

Big Data Algorithms and Applications in Urban Intelligent Transportation Systems

Meitong Wang

The Reading Academy, Nanjing University of Information Science and Technology,
Nanjing, China

3273809783@qq.com

Abstract. In recent years, with the rapid development of Big Data algorithms, more and more advanced techniques such as 3G wireless transmission, GPS positioning and car-sharing have arisen and brought opportunities and challenges to the innovation and development of Urban Intelligent Transportation Systems. Both academic and industrial sectors of ITS have shown interest in these applications. However, defects still exist in current data collection and require improvement. This paper is intended to provide a comprehensive analysis of the applications of Big Data in ITS, showing diverse fields of those applications, integrating different algorithms and models, analyzing potential problems and putting forward future prospects. By comparing different algorithms and their applications according to relevant research, results can be found that the algorithms discussed are able to help each other with traffic issues. Meanwhile, ITS is becoming increasingly accurate and efficient. In addition, although modern data collection has brought a great deal of convenience and transformation to ITS, there are still quite a few challenges that demand immediate solutions. The results of the study are considered to play a significant role in the promising directions for future research.

Keywords: Big Data algorithms, Intelligent Transportation Systems, Applications, Data models, Challenges.

1. Introduction

Nowadays, the rapid advancement of information, communication and computer technology has led to the emergence of Digital Cities and Smart Cities[1]. Under the influence of these technologies, significant changes which include improvement of traffic flow, reduction of congestion and enhancement of transportation safety have occurred in transportation areas, promoting the development of Intelligent Transportation Systems. Furthermore, Big Data algorithms in ITS are widely applied in signal recognition, object detection, traffic flow prediction, travel time planning, travel route planning, safety of vehicles and roads and other urban traffic applications[2]. In most advanced cities, a great deal of traffic big data is generated, which indicates that big data algorithm play a crucial role in accelerating ITS development. By analyzing massive amounts of data, such algorithm can provide valuable insights for decision-making in Urban ITS. Moreover, since data is the foundation of ITS, it is of great importance to conduct in-depth analysis of Big Data algorithms, clarify the basic concepts, discuss the problems brought by applications of traffic Big Data and understand the mathematical modeling methods driven by Big Data. This study will analyse several recognized Big Data algorithms and their

applications in ITS, describe corresponding data models and integrate them to get a result for the prospect of Urban ITS. By using methods of current situation countermeasure, we are able to seek out the development of ITS and potential issues caused by current algorithms, analyzing possible solutions and make improvements.

2. The Development of Intelligent Transportation Systems

ITS, which stands for Intelligent Transportation Systems, refers to the integration of information and communication technology with transportation infrastructure. Such tools are aimed at improving the efficiency, safety, and sustainability of traffic. Additionally, these systems typically involve a variety of technologies, such as sensors, control systems, and communication networks, to enhance traffic management, traveller information, vehicle safety, and transportation infrastructure maintenance.

The history of ITS can be dated back to the late 20th century. Early in the 1970s, the National Highway Traffic Safety Administration(NHTSA) established by the Highway Safety Act of 1970 predated a domestic ITS program[3]. After such a program was set out, developed countries including the USA, Europe and Japan established policies about ITS one after another in the period between 1990 and 2000. ITS gradually took an important role in traffic systems.

Recently, ITS continued to evolve toward intelligence and networking, with autonomous vehicles, vehicle networking technology, intelligent transportation management systems, and other areas becoming major topics in research and application. Researchers are now able to use ITS to deal with environmental issues in transportation. According to a study, with the help of ITS, the negative effects brought by transportation such as air pollution and greenhouse gas emissions will decrease,as data showed that utilizing ITS could decrease emissions by almost half[4]. Such results prove that ITS is essential to reducing the adverse environmental impact of transportation. Moreover, with the rapid development of Big Data Algorithms, more and more advanced techniques have arisen and brought considerable opportunities to the innovation and development of urban ITS. For instance, in order to offer services to satisfy both drivers and passengers, promote road safety and cut down traffic congestion, sensors and actuators like tire pressure sensor and rear-view visibility systems have become compulsory in manufacturing vehicles and ITS implementation[5]. It is obvious that ITS has evolved into a sophisticated and practical technique.

3. Transportation Big Data

Transportation big data refers to the large volumes of information collected from various transportation systems, including vehicle flows, public transit operations, and pedestrian movements. It encompasses data on objects like traffic patterns, congestion levels, accident occurrences and travel behaviors. This data is characterized by its volume, velocity, and variety, which is often referred to as the 3Vs of big data[1]. Researchers showed that big data has been incorporated into transportation research across a great deal of topics, containing service delivery, traffic, environment, infrastructure, route advancements, and behaviour of both services and individuals[6]. There are diverse major algorithms that are often utilized in Transportation Big Data such as clustering, deep learning and regression analysis.

3.1. Clustering

A cluster algorithm is a kind of unsupervised learning algorithm that automatically classifies similar samples into a category. Its purpose is to identify structures within data, where similar data points are categorized together based on their attributes or features. There are various types of cluster algorithms, consisting of partitioning methods, hierarchical methods, density-based methods and centroid-based methods. Take K-means as a typical example. Researchers iteratively divide data into k clusters to make the data points as close to their center as possible. According to figure 1, the sample is divided into k points as the initial centroid of the clusters randomly. Each point is assigned to the nearest centroid, creating 4 clusters. Then, calculate the new centroid of each cluster by taking the mean of all points assigned to that cluster. Eventually, repeat the former steps until the centroids no longer change significantly. The whole process is beneficial to reduce data volume and mine data structure.

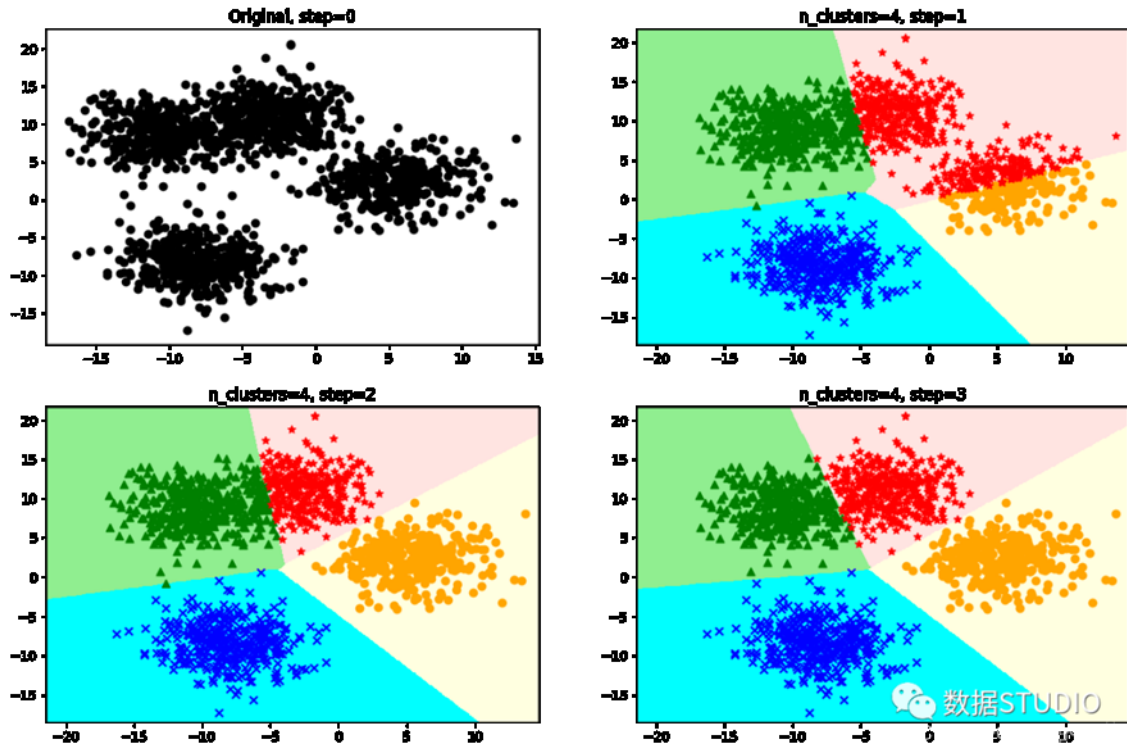


Figure 1. The main process of K-means clustering(cited from Tencent Cloud)

In urban transportation systems, clustering algorithms are often used in detecting abnormal events such as traffic accidents or congestion from a large amount of traffic data. One major example is scheduling and predicting urban traffic by using K-means clustering and other machine learning methods. As is shown in figure 2, by using tools such as RSUs and mobile sensors, the police are able to predict actual density. Then they can get information on how many cops are needed in different regions where junctions occur, by using K-means clustering or deep neural networks. At the same time, they will notify the next location that may cause a junction[7]. With the help of clustering methods, researchers can get both accuracy and efficiency in detecting traffic accidents or conjunctions.

3.2. Deep learning

The deep learning algorithm is a sophisticated type of machine learning technique that involves neural networks with numerous layers. Multi-layer network structures operate these algorithms, classifying or regressing unknown data to model complex patterns in large datasets, enabling the system to perform accurate predictions or decisions without explicit programming. Common types of deep learning algorithms include convolutional neural networks (CNNs) for image and video recognition, recurrent neural networks (RNNs) for sequence processing and natural language understanding, and generative adversarial networks (GANs) for generating new data.

Given the complexity of the deep learning process, we can conceptualize it as a water flow system, as illustrated in Figure 3. The deep learning network for processing data is a huge water pipe network composed of pipes and valves. The entrance of the network is several pipeline openings, while the exit is also several pipeline openings. There are considerable layers on this water pipe network, with each layer containing numerous regulating valves that can control the direction and flow of water. According to the needs of different tasks, the number of layers in the water pipe network and the number of regulating valves in each layer can have different combinations of changes. For more complex tasks, the total number of regulating valves can reach thousands or even higher. In the water pipe network, each regulating valve in each layer is connected to all regulating valves in the next layer through water pipes,

forming a fully connected water flow system from front to back, layer by layer[8]. Applying examples from everyday life makes understanding the application of deep learning easier.

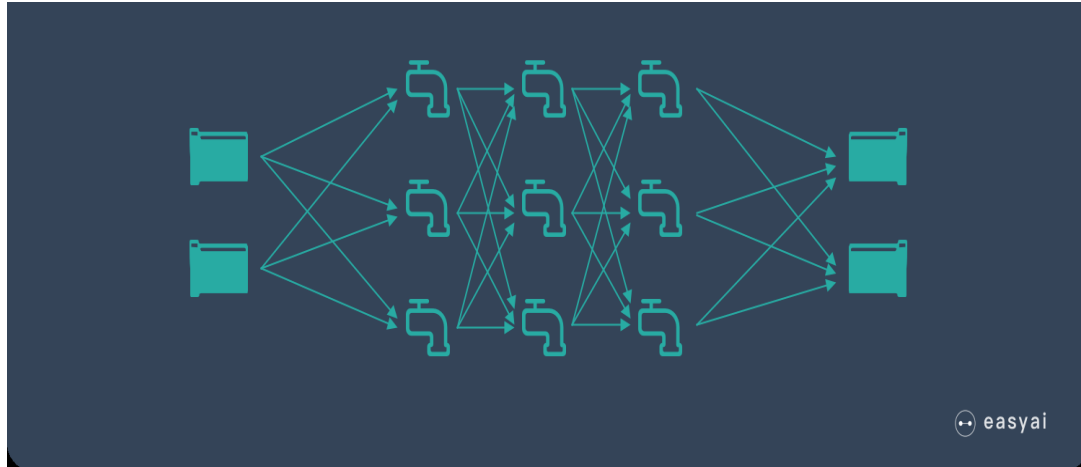


Figure 2. Assuming that the process of deep learning is "water flow"(from CSDN)

Researchers often use deep learning methods to ensure high accuracy in vehicle, traffic sign and license plate recognition. By analyzing traffic flow detection, researchers found that although the detection accuracy of the algorithm shows a downward trend, the average accuracy is higher than 92.0% at all scales. The average accuracy of all test videos and all YOLO vehicle detector input image normalization scales is 94.6%[9]. It seems that such algorithms bring more efficiency than challenges in detection.

3.3. Regression analysis

Regression analysis is a statistical method used to learn the relationship between a dependent variable and one or more independent variables. It is employed to predict the output of a continuous value based on one or more inputs in order to understand the strength of the relationship, and to identify independent variables. Common methods include linear regression and logistic regression.

Researchers mainly use regression algorithms to predict traffic flow and plan optimal routes. The recommended measure can be extended by providing a mobile route guidance system or application that can automatically gather the information and further implement fuzzy logic to take the best decision to select the path to reach the destination with the help of Google Map[10]. It can be concluded that regression analysis enhances the comfort and efficiency of the process.

Similarly, such algorithms are able to be utilized for GPS sensors in order to control the velocity of each vehicle. In accordance with figure 4, researchers sorted out the information of velocity offered by the GPS sensor. There are discernible differences among diverse traffic transportation. Based on the message, the data provided by GPS sensor can be useful for recognizing activity or detecting transportation mode[11]. This study showed that logistic regression classification enables people to better operate the speed of different types of vehicles.

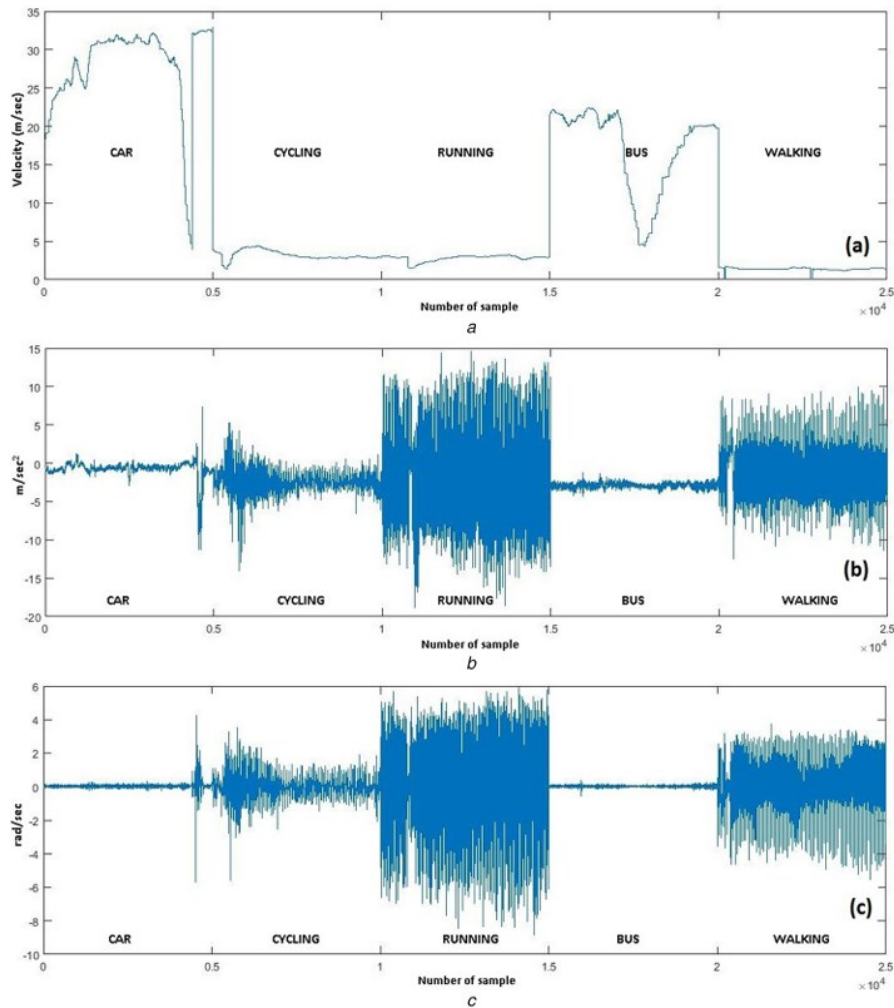


Figure 3. Sensor data sample (a) Velocities of transportation modes, (b) Amplitude change of accelerometer X-axis, (c) Amplitude change of gyroscope Y-axis [11]

4. Collaboration and challenges of algorithms in urban intelligent transportation system

By analyzing the above algorithms and their applications, the study shows that different algorithms assist each other in dealing with traffic problems. The K-means cluster, combining with RNN, can help better predicting traffic flow. Diverse algorithms are required to complete one task and all of them are significant and essential. Furthermore, with the help of Big Data algorithms, data collection in ITS is becoming more and more efficient and precise. By integrating machine learning methods, ITS has improved to a large extent. Nowadays, detecting traffic congestion is easier than it was in the past. While the efficiency promotes, the accuracy of the results gets improved. Compared to traditional data collection, modern data models have more intensive traffic sensor design and steady multi-source data fusion[1]. It proves that these algorithms can reduce processing and increase speed to collect data. The collection used in ITS now has more motivation to promote the innovation and development of urban ITS.

However, while big data algorithms present numerous opportunities for Information and Communication Technology (ITS), they also present various challenges. First and foremost, in the aspect of data security, as the volume and variety of traffic data continue to grow, problems of data privacy, security, and interoperability begin to emerge. Collecting and storing large amounts of personal

and vehicle information can lead to privacy violations and data security issues. Transportation big data may contain national security like data transmitted through the public security network and personal privacy such as vehicle trajectory data detected by checkpoint systems[1]. Similarly, the security and privacy of the VFC network are also highly critical problems. VFC is proved to be more vulnerable to security and privacy threats because of its temporary property[12]. The leakage of this data will lead to huge losses.

Secondly, in terms of quality and completeness, the effectiveness of big data relies heavily on the quality and completeness of the data. In ITS, data can originate from a variety of sources, including sensors, cameras, and GPS, all of which can potentially contain errors, omissions, or inconsistent formats. For instance, toll areas of the same highway are yielded to different operators, causing different highway traffic data reporting schemes[13]. Fusing and harmonizing large amounts of data from diverse sources at the same time may cause less quality and completeness of data.

Finally, data storage poses a significant issue. With the continuous growth of data volumes, effectively storing, managing, and maintaining this data becomes a problem for ITS. Currently, the data volume has increased from TB level to PB level while the growth in data storage capacity is far behind data growth[14]. This demands efficient data storage solutions and effective data management strategies.

5. Conclusion

To draw a conclusion, big data algorithms have revolutionized urban ITS by providing valuable insights for decision-making and enhancing system efficiency. With the continuous advancement of technology, these algorithms especially clustering, deep learning and regression are expected to play an even more significant role in urban transportation management. This essay has proved that these algorithms all make it easier and more convenient to regulate urban ITS efficiently and accurately. Additionally, researchers have found that there is a deep correlation among these algorithms. Different algorithms have the ability to support each other in handling traffic issues. Nevertheless, it is crucial to concentrate more on the challenges that advanced big data have brought to urban ITS. Despite advanced development of algorithms and techniques, current researches are still lack of consideration of data security, quality and storage. Regarding these issues, certain measures ought to be put forward. As for privacy problems, government should make more data protection regulations and anonymization processing is recommended. Additionally, it is crucial to intensify the inspection and promotion of facilities to ensure quality and completeness. Moreover, it is possible to improve storage of data by using cloud services or distributed storage.

Based on opportunities and challenges brought by algorithms in urban ITS, future researches should concentrate on promoting the current data system, strengthening security protection tools and developing more sophisticated algorithms to address the challenges posed by the ever-growing volume and complexity of transportation data. Meanwhile, promoting and perfecting available ITS resources, integrating diverse machine learning ways and innovating more algorithms are crucial.

References

- [1] LUHua-pu, SUN Zhi-yuan, QU Wen-cong.: 'Big Data and Its Applications in Urban Intelligent Transportation System', *Journal of Transportation Systems Engineering and Information Technology*, 2015, 15, (5)
- [2] Kaffash, S., Nguyen, A.T., Zhu, J.: 'Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis', *International Journal of Production Economics*, 2020, doi:<https://doi.org/10.1016/j.ijpe.2020.107868>.
- [3] Ashley Auer, Shelley Feese, Stephen Lockwood.: 'History of Intelligent Transportation Systems', U.S. Department of Transportation, 2016, pp. 12
- [4] Alrawi, Firas.: 'The importance of intelligent transport systems in the preservation of the environment and reduction of harmful gases', *Transportation Research Procedia*, 2017, Vol. 24, pp. 197–203. doi:10.1016/j.trpro.2017.05.108

- [5] Juan Guerrero-Ibáñez, Sherali Zeadally, Juan Contreras-Castillo.: ‘Sensor Technologies for Intelligent Transportation Systems’, *Sensors*, 2018, 18(4), pp. 1212 doi:10.3390/s18041212
- [6] Timothy F. Welch, Alyas Widita.: ‘Big data in public transportation: a review of sources and methods’, *Transport Reviews*, 2019, 39(6), pp. 795-818 doi:10.1080/01441647.2019.1616849
- [7] Mahima Nama, Ankita Nath, Nancy Bechra, *et al.*: ‘Machine learning - based traffic scheduling techniques for intelligent transportation system: Opportunities and challenges’, *International Journal of Communication Systems*, 2021 doi:10.1002/dac.4814
- [8] Kai-Fu Lee, 2017. *Artificial Intelligence*, no data
- [9] Chen Chen, Bin Liu, Shaohua Wan, *et al.*: ‘An Edge Traffic Flow Detection Scheme Based on Deep Learning in an Intelligent Transportation System’, *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(3), pp. 1840-1852 doi:10.1109/TITS.2020.3025687
- [10] Anurag Singh Tomar, Mridula Singh, Girish Sharma, *et al.*: ‘Traffic Management using Logistic Regression with Fuzzy Logic’, *Procedia Computer Science*, 2018, Vol.132, pp. 451–460 doi: 10.1016/j.procs.2018.05.159
- [11] Serkan Ball, Ensar Arif Sağbaş.: ‘Diagnosis of transportation modes on mobile phone using logistic regression classification’, *IET Software*, 2018, 12(2), pp.142–151 doi:10.1049/iet-sen.2017.0035
- [12] Tasneem S. J. Darwish, Kamalrulnizam Abu Bakar.: ‘Fog Based Intelligent Transportation Big Data Analytics in The Internet of Vehicles Environment: Motivations, Architecture, Challenges and Critical Issues’, *IEEE Access*, 2018, Vol 6, pp. 15679-15701 doi:10.1109/ACCESS.2018.2815989
- [13] Guilherme Guerreiro, Paulo Figueiras, Ricardo Silva, *et al.*: ‘An architecture for big data processing on intelligent transportation systems’, *An application scenario on highway traffic flows*, 2016, pp. 65–72 doi:10.1109/IS.2016.7737393
- [14] Li Zhu, Fei Richard Yu, Yige Wang, *et al.*: ‘Big Data Analytics in Intelligent Transportation Systems: A Survey’, *IEEE Transactions on Intelligent Transportation Systems*, 2018, 20(1), pp.383-398 doi:10.1109/TITS.2018.2815678