

Методи та прилади контролю якості. – 2004. – Вип.12, – С. 42–48. 3. Френкель Б.А. Измерение расхода жидкостей и газов в малотоннажных производствах и на экспериментальных установках. – М., 1989. 4. Коллеров Д.К. Газоанализаторы. Проблемы практической метрологии. – М., 1980. 5. Залманзон Л.А. Аэрогидродинамические методы измерения входных параметров автоматических систем. Измерительные элементы пневмоники. – М., 1973. 6. Голубев И.Ф. Вязкость газов и газовых смесей: Справочное руководство. – М., 1959. 7. Кивилис С.С. Плотномеры. – М., 1980. 8. Бретишайдер С. Свойства газов и жидкостей. – М., 1966. 9. Голубев И.Ф., Агаев Н.А. Вязкость предельных углеводородов. – Баку, 1964. 10. Голубев И.Ф., Гнездилов Н.Е. Вязкость газовых смесей. – М., 1971. 11. Рид Р., Шервуд Т. Свойства газов и жидкостей (определение, корреляция). – Л., 1971. 12. Варгафтик Н.Б. Справочник по теплофизическим свойствам газов и жидкостей. – М., 1972. 13. Динамическая вязкость и теплопроводность гелия, неона, аргона, криптона и ксенона при атмосферном давлении в интервалах температур от нормальных точек кипения до 2500 К. ГСССД 17-81. – М., 1982. 14. Морачевский А.Г., Сладков И.Б. Физико-химические свойства молекулярных неорганических соединений (Экспериментальные данные и методы расчета): Справ. изд. – Л., 1987. 15. Рид Р., Праусниц Дж., Шервуд Т. Свойства газов и жидкостей: Справочное пособие. 3-е изд., перераб. и доп. – Л., 1982. 16. Теплюх З.М., Друль Я.Г. Апроксимація температурної залежності в'язкості основних промислових газів // Вісник ДУ "Львівська політехніка" "Теплоенергетичні системи та пристрої". – 1993. – № 273. – С. 44–46. 17. Теплюх З.М. Синтезатори газових сумішей для перевірки аналізаторів складу димових газів // Методи та прилади контролю якості. – Ів.-Франківськ. – 2002. – Вип. 8. – С. 83–85.

УДК 681.5:666.3/7

О. Бешта, Т. Бабенко

Національний гірничий університет

КЛАСТЕРНИЙ АНАЛІЗ В КОМП'ЮТЕРНОМУ МОДЕЛЮВАННІ ТЕХНОЛОГІЧНОГО ПРОЦЕСУ ВИРОБНИЦТВА КЕРАМІКИ

© Бешта О., Бабенко Т., 2005

The article is devoted to the revealing with use of the cluster analysis the hidden relations in the data needed for the synthesis of computer simulating models of the ceramics manufacture process.

Технологічні процеси керамічного виробництва внаслідок складності вивчення їх фізико-хімічної природи здебільшого характеризуються значною розмірністю простору станів, недостатніми теоретичними обґрунтуваннями причинно-наслідкових зв'язків між значеннями окремих технологічних параметрів і їх впливом на показники якості на проміжних та кінцевій стадіях виробництва. При цьому необхідно також враховувати, що параметри стану об'єкта досліджень формуються під дією значної кількості зовнішніх чинників (температура, вологість зовнішнього середовища) та внутрішніх (термодинамічні джерела тощо), а також збурень різної фізичної природи, які діють в різних напрямках і не є завжди відомими. Тому традиційно керамічна промисловість характеризується досить великою часткою зворотного браку на різних технологічних ділянках, що при значній ресурсо- і енергоємності виробництва обумовлює його низьку ефективність та конкурентоспроможність продукції.

У загальному випадку однією з основних і найбільш складних задач комп'ютерного моделювання технологічного процесу є задача прогнозування кінцевого стану системи (об'єкта). Будь-яке прогнозування (екстраполяція) ґрунтується на формалізованому уявленні про існування зв'язків між причиною і наслідком. Статистичний аналіз зазначених процесів включає дослідження взаємозв'язків параметрів стану системи як в статичному стані, так і в часі, а інформацією для вивчення зазначених взаємозв'язків, як правило, є часові ряди показників, що характеризують розвиток об'єктів. Найбільш розповсюдженим підходом до розв'язання цього класу задач вважається екстраполяція відомих зв'язків і закономірностей на майбутнє. Моделі прогнозування, побудовані за цим принципом, відрізняються лише гіпотезами щодо конкретних видів зв'язків. Проте у випадку неповноти знань про причинно-наслідкові зв'язки в станах системи синтезувати і обґрунтувати математичну модель, яка б адекватно відтворювала властивості об'єкта, досить складно.

Виходячи з зазначеного, на початковому етапі досліджень доцільно використовувати технології Data Mining (discovery – driven data mining), в основу яких покладено концепцію шаблонів, що відтворюють закономірності, властиві підвбіркам даних, а також фрагменти багатоаспектних взаємозв'язків у даних. Знайдені в такий спосіб шаблони дають змогу виявити неочевидні регулярності в даних. Одними із стандартних типів виявлення залежностей в даних методами Data Mining є методи кластерного аналізу.

У зв'язку з зазначеним, метою першого етапу досліджень повинна бути редукція простору станів і виявлення значущих залежностей, які уможливають максимально повно пояснити дію певного параметра чи групи параметрів на залежну змінну – брак на конкретних ділянках технологічного процесу.

Розглянемо загальну постановку задачі досліджень. Маємо n об'єктів, які характеризуються двома групами показників:

- x_1, \dots, x_n – результуючі значення параметрів технологічного процесу;
- y_1, \dots, y_m – показники якості на різних стадіях технологічного процесу.

Отже, об'єкти можна подати у вигляді точки в n -вимірному просторі. При цьому отримані точки, як правило, мають деякі області згущення, не утворюючи однорідної множини. Відповідно множина об'єктів, поданих точками, які належать одній області згущення, є більш однорідними порівняно з усією сукупністю об'єктів, що досліджуються. Відповідно множина об'єктів, що вивчаються, розбивається на однорідні групи параметрів, які повинні характеризуватися загальними для групи залежностями.

Методи кластерного аналізу досить різноманітні, однак критерій якості кластеризації тією чи іншою мірою має задовольняти таким формальним вимогам [1]:

- всередині виділених груп об'єкти повинні бути тісно пов'язані між собою;
- об'єкти різних груп повинні бути далекими один від одного;
- за інших рівних умов розподіл об'єктів за групами повинен бути рівномірним.

Вузловим моментом в кластерному аналізі вважається вибір міри близькості об'єктів (метрики), від якого у вирішальний спосіб залежить кінцевий результат розбиття об'єктів на групи за заданого алгоритму кластеризації. Найбільш розповсюдженими відстанями, що характеризують розташування окремих груп об'єктів у просторі, є Евклідова і квадрат Евклідової відстані, відстані Махаланобіса, Хеммінга, Чебешева і Мінковського. Вибір метрики в кожному конкретному випадку визначається, виходячи з головних цілей досліджень, фізичної і статистичної природи інформації. Найбільш широко як метрики застосовуються Евклідова та нормалізована Евклідова відстані:

$$dist = \sqrt{\sum_1^n (x_i - y_i)^2} ; \quad (1)$$

$$dist = \sum_1^n (x_i - y_i)^2 , \quad (2)$$

де x_i y_i – точки у n -вимірному просторі.

Проте, застосовуючи нормалізовану Евклідову відстань, краще враховуються великі різниці.

Іншою доволі важливою величиною під час виконання кластерного аналізу є визначення відстані між кластерами. Найбільш розповсюдженими відстанями або стратегіями об'єднання об'єктів у групи є :

- відстань між найближчими об'єктами кластерів (найближчого сусіда);
- відстань між найвіддаленішими об'єктами кластерів (найдалшого сусіда);
- відстань між центральними точками кластерів (центрів ваги);
- відстань між усіма можливими парами спостережень.

Інформація, необхідна для розрахунку зазначеної відстані, знаходиться на основі усіх теоретично можливих пар спостережень.

З погляду на наявність багатьох методів кластерного аналізу оцінка зазначених процедур була виконана Меззихом [2], який прийшов до висновку, що з найбільш розповсюджених, найкраще зарекомендували себе ієрархічні алгоритми, що поєднують об'єкти в групи за принципом центрів ваги груп і принципом середнього зв'язку. Досвід інших авторів [3] здебільшого підтверджує зазначений висновок.

Виходячи з інформаційної теорії Шеннона [4], найкращим методом класифікації необхідно вважати метод, який приводить до найбільшої ентропії (невизначеності), що отримується цим методом класифікації. Ентропія класифікації N об'єктів, що розбивається на N класів, визначається за формулою

$$H = -\sum_{i=1}^n \frac{N_i}{N} \log_2 \frac{N_i}{N} = \log_2 N - \frac{1}{N} \sum_{i=1}^n N_i \log_2 N_i, \quad (3)$$

де N_i – кількість об'єктів в i -му класі.

Отже, найкращою буде класифікація, за якої усі або більшість класів заповнені, оскільки в протилежному випадку, коли заповненими є один чи два кластери, а інші вміщують по 1–2 об'єкти, відбувається нерозбиття усієї сукупності об'єктів на класи, а виявлення об'єктів, що порушують однорідність.

Для проведення кластеризації параметрів технологічного процесу виробництва кераміки методом лиття з водних суспензій нами були використані дані підприємства “Дніпрокераміка”. Аналіз виконувався засобами програмного комплексу SPSS. Цей пакет вміщає сім методів ієрархічного кластерного аналізу: Between-groups linkage, Within-groups linkage, Nearest neighbor, Furthest neighbor, Centroid clustering, Median clustering, Ward's method. При цьому під час проведення кластеризації допускається використання восьми різних метрик: Euclidean distance, Squared Euclidean distance, Cosin, Pearson correlation, Chebychev, Block, Minkowski, Customized. Нам здається, що найбільш доцільним є використання таких стратегій об'єднання об'єктів в групи: Between-groups linkage, Centroid clustering, и Ward's method. Цей вибір обумовлений тим, що зазначені методи дають змогу отримувати найбільш однорідні в статистичному розумінні кластери.

Таблиця початкових даних вміщає 44 змінних, одна з яких характеризує якість виконання технологічних операцій (брак) на ділянках масопідготовки і лиття. Дані взяті за три місяці роботи підприємства “Дніпрокераміка”

Кластерний аналіз виконувався за такою схемою: Between-groups linkage; Within-groups linkage; Single Linkage; Complete Linkage; Centroid Linkage; Ward Linkage з використанням таких метрик: Squared Euclidean Distance; Euclidean Distance; Chebychev Distance; Minkowski Distance.

Для визначення кількості кластерів, яку необхідно вважати оптимальною, вирішальне значення має визначена відстань між двома кластерами, що отримується на основі вибраної дистанційної метрики з урахуванням прийнятої стратегії об'єднання. Як правило, процес кластеризації доцільно проводити, доки відстань між об'єктами кластерів в середньому за усіма методами і відстанями не перевищує 5–10 %. В іншому випадку в кластери будуть об'єднані об'єкти, що знаходяться на порівняно далекій відстані один від одного. Проте кінцеве рішення про зупинку кластеризації і відповідно про кількість кластерів, на які розділяється вся сукупність об'єктів, має прийматись лише після їх додаткового змістовного аналізу.

У ході проведених досліджень встановлено, що стрибкоподібне збільшення відстані між об'єктами-кластерами, що перевищує поріг у 10 %, спостерігається лише при використанні таких стратегій поєднання об'єктів у групи: Between-groups Linkage; Within-groups Linkage, Ward Linkage і відстаней, що характеризують об'єкти в гіперпросторі: Squared Euclidean Distance, Euclidean Distance. При використанні інших стратегій і методів спостерігається поступове зростання зазначеної відстані до зупинки роботи методу кластеризації.

Доречно зазначити, що ступінь порівняно рівномірного розподілу кількості технологічних параметрів на ділянках масопідготовки і лиття керамічних виробів і відповідно максимально можлива для даної вибірки ентропія досягається тільки при використанні таких методів кластерного аналізу: Between-groups Linkage - Squared Euclidean Distance; Within-groups Linkage - Squared Euclidean Distance; Ward Linkage - Squared Euclidean Distance; Ward Linkage - Euclidean Distance. Результати кластеризації з використанням зазначених методів наведені в таблиці.

**Групування 44 змінних для даних технологічного процесу
виробництва керамічних виробів
за допомогою чотирьох різних методів кластеризації**

Кластери	Використання методу Between-groups Linkage	Within-groups Linkage	Ward Linkage	Ward Linkage
	Squared Euclidean Distance			Euclidean Distance
1	2	3	4	5
1	1. Жорсткість води 19. Хімічний склад, вміст Na ₂ O	Ті самі змінні	Ті самі змінні, об'єднані з 6;17;18; 22;40;43	1;19 ті самі змінні, об'єднані з 4;5;14;31;38;44
2	2. Жорсткість фільтрованої води 3. Вміст суспензії в рекуперативній воді 11. Гранулометрія, прохід через сито 0063 12. Хімічний склад (залишок після прокалювання) 28. Корегування властивостей шлікеру, вода (накопичувальна мішалка) 35. Текучість шлікеру 1 36. Текучість шлікеру 2 37. Коефіцієнт загустіння шлікеру 39. Температура шлікеру 41. Залишок після прокалювання керамічної маси 42. Мікроклімат на ділянці лиття, температура	2;11;12; 39;41; 42 ті самі змінні об'єднані з 16;17;25; 32;33;34	3;11;28;39;42 ті самі змінні, об'єднані з 20;23	2;12;41 ті самі змінні, об'єднані з 15;16;32
3	4. Щільність, рекуперативної води 5. РН, рекуперативної води 27. Корегування властивостей шлікеру, рідке скло (мішалка) 38. РН, шлікеру	4;5;38; ті самі змінні, об'єднані з 15	7;8;9;10	7;8;9;10
4	6. Гранулометрія, залишок на ситі 025 13. Хімічний склад, вміст SiO ₂ 21. Хімічний склад, вміст вільного кварцу 40. Пластичність шлікеру керамічної маси	13;21 ті самі змінні, об'єднані з 20;23;24; 27;29;30	13;21; ті самі змінні, об'єднані з 24;27;29;30	3;11;28;39;42
5	7. Гранулометрія, залишок на ситі 0125 8. Гранулометрія, залишок на ситі 0063 9. Гранулометрія, залишок на ситі 0045 10. Гранулометрія, прохід через сито 0045	7;8;9;10;18;2 2; 26;43 ті самі змінні, об'єднані з 6	15;16;25;26 ті самі змінні, об'єднані з 2	17;18;22;25;26;33; 34 ті самі змінні, об'єднані з 6;40; 43

	15. Хімічний склад вміст Fe ₂ O ₃ 16. Хімічний склад, вміст TiO ₂ 17. Хімічний склад, вміст CaO 18. Хімічний склад, вміст MgO 22. Корегування властивостей шлікеру, вода (млин) 25. Корегування властивостей шлікеру, вода (мішалка) 26. Корегування властивостей шлікеру, дефлон (мішалка) 32. В'язкість шлікеру 33. Тиксотропія шлікеру 1 34. Тиксотропія шлікеру 5 43. Мікроклімат на ділянці лиття, вологість			
6	14. Хімічний склад, вміст Al ₂ O ₃ 31. Щільність шлікеру 44. Брак лиття	Ті самі змінні	14;31;44 ті самі змінні з 4;5;38	35;36;37
7	20. Хімічний склад, вміст K ₂ O 23. Корегування властивостей шлікеру, дефлон (млин) 24. Корегування властивостей шлікеру, рідке скло (млин) 29. Корегування властивостей шлікеру, дефлон (накопичувальна мішалка) 30. Корегування властивостей шлікеру, рідке скло (накопичувальна мішалка)	3;28;35;36;37; 40	35;36;37	20;23;24;30 ті самі змінні, з 13;21; 27;29

Змістовний аналіз отриманих результатів проводився за усіма наведеними методами шляхом експертної оцінки однорідності отриманих кластерів (з точки зору несуперечливості відомим теоріям та сутності фізико-хімічних процесів виробництва керамічних виробів методом лиття з водних суспензій). Виділення однорідних груп кластерів за усіма методами уможливило визначити кращий (з формального погляду) метод кластеризації у відповідному багатовимірному просторі. На нашу думку, найбільш змістовними можна вважати методи об'єднання "Зв'язок всередині груп" (Within-groups linkage) та метод Варда (Ward Linkage) у нормалізованому Евклідовому просторі (Squared Euclidean Distance). Результати інших методів також можуть виявитися корисними в подальших дослідженнях.

Оцінка сталості отриманих результатів і уточнення виділених кластерів у відповідному багатовимірному просторі виконувалась шляхом порівняння результатів кластерного аналізу, виконаного на аналогічних даних за інші рівні періоди роботи підприємства "Дніпрокераміка". Аналіз отриманих результатів підтверджує їх сталість відтворення у 96 % випадків.

Отже, в результаті проведених досліджень встановлено, що початкову розмірність простору станів можливо зменшити до розмірності, яка дорівнює кількості об'єктів у кожному конкретному кластері.

У подальших дослідженнях із застосуванням методик інтелектуальних технологій, зокрема апарата нейронних мереж та систем нечіткої логіки, необхідно встановити причинно-наслідкові взаємозв'язки між елементами всередині виділених груп і групами об'єктів та виявити їх вплив на залежні змінні (дефекти напівфабрикатів виробів на ділянці лиття) і відповідно побудувати модель залежностей між набором параметрів стану і якістю виконання технологічних операцій, перший етап контролю якої здійснюється на ділянці лиття. Це, своєю чергою, дасть можливість прогнозувати область параметрів стану на етапі виконання проміжних технологічних операцій і результати їх виконання, що необхідно для синтезу інтелектуальної системи підтримки прийняття рішень.

1. *Классификация и кластер* / Под ред. Дж. Вэн Райзин. – М., 1980. 2. *Mezzich J.E. An evaluation of Quantitative Taxonomic Methods. Ph. D. dissertation, Ohio State University, 1975.* 3. *Александров В.В., Алексеев А.И., Горский Н.Д. Анализ данных на ЭВМ (на примере системы сито).* – М., 1990. 4. *Шеннон К. Работы по теории информации и кибернетике / Пер. с англ.; Под ред. Р.Л. Домбрушина и О.Б. Лупанова.* – М., 1964.