

УДК 519.816

С. В. Каденко, В. В. Циганок

Інститут проблем реєстрації інформації НАН України
вул. М. Шпака 2, 03113 Київ, Україна

Підходи до оцінки ефективності експертних методів

Показано, що визначення точності експертних методів становить суттєву проблему внаслідок відсутності еталонних значень оцінок об'єктів в умовах реальних експертиз. Натомість, запропоновано використовувати показник ефективності методів, заснований на їхній стійкості до збурень вихідних даних. На прикладі порівняння комбінаторного методу агрегації експертних оцінок з урахуванням і без урахування ваг покривних дерев досліджено два підходи до оцінки ефективності експертних методів. Один підхід ґрунтується на використанні даних реального експертного оцінювання заданих еталонних об'єктів; другий — на імітаційному моделюванні усього циклу експертизи, в тому числі, й самих експертних оцінок. У ході дослідження ефективності двох указаних методів агрегації експертних оцінок імітаційний підхід показав себе як більш коректний і репрезентативний. Отримано експериментальні результати, які емпірично підтверджують перевагу комбінаторного методу з урахуванням ваг покривних дерев над комбінаторним методом без урахування ваг (і, відповідно, над методами геометричного середнього по рядках і логарифмічним методом найменших квадратів).

***Ключові слова:** підтримка прийняття рішень, слабкоструктурована предметна область, експертна оцінка, попарні порівняння.*

Вступ: сутність, актуальність і стан проблеми визначення ефективності експертних методів

Методи експертного оцінювання являють собою дієвий інструмент підтримки прийняття рішень у слабкоструктурованих предметних областях, ознаки яких наводяться у багатьох джерелах, зокрема, в [1]. У даному дослідженні пропонується зосередитися на таких ознаках слабкоструктурованих предметних областей як відсутність еталонів, неповнота відомостей про об'єкти, та значна роль людського фактору. Саме вони зумовлюють проблему визначення ефективності методів експертного оцінювання.

© С. В. Каденко, В. В. Циганок

У той час як вимірювання об'єктів за кількісними параметрами (такими як довжина, маса тощо) — це їхнє порівняння з певними еталонними величинами (метр, кілограм), експертне оцінювання застосовується як альтернатива вимірюванням у ситуаціях, де об'єкти неможливо виміряти. За таких умов єдиним джерелом кількісної інформації про об'єкти стають експертні оцінки цих об'єктів. Експерти можуть давати оцінки як безпосередньо у певній шкалі (ординальній чи кардинальній, чисельній чи вербальній, згоди-незгоди тощо), так і шляхом попарних порівнянь об'єктів. На думку багатьох дослідників, найкращий спосіб оцінити низку об'єктів за «неосяжним» (intangible) критерієм — саме порівняти їх між собою. З цього припущення постала велика кількість методів, які базуються на використанні парних порівнянь об'єктів. Зокрема, слід згадати методи аналізу ієрархій (MAI) та мереж [2, 3], TOPSIS [4], методи «лінія», «трикутник», «квадрат» [5], комбінаторний метод [6], метод логарифмічних найменших квадратів (ЛМНК) [7] тощо.

Наголосимо, що показниками ефективності методів, які оперують детермінованими даними, зазвичай, слугують різноманітні міри відхилення реальних (експериментальних) значень від еталонних (середнє відхилення, середньоквадратичне відхилення, математичне сподівання похибки, евклідова відстань, тощо). У випадку методів, які оперують експертними даними, постає питання: «з чим порівнювати ці дані та результати їхньої обробки?» (адже в загальному випадку еталонні значення відсутні). Слід також зазначити, що, організовуючи експертизу, особа, що приймає рішення (ОПР), виходить з евристичного припущення щодо існування певних «істинних» значень оцінок об'єктів, чи кількісних співвідношень між ними (так званої «глибинної істини», ground truth). Отже, ключове питання полягає в тому, які індикатори можуть вказувати на ступінь точності та достовірності експертних даних і методів їхньої агрегації? Від цих індикаторів залежатиме рівень довіри ОПР до експертних рекомендацій, і (як наслідок) довіри до рішення, прийнятого на їхній основі.

У літературі викладено порівняно мало підходів до визначення точності експертних методів (якщо поняття «точність» взагалі коректно використовувати, з огляду на характер сфери їхнього застосування).

Наукова школа Т. Сааті вимагає від експертних даних, що вводяться до матриць попарних порівнянь (МПП), у першу чергу, узгодженості або ж відсутності порушень транзитивності (в ординальному та кардинальному сенсі) [2]. У праці [3] фігурує поняття «легітимності» результатів експертиз, тобто їхнього співпадіння з незалежним вибором ОПР (у [3] автор наводить цікавий приклад прийняття рішення щодо вибору локації для будівництва Диснейленду в Китаї). Н.Д. Панкратова та Н.І. Недашківська у праці [8] демонструють, що результати, отримані методом аналізу мереж та авторським гібридним методом, співпадають, що, на думку авторів, свідчить про достовірність результатів роботи гібридного методу. У праці Еліота [9] експерти порівнюють декілька шкал оцінювання та самі визначають, яка шкала дозволяє їм найточніше висловити їхні переваги. Аналогічний підхід використаний у праці [10], де експерти самі обирають той результат агрегації їхніх оцінок, який, найбільш достовірно відображає їхні уявлення про предметну область.

Спільним для щойно перелічених підходів є відсутність еталонних значень оцінок (в умовах реальних експертиз еталони дійсно, зазвичай, відсутні). Принципово інший підхід — тестування експертних методів на спеціально згенерованій множині «еталонних» об'єктів, для яких точно відомі значення оцінок за певним критерієм. Варто згадати експеримент, що описаний у [2], в якому респондентам пропонується оцінити співвідношення площ кількох фігур. При цьому реальні співвідношення цих площ відомі лише організатору експертизи. З цими реальними співвідношеннями порівнюються значення, отримані на основі експертних оцінок груповим МАІ.

В Україні подібний підхід застосовано в [11, 12]: близько 20-ти експертних методів порівняно за трьома критеріями ефективності: точністю, тривалістю роботи та узгодженістю результатів оцінювання. Експерти почергово оцінюють різними методами 7 об'єктів (кількість обрано з урахуванням психофізіологічних обмежень людини [13]). Реальні (еталонні) співвідношення ваг об'єктів (частково зафарбованих зображень) відомі лише організатору експертизи. Агреговані результати оцінювання порівнюються з еталонними значеннями. Методи, які дають меншу середню похибку, вважаються кращими за точністю.

Ще один підхід до визначення точності експертних методів не передбачає участі експертів: йдеться про моделювання (simulation) експертиз. Моделюються не лише ваги еталонних об'єктів, а й самі оцінки. Зокрема, імітаційне моделювання МПП використовується для визначення порогів CI та CR (значення порогів постійно оновлюються зі зростанням кількості згенерованих МПП [3]). Приклади моделювання МПП можна також знайти в праці [14].

Комбінаторний метод: огляд

Вперше комбінаторний метод агрегації парних порівнянь був запропонований на початку 2000-х років [6]. Задача, як і в МАІ, полягає у знаходженні вектора відносних ваг n об'єктів за МПП цих об'єктів, які задані одним або кількома експертами. Основна ідея методу — максимально повне використання інформації, заданої експертами у вигляді МПП. У загальному випадку ця інформація є надлишковою. Тому комбінаторним шляхом почергово перебираються усі можливі базисні (не надлишкові, але інформативно змістовні) множини парних порівнянь, які можна сформуувати з елементів заданої МПП. Іноді, користуючись аналогією з теорією графів, ці множини називають покривними деревами (spanning trees). З відомої теореми Кейлі про дерева [15] випливає, що за повною МПП n об'єктів (розмірністю $n \times n$) можна сформуувати n^{n-2} таких базисних множин. За кожною із базисних множин можна знайти вектор відносних ваг об'єктів. Після цього шляхом звичайного або ж зваженого усереднення можна знайти агрегований вектор ваг об'єктів.

Практичне значення комбінаторного методу полягає в тому, що він є ключовим інструментом агрегації експертних оцінок, яка використовується у технології стратегічного планування у слабкоструктурованих предметних областях, що описана, зокрема, в [16].

У 2010 році показано ефективність комбінаторного підходу порівняно з іншими методами агрегації парних порівнянь [17]. У 2012 році метод був «відкритий вдруге» групою зарубіжних науковців [18, 19]. У 2017 році запропоновано

удосконалені формули для обчислення відносної компетентності експертів з урахуванням якості інформації, що вводиться цими експертами до МПП [20]. Протягом останніх кількох років доведено еквівалентність комбінаторного методу та ЛМНК [7] (для повних і неповних, адитивних і мультиплікативних МПП), а також середнього геометричного по рядках МПП [21]. Втім, методи є еквівалентними лише у випадку, коли у комбінаторному методі для визначення агрегованого вектора ваг альтернатив використовується звичайне, а не зважене геометричне середнє:

$$w_j^{aggregate} = \left(\prod_{q=1}^T w_j^q \right)^{\frac{1}{T}}, j = 1..n, T \in [1..mn^{n-2}], \quad (1)$$

де T — загальна кількість базисних множин парних порівнянь (покривних дерев); $\{w_j^q; j = 1..n, q = 1..T\}$ — множина відносних ваг n об'єктів, обчислених за базисною множиною парних порівнянь з номером q ; $(w_j^{aggregate}, j = 1..n)$ — агреговані ваги об'єктів; m — кількість експертів.

Принциповою відмінністю модифікованого комбінаторного методу є використання «рейтингів» базисних множин парних порівнянь, що визначаються повнотою, детальністю, та узгодженістю (внутрішньою та взаємною) даних, уведених експертами до індивідуальних МПП. Відповідно, якщо у формулі (1) врахувати рейтинги векторів відносних ваг альтернатив, отриманих на основі ідеально узгоджених матриць, що побудовані за кожною базисною множиною попарних порівнянь (для кожного покривного дерева), то отримаємо наступний вираз для агрегованих значень ваг об'єктів:

$$w_j^{aggregate} = \prod_{k,l=1}^m \left(\prod_{q^k=1}^{T_k} (w_j^{(kq^k l)})^{\frac{R_{kq^k l}}{\sum R_{upv}}} \right), j = 1..n. \quad (2)$$

Адитивний вигляд формули рейтингу базисної множини попарних порівнянь:

$$R_{kql} = \frac{c_k c_l s^{kq} s^l}{\ln(\sum_{u,v} |a_{uv}^{kq} - a_{uv}^l| + e)}. \quad (3)$$

Мультиплікативний вигляд формули рейтингу базисної множини попарних порівнянь:

$$R_{kql} = \frac{c_k c_l s^{kq} s^l}{\ln(\prod_{u,v} \max(\frac{a_{uv}^{kq}}{a_{uv}^l}, \frac{a_{uv}^l}{a_{uv}^{kq}}) + e - 1)}. \quad (4)$$

У формулах (3) та (4) k, l — номери експертів ($k, l = 1..m$), чий МПП порівнюються; c_k, c_l — апіорні значення відносної компетентності цих експертів; k та l

можуть бути рівними чи різними; q — номер «копії» ідеально узгодженої МПП ($q = 1..mT_k, T_k \in (0, \dots, n^{n-2})$); s^{kq} — відносна усереднена вага шкал, у яких вводилися парні порівняння з відповідної базисної множини, що визначається на основі формули Хартлі [22]; s^l — середня вага шкал, у яких вводилася відповідна індивідуальна МПП експерта номер l .

$$s^{kq} = \left(\prod_{u=1}^{n-1} \log_2 N_u^{(kq)} \right)^{\frac{1}{n-1}} \quad (5)$$

$$s^k = \left(\prod_{\substack{u,v=1 \\ v>u}}^n \log_2 N_{uv}^{(k)} \right)^{\frac{2}{n(n-1)}} \quad (6)$$

У формулах (5) та (6) N — кількість поділок у шкалі, в якій задане відповідне парне порівняння. Більш детальні пояснення щодо вигляду формул рейтингів базисних множин парних порівнянь наведені у [20].

Для ілюстративності пропонується розглянути чисельний приклад. Нехай 3 рівно-компетентних експерти E_1, E_2, E_3 ($c_1 = c_2 = c_3 = 1$) порівнюють 4 об'єкти (A_1, A_2, A_3, A_4) у цілочисельних шкалах. За МПП, які задані експертами, треба обчислити відносні ваги об'єктів. Кількість градацій шкал, обраних експертами наведено в табл. 1. У табл. 2 наведено номери поділок відповідних шкал, обраних експертами; у табл. 3 — власне, МПП, приведених до єдиної шкали.

Таблиця 1. Кількість градацій у шкалах, обраних експертами для здійснення порівнянь відповідних пар об'єктів

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	9	8	7	1	3	4	5	1	9	9	8
A_2		1	6	5		1	6	7		1	3	9
A_3			1	4			1	8			1	7
A_4				1				1				1

Таблиця 2. Номери поділок обраних експертами шкал парних порівнянь

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	4	7	1	3	4	5	1	2	4	8
A_2		1	2	4		1	2	3		1	2	5
A_3			1	2			1	2			1	3
A_4				1				1				1

Таблиця 3. Значення парних порівнянь, приведені до єдиної шкали

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	4 1/3	8 5/6	1	7 1/2	8 1/6	8 1/2	1	2	4	9
A_2	1/2	1	2 2/7	6 1/2	1/7	1	2 2/7	3 1/2	1/2	1	3 1/2	5
A_3	2/9	3/7	1	2 5/6	1/8	3/7	1	2	1/4	2/7	1	3 1/2
A_4	1/9	1/6	1/3	1	1/8	2/7	1/2	1	1/9	1/5	2/7	1

За трьома МПП з табл. 3 будуються 48 ідеально узгоджених МПП (ІУМПП) ($mn^{n-2} = 3 \times 4^2 = 48$). Кожна ІУМПП будується на основі відповідної базисної множини парних порівнянь (покривного дерева). Наприклад, базисній множині порівнянь пар об'єктів (A_1, A_2) , (A_1, A_3) та (A_2, A_4) відповідає покривне дерево, що показано на рис. 1.

За множиною відповідних елементів МПП, заданої другим експертом E_2 , будується ІУМПП наступного вигляду (жирним курсивом позначені базисні порівняння, решта порівнянь — обраховується за принципом транзитивності):

Аналогічним чином будуються ІУМПП за усіма базисними множинами парних порівнянь з МПП кожного з 3 експертів. При цьому, для перевірки внутрішньої та взаємної узгодженості матриць, ІУМПП порівнюються з початковими МПП експертів. Тож, будується по 3 копії кожної ІУМПП. Отже, загальна кількість ІУМПП, яку слід аналізувати, становить $T = m^2 n^{n-2} = 144$. Кожній ІУМПП присвоюється рейтинг, що обчислюється за формулою (4). Наприклад, для ІУМПП, наведеної у табл. 4, при порівнянні її з МПП першого експерта E_1 , ненормоване значення рейтингу дорівнює 1,191. Коли обчислені рейтинги усіх ІУМПП, вони нормуються за сумою (див. показник степеня у формулі (2)). За кожною «копією» ІУМПП з номером $q = 1..144$ будується вектор відносних ваг об'єктів — $w_1^q, w_2^q, w_3^q, w_4^q$ (оскільки матриця ідеально узгоджена, як такий вектор можна брати будь-яку базисну множину парних порівнянь, наприклад, перший рядок). Після цього вектори ваг об'єктів, отримані за всіма ІУМПП, усереднюються за формулою (2).

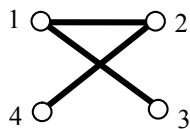


Рис. 1. Приклад графа покривного дерева для чотирьох об'єктів

Таблиця 4. Приклад ІУМПП

1	7 1/2	8 1/6	26 1/4
1/7	1	1	3 1/2
1/8	1	1	3 1/5
0	2/7	1/3	1

Нормовані значення ваг w_1, w_2, w_3, w_4 , що отримані модифікованим комбінаторним методом (формула (2)) за даними наведеного прикладу дорівнюють (0,563734299; 0,263382041; 0,120820159; 0,052063501). Значення ваг, отриманих звичайним комбінаторним методом (формула (1)), дорівнюють (0,590174795; 0,243658012; 0,114086692; 0,052080501).

Доведено, що звичайний комбінаторний метод агрегації ваг альтернатив (формула (1)) є еквівалентним, відповідно, геометричному середньому по рядках [21] і ЛМНК [7]. У той же час, як показано у [20], результати роботи методу з урахуванням ваг покривних дерев (формула (2)) та без їхнього урахування суттєво відрізняються, отже ці методи не є еквівалентними.

Як метод геометричного середнього по рядках, так і ЛМНК мають значно нижчу обчислювальну складність, аніж комбінаторний метод з урахуванням ваг покривних дерев, і в цьому їхня безперечна перевага. Наприклад, треба визначити ваги n об'єктів за заданою МПП. Для розв'язання задачі ЛМНК передбачає розв'язання системи лінійних рівнянь з n невідомими (кількість обчислювальних

операцій порядку n^2). Метод геометричного середнього передбачає обчислення ваг за n рядками МПП та подальше усереднення отриманих значень (кількість операцій порядку n). Комбінаторний метод передбачає обчислення ваг за n^{n-2} покривними деревами (кількість операцій порядку n^{n-2}). Якщо ж враховувати ваги, тобто рейтинги, дерев (формули (3) або (4)), то порядок кількості обчислювальних операцій у комбінаторному методі суттєво збільшується.

Основна ж перевага «зваженого» комбінаторного методу в тому, що він дозволяє врахувати якість експертної інформації на етапі її агрегації. Відтак, результати його роботи, загалом, більш адекватно відображають рівень обізнаності експертів у питанні, що розглядається.

Дане дослідження — спроба емпірично підтвердити перевагу зваженого комбінаторного методу над звичайним, а відтак, і над методами ЛМНК і геометричного середнього по рядках.

Вибір підходів до визначення ефективності звичайного та модифікованого методів

З-поміж підходів до визначення ефективності експертних методів (та їхнього порівняння), окреслених на початку даної статті, в даному випадку можна застосувати далеко не всі.

1. Проведення справжніх експертиз у реальних умовах («легітимізація» за Сааті [3]) з метою емпіричної перевірки певних гіпотез — це «розкіш», яку не може собі дозволити пересічний дослідник, адже пошук справжніх експертів та оплата їхнього часу пов'язані з дуже значними витратами.

2. Порівняння результатів кількох методів (підхід Панкратової [8]) на предмет співпадіння не відповідає поставленій задачі: в даному випадку ми намагаємося з'ясувати, який з двох методів дає *кращий* результат.

3. Проведення тестових експертиз (наприклад, зі студентами), на зразок тих, що описані у [9, 10], де експерти самі визначали б, які результати точніше відображають їхні уявлення про предметну область — знов-таки, не відповідає характеру задачі. Хотілося би отримати більш об'єктивні характеристики методів, які би не залежали від ставлення учасників експертиз, і порівняти методи саме за цими характеристиками.

4. Тестування методів на даних реального експертного оцінювання спеціально змодельованих еталонних об'єктів [11, 12] — здійсненне.

5. Моделювання еталонних ваг об'єктів і самих експертних оцінок співвідношень цих ваг [17] — здійсненне.

Отже, пропонується зосередити увагу на двох підходах: а) порівняння двох методів на реальних експертних оцінках тестової (еталонної) множини об'єктів; б) моделювання як еталонних значень об'єктів, так і самих експертних оцінок.

Порівняння звичайного та модифікованого комбінаторних методів на реальних даних оцінювання еталонних об'єктів

Як еталонні об'єкти було використано зображення з різною кількістю зафарбованих пікселів (відомою організатору експертизи). Кількість таких зображень,

які експерти порівнювали за ступенем зафарбування, дорівнювала 7 ($n = 7$) (з урахуванням вищезазначених психофізіологічних обмежень людського сприйняття [13]).

Було проведено 18 незалежних сеансів парних порівнянь об'єктів реальними респондентами (експертами). Парні порівняння — мультиплікативні, тобто, правило транзитивності (внутрішньої узгодженості) у МПП виглядає наступним чином: $a_{ij} = w_i/w_j = (w_i/w_k) \times (w_k/w_j) = a_{ik} \times a_{kj}; i, j, k = 1..n$, де a_{ij} — значення парного порівняння об'єктів з номерами i та j ; w_i, w_j, w_k — ваги об'єктів з відповідними номерами. На основі побудованих експертами МПП було проведено 4 серії розрахунків ваг об'єктів: відповідно, індивідуальним і груповим, звичайним (формула (1)) та модифікованим (формула (2)) методами.

Зазначимо, що усі оцінки вводилися в однаковій шкалі, а експерти апріорно вважалися рівно-компетентними, тож чисельник дроби у формулі мультиплікативного рейтингу (4) було прийнято за 1.

Отримані ваги порівнювалися з еталонними значеннями та обраховувалися відносні похибки оцінювання кожного з 7 об'єктів (7):

$$\delta_k = \frac{|w_k - w_k^{etalonne}|}{w_k^{etalonne}}, k = 1..n. \quad (7)$$

Групові сеанси оцінювання моделювалися шляхом комбінування груп по 3 експерти з 18-ти наявних прецедентів індивідуального оцінювання. Відповідно, таких групових сеансів було реалізовано $C_{18}^3 = 816$.

Як узагальнений показник точності методів було обрано усереднену відносну похибку оцінювання за всіма об'єктами (8):

$$\delta = \frac{1}{n} \sum_{k=1}^n \delta_k = \frac{1}{n} \sum_{k=1}^n \frac{|w_k - w_k^{etalonne}|}{w_k^{etalonne}}. \quad (8)$$

Для проведення експерименту використовувався спеціально створений автотестовим програмний модуль (рис. 2).

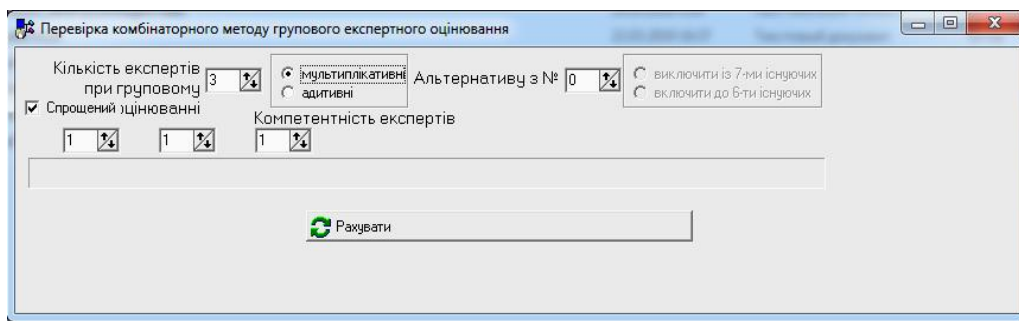


Рис. 2. Екранна форма програмного модуля для експериментального дослідження комбінаторного методу на реальних оцінках

Стислий алгоритм проведення експерименту після запуску програмного модуля — наступний.

1. Обирається тип парних порівнянь (адитивні чи мультиплікативні), кількість експертів у групі, задаються значення відносної компетентності експертів. Визначається, яку модифікацію комбінаторного методу слід запускати — спрощений метод (формула (1)), чи метод зі зважуванням базисних множин парних порівнянь (формула (2)). Передбачається також можливість виключення певної альтернативи з множини наявних 7 еталонних об'єктів.

2. Модуль зчитує з текстового файлу значення парних порівнянь об'єктів і проводить розрахунки — по чергово розраховує значення ваг об'єктів на основі кожного прецеденту оцінювання за допомогою обраної на попередньому кроці модифікації комбінаторного методу. Результати розрахунків (значення відносних ваг 7 еталонних об'єктів) у реальному часі записуються в інший текстовий файл.

3. Програма закінчує роботу, коли перебрані усі можливі варіанти експертних груп. Наприклад, якщо загальна кількість прецедентів індивідуального експертного оцінювання дорівнює 18, а кількість експертів у групі дорівнює 3, то розрахунки проводяться для $C_{18}^3 = 816$ сеансів групового оцінювання (1-й сеанс — 1-й, 2-й, 3-й прецеденти; 2-й сеанс — 1-й, 2-й, 4-й прецеденти; ... 816-й сеанс — 16-й, 17-й, 18-й прецеденти).

Розраховані значення похибок індивідуального оцінювання для модифікованого («зваженого») та звичайного (спрощеного) методів показані на рис. 3. Значення похибок групового оцінювання для цих методів показані на рис. 4. На обох рисунках по осі абсцис відкладаються номери прецедентів експертного оцінювання 7 еталонних об'єктів. По осі ординат відкладаються значення середніх похибок оцінювання у долях одиниці (формула (8)). Кожному номеру сеансу оцінювання відповідають 2 значення середніх відносних похибок (похибка методу зі зважуванням і без зважування покривних дерев) — 2 точки на координатній площині. Якщо для заданого сеансу оцінювання точка, що відповідає одному з двох методів, знаходиться вище (має більшу ординату), це означає, що даний метод дає більшу відносну похибку (гірший результат у порівнянні з еталоном) на відповідному наборі експертних оцінок.

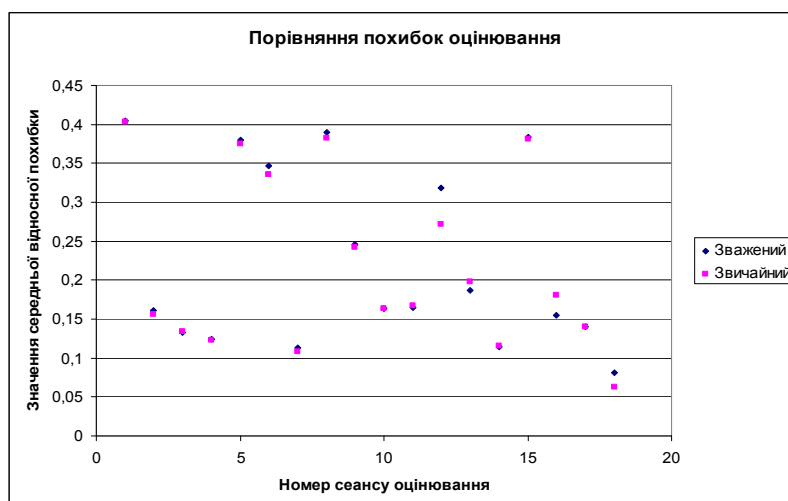


Рис. 3. Похибки оцінювання 7 еталонних об'єктів при агрегації оцінок звичайним і зваженим індивідуальним комбінаторним методом (18 сеансів оцінювання)

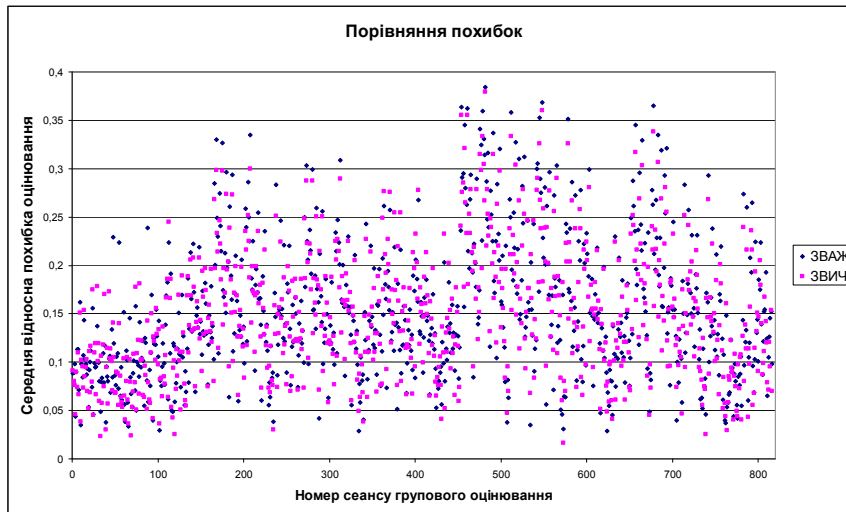


Рис. 4. Похибки оцінювання 7 еталонних об'єктів при агрегації оцінок звичайним і зваженим груповим комбінаторним методом (816 сеансів оцінювання)

Діапазони та середні значення похибок (формула (8)) для 816 прецедентів групового оцінювання наведено у табл. 5.

Таблиця 5. Характеристики роботи модифікованого та звичайного комбінаторного методів на реальних даних групового експертного оцінювання

	Максимальна середня похибка	Мінімальна середня похибка	Середня похибка за всіма прецедентами
Звичайний метод	0,379730445	0,016795423	0,126609639
Модифікований метод	0,384422003	0,028881538	0,130089121

Результати експериментального дослідження звичайного та зваженого комбінаторного методів на реальних даних експертного оцінювання еталонних об'єктів, на жаль, не дозволяють зробити однозначного висновку щодо переваг одного з двох методів за точністю. Як видно з рис. 3 та 4, на одних прецедентах індивідуального та групового оцінювання більш точним виявляється метод, що не враховує ваг базисних множин попарних порівнянь, на інших — метод з урахуванням ваг.

Вигляд відповідних формул агрегації ваг (1) та (2), а також дані 816 прецедентів групового оцінювання вказують на те, що метод з урахуванням рейтингів покривних дерев демонструє вищу ефективність у випадках, коли точніших (ближчих до еталонних) парних порівнянь більше, ніж помилкових. Якщо ж більшість порівнянь — неточні (далекі від еталонних), але узгоджені між собою, то вони, за рахунок зважування, «відтягують» на себе агреговане значення оцінки об'єкта, і, в результаті, більшу відносну точність демонструє звичайний метод. Більше того, обидва методи агрегації можуть створювати казуси, коли усереднення неузгоджених та віддалених від еталонів значень дає точний результат.

Загальний висновок: на результати роботи методу впливає не стільки обрана процедура агрегації оцінок, скільки самі значення цих оцінок. Під час сеансу оцінювання свідомість експерта створює власний еталонний вектор ваг об'єктів. Ре-

зультат і відносна точність роботи методу залежить, у першу чергу, від того, наскільки це суб'єктивне уявлення близьке до реального еталонного вектора. Отже, різниця між вагами, що отримані в результаті агрегації реальних експертних даних, та еталонними вагами не може служити показником ефективності самих методів агрегації. Адже точний результат роботи методу свідчить, у першу чергу, про точність вихідних експертних даних.

Відносну точність реальних експертних оцінок еталонних об'єктів доцільно використовувати у якості критерію ефективності для порівняння концептуально різних методів (наприклад, таких, що використовують адитивну та мультиплікативну шкалу; вербальний, графічний, або ж числовий спосіб введення даних, вектори парних порівнянь, трикутні чи квадратні МПП; методів із зворотним зв'язком та без такого; тощо), як це робиться у [11, 12]. Під час порівняння таких методів використовуються, почасти, різні вихідні дані та процедури їхньої агрегації, що суттєво різняться між собою. Зручність процесу вводу експертних даних до системи підтримки прийняття рішень (СППР) [23] становить окреме питання, яке, знов-таки, не пов'язане з конкретними процедурами агрегації. У випадку ж поточного дослідження у якості вихідних даних використовувалися одні й ті самі множини експертних оцінок, і принципово подібні процедури агрегації. Тому застосування підходу, запропонованого у [11, 12], в контексті поточного дослідження виявляється некоректним, а його результати — не репрезентативними.

Людський фактор (суб'єктивне уявлення експерта про співвідношення об'єктів), додає до умов експерименту зайвий ступінь свободи. Єдиний підхід, який дозволить під час експерименту контролювати віддаленість експертних оцінок від еталонних значень (а відтак, нівелювати вплив людського фактору) — моделювання самих експертних оцінок. Тому, доцільно звернутися до способу порівняння методів за ефективністю, запропонованому в [17].

Порівняння звичайного та модифікованого комбінаторного методів шляхом імітаційного моделювання процесу експертного оцінювання еталонних об'єктів

Основна ідея визначення ефективності методів агрегації полягає у перевірці їхньої стійкості до збурень вихідних даних (МПП). Як уже сказано у вступі, передбачається, що існує певне істинне значення оцінки об'єкта за заданим критерієм (таке як, наприклад, точна кількість зафарбованих чи не зафарбованих точок на певному зображенні). Оцінка, яку дає експерт, відрізняється від цього істинного значення на величину помилки оцінювання. Нехай, за одних і тих самих значень помилок експерта один метод агрегації дає результат (вектор ваг об'єктів), що є більш близьким до еталонного вектора, аніж інший метод. Тоді можна буде стверджувати, що перший метод є більш ефективним.

Для того, щоб мати можливість відслідковувати величину помилок експертів та відхилень отриманих різними методами векторів ваг об'єктів від еталонних значень (у контексті порівняння методів агрегації), пропонується моделювати процес експертного оцінювання наступним чином.

1. Задаються еталонні ваги об'єктів.

2. Будується повністю (ідеально) узгоджена МПП (consistency matrix) A (на основі співвідношення $a_{ij} = w_i/w_j$ для мультиплікативних або $a_{ij} = w_i - w_j$ для адитивних порівнянь, де a_{ij} – елемент матриці A).

3. Здійснюється «зашумлення» матриці A таким чином, що кожен елемент матриці A , окрім діагональних, може бути змінений за наступною формулою $a'_{ij} = a_{ij} \pm a_{ij} \cdot \delta / 100\%$, де $\delta > 0$ – наперед задана величина, що характеризує максимальне відносне відхилення результатів парних порівнянь експерта (елементів матриці A) у відсотках від еталонних значень. Таким чином моделюються (імітуються) помилки експертного оцінювання. Величина δ , у цьому випадку, являє собою можливу відносну помилку експерта при проведенні парних порівнянь.

4. «Зашумлена» МПП (позначимо її A') подається на вхід одного з методів агрегації з метою одержання узагальнених (агрегованих) ваг альтернатив w'_i у вигляді вектора ваг (пріоритетів за Сааті). Ефективність методу агрегації попарних порівнянь пропонується визначати як максимально можливе відносне відхилення отриманої в результаті розрахунку ваги альтернативи від еталонної ваги цієї ж альтернативи (чим менше відхилення, тим краще метод):

$$\Delta = \max_i \left| \frac{w'_i - w_i}{w_i} \right| \cdot 100 \% . \quad (9)$$

Слід зазначити, що обчислені значення показника Δ залежатимуть як від заданої δ , так і від відносних значень самих заданих еталонів. Тому, визначені величини показника ефективності кожного методу подаються у вигляді графіків залежності $\Delta(\delta)$ для кожного з варіантів еталонного вектора ваг.

Залежність $\Delta(\delta)$ визначається для кожного з досліджуваних методів на інтервалі $\delta \in]0;100[$, виходячи з припущення, що відносна помилка оцінювання при парних порівняннях експерта не перевищує 100 %. Хоча, у загальному випадку функція $\Delta(\delta)$ визначена на більш широкому діапазоні — $\delta \in]0;\infty[$. Знаходження максимально можливого відносного відхилення Δ для кожного δ пропонується проводити із застосуванням генетичного алгоритму (ГА) [24].

Як і в попередньому випадку, для проведення експерименту використовувався спеціальний програмний модуль, у якому реалізовані генерація та зашумлення МПП та, власне, ГА (рис. 5).

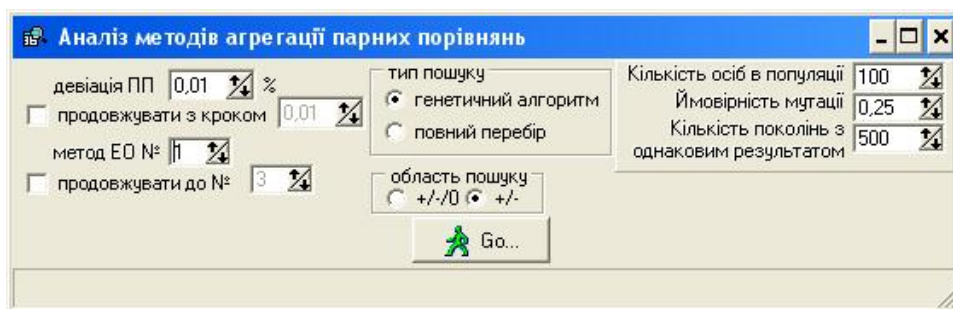


Рис. 5. Екранна форма програмного модуля для експериментального дослідження комбінаторного методу за допомогою ГА

У термінах ГА «особинами» є зашумлені МПП із заданим значенням відносної похибки оцінювання δ . У ролі «функції придатності» (fitness function) виступає максимальне відносне відхилення від еталона значень ваг об'єктів Δ (формула (9)). ГА працює наступним чином.

1. Для заданих еталонних значень ваг об'єктів і значення δ генерується «популяція» особин (зашумлених МПП). (Потужність повного перебору матриць розмірністю $n \times n$, елементи яких відрізняються від еталонних значень на задане δ — дуже велика, тож популяція являє собою лише певну частину МПП з цієї множини. Втім, для $n \leq 5$ передбачається можливість повного перебору).

2. З популяції вибираються особини, для яких значення функції придатності є максимальними. Тобто вибираються МПП, які після агрегації дають найбільше відхилення вектора ваг об'єктів від еталонних значень.

3. Особини з вибраної підмножини «схрещуються» між собою шляхом зваженого додавання (опуклої комбінації) і мутації. Таким чином, отримується нове покоління особин.

4. Якщо для нового покоління значення функції придатності Δ збільшується, слід переходити на крок 2. Якщо ж протягом заданої фіксованої кількості поколінь Δ не збільшується, то алгоритм зупиняється і завершує роботу. В результаті отримуємо максимальне значення Δ для заданого δ .

По суті, йдеться про знаходження максимуму функції багатьох змінних $f(a'_{ij})$, $i, j = (1, n)$, аргументами якої є елементи МПП A' . Значення $\Delta(\delta)$ для кожного методу агрегації залежить також від конкретних значень вхідних еталонних ваг альтернатив $w_i, i = (1, n)$. Приклади графіків залежностей $\Delta(\delta)$ для заданих еталонних значень ваг 5-ти альтернатив наведено на рис. 6. Варіанти еталонних векторів ваг задаються таким чином, щоб проілюструвати різні співвідношення між вагами об'єктів (однакові, попарно однакові, арифметична прогресія, геометрична прогресія, крайні значення поділок шкали тощо — див. рис. 6).

У праці [17] індивідуальний зважений комбінаторний метод порівнювався з кількома іншими методами агрегації індивідуальних парних порівнянь (рис. 6). У контексті ж поточного дослідження важливо відзначити, що середнє геометричне по рядках (яке входить до множини методів, досліджуваних у [17]) — еквівалентне звичайному комбінаторному методу (як доведено у [21]). Отже, з переваги зваженого комбінаторного методу над середнім геометричним по рядках [17, 20] впливає його перевага над звичайним комбінаторним методом і ЛМНК (який також еквівалентний йому — див. [7]).

Відповідно, можна стверджувати, що емпіричні дані порівняння різних модифікацій методів агрегації попарних порівнянь (саме шляхом моделювання процесу експертного оцінювання заданих еталонних ваг об'єктів), вказують на перевагу модифікованого комбінаторного методу над іншими. Критерій ефективності методу — його стійкість до збурень МПП (тобто помилок експерта).

Втім, слід зауважити, що експериментальні результати, які ілюструє рис. 6, отримані лише для індивідуальних методів оцінювання альтернатив ($m = 1$), переважно, в адитивній шкалі. Тобто, у проведеному експерименті у формулах (1) та (2) використовується сума, а не добуток. Відповідно:

$$w_j^{aggregate} = \frac{1}{T} \sum_{q=1}^T w_j^q, j = 1..n, T \in [1..mn^{n-2}], \quad (10)$$

$$w_j^{aggregate} = \frac{1}{\sum_{u,p,v} R_{upv}} \sum_{k,l=1}^m \left(\sum_{q^k=1}^{T_k} R_{kq_k l} w_j^{(kq_k l)} \right), j = 1..n. \quad (11)$$

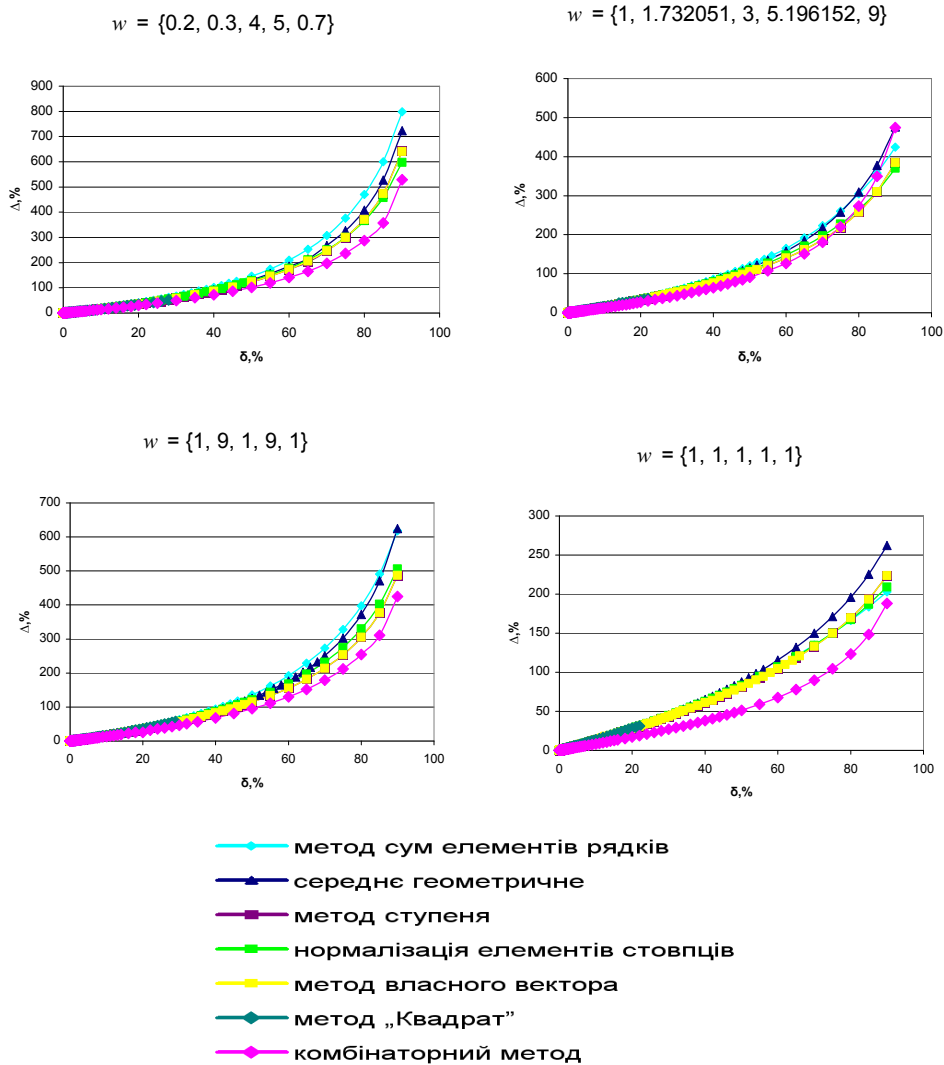


Рис. 6. Приклади залежностей Δ від δ для різних ілюстративних наборів еталонних значень ваг 5-ти альтернатив

Отже, наступним кроком дослідження буде моделювання процесу агрегації групових експертних оцінок об'єктів, що задані як в адитивній, так і в мультиплікативній шкалах. Перехід від мультиплікативних переваг до адитивних і навпаки (зокрема, під час моделювання) можна здійснювати, застосовуючи операції логарифмування та піднесення до степеня. Обидві функції монотонно зростають, тож

є підстави сподіватися, що властивості адитивного та мультиплікативного методів будуть співпадати.

Висновки

У ході дослідження розглянуто два можливих способи оцінювання ефективності методів агрегації парних порівнянь, а саме, комбінаторних методів агрегації попарних порівнянь об'єктів з урахуванням і без урахування відносної вагомості покривних дерев. Показано, що традиційне поняття точності проблематично застосовувати до експертних методів. У результаті, імітаційне моделювання (simulation) виявляється більш коректним способом оцінки точності методів, аніж їхнє тестування на реальних експертних даних. Запропоновано два підходи до оцінки ефективності модифікацій комбінаторного методу агрегації попарних порівнянь (індивідуального та групового, адитивного та мультиплікативного), а саме: визначення відносної точності результатів агрегації реальних попарних експертних порівнянь еталонних об'єктів та імітаційне моделювання усього циклу експертизи. Отримано експериментальні результати, які емпірично підтверджують перевагу модифікованого комбінаторного методу над звичайним (і, відповідно, над методами геометричного середнього по рядках МПП та ЛМНК).

Отримані експериментальні результати дозволяють зробити кілька принципових висновків.

1) слід чітко розрізняти поняття точності експертних оцінок та ефективності методів їхньої агрегації; часто низька точність результатів роботи методу визначається помилками експертів, а не вадами самого методу;

2) важлива не просто точність, а саме «узгоджена точність», як в рамках множини оцінок (МПП) 1-го експерта, так і групи експертів; тобто більш достовірними слід вважати більш узгоджені результати;

3) на реальних експертних оцінках доцільно порівнювати принципово різні методи, в яких використовується експертна інформація різних типів; якщо ж вихідна інформація для кількох методів — однакова, а відрізняються тільки процедури (алгоритми) її агрегації, то для оцінки методів необхідно моделювати весь цикл експертизи, в тому числі й самі оцінки.

Подальші дослідження за даною тематикою будуть спрямовані на моделювання процесу агрегації групових експертних оцінок об'єктів, що задані як в адитивній, так і в мультиплікативній шкалах, з метою дослідження ефективності комбінаторного методу агрегації з урахуванням ваг покривних дерев.

1. Таран Т.А., Зубов Д.А. Искусственный интеллект. Теория и приложения. Луганск: ВНУ им. В. Даля, 2006. 239 с.

2. Saaty T.L. The Analytic Hierarchy Process. McGraw-Hill, New York. 1980. 287 p.

3. Saaty T.L. Decision Making with Dependence and Feedback: The analytic Network Process. Pittsburgh. RWS Publications, 1996. 370 p.

4. Hwang C.L., Yoon K. Multiple attribute decision making: methods and applications: a state-of-the-art survey. Berlin; New York: Springer-Verlag, 1981. 259 p.

5. Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект. ИПРИ НАНУ. Київ: Наук. думка, 2002. 382 с.

6. Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом. *Ресстрація, зберігання і оброб. даних*. 2000. Т. 2. № 2. С. 92–102.

7. Bozoki S., Tsyganok V. The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. *International Journal of General Systems*. 2019. Vol. 48. Issue 4. P. 362–381.
8. Панкратова Н.Д., Недашковская Н.И. Гибридный метод многокритериального оценивания альтернатив принятия решений. *Кибернетика и системный анализ*. 2014. Т. 50. № 5. С. 58–70.
9. Elliott M.A. Selecting numerical scales for pair-wise comparisons. *Reliability Engineering and System Safety*. 2010. Vol. 95. P. 750–763.
10. Циганок В.В., Качанов П.Т., Каденко С.В., Андрійчук О.В., Гоменюк Г.А. Експериментальний аналіз технології експертного оцінювання. *Реєстрація, зберігання і оброб. даних*. 2012. Т. 14. № 1. С. 91–100.
11. Тоценко В.Г., Цыганок В.В., Качанов П.Т., Деев А.А., Качанова Е.В., Торба Л.Т. Экспериментальное исследование методов получения кардинальных экспертных оценок альтернатив. Ч. 1. Методы без обратной связи с экспертом. *Проблемы управления и информатики*. 2003. № 1. С. 34–48.
12. Тоценко В.Г., Цыганок В.В., Качанов П.Т., Деев А.А., Качанова Е.В., Торба Л.Т. Экспериментальное исследование методов получения кардинальных экспертных оценок альтернатив. Ч. 2. Методы с обратной связью с экспертом. *Проблемы управления и информатики*. 2003. № 2. С. 112–125.
13. Miller G.A. The Magical Number Seven, Plus or Minus Two. *The Psychological Review*. 1956. Vol. 63. P. 81–97.
14. Tsyganok V., Kadenko S., Andriichuk O. Significance of expert competence consideration in group decision making using AHP. *International Journal of Production Research*. 2012. Vol. 50. Issue 17. P. 4785–4792.
15. Cayley A.A Theorem on Trees. *Quarterly Journal of Mathematics*. 1889. Vol. 23. P. 376–378.
16. Циганок В.В., Каденко С.В., Андрійчук О.В., Качанов П.Т., Роїк П.Д. Інструментарій підтримки прийняття рішень як засіб стратегічного планування. *Озброєння та військова техніка*. 2015. № 3(7). С. 59–66.
17. Tsyganok V. Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. *Mathematical and Computer Modeling*. 2010. 52(3–4). P. 538–544.
18. Siraj S., Mikhailov L., Keane J.A. Enumerating all spanning trees for pairwise comparisons. *Computers & Operations Research*. 2012. 39(2). P. 191–199.
19. Siraj S., Mikhailov L., Keane J.A. Corrigendum to «Enumerating all spanning trees for pairwise comparisons [Comput. Oper. Res. 39 (2012) 191–199]». *Computers & Operations Research*. 2012. 39(9). P. 2265.
20. Каденко С.В., Циганок В.В. Визначення відносної компетентності експертів під час агрегації парних порівнянь. *Реєстрація, зберігання і оброб. даних*. 2017. Т. 19. № 2. С. 69–83.
21. Lundy M., Siraj S., and Greco S. The Mathematical Equivalence of the «Spanning Tree» and Row Geometric Mean Preference Vectors and its Implications for Preference Analysis. *European Journal of Operational Research*. 2017. 257(1). P. 197–208.
22. Hartley R.V.L. Transmission of information. *Bell System Technical Journal*. 1928. Vol. 7. P. 535–563.
23. Каденко С.В. Проблеми представлення експертних даних у системах підтримки прийняття рішень. *Реєстрація, зберігання і оброб. даних*. 2016. Т. 18. № 3. С. 67–74.
24. Holland J.H. *Adaptation in Natural and Artificial Systems*. University of Michigan Press. Ann Arbor, 1975. 183 p.

Надійшла до редакції 02.05.2019