
УДК 681.51

О.Я. КУЗЬОМІН, В.В. ЛЯШЕНКО

ІДЕНТИФІКАЦІЯ СКЛАДОВИХ ЕКОНОМІЧНИХ КЛАСТЕРІВ У ПРОСТОРІ ПЕРЕТВОРЕНИХ ДАНИХ

Розглянуто формалізоване подання сутності економічного кластеру. Запропоновано підхід щодо ідентифікації елементів економічного кластеру та надано його імовірнісну інтерпретацію. Визначено похибку ідентифікації елементів економічного кластеру на основі запропонованого підходу.

Вступ

Сучасні методи аналізу діяльності суб'єктів господарювання передбачають застосування досить широкого кола статистичних даних, що визначається наявністю цілої сукупності внутрішніх та зовнішніх впливів на сталість функціонування суб'єктів ринкових відносин. Разом з цим обробка статистичних даних визначається не лише простим узагальненням діяльності функціонування окремих суб'єктів господарювання з погляду одного з показників. Для отримання зважених висновків доцільно використовувати для відповідного аналізу як низку показників за одним із досліджуваних суб'єктом, так й низку різних суб'єктів господарювання, які у підсумку і визначають економічну змістовність наявних зв'язків у ринкових умовах ведення господарської діяльності.

Загалом відмічений напрям дослідження проводиться з погляду так званих економічних кластерів [1–4], що і дозволяє узагальнити розгляд різноманітних економічних даних. Однак, незважаючи на застосування поняття кластеру в економічних роботах, й досі не доопрацьованими залишаються питання формалізованого застосування кластерів як інструменту дослідження. Проте формалізація розгляду кластерів дозволяє не лише більш чіткіше описувати явища та процеси економічних відносин, а й визначати методи порівняльного оцінювання діяльності різних угруповань суб'єктів господарювання на основі методів розпізнавання, які мають значне обґрунтування з погляду їх використання у різних сферах діяльності людини. Разом з цим слід враховувати, що узагальнення економічних кластерів можливе як на підставі первісних економічних даних, так й перетворених даних, що розширює можливості проведення відповідного економетричного аналізу. Тому, в даному випадку постає задача ідентифікації економічних даних з погляду на формалізоване їх уявлення щодо таких угруповань суб'єктів ринкової економіки як кластери.

Таким чином, *головна мета даної роботи* – формалізоване узагальнення сутності економічного кластеру та визначення підходу щодо ідентифікації складових економічного кластеру у просторі перетворених даних на основі економічних показників діяльності різноманітних суб'єктів господарювання.

Загальне та формалізоване узагальнення сутності економічного кластеру

В економічному сенсі під кластером загалом розуміють певну концентрацію суб'єктів ринку, пов'язаних між собою економічними зв'язками, які визначають як розвиток такого угруповання, так й розвиток інших кластерів, що можуть знаходитися як у прямій, так і опосередкованій взаємодії між собою. При цьому підкреслення концентрації суб'єктів ринку, пов'язаних між собою економічними зв'язками, визначає їх компактність з погляду подання формалізованого узагальнення економічного кластеру. Інакше кажучи, у певному фазовому просторі економічний кластер може розглядатися як компактна множина, що складається з окремих елементів, кожен з яких визначає той або інший суб'єкт господарювання, що входить до такого кластеру.

Як економічний кластер можна, наприклад, розглядати сукупність банків, які здійснюють свою діяльність у межах територіально-адміністративних одиниць країни. Кластером також є, наприклад, і сукупність виробничих підприємств, які здійснюють свою діяльність у межах відповідних територіально-адміністративних одиниць країни. Поруч із цим як кластер можна визначити й сукупність усіх суб'єктів господарювання, які діють у межах територіально-адміністративних одиниць країни, або в межах окремих секторів економіки, галузях виробництва. При цьому в будь-якому з визначених випадків класичного подання економічного кластеру його можна формалізувати з погляду множини показників діяльності суб'єктів господарювання, що утворюють такий кластер. Так, якщо це сукупність банків, яка діє в межах окремої територіально-адміністративної одиниці країни, то відповідний кластер можна охарактеризувати через сукупність показників банківської діяльності, таких як обсяги наданих кредитів, залучених ресурсів, значень отриманого фінансового результату тощо. Тобто, у формалізованому вигляді економічний кластер через первісні дані про діяльність суб'єктів ринкових відносин, які його визначають, може бути описаний як вектор абсолютних показників діяльності таких суб'єктів господарювання. В такому опису кластеру можуть бути й такі дані, що не відповідають його загальноекономічній інтерпретації. Наприклад, якщо ми говоримо про діяльність банків за ознакою розташування їх головної установи, то в межах територіально-адміністративних одиниць країни можуть діяти й такі філії банків, головні установи яких не належать до визначеної територіально-адміністративної одиниці. Тому, в даному випадку слід говорити про похибку кластеризації даних та важливість проведення ідентифікації окремих складових економічних кластерів.

Якщо ми маємо справу з перетвореними даними про діяльність суб'єктів господарювання, то в даному випадку опис кластеру таких суб'єктів можуть бути використані відносні дані щодо їх діяльності. Так як продовження наведеного вище прикладу з погляду банків це можуть бути відносні дані щодо обсягів наданих кредитів, залучених ресурсів тощо з погляду певного регіону, або з урахуванням банківського сектору економіки тощо. При цьому також можлива похибка у визначенні складових окремих кластерів.

З метою ж уніфікації перетворених даних для всієї сукупності суб'єктів господарювання, що розглядаються (вважають, наприклад, для всієї галузі, всієї країни), доцільно провести відображення отриманих даних на деяку шкалу, яка дозволяє оперувати перетвореними даними у межах встановленого діапазону. Тому в такому випадку загальний економічний простір, в якому діє безліч різноманітних суб'єктів господарювання, можна визначити як загальний економічний кластер через множину даних $V(i)$, значення яких належить визначеній області $Q = \{0, 1, \dots, q - 1\}$, де q – розмір шкали перетворення відносних даних про діяльність суб'єктів господарювання, i – окремий елемент загального кластеру, $V(\dots)$ – значення окремого елемента загального кластеру. Це надає можливість розглядати кластер через залежність частоти попадання елементів, що розглядаються у відповідні інтервали їх групування [5]. Тобто, мова йдеться про опис кластеру у вигляді гістограми можливих значень про опис діяльності суб'єктів господарювання.

При цьому на загальному кластері можна визначити й окремі кластери, які було розглянуто вище. Отже загальний економічний кластер – це сукупність різноманітних кластерів, певна кількість яких може визначатися як помилково визначені кластери, якщо визначається заданий кластер:

$$V(i) = K_1(j_1) \cup K_2(j_2) \cup \dots \cup K_n(j_n), \quad (1)$$

де K_1, K_2, \dots, K_n – різноманітні економічні кластери, що визначаються їх окремими елементами j_1, j_2, \dots, j_n . Тобто, якщо кластер K_1 – заданий для визначення кластер, то інші є помилковими кластерами. Виходячи з того, що також можлива похибка у визначенні складових окремих кластерів загалом деякий кластер K_n має елементи, що дійсно йому відповідають \tilde{K}_n та ті, що є помилковими \hat{K}_n .

Тоді, відповідно до поданого вище, як загальний економічний кластер, так й окремі кластери, що його складають, можна подати у вигляді окремих гістограм.

Разом з цим для визначення окремого кластеру із загального економічного кластеру можна скористатися таким відображенням π :

$$\pi = \begin{cases} 1, & j_n \in K_n, \\ 0, & j_n \notin K_n. \end{cases} \quad (2)$$

Для більш чіткої формалізації відображення π розглянемо такі функції уявлення кластерів:

$$H(K_n) = \frac{\Gamma_n(K_n)}{\Gamma(B)}, \quad (3)$$

$$H(Z) = \frac{\Gamma(Z)}{\Gamma(B)}, \quad (4)$$

де $\Gamma_n(K_n)$ – гістограма значень відносних даних про діяльність суб'єктів господарювання, за допомогою яких описується K_n кластер; $\Gamma(B)$ – гістограма значень відносних даних про діяльність суб'єктів господарювання, за допомогою яких описується загальний економічний кластер; $\Gamma(Z)$ – гістограма значень відносних даних про діяльність суб'єктів господарювання, за допомогою яких можуть бути описані помилково визначені кластери. Загалом

$$\Gamma(Z) = \Gamma(B) - \Gamma_n(K_n). \quad (5)$$

Отже формули 3 та 4 визначають приналежність перетворених відносних даних про діяльність суб'єктів господарювання значенням кластеру K_n та іншим угрупованням суб'єктів, які розглядаються у загальному економічному кластері.

Відповідно до зроблених зауважень відображенням π за формулою 2 можна подати так:

$$\pi_r = \begin{cases} 1, & H(K_n) - H(Z) > r, \\ 0, & H(K_n) - H(Z) \leq r, \end{cases} \quad (6)$$

де r – є граничне значення приналежності окремих складових загального економічного кластеру до складових визначеного кластеру K_n .

Отже, відображення π можна вважати основою ідентифікації складових визначеного економічного кластеру в просторі перетворених даних. Проте відкритим залишається питання якості побудови зазначеного відображення, що може бути визначено на основі врахування можливої похибки ідентифікації. Однак перш ніж визначити таку похибку узагальнимо імовірнісну інтерпретацію відображення π з метою його порівняння з іншими відомими методами ідентифікації даних.

Імовірнісна інтерпретація відображення π

Для вирішення поставленого питання подамо умову $H(K_n) - H(Z) > r$ таким чином

$$\frac{N(K_n/x) - N(Z/x)}{N(B/x)} > r, \quad (7)$$

де $N(K_n/x)$, $N(B/x)$ – кількість окремих елементів кластерів K_n та B із значенням x .

Виходячи з того, що

$$N(Z/x) = N(B/x) - N(K_n/x), \quad (8)$$

то вираз (7) можна записати так:

$$\frac{2N(K_n/x) - N(B/x)}{N(B/x)} > r, \quad (9)$$

або

$$\frac{N(K_n/x)}{N(Z/x)} > \frac{1+r}{1-r}, \quad (10)$$

Для подальшого викладення матеріалу розглянемо класичний Байєсівський класифікатор віднесення елементів загального кластеру із значення x до визначеного кластеру [6]. Таке правило віднесення елементів загального кластеру із значення x до визначеного кластеру можна подати так:

$$P(K_n/x) > P(Z/x) \times \Psi, \quad (11)$$

де $P(K_n/x)$ – імовірність того, що складовий елемент загального кластеру є елементом кластеру K_n за умови, що значення такого елемента дорівнює x ; $P(Z/x)$ – імовірність того, що складовий елемент загального кластеру є елементом кластеру Z за умови, що значення такого елемента дорівнює x ; Ψ – параметр вартості помилкової ідентифікації.

Вираз (11) також можна подати і таким чином:

$$\frac{P(x/K_n)}{P(x/Z)} > \frac{\alpha}{1-\alpha} \times \Psi, \quad (12)$$

де $P(x/K_n)$, $P(x/Z)$ – імовірність того, що елемент кластеру має значення x за умови того, що він належить кластеру K_n або Z ; α – апіорна імовірність віднесення помилкового елемента з кластеру Z до кластеру K_n .

Продовжимо перетворення виразу (10) через $P(K_n/x)$ та $P(Z/x)$:

$$N(K_n/x) = N(\tilde{K}_n/x) + N(\hat{K}_n/x), \quad (13)$$

$$\frac{N(\tilde{K}_n/x) + N(\hat{K}_n/x)}{N(Z/x)} > \frac{1+r}{1-r}, \quad (14)$$

або

$$\frac{N_d P(K_n/x)}{N_f P(Z/x)} > \frac{\sigma + r}{1 - r}, \quad (15)$$

де σ – можлива кількість помилкових значень при визначенні кластеру K_n : $\sigma = \text{card}\hat{K}_n$; N_d, N_f – загальна кількість елементів кластерів K_n та Z відповідно.

Якщо $N(\hat{K}_n / x) = N(Z / x)$, то $\sigma = 0$. Якщо $N(\hat{K}_n / x) = 0$, то $\sigma = 1$.

Перепишемо вираз (15) так:

$$P(K_n / x) > P(Z / x) \times L, \quad L = \frac{\sigma + r}{1 - r} \times \frac{N_f}{N_d}. \quad (16)$$

Порівнюючи вираз (16) та (11) можна відмітити, що параметр L у виразі (16) виконує функцію параметру ψ у виразі (11). При цьому значення параметру L на відміну від значення параметру ψ враховує вартість помилкової класифікації з погляду окремих значень елементів кластеру, який розглядається, що робить таку вартість більш зваженою та ґрунтовною.

Похибки ідентифікації економічного кластеру на основі використання відображення π

Для з'ясування можливої похибки ідентифікації елементів визначеного кластеру розглянемо деякі відображення π в залежності від параметру r . Для цього визначатимемо r як

$$r = \frac{N_K}{N_B}, \quad (17)$$

де N_K – загальна кількість елементів у деякому кластері K , який ідентифікується; N_B – кількість елементів у загальному кластері B .

Припустимо, що окремі значення елементів кластеру K , які ідентифікуються, належать деякій множині $\{\alpha_s\}$, а значення елементів кластеру B , які можуть бути визначені як помилкові елементи, тобто значення елементів кластеру Z належать деякій множині $\{\beta_t\}$ і разом з цим $\forall s, t; \alpha_s \neq \beta_t$. Тоді ідентифіковані елементи із значенням α_s будуть визначені як елементи кластеру K . Похибка ж визначення елементів із загального кластеру B із значенням β_t може бути визначена на основі такого виразу:

$$\frac{N_B}{N_K} > \max_t \frac{\sigma_t + 1}{\sigma_t - 1}, \quad (18)$$

де $\frac{N_B}{N_K} = \frac{1}{r}$.

Висновок виразу (18) базується на наступному. Припустимо, що усі елементи кластеру K визначаються із значенням α_1 , а елементи кластеру Z лише із значенням β_1 . Тоді для елементів із значенням α_1 $H(K_n) = 1$, $H(Z) = 0$, а $H(K_n) - H(Z) = 1$ і відображення π правильно ідентифікує елементи кластеру K за будь-яких співвідношень N_K та N_B , бо $N_B > N_K$.

Для елементів із значенням β_1 маємо: $H(K_n) = \frac{N_K^{\beta_1}}{N_B^{\beta_1}}$, $H(Z) = \frac{N_Z^{\beta_1}}{N_B^{\beta_1}}$. Припустивши, що

$N_K^{\beta_1} = \sigma \cdot N_Z^{\beta_1}$, можемо записати похибку ідентифікації так:

$$\frac{N_B}{N_K} > \frac{\sigma + 1}{\sigma - 1}. \quad (19)$$

Перебираючи усі значення елементів кластерів із їх можливої множини $\{\alpha_s\}$ та $\{\beta_t\}$ отримаємо відповідне узагальнення виразу (19), яке і подано формулою (18).

У разі, якщо окремі значення елементів кластеру K , які ідентифікуються, належать деякій множині $\{\alpha_s\}$, а значення елементів кластеру B , які можуть бути визначені як

помилкові елементи, тобто значення елементів кластеру Z належать деякій множині $\{\beta_t\}$ і разом з цим $\{\alpha_s\} \cap \{\beta_t\} = \{\gamma_r\} \neq \emptyset$, то ідентифіковані елементи із значенням $\alpha_s \in \{\gamma_r\}$ визначатимуть елементи кластеру K . Для елементів кластеру із значенням $\beta_t \in \{\gamma_r\}$ похибка ідентифікації визначатиметься відповідно до виразу (18), бо в даному випадку узагальнення множин значень $\{\alpha_s\}$ та $\{\beta_t\}$ відповідних кластерів є ідентичним до розглянутого вище. Отже, залишається оцінити похибку ідентифікації для елементів кластеру із значенням $\alpha_s \in \{\gamma_r\}$.

Припустимо, що $N_{\gamma_r}^K$ – кількість елементів кластеру K із значенням γ_r , $N_{\gamma_r}^Z$ – кількість елементів кластеру Z із значенням γ_r , $N_{\gamma_r}^K = d_r N_{\gamma_r}^Z$, $d_r > 1$. Тоді елемент із значенням $\alpha_s \in \{\gamma_r\}$ буде ідентифіковано як елемент кластеру у такому випадку:

$$\frac{N_B}{N_K} > \min_t \frac{d_r + 1}{d_r - 1}. \quad (20)$$

Тобто, саме вираз (20) і визначає похибку можливої ідентифікації.

Висновки

Отже, в роботі розглянуто загальноекономічну сутність кластерів, що дозволило надати їх формалізоване узагальнення. Разом з цим запропоновано підхід щодо ідентифікації елементів кластерів. Показано, що такий підхід у порівнянні з класичними імовірнісними методами ідентифікації даних має перевагу у сенсі визначення зваженої вартості помилки класифікації з погляду окремих значень елементів кластеру. До того ж у роботі наведено узагальнення щодо визначення похибки ідентифікації економічного кластеру на основі запропонованого підходу. Виходячи з того, що на ідентифікацію елементів економічних кластерів суттєвий вплив справляють збурення зовнішнього середовища, слід визначити за доцільним розгляд похибки ідентифікації в залежності від різновидів таких збурень. Таким чином, відмічене можна вважати як напрям подальших досліджень.

Список літератури: 1. Ковальова Ю. М. Класифікаційна характеристика, типологія кластерів / Н. В. Алишева, Ю. М. Ковальова // Економіка: проблеми теорії та практики. Дніпропетровськ, 2008. Т. 1, вип. 239. С. 207–217. 2. Канищенко Н. Г. Кластери в системі національної конкурентоспроможності / Н. Г. Канищенко // Вісник КНУ. Серія Економіка. 2006. Вип. 85. С. 14–16. 3. Богма О. С. Застосування кластерного підходу у якості стратегії підвищення конкурентоспроможності підприємств національної економіки / О. С. Богма // Прометей: регіональний збірник наукових праць з економіки. Донецьк: ТОВ “Юго-Восток, Лтд”, 2006. С. 157–159. 4. Крпельницька С. О. Методологічні основи створення кластерів в українській економіці: регіональний аспект / С. О. Крпельницька // Вісник Прикарпатського ун-ту. Серія: Економіка. Івано-Франківськ: Плай, 2005. Вип. 2. С. 54–58. 5. Общая теория статистики: Учебник / Т. В. Рябушкин, М. Р. Ефимова, И. М. Ипатова, Н. И. Яковлева. М.: Финансы и статистика, 1981. 342 с. 6. Comer M. L. Segmentation of Textured Image Using a Multiresolution Gaussian Autoregressive Model / M. L. Comer, E. J. Fellow // IEEE TRANSACTIONS ON IMAGE PROCESSING. 1999, Vol. 8., No. 3. P. 408–420.

Надійшла до редколегії 15.08.2010

Кузьомін Олександр Якович, д-р техн. наук, професор кафедри інформатики ХНУРЕ, начальник інформаційно-маркетингового відділу ХНУРЕ. Наукові інтереси: управління ризиками, геоінформаційні системи. Адреса: Україна, 61166, Харків, пр. Леніна, 14, тел. 70-21-515, e-mail: kuzy@kture.kharkov.ua.

Ляшенко Владислав Вікторович, завідувач лабораторії «Трансфер інформаційних технологій щодо систем скорочення ризиків лих» ХНУРЕ. Адреса: Україна, 61166, Харків, пр. Леніна, 14, тел. 70-21-515, e-mail: kuzy@kture.kharkov.ua.