

## ЗАСОБИ ОПТИМІЗАЦІЇ КІЛЬКОСТІ КЛАСТЕРІВ ПРИ КЛАСТЕРИЗАЦІЇ МРТ-ЗНІМКІВ РАКОВИХ ПУХЛИН МОЗКУ ЗА ДОПОМОГОЮ АЛГОРИТМУ K-MEANS

Місоченко С.Ю.

Науковий керівник – к.т.н. доцент Селіванова К.Г.

Харківський національний університет радіоелектроніки, каф. БМІ  
м. Харків, Україна

e-mail: [serhii.misochenko@nure.ua](mailto:serhii.misochenko@nure.ua)

In this paper, the working principle of the medical image-clustering algorithm for image medical images has been reviewed, and ways to overcome its shortcomings have been proposed. As these ways we analyzed the application of elbow method and silhouette method to find the optimal number of clusters. The accuracy of these methods has been evaluated, and the peculiarities of their application to improve the quality of image segmentation by k-means method have been assessed.

Методи кластеризації використовують необроблені дані для формування кластерів на основі спільних факторів між різними точками даних. Оскільки цифрове зображення представляє собою масив однорідних числових даних, кластеризація представляє собою один з доволі популярних методів сегментації зображення [1]. Найбільш популярним методом кластеризації зображення є метод К-середніх (k-means). Основним недоліком кластеризації є те, що оптимальна кількість кластерів для найбільш якісної сегментації зображення невідома на початку. Для визначення оптимальної кількості кластерів будуть використовуватися метод ліктьової кривої і метод силуету [2].

Кількість кластерів методом k-середніх найчастіше оцінюється за допомогою «методу ліктя». Він передбачає багаторазове циклічне виконання алгоритму зі збільшенням кількості кластерів, що обираються, а також подальшим відкладанням на графіку балів кластеризації, обчисленого як функція від кількості кластерів. Основним показником «методу ліктя» є SSE (сума квадратів помилок).

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2, \quad (1.1)$$

де  $C_i$  –  $i$ -й кластер;

$p_i$  – точка вибірки у  $C_i$ ;

$m_i$  – центр тяжкості  $C_i$  (середнє значення для всіх вибірок у  $C_i$ );

$SSE$  – помилка кластеризації всіх вибірок.

Саме SSE відображає якість кластеризації. Основна ідея «методу ліктя» полягає в тому, що в міру збільшення числа кластерів  $k$ , розподіл вибірок

буде більш точним, а ступінь агрегації кожного кластера поступово збільшуватиметься, тому квадрат помилок і SSE стануть меншими.

Метод силуету – це також метод пошуку оптимальної кількості кластерів, інтерпретації та перевірки узгодженості усередині кластерів даних. Обчислюються коефіцієнти силуету кожної точки, які вимірюють, наскільки точка схожа одного кластеру проти іншими кластерами. Шляхом надання короткого графічного уявлення у тому, наскільки добре кожен об'єкт було класифіковано. Було проведено почерговий пошук оптимальної кількості кластерів двома методами для декількох зразків (рис. 1):

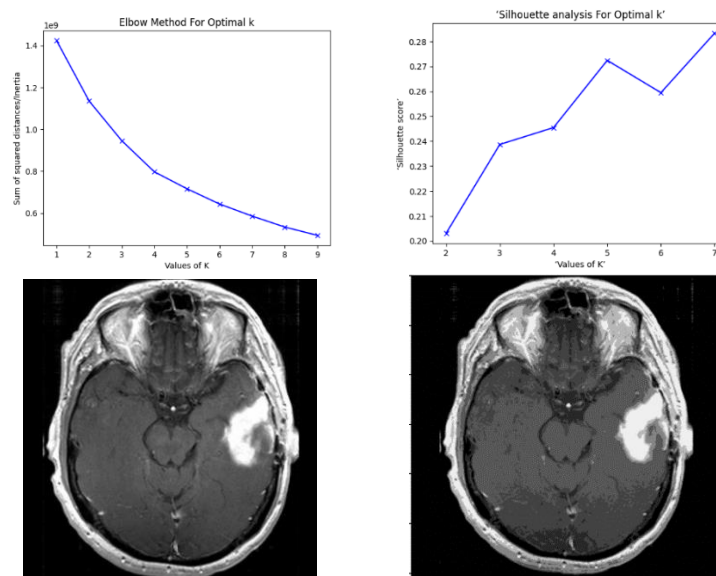


Рисунок 1 – Результати роботи алгоритмів методу ліктя та методу силуету, початкове та кластеризоване зображення (при  $K = 7$ )

Виходячи з отриманих результатів, «метод ліктя» підходить для більш контрастних зображень, метод силуету доповнює метод ліктя у випадку, якщо через «пологість спуску» графіка «лікоть» важко визначити.

#### Список використаних джерел:

1. Місоченко С. Ю. Дослідження використання вірогіднісних методів у сфері обробки біомедичних зображень / С. Ю. Місоченко, К. Г. Селіванова, О. Г. Аврунін // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: тези доповідей XXX міжнародної науково-практичної конференції MicroCAD-2022, 19-21 жовтня 2022 р. – Харків : НТУ «ХПІ», 2022. – С. 902.

2. Tymkovych, M. *et al.* (2021). Application of SOFA Framework for Physics-Based Simulation of Deformable Human Anatomy of Nasal Cavity. In: Jarm, T., Cvetkoska, A., Mahnič-Kalamiza, S., Miklavcic, D. (eds) 8th European Medical and Biological Engineering Conference. EMBEC 2020. IFMBE Proceedings, vol 80. Springer, Cham. [https://doi.org/10.1007/978-3-030-64610-3\\_14](https://doi.org/10.1007/978-3-030-64610-3_14)