



UDC 004.51

ЕКСПЕРТНА ОЦІНКА РЕВАЛЕНТНОСТІ ВІДПОВІДЕЙ АВТОМАТИЗОВАНОЇ СИСТЕМИ ПІДТРИМКИ ДІАЛОГУ ДЛЯ ДИСТАНЦІЙНОГО НАВЧАННЯ

Expert assessment of a revalent of replies of the automated system of support of dialogue for distance learning.

Лісовенко А.І.¹, Бісікало О.В.²

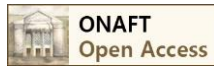
^{1,2} Vinnytsia National Technical University, Vinnytsia

E-mail: sweettee@ukr.net

Copyright © 2014 by author and the journal “Automation technological and business - processes”.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Анотація

В статті представлено опис та аналіз існуючих методів експертної оцінки ревалентності відповідей автоматизованої діалогової системи, наведено переваги та недоліки кожного з них, а також обґрунтовано метод, що найбільше підходить для оцінки ревалентності відповідей нової діалогової системи для дистанційного навчання.

Ключові слова

Автоматизована діалогова система, експертна оцінка, навчальний контент, ревалентність відповіді.

Експертною оцінкою вважають процедуру отримання оцінки певної задачі (проблеми) на основі думки спеціалістів (експертів) з метою подальшого прийняття рішення [1]. Даний метод використовується у випадках значної складності задачі (проблеми), її новизни, недостатньої кількості наявної інформації про неї, неможливості математичної формалізації процесу вирішення. У таких випадках постає необхідність звернутись до рекомендацій компетентних спеціалістів, що добре знають проблему – сукупність процесів вирішення ними задачі, аргументації, формування кількісних оцінок, оброблення цих оцінок формальними методами отримала назву методу експертних оцінок.

Існує дві групи експертних оцінок:

- індивідуальні оцінки – основані на використанні незалежних одна від одної думок окремих експертів;
- колективні оцінки – ґрунтуються на використанні колективної думки експертів.

Розрізняють такі способи вимірювання об'єктів:

- ранжування;
- попарне порівняння;
- безпосередня оцінка.

Розглянемо ці методи більш детально. Метод ранжування – це процедура впорядкування об'єктів, яку виконує експерт. На основі знань та досвіду, експерт розташовує об'єкти в порядку надання їм переваги, керуючись одним чи кількома обраними показниками порівняння [2]. Існують різні варіанти впорядкування, в залежності від типу відношень між об'єктами. Нехай серед об'єктів порівняння немає однакових за показниками порівняння, тобто, еквівалентних. У такому випадку, між об'єктами є лише відношення строгого порядку. Внаслідок порівняння, складається впорядкована строга послідовність $a_1 \succ a_2 \succ \dots \succ a_n$, де об'єкт a_1 має найбільшу перевагу з усіх об'єктів, другий рівень переваги за об'єктом a_2 тощо. Тому система об'єктів із відношенням строгого порядку за умовою порівняння всіх об'єктів за даним відношенням утворює повний строгий порядок. Для даного типу відношень властиво застосування дійсних чисел із застосуванням відношень типу «>» чи «<».

Відповідність перерахованих властивостей, тобто гомоморфізм можна здійснити обираючи будь-які числові послідовності. Головною особливістю є монотонність переходу числових представлень утвореного ряду.

Суттєвим недоліком методу ранжування є практична неможливість його застосування до великої кількості об'єктів. Досвід показує, що в процесі застосування методу до великої кількості об'єктів (15-20) експертам важко



визначитись у ранжуванні. Збільшення кількості об'єктів призводить до збільшення у квадраті кількості зв'язків між ними. Тому при великій кількості об'єктів експерти можуть допускати велику кількість помилок [3].

Метод попарних порівнянь – це встановлення переваги об'єкту при порівнянні всіх можливих пар. На відміну від метою ранжування, метод попарного порівняння являє собою простішу задачу: при порівнянні пари об'єктів можливе відношення строгого порядку або відношення еквівалентності, тобто не потрібно впорядковувати всі об'єкти. Звідси випливає, що попарне порівняння, як і ранжування – вимірювання в порядковій шкалі [4].

Сутність методу така: припустимо, нас цікавить, як експерти оцінять певні об'єкти, які ми позначимо a_1, a_2, \dots, a_n , де n – кількість об'єктів, що підлягають ранжуванню. Метод дозволяє отримати ефективну процедуру оцінки, оскільки кожному експерту пропонуються пари об'єктів, що розглядаються. Він повинен визначити відносно кожної пари, який об'єкт із пари йому подобається більше. Усі відповіді фіксуються і заносяться в квадратну матрицю, розмірністю $n \times n$, яка буде складатися із 0 та 1 в залежності від вподобань експерта. Таку матрицю називають матрицею попарних порівнянь.

Надалі, замість виразу «об'єкт a_i кращий, ніж a_j » будемо використовувати « $a_i > a_j$ ». В загальному вигляді матрицю для експерта r_1 , що входить до групи з N експертів позначимо через $\|\delta^1_{ij}\|$, де

$$\delta^1_{ij} = \begin{cases} 1, & \text{якщо респондент } r_1 \text{ сказав, що } a_i > a_j \\ 0, & \text{якщо респондент } r \text{ сказав, що } a_i < a_j \end{cases} \quad (1)$$

Головною характеристикою матриці попарних порівнянь є асиметричність та транзитивність. До того ж, варто зазначити, що слід враховувати особисте сприйняття експерта, тобто він не завжди може визначити однозначно – який об'єкт йому подобається більше. Тоді виникають різні випадки, наприклад, експерт може вважати об'єкти неможливими для порівняння, а може припускати, що вони рівні. Отже, двох чисел (міток) вже не достатньо для оцінювання, а доцільно обрати чотири міри оцінювання: «більше», «менше», «рівно», «не можливо порівняти». Але, в різних ситуаціях можливі різні варіанти впевненості експерта, так, наприклад, він може давати різну ступінь оцінки. Тому доволі простий метод ускладнюється в залежності від ситуації [5].

Метод безпосередньої оцінки – у багатьох випадках бажано не лише впорядкувати (ранжувати) об'єкти аналізу, але і визначити, на скільки один фактор більш значимий за інші. У такій ситуації діапазон зміни характеристик об'єкту розбивається на окремі інтервали, кожному з яких присвоюється власна оцінка (бал), наприклад, від 0 до 10. Саме тому метод безпосередньої оцінки іноді також називають методом балів [6].

В даному методі експерту необхідно поставити, у відповідності до кожного об'єкта, точку на певному відрізку числової осі, а також, вимагати, щоб еквівалентним об'єктам надавалися однакові числа. Результати зручно представити графічно [7].

Експерт приписує кожному об'єкту певну оцінку в балах із інтервалу, наприклад від 0 до 10, де 0 балів – відсутність значення параметру. Верхня межа – 10 балів – найвище можливе значення параметру. Використовуючи подібну шкалу, експерт має надати кожному параметру певне числове значення P_{ij} у межах бальної шкали, що застосовується. Зрештою обчислюється середнє за всіма експертами значення P_i .

Для аналізу результатів використовують різноманітні методи математичної статистики, які можуть варіюватися та комбінуватися в залежності від типу задачі та необхідного результату.

Формування загальної оцінки. Отже, нехай група експертів оцінила певний об'єкт, тоді x_j – оцінка j -го експерта, де m – кількість експертів. Для формування загальної оцінки групи експертів найчастіше використовується така оцінка, по відношенню до якої кількість більших оцінок прирівнюється кількості менших.

Визначення відносних вагів об'єктів. Іноді необхідно визначити, на скільки той чи інший фактор (об'єкт) важливий з точки зору якого-небудь критерію. В такому випадку говорять, що потрібно визначити вагу кожного фактора. Даний метод відрізняється від методу узагальненої оцінки тим, що визначається не загальна оцінка об'єкту, а оцінка для кожної її ознаки.

Існує велика кількість можливих методів обробки оцінок. Загалом результат може складатися із декількох алгоритмів, що перетинаються з іншими. Наприклад, алгоритм розрахунку коефіцієнту компетентності експерта може вплинути на середньостатистичну оцінку даного експерта [6].

Встановлення степені узгодженості думки експертів. У випадку участі декількох експертів розбіжності в їх оцінках неминучі, однак, величина цього розходження має важливе значення. Групова оцінка може вважатися достатньо надійною лише за умови високого рівня узгодженості відповідей окремих спеціалістів.

Для аналізу розкиду та узгодженості оцінок застосовують статистичні характеристики – міри розкиду чи статистична варіація.

Для визначення даної міри використовують ряд характеристик, що надають інформацію про розкид, а саме, варіаційний розмах (2), середнє лінійне відхилення (3), середньоквадратичне відхилення (4), дисперсія (5) та коефіцієнт рангової кореляції Спірмена (6).

$$R = x_{\max} - x_{\min} \quad (2)$$

$$a = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (3)$$



$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5)$$

$$\rho = 1 - \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n^2 - 1)} = 1 - \frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (6)$$

Коефіцієнт (величина ρ) може змінюватися в діапазоні від -1 до $+1$. При повній узгодженості думок експертів, коефіцієнт конкордації (множинної рангової кореляції) рівний одиниці, при повній розбіжності – нулю. Найреальнішим є випадок часткової узгодженості думок експертів.

Говорячи про узгодженість думок експертів, варто згадати, що ранжування не має на увазі відстань. Тобто, у одного експерта $A > B > C$ означає, що $A >> B > C$, а у другого $A > B >> C$. І будь-які кореляції та розрахунки середніх оцінок тут не допоможуть. Як варіант, потрібно рахувати індекс узгодженості. Рейтинги зазвичай базуються на деякій ймовірнісній моделі, тому важливо ретельно враховувати область їх можливого застосування.

Для визначення адекватності запропонованих у [8] методу та автоматизованої системи підтримки діалогу для дистанційного навчання потрібно обрати метод оцінки відповідей. Особливістю даної системи є те, що відповіді на питання користувачів надаються у вигляді обраних фрагментів (слів, словосполучень) з кількох речень закладеного в систему навчального текстового контенту. Проаналізувавши існуючі методи експертної оцінки релевантності відповіді, обираємо бальний метод встановлення пріоритетів. Пропонується побудувати таблицю, в якій експерти виставляють бали згідно такої шкали: 1 бал – відповідь не є коректною, 2 бали – відповідь майже задовольняє запитання та 3 бали – відповідь на запитання вірна. В таблиці 1.1 наведено приклади запитань та варіанти відповідей на них [8].

Табл. 1.1 - Запитання та варіанти відповідей для експертів

Запитання	Варіант відповіді а	Варіант відповіді б
Що означає онтологічний ресурс?	Кожна онтологічна мережа означає мережевий ресурс RDF	Онтологічна мережа означає онтологічний інтернет-ресурс, онтологічні дані означають, онтологічна мова означає, онтологія може бути корисним означенням, довільним онтологічним означенням, деяким онтологічним означенням окремих ресурсів
Які мови дозволяють мережеві ресурси?	Кожна мережа дозволяє онтологічний мовний ресурс RDF та XML з дозволом людини	Дані інтернет мережі дозволяють бути корисними деяким мовам мережі, що дозволяють комп'ютеру три ключі дозволу OWL, виконану RDF мову і виконаний RDF ресурс
Як будується онтологічний ресурс?	Дані інтернет мережі є деякою довільною мовою мережі, де створена концепція павутини ресурсів впроваджує класифікацію людського каталогу ресурсів, онтологічна мережа є зрозумілою корисною інтернет мовою деякої довільної онтології	Кожна онтологія будується на RDF онтології
Які мови обробляють мережеві ресурси?	Дозвіл на обробку кожної онтології людської обробки є XML обробка та RDF обробка	Дані інтернет мережі комп'ютерної обробки мовної інформації є мережею доповнення представлення мовної розмітки, основної та другої обробки, призначення такої онтологічної обробки три ключі обробки OWL, деяка довільна корисна мережа автоматичної контентної мови

В таблиці 1.2 представлено отримані у групи з 4-х експертів оцінки релевантності відповідей системи на поставлені запитання.



Табл. 1.2 – Експертні оцінки релевантності відповідей

	Експерт 1		Експерт 2		Експерт 3		Експерт 4		Σа, %	Σб, %	Варіант відповіді
	а	б	а	б	а	б	а	б			
Запитання 1	3	1	2	2	1	2	3	2	56,25	43,75	а
Запитання 2	1	3	3	1	2	1	3	1	60	40	а
Запитання 3	1	3	1	3	3	2	1	3	35,3	64,7	б
Запитання 4	3	1	3	2	2	1	3	2	64,7	35,3	а

Аналіз результатів експертних оцінок релевантності відповідей системи демонструє досить високий рівень її адекватності. Запропонований метод бальної оцінки забезпечує отримання чисельних результатів для кожного варіанту відповідей за даними групи експертів, що дозволить застосувати методи параметричної ідентифікації для покращення роботи системи підтримки діалогу.

Література

- [1] Peter D. Grogono. A Review of Expert Systems Evaluation Techniques / Peter D. Grogono, Alun D. Preece, Rajjan Shinghal, and Ching Y. Suen // American Association for Artificial Intelligence in TECHNICAL REPORT- AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE WS. 1994. – pp.113-118.
- [2] Fulop J. Introduction to Decision Making Methods [Електронний ресурс] / JaponFulop // Working Paper 05-6, Laboratory of Operations Research and Decision Systems, Computer and Automation Institute, Hungarian Academy of Sciences, Budapest, November 2005. – 15 p. – Режим доступу: <http://academic.evergreen.edu/projects/bdei/documents/decisionmakingmethods.pdf>
- [3] Wilcoxon F. Individual Comparisons by Ranking Methods / Frank Wilcoxon // Biometrics Bulletin, Vol. 1, No. 6. – 1945. – pp. 80-83.
- [4] Pankratova N. Methods of evaluation and improvement of consistency of expert pairwise comparison judgments / N. Pankratova, N. Nedashkovskaya // International Journal “Information Theories and Applications”. – Vol. 22, Number 3, 2015– pp. 203-223.
- [5] Метод попарних порівнянь (ПС). Лекція №6. [Електронний ресурс] – Режим доступу: <http://ecsocman.hse.ru/data/466/641/1219/ch6.pdf>.
- [6] Методи експертних оцінок [Електронний ресурс] – Режим доступу: <http://habrahabr.ru/post/189626/>.
- [7] Allen M.J. Direct Method to Assess Student Learning / Mary J. Allen // WASC/AAHE Collaborative Workshop on Building Learner-Centered Institutions-Developing. – Institutional Strategies for Accessing and Improving Student Learning. – 2004.– 8 p.
- [8] Лісовенко А.І. Підтримка діалогу з навчальним контентом [Електронний ресурс] / А.І. Лісовенко, О.В. Бісікало, О.В. Яхимович, С.С. Траченко // Адаптивні технології управління навчанням: матеріали першої міжнародної конференції. Одеса, 23-25 вересня 2015 р. – Одеса, 2015. – С. 97-100. – Режим доступу: <https://drive.google.com/file/d/0B34KZFqaGoAyOVRGTlc2c1J0eUk/view>.

Reference

- [1] Peter D. Grogono. A Review of Expert Systems Evaluation Techniques / Peter D. Grogono, Alun D. Preece, Rajjan Shinghal, and Ching Y. Suen // American Association for Artificial Intelligence in TECHNICAL REPORT- AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE WS. 1994. – pp.113-118.
- [2] Fulop J. Introduction to Decision Making Methods [Electronic resource] / JaponFulop // Working Paper 05-6, Laboratory of Operations Research and Decision Systems, Computer and Automation Institute, Hungarian Academy of Sciences, Budapest, November 2005. – 15 p. – Access mode: <http://academic.evergreen.edu/projects/bdei/documents/decisionmakingmethods.pdf>
- [3] Wilcoxon F. Individual Comparisons by Ranking Methods / Frank Wilcoxon // Biometrics Bulletin, Vol. 1, No. 6. – 1945. – pp. 80-83.
- [4] Pankratova N. Methods of evaluation and improvement of consistency of expert pairwise comparison judgments / N. Pankratova, N. Nedashkovskaya // International Journal “Information Theories and Applications”. – Vol. 22, Number 3, 2015– pp. 203-223.
- [5] The method of pairwise comparisons (PC). Lecture №6. [Electronic resource] – Access mode: <http://ecsocman.hse.ru/data/466/641/1219/ch6.pdf>.
- [6] Methods of expert estimates [Electronic resource] – Access mode: <http://habrahabr.ru/post/189626/>.
- [7] Allen M.J. Direct Method to Assess Student Learning / Mary J. Allen // WASC/AAHE Collaborative Workshop on Building Learner-Centered Institutions-Developing. – Institutional Strategies for Accessing and Improving Student Learning. – 2004.– 8 p.
- [8] Lisovenko A.I. Support of Dialogue with Educational Content [Electronic resource] / A.I. Lisovenko, O.V. Bisikalo, O.V. Yakhimovich, S.S. Trachenko // Adaptive Technologies of Management of Training: Materials of the First International Conference. Odessa, 23-25 september 2015. – Odessa, 2015. – Pp. 97-100. – Access mode: <https://drive.google.com/file/d/0B34KZFqaGoAyOVRGTlc2c1J0eUk/view>.