On the analysis of individual data on transport usage*

M.V. Bulygin[I], D.E. Namiot[I], O.N. Pokusaev[II]

[I] Lomonosov Moscow State University, Moscow, Russia
[II] Russian Transport University (MIIT), Moscow, Russia

Abstract. The percentage of the world's urban population is currently more than 50\% and will increase according to UN forecasts. Urban infrastructure must develop along with population growth. This article provides an overview of methods for improving the city's transport infrastructure based on data analysis. The article presents methods for reducing harmful emissions, optimizing the operation of taxis and public transport, as well as recognizing transportation modes and some other tasks. These methods operate with data describing the transport behavior of individual users of the transport network. The sources of such data are smart card validators, GPS sensors, and smartphone accelerometers. The article reveals the advantages and disadvantages of using each of the data types, as well as presents alternative ways to obtain them. These methods, along with methods for aggregated data analysis, can become the main part of a single platform that will allow city authorities in the process of improving the transport infrastructure. We propose architecture of this platform which will allows developers to extend range of available algorithms and methods dynamically.

Keywords: *transport data analysis, Data on transport usage, Smart city, Digital urbanism, Smart card data analysis, GPS data analysis.*

DOI: 10.14357/20790279230104

Introduction

According to the Digital 2021 report, more than half of the world's population (56%) lives in cities [1]. According to UN forecasts, the city population will continue to increase further to 6.3 billion (about 65%) by 2050 [4]. According to the study [17], population density affects the quality of life. It is very important to compensate for the negative impact of urban population growth on the comfort of their inhabitants. One of the concepts that make the life of urban residents more comfortable is the concept of a smart city. A smart city is a concept for managing the resources and infrastructure of a city with the widespread use of information technology and Internet of Things technologies. In article [13], a smart city is defined as a well-defined geographical area, in which high technologies such as ICT, logistics, energy production, and so on, cooperate to create benefits for citizens in terms of well-being, inclusion, and participation, environmental quality, intelligent development. It is governed by a well-defined pool of subjects, able to state the rules and policy for the city government and development. Public transport is one of the main components of the urban infrastructure that provides a comfortable city life. In [15], the authors emphasize that smart projects in a technocity should be aimed at transport improvement. To build an easy-to-use public transport system, it is necessary to take into account the transport needs of citizens. In a time before information technology penetrated everyday life, surveys and censuses were used to study the needs of citizens. The collection of such data is expensive, requires the participation and time of citizens, and the data obtained quickly become outdated. In the modern world, new data sources describe the movements of citizens. They provide data at a lower cost and with accuracy and speed that was not possible with the methods of the past. One such data source is mobile phones. During their operation, mobile phones and smartphones exchange information about signal strength and delay with base stations. This information is stored in base stations and can be used to determine the location of devices. The main advantage of these cellular operators is mass character. According to [1], the penetration rate of mobile phones is more than 65% and is constantly increasing. In large cities

* This research has been supported by the Interdisciplinary Scientific and Educational School of Moscow University "Brain, Cognitive Systems, Artificial Intelligence".

and developed countries, according to [3], almost the entire population is subscribers of mobile networks. The important advantage of cellular operators' data is collecting at the operators' base stations, centrally and not visible to users. The disadvantages of such data include their low spatial accuracy, which is sufficient to determine, for example, the area of departure and arrival, but not a specific object. The data of cellular operators are convenient for analysis and mostly used in the form aggregated by city districts and time intervals. Example of aggregated transport data is presented in Table 1. Thus, researchers have access to data on the number of people in each of the city districts. Traffic flow data between each pair of districts for each time interval is also available. Due to the presence of base stations of cellular operators in the subways of some cities, for example, in Moscow, it is possible to measure the traffic flow of the subway separately. The development of data analysis algorithms also makes it possible to obtain information on the number of people traveling from home to work and from work to home.

Table 1

Aggregated transport data example

Timestamp	Source district id	Destination district id	Count of customers
20/01/22 10:00	1	2	1012
20/01/22 10:00	1	3	1258
20/01/22 10:30	1	2	521
20/01/22 10:30	1	3	620

An overview of cellular operators' data analysis methods is presented in the article [10]. Our paper [8] presents an anomaly detection method for aggregated data of cellular operators. Such anomalies correspond to important social events. These events do not need to be predicted. They are either known in advance (for example, large concerts, social events) or are unpredictable (for example, major accidents, fires). City authorities need to respond to such events on time and take measures to improve the traffic situation, so it is necessary to identify changes in traffic flows and measure them. The paper [9] proposes a method for clustering city districts based on aggregated data from cellular operators. As a result of clustering, the city areas were divided into five clusters (from residential to working areas). Urban infrastructure in districts of different clusters should have various development vectors to meet the needs of citizens living in these areas. This article is devoted to an overview of methods for analyzing individual transport data. Such data, in contrast to aggregated data, make it possible to obtain more accurate and granular information about the nature of movements in the city.

1. Main Part

Most modern mobile phones are smartphones. They have many sensors such as GPS [20] and accelerometers. Unlike location data received by cellular operators, data from GPS sensors has higher spatial accuracy. It is important to note that GPS data has a lower penetration rate than cellular data, as not all mobile phones have a GPS sensor. To transfer data, you need to use a special application. Crowdsensing can be used to motivate smartphone users to install applications. It is the provision of some bonuses to users, for example, additional free Internet traffic or free access to content, in exchange for data. The collection of trajectory information can also be embedded in applications that provide transportation services, such as a taxi/bus depot application. The data received by smartphone accelerometers are also collected on the user's side. It can also be obtained utilizing crowdsensing. These data are successfully used to predict the type of activity of a smartphone user and the type of transport they use [35] [29]. The solution to such problems helps to understand the distribution of the traffic flow between different transportation modes. It allows city governments to organize better the transport network of the city. It is possible to use data from smart card validators to analyze the movements of public transport passengers. Smart cards (transport cards) are a popular means of payment in transport systems. Such cards are used to pay for travel in cities such as Moscow, Beijing, Melbourne, and many others. They record information about the balance or the number of remaining trips. Admission to public transport is carried out after passing the validation process. If the cost of the trip depends on the travel distance or the endpoint, then the smart card can also be validated at the end of the trip. This solution is less convenient for passengers. The data of the validators, in this case, allows researchers to accurately determine the trip destination. If destination data is not available, there are heuristics to retrieve it. The paper [33] is devoted to an overview of the data formats collected by the validators and some processing methods. The advantages and disadvantages of using data from various sources are presented in the Table 2.

Data type Advantages Disadvantages GPS data Ability to obtain trajectories with high spatial accuracy, the ability to determine the speed. Low penetration rate, the need to install special applications to obtain Accelerometer data High accuracy in solving transportation mode prediction problems Low penetration rate, the need to install special applications to obtain, do not contain information about the location, a narrow range of tasks to be solved Smart card data Collected centrally, high penetration degree, accurate spatio-temporal marks Reflect only data on the traffic flow in public transport, the endpoint of the trip and the exact trajectory are not always known

Table 2

Advantages and disadvantages of using data from various sources

Data type	Advantages	Disadvantages
GPS data	Ability to obtain trajectories with high spatial accuracy, the ability to determine the speed	Low penetration rate, the need to install special applications to obtain
Accelerometer data	High accuracy in solving transportation mode prediction problems	Low penetration rate, the need to install special applications to obtain, do not contain information about the location, a narrow range of tasks to be solved
Smart card data	Collected centrally, high penetration degree, accurate spatio-temporal marks	Reflect only data on the traffic flow in public transport, the endpoint of the trip and the exact trajectory are not always known

2. GPS data

GPS data are collected from smartphones or navigators using specialized applications. Such an application can be installed on the device by the user himself to exchange his data for some bonuses or useful functionality. Also, applications that collect GPS trajectory data are often installed on the devices of employees of transport companies.

Taxi fleet data. There are two main directions in the analysis of taxi fleet data on trajectories collected using GPS. The purpose of the first direction is to optimize the operation of taxis. Taxi companies are most interested in the development of this area. It includes a behavior analysis of drivers, an analysis of passenger demand, and other tasks. The solution of these tasks will reduce costs and increase the income of taxi services. The second direction is the analysis of the transport network of the city. Within the framework of this direction, the tasks of assessing the total transport flow and its characteristics on the city streets are being solved. The possibilities of reducing harmful emissions from transport services are being analyzed. The latter challenge is especially relevant in some metropolitan areas where high levels of CO_2 emissions harm the health of citizens.

In the study [22], the authors, based on the analysis of GPS data on the trajectories of taxi cars, proposed a method for optimizing taxi ranks, taking into account the spatio-temporal distribution of passenger demand and the waiting time for taxi customers. The proposed methodology is demonstrated on 204 hours of trajectory records in Shenzhen (China). The research materials in this paper are useful to city authorities when planning taxi stops, taxi companies to predict demand, as well as taxi drivers and users who can better estimate the waiting time for a taxi. In [24], GPS data on the trajectories of Beijing taxis are used to study the influence of network companies such as Uber and Didi on taxis in the city. It is concluded that with the advent of taxi network services, the number of trips per day per taxi decreased by 18.08%, and the average daily income by more than 19%. The authors also compare strategies for searching and delivering passengers by taxi drivers with high and low efficiency in different periods. Research [48] is devoted to the detection of anomalous trajectories. According to the authors, anomalous trajectories are trajectories chosen by a small number of drivers that differ from the normal trajectories chosen by other drivers. Such trajectories can serve as signals of incidents in urban traffic systems or fraudulent activities on the part of drivers and passengers. The researchers propose an anomaly detection algorithm based on hierarchical clustering of different trajectories with the same source-destination pairs. They identify four main patterns of abnormal behavior. Computational experiments show the effectiveness of the proposed method for detecting fraudulent routes and adverse road events. The study [12] examines the impact of precipitation on taxi demand. The authors introduce the concept of a high-income passenger to study the demand for a taxi and the driver's efficiency. The researchers show that increased rainfall intensity has the opposite effect on passenger demand for taxis during weekday peak hours and off-peak hours. It is noted that a deeper study of the activities of a taxi is possible using data from taxi calling applications, sensors, as well as financial reports of drivers. Data on the taxi trajectories obtained from GPS sensors also are used to estimate the congestion of city roads. The study [30] proposes a method for calculating the road congestion factor based on the average speed of taxi drivers. As a result of a computational experiment carried out using data on the trajectories of Beijing taxi drivers, the authors set the time limits for the morning and evening rush hours. They estimate the intensity of traffic on the city roads at various time intervals. It notes that the workload in the morning peak hours is higher than in the evening. There are no peak hours on weekends. In the study [28], GPS data on the trajectories of taxi drivers in the city are also used to estimate traffic intensity. Based on this assessment, a model for the analysis of urban transport emissions is proposed. The territory of the city is divided into traffic analysis zones (TAZ). Within them, instantaneous emissions of CO_2, NO_x are estimated. The authors show the relationship between the density of roads and the number of emissions in the TAZ. They conclude that the highest emissions are in TAZs with large business centers. The authors show that within Beijing's fifth ring road, emissions are higher in the north than in the south. The results of this study can be used by the city authorities for lowcarbon urban transport planning, the promotion of alternative energy vehicles, and the design of charging stations. The article [52] is also devoted to the study of emission levels in megacities. The authors analyze data from more than three million GPS trajectories of mobile subscribers obtained in Setagaya, Tokyo. They propose a method to reduce harmful emissions by changing the bike-sharing system. The proposed method showed a reduction in emissions by more than 6% compared to previous methods. The materials of this study are useful in the design and improvement of bicycle-sharing systems around the world. The study [40] is devoted to the comparison of harmful emissions from classic taxis and taxi drivers working with aggregators. In it, the authors compare the trajectories of Didi taxi drivers with ordinary taxi drivers working without an aggregator. The study found that DiDi drivers drive less in search of passengers, heading directly to the pick-up point. Fuel consumption and carbon monoxide emissions, nitrogen oxides, and hydrocarbons per passenger-kilometer are about 1.36 times higher in classic taxi rides.

Bus fleet data. GPS data on vehicle trajectories are massively collected not only in taxi services but also in bus networks. A model for reducing public transport emissions based on GPS trajectory data using the concept of individual buses is proposed in [50]. During a computational experiment conducted on trajectory data recorded by mobile phones in Tokyo, 29 potential individual bus routes are calculated. The researchers identify three types of routes: radial, circular and suburban. The estimated emissions reduction is 13%. The authors find convenient places for the proposed stops of individual buses. This study may be useful to city authorities in the implementation of the concept of individual buses. In [27], the relationship between critical driving events (long stop, hard acceleration, and hard deceleration) and crashes is investigated. The authors use Spearman's rank correlation coefficient based on data on the trajectories of 300 Orlando buses. They find that sudden acceleration and long stops are positively associated with traffic accidents involving pedestrians and cyclists. The authors of the article propose to use the materials of their research in the design and implementation of proactive traffic safety management systems. Paper [53] proposes a framework for evaluating the performance of bus routes based on GPS trajectory data collected in Jinan, China. Several important bus performance metrics are studied, including route times, stop times, idle times, and groupings of buses. The results show that the travel times of the routes follow a correct skewed distribution. In addition, the passage time of a section between two successive stops varies at different periods and is longer during the evening peak hours. The article [41] is devoted to the solution to the problem of planning dedicated lanes for buses. The authors formulate the problem of planning dedicated lanes for buses as a multiobjective optimization problem in which road conditions, traffic flow, bus lane connectivity, and construction cost act as constraints. The use of the road and the punctuality of the bus are taken as objective functions. A method based on an evolutionary algorithm is presented for solving the problem. The operation of the method is illustrated by a computational experiment conducted on GPS data from buses in Shenzhen, China. Often the trajectories of buses following a fixed route are stored in a compressed form. The data includes only relevant bus station arrival and departure records. The article [37] presents a BVI system for visual data analysis. This system contains four data analysis modules. The first module cleans and displays sparse trajectory data. The second module is responsible for analyzing the state of global traffic and traffic patterns of road sections. In the third module, an analysis is made of bus station congestion patterns, and in the fourth, an analysis is made of the importance of bus stations in a complex public transport network. The authors demonstrate the performance and efficiency of the proposed system in three experiments using a data set of real bus GPS trajectories.

GPS data from other sources. The study [7] is devoted to ensuring safety at construction sites based on the analysis of GPS trajectories of builders. The paper proposes a system that processes GPS data, calculates stopping points, trajectory intersections, and provides this information to safety managers. The main distinguishing feature of the proposed system is that it does not use the GPS data directly, but the processed spatio-temporal trajectory data. The materials of this article are useful to identify potentially unsafe behavior at construction and other facilities. In [16], a method for transportation mode prediction based on GPS trajectory data is presented. The main part of the proposed method is preparing data for classifiers (random forests, decision trees, nearest neighbors, Naive Bayes). It consists of five stages. At the first stage, points of GPS trajectories are grouped. Then points signs of trajectories are generated. Trajectory characteristics (percentiles, medians, etc.) are highlighted. Noise is removed from the obtained data. Normalization is carried out. Classifiers built on prepared data show higher accuracy, surpassing classifiers built using other data preparation methods. City authorities need to create a comfortable urban environment for pedestrians. Active walking helps to improve the health of the city's population and improve the quality of life in the city. The study [38] is devoted to the analysis of pedestrian behavior. The authors use GPS data on the trajectories of pedestrians collected using a special application for smartphones. The authors show the influence of various street attributes, which, as is known from previous studies, influence the choice of a walking route. Unlike most studies where the data is limited to a specific type of destination (such as public transit stops), this paper examines a set of trajectories from a wide range of destinations and geographic regions. To evaluate the choice of paths, a new method of alternative path creating is used. The proposed approach obtains information about the attributes of the route using Google Street View image analysis.

The paper [55] proposes a system for constructing a city pedestrian network. This system includes three key modules. The first module is filtering data on walking trajectories. The second is building a pedestrian network and the third is its evaluation. Data for analysis is obtained using crowd-sensing from the GPS sensors of phones. The authors conduct an experiment showing that the pedestrian network extracted using the proposed system is accurate and complete. The work [25] is devoted to modeling the trajectories of pedestrians in the city. An important problem in the analysis of pedestrian trajectory data is large errors in positioning caused by large buildings and frequent stops, and direction changes. The authors proposed a system for modeling pedestrian trajectories that solve such problems. The paper describes in detail the architecture of the system, and the tools necessary to implement such an architecture. The study [46] proposes a method for people crowds detection based on GPS movement trajectories data. The main feature of the proposed method is resistance to noise and missing data, which is typical for data collected in urban areas. The results of the computational experiment show that the method accuracy of detecting crowds and isolating their members is 91.3

3. Smart card data

Currently, smart cards are widely used to pay fares in the transport systems of many cities. Data on the validation of transport cards are collected by special systems and are utilized mainly for invoicing. These data can also be used to solve other applied problems. Example of individual transsport data is presented in Table 3 A large number of individual transport contributes to an increase in the number of harmful emissions into the atmosphere. The main alternative to individual transport is public transport. The use of it is more environmentally friendly. For increasing public transport use, it is necessary to identify the reasons why urban transport is not attractive to citizens. The study [36] is devoted to identifying factors contributing to the public transport use reduction. The authors built a Cox regression model on features obtained from smart card data. These features include for each passenger the share of weekdays using public transport, the number of pairs of places of departure and arrival, the share of tram use, the type of transport card, and others. Based on the analysis, the authors propose incentive measures to maintain and increase the use level of public transport for various population groups.

Table 3

Smart card data example

Timestamp	Validator id	Smart card id	Balance
20/01/22 09:30:55	0	125558	500
20/01/22 09:31:32	1	136472	362
20/01/22 09:32:05	0	123657	414
20/01/22 09:32:55	0	130058	580

In Brisbane, Queensland, Australia, smart cards are used to access CityCat ferry transport. The article [39] provides a detailed study of the data of more than 1.5 million smart card data transactions. The authors establish that, despite the presence of only one route, more than 2% of trips on public transport are made on these ferries. The article notes that the use of ferries is mainly for going to work and school during rush hours. The ferry use level at the weekend remained at the same level. A cluster of users using the ferries for one-time recreational purposes is identified. A high degree of integration of the CityCat system with other transportation modes is established. More than 15% of trips continued using other transportation modes. The paper [54] is devoted to the transfer identification problem. This paper investigates transfers between subways and bike-sharing systems. An important difficulty in conducting such studies is the use of various transport cards for metro travel and bicycle rental. The authors proposed a method for comparing data from smart cards of two different transport systems for one passenger. In the computational experiment, this method showed an accuracy of 100% for 573 passengers under study. The authors examine the identified transfers and conclude the movement of passengers. The authors note that 2/3 of bike-to metro transfers and vice versa take place during peak hours, and bike-to-bike and bike trip times account for an average of 27% of the total trip time. For the correct distribution of transport resources in the city and the prompt response to various incidents, the city administration needs to understand how traffic flows change during the day, and in emergencies. The article [51] is devoted to the subway daily traffic fluctuation study. The authors use data from more than a million trips over five working days with normal weather conditions in Nyanzhin city. The researchers identify the coefficient of increase in traffic during peak hours, analyze and compare passenger traffic on different metro lines and at different stations. The authors use thermodynamic diagrams to visualize the inflow and outflow of passengers at stations. This representation allows researchers to identify congestion in the metro in the city. Also, it helps to visualize the features of passenger traffic in the city. The article [34] is devoted to the accident impact study on the city transport system. The authors found that medium-term disruptions can have long-term consequences for the travel patterns of long-term users of the affected infrastructure. They note that their method is one of the first in this area, using passively collected data. Other studies in this area use data from questionnaires or surveys that require participation from passengers. One of the main problems of modern megacities is their monocentric. According to our study [9], the working districts of Moscow are concentrated in the center of the city. This leads to the fact that urban residents from peripheral areas are forced to spend a lot of time traveling to their places of work and also stimulates the use of personal transport. This leads to an increase in harmful emissions into the atmosphere in the city center. The study [31] is devoted to identifying spatiotemporal patterns of trips to work for public transport passengers in Beijing. Using one month's smart card validation data, the authors identify the places of work and residence of individual passengers, as well as the time of their departure. Visualization of the obtained data showed significant differences between workplaces and residences in Beijing. The study materials are useful to build a balanced transport system in a monocentric city. The validator data contains not only information on general traffic flows and individual traffic behavior, but also indirect features describing the transport network users. For example, information on the smart card validation can also be used to obtain socioeconomic features describing their users. Such data are useful for the rational allocation of social subsidies, the detection of potential fraud with social smart cards, and for solving other applied problems. Modern methods for assessing socioeconomic behavior are based on data on the behavior of users in cyberspace, for example, on data from social networks [6]. Article [14] proposes an approach to assess the socioeconomic status of a smart card user based on deep learning. This approach combines the mass nature of modern methods based on data, but at the same time works with data on real, not virtual, user behavior. The paper illustrates the application of the proposed method on the Shanghai SCD dataset containing data on more than a million smart card users. Data similar in nature to smart card validator data can be collected in other ways. The paper [19] proposes an approach for calculating travel time based on data from Wi-Fi scanners. With this approach, passenger devices with a Wi-Fi receiver, such as a smartphone or tablet, act as a smart card, the MAC addresses of their devices act as a passenger identifier, and a Wi-Fi scanner acts as a validator. This approach makes it possible to obtain data that can be used for analysis by the methods described above and expand their scope. It is important to note that modern mobile operating systems may prevent such tracking of device owners to increase privacy. When connected to Wi-Fi, such devices generate random MAC addresses, which limits the application of the method proposed above. Alternatively, an approach can be proposed that uses unique Bluetooth addresses instead of Wi-Fi MAC addresses. Quite a lot of popular wearable wireless devices use this technology (wireless headphones, fitness bracelets, and others), while the Bluetooth address of these devices does not change when connected. Currently, the development and implementation of alternative payment methods to smart cards are underway. An example is the introduction of a payment system using face scanning in the Moscow metro. When using such systems, data on the validation of the transport system user is also saved, so smart card data analysis methods will not lose their relevance. Let us summarize all the presented methods for analyzing individual transport data in the Table 4.

Table 4

Problems and bibliography

Problem	Bibliography
Taxi research and optimization	[22], [24], [45], [48], [12]
Public transport research and optimization	[27], [53], [41], [37], [39], [36], [51], [34], [31
Identification and research of transfers	[23],[54]
City road congestion assessment	[30], [28]
Reducing harmful emissions in megacities	[28], [52], [40], [50]
Transportation mode prediction	[7], [16], [35], [29]
Pedestrian network research	[38], [55], [46], [25]

4. Discussion

Our literature review shows that there are many methods for solving urban transport management problems. In addition, there are ready-made platforms that combine groups of such them. Typically, such platforms, for example [18], con tain a fixed set of algorithms that solve some previously known set of problems. Other platforms, such as [26], fix a specific problem-solving method. In [26], an architecture based on neural networks is fixed for solving regression problems. Such algorithms have low explainability and can be difficult to tune. We propose a new approach to the platform architecture for data analysis in digital urbanism. Amount of tasks that arise in urban infrastructure management reduces to the tasks of finding deviations from some given "normality". Thus, to solve them, the platform must provide modules for working with the initial data, and modules for describing/obtaining normality and searching for deviations from it. There are two data processing modules. The first is for reading and loading data, storing it, and also checking the correctness of the data type. It is important to note that in urban studies, most of the analyzed data can be reduced either to correspondence matrices (cellular operator data, aggregated GPS, or smart card data) or to a form similar to that presented in Table 1 (individual GPS data, smart cards). This feature allows you to work with data from different sources in a unified way. The second module is responsible for the semantic validation of the submitted data according to heuristic rules. The third module of the platform provides some standard ways of describing normalities and methods for detecting deviations. As an example, the normality model presented in [8] can be taken. The API module will allow developers to define their own "normalities" and deviation-detecting methods. This module will allow the platform to overcome the limitations of existing platforms in the form of a fixed range of tasks to be solved and a fixed architecture for their solution since new normality and deviation-detecting methods can be defined on any data of one of two standard types. The results of the algorithms must be visualized in a form convenient for perception. The platform must contain a data visualization module to do this. One of the important parts of such a platform is methods for implementing visualization on city maps, and graphs. Ready-made solutions, such as [2], can be used as the basis for such methods. The platform architecture diagram is shown in Fig. 1

According to this architecture, it is possible to create a comprehensive platform for data analysis in urban studies. The main advantages of such a platform are the ability to work with different data sources, and the ability to dynamically expand the base of platform analysis methods using the API, which will eventually expand the range of tasks to be solved with low labor costs.

Conclusion

In this article, individual transport data analysis methods are considered. Many of these methods can help cities design transportation systems. Data sources, combined with new methods of analysis, help to better understand the transport needs of urban residents and improve the comfort of their movement, as well as reduce travel time. Modern data sources allow answering questions about when, where, and how citizens move. We propose a platform architecture that will allow us to combine many urban data analysis algorithms from different sources. This architecture allows developers to describe the concept of normality and deviations themselves, which makes it possible to dynamically expand the range of available algorithms. The next step in the development of transport systems in the cities of the future may be the development of self-driving vehicles, which will provide a higher quality of service and lower operating costs. In recent years, the development of robotic cars [11], as well as the concepts of their interaction with regular cars driven by people and pedestrians [43,47], are actively carried out. In many respects, this became possible due to the development of algorithms for computer vision [21,42], depth estimation [32,44],

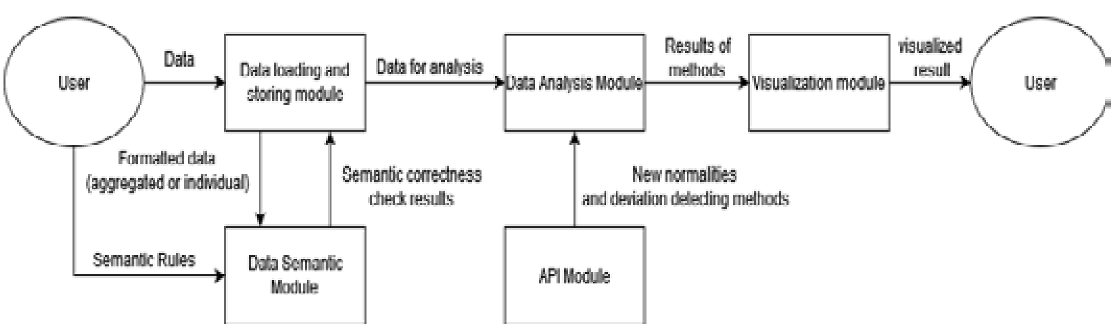

Fig. 1. The platform architecture diagram

and three-dimensional object detection using LiDAR data (3D-object-detection) [49]. With the widespread use of self-driving vehicles, there will be opportunities for building automated logistics systems (autonomous logistics), where artificial intelligence solves not only the tasks of traffic flow planning but also the direct transportation of goods or passengers [5].

References

1. Digital 2021. Report by “We are social” agency, https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/
2. Kepler.gl, https://kepler.gl
3. Measuring digital development Facts and figures 2021, https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf
4. New thinking urgently required to shift urbanisation onto a more sustainable path, https://fennerschool.anu.edu.au/news-events/news/new-thinking-urgentlyrequired-shift-urbanisation-more-sustainable-path
5. *Abosuliman, S.S., Almagrabi, A.O.*: Routing and scheduling of intelligent autonomous vehicles in industrial logistics systems. Soft Computing 25(18), 11975-11988 (2021)
6. *Ajrouch, K.J., Blandon, A.Y., Antonucci, T.C.*: Social networks among men and women: The effects of age and socioeconomic status. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 60(6), S311-S317 (2005)
7. *Arslan, M., Cruz, C., Roxin, A.M., Ginhac, D.*: Spatio-temporal analysis of trajectories for safer construction sites. Smart and Sustainable Built Environment (2018)
8. *Bulygin, M., Namiot, D.*: Anomaly detection method for aggregated cellular operator data. In: 2021 28th Conference of Open Innovations Association (FRUCT). pp. 42-48. IEEE (2021)
9. *Bulygin, M., Namiot, D.*: A new approach to clustering districts and connections between them based on cellular operator data. In: 2021 29th Conference of Open Innovations Association (FRUCT). pp. 71-80. IEEE (2021)
10. *Bulygin, M., Namiot, D.*: On the possibilities of using the data of cellular operators to solve the problems of digital urbanism. International Journal of Open Information Technologies 9(1), 48-57 (2021)
11. *Chan, T.K., Chin, C.S.*: Review of autonomous intelligent vehicles for urban driving and parking. Electronics 10(9), 1021 (2021) 12.
12. *Chen, D., Zhang, Y., Gao, L., Geng, N., Li, X.*: The impact of rainfall on the temporal and spatial distribution of taxi passengers. Plos one 12(9), e0183574 (2017)
13. *Dameri, R.*: Searching for smart city definition: a comprehensive proposal. International Journal of Computers & Technology 11, 2544 (10 2013). https://doi.org/10.24297/ijct.v11i5.1142
14. *Ding, S., Huang, H., Zhao, T., Fu, X.*: Estimating socioeconomic status via temporal-spatial mobility analysis-a case study of smart card data. In: 2019 28th international conference on computer communication and networks (ICCCN). pp. 1-9. IEEE (2019)
15. *Downey, J., McGuigan, J.*: Technocities: The Culture and Political Economy of the Digital Revolution. Sage (1999)
16. *Etemad, M., Soares Júnior, A., Matwin, S.*: Predicting transportation modes of gps trajectories using feature engineering and noise removal. In: Canadian conference on artificial intelligence. pp. 259-264. Springer (2018)
17. *Fassio, O., Rollero, C., De Piccoli, N.*: Health, quality of life and population density: A preliminary study on contextualized quality of life. Social indicators research 110(2), 479-488 (2013)
18. *Golubev, A., Chechetkin, I., Solnushkin, K.S., Sadovnikova, N., Parygin, D., Shcherbakov, M.*: Strategway: web solutions for building public transportation routes using big geodata analysis. In: Proceedings of the 17th international conference on information integration and web-based applications & services. pp. 1-4 (2015)
19. *Hidayata, A., Terabea, S., Yaginumaa, H.*: Time travel estimations using mac addresses of bus, passengers: A point to path-qgis analysis. Geoplanning: Journal of Geomatics and Planning 5(2), 259-268 (2018)
20. *Hofmann-Wellenhof, B., Lichtenegger, H., Collins, J.*: Global positioning system: theory and practice. Springer Science & Business Media (2012)
21. *Hong, Y., Pan, H., Sun, W., Jia, Y.*: Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes. arXiv preprint arXiv:2101.06085 (2021)
22. *Hu, X., An, S., Wang, J.*: Taxi driver’s operation behavior and passengers’ demand analysis based on gps data. Journal of advanced transportation 2018 (2018)
23. *Huang, Z., Xu, L., Lin, Y., Wu, P., Feng, B.*: City-wide metro-to-bus transfer behavior identification based on combined data from smart cards and gps. Applied Sciences 9(17), 3597 (2019)
24. *Jiang, W., Lian, J., Shen, M., Zhang, L.*: A multi-period analysis of taxi drivers’ behaviors based on gps trajectories. In: 2017 IEEE 20th International

Conference on Intelligent Transportation Systems (ITSC). pp. 1-6. IEEE (2017)
25. *Karim, L., Boulmakoul, A., Zeitouni, K.*: From raw pedestrian trajectories to semantic graph structured model towards an end-to-end spatiotemporal analytics framework. Procedia Computer Science 184, 60-67 (2021)
26. *Ke, S., Song, L., Bao, K., Pan, Z., Zhang, J., Zheng, Y.*: East: An enhanced automated machine learning library for spatio-temporal forecasting
27. *Li, P., Abdel-Aty, M., Yuan, J.*: Using bus critical driving events as surrogate safety measures for pedestrian and bicycle crashes based on gps trajectory data. Accident Analysis & Prevention 150, 105924 (2021)
28. *Li, T., Wu, J., Dang, A., Liao, L., Xu, M.*: Emission pattern mining based on taxi trajectory data in beijing. Journal of cleaner production 206, 688-700 (2019)
29. *Liang, X., Zhang, Y., Wang, G., Xu, S.*: A deep learning model for transportation mode detection based on smartphone sensing data. IEEE Transactions on Intelligent Transportation Systems 21(12), 5223-5235 (2019)
30. *Liu, C., Wang, S., Cuomo, S., Mei, G.*: Data analysis and mining of traffic features based on taxi gps trajectories: A case study in beijing. Concurrency and Computation: Practice and Experience 33(3), e5332 (2021)
31. *Ma, X., Liu, C., Wen, H., Wang, Y., Wu, Y.J.*: Understanding commuting patterns using transit smart card data. Journal of Transport Geography 58, 135-145 (2017)
32. *Murez, Z., As, T.v., Bartolozzi, J., Sinha, A., Badrinarayanan, V., Rabinovich, A.*: Atlas: End-to-end 3d scene reconstruction from posed images. In: European Conference on Computer Vision. pp. 414-431. Springer (2020)
33. *Namiot, D., Sneps-Sneppe, M.*: A survey of smart cards data mining. In: AIST (Supplement). pp. 314-325 (2017)
34. *Nazem, M., Lomone, A., Chu, A., Spurr, T.*: Analysis of travel pattern changes due to a medium-term disruption on public transit networks using smart card data. Transportation Research Procedia 32, 585-596 (2018)
35. *Ni, Q., Fan, Z., Zhang, L., Zhang, B., Zheng, X., Zhang, Y.*: Daily activity recognition and tremor quantification from accelerometer data for patients with essential tremor using stacked denoising autoencoders. International Journal of Computational Intelligence Systems 15(1), 1-13 (2022)
36. *Nishiuchi, H., Chikaraishi, M.*: Identifying passengers who are at risk of reducing public transport use: A survival time analysis using smart card data. Transportation research procedia 34, 291-298 (2018)
37. *Pei, W., Wu, Y., Wang, S., Xiao, L., Jiang, H., Qayoom, A.*: Bvis: urban traffic visual analysis based on bus sparse trajectories. Journal of Visualization 21(5), 873-883 (2018)
38. *Sevtsuk, A., Basu, R., Li, X., Kalvo, R.*: A big data approach to understanding pedestrian route choice preferences: Evidence from san francisco. Travel behaviour and society 25, 41-51 (2021)
39. *Soltani, A., Tanko, M., Burke, M.I., Farid, R.*: Travel patterns of urban linear ferry passengers: Analysis of smart card fare data for brisbane, queensland, australia. Transportation Research Record 2535(1), 79-87 (2015)
40. *Sui, Y., Zhang, H., Song, X., Shao, F., Yu, X., Shibasaki, R., Sun, R., Yuan, M., Wang, C., Li, S., et al.*: Gps data in urban online ride-hailing: A comparative analysis on fuel consumption and emissions. Journal of Cleaner Production 227, 495-505 (2019)
41. *Sun, Y., Wu, M., Li, H.*: Using gps trajectories to adaptively plan bus lanes. Applied Sciences 11(3), 1035 (2021)
42. *Tao, A., Sapra, K., Catanzaro, B.*: Hierarchical multi-scale attention for semantic segmentation. arXiv preprint arXiv:2005.10821 (2020)
43. *Tran, T.T.M., Parker, C., Tomitsch, M.*: A review of virtual reality studies on autonomous vehicle-pedestrian interaction. IEEE Transactions on Human-Machine Systems (2021)
44. *Wang, L., Zhang, J., Wang, O., Lin, Z., Lu, H.*: Sdc-depth: Semantic divideand-conquer network for monocular depth estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 541-550 (2020)
45. *Wang, Y., Qin, K., Chen, Y., Zhao, P.*: Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi gps data. ISPRS International Journal of Geo-Information 7(1), 25 (2018)
46. *Wirz, M., Schläpfer, P., Kjærgaard, M.B., Roggen, D., Feese, S., Tröster, G.*: Towards an online detection of pedestrian flocks in urban canyons by smoothed spatio-temporal clustering of gps trajectories. In: Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks. pp. 17-24 (2011)
47. *Xing, Y., Lv, C., Cao, D., Hang, P.*: Toward human-vehicle collaboration: Review and perspectives on human-centered collaborative automated driving. Transportation research part C: emerging technologies 128, 103199 (2021)

48. *Yang, S., Bi, S., Athanase, N., Huang, T., Wan, L.*: Spatial clustering method for taxi passenger trajectory. Computer Engineering and Applications 54(14), 249-255 (2018)
49. *Yin, T., Zhou, X., Krahenbuhl, P.*: Center-based 3d object detection and tracking. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11784-11793 (2021)
50. *Yu, Q., Zhang, H., Li, W., Song, X., Yang, D., Shibasaki, R.*: Mobile phone gps data in urban customized bus: Dynamic line design and emission reduction potentials analysis. Journal of Cleaner Production 272, 122471 (2020)
51. *Yu, W., Bai, H., Chen, J., Yan, X.*: Analysis of space-time variation of passenger flow and commuting characteristics of residents using smart card data of nanjing metro. Sustainability 11(18), 4989 (2019)
52. *Zhang, H., Song, X., Long, Y., Xia, T., Fang, K., Zheng, J., Huang, D., Shibasaki, R., Liang, Y.*: Mobile phone gps data in urban bicycle-sharing: Layout optimization and emissions reduction analysis. Applied Energy 242, 138-147 (2019)
53. *Zhang, H., Shi, B., Song, S., Zhao, Q., Yao, X., Wang, W.*: Statistical analysis of the stability of bus vehicles based on gps trajectory data. Modern Physics Letters B 33(03), 1950015 (2019)
54. *Zhao, D., Wang, W., Ong, G.P., Ji, Y.*: An association rule based method to integrate metro-public bicycle smart card data for trip chain analysis. Journal of Advanced Transportation 2018 (2018)
55. *Zhou, B., Zheng, T., Huang, J., Zhang, Y., Tu, W., Li, Q., Deng, M.*: A pedestrian network construction system based on crowdsourced walking trajectories. IEEE Internet of Things Journal 8(9), 7203-7213 (2020)

Bulygin M.V. PhD student, Lomonosov Moscow State University, MSU, Faculty of Computational Mathematics and Cybernetics, Russia, 119991, Moscow, GSP-1, 1-52, Leninskiye Gory, e-mail: messimm@yandex.ru (correspondent author)

Namiot D.E. Dr. of Sci., Lomonosov Moscow State University, MSU, Faculty of Computational Mathematics and Cybernetics, Russia, 119991, Moscow, GSP-1, 1-52, Leninskiye Gory, email: dnamiot@gmail.com

Pokusaev O.N. PhD, docent, Russian Transport University, MIIT, Higher Engineering School, Russia, 127055 Moscow, Novosushchevskaya st., 22, building 2, email: dnamiot@gmail.com