

УДК 004.032.26(045)

## СИСТЕМА КЛАСТЕРНОГО АНАЛІЗУ РЕЗУЛЬТАТІВ НЕРУЙНІВНОГО КОНТРОЛЮ ВИРОБІВ ІЗ КОМПОЗИЦІЙНИХ МАТЕРІАЛІВ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ КОХОНЕНА

В. С. Єременко, А. В. Переїденко, В. О. Роганьков

Національний авіаційний університет

nau\_307@ukr.net

*Наведено загальну характеристику процедури кластерного аналізу даних. Подано результати дослідження різних функцій відстані як критерію про схожість об'єктів. Описано систему проведення кластерного аналізу і дослідження достовірності кластеризації із застосуванням алгоритмів на основі описаних мір близькості. Систему реалізовано в середовищі LabVIEW 8.5.*

*This article is devoted to realization system of the cluster analysis without etalon samples. Main different vector space metrics are analyzed using the special control system. System was created with NI LabVIEW 8.5.*

### Постановка проблеми

На сьогодні розробка системи розпізнавання (класифікатора) стану виробів із композиційних матеріалів для своєчасного виявлення пошкоджень є важливою та актуальною задачею. У реальних умовах проведення діагностичного експерименту на його результат істотно впливають випадкові складові похибки вимірювання інформаційних параметрів. Під час діагностики методом низькошвидкісного удару випадкові складові похибки вимірювання параметрів імпульсів сили ударної взаємодії визначаються не тільки шумами датчика сили, узгоджувального підсилювача і аналого-цифрового перетворювача, але і нестабільністю енергії ударної дії на об'єкт контролю, відхиленням вектора сили удару від нормалі до контрольованої поверхні, неоднорідністю структури та ступенем шорсткості поверхні виробу, а також іншими чинниками, врахувати вплив яких достатньо складно. За наявності мінімальної кількості інформації про образи, що розпізнаються, та обмеженої кількості образів для навчання, а також враховуючи, що однією зі складних для вирішення завдань є виготовлення спеціальних еталонних зразків з різними типами дефектів, притаманних контрольованому матеріалу, в роботі запропоновано вирішити поставлене завдання безеталонної діагностики технічного стану виробів із композитів на основі використання штучних нейронних мереж, які здатні проводити нелінійну кластеризацію і класифікацію, а також є гнучкими та здатними до розпізнавання за ознаками на основі сучасних методів обробки інформації.

### Основні цілі

Мета роботи — дослідження можливості проведення кластерного аналізу результатів неруйнівного контролю виробів із композиційних матеріалів на основі апарата штучних нейронних мереж. Такий підхід підвищує точність визначення стану об'єкта контролю, дає змогу зменшити вплив випадкових факторів на результати контролю і дозволяє провести якісний кластерний аналіз експериментальних даних в умовах об-

меженої початкової інформації про досліджуваній об'єкт контролю. В такому випадку зникає необхідність мати еталонні об'єкти. Зразки досліджуваних композиційних матеріалів подано на рис. 1.



Рис. 1. Зразки композиційних матеріалів

### Вирішення поставленого завдання

Особливість кластеризації полягає в тому, що класи об'єктів спочатку не відомі. Результатом кластеризації є розбиття об'єктів на групи, що задовольняють деякий критерій оптимальності [1]. Цей критерій може бути деяким функціоналом, що виражає рівні бажаності різних варіантів розбиття і об'єднання. На відміну від задач класифікації, кластерний аналіз не потребує апріорних припущень про набір даних, не накладає обмеження на представлення досліджуваних об'єктів, дає змогу аналізувати показники різних типів даних (інтервальні дані, частоти, бінарні дані). При цьому необхідно пам'ятати, що змінні повинні вимірюватися в конгруентних (порівнянних) шкалах.

Формально задача кластеризації формується таким чином. Нехай  $X$  — множина об'єктів;  $Y$  — множина номерів (імен, міток) кластерів. Задано функцію відстані між об'єктами  $\rho(x, x')$ , також є кінцева навчальна вибірка об'єктів  $X^m = \{x_1, \dots, x_m\} \in X$ . Потрібно розбити вибірку на непересічні підмножини (кластери) таким чином, щоб кожен кластер складався з об'єктів, близьких за метрикою  $\rho$ , а об'єкти різних клас-

терів істотно відрізнялися. При цьому кожному об'єкту  $x_i \in X^m$  приписується номер кластера  $y_i$ . Алгоритм кластеризації — це функція  $a: X \rightarrow Y$ , яка будь-якому об'єкту  $x \in X$  ставить у відповідність номер кластера  $y \in Y$ . Множина  $Y$  у деяких випадках відома заздалегідь, проте частіше потрібно визначити оптимальну кількість кластерів, з погляду одного або іншого критерію якості кластеризації.

Методи кластеризації розрізняються за правилами побудови кластерів. Як такі правила виступають критерії, що використовуються при вирішенні питання про «схожість» об'єктів. Критерієм для визначення схожості і відмінності кластерів є відстань між векторами на діаграмі розсіювання (рис. 2).

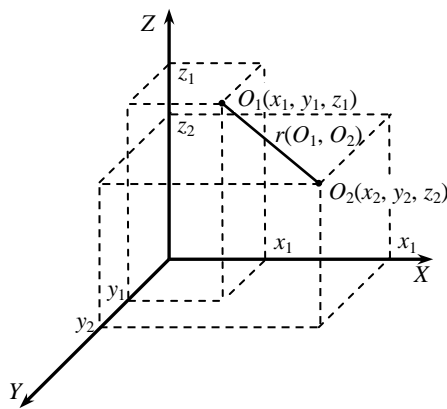


Рис. 2. Відстань між двома векторами в просторі

Для обчислення відстані між об'єктами використовуються різні міри схожості (міри подібності), які також називаються метриками або функціями відстаней [2]. Для визначення відстані між кластерами в реалізованій нейронній мережі використовуються декілька мір близькості.

1. Перша і найпоширеніша міра близькості — евклідова відстань між двома векторами  $X$  і  $Y$  у  $n$ -вимірному просторі з відомими координатами цих векторів  $x_i$  і  $y_i$ , де  $i = \overline{1, n}$ . Ця міра близькості просто є геометричною відстанню в багатовимірному просторі і обчислюється так:

$$\rho(X, Y) = \sqrt{\sum_i (x_i - y_i)^2}.$$

Слід зазначити, що евклідова відстань (та її квадрат) розраховується за початковими, а не за стандартизованими даними. Це звичайний спосіб його обчислення, який має певні переваги (наприклад, відстань між двома об'єктами не змінюється при введенні в аналіз нового об'єкта, який може виявитися викидом). Проте на відстані можуть сильно впливати відмінності між осями, по координатах яких розраховуються ці відстані.

2. Щоб додати великої ваги віддаленішим один від одного об'єктам було використано квад-

рат стандартної евклідової відстані. Ця відстань обчислюється за формулою:

$$\rho(X, Y) = \sum_i (x_i - y_i)^2.$$

3. Коли необхідно визначити два об'єкти як «різні», якщо вони відрізняються за якимось одним вимірюванням у розробленій мережі Кохонена, використовується відстань Чебишева. Відстань Чебишева обчислюється за формулою:

$$\rho(X, Y) = \max(|x_i - y_i|).$$

4. Для значного збільшення або зменшення ваги, що належать до розмірності, для якої відповідні об'єкти сильно відрізняються в мережі використовується степенева відстань:

$$\rho(X, Y) = \sqrt[r]{\sum_i |x_i - y_i|^p},$$

де  $r$  і  $p$  — параметри, що визначаються дослідником.

Параметр  $p$  відповідає за поступове зважування різниць за окремими координатами, параметр  $r$  відповідає за прогресивне зважування великих відстаней між об'єктами. Якщо параметри  $r$  і  $p$  дорівнюють двом, то ця відстань збігається з евклідовою відстанню.

5. У системі також реалізовано відстань Махаланобіса, яка обчислюється таким чином:

$$\rho(X, Y) = (X - Y)^T C^{-1} (X - Y),$$

де  $X, Y$  — вектори середніх значень змінних однієї та другої групи відповідно;  $C^{-1}$  — обернена коваріаційна групової матриця;  $()^T$  — оператор транспонування.

Для вирішення задачі попередньої кластеризації та подальшої класифікації дефектних і бездефектних ділянок досліджуваних зразків під час проведення неруйнівного контролю виробів з композиційних матеріалів було реалізовано нейронну мережу Кохонена. Нейронні мережі Кохонена — це клас нейронних мереж, основним елементом яких є шар Кохонена. Шар Кохонена складається з деякої кількості  $n$  адаптивних лінійних суматорів, що діють паралельно (лінійних формальних нейронів) [3]. Усі вони мають однакову кількість входів  $m$  і отримують на свої входи один і той же вектор входних сигналів  $x = (x_1, \dots, x_m)$ . На виході  $j$ -го лінійного елемента отримуємо сигнал:

$$y_j = w_{j0} + \sum_{i=1}^m w_{ji} x_i,$$

де  $w_{ji}$  — ваговий коефіцієнт  $i$ -го входу  $j$ -го нейрона;  $w_{j0}$  — пороговий коефіцієнт.

Після проходження шару лінійних елементів сигнали посилаються на обробку за правилом «переможець забирає все»: серед вихідних сигналів  $y_j$  обирається максимальний, за номером  $j_{\max} = \arg \max_j \{y_j\}$ .

Остаточо, на виході сигнал з номером  $j_{\max}$  дорівнює одиниці, інші — нулю. Якщо максимум одночасно досягається для декількох  $j_{\max}$ , то

обирають усі відповідні сигнали рівними одиницями або тільки перший у списку.

Залежно від способу налаштування вхідних ваг суматорів і за розв'язуваними задачами в роботі розглянуто такі різновиди мереж Кохонена:

— мережі векторного квантування сигналів, що тісно пов'язані з простим базовим алгоритмом кластерного аналізу (метод динамічних ядер або  $K$ -середніх, тобто  $K$ -means);

— карти Кохонена, що самоорганізуються (*Self-Organising Maps, SOM*).

Робота нейронної мережі побудована таким чином, що вектори вхідних сигналів  $X^p$  обробляються по черзі, для кожного з них знаходиться найближчий вектор («переможець», який «забирає все»)  $W_{j(x)}$ . Після цього ваги вектора-переможця перераховуються за формулою:

$$W_i^{new} = W_i^{old} + \eta \rho(X^p, W_i^{old}),$$

де  $\eta \in (0,1)$  — темп навчання;  $W_i^{old}$ ,  $W_i^{new}$  — ваги нейрона-переможця до і після модифікації відповідно;  $\rho(X^p, W_i^{old})$  — функція відстані між вхідним вектором та ядром кластера.

Дані, що надходять на вхід нейронної мережі, мають бути правильно підготовлені для подальшої їх обробки. Один з поширених способів — масштабування.

У роботі попередня обробка вхідних даних виконувалася за таким алгоритмом:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}},$$

де  $x_i^*$  — нормоване значення ознаки;  $x_i$  — дійсне значення ознаки;  $x_{\min}$  та  $x_{\max}$  — мінімальне та максимальне значення ознаки відповідно.

Як експериментальні дані для дослідження роботи нейронних мереж було використано дані, отримані під час проведення контролю зразків композиційних матеріалів методом низькошвидкісного удару [4].

Досліджуваний зразок мав п'ять характерних зон — бездефектну і чотири зони з різним ступенем пошкодження (дефекту). Інформативними параметрами для аналізу були амплітуда та тривалість імпульсу одержаного сигналу.

У результаті роботи нейронної мережі Кохонена було проведено кластерний аналіз експериментальних даних із застосуванням алгоритмів визначення функцій відстані на основі різних мір близькості, знайдено центри скупчення точок (центри кластерів) у двовимірному просторі для кожної із зон (рис. 3), що характеризують дефектні та бездефектні ділянки досліджуваного зразка.

Результати проведення кластерного аналізу за допомогою штучної мережі Кохонена з використанням алгоритмів пошуку відстаней між векторами і відповідними кластерами на основі описаних мір близькості можна навести у вигляді таблиці.

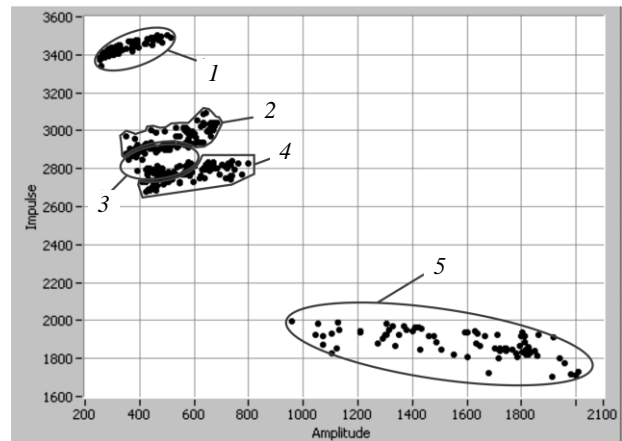


Рис. 3. Розміщення векторів, що характеризують різні ділянки зразка композиту:  
1—4 — ділянки з різним ступенем дефекту;  
5 — бездефектна ділянка

У таблицю занесено достовірність приєднання вектора до певного кластеру (ділянки). У таблиці наведено результати застосування алгоритмів кластеризації на основі тих мір близькості, застосування яких дає змогу отримати достовірність кластеризації понад 85 %.

#### Достовірність кластерного аналізу із застосуванням нейронної мережі Кохонена

Тип ділянки	Міри близькості				
	Чебишева	Махаланобіса	Евкліда	Степенева	Квадрат Евкліда
Без дефекту	0,95	0,95	0,95	0,95	0,95
Дефект 1	0,95	0,67	0,84	0,95	0,90
Дефект 2	0,89	0,95	0,67	0,84	0,83
Дефект 3	0,80	0,95	0,88	0,77	0,62
Дефект 4	0,95	0,95	0,95	0,95	0,95
Загальна точність	0,92	0,91	0,87	0,87	0,87

На рис. 4 зображено достовірність віднесення об'єкта до кластера із застосуванням різних мір близькості. Як видно з рис. 4 найкраща достовірність проведення кластерного аналізу досягається із застосуванням міри близькості Чебишева. Високу достовірність також можна отримати із застосуванням міри близькості Махаланобіса та Евкліда. Під час розв'язання задачі класифікації стану стільникових панелей із застосуванням нейронної мережі Кохонена достовірність віднесення об'єкта до певного класу становила 92 %. Тому її використання для розв'язання задач є доцільним.

Інтерфейс і програмний код системи кластерного аналізу з використанням нейронних мереж Кохонена представлено на рис. 5 і 6 відповідно.

Систему було розроблено в програмному пакеті NI LabVIEW 8.5.

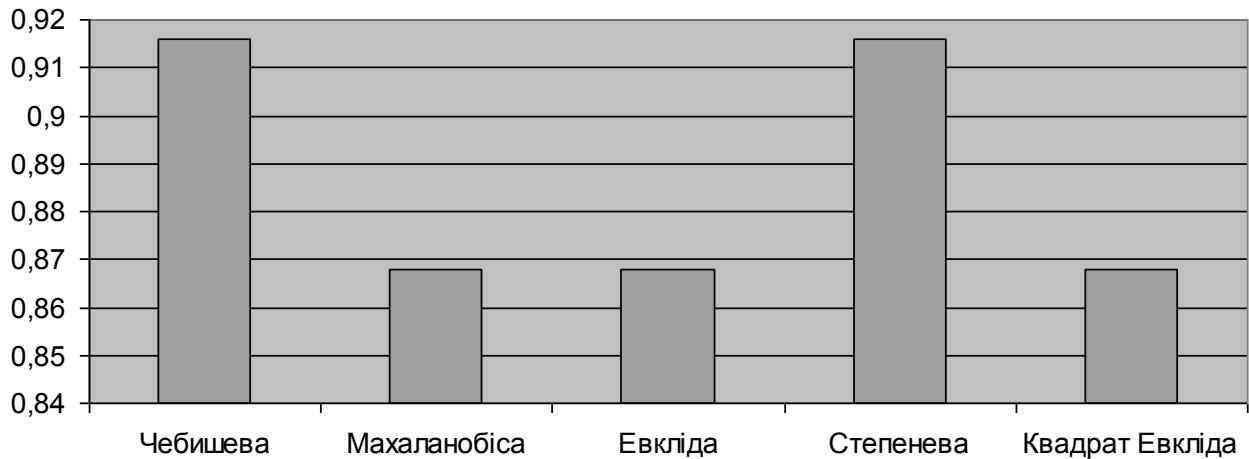


Рис. 4. Достовірність кластерного аналізу із застосуванням нейронної мережі Кохонена

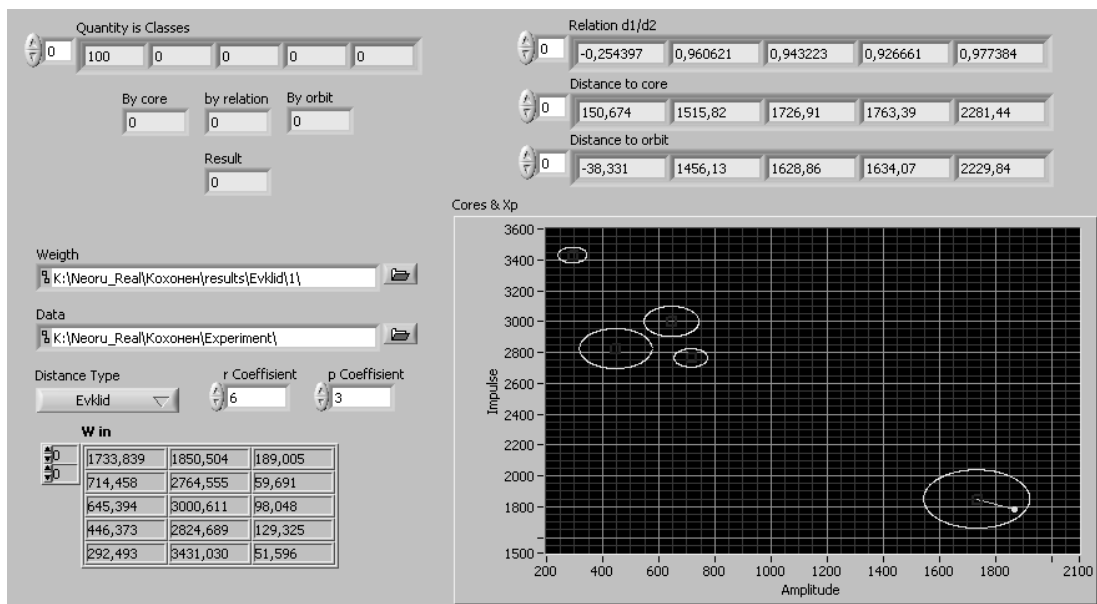


Рис. 5. Процес навчання багатшарового персеプトна

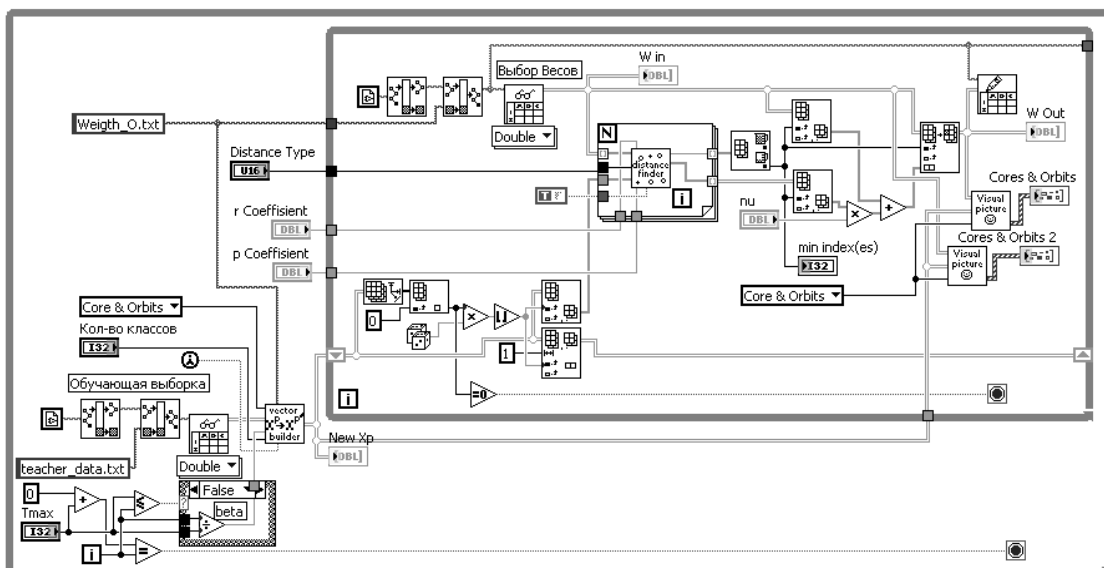


Рис. 6. Процес роботи нейронної мережі Кохонена

### Висновки

На основі отриманих результатів можна зазначити, що для вирішення задачі безеталонної дефектоскопії без попереднього навчання на еталонних зразках для знаходження відстані між вектором, що характеризує властивості об'єкта контролю, і центром відповідного кластера найбільш доцільно використовувати міри близькості Чебишева, Махаланобіса або Евкліда.

Застосування алгоритмів пошуку відстаней на основі даних мір близькості дає змогу отримати достовірність віднесення вектору до необхідного кластера 94, 93 і 88 % відповідно.

Високий ступінь універсальності системи досягається завдяки модульній структурі підсистеми аналізу та обробки даних, яка дає змогу без ускладнень додавати необхідні або усувати непотрібні методики аналізу сигналу.

Принципи, закладені в систему на етапі проектування, дозволяють виконувати її модернізацію і адаптацію до зміни умов роботи без внесення значних змін до структури програмного забезпечення.

Наведена система успішно застосовується в навчальному і дослідницькому процесі з тематики неруйнівного контролю в науково-дослідній лабораторії на кафедрі інформаційно-вимірвальних систем Національного авіаційного університету.

### ЛІТЕРАТУРА

1. Дюран Б. Кластерный анализ / Б. Дюран, П. Оделл : пер. с англ. Е.З. Демиденко; под ред. А. Я. Боярского. — М. : Статистика, 1977. — 128 с.
2. Скворцов В. А. Примеры математических пространств / В. А. Скворцов. — М. : МЦНМО, 2002. — 24 с.
3. Хайкин Саймон Нейронные сети: полный курс, 2-е издание : пер. с англ. / Саймон Хайкин. — М. : Издательский дом «Вильямс», 2006. — 1104 с.
4. Еременко В.С. Обнаружение ударных повреждений сотовых панелей методом низкоскоростного удара / В.С. Еременко, В.М. Мокійчук, А.М. Овсянкин // Техническая диагностика и неразрушающий контроль. — К., 2007. — № 1. — С. 24—27.

Стаття надійшла до редакції 01.12.09.