**МОДЕЛЬ ІЄРАРХІЧНОЇ КЛАСТЕРИЗАЦІЇ ЯК ІННОВАЦІЙНА СКЛАДОВА ЕКОНОМІЧНОГО АНАЛІЗУ**

**к.е.н. Ключенко Аделіна Вікторівна,**

**Макойда Богдан Теодозійович**

*Івано-Франківський навчально-науковий*

*інститут менеджменту*

*Тернопільського національного економічного університету,*

*Україна*

Термін «кластерний аналіз» першим використовує Тірон у 1939 році, а Сокал, Снет, Ланс та Вільямс (1963-1967рр.) [1,2] провели перші суттєві дослідження в цій області. Згодом кластерний аналіз отримав значного поширення та диверсифікації, синтезуючи дві основні наукові тенденції, представлені американськими та французькими школами.

За допомогою кластерного аналізу в межах груп об'єктів або форм ведеться пошук ідентифікації класів, груп або кластерів з найбільш подібними елементами всередині одного класу (мінімальна мінливість у класах) і якомога різними, якщо ці елементи належать до різних класів (максимальна мінливість між класами) [3,4]. Звідси випливає, що кластерний аналіз дозволяє досліджувати подібності та відмінності між об'єктами, що належать до певної множини, для того, щоб згрупувати їх у вигляді окремих класів і забезпечити однорідність усередині них. Кожен об'єкт в аналізованому наборі присвоюється одному класу, а набір класів є дискретною, невпорядкованою множиною. Класи або групи, в яких структуруються набори об'єктів називаються кластерами.

 Враховуючи те, що в сучасній економіці інтенсивного розвитку отримує поняття «кластеру» як економічної системи (рекреаційні кластери, будівельні кластери, медичні кластери тощо), то суміжний із ним кластерний метод економічного аналізу набуває особливої актуальності.

Кластерний аналіз передбачає фіксацію форм або об'єктів у кластерах або групах поступово, без попереднього знання кількості класів і відповідно до двох основних критеріїв:

а) Кожен клас повинен бути однорідним, щоб містити об'єкти або форми якомога ближче до характеристик, що розглядаються для класифікації об'єктів;

б) Кожен клас повинен містити класифіковані об'єкти, які максимально відрізняються з точки зору класифікаційних характеристик об'єктів, які присутні в будь-якому з інших класів.

Залежно від характеристик використовуваних моделей, початкових припущень і характеру отриманих результатів, кластерні методи поділяються на кластеризацію за допомогою:

1. *Агрегації;*
2. *Розподілу.*

В даному випадку відомі такі конкретні моделі: простий метод агрегації, комплексний метод агрегації, метод середньої агрегації, метод Ward і т.д.

Втім, у разі аналізу великих обсягів даних, що характеризуються високим ступенем неоднорідності доцільним вбачаємо застосування методу ієрархічної кластеризації з метою оптимальної систематизації, групування та синтезу інформації. Оскільки даний метод передбачає використання концепції відстані, що є ефективними при попередньому аналізі даних та дозволяє більш ефективно організовувати гетерогенні дані, а також полегшувати і більш послідовно отримувати та інтерпретувати інформацію в структурованих системах даних.

Ієрархічна кластеризація є неконтрольованою системою розпізнавання, оскільки процес класифікації починається без відома про кількість класів і складності форм у цих класах. У цьому випадку класи формуються, коли кількість аналізованих форм збільшується, кількість кінцевих класів визначається наприкінці процесу розпізнавання. Деякі неконтрольовані алгоритми класифікації, такі як алгоритми розбиття, включають апріорну фіксацію кількості класів, на яких будуть розділені аналізовані об'єкти. Це не означає, що кількість класів фактично відома, але лише зроблено таке припущення.

Ієрархічний метод передбачає створення декількох кластерних рішень, які називаються ієрархіями кластерів. Головною особливістю є те, що кількість кластерів не відома заздалегідь, а також такий параметр не запропоновано користувачем.

Ієрархії кластерів - це кластерні структури зі змінною кількістю багаторівневих кластерів, які диференціюються за кількістю кластерів, які вони включають та їх агрегацією. Таким чином, з Т-об'єктами ми матимемо T кластерних рішень, кожне рішення містить з кожним кроком все більше кластерів, які відповідно з кожним етапом мають все вищий рівень агрегації. Ієрархія кластера має структуру наступної форми:

рівень 0: $ω\_{1}^{(0)},ω\_{2}^{(0)}, …,ω\_{k\_{0}}^{(0)}$

Рівень 1: $ω\_{1}^{(1)},ω\_{2}^{(1)}, …,ω\_{k\_{1}}^{(1)}$

Рівень 2: $ω\_{1}^{(2)},ω\_{2}^{(2)}, …,ω\_{k\_{2}}^{(2)}$ (1),

 …….…………….....

Рівень Т-1: $ω\_{1}^{(T-1)},ω\_{2}^{(T-1)}, …,ω\_{k\_{T-1}}^{(T-1)}$

де $ K\_{i}$ - число кластерів у кластерній сукупності i-го рівня.

Оскільки кластерне рішення, представлене списком об'єктів, що підлягають класифікації є першим розділом, випливає, що можливе число рішень у структурі кластера, отриманих ієрархічними алгоритмами, буде на 1 менше ніж кількість об'єктів. Це число задається наступним співвідношенням:

$N\_{s}=T-1 $ (2).

Вибір найбільш відповідного кластерного рішення Т-1 базується на цілях, які супроводжуються в аналізі. В умовах сучасної економіки, на такому етапі можуть бути вмонтовані складові алгоритму, які б генерували рішення суміжними із стратегічними цілями розвитку в межах кластерних утворень (наприклад екологоспрямовані складові економічного аналізу в процесі становлення та розвитку рекреаційного кластеру; дотримання соціоорієнтованих пріоритетів при функціонуванні будівельних кластерів тощо).

Розглянемо дві можливі категорії алгоритмів ієрархічної кластеризації:

*• Алгоритми агрегації.* Для цих алгоритмів кількість кластерів у першому розділі дорівнює кількості об'єктів, тобто $ K\_{0}=T$. Також кількість кластерів у розбитті на певному рівні на 1 менше, ніж кількість кластерів у попередньому рівні та на 1 більше ніж число кластерів у наступному рівні, відповідно:

 $ K\_{T-1}=K\_{T-2}-1=K\_{T-3}-1=…=K\_{2}-1=K\_{1}-1=K\_{0}-1$ (3).

• *Алгоритми дезагрегації.* Ці методи по суті складаються з тих самих операцій, що використовуються алгоритмами агрегації, але в зворотному порядку. Таким чином, перший розглянутий розділ представлений єдиним кластером, що містить всі об'єкти, другий розділ, що складається з двох кластерів і так далі.

Ієрархічні методи кластеризації являються евристичними методами («евристика» виводиться правилами на основі теоретичних міркувань або статичних спостережень), які включають класифікаційні процедури, розроблені на основі певного інтуїтивного способу вирішення певної (евристичної) задачі.

Серед цих методів можна згадати простий метод агрегації, метод повної агрегації, метод середньої агрегації, метод центроїда або метод Уорда.

Відстань Уорда між двома кластерами вимірює кумулятивну внутрішньоклітинну мінливість, індукована кластеризацією двох кластерів у результуючій кластерній конфігурації. Об'єднуючи два кластери, він призначений для досягнення максимальної однорідності у всіх кластерах, що належать до заданої конфігурації кластерних об'єктів.

Звідси випливає, що відстань Уорда є єдиною, що розглядає мінімізацію внутрішньокластерної мінливості або, іншими словами, максимізацію міжкластерної мінливості, тобто ступеня однорідності кластерів. Слід зазначити, що ступінь однорідності кластера максимізується шляхом мінімізації загальної суми квадратів міжкластерних відхилень.

Якщо$ w\_{12}$ є новим кластером, отриманий шляхом злиття кластерів $ w\_{1}$ та $ w\_{2}$, то суми відстаней між кластерами будуть:

$$SSE\_{w\_{1}}=\sum\_{i=1}^{T\_{w\_{1}}}\left(y\_{i}-\overbar{y\_{w\_{1}}}\right)^{t}\left(y\_{i}-\overbar{y\_{w\_{1}}}\right)$$

$SSE\_{w\_{2}}=\sum\_{i=1}^{T\_{w\_{2}}}\left(y\_{i}-\overbar{y\_{w\_{2}}}\right)^{t}\left(y\_{i}-\overbar{y\_{w\_{2}}}\right)$ (4).

$$SSE\_{w\_{12}}=\sum\_{i=1}^{T\_{w\_{12}}}\left(y\_{i}-\overbar{y\_{w\_{12}}}\right)^{t}\left(y\_{i}-\overbar{y\_{w\_{12}}}\right)$$

Ці два кластери $ w\_{1}$ та $ w\_{2}$ підлягатимуть об’єднанню коли вони мінімізують зростання суми квадратів відхилень, яка визначається як :

$I\_{w\_{12}}=SSE\_{w\_{12}}-(SSE\_{w\_{1}}-SSE\_{w\_{2}})$ (5).

Ключовими етапами ієрархічного кластерного аналізу на шляху класифікації та систематизації об'єктів є:

• вибір характеристик, згідно з якими буде здійснена класифікація;

• вибір типу моделі для оцінки близькості об'єктів;

• встановлення правил формування класу або кластера;

• побудова класів, тобто об’єднання об'єктів у класах;

• перевірка послідовності та значимості класифікації;

• вибір оптимальної кількості кластерів залежно від характеру проблеми класифікації та цілей, що супроводжуються;

• тлумачення значення окремих кластерів.

В результаті, ієрархічний кластерний аналіз створює можливість ідентифікації вихідних даних, груп, класів або кластерів, в залежності від схожості і відмінностей між об'єктами до яких відносяться дані. Що стосується застосованої методики, то кластерний аналіз для класифікації об’єктів оцінює відстані між парами об'єктів, а кластерний аналіз для класифікації змінних оцінює відстані між парами змінних.

Ієрархічний кластерний аналіз може розглядатися як інструмент, який спрямований на зведення множини об'єктів або навіть змінних до меншої кількості інформаційних об'єктів, які є класами або кластерами. Відповідний інструмент може бути застосованим в економічній та фінансовій сфері з метою розподілу підприємства на класи, дотримуючись його існуючої латентної структури. В межах інноваційних кластерних структур та різного типу корпорацій, досліджений економіко-математичний інструмент аналізу може бути застосованим для оцінки синергетичного ефекту суміжного функціонування суб’єктів (фінансового, екологічного, соціального тощо), в залежності від постановки завдання, що формує перспективу подальшого дослідження.

Список використаних джерел:

1. Everitt B. Cluster Analysis / B. Everitt. – NewYork : Wiley,1993. – 283 p.

2. Porter M.F. An algorithm for suffix stripping [Електронний ресурс]. M.F. Porter// Program–1980. – №4(3). – Р. 130-137. – Режим доступу: <http://tartarus.org/~martin/Porter> .

3. Bradley, P. S. Constrained k-means clustering [Електронний ресурс] / Bradley, P. S., Bennett, K. P. Demiriz, A. // Microsoft Research. MSR-TR-2000-65. 2000. — Redmond, W. A. — Режим доступу до статті.: <http://www.litech.org/~wkiri/Papers/wkiri.html> .

4. Чубукова І. А. Методи кластерного аналізу. Ієрархічні методи [Електронний ресурс]: (Інтернет-Університет Інформаційних Технологій. Дистанційна освіта. – 2003-2008)/ І.А. Чубукова// Data Mining: (лекція No13). – 2006. – Режим доступу: <http://www.intuit.ru/department/database/datamining/13/2.html> .