Chapter 5
Hybrid Statistical and Machine Learning Methods for Road Traffic Prediction: A Review and Tutorial

Bdoor Alsolami, Rashid Mehmood, and Aiiad Albeshri

5.1 Introduction

Mobility is one of the major dimensions of smart city design and development. Many approaches have been proposed to address smart mobility–related challenges [1]—for example, social media–based approaches [2–4], location-based services [5, 6], telematics [7], modeling and simulation–based approaches [8, 9], approaches based on vehicular networks (VANETs) and systems [10–12], autonomic mobility management [13–15], autonomous driving [16], mobility in emergency situations [17–22], approaches to improve urban logistics [2, 23–25], and big data–based approaches [2–4, 26, 27]. A recent book has covered a number of topics related to smart cities, including smart mobility [28].

Traffic flow modeling and prediction play important roles in smart city transportation systems. The modeling of transportation traffic is usually done by using data acquired through various sensors [10], such as inductive loops and Motorway Incident Detection and Automatic Signalling (MIDAS) [9], use of surveys [17], vehicular ad hoc networks [10], and social networks [2–4]. Various methods are in practice to model and predict traffic, including mathematical modeling [9, 17, 18], simulations [8, 19–21], and deep learning [22]. Accurate prediction of the transport network state can improve information services for travelers and help them

B. Alsolami (✉) · A. Albeshri
Department of Computer Science, FCIT, King Abdulaziz University, Jeddah, Saudi Arabia
e-mail: balsolami0069@stu.kau.edu.sa; aaalbeshri@kau.edu.sa

R. Mehmood
High Performance Computing Center, King Abdulaziz University, Jeddah, Saudi Arabia
e-mail: RMehmood@kau.edu.sa

R. Mehmood et al. (eds.), *Smart Infrastructure and Applications*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-030-13705-2_5

to make informed travel decisions. Furthermore, precise prediction of road traffic can improve road safety by decreasing congestion problems, air pollution, traffic costs, and accidents [29]. The predicted information leads to good planning of the traffic infrastructure.

Today, the transportation sector has truly entered the big data era. The rapid increase of Global Positioning System (GPS) device use in vehicles and on smartphones has provided opportunities for researchers to use transport data for studying traffic states and solving traffic problems. In addition, social networking applications such as Twitter and Facebook have become sources of traffic data because most people share their various statuses and environmental conditions, including the status of road traffic. Many governments make their transport data available so that researchers can analyze it and propose solutions to traffic problems [30].

In the last few years, many research attempts have emerged to provide accurate and timely traffic flow prediction models. However, most of the existing prediction models are based only on a single prediction method, such as statistical methods or machine learning methods. Neither statistical nor machine learning methods can completely capture traffic flow patterns, because of the complex relationship of traffic data. Statistical methods provide good performance when traffic data have a linear relationship, while machine learning methods work well with nonlinear traffic data [31]. Furthermore, most of the existing prediction models are built on stand-alone models because of the data and compute-intensive nature of the complex models. There is a need for novel prediction methods that provide higher accuracy for prediction of traffic with diverse characteristics. Moreover, there is a need to use distributed and parallel big data platforms for traffic prediction [32]. This chapter:

– Gives a review of traffic flow prediction and modeling methods.
– Discusses the limitation of each method.
– Introduces a review of various types of traffic data sources.
– Describes notable big data analysis tools.
– Describes a hybrid method for road traffic prediction and provides a tutorial on the process of hybrid traffic flow prediction. This method is based on the autoregressive integrated moving average (ARIMA) and support vector machine (SVM) methods.

The rest of the chapter is organized as follows. Section 5.2 reviews methods for road traffic analysis and prediction, and discusses a classification for the methods. We also discuss the limitation of using a single method type. In Sect. 5.3, we discuss a classification of the available traffic data sources. Big data analysis tools are introduced in Sect. 5.4. Section 5.5 gives a tutorial on the process of hybrid traffic flow prediction. In Sect. 5.6, we provide our conclusions on the chapter.

5.2 Traffic Flow Prediction and Modeling Methods

Road traffic modeling and prediction are important issues that face both individuals and governments because of increases in traffic congestion, accidents, and air pollution [33]. Because of their importance, many researchers have published several methods for traffic flow prediction. In general, and from the academic literature, we can categorize the existing methods into two types: long-term prediction and short-term prediction. Long-term prediction accuracy is affected by external factors such as weather conditions, road construction, and changes in road infrastructure. For these reasons, long-term prediction is not widely used. Conversely, short-term prediction has been widely used for road traffic prediction. Many methods have been proposed for short-term traffic prediction. As shown in Fig. 5.1, short-term prediction can be categorized into three types: statistically based methods, machine learning methods, and hybrid methods.

5.2.1 Statistical Methods

In this section, we discuss Kalman filtering (KF) and ARIMA methods.

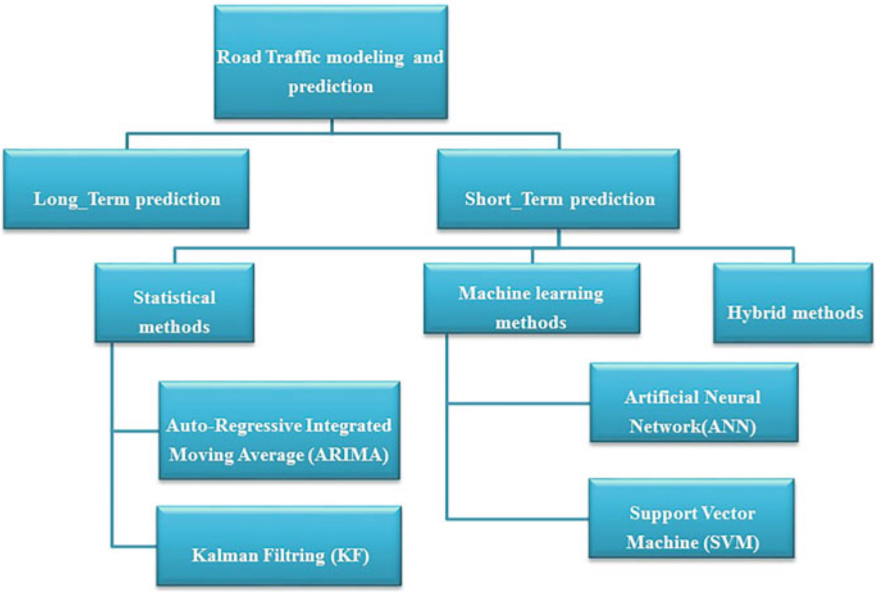


Fig. 5.1 Classification of traffic flow prediction methods

Kalman Filtering

Kalman filtering is a statistical method involving an algorithm that uses a set of measurements observed over time, which contain noise, and predicted outputs of unknown variables. In a KF approach, the estimated future state is based only on the estimated state of the previous time step [34]. Wang et al. [35] carried out research on the possibility of using KF for traffic flow prediction. Jinxing et al. [36] used a KF method to remove errors and redundancy from data sets. They concluded that filtering can increase the accuracy of the prediction model. Ahmad et al. [37] used social network traffic data (Twitter data) with a KF method for arrival time prediction. The results showed that the KF model has the ability to forecast vehicle arrival time with reasonable accuracy.

Autoregressive Integrated Moving Average

An ARIMA model is a tool for predicting the future values of time series [38]. Since it can represent traffic flow by a time series, we can fit an ARIMA model to predict future traffic flow. ARIMA (p, d, q) has been used for prediction by many researchers in many sectors, such as economics and transportation. In an ARIMA model the training data preprocessing to exclude the error data. After that, the data are classified into multiple data sets based on multiple time periods, and predictions are made through extraction of the correlation between sequences [39]. An ARIMA model is based on linear analysis [40]. Wang et al. [40] used ARIMA for short-term traffic flow prediction (see Sect. 5.2.4), and their ARIMA model provided good accuracy. Because the pattern of traffic flow appears as a seasonal pattern according to peak and off-peak traffic conditions, seasonal ARIMA (SARIMA) models are very suitable for modeling traffic flow behavior [41]. Kumar et al. [41] used 3 days of data to forecast the flow for the next day. The advantage of this type of model is the limited data set used for prediction.

5.2.2 Machine Learning Methods

Artificial Neural Networks

Artificial neural networks (ANNs) are the most widely used type of algorithm for traffic flow prediction. ANNs are a set of statistical learning algorithms. They have the ability to deal with complex problems and with missing and noisy data [42, 43]. ANNs contain multiple layers; the most widely used model is a multilayer perceptron (MLP). Changqing et al. [43] presented a classification and prediction framework for taxi hailing. They used K-means clustering to divide the data set into several clusters and used a neural network based on the clustering result to generate the prediction result. Florido et al. [44] used an NN algorithm for congestion prediction in road networks. Achieving an accurate prediction result with a minimum value of square error (MSE) is the objective of an ANN. To achieve this aim, many researchers have developed a number of algorithms, such as the back propagation neural networks (BPNNs) proposed by Park and Rilett [45]. Pamuła [46] conducted research on analysis of traffic flow data using BPNNs. The traffic flow was represented by four classes of time series. The results showed the capability of neural networks to be used in intelligent transportation systems (ITSs). In ANNs, there are different processing elements called neurons; each of them takes multiple inputs and, on the basis of an internal weighting, only one output is produced. The neurons are organized into layers [39]. Ban et al. [47] used new neural networks named extreme learning machines (ELMs) to predict traffic states. The algorithm provided good performance in comparison with other prediction algorithms. However, time is the cost of this algorithm. ANNs need parallel architecture to process large data volumes in small amounts of time, and the cost of using ANNs is the large number of training data sets, which need large amounts of storage.

Support Vector Machines

An SVM comes under the statistical learning algorithm category and is used in many studies for traffic flow prediction. The SVM process involves getting the optimal separating hyperplane. It can work with any number of dimensions [48]. Deshpande et al. [48], in their research, presented the potential use of SVM for traffic flow prediction. Zhou et al. [49] conducted research on traffic flow analysis based on GPS data on floating cars using a least squares (LS)–SVM method. Li et al. [50] employed SVM for bus arrival time prediction based on GPS data. Their results indicated that SVM was robust, adaptive, and able to provide good prediction accuracy.

5.2.3 Limitations of Using a Single Prediction Method

Our review reveals that several methods have been used for traffic flow prediction and modeling. These methods are classified into three categories: statistical, machine learning, and hybrid methods. The statistical methods work well only with linear traffic flow, while the machine learning methods have the ability to work with nonlinear traffic flow. Each of the above methods works well under specific conditions; when the conditions change, the performance of the predictive method is affected and the accuracy decreases.

KF exhibits high prediction accuracy, but it is a linear prediction model and it is not suitable for nonlinear traffic flow. Furthermore, it is not adaptive for dynamic traffic conditions. The KF method has limited accuracy when it deals with noisy traffic data [51].

ARIMA is a popular and widely used statistical method for traffic flow prediction. However, it has the disadvantage of being unable to capture rapid changes in traffic data. In addition, a SARIMA model requires a lot of time for estimation of parameters. ARIMA provides good performance only with static traffic conditions and does not reflect the dynamics.

Many kinds of neural networks have been used for traffic flow prediction, such as fuzzy neural models (FNMs), genetic algorithms (GAs), and multilayer perceptrons (MLPs). Large volumes of historical data and many computational resources are needed. Neutral networks are suitable for nonlinear features. Neural networks constantly demonstrate high accuracy but take more time for training and require large storage space. They are based on training using part of the historical data to find the relationship between the input and the output. Moreover, ANNs are not able to find an optimal solution for nonconvex problems.

SVMs can overcome the limitations of ANNs because they are able to map nonlinear problems in a low dimension to linear problems in a high dimension. However, they fail to give high prediction accuracy when the data contain noise.

5.2.4 Hybrid Traffic Flow Analysis and Prediction Methods

The hybrid prediction model combines the advantages of both statistical and machine learning methods. In previous studies, hybrid models have outperformed both statistical and machine learning models. However, the costs of using a hybrid model are computational complexity and large storage needs. Most of the existing hybrid models are performed using stand-alone platforms. Some existing hybrid models have been performed using a distributed platform such as Hadoop; however, the current state of work on hybrid models in distributed environments is very basic. Apache Spark has recently emerged as another big data platform with much better performance than Hadoop [52].

Wang et al. [40] conducted research on the potential use of a combination of ARIMA and SVM for traffic flow prediction. They found the characteristics of the data by using feature analysis and, on the basis of the analysis results, they used hybrid ARIMA and SVM methods. The results showed that the hybrid methods did improve the prediction accuracy.

Meng et al. [53] carried out research on the potential use of a hybrid K-nearest neighbor (K-NN) method with a balanced binary (AVL) tree for short-term traffic flow prediction to increase the accuracy of the prediction result. The results showed that the hybrid K-NN method with AVL increased the speed of the search and the accuracy outperformed both K-NN and AVL.

Xie et al. [54] proposed a novel hybrid prediction model combining the advantages of an ARIMA model and a periodical moving average (PMA) model. The model was evaluated using real-time data as well as historical data. The results showed that the forecasting performance was improved by use of the hybrid prediction model.

Li et al. [55] applied both ARIMA and a radial basis function ANN (RBF-ANN) for traffic flow prediction. The results indicated that the hybrid model had better performance than use of a single ARIMA or RBF-ANN.

5.3 Transportation Data Sources

Because of the rapid increase in the population and the numbers of vehicles, several problems have emerged in transportation systems, such as traffic congestion and traffic accidents. More recently, many types of transportation data have emerged and can be used for studying traffic status and solving transportation-related issues. Governments, city planners, and researchers have used these data for various purposes such as predicting traffic flow and identifying traffic congestion and traffic accidents. To find out the available traffic-related data, we widely review technologies used for collection and acquisition of traffic data.

On the basis of the work done by Dabiri and Heaslip [30], we can classify traffic data sources into six categories (Fig. 5.2). These are explained in Sects. 5.3.1–5.3.6.

5.3.1 Traffic Flow Sensors

Traffic flow sensors are devices for capturing the passage of vehicles over a particular road so as to capture traffic parameters. They are classified into two categories: the first category is sensors that are attached to the road pavement or road surface, such as inductive loop and magnetometers sensors; the second category is sensors that are placed above the road surface, such as infrared sensors, video image processors, and microwave radar. Both ultrasonic sensors and passive infrared sensors are widely used to collect unprocessed data, which is used by most of the existing prediction models. Prathilothamai et al. [56] proposed a system for road traffic prediction; they used traffic data collected by ultrasonic as well as passive infrared sensors, and they suggested the use of traffic video in future studies.

Fig. 5.2 Classification of traffic data sources

5.3.2 Video Image Processors

A video image processor (VIP) is a camera mounted on poles on the road pavement or traffic signal for taking images or video of passing vehicles. Microprocessors store and process these images and videos to apply computer vision algorithms for extracting traffic parameters that are used in traffic management operations.

5.3.3 Probe Vehicles and People Data

Traffic sensors and VIPs capture traffic data only in a limited area and location, which leads to data collection that is unrepresentative of the network as a whole. To overcome this limitation, individuals' vehicles equipped with GPS devices can be used for collecting representative traffic data. Also, smartphones can used to capture spatial data and can be used for tracking vehicles and people's trajectories.

Floating car data (FCD) is an important source of traffic data in smart cities and the transportation sector. It is a set of GPS entries containing information about driving status. By using FCD, traffic congestion can be identified, travel time can be computed, and traffic flow can be predicted. Castro et al. [57] proposed a method to predict future traffic conditions. They evaluated their method by using large-scale taxi GPS data obtained from around 5000 taxis in Hangzhou, China, over a period of a month (February 2010). Wang et al. [58] proposed a three-phase framework to explore the congestion correlation between road segments from three data sources: GPS trajectories of taxis, road network data, and point of interest (POI) data.

5.3.4 Social Network Data

Today, millions of people share data and communicate using social networks such as Twitter, Facebook, and Instagram. People share their images, locations, and video on social media networks. These data contain hidden knowledge and can be used in transportation systems. In social media, traffic data acquisition is performed using an application programming interface (API) by using queries to access historical and real-time information. Petalas et al. [29] proposed big data architecture for road traffic prediction using multiple sources of data. The data they used for traffic prediction were urban data and social media data. They utilized data from multiple heterogeneous sources. The results showed that the performance of the prediction model depends on the type of traffic data used.

5.3.5 Smart Card Data

Smart cards are one traffic data source, using technology for capturing transit data and passenger behavior. There are two types of smart cards: automated passenger counter (APC) cards and automated fare collection (AFC) cards. They are designed for controlling passenger movement in and out of buses and subways.

5.3.6 Environmental Data

Traffic exhibits sudden shifts due to various factors such as weather status. Meteorological data—including temperature, wind speed, and precipitation—must be taken into consideration when analyzing traffic flow. There are numerous public websites that provide metrological data, such as the [US] National Weather Service (NWS; https://www.weather.gov) and the [US] National Climatic Data Center (NCDC; https://www.ncdc.noaa.gov).

5.4 Big Data Analysis and Processing Tools

5.4.1 Hadoop

Hadoop is a powerful open-source framework (developed by Apache) used by many organizations and companies for storing, analyzing, and processing big data. Hadoop provides good availability and scalability of data [59]. Also, Hadoop is considered reliable and able to detect bugs. It is a distributed tool comprising two components: MapReduce and the Hadoop distributed file system (HDFS). Large data sets are divided into small pieces and stored in blocks of 64 MB in size; each block is called a DataNode. Those DataNodes are indexed by NameNode in HDFS [60]. MapReduce is a big data processing tool for distributed computing, which was developed by Google. MapReduce works on a divide-and-conquer principle, dividing big data problems into small problems and processing them in parallel. Hadoop also contains a data warehousing application called Hive. Hive uses structured query language (SQL) and HiveQL as query languages.

Hadoop_GIS Tool

The main components of a traditional geographic information system (GIS) are a database for storage and an analyzing model. The data are represented in a relational database or geodatabase. The data are transported to the ArcGIS environment for analysis, and the ArcGIS toolbox uses spatial analysis jobs. The traditional GIS is single threaded, which means there is only one module to process and analyze the data stream. For this reason, the traditional GIS is not suitable for processing large data sets; the cost of processing a large data set is very high and it takes a lot of time [60]. To overcome the limitations of the traditional GIS, the Hadoop_GIS tool was proposed by Deng and Bai [60]. The Hadoop_GIS tool is a package containing a spatial framework and geoprocessing tools. The spatial framework consists of functions such as ST_Geometry. The Hadoop_GIS tool adds geometry functions and a geoprocessing toolbox for Hadoop. Hadoop_GIS is considered more efficient than the traditional GIS for processing large data sets; however, it is just a processing model with no visualization feature.

5.4.2 Apache Spark

Apache Spark is an open-source framework designed for cluster computing. It is designed to be fast and general purpose and to overcome the limitations of MapReduce. Since time is a very important factor in big data processing, Spark supports in-memory processing, which is faster than disk-based processing [61]. This feature make it faster to query big data than in a traditional disk-based engine such as Hadoop. Spark can run multiple applications in the same engine. Also, Spark has the ability to process different types of workloads that need separate systems, including queries, iterative algorithms, and streaming. This feature leads to a reduced management cost of multiple big data tools. Furthermore, the accessibility of Spark is considered very high because it provides easy application programming interfaces (APIs) in Java, Python, Scala, and SQL. Spark can be integrated with any big data tools such as Hadoop. Moreover, Spark supports additional machine learning tools such as M-Lib, a tool for graph processing (named GraphX), a tool for streaming processing (named Spark Streaming), and Spark SQL for processing structured data. All Spark components and supported tools are shown in Fig. 5.3. When one compares Spark with other big data tools such as Hadoop, Spark is faster and has greater ability to process and write data than Hadoop [61].

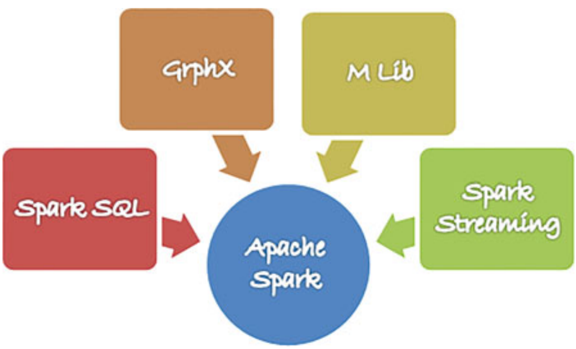


Fig. 5.3 Apache Spark components

Apache Spark features Apache Spark is considered one of the high-performance frameworks that are designed for cluster computing and real-time streaming processing [56]. It is distinguishable from all other available tools because:

- It is a general purpose framework.
- It is easy to install and configure.
- It is simple to use because it support APIs with several programming languages (such as Java, Scala, and SQL).
- It is faster than Hadoop because it supports in-memory computing.
- It supports Java and Scala, which are powerful languages for object-oriented programming.
- It has the ability to aggregate multiple data sets from different sources.
- It has the ability to join and work with other big data tools such as Hadoop.

5.5 Process of Hybrid Prediction

Hybrid methods for road traffic prediction and analysis combine both statistical and machine learning analysis. Researchers have proposed many hybrid methods with different combinations of prediction methods for several kinds of traffic data. In this section we introduce a tutorial for combining statistical analysis and machine learning analysis. Furthermore, we propose the methodology of a hybrid prediction model combining a widely used ARIMA model from the statistical category with SVM from the machine learning category.

5.5.1 Statistical Analysis

In this section, we discuss the ARIMA model as an example of a time series–forecasting method.

Autoregressive Integrated Moving Average

ARIMA was proposed by Box and Jenkins (1976) and is also called the Box–Jenkins model [55]. It is widely used in predicting stationary time series. If the series is not stationary, then it is transformed into a stationary series by differencing. The order of differencing is denoted by the parameter d. The steps for predicting traffic flow by using an ARIMA model are shown in Fig. 5.4, and an ARIMA flow chart is shown in Fig. 5.5.

1. *Data visualization*: The goal of this step is to explore any trend in the data and to decide what type of ARIMA we should use. If there is a seasonal trend in the data, we should use a seasonal ARIMA model; if there is no seasonal trend, we can use a normal ARIMA model.

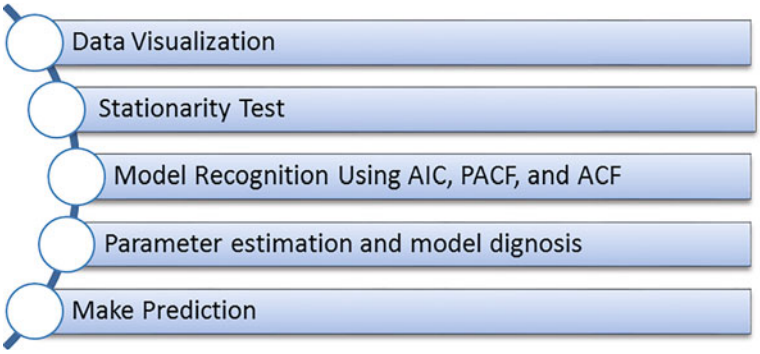
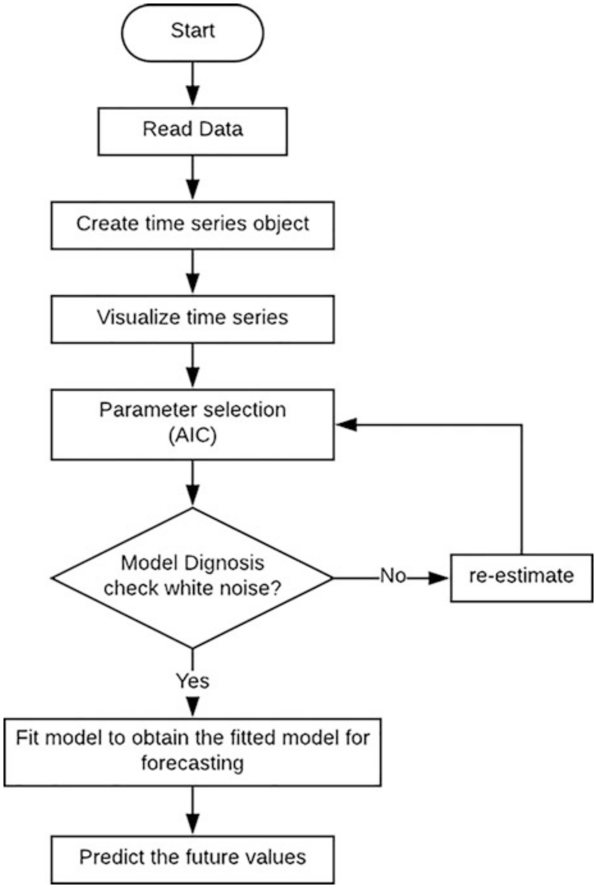


Fig. 5.4 Autoregressive integrated moving average (*ARIMA*) framework

Fig. 5.5 Autoregressive integrated moving average (*ARIMA*) flow chart



2. *Stationarity test*: Series stationary means that the mean and the variance of the series should not be a function of time. The mean of the series should not increase over time. The series is considered not stationary if there is a varying spread of the data over time. Also, the covariance of the series should not be a function of time. The covariance should be constant with time; if the spread of the data becomes closer as the time increases, then the series violates the stationary property.
3. *Model recognition*: The parameters p, d, and q are determined on the basis of the Akaike information criterion (AIC) minimum criterion, autocorrelation function (ACF), or partial autocorrelation function (PACF).
4. *Parameter estimation and model diagnosis*: The fourth step in building an ARIMA model is checking the accuracy of the model by a diagnostic test such as the Q statistic [62]. Then, we see if the chosen model and its parameters fit the data reasonably or not. If not, the parameters and the model must be re-estimated.
5. *Making the prediction*: The chosen model is used with suitable parameters to predict traffic flow.

5.5.2 Machine Leaning Analysis

Support Vector Machine

This is an advanced machine learning method and is widely used for short-term prediction such as travel time prediction and traffic flow prediction [40]. The SVM method has an algorithm called support vector regression (SVR), which is used to solve classification and regression problems [51]. SVM has the ability to predict unknown data on the basis of the given pattern. It is superior to ANN in terms of generalization and learning ability [63].

Suppose we have the training data set: $\{(x_1, y_1), \ldots (x_n, y_n)\}$. SVR can find the function that represents the relationship of x and y; also, the function gives the forecasted value of the new x. The SVR function can be represented as:

$$f(x) = w.\, \phi\,(x_i) + b \tag{5.1}$$

where w and b are the final study variables of SVR, and $\phi\,(x_i)$ is nonlinear mapping to high-dimensional space.

5.5.3 Hybrid Autoregressive Integrated Moving Average–Support Vector Machine Methodology

As we know, most real-world time series contain linear and nonlinear correlation structures. Moreover, neither ARIMA nor SVM can capture all characteristics of traffic flow patterns and provide reasonable prediction accuracy. Thus, we need to combine ARIMA and SVM to improve the prediction accuracy. Traffic flow data contain nonlinear time series with white noise and linear time series, which can represented as:

$$Y_t = L_t + N_t \tag{5.2}$$

where L_t and N_t denote the linear and nonlinear parts, respectively. Thus, we can represent the hybrid prediction model as follows:

1. *Step 1*: Fit the ARIMA model to the linear time series, and the corresponding predicted $\acute{L}_t$ at time t is obtained.
2. *Step 2*: Compute the residual e_t from the ARIMA model as follows:

$$e_t = Y_t - \acute{L}_t \tag{5.3}$$

3. *Step 3*: Model the residual e_t by using the SVM model. Thus, the nonlinear traffic flow time series can be captured. $\acute{N}_t$ is the result of the prediction of the SVM model.
4. *Step 4*: Finally, the overall prediction value of the traffic flow time series can estimated as:

$$\acute{Y}_t = \acute{L}_t + \acute{N}_t \tag{5.4}$$

A flow diagram of the hybrid model is shown in Fig. 5.6.

5.5.4 Model Evaluation

To measure the model performance and the accuracy of the prediction results in order to compare the proposed model with other models, the following performance indexes can be used:

1. Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| y(t) - y'(t) \right| \tag{5.5}$$

2. Mean square error (MSE):

$$\text{MSE} = \frac{1}{N} \sqrt{\sum_{i=1}^{N} (y(t) - y'(t))\, 2} \tag{5.6}$$

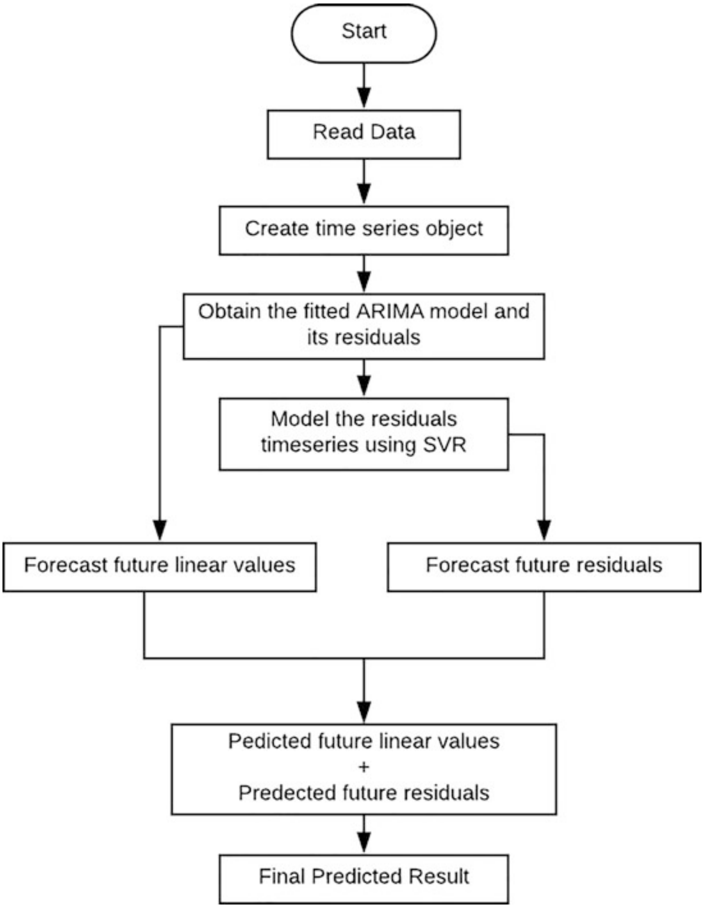


Fig. 5.6 Hybrid autoregressive integrated moving average (*ARIMA*)–support vector machine (*SVM*) model

3. Mean relative error (MRE):

$$\mathrm{MRE} = \frac{1}{N}\left(\sum_{i=1}^{N}\frac{\left|y(t) - y'(t)\right|}{y(t)}\right) \times 100\% \tag{5.7}$$

5.6 Conclusions

In this work, we have reviewed the widely used traffic flow prediction methods along with the limitations of each method. The traffic data sources have been classified into six categories and discussed. We have also reviewed high-performance big data analysis tools with their advantages and disadvantages. Finally, we have introduced a tutorial on the process of hybrid modeling for traffic flow prediction. The hybrid model combines both statistical and machine learning methods. In the proposed hybrid model, we combine ARIMA for linear time series and SVM for nonlinear components. The accuracy of the hybrid model can be measured using the discussed performance metrics, and this is our future work.

Acknowledgements The work carried out in this chapter is supported by the High Performance Computing (HPC) Center at the King Abdulaziz University, Jeddah.

References

1. Büscher, M., Coulton, P., Efstratiou, C., Gellersen, H., Hemment, D., Mehmood, R., Sangiorgi, D.: Intelligent mobility systems: some socio-technical challenges and opportunities. In: Communications Infrastructure. Systems and Applications in Europe, Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST 16. pp. 140–152 (2009)
2. Suma, S., Mehmood, R., Albugami, N., Katib, I., Albeshri, A.: Enabling next generation logistics and planning for smarter societies. In: Procedia Computer Science. pp. 1122–1127 (2017)
3. Suma, S., Mehmood, R., Albeshri, A.: Automatic event detection in smart cities using big data analytics. In: International Conference on Smart Cities, Infrastructure, Technologies and Applications (SCITA 2017): Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Volume 224. pp. 111–122. Springer, Cham (2018)
4. Alomari, E., Mehmood, R.: Analysis of tweets in Arabic language for detection of road traffic conditions. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Volume 224. pp. 98–110. Springer, Cham (2018)
5. Ayres, G., Mehmood, R., Mitchell, K., Race, N.J.P.: Localization to enhance security and services in Wi-Fi networks under privacy constraints. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Volume 16. pp. 175–188. Springer (2009)
6. Ayres, G., Mehmood, R.: LocPriS: a security and privacy preserving location based services development framework. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNAI, Volume 6279, Part 4. pp. 566–575. Springer (2010)
7. Elmirghani, J.M.H., Badic, B., Li, Y., Liu, R., Mehmood, R., Wang, C., Xing, W., Garcia Zuazola, I.J., Jones, S.: IRIS: an intelligent radio-fibre telematics system. In: Proceedings of the 13th ITS World Congress, London, 8–12 October (2006)
8. Ayres, G., Mehmood, R.: On discovering road traffic information using virtual reality simulations. In: 11th International Conference on Computer Modelling and Simulation, UKSim 2009. pp. 411–416 (2009)
9. Mehmood, R.: Towards understanding intercity traffic interdependencies. In: 14th World Congress on Intelligent Transport Systems, ITS 2007. pp. 1793–1799. ITS America, Beijing (2007)
10. Mehmood, R., Nekovee, M.: Vehicular ad hoc and grid networks: discussion, design and evaluation. In: 14th World Congress on Intelligent Transport Systems, ITS 2007. pp. 1555–1562. ITS America, Beijing (2007)

11. Gillani, S., Shahzad, F., Qayyum, A., Mehmood, R.: A survey on security in vehicular ad hoc networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 59–74 (2013)
12. Alvi, A., Greaves, D., Mehmood, R.: Intra-vehicular verification and control: a two-pronged approach. In: 7th IEEE International Symposium on Communication Systems, Networks and Digital Signal Processing, CSNDSP 2010. pp. 401–405 (2010)
13. Schlingensiepen, J., Nemtanu, F., Mehmood, R., McCluskey, L.: Autonomic transport management systems—enabler for smart cities, personalized medicine, participation and industry Grid/Industry 4.0. In: Intelligent Transportation Systems—Problems and Perspectives, Volume 32 of the series Studies in Systems, Decision and Control. pp. 3–35. Springer (2016)
14. Schlingensiepen, J., Mehmood, R., Nemtanu, F.C.: Framework for an autonomic transport system in smart cities. Cybern. Inf. Technol. **15**, 50–62 (2015)
15. Schlingensiepen, J., Mehmood, R., Nemtanu, F.C., Niculescu, M.: Increasing sustainability of road transport in European cities and metropolitan areas by facilitating autonomic road transport systems (ARTS). In: Wellnitz, J., Subic, A., Trufin, R. (eds.) Sustainable Automotive Technologies 2013 Proceedings of the 5th International Conference ICSAT 2013, pp. 201–210. Springer, Ingolstadt (2014)
16. Alam, F., Mehmood, R., Katib, I.: D2TFRS: an object recognition method for autonomous vehicles based on RGB and spatial values of pixels. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Volume 224. pp. 155–168. Springer, Cham (2018)
17. Alazawi, Z., Altowaijri, S., Mehmood, R., Abdljabar, M.B.: Intelligent disaster management system based on cloud-enabled vehicular networks. In: 2011 11th International Conference on ITS Telecommunications, ITST 2011. pp. 361–368. IEEE (2011)
18. Alazawi, Z., Abdljabar, M.B., Altowaijri, S., Vegni, A.M., Mehmood, R.: ICDMS: an intelligent cloud based disaster management system for vehicular networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, Volume 7266. pp. 40–56. Springer, Vilnius (2012)
19. Alazawi, Z., Alani, O., Abdljabar, M.B., Altowaijri, S., Mehmood, R.: A smart disaster management system for future cities. In: Proceedings of the 2014 ACM International Workshop on Wireless and Mobile Technologies for Smart Cities—WiMobCity '14. pp. 1–10. ACM, New York (2014)
20. Alazawi, Z., Alani, O., Abdljabar, M.B., Mehmood, R.: An intelligent disaster management system based evacuation strategies. In: 2014 9th International Symposium on Communication Systems, Networks and Digital Signal Processing, CSNDSP 2014. pp. 673–678 (2014)
21. Alazawi, Z., Alani, O., Abdljabar, M.B., Mehmood, R.: Transportation evacuation strategies based on VANET disaster management system. Procedia Econ. Financ. **18**, 352–360 (2014)
22. Aqib, M., Mehmood, R., Albeshri, A., Alzahrani, A.: Disaster management in smart cities by forecasting traffic plan using deep learning and GPUs. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Volume 224. pp. 139–154 (2018)
23. Mehmood, R., Meriton, R., Graham, G., Hennelly, P., Kumar, M.: Exploring the influence of big data on city transport operations: a Markovian approach. Int. J. Oper. Prod. Manag. **37**, 75–104 (2017)
24. Mehmood, R., Lu, J.A.: Computational Markovian analysis of large systems. J. Manuf. Technol. Manag. **22**, 804–817 (2011)
25. Mehmood, R., Graham, G.: Big data logistics: a health-care transport capacity sharing model. In: Procedia Computer Science. pp. 1107–1114 (2015)
26. Arfat, Y., Mehmood, R., Albeshri, A.: Parallel shortest path graph computations of United States road network data on Apache Spark. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Volume 224. pp. 323–336. Springer, Cham (2018)

27. Arfat, Y., Aqib, M., Mehmood, R., Albeshri, A., Katib, I., Albogami, N., Alzahrani, A.: Enabling smarter societies through mobile big data fogs and clouds. Procedia Computer Science 109: 1128–1133 (2017)
28. Mehmood, R., Bhaduri, B., Katib, I., Chlamtac, I. eds: Smart Societies, Infrastructure, Technologies and Applications, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (LNICST), Volume 224. Springer, Cham (2018)
29. Petalas, Y.G., Ammari, A., Georgakis, P., Nwagboso, C.: A big data architecture for traffic forecasting using multi-source information. Presented at the (2017)
30. Via, C.E., Hall, P., Tech, V., States, U., Author, C.: Transport-domain applications of widely used data sources in the smart transportation: a survey Sina Dabiri. 1–52 (2018)
31. Hadi Hosseini, S., Hadi Hosseini, S., Rahimi-kian, A.: Traffic flow prediction using MI algorithm and considering noisy and data loss conditions: an application to Minnesota traffic flow prediction. 26, 393–403 (2014)
32. Gandewar, R., Phakatkar, A.: Classification approach for big data driven traffic flow prediction using apache spark. Int. Res. J. Eng. Technol. **4**(9), 2395–2356 (2017)
33. Zhang, R., Shu, Y., Yang, Z., Cheng, P., Chen, J.: Hybrid traffic speed modeling and prediction using real-world data. In: 2015 IEEE International Congress on Big Data (BigData Congress), New York, pp. 230–237 (2015)
34. C.P.IJ. van Hinsbergen, J.W.C. van Lint, F.M.S.: Short term traffic prediction models deliverable DIIF-1a
35. Coifman, B.: Vehicle level evaluation of loop detectors and the remote traffic microwave sensor. J. Transp. Eng. **132**, 213–226 (2006)
36. Hu, J.H.J., Cao, W.C.W., Luo, J.L.J., Yu, X.Y.X.: Dynamic modeling of urban population travel behavior based on data fusion of mobile phone positioning data and FCD. 2009 17th Int. Conf. Geoinformatics. 1–5 (2009)
37. Abidin, A.F., Kolberg, M., Hussain, A.: Integrating Twitter traffic information with Kalman filter models for public transportation vehicle arrival time prediction. In: Big-Data Analytics and Cloud Computing. pp. 67–82. Springer, Cham (2015)
38. Yan, D., Zhou, J., Zhao, Y., Wu, B.: Short-term subway passenger flow prediction based on ARIMA. Presented at the December (2018)
39. Gandewar, R.R., Phakatkar, A.G., Student, M.E.: Survey on classification and prediction approaches in traffic flow. Int. J. Innov. Res. Comput. Commun. Eng. 5, (2017)
40. Wang, Y., Li, L., Xu, X.: A piecewise hybrid of ARIMA and SVMs for short-term traffic flow prediction. In: International Conference on Neural Information Processing. pp. 493–502. Springer, Cham (2017)
41. Vasantha Kumar, S., Vanajakshi, L.: Short-term traffic flow prediction using seasonal ARIMA model with limited input data. Eur. Transp. Res. Rev. 21 (2015)
42. Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y.: Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. Part C Emerg. Technol. **54**, 187–197 (2015)
43. Yin, C., Lin, Y., Yang, C.: A classification and predication framework for taxi-hailing based on big data. In: International Conference on Intelligent Computing ICIC 2017: Intelligent Computing Methodologies. p. pp 747–758. Springer (2017)
44. Florido, E., Castaño, O., Troncoso, A., Martínez-Álvarez, F.: Data mining for predicting traffic congestion and its application to Spanish data. In: 10th International Conference on Soft Computing Models in Industrial and Environmental Applications. pp 341–351. Springer, Cham (2015)
45. Park, D., Rilett, L.R.: Forecasting freeway link travel times with a multilayer feedforward neural network. Comput. Civ. Infrastruct. Eng. **14**, 357–367 (1999)
46. Pamuła, T.: CCIS 329—Traffic flow analysis based on the real data using neural networks. In: CCIS. pp. 364–371. Springer, Berlin (2012)

47. Ban, X., Guo, C., Li, G., Ban, X., Guo, C, Guo, C., Li, G.: Application of extreme learning machine on large scale traffic congestion prediction high-speed. In: Proceedings in Adaptation, Learning and Optimization. pp. 293–305. Springer, Cham (2016)
48. Deshpande, M., Bajaj, P.R.: Performance analysis of support vector machine for traffic flow prediction. In: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC). pp. 126–129. IEEE (2016)
49. Zhou, X., Wang, W., Yu, L., Zhou, X., Yu, Á.L., Wang, W., Lu, W.: Traffic flow analysis and prediction based on GPS data of floating cars. In: Lecture Notes in Electrical Engineering. pp. 497–508. Springer, Berlin (2012)
50. Li, Y., Huang, C., Jiang, J.: Research of bus arrival prediction model based on GPS and SVM. In: 2018 Chinese Control And Decision Conference (CCDC). pp. 575–579. IEEE (2018)
51. Hu, W., Yan, L., Liu, K., Wang, H.: A short-term traffic flow forecasting method based on the hybrid PSO-SVR. Neural. Process. Lett. **43**, 155–172 (2016)
52. Shoro, A.G., Soomro, T.R.: Big data analysis: Ap spark perspective. Glob. J. Comput. Sci. Technol. **15**, 7–14 (2015)
53. Meng, M., Shao, C.-F., Wong, Y.-D., Wang, B.-B., Li, H.-X.: A two-stage short-term traffic flow prediction method based on AVL and AKNN techniques. J. Cent. South Univ. **22**, 779–786 (2015)
54. Xie, J., Choi, Y.-K.: Hybrid traffic prediction scheme for intelligent transportation systems based on historical and real-time data. Int. J. Distrib. Sens. Networks. **13**, 1–11 (2017)
55. Li, K.-L., Zhai, C.-J., Xu, J.-M.: Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ANN. In: 2017 Chinese Automation Congress (CAC). pp. 2804–2807. IEEE (2017)
56. Prathilothamai, M., Sree Lakshmi, A.M., Viswanthan, D.: Cost effective road traffic prediction model using apache spark. Indian J. Sci. Technol. **9**, (2016). https://doi.org/10.17485/ijst/2016/v9i17/87334
57. Castro, P.S.., Zhang, D.., Li, S: Urban traffic modelling and prediction using large scale taxi GPS traces. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 57–72. Springer, Berlin (2012)
58. Wang, Y., Cao, J., Li, W., Gu, T., Shi, W.: Exploring traffic congestion correlation from multiple data sources. Pervasive Mob. Comput. **41**, 470–483 (2017)
59. Usama, M., Liu, M., Chen, M.: Job schedulers for big data processing in Hadoop environment: testing real-life schedule with benchmark programs. Digit. Commun. Networks. **3**(4), 260–273 (2017)
60. Deng, Z., Bai, Y.: Floating car data processing model based on Hadoop-GIS tools. In: 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics). pp. 1–4. IEEE (2016)
61. Karau, H., Konwinski, A., Wendell, P., Zaharia, M.: Learning spark: lightning-fast big data analysis. (2015)
62. Zeng, D., Xu, J., Gu, J., Liu, L., Xu, G.: Short term traffic flow prediction using hybrid ARIMA and ANN models. In: 2008 Workshop on Power Electronics and Intelligent Transportation System. pp. 621–625. IEEE (2008)
63. Yang, Y., Lu, H.: Short-term traffic flow combined forecasting model based on SVM. In: 2010 International Conference on Computational and Information Sciences. pp. 262–265. IEEE (2010)