



# Automated Evaluation of Smart City Data from Cloud-System

Iurii Koniaev-Gurchenko

Yuri Gagarin Saratov State Technical University  
Information systems and technologies department  
E-mail: konyaev\_yurii@mail.ru

Hardt Wolfram

Technische Universität Chemnitz  
Department of Computer Science  
E-mail: wolfram.hardt@informatik.tu-chemnitz.de

**Abstract** — Smart city data processing is an important task for the promotion and development of smart cities. The article describes and presents the types of smart city data, discusses the existing modern methods and approaches to the processing of smart city data, such as pre-processing, assessment and analysis, and their tasks. This article contains architectural solutions and methods used in the developed automated smart city data evaluation system. There is also a detailed description of the integration of the developed system with the DriveCloud cloud server for receiving and storing smart city data.

**Keywords** — Smart City, data evaluation, data processing, data pre-processing, data assessment, smart city data.

## I. INTRODUCTION

At present, there is a global growth of the urban population, which leads to the growth of large cities. This leads to changes and problems in all areas of life. The number of resources consumed increases, water, and air pollution increases, traffic jams increase, and crime rates increase. The changes also affect urban governance, and it is becoming increasingly difficult to plan resources and optimize costs. Previously, due to the complex and dynamic structure, it seemed that this was an unsolvable problem, but the widespread digitalization and the emergence of a mass of technologies that allow you to collect up-to-date data on the life of the city, make it possible to improve people's lives. Even though this data is of great value, it is useless without proper processing, so there are more and more systems that deal with this.[1]

## II. SMART CITY DATA

First of all, let's talk about what is smart city data and how are they generating. Urban conditions are extremely extensive and dynamic, and data is produced from completely different sources in different formats, example you can see in Fig.1.



Fig. 1. Smart City Data generating[2]

**Traffic Data:** Data on vehicles and road conditions. This data is collected using various sensors installed in mobile phones, navigators, vehicles, as well as sensors and cameras installed on the roads.

**Mobile Phone Data:** Modern smartphones contain many different sensors that generate a lot of useful information. For example: GPS, telephone statistics, etc.

**Geographic Data:** Geographical data is more static than previous types. They allow you to find out the location of organizations, buildings, and roads.

**Commuting Data:** Such data is generated when a person leaves some trace in their daily life, for example pays for parking.

**Environmental Data:** This type of data is collected using sensors throughout the city or satellites. This can be used to assess the state of the air, the level of pollution, and track climate change.

**Social network Data:** Allows you to get statistical information about the activity of people on the network. Good

example of the representation of such data is the social graph, like in Fig.2.

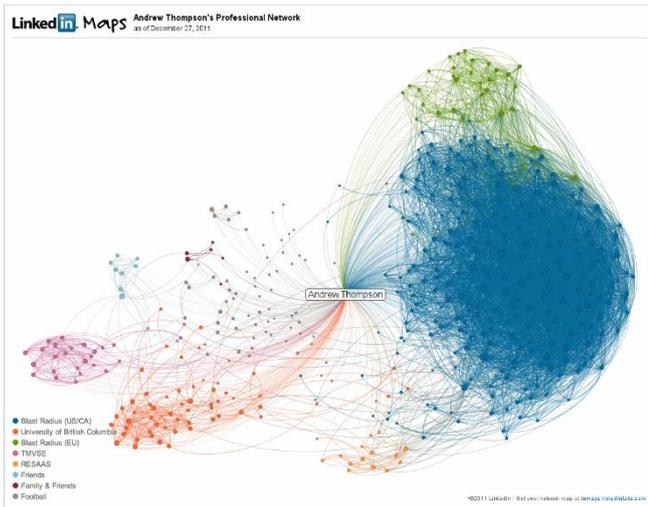


Figure 2: LinkedIn Social Graph[3]

**Economic Data:** This type of data represents any things related to money. For example, prices for goods and services.

**Energy Data:** Data on resource consumption. They are generated by various sensors in homes, gas stations, and factories.

**Health Data:** This is data produced by medical institutions and various wearable devices (for example, a fitness bracelet). This helps you track statistics about the health status of the population.

### III. OVERVIEW OF DATA EVALUATION METHODS

Evaluation can be divided into three main approaches:

1. Pre-processing –the preparation of data to get rid of incorrect data and make it the most suitable for further processing.
2. Assessment – companies conduct data assessment and determine how useful they are for different areas of research
3. Analysis - the processing of data for a specific task. There are a lot of different approaches and tools that are suitable for different purposes. One of the most popular ways to process data is neural networks.

#### A. Pre-processing.

Data sets collected from various sources is very rarely suitable for use in data processing systems. Initially, they usually have many factors due to which they are not applicable in such algorithms. They may contain some incorrect data, be incomplete, or have different formats and structures. To solve these problems, there are preprocessing algorithms. [4-5]

Preprocessing involves several different tasks, that you can see in Fig.3.

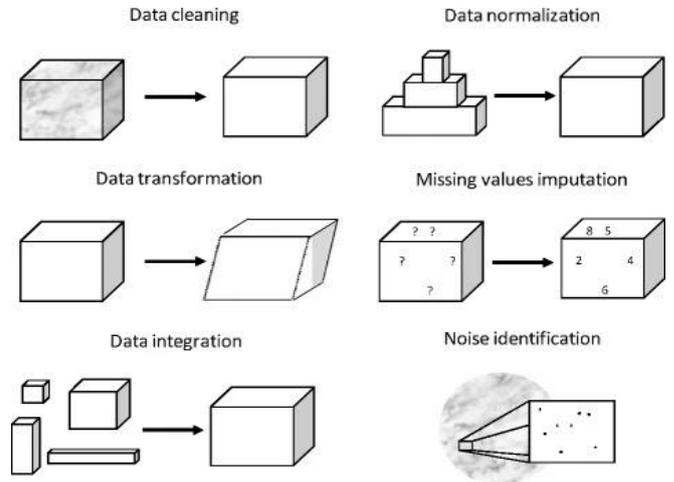


Fig. 3. Pre-processing tasks[6]

**Data cleaning:** This is when we have a set of data that has any defects. For example, extra or incorrect data.

**Data normalization:** For example, we have clear dataset, that has invalid structure that is not suitable for the target processing algorithm.

**Data transformation:** This is when we have data in one format, but the target processing algorithm used another one. For example, we have data in JSON but need XML, and we have to convert it.

**Missing values imputation:** There are missing values and algorithm have to restore them before use.

**Data integration.** it's, for example, we have data from difference sources and have to merge it.

**Noise identification:** In this case, there are noise data in dataset, that can make negative impact. So before use we have to clean it.

#### B. Assessment.

As described earlier, a smart city collects a huge amount of data from various sources [7-8]. They have a completely different format and structure and are applicable in different areas of the city's life. To structure this data, understand which areas of the city are not sufficiently covered, and optimize data collection, large companies conduct a smart city data assessment. Assessment systems are based on the index systems they develop. Index systems are compiled by dividing all areas of activity into smaller parts Fig.4 and then analyzing them.

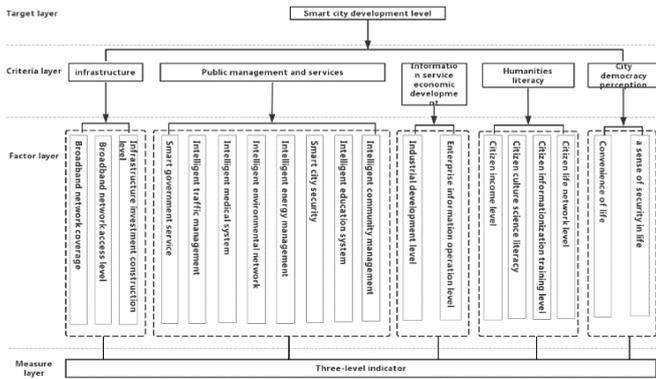


Fig. 4. Example of evaluation factors [9]

C. Analysis.

In order for the data to become valuable for any task, it must be processed. Modern data processing approaches are divided into 2 main types: Machine learning and Data mining.

1) *Machine Learning*: Machine learning is a method approach in which a network is taught to anticipate subsequent data based on available data. This is based on a model that is trained according to a certain given algorithm.[10] Fig.5 shows tasks of machine learning. Machine learning is divided into:

Learning with a teacher. The training of the model is based on the construction of an algorithm that relies on the dependence of the output data on the input data. In this case, the quality of the model is determined by checking for unsuccessful recognition.

Learning without a teacher. In this case, the ideal output data is not set and the training is based on finding dependencies between input objects.

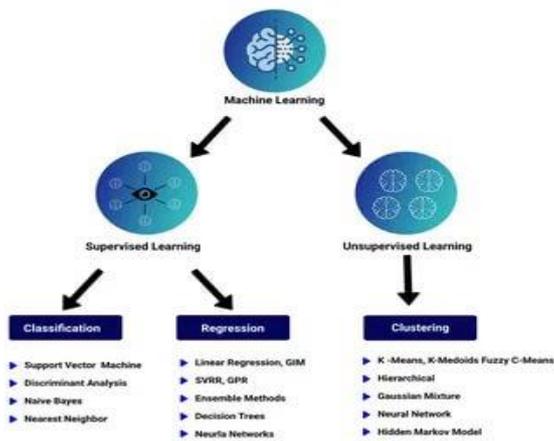


Fig. 5. Machine Learning tasks[11]

2) *Data Mining*: The main idea of Data Mining is determining the most accurate dependencies from an extensive data set. This allows you to find new, non-trivial relationships between different objects and use them for subsequent data processing in new areas. Data mining combines various mathematical functions, modern methods and approaches of data processing.[12] Fig.6 shows basic Data Mining structure.

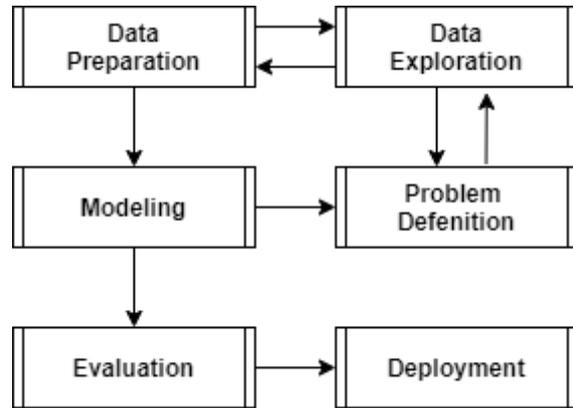


Fig. 6. Data mining structure [13]

D. Applied method

When developing the system, data analysis using panoptic segmentation was used for created models.

The main mechanism of panoptic segmentation is the assignment of a unique value to each pixel in the image, followed by the encoding of a semantic label and an object identifier. This method allows for high-quality detection of various objects in the image

IV. AUTOMATED EVALUATION SYSTEM

A. Shared system.

The proposed automated data processing system is part of a large collective effort. The main task of the overall system is pipeline for smart city data collection, processing and visualization. You can see shared system diagram in Fig.7.

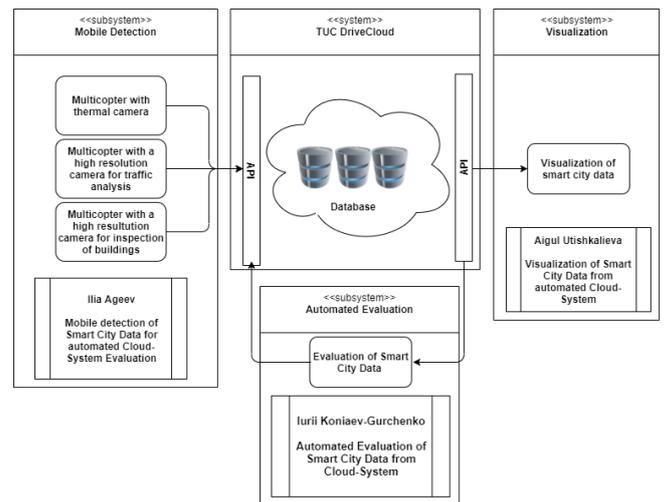


Fig. 7. Diagram of the shared system

B. Automated evaluation part

There are images obtained from multicopters that are stored in DriveCloud. These images should be processed automatically, depending on the selected task, and the processed data should be saved in DriveCloud.

There are 3 main components in the system:

- Data synchronization component with DriveCloud

- Data evaluation component
- Web interface for testing models, starting automated evaluating process, and changing the application settings

The data evaluates using trained models. The user chooses the model that will be used in automated processing. At the moment, 3 trained models have been added to the system:

- heat leakages in hot water pipes detection
- cars detection
- people detection.

But the system is designed in such a way that the user can easily add new models to detect any objects, and then use these models for automated evaluation. This makes the system more useful and easily scalable. Fig.8 shows automated evaluation system diagram.

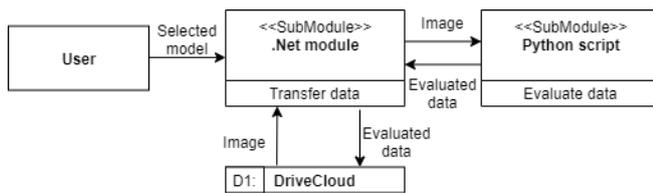


Fig. 8. Automated evaluation data diagram

### C. Tools

The web interface and the part of the application responsible for receiving and storing data, developed using C# on the platform .Net 5.

As described above, the application contains 3 trained models by default. These models were created using the Detectron2 tool in Python on Google Colab. In Detectron2 training was used COCO Panoptic Segmentation with Panoptic-DeepLab (DSCnv) method and ImageNet pretraining.

### D. DriveCloud implementation

First of all, we have to create: Vehicle with vin “AEVin”, Sensor “[AutomatedEvaluation]Image” Type:Image, Sensor “[AutomatedEvaluation]Count” Type:Int.

When automated processing is started, authorization data is sent to the DriveCloud API. Then system checks requirement setup described above.

After checking long polling starts and every minute checks for unevaluated drives. This is done using labels. If a drive has a label corresponding to the pattern AutomatedEvaluation:{ModelName}, it means that this drive prepared for automated evaluating by a model named ModelName.

Next, if not evaluated drive is found, we start a new drive AutomatedEvaluation:{ModelName}\_{driveName}, where the evaluated image will be saved, as well as the number of objects found in this image. Then we evaluated each image and save evaluated data for all of them.

Also important note, we store the processed data with the same timestamp as the raw data contain. This allows us to further compare the processed data with the original data.

### E. System requirements

The technical support of the system should use the technical means as much as possible and in the most effective way.

Requirements for the system:

- DriveCloud server
- Stable ethernet connection
- Web browser

Requirements for the DriveCloud:

- Vehicle with vin: AEvIn
- Sensor:[AutomatedEvaluation]Image Type:Image
- Sensor: [AutomatedEvaluation]Count Type:Int

Minimum requirements for the workstation:

- Processor - Intel or AMD 1500 MHz or higher;
- RAM – 2048 MB or more;
- hard disk space – 3 GB;
- ethernet network card;
- a video adapter with a resolution of 800\*600 or more.

### CONCLUSION

As a result of the work, the data types of the smart city were considered, modern approaches and methodologies used in data evaluation were analyzed. For each of the methodologies, the main goals and objectives are considered. The developed automated data evaluation system is considered. The approaches, architectural solutions and tools used in the development of this system are described.

The models used in the system were tested by the iterative method, accuracy of models:

- heat leakages detection: ~70%
- cars detection: ~95%
- people detection: ~95%

The next possible step in the development of this system is adding the ability to process data of new types and formats.

### REFERENCES

- [1] Al-Nuaimi, E., Al-Neyadi, H., Al-Jaroodi, J., Mohamed, N., "Applications of big data to smart cities". *J Internet Serv Appl* 6(1) p. 1–15, 2015.
- [2] "A set of innovative applications and services around road traffic information". Accessed: May 29, 2021. Available online at <https://www.aplana.com/case-studies/inrix/>
- [3] "Andrew thompson". Accessed: May 29, 2021. Available online at <https://www.andrewthompson.co/2012/01/>

- [4] Pyle, D., "Data preparation for data mining". *San Francisco: Morgan Kaufmann Publishers Inc. (1999)*
- [5] S García, J Luengo, F.H., "Data preprocessing in data mining". *Berlin: Springer(2015)*
- [6] S García, S Ramírez-Gallego, J.L., "Big data preprocessing: methods and prospects". *Big Data Analytics (2016)*
- [7] Yuan, N., Zheng, Y., Zhang, L., Xie, X., "T-finder: a recommender system for finding passengers and vacant taxis". *IEEE Trans Knowl Data Eng 25(10) p.2390–2403 (2013)*
- [8] Zheng, Y., Capra, L., Wolfson, O., Yang, H., "Urban computing: concepts, methodologies, and applications". *ACM Trans Intell Syst Technol 5(3) p. 38:1–38:55 (2014)*
- [9] Wang, C., Li, S., Cheng, T. et al, "A construction of smart city evaluation system based on cloud computing platform". *Evol. Intel. 13, 119–129 (2020). Accessed: May 29, 2021. Available online at <https://doi.org/10.1007/s12065-019-00259-w>*
- [10] Bergelt, René, and Norbert Englisch. "Towards Cloud-supported Automotive Software Development and Test", IN *Journal of Embedded Selforganising Systems (ESS)*, 2020, Vol.2, Iss.2, pp.8-12, doi:10.14464/ess.v7i2.440
- [11] "What is Machine Learning". *Accessed: May 29, 2021. Available online at <https://www.indiumsoftware.com/blog/what-is-machine-learning/>*
- [12] Englisch, Norbert, René Bergelt, and Wolfram Hardt. "An Educational Platform for Automotive Software Development and Test." In *2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE&T)*, pp. 1-4. *IEEE, 2020.*
- [13] Fasihuddin, M., Syed, M.F.U.H., "Machine learning technique for smart city development-focus on smart mobility" pp. 169–176 (2020)