

ЭКОНОМИКА

Лободина Ольга Николаевна

аспирант

ФГАОУ ВПО «Дальневосточный федеральный университет»

г. Владивосток, Приморский край

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДИК ОЦЕНКИ УРОВНЯ ИННОВАЦИОННОГО РАЗВИТИЯ ТЕРРИТОРИЙ

***Аннотация:** в статье рассматриваются методические схемы построения рейтингов развития регионов РФ, которые затем сравниваются между собой методами кластерного анализа. Оценивается влияние разного набора показателей методик, характеризующих инновационное развитие, и алгоритмов, используемых при составлении рейтингов.*

***Ключевые слова:** уровень инновационного развития, кластеризация, метрики кластеризации, алгоритмы кластеризации.*

Рейтинги инновационного развития регионов на региональном и федеральном уровнях используются для [1]:

- оценка проводимой в субъекте Российской Федерации политики по стимулированию инновационной деятельности;
- определение группы регионов-лидеров по уровню инновационного развития;
- анализ факторов успеха отдельных регионов в сфере инновационной деятельности и распространение лучшей практики;
- использование результатов оценки инновационного развития субъектов РФ при распределении субсидий и дотаций из федерального бюджета.

Однако в настоящее время нет законодательно оформленной методики оценки инновационной развития, а полученные различными методиками рейтинги значительно различаются. Естественно, что региональные власти будут

опираться на тот рейтинг, в котором их регион находится на более высоком месте, а остальные оспаривать. На федеральном уровне для принятия решений в зависимости от различных причин могут быть выбраны различные рейтинги, что также не способствует доверию к принятым решениям. Поэтому актуальной представляется задача провести сравнение методических схем расчета рейтингов инновационного развития, чтобы выявить проблемы и обобщить инструменты, используемые при формировании рейтингов.

В первой главе данной работы были рассмотрены несколько методик оценки инновационного развития регионов:

– «Рейтинг инновационного развития субъектов РФ» [3], разработанный Институтом статистических исследований и экономики знаний Национального исследовательского университета «Высшая школа экономики» (I);

– «Системы оценки инновационного развития регионов», разработанные Ассоциацией инновационных регионов под руководством Бортника И.М. (II, III) [1; 8];

– рейтинг оценки, основанный на концепции региональных инновационных систем (1 способ), представленный в работе Н.Н. Михеевой (IV) [2];

– рейтинг оценки, основанный на концепции региональных инновационных систем (2 способ), представленный в работе Н.Н. Михеевой (V) [2].

Эти методики обозначены соответственно: I, II, III, IV, V. и будут использоваться для компактного представления данных в таблицах.

Во всех рассмотренных методиках различаются входные данные по составу показателей и их количеству, а также методические схемы построения рейтинга. Результаты рейтинга, хотя и имеют некоторые сходства, однако они тоже значительно различаются. Например, по результатам методики I Чувашская республика относится к группе сильных инноваторов и занимает общее 6 место, согласно же рейтингу III, этот регион относится к группе среднесильных инноваторов и занимает общее 18 место, Приморский край по обеим методикам входит в группу средних инноваторов и занимает 52-ое и 42-ое место соответственно (рейтинги построены по данным за 2012–2013 годы).

Проанализировав методические схемы рассматриваемых методик, был построен обобщенный универсальный алгоритм для создания рейтингов (рисунок 1). Все методические схемы можно описать с помощью этого алгоритма, при этом часть этапов в некоторых методиках будет пропущена, а часть может выполняться различными инструментами. Например, методику I, с помощью представленной схемы, можно описать следующим образом: этап 1 (1.1, 1.2, 1.3, 1.4, 1.7) → этап 2 (2.1, 2.2) → этап 3 (3.1) → этап 4 (4.1) → этап 6 (6.1) → этап 8 (8.1) → этап 9 (9.1, 9.2). Аналогично описываются все остальные методики.

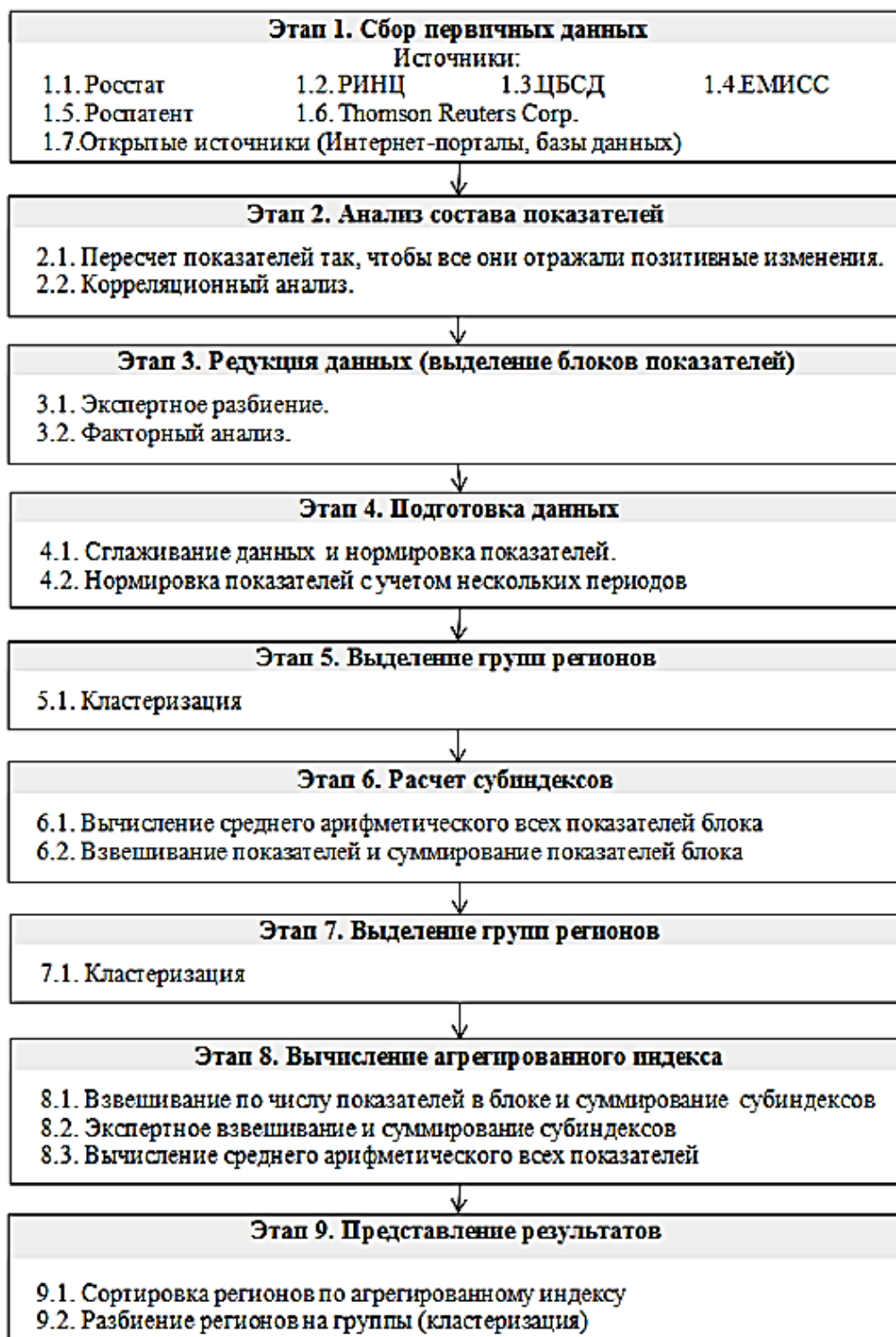


Рис. 1. Обобщенный алгоритм методик построения рейтинга инновационного развития регионов

В работах [5] рассматриваются различные инструменты для оценки методических схем, однако в большинстве своем они анализируют зависимость конеч-

ного результата рейтинга от отдельно взятых инструментов (нормирования, взвешивания показателей). Суть оценки заключается в том, что для некоторых показателей выполняется нормировка и взвешивание по различным алгоритмам, а затем результаты сравниваются, и визуально определяется их близость.

В данной работе предлагается оценку методик рейтинга инновационного развития провести методами кластерного анализа, используя следующие инструменты [4; 6].

Компактность (homogeneity) – элементы одного кластера должны быть как можно ближе друг к другу. Компактность можно вычислить по следующей метрике:

$$H = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^n u_{ij} \cdot \|x_i - \mu_j\|^2.$$

Отделимость (separation) – элементы разных кластеров должны быть как можно дальше друг от друга, соответствующая метрика:

$$S = \sum_{l=1}^k \frac{1}{\sum_{j \neq l} n_j \cdot n_l} \sum_{j \neq l} n_j \cdot n_l \cdot \|\mu_j - \mu_l\|^2$$

Чем значение H меньше, тем качество кластеризации лучше, чем значение S больше, тем качество кластеризации лучше.

Ещё одним способом оценки качества кластеризации является расстояние между кластеризациями, обозначается IV [7]. Если исходное множество кластеризовать несколькими методами, получив кластеризации \mathbb{C} и \mathbb{C}' , то чем меньше расстояние между ними, тем качество кластеризации лучше. Метрика для вычисления расстояния следующая:

$$VI(\mathbb{C}, \mathbb{C}') = H(\mathbb{C}) + H(\mathbb{C}') - 2 \cdot I(\mathbb{C}, \mathbb{C}'),$$

$$H(\mathbb{C}) = - \sum_{j=1}^k P(j) \log(P(j)), \quad I(\mathbb{C}, \mathbb{C}') = \sum_{j=1}^k \sum_{j'=1}^{k'} P(j, j') \cdot \log \left(\frac{P(j, j')}{P(k) \cdot P'(j')} \right),$$

$$P(j) = \frac{n_j}{n}, \quad P(j, j') = \frac{|C_j \cap C_{j'}|}{n}, \quad n_j = \sum_{i=1}^n u_{ij}.$$

Для большинства этапов методик происходит некоторое преобразование входных данных (рисунок 2). Также известно, на сколько кластеров разделены регионы в итоговом рейтинге. Следовательно, проводя кластеризацию многомерных исходных данных и результатов рейтинга можно оценить влияние инструмента на входные данные и сравнить итоговые рейтинги между собой.

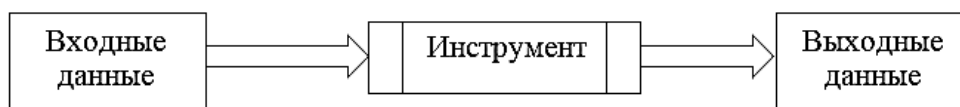


Рис. 2. Преобразование данных на каждом этапе методики

Кроме того, можно проверить их на устойчивость и чувствительность. Чтобы оценить устойчивость, необходимо осуществить разбиение данных на кластеры несколькими алгоритмами. Если разные алгоритмы выдают схожие кластеризации (оценивается метрикой расстояния IV), разбиение на кластеры можно считать устойчивым. Чувствительность же интегрального индекса проверяется исключением из исходных данных некоторого показателя. Если такое исключение не сильно влияет на конечную кластеризацию (также можно оценить метрикой IV), то можно считать индекс малочувствительным к конкретным показателям.

Анализ рейтингов, полученных несколькими методиками

Результатом применения большинства методик является рейтинг инновационного развития регионов, рассчитанный как интегральный индекс, и кластеризация регионов, полученная на разных этапах методик. В таблице 1 приведены основные характеристики исходных данных и результирующих кластеризаций.

Таблица 1

Характеристики входных данных и итоговых кластеризаций

Характеристики кластеризаций	Методики				
	I	II	III	IV	V
<i>Количество показателей</i>	36	16	23	34	34
<i>Количество кластеров</i>	4	5	5	6	6
Входные данные					
<i>H</i>	3.535E+10	1.91575E+6	9.42817E+6	2.98604E+10	2.98604E+10
<i>S</i>	4.343E+12	6.99454E+7	3.04877E+8	2.54387E+12	2.54387E+12
<i>Устойчивость</i>	1.07016	0.610562	0.844394	0.79512	0.79512
<i>Чувствительность</i>	0.936957	0.845682	0.566194	0.648213	0.648213
Итоговые кластеризации					
<i>H</i>	0.00048	0.00039	0.00054	0.741523	0.02199
<i>S</i>	0.01254	0.00979	0.01332	1.21841	0.06997
<i>Устойчивость</i>	0.315932	1.08654	0.390011	1.63796	1.44767
<i>Чувствительность</i>	0.60038	0.81272	0.495264	2.53416	1.67454
<i>IV</i>	1.86849	2.20666	2.1273	1.99187	2.70454

Также была выполнена оценка чувствительности и устойчивости результатов кластеризации и входных данных. Оценка устойчивости осуществлялась по схеме, в которой каждый набор данных разбивался на кластеры несколькими алгоритмами, после чего были вычислены попарные расстояния (метрика IV) между полученными кластеризациями и найдено среднее арифметическое полученных расстояний.

Для оценки чувствительности случайным образом были выбраны несколько показателей (5% от общего числа), которые затем были исключены из исходной системы показателей. К оставшимся показателям применялась соответствующая методика, после чего вычислялось расстояние (метрика IV) между кластеризациями полного и неполного наборов данных. Исключение показателей производилось m раз (по числу показателей в рейтинге), затем вычислялось среднее расстояние.

Для сравнения различных методик также были вычислены перекрестные расстояния между всеми кластеризациями (таблица 2), а также найдены пересечения исходных показателей (таблица 3)

Таблица 2

Попарные расстояния между кластеризациями

Методики	I	II	III	IV	V
I		2.49258	1.82589	2.34469	2.33338
II			2.28936	2.37556	2.7952
III				2.19241	2.30808
IV					2.35559

Таблица 3

Количественное пересечение входных данных рейтингов

Методики	I (36)	II (16)	III (23)	IV–V (34)
I (36)		5 (14%)	16 (45%)	18 (50%)
II (16)	5 (30%)		5 (31%)	3 (18%)
III (23)	16 (70%)	5 (21%)		10 (44%)
IV–V (34)	18 (53%)	3 (9%)	10 (30%)	

Первым (важнейшим) этапом каждой методики являлся отбор показателей для рейтинга. Разработчики рассматриваемых рейтингов [1; 2; 3; 8] указывают, что при отборе был проведен корреляционный анализ, позволивший устранить мультиколлинеарность показателей. В результате, как видно из таблицы 3, в качестве входных данных рассматривается большое число одинаковых показателей. Однако, результирующие кластеризации значительно различаются (таблица 2). Следовательно, на итоговый рейтинг влияет не только выбор показателей, но и последовательность этапов и применяемые на них инструменты. Представляется целесообразным при отборе показателей использовать метрики устойчивости и чувствительности исходных данных. Среди исследуемых методик наилучшей устойчивостью и чувствительностью обладают исходные показатели, применяемые в методиках IV и V.

Следующим этапом методик является выделение групп показателей. Как показывает оценка, лучшие результаты имеют методики, в которых число показателей по блокам не сильно различается (методики I–III).

Результатом каждой методики была кластеризация регионов, выполненная по данным разных этапов. Так методики I–III осуществляли кластеризацию по интегральному индексу, в методике IV данные кластеризовались после нормализации и сглаживания, а в методике V кластеризация проводилась по вычисленным субиндексам. Наилучшие показатели устойчивости и чувствительности имеют методики I и III, в которых кластеризации проводилась по итоговому индексу, однако по близости к входным данным методика I значительно лучше.

По близости к входным данным выделяются методики I и IV, причем в методике IV кластеризация данных проводилась сразу после нормализации и сглаживания, в то время как в I-ой методике к входным данным применялось 3 инструмента (нормализации и сглаживание, расчет субиндексов, расчет интегрального индекса). При этом расстояния, вычисленные после этапа нормализации и после вычисления интегрального индекса, практически совпадают. Это еще раз подтверждает качество применения инструментов I-ой методики. Однако для методик II, III, V расстояния между входными и результирующими данным значительно больше. Если бы для этих методик кластеризация была проведена по данным после нормализации и сглаживания, результаты были бы сравнимы с методиками I и III.

Следует отметить, что во многих зарубежных методиках оценки инновационного развития [9; 10] применяется именно такой же подход, при котором регионы разбиваются на кластеры после нормализации и сглаживания, а затем различными способами вычисляются интегральные индексы.

Список литературы

1. Бортник И. М. Система оценки и мониторинга инновационного развития регионов России / И.М. Бортник, Г.И. Сенчена, Н.Н. Михеева, А.А. Здунов, П.А. Кадочников, А.В. Сорокина // Инновации. – 2012. – №9 (167). – С. 48–61.
2. Михеева Н.Н. Сравнительный анализ инновационных систем российских регионов // Пространственная экономика. 2014. – №4. – С. 61–81.

3. Рейтинг инновационного развития субъектов Российской Федерации: Выпуск 2 / Под ред. Л.М. Гохберга. – М., Национальный исследовательский университет «Высшая школа экономики», 2014. – 88 с.
4. Сивоголовко Е. Методы оценки качества четкой кластеризации // Компьютерные инструменты в образовании. – Вып. 4 (96). – Тверь, 2011. – С. 14–31.
5. Унтура Г. А. Проблемы и инструменты аналитики инновационного развития субъектов РФ / Г.А. Унтура, Т.Н. Есикова, И.Д. Зайцев, О.Н. Морошкина // Вестник Новосибирского государственного университета. – Серия: Социально-экономические науки. – Т.14, Вып. 1, 2014. – С. 81–100.
6. Halkidi M., Batistakis Y., Vazirgiannis M.: Cluster validity methods: part II, SIGMOD Rec, 2002. – Vol. 31, №3. – P. 19–27.
7. Meilă M. Comparing clusterings – an information based distance // Journal of Multivariate Analysis. – May 2007. – Vol. 98, Issue 5. – P. 873–895.
8. Сайт Ассоциации инновационных регионов России [Электронный ресурс]. – Режим доступа: <http://www.i-regions.org/upload/internen.airr.booklet.pdf/>
9. Regional Innovation Scoreboard 2012 [Электронный ресурс]. – Режим доступа: <http://ec.europa.eu/growth/industry/innovation/facts-figures/scoreboards/index.en.htm>.
10. The Global Innovation Index 2015: Effective Innovation Policies for Development [Электронный ресурс]. – Режим доступа: <https://www.globalinnovationindex.org/content/page/gii-full-report-2015/>