

УДК 658.5

DOI 10.31471/1993-9981-2022-1(48)-109-114

## ПРОГНОЗУВАННЯ ОБ'ЄМНОЇ ПРОДУКТИВНОСТІ НАГНІТАЧА ІЗ ВИКОРИСТАННЯМ МЕТОДІВ КЛАСТЕРИЗАЦІЇ

*М. І. Козуляк, В. М. Гарасимів, Д. М. Магас*

*Інститут інформаційних технологій, Івано-Франківський національний технічний університет нафти і газу, вул. Карпатська, 15, м. Івано-Франківськ, 76019, e-mail: viraharasymiv78@gmail.com*

В даній роботі розглянуто методи прогнозування значень об'ємної продуктивності відцентрового нагнітача, в основі яких лежить підхід кластеризації. Їхня перевага над існуючими методами прогнозування полягає в тому, що вони враховують хаотичну складову експериментальних даних та включають в себе сучасні інформаційні технології (нечітку логіку, нейронні мережі, генетичні алгоритми). Завдяки великій кількості ітераційних підходів значення похибки цільової функції поступово зменшується до бажаного значення, що забезпечує досить хорошу якість прогнозування.

Метод прогнозування на основі алгоритму Fuzzy C-Means, який базується на розв'язанні нечітких с-середніх, реалізовано в програмному середовищі Matlab. Для даного методу кількість кластерів рівна двадцяти, а кількість ітераційних підходів коливається в межах ста. Обчислено величину середньоквадратичного відхилення, яка рівна 0.036816.

Також розглянуто інший метод прогнозування значень об'ємної продуктивності відцентрового нагнітача, який побудовано на основі алгоритму субтрактивної нечіткої кластеризації. Нечіткі правила синтезовано автономно із експериментальних даних, а для побудови нечіткої бази правил використано модель типу Сугено. Для кожної точки розраховано значення потенціалу, що показує можливість формування кластера в його околиці. Центри кластерів вибрано ітераційно серед точок з максимальними потенціалами. Встановлено, що значення середньоквадратичного відхилення для даного методу рівне 0.024289. Отримані результати підтверджують хорошу степінь збіжності експериментальних значень до прогнозованих.

Експериментальні дані, які використовувалися для апробації запропонованих методів, отримано при експлуатації відцентрового нагнітача типу 16 ГЦ2-395/53-76С у Долинському лінійному виробничому управлінні магістральними газопроводами за листопад місяць 2014 року.

**Ключові слова:** відцентровий нагнітач, об'ємна продуктивність, кластеризація, прогнозування, алгоритм Fuzzy C-Means, алгоритм субтрактивної нечіткої кластеризації.

Methods for predicting the volumetric productivity of a centrifugal supercharger, which is based on the clustering approach, are performed in this paper. Their advantage over existing forecasting methods is that they take into account the chaotic component of experimental data and include modern information technology (fuzzy logic, neural networks, genetic algorithms). Due to the large number of iterative approaches, the value of the error of the objective function is gradually reduced to the desired value, which provides a fairly good quality of forecasting.

The prediction method based on the Fuzzy C-Means algorithm, which is based on solving fuzzy c-means, is implemented in the Matlab software environment. For this method, the number of clusters is twenty, and the number of iterative approaches ranges from one hundred. The value of the standard deviation equal to 0.036816 is calculated.

Another method for predicting the volumetric performance of a centrifugal supercharger, which is based on the subtractive fuzzy clustering algorithm, is also considered. Fuzzy rules were synthesized autonomously from experimental data, and a Sugeno-type model was used to construct a fuzzy rule base. For each point, the value of the potential is calculated, which shows the possibility of forming a cluster in its vicinity. The centers of the clusters are selected iteratively from the points with maximum potentials. It is established that the value of the standard deviation for this method is 0.024289. The obtained results confirm a good degree of convergence of the experimental values to the predicted ones.

Experimental data used for testing of the proposed methods were obtained during the operation of the centrifugal supercharger type 16 ГЦ2-395/53-76С of Dolyna linear production administration of gas transmittal pipelines in November 2014.

**Keywords:** centrifugal supercharger, volumetric productivity, clustering, forecasting, Fuzzy C-Means algorithm, subtractive fuzzy clustering algorithm.

### Вступ

На даний час збільшується запит підприємств та організацій на системи управління, аналізу та прогнозування, які зможуть не лише накопичувати дані, а сприяти досягненню нових вказівок та знань у відповідності зі стандартами Data Mining.

Методи, які традиційно використовуються для прогнозування, базуються на побудові єдиної моделі, в якій число параметрів являється досить малим і залежить від числа експериментальних даних. Досвід показує, що таку модель одержати досить складно, а іноді є неможливим.

Тому на даний час актуально використовувати методи прогнозування, які враховують хаотичну складову даних із використанням сучасних підходів кластеризації. Існуючі сучасні підходи можна поділити на три основні групи відповідно до теорій штучного інтелекту, на які вони опираються [1]. До першої групи відносяться нейронні мережі, які є універсальними адаптивними апроксиматорами, що можуть виявляти та апроксимувати локальні тенденції у поведінці даних. До другої групи віднесемо нечіткі та нейронно-нечіткі підходи, які використовуються для побудови надійних та логічно прозорих моделей прогнозування. Третя група – це методи розподіленого штучного інтелекту: генетичні алгоритми, метод «мурашиних колоній», інтелект рою. Також ці методи можна застосувати для пошуку величин ваг в нейронних мережах.

### Аналіз літературних джерел

На сьогодні вирішення багатьох задач прогнозування та моделювання у різноманітних сферах науки реалізується із використанням методів кластеризації. Результуючі значення методів прогнозування із використанням часових рядів є актуальними для розв'язання задач планування, в економіці, торгівлі, керування, а також широко застосовуються для побудови інтелектуальних систем підтримки прийняття рішень [2-4].

Основоположниками теорії прогнозування з використанням інтелектуальних підходів є Н. Вінер, Н. А. Колмогоров, Р. Калман [5]. Для формалізації невизначеності класична теорія прогнозування використовує методи статистики та випадкових процесів.

Оскільки більшість процесів містять у собі хаотичну складову, то виникає необхідність розробки методів прогнозування, які дають можливість об'єднати у собі множину моделей, що відповідатимуть різним патернам, які міститимуться в експериментальних даних [6]. Тому серед багатьох методів набули популярності методи прогнозування, які базуються на основах кластеризації, де моделі формуються на основі кластерів послідовності експериментальних значень.

### Основна частина

Для забезпечення надійної та ефективної роботи ГПА (газоперекачувального агрегату) в умовах його експлуатації диспетчери КС (компресорної станції) контролюють теплотехнічні характеристики агрегату: потужність і к. к. д. (коефіцієнт корисної дії) ГТД (газотурбінного двигуна), продуктивність, політропний к.к.д., потужність ВЦН (відцентрового нагнітача), наявну приведену потужність ГТД, коефіцієнт технічного стану ГТД за потужністю, коефіцієнт навантаження ГТД, коефіцієнт технічного стану ВЦН за політропним к.к.д., коефіцієнт режиму ВЦН [7]. З метою зменшення ризику при прийнятті рішень використовують прогнозування теплотехнічних характеристик ГПА, точність якого залежить від використаних ресурсів та методів прогнозування. У випадку, коли розглядають об'єкти, які працюють в умовах хаотичної та випадкової невизначеності теорія прогнозування використовує апарат статистики та випадкових процесів.

Для апроксимації нелінійних залежностей використано нечітку логіку, основу якої становлять правила «ЯКЩО вхід, ТО вихід». Точність та якість нечіткої моделі залежить від форми і взаємного розташування функцій належності. Тому при використанні нейромережових нечітких систем, налаштування параметрів нечіткої моделі в значній мірі автоматизується, нейронна (адаптивна) мережа виступає в якості механізму налаштування параметрів нечіткої моделі. Нечіткі множини є параметрами нечіткої моделі, а структура правил та їхня кількість визначають структуру моделі.

В даний роботі використано експериментальні дані, які були зібрані на КС «Долина», отриманих при експлуатації

відцентрового нагнітача природного газу 16 ГЦ2-395/53-76С за листопад місяць 2014 року.

Об'ємна продуктивність компресора за умовами всмоктування в м<sup>3</sup>/хв визначалася за формулою:

$$Q_{ec} = A_k \cdot \sqrt{\frac{\Delta P \cdot Z_1 \cdot R \cdot T_1}{P_1}}, \quad (1)$$

де  $A_k$  – коефіцієнт конфузору (паспортні дані), м<sup>2</sup>;

$\Delta P$  – перепад тиску на конфузори, МПа;

$Z_1$  – коефіцієнт стисливості компримуємого газу на вході в нагнітач;

$R$  – газова стала газу, кДж/(кг·К);

$T_1$  – температура на вході в нагнітач, К;

$P_1$  – тиск газу на вході в нагнітач, МПа.

Спочатку синтезуються нечіткі правила з експериментальних даних з використанням субтрактивної кластеризації. На другому етапі налаштовуються параметрів нечіткої моделі за допомогою ANFIS-алгоритму. Субтрактивна кластеризація може використовуватися як швидкий автономний метод синтезу нечітких правил з даних. За результатами нечіткої кластеризації можна синтезувати нечіткі правила різних баз знань (синглтонної, Мамдані, Сугено). Ідея методу полягає в тому, що спочатку визначають точки, які можуть бути центрами кластерів. Для кожної точки розраховується значення потенціалу, що показує можливість формування кластера в її околиці. Після цього ітераційно вибираються центри кластерів серед точок з максимальними потенціалами.

Важливою перевагою застосування кластеризації для синтезу нечіткої моделі є те, що правила бази знань виходять об'єктно-орієнтованими.

У системі Matlab для розв'язання задачі нечіткої кластеризації реалізовано два алгоритми:

- алгоритм FCM (Fuzzy Classifier Means) - спеціальна функція системи Matlab, яка використовує алгоритм розв'язання методом нечітких с-середніх;

- алгоритм субтрактивної нечіткої кластеризації.

Метою FCM-алгоритму кластеризації є автоматична класифікація множини об'єктів, тобто даний алгоритм визначає кластери і

відповідно до них здійснює класифікацію об'єктів. Кластери представляються нечіткими множинами. Границі між кластерами також є нечіткими. Степінь належності об'єкта визначається відстанню від нього до відповідних кластерних центрів. Даний алгоритм ітераційно обчислює центри кластерів і нові степені належності об'єктів.

Для множини експериментальних значень  $K=336$  вхідних векторів  $x_k$  та  $N$  виділених кластерів  $c_j$  припустимо, що ступінь належності будь-якого вектору  $x_k$  до відповідного кластеру  $c_j$  рівна  $\mu_{jk} \in [0,1]$ , де  $j$  – номер кластера, а  $k$  – номер вхідного вектора.

Запишемо умови нормування для  $\mu_{jk}$ :

$$\sum_{j=1}^N \mu_{jk} = 1, \forall k = 1, \dots, K; \quad (2)$$

$$0 < \sum_{k=1}^N \mu_{jk} \leq K, \forall j = 1, \dots, N.$$

Метою алгоритма є мінімізація суми усіх зважених відстаней:

$$\sum_{j=1}^N \sum_{k=1}^K (\mu_{jk})^q \|x_k - c_j\| \rightarrow \min, \quad (3)$$

де  $q$  – фіксований параметр, заданий перед ітераціями.

Із врахуванням умови (3) необхідно розв'язати наступну систему рівнянь:

$$\begin{cases} \frac{\partial}{\partial \mu_{jk}} \left( \sum_{j=1}^N \sum_{k=1}^K (\mu_{jk})^q \|x_k - c_j\| \right) = 0, \\ \frac{\partial}{\partial c_j} \left( \sum_{j=1}^N \sum_{k=1}^K (\mu_{jk})^q \|x_k - c_j\| \right) = 0. \end{cases} \quad (4)$$

Відповідно до умов (2) розв'язок системи (4) запишемо у такому вигляді:

$$c_j = \frac{\sum_{k=1}^K (\mu_{jk})^q \cdot x_k}{\sum_{k=1}^K (\mu_{jk})^q}, \quad (5)$$

$$\mu_{jk} = \frac{1}{\|x_k - c_j\|^{\frac{1}{q-1}} \left( \sum_{j=1}^N \frac{1}{\|x_k - c_j\|^{\frac{1}{q-1}}} \right)}. \quad (6)$$

Реалізація алгоритму FCM включає виконання наступних кроків.

Крок 1. Ініціалізація. Тут вибираємо необхідну кількість кластерів  $N$ , кількість ітерацій та фіксований параметр  $q$ . Оскільки  $2 < N < K$ , то припустимо, що  $N=20$ . Кількість ітерацій рівна 100, а фіксований параметр  $q=1.5$ . Початкова матриця належності рівна  $U^{(0)} = (\mu_{jk})^{(0)}$ .

Крок 2. Обчислення значень центрів кластерів  $c_j^{(t)}$ . На ітераційному кроці  $t$  при відомій матриці  $\mu_{jk}^{(t)}$  виконуємо обчислення  $c_j^{(t)}$  із використанням формули (5).

Крок 3. Визначення значень функції належності  $\mu_{jk}$ . Якщо  $x_k \neq c_j$  обчислюємо  $\mu_{jk}^{(t)}$ , враховуючи відомі значення  $c_j^{(t)}$ . В іншому випадку:

$$\mu_{jk}^{(t+1)} = \begin{cases} 1, & l = j, \\ 0, & \text{інакше} \end{cases}$$

Крок 4. Зупинка алгоритму. Алгоритм нечіткої кластеризації зупиняється у випадку виконання наступної умови:

$$|U^{(t+1)} - U^t| \leq \varepsilon,$$

де  $\varepsilon = 10^{-5}$  - задана точність виконання алгоритму.

Результат виконання писаного вище алгоритму показано на рисунку 1.

Для оцінювання точності прогнозування розраховано середньоквадратичне відхилення прогнозованих значень від розрахованих значень об'ємної продуктивності за такою формулою [2]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1, n} (y_i - \hat{y}_i)^2}, \quad (7)$$

де  $n=336$  - кількість досліджень;

$y_i$  - розрахункові значення об'ємної продуктивності нагнітача,  $\text{м}^3/\text{с}$ ;

$\hat{y}_i$  - прогнозовані за алгоритмом FCM значення об'ємної продуктивності нагнітача,  $\text{м}^3/\text{с}$ .

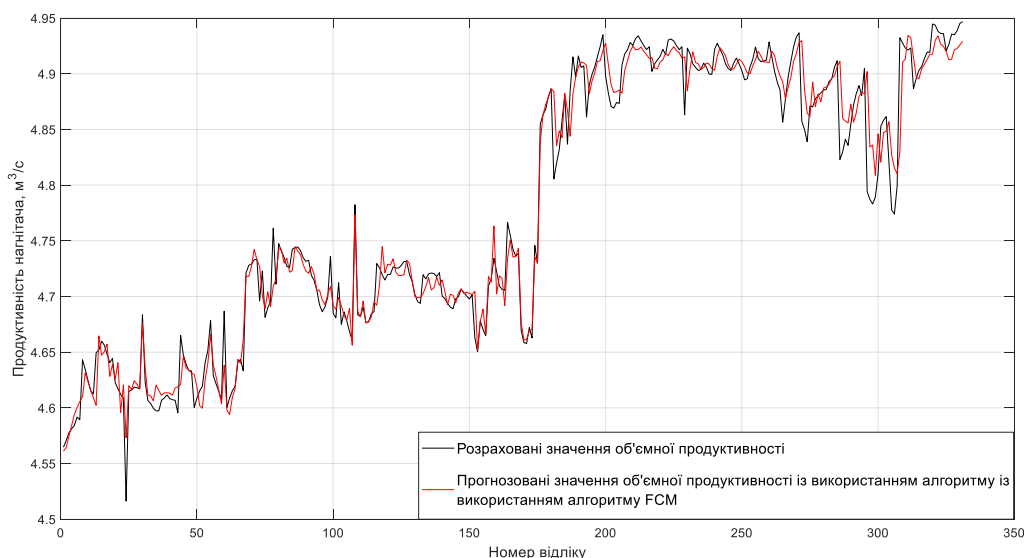


Рисунок 1 – Розрахункові та прогнозовані за алгоритмом FCM значення об'ємної продуктивності нагнітача

Одержали значення середньоквадратичного відхилення рівне 0.036816.

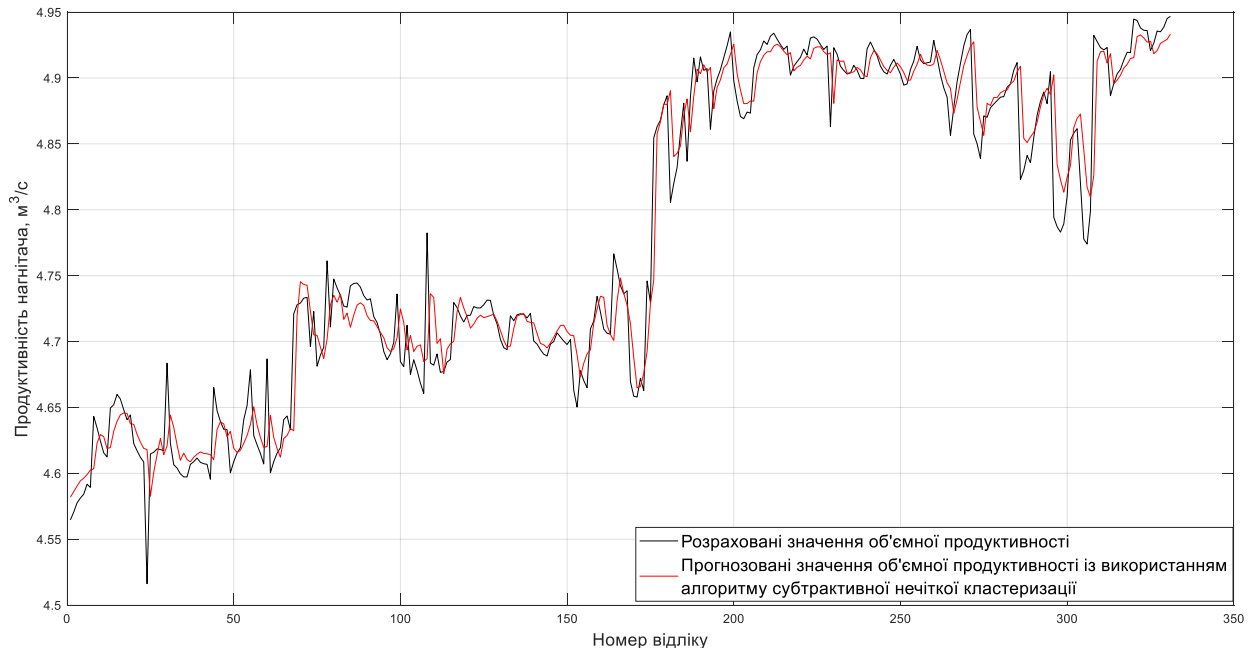
Для реалізації алгоритму субтрактивної нечіткої кластеризації використано пакет Matlab Fuzzy Logic Toolbox, який автоматично синтезує

нечітку базу знань, використовуючи експериментальні дані, для моделі типу Сугено. В результаті генеруються правила, які відповідають областям найбільшої концентрації експериментальних даних. Даний алгоритм реалізований у вигляді функції genfis2.

Результат виконання алгоритму субтрактивної нечіткої кластеризації показано на рисунку 2.

Процес екстракції правил із експериментальних даних у функції *genfis2* можна поділити на два етапи. Спочатку

використовуємо функцію *subclust* для визначення кількості правил. Потім із використанням МНК (методу найменших квадратів) визначається частина «...то...» для кожного правила.



**Рисунок 2 – Розрахункові та прогнозовані за алгоритмом субтрактивної нечіткої кластеризації значення об'ємної продуктивності нагнітача**

Одним із вхідних параметрів функції є вектор  $r$ , який визначає розміри області правил за кожною координатою. Відомо, що значення вектора  $r$  повинно знаходитися в межах  $[0,1]$  так, як під час виконання функції *subclust* дані масштабуються на одиничний гіперкуб. Вибір радіуса  $r$  помітно впливає на результат прогнозування. Якщо прийняти невелике значення  $r$ , то нечітка база знань буде більш повною, але чутливо до виключень і неточностей. Якщо прийняти значення радіуса досить великим, то можливо втратити деякі правила під час синтезу моделі. За рекомендаціями праці [8] вибираємо  $r=0.55$ . Кількість підходів вибрано 100.

Для оцінювання точності прогнозування використано формулу (7) та одержано значення середньоквадратичного відхилення рівне 0.024289.

## Висновки

В даній роботі запропоновано метод прогнозування значень об'ємної продуктивності нагнітача, основою якого є алгоритм Fuzzy C-Means. Обчислено величину середньоквадратичного відхилення, яка рівна 0.036816.

Також виконано прогнозування значень об'ємної продуктивності нагнітача із застосуванням методу на основі алгоритму субтрактивної нечіткої кластеризації. Нечіткі правила синтезовано автономно із експериментальних даних, а для побудови нечіткої бази правил використано модель типу Сугено. Встановлено, що значення середньоквадратичного відхилення для даного методу рівне 0.024289. Отримані результати підтверджують хорошу якість прогнозування значень об'ємної продуктивності нагнітача.

### Список використаних джерел

1. Wang J., Chi D., Wu J., Lu H. Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting. *Expert Systems with Applications*. 2011. № 38. P. 8419-8429.
2. Громов А.В., Воронин И.М., Гатыло В.Р., Прокопало Е.Т. Оценка гиперпараметров в задачах прогнозирования на основе кластеризации. *Питання прикладної математики і математичного моделювання*. 2016. № 16. С. 25-46.
3. Widiputra H., Kho H., Pears R., Kasabov N. A novel evolving clustering algorithm with polynomial regression for chaotic time-series prediction. *Neural Inf Process*. 2009. Vol. 5864. P. 114-121.
4. Пахомова В.М., Дмитрієв С.Ю. Розробка підсистеми оперативного прогнозування простоїв прибуваючих поїздів на основі ANFIS-системи. *Інформаційно-керуючі системи на залізничному транспорті*. 2013. №4. С. 46-55.
5. Chatfield C. *The Analysis of Time Series : An Introduction, Sixth Edition*. CRC Press. 2009. 352 p.
6. Phu L., Anh D.T. Mostif-based method for initialization the K-means clustering for time series data. *Springer*. 2011. Vol. 7106. P. 11-20.
7. Довідник інженера диспетчерської служби / за заг. ред. канд. техн. наук Ю.В. Пономарьова та М.П. Химка. Київ : УЦЕБОПнафтогаз, 2009. 245 с.
8. Ремнева В.В. Особенности применения критериев оценки качества разбиения в алгоритме кластеризации Fuzzy C-Means. *Сучасні інформаційні та інноваційні технології на транспорті*. 2013. № 1. С. 157-158.

### References

1. Wang J., Chi D., Wu J., Lu H. Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting. *Expert Systems with Applications*. 2011. № 38. P. 8419-8429.
2. Hromov A.V., Voronyn Y.M., Hatylo V.R., Prokopalo E.T. Otsennka hyperparametrov v zadachakh prohozyrovanyya na osnove klasteryzatsyy. *Pytannya prykladnoyi matematyky i matematychnoho modelyuvannya*. 2016. № 16. S. 25-46.
3. Widiputra H., Kho H., Pears R., Kasabov N. A novel evolving clustering algorithm with polynomial regression for chaotic time-series prediction. *Neural Inf Process*. 2009. Vol. 5864. P. 114-121.
4. Pakhomova V.M., Dmitriyev S.YU. Rozrobka pidsystemy operatyvnoho prohozuvannya prostoyiv prybuyayuchykh poyzdiv na osnovi ANFIS-systemy. *Informatsiyno-keruyuchi systemy na zaliznychnomu transporti*. 2013. №4. S. 46-55.
5. Chatfield C. *The Analysis of Time Series : An Introduction, Sixth Edition*. CRC Press. 2009. 352 p.
6. Phu L., Anh D.T. Mostif-based method for initialization the K-means clustering for time series data. *Springer*. 2011. Vol. 7106. P. 11-20.
7. Dovidnyk inzhenera dyspetchers'koyi sluzhby / za zah. red. kand. tekhn. nauk YU.V. Ponomar'ova ta M.P. Khymka. Kyiv : UTSEBOPnaftohaz, 2009. 245 s.
8. Remneva V.V. Features of application of partition quality evaluation criteria in Fuzzy C-Means clustering algorithm. *Modern information and innovative technologies in transport*. 2013. № 1. pp. 157-158.