Information Fusion 59 (2020) 1–12

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Urban flow prediction from spatiotemporal data using machine learning: A survey

Peng Xie[a,b,c], Tianrui Li[a,b,c,*], Jia Liu[a,b,c], Shengdong Du[a,b,c], Xin Yang[d], Junbo Zhang[e,f,b]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China
[b] Institute of Artificial Intelligence, Southwest Jiaotong University, Chengdu, China
[c] National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu, China
[d] School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, China
[e] JD Intelligent Cities Business Unit, JD Digits, Beijing, China
[f] JD Intelligent Cities Research, Beijing, China

A R T I C L E I N F O

Keywords:
Urban flow prediction
Spatiotemporal data mining
Data fusion
Deep learning
Urban computing

A B S T R A C T

Urban spatiotemporal flow prediction is of great importance to traffic management, land use, public safety. This prediction task is affected by several complex and dynamic factors, such as patterns of human activities, weather, events, and holidays. Datasets evaluated the flow come from various sources in different domains, e.g. mobile phone data, taxi trajectories data, metro/bus swiping data, bike-sharing data. To summarize these methodologies of urban flow prediction, in this paper, we first introduced four main factors affecting urban flow. Second, in order to further analyze urban flow, we partitioned the preparation process of multi-source spatiotemporal data related with urban flow into three groups. Third, we chose the spatiotemporal dynamic data as a case study for the urban flow prediction task. Fourth, we analyzed and compared some representative flow prediction methods in detail, classifying them into five categories: statistics-based, traditional machine learning-based, deep learning-based, reinforcement learning-based, and transfer learning-based methods. Finally, we showed open challenges of urban flow prediction and discussed many recent research works on urban flow prediction. This paper will facilitate researchers to find suitable methods and public datasets for addressing urban spatiotemporal flow forecast problems.

1. Introduction

Forecasting urban flow is strategically important for traffic management [1], land use [2], public safety [3], etc. For city managers, they can pre-discover the traffic congestion that may occur in the city, deploy traffic in advance, and ease traffic congestion. For businessmen, they can find the crowded regions or potential business investment areas to gain greater business benefits. For public, they can improve their own travel plans in advance, stagger the peak of travel, and choose a more convenient way to travel. From the perspective of people's travel mode, urban flow contain crowd flow [4–6], traffic flow [7–9] and public transit flow [10–12], etc.

However, urban flow prediction is not an easy issue [5,11–13]. First, there are some main factors affecting urban flow, which can be classified into four groups. 1. *Daily flow activity patterns:* They are the main patterns of urban flow, including working day commute, going to and back from school and other daily repeated activities. 2. *Anomalies of flow activity patterns:* Although daily flow activity patterns are main patterns of urban flow, our mostly concern is anomalies of flow activity patterns and certain areas, e.g. an increase of urban traffic anomaly, since this may lead to the phenomena of traffic congestion, social security, etc. For this kind of phenomenon, we can improve traffic deployment to make people travel more convenient and enhance security emergency to make social order more harmonious. Some urban events or activities will also affect the flow in the city. When temporary traffic control is imposed on an area because of the construction of roads, there will be a corresponding decrease in flow in that area. 3. *Weather:* It also has a certain impact on urban flow, e.g. heavy rain, smog and other extreme weather conditions, which may cause the number of people going out to decrease, while the number of people going out may increase when the weather is sunny. 4. *Holidays:* For holidays, e.g. National Day holiday and Spring Festival holiday, there may be cross-regional and surging flow, which is periodic in years. In terms of time, it also has a certain impact on the urban flow on the date near the holiday, making the fluctuation of crowd flow last for a period of time.

Datasets we can obtain are almost spatiotemporal data, e.g. mobile phone data, taxi trajectories data, metro/bus swiping data, all of which

* Corresponding author at: Mendeley Chengdu 611756 China.
E-mail address: trli@swjtu.edu.cn (T. Li).

https://doi.org/10.1016/j.inffus.2020.01.002
Received 22 April 2019; Received in revised form 30 November 2019; Accepted 3 January 2020
Available online 10 January 2020

Table 1
Types of spatiotemporal datasets.

Data type	Spatiotemporal static	Spatial static temporal dynamic	Spatiotemporal dynamic
Point data	POI and Events	Air quantity data and Road monitoring data	User check-in data and Phone signal data
Network data	Road structure data	Traffic flow data and Crowd flow data	Taxi trajectory data and Public transit data

have temporal dependence and spatial correlation. How to deal with these spatiotemporal data is a challenging issue [14]. We will further elaborate it and review widely used models for urban flow prediction problems in Section 4.

In addition, predicting urban flow requires determining the granularity of forecasting [3,11,15–17], in terms of space, e.g. the entire city, regional, and street-wide level; in terms of time, e.g. the next 15 minutes, the next hour, the next 24 hours of urban flow in each regions. Different prediction granularity requires different precision and different processing methods. Apart from the objective factors mentioned above, there are also some unmeasured subjective travel intentions of the population, which are challenging to study and haven't any breakthrough been achieved.

The contribution of this paper lies in three folds. First, urban flow prediction from spatiotemporal data is systematically reviewed. Second, we divide the typical and representative methods into five categories for urban flow prediction and mainly analysis the deep learning-based methods. Third, some public spatiotemporal datasets for urban flow forecasting are shared for facilitating research.

The rest of this paper was organized as follows. In Section 2, we partition the multi-sources spatiotemporal data preparation process into three groups. In Section 3, we choose the spatiotemporal dynamic data as a case study for the urban flow prediction task. In Section 4, we analyze and compare some well-known and state-of-the-art flow prediction methods in detail by classifying them into five categories, i.e. statistics-based, traditional machine learning-based, deep learning-based, reinforcement learning-based and transfer learning-based methods. Finally, we show open challenges of urban flow prediction and discuss many recent research works on urban flow prediction.

2. Urban flow prediction preparation

Multi-sources data from urban must be processed and prepared for further data analysis. In this section, we partition the preparation process into three groups.

2.1. Spatiotemporal datasets

The datasets used for urban flow prediction are most spatiotemporal data. From the characteristic perspectives of spatial and temporal, we can divide the datasets into three categories, e.g. spatiotemporal static data, spatial static temporal dynamic data and spatiotemporal dynamic data. According to data type of the spatiotemporal datasets, there are point data and network data [18]. The spatiotemporal datasets are illustrated in Table 1.

2.2. Map decomposition

It can be found that in cities, a large amount of data is spatiotemporal data, e.g. traffic data including bicycle renting and returning data, taxi track data, metro card swiping data, etc. These data have time and space properties, and are in a constantly changing state. Therefore, we need to represent and measure these data at the time and space level. In order to better process these spatiotemporal data, we should decompose the city map first. One decomposition method is grid-based decomposition. For example, a DNN-based prediction model was proposed for spatiotemporal data [5], which can capture both temporal and spatial properties. They defined a grid map based on the longitude and latitude and partitioned the Beijing city in to an $M^{*}N$ grip map. For example, there is an entertainment region (i, j) that lies at the i^{th} row and the j^{th} column in the grip map, as shown in Fig. 1. This is a good presentation of dividing region into some grids. Then they used in-flow and out-flow to measure the crowd flow in a region [3]. Another is the road network-based map decomposition method in which the vehicles' GPS trajectories are mapped onto the city road network [19]. Unlike the method mentioned before, it can sufficiently take advantage of the road network's information and apply classical clustering approach for further refining. However, it is not as convenient and simple as grid-based decomposition method.

2.3. Dealing with data problems

When dealing with spatiotemporal data, we may face many data problems, e.g. data missing [4,20,21], data imbalance [22–25] and data uncertainty [26–30]. These data problems appear alone or in combination, which will reduce the accuracy of analysis result and efficiency of prediction model. In the following, we review some methods to overcome them.

2.3.1. Data missing

Due to sensor failures, communication errors, and other human factors, spatiotemporal data is often missing. Data missing often brings negative impact to the subsequent data analysis, so it is necessary to study the problem of data missing. At present, the main data missing processing method is to fill the missing value. For example, Lee et al. [20] proposed a factorial hidden Markov model to recover missing values. Hoang et al. [4] divided a city into low-level regions based on road network, and grouped adjacent these regions with similar crowd flow patterns using graph clustering. It's a novel and effective solution, but it's not sure how to define *similar crowd flow patterns* accurately. To consider temporal and spatial correlations, Yi et al. [21] proposed a spatiotemporal multi-view-based learning (ST-MVL) method to collectively fill missing value in a collection of geo-sensory time series data. This method received excellent performance because of combining empirical statistic models, which consist of Inverse Distance Weighting, Simple Exponential Smoothing, and User-based and Item-based Collaborative Filtering. Hence, data missing in spatiotemporal datasets may have an implicit spatiotemporal correlation.

2.3.2. Data imbalance

Spatiotemporal data imbalance is mainly manifested in two aspects: data distribution imbalance and data label imbalance. First of all, for the problem of imbalanced data distribution, Zheng et al. [22] proposed a semi-supervised learning algorithm to deal with the problem of sparse training data caused by the lack of air monitoring stations. Then, for the problem of imbalanced data label, Beckmann et al. [23] studied a KNN-based undersampling methods for data balancing. Wang et al. [24] used a K-labelsets ensemble method based on mutual information and joint entropy to deal with inblanced data. Gong et al. [25] presented a ensemble method using random undersampling and ROSE sampling to solve the imbalance classification problem. So when we face the data imbalance problem, it's a good choice to determine data distribution or data label imbalance, and then apply these corresponding methods.

2.3.3. Data uncertainty

In the actual deployment of machine learning algorithm, in order to better explain the model and effectively deal with the risk caused

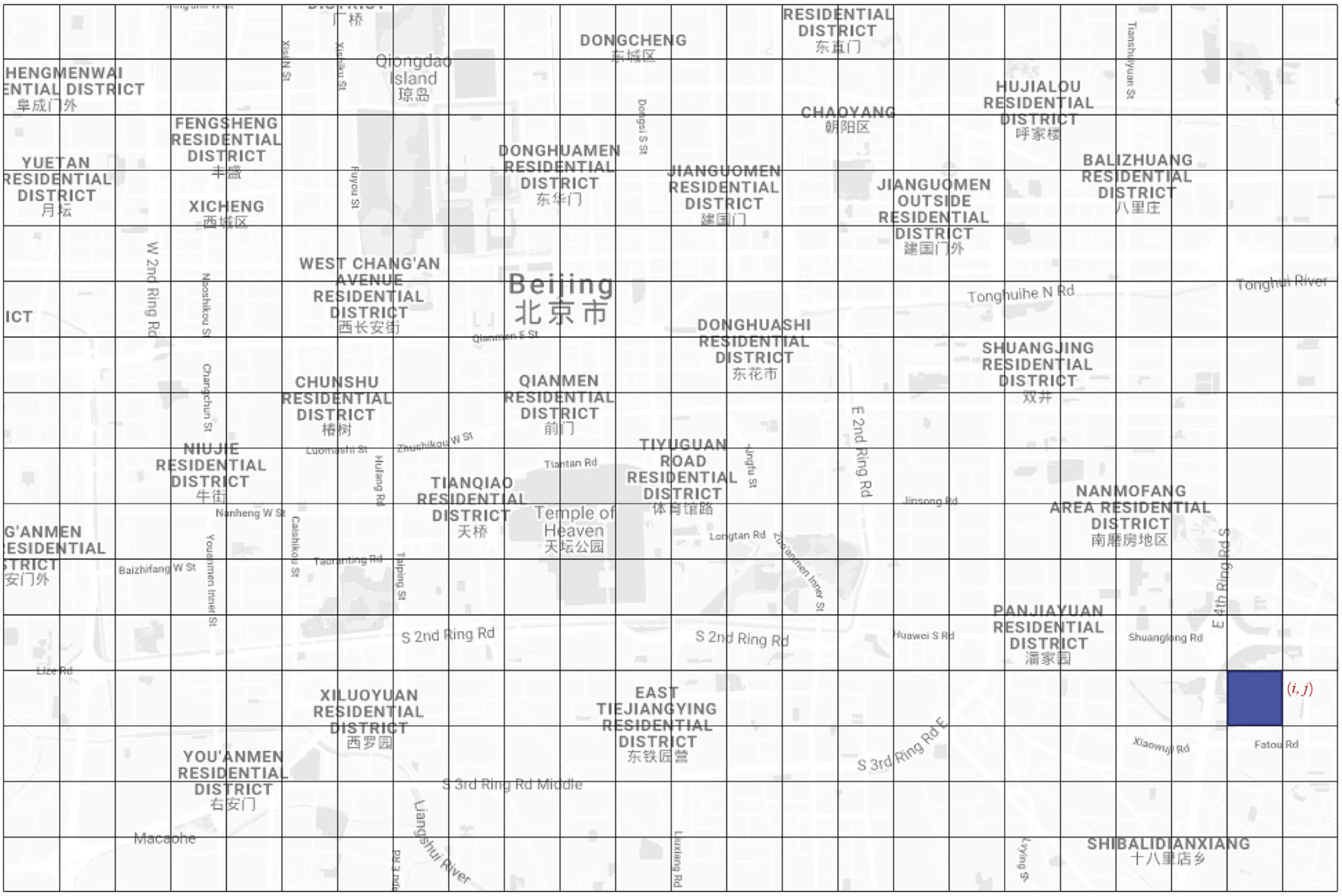

Fig. 1. Grid-based map segmentation in Beijing and the blue region is an entertainment area (Happy Valley Beijing). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by data uncertainty, researchers proposed to adopt uncertainty quantification to alleviate the problem of data uncertainty [26,27]. Bayesian Deep Learning [28] is a kind of uncertainty quantification technique which can learn the weight distribution of networks. Quantifying predictive uncertainty in neural networks is a challenging problem. Lakshminarayanan et al. [29] proposed a model which puts uncertainty into the loss function and is directly optimized through BP algorithm. Rangapuram et al. [30] combined state space models with deep learning for probabilistic time series forecasting. This method keeps properties of state space models such as data interpretability. Recently, more and more uncertainty quantification research works emerge in the traffic forecast field [31–33]. Due to the uncertainty and chaos of the traffic system itself, the point prediction methods will cause high prediction errors [32]. In order to quantify the uncertainty of prediction and make more robust decisions, it is necessary to introduce predictive uncertainty quantification in traffic prediction. By quantifying ideas, it can give an accurate and reliable prediction interval with upper and lower bounds with a certain degree of confidence [32]. It hopefully becomes to be a novel direction for traffic forecast research.

3. Trajectories data preprocessing case

In this section, we choose the spatiotemporal dynamic data (trajectories data) as a case study for the urban flow prediction task because it is one of widely studied spatiotemporal data types in urban flow forecasting and it relates more with our urban crowd flow and traffic flow prediction topic.

In the spatiotemporal data mining field, spatiotemporal data types can be divided into four categories [34], which are different from the classification we mentioned in Table 1: (i) event data, which often occurs at point locations and times (e.g. a concert or a car accident), (ii) trajectory data, which refers to the trajectories of moving objects (e.g. human, vehicle, animals), (iii) point reference data, where a continuous spatiotemporal field is being measured at moving spatiotemporal reference sites (e.g. surface temperature are measured by using weather balloons), and (iv) raster data, whose measurements of an spatiotemporal field are collected at fixed spatiotemporal grids (e.g. air quality of Earth's surface collected by ground-based sensors). We can find that the spatiotemporal event data and spatiotemporal point reference data are point data, while trajectories data and spatiotemporal raster data are network data. They can be merged into the categories of spatiotemporal datasets showed in Table 1. As we know, the derivation of trajectories data can be classified into four main categories, which are human mobility, transportation vehicles mobility, animals mobility and natural phenomena [35]. In this paper, the urban flow prediction task mainly aims at estimating and predicting human mobility and transportation vehicles mobility. Before starting data mining tasks for urban flow prediction, we should preprocess these spatiotemporal dynamic trajectory data. The raw location traces are often collected by smartphones with GPS and WiFi or taxis equipped with a GPS sensor. There are some spatiotemporal trajectory data preprocessing methods, consisting of extracting the history of place visits, data filtering and statistics, trajectory compression, trajectory segmentation and map matching [35,36].

Extracting the history of place visits. From the analysis of the location traces, we can find that there are transitions and stay points in

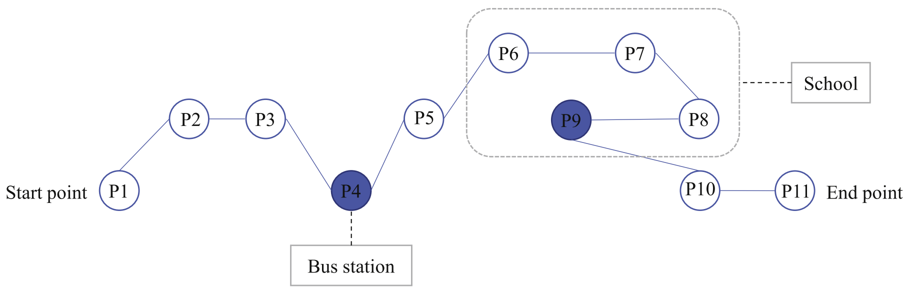

Fig. 2. Stay points and regions in trajectories.

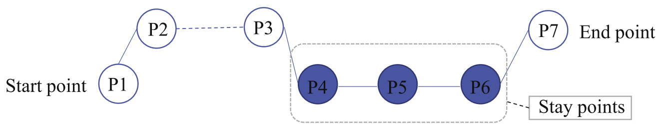

Fig. 3. Time-interval and stay points trajectory segmentation.

trajectories data. Then using a gird clustering algorithm [36], the stay regions will be generated from the set of stay points based on a given radius. So the history of place that users visited can be extracted. As demonstrated in Fig. 2, if we set the minimum time of stay points to 5 min, the bus station (P4) can be found in this trajectory, and the stay region can be generated around the school (P9) about 1km radius. Based on these places with timestamp information, some applications may appear, e.g. travel recommendation, business location and travel time estimation.

Data filtering and statistics. Because of sensor's error or other technical issues, there are some incomplete instances or outliers in the trajectories data. To forecast urban flow well, we need to filter out these incomplete instances and outliers before prediction task. The filter step is significant to avoid biased estimates of prediction performance. In order to find the trajectories distribution, we also need to perform some statistics analysis, such as mean, median, the ratio between location and transitions, and how many places visited by a user during a given recording interval.

Trajectory compression. We can collect time-stamped location every second even more accurate time measurements for moving objects. It will cost plenty of communication, computing and storage sources. To efficiently collect and leverage these data, we need to compress the trajectory data. There are two major categories of trajectory compression methods. One is offline compression, such as Douglas-Peucker algorithm [37], which replaces the original trajectory by an approximate line segment until the negligible error is below a specified error. The other is online compression, such as Sliding Window algorithm [38] and Open Window algorithm [39], which transmits trajectory data timely. They are window-based algorithms which fit the trajectory points in a sliding window with a valid line segment and expand the sliding window until it exceeds the specified error bound.

Trajectory segmentation. In order to classify or cluster trajectories to mine more useful knowledge, we need to study trajectory segmentation before mining tasks. There are three common types of trajectory segmentation methods. They are time interval-based, shape of a trajectory-based and semantic meaning-based methods. The first one is that a trajectory is divided into some segments based on a given time-length (larger than the given threshold or the same time interval), as illustrated in Fig. 3. Due to the time length between p2 and p3 being larger than a given time interval, so we can divide the trajectory into two segments (p1p2 and p3p7). The second one is that we can partition a trajectory by the turning points with heading direction changing over a threshold [35]. The last one is based on the semantic meaning of points in a trajectory. For example, in the travel speed estimation task, we often remove the stay points from the GPS trajectories because the stay points may be location where taxi is waiting for passengers [40]. For example, from Fig. 3, the trajectory points in the dotted box can be removed because of stay points (p4p5p6) and the trajectory can be divided into two segments (p1p3 and p6p7). To find a walk-based segmentation [41,42], we also need to combine with the human mobility patterns and employ further semantic meaning-based trajectory segmentation research.

Map matching. There are two major categories of map matching methods. One is the additional information-based method, and the other is the range of sampling points-based method. The first type of methods can be divided into four groups: geometric [43], topological [44,45], probabilistic [46,47] and other advanced methods [48–50]. The second type of methods can be divided into two categories: local and global methods. The local methods aim to find a local optimal point based on the distance and orientation similarity. The global methods [51,52] try to match an entire trajectory with a road network.

4. Techniques for urban flow prediction

Urban flow prediction is one of the spatiotemporal prediction tasks in the intelligent transportation field, which generally aims to predict the urban flow (e.g. the traffic of crowds, vehicles, and bikes) in each region at the time interval when given the historical observations and other influence factors. There are some related research works, e.g. air quality prediction [53–55], traffic flow prediction [8,56,57], and travel demand prediction [58–60]. As urban flow prediction can provide traffic patterns for urban traffic management to improve the efficiency of public transportation, and give congestion warnings in advance for public safety emergency management [8,61,62], it has become more and more important in traffic management and public safety. From the perspective of spatial forecasting measure, urban flow prediction can be divided into three categories, e.g. citywide-level, region-level and road-level [4,6,63]. From the perspective of temporal prediction, urban flow prediction can be parted into three categories, e.g. short-term, mid-term and long-term flow prediction [3,63]. And we can find that the problem of urban flow prediction has both spatial relation and temporal dependencies.

To solve the urban flow prediction problem, in recent years, there are many novel methods have been proposed. The major methods can be classified into five categories: statistics-based methods [64–66], traditional machine learning methods [7,16,17,62,67], deep learning-based methods [5,6,12,15,57,68–70], reinforcement learning methods [71,72] and transfer learning methods [73,74].

4.1. Statistics-based methods

In statistics-based methods, ARMA (Autoregressive Moving Average) [64] is a fundamental time series prediction method, and the variant method is ARIMA (Autoregressive Integrated Moving Average) [65]. An integrated version of ARMA model is also very popular in time-series prediction problems. The Hyndman-Khandakar algorithm can be applied for automatic ARIMA modelling in R [75]. The default procedure contains two steps [76] : (i) The number of differences ($0 \le d \le 2$) is determined using repeated KPSS tests; (ii) The value of p and q are then

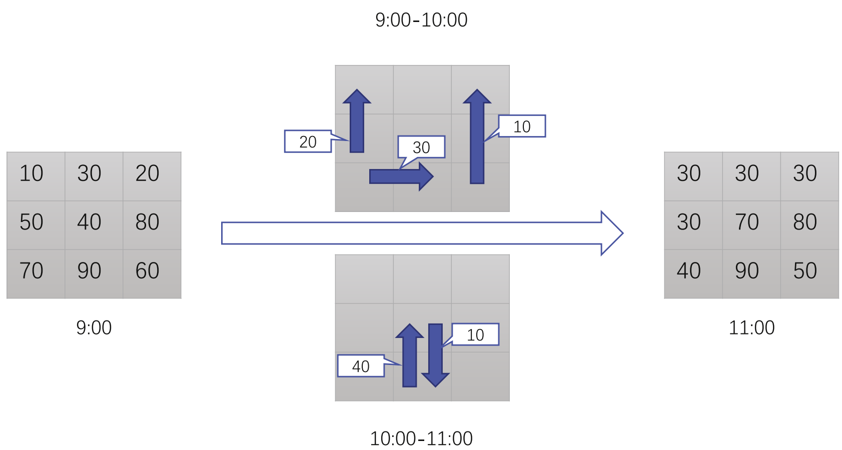

Fig. 4. The task of estimating people flow between cells. (The original task description is available in [62]). Input: population of each grid cell at each time; Output: the number of people who move between cells over time. The map is divided by cells based on latitude and longitude.

chosen by minimizing the AICc (AICc is AIC (Akaike Information Criterion) with a correction for small sample sizes) after differencing the data d times. Rather than considering every possible combination of p and q, the algorithm uses a stepwise search to traverse the model space. And the process for forecasting is summarized in Fig. 5. As an another extension of the ARMA method, Seasonal Auto-Regressive Integrated Moving Average (SARIMA) method [66] can catch intrinsic correlations in time series data, especially fit for modeling seasonal, stochastic time series that always occur in traffic flow data. Although these classical time-series methods can capture temporal dependencies in time series data, they can't depict the spatial influence in urban flow prediction problems.

4.2. Traditional machine learning methods

Support vector regression (SVR) model is usually used in traffic flow prediction. For example, SVR with RBF kernel multiplied by a seasonal kernel has been used in traffic flow forecasting with high prediction accuracy and computational efficiency [16]. As one of non-parametric and data-driven methods, an enhanced K-nearest neighbor (K-NN) algorithm was applied in short-term traffic flow prediction based on identify similar traffic patterns [67]. Zhu et al. studied a linear conditional Gaussian (LCG) Bayesian network (BN) model for short-term traffic flow prediction, which considers spatiotemporal characteristics as well as speed information [7]. To tackle the task of estimating the number of people who moved between cells, Akagi et al. [62] developed a probabilistic model based on collective graphical models, which has considered movements to remote cells. As presented in Fig. 4, the proposed method is an unsupervised learning method and only needs input variables. The input variables are spatiotemporal population data [62]. Liu et al. [17] developed a graph processing framework based traffic estimation (GPTE), which can capture traffic correlation from taxi data and enable advanced traffic estimation at city-scale based on graph-parallel processing method. From these previous traditional machine learning methods, we can conclude that these models mainly focus on short-term traffic flow prediction and receive high prediction accuracy. However, traffic data explosion is due to the increase of traffic sensors and the rapid development of intelligent transport systems in recent years. Traditional machine learning methods are restricted with mining the deep, implicit spatiotemporal correlations in the big traffic data.

4.3. Deep learning-based methods

Deep learning-based methods are becoming popular methods for traffic spatiotemporal tasks. Due to big data and strong computing power, the success of deep learning in many application scenarios has motivated it to be used in a large number of different areas, such as CNNs in computer vision [77,78] and RNNs in sequence learning tasks [79,80].

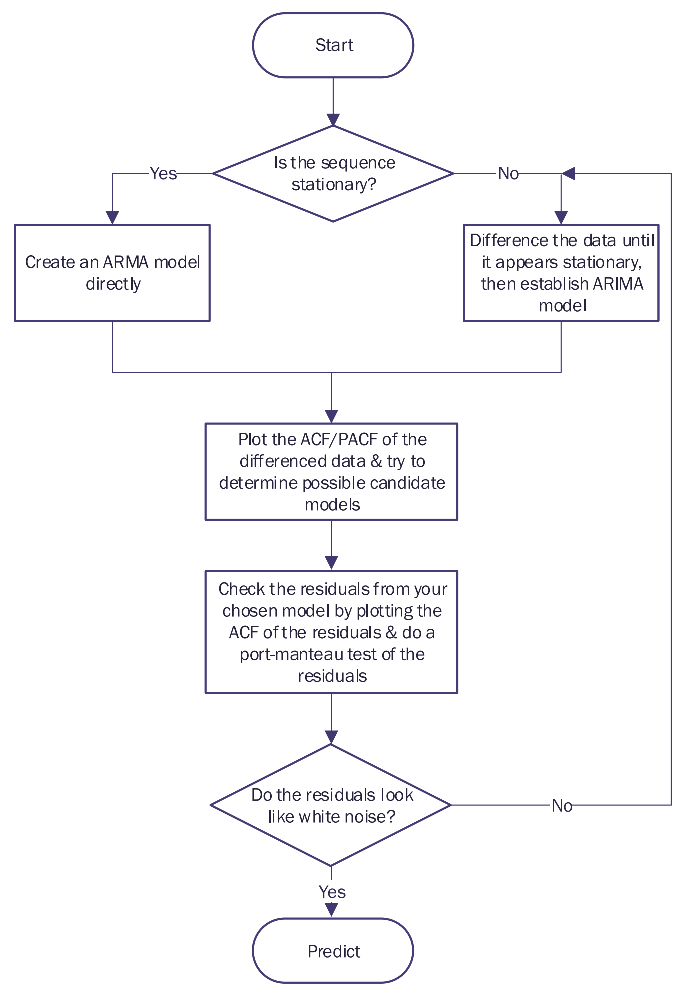

Fig. 5. General process for forecasting using an ARIMA model.

In previous research work, there are many papers on human mobility prediction based on their history location trajectories data [36,81,82]. They aim at providing context-aware services and other location-based services for users. However, it may expose the privacy of users to others. So these data may be unavailable because of the policy of protecting privacy. Compared with the human mobility prediction problem, we can obtain more related datasets based on crowd-scale rather than individual-scale, e.g. taxi trajectories data, public transportation system data (metro or bus card swiping data), bike-sharing data, road network data, weather data, to forecast urban crowd flow. And it is also of great importance to traffic management and public safety.

From the perspectives of people's travel mode and urban flow type, we can divide the urban flow into three categories: crowd flow, traffic flow, and public transit flow. Crowd flow usually can be concluded from users' phone signals and other vehicles GPS trajectories separately or synthetically. Traffic flow can mainly be estimated by using taxi trajectories data. Public transport flow indicates that the movement passengers measured by public transit card swiping data or bike-sharing data. If we want to measure the urban flow accurately, all the flow types need to be prepared and analyzed synthetically. However, in practical scenario, it's hard to receive all the urban flow types of data in a city and there are some complex relationships among these flow. Next, we will review and analysis the three types of urban flow separately.

4.3.1. Crowd flow prediction

In recent years, there are many researchers focus on citywide-level traffic flow prediction [4,10,83], to forecast the citywide-level crowd flow. In the following, before starting touch deep learning models, we will follow the related definitions of the crowd flow prediction problem in advance [5].

Region [5]. There are many definitions of Region in terms of different scales and semantic meanings. In most studies, they often partition a city into an I^*J grid map based on longitude and latitude and a grid cell called a region, as shown in Fig. 1.

Inflow/Outflow [5]. Let $\mathbb{P}$ be a collection of trajectories at the t^{th} time interval. A grid cell (i, j) means the i^{th} row and the j^{th} column. The inflow and outflow of the crowds at the time interval t are defined respectively as follows.

$$x_t^{in,i,j} = \sum_{T_r \in \mathbb{P}} \left| \left\{ k > 1 | g_{k-1} \notin (i,j) \wedge g_k \in (i,j) \right\} \right|, \tag{1}$$

$$x_t^{out,i,j} = \sum_{T_r \in \mathbb{P}} \left| \left\{ k \geqslant 1 | g_k \in (i,j) \wedge g_{k+1} \notin (i,j) \right\} \right|, \tag{2}$$

where $T_r : g_1 \to g_2 \to \ldots \to g_{|T_r|}$ is a trajectory in $\mathbb{P}$, and g_k is the geospatial coordinate, $g_k \in (i, j)$ means the point g_k lies within grid (i, j), and vice versa, and $|\cdot|$ denotes the cardinality of a set.

At the tth time interval, inflow and outflow in all I^*J regions denote as a tensor $X_t \in \mathbb{R}^{2*I*J}$, where $(X_t)_{0,i,j} = x_t^{in,i,j}$, $(X_t)_{1,i,j} = x_t^{out,i,j}$.

Prediction Target [5]. Given the historical observation $\{X_t | t = 0, \ldots, n-1\}$, predict X_n.

In 2016, a deep neural network (DNN) prediction model called DeepST was proposed, which can capture spatial and temporal properties to predict citywide crowd flow [5]. The architecture of DeepST contains two parts: time sequence part and external factors part. From the history observation, the time serials contains temporal closeness, period and seasonal trend properties. The external factors have some related information with crowd flow prediction, e.g. dayofweek, weekday/weekend and meteorological condition. Then convolution layers are employed to capture the temporal closeness, period and seasonal trend properties, and the convolution layers output is fused followed by three sequential convolutional layers. At last, this result is fused with the output of the external factors captured by fully-connected layers and the prediction target X_n is obtained. Finally, they built a real-time flow forecasting system (called as UrbanFlow) based on the DeepST model.

Example 1: Recent years, a novel deep learning-based model (ST-ResNet) [6] for citywide crowd flow prediction was presented, which is shown in Fig. 6. The city is partitioned by using a grid-based method for forecasting the crowd flow in each and every region of a city [3,6]. Note that the model outperforms other classical time-series and deep learning prediction methods. Mostly like the DeepST model, the ST-ResNet adds Residual Units and only one fusion component. The fusion step uses a parameter-matrix-based fusion method,

$$X_{Res} = W_c \circ X_c^{L+2} + W_p \circ X_p^{L+2} + W_q \circ X_q^{L+2} \in \mathbb{R}^{2*I*J}, \tag{3}$$

where $\circ$ is Hadamard product (i.e., element-wise multiplication), W_c, W_p and W_q are the learnable parameters that adjust the degrees affected by closeness, period and trend, respectively. They concatenate the output of the three components after fusion with the external component. To model citywide dependencies, they employ residual learning in this ST-ResNet model, which has been demonstrated to be very effective for training super deep neural networks of over 1000 layers [6,84]. The residual unit used in the ST-ResNet is shown in Fig. 7. But in short-term crowd flow prediction problem, the residual network structure of ST-ResNet can be removed to get much more better performance, because it's not necessary to use the residual network structure to capture the distant spatial dependencies far away from the target region. And the ST-ResNet also needs too many data to train the model, so it has not good performance if we can't get much available data [63]. Some researchers chose the ST-ResNet model as baseline to do further urban crowd flow prediction tasks [6,63], because it outperforms other deep learning-based methods before. In the urban flow prediction task, it's usual to use Mean Absolute Error (MAE), Mean Squared Error (MSE), Rooted Mean Square Error (RMSE) and Mean Average Percentage Error (MAPE) as evaluation metrics (the smaller the result, the better the algorithm),

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|, \tag{4}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}, \tag{6}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i}, \tag{7}$$

where n is the number of instances, $\hat{y}_i$ is the prediction result and y_i is the ground truth.

4.3.2. Traffic flow prediction

Traffic flow prediction plays an important role in urban flow forecasting [9,13,85,86], because the taxi GPS data can access easily and it can represent the citizens's traffic behaviors without the limitation of fixed lines. For example, Jiang et al. [68] developed a deep learning framework which transforms geospatial data to images using Convolutional Neural Network (CNN) and residual networks for traffic prediction. Wu et al. [57] proposed a novel model, which combines the CNN and RNN to capture the spatiotemporal features and learn the importance of past traffic flow using attention mechanism. This method makes full use of the temporal and spatial characteristics of traffic flow to model and improve prediction performance. Since the forecast of traffic flow is affected by complex factors, e.g. temporal relationship, spatial correlation, and other external factors (weather and events), it is more challenging to accurately predict traffic flow. Zhang et al. [15] studied a multitask deep learning framework to simultaneously forecast the node flow and edge flow in the spatiotemporal networks. The model outperforms 11 baselines and shows great prediction performance in traffic flow forecast.

Example 2: As illustrated in Fig. 8, Yao et al. [87] first learned spatial dynamic similarity and handled long-term periodic temporal shifting in a unified framework called Spatial-Temporal Dynamic Network (STDN). The long-term periodic representation and temporal shifting are captured by periodically shifted attention mechanism, as depicted in Fig. 8 (a), where they used Long Short-term Memory (LSTM) module to model sequential information for each day. As shown in Fig. 8 (b), the short-term temporal representation generates from one LSTM. The flow gating mechanism is shown in Fig. 8 (c), which controls the spatial information propagation via a flow gate to infer the flow gated spatial representation. Conv and FC mean several convolutional layers and

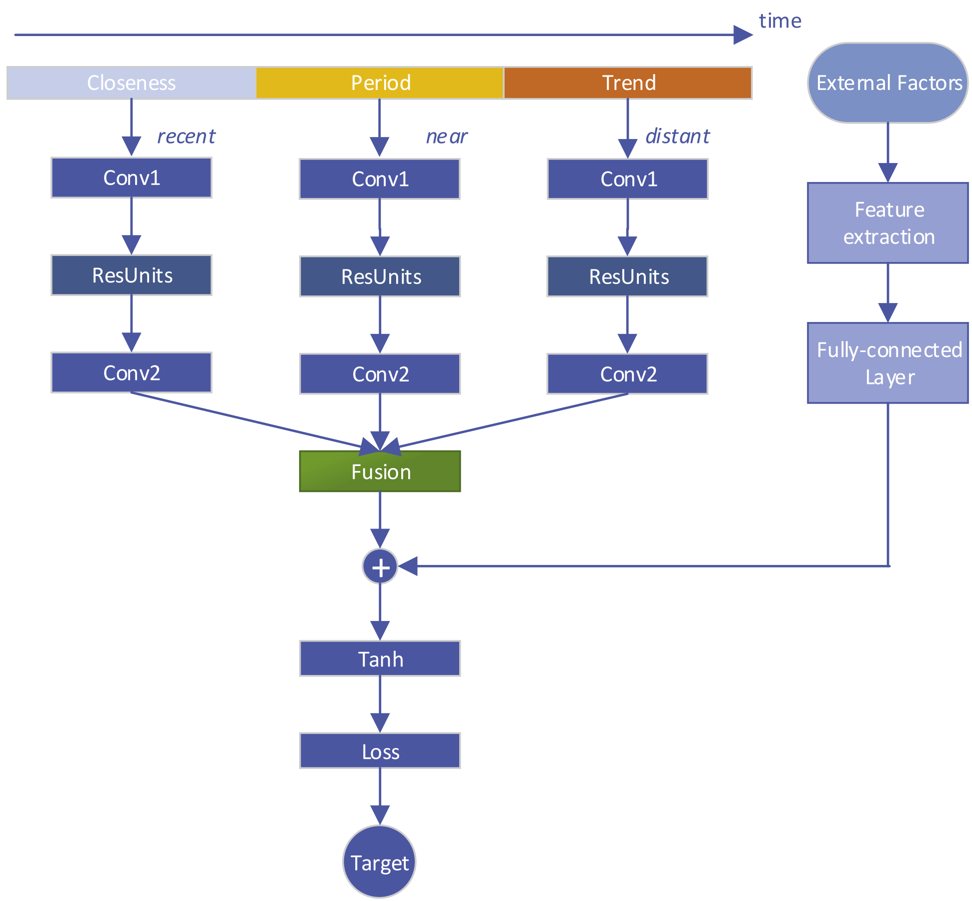

Fig. 6. The architecture of ST-ResNet. (The original ST-ResNet architecture is available in [6]). The model outperforms other classical time-series and deep learning prediction methods.

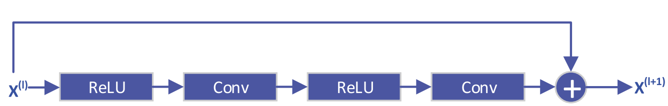

Fig. 7. Residual Unit. (The original Residual Unit is available in [6]).

fully connected layers, respectively. Fig. 8 (d) presents the paradigm of joint training. The short-term representation and long-term representation are concatenated and then fed to a fully connected layer to get the final prediction value of start and end traffic volume for each region. This work aims to predict traffic flow from spatiotemporal datasets using a Spatial-Temporal Dynamic Network. They evaluated the STDN method on two large-scale real world public datasets including taxi data of Newer York City (NYC) and bike-sharing data of NYC. It's a good example for urban flow prediction from spatiotemporal data using machine learning and the datasets are also given in the Section 7.

4.3.3. Public transit flow prediction

As an important component of the urban public transportation system, the metro has been rapidly deployed in the city because of its large capacity, high speed and high reliability, and it has attracted a large number of passengers [11,69,70,88]. Therefore, doing a good job of metro passenger flow forecast will not only help the metro management department to manage passenger travel demand and optimize metro dispatch, but also help passengers choose travel time and travel mode. Liu et al. [69] proposed an end-to-end deep learning model, named as DeepPF, to forecast the metro inbound and outbound passenger flow. They combined all the influence factors, such as temporal dependencies, spatial characteristics, metro operation properties and external environment factors to predict short-term metro passenger flow. The experiment showed that the model has good prediction performance and can be applied to general conditions. We can find that it is feasible and effective to use the deep learning methods to capture the temporal and spatial characteristics of metro data and predict the metro passenger flow. Ma et al. [70] analyzed the metro data's spatiotemporal characteristics and then developed a parallel framework which comprises convolutional neural network (CNN) and bi-directional long short-term memory network (BLSTM) to forecast metro passenger flow. The model was evaluated by Beijing metro network data and it outperformed traditional statistics methods. In recent years, bike-sharing is becoming more popular in urban transportation because of providing flexible transport mode and reducing the production of greenhouse gas.

Example 3: As illustrated in Fig. 9, Chai et al. [12] proposed a novel multi-graph CNN method to predict bike flow at station-level. This method gives us a novel graph neural network perspective to study traffic prediction, which includes three parts: graph generation, multi-graph convolution, and prediction network. It is worth noting that the construction of the graph is very novel. In addition to the general distance graph, an interaction graph and a correlation graph are also introduced. The distance graph refers to the weight of the connected edge measured by the reciprocal of the distance between the two stations; the interaction graph measures the weight of the connected edge by the number of historical travel records between the two sites; the correlation graph is based on the number of two stations historical travel records, using the Pearson coefficient to calculate the relationship between the two stations. Before the multi-graph convolution operation, these three types of graphs generate one fused graph by a graph fusion method, which normalizes the three different graphs and makes a fusion graph by weighted summing adjacency matrices. Next, the convolution opera-

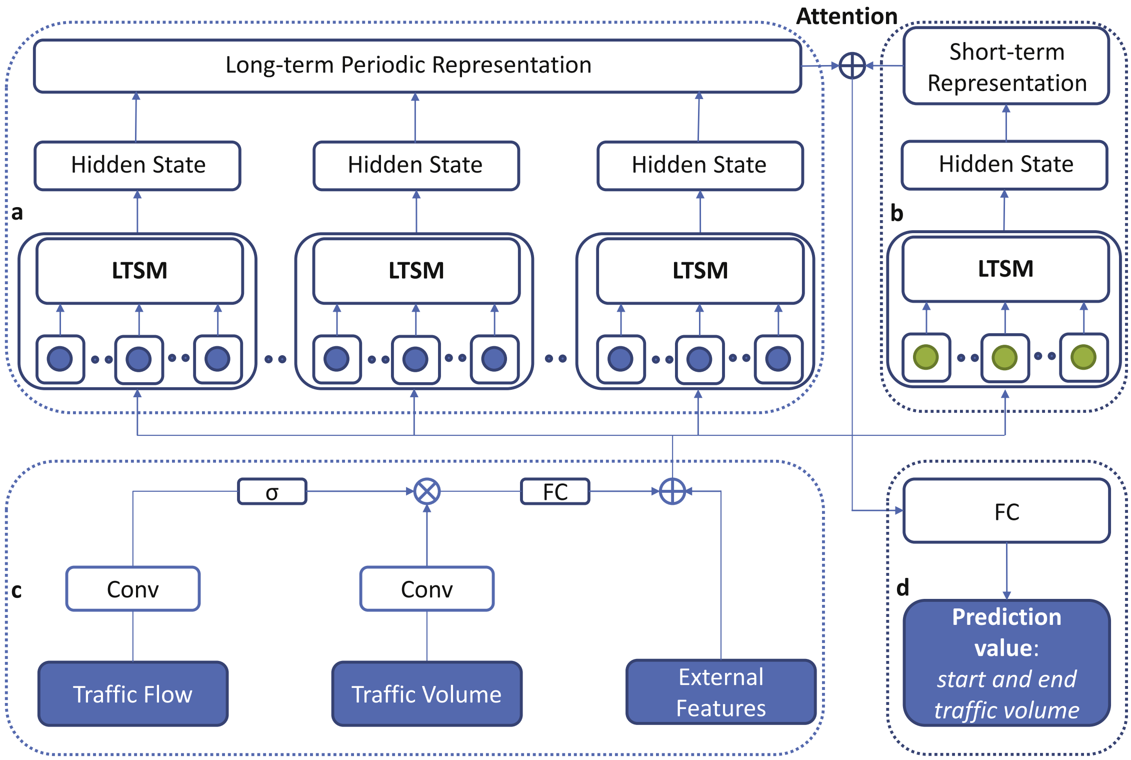

Fig. 8. The architecture of STDN. (The original STDN architecture is available in [87]).

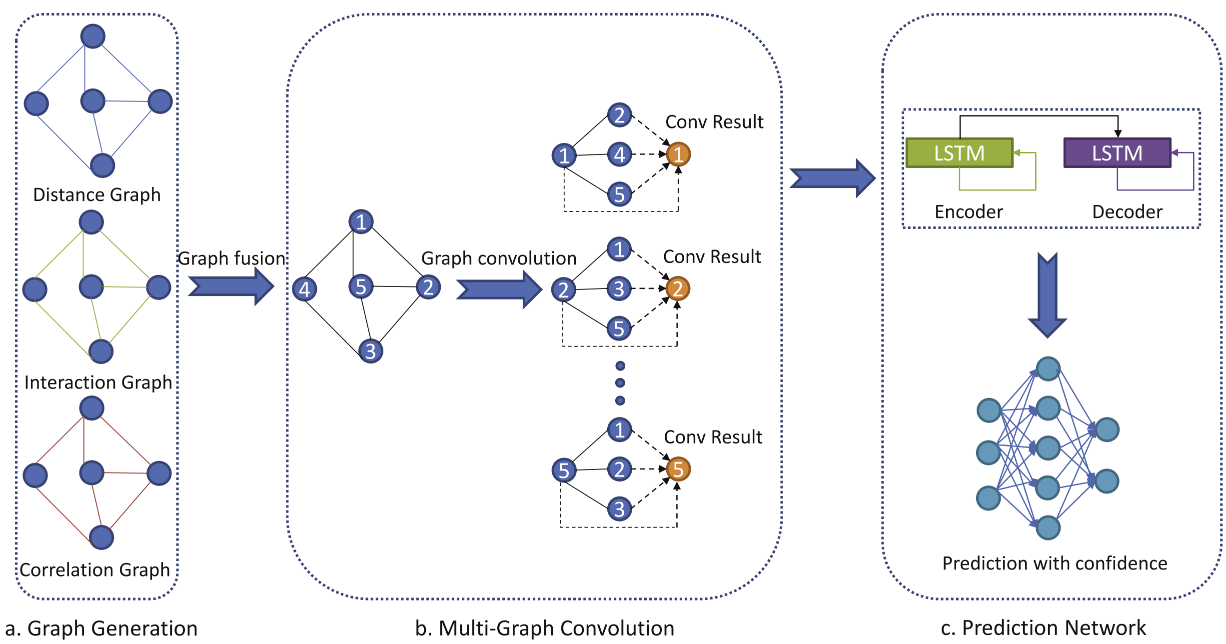

Fig. 9. The architecture of Multi-Graph Convolutional Neural Network Model for Bike Flow Prediction. (The original Multi-Graph Convolutional Neural Network Model is available in [12]).

tion is performed on the fusion graph, as shown in Fig. 9 (b). Finally, the graph convolution results are input into the prediction network through a pre-trained encoder-decoder network. Then other external context features (e.g., temperature, wind speed, weekday/weekend) are combined to output the bike flow prediction result with confidence through a fully connected network. The datasets used in the experiments (NYC bike data, Chicago bike sharing data, weather and climate data) are shown in the Section 7.

4.4. Reinforcement learning-based methods

The reinforcement learning methods can usually be applied in traffic flow optimization problems. As we know, traffic congestion is a tricky

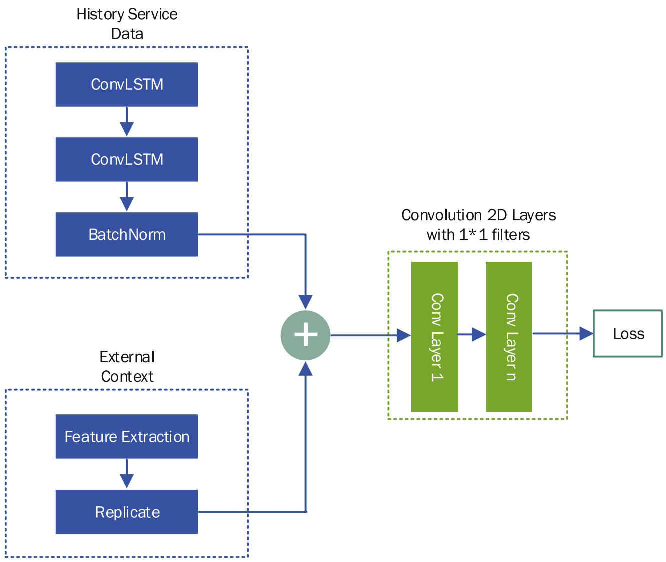

Fig. 10. Deep Spatiotemporal Neural Network with Region Representations. (The original model is available in [73]).

problem in urban and a large amount of traffic flow in a period is an essential factor causing traffic congestion. Hence, it is necessary to optimize traffic flow, make traffic control in advance, and alleviate traffic congestion. To tackle these challenges, Erwin et al. [71] proposed a method to optimize traffic flow based on reinforcement learning, which uses Q-learning to obtain policies dictating the maximum driving speed allowed on a highway. The model takes traffic prediction into account and controls traffic flow proactively. More importantly, it can further help alleviate traffic congestion. Another example is to coordinate passenger inflow control problem on an urban rail transit line in Shanghai. In order to reduce the frequency of metro passengers being stranded and ensure public safety, Jiang et al. [72] presented a reinforcement learning-based method applied to study metro passenger inflow control strategy in peak hours. These methods aim at finding better optimization strategies to optimize current traffic flow using reinforcement learning algorithms, and improving the efficiency of intelligent transportation systems. However, few studies combine deep learning methods and reinforcement learning approaches to predict and optimize traffic flow jointly, and it seems to be a promising research direction for the future.

4.5. Transfer learning-based methods

Using transfer learning methods to predict crowd flow is the novel direction for urban flow forecast in a data-scarce city [73,74]. A novel network for spatiotemporal prediction with region representation was developed [73], as shown in Fig. 10. The model's objective is to minimize the squared error between predicted $\tilde{y}_t$ and real y_t.

$$\min_{\theta} \sum_{t\in\mathbb{T}} \left\| \tilde{y}_t - y_t \right\|_F^2 \tag{8}$$

This method focuses on finding inter-city region pairs that share similar patterns and then transfers knowledge from data-adequate city (source city) to data-scarce city (target city). This method is different from the deep learning method that requires a large amount of data to train, which does not require too much data but takes full advantage of knowledge transfer from the source city. But the model only shows good performance between two cities with similar patterns and it's not easy to find the matching function of the source regions and target regions.

5. Open challenges

In this paper, how to forecast urban flow from spatiotemporal datasets were reviewed comprehensively. In dealing with these datasets and choosing suitable analysis methods, some challenges of urban spatiotemporal flow prediction could be divided into three aspects.

Dealing with multiple influence factors. When we design our model, we may face the following problems: What is the main factor affects urban spatiotemporal flow and what are the minor but necessary factors for our given tasks. For example, in this crowd flow prediction problem, we need consider the influence factors of spatial and temporal scale, and other factors, such as weather, holidays, social events, traffic accidents, traffic jams and so on.

Finding suitable data fusion methods. In order to improve our model's performance, we often need more available datasets from various domains (traffic, weather, social) to train the deep learning-based model. However, we may face those problems: how to choose the suitable datasets for our tasks and how to fuse these heterogenous data as the model input. Hence, it's not trivial to investigate cross-domain data fusion methods to speed up the research.

Limitations in data sparsity. In practical scenario, some data are missing and failure due to sensor error, transmission failure and storage loss. These incomplete/inaccurate flow data will reduce the prediction accuracy. Moreover, these flow data are often time series data with geographic location information. At present, the current missing value filling methods are difficult to directly apply to the filling tasks of missing values of these flow data. It is necessary to further study the missing value processing method for spatiotemporal flow data to support subsequent data mining tasks.

6. Discussion

Many studies on urban flow prediction spring up during recent years, with the development of machine learning algorithms, especially the still active deep learning methods, and more open urban spatiotemporal datasets for research use. In order to give researchers a clear research progress outline in the field and help them with further research, we summarized the classic and representative works on urban flow forecast research during recent five years. We listed these recent works mentioned in Section 4, as shown in Table 2.

In Table 2, we categorized these works into three categories by tasks they faced, datasets and used methods. And then, we listed some open datasets to facilitate researchers to further dig into the field.

From Table 2, we can find those deep learning methods (e.g. DNN, CNN, LSTM, and GRU, etc) are frequently used in solving many urban flow prediction tasks, such as crowd flow prediction and traffic flow prediction. Due to more public transit data are becoming available on research under the privacy protection agreement, we can see that there are some public transit flow forecast papers using advanced machine learning algorithms. The works list in the Table 2 also give researchers good examples for dealing with urban spatiotemporal flow forecast problems. From these representative previous works, it's better to use statistics-based methods (e.g. SVR) and traditional machine learning algorithms (e.g. Linear conditional Gaussian Bayesian network) when addressing short-term traffic flow prediction tasks. If you want to capture urban flow temporal dependency, spatial correlation simultaneously and further improve the forecast performance, you can try to use deep learning-based methods (e.g. DNN, CNN, and LSTM), even a hybrid deep learning-based method. These methods have shown excellent results in spatiotemporal flow forecast tasks. Recently, reinforcement learning-based methods (e.g. Q-learning) have been applied in urban traffic flow optimization problem. We also find that the method combines traffic flow prediction using deep learning and traffic flow optimization using reinforcement learning, which shows a promising direction for urban flow study. At last, if you have a small amount of traffic data at hand, you can consider using a novel transfer learning approach, which has the ability to take advantage of the knowledge learned from the source domain.

Table 2
A summary of previous works about urban flow prediction.

Urban flow type	Task	Dataset	Method	Reference
Crowd flow	Human mobility prediction	Mobile phone data	Probabilistic kernel method	[36]
	Forecasting citywide crowd flow	Beijing taxi GPS data, NYC taxi trajectory data and NYC bike data	IGMRF[a] and Bayesian network transit model	[4]
	Crowd flow forecasting	TaxiBJ15, TaxiGY16, LoopGY16 and BikeNYC14	DNN	[5]
	Predicting citywide crowd flow	TaxiBJ and BikeNYC	CNN and Residual networks	[6]
	Citywide short-term crowd flow prediction	MobileBJ and TaxiBJ	CNN and LSTM	[63]
	Estimating People Flow	GPS trace data of cars	A probabilistic model based on collective graphical models	[62]
	Crowd flow prediction	Bike data of New York City and Chicago	Transfer learning-based framework	[73]
	Flow prediction in spatio-temporal networks	TaxiBJ and TaxiNYC	Multitask deep-learning framework	[15]
Traffic flow	Short-term traffic flow forecasting	PeMS sensors data	SVR	[16]
	Short-term traffic flow prediction	Traffic simulation data	Linear conditional Gaussian Bayesian network	[7]
	Taxi movement	NYC taxi data and Road network data	Parallelization of the Dijkstra algorithm and statistics computations	[85]
	Traffic flow optimization	Traffic simulation data	Q-learning	[71]
	Short-term traffic flow prediction	Traffic flow data	DNN	[8]
	Short-term traffic forecast	Traffic data	LSTM	[9]
	Estimate citywide traffic volume	Road network data, POI data, GPS trajectories, Weather conditions and Videos clips	A hybrid framework(machine learning techniques+traffic flow theory)	[86]
	Short-term traffic flow forecasting	UK traffic flow datasets	A hybrid multimodal deep learning method	[56]
	Traffic flow prediction	Traffic flow data	CNN and GRU[b]	[57]
	Traffic flow prediction	Taxi GPS trajectory data	CNN and LSTM	[13]
	Real-time traffic estimation at city-scale	Road network data and SG taxi dataset	Graph-parallel processing framework	[17]
	Taxi pick-up/drop-off	NYC taxi trip data	CNNs and Residual networks	[68]
Public transit flow	Traffic prediction of bike-sharing system	Four datasets (bike data and meteorology data) from NYC and D.C	Hierarchical prediction model	[10]
	Metro stations crowd flow forecast	Metro AFC record data	Residual neural network framework	[11]
	Network-wide metro ridership prediction	Metro ridership data	CNN and BLSTM	[70]
	Optimize metro passenger inflow volume	Metro data	Reinforcement learning-based method	[72]
	Bike flow prediction	Bike data of New York city and Chicago, Weather data	Multi-graph convolutional networks	[12]
	Metro passenger flow prediction	Metro data	DNN	[69]

[a] IGMRF is the abbreviations of Intrinsic Gaussian Markov Random Fields.
[b] GRU is the abbreviations of Gated Recurrent Neural Network.

7. Public urban spatiotemporal datasets

In order to help other researchers further participate and make more valuable works, we collect and organize several related open datasets on this urban spatiotemporal forecast topic. Here are links of these datasets:

- NYC taxi data: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page. This type of data can be used for citywide crowd flow prediction, taxi movement and Taxi pick-up/drop-off tasks.
- NYC bike data: https://www.citibikenyc.com/system-data. Such kind of data can be used for crowd flow forecasting, traffic prediction of bike-sharing system and bike flow prediction tasks.
- San Francisco taxi data: https://crawdad.org/~crawdad/epfl/mobility/20090224/. The taxi data can be used with NYC taxi data before.
- Weather and events data: https://www.wunderground.com/. The weather and events data can be used with other traffic flow data for estimating citywide traffic volume, like [86].
- UK traffic flow datasets: http://data.gov.uk/dataset/highways-england-network-journey-time-and-traffic-flow-data. The traffic flow data is suitable for short-term traffic flow forecasting, and it has been used in [56].
- Traffic flow data is available from the Illinois Department of Transportation: http://www.travelmidwest.com/. The traffic flow data can be used as UK traffic flow datasets.
- Weather and climate data: https://www.ncdc.noaa.gov/data-access. The weather and climate data are suitable for weather forecasting related tasks than urban flow prediction tasks.
- Chicago bike sharing data: https://www.divvybikes.com/system-data. The bike sharing data is often used for urban bike flow prediction, like [12].
- NSW POI data: https://sdi.nsw.gov.au/catalog/search/resource/details.page?uuid=%7BC41F6FE5-1C56-4556-9EC6-EC9BD7094BBB%7D. The POI data are frequently combined with road network data, GPS trajectories for estimating citywide traffic volume, like [86].
- Road network data: http://networkrepository.com/road.php. The road network data is one of the vital data constituent parts for urban flow prediction, like [17,85,86].

8. Conclusion and future work

In this paper, we conducted a comprehensive overview of the recent development methods for urban spatiotemporal flow prediction during 2014–2019, which plays an increasingly significant role in urban computing research and has a close relationship with traffic management, land use, and public safety. We first categorized the complex and dynamic influential factors of urban spatiotemporal flow and introduced the general preparation process of spatiotemporal data for spatiotemporal prediction. Then we surveyed current works on well-known and state-of-the-art urban flow prediction methods for urban spatiotemporal flow prediction tasks including statistics-based, traditional machine learning-based, deep learning-based, reinforcement learning-based, and transfer learning-based methods. Next, we listed some open challenges and discussed how to deal with urban spatiotemporal flow forecast problems. Finally, we gave a list of public urban spatiotemporal datasets and a few suggestions for use.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Peng Xie: Conceptualization, Writing - original draft. **Tianrui Li:** Supervision, Project administration. **Jia Liu:** Investigation. **Shengdong Du:** Writing - review & editing. **Xin Yang:** Visualization. **Junbo Zhang:** Methodology.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 61773324).

References

[1] Q. Chen, X. Song, H. Yamada, R. Shibasaki, Learning deep representation from big and heterogeneous data for traffic accident inference, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 338–344.
[2] K. Jayarajah, A. Tan, A. Misra, Understanding the interdependency of land use and mobility for urban planning, in: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, in: UbiComp '18, ACM, 2018, pp. 1079–1087.
[3] J. Zhang, Y. Zheng, D. Qi, Deep spatio-temporal residual networks for citywide crowd flows prediction, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 1655–1661.
[4] M.X. Hoang, Y. Zheng, A.K. Singh, Fccf: Forecasting citywide crowd flows based on big data, in: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2016, p. 6.
[5] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, Dnn-based prediction model for spatio-temporal data, in: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2016, p. 92.
[6] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, T. Li, Predicting citywide crowd flows using deep spatio-temporal residual networks, Artif. Intell. 259 (2018) 147–166.
[7] Z. Zhu, B. Peng, C. Xiong, L. Zhang, Short-term traffic flow prediction with linear conditional gaussian Bayesian network, J. Adv. Transp. 50 (6) (2016) 1111–1123.
[8] N.G. Polson, V.O. Sokolov, Deep learning for short-term traffic flow prediction, Transp. Res. Part C 79 (2017) 1–17.
[9] Z. Zhao, W. Chen, X. Wu, P.C. Chen, J. Liu, Lstm network: a deep learning approach for short-term traffic forecast, IET Intel. Transp. Syst. 11 (2) (2017) 68–75.
[10] Y. Li, Y. Zheng, H. Zhang, L. Chen, Traffic prediction in a bike-sharing system, in: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2015, pp. 33:1–33:10.
[11] Y. Ning, Y. Huang, J. Li, Q. Liu, D. Yang, W. Zheng, H. Liu, St-drn: Deep residual networks for spatio-temporal metro stations crowd flows forecast, in: Proceedings of 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, pp. 1–8.
[12] D. Chai, L. Wang, Q. Yang, Bike flow prediction with multi-graph convolutional networks, in: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2018, pp. 397–400.
[13] Z. Duan, Y. Yang, K. Zhang, Y. Ni, S. Bajgain, Improved deep hybrid networks for urban traffic flow prediction using trajectory data, IEEE Access 6 (2018) 31820–31827.
[14] S. Wang, J. Cao, P.S. Yu, Deep learning for spatio-temporal data mining: a survey, arXiv Preprint arXiv:1906.04928 (2019).
[15] J. Zhang, Y. Zheng, J. Sun, D. Qi, Flow prediction in spatio-temporal networks based on multitask deep learning, IEEE Trans. Knowl. Data Eng. (2019) 1.
[16] M. Lippi, M. Bertini, P. Frasconi, Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning, IEEE Trans. Intell. Transp. Syst. 14 (2) (2013) 871–882.
[17] Z. Liu, P. Zhou, Z. Li, M. Li, Think like a graph: real-time traffic estimation at city-scale, IEEE Trans. Mob. Comput. PP (99) (2018) 1.
[18] Y. Zheng, Urban Computing, MIT Press, 2019.
[19] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, H. Li, T-Gcn: a temporal graph convolutional network for traffic prediction, IEEE Trans. Intell. Transp. Syst. (2019) 1–11.
[20] D. Lee, D. Kulic, Y. Nakamura, Missing motion data recovery using factorial hidden Markov models, in: 2008 IEEE International Conference on Robotics and Automation, IEEE, 2008, pp. 1722–1728.
[21] X. Yi, Y. Zheng, J. Zhang, T. Li, St-mvl: filling missing values in geo-sensory time series data, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016, pp. 2704–2710.
[22] Y. Zheng, F. Liu, H.-P. Hsieh, U-air: when urban air quality inference meets big data, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2013, pp. 1436–1444.
[23] M. Beckmann, N.F. Ebecken, B.S.P. de Lima, A knn undersampling approach for data balancing, J. Intell. Learn. Syst. Appl. 7 (04) (2015) 104.
[24] R. Wang, S. Kwong, Y. Jia, Z. Huang, L. Wu, Mutual information based k-labelsets ensemble for multi-label classification, in: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2018, pp. 1–7.
[25] J. Gong, H. Kim, Rhsboost: improving classification performance in imbalance data, Comput. Stat. Data Anal. 111 (2017) 1–13.
[26] A. Khosravi, S. Nahavandi, D. Creighton, A.F. Atiya, Comprehensive review of neural network-based prediction intervals and new advances, IEEE Trans. Neural Netw. 22 (9) (2011) 1341–1356.
[27] E. Begoli, T. Bhattacharya, D. Kusnezov, The need for uncertainty quantification in machine-assisted medical decision making, Nat. Mach. Intell. 1 (1) (2019) 20.
[28] H. Wang, D.-Y. Yeung, Towards Bayesian deep learning: a framework and some existing methods, IEEE Trans. Knowl. Data Eng. 28 (12) (2016) 3395–3408.
[29] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, in: Advances in Neural Information Processing Systems, 2017, pp. 6402–6413.
[30] S.S. Rangapuram, M.W. Seeger, J. Gasthaus, L. Stella, Y. Wang, T. Januschowski, Deep state space models for time series forecasting, in: Advances in Neural Information Processing Systems, 2018, pp. 7796–7805.
[31] W. Huang, W. Jia, J. Guo, B.M. Williams, G. Shi, Y. Wei, J. Cao, Real-time prediction of seasonal heteroscedasticity in vehicular traffic flow series, IEEE Trans. Intell. Transp. Syst. 19 (10) (2017) 3170–3180.
[32] L. Lin, J.C. Handley, Y. Gu, L. Zhu, X. Wen, A.W. Sadek, Quantifying uncertainty in short-term traffic prediction and its application to optimal staffing plan development, Transp. Res. Part C 92 (2018) 323–348.
[33] Y. Liu, Z. Liu, X. Li, W. Huang, Y. Wei, J. Cao, J. Guo, Dynamic traffic demand uncertainty prediction using radio-frequency identification data and link volume data, IET Intel. Transport Syst. (2019) 1309–1317.
[34] G. Atluri, A. Karpatne, V. Kumar, Spatio-temporal data mining: a survey of problems and methods, ACM Comput. Surv. 51 (4) (2018) 1–37.
[35] Y. Zheng, Trajectory data mining: an overview, ACM Trans. Intell. Syst. Technol. (TIST) 6 (3) (2015) 1–41.
[36] T.M.T. Do, O. Dousse, M. Miettinen, D. Gatica-Perez, A probabilistic kernel method for human mobility prediction with smartphones, Pervasive Mob. Comput. 20 (2015) 13–28.
[37] D.H. Douglas, T.K. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Class. Cartogr. (2011) 15–28.
[38] E. Keogh, S. Chu, D. Hart, M. Pazzani, An online algorithm for segmenting time series, in: Proceedings of the IEEE International Conference on Data Mining, IEEE, 2001, pp. 289–296.
[39] N. Meratnia, A. Rolf, Spatiotemporal compression techniques for moving point objects, in: Proceedings of the International Conference on Extending Database Technology, Springer, 2004, pp. 765–782.
[40] J. Yuan, Y. Zheng, X. Xie, G. Sun, T-Drive: enhancing driving directions with taxi drivers' intelligence, IEEE Trans. Knowl. Data Eng. 25 (1) (2012) 220–232.
[41] Y. Zheng, Y. Chen, Q. Li, X. Xie, W.-Y. Ma, Understanding transportation modes based on GPS data for web applications, ACM Trans. Web 4 (1) (2010) 1–36.
[42] Y. Zheng, L. Liu, L. Wang, X. Xie, Learning transportation mode from raw GPS data for geographic applications on the web, in: Proceedings of the 17th International Conference on World Wide Web, ACM, 2008, pp. 247–256.
[43] J.S. Greenfeld, Matching GPS observations to locations on a digital map, in: Proceedings of the 81th Annual Meeting of the Transportation Research Board, Transportation Research Board, 2002, pp. 164–173.
[44] W. Chen, M. Yu, Z. Li, Y. Chen, Integrated vehicle navigation system for urban applications, in: Proceedings of International Conference Global Navigation Satellite System, 2003, pp. 15–22.
[45] H. Yin, O. Wolfson, A weight-based map matching method in moving objects databases, in: Proceedings of the International Conference on Scientific & Statistical Database Management, IEEE, 2004, pp. 437–438.
[46] M.A. Quddus, R.B. Noland, W.Y. Ochieng, A high accuracy fuzzy logic based map matching algorithm for road transport, J. Intell. Transp. Syst. 10 (3) (2006) 103–115.
[47] O. Pink, B. Hummel, A statistical approach to map matching using road network geometry, topology and vehicular motion constraints, in: Proceedings of the 11th International IEEE Conference on Intelligent Transportation, IEEE, 2008, pp. 862–867.
[48] P. Newson, J. Krumm, Hidden markov map matching through noise and sparseness, in: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2009, pp. 336–343.
[49] J. Yuan, Y. Zheng, C. Zhang, X. Xie, G.-Z. Sun, An interactive-voting based map matching algorithm, in: Proceedings of the 2010 Eleventh International Conference on Mobile Data Management, IEEE Computer Society, 2010, pp. 43–52.
[50] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, Y. Huang, Map-matching for low-sampling-rate GPS trajectories, in: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances In Geographic Information Systems, ACM, 2009, pp. 352–361.
[51] H. Alt, A. Efrat, G. Rote, C. Wenk, Matching planar maps, J. Algo. 49 (2) (2003) 262–283.
[52] S. Brakatsoulas, D. Pfoser, R. Salas, C. Wenk, On map-matching vehicle tracking data, in: Proceedings of the 31st International Conference on Very Large Data Bases, VLDB Endowment, 2005, pp. 853–864.
[53] X. Yi, J. Zhang, Z. Wang, T. Li, Y. Zheng, Deep distributed fusion network for air quality prediction, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018, pp. 965–973.
[54] X. Li, L. Peng, Y. Hu, J. Shao, T. Chi, Deep learning architecture for air quality predictions, Environ. Sci. Pollut. Res. 23 (22) (2016) 22408–22417.
[55] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, T. Chi, Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation, Environ. Pollut. 231 (2017) 997–1004.

[56] S. Du, T. Li, X. Gong, Z. Yu, Y. Huang, S.-J. Horng, A hybrid method for traffic flow forecasting using multimodal deep learning, arXiv preprint arXiv:1803.02099 (2018).
[57] Y. Wu, H. Tan, L. Qin, B. Ran, Z. Jiang, A hybrid deep learning based traffic flow prediction method and its understanding, Transp. Res. Part C 90 (2018) 166–180.
[58] D. Wang, Y. Yang, S. Ning, Deepstcl: A deep spatio-temporal convlstm for travel demand prediction, in: 2018 International Joint Conference on Neural Networks, IEEE, 2018, pp. 1–8.
[59] J. Ke, H. Zheng, H. Yang, X.M. Chen, Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach, Transp. Res. Part C 85 (2017) 591–608.
[60] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, Deep multi-view spatial-temporal network for taxi demand prediction, arXiv:1802.08714. 2018.
[61] Y. Zheng, L. Capra, O. Wolfson, H. Yang, Urban computing: concepts, methodologies, and applications, ACM Trans. Intell. Syst. Technol. 5 (3) (2014) 38–55.
[62] Y. Akagi, T. Nishimura, T. Kurashima, H. Toda, A fast and accurate method for estimating people flow from spatiotemporal population data, in: Proceedings of the International Joint Conferences on Artificial Intelligence, 2018, pp. 3293–3300.
[63] W. Jin, Y. Lin, Z. Wu, H. Wan, Spatio-temporal recurrent convolutional networks for citywide short-term crowd flows prediction, in: Proceedings of the 2nd International Conference on Compute and Data Analysis, ACM, 2018, pp. 28–35.
[64] S.E. Said, D.A. Dickey, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 71 (3) (1984) 599–607.
[65] B.M. Williams, P.K. Durvasula, D.E. Brown, Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models, Transp. Res. Rec. 1644 (1) (1998) 132–141.
[66] N. Zhang, Y. Zhang, H. Lu, Seasonal autoregressive integrated moving average and support vector machine models: prediction of short-term traffic flow on freeways, Transp. Res. Rec. 2215 (1) (2011) 85–92.
[67] F.G. Habtemichael, M. Cetin, Short-term traffic flow rate forecasting based on identifying similar traffic patterns, Transp. Res. Part C 66 (2016) 61–78.
[68] W. Jiang, L. Zhang, Geospatial data to images: a deep-learning framework for traffic forecasting, Tsinghua Sci. Technol. 24 (1) (2019) 52–64.
[69] Y. Liu, Z. Liu, R. Jia, Deeppf: a deep learning based architecture for metro passenger flow prediction, Transp. Res. Part C 101 (2019) 18–34.
[70] X. Ma, J. Zhang, B. Du, C. Ding, L. Sun, Parallel architecture of convolutional bi-directional lstm neural networks for network-wide metro ridership prediction, IEEE Trans. Intell. Transp. Syst. (2018) 1–11.
[71] E. Walraven, M.T. Spaan, B. Bakker, Traffic flow optimization: a reinforcement learning approach, Eng. Appl. Artif. Intell. 52 (2016) 203–212.
[72] Z. Jiang, W. Fan, W. Liu, B. Zhu, J. Gu, Reinforcement learning approach for coordinated passenger inflow control of urban rail transit in peak hours, Transp. Res. Part C 88 (2018) 1–16.
[73] L. Wang, X. Geng, X. Ma, F. Liu, Q. Yang, Crowd flow prediction by deep spatio-temporal transfer learning, arXiv:1802.00386. 2018.
[74] B. Wang, Z. Yan, J. Lu, G. Zhang, T. Li, Road traffic flow prediction using deep transfer learning, in: Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference (FLINS 2018), 11, World Scientific, 2018, pp. 331–338.
[75] R.J. Hyndman, Y. Khandakar, Automatic time series for forecasting: the forecast package for R, Monash University, Department of Econometrics and Business Statistics, 2007.
[76] G.A. R. J. Hyndman, Forecasting: principles and practice, Otexts.com (2014).
[77] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[78] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[79] R.J. Williams, D. Zipser, A learning algorithm for continually running fully recurrent neural networks, Neural Comput. 1 (2) (1989) 270–280.
[80] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, in: Advances in Neural Information Processing Systems, 2014, pp. 3104–3112.
[81] S. Jiang, G.A. Fiore, Y. Yang, J. Ferreira Jr, E. Frazzoli, M.C. González, A review of urban computing for mobile phone traces: current methods, challenges and opportunities, in: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, ACM, 2013, pp. 1–9.
[82] F. Calabrese, L. Ferrari, V.D. Blondel, Urban sensing using mobile phone network data: a survey of research, ACM Comput. Surv. 47 (2) (2015) 1–20.
[83] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, J. Zhang, Urban traffic prediction from spatio-temporal data using deep meta learning, in: Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2019.
[84] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[85] J.A. Deri, F. Franchetti, J.M. Moura, Big data computation of taxi movement in new york city, in: 2016 IEEE International Conference on Big Data, IEEE, 2016, pp. 2616–2625.
[86] X. Zhan, Y. Zheng, X. Yi, S.V. Ukkusuri, Citywide traffic volume estimation using trajectory data, IEEE Trans. Knowl. Data Eng. 29 (2) (2017) 272–285.
[87] H. Yao, X. Tang, H. Wei, G. Zheng, Z. Li, Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction, in: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, 2019.
[88] D. Nam, H. Kim, J. Cho, R. Jayakrishnan, A model based on deep learning for predicting travel mode choice, in: Proceedings of the Transportation Research Board 96th Annual Meeting Transportation Research Board, Washington, DC, USA, 2017, pp. 8–12.