

БАГАТОВИМІРНИЙ АНАЛІЗ СОЦІАЛЬНО-ЕКОНОМІЧНОГО РОЗВИТКУ РЕГІОНІВ УКРАЇНИ

На сьогодні ключовим питанням вітчизняної регіональної політики є подолання диспропорцій розвитку територій. Диспропорції в економічному і соціальному аспектах є основним індикатором послаблення результативності й ефективності функціонування суб'єктів державного управління різного рівня. Аналіз тенденцій та дослідження факторів, що впливають на соціально-економічний розвиток (СЕР) регіонів доцільно проводити на основі економіко-математичного моделювання. Даний інструментарій дає змогу оцінити вплив факторів на економічне та соціальне становище територій для успішного управління їх розвитком.

Об'єктом дослідження є регіони (області) України за виключенням м. Севастополя та АР Крим. Предметом дослідження виступають методи багатовимірного статистичного аналізу: факторний аналіз, метод таксономії та кластерний аналіз. Період дослідження – з 2002 по 2017 рр. У дослідженні поставлені такі цілі: ідентифікація положення кожного регіону за показниками СЕР і визначення перспектив подальшого СЕР регіонів.

На першому етапі дослідження було сформовано множину вихідних даних, яка містила 40 показників, що характеризують соціальний розвиток регіонів та 17 показників, що характеризують економічний розвиток. За допомогою методів факторного та кореляційного аналізу з цієї множини було відібрано 15 змістовних, інформативних показників, які не корелюють між собою, з них 10 – соціальні, а 5 – економічні.

На другому етапі дослідження на основі відібраних показників було побудовано індекси СЕР регіонів за допомогою зваженого методу таксономії. Для визначення вагових коефіцієнтів (ВК) було використано метод головних компонент, перевагою якого є те, що він, не зважаючи на свою простоту, дозволяє отримати обґрунтовані значення ВК. Розраховані значення ВК показують, що вплив показників соціального блоку на величину інтегрального індексу складає понад 60%, а економічного – близько 40%. Розраховані індекси СЕР дозволили провести ранжування регіонів та прослідкувати зміну їх положення відносно інших регіонів протягом усього досліджуваного періоду.

На третьому етапі було проведено групування регіонів у багатовимірному просторі ознак за допомогою кластерного аналізу з метою визначення спільних рис розвитку територій. Серед множини

методів кластерного аналізу було обрано ітеративні алгоритми. Вони дають змогу декілька разів передивитися вихідні дані та компенсувати таким чином наслідки невдалого початкового розбиття об'єктів на кластери. Окрім цього, більшість цих алгоритмів не припускають ситуації перекриття кластерів. Недоліком є відсутність можливості довести, що знайдений розподіл об'єктів є оптимальним, а не субоптимальним рішенням, а також наявність обов'язкового апріорного визначення кількості кластерів [2]. Більшість ітеративних алгоритмів містять наступні кроки: 1) обирається початковий варіант групування об'єктів на задану апріорі кількість кластерів; 2) обчислюються центри тяжіння кожного кластера; 3) кожен об'єкт переміщується в кластер з найближчим центром тяжіння; 4) обчислюються нові центри тяжіння; 5) кластери не замінюються на нові, доки не будуть розглянуті всі об'єкти. Кроки 3-4 повторюються до тих пір, поки не припинять змінюватися кластери.

До найбільш уживаних ітеративних алгоритмів відносяться алгоритм k-середніх (k-means), алгоритм PAM (partitioning around medoids або k-medoids), алгоритм CLARA (Clustering Large Applications).

В алгоритмі k-середніх координати центрів тяжіння визначаються як середні значення відповідних координат об'єктів, що містяться в кластері. Перевагами цього алгоритму є відносна простота використання та швидкість роботи. Але він занадто чутливий до викидів, які можуть впливати на середнє. Можливим вирішенням цієї проблеми є використання алгоритму PAM. Його особливістю є те, що у якості центрів тяжіння обираються саме об'єкти кластеризації, а не обчислюється штучний центроїд.

Алгоритм PAM складається з двох етапів.

Етап 1. Формується вихідна множина медоїдів S на основі припущення, що медоїд – це об'єкт, для якого сума відстаней до усіх інших об'єктів кластеру є мінімальною. Припустимо, що об'єкт $i \in U$ – кандидат на включення до множини медоїдів S . Тоді для об'єкта $j \in (U - \{i\})$ обчислюється D_j – відстань між ним та найближчим об'єктом в множині S . Наявність ситуації, коли $D_j > d(i, j)$ означає сприятливість включення об'єкту i до множини медоїдів. Далі для кожного j обчислюємо $C_{ji} = \max\{D_j - d(j, i), 0\}$. Сукупний вигравш від включення об'єкту i до множини медоїдів визначається за формулою:

$g_i = \sum_{j \in U} C_{ji}$. До множини медоїдів включається об'єкт, що має максимальне значення g_i . Таким чином обираються k кластерних центрів та формується початкова множина медоїдів U .

Етап 2. Метою цього етапу є покращення якості кластеризації шляхом змін у складі множини медоїдів. Для кожної пари об'єктів i та h , таких що $i \in S$, $h \in U$ обчислюємо показник T_{ih} - сукупний ефект від їх перестановки місцями, в результаті якої отримуємо $i \in U$, $h \in S$: $T_{ih} = \sum \{K_{jih} | j \in U\}$, де K_{jih} - внесок кожного об'єкту $j \in (U - \{h\})$ у сукупний ефект від перестановки місцями об'єктів i та h . K_{jih} обчислюються через показники парних відстаней між об'єктами. Якщо усі отримані T_{ih} є від'ємними, для здійснення перестановки обирається така пара об'єктів (i, h) , яка має мінімальне значення цього показника. Алгоритм припиняється, коли $\min(T_{ih}) > 0$.

Для оцінки ефективності кластеризації виділяють два типи показників: зовнішні показники, які використовують додаткову (зовнішню) інформацію про справжній розподіл об'єктів по класах та внутрішні показники, що використовують тільки інформацію про саму кластеризацію [1]. Одним з таких методів є Гар-статистика. Її застосування є ефективним методом оцінки оптимального числа кластерів у разі незалежних (або слабо пов'язаних) ознак. Метод Гар використовує статистику розбіжностей.

Розглянемо результати групування регіонів за допомогою алгоритму РАМ. Для визначення оптимального числа кластерів скористалися індексом оцінки силуету. Оскільки дана оцінка є чутливою до викидів, то м. Київ було виключено із розгляду. Цілком очевидно, що Київ буде утворювати окремий кластер, тому після отримання оптимального числа кластерів за допомогою індексу оцінки силуету число кластерів необхідно збільшити на одиницю.

Найчастіше оптимальне число кластерів для досліджуваного періоду оцінювалось у межах двох-трьох за виключенням 2011 р. (4), 2013 р. (5) та 2015 р. (4). Було вирішено використовувати лише 3 класи для проведення кластерного аналізу, що дозволяє досить чітко розділити території на регіони з низьким, середнім та високим ступенем розвитку. Також, розрахунки інтегральних показників показують, що у 2015 – 2017 рр. соціально-економічне положення Донецької та Луганської області значно відрізняється від інших регіонів, тому для цього періоду було виділено окремий клас для проблемних територій.

Результати кластерного аналізу показують, що як і очікувалося м. Київ утворив окремий кластер з високим рівнем соціально-економічного розвитку. Донецька та Луганська області у 2015-2017 рр. також утворюють окремий кластер. Також із результатів кластерного аналізу видно, що більшість областей протягом всього досліджуваного періоду перебувають на одному й тому самому рівні розвитку. Так Вінницька, Волинська, Житомирська,

Кіровоградська, Рівненська, Сумська, Тернопільська, Херсонська, Хмельницька та Чернігівська області постійно знаходяться в групі територій з низьким СЕР. Ранжування цих регіонів показує, що вони не підіймалися вище 10 сходинки у рейтингу регіонів. Такі регіони, як Дніпропетровська, Запорізька, Київська, Одеська, та Харківська області кожного року потрапляли до груп з середнім рівнем СЕР (за виключенням Запорізької у 2014 році). У 2012 – 2014 рр. Чернівецька область перейшла з класу територій з низьким розвитком до класу з середнім розвитком, подібна ситуація притаманна і для Закарпатської області у 2014 – 2017 рр.

Результати розрахунків показують, що до 2008 р. всі регіони демонстрували зростання інтегральних індексів СЕР. Внаслідок економічної кризи 2008 р. темп СЕР у багатьох регіонів сповільнився. У деяких регіонів спостерігалось погіршення СЕР (Волинська, Дніпропетровська, Закарпатська, Запорізька, Луганська, Одеська, Рівненська, Харківська та Черкаська області). У 2009 р. у всіх регіонах відновився сталий розвиток. Приріст інтегральних індексів зберігався до 2013 р., за підсумками якого індекси СЕР значно зменшилися у Дніпропетровської, Запорізької, Івано-Франківської та Харківської областей.

У 2013 році у цих областях спостерігалось значне міграційне скорочення населення, скорочення показників росту наявного реального доходу населення, а також зниженням обсягів експорту товарів та послуг.

Внаслідок активних бойових дій на території України у 2014 році відбулося значне зниження показників СЕР у всіх регіонів, окрім Вінницької області та м. Києва. В першу чергу зменшення інтегральних показників у соціальній сфері викликане погіршенням благополуччя населення (зменшенням реальних розмірів пенсії, заробітної плати та наявних доходів населення), в економічній сфері – зменшенням показників експорту товарів та послуг. Показники розвитку медицини, освіти, житлово-комунального господарства не мали значного впливу на інтегральний індекс СЕР регіонів.

Загалом, проведені розрахунки показують, що різниця у СЕР більшості областей не значна. Лише м. Київ суттєво відрізняється у СЕР від інших територій. Значення індексів протягом всього періоду дослідження для регіонів змінювалися у діапазоні 0,13 – 0,32, що говорить про низький загальний СЕР регіонів України.

Список літератури

- [1] П. А. Пархоменко, А. А. Григорьев, Н. А. Астраханцев, "Обзор и экспериментальное сравнение методов кластеризации текстов", *Труды ИСП РАН*, т. 29, вып. 2, 2017 г., стр. 161-200. DOI: 10.15514/ISPRAS-2017-29(2)-6
- [2] Дж.- О. Ким, Ч. У. Мьюллер, У. Р. Клекка *Факторный, дискриминантный и кластерный анализ* - М.: Финансы и статистика, 1989.