

МЕТОДЫ КЛАСТЕРИЗАЦИИ ДЛЯ ОЦЕНКИ СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ РЕГИОНА

Т.А. Свиридова

Воронежский государственный технический университет

Россия, 394006, г. Воронеж, ул. 20-летия Октября, 84

E-mail: cviridova81@mail.ru

Ключевые слова: кластеризация, кластерный анализ, статистические модели исследования, методы и модели оценки социальных и экономических показателей региона.

Аннотация: В данной статье произведен краткий обзор социально-экономической ситуации РФ на макро и микроуровне, приоритетной для анализа была взята ситуация сложившаяся в регионах РФ. Обозначена общая ситуация относительно существующих методов и моделей оценки социальных и экономических показателей региона, которая диктует необходимость их совершенствования, в том числе методик статистического исследования. Обосновываются причины и актуальность применения кластерного анализа в качестве метода (модели), в зависимости от поставленной задачи, для социально-экономической оценки региона. Кластеризация актуализируется как одна из самых используемых методологий для группировки при решении социально-экономических задач.

1. Введение

Характерной особенностью социально-экономической ситуации в РФ на макро и микроуровне в настоящее время является интенсивный поиск идей, подходов, методов и средств роста темпов социально - экономического развития региона. Особое внимание в связи с этим следует уделять анализу принимаемых управленческих решений, которые должны быть гуманными и направленными, прежде всего, на обеспечение стабильного, равновесного развития области и создания социально-экономического климата, благоприятного для свободной реализации агентами рынка собственных целей и повышения качества жизни населения. [1] Описанная ситуация приводит к необходимости разработки принципиально нового или системной реорганизации старого, не содержащего элементов жесткого регулирования, инструментария системного анализа состояния и динамики развития регионов. Необходимо сказать о том, что такие инструментарии имеют место, они разрабатываются, видоизменяются ежегодно находятся новые подходы к данной оценке, если сказать емче эволюционируют. Соответственно возникает вполне логичный и закономерный вопрос зачем необходимы новые подходы, если все уже продумано и разработано. Однако все мы замечаем, что в некоторых случаях разработанные методы, модели и процедуры их приложения не всегда эффективны для решения практических задач разработки региональных программ развития, согласования интересов экономических агентов, оценки бюджетной эффективности этих программ, задач распределения господдержки и т.д. [5] Это в большинстве случаев не

является недостатком представленного метода или модели, а охарактеризовано просто частичной недостоверностью или недостаточностью исходных данных, отсутствием оперативного внесения изменений к подходам и методикам оценки уже имеющихся показателей в связи с нестабильностью и постоянной изменчивости экономической ситуации в региональной структуре РФ. [2] Именно эта изменчивость экономики диктует необходимость появления новых методов и моделей для оценки социальных и экономических показателей развития региона, а также, что зачастую является более прогрессивным направлением, применение уже отработанных методик, которые вследствие традиционности и устоявшегося мнения относительно их применения не применялись ранее для подобной оценки. К такой методике или способу оценки относится кластеризация или кластерный анализ.

2. Методы кластеризации для оценки социальных и экономических показателей региона

Актуальность данной проблемы диктует необходимость совершенствования методик статистического исследования в разрезе социального и экономического развития региона. Целью данного доклада является обратить внимание на понятие кластеризации и кластерного анализа в разрезе оценки социальных и экономических показателей развития региона. В последнее время ведущие экономисты обратили внимание на то, что именно на развитии таких кластеров сосредоточены интересы региональных политиков и всех тех, кто озабочен региональным развитием (и вообще развитием экономики). Региональные кластерные политики и стратегии обычно нацелены на эту категорию кластеров. [4] Лишь отчасти эту асимметрию интересов можно отнести на счет естественного консерватизма. Корни находятся глубже.

Кластерный анализ имеет огромное достоинство, которое заключается в том, что он производит разделение всех объектов не только по единичному параметру, а по целому ряду признаков. И, кроме того, кластерный анализ в отличие от большинства других математико-статистических методов не накладывает никаких ограничений на вид рассматриваемых объектов, и позволяет рассматривать множество исходных данных практически произвольной природы. Существует три основных метода кластерного анализа: иерархические алгоритмы, Метод К-средних, двухходовое объединение. [3] Приведенное теоретическое обоснование, безусловно, требует практического примера. Упростим пример, взяв небольшое число исходных данных для расчета кластерного анализа. Воспользуемся иерархическим алгоритмом классификации. К примеру, известно две переменные x_1 и x_2 (заработная плата работников) и количество работников – 10. Необходимо объединить данные и изучить, кому требуется повысить заработную плату. Кому необходим соц. пакет и т.д. С помощью принципа «ближайшего соседа» образуем 3 кластера. Измерения в тыс.р. Данные приведены в таблице.

Таблица 1. «Исходные данные к задаче 2»

№ п/п	1	2	3	4	5	6	7	8	9	10
x_1	5	12	34	43	21	8	32	54	14	12
x_2	13	14	56	34	7	76	23	37	23	76

1. Воспользуемся агломеративным иерархическим алгоритмом классификации. В качестве расстояния между объектами примем обычное евклидовое расстояние. Тогда согласно формуле:

$$p(x_i x_j) = \sqrt{\sum (x_{ij} x_{ji})^2}$$

где l – признаки; k – количество признаков.

$$p(x_i x_j) = \sqrt{\sum (5 - 12)^2 + (13 - 14)^2} = 7,07$$

$$p(1^3) = \sqrt{\sum (5 - 34)^2 + (13 - 56)^2} = 51,87$$

$$p(1^{344}) = \sqrt{\sum (5 - 43)^2 + (13 - 34)^2} = 43,42$$

2. Полученные данные помещаем в таблицу (матрицу расстояний).

Таблица 2. «Матрица расстояний к задаче 2».

№ п/п	1	2	3	4	5	6	7	8	9	10
1	0	7.07	51.87	43.42	17.09	63.07	28.79	54.56	13.45	63.39
2	7.07	0	47.41	36.89	11.4	62.13	21.93	47.89	9.22	62
3	51.87	47.41	0	23.77	50.7	32.8	33.06	27.59	38.59	29.73
4	43.42	36.89	23.77	0	34.83	54.67	15.56	11.4	31.02	52.2
5	17.09	11.4	50.7	34.83	0	70.21	19.42	44.6	17.46	69.58
6	63.07	62.13	32.8	54.67	70.21	0	58.18	60.31	53.34	4
7	28.79	21.93	33.06	15.56	19.42	58.18	0	26.08	18	56.65
8	54.56	47.89	27.59	11.4	44.6	60.31	26.08	0	42.38	57.31
9	13.45	9.22	38.59	31.02	17.46	53.34	18	42.38	0	53.04

3. Поиск наименьшего расстояния.

Из матрицы расстояний следует, что объекты 6 и 10 наиболее близки $P_{6,10} = 4$ и поэтому объединяются в один кластер.

Таблица 3. «Расчет к задаче 2».

№ п/п	1	2	3	4	5	[6]	7	8	9	[10]
1	0	7.07	51.87	43.42	17.09	63.07	28.79	54.56	13.45	63.39
2	7.07	0	47.41	36.89	11.4	62.13	21.93	47.89	9.22	62
3	51.87	47.41	0	23.77	50.7	32.8	33.06	27.59	38.59	29.73
4	43.42	36.89	23.77	0	34.83	54.67	15.56	11.4	31.02	52.2
5	17.09	11.4	50.7	34.83	0	70.21	19.42	44.6	17.46	69.58
[6]	63.07	62.13	32.8	54.67	70.21	0	58.18	60.31	53.34	4
7	28.79	21.93	33.06	15.56	19.42	58.18	0	26.08	18	56.65
8	54.56	47.89	27.59	11.4	44.6	60.31	26.08	0	42.38	57.31
9	13.45	9.22	38.59	31.02	17.46	53.34	18	42.38	0	53.04

При формировании новой матрицы расстояний, выбираем наименьшее значение из значений объектов № 6 и № 10. В результате имеем 9 кластеров: $S_{(1)}$, $S_{(2)}$, $S_{(3)}$, $S_{(4)}$, $S_{(5)}$, $S_{(6,10)}$, $S_{(7)}$, $S_{(8)}$, $S_{(9)}$. Из матрицы расстояний следует, что объекты 1 и 2 наиболее близки $P_{1,2} = 7.07$ и поэтому объединяются в один кластер.

Таблица 4. «Расчет к задаче 2».

№ п/п	[1]	[2]	3	4	5	6,10	7	8	9
[1]	0	7.07	51.87	43.42	17.09	63.07	28.79	54.56	13.45
[2]	7.07	0	47.41	36.89	11.4	62	21.93	47.89	9.22
3	51.87	47.41	0	23.77	50.7	29.73	33.06	27.59	38.59
4	43.42	36.89	23.77	0	34.83	52.2	15.56	11.4	31.02

При формировании новой матрицы расстояний, выбираем наименьшее значение из значений объектов № 1 и № 2.

В результате имеем 8 кластеров: $S_{(1,2)}$, $S_{(3)}$, $S_{(4)}$, $S_{(5)}$, $S_{(6,10)}$, $S_{(7)}$, $S_{(8)}$, $S_{(9)}$.

Из матрицы расстояний следует, что объекты 1,2 и 9 наиболее близки $P_{1,2,9} = 9.22$ и поэтому объединяются в один кластер.

Таблица 5. «Расчет к задаче 2».

№ п/п	[1,2]	3	4	5	6,10	7	8	[9]
[1,2]	0	47.41	36.89	11.4	62	21.93	47.89	9.22
3	47.41	0	23.77	50.7	29.73	33.06	27.59	38.59
4	36.89	23.77	0	34.83	52.2	15.56	11.4	31.02
5	11.4	50.7	34.83	0	69.58	19.42	44.6	17.46
6,10	62	29.73	52.2	69.58	0	56.65	57.31	53.04
7	21.93	33.06	15.56	19.42	56.65	0	26.08	18
8	47.89	27.59	11.4	44.6	57.31	26.08	0	42.38
[9]	9.22	38.59	31.02	17.46	53.04	18	42.38	0

При формировании новой матрицы расстояний, выбираем наименьшее значение из значений объектов №1,2 и №9. В результате имеем 7 кластеров: $S_{(1,2,9)}$, $S_{(3)}$, $S_{(4)}$, $S_{(5)}$, $S_{(6,10)}$, $S_{(7)}$, $S_{(8)}$. Из матрицы расстояний следует, что объекты 1, 2, 9 и 5 наиболее близки $P_{1,2,9,5} = 11.4$ и поэтому объединяются в один кластер.

Таблица 6. «Расчет к задаче 2».

№ п/п	[1,2,9]	3	4	[5]	6,10	7	8
[1,2,9]	0	38.59	31.02	11.4	53.04	18	42.38
3	38.59	0	23.77	50.7	29.73	33.06	27.59
4	31.02	23.77	0	34.83	52.2	15.56	11.4
[5]	11.4	50.7	34.83	0	69.58	19.42	44.6
6,10	53.04	29.73	52.2	69.58	0	56.65	57.31
7	18	33.06	15.56	19.42	56.65	0	26.08
8	42.38	27.59	11.4	44.6	57.31	26.08	0

При формировании новой матрицы расстояний, выбираем наименьшее значение из значений объектов № 1, 2, 9 и № 5. В результате имеем 6 кластера: $S_{(1,2,9,5)}$, $S_{(3)}$, $S_{(4)}$, $S_{(6,10)}$, $S_{(7)}$, $S_{(8)}$. Из матрицы расстояний следует, что объекты 4 и 8 наиболее близки $P_{4,8} = 11.4$ и поэтому объединяются в один кластер.

Таблица 7. «Расчет к задаче 2».

№ п/п	1,2,9,5	3	[4]	6,10	7	[8]
1,2,9,5	0	38.59	31.02	53.04	18	42.38
3	38.59	0	23.77	29.73	33.06	27.59
[4]	31.02	23.77	0	52.2	15.56	11.4
6,10	53.04	29.73	52.2	0	56.65	57.31
7	18	33.06	15.56	56.65	0	26.08
[8]	42.38	27.59	11.4	57.31	26.08	0

При формировании новой матрицы расстояний, выбираем наименьшее значение из значений объектов № 4 и № 8. В результате имеем 5 кластеров: $S_{(1,2,9,5)}$, $S_{(3)}$, $S_{(4,8)}$, $S_{(6,10)}$, $S_{(7)}$. Из матрицы расстояний следует, что объекты 4, 8 и 7 наиболее близки $P_{4,8,7} = 15.56$ и поэтому объединяются в один кластер.

Таблица 8. «Расчет к задаче 2».

№ п/п	1,2,9,5	3	[4,8]	6,10	[7]
1,2,9,5	0	38.59	31.02	53.04	18
3	38.59	0	23.77	29.73	33.06
[4,8]	31.02	23.77	0	52.2	15.56
6,10	53.04	29.73	52.2	0	56.65
[7]	18	33.06	15.56	56.65	0

При формировании новой матрицы расстояний, выбираем наименьшее значение из значений объектов № 4, 8 и № 7. В результате имеем 4 кластера: $S_{(1,2,9,5)}$, $S_{(3)}$, $S_{(4,8,7)}$, $S_{(6,10)}$. Из матрицы расстояний следует, что объекты 1, 2, 9, 5 и 4, 8, 7 наиболее близки $P_{1,2,9,5,4,8,7} = 18$ и поэтому объединяются в один кластер.

Таблица 9. «Расчет к задаче 2».

№ п/п	[1,2,9,5]	3	[4,8,7]	6,10
[1,2,9,5]	0	38.59	18	53.04
3	38.59	0	23.77	29.73
[4,8,7]	18	23.77	0	52.2
6,10	53.04	29.73	52.2	0

При формировании новой матрицы расстояний, выбираем наименьшее значение из значений объектов № 1, 2, 9, 5 и № 4, 8, 7. В результате имеем 3 кластера: $S_{(1,2,9,5,4,8,7)}$, $S_{(3)}$, $S_{(6,10)}$.

Таблица 10. «Расчет к задаче 2».

№ п/п	1,2,9,5,4,8,7	3	6,10
1,2,9,5,4,8,7	0	23.77	52.2
3	23.77	0	29.73
6,10	52.2	29.73	0

Таким образом, стоит отметить, что 1 группу (кластер) составляют те рабочие, которые явно недовольны своей заработной платой и их стоит поощрять к труду, следовательно, им требуется повысить заработную плату, им же необходим соц. пакет. Возможно, это новые рабочие работают не на полную ставку. Вторая группа, пожалуй, самая довольная. В нее вошел только один рабочий (по списку третий). И в итоге третий кластер – это рабочие под номером 6 и 10. Эти рабочие вероятнее всего будут довольны своей заработной платой и дополнительные поощрения им не требуются. [3] Соизмеримо с приведенным практическим применением такое направление расчета будет достаточно достоверным и для охвата более объемных сегментов исследования социально-экономического развития региона.

3. Заключение

Подводя итог вышесказанного можно утверждать, что кластерный анализ отлично зарекомендовал себя во всех сферах жизнедеятельности человека.

Прелесть предоставленного способа – он работает даже тогда, когда данных недостаточно и не выполняются запросы нормальности распределений случайных величин и остальных запросов классических способов статистического разбора. [3]

Таким образом, можно сделать вывод, что, несмотря на достаточно огромное количество существующих способов группировки, кластерный анализ остается одной из самых актуальных методологий для оценки социальных и экономических показателей развития региона при постановке социально-экономических задач и бесспорно имеет право на дальнейшую модификацию и развитие.[3]

Список литературы

1. Буреева Н.Н. Многомерный статистический анализ с использованием ППП «STATISTICA». 2007.
2. Савченко Е. А. Кластерный анализ как метод управления дебиторской задолженностью организации // Концепт. 2013. № 12. ART 13266. 0,4 п.л. URL: <http://e-koncept.ru/2013/13266.htm>. Гос. рег. Эл № ФС 77-49965. ISSN 2304-120X (научная публикация автора).
3. Свиридова Т.А., Глушенкова У.Г. Кластерный анализ, как перспективный многомерный метод классификации в социально-экономической сфере // Сборник «Моделирование и наукоемкие информационные технологии в технических и социально-экономических системах». Труды IV Всероссийской научно-практической конференции с международным участием / Под общей редакцией В.П. Цымбала, Т.В. Киселевой. 2016. С. 309-316.
4. Свиридова Т.А., Глушенкова У.Г., Карасева О.В., Пахомова А.В. Кластерный анализ, как перспективный многомерный метод классификации в строительстве // Студенческий научный вестник Воронежского государственного архитектурно-строительного университета. Серия: Управление строительством и недвижимостью. 2016. № 1 (2). С. 80-90.
5. Свиридова Т.А., Агеева А.В., Котова И.А. Роль статистических показателей при анализе финансово-хозяйственной деятельности организации в условиях современной рыночной экономики // Научный вестник Воронежского государственного архитектурно-строительного университета. Серия: Управление строительством. 2015. № 2 (7). С. 7-12.