2020 Global Smart Industry Conference (GloSIC)

Digital Twin of City: Concept Overview

Sergey Ivanov
School of Electronic Engineering and Computer Science
South Ural State University (national research university)
Chelyabinsk, Russian Federation
saivanov@susu.ru

Ksenia Nikolskaya
School of Electronic Engineering and Computer Science
South Ural State University (national research university)
Chelyabinsk, Russian Federation
nikolskaiaki@susu.ru

Gleb Radchenko
School of Electronic Engineering and Computer Science
South Ural State University (national research university)
Chelyabinsk, Russian Federation
gleb.radchenko@susu.ru

Leonid Sokolinsky
School of Electronic Engineering and Computer Science
South Ural State University (national research university)
Chelyabinsk, Russian Federation
leonid.sokolinsky@susu.ru

Mikhail Zymbler
School of Electronic Engineering and Computer Science
South Ural State University (national research university)
Chelyabinsk, Russian Federation
mzym@susu.ru

***Abstract*—This article presents the concept of the digital twin of a city. The urban economy is a complex multi-vector system. Creating a single digital twin of such a system is now a difficult task to solve. In this article the concept of a digital twin of a city is presented. The authors propose an evolutionary approach to this problem, according to which digital twins of individual elements of the urban environment are consistently built on a single hardware and software platform. These digital twins are linked in a single cooperative system that allows one digital twin to use data produced by other digital twins. The article gives the definition and architecture of such a system. It describes the classes of models that can be used to create digital twins. Special attention is paid to neural network models and models for data analysis. The information infrastructure of the digital twin of a city, including sensory networks, data cleansing mechanisms and nebulous calculations, is considered.**

***Keywords*—digital twin, smart city, urban management, neural networks, data mining, sensors, fog computing**

I. Introduction

Information technologies have become the main driving force of social and economic development today. Innovative methods of data gathering, and analysis are gradually replacing well-established mechanisms of city management. Unlike statistical samples that tend to become outdated by the time they are analyzed, "big data" can be processed in real-time, improving the quality and speed of decision making. Big Data in urban governance complements and expands the traditional types of information about a city. Thus, thanks to Big Data, it became possible to monitor behavioral patterns and analyze urban lifestyle at the intersection of such well-known categories as population, economic development, construction, and infrastructure, etc.

The digital revolution and the global spread of the Internet have given rise to such phenomena as "Data-Driven City" (DDC) [14] and "Smart City". Smart City can be defined as a strategic approach to integrating data and digital technologies to ensure sustainability, citizen welfare and economic development of the urban environment [37]. The Smart City concept defines a space in which the key components of the urban infrastructure (environment, emergency management, traffic management, and power) are integrated in such a way that their functions and capabilities can easily be combined with each other as well as with new systems [29].

The Russian Ministry of Construction Industry in cooperation with Lomonosov Moscow State University has developed the "IQ of cities" index aimed to digitalization of urban economy within the framework of the "Smart City" project, which is being implemented within the framework of two national projects, namely "Housing and Urban Environment", and "Digital Economy". "IQ of cities" is calculated in ten areas (urban management, intelligent housing and utilities, innovation for the urban environment, smart urban transport, intelligent systems of public and environmental safety, tourism and services, intelligent systems of social services, economic conditions and investment climate, infrastructure communications networks), and contains 47 indicators [17]. One of the important indicators of "IQ of cities" is the presence of a digital twin of the city.

The "Digital Twin" (DT) concept ensures the development and support of virtual models of real-world objects and processes. The DT approach relies on the ability to receive and effectively process data flows collected automatically through distributed "Internet of Things" (IoT) sensor systems. The DT of the city is gradually filled with the data of the real city, collected in real-time from deployed IoT infrastructure and urban information systems. DT supports forecasting of changes in the state of urban infrastructure, and to offer optimal solutions by analyzing information on the dynamics of people and transport, their interdependence and their fluctuations in time and space. In addition, regardless of the current state, the digital twin allows analysts to answer "what if" questions, helping to understand how cities equipped with intelligent technology, will function in a particular economic, environmental and social conditions, and identify the factors that contribute to possible failures [28]. The article [32] considers the results of the analysis carried out by the Urban

This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (government order FENU-2020-0022) and by the Russian Foundation for Basic Research (grants No. 20-07-00092, 20-07-00140, and 18-07-01224).

978-1-7281-8075-5/20/$31.00 ©2020 IEEE

Planning Institute of Spatial Modeling and Development "Giprogor Project", in which 10 most complex solutions for creating the digital twins of cities were considered. The list includes Singapore, Amaravati, Boston, Newcastle, Jaipur, Helsinki, Rotterdam, Stockholm, Rennes and Antwerp. As noted in the paper, Russia also faces the inevitable process of digitalization of cities, as it leads to a significant increase in the quality of life of citizens. Among the Russian solutions approaching the vision of "Smart City" as a digital service that provides a qualitatively new level of service for the citizens of the city is the "Smart Cities of Rosatom" project [36]. Services for searching for relevant information, tracking the work of urban services, collecting feedback on current urban problems, etc. provided to residents of several cities in Russia as part of this project.

The purpose of this article is to study the existing technologies for building digital twins of cities. For this purpose, we need to define the concept of a digital twin of the city, to analyze the simulation methods that can be used to build digital twins, as well as to systematize information infrastructure, which should ensure the functioning of the digital twin of the city. The article has the following structure. Section I defines a digital twin of the city, describes its architecture and functional requirements. Section II discusses the classes of models used to create digital twins. Section III describes the information infrastructure of the digital twin of the city, including sensor networks, data cleansing and fog computing. The results are summed up in the conclusion.

II. Digital Twin of the City: Definition, Structure, Functionality

Digital Twin (DT) is an integrated multi-physical, multi-scale probabilistic simulation of a complex object that uses physical, mathematical, simulative and other models to obtain the most accurate representation of the corresponding real object based on analysis of data from sensor networks and other sources [21].

The digital twin of a city is a system of interconnected digital twins, representing certain aspects of the functioning and development of the urban environment. These digital twins support fine-tuning and synchronization with the real state of urban infrastructure through data from various sources in real-time [32], [35]. A continuous flow of data generated by different sources in the digital infrastructure of a smart city is the key to the effective functioning of the city digital twin. We can highlight the following as an example of such sources (see Fig. 1).

- Information about traffic flow of city residents, including information about the traffic of private, commercial, and public transport, as well as traffic congestion collected through various mechanisms (information about transactions of the "single travel card", the results of traffic monitoring, etc.).
- Information about physical parameters of the urban environment obtained in real-time from arrays of intelligent sensors (both private and public) allows monitoring and analysis of such parameters as air temperature and humidity, the number of suspended particles and chemical composition of air, noise pollution, radiation level, the chemical composition of water, etc. linked to the geographical position.
- Data from outdoor surveillance cameras allow for intelligent analysis of such characteristics of the urban environment, which is impossible or difficult to collect by other means (traffic congestion along freeways and pedestrian roads, contamination, and quality of the road network, identification of individual objects and events).
- Data from open sources (such as open state portals and services, data on meteorological conditions, open reporting information of business entities, etc.) allow us to enrich models of intelligent data analysis.

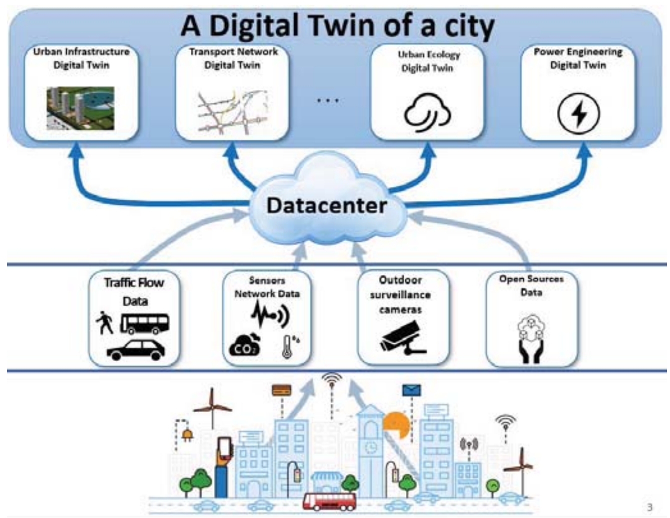

Fig. 1. An example of the interface's appearance

The collected data allow us to provide the functioning of the complex of digital twins of the city. We can point out the following examples of such digital twins [8]:

- A digital twin of urban infrastructure, which is an interactive 3D model of buildings, structures, engineering communications and other urban infrastructure.
- A digital twin of the transport network, which provides monitoring and forecasting of development of the situation of transport availability, the efficiency of public transport, etc.
- A digital twin of urban ecology that provides mechanisms for monitoring and forecasting the environmental condition of the urban environment, including the quality of soil, water, air, etc.
- A digital twin of power engineering, etc.

The digital twin of the city provides the following features:

- Monitoring of the current state of the urban environment.

- Rapid response to emergencies.
- Efficiency assessment of design solutions.
- Identification of sources of potential risks.
- Forecasting of situation development considering historical data.

The digital twin of the city is an extremely complex, integrated solution, the development of which can and should be implemented in stages, through evolutionary development and integration of specific solutions aimed at solving the most urgent ("point") problems. The analysis shows that such solutions are now being implemented everywhere in the world, not only in advanced metropolitan areas but also in large and medium-sized cities. The application of such systems, together with the "Internet of Things" technologies makes it possible to solve several problems on a completely different level [26].

- Pollution control and regulatory impact analysis through an environmental survey.
- Microclimatic weather forecasts based on an urban sensor network.
- Efficiency improvement and cost reduction through waste removal and recycling on demand rather than on a schedule.
- Improving the situation on the roads and saving fuel through smart traffic lights and road markings.
- Rational energy consumption thanks to city lighting on demand.
- Optimization of snow removal operations thanks to real-time data about the situation on the roads, weather conditions and the nearest snow removal vehicles.
- A smart irrigation system in parks and public areas that considers weather conditions and current conditions.
- Smart surveillance cameras for crime tracking and the AMBER Alert automated alert system in real-time.
- Smart parking solutions that help you automatically find the best parking spot.
- Monitoring the wear and tear and condition of bridges, streets, and urban infrastructure to ensure timely maintenance and extended service life.

As an example, we can point out the digital twin of the Takamatsu city, Japan, with a population of 420 thousand people. As a pilot project, the city authorities have implemented two digital twins: the digital twin of monitoring and prevention of emergencies (especially flooding), and the digital twin of tourist attraction of the city [22].

The digital twin for monitoring and prevention of emergency situations is based on the collection and analysis of data from water level sensors located throughout the city and provides real-time monitoring of flood risk in each of the city districts. Also, the system provides monitoring of the condition of shelters for city residents by collecting information from humidity sensors and electricity consumption in each of the shelters. Mobile application available to every resident of the city allows not only to notify in advance the residents of the area at risk but also to provide all necessary information in case of emergency, including the way to the nearest functioning shelter.

Monitoring of the movement of rented bicycles (as the most attractive method of moving around the city for tourists) was introduced in terms of the digital twin of tourist attractiveness of the city. Comparison of the geo-track movement of rented bikes with the data on the tourist (age, gender, nationality, etc.) allows forming a map of the most tourist important points of the city, to plan and evaluate the effectiveness of marketing campaigns to promote tourist attractions (including support for the multi-lingual environment, etc.).

III. Model Classes

Digital twins use various approaches to model real objects and technological processes, including statistical and intelligent data analysis methods, computational modeling methods such as finite element method, etc. [25]. Each of these methods needs specific computational resources. For example, data mining methods require a warehouse with large storage capacity and high bandwidth to collect and access the analytical data, as well as high scalability of the computer system to process them; machine learning methods require computer nodes with graphic accelerators installed, and models employed the finite element method require high performance CPUs with large amount of RAM [4]. This section discusses different classes of models used to create digital twins.

The *physical model* provides computer simulation of physical processes running in time. The construction of such models in practice is associated with the use of modern modeling packages, such as ANSYS. Basic physical models used to build digital twins:

- models based on the Navier–Stokes equation (simulation of hydro-gasodynamic processes);
- models based on the finite element method (simulation of structures made of solid materials).

An example of physical model application for building a digital twin can be found in [27]. The authors present the results of numerical modeling of dispersion in the atmosphere of air pollutants generated by flue gas extraction from fuel combustion at thermal power plants. The model employs the Navier–Stokes equation and is implemented through the ANSYS CFX-CFD package.

The *optimization model* provides search of target function optimums in case of restrictions with the use of mathematical methods. The structure of the optimization model consists of the target function, an area of acceptable solutions and a system of constraints defining this area. The main task of the optimization modeling is to find the function extremum under constraints in the form of equations and inequalities.

The basic approaches to the optimization modeling are as follows:

- Linear programming;
- Mixed Integer Programming;
- Non-linear programming.

An example of the application of optimization models is the optimization of the city street gardening program [41]. This study attempts to define a planning model that determines the location and type of green spaces based on their multiple effects (e.g., cooling and connectivity improvement) and calculates the cost of implementation using meta-heuristic optimization algorithms. Another example is optimization of road construction [6]. The article solves the problem of searching for the best construction plan taking into account preconditions and interdependencies of benefit from project completion. The problem is formulated as a two-level problem where the objective function is to minimize generalized costs, and the lower level takes into account the route selection by drivers.

The *simulation modeling* is a method of research, in which the system under study is replaced by a computer model, simulating with sufficient accuracy the processes occurring in the real system, with which experiments are conducted to obtain information about this system [24]. Bright representatives of simulation models can serve as models of distribution of energy resources to end users [19]. Another striking example is the prototype of the system presented in the article [9]. The prototype includes a three-dimensional model of the built-up environment, a street network model using the theory and method of space syntax, urban mobility modeling, wind flow modeling and a number of empirical quantitative and qualitative data using voluntary geographic information (VGI). In addition, the urban digital twin was implemented in a visualization platform for virtual reality.

A. Data Mining Based Models

Models based on data mining are used to discover previously unknown, non-trivial, practically useful, and feasible for interpretation knowledge from the data, which are necessary for making strategic important decisions in various spheres of human activity [13]. As a part of development a digital twin of the city, the data mining based model allows for solving the following main tasks:

- discovery of the impact of factors on each other;
- discovery of the impact of factors on the indicator;
- prediction the values of factors and indicators.

Let us consider methods for solving the tasks above through the example of analyzing the concentration of fine suspended particles in the city air (hereinafter denoted as n and measured in MACs where MAC stands for maximum allowable concentration). Let we have a database containing the sensor data given in Table 1.

Discovery of the impact of factors on each other is performed by the pattern mining methods [1]. The pattern is a stable association rule of the form "IF antecedent, THEN consequent", where non-empty and non-overlapping sets of factors appear as antecedent and consequent. An example of two such rules is presented below:

- IF Wind force>10 m/s AND Humidity>65% THEN Dirty vehicles>75%;
- IF Day=Monday THEN Emissions from city plants>1.5 Ktons AND Emissions from regional plants>3 Ktons.

TABLE I. SENSOR DATA

Day	Meteo factors			Internal factors			External factors		…	Indicator
	Wind force, m/s	***Air humi-dity, %***	***Air t°, °C***	***City traffic intensity, 1..10***	***Emissions from city industrial plants, Ktons***	***Clean streets, 1..10***	***Emissions from regional industrial plants, Ktons***	***Entering dirty vehicles, %***	…	***Concent-ration of fine suspended particles n, MAC***
10 Apr	5	40	12	5	1.3	4	4	65		2
11 Apr	7	60	14	4	0.6	8	2.1	40		1
12 Apr	8	65	13	3	0.5	9	1.2	30		0.3
13 Apr	10	80	16	10	1.6	3	3.9	85	…	3
…										

The stability of a rule is determined through two basic measures: support and confidence. Rule support is the probability of records in the database that contain both antecedent and consequent. Rule confidence is the conditional probability of the presence in the database of records in which the consequent is present, under condition that these records contain antecedent. In practice, rules are chosen as patterns, if their support and confidence exceed the thresholds (*minsup* and *minconf*, respectively) predefined by an expert in a given subject area. For pattern mining, the main algorithms are Apriori [2], FP-Growth [15], and Eclat [43], and their numerous modifications including parallel versions [45].

Discovery of the impact of factors on the indicator can be performed as constructing a hierarchy of factors reflecting the degree of influence of a factor on the indicator: the higher the factor in the hierarchy, the greater its influence on the indicator. The solution of this task assumes classification of data in the sensor database based on decision trees [33]. Fig. 2 shows an example of a decision tree that classify the days of observations by the "Concentration of fine suspended particles" indicator with the following classes: "less than one MAC", "between one and two MACs", "between two and three MACs", and "more than three MACs".

Prediction the values of factors and indicators can be performed based on both neural network models (see Section B) and data mining models. In the latter case, the solution of the problem involves two steps (see Fig. 3). At the first step, the forecast of the values of the factors is carried out based on the regression models [10]. At the second step, using the obtained predicted values of the indicators and the previously constructed classification model, the predicted value of the indicator is calculated.

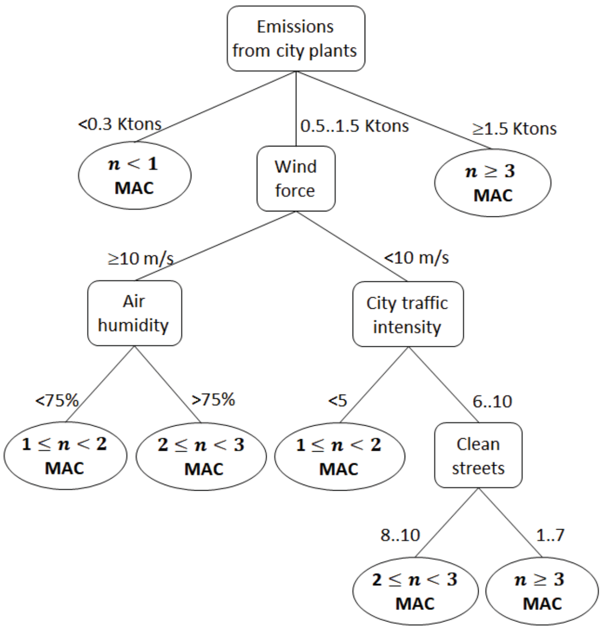

Fig. 2. An example of classification model of data mining

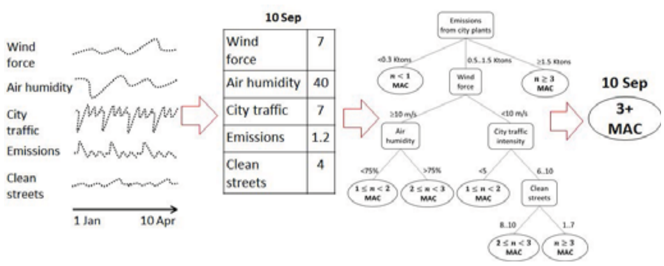

Fig. 3. Model for prediction the concentration of fine suspended particles

B. Neural Network Models

Artificial neural network (ANN) is a mathematical model, as well as its software or hardware implementation, built on the principle of organization and functioning of biological neural networks [44]. The ANN is a system of connected and interacting artificial neurons. Each neuron of such a network deals only with signals that it periodically receives and signals that it periodically sends to other neurons. As a result of training, the neural network is able to detect complex relationships between input and output data, as well as to perform generalization and will be able to return the correct result based on data that was absent in the training sample, as well as incomplete and/or "noisy", partially distorted data.

Convolutional neural network (CNN) - special architecture of artificial neural networks aimed at image analysis. A convolution neural network consists of several layers: convolution layers, sub-sampling layers, and layers of an "ordinary" neural network (full-coupled), a perceptron. The first two layers (convolution and sub-sampling), alternating among themselves, form an input vector of features for the perceptron [3]. The convolution layer includes cores of convolution, which are weight matrices processing the previous layers by fragments. At present, there are many implementations of convolution neural networks for solving tasks of a smart city. Worth-noting examples are as follows. Detectron2 [40] is a software system Facebook AI Research [11] that implements modern algorithms of object detection such as Faster R-CNN, Mask R-CNN, RetinaNet, DensePose and others. YOLOv3 [34] is a real-time object detection system. Thanks to the wide range of available options, you can choose the version most suitable for the application. For example, Tiny YOLO [12] is the most "compact" version, which can work even on smartphones.

Neural network models are used to solve the tasks of the following classes.

- Detection of objects in the image. For example, detection of motor transport type and presence of dirt or garbage on it, including several objects in one image at once.
- Detection of complex non-linear dependencies between the initial data and target values. For example, building models of dependence of fine-dispersed particles concentration on various factors, assessment of the degree of factors' influence.
- Construction of recommendation systems. For example, decision support systems to manage urban traffic to reduce harmful emissions into the atmosphere.

An example of a neural network model to assess urban traffic at an intersection is AIMS (Artificial Intelligence Monitoring System) [38]. The aim of the project is to develop and implement a system to assess the efficiency of road infrastructure use, forecast traffic congestion and total toxic emissions from motor vehicles. The use of deep-learning neural networks allows collecting, interpreting and aggregation of data on road traffic intensity and classification in real time.

Intelligent system for monitoring traffic flows and road infrastructure allows you to solve a number of complex problems at once, such as collection, interpretation and aggregation of road traffic data, identifying underutilized resources in the transport infrastructure, reducing capital and operating costs. The system makes it possible to evaluate the efficiency of real-time traffic management solutions (changes in traffic signal cycles, transfer of markings, etc.), forecast total emissions of toxic exhaust gases from motor vehicles, taking into account atmospheric and climatic conditions. The program will also warn about exceeding MAC at nodes (intersections) of the street and road network, which will prevent a negative scenario (by increasing the capacity of the node, limiting traffic for trucks).

The technology using trained neural networks, does not require large costs for server equipment and video cameras. To monitor a large intersection, it is often enough to install one street surveillance camera [38]. An example of such system operation is shown in Fig. 4.

Another example of using neural network technologies is the task of detecting garbage on city streets [42]. The general structure of the model operation scheme is shown in Fig. 5.

Fig. 4. Example of neural network operation for car recognition

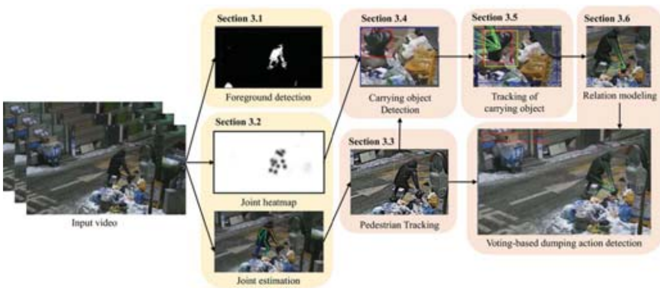

Fig. 5. Example of a neural network operation to detect garbage in the image

The foreground area, joint thermal maps and joint positions are obtained by subtracting the background and joint evaluation. The algorithm then tracks the pedestrians and finds the object they are carrying. Once an object is detected, it is tracked with a correlation filter tracker that works in real time. At the same time, the distance is gradually simulated between the joint of each person and the object being tracked. If a change in distance between an object and a person is detected, the algorithm generates a warning about garbage disposal. Experimental results have shown high efficiency of the proposed method.

IV. Information Infrastructure of the Digital Twin of the City

The information infrastructure of the digital twin of the city includes the following components:

- sensors that collect information,
- networks that transmit data,
- software systems that deal with data cleansing,
- software complexes, which are engaged in intelligent data analysis.

In order for the DT to ensure synchronization between the state of the process in the real world and its virtual copy, we must ensure that the DT can receive, transmit and analyze the data stream from the IoT devices.

A. Sensor Networks

Both existing sensor infrastructure and new sensor clusters can be used to collect information. The following typical types of data sources can be distinguished:

- surveillance cameras;
- pressure sensors;
- humidity sensors;
- temperature sensors;
- air pollution sensors;
- geolocation data of vehicles;
- information from passengers' travel cards, etc.

There are several layers of communication through which data collected from sensor networks pass. One of the variants of such classification is the division of communication channels by the coverage area into the levels of field, local and global computer networks [8].

The field level of communication provides data transmission at distances from dozens of centimeters to hundreds of meters, from the place of data generation to the first node at the edge of the network, responsible for data collection and further retransmission. Low-speed, low-power information channels are used for data exchange at this level. Communication protocols at this level include industrial network protocols, such as Modbus and HART, which are focused on the operation of wired networks such as RS-232/RS-485 and “Current Loop”. Wireless communication technology has also expanded significantly in the last decade, providing opportunities for wireless data transmission from sensor systems. Among the most famous wireless protocols of the field level are Bluetooth; WirelessHART; ZigBee (IEEE 802.15.4); Z-Wave; NFC; RFID and others (see Fig. 6).

The level of local area networks provides data transmission within a local network deployed near data sources (at distances of hundreds of meters to kilometers). Standard network solutions based on the TCP/IP communications protocol stack are mostly used today at this level. The family of Ethernet technologies is used to organize wired networks, and wireless networks are implemented based on a family of Wi-Fi protocols (IEEE 802.11).

At the level of global computer networks, data are transmitted over the Internet. Tele-communication operators, including operators of cellular networks, satellite communications, Low-Power Wide-Area Network (LPWAN, see Fig. 6) are responsible for this level of operation. Such protocols as MQTT, CoAP, HTTP, which provide opportunities for synchronous or asynchronous data processing, are commonly used to organize transmission and efficient data processing in global networks [26]. Backbone telecommunications solutions such as Gigabit Ethernet, EPON, and GPON are used for wired data transmission. Low-power global networks, including those based on LoRa protocol, can be organized for wireless IoT data transmission. It should be noted separately that cellular networks can be used to provide direct connection of IoT devices to Internet nodes based on 2G (GSM), 3G, LTE, 5G technologies.

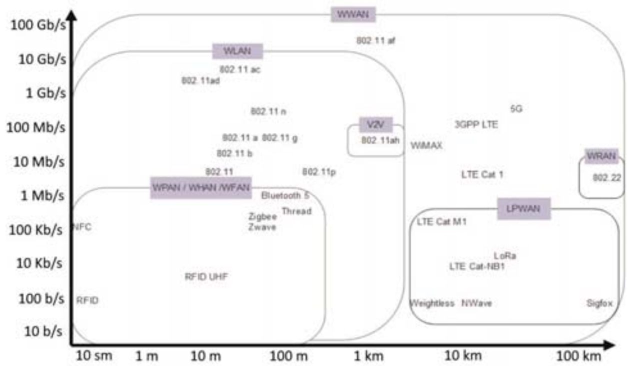

Fig. 6. Diagram of comparison of speed and range of data transmission by different communication channels [26]

B. Data Cleaning

During its work, a sensor may often produce the "dirty" data that contain incorrect, abnormal or missing values. This can be caused by human factors, scheduled sensor maintenance, sensor or communication line failures, etc. The dirty data cannot be subsequently used in the models based on data mining and neural networks since it results in an unreliable output. Thus, for use in the digital twin of the city, the sensor data need to be cleaned [39], [47]. Cleaning sensor data involves the following basic operations: outlier detection and imputation of missing values.

Outlier detection supposes finding single values and subsequences in the time series of the sensor data that are significantly deviate from the rest data of the series [7]. Imputation of missing values involves the generation of the synthetic values of the time series of the sensor data instead of incorrect or missing ones based on analysis of historical data of the given sensor and/or sensors geographically/semantically related to this one [5]. The solution of these tasks can be performed based on both data mining models and neural network models [23], [31] including parallel algorithms [46], [48].

Thus, a digital twin of a city should include an important component, namely Data Cleansing Module (DCM), which provides outlier detection and imputation of missing values in the data streams coming from the system's sensors. DCM is organized in accordance with the following basic principles. First, before entering the Data Center, the data stream of each group of sensors must be processed according to its own cleaning rules. Second, the cleaned data stream of each sensor group must be regularly monitored according to its own verification rules.

The cleaning rule determines the set of correct values for each group of sensors (for example, the minimum and maximum values, the dependence of the values of a given group on the values of sensors of other groups, etc.). The verification rule for a given group of sensors determines the frequency and method for assessing the likelihood of synthetically generated values instead of missing or anomalous ones. To assess the likelihood, DCM exploits various measures (for example, the normalized mean absolute error, the mean square error, etc. [16]) and corresponding thresholds. The low likelihood of synthetic values revealed as a result of verification is a reason to revise methods and algorithms implemented in DCM.

C. Fog Computing

The concept of Fog Computing is useful for processing information from multiple Internet of Things (IoT) sources.

Fog computing is a multi-level model, which is an extension of the concept of cloud computing, facilitating the deployment of distributed applications and services, taking into account network latency, on so-called fog nodes (physical or virtual), located between intelligent end devices and centralized (cloud) services. Fog nodes are context-dependent and support a unified data management and communication system. They can be organized into clusters vertically (to support isolation), horizontally (to support service federations) or linked to the network proximity of smart end devices. Fog Computing minimizes network response time of supported applications and provides end devices with local computing resources and, if necessary, a network connection to centralized services [18].

While cloud computing provides potentially limitless resources for solving tasks that require substantial computing resources, fog computing provides:

- Provisioning of computing resources near endpoint devices.
- Pre-processing of data before it is sent to the cloud.
- Solving tasks that require minimal response time.

The task of data pre-processing from CCTV cameras can serve as an example of using Fog Computing to organize effective data processing. Today, the city can accommodate hundreds of outdoor surveillance cameras, which provide video monitoring services to cover a large area of objects of special interest to urban management. If the resources of a single data processing center (cloud) are used for data analysis, the task of transferring the video data stream through the main channels from such a large array of cameras can be significantly complicated. To solve this problem, one must resort to certain compromises, which include:

- Strong compression of the video stream.
- Reduced resolution of transmitted video information.
- Limitation of the number of transmitted frames per second, etc.

Such solutions result in the lack of quality information to ensure efficient operation of the intelligent model used in the cloud system to analyze incoming data. A possible solution to this problem is to deploy a network of fog nodes near video sources. These nodes consume and perform a preliminary analysis of video data in its original quality, providing solutions to such problems of video data pre-processing as:

- Segmentation and selection of objects of interest.
- Background removal;

- Intellectual compression of the video stream.

This significantly reduces the volume and quality of data transferred to the cloud, improving the quality of the intelligent data analysis system [20], [25], [30].

V. Conclusion

In this article, we analyzed the existing technologies for the construction of digital twins of cities. The definition and key components that make up the city digital twin have been considered. We also examined examples of solutions that can already be considered the first steps to building a full-featured digital twin.

We have considered the key methods for simulation of real objects and technological processes, which find their application in the implementation of digital twins. Separately, based on practical examples, we considered models based on intelligent data analysis, and neural network models.

We also described the key elements of the information infrastructure of the digital twin of a city and considered the key elements of that infrastructure, such as sensor networks, combined with raw data cleaning systems. It was separately noted that for the successful construction of such a system it is necessary to form an information environment based on the concept of the Fog Computing Model, since the tasks of pre-processing data flows often require substantial computational resources for pre-processing data located close to the sources.

References

[1] C. C. Aggarwal and J. Han, Frequent pattern mining. Springer, 2014. DOI: 10.1007/978-3-319-07821-2

[2] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," Proc. of 20th Int. Conf. on Very Large Data Bases, pp. 487–499, 1994. [Online]. Available: http://www.vldb.org/conf/1994/P487.PDF.

[3] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," The Int. Conf. on Engineering and Technology, pp. 1–6, 2017. DOI: 10.1109/ICEngTechnol.2017.8308186

[4] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, and G. K. Karagiannidisa, "Efficient Machine Learning for Big Data: A Review," Big Data Research, vol. 2, no. 3, pp. 87–93, 2015. DOI: 10.1016/j.bdr.2015.04.001

[5] G. E. A. P. A. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning," Appl. Artif. Intell., vol. 17, no. 5–6, pp. 519–533, 2003. DOI: 10.1080/713827181

[6] S. A. Bagloee, M. Sarvi, M. Patriksson, and M. Asadi, "Optimization for roads' construction: selection, prioritization, and scheduling," Computer–Aided Civil and Infrastructure Engineering, pp. 833–848, 2018. DOI: 10.1111/mice.12370

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surv., article no. 15, 2009. DOI: 10.1145/1541880.1541882

[8] Data-Driven Smart Cities: Big Data, Analytics and Security. Skelia. 2018. [Online]. Available: https://skelia.com/articles/data-driven-smart-cities-big-data-analytics-and-security/

[9] F. Dembski, U. Wössner, M. Letzgus, and M. Ruddat, "Urban digital twins for smart cities and citizens: the case study of Herrenberg, Germany," Sustainability. vol. 12, article 2307, 2020. DOI: 10.3390/su12062307

[10] N. R. Draper and H. Smith, Applied Regression Analysis. Wiley, 1981.

[11] Facebook AI Research. [Online]. Available: https://ai.facebook.com/

[12] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: A real-time object detection method for constrained environments," IEEE Access pp. 1935–1944. DOI: 10.1109/ACCESS.2019.2961959

[13] W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge Discovery in databases: an overview," Knowledge Discovery in Databases, pp. 1–30, 1991.

[14] From concept to application solutions. Data-driven cities. PwC. [Online]. Available: https://www.pwc.ru/ru/government-and-public-sector/assets/ddc_rus.pdf

[15] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," ACM SIGMOD Record, vol. 29, no. 2, pp. 1–12, 2000.

[16] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," Int. Journal of Forecasting, vol. 22, no. 4, pp. 679–688, 2006. DOI: 10.1016/j.ijforecast.2006.03.001

[17] Index-IQ of large Russian cities. [Online]. Available: https://www.minstroyrf.ru/upload/iblock/2a5/Indeks-IQ_Krupneyshie.pdf.

[18] M. Iorga, L. Feldman, R. Barton, and M. J. Martin, "Fog computing conceptual model Special Publication (NIST SP)-500-325". [Online]. Available: https://www.nist.gov/publications/fog-computing-conceptual-model

[19] M. N. Iqbal and L. Kütt, "End-user electricity consumption modelling for power quality analysis in residential building," 19th Int. Scientific Conf. on Electric Power Engineering, pp. 1–6, 2018. DOI: 10.1109/EPE.2018.8396030

[20] G. Gardašević, L. Berbakov, and A. Mastilović, "Cybersecurity of Industrial Internet of Things," Cyber Security of Industrial Control Systems in the Future Internet Environment, pp. 47–68, 2020. DOI: 10.4018/978-1-7998-2910-2.ch003.

[21] E. Glaessgen and D. Stargel, "The Digital Twin Paradigm for Future NASA and U.S.," Air Force Vehicles. 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conf., 2012, 2012. DOI: 10.2514/6.2012-1818

[22] I. Kazuhiko and Y. Atsush, "Building a common smart city platform utilizing FIWARE (case study of Takamatsu city)," NEC Tech. J., vol. 13, no. 1, pp. 28–31, 2018.

[23] J. Kim, D. Tae, and J. Seok, "A survey of missing data imputation using generative adversarial networks," Proc. of the 2020 Int. Conf. on Artificial Intelligence in Information and Communication, pp. 454–456, 2020. DOI: 10.1109/ICAIIC48513.2020.9065044

[24] W. Kin and V. Chan, Foundations of Simulation Modeling, Wiley Encyclopedia of Operations Research and Management Science. Wiley, 2011.

[25] P. Korambath, J. Wang, A. Kumar, and J. Davis, "A smart manufacturing use case: furnace temperature balancing in steam methane reforming process via Kepler workflows," Procedia Computer Science, vol. 80, pp. 680–689, 2016. DOI: 10.1016/j.procs.2016.05.357

[26] P. Li, Architecture of the Internet of Things. Litres, 2019.

[27] O. Mărunţălu, G. Lăzăroiu, E. E. Manea, and D. A. Bondrea, "Numerical simulation of the air pollutants dispersion emitted by CHP using ANSYS CFX," World Academy of Science, Engineering and Technology, Int. Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering, vol. 9, pp. 1058–1064, 2015. DOI: 10.5281/zenodo.1108252

[28] N. Mohammadi and J. E. Taylor "Smart city digital twins," IEEE Symposium Series on Computational Intelligence, pp. 1–5, 2017. DOI: 10.1109/SSCI.2017.8285439

[29] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions," Proc. of the 12th annual int. digital government research conference: digital government innovation in challenging times, pp. 282–291, 2011. DOI: 10.1145/2037556.2037602

[30] S. Y. Nikouei, Y. Chen, A. Aved, and E. Blasch, "I-ViSE: Interactive video surveillance as an edge service using unsupervised feature queries," IEEE Internet of Things Journal, pp. 1–1. DOI: 10.1109/jiot.2020.3016825

[31] M. S. Osman, A. M. Abu-Mahfouz, and P. R. Page, "A survey on data imputation techniques: water distribution system as a use case," IEEE

Access, vol. 6, pp. 63279–63291, 2018. DOI: 10.1109/ACCESS.2018.2877269

[32] E. Povh, “10 digital city twins,” RBC-Realty. [Online]. Available: https://realty.rbc.ru/news/5e297b079a79478024d54ff6.

[33] J. R. Quinlan, “Induction of decision trees,” Machine Learning. vol. 1, no. 1, pp. 81–106, 1986. DOI: 10.1023/A:1022643204877

[34] J. Redmon and A. Farhadi, “You only look once: unified: real-time object detection,” Conf. on Computer Vision and Pattern Recognition, pp. 779–789, 2016. DOI: 10.1109/CVPR.2016.91

[35] T. Ruohomaki, E. Airaksinen, P. Huuska, and O. Kesäniemi, “Smart City Platform Enabling Digital Twin,” Int. Conf. on Intelligent Systems, pp. 155–161, 2018. DOI: 10.1109/IS.2018.8710517

[36] Smart cities of Rosatom. [Online]. Available: https://rosatom.city/

[37] Smart Cities Readiness: Smart Cities Maturity Model and Self-Assessment Tool, Scottish Cities Alliance (October 2014). [Online]. Available: https://www.scottishcities.org.uk/site/assets/files/1103/smart_cities_readiness_assess ment_-_guidance_note.pdf

[38] V. Shepelev, S. Aliukov, A. Glushkov, and S. Shabiev, “Identification of distinguishing characteristics of intersections based on statistical analysis and data from video cameras,” Journal of Big Data. vol. 7, no. 1, pp. 1–23, 2020. DOI: 10.1186/s40537-020-00324-7

[39] X. Wang and C. Wang, “Time series data cleaning: A survey,” IEEE Access, vol. 8, pp. 1866–1881, 2020. DOI: 10.1109/ACCESS.2019.2962152

[40] Y. Wu, A. Kirillov, F. Massa, and W. Lo, Detectron2. [Online]. Available: https://github.com/facebookresearch/detectron2

[41] E. J. Yoon, B. Kim, and D. K. Lee, “Multi-objective planning model for urban greening based on optimization algorithms,” Urban Forestry & Urban Greening, vol. 40, pp. 183–194, 2019. DOI: 10.1016/j.ufug.2019.01.004

[42] K. Yun, Y. Kwon, S. Oh, and J. Moon, “Vision‐based garbage dumping action detection for real‐world surveillance platform,” ETRI Journal, vol. 41, no. 4, pp. 494–505, 2019. DOI: 10.4218/etrij.2018-0520

[43] M. J. Zaki, “Scalable algorithms for association mining,” IEEE Transactions on Knowledge and Data Engineering, vol. 12, no. 3, pp. 372–390, 2000. DOI: 10.1109/69.846291

[44] J. Zupan, “Introduction to artificial neural network (ANN) methods: what they are and how to use them,” Acta Chimica Slovenica, vol. 41, no. 3, pp. 327–352, 1994.

[45] M. Zymbler, “Parallel algorithm for frequent itemset mining on Intel many-core systems,” Journal of Computing and Information Technology, vol. 26, no. 4, pp. 209–221, 2018. DOI: 10.20532/cit.2018.1004382

[46] M. Zymbler and Ya. Kraeva, “Discovery of time series motifs on Intel many-core systems,” Lobachevskii Journal of Mathematics, vol. 40, no. 12, pp. 2124–2132, 2019. DOI: 10.1134/S199508021912014X

[47] M. Zymbler, Ya. Kraeva, E. Latypova, S. Kumar, “Cleaning sensor data in smart heating control system,” Global Smart Industry Conf., 2020.

[48] M. Zymbler, A. Polyakov, and M. Kipnis, “Time series discord discovery on Intel many-core systems,” 13th Int. Conf., Revised Selected Papers. Communications in Computer and Information Science, vol. 1063, pp. 168–182, 2019. DOI: 10.1007/978-3-030-28163-2_12