



2. Pitas I., Venetsanopoulos A. Nonlinear Digital Filters: Principles and Applications. – Boston, MA: Kluwer, 1990.
3. Методы компьютерной обработки изображений / М.В. Гашников и др.; под ред. В.А. Сойфера. – Изд. 2-е, испр. – М.: Физматлит, 2003. – 784 с.
4. Ren, X. Learning a classification model for segmentation / X. Ren, J. Malik // Proceedings of the Ninth IEEE International Conference on Computer Vision in ICCV. – 2003. – P. 10-17.
5. Achanta, R. SLIC Superpixels compared to state-of-the-art superpixel methods / R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2012. – Vol. 34 (11). – P. 2274-2282.
6. Stutz, D. Superpixels: An evaluation of the state-of-the-art / D. Stutz, A. Hermans, B. Leibe // Computer Vision and Image Understanding. – 2018. – Vol. 166. – P. 1-27.
7. Сергеев, В.В. Имитационная модель изображения и метод сжатия данных / В.В. Сергеев, В.А. Сойфер // Автоматика и вычислительная техника. – 1978. – № 3. – С.76-78.
8. Егорова, А.А. Исследование эффективности суперпиксельной фильтрации изображений / А.А. Егорова // Сборник трудов ИТНТ-2020. – Самара: Самарский университет, 2020. – Т. 2. – С. 32-39.

Б.А. Есипов

## АЛГОРИТМ КЛАСТЕРИЗАЦИИ С ПРОЕКЦИЕЙ НА ОСНОВЕ МАТРИЦЫ БЛИЗОСТИ

(Самарский университет)

Во многих задачах практики приходится решать задачи оптимального размещения «центров» обслуживания «объектов». Например, это выбор мест расположения контейнерных пунктов для промышленных предприятий, размещение камер видеонаблюдения, организация центров ремонта, станций скорой помощи и многое другое. В качестве математической модели решения задач оптимизации выбора центров обслуживания можно применять известный аппарат кластерного анализа. Особенностью перечисленных примеров задач является то, что выбираемые «центры», должны находиться не в произвольных точках местности, а в определенных местах (на ж/д станциях, перекрестках дорог или на шоссе, или просто заранее заданных удобных точках пространства). Для решения таких задач разработан и изучен алгоритм кластеризации с проекцией *k-means pro*, являющийся модификацией широко известного алгоритма *k-means* (к-средних) [1,2]. В качестве реального примера рассмотрена задача выбора мест расположения контейнерных пунктов для предприятий Приволжского федерального округа так, чтобы центры кластеров находились не в любой географической точке, а обязательно в одной из ж/д станций. Для этого в алгоритме *k-*



*means pro* на каждой итерации получаемый геометрический центр проектируется на ближайшую станцию [2].

Кластеризуемыми могут быть как объекты, так и параметры, причем в качестве меры близости могут быть использованы любые меры типа расстояния. Известно, что меры различия кластеризуемых объектов накладывают жесткие ограничения на применяемые методы кластерного анализа [1].

Несмотря на большие достоинства кластерного подхода и разработанного алгоритма *k-means pro*: математическую обоснованность оптимальности, невысокий рост сложности вычислений при больших размерностях задач, - в предложенной модели есть недостаток, затрудняющий ее применение на практике. Целевая функция при оптимизации использует меру близости как евклидово расстояние между двумя точками  $i$  и  $j$

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Фактическое расстояние между реальными точками-производств и точками-станциями не совпадает с евклидовым, т.к. не учитывает извилистость дорог, наличие рек, мостов, загруженность дорог, необходимость повышенной скорости доставки и другое. Применение других метрик, в частности метрики «расстояние городских кварталов» (Манхэттенское расстояние) не подходит для шоссейных и железнодорожных перевозок. Более того в других прикладных задачах могут быть совершенно другие по природе меры близости объектов и их центров.

В настоящей работе развивается модель и предлагаются алгоритмы решения, устраняющие этот недостаток.

Будем считать, что кластеризуемые объекты в общем случае определяются множеством точек  $I$  со своими номерами  $i=(1,m)$  и весами этих точек  $v_i$ . Задано множество точек проекций  $J$  с номерами  $j=(1,n)$ . Введем в рассмотрение матрицу близости  $D = \{d_{ij}\}$ . Каждый элемент этой матрицы выражает величину того критерия, которым измеряются потери, если для объекта  $i$  назначается центр  $j$ . В простейшем случае это могут быть расстояния в километрах дорог, но можно учесть и все многообразие факторов, влияющих на издержки при назначении  $i$  на  $j$  (качественные характеристики пути, нелинейность зависимости от объема, время доставки, качество видеонаблюдения и другое). Далее мы будем называть величину  $d_{ij}$  по традиции «расстоянием».

Тогда задача выбора мест расположения центров ставится так: найти такие непересекающиеся подмножества точек множества  $I$  – кластеры  $S_l$ , чтобы суммарное взвешенное расстояние от объектов до своих центров кластеров, принадлежащих множеству проекций  $J$  было минимальным.

$$F = \sum_{l=1}^K \sum_{i \in S_l} d_{il} v_i \rightarrow \min$$

Где  $d_{il}$  - расстояние между  $i$ -ой точкой объекта, входящей в  $l$ -й кластер и одной из точек проекций, являющейся центром  $l$ -го кластера. Здесь величина  $F$  зависит от разбиения на кластеры  $S_l$ .



Рассмотрим алгоритм кластеризации на основе матрицы расстояний  $D=\{d_{ij}\}$  при каждом фиксированном  $K$ . Исходные данные для работы алгоритма: множество точек объектов  $I, i=(1,m)$ ; множество номеров точек проекций  $J, j=(1,n)$ ; матрица расстояний  $D=\{d_{ij}\}$ , «веса» точек объектов  $v_i$ , матрица расстояний между точками проекциями  $B=\{b_{jj'}\}$ .

1. Выбираем  $K$  проекций и объявляем их центрами на первой итерации – эталоном. Выбор этого первого эталона, строго говоря, влияет на результат работы алгоритма. В принципе, как это делалось в [3], можно выбрать состав эталона случайным образом. (Ниже будет рассмотрен более совершенный метод построения эталона).

2. Выбираем каждый номер  $i$  и находим центр из множества центров до которого расстояние  $d_{ii}$  минимально. Таким образом привязываем все точки-производства к своим центрам и получаем кластеры  $S_l, l=(1,K)$ .

3. Для каждого кластера  $l$  находим новый центр, для чего в каждом кластере для  $i \in S_l$  находим точку из множества  $J$ , для которой суммарное взвешенное расстояние  $D_j^l$  будет минимально

$$D_j^l = \sum_{i \in S_l} d_{ij} v_i; l = (1, \dots, K)$$

4. Получив новые центры, переходим к пункту 2 и получаем новые кластеры и т.д. повторяем процедуры 2,3,4 до тех пор, пока полученные кластеры и их центры начнут повторяться.

Заметим, что строго говоря, чтобы матрица меры близости  $D$  была бы метрикой, необходимо выполнение аксиом симметрии, неравенства треугольника, различимости нетождественных и неразличимости тождественных объектов.

Проведены эксперименты с матрицами, для которых указанные условия частично не выполняются. Полученные результаты говорят об эффективной работе алгоритма даже при несимметрии и невыполнении неравенства треугольника.

### Алгоритм получения первого эталона

Для улучшения сходимости можно применить следующий алгоритм получения первого эталона. Идея его заключается в том, что более планомерное «движение» кластеров к оптимальному варианту будет, если начальные центры будут наиболее сильно разнесены друг от друга.

1. Выбираем случайный номер из множества номеров точек проекций  $J$  и находим другую точку, наиболее удаленную от первой по величине  $d_{jj}$ , эти две точки проекции уже являются элементами начального эталона.

2. Для каждой точки проекции, не входящей в эталон, по величине  $d_{jj}$  находим ближайшую точку из эталона. Получаем два кластера проекций.

3. Находим в каждом таком кластере самую удаленную точку проекцию от элемента эталона и включаем в эталон ту точку, у которой расстояние до точки

-эталона в кластере наибольшая. Получаем третью точку эталона.



4. Переходим к п.2,3 и так получаем 4-ю, 5-ю и т.д. К-ю точку эталона.

В среднем количество итераций основного алгоритма кластеризации при применении алгоритма получения первого эталона уменьшается более чем в два раза.

### Литература

1. Айвязан, С.А. Прикладная статистика: классификация и снижение размерности: справочное издание / С.А.Айвазян, В.М.Бухштабер, И.С.Енюков, Л.Д.Мешалкин; под ред. С.А.Айвазяна. – Москва: Финансы и статистика, 1989. – 607 с.

2. Есипов, Б.А. Программные средства решения задач оптимизации размещения транспортных объектов на основе алгоритма кластеризации с проекцией. Программные продукты и системы, 2018. Т. 31. № 1. С. 561–566.

3. Есипов, Б.А., Москвичев 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.

4. 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 - Т. 16, № 4. - 2017.

Н.В. Жиркович

## РАЗРАБОТКА ВИРТУАЛЬНОЙ ПЛАТФОРМЫ ДЛЯ ПУБЛИЧНЫХ ОБСУЖДЕНИЙ ПРОЕКТОВ СТРОЯЩИХСЯ ЖИЛЫХ КОМПЛЕКСОВ

(Самарский университет)

Улучшение качества условий проживания россиян, их непосредственное участие в определении вектора развития городов, является одной из государственных инициатив в рамках нацпроектов, принятых на период 2018-2024 гг. Одним из вариантов участия населения в развитие городских систем является создание виртуальной платформы для публичных обсуждений планируемых проектов. Применение информационных технологий в процессе участия населения в развитии городских систем предоставляет возможность организации взаимодействия населения и органов, отвечающих за развитие городской среды, в едином информационном пространстве.

Интерес горожан к вопросам урбанистики постоянно растет, и в интересах обеих сторон стремиться к городу, удобному и комфортному для жителей и гостей Самары, где каждый человек может реализовать свои образовательные, культурные и духовные запросы в полной мере. Проявление активной гражданской позиции горожан в вопросе развития города Самары поможет выработке