

# Многоуровневый профиль кластерного разбиения

Дронов С.В.

*Алтайский государственный университет, г. Барнаул*  
*Алтайский государственный технический университет, г. Барнаул*  
*dsv@math.asu.ru*

## Аннотация

Предлагается новый, по сути нечисловой, подход к изучению структуры кластерного разбиения с возможностью сравнения нескольких кластеризаций одного и того же конечного множества объектов. Подход основан на представлении кластеризуемых объектов и формирующих признаков этих объектов точками одного и того же искусственно построенного универсального пространства. При этом предложены как количественные характеристики рассматриваемых кластеризаций, так и способ чисто визуального анализа даже в случае, когда количество формирующих разбиения показателей достаточно велико, поскольку размерность универсального пространства может выбираться практически произвольно.

*Ключевые слова:* визуализация кластеров, метод нечисловой оценки структуры кластеров, однородность кластеров, сравнение кластерных разбиений, анализ соответствий.

## 1. Задача анализа кластерного разбиения

В современных статистических исследованиях интенсивно применяется аппарат так называемого кластерного анализа. Речь идет о разбиении некоторого конечного множества  $U$  объектов на непересекающиеся части. При этом объекты, оказавшиеся в одной части, должны быть в каком-либо смысле более похожи, чем объекты из разных частей этого разбиения. Процесс построения разбиения называется кластеризацией, а сами получающиеся части кластерами. При постановке и решении задач кластерного анализа, как правило, считается, что каждый из объектов задан набором некоторых своих числовых показателей  $X_1, \dots, X_p$  и, более того, объект отождествляется с точкой в  $p$ -мерном евклидовом пространстве, координатами которой являются значения этих его показателей. Образующаяся в результате кластеризация, таким образом, определяется исключительно значениями показателей  $X^{(1)}, \dots, X^{(p)}$ , которые в силу этого мы далее будем называть формирующими. Используя различные кластерные алгоритмы, описания которых в большом количестве содержатся, например, в [1, 2], для одного и того же множества объектов, исследователь может получить достаточно широкое разнообразие его кластеризаций. Исключение, видимо, возможно только в случае, когда множество  $U$  (а точнее, отождествляемое с ним множество точек  $R^p$ ) однозначно разбивается на несколько четко разделенных подмножеств. Поэтому проблема выбора наиболее подходящего кластерного разбиения в практическом исследовании может стоять достаточно остро. В связи с этим актуальной является задача сравнения между собой различных кластерных разбиений. Разумеется, сегодня разработан ряд методик для таких сравнений. К ним, например, относится сравнение отношений суммарной внутрикластерной изменчивости к изменчивости средних элементов построенных кластеров, подробно описанное в [3] и [4], а также изучение кластерных силуэтов элементов  $U$  (см., к примеру, [5, 6]). Тем не менее, современной тенденцией в любых задачах является предоставление исследователю не только количественных, но и качественных, более того, визуальных методов анализа допустимых

решений. При этом стандартные методы, упомянутые выше, используют формирующие показатели только в качестве вспомогательного инструмента, а иногда правильным должно быть отношение к ним как к объектам статистического исследования, равноправным с элементами основного множества  $U$ . Цель настоящей работы – описание нового качественного метода анализа кластерных разбиений одного и того же множества, который является в некотором смысле симметричным по отношению к объектам и формирующим признакам, – и те, и другие участвуют в предлагаемой процедуре равноправно.

## 2. Искусственное универсальное пространство

Далее, следовательно, мы будем заниматься визуализацией статистических данных в универсальном пространстве, которое по ходу дела предстоит сконструировать. Одним из традиционных требований, накладываемых на визуализацию данных, является максимально возможная разбросанность получающихся изображений объектов. Это требование возникает, для того, чтобы, при наличии нескольких решений, выбранная визуализация выглядела бы наиболее наглядной. Такая мотивация предлагается, например, при обосновании методов главных компонент и факторного анализа, подробнее в [7]. Для задач же многомерного развертывания (multidimensional unfolding problem) выдвинутое требование является принципиально необходимым. В задачах развертывания предположение о максимальном возможном минимальном расстоянии между визуализациями, наряду с очевидными, допускает и специфические толкования, см. [8].

Следующий факт на интуитивном уровне может показаться очевидным, например, в силу его близости к теореме 21 из [9, с. 158]. Но, чисто формально, мы собираемся говорить не о классической ковариационной матрице данных, а о матрице рассеивания данных относительно начала координат. Поэтому сейчас дадим точную формулировку и схему доказательства, которое фактически совпадает с доказательством из [9]. Предположим, что  $n$  точек-объектов заданы значениями  $p$  своих формирующих показателей:  $A^{(i)} = (x_{i,1}, \dots, x_{i,p})$ ,  $i = 1, \dots, n$ , где  $x_{i,j}$  – значение показателя  $X^{(j)}$  на  $i$ -м объекте. Соберем все строки  $A^{(i)}$  в матрицу данных  $F$  из  $n$  строк и  $p$  столбцов. Введем в рассмотрение  $n \times n$  матрицу  $\Phi = FF^t$ , где индекс  $t$  вверху будет обозначать транспонирование.

**Предложение 1.** *Наибольший разброс проекций точек  $p$ -мерного пространства на его  $q$ -мерное ( $q < p$ ) подпространство относительно начала координат достигается тогда, когда это подпространство натянуто на  $q$  собственных векторов матрицы рассеивания  $\Phi$ , отвечающих  $q$  ее наибольшим собственным числам.*

*Доказательство.* Рассмотрим случай  $q = 1$ . Пусть  $\vec{a} \in R^p$  – вектор единичной длины. Будем искать его координаты из заявленного условия

$$G = G(\vec{a}) = \sum_{i=1}^n \langle \vec{a}, A^{(i)} \rangle^2 \rightarrow \max. \quad (1)$$

Здесь  $\langle \cdot, \cdot \rangle$  обозначает скалярное произведение векторов. Это условие формализует максимальный разброс длин проекций радиус-векторов имеющихся точек на направление  $\vec{a}$ . Учитывая то, что искомый вектор должен иметь единичную длину, можно найти условный экстремум  $G$ , используя метод неопределенных множителей Лагранжа, т.е. изучая функцию

$$L(\vec{a}, \lambda) = G(\vec{a}) - \lambda (|\vec{a}|^2 - 1).$$

Вычисляя по координатным частным производным и приравнявая каждую из них к 0, приходим к уравнению в матричной форме  $\Phi \vec{a} = \lambda \vec{a}$ , что означает, что  $\vec{a}$  обязан быть собственным вектором  $\Phi$ . Если учесть легко проверяемое равенство

$$G(\vec{a}) = \|\Phi \vec{a}\|^2 = \lambda^2 \|\vec{a}\|^2 = \lambda^2,$$

то решением (1) служит вектор  $\vec{a}$ , отвечающий максимальному собственному числу  $\lambda$  матрицы рассеивания. Дальнейшее рассуждение проводится снижением размерности путем перехода в линейное подпространство  $R^p$ , ортогональное найденному вектору  $\vec{a}$ . Существенным моментом полного доказательства является то, что при ортогональном проецировании результат проецирования  $R^p$ , скажем на  $R^w$ , а после этого на  $R^q$  ( $p > w > q$ ) совпадает с результатом непосредственного проецирования  $R^p$  на  $R^q$ .  $\square$

При описанном в предложении проецировании максимизируется сумма расстояний проекций точек до нуля, что вовсе не обязательно означает максимизацию их разброса относительно друг друга. Этот разброс будет максимизирован, если центр множества проекций совпадет с началом координат. Мы собираемся строить проекции как строк, так и столбцов матрицы  $F$ , поэтому есть смысл начать с ее преобразования к такому виду, когда среднее каждой ее строки и каждого столбца равно 0. Подобные матрицы называют матрицами с двойным центрированием, и способ их получения известен. Следующее утверждение легко проверить вычислением сумм строк и столбцов преобразованной матрицы.

**Предложение 2.** Пусть  $F_{i,\cdot}$ ,  $F_{\cdot,j}$  – средние значения  $i$ -й строки и  $j$ -го столбца матрицы  $n \times p$  соответственно,  $F_{\cdot,\cdot}$  – среднее всех ее элементов. Тогда матрица  $\widehat{F}$  с элементами

$$\widehat{F}_{i,j} = F_{i,j} - F_{i,\cdot} - F_{\cdot,j} + F_{\cdot,\cdot}, \quad i = 1, \dots, n; \quad j = 1, \dots, p \quad (2)$$

является матрицей с двойным центрированием.

Преобразуем элементы матрицы  $F$  по формуле (2), но сохраним за этой матрицей первоначальное обозначение. Заметим, что если мы транспонируем матрицу исходных данных  $F$ , то строки новой матрицы будут соответствовать формирующим показателям изучаемых объектов и мы можем в точности так же спроецировать эти (на этот раз  $n$ -мерные точки) в пространство, пусть той же размерности  $q$ . Для этого используемое пространство натянуто на  $q$  собственных векторов, отвечающих наибольшим собственным числам матрицы рассеивания векторов-показателей  $\Phi_p = F^t F$ . Это пространство, хотя и имеет ту же размерность, что и пространство проекций векторов-объектов, не обязано совпадать с ним. Но следующий факт позволяет произвести отождествление этих пространств или, точнее, построить естественное линейное отображение одного из них на другое.

**Предложение 3.** Матрицы  $\Phi$  и  $\Phi_p$  имеют один и тот же набор ненулевых собственных чисел. Если  $\vec{u}$  – единичный собственный вектор  $\Phi$ , отвечающий ее ненулевому собственному числу  $\lambda$ , то вектор

$$\vec{v} = \frac{1}{\sqrt{\lambda}} F^t \vec{u} \quad (3)$$

является единичным собственным вектором  $\Phi_p$  и отвечает тому же собственному числу.

*Доказательство.* Заметим, что

$$\Phi_p F^t \vec{u} = F^t F F^t \vec{u} = F^t \Phi \vec{u} = \lambda F^t \vec{u}.$$

Следовательно, вектор  $F^t \vec{u}$  является собственным для  $\Phi_p$  и отвечает собственному числу  $\lambda$ . Осталось вспомнить, что вектор  $\vec{u}$  имел единичную длину, откуда

$$\|F^t \vec{u}\|^2 = \langle F^t \vec{u}, F^t \vec{u} \rangle = \langle \Phi \vec{u}, \vec{u} \rangle = \lambda \langle \vec{u}, \vec{u} \rangle = \lambda,$$

и, для получения единичного вектора, следует разделить  $F^t \vec{u}$  на  $\sqrt{\lambda}$ .  $\square$

Аналогичным образом проверяется, что, если  $\vec{v}$  единичный собственный вектор  $\Phi_p$ , отвечающий ненулевому собственному числу  $\lambda$ , то

$$\vec{u} = \frac{1}{\sqrt{\lambda}} F \vec{v} \quad (4)$$

отвечает тому же собственному числу матрицы  $\Phi$  и имеет длину 1.

Формулы (3–4) позволяют одновременно построить проекции и строк, и столбцов матрицы  $F$  (в нашем случае это объекты и формирующие показатели соответственно) в одном пространстве, которое выше было названо универсальным. Для построения этого пространства предлагается отождествить пространства одной и той же размерности с базами из собственных векторов матриц рассеивания  $\Phi$ ,  $\Phi_p$ . При этом в качестве базиса такого пространства берутся не все собственные вектора, а только два или три, которые отвечают наибольшим собственным числам. Это бывает удобно для более наглядной визуализации, да, к тому же, все равно придется исключать из числа базисных те вектора, которые отвечают нулевым собственным числам, которые обязательно имеются у каждой из матриц рассеивания  $\Phi$ ,  $\Phi_p$  в силу произведенного двойного центрирования матрицы  $F$ .

Выберем, например, в качестве универсального пространство  $R_u^q$ , натянутое на собственные векторы  $\vec{v}_1, \dots, \vec{v}_q$  матрицы  $\Phi_p$ . Тогда координаты визуализации показателей  $X_{(v)}^{(i)} = (X_{(v),1}^{(i)}, \dots, X_{(v),q}^{(i)})$  могут быть получены путем проецирования столбцов матрицы  $F$  на каждый из выбранных собственных векторов:

$$X_{(v),j}^{(i)} = \langle (x_{1,i}, \dots, x_{n,i})^t, \vec{v}_j \rangle, \quad j = 1, \dots, q, \quad i = 1, \dots, p. \quad (5)$$

Чтобы получить визуализацию объекта  $A^{(i)}$  в том же пространстве, нужно произвести проецирование  $i$ -й строки исходной матрицы  $F$  в пространство  $R_u^q$ , натянутое на собственные векторы матрицы  $\Phi$ , которые находятся применением к  $\vec{v}_1, \dots, \vec{v}_q$  формул (4). После этого полученные координаты проекции объекта  $A_{(u)}^{(i)}$  можно принять за координаты искомым его визуализации  $A_{(v)}^{(i)}$  в универсальном пространстве.

Действительно, линейное отображение  $H : R_u^q \rightarrow R_v^q$ , определяемое на базисных векторах формулой (3), переводит ортонормированный базис одного пространства в ортонормированный базис другого, а значит, является изометрическим. Таким образом, в множестве точек  $H(A_{(u)}^{(i)})$ ,  $i = 1, \dots, n$ , координаты которых в универсальном пространстве совпадают с координатами  $A_{(u)}^{(i)}$ ,  $i = 1, \dots, n$  в  $R_u^q$  соответственно, все попарные расстояния между точками не будут отличаться от расстояний между их образами.

Вместо описанного можно было произвести двойственное построение, выбирая за универсальное пространство то, которое натянуто на подходящие собственные векторы  $\Phi$ . Тогда проекции на это пространство объектов (строк матрицы  $F$ ) находятся непосредственно, а координаты визуализаций столбцов – признаков в нем должны отождествляться с соответствующими проекциями в пространство собственных векторов  $\Phi_p$ . Ясно, что результат в итоге получится тот же, что и при первом способе.

Как правило, для сокращения объема вычислений рекомендуется в качестве базиса универсального пространства использовать элементы системы собственных векторов той из матриц  $\Phi$ ,  $\Phi_p$ , размерность которой меньше, а собственные векторы двойственного пространства находить по формулам перехода (3) или (4). Собственно, та же идея используется в известном алгоритме анализа соответствий, см. [10].

### 3. Основной алгоритм

Изложенные выше соображения далее применяются для построения таких изображений объектов и их формирующих показателей в рамках выбранных кластеров, что расстояния между этими изображениями соответствуют различиям между ними, – чем ближе

изображения, тем более сильно взаимодействуют их прообразы, и наоборот. Опишем алгоритм. Пусть у нас имеется некоторое кластерное разбиение основного множества  $U$  на  $m$  кластеров. Будем считать, что и число формирующих показателей, и число объектов в каждом из кластеров не менее 3. Это позволит нам построить визуализацию на плоскости ( $q = 2$ ), что с точки зрения наглядности и информативности следует, видимо, считать оптимальным. Далее предлагается проделать следующие шаги.

1. Для выбранного кластера построим матрицу данных  $F$ , расположив значения формирующих показателей объектов кластера в строках этой матрицы. Превратим ее в матрицу с двойным центрированием, используя формулы (2). За полученной матрицей сохраним исходное обозначение.

2. Найдем два наибольших собственных числа и единичные собственные векторы, отвечающие, для каждой из матриц  $\Phi = FF^t$ ,  $\Phi_p = F^tF$ . На самом деле достаточно сделать это для той из них, которая имеет меньшую размерность. Все собственные числа другой матрицы, отличные от 0, окажутся теми же согласно предложению 3, а ее собственные векторы находятся с помощью формул (3) или (4). Найденные векторы пронумеруем в соответствии с убыванием величин собственных чисел.

3. Находя координаты проекций столбцов  $F$  на плоскость, натянутую на два найденных собственных вектора матрицы  $\Phi$  и координаты проекций строк  $F$  на плоскость, натянутую на два собственных вектора  $\Phi_p$  в базисах этих векторов соответственно, объявляем их координатами визуализаций формирующих показателей и объектов кластера в универсальном двумерном пространстве. Поскольку все собственные векторы имеют единичные длины, поиск соответствующих координат сводится к вычислению скалярных произведений, см. (5).

4. Повторяем шаги 1 – 3 для каждого из  $m$  кластеров имеющегося разбиения.

После окончания описанной процедуры у нас получаются  $m$  плоских рисунков, которые и предлагается рассматривать как многоуровневый профиль имеющейся кластеризации. Каждый рисунок соответствует одному из уровней профиля. При необходимости сравнить между собой два или более кластерных разбиения одного и того же множества на качественном, визуальном уровне, можно построить подобные профили каждого из этих разбиений и сравнивать их между собой.

Дополнительным источником информации (уже в числовом виде) служат значения координат изображений признаков и объектов в универсальных пространствах. Эти пространства не обязаны совпадать при изучении различных кластеров, но визуализация в каждом из них описывает взаимодействие признаков с объектами внутри кластера, поэтому их сравнение между собой вполне законно.

#### 4. Два примера многоуровневых профилей

Для иллюстрации результатов работы предложенного алгоритма рассмотрим данных анализа лимфоцитарного профиля периферической крови 10 пациентов по 3 показателям во время обследования их в Алтайском краевом диагностическом центре. Данные приведены в таблице 1. Формирующими показателями являются Т4 – свободный тироксин (пмоль/л); В – количество В-лимфоцитов (Млрд/л); ФНО – уровень кахектина (фактор некроза опухоли).

Сначала пациенты были разбиты на 3 кластера методом  $k$ -средних (см., например, [9, с. 67], для чего использован компьютерный пакет IBM SPSS 23. Первый кластер образовали пациенты 4, 7 и 9, второй 2, 5, 6 и 10, третий – 1, 3 и 8. После применения предлагаемого алгоритма координаты визуализаций объектов и показателей в каждом из кластеров приведены в таблице 2. Трехуровневый профиль кластерного разбиения показан на рис. 1. Признаки на этом рисунке показаны рыжеватыми точками, а объекты синими.

Таблица 1

## Данные анализа крови

Пациент	Т4	В	ФНО	Пациент	Т4	В	ФНО
1	16,57	8,13	29,52	6	14,93	7,76	31,04
2	14,97	7,78	31,44	7	16,07	7,74	30,95
3	16,72	8,66	29,55	8	15,29	8,24	32,65
4	15,52	8,36	30,75	9	16,47	8,53	34,41
5	15,22	8,66	31,34	10	15,88	7,94	32,06

Таблица 2

## Координаты визуализаций разбиения 1

Кластер 1		Кластер 2		Кластер 3				
<b>об1</b>	0,726	0,372	<b>об1</b>	0,108	0,857	<b>об1</b>	0,664	-0,472
<b>об2</b>	-0,686	0,443	<b>об2</b>	-0,700	-0,169	<b>об2</b>	0,080	0,807
<b>об3</b>	-0,041	-0,815	<b>об2</b>	-0,106	-0,334	<b>об2</b>	-0,744	-0,336
			<b>об4</b>	0,698	-0,353			
<b>пр1</b>	0,009	0,816	<b>пр1</b>	0,360	0,732	<b>пр1</b>	-0,111	0,804
<b>пр2</b>	0,702	-0,416	<b>пр2</b>	-0,815	-0,054	<b>пр2</b>	0,756	-0,306
<b>пр3</b>	-0,712	-0,400	<b>пр3</b>	0,455	-0,678	<b>пр3</b>	-0,645	-0,498



Рисунок 1. Профиль после применения метода 3-средних

Для того, чтобы иметь возможность сравнить профили, разобьем данные таблицы 1 практически наугад, по порядку пациентов в списке. Оставим численный состав кластеров без изменений – первый кластер состоит из пациентов с первого до третьего, второй с четвертого по седьмого, последние 3 пациента составят третий кластер. Визуализации этого разбиения приведены на рис. 2, а координаты соответствующих точек – в таблице 3.

Таблица 3

## Координаты визуализаций разбиения 2

Кластер 1		Кластер 2		Кластер 3				
<b>об1</b>	-0,369	0,728	<b>об1</b>	0,635	-0,152	<b>об1</b>	-0,776	0,249
<b>об2</b>	0,815	-0,044	<b>об2</b>	-0,570	0,554	<b>об2</b>	0,607	0,545
<b>об3</b>	-0,446	-0,684	<b>об2</b>	-0,400	-0,744	<b>об2</b>	0,170	-0,794
			<b>об4</b>	0,335	0,342			
<b>пр1</b>	0,592	-0,562	<b>пр1</b>	-0,148	-0,803	<b>пр1</b>	-0,377	0,721
<b>пр2</b>	0,191	0,7946	<b>пр2</b>	-0,621	0,530	<b>пр2</b>	-0,439	-0,685
<b>пр3</b>	-0,783	-0,232	<b>пр3</b>	0,769	0,273	<b>пр3</b>	0,816	-0,037



Рисунок 2. Трехуровневый профиль второго разбиения

## 5. Обсуждение результатов

Простое визуальное сравнение рисунков 1 и 2 показывает, что существенной разницы между ними нет. Тем не менее, понятно, что кластерное разбиение методом 3-средних должно давать «более правильный» результат, чем просто фактически случайное разбиение множества объектов. Отсюда вытекает, что приведенные рисунки показывают нечто иное, чем качество кластерного разбиения, какой бы смысл в это понятие не вкладывался. Это же соображение может быть подкреплено тем, что при добавлении произвольной константы ко всем элементам исходной матрицы данных  $F$  не изменяет итоговой матрицы с двойным центрированием, что ясно из формулы (2). Поэтому, если мы рассмотрим два кластера, все объекты одного из которых получены сдвигом объектов второго в пространстве показателей  $R^p$  на один и тот же вектор  $(x, \dots, x)$ , то матрицы рассеивания, а значит, и визуализации этих кластеров окажутся идентичными. При этом, если число  $x$  достаточно велико, то имеющуюся кластеризацию следует признать качественной, а если это число близко к 0, то, напротив, абсолютно неудовлетворительной.

Заметим, что, переходя к матрице данных с двойным центрированием, мы фактически сместили центр рисунка в начало координат универсального пространства. Поэтому полученный рисунок будет показывать структуру взаимодействий объектов и показателей именно внутри кластера, никак не касаясь кластерного разбиения в целом, что подтверждается рассмотренным только что примером.

Следовательно, сравнение уровней полученного с помощью предлагаемого алгоритма профиля кластерного разбиения дает лишь возможность сравнить способы взаимодействия объектов и формирующих показателей в каждом из имеющихся кластеров. Например, если рисунки двух уровней похожи, то это означает в определенном смысле однородность соответствующих кластеров, – их внутреннее строение имеет одинаковую структуру, организовано по тем же самым принципам.

При этом, наряду с чисто визуальным способом оценки структуры уровней профиля, вполне можно получить выводы в более традиционной числовой форме, если использовать классические критерии однородности данных. В качестве проверяемых на однородность выборок можно взять первые координаты визуализаций из таблиц 2 или 3. При этом разное число объектов на разных уровнях профиля не является препятствием, что облегчает такой анализ по сравнению с визуальным.

## 6. Некоторые выводы

В работе фактически был предложен способ сравнения кластеров по характеру взаимодействия между собой объектов и показателей этих кластеров. Речь, в основном, шла о сравнении кластеров в рамках одной и той же кластеризации, и предлагалось сравнивать схожесть характеров взаимодействий внутри сравниваемых кластеров. Конечно же,

описанный способ можно применить и при сравнении различных кластерных разбиений, попытавшись найти в одном из них «двойника» каждому из кластеров в другом по предложенному критерию. Из анализа должны быть исключены одноэлементные кластеры. Если такой двойник для каждого кластера успешно находится, то кластерные разбиения можно объявить схожими.

## Список литературы

1. Rui Xu, Wunsch D. II. Survey of Clustering Algorithms // IEEE Transactions on Neural Networks. — 2005. — Vol. 16, no. 3. — P. 645–678.
2. Xu D., Tian Y. A. Comprehensive Survey of Clustering Algorithms // Ann. Data. Sci. — 2015. — Vol. 2. — P. 165–193.
3. Halkidi M., Batistakis Y., Vazirgiannis M. On clustering validation techniques // Journal of intelligent information systems. — 2001. — Vol. 17(2-3). — P. 107–145.
4. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Классификация и снижение размерности. — М. : Финансы и статистика, 1989.
5. Rousseeuw P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis // Journal of Computational and Applied Mathematics. — 1987. — Vol. 20. — P. 53–65.
6. Журавлева В.В., Куракина А.А. Упрощенный показатель силуэта кластерной структуры // МАК: Математики – Алтайскому краю. сборник трудов всероссийской конференции по математике с международным участием / Главный редактор профессор Н.М. Оскорбин. — Барнаул : Изд-во Алт. ун-та, 2019. — С. 254–255.
7. Pages J. Multiple Factor Analysis by Example Using R. (CRC The R Series). — London : Chapman & Hall, 2014.
8. Dronov S.V., Leongardt K.A. Multidimensional unfolding problem solution in the case of a single target // IOP Conf. Series: Journal of Physics: Conf. Series 1210. — 2019. — 012034.
9. Дронов С.В. Методы и задачи многомерной статистики. — Барнаул : Изд-во Алт. ун-та, 2015.
10. Greenacre M. Correspondence Analysis in Practice. — 3rd edition. — Boca Raton : CRC Press, 2021.