

ЗАСТОСУВАННЯ КЛАСТЕРНОГО АНАЛІЗУ ПРИБУДОВІ РЕГРЕСІЙНИХ МОДЕЛЕЙ

А. І. Купін, д-р техн. наук; І. О. Музика, аспірант,
Криворізький національний університет
MusicVano@mail.ru

Одним із головних завдань інтелектуального керування технологічними процесами (ТП) згідно із концепцією «чорної скриньки» є побудова математичної моделі за статистичними даними. Тому метою роботи є підвищення ефективності навчання регресійної моделі в режимі online за умов неоднорідності вхідної інформації. Для підвищення якості прогнозування отриманої математичної моделі доцільно перед процедурою навчання попередньо проводити кластерний аналіз статистичних даних (рис. 1).



Рис. 1. Етапи побудови математичної моделі

Аналіз існуючих алгоритмів кластеризації (K-середніх, мінімального покриваючого дерева, нечіткої, пошарової та ієрархічної кластеризації) показав властивий більшості методів недолік – необхідність апріорного визначення кількості кластерів або їх діаметра, що в умовах значних обсягів статистичних даних (понад 200) зробити досить складно. Дослідження показали, що ефективним в умовах неперервних функцій розподілу технологічних параметрів є алгоритм нечіткої кластеризації з динамічним настроюванням кількості та розмірів кластерів [1, 2].

Як критерій кластеризації використовується мінімум суми всіх зважених відстаней (1)

$$F = \sum_{j=1}^K \sum_{i=1}^N (\mu_{ji})^q \|X_i - C_j\| \rightarrow \min, \quad (1)$$

де i, j – номер вхідного вектора та кластера відповідно;

K, N – кількість кластерів та векторів відповідно; μ_{ji} – матриця приналежності i -го вхідного вектора X_i до j -го кластера заданого своїм центром C_j ; q – фіксований параметр, що настроюється; $\|X_i - C_j\|$ – евклідова міра відстані. Визначення кількості кластерів проводиться на основі аналізу екстремумів функцій густини розподілу статистичних даних за окремими змінними $f(x_i)$ (рис. 2).

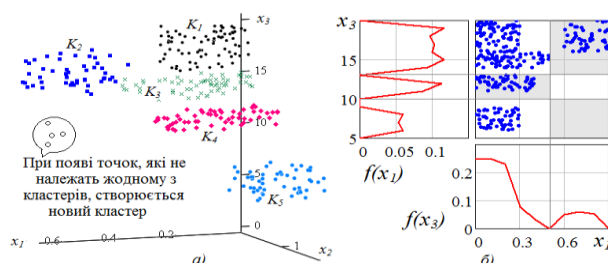


Рис. 2. Ідентифікація кластерів у статистичній вибірці (а) та визначення кількості кластерів на основі функцій густини розподілу (б)

Таким чином, за попередніми оцінками, застосування кластерного аналізу перед процедурою побудови регресійного рівняння дозволить підвищити якість прогнозування моделлю на 5–10%.

2. Вятчин Д. А. Нечеткие методы автоматической классификации: монография / Д. А. Вятчин. – Минск: Технопринт, 2004. – 219 с.

