TERMINAL INFRASTRUCTURE AND CONTAINER TRAINS: OBJECT CLUSTERING

Moskvichev, Oleg V., *Samara State Transport University, Samara, Russia.*

ABSTRACT

The task is considered based on the long-term planning of development of a container transport system (CTS) as a part of a transport complex and associated with rational placement of the terminal infrastructure with respect to the areas of industrial production and consumption of container-compatible products. A methodological approach to clustering of objects of the terminal and logistics infrastructure was developed to create conditions for the mass use of container trains on the railway network. The enterprises are divided into clusters with geometric centers, where stations and container points are located.

Keywords: *cluster analysis, container transportation, container terminal, placement of transport facilities, container train, storage and distribution center.*

Background. *One of the areas related to improving the efficiency of organization of container transportation on railway transport has been introduction of new high-tech transport products.*

In recent decades, the forms of inter-industry and interregional cooperation have strengthened, the configuration of production chains has changed. This led to expansion of a range of products, reorientation of markets and territorial redistribution of production. And the changes themselves ultimately affected the loading of terminal facilities, many of which remained unclaimed or unable to consolidate existing volumes of container-compatible products into accelerated container trains [1].

There is a situation when there are attempts to implement modern promising technologies on the infrastructure built in the 40es of the last century and designed for solving completely different tasks and implementation of other technologies. The new stage of globalization and the trend towards intercountry free trade associations involve a change in the geography and structure of commodity and transport flows, which in turn requires a revision of the existing container transport system (CTS) in a structural and territorial sense.

Objective. *The objective of the author is to consider clustering of terminal infrastructure and container trains.*

Methods. *The author uses general scientific and engineering methods, modeling, evaluation approach, comparative analysis, scientific description.*

Results.

1.

To implement perspective transport technologies and increase the level of containerization, a new model for formation and functioning of the container transport system of rail transport is proposed. It is based on creation of a two-level network of terminal facilities that allow consolidation of the freight base of individual consignors / consignees into accelerated container trains. The implementation of the model will contribute to the emergence of the infrastructure of the CTS, properly balanced not only by the number of terminal facilities, but also by the place of location relative to industrial production [2].

To this end, in each region, a container storage and distribution center (CSDC), optimally located in relation to the network of container points (CP), should be created, which, in turn, should be optimized in terms of placement for customers, i.e. for specific loads.

A universal methodology for partitioning a set of objects with given properties into subsets under given criteria and obtaining «centers» of these subsets possessing optimal properties will help to make an optimal choice of locations of CP and CSDC. As such a universal procedure, it is proposed to use mathematical methods of clustering of objects, known as cluster analysis [3].

The analysis of known clustering algorithms showed that in them the cluster center is determined only by the properties of objects, and in the clustering procedure it is not possible to impose restrictions on the choice of such centers. So, for example, when using clustering algorithms using the known k-means method, the optimal center can be located at any point in the parameter space that defines objects. If the parameters are the geometric coordinates of productions, then the center can lie anywhere in the plane. In practice, one should consider a case when the center must necessarily be at one of the given points (say, at the railway station). That is, when choosing the location of CP and CSDC, it is necessary to solve the task of clustering «with projection to a function», when the center must necessarily be at the station or «with a projection to the points».

In this regard, a new clustering method with a projection to the set of points «k-means pro» is proposed and the possibility of its application in the design of transport infrastructure is being studied [4].

The input data is the set of clustering objects $X = \{x_1, \ldots, x_n\}$*, their weights* $V = \{v_1, \ldots, v_n\}$ *and admissible set of projections* $Y = \{y_1, \ldots, y_p\}$*. Each j-th object and each every admissible point-projection are given in the G-dimensional space* R^G*, or* $x_j = (x_{j1}, \ldots, x_{jG})$ *and* $y_r = (y_{r1}, \ldots, y_{rG})$*.*

The only controlling parameter is the number of clusters k into which the partition $S = \{S_1, \ldots, S_k\}$ *of the set X is carried out. As a result, we obtain an unbiased partition* $S^* = \{S_1^*, \ldots, S_k^*\}$*, the centers of which are optimal set of projections* $C^* \subseteq Y$ *.*

We introduce the notation: n – the number of clustering objects, p – the number of points of the permissible set of projections, i, i' – the cluster number, j – the object number, r – the point number of the projection set, l – the coordinate number of the point, m – the current iteration, G – dimension of the space in which clustering is performed.

The distance between points in a given G-dimensional space is found from the Euclidean metric, where t_1 *and* t_2 *are two arbitrary points of the space* R^G*:*

$$d(t_1, t_2) = \sqrt{\sum_{l=1}^{G} (t_{1l} - t_{2l})^2}. \qquad (1)$$

1. We choose an initial partition $S^0 = \{S_1^0, \ldots, S_k^0\}$*:*

$$S_i^0 = \{x_{i1}^0, \ldots, x_{in}^0\},\ \bigcup_{i=1}^{k} S_i^0 = X,\ S_i^0 \cap S_{i'}^0 = \varnothing,\ i \neq i'. \qquad (2)$$

No.	Stations of CP	Number of enterprises	Numbers in the list of enterprises	Volume	% of the total volume	Average distance to CP
k = 25. Total volume of cargo – 558296 t. Total distance – 19019,14 km. Transportation volume – 18234 000 t • km. Average distance to CP – 22,4 km. Average distance between CP – 98,1 km.						
1	Trofimovsky 2	54	788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 828, 829, 830, 831, 832, 835, 837, 838, 839, 840, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852	54713	9,8	2
2	Krotovka	12	150, 151, 199, 200, 208, 576, 577, 602, 611, 723, 724, 749	35373	6,3	16

Pic. 1 shows a fragment of the table, issued by the program.

2. Assume that the m-th partition $S^m = \{S_1^m, \dots, S_k^m\}$ *is constructed.*

Let's calculate the set of average vectors $E^m = \{e_1^m, \dots, e_k^m\}$ *i.e.* $e_i^m = \{e_{i1}^m, \dots, e_{iG}^m\}$, *where*

$$e_{il}^m = \frac{\sum_{j=1}^{n_i} v_j x_{ij}}{\sum_{j=1}^{n_i} v_j}, \qquad (3)$$

n_i *– number of points of the i-th cluster.*

3. We define the set of projections, which are average for the current partition:

$$C^m = \left\{ y \in Y : \forall i, d^*\left(y, e_i^m\right) = \min_{1 \le r \le p} d(y_r, e_i^m) \right\}. \qquad (4)$$

4. We construct the minimal distance partition generated by the set C^m, *and take it as* $S^{m+1} = (S_1^{m+1}, \dots, S_k^{m+1})$, *i.e. for the first*

$$S_i^{m+1} = \left\{ x \in X : d\left(x, c_i^m\right) = \min_{1 \le i \le k} d(x, c_i^m) \right\}, 1 \le i \le k. \qquad (5)$$

5. If $S^{m+1} \neq S^m$, *then proceed to step 2, replacing m by m+1; if* $S^{m+1} = S^m$, *then we set* $S^m = S^*$, $C^m = C^*$ *and finish the algorithm.*

The criterion for clustering in this algorithm is the functional [5, 6]:

$$F(S) = \sum_{i=1}^{k} \sum_{X \in S^i} \left\| X - e_i(S) \right\|^2. \qquad (6)$$

Since the functional F(S) does not increase on the sequence of partitions $S^0, S^1, \dots, S^m, \dots$, *which is constructed in the k-means algorithm, and* $F(S^m) = F(S^{m+1})$. *Only if* $S^m = S^{m+1}$, *then for any initial partition* S^0 *the algorithm finishes the work in a finite number of steps.*

The result of the classification depends on the choice of e^0, *so each time we have a local minimum of F(S). The coordinates* e^0 *were obtained as random numbers uniformly distributed in the rectangle of possible coordinates of the original points. To check the stability of the results and obtain various dependencies, the choice of* e^0 *changes.*

2.

The validity of the formed clusters, i.e. the acceptability of the results obtained for determining the location of container points is determined by validation of clusters.

There are two types of validation: internal – by how much clusters correspond to the initial data, and external (target) – by how much clusters correspond to information not taken into account in their construction, but known to specialists who use clustering for their own purposes.

For internal validation, a wide variety of indices are used that express the quality of clustering results.

The most popular is the Davis–Boldin index. The smaller is the value of this index, the more compact and more distant are clusters from each other. This allows us to justify the number of clusters k, which is important, because when clustering of production, the number of center-container points can in general not be specified and should be determined from on the optimization condition for some additional criterion.

As the target criteria for determining the quality of clustering, two options are considered:

1. The number of centers is set. This will be in case, when designing, resources are assigned for the creation of all CP and the average normative cost of one point is known. In this case, the costs of creating all the CP themselves are not optimized and the criterion is the cost of transporting goods from all customers to their container points.

2. The number of CP is not specified (k is unknown), but the average cost of one CP is known – c. Then the optimization criterion for clustering is the sum of the total transportation costs and the cost of creating the CP.

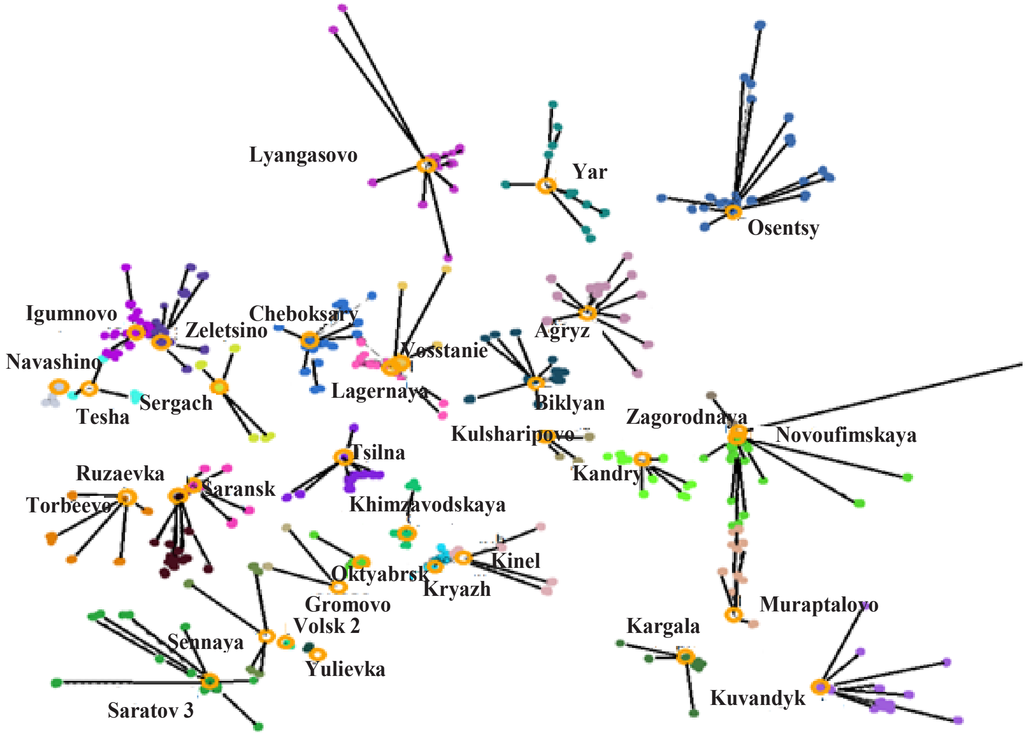
To carry out experimental and practical calculations, a JavaScript program was written. The program is implemented in several modes: with a given or an arbitrary number of clusters.

In the first mode, the implementation consists in applying the chosen algorithm and specifying the number of clusters k.

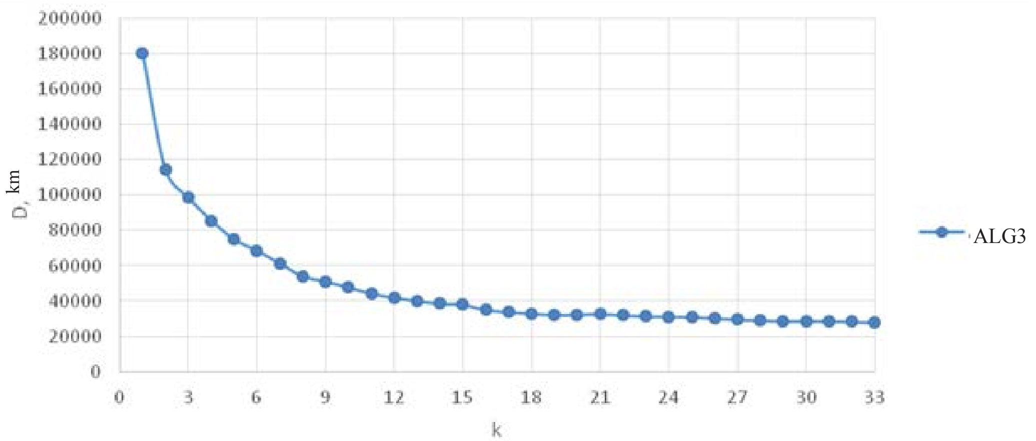
Solution of the set task is possible using three different algorithms:

1. The classical k-means algorithm [5, 6]. The result will be the location of container points in geometric «centers», providing optimal properties from the point of view of the lowest total traffic for the whole network from the point-enterprises to the CP. Such a clustering is called «free» and the algorithm realizing it is denoted as Algorithm 1.

2. k-means with the projection at the last step. First, we cluster objects with the help of the k-means algorithm and get a breakdown of enterprises in the form of clusters with geometric centers, and then for



Pic. 2. The result of the program for k = 33 in the clustering mode with the projection and the total distance criterion.



Pic. 3. Graph of the dependence of the total distance on the number of clusters.

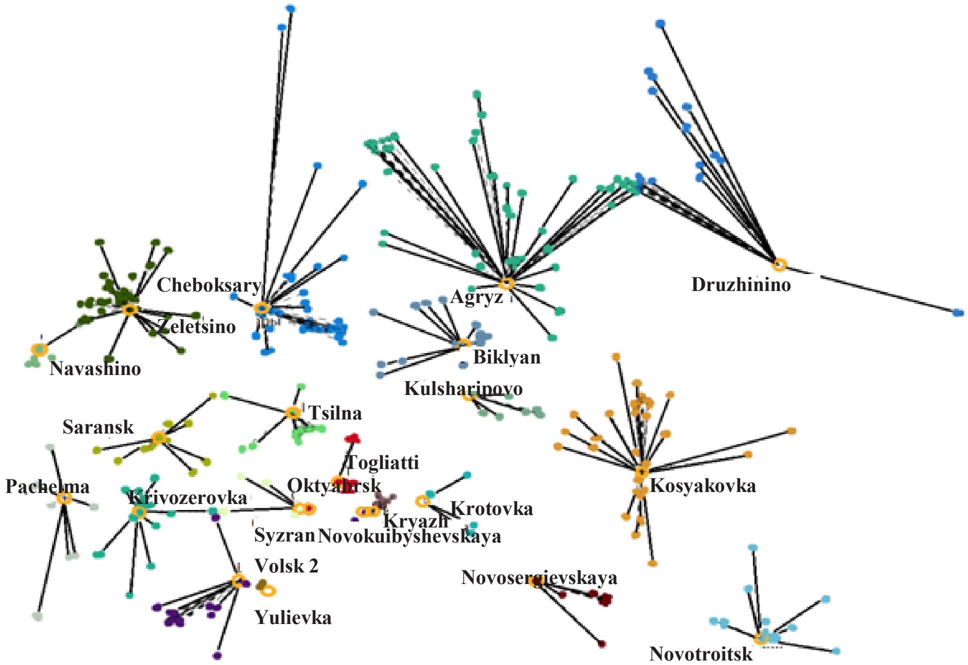
each center we find the nearest railway station and assume that there should be a container point here. Such an algorithm with a projection at the last iteration will be called Algorithm 2.

3. Modified k-means algorithm with projection (k-means pro). We select the number k and in the first step we throw out k random points, called «cluster centers». Then, each production is tied to the nearest center. As a result, each object is assigned to a specific cluster. We calculate the new «centers» as coordinate-wise average clusters, and then project them onto a multitude of railway stations. The received set is considered to be the new cluster centers, and the objects are redistributed again. The process of calculating centers and redistributing objects continues until the cluster centers have stabilized, i.e. all observations belong to the clusters to which they belonged prior to the current iteration. And this is Algorithm 3.

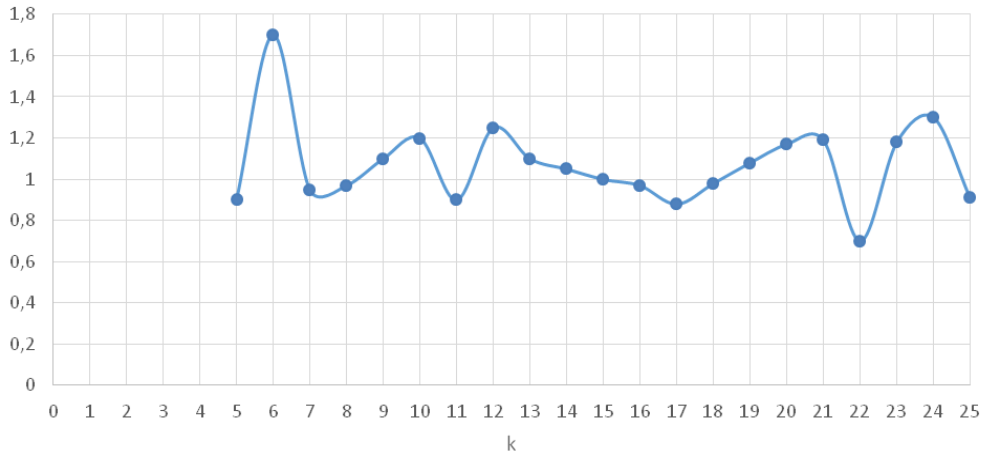
In the second mode, the number of clusters is determined according to the selected criterion (the minimum of the total transportation costs and the costs of creating a CP by means of the Davis-Boldin index) by looking through the options for each k.

It should be noted that such a variety of operating modes allows in an experimental way to achieve the best result of the partition.

On the basis of the developed models, algorithm and program, multiple experiments were



Pic. 4. The result of the program in the clustering mode with the projection and Davis–Boldin criterion.



Pic. 5. Graph of the dependence of the Davis–Boldin index on the number of clusters.

conducted in various regimes using the example of Privolzhsky Federal District (PFD). 900 industrial enterprises and 137 railway stations were considered. Production was determined by geographical coordinates and the volume of produced and / or extracted container-compatible products. A lot of stations are set on the network of six railways passing through the territory of Privolzhsky Federal District.

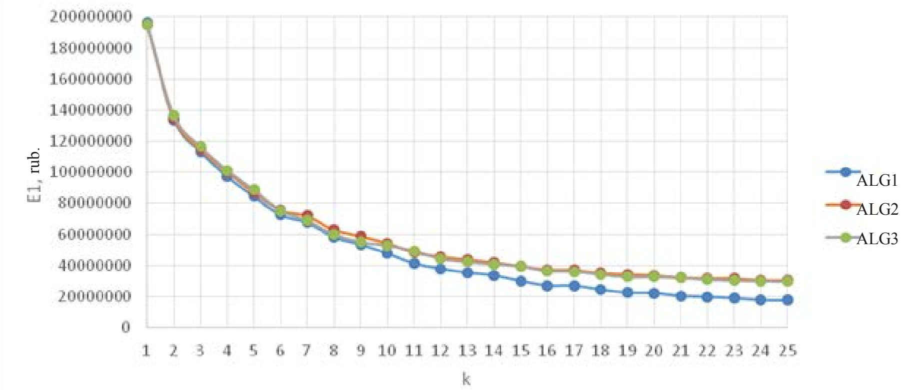
Let's consider some of the obtained results for PFD in the mode of the k-means algorithm with projection; the criterion is the total distance from all points to their centers $D=\sum_{i=1}^{n} D_i$ *. The result of the program for k = 33 is shown in Pic. 2. Data received: D = 28242 km. The graph of the dependence of criterion D on k is shown in Pic. 3.*

From the graph shown in Pic. 3, it follows that as the number k increases, the total distance $D=\sum_{i=1}^{n} D_i$ *is reduced.*

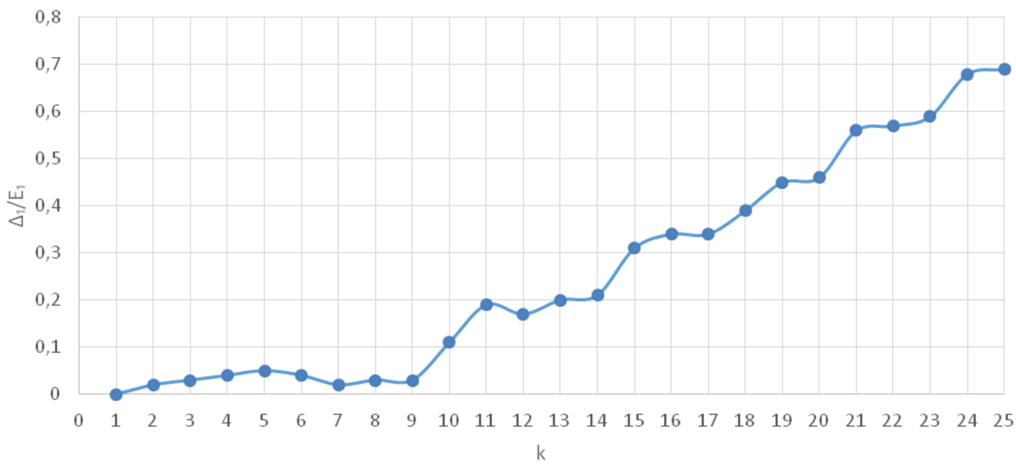
Further, to optimize the number of clusters k, we select the Davis–Boldin index, and we leave the algorithm unchanged-clustering with projection. The obtained data: k = 22, DB = 0.71. The result of the program for k = 22 is shown in Pic. 4. The graph of the dependence of the DB criterion is shown in Pic. 5.

3.

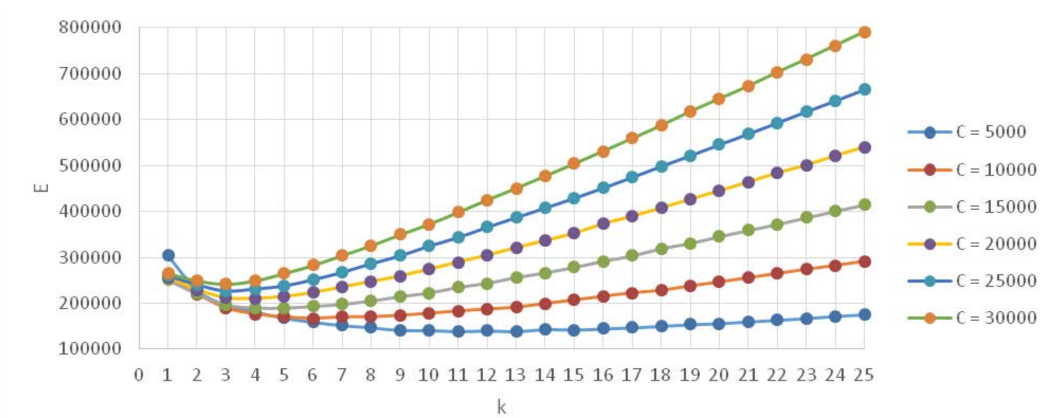
We perform clustering provided that the target criterion is specified in the case when the number k



Pic. 6. Graph of the dependence of transportation costs on the number of clusters.



Pic. 7. Dependence of ∆/E1 on k for production of PFD.



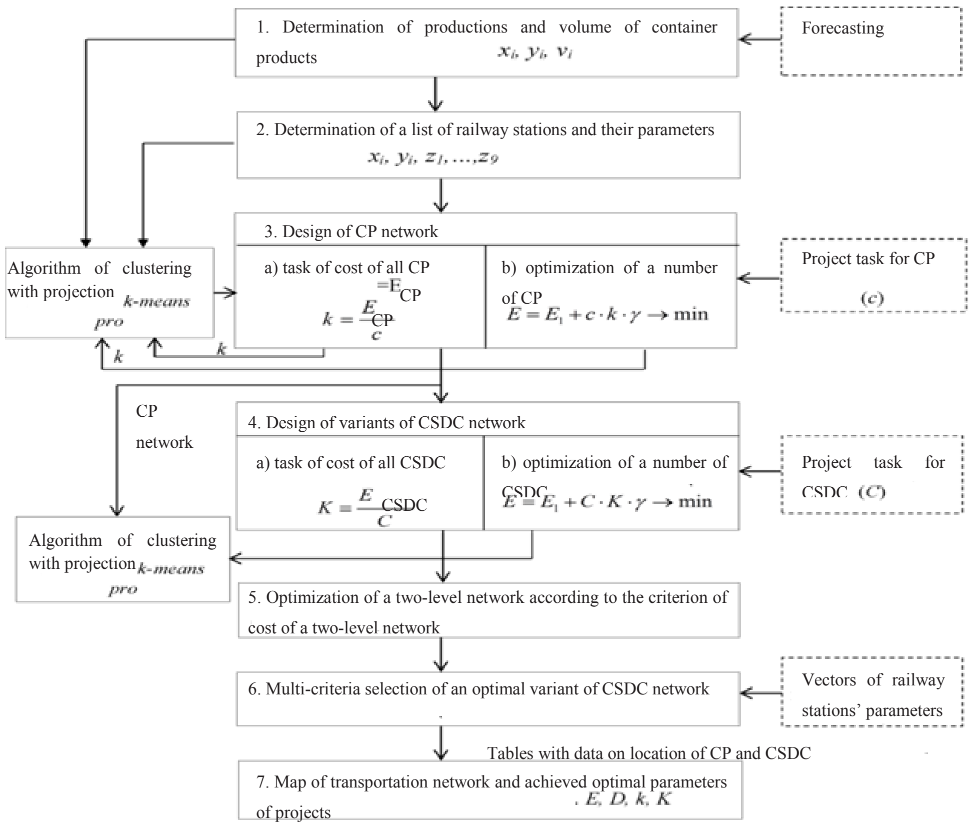
Pic. 8. Graphs of the dependence of total costs on k for different c (algorithm 3).

(CP) is known, i.e. when designing resources to create all CP and the average normative cost of one container point is given. In this case, the costs of creating all CP are not optimized, as the criterion is the cost of transporting goods from all customers to their CP, i.e.

$$E_1 = \sum_{i=1}^{n} D_i \cdot V_i \cdot s \to \min, \qquad (7)$$

where E_1 – transportation costs; D_i – distance from a production point to CP; V_i – volume of production of a container-compatible products; s – freight rate.

Clustering was carried out for given quantities k, which varied from 1 to 25 for all three algorithms. The results of the execution of Algorithm 1, Algorithm 2



Pic. 9. Algorithm for modeling of a two-level structure of the CTS of rail transport.

and Algorithm 3 in the form of a graph of dependencies are shown in Pic. 6.

As can be seen from Pic. 6, the total costs of transportation fall with increasing k – the number of CP. From this point of view, the more CP are there, the lower are the costs of transportation from production to container points.

In our case, the cluster centers must necessarily be on the railway line and this is a limitation for the very process of clustering. The k-means pro algorithm projects the cluster centers to the railway station each time.

We call the projection defect the difference between the criterion values of the quality of free clustering and the clustering «with projection» $\Delta = E_{1pr} - E_1$. The dependence $\Delta / E1$ on k for the production of PFD is shown in Pic. 7.

From here one can observe the following dependence: with a significant increase in the number of CP, the projection defect increases, i.e. in some cases, when the difference reaches 30–40 %, it is probably more profitable to construct a CP, to create a new infrastructure, rather than to place them on the existing infrastructure.

Next, we consider clustering under the condition that the number of CP is not specified (k is unknown), but the average cost of one CP – c is known. Then the optimization criterion for clustering is the sum of the total transportation costs and the cost of creating the CP, i.e.

$$E = \sum_{i=1}^{n} D_i \cdot V_i \cdot s + c \cdot k \cdot \gamma \to \min, \qquad (8)$$

where γ – normative efficiency coefficient.

The results of operation of the new algorithm k-means pro (algorithm 3) for PFD in the form of graphs of the dependence of the total costs E on k for various conditional c are shown in Pic. 8.

The optimal variant for different c will look like this:
for c = 5000 – 11 clusters; for c = 20000 – 4 clusters;
for c = 10000 – 6 clusters; for c = 25000 – 2 clusters;
for c = 15000 – 5 clusters; for c = 30000 – 2 clusters.

To optimize the location of CSDC with the use of clustering algorithms, a mathematical model was developed.

Let the coordinates and parameters of the railway stations (their numbers l = 1, 2,...L) be given and k station numbers from L, in which there will be CP, be found. It is necessary to find r station numbers where CSDC will be located.

At the first stage, we will assume that candidates for the deployment of CSDC can be any station from the general list. Let us imagine the search as a task of finding cluster centers of stations-CP. The solution is based on the k-means pro algorithm. At the same time, it is possible to obtain not only the optimal variant based on the cost criterion for the transportation of goods from all CP to their CSDC, but also a number of suboptimal options having very close criteria values.

In the second stage, we consider the properties of station points and additional criteria for optimizing clustering. That is, the point- station has q coordinates, which determine its properties with respect to how much it satisfies the purposes of creating a CSDC in it. The first two coordinates are the coordinates of the terrain in a planar system (x, y). Next are the components of the vector of coordinates, quantitatively measuring the nine criteria for the creation of the CSDC.

Each component should be expressed in conventional units, comparable to other coordinates. For this, the values of the criteria $z_1,\dots, z_9$ must be converted into dimensionless quantities. The most common normalization method is z:

$$Z_{\text{norm}} = \frac{z - \overline{z}}{\sigma_z}, \qquad (9)$$

where z – average value; s_z – standard deviation of the values z.

So, as a result of solving the problem of clustering of the first level, the CP is determined. Each such point-CP in addition to the coordinates is characterized by «weight», which is determined by the volume of the processed containers.

Then the stations are found for CSDC as centers of clusters for points-CP. Clustering in this case should be such that, taking into account the volumes of CP, clusters that are as compact as possible in the (x, y) coordinates are obtained, and their centers are as far apart as possible from each other. Thus, only the coordinates (x, y) become clustered traits, and all other attributes will be additional. Taking into account the volumes, this leads to minimization of the total costs for transportation of goods from the CP to the CSDC. All other criteria $z_1,\dots, z_9$ are additional for clustering and act as restrictions on the choice of CSDC points.

At the level of preliminary design of the locations of CSDC, two options for setting the task are possible.

1. With the available means A to construct $r = A/C$ CSDC in the places optimizing the integral efficiency index B, where C – average cost of CSDC.

2. To construct an optimal number of CSDC, optimizing the integrated efficiency index B taking into account the costs of crating CSDC network. These costs will be $r \cdot C$.

In order to take into account all properties of points-stations, it is necessary to ensure the implementation of the main rule – the greater are the values of the components of the coordinates $z_1,\dots, z_9$ for points-stations, the more expedient is it to select the center-CSDC at this point-station.

Pic. 9 shows the algorithm for modeling the design of a two-level structure of the CTS of rail transport.

Conclusion. In the course of the research, mathematical models, methods, a clustering algorithm with a projection and a software product were developed that can be used to solve problems related to the design of the location of transport facilities. Algorithm and software product were tested in the practical calculations of the optimum from the point of view of the specified criteria location of objects of the terminal infrastructure of Privolzhsky Federal District. The results obtained make it possible to recommend a unified methodology based on cluster analysis methods as a means for rational organization of the terminal infrastructure serving container trains.

REFERENCES

1. Rezer, S. M., Moskvichev, O. V., Moskvicheva, E. E. Optimization of the model of formation and functioning of the container-transport system of the country [*Optimizacija modeli formirovanija i funkcionirovanija kontejnerno-transportnoj sistemy strany*]. *Transport: nauka, tehnika, upravlenie*, 2016, Iss. 7, pp. 3–7.

2. Moskvichev, O. V. About a new approach to the organization of container trains in the inland traffic [*O novom podhode k organizacii kontejnernyh poezdov vo vnutrennem soobshhenii*]. *Zheleznodorozhnyj transport*, 2014, Iss. 2, pp. 56–59.

3. Mirkin, B. G. Methods of cluster analysis for decision support [*Metody klaster-analiza dlja podderzhki prinjatija reshenij*]. Moscow, Vysshaja shkola ekonomiki publ., 2011, 88 p.

4. Moskvichev, O. V., Esipov, B. A., Skladnev, N. S., Alyoshintsev, A. O. Development and research of a clustering algorithm with projection for solving the problems of transport infrastructure optimization [*Razrabotka i issledovanie algoritma klasterizacii s proekciej dlja reshenija zadach optimizacii transportnoj infrastruktury*]. *Perspective information technologies: Proceedings of the international scientific-technical conference*. Samara, Samara Scientific Center of the Russian Academy of Sciences, 2017, pp. 633–637.

5. Dragan, D. D. Research of algorithms and methods of intellectual support of decision-making [*Issledovanie algoritmov i metodov intellektual'noj podderzhki prinjatija reshenij*]. [Electronic resource]: http://cad.kpi.ua/attachments/diplomas/presentations/2011_B_05_Dragan.pdf. Last accessed 10.06.2017.

6. Aivazyan, S. A., Buchstaber, V. M., Enyukov, I. S., Meshalkin, L. D. Classification and diminution of dimensionality [*Klassifikacija i snizhenie razmernosti*]. Moscow, Finansy i statistika publ., 1989, 607 p.

7. Moskvichev, O. V., Tretyakov, G. M., Moskvicheva, E. E. New approach to development of the container-transport system of Russia on the basis of formation of a container cluster [*Novyj podhod v razvitii kontejnerno-transportnoj sistemy Rossii na osnove formirovanija kontejnernogo klastera*]. *Vestnik transporta Povolzh'ja*, 2012, Iss. 3, pp. 35–39.

8. Moskvichev, O. V., Nikonov, Yu. S. Estimation of the potential and perspectives of development of a container transport system [*Ocenka potenciala i perspektiv razvitija kontejnernoj transportnoj sistemy*]. *Zheleznodorozhnyj transport*, 2013, Iss. 4, pp. 37–39.

9. Moskvichev, O. V., Moskvicheva, E. E., Nikonov Yu. S. To the issue of consolidation and delivery of single and group container shipments as a part of container trains [*K voprosu konsolidacii i dostavki odinochnyh i gruppovyh kontejnernyh otpravok v sostave kontejnernyh poezdov*]. *Transport Urala*, 2014, Iss. 2, pp. 15–18.

10. Moskvichev, O. V. Multicriteria assessment of the containerability of manufactured products as one of the factors determining placement of the terminal-logistics infrastructure [*Mnogokriterial'naja ocenka kontejneroprigodnosti proizvodimoj produkcii kak odin iz faktorov, opredeljajushhih razmeshhenie terminal'no-logisticheskoj infrastruktury*]. *Vestnik transporta Povolzh'ja*, 2015, Iss. 1, pp. 74–80. ●

Information about the author:
Moskvichev, Oleg V. – Ph.D. (Economics), associate professor, head of Operational management department, director of the Center for Modeling of Transport and Logistics Processes of Samara State Transport University, Samara, Russia, moskvichev063@yandex.ru.

Article received 21.03.2017, revised 29.05.2017, accepted 10.06.2017.