

Impact Factor:

ISRA (India) = 6.317
ISI (Dubai, UAE) = 1.582
GIF (Australia) = 0.564
JIF = 1.500

SIS (USA) = 0.912
ПИИИ (Russia) = 3.939
ESJI (KZ) = 9.035
SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
PIF (India) = 1.940
IBI (India) = 4.260
OAJI (USA) = 0.350

SOI: [1.1/TAS](#) DOI: [10.15863/TAS](#)

International Scientific Journal Theoretical & Applied Science

p-ISSN: 2308-4944 (print) e-ISSN: 2409-0085 (online)

Year: 2022 Issue: 02 Volume: 106

Published: 28.02.2022 <http://T-Science.org>

QR – Issue



QR – Article



S. U. Zhanatauov

noncommercial joint-stock company "Kazakh national agrarian research university"

Academician of International Academy

of Theoretical and Applied Sciences (USA),

Candidate of physics and mathematical sciences,

Department «Information technologies and automatization»,

Professor, Kazakhstan

sapagtu@mail.ru

METHOD FOR REGULATING THE PROPORTION OF DOMINANT EIGENVALUES FOR A FIXED MATRIX OF EIGENVECTORS

Abstract: In the article, quantitative parameters and variables of the method for regulating the share of dominant eigenvalues are found for a fixed matrix of eigenvectors of the correlation matrix R_{nn} . Changing the variances (eigenvalues from the matrix A_{nn}) and increasing the proportion of variances $f_4(A_{nn}, 2) = (\lambda_1 + \lambda_2)/n$ did not lead to an increase in the number $\ell=2$ of dominant dispersions (according to the Kaiser-Dickman criterion), but increased the share of extracted knowledge: an increase in the values of off-diagonal elements by 20% leads to parallel shifts of the curves (y -variables) and to an increase in the share of information by 7.4%, from which hidden knowledge is extracted. The hidden knowledge extracted from the triplet of matrices ($C_{nn}, R(s)_{nn}, A(s)_{nn}$) refers to the calculated matrices of z -variability values $Z_{mn} = \{z_{ij}\}$ of correlated z -variables) and to the matrix $Y_{mn} = Z_{mn} C_{nn} = \{y_{ij}\}$ y -variability of uncorrelated y -variables.

Key words: matrix, variability, eigenvalues, eigenvectors.

Language: Russian

Citation: Zhanatauov, S. U. (2022). Method for regulating the proportion of dominant eigenvalues for a fixed matrix of eigenvectors. *ISJ Theoretical & Applied Science*, 02 (106), 601-613.

Soi: <http://s-o-i.org/1.1/TAS-02-106-64> **Doi:** <https://dx.doi.org/10.15863/TAS.2022.02.106.64>

Scopus ASCC: 2604.

СПОСОБ РЕГУЛИРОВАНИЯ ДОЛИ ДОМИНИРУЮЩИХ СОБСТВЕННЫХ ЧИСЕЛ ПРИ ФИКСИРОВАННОЙ МАТРИЦЕ СОБСТВЕННЫХ ВЕКТОРОВ

Аннотация: В статье найдены количественные параметры и переменные метода регулирования доли доминирующих собственных чисел при фиксированной матрице собственных векторов корреляционной матрицы R_{nn} . Изменение дисперсий (собственных чисел из матрицы A_{nn}) и увеличение доли дисперсий $f_4(A_{nn}, 2) = (\lambda_1 + \lambda_2)/n$ не привело к увеличению количества $\ell=2$ доминирующих дисперсий (по критерию Кайзера-Дикмана), но повысило долю извлекаемых знаний: увеличение значений недиагональных элементов на 20% приводит к параллельным сдвигам кривых (y -переменных) и к увеличению доли информации на 7.4%, из которой извлекаются скрытые знания. Извлеченные из тройки матриц ($C_{nn}, R(s)_{nn}, A(s)_{nn}$) скрытые знания относятся к вычисленным матрицам значений z -изменчивостей $Z_{mn} = \{z_{ij}\}$ коррелированных z -переменных) и к матрице $Y_{mn} = Z_{mn} C_{nn} = \{y_{ij}\}$ y -изменчивостей некоррелированных y -переменных.

Ключевые слова: матрица, изменчивость, собственные числа, собственные вектора.

Введение

Матрица (z_1, z_j) -корреляций $R_{nn} = \{r_{ij}\} = (1/m) Z_{mn}^T Z_{mn}$, $r_{ij} = \text{corr}(z_i, z_j)$, $i, j = 1, \dots, n$, является функцией, зависящей от матрицы

$Z_{mn} = \{z_{ij}\}$, $z_{ij} = (x_{ij}^0 - x_j^{me})/s_j$ изменчивостей z -переменных, вычисленных по значениям x_{ij}^0 реальных данных. У известной корреляционной матрицы $R_{nn} = \{r_{ij}\}$, $i, j = 1, \dots, n$, ее собственные

Impact Factor:

ISRA (India) = 6.317
 ISI (Dubai, UAE) = 1.582
 GIF (Australia) = 0.564
 JIF = 1.500

SIS (USA) = 0.912
 ПИИЦ (Russia) = 3.939
 ESJI (KZ) = 9.035
 SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
 PIF (India) = 1.940
 IBI (India) = 4.260
 OAJI (USA) = 0.350

числа $\lambda_j, j=1, \dots, n$, зависят от величин всех коэффициентов корреляции. Необходим поиск постоянных и мало меняющихся матриц, таких, что они не меняют состав и мозаику индикаторов присутствия скрытых, но извлекаемых знаний. Такой матрицей является матрица C_{nn} собственных векторов анализируемой ниже симметрической матрицы $R^{(s)}_{nn}$ и ее спектр $\Lambda^{(s)}_{nn}$, который является (f_1, f_4) -спектром [1], адекватным “реальному” (f_1, f_4) -спектру Λ_{nn} “реальной” симметрической матрицы R_{nn} . Они не меняются, как показано ниже, при специальном изменении (применении регулируемого параметра s) значений элементов матрицы R_{nn} . Использование матрицы собственных векторов C_{nn} (матрицы (z, y) -корреляций) и новой матрицы (z, z) -корреляций $R^{(s)}_{nn}$, имеющей матрицу собственных векторов, равной прежней матрице C_{nn} , позволило нам изменить дисперсии (матрицу собственных чисел $\Lambda^{(s)}_{nn}$) и увеличить долю дисперсий прежнего количества ℓ доминирующих дисперсий $f_4(\Lambda_{nn, 2}) = (\lambda_1 + \lambda_2)/n$. Мы рассматриваем ситуацию моделирования и извлечения когнитивных знаний [4-8] из многомерных данных типа «объект-свойства», в случае когда матрицы (z, y) -корреляций предпочтительней классической матрицы (z, z) -корреляций R_{nn} . Но матрица (z, z) -корреляций R_{nn} является исходным объектом, характеризующим исходные реальные многомерные данные X^0_{mn} (относящихся к разным предметным областям) без которой невозможно провести тот или иной глубокий (интеллектуальный [2-8]) анализ данных. Ниже мы найдем количественные параметры и переменные метода регулирования доли доминирующих собственных чисел при фиксированной матрице собственных векторов. Глубинный анализ будет проводиться над матрицами C_{nn} и $\Lambda^{(s)}_{nn}$. Извлеченные из них знания относятся к вычисленным матрицам значений z -

изменчивостей $Z_{mn} = \{z_{ij}\}$ коррелированных z -переменных) и к матрице $Y_{mn} = Z_{mn}C_{nn} = \{y_{ij}\}$ y -изменчивостей некоррелированных y -переменных.

Исходные данные

Имеем значения удельных масс n газообразных веществ, вычисленные по реальным данным службы мониторинга загрязнений воздуха. Загрязняющие атмосферу газообразные вещества (единица измерения – $mg/(l \cdot m^3)$) рассчитывались в 2005 году в начале каждого часа в течение 24 часов (одного дня) по n наименованиям: формальдегид, бензопирен, диоксид азота, фенол, оксид углерода, аммиак. Матрица (z, y) -корреляций C_{nn} (Таблица 2) соответствует как паре матриц (R_{nn}, Λ_{nn}) таких, что: $R_{nn}C_{nn} = C_{nn}\Lambda_{nn}$, вычислена при решении ПСЗ: $R_{nn} = \langle C_{nn}\Lambda_{nn} \rangle$, где $R_{nn} = R^T_{nn}$ - исходная для нашего метода матрица¹ (z, z) -корреляций (Таблица 1) $R_{nn} = \{r_{ij} = \text{corr}(z_i, z_j)\}$, $i=1, \dots, n$; $j=1, \dots, n$, диагональная матрица $\Lambda_{nn} = \text{diag}(2.8198, 1.3987, 0.8343, 0.465, 0.2856, 0.1965)$. Многомерными данными для матрицы (z, z) -корреляций (Таблица 1) $R_{nn} = \{r_{ij} = \text{corr}(z_i, z_j)\}$ служит матрица $Z_{mn} = \{z_{ij}, z_{ij} = (x^0_{ij} - x^{me}_{ij})/s_j\}$ значений z -изменчивостей, вычисленных по $m=24$ значениям удельных масс x^0_{ij} $i=1, \dots, m$; $j=1, \dots, n$, n загрязняющих газообразных веществ¹. Диагональные элементы матрицы $\Lambda_{nn} = \text{diag}(2.8198, 1.3987, 0.8343, 0.465, 0.2856, 0.1965)$ является дисперсиями n y -изменчивостей: $\Lambda_{nn} = (1/m)Y^T_{mn}Y_{mn}$, $Y_{mn} = Z_{mn}C_{nn}$ – матрица значений y -изменчивостей. Вычислительная схема $R_{nn} = \langle C_{nn}\Lambda_{nn} \rangle$ предшествует схеме вычисления многомерной выборки $Y_{mn} = Z_{mn}C_{nn}$.

Таблица 1. Матрица $R_{nn} = \{r_{ij} = \text{corr}(z_i, z_j)\}$ (z, z) -корреляций

	«z1»	«z2»	«z3»	«z4»	«z5»	«z6»
	1,0000	-0,6748	-0,6767	-0,3916	0,2168	-0,1852
	-0,6748	1,0000	0,5379	0,1711	0,0538	0,3818
	-0,6767	0,5379	1,0000	0,4740	-0,4042	0,4064
	-0,3916	0,1711	0,4740	1,0000	-0,4861	0,2131
	0,2168	0,0538	-0,4042	-0,4861	1,0000	0,1418
	-0,1852	0,3818	0,4064	0,2131	0,1418	1,0000

Impact Factor:

ISRA (India) = 6.317	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 1.582	ПИИЦ (Russia) = 3.939	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 9.035	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 7.184	OAJI (USA) = 0.350

Таблица 2. Матрица $C_{nn}=C^{(s)}_{nn}=\{c^{(s)}_{ij}=\text{corr}(z^{(s)}_i, y^{(s)}_j)\}$ ($z^{(s)}, y^{(s)}$)-корреляций

№	c1	c2	c3	c4	c5	c6
1	-0.4969	-0.0782	-0.4504	0.1747	0.3496	-0.6256
2	0.4294	0.4172	0.3283	-0.0565	0.6779	-0.2665
3	0.5296	-0.0368	-0.0215	0.4075	-0.4884	-0.5596
4	0.3839	-0.3977	-0.3623	-0.7292	0.0492	-0.1705
5	0.2445	0.6891	0.0065	-0.4748	-0.4146	-0.2609
6	0.2848	0.4307	-0.7467	0.2076	0.0738	0.3567

Таблица 3. Матрица $R^{(s)}_{nn}=\{r^{(s)}_{ij}=\text{corr}(z^{(s)}_i, z^{(s)}_j)\}$ (z, z)-корреляций

	1	2	3	4	5	6
1	1.0000	-0.8097	-0.8121	-0.4699	0.2602	-0.2223
2	-0.8097	1.0000	0.6454	0.2053	0.0645	0.4581
3	-0.8121	0.6454	1.0000	0.5688	-0.4850	0.4876
4	-0.4699	0.2053	0.5688	1.0000	-0.5833	0.2557
5	0.2602	0.0645	-0.4850	-0.5833	1.0000	0.1701
6	-0.2223	0.4581	0.4876	0.2557	0.1701	1.0000

Таблица 4

	1	2	3	4	5
1	0.4605	-0.4758	-0.2035	0.1420	0.7071
2	0.4679	0.0132	0.5397	-0.6997	-0.0000
3	0.4030	0.6026	-0.6520	-0.2221	-0.0000
4	0.4411	0.4289	0.4480	0.6486	-0.0000
5	0.4605	-0.4758	-0.2035	0.1420	-0.7071

Применяемые вычислительные модели

Метод регулирования доли доминирующих собственных чисел при фиксированной матрице собственных векторов состоит из реализации нескольких вычислительных схем. Для преобразованной матрицы (z, z)-корреляций $R^{(s)}_{nn}=\{r^{(s)}_{ij}=\gamma_{ij} * s\}$, $i \neq j$, $r^{(s)}_{ii}=\{1\}=R^{(s)T}_{nn}$ решается классическая однородная спектральная задача - Прямая Спектральная Задача (ПСЗ): $R^{(s)}_{nn} \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn})$, где выполняется равенство $R^{(s)}_{nn} C_{nn} = C_{nn} \Lambda^{(s)}_{nn}$, $s \geq 2$. Для новой пары матриц ($C_{nn}, \Lambda^{(s)}_{nn}$) решается Обратная Задача (ОЗ, вариант №2) из ОМ АГК [9-10] со схемой вычислений (моделирования) вида: $(C_{nn}, \Lambda^{(s)}_{nn}) \Rightarrow (Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn})$, $t=1, \dots, k_t$. Здесь повторно вычисляется матрица C_{nn} при решении ПСЗ 2 вида: $R^{(s)}_{nn} \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn})$, где $R^{(s)}_{nn}=\{r^{(s)}_{ij}=\gamma_{ij} * s\}$, $i \neq j$, $r^{(s)}_{ii}=\{1\}$. Кроме ПСЗ 2 [9]: $R^{(s)}_{nn} \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn})$ существует ОСЗ 1 [11] со схемой вычислений $\Lambda_{nn} \Rightarrow (C^{(l)}_{nn}, R^{(l)}_{nn})$, номер

$l=1, 2, \dots, k_l$. ОСЗ 1 $\Lambda_{nn} \Rightarrow (C^{(l)}_{nn}, R^{(l)}_{nn})$ используется в ОМ АГК [9,10]: $\Lambda_{nn} \Rightarrow (C^{(l)}_{nn}, R^{(l)}_{nn}, Y^{(t)}_{mn}, Z^{(t,l)}_{mn})$, $t=1, \dots, k_t$, $l=1, 2, \dots, k_l$. В ОМ АГК реализуются равенства $R^{(l)}_{nn} C^{(l)}_{nn} = C^{(l)}_{nn} \Lambda_{nn}$, $R_{nn} = (1/m) Z^{(t,l)T}_{mn} Z^{(t,l)}_{mn}$, $Y^{(t)}_{mn} = U^{(t)} \Lambda^{1/2}_{nn}$, $Z^{(t,l)}_{mn} = Y^{(t)}_{mn} C^{(l)}_{nn}$. В нашем же методе вычислительная схема $(R_{nn}, s) \Rightarrow (R^{(s)}_{nn}, \Lambda^{(s)}_{nn})$ предшествует схеме моделирования многомерных выборок $Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn}$: $(R_{nn}, s) \Rightarrow (R^{(s)}_{nn}, \Lambda^{(s)}_{nn}) \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn}) \Rightarrow (Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn})$, $t=1, \dots, k_t$.

Применяемые вычислительные модели [1-8] соответствуют 3 типам решаемых задач:

а) вычисление матрицы $R_{nn} = (1/m) Z^T_{mn} Z_{mn}$ и новой матрицы (z_i, z_j)-корреляций $R^{(10)}_{nn} = \{r^{(1)} * s\}$;

б) вычисление матрицы собственных векторов $C_{nn} = C^{(s)}_{nn}$ с индикаторами наличия знаний [12,131] и матрицы собственных чисел таких, что $R_{nn} C_{nn} = (C_{nn} \Lambda_{nn})$, $R^{(s)}_{nn} C^{(s)}_{nn} = (C^{(s)}_{nn} \Lambda^{(s)}_{nn})$, $\Lambda_{nn} = \text{diag}(2.8198, 1.3987, 0.8343, 0.465, 0.2856,$

Impact Factor:

ISRA (India) = 6.317
ISI (Dubai, UAE) = 1.582
GIF (Australia) = 0.564
JIF = 1.500

SIS (USA) = 0.912
ПИИЦ (Russia) = 3.939
ESJI (KZ) = 9.035
SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
PIF (India) = 1.940
IBI (India) = 4.260
OAJI (USA) = 0.350

0.1965), $\Lambda^{(s)}_{nn} = \text{diag}(\lambda^{(s)}_1, \dots, \lambda^{(s)}_n) = \text{diag}(3.1839, 1.4787, 0.8011, 0.3580, 0.1426, 0.0358)$;

в) вычисление матрицы Y_{mn} , исходя из матриц Z_{mn}, C_{nn} ;

г) вычисление матриц $Y^{(s)}_{mn}, Z^{(s)}_{mn}$, исходя из матрицы собственных чисел $\Lambda^{(s)}_{nn}$: $\Lambda^{(s)}_{nn} \rightarrow Y^{(s)}_{mn}, Z^{(s)}_{mn}$.

Постоянные веса z-переменных и случайные изменчивости весов

«Вес» $c_{ik} = \text{corr}(z_i, y_k)$, умноженный на величину изменчивости z_{kj} образует слагаемое вида $(+c_{ik} * z_{kj})$, входящее в формулы смысловых u-переменных. Матрица (z_1, z_j) -корреляций $R_{nn} = \{r_{ij}\} = (1/m) Z_{mn}^T Z_{mn}$, $r_{ij} = \text{corr}(z_i, z_j)$, $i, j = 1, \dots, n$, является функцией ей, зависящей от матрицы $Z_{mn} = \{z_{ij}\}$ изменчивостей z-переменных. Нам надо управлять изменчивостями z-переменных. Мы добиваемся регулирования изменчивостями z-переменных через управление элементами матрицы (z_1, z_j) -корреляций R_{nn} , равной функции вида $(1/m) Z_{mn}^T Z_{mn}$, зависящей от матрицы изменчивостей n z-переменных Z_{mn} . Матрица Z_{mn} (изменчивости z-переменных) влияет на матрицу $Y_{mn} = Z_{mn} C_{nn}$ через постоянные «веса» - матрицу C_{nn} . Неизменяемость элементов матрицы C_{nn} при изменении элементов матрицы R_{nn} будет показана ниже. «Вес» $c_{ik} = \text{corr}(z_i, y_k)$, умножается на разную случайную величину изменчивости z_{kj} , образуется слагаемое вида $(+c_{ik} * z_{kj})$. Изменения в матрице R_{nn} приведут к изменениям в матрице Z_{mn} . В ПМ ГК для вычисленной матрицы Z_{mn} вычисляется матрица R_{nn} , а в ОМ ГК - наоборот: для заданной матрицы R_{nn} моделируется бесконечное множество матриц $Z^{(t)}_{mn} = \{z^{(t)}_{ij}\}$, $t = 1, \dots, \infty$, изменчивостей z-переменных. Если в ОМ ГК изменим элементы матрицы R_{nn} , то при моделировании (решении ОЗ АГК) получатся матрицы $Z^{(t)}_{mn} = \{z^{(t)}_{ij}\}$, $t = 1, \dots, \infty$, с измененными свойствами. Применим один из простых способов управления изменением матрицей R_{nn} . Пусть в новой матрице (z_1, z_j) -корреляций $R^{(s)}_{nn} = \{r^{(s)}_{ij}\}$, где $r^{(s)}_{ij} = s * r_{ij}$, $s > 1$, $r_{\max} = \text{abs}(s * r_{ij}) < 1$. Для соблюдения этих условий достаточно найти максимальную величину (z_i, z_j) -корреляций $\max\{\text{abs}(r_{ij})\}$ и вычислить величину $1 < s \leq 1 / r_{\max}$. Увеличение всех коэффициентов (z_1, z_j) -корреляций в $s > 1$ раз позволит поднять общий уровень парных связей в множестве z-переменных. Увеличится значение функции $f_4(R^{(s)}_{nn}, \ell) = f_4(\Lambda^{(s)}_{nn}, \ell) = (\lambda_1 + \dots + \lambda_\ell) / n$ и значение количества ℓ доминирующих элементов спектра. Увеличение собственных векторов нам необходимо. Этим мы увеличиваем количество собственных векторов, имеющих заметные значения компонент. Эти собственные векторы располагались в $(\ell+1)$ -ом, в $(\ell+2)$ -ом, ..., в n-ом столбце матрицы C_{nn} . Заметные значения компонент собственных векторов являются индикаторами присутствия [12,13] извлекаемых

знаний из матрицы Z_{mn} и из множества когнитивных смыслов имен z-переменных, количество которых будут превышать ℓ . При $\ell = n$ работа аналитика существенно усложняется, ибо потребуются осмысление и конструирование одной фразы, равной сумме из n смыслов. Привлечение новой матрицы (z_1, z_j) -корреляций $R^{(s)}_{nn} = \{r^{(s)}_{ij}\}$, имеющей ту же матрицу C_{nn} индикаторов, что и матрица (z_1, z_j) -корреляций R_{nn} не увеличивает количество смысловых u-переменных. Чем больше имеем фраз-имен u-переменных, тем больше имеем извлеченных знаний. Особенно ценятся имена-смыслы валидных переменных, повышающих точность и адекватность моделирования изучаемого явления или процесса.

Другим примером целесообразности увеличения количества смысловых u-переменных является матрица собственных векторов C_{55} [2]. Здесь выделена 1 смысловая u-переменная, ибо по критерию Кайзера-Дикмана $\ell = 1$ из спектра $\Lambda_{55} = \text{diag}(3.9125, 0.6944, 0.2476, 0.1456, 0.0000)$ выделена 1 дисперсия $\lambda_1 = 3.9125$, соответствующая ей смысловая u-переменная имеет формулу $y_1 = 0.4479 * z_1 + 0.3961 * z_2 + 0.4444 * z_4 + 0.4129 * z_5 + 0.4259 * z_6$, $\text{corr}(z_1, y_1) = 0.4479$, $\text{corr}(z_2, y_1) = 0.3961$, $\text{corr}(z_4, y_1) = 0.4444$, $\text{corr}(z_5, y_1) = 0.4129$, $\text{corr}(z_6, y_1) = 0.4259$. Здесь руководствовались одной доминирующей дисперсией $\lambda_1 = 3.9125$, пренебрегали индикаторами присутствия знания в других столбцах матрицы C_{55} (Таблица 4).

Необходимо более глубоко анализировать C_{55} и присвоить имена-смыслы другим смысловым u-переменным с меньшими дисперсиями. Для этого надо анализировать доминирующие элементы спектра $\Lambda^{(s)}_{nn}$ новой матрицы $R^{(s)}_{nn} = \{r^{(s)}_{ij}\}$, собственные векторы которой остались неизменными и содержат те же индикаторы присутствия извлекаемых знаний из той же матрицы Z_{mn} и из того же множества когнитивных смыслов имен z-переменных.

Перечень имен-смыслов 5 коррелированных z-переменных весьма богатый, он позволит сконструировать имена-смыслы нескольким u-переменным. Смысл(z_1)=(Т4)=«Доля прибыльных предприятий»; Смысл(z_2)=«Объем промышленного производства на 1 предприятие»; Смысл(z_3)=«Дебиторская задолженность и задолженность по обязательствам на 1 предприятие»; Смысл(z_4)=«Количество междугородных разговоров на 1 предприятие»; Смысл(z_5)=«Международный трафик на СНГ (минуты) для предприятий».

Impact Factor:

ISRA (India) = 6.317
 ISI (Dubai, UAE) = 1.582
 GIF (Australia) = 0.564
 JIF = 1.500

SIS (USA) = 0.912
 ПИИЦ (Russia) = 3.939
 ESJI (KZ) = 9.035
 SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
 PIF (India) = 1.940
 IBI (India) = 4.260
 OAJI (USA) = 0.350

**Параметры, определяющие значения
 недиагональных элементов корреляционной
 матрицы**

Матрица C_{nn} соответствует как паре матриц (R_{nn}, Λ_{nn}) таких, что: $R_{nn}C_{nn} = C_{nn}\Lambda_{nn}$, так и паре матриц $(R^{(s)}_{nn}, \Lambda^{(s)}_{nn})$ таких, что $R^{(s)}_{nn}C_{nn} = C_{nn}\Lambda^{(s)}_{nn}$. Параметр s – вещественное число, определяющее значения элементов как матрицы $R^{(s)}_{nn}$, так и ее спектра $\Lambda^{(s)}_{nn}$. Число s не влияет на матрицу C_{nn} собственных векторов симметрических матриц $R_{nn}, R^{(s)}_{nn}$. Вычислительная схема $(R_{nn}, s) \Rightarrow (R^{(s)}_{nn}, \Lambda^{(s)}_{nn})$ предшествует схеме моделирования многомерных выборок $Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn}$:

$(R_{nn}, s) \Rightarrow (R^{(s)}_{nn}, \Lambda^{(s)}_{nn}) \Rightarrow (C_{nn}) \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn}) \Rightarrow (Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn}), t=1, \dots, k_t$. Так как в ПСЗ 2 выполнено равенство $\Lambda^{(s)}_{nn} = C_{nn}^T R^{(s)}_{nn} C_{nn}$, то вычисляется в ПСЗ2 по этой формуле (по заданной матрице $R^{(s)}_{nn}$) спектр $\Lambda^{(s)}_{nn}$. Вычисление спектра $\Lambda^{(s)}_{nn}$ по формуле предпочтительнее сложных итерационных вычислений (программа КИМ) спектра Λ_{nn} (по заданной матрице R_{nn}).

Из решения Прямой спектральной задачи: $R^{(s)}_{nn}C_{nn} = C_{nn}\Lambda^{(s)}_{nn}$ мы находим матрицу C_{nn} – главный объект, содержащий индикаторы присутствия извлекаемых знаний. Смысл этих знаний выбираем из смыслов-имен z -переменных с заметными «весами» $c_{ij} \in C_{nn} = \{c_{ij}\}$. Отдельные знания содержатся в смыслах-именах u -переменных, они образуют когнитивную сумму смыслов-имен z -переменных, имеющих заметные «веса». Учитываются 1 или 2 или 3 собственных вектора (1 или 2 или 3 столбца матрицы C_{nn}), имеющие 1 или 2 или 3 доминирующих чисел из матрицы $\Lambda^{(s)}_{nn}$. Две матрицы $R_{nn}, R^{(s)}_{nn}$ имеют отличающиеся в s раз коэффициенты корреляции $R^{(12)}_{nn} = \{r^{(s)}_{ij} = r_{ij}/s\}$, $i \neq j$, $r_{ii} = 1$, $r^{(s)}_{ii} = 1$, но одинаковые собственные векторы, одинаковые индикаторы присутствия извлекаемых знаний. Такая ситуация с корреляционными матрицами характеризует 2 разные ситуации одного специфического явления. Одним из таких явлений является процесс абсорбции загрязняющих газообразных веществ¹.

Метод регулирования доли доминирующих собственных чисел при фиксированной матрице собственных векторов состоит из реализации нескольких вычислительных схем.

Решается ПСЗ: $R^{(s)}_{nn} \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn})$, где выполняется равенство $R^{(s)}_{nn}C_{nn} = C_{nn}\Lambda^{(s)}_{nn}$, $s \geq 2$. Для новой пары матриц $(C_{nn}, \Lambda^{(s)}_{nn})$ решается ОЗ (вариант №2) из ОМ АГК со схемой вычислений (моделирования) вида:

$(C_{nn}, \Lambda^{(s)}_{nn}) \Rightarrow (Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn}), t=1, \dots, k_t$. Здесь повторно вычисляется матрица C_{nn} при решении ПСЗ вида: $R^{(s)}_{nn} \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn})$, где $R^{(s)}_{nn} = \{r^{(s)}_{ij} = r_{ij} * s\}$, $i \neq j$, $r^{(s)}_{ii} = 1$. Кроме ПСЗ 2: $R^{(s)}_{nn} \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn})$ существует ОСЗ со схемой вычислений $\Lambda_{nn} \Rightarrow (C^{(\ell)}_{nn}, R^{(\ell)}_{nn})$, номер $\ell = 1, 2, \dots, k_\ell$. ОСЗ 1 $\Lambda_{nn} \Rightarrow (C^{(\ell)}_{nn}, R^{(\ell)}_{nn})$ используется в ОМ АГК:

$\Lambda_{nn} \Rightarrow (C^{(\ell)}_{nn}, R^{(\ell)}_{nn}, Y^{(t)}_{mn}, Z^{(\ell,t)}_{mn}), t=1, \dots, k_t, \ell=1, 2, \dots, k_\ell$. В ОМ АГК реализуются равенства $R^{(\ell)}_{nn}C^{(\ell)}_{nn} = C^{(\ell)}_{nn}\Lambda_{nn}$, $R_{nn} = (1/m)Z^{(\ell,t)}_{mn}Z^{(\ell,t)T}_{mn}$, $Y^{(t)}_{mn} = U^{(t)}\Lambda^{1/2}_{nn}$, $Z^{(\ell,t)}_{mn} = Y^{(t)}_{mn}C^{(\ell)}_{nn}$.

В нашем же методе вычислительная схема $(R_{nn}, s) \Rightarrow (R^{(s)}_{nn}, \Lambda^{(s)}_{nn})$ предшествует схеме моделирования многомерных выборок $Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn}$: $(R_{nn}, s) \Rightarrow (R^{(s)}_{nn}, \Lambda^{(s)}_{nn})$, $(R^{(s)}_{nn}) \Rightarrow (C_{nn}, \Lambda^{(s)}_{nn}) \Rightarrow (Y^{(s,t)}_{mn}, Z^{(s,t)}_{mn}), t=1, \dots, k_t$. Так как в ПСЗ 2 выполнено равенство $\Lambda^{(s)}_{nn} = C_{nn}^T R^{(s)}_{nn} C_{nn}$, то вычисляется в ПСЗ 2 по этой формуле (по заданной матрице $R^{(s)}_{nn}$) спектр $\Lambda^{(s)}_{nn}$. Вычисление спектра $\Lambda^{(s)}_{nn}$ по формуле предпочтительнее сложных итерационных вычислений (программа КИМ) спектра Λ_{nn} по заданной матрице R_{nn} .

Вычисленные по известной корреляционной матрице R_{nn} ее собственные числа $\lambda_j, j=1, \dots, n$, зависят от величин всех коэффициентов корреляции (ПСЗ). В ОМ ГК входным объектом является спектр $\Lambda_{nn} = \text{diag}(\lambda_1, \dots, \lambda_n)$, $n > 2$, вычисленный при решении ПСЗ $R_{nn} \Rightarrow (C_{nn}, \Lambda_{nn})$. в случае, если нет реальных данных и нет корреляционной матрицы R_{nn} , но имеется спектр $\Lambda_{nn} = \text{diag}(\lambda_1, \dots, \lambda_n)$, то можно смоделировать модельные матрицу Y_{mn} u -изменчивостей (с дисперсиями $\text{diag}(\lambda_1, \dots, \lambda_n) = \Lambda_{nn} = (1/m)Y_{mn}^T Y_{mn}$) и матрицу $Z_{mn} = Y_{mn}C_{nn}^T$ значений z -изменчивостей с одинаковыми (=1) дисперсиями. Если не имеется спектр, но известны число $n=6$ и значение f_4 функции $f_4(\Lambda_{nn}) = (\lambda_1 + \dots + \lambda_n)/n < 1$ (доли доминирующих собственных чисел), то можно смоделировать модельный спектр Λ_{nn} , (f_1, f_4) -адекватный “реальному” спектру: $(f_1, f_4) \Rightarrow \Lambda_{nn}$, такому что его f -параметры f_1, f_4 равны заданным значениям, а остальные f -параметры имеют произвольные значения: $f_1(\Lambda_{nn}) = \lambda_1 + \dots + \lambda_n = n$, $f_2(\Lambda_{nn}) = (\lambda_1^2 + \dots + \lambda_n^2)/n$, $f_3(\Lambda_{nn}) = \lambda_1/\lambda_n$, $f_4(\Lambda_{nn}, \ell) = (\lambda_1 + \dots + \lambda_\ell)/n < 1$, $f_5(\Lambda_{nn}) = \lambda_1 \times \lambda_2 \times \lambda_3 \times \dots \times \lambda_n$, $f_n(\Lambda_{nn}) = (\lambda_1/\lambda_2 + \dots + \lambda_{n-1}/\lambda_n)$ [1,9]. Так как $\text{tr}(R^T R) = \text{tr}(\Lambda^2)$, то соотношения между f -параметрами (f_1, f_4) функционально влияют на значения элементов r_{ij} корреляционной матрицы $R_{nn} = \{r_{ij} = \text{corr}(z_i, z_j)\}$ [6, стр.37-46]. В нашем методе мы используем влияние значений всех недиагональных элементов r_{ij} корреляционной матрицы R_{nn} на регулируемое значение f -параметра $f_4(\Lambda_{nn}, \ell) = (\lambda_1 + \dots + \lambda_\ell)/n$. Для повышения доли доминирующих собственных чисел, приводящему к увеличению значения f -параметра $f_4(\Lambda_{nn}, \ell)$, Ниже применяем мультипликативный параметр s – вещественное число $s > 1$, при этом для значения f -параметра $f_4(\Lambda_{nn}, \ell)$ нет формулы зависимости от корреляционной матрицы R_{nn} . нами установлена зависимость значения f -параметра $f_4(\Lambda_{nn}, \ell)$ от Λ_{nn} (при неизменяемом количестве ℓ доминирующих собственных чисел) от элементов матрицы R_{nn} при фиксированной матрице собственных векторов $C^{(s)}_{nn} = C_{nn}$.

Impact Factor:

ISRA (India) = 6.317
 ISI (Dubai, UAE) = 1.582
 GIF (Australia) = 0.564
 JIF = 1.500

SIS (USA) = 0.912
 ПИИЦ (Russia) = 3.939
 ESJI (KZ) = 9.035
 SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
 PIF (India) = 1.940
 IBI (India) = 4.260
 OAJI (USA) = 0.350

Поясним причину такого регулирования. Значение $\lambda_1=2.8198$ вычислено для заданной симметрической матрицы R_{nn} максимальным из-за решаемой при этом оптимизационной задачи, оно может быть увеличено за счет увеличения значений элементов матрицы $R_{nn}=R_{nn}^T$. В новой матрице $R^{(s)}_{nn}=\{r^{(s)}_{ij}=r_{ij}^s, i \neq j, r^{(s)}_{ii}=1, s=1.2\}$ (Таблица 2) все начения недиагональных элементов увеличены на 20%, при этом значение элемента λ_1 увеличено с $\lambda_1=\lambda_j=(1/24)*(y^2_{1j}+\dots+y^2_{mj})=2.8198$ до величины $\lambda^{(s)}_1=(1/24)*(y^{(s)2}_{1j}+\dots+y^{(s)2}_{mj})=3.1839$. Ее вычисленная ортонормированная матрица собственных векторов $C^{(s)}_{nn}$ осталась равной прежней: $C^{(s)}_{nn}=C_{nn}$ (Таблица 2), а матрица собственных чисел равна $\Lambda^{(s)}_{nn}=(3.1839, 1.4787, 0.8011, 0.3580, 0.1426, 0.0358)$ имеет 1-ый (максимальный) элемент, увеличенной величины: $\lambda^{(s)}_1=3.1839 > 2.8198 = \lambda_1$. Матрица $R_{nn}=R_{nn}^T$ имеет матрицу собственных чисел равную $\Lambda_{nn}=\text{diag}(2.8198, 1.3987, 0.8343, 0.465, 0.2856, 0.1965)$. Для новой матрицы $R^{(s)}_{nn}=R^{(s)T}_{nn}$ ее матрица собственных чисел равна $\Lambda^{(s)}_{nn}=(3.1839, 1.4787, 0.8011, 0.3580, 0.1426, 0.0358)$, для нее также выполнен критерий Кайзера-Дикмана: $\lambda^{(s)}_2=1.4787 > 1, \ell=2$. При этом доля 2-х доминирующих дисперсий $\lambda^{(s)}_1=3.1839, \lambda^{(s)}_2=1.4787$ увеличилась до значения $f_4(\Lambda_{nn}, 2) = (\lambda_1+\lambda_2)/n = (3.1839+1.4787)/6 = 0,777 (=77.7\%)$. Увеличение произошло с 70.3% до 77.7%. А увеличение на 30% ($s=1.3$) значений всех внедиагональных элементов матрицы R_{nn} дает увеличение доли 2-х дисперсий $f_4(\Lambda_{nn}, 2)$ (при $s=1.3$) до 81.4% (прирост равен $81.4\% - 0,70,3\% = 11.1\%$), но появляются отрицательные недоминирующие собственные числа, что недопустимо. Прирост $7.4\% = 77.7\% - 70.3\%$ – существенный дополнительный объем знаний, извлекаемый из многомерных данных. При этом общий уровень коррелированности переменных возрос с $\varphi=0.36471$ до $\varphi=0,4371$, т е его значение не перешло в другой интервал шкалы Чеддока - умеренная — от 0,3 до 0,5. Такое соотношение уровня коррелированности с 7.4% -ым приростом объема знаний соответствует хорошей устойчивости извлеченных знаний к

изменчивостям в данных, из которых извлекаются знания с применением нашей тройки матриц $(C_{nn}, R^{(s)}_{nn}, \Lambda^{(s)}_{nn})$. Прирост доли $f_4(\Lambda_{nn}) = (\lambda_1 + \dots + \lambda_\ell) / n$ не привело к увеличению числа $\ell=2$ по критерию Кайзера-Дикмана. Динамики 2-х пар кривых у-переменных $(y^{(s)}_1, y_1), (y^{(s)}_2, y_2)$ с дисперсиями $\lambda^{(s)}_1 = (1/m) * (y^{(s)2}_{11} + \dots + y^{(s)2}_{m1}) = 3.1839,$
 $\lambda_1 = (1/m) * (y^2_{11} + \dots + y^2_{m1}) = 2.8198,$
 $\lambda^{(s)}_2 = (1/m) * (y^{(s)2}_{12} + \dots + y^{(s)2}_{m2}) = 1.4787,$
 $\lambda_2 = (1/m) * (y^2_{12} + \dots + y^2_{m2}) = 1.1839$ представляют собой 2 пары параллельных кривых (Рисунок 1, Рисунок 2). Значения 4-х дисперсий для 4-х у-или $y^{(s)}$ -переменных в их формулах не меняются от любой перестановки слагаемых, поэтому для наглядности динамик элементы 4-х столбцов из Таблиц 6,7 упорядочены по возрастанию. Каждая кривая из пары кривых получается приблизительно параллельным переносом другой кривой. Параллельность кривой у-переменной к кривой $y^{(s)}$ -переменной означает наличие приблизительно постоянной величины отличия значения у-переменной от значения $y^{(s)}$ -переменной. Увеличение значений недиагональных элементов на 20% приводит к сдвигам кривых у-переменных и к увеличению доли информации (на 7.4%), из которой извлекается скрытые знания.

В других столбцах матрицы $C_{nn}=C^{(s)}_{nn}$ с номерами $j > 2$ остались индикаторы присутствия скрытых и извлекаемых знаний. Нужны другие критерии и другие методы для их извлечения при таких индикаторах. Смысл ограничений на величины доминирующих дисперсий $\lambda_1 > 1, \dots, \lambda_\ell > 1$ в критерии Кайзера-Дикмана состоит в отборе только таких у-переменных u_1, \dots, u_ℓ , дисперсии которых превышают по величине дисперсию любой z-переменной (равной 1). Изменчивости с большими дисперсиями содержат знания, чем изменчивости с меньшими дисперсиями. Пока мы применяем только такие индикаторы присутствия извлекаемых знаний (матрица C_{nn}) и доминирующие дисперсии $\lambda_1, \dots, \lambda_\ell$ являются объектами когнитивного (интеллектуального) анализа многомерных данных.

Таблица 5. Матрица Y_{mn} у-изменчивостей

1	-1.8506	3.0476	0.6410	-0.1496	-0.3494	-0.9996
2	0.3897	-1.3875	-0.2529	1.6490	-0.2404	-0.6589
3	0.8519	-1.8874	1.5437	1.3308	-0.1785	-0.0524
4	0.0836	0.6494	-0.5616	-0.0275	-0.5335	0.5015
5	-0.5096	-0.0988	-0.9408	0.0309	-0.3278	0.6647
6	-0.3109	-0.7781	-0.9166	0.6355	-0.3727	0.2482

Impact Factor: ISRA (India) = 6.317 SIS (USA) = 0.912 ICV (Poland) = 6.630
 ISI (Dubai, UAE) = 1.582 ПИИЦ (Russia) = 3.939 PIF (India) = 1.940
 GIF (Australia) = 0.564 ESJI (KZ) = 9.035 IBI (India) = 4.260
 JIF = 1.500 SJIF (Morocco) = 7.184 OAJI (USA) = 0.350

7	-0.1641	0.8611	0.3037	-0.5784	0.4131	0.1527
8	-1.7232	1.2843	-0.7375	1.2822	0.6160	0.3104
9	2.0674	-0.8517	0.3213	-0.7024	0.4431	0.3617
10	3.8821	1.1686	-0.3706	-0.0716	0.2843	0.1093
11	-0.0588	0.4585	1.1039	-0.3609	-0.6072	0.2225
12	1.0905	0.9767	1.9074	-0.0381	0.8157	0.3041
13	-0.5102	0.7829	-0.7048	-0.1429	0.1256	0.6272
14	0.5594	-2.2191	0.3052	-1.1115	0.1484	-0.1987
15	2.2975	0.0693	-1.2764	0.2291	-0.0169	0.2418
16	-1.2813	-1.0757	-1.4426	-0.4574	0.7682	-0.1850
17	-1.7128	0.6749	0.6936	0.8855	-0.1191	0.2799
18	3.8061	0.3950	0.7140	-0.0901	-0.7301	-0.4288
19	-0.4689	-0.1908	-0.9152	-0.4919	-0.9510	-0.3535
20	-1.0259	0.1121	-0.6701	-0.7993	-0.8485	-0.2000
21	-0.4093	0.0425	-0.4680	-0.2825	0.9108	-0.8104
22	-2.5486	-0.0708	1.3485	-0.4502	-0.0651	0.4754
23	0.1175	0.2054	-0.4065	0.0995	0.7850	-0.5282
24	-2.5716	-2.1684	0.7812	-0.3881	0.0301	-0.0838
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2.8198	1.3987	0.8343	0.4650	0.2856	0.1965

Таблица 6. Матрица Z_{mn} z-изменчивостей

1	0.8696	0.7253	-0.4371	-1.8925	3.0333	-0.1067
2	0.6451	-0.5752	1.4210	-0.3088	-1.5643	-0.2081
3	-0.7681	-0.0971	1.1463	-0.4518	-2.0430	-1.4785
4	-0.3445	-0.3713	0.0011	-0.1145	0.5267	0.8567
5	0.1596	-0.9700	-0.4453	0.0325	-0.0019	0.7341
6	0.4536	-1.1138	0.1858	-0.0019	-0.6781	0.4538
7	-0.1748	0.6606	-0.6481	-0.0994	0.6990	0.0622
8	1.3331	-0.1839	-0.8961	-1.8627	0.3563	1.0354
9	-1.2994	0.8816	0.4143	1.4884	-1.0347	-0.0020
10	-1.8350	2.2005	1.7916	1.2074	-0.2587	1.9308
11	-0.9184	0.0779	-0.0468	-0.4095	0.7026	-0.6839
12	-1.3890	1.9761	-0.0836	-0.6447	0.0195	-0.5324
13	0.1363	-0.1978	-0.7545	-0.2485	0.5117	0.9215
14	-0.2598	-0.3691	-0.0427	1.8384	-1.1458	-1.3150
15	-0.6894	0.5075	1.2079	1.1077	-0.6872	1.7699
16	1.6749	-0.8767	-1.0659	0.8614	-0.4904	0.1447
17	0.4238	-0.4315	-0.6845	-1.8765	0.4442	-0.4403
18	-2.2464	1.6579	2.5456	1.1483	-0.1963	0.4955
19	0.4628	-1.1041	0.2402	0.5996	0.6972	0.1693
20	0.4916	-1.0905	-0.3324	0.3795	1.1071	-0.0434

Impact Factor: ISRA (India) = 6.317 SIS (USA) = 0.912 ICV (Poland) = 6.630
 ISI (Dubai, UAE) = 1.582 ПИИЦ (Russia) = 3.939 PIF (India) = 1.940
 GIF (Australia) = 0.564 ESJI (KZ) = 9.035 IBI (India) = 4.260
 JIF = 1.500 SJIF (Morocco) = 7.184 OAJI (USA) = 0.350

21	1.1869	0.5377	-0.3146	0.3846	0.0943	-0.0294
22	0.2657	-0.8265	-1.7938	-1.1948	0.6998	-1.6921
23	0.7309	0.6700	0.0162	0.1668	-0.1247	0.3156
24	1.0907	-1.6877	-1.4247	-0.1090	-0.6666	-2.3579
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Таблица 7. Матрица $Y^{(s)}_{mn}$ у-изменчивостей

№						
1	-1.9664	3.1333	0.6281	-0.1313	-0.247	-0.427
2	0.414	-1.4265	-0.2478	1.4469	-0.17	-0.2814
3	0.9052	-1.9405	1.5128	1.1677	-0.126	-0.0224
4	0.0888	0.6677	-0.5503	-0.0241	-0.377	0.2142
5	-0.5415	-0.1016	-0.9219	0.0271	-0.232	0.2839
6	-0.3304	-0.8	-0.8983	0.5576	-0.264	0.106
7	-0.1744	0.8853	0.2976	-0.5075	0.292	0.0652
8	-1.831	1.3204	-0.7227	1.125	0.4355	0.1326
9	2.1968	-0.8756	0.3149	-0.6163	0.3133	0.1545
10	4.1251	1.2014	-0.3631	-0.0629	0.201	0.0467
11	-0.0625	0.4714	1.0817	-0.3167	-0.429	0.095
12	1.1588	1.0042	1.8691	-0.0334	0.5766	0.1299
13	-0.5422	0.805	-0.6907	-0.1254	0.0888	0.2679
14	0.5944	-2.2815	0.2991	-0.9753	0.1049	-0.0849
15	2.4413	0.0712	-1.2508	0.201	-0.012	0.1033
16	-1.3615	-1.106	-1.4136	-0.4013	0.5431	-0.079
17	-1.82	0.6939	0.6797	0.777	-0.084	0.1195
18	4.0443	0.4061	0.6996	-0.0791	-0.516	-0.1832
19	-0.4982	-0.1962	-0.8968	-0.4316	-0.672	-0.151
20	-1.0901	0.1152	-0.6567	-0.7014	-0.6	-0.0854
21	-0.4349	0.0437	-0.4586	-0.2479	0.6439	-0.3462
22	-2.7081	-0.0728	1.3215	-0.395	-0.046	0.203
23	0.1249	0.2112	-0.3983	0.0873	0.555	-0.2256
24	-2.7325	-2.2293	0.7655	-0.3405	0.0213	-0.0358
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3.1838	1.4785	0.8011	0.3580	0.1428	0.0359

Таблица 8. Матрица $Z^{(s)}_{mn}$ z-изменчивостей

№						
1	0.6070	0.6229	-0.8643	-2.0723	2.9200	0.1225
2	0.3869	-0.6207	1.1073	-0.1993	-1.6289	-0.1238

Impact Factor:

ISRA (India) = 6.317 **SIS (USA) = 0.912** **ICV (Poland) = 6.630**
ISI (Dubai, UAE) = 1.582 **ПИИИ (Russia) = 3.939** **PIF (India) = 1.940**
GIF (Australia) = 0.564 **ESJI (KZ) = 9.035** **IBI (India) = 4.260**
JIF = 1.500 **SJIF (Morocco) = 7.184** **OAJI (USA) = 0.350**

3	-0.8054	-0.0698	1.0683	-0.2826	-2.0448	-1.4823
4	-0.1186	-0.1754	0.0888	-0.0696	0.5467	0.7674
5	0.4383	-0.8119	-0.2978	0.0870	0.0655	0.5802
6	0.5703	-1.0089	0.1704	0.0791	-0.6595	0.3663
7	-0.1440	0.6015	-0.5173	-0.1535	0.7575	0.0488
8	1.3979	-0.2764	-0.8312	-1.7878	0.6035	0.8998
9	-1.2596	0.8874	0.6983	1.5160	-1.0159	-0.0363
10	-1.9500	2.2806	1.9982	1.2851	-0.2487	1.9819
11	-0.7579	0.2266	-0.0463	-0.4098	0.6507	-0.6860
12	-1.3816	1.8884	0.1685	-0.6011	0.1638	-0.5513
13	0.3590	-0.1278	-0.5463	-0.2280	0.6355	0.7841
14	-0.3322	-0.4496	-0.0087	1.7580	-1.2737	-1.2617
15	-0.6890	0.6203	1.3471	1.1972	-0.6734	1.7377
16	1.5689	-1.0982	-1.0344	0.7622	-0.4524	0.1201
17	0.5754	-0.4017	-0.7132	-1.8120	0.5623	-0.5294
18	-2.4361	1.8391	2.4341	1.2011	-0.4051	0.6846
19	0.4508	-0.9813	-0.0004	0.5190	0.5038	0.2502
20	0.5496	-0.9798	-0.5125	0.2700	0.9456	0.0091
21	0.8176	0.2236	-0.4438	0.2533	0.0745	0.1100
22	0.5440	-0.8223	-1.7120	-1.2383	0.7742	-1.8025
23	0.4513	0.4423	-0.0422	0.1104	-0.1003	0.4026
24	1.1577	-1.8088	-1.5105	-0.1842	-0.7009	-2.3920
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

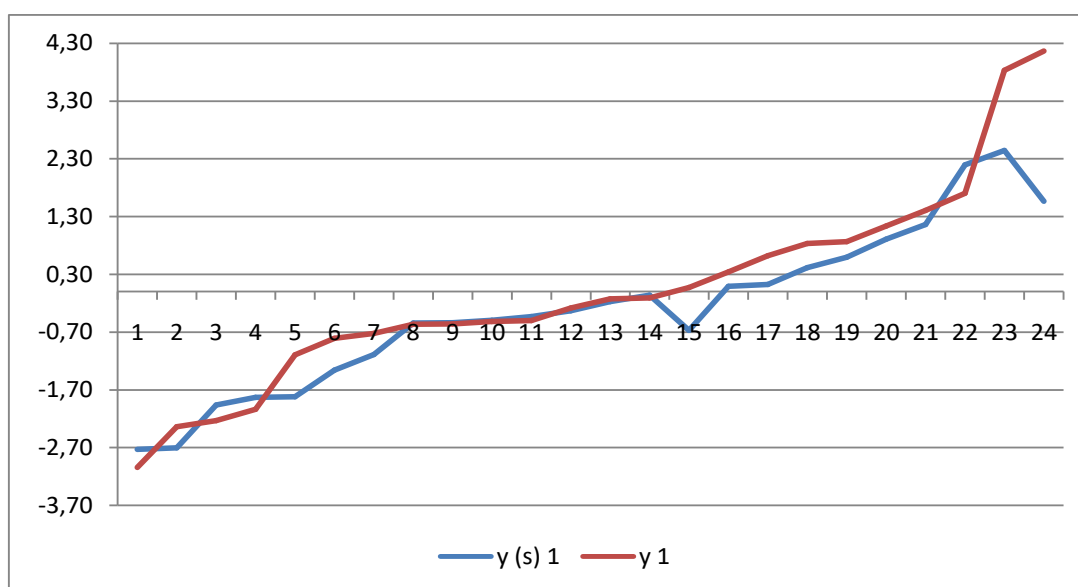


Рисунок 1. Динамики 2-х у-переменных $y^{(s)}_1, y_1$

Impact Factor:

ISRA (India) = 6.317	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 1.582	ПИИЦ (Russia) = 3.939	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 9.035	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 7.184	OAJI (USA) = 0.350

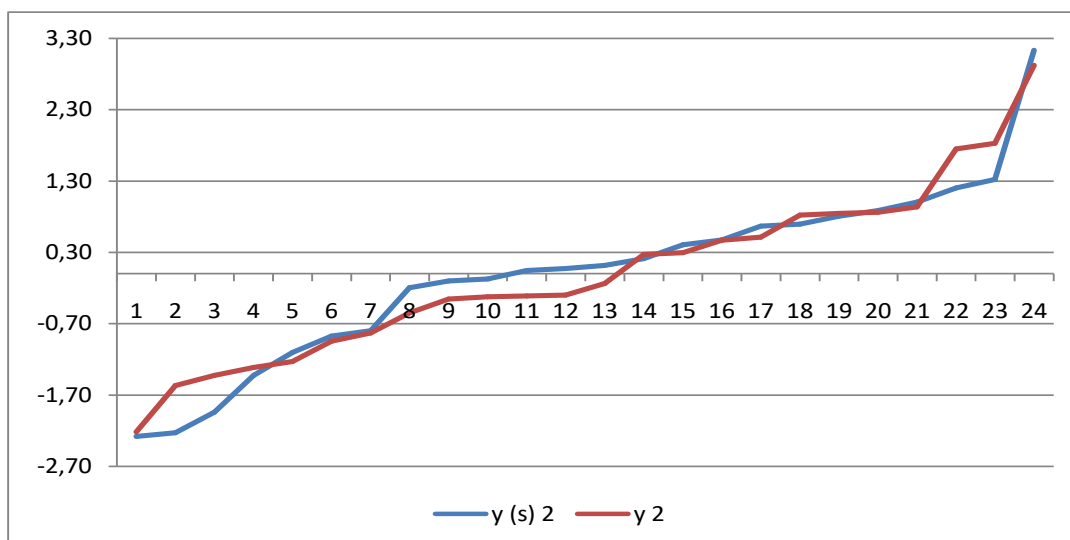


Рисунок 2. Динамики 2-х у-переменных $y^{(s)2}$, y_2

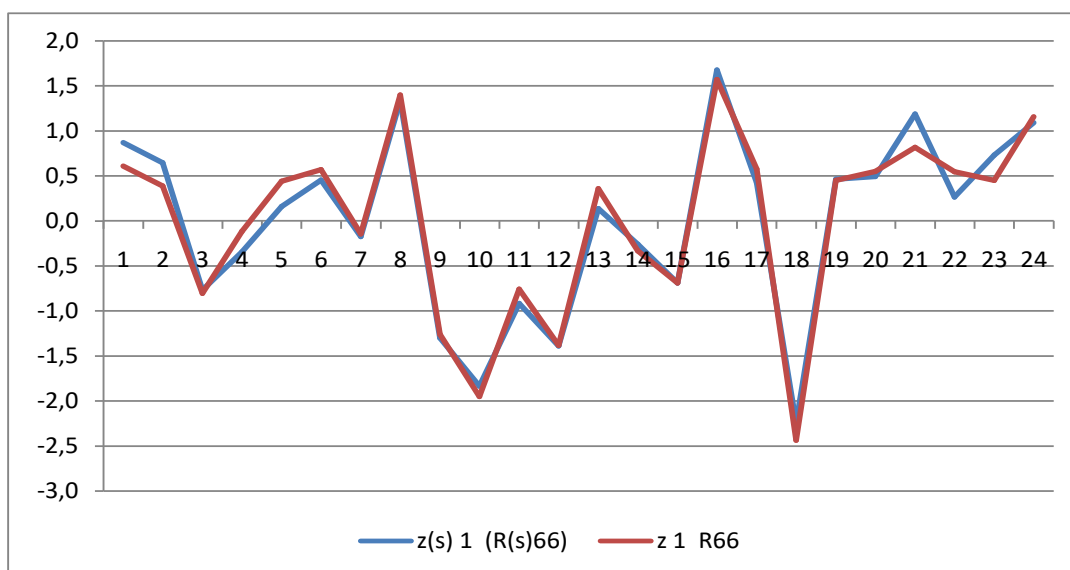


Рисунок 3. Динамики одной z-переменной z_1 , $z^{(s)1}$ из 2-х матриц Z_{m6} , $Z^{(s)mn}$

Использование матрицы собственных векторов C_{nn} (матрицы (z,y)-корреляций) и новой матрицы (z,z)-корреляций $R^{(s)nn}$, имеющей матрицу собственных векторов, равной прежней матрице C_{nn} , позволило нам изменить дисперсии (матрицу собственных чисел $\Lambda^{(s)nn}$) и увеличить долю дисперсий прежнего количества ℓ доминирующих дисперсий $f_4(\Lambda_{nn}, 2) = (\lambda_1 + \lambda_2)/n$. Мы привели очередную ситуацию моделирования и извлечения когнитивных знаний [4-8] из многомерных данных типа «объект-свойства», когда матрицы (z,y)-корреляций предпочтительней классической матрицы (z,z)-корреляций R_{nn} . Но матрица (z, z)-корреляций R_{nn} является исходным объектом, характеризующим исходные реальные многомерные данные X_{mn}^0 (относящихся к разным предметным областям) без которой невозможно провести тот или иной

глубинный (интеллектуальный [4-8]) анализ данных. Глубинный анализ проводится с вычисленными матрицами значений z-изменчивостей (с матрицей $Z_{mn} = \{z_{ij}\}$ изменчивостей коррелированных z-переменных) и y-изменчивостей (с матрицей $Y_{mn} = Z_{mn} C_{nn} = \{y_{ij}\}$ изменчивостей некоррелированных y-переменных). В данной статье фиксированным объектом является матрица (z,y)-корреляций) C_{nn} – матрица преобразования z-изменчивостей в y-изменчивости. Регулятор – число $s > 0$ в электронной таблице Excel (с применением или без применения надстройки «Поиск решения» (Solver)) автоматически изменяет недиагональные элементы матрицы R_{nn} : $R^{(s)nn} = \{r^{(s)ij} = r_{ij} * s\}$, $i \neq j$, $r_{ii} = 1$, $r^{(s)ij} = 1$, вычисляются ее собственные числа $\lambda^{(s)j}$, $j = 1, \dots, n$, образуя диагональную матрицу $\Lambda^{(s)nn} = C_{nn}^T R^{(s)nn} C_{nn} = \text{diag}(\lambda^{(s)1}, \dots, \lambda^{(s)n})$, $n > 2$. При этом

Impact Factor:

ISRA (India) = 6.317
ISI (Dubai, UAE) = 1.582
GIF (Australia) = 0.564
JIF = 1.500

SIS (USA) = 0.912
ПИИЦ (Russia) = 3.939
ESJI (KZ) = 9.035
SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
PIF (India) = 1.940
IBI (India) = 4.260
OAJI (USA) = 0.350

полученный модельный спектр $\Lambda^{(s)}_{nn}$ является (f_1, f_4) -адекватным “реальному” (f_1, f_4) -спектру. Наш (f_1, f_4) -спектр $\Lambda^{(s)}_{nn}$, такой что его f -параметры f_1, f_4 равны заданным значениям, а остальные f -параметры произвольны, где: $f_1(\Lambda_{nn}) = \lambda_1 + \dots + \lambda_n = n$, $f_2(\Lambda_{nn}) = (\lambda_1^2 + \dots + \lambda_n^2)/n$, $f_3(\Lambda_{nn}) = \lambda_1/\lambda_n$, $f_4(\Lambda_{nn}) = (\lambda_1 + \dots + \lambda_\ell)/n < 1$, $f_5(\Lambda_{nn}) = \lambda_1 \times \lambda_2 \times \lambda_3 \times \dots \times \lambda_n$, $f_n(\Lambda_{nn}) = (\lambda_1/\lambda_2 + \dots + \lambda_{n-1}/\lambda_n)$. Так как $\text{tr}(\mathbf{R}^T \mathbf{R}) = \text{tr}(\Lambda^2)$, то соотношения между f -параметрами функционально зависят от элементов корреляционной матрицы \mathbf{R}_{nn} [6, стр.37-46].

Матрица \mathbf{C}_{nn} соответствует как паре матриц $(\mathbf{R}_{nn}, \Lambda_{nn})$ таких, что: $\mathbf{R}_{nn} \mathbf{C}_{nn} = \mathbf{C}_{nn} \Lambda_{nn}$, так и паре матриц $(\mathbf{R}^{(s)}_{nn}, \Lambda^{(s)}_{nn})$ таких, что $\mathbf{R}^{(s)}_{nn} \mathbf{C}_{nn} = \mathbf{C}_{nn} \Lambda^{(s)}_{nn}$. Параметр s – вещественное число, определяющее значения элементов как матрицы $\mathbf{R}^{(s)}_{nn}$, так и ее спектра $\Lambda^{(s)}_{nn}$. Число s не влияет на матрицу \mathbf{C}_{nn} собственных векторов симметрических матриц $\mathbf{R}_{nn}, \mathbf{R}^{(s)}_{nn}$. Две разные симметрические матрицы существенно разными элементами не имеют разные матрицы собственных векторов, но имеют разные спектры, соответствующие разным объемам скрытых, но извлекаемых знаний из одной и той же матрицы реальных многомерных данных, для которой анализу подвергается матрица значений z -изменчивости \mathbf{Z}_{mn} . Возможность извлечения немного больших знаний из той же матрицы реальных многомерных данных является заметным преимуществом перехода к матрицам $\mathbf{R}^{(s)}_{nn}$ и $\Lambda^{(s)}_{nn}$.

Матрица \mathbf{C}_{nn} содержит индикаторы присутствия извлекаемых знаний из матрицы \mathbf{Z}_{mn} и знаний из множества когнитивных смыслов имен z -переменных, количество которых будет превышать ℓ . При $\ell = n$ работа аналитика существенно усложняется, ибо потребуются осмысление и конструирование одной фразы, равной сумме из n смыслов. Привлечение новой матрицы (z_1, z_j) -корреляций $\mathbf{R}^{(s)}_{nn} = \{r^{(s)}_{ij}\}$, имеющей ту же матрицу \mathbf{C}_{nn} индикаторов, что и матрица (z_1, z_j) -корреляций \mathbf{R}_{nn} не увеличивает количество смысловых u -переменных.

Изменения значений чисел не учитывалось до сих пор. Наш глубинный анализ позволил определить 2 параметра изменения величины x^{0}_{ij} . И обнаружить за этими изменениями чисел реальные химические реакции абсорбции, иные поглощения или иную реакцию соединения.

Заключение

Разработанный способ регулирования доли доминирующих собственных чисел при фиксированной матрице собственных векторов прибавил общий уровень коррелированности переменных в матрицах \mathbf{C}_{nn} , $\mathbf{R}^{(s)}_{nn}$, $\Lambda^{(s)}_{nn}$, а не в матрице исходных реальных данных. Наш способ позволил увеличить уровень коррелированности с $\varphi = 0.36471$ до $\varphi = 0.4371$, т.е. значение не уровня не перешло в другой интервал шкалы Чеддока -

умеренная (от 0,3 до 0,5). Такое соотношение уровня коррелированности с 7.4% -ым приростом объема знаний соответствует хорошей устойчивости извлеченных знаний к изменчивостям в данных, из которых извлекаются знания с применением нашей тройки матриц $(\mathbf{C}_{nn}, \mathbf{R}^{(s)}_{nn}, \Lambda^{(s)}_{nn})$ и мышления формулами и фразами-смыслами переменных. Прирост доли $f_4(\Lambda_{nn}) = (\lambda_1 + \dots + \lambda_\ell)/n$ не привело к увеличению числа $\ell = 2$ по критерию Кайзера-Дикмана. Изменились значения дисперсий u -переменных. Но измененные значения u -переменных имеют «параллельные» динамики: каждая кривая из пары кривых получается приблизительно параллельным переносом другой кривой. Параллельность кривой u -переменной к кривой $u^{(s)}$ -переменной означает наличие приближенно постоянной величины отличия значения u -переменной от значения $u^{(s)}$ -переменной. Увеличение значений недиагональных элементов на 20% приводит к сдвигам кривых u -переменных и к увеличению доли информации (на 7.4%), из которой извлекается больше скрытых знаний.

Использование матрицы собственных векторов \mathbf{C}_{nn} (матрицы (z, y) -корреляций) и новой матрицы (z, z) -корреляций $\mathbf{R}^{(s)}_{nn}$, имеющей матрицу собственных векторов, равной прежней матрице \mathbf{C}_{nn} , позволило нам изменить дисперсии (матрицу собственных чисел $\Lambda^{(s)}_{nn}$) и увеличить долю дисперсий прежнего количества ℓ доминирующих дисперсий $f_4(\Lambda_{nn}, 2) = (\lambda_1 + \lambda_2)/n$. Мы привели очередную ситуацию моделирования и извлечения когнитивных знаний [4-8] из многомерных данных типа «объект-свойства», когда матрицы (z, y) -корреляций предпочтительней классической матрицы (z, z) -корреляций \mathbf{R}_{nn} . Но матрица (z, z) -корреляций \mathbf{R}_{nn} является исходным объектом, характеризующим исходные реальные многомерные данные \mathbf{X}^0_{nn} (относящихся к разным предметным областям) без которой невозможно провести тот или иной гдубинный (интеллектуальный [4-8]) анализ данных. Глубинный анализ проводится с вычисленными матрицами значений z -изменчивостей (с матрицей $\mathbf{Z}_{mn} = \{z_{ij}\}$ изменчивостей коррелированных z -переменных) и u -изменчивостей (с матрицей $\mathbf{Y}_{mn} = \mathbf{Z}_{mn} \mathbf{C}_{nn} = \{y_{ij}\}$ изменчивостей некоррелированных u -переменных). В данной статье фиксированным объектом является матрица (z, y) -корреляций \mathbf{C}_{nn} – матрица преобразования z -изменчивостей в u -изменчивости. Регулятор – число $s > 0$ в электронной таблице Excel (с применением или без применения надстройки «Поиск решения» (Solver)) автоматически изменяет недиагональные элементы матрицы \mathbf{R}_{nn} : $\mathbf{R}^{(s)}_{nn} = \{r^{(s)}_{ij} = r_{ij} * s\}$, $i \neq j$, $r_{ii} = 1$, $r^{(s)}_{ii} = 1$, вычисляются ее собственные числа $\lambda^{(s)}_j$, $j = 1, \dots, n$, образуя диагональную матрицу $\Lambda^{(s)}_{nn} = \mathbf{C}^T_{nn} \mathbf{R}^{(s)}_{nn} \mathbf{C}_{nn} = \text{diag}(\lambda^{(s)}_1, \dots, \lambda^{(s)}_n)$, $n > 2$. При

Impact Factor:

ISRA (India) = 6.317
ISI (Dubai, UAE) = 1.582
GIF (Australia) = 0.564
JIF = 1.500

SIS (USA) = 0.912
ПИИЦ (Russia) = 3.939
ESJI (KZ) = 9.035
SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
PIF (India) = 1.940
IBI (India) = 4.260
OAJI (USA) = 0.350

этом полученный модельный спектр $\Lambda_{nn}^{(s)}$ является (f_1, f_4) -адекватным “реальному” (f_1, f_4) -спектру. Наш (f_1, f_4) -спектр $\Lambda_{nn}^{(s)}$, такой что его f -параметры f_1, f_4 равны заданным значениям, а остальные f -параметры произвольны, где:
 $f_1(\Lambda_{nn}) = \lambda_1 + \dots + \lambda_n = n$, $f_2(\Lambda_{nn}) = (\lambda_1^2 + \dots + \lambda_n^2)$,
 $f_3(\Lambda_{nn}) = \lambda_1 / \lambda_n$, $f_4(\Lambda_{nn}) = (\lambda_1 + \dots + \lambda_n) / n < 1$,
 $f_5(\Lambda_{nn}) = \lambda_1 \times \lambda_2 \times \lambda_3 \times \dots \times \lambda_n$, $f_n(\Lambda_{nn}) = \lambda_1 / \lambda_2 + \dots + \lambda_{n-1} / \lambda_n$.
Так как $\text{tr}(R^T R) = \text{tr}(\Lambda^2)$, то соотношения между f -параметрами функционально зависят от элементов корреляционной матрицы R_{nn} [9, стр.37-46].

Прирост 7.4%=77.7%-70.3% – существенный дополнительный объем знаний, извлекаемый из многомерных данных. При этом общий уровень коррелированности переменных возрос с $\varphi=0.36471$ до $\varphi=0.4371$, т.е. его значение не перешло в другой интервал шкалы Чеддока : умеренная - от 0,3 до 0,5. Такое соотношение уровня коррелированности с 7.4% -ым приростом объема знаний соответствует хорошей устойчивости извлеченных знаний к изменчивостям в данных, из которых извлекаются знания с применением нашей тройки матриц $(C_{nn}, R_{nn}^{(s)}, \Lambda_{nn}^{(s)})$. Прирост доли $f_4(\Lambda_{nn}) = (\lambda_1 + \dots + \lambda_n) / n$ не привело к увеличению числа $\ell=2$ по критерию Кайзера-Дикмана. Динамики 2-х пар кривых u -переменных $(y^{(s)_1}, y_1)$, $(y^{(s)_2}, y_2)$ с дисперсиями $\lambda^{(s)_1} = (1/m) * (y^{(s)_2}_{11} + \dots + y^{(s)_2}_{m1}) = 3.1839$, $\lambda_1 = (1/m) * (y^2_{11} + \dots + y^2_{m1}) = 2.8198$, $\lambda^{(s)_2} = (1/m) * (y^{(s)_2}_{12} + \dots + y^{(s)_2}_{m2}) = 1.4787$, $\lambda_2 = (1/m) * (y^2_{12} + \dots + y^2_{m2}) = 1.1839$ представляют собой 2 пары параллельных кривых (Рисунок 1, Рисунок 2). Значения 4-х дисперсий для 4-х u -или $y^{(s)}$ -переменных в их формулах не меняются от любой перестановки слагаемых, поэтому для наглядности динамик элементы 4-х столбцов из Таблиц 6,7 упорядочены по возрастанию. Каждая кривая из пары кривых получается приблизительно параллельным переносом другой кривой. Параллельность кривой u -переменной к

кривой $y^{(s)}$ -переменной означает наличие приблизительно постоянной величины отличия значения u -переменной от значения $y^{(s)}$ -переменной. Увеличение значений недиагональных элементов на 20% приводит к сдвигам кривых u -переменных и к увеличению доли информации (на 7.4%), из которой извлекается скрытые знания. В других столбцах матрицы $C_{nn} = C_{nn}^{(s)}$ с номерами $j > 2$ остались индикаторы присутствия извлекаемых знаний.

Нужны другие критерии и другие методы извлечения знаний для таких индикаторов. Смысл ограничений в критерии Кайзера-Дикмана состоит в отборе только таких u -переменных u_1, \dots, u_ℓ , дисперсии которых превышают по величине дисперсию любой z -переменной (равной 1). Изменчивости с большими дисперсиями содержат знания, чем изменчивости с меньшими дисперсиями. Пока мы применяем только такие индикаторы присутствия извлекаемых знаний (матрица C_{nn}) и доминирующие дисперсии u_1, \dots, u_ℓ являются объектами когнитивного (интеллектуального) анализа многомерных данных [14].

Модель извлечения цифровых знаний из числовых реальных данных – применение мышления формулами и мышления фразами, выражающих смысл терминов предметной области. На этапах абстрагирования и формализации количественных параметров и переменных применяется мышление именами-смыслами постоянных и переменных величин. В нашем методе часть постоянных и переменных образует функцию, требующую достижения конкретного значения при ограничивающих соотношениях на допустимые значения неизвестных переменных (или при комбинациях переменных) при фиксированной матрице собственных векторов. Следующий этап – извлечение из данных значимого знания за счет применения разработанной математической модели и системы многомерных уравнений когнитивных имен-смыслов переменных.

References:

1. Zhanatauov, S.U. (2017). The optimization problem with linearized equations f -parameters $(f_1, f_2, f_3, f_4, f_5, f_6)$ -spectrum. *ISJ Theoretical & Applied Science*, №11, vol.55, pp.251-267. www.t-science.org
2. Zhanatauov, S.U. (n.d.). *Kognitivnaja karta i kognitivnaja model' analiza glavnyh komponent*

(*telekommu nika cionnaja otrasl'*). *Nacional'naja asociacija uchenyh (NAU)*. IX Mezhd.nauch.-prakt. konf.: «Otechestvennaja nauka v jepohu izmenenij: postulaty proshlogo i teorii novogo vremeni». Rossija, g.Ekaterinburg, 16-17 maja 2015 g. (pp.55-58). Retrieved from <http://national-science.ru/>

Impact Factor:

ISRA (India) = 6.317
ISI (Dubai, UAE) = 1.582
GIF (Australia) = 0.564
JIF = 1.500

SIS (USA) = 0.912
ПИИИ (Russia) = 3.939
ESJI (KZ) = 9.035
SJIF (Morocco) = 7.184

ICV (Poland) = 6.630
PIF (India) = 1.940
IBI (India) = 4.260
OAJI (USA) = 0.350

3. Zhanatauov, S.U. (2013). Kognitivnaja karta i model` social`no-jekonomicheskikh faktorov kar`ernoj uspeshnosti shkol`nikov municipal`nyh shkol SShA. *Sibirskij pedagogicheskij zhurnal*, №6, pp. 28-33.
4. Zhanatauov, S.U. (n.d.). *Analiz budushhih debitorskoj i kreditorskoj zadolzhen nostej municipalitetov gorodov*. Jekonomicheskij analiz:teorija i praktika.
5. Zhanatauov, S.U. (2020). Cognitive simulation of price changes and money costs of the population of the Republic of Kazakhstan. *ISJ «Theoretical & Applied Science»*, № 1, vol.81, pp.135-143. www.t-science.org
6. Zhanatauov, S.U. (2020). Cognitive modeling of dependence of quantities of its in apartments from changes in income and expenditures of population Republic of Kazakhstan. *ISJ «Theoretical & Applied Science»*, №1, vol.81, pp.543-555. www.t-science.org
7. Zhanatauov, S.U. (2020). Cognitive modeling of dependence of number of individual telephones at enterprises on changes in structures of income and expenditure of enterprises. *ISJ «Theoretical & Applied Science»*, № 2, vol.82, pp.213-221. www.t-science.org
8. Zhanatauov, S.U. (2020). Formula of the key indicator “power of a profitable enterprise”. *ISJ «Theoretical & Applied Science»*, №2, vol.82, pp.222-236. www.t-science.org
9. Zhanatauov, S.U. (2013). *Obratnaja model` glavnyh komponent*. (p.201). Almaty: Kazstatinform.
10. Zhanatauov, S.U. (2017). Theorem on the Λ -samples. *International scientific journal «Theoretical & Applied Science»*, № 9, vol. 53, pp. 177-192. www.T-Science.org .
11. Chalmers, C.P. (n.d.). Generation of correlation matrices with a given eigen-structure. –*J. Stat. Comp. Simul.*, 195, vol.4, pp.133-139.
12. Zhanatauov, S.U. (2019). A matrix of values the coefficients of combinational proportionality. *ISJ Theoretical & Applied Science*, №3(68), 401-419. www.t-science.org
13. Zhanatauov, S.U. (2018). Modeling eigenvectors with given the values of their indicated components. *ISJ Theoretical & Applied Science*, №11(67), pp.107-119. www.t-science.org
14. Zhanatauov, S.U. (2021). Cognitive computing: models, calculations, applications, results. *ISJ «Theoretical & Applied Science»*, №5, vol.97, pp.594-610. www.t-science.org