

*М.М. МИКИЙЧУК*, канд. техн. наук, доц., НУ«Львівська політехніка»

## **ПЕРСПЕКТИВИ ЗАСТОСУВАННЯ КЛАСТЕРНОГО АНАЛІЗУ ПРИ КОНТРОЛІ ЯКОСТІ ПРОДУКЦІЇ**

Здійснено дослідження перспектив застосування кластерного аналізу при вибірковому контролі якості продукції з метою підвищення достовірності ідентифікації її відповідності, що дозволить ефективно управляти ризиками виробника та споживача.  
Ключові слова: кластерний аналіз, якість продукції, контроль якості.

Осуществлено исследование перспектив применения кластерного анализа при выборочном контроле качества продукции с целью повышения достоверности идентификации ее соответствия, что позволит эффективно управлять рисками производителя и потребителя.  
Ключевые слова: кластерный анализ, качество продукции, контроль качества.

Research of prospects of application of cluster analysis is carried out at the sampling test of quality of products with the purpose of increase of authenticity of authentication of its accordance, that will allow effectively to manage the risks of producer and user.  
Keywords: cluster analysis, quality of products, control of quality.

### **Вступ**

Сьогодні кластерний аналіз широко використовується в медицині, психології, державному управлінні, маркетингу, соціології [1]. Ефективним кластерний аналіз може бути у випадках, коли необхідно розподілити велику кількість інформації на придатні для подальшої опрацювання масиви даних [2]. Особливо це актуально для оперативного аналізу потоків інформації про значення показників якості продукції з метою подальшої її ідентифікації та віднесення до відповідного класу [3]. При аналізі рівня якості продукції достовірною її кластеризація може стати вагомим шляхом для зменшення наслідків від виробничих ризиків, а також ефективним інструментом їх запобігання. Проблема класифікації об'єктів за певним набором кількісних чи якісних ознак була і залишається актуальною з огляду на потребу їх автоматизованого групування та ідентифікації.

### **Сучасний стан проблеми**

Кластерний аналіз — це багатовимірна статистична процедура, що виконує збір даних, що містять інформацію про вибірку об'єктів, з наступним упорядкуванням об'єктів у порівняно однорідні групи.

Всі методи кластеризації поділяють на дві групи: ієрархічні та неієрархічні. Неієрархічні методи кластеризації використовують ітеративні процедури для виконання яких потрібно означити такі параметри як початкова точка, правило формування нових кластерів, правило зупинки. Неієрархічні методи можуть працювати з потужними базами даних, вони є складними, тому що потрібно до початку кластеризації визначити приблизну кількість кластерів та кількість ітерацій. На відміну від попередніх, ієрархічні методи не тільки враховують особливості реальних об'єктів дослідження, а й дозволяють знайти найкращі варіанти утворення кластерів, орієнтуючись на задані критерії оптимізації без

побудови повного дерева рішень, вони є наочними та простими, при їх допомозі можливо отримати детальну структурування даних [5].

Сучасні технологічні процеси (ТП) характеризуються великою кількістю контрольованих технологічних параметрів та показників якості готової продукції [4]. Тому для управління ТП широко використовують методи математичної статистики. Однак необхідність підвищення рівня та оперативності управління сучасними ТП ставить вимогу переходу від статистичного оцінювання одномірних залежностей – вплив окремого технологічного параметра на показник якості готової продукції, до оцінювання, багатомірному впливу сукупності технологічних параметрів на сукупність показників якості. Традиційно проблему оцінювання якості виробів, які характеризуються значною кількістю показників якості вирішують шляхом визначення узагальненого показника якості. Основною вимогою до узагальненого показника якості є максимальна повнота відображення всіх характерних властивостей продукції, що в багатьох випадках або неможливо або виникають недопустимі похибки оцінювання.

Враховуючи ефективність застосування кластерного методів для аналізу багатовимірних статистичних даних доцільно використовувати його при контролі якості продукції на стадії виготовлення. Тому використання кластерного методу для аналізу даних про значення показників якості продукції дозволить підвищити достовірність ідентифікації придатності продукції, а отже, зменшити ризики виробництва.

#### **Постановка завдання**

Розглянути перспективи застосування кластерного аналізу при здійсненні вибіркового контролю якості продукції з метою підвищення достовірності контролю.

#### **Перспективи застосування кластерного аналізу для контролю якості продукції**

Задачу кластеризації при контролі якості продукції можна визначити в такий спосіб. Нехай множина  $I = \{I_1, \dots, I_j, \dots, I_n\}$  визначає  $n$  об'єктів готової продукції які відносяться до певної локальної сукупності  $\pi_l$ . Відома, також множина показників якості  $Y = \{Y_1, \dots, Y_i, \dots, Y_p\}$ , які характеризують кожен об'єкт множини  $I$  та контролюються при здійсненні вихідного контролю якості. Результат вимірювання  $i$ -го параметра під час виготовлення  $I_j$  об'єкту позначатимемо  $x_{ij}$ , а вектор  $X_i = \{x_{ij}\}$  буде відповідати кожному ряду вимірювань  $Y_i$  показника якості для всіх контрольованих об'єктів. Таким чином, для множини об'єктів  $I$  матимемо множину векторів вимірювань  $X = \{X_1, X_2, \dots, X_j, \dots, X_n\}$ , які описують умови виготовлення множини продукції  $I$ .

Завдання кластерного аналізу при оцінюванні якості продукції полягає в тому, щоб на основі даних, які містяться в  $X$ , розділити множину об'єктів  $I$  на  $m$  кластерів  $\pi_1, \pi_2, \dots, \pi_m$  так, щоб кожен об'єкт  $I_j$  належав лише одному з кластерів. Важливо при цьому забезпечити умову, що об'єкти, які належать одному кластеру є максимально подібні, а об'єкти, які належать різним кластерам є мінімально подібні, тобто відмінні. Поділ вибірки на групи схожих об'єктів (кластери) дозволяє спростити подальшу обробку даних і ухвалення рішень, а також застосовувати до кожного кластера свій метод аналізу.

Очевидно, що для вирішення задачі кластеризації необхідно визначити ознаки подібності та відмінності.

У літературі описується багато різних методів кластеризації, які базуються на використанні матриць подібності, оцінюванні функцій щільності статистичного розподілу, евристичних алгоритмах перебору, ідеях математичного програмування і ін. Значна частина цих алгоритмів, при всій їх несхожості, методично засновані на одній тезі -гіпотезі компактності, тобто визначенні подібності об'єктів за мірою близькості оцінюваних ознак. Ознака подібності виникає коли відстань між результатами вимірювань показників якості  $j$ -го та  $n$ -го об'єктів є достатньо малою, і навпаки, відмінність між  $j$ -м та  $n$ -м об'єктами виникає при достатньо великій різниці між вказаними вище результатами вимірювань [6].

Евклідова метрика, в порівнянні з іншими методами оцінювання міри подібності, є достатньо простою та добре інтегрується з існуючими методиками контролю якості. В евклідовій метриці міру подібності  $d(X_j, X_n)$  визначають як відстань між  $X_j$  та  $X_n$ , що еквівалентно відстані між  $I_j$  та  $I_n$ . Тоді можна записати міру подібності для евклідового простору:

$$d(X_j, X_n) = \sqrt{\sum_{i=1}^p (x_{ij} - x_{in})^2} \quad (1)$$

Для сукупності об'єктів контролю  $I = \{I_1, \dots, I_j, \dots, I_n\}$ , кожен з яких характеризується множиною показників якості  $\{Y_1, \dots, Y_j, \dots, Y_p\}$ , результати  $n$  векторів вимірювань  $\{X_1, X_2, \dots, X_n\}$  можна представити у вигляді матриці результатів  $\{x_{ij}\}$  розміром  $p \times n$ :

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \cdot & \cdot & \dots & \cdot \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix} = (X_1, X_2, \dots, X_n) \quad (2)$$

Відстань між парами векторів  $d(X_j, X_n)$ , з врахуванням (1), можна представити у вигляді симетричної матриці відстаней :

$$D = \begin{pmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \cdot & \cdot & \dots & \cdot \\ d_{p1} & d_{p2} & \dots & 0 \end{pmatrix} \quad (3)$$

Аналіз методів числового визначення коефіцієнтів подібності проведений в [7] показав, що їх оцінювання можливо при забезпеченні умови нормалізації значень  $X_j$  та  $X_n$ . Така умова забезпечується при організації контролю якості продукції за альтернативною ознакою. Зокрема в статистиці широко використовується лінійна міра подібності  $r_{jn}$ , що називається коефіцієнтом кореляції та знаходиться з виразу:

$$r_{jn} = \frac{\sum_{i=1}^p x_{ij} \cdot x_{in}}{\sqrt{\sum_{i=1}^p x_{ij}^2 \cdot \sum_{i=1}^p x_{in}^2}} \quad (5)$$

Таким чином, використовуючи приведені вище залежності можна визначати ступінь близькості результатів вимірювань показників якості окремих зразків готової продукції.

Наступним кроком кластерного оцінювання якості продукції є означення властивостей кластерів та встановлення відповідності окремих зразків продукції цим кластерам. Очевидно, що властивості кластерів визначаються переліком оцінюваних показників якості продукції  $\{Y_1, \dots, Y_b, \dots, Y_p\}$  та допустимими межами їх зміни. Сьогодні інформація про показники якості конкретної продукції та допустимі межі їх зміни зосереджена в технічних умовах, а отже труднощів з визначенням властивостей кластерів не повинно виникати.

Виконання основного завданням кластерного аналізу - розподіл множини значень показників якості партії продукції до певного кластера можна здійснювати за обраним критерієм оптимальності (цільовою функцією). Одним з варіантів є оцінка області допусків параметрів у вигляді еліпсоїда. Такі оцінки особливо ефективні у випадку «втягнутості» області параметрів, або у випадках коли задана область розсіювання відхилень параметрів від номінальних [6, 8].

Оскільки при здійсненні вибіркового контролю за альтернативною ознакою рівень якості продукції визначається за відповідністю виміряних значень  $X = \{X_1, X_2, \dots, X_j, \dots, X_n\}$  оцінюваних показників якості  $\{Y_1, \dots, Y_b, \dots, Y_p\}$  встановленим допуском, то природньо прийняти за основні кластери зону значень окремих показників якості  $Y_i$  за яких продукція вважається придатною до використання.

Для означення властивостей кластера потрібно знайти його центр за результатами вимірювань значень  $i$ -го параметра для всієї сукупності об'єктів контролю:

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n x_{ij} \quad (6)$$

Також необхідно визначити допустиму область розповсюдження кластера, наприклад через середнє квадратичне відхилення:

$$\sigma_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{X}_i)^2} \quad (7)$$

Таким чином, правило належності  $j$ -го об'єкта контролю за окремим  $i$  – м показником якості буде визначатися умовою:

$$\sigma_{ij} < \sigma_i \quad (8)$$

Провівши процедуру кластеризації всіх об'єктів контролю за всіма показниками якості можна всі об'єкти контролю розділити на три групи:

- придатні – всі показники якості увійшли в допустимі кластери;
- непридатні – всі показники якості опинилися за межами допустимих кластерів;
- сумнівні – частина показників якості увійшла в допустимі кластери, а частина опинилася за межами допустимих кластерів.

Достовірність класифікації рівнів якості продукції залежить від розміщення еліпсоїдів розсіювання показників якості, що належать відповідному кластеру. При ідентифікації продукції з метою віднесення її до певного рівня якості доцільно використовувати існуюче нормативно-методичне забезпечення

контролю якості, в першу чергу методи контролю за альтернативною ознакою. Ці методи використовуються в системах контролю якості завдяки широкому впровадженню статистичного вибіркового контролю і базуються на побудові операційної характеристики контролю за допомогою якої погоджуються ризики виробника та споживача. Існує ряд нормативних документів, які регламентують проведення контролю якості за альтернативною ознакою. Тому важливим моментом впровадження кластерного аналізу є інтегрування його елементів в існуючі методики контролю за альтернативною ознакою.

Традиційно при вибіркового контролі партію виробів розділяють на дві групи: придатні та непридатні, за допомогою двох чисел:  $q_0$  та  $q_m$ . Причому  $q_0$  визначає гранично допустиму долю непридатної продукції, а  $q_m$  встановлює межу ідентифікації партії продукції як непридатної. Продукція вважається придатною при  $q < q_0$  та непридатною якщо  $q > q_m$ . Задачу вибору значень  $q_0$  та  $q_m$  вирішують шляхом погодження з показниками ризику виробника -  $\alpha$  та ризику споживача -  $\beta$ . Погодження відбувається шляхом побудови плану контролю до якого ставляться такі вимоги:

$$\begin{aligned} P(q) &\geq 1 - \alpha \text{ при } q \leq q_0 \\ P(q) &\leq \beta \text{ при } q \geq q_m \end{aligned} \quad (9)$$

При цьому виникає зона невизначеності при  $q_0 < q < q_m$  коли виникають сумніви в придатності продукції. В більшості випадків таку продукцію визнають придатною, що збільшує ризик споживача. У випадку необхідності підвищення достовірності ідентифікації продукції здійснюють повторні випробування, що зумовлює зростання затрат виробництва.

Ефективним шляхом уникнення вказаних затрат виробництва при одночасному підвищенню достовірності ідентифікації продукції є використання кластерного аналізу результатів контролю продукції. Інтегрування кластерного методу ідентифікації продукції, якість якої піддається сумніву, можна шляхом побудови дерева рішень з використання операційної характеристики контролю. Графічну інтерпретацію інтегрованого методу оцінювання якості сумнівної продукції можна

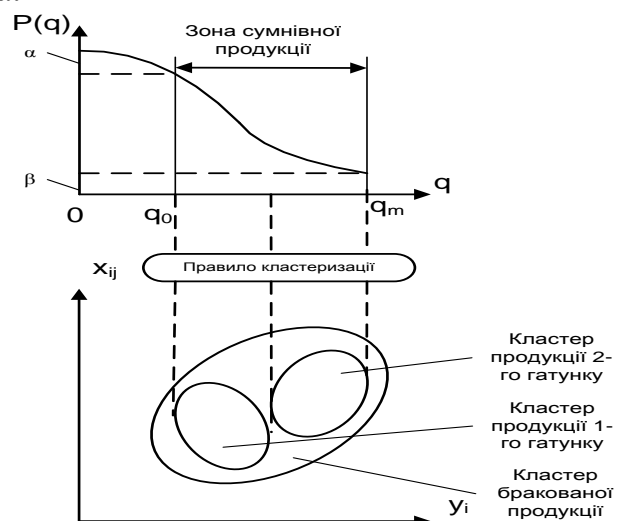


Рис.1. Графічна інтерпретація інтегрованого методу оцінювання якості сумнівної продукції.

представити в наступному виді (див. рис. 1).

Важливим завданням застосування кластерного аналізу при контролі якості є розроблення вирішального правила яке визначатиме умови ідентифікації продукції.

### Висновки

Таким чином, проведений аналіз перспектив застосування кластерного аналізу для контролю якості продукції показує доцільність його інтегрування в

існуючі методи вибіркового контролю. Застосування інтегрованого методу при контролі якості продукції дозволить зменшити затрати на проведення контролю при одночасному підвищенні його достовірності.

**Список літератури:**1. *Савчук Т.О.* Порівняльний аналіз використання методів кластеризації для ідентифікації надзвичайних ситуацій на залізничному транспорті / Т.О. Савчук, С.І. Петришин // Наукові праці ДонНТУ, Серія "Інформатика, кібернетика та обчислювальна техніка". - 2010. - Випуск 12(165). - С.135-140.2.Кластерный анализ. [Електронний ресурс]. Режим доступу: <http://www.statsoft.ru/home/textbook/modules/stcluan.html#statistical>. 3.*Дунаев Б.Б.* Точность измерений при контроле качества. [Текст]: – К.: Техніка. - 1981. - 152 с.4.*Микийчук М.М.* Параметрична модель для оцінювання якості технологічних процесів / М.М. Микийчук, Т.З. Бубела, Т.Г. Бойко // Автоматика, вимірювання та керування. Вісник НУ "Львівська політехніка". - 2007. - №574. - С. 89-93.5.*Савчук Т.О.* Використання ієрархічних методів кластеризації для аналізу надзвичайних ситуацій на залізничному транспорті / Савчук Т.О., Петришин С.І. // Вимірювальна та обчислювальна техніка в технологічних процесах. - 2009. - № 1. - С.193-198.6.*Дюран Б.* Кластерный анализ [Текст]: / Б. Дюран , П. Одел // Пер. с англ. – М.: Статистика. - 1977. – 128 с.7. *Айвазян С.А.* Классификация многомерных наблюдений [Текст]: / С.А. Айвазян, З.И. Бежаева, О.В. Староверов.// - М.: «Статистика». – 1974. – 238 с.8.*Козак О.Л.* Метод ітераційного формування допускової еліпсоїдної оцінки в задачах параметричної ідентифікації інтервальних моделей / О.Л. Козак // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2008. - № 2 – с. 254-261.

*Поступила в редколлегию 12.05.2011*