УДК 656.073.7

DOI: 10.15827/0236-235X.121.005-011

Дата подачи статьи: 12.12.17

2018. Т. 31. № 1. С. 005–011

ПРОГРАММНЫЕ СРЕДСТВА РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ РАЗМЕЩЕНИЯ ТРАНСПОРТНЫХ ОБЪЕКТОВ НА ОСНОВЕ АЛГОРИТМА КЛАСТЕРИЗАЦИИ С ПРОЕКЦИЕЙ

Б.А. Есипов¹, к.т.н., доцент, bobpereira@yandex.ru

¹ Самарский национальный исследовательский университет им. академика С.П. Королева,
Московское шоссе, 34, г. Самара, 443086, Россия

Выбор оптимального расположения логистических центров является актуальной задачей для всех видов перевозок. Решение ее для реальных задач региона или страны приводит к математическим алгоритмам высокой сложности.

В статье предлагаются математическая модель и новый метод решения задачи оптимального размещения логистических центров двухуровневой сети перевозок на основе применения математического аппарата кластерного анализа. При заданных геоинформационных параметрах производств-поставщиков, а также заданной топологии дорог, станций или других объектов транспортной инфраструктуры ставится задача оптимального выбора логистических центров. Для сети железных дорог логистическими центрами могут быть, например, контейнерные пункты. Критерием оптимизации является минимизация общего суммарного объема перевозок в тонно-километрах от производств до контейнерных пунктов. Для этого в качестве оптимизационной математической модели используется модель разбиения объектов на кластеры. Искомыми кластерами являются подмножества точек-производств со своими центрами – контейнерными пунктами. Поскольку центры кластеров обязательно должны находиться на железнодорожных станциях, предложен новый алгоритм кластеризации с проекцией. Исследованы возможности такого алгоритма кластеризации, названного k-means pro.

Рассмотрена методика выбора количества центров (кластеров) по обобщенному экономическому показателю затрат на перевозки и создание логистических центров. Приведены примеры расчетов для предприятий и железных дорог Приволжского федерального округа на основе созданного ПО.

Ключевые слова: двухуровневая сеть контейнерных перевозок, логистические центры, кластеризация с проекцией, дефект проекции.

В настоящее время задача реконструкции и развития транспортной системы страны сталкивается 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. Основной проблемой является то, что строительство транспортных объектов производится однократно, в то время как эксплуатация – в течение многих десятилетий [1].

Основной идеей для повышения эффективности транспортных сетей является создание многоуровневой инфраструктуры (см. рис. 1) с центрами обслуживания на каждом уровне [2].

В работе решается задача выбора мест расположения транспортных объектов при заданных координатах точек производств, их объемах продукции, а также при заданной сети дорог в виде множества точек (станций), где возможно построение логистических центров. В качестве критерия оптимизации выступают общие затраты на перевозку и создание новых объектов инфраструктуры. Поскольку объемы перевозимой продукции фиксированные, оптимизация связана прежде всего с сокращением расстояний при перевозках от производств до логистических центров и обратно [1].

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 [2–4].

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

Рис. 1. Двухуровневая сеть контейнерных перевозок (точки – производства, квадраты и треугольники – логистические центры первого и второго уровней)

Fig. 1. A two-tier network of container transportations (points are production, squares and triangles are logistics centers of the first and second level)

ших n полный перебор таких вариантов определяется величиной порядка k^{n-1} , где k – количество кластеров, а n – количество точек (производств). Поэтому для практических задач необходимы методы, радикально уменьшающие перебор вариантов [4, 5].

Математическая модель размещения транспортных объектов на основе кластерного анализа

В данной работе для оптимального выбора мест расположения *контейнерных пунктов* (КП) и *контейнерных накопительно-распределительных центров* (КНРЦ) предлагается применить универсальную методологию разбиения множества объектов с заданными свойствами на подмножества при заданных критериях разбиения и получения центров этих подмножеств, обладающих оптимальными свойствами. В качестве такой универсальной процедуры предлагается использовать математические методы кластеризации объектов, известные как *кластерный анализ* [5, 6].

Геометрическая близость объектов от центра гарантирует минимизацию расстояний при перевозке, а учет веса каждого объекта, выражающего объем перерабатываемой объектом продукции, оптимизирует общие затраты на перевозки. Большое достоинство кластерного анализа и в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. Анализ литературы по кластерному анализу и опыт использования стандартных программных средств кластерного анализа позволяют утверждать, что принципиально возможно решать поставленные задачи для практических задач большой размерности (для федеральных округов и всей страны в целом) [7, 8].

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* [6] 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, на железнодорожной линии или станции). Таким образом, 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.

В работе предлагается новый алгоритм кластеризации с проекцией на множество точек, названный *k-means pro*, и исследуется возможность его применения в практических задачах проектирования транспортной инфраструктуры [4].

Входными данными алгоритма являются 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 , то есть $x_j = (x_{j1}, \dots, x_{jG})$ 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 несмещенное разбиение $S^* = \{S^*_1, \dots, S^*_k\}$, центры которого – это оптимальное множество проекций $C^* \subseteq Y$.

Введем следующие 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; m – 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

$$E^m = \{e_1^m, \dots, e_k^m\}, \text{ то есть } e_i^m = (e_{i1}^m, \dots, e_{iG}^m),$$

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

кластера.

3. Определим множество проекций средних для текущего разбиения:

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

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

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

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.

Поскольку на последовательности разбиений $S^0, S^1, \dots, S^m$, строящейся в алгоритме, функционал

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

возрастает, причем $F(S^m) = F(S^{m+1})$, только если $S^m = S^{m+1}$, 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 так: из всего множества координат объектов выбираются минимальное и максимальное число по x и по y , затем в этом диапазоне выбираются k точек (начальных центров кластеров). В данной работе координаты этих центров e^0 получены как случайные числа, равномерно распределенные в прямоугольнике возможных координат исходных точек в диапазонах $[x_{\min}, x_{\max}]$ и $[y_{\min}, y_{\max}]$ соответственно. Затем для каждого объекта кластеризации выбирается ближайшая из этих точек, таким образом получается начальное случайное разбиение на непересекающиеся подмножества S^0 .

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

Исследование алгоритма k-means pro

Разработана программа, реализующая приведенный алгоритм с возможностью визуализации получаемых кластеров и вычисления разнообразных параметров. Для сравнения взят классический алгоритм k-means Мак-Куина [6, 9, 10]. Сравнение производилось на тестовых выборках точек, распределенных равномерно на плоскости. На рисунке 2 показаны кластеры, полученные классическим k-means, а на рисунке 3 – алгоритмом k-means pro. В данном примере железнодорожная магистраль представлена в виде синусоиды (+ отмечены центры кластеров, о – железнодорожные станции). Из рисунков 2 и 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).

Разработанный алгоритм применен для решения задачи оптимального выбора мест расположе-

Рис. 2. Свободная кластеризация

Fig. 2. Free clustering

Рис. 3. Кластеризация с проекцией

Fig. 3. Clustering with a projection

ния КП для заданных 900 производств и 137 железнодорожных станций Приволжского федерального округа (ПФО). Производства определялись географическими координатами и объемом контейнеро-пригодной продукции. Множество железнодорожных станций задано на сети 7 железных дорог, расположенных на территории ПФО. Результат при $k = 42$ показан на рисунке 4, где точками изображены заданные производства, малыми квадратами – железнодорожные станции, большими квадратами – найденные станции-КП. Показаны кластеры производств.

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 . Такую кластеризацию в контексте данной работы можно назвать свободной.

В рассматриваемом случае центры кластеров обязательно должны находиться на железнодорожной линии, и это является ограничением для самого процесса кластеризации. Алгоритм каждый раз проектирует центры кластеров на железнодорожную линию. В результате получаем вариант класте-

Рис. 4. Кластеризация предприятий ПФО

Fig. 4. Clustering of Volga Federal District companies

ризации с проекцией и, очевидно, с другим значением критерия D_{np} .

Для классического алгоритма k-means

$$D = \sum_{i=1}^k \sum_{j=1}^{n_i} d(x_{ij}, e_i),$$

$$d(x_{ij}, e_i) = \sqrt{(x_{ij1} - e_{i1})^2 + (x_{ij2} - e_{i2})^2}.$$

Для k-means pro

$$D_{np} = \sum_{i=1}^k \sum_{j=1}^{n_i} d(x_{ij}, c_i^*),$$

$$d(x_{ij}, c_i^*) = \sqrt{(x_{ij1} - c_{i1}^*)^2 + (x_{ij2} - c_{i2}^*)^2}.$$

На рисунке 5 изображена зависимость D и D_{np} от k для вышеприведенного примера, а на рисунке 6 для ПФО. Видно, что суммарное расстояние сначала резко падает при увеличении числа кластеров, а затем меняется слабо.

Назовем дефектом проекции разницу критериевых величин качества свободной кластеризации и кластеризации с проекцией: $\Delta = D_{np} - D$. Зависи-

мость Δ / D от k для производств ПФО показана на рисунке 7.

Из графика видно, что при значительном увеличении числа КП растет дефект проекции, то есть в некоторых случаях, когда разница существенна, возможно, выгоднее строить КП на новых станциях, а не размещать их в существующей инфраструктуре. Вышеприведенные графики дают возможность количественно оценить такие варианты.

Выбор оптимального количества кластеров

Выбор оптимального числа кластеров является проблемным вопросом кластерного анализа [5, 7]. Рассмотрим экономический подход к выбору оптимального количества кластеров k . В данном случае это число определяет количество логистических центров (КП).

Итак, пусть количество КП не задано (k неизвестно), но известна приведенная к одному году

Рис. 5. Зависимость D от k

Fig. 5. A dependence of D on k

Рис. 6. Зависимость D от k для ПФО

Fig. 6. A dependence of D on k for Volga Federal District companies

Рис. 7. Дефект проекции

Fig. 7. A projection defect

нормативная стоимость затрат на создание одного КП – c . Тогда в качестве критерия оптимизации при кластеризации нужно взять величину общих затрат на перевозку плюс затраты на создание КП:

$$E = \sum_{i=1}^n D_i \cdot V_i \cdot s + c \cdot k \rightarrow \min, \quad (1)$$

$$k \leq \frac{A}{c}, \quad (2)$$

$$k \geq \frac{V}{v_{1КП}}, \quad (3)$$

$$E_1 = \sum_{i=1}^n D_i \cdot V_i \cdot s \quad - \quad \text{затраты на перевозку};$$

Рис. 8. Оптимизация затрат при создании центров инфраструктуры перевозок

Fig. 8. Costs optimization when creating transportation infrastructure centers

D_i – расстояние от точки-производства до КП; V_i – объем производства контейнеропригодной продукции i -го производства; s – тариф на перевозку 1 тонно-километра.

Поскольку первое слагаемое в (1) уменьшается при увеличении k , а второе увеличивается от k , величина E имеет минимум, который и определяет оптимальное k при условии, что эта величина попадает в интервал, определяемый ограничениями (2) и (3). На рисунке 8 показаны возможные случаи решения оптимизационной задачи (1), (2), (3). Очевидно, что ограничения (2) и (3) существенно влияют на выбор оптимального k .

Если объем V такой, что выделенных денег не хватает, то нет никакого решения задачи проектирования сети КП (рис. 8а). Если объем V очень мал и выделенных средств хватает на создание необходимого количества КП, то оптимальное k находится так, как показано на рисунке 8б (минимум обозначен кружком). Здесь оптимальная точка определяется ограничением (2). Рисунок 8в показывает выбор оптимального k , когда объем грузов очень большой, но выделяемых на создание КП средств хватает. Здесь оптимальная точка определяется ограничением (3). На рисунке 8г показан выбор оптимального k , когда объем грузов невелик, а средств на создание КП хватает с запасом. Оптимальная точка определяется минимумом (1). Очевидно, что средства A могут быть уменьшены до величины $A = ck_{opt}$.

Рассмотрим результаты оптимизации количества КП по критерию E для ПФО в виде графика зависимости общих затрат от k при различных s .

На рисунке 9 изображены кривые затрат на проект E : $E_{пр1}$ ($s = 1,5$ млн руб.), $E_{пр2}$ ($s = 5$ млн руб.), $E_{пр3}$ ($s = 12$ млн руб.), $s = 8$ руб./т-км. Из графика видно, что, например, для $s = 12$ млн руб. оптимальным решением будет создание 15 КП.

Рис. 9. Затраты на проект

Fig. 9. Project costs

В предложенной методике оптимальное количество КП зависит от величины s . Эта величина является, во-первых, нормативной, то есть усредненной и зависящей от даты проектирования, во-вторых, приведенной, так как увязывает текущие затраты E_1 и капитальные (единовременные) затраты ck . Как известно, это связано также с необходимостью оценки срока окупаемости КП. Все это показывает, что оптимизация числа k возможна только при тщательной экспертной оценке величины s .

Заключение

Полученные результаты показывают возможность и порядок расчетов различных вариантов постановок задач оптимизации размещения транспортных объектов на основе применения алгоритмов кластерного анализа. Разработанный алгоритм k-means pro и программные средства его реализации показали высокую эффективность при решении прикладных задач. Созданные программные средства применены для решения задач проектирования инфраструктуры перевозок на примере ПФО. Предложенные в статье модели, алгоритм и программы позволят лицам, принимающим решения, количественно оценить оптимальные и подоптимальные варианты в конкретных условиях проектирования при оптимизации количества терминальных объектов и мест их размещения.

Литература

1. Кириллова А.Г. Оптимальный выбор расположения терминала как логистического центра при организации контейнерных и контрейлерных перевозок // Транспорт: наука, техника, управление. 2010. № 10. С. 22–25.
2. Москвичев О.В. Модели, методы и алгоритмы оптимизации контейнерно-транспортной системы железнодорожного транспорта на основе кластерного подхода // Транспорт Урала. 2017. № 2. С. 36–60.
3. Михалевич В.С., Трубин В.А., Шор Н.З. Оптимизационные задачи производственно-транспортного планирования. М.: Наука, 1986. 260 с.
4. Есипов Б.А., Москвичев О.В., Складнев Н.С., Алешинцев А.О. Разработка и исследование алгоритма кластеризации с проекцией для решения задач оптимизации транспортной инфраструктуры // Перспективные информационные технологии: тр. Междунар. науч.-технич. конф. Самара: Изд-во СамНЦ РАН, 2017. С. 633–637.
5. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика, 1988. 177 с.
6. Прикладная статистика. Классификация и снижение размерности; [под ред. С.А. Айвазяна]. М.: Финансы и статистика, 1989. 607 с.
7. Миркин Б.Г. Методы кластер-анализа для поддержки принятия решений: обзор: препр. WP7/2011/03. М.: Изд-во ВШЭ, 2011. 88 с.
8. Бериков В.Б., Лбов Г.С. Современные тенденции в кластерном анализе. Н.: Изд-во Ин-та матем. им. С.Л. Соболева СО РАН, 2009. С. 16–21.
9. MacQueen J. Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. on Math. Statistics and Probability, 1967, pp. 281–297.
10. Arthur D. and Vassilvitskii S. How Slow is the k-means Method? Proc. 2006 Symp. on Comp. Geometry (SoCG), ACM Press., pp. 144–153.

**SOFTWARE TO SOLVE THE PROBLEMS OF OPTIMIZING TRANSPORT OBJECTS PLACEMENT
BASED ON THE ALGORITHM OF PROJECTION CLUSTERING****B.A. Esipov**¹, Ph.D. (Engineering), Associate Professor, bobpereira@yandex.ru¹ Samara National Research University, Moskovskoe Highway 34, Samara, 443086, Russian Federation

Abstract. The choice of the logistics centers optimal placing is an important task for all types of transportations. Solving it for real problems of a region or country leads to mathematical algorithms of high complexity.

The paper proposes a mathematical model and a new method for solving the problem of logistics centers optimal placement of a two-tier transportation network based on the mathematical apparatus of a cluster analysis. As geoinformation parameters of supplier plants, as well as a road topology, stations or other objects of transport infrastructure are given, there is the problem of an optimal choice of logistics centers. For example, logistics centers for a railways network can be container points. The optimization criterion is minimization of the total traffic volume in ton-kilometers from production to container points. For this purpose, the model of partitioning objects into clusters is used as an optimization mathematical model. The required clusters are subsets of production points with their own centers (container points). Since cluster centers must be located at railway stations, the article suggests a new clustering algorithm with a projection.

The paper investigates the possibilities of such clustering algorithm called k-means pro. It also considers the methodology of selecting the number of centers (clusters) by a generalized economic indicator of transportation costs and costs for creating logistics centers. Based on the created software, there are some examples of calculations for enterprises and railways of the Volga Federal District.

Keywords: two-tier network container transportations, logistic centers, projection clustering, projection defect.

References

1. Kirillova A.G. Optimal choice of locating a terminal as a logistics center when organizing container and piggyback shipments. *Transport: nauka, tekhnika, upravlenie* [Transport: Science, Technology, Management]. 2010, no. 10, pp. 22–25 (in Russ.).
2. Moskvichev O.V. Models, methods and algorithms for optimizing the container transport system of railway transport based on the cluster approach. *Transport Urala* [Transport of the Urals]. 2017, no. 2, pp. 36–60 (in Russ.).
3. Mikhalevich V.S., Trubin V.A., Shor N.Z. *Optimizatsionnye zadachi proizvodstvenno-transportnogo planirovaniya* [Optimization Problems of Production-Transportation Planning]. Moscow, Nauka Publ., 1986, 260 p.
4. Esipov B.A., Moskvichev O.V., Skladnev N.S., Alyoshintsev A.O. Development and investigation of a clustering algorithm with a projection for solving the problems of transport infrastructure optimization. *Perspektivnye informatsionnye tekhnologii: tr. Mezhdunar. nauch.-tekhnich. konf.* [Perspective Information Technologies: Proc. Int. Science and Tech. Conf.]. Samara, SSC RAS Publ., 2017, pp. 633–637 (in Russ.).
5. Mandel I.D. *Klasterny analiz* [Cluster Analysis]. Moscow, Finansy i statistika Publ., 1988, 177 p.
6. *Prikladnaya statistika. Klassifikatsiya i snizhenie razmernosti* [Applied Statistics. Classification and Dimensionality Reduction]. S.A. Ayvazyan (Ed.). Moscow, Finansy i statistika Publ., 1989, 607 p.
7. Mirkin B.G. *Metody klaster-analiza dlya podderzhki prinyatiya resheny: obzor: prepr. WP7/2011/03* [Cluster Analysis Methods for Decision Support: Overview: Preprint WP7/2011/03]. Moscow, VShE Publ., 2011, 88 p.
8. Berikov V.B., Lbov G.S. *Sovremennye tendentsii v klasternom analize* [Modern Trends in cluster Analysis]. Sobolev Institute of Mathematics Publ., 2009, pp. 16–21 (in Russ.).
9. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symp. on Math. Statistics and Probability*. 1967, pp. 281–297.
10. Arthur D., Vassilvitskii Se. How Slow is the k-means Method? *Proc. 2006 Symp. on Computational Geometry (SoCG)*. ACM Press., 2006.