DSI DATA SCIENCE AND INFORMETRICS
VOL.2, NO.2, MAY. 2022

Traffic load predictions for cellular networks: A comparative study based on deep learning

Miao Zhang, Chunshan Liu, Lou Zhao

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China

ABSTRACT

Prediction of the traffic load of cellular networks is important for network planning, load balancing, and operational optimization. In this paper, a comparative study of cellular traffic load prediction models based on deep learning is performed and a prediction method built on a multi-channel Gated Recurrent Unit (GRU) model is proposed. The proposed method uses multiple channels to extract the daily and weekly variation feature as well as the variation feature of the peak period of the BS load and can be used to provide 24-hour ahead predictions. Experimental results obtained from real dataset show that the proposed multi-channel model can effectively capture the temporal-variations of BS load and reduce the prediction error. Compared with conventional prediction algorithms such as Convolutional Neural Network (CNN), Long Short-Term Memory network, GRU and combination of CNN and GRU (CNN-GRU), the proposed model can achieve better prediction accuracies.

KEYWORDS

Traffic prediction; 24-hour ahead prediction; GRU

1 INTRODUCTION

The fast development of cellular networks has posed significant challenges in network management and operational optimization. Conventional business analysis approaches based on mathematical models and human experts cannot catch up with the speed of fast evolution of modern cellular networks. Data-driven approaches have thus gained more and more attention from both academia and industry. For cellular networks, accurate predictions of network load, e.g., the throughput of Base Stations (BSs) or the number of active calls, are helpful for network planning, load balancing and energy saving optimization (Xu et al., 2016). For instance, Long Short-Term Memory (LSTM) network combined with historical BS load data has been used to forecast the BS load, based on which energy saving strategies have been developed for 5G networks (Wang, 2021). However, accurate prediction of the load of cellular BSs is challenging due to the randomness of user behavior, the influences of seasonalities and special events and the existence of diverse types of BSs.

Generally, the load of cellular BSs changes over time. Hence, the historical load of a BS can be considered as a time series. In this regard, the prediction of BS load can be viewed as a time-series forecasting problem, for which there are many classical models including such as linear regression (Moghaddas-Tafreshi & Farhadi, 2008), autoregressive Integrated Moving Average model (ARIMA) (Wei & Zhen-gang, 2009; Shu et al., 2003), exponential smoothing

(Song et al., 2005), and multiple linear regression. However, classical models are often limited to solve linear problems, thus may not be adequate for the BS load prediction. As an alternative and promising route, deep learning (DL) models can deal with non-linear problems and hence can be used to perform the BS load predictions.

For DL based time-series prediction, choosing an appropriate neural network structure is crucial since different models are suitable for different types of data. For time-series prediction, existing works have made different attempts that have used various DL models. In (Liang et al., 2019), a Convolutional Neural Network (CNN)-based mobile traffic prediction method was developed to predict the mobile traffic of BSs deployed along highways. Again based on CNN, an ultra-shortterm wind power probabilistic prediction model was proposed in (H.-z. Wang et al., 2017), and its prediction accuracy was demonstrated with real dataset. In (Kong et al., 2019), a load forecasting framework based on LSTM recurrent neural networks was developed to predict household power consumption. The LSTM based framework was thoroughly tested against multiple benchmarking schemes including k-Nearest Neighbor (Zhang et al., 2016), extreme learning machine (Zhang et al., 2013) and a sophisticated input selection scheme combined with a hybrid forecasting framework (Ghofrani et al., 2015). It turns out that the LSTM framework usually achieved the best prediction accuracy in the short-term forecast of household energy consumption. In (Duan et al., 2018), LSTM is combined with seasonal and trend decomposition to predict the traffic loads of BSs. Reference (Zhang et al., 2017) proposed to first decompose the original time series and then train Gated Recurrent Unit (GRU) networks to predict the decomposed subsequences, from which a prediction of the overall sequence can be produced. In (Sajjad et al., 2020), a CNN-GRU energy prediction model based on hybrid sequential learning was proposed to obtain energy predictions. The numerical results there demonstrate that the method has better accuracy than CNN, LSTM and combination of CNN and LSTM (CNN-LSTM). CNN and LSTM were also combined to construct a neural network for fault prediction (Zheng et al., 2019), where multiple features were used to improve the accuracy and reliability of the predictions. In (Huang et al., 2017) , Recurrent Neural Network (RNN), three-dimensional CNN and the combination of CNN and RNN (CNN-RNN) for traffic load prediction were compared, where the numerical results obtained from the dataset provided by Telecom Italia show that CNN-RNN had the best prediction performance. Furthermore, in (Lin et al., 2021), a multi-channel input model combined with attention mechanism was used to capture the temporal characteristics of BS traffic data. While there has already been extensive effort devoted to developing neural-network based methods for time-series predictions, using a single universal model for the prediction of different types of time series is still challenging.

In this paper, we perform a comparative study on DL-based cellular traffic predictions and propose a new DL model that improves the GRU model by utilizing multiple channels to capture the complicated characteristics of BS traffic load. The proposed model can be trained to learn the characteristics of different types of BS traffic patterns at the same time, thus serves as a universal predictor for BS traffic load in cellular networks. The model is used to provide 24-hour ahead predictions, i.e., prediction the traffic load of a BS in the next 24 hours. The proposed model is compared with several commonly adopted DL models and its performance advantages are demonstrated using dataset obtained in real cellular networks. The contributions of this work are summarized as follows.

1. We propose an improved GRU model that utilizes multiple channels to extract the daily and weekly variation features as well as the variation feature of the BS load in peak time. The proposed model shows good universality for different types of time-series.

2. We conduct a systematic evaluation of the performance of the proposed model, with comparisons to commonly adopted DL models including CNN, LSTM, GRU, CNN-GRU. The performance advantages of the proposed model are demonstrated using real dataset.

2 Description of the Dataset and Preprocessing

The BS load data used to train and test the proposed model to be detailed is the counter values of 'pmRrcConnLevSum', a BS counter that reflects the total number of active Radio Resource Control (RRC) connection attempts (Li et al., 2020) to a BS, collected by a fourth generation long term evolution (4G-LTE) service provider from 15/01/2018 to 15/11/2018. The counter was collected once per hour at each BS, thus each value reflects the number of RRC connection attempts in an hour made to the corresponding BS. Preliminary inspections show that most of the time-series exhibit daily/weekly periodicity, as expected. However, abrupt changes, trend (e.g., increasing trend) and irregular daily variations are also common on top of the daily/weekly pattern.

Figure 1 plots four representative examples of the data, obtained from four different BSs. In this work, we categorize the original time-series of all BSs into four groups to match the characteristics shown in Figure 1 (a)-(d), and then randomly select five BSs from each group to make a dataset of 20 time-series. The dataset is then divided into three parts according to the time, with the ratio 6:2:2 to form the training (60%), validation (20%) and test (20%) dataset. Model training and evaluation are performed using this dataset.

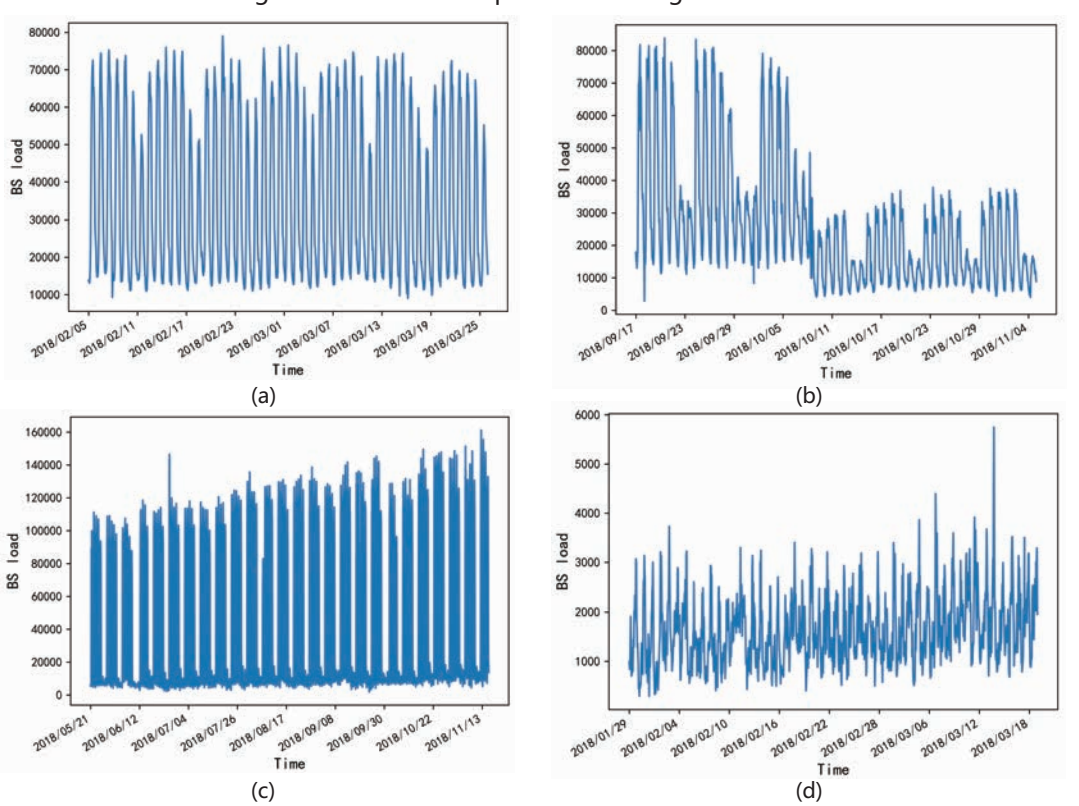

Figure 1 An illustration of four representative types of the counter data time-series with: (a) daily and weekly periodicity, (b) abrupt changes, (c) trend (increasing) and (d) irregular variations.

2.1 Data Processing

Denote a complete time-series across the training, validation and testing period as X_{raw}= $\{X_1^{raw},X_2^{raw},X_3^{raw}\ldots X_n^{raw}\}$, where n is the total number of time instances (the number of hours in the current dataset). The following two steps are performed for missing value imputation and data transformation to enhance the integrity of the data.

1. Missing value imputation: Missing values are presented as 'Null' in X_{raw} . Fortunately, the original data is of high quality and the percentage of Null values is low (0.06%). Suppose sample ,X_j^{raw} is 'Null', then it is replaced by $X_j^{raw} \leftarrow (X_i^{raw}+X_k^{raw})/2$, where X_i^{raw} and X_k^{raw} are the nearest valid sample before ($i < j$) and after ($k > j$) sample X_j^{raw}, respectively. In case that j is the first or the last sample of the time series, it is replaced by the nearest valid sample.

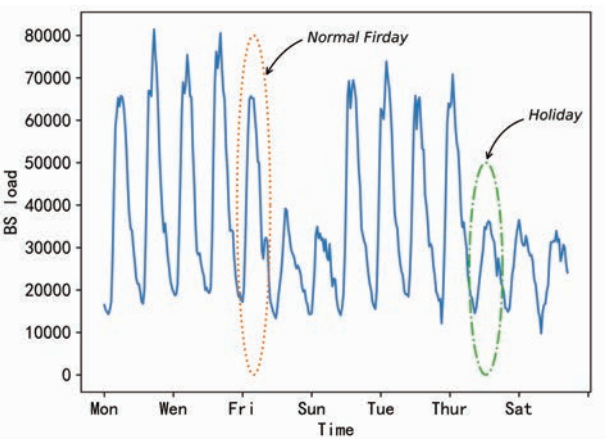

Figure 2 An illustration on the impact of public holidays on the BS load.

2. Logarithm data transformation: As a further step to pre-process the data, we take the logarithm transformation of the data after imputation: $X'_{raw}=\log X_{raw}$ where X'_{raw} represents the data after transformation. This step is to reduce the variation order to facilitate the training of the prediction model.

2.2 Feature Selection

As illustrated in Figure 1, the counter value generally varies periodically every day and every week. Additionally, as expected, public holidays also have a significant impact on the value of the data. For instance, as shown in Figure 2, the counter value on a Friday which is a public holiday is different from that of a normal Friday and behaves similarly to a weekend. Since the dataset only covers a period of about 11 months and the number of public holidays (non-weekend holidays) is relatively small, it is difficult for the DL model to learn the characteristics of non-weekend holidays and give accurate traffic load predictions. Therefore, we include the indicator of working day and the indicator of holiday as additional input to the DL models. Furthermore, we also add the indicator of holiday for the predicted day as an input feature.

3 DL Baseline Models for BS Load Prediction

In this section, we give a brief review of several classical DL models that will be used as

baselines in cellular BS load prediction, including CNN, LSTM, GRU and CNN-GRU. Denote the BS load at time instance t as x_t, the corresponding encoded weekday as $w_t \in \{1,2,\ldots,7\}$, the indicator of working days as $iw_t \in \{0,1\}$, the indicator of public holiday as $ih_t \in \{0,1\}$, and the indicator whether the predicted values are on a public holiday as $fw_t \in \{0,1\}$. With the notation described, the input features for time instance t can be represented by vector $X_t=[x_t, w_t, iw_t, ih_t, fw_t]$, which is of dimension $N \times 1$ and $N = 5$. Then the BS load prediction problem can be formulated as using a set of T input samples $X_t = \{X_{t-T+1}, \ldots, X_{t-T+2}, X_{t-2}, X_t\}$ to predict m BS load values beyond time instance t, i.e., $y_t = \{x_{t+1}, x_{t+2}, \ldots, x_{t+m}\}$, where m =24 for the 24-hour ahead predictions.

3.1 Convolution Neural Network (CNN) –Based Model

Figure 3 presents the CNN model used for predicting the BS load, which consists of two convolutional layers.

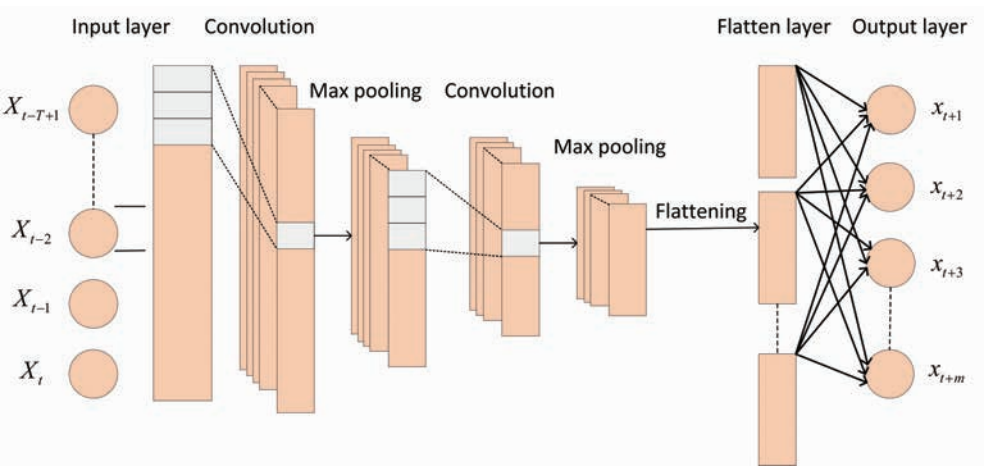

Figure 3 Architecture of the CNN prediction model.

The input is x_t and is of dimension $T \times N$, while the shape of the convolution kernel is $S \times N$. Here, S represents the size of the convolution kernel. After two convolution and pooling operations, the predicted BS load $y_t = \{x_{t+1}, x_{t+2}, \ldots, x_{t+m}\}$ are generated through the fully connected layer.

3.2 Long Short–Term Memory (LSTM) –Based Model

LSTM networks are extensions of recurrent neural networks that are commonly used for timeseries predictions. Each LSTM cell has an input gate, and an forget gate to control the internal information flow:

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned}, \quad (1)$$

where σ denotes the logistic sigmoid function, x_t represents the input vector at the current moment, h_t is the hidden state, W is the weight matrices (e.g., W_i represents the input gate weight matrix) and b is the bias vectors (e.g.,b_i represents the input gate bias vector) for the three gates. Figure 4 shows the structure of an LSTM cell.

In this work, we use a two-layer LSTM network for BS load prediction. The overall architecture of the LSTM network is shown in Figure 5. With this network, the input vector $X_{t\text{-}T+1}$ of the first time instance is passed through the first LSTM cell in the first LSTM layer to obtain the output state vector $h^1_{t\text{-}T+1}$, which, together with the input vector $X_{t\text{-}T+2}$ at the next moment, form the input into the next LSTM cell in the first layer to generate $h^1_{t\text{-}T+2}$; this process continues until the last input (the latest time instance) is completed.

The output of the first LSTM layer is fed into the second LSTM layer as the input and generates $h^2_{t\text{-}k+1}$,k={T,T-1,… ,1}. The output state vector $h^2_{t\text{-}1}$ generated by the last LSTM-cell of the second LSTM layer is finally flattened into a fully connected layer to obtain the predictions y_t= { x_{t+1}, x_{t+2}, … , x_{t+m}} .

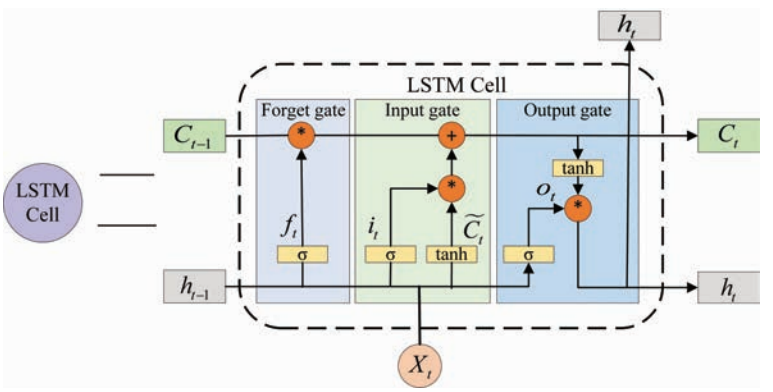

Figure 4 Structure of an LSTM CELL.

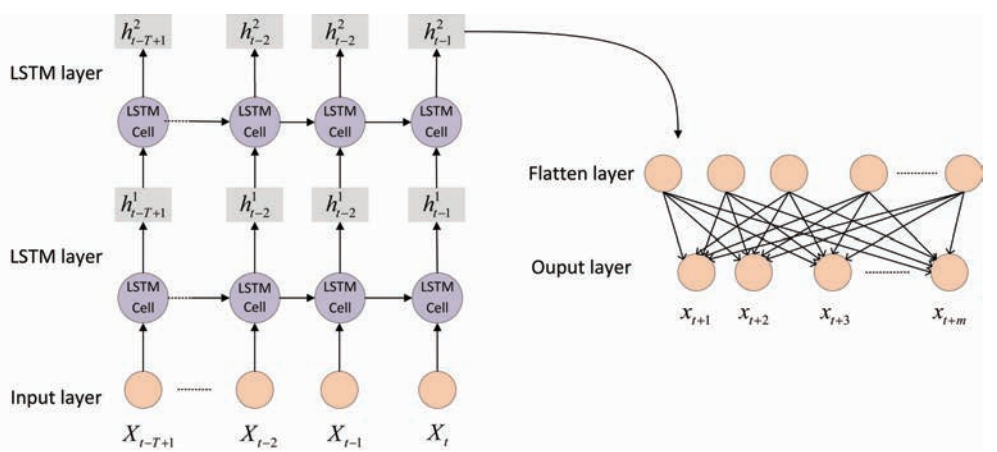

Figure 5 Structure of the LSTM-based prediction model.

3.3 Gated Recurrent Unit (GRU) –Based Model

The principle of using GRU networks to predict the BS load is similar to that of the LSTM networks. Compared with LSTM, GRU combines the internal state vector and output vector

into one state vector, and uses a reset gate and update gate to replace the LSTM gates. GRU can achieve similar prediction accuracies as LSTM in sequence learning. However, it is easier to train and has lower computational complexities. A GRU cell has the following internal information flow:

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \\ \tilde{h}_t &= \tan h(W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \\ z_t &= \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{aligned} \tag{2}$$

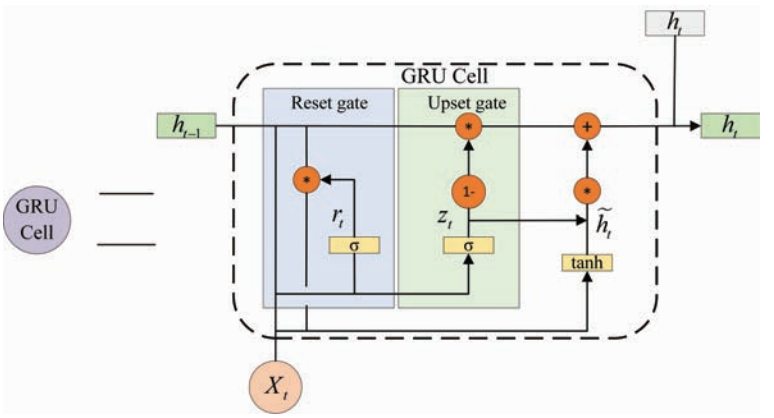

Figure 6 Structure of a GRU CELL.

where σ denotes the logistic sigmoid function, x_t represents the input vector at the current moment, h_t is the hidden state, W is the weight matrices and b the bias vectors.

Figure 7 presents the GRU model used for predicting the BS load. Similar to the LSTM network training process, the features extracted from the historical data are transformed to the last GRU cell through two layers of GRU, and the outputs of the last GRU Cell at the second layer are flattened and then fed into the fully layer to obtain the final predictions.

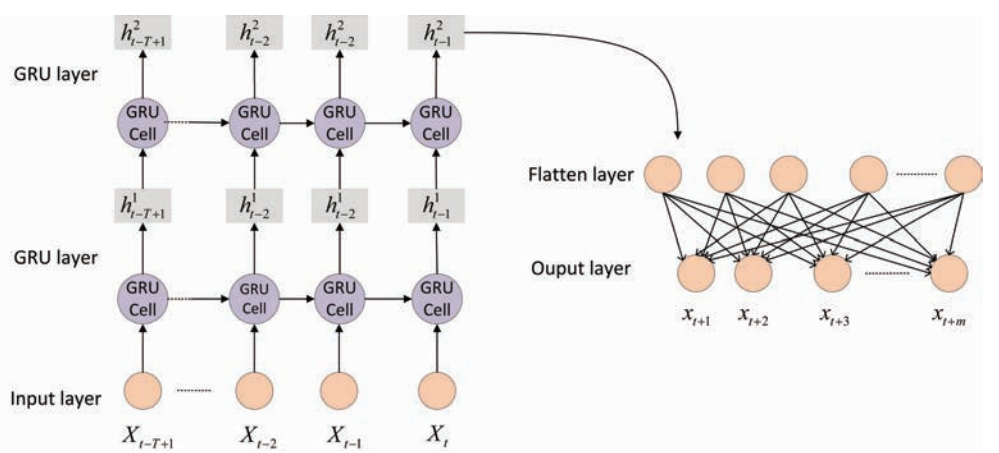

Figure 7 The structure of the GRU prediction model

3.4 CNN–GRU Based Model

The CNN-GRU model combines the effectiveness of CNN in extracting features and the low computational complexity and powerful learning of GRU in time series learning.

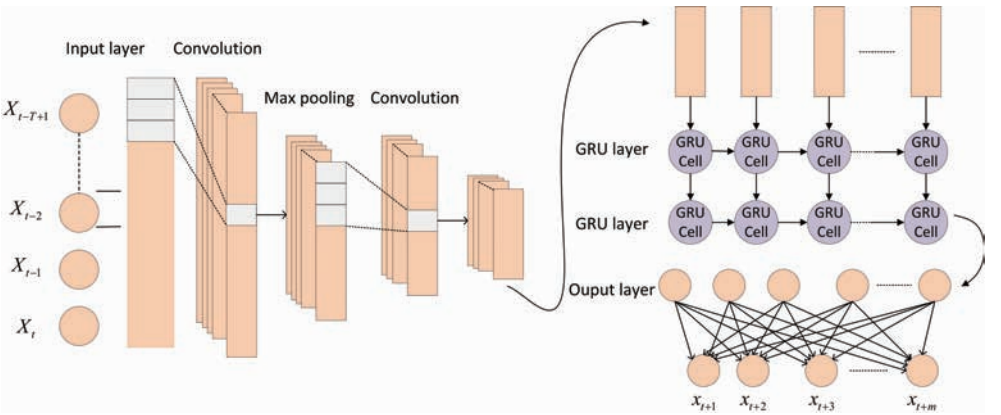

Figure 8 Structure of the CNN-GRU prediction model.

In CNN-GRU, a CNN network is first used to extract the features of the input sequence $\{X_{t-T+1} \ldots X_{t-T+2}, X_{t-2}, X_{t-1}\}$ by performing two convolutional and pooling operations. Then the extracted features are put into a GRU network for training, and the final feature extraction results are obtained after the two-layer GRU network. The outputs of the GRU network are flattened and passed through a fully connected layer to obtain the output $y_t = \{x_{t+1}, x_{t+2}, \ldots, x_{t+m}\}$. Figure 8 depicts the structure for CNN-GRU network we used for BS load prediction.

4 The proposed model: Multi–channel GRU

Preliminary investigations show that the models reviewed above, which mainly use the original historical data as the input feature, do not predict the peak BS load accurately. To solve this problem, in this section, we propose a multi-channel GRU model. The proposed model uses three channels to extract the daily and weekly fluctuation characteristics of the BS load, as well as the characteristics of the peak load to enhance the prediction accuracy. It is noted that the input feature used to extract the characteristics of the peak load is in fact the historical BS load during peak time, which is part of the input feature used for learning the daily fluctuation pattern. The repeated use of the historical data of this period is to augment the GRU model such that the BS load prediction can better capture the significant fluctuations of the daily peak values.

The input of the first channel is the historical data of consecutive days, defined as follows:

$$\begin{aligned} \mathcal{X}_d \triangleq (&X_{t-24*T_d+1}, \cdots, X_{t-24*T_d+24}, \\ &X_{t-24*(T_d-1)+1}, \cdots, X_{t-24*(T_d-1)+24}, \\ &\cdots X_{t-24+1}, \cdots, X_{t-1}), \end{aligned} \tag{3}$$

where T_d represents the number of days. In this work, $T_d = 7$, thus X_d contains 7 days of data.

The input of the second channel contains historical data from the same weekday from the predicted day:

$$\begin{aligned}\mathcal{X}_w \triangleq (&X_{t-24*7*T_w+1}, \cdots, X_{t-24*7*T_w+24},\\ &X_{t-24*7*(T_w-1)+1}, \cdots, X_{t-24*7*(T_w-1)+24},\\ &\cdots X_{t-24*7+1}, \cdots, X_{t-24*7+24}),\end{aligned} \tag{4}$$

where T_w is the number of weeks considered. In this work, T_w =3 hence historical data from the 3 same weekdays of the weeks before the prediction time t is used.

The input of the third channel contains data between hour T_s and T_e on each of the past T_r days:

$$\begin{aligned}\mathcal{X}_h \triangleq (&X_{t-24*T_r+1+T_s}, \cdots, X_{t-24*T_r+1+T_e},\\ &X_{t-24*(T_r-1)+1+T_s}, \cdots, X_{t-24*(T_r-1)+1+T_e},\\ &\cdots X_{t-24+1+T_s}, \cdots, X_{t-24+1+T_e}),\end{aligned} \tag{5}$$

Here T_s= 7 and T_e =16 when performing the 24-hour ahead prediction.

Figure 9 presents the architecture of the proposed multi-channel GRU network. Each channel is in fact a two-layer GRU network separated from each other. The input sequences are fed into the corresponding models. The outputs of the three GRU networks are then concatenated and passed through a fully connected layer to obtain the BS load predictions. In this way, the final results are a weighted combination of forecasts made according to the daily pattern, weekly pattern and the peak period of each day.

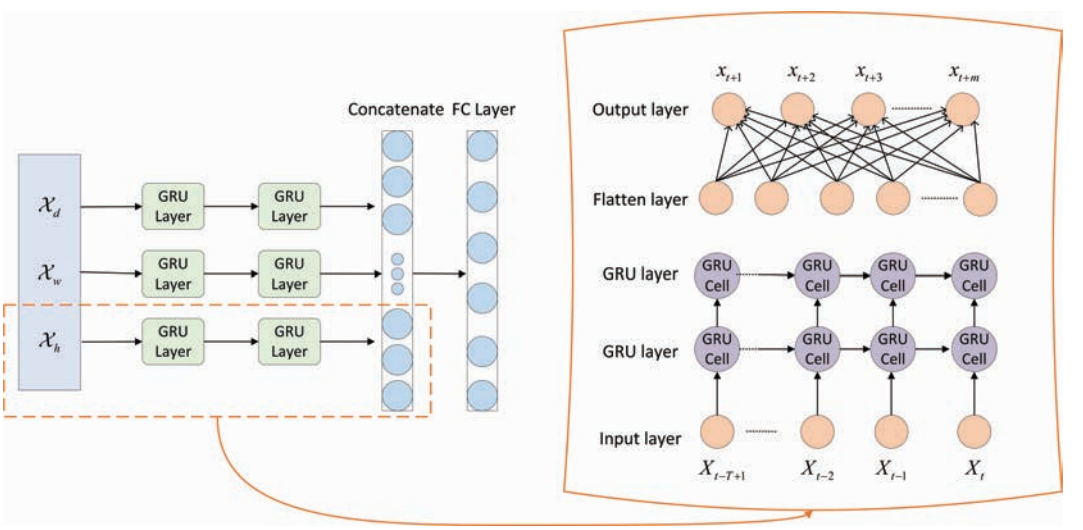

Figure 9 The architecture of the proposed multi-channel GRU prediction model.

5 Numerical Results

In this section, we evaluate the performance of the proposed multi-channel GRU model for BS load prediction. The proposed model is compared to the baseline schemes described in Section 3, including CNN, LSTM, GRU and CNN-GRU. Table 1 presents the hyper parameters of the models considered as well as the parameters for training the models.

Table 1 The hyper parameters of the models considered as well as the parameters for training the models.

Parameters	CNN	LSTM	GRU	CNN–GRU		Multi–channel GRU		
				CNN	GRU	channel1	channel2	channel3
Depth	2	2	2	2	2	2	2	2
Input Length	168	168	168	168	40	168	72	210
Hidden neural	None	100	100	None	100	100	50	100
Kernel size	64,32	None	None	64,32	None	None	None	None
Dropout rate	None	0.2	0.2	None	0.2	0.2	0.2	0.2
Learning rate	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
Loss function	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam

Table 2 Performance comparison of different prediction algorithms for multi-output prediction of four types of base station data.

Metrics	CNN	LSTM	GRU	CNN–GRU	Multi–channel GRU
MAPE	17.12%	16.79%	16.20%	17.39%	15.49%
NMAPE	14.69%	15.18%	14.38%	15.88%	13.49%
R2	0.7974	0.7830	0.8025	0.7657	0.8179

To evaluate the prediction accuracy of the proposed model and the baselines, the Mean Absolute Percentage Error (MAPE), Normalized MAPE (NMAPE) (Ahmed et al., 2020) and the R-square (R2) are adopted as the performance metrics. It is noted that absolute error metrics such as the mean absolute error are not adopted since the BS load varies significantly across different BSs, thus the absolute metrics cannot accurately reflect the performance of the prediction models.

Denote y_i as the true BS load and $\hat{y}_i$ as the corresponding prediction, then the MAPE can be calculated as:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|, \tag{6}$$

The NMAPE can be calculated as:

$$\mathrm{NMAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{\overline{y}}, \tag{7}$$

where $\overline{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$ is the average BS load. R2 is the coefficient of determination, which reflects the proportion of the total variance of the dependent variable that the independent variable can explain through the regressed relationship, i.e., the trained DL model. It can be calculated as:

$$R2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}. \tag{8}$$

In the following, we present the numerical results obtained for the 24-hour ahead

predictions.

Figure 10(a)-(d) present the predicted BS load series obtained from the proposed model and the baselines for the four different types explained in Figure 1, respectively. As can be seen from the figure, all the DL models considered perform reasonably well in capturing the differences between working days and non-working days, thus provide good prediction accuracies. The proposed model provides visually more accurate predictions on the peak BS load than the baselines.

Moreover, for the more challenging task of predicting the BS load in the presence of abrupt changes, as illustrated in Figure 10(b), the proposed multi-channel GRU exhibits the best adaptability to the changes. As can be seen from Figure 10(b), all the DL models have significant prediction errors when the change occur, which is inevitable. However, after the change, the multi-channel GRU quickly captures the new characteristics of the BS load and provide more accurate predictions.

For trending data as in Figure 10(c) (see Figure 1(c) for the complete time-series), LSTM and GRU models fail to give accurate predictions of the BS load on weekends, due to the orders of magnitude difference between the peak loads of working days and non-working days. The prediction errors of CNN are also significant on weekend. As a comparison, the proposed model provides consistently accurate predictions on both working and non-working days.

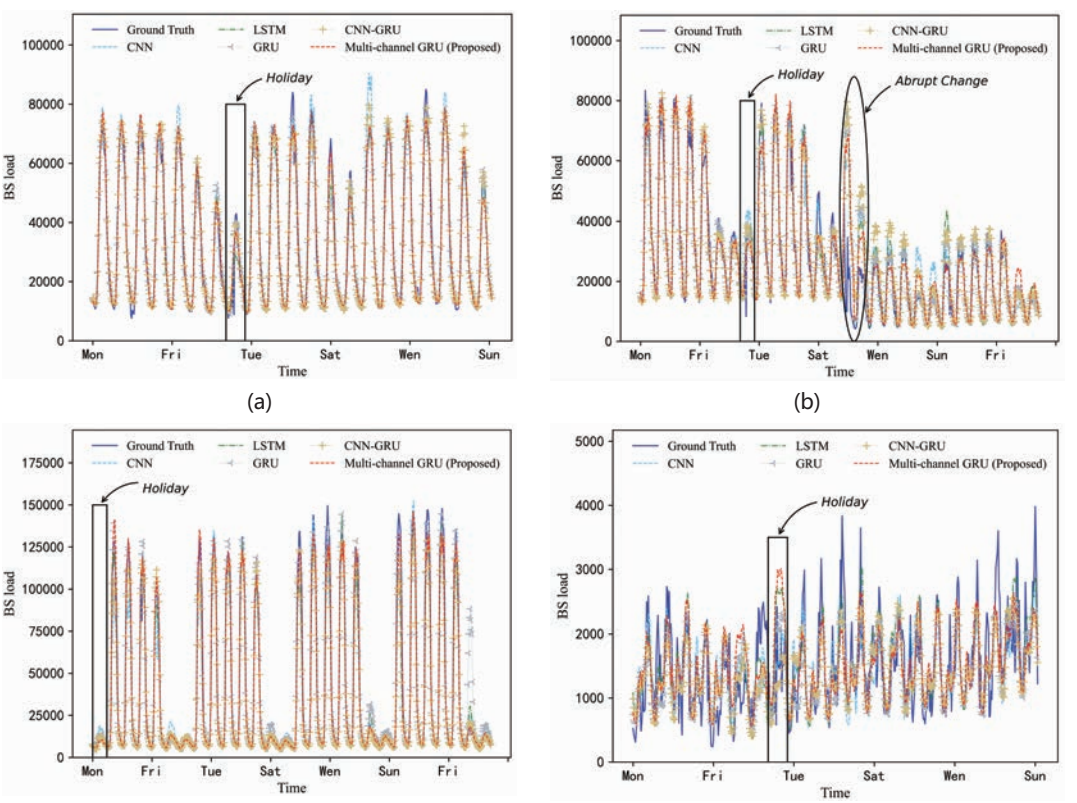

Figure 10 A Comparison of predicted and valid values for all models in Multi-Output Prediction.

Table 2 presents the overall performance of the proposed model and the baselines. As can be seen, for all the three metrics considered, the proposed multi-channel GRU model performs the best. Compared to CNN, LSTM, GRU and CNN-GRU, the proposed model reduces the MAPE by 9.52%, 7.74%, 4.38%, and 10.93%, and reduces the NMAPE by 8.17%, 11.13%, 6.45%, and 15.05%, respectively.

Figure 11(a)-(d) presents the cumulative distribution function (CDF) of NMAPE of all the tested instances (each time instance corresponds to a 24-hour prediction) for the four different types of time-series. As can be seen, the performance improvement from the proposed model mainly comes from the first three types of time-series. For the last type with irregular fluctuations which is more challenging to predict, the proposed model performs similarly to the baselines.

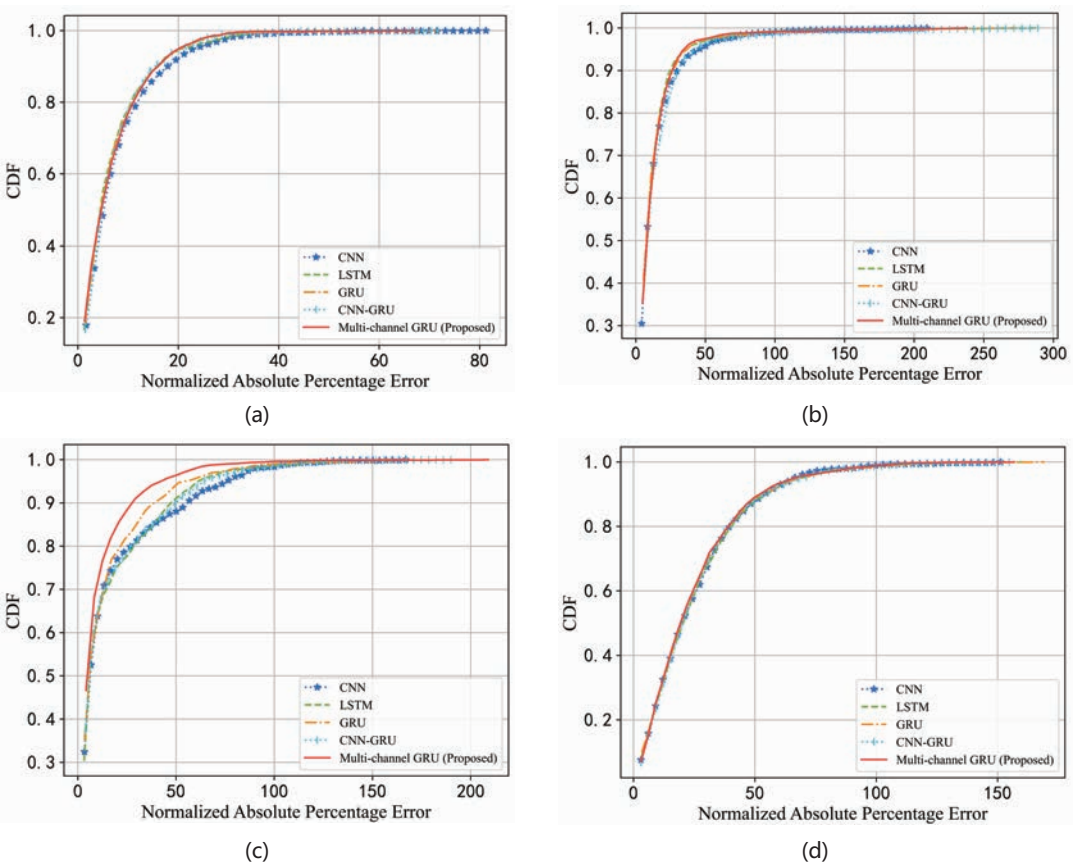

Figure 11 Cumulative Distribution Function (CDF) of NMAPE of the predictions.(a) daily and weekly periodicity, (b) abrupt changes, (c) trend (increasing) and (d) irregular variations.

6 CONCLUSIONS

In this work, a multi-channel GRU model has been proposed for the prediction of the traffic load in cellular networks. The proposed model adopts three GRU networks to learn the daily and weekly pattern of the traffic load as well as the fluctuation characteristics of the load in peak times. A comprehensive comparison to popular DL based models, including CNN, LSTM, GRU and CNN-GRU has been performed with real dataset collected by a LTE

mobile service provider. Numerical results demonstrate 6% to 15% improvement over CNN, LSTM, GRU and CNN-GRU, when performing the 24-hour ahead traffic load predictions for individual BSs. Future extensions of the current work may include the utilization of data from adjacent BSs to learn the spatiotemporal characteristics of the BS load so as to enhance the prediction accuracies. The prediction accuracy may also be improved by using multiple counters that are partially correlated with each other.

Acknowledgement

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LZ22F010001.

References

Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. (2020). A review and evaluation of the state–ofthe–art in PV solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews, 124,* 109792. doi: 10.1016/j.rser.2020.109792

Duan, Q., Wei, X., Gao, Y., & Zhou, F. (2018). Base station traffic prediction based on STL–LSTM networks. In *2018 24th Asia–Pacific Conference on Communications (APCC)* (p. 407–412). doi:10.1109/APCC.2018.8633565

Ghofrani, M., Ghayekhloo, M., Arabali, A., & Ghayekhloo, A. (2015). A hybrid short–term load forecasting with a new input selection framework. *Energy, 81* (119), 777–786. doi: 10.1016/j.energy.2015.01.028

Huang, C.–W., Chiang, C.–T., & Li, Q. (2017). A study of deep learning networks on mobile traffic forecasting. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (pp. 1–6). doi: 10.1109/PIMRC.2017.8292737

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short–Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Transactions on Smart Grid, 10* (1), 841–851. doi: 10.1109/TSG.2017.2753802

Li, M., Wang, Y., Wang, Z., & Zheng, H. (2020). A deep learning method based on an attention mechanism for wireless network traffic prediction. *Ad Hoc Networks, 107*, 102258. doi: 10.1016/j.adhoc.2020.102258

Liang, D., Zhang, J., Jiang, S., Zhang, X., Wu, J., & Sun, Q. (2019). Mobile Traffic Prediction Based on Densely Connected CNN for Cellular Networks in Highway Scenarios. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)* (p. 1–5). doi: 10.1109/WCSP.2019.8927980

Lin, J., Chen, Y., Zheng, H., Ding, M., Cheng, P., & Hanzo, L. (2021). A data–driven base station sleeping strategy based on traffic prediction. *IEEE Transactions on Network Science and Engineering,* 1–1. doi: 10.1109/TNSE.2021.3109614

Moghaddas–Tafreshi, S., & Farhadi, M. (2008). A linear regression–based study for temperature sensitivity analysis of iran electrical load. In *2008 IEEE International Conference on Industrial Technology* (p. 1–7). doi: 10.1109/ICIT.2008.4608590

Sajjad, M., Khan, Z. A., Ullah, A., Hussain, T., Ullah, W., Lee, M. Y., & Baik, S. W. (2020). A Novel CNN–GRU–Based Hybrid Approach for Short–Term Residential Load Forecasting. *IEEE Access, 8*, 143759–143768. doi: 10.1109/ACCESS.2020.3009537

Shu, Y., Yu, M., Liu, J., & Yang, O. (2003). Wireless traffic modeling and prediction using seasonal ARIMA models. *In IEEE International Conference on Communications, 2003.* ICC ´03. (Vol. 3, p. 1675–1679 vol.3). doi: 10.1109/ICC.2003.1203886

Song, K.–B., Baek, Y.–S., Hong, D. H., & Jang, G. (2005). Short–term load forecasting for the holidays using fuzzy linear regression method. *IEEE Transactions on Power Systems, 20* (1), 96–101. doi: 10.1109/TPWRS.2004.835632

Wang, H.–z., Li, G.–q., Wang, G.–b., Peng, J.–c., Jiang, H., & Liu, Y.–t. (2017). Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied energy, 188*, 56–70. doi: 10.1016/j.apenergy.

2016.11.111

Wang, Y. (2021, nov). Energy–saving Scheme of 5G Base Station Based on LSTM Neural Network. *Journal of Physics: Conference Series, 2083* (3), 032026. doi: 10.1088/1742–6596/2083/3/032026

Wei, L., & Zhen–gang, Z. (2009). Based on Time Sequence of ARIMA Model in the Application of Short–Term Electricity Load Forecasting. In *2009 International Conference on Research Challenges in Computer Science* (p. 11–14). doi: 10.1109/ICRCCS.2009.12

Xu, F., Lin, Y., Huang, J., Wu, D., Shi, H., Song, J., & Li, Y. (2016). Big data driven mobile traffic understanding and forecasting: A time series approach. *IEEE Transactions on Services Computing, 9* (5), 796–805. doi: 10.1109/TSC.2016.2599878

Zhang, R., Dong, Z. Y., Xu, Y., & Meng, K. (2013). Short–term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine. *Iet Generation Transmission & Distribution, 7* (4), 391–397. doi: 10.1049/iet–gtd.2012.0541

Zhang, R., Xu, Y., Dong, Z. Y., Kong, W., & Wong, K. P. (2016). A composite k–nearest neighbor model for day–ahead load forecasting with limited temperature forecasts. In *2016 IEEE Power and Energy Society General Meeting (PESGM)* (p. 1–5). doi: 10.1109/PESGM.2016.7741097

Zhang, X., Shen, F., Zhao, J., & Yang, G. (2017). Time series forecasting using GRU neural network with multi–lag after decomposition. *In International Conference on Neural Information Processing* (pp. 523–532). doi: 10.1007/978–3–319–70139–4_53

Zheng, L., Xue, W., Chen, F., Guo, P., Chen, J., Chen, B., & Gao, H. (2019). A fault prediction of equipment based on CNN–LSTM network. *In 2019 IEEE International Conference on Energy Internet (ICEI)* (p. 537–541). doi: 10.1109/ICEI.2019.00101