



Судово-експертна ідентифікація об'єктів за їх багатовимірними ознаками на підставі критерію косинусної подібності та методів кластерного аналізу

Сергій Чорний

кандидат технічних наук, доцент, провідний науковий співробітник лабораторії інженерно-технічних досліджень, Національний науковий центр «Інститут судових експертиз ім. Засл. проф. М. С. Бокаріуса», м. Харків, Україна, ORCID: <http://orcid.org/0000-0002-0445-2886>, e-mail: svch1mail@yahoo.com

Олександр Смирнов

кандидат технічних наук, старший дослідник, завідувач сектору лабораторії криміналістичних та військових досліджень, Національний науковий центр «Інститут судових експертиз ім. Засл. проф. М. С. Бокаріуса», м. Харків, Україна, ORCID: <http://orcid.org/0000-0001-6398-2855>, e-mail: smirnov_alexa@ukr.net

Сергій Омельник

реставратор вищої кваліфікаційної категорії творів з металу, завідувач, Національний науково-дослідний реставраційний центр України, Харківська філія, м. Харків, Україна, e-mail: osv110267@gmail.com

Розглянуто підхід до судово-експертної ідентифікації об'єктів на основі їх багатовимірних ознак із використанням критерію косинусної подібності та методів кластерного аналізу.

Ключові слова: косинусна подібність; кластерний аналіз; вектор ознак.

Forensic Identification of Objects by their Multidimensional Features Based on the Cosine Similarity Criterion and Cluster Analysis Methods

Sergiy Chorny, Oleksandr Smyrnov, Sergiy Omelnyk

An approach to forensic identification of objects based on their multidimensional features using the cosine similarity criterion and cluster analysis methods is considered.

Keywords: cosine similarity; cluster analysis; feature vector.

Сучасна судово-експертна діяльність дедалі частіше стикається з необхідністю аналізу складних, багатовимірних даних. Традиційні підходи, засновані на порівнянні окремих ознак або порогових значень, не завжди забезпечують достатню точність та обґрунтованість висновків, особливо у випадках, коли об'єкти характеризуються великою кількістю взаємопов'язаних параметрів. Прикладами таких об'єктів є: хімічний склад металевих, полімерних або мінеральних зразків; вектори ознак зображень, отримані за допомогою нейронних мереж; спектральні або тембральні ознаки аудіозаписів; багатоканальні сигнали з технічних пристроїв. У цих випадках природним є подання кожного об'єкта у вигляді вектора ознак у багатовимірному просторі. Завдання ідентифікації зводиться до оцінювання подібності між дослідним об'єктом та відомими зразками з бази даних.

Метою роботи є дослідження методу ідентифікації об'єктів на основі критерію подібності векторів у поєднанні з сучасними методами кластерного аналізу.

Припустимо, що маємо дослідний об'єкт, для якого виміряний вектор ознак та базу даних еталонних об'єктів де кожен вектор має таку саму розмірність. Метою судово-експертної ідентифікації є: визначення ступеня подібності між вектором об'єкта та кожним вектором з бази даних, встановлення найбільш імовірного класу або групи, до якої належить дослідний об'єкт, оцінка узгодженості результатів за різними методами аналізу.

Для порівняння двох багатовимірних векторів у багатовимірному нормованому просторі точним математичним рішенням є обчислення кута між ними на основі їх скалярного добутку, але достатньо обчислити лише косинус цього кута. Це і є у даному

випадку критерієм косинусної подібності векторів [1]. Значення косинусної подібності лежить у межах від -1 до 1 , однак у більшості прикладних експертних задач (де всі ознаки невід'ємні) — від 0 до 1 . Косинусна подібність має низку властивостей, які роблять її придатною для судово-експертної ідентифікації: інваріантність до масштабу (важливим є співвідношення ознак, а не їх абсолютні значення), стійкість до варіацій концентрацій, інтерпретованість (значення близькі до 1 вказують на високу структурну подібність).

Оцінка подібності дає матрицю парних відстаней. Проте для формування експертного висновку важливо виявити групи взаємно подібних об'єктів, оцінити стабільність належності дослідного зразка до певної групи, порівняти результати за різними методами. Саме ці завдання розв'язуються методами кластерного аналізу.

Станом на сьогодні існує широкий вибір програмно реалізованих методів кластерного аналізу. Найбільш зручним, на думку авторів, виглядає пакет `scikit-learn` [2], у якому реалізовано наступні методи кластерного аналізу: `MiniBatch KMeans`, `Affinity Propagation (Af)`, `MeanShift`, `Spectral Clustering`, `Ward`, `Agglomerative Clustering (Ag)`, `DBSCAN`, `HDBSCAN`, `OPTICS`, `BIRCH`, `Gaussian Mixture Models (G)`. Як практичний приклад застосування пропонованого підходу розглянемо ідентифікацію металевих зразків історичного значення які були елементами важливих будівель міста Харкова (Палац культури «Залізничник», Держпром), що зазнали руйнувань під час поточної війни росії проти України. Необхідність точної ідентифікації цих зразків обумовлена включенням цих будівель до переліку об'єктів посиленого захисту ЮНЕСКО [3], що пов'язано з категоричними

вимогами щодо реставрації їх відповідними аутентичними будівельними матеріалами. Отже, маємо 5 дослідних металевих зразків (див. табл. 1), для яких виміряні вектори їх хімічного складу (концентрації металів: Fe, C, Mn, Cr, Ni, Mo тощо). Для ідентифікації цих зразків використано базу даних з 2678 сплавів [4]. Для кожного зразка обчислено максимальне значення косинусної подібності з базою даних, після чого результати було проаналізовано означеними методами кластеризації (див. табл. 2). Обчислення проведено за допомогою розробленої авторами комп'ютерної програми ідентифікації на мові Python використанням відповідних бібліотек `scikit-learn` [2]. Однак у табл. 2 відображені результати кластеризації (номери кластерів) лише тих методів (Af, W, Ag, G), які узгоджуються з оцінками отриманими по максимальних значеннях косинусної подібності (містять найближчі аналоги).

Таблиця 1

Узагальнені результати ідентифікації

№ з/п	Назва об'єкта	Походження об'єкта
1	Кутова накладка віконної рами	Держпром
2	Кутова накладка віконної рами	Палац культури «Залізничник»
3	Завіс віконний	Держпром
4	Елемент вітража	Держпром
5	Завіс квартирки	Електрична підстанція

Отже, косинусна подібність у поєднанні з методами кластерного аналізу є певною мірою універсальним (щодо природи об'єктів) ефективним інструментом судово-експертної ідентифікації групової належності об'єктів за їх багатовимірними ознаками.

Таблиця 2

Узагальнені результати ідентифікації об'єктів

№ з/п	Найближчий аналог за максимумом критерію з бази даних [4]	max cos	Метод кластеризації				Клас за [5]	Міць, МПа за [5]
			Af	W	Ag	G		
1	EN 1.5507(23MnB3), 1.5535 23MnB4	0.96	26	21	2	33	Конструкційна сталь (MnB)	640
2	SAE-AISI 1211 (G12110), EN 1.0314 (C2C)	0.97	27	26	22	13	Нелегована сталь	450



Завершення таблиці 2

№ з/п	Найближчий аналог за максимумом критерію з бази даних [4]	max cos	Метод кластеризації				Клас за [5]	Міць, МПа за [5]
			Af	W	Ag	G		
3	EN 1.0255 (P235TR2)	0.92	27	26	22	13	Нелегована сталь	500
4	EN 1.0314 (C2C)	0.96	17	23	46	31	Нелегована сталь	450
5	EN 1.6228 (15NiMn6)	0.98	11	15	32	19	Конструкційна сталь (NiMn)	640

Перелік джерел посилання

1. Cosine similarity. Scikit learn: web. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_similarity.html (дата звернення: 31.01.2026).
2. Sklearn cluster. Там само. URL: <https://scikit-learn.org/stable/api/sklearn.cluster.html> (дата звернення: 31.01.2026).
3. Держпром / Вікіпедія : вебсайт. URL: <https://uk.wikipedia.org/wiki/%D0%94%D0%B5%D1%80%D0%B6%D0%BF%D1%80%D0%BE%D0%BC> (дата звернення: 31.01.2026).
4. Alloy Dataset. Kaggle : web. URL: <https://www.kaggle.com/datasets/sohamumbare/alloy-dataset> (дата звернення: 31.01.2026).
5. Comparison of materials. Aloro : web. URL: <https://aloro.org/grades/de> (дата звернення: 31.01.2026).