

ПОРІВНЯННЯ АЛГОРИТМІВ КЛАСТЕРИЗАЦІЇ ДЛЯ ЗАДАЧ  
РОЗПІЗНАВАННЯ ГОЛОСУ

Вінницький національний технічний університет

**Анотація**

У роботі розглядаються питання порівняння алгоритмів кластеризації для задачі розпізнавання голосів. Проаналізовано ефективність різних підходів, таких як агломеративна ієрархічна кластеризація, спектральна кластеризація та методи на основі глибокого навчання. Визначено перспективи використання цих методів у реальних умовах та їх вплив на точність розпізнавання голосів.

**Ключові слова:** кластеризація, розпізнавання голосів, штучний інтелект, глибоке навчання, агломеративна ієрархічна кластеризація, спектральна кластеризація.

**Abstract**

The paper considers the comparison of clustering algorithms for speaker recognition tasks. The efficiency of different approaches, such as agglomerative hierarchical clustering, spectral clustering, and deep learning-based methods, is analyzed. The prospects of using these methods in real-world conditions and their impact on the accuracy of speaker recognition are determined.

**Keywords:** clustering, speaker recognition, artificial intelligence, deep learning, agglomerative hierarchical clustering, spectral clustering.

**Вступ**

Останніми роками технології здійснили значний прорив у галузі розпізнавання голосів. Кластеризація є важливим етапом цього процесу, оскільки дозволяє визначати окремих мовців у записах. Існує кілька основних підходів до кластеризації даних, серед яких агломеративна ієрархічна кластеризація, спектральна кластеризація та методи на основі глибокого навчання. Метою даної роботи є порівняння цих методів з точки зору точності та ефективності для задачі розпізнавання голосів.

**Основна частина**

Для порівняння алгоритмів кластеризації для задачі розпізнавання голосів використовувалися три основні методи: агломеративна ієрархічна кластеризація (АНС), спектральна кластеризація та методи на основі глибокого навчання (DL). Опис та деталі кожного з методів наведено нижче.

Агломеративна Ієрархічна Кластеризація (АНС) є класичним підходом до кластеризації, який починається з кожного сегмента як окремого кластера і поступово з'єднує найбільш подібні кластери до тих пір, поки не буде досягнуто певного критерію зупинки. Для визначення подібності між кластерами використовувався Байєсів інформаційний критерій (BIC). Формула для BIC:

$$BIC = \log(L(X)) - \frac{1}{2}P \log(N),$$

де  $L(X)$  — максимальна ймовірність даних  $X$  для моделі,  $P$  — число параметрів моделі,  $N$  — кількість даних.

Спектральна кластеризація базується на використанні спектральної інформації з матриці суміжності (афінності) для поділу даних на кластери. Основний етап включає розкладання матриці афінності на власні вектори і використання цих векторів для подальшої кластеризації. Основні кроки:

- побудова матриці афінності  $A$ , де  $A_{ij}$  — міра подібності між  $i$ -тим і  $j$ -тим сегментами;
- обчислення власних векторів і власних значень для матриці Лапласа  $L = D - A$ , де  $D$  — діагональна матриця ступенів;
- вибір  $k$  найбільших власних векторів і застосування методу  $k$ -середніх для кластеризації.

Методи на основі глибокого навчання використовують нейронні мережі для отримання високорозмірних векторів-ознак (наприклад,  $X$ -векторів), які потім кластеризуються за допомогою традиційних методів, таких як  $k$ -середніх. Ці методи показують високу точність завдяки здатності навчатися складним патернам в даних. Основні кроки:

- формування  $X$ -векторів;
- використання нейронної мережі для отримання векторів-ознак для кожного сегмента;
- використання методу  $k$ -середніх для кластеризації отриманих векторів-ознак.

Для порівняння алгоритмів була проведена серія експериментів на датасеті, що містить записи голосів різних мовців. Оцінка результатів здійснювалася за допомогою показника помилки діаризації (DER):

$$DER = \frac{False\ Alarm + Missed\ Detection + Confusion}{Total\ Time}$$

За результатами дослідження, методи на основі глибокого навчання показали найкращу точність, знижуючи DER на 30% порівняно з АНС та зі спектральною кластеризацією.

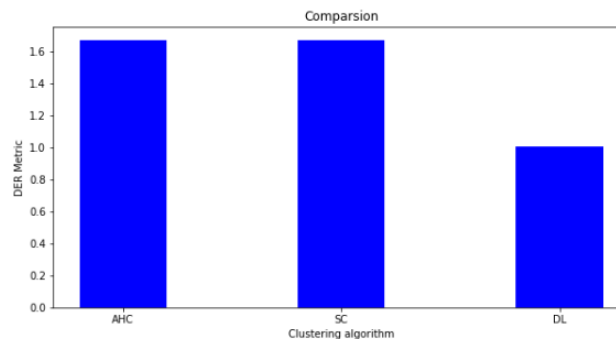


Рисунок 1 — Порівняння DER метрики для різних алгоритмів кластеризації

Це зумовлено здатністю глибоких моделей краще захоплювати складні патерни в голосових даних. Спектральна кластеризація продемонструвала добрі результати за умови наявності довгих сегментів, але її ефективність знижувалася при коротких сегментах. АНС показала непогані результати, завдяки своїй простоті та низьким вимогам до обчислювальних ресурсів.

## Висновок

У підсумку, методи на основі глибокого навчання є найкращими для задачі розпізнавання голосів завдяки їх високій точності та здатності ефективно обробляти великі обсяги даних. Спектральна

кластеризація також є потужним інструментом, особливо для довших сегментів, тоді як агрегативна ієрархічна кластеризація залишається корисною для менш вимогливих застосувань. Подальші дослідження можуть бути спрямовані на комбінування цих методів для досягнення оптимальної продуктивності.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Reynolds, D.A., Quatieri, T.F., Dunn, R.B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10(1-3), 19-41.
2. Shi, J., Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888-905.
3. Snyder, D., Garcia-Romero, D., Povey, D., Khudanpur, S. (2018). X-vectors: Robust DNN embeddings for speaker recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5329-5333.
4. Zhou, B., Hansen, J.H.L. (2000). Unsupervised speaker change detection using probabilistic models. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 423-426.
5. Kinnunen, T., Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*, 52(1), 12-40.

*Дар'я Вікторівна Водолазька* – студентка групи ІКІ-23м, факультет інформаційних технологій і комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, [dvodolazska@gmail.com](mailto:dvodolazska@gmail.com).

*Крупельницький Леонід Віталійович* – канд. техн. наук, доцент, доцент кафедри обчислювальної техніки, Вінницький національний технічний університет, м. Вінниця, [krupost@gmail.com](mailto:krupost@gmail.com).

*Vodolazska Daria V.* – student of group ІСІ – 23m, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, [dvodolazska@gmail.com](mailto:dvodolazska@gmail.com).

*Krupelnitskyi, Leonid V.* – Cand. Sc. (Eng.), Assistant Professor of the Computer Techniques Chair, Vinnytsia National Technical University, Vinnytsia, [krupost@gmail.com](mailto:krupost@gmail.com).