

ної енергії вугільних шахт // Техн. електродинаміка. – 2008. – № 4. – С. 61–64. 6. Трифонов В.Д., Трифонов Д.В. Теоретико-информационный подход к решению задач энергетического аудита // Гірничя електромеханіка та автоматика. – 2005. – Вип. 75. – С. 30–37. 7. Трифонов Д.В. Моделі та методи підвищення ефективності використання електроенергії у виробничих системах (на прикладі вугільних шахт): автореф. дис. ...канд. техн. наук: спец. 05.14.01 “Енергетичні комплекси та системи” / Д.В. Трифонов. – К., 1998. – 24 с. 8. Дубров А.М. Обработка статистических данных методом главных компонент. – М.: Статистика, 1978. – 135 с. 9. Иберла К. Факторный анализ. – М.: Статистика. – 1980. – 398 с. 10. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика : Классификация и снижение размерности. – М. : Финансы и статистика, 1989. – 607 с. 11. Сошникова Л.А., Тамашевич В.Н., Узбе Г., Шефер М. Многомерный статистический анализ в экономике: учеб. пособ. для вузов / Под ред. В.Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с. 12. Таблицы математической статистики / С.Н. Большев, Н.В. Смирнов. – М.: Наука, 1983. – 416 с.

УДК 658.26:338.008

В.П. Розен, П.П. Іщук, Л.В. Давиденко
Національний технічний університет України
“Київський політехнічний інститут”,
кафедра АУЕК
Луцький національний технічний університет,
кафедра електропостачання

ВИКОРИСТАННЯ МЕТОДУ К-СЕРЕДНІХ КЛАСТЕРНОГО АНАЛІЗУ ПІД ЧАС РОЗВ’ЯЗАННЯ ЗАДАЧ ЕНЕРГЕТИЧНОЇ БЕЗПЕКИ ТЕРИТОРІЙ

© Розен В.П., Іщук П.П., Давиденко Л.В., 2010

Описано метод кластеризації з використанням методу К-середніх для завдань енергетичної безпеки. Цей метод адаптований для розв’язання конкретної групи завдань.

Ключові слова: енергетична безпека, кластерний аналіз, метод К-середніх, управління енергетичною галуззю

In this article it is described a method of clustering with use of a method of K-averages for problems of power safety. The given method is adapted for the decision of the given circle of problems.

Keywords: power safety, cluster analysis, K-means method, power management

Вступ

У сучасних умовах внутрішнє й зовнішнє середовище економіко-енергетичних об’єктів (територій), стосовно їх енергетичної безпеки, усе більше набуває властивості нестабільності й невизначеності. Нестабільність проявляється в тому, що темпи зміни середовища зростають, а невизначеність – у тому, що виникаючі ситуації все частіше стають непізнаваними, точніше зовсім новими. Цей факт висуває вимогу до підвищення ефективності управлінських рішень, які обумовлені впливом стихійних і динамічних факторів внутрішнього й зовнішнього середовища території. Однак визначення ступеня впливу зазначених факторів і встановлення серед них ключових утруднене, оскільки існуючі методи визначення факторного впливу не завжди відповідають сучасним умовам. У зв’язку із цим виникає методична проблема прийняття

управлінських рішень, заснована на обліку ключових факторів, що покладає весь тягар керування на інтуїцію управлінця. Вирішення цієї проблеми дозволить виробляти управлінські рішення, адекватні динамічному факторному впливу.

Аналіз останніх досягнень та публікацій

Кластерний аналіз включає методи, адекватні для тієї або іншої сукупності даних. Найпоширеніший метод об'єднання (деревоподібної кластеризації). Призначення цього алгоритму полягає в об'єднанні об'єктів у доволі більші кластери, використовуючи деяку міру подібності або відстань між об'єктами. Типовим результатом такої кластеризації слугує деревоподібна діаграма.

Метод деревоподібної кластеризації використовується під час формування кластерів відмінності або відстані між об'єктами. Ці відстані можуть визначатися в одномірному або багатомірному просторі. Найбільш прямий шлях обчислення відстаней між об'єктами в багатомірному просторі полягає в обчисленні евклідових відстаней. Якщо існує дво- або тривимірний простір, то цей захід має реальну геометричну відстань між об'єктами в просторі. Однак алгоритм об'єднання не “турбується” про те, чи є “дані” для цієї відстані справжніми або деякими іншими похідними мірами відстані, що дуже важливо під час дослідження, тому виникає необхідність розв'язання задачі з добору правильного методу для специфічних застосувань.

Метод Варда використовує методи дисперсійного аналізу для оцінки відстаней між кластерами. Метод мінімізує суму квадратів для будь-яких двох (гіпотетичних) кластерів, які можуть бути сформовані на кожному кроці [3]. Загалом метод представляється ефективним, однак він прагне створювати кластери малого розміру і застосовується для порівняно невеликих сукупностей даних.

Задача дослідження

Для проведення детального аудиту енергетичної безпеки територіальних утворень необхідно провести кластерний аналіз [1]. Основна мета аналізу – виділити у вихідних багатомірних даних такі однорідні підмножини, щоб об'єкти усередині груп були схожі у відомому змісті один на одного, а об'єкти з різних груп – не схожі. Під “подібністю” розуміється близькість об'єктів у багатомірному просторі ознак, і тоді задача зводиться до виділення в цьому просторі природніх скупчень об'єктів, які й вважаються однорідними групами. Розчленувавши на групи території з подібними ознаками, зможемо робити подальший аналіз стану енергобезпеки та рекомендувати до виконання ті чи інші заходи для усунення нестабільності в енергетичній сфері цього утворення.

Виклад основного матеріалу

Широке застосування для задач кластеризації з великою кількістю вхідної інформації одержав метод К-середніх [1]. Цей метод кластеризації істотно відрізняється від таких агломеративних методів, як деревоподібна кластеризація. Припустимо, існують гіпотези щодо кількості кластерів (за станом енергетичної безпеки). У такому разі можливо вказати алгоритму утворювати рівно три кластери так, щоб вони були настільки різні, наскільки це можливо. Це саме той тип задач, які розв'язує алгоритм методу К-середніх. У загальному випадку метод К-середніх будує рівно К різних кластерів, розташованих на якнайбільших відстанях один від одного.

Для розрахунку використаємо математичний пакет NCSS (рис. 1) [5].

Першим етапом для проведення цього аналізу є необхідність визначення діагностичних ознак їх кількості та збору статистичної інформації. За статистичні одиниці приймаємо окремі райони, описані набором діагностичних ознак (таблиця), що взятий з [4] і має такий склад: використання котельно-пального палива (тис. т у. п.) (1); використання теплоенергії (тис. Гкал) (2); використання електроенергії (млн. кВт·год) (3); використання автомобільного бензину (т) (4); використання дизельного палива (т) (5); видобуток котельно-пального палива (тис. т у. п.) (6); виробництво електроенергії (млн. кВт·год) (7); потужність електростанцій (тис. кВт) (8); відпуск теплоенергії котельними (Гкал) (9); відпуск теплоенергії електростанціями (Гкал) (10).

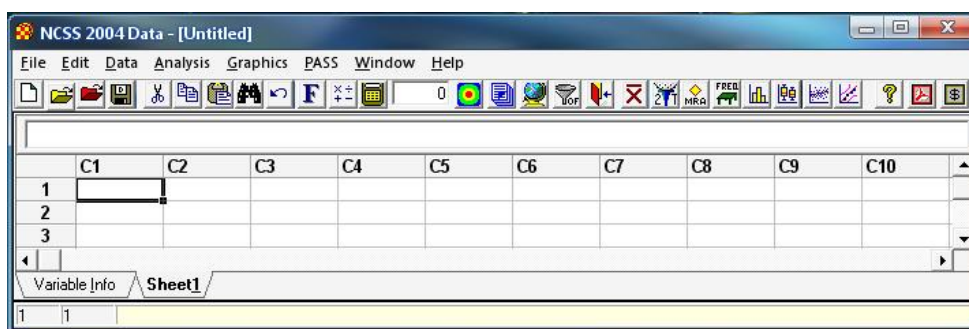


Рис. 1. Загальний вигляд вікна програми NCSS

Нехай в нас є N – рядки спостережень, які поділимо на K груп. K -тий кластер містить n_k спостережень. Кожен рядок складається з змінних P . Недостаюче (пропале) значення в i -й змінній j -го рядка k -ї групи визначається як ijk . Стандартизовані дані позначаються як z_{ij} .

Таблиця 1

Набір діагностичних ознак районів Волинської області (усереднені значення за 1996–2002 рр.)

№ з/п	Райони	№ ознаки									
		1	2	3	4	5	6	7	8	9	10
1	Володимир-Волинський	34,21	93,50	27,28	1874,84	4729,58	0,00	5,40	11,00	98442,75	55857,50
2	Горохівський	23,01	71,05	29,14	2649,46	7050,78	0,00	5,85	6,50	32816,25	58976,25
3	Іваничівський	69,33	181,02	100,96	2874,14	4035,40	318,72	3,10	6,60	507527,50	33277,75
4	Камінь-Каширський	5,78	12,94	8,43	838,64	1636,28	0,00	0,00	0,40	13406,00	0,00
5	Ківерцівський	9,46	33,09	13,02	1166,14	2843,06	0,00	0,00	0,60	29184,50	0,00
6	Ковельський	81,48	93,45	111,99	2971,16	5276,82	0,00	0,00	2,10	223714,25	0,00
7	Локачинський	2,57	1,16	7,78	1063,48	2878,46	44,74	0,20	0,50	2736,50	0,00
8	Луцький	275,78	755,68	240,18	17214,70	20447,82	0,00	9,75	15,30	1181711,00	106522,00
9	Любешівський	4,77	8,22	5,53	577,38	1618,26	0,00	0,00	0,10	8636,50	0,00
10	Любомльський	5,53	10,47	7,94	963,04	1811,36	0,00	0,15	0,90	19090,25	0,00
11	Маневийський	45,03	56,73	16,74	960,36	2462,20	95,96	0,00	0,10	63973,50	0,00
12	Ратнівський	5,97	8,77	8,85	1116,96	2915,16	0,00	0,00	0,10	13940,00	0,00
13	Рожищенський	6,68	18,05	14,57	1441,08	3502,48	0,00	0,00	0,70	22622,50	0,00
14	Старовижівський	4,94	4,03	5,58	513,72	1642,44	0,00	0,00	0,80	13211,25	0,00
15	Турійський	2,69	4,34	7,35	996,36	2433,98	0,00	0,00	0,20	7996,75	0,00
16	Шацький	3,28	6,82	3,33	385,20	1049,68	0,00	0,00	0,10	9282,50	0,00

Метод ініціалізації кластерів впливає на остаточне рішення. Тому здійснюємо низку випадкових стартових конфігурацій для збільшення ймовірності глобального оптимального рішення для специфічної кількості кластерів. Метод починає з K випадково обраних кластерів, а потім змінює належність об'єктів до них, щоб, по-перше, мінімізувати мінливість усередині кластерів і, по-друге, максимізувати мінливість між кластерами. Цей спосіб аналогічний методу "дисперсійний аналіз навпаки" у тому розумінні, що критерій значимості в дисперсійному аналізі порівнює міжгрупову мінливість із внутрігруповою при перевірці гіпотези про те, що середні в групах відрізняються одна від одної. У кластеризації методом K -середніх алгоритм переміщує об'єкти з одних кластерів в інші для того, щоб одержати найбільш значимий результат.

Критерій згоди підганяння різних кластерних конфігурацій для порівняння визначається за формулою

$$WSS_k = \left(\frac{NP}{NP - m} \right) \sum_{k=1}^K \sum_{i=1}^P \sum_{j=1}^n (1 - d_{ijk}) (z_{ij} - c_{ik})^2, \quad (1)$$

де c_{ik} – середнє значення i -ї змінної в k -му кластері.

Процент розкиду визначається за формулою

$$PV_k = 100 \frac{WSS_k}{WSS_1}. \quad (2)$$

Програма видає графіки кластерного розбиття, які умовно розчленовують N -вимірний простір на ряд двовимірних графіків, що є прийнятним для зорового сприйняття яких зображено кластери і розділювальні лінії, що символізують межу між кластерами. Для прикладу наведемо декілька з них (рис. 2).

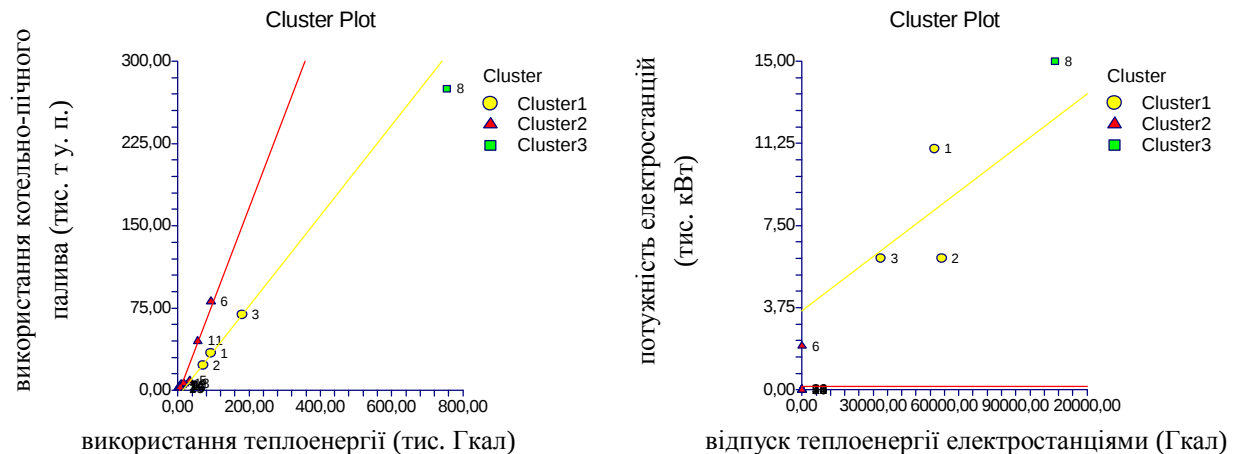


Рис. 2. Двовимірні графіки кластерів

Звичайно, коли результати кластерного аналізу методом К-Середніх отримані, розраховуються середні для кожного кластера по кожному виміру, оцінюється наскільки кластери відрізняються один від одного. В ідеалі повинні вийти середні, що сильно відрізняються, для більшості, якщо не для всіх вимірів, використовуваних в аналізі. Значення F-Статистики, отримані для кожного виміру, є індикатором того, наскільки добре відповідний вимір дискримінує кластери.

Результати цього дослідження вказують на правильність вибору цього методу для кластеризації енергетичних територіальних об'єктів з дослідженнями методами класичної статистики, а саме методом сфер, яке було описано в [6].

Якщо прийняти як визначальну задачу стратегічного керування виробляти управлінські розв'язки, керуючись впливом ключових факторів, то для цих цілей необхідно рахувати те, що ці фактори мають властивість змінюватися в значеннях і напрямках відповідно динаміки й стану енергобезпеки. Отже, неодмінною якістю для прийняття тих або інших стратегічних орієнтирів території є здатність сприймати дані зміни факторів і відповідно на них реагувати, тобто перебувати в деякому рухливому стані, у якому реакції цієї території будуть варіюватися.

Висновки

Безумовно, кластерний метод має й деякі межі для точності прийнятих на його основі розв'язків, але він і дозволяє розширити категорію стратегії стосовно більшої кількості прийнятих управлінцями стратегічних розв'язків, тому що добре погоджується із системним представленням про енергобезпеку території, що перебуває в стані постійних змін.

Застосування кластерного аналізу дозволяє розробляти управлінські розв'язки, на підставі яких можуть бути оцінені ті або інші стратегічні дії стосовно діяльності управлінців на територіях. При цьому вибір уже сформульованої альтернативи ґрунтується не на двох, трьох, а на багатьох факторах, що являють собою інтегральні або результуючі характеристики. Перевага використання кластерного аналізу полягає в застосуванні значно більшої кількості факторів, що

відносяться й до внутрішнього, і до зовнішнього середовищ території. Більше того, кластерний аналіз за допомогою своїх алгоритмів дозволяє визначити значимість того або іншого фактора на момент дослідження.

Усе це дає змогу найефективніше реалізувати систему стратегічного керування. Позитивні ознаки використання кластерного аналізу в системі стратегічного керування полягають у такому:

- дозволяє досягати стабільності, упорядкованості в стратегічному плануванні, тому що основним моментом є зміна стратегії залежно від впливу факторів;
- підвищує ефективність у подоланні опору змінам за рахунок планування стратегії.

Підхід до розроблення управлінських розв'язків за допомогою кластерного аналізу має можливість якісно поліпшити систему стратегічного керування енергетичною галуззю території за допомогою адекватного реагування на вплив ключових факторів, що впливають на енергетичну безпеку територіальних утворень.

1. Мандель И.Д. Кластерный анализ / И.Д. Мандель. – М.: Финансы и статистика, 1988. – 176 с. 2. Hartigan, J.A. Clustering algorithms. – New York: Wiley, 1975. 3. Ward J.H. Hierarchical grouping to optimize an objective function // *Journal of the American Statistical Association*. – 1963. – 236 p. 4. Влияние энергетического фактора на экономическую безопасность регионов Российской Федерации / Л.Л. Богатырев, В.В. Бушуев, А.А. Куклин, А.Л. Мызин, А.И. Татаркин и др. – Екатеринбург: Изд-во Уральского университета, 1998. – 288 с. 5. Internet: ncss.com. 6. Розен В.П. Районування адміністративних одиниць Волинської області відповідно до стану їх енергобезпеки / В.П. Розен, П.П. Іщук // *Промелектро*. – 2005. – № 4. – С. 40–45.