УДК 004.62:517.977.18

ТЕХНОЛОГИИ И МЕТОДЫ BIG DATA ОБРАБОТКИ ЭКСПЕРТНОЙ ИНФОРМАЦИИ – АЛГЕБРАИЧЕСКИЙ МЕТОД И МЕТОД ШКАЛИРОВАНИЯ

В.С. Гладкая,
Ассистент кафедры инженерной психологии и эргономики БГУИР, магистр технических наук

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: v.gladkaya@bsuir.by

Аннотация. В данной работе были поставлены следующие цели: изучить и применить в работе научную литературу по теме; сопоставить результаты измерения величины с точками числовой прямой по алгебраическому методу шкалирования, выполнить операции над параметрами, измеренными в разных шкалах по методу одномерного шкалирования. Все поставленные цели были рассмотрены и успешно достигнуты.

Ключевые слова: Экспертная оценка, экспертный анализ, шкалирование, алгебраический метод, одномерное шкалирование, ранжировка, медиана Кемени-Снелла.

Сопоставление результатов измерения величины с точками числовой прямой. Алгебраический метод. Суть алгебраических методов обработки экспертной информации состоит во введении некоторого расстояния между оценками. Задача состоит в сопоставлении системе нестрогой ранжировки. Для ее решения используется экспертиза Э7:

Ω – множество всех нестрогих ранжировок *n* объектов (вид Ω будет указан ниже);

$\Omega_Э = \Omega$

L – эксперты изолированы;

Q – обратная связь отсутствует.

Отображение φ определяется следующим образом. Результирующая оценка A_0 находится из формулы

$$\mathrm{A}_0 \in \operatorname{Arg}\min_{\mathrm{A}\in\Omega}\sum_{\mathrm{i}=1} \mathrm{d}(\mathrm{A},\mathrm{A}^{\mathrm{j}})$$

где: *d* – расстояние между ранжировками.

Дадим описание нестрогих ранжировок и определим расстояние *d* между ними. Ранжировки (т. е. элементы Ω) будем задавать матрицами $A = \left(a_{ij}\right)$, в которых $a_{ij} = 1$ тогда и только тогда, когда i-й объект предшествует j-му; если объекты i и j равноценны, то $a_{ij} = 0$; кроме того, $a_{ii} = 0 \left(i = \overline{1,n}\right)$; $a_{ij} = 1 \Rightarrow a_{ji} = -1$.

Будем говорить, что ранжировка *С* находится *между ранжировками А* и *В*, если

$a_{ij} \le c_{ij} \le b_{ij}$ (или $a_{ij} \ge c_{ij} \ge b_{ij}$) для всех $i, j = \overline{1,n}$.

Расстояние *d* между ранжировками вводится аксиоматически:

1) *d(A, B)* > 0, причем *d(A, B)* = 0 тогда и только тогда, когда *A = B*.

2) d(A, B) = d(B, A).

3) $d(A,B)+d(B,C)\ge d(A,C)$, причем равенство достигается тогда и только тогда, когда *B* находится между *A и C*.

4) Расстояние *d* инвариантно относительно обозначений. Эта аксиома означает, что при одинаковых перестановках объектов внутри ранжировок А и *B* расстояние между новыми ранжировками *A'* и *B'* равно *d(A, B)*.

5) Если две ранжировки отличаются друг от друга только на части объектов, то расстояние между исходными ранжировками равно расстоянию между ранжировками только этих объектов.

6) Минимальное положительное расстояние между ранжировками равно единице.

Утверждение 1. Аксиомы 1) – 6). однозначно определяют расстояние d(A, B) при любой длине ранжировок $n \ge 2$, а формула

$$d(A,B) = 0.5\sum_{i=1}\sum_{j=1}\left|a_{ij} - b_{ij}\right| \tag{1}$$

определяет единственное расстояние d, удовлетворяющее аксиомам 1)-6).

Ранжировка A_0 называется *медианой Кемени-Снелла* ранжировок A^1,.., A^n. Наряду с медианой Кемени-Снелла в качестве группового мнения используют ранжировку

$$A_0^{'} = \operatorname{Arg}\min_{A\in\Omega}\sum_{i=1} d^2(A, A^j) \tag{2}$$

которая называется *средним значением.*

<u>Пример 1.</u> Пусть три эксперта упорядочили три объекта и полученные ранжировки имеют вид A = <x, *y,* z>, B = <x, *y,* z>, C = <y, *x*, z>. Выпишем соответствующие матрицы, обозначив их теми же буквами, что и ранжировки:

$$A = B = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}.$$

Пользуясь выражением (1), вычислим расстояния; *d(A,* B)=0, *так как* $a_{ij} = b_{ij}$ ($i, j = \overline{1,3}$), d(A, C) = d(B, *C) = 2.*

Если взять в качестве медианы ранжировку *A* или *B,* то сумма расстояний от нее до всех ранжировок равна 2. Если взять любую другую ранжировку, то сумма расстояний от нее до всех остальных окажется больше. Например, для ранжировки *D* = <*x, z, y*> матрица имеет вид

$$D = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix},$$

расстояния *d(A, D) = d(B, D)* = 2, *d (C,* D) = 4. Сумма расстояний от D до исходных ранжировок *A, B, C* равна 8.

Методы шкалирования. Одномерное шкалирование. Выполнение операции над параметрами, измеренными в разных шкалах

Введем экспертизу Э9:

$\Omega = E_n$

$\Omega_Э = \Omega$,

L – эксперты изолированы;

Q – обратная связь отсутствует.

Отображение $\varphi : \Omega_Э \to \Omega$ будет описано ниже.

Приведем операции, которые необходимо осуществить для построения φ.

1) Вычисляют матрицу $P = \sum_{j=1}^{N} A_j / N$, где A^j – ранжировка, данная j-м экспертом Элемент p_{ij} матрицы P интерпретируют как вероятность предпочтения i-го объекта j-му.

2) Находят Z_{ij} по формуле

$$G(Z_{ij}) = p_{ij} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_{ij}} e^{-t^2/2} dt \qquad (3)$$

с использованием таблиц нормального распределения, исходя из известных p_{ij}. Связь между Z_{ij} и p_{ij} показана на рис. 3 где заштрихованная площадь под кривой равна p_{ij}. Величина Z_{ij} измеряется в единицах стандартного отклонения.

3) Образуют матрицу Z = (Z_{ij}) Подсчитывают сумму оценок $Z_i = \sum_{j=1}^{n} Z_{ij}$ и среднее значение $\overline{Z}_i = Z_i / n$. Величину $\overline{Z}_i$ принимают за искомую оценку объекта A_i $\left(i = \overline{1,n}\right)$

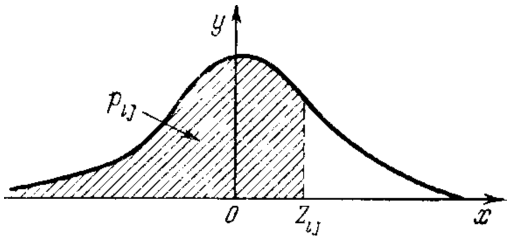

Рисунок 1. Ранжировка

4) Определяют величины $\overline{P}_i = G\left(\overline{Z}_i\right)$ по формуле (3), которые нормируют по формуле

$$P_i^* = \overline{P}_i / \sum_{j=1}^{n} \overline{P}_j$$

P_i^* называют *показателями относительной важности объектов.*

5) Осуществляют проверку на непротиворечивость. Для этого по формуле (3) находят $\overline{p}_{ij} = G\left(\overline{Z}_i - \overline{Z}_j\right)$ и вычисляют разности $\Delta_{ij} = \overline{p}_{ij} - p_{ij}$. Определяют среднее отклонение, если оно мало, то это свидетельствует о непротиворечивости полученных экспертных ранжировок.

$$\delta = \sum_{j=1}^{n} \sum_{i<j} \left| \Delta_{ij} \right| \Big/ (n(n-1))$$

Таблица 1

Ранжирование параметров

Эксперты	Параметры			
	S	R	c	p
1	3	2	1	4
2	1	2	3	4
3	3	1	2	4
4	1	2	3	4
5	3	1	2	4
6	3	1	2	4
7	3	2	4	1
8	3	4	1	2
9	2	4	1	3
10	2	1	3	4
Сумма рангов	24	20	22	34
Средний ранг	2,4	2,0	2,2	3,4

Пример 2 (использование Э9 для оценки относительной важности параметров самолета). В качестве экспертов приглашены 10 специалистов. Рассматривались следующие параметры: S – скорость полета, *R* – радиус действия, С – боевой потолок, *P* – полезная нагрузка. Экспертам было предложено ранжировать названные параметры в порядке их важности.

Результаты показаны в табл.1. Эксперт 7 расположил параметры по степени важности в порядке *P, R, S, C,* а эксперт 2 – в порядке S, *R, C, P.* Сформируем матрицу А размера 4х4, в которой указано число случаев, когда один параметр важнее другого (табл. 2) Разделив элементы матрицы. А на 10, получим матрицу *P.* С ее элементами осуществим все указанные выше операции. Матрица Z**,** величины Z_i и оценки $\overline{Z}_i$ приведены в табл.3. Величины, необходимые для расчета относительной важности параметров, приведены в табл.4.

Таблица 2

Матрица А

	S	R	C	P
S	-	4	4	8
R	6	-	7	7
C	6	3	-	9
P	2	3	1	-

Таблица 3

Параметры

Параметр i	Параметр j					
	S	R	C	P	Z_i	$\overline{Z}_i$
S	0	–0,25334	–0,25334	0,84161	0,33493	0,08373
R	0.25334	0	0,52441	0,52441	1,30216	0,35554
C	0,25334	–0,52441	0	1,28155	1.01048	0,25262
P	0.84161	–0,52441	–1,28155	0	-2,64757	-0,66189

Осуществим проверку на непротиворечивость. Необходимые данные собраны в табл.5.

Среднее отклонение в данном случае δ_{sr}=0,45482/6=0,0758. Наибольшее по абсолютной величине расхождение между p_{ij} и $\overline{p}_{ij}$ равно 0,17094; это свидетельствует о непротиворечивости ранжировок.

Таблица 4

Параметры

Параметр i	$\overline{Z}_i$	$\overline{p}_i = G(\overline{Z}_i)$	Нормированная относительная важность p_i^*
S	0,08373	0,53336	0,2647
R	0,32554	0,62761	0,3115
C	0,25262	0,59972	0,2977
P	–0,66189	0,25403	0,1261

Таблица 5

Проверка на непротиворечивость

$\overline{Z}_i - \overline{Z}_j$	$\overline{p}_{ij}$ расчетное	p_{ij} исходное	Отклонение Δ_{ij}
$\overline{Z}_1 - \overline{Z}_2$ = 0.08373 – 0,32554 = – 0,24181	0.40447	0,400	0,0047
$\overline{Z}_1 - \overline{Z}_3$ = 0,08373 – 0,25262 = – 0,16889	0,43297	0.400	0,032295
$\overline{Z}_1 - \overline{Z}_4$ = 0,08373 – (–0,6189) = 0,74562	0.77205	0,800	– 0,02795
$\overline{Z}_2 - \overline{Z}_3$ = 0.32554 – 0,25262 = 0,07292	0,52906	0,700	– 0.17094
$\overline{Z}_2 - \overline{Z}_4$ = 0.32554 – (–0,66189) = 0,98743	0,83828	0,700	0.13828
$\overline{Z}_3 - \overline{Z}_4$ = 0,25262 – (–0,66189) = 0,91451	0.81977	0,900	– 0,08023

Список литературы

[1]. Верескун В.Д. Экспертные оценки в производственно-транспортных процессах: вопросы организации, моделирования и управления // Фундаментальные исследования. – 2016. – № 4-3. – С. 485-489;

[2]. Варфоломеева А.О. Информационные системы предприятия: Учебное пособие / А.О. Варфоломеева, А.В. Коряковский, В.П. Романов. - М.: НИЦ ИНФРА-М, 2016. – 283 с.

[3]. Евграфов, М.А. Асимптотические оценки и целые функции / М.А. Евграфов, 2013. – 299 с.

[4]. Боровков, А.А. Математическая статистика: оценка параметров, проверка гипотез / А.А. Боровков, 2013. – 125 с

[5]. Акперов И. Г. Информационные технологии в менеджменте: учебник для вузов/ И. Г. Акперов, А. В. Сметанин, И. А. Коноплева. - Москва: ИНФРА-М, 2013. - 400 с.

[6]. Гуцыкова С.В. Метод экспертных оценок: теория и практика. – М.: Институт психологии РАН, 2011. – 144 с.

TECHNOLOGIES AND METHODS OF BIG DATA PROCESSING EXPERT INFOR-MATION - ALGEBRAIC METHOD AND METHOD OF SCALING

V.S.HLADKAYA
Assistant to department of engineering psychology and ergonomics BSUIR, master of technical science

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
E-mail: v.gladkaya@bsuir.by

Abstract. In this paper, the following goals were set: to study and apply scientific literature on the topic; compare the results of measuring the value with the points of the number line in accordance with the algebraic scaling method, perform operations on the parameters measured in different scales using the one-dimensional scaling method. All the goals set were considered and successfully achieved.

Key words: expert evaluation, expert analysis, scaling, algebraic method, one-dimensional scaling, ranking, median Kemen-Snell.