

# ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ ДАНИХ БАГАТООБ'ЄКТНОЇ ЕКСПЕРТИЗИ

О. М. Чмерук<sup>1, а</sup>, О. Є. Архипов<sup>1, б</sup>

<sup>1</sup> Національний технічний університет України  
«Київський політехнічний інститут» імені Ігоря Сікорського,  
Фізико-технічний інститут

## Анотація

Експертне оцінювання – здавна одна з найбільш поширених інформаційних технологій, яка і в наш час приваблює широке коло фахівців – як практиків, так і теоретиків. Пояснюється це рядом особливостей, притаманних методу експертних оцінок. Даний метод – найбільш доступний, універсальний, а іноді просто єдиний можливий для отримання та аналізу інформації, що використовується для вирішення широкого спектра задач управління, прогнозування, планування в найрізноманітніших сферах діяльності. При використанні методу експертних оцінок основним джерелом інформації є експерт – його судження, якісні та кількісні оцінки. Тобто експертні методи ґрунтуються виключно на оцінках експертів, зроблених щодо проблеми, яку вони вичерпно знають. При цьому механізм продукування цих оцінок лишається невизначеним. Як правило, він невідомий навіть самому експертові, має виключно індивідуальний, особистий характер і не може бути повторений чи відтворений кимсь іншим. Це обумовлює особливі вимоги щодо вибору складу експертів, зокрема рівня їх компетентності, адже недостатній рівень компетентності експерта може призвести до появи грубих (аномальних) помилок у даних експертизи чи просто зумовити високий рівень неоднорідності цих даних. Тому ефективна, якісна обробка експертних даних значною мірою визначає коректність й правильність виконання всієї експертизи в цілому. У цій ситуації особливу важливість і актуальність набуває проблема аналізу та обробки експертної інформації, так як виконання тільки цих заходів дозволяє забезпечити якість рішень, що приймаються.

**Ключові слова:** оцінки, компетентність, експертиза, методи, обробка

## Вступ

Відсутність загально визнаних формально-теоретичних і методичних положень, що пояснюють механізм формування експертних даних (і в тому числі похибок у цих даних), обумовлює той факт, що загальні рекомендації по обробці експертних даних, тим більш реалізовані в формі програмного продукту, на сьогодні відсутні. Оброблювач, що використовує експертні дані, найчастіше сам вирішує, як з ними працювати. Математичні методи обробки, що застосовуються, як правило, дуже прості, бо використання більш складних методик обробки потребує залучення додаткової інформації, зазвичай відсутньої. Все сказане вище обумовлює актуальність досліджень, пов'язаних з вивченням і розробкою методів або методик обробки експертних даних. Особливо актуальна проблема обробки даних експертизи в нових сферах людської діяльності, в яких ще недостатньо сформувався формально-теоретичний базис, не структурована різноманітність властивостей і особливостей об'єктів, що вивчаються, не сформувалась достатня кількість спеціалістів, за якістю своєї підготовки адекватним вимогам, що пред'являються рівню експерта [1, стор. 214].

## 1. Оцінювання якості роботи експертів за даними багатоб'єктної експертизи

Серед множини експертних технологій, що залучаються для вирішення різноманітних завдань, можна виділити досить розповсюджений вид колективної експертизи, яка називається багатоб'єктною експертизою (БОЕ) [2, 3]. У БОЕ бере участь група з  $N$  експертів, кожен з яких здійснює індивідуальну експертизу  $M$  об'єктів, які складають сукупність, що експертується. Отримані в ході індивідуальних експертиз підмножини з експертних оцінок зводяться в загальну матрицю даних, що підлягають наступній спільній обробці:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1N} \\ z_{12} & z_{22} & \cdots & z_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ z_{M1} & z_{M2} & z_{M3} & z_{MN} \end{bmatrix} = [Z_1, Z_2, \dots, Z_n]. \quad (1)$$

Особливістю БОЕ є достатньо великі обсяги об'єктів, що підлягають експертизі, у якості яких можуть виступати зразки певних типів продукції, виробів, інформаційні продукти (зокрема, програмне забезпечення), художні або літературні твори, списки питань (опитувальники) у соціологічних або психологічних дослідженнях, прислані на конкурс проекти і т.п. Зокрема, робота постійно діючих груп

<sup>а</sup>choleks2@mail.com

<sup>б</sup>sonet0515@mail.com

(комісії) експертів, що діють у певній фаховій сфері, реалізуючи експертизу більш-менш усталеного класу (чи класів) інформаційних продуктів, є типовим зразком БОЕ. Отримані в цьому випадку відносно великі об'єми індивідуальних експертних оцінок серед іншого містять певну інформацію про особисті якості експертів, зокрема, про рівні їхньої компетентності, знання яких досить актуальні для організації ефективної обробки результатів експертизи. Результати індивідуальної експертизи, здійсненої  $j$ -м експертом, являють собою випадкову послідовність  $Z_i = [z_{1j}, z_{2j}, \dots, z_{Mj}]^T$ , кожен елемент якої містить інформативну складову  $x_{i0}$ , спільну для всіх експертних оцінок  $z_{ij}$ , і випадкову похибку  $e_{ij}$ , характеристики якої індивідуальні у кожного конкретного експерта:

$$z_{ij} = x_{i0} + e_{ij}, i = \overline{1, M}, j = \overline{1, N}. \quad (2)$$

Поставивши у відповідність елементам послідовності  $Z_j$  цілочислові моменти часу  $t_i = 1, 2, \dots, M$ , отримаємо певний аналог часового ряду  $(z_{ij})$ , у загальному випадку нестационарного. Однак якщо припустити, що характеристики експерта як деякої інформаційно-аналітичної оцінюючої системи залишаються незмінними впродовж процедури експертування, а всі похибки, помилки та неточності в експертних оцінках визначаються винятково властивостями і станом експерта на момент проведення експертизи, то справедливим уявляється припущення щодо стаціонарності та ергодичності випадкових послідовностей  $E_j = e_{1j}, e_{2j}, \dots, e_{Mj}, j = \overline{1, N}$ .

За аналогією з відомими положеннями кластерного аналізу [3, 4] введемо поняття образу експерта як деякої точки  $Z_j = [z_{1j}, z_{2j}, \dots, z_{Mj}]^T, j = \overline{1, N}$  у  $M$ -вимірному просторі результатів БОЕ. При повному збігові думок експертів їхні образи однакові, тобто всі результати експертизи будуть представлені єдиною точкою в просторі результатів БОЕ. Наявність помилок експертів приводить до розщеплення точки в хмару (кластер), щільність якої (якого) неоднорідна і зазвичай максимальна в області, що прилягає до центру кластера з координатами  $Z_0 = [z_{10}, z_{20}, \dots, z_{M0}]^T$ , визначеними співвідношенням [4, 5]:

$$Z_0 = \underset{Z_j, Z_0 \in R^M}{\operatorname{argmin}} \sum_{j=1}^N r_j(Z_j, Z_0), \quad (3)$$

де  $r_j(Z_j, Z_0)$  – відстань між образом  $j$ -го експерта і центром  $Z_0$  кластера в  $M$ -вимірному просторі результатів БОЕ. При використанні для знаходження  $r_j(Z_j, Z_0)$  евклідової метрики:

$$r_j = r_j(Z_j, Z_0) = \left( \sum_{i=1}^M (z_{ij} - z_{i0})^2 \right)^{\frac{1}{2}}, j = \overline{1, N}, \quad (4)$$

мінімізація співвідношення (3) досягається при:

$$z_{i0} = \frac{1}{N} \sum_{j=1}^N z_{ij}. \quad (5)$$

Наявність анамальних даних у рядках матриці  $Z$  спричиняє зміщення оцінок (4), у зв'язку із чим

більш надійний результат (при ймовірній наявності «аномальних» експертів) дає застосування робастних медіанних оцінок виду:

$$z_{i0} = \operatorname{med}(Z_i) = \operatorname{med}(z_{i1}, z_{i2}, \dots, z_{iN}). \quad (6)$$

При відсутності анамальних експертів справедливе припущення про рівність нулю математичних очікувань помилок експертизи:  $\mu(E_j) = 0, j = \overline{1, N}$ , що приводить до виконання рівності  $\mu(Z_0) = X_0 = [X_{10}, X_{20}, \dots, X_{M0}]^T$  і дозволяє обґрунтувати гіпотезу незміщеності середньогрупових експертних оцінок:

$$\begin{aligned} \mu(\overline{Z}_i) &= \mu\left(\frac{1}{N} \sum_{j=1}^N Z_{ij}\right) = \frac{1}{N} \mu\left(\sum_{j=1}^N x_{i0} + e_{ij}\right) = \\ &= \frac{1}{N} (N x_{i0}) = x_{i0}. \end{aligned} \quad (7)$$

Метризація віддаленості образів експертів від центра дозволяє представити в інтегрованій формі інформацію про помилки кожного з експертів і допускає можливість існування шкального перетворення  $c_j = f(r_j)$ , що забезпечує взаємодозначне відображення елементів множини  $R(r_j \in R, j = \overline{1, N})$  у відповідні оцінки компетентності експертів  $f: R \rightarrow C; c_j \in C$ . Вибір структури й параметрів відображення  $f$  представляє нетривіальну задачу, що вимагає окремого розгляду. При формуванні вимог до шкального перетворення  $c = f(r)$  будемо виходити з наступних міркувань. По-перше, очевидно, що із ростом компетентності  $C$  значення  $r$  зменшуються, тобто похідна  $\frac{dc}{dr} < 0$ . Із цього виходить також твердження про монотонний характер залежності  $c = f(r)$ . По-друге, при побудові шкали виміру компетентності, множини можливих значень  $r$ , визначену на напіввідкритому інтервалі  $R = [0, \infty)$  зручно відображати в замкнутий інтервал  $C = [1, 0]$ , що відповідає типовій шкалі компетентності. При цьому значенням  $r \rightarrow \infty$  відповідає права гранична відмітка  $= 0$  шкали компетентності, а значенню  $r = 0$  – ліва,  $= 1$ . Для малих значень оцінок  $r$ , враховуючи, що величина похибки оцінювання в цьому випадку може бути зіставимою або навіть істотно перевищувати невідоме істинне значення відстані  $r$ , з метою зменшення впливу похибки на точність значень компетентності, доцільно ввести умову:

$$\frac{dc}{dr} \approx 0. \quad (8)$$

При цьому для області малих значень  $r$  буде справедливе співвідношення  $f(r) = c \approx 0$ . Умову, аналогічну (8), варто ввести і для області великих значень  $r$ , яка прилягає до правого кінця інтервалу  $R = [0, \infty)$ . Тоді точки цієї досить протяжної області великих значень  $r$  (що відповідають суттєво віддаленим образам малокомпетентних експертів від центра  $Z_0$ ) будуть відображатися в значеннях компетентності, рівні або близькі 0. У підсумку, якщо припустити, що значення похідної  $\frac{dc}{dr}$  максимальні (за модулем) у центральній частині шкали й зменшуються, прагнучи до 0, з наближенням до периферії

шкали, справедливе співвідношення:

$$\frac{dc}{dr} = -c(b_0 - b_1c), b_0, b_1 > 0, b_0 \geq b_1. \quad (9)$$

Квадратичний зсув  $b_1c^2$  у правій частині (9) дозволяє реалізувати виконання умови (8) в області великих значень  $r$ . У цілому співвідношення (9) являє собою диференціальне рівняння з відокремлюваними змінними, яке розв'язавши матимемо:

$$\ln \frac{c}{b_0 - b_1c} = -b_0r + \ln A, \quad (10)$$

де  $A$  – постійна інтегрування. З урахуванням граничної умови  $c(0) = 1$ , після потенціювання і ряду перетворень, вводячи сталу  $B = \frac{b_1}{b_0}$ , отримаємо:

$$c = f(r) = \frac{1}{(1 - \frac{b_1}{b_0}) \exp(b_0r) + \frac{b_1}{b_0}} = \quad (11)$$

$$= \frac{1}{(1 - B) \exp(b_0r) + B}. \quad (12)$$

Графік залежності  $c(r)$ , наведений на Рис. 1, за своїм характером – «перевернута» логістична крива.

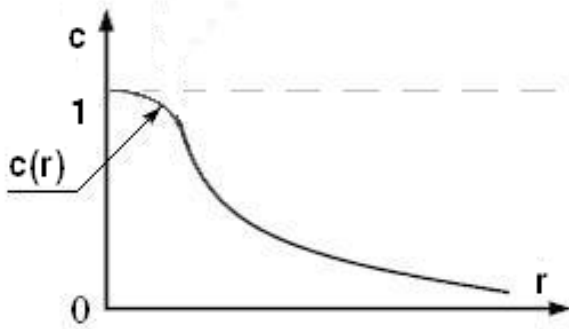


Рис. 1. Графік залежності  $c(r)$

Отримані вище результати, на жаль, носять суто прикладний характер. Розраховані за формулою (4) оцінки  $r$  дозволяють зіставляти рівні помилок експертів тільки в рамках конкретної задачі. Це обумовлюється тим, що знайдені оцінки  $r$  залежать від числа об'єктів експертизи  $M$  та кількісних характеристик прийнятої шкали оцінок. Отримана на базі достатньо загальних і об'єктивних передумов структура шкального перетворення для своєї конкретизації й прикладного застосування вимагає завдання кількісних значень параметрів  $b_0$ ,  $B$ , що стає можливим лише при визначенні конкретного типу експертизи.

Тому далі, для одержання можливості більш деталізованого врахування особливостей і характеру БОЕ, трохи звужимо клас досліджуваних процедур експертизи Розглянемо досить розповсюджену на практиці процедуру БОЕ, у якій використовується бальне оцінювання.

Якщо процедура БОЕ полягає в оцінюванні кожного з об'єктів експертизи в бальній шкалі  $0, 1, 2, \dots, l_{max}$ , тобто  $z_{ij} \in 0, 1, \dots, l_{max} = L$ , то тео-

ретично можливими мінімальними й максимальними значеннями  $r$  будуть  $r_{min} = 0$  і  $r_{max} = l_{max}\sqrt{M}$ . Вводячи у вираз (4) нормуючі множники  $\frac{1}{l_{max}}$  та  $\frac{1}{\sqrt{M}}$ , отримаємо формулу для обчислення нормованої відстані образу  $j$ -ого експерта від центру кластера  $Z_0$ :

$$r_{Hj} = \frac{r_j}{l_{max}\sqrt{M}}. \quad (13)$$

У разі, коли поділки бальної шкали починаються з одиниці, тобто  $1, 2, \dots, l_{max}$ , нормована відстань образу  $j$ -ого експерта від центру кластера  $Z_0$  обчислюється за формулою:

$$r_{Hj} = \frac{r_j}{(l_{max} - 1)\sqrt{M}}. \quad (14)$$

Нормована відстань не залежить від числа  $M$  об'єктів, що підлягають експертизі, і кількості відліків бальної шкали, тобто від  $l_{max}$ , будучи індивідуалізованою оцінкою експерта, що враховує тільки величину і характеристики розподілу помилок експерта.

Досвід практичної роботи з даними БОЕ свідчить, що значення  $r_H \leq 0,2$  характерні для експертів досить високої кваліфікації, значення  $r_H \gg 0,3 \div 0,35$  відповідає образам експертів, що мають відносно невисокий рівень професійної підготовки, нерівно проводять експертизу і припускають у своїх оцінках помилки досить значної величини.

Зокрема, подібну характеристику шкальному перетворенню забезпечують наступні значення параметрів:  $b_0 = 15$ ,  $B = 0,967$  (Рис. 2). Слід зазначити, що у цьому випадку безпосереднє виявлення й виключення з обробки даних «аномальних» експертів відсутнє. Це пояснюється тим, що при здійсненні обробки із введенням ваг, пропорційних компетентності експертів, дані, отримані від експертів, для яких  $r_H > 0,4$ , фактично обнуляються.

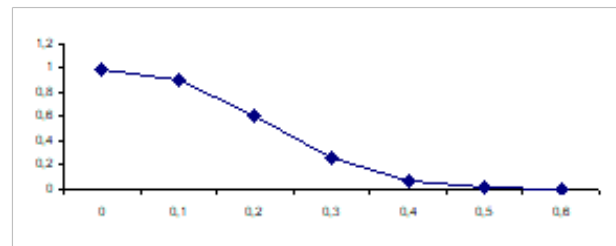


Рис. 2. Графік характеристики шкального перетворення (11) з параметрами:  $b_0 = 15$ ,  $B = 0,967$

Очевидно, що завдання параметрів шкального перетворення містить істотний суб'єктивний момент і визначається цілями перетворення, особливостями прийнятої моделі розподілу похибок оцінок експертизи, застосовуваним способом кількісного оцінювання рівня компетентності (бальна шкала, шкала з одностороннім обмеженням, шкала із двостороннім обмеженням і т.п.).

## 2. Методи обробки

10 рядків експертних оцінок, кожний рядок – вектор  $Z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$ ,  $i = \overline{1, M}$ ,  $M = 10$ ,  $N = 6$

– дані групової експертизи чергового об'єкту експертизи, по одній оцінці від кожного з шести експертів ( $N = 6$ ), склад яких залишається незмінним впродовж всієї серії з 10 експертиз).

Обираємо наступні вихідні дані:

242453	334433
221942	343467
321365	584445
654766	889771
989899	879658

Вибрані дані цікаві тим, що в них присутня стрічка із аномальним експертом: 221942.

## 2.1. Оцінка компетентності експерта за результатами обробки даних багатооб'єктної експертизи

Заносимо обраховані значення до таблиць 1-4.

Табл. 1. Рівні компетентності

Номер експер.	$r_j$	$r_{nj}$	$c_j$
1	2,95	0,12	0,86
2	4,01	0,16	0,76
3	5,3	0,21	0,58
4	6,17	0,24	0,45
5	4,33	0,17	0,72
6	6,64	0,26	0,38

Табл. 2. Середньозважені бали

Номер об'єк.	$q_{jc}$	$q_j$
1	5,35	3,33
2	5,35	3,33
3	5,35	3,33
4	7,23	4,5
5	5,35	3,33
6	8,03	5,0
7	9,1	5,67
8	10,71	6,67
9	13,92	8,67
10	11,51	7,17

Табл. 3. Рівні компетентності з використанням медіанних оцінок

Номер експер.	$r_j$	$r_{nj}$	$c_j$
1	2,0	0,08	0,93
2	4,0	0,16	0,76
3	4,24	0,17	0,73
4	7,42	0,29	0,27
5	5,1	0,2	0,61
6	7,48	0,3	0,27

Табл. 4. Середньозважені бали з використанням медіанних оцінок

Номер об'єк.	$q_{ic}$	$q_j$
1	5,61	3,33
2	5,61	3,33
3	5,61	3,33
4	7,58	4,5
5	5,61	3,33
6	8,42	5,0
7	9,54	5,67
8	11,23	6,67
9	14,6	8,67
10	12,07	7,17

## 2.2. Медіанні оцінки

Обраховуємо значення медіанних оцінок та заносимо результати до таблиці (табл. 5).

Табл. 5. Медіанні оцінки

Номер об'єк.	Значення	Дані
1	3,5	[242453]
2	3,0	[334433]
3	2,0	[221942]
4	4,0	[343467]
5	3,0	[321365]
6	4,5	[584445]
7	6,0	[654766]
8	7,5	[889771]
9	9,0	[989899]
10	7,5	[879658]

## 2.3. Оцінки середнього

Обраховуємо середнє значення оцінок групової експертизи та заносимо результати до таблиці

Табл. 6. Оцінки середнього

Номер об'єк.	Значення	Дані
1	3,33	[242453]
2	3,33	[334433]
3	3,33	[221942]
4	4,5	[343467]
5	3,33	[321365]
6	5,0	[584445]
7	5,67	[654766]
8	8,67	[889771]
9	8,67	[989899]
10	7,17	[879658]

## 2.4. Збіжність ітераційних процедур

1-й крок:

$$j = 1; x_{01} = \frac{1}{n} \sum_{i=1}^n x_i; w_{i1} = \frac{1}{|x_i - x_{01}|}$$

2-й крок:

$$j = 2; x_{02} = \frac{1}{\sum_i w_{i1}} \sum_i w_{i1} x_i; w_{i2} = \frac{1}{|x_i - x_{02}|}$$

j-й крок:

$$x_{0j} = \frac{1}{\sum w_{ij-1}} \sum w_{ij-1} x_i$$

Умова завершення ітеративного циклу:

$$|x_{ij} - x_{0j}| \approx |x_{ij-1} - x_{0j-1}|,$$

аналогічно  $x_{0j-1} \approx x_{0j}$ .

Табл. 7. Обрахунок збіжності ітераційних процедур

Номер об'єк.	Знач. X	К-сть ітер.	Дані
1	3,33333	2	[242453]
2	3,00098	10	[334433]
3	2,00041	10	[221942]
4	4,0	6	[343467]
5	3,0	5	[321365]
6	4,95385	3	[584445]
7	5,99967	8	[654766]
8	7,16019	3	[889771]
9	8,99902	10	[989899]
10	7,16667	2	[879658]

## 3. Висновки

Одним із ключових питань обробки експертних даних є визначення рівня показника компетентності експертів й урахування цього рівня при остаточній обробці висновків експертизи. Вага показника компетентності зумовлена тим фактом, що однастайність висновків експертів з проблеми, яка аналізується – це ймовірне свідомство або тривіальності означеної проблеми, або відсутності можливості вільного висловлення кожним з експертів своєї особистої думки. Кожен з використаних методів оцінки результатів експертизи дав однакові результати – найвищі оцінки отримав об'єкт під номером – 9.

Принциповою особливістю процедури експертизи є можливість отримання істотно відмінних чи навіть несумісних оцінок за темою експертування, і як наслідок цього – проблема коректного синтезу остаточного експертного висновку, розв'язання якої можливе тільки з урахуванням об'єктивно визначеного рівня компетентності експертів.

Обчислені значення нормованої відстані виставлених експертами оцінок від центра кластера свідчить про високу кваліфікацію чотирьох експертів, натомість двоє експертів мають відносно невисокий рівень професійної підготовки.

В результаті проведеного дослідження методів обробки було з'ясовано, що метод збіжності ітеративних процедур дає найбільш точні оцінки при існуванні розкиду в виборці даних. У порівнянні з ним звичайні медіанні та середні оцінки дають дуже грубі значення.

## Перелік використаних джерел

1. О. Архипов Є. Вступ до теорії ризиків: інформаційні ризики. — 2015. — С. 248.
2. О. Архипов Є. Критерії визначення можливої шкоди національній безпеці України у разі розголошення інформації, що охороняється державою: монографія / О.Є. Архипов, О.Є. Муратов. — К: Наук.-вид. відділ НА СБ України. — 2011. — С. 195.
3. О. Архипов Є., С. Архипова. А. Оцінювання якості роботи експертів за даними багатооб'єктної експертизи // Захист інформації. — 2011. — №4(53) — С. 45-54.
4. А. Архипов Е. Применение методов кластеризации в задаче обработки данных экспертного опроса / А.Е. Архипов, С.А. Архипова, С.А. Носок, И.В. Пишко // Радиоелектроніка, інформатика, управління. — 2003. — №2(10) — С. 104-108.
5. А. Архипов Е. Модели компетентности эксперта / А.Е. Архипов, С.А. Архипова, С.А. Носок // Міжнародна наукова конференція Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій (ISMIT-2006), м. Євпаторія, 15-19 травня 2006. — том 1. — 2006. — С. 22-25.