

О.М. КОСТЕНКО*(Національний університет біоресурсів і природокористування України, м. Київ, Україна)***В.М. КРАЄВСЬКИЙ***(Національний університет біоресурсів і природокористування України, м. Київ, Україна)*

Кластерний аналіз як метод ідентифікації однорідності об'єктів інформаційно-аналітичної системи управління

В статті досліджено теоретико-методичні засади використання кластерного аналізу як методу обґрунтування класифікаційних ознак об'єктів інформаційно-аналітичної системи управління, що забезпечує удосконалення її параметрів з врахуванням вимог менеджменту. Зокрема, всебічно розглянуто сутність поняття «кластерний аналіз», поглиблено його характеристику як методу багатомірного статистичного дослідження, що передбачає збір інформації про об'єкти інформаційно-аналітичної системи управління та впорядкування їх в однорідні, схожі між собою групи (кластери) за допомогою проведення ієрархічних (агломеративних) чи неієрархічних (ітераційних) процедур. Імплементовано розглянуті теоретико-методичні положення кластерного аналізу в емпіричну площину за допомогою наведеного прикладу визначення із загальної сукупності вітчизняних господарюючих суб'єктів аграрної сфери типового об'єкта, що має узагальнені характеристики. Це дає можливість порівнювати результативні показники власного підприємства з індикаторами знайденого типового аналогу, що дозволяє достовірно оцінювати результати та визначати потенційні резерви своєї діяльності.

Ключові слова: інформація, показник, кластерний аналіз, інформаційно-аналітична система, управління.

А.Н. КОСТЕНКО*(Национальный университет биоресурсов и природопользования Украины, г. Киев, Украина)***В.Н. КРАЕВСКИЙ***(Национальный университет биоресурсов и природопользования Украины, г. Киев, Украина)*

Кластерный анализ как метод идентификации однородности объектов информационно- аналитической системы управления

В статье исследованы теоретико-методические основы использования кластерного анализа как метода обоснования классификационных признаков объектов информационно-аналитической системы управления, что обеспечивает усовершенствование ее параметров с учетом требований менеджмента. В частности, всесторонне рассмотрена сущность понятия «кластерный анализ», углубленно его характеристику как метода многомерного статистического исследования, который предполагает сбор информации об объектах информационно-аналитической системы управления и упорядочение их в однородные, похожие между собой группы (кластеры) посредством проведения иерархических (агломеративных) или неиерархических (итерационных) процедур. Имплементировано рассмотренные теоретико-методические положения кластерного анализа в эмпирическую плоскость с помощью примера определения из общей совокупности отечественных хозяйствующих субъектов аграрной сферы типового объекта, который имеет обобщенные характеристики. Это дает возможность сравнивать результативные показатели собственного предприятия с индикаторами найденного типичного аналога, что позволяет достоверно оценивать результаты и определять потенциальные резервы своей деятельности.

Ключевые слова: информация, показатель, кластерный анализ, информационно-аналитическая система, управление.

Cluster Analysis as Method of Objects Uniformity Identification of Information-Analytical Management System

The article studies the theoretical and methodological foundations of the use of cluster analysis as a method of validation of classification features of objects of information-analytical management system that provides the improvement of its parameters with regard to management requirements. In particular, comprehensively discussed the essence of the concept of 'cluster analysis', supplied its in-depth description as the method of multivariate statistical analysis, involves gathering information about the objects of information-analytical management system and ordering them in uniform, similar to each other groups (clusters) with the help of the hierarchical (agglomerative) or non-hierarchical (iterative) procedures. Implemented the examined theoretical and methodological positions of cluster analysis in the empirical plane using the example of the definition of the model object with generalized characteristics from a common set of domestic economic entities of the agrarian sector. This makes it possible to compare the performance indicators of own productive enterprise with indicators of found typical analog that can reliably evaluate the results and identify potential reserves of the activities.

Keywords: information, indicator, cluster analysis, information-analytical management system.

Постановка проблеми. Об'єктивність дослідження і вирішення проблемних питань в діяльності підприємств визначається обсягом і достовірністю даних інформаційно-аналітичної системи управління їх функціонуванням. На основі такої інформації розробляються обґрунтовані управлінські рішення щодо поліпшення результатів господарювання як окремих організаційних формувань, так і регіонів в цілому. При цьому, рівень виваженості, а відтак, і прийнятності до використання зазначених рішень суттєво зростає у випадках автоматизації процесів роботи з інформаційно-аналітичним забезпеченням, застосування системи новітніх статистичних методів, зокрема, кластерного аналізу та вибору з його допомогою типового підприємства для проведення порівняльних процедур.

Аналіз останніх досліджень і публікацій. Суттєвим науковим доробком в розвиток методик кластеризації відзначилися такі вітчизняні та зарубіжні вчені як: Ф.Ф. Бутинець, І.І. Каракоз, Г.Г. Кірейцев, В.В. Ковальов, Л.І. Кравченко, В.М. Краєвський, І.Д. Лазаришина, Ю.Я. Литвин, Г.В. Савицька, В.К. Савчук, В.І. Самборський, Р.А. Фатхутдінов, М.Г. Чумаченко та інші. В роботах згаданих економістів зазначена тематика наукового пошуку відзначається дотичністю переважно до теоретичної складової, емпірична ж розглянута менше, зокрема, залишаються недостатньо дослідженими проблемні питання щодо призначення алгоритмів кластерного аналізу як інструментів дисперсійного моніторингу об'єктних елементів інформаційно-аналітичної системи управління діяльністю підприємств.

Мета статті є визначення теоретико-методичних засад використання кластерного аналізу як методу

ідентифікації однорідності об'єктів інформаційно-аналітичної системи управління, що забезпечує удосконалення її параметрів з врахуванням вимог менеджменту.

Виклад основного матеріалу. Аналіз, як функція управління, на різних етапах створення і реалізації його якісної інформаційно-аналітичної системи, передбачає наступний інструментарій дослідження ефективності функціонування господарюючого суб'єкта:

1) традиційні логічні прийоми та способи (порівняння, абсолютних, відносних та середніх величин, табличного та графічного оформлення даних, групування, балансовий, евристичні методи вирішення економічних задач на основі інтуїції, досвіду, експертних оцінок тощо);

2) способи детермінованого факторного аналізу (ланцюгової підстановки, абсолютних різниць, відносних різниць, індексний, пропорційного ділення і дольової участі, інтегральний, логарифмічний та інші) і стохастичного факторного аналізу (дисперсійний, регресійний, кореляційний, компонентний, дискримінантний, кластерний, багатомірний математичний факторний аналіз та інші), а також оптимізаційного вирішення економічних задач (лінійне та нелінійне програмування, теорія ігор, теорія масового обслуговування, дослідження операцій та інші);

3) SWOT аналіз;

4) функціонально-вартісний аналіз;

5) маржинальний аналіз та інші.

В наведеному переліку одним з основоположних методів формування зазначеної інформаційно-аналітичної системи управління є кластерний аналіз, оскільки він обумовлює необхідність збору і обробки

інформації про підприємство, а також використовується як метод в підготовці та прийнятті виважених управлінських рішень, зокрема, за рахунок дослідження впливу факторів на результати його діяльності та визначення потенційних резервів.

Кластеризація (від англ. *cluster* – група, пучок, гроно, згусток) – розбиття заданої вибірки об'єктів (ситуацій) на підмножини, що називаються кластерами, так, щоб кожний кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися. Звідси, кластерний аналіз – це багатовимірний статистичний процес, який виконує збір даних, що містять інформацію про вибірку об'єктів, а потім упорядковує останні в однорідні групи (кластери) [2, с. 5; 3, с. 46; 4, с. 3].

Спектр застосування кластерного аналізу дуже широкий: в першу чергу, економіка, крім того, соціологія, філологія, хімія, біологія, медицина, археологія та інші галузі наукових знань. Така універсальність стала передумовою появи великої кількості алгоритмів і підходів (сьогодні їх налічується більше 150), що інколи ускладнюють несуперечливу інтерпретацію кластерного аналізу, проте, розширюють межі його використання.

Основна мета кластерного аналізу – знаходження груп схожих об'єктів у вибірці. Така цільова установка визначає наступні його завдання:

- 1) розробка типології або класифікації;
- 2) визначення концептуальних положень групування об'єктів;
- 3) породження гіпотез на основі дослідження даних;
- 4) перевірка гіпотез для визначення, чи дійсно групи, виділені тим чи іншим способом, присутні в даних [5, с. 10].

Незалежно від виду алгоритму та сфери застосування кластерного аналізу, його методика включає етапи:

- а) формування вибірки для кластеризації;
- б) визначення множини характеристик, за якими будуть оцінюватися об'єкти у вибірці;
- в) обчислення значень тієї чи іншої міри схожості між об'єктами;
- г) застосування одного з методів кластерного аналізу для створення груп об'єктів;
- д) перевірка достовірності результатів кластеризації.

Якщо кластерному аналізу передують факторний аналіз, то вибірка не потребує корегування – вимога до кластеризації (схожість між об'єктами) виконується автоматично самою процедурою факторного моделювання. В іншому випадку вибірку потрібно корегувати.

Об'єднання схожих об'єктів у групи може здійснюватися різними способами. Для цього розроблено такі методи:

- 1) к-середніх (K-means);
- 2) нечіткої кластеризації с-середніх (C-means);
- 3) графові алгоритми кластеризації;
- 4) статистичні алгоритми кластеризації;
- 5) алгоритми сімейства FOREL;

- 6) ієрархічна кластеризація або таксономія;
- 7) нейронна мережа Кохонена;
- 8) ансамбль кластеризаторів;
- 9) алгоритми сімейства KRAB;
- 10) EM-алгоритм;
- 11) метод просіювання та інші [1, с. 12; 9, с. 6].

Вхідними даними для кластерного аналізу є набір об'єктів. В залежності від способів представлення цих об'єктів, розрізняють типи вхідних даних:

- вектор характеристик (кожний об'єкт описується набором власних характеристик; які можуть бути числовими або нечисловими);
- матриця відстаней (кожний об'єкт описується відстанями до всіх інших об'єктів вибірки).

Кластерний аналіз висуває наступні вимоги до даних:

- а) об'єкти не повинні корелювати між собою;
- б) безрозмірність об'єктів;
- в) близькість розподілу об'єктів до нормального;
- г) відповідність об'єктів вимозі стійкості, під якою розуміють відсутність впливу на їх значення випадкових чинників;
- д) однорідність вибірки.

Рішення задач кластеризації принципово неоднозначні, і цьому є декілька підстав:

- не існує найкращого критерія якості кластеризації (розроблено цілий ряд евристичних критеріїв, а також алгоритмів, що не мають чітко вираженого критерія, але здійснюють достатньо обґрунтовану кластеризацію «по побудові», всі вони можуть давати різні результати);
- число кластерів, як правило, невідоме заздалегідь і встановлюється здебільшого суб'єктивно;

- результат кластеризації істотно залежить від метрики, вибір якої, як правило, також суб'єктивний і визначається експертом [6, с. 21; 7, с. 75].

Результатом кластеризації є групи об'єктів, узагальнені за однією чи декількома характеристиками. Однак ці результати можуть бути інтерпретовані по-різному. Так, при оцінці результатів монографічних досліджень, аналіз рекомендується здійснювати ієрархічними методами, наприклад, методом Уорда, згідно з яким всередині кластерів оптимізується мінімальна дисперсія і в результаті створюються кластери приблизно рівних розмірів. В якості міри відмінності між кластерами використовується квадратична евклідова відстань, що сприяє збільшенню контрастності кластерів.

На завершальному етапі аналізу може виникнути питання стійкості знайденого кластерного рішення. По суті, перевірка стійкості кластеризації зводиться до оцінки її достовірності. В таких випадках застосовують емпіричне правило – стійка типологія зберігається при зміні методів кластеризації. Результати ієрархічного кластерного аналізу можна перевірити ітеративним кластерним аналізом (методом к-середніх). Якщо при порівнянні груп та їх складу відмічається співпадання більше, ніж на 70 % (понад 2/3 збігів), то кластерне рішення приймається.

Перевірити адекватність такого управлінського рішення, не вдаючись до допомоги інших видів аналізу, неможливо. Принаймні, в теоретичному плані ця проблема не вирішена. Деякі додаткові методи перевірки стійкості відкидаються з наступних причин:

- 1) кофенетична кореляція – не рекомендується і обмежена у використанні;
- 2) тести значущості (дисперсійний аналіз) – завжди дають значущий результат;
- 3) метод повторних випадкових вибірок – не доводить правильність рішення;
- 4) тести значущості для зовнішніх ознак – придатні тільки для повторних вимірювань;
- 5) методи Монте-Карло – дуже складні і доступні тільки досвідченим математикам [10, с. 22; 11, с. 31].

Імплементуємо розглянуті теоретико-методичні положення кластерного аналізу в емпіричну площину, розглянувши на прикладі визначення із загальної сукупності вітчизняних господарюючих суб'єктів аграрної сфери типового об'єкта, що має узагальнені характеристики. Це дасть можливість порівнювати результативні показники власного підприємства з індикаторами знайденого типового аналогу, що дозволить достовірно оцінити результати своєї діяльності та визначати потенційні резерви.

Із багаточисельних алгоритмів кластеризації нами обрано неієрархічний ітеративний метод к-середніх та ієрархічний агломеративний метод аналізу (перевірка результатів останнього проводилася неформальним методом, а саме повторною кластеризацією по невеликих часткових вибірках при використанні різних метрик відстаней та правил об'єднання (зв'язку): метод Уорда і відстань Евкліда; метод повного об'єднання (зв'язку) і відстань міських кварталів (манхетенська відстань); метод попарного середнього зваженого і відстань Чебишева), а з таких найвідоміших та поширених вітчизняних і зарубіжних статистичних пакетів як SIGAMD, DataScore, STADIA, COMI, ПНП-БИМ, COPPA-2, СИТО, SAS, SPSS, STATISTICA, BMDP, STATGRAPHICS, GENSTAT, S-PLUS та інших розрахунки проводилися в модулі «Кластерний аналіз» автоматизовано-інтегрованого середовища SPSS.

Вихідні дані для аналізу включають 17 показників:

- площа сільськогосподарських угідь;
- середньорічна вартість активів;
- середньооблікова чисельність працівників, зайнятих у сільськогосподарському виробництві всього по підприємству, в тому числі в рослинництві та в тваринництві;
- кількість вироблених видів продукції рослинництва;
- кількість реалізованих видів продукції рослинництва;
- кількість вироблених видів продукції тваринництва;
- кількість реалізованих видів продукції тваринництва;

- доход (виручка) від реалізації продукції всього по підприємству, в тому числі в рослинництві та в тваринництві;

- витрати на виробництво продукції всього по підприємству, в тому числі в рослинництві та в тваринництві;

- витрати на оплату праці всього по підприємству, в тому числі в рослинництві та в тваринництві;

- матеріальні витрати, які увійшли в собівартість продукції всього по підприємству, в тому числі в рослинництві та в тваринництві;

- амортизація основних засобів всього по підприємству, в тому числі в рослинництві та в тваринництві;

- прямі затрати праці на продукцію всього по підприємству, в тому числі в рослинництві та в тваринництві;

- повна собівартість продукції всього по підприємству, в тому числі собівартість продукції рослинництва та собівартість продукції тваринництва;

- прибуток всього по підприємству, в тому числі в рослинництві та в тваринництві;

- рівень рентабельності виробництва всієї продукції, в тому числі продукції рослинництва та продукції тваринництва;

- продуктивність праці всього по підприємству, в тому числі в рослинництві та в тваринництві).

Зазначені показники характеризують діяльність 453 довільно обраних сільськогосподарських підприємств усіх адміністративних одиниць України. Систему наведених індикаторів таких суб'єктів господарювання поділено на 4 групи:

- 1) оцінки ресурсного потенціалу та розміру;
- 2) оцінки спеціалізації та галузевої приналежності;
- 3) оцінки технологій;
- 4) результативності діяльності.

Функціональна залежність між окремими коефіцієнтами була виключена забезпеченням їх співставності, тобто шляхом перерахунку кількісних показників у якісні.

Як зазначалося вище, для вирішення задачі кластерного аналізу потрібний поділ на кластери (за власним бажанням аналітика з можливим підтвердженням графіком агломерації (об'єднання)), який задовольняє відповідний критерій оптимальності. Цей критерій вважається функціоналом, що визначає необхідність кількості групувань. В нашому монографічному дослідженні такою цільовою функцією було обрано ефективність діяльності сільськогосподарських підприємств, відтак, сформовані 3 кластери мають вигляд:

- 1) низький рівень;
- 2) середній рівень;
- 3) високий рівень ефективності.

Типовим ідентифікуємо те підприємство, яке входить до більшої кількості кластерів та до найчисельнішої сукупності підприємств відповідного кластеру як за всіма показниками, так і за окремими їх групами (табл. 1, 2, 3).

Приналежність підприємств до кластерних груп

Об'єкт монографічного дослідження в загальній сукупності підприємств	Кількість кластерів, до яких входить об'єкт дослідження за всіма групами показників
123	2
208	2
247	3
443	3

Джерело: розроблено авторами.

Як свідчать дані табл. 1, тільки чотири суб'єкти господарювання із загальної сукупності сільськогосподарських підприємств за 17 обраними показниками входять більш ніж до 1 кластеру. Для підтвердження достовірності отриманих результатів і усередненості індивідуальних характеристик

відібраних організаційних формувань, а також виділення серед них типового підприємства за визначеними показниками, проведемо повторне групування, а саме двоетапний кластерний аналіз (табл. 2, 3).

Таблиця 2

Кластеризація підприємств за загальною кількістю показників

Групи кластерів	Кількість підприємств, одиниць	Питома вага в загальній сукупності, %
Низький рівень ефективності	428	94,5
Середній рівень ефективності	12	2,6
Високий рівень ефективності	13	2,9
Всього	453	100,0

Джерело: розроблено авторами.

Таблиця 3

Кластеризація підприємств за окремими групами показників

Групи кластерів	Показники функціонування сільськогосподарських підприємств							
	ресурсний потенціал та розмір		спеціалізація та галузевість		оцінка технологічного процесу		результативність діяльності	
	кількість підприємств, одиниць	питома вага в загальній сукупності, %	кількість підприємств, одиниць	питома вага в загальній сукупності, %	кількість підприємств, одиниць	питома вага в загальній сукупності, %	кількість підприємств, одиниць	питома вага в загальній сукупності, %
Низький рівень ефективності	426	94,1	435	96,0	421	93,0	431	95,1
Середній рівень ефективності	19	4,2	11	2,5	9	2,0	10	2,2
Високий рівень ефективності	8	1,7	7	1,5	23	5,0	12	2,7
Всього	453	100,0	453	100,0	453	100,0	453	100,0

Джерело: розроблено авторами.

Згідно даних табл. 2 найбільше сільськогосподарських підприємств, в тому числі два з чотирьох раніше виділених суб'єктів господарювання, входять до 1 кластеру за загальною кількістю показників. Це свідчить про низький рівень

ефективності діяльності організаційних формувань відібраної сукупності.

Проведена кластеризація за окремими групами показників (табл. 3), підтвердила попередній висновок щодо незадовільних результатів функціонування об'єктів монографічного

дослідження та дозволила визначити в загальній їх кількості один господарюючий суб'єкт з чіткими ознаками типізації індивідуальних параметрів. Ним виявився ВП НУБіП України «Великосітинське НДГ імені О.В. Музиченка» Фастівського району Київської області, висновки та пропозиції щодо підвищення ефективності діяльності якого можуть використовуватися іншими сільськогосподарськими підприємствами найчисельнішого їх спільного фракційного об'єднання відповідного кластеру.

Методику кластерного аналізу в аналогічний спосіб слід застосовувати й у внутрішньосистемній діагностиці, зокрема, щодо поглибленої оцінки показників виробничо-фінансової діяльності як власного, так і типового (модельного, тестового) підприємства для подальшого відповідного їх порівняння. Це дозволить достовірно визначити зони ризику, наприклад, менш успішну за рівнем ефективності галузь, що першочергово потребує нових ідей та розробок, вимагає впровадження сучасних, науково обґрунтованих способів та прийомів ведення діяльності.

Отримана інформація доповнить змістовну характеристику даних інформаційно-аналітичної системи про зовнішнє та внутрішнє оточення підприємства, що дозволить приймати виваженіші управлінські рішення з підвищення результативності його функціонування.

Висновки. Проведене дослідження в зазначеному аспекті свідчить, що під час оцінки діяльності підприємства кластерний аналіз, на відміну від інших методів, дає можливість класифікувати об'єкти згаданої інформаційно-аналітичної системи не тільки за однією, а й за декількома ознаками одночасно. Формування кластерів означає структурування елементів вибірки, тобто алгоритми кластеризації необхідні для виявлення структури в даних, яку нелегко знайти при візуальному обстеженні або за допомогою експертних прийомів.

Таким чином, кластерний аналіз дозволяє перетворити великий обсяг різнобічної інформації в упорядкований, компактний вигляд на основі ідентифікації однорідності об'єктів, що сприяє поліпшенню якісних параметрів інформаційно-аналітичної системи управління, зокрема, за рахунок

підвищення рівня наочності, зрозумілості та сприйняття отриманих висновків про результати діяльності підприємства.

4 Список використаних джерел

1. Вуколов В.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL / В.А. Вуколов. – М.: ФОРУМ ИНФРА-М, 2004. – 464 с.
2. Дюран Б. Кластерный анализ / Б. Дюран, П. Оделл. – М.: Финансы и статистика, 1977. – 128 с.
3. Економічний аналіз: навч. посіб. / [М.А. Болюх, В.З. Бурчевський, М.І. Горбаток та ін.]; за ред. М.Г. Чумаченка. – [2-ге вид.]. – К.: КНЕУ, 2003. – 556 с.
4. Мандель И.Д. Кластерный анализ / И.Д. Мандель. – М.: Финансы и статистика, 1981. – 133 с.
5. Олдендерфер М.С. Кластерный анализ / Факторный, дискриминантный и кластерный анализ / М.С. Олдендерфер, Р.К. Блэшфилд. – М.: Финансы и статистика, 1989. – 215 с.
6. Савицкая Г.В. Экономический анализ: учебник / Г.В. Савицкая. – [9-е изд.]. – М.: Новое знание, 2004. – 640 с.
7. Савчук В.К. Аналіз господарської діяльності сільськогосподарських підприємств / В.К. Савчук. – К.: Урожай, 1995. – 328 с.
8. Федорченко Б.С. Кластерний аналіз біоенергетичного потенціалу областей України / Б.С. Федорченко // Облік і фінанси. – 2014. – № 2(64). – С. 173-178.
9. Шуметов В.Г. Кластерный анализ: подход с применением ЭВМ / В.Г. Шуметов, Л.В. Шуметова. – Орел: Орел ГТУ, 2000. – 118 с.
10. Karlsson Ch. Spatial ICT Clusters in Sweden – An Empirical Method to Identify Necessary conditions for existence / Karlsson Ch., Nellander Ch., Paulsson T. – London: Routledge, 2004. – 132 p.
11. Maggioni M. High-Tech Firms and the Dynamics of Innovative Industrial Clusters / M. Maggioni, M. Riggi. – Copenhagen: Regions and Institutions, 2006. – 97 p.