

Optimization of urban freight transportation based on evolutionary modelling

E A Gladchenko¹, O N Saprykin¹ and A N Tikhonov¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

e-mail: liz-ok_033@mail.ru, saprykinon@ssau.ru

Abstract. Logistics problems require special attention, because every year they become more complicated and multivariable. On the one hand, a supply chain management includes incessant monitoring of such issues as requests elaboration, paths determination, routing of shipments, multimodal choice, set up of transshipments, fleet choice and maintenance, warehousing, packaging and others. On the other hand, dozens of people are involved in the logistics process. All these moments complicate the decision-making that is why data driven decisions are required nowadays. As well as shipment problems are NP-hard, the heuristic methods should be applied to resolve them. In this article we propose a genetic algorithm to solve the complex problem that consists of the Travelling Salesman Problem combined with the Knapsack Problem. We have developed an urban freight transportation model which is focused on the minimization of the underway time as well as on the maximization of the truck's loading. A significant contribution in our method is the census of traffic frequency by using traffic zoning. The developed approach has been implemented using the Python programming language in the Zeppelin environment. The first version of the system has been approved in the city of Samara (Russia) with test demand dataset.

1. Introduction

Nowadays, the competition between logistics companies is extremely high. To stay alive each of them competes for clients, strives to reduce costs, improve quality and efficiency of work, in other words they try to optimize the process. In the 21st century, such technological advances as Big Data or Internet of Things [1], are available to be used for the above listed purposes. The potential of Big Data techniques with state-of-the-art software can significantly enhance the data-driven decision-making process.

Logistics optimization is a complex and dynamic problem that requires a systematic approach. On the one hand, this is determined by the large membership involved in the process: shippers, carriers, recipients, third party logistic managers and others. Their decisions and actions should be well coordinated, that is why it is necessary to have a good communication. On the other hand, the supply chain consists of several stages. The first is a distribution that includes tracking the demand level and types of commodities, booking and consolidation of requests. Transportation operations take place at the second stage which includes layout of the route, fleet choice, schedule preparing and adjustment, organization of the multimodal transshipment. This is the most sensitive and vast part of the logistics process, because of the wide variety of stochastic issues that need to be predicted. Warehousing, with

the best depots locations, as well as using the interior automation equipment and mechanic means, is the third stage. The last one is packaging, which is especially considerable in perishable, fragile and out of gauge goods shipments.

Outlined stages are not independent and affect each other. Thus, routing is NP-hard optimization problem that can be formally solved only in limited cases and has low performance. To avoid these limitations, heuristic methods are commonly applied. This paper is focused on the genetic analysis in terms of integrated transportation optimization.

The rest of the paper is organized as follows. Section 2 gives a brief background to the recent research, both theoretic and empirical; and problem statement. Section 3 contains basic model's features, presents methodological description, including genetic analysis. In Section 4 the implementation is displayed. Section 5 concludes and summarizes the paper.

2. Background

2.1. Literature Review

There is a huge amount of research related to supply chain optimization. For example, M. Grazia Speranza [1] outlines the main tendencies in the logistics management, one of which is the complex data handling. The author points out three directions of the extension: systemic direction (the most applicable for this research), collaborative direction and dynamic direction. The paper presents some feasible ways of using Big Data in terms of transportation process optimization. Big Data technologies make possible to build predictive models of transportation processes too [2]. Generally, articles which describe different methodologies, rested on the model formulization. For instance, Shui Lam and Burkhard Englert [3] represent a basic mathematical model, which takes into account only two shipment offers. The proposed method compares shipment schedules, transportation cost and terminal charges that helps to find out optimal deliveries.

Nevertheless, heuristic analysis allows estimating multiple decisions applying Genetic Algorithms [4]. Zbigniew Michalewicz [5] describes the processes of evolving with the parents' selection and the producing of offspring. This study contains several modes of application Mutation and Crossover operators, as well as the representation in matrix encoded data. For example, genetic analysis, with the unique combination of operators, was used by Gintaras Vaira [6] to solve the Rich Vehicle Routing Problem (RVRP). The general VRP is different from the typical VPR, because it restricts extra constraints for a task, a vehicle, a cargo and so on. The genetic algorithm allows resolving VRP with pick-up and deliveries as well as VRP with time windows.

As we have already noted the supply chain optimization is a complex subject that is why solving only the routing problem apart from real life constraints is not efficient. The next research tries to take this fact into account. The aim of the next model is not to create new optimal routes, but to find the best combination of existing routes and services in maritime transportation. Anantaram Balakrishnan and Christian Vad Karsten [7] report the LP-based heuristic procedure for containerized cargo that may require the transshipment during the delivery. The approach is based on a selection from the limited number of possible transshipments using the augmented multi-commodity flow network.

Another complicated problem for freight operators is to control the booking rate to provide the maximum profit and execute all shipments when the product demand and resource capacities are not determined. The stochastic resource allocation problem which is formulated by Xinchang Wang [8] as a stochastic programming (STOC) model that incorporates three important features: random resource capacities, stochastic product demands, and network effects. The algorithm embeds a nonlinear optimization solver that uses an iterative-search type algorithm.

Furthermore, there is a strong relationship between the shipment size and fleet size. Lots of papers propose various models to find the optimal mapping function. One of them is based on the discrete-continuous econometric model where the shipment size is a continuous variable and the size of the vehicle fleet is a discrete variable [9]. Results that are represented in the paper prove a hypothesis: with the increase of the distance and total demand it is become more efficient to exploit heavier, except for old vehicles.

Latent Class Analysis (LCA) is also applicable to shipment size optimization. Raphael Piendl, Gernot Liedtke and Tilman Matteis [10] combine LCA and minimization of the total logistics costs. The proposed model divides transport cases into classes according to similar response patterns. Four homogeneous segments were identified which were integrated as an additional component to the systematic framework of the total logistics cost formulation. Concepción Román, Ana Isabel Arencibia, María Feo-Valero [11] represent fascinating research, that outlines the quantification of attribute cut-offs combined with the estimation of taste heterogeneity through the Latent Class Model (LCM). The analysis is based on the empirical data about decision-making in the area of logistics management and based on the specification of the penalized utility.

Another application of the heuristic algorithm was introduced by Wu-Yang Liu and Li-Ning Xing [12] to reduce costs in the processing of express logistics system as a type of capacitated arc routing problem (CARP). The first step in the model is performing architecture optimization. The second step is assuming extended random path-scanning (drawing between 5 Rules). The paper illustrates the dependence between the number of depots, vehicles, total cost and different characteristics, as vehicles load, speed, acquisition cost, depot construction post, maximum service time, etc.

Nevertheless, when the optimal number of depots is identified it can be not conformed to the best way between stores and final destinations, their location should be taken into account in this case. Xiang Hua, Xiao Hu and Wuwei Yuan [13] present a mean to solve this problem by applying the Adaptive particle swarm optimization (APSO). The APSO is a supplement of nonlinear inertia weight as well as time-varying acceleration coefficients as opposite to the Particle swarm optimization (PSO), which is a common method of decision-making. This fact allows a quick and accurate finding of the global optimum. As a result, the best locations may be determined using the knowledge of customer's locations and demand. However, logistics managers more often deal with existing depots, so the real problem is to create optimal paths under existing circumstances.

To achieve the maximum effect, the logistic problem should be solved as a complex task, involving different aspects of the transportation process. One of the papers that consider this conception is a review written by Pieter Vansteenwegen, Wouter Souffriau and Dirk Van Oudheusden [14]. The authors outline different combinations of logics problems, which are put together into the Orienteering Problem (OP). They investigate in more detail three of them (Team OP, OP with time windows, Team OP with time windows), give several examples of their mathematical formalisations and ways of solving. Another example of solving the complex logistic task is the study of Mohamed El Yafrani and Belaïd Ahiod [15]. They look into the Travelling Thief Problem (TTP) which includes two interdependence combinatorial optimization subproblems: the Travelling Salesman Problem (TSP) and the Knapsack Problem (KP). The authors compare iterative neighborhood algorithms to other methods, present deterministic procedures for each problem combination reinforce their assumptions by experimental results.

However, recent research in the considered area does not take into account city traffic flows. This constraint is very vital for modern cities with high motorization level. Our study extends the described above conception, and focused on the routing of the city freight trucking in urban traffic conditions. To solve this problem the transportation planning techniques were involved in a model and combined with an evolutionary approach.

2.2. Problem statement

The study addresses the following problem: identify the optimal route from the distribution center to destination points with the fleet's capacity constraint and taking into account urban traffic. The problem differs from a typical TSP, because the route is based on the road traffic. This supplement promises to minimize the delivery time and consequently the total logistics costs. In addition, the described problem requires combining the TSP with KP.

To solve the described above problem we propose the evolutionary approach that is based on a modified genetic algorithm. This approach allows reflecting the latent organizing process of a complex technical system that is close to nature system [16]. The input data for the algorithm are city transport infrastructure and transportation demand. The idea is that we have a city where vehicle density is varying in different time-windows and districts. It means that the underway time of one route during

various times of the day will be different. This information can be provided by modelling system that is based on the origin-destination matrices.

Therefore, the algorithm takes into account origin and destination locations, schedule time, transportation demand, fleet capabilities. Moreover, to find the optimal loading the KP is solving on every iteration of TSP algorithm. The KP algorithm is based on the genetic approach too. The overall solution represents the combination of two genetic algorithms: the KP genetic algorithm is nested into the TSP genetic algorithm.

3. Methodology

3.1. Mathematical Model

The model describes the city freighting planning for helping logistics managers in decision-making, i.e. to optimize the whole route (TSP), with maximum trucks loading (KP). A solution consists of determining the final routes of delivery with the sequence of destination points which start and end in the depots. The goal is to minimize the underway time and to maximize the truck's loading.

To get an accurate and realistic solution the impact of the urban vehicle density should be taken into account. However, traffic datasets for most of the cities are unobtainable. For this reason, we use Transport Zoning and macroscopic transport simulation to make our model more flexible.

The process of transportation simulation of the city is based on the gravity law. Marc Barthelemy [17] presents detailed information about creating origin-destination matrices and processing mobility networks. Simulation allows estimating the flow value between different areas in the city by using population and infrastructure datasets. Figure 1 shows a visual representation of the model that takes into account traffic zones. The micromodel that is based on this traffic zones allocation shows different transportation velocity on different branches of the road network. We propose the space-time model because the velocity depends on time-windows too. Therefore, we can predict the average velocity of trucks during transportation.

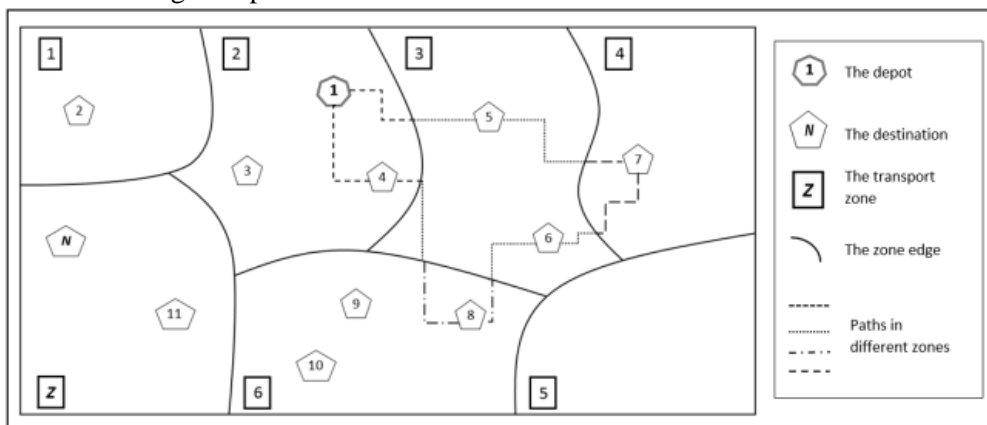


Figure 1. Example of the delivery route that starts and ends in the depot (1-5-7-6-8-4-1).

In our paper we are using the following designation indicators:

- N is the number of destination points.
- i is the origin point, ($i=1\dots N$).
- j is the destination point, ($j=1\dots N$).
- d_j is the quantity demand in the point j , ($j=2\dots N$).
- l_{ij} is the distance between i and j , ($i=1\dots N$), ($j=1\dots N$).
- $l_{ij} = 0$ if $i = j$, which means there is no distance between the same origin-destination points.
- ts is the average service time that is needed for unloading one package, is a set value, e.g. $ts = 2$ min.
- te is the average extra time in the destination point, for instance drawing up of documents etc., is a set value, e.g. $te = 7$ min.

- c is the capacity of the truck.
- z is the transport zone of the city ($z=1 \dots Z$).
- $\overline{V}_z^I, \overline{V}_z^J$ is the average speed in z -th zone.
- M_k is the matrix chromosome with the whole route, ($k=1 \dots K$).
- x_{ij} is the variable for the matrix.
- P_k is the vector form of the matrix's chromosome ($k=1 \dots K$).
- p_n^P is the value (number) of the destination point, where n is an atomic number of the gene in the P -th chromosome.

The fitness function (FF) for the TSP is T_k shows the total underway time of the route, where the sequence of m is a reverse sequence of loading the k th truck in the depot. FF1 for TSP looks like:

$$T_k = \left(\sum_{j=p_1}^{p_{n-2}} \sum_{z=1}^Z \left(\frac{2 \times l_{j,j+1}}{\overline{V}_z^J + \overline{V}_z^{J+1}} + d_{j+1} \times ts + te \right) + \frac{2 \times l_{p_{n-1},p_n}}{\overline{V}_z^{p_{n-1}} + \overline{V}_z^{p_n}} \right) \rightarrow \min \quad (1)$$

The constraint is:

$$\sum_{j=p_2}^{p_{n-1}} d_j \leq c \quad (2)$$

It means that the total quantity of freight during one route cannot exceed the capacity of a truck. To relate the points to transport zones (z), we assign a zone number to each origin and destination points.

The fitness function for the Knapsack Problem (FF2):

$$\sum_{j=p_2}^{p_{n-1}} d_j \rightarrow \max \quad (3)$$

(3) can be treated as the more loading leads to a better economic effect for the final logistics costs.

3.2. Genetic Algorithm

To solve the designated in Section II B problem we adopted genetic analysis which is based on the evolutionary process. Genetic modelling is one among many heuristic ways in decision-making that uses encoded information, where variables have discrete values 0 and 1. Basic stages of the program are chromosomes generation and selection. Genotypes consist of genes with straight positions. When the first population is built, these potential solutions are tested by the fitness function, where the best ones are selected for the following iterations using mutation and/or crossover operators.

We randomly create the first K "parents" populations that are shaped into Binary Matrixes Chromosomes. Each matrix represents the encoded route, includes rows i and columns j , if $x_{ij}=1$ there is a sub-path between i and j , otherwise $x_{ij}=0$. Matrices are completed with some special conditions, such as:

$$x_{ij} \in \{0,1\}; \quad i = 1 \dots N, j = 1 \dots N, \quad (4)$$

$$x_{1j} = x_{i1} = 1; \quad i = 1 \dots N, j = 1 \dots N, \quad (5)$$

$$x_{ij} = 0; \quad \forall (i = j); \quad i = 1 \dots N, j = 1 \dots N, \quad (6)$$

$$\sum_{i=1}^N x_{ij} \leq 1; \quad \forall j = 1 \dots N, \quad (7)$$

$$\sum_{j=1}^N x_{ij} \leq 1; \quad \forall i = 1 \dots N. \quad (8)$$

For example, figure 2 displays the matrix for M_1 .

The route from figure 1 can be treated as vector $P_1: (1-5-7-6-8-4-1)$, where $p_1^P = p_n^P = 1$ in any way. It is because the delivery is always started and ended in the depot, which number is $i = j = 1$.

$i \backslash j$	1	2	3	4	5	6	7	8	...	N-1	N
1	0	0	0	0	1	0	0	0		0	0
2	0	0	0	0	0	0	0	0		0	0
3	0	0	0	0	0	0	0	0		0	0
4	1	0	0	0	0	0	0	0		0	0
5	0	0	0	0	0	0	1	0		0	0
6	0	0	0	0	0	0	0	1		0	0
7	0	0	0	0	0	1	0	0		0	0
8	0	0	0	1	0	0	0	0		0	0
...											
N-1	0	0	0	0	0	0	0	0		0	0
N	0	0	0	0	0	0	0	0		0	0

Figure 2 Example of the Binary Matrix (M_1).

After that, the FF1 (1) is calculated for all K individuals, and the algorithm checks the constraint (2). Some of the structures from the first population will be removed due to inconsistency with the capacity limitation.

After these steps, the standard algorithm may proceed to the next iterations, but our approach contains additional KP step. The rest of K vectors are estimated by FF2 (3). Figure 3 shows an example of one of the possible solutions of the KP task.

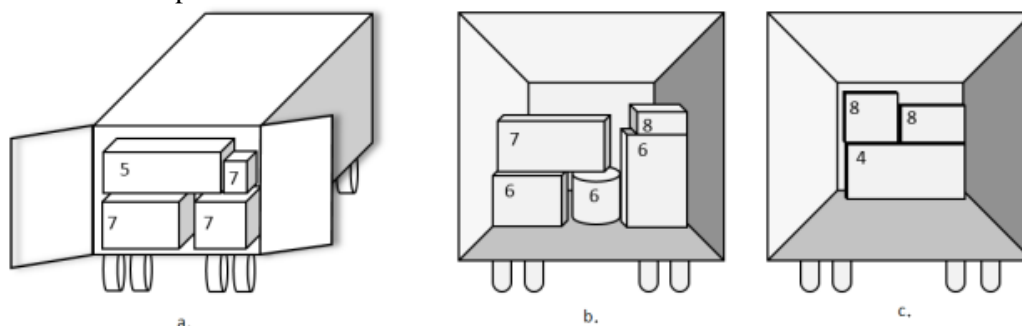


Figure 3. Example of the truck loading for the route (1-5-7-6-8-4-1): a) the third layer of loading; b) the second layer of loading; c) the first layer of loading.

Subsequently the best representatives P_k are selected for further simulations with M_k using the crossover operator. The outcome of this procedure is the new offspring, which is used for the next loop iteration. Figure 4 represents the schema of the proposed genetic algorithm.

4. Implementation

The solution is implemented as Python notebook that allows working with the model iteratively and improving it step by step. As a host environment, we are using Zeppelin, because it provides capabilities to build scalable interactive notebooks with a rich user interface. The developed software is implemented as a Docker container, which contains all configured modules and packages and can be deployed on most common operating systems. The developed scripts and notebooks are stored in the Bitbucket git repository. Figure 5 presents the architecture of the developed solution.

The Python programming language is a flexible tool for systems prototyping. There are a lot of packages for solving different applied and scientific problems. To work with spatial data in our project the following Python packages were used:

- Shapely (low level processing of geometries).
- GeoPandas (processing of data frames with geometry column).
- NetworkX (working with spatial graphs).
- OSMnx (loading network for some region from OpenStreetMap resource).
- Matplotlib (power tool for plotting with supporting of geographic maps).

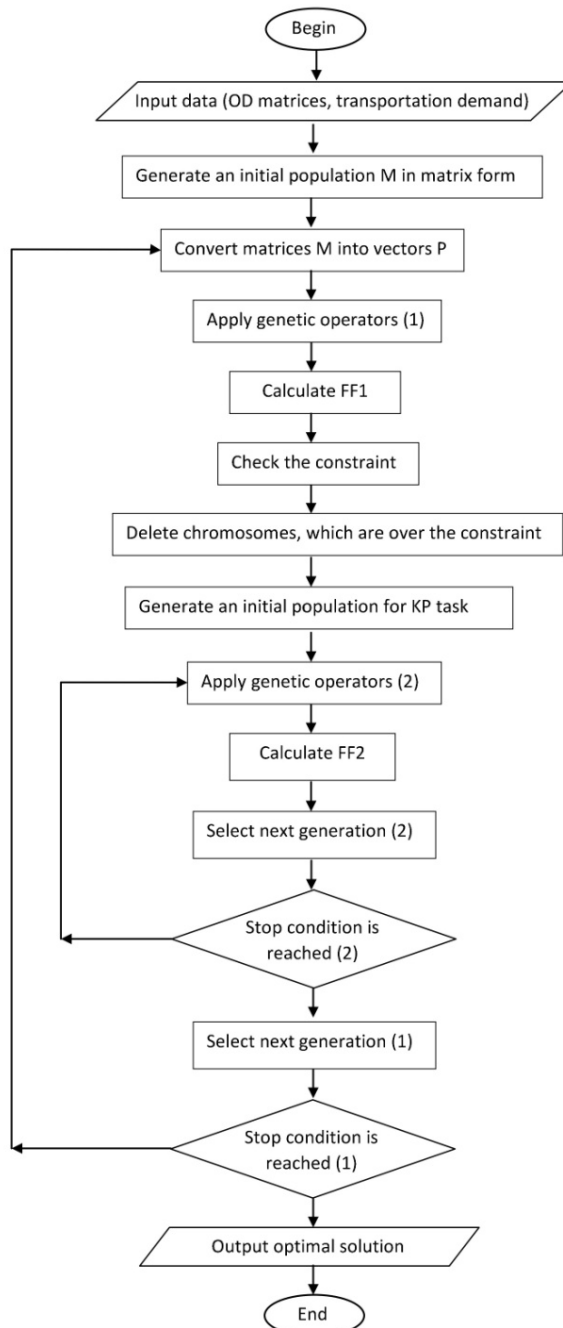


Figure 4. The schema of the heuristic algorithm.

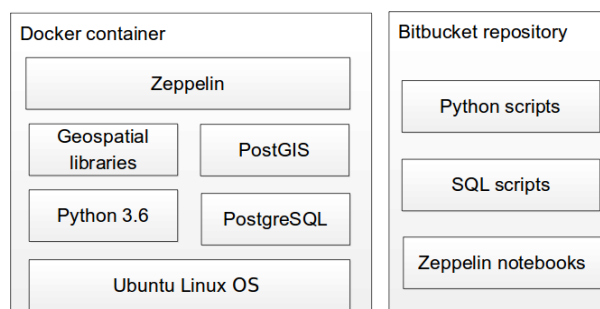


Figure 5. Schema of the solution environment.

A genetic algorithm was implemented using Deap package that contains genetic operators and means to construct any evolutionary model. Fitness function (1) was implemented as a custom function and registered in Deap framework as “evaluate” parameter. Inside the main evolution loop that solves TSP task, we have added the nested evolution loop that solves KP task for each “individual” that is selected as a population for the next generation. Therefore, the developed Python script fully implements the method proposed in Section III.

The solution was debugged using the map of the city of Samara (Russia). The information about dividing the city into traffic zones and building Origin-Destination matrices was taken from the previous study [18]. For other task parameters (origin and destination points, transportation demand, freight fleet and its capacity) we have used the test data. We have used SUMO as a simulation engine in our research (figure 6).

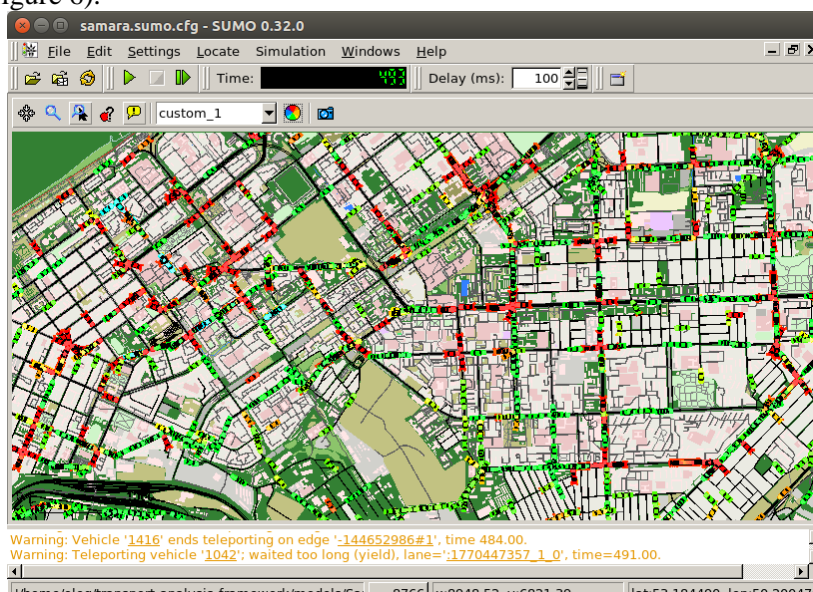


Figure 6. Simulation of transportation flows in Samara.

5. Conclusion

The genetic analysis is used as a tool for freight transportation decision making support. The urban freight transportation model allows us to determinate the optimal origin-destination route of the delivery by minimizing underway time and maximizing truck’s loading. The main advantage of the proposed approach is applying of transport zoning, which promises to bring solutions closer to the real-life situation.

The further research may take into account time windows of deliveries and fleet capability differences. Moreover, the genetic algorithm requires tuning to get optimal execution performance. Additionally, the simulation stage should be used to test the model performance within the urban traffic circumstances.

6. References

- [1] Speranza M G 2016 Trends in transportation and logistics *European Journal of Operational Research* **264** 830-836
- [2] Agafonov A A, Yumaganov A S and Myasnikov V V 2018 Big Data analysis in a geoinformatic problem of short-term traffic flow forecasting based on a k nearest neighbors method *Computer Optics* **42(6)** 1101-1111 DOI: 10.18287/2412-6179-2018-42-6-1101-1111
- [3] Lam S and Englert B 2009 Cost-Based Optimization of Shipment Scheduling *IFAC Proceedings Volumes* **42(15)** 143-148
- [4] Mokshin A V, Mokshin V V and Sharnin L M 2019 Adaptive genetic algorithms used to analyze behavior of complex system *Communications in Nonlinear Science and Numerical Simulation* **71** 174-186
- [5] Michalewicz Z 1996 *Genetic Algorithms + Data Structures = Evolution Programs* (Spring-Verlag Berlin Heidelberg, New York)
- [6] Vaira G 2014 Genetic algorithm for vehicle routing problem *Doctoral Dissertation, Technological Sciences, Informatics Engineering* (Vilnius)
- [7] Balakrishnan An and Karsten C V 2017 Container shipping service selection and cargo routing with transshipment limits *European Journal of Operational Research* **263** 652-663
- [8] Wang X 2016 Stochastic resource allocation for containerized cargo transportation networks when capacities are uncertain *Transportation Research Part E* **93** 334-357
- [9] Abate M and Jong G 2013 The optimal shipment size and truck size choice – The allocation of trucks across hauls *Transportation Research Part A* **59** 262-277
- [10] Piendl R, Liedtke G and Matteis T 2016 A logit model for shipment size choice with latent classes – Empirical findings for Germany *Transportation Research Part A* **102** 188-201
- [11] Román C, Arencibia An Is and Feo-Valero M 2016 A latent class model with attribute cut-offs to analyse modal choice for freight transport *Transportation Research Part A* **102** 212-227
- [12] Liu W-Y and Xing L-N 2013 The double layer optimization problem to express logistics systems and its heuristic algorithm *Expert Systems with Applications* **41** 237-245
- [13] Hua X, Hub X and Yuan W 2016 Research optimization on logistics distribution centerlocation based on adaptive particle swarm algorithm *Optik* **127** 8443-8450
- [14] Vansteenwegen P, Souffriau W and Oudheusden D V 2010 The orienteering problem: A survey *European Journal of Operational Research* **209** 1-10
- [15] Yafrani E M and Ahiod B 2016 A local search based approach for solving the Travelling Thief Problem: The pros and cons *Applied Soft Computing* **52** 795-804
- [16] Saprykina O and Saprykin O 2017 Transport Infrastructure Optimization Method Based on a Memetic Algorithm *Proceedings of the IEEE 20th International Conference on Intelligent Transportation Systems* (Yokohama, Japan) 1636-1641
- [17] Barthelemy M 2010 Spatial Networks *Physics Reports* **499** 1-101
- [18] Saprykina O and Saprykin O 2017 Validation of Transport Infrastructure Changes via Microscopic Simulation: A Case Study for the City of Samara, Russia *Proceedings of the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems* (Naples, Italy) 788-793