Zhu *et al. BMC Bioinformatics* 2019, **20**(Suppl 18):575
https://doi.org/10.1186/s12859-019-3131-8

BMC Bioinformatics

RESEARCH Open Access

Attention-based recurrent neural network for influenza epidemic prediction

Xianglei Zhu[1†], Bofeng Fu[1†], Yaodong Yang[1], Yu Ma[3], Jianye Hao[1*], Siqi Chen[1], Shuang Liu[1], Tiegang Li[3], Sen Liu[2], Weiming Guo[2] and Zhenyu Liao[4,5]

From Biological Ontologies and Knowledge bases workshop at IEEE BIBM 2018
Madrid, Spain. 3-6 December 2018

Abstract

Background: Influenza is an infectious respiratory disease that can cause serious public health hazard. Due to its huge threat to the society, precise real-time forecasting of influenza outbreaks is of great value to our public.

Results: In this paper, we propose a new deep neural network structure that forecasts a real-time influenza-like illness rate (ILI%) in Guangzhou, China. Long short-term memory (LSTM) neural networks is applied to precisely forecast accurateness due to the long-term attribute and diversity of influenza epidemic data. We devise a multi-channel LSTM neural network that can draw multiple information from different types of inputs. We also add attention mechanism to improve forecasting accuracy. By using this structure, we are able to deal with relationships between multiple inputs more appropriately. Our model fully consider the information in the data set, targetedly solving practical problems of the Guangzhou influenza epidemic forecasting.

Conclusion: We assess the performance of our model by comparing it with different neural network structures and other state-of-the-art methods. The experimental results indicate that our model has strong competitiveness and can provide effective real-time influenza epidemic forecasting.

Keywords: Influenza epidemic prediction, Attention mechanism, Multi-channel LSTM neural network

Background

Influenza is an infectious respiratory disease that can cause serious public health hazard. It can aggravate the original underlying disease after infection, causing secondary bacterial pneumonia and acute exacerbation of chronic heart and lung disease. Furthermore, the 2009 H1N1 pandemic caused between 151,700 and 575,400 deaths in worldwide during the first year the virus circulated [1]. Therefore, precise on-line monitoring and forecasting of influenza epidemic outbreaks has a great value to public health departments. Influenza detection and surveillance systems provide epidemiologic information that can help public health sectors develop preventive measures and assist local medical institutions in deployment planning [2].

Influenza-like-illness (ILI) is an infectious respiratory infection measurement defined by the World Health Organization (WHO). ILI with a measured fever higher than 38°C, and cough, with onset within the previous 10 days [3]. Our prediction target, ILI%, is equal to the ratio of the influenza-like cases number to the visiting patients' number. In the field of influenza surveillance, ILI% is often used as an indicator to help determine if there is a possible influenza epidemic. When the ILI% baseline is exceeded, the influenza season has arrived, reminding the health administrations to take timely preventive measures.

In recent years, more and more researchers have concentrated on precise on-line monitoring, early detection and influenza epidemic outbreaks forecasting. Thus, influenza epidemic outbreaks forecasting has become the most active research direction. The information from

*Correspondence: haojianye@gmail.com
†Xianglei Zhu and Bofeng Fu contributed equally to this work.
[1]College of Intelligence and Computing, Tianjin University, Peiyang Park Campus: No.135 Yaguan Road, Haihe Education Park, 300350 Tianjin, China
Full list of author information is available at the end of the article

website search or social network applications, such as Twitter and Google Correlate[4–6], provides sufficient data support for this research area. Previous methods are commonly built on linear models, such as least absolute shrinkage and selection operator (LASSO) or penalized regression[4, 6, 7]. Some people also implement deep learning models when solving influenza epidemic forecasting problems[8, 9]. However, these methods can't efficiently provide the precise forecasting of ILI% one week in advance. First, the online data is not accurate enough and lacks necessary features, which cannot fully reflect the trend of the influenza epidemic. Second, influenza epidemic data is usually very complex, non-stationary, and very noisy. Traditional linear models cannot handle multivariable inputs appropriately. Third, previously proposed deep learning methods didn't consider the time-sequence property of influenza epidemic data.

In this paper, we use influenza surveillance data as our data set, which is provided by the Guangzhou Center for Disease Control and Prevention. This data set includes multiple features and is count separately of each district in Guangzhou. Our approach takes advantage of these two characteristics. Meanwhile, we consider the time-sequence property, making our approach solve the influenza epidemic forecasting problem in Guangzhou with pertinence. Due to the relevant specifications of data collection, our method is also applicable in other regions.

We concentrate on implementing deep learning models to address the influenza outbreaks forecasting problem. Recently, deep learning methods have obtained remarkable performances in various research areas from computer vision, speech recognition to climate forecasting[10–12]. We implement long-short term memory (LSTM) neural networks[13] as a fundamental method for forecasting, because the influenza epidemic data naturally has time series attribute. Considering that different types of input data correspond to different characteristics, one single LSTM with a specific filter may not capture the time series information comprehensively. By using a multi-channel architecture, we can better capture the time series attributes from the data. Not only ensures the integration of various relevant descriptors in the high-level network, but also ensures that the input data will not interfere with each other in the underlying network. The structured LSTM can provide robust fitting ability that has been provided in several papers [14, 15]. We further enhance our method using attention mechanism. In attention layer, the probability of occurrence of each value in the output sequence depends on the values in the input sequence. By designing this architecture, we can better deal with input stream relationships among multiple regions more appropriately. We named our model as Att-MCLSTM, which stands for attention-based multi-channel LSTM.

Our main contributions can be summarized as follows: (1) We test our model on Guangzhou influenza surveillance data set, which is authentic and reliable. It contains multiple attributes and time series features. (2) We propose an attention-based multi-channel LSTM structure that associates different well-behaved approaches. The structure takes the forecasting problem and the influenza epidemic data attributes into account. The proposed model can be seen as an alternative to forecast influenza epidemic outbreaks in other districts. (3) The proposed model makes full use of information in the data set, solving the actual problem of influenza epidemic forecasting in Guangzhou with pertinence. The experimental results demonstrate the validity of our method. To the best of our knowledge, this is the first study that applies LSTM neural networks to the influenza outbreaks forecasting problem.

The rest of this paper is organized as follows. In the second section, we illustrate details of our method. In the third section, we evaluate performances of our method by comparing it with different neural network structures and other prior art methods. In the fourth section, we discuss conclusions and prospects for future works.

Methods

The accurateness of the forecasting problems can be enhanced by combining multiple models[16–26]. In this paper, we devise an novel LSTM neural network structure to settle the influenza epidemic forecasting problem in Guangzhou, China. Our model can extract characteristics more effectively from time series data, and take different impacts of different parts of data into consideration. In order to illustrate our model clearly, we illustrate our data set first. The following sections will give further illustrations on the data set, the overall idea of our model, details of LSTM neural networks, attention mechanism, attention-based multi-channel LSTM, data normalization, and evaluation method.

Data set description

The influenza surveillance data we used includes 9 years data. Statistics on influenza epidemic data in 9 regions are counted each year. The data set includes 6 modules, and each of these modules has multiple features. The data set has one record each week, and data for 52 weeks is counted each year.

Design of the proposed model

In Fig. 1, we demonstrate the flow diagram of our method. The integrated flow diagram has two parts, training part and test part. In the training part, first, we select 19 relevant features after data cleaning and normalization processes. We further illustrate the chosen modules and

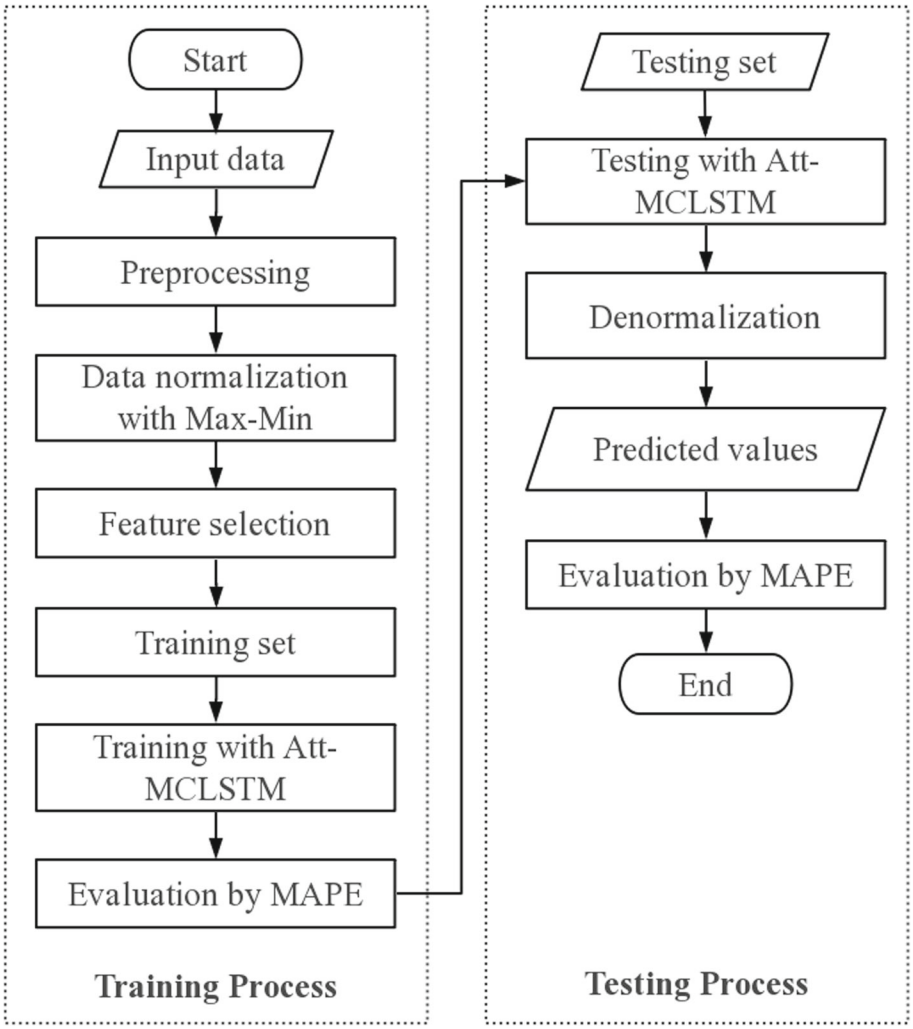

Fig. 1 The flowchart of Attention-based multi-channel LSTM

features in Table 1. Table 1 doesn't include basic information module, which includes time information, districts, and population. We use model-based ranking method as our feature selection method. In order to implement model-based ranking method, we delete one feature at a time, and input the rest of features into the same forecasting model every time. If the forecasting accuracy is low, this means that the feature we removed is relevant to our forecasting objective. After ranking all the forecasting accuracy, we select 19 features that are relevant to the forecasting objective. Then we separate the data set into training data set and test data set. The training data set contains 80 percent of data to extract annual trend and seasonal periodicity. In the test part, we test our model on the test data set. Then, we preform denormalization process to reconstruct the original values. Finally, we assess our model and compare it with other models.

Data normalization

Min-Max normalization is a linear transformation strategy[27]. This method maintains the relationship among all the original data. Min-Max normalization transforms a value x to y, y is defined as Eq. 1.

$$y = \frac{(x - min)}{(max - min)} \tag{1}$$

Where min is the smallest value in the data, max is the biggest value in the data. After data normalization, the features of data will be scaled between 0 and 1.

We preform denormalization process to reconstruct the original data. Given a normalized value y, its original value x is defined as Eq. 2.

$$x = (max - min)y + min \tag{2}$$

Table 1 Modules and features description for Section 2.1

Module name	Feature name	Description
Legal influenza cases report module	Legal influenza cases numbers	The number of influenza cases in the national infectious disease reporting system.
Epidemic monitoring module	Influenza outbreaks numbers	More than 10 influenza-like cases occurred within one week in the same unit.
	Affected cases numbers	The total number of people affected by the epidemic.
Symptom monitoring module	Influenza-like cases numbers (0-5 age)	The number of influenza-like cases (0-5 age).
	Influenza-like cases numbers (5-15 age)	The number of influenza-like cases (5-15 age).
	Influenza-like cases numbers (15-25 age)	The number of influenza-like cases (15-25 age).
	Influenza-like cases numbers (25-60 age)	The number of influenza-like cases (25-60 age).
	Influenza-like cases numbers (> 60 age)	The number of influenza-like cases (over 60 age).
	Total influenza-like cases numbers	The total number of influenza-like cases.
	Total visiting patients numbers	The total number of visiting patients.
	Upper respiratory tract infections numbers	The number of upper respiratory tract infections.
Pharmacy monitoring module	Chinese patent cold medicines	Sales of Chinese patent cold medicines.
	Other cold medicines	Sales of other cold medicines.
Climate data module	Average temperature (°C)	Average temperature.
	Maximum temperature (°C)	Maximum temperature.
	Minimum temperature (°C)	Minimum temperature.
	Rainfall (mm)	Rainfall.
	Air pressure (hPa)	Air pressure.
	Relative humidity (%)	Relative humidity.

Long-short term memory neural network

Recurrent neural networks have the ability to dynamically combine experiences because of their internal recurrence[28]. Different from other traditional RNNs, LSTM can deal with the gradient vanishing problem[29]. The memory units of LSTM cells retain time series attributes of given context[29]. Some researches have proven that LSTM neural networks can yield a better performance compared with other traditional RNNs when dealing with long-term time series data[30].

The structure of a single LSTM cell illustrate in Fig. 2. The gates control the flow of information, that is, interactions between different cells and cell itself. Input gate controls the memory state updating process. Output gate controls whether the output flow can alter other cells' memory state. Forget gate can choose to remember or forget its previous state. LSTM is implemented by following composite functions:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (4)$$

$$c_t = f_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$

$$h_t = o_t \tanh(c_t) \quad (7)$$

Where σ represent the logistic sigmoid function. i, f, o, and c represent the input gate, forget gate, output gate,

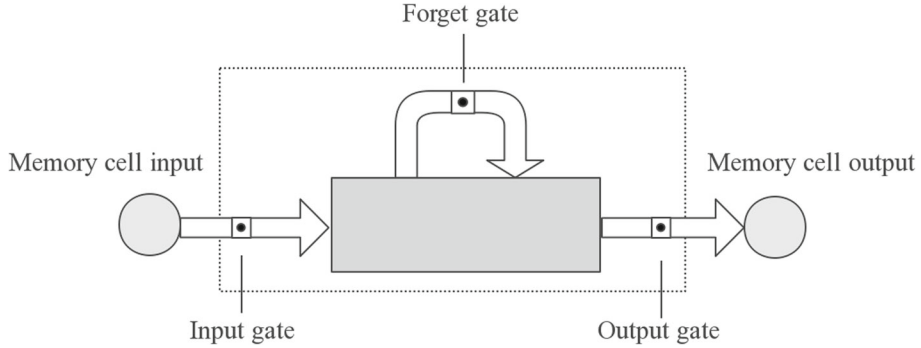

Fig. 2 The structure of single LSTM cell

cell input activation vectors respectively. h represents the hidden vector. The weight matrix subscripts have the intuitive meaning. Like, W_{hi} represents the hidden-input gate matrix etc.

Attention mechanism

Traditional Encode-Decode structures typically encode an input sequence into a fixed-length vector representation. However, this model has drawbacks. When the input sequence is very long, it is difficult to learn a feasible vector representation.

One fundamental theory of attention mechanism[31] is to abandon the conventional Encoder-Decoder structure. Attention mechanism trains a model that selectively learns the input streams by conserving the intermediate outputs of LSTM. In attention structure, the output sequences are affiliated with the input sequences. In other words, the probability of occurrence of each value in the output sequence depends on the value in the input sequence. Figure 3 illustrates the attention mechanism.

Attention layer calculates the weighted distribution of $X_1, \ldots, X_T$. The input of S_t contains the output of the attention layer. The probability of occurrence of the output sequence $\ldots, y_{t-1}, y_t, \ldots$ depends on input sequence $X_1, X_2, \ldots, X_T$. h_i represents the hidden vector. $A_{t,i}$ represents the weight of i^{th} input at time step t. 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:

$$m_i = \tanh(W_{cm}c + W_{ym}y_i) \tag{8}$$

$$s_i \propto \exp(\langle w_m, m_i \rangle) \tag{9}$$

$$\sum_i s_i = 1 \tag{10}$$

$$z = \sum_i s_i y_i \tag{11}$$

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

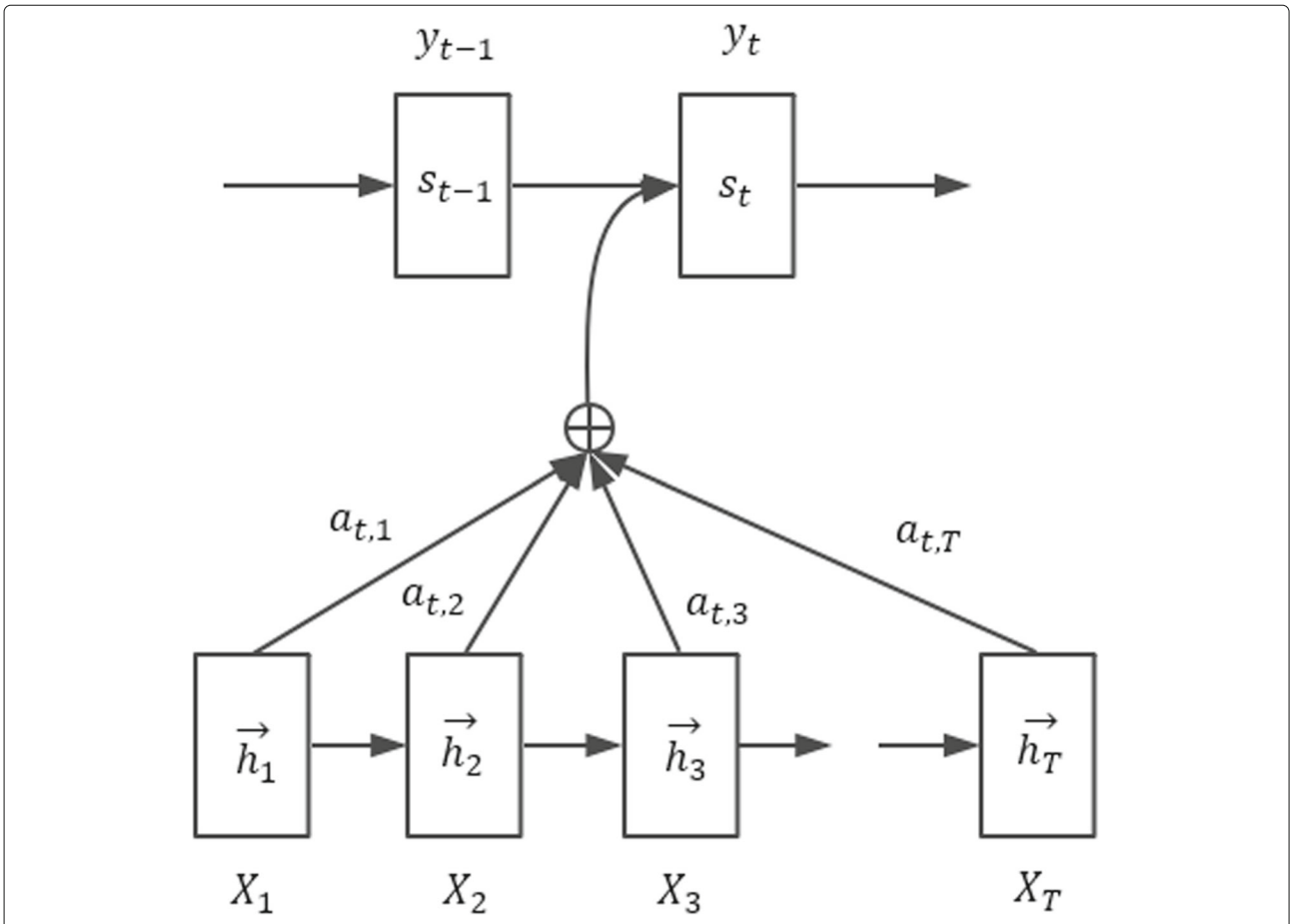

Fig. 3 The diagram of attention mechanism. Attention layer calculates the weighted distribution of $X_1, \ldots, X_T$. The input of S_t contains the output of the attention layer. The probability of occurrence of the output sequence $\ldots, y_{t-1}, y_t, \ldots$ depends on input sequence $X_1, X_2, \ldots, X_T$. h_i represents the hidden vector. $A_{t,i}$ represents the weight of i^{th} input at time step t

the weighted arithmetic mean of all y_i, W represents the relevance for each variable according to the context c.

Attention-based multi-channel LSTM

In Fig. 4, we illustrate the overall architecture of our model. We separate our data set into two categories. First, we classify average temperature, maximum temperature, minimum temperature, rainfall, air pressure and relative humidity together as climate-related data category. Then, the rest of features are classified together as influenza-related data category. In our data set, each region has its own influenza-related data, and they share the same climate-related data every week.

Because our data set has the above characteristics, the inputs of Att-MCLSTM contains two parts. First, the influenza-related data is input into a series of LSTM neural networks (LSTM 1, ..., LSTM 9) to capture correlative features. Second, the climate-related data is input into a single LSTM neural network (LSTM 10) to capture the long-term time series attribute of influenza epidemic data. For the first part, each LSTM neural network acquires the influenza-related data from one distinct region. In order to make full use of the complementarity among every regions, the outputs of LSTM neural networks (LSTM 1, ..., LSTM 9) are concatenated in a higher layer (Merge 1). This higher layer can obtain the fused descriptors of underlying neural networks. After we capture the features of every regions, we still want to weight intermediate sequences. The reason is that the data of each region has different influences on the final forecasting result. Therefore, the intermediate sequences pass through an attention layer (Attention) and a fully connected layer (Dense 1) in turn. Thereafter, we concatenate the outputs of these two parts together (Merge 2). Finally, the intermediate sequences are passed through two fully connected layers (Dense 2, Dense 3). So far, we acquire the high-level features of the input data, and they are used to solve the influenza epidemic forecasting.

By designing a multi-channel structure, we can better extract the time-sequence property of each type of data. Not only ensures the integration of various relevant descriptors in the high-level network, but also ensures that input data will not interfere with each other in the underlying network. In the attention layer, the probability of occurrence of each value in the output sequence depends on the value in the input sequence. This structure allows us to handle the relationship of input data between different districts more appropriately.

Evaluation method

To evaluate our method, we use the mean absolute percentage error (MAPE) as the criteria standard. Its formula is express as Eq. 12.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - x_i}{y_i}\right| \times 100 \tag{12}$$

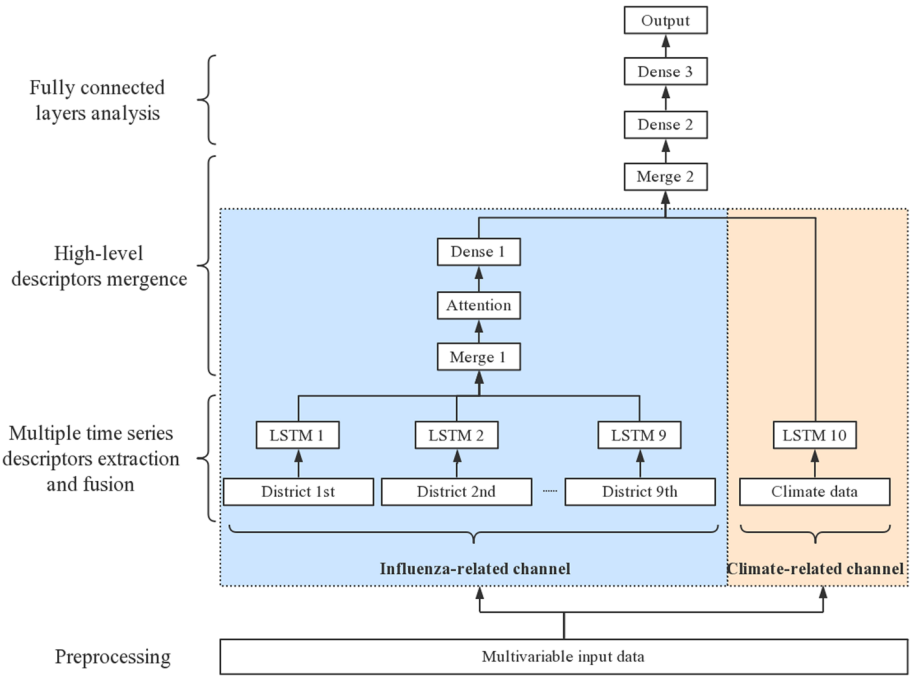

Fig. 4 The structure of Attention-based multi-channel LSTM

Where y_i denotes the i^{th} actual value, and x_i denotes the i^{th} predicted value. If the value of MAPE is low, the accuracy of the method is high.

Experiments

In this section, we did two experiments to verify the Att-MCLSTM model. In the first experiment, we evaluate the numbers of consecutive weeks of data that we need to forecast ILI% for the next week. In the second experiment, we compare our model with different neural network structures and other methods. Each experiment result is the average of 10 repeated trials.

Selection of consecutive weeks

In this experiment, we set the numbers of consecutive weeks as 6, 8, 10, 12, 14 respectively. The hyper-parameters of each layer are listed in Table 2. The activation functions we used are linear activation function. The loss function and optimizer are *mape* and *adam* respectively.

We use the first 370 consecutive weeks' data in training phase and the remaining data in the test phase. Each data sample includes 6 features in climate-related data category and 9 different districts' influenza-related data. Each influenza-related data contains 13 features. The climate-related data and each district's influenza-related data are input into the climate-related channel and the influenza-related channel respectively. The forecasting results are shown in Table 3.

Performance validation

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

- First, we compare Att-MCLSTM with MCLSTM by comparing their forecasting accuracy. The purpose of doing this is to verify the effect of the attention mechanism. For both models, we use the same multi-channel architecture (as shown in Fig.4). The only difference between these two models is that we delete the attention layer in MCLSTM. The parameters settings and data inputs method are as described in the first experiment.

Table 2 The size of every unit in Att-MCLSTM neural network for Section 3.1

Layer name	Units number
LSTM 1, …, LSTM 9	32
LSTM 10	32
Dense 1	16
Dense 2	10
Dense 3	1

Table 3 The MAPE of the prediction results for Section 3.1

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

- Second, we compare MCLSTM with LSTM by comparing their forecasting accuracy. The purpose of doing this is to verify the effect of the multi-channel structure. For MCLSTM, parameters settings and data inputs method are as described in the first experiment. For LSTM, we input entire features into one LSTM layer to capture the fused descriptors. Instead of separating data set according to different regions, we sum corresponding influenza-related features in each week from every regions together. Therefore, each data record includes 19 selected features. The data that contains these 19 features are passed through a fully connected layer to acquire high-level features. The units' number of LSTM layer and fully connected layer are 32 and 1 respectively.
- Third, we demonstrate that LSTMs can yield better performance than RNNs when dealing with time series data.

Results

(1) As can be seen from Table 3, 10 consecutive weeks' data yields the best performance. (2) Table 4 shows that Att-MCLSTM has strong competitiveness and can provide effective real-time influenza epidemic forecasting.

Discussion

The results of the first experiment indicate that 10 consecutive weeks data can appropriately reflect the time series attribute of influenza data. If the length of input data is shorter than 10, the input data doesn't contain enough time series information. On the contrary, if the length of input data is longer than 10, the noise inside the input data increased, leading to a decrease in forecasting accuracy. Therefore, in our experiments, each data record includes 10 consecutive weeks' data.

Table 4 The MAPE of the prediction results for Section 3.2

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

The results of the second experiment show that Att-MCLSTM can yield the best performance. In Table 4, from the first two rows, we can conclude that using attention mechanism can improve the MAPE from 0.105 to 0.086. The reason is that the attention layer can better deal with the relationships of input streams among every regions more appropriately. From the second row and the third row, we can conclude that using multi-channel structure can improve the MAPE from 0.118 to 0.105. The reason is that the multi-channel structure can better capture the time series attributes from different input streams. From the last two rows, we can conclude that using LSTM can improve the MAPE from 0.132 to 0.118. The reason is that LSTM neural network can better deal with time series data. This result also demonstrates the time series attribute of influenza epidemic data.

Figure 5 shows the actual values and predicted values of four models. We can see that the result of Att-MCLSTM is close to the actual output. There are more obvious differences between the predicted results and the actual value by using the other three models. So, this can verify that adopting Att-MCLSTM to analyze the sequential information can help to extract time-sequence characteristic more accurately and comprehensively.

Conclusion and future work

In this paper, we propose a new deep neural network structure (Att-MCLSTM) to forecast the ILI% in Guangzhou, China. First, we implement the multi-channel architecture to capture time series attributes from different input streams. Then, the attention mechanism is applied to weight the fused feature sequences, which allows us to deal with relationships between different input streams more appropriately. Our model fully consider the information in the data set, targetedly solving the practical problem of influenza epidemic forecasting in Guangzhou. We assess the performance of our model by comparing it with different neural network structures and other state-of-the-art models. The experimental results indicate that our model has strong competitiveness and can provide effective real-time influenza epidemic forecasting. To the best of our knowledge, this is the first study that applies LSTM neural networks to the influenza outbreaks forecasting. Continuing work will further improve the expansion ability of our model by introducing transfer learning.

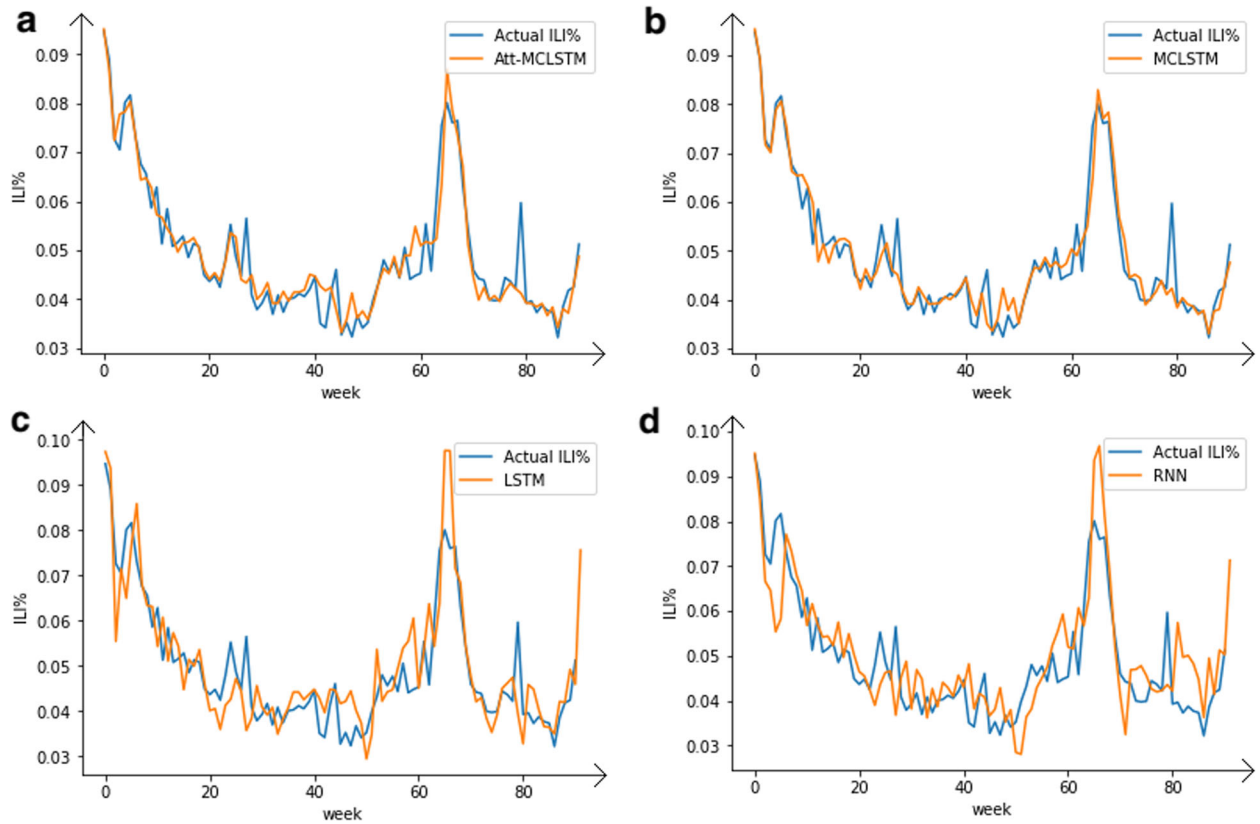

Fig. 5 The results of one-week ahead prediction by using four individual models. **a** shows the comparison of Att-MCLSTM and real data; **b** shows the comparison of MCLSTM and real data; **c** shows the comparison of LSTM and real data; **d** shows the comparison of traditional RNN and real data. In each figure, the blue line denotes the actual values, and the orange line denotes the predicted values

Abbreviations
ILI: Influenza-like illness; LSTM: Long short-term memory; LASSO: Least absolute shrinkage and selection operator; MAPE: Mean absolute percentage error

Acknowledgements
We thank the reviewers' valuable comments for improving the quality of this work.

About this supplement
This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 18, 2019: Selected articles from the Biological Ontologies and Knowledge bases workshop 2018*. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-18.

Authors' contributions
XZ and BF contributed equally to the algorithm design and theoretical analysis. YY, YM, JH, SC, SL, TL, SL, WG, and ZL contributed equally to the the quality control and document reviewing. All authors read and approved the final manuscript.

Funding
Publication costs are funded by The National Natural Science Foundation of China (Grant Nos.: U1836214), Tianjin Development Program for Innovation and Entrepreneurship and Special Program of Artificial Intelligence of Tianjin Municipal Science and Technology Commission (NO.: 17ZXRGGX00150).

Availability of data and materials
All data information or analyzed during this study are included in this article.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Intelligence and Computing, Tianjin University, Peiyang Park Campus: No.135 Yaguan Road, Haihe Education Park, 300350 Tianjin, China. [2]Automotive Data Center, China Automotive Technology & Research, 300300 Tianjin, China. [3]Guangzhou Center for Disease Control and Prevention, 510440 Guangzhou, China. [4]Pony Testing International Group, 300051 Tianjin, China. [5]Tianjin FoodSafety Inspection Technology Institute, 300300 Tianjin, China.

Published: 25 November 2019

References
1. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using google search data via argo. Proc Natl Acad Sci. 2015;112(47):14473–8.
2. Brownstein JS, Mandl KD. Reengineering real time outbreak detection systems for influenza epidemic monitoring. In: AMIA Annual Symposium Proceedings, vol. 2006. American Medical Informatics Association; 2006. p. 866.
3. Organization WH, et al. Who interim global epidemiological surveillance standards for influenza. 20121–61.
4. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) google flu trends? Am J Prev Med. 2014;47(3):341–7.
5. Achrekar H, Gandhe A, Lazarus R, Yu S.-H., Liu B. Predicting flu trends using twitter data. In: Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference On. IEEE; 2011. p. 702–7. https://doi.org/10.1109/infcomw.2011.5928903.
6. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. PLoS ONE. 2013;8(12):83672.
7. Santillana M, Nsoesie EO, Mekaru SR, Scales D, Brownstein JS. Using clinicians' search query data to monitor influenza epidemics. Clin Infect Dis Off Publ Infect Dis Soc Am. 2014;59(10):1446.
8. Xu Q, Gel YR, Ramirez LLR, Nezafati K, Zhang Q, Tsui K.-L. Forecasting influenza in hong kong with google search queries and statistical model fusion. PLoS ONE. 2017;12(5):0176690.
9. Hu H, Wang H, Wang F, Langley D, Avram A, Liu M. Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network. Sci Rep. 2018;8(1):4895.
10. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. science. 2006;313(5786):504–7.
11. Zou B, Lampos V, Gorton R, Cox IJ. On infectious intestinal disease surveillance using social media content. In: Proceedings of the 6th International Conference on Digital Health Conference. ACM; 2016. p. 157–61. https://doi.org/10.1145/2896338.2896372.
12. Huang W, Song G, Hong H, Xie K. Deep architecture for traffic flow prediction: Deep belief networks with multitask learning. IEEE Trans Intell Transp Syst. 2014;15(5):2191–201.
13. How DNT, Loo CK, Sahari KSM. Behavior recognition for humanoid robots using long short-term memory. Int J Adv Robot Syst. 2016;13(6):1729881416663369.
14. Yang Y, Hao J, Sun M, Wang Z, Fan C, Strbac G. Recurrent deep multiagent q-learning for autonomous brokers in smart grid. In: IJCAI, vol. 18; 2018. p. 569–75. https://doi.org/10.24963/ijcai.2018/79.
15. Yang Y, Hao J, Wang Z, Sun M, Strbac G. Recurrent deep multiagent q-learning for autonomous agents in future smart grid. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. International Foundation for Autonomous Agents and Multiagent Systems; 2018. p. 2136–8. https://doi.org/10.24963/ijcai.2018/79.
16. Shafie-Khah M, Moghaddam MP, Sheikh-El-Eslami M. Price forecasting of day-ahead electricity markets using a hybrid forecast method. Energy Convers Manag. 2011;52(5):2165–9.
17. Xiaotian H, Weixun W, Jianye H, Yaodong Y. Independent generative adversarial self-imitation learning in cooperative multiagent systems. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems; 2019. p. 1315–1323. International Foundation for Autonomous Agents and Multiagent Systems.
18. Yaodong Y, Jianye H, Yan Z, Xiaotian H, Bofeng F. Large-scale home energy management using entropy-based collective multiagent reinforcement learning framework. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems; 2019. https://doi.org/10.24963/ijcai.2019/89.
19. Hongyao T, Jianye H, Tangjie Lv, Yingfeng C, Zongzhang Z, Hangtian J, Chunxu R, Yan Z, Changjie F, Li W. Hierarchical deep multiagent reinforcement learning with Temporal Abstraction. In: arXiv preprint arXiv:1809.09332; 2018.
20. Peng J, Guan J, Shang X. Predicting parkinson's disease genes based on node2vec and autoencoder. Front Genet. 2019;10:226.
21. Peng J, Zhu L, Wang Y, Chen J. Mining relationships among multiple entities in biological networks. IEEE/ACM Trans Comput Biol Bioinforma. 2019. https://doi.org/10.1109/tcbb.2019.2904965.
22. Peng J, Xue H, Shao Y, Shang X, Wang Y, Chen J. A novel method to measure the semantic similarity of hpo terms. IJDMB. 2017;17(2):173–88.
23. Cheng L, Hu Y, Sun J, Zhou M, Jiang Q. Dincrna: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncrna function. Bioinformatics. 2018;34(11):1953–6.
24. Cheng L, Wang P, Tian R, Wang S, Guo Q, Luo M, Zhou W, Liu G, Jiang H, Jiang Q. Lncrna2target v2. 0: a comprehensive database for target genes of lncrnas in human and mouse. Nucleic Acids Res. 2018;47(D1):140–4.
25. Hu Y, Zhao T, Zang T, Zhang Y, Cheng L. Identification of alzheimer's disease-related genes based on data integration method. Front Genet. 2018;9:. https://doi.org/10.3389/fgene.2018.00703.
26. Peng J, Hui W, Li Q, Chen B, Jiang Q, Wei Z, Shang X. A learning-based framework for mirna-disease association prediction using neural networks. bioRxiv. 2018276048. https://doi.org/10.1101/276048.
27. Panda SK, Jana PK. Efficient task scheduling algorithms for heterogeneous multi-cloud environment. J Supercomput. 2015;71(4):1505–33.
28. Murtagh F, Starck J-L, Renaud O. On neuro-wavelet modeling. Dec Support Syst. 2004;37(4):475–84.
29. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.

30. Palangi H, Deng L, Shen Y, Gao J, He X, Chen J, Song X, Ward R. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. IEEE/ACM Trans Audio Speech Lang Process (TASLP). 2016;24(4):694–707.
31. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Advances in Neural Information Processing Systems; 2017. p. 5998–6008.

Publisher's Note