

Мулеца О. Ю.

## РОЗРОБКА ЕВОЛЮЦІЙНИХ МЕТОДІВ СТРУКТУРНОЇ І ПАРАМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ ТАБЛИЧНИХ ЗАЛЕЖНОСТЕЙ

*Розглядається задача структурної і параметричної ідентифікації табличних залежностей. Розроблено еволюційний метод структурної ідентифікації, який визначає оптимальну за заданим критерієм структуру функції. Розроблено еволюційний метод параметричної ідентифікації, який, на основі дослідження характеру вхідних даних, визначає параметри функції, яка їх характеризує. Виконано експериментальну верифікацію розроблених методів за допомогою однофакторного та двофакторного аналізу.*

**Ключові слова:** структурна ідентифікація, параметрична ідентифікація, таблична залежність, еволюційний метод.

### 1. Вступ

Задачі структурної і параметричної ідентифікації табличних залежностей розглядаються, як правило, в контексті розв'язування задач штучного інтелекту. Вони переважно пов'язані з дослідженням параметрів медичних, економічних, соціальних та інших систем.

Розробці моделей і методів ідентифікації присвячені численні наукові публікації. Проте підбір і застосування методів для розв'язання реальних прикладних задач часто супроводжуються труднощами, пов'язаними з певними особливостями вхідних даних (наприклад, їх невеликий об'єм). Також, зважаючи на зміст вхідних даних, часто виникає необхідність не лише ідентифікувати функцію в аналітичному вигляді, яка характеризує ці дані, а і задати такі характеристики функції, як парність, періодичність, монотонність, область значень тощо.

Важливою особливістю великої частини методів ідентифікації є те, що в результаті їх роботи неможливо визначити аналітичний вигляд результуючої функції, що не дозволяє проводити подальший її аналіз. До таких методів, наприклад, належать нейромережні методи ідентифікації. Методи ж, результатом застосування яких є функція в явному виді, часто характеризуються обчислювальною складністю та необхідністю використання додаткових інструментів для наближеного розв'язування систем нелінійних рівнянь.

Таким чином, актуальним є розробка моделей і методів структурної і параметричної ідентифікації, які б дозволяли оперувати невеликими вибірками та не вимагали використання додаткового математичного апарату.

Дослідження присвячене розробці еволюційних методів структурної і параметричної ідентифікації, які дозволяють ідентифікувати структуру функції із заданих базисних функцій, а також визначити параметри функції вже заданої структури. В основі розроблених алгоритмів лежить еволюційна стратегія, застосування якої дозволяє уникнути складних математичних перетворень, розв'язування систем нелінійних рівнянь багатьох змінних та підвищує оптимальність результатів ідентифікації.

### 2. Об'єкт дослідження та його технологічний аудит

Об'єкт дослідження — моделі і методи структурної і параметричної ідентифікації табличних залежностей та їх застосування для розв'язання прикладних задач ідентифікації параметрів системи в різних початкових умовах:

- при вхідних даних, що мають невеликий об'єм;
- коли є необхідність отримання результуючої залежності в аналітичному вигляді;
- з метою забезпечення мінімізації критерію оптимальності та додаткових властивостей результуючої функції.

Важливою особливістю моделей і методів структурної і параметричної ідентифікації є їх неефективність при розв'язуванні задач з вхідними даними малих об'ємів. Також у випадках, коли залежність виражається складними математичними виразами, елементами яких є зокрема тригонометричні, ірраціональні, логарифмічні та інші функції, для ідентифікації її параметрів, як правило, необхідним є розв'язування систем нелінійних рівнянь багатьох змінних, що саме по собі часто є нетривіальною оптимізаційною задачею.

### 3. Мета та задачі дослідження

Метою роботи є розробка релевантних моделей і методів структурної і параметричної ідентифікації табличних залежностей для ефективного розв'язання задач з невеликими вхідними даними та ідентифікації аналітичних залежностей складної структури.

В межах зазначеної мети було поставлено такі задачі:

1. Виконати математичну постановку задачі структурної і параметричної ідентифікації табличних залежностей.
2. Розробити методи структурної і параметричної ідентифікації на основі еволюційної технології.
3. Виконати експериментальну верифікацію розроблених методів для розв'язання прикладних задач.

4. Здійснити порівняльний аналіз результатів роботи розроблених методів ідентифікації з результатами ідентифікації іншими відомими методами.

**4. Аналіз літературних даних**

Задача ідентифікації невідомих залежностей, як правило, виникає при дослідженні слабкоструктурованих та погано формалізованих задач штучного інтелекту [1]. Аналіз наукових джерел показав, що більшість методів, які використовуються при розв'язуванні таких задач, базуються на регресійних моделях і методах. Перевагою таких підходів є добре розвинений математичний апарат, за допомогою якого можна оцінити точність та адекватність побудованих моделей. Тому, поряд з класичними регресійними методами, активно розвиваються і сучасні регресійні методи ідентифікації систем. Так, в [2] представлено метод побудови нелінійної моделі авторегресії, яку можна використовувати при описі широкого класу динамічних систем. Представлений метод є узагальненням алгоритму крокової регресії LARS [3], який, в свою чергу, полягає в послідовному додаванні ознак при побудові регресійної моделі. Широкого застосування при розв'язуванні задач ідентифікації набули нейромережні методи. Так, в роботі [4] пропонується нейромережний метод ідентифікації параметрів системи, який встановлює математичні співвідношення між вагами мережі та параметрами функції активації. Такий підхід покращує ефективність роботи мереж зі зворотнім поширенням похибок. Для ідентифікації параметрів систем, які описуються диференціальними рівняннями, в [5] запропонований метод синтезу та навчання диференціальної нейронної мережі. Використання нечітких нейронних мереж для ідентифікації параметрів системи, як показано в наукових джерелах, дозволяє підвищити ефективність процесів ідентифікації. Так, в [6] представлений гібридний інтелектуальний метод ідентифікації нечіткими нейронними мережами, у яких функції активації представлені як вейвлет перетворення. Нечітка стохастична модель нейронної мережі для структурної ідентифікації системи описана в [7]. Особливостями нейромережних методів є те, що вони, як правило, не дозволяють отримати залежності в аналітичному виді. Також для успішного навчання мережі потребуються вибірки значних об'ємів.

Ще одним класом алгоритмів, які використовуються для розв'язування задач штучного інтелекту, є еволюційні алгоритми. Успішне їх використання дозволяє підвищувати ефективність розв'язків. Теоретичні викладки еволюційних стратегій, які можуть використовуватися при розв'язуванні задач ідентифікації, наведені в [8]. Робота містить теоретичне обґрунтування переваг еволюційних алгоритмів в порівнянні з класичними методами ідентифікації. В свою чергу у [9] наведено каскадний еволюційний алгоритм ідентифікації, що поєднує в собі класичний генетичний алгоритм та еволюційні алгоритми, які використовуються для вибору входів та оцінки параметрів RBF-нейронних мереж.

Таким чином, аналіз наукових джерел показав, що задача ідентифікації параметрів системи загалом та ідентифікації табличних залежностей зокрема, широко досліджується та описана в науковій літературі. Проте, наведені алгоритми переважно не дозволяють отримати залежність в аналітичному вигляді, а також погано

працюють для вибірок малого об'єму. Тому актуальною є розробка моделей і методів структурної і параметричної ідентифікації залежностей, які б дозволяли враховувати вказані особливості.

**5. Матеріали та методи дослідження**

Розглянемо задачу структурної і параметричної ідентифікації табличних залежностей у такій постановці [1]: нехай дано початкові дані у вигляді табл. 1, де  $X_1, \dots, X_n$  – вектор вхідних факторів,  $Y$  – результуюча характеристика,  $m$  – кількість статистичних спостережень або експериментів. Необхідно апроксимувати дану табличну залежність аналітичними виразами, оптимальними за заданими критеріями.

Таблиця 1

Початкові дані для ідентифікації

$X_1$	$X_2$	...	$X_n$	$Y$
$x_{11}$	$x_{12}$	...	$x_{1n}$	$y_1$
$x_{21}$	$x_{22}$	...	$x_{2n}$	$y_2$
...	...	...	...	...
$x_{m1}$	$x_{m2}$	...	$x_{mn}$	$y_m$

Таким чином, необхідно побудувати таку функцію виду (1).

$$Y = F(X_1, X_2, \dots, X_n), \tag{1}$$

яка б була оптимальною за деякими критеріями.

Серед критеріїв оптимальності можна виділити критерії мінімізації середньоквадратичної та середньої відносної похибок та інші. Тобто, функція  $F$  вибирається за одним з правил (2)–(5):

$$F \in \text{Arg} \left\{ \min_{F \in \Phi} \sqrt{\frac{\sum_{i=1}^m (F(x_{i1}, x_{i2}, \dots, x_{in}) - y_i)^2}{m}} \right\}, \tag{2}$$

$$F \in \text{Arg} \left\{ \min_{F \in \Phi} \frac{1}{m} \cdot \sum_{i=1}^m \left| \frac{y_i - F(x_{i1}, x_{i2}, \dots, x_{in})}{y_i} \right| \right\}, \tag{3}$$

$$F \in \text{Arg} \left\{ \min_{F \in \Phi} \left\{ \max_{i=1, n} \left| \frac{y_i - F(x_{i1}, x_{i2}, \dots, x_{in})}{y_i} \right| \right\} \right\}, \tag{4}$$

$$F \in \text{Arg} \left\{ \min_{F \in \Phi} \left\{ \max_{i=1, n} |y_i - F(x_{i1}, x_{i2}, \dots, x_{in})| \right\} \right\}, \tag{5}$$

де  $\Phi$  – деякий простір всіх допустимих функцій.

Як правило, якщо задача полягає у визначенні аналітичного вигляду функції (1), то процес ідентифікації залежності складається з двох етапів:

1. Структурна ідентифікація – визначення структури функції (1).
2. Параметрична ідентифікація – уточнення параметрів функції.

Якщо відомо, що функція (1) має лінійний вид, то задача зводиться до побудови множинної лінійної регресії і може бути розв'язаною відомими методами. В загальному ж випадку задача побудови нелінійної функції (1) аналітичними методами математичної статистики не розв'язується. Таким чином, актуальною є розробка евристичних методів розв'язання задачі структурної і параметричної ідентифікації, які дають адекватні результати.

## 6. Результати дослідження

**6.1. Розробка еволюційного методу параметричної ідентифікації.** Нехай початкові дані задані табл. 1, а також відомо, що таблична залежність апроксимується функцією виду (6):

$$F(a_1, \dots, a_q) = \sum_{s=1}^q a_s \cdot f_s(X_1, X_2, \dots, X_n), \quad (6)$$

де  $f_s$  – функції  $n$  змінних.

Необхідно визначити такі значення параметрів  $a_1, \dots, a_q$ , при яких (6) буде оптимальною за певним критерієм. Серед критеріїв оптимальності, як правило використовують критерій мінімізації середньоквадратичної похибки, середньої відносної похибки та інші.

Пропонується еволюційний метод параметричної ідентифікації, в основі якого лежить генетичний алгоритм, основні етапи якого подано в [10]. В такому випадку задача параметричної ідентифікації полягає у знаходженні мінімального значення однієї з функцій:

$$Z_1(a_1, a_2, \dots, a_q) = \sqrt{\frac{\sum_{i=1}^m \left( y_i - \sum_{s=1}^q a_s \cdot f_s(x_{i1}, x_{i2}, \dots, x_{in}) \right)^2}{m}}, \quad (7)$$

у випадку мінімізації середньоквадратичної похибки;

$$Z_1(a_1, a_2, \dots, a_q) = \frac{1}{m} \cdot \sum_{i=1}^m \left| \frac{y_i - \sum_{s=1}^q a_s \cdot f_s(x_{i1}, x_{i2}, \dots, x_{in})}{y_i} \right|, \quad (8)$$

у випадку мінімізації середньої відносної похибки;

$$Z_1(a_1, a_2, \dots, a_q) = \max_{i=1, m} \left| y_i - \sum_{s=1}^q a_s \cdot f_s(x_{i1}, x_{i2}, \dots, x_{in}) \right|, \quad (9)$$

при мінімізації максимального відхилення від точного значення, тощо.

Варто зазначити, що зменшення значення цільових функцій (7)–(9) можна забезпечити багаторазовим виконанням запропонованого алгоритму для одних і тих же вхідних даних та вибору того набору параметрів  $a_1, \dots, a_q$ , який забезпечує мінімальне значення цільової функції.

**6.2. Еволюційний метод структурної ідентифікації.** Нехай початкові дані задані табл. 1. Розглянемо схему структурної ідентифікації табличних залежностей.

Нехай на початковому етапі задано систему лінійно незалежних базисних функцій однієї чи декількох змінних (10):

$$\{f_1(X_1, X_2, \dots, X_n), f_2(X_1, X_2, \dots, X_n), \dots, f_k(X_1, X_2, \dots, X_n)\}. \quad (10)$$

Розглянемо функціонал виду:

$$L(\alpha_1, \alpha_2, \dots, \alpha_k) = \sum_{s=1}^k \alpha_s \cdot a_s \cdot f_s(X_1, X_2, \dots, X_n), \quad (11)$$

де  $\alpha_s \in \{0, 1\}$ ,  $\alpha_s \in R$ ,  $s = 1, k$ .

Необхідно знайти такі значення компонент биевального вектору  $(\alpha_1, \alpha_2, \dots, \alpha_k)$ , при яких функціонал (11) задаватиме сім'ю функцій, що найбільш оптимально апроксимують дані задані табл. 1.

Задамо функцію  $g(\alpha_1, \alpha_2, \dots, \alpha_k)$ , яка кожному набору значень компонент вектора  $(\alpha_1, \alpha_2, \dots, \alpha_k)$  ставить у відповідність ціле число за таким правилом:

$$g(\alpha_1, \alpha_2, \dots, \alpha_k) = \sum_{s=1}^k \alpha_s \cdot 2^{s-1}. \quad (12)$$

Тоді, для всіх  $g = \overline{1, 2^k - 1}$  виконаємо послідовно такий алгоритм:

Крок 1. Визначаємо структуру функції  $F$  у вигляді (6), виключивши з розгляду доданки, які перетворюються в 0, та перенумерувавши їх від 1 до  $q$ , де  $q$  – кількість ненульових елементів вектора  $(\alpha_1, \alpha_2, \dots, \alpha_k)$ .

Крок 2.  $\tau = \tau_0$  ( $1 \leq \tau_0 < m$ ).

Крок 3. Виконуємо алгоритм еволюційного методу параметричної ідентифікації для даних табл. 1 з номерами від 1 до  $\tau$ . Нехай  $F = \sum_{s=1}^q a_s^\tau \cdot f_s(X_1, X_2, \dots, X_n)$ .

Крок 4. Обчислюємо середньоквадратичну похибку на рядках табл. 1 з номерами від 1 до  $\tau$ :

$$\Delta_{g\tau} = \sqrt{\frac{\sum_{i=1}^{\tau} \left( y_i - \sum_{s=1}^q a_s^\tau \cdot f_s(x_{i1}, x_{i2}, \dots, x_{in}) \right)^2}{\tau}}. \quad (13)$$

Крок 5. Обчислюємо похибку ідентифікованої функції  $F$  на рядках табл. 1 з номерами від  $\tau + 1$  до  $m$ :

$$\delta_{g\tau} = \sqrt{\frac{\sum_{i=\tau+1}^m \left( y_i - \sum_{s=1}^q a_s^\tau \cdot f_s(x_{i1}, x_{i2}, \dots, x_{in}) \right)^2}{m - \tau}}. \quad (14)$$

Крок 6.  $\tau = \tau + 1$ . Якщо  $\tau < m$ , то переходимо до Кроку 3, інакше – кінець алгоритму.

З даними, отриманими в результаті виконання такого алгоритму будемо таблицю у виді табл. 2.

Таблиця 2

Похибки ідентифікації

g	$\Delta_{g\tau_0}$	$\delta_{g\tau_0}$	$\Delta_{g\tau_0+1}$	$\delta_{g\tau_0+1}$	...	$\Delta_{gm-1}$	$\delta_{gm-1}$
1	$\Delta_{1\tau_0}$	$\delta_{1\tau_0}$	$\Delta_{1\tau_0+1}$	$\delta_{1\tau_0+1}$	...	$\Delta_{1m-1}$	$\delta_{1m-1}$
2	$\Delta_{2\tau_0}$	$\delta_{2\tau_0}$	$\Delta_{2\tau_0+1}$	$\delta_{2\tau_0+1}$	...	$\Delta_{2m-1}$	$\delta_{2m-1}$
...	...	...	...	...	...	...	...
$2^k - 1$	$\Delta_{2^k-1,\tau_0}$	$\delta_{2^k-1,\tau_0}$	$\Delta_{2^k-1,\tau_0+1}$	$\delta_{2^k-1,\tau_0+1}$	...	$\Delta_{2^k-1,m-1}$	$\delta_{2^k-1,m-1}$

Для визначення оптимальної структури функціоналу (11) знайдемо номер  $g^* \in \{1, 2, \dots, 2^k - 1\}$  за таким правилом:

$$g^* \in \text{Arg} \min_{g=1, 2^k} \sum_{\tau=\tau_0}^{m-1} (\Delta_{g\tau} + \delta_{g\tau}). \quad (15)$$

Тоді оптимальною в системі базисних функцій (10) буде функція виду (6), яка отримується з (11) при визначеному  $g^*$ .

Таким чином, загальний метод структурної і параметричної ідентифікації табличних залежностей може бути представлений такою послідовністю кроків:

Крок 1. Завдання системи лінійно незалежних базисних функцій (10).

Крок 2. Виконання алгоритму еволюційного методу структурної ідентифікації для визначення (6).

Крок 3. Виконання алгоритму еволюційного методу параметричної ідентифікації для даних табл. 1.

У випадку, коли у записах функцій з (10) є невізначені параметри, вони визначаються еволюційним методом параметричної ідентифікації поряд з параметрами структури функціоналу (10), що не впливає на опис та роботу запропонованих алгоритмів.

**6.3. Експериментальна верифікація еволюційного методу параметричної ідентифікації.** Для виконання експериментальної верифікації розробленого еволюційного методу ідентифікації табличних залежностей розглянемо декілька прикладів.

Приклад 1 – ідентифікація впливу явища трудової міграції дорослого населення на показники захворюваності на ВІЛ-інфекцію [11] за допомогою однофакторного аналізу. Проаналізуємо статистичні дані, які характеризують захворюваність на ВІЛ-інфекцію в Закарпатській області в різні роки (табл. 3).

Таблиця 3

Захворюваність на ВІЛ-інфекцію трудових мігрантів і дорослого населення Закарпатської області за період 2007–2013 років (на 100 тис. дорослого населення)

Рік	Захворюваність на ВІЛ-інфекцію серед дорослих осіб (Y)	Захворюваність на ВІЛ-інфекцію трудових мігрантів (X)
2007	3,05	1,94
2008	2,74	1,40
2009	2,42	1,51
2010	4,74	2,59
2011	5,37	2,70
2012	6,74	3,67
2013	7,68	4,1

Після виконання методу структурної ідентифікації в системі базисних функцій  $\{1, X, \sin(a \cdot X + b), X^a\}$ , ( $X > 1$ ), було визначено характер залежності у вигляді:

$$Y(a_1, a_2, a_3, a_4, a_5) = a_1 X + a_2 \cdot \sin(a_3 X + a_4) + a_5. \quad (16)$$

Для аналізу результатів роботи еволюційного методу структурної і параметричної ідентифікації розглянемо три випадки:  $\tau = 5$ ,  $\tau = 6$  та ідентифікація функції для всіх записів табл. 3 ( $\tau = 7$ ).

Результати ідентифікації при запуску еволюційного методу параметричної ідентифікації з верхнім значенням параметру циклу 100 були такими:

для  $\tau = 5$ :

$$Y_{e5}(X) = 1,3084 \cdot X + 0,6878 \cdot \sin(13,0125 \cdot X - 39,7216) + 1,2067; \quad (17)$$

для  $\tau = 6$ :

$$Y_{e6}(X) = 1,6884 \cdot X - 0,5759 \cdot \sin(-12,8763 \cdot X - 10,8845) + 0,299; \quad (18)$$

для  $\tau = 7$ :

$$Y_{e7}(X) = 2,0175 \cdot X + 0,5638 \cdot \sin(-20,4792 \cdot X - 69,6521) - 0,5175. \quad (19)$$

Обчислимо значення середньоквадратичних похибок за формулами:

$$\Delta_{v\tau} = \sqrt{\frac{\sum_{s=1}^{\tau} (y_s - Y_{v\tau}(x_{s1}, x_{s2}, \dots, x_{sn}))^2}{\tau}}, \quad (20)$$

$$\delta_{v\tau} = \sqrt{\frac{\sum_{s=\tau+1}^m (y_s - Y_{v\tau}(x_{s1}, x_{s2}, \dots, x_{sn}))^2}{m - \tau}}, \quad (21)$$

де  $v \in \{q, r\}$ , та для аналізу отриманих результатів порівняємо їх з результатами отриманими за допомогою побудови рівняння прямої лінії регресії, які при різних значеннях  $\tau$  будуть мати такий вид:

для  $\tau = 5$ :

$$Y_{r5}(X) = 2,107158 \cdot X - 0,60932; \quad (22)$$

для  $\tau = 6$ :

$$Y_{r6}(X) = 1,964574 \cdot X - 0,34513; \quad (23)$$

для  $\tau = 7$ :

$$Y_{r7}(X) = 1,957501 \cdot X - 0,33126. \quad (24)$$

Результати обчислень наведено в табл. 4.

Таблиця 4

Зведені значення результатів ідентифікації залежності захворюваності на ВІЛ-інфекцію трудових мігрантів і дорослого населення Закарпатської області за період 2007–2013 років (на 100 тис. дорослого населення) різними методами

Рік	Y	$Y_{e5}(X)$	$Y_{r5}(X)$	$Y_{e6}(X)$	$Y_{r6}(X)$	$Y_{e7}(X)$	$Y_{r7}(X)$
2007	3,05	3,096611	3,47857	3,018518	3,466146	3,090312	3,466288
2008	2,74	2,716556	2,340705	2,320229	2,405276	2,760166	2,409238
2009	2,42	2,535726	2,572492	2,338352	2,621379	2,503716	2,624563
2010	4,74	4,774907	4,848223	4,815451	4,743119	4,803829	4,738664
2011	5,37	5,421853	5,08001	5,43085	4,959222	5,300563	4,953989
2012	6,74	6,68518	7,123953	7,0712	6,864859	6,721457	6,852766
2013	7,68	7,172246	8,030031	7,652379	7,709626	7,576245	7,694491
$\Delta_{\tau}$		0,06328	0,304029	0,224705	0,291579	0,064415	0,270063
$\delta_{\tau}$		0,361123	0,367384	0,027621	0,029626	—	—

Примітка:  $v \in \{q, r\}$ .

Перевіримо моделі (17)–(19) на адекватність за критерієм Фішера, а також порівняємо їх з такими ж результатами для моделей (22)–(24). Результати обчислень коефіцієнтів Фішера наведено у табл. 5.

Таблиця 5

Розраховані значення коефіцієнтів Фішера для побудованих моделей при ідентифікації залежності захворюваності на ВІЛ-інфекцію трудових мігрантів і дорослого населення Закарпатської області за період 2007–2013 років (на 100 тис. дорослого населення)

$\tau$	Табличне значення коефіцієнта Фішера ( $F_{\text{табл}}, p = 0,01$ )	$F_{e\tau}$	$F_{r\tau}$
5	4,604095	1019,867	41,44165
6	4,032143	244,6005	111,5152
7	3,707428	4115,244	242,3011

Як бачимо з табл. 4 і табл. 5 для моделей, які побудовані еволюційним методом параметричної ідентифікації залежності захворюваності на ВІЛ-інфекцію трудових мігрантів і дорослого населення Закарпатської області за період 2007–2013 років (на 100 тис. дорослого населення) коефіцієнти Фішера перевищують відповідні коефіцієнти Фішера лінійних регресійних моделей. Це свідчить про те, що ступінь адекватності запропонованих в роботі моделей є кращим за ступінь адекватності лінійних регресійних моделей. Також еволюційним моделям відповідають менші значення середньоквадратичних похибок.

Приклад 2 — ідентифікація впливу трудової міграції на показники захворюваності на ВІЛ-інфекцію за допомогою двофакторного аналізу (коли залежність поширення ВІЛ-інфекції визначається від двох факторів, а саме від кількості трудових мігрантів та напрямку їх міграції за адміністративними територіями Закарпатської області [12]). Розглянемо статистичні дані, які характеризують трудову міграцію та поширеність ВІЛ-інфекції в різних районах Закарпатської області (табл. 6).

Таблиця 6

Дані по трудовій міграції та поширеності ВІЛ-інфекції в Закарпатській області

№ п/п	Район	Загальний відсоток осіб, що виїжджають за межі області ( $X_1$ )	Відсоток осіб, що виїжджають в Росію ( $X_2$ )	Поширеність ВІЛ-інфекції (на 100 тис. населення) ( $Y$ )
1	Перечинський	3,68	0,58	9,4
2	В. Березнянський	13,9	4,43	14,9
3	Виноградівський	11,8	2,26	15,7
4	Воловецький	9,11	1,64	20,4
5	Хустський	23,79	8,96	59,6
6	Мукачівський	5,87	1,48	38,1
7	Берегівський	5,84	0,83	25
8	Тячівський	33,09	14,14	47,5
9	Рахівський	23,8	3,96	20,5
10	Свалявський	15,84	4,36	49,9
11	Іршавський	16,16	6,12	17
12	Міжгірський	6,43	2,53	10,4
13	Ужгородський	16,09	2,9	57,9

Аналогічно до Прикладу 1, розглянемо випадки при  $\tau = 8$ ,  $\tau = 10$ ,  $\tau = 13$  та проведемо дослідження за такою ж схемою.

Після виконання методу структурної ідентифікації в системі базисних функцій  $\{1, X_1, X_2, X_1^a, X_2^a, \sin(a \cdot X_1 + b \cdot X_2 + c)\}$ , ( $X_1, X_2 > 1$ ), було визначено характер залежності у вигляді:

$$Y(a_1, a_2, a_3, a_4, a_5) = a_1 \cdot X_1^{a_2} + a_3 \cdot X_2^{a_4} + a_5. \quad (25)$$

Результати ідентифікації при запуску еволюційного методу параметричної ідентифікації з верхнім значенням циклу 100 були такими:

для  $\tau = 8$ :

$$Y_{e8} = -762,14 \cdot X_1^{-3,6156} + 2,6867 \cdot X_2^{0,9655} + 19,2784; \quad (26)$$

для  $\tau = 10$ :

$$Y_{e10} = -717,684 \cdot X_1^{-3,718} + 6,2992 \cdot X_2^{0,6851} + 15,3361; \quad (27)$$

для  $\tau = 13$ :

$$Y_{e13} = 8,0512 \cdot X_1^{0,4967} - 0,307 \cdot X_2^{-5,7723} + 1,8873. \quad (28)$$

Рівняння множинної лінійної регресії при різних значеннях  $\tau$  запишуться так:

для  $\tau = 8$ :

$$Y_{r8} = -1,06367 \cdot X_1 + 4,865611 \cdot X_2 + 22,18877; \quad (29)$$

для  $\tau = 10$ :

$$Y_{r10} = -0,55066 \cdot X_1 + 3,813433 \cdot X_2 + 21,91884; \quad (30)$$



для  $\tau = 13$ :

$$Y_{r13} = 1,017841 \cdot X_1 + 0,29917 \cdot X_2 + 13,95233. \quad (31)$$

Результати обчислень значень функцій та похибок наведені в табл. 7.

**Таблиця 7**

Зведені значення результатів ідентифікації залежності поширеності ВІЛ-інфекції від трудової міграції різними методами

№ п/п	$\gamma$	$Y_{\text{вб}}(X)$	$Y_{r\text{в}}(X)$	$Y_{\text{г10}}(X)$	$Y_{r10}(X)$	$Y_{\text{г13}}(X)$	$Y_{r13}(X)$
1	9,4	14,00891	21,09651	14,02249143	22,1042	10,14226	17,8715
2	14,9	30,52859	28,95839	32,75942968	31,15814	31,64472	29,42564
3	15,7	25,0804	20,63373	26,27433242	24,03938	29,31696	26,63898
4	20,4	23,35132	20,47832	23,98227407	23,15634	25,99387	23,7155
5	59,6	41,58924	40,47991	43,62574843	42,98694	40,74846	40,84733
6	38,1	21,9338	23,14612	22,58046639	24,33033	21,24828	20,36983
7	25	20,23155	20,01538	19,86551877	21,86812	20,33092	20,14483
8	47,5	53,94789	55,79163	54,01198464	57,61936	47,66922	51,86294
9	20,5	29,41634	16,14121	31,50273029	23,91427	40,75646	39,36165
10	49,9	30,37719	26,55428	32,58544074	29,82292	33,63982	31,37931
11	17	34,69225	34,77738	37,10461913	36,35835	33,95688	32,23156
12	10,4	24,94979	27,65936	26,5242214	28,02607	22,17665	21,25395
13	57,9	26,75574	19,18457	28,37660266	24,11764	33,88716	31,19698
$\Delta_{\text{vт}}$		10,06103	10,20583	12,04421589	12,22874	14,77107	14,89368
$\delta_{\text{vт}}$		19,77533	23,13796	22,62607063	24,67571	—	—

**Примітка:**  $v \in \{q, r\}$ .

Результати обчислень коефіцієнтів Фішера наведено у табл. 8.

**Таблиця 8**

Розраховані значення коефіцієнтів Фішера для побудованих моделей при ідентифікації залежності поширеності ВІЛ-інфекції від трудової міграції

$\tau$	$p_0$	Табличне значення коефіцієнта Фішера ( $F_{\text{табл}}, p = p_0$ )	$F_{\text{вт}}$	$F_{r\tau}$
8	0,05	2,446912	2,918819	2,806785
10	0,05	2,306004	2,996092	2,838484
13	0,10	1,795885	2,171963	2,002883

Таким чином, ступінь адекватності моделей, в основі яких є еволюційний метод структурної і параметричної ідентифікації, за критерієм Фішера, перевищує ступінь адекватності лінійних регресійних моделей і забезпечує ідентифікацію з меншою середньоквадратичною похибкою.

Отже, розроблені еволюційні методи структурної і параметричної ідентифікації можуть ефективно використовуватися при розв'язуванні прикладних задач, пов'язаних з ідентифікацією табличних залежностей.

### 7. SWOT-аналіз результатів дослідження

В результаті проведеного дослідження моделей і методів структурної і параметричної ідентифікації табличних

залежностей були розроблені релевантні еволюційні методи. Виконана експериментальна верифікація розроблених методів, а також проведений порівняльний аналіз їх результатів з результатами застосування регресійних методів, довели їх переваги: шляхом застосування розроблених методів вдалося зменшити середньоквадратичну похибку, а також досягти більшого значення коефіцієнту Фішера в порівнянні з аналогічними показниками для регресійних методів. Також, було показано, що при використанні розроблених методів можливим є задання таких характеристик результуючої функції, як періодичність, парність, монотонність, область значень тощо, а також проведення уточнення параметрів результуючої функції без необхідності виконання складних математичних перетворень та розв'язування систем нелінійних рівнянь багатьох змінних.

Особливістю запропонованих еволюційних методів є те, що результати ідентифікації залежать від початково заданої системи лінійно незалежних базисних функцій, а також від кількості запусків алгоритмів еволюційних методів для одних і тих же вхідних даних з метою покращення критерію оптимальності, що ускладнює повну автоматизацію процесу структурної і параметричної ідентифікації.

Подальші дослідження еволюційних методів ідентифікації мають бути направлені на вивчення та розробку рекомендацій і правил щодо вибору системи базисних функцій для визначення структури результуючої функції, що має привести до зменшення трудомісткості проведення ідентифікації та підвищення її ефективності.

Розроблені еволюційні методи ідентифікації можуть показати дещо гірші з точки зору критерію оптимальності результати, в порівнянні з, наприклад, нейромережними методами ідентифікації, у випадках, коли початкові дані мають достатній для нейромережних методів об'єм, проте, останні не дозволяють отримати аналітичний вигляд результуючої залежності, що для деяких задач є значним недоліком.

### 8. Висновки

В ході дослідження було:

1. Виконано математичну постановку задачі структурної і параметричної ідентифікації табличних залежностей для випадку, коли метою ідентифікації є побудова функції, яка оптимально, за деяким критерієм екстраполює залежність. Відзначено, що у такому випадку задача може бути зведена до послідовного розв'язання двох задач: задачі структурної ідентифікації, в якій визначається структура функції, та задачі параметричної ідентифікації для налаштування параметрів функції.

2. Розроблено методи структурної і параметричної ідентифікації на основі еволюційної технології. Еволюційний метод структурної ідентифікації дозволяє побудувати функцію будь-якої структури із заданого числа лінійно незалежних базисних функцій шляхом дослідження похибок функцій на різних частинах навчальної вибірки. Метод параметричної ідентифікації призначений для налаштування параметрів функцій заданої структури відповідно до заданого критерію оптимальності: середньоквадратичної похибки, середньої відносної похибки тощо. Перевагою розроблених методів є те, що через задання системи базисних функцій для визначення структури функції, що відповідає вхідним

даним, виникає можливість контролювати такі її характеристики як парність, періодичність, монотонність, область значень тощо.

3. Виконано експериментальну верифікацію розроблених методів для розв'язання прикладних задач ідентифікації залежностей за допомогою однофакторного та двофакторного аналізу. В ході дослідження були побудовані функції, які відображають задані табличні залежності. Побудовані функції були адекватні за критерієм адекватності Фішера.

4. Здійснено порівняльний аналіз результатів роботи розроблених еволюційних методів ідентифікації з результатами застосування регресійних методів за двома показниками: значення середньоквадратичних похибок, а також значення коефіцієнта Фішера. Побудовані еволюційним методом структурної і параметричної ідентифікації моделі показали кращі результати, ніж лінійні регресійні моделі для всіх випадків, що розглядалися. Таким чином було доведено переваги розробленого еволюційного методу.

### Література

- Снитюк, В. Є. Прогнозування. Моделі. Методи. Алгоритми [Текст]: навчальний посібник / В. Є. Снитюк. — К.: Маклаут, 2008. — 364 с.
- Zhang, L. Forward and backward least angle regression for nonlinear system identification [Text] / L. Zhang, K. Li // Automatica. — 2015. — Vol. 53. — P. 94–102. doi:10.1016/j.automatica.2014.12.010
- Iturbide, E. A Comparison between LARS and LASSO for Initialising the Time-Series Forecasting Auto-Regressive Equations [Text] / E. Iturbide, J. Cerda, M. Graff // Procedia Technology. — 2013. — Vol. 7. — P. 282–288. doi:10.1016/j.protcy.2013.04.035
- Tutunji, T. A. Parametric system identification using neural networks [Text] / T. A. Tutunji // Applied Soft Computing. — 2016. — Vol. 47. — P. 251–261. doi:10.1016/j.asoc.2016.05.012
- Aguilar-Leal, O. Distributed parameter system identification using finite element differential neural networks [Text] / O. Aguilar-Leal, R. Q. Fuentes-Aguilar, I. Chairez, A. García-González, J. C. Huegel // Applied Soft Computing. — 2016. — Vol. 43. — P. 633–642. doi:10.1016/j.asoc.2016.01.004
- Loussifi, H. A new efficient hybrid intelligent method for nonlinear dynamical systems identification: The Wavelet Kernel Fuzzy Neural Network [Text] / H. Loussifi, K. Nouri, N. Benhadj Braiek // Communications in Nonlinear Science and Numerical Simulation. — 2016. — Vol. 32. — P. 10–30. doi:10.1016/j.cnsns.2015.08.010
- Jiang, X. Fuzzy stochastic neural network model for structural system identification [Electronic resource] / X. Jiang, S. Mahadevan, Y. Yuan // Mechanical Systems and Signal Processing. — 2016. — Available at: \www/URL: http://doi.org/10.1016/j.ymssp.2016.05.030
- Yan, J. NARMAX model identification using a set-theoretic evolutionary approach [Text] / J. Yan, Jr. J. R. Deller // Signal Processing. — 2016. — Vol. 123. — P. 30–41. doi:10.1016/j.sigpro.2015.12.001
- Ayala, H. V. H. Cascaded evolutionary algorithm for nonlinear system identification based on correlation functions and radial basis functions neural networks [Text] / H. V. H. Ayala, L. dos S. Coelho // Mechanical Systems and Signal Processing. — 2016. — Vol. 68–69. — P. 378–393. doi:10.1016/j.ymssp.2015.05.022
- Mulesa, O. Designing fuzzy expert methods of numeric evaluation of an object for the problems of forecasting [Text] / O. Mulesa, F. Geche // Eastern-European Journal of Enterprise Technologies. — 2016. — № 3/4(81). — P. 37–43. doi:10.15587/1729-4061.2016.70515
- Миронюк, І. С. Результати вивчення ролі трудової міграції у поширенні ВІЛ-інфекції в Закарпатті [Текст] / І. С. Миронюк, В. Й. Шатило // Україна. Здоров'я нації. — 2011. — № 1(17). — С. 58–62.
- Миронюк, І. С. Особливості ризикованої поведінки ВІЛ-інфікованих трудових мігрантів Закарпатської області залежно від регіону міграції [Текст] / І. С. Миронюк // Науковий вісник Ужгородського університету. Серія «Медицина». — 2012. — Вип. 1(43). — С. 146–151.

### РАЗРАБОТКА ЭВОЛЮЦИОННЫХ МЕТОДОВ СТРУКТУРНОЙ И ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ТАБЛИЧНЫХ ЗАВИСИМОСТЕЙ

Рассматривается задача структурной и параметрической идентификации табличных зависимостей. Разработан эволюционный метод структурной идентификации, который определяет оптимальную по заданному критерию структуру функции. Разработан эволюционный метод параметрической идентификации, который, на основе исследования характера входных данных, определяет параметры функции, её характеризующей. Выполнена экспериментальная верификация разработанных методов с помощью однофакторного и двухфакторного анализа.

**Ключевые слова:** структурная идентификация, параметрическая идентификация, табличная зависимость, эволюционный метод.

*Мулеса Оксана Юрьевна, кандидат технических наук, доцент, кафедра кибернетики і прикладної математики, ДВНЗ «Ужгородський національний університет», Україна,*  
**e-mail: mulesa.oksana@gmail.com.**

*Мулеса Оксана Юрьевна, кандидат технических наук, доцент, кафедра кибернетики и прикладной математики, ГВУЗ «Ужгородский национальный университет», Украина.*

**Mulesa Oksana, Uzhgorod National University, Ukraine,**  
**e-mail: mulesa.oksana@gmail.com**