
РЕГИОНАЛЬНЫЕ ПРОБЛЕМЫ

Кластерный анализ и классификация промышленно ориентированных регионов РФ по экономической специализации

© 2022 г. О.М. Шаталова, Е.В. Касаткина, В.Н. Лившиц

О.М. Шаталова,

Удмуртский филиал Института экономики УрО РАН, Ижевск; e-mail: oshatalova@mail.ru

Е.В. Касаткина,

Ижевский государственный технический университет имени М.Т. Калашникова, Ижевск; e-mail: e.v.trushkova@gmail.com

В.Н. Лившиц,

Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва; e-mail: livchits@isa.ru

Поступила в редакцию 14.09.2021

Статья подготовлена в соответствии с Программой фундаментальных научных исследований государственных академий наук и планом НИР Института экономики Уральского отделения РАН на 2021–2023 гг. по теме «Методология инновационного развития регионально-ориентированных систем в условиях нестабильной экономической конъюнктуры».

Аннотация. Статья посвящена вопросам организации промышленного сектора в пространственном развитии национальной экономики. Цель исследования состояла в классификации промышленно ориентированных регионов РФ по характеру их экономической специализации и развития необходимых для этого экономико-математических методов. В условиях высокой социально-экономической дифференциации субъектов РФ исследовательские задачи их классификации приобретают высокую актуальность, поскольку способствуют выявлению однородных (по задаваемым признакам) групп регионов, каждой из которых присущи особые закономерности функционирования. Классификация промышленных регионов по критерию их экономической специализации была проведена с помощью интеллектуальных математических методов кластерного анализа. Для решения задачи был разработан адаптированный алгоритм кластеризации, основанный на использовании метода k-средних. Результатом кластеризации явилось разделение совокупности промышленных регионов на 12 классов, устойчивых по признаку экономической специализации, и их содержательная интерпретация, которая раскрывается следующими положениями: а) составлены описательные характеристики, идентифицирующие каждый класс регионов на основании концепции трехсекторной модели экономики (Фишера–Кларка) и результатов количественной оценки диверсификации экономики регионов; б) выраженная однородность регионов внутри кластеров позволила составить дифференцированную оценку социально-экономических показателей, учитывающую специфику экономики регионов в каждом классе. Полученные результаты позволили уточнить некоторые существенные закономерности, которые могут иметь значение для повышения эффективности государственного управления.

Ключевые слова: региональная экономика, устойчивое развитие, пространственное развитие, промышленно ориентированные регионы, экономическая специализация, диверсификация экономики региона, кластерный анализ, машинное обучение.

Классификация JEL: C38, O14, R11.

Цитирование: Шаталова О.М., Касаткина Е.В., Лившиц В.Н. (2022). Кластерный анализ и классификация промышленно ориентированных регионов РФ по экономической специализации // Экономика и математические методы. Т. 58. № 1. С. 80–91. DOI: 10.31857/S042473880018971-7

ВВЕДЕНИЕ

Важной функцией государственной региональной политики выступает обеспечение устойчивого экономического развития регионов. В сложившихся условиях высокой межрегиональной социально-экономической дифференциации для реализации этой функции становятся необходимы научно-обоснованные методы классификации регионов. «Стратегия пространственного развития

РФ на период до 2025 года»¹ определила ряд задач, в решении которых необходимо учитывать специфику региональной экономики субъектов РФ, в том числе такое ее проявление, как *экономическая специализация* — совокупность отраслей, имеющих приоритетное значение для социально-экономического развития региона. Экономическая специализация субъектов РФ является существенной предпосылкой к обоснованной оценке их макроэкономических параметров и выявлению действенных приоритетов экономического развития как основаниям для формирования региональной, инновационной, промышленной политики государственного управления. Данное положение определяет необходимость классификации регионов по признаку их экономической специализации. Актуальность исследования *промышленно ориентированных регионов* РФ обусловлена тем, что промышленность играет центральную роль в экономике современного общества². Индустриальный сектор экономики способствует повышению занятости, обеспечивает материальную основу повышения уровня жизни, служит базой для развития торговли и сельского хозяйства, стимулирует научно-исследовательскую и инновационную деятельность.

Представляется обоснованным классифицировать регионы средствами современных математических методов кластерного анализа, который направлен на многомерную обработку данных и оптимальное разбиение множества исследуемых объектов на однородные группы по значимым для исследования признакам. Назначение кластерного анализа — снижение размерности изучаемой совокупности объектов за счет их обоснованной классификации, что создает предпосылки к детализированному исследованию совокупности явлений для выявления актуальных причинно-следственных отношений и закономерностей (Айвазян и др., 1989).

В отечественных научных исследованиях регионы являются наиболее типичными объектами кластеризации и классификации данных по различным критериальным признакам: инновационное развитие (Голова, Суховой, 2019), качество жизни (Локосов, Рюмина, Ульянов, 2015), уровень человеческого капитала (Петрыкина, 2013; Локосов, Рюмина, Ульянов, 2019) и т.д. Большое значение для проводимого нами исследования имеют результаты кластеризации регионов РФ *по признаку структуры ВРП* в разрезе укрупненных ВЭД (Айвазян, Афанасьев, Кудров, 2016). Сложившиеся научные результаты показывают, что классификация регионов методами кластерного анализа способствует извлечению новых знаний о специфике региональной экономики в контексте решаемых научных и прикладных задач.

Цель нашего исследования состоит в классификации промышленно ориентированных регионов РФ по характеру их экономической специализации. Полученные результаты (выявленные методами кластерного анализа группы однородных по своей экономической специализации промышленных регионов РФ) будут способствовать, по нашему мнению, систематизации экономических исследований таких регионов, уточнению существенных закономерностей их функционирования и развития.

1. КЛАСТЕРНЫЙ АНАЛИЗ В КЛАССИФИКАЦИИ ПРОМЫШЛЕННО ОРИЕНТИРОВАННЫХ РЕГИОНОВ РФ

Методика исследования и решения задачи классификации промышленно ориентированных регионов РФ по характеру их экономической специализации раскрывается в следующих основных положениях.

Совокупность промышленно ориентированных регионов РФ сформирована с учетом представленных в статье (Айвазян, Афанасьев, Кудров, 2016) результатов кластеризации, которые получены методом максимального правдоподобия на основе модели граничного производственного потенциала по вектор-характеристике структуры региональной экономики, включающей индекс промышленной специализации и индекс индустриализации. Результаты кластеризации позволили выделить группу регионов со специализацией в промышленности (по укрупненным, агрегированным ВЭД). При формировании совокупности промышленных регионов учитывалась региональная структура ВРП (по критерию превышения доли ВЭД, относимых к промышленному сектору, над

¹ «Стратегия пространственного развития РФ на период до 2025 года» (утв. Распоряжением Правительства РФ от 13.02.2019 № 207-п) (<http://government.ru/docs/35733>).

² «Наше общее будущее: доклад Всемирной комиссии по окружающей среде и развитию» (официальный отчет Генеральной Ассамблеи ООН, 1987) (<https://www.un.org/ru/ga/pdf/brundtland.pdf>).

среднероссийским значением). Кластерный анализ проводился на основании данных об объеме отгруженной продукции в разрезе отраслей промышленности (ВЭД второго уровня).

Для определения достаточного числа кластеризационных признаков, при которых не будет ярко выраженной мультиколлинеарности, была проанализирована зависимость доли объясненной дисперсии от числа признаков. По результатам такого анализа были исключены данные по тем ВЭД, которые очевидно несущественны для проводимого исследования.

Основным методом кластерного анализа явился метод k -средних. Этот метод принадлежит к группе итеративных методов эталонного типа (Охупкина, Охупкин, 2015). Сформированный в рамках исследования адаптированный алгоритм кластеризации, основанный на методе k -средних, включает интеллектуальную процедуру « k -means++» (Arthur, Vassilvitskii, 2007), которая обеспечивает решение проблемы чувствительности к выбору начального приближения и ускорение конвергенции задания центров кластеров, позволяет инициировать максимально дифференцированные друг от друга устойчивые центры кластеров и достигать высокой и выраженной отделимости кластеров. Необходимое для реализации метода k -средних число кластерных разбиений (K) было установлено на основании анализа зависимости целевой функции кластеризации от K , а также в ходе экспериментальных вычислительных процедур, направленных на проверку выполнимости условий компактности и отделимости формируемых кластеров.

Проверка устойчивости результатов кластеризации, полученных итерационным методом k -средних, проводилась в соответствии с эмпирическим правилом, согласно которому устойчивость решения доказывается при изменении методов (Олдендерфер, Блэшфилд, 1989). Для проверки был использован метод иерархического кластерного анализа.

2. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Информационную основу исследования составили данные государственной статистики³. Преобработка данных включала оценку отраслевой структуры промышленного сектора и оценку уровня диверсификации экономики промышленных регионов.

Для оценки отраслевой структуры промышленного сектора регионов были использованы следующие исходные данные:

1) объем отгруженной продукции (ОП) по каждому ВЭД⁴ $V_i^{(r)}$; для исследования приняты ВЭД (r), относимые к промышленности, в разрезе регионов (i); $r = 1, \dots, m$, $i = 1, \dots, N$, где m — число ВЭД, относимых к промышленности, $m = 4$, N — число входящих в выборку промышленных регионов (в соответствии с принятой методикой исследования $N = 40$);

2) структура объема ОП по каждому виду экономической деятельности⁵ $y_{ij}^{(r)}$; показатель $y_{ij}^{(r)}$ представляет отношение объема ОП в отрасли j (V_{ij}) к объему ОП по соответствующему ВЭД ($V_i^{(r)}$):

$$y_{ij}^{(r)} = V_{ij} / V_i^{(r)}, \quad j = 1, \dots, L,$$

где L — число кластеризационных признаков (отраслей промышленности); при $r = 1$ рассматриваются отрасли $j = 1, \dots, 4$, при $r = 2$ — $j = 5, \dots, 16$, при $r = 3$ — $j = 17, 18, 19$, при $r = 4$ — $j = 19, \dots, 23$.

На основании данных по $V_i^{(r)}$ и $y_{ij}^{(r)}$ можно определить объем ОП отрасли j региона i V_{ij} : $V_{ij} = V_i^{(r)} y_{ij}^{(r)}$ и применить V_{ij} в расчете показателя «доля отрасли в объеме ОП по промышленности, в целом» x_{ij} (для характеристики отраслевой структуры промышленного сектора регионов):

$$x_{ij} = V_{ij} / \sum_{r=1}^m V_i^{(r)}. \quad (1)$$

Для выявления существенных критериев кластеризации L был использован метод главных компонент. С помощью этого метода анализировалась зависимость доли объясненной дисперсии от числа принимаемых признаков; проведенный анализ показал, что 13 из 23 признаков позволяют объяснить 99,9% дисперсии. С учетом данного положения, при исследовании нами принято — в качестве критериев кластеризации — $L = 13$ отраслей промышленности, по которым формируются основной объем ОП для рассматриваемых регионов.

³ «Регионы России. Социально-экономические показатели 2020» (https://gks.ru/bgd/regl/b20_14p/Main.htm).

⁴ Раздел 13.1. «Объем ОП <...> по видам экономической деятельности» статистического сборника «Регионы России ...».

⁵ Данные разделов 13.04, 13.05, 13.06, 013.07 статистического сборника «Регионы России...».

Недостаток информации в разделах 13.4, 13.5, 13.6, 13.7 статистического сборника «Регионы России. Социально-экономические показатели 2020» был компенсирован результатами регрессионного анализа. В качестве алгоритма регрессии применен метод Random Forests Regressor, поскольку он способен эффективно обрабатывать данные с большим числом признаков и может строить деревья по данным с пропущенными значениями признаков⁶.

Оценка диверсификации основана на индексе Хиршмана–Херфиндаля (I_{HH}). Данный показатель признается применимым для региональных исследований (Михеева, 2013). Индекс рассчитывается по формуле $I_{HH} = \sum_j x_{ij}^2$ и отражает степень доминирования отдельных отраслей в экономике региона. Высокое значение I_{HH} может свидетельствовать о выраженной специализации экономики региона, а низкое — о ее диверсификации. В оценке диверсификации необходимо учитывать, что в некоторых регионах сложились межотраслевые территориально-производственные комплексы (ТПК). Объемы производства по технологически смежным отраслям означают не диверсификацию экономики региона, а ее специализацию по соответствующему ТПК (Тимошенко, 2012). Данное положение было учтено через скорректированный расчет I_{HH} в регионах со сложившимися ТПК. Такой расчет проводился на основе объединения данных по смежным отраслям (выделены серым цветом в табл. 1).

Таблица 1. Результаты предварительной обработки данных для кластерного анализа промышленно ориентированных регионов РФ по характеру экономической специализации

i	Регион	Доля отраслей в объеме отгруженной продукции, %													I_{HH}
		Д ₁	Д ₂	Д ₃	Д ₄	ОП ₁	ОП ₃	ОП ₄	ОП ₅	ОП ₆	ОП ₈	ОП ₉	ОП ₁₀	ОЭ ₁	
0	Белгородская область	0	0	19	0	43	0	1	1	3	23	1	1	2	0,28
1	Владимирская область	0	0	0	1	28	3	1	6	11	13	9	7	4	0,13
2	Калужская область	0	0	0	0	16	0	2	7	6	9	12	39	1	0,21
3	Курская область	0	0	29	0	28	0	3	5	7	1	5	3	14	0,19
4	Липецкая область	0	0	0	1	21	0	2	2	1	61	4	2	2	0,42
5	Смоленская область	0	0	0	1	10	9	1	9	11	11	8	7	17	0,09
6	Тульская область	0	0	0	1	17	0	4	2	17	37	4	3	5	0,20
7	Республика Карелия	0	0	31	6	7	9	25	0	1	3	0	2	6	0,23
8	Республика Коми	6	58	0	0	1	5	9	9	0	4	0	0	4	0,36
9	Ненецкий АО	0	98	0	0	0	0	0	0	0	0	0	0	1	0,96
10	Архангельская область без АО	0	0	7	6	5	15	33	0	2	3	1	11	7	0,26
11	Вологодская область	0	0	0	0	7	5	1	2	19	55	0	2	3	0,35
12	Ленинградская область	0	0	0	1	17	2	8	24	7	4	3	12	11	0,13
13	Новгородская область	0	0	0	1	16	10	6	1	32	2	4	3	5	0,15
14	Астраханская область	0	75	0	1	2	0	0	10	1	0	0	2	4	0,73
15	Волгоградская область	0	6	0	0	7	0	0	47	6	17	0	3	5	0,34
16	Республика Башкортостан	0	14	2	1	5	1	1	36	13	3	2	10	4	0,41
17	Республика Марий Эл	0	0	0	0	28	2	7	12	2	9	18	9	4	0,15
18	Республика Татарстан	0	25	0	0	7	1	1	26	11	4	3	13	3	0,41
19	Удмуртская Республика	0	36	0	0	9	2	1	1	1	19	9	14	3	0,20
20	Пермский край	0	23	0	1	4	1	3	24	19	5	3	5	6	0,45
21	Нижегородская область	0	0	0	0	8	1	2	27	8	16	4	21	4	0,16
22	Оренбургская область	0	55	3	1	5	0	0	10	1	13	0	1	5	0,34
23	Самарская область	0	22	0	0	9	0	0	6	12	7	3	29	5	0,17
24	Саратовская область	0	6	0	1	17	0	1	5	16	9	6	13	15	0,11
25	Свердловская область	0	0	3	1	6	1	0	2	4	52	4	10	8	0,30
26	Ханты-Мансийский АО	0	80	0	0	0	0	0	13	0	0	0	0	4	0,87
27	Ямало-Ненецкий АО	0	86	0	0	0	0	0	11	0	0	0	0	1	0,94
28	Тюменская область без АО	0	25	0	0	4	0	0	33	20	3	1	2	4	0,61
29	Челябинская область	0	0	4	1	8	0	0	2	3	53	2	9	5	0,30
30	Республика Хакасия	26	0	5	1	5	0	0	0	1	32	1	1	20	0,54

⁶ Машинное обучение на Python (<https://scikit-learn.org>).

Окончание таблицы 1

<i>i</i>	Регион	Доля отраслей в объеме отгруженной продукции, %													<i>I_{HH}</i>
		Д ₁	Д ₂	Д ₃	Д ₄	ОП ₁	ОП ₃	ОП ₄	ОП ₅	ОП ₆	ОП ₈	ОП ₉	ОП ₁₀	ОЭ ₁	
31	Красноярский край	1	27	3	0	2	2	0	1	2	49	0	1	5	0,32
32	Иркутская область	1	39	9	0	4	5	5	3	5	10	1	4	8	0,19
33	Кемеровская область	55	0	0	0	3	0	0	9	5	16	0	2	4	0,57
34	Омская область	0	0	0	0	9	0	0	69	6	2	2	3	2	0,49
35	Томская область	0	45	0	0	9	4	1	9	7	3	9	1	4	0,24
36	Республика Саха (Якутия)	8	46	12	22	1	0	0	0	0	0	0	0	4	0,28
37	Магаданская область	1	0	14	1	1	0	0	0	1	73	0	0	6	0,76
38	Сахалинская область	4	87	0	0	6	0	0	0	0	0	0	0	1	0,76
39	Чукотский АО	3	0	13	0	2	0	0	0	0	71	0	0	5	0,71

Условные обозначения: Д₁ — добыча угля; Д₂ — добыча нефти и природного газа; Д₃ — добыча металлических руд; Д₄ — добыча прочих полезных ископаемых; ОП₁ — производство пищевых продуктов <...>; ОП₃ — обработка древесины и производство изделий из дерева <...>; ОП₄ — производство бумаги и бумажных изделий <...>; ОП₅ — производство кокса и нефтепродуктов <...>; ОП₆ — производство химических веществ и химических продуктов <...>; ОП₈ — производство металлургическое <...>; ОП₉ — производство компьютеров, электронных и оптических изделий; производство электрического оборудования; ОП₁₀ — производство машин и оборудования, не включенных в другие группировки <...>; ОЭ₁ — производство, передача и распределение электроэнергии.

Источник: расчеты авторов.

В ходе кластерного анализа решалась задача разбиения множества промышленных регионов $X = \{x_i\}_{i=1, \dots, N}$, определяемых вектором категориальных признаков $x_i = (x_{ij}) = (x_{i1}, \dots, x_{iL})$, на кластеры таким образом, чтобы суммарное отклонение регионов кластеров от центров кластеров F (внутриклассовое расстояние) было минимальным. То есть целевая функция кластеризации может быть представлена:

$$F = \sum_{k=1}^K \sum_{i=1}^{n_k} \|x_i^k - \mu^k\| \rightarrow \min, \quad (2)$$

где $\mu^k = (\mu_1^k, \dots, \mu_L^k)$ — центр кластера k ; x_i^k — регион i , входящий в кластер k ; n_k — число регионов, входящих в кластер k . В качестве расстояния $\rho(x_i^k, \mu^k) = \|x_i^k - \mu^k\|$ использовалось Евклидово расстояние.

Оптимальное число кластеров определяется в результате анализа зависимости целевой функции F (рис. 1). График показывает, что при $K \geq 12$ стабилизируется значение целевой функции; это гарантирует, что при таком числе кластеров выполняется свойство компактности. Экспериментально было установлено, что при $K > 12$ нарушается свойство отделимости кластеров (уже при $K = 13$ образуются кластеры, состоящие из 1 объекта). Учитывая такую проблему в использовании метода k -средних, как высокая чувствительность к выбросам, а также значительную дифференциацию регионов РФ, было задано достаточно высокое значение числа k ($K = 12$).

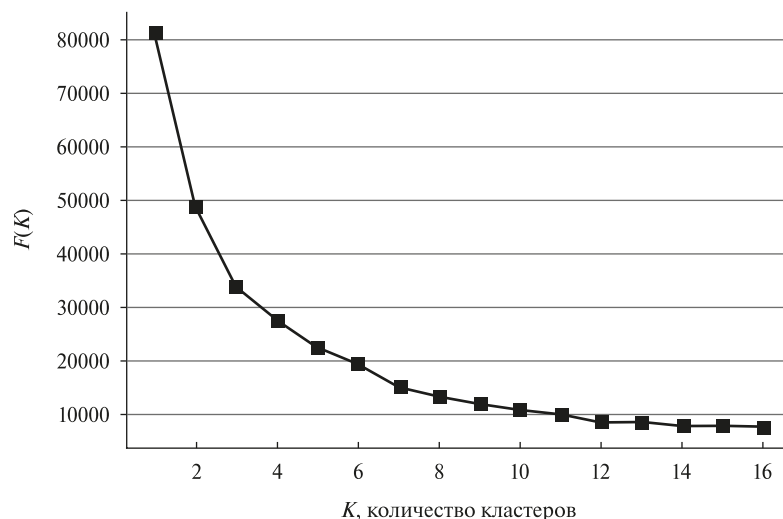


Рис. 1. Зависимость целевой функции кластеризации от числа кластеров

Алгоритм кластеризации, основанный на использовании метода k-средних и адаптированный под решаемую задачу.

1. Задается число кластеров K .
2. Инициализируются центры кластеров методом k-means++:
 - 2.1) первый центроид инициализирован случайным методом $\mu^k = x_p, p = \text{random}(1, N), k = 1$;
 - 2.2) расстояние каждого объекта до ближайшего центроида определяется по формуле $R_i = \min_k (\rho(x_i, \mu^k)), i = 1, \dots, N$;
 - 2.3) проводится расчет вероятности P_i каждого региона быть выбранным центром следующего кластера: $P_i = R_i^2 / \sum_j R_j^2, i = 1, \dots, N$;
 - 2.4) рассчитываются накопленные вероятности выбора: $S_i = \sum_{j=1}^i P_j, i = 1, \dots, N$;
 - 2.5) генерируется равномерная случайная величина $R_{nd} = \text{random}(0, 1)$;
 - 2.6) увеличивается номер кластера ($k = k + 1$), для которого определяется центроид как регион, которому соответствует интервал накопленных частот, содержащий величину R_{nd} : $p: S_{p-1} \leq R_{nd} < S_p, \mu^k = x_p$;
 - 2.7) если $k < K$, то алгоритм повторяется (с п.п. 2.2), иначе осуществляется переход на п. 3, так как инициализированы все центры кластеров.

3. На основании минимума расстояния ρ до центров кластеров каждый регион x_i закрепляется за определенным кластером $T^k = \{x_j^k\}, k \in \{1, 2, \dots, K\}$:

$$x_i \rightarrow x_j^k, k = \arg \min_{k=1, \dots, K} (\rho(x_i, \mu^k)), i = 1, \dots, N.$$

4. Пересчитываются центры кластеров: $\mu^k = (\mu_1^k, \dots, \mu_L^k), \mu_j^k = \frac{1}{n_k} \sum_{l=1}^{n_k} x_{jl}^k, j = 1, \dots, L$.

5. Если произошло изменение центров кластеров или перераспределение объектов, то осуществляется переход на п. 3, если не произошло, то полученные центры кластеров $\mu^k, k = 1, \dots, K$ и распределение объектов по кластерам $T^k = \{x_j^k\}, j = 1, \dots, n_k$ считаются оптимальными.

Численная реализация представленного алгоритма проведена по данным табл. 1 на языке Python с использованием библиотеки машинного обучения Scikit-Learn⁷.

Результаты кластерного анализа позволили классифицировать совокупность промышленных регионов за счет ее разбиения на 12 групп (кластеров) однородных по экономической специализации регионов (табл. 2).

Проверка устойчивости результата кластеризации проводилась иерархическим агломеративным методом. Схожесть результатов составила 97,5%, что подтверждает устойчивость решения, полученного на основе адаптированного алгоритма метода k-средних.

Для выявления значимых отраслей в составе каждого кластера была проведена численная оценка уровня связи между каждым кластером и критерием кластеризации. Такая оценка составлена на основании коэффициента корреляции между долями отраслей промышленности в объеме отгруженной продукции (x_{ij} из табл. 1) и фиктивной переменной, характеризующей отношение региона к определенному кластеру:

$$r_{jk} = \left(\sum_{i=1}^n (x_{ij} - \bar{x}_j) (C_{ik} - \bar{C}_k) \right) \left\{ \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (C_{ik} - \bar{C}_k)^2 \right\}^{-1/2}, \quad (3)$$

где C_k — фиктивная переменная принадлежности региона к определенному кластеру k :

$$C_{ik} = \begin{cases} 1, & \text{если регион } i \text{ вошел в кластер } k, \\ 0 & \text{иначе.} \end{cases}$$

Результат оценки приведен на рис. 2 в форме HeatMap-отображения.

Интерпретируя согласно шкале Чеддока представленные на рис. 2 результаты, можно сделать следующие заключения: большая часть кластеров включает регионы с выраженной экономической специализацией, в основном сырьевой; только в двух кластерах — (1) и (8) — региональная экономика ориентирована на высокотехнологичные отрасли промышленности.

⁷ Машинное обучение на Python (<https://scikit-learn.org>).

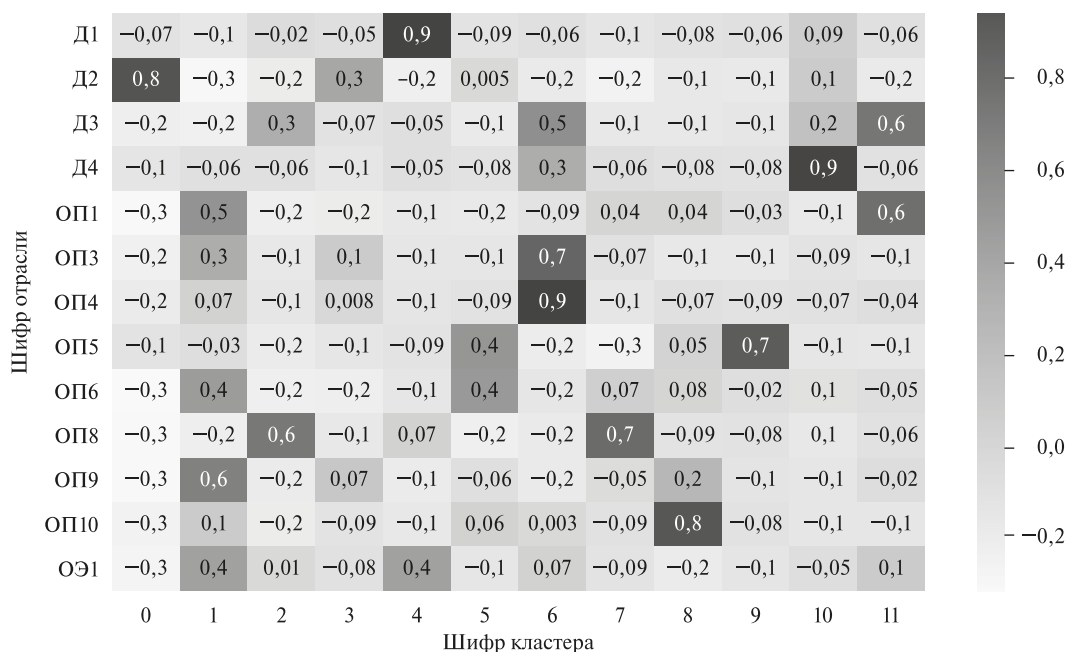


Рис. 2. Численная характеристика характера отраслевой специализации кластеров (классов) промышленно ориентированных регионов РФ

Более развернутая характеристика кластеров приведена в табл. 2. Составленная характеристика включает: а) количественные оценки уровня диверсификации экономики регионов, которые представлены средними для каждого кластера значениями коэффициента I_{HH} ($E(I_{HH})$) и значениями коэффициента вариации ($V(I_{HH})$); б) качественные (описательные) характеристики каждого кластера, в том числе идентификацию кластеров в соответствии с теоретической концепцией трехсекторной модели экономики Фишера–Кларка (Fisher, 1939) и состав отраслевых доминант; в) численные значения показателя «доля доминирующих отраслей» (d) в средней для каждого кластера оценке ($E(d)$ и $V(d)$).

Таблица 2. Характеристика кластеров промышленно ориентированных регионов РФ по экономической специализации

Кластер	Регионы РФ, сгруппированные по кластерам	Характеристика экономической специализации регионов			Доля доминирующих отраслей, d		
		Количественная оценка, I_{HH}		Описательная характеристика	Шифр	$E(d)$	$V(d)$
		$E(I_{HH})$	$V(I_{HH})$				
0	Ханты-Мансийский АО, Ямало-Ненецкий АО, Ненецкий АО; Сахалинская и Астраханская области	0,85	12	Выраженная экономическая специализация регионов; преобладает первичный сектор экономики (отрасль «Добыча НГ», «Производство кокса, нефтепродуктов»)	Д ₂ , ОП ₅	85	10
1	Владимирская, Новгородская, Саратовская, Ленинградская, Смоленская области; Республика Марий Эл	0,13	17	Широкая отраслевая диверсификация экономики регионов; преобладает вторичный сектор экономики (обрабатывающие отрасли промышленности). Значительная роль высокотехнологичных отраслей: «Производство компьютеров, электронных и оптических изделий, электрооборудования»; «Машиностроение». Развитые отрасли: «Металлургия», «Производство пищевых продуктов». При общем свойстве — широко диверсифицированная экономика и высокая роль высокотехнологичных отраслей, отсутствуют общие доминанты в экономической специализации регионов			
2	Магаданская область; Чукотский АО	0,74	5	Выраженная экономическая специализация регионов; преобладает первичный сектор экономики — основа экономики представлена ТПК, включающим элементы «Добыча металлических руд» и «Металлургия»	Д ₃ , ОП ₈	85	5

Окончание таблицы 2

Кластер	Регионы РФ, сгруппированные по кластерам	Характеристика экономической специализации регионов		Описательная характеристика	Доля доминирующих отраслей, d		
		Количественная оценка, I_{HH}			Шифр	$E(d)$	$V(d)$
		$E(I_{HH})$					
3	Республика Карелия; Иркутская, Оренбургская и Томская области; Удмуртская Республика	0,27	30	Диверсифицированная экономика регионов; преобладает первичный сектор экономики (отрасль «Добыча НГ»). Важную роль играют также отрасли обрабатывающей промышленности, связанные с первичным сектором: «Переработка нефти» (Республика Карелия, Оренбургская и Томская области) и «Металлургия» (Удмуртская Республика, Иркутская область)	D_2	46	21
4	Республика Хакасия; Кемеровская область	0,56	5	Выраженная экономическая специализация регионов; преобладает первичный сектор экономики — ТПК, включающий элементы «Добыча угля» и «Металлургия»	D_1, OP_8	64	13
5	Тюменская область без АО; Республики Башкортостан и Татарстан; Пермский край	0,47	20	Диверсифицированная экономика регионов; преобладает первичный сектор экономики — ТПК, включающий отрасли «Добыча НГ», «Переработка нефти». В Республиках Татарстан и Башкортостан и в Пермском крае в качестве доминант экономики выступает также отрасль «Машиностроение»	D_2, OP_5, OP_6	22 45	23 16
6	Республика Карелия; Архангельская область без АО	0,24	10	Диверсифицированная экономика регионов; преобладает первичный сектор экономики — ТПК «Переработка древесины, производство бумаги»	OP_3, OP_4	40	24
7	Красноярский край; Вологодская, Свердловская, Челябинская, Липецкая и Тульская области	0,31	23	Диверсифицированная экономика регионов; преобладает первичный сектор экономики (отрасль «Металлургия»). Также развиты отрасли: «Машиностроение» (8–10%) — Свердловская, Челябинская области; «Химическое производство» (17–19%) — Вологодская, Тульская области	OP_8	50	15
8	Калужская, Самарская, Нижегородская области	0,18	15	Диверсифицированная экономика регионов; преобладает вторичный сектор экономики (обрабатывающие отрасли промышленности); выраженная доминанта — «Машиностроение»	OP_{10}, OP_8	30 11	31 46
9	Волгоградская, Омская области	0,46	36	Диверсифицированная экономика регионов; преобладает первичный сектор экономики (доминирующая отрасль «Переработка нефти»)	OP_5, OP_6	64	27
10	Республика Саха (Якутия)	0,28	—	Диверсифицированная экономика регионов; представлена, в основном, первичным сектором — добывающими отраслями	D_1, D_2, D_3, D_4	88	—
11	Белгородская, Курская области	0,20	5	Диверсифицированная экономика регионов; преобладает первичный сектор — «Пищевое производство». «Добыча металлических руд»	OP_1, D_3	35 23	31 29

Источник: составлено авторами.

Из данных табл. 2 можно сделать два основных заключения.

1. Исследуемые кластеры неоднородны по уровню диверсификации экономики:

- кластеры 0, 2, 4 представляют регионы с выраженной экономической специализацией, причем основу их экономики составляют отрасли первичного сектора;
- регионы, входящие в кластеры 3, 5, 6, 7, 9, 10, 11, имеют диверсифицированную экономику, однако ее основа представлена первичным сектором;
- два кластера — 1 и 8 — характеризуются диверсифицированной экономикой регионов и преобладанием отраслей вторичного сектора.

2. Наличие выраженной общей специфики регионов внутри каждого кластера позволяет провести дифференцированную (учитывающую эту специфику) оценку социально-экономических показателей региональной экономики.

ОЦЕНКА СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ РЕГИОНАЛЬНОЙ ЭКОНОМИКИ

Для исследования были приняты формируемые в системе государственной статистики «Основные социально-экономические показатели»⁸. Интерпретация этих показателей, проводимая в целом по совокупности промышленных регионов, затруднена из-за высокого разброса их значений, что обусловлено неоднородностью регионов. Кластеризация позволила сформировать однородные по экономической специализации группы регионов и реализовать дифференцированный подход к оценке социально-экономических показателей. Для оценки были рассчитаны средние (E) по каждому кластеру значения соответствующих показателей и коэффициенты вариации (V , %) (табл. 3).

Таблица 3. Социально-экономические показатели — характеристики кластеров экономической специализации промышленно ориентированных регионов РФ (2019 г.)

Кластер	Регионы РФ, сгруппированные по кластерам	Доля занятых в общей численности населения		Среднемесячная номинальная начисленная заработная плата, тыс. руб.		ВРП на душу населения, тыс. руб./ человек		Доля инвестиций в основной капитал в общей среднегодовой сумме основного капитала		Рентабельность отгруженной продукции	
		E	V	E	V	E	V	E	V	E	V
0, 10	Ненецкий АО, Астраханская, Сахалинская области, Ямало-Ненецкий АО, Ханты-Мансийский АО, Республика Саха (Якутия)	0,59	20	76,8	29	2253	113	0,07	33	0,2	66
1	Владимирская, Смоленская, Ленинградская, Саратовская области, Республика Марий Эл	0,43	5	34,3	20	378	33	0,06	33	0,08	70
2	Магаданская область, Чукотский АО	0,59	20	101	9	158	150	0,09	37	0,27	8
3	Республика Коми, Удмуртская Республика, Оренбургская, Иркутская, Томская области	0,46	4	42,5	21	551	26	0,06	39	0,17	32
4	Республика Хакасия, Кемеровская область	0,44	5	41,2	2	463	4	0,07	33	0,07	24
5	Республики Башкортостан и Татарстан, Пермский край, Тюменская область	0,45	9	40,4	14	557	31	0,06	25	0,21	89
6	Республика Карелия, Архангельская область	0,44	1	46,2	10	465	2	0,04	1	0,2	31
7	Липецкая, Тульская, Вологодская, Свердловская, Челябинская области, Красноярский край	0,48	4	40	13	543	25	0,07	19	0,22	63
8	Калужская, Нижегородская, Самарская области	0,51	1	37,7	9	453	6	0,05	5	0,08	15
9	Волгоградская, Омская области	0,45	2	34,4	4	347	2	0,08	11	0,06	56
11	Белгородская, Курская области	0,47	4	33,7	4	488	25	0,07	33	0,22	1
В целом, по совокупности промышленно ориентированных регионов		0,47	18	48	44	622	219	0,06	33	0,17	67
В целом по РФ		0,48	16	47,9	38	579	169	0,06	49	0,23	85

Источник: расчеты авторов.

⁸ Регионы России. Социально-экономические показатели 2020, раздел 1.1 (https://gks.ru/bgd/regl/b20_14p/Main.htm).

По данным табл. 3 могут быть сделаны следующие выводы:

- коэффициенты вариации по большей части показателей имеют допустимые значения, что подтверждает однородность регионов в составе каждого кластера и возможность необходимых обобщений при анализе социально-экономических показателей; при этом из анализа были исключены несколько показателей, по которым отмечается высокий уровень *V* (эти показатели выделены в табл. 3 серым цветом);
- уровень средней заработной платы выше среднероссийского уровня только для кластеров 0 и 2; кластеры 1, 8, 9, 11 характеризуются низким уровнем этого показателя, что определяет угрозы отрицательной миграции кадров из тех регионов, которые обеспечивают технологичный потенциал промышленности страны;
- характеристика инвестиционной активности (в основной капитал) имеет очевидно недостаточный уровень, принимая во внимание важную роль основного капитала и значительный уровень износа основных фондов в промышленности;
- показатель рентабельности в промышленном секторе существенно ниже среднего по РФ уровня, особенно по кластерам 1, 4, 8, 9; это может свидетельствовать о недостаточных инвестиционных возможностях предприятий обрабатывающей промышленности в обеспечение их технологического роста.

ЗАКЛЮЧЕНИЕ

В проведенном нами исследовании составлена классификация промышленно ориентированных регионов РФ с учетом их экономической специализации. Для этого был разработан адаптированный алгоритм кластеризации и выполнена его численная реализация.

Кластеризация промышленных регионов РФ позволила выявить некоторые существенные закономерности:

- а) при высокой неоднородности промышленных регионов по уровню диверсификации экономики большая часть кластеров представлена регионами с выраженной специализацией, ориентированной на сырьевые товары, и преобладанием отраслей первичного сектора;
- б) диверсифицированная экономика и ведущее значение отраслей обрабатывающей промышленности характерны только для двух кластеров — 1, 8;
- в) для регионов, входящих в кластеры 1 и 8, характерно невысокое значение основных экономических показателей.

Содержание кластеров промышленных регионов наглядно иллюстрирует негативные проявления структурных сдвигов в национальной экономике и диспропорции ее пространственного развития, отмечаемые многими российскими учеными (Вызовы и политика..., 2020; Лившиц, 2013; и др.). В частности, из результатов проведенного нами кластерного анализа явно прослеживается, что ряд регионов, относимых ранее к числу наиболее промышленно развитых и включающих в свою отраслевую специализацию (в качестве выраженных доминант) высокотехнологичные производства, в настоящее время относятся к кластерам с преобладающим первичным сектором экономики. В числе таких регионов следует отметить Удмуртскую Республику, Томскую, Оренбургскую, Свердловскую области. Одним из результатов структурных сдвигов явилось преобладание в экономическом пространстве РФ первичного сектора, включающего добывающие отрасли и низкотехнологичные отрасли обрабатывающей промышленности.

Специализация на отраслях первичного сектора экономики, как правило, редко создает предпосылки к устойчивому экономическому росту. Диверсификация экономики — расширение ее структуры — становится необходимой стадией развития промышленного сектора национальной экономики. Важную роль в диверсификации экономики играет государственная стратегия пространственного поляризованного развития (Швецов, 2018). В реализации этого приоритета необходим взвешенный подход. Чтобы сформировать полюса роста национального масштаба, необходимо учитывать сложившиеся центры концентрации производительных сил и перспективы их развития в условиях неоиндустриальной экономики (Минакир, 2018; Швецов, 2018 и др.). Разработанный алгоритм кластеризации позволяет систематизировать исследования по данному вопросу, увидеть состав промышленных регионов с позиции перспектив их развития и актуальности государственных регулирующих воздействий.

Таким образом, представленный в статье адаптированный алгоритм кластеризации позволил классифицировать промышленные регионы РФ по характеру их специализации и обозначить ряд важных закономерностей. Целесообразность продолжения данного исследования видится, в частности, в необходимости расширить пространство критериев кластеризации, а также использовать подход кластерной динамики для уточнения причин перехода регионов из одних кластеров в другие, что позволит повысить эффективность государственной политики индустриального развития РФ.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Айвазян С.А., Афанасьев М.Ю., Кудров А.В.** (2016). Метод кластеризации регионов РФ с учетом отраслевой структуры ВРП // *Прикладная эконометрика*. № 1 (41). С. 24–46. [Aivazian S., Afanasiev M., Kudrov A. (2016). Clustering methodology of the Russian Federation regions with account of sectoral structure of GRP. *Applied Econometrics*, 1, 41, 24–46 (in Russian).]
- Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.** (1989). Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с. [Aivazyan S.A., Bukhshtaber V.M., Enjukov I.S., Meshalkin L.D. (1989). *Applied statistics. Classification and dimension reduction*. Moscow: Finansy i statistika (in Russian).]
- Вызовы и политика пространственного развития России в XXI веке (2020). В.М. Котляков, А.Н. Швецов, О.Б. Глезер (ред.). М.: Товарищество научных изданий КМК. 365 с. [Challenges and policy of Russia's spatial development in the 21st century (2020). V.M. Kotlyakov, A.N. Shvetsov, O.B. Glezer (Eds). Moscow: KMK Scientific Press Ltd. 365 p. (in Russian)]
- Голова И.М., Суховой А.Ф.** (2019). Дифференциация стратегий инновационного развития с учетом специфики российских регионов // *Экономика региона*. Т. 15. № 4. С. 1294–1308. DOI: 10.17059/2019-4-25. [Golova I.M., Sukhovey A.F. (2019). Differentiation of innovative development strategies considering specific characteristics of the Russian regions. *Economy of Region*, 15, 4, 1294–1308 (in Russian). DOI: 10.17059/2019-4-25 (in Russian).]
- Лившиц В.Н.** (2013). Системный анализ рыночного реформирования нестационарной экономики России: 1992–2013. М.: ЛЕНАНД. 640 с. [Livshits V.N. (2013). *System analysis of the market reform of Russia's non-stationary economy: 1992–2013*. Moscow: LENAND. 640 p. (in Russian).]
- Локосов В.В., Рюмина Е.В., Ульянов В.В.** (2015). Региональная дифференциация показателей человеческого потенциала // *Экономика региона*. № 4. С. 185–196. DOI: 10.17059/2015-4-15 [Lokosov V.V., Ryumina Ye.V., Ulyanov V.V. (2015) Regional differentiation of human potential indicators. *Economy of Region*, 4, 185–196. DOI: 10.17059/2015-4-15 (in Russian).]
- Локосов В.В., Рюмина Е.В., Ульянов В.В.** (2019). Кластеризация регионов России по показателям качества жизни и качества населения // *Народонаселение*. Т. 22. № 4. С. 4–17. DOI: 10.24411/1561-7785-2019-00035. [Lokosov V.V., Ryumina E.V., Ulyanov V.V. (2019). Clustering of regions by indicators of quality of life and quality of population. *Population*, 22, 4, 4–17. DOI: 10.24411/1561-7785-2019-00035 (in Russian).]
- Минакир П.А.** (2018). «Стратегия пространственного развития» в интерьере концепций пространственной организации экономики // *Пространственная экономика*. № 4. С. 8–20. DOI: 10.14530/se.2018.4.008-020. [Minakir P.A. (2018). “The strategy of spatial development” in the interior of the concepts of the spatial organization of the Economy. *Spatial Economics*, 4, 8–20. DOI: 10.14530/se.2018.4.008-020 (in Russian).]
- Михеева Н.Н.** (2013). Структурные факторы региональной динамики: измерение и оценка // *Пространственная экономика*. № 1. С. 11–32. [Mikheeva N.N. (2013). Structural factors of regional dynamics: Measurement and assessment. *Spatial Economics*, 1, 11–32 (in Russian).]
- Олдендерфер М.С., Блэшфилд Р.К.** (1989). Кластерный анализ. Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика. 215 с. [Oldenderfer M.S., Blashfield R.K. (1989). *Cluster analysis. Factorial, discriminant and cluster analysis*. Moscow: Finansy i statistika (in Russian).]
- Охупкина Е.П., Охупкин В.П.** (2015). Подходы к кластеризации групп социальной сети // *Компьютерные исследования и моделирование*. Т. 7. № 5. С. 1127–1139. [Okhapkina E.P., Okhapkin V.P. (2015). Approaches to a social network groups clustering. *Computer Research and Modeling*, 7, 5, 1127–1139 (in Russian).]
- Петрыкина И.Н.** (2013). Кластерный анализ регионов Центрального федерального округа по уровню развития человеческого капитала // *Вестник ВГУ. Серия: экономика и управление*. № 1. С. 72–80. [Petrykina I.N. (2013). Cluster analysis of regions of the Central Federal District in terms of human capital development. *Proceedings of VSU. Series: Economics and Management*, 1, 72–80 (in Russian).]
- Тимошенко А.И.** (2012). Сибирские территориально-производственные комплексы: исторический опыт и новые проекты // *Историко-экономические исследования*. Т. 13. № 1. С. 71–93. [Timoshenko A.I. (2012). Siberian territorial-production complexes: Historical experience and new projects. *Historical and Economic Research*, 13, 1, 71–93 (in Russian).]

- Швецов А.Н. (2018). Инструменты политики поляризованного пространственного развития // *Федерализм*. № 1 (89). С. 82–103. [Shvetsov A.N. (2018). Policy instruments of polarized spatial development. *Federalism*, 1 (89), 82–103 (in Russian).]
- Arthur D., Vassilvitskii S. (2007). K-means++: The advantages of careful seeding. *Conference: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms, SODA 2007*. New Orleans, Louisiana, USA. DOI: 10.1145/1283383.1283494
- Fisher A.G.B. (1939). Production, primary, secondary and tertiary. *Economic Record*, 15, 1, 24–38. Available at: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1475-4932.1939.tb01015.x>

Cluster analysis and classification of Russia's industrial oriented regions by economic specialization

© 2022 O.M. Shatalova, E.V. Kasatkina, V.N. Livchits

O.M. Shatalova,

Udmurt Branch of the Institute of Economics of the Ural Branch of the Russian Academy of Sciences Izhevsk, Russia; e-mail: oshatalova@mail.ru

E.V. Kasatkina,

Kalashnikov Izhevsk State Technical University, Izhevsk, Russia; e-mail: e.v.trushkova@gmail.com

V.N. Livchits

Federal State Institution "Federal Research Center 'Informatics and Management'" of the Russian Academy of Sciences, Moscow, Russia; e-mail: livchits@isa.ru.

Received 14.09.2021

The article was prepared according to the Program of Fundamental Scientific Research of the State Academies of Sciences and the research plan of the Institute of Economics of the Ural Branch of the Russian Academy of Sciences for 2021–2023 (theme "Methodology of innovative development of region-oriented systems in an unstable economic environment").

Abstract. In the context of high socio-economic differentiation of the constituent entities of the Russian Federation (RF), the research tasks of their classification are highly relevant, since such classification contributes to the identification of homogeneous groups of regions; each group has its own particularity. The purpose of the research was to classify the industrially oriented regions of the RF by the nature of their economic specialization and to develop the necessary economic and mathematical methods. The classification of regions according to the criterion of their economic specialization was carried out using intelligent mathematical methods of cluster analysis. To solve the research problem, an adapted k-means algorithm was developed. The result of clustering was the division of industrial regions' aggregate into twelve classes. Also, the result of the research was a meaningful interpretation of the identified clusters. In particular, a differentiated assessment of the economic indicators of the regions was compiled, taking into account the specifics of each cluster. The clustering also made it possible to compile descriptive characteristics of clusters which correspond to the concept of a three-sector model of the economy (Fisher–Clark) and was based on a quantitative assessment of the of diversification of the regions. The results of our research made it possible to clarify some regularities that are essential for the spatial development of the country's economy, which may be important for increasing the effectiveness of industrial, innovation, and fiscal policy of state administration.

Keywords: regional economy, sustainable development, spatial development, industrially oriented regions, economic specialization, diversification of the region's economy, cluster analysis, machine learning.

JEL Classification: C38, O14, R11.

Quoting: Shatalova O.M., Kasatkina E.V., Livchits V.N. (2022). Cluster analysis and classification of Russia's industrial oriented regions by economic specialization. *Economics and Mathematical Methods*, 58, 1, 80–91. DOI: 10.31857/S042473880018971-7