

К ПРОБЛЕМЕ КЛАСТЕРИЗАЦИИ БИНАРНЫХ МАТРИЦ НА ПРИМЕРЕ ЗАДАЧИ ПРОСТРАНСТВЕННОГО ЭКОНОМИЧЕСКОГО АНАЛИЗА**Владимир Михайлович МОСКОВКИН^а, Эрнелту КАЗИМИРУ^б**

^а доктор географических наук, профессор кафедры мировой экономики, Белгородский государственный национальный исследовательский университет, Белгород, Российская Федерация
moskovkin@bsu.edu.ru
ORCID: отсутствует
SPIN-код: 2719-8360

^б аспирант кафедры экономики и моделирования производственных процессов, Белгородский государственный национальный исследовательский университет, Белгород, Российская Федерация
618915@bsu.edu.ru
ORCID: отсутствует
SPIN-код: отсутствует

* Ответственный автор

История статьи:

Получена 13.03.2018
Получена в доработанном виде 23.03.2018
Одобрена 30.03.2018
Доступна онлайн 29.05.2018

УДК 004.02

JEL: C51, C53, C81, C87

Ключевые слова:

матричная кластеризация, бинарная матрица, пороговая бинаризация матриц, многокритериальная бинаризация матриц, экспортная конкурентоспособность стран

Аннотация

Предмет. Бинаризация и кластеризация матриц состояний произвольных объектов на примере показателей экспортной конкурентоспособности сектора «свежие продукты» стран Субсахарской Африки.

Цели. Разработка методологии матричной кластеризации и ее апробация на примере решения задачи пространственного экономического анализа.

Методология. Предложена методология матричной кластеризации объектов, состоящая в построении матрицы их состояний, многокритериальной пороговой бинаризации матрицы состояний и кластеризации полученной бинарной матрицы на субматрицы с разной плотностью нулевых или единичных элементов.

Результаты. Разработанная методология матричной кластеризации апробирована на основе индикаторов экспортной конкурентоспособности стран Субсахарской Африки для сектора «свежие продукты» (Fresh food).

Выводы. На основе разработанной методологии матричной кластеризации, которая апробирована для показателей экспортной конкурентоспособности сектора «свежие продукты» стран Субсахарской Африки, показано, что большинство из этих стран сосредоточено во втором (количество нулевых элементов бинарной матрицы изменяется от 25 до 50%) и третьем (количество нулевых элементов бинарной матрицы изменяется от 50 до 75%) кластерах. Здесь конкурентоспособность экспортного сектора «свежие продукты» возростала в направлении от четвертого кластера к первому. Показано, что кластеризация бинарной матрицы является достаточно устойчивой к изменению пороговых критериев исходной матрицы состояний.

© Издательский дом ФИНАНСЫ и КРЕДИТ, 2018

Для цитирования: Московкин В.М., Казимиру Э. К проблеме кластеризации бинарных матриц на примере задачи пространственного экономического анализа // *Экономический анализ: теория и практика*. – 2018. – Т. 17, № 5. – С. 967 – 980.

<https://doi.org/10.24891/ea.17.5.967>

Введение

Имеется много публикаций, в которых матричная кластеризация состоит в выделении плотной субматрицы из большой разреженной бинарной матрицы, элементы которой состоят из нулей и единиц. При этом плотная

субматрица состоит преимущественно из единиц. Такие задачи возникают в Data Mining, Web-анализе и анализе изображений [1–4], анализе библиографических потоков информации [5], промышленном проектировании [6], анализе генов [7, 8] и в других областях. Для решения таких задач в

работе [3] был предложен ring-pong алгоритм, суть которого состоит в оптимальной перестановке строк и столбцов в исходной разреженной бинарной матрице. Отметим, что обзор такого рода задач приведен нами в работах [9, 10].

Примером такой задачи в пространственном экономическом анализе может служить задача построения матрицы наличия значений статистических данных по какому-либо набору показателей. Например, можно взять Европейское инновационное табло (European Innovation Scoreboard, EIS), которое представляет собой матрицу

$$I = (I_{ij}), 1 \leq i \leq m, 1 \leq j \leq n,$$

где m – количество стран в EIS;

n – количество инновационных индикаторов.

Тогда преобразование исходной матрицы (I_{ij}) в бинарную (B_{ij}) будет состоять в следующем:

$$B_{ij} = \begin{cases} 0, & \text{если не существуют статистические} \\ & \text{данные по индикатору } I_{ij} \\ 1, & \text{если существуют статистические} \\ & \text{данные по индикатору } I_{ij} \end{cases}.$$

После этого представляется возможным выделить плотную субматрицу, состоящую преимущественно из единиц, которая соответствует группе стран с наибольшим количеством индикаторов, для которых имеются статистические данные. В этом состоит суть матричной кластеризации в понимании работ [1–8].

В то же время возникает класс задач, в которых бинаризация исходной матрицы состояний может происходить с помощью критериев, наложенных на множество значений всех индикаторов. Такую задачу мы назвали задачей многокритериальной пороговой бинаризации матрицы состояний. Кластеризацию полученной бинарной матрицы предлагается проводить по интервалам изменения числа нулевых или единичных элементов этой матрицы. Опишем методологию кластеризации бинарных матриц.

Методология кластеризации бинарных матриц

Такая методология состоит из трех этапов.

1. Построение матрицы состояний

$$A = (A_{ij}) = \begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1j} & \cdots & A_{1n} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2j} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ A_{i1} & A_{i2} & A_{i3} & \cdots & A_{ij} & \cdots & A_{in} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mj} & \cdots & A_{mn} \end{pmatrix},$$

где A_{ij} – значение j -го индикатора для i -го объекта;

m – количество объектов;

n – количество индикаторов;

$$1 \leq i \leq m, 1 \leq j \leq n.$$

2. Многокритериальная пороговая бинаризация матрицы состояний.

Предположим, что на все индикаторы матрицы состояний наложены некоторые пороговые критерии K_j , которые позволяют преобразовать эту матрицу в бинарную (B_{ij}) .

Такое преобразование будет иметь вид

$$\begin{cases} (A_{ij}) \sim (B_{ij}) \\ B_{ij} = \begin{cases} 0, & \text{если } A_{ij} \leq K_j \text{ или } A_{ij} \geq K_j > 0. \\ 1, & \text{если } A_{ij} > K_j \text{ или } A_{ij} < K_j > 0 \end{cases} \end{cases}$$

Здесь при нулевом значении B_{ij} берется знак меньше или равно, если A_{ij} стимулятор, и противоположный знак больше, если A_{ij} дестимулятор. При единичном значении B_{ij} берутся противоположные знаки. Эти критерии вводятся, чтобы абстрагироваться от незначимых значений индикаторов.

3. Кластеризация бинарной матрицы.

Кластеризацию бинарной матрицы предлагается проводить так, как это представлено в *табл. 1*.

Если распределение числа нулей брать по двадцатипроцентным интервалам, то придем к разбиению бинарной матрицы на пять кластеров или бинарных субматриц, отличающихся разной плотностью нулей.

Пример кластеризации бинарной матрицы

В качестве примера возьмем базу данных Trade Competitiveness Map ВТО и применим ее для стран Субсахарской Африки. В этой базе данных имеется 14 экспортных секторов экономики, каждый из которых характеризуется 19 индикаторами. Возьмем первый сектор – Fresh food (свежие продукты). Построим для него исходную матрицу состояний (A_{ij}) на уровень 2016 г. (табл. 2). Эта матрица имеет размерность $m \times n = 45 \times 19$, где m – количество стран Субсахарской Африки, которые имели статистические данные по всем 19 индикаторам. Отметим, что Центрально-Африканская Республика, Республика Конго и Танзания не имели таких данных в рассматриваемом секторе экономики. Здесь мы не будем давать детальную расшифровку 19 индикаторов, она доступна на сайте ИТС (International Trade Center, WTO), а приведем только их оригинальные обозначения (см. примечание к табл. 2).

Приведем гипотетические критерии, полученные из эвристических соображений (они могут быть выбраны и другими).

Если $G_1 \leq 30\,000$, $G_2 \leq 2\%$, $G_3 \leq 10\%$, $G_4 \geq 10\%$, $G_5 \leq 10\%$, $G_6 \leq 1$, $P_1 \leq 50$, $P_2 \leq 10$, $P_3 \leq 0,05\%$; $P_{4a} \leq 2$, $P_{4b} \geq 100$, $P_{5a} \leq 5$, $P_{5b} \geq 100$, $C_1 \leq 5\%$, $C_{1a} \leq 5\%$, $C_{1b} \leq 2\%$, $C_{1c} \leq 5\%$, $C_{1d} \leq 5\%$, $C_2 \geq 100$, тогда $B_{ij} = 0$, в противном случае $B_{ij} = 1$.

Здесь G_1 , P_1 даны в абсолютных величинах (тыс. долл. США), G_6 , P_{4a} , P_{4b} , P_{5a} , P_{5b} , C_2 – в относительных непроцентных единицах, P_2 – в удельной единице (экспорт на душу населения), G_2 , G_3 , G_4 , G_5 , P_3 , C_{1a} , C_{1b} , C_{1c} , C_{1d} – в относительных процентных единицах, P_{4b} , P_{5b} , C_2 – в рангах (места в ранжировке всех стран мира в рассматриваемом секторе по значениям этих индикаторов). В качестве дестимуляторов брались ранговые индикаторы

P_{4b} , P_{5b} , C_2 и индикатор G_4 , отвечающий за импорт (доля рассматриваемого сектора в национальном импорте).

Наложив эти критерии на исходную матрицу состояний (табл. 2), получим бинарную матрицу (табл. 3). Кластеризация этой матрицы по четырем квартилям привела к четырем кластерам (табл. 4). Большинство стран Субсахарской Африки сосредоточено во втором и третьем кластерах. Выделенные кластеры укрупненно показывают конкурентоспособность экспортного сектора «свежие продукты» стран Субсахарской Африки, которая возрастает в направлении от четвертого кластера к первому.

Ужесточим критерии, изменив их значения в два раза за исключением G_6 : $G_1 \leq 60\,000$, $G_2 \leq 4\%$, $G_3 \leq 20\%$, $G_4 \geq 5\%$, $G_5 \leq 20\%$, $G_6 \leq 1$, $P_1 \leq 100$, $P_2 \leq 20$, $P_3 \leq 0,1\%$; $P_{4a} \leq 4$, $P_{4b} \geq 50$, $P_{5a} \leq 10$, $P_{5b} \geq 50$, $C_1 \leq 10\%$, $C_{1a} \leq 10\%$, $C_{1b} \leq 4\%$, $C_{1c} \leq 10\%$, $C_{1d} \leq 10\%$, $C_2 \geq 50$, тогда получим новую кластеризацию стран Субсахарской Африки по рассматриваемому сектору (табл. 5). Как и следовало ожидать, большая часть стран, по сравнению с предыдущей кластеризацией, перешла в менее конкурентоспособные кластеры.

Заключение

Нами предложена методология матричной кластеризации, состоящая из построения исходной матрицы состояний объектов, многокритериальной пороговой бинаризации этой матрицы и кластеризации полученной бинарной матрицы на субматрицы с разной плотностью нулевых или единичных элементов. Методология апробирована на показателях экспортной конкурентоспособности всех стран Субсахарской Африки для сектора «свежие продукты». В дальнейшем мы планируем разработать алгоритм и стандартную программу для многокритериальной пороговой бинаризации и кластеризации произвольных матриц состояний и провести расчеты для 13 остальных экспортных секторов стран Субсахарской Африки по базе данных Trade Competitiveness Map.

Таблица 1**Кластеризация бинарной матрицы (B_{ij}) по количеству нулей в ее строках****Table 1****Binary matrix clustering (B_{ij}) by the number of zeros in its lines**

Кластер	Кластеризация бинарной матрицы
Кластер 1 (Q_1 – первая квартиль)	От 0 до 25% нулей (от 0 до 4 нулей)
Кластер 2 (Q_2 – вторая квартиль)	От 25 до 50% нулей (от 5 до 9 нулей)
Кластер 3 (Q_3 – третья квартиль)	От 50 до 75% нулей (от 10 до 14 нулей)
Кластер 4 (Q_4 – четвертая квартиль)	От 75 до 100% нулей (от 15 до 19 нулей)

Источник: авторская разработка*Source:* Authoring

Таблица 2

Сектор экономики «свежие продукты» стран Субсахарской Африки, представленный в виде матрицы состояний (A_{ij}) на уровень 2016 г.

Table 2

The 'Fresh food' sector of sub-Saharan African countries presented in the form of a state matrix (A_{ij}) for 2016

Страна	G_1 , Value	G_2 , Value	G_3 , Value	G_4 , Value	G_5 , Value	G_6 , Value	P_1 , Value	P_2 , Value	P_3 , Value
А	1	2	3	4	5	6	7	8	9
Ангола	42 328	-1	0	6	-87	1,4	-620 510	1,5	0,01
Бенин	245 029	1	59	40	-62	0,9	-826 527	22,5	0,03
Ботсвана	114 647	6	1	3	-33	1,4	-113 481	50,9	0,02
Буркина-Фасо	720 421	11	28	5	57	0,8	528 273	38,6	0,1
Бурунди	61 396	-8	49	7	11	0,5	13 136	5,8	0,01
Кабо Верде	23 328	-17	37	10	-43	1,5	-35 215	43,2	0
Камерун	1 096 142	4	51	14	20	0,9	379 396	46,8	0,15
Чад	79 734	-3	5	3	55	1,1	56 798	5,5	0,01
Коморы	49 967	48	71	11	5	1,5	4 980	62,8	0,01
Демократическая Республика Конго	53 741	0	1	7	-70	1,4	-256 417	0,7	0,01
Кот-д'Ивуар	7 681 311	10	74	15	72	0,7	6 432 252	324,2	1,05
Джибути	39 219	-8	38	6	-74	1,4	-234 401	41,6	0,01
Экваториальная Гвинея	3,175	4	0	6	-9	0,9	-63 824	2,6	0
Эритрея	6 072	0	2	4	-33	1,4	-6 015	1,3	0
Эфиопия	2 214 770	-1	84%	6	33	0,8	1 110 966	21,6	0,3
Габон	20 875	-23	0	11	-83	0,9	-205 048	10,5	0
Гамбия	19 992	23	21	16	-51	0,7	-42 940	9,8	0
Гана	3 021 994	4	28	8	50	1	2 037 344	107,1	0,41
Гвинея	219 664	-8	10	10	-23	1,4	-133 343	17,7	0,03
Гвинея-Бисау	265 832	12	97	10	76	0,8	229 709	146,4	0,04
Кения	2 705 891	1	56	6	47	2	1 732 300	55,8	0,37
Лесото	55 747	8	6	9	-40	0,9	-76 537	25,3	0,01
Либерия	239 067	-6	25	1	20	1,1	79 757	51,8	0,03
Мадагаскар	811 057	25	35	6	62	1	624 509	32,6	0,11
Малави	737 905	-5	84	6	81	1,1	661 389	40,8	0,1
Мали	289 929	-13	26	6	21	0,8	100 887	16,1	0,04
Мавритания	607 716	1	35	8	52	1,1	416 749	141,3	0,08
Маврикий	231 685	17	10	14	-47	1,8	-424 134	183,4	0,03
Мозамбик	433 613	1	12	8	-4	0	-40 905	15	0,06
Намибия	780 503	-6	16	4	47	0,8	504 500	314,8	0,11
Нигер	160 462	-1	17	10	-10	0,2	-35 832	7,8	0,02
Нигерия	573 596	-56	1	6	-58	0,0	-1 587 052	3,1	0,08
Руанда	179 585	-2	28	7	12	1	39 157	15,1	0,02
Сан-Томе и Принсипи	9 063	15	86	8	-14	0,9	-3 112	45,3	0
Сенегал	574 496	8	21	12	-8	0,8	-100 564	37,7	0,08
Сейшелы	221 887	157	39	16	28	0,8	98 874	2 343,6	0,03
Сьерра-Леоне	171 550	21	36	14	10	1,2	32 356	23,2	0,02
Сомали	535 640	7	94	22	9%	0,8	96 574	37,4	0,07
Южно- Африканская Республика	5 143 152	1	6	4	19	1,5	1 669 567	92	0,7
Свазиленд	26 541	2	1	7	-61%	1,1	-84 469	19,8	0
Судан	1 131 929	-6	35	9	24	1	443 612	21,8	0,16
Того	81 604	-8	11	5	-4	0,4	-8 080	10,7	0,01
Уганда	1 002 562	4	40	4	62	0,8	773 795	24,2	0,14
Замбия	398 905	-15	7	4	41	1	232 698	24	0,05
Зимбабве	995 859	-1	35	11	23	0,9	374 312	61,7	0,14

Продолжение таблицы

Страна	P_4 , Value	P_3 , Rank	P_{5a} , Value	P_{5b} , Rank	C_1 , Value	C_{1a} , Value	C_{1b} , Value	C_{1c} , Value	C_{1d} , Value	C_2 , Rank
А	10	11	12	13	14	15	16	17	18	19
Ангола	4	118	5	125	0,08	-5,16	0,77	-14,19	18,67	26
Бенин	2	163	8	78	2,93	10,81	2,34	-6,66	-3,57	90
Ботсвана	2	151	4	142	7,49	20,13	0,23	-2,38	-10,49	159
Буркина-Фасо	3	145	6	107	12,75	19,2	-0,04	-0,22	-6,19	9
Бурунди	2	168	5	129	-4,9	4,46	-10,26	-8,93	9,83	63
Кабо Верде	3	177	1	177	-10,32	-13,99	0,34	5,52	-2,18	10
Камерун	2	150	5	117	5,31	9,63	0,83	-0,8	-4,35	38
Чад	3	142	6	102	-1,49	5,26	1,05	-7,67	-0,13	3
Коморы	2	174	4	130	82,49	50,25	1,07	31,66	-0,5	162
Демократическая Республика Конго	4	117	11	48	0,84	0,79	0,74	-1,77	1,08	12
Кот-д'Ивуар	3	130	12	41	11,86	13,33	1,49	0,84	-3,8	148
Джибути	6	94	5	124	-5,09	-9,04	-0,02	-1,18	5,15	35
Экваториальная Гвинея	3	158	2	162	4,87	-6,95	-1,29	-0,81	13,92	175
Эритрея	4	170	2	173	1,15	6,03	-2,95	4,99	-6,91	6
Эфиопия	6	89	15	18	0,1	-1,2	0,54	1,66	-0,9	146
Габон	3	152	8	98	-12,64	7,47	0,93	-8,34	-12,7	118
Гамбия	5	107	3	149	28,83	44,2	3,25	12,9	-31,52	61
Гана	2	159	10	54	5,65	2,87	0,67	3,07	-0,96	107
Гвинея	8	76	9	77	-5,26	-3,84	0,29	-17,55	15,84	15
Гвинея-Бисау	2	167	2	165	14,31	-1,31	9,72	9,36	-3,46	172
Кения	5	100	11	49	2,17	0,1	-0,03	-0,56	2,66	115
Лесото	6	103	2	159	9,61	14,7	0,41	-10,81	5,32	64
Либерия	2	149	5	127	-3,77	6,2	1,28	-4,83	-6,42	163
Мадагаскар	4	124	7	88	31,81	5,22	2,66	9,99	13,93	140
Малави	2	160	12	37	-2,83	-2,97	0,07	-2,68	2,76	138
Мали	3	141	5	110	-8,35	-9,93	-0,08	-17,73	19,39	128
Мавритания	8	153	8	82	2,69	-1,27	-1,73	-1,56	7,26	111
Маврикий	6	90	10	66	21,04	9,74	0,11	7,15	4,04	156
Мозамбик	4	127	19	8	2,92	7,88	1,43	-0,12	-6,26	129
Намибия	9	61	7	95	-3,66	-4,11	-0,28	-1,87	2,6	143
Нигер	2	162	2	157	-0,22	8,34	-3,91	11,41	-16,06	57
Нигерия	4	129	7	100	19,23	-18,75	-3,55	-10,61	13,68	167
Руанда	5	109	4	131	-0,46	3,94	-3,43	-4,42	3,45	91
Сан-Томе и Принсипи	1	176	1	175	18,17	-16,5	0,66	5,48	28,53	177
Сенегал	16	35	15	21	8,99	4,24	1,13	5,67	-2,05	60
Сейшелы	3	148	5	113	919,92	947,54	1,84	7,75	-37,21	17
Сьерра-Леоне	3	155	2	161	26,86	-1,65	0,24	4,14	24,12	44
Сомали	4	116	2	152	7,85	-0,05	2,69	2,84	2,37	11
Южно- Африканская Республика	24	17	22	4	2,26	2,9	-0,58	3,34	-3,39	33
Свазиленд	8	73	6	108	3,74	6,73	0,69	-1,14	-2,54	100
Судан	5	99	4	132	-3,54	-3,22	-2,54	-3,65	5,87	126
Того	2	156	11	64	-5,18	-3,33	2,67	-4,07	-0,45	13
Уганда	6	83	12	39	4,85	3,75	-0,4	-1,13	2,63	105
Замбия	4	114	8	80	-9,16	-3,56	-1,97	2,65	-6,29	56
Зимбабве	1	173	1	172	-0,37	-1,82	0,7	-1,55	2,31	124

Примечание. Оригинальные обозначения индикаторов: *Current performance*: P_1 – Value of net exports, thousand USD; P_2 – Per capita exports, USD; P_3 – Share in the world market, percentage share of world exports; P_{4a} – Product diversification, number of equivalent products; P_{4a} – Product diversification, number of equivalent products; P_{4b} – Product concentration, Spread; P_{5a} – Market diversification, number of equivalent markets; P_{5b} – Market concentration, Spread;

General profile: G_1 – Value of exports, thousand USD; G_2 – Trend growth of exports (last 5 years), percentage; G_3 – Share in national exports, percentage; G_4 – Share in national imports, percentage; G_5 – Growth in per capita exports (last 5 years), percentage; G_6 – Level in relative unit values (world average = 1); *Decomposition of changes in world market share (last 5 years):* C_1 – Relative change of world market share, decomposed into: C_{1a} – Competitiveness effect, percentage; C_{1b} – Initial geographic specialization, percentage; C_{1c} – Initial product Specialization, percentage; C_{1d} – Adaptation effect, percentage; C_2 – Matching with dynamics of world demand

Источник: авторская разработка

Source: Authoring

Таблица 3**Бинарная матрица, полученная с помощью наложения гипотетических критериев на исходную матрицу состояний****Table 3****A binary matrix built by applying hypothetic criteria to the initial state matrix**

Страна	G_1	G_2	G_3	G_4	G_5	G_6	P_1	P_2	P_3
А	1	2	3	4	5	6	7	8	9
Ангола	1	0	0	1	0	1	0	0	0
Бенин	1	0	1	0	0	0	0	1	0
Ботсвана	1	1	0	1	0	1	0	1	0
Буркина-Фасо	1	1	1	1	1	0	1	1	1
Бурунди	1	0	1	1	1	0	1	0	0
Кабо Верде	0	0	1	0	0	1	0	1	0
Камерун	1	1	1	0	1	0	1	1	1
Чад	1	0	0	1	1	1	1	0	0
Коморы	1	1	1	0	0	1	1	1	0
Демократическая Республика Конго	1	0	0	1	0	1	0	0	0
Кот-д'Ивуар	1	1	1	0	1	0	1	1	1
Джибути	1	0	1	1	0	1	0	1	0
Экваториальная Гвинея	0	1	0	1	0	0	0	0	0
Эритрея	0	0	0	1	0	1	0	0	0
Эфиопия	1	0	1	1	1	0	1	1	1
Габон	0	0	0	0	0	0	0	1	0
Гамбия	0	1	1	0	0	0	0	0	0
Гана	1	1	1	1	1	0	1	1	1
Гвинея	1	0	0	0	0	1	0	1	0
Гвинея-Бисау	1	1	1	0	1	0	1	1	0
Кения	1	0	1	1	1	1	1	1	1
Лесото	1	1	0	1	0	0	0	1	0
Либерия	1	0	1	1	1	1	1	1	0
Мадагаскар	1	1	1	1	1	0	1	1	1
Малави	1	0	1	1	1	1	1	1	1
Мали	1	0	1	1	1	0	1	1	0
Мавритания	1	0	1	1	1	1	1	1	1
Маврикий	1	1	0	0	0	1	0	1	0
Мозамбик	1	0	1	1	0	0	0	1	1
Намибия	1	0	1	1	1	0	1	1	1
Нигер	1	0	1	0	0	0	0	0	0
Нигерия	1	0	0	1	0	0	0	0	1
Руанда	1	0	1	1	1	0	1	1	0
Сан-Томе и Принсипи	0	1	1	1	0	0	0	1	0
Сенегал	1	1	1	0	0	0	0	1	1
Сейшелы	1	1	1	0	1	0	1	1	0
Сьерра-Леоне	1	1	1	0	0	1	1	1	0
Сомали	1	1	1	0	0	0	1	1	1
Южно-Африканская Республика	1	0	0	1	1	1	1	1	1
Свазиленд	0	0	0	1	0	1	0	1	0
Судан	1	0	1	1	1	0	1	1	1
Того	1	0	1	1	0	0	0	1	0
Уганда	1	1	1	1	1	0	1	1	1
Замбия	1	0	0	1	1	0	1	1	0
Зимбабве	1	0	1	0	1	0	1	1	1

Продолжение таблицы

Страна	P_4	P_5	P_{5a}	P_{5b}	C_1	C_{1a}	C_{1b}	C_{1c}	C_{1d}	C_2
А	10	11	12	13	14	15	16	17	18	19
Ангола	1	0	0	0	0	0	0	0	1	1
Бенин	0	0	1	1	0	1	1	0	0	1
Ботсвана	0	0	0	0	1	1	0	0	0	0
Буркина-Фасо	1	0	1	0	1	1	0	0	0	1
Бурунди	0	0	0	0	0	0	0	0	1	1
Кабо Верде	1	0	0	0	0	0	0	1	0	1
Камерун	0	0	0	0	1	1	0	0	0	1
Чад	1	0	1	0	0	1	0	0	0	1
Коморы	0	0	0	0	1	1	0	1	0	0
Демократическая Республика Конго	1	0	1	1	0	0	0	0	0	1
Кот-д'Ивуар	1	0	1	1	1	1	0	0	0	0
Джибути	1	1	0	0	0	0	0	0	1	1
Экваториальная Гвинея	1	0	0	0	0	0	0	0	1	0
Эритрея	1	0	0	0	0	1	0	0	0	1
Эфиопия	1	1	1	1	0	0	0	0	0	0
Габон	1	0	1	1	0	1	0	0	0	0
Гамбия	1	0	0	0	1	1	1	1	0	1
Гана	0	0	1	1	1	0	0	0	0	0
Гвинея	1	1	1	1	0	0	0	0	1	1
Гвинея-Бисау	0	0	0	0	1	0	1	1	0	0
Кения	1	0	1	1	0	0	0	0	0	0
Лесото	1	0	0	0	1	1	0	0	1	1
Либерия	0	0	0	0	0	1	0	0	0	0
Мадагаскар	1	0	1	1	1	1	1	1	1	0
Малави	0	0	1	1	0	0	0	0	0	0
Мали	1	0	0	0	0	0	0	0	1	0
Мавритания	1	0	1	1	0	0	0	0	1	0
Маврикий	1	1	1	1	1	1	0	1	0	0
Мозамбик	1	0	1	1	0	1	0	0	0	0
Намибия	1	1	1	1	0	0	0	0	0	0
Нигер	0	0	0	0	0	1	0	1	0	1
Нигерия	1	0	1	0	1	0	0	0	1	0
Руанда	1	0	0	0	0	0	0	0	0	1
Сан-Томе и Принсипи	0	0	0	0	1	0	0	1	1	0
Сенегал	1	1	1	1	1	0	0	1	0	1
Сейшелы	1	0	0	0	1	1	0	1	0	1
Сьерра-Леоне	1	0	0	0	1	0	0	0	1	1
Сомали	1	0	0	0	1	0	1	0	0	1
Южно-Африканская Республика	1	1	1	1	0	0	0	0	0	1
Свазиленд	1	1	1	0	0	1	0	0	0	0
Судан	1	1	0	0	0	0	0	0	1	0
Того	0	0	1	1	0	0	1	0	0	1
Уганда	1	1	1	1	0	0	0	0	0	0
Замбия	1	0	1	1	0	0	0	0	0	1
Зимбабве	0	0	0	0	0	0	0	0	0	0

Источник: авторская разработка

Source: Authoring

Таблица 4

Кластеризация бинарной матрицы

Table 4

Binary matrix clustering

Кластер	Страны
I	Мадагаскар
II	Буркина-Фасо, Камерун, Кот-д'Ивуар, Эфиопия, Гана, Кения, Малави, Мавритания, Маврикий, Намибия, Сенегал, Сейшелы, Сьерра-Леоне, Сомали, ЮАР, Судан, Уганда
III	Ангола, Бенин, Ботсвана, Бурунди, Кабо Верде, Чад, Коморы, Демократическая Республика Конго, Джибути, Эритрея, Габон, Гамбия, Гвинея, Гвинея-Бисау, Лесото, Либерия, Мали, Мозамбик, Нигер, Нигерия, Руанда, Сан-Томе и Принсипи, Свазиленд, Того, Замбия, Зимбабве
IV	Экваториальная Гвинея

Источник: авторская разработка

Source: Authoring

Таблица 5

Кластеризация бинарной матрицы по сектору экономики «свежие продукты» при изменении критериев

Table 5

Binary matrix clustering of the ‘Fresh food’ sector when changing the criteria

Кластер	Страны
I	-
II	Кот-д’Ивуар, Кения, Уганда, ЮАР
III	Ботсвана, Буркина-Фасо, Камерун, Чад, Коморы, Джибути, Эфиопия, Гамбия, Гана, Гвинея, Гвинея-Бисау, Либерия, Мадагаскар, Малави, Мали, Мавритания, Маврикий, Сан-Томе и Принсипи, Сенегал, Сейшелы, Сьерра-Леоне, Сомали, Намибия, Замбия, Зимбабве, Судан
IV	Ангола, Бенин, Бурунди, Кабо Верде, Демократическая Республика Конго, Экваториальная Гвинея, Эритрея, Габон, Лесото, Мозамбик, Нигер, Нигерия, Руанда, Свазиленд, Того

Источник: авторская разработка

Source: Authoring

Список литературы

1. Oyanagi S., Kubota K., Nakase A. Matrix Clustering: A New Data Mining Method for CRM. *Trans. IPSJ*, 2001, vol. 42(8), pp. 2156–2166.
2. Oyanagi S., Kubota K., Nakase A. Application of Matrix Clustering to Web Log Analysis and Access Prediction. In: WEBKDD 2001 – Mining Web Log Data Across All Customers Touch Points, Third International Workshop, 2001, pp. 13–21.
3. Oyanagi S., Kubota K., Nakase A. Mining WWW Access Sequence by Matrix Clustering. In: WEBKDD 2002 – Mining Web Data for Discovering Usage Patterns and Profiles, LNAI 2703, 2003, pp. 119–136.
4. Sudhamathy G., Venkateswaran C.J. Matrix Based Fuzzy Clustering for Categorization on Web Users and Web Pages. *International Journal of Computer Applications*, 2012, vol. 43, no. 14, pp. 43–47. URL: <https://doi.org/10.5120/6175-8602>
5. Kuo J.J., Zhang Y.J. A Library Recommender System Using Interest Change over Time and Matrix Clustering. Taipei, Taiwan, 2012, pp. 259–268.
6. Nagaraj G., Sheik Syed Abuthahir S., Manimaran A., Paramasamy S. Comparison of Matrix Clustering Methods to Design Cell Formation. *International Journal of Applied Engineering Research*, 2015, vol. 10, iss. 28, pp. 21900–21904.
7. Zhang Z.Y., Li T., Ding C., Ren X.W., Zhang S. Binary Matrix Factorization for Analyzing Gene Expression Data. *Data Mining and Knowledge Discovery*, 2010, vol. 20, iss. 1, pp. 28–52. URL: <https://doi.org/10.1007/s10618-009-0145-2>
8. Oyelade J., Isewon I., Oladipupo F. et al. Clustering Algorithms: Their Application to Gene Expression Data. *Bioinformatics and Biology Insights*, 2016, vol. 10, pp. 237–253.
9. Московкин В.М., Казимиру Э. Кластеризация многомерных объектов различной природы: постановка исследовательской задачи // Современные проблемы социально-экономических систем в условиях глобализации. Белгород: НИУ БелГУ, 2017. С. 27–30.
10. Московкин В.М., Казимиру Э. Матричная кластеризация как кластеризация матриц одинаковой размерности // Научные ведомости Белгородского государственного университета. Сер.: Экономика. Информатика. 2017. № 23. Вып. 44. С. 123–127.

Информация о конфликте интересов

Мы, авторы данной статьи, со всей ответственностью заявляем о частичном и полном отсутствии фактического или потенциального конфликта интересов с какой бы то ни было третьей стороной, который может возникнуть вследствие публикации данной статьи. Настоящее заявление относится к проведению научной работы, сбору и обработке данных, написанию и подготовке статьи, принятию решения о публикации рукописи.

ON THE PROBLEM OF CLUSTERING BINARY MATRICES: THE SPATIAL ECONOMIC ANALYSIS CASE

Vladimir M. MOSKOVKIN^{a*}, Herinelto CASIMIRO^b

^a Belgorod State University, Belgorod, Russian Federation
moskovkin@bsu.edu.ru
ORCID: not available

^b Belgorod State University, Belgorod, Russian Federation
618915@bsu.edu.ru
ORCID: not available

* Corresponding author

Article history:

Received 13 March 2018
Received in revised form
23 March 2018
Accepted 30 March 2018
Available online
29 May 2018

JEL classification: C51, C53,
C81, C87

Keywords: matrix clustering,
binary matrix, threshold
binarization, multi-criteria
binarization, export
competitiveness

Abstract

Importance The article addresses binarization and clusterization of arbitrary objects' state matrices on the case of export competitiveness indicators in the 'Fresh food' sector of sub-Saharan African countries.

Objectives The purpose of the study is to develop a methodology of matrix clustering and to test it by solving the problem of spatial economic analysis.

Methods The article presents a methodology for matrix clusterization of objects that involves creating objects' state matrix, multi-criteria threshold binarization of state matrix and clustering the resulting binary matrix into sub-matrices with different density of zero or single elements.

Results The developed matrix clustering methodology was tested on export competitiveness indicators in the 'Fresh food' sector of sub-Saharan African countries.

Conclusions The developed matrix clustering methodology tested on export competitiveness indicators in the 'Fresh food' sector of sub-Saharan African countries shows that most of these countries are found in the second cluster (the number of zero elements in binary matrix ranges from 25 to 50%) and the third cluster (the number of zero elements in binary matrix ranges from 50 to 75%). The competitiveness of the 'Fresh food' sector grew from the fourth cluster to the first one. It is shown that binary matrix clustering is rather resistant to changes in threshold criteria in the initial state matrix.

© Publishing house FINANCE and CREDIT, 2018

Please cite this article as: Moskovkin V.M., Kazimiru E. On the Problem of Clustering Binary Matrices: The Spatial Economic Analysis Case. *Economic Analysis: Theory and Practice*, 2018, vol. 17, iss. 5, pp. 967–980.
<https://doi.org/10.24891/ea.17.5.967>

References

1. Oyanagi S., Kubota K., Nakase A. Matrix Clustering: A New Data Mining Method for CRM. *Trans. IPSJ*, 2001, vol. 42(8), pp. 2156–2166.
2. Oyanagi S., Kubota K., Nakase A. Application of Matrix Clustering to Web Log Analysis and Access Prediction. In: *WEBKDD 2001 – Mining Web Log Data Across All Customers Touch Points*, Third International Workshop, 2001, pp. 13–21.
3. Oyanagi S., Kubota K., Nakase A. Mining WWW Access Sequence by Matrix Clustering. In: *WEBKDD 2002 – Mining Web Data for Discovering Usage Patterns and Profiles*, LNAI 2703, 2003, pp. 119–136.

4. Sudhamathy G., Venkateswaran C.J. Matrix Based Fuzzy Clustering for Categorization on Web Users and Web Pages. *International Journal of Computer Applications*, 2012, vol. 43, no. 14, pp. 43–47. URL: <https://doi.org/10.5120/6175-8602>
5. Kuo J.J., Zhang Y.J. A Library Recommender System Using Interest Change over Time and Matrix Clustering. Taipei, Taiwan, 2012, pp. 259–268.
6. Nagaraj G., Sheik Syed Abuthahir S., Manimaran A., Paramasamy S. Comparison of Matrix Clustering Methods to Design Cell Formation. *International Journal of Applied Engineering Research*, 2015, vol. 10, iss. 28, pp. 21900–21904.
7. Zhang Z.Y, Li T., Ding C., Ren X.W., Zhang S. Binary Matrix Factorization for Analyzing Gene Expression Data. *Data Mining and Knowledge Discovery*, 2010, vol. 20, iss. 1, pp. 28–52. URL: <https://doi.org/10.1007/s10618-009-0145-2>
8. Oyelade J., Isewon I., Oladipupo F. et al. Clustering Algorithms: Their Application to Gene Expression Data. *Bioinformatics and Biology Insights*, 2016, vol. 10, pp. 237–253.
9. Moskovkin V.M., Casimiro H. *Klasterizatsiya mnogomernykh ob"ektov razlichnoi prirody: postanovka issledovatel'skoi zadachi. V kn.: Sovremennye problemy sotsial'no-ekonomicheskikh sistem v usloviyakh globalizatsii* [Clustering the multidimensional objects of different nature: A statement of research problem. In: Modern problems of socio-economic systems under globalization]. Belgorod, NRU BelSU Publ., 2017, pp. 27–30.
10. Moskovkin V.M., Casimiro H. [Matrix clustering as a clustering of matrices of the same dimension]. *Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta. Ser.: Ekonomika. Informatika = Belgorod State University Scientific Bulletin. Economics. Information Technologies*, 2017, no. 23, iss. 24, pp. 123–127.

Conflict-of-interest notification

We, the authors of this article, bindingly and explicitly declare of the partial and total lack of actual or potential conflict of interest with any other third party whatsoever, which may arise as a result of the publication of this article. This statement relates to the study, data collection and interpretation, writing and preparation of the article, and the decision to submit the manuscript for publication.