

АЛГОРИТМ КЛАСТЕРИЗАЦИИ С ПРОЕКЦИЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ РАЗМЕЩЕНИЯ ТРАНСПОРТНЫХ ОБЪЕКТОВ

© 2017

Б. А. Есипов кандидат технических наук, доцент кафедры информационных систем и технологий;
Самарский национальный исследовательский университет
имени академика С.П. Королёва;
bobpereira@yandex.ru

О. В. Москвичев кандидат экономических наук, доцент, заведующий кафедрой
«Управление эксплуатационной работой»;
Самарский государственный университет путей сообщения;
moskvichev063@yandex.ru

Н. С. Складнев студент;
Самарский национальный исследовательский университет
имени академика С.П. Королёва;
author_skn@gmail.com

А. О. Алёшинцев студент;
Самарский национальный исследовательский университет
имени академика С.П. Королёва;
fen1r16n@yandex.ru

Предлагается математическая модель и новый метод решения задачи оптимального размещения логистических центров двухуровневой сети железнодорожных перевозок на основе применения математического аппарата кластерного анализа. При заданных геоинформационных параметрах производств-поставщиков, а также заданных железнодорожных станциях сети железных дорог ставится задача оптимального выбора станций – контейнерных пунктов (КП). Критерием является минимизация общего суммарного объёма перевозок в тонно-километрах от производств до КП. Для этого в качестве оптимизационной математической модели используется модель разбиения объектов на кластеры. Искомыми кластерами являются подмножества точек-производств со своими центрами-КП. Поскольку центры кластеров обязательно должны находиться в железнодорожных станциях, в статье предложен новый алгоритм кластеризации «с проекцией». Исследованы возможности такого алгоритма кластеризации, названного *k-means pro*. Рассмотрена методика оптимизации выбора мест расположения контейнерных накопительных распределительных центров как центров кластеров второго уровня для двухуровневой сети перевозок. Приведены примеры расчётов для предприятий и железных дорог Приволжского федерального округа на основе созданного программного обеспечения.

Размещение транспортных объектов; двухуровневая сеть контейнерных перевозок; логистические центры; кластеризация с проекцией; дефект проекции.

Цитирование: Есипов Б.А., Москвичев О.В., Складнев Н.С., Алёшинцев А.О. Алгоритм кластеризации с проекцией для решения задач оптимизации размещения транспортных объектов // Вестник Самарского университета. Аэрокосмическая техника, технологии и машиностроение. 2017. Т. 16, № 4. С. 108-117. DOI: 10.18287/2541-7533-2017-16-4-108-117

Введение

В настоящее время реконструкция и развитие транспортной системы страны сталкиваются с трудностями поиска оптимальных мест размещения транспортных коммуникаций. Современные решения по размещению элементов транспортных систем не всегда отвечают требованиям рациональности вследствие сложности и многовариантности задач. Проблемой является то, что строительство транспортных объектов производится однократно, в то время как эксплуатация – в течение многих десятилетий.

Основной идеей для повышения эффективности транспортных сетей является создание многоуровневой инфраструктуры с центрами обслуживания на каждом уровне. Так, например, для решения общей задачи выбора оптимальной двухуровневой сети транспортных объектов, реализующих технологию контейнерных поездов, предлагается на первом уровне все производства с контейнеропригодной продукцией привязать к железнодорожным контейнерным пунктам, а на втором уровне создать контейнерные накопительно-распределительные центры (КНРЦ), к которым будут привязаны подмножества КП [1;2].

КП и КНРЦ являются центрами обслуживания грузопотоков своего уровня: первый уровень – от центров производства до КП и второй уровень от КП до КНРЦ (рис. 1). Предполагается, что между КНРЦ грузоперевозки осуществляются контейнерными поездами, а далее до получателей в обратном порядке.

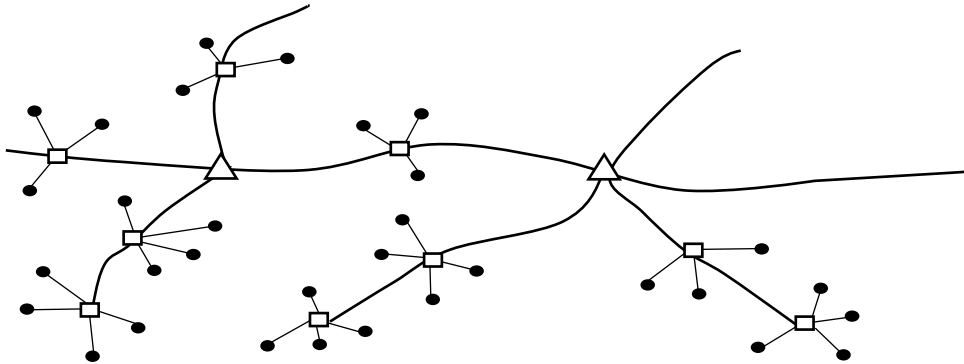


Рис. 1. Двухуровневая сеть контейнерных перевозок:
точки – центры производства; квадраты – КП; треугольники – КНРЦ

В работе решается задача выбора мест расположения указанных транспортных объектов при заданных координатах точек производств, их объемах контейнеропригодной продукции и при заданной сети железных дорог в виде множества станций, где возможно построение КП или КНРЦ. В качестве критерия оптимизации выступают общие затраты на перевозку и создание новых объектов инфраструктуры. Поскольку объемы перевозимой продукции фиксированы, то оптимизация связана в первую очередь с сокращением расстояний при перевозках от производств до КП и от КП до КНРЦ.

Постановка задачи оптимального размещения транспортных объектов на основе теории центров обслуживания потребителей приводит к многомерным задачам дискретной оптимизации на основе переборных алгоритмов NP сложности и не позволяет решать поставленные задачи для десятков тысяч предприятий и тысяч железнодорожных станций [3;4].

Классические задачи о кратчайшем расстоянии между точками и их центром на плоскости решаются, когда точки, для которых находится центр, заданы [5]. В общем случае оптимизация расстояний от центров подмножеств точек должна достигаться вариацией самих подмножеств точек, что приводит к задаче оптимальной кластеризации исходного множества точек и определения оптимальных центров кластеров.

Математическая модель размещения транспортных объектов

Предлагается для целей оптимального выбора мест расположения КП и КНРЦ применить универсальную методологию разбиения множества объектов с заданными свойствами на подмножества при заданных критериях разбиения и получения «цен-

тров» этих подмножеств, обладающих оптимальными свойствами. В качестве такой универсальной процедуры предлагается использовать математические методы кластеризации объектов, известные как *кластерный анализ* [6;7].

Геометрическая близость объектов от центра гарантирует минимизацию расстояний при перевозке, а учёт «веса» каждого объекта, выражающего объём перерабатываемой объектом продукции, оптимизирует общие затраты перевозок. Большое достоинство кластерного анализа заключается в том, что он позволяет производить разбиение объектов не по одному параметру, а по их целому набору. Анализ литературы по кластерному анализу и опыт использования стандартных программных средств кластерного анализа позволяет утверждать, что принципиально возможно решать практические задачи большой размерности (для федеральных округов и всей страны в целом).

Однако при подходе к решению практических задач оптимизации местоположения КП и КНРЦ на основе идеи кластеризации возникла новая проблема, решение которой развивает сами методы кластерного анализа.

Так при применении алгоритмов кластеризации по известному методу *k-means* считается, что оптимальный «центр» может находиться в любой точке пространства параметров, определяющих объекты. Если параметры – это геометрические координаты центров производства, то «центр» лежит в любой точке плоскости. На практике следует рассмотреть случай, когда «центр» обязательно должен находиться в одной из заданных точек (например, железнодорожной линии, железнодорожной станции). Поэтому при решении задачи определения мест КП и КНРЦ приходится решать задачу кластеризации «с проекцией на функцию», когда «центр» обязательно должен находиться на железнодорожной магистрали или «с проекцией на точки» – на железнодорожной станции.

Предлагается новый алгоритм кластеризации с проекцией на множество точек, названный «*k-means pro*», и исследуется возможность его применения в практических задачах проектирования транспортной инфраструктуры [8].

Входными данными алгоритма является множество объектов кластеризации $X = \{x_1, \dots, x_n\}$, их веса $V = \{v_1, \dots, v_n\}$ и допустимое множество проекций $Y = \{y_1, \dots, y_p\}$. Каждый j -й объект и каждая допустимая точка-проекция заданы в G -мерном пространстве R^G , т.е. $x_j = (x_{j1}, \dots, x_{jG})$ и $y_r = (y_{r1}, \dots, y_{rG})$.

Единственным управляющим параметром является число кластеров k , на которые производится разбиение $S = \{S_1, \dots, S_k\}$ множества X . В результате получается не смещённое разбиение $S^* = \{S_1^*, \dots, S_k^*\}$, центры которого являются оптимальным множеством проекций $C^* \subseteq Y$.

Введём обозначения:

- n – количество объектов кластеризации;
- p – количество точек допустимого множества проекций;
- i, i' – номер кластера;
- j – номер объекта;
- r – номер точки множества проекций;
- l – номер координаты точки;
- t – номер текущей итерации;
- G – размерность пространства, в котором выполняется кластеризация.

Расстояние между точками в G -мерном пространстве определяется по евклидовой метрике, где t_1 и t_2 – две любые точки пространства R^G : $d(t_1, t_2) = \sqrt{\sum_{l=1}^G (t_{1l} - t_{2l})^2}$.

Алгоритм «*k-means pro*»

1. Выберем начальное разбиение $S^0 = \{S_1^0, \dots, S_k^0\}$:

$$S_i^0 = \{x_{i1}^0, \dots, x_{in}^0\}, \bigcup_{i=1}^k S_i^0 = X, S_i^0 \cap S_{i'}^0 = \emptyset, i \neq i'.$$

2. Пусть построено m -е разбиение $S^m = \{S_1^m, \dots, S_k^m\}$.

Вычислим набор векторов средних $E^m = \{e_1^m, \dots, e_k^m\}$:

$$e_i^m = (e_{i1}^m, \dots, e_{iG}^m),$$

где $e_{il}^m = \frac{\sum_{j=1}^{n_i} v_j x_{jl}}{\sum_{j=1}^n v_j}$; n_i – количество точек i -го кластера.

3. Определим множество проекций средних для текущего разбиения:

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

4. Построим минимальное дистанционное разбиение, порожаемое множеством C^m , и возьмём его в качестве $S^{m+1} = (S_1^{m+1}, \dots, S_k^{m+1})$:

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

5. Если $S^{m+1} \neq S^m$, то переходим к п. 2, заменив m на $m+1$, если $S^{m+1} = S^m$, то полагаем $S^m = S^*$, $C^m = C^*$ и заканчиваем работу алгоритма.

Так как на последовательности разбиений S^0, S^1, \dots, S^m , которая строится в алгоритме, функционал качества разбиений $F(S)$:

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

не возрастает (причём $F(S^m) = F(S^{m+1})$, только если $S^m = S^{m+1}$), то для любого начального разбиения S^0 алгоритм через конечное число шагов заканчивает работу. Сложность вычислений по этому алгоритму оценивается как $O(nkm)$, где n – количество кластеризуемых объектов; k – количество кластеров, m – количество итераций.

Начальное разбиение $S^0 = \{S_1^0, \dots, S_k^0\}$ выбирается произвольным образом. Например, из всего множества координат объектов выбирается минимальное и максимальное число по x и по y , затем в этом диапазоне выбирается k точек (начальных центров кластеров). В работе координаты этих центров e^0 получены как случайные числа, равномерно распределённые в прямоугольнике возможных координат исходных точек в диа-

пазоне $[x_{min}, x_{max}]$ и $[y_{min}, y_{max}]$ соответственно. Затем для каждого объекта кластеризации выбиралась ближайшая из этих точек и таким образом получалось начальное случайное разбиение на непересекающиеся подмножества S^0 .

Для проверки устойчивости результатов и получения различных зависимостей менялся выбор e^0 , получались 100 различных реализаций, выбиралась наилучшая и усреднялись полученные параметры для построения графиков.

Исследование алгоритма «*k-means pro*»

Разработана программа, реализующая данный алгоритм с возможностью визуализации получаемых кластеров и вычисления разнообразных параметров. Для сравнения был взят классический алгоритм *k-means* Мак-Куина [5].

Сравнение проводилось на тестовых выборках точек, распределённых равномерно на плоскости. На рис. 2 показаны кластеры, полученные классическим *k-means*, а на рис. 3 – алгоритмом *k-means pro*. В данном примере железнодорожная магистраль представлена в виде «синусоиды» («+» отмечены центры кластеров, «o» – железнодорожные станции). Из рис. 2, 3 следует, что оптимальное разбиение точек на кластеры для обычного алгоритма приводит к тому, что центры располагаются произвольно (рис. 2), в то время как с использованием алгоритма *k-means pro* оптимальная кластеризация допускает только центры, лежащие на железнодорожных станциях (рис. 3).

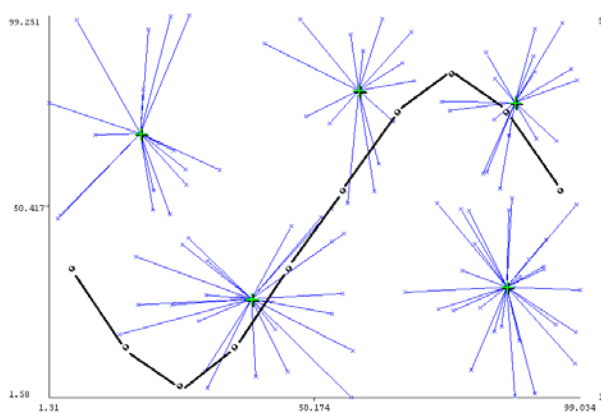


Рис. 2. Свободная кластеризация

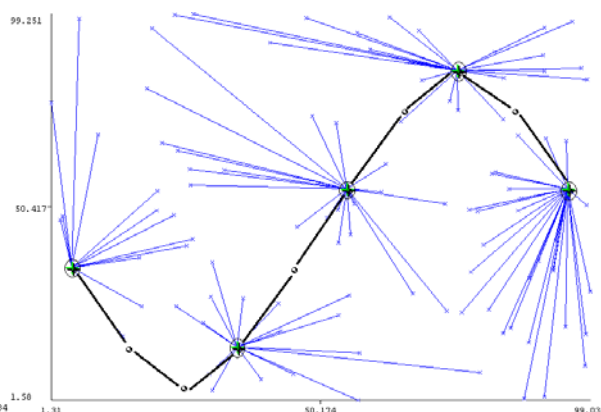


Рис. 3. Кластеризация с проекцией

Разработанный алгоритм применён для решения задачи оптимального выбора мест расположения КП для заданных 900 производств и 137 железнодорожных станций Приволжского федерального округа (ПФО). Производства определялись географическими координатами и объёмом контейнеропригодной продукции. Множество железнодорожных станций задано на сети 7 железных дорог, расположенных на территории ПФО. Результат при $k = 42$ представлен на рис. 4. Точками изображены заданные производства, малыми квадратами – железнодорожные станции, большими квадратами – найденные станции-КП. Показаны кластеры производств.

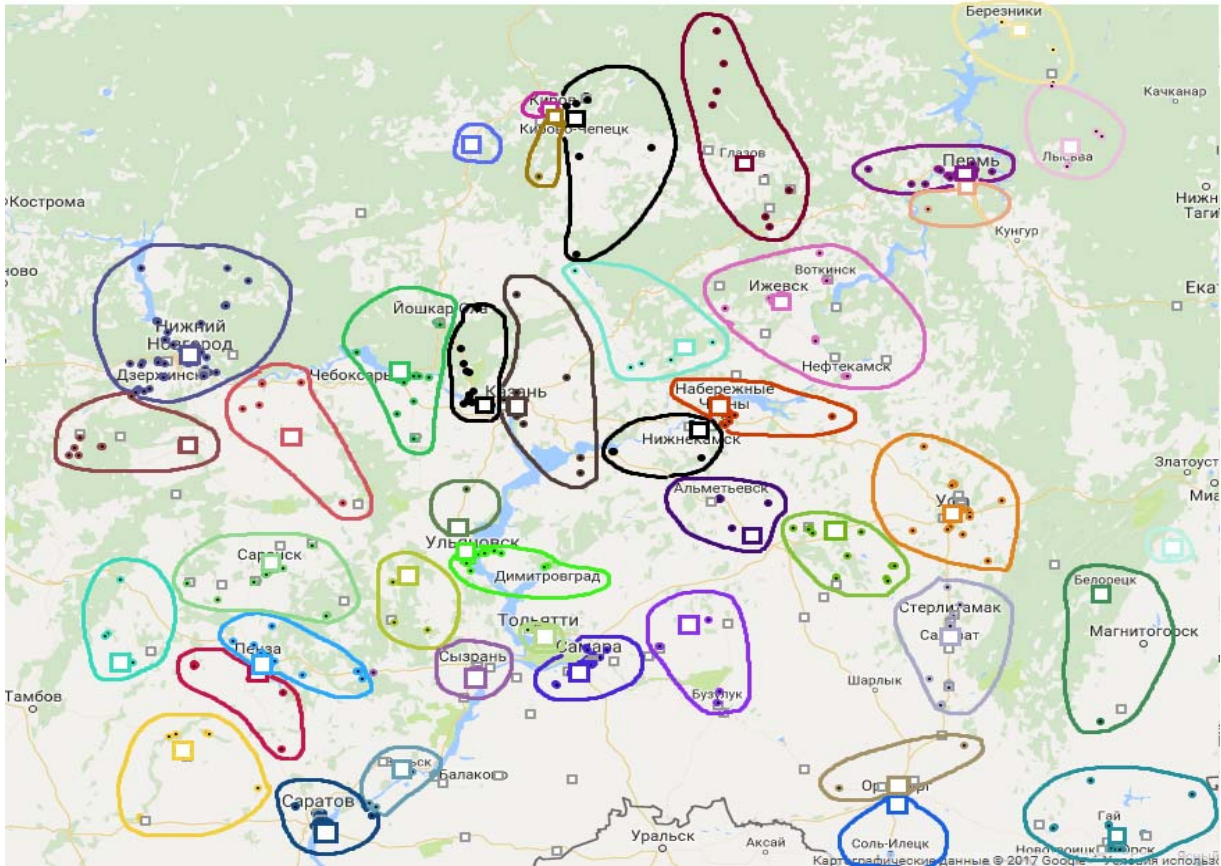


Рис. 4. Кластеризация предприятий ПФО

Представляет интерес соотношение показателей качества кластеризации для обычного алгоритма и алгоритма «кластеризации с проекцией». В первом случае центры кластеров определяются исключительно из свойств расположения объектов (точек) и критерия оптимальности кластеризации D . Таковую кластеризацию в контексте нашей работы можно назвать свободной.

В рассматриваемом случае центры кластеров обязательно должны находиться на железнодорожной линии, что является ограничением для процесса кластеризации. Алгоритм каждый раз проектирует центры кластеров на эту линию. В результате получаем вариант кластеризации с проекцией и, очевидно, с другим значением критерия D_{np} .

Для классического алгоритма k -means

$$D = \sum_{i=1}^k \sum_{j=1}^{n_i} d(x_{ij}, e_i), \quad d(x_{ij}, e_i) = \sqrt{(x_{ij1} - e_{i1})^2 + (x_{ij2} - e_{i2})^2}.$$

Для алгоритма k -means pro

$$D_{np} = \sum_{i=1}^k \sum_{j=1}^{n_i} d(x_{ij}, c_i^*), \quad d(x_{ij}, c_i^*) = \sqrt{(x_{ij1} - c_{i1}^*)^2 + (x_{ij2} - c_{i2}^*)^2}.$$

На рис. 5 изображены зависимости D и D_{np} (в км) от k для рассмотренного примера, а на рис. 6 – для ПФО. Видно, что суммарное расстояние сначала резко падает при увеличении числа кластеров, а затем меняется слабо.

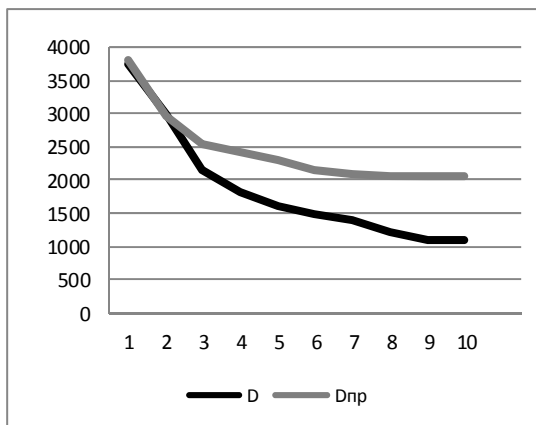


Рис. 5. Зависимость D от k

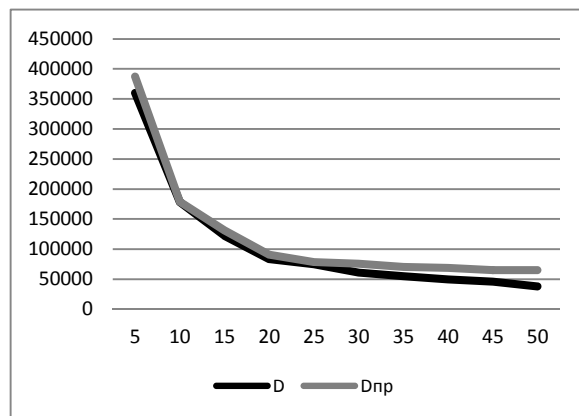


Рис. 6. Зависимость D от k для ПФО

Назовём дефектом проекции разницу критериальных величин качества свободной кластеризации и кластеризации «с проекцией»: $\Delta = D_{пр} - D$. Зависимость Δ / D от k для производств ПФО представлена на рис. 7.

Из графика видно, что при значительном увеличении числа КП растёт дефект проекции. В некоторых случаях, когда величина Δ значительна, целесообразно размещать КП в новых станциях, а не на существующей инфраструктуре.

Вышеприведённые графики дают возможность количественно оценить такие варианты.

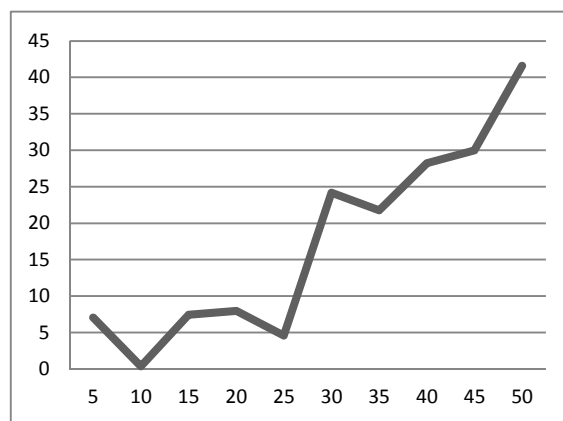


Рис. 7. Дефект проекции

Оптимизация размещения КНРЦ

Задача оптимального выбора мест расположения КНРЦ может быть решена на основе алгоритма кластеризации с проекцией, где исходными точками являются уже полученные точки-КП.

Поэтому для этой задачи можно использовать разработанные алгоритмы и программные средства. Для демонстрации решения задачи оптимального выбора мест расположения КНРЦ возьмём за основу проект оптимального расположения КП для $k = 42$.

Сначала решим задачу оптимального размещения сети КНРЦ по критерию минимизации затрат на перевозку всех объёмов перерабатываемой продукции от КП до КНРЦ для некоторых значений числа кластеров K (количества КНРЦ).

Зависимость суммарных расстояний перевозки $D_{КНРЦ}$ (в км) от КП до КНРЦ от количества КНРЦ представлена на рис. 8. Из неё следует, что уменьшение суммарного расстояния с ростом K замедляется, что позволяет выбрать для ПФО три-четыре КНРЦ.

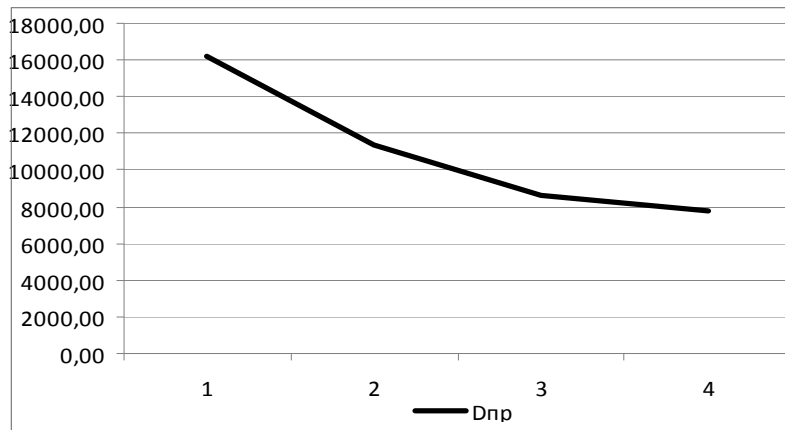


Рис. 8. Зависимость суммарного расстояния от КП до КНРЦ от K

Работа с алгоритмом кластеризации *k-means pro* показала одну особенность. При повторении кластеризации с выбором другого начального множества центров e^0 получаются другие станции-КНРЦ, хотя значения критерия D близки. Это можно использовать для выбора КНРЦ с использованием дополнительных критериев [9].

Заключение

Приведённые примеры показывают возможность и порядок расчётов для различных вариантов постановок задач оптимизации размещения транспортных объектов на основе применения алгоритмов кластерного анализа. Разработанный алгоритм *k-means pro* показал высокую эффективность при решении прикладных задач оптимизации. Созданные программные средства применены для решения задач проектирования инфраструктуры перевозок на примере ПФО. Предложенные модели, алгоритмы и методики позволят лицам, принимающим решения, количественно оценить различные варианты в конкретных условиях проектирования при оптимизации количества терминальных объектов и мест их размещения.

Библиографический список

1. Резер С.М., Москвичев О.В., Москвичева Е.Е. Оптимизация модели формирования и функционирования контейнерно-транспортной системы страны // Транспорт: наука, техника, управление. 2016. № 7. С. 3-7.
2. Москвичев О.В. О новом подходе к организации контейнерных поездов во внутреннем сообщении // Железнодорожный транспорт. 2014. № 2. С. 56-59.
3. Разина М.А. Математические модели и оптимизация размещения станций скорой помощи для обслуживания населения заданной области. Дис. ... канд. физ.-мат. наук. Казань, 2005. 110 с.
4. Михалевич В.С., Трубин В.А., Шор Н.З. Оптимизационные задачи производственно-транспортного планирования: Модели, методы, алгоритмы. М.: Наука, 1986. 260 с.
5. Протасов В.Ю. Максимумы и минимумы в геометрии. М.: Московский центр непрерывного математического образования. 2005. 56 с.
6. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
7. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика, 1988. 176 с.

8. Есипов Б.А., Москвичёв О.В., Складнев Н.С., Алёшинцев А.О. Разработка и исследование алгоритма кластеризации с проекцией для решения задач оптимизации транспортной инфраструктуры // Сборник трудов международной научно-технической конференции «Перспективные информационные технологии (ПИТ 2017)». Самара: Самарский научный центр РАН, 2017. С. 633-637.

9. Москвичев О.В. Модели, методы и алгоритмы оптимизации контейнерно-транспортной системы железнодорожного транспорта на основе кластерного подхода. // Транспорт Урала. 2017. № 2 (53). С. 18-27.

CLUSTERING ALGORITHM WITH PROJECTION FOR SOLVING PROBLEMS OF OPTIMAL ALLOCATION OF TRANSPORT FACILITIES

© 2017

B. A. Esipov Candidate of Science (Engineering), Associate Professor,
Department of Information Systems and Technologies;
Samara National Research University, Samara, Russian Federation;
bobpereira@yandex.ru

O. V. Moskvichev Candidate of Science (Economics), Associate Professor,
Head of the Department of Operational Work Management;
Samara State Transport University, Samara, Russian Federation;
moskvichev063@yandex.ru

N. S. Skladnev Student;
Samara National Research University, Samara, Russian Federation;
author.skn@gmail.com

A. O. Alyoshintsev Student;
Samara National Research University, Samara, Russian Federation;
fen1r16n@yandex.ru

The paper proposes a mathematical model and a new method for solving the problem of optimal location of logistic centers for a two-level railway transportation network based on the application of the mathematical apparatus of cluster analysis. Given the geo-information parameters of the supplier plants, as well as the railroad networks specified by the railroad stations, the task of the optimal choice of railway stations – container points (CP) – is set. The criterion is to minimize the total amount of traffic in ton-kilometers from the production plant to the CP. For this purpose, the model of dividing the object into clusters is used as an optimization mathematical model. The required clusters are subsets of production points with their own CP centers. Since cluster centers must necessarily be located at railway stations, the article suggests a new clustering algorithm “with projection”. The possibilities of such a clustering algorithm, called k-means pro, are explored. A method of optimizing the choice of location of container storage distribution centers as second-level centers for a two-level transportation network is described.

Location of transportation facilities; two-layer network of container transportations; logistic centers; clustering with projection; projection defect.

Citation: Esipov B.A., Moskvichev O.V., Skladnev N.S., Alyoshintsev A.O. Clustering algorithm with projection for solving problems of optimal allocation of transport facilities. *Vestnik of Samara University. Aerospace and Mechanical Engineering*. 2017. V. 16, no. 4. P. 108-117. DOI: 10.18287/2541-7533-2017-16-4-108-117

References

1. Rezer S.M., Moskvichev O.V., Moskvicheva E.E. Optimization of Formation and Functioning of Container Transportation System of the Country. *Transport: science, technique, management*. 2016. No. 7. P. 3-7. (In Russ.)
2. Moskvichev O.V. A new approach to the organization of container train traffic in inland transport. *Zheleznodorozhnyy transport*. 2014. No. 2. P. 56-59. (In Russ.)

3. Razina M.A. *Matematicheskie modeli i optimizatsiya razmeshcheniya stantsiy skoroy pomoshchi dlya obsluzhivaniya naseleniya zadannoy oblasti. Dis. ... kand. fiz.-mat. nauk* [Mathematical models and optimization of location of ambulance stations to provide healthcare services to the population of a given region: Thesis for a Candidate Degree in Physics and Mathematics]. Kazan, 2005. 110 p.

4. Mikhalevich B.C., Trubin V.A., Shor N.Z. *Optimizatsionnye zadachi proizvodstvenno-transportnogo planirovaniya: Modeli, metody, algoritmy* [Optimization problems of production-transportation planning]. Moscow: Nauka Publ., 1986. 260 p.

5. Protasov V.Yu. *Maksimumy i minimumy v geometrii* [Maxima and minima in geometry]. Moscow: Moskovskiy Tsentr Nepreryvnogo Matematicheskogo Obrazovaniya Publ., 2005. 56 p.

6. Ayvazyan S.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D. *Prikladnaya statistika. Klassifikatsiya i snizhenie razmernosti* [Applied statistics. Classification and reduction of dimensionality]. Moscow: Finansyi Statistika Publ., 1989. 607 p.

7. Mandel I.D. *Klasternyy analiz* [Cluster analysis]. Moscow: Finansy i Statistika Publ., 1988. 176 p.

8. Esipov B.A., Moskvichev O.V., Skladnev N.S., Aleshintsev A.O. Development and investigation of a clustering algorithm with a projection for solving the problems of optimization of transport infrastructure. *Proceedings of the International Scientific Conference «Advanced Information Technologies and Scientific Computing (PIT 2017)»*. Samara: Samarskiy Nauchnyy Tsentr RAN Publ., 2017. P. 633-637. (In Russ.)

9. Moskvichev O.V. Models, methods and algorithms of optimization of container transport system of railway transport based on the cluster approach. *Transport Urala*. 2017. No. 2 (53). P. 18-27. (In Russ.)