Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 61 (2022) 462–466

Transportation Research Procedia

www.elsevier.com/locate/procedia

XII International Conference on Transport Infrastructure: Territory Development and Sustainability

Big Data and Data Mining Technologies Application at Road Transport Logistics

Pavel Mitroshin[a,]*, Yulia Shitova[b], Yury Shitov[c], Dmitry Vlasov[d], Anton Mitroshin[b]

[a]*Dmitrov Institute of Continuing Education branch of Dubna State University, 23, mikrorayon DZFS, Dmitrov, Russia*
[b] *Russian State University for the Humanities, 6, Miusskaya Sq., Moscow, Russia*
[c] *Joint Institute for Nuclear Research, 6, Jolliot Curie St., Dubna, Russia*
[d] *Dubna State University, 19, Universitetskaya St., Dubna, Russia*

Abstract

The article deals with the application of modern Big Data and Data Mining technologies in logistics. The analysis of studies related to Big Data and Data Mining in logistics has been carried out. The problems of collecting, analyzing and interpreting data on the workload of the transport system of the Moscow region are considered. The paper presents prospects for the application of the results for modern transport companies and government agencies in the planning of transport infrastructure construction.

Keywords: dig data; data mining; road transport logistics.

1. Introduction

The economic development of the country is the most important task in the management of the state. Creation of favorable conditions for running and developing business allows attracting additional investments in a market economy. According to The Global Competitiveness Report 2019, the Russian Federation is ranked 43rd based on the Global Competitiveness Index 4.0 (GCI). Looking at the metrics that affect the overall ranking of a country Road connectivity and Quality of road infrastructure one of it. Russia ranks 41 - in the road connectivity and 99th - in quality of road infrastructure. Thus, it can be argued that the road transport infrastructure is the most important factor in the development of the economic component. In logistics the development of the road transport system and its congestion directly affect for costs.

In Russia there is an intensive development of the road transport system but transportation development should be implemented not as a set of separate transport corridors, but as a single transport and logistics network (Gnezdova et al, 2017). In accordance with the quality of road infrastructure rank formed in The Global Competitiveness Report within the World Economic Forum (Fig. 1), there is an annual improvement in this indicator in Russia. Despite the development of the road transport system and, as a consequence, the predicted improvement in the transport system, approaches and methods are needed to analyze the current situation.

* Corresponding author. Tel.: +7-916-513-91-01.
E-mail address: mitpa@yandex.ru

2352-1465
10.1016/j.trpro.2022.01.075

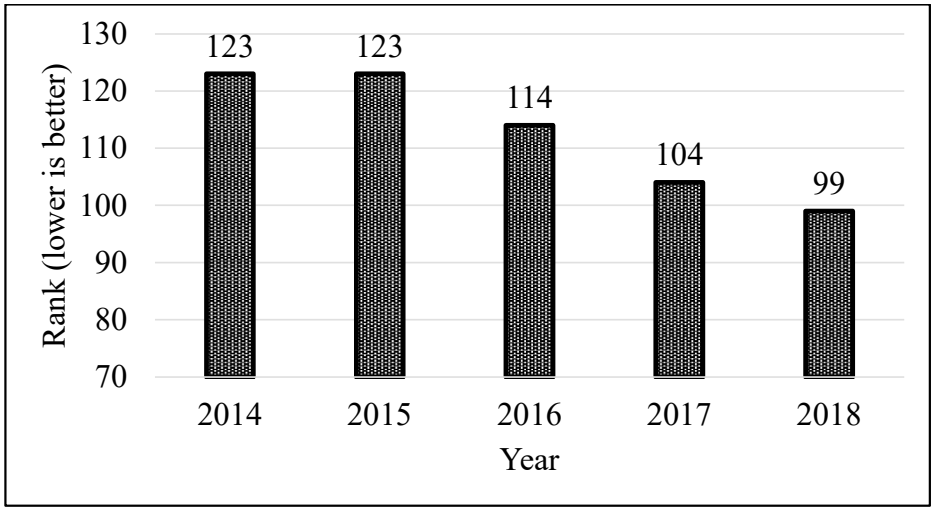

Fig. 1. Russia quality of road infrastructure rank by The Global Competitiveness Report (World Economy Forum).

The introduction of modern information systems and technologies into business processes, as well as the constant development of information processing tools, makes it possible to talk about the presence of large amounts of data that can be used for effective development. Due to the emergence of large amounts of data in various fields of human activity that can be collected and processed, we can talk about the rapid development of Bid Data and Data Mining technologies.

To ensure a competitive advantage it is necessary to develop and introduce new advanced knowledge into economic processes related to the management of transport. Considering the issues of processing and generating big data we can talk about the following areas in road transport logistics:

- analysis of statistical data on the loading of transport hubs;
- analysis and assessment of time losses as a result of erroneous supply chains;
- research of methods for forecasting the congestion of transport systems;
- formation of methods and algorithms for assessing the physiological state of drivers with predicting behavioral aspects in logistic distribution;
- analysis and prediction of the influence of weather factors on route sections;
- research of emergency sections of transport infrastructure based on statistics of road traffic accidents;
- analysis and research of advanced technologies for the distribution and movement of goods in storehouse.

The issue of saving resources of a logistics organization based on data collection and analysis is always very relevant. Currently, there are methods for obtaining transport network loading and predicting best routes to reduce delivery costs. The problems associated with Data Mining analytics are considered in work (Paul et al, 2011), where the issue of minimize logistics cost is solved, but the model does not take into account the issue of loading of transport nodes. The problems of modeling and analysis of logistical state data based on Big Data and Data Mining technologies are discussed in article (Brandau and Tolujevs, 2013), where an efficient and effective supply chain with a continuous and uninterrupted flow of goods is required for perfect management.

An important aspect of data processing is collection and formation. As part of the research carried out with the support of the grant RFBR 19-010-00794, a software and hardware complex was developed to collect data about the transport system loading of the Moscow region.

2. Methods for collecting data of the transport infrastructure loading for the Moscow region

The paper (Zaitsev and Danilenko, 2018) considers the problem of monitoring and, as a consequence, the collection of statistical data of the road transport system. In general, the following principles of data collection are highlighted:

- getting traffic statistics in a structured format from city road transport services;

- calculation of the throughput of the city transport network by processing video recordings from traffic cameras installed at crossroads;
- installation of additional equipment at streets or the use of other special means.

Presented above methods have proven themselves well at the beginning of the 21st century and as a solution to individual problems, for example, monitoring individual crossroads. Scaling the problem of monitoring the transport system to a region or country, such data collection approaches have a number of limitations:

- problems of efficiency for receipt of information;
- limited data on time intervals;
- coverage limits;
- low scalability;
- high price.

Considering the options for collecting statistical it possible to make conclusion data geographic information systems and services play a special role in modern society. The result of the introduction of GIS into society was the widespread use of navigation systems with the display of road loading. GIS services automatically collect and process information received from the users themselves on the intensity of traffic along a certain route. Examples of such services are Yandex.Maps and Google.Maps (Yandex Map Platform, 2021; Yandex Traffic Platform, 2021; Google Map Platform, 2021). A distinctive feature of the services presented is an open API that allows to contact the service for information. An example of information received from services is the length of the route between two points and the time of the trip. Using of map and navigation services to assess for loading of transport infrastructure is an effective source of large data. Analysis and processing of such data can be used to solve logistic problems of the region and reduce transportation costs in logistics.

As part of the research in the Moscow region (Mitroshin et al, 2021; Shitova et al, 2021), 20,000 points of departure were evenly distributed to build routes. The final destination of the route is the center of Moscow. On the basis of the data obtained the problem of time losses during the passage of the route is investigated. The collection of statistics has been carried out since mid-2015. Several times per hour, data are collected on previously defined routes based on requests to the navigation GIS service. The accumulated data make it possible to make conclusions regarding the loading of the transport system of the Moscow region.

3. Analysis of the received data

An algorithm and methodology that allows collecting and analyzing information about traffic congestion was developed and tested in the course of earlier research (RFBR grants No. 11-06-00323 and No. 14-06-00249). Taking into account of traffic's big data, a complex of software solutions was developed that allows collection and analysis in a semi-automatic mode. The data required for further research characterizing the workload of the transport system are collecting in the following form:

- in accordance with the uniform distribution of the starting points over the study area, a unique number is determined;
- for further analysis, in accordance with the time reference, information about the date of data receipt is generated in mysql format YYYY-MM-DD HH: MM: SS;
- in accordance with the constructed route on the house-work section, the length in meters is saved for further analysis;
- information about the travel time by car is saved, taking into account traffic jams, based on data from the Yandex.Maps service.

A variable generated for each route in a single copy as a constant:

- travel time by car without traffic jams, based on data from the Yandex.Maps service).

The collected data allows to make some conclusions regarding the loading and time losses on roads within the day and week. Considering the intraday time losses it can be argued that the congestion of the transport infrastructure sharply increases between 5-9 am and 16-18 pm in the direction of the Moscow region to Moscow. The opposite direction is characterized by a sharp increase in time losses due to road congestion in the period from 2 pm to 8 pm and from 5 am there is a constant increase of roads loading until 6 pm (Fig. 2).

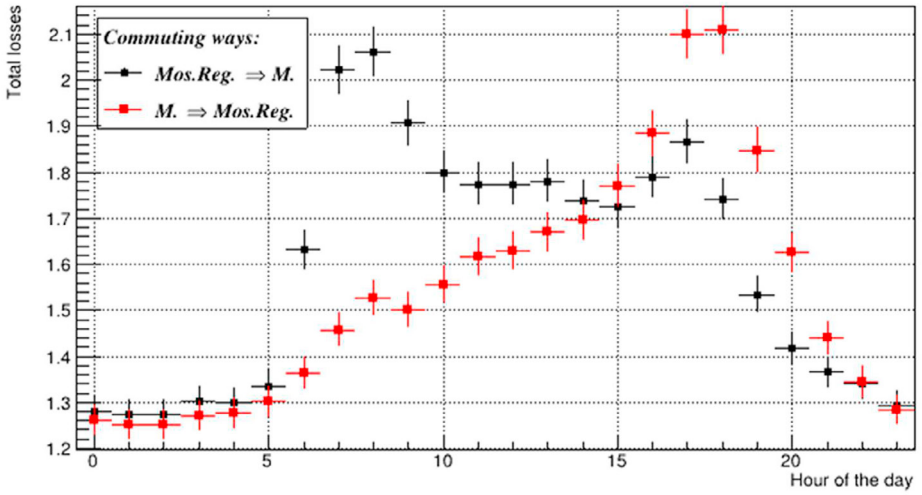

Fig. 2. Intraday time losses depending on the time of day (routes between 20 km from Moscow). Black dots for direction from Moscow region to Moscow. Red dots for direction from Moscow to Moscow region.

Considering the intra-week distribution for the Moscow region it is reasonable to conclude that the busiest traffic days are Tuesday and Thursday. The largest number of time losses in comparison also falls on Tuesday and Thursday (Fig. 3).

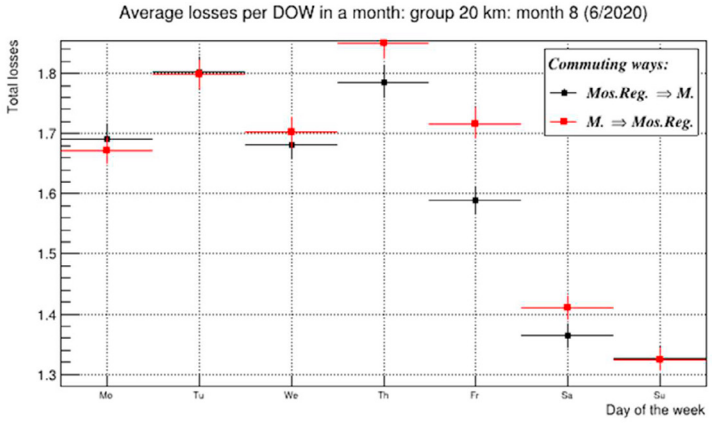

Fig. 3. Distribution of time losses within a week.

4. Summary

The presented method of collecting information and the results obtained allow to make conclusion of using of such data processing technologies in logistics processes. An important problem in logistics is time losses, based on current statistical data and predictive models, conclusions can be drawn about traffic planning. By choosing time intervals for road movement with the least load by statistic it possible saves significant time resources. By integrating data on weather conditions and road traffic accidents into the presented model it is possible to form a predictive model for the road transport system. The presented model makes it possible to estimate the congestion by directions which can be used as a tool in planning the reconstruction and development of transport infrastructure.

References

Brandau, A., Tolujevs, J., 2013. Modelling and Analysis of Logistical State Data. Transport and Telecommunication Journal. 14(2), 102-115.

Gnezdova, Y. V., Glekova, V. V., Adamov, N. A., Bryntsev, A. N., Kozenkova, T. A., 2017. Development prospects of the transport infrastructure of Russia in the conditions of development of the market of transport and logistic services. European Research Studies Journal. 20 (4A), 619-631.

Google Map Platform https://www.google.ru/maps (accessed 20 September 2021).

Mitroshin, P. A., Shitova, Yu. Yu., Shitov, Yu. A., Mitroshin, A. A., Vlasov, D. N., 2021. GIS-monitoring of Regional Transport Network Traffic as a Method to Study Commuting: Moscow Region Case. Journal of Physics: Conference Series. 1828(1), 012073. http://dx.doi.org/10.1088/1742-6596/1828/1/012073.

Paul, A., Saravanan, V., Thangaiah, P. R., 2011. Data mining analytics to minimize logistics cost. International Journal of Advances in Science and Technology. 2(3), 89-107.

Shitova., Yu., Shitov, Yu., Vlasov, D., Mitroshin, A., Mitroshin, P., 2021. Gis-Monitoring Of Regional Transport Network. 916-926. http://dx.doi.org/10.15405/epsbs.2021.04.97.

Yandex Map Platform https://yandex.ru/maps (accessed 20 September 2021).

Yandex Traffic Platform http://maps.yandex.ru/traffic (accessed 20 September 2021).

Zaitsev, I. Yu., Danilenko, I. N., 2018. Development of mobile application for the urban-traffic statistics gathering. Proceedings in Cybernetics. 4, 170-176.