

МЕТОДЫ ОПИСАНИЯ ЭКОНОМИЧЕСКОГО ПРОФИЛЯ РЕГИОНА НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА

А. Н. Кисляков

Владимирский филиал Российской академии народного хозяйства
и государственной службы при Президенте Российской Федерации,
Владимир, Москва

Статья посвящена решению актуальной задачи формирования экономического профиля региона в рамках задачи мастер-планирования развития территорий с применением методов пространственно-временной кластеризации данных, характеризующих хозяйственную деятельность бизнес-субъектов. Цель исследования – анализ моделей пространственно-временной кластеризации, применяемых в задачах описания экономического профиля территорий для выявления закономерностей и тенденций в экономических данных, а также разработки рекомендаций по их эффективному использованию для оптимизации управления и стратегического планирования на уровне регионов. Автором приведен пример интеграции нелинейных методов уменьшения размерности и моделей автокодировщиков, позволяющих упростить структуру данных, улучшить визуализацию, сохранить важные характеристики и устранить шум, что в конечном итоге способствует более точному и эффективному выявлению кластеров. Подход может применяться в задачах, где данные имеют сложную структуру и требуют как нелинейного уменьшения размерности, так и сохранения локальных и глобальных свойств. В результате исследования даны рекомендации по расширению инструментария описания экономического профиля региона.

Ключевые слова: стратегическое планирование, модели пространственно-временной регрессии, конкурентоспособность бизнеса.

METHODS OF DESCRIBING REGION ECONOMIC PROFILE ON THE BASIS OF CLUSTER ANALYSIS

Aleksey N. Kislyakov

Vladimir branch of Russian Presidential Academy
of National Economy and Public Administration, Vladimir, Russia

The article deals with the acute objective of drawing-up economic profile of region within the frames of master-planning of territory development with the help of methods of space-time data clusterization that characterize economic activities of business entities. The goal of the research is to analyze models of space-time clusterization used in descriptions of economic profile of territory in order to identify regularities and trends in economic data, as well as to work out recommendations on their effective use to optimize management and strategic planning on regional level. The author provided an example of integration of non-linear methods of reducing size and models of auto encoders, which help simplify data structure, improve visualization, preserve important characteristics and eliminate noise, which can provide more accurate and effective cluster reveal. The approach can be used in tasks where complicated structures are present and which require both non-linear size reduction and keeping local and global features. As a result recommendations were prepared dealing with extension of tools for describing region profile.

Keywords: strategic planning, models of space-time regression, business competitiveness.

Введение



Формирование экономического
профиля территорий как часть
задачи мастер-планирования [15]

становится основным инструментом для
принятия обоснованных решений в обла-
сти стратегического управления развитием

территорий на региональном и муниципальном уровне.

Эффективное использование методов кластерного анализа позволяет выявлять скрытые и неочевидные закономерности и тенденции в данных, характеризующих хозяйственную деятельность бизнес-субъектов по основным видам деятельности с учетом их географического расположения, что в свою очередь способствует более глубокому пониманию динамики развития экономики регионов.

Современные методы и инструменты сбора и анализа данных, включая геомаркетинг и геостатистику [1; 2], открывают новые горизонты для изучения взаимосвязей между экономическими показателями, представленными в форме временных рядов, и пространственными характеристиками субъектов хозяйственной деятельности, расположенных на территории региона и муниципалитетов. Кластеризация данных по пространственным и временным признакам позволяет не только сегментировать территории по уровню экономического развития, но и определить закономерности изменения условий ведения бизнеса, которые могут быть использованы для оптимизации ресурсов и повышения эффективности управления на региональном уровне.

Экономический профиль [3] как комплексная характеристика экономического состояния и динамики изменчивости структуры деловой активности бизнеса на определенной территории региона или страны позволяет оценить конкурентоспособность бизнеса, выявить его сильные и слабые стороны, а также определить направления для дальнейшего развития и улучшения экономической ситуации.

Методы исследования

Анализ существующих моделей пространственно-временной кластеризации включает в себя набор как классических, так и специализированных подходов и методов, позволяющих выявлять закономерности в данных и выполнять группировку

объектов, характеризующихся пространственным расположением, а также социально-экономическими показателями, представленными в форме временных рядов.

Наиболее популярными для решения задач моделирования пространственно-временных зависимостей являются модели пространственно-временной регрессии [10]. В основе работы подобных регрессионных моделей лежит использование кригинговой интерполяции [8; 10] – метода, который позволяет оценивать значения экономических показателей в наблюдаемых точках на основе пространственной корреляции. Кригинговая интерполяция часто используется для создания пространственных карт экономических показателей. Также в ряде случаев не теряют свою актуальность и модификации классических авторегрессионных моделей, такие как SARIMA (Seasonal Autoregressive Integrated Moving Average) [5], которые могут быть расширены для учета пространственных зависимостей, что позволяет анализировать временные ряды экономических данных с учетом их пространственного контекста. Широкое распространение получили также модели пространственной автокорреляции [5; 7], основанные на использовании индекса Морриса или теста Лоренца. Подобные модели чувствительны к выбросам в данных, а интерпретация результатов может быть осложнена в контексте работы с высокоразмерными данными.

Современные тенденции показывают возможности адаптации нейронных сетей, таких как рекуррентные нейронные сети (RNN) и сверточные нейронные сети (CNN), для анализа пространственно-временных данных [6; 19]. Эти методы могут эффективно обрабатывать большие объемы данных и выявлять сложные зависимости, однако имеют недостатки в интерпретации полученных результатов. Кроме того, эффективное обучение таких моделей требует большого количества данных. При этом классические алгоритмы кластерного анализа, такие как K-means и K-medoids [18], базирующиеся

на стандартных метриках расстояний в многомерном пространстве признаков, предполагают, что кластеры имеют сферическую форму и одинаковую плотность, что не всегда соответствует реальности и делает их непригодными для решения поставленной задачи. Методы кластеризации на основе плотности, например, DBSCAN, также неэффективно работают с кластерами, имеющими различную плотность и аномалии.

При выборе эффективной модели пространственно-временной кластеризации важно учитывать специфику задачи, доступные данные и цели исследования, чтобы минимизировать влияние указанных недостатков на результаты анализа. В этом смысле необходимо определить, какие конкретно объекты и в каком контексте необходимо рассмотреть.

Экономический профиль региона формируется по результатам хозяйственной деятельности различных экономических субъектов, включая крупные промышленные предприятия, субъекты малого и среднего предпринимательства (МСП), а также сельскохозяйственные предприятия, крестьянско-фермерские хозяйства и т. п., т. е. все субъекты, производящие готовые продукты и услуги и формирующие рабочие места. Анализ существующих стратегий ведения хозяйственной деятельности крупных предприятий, особенно в сфере промышленного производства и добычи полезных ископаемых, показывает высокое

влияние факторов близости ресурсной базы, условий ведения внешнеэкономической деятельности региона [13], внешнеполитических факторов. Следует также отметить, что крупный бизнес зачастую подвержен более строгому регулированию и контролю со стороны государственных органов, и его стратегическое развитие чаще рассматривается уже на макрорегиональном уровне. Поэтому в качестве основы анализа структуры и видов экономической деятельности на территории региона и муниципалитетов следует рассматривать субъекты МСП, формирующие активную деловую среду территории, для которых наличие крупного бизнеса является внешним фактором развития экономической среды.

Данные о субъектах малого и среднего предпринимательства в динамике формируются на основе налоговых деклараций, подаваемых ежегодно субъектами МСП, предоставляющих информацию о доходах, расходах и других финансовых показателях, что позволяет анализировать динамику их деятельности. Использование геоинформационных систем (ГИС) позволяет визуализировать и анализировать данные о субъектах МСП в пространственном контексте, что помогает выявить тенденции и закономерности в их распределении и динамике, в том числе с помощью методов геостатистики, картографии и экономико-математических методов, как, например, показано на рис. 1, а.

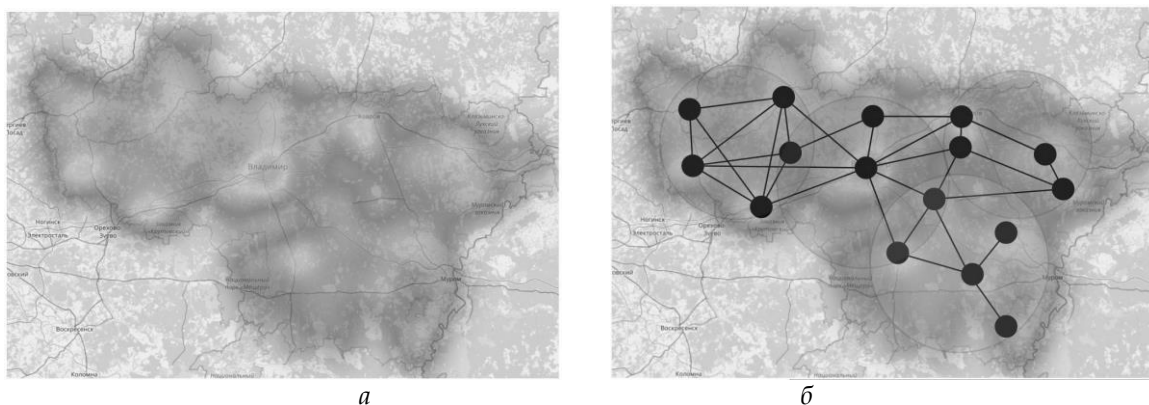


Рис. 1. Тепловая карта плотности размещения субъектов МСП на территории региона на примере Владимирской области (а) и сетевая модель кластеризации пространственных объектов (б)

Ввиду изменчивости количества значимых показателей, характеризующих результативность деятельности субъектов МСП во времени с привязкой к географическим координатам, необходима разработка универсальных методов кластеризации, позволяющих работать с большим количеством разнородных признаков и имеющих высокую степень унификации для решения поставленной задачи. С этой проблемой помогает справиться комплексный подход, включающий не только использование универсальных методов кластеризации, способных работать на размеченных данных с применением специализированных метрик оценки сходства объектов, но и методов уменьшения размерности пространства признаков, снижающих информационную избыточность данных и позволяющих легче интерпретировать результаты кластеризации и анализировать, какие признаки оказывают наибольшее влияние на формирование кластеров, что особенно полезно для принятия решений и разработки стратегий.

В свою очередь методы уменьшения размерности в комбинации с методами теории графов и анализа сетевых сообществ позволяют визуализировать высокоразмерные данные, что облегчает анализ и интерпретацию пространственно-временных кластеров. Так, например, метод анализа главных компонент (PCA) или метод t-SNE (t-distributed Stochastic Neighbor Embedding) [11; 12] помогают выявить и удалить нерелевантные или избыточные признаки, которые могут вносить шум в данные и ухудшать качество кластеризации.

Кроме того, уменьшение размерности может улучшить производительность алгоритмов машинного обучения, используемых для кластеризации, за счет уменьшения переобучения и повышения обобщающей способности моделей.

UMAP (Uniform Manifold Approximation and Projection) – это алгоритм уменьшения размерности, который был предложен в 2018 г. [9; 17]. Он быстро завоевал попу-

лярность благодаря своей способности сохранять глобальную и локальную структуру данных, а также обеспечивать высокую скорость работы и эффективность при визуализации высокоразмерных данных с применением графовых методов. UMAP основывается на теории многообразий и топологии, что позволяет ему эффективно моделировать сложные структуры в данных. Он использует графы для представления данных и оптимизирует их проекцию в пространство меньшей размерности.

Для решения поставленной задачи предлагается использовать подход на основе комбинированных с UMAP методов, включая методы глубокого обучения – модели автокодировщиков (или автоэнкодеров) [14; 16] для улучшения результатов в специфических задачах, таких как кластеризация и классификация.

Задача автокодировщика – обучаться восстановлению входных данных. Подобные модели состоят из двух частей: кодера – преобразует входные данные x в скрытое представление $z = f(x, \theta_e)$, и декодера – восстанавливает входные данные $\hat{x} = g(z, \theta_d)$ из скрытого представления z , где f – функция кодера с параметрами θ_e ; g – функция декодера с параметрами θ_d .

Обучение автокодировщика происходит путем минимизации функции потерь, например, среднеквадратичной ошибки (MSE):

$$L(\theta_e, \theta_d) = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2.$$

После того как было сформировано скрытое представление z автокодировщика, необходимо применить UMAP для дальнейшего уменьшения размерности. Для каждого объекта z_i в скрытом пространстве находится k ближайших соседей и строится граф, где узлы представляют объекты, а ребра – расстояния между ними.

Для каждого узла z_i скрытого представления вычисляются вероятности p_{ij} для соседей z_j :

$$p_{ij} = \frac{\exp\left(-\frac{\|z_i - z_j\|^2}{\sigma_i^2}\right)}{\sum_{k \in N_i} \exp\left(-\frac{\|z_i - z_k\|^2}{\sigma_i^2}\right)},$$

где $\|z_i - z_j\|^2$ обозначим как $d(z_i, z_j)$ – расстояние между точками z_i и z_j ;

N_i – множество соседей узла z_i ;

σ_i – параметр, определяющий ширину гауссова распределения (рис. 2).

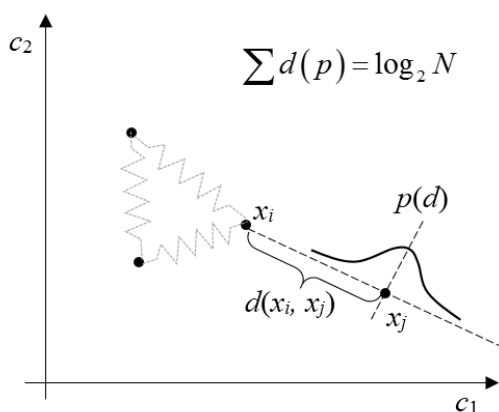


Рис. 2. Нормировка расстояний между точками UMAP в пространстве главных компонент c_1 и c_2

UMAP минимизирует разницу между вероятностями в исходном пространстве x и в низкоразмерном пространстве y . Это достигается с помощью минимизации функции потерь, которая представляет собой дивергенцию Кульбака – Лейблера:

$$D_{KL} = \sum_i \sum_j p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right),$$

где q_{ij} – вероятности в низкоразмерном пространстве y , расстояния между которыми вычисляются аналогично.

$$q_{ij} = \frac{(1 + d(y_i, y_j))^{-1}}{\sum_{k \neq i} (1 + d(y_i, y_k))^{-1}},$$

где y_i и y_j – проекции точек в низкоразмерном пространстве.

Таким образом, процесс комбинирования UMAP и моделей автокодировщиков можно описать следующим образом:

- обучение автокодировщика на исходных данных x для получения скрытого представления z ;

- применение UMAP к скрытому представлению z для дальнейшего уменьшения размерности и визуализации: $y = \text{UMAP}(z)$.

Часть классической схемы автокодировщика, необходимая для декодирования (восстановления) исходных данных из скрытого состояния, в этом случае работает только для обучения скрытого состояния, на основе которого формируется граф связности UMAP и выполняется дальнейшая кластеризация (рис. 3).

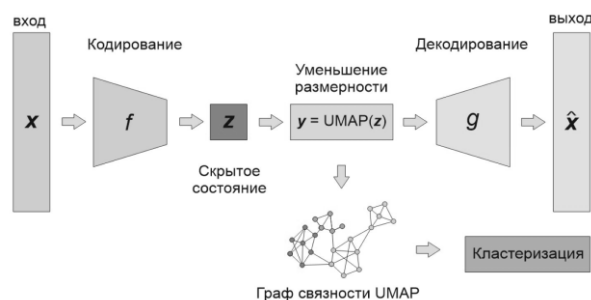


Рис. 3. Структурная схема комбинирования автокодировщика и алгоритма UMAP для задачи кластеризации

Последовательная комбинация модели автокодировщика и UMAP выбрана из соображений необходимости выполнения процедуры извлечения признаков и уменьшения информационной избыточности данных, а затем уменьшения размерности и изучения топологии связей объектов.

Таким образом, автокодировщики хорошо работают с данными, которые имеют сложные структуры и могут быть представлены в виде векторов (например, изображения, текстовые данные, пространственно-временные данные), а UMAP эффективен для визуализации данных высокой размерности с большим количеством признаков и сохранения глобальной и локальной топологии данных. Оба метода

используются на стадии предварительной обработки данных.

Результаты и обсуждение

Основной гипотезой для проведения исследований является тезис, что комбинирование автокодировщиков и UMAP позволит эффективно уменьшить размерность пространства признаков и сохранить их структуру, что приведет к повышению качества и интерпретируемости результатов кластерного анализа.

Рассмотрим результаты применения описанного комбинированного подхода, основанного на моделях автокодировщиков и UMAP, для уменьшения размерности, кластеризации и визуализации высокоразмерных данных о деятельности субъектов МСП. Эксперименты проводились на основе набора данных, содержащих информацию из реестра субъектов малого и среднего предпринимательства, дополненную сведениями о доходах, расходах, среднесписочной численности работников, а также географическими координатами населенных пунктов, где зарегистрированы субъекты МСП.

Указанный набор данных был сформирован на базе выгрузок реестра МСП и других открытых данных Федеральной налоговой службы России за 2018–2023 гг. Обработка данных выполнялась с применением специализированных программных пакетов на языке программирования Python. Для большей наглядности данные были взяты на примере одного региона – Владимирской области.

Уровень детализации (грануляция) данных предполагает наблюдение характеристик одного уникального субъекта МСП в привязке ко времени и локации. Таким образом, после предварительного анализа данных на территории указанного региона за 2018–2023 гг. насчитывалось порядка 830 средних, 10 тыс. малых и 77 тыс. микропредприятий – уникальных объектов, которые имели при этом свыше 1 200 уникальных наименований видов деятельности (ОКВЭД от 4 до 6 знаков).

Для унификации представления видов деятельности все идентификаторы ОКВЭД были укрупнены до групп (4 знака), и было получено порядка 670 уникальных видов деятельности за все время наблюдения.

По сформированным видам деятельности была построена сводная таблица, отражающая динамику суммарных доходов МСП по каждому виду деятельности, что дает возможность проведения агрегированной оценки результативности работы МСП по видам деятельности. В перспективе необходимо провести дополнительные исследования на предмет выбора показателя результативности деятельности предприятий [4].

На данном этапе такой показатель дает достаточное представление об относительной динамике деловой активности по отдельным видам экономической деятельности.

Для упрощения анализа и наглядности представления полученных кластеров была выполнена фильтрация по суммарным доходам за весь период времени по каждому виду деятельности (пороговое значение – 100 млн рублей за весь период), что сократило количество наблюдаемых объектов (видов деятельности МСП) до 342. Именно этот список самых результативных видов деятельности в дальнейшем использовался для построения вторичного набора данных и анализа, включающего суммарные значения среднесписочной численности работников, доходов и расходов организации за год, а также усредненные значения геокоординат (широта и долгота), характеризующих размещение субъектов МСП и их категорию по каждому из укрупненных временных периодов (по годам). Это позволило получить набор, содержащий 6 обособленных признаков за 6 периодов – всего 36 столбцов и 342 строки для ОКВЭД.

Полученный набор данных использован в качестве входного тензора x для обучения модели автокодировщика. Характеристики структуры модели автокодировщика представлены на рис. 4, а.

Результаты обучения автокодировщика по метрике MSE представлены на рис. 4, б.

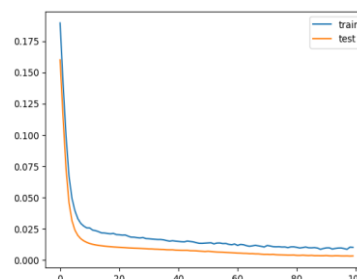
Результаты позволяют утверждать, что модель с достаточно высокой точностью

обучена восстановлению характеристик входной последовательности данных.

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 36)	0
dense (Dense)	(None, 8)	296
dropout (Dropout)	(None, 8)	0
dense_1 (Dense)	(None, 36)	324

Total params: 1,862 (7.28 KB)
Trainable params: 620 (2.42 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 1,242 (4.86 KB)



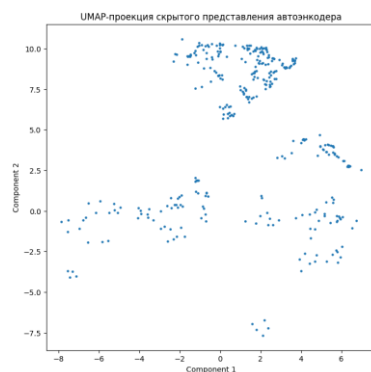
а

б

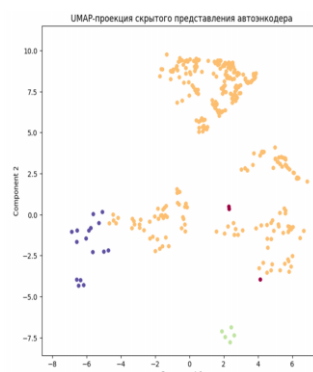
Рис. 4. Характеристики структуры модели (а) и результаты обучения (б) автокодировщика

На основе скрытого состояния автокодировщика – слоя, состоящего из 8 нейронов, осуществляется представление входных данных в пространстве меньшей размерности, которые затем переносятся с помощью алгоритма UMAP в топологическое представление в форме компактных

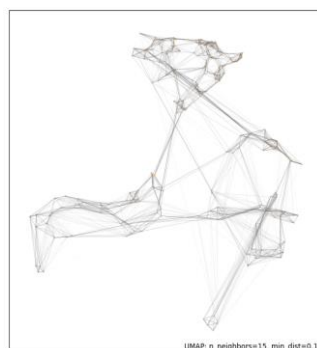
связных множеств, формируя тем самым проекцию в пространство еще меньшей размерности, но с сохранением структуры взаимосвязей между объектами (видами деятельности) по их характеристикам (рис. 5, а).



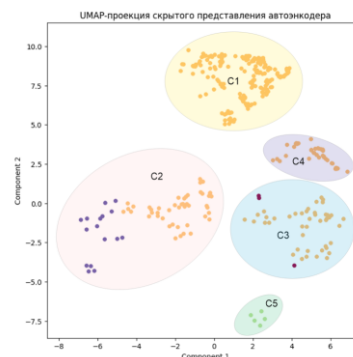
а



б



в



г

Рис. 5. UMAP-проекция скрытого состояния автокодировщика (а), маркировка по результатам кластеризации исходных данных по методу K-means (4 кластера) (б), граф связности UMAP в проекции двух главных компонент (в), маркировка по выделенным кластерам после кластеризации графа связности UMAP (5 кластеров) (г)

Для того чтобы наглядно показать эффект применения автокодировщика и алгоритма UMAP, вначале была выполнена кластеризация входных данных только классическим вариантам на основе алгоритма K-means: после выбора количества кластеров путем оценки внутрикластерного разнообразия и межкластерных расстояний было определено оптимальное разбиение видов деятельности на 4 кластера, содержащих 318, 16, 5 и 3 вида деятельности соответственно.

Кластер с условным номером 2 (16 объектов): предприятия с видами деятельности 41.XX, 45.XX–47.XX – подавляющее большинство секторов оптовой и розничной торговли.

Кластер с условным номером 3 (5 объектов): эти виды деятельности можно отнести к вымирающим. В них динамика всех основных показателей идет на спад. Это в основном предприятия по производству ручных инструментов, спортивных товаров, стали и ферросплавов.

Кластер с условным номером 4 (3 объекта): растущие отрасли – производство офисной техники и оборудования, велосипедов, обработка цветных металлов.

Разумеется, самый большой кластер является неинформативным и не позволяет сделать однозначные выводы о причинах группировки объектов, что делает указанный подход слабо пригодным для практического применения в данном случае.

На рис. 5, б показано применение алгоритма UMAP для визуализации точек в пространстве двух главных компонент на основе скрытого состояния автокодировщика, а также представлены результаты маркировки видов деятельности по полученным ранее с помощью K-means кластерам. Для кластера вымирающих видов деятельности ситуация практически не менялась, однако уже прослеживается тенденция к увеличению количества кластеров и частичному несоответствию с результатами кластеризации на основе K-means.

Теперь покажем, как реализуется разбиение с применением алгоритма UMAP с использованием графы для представления данных, что позволяет лучше сохранять как локальную, так и глобальную структуру взаимосвязей объектов (рис. 5, в). UMAP имеет несколько параметров, которые можно настраивать, таких как `n_neighbors` (количество соседей для определения локальной структуры) и `min_dist` (минимальное расстояние между точками в низкоразмерном пространстве). UMAP начинает с построения графа связности, который представляет данные в виде точек (объектов) и ребер (связей между объектами). Для каждой точки вычисляются ее ближайшие соседи и создается граф, где ребра соединяют точки, которые близки друг к другу. Вес каждого ребра определяется метрическим расстоянием между вершинами графа. В данном случае использовалась мера косинусного расстояния [16; 17].

Теперь к этой структуре можно вновь применить алгоритм кластеризации по методу K-means и проанализировать результаты. В этом случае оптимальным является разбиение уже на 5 более сбалансированных по количеству точек кластеров, интерпретировать которые можно следующим образом (рис. 5, г):

- кластер 1 (144 точки): обрабатывающие производства, строительство, оптовая и розничная торговля объединяются больше по пространственному принципу и имеют схожие географические координаты, совпадающие с координатами областного центра (Владимир) и его окрестностями;

- кластер 2 (89 точек): обрабатывающие производства (60%) с положительной динамикой и высокими экономическими показателями, но расположенные на территории всего региона – преимущественно малые предприятия, но с относительно удаленными друг от друга локациями. Сюда вошли 16 ранее выявленных субъектов;

- кластер 3 (64 точки): обрабатывающие производства (30%), торговля оптовая

и розничная, расположенные на территории всего региона, преимущественно средние предприятия – субъекты МСП разнообразных сфер деятельности с положительной динамикой среднесписочной численности сотрудников в год и растущих показателей доходности. Содержит в том числе ранее выявленные объекты растущих отраслей;

- кластер 4 (40 точек): обрабатывающие производства (50%) и предприятия с динамикой снижения среднесписочной численности сотрудников в год, но при этом стабильными показателями годовой доходности;

- кластер 5 (5 точек): сюда вошли предприятия ранее указанного вымирающего кластера.

Следует отметить тот факт, что модель автокодировщика обладает некоторой степенью вариативности, поэтому при проведении экспериментов количество и расположение точек в проекциях могут несущественно изменяться, сохраняя при этом баланс на уровне оптимального разбиения на 5–6 кластеров с относительно стабильными пропорциями количества объектов и сохранением общих характеристик объектов внутри каждого кластера, что позволяет интерпретировать полученные результаты весьма однозначно.

Заключение

Результаты исследования и интерпретации пространственно-временных паттернов структуры деловой активности в регионе позволяют сформировать более логичную и сбалансированную картину изменчивости и локализации факторов, описывающих экономическое развитие региона. Применение методов кластерного анализа, уменьшение размерности пространства признаков и моделей автокодировщиков позволяют глубже понять взаимосвязи между характеристиками различных субъектов бизнеса и более информативно выявить ключевые тренды.

Модель автокодировщика использована для извлечения значимых признаков из

данных. Их скрытое (латентное) пространство содержит более информативные представления данных, что позволяет в конечном итоге представить входной набор данных с уменьшением информационной избыточности.

Методы уменьшения размерности играют ключевую роль в процессе кластеризации, особенно в условиях работы с высокоразмерными и сложными данными. Они позволяют упростить структуру данных, улучшить визуализацию, сохранить важные характеристики и устранить шум, что в конечном итоге способствует более адекватному и эффективному выявлению кластеров.

Использование таких методов, как PCA, t-SNE и UMAP, помогает исследователям и аналитикам не только ускорить вычислительные процессы, но и повысить качество кластеризации, избегая проблем, связанных с «проклятием размерности». Эти подходы позволяют сосредоточиться на наиболее значимых признаках, что делает результаты анализа более интерпретируемыми и полезными для принятия обоснованных решений.

Комбинирование UMAP и автокодировщиков позволяет реализовать уменьшение размерности и извлечение признаков из высокоразмерных данных, формируя при этом более информативные признаки, выполняя дальнейшую кластеризацию объектов и эффективно визуализируя их в низкоразмерном пространстве. Этот подход оказывается особенно полезным в задачах, где данные имеют сложную структуру и требуют как нелинейного уменьшения размерности, так и сохранения локальных и глобальных свойств. Таким образом, интеграция методов уменьшения размерности в процесс кластеризации является важным шагом для достижения более глубокого понимания структуры данных и выявления скрытых закономерностей при формировании экономического профиля региона.

Список литературы

1. Грекусис Д. Методы и практика пространственного анализа. Описание, исследование и объяснение с использованием ГИС / пер. с англ. А. Н. Киселева. – М. : ДМК Пресс, 2021.
2. Корчагина И. В., Пытченко К. В. Социально-экономическая система регионального предпринимательства как объект стратегирования // Экономика промышленности. – 2023. – № 16 (4). – С. 361–371.
3. Кричевский М. Л., Мартынова Ю. А. Использование методов машинного обучения для оценки инвестиционной деятельности различных регионов России // Вопросы инновационной экономики. – 2019. – Т. 9. – № 4. – С. 1557–1572.
4. Медведева О. А. Оценка экономического потенциала региона для развития кластеров // Развитие территорий. – 2023. – № 3. – С. 25–31.
5. Окунев И. Ю. Основы пространственного анализа : монография. – М. : Аспект Пресс, 2020.
6. Яковлев И. В., Яковлева О. А. Факторный анализ и кластеризация данных состояния инженерной инфраструктуры сельских территорий субъектов Российской Федерации // Вестник Академии права и управления. – 2019. – № 4 (57). – С. 86–90.
7. Assuncao R. M., Reis E. A. A New Proposal to Adjust Moran's I for Population Density // Statistics in Medicine. – 1999. – Vol. 18 (16). – P. 2147–2162.
8. Basaraner M., Cetinkaya S. Performance of Shape Indices and Classification Schemes for Characterising Perceptual Shape Complexity of Building Footprints in GIS // International Journal of Geographical Information Science. – 2017. – Vol. 31 (10). – P. 1952–1977.
9. Becht E. et al. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP // Nature Biotechnology. – 2019. – Vol. 37. – N 1. – P. 38–44.
10. Church R. L., Murray A. T. Business Site Selection, Location Analysis and GIS. – John Wiley & Sons, Inc., 2009.
11. Franklin J. The Elements of Statistical Learning: Data Mining, Inference and Prediction // The Mathematical Intelligencer. – 2003. – Vol. 27. – P. 83–85.
12. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. – 2nd edition. – Springer, 2017.
13. Hausmann R., Hidalgo C. A. The Network Structure of Economic Output // Journal of Economic Growth. – 2011. – Vol. 16 (4). – P. 309–342.
14. Kim Yongho, Jan Heiland. Convolutional Autoencoders, Clustering and POD for Low-Dimensional Parametrization of Navier-Stokes Equations. – DOI: 10.48550/arXiv.2302.01278
15. Kuk G., Marijn J. The Business Models and Information Architectures of Smart Cities // Journal of Urban Technology. – 2011. – Vol. 18 (2). – P. 39–52.
16. Liu Aofu, Zexuan Ji. Deep Mixture of Adversarial Autoencoders Clustering Network // Pattern Recognition and Computer Vision. – Springer, 2021. – P. 191–202.
17. McInnes L., Healy J., Melville J. Umap: Uniform Manifold aProximation and Projection for Dimension Reduction. – URL: arXiv preprint arXiv:1802.03426
18. Wolf L. J., Knaap E., Rey S. Geosilhouettes: Geographical Measures of Cluster Fit // Environment and Planning. – 2019. – Vol. 48 (1).
19. Zhu Jia, Baofeng Li, Hong Chen. AQI Multi-Point Spatiotemporal Prediction Based on K-Mean Clustering and RNN-LSTM Model // Journal of Physics Conference Series. – 2021. – Vol. 2006 (1).

References

1. Grekuis D. Metody i praktika prostranstvennogo analiza. Opisanie, issledovanie i obyasnienie s ispolzovaniem GIS [Methods and Practice of Spatial Analysis. Description, Research, and Explanation Using GIS], translated from English by A. N. Kiselev. Moscow, DMK Press, 2021. (In Russ.).
2. Korchagina I. V., Pytchenko K. V. Sotsialno-ekonomicheskaya sistema regionalnogo predprinimatelstva kak obekt strategirovaniya [The Socio-Economic System of Regional Entrepreneurship as an Object of Strategic Planning]. *Ekonomika promyshlennosti* [Industrial Economics], 2023, No. 16 (4), pp. 361–371. (In Russ.).
3. Krichevskiy M. L., Martynova Yu. A. Ispolzovanie metodov mashinnogo obucheniya dlya otsenki investitsionnoy deyatel'nosti razlichnykh regionov Rossii [Using Machine Learning Methods to Assess Investment Activity in Various Regions of Russia]. *Voprosy innovatsionnoy ekonomiki* [Innovation Economy Issues], 2019, Vol. 9, No. 4, pp. 1557–1572. (In Russ.).
4. Medvedeva O. A. Otsenka ekonomicheskogo potentsiala regiona dlya razvitiya klasterov [Assessment of the Economic Potential of the Region for Cluster Development]. *Razvitie territoriy* [Territorial Development], 2023, No. 3, pp. 25–31. (In Russ.).
5. Okunev I. Yu. Osnovy prostranstvennogo analiza: monografiya [Fundamentals of Spatial Analysis: monograph]. Moscow, Aspekt Press, 2020. (In Russ.).
6. Yakovlev I. V., Yakovleva O. A. Faktorniy analiz i klasterizatsiya dannykh sostoyaniya inzhenernoy infrastruktury selskikh territoriy subektov Rossiyskoy Federatsii [Factor Analysis and Clustering of Data on the Condition of Engineering Infrastructure in Rural Areas of the Subjects of the Russian Federation]. *Vestnik Akademii prava i upravleniya* [Bulletin of the Academy of Law and Management], 2019, No. 4 (57), pp. 86–90. (In Russ.).
7. Assuncao R. M., Reis E. A. A New Proposal to Adjust Moran's I for Population Density. *Statistics in Medicine*, 1999, Vol. 18 (16), pp. 2147–2162.
8. Basaraner M., Cetinkaya S. Performance of Shape Indices and Classification Schemes for Characterising Perceptual Shape Complexity of Building Footprints in GIS. *International Journal of Geographical Information Science*, 2017, Vol. 31 (10), pp. 1952–1977.
9. Becht E. et al. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nature Biotechnology*, 2019, Vol. 37, No. 1, pp. 38–44.
10. Church R. L., Murray A. T. Business Site Selection, Location Analysis and GIS. John Wiley & Sons, Inc., 2009.
11. Franklin J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. *The Mathematical Intelligencer*, 2003, Vol. 27, pp. 83–85.
12. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition. Springer, 2017.
13. Hausmann R., Hidalgo C. A. The Network Structure of Economic Output. *Journal of Economic Growth*, 2011, Vol. 16 (4), pp. 309–342.
14. Kim Yongho, Jan Heiland. Convolutional Autoencoders, Clustering and POD for Low-Dimensional Parametrization of Navier-Stokes Equations. DOI: 10.48550/arXiv.2302.01278
15. Kuk G., Marijn J. The Business Models and Information Architectures of Smart Cities. *Journal of Urban Technology*, 2011, Vol. 18 (2), pp. 39–52.
16. Liu Aofu, Zexuan Ji. Deep Mixture of Adversarial Autoencoders Clustering Network. *Pattern Recognition and Computer Vision*. Springer, 2021, pp. 191–202.
17. McInnes L., Healy J., Melville J. Umap: Uniform Manifold aProximation and Projection for Dimension Reduction. Available at: arXiv preprint arXiv:1802.03426

18. Wolf L. J., Knaap E., Rey S. Geosilhouettes: Geographical Measures of Cluster Fit. *Environment and Planning*, 2019, Vol. 48 (1).

19. Zhu Jia, Baofeng Li, Hong Chen. AQI Multi-Point Spatiotemporal Prediction Based on K-Mean Clustering and RNN-LSTM Model. *Journal of Physics Conference Series*, 2021, Vol. 2006 (1).

Поступила: 04.02.2025

Принята к печати: 21.05.2025

Сведения об авторе

Алексей Николаевич Кисляков

доктор экономических наук,
кандидат технических наук,
доцент, профессор кафедры
информационных технологий
Владимирского филиала РАНХиГС.
Адрес: Владимирский филиал
ФГБОУ ВО «Российская академия
народного хозяйства и государственной
службы при Президенте
Российской Федерации»,
600017, Владимирская область,
Владимир, ул. Горького, д. 59а.
E-mail: ankislyakov@mail.ru
ORCID: 0000-0001-8790-6961

Information about the author

Aleksey N. Kislyakov

Doctor of Economics, PhD,
Associate Professor,
Professor of the Department
of Information Technology
of the Vladimir branch of RANEPa.
Address: Vladimir Branch
of the Russian Presidential Academy
of National Economy
and Public Administration,
59a Gorky Str., Vladimir,
Vladimir region, 600017,
Russian Federation.
E-mail: ankislyakov@mail.ru
ORCID: 0000-0001-8790-6961