset Description

The influenza surveillance data is provided by the Guangzhou Center for Disease Control and Prevention. It contains 9 years of influenza epidemic data in 9 districts of Guangzhou from 2009 to 2017. This dataset includes 6 modules. Each module has multiple features. It has one record per week and includes 52 weeks per year. After preprocessing, normalization and feature selection, we choose 19 features that are more relevant to prediction target.

B. Design of the Proposed Model

Fig.1 shows the flowchart of the proposed model. The overall framework has two processes, training and testing. In the training process, after data cleaning and normalization, we select 19 relevant features. The feature selection method we use is the model-based ranking method. After that, we divide the dataset into training and testing sets. The training process uses 80 percent of data to capture the yearly trend and seasonal pattern. At last, we train our model using the training set. In the testing process, we test our model on the testing set. To restore the original values, the denormalization process is performed. Finally, we evaluate our method and compare with other methods.

C. Long-short Term Memory Network

The recurrent neural network has the capability to dynamically incorporate experience due to internal recurrence [15]. Unlike conventional RNN, LSTM can solve the problem of a vanishing gradient [16]. The memory unit of LSTM cell retains the sequential information of given context [16]. It has been proved that LSTM has better performance than the conventional RNN [17].

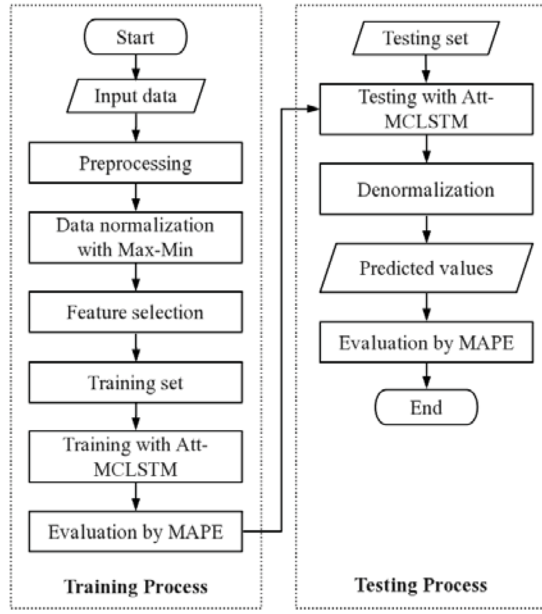

Fig. 1. The flowchart of Attention-based multi-channel LSTM.

LSTM memory cell has four units: input gate, output gate, forget gate, and self-recurrent neuron. LSTM is implemented by following composite function:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{2}$$

$$c_t = f_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{3}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{4}$$

$$h_t = o_t \tanh(c_t) \tag{5}$$

Where σ represent the logistic sigmoid function. i, f, o, and c represent the input gate, forget gate, output gate, cell input activation vectors respectively. h represents the hidden vector. The weight matrix subscripts has the intuitive meaning. Like, W_{hi} represents the hidden-input gate matrix etc.

D. Attention Mechanism

The basic idea of the Attention mechanism [18] is to break the traditional Encode-Decode structure and train a model to selectively learn these inputs by preserving the LSTM encoder's intermediate output. The output sequence associated with the input sequence.

The Attention layer inputs n parameters $y_1, \ldots, y_n$, context sequence c, and outputs vector z, z is the weighted distribution of y_i for a given context c. Attention mechanism is implemented by following composite function:

$$m_i = \tanh(W_{cm}c + W_{ym}y_i) \tag{6}$$

$$s_i \propto \exp(\langle w_m, m_i \rangle) \tag{7}$$

$$\sum_i s_i = 1 \tag{8}$$

$$z = \sum_i s_i y_i \tag{9}$$

Where m_i is calculated by $\tanh$ layer, s_i is the $softmax$ of the m_i projected on a learned direction. The output z is the weighted arithmetic mean of all y_i, W represents the relevance for each variable according to the context c.

E. Attention-based Multi-channel LSTM

Fig.2 shows the structure of the proposed model. Among all the features in the dataset, we divide them into two categories. We cluster average temperature, maximum temperature, minimum temperature, rainfall, air pressure, relative humidity together as climate-related data. The remaining features are grouped together as influenza-related data. Every week, each district has its own separate influenza-related data, and they share the same climate-related data.

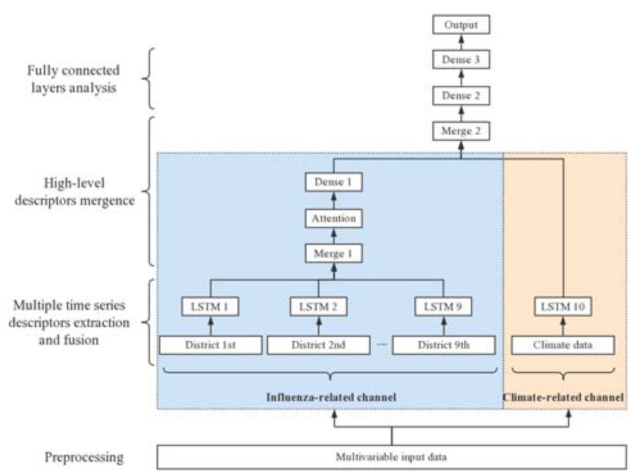

Fig. 2. The structure of Attention-based multi-channel LSTM.

Therefore, the input of the entire neural network has two parts. First, we use a group of LSTM networks (LSTM 1, ..., LSTM 9) to receive the influenza-related data and obtain the relevant descriptor. Second, a single LSTM network (LSTM 10) is applied to the climate-related data to extract its long-term structure property. In the first part, each LSTM network receives the influenza-related data from one district. To take advantage of the complementarity among every district, the outputs of these LSTM networks concatenate in the higher layer (Merge 1) to obtain the fused descriptor. Although this group of LSTM networks extracts characteristics of every district, we still need to weight the intermediate sequence, because the data of each district has a different influence on the municipal prediction result. So, the intermediate sequence passes through an attention layer (Attention) and a fully connected layer (Dense 1) in turn. After that, we concatenate the outputs of these two parts together (Merge 2). After passing through two fully connected layers (Dense 2, Dense 3), we obtain the high-level descriptor, and it's applied to the influenza epidemic prediction.

III. Experiments

In this section, we describe how we get the prediction value and evaluate the performance of our model. To verify our Att-MCLSTM model, we did two experiments. In the first experiment, we focus on how many consecutive weeks of data should we use to predict ILI% for the next week. In the second experiment, we verify the validity of our model. Every experimental result is the average of 10 repeated trials.

A. Selection of consecutive weeks

In this experiment, we test 6, 8, 10, 12, 14 weeks respectively. The structure of Att-MCLSTM is as shown in Fig.2.

We use the first 370 consecutive weeks' data for training and the rest of data in the testing process. Each data record contains 6 features of climate-related data and 9 districts' influenza-related data. Each influenza-related data includes 13 features. The climate-related data feeds into the climate-related channel, and each district's influenza-related data feeds into the influenza-related channel. The prediction results show in Table I.

TABLE I
The MAPE of the prediction results

Number of Weeks	MAPE
6	0.107
8	0.092
10	0.086
12	0.106
14	0.109

From Table I, we can see that 10 weeks data has the best result. This result indicates that 10 weeks data can best reflect the time-sequence property of influenza epidemic data. If input data is small, there is a lack of valid sequential information. In contrast, the noise of data increased. In the following experiments, each sample contains 10 weeks of data.

B. Performance Validation

In the second experiment, we verify the validity of our model.

- First, we verify the validity of attention mechanism by comparing the prediction accuracy of Att-MCLSTM and MCLSTM. We use the same multi-channel structure (as shown in Fig.2), and the only difference is whether there is an attention layer. The data input method is as described in the first experiment.
- Second, we verify the validity of the multi-channel structure by comparing the prediction accuracy of MCLSTM and LSTM network. For MCLSTM, data input method is as described in the first experiment. For the LSTM

network, a single LSTM layer is applied to all features to get fused descriptor. Instead of considering every district individually, we sum the corresponding features of all districts in the same week except the climate-related data. So, every data record contains 19 features. Then passing through a fully connected layer to obtain the high-level descriptor.

- Third, we prove that the LSTM is better than RNN. The parameter settings and data input method of these two models are the same as described above.

Table II shows the MAPE value of four models. Att-MCLSTM appears to deliver the most competitive prediction performance. The first two rows indicate that the introducing of attention mechanism enhance the MAPE from 0.105 to 0.086. Because attention layer allows us to handle the relationship of input data between different districts more appropriately. The second row and the third row show an improvement of MAPE from 0.118 to 0.105. By designing the multi-channel structure, we can better extract the time-sequence information of each input stream. The MAPE of LSTM and RNN is 0.118 and 0.132 respectively. The last two rows prove that LSTM has an advantage over conventional RNN when input data has the long-term structure property. They also indicate the time-sequence property of influenza epidemic data.

TABLE II
THE MAPE OF THE PREDICTION RESULTS

Schemes	MAPE
Att-MCLSTM	0.086
MCLSTM	0.105
LSTM	0.118
RNN	0.132

IV. CONCLUSION AND FUTURE WORK

In this paper, we present a novel deep neural network architecture (Att-MCLSTM) to predict the ILI% in Guangzhou, China. First, we use the Multi-channel structure to abstract sequential information from multiple input streams. Then, the attention mechanism performs on the fused descriptor sequence, which allows us to handle the relationship of multiple input streams more appropriately. The proposed model makes full use of information in the dataset, solving the actual influenza epidemic prediction problem in Guangzhou with pertinence. To verify the efficiency of the proposed model, we compare Att-MCLSTM with different architectures and other state-of-art methods. Att-MCLSTM enhances the MAPE to 0.086, delivers the most competitive prediction result. The experimental results prove the effectiveness of our model. To the best of our knowledge, this is the first study that applies LSTM architecture to the field of influenza epidemic prediction. Continuing work will further improve the expansion ability of our model by introducing transfer learning.

ACKNOWLEDGMENT

The work is supported by Tianjin Development Program for Innovation and Entrepreneurship and Special Program of Artificial Intelligence of Tianjin Municipal Science and Technology Commission (NO.: 17ZXRGGX00150).

REFERENCES

[1] N. Cox and K. Subbarao, “Global epidemiology of influenza: past and present,” *Annual review of medicine*, vol. 51, no. 1, pp. 407–421, 2000.
[2] I. Stephenson, J. Wood, K. Nicholson, and M. Zambon, “Sialic acid receptor specificity on erythrocytes affects detection of antibody to avian influenza haemagglutinin,” *Journal of medical virology*, vol. 70, no. 3, pp. 391–398, 2003.
[3] J. S. Brownstein and K. D. Mandl, “Reengineering real time outbreak detection systems for influenza epidemic monitoring,” in *AMIA Annual Symposium Proceedings*, vol. 2006. American Medical Informatics Association, 2006, p. 866.
[4] M. Santillana, D. W. Zhang, B. M. Althouse, and J. W. Ayers, “What can digital disease detection learn from (an external revision to) google flu trends?” *American journal of preventive medicine*, vol. 47, no. 3, pp. 341–347, 2014.
[5] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu, “Predicting flu trends using twitter data,” in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*. IEEE, 2011, pp. 702–707.
[6] D. A. Broniatowski, M. J. Paul, and M. Dredze, “National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic,” *PloS one*, vol. 8, no. 12, p. e83672, 2013.
[7] M. Santillana, E. O. Nsoesie, S. R. Mekaru, D. Scales, and J. S. Brownstein, “Using clinicians search query data to monitor influenza epidemics,” *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, vol. 59, no. 10, p. 1446, 2014.
[8] Q. Xu, Y. R. Gel, L. L. R. Ramirez, K. Nezafati, Q. Zhang, and K.-L. Tsui, “Forecasting influenza in hong kong with google search queries and statistical model fusion,” *PloS one*, vol. 12, no. 5, p. e0176690, 2017.
[9] H. Hu, H. Wang, F. Wang, D. Langley, A. Avram, and M. Liu, “Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network,” *Scientific reports*, vol. 8, no. 1, p. 4895, 2018.
[10] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *science*, vol. 313, no. 5786, pp. 504–507, 2006.
[11] B. Zou, V. Lampos, R. Gorton, and I. J. Cox, “On infectious intestinal disease surveillance using social media content,” in *Proceedings of the 6th International Conference on Digital Health Conference*. ACM, 2016, pp. 157–161.
[12] W. Huang, G. Song, H. Hong, and K. Xie, “Deep architecture for traffic flow prediction: Deep belief networks with multitask learning.” *IEEE Trans. Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.
[13] D. N. T. How, C. K. Loo, and K. S. M. Sahari, “Behavior recognition for humanoid robots using long short-term memory,” *International Journal of Advanced Robotic Systems*, vol. 13, no. 6, p. 1729881416663369, 2016.
[14] M. Shafie-Khah, M. P. Moghaddam, and M. Sheikh-El-Eslami, “Price forecasting of day-ahead electricity markets using a hybrid forecast method,” *Energy Conversion and Management*, vol. 52, no. 5, pp. 2165–2169, 2011.
[15] F. Murtagh, J.-L. Starck, and O. Renaud, “On neuro-wavelet modeling,” *Decision Support Systems*, vol. 37, no. 4, pp. 475–484, 2004.
[16] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[17] H. Palangi, L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, and R. Ward, “Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval,” *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 24, no. 4, pp. 694–707, 2016.
[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.