

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ Й ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

УДК 004.04; 681.3; 332.1

ВИКОРИСТАННЯ ФАКТОРНОГО АНАЛІЗУ В СИСТЕМІ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ДАНИХ МОНІТОРИНГУ СОЦІАЛЬНО-ЕКОНОМІЧНОГО РОЗВИТКУ РЕГІОНІВ

О. І. Пурський, доктор фізико-математичних наук. І. О. Мороз
Київський національний торговельно-економічний університет.

© Пурський О. І., 2013.

© Мороз І. О., 2013

Стаття отримана редакцією 13.03.2013 р.

Вступ. Управління регіональним розвитком тісно пов'язане з виявленням, описом і аналізом внутрішніх закономірностей його формування та розвитку. Характерні закономірності функціонування регіону проявляються в динаміці показників, які описують стан регіону в конкретно визначений момент часу. Взаємозв'язки, що відображаються у зміні великої кількості окремо взятих показників, підтверджують дію певної кількості факторів, які впливають на ефективність функціонування регіону. Визначення впливу цих факторів дозволяє розв'язати такі важливі завдання, як [1–3] перехід від великої кількості показників до меншого числа факторів без втрати змістовної інформативності та виявлення структури сукупності показників, що повністю характеризують стан регіону. Визначення факторів, які безпосередньо впливають на механізми функціонування регіону, дозволяє оптимізувати процес прийняття управлінських рішень і, як наслідок, підвищити загальну ефективність системи регіонального управління, що і визначає актуальність цього дослідження.

Огляд останніх джерел досліджень і публікацій. Дослідженню питань соціально-економічного розвитку присвячена значна кількість праць як закордонних дослідників Н. Blair, М. Donald Fisk, М. John Greiner, R. John Hall, S. Philip Schaanman, М. Кастельса, так і вітчизняних економістів-науковців, зокрема В. Г. Андрійчука, Я. О. Побурка та інших [4–6]. У роботах зазначається, що розв'язання проблеми управління соціально-економічним розвитком регіону має спиратися на повне й адекватне відображення економічної та соціальної динаміки по кожному окремо визначеному показникові системи адміністративно-територіального устрою. При аналізі даних моніторингу соціально-економічного розвитку регіонів особливу актуальність представляють ситуації, що описуються великою кількістю різних властивостей, кожна з яких має суттєвий вплив на їхні характеристики. Такі багатофакторні ситуації

виникають при аналізі соціально-економічних систем (СЕС), що пояснюється їх складністю та слабою дослідженістю [7]. Інформація про таку систему, як правило, фіксується у вигляді сукупності властивостей виділених одиниць (об'єктів) спостереження. Також характерним є факт взаємозалежності багатьох ознак і їх взаємного дублювання. У той же час нерідко ознаки тільки в опосередкованій формі відображають найбільш суттєві властивості СЕС, які не підлягають безпосередньому спостереженню і вимірюванню. У таких випадках виконуються спроби сконцентрувати інформацію, виражаючи велику кількість початкових опосередкованих ознак через меншу кількість більш смісних внутрішніх характеристик соціально-економічного явища [3, 8, 9]. При цьому вважається, що найбільш смісні характеристики виявляються одночасно найбільш суттєвими і визначальними. Таким чином, виникає задача побудови інтегрального (комплексного) показника, який би ставив у відповідність кожному досліджуваному об'єктові певну скалярну величину [10]. Такі показники є індикаторами соціально-економічної регіональної політики та інструментарієм для виявлення проблем у досліджуваних СЕС.

Постановка завдання. Завданням дослідження є розробка механізму інтелектуальної обробки даних соціально-економічного моніторингу регіону на основі методів факторного аналізу.

Основний матеріал і результати. Ураховуючи визначення Data Mining (зокрема, що знання які виявляються, повинні бути раніше невідомими, нетривіальними, практично корисними і доступними для інтерпретації людиною), задачу побудови інтегральних показників можна віднести до задач інтелектуального аналізу даних (ІАД) [11, 12]. У нашому випадку для побудови інтегральних показників соціально-економічного розвитку запропоновано використовувати двоетапний підхід. На першому етапі відбувається скорочення розмірності початкового простору ознак (факторів). Зменшення розмірності простору ознак засновано на застосуванні методів факторного аналізу [3, 13]. Їх суть полягає в переході від опису деякої множини досліджуваних об'єктів, заданого великим набором опосередкованих, безпосередньо вимірюваних ознак, до опису меншим числом максимально інформативних змістовних змінних (факторів), які відображають найбільш важливі властивості соціально-економічного явища. З метою отримання такого скороченого набору факторів ми використовуємо один з методів факторного аналізу – метод головних компонент [3, 14]. Ці фактори є деякими функціями початкових ознак. Отриманий у результаті застосування цього методу скорочений набір факторів має важливу особливість їх незалежності всередині системи. Наступний етап полягає в отриманні на основі скороченого набору незалежних факторів одного інтегрального показника, який об'єднував би в собі найкращим чином усі ці фактори [3]. Можливі три варіанти його визначення: важливість (вага) кожного фактора визначається експертом у відповідній предметній області, коефіцієнти важливості кожного фактора розраховуються автоматично на основі аналізу факторів, комбінований варіант поєднує в собі два попередніх варіанти [3]. Останній експертно-статистичний варіант є найбільш придатним для нашого випадку, оскільки врахування при розрахунках знань і досвіду експертів предметної області суттєво підвищує надійність отриманих у результаті досліджень висновків.

Важливою особливістю факторного аналізу є можливість одночасного дослідження необмежено великої кількості взаємопов'язаних змінних, завдяки чому факторний аналіз є цінним інструментом дослідження соціально-економічних явищ у всьому різноманітті їх реальних взаємозв'язків. Для соціально-економічних досліджень відсутність обмежень на число і взаємозалежність змінних є надзвичайно важливою обставиною, оскільки виділити вплив окремих факторів на поведінку всієї СЕС виявляється досить складним завданням. Тому в цьому випадку з успіхом можна застосовувати методи факторного аналізу, які дозволяють здійснити чітку структурування соціально-економічних даних і шляхом розрахунків виділити відносно незалежні фактори, що визначають сутність явища.

У нашому дослідженні фактор є розрахунковою величиною, тобто являє певну нову характеристику досліджуваної множини об'єктів. Опис фактора в термінах його зв'язку з набором початкових ознак знаходиться у вигляді матриці факторів A розмірністю $n \times m$, де n – число ознак, m – число факторів [3]. Базою для побудови матриці факторів A є матриця попарних кореляцій R розмірністю $n \times n$. Вона відображає ступінь взаємозв'язку між кожною

парою ознак, тоді як матриця факторів характеризує ступінь зв'язку між кожною із n ознак та m факторами, виявленими в процесі аналізу. При цьому число факторів m повинно бути значно меншим за n , а рівень втрати інформативності незначним. Матриця факторів дозволяє нам для кожного фактора виділити групу найбільш тісно з ним пов'язаних ознак. Тим самим надається можливість зіставити різні фактори, дати їм змістовну трактовку, тобто здійснити інтерпретацію факторів у рамках СЕС. Варто зазначити, що перетворення матриці попарних кореляцій у факторну не є однозначним, оскільки одна і та ж матриця кореляцій може бути перетворена в різні факторні [3]. Для нашого випадку, матриця факторів визначається таким чином, щоб певна множина початкових ознак була значно пов'язана тільки з одним із набору факторів і слабо пов'язана з усіма іншими, із цією метою за допомогою методів ІАД, автоматично вибираються найбільш інформаційні ознаки з точки зору їх впливу на цільовий фактор. Для виявлення латентних ознак використовуються такі методи ІАД як дерево рішень, що вибудовує ієрархічну структуру класифікуючих правил продукційного типу, та метод асоціативних правил, призначений для пошуку груп спільних характеристик пов'язаних між собою ознак.

Вважаємо, що є множина N ($i = 1, 2, \dots, N$) спостережень певного досліджуваного соціально-економічного явища. При цьому явище описується набором із n ($j = 1, 2, \dots, n$) ознак. Тобто наведена в соціально-економічному дослідженні інформація може бути описана у вигляді матриці X розмірністю $N \times n$

$$\begin{matrix}
 x_{11} & \dots & x_{1j} & \dots & x_{1n} \\
 \dots & \dots & \dots & \dots & \dots \\
 x_{i1} & \dots & x_{ij} & \dots & x_{in} \\
 \dots & \dots & \dots & \dots & \dots \\
 x_{N1} & \dots & x_{Nj} & \dots & x_{Nn}
 \end{matrix} \tag{1}$$

Рядки матриці відповідають спостереженням, тобто множині процесів, об'єктів та ін., що представляють соціально-економічне явище. Об'єктами таких спостережень в економічних дослідженнях можуть бути підприємства, галузеві об'єкти, види продукції, економічні процеси. При регіональних дослідженнях об'єкти спостережень – це країни, міста, селища, муніципальні утворення. У соціальних дослідженнях об'єктам спостережень, як правило, виступають анкетні дані респондентів. Варто зазначити, що в деяких випадках, досліджуване соціально-економічне явище спостерігається як єдине і неподільне ціле та не може бути представлено сукупністю об'єктів спостережень. У цих випадках дослідник може мати множину спостережень станів соціально-економічного явища, визначених уведенням часового розподілу проведення спостережень (години, дні, місяці, роки) [3]. Колонки матриці даних відповідають ознакам, які характеризують соціально-економічне явище і разом утворюють профіль опису явища. Набір ознак, як правило, включає в себе найбільш суттєві та найбільш якісно вимірювані в рамках конкретного дослідження характеристики об'єктів. Так, наприклад, кожне з муніципальних утворень може бути схарактеризовано такими ознаками, як кількість населення, чисельність зайнятого і безробітного населення, кількість підприємств прибуткових та збиткових, площа території, кількість населених пунктів тощо, тобто кожна ознака з досліджуваного набору ознак задається у вигляді статистичного ряду. Якщо рядки матриці являють собою множину об'єктів, то це ряди даних, які характеризують зміну явища в просторовому розподілі, якщо рядки матриці відповідають часовому розподілу, то ознаки задані у вигляді динамічного ряду. Таким чином, як дані соціально-економічного моніторингу, що підлягають аналізу, можуть виступати різного роду перерізи багатовимірних даних тривимірного простору з координатами «об'єкти – ознаки – час» при фіксованому часі [3]. При формуванні матриці початкових даних варто виділити найбільш важливі моменти. У нашому дослідженні матриця даних – це початковий інформаційний масив даних соціально-економічного моніторингу. Він формується заданням списку ознак і списку об'єктів, ці списки повинні визначатися у чіткій відповідності до мети дослідження, оскільки кінцеві результати

багато в чому визначаються якраз на етапі відбору ознак та об'єктів, що характеризують стан соціально-економічного розвитку регіонів. Як правило, в списках ураховують тільки ті ознаки й об'єкти, які містять у собі найбільш змістовну і суттєву інформацію про досліджувані явища або системи.

При побудові інтегральних показників завжди виникає проблема вибору вимірювальної шкали для ознак дослідження. Для її розв'язання використовуються певні перетворення змінних з метою: ослаблення впливу екстремальних значень ознак на результати розрахунків, компенсації впливу можливих помилок у початкових даних, забезпечення можливостей зіставлення змін значень ознак на різних ділянках вимірювальної шкали. Перш за все потрібно навести ознаки, значення яких вимірюються якісно до числового виду. Наприклад, якісні оцінки «позитивний», «негативний», «без змін» можна представити числами +1, -1, 0 відповідно. Потім потрібно привести всі ознаки до єдиної цільової функції, тобто застосувати до кожної з ознак таке перетворення, в результаті якого найменше значення перетвореного показника буде відповідати найгіршому значенню інтегрального показника, а максимальне – найкращому. Вибір конкретного уніфікуючого перетворення залежить від того, до якого із трьох основних типів належить показник [8–10]:

- у випадку, якщо початковий показник x пов'язаний з відповідним інтегральним показником монотонно зростаючою залежністю, тобто чим більше значення x , тим вище значення інтегрального показника, значення уніфікуючої змінної x^* залишається без змін;
- якщо початковий показник x пов'язаний з відповідним інтегральним показником монотонно спадною залежністю, тобто чим більше значення x_{max} , тим менше значення інтегрального показника, то значення уніфікуючої змінної x^* визначається за формулою

$$x^* = x_{max} - x, \quad (2)$$

- якщо початковий показник x пов'язаний з відповідним інтегральним показником немонотонною залежністю, тобто між мінімальним x_{min} і максимальним x_{max} значенням показника x існує певне оптимальне значення x_{opt} , при якому отримується найкраще значення інтегрального показника, то значення уніфікуючої змінної x^* визначається за формулою:

$$x^* = x_{opt} - |x_{opt} - x|. \quad (3)$$

У процесі застосування описаного способу уніфікації вимірювальної шкали може виникнути необхідність проведення корегування в ситуаціях, коли у множині початкових значень ознак об'єкта є значні відхилення – значення, що значно відрізняються від основної маси. У цьому випадку для усунення недоліку використовуємо положення статистичної теорії похибок, згідно з якими, значні відхилення вважаються грубими помилками і виключаються з подальшого аналізу або замінюються середніми значеннями.

Як правило, ознаки, відібрані для опису соціально-економічного явища, мають різну розмірність, а тому й різну масштабність. З метою забезпечення можливості зіставлення ознак об'єкта та уникнення впливу їх розмірності матрицю початкових даних X зазвичай трансформують (нормують), вводячи єдиний для всіх ознак масштаб. Найбільш розповсюджені способи отримання нормованої матриці даних Z_{ij} – центрування (4) або приведення до стандартної форми (5)

$$Z_{ij} = x_{ij} - \bar{x}_j, \quad (4)$$

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j \bar{x}_j}, \quad (5)$$

де x_{ij} – значення j -ї ознаки i -го об'єкта; \bar{x}_j – середнє арифметичне значення j -ї ознаки; s_j – середньоквадратичне відхилення j -ї ознаки (дисперсія j -ї ознаки) [3].

У результаті стандартизації показників отримуємо матрицю розміром $m \times n$ нормованих значень спостережень. У цьому випадку кожний регіон можна інтерпретувати як деяку точку P_i у n -вимірному векторному просторі, координатами якого є величини Z_{ij} , $j = 1, 2, \dots, n$. Таким

чином, отримується нормована матриця, що складається з векторів, координатами яких є соціально-економічні показники регіонального розвитку. У сукупності векторів нормованої матриці, відповідно до заданих критеріїв, можна вибрати оптимальний вектор, відносно котрого доцільно здійснювати оцінювання стану регіонального розвитку. Якщо базис векторів утворений трьома соціально-економічними показниками, то для подальшого аналізу даних корисно побудувати просторову візуалізацію векторів. У рамках візуального аналізу даних їх просторове представлення обмежується тривимірним випадком, що є суттєвим недоліком у соціально-економічних дослідженнях, які містять у собі, як правило, значно більше показників. Як уже згадувалося раніше, важливою перевагою факторного аналізу є можливість одночасного дослідження необмежено великої кількості показників. Тому подальші дії у визначенні інтегральних показників будуть ґрунтуватися на визначенні факторів, які є відображенням великої кількості показників, що характеризують стан об'єкта.

У факторному аналізі вважається, що ознаки, які входять у досліджуваний набір, не є незалежними, більш того – вони мають корелятивну залежність. У той же час зміна значень різних ознак при переході від одного об'єкта до іншого не може бути описана як функція якогось одного з параметрів. Тобто всі елементи матриці попарних кореляцій, розрахованої для досліджуваного набору ознак, відрізняються по модулю від її граничних значень 0 та 1. Застосування цього положення стосовно набору характеристик, які описують соціально-економічні явища не викликає заперечень з таких причин. Статистичні ряди, отримані при перехресних дослідженнях, так, як і динамічні ряди таких соціально-економічних характеристик, тісно взаємопов'язані між собою і в той же час не є функціями якогось одного з параметрів. Варіація окремих пар або груп взаємопов'язаних ознак може розглядатися як результат впливу декількох складових, зокрема складової, спільної для цієї групи ознак і складових, що описують специфіку кожної з ознак. Наявність кореляцій між двома чи декількома ознаками може бути наслідком однієї з двох причин: або одна з ознак визначає всі інші, або існує певний не включений у досліджуваний набір прихований параметр, що впливає на кореляцію ознак. Такі явно не виражені параметри прийнято називати загальними факторами. Методи факторного аналізу якраз і направлені на виявлення таких факторів. Головне припущення факторного аналізу полягає в наступному: ознаки досліджуваного набору можуть бути описані в термінах невеликої кількості основних внутрішніх параметрів – загальних факторів. Тобто в основі складних взаємозв'язків між ознаками міститься більш проста, прихована за зовнішніми проявами схема, яка відображає найбільш характерні та часто повторювані взаємозв'язки [3].

Вважаємо, що кожна з ознак Z_j , котрі входять у досліджуваний набір, може бути представлена як функція невеликої кількості загальних факторів F_1, F_2, \dots, F_m і характерного фактора U_j

$$Z_j = f(F_1, F_2, \dots, F_m, U_j). \quad (6)$$

Методи факторного аналізу переважно ґрунтуються на припущенні про лінійність взаємозв'язків:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jp}F_p + \dots + a_{jm}F_m + d_jU_j. \quad (7)$$

Коефіцієнти a_{jm} називаються факторними навантаженнями і характеризують значимість кожного з факторів для опису j -ї ознаки. Припущення стосовно лінійності взаємозв'язків у факторних моделях не є очевидним і може викликати заперечення з причини того, що основні параметри, які визначають соціально-економічне явище, впливають і взаємодіють один з одним значно складнішим чином. У цьому випадку варто зазначити, що модель факторного аналізу тільки в першому наближенні відображає реальні процеси, тому завжди потрібно оцінювати, наскільки вона адекватна й ефективно описує соціально-економічні явища [3]. Також варто зазначити, що припущення лінійності значно спрощує програмну реалізацію процедур визначення показників соціально-економічного розвитку.

Розглянемо тепер модель для представлення початкових параметрів через фактори, що відповідає методу головних компонент

$$Z_j = \sum_{p=1}^n a_{jp} F_p. \quad (8)$$

Відмінність такої моделі від моделі (7) полягає в тому, що в цій моделі число факторів приймається рівним числу початкових параметрів n . Тому в ній немає ніяких характерних факторів, і, по суті, ми маємо стандартну систему перетворення одних параметрів на інші. Оскільки число факторів дорівнює числу вхідних параметрів, задача знаходження потрібного перетворення також вирішується однозначно. Запишемо вираз (8) у векторному вигляді

$$\mathbf{Z} = \mathbf{A}\mathbf{F}, \tag{9}$$

де $\mathbf{F}=(F_1, F_2, \dots, F_n)^T$ – центрований випадковий вектор-стовпець некорельованих головних компонент; $\mathbf{Z}=(Z_1, Z_2, \dots, Z_n)^T$ – центрований випадковий вектор-стовпець початкових ознак; $\mathbf{A}=(a_{ij})$ – не випадкова матриця факторних навантажень випадкових величин Z_i на компоненти F_j ($i=1, 2, \dots, n; j=1, 2, \dots, n$).

Нехай $\mathbf{\Omega}=\mathbf{M}(\mathbf{Z}\mathbf{Z}^T)$ коваріаційна матриця вектора \mathbf{Z} . Будучи симетричною і позитивно визначеною, вона має n додатних власних значень $\lambda_1, \lambda_2, \dots, \lambda_n$. Припустимо, що $\lambda_1 > \lambda_2 > \dots > \lambda_n$. Позначимо

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}. \tag{10}$$

Нехай $\mathbf{v}_j=(v_{1j}, v_{2j}, \dots, v_{nj})^T$ – нормовані власні вектори-стовпчики матриці $\mathbf{\Omega}$, що відповідають власним значенням λ_j ($j=1, 2, \dots, n$). Тоді для всіх $j=1, 2, \dots, n$ справедливі такі рівності:

$$\det |\mathbf{\Omega} - \lambda_j \mathbf{I}| = 0, \tag{11}$$

де \mathbf{I} – одинична матриця n -го порядку;

$$\mathbf{\Omega} \mathbf{v}_j = \lambda_j \mathbf{v}_j, \tag{12}$$

$$\mathbf{v}_p^T \mathbf{v}_j = \sum_{i=1}^n v_{ip} v_{ij} = \delta_{pj} = \begin{cases} 1, & p = j \\ 0, & p \neq j \end{cases} \tag{13}$$

Уведемо тензор $\mathbf{V}=(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$. Оскільки при врахуванні формул (12) і (13)

$$\mathbf{v}_j^T \mathbf{\Omega} \mathbf{v}_p = \lambda_j \mathbf{v}_j^T \mathbf{v}_p = \begin{cases} \lambda_j, & p = j \\ 0 & p \neq j \end{cases} \tag{14}$$

то

$$\mathbf{V}^T \mathbf{\Omega} \mathbf{V} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} = \mathbf{\Lambda}. \tag{15}$$

Припустимо, що

$$\mathbf{F}' = \mathbf{V}^T \mathbf{Z}, \tag{16}$$

при цьому оскільки

$$\mathbf{M}\mathbf{F}' = \mathbf{M}(\mathbf{V}^T \mathbf{Z}) = \mathbf{V}^T \mathbf{M} \mathbf{Z}, \quad (17)$$

то \mathbf{F}' – центрований вектор, а оскільки

$$\mathbf{M}(\mathbf{F}' \mathbf{F}'^T) = \mathbf{M}(\mathbf{V}^T \mathbf{Z} \mathbf{Z}^T \mathbf{V}) = \mathbf{V}^T \mathbf{M}(\mathbf{Z} \mathbf{Z}^T) \mathbf{V} = \mathbf{V}^T \mathbf{\Omega} \mathbf{V}, \quad (18)$$

Внаслідок формули (15) компоненти вектора \mathbf{F}' некорельовані, а дисперсія факторів дорівнює

$$\mathbf{D} F'_j = \lambda_j \quad (j=1, 2, \dots, n), \quad (19)$$

тобто \mathbf{F}' є вектором головних компонент \mathbf{F} , який розраховується відповідно до формули (16) таким чином:

$$F_j = \sum_{i=1}^n v_{ij} Z_i, \quad j=(1, 2, \dots, n) \quad (20)$$

Знайдемо матрицю факторних навантажень \mathbf{A} . Використовуючи ортогональність матриці \mathbf{V} та рівняння (16), отримаємо

$$\mathbf{V} \mathbf{F} = \mathbf{V} \mathbf{V}^T \mathbf{Z} = \mathbf{V} \mathbf{V}^{-1} \mathbf{Z} = \mathbf{Z}, \quad (21)$$

ураховуючи формулу (9), вираз (21) можна записати як [3]

$$\mathbf{A} = \mathbf{F}. \quad (22)$$

Реально для аналізу використовують $n' < n$ перших головних компонент, якими вичерпується не менше 70% початкових випадкових величин. Досить легко можна довести [15], що за допомогою компонент F_1, F_2, \dots, F_n досягається найкращий, згідно з методом найменших квадратів, прогноз величин Z_1, Z_2, \dots, Z_n серед всіх прогнозів, які можна побудувати за допомогою n' лінійних комбінацій набору факторів із n довільних величин, при цьому відносна похибка прогнозу на практиці становить $n'/n \approx 10 \div 20\%$ і визначається за формулою

$$\delta = \frac{\sum_{i=n'+1}^n \lambda_i}{\sum_{i=1}^n \lambda_i}. \quad (23)$$

Оскільки в результаті факторного аналізу отримані ознаки мають відносну незалежність, то для розрахунку інтегрального показника доцільно використати їх лінійну комбінацію. При цьому виникає проблема оцінки вагових коефіцієнтів знайдених факторів. Як зазначалося раніше, експертно-статистичний варіант є найбільш придатним для нашого випадку, тому розглянемо його застосування для визначення вагових коефіцієнтів. Якщо оцінка дається декількома експертами, необхідно застосувати один з методів групових експертних оцінок [16]. При цьому кожний експерт може надати або безпосередні оцінки факторів, або результати їх попарних порівнянь. Як наслідок, отримується набір безрозмірних коефіцієнтів $v_i = 1, 2, \dots, n$ (n – число факторів), значення яких знаходяться в межах від 0 до 1.

Чим вищі значення безрозмірних оцінок, тим кращі показники моніторингу (по будь-яких об'єктах). Потім знаходиться вага факторів (по суті, в соціально-економічних дослідженнях регіонального розвитку зміна значення конкретного фактора визначає вагу його впливу на інтегральний показник), що відображає зміну його значень. Чим більша різниця в значенні конкретного фактора, тим більший його вплив на зміну стану СЕС і, як наслідок, на значення оцінки соціально-економічного розвитку. При автоматизованих обчисленнях у як такі вагові коефіцієнти можна $z_i = \sqrt{DF_i}$, використати середньоквадратичні відхилення факторів [3]

$$i=1, 2, \dots, n \quad (24)$$

або при врахуванні формули (19) отримаємо

$$z_i = \sqrt{\lambda_i}, \quad i=1, 2, \dots, n. \quad (25)$$

Тобто для визначення вагових коефіцієнтів можна використовувати розраховані раніше власні числа коваріаційної матриці початкових показників. Чим більша різниця в значеннях об'єктів по фактору, тим більша вага цього фактора. Таким чином, фактори, стосовно яких значення об'єктів суттєво відрізняються, вважаються найбільш важливими. Узагальнену вагу факторів (експертно-статистичну), яка б враховувала як думку експерта, так і різницю в значеннях по цьому фактору, можна отримати, розрахувавши середнє арифметичне цих двох оцінок;

$$w_i = (v_i + z_i)/2, \quad i=1, 2, \dots, n. \quad (26)$$

Таким чином, інтегральний показник розраховується як середньозважена сума факторів з ваговими коефіцієнтами w_i :

$$W_j = \sum_{i=1}^n w_i p_{ij}, \quad j=1, 2, \dots, n, \quad (27)$$

де n – кількість факторів; N – кількість об'єктів; p_{ij} – значення i -го фактора для j -го об'єкта.

Кращим є об'єкт з більшим значенням інтегрального показника. Для автоматизації експертних оцінок у рамках розрахункового механізму інтегральних показників доцільно використовувати експертну систему (ЕС) продукційного типу. У нашому випадку це програмно реалізований механізм надання експертного висновку шляхом використання створеної бази знань у сфері соціально-економічного розвитку регіонів, основою якої є, представлені у вигляді продукційних правил, досвід і знання фахівців-експертів в соціально-економічних дослідженнях. Наведений механізм визначення інтегральних показників розвитку регіонів формалізує розрахункові процедури і дозволяє розробити програмний алгоритм інтелектуальної обробки даних соціально-економічного моніторингу. Блок-схему алгоритму інтелектуальної обробки даних наведено на рис. 1. Початковий етап алгоритму характеризується введенням даних соціально-економічного моніторингу регіону. Подальші дії в рамках наведеного алгоритму пов'язані із програмно реалізованими процедурами визначення N факторів та їх змістовного наповнення (перелік ознак, які входять до складу фактора). На основі кореляційної матриці проводиться також інтелектуальний аналіз даних на предмет виявлення латентних ознак, котрі можуть бути підґрунтям для встановлення нових **закономірностей** процесів соціально-економічного розвитку. Важливим **моментом** інтелектуалізації обробки даних соціально-економічного моніторингу є використання

експертної системи визначення ваги факторів. Функціонування експертної системи базується на використанні бази знань продукційного типу, вона здатна генерувати відповіді, які за своєю якістю та надійністю не поступаються відповідям експертів. Завершальний етап алгоритму – це розрахунок інтегральних показників і візуалізація результатів інтелектуальної обробки даних.

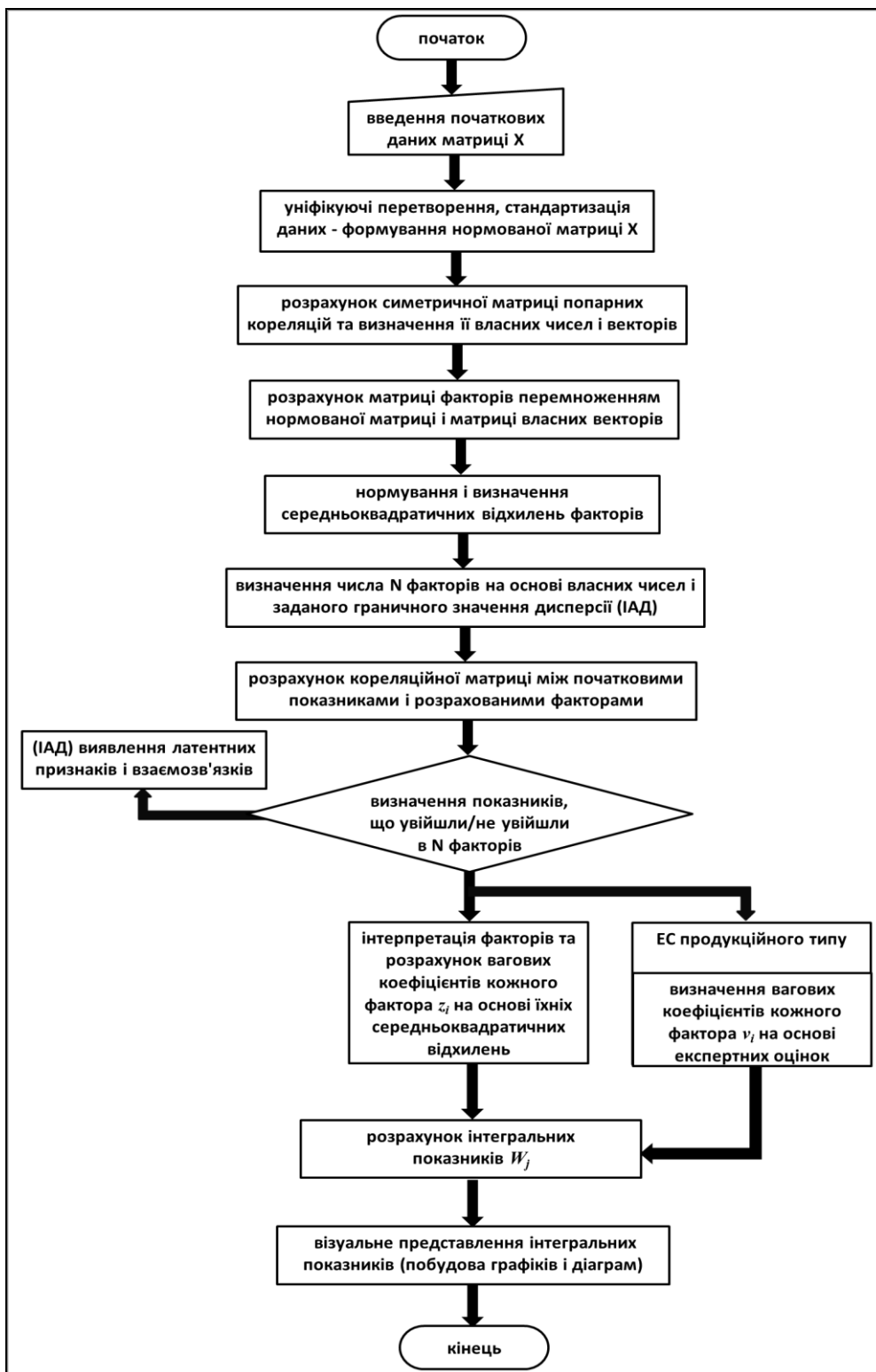


Рис. 1. Блок-схема алгоритму визначення інтегральних показників соціально-економічного розвитку регіонів

Висновки. Характерною ознакою засобів Data Mining є той факт, що вони на основі наявних даних здатні самостійно продукувати моделі, які дозволяють якісно оцінити ступінь впливу різних досліджуваних факторів на задану властивість або, іншими словами, здатні проводити інтелектуальний аналіз даних. Крім цього, засоби Data Mining дозволяють продукувати нові гіпотези про характер невідомих, але реально існуючих відношень у даних. У зв'язку із цим, у роботі розглянуто метод інтелектуального аналізу даних соціально-економічного моніторингу, що дозволяє визначати інтегральні показники регіонального розвитку. Отримані інтегральні показники можна використовувати, наприклад, для виявлення латентних ознак, порівняння і ранжування об'єктів або виявлення ознак, які обумовлюють соціально-економічний розвиток (регресію) окремих регіонів і т.п. Наведена схема побудови інтегральних показників має такі переваги: єдиний підхід до проведення інтелектуального аналізу даних; інтеграція в загальний процес технології виявлення знань у базах даних; оперативність побудови інтегральних показників та ін.

Головними перевагами цього методу побудови інтегральних показників є: використання всієї множини початкових даних, що виключає можливість викривлення змісту регіональної соціально-економічної моделі; забезпечення можливості роботи з великими масивами даних соціально-економічного моніторингу, що є характерним для інтелектуального аналізу даних; урахування знання і досвіду експертів при побудові єдиного інтегрального показника.

ЛІТЕРАТУРА:

1. Иберла, К. Факторный анализ / К. Иберла. – М.: Статистика, 1980. – 398 с.
2. Факторный, дискриминантный и кластерный анализ, пер. с англ. / Дж. Ким., Ч. У. Мьюллер, У. Р. Клекка. – М.: Финансы и статистика, 1989. – 215 с.
3. Жуковская, В. М. Факторный анализ в социально-экономических исследованиях / В. М. Жуковская, И. Б. Мучник. – М.: Статистика, 1976. – 152 с.
4. Біла, С. О. Стратегії розвитку регіонів: шляхи забезпечення дієвості / С. О. Біла, О. В. Шевченко, М. О. Кушнір та ін.; ред. С. О. Біла – К.: НІСД, 2011. – 88 с.
5. Моніторингові оцінювання складних соціально-економічних явищ розвитку регіону / [за наук. ред. Я. О. Побурка]. – Львів : НАН України. Ін-т регіональних досліджень, 2006. – 306 с.
6. Кастельс, М. Информационная эпоха: экономика, общество и культура [пер. с англ. под науч. ред. О. И. Шкаратана] / М. Кастельс. – М.: ГУ ВШЭ, 2000. – 608 с.
7. Миркин, Б. Г. Анализ качественных признаков и структур / Б. Г. Миркин. – М.: Статистика, 1980. – 319 с.
8. Розин, Б. Б. Конструирование экономико-статистических моделей с заданными свойствами / Б. Б. Розин, М. А. Ягольницер. – Новосибирск: Наука, 1981. – 176 с.
9. Розин, Б. Б. Статистическое моделирование экономических показателей / Б. Б. Розин. – Новосибирск: Наука, 1976. – 136 с.
10. Айвазян, С. А. Разработка и анализ интегральных индикаторов качества жизни населения Самарской области / С. А. Айвазян. – М.: ЦЭМИ РАН, 2005. – 124 с.
11. Арсеньев, С. Б. Использование технологии анализа данных в интеллектуальных информационных системах / С. Б. Арсеньев, В. Б. Бритков, Н. А. Маленкова // Управление информационными потоками: сб. тр. Института системного анализа РАН. – М.: Эдиториал УРСС, 2002. – С. 47-68.
12. Барсемян, А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсемян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб.: БХВ-Петербург, 2004. – 336 с.
13. Калинина, В. Н. Введение в многомерный статистический анализ: учеб. пособ. / В. Н. Калинина, В. И. Соловьев. – М.: ГУУ, 2003. – 66 с.
14. Баранов, С. В. Новые методики и результаты исследования межрегиональной дифференциации на основе метода главных компонент / С. В. Баранов, Т. П. Скуфьина // Вестник МГТУ, том 11. – 2008. – № 2. – С 201–210.
15. Айвазян, С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022 с.
16. Смородинский, С. С. Методы анализа и принятия управленческих решений: учеб. пособ. / С. С. Смородинский, Н. В. Батин. – Мн.: БКУУ, 1999. – 101 с.

УДК 004.04; 681.3; 332.1

Пурський Олег Іванович, д. ф.-м. н., професор кафедри економічної кібернетики та інформаційних систем. Київський національний торговельно-економічний університет. **Мороз Ірина Олегівна**, аспірантка кафедри економічної кібернетики та інформаційних систем. Київський національний торговельно-економічний університет. **Використання факторного аналізу в системі інтелектуальної обробки даних моніторингу соціально-економічного розвитку регіонів.** Наведено результати застосування факторного аналізу з метою розроблення механізму інтелектуальної обробки даних соціально-економічного моніторингу регіонального розвитку. Розрахунковий механізм ґрунтується на застосуванні методу головних компонент. Інтегральний показник соціально-економічного розвитку розраховується як середньозважена сума факторів з відповідними ваговими коефіцієнтами.

Ключові слова: соціально-економічний розвиток регіону, факторний аналіз, метод головних компонент, інтегральний показник розвитку.

УДК 004.04; 681.3; 332.1

Пурский Олег Иванович, д. ф.-м. н., профессор кафедры экономической кибернетики и информационных систем. Киевский национальный торгово-экономический университет. **Мороз Ирина Олеговна**, аспирантка кафедры экономической кибернетики и информационных систем. Киевский национальный торгово-экономический университет. **Использование факторного анализа в системе интеллектуальной обработки данных мониторинга социально-экономического развития регионов.** Представлены результаты применения факторного анализа для разработки механизма интеллектуальной обработки данных социально-экономического мониторинга регионального развития. Расчетный механизм базируется на использовании метода главных компонент. Интегральный показатель социально-экономического развития рассчитывается как средневзвешенная сумма факторов с соответствующими весовыми коэффициентами.

Ключевые слова: социально-экономическое развитие региона, факторный анализ, метод главных компонент, интегральный показатель развития.

UDC 004.04; 681.3; 332.1

Pursky Oleg I., doctor of Sciences (Physics&Mathematics), Professor of Department of Economic cybernetics and information systems. Kiev National University of Trade and Economic. **Moroz Irina O.**, postgraduate Student of Department of Economic cybernetics and information systems. Kiev National University of Trade and Economic. **The use of factor analysis in system of intellectual data processing of monitoring of social and economic development of regions.** In this article the results of factor analysis application for working out of the mechanism of intellectual data processing of social and economic monitoring of regional development are presented. The calculation mechanism is based on the use of main component method. The integrated indicator of social and economic development is defined as the sum of factors with corresponding weight factors.

Keywords: social and economic development of region, factorial analysis, main component method, integrated indicator of development.