

## КЛАСТЕРИЗАЦИЯ КАК ИНСТРУМЕНТ АНАЛИЗА УЛИЧНО-ДОРОЖНОЙ СЕТИ МЕГАПОЛИСА

Михеева Т.И.<sup>1</sup>, Сапрыкина О.В.<sup>1</sup>

<sup>1</sup>ФГАОУ ВО «Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет)», Самара, Россия (443086, Самара, Московское шоссе, 34), e-mail: Olga Grineva @mail.ru

Проведен кластерный анализ участков улично-дорожной сети, основанный на методе Уорда на примере города Самара. Применение кластерного анализа при проектировании транспортной инфраструктуры позволяет выявить участки улично-дорожной сети с высокой интенсивностью транспортного потока, геометрические параметры которых требуют модификации. При проведении анализа рассматривались следующие параметры кластеризации: количество полос движения участка, длина магистрали, интенсивность транспортного потока на участке, мощности точек тяготения в зоне нормативной транспортной доступности участка, мощности объектов рекреации, параметр загруженности транспортных развязок. В результате исследования сделаны выводы об эффективности применения кластерного анализа на этапе предварительного проектирования транспортной сети и для классификации участков улично-дорожной сети, согласно их геометрическим параметрам. Использование различных параметров при кластеризации дает возможность выделить кластеры с участками улично-дорожной сети, где необходимо расширения ширины дорожного полотна по причине высокой интенсивности транспортного потока, кластеры с участками, использующимися с целью рекреации, а также зоны урбанизированной территории, подверженные наиболее интенсивной транспортной нагрузке. Данные, полученные в результате исследования, используются при оптимизации конфигурации улично-дорожной сети.

Ключевые слова: кластер, параметр кластеризации, интенсивность транспортного потока, транспортная доступность, точка тяготения.

## CLUSTERING AS A TOOL FOR ANALYSIS OF MEGAPOLICE ROAD NETWORK

Mikheeva T.I.<sup>1</sup>, Saprykina O.V.<sup>1</sup>

<sup>1</sup>Samara State Aerospace University, Samara, Russia (443086, Samara, Moskovskoe shosse, 34), e-mail: Olga Grineva @mail.ru

We have done cluster analysis for the sections of the road network of Samara, based on the method of Ward. Use of cluster analysis in the design of transport infrastructure reveals sections of the road network, which has high intensity of the traffic flow and its geometrical parameters need to be modified. In the analysis, the following clustering parameters were considered: number lanes of road, highway length, the intensity of the traffic flow in the section of the road, power of gravity points in the area of transport accessibility around the road, power of objects recreational facilities, boot parameter for transport interchanges. The research results show the effectiveness of cluster analysis on the pre-design stage of the transport network and possible to classify sections of the road network, according to their geometrical parameters. The use of different parameters for clustering make possible to identify clusters with sections of the road network, where it is necessary to expand the width of the roadway due to the high intensity of the traffic flow, clusters with sections used for the purpose of recreation, as well as identify intensive urban areas. The data obtained from study are used for the optimization of the configuration of the road network.

Keywords: cluster, clustering parameter, the intensity of the traffic flow, accessibility, gravity point.

Проектирование и реконструкция конфигурации улично-дорожной сети (УДС) мегаполиса требует проведения фазы анализа во избежание излишних некомпенсируемых затрат в случае ошибок. Анализуются участки улично-дорожной сети для выявления узких мест сети. Использование методов кластеризации обуславливается отсутствием априорных гипотез относительно классов, которые необходимо получить в результате.

Целью кластеризации является объединение объектов в классы на основе максимального сходства. Каждый объект представляется вектором из  $n$  значений параметров и

рассматривается как точка в  $n$ -мерном пространстве. За меру сходства двух объектов принимается функция расстояния между ними в этом пространстве.

### Задача кластерного анализа

Задача кластеризации заключается в следующем: пусть имеется множество объектов  $S_n = \{S_1, \dots, S_n\} \subset S$  и функция расстояния между объектами  $\rho(s_i, s'_j)$ . Требуется разбить выборку на непересекающиеся подмножества, называемые кластерами, таким образом, что каждый кластер состоит из объектов, близких по метрике  $\rho$ , а объекты разных кластеров существенно отличаются. При этом каждому объекту  $S_i \in S_n$  приписывается номер кластера  $n_k$ . За алгоритм кластеризации принимается функция  $f: S \rightarrow K$ , которая любому  $S_i \in S_n$  ставит в соответствие номер кластера  $n_k \in K$ .

Поскольку число кластеров для задачи анализа участков улично-дорожной сети заранее неизвестно, то для проведения кластерного анализа выбираем иерархический агломеративный метод Уорда. Метод является наилучшим по результатам экспериментального сравнения на представительном наборе модельных задач и дает возможность получения кластеров с минимальной внутрикластерной вариацией [5].

Алгоритм метода Уорда направлен на объединение близко расположенных кластеров и заключается в следующем: первоначально каждый объект формирует свой собственный кластер, а затем на каждом шаге алгоритма объединяются такие два кластера, которые приводят к минимальному приращению функции расстояния между кластерами.

Алгоритм кластеризации на основе иерархического агломеративного метода Уорда состоит из следующих шагов [1]:

*Шаг 1.* Создать первичный набор кластеров, каждый из которых содержит один элемент набора кластеризации  $K_t = \{\{S_1\}, \dots, \{S_n\}\}$ ,  $t=1$ , где  $S_i$  – объект кластеризации,  $t$  – номер итерации;

*Шаг 2.* Для всех  $t=2, \dots, n$ , где  $t$  – номер итерации:

*Шаг 3.* найти в  $K_{t-1}$  два ближайших кластера  $U, V$

$$(U, V) := \arg \min_{U \neq V} R(U, V), \quad R_t = R(U, V),$$

где  $R(U, V)$  – расстояние между кластерами  $U, V$ ;

*Шаг 4.* Изъять кластеры  $U$  и  $V$ , добавить слитый кластер  $W = U \cup V$ :  $K_t = K_{t-1} \cup \{W\} \setminus \{U, V\}$ ;

*Шаг 5.* Для всех  $S_i \in K_t$  вычислить расстояние  $R(W, S)$ .

## Параметры исследования

В анализе улично-дорожной сети кластеризация применяется как средство выявления участков сети, требующих модификации. При решении задач проектирования развития мегаполиса улично-дорожную сеть резонно представлять в виде многослойной карты с географической привязкой объектов городской и транспортной инфраструктуры. В связи с этим, в качестве метода кластеризации выбран алгоритм Уорда [2], как наиболее применимый к кластеризации географических объектов.

За объект кластеризации принимаем участок улично-дорожной сети  $S_i$ . Количество кластеров, которые необходимо получить заранее не известно, а максимальное количество кластеров устанавливаем равным пяти. В качестве функции расстояния, для определения близости объектов (участков УДС), выбрана классическая евклидова метрика [3]:

$$\rho(s_i, s_j) = \sum_{p=1}^{n_p} (\mu_k(s_i) - \mu_k(s_j))^2, \quad (1)$$

где  $\mu_k(s_i), \mu_k(s_j)$  – меры обладания  $p$ -м параметром  $i$ -го и  $j$ -го участков УДС, соответственно.

Эмпирическим путем было выявлено, что добавление избыточного количества параметров кластеризации (более пяти) даёт некорректные результаты, не учитывающие в полной мере все характеристики улично-дорожной сети. Поэтому рационально проводить кластеризацию по нескольким видам, каждый из которых содержит собственный набор входных параметров. Проведено четыре вида кластеризации:

- по интенсивности транспортных потоков участков УДС;
- по мощности точек тяготения урбанизированной территории;
- по площади рекреационных зон урбанизированной территории;
- по загрузки транспортных узлов.

Входные параметры разбиты на два класса – базовые и дополнительные. Базовые параметры использованы в каждом виде кластерного анализа. В кластеризации по интенсивности транспортных потоков используется только базовый набор параметров. Для остальных видов кластеризации, кроме базового, используется один из дополнительных параметров.

В качестве базовых параметров выбраны наиболее важные количественные характеристики участка улично-дорожной сети  $S_i$ :

- количество полос движения  $q^r(S_i)$ ;
- длина магистрали улично-дорожной сети  $L(S_i)$  на которой дислоцируется участок;

- интенсивность транспортного потока  $I(S_i)$  на участке.  
Дополнительными входными параметрами решения задачи кластеризации являются:
- количество точек тяготения  $q_{fp,k}^p(S_i)$  в зоне нормативной шаговой доступности от участка УДС с указанием мощности  $f_{p,k}$  точки тяготения;
- количество объектов тяготения  $q_{fo,k}^o(S_i)$ , имеющих социальную привлекательность в выходные и праздничные дни, в зоне транспортной доступности от участка улично-дорожной сети  $S_i$  с указанием мощности  $f_{o,k}$  объекта рекреации;
- количество транспортных развязок  $q_{pk}^c(S_i)$  с указанием мощности загруженности  $P_k$  в зоне нормативной транспортной доступности от участка улично-дорожной сети  $S_i$ .

### Результаты исследований

В результате применения метода кластерного анализа Уорда к улично-дорожной сети города Самара с использованием различных параметров получены результаты, представленные на карте города Самара (рисунки 1-4). Кластеризация по интенсивности транспортного потока участков УДС с использованием набора базовых параметров приведена на рисунке 1.

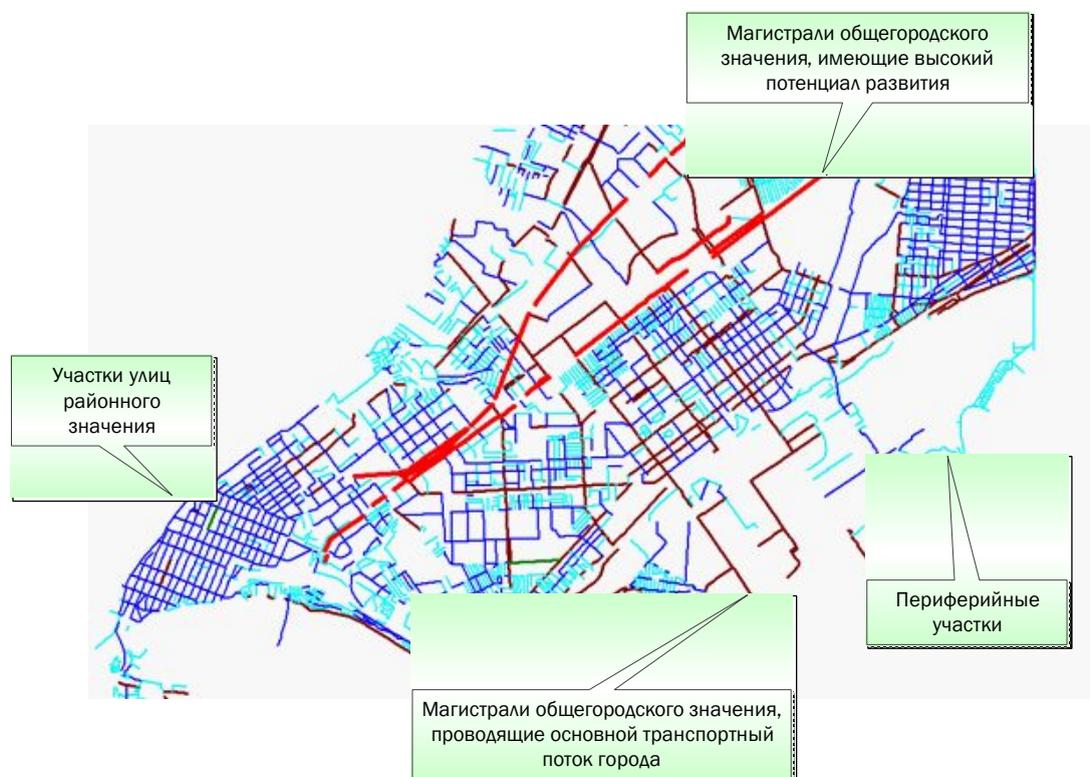


Рис. 1. Кластеры участков УДС города при использовании базовых параметров

В результате данного вида кластеризации были получены 4 кластера. Анализ полученных кластеров позволил сделать вывод об их содержании:

**кластер 1 ( )** - магистрали общегородского значения, имеющие высокий потенциал развития. К данному кластеру отнесены такие улицы как Московское шоссе и проспект Карла Маркса;

**кластер 2 ( )** - магистрали общегородского значения, проводящие основной транспортный поток города, предназначенные для транспортной связи жилых и промышленных районов города с объектами общегородского значения. К улицам, данного кластера относятся проспект Кирова, проспект Metallургов, ул. Гагарина и др.;

**кластер 3 ( )** - городские улицы районного значения, предназначенные для транспортной связи;

**кластер 4 ( )** - внутридворовые проезды, предназначенные для связи отдельных групп зданий между собой.

Для выявления участков улиц, имеющих наибольшую нагрузку в выходные и праздничные дни, к базовым параметрам, добавлен параметр мощности объектов тяготения, расположенных в зоне транспортной доступности от участка УДС и имеющих рекреационное и культурно-развлекательное назначение. Результаты данного вида кластеризации представлены на рисунке 2.

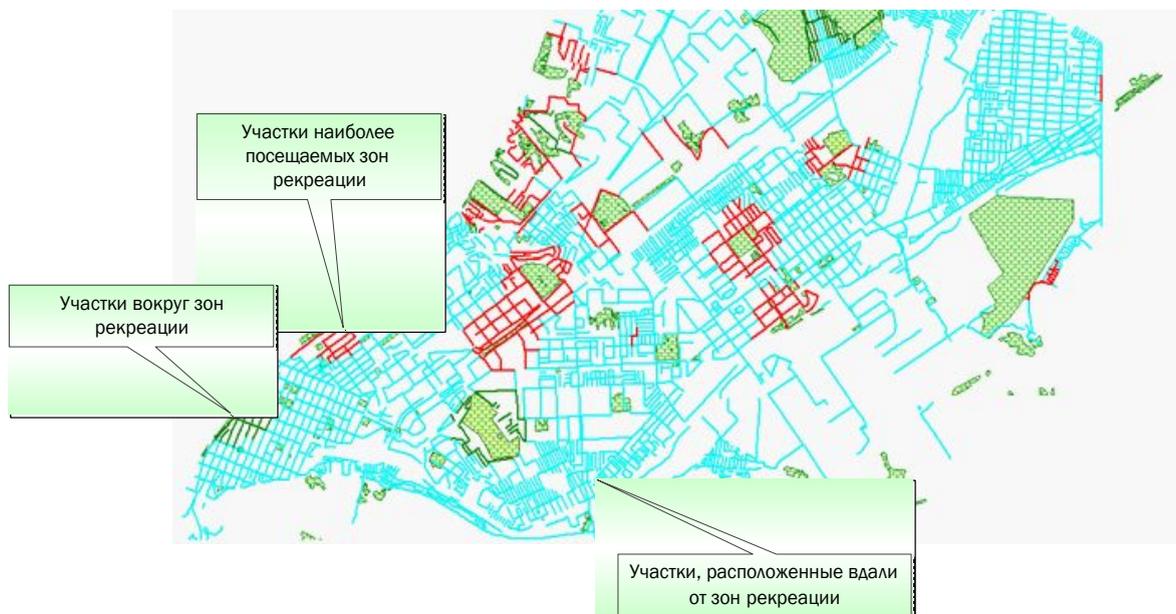


Рис. 2. Кластеры участков при использовании параметра мощностей зон рекреации

При кластеризации с добавлением параметра мощностей зон рекреации получено три кластера, функциональные значения которых:

**кластер 1** ( ) – участки УДС вокруг наиболее посещаемых парков города и набережной реки Волга. Примером служит - Волжский проспект, Московское шоссе в районе парка им. Ю. Гагарина, ул. Ново-Садовая в районе Загородного парка, ул. Самарская в районе площади Славы и др.;

**кластер 2 ( )** - участки УДС вокруг часто посещаемых скверов. К данному кластеру отнесены участки вокруг первой очереди набережной, а также городское кладбище;

**кластер 3 ( )** - участки, не используемые активно в выходные и праздничные дни.

С целью выявления наиболее развитых областей города, обладающих значительной численностью населения и наиболее плотным расположением социальной инфраструктуры проведена кластеризация с добавлением параметра мощности точек тяготения. Точкой тяготения является объект урбанизированной территории, несущий социальную функцию [4]. Примером точек тяготения могут служить места приложения труда, крупные торговые центры, больницы и поликлиники, и т.д. Данные о точках тяготения взяты из электронного справочника организаций Самарской области. Для каждого подобного объекта учтен фактор значимости, исходя из информации о площади объекта, посещаемости объекта и его масштаба активности (региональный, областной и т.д.) Результаты кластерного анализа с учетом параметра мощности точек тяготения представлены на рисунке 3.

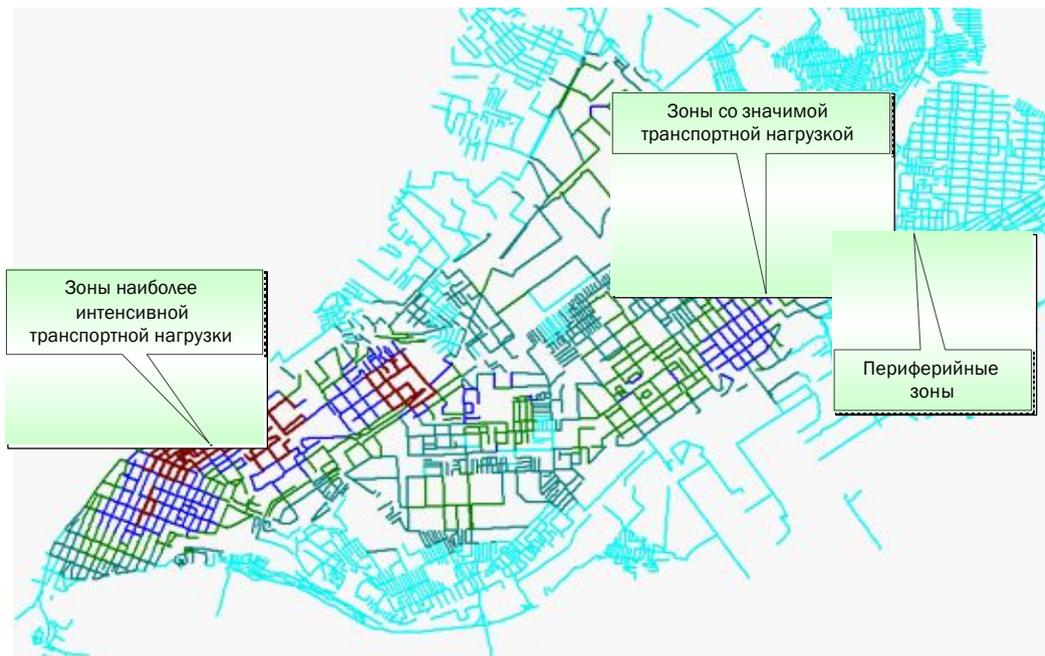


Рис. 3. Кластерный анализ участков УДС с параметром мощности точек тяготения для выявления наиболее развитых зон города

В результате кластеризации с параметром мощности точек тяготения получены четыре

кластера. В первый кластер ( ) попали участки УДС, входящие в зоны наиболее интенсивной транспортной нагрузки. К таким зонам отнесены географический центр города (Московское шоссе в районе Аэрокосмического университета) и исторический центр города

(ул. Самарская, ул. Молодогвардейская и т.д.). Второй кластер ( ) характеризует районы города со значимой транспортной нагрузкой. К данному кластеру отнесены участки улиц в историческом центре города (ул. Льва Толстого, ул. Алексея Толстого и т.д.), в географическом центре города (просп. Масленникова, ул. Мичурина и т.д.), в районе металлургического завода (пр. Metallургов и т.д.), в районе торгового центра «Мега» (Московское шоссе), в 116-ом микрорайоне (Пугачевский тракт и т.д.). В третий и четвертый кластер отнесены периферийные участки улиц спальных районов города.

Данные для кластерного анализа на основе загрузки транспортных узлов собраны путем телефонного социологического опроса населения Самарской области в 2011 году. Выборка составляла – 10 000 человек. Анкетиремым предлагалось назвать пятерку наиболее загруженных транспортных развязок города. Графическая интерпретация результата кластеризации представлена на рисунке 4.

Для этого вида кластерного анализа представлены графики зависимостей входных параметров кластеризации. Кластеризация с использованием данных о наиболее загруженных транспортных развязках выявила четыре кластера:

**кластер 1 ( )** – участки магистралей, имеющие перспективу развития. К кластеру отнесены:  
Московское шоссе и пр. Карла Маркса;

**кластер 2 ( )** - участки, не получившие должного развития и требующие расширения. К кластеру отнесены – ул. Аврора, пр. Кирова, ул. Советской Армии, ул. Ново-Садовая, ул. Гагарина, ул. Куйбышева, ул. Самарская и т.д.;

**кластер 3 и 4** - периферийные улицы.



Рис. 4. Кластерный анализ с использованием данных о дислокации транспортных узлов УДС

### Заключение

Анализ результатов кластеризации участков улично-дорожной сети города Самара на основе различных входных параметров позволил сделать основные выводы:

- кластеризация участков города по параметру интенсивности движения транспортных потоков доказывает острую необходимость расширения проспекта Карла Маркса и Московского шоссе;
- при кластерном анализе с использованием параметра площади рекреационных зон определяется необходимость развития улиц, прилегающих к зеленым зонам города;
- зоны наиболее интенсивной и значимой транспортной нагрузки, а также участки УДС, выделенные в первый и второй кластер при кластеризации с параметром мощности точек тяготения, необходимо связать магистралями общегородского значения;
- к основным магистралям города следует отнести все участки УДС, полученные во втором кластере при кластеризации с учетом данных о крупнейших транспортных узлах города, как наиболее важные связи города, соединяющие основные развязки.

Полученные в исследовании выводы используются в работе при оптимизации улично-дорожной сети города Самара.

### Список литературы

1. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод - СПб.: БХВ-Петербург, 2004. - 336с.
2. Дюран Б. Кластерный анализ. Пер. с англ. Е.З. Демиденко / Б. Дюран, П. Оделл– М.:

Статистика, 1977. – 128 с.

3. Мандель И. Д. Кластерный анализ. — М.: Финансы и Статистика, 1988. - 176 с.

4. Сапрыкина О.В. К вопросу об определении фокусов тяготения на основе геоинформационной составляющей интеллектуальной транспортной системы / О.В. Сапрыкина, Т.И. Михеева // Актуальные проблемы автотранспортного комплекса: межвуз. сб. науч. статей. – Самара: Самар. гос. техн. ун-т., 2010. – С. 156-160.

5. Lam B. A new cluster validity index for data with merged clusters and different densities/ B. Lam, H. Yan // Systems, Man and Cybernetics: IEEE International Conference. – 2005. – Vol. 1. – P. 798-803.

**Рецензенты:**

Титов Б.А., д.т.н., профессор, заведующий кафедрой организации и управления перевозками на транспорте, ФГАОУ ВО «Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет)», г.Самара.

Кораблин М.А., д.т.н., профессор, декан факультета информационных систем и технологий, заведующий кафедрой вычислительная техника ФГОБУ ВПО «Поволжский государственный университет телекоммуникаций и информатики», г.Самара.