Inter- and Intra-Series Embeddings Fusion Network for Epidemiological Forecasting

Feng Xie
National University of Defense Technology
Changsha, China
xiefeng@nudt.edu.cn

Zhong Zhang
National University of Defense Technology
Changsha, China
zhangzhong@nudt.edu.cn

Xuechen Zhao
National University of Defense Technology
Changsha, China
zhaoxuechen@nudt.edu.cn

Bin Zhou
* National University of Defense Technology
Changsha, China
binzhou@nudt.edu.cn

Yusong Tan
National University of Defense Technology
Changsha, China
ystan@nudt.edu.cn

ABSTRACT
The accurate forecasting of infectious epidemic diseases is the key to effective control of the epidemic situation in a region. Most existing methods ignore potential dynamic dependencies between regions or the importance of temporal dependencies and inter-dependencies between regions for prediction. In this paper, we propose an Inter- and Intra-Series Embeddings Fusion Network (SEFNet) to improve epidemic prediction performance. SEFNet consists of two parallel modules, named Inter-Series Embedding Module and Intra-Series Embedding Module. In Inter-Series Embedding Module, a multi-scale unified convolution component called Region-Aware Convolution is proposed, which cooperates with self-attention to capture dynamic dependencies between time series obtained from multiple regions. The Intra-Series Embedding Module uses Long Short-Term Memory to capture temporal relationships within each time series. Subsequently, we learn the influence degree of two embeddings and fuse them with the parametric-matrix fusion method. To further improve the robustness, SEFNet also integrates a traditional autoregressive component in parallel with nonlinear neural networks. Experiments on four real-world epidemic-related datasets show SEFNet is effective and outperforms state-of-the-art baselines.

KEYWORDS
deep learning; epidemiological forecasting; time series

1 INTRODUCTION
The outbreak of an epidemic will bring huge disasters to a region and even a country. The World Health Organization (WHO) estimates that influenza annually causes approximately 3-5 million severe cases and 290,000-650,000 deaths.¹ In recent years, the COVID-19 pandemic has spread to more than 200 countries and territories around the world,² and the number of infections and deaths in almost all affected countries is increasing at an alarming rate. Accurately forecasting epidemics plays an essential role in allocating healthcare resources and promoting administrative planning.

¹https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)
²https://covid19.who.int/

Corresponding Author: binzhou@nudt.edu.cn
DOI reference number: 10.18293/SEKE2022-109

Figure 1: The blue boxes indicate the temporal dependency between time points, while the green boxes indicate the inter-dependency between regions.

- **Inter-series embedding modeling.** There are dynamic dependencies among regions/time series. When epidemics spread in different geographic regions, it is highly likely that similar progression patterns are shared among multiple regions owing to various factors [9] (e.g., similar geographic topology or climate), and these similar patterns can aid in prediction.
- **Intra-series embedding modeling.** There are temporal dependencies within a region/time series (e.g., seasonal influenza). The epidemic development trend of one region also can be distinguished from others, this is due to region-specific factors such as government intervention, healthcare quality, climate, etc.

To date, various methods have been proposed for epidemic forecasting, but they suffer from some limitations that are bad for performance. First, using vanilla Recurrent Neural Networks (RNNs) [2,
10] or single-scale Convolution Neural Networks (CNNs) [9, 20] is hard to capture multi-scale and complex patterns, thus resulting in a certain degree of distortion, making it difficult to extract dynamic dependencies between time series. Second, some methods are dedicated to capturing dependencies between regions by introducing the “Attention Mechanism” [2, 9, 10], but these dependencies may misguide the final prediction because the progression patterns or data distribution of different regions is not fully consistent. Therefore, we believe both inter-series dependencies and intra-series dependencies jointly contribute to epidemic forecasting, and their influence degree on prediction results varies by region.

To tackle these challenges, we propose a novel deep learning model called Inter- and Intra-Series Embeddings Fusion Network (SEFNet) that extract inter- and intra-series embeddings through two parallel modules respectively and fuse them using parametric-matrix fusion [22]. To further improve the robustness, we also integrate autoregressive component parallel to the model. Our contributions are summarized below:

- We propose a new model that extracts inter-series correlations and intra-series temporal dependencies through two separate neural networks and uses parametric-matrix fusion to emphasize the importance of each information for epidemic prediction.
- We propose a multi-scale unified convolution component called Region-Aware Convolution that is capable of extracting local, periodic, and global patterns to better obtain feature representation and capture potential dependencies between regions.
- We conduct extensive experiments on four real-world epidemic-related datasets. The results show that our model achieves better performance than other state-of-the-art methods and demonstrates the effectiveness of each component.

2 RELATED WORK

There has been a large body of work focusing on epidemic forecasting in literature, including statistical models [8, 15, 18], compartment models [4, 19], and swarm intelligence models [3]. In recent years, deep learning models have shown excellent performance in various prediction tasks due to their powerful training and data-driven capabilities. CNNRNN-Res [20] is the first to apply deep learning for epidemiological prediction. Deng et al. [2] proposed Cola-GNN that treats regions as nodes in a graph and applies Graph Neural Networks (GNNs) to capture dependencies among regions. Jung et al. [10] proposed SAIFlu-Net that combines Long Short-Term Memory and self-attention to capture inter-dependencies between regions. Jin et al. [9] developed ACTs based on inter-series attention for COVID-19 forecasting. Cui et al. [1] designed a multi-range encoder-decoder framework for COVID-19 prediction. Nowadays, improving epidemic prediction is an open research problem to help the world mitigate the crisis that threatens public health.

3 THE PROPOSED METHOD

3.1 Problem Formulation

We formulate the epidemic prediction problem as a time series forecasting task. We have a total of $N$ regions, and each region is associated with a time series input for a window $T$, where $T$ is the length of historical observation data. Furthermore, we denote the epidemiology profiles $X = \{x_{i, T+1},..., x_i\} \in R^{N \times T}$ at time point $t$. An instance for region $i$ is represented by $x_i = \{x_{i,t-T+1},...,x_{i,t}\} \in R^T$. The goal of this task is to predict the epidemiology profile of the future time point $t + h$, where $h$ is the horizon also called lead time. The proposed model SEFNet is shown in Figure 2. In the following sections, we introduce the building blocks of SEFNet in detail.

3.2 Intra-Series Embedding Module

The first module is Intra-Series Embedding module, which uses the historical information of time series to focus on the autocorrelation also called temporal dependencies of a single time series. In this work, we apply the Long Short-Term Memory (LSTM) [5] to capture temporal sequential dependency. The Recurrent Neural Networks (RNNs) are shown effective in sequence modeling and LSTM is a variant of RNNs, which can solve the vanishing gradients and exploding gradients problems in traditional RNNs [15]. Let $D$ be
the dimension of the hidden state of LSTM, we use the original version of LSTM and formulate it as:

$$h_{i,t} = \text{LSTM}(x_{i,t}, h_{i,t-1}),$$  \hspace{1cm} (1)

where $h_{i,t}$ is the output representation of region $i$ at time point $t$. For each region, we use the last output of LSTM as region’s intra-series embedding $h_{i}^{\text{intra}} \in \mathbb{R}^{D}$.

3.3 Inter-Series Embedding Module

The second module is Inter-Series Embedding module which focuses on dependencies between time series. First, we obtain temporal patterns through the proposed Region-Aware Convolution (RAConv), which is a multi-scale unified convolution component. Next, we feed the output of RAConv into an attention layer to generate embeddings of dynamic dependencies between regions.

Figure 3 shows different temporal patterns of influenza case trends in different Health and Human Services (HHS) regions in the United States. Inspired by the Inception [16] in computer vision, we propose a multi-scale unified component called Region-Aware Convolution (RAConv) that can extract local, periodic, and global patterns simultaneously. The structure of RAConv is shown in Figure 4. RAConv consists of three branches that apply convolution blocks with different scales or different types, thus is capable of capturing multi-scale and more complex feature patterns. Each convolution block has $K$ filters. The local pattern branch applies standard convolution with some small kernel sizes to extract local patterns in the time series through local mapping. The periodic pattern branch inspired by skip-RNN in LSTMNet [11] applies dilated convolution that enables a large receptive field via dilation factor to capture the periodic pattern. Formally, the dilated convolution is a standard convolution applied to input with defined gaps. The global pattern branch applies standard convolution with the same size as $T$ to extract time-invariant patterns of all time steps for regions [6] (e.g., time series uptrend in Figure 3c). We denote convolution filter in RAConv as $f_{j \times k \times d}$ where $s$ is kernel size and $d$ is dilated factor. We empirically choose the kernel size $s$ to $\{3,5,7\}$, and dilated factor $d$ to $\{1,2\}$. We can get the local, periodic, and global features of region $i$ by following equations:

$$h_{i}^{l} = \text{Pool}(\text{BN}(x_{i} \ast f_{1 \times 3,1})); \text{Pool}(\text{BN}(x_{i} \ast f_{1 \times 5,1}))),$$  \hspace{1cm} (2)

$$h_{i}^{p} = \text{Pool}(\text{BN}(x_{i} \ast f_{1 \times 3,2})); \text{Pool}(\text{BN}(x_{i} \ast f_{1 \times 5,2}))),$$  \hspace{1cm} (3)

where $\ast$ is convolution operator, and $\{\}$ is concatenation operation. $\text{Pool}(\cdot)$ is the Adaptive Max Pooling layer that can not only capture the most representable features, but also effectively reduce the amount of parameters. Adaptive Max Pooling is able to control the output size same as parameters $P$. $\text{BN}(\cdot)$ is the Batch Normalization [7] layer that normalize the data and speed up convergence.

The convolution operation of $x$ with $f$ at step $j$ is represented as:

$$x \ast f_{1 \times s,d}(j) = \sum_{i=1}^{s} f_{1 \times k}(i)x(j - d \times i).$$  \hspace{1cm} (5)

Next, we concatenate three patterns and apply an element-wise activation function (e.g., hyperbolic tangent):

$$h_{i}^{\text{dev}} = \text{tanh}(\{h_{i}^{l}; h_{i}^{p}; h_{i}^{g}\}).$$  \hspace{1cm} (6)

For each time series, we execute the above process and get the intermediate matrix called $H^{\text{dev}} \in \mathbb{R}^{N \times K \times K}$.

Due to the powerful feature extraction capability of the self-attention network, we apply a typical self-attention network inspired by the Transformer [17] to capture the dependencies among regions. Let $A$ be the dimension of inter-series embedding. We can calculate attention distribution $A$ and inter-series embedding matrix $H^{\text{inter}} \in \mathbb{R}^{N \times A}$ by following equations:

$$A = \text{softmax}((H^{\text{dev}} \mathbf{W}^{Q})(H^{\text{dev}} \mathbf{W}^{K})),$$ \hspace{1cm} (7)

$$H^{\text{inter}} = A H^{\text{dev}} \mathbf{W}^{V},$$ \hspace{1cm} (8)

where $\mathbf{W}^{Q}, \mathbf{W}^{K},$ and $\mathbf{W}^{V}$ are the weight matrices that linearly map the $H^{\text{dev}}$ to query, key, and value matrices.

3.4 Fusion

Directly concatenating or summing inter-series embedding and intra-series embedding will have the following problems: (1) **Inconsistent scale.** Since two feature embeddings come from different neural network modules, the structural differences of each module (e.g., activation function) will lead to inconsistent scales of feature embeddings; (2) **Different importance.** Two feature embeddings describe different feature information of time series so the...
importance of two feature embeddings is very different in the process of epidemiology forecasting (e.g., temporal dependency is more significant for a region with periodic recurrence of an epidemic, although there may be similar development patterns to others). Therefore, to address these problems, we adopt parametric-matrix fusion [22] to adaptively control the flow of inter-series embedding and intra-series embedding and fuse them together:

\[ H^{ fus} = W^{ inter} \circ H^{ inter} ; W^{ intra} \circ H^{ intra} , \]

where \( H^{ fus} \in R^{N \times (D+A)} \) is the output of fusion operation. \( \circ \) is element-wise multiplication. \( W^{ inter} \) and \( W^{ intra} \) are the learnable parameters that adjust the degrees affected by inter-series embedding and intra-series embedding respectively.

3.5 Prediction

Due to the nonlinear characteristics of Convolutional, Recurrent and self-attention components, the scale of neural network output is not sensitive to input. Meanwhile, the historical infection cases of each region are not purely nonlinear, which cannot be fully handled well by neural networks. To address these drawbacks, we retain the advantages of traditional linear models and neural networks by combining a linear part to design a more accurate and robust prediction framework inspired by [11, 14]. Specifically, we adopt the classical AutoRegressive (AR) model as the linear component in a parallel manner. Denote the forecasting result of AR component as \( \hat{y}^l \) and the forecasting result of the final prediction as \( \hat{y} \).

\[ \hat{y}^l = \hat{y}^l + \hat{y}^l = \sum_{m=0}^{q-1} W^{ar,m} x_{t-m} + b^{ar} . \]

where \( W^{ar} \) is the weight matrix and \( b^{ar} \) is the bias. \( q \) is the look-back window of AR that need be less than or equal to input window size \( T \). Then, we feed the output after fusion operation to a dense layer to get the nonlinear part of the final prediction:

\[ \hat{y}_t^l = H^{ fus} W^a + b^a . \]

The final prediction of model is then obtained by summing the nonlinear part and the linear part got by AR component:

\[ \hat{y}_{t+h} = \hat{y}_{t+h}^l + \hat{y}_{t+h}^l . \]

In the training process, we adopt the Mean Square Error as the loss function that defined as:

\[ \min _{\theta} ||y_{t+h} - \hat{y}_{t+h}||^2 , \]

where \( y_{t+h} \) is the normalized values into the real range. PCC is a measure of the linear dependence between time series. For RMSE lower value is better, while for PCC higher value is better.

4.2 Methods for Comparison

We compared the proposed model with the following methods.

- **AR**: The most classic statistical methods in time series analysis.
- **LRidge**: The vector autoregression (VAR) with L2-regularization.
- **LSTNet [11]**: A deep learning model that combines CNN and RNN to extract short- and long-term patterns.
- **TPA-LSTM [14]**: An attention based LSTM network that employs CNN for pattern representations.
- **CNNRNN-Res [20]**: A deep learning model that combines CNN, RNN, and residual links for epidemiological prediction.
- **SAIFlu-Net [10]**: A self-attention based deep learning model for regional influenza prediction.
- **Cola-GNN [2]**: A deep learning model that combines CNN, RNN and GNN for epidemiological forecasting.

4.3 Experimental Details

All programs are implemented using Python 3.8.5 and PyTorch 1.9.1 with CUDA 11.1 (1.9.1 cu111) in an Ubuntu server with an Nvidia Tesla K80 GPU. Our source codes are publicly available.\(^6\)

**Experimental setting.** All datasets have been split into training set(50%), validation set(20%) and test set(30%). The batch size is set to 128. We use Min-Max normalization to convert data to [0,1] scale and after prediction, we denormalize the prediction value and use it

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \]

\[ \text{PCC} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2}} \]

Table 1: Dataset statistics: min, max, mean, and standard deviation (SD) of patient counts; dataset size means the number of regions multiplied by the number of samples.

| Datasets         | Size    | Min   | Max   | Mean   | SD    |
|------------------|---------|-------|-------|--------|-------|
| Japan-Prefectures| 47×348  | 0     | 26635 | 655    | 1711  |
| US-Regions       | 10×785  | 0     | 16526 | 1009   | 1351  |
| US-States        | 49×360  | 0     | 9716  | 223    | 428   |
| Canada-Covid     | 13×717  | 0     | 127199| 3082   | 8473  |

\(^1\)https://tinyurl.com/y5d7t8m
\(^2\)https://tinyurl.com/y390q3h
\(^3\)https://github.com/CSSEGISandData/COVID-19
\(^4\)https://github.com/Xiefeng69/SEFNet

4.4 EXPERIMENTS

4.1 Datasets and Metrics

We prepare four real-world epidemic-related datasets as follows, and their data statistics are shown in Table 1.

- **Japan-Prefectures.** This dataset is collected from the Infectious Diseases Weekly Report (IDWR) in Japan,\(^3\) which contains weekly influenza-like-illness statistics from 47 prefectures, ranging from August 2012 to March 2019.
- **US-Regions.** This dataset is the ILINet portion of the US-HHS dataset,\(^4\) consisting of weekly influenza activity levels for 10 HHS regions of the U.S. mainland for the period of 2002 to 2017. Each HHS region represents some collection of associated states.
- **US-States.** This dataset is collected from the Center for Disease Control (CDC).\(^4\) It contains the count of patient visits for ILI (positive cases) for each week and each state in the United States from 2010 to 2017. After removing a state with missing data we keep 49 states remaining in this dataset.
- **Canada-Covid.** This dataset is publicly available at JHU-CSSE.\(^5\) We collect daily COVID-19 cases from January 25, 2020 to January 10, 2022 in Canada (including 10 provinces and 3 territories).

We adopt two metrics for evaluation that are widely used in epidemiologic forecasting, including the **Root Mean Squared Error (RMSE)** and the **Pearson’s Correlation (PCC)**. RMSE measures the difference between predicted and true value after projecting the normalized values into the real range. PCC is a measure of the linear dependence between time series. For RMSE lower value is better, while for PCC higher value is better.
Table 2: RMSE and PCC performance of different methods on four datasets with horizon = 3, 5, 10. Bold face indicates the best result of each column and underlined the second-best. For RMSE lower value is better, while for PCC higher value is better.

| Dataset          | Methods | Metrics | Horizon | Horizon | Horizon | Horizon | Horizon |
|------------------|---------|---------|---------|---------|---------|---------|---------|
|                  |         | 3       | 5       | 10      | 3       | 5       | 10      |
|                  |         |         |         |         |         |         |         |
| Japan-Prefectures| AR      | RMSE    | 1705    | 2013    | 2307    | 757     | 997     | 1330    |
|                  |         | PCC     | 0.579   | 0.310   | 0.238   | 0.878   | 0.792   | 0.612   |
|                  | LRidge  | RMSE    | 1711    | 2025    | 1942    | 870     | 1059    | 1270    |
|                  |         | PCC     | 0.308   | 0.429   | 0.238   | 0.878   | 0.792   | 0.612   |
|                  | LSTMNet | RMSE    | 1459    | 1883    | 1811    | 801     | 998     | 1157    |
|                  |         | PCC     | 0.728   | 0.432   | 0.518   | 0.868   | 0.746   | 0.609   |
|                  | TPA-LSTM| RMSE    | 1142    | 1192    | 1677    | 761     | 950     | 1388    |
|                  |         | PCC     | 0.879   | 0.868   | 0.644   | 0.847   | 0.814   | 0.675   |
|                  | CNNRNN-Res | RMSE  | 1550    | 1942    | 1865    | 738     | 936     | 1233    |
|                  |         | PCC     | 0.673   | 0.380   | 0.438   | 0.862   | 0.782   | 0.552   |
|                  | SAIFlu-Net | RMSE  | 1356    | 1430    | 1527    | 661     | 871     | 1158    |
|                  |         | PCC     | 0.765   | 0.654   | 0.592   | 0.903   | 0.800   | 0.674   |
|                  | Cola-GNN | RMSE    | 1051    | 1117    | 1372    | 636     | 855     | 1134    |
|                  |         | PCC     | 0.901   | 0.890   | 0.813   | 0.909   | 0.835   | 0.717   |
|                  | SEFNet  | RMSE    | 1020    | 1123    | 1319    | 618     | 821     | 1036    |
|                  |         | PCC     | 0.904   | 0.893   | 0.826   | 0.909   | 0.842   | 0.725   |

4.4 Main Results

We evaluate our model in short-term (horizon = 3) and long-term (horizon = 5, 10) settings. Table 2 summarizes the results of all methods. The large difference in RMSE values across different datasets is due to the scale and variance of the datasets, i.e., the scale of the Japan-Prefectures and Canada-Covid is greater than the US-Regions and US-States datasets, which is closely related to the prevalence of epidemics and population density. There is an overall trend that the prediction accuracy drops as the prediction horizon increases because the larger the horizon, the harder the problem.

We observe that the proposed SEFNet achieves the state-of-the-art results on most of the tasks. Traditional statistics methods (AR and LRidge) do not perform well in influenza-related datasets. The main reason is that they are based on oversimplified assumptions and only rely on historical records, they cannot model the strong seasonal effects in influenza datasets. For deep learning-based methods, the performance is improved since they make efforts to deal with nonlinear characteristics and complex patterns behind time series. However, some deep learning-based models work well on some datasets, while not well on others.

(1) **Performance.** The methods mainly focused on dependencies between regions/time series (Cola-GNN and TPA-LSTM) have better performance than the method mainly focused on dependencies between time points (LSTNet), which can point out that inter-series dependencies are quite valuable information. Our proposed model takes inter-series correlations between different regions and temporal relationships in a single region into consideration and carefully fuses them for prediction to achieve better performance.

(2) **Stability.** For Japan-Prefectures and Canada-Covid datasets, the prediction performance of many compared methods greatly decreases when the horizon increases. Because as the variance of the dataset increases, the fluctuation within time series and dependencies between regions are more intricate. SEFNet can make full use of dependencies information, so its prediction error rises smoothly and slowly within the growing prediction horizon. Therefore, SEFNet has a better stability.

Through this experiment, it can be concluded that SEFNet has better performance and stability in epidemic forecasting, especially in long-term prediction.

4.5 Ablation Study

In order to clearly verify that the above improvement comes from each added component, we conduct an ablation study on the US-Regions and US-States datasets. Specifically, we remove each component one at a time in SEFNet. We name the model without different components as follows:

- w/oInter The model without Inter-Series Embedding module.
- w/oIntra The model without Intra-Series Embedding module.
- w/oAR The model without the AR component.
- w/oRAConv The model uses 1×3 convolution blocks only instead of Region-Aware Convolution.
In this paper, we propose an Inter- and Intra-Series Embeddings Fusion Network (SEFNet) for epidemic forecasting. We first extract inter- and intra-series embeddings from two parallel modules. Specifically, in inter-series embedding module, we design a Region-Aware Convolution component that is better to extract feature representations of time series and capture the dynamic dependencies among regions. Then, we fuse two embeddings through parametric-matrix fusion for prediction. To further enhance the robustness, we apply an AutoRegressive component as the linear part. Experiments on four real-world epidemic-related datasets show the proposed model outperforms the state-of-the-art baselines in terms of performance and stability. In future work, we plan to delve into the dynamic dependencies and mutual influences among regions.

ACKNOWLEDGMENTS
This work is supported by the Key R&D Program of Guangdong Province No.2019B010136003 and the National Natural Science Foundation of China No. 62172428, 61732004, 61732022.

REFERENCES
[1] Yue Cui, Chen Zhu, Guanyu Ye, Zewei Wang, and Kai Zheng. 2021. Into the Unobservables: A Multi-range Encoder-decoder Framework for COVID-19 Prediction. In Proc. of CIKM.
[2] Songgaojun Deng, Shusen Wang, Huzefa Rangwala, Lijing Wang, and Yue Ning. 2020. Cda-Rgb: Cross-location attention based graph neural networks for long-term iil prediction. In Proc. of CIKM.
[3] Shuhua Guo, Fan Fang, Tao Zhou, Wei Zhang, Qiang Guo, Rui Zeng, Xiaohong Chen, Jianguo Liu, and Xin Lu. 2021. Improving Google Flu trends for COVID-19 estimates using Weibo posts. Data Science and Management (2021).
[4] Tiberiu Harko, Francisco SN Lobo, and MK197716 Mak. 2014. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. Appl. Math. Comput. (2014).
[5] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural computation (1997).
[6] Siteng Huang, Donglin Wang, Xuehan Wu, and Ao Tang. 2019. DSANet: Dual self-attention network for multivariate time series forecasting. In Proc. of CIKM.
[7] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proc. of ICLR.
[8] Jayson S Jia, Xin Lu, Yun Yuan, Ge Xu, Jianmin Jia, and Nicholas A Christakis. 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature (2020).
[9] Xiaoyong Jin, Yu-Xiang Wang, and Xifeng Yan. 2021. Inter-series attention model for covid-19 forecasting. In Proc. of SDM.
[10] Seungwon Jung, Jaeok Moon, Sungwoo Park, and Eunjun Hwang. 2021. Self-attention-based Deep Learning Network for Regional Influenza Forecasting. IEEE Journal of Biomedical and Health Informatics (2021).
[11] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling long-and short-term temporal patterns with deep neural networks. In Proc. of AAAI.
[12] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. 2022. Deep Graph Clustering via Dual Correlation Reduction. In Proc. of AAAI.
[13] Edison Zangiacomi Martinez, Elisângela Aparecida Soares da Silva, and Amaury Leîis Dal Fabbro. 2011. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. Revista da Sociedade Brasileira de Medicina Tropical (2011).
[14] Shun-Yao Shih, Fan-Keng Wu, and Hung-yr Lee. 2019. Temporal pattern attention for multivariate time series forecasting. Machine Learning (2019).
[15] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami-Namini. 2019. The performance of LSTM and BiLSTM in forecasting time series. In 2019 IEEE International Conference on Big Data (Big Data).
[16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vladimir Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In Proc. of CVPR.
[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. Proc. of NeurIPS (2017).
[18] Zheng Wang, Prithwish Chakraborty, Sumiko B Meekar, John S Brownstein, Jieping Ye, and Naren Ramakrishnan. 2015. Dynamic poisson autoregression for influenza-like-illness case count prediction. In Proc of KDD.
[19] Miguel Won, Manuel Marques-Pita, Carlota Louro, and Ioana Gomolch-Sá. 2017. Early and real-time detection of seasonal influenza onset. Flus computational biology (2017).
[20] Yuexin Wu, Yiming Yang, Hiroshi Nishiura, and Masaya Saitoh. 2018. Deep learning for epidemiological predictions. In Proc. of SKG.
[21] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaoqun Chang, and Chengqi Zhang. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proc. of KDD.
[22] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In Proc. of AAAI.