Idle Time Window Prediction in Cellular Networks With Deep Spatiotemporal Modeling

Luoyang Fang, *Student Member, IEEE*, Xiang Cheng, *Senior Member, IEEE*, Haonan Wang, and Liuqing Yang, *Fellow, IEEE*

***Abstract*— Idle time windows (ITWs) consist of one critical trigger for various functions in green intelligent network management and traffic scheduling in mobile networks. In this paper, we study the ITW prediction in mobile networks based on network subscribers' demand and mobility behaviors observed by network operators. We first innovatively formulate the ITW prediction into a regression problem with an ITW presence confidence index that facilitates direct ITW detection and estimation. Feature extraction on the demand and mobility history is then proposed to capture the current trends of subscribers' demand and mobility as well as to account for the periodicity underlying subscribers' demand and mobility patterns as exogenous inputs. In light of feature engineering, a deep learning-based ITW prediction model is proposed, which consists of two components, namely the representation learning network and the output network. The representation learning network is aimed to learn effective patterns, whereas the output network is designed to produce the desired ITW presence confidence index and the ITW estimate by integrating the learned representation and exogenous inputs. In this paper, a novel temporal graph convolutional network (TGCN) for the representation learning network is proposed to effectively capture the graph-based spatiotemporal input features. The experiment results validate the proposed direct ITW prediction formulation and demonstrate the superiority of the proposed TGCN in terms of both ITW detection and ITW estimation performance, which can achieve a significant intersection-over-union (IoU) improvement compared with baselines.**

***Index Terms*— Machine learning, mobile communication, predictive models.**

Manuscript received July 21, 2018; revised December 20, 2018; accepted March 3, 2019. Date of publication March 14, 2019; date of current version May 15, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61622101 and Grant 61571020, in part by the Shenzhen Fundamental Research Fund under Grant JCYJ20170411102217994, in part by the Shenzhen Peacock Plan under Grant KQTD2015033114415450, in part by the Guangdong Province under Grant 2017ZT07X152, and in part by the National Science Foundation under Grant DMS-1521746 and Grant DMS-1737795. *(Corresponding author: Xiang Cheng.)*

L. Fang is with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: luoyang.fang@colostate.edu).

X. Cheng is with the State Key Laboratory of Advanced Optical Communication Systems and Network, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China, and also with the Shenzhen Research Institute of Big Data, Shenzhen 518172, China (e-mail: xiangcheng@pku.edu.cn).

H. Wang is with the Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: wanghn@stat.colostate.edu).

L. Yang is with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA, and also with the Shenzhen Research Institute of Big Data, Shenzhen 518172, China (e-mail: lqyang@engr.colostate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2019.2904367

I. Introduction

As MOBILE phones (smartphones) have successfully penetrated nearly every aspect of human life due to the flourished mobile applications and services, massive data collected at mobile devices and in mobile networks, termed as mobile big data [1]–[3], has attracted remarkable attention from various research communities and industries. At the same time, with the virtualization and cloudization of network functions in future cellular networks [4]–[6], the mobile network architecture is undergoing a vast transformation facilitated by emerging technologies, such as software-defined networking, network function virtualization, etc., to support various traffic demands and to fulfill diverse quality-of-service (QoS) requirements [7]. Mobile big data collected by mobile network operators revealing user' behavior patterns [8] can significantly benefit the resource-constrained network automation, from network planning and network traffic monitoring to network management in mobile networks.

In recent years, self-organizing networks (SON) is widely studied to automatically manage and organize networks with much less manual interventions so that network operational expenditures (OPEX) and capital expenditures (CAPEX) can both be reduced [9], [10]. Mobile data plays a critical role in cellular SONs, providing system observability and predictability for network management. In fact, one of the largest portions of OPEX is the power consumption of cell towers [11]. To switch cells off when traffic loads across the network are extremely low [11], [12] is a potential approach to lower the power consumption of mobile networks. However, cell towers may not be able to switch on to meet the traffic demands in real time. As a result, the cell on/off switching requires reliable predictive knowledge of ITW for each cell in a mobile network to reduce power consumption with subscribers' quality of experience ensured. In fact, ITWs in the network may vary spatially, which needs to be carefully learned [13]. In addition, the ITW prediction is not limited to the application of cell switching on/off but can facilitate flexible traffic scheduling and network management applications. For example, time-dependent pricing [14] is one of the recently proposed solutions to reduce the peak-to-average (PAR) by motivating delay-tolerant traffic consumed in idle time windows so that the network congestion could be alleviated during the peak time. The dynamic pricing mechanism highly relies on the predictive knowledge of the ITWs across the network, as traffics and ITWs may vary temporally and spatially [15].

In this paper, we propose to study the ITW prediction for each cell in mobile networks, based on both the recent history and periodic factors of mobile demands and subscribers' aggregated mobility behaviors observed by mobile operators. Prediction of the ITWs essentially amounts to answer the following questions:

1) Will the ITW start within the prediction horizon in the future?
2) When will the ITW start and how long will it last?

In fact, the first question is essentially a detection problem, while the second question leads to a regression or localization problem. The intuitive approach of ITW prediction is to first perform a long-term demand forecasting for each cell in the network, based on which ITWs in the future could be extracted. However, the long-term demand forecasting is usually formulated as one-step-ahead prediction [13], [16]–[20] and then generates long-term forecasts one-by-one sequentially based on predicted results, leading to mediocre ITW prediction due to error accumulation during forecasting. Furthermore, to predict ITW based on long-term demand forecasting may be expensive, as it needs to generate far more estimates than actually desired (i.e., start time and duration).

Differing from the aforementioned approaches, we propose to directly predict ITW for each mobile network cell in this paper. Specifically, the start time and duration within the prediction horizon in the future will be directly estimated. To the best of our knowledge, this paper is the first work on direct ITW prediction. The ITW prediction is formulated as a regression problem with an *ITW presence confidence index*, which simultaneously tackles the ITW detection problem (whether an ITW will present or not) and the ITW estimation (where the ITW is located within the prediction horizon). In fact, the novel ITW presence confidence index proposed in this work can effectively indicate the presence of ITWs in the forecasting horizon, and also provide the flexibility and capability to control the robustness of the prediction model in different practical scenarios. In terms of feature engineering, we first propose an innovative feature extraction scheme to obtain multiple demands and mobility features from the raw signaling datasets so that reliable ITW prediction can be facilitated. In addition, the day-ahead and week-ahead periodic observations will be regarded as exogenous inputs to account for the strong temporal seasonality. Furthermore, the spatiotemporal semivariogram demonstrates that mobile demand at the cell level has a strong temporal relevancy yet relatively weak spatial relevancy, where the spatial relevancy among cells is modeled as a relevancy graph as in our previous work [19].

In light of the proposed feature engineering, exogenous inputs, and desired prediction output, we propose a novel ITW prediction model, consisting of the representation learning network and the output network. The representation learning network is aimed to learn the useful patterns from the recent demand and mobility history for ITW prediction via effective graph sequence modeling. The output network is aimed to combine the learned patterns via representation learning networks and exogenous inputs to generate the desired ITW presence confidence index and ITW location. In the literature, two graph-sequence spatiotemporal models, GeoMAN [21] and DCRNN [22] were recently proposed to deal with the sequence-to-sequence environmental pollution prediction and road traffic prediction, respectively. However, these two models were designed for strong spatial and strong temporal relevancy, which may not be ideally suitable for the mobile demand traffic with strong temporal relevancy yet relatively weak spatial relevancy. As a result, we further propose a novel prediction model, termed as temporal graph convolutional networks (TGCN), based on the cutting-edge temporal modeling [23] and graph modeling [24] techniques.

To effectively evaluate the performance of ITW estimation, we propose to employ a metric called intersection-over-union (IoU) borrowed from the object detection task in the field of computer vision [25], to assess how well the predicted time window overlaps with the ground truth. Experiment results demonstrate that our proposed general ITW prediction model can achieve a significant performance improvement compared with baselines, and the proposed TGCN-based model can outperform the temporal convolutional networks (TCN) [23], long short-term memory (LSTM) [26], and graph convolutional LSTM (GCLSTM) [19]. The good prediction performance suggests the validity of the proposed problem formulation and feature engineering as well as the superiority of our proposed TGCN model. The key contributions of this paper are summarized as follows:

- Direct ITW prediction is proposed for the first time. It is formulated as a regression problem with an ITW presence confidence index, which can simultaneously tackle both the ITW detection and estimation tasks.
- A feature extraction scheme is proposed to extract the demand and mobility features from the raw signaling dataset as well as the exogenous inputs to account for the inherent characteristics of the demand and mobility behaviors.
- A novel ITW prediction model consisting of representation learning network and output network is proposed to account for both the spatiotemporal feature and the exogenous inputs. In addition, a graph-sequence representation network model, TGCN, is proposed to characterize both the spatial and temporal relevancy in subscribers' demand and mobility behaviors to facilitate a good ITW prediction.
- A cost function combining the cross entropy loss for ITW prediction and the mean absolute error for ITW estimation is proposed to effectively train the prediction model. In addition, an evaluation metric, intersection-over-union (IoU), is proposed to assess the performance of the ITW estimation.

The rest of this paper is organized as follows. In Section II, the dataset employed in this paper is first described and then the direct ITW prediction problem is formulated. In Section III, feature engineering with respect to the characteristics of the demand and mobility behaviors of the network will be discussed. In Section IV, the proposed prediction model will be presented and discussed in detail. In Section V, experiment results will be demonstrated. Finally, concluding remarks will be made in Section VI.

II. Dataset and Problem Formulation

In this section, we will first describe the signaling dataset studied in this paper, based on which two semantic time series from the perspective of cells (base stations) can be extracted from the raw data, namely the demand time series and the mobility time series. According to these two semantic time series, the problem formulation of ITW prediction will be discussed.

A. Studied Dataset

The signaling data is a typical example of control-plane data collected from mobile networks [2], which is collected at the mobility management entity of LTE networks. The signaling dataset records every communication/location update event of all active subscribers in a mobile network. Data fields of the signaling data include 1) *subscriber's anonymized identifier*, 2) *time stamp* (e.g., 20160101184312), 3) *location coordinates* (i.e., the longitude and latitude of the base station), 4) *event type*, and 5) *cell type* (i.e., small cell or macro cell). The longitude and latitude coordinates where the base station of each cell is located are accurate to 6 decimal places and timestamps are accurate to seconds. In addition, the signaling data logs event type as well as the direction of the event (e.g., initiating a call or being called). Compared with the commonly used call detail record (CDR) data, the signaling data does not record the duration information of voice services. However, the signaling data further logs two types of location update events in addition to the regular event types (calls or texts), namely the regular location update and the periodic location update. In cellular networks, location updating is a fundamental technique of idle mobile device mobility management. The regular location updates are triggered by tracking area crossing, while the periodic location update is prompted by a timeout event when no event occurs for a subscriber within a predefined time period. In the studied dataset, the time-out interval is about 1 hour, which can guarantee that any active subscriber in the mobile network has at least one observation per hour in the dataset. In this datasets, around three millions of subscribers are recorded. The time period of the studied signaling data utilized in this case study is from Aug 1st, 2016 to Dec 19th, 2016. To understand subscribers' behaviors in urban areas, we extract around 700 macro cells covering the main urban area of a city in China.

B. Demand and Mobility Time Series

According to the event type, user pseudo-ID, timestamp and location information recorded in the studied signaling dataset, one can extract two semantic time series for each cell in the mobile network to understand the aggregated spatiotemporal behaviors of subscribers, namely

- *Demand Time Series.* The demand time series can be extracted by counting the *number of communication events* occurring in a counting time window for a specific cell; that is,

$$d_i = \{\cdots,\ d_{t-l,i},\ d_{t-l+1,i},\ \cdots,\ d_{t-1,i},\ d_{t,i},\ \cdots\};\quad (1)$$

- *Mobility Time Series.* The mobility time series is obtained by counting the *number of unique subscribers* observed in a counting time window for a specific cell; that is,

$$m_i = \{\cdots,\ m_{t-l,i},\ m_{t-l+1,i},\ \cdots,\ m_{t-1,i},\ m_{t,i},\ \cdots\};\quad (2)$$

where subscripts i and t denote the i-th cell and the t-th counting time window (i.e., $[t\Delta, (t+1)\Delta)$), respectively, and Δ denotes the counting time window length ($\Delta = 20$ minutes in this paper). The demand time series, counting voice and text service events of each cell, can directly illustrate the load of mobile networks both spatially and temporally, while the mobility time series can capture the mobility behavior of subscribers in an aggregated manner, shedding lights on the crowd flow of network subscribers. Both mobility and demand time series extracted from the studied signaling dataset can lead to better understanding of mobile network subscribers.

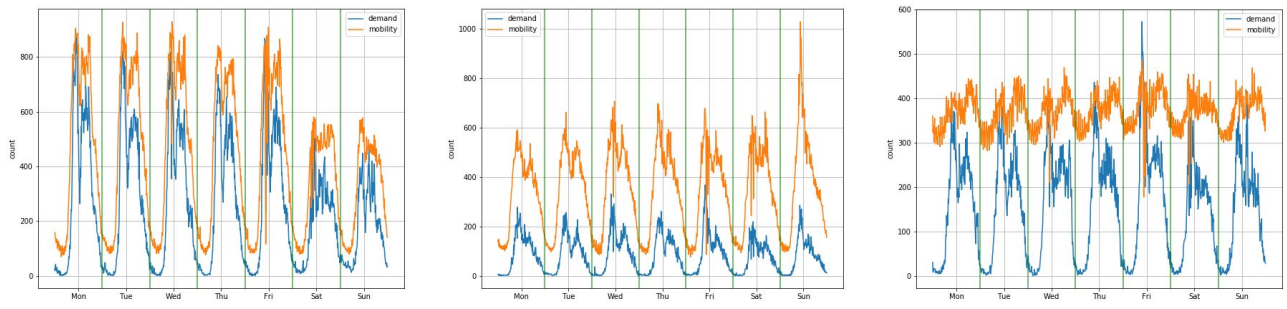
Three demand and mobile time series examples of selected cells located at different typical points of interest are demonstrated in Fig. 1. The cells of business, entertainment, and residence are located at the central business district (CBD), zoo, and residential area, respectively. One can easily observe that both the mobile time series and demand time series are daily periodic in all three examples, while demands and mobility in the business and the entertainment cells behave differently during weekdays and weekends, which may be regarded as weekly periodicity in both time series. In fact, both the demand and the number of visited subscribers (mobility) of the business-type cell clearly decline during weekends, while the number of visited subscribers in the entertainment-type cell increases, compared with the ones during weekdays. In general, demands tend to increase as the number of visited subscribers rises, but the relationship between the two is nonlinear and depends on the cell type and the time within a day. Two peaks can be clearly observed within a day for both the demand and mobile time series, and the demands of cells can drop to zero after midnight. As a result, one can easily find that the ITWs typically occur in the early morning in terms of the loads of each cell as illustrated by demand time series, regardless of the mobility pattern shown by the mobility time series.

C. Idle Time Window Prediction Problem Formulation

In this paper, we propose to predict the ITW in the near future based on the features extracted from the demand and mobility time series in recent history, where the definition of ITWs is given in **Definition 1**.

Definition 1 (Idle Time Window): The idle time window (ITW) of cell i at time t is represented as a tuple, $(S_{t,i}, D_{t,i})$, indicating that a consecutive time period during which demands fall below a predefined threshold ζ_i within the prediction horizon H, i.e.,

$$\begin{aligned} S_{t,i}, D_{t,i} = \underset{S,D}{\text{argmax}} & D \\ \text{s.t.}\quad & d^i_{t+l} < \zeta_i, \quad \forall l \in [S, S+D) \\ & 1 \le S, D \le H \end{aligned}$$



(a) business (b) entertainment (c) residence

Fig. 1. Mobility and demand time series of typical cells, including (a) business, (b) entertainment, and (c) residence.

where S and D denote the start time and duration of the ITW within the time horizon $[t+1, t+H]$ in the future, respectively.

To predict the per-cell ITW, one needs to answer the following questions:

1) Will an ITW present within the prediction horizon, i.e., $[t+1, t+H]$?
2) When will the ITW start (S) and how long (D) will it last?

In fact, the first question is essentially a binary classification or detection problem to determine whether an ITW will present within the future horizon H. The second one can be regarded as a regression problem, which is to estimate the start time S and the duration D for each cell, respectively.

Inspired by the object detection algorithm in the field of computer vision [25], we formulate the ITW prediction as a regression problem with a confidence index to simultaneously account for both the ITW detection and the ITW regression tasks, i.e.,

$$C_{t,i}, S_{t,i}, D_{t,i} = f(X_t, E_t, A) \tag{3}$$

where $S_{t,i} = (S_{t,i} - 1)/H$ and $D_{t,i} = D_{t,i}/H$[1] denote the normalized start time and duration with respect to the horizon H, respectively, when the ITW presents within the future horizon H. In addition, $C_{t,i}$ denotes the confidence index to suggest the confidence that an ITW presents within the future horizon. Hence, $C_{t,i}$, $S_{t,i}$, $D_{t,i} \in [0, 1)$. In (3) and (5), X_t, E_t, and A represent the input features, the exogenous inputs at time t, and the geospatial foreknowledge, respectively. And $f(\cdot)$ represents the predictor or the mapping that is trainable in a supervised manner based on extracted input-output pairs from the raw data.

Furthermore, the presence of ITWs can be determined by the confidence index $C_{t,i}$, i.e.,

$$\begin{cases} \mathcal{H}_0, & C_{t,i} < \tau \\ \mathcal{H}_1, & C_{t,i} \geq \tau \end{cases}, \tag{4}$$

where $\mathcal{H}_1$ and $\mathcal{H}_0$ represent the presence and absence of ITWs, respectively. And τ denotes the threshold for the confidence index to determine the presence of ITWs. In this paper, we also aim to simultaneously predict the ITWs for all cells across the network. As a result, the per-cell ITW prediction (3) could be further rewritten as follows,

$$C_t, S_t, D_t = f(X_t, E_t, A) \tag{5}$$

where $C_t = [C_{t,1}, \cdots, C_{t,N}]^T$, $S_t = [S_{t,1}, \cdots, S_{t,N}]^T$, and $D_t = [D_{t,1}, \cdots, D_{t,N}]^T$. Based on the problem formulation in (5), all components of ITW prediction will be thoroughly discussed in next sections.

D. Combined Cost Function for Model Training

Although the value of all three outputs in our problem formulation ranges between 0 and 1, the meanings underlying these three outputs are different. In fact, the confidence index $C_{t,i}$ serves as a detection statistics to determine the presence of ITWs, where $S_{t,i}$ and $D_{t,i}$ are to estimate where an ITW is located. As a result, the loss of these outputs in training should be specifically designed. Due to the underlying meaning of confidence index, we employ the cross entropy loss for binary classification (or detection) to train the model with respect to the confidence index output, i.e.,

$$\text{closs}(\widehat{C}_{t,i}, \tilde{C}_{t,i}) = \tilde{C}_{t,i} \log(\widehat{C}_{t,i}) + (1 - \tilde{C}_{t,i}) \log(1 - \widehat{C}_{t,i}) \tag{6}$$

where $\widehat{C}_{t,i}$ and $\tilde{C}_{t,i}$ denote estimated confidence index $C_{t,i}$ and its ground truth, respectively. As the true value of $\tilde{C}_{t,i}$ is either 1 or 0, the cross entropy loss function could be reduced to

$$\text{closs}(\widehat{C}_{t,i}, \tilde{C}_{t,i}) = \begin{cases} \log(\widehat{C}_{t,i}) & \tilde{C}_{t,i} = 1 \\ \log(1 - \widehat{C}_{t,i}) & \tilde{C}_{t,i} = 0 \end{cases} \tag{7}$$

With respect to the estimation of ITW start time and duration, the absolute error is employed to evaluate the estimates as follows,

$$\text{bloss}(\widehat{S}_{t,i}, \tilde{S}_{t,i}, \widehat{D}_{t,i}, \tilde{D}_{t,i}) = |\widehat{S}_{t,i} - \tilde{S}_{t,i}| + |\widehat{D}_{t,i} - \tilde{D}_{t,i}| \tag{8}$$

Accordingly, a cost function combining the above two loss functions is employed for model training as follows,

$$\text{cost} = \frac{1}{T \times B} \sum_t \sum_i \Big\{ \text{closs}\left(\widehat{C}_{t,i}, \tilde{C}_{t,i}\right) + \lambda \times \mathbb{1}\left[\text{bloss}\left(\widehat{S}_{t,i}, \tilde{S}_{t,i}, \widehat{D}_{t,i}, \tilde{D}_{t,i}\right)\right] \Big\} \tag{9}$$

where $\mathbb{1}[\cdot]$ is an indicator function to let the cost function only consider the start time and duration estimation when

[1]Without confusion, $S_{t,i}$ and $D_{t,i}$ are employed to denote the start time and duration variable and also their normalized ones.

an ITW presents. In addition, a weight hyperparameter λ is further employed to help the cost function emphasize the model trained on the start time and duration prediction when an ITW exists.

III. Feature and Foreknowledge Engineering

A. Input Features X_t

As ITWs are directly defined based on demand time series as shown in **Definition 1**, the trend of demands in each cell captured by the demand time series should be a key feature for the ITW prediction. In addition, the mobility time series (2), describing the number of subscribers observed by each cell within a counting time window, contains the information of aggregated crowd mobility behavior trend in the network. As a result, a series of demand and mobility observation of each cell should be regarded as features to predict the ITWs, i.e.,

$$X_{t,i}^{\text{dm}} = \begin{bmatrix} d_{t-L+1,i} & d_{t-L+2,i} & \cdots & d_{t,i} \\ m_{t-L+1,i} & m_{t-L+2,i} & \cdots & m_{t,i} \end{bmatrix}, \tag{10}$$

where L denotes the length of recent history considered for ITW prediction. According to the characteristics of the signaling data, the mobility time series can only observe active subscribers with communication demands or location updates. However, the counting time window (20 minutes) is much smaller than the periodical location update interval (60 minutes). Hence, the mobility time series may not be able to capture all subscribers attached to cells within one counting window, as inactive subscribers might stay in the same cell after a location update but unobserved at time t.

In this paper, we propose an innovative feature extraction scheme to largely characterize the aggregated subscriber's mobility in each cell. Let $\mathcal{U}_{t,i}^{1h}$ denote a subscriber set of cell i in a one-hour time window W_t^{1h}, i.e., $W_t^{1h} = [(t-2)\Delta, (t+1)\Delta)$ based on 20-minute counting windows employed in this paper. We then propose to extract the following semantic feature time series based on the subscriber set $\mathcal{U}_{t,i}^{1h}$ and its one-step past $\mathcal{U}_{t-1,i}^{1h}$ as follows,

- *Arriving* $\delta_{t,i}^{m,+}$*:* The new arrival subscriber number is defined as the number of subscribers that are only observed in the counting time window t but not present in its one-step past subscriber set $\mathcal{U}_{t-1,i}^{1h}$, that is

$$\delta_{t,i}^{m,+} = \left|\mathcal{U}_{t,i}^{1h} - \mathcal{U}_{t-1,i}^{1h}\right|.$$

  Thus, the demand $\delta_{t,i}^{d,+}$ generated by the newly arrived subscribers $\mathcal{U}_{t,i}^{1h} - \mathcal{U}_{t-1,i}^{1h}$ can also be extracted in the counting time window $[t\Delta, (t+1)\Delta)$.
- *Staying* $\delta_{t,i}^{m,=}$*:* The subscribers observed in both sets, $\mathcal{U}_{t,i}^{1h}$ and $\mathcal{U}_{t-1,i}^{1h}$, are assumed to be the subscribers staying at cell i in the past one-hour time window, that is

$$\delta_{t,i}^{m,=} = \left|\mathcal{U}_{t,i}^{1h} \cap \mathcal{U}_{t-1,i}^{1h}\right|.$$

- *Departing* $\delta_{t,i}^{m,-}$*:* The subscribers observed only in one-step ahead set $\mathcal{U}_{t-1,i}^{1h}$, but do not appear in current subscriber set $\mathcal{U}_{t,i}^{1h}$, that is

$$\delta_{t,i}^{m,-} = \left|\mathcal{U}_{t-1,i}^{1h} - \mathcal{U}_{t,i}^{1h}\right|.$$

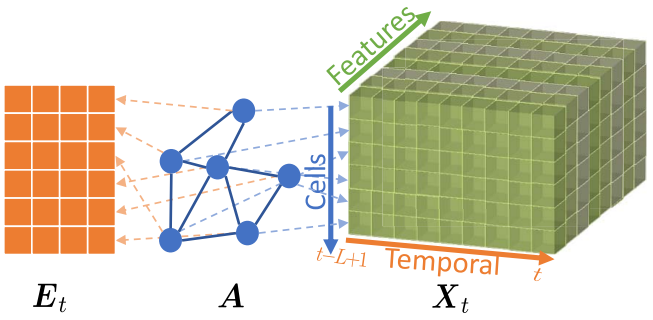


Fig. 2. Features, exogenous inputs and relevancy graph for ITW prediction at time t.

Here, the operation $|\cdot|$ denotes the set cardinality. Clearly, $\delta_{t,i}^{m,+}, \delta_{t,i}^{m,=}, \delta_{t,i}^{m,-}, \delta_{t,i}^{d,+} \geq 0$. Accordingly, each cell could provide multiple features for its ITW prediction based on above operations on subscriber sets as follows,

$$X_{t,i}^{\text{diff}} = \begin{bmatrix} \delta_{t-L+1,i}^{m,+} & \delta_{t-L+2,i}^{m,+} & \cdots & \delta_{t,i}^{m,+} \\ \delta_{t-L+1,i}^{m,=} & \delta_{t-L+2,i}^{m,=} & \cdots & \delta_{t,i}^{m,=} \\ \delta_{t-L+1,i}^{m,-} & \delta_{t-L+2,i}^{m,-} & \cdots & \delta_{t,i}^{m,-} \\ \delta_{t-L+1,i}^{d,+} & \delta_{t-L+2,i}^{d,+} & \cdots & \delta_{t,i}^{d,+} \end{bmatrix}. \tag{11}$$

Similar to our previous work on mobile demand forecasting [19], we also add the one-day ahead and 7-day ahead demand observations as features in order to capture both the daily periodic and weekly periodic effects as observed in Fig. 1, i.e.,

$$X_{t,i}^{\text{period}} = \begin{bmatrix} d_{t-L+1-n_d,i} & d_{t-L+2-n_d,i} & \cdots & d_{t-n_d,i} \\ d_{t-L+1-7n_d,i} & d_{t-L+2-7n_d,i} & \cdots & d_{t-7n_d,i} \end{bmatrix} \tag{12}$$

where n_d denotes the number of observations in one day (i.e., $n_d = 72$). In this paper, we take periodic demands to predict ITWs, while only recent mobility information is considered for ITW prediction.

In summary, the input features for the ITW prediction of cell i at time t can be expressed by stacking $X_{t,i}^{\text{dm}}$, $X_{t,i}^{\text{diff}}$, and $X_{t,i}^{\text{period}}$ as follow,

$$X_{t,i} = \left[(X_{t,i}^{\text{dm}})^T, (X_{t,i}^{\text{period}})^T, (X_{t,i}^{\text{diff}})^T\right]^T. \tag{13}$$

By stacking $X_{t,i}$ of all cells in the network, one can easily obtain a three-dimensional tensor,

$$X_t \in \mathcal{R}^{N \times L \times 8},$$

each axis of which represents cells, temporal sequence, and features, respectively, as shown in Fig. 2. The one-day plot of the input features X_t is shown in Fig. 3, in which subscriber movement can be rarely observed at the early morning from 3 am to 6 am by both $\delta_{t,i}^{m,+}$ and $\delta_{t,i}^{m,-}$. Also, the quantity $(\delta_{t,i}^{m,+} - \delta_{t,i}^{m,-})$ is positive from 6 am to 9 am, meaning that subscribers move into this cell in this interval. The quantity $(\delta_{t,i}^{m,+} - \delta_{t,i}^{m,-})$ is negative from 5 pm to 8 pm, indicating that subscribers move out from this cell during this interval. Hence, time series $\delta_{t,i}^{m,+}$ and $\delta_{t,i}^{m,-}$ can effectively capture the movement of subscribers.

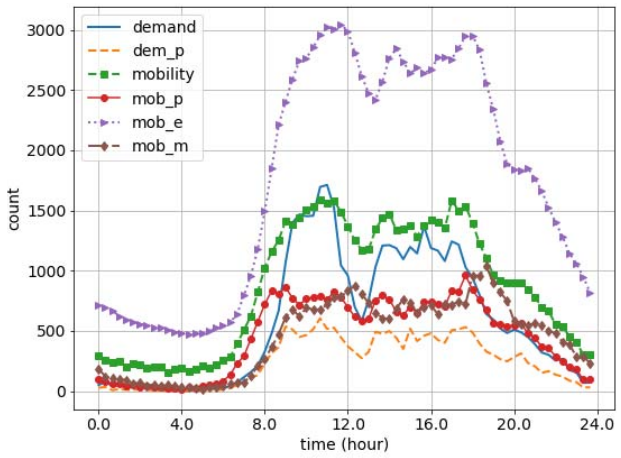


Fig. 3. One-day plot of $X_{t,i}$ at the cell corresponding to Fig. 1(a): demand ($d_{t,i}$), dem_p ($\delta_{t,i}^{d,+}$), mobility ($m_{t,i}$), mob_p ($\delta_{t,i}^{m,+}$), mob_e ($\delta_{t,i}^{m,=}$), mob_m ($\delta_{t,i}^{m,-}$).

B. Exogenous Inputs E_t

As shown in Fig. 1, it can be clearly observed that both demands and mobility are daily periodic and weekly periodic. As a result, both the one-day-ahead and the 7-day ahead ITWs of each cell at the corresponding time point could provide valuable information to guide the ITW predictor to obtain more accurate estimates. In fact, the information of the one-day ahead and 7-day ahead ITWs in each cell can serve as relatively good starting point for ITW estimation, which will be employed as baselines in comparison with our proposed predictors. Hence, the information of the one-day and 7-day ahead ITWs in each cell at time t will be regarded as exogenous inputs to our proposed predictors, i.e.,

$$E_t = [E_{t,1}, E_{t,2}, \ldots, E_{t,N}]^T \in \mathcal{R}^{N\times 4}, \tag{14}$$

where $E_{t,i} = [S_{t-n_d,i}\ D_{t-n_d,i}\ S_{t-7n_d,i}\ D_{t-7n_d,i}]^T$.

C. Geospatial Modeling via Graph A

Based on the spatiotemporal semivariogram analysis of the demand time series across the network as in our previous work [19], it can be concluded as shown in Fig. 4 that the demand relevancy between two cells declines when their spatial distance increases. Hence, we employ the relevancy graph as in [19] to capture the spatial relevancy between cells across the network. The adjacency matrix A of the dependency graph can be obtained based on the spatial distance between cells as follows,

$$A_{ij} = \begin{cases} 1, & \text{dist}(s_i, s_j) \le \eta \\ 0, & \text{otherwise} \end{cases}, \tag{15}$$

where s_i denotes the geolocation of cell i, and η is the distance threshold, that is a hyperparameter that could be tuned. We set $\eta = 2$ km in this paper. In fact, the threshold suggests that any two cells whose distance is beyond the threshold will be considered irrelevant. Such graph modeling could successfully make the cell relevancy sparse (from N^2 to $\sum_{i,j} A_{i,j}$). As a result, each cell could be regarded as a vertex in the spatial dependency graph and the input $X_{t,i}$ is viewed as the signal observed at node i of the graph at time t.

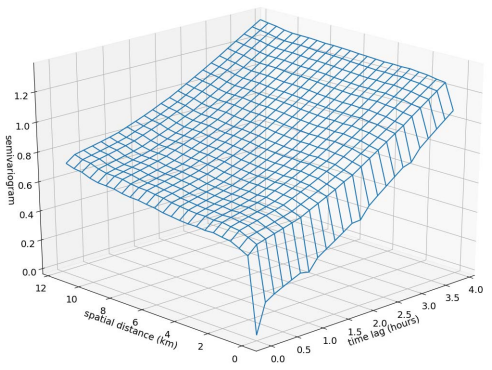


Fig. 4. Spatiotemporal semivariogram analysis on demand time series.

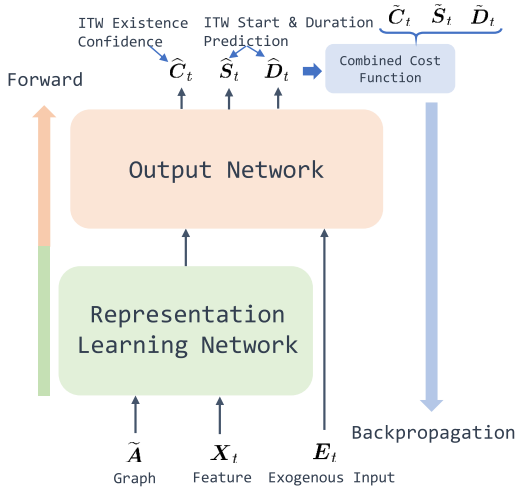


Fig. 5. ITW prediction model.

IV. ITW Prediction Model

To predict the ITW for all cells in the network at each time t, a deep learning based ITW prediction model is proposed to account for the inherited structure of input features X_t and exogenous features E_t as well as the spatial relevancy foreknowledge encoded in the graph adjacency matrix A. As demonstrated previously, the input features take the form of a three-dimensional tensor with both the temporal and spatial structures as shown in Fig. 3. The proposed ITW predition model comprises two main components as shown in Fig. 5, namely

- *Representation Learning Network:* The representation learning network is aimed to learn the high-level representations from the spatiotemporal input tensor X_t with the foreknowledge provided by the relevancy graph A,
- *Output Network:* The output network is responsible to integrate the learned high-level representations (obtained from the representation learning network) and the exogenous inputs E_t to generate the potential ITWs with confidence index (i.e., $[C_t, S_t, D_t]$).

In this paper, we employ the *feedforward neural networks (FNN)* as the output network structure. As for the representation learning network, a *temporal graph convolutional network (TGCN)* is proposed to account for the spatiotemporal structure of the input features X_t and to incorporate the spatial relevancy preknowledge A. In the proposed TGCN, we innovatively integrate the temporal convolutional networks (TCN) [23] and graph convolutional network (GCN) [24] into a spatiotemporal model. It is worth noting that we propose to use exactly the same network (the same network architecture and network parameters) at each cell in the network to predict their respective ITWs, as the powerful prediction model could simultaneously learn the representations of all cells in a mobile network.

A. Representation Learning Network

To learn the high-level representations from the input features X_t, both the sequence and graph structures in X_t need to be sophisticatedly modeled, to prevent the overall prediction model from overfitting. As a result, we will discuss both the sequence and graph modeling for representation learning as follows.

A1. Temporal Modeling

In this paper, we propose to employ the temporal convolutional network (TCN) to model the temporal structure of input X_t in our proposed representation learning network, which has been demonstrated as a good generic temporal (sequence) modeling architecture in [23].

The TCN comprises two key operations, namely the *dilated casual convolution (DC-Conv)* and the *residual connection*, both of which are aimed to deal with the training difficulty issue of very deep networks in different manners and discussed as follows.

Dilated Casual Convolution (DC-Conv): The dilated casual convolution operation takes the form as follows,

$$y(t) = (Z *_d F)(t) = \sum_{i=0}^{k-1} f_i \times z_{t-d*i} \tag{16}$$

where $Z = [z_1, z_2, \cdots, z_L]$ denotes a sequence of temporal signals, each of which z_i is a vector signal. And $F = [f_0, \cdots, f_{k-1}]$ represents a trainable filter with size k, each tap of which f_i is also a vector with the same size as z_i, as shown in Fig. 6(a). Thus, the output of a 1-D convolution is $Y = [y_i, \cdots, y_F]$, where F denotes the number of filters. As illustrated in (16), the DC-Conv operation is different from the conventional 1-D convolution operation in terms of two important concepts, namely causality and dilation:

- *Causality*: The causality is a fundamental requirement for temporal signal processing, which prevents the leakage of future information to the past. In other words, the current signal is completely dependent on its past but not relies on its future.
- *Dilation*: The dilation is to relax the consecutiveness restriction in convolution operations, i.e., $f_i \times z_{t-d*i}$, in the DC-Conv operation as shown in (16), where d denote the dilation factor. That is, the conventional one-dimensional convolution is a special case of dilated convolution with $d = 1$.

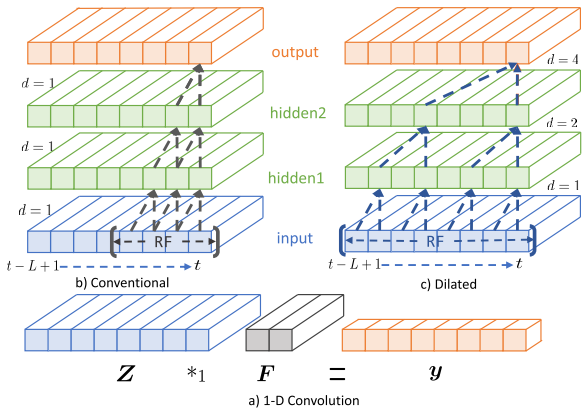


Fig. 6. An example of dilated casual convolution: a) 1-D convolution with $d = 1$ and $k = 2$, b) conventional convolution receptive field (RF) of a three-layer network with $d = 1$ and $k = 2$, c) dilated convolution receptive field (RF) of a three-layer network $d = 2^{j-1}$ and $k = 2$.

The advantage of dilation is to enable a deep network to look back history of inputs much faster than that of the conventional 1-D convolution as shown in Fig. 6 [23], [27], as the dilation factor d is designed to grow exponentially with respect to the depth of the network. Accordingly, the DC-Conv networks with exponential dilation factor could achieve the same receptive field as the conventional 1-D convolution without a very deep network structure.

Residual Connection: The residual connection is to add a path bypassing some layers in a very deep network as shown in Fig. 7, which has become a prominent architecture in deep learning [28]. The mapping enabled by the residual path could be expressed as follows,

$$y = \text{Activation}(\mathcal{M}(x)) = \text{Activation}(\mathcal{F}(x) + x) \tag{17}$$

where $\mathcal{M}(\cdot)$ denotes the underlying mapping, while $\mathcal{F}(\cdot)$ denotes the actual mapping to be learned in training. In (17), it can be observed that the term "residual" originates from the mapping $\mathcal{F}$ actually learning the residual between $\mathcal{M}(x)$ and x, i.e., $\mathcal{F}(x) = \mathcal{M}(x) - x$, rather than the underlying mapping $\mathcal{M}(x)$. It is demonstrated in [29] that the residual connection would lead to the easy training of very deep networks with high accuracy gain, compared with the counterpart without the residual connection. With the residual shortcuts, the entire neural network could be formed in terms of blocks [23], each of which consists of DC-Conv layers and residual connection parallelly coupled as shown in Fig. 7. It is worth noting that two DC convolution layers share the same dilation factor in one TCN block. In addition, zero padding is employed to ensure the same sequence length in both the input and the output of TCN blocks.

A2. Spatiotemporal Modeling:

TCN blocks are employed to account for the temporal underlying structure of the input X_t. However, the spatial relevancy among cells encoded by graph adjacency matrix A is not yet touched so far. In our previous work [19], the graph convolutional networks (GCN) has been demonstrated that it

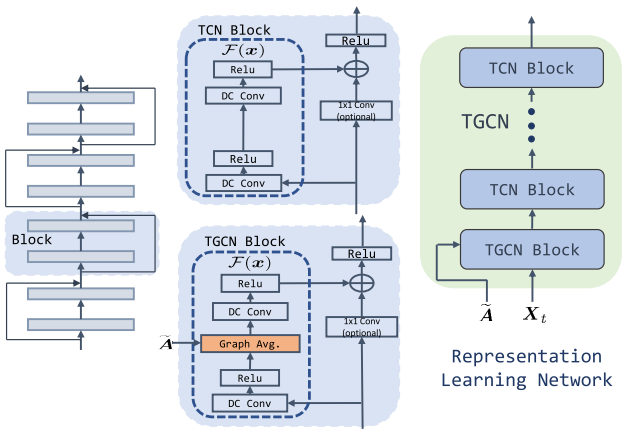


Fig. 7. Representation learning network: temporal graph convolutional network (TGCN).

can successfully capture the spatial relevancy on the task of mobile demand forecasting. In this work, we propose to further employ the concept beneath GCN combined with TCN blocks as discussed previously, to introduce the proposed *TGCN block* that can account for both the temporal and spatial structures of the input.

The approximated graph convolution operation proposed by Kipf and Welling [24] takes the form as follows,

$$y = g_\theta^{(1)}(\widetilde{L}) \star Z = \widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}} Z theta, \tag{18}$$

where $\widetilde{A} = I + A \in \mathcal{R}^{N\times N}$ and $\widetilde{D}$ is a diagonal matrix, $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$. And $Z \in \mathcal{R}^{N\times F}$ is the node-based input matrix, each row of which is the feature vector of each node in the graph, while $theta$ denotes the graph filter. In (18), it can be observed that the approximated graph convolution is essentially to first filter the feature of each node independently (i.e., $z_i theta$, where z_i is a row vector representing the feature vector of node i), and the output of each node by the graph convolution (18) is the average of the filtered results among itself and its neighbors as follow,

$$y_i = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} z_i theta, \tag{19}$$

where $\mathcal{N}_i$ denote the neighbor set of node i including itself. In this paper, this operation is termed as *graph average*.

As a result, we propose to incorporate the graph convolution into the TCN block, to create a model that can simultaneously capture both the temporal and the spatial underlying structure of the input, termed as *temporal graph convolutional networks (TGCN)*. The TGCN block is to add the graph average operation in the midst of two dilated casual convolution layers in TCN blocks as shown in Fig. 7. Let $H^{j-1} \in \mathcal{R}^{N\times L\times F_{j-1}}$ denote the input of j-th TGCN block. The first DC-Conv layer in TGCN blocks will filter the input H^{j-1} along the temporal axis cell-by-cell and generate $H^{j,1} \in \mathcal{R}^{N\times L\times F_j}$ after activation function. The output $H^{j,1}$ will be further inputted to graph average based on the sparsity information provided by the graph adjacency $\widetilde{A}$ as follows,

$$H_i^{j,2} = \frac{1}{|\mathcal{N}_i|}\sum_{n\in\mathcal{N}_i} H_n^{j,1}, \tag{20}$$

TABLE I
NOTATION

Notation	Description
H	Prediction horizon
L	Length of history
$S_{t,i}$, $\hat{S}_{t,i}$, $\tilde{S}_{t,i}$	Start time of ITW of cell i at time t
$D_{t,i}$, $\hat{D}_{t,i}$, $\tilde{D}_{t,i}$	Duration of ITW of cell i at time t
$\boldsymbol{X}_t \in \mathcal{R}^{N\times L\times 8}$	Input features
$d_{t,i}$, $m_{t,i}$	demand and mobility observation of cell i at time t
$\delta_{t,i}^{m,+}$	New arrival subscriber num. of cell i at time t
$\delta_{t,i}^{m,=}$	Keep staying subscriber num. of cell i at time t
$\delta_{t,i}^{m,=}$	Departed subscriber num. of cell i at time t
$\delta_{t,i}^{d,+}$	demand of new arrival subscriber of cell i at time t
$\boldsymbol{A}$	Adjacency matrix of relevancy graph
$\boldsymbol{E}_t$	Exogenous input at time$t$
$\boldsymbol{F}$	Filters for DC convolution
λ	Weight for start and duration estimation in cost function
IoU	Intersection over union

where $H_i^{j,2} \in \mathcal{R}^{L\times F_j}$ denotes the output of node i after graph average operation and tensor $H^{j,2} \in \mathcal{R}^{N\times L\times F_j}$ can be obtained by stacking $H_i^{j,2}$ across all the nodes in the graph. And $H_i^{j,2}$ will be further fed to the second DC-Conv layer and then combined with the result via the residual connection to generate the final output of TGCN H^j. Details of TGCN is shown in Fig. 7.

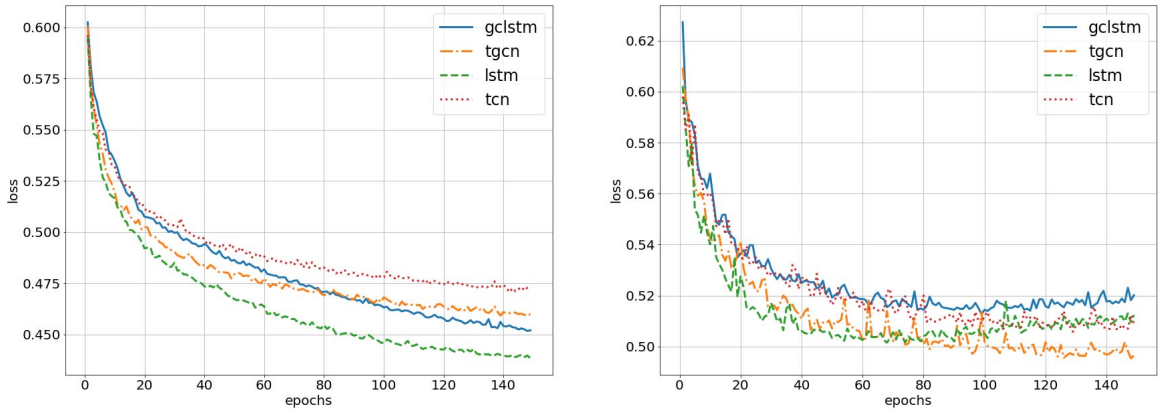
B. Prediction Model Assembly

As observed in Fig. 4, both demand and mobility time series are more relevant to its own history than the one from its neighbors. As a result, the proposed representation learning network first emphasizes the temporal relevancy by employing three TCN blocks to learn the high-level representation from inputs as shown in Fig 7. In addition, one TGCN block is introduced as the first block in the representation network to embed the graph information in the proposed prediction model so that the spatial relevancy among cells could be accounted for by the proposed ITW prediction model.

The output network is a two-layer full-connection forward neural networks, in which the rectifier (ReLu) function is employed as the activation function in the input and hidden layers. In addition, we employ sigmoid function as activation functions in the final layer, as the value of the desired outputs (C_t, S_t, and D_t) is bounded between zero and one. In addition, it is worth noting that the output of representation learning network will be flattened cell-by-cell (i.e., the output will be formatted from $\mathcal{R}^{N\times L\times F_4}$ into $\mathcal{R}^{N\times F_f}$, where $F_f = L\times F_4$). And the output layer will generate the final result also in a cell-by-cell manner. Details of the proposed predictive model architecture are illustrated by Fig. 7.

V. EXPERIMENTS RESULTS

In this section, we validate the proposed problem formulation on cell ITW prediction and also compare the proposed temporal graph convolutional network (TGCN) with other sequence and sequence-graph models, namely long short-term memory (LSTM) [30], temporal convolutional networks (TCN) [23], and graph-convolutional LSTM



(a) training loss (b) validation loss

Fig. 8. (a) Training and (b) validation comparisons in terms of designed cost and epochs.

TABLE II
TGCN, TCN, LSTM, AND GCLSTM SPECIFICATIONS

	TGCN	TCN	LSTM	GCLSTM
Input	$\boldsymbol{X}_t, \boldsymbol{E}_t, \boldsymbol{A}$	$\boldsymbol{X}_t, \boldsymbol{E}_t$	$\boldsymbol{X}_t, \boldsymbol{E}_t$	$\boldsymbol{X}_t, \boldsymbol{E}_t, \boldsymbol{A}$
RL Networks	$[128, 128, 8]$	$[128, 128, 8]$	$[90, 90]$	$[90, 90]$
Kernel Size	2	2	N/A	N/A
Trainable Para.	$106,304$	$106,304$	$103,152$	$103,152$

(GCLSTM) [19]. The GCLSTM model in our previous work is an intuitive extension of the classic convolutional LSTM (convLSTM) [31] from grid-like spatiotemporal data to graph-based spatiotemporal data. The output network among compared models, including TGCN, TCN, LSTM, and GCLSTM, is kept the same as shown in Fig. 5 (one 32-unit hidden layer), while the specifications of their representation learning networks are detailed in Table II. It is easily observed in Table II that the total trainable parameters are designed to be similar for fair comparisons. In addition, two baselines are also employed to validate the problem formulation and the performance of the proposed predictive model, namely the *baseline-Y* and *baseline-W*, whose performances are shown in Table III. Specifically, the baseline-Y is to directly use the time window at the same time but one-day before as the ITW estimate, while baseline-W is to use the time window of one-week before. These two baselines are chosen based on the fact that both the demand and mobility time series are daily periodic and weekly periodic.

A. Model Training

All experiments in this paper are carried out by the PyTorch deep learning framework [32]. In all experiments, the training-validation data are cleaned data extracted from the previously discussed data starting from Aug. 1st, 2016 to Nov. 30th, 2016, while the test data is extracted from Dec. 4th to Dec. 19th. The training and validation datasets are uniform-randomly selected from the entire train-validation data with the 95% and 5% of total samples, respectively. The length of history for ITW prediction is 6 or equivalently 2 hours, while the future horizon H is 18 or equivalently 6 hours. In this paper, we employ the combine cost function as stated in (9) to train all the compared models. The weight λ for ITW estimation in the cost function (9) is set to be 5.

Before training, all the features and inputs are first normalized by their mean and standard deviation. The dropout technique [33] is employed as the regularization during training the proposed predictive model. The optimization method utilized for training is Adam [34], which generally has a relatively better performance compared with the commonly used stochastic gradient descent (SGD) algorithm. In addition, weight normalization [35] is added to every DC-Conv layer in the TCN and TGCN to expedite the training speed as in [23]. Fig. 8 shows the loss of both the training dataset and validation dataset during training versus epochs, which suggests that the prediction model converge easily. It could be observed that the LSTM and GCLSTM could easily outperform the convolution-based models (TCN and TGCN) in terms of training loss, however, the validation loss suggests that the lower training loss does not necessarily lead to a good generalization. In fact, the LSTM and GCLSTM easily overfit starting around the 60th and 100th epoch, respectively.

B. Testing: Performance Evaluation

In this paper, two tasks are simultaneously fulfilled by our proposed ITW prediction model, namely ITW detection and ITW estimation. As a result, both the ITW detection and ITW estimation performance will be evaluated differently as follows.

ITW Detection The receiver operating characteristics (RoC) and precision-recall curves are typically employed to assess the performance of detection or binary classification problems. The RoC curve consists of two core metrics, namely detection and false alarm. In fact, the detection results could be categorized into the following 4 types (as shown in the table below), based on which the evaluation metrics (detection/recall, false

TABLE III
ITW PREDICTION TEST RESULTS

			GCLSTM	TGCN	LSTM	TCN	baseline-Y	baseline-W
ITW Detection	Recall (%)		93.5899	94.1442	93.7525	**94.1974**	92.0651	92.0906
	Precision (%)		94.1415	**94.5600**	94.5171	94.3749	93.0336	92.5659
	F1 Score (%)		93.8649	**94.3516**	94.1332	94.2861	92.5468	92.3276
	False Alarm (%)		5.6931	**5.2942**	5.3162	5.4881	6.7387	7.2295
ITW Estimation	ITW Presence	IoU (%)	78.3439	**79.8565**	79.1544	79.1415	69.3181	68.8176
		Start Accuracy (%)	58.5265	**59.3068**	58.5818	58.8425	52.1011	50.6305
		Dura Accuracy (%)	27.0026	**30.6853**	29.9848	30.3642	26.5762	25.6742
		Start MAE	0.0589	**0.0541**	0.0577	0.0558	0.0991	0.1005
		Dura MAE	0.0924	**0.0840**	0.0886	0.0861	0.1080	0.1113
	ITW Detected	IoU (%)	73.1486	**74.9044**	74.3670	74.2978	70.6629	69.6650
		Start Accuracy (%)	55.2978	**56.4903**	56.1036	56.1707	52.6492	50.8917
		Dura Accuracy (%)	23.6972	**27.6653**	27.3187	27.5659	26.8558	25.8067
		Start MAE	0.0851	**0.0775**	0.0819	0.0792	0.0928	0.0970
		Dura MAE	0.1060	**0.0949**	0.0992	0.0975	0.1076	0.1126
	ITW Correctly Detected	IoU (%)	79.3999	**80.8638**	80.4468	80.3472	77.8243	77.1938
		Start Accuracy (%)	58.7390	**59.7402**	59.3582	59.5186	56.5916	54.9790
		Dura Accuracy (%)	25.1150	**29.2438**	28.8446	29.2013	28.8667	27.8793
		Start MAE	0.0576	**0.0527**	0.0558	0.0542	0.0664	0.0684
		Dura MAE	0.0946	**0.0858**	0.0901	0.0878	0.0997	0.1032

alarm, and precision) are defined as follows,

$$\text{Detection/Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}},$$
$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}},$$
$$\text{False Alarm} = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}. \quad (21)$$

The F1 score is essentially the harmonic average of the recall and precision as follows,

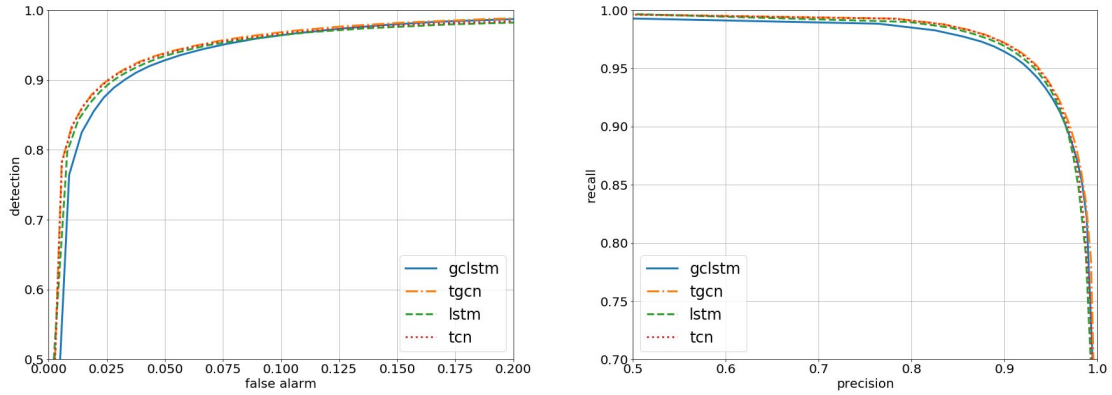
$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (22)$$

Based on our proposed prediction model, one can obtain the RoC and precision-recall curves by adjusting the confidence threshold τ in the decision rule as stated in (4). Fig. 9 shows the RoC and precision-recall performance of all the compared models. It could be observed that all the models could well detect ITW in future horizons in terms of both RoC and precision-recall metrics, which verifies the validness of the proposed problem formulation, feature engineering, and prediction model structure. In addition, the proposed TGCN can outperform others from the perspective of both the RoC and precision-recall.

ITW Estimation

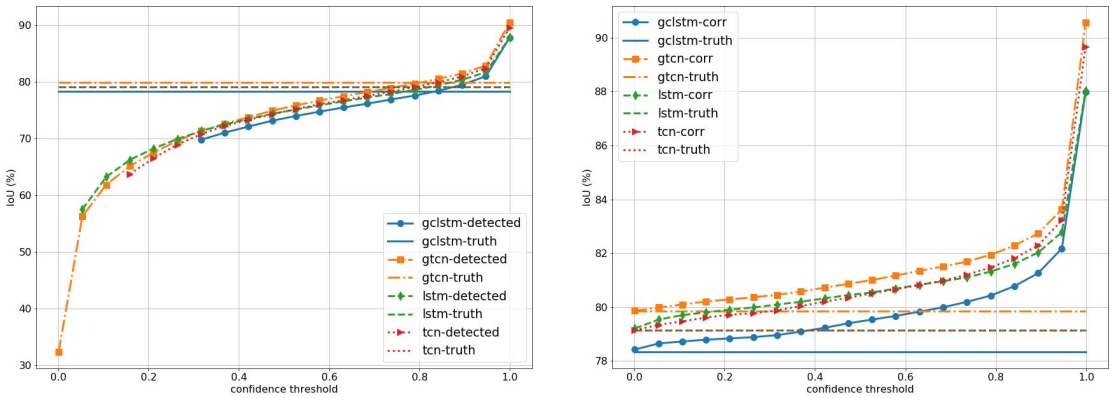
As ITW start time and duration estimation are only meaningful when an ITW is predicted to appear, ITW estimation performance should be evaluated in terms of ITW detection. Hence, the ITW estimation performance will be evaluated in terms of three cases, namely *ITW Presence*, *ITW Detected*, and *ITW Correctly Detected*. The ITW Presence case is to assess the model on all the samples that ITW truly presents, regardless of whether the prediction model can detect the ITW, which is aimed to evaluate the overall performance that how a prediction model can estimate ITW. The ITW detected case is to test the model on the samples that a prediction model claims ITW presence in the future horizon H, aimed to evaluate a prediction model in practice, while the ITW correctly detected is to assess a model on the samples that correctly identified. As the detection of all the models is not perfect, false alarms will appear in the ITW detected cases, which will be penalized when calculating IoU detailed later. In these three cases, the employed evaluation metrics on ITW estimation will be discussed as follows.

- *Accuracy.* As both the state time and duration of ITWs as defined in **Definition 1** are discretized, the accuracy is to assess how many start time and duration can be exactly predicted by each predictor as follows,
$$\text{sacc.} = \frac{\#(\tilde{S} = \hat{S})}{\text{total } \#} \text{ or } \text{dacc.} = \frac{\#(\tilde{D} = \hat{D})}{\text{total } \#} \quad (23)$$
- *Error.* The absoluble error between ground truth and a predict is also employed as assessment metrics, i.e., $|\hat{S} - \tilde{S}|$ or $|\hat{D}X - \tilde{D}|$. To analysis the prediction error, we will show mean absolute error (MAE) of both the start time and the duration estimates.
- *Intersection over Union (IoU).* The previous two metrics only assess the start time and duration prediction independently, but not the quality of the overall ITW estimation. In this paper, we borrow the intersection-over-union (IoU)



(a) RoC (b) Precision-Recall

Fig. 9. (a) RoC and (b) precision-recall comparison.



(a) ITW Detected: IoU (b) ITW Correctly Detected: IoU

Fig. 10. IoU comparison, where line legend "-detected", "-truth", and "corr" denote the ITW detected case, ITW presence case, ITW correctly detected case, respectively. (a) ITW detected: IoU. (b) ITW correctly detected: IoU.

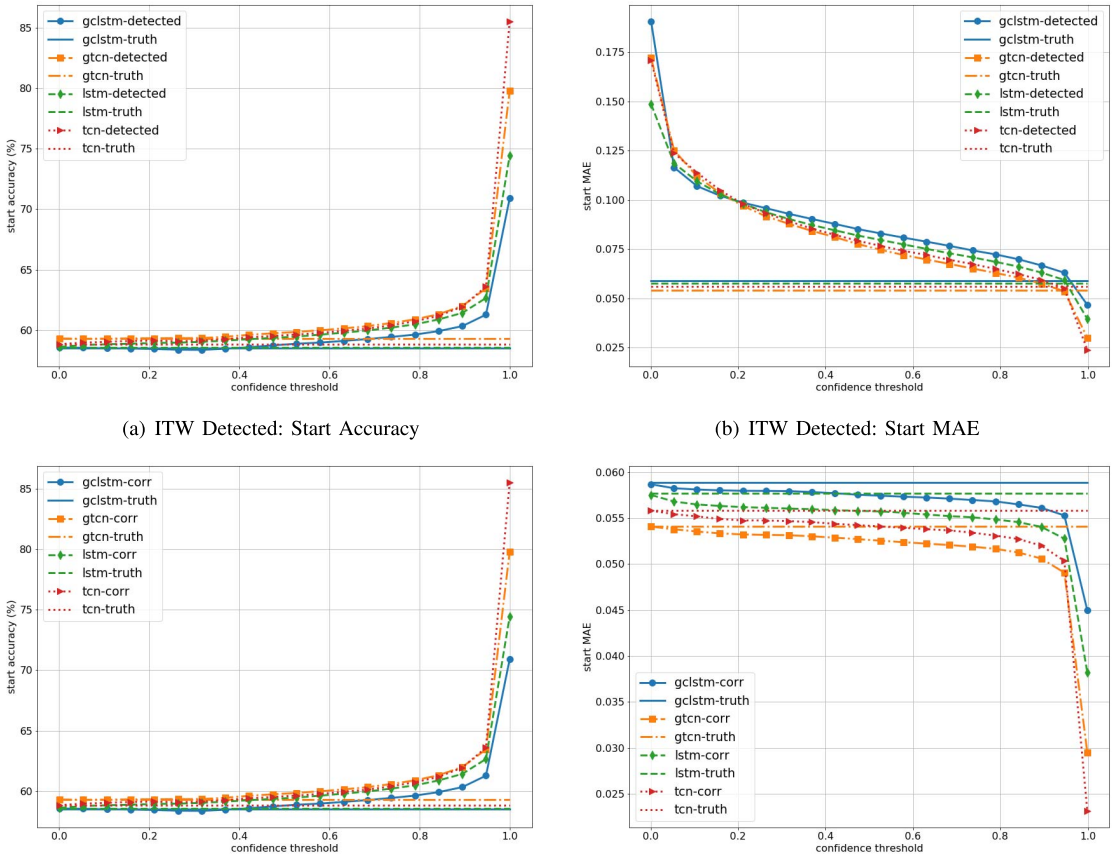
metric from the object detection task from the field of computer vision [25] to assess how well the predicted time window overlaps with the ground truth as follows,

$$\begin{aligned} \text{Intersection} &= \min\{\hat{S}+\hat{D}, \tilde{S}+\tilde{D}\} - \max\{\hat{S}, \tilde{S}\} \\ \text{Union} &= \max\{\hat{S}+\hat{D}, \tilde{S}+\tilde{D}\} - \min\{\hat{S}, \tilde{S}\} \\ \text{IoU} &= \text{Intersection}/\text{Union} \end{aligned} \tag{24}$$

Among the above three evaluation metrics, the IoU shall be the most important one, as it directly reflects how a predictor performs in terms of window estimation. As for false alarms, the IoU will be directly set to be zero as the penalty, since the IoU metric when ITW is absent is meaningless.

Fig. 10 shows the IoU comparisons in three cases discussed previously in terms of confidence threshold τ. It is intuitive that the IoU of all compared prediction models does not vary with the confidence threshold, which is also shown as a horizontal line in the figure. In fact, the tradeoff between precision and recall could be demonstrated by adjusting the confidence threshold. That is, the high confidence threshold suggests the high precision performance but relatively low recall performance and vice versa. As a result, one can clearly observe that the IoU in both the ITW detected and ITW correctly detected cases can grow with the increase of confidence threshold. Such phenomenon demonstrates the flexibility of our proposed prediction model facilitated by the designed confidence index. The IoU can reach 90% even in the ITW detected case when the confidence threshold is high. Compared with other representation learning models, our proposed TGCN model is the best in terms of IoU.

Fig. 11 illustrates the ITW estimation performance in terms of start time and duration accuracy and MAE. It can be observed that the accuracy and MAE have the similar pattern as IoU performance for each compared prediction model. In Table III, all the metrics discussed previously of all compared prediction models are compared with baselines. The confidence threshold for each prediction model is selected based on its optimal F1 score in Table III. It can be observed that the proposed prediction model can have about 10% IoU improvement compared with two baselines employed (as shown by IoU in ITW presence case). By relaxing the detection performance to be the same level as the baselines via adjusting the confidence threshold, the IoU performance can be further improved according to the relationship between IoU and confidence threshold as shown in Fig. 10.



(a) ITW Detected: Start Accuracy

(b) ITW Detected: Start MAE

(c) ITW Correctly Detected: Start Accuracy

(d) ITW Correctly Detected: Start MAE

Fig. 11. Accuracy and MAE comparisons of ITW estimation, where line legend "-detected", "-truth", and "corr" denote the ITW detected case, ITW presence case, ITW correctly detected case, respectively. (a) ITW detected: start accuracy. (b) ITW detected: start MAE. (c) ITW correctly detected: start accuracy. (d) ITW correctly detected: start MAE.

C. Discussion

The experiment results discussed previously have shown the validness and effectiveness of our proposed feature engineering scheme, ITW prediction model network structure, and temporal-graph convolutional networks. In fact, ITW can be well predicted by one-day or one-week ahead observations at the same time, due to the strong seasonality exhibited in mobile demand time series across the network. In this paper, our proposed prediction model can effectively learn the pattern from recent history via representation learning network and project a better ITW detection and estimation by taking the periodic observations as exogenous inputs.

In the proposed ITW prediction model, the ITW existence confidence index plays an important role, as it not only indicates the confidence of ITW presence during the model inference, but also helps eliminate the negative impact of ITW non-presence samples on the ITW estimation performance during the model training. In fact, such ITW estimation performance enhancement during model training is facilitated by the indicator function for ITW presence in the designed cost function (9) for model training as well as its tunable weight hyperparameter λ emphasizing ITW starting time and duration estimation. In addition, the confidence index threshold can be further employed to control the tradeoff between the recall performance and the IoU performance (or the precision performance) of the predictor as shown in Figs. 10 and 11. In such a manner, our proposed model is flexible and can fulfill different robustness requirements in practice.

As for representation learning models, both the TCN and LSTM can effectively learn useful patterns for ITW prediction as demonstrated by the experiment results discussed previously. In addition, the spatial modeling by graph averaging employed in TGCN can further improve the ITW prediction, compared with the one by TCN and LSTM. However, due to the strong individual temporal relevancy of mobile demand time series—semivariogram is much smaller at 0 spatial distance as shown in Fig. 4—the overwhelming spatial modeling may lead to a deteriorated prediction performance. This is the reason why the GCLSTM has a worse performance compared with LSTM, where GCLSTM employs the graph averaging operation (19) in both two layers. In fact, TCGN employs one TGCN block (Fig. 7) involving the graph averaging operation (19) with other TCN blocks for temporal modeling, which could effectively capture both the spatial and temporal characteristics. Since the TGCN is similar to TCN yet with one additional graph averaging operation, the TGCN inherits the advantages and limitation of TCN. The advantages and disadvantages of TCN compared with LSTM has been

thoroughly discussed in [23] in terms of model training and inference. Overall, the proposed TGCN model demonstrates its superiority by experiment results by properly capturing both the spatial and temporal patterns.

VI. Conclusions

In this paper, we proposed to directly predict the idle time window based on subscribers' demand and mobility in mobile networks. A novel feature extraction scheme has been discussed to capture current trends of demand and mobility as well as the exogenous inputs accounting for the periodicity inherited in subscribers' demands. By modeling the spatial relevancy among cells as a graph, an ITW prediction model consisting of the representation learning network and the output network has been proposed, in which the temporal graph convolutional networks (TGCN) was further proposed to learn the high-level spatiotemporal patterns for the ITW prediction. Experiment results validated the effectiveness of the proposed idle time window prediction formulation and demonstrated the superiority of the proposed TGCN.

References

[1] X. Cheng, L. Fang, X. Hong, and L. Yang, "Exploiting mobile big data: Sources, features, and applications," *IEEE Netw.*, vol. 31, no. 1, pp. 72–79, Jan./Feb. 2017.

[2] X. Cheng, L. Fang, L. Yang, and S. Cui, "Mobile big data: The fuel for data-driven wireless," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1489–1516, Oct. 2017.

[3] X. Cheng, L. Fang, L. Yang, and S. Cui, *Mobile Big Data*. Cham, Switzerland: Springer, 2018.

[4] E. J. Kitindi, S. Fu, Y. Jia, A. Kabir, and Y. Wang, "Wireless network virtualization with SDN and C-RAN for 5G networks: Requirements, opportunities, and challenges," *IEEE Access*, vol. 5, pp. 19099–19115, 2017.

[5] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, Aug. 2017.

[6] X. Cheng, L. Fang, and L. Yang, "Mobile big data based network intelligence," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4365–4379, Dec. 2018.

[7] K. Wang, Y. Wang, D. Zeng, and S. Guo, "An SDN-based architecture for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 25–31, Feb. 2017.

[8] F. Malandrino, C.-F. Chiasserini, and S. Kirkpatrick, "Understanding the present and future of cellular networks through crowdsourced traces," in *Proc. 18th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Macau, China, Jun. 2017, pp. 1–9.

[9] M. Peng, D. Liang, Y. Wei, J. Li, and H.-H. Chen, "Self-configuration and self-optimization in LTE-advanced heterogeneous networks," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 36–45, May 2013.

[10] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 336–361, 1st Quart., 2013.

[11] M. Ismail, W. Zhuang, E. Serpedin, and K. Qaraqe, "A survey on green mobile networking: From the perspectives of network operators and mobile users," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1535–1556, 3rd Quart., 2014.

[12] H. Ghazzai, M. J. Farooq, A. Alsharoa, E. Yaacoub, A. Kadri, and M. S. Alouini, "Green networking in cellular hetnets: A unified radio resource management framework with base station on/off switching," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5879–5893, Jul. 2017.

[13] S. Zhang *et al.*, "Traffic prediction based power saving in cellular networks: A machine learning method," in *Proc. 25th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst.*, Redondo Beach, CA, USA, 2017, pp. 29-1–29-10.

[14] H. Shi and Y. Li, "Discovering periodic patterns for large scale mobile traffic data: Method and applications," *IEEE Trans. Mobile Comput.*, vol. 17, no. 10, pp. 2266–2278, Oct. 2018.

[15] J. Ding, R. Xu, Y. Li, P. Hui, and D. Jin, "Measurement-driven modeling for connection density and traffic distribution in large-scale urban mobile networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1105–1118, May 2018.

[16] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "DNN-based prediction model for spatio-temporal data," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst.*, Burlingame, CA, USA, Oct./Nov. 2016, pp. 92-1–92-4.

[17] F. Xu *et al.*, "Big data driven mobile traffic understanding and forecasting: A time series approach," *IEEE Trans. Services Comput.*, vol. 9, no. 5, pp. 796–805, Sep./Oct. 2016.

[18] J. Wang *et al.*, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Atlanta, GA, USA, May 2017, pp. 1–9.

[19] L. Fang, X. Cheng, H. Wang, and L. Yang, "Mobile demand forecasting via deep graph-sequence spatiotemporal modeling in cellular networks," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3091–3101, Aug. 2018.

[20] C. Zhang and P. Patras, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Los Angeles, CA, USA, 2018, pp. 231–240.

[21] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "Geoman: Multi-level attention networks for geo-sensory time series prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, Jul. 2018, pp. 3428–3434. doi: 10.24963/ijcai.2018/476.

[22] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, Apr./May 2018, pp. 1–16. [Online]. Available: https://openreview.net/forum?id=SJiHXGWAZ

[23] S. Bai, J. Z. Kolter, and V. Koltun. (2018). "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." [Online]. Available: https://arxiv.org/abs/1803.01271

[24] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Paris, France, Apr. 2017, pp. 1–14.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, May 2016, pp. 1–13.

[28] S. Jastrzebski, D. Arpit, N. Ballas, V. Verma, T. Che, and Y. Bengio, "Residual connections encourage iterative inference," in *Proc. Int. Conf. Learn. Represent.*, Vancouver, BC, Canada, Apr./May 2018, pp. 1–14. [Online]. Available: https://openreview.net/forum?id=SJa9iHgAZ

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[31] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. New York, NY, USA: Curran Associates, 2015, pp. 802–810. [Online]. Available: http://papers.nips.cc/paper/5955-convolutional-lstm-network-a-machine-learning-approach-for-precipitation-nowcasting.pdf

[32] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. NIPS Autodiff Workshop*, Long Beach, CA, USA, Dec. 2017, pp. 1–4.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

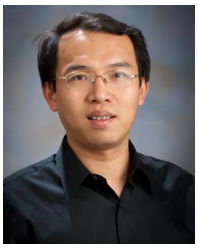
[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–15.

[35] T. Salimans and D. P. Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 901–909.

Luoyang Fang (S'12) received the B.S. degree from the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2011. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Colorado State University. His research interests include big data, mobile data, location privacy, data mining, distributed storage systems, and information-centric networking.

Xiang Cheng (S'05–M'10–SM'13) received the Ph.D. degree from Heriot-Watt University and The University of Edinburgh, Edinburgh, U.K., in 2009, where he received the Post-Graduate Research Thesis Prize.

He is currently a Professor with Peking University. His general research interests are in the areas of channel modeling and mobile communications, subject on which he has published more than 160 journal and conference papers, three books, and six patents. He was a recipient of the IEEE Asia Pacific (AP) Outstanding Young Researcher Award in 2015 and a co-recipient of the 2016 IEEE JSAC Best Paper Award: Leonard G. Abraham Prize, the NSFC Outstanding Young Investigator Award, the First-Rank and Second-Rand Awards in Natural Science, Ministry of Education in China, and the Best Paper Awards at IEEE ITST 2012, ICCC 2013, ITSC 2014, ICC 2016, and ICNC 2017. He has served as a Symposium Leading-Chair, a Co-Chair, and a member of the Technical Program Committee for several international conferences. He is currently an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and the *Journal of Communications and Information Networks*.



Haonan Wang received the Ph.D. degree in statistics from the University of North Carolina, Chapel Hill, NC, USA, in 2003. He is currently a Professor of statistics with Colorado State University, Fort Collins, CO, USA. His research interests include object-oriented data analysis, statistical analysis on tree-structured objects, functional dynamic modeling of neuron activities, and spatiotemporal modeling.

Liuqing Yang (S'02–M'04–SM'06–F'15) received the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA, in 2004.

Her main research interests include communications and signal processing. She received the Office of Naval Research Young Investigator Program Award in 2007, the National Science Foundation Career Award in 2009, the IEEE GLOBECOM Outstanding Service Award in 2010, the George T. Abell Outstanding Mid-Career Faculty Award, and the Art Corey Outstanding International Contributions Award from CSU in 2012 and 2016, respectively, and Best Paper Awards at IEEE ICUWB 2006, ICCC 2013, ITSC 2014, GLOBECOM 2014, ICC 2016, and WCSP 2016. She has been actively serving in the technical community, including the organization of many IEEE international conferences, and on the editorial boards of a number of journals, including the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. She is currently serving as the Editor-in-Chief for *IET Communications*.