ARTICLE

DOI 10.36074/grail-of-science.07.05.2021.051

# ЕКСПЕРТНО-СТАТИСТИЧНИЙ МЕТОД ПОБУДОВИ ІНТЕГРАЛЬНИХ ІНДИКАТОРІВ СОЦІАЛЬНИХ І ТЕХНІКО-ЕКОНОМІЧНИХ СИСТЕМ

Грінберг Галина Леонідівна
кандидат технічних наук, доцент, доцент кафедри
економічної кібернетики і маркетингового менеджменту
*Національний технічний університет*
*«Харківський політехнічний інститут», Україна*

Водовозова Єлизавета Євгеніївна
здобувач вищої освіти
факультету комп'ютерних наук і програмної інженерії
*Національний технічний університет*
*«Харківський політехнічний інститут», Україна*



***Анотація.*** *Розглянуто задачу побудови інтегрального індикатору на основі агрегування набору часткових показників. В рамках концепції машинного навчання передбачається, що набір вихідних даних складається з набору статистичних даних вимірювань часткових показників та експертних оцінок відповідних значень інтегрального індикатору. Для оцінки параметрів лінійної моделі інтегрального індикатору використовується метод оптимального узгодження експертних і статистичних оцінок.*

***Ключові слова.*** *Інтегральний індикатор, координація оцінок, регуляризовані оцінки, часткові показники, експертні оцінки.*

Проблема побудови узагальнених інтегральних показників (індикаторів) якості на основі агрегування окремих часткових показників ефективності є однією з центральних в системному аналізі складних соціальних та техніко-економічних систем [1, 2]. Подібні завдання виникають при вибрі ефективних варіантів, аналізу ефективності інвестицій, у маркетингових дослідженнях тощо. Під інтегральним індикатором зазвичай розуміють скалярну функцію, яка в певному сенсі визначає узагальнену характеристику якості об'єктів і залежить від набору часткових показників, що характеризують їхні індивідуальні особливості чи властивості [3]. Різні методи агрегування набору часткових показників в одному узагальненому розроблялись в рамках загальної теорії багатокритеріального прийняття рішень [4]. Загальновідомо, що для цього необхідно залучати додаткову інформацію у вигляді експертних оцінок відносної важливості певних окремих показників та (або) оцінки інтегрального показника в цілому. Ці експертні оцінки зазвичай формуються з використанням вихідних даних у вигляді набору статистичних даних вимірювань часткових показників для даного набору об'єктів. Тоді проблема побудови інтегрального показника зводиться до відновлення деякої невідомої скалярної функції, яка відображає переваги експертів і значення якої експертно оцінюються для кожного вимірюваного набору часткових показників. У цьому формулюванні задачу побудови інтегрованих індикаторів можна розглядати в рамках загальної концепції машинного навчання, точніше, навчання перевагам (preference learning), оскільки модель інтегральних показників повинна відображати уподобання експертів [5].

Вибір методу вирішення цієї задачі значною мірою визначається наявними вихідними даними, що трактуються в даному випадку як навчальний набір даних (training dataset). Якщо доступні лише результати вимірювань окремих часткових показників ефективності, то вочевидь використовується концепція неконтрольованого навчання, яка для проблеми побудови інтегрального показника зазвичай реалізується за допомогою аналізу головних компонент [6]. Якщо є додаткова інформація, така, як експертні оцінки інтегрального показника, та оцінки відносної значущості часткових показників, проблема може трактуватися як задача відновлення регресії. Тоді, як пропонується в [7], експертні оцінки ваг відносної значущості певних показників можуть бути використані у якості апріорної інформації для регуляризації задачі, і такий підхід можна розглядати як розвиток ідеї узгодження (координації) експертних оцінок значень інтегрального показника та відносних ваг часткових показників [8, 9].

В якості моделі інтегрального індикатору найчастіше використовується проста адитивна модель у вигляді зваженої суми часткових показників, тобто модель, лінійна за параметрами; тоді оцінка параметрів моделі інтегрального показника здійснюється за схемою лінійної регресії.

Розглянемо задачу оцінки параметрів лінійної моделі інтегрального показника з оптимальною координацією експертних оцінок інтегральних показників, а також ваг відносної значущості часткових показників.

Нехай задана множина об’єктів $V = \{v_1, v_2, \dots, v_n\}$ та множина часткових показників $B = \{b_1, b_2, \dots, b_m\}$. Довільний об’єкт $v_i$ можна описати за допомогою вектору-строки значень часткових показників $x_i = (x_{i1}, x_{i2} \dots, x_{im})$. Інтегральний індикатор об’єкта $v_i \in V$ із номером *i* - це скаляр $q_i$, поставлений у відповідність набору $x_i$ значень часткових показників об’єкта.

В якості вихідних даних в задачі побудови моделі інтегрального показника будемо використовувати:

– статистичні дані у вигляді матриці $X = \{x_{ij}\}_{i,j=1}^{n,m}$ вимірювань часткових показників, де елемент $x_{ij}$ - значення j-го показника $b_j$ для i-го об’єкта $v_i$;

– експертні оцінки інтегральних показників $q = (q_1, \dots, q_n)^T$ для кожного об’єкту з параметрами $x_i = (x_{i1}, x_{i2} \dots, x_{im})$.

– експертні оцінки ваг $w_0$ , що характеризують відносну значущість часткових показників.

 

Зазначемо, що оцінки вектору $w_0$ можуть бути отримані на основі методу аналізц ієрархій в вигляді процедури попарного порівняння [10]. Оцінка відносної важливості часткових показників $w_0$ у подальшому використовується як апріорна оцінка вагових коефіцієнтів для покращення якості оцінок.

Модель інтегрального індикатора приймається у вигляді лінійної згортки:

$$\mathrm{J}(\mathrm{x_i}) = \bar{\mathrm{w}}^{\mathrm{T}}\mathrm{x_i} \tag{1}$$

*де $\bar{w}$ - вектор вагових коефіцієнтів, що підлягають визначенню.*

З використанням зазначених оцінок можуть бути знайдені прогнозовані значення інтегральних індикаторів $\bar{q}$, що відповідають наявним статистичним даним вимірювань часткових показників:

$$\bar{\mathrm{q}} = (\mathrm{J}(\mathrm{x_1}), \dots, \mathrm{J}(\mathrm{x_n}))^{\mathrm{T}} = \mathrm{X}\bar{\mathrm{w}}. \tag{2}$$

Оцінка вагових коефіцієнтів моделі інтегрального індикатора *(1), (2)* з використанням вхідних даних $q^T = (q_1, \dots, q_n)$, $X = \{x_{ij}\}_{i,j=1}^{n,m}$ та попередніх оцінок вагових коефіцієнтів $w_0$ здійснюється на основі резулярізованого методу найменших квадратів з урахуванням умов нормування $e^T w = 1$, де $e^T = (1, 1, \dots, 1)$.

Резулярізуюча складова, що забезпечує стійкість оцінок, формується з використанням апріорної інформації щодо відносних ваг показників.

Тоді маємо оптимізаційну задачу з обмеженнями:

$$\mathrm{Q}(\mathrm{q}, \mathrm{w}) = \frac{1}{2}\left||\mathrm{q} - \mathrm{Xw}|\right|^2 + \frac{1}{2}\alpha||\mathrm{w} - \mathrm{w_0}||^2 \tag{3}$$
$$e^T w = 1$$

Рішення задачі *(3)* отримуємо з використанням функції Лагранжа:

$$\mathrm{L}(\mathrm{q}, \mathrm{w}, \alpha) = \frac{1}{2}\left||\mathrm{q} - \mathrm{Xw}|\right|^2 + \frac{1}{2}\alpha||\mathrm{w} - \mathrm{w_0}||^2 + \mu(1 - \mathrm{e^T w}) \tag{4}$$

Тоді з умов мінімізації функції *(4)* оцінка параметрів моделі інтегрального індикатору набуває вигляду:

$$\bar{\mathrm{w}} = \Pi_{\mathrm{m}}(\alpha \mathrm{I_m} + \mathrm{X^T X})^{-1}\mathrm{X^T q} + \alpha \Pi_{\mathrm{m}} \mathrm{w_0} + \mathrm{m}^{-1}\mathrm{e}, \tag{5}$$

*де $\Pi_m = I_m - m^{-1}ee^{\mathrm{T}}$, $I_m$ – одинична матриця відповідного розміру*

Оцінки інтегральних показників і оцінки параметрів моделі *(5)* називаються узгодженими [8, 9], якщо вони задовольняють вимогам $q = \bar{q} = X\bar{w}$ . Зрозуміло, що для довільних вихідних даних ця умова не виконується, що вимагає додаткової корекції (координації) оцінок параметрів моделі інтегрального індикатора для забезпечення узгодженості оцінок.

Відповідно до запропонованої у [7] методиці оптимальної координації експертних оцінок, знайдемо уточнені значення параметрів моделі інтегрального показника із умов мінімізації функціоналу з додатковими обмеженнями:

$$\mathrm{M}(\mathrm{q}, \mathrm{w}) = \frac{1}{2}\left||\mathrm{q} - \mathrm{Xw}|\right|^2 + \frac{1}{2}\alpha||\mathrm{w} - \mathrm{w_0}||^2, \tag{6}$$
$$q = Xw, \qquad e^T w = 1.$$

Обмеження у *(6)* відповідають взаємозв’язку між експертними оцінками інтегральних та часткових показників для лінійної функції моделі інтегрального індикатору та вимогам нормалізації вагових коефіцієнтів.

 

Розв'язання оптимізаційної задачі *(6)* з обмеженнями отримуємо за допомоги функції Лагранжу:

$$\mathrm{L}(\mathrm{q}, \mathrm{w}, \alpha, \mu) = \mathrm{M}(\mathrm{q}, \mathrm{w}) + \alpha^{\mathrm{T}}(\mathrm{q} - \mathrm{Xw}) + \mu\left(1 - \mathrm{e}^{\mathrm{T}}\mathrm{w}\right). \quad (7)$$

Використовуючи відомі умови оптимальності Куна-Таккеру для функції *(7),* знайдемо співвідношення між оптимальними значеннями множників Лагранжу:

$$\mu = \mathrm{m}^{-1}\mathrm{e}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\alpha, \alpha = \mathrm{P}_{\mathrm{n}}^{-1}(\mathrm{q} - \mathrm{X}_{\mathrm{n}}\mathrm{w}_0), \quad (8)$$

*де* $P_n = \alpha I_n + X_n \Pi_m X_n^T, \quad \Pi_m = I_m - m^{-1}ee^{\mathrm{T}}, I_n, I_m$ *– одиничні матриці.*

З урахуванням отриманої залежності *(8),* розв'язання задачі оптимального узгодження експертних оцінок має наступний вигляд:

$$\overline{\mathrm{w}} = \mathrm{w}_0 + \Pi_{\mathrm{m}}\mathrm{X}^{\mathrm{T}}\mathrm{P}_{\mathrm{n}}^{-1}(\mathrm{q} - \mathrm{Xw}_0), \quad (9)$$

$$\bar{q} = Xw_0 + \Psi_n P_n^{-1}(q - Xw_0),$$

*де* $\Psi_n = X_n \Pi_m X_n^T, \quad P_n = \alpha I_n + \Psi_n.$

У сукупності з *(1),* формула *(9)* визначає параметри узгодженої моделі інтегрального індикатору.

Запропонований метод проілюстровано на прикладі побудови інтегрального індикатору якості міста [11]. Для побудови показників та оцінки якості використовуються реальні данні із набору даних MoveHub City Rankings https://www.kaggle.com/blitzr/movehub-city-rankings.

Набір навчальних даних містить дані про 198 міст, і кожне місто із набору даних описується 11 числовими ознаками часткових показників та інтегральною рейтинговою оцінкою MoveHub.

На рис. 1 представлена візуалізація побудованої лінійної моделі интегрального індикатору, перетворена до перших двох головних компонент набору статистичних даних.

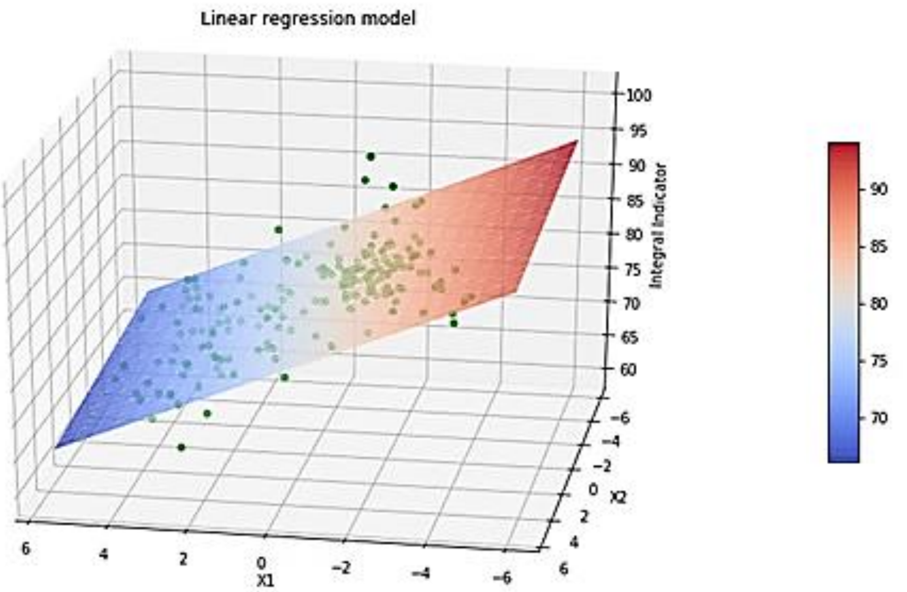


Рис. 1. Візуалізація лінійної моделі інтегрального індикатору.

Таким чином, запропонована методика дозволяє ефективно поєднувати експертну та статистичну інформацію при побудові інтегральних індикаторів.

 

**Список використаних джерел:**

[1]. Guihuana Z. & Yu W. (2014) Financial Conditions Index Construction and Its Application on Financial Monitoring and Economic Forecasting. *Procedia Computer Science*, (31), 32–39.

[2]. Chibilev A., Petrishchev V., Levykin S., Ashikkaliev A. & Kazachkov G. (2016) The social-ecological index as an integral indicator for the optimization of land-use structure. *Geography and Natural Resources,* (71–4), 3481–3514.

[3]. Aivazian S. & Mkhitariyan V. (1998) *Applied statistics and essential econometrics*. Moscow, UNITI.

[4]. Figueira J., Greco S. & Ehrgott M. (2005) *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer.

[5]. Barzilai J. (2005) Measurement and Preference Function Modeling. *Intern. Trans. in Operational Research*, (12), 1731–183.

[6]. Ruan L. & Yuan M. (2010) Dimension reduction and parameter estimation for additive index models. *Statistics and Its Interface*, (3), 4931–499.

[7]. Lyubchyk L. & Grinberg G. (2014) Preference Function Reconstruction for Multiple Criteria Decision Making Based on Machine Learning Approach. L.A. Zadeh et al. (eds). *Recent Developments and New Directions in Soft Computing* (53–63), Switzerland: Springer.

[8]. Strijov V. & Shakin V. (2003) Index construction: the expert-statistical method. *Environmental research, engineering and management*, (26–4), 51–55.

[9]. Kuznetsov V. & Strijov V. (2014) Methods of expert estimations concordance for integral quality estimation. *Expert Systems with Applications*, (41), 1988–1996.

[10]. Saaty T. (2016) The Analytic Hierarchy and Analytic Network Processes for the Measurement of Intangible Criteria and for Decision-Making. In: Greco S., Ehrgott M., Figueira J. (eds) Multiple Criteria Decision Analysis. *International Series in Operations Research & Management Science*, (233). Springer, New York, NY.

[11]. Lyubchyk L., Grinberg G. & Yamkovy K. (2018) Integral indicator for complex system building based on semi-supervised learning. *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, 134 –140.