



Кластерный анализ объектов инфраструктуры



Александр КУТЫРКИН
Alexander V. KUTYRKIN

Елена ОВЧИННИКОВА
Elena A. OVCHINNIKOVA



Светлана СУДОРГИНА
Svetlana S. SUDORGINA

*Кутыркин Александр Васильевич – доктор технических наук, профессор Московского государственного университета путей сообщения (МИИТ).
Овчинникова Елена Александровна – ассистент МИИТ.
Судоргина Светлана Сергеевна – технолог Главного вычислительного центра – филиала ОАО «РЖД».*

Методология и алгоритмы кластеризации. Их участие в классификации предметной среды. Кластерный анализ объектов транспортной инфраструктуры на примере железнодорожных вокзальных комплексов. Возможности автоматизации процесса.

Ключевые слова: искусственные нейронные сети, кластерный анализ, топологическая карта Кохонена, железнодорожный вокзальный комплекс.

Для любой научной деятельности кластерный анализ является одной из фундаментальных составляющих ее содержания и методологических особенностей. Исследуемые объекты и явления всегда должны быть упорядочены или сгруппированы по их схожести, то есть классифицированы, прежде чем начнут разрабатываться научные гипотезы и теории, объясняющие их поведение и взаимную связь. В процессе кластерного анализа осуществляется разбиение исследуемого множества объектов, представленных многомерными данными, на группы похожих в определенном смысле объектов, называемых кластерами. Ценность современного кластерного анализа, имеющего мощную компьютерную базу, заключается в том, что он может производить группировку объектов не только по одному или нескольким параметрам, как при субъективной классификации, а и по целому набору признаков, что требует сложных алгоритмов кластеризации [1].

1.

В нашем случае кластерный анализ объектов транспортной инфраструктуры осуществляется на примере железнодорож-

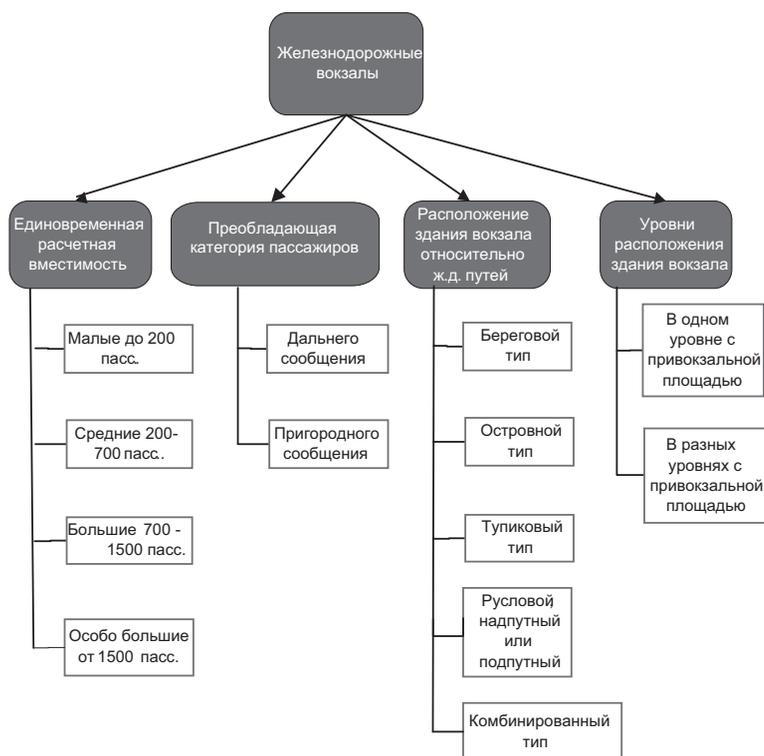


Рис. 1. Существующая классификация железнодорожных вокзалов РФ.

ных вокзальных комплексов, включающих привокзальную площадь, пассажирское здание, посадочные платформы, тоннели, переходные мостики через железнодорожные пути и другие коммуникации, различные малые архитектурные формы для отдыха пассажиров и занятия детей и др.

Начиная с середины XX века железнодорожные вокзалы стали интенсивно развиваться как в функциональном, так и градостроительном плане. Исторически сложившееся месторасположение их зданий в центральной части городов послужило основанием определять вокзал как «центр притяжения» городской инфраструктуры.

Постепенное развитие общественно-деловой функции, не типичной для железнодорожного вокзала, упорядочение пешеходных коммуникаций, освоение надземного, подземного пространства, увязка работы с привокзальной площадью, развитие рекреационных зон привели к преобразованию вокзала как объекта транспортной инфраструктуры в вокзальный комплекс в составе городских транспортных узлов.

Требования, предъявляемые к железнодорожному вокзальному комплексу,

не просто существенно изменились, появилась насущная потребность в инновационных технологиях, новых методах управления, разработке проектов комплексного развития вокзала и прилегающих к нему территорий и т. д. [2].

На сегодняшний день существует классификация железнодорожных вокзалов, созданная в середине XX века (рис. 1).

Проанализировав общепринятую классификацию, можно убедиться, что она не отражает в полной мере изменения, произошедшие за последнее время в данной сфере.

В связи с этим есть необходимость выделить требующие нового осмысления недостатки и достоинства сложившейся классификации.

Недостатки:

- не дает полной картины о всем многообразии вокзальных комплексов и их функций;

- строилась исключительно с помощью эвристических методов, не позволяющих охватить все скрытые особенности вокзалов;

- не несет в себе достаточно значимой информации для глубокого изучения про-





цессов формирования вокзальных комплексов;

– не применима для принятия комплексных решений по развитию и модернизации тех или иных групп вокзалов;

– не учитывает изменившиеся требования к структуре и функциям вокзального комплекса;

– не принимает во внимание прогнозы развития той городской среды, в которой находится вокзал.

Достоинства:

– проста для понимания;

– наглядна;

– удобна для первичного, самого общего ознакомления со структурой железнодорожных вокзалов и их отдельных элементов.

При таком раскладе оценок закономерно считать, что существующая классификация железнодорожных вокзалов не отвечает реальному состоянию исследуемого объекта и тенденциям реформирования вокзалов в вокзальные комплексы. Основываясь на выделенных недостатках классификации и учитывая дефицит разработанности темы в научной литературе, правомерно сделать вывод об актуальности постановки задачи, которая предусматривала бы автоматизацию кластерного анализа вокзальных комплексов.

Важно отметить, что такая задача ставится впервые и она не имеет аналогов в транспортной сфере.

2.

Первый шаг к созданию условий для автоматизированного кластерного анализа — целевое программное приложение с заданными функциями.

Разработанное в рамках поставленной задачи приложение позволит сформировать адаптированную структуру вокзальных комплексов, которая будет иметь качественно новое значение при подготовке решений, связанных с реконструкцией и строительством интермодальных транспортных узлов. Приложение предполагает разбиение на кластеры вокзалов по различным блокам признаков, характеризующих как функциональные особенности, так и качественные показатели. За счет этого на выходе мы получим такую модель вокзальных комплексов, которая поможет

ликвидировать недостатки существующей классификации и даст возможность преимущественным образом осуществлять техническое и технологическое планирование.

При этом порядке процесс принятия решений значительно упрощается. Достаточно выявить типичного представителя одного из кластеров вокзалов, определить для него пути совершенствования и применить тот же самый рецепт для остальных членов кластерной ячейки. Подобный способ действий позволяет сократить массив объектов, поскольку нет необходимости в изучении, анализе и выборе определенной политики развития в отношении каждого вокзала в отдельности. Приложение вполне годится для первичной обработки данных и накопления информации для дальнейшего более глубокого анализа складывающихся тенденций.

В качестве исходных для функционирования приложения выступают данные по дорогам России, предоставленные дирекцией железнодорожных вокзалов.

В границах кластеризации следует произвести выбор параметров, по которым будут оцениваться объекты — железнодорожные вокзальные комплексы (в нашем распоряжении оказалось 139 таких адресов).

В качестве параметров были выделены следующие блоки:

- вокзал и город {население города, расположение здания вокзала в городской черте, средняя длительность поездки, наличие метрополитена в городе};
- привокзальная площадь {наличие остановок городского вида транспорта, зон парковки, такси, службы проката автотранспорта, рекреационных зон};
- архитектурные особенности здания вокзала {тип здания в плане и его расположение по отношению к железнодорожным путям, тип здания в разрезе и его расположение по отношению к железнодорожным путям, единовременная расчётная вместимость вокзала, является ли вокзал объединённым с другими видами транспорта, площадь вокзальных помещений, разделение здания на основное и пригородное};
- пассажиропоток {количество отправляемых пассажиров в сутки};
- размеры движения {среднесуточное количество пар поездов};

Наименование региональной ДЖВ	БЛОКИ →	ПРИВOKЗАЛЬНАЯ ПЛОЩАДЬ			АРХИТЕКТУРНЫЕ ОСОБЕННОСТИ ЗДАНИЯ ВОКЗАЛА				
		Название вокзала	Наличие остановок городского вида транспорта	Наличие зон парковки, такси, службы проката автомобилей	Наличие рефрежилированных зон	Год постройки здания вокзала год капитальной реконструкции	Тип здания вокзала в плане и его расположение по отношению к ж.д. путям	Тип здания вокзала в разрезе и его расположение по отношению к ж.д. путям	Единовременная расчетная вместимость вокзала, пассажиров
Северо- Западная	Балтийский	да	нет	нет	1857	тушковый	В одном уровне с привокзальной площадью	2200	нет
	Витебский	да	да	нет	1904	тушковый	многоуровневый	2000	нет
	Калининград - Южный	нет	да	нет	1929	береговой	многоуровневый	1650	нет
	Ленинградский	да	да	да	1849	тушковый	В одном уровне с привокзальной площадью	15900	нет
	Ладужский	да	да	да	2003	руслый	многоуровневый	3400	нет
	Московский	да	нет	нет	1851	тушковый	В одном уровне с привокзальной площадью	19 114	нет
	Финляндский	нет	да	нет	1961	тушковый	В одном уровне с привокзальной площадью	2200	нет

Рис. 2. Фрагмент исходной матрицы данных.

- устройства на территории вокзала {количество пассажирских платформ, пешеходных переходов в разных уровнях через пути, наличие конкорса над путями, турникетов для пригородных пассажиров};
- бесплатные услуги {билетно-кассовое обслуживание, наличие зала ожидания, современное справочно-информационное обслуживание, услуги санитарных комнат, медицинский пункт};
- платные услуги {камеры хранения, услуги связи, залы повышенной комфортности, комнаты длительного отдыха и комнаты матери и ребенка, услуги сервисных центров, услуги носильщиков};
- социальная сфера {адаптивность под определенные потребности помещений вокзала и прилегающих территорий};
- безопасность на вокзале {оборудование помещений охранно-пожарной сигнализацией, системой пожаротушения}.

В общей сложности сформировалось 38 параметров, обоснованных по каждому вокзальному комплексу. Все искомые данные сведены в итоговую таблицу в формате *xlsx*. Фрагмент матрицы подобного типа представлен на рис. 2.

В список вошли все внеклассные вокзалы, вокзалы I класса, выборочно — вокзалы II и III классов.

3.

Кластеризация вокзальных комплексов осуществлялась с помощью самоорганизу-

ющейся карты Т. Кохонена (англ. *Self-organizing map* — SOM), построенной в виде соревновательной нейронной сети с обучением без учителя, выполняющей одновременно задачу кластеризации и визуализации её результатов. Идея сети основана на проецировании многомерного пространства в пространство с более низкой размерностью (двумерное) [3].

Структуру самоорганизующейся карты формируют узлы или нейроны, количество которых задаётся. Каждый из узлов описывается двумя векторами. Первый — вектор веса, имеющий такую же размерность, что и входные данные (W_j). Второй — вектор, представляющий собой координаты узла на карте (r). По координатам узла на карте можно определить его соседей — ближайшие узлы. Веса — это главный элемент модели Кохонена. Веса рассчитываются с помощью итеративной процедуры, в ходе которой параметры карты подстраиваются под наблюдение. После её завершения самоорганизующаяся карта построена, а полученные кластеры изображаются на двухмерном рисунке, что позволяет наглядно и легко интерпретировать структуру кластеров.

Векторы входных сигналов (x) обрабатываются по одному, для каждого из них находится ближайший кодовый вектор «победитель» — $W_{j(x)}$. После этого все кодовые векторы W_i соседних узлов пересчитываются по формуле [3]:



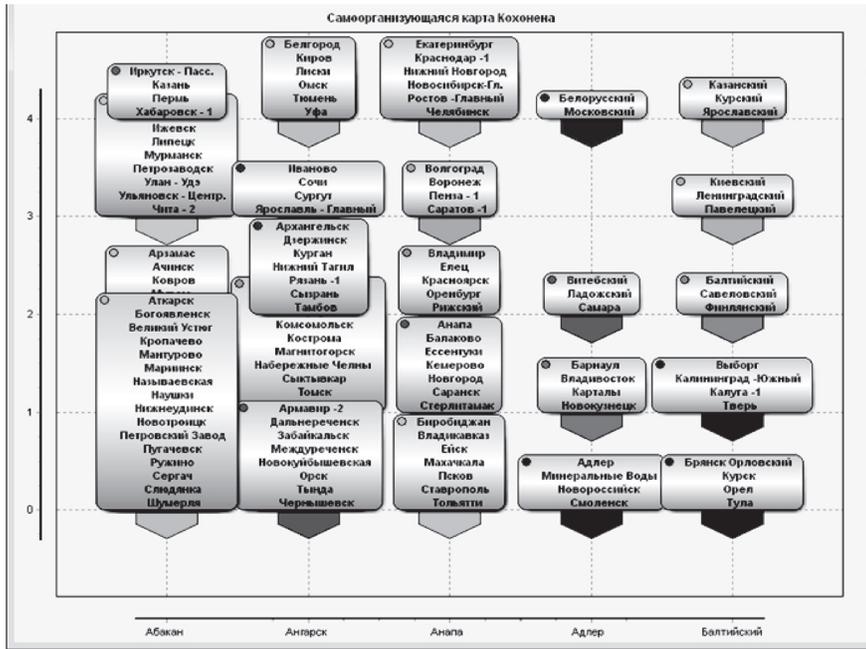


Рис. 3. Построенная на данных вокзальных комплексов топологическая карта Кохонена.

$$W_l^{новый} = W_l^{старый} (1 - \eta_{j(x)l} \theta) + x \eta_{j(x)l} \theta, \quad (1)$$

где $\theta \in (0,1)$ – шаг обучения;

$\eta_{j(x)l}$ – мера «соседства» узлов карты для каждой пары j, l ($j, l = 1, \dots, k$), определено число η_{jl} ($0 \leq \eta_{jl} \leq 1$, при этом диагональные элементы таблицы близости равны единице ($\eta_{jj} = 1$)).

Соседи кодового вектора-победителя (по заранее заданной таблице близости) сдвигаются в ту же сторону, что и этот вектор, пропорционально мере близости.

Чаще всего таблица кодовых векторов представляется в виде фрагмента квадратной решётки на плоскости, а мера близости определяется, исходя из евклидового расстояния на плоскости.

Работа SOM построена по заданному принципу [3, 4]:

- инициализация карты, то есть первоначальное задание векторов веса для узлов;
- организация цикла:

- 1) выбор следующего наблюдения (вектора из множества входных данных);
- 2) нахождение для него лучшей единицы соответствия (best matching unit, BMU, или Winner) – узла на карте, вектор веса которого меньше всего отличается от наблюдения (в метрике, задаваемой аналитиком, чаще всего – евклидовой);

- 3) определение количества соседей и обучение – изменение векторов веса и его соседей с целью их приближения к наблюдению;

- 4) выявление ошибки карты.

Алгоритм кластеризации:

а) Инициализация

Возможны следующие способы задания первоначальных весов узлов:

- задание всех координат случайными числами;
- присваивание вектору веса значение случайного наблюдения из входных данных.

б) Цикл

Пусть t – номер итерации (инициализация соответствует номеру 0) необходимо:

- 1) выбрать произвольное наблюдение $x(t)$ из множества входных данных;

2) найти расстояния от него до векторов веса всех узлов карты и определить ближайший по весу узел $M_c(t)$. Это – BMU или Winner. Условие на $M_c(t)$:

$$x(t) - m_c(t) \leq x(t) - m_i(t)$$

для любого $m_i(t)$, где $m_i(t)$ – вектор веса узла $M_i(t)$. Если находится несколько узлов, удовлетворяющих условию, BMU выбирается случайным образом среди них;

- 3) определить с помощью функции h (функции соседства) соседей M_c и изменение их векторов веса.

Исходные данные	Результат	Нормализованные данные	Итерации	
Название	X	Y	Кластер	Маркировка
Перь	0	4	22	
Петровский Завод	0	0	9	
Петрозаводск	0	3	8	
Псков	2	0	14	
Пугачевск	0	0	9	
Рижский	2	2	17	
Ростов -Главный	2	4	20	
Ружино	0	0	9	
Рязань -1	1	2	7	
Савеловский	4	2	10	
Самара	3	2	16	
Саранск	2	1	3	
Саратов -1	2	3	18	

Рис. 4. Таблица координат, номер кластера и его цвет.

Функция h выявляет «меру соседства» узлов M_i и M_c и изменение векторов веса. Она должна постепенно уточнять их значения, сначала у большего количества узлов и сильнее, потом у меньшего и слабее. Часто в качестве функции соседства используется гауссовская функция:

$$h_{ci}(t) = \alpha(t) \cdot \exp\left(-\frac{r_c - r_i^2}{2\sigma^2(t)}\right), \quad (2)$$

где $0 < \alpha(t) < 1$ – обучающий множитель, монотонно убывающий с каждой последующей итерацией;

r_c, r_i – координаты узлов $M_i(t)$ и $M_c(t)$

на карте;

$\sigma^2(t)$ – множитель, уменьшающий количество соседей с итерациями, монотонно убывает.

Параметры α, σ и характер их убывания задаются.

Более простой способ задания функции соседства: $h_{ci}(t) = \alpha(t)$, если $M_i(t)$ находится в окрестности $M_c(t)$ заранее заданного аналитиком радиуса, и 0 – в противном случае.

Функция $h(t)$ равна $\alpha(t)$ для ВМУ и уменьшается с удалением от ВМУ.

4) изменить векторов веса по формуле: $m_i(t) = m_i(t-1) + h_{ci}(t) (x(t) - m_i(t-1))$. (3)

Вектора веса всех узлов, являющихся соседями ВМУ, приближаются к рассматриваемому наблюдению.

5) вычислить ошибку карты как среднее арифметическое расстояний между наблюдениями и векторами веса соответствующих им ВМУ:

$$\frac{1}{N} \sum_{i=1}^N x_i - m_c, \quad (4)$$

где N – количество элементов набора входных данных.

4.

Представленная методология самоорганизующихся карт Кохонена [3] была реализована в виде программы кластерного анализа объектов транспортной инфраструктуры.

Программа предусматривает выполнение следующих двух этапов:

1) Подготовка статистически значимых данных.

Подготовка исходных (статистически значимых) данных осуществляется в Excel. Исходные данные представляют собой таблицу, первая строка и столбец которой содержат наименования объектов транспортной инфраструктуры и их характеристики соответственно. Остальные ячейки таблицы заполняются значениями. Таблицы могут содержать как количественную, так и качественную информацию. Перед проведением кластерного анализа требуется провести преобразование качественных признаков, путем их кодирования.

2) Анализ подготовленных данных.

Материалы для анализа, подготовленные согласно требованиям этапа 1, анали-





зируются при помощи нейросетевого метода, основанного на SOM – соревновательной нейронной сети с обучением без учителя, выполняющей задачу визуализации и кластеризации. Анализ производится в программе Neuron.exe.

Кластерный анализ в зависимости от целей исследования можно проводить как по отдельным блокам параметров, так и по всей совокупности. Далее описан пример кластерного анализа железнодорожных вокзальных комплексов по всей совокупности выделенных параметров: 139 вокзалов и 38 параметров.

Программа представляет собой результаты анализа в виде топологической карты Кохонена и соответствующей ей таблицы координат с указанием кластера, к которому относится объект.

При этом схожие объекты будут находиться близко друг от друга на карте и образовывать кластеры. На карте каждый кластер помечен своим цветом и сопровождается списком входящих в него объектов. При размерности карты 5×5 получаем 24 кластера. Построенная топологическая карта Кохонена показана на рис. 3.

Топологическая карта строится по координатам, вычисленным в результате анализа. В таблице координат, показанной на рис. 4, также указаны кластеры, к которым принадлежат объекты, и цвет, которым они обозначены на карте.

Результаты кластерного анализа, данные, полученные в программе Neuron, можно сохранить в файле формата «.xls» и «.xlsx». Данный файл можно использовать для дальнейшего анализа объектов транспортной инфраструктуры.

Таким образом, получив результаты кластеризации вокзальных комплексов, можно сделать выводы о том, что:

- ❖ был проведён разведочный анализ первичных данных (параметров исходной матрицы), формирующих понятие вокзального комплекса;
- ❖ произведена проверка предположений о наличии структуры изучаемой исходной матрицы данных;
- ❖ подготовлены данные для дальнейшего более глубокого анализа структуры вокзальных комплексов;
- ❖ сократился размер анализируемых данных, а следовательно и время исследовательской работы;
- ❖ произведена визуализация полученной структуры.

Полученные в результате исследования данные могут быть подвергнуты дальнейшему исследованию, более глубокому анализу традиционными статистическими и другими методами, могут использоваться для научно-исследовательских работ, при разработке программ стратегического развития объектов транспортной инфраструктуры ОАО «РЖД».

ЛИТЕРАТУРА

1. Кутыркин А. В., Сёмин А. В., Кластерный анализ, – М.: МИИТ, 2009.
2. Концепция эффективного использования и развития железнодорожных вокзалов – филиала ОАО «РЖД» до 2015 года. Утверждена ОАО «РЖД» 17 сентября 2008 г. № 15098.
3. Kohonen T. The self-organizing map, Proceedings of the Institute of Electrical and Electronics Engineers, 1990, vol. 78, p. 1464–1480.
4. Головенко В. А. Нейронные сети: обучение, организация и применение. – Кн. 4: Учеб. пособие для вузов. – М.: ИПРЖР, 2001. ●

CLUSTER ANALYSIS OF INFRASTRUCTURE PROJECTS

Kutyркин, Alexander V. – D.Sc. (Tech), professor of Moscow State University of Railway Engineering (MIIT).
Ovchinnikova, Elena A. – assistant lecturer of Moscow State University of Railway Engineering (MIIT).
Sudorgina, Svetlana S. – production engineer of Main computation center – a subsidiary to JSC Russian Railways.

The authors describe methodology and algorithms of clustering, substantiate their role for classification of an object environment. They analyze transport infrastructure taking an example of railway stations. They also study possibilities of process automation.

Key words: artificial neuronets, cluster analysis, self-organizing map, Kohonen's topology-preserving map, railway station.

Координаты авторов (contact information): Кутыркин А. В. – (499) 241–92–83, Овчинникова Е. А. – bogdanelena@yandex.ru, Судоргина С. С. – lana1911@gmail.com.