УДК 303.094.5

МЕТОДИ ПОБУДОВИ ІНТЕГРАЛЬНОГО ПОКАЗНИКА

ГРИГОРУК П. М.
кандидат технічних наук
ТКАЧЕНКО І. С.
доктор економічних наук
Хмельницький

У ході практичної діяльності часто виникає завдання ранжування об'єктів спостереження за показниками, отриманими в результаті проведеного дослідження. Якщо кількість показників невелика, теоретично можна обрати з них найбільш значущий і провести відповідне ранжування за його значеннями. Якщо кількість показників є досить великою, вони мають різну природу, відображають різні характеристики досліджуваного явища або підстав для виявлення найбільш вагомого показника немає, вирішення завдання істотно ускладнюється. Більше того, якість, за якою необхідно провести ранжування або зіставлення об'єктів, може мати латентну природу, і, відповідно, не мати явного відображення серед відібраних показників. У такому випадку виникає завдання побудови узагальненого (інтегрального, зведеного) показника, за допомогою якого можна було б провести необхідне ранжування. Це завдання наведено лише як приклад, який ілюструє необхідність і доцільність побудови інтегрального показника.

Проблема розробки інтегрального показника останнім часом привертає увагу багатьох дослідників з різних наукових галузей, про що свідчать численні публікації з розробки та використання таких показників в економіці, соціології, педагогіці, медицині екології, житлово-комунальній сфері, військовій галузі тощо. Популярність даного напрямку зумовлена, на наш погляд, досить широким колом завдань, які можуть бути вирішені з його допомогою, серед яких можна виділити зіставлення об'єктів між собою, визначення структури об'єктів, класифікації об'єктів стосовно рівня досліджуваної якості, визначення загального рівня якості, класифікації нових об'єктів стосовно визначеної структури, виявлення ступеня відповідності досліджуваних об'єктів деякому уявному «ідеалу» та визначення напрямків покращення ситуації тощо.

Окремо слід сказати ще про декілька напрямків досліджень, специфіка вирішення завдань у яких також вимагає побудови інтегрального показника.

Якщо вихідні показники, якими описується досліджуване явище, мають оціночний характер, тобто, являють собою критерії, а сукупність об'єктів є альтернативами, з яких обирається найкраще рішення стосовно цих критерії, то ми приходимо до вирішення багатокритеріального завдання. Його складність полягає в тому, що альтернативи, прийнятні за одними критеріями, є непридатними за іншими. Можливим шляхом вирішення завдання є використання редукції критеріїв з урахуванням їх відносної значущості. Отриманий інтегральний показник називається критеріальним.

У ролі вихідних даних, які використовуються для побудови інтегрального показника, може виступати матриця симетричних бінарних відношень. Отриманий в

www.business-inform.net

результаті її опрацювання узагальнений показник може використовуватись для вирішення завдань подібних до завдань розпізнавання образів та автоматичної класифікації об'єктів. Тому такий показник називається вирішальним.

Ще одне завдання агрегації показників у єдиний узагальнений пов'язане з дискримінантним аналізом. Крім даних у формі «об'єкт – властивість», у ньому також фігурує деяка задана структура, що визначає відношення між об'єктами, розподіляючи їх на класи, що не перетинаються. У такому випадку постає завдання апроксимації заданої структури з метою можливості класифікації нових об'єктів, яке вирішується методами дискримінантного аналізу [1].

Під інтегральним показником будемо розуміти деякий умовний числовий вимірювач латентної якості досліджуваного явища. Реалізація ідеї побудови інтегрального показника пов'язана з трьома основними складовими, які становлять його фундаментальну базу: визначення його концепції; формування інформаційної бази; визначення алгоритму його розрахунку.

Метою побудови інтегрального показника є компактний опис деякої якості досліджуваного явища їх збереженням основних властивостей структури досліджуваних об'єктів.

Побудова інтегрального показника передбачає врахування певних вимог, серед яких найбільш істотними на наш погляд є такі:

- показник повинен чітко відображати мету його побудови і дозволяти вирішувати поставлені завдання;
- він повинен бути достатньо інформативним і володіти достатньою роздільною здатністю для досліджуваних об'єктів;
- він повинен піддаватись простій і зрозумілій інтерпретації;
- зміна його позитивної якості повинна відповідати напрямкам «переваги» його складових;
- показник повинен максимально враховувати інформативність його складових і при цьому допускати стиснення надлишкової інформації, що в них міститься;
- він повинен бути інваріантним стосовно одиниць вимірювання його складових;
- він повинен максимально відтворювати варіацію його складових;
- вагові коефіцієнти при його складових повинні бути статистично значущими.

Наведений перелік може бути доповнений з урахуванням специфіки застосування показника та особливостей його побудови.

Наявність великої кількості вихідних показників ускладнює процедуру побудови інтегрального показника, робить його громіздким, знижує його інформативність та дискримінуючи здатність, негативно впливає на значущість вагових коефіцієнтів. Виходом із ситуації може бути процедура послідовної згортки, в якій вихідні показники спочатку групуються за певною характеристикою. При цьому до кожної групи висуваються такі умови:

- показники кожної групи повинні відображати одну характеристику досліджуваних об'єктів;
- між показниками всередині групи повинні спостерігатись досить щільні кореляційні зв'язки;
- між показниками різних груп кореляційні зв'язки повинні бути незначними.

Для кожної групи визначається частковий узагальнений показник. Цей процес може відбуватись у декілька етапів, поки не буде досягнута прийнятна кількість базових показників. При цьому потрібно зауважити, що на наступних етапах перша та друга умови виконуватись не будуть. Недоліком даного підходу є втрата безпосереднього зв'язку між вихідними показниками і кінцевим інтегральним показником, що ускладнює його використання стосовно нових об'єктів.

Іншим способом є відбір з кожної групи найбільш «інформативного» показника за евристичними методами зниження розмірності [2]. Однак в такому випадку неодмінно втрачається частина інформації, що знижує цінність кінцевого результату.

Найбільш поширеним підходом до побудови залежності між інтегральним і базовими показниками є використання їх згортки. Найчастіше використовується дві її форми: адитивна і мультиплікативна. Процедура побудови конкретного вигляду інтегрального показника визначається як формою згортки, так і наявністю експертної інформації щодо ранжування об'єктів стосовно оцінюваної якості.

Визначення та обґрунтування вагових коефіцієнтів є можливо самим складним завданням в процедурі побудови інтегрального показника. Найбільш поширеним є використання експертного оцінювання вагомості кожного базового показника. Однак більш об'єктивним є використання оцінок, що ґрунтуються на статистичному опрацюванні даних.

Спрощений алгоритм побудови інтегрального показника, який має такий вигляд [3]:

1) Формування сукупності вихідних характеристик досліджуваного явища:

$$X = \{X_1, X_2, ..., X_n\}, \quad (1)$$

кожна з яких являє собою показник, виміряний за метричною шкалою. Ознаки реалізовані на деякій сукупності об'єктів:

$$O = \{O_1, O_2, ..., O_m\}. \quad (2)$$

2) Формування вектора $q = \{q_1(X), q_2(X), ..., q_s(X)\}$ окремих показників, які являють собою функції від сукупності вихідних показників і призначені для оцінювання окремих аспектів досліджуваних об'єктів з використанням s різноманітних критеріїв.

3) Вибір вигляду синтезуючої функції $Q = Q(w_1, w_2, ..., w_s, q_1, q_2, ..., q_s)$, яка ставить у відповідність вектору q значення зведеного показника Q, який характеризує об'єкт у цілому, з урахуванням вектора деяких додатних параметрів $w = \{w_1, w_2, ..., w_s\}$, що відображають значущість окремих складових вектора q.

www.business-inform.net

4) Вибір значень вектора *w*, який надалі будемо називати вектором вагових коефіцієнтів. Зазвичай на його компоненти додатково накладається нормувальна умова $w_1 + w_2, ..., + w_s = 1$, що дає підстави говорити про відносну важливість кожної складової інтегрального показника.

Зазвичай результати дослідження являють собою результати вимірювання деякої сукупності показників на об'єктах спостереження. При цьому зібрані дані надають певні уявлення стосовно взаємозв'язків (відношень) між об'єктами, однак будь-які відомості щодо значень або структури досліджуваної латентної якості на об'єктах відсутні. Метою побудови інтегрального показника в такому випадку є вирішення одного або декількох наведених вище завдань в умовах інформаційного дефіциту щодо досліджуваної якості.

Відомо, що ознака X_i називається стимулятором (має монотонно зростаючу залежність якості), якщо вищим значенням ознаки відповідає краща якість об'єкта. Ознака X_i називається дестимулятором (має монотонно спадаючу залежність якості), якщо нижчим значенням ознаки відповідає краща якість об'єкта. Ознака X_i називається номінатором (має немонотонну залежність якості), якщо існує деяке значення на проміжку зміни ознаки, яке відповідає найкращій якості об'єкта. Віддалення від цього значення призводить до погіршення якості.

У більшості випадків побудова інтегрального показника ґрунтується на вимозі подання всіх ознак як стимуляторів. У цьому випадку зберігається позитивний кореляційний зв'язок з тією якістю, яка буде досліджуватись. Крім того, як було зазначено вище, значення інтегрального показника не повинні залежати від одиниць вимірювання ознак. Для дотримання цих вимог провести уніфікацію шкал, за якими початково виміряні вихідні ознаки [3]. Ця процедура являє собою таке перетворення шкали (перенесення початку відліку і зміну масштабу), в результаті якого область можливих значень вимірювання завжди обмежується відрізком [0; *N*], де число *N*, що визначає розмах нової шкали і обирається дослідником із змістовних міркувань. При цьому нульове значення перетвореного показника повинне відповідати найнижчій якості за даною властивістю, а значення, рівне *N*,– найвищій. Зазвичай інтегральний показник будується таким чином, щоб його значення знаходились в межах від 0 до 1. Це покращує змістовну інтерпретацію його значення та дозволяє проводити зіставлення різних об'єктів. В такому випадку при уніфікації шкали значення *N* також обирається рівним одиниці. Надалі будемо розглядати саме такий випадок.

Якщо ознака є стимулятором, то перетворення здійснюється за правилом:

$$\tilde{x}_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}, \tag{3}$$

де $x_{j\min} = \min_i x_{ij}$; $x_{j\max} = \max_i x_{ij}$.

Якщо ознака є дестимулятором, то перетворення здійснюється за правилом:

$$\tilde{x}_{ij} = \frac{x_{j\max} - x_{ij}}{x_{j\max} - x_{j\min}}. \tag{4}$$

Зазначимо, що для деяких показників-дестимуляторів може мати місце, коли покращенню якості відповідає зниженню показників, але за умови їх додатних значень. Нульове чи від'ємне їх значення свідчить про погіршення якості. Відповідні перетворення мають такий вигляд:

$$\begin{cases} x_{j\min} = \begin{cases} \min\limits_{x_{ij}\ge 0} x_{ij}, & \text{якщо серед } x_{ij} \text{ є додатні,} \\ 0, & \text{якщо серед } x_{ij} \text{ всі від'ємні,} \end{cases} \\ x_{j\max} = \max\limits_i \left| x_{ij} - x_{j\min} \right|, \\ \tilde{x}_{ij} = \left(1 - \dfrac{\left| x_{ij} - x_{j\min} \right|}{x_{j\max}} \right), \end{cases} \tag{5}$$

У випадку поведінки ознаки як номінатора, перетворення має вигляд:

$$\tilde{x}_{ij} = \left(1 - \frac{\left| x_{ij} - x_{jnom} \right|}{\max\{(x_{j\max} - x_{jnom}), (x_{jnom} - x_{j\min})\}} \right), \tag{6}$$

де x_{jnom} – граничне (номінаторне) значення ознаки, для якого має місце найвища якість.

Стосовно наведених формул потрібно зробити декілька зауважень. По-перше, у вказаних перетвореннях найбільше та найменше значення за кожною ознакою розраховується за вибіркою. У деяких випадках вони можуть задаватись як нормативні значення показників. Тоді наведені перетворення у випадку стимуляторів матимуть вигляд:

$$\tilde{x}_{ij} = \begin{cases} \dfrac{x_{jnorm} - x_{ij}}{x_{jnorm} - x_{j\min}}, & x_{ij} \le x_{jnorm} \\ 1, & x_{ij} > x_{jnorm} \end{cases}, \tag{7}$$

для дестимуляторів:

$$\tilde{x}_{ij} = \begin{cases} \dfrac{x_{ij} - x_{jnorm}}{x_{j\max} - x_{jnorm}}, & x_{ij} \ge x_{jnorm} \\ 1, & x_{ij} < x_{jnorm} \end{cases}, \tag{8}$$

де x_{jnorm} – нормативне значення показника.

Важливим є питання визначення граничного значення x_{jnom} у випадку показників-номінаторів. Воно може обиратись, виходячи з додаткових відомостей стосовно оцінювання якості досліджуваного явища за такими показниками, або встановлюватись, як нормативне. Якщо така інформація відсутня, для визначення цього значення за даними вибірки можна використати таку залежність:

$$x_{jnom} = \frac{x_{j\min}^{(1)} + x_{j\max}^{(1)}}{2}, \tag{9}$$

де $x_{j\min}^{(1)}$ – найменше значення показника, при якому починає спостерігатись погіршення якості;

$x_{j\max}^{(1)}$ – найбільше значення показника, при якому якість покращується .

www.business-inform.net

Зазначимо, що перехід до уніфікованої шкали в термінах представленого вище алгоритму можна розглядати як вибір функцій q_j:

$$\tilde{X}_j = q_j(X_j). \qquad (10)$$

Наведені формули (3)– (9) можуть мати інший вигляд залежно від специфіки поведінки показника. Наприклад, якщо показники відображають відносні характеристики динаміки зміни деяких інших показників, або структурні характеристики, вигляд функцій q_j буде іншим, хоча кінцевим результатом все одно повинна бути уніфікація шкал.

Формування синтезуючої функції Q здійснюється шляхом використання лінійної адитивної або мультиплікативної згортки:

$$Q_A = \sum_{j=1}^{n} w_j \tilde{X}_j, \qquad (11)$$

$$Q_M = \prod_{j=1}^{n} \tilde{X}_j^{w_j}. \qquad (12)$$

Відзначимо, що в такому випадку значення інтегрального показника також будуть мати діапазон зміни значень [0; 1].

Адитивна згортка є більш поширеною і використовується, коли уніфікація показників проводилась за формулами (3) – (9), і є підстави вважати, кожна складова лінійно і адитивно впливає на досліджувану якість об'єктів. При цьому практично немає обмежень на кількість складових залежності.

Мультиплікативна згортка використовується тоді, коли базисні показники характеризують відносні величини. Зазвичай кількість складових обирається не більше семи. Крім того, нижнє значення уніфікованою шкали для оцінювання показників, рівне нулю, зазвичай не використовується. Слід також зауважити, що така згортка є надто чутливою до низьких значень базисних показників: близьке до нуля значення одного з них фактично може нівелювати вплив інших показників, що погіршує диференціюючи знатність інтегрального показника.

Методи використання адитивної та мультиплікативної згорток у науковій літературі одержали назву метода сум та метода добутків відповідно.

У тому випадку, якщо при оцінюванні якості використовується деякий еталонний об'єкт, який володіє найкращими значеннями (отриманими з вибірки, або гіпотетично) всіх показників (в термінах уніфікованої шкали – рівних одиниці), то вираз для адитивної згортки трансформується у вигляд:

$$Q_\Sigma = \left(\sum_{j=1}^{n} w_j \left| \tilde{x}_{ij} - 1 \right|^p \right)^{1/p}, \qquad (13)$$

де p – деякий показник ступеня.

Вираз (13) являє собою зважену відстань Мінковського, тому відповідний метод розрахунку отримав назву методу відстаней. Зростання значення p спричиняє більшу вагу максимального відхилення від еталону в загальній сумі, зменшення – навпаки. Найбільш поширеними є використання значень p = 1 (зважена відстань Хемінга), та p = 2 (зважена Евклідова відстань).

При використанні даного методу часто вагові коефіцієнти беруться однаковими, рівними $1/n$, що дещо спрощує розрахунки.

Певну модифікацію адитивної згортки можна отримати і при використанні методів зниження розмірності за умови повної редукції ознак, зокрема методу головних компонент та методу багатовимірного шкалювання.

Якщо в методі головних компонент перше власне значення коваріаційної матриці вихідних ознак задовольняє умові [4]:

$$\frac{\lambda_1}{\sum_{j=1}^{n} \lambda_j} \geq 0,55, \qquad (14)$$

то першу головну компоненту, що відповідає цьому власному значенню, можна використати в ролі інтегрального показника. Зауважимо, що алгоритм методу не передбачає уніфікацію шкал, за якими виміряні базисні показники. У даному випадку така умова є обов'язковою. Тоді вагові коефіцієнти w_j визначаються за формулою:

$$w_j = v_{j1}^2, \qquad (15)$$

де v_{j1} – компоненти нормованого власного вектора, що відповідає власному значенню γ_1.

У тому випадку, якщо умова (15) не виконується, згортка здійснюється у два етапи. На першому етапі відбувається розподіл вихідної сукупності ознак на p однорідних груп. Значення p визначається як найменша кількість перших власних значень коваріаційної матриці ознак $\tilde{X}_j$, для яких виконується умова:

$$\frac{\sum_{j=1}^{p} \lambda_j}{\sum_{j=1}^{n} \lambda_j} \geq 0,55. \qquad (16)$$

Для кожної виявленої таким чином однорідної групи розраховується своя коваріаційна матриця, і для неї обчислюється перша головна компонента, яка і буде виступати в ролі часткового інтегрального показника:

$$Q^{(i)} = \sum_{j=1}^{n_i} w_j^{(i)} \tilde{X}_j^{(i)}, \qquad (17)$$

де $i = 1..p$ – номер групи;

n_i – кількість ознак i-тої групи;

$Q^{(i)}$ – частковий інтегральний показник;

$\tilde{X}_j^{(i)}, w_j^{(i)}$ – базисні показники, що увійшли до i-тої групи та відповідні їх вагові коефіцієнти, які розраховуються за формулою (17).

На другому етапі розраховується зважена евклідова відстань d_i в просторі часткових інтегральних показників від вихідних об'єктів до еталону:

$$d_i = \sqrt{\sum_{j=1}^{p} w_j (Q_i^{(j)} - 1)^2}, \qquad (18)$$

де вагові коефіцієнти w_j обчислюються за формулою:

www.business-inform.net

$$w_j = \frac{\sum_{i=1}^{m}(Q_i^{(j)} - \bar{Q}^{(j)})^2}{\sum_{j=1}^{p}\sum_{i=1}^{m}(Q_i^{(j)} - \bar{Q}^{(j)})^2}, \quad (19)$$

тобто, визначаються пропорційно вибірковим дисперсіям часткових інтегральних показників.

Значення узагальненого інтегрального показника визначається за формулою:

$$Q_i = 1 - d_i. \quad (20)$$

Наведений метод являє собою приклад нелінійної згортки.

Побудову узагальненого показника можна здійснити і за допомогою методів метричного шкалювання, спроектувавши вихідний багатомірний простір на одну шкалу [5]. Однак недоліком такого підходу є те, що отриманий інтегральний показник втрачає зв'язок з базисними показниками, що ускладнює інтерпретацію результату. Разом з тим цей підхід дозволяє вирішувати завдання ранжування об'єктів, їх зіставлення, виявлення їх структури та деякі інші, що цілком відповідає призначенню інтегрального показника.

Очевидно, що використання різних підходів до побудови інтегрального показника приводить, взагалі кажучи, до різних результатів. При цьому природним виникає завдання пошуку «найкращої» шкали для вимірювання за побудованим показником. Таке завдання можна вирішити в рамках теорії функціонального шкалювання [6].

Використання уніфікованої шкали для інтегрального показника має свої особливості в інтерпретації його результатів. З одного боку, це полегшує цей процес, оскільки дозволяє порівнювати результати за різні періоди часу, обґрунтовано визначати підходи щодо поділу шкали на сукупність рівнів для проведення ранжованого групування, зіставляти результати розрахунків за різними показниками (за умови коректності такого зіставлення).

Однак такий формат шкали має і певні недоліки. У першу чергу вони визначаються тим, що шкала не має одиниць вимірювання. Крім того, вона не допускає арифметичних перетворень. Наприклад, твердження, що значення показника одного об'єкта, удвічі більше за значення для іншого об'єкта, свідчить про удвічі більшу його якість, взагалі кажучи, безпідставне. Так само фактично втрачає сенс визначення відстані між об'єктами на цій шкалі. Тому інтерпретація результатів допускає лише їх рейтингування, для чого можна визначати не лише відносний порядок розташування об'єктів, але і їх групування. ■

ЛІТЕРАТУРА

1. Факторный, дискриминантный и кластерный анализ / [Дж-О. Ким, Ч. У. Мюллер, У. Р. Клекка и др. ; пер. с англ ; под. ред. И. С. Енюкова].– М. : Финансы и статистика, 1989.– 215 с. : ил. – ISBN 5-279-00247-X.
2. **Айвазян С. А.** Прикладная статистика и основы эконометрики. Учебник для вузов / С. А. Айвазян, В. С. Мхитарян.– М. : ЮНИТИ, 1998.–1022 с.– ISBN 5-238-00013-8.
3. **Хованов Н. В.** Анализ и синтез показателей при информационном дефиците / Н. В. Хованов.– СПб. : Изд-во С.-Петербуржского ун-та, 1996.– 196 с.– ISBN 5-288-01533-3.
4. **Бородкин Ф. М.** Социальные индикаторы : учебник / Ф. М. Бородкин, С. А. Айвазян.– М. : ЮНИТИ-ДАНА, 2006.– 607 с.– ISBN 5-238-01094-X.
5. **Перекрест Ф. Т.** Функциональный подход в метрическом одномерном шкалировании / Ф. Т. Перекрест // Анализ нечисловой информации в социологических исследованиях / Отв. ред. В. Г Андреенков, А. И. Орлов, Ю. Н. Толстова.– М. : Наука, 1985.– С. 113 – 131.
6. **Авен П. О.** Функциональное шкалирование / П. О. Авен, И. Б. Мучник, А. А. Ослон ; [отв. ред. д. ф.-м. н. Б. А. Березовский].– М. : Наука, 1988.– 177 с.– ISBN 5-02-006597-8.

www.business-inform.net