ISSN : 0973 - 8355

www.ijmmsa.com

INTERNATIONAL JOURNAL OF MATHEMATICAL, MODELLING, SIMULATIONS AND APPLICATIONS

IJMMSA

E-MAIL
editor.ijmmsa@gmail.com
editor@ijmmsa.com

INTERNATIONAL JOURNAL OF MATHEMATICAL, MODELLING, SIMULATIONS AND APPLICATIONS
IJMMSA

ISSN 0973-8355www.ijmmsa.com
Vol 10,Issue 4.Nov 2022

TRAFFIC FORECASTING FOR AN INTELLIGENT TRANSPORTATION SYSTEM USING ML

DR SMITHA CHOWDARY CHAPARALA[1], KOYA JAYASREE[2], KUMARAM SATHWIKA[3], NANDALA SANDHYA RANI[4]

ABSTRACT

In recent years, traffic congestion prediction has led to a growing research area, especially of machine learning of artificial intelligence (AI). With the introduction of big data by stationary sensors or probe vehicle data and the development of new AI models in the last few decades, this research area has expanded extensively. Traffic congestion prediction, especially short-term traffic congestion prediction is made by evaluating different traffic parameters. Most of the researches focus on historical data in forecasting traffic congestion. However, a few articles made real-time traffic congestion prediction. This paper systematically summarises the existing research conducted by applying the various methodologies of AI, notably different machine learning models. The paper accumulates the models under respective branches of AI, and the strength and weaknesses of the models are summarised.

1. INTRODUCTION

Artificial intelligence (AI) is the most important branch of computer science in this era of big data. AI was born 50 years ago and came a long way, making encouraging progress, especially in machine learning, data mining, computer vision, expert systems, natural language processing, robotics, and related applications [1]. Machine learning is the most popular branch of AI. Other classes of AI include probabilistic models, deep learning, artificial neural network systems, and game theory. These classes are developed and applied in a wide range of sectors. Recently, it has been the leading research area in transportation engineering, especially in traffic congestion prediction.Traffic congestion has a direct and indirect impact on a country's economy and its dwellers' health. According to Ali et al. [2], traffic congestion causes Pak Rs. 1 million every day in terms of opportunity cost and fuel consumption due to traffic congestion. Traffic congestion affects on individual level as well. Time loss, especially during peak hours, mental stress, and the added pollution to the global warming are also some important factors caused due to traffic congestion.

ASSOSCIATE PROFESSOR[1], UG SCHOLAR [2,3&4]
DEPARTMENT OF CSE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN,MAISAMMAGUDA, DHULAPALLY KOMPALLY, MEDCHAL RD, M, SECUNDERABAD, TELANGANA 500100

Ensuring economic growth and the road users' comfort are the two requirements for the development of a country, which is impossible without smooth traffic flow. With the development in the transportation sector by collecting traffic information, authorities are putting more attention on traffic congestion monitoring. Traffic congestion prediction provides the authorities with the required time to plan in the allocation of resources to make the journey smooth for travellers. Traffic congestion prediction problem discussed in this paper can be defined as an estimation of parameters related to traffic congestion into the short-term future, e.g., 15 minutes to a few hours by applying different AI methodologies by using collected traffic data. There are usually five parameters to evaluate, including traffic volume, traffic density, occupancy, traffic congestion index, and travel time while monitoring and predicting traffic congestions. Depending on the nature of the collected data, a variety of AI approaches are applied to evaluate the congestion parameters. This article systematically discusses the models and their advantage and disadvantages. The primary motivation of this review is to gather the articles focusing solely on traffic congestion prediction models. The keywords used in the search process included "traffic congestion prediction" OR "traffic congestion estimation" OR "congestion prediction modelling" OR "prediction of traffic congestion" OR "road congestion forecast" OR "traffic congestion forecast." For efficient screening, research paper search was done according to year using search engines like Scopus, Google Scholar, and Science Direct.

2. GENERAL LAYOUT

Traffic congestion forecasting has two basic steps of data collection and prediction model development. Every step of the methodology is important and may affect the results if not done correctly. After data collection, data processing plays a vital role to prepare the training and testing datasets. Case area differs for different research. After developing the model, it is validated with other base models and ground true results. Figure 1 shows the general components of traffic congestion prediction studies. These branches were further divided into more specific sub-branches and are discussed in the following sections.

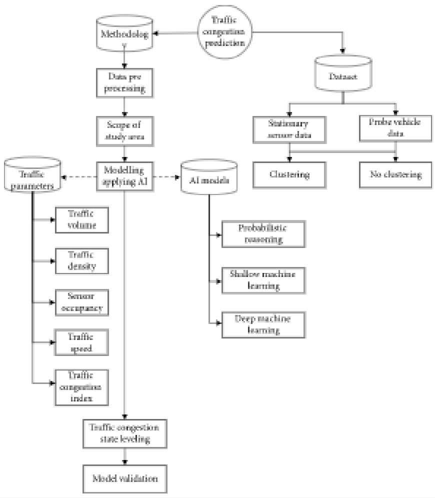

Fig :The layout of traffic congestion prediction system.

3. Data Source

Traffic datasets used in different studies can be mainly divided into two classes, including stationary and probe data. Stationary data can be further divided into sensor data and fixed cameras. On the other hand, probe data that were used in the studies were GPS data mounted on vehicles.Stationary sensors continuously capture spatiotemporal data of traffic. However, sensor operation may interrupt anytime. Authorities should always consider this temporary failure of the sensor while planning by using this data. The advantage of the sensor data is that there is no confusion on the location of the vehicles. The most used dataset was Performance Measurement System (PeMS) that collects highway data across all major metropolitan areas of the State of California of traffic flow, sensor occupancy, and travel speed in real-time. Most of the studies used dataset from the I-5 highway, in San Diego, California, every 5 minutes [3–6]. Other systems included the Genetec blufaxcloud travel-time system engine (GBTTSE) [7] and the Topologically Integrated Geographic Encoding and Referencing (TIGER) line graph [8].On the other hand, probe data has the advantage of covering the entire road network. A network consists of different structured roads. Therefore, studies, especially those that considered the network wide area, used probe data. The most used dataset was GPS data collecting every second from approximately 20000 taxies of Beijing, China. Data included the taxi number, the latitude-longitude of the vehicle, timestamp when sampling, and whether there was a passenger or not. Data updating frequency of this dataset varies from 10 s to 5 min according to the quality of GPS device [4, 5, 9]. Other probe data included low-frequency Probe Vehicle Data (PVD) [10] and bus GPS data [11, 12]. However, sometimes probe data show significant fluctuation. Besides, map matching is usually a must for probe data. But data can minimize this limitation. Probe data collected from one city cannot be used directly for modelling other city networks. This is because the data collected from Beijing, China, includes latitude-longitude of the vehicle, which is unique. However, a generalised model using probe data can be generated for different cities.Other data sources, e.g., data from tolling system and data provided by transportation authority, will add more reliable data as the sources are dependable. However, a lot of the times, study area needs to be adjusted as in most cases, tolled road information is not available. Tracking cellular phone movements without privacy breach can also be a source of data. However, the heterogeneity of the vehicle distribution will be hard to determine from this dataset, if not impossible. Besides, due to pedestrian or cyclists travelling through the sidewalk, there might be many outliers in the dataset if modelling is done for a road network. Data collected from a questionnaire to the general public/drivers may provide a misleading result [13].

3.1. Clustering Algorithms

Some studies use clustering the acquired data before applying the main congestion models of prediction. This hybrid modelling technique is applied to fine-tune the input values and to use them in the training phase. Figure 2 shows the commonly used AI clustering models in this field of research[12][13][14]. The models are described briefly in this section.

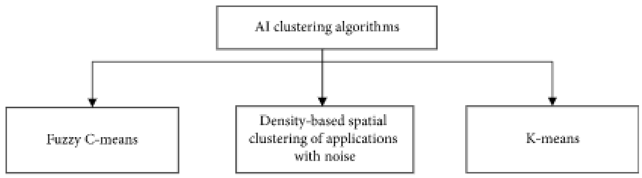

Figure 2
Commonly used AI clustering algorithms.

Fuzzy C-Means (FCM) is a popular nondeterministic clustering technique in data mining. In traffic engineering researches, traffic pattern recognition plays an important role. Besides, these studies often face the limitation of missing or incomplete data. To deal with these constraints, FCM has become a commonly applied clustering technique. The advantage of this approach is, unlike original C-means clustering methods, it can overcome the issue of getting trapped in the local optimum [14]. However, FCM requires setting a predefined cluster number, which is not always possible while dealing with massive data without any prior knowledge of the data dimension. Besides, this model becomes computationally expensive with data size increment. Different studies have applied FCM successfully by improving its limitations. Some studies changed the fuzzy index value for each FCM algorithm execution [15], some calculated the Davies-Bouldin (DB) index [10], while

others applied the K-means clustering algorithm [6][7].
K-means clustering is an effective and relatively flexible algorithm while dealing with large datasets. It is a popular unsupervised machine learning algorithm. Depending on the features, cluster number varied from two to 50 Like FCM, K-means clustering requires a predefined cluster number and selecting K original cluster centres. GAP and WEKA toolboxwere used to define the value. For large datasets, as the sample distribution is unknown in the beginning, it is not always possible to fulfil these two requirements. A few studies used adaptive K-means clustering overcoming the limitations and exploited the pattern using principal component analysis (PCA) [15][16]
DBSCAN is more of a general clustering application in machine learning and data mining. This method overcomes the limitation of FCM of predefining the cluster number. It can automatically generate arbitrary cluster shapes surrounded by clusters of different characteristics and can easily recognise outlier. However, it requires two parameters to preset. A suitable parameter determination method, e.g., trial and error method [8] and human judgement [26] makes the model computationally expensive and requires a clear understanding of the dataset.
From the above discussion, it is concluded that only 16 out of 48 studies have done clustering before applying prediction models. Several time-series models and shallow machine learning (SML) algorithms have used clustering approach. However, deep learning algorithms can process input data on different layers of the model, thus may not need clustering beforehand.

4. Applied Methodology

Traffic flow is a complex amalgamation of heterogenous traffic fleet. Thus, traffic pattern prediction modelling could be an easy and efficient congestion prediction approach. However, depending on the data characteristics and quality, different classes of AI are applied in various studies. Figure 3 shows the main branches—probabilistic reasoning and machine learning (ML). Machine learning comprised of both shallow and deep learning algorithms. However, with the progress of this article, these sections were subdivided into detailed algorithms.

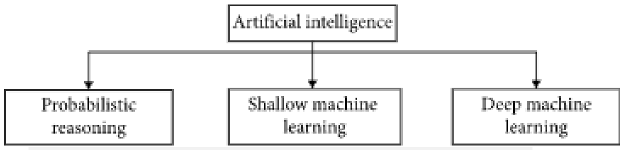

Fig:Branches of artificial intelligence in this article.

CONCLUSIONS

Traffic congestion prediction is getting more attention from the last few decades. With the development of infrastructure, every country is facing traffic congestion problem. Therefore, forecasting the congestion can allow authorities to make plans and take necessary actions to avoid it. The development of artificial intelligence and the availability of big data have led researchers to apply different models in this field. This article divided the methodologies in three classed. Although probabilistic models are simple in general, they become complex while different factors that affect traffic congestion, e.g., weather, social media, and event, are considered. Machine learning, especially deep learning, has the benefit in this case. Therefore, deep learning algorithms became more popular with time as they can assess a large dataset. However, a wide range of machine learning algorithms are yet to be applied. Therefore, a vast opportunity of research in the field of traffic congestion prediction still prevails

REFERENCES

1. Z. Shi, *Advanced Artificial Intelligence*, World Scientific, Singapore, 2011.
2. M. S. Ali, M. Adnan, S. M. Noman, and S. F. A. Baqueri, “Estimation of traffic congestion

cost-A case study of a major arterial in karachi," *Procedia Engineering*, vol. 77, pp. 37–44, 2014.
3. W. Cao and J. Wang, "Research on traffic flow congestion based on Mamdani fuzzy system," *AIP Conference Proceedings*, vol. 2073, 2019.View at: Publisher Site | Google Scholar
4. X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," *Future Generation Computer Systems*, vol. 61, pp. 97–107, 2016.View at: Publisher Site | Google Scholar
5. Q. Yang, J. Wang, X. Song, X. Kong, Z. Xu, and B. Zhang, "Urban traffic congestion prediction using floating car trajectory data," in *Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing*, pp. 18–30, Springer, Zhangjiajie, China, November 2015.View at: Google Scholar
6. W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1688–1711, 2019.View at: Publisher Site | Google Scholar
7. T. Adetiloye and A. Awasthi, "Multimodal big data fusion for traffic congestion prediction," *Multimodal Analytics for Next-Generation Big Data Technologies and Applications*, Springer, Berlin, Germany, 2019.View at: Publisher Site | Google Scholar
8. F. Wen, G. Zhang, L. Sun, X. Wang, and X. Xu, "A hybrid temporal association rules mining method for traffic congestion prediction," *Computers & Industrial Engineering*, vol. 130, pp. 779–787, 2019.View at: Publisher Site | Google Scholar
9. J. Wang, Y. Mao, J. Li, Z. Xiong, and W.-X. Wang, "Predictability of road traffic and Congestion in urban areas," *PLoS One*, vol. 10, no. 4, Article ID e0121825, 2015.View at: Publisher Site | Google Scholar
10. Z. He, G. Qi, L. Lu, and Y. Chen, "Network-wide identification of turn-level intersection congestion using only low-frequency probe vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 320–339, 2019.View at: Publisher Site | Google Scholar
11. H. Manoharan, et al., "Examining the effect of aquaculture using sensor-based technology with machine learning algorithm", Aquaculture Research, 51 (11) (2020), pp. 4748-4758.
12. S. Sundaramurthy, S. C and P. Kshirsagar, "Prediction and Classification of Rheumatoid Arthritis using Ensemble Machine Learning Approaches," *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 2020, pp. 17-21, doi: 10.1109/DASA51403.2020.9317253.
13. P. R. Kshirsagar, H. Manoharan, F. Al-Turjman and K. Kumar, "DESIGN AND TESTING OF AUTOMATED SMOKE MONITORING SENSORS IN VEHICLES," in *IEEE Sensors Journal*, doi: 10.1109/JSEN.2020.3044604.
14. P. Kshirsagar and S. Akojwar, "Optimization of BPNN parameters using PSO for EEG signals," ICCASP/ICMMD-2016.

Advances in Intelligent Systems Research. Vol. 137, Pp. 385-394,2016
15. Dr. Sudhir Akojwar, Pravin Kshirsagar, Vijetalaxmi Pai "Feature Extraction of EEG Signals using Wavelet and Principal Component analysis", National Conference on Research Trends In Electronics, Computer Science & Information Technology and Doctoral Research Meet, Feb 21st & 22nd ,2014.
16. D. Jose, P. N. Kumar and A. A. Hussain, "Fault tolerant adaptive neuro-fuzzy based automated cruise controller on FPGA," *2013 Annual IEEE India Conference (INDICON)*, 2013, pp. 1-5, doi: 10.1109/INDCON.2013.6725977.